BMC Genomics

RESEARCH

Open Access

Chromosome-level comparative genomics and host-specific fungal transcriptomics uncover adaptive virulence strategies in the sugarcane smut pathogen

Pedro Fernando Vilanova[1], Lucas Mitsuo Taniguti[2], Marcella Ferreira[1,4], Thaís Carolina da Silva Dal'Sasso[1], Gustavo Schiavone Crestana[1], Lâina da Silva de Oliveira[1], Renato Gustavo Hoffmann Bombardelli[1], Tatiana Caroline Pisetta[5], Paula Turini[5], João Paulo Kitajima[2], Silvana Creste[3], Luis Eduardo Aranha Camargo[4], Marie-Anne Van Sluys[5] and Claudia Barros Monteiro-Vitorello[1*]

Abstract

Background *Sporisorium scitamineum* is the causal agent of sugarcane smut, affecting global sugarcane production. Despite advances in smut genomics, the relationships between fungal genetic diversity, host adaptation, and virulence remain elusive.

Methods We applied chromosome-level genome sequencing (Oxford Nanopore and Illumina technologies) of two haploid strains (*MAT-1* x *MAT-2*) per isolate and high-depth transcriptomic profiling (Illumina) during early infection (48 hpi) of a more virulent isolate (SSC04) in resistant (SP80-3280) and susceptible (IAC66-6) sugarcane genotypes.

Results Despite the overall genomic similarity (99.9%), we identified nine highly polymorphic genomic islands (HPIs). The most variant HPI harbors the mating-type loci, where dense sequence variation, intrachromosomal rearrangements, and inversions, potentially linked to transposable element remnants, were observed. Additionally, the genome-wide analysis revealed non-synonymous single-nucleotide variants (SNVs) in 160 genes, including those involved in vesicular trafficking and candidate-secreted effectors. Transcriptomic profiling of the more virulent isolated revealed host-dependent transcriptional reprogramming in response to immune and metabolic cues, driving distinct infection strategies: in resistant plants, the fungus upregulated genes associated with detoxification, nitrogen starvation responses, and cell wall-degrading enzymes, while in susceptible hosts, it induced genes related to hyphal growth, lipid catabolism, and the unfolded protein response. The repertoire of expressed candidate effector genes also varied according to host and isolate genotypes.

Conclusions These findings uncover genomic signatures and context-dependent transcriptional regulation shaping the adaptive landscape of *S. scitamineum* virulence, identifying targets for pathogen monitoring and breeding for resistance.

*Correspondence:
Claudia Barros Monteiro-Vitorello
cbmontei@usp.br

Full list of author information is available at the end of the article

Keywords Virulence, Mating-Type, Effectors, Chromosome-level sequencing, *Sporisorium scitamineum*

Background

Ustilaginaceae fungi are primarily known for their association with smut diseases, affecting several host plants, including grasses, and dicotyledonous plants [1, 2]. The causal agent of sugarcane smut is the biotrophic basidiomycete *Sporisorium scitamineum* (Syd.) Piepenbr. & Oberw. 2002 [3]. The disease holds economic significance, causing substantial yield losses ranging from 12 to 75% and, in severe cases, leading to total crop failure [4, 5]. Fungal colonization induces the development of a whip-like structure in the plant's apical and lateral meristems responsible for the sporogenesis and spread of teliospores [6]. The disease severely stunts the affected plants and results in slender canes with smaller narrow leaves and impacts sucrose content, juice quantity and quality, affecting productivity [4, 7].

Genomic studies have provided insights into virulence factors, the influence of secondary metabolism, host-pathogen genotypes interactions, and adaptive evolution of Ustilaginaceae members [5, 8–12]. For most smut species, genome sequences are available for a limited number of strains. However, *Ustilago maydis* stands out as a model organism with extensive genomic data [13, 14], alongside *Ustilago hordei* [15], providing insights into genetic processes. So far, complete genome sequences of six *S. scitamineum* strains are available in the NCBI database [12, 16–18], but only the genome reported by Taniguti et al. (2015) [17] is assembled at the chromosome level.

Genome rearrangement events have been reported in smut species both before and after speciation [9, 12, 16]. These rearrangements are thought to play a key role in the evolution of mating systems, contributing to the transition between tetrapolar and bipolar organizations observed in different smut fungi [19–22]. Most of the genetic diversity of *S. scitamineum* was reported for Asian isolates, including race descriptions and pathogenic variants [23–28]. Genetic variation was described for Australian isolates [29] and a few reports from Brazil described pathogenic variation [30, 31]. Although not yet thoroughly explored, understanding the relationship between fungal genetic variability and pathogenicity infecting different sugarcane genotypes is critical for improving the selection of resistant varieties [4, 11, 32].

Previous studies have explored smut disease biology and global gene expression, primarily focusing on the plant's response to infection [17, 32–36], with a few exceptions focusing on analyzing the fungal transcriptome through RNA-seq [17, 37]. Dual-transcriptome analysis often faces challenges due to sequencing depth limitations, as fungal gene expression is typically detected at lower levels when compared to host expression [38].

Here, we sequenced, assembled and compared the chromosome-level genomes of mating-compatible cells (haplotype strains *MAT-1* x *MAT-2*) from two *S. scitamineum* isolates (SSC04 and SSC39), each representing distinct haplotype groups within a collection of Brazilian isolates [30]. In addition, we achieved an RNA-seq sequencing during *in planta* infection with a depth of approximately 10 times higher than our previous works [39] providing the most comprehensive view of the sugarcane smut fungus transcriptome to date. The enhanced resolution revealed key differences between gene expression when a more-virulent *S. scitamineum* isolate infected resistant and susceptible host genotypes. Our analysis focused on identifying single nucleotide variants (SNVs) within strains and between isolates, characterizing structural variants in mating-compatible cells, and investigating differential gene expression from RNA-seq of the fungus in its interaction with susceptible and resistant plants.

Results

The SSC04 isolate induces a higher incidence of whip in a susceptible sugarcane genotype

We employed a greenhouse experiment using needle-bud puncture as an inoculation method to test differences in virulence levels between sugarcane smut isolates with distinct genetic backgrounds when inoculated in the susceptible genotype RB925345. We quantified fungal DNA by qPCR at four time points (0, 48, 72, 120 h after inoculation) and found no significant difference ($p < 0.05$) in the mean DNA quantities between isolates at any of the measured time points (Additional file 1: Fig. S1). Morphometric comparative analyses revealed no significant differences in plant height, culm diameter, and number of tillers in plants inoculated either with SSC04 or SSC39 isolates. Notwithstanding, the two isolates differed with respect to the occurrence of whips. SSC04 induced whips a month earlier (~60 days post-inoculation) and in greater abundance than those plants inoculated with SSC39 (16 vs. 4 plants with whip/total plants, respectively) (~90 days post-inoculation) (Fig. 1). Based on whip emission, which is the most detrimental symptom of smut, we classified SSC39 as a less-virulent isolate (LV) and SSC04 as a more-virulent isolate (MV) for comparative purposes. Because *S. scitamineum* is dikaryotic during the infection stage, we individually sequenced and analyzed the haploid strains (*MAT-1* and *MAT-2*) derived from the SSC04 and SSC39 teliospores to assess

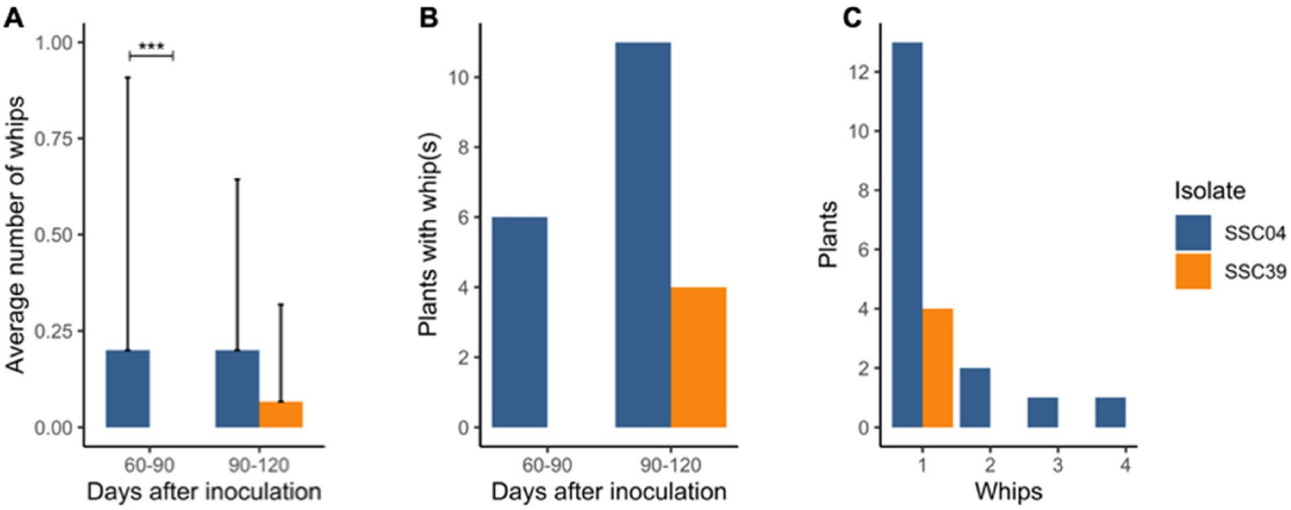

Fig. 1 Analysis of whip emissions in plant population. **A** Average number of whips among plants inoculated with two different *S. scitamineum* isolates (statistical significance was obtained through mixed-effect ANOVA followed by Tukey's test, asterisk represent *p*-value < 0.001). **B** Inoculated plants with whips according to isolate and number of days after inoculation. **C** Total number of whips per plant, according to the inoculated isolate. Control group plants did not show whips during the experimental period

their collinearity and similarity, with a particular focus on polymorphic genes and their potential role in virulence.

The genomes of SSC04 and SSC39 are highly similar and collinear except at the mating-type loci

We assembled three new chromosome-level genomes from telomere-to-telomere (T2T) (SSC04 *MAT-1* x *MAT-2* and SSC39 *MAT-2*) through *de novo* assembly of a combination of long Oxford Nanopore MinION reads and Illumina high-quality short reads and used the SSC39 *MAT-1* v.2 as the reference genome (Material and Methods). BUSCO analysis assessed the completeness of the assemblies, with all genomes exceeding 98%. The mitochondrial genome assembled to one contig for each strain is identical. Table 1 compares the basic features of the genome assemblies across all strains. Genomes comprised approximately 20 Mb, distributed across 26 chromosomes with a GC content of ~55% and gene density ranging from 331 to 334 per Mbp (Fig. 2). Assemblies shared more than 99.9% average nucleotide identity (ANI) in any pairing combinations (Additional file 2: Fig. S1) and were highly collinear (Additional file 2: Fig. S2–3), except at the mating-type loci (Additional file 2: Fig. S4A-D). Mating-type regions diverged between *MAT-1* and *MAT-2* strains of the same isolate and between strains of the same mating-type from different isolates. We found most contigs to be telomere-to-telomere complete across all assemblies (Additional file 2: Fig. S5–6). The mean length of telomeric repeats was 26 bp, ranging from 5 bp to 37 bp (Additional file 3: Table S1–3).

The sequenced genomes contained relatively few repetitive elements, with interspersed (related to transposable elements) (~ 4%) and simple repeats (short single-nucleotide motifs) (~ 1.6%) being the most abundant classified types. Noticeably, chromosomes 24, 25 and 26 were amongst the shortest (< 100 kbp) with a lower GC content (~ 50%), lower gene density (Chr24: 251.89 per Mb, Chr25: 170.24 per Mb, Chr26: 211.75 per Mb) and harbored larger amounts of repetitive elements (Chr24: 45.68%, Chr25: 22.84%, and Chr26: 23.39%) not detected extensively in other chromosomes (Additional file 2: Fig. S7). Despite having few annotated genes, we consistently detected the expression of 14 putative genes across all our tested conditions for both isolates (Additional file 2: Fig. S8). Using *ab initio* prediction, we annotated approximately 6,600 protein-encoding genes, representing 63% of the complete genome sequence in length. Most of them were part of single copy orthogroups among the four strains (Additional file 4: Table S1). By functionally re-annotating the reported proteome of Taniguti et al. (2015) [17], we expanded the set of candidate secreted effector proteins (CSEPs) to ninety-five putative proteins (Additional file 4: Table S2).

The mating-type loci *a* and *b* were tightly linked, approximately 50 kbp apart, in Chromosome 2. This region between the two loci contains genes with specific functions (Additional file 4: Table S3) alongside relics of transposable elements (TEs) (Fig. 3A). Additionally, CENPB-like (Centromere Protein B-like) sequences associated with repeats were identified, suggesting potential centromere locations. These repeated sequences and CENPB-like elements were scanned across all other chromosomes, marking potential centromere-associated regions (Additional file 2: Fig. S7). The loss of collinearity between loci *a* and *b* among strains and isolates included five genes adjacent to remnants of TEs (Fig. 3A). These

Table 1 Comparative overview of genomic features for four assemblies: SSC04 *MAT-1*, SSC04 *MAT-2*, SSC39 *MAT-1*, and SSC39 *MAT-2*. It includes metrics related to raw sequencing data (Nanopore and Illumina), assembly statistics (genome size, N50, number of contigs, GC content), and genome completeness (BUSCO scores in genome and proteome modes). Assemblies demonstrate high completeness and continuity, suitable for comparative genomic analyses between mating-types and isolates

Feature	SSC04 MAT-1	SSC04 MAT-2	SSC39 MAT-1	SSC39 MAT-2
Estimated bases of raw Nanopore reads (Mb)	656.93	3,936.4	-	846.19
Estimated bases of raw Illumina reads (Mb)	1,611.4	2,647.6	1,590.4	2,043.6
Genome size (bp)	19,975,378	19,970,026	20,073,740	19,998,869
Largest contig (bp)	2,011,902	2,011,573	2,010,363	2,010,361
Illumina mean depth of coverage	65x	113x	71x	82x
ONT mean depth of coverage	19x	170x	-	25x
Number of contigs	26	26	26	26
Number of mitochondrial contigs	1	1	1	1
Mitochondrial genome size (bp)	88,018	88,018	88,018	88,018
Genome N50 (bp)	870,864	884,916	876,086	864,988
GC content (%)	55.14	55.16	55.06	55.15
BUSCO completeness (genome mode) (%)	98.4	98.4	98.4	98.3
Number of protein-encoding genes	6,659	6,677	6,693	6,680
BUSCO completeness on predicted proteome (%)	98.2	98.1	98.2	98.2
Repeat elements (%)	6.06	5.74	5.75	6.13
Clustered rDNA copies	11	3	3	6
tRNAs	111	112	112	112

genes encode homologs of a putative N-acetyltransferase (*g972*), an uncharacterized protein (*g973*), a fungal-specific membrane protein (*g974*), a cytoskeleton assembly control protein (Sla2 family; *g976*), and the RPN5–26 S proteasome regulatory subunit (*g977*). A second inversion event altered the position of genes encoding homologs of an uncharacterized protein (*g985*) and the DNA replication regulator SLD2 (*g986*). Additionally, the CKI kinase gene *hhp1* (*g979*) exhibited an inversion only in the LV isolate SSC39 genomes, whereas the MV isolate SSC04 retained a conserved orientation in both strains.

Most genes within the mating-type loci were expressed under *in vitro* and *in planta* (48 h post-inoculation, hpi) conditions by both isolates (Fig. 3B) and only the actin-related protein 2/3 complex subunit 5 (*g983*) was differentially regulated. However, on the resistant cultivar, SSC04 *MAT-1* expressed only the first copy of the mating pheromone *g988.1* but not the second *g988.2* (Fig. 3). The latter lacks the cysteine in the canonical Cys–A-A-X sequence (where A represents aliphatic amino acids and X any amino acid) but has an upstream pheromone response element (PRE) sequence (sequences in Additional file 5: Fig. S1). By contrast, SSC04 *MAT-2* expressed both copies despite the absence of a canonical start codon (ATG) and the PRE sequence in one of the copies (Additional file 5: Fig. S1 & Table S1) (Fig. 3).

Using SSC39 *MAT-1* v.2 as a reference, we identified single nucleotide variants (SNVs) in SSC04 distributed across chromosomes (Fig. 4). A total of 2,033 SNVs were identified between SSC04 and SSC39, corresponding to, approximately, 1 SNV per 10 kb. Bi-allelic and multi-allelic variants, as well as multiple nucleotide polymorphisms (MNPs), were filtered out and corresponded predominantly to indels located in repetitive regions. The remaining 1,476 SNVs retained were categorized as shared by both *MATs* (907) or specific to a single *MAT* (569) in SSC04 and further classified based on their presence in coding (575) and non-coding regions (901) (Additional file 4: Table S4). Genetic differences were more pronounced between isolates (e.g., SSC04 vs. SSC39) than among strains of the same isolate (*MAT-1* vs. *MAT-2*) (Additional file 4: Table S4–5). The mating-type loci harbored 31.46% of all genomic SNVs (60 SNVs per 10 kbp), approximately 60 times higher than the genome-wide average (Fig. 4). Most SNVs within the mating-type loci were *MAT*-specific, detected between *MAT-1* and *MAT-2* strains of the same isolate (heterozygous-SNVs).

Besides these variants in the mating loci, we identified clusters of SNVs interspersed in various chromosomes. We designated these regions as 'highly polymorphic islands' (HPIs) 1–9, with HPI1 located on chromosome 2, containing the mating loci, and HPI 2–9 distributed across chromosomes 4, 6, 10, 11, 16, and 19 (Fig. 4). Variants in HPIs affected over a hundred genes, either by directly altering their coding sequences or impacting the surrounding flanking regions (Additional file 4: Table S6).

The total of 575 SNVs located in coding sequences mapped to 258 genes, with 80 genes containing more than one SNV. These variants resulted in 303 non-synonymous substitutions, 262 synonymous substitutions, and 10 additional variants classified as other types (e.g., splice-site or stop-gain; Fig. 5; Additional file 4: Table S4–5). Among the 303 non-synonymous substitutions identified between isolates, 160 genes carried homozygous SNVs, meaning the same variant was present in both *MAT-1* and *MAT-2* nuclei of the same isolate. However, within the mating-type region, 50 SNVs across 10

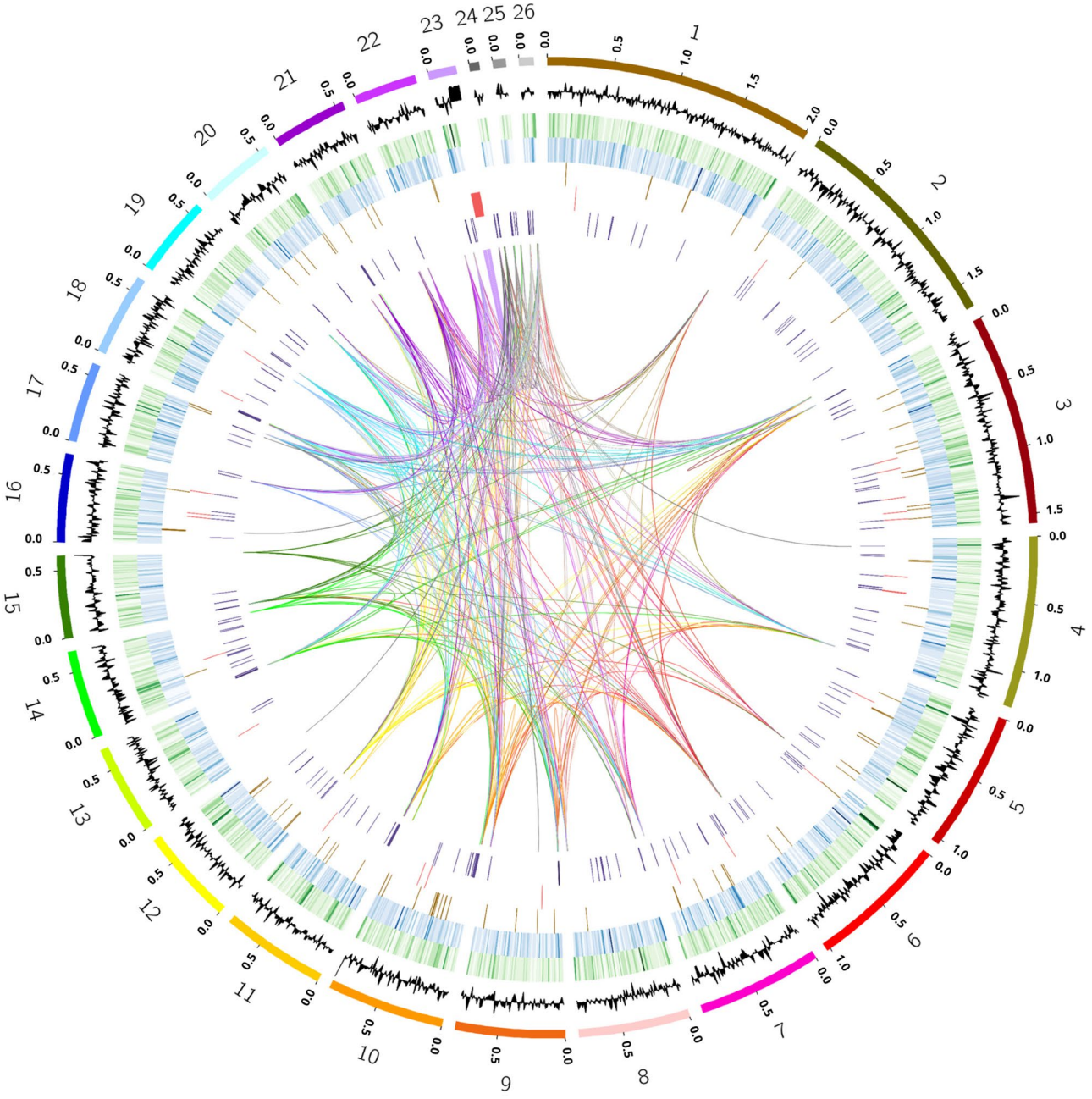

Fig. 2 Circos plot displaying genome-wide features among the 26 chromosomes of *Sporisorium scitamineum* SSC04 *MAT-1* genome. Track depict as follows: chromosomes (outermost ring), GC skew, coding sequence density on the plus strand, coding sequence density on the minus strand, genes encoding candidate secreted effector proteins, rDNA sequences, repetitive sequences (only DNA transposons and retroelements are shown), inner colored ribbons represent alignments of repetitive sequences using BLASTn (e-value < 1e-10 and identity of 95%) limited to chromosome ends (500 bp in length)

coding sequences (*g970*, *g971*, *g972*, *g974*, *g976*, *g982*, *g984*, *g985*, *g986*, and *g988*) were heterozygous, exhibiting sequence variation between the two nuclei of the same isolate. Only three other genes outside mating-loci exhibited heterozygous SNVs: *g406*, encoding a homolog of *U. maydis UMAG_00389* (Hsk1-interacting molecule, Him1); *g994*, a homolog of the oligopeptide transporter 4 (OPT4) (*U. maydis UMAG_02387*); and an aquaporin homolog (*UMAG_10452*).

A functional analysis of 160 polymorphic genes, which harbored homozygous SNVs leading to non-synonymous amino acid substitutions, revealed functions related to membrane trafficking, cell cycle regulation, signal transduction, cell wall remodeling and stress response. Notably, these included genes involved in vesicular trafficking,

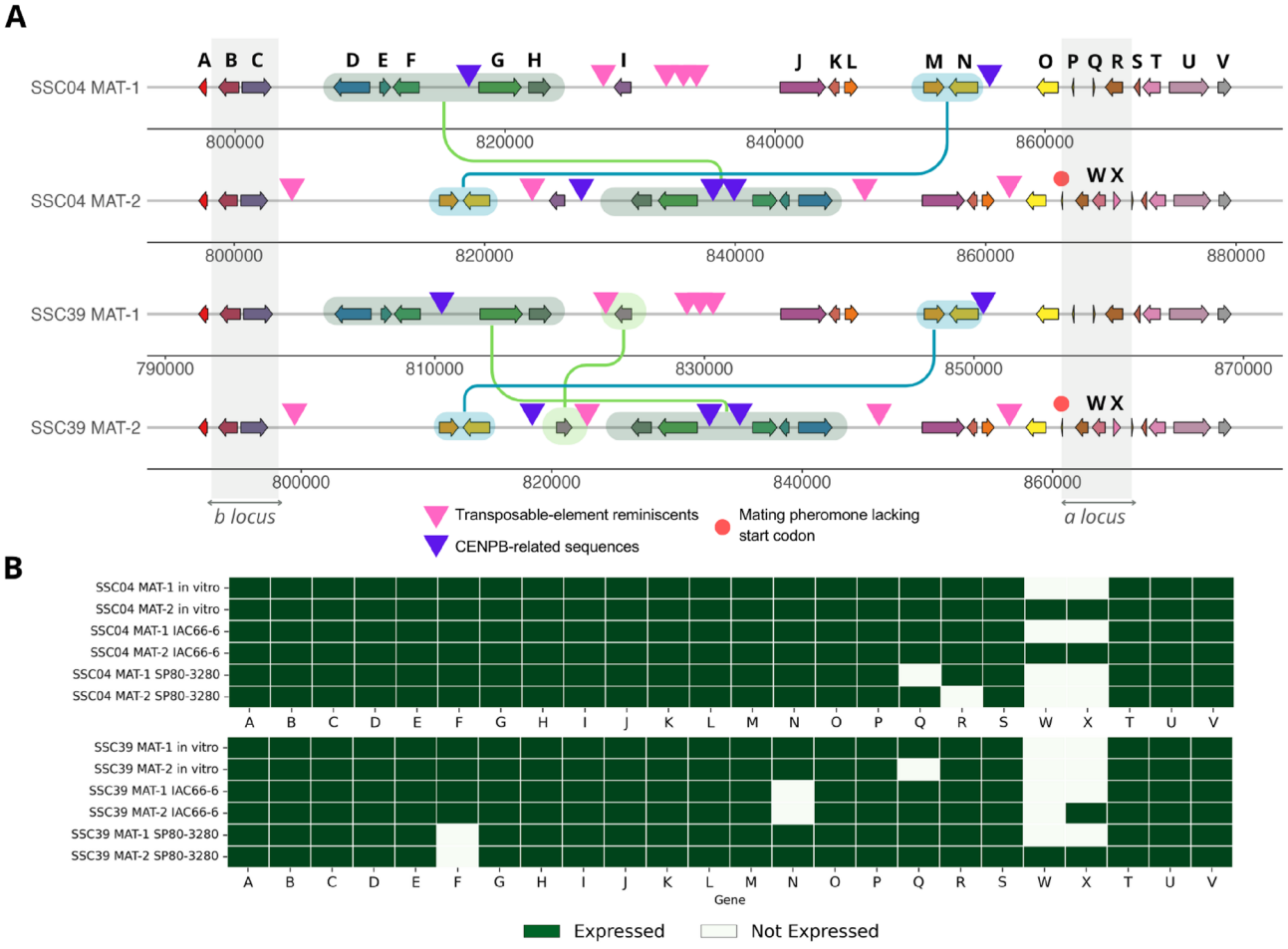

Fig. 3 Structural variation and expression profiling of mating-type genes among strains of SSC04 and SSC39 isolates. **A** Schematic representation of structural variants identified in the mating-type loci region on chromosome 2 across the four assemblies (SSC04 *MAT-1* and *MAT-2*, SSC39 *MAT-1* and *MAT-2*). **B** Transcript counts for annotated ORFs in this region were obtained using featureCounts v1.6.0. Expression was considered detected for a gene if at least one count was observed in any biological replicate. A (*g969*) - U3 ribonucleoprotein complex (Sas10/Utp3 domain), B (*g970*) - B mating-type locus, bE allele, C (*g971*) - B mating-type locus, bW allele, D (*g972*) - N-acetyltransferase, E (*g973*) - Uncharacterized protein, F (g974) - Fungal-specific membrane protein, G (g976) - Cytoskeleton assembly control protein (Sla2 family protein), H (*g977*) - RPN5-26 S proteasome regulatory subunit, I (*g979*) - Serine/Threonine protein kinase hhp1 (CK1 protein kinase), J (*g982*) - Fungal specific transcription factor domain, K (g983) - Actin-related protein 2/3 complex subunit 5, L (*g984*) - tRNA-splicing endonuclease subunit Sen15 domain-containing protein, M (g985) - Uncharacterized protein, N (*g986*) - DNA replication regulator SLD2, O (*g988*) - Chitin deacetylase, P (*g988.1*) - mating factor (pheromone), Q (*g988.2*) - mating factor (pheromone), R (*g988.3*) - Pheromone receptor, S (g988.4) - Rba protein, T (*g989*) - Pantoate-beta-alanine ligase (PanC superfamily), U (*g990*) - uncharacterized protein, V (*g991*) - CBP4 family protein, W (*g988.5*) - lga2, X (*g988.6*) - rga2

especially related to the exocyst complex (*g1974*/Sec3, *g4764*/Sec6, and *g5005*/Sec34-like); intracellular trafficking (*g1876*, Vps4-associated protein 1; *g1877*, Vps9; *g1835*, Vps16; *g1888*, Tgl2/syntaxin); metabolic adaptation (*g4284*, class 3 lipase), and stress responses (*g1971*, ATR serine/threonine-protein kinase homolog; *g4167*, *Mus42* homolog). Among those, the most variable were the Sec3 homolog (*g1974*) and DNA damage repair protein (*g4167*). We also identified various genes related to carbohydrate metabolism, comprising genes associated with galactose metabolism (*g391* and *g631*), predicted to encode orthologs of a potential raffinose synthase and a galactose oxidase domain-containing protein, respectively, fructose and mannose metabolism (*g4283*), and the regulation of 1,3-beta-glucan synthase activity (*g2864*). The full list, including the number and location of polymorphic sites, is provided in Additional file 4: Table S4–5.

We highlight twelve genes harboring more than four SNVs. This set includes eight genes encoding candidate secreted effectors (CSEPs): *g1247*, *g3870*, *g4260*, *g4278* with signal peptides and non-cytoplasmic domains; *g5359*, encoding a protein with an intrinsically disordered region (Mobi DB-lite); and three others previously studied [40, 41], *g1052*, *g4257*, and *g5159*. Selected variants in these three genes were further validated by Sanger sequencing in a pool of over 40 isolates from Brazil and Argentina (Additional file 6: Fig. S1–4).

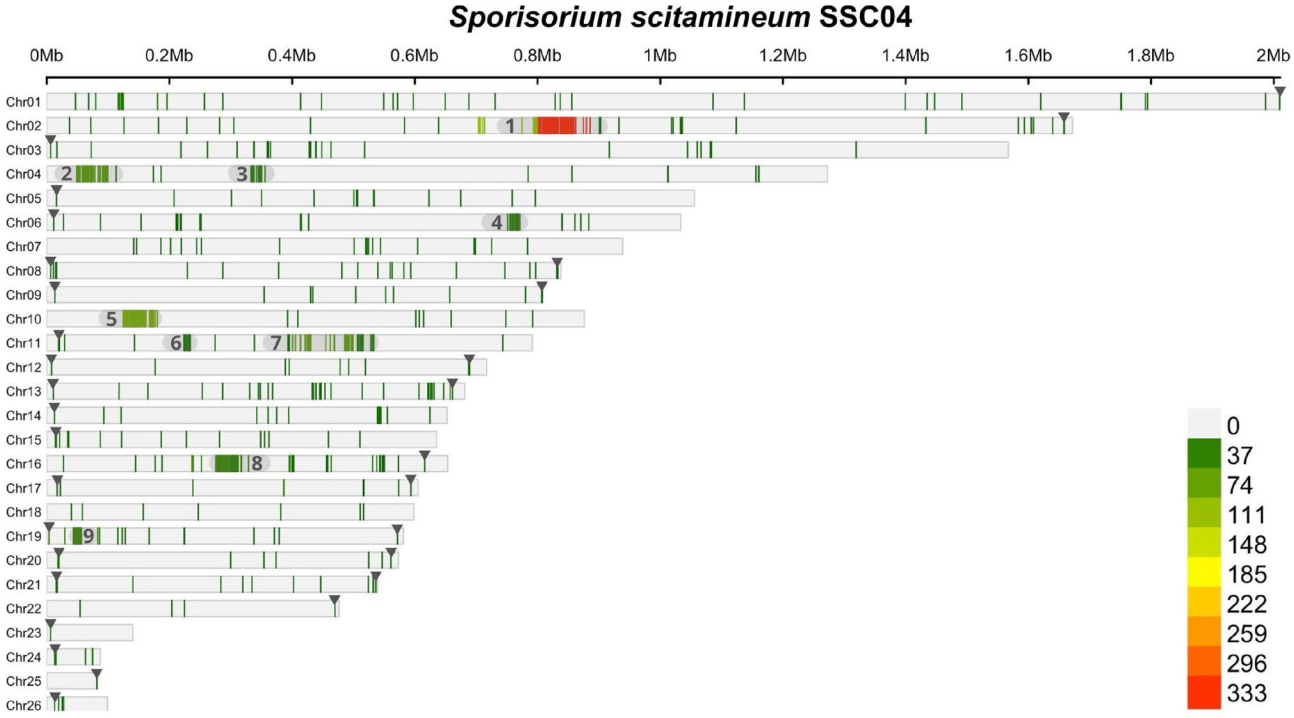

Fig. 4 Single-nucleotide variant (SNV) density plot of both homozygous and heterozygous polymorphic sites across SSC04 *MAT* strains. SSC39 *MAT-1* v.2 was used as the reference genome. SNVs were identified using GATK v.4.5.0 and plotted in 10-kbp windows with CMplot v.4.5.1. The color gradient visually reflects variant density, from dark green (low density) to dark red (high density). The Highly Polymorphic Islands (HPIs 1 to 9) regions correspond to continuous genomic segments with a high density of adjacent single-nucleotide variants. Black downward arrows indicate subtelomeric polymorphisms

Sugarcane genotype influences fungal transcriptomes and uncovers adaptive pathogenic strategies

To achieve high-resolution insights into the transcriptional dynamics of the more virulent isolate during early infection of sugarcane genotypes with contrasting resistance levels (smut-resistant SP8-03280 and smut-susceptible IAC66-6), we performed high-coverage total RNA sequencing 48 h post-inoculation (hpi) with SSC04 teliospores. Sequencing yielded 1,027,426,235 paired-end reads from inoculated plants, of which 868,330 mapped to the reference genome. Statistics regarding sequencing can be seen in Table 2.

Fungal quantification by qPCR confirmed SSC04 infection in both genotypes, revealing a statistically significant difference in fungal growth between them ($p = 0.0392$; Additional file 7: Fig. S1). Fungal reads accounted for 0.1% of RNA-seq reads in the susceptible genotype and 0.06% in the resistant genotype (Additional file 7: Table S1). A total of 5,391 predicted genes were expressed by SSC04 during plant infection, representing 81% of all predicted *S. scitamineum* genes (6,673) (Additional file 8: Table S1). In this context, we also performed transcriptome sequencing of SSC04 cells grown under axenic conditions for comparative purposes. Sequencing produced 46,145,863 paired-end reads and we identified 6,319 predicted genes during *in vitro* growth, representing 94% of coding sequences (Additional file 8: Table S7). Notably, we detected the expression of 41 CSEPs during biotrophic growth, representing 43.16% of our predicted effectorome (Additional file 7: Fig. S3–4). Among them, *g1344*, *g1450*, *g2498*, *g3902*, and *g4839*, harbored expansin-like domains (Plant Cell Wall Expansin (IPR051477)/DPBB_RlpA_EXP_N-like).

To explore differences in fungal activity between host genotypes, we performed differential expression analysis of SSC04 genes (SSC04-DEGs) (Fig. 6), comparing inoculated smut-resistant vs. smut-susceptible plants. We prepared two lists: one where we considered an FDR < 0.05 and |log2FC| >1, resulting in 98 DEGs (Additional file 8: Table S2), and another with less stringent criteria considering $p < 0.05$ and |log2FC| >1, resulting in 710 DEGs (Fig. 6; Additional file 8: Table S3). This dual approach enabled the detection of potentially important genes that might have been missed under stricter filters, ensuring biologically relevant findings were preserved.

Among the induced SSC04-DEGs in the resistant genotype, we identified genes encoding cell wall-degrading enzymes (CWDEs), candidate secreted effector proteins (CSEPs), and biosynthetic clusters for mannosylerythritol lipids (MELs) and ustilagic acid (UA), as well as genes involved in xenobiotic metabolism and detoxification (Table 3). Considering the CWDEs, we highlight an *Egl1* homolog (*g3790*, endoglucanase) showing an 80-fold induction. Additionally, nine CSEPs were detected as

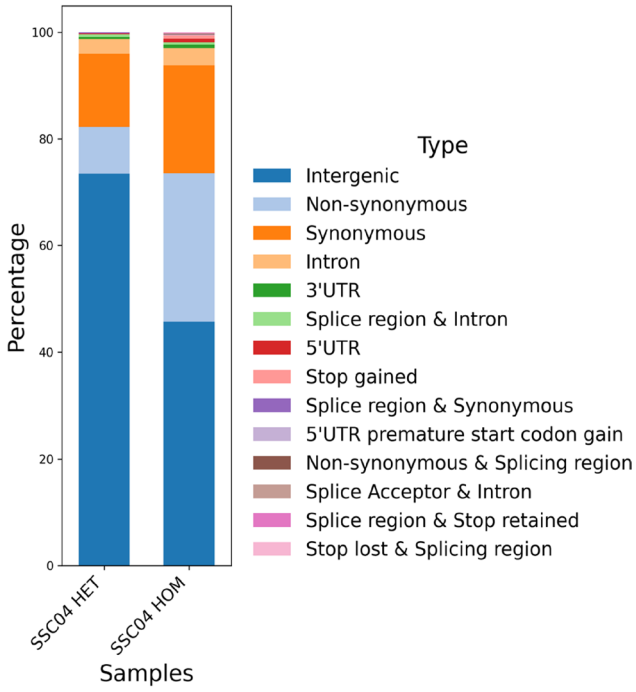

Fig. 5 Stacked bar of genetic variation of SSC04 *MAT* strains. Plot showing the percentage distribution of genetic variants categorized as shared by both *MAT* (HOM: homozygous, $n=907$) or specific to a single *MAT* (HET: heterozygous, $n=569$) in SSC04 when the fungus exists in its dikaryotic form. The x-axis represents the variant types, while the y-axis shows the percentage of each variation type within the samples. Each color in the stacked bars corresponds to a different type of genetic variation

Table 2 Comparative transcriptome profiling of *Sporisorium scitamineum* isolates SSC04 and SSC39

Features	SSC04[1]	SSC39[1]
Total mapped paired-reads *in planta*	868,330	99,341*
Total mapped paired-reads *in vitro*	41,399,714	23,495,162**
Expressed *in planta*	5,391 (81%)	3,452 (52%)
Detected only in S plants	111 (1.66%)	4 (0.06%)
Detected only in R plants	18 (0.27%)	139 (2%)
Expressed *in vitro*	6,319 (94.74%)	6,301 (94.42%)
Detected only *in vitro* in *SSC04*	5 (0.07%)	-
Detected only *in vitro* in *SSC39*	-	6 (0.09%)
Restricted to *in vitro* growth	529	836

[1]Data were analyzed independently for each isolate. Total gene annotation comprises 6,673 genes [17]

*Analyzed data obtained by Rody et al. (2019)

**Reanalyzed data obtained by Taniguti et al. (2015) [17]

DEGs upregulated in the resistant cultivar (Table 3). We also identified genes involved in the activation of nitrogen starvation responses, *Nit2* (*g5371*), which showed an 8-fold induction, and *Opt2* (*g5318*), a peptide transporter upregulated 32-fold in the resistant genotype. The most highly induced gene (200-fold) was *g3970*, a CSEP previously described by Teixeira-Silva et al. (2020) [41] and Taniguti et al. (2015) [17].

In contrast, we emphasize the set of SSC04-DEGs upregulated in the susceptible genotype. Genes associated with growth and cell division, cell wall remodeling, and vesicle trafficking, including vacuolar sorting proteins (VSPs), iron homeostasis, oxidative stress, DNA replication, membrane-associated components, and a distinct set of CSEPs, were upregulated (Table 4). The susceptible interaction also triggered significant activation of the unfolded protein response (UPR) pathway (GO:0051082, $p=0.0075$). We identified highly expressed genes *g3729*, an UPR protein; *g1608*, a glutathione peroxidase; genes encoding proteins of chromatin regulation, *g2695* and *g4330*, related to histones acetylation and deacetylation (Table 4).

Candidate effector genes have contrasting expression profiles according to isolate virulence and host susceptibility

For comparative purposes, we analyzed the DEGs of the less-virulent strain SSC39 during infection of the same sugarcane genotypes, smut-resistant SP80-3280 and smut-susceptible IAC66-6, using RNA-seq data previously generated but not yet reported [39] (Additional file 7: Fig. S5). The SSC39 experiment yielded 47,902,014 paired-end reads from inoculated plants, of which 99,341 aligned to the reference genome and were included in this analysis (Additional file 7: Table S1). In infected tissues, 52% of SSC39 genes (3,452) showed detectable expression, most of which (96%) were also expressed by the more-virulent SSC04 isolate under similar conditions (Additional file 8: Table S4).

Differential expression analysis revealed 22 SSC39-DEGs (FDR < 0.05 and $|\log_2 FC| > 1$) (Additional file 8: Table S5) and 419 SSC39-DEGs ($p < 0.05$ and $|\log_2 FC| > 1$) (Additional file 8: Table S6) when comparing resistant and susceptible genotypes. In this comparison, we focused only on candidate secreted effector proteins (CSEPs), identifying fourteen CSEP-DEGs in SSC39 and eleven in SSC04, with only three genes (*g1450*, *g3870*, *g6107*) shared between isolates but displaying opposite expression patterns. Among them, *g3870* stands out as a particularly relevant effector candidate due to the presence of a non-synonymous SNV and its contrasting transcriptional profile across genotypes and isolates.

Discussion

Resistance to smut is a quantitative trait with moderate heritability, influenced by environmental conditions and crop management practices, such as whether sugarcane is newly planted or ratooned [4, 5]. In our greenhouse study, using injury as the inoculation method, SSC04 spores induced whip emergence significantly earlier than SSC39. Even though greenhouse conditions may not fully reflect genotype responses in the field, they offer

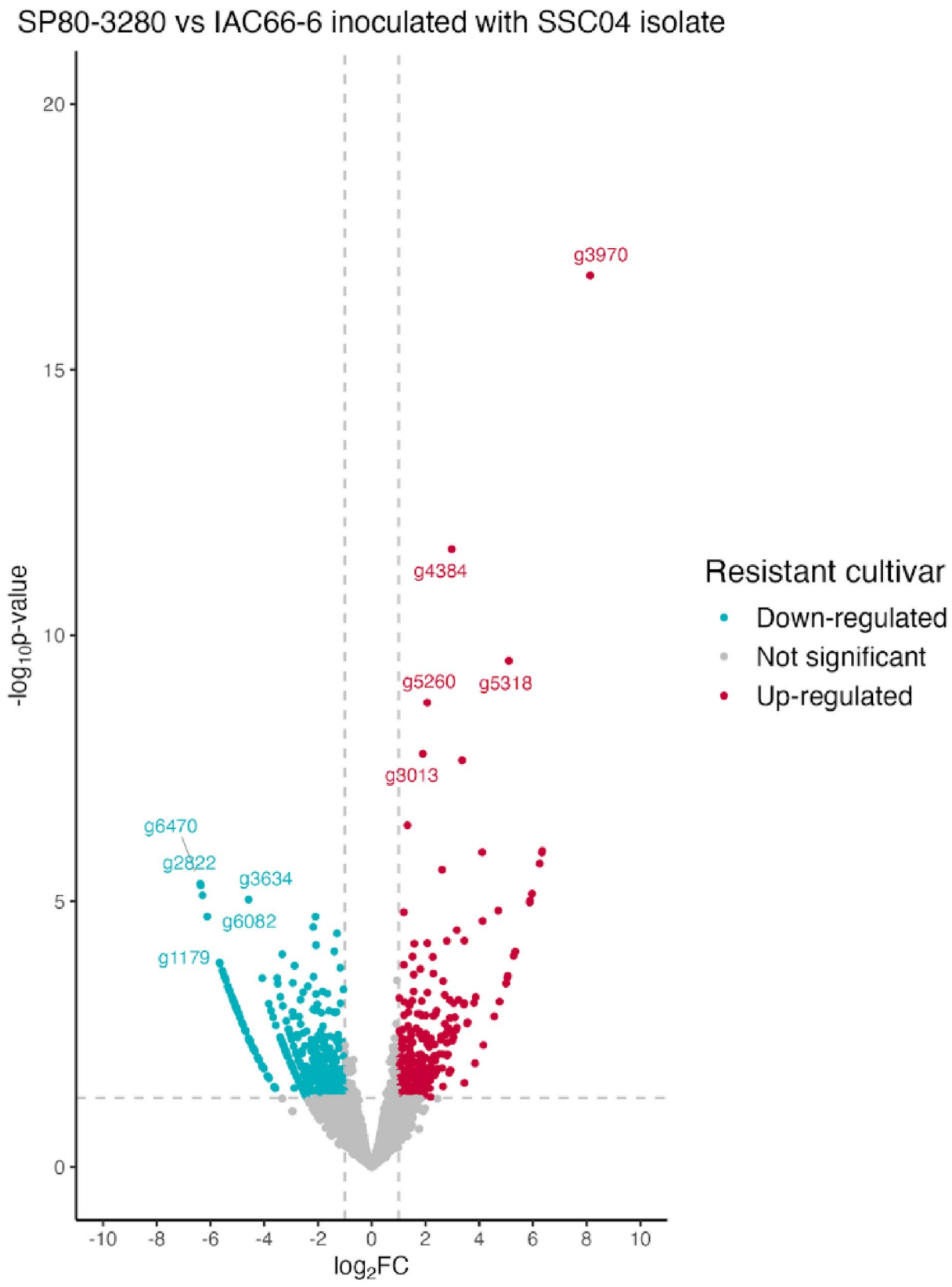

Fig. 6 Volcano plot of differentially expressed genes (DEGs) in SSC04 isolate when comparing colonization of SP80-3280 resistant genotype and IAC66-6 susceptible genotype. Red dots represent induced genes in the resistant genotype, while blue dots represent repressed genes in the resistant. Labeled points correspond to selected genes with the most significant expression changes

Table 3 Functional categorization and expression levels of upregulated SSC04 genes during infection of the resistant sugarcane genotype

Gene ID	Product	Log_2FC (R/S)
Cell Wall Degrading Enzymes (CWEDs)		
g3790	Endoglucanase 1 (EGL1)*	6.34592
g4394	Endo-1,4-beta-xylanase Xyn11A	4.75559
g3529	Pectin lyase*	4.15459
g4463	Alpha-galactosidase	3.84324
g1655.1	Alpha-L-arabinofuranosidase I (ERC1)	4.58165
g5941	Endo-1,4-beta-xylanase b	3.54798
g2736	Ribonuclease U1*	2.09640
g141	Glucan 1,3-beta-glucosidase	1.23247
Candidate Secreted Effector Proteins (CSEPs)		
g3970	RSP3	8.12969
g1513	PELE1	5.96570
g2338	PIT1	2.65485
g1777	SUC2-invertase	2.13904
g3870	Uncharacterized	1.41489
g1612	Uncharacterized	1.36606
g6107	Uncharacterized	1.34995
g6307	CMU1	1.08899
g1642	CPL1	0.77027
Ustilagic acid (UA) biosynthesis cluster		
g4377	Glycosyltransferase UGT1	5.88036
g4379	ORF2	2.37989
g4380	Probable alcohol acetyltransferase ORF1	4.70479
g4381	Cytochrome P450 monooxygenase CYP1	4.12421
g4383	ABC-type transporter ATR1	1.66867
g4382	Acyltransferase UAT1	1.34499
g4384	Fatty acid synthase FAS2	2.97657
g4385	Cytochrome P450 monooxygenase CYP2	5.27931
Mannosylerythritol lipids (MELs)		
g2945	Erythritol-mannosyl-transferase 1 EMT1	1.84090
g2946	Acyl-CoA-dependent acyltransferase MAC1	-
g2947	MFS-type efflux pump MMF1	1.02151
g2948	Acetyltransferase MAT1	-
Xenobiotic Metabolism		
g1378	Related to TNA1 - high-affinity nicotinic acid plasma membrane permease	1.34758
g1559	Related to short-chain dehydrogenase	1.58631
g1791	Related to aminotriazole resistance protein	1.56008
g1841	Major facilitator superfamily (MFS) profile domain-containing protein	1.31837
g1945	Citrate transporter-like domain-containing protein	1.56460
g2927	ABC drug exporter AtrF	1.19145
g2947	MFS-type efflux pump MMF1	1.02151
g3013	YBT1-Vacuolar, ABC protein transporting bile acids	1.90188
g3328	Major facilitator superfamily (MFS) profile domain-containing protein	1.47902
g3346	Related to SNG1-protein involved in resistance to nitrosoguanidine and 6-azauracil	2.29307
g4240	Related to positive effector protein GCN20	1.73063
g4334	Cytochrome P450	1.23174
g5019	Putative aflatoxin efflux pump AFLT	4.11055
g5260	ABC transporter domain-containing protein	2.06738
g5863	Related to FLR1 - Putative H + antiporter involved in multidrug resistance	1.19183
g6414	Probable SNQ2 - ABC transporter involved in multidrug resistance	1.40560

*Protein-encoding genes also studied by Nalayeni et al. (2016)

Table 4 Functional categorization and expression levels of upregulated SSC04 genes during infection of the susceptible sugarcane genotype

Gene ID	Product	Log_2FC (*R*/S)
UPR-related		
g639	Pdi1 (Protein disulfide isomerase 1)	1.04483
g3729	Related to SPC3-signal peptidase subunit	3.85806
Cell-wall associated, cell-cortex growth and vesicle trafficking		
g5897	Related to SPA2 protein	1.26549
g619	Related to CDC24-GTP/GDP exchange factor for Cdc42p	1.12074
g1850	Probable Chitin deacetylase	1.29402
g4597	F-actin	1.74618
g5512	C-terminal kinesin	5.17834
g5563	Myosin V	0.82535
g3892	Probable Chs8-Chitin Synthase 8	0.63336
g4585	Chitin Synthase 2	1.21343
g4702	Related to CHS7-control of protein export from the ER (Like chitin synthase III)	2.55353
Oxidative stress responses		
g1608	Glutathione peroxidase	5.28040
g2735	Glutathione-S-transferase	1.90947
g4476	Probable PEX1-peroxisomal assembly protein-peroxin	1.83827
DNA replication, repair and chromatin regulation		
g117	Related to DNA polymerase	6.12412
g2052	DNA polymerase	3.34422
g494	Related to MutS protein homolog 5	2.64399
g734	Related to RAD5-DNA helicase	2.30639
g1233	Histone acetyltransferase	1.02976
g2695	Related to histone acetyltransferase subunit HAT1	5.00893
g4330	Histone deacetylase	5.10873
g6310	Related to HOS3-Trichostatin A-insensitive homodimeric histone deacetylase (HDAC)	1.13872
Candidate secreted effector proteins		
g1450	Uncharacterized	0.898680
g2894	Uncharacterized	3.821870
g6254	Uncharacterized	5.071000
Iron homeostasis		
g3634	L-Ornithine N5-oxygenase	4.58158
g4510	Probable glutaredoxin	2.04253

an opportunity to evaluate them comparatively in short periods [42]. SSC04 also induced symptoms approximately 30 days earlier than SSC39 ($p < 0.001$), affecting 28% of infected stools, compared to ~ 6% of SSC39, supporting its classification as a more virulent isolate [11, 32]. Following Latiza's (1980) [43] smut rating scale, the disease incidence caused by SSC04 classifies the genotype RB925345 as highly susceptible, contrasting with its previous classification as intermediately resistant when tested with SSC39 spores [44]. Smut-resistant plants respond to fungal infection and thus fungal development is delayed when compared to growth in smut-susceptible plants [45] as schematically demonstrated in Fig. 7. Host genotype-specific conditions, such as plant developmental timing and capacity for immune responses may influence the underlying molecular landscape of *S. scitamineum* for infection.

To investigate the molecular basis of our observed variation, we analyzed fungal genomic and transcriptomic data to propose a model in which *S. scitamineum* adopts distinct interaction strategies depending on whether the host genotype is resistant or susceptible.

Genomic variation signatures of more-and less-virulent isolates

This study represents a genome-wide assessment of genetic variation in *S. scitamineum* haploid strains (haplotypes) derived of isolates with contrasting virulence levels. We generated chromosome-level assemblies of three haploid genomes (SSC04 *MAT-1*; SSC04 *MAT-2*; SSC39 *MAT-2*) and a revised version of a fourth genome (SSC39 *MAT-1* v.2) [17] combining long- and short-read sequencing technologies. These assemblies revealed high genetic similarity (homozygosity) between strains of the same isolate and strains of different isolates (>99% average nucleotide identity), consistent with prior reports of genetic homogeneity in populations of *S. scitamineum* [9, 25, 26, 30] and reflecting its predominant reproduction strategy [26]. The smallest chromosomes identified in our four assemblies shared structural features (% GC, repetitive sequences, low gene density) with accessory chromosomes observed in other plant-pathogenic fungi [46]. While accessory chromosomes are often linked to adaptive traits such as virulence [47], their characterization and functional role in *S. scitamineum* remains speculative and warrants further investigation.

Despite this overall conservation, single nucleotide variations (SNVs) were identified among isolates, with most polymorphisms localized to non-coding regions (upstream/downstream of genes). Tendency for non-coding variation may be a sign of purifying selection on coding regions, particularly those crucial housekeeping genes [48]. Additionally, areas of the genome that are highly polymorphic may serve as hotspots for mutation. A subset of polymorphic genes resulted in proteins with amino acid substitutions, including those associated with mating-type determination and candidate secreted effector proteins (CSEPs), functional categories linked to virulence in smut fungi [10, 37]. We also detected nine highly polymorphic islands (HPIs 1–9) in different chromosomes comparing strains. HPI-1 encompasses the region containing the mating-type loci, where the accumulation of polymorphisms is consistent with previously reported recombination suppression between these loci [49]. In contrast, the high density of variants observed

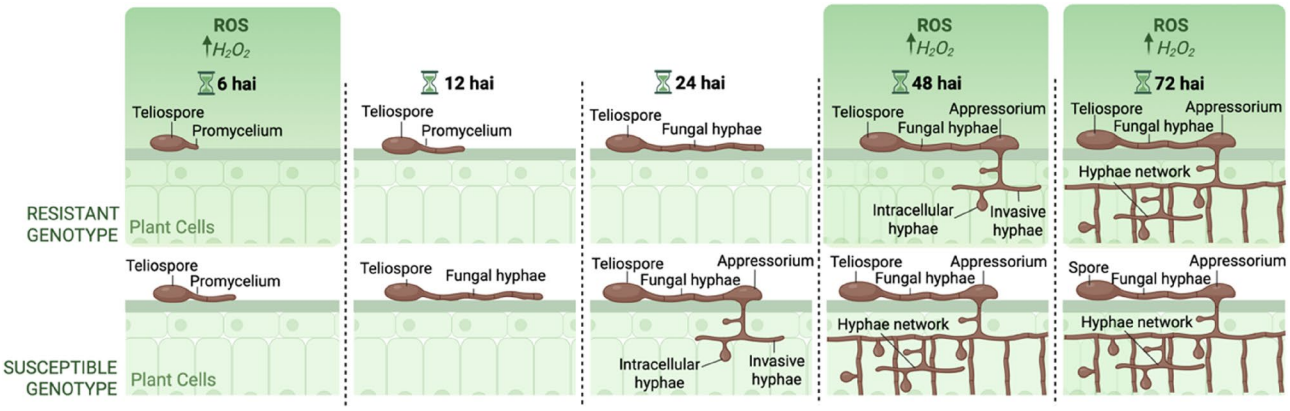

Fig. 7 Host genotype-specific fungal development. Schematical representation of *Sporisorium scitamineum* development when infecting smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) sugarcane genotypes based on Peters et al. (2017) [45]. Created in BioRender. Vilanova, P. (2025) https://biorender.com/eatzxxm

in other chromosomal regions does not appear to correspond to known compartmentalized genome segments in *S. scitamineum* [12]. These HPIs encompass differentially expressed genes (DEGs) involved in DNA repair, lipid/nitrogen metabolism, and vesicular trafficking pathways, often located near polymorphic sites of non-coding sequences. While some overlap with DEGs, our study does not establish eQTLs, as expression-genotype associations at the population level were not assessed. Still, the localization of dense polymorphisms near functionally relevant genes raises important questions about the role of mutation and selection in shaping genome evolution in *S. scitamineum*.

The HPI-1 contains multiple polymorphisms across nearly all genes and intergenic regions. This pattern likely reflects suppressed recombination between mating loci, a mechanism that promotes divergence between homologous chromosomes while preserving allelic combinations critical for non-self-recognition [49]. Noteworthy, the presence of CENP-like sequences potentially marks the centromere linked to mating loci. The presence of centromeres between the two mating loci has been reported in other systems, where detailed studies have described the evolution of tetrapolar systems driven by centromere fission or translocations in centromere-flanking regions [20, 50–52]. The compact architecture of this region in *S. scitamineum* (50 kbp between *a* and *b* loci) contrasts sharply with the ~ 500 kbp TE-rich regions in *Ustilago hordei* and *U. esculenta* [53, 54], also with a bipolar mating structure, suggesting distinct evolutionary trajectories. In *S. scitamineum*, fragmented transposable element (TE) relics, rather than intact LTR-type TEs, are interspersed among twelve predicted genes. The synteny breakpoint between the bipolar *S. scitamineum* and the closely related tetrapolar *S. reilianum* aligns precisely with transposable element remnants located between the fungal-specific transcription factor *g983* and the serine/threonine kinase gene *hhp1* (*g982*) [17]. This spatial association strongly implicates TE-mediated structural rearrangements in driving architectural reorganization of the mating locus during speciation [49].

The *hhp1* gene (*g982*) also exhibited a structural rearrangement: an inversion in the less-virulent isolate, contrasting with the conserved orientation in the more-virulent isolate. The Hhp1 protein is a CKI kinase that regulates vesicular trafficking, DNA repair, and cell cycle progression [55, 56]. Although these and other rearrangements at mating loci did not significantly alter the expression of the affected genes under our experimental conditions, structural variation in mating loci is increasingly recognized as related to adaptive evolution in plant pathogens [19]. Whether the observed inversion in *hhp1* and other intragenomic rearrangements alter expression during infection, potentially impacting host-pathogen interaction dynamics, remains a question for future investigation.

Functional annotation suggested that genes flanking the mating-type loci are associated with cell cycle regulation and filamentous growth, both critical for fungal mating and host colonization [57, 58]. We did not identify DEGs among most of these genes, except for an actin-related protein 2/3 complex subunit 5 (*g983*), significantly induced during infection of the susceptible sugarcane genotype, possibly associated with fungal proliferation [59]. However, we observed variations in expression levels in some of the genes across host genotypes and isolates. This variability suggests a context-dependent regulation, suggesting fine-tuning dimorphic transitions in response to host-specific cues. Alongside *hhp1*, other genes in this region may contribute to fungal growth and colonization. For instance, *g986*, which encodes a homolog of *SLD2*, plays a pivotal role in G1 phase progression by assembling the pre-replicative complex [60] and has been shown to interact with G1 cyclins/CDKs during

mating-type switching in *S. cerevisiae* [61]. Although the role of SLD2 in mating remains uncharacterized in Ustilaginaceae, its CDK dependence, key regulators of mating progression, suggests potential functional conservation [57]. Similarly, *g974*, which likely encodes a protein with a SUR7/Rim9-like domain (IPR009571), has been associated with cell morphogenesis [62]. The clustering of these genes positioned between mating loci may reflect an evolutionary strategy that links morphological adaptability with mating and host colonization processes.

We also identified non-synonymous mutations in vesicular trafficking genes, particularly in components of the exocyst complex and in endosomal trafficking genes, in the more virulent *S. scitamineum* isolate. Given the known roles of the exocyst in polarized effector secretion and hyphal growth [63], these polymorphisms may represent candidate genetic variations relevant to virulence [64]. The class 3 lipase (*g4284)*, for instance, may have a dual role of hydrolyzing host lipids for nutrient scavenging while disrupting plant membrane integrity, a recognized virulence strategy in phytopathogens [65].

Although we did not detect DEGs for all the polymorphic genes mentioned in our transcriptome experiment 48 hpi, this does not preclude their functional relevance to other infection phases (e.g., sporulation) or under unexamined stressors (e.g., nutrient deprivation). They represent high-priority candidates for functional validation to dissect how genetic variation shapes isolate-specific virulence strategies.

Transcriptome dynamics reflecting host-specific adaptation

The in-depth RNA sequencing of smut-inoculated sugarcane at 48 h post-inoculation (hpi) with SSC04 teliospores revealed host-dependent transcriptional profiles, highlighting genotype-specific interactions between sugarcane and *S. scitamineum*. While a large number of DEGs were identified, we focused our discussion on genes with known or putative functions relevant to smut fungi biology. Uncharacterized DEGs, although not addressed individually here, remain valuable targets for future functional studies.

In resistant plants, we detected strong induction of genes encoding secreted CAZymes, including cell wall-degrading enzymes (CWDEs) such as Egl1 (*g3790*, 80-fold), essential for breaching host barriers [66]; Erc1 (*g1655.1*), involved in suppressing β-glucan-triggered defenses [1]; and Xyn1 and Xyn11A (*g5941* and *g4394*, respectively), both associated with fungal virulence [67]. A homolog of *U. maydis pit4*, encoding a protein containing an expansin domain, was the third most highly induced gene (~ 64-fold). Expansins act via a non-enzymatic mechanism to loosen host cell walls and facilitate penetration [68, 69]. In addition, we identified induced genes encoding known effectors Cpl1 (*g1642*) and Rsp3 (*g3970*) of smut fungi. In *U. maydis*, these proteins appear to interact to defeat host immune responses at the cell wall level [13, 70]. Altogether, the induction of CWDEs, expansin-like proteins, and effectors in resistant sugarcane genotypes supports a layered infection strategy by *S. scitamineum*, combining mechanical penetration with active immune suppression [66, 71]. More recently, *g1642* was named SsPE15 and described in *S. scitamineum* to hijack the host's vesicular trafficking system to avoid immune detection [72].

In parallel, SSC04 upregulated *nit2* (*g5371*, 8-fold), a master regulator of nitrogen assimilation and virulence [73], and *opt2* (*g5318*, 32-fold), a peptide transporter associated with biotrophic development [74]. Nutrient availability, particularly nitrogen, is a key regulator of fungal pathogenicity, influencing mating and biotrophic development in smut fungi [75]. These responses coincided with the expression of the UA and the MEL biosynthesis pathways, which facilitate adhesion under nitrogen-limited conditions [16, 76] and underline that nitrogen scarcity primes smut fungi for pathogenic development [77].

Finally, we observed induced genes related to xenobiotic metabolism encoding transporters, such as ATP-binding cassette (ABC) efflux transporters and members of the Major Facilitator Superfamily (MFS), that may act in the excretion of toxic compounds or nutrient uptake [18]. These may reflect stress responses triggered by the resistant genotype, which, over time, negatively affects fungal colonization [78]. Peters et al. (2017) [45] reported hydrogen peroxide accumulation in resistant sugarcane genotypes at several point post-infection, including 48 hpi, however, we did not detect fungal DEGs involved in reactive oxygen species (ROS) detoxification at this stage. The proposed model for the strategies used by *S. scitamineum* during infection of the smut-resistant genotype can be seen in Fig. 8.

In contrast, when infecting a susceptible host, *S. scitamineum* SSC04 induced genes related to hyphal proliferation and stress adaptation. The induction of genes related to polarized growth emphasizes the dependence on tip-focused vesicle transport and cell wall remodeling for hyphal spread [79]. In the susceptible genotype we identified genes linked to host-derived toxin neutralization, probably responding to distinct stress cues beyond canonical oxidative stress [80, 81]. Furthermore, the activation of fatty acid catabolism pathways highlights the pathogen's reliance on host lipid reserves to sustain hyphal growth and colonization.

One of the most significant processes activated in the susceptible genotype was that related to the Unfolded Protein Response (UPR) pathway. The UPR is critical to maintaining cellular homeostasis under stress conditions.

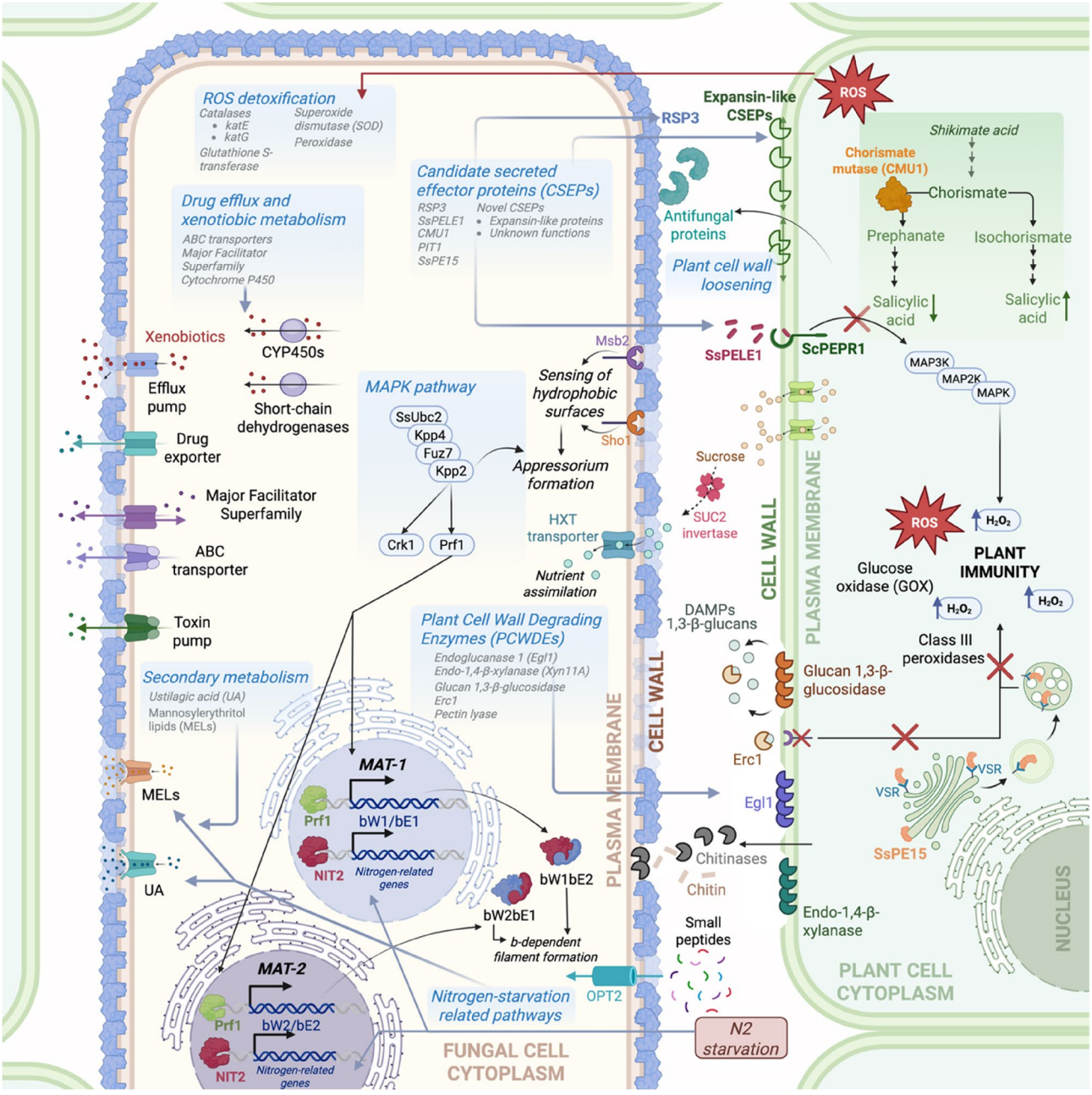

Fig. 8 Model illustrating strategies employed by *Sporisorium scitamineum* during infection of a smut-resistant sugarcane genotype. Simplified schematic representation of molecular events occurring during infection of a smut-resistant sugarcane genotype, based on transcriptional profiling at 48 h post-inoculation (hpi). Functional pathways are inferred from *S. scitamineum* homologs with characterized roles in other smut fungi systems. Host responses are illustrated based on previous studies [45]. Created in BioRender. Vilanova, P. (2025) https://biorender.com/g4wscp8

UPR activation is related to excess misfolded proteins in the endoplasmic reticulum (ER) [82]. In our dataset, we observed the induction of several UPR-related genes, including Pdi1 (*g639*) and Spc3 (*g3729*), both implicated in ER stress responses in fungi [83, 84]. UPR activation was previously linked to cell cycle progression and vesicle trafficking and the exact timing of UPR is required for virulence and controls the switch from budding to filamentous growth [85]. In *U. maydis*, the UPR synchronizes release from cell cycle arrest and expression of effector genes, since they contain UPR elements in their promoter regions and thus have their expression activated in an ER stress-dependent manner [86].

Other genes involved in iron acquisition and DNA replication were induced probably to sustain hyphal proliferation [87], alongside chromatin regulators ensuring genome integrity during rapid growth and fine-tuned transcriptional remodeling [88]. In general, we have seen

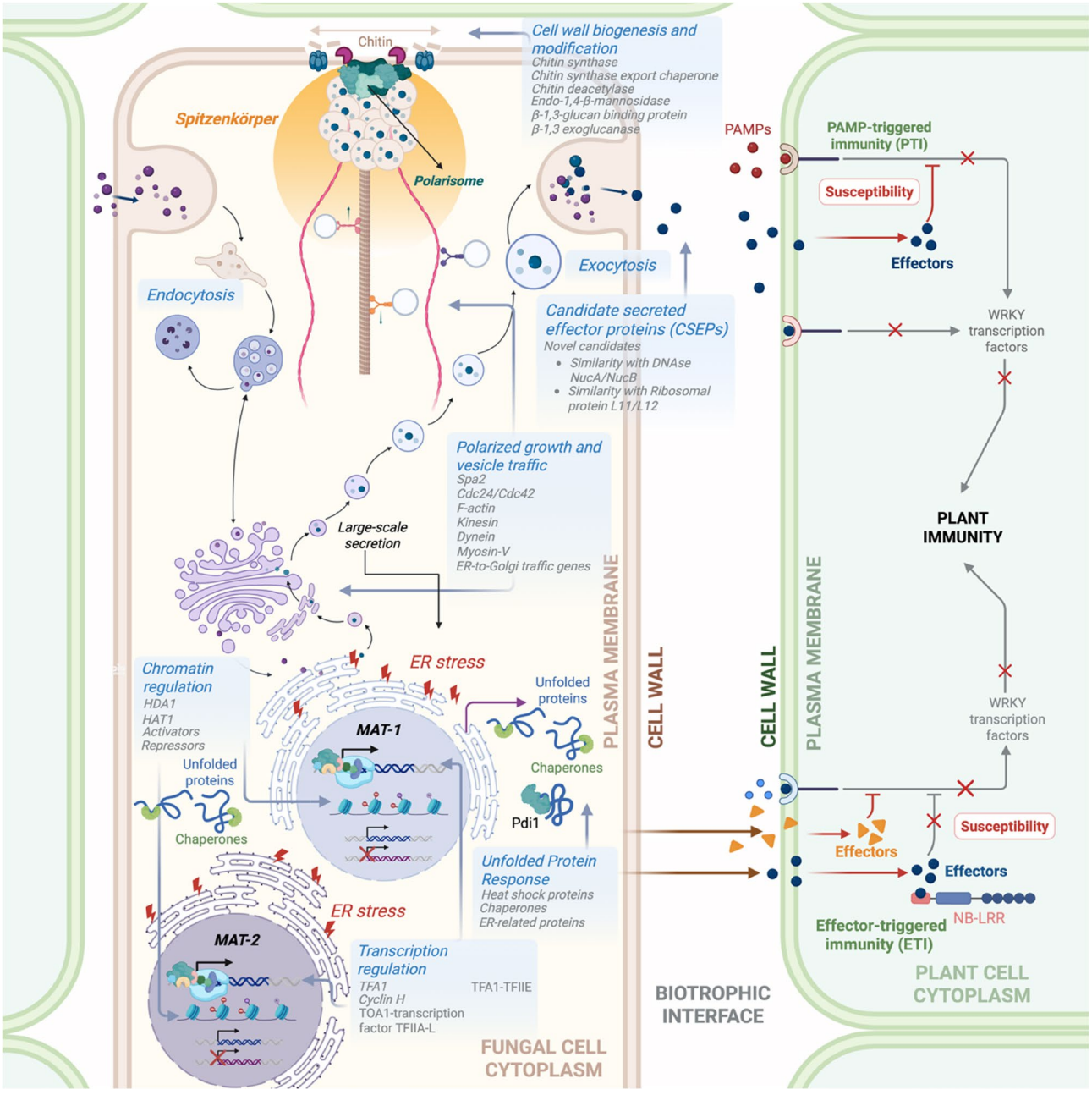

Fig. 9 Model illustrating strategies employed by *Sporisorium scitamineum* during infection of a smut-susceptible sugarcane genotype. Simplified schematic representation of molecular events occurring during infection of a smut-susceptible sugarcane genotype, based on transcriptional profiling at 48 h post-inoculation (hpi). Functional pathways are inferred from *S. scitamineum* homologs with characterized roles in other smut fungi systems. Suppression of host genotype immune responses is based on effector-triggered susceptibility (ETS). Created in BioRender. Vilanova, P. (2025) https://biorender.com/ve03s20

a group of molecular events related to fungal proliferation as the major effect in susceptible plants inoculated with SSC04 and we compiled these events in a model demonstrated in Fig. 9.

Host genotype influences effector gene expression in *S. scitamineum* isolates

While mating-associated genes provide the basis for sexual reproduction and dimorphic switching related to pathogenicity, effector diversification and their context-dependent expression are potential strategies for *S. scitamineum* to overcome host defenses. In this study, comparative transcriptomic analyses of two fungal

isolates with contrasting virulence profiles, SSC04 (more virulent) and SSC39 (less virulent), infecting the same sugarcane genotypes revealed host-genotype-specific expression patterns of candidate secreted effector proteins (CSEPs). Some of them were CSEPs with unknown functions, specific to *S. scitamineum*, that have been recurrently observed in inoculated plants across distinct sugarcane genotypes [40, 41]. For example, *g1052* and *g5159*, which were previously detected as differentially expressed in smut-resistant and -susceptible sugarcane in other studies had the ability to suppress PTI/ETI responses [40]. The effector repertoire also included homologs of known virulence factors from related smut fungi, suggesting conserved mechanisms for host manipulation. Notably, *g3970*, encoding the CSEP Rsp3 homolog, was strongly upregulated (200-fold) in resistant plants inoculated with SSC04. In *U. maydis*, Rsp3 interacts with Cpl1, as previously discussed, and protects fungal hyphae from host antifungal proteins [89]. Other homologs such as Cmu1, Pit1, and Suc2 were also detected and are known to modulate host metabolism and nutrient acquisition [68, 90]. Additionally, SsPele1 (*g1513*), previously shown to attenuate receptor-mediated immunity in *S. scitamineum* [10], was induced in our dataset. We also uncovered a pool of novel expressed candidate effectors *in planta* containing expansin-like domains, a feature similar to ceratoplatanin (CP) proteins. These CSEPs may be acting on mechanical penetration, chitin-binding to suppress chitin-triggered responses or fungal cell wall remodeling [91]. Importantly, the expression profiles of CSEPs differed between resistant and susceptible genotypes 48 h post-inoculation.

Conclusions

This work proposes a detailed overview of the complex interplay between *Sporisorium scitamineum* genetic diversity and sugarcane host responses, indicating critical insights of fungal pathogenicity. The comparative virulence assay of two isolates demonstrated the impact of haplotype-specific variation. Genomic analyses revealed high similarity among isolates, yet identified polymorphisms in coding regions, including effector genes and vesicular trafficking components, alongside with structural rearrangements in mating loci. Notably, two new findings suggest mechanisms driving pathogen diversification: the compact mating region in *S. scitamineum* strains shaped by TE-mediated rearrangements and presumed accessory-like chromosomes. Transcriptomic profiling highlighted host-dependent fungal strategies: the more virulent isolate SSC04 adopts stealthy tactics in resistant genotypes, upregulating detoxification and effector genes while shifting to aggressive hyphal proliferation and stress adaptation in susceptible hosts. Because resistant plants respond to pathogen perception, smut development *in planta* is delayed. Effector diversification, nutrient scavenging, and UPR pathway activation emerged as central to infection. Moreover, we uncovered various candidate effectors opening future perspectives for studies on their functional role in suppressing the host immune response. Although polymorphic genes among isolates frequently differed from those differentially regulated genes in our tested conditions, the same functional categories arose, suggesting that processes, such as vesicle trafficking, may be hallmarks of *S. scitamineum* plasticity. These findings emphasize the role of genetic variation and transcriptional flexibility in fungal virulence, which we should further explore to consider efficient smut management strategies.

Materials and methods

Fungal strains and mating-type confirmation

SSC04 and SSC39 teliospores, referred to as "BR_04" and "BR_39" by Benevenuto et al. (2016), were isolated following the methodology previously described. SSC04 *MAT-1* (formerly BR_04B) and SSC04 *MAT-2* (BR_04A) haploid cells were deposited in The Environmental and Health Fungi Collection of Oswaldo Cruz Foundation (FIOCRUZ) (CFAS 40478 e CFAS 40479, respectively), whereas SSC39 *MAT-1* (BR_39B) and SSC39 *MAT-2* (BR 39 A) were deposited previously (INCQS 40413 and INCQS 40412, respectively) (Taniguti et al., 2015) [17]. Mating-type *in vitro* confirmation was conducted according to Taniguti et al., (2015) [17] in YM solid medium (3 g/L of yeast extract, 3 g/L of malt extract, 5 g/L of peptone, 10 g/L of glucose, 17 g/L agar) (Additional file 9: Fig. S1). Molecular determination of mating-type was performed according to Agisha et al. (2022) [32] (Additional file 9: Table S1). A schematic representation of the sexual mating process can be seen in Additional file 9: Fig. S2. Haplotype confirmation is described in Additional file 9: Table S2.

Fungal virulence evaluation

We used 10-month-old sugarcane plants of the susceptible genotype RB925345, kindly provided by the RIDESA Breeding Program. SSC04 and SSC39 teliospores were used in the fungal virulence evaluation. A total of 180 single bud sets were disinfected, according to Taniguti et al. (2015) [17], and placed in a vermiculite-filled tray. Following Benevenuto et al. (2016) [30], fungal teliospores obtained with a germination rate of >= 70% were prepared in a suspension (10^{-7} teliospores. mL^{-1} in 0.85% sterile saline with 0.01% tween) for inoculation. Controls were mock-inoculated with saline and tween. Inoculation of sugarcane single-bud sets followed the procedure described by Carvalho et al. (2016) [44]. Trays were kept in a growth chamber at 29 °C for seven days before

transplanting each bud to a 3 L pot filled with BasePlant substrate and transferring to a greenhouse. The experiment followed a randomized block design with three replicates (Additional file 1: Fig. S2) and occurred between October and January. Plants were evaluated monthly for the following traits: plant height, culm diameter, and the number of tillers. For whip emission, we assessed daily. All plants were measured from the substrate to leaf + 1 to determine plant height. Disease incidence was calculated as the proportion of plants that emitted whips relative to the total number of plants assessed [4, 42, 92].

A mixed-effect ANOVA followed by Tukey's test ($p < 0.05$) to determine statistical differences for each target trait was used. Normality of the data was assessed using the Shapiro-Wilk test [93]. We examined the fixed effects of isolate, time, and interaction (isolate × time), and incorporated random effects associated with the different blocks. Statistical analyses were conducted in R v.4.2.1 (http://www.R-project.org/) using *tidyr* v.1.3.0 and *emmeans* v.1.8.4 packages.

qPCR-based quantification of fungal DNA in infected plants

We conducted an experiment with a similar design as described above but using qPCR to quantify fungal growth of both isolates at five time points: 0, 48, 72, and 120 h after-inoculation. We used a pool of 6 culms from the RB925345 sugarcane genotype per biological replicate. We used three biological replicates and three technical replicates for three different treatments: plants inoculated with more-virulent SSC04, less-virulent SSC39 and mock-inoculation. We inoculated plants by bud-piercing, as described in Carvalho et al. (2016) [44].

DNA was extracted from 100 mg of infected plant tissues (buds) using the CTAB method [94]. Fungal genomic DNA was obtained from teliospores using a modified Doyle & Doyle (1987) protocol [30] and used to establish the standard curve. DNA concentration and quality were verified in NanoDrop® 2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and by electrophoresis. Detection and quantification were performed in a 7300 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA) using GoTaq® qPCR Master Mix (Promega Corporation). The Intergenic Spacer (IGS) of *S. scitamineum* ribosomal locus was used as a target for pathogen detection (FP: 5'-CGGCTATTGTCGCACATCTC-3'; RP: 5'-CCAAACG CAGGTCACAGTCT-3') [45]. Standard regression lines were obtained from the fungal DNA dilution series by plotting the threshold cycle (Ct) versus logarithmic values of known DNA concentrations. Fungal DNA quantity was calculated in infected plants at each time point in 100 ng of total DNA (Additional file 1: Fig. S1) regarding the virulence assay. Levene's test was used to evaluate the homogeneity of variances between samples to proceed to Student's t-test to assess the statistical significance of the observed differences ($p < 0.05$). All tests were performed using the *scipy.stats* module from SciPy v.1.11.4 [95].

We conducted a second experiment using SSC04 inoculated in a smut-susceptible sugarcane genotype IAC66-6 and a smut-resistant genotype SP80-3280, both of which are widely used as model genotypes for studying sugarcane smut responses [39, 44, 45]. The experiment was conducted in three replicates, combining ten buds of inoculated plants per replicate and three replicates of ten buds for mock-inoculated plants. We used qPCR to quantify fungal growth in both sugarcane genotypes 48 after inoculation, as described above (Additional file 7: Fig. S1).

DNA extraction, library preparation, and sequencing strategies

A modified Doyle & Doyle (1987) [94] protocol and the Genomic-tip 20/G (QIAGEN, Hilden, Germany) kit were used to extract DNA for short read Illumina platform paired-end sequencing and Oxford Nanopore MinION platform (Oxford, UK), respectively. Library construction was performed using the Illumina DNA Prep kit, followed by sequencing on the NextSeq 2000 Sequencing System with 2 × 150 bp runs at the Functional Genomics Center, ESALQ, University of São Paulo, Piracicaba, Brazil. For long reads, MinION libraries and sequencing were performed in-house using the Ligation kit (SQK-LSK109) (Oxford Nanopore Technologies, Oxford, UK), NEBNext Ultra II End-prep reaction/FFPE DNA Repair (New England Biolabs, Ipswich, MA, USA), Ligation Buffer (LNB), NEBNext Quick T4 DNA Ligase, and Adapter Mix (AMX), following manufacturer's instructions. The final library was quantified and used to load the flow cell (MinION Flow Cells (R9.4.1). We used Guppy v. 6.2.1 (Oxford Nanopore Technologies, Oxford, UK) super accuracy mode for base-calling long-read sequencing. Additional file 10: Table S1 shows data obtained from each sequencing platform and SRA accession numbers to raw data.

Sanger sequencing of polymorphic candidate effectors among Brazilian and Argentinian strains

DNA from *S. scitamineum* strains from different Brazilian states and Argentina was obtained according to Benevenuto et al. (2016) [30]. Specific primers for annealing to candidate genes of interest were designed using the *S. scitamineum* reference genome [17] and can be found in Additional file 6: Table S1. Gene amplification was performed using Taq DNA Polymerase (Thermo Fisher Scientific Inc). Reactions were conducted in a Veriti® 96-Well thermocycler (AppliedBiosystems, Waltham, MA) and PCR products purified with the Wizard® SV Gel and PCR Clean-Up System kit (Promega Corporation).

The sequences obtained were assembled using the Sequencher v.5.0.0 software (Gene Codes Corporation) and subsequently aligned in the CLC Genomics Workbench v.20.0.4 (QIAGEN) using reference genomes from Brazil, Australia [17], China [18] and South Africa [12] to detect polymorphic sites (Additional file 6: Fig. S1–4).

Comprehensive genome assembly and annotation workflow

De novo assembly using Nanopore reads was conducted with Canu v.2.2 [96] with "genomeSize = 20 M" and "-nanopore-raw" parameters. Illumina short-reads were used to polish the genomes with Medaka v.1.2.2 and Pilon v.1.24. We used the genome reference from Taniguti et al. (2025) [97] referred as "SSC39 MAT-1 v.2" for variant discovery and gene annotation using the FunGAP [98] pipeline to predict genes using RNA-seq data from *in vitro* expression studies [17] and *U. maydis* as a sister proteome [99]. Genome completeness was verified by BUSCO v.5.4.7 using the "basidiomycota_odb10" database (1,764 orthologs) as a reference [100]. For comparative purposes, *ab initio* gene prediction was executed with Augustus v.3.3.3 [101] in all assemblies. We kept the original annotation of Taniguti et al. (2015) [17] as a reference for further functional annotation and expression analysis. Canonical telomeric repeats in the assembled contigs were detected using the Tandem Repeat Finder Tool (https://tandem.bu.edu/), the Telomere Identification toolKit (tidk) (https://github.com/tolkit/telomeric-identifier) with the canonical sequence "TTAGGG" [17] and were also manually inspected. Average nucleotide identity was assessed using the OrthoANIu algorithm [102] and dotPlotly (https://github.com/tpoorten/dotPlotly) was used to produce dot plots comparing assemblies. We used barrnap v.1.2.2 (https://github.com/tseemann/barrnap) and tRNAscan v.0.4 with the "eukaryotic" option, as implemented in the Galaxy web platform (usegalaxy.org) [103] to scan genomes for rDNA copies and tRNAs, respectively.

Discovery and functional annotation of variants

High-quality reads were aligned to the *S. scitamineum* SSC39 MAT-1 v.2 reference genome using the Burrows-Wheeler Aligner (BWA-MEM) 0.7.17-r1188 with default parameters [104]. Aligned reads were sorted and indexed using SAMtools *sort* v.1.3.1 [105]. Variant calling was performed using the Genome Analysis Toolkit (GATK) v4.5.0.0 with the ploidy parameter set to one [106], considering that the DNA was extracted from haploid cells. Pipelines are available at GitHub (https://github.com/lmtani/s-scitamineum-pipelines). To create a diploid phased VCF, we used a custom script also available at GitHub (https://github.com/lmtani/synthetic-diploid). This script combines variants from two haploid VCF files into a single diploid one, ensuring accurate phasing. We utilized SnpEff v.5.2c [107] to functionally annotate variants from the diploid VCF files, using gene predictions generated with extrinsic evidence by FunGAP as previously described. The SnpSift toolbox [107] was then employed to parse these variant annotations.

Functional annotation and mating-type analyses

To improve the reference data, the predicted SSC39 *MAT-1* proteome [17] was reanalyzed using PfamScan with Pfam v.35.0 [108] and InterProScan v.5.55.88 [109] with the following parameters: SMART v.7.1, SUPERFAMILY v.1.75, CDD v.3.18, TIGRFAM-15.0, Pfam v.34.0, and Gene3D v.4.3.0. Distributions of euKaryotic Orthologous Group (KOG) terms were assigned using the eggNOG-mapper [110] online tool against the eggNOG v.5.0 database [111]. Carbohydrate-active enzymes (CAZymes) were predicted using the dbCAN2 HMM profile database v.7.0 [112] and hmmscan from HMMER v.3.1b2 (http://hmmer.org). Lipases were predicted using HMM profiles from the Lipase Engineering Database v.3.0 [113] and hmmscan. Significant matches for CAZymes and lipases were parsed using thresholds set as e-value = 1e-4 and coverage = 0.35. Proteases were predicted using the predicted proteins as queries and the MEROPS database v.12.1 [114] with BLASTp run locally (thresholds set as e-value = 1e-4, minimum identity = 0.40, coverage = 0.85). The predicted secretome was defined by a signal peptide and absence of transmembrane domains using SignalP v.4.1 and Phobius v.1.01, and subcellular localization was predicted using TargetP v.2.0 and EffectorP v.3.0. Finally, OrthoFinder v.2.5.4 [115] was employed to establish the orthologous relationships among the predicted proteomes.

Genes of mating-type loci of all four genomes were manually examined through the Integrative Genomics Viewer (IGV) [116] using gene predictions from Augustus v.3.3.3 [101]. and aligned RNA-seq expression data. We used BLASTx and ExPASy [117] for unpredicted ORFs. Repeated elements were manually analyzed to determine their lengths and potential relevance for rearrangements.

RNA extraction and transcriptome sequencing of *in vitro* fungal cultures

Total RNA was extracted from mating-compatible haploid cells from the SSC04 isolate individually grown in a YM liquid medium (3 g/L of yeast extract, 3 g/L of malt extract, 5 g/L of peptone, 10 g/L of glucose) for 15 h at 28 °C under 200 rpm. Cells were pooled and centrifuged at 8,000 rpm for 15 min. The concentration of each mating-type sample was stoichiometrically adjusted. TRIzol®Plus RNA Purification Kit (Life Technologies, Carlsbad, CA, USA) protocol was used, followed by a

DNAse I treatment (Sigma Aldrich, St. Louis, MO, USA). All samples were evaluated for RNA integrity both by RNA integrity number (RIN) ≥ 8 (Agilent 2100 Bioanalyzer, Agilent Technologies, USA) and gel electrophoresis (Additional file 7: Fig. S2) and quantified using a Qubit 4 fluorometer (Life Technologies, Thermo Fisher Scientific, USA). Paired-end RNA libraries were prepared according to the manufacturer's instructions for the Illumina® Stranded mRNA Prep Ligation kit, and RNA paired-end sequencing was performed by NextSeq 2000 (Illumina) in 2 × 100 bp runs at the Functional Genomics Center, ESALQ, University of São Paulo, Piracicaba, BR. Approximately 30 million reads per replicate were obtained (Additional file 7: Table S1).

RNA extraction and sequencing of inoculated smut-resistant and smut-susceptible sugarcane

S. scitamineum isolate SSC04 teliospores suspended in saline solution with 0.1% Tween 20, at a final concentration of 10^{-7} teliospores.mL^{-1} (>70% germination rate) were used to inoculate single-bud sets of smut-resistant (SP80-3280) and smut-susceptible (IAC66-6) sugarcane, according to Taniguti et al. (2015) [17]. The experiment was conducted in three replicates of ten buds for both inoculated and mock-inoculated plants. Forty-eight hours post-inoculation buds were collected in liquid nitrogen for RNA extraction with a similar experiment design as described by Rody et al. (2019) [39]. Total RNA was extracted using the LiCl-based method [118]. RNA concentration was determined using a NanoDrop® 2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and the Qubit RNA HS Assay Kit with Qubit 4 fluorometer (Life Technologies, Thermo Fisher Scientific, USA) fluorometric method. The RNAseq experiment used a depletion of rRNA from the fungus and the plant before library preparation and sequencing of total RNA, according to Morlan et al. (2012) [119] and Miranda et al. (2023) [120]. TruSeq Stranded mRNA library prep, Ligation kit, and RNA was sequenced on the Illumina NovaSeq 6000 platform at Genoma (Centro de Estudos do Genoma Humano e Células Tronco), USP, São Paulo, BR to produce 2 × 100 bp. Data regarding RNAseq experiments can be found in Additional file 7: Table S1.

Differential expression analysis

FastQC v.0.11.5 [121] was used to assess the quality of sequencing reads. Low-quality reads (Q < 20), reads containing ambiguous bases (N >0), and library adapters were removed using Cutadapt v.4.9 [122]. Transcripts were mapped to the reference genome SSC39 *MAT-1* [17] using Hisat2 v.2.1.0 [123]. Read counts were generated with featureCounts v.2.0.6 from the Subread package [124]. Genes with at least one count per million (CPM) in all three biological replicates of a given treatment were considered expressed. Genes detected uniquely in one condition were defined as those with at least one CPM in all replicates of a specific treatment and no detectable counts in the compared treatment (Additional fFile 11: Tables S1-9).

Differentially expressed genes (DEGs) *in planta* were analyzed within sequencing batches for SSC04 and SSC39 to minimize batch effect biases. DEGs were identified using the likelihood ratio test in edgeR v.4.0.16 [125, 126] from R Bioconductor. For *in planta* comparisons (fungus infecting a smut-resistant x smut-susceptible genotypes), DEGs were divided into two lists: one accepting $p < 0.05$ and $|\log_2 FC| > 1$, and another with more stringent criteria of False Discovery Rate (FDR) < 0.05 and $|\log_2 FC| > 1$. Finally, DEGs from all experiments were cross-referenced to the reference annotation [17] using reference gene IDs for functional comparisons.

Abbreviations

ABC	ATP-binding cassette efflux transporters
ANI	Average nucleotide identity
CP	Cerato-platanin proteins
CSEP	Candidate secreted effector protein
CWDEs	Cell wall-degrading enzymes
ER	Endoplasmic reticulum
ETS	Effector-triggered susceptibility
hpi	Hours post-inoculation
HPI	Highly polymorphic islands
LV	Less-virulent
MELs	Mannosylerythritol lipids
MFS	Major Facilitator Superfamily
MV	More-virulent
PRE	Pheromone response element
ROS	Reactive oxygen species
SNV	Single nucleotide variant
T2T	Telomere-to-telomere
TE	Transposable element
UA	Ustilagic acid
UPR	Unfolded protein response

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12160-1.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Supplementary Material 8.

Supplementary Material 9.

Supplementary Material 10.

Supplementary Material 11.

Supplementary Material 12. Additional file 1: Virulence assay experimental design and qPCR quantification. Additional file 2: Assembly statistics and structural variants. Additional file 3: Telomeric annotation. Additional file

4: Comparative genomics analysis. Additional file 5: Pheromone sequence analysis. Additional file 6: Candidate effector polymorphic sites in Brazilian and Argentinian isolates.Additional file 7: Transcriptome analysis. Additional file 8: Gene expression and differential expression analysis (DEA). Additional file 9: Mating-type ***in vitro*** and molecular confirmation. Additional file 10: DNA sequencing. Additional file 11 : Exclusively detected genes in each condition

Acknowledgements
Acknowledgments: We thank Elaine Vidotto Batista (Genomics Group Lab, ESALQ/USP) and Tatiana Caroline Silveira Corrêa (GaTE Lab, USP) for technical support. We also acknowledge Tatiana Caroline Pisetta (GaTE Lab) and Paula Turrini (Molecularis Biotecnologia) for RNA-seq library preparation. Infrastructure was provided by GaTE Lab and the Centro de Facilidades de Apoio à Pesquisa (CEFAP/USP). RNA-seq was performed at the Centro de Estudos do Genoma Humano e Células-Tronco (USP), which we gratefully acknowledge.

Authors' contributions
CBM-V conceived the study and edited the manuscript. PFV performed the experiments, analyzed the data and wrote the manuscript. LMT, TCSDS, RGHB participated in the data analysis. MF, LSO, GSC, TCP, PT participated in experiments. SC contributed with plant materials. JPK, SC, LEAC and M-AVS provided scientific expertise and manuscript editing. All authors reviewed and approved the final version of the manuscript.

Funding
This study was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – 2016/17545–8; 2022/03965). Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) supported CBMV (CNPq 395961/2021-7; 405314/2021-3); LEAC (CNPq 305542/2021-4); MAVS (CNPq). Fellowships to PFV (FAPESP 2023/13474-2). Also, this study was financed in part by the Coordenação de
Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Mendelics provided support in the form of a salary for authors LMT and JPK. These funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability
The datasets generated during the current study are available in the NCBI BioProject database (http://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA998034. Accession numbers of other analyzed data are available in Additional file 7: Table S2. Code for the reproduction of the analyses within this paper is available on GitHub (https://github.com/lmtani/s-scitamineum-pipelines and https://github.com/pedrofvilanova/sscitamineum_omics).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil
[2]Mendelics, São Paulo, São Paulo, Brazil
[3]Sugarcane Center, IAC, Ribeirão Preto, São Paulo, Brazil
[4]Department of Phytopathology and Nematology, Luiz de Queiroz College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil
[5]GaTE Lab (Genomics and Transposable Elements Laboratory), Department of Botany, Institute of Biosciences, University of São Paulo, São Paulo, São Paulo, Brazil

Received: 25 April 2025 / Accepted: 29 September 2025
Published online: 05 November 2025

References

1. Ökmen B, Jaeger E, Schilling L, Finke N, Klemd A, Lee YJ, et al. A conserved enzyme of Smut fungi facilitates cell-to-cell extension in the plant bundle sheath. Nat Commun. 2022;13(1). https://doi.org/10.1038/s41467-022-33815-7.
2. van der Linde K, Göhre V. How do Smut fungi use plant signals to Spatiotemporally orientate on and in planta? J Fungi. 2021;7(2). https://doi.org/10.3390/jof7020107.
3. Piepenbring M, Stoll M, Oberwinkler F. The generic position of *Ustilago maydis*, *Ustilago scitaminea*, and *Ustilago esculenta* (Ustilaginales). Mycological Progress. 2002;1(1):71–80. https://doi.org/10.1007/s11557-006-0006-y.
4. Rajput MA, et al. Screening of sugarcane germplasm against *Sporisorium Scitamineum* and its effects on setts germination and tillering. Sci Rep. 2024;14(1). https://doi.org/10.1038/s41598-024-64810-1.
5. Bhuiyan SA, Magarey RC, McNeil MD, Aitken KS. Sugarcane Smut, caused by *Sporisorium scitamineum*, a major disease of sugarcane: A contemporary review. Phytopathology®. 2021;111(11):1905–17. https://doi.org/10.1094/phyto-05-21-0221-rvw.
6. Sundar AR, Barnabas EL, Malathi P, Viswanathan R. A Mini-Review on Smut disease of sugarcane caused by Sporisorium Scitamineum. Botany. 2012;5:107–28.
7. Rajput MA, Rajput NA, Syed RN, Lodhi AM, Que Y. Sugarcane smut: current knowledge and the way forward for management. J Fungi. 2021;7(12). https://doi.org/10.3390/jof7121095.
8. Yu C, Wang DC, Han H, Wang P, Liu C. Progress in pathogenesis research of *Ustilago maydis*, and the metabolites involved along with their biosynthesis. Mol Plant Pathol. 2023. https://doi.org/10.1111/mpp.13307.
9. Steins L, Duhamel M, Klenner-Koch S, Begerow D, Kemler M. Resources and tools for studying convergent evolution in different lineages of Smut fungi. Mycological Progress. 2023;22(11). https://doi.org/10.1007/s11557-023-01918-0.
10. Ling H, Fu X, Huang N, Zhong Z, Su W, Lin W, et al. A sugarcane Smut fungus effector simulates the host endogenous elicitor peptide to suppress plant immunity. New Phytol. 2022;233(2). https://doi.org/10.1111/nph.17835.
11. Nalayeni K, Ashwin NMR, Barnabas L, Vinodhini T, Agisha VN, Sundar AR, et al. Comparative expression analysis of potential pathogenicity-associated genes of high- and low-virulent *Sporisorium Scitamineum* isolates during interaction with sugarcane. 3 Biotech. 2021;11(7). https://doi.org/10.1007/s13205-021-02893-7.
12. Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Münsterkötter M, Güldener U, et al. A Tale of genome compartmentalization: the evolution of virulence clusters in Smut fungi. Genome Biol Evol. 2016;8(3):681–704. https://doi.org/10.1093/gbe/evw026.
13. Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R. *Ustilago Maydis* as a pathogen. Annu Rev Phytopathol. 2009;47(1):423–45. https://doi.org/10.1146/annurev-phyto-080508-081923.
14. Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago Maydis*. Nature. 2006;444(7115):97–101. https://doi.org/10.1038/nature05248.
15. Depotter JRL, Ökmen B, Ebert MK, Beckers J, Kruse J, Thines M, et al. High nucleotide substitution rates associated with retrotransposon proliferation drive dynamic secretome evolution in Smut pathogens. Microbiol Spectr. 2022;10(5). https://doi.org/10.1128/spectrum.00349-22.
16. Ullmann L, Wibberg D, Busche T, Rückert C, Müsgens A, Kalinowski J, et al. Seventeen ustilaginaceae High-Quality genome sequences allow phylogenomic analysis and provide insights into secondary metabolite synthesis. J Fungi. 2022;8(3):269. https://doi.org/10.3390/jof8030269.
17. Taniguti LM, Schaker PDC, Benevenuto J, Peters LP, Carvalho G, Palhares A, et al. Complete genome sequence of *Sporisorium Scitamineum* and biotrophic interaction transcriptome with sugarcane. PLoS ONE. 2015;10(6). https://doi.org/10.1371/journal.pone.0129318.
18. Que Y, Xu L, Wu Q, Liu Y, Ling H, Liu Y, et al. Genome sequencing of *Sporisorium Scitamineum* provides insights into the pathogenic mechanisms of sugarcane Smut. BMC Genomics. 2014;15(1):996. https://doi.org/10.1186/1471-2164-15-996.
19. Hartmann FE. Using structural variants to understand the ecological and evolutionary dynamics of fungal plant pathogens. New Phytol. 2021;234:43–9. https://doi.org/10.1111/nph.17907.

20. Carpentier F, Rodríguez RC, Branco S, Alodie Snirc, Coelho MA, Hood ME, et al. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. Genome Res. 2019;29(6):944–53. https://doi.org/10.1101/gr.242578.118.
21. Benevenuto J, Teixeira-Silva NS, Kuramae EE, Croll D, Monteiro-Vitorello CB. Comparative genomics of Smut pathogens: insights from orphans and positively selected genes into host specialization. Front Microbiol. 2018;9. https://doi.org/10.3389/fmicb.2018.00660.
22. Bakkeren G, Kämper J, Schirawski J. Sex in Smut fungi: Structure, function and evolution of mating-type complexes. Fungal Genet Biol. 2008;45:S15–21. https://doi.org/10.1016/j.fgb.2008.04.005.
23. Thushari ANWS, De Costa DM. Molecular and genetic variability of *Sporisorium Scitamineum* (Sugarcane Smut Pathogen) in sugarcane plantations in Sri Lanka. Sugar Tech. 2023. https://doi.org/10.1007/s12355-022-01239-8.
24. Barnabas L, Ashwin N, Nalayeni K, Sundar AR, Malathi P, Viswanathan R. Genetic and pathogenic variability among the Indian isolates of *Sporisorium Scitamineum* causing sugarcane Smut. J Sugarcane Res. 2019;8:138–54.
25. Que YK, Xu L, Lin J, Chen R, Grisham M. Molecular variation of *Sporisorium Scitamineum* in Mainland China revealed by RAPD and SRAP markers. 2012;96(10):1519–25. https://doi.org/10.1094/pdis-08-11-0663-re.
26. Raboin LM, Selvi A, Oliveira KM, Paulet F, Calatayud C, Zapater MF, et al. Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago Scitaminea*. Fungal Genet Biol. 2007;44(1):64–76. https://doi.org/10.1016/j.fgb.2006.07.004.
27. Lee C, Yuan C, Liang Y. Occurrence of a new pathogenic race of culmicolous smut of sugarcane in Taiwan. Proc Int Soc Sugarcane Technol. 1999;23:406–7.
28. Comstock JC. Hawaii's approach to control of sugarcane Smut. Plant Dis. 1983;67(4):452. https://doi.org/10.1094/pd-67-452.
29. Bhuiyan SA, Croft BJ, Stringer JK, Deomano EC. Pathogenic variation in spore populations of *Sporisorium scitamineum*, causal agent of sugarcane Smut in Australia. Plant Dis. 2015;99(1):93–9. https://doi.org/10.1094/pdis-12-13-1257-re.
30. Benevenuto J, Longatto DP, Reis G, Mielnichuk N, Palhares AC, Carvalho G et al. Molecular variability and genetic relationship among Brazilian strains of the sugarcane Smut fungus. 2016;363(24). https://doi.org/10.1093/femsle/fnw277.
31. Rago AM, et al. Variabilidade patogênica de *Ustilago Scitaminea* no Estado de São Paulo. Summa Phytopathologica. 2009;35(2):93–7. https://doi.org/10.1590/s0100-54052009000200002.
32. Agisha VN, et al. Transcriptome analysis of sugarcane reveals differential switching of major defense signaling pathways in response to *Sporisorium Scitamineum* isolates with varying virulent attributes. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.969826.
33. Bedre B, Irigoyen S, Carvalho D, Monteiro-Vitorello CB, Silva JA, Mandadi KK. Genome-wide alternative splicing landscapes modulated by biotrophic sugarcane Smut pathogen. Sci Rep. 2019;9(1). https://doi.org/10.1038/s41598-019-45184-1.
34. McNeil MD, Bhuiyan SA, Berkman PJ, Croft BJ, Aitken KS. Analysis of the resistance mechanisms in sugarcane during *Sporisorium Scitamineum* infection using RNA-seq and microscopy. PLoS ONE. 2018;13(5):e0197840. https://doi.org/10.1371/journal.pone.0197840.
35. Schaker PDC, Palhares AC, Taniguti LM, Peters LP, Creste S, Aitken KS, et al. RNAseq transcriptional profiling following whip development in sugarcane Smut disease. PLoS ONE. 2016;11(9). https://doi.org/10.1371/journal.pone.0162237.
36. Que Y, Su Y, Guo J, Wu Q, Xu L. A Global View of Transcriptome Dynamics during *Sporisorium scitamineum* Challenge in Sugarcane by RNA-seq. Fu B, editor. PLoS One. 2014;9(8). https://doi.org/10.1371/journal.pone.0106476.
37. Yan M, Dai W, Cai E, Deng YZ, Chang C, Jiang Z, et al. Transcriptome analysis of *Sporisorium Scitamineum* reveals critical environmental signals for fungal sexual mating and filamentous growth. BMC Genomics. 2016;17(1). https://doi.org/10.1186/s12864-016-2691-5.
38. Naidoo S, Visser EA, Zwart L, Toit Y, du, Bhadauria V, Shuey LS. Dual RNA-Sequencing to elucidate the Plant-Pathogen duel. Curr Issues Mol Biol. 2018. https://doi.org/10.21775/cimb.027.127.
39. Rody HVS et al. Genome survey of resistance gene analogs in sugarcane: genomic features and differential expression of the innate immune system from a smut-resistant genotype. 2019;20(1). https://doi.org/10.1186/s12864-019-6207-y.
40. Maia T, Rody HVS, Bombardelli RGH, Souto TG, Camargo LEA, Monteiro-Vitorello CB. A bacterial type three Secretion-Based delivery system for functional characterization of *Sporisorium Scitamineum* plant immune suppressing effector proteins. Phytopathology. 2022;112(7):1513–23. https://doi.org/10.1094/PHYTO-08-21-0326-R.
41. Teixeira-Silva NS, Schaker PDC, Rody HVS, Maia T, Garner CM, Gassmann W, et al. Leaping into the unknown world of *Sporisorium Scitamineum* candidate effectors. J Fungi. 2020;6(4):339. https://doi.org/10.3390/jof6040339.
42. Santos LF, Silva FB, Anjos LOS, Júnior LLR, Anjos IA, Fernandes TC, Silva MF, Perecin D, Goes A, Pinto LR. Screening of sugarcane genotypes for Smut (*Sporisorium scitamineum*) resistance under greenhouse conditions. Agronomy. 2025;15(2):448–8. https://doi.org/10.3390/agronomy15020448.
43. Latiza AS, Ampusta DC, Rivera J.R, et al. Reaction of sugarcane clones to strain B of *Sporisorium Scitamineum*. International Society of Sugarcane Technologists.
44. Carvalho G, Quecine MC, Longatto DP, Peters LP, Almeida JR, Shyton TG, et al. *Sporisorium Scitamineum* colonisation of sugarcane genotypes susceptible and resistant to Smut revealed by GFP-tagged strains. Ann Appl Biol. 2016;169(3):329–41. https://doi.org/10.1111/aab.12304.
45. Peters LP, Carvalho G, Vilhena MB, Creste S, Azevedo RA, Monteiro-Vitorello CB. Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. Planta. 2017;245(4):749–64. https://doi.org/10.1007/s00425-016-2642-z.
46. Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan KC, Hane JK. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in Plant-Pathogenic fungi. Mol Plant-Microbe Interactions®. 2018;31(8):779–88. https://doi.org/10.1094/MPMI-06-17-0135-FI.
47. Witte TE, Villeneuve N, Boddy CN, Overy DP. Accessory Chromosome-Acquired secondary metabolism in plant pathogenic fungi: the evolution of biotrophs into Host-Specific pathogens. Front Microbiol. 2021;12. https://doi.org/10.3389/fmicb.2021.664276.
48. Zhong Z, Chen M, Lin L, Chen R, Liu D, Norvienyeku J, et al. Genetic variation bias toward noncoding regions and secreted proteins in the rice blast fungus *Magnaporthe oryzae*. mSystems. 2020;5(3). https://doi.org/10.1128/mSystems.00346-20.
49. Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, et al. Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and Understanding evolutionary and mechanistic causes. New Phytol. 2020;229(5):2470–91. https://doi.org/10.1111/nph.17039.
50. Narayanan A, Reza MH, Sanyal K. Behind the scenes: Centromere-driven genomic innovations in fungal pathogens. PLoS Pathog. 2024;20(3):e1012080–0. https://doi.org/10.1371/journal.ppat.1012080.
51. Coelho MA, Ianiri G, David-Palma M, Theelen B, Goyal R, Narayanan A et al. Frequent transitions in mating-type locus chromosomal organization in *Malassezia* and early steps in sexual reproduction. Proceedings of the National Academy of Sciences. 2023;120(32). https://doi.org/10.1073/pnas.2305094120.
52. Sun S, Yadav V, Billmyre BR, Cuomo CA, Minou Nowrousian, Wang L, et al. Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. PLoS Biol. 2017;15(8):e2002527–7. https://doi.org/10.1371/journal.pbio.2002527.
53. Ye Z, Pan Y, Zhang Y, Cui H, Jin G, McHardy AC, et al. Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome. DNA Res. 2017;24(6):635–48. https://doi.org/10.1093/dnares/dsx031.
54. Laurie JD, Ali S, Linning R, Mannhaupt G, Wong P, Güldener U, et al. Genome comparison of barley and maize Smut fungi reveals targeted loss of RNA Silencing components and species-specific presence of transposable elements. Plant Cell. 2012;24(5):1733–45. https://doi.org/10.1105/tpc.112.097261.
55. Gross SD, Anderson RA, Casein Kinase I. Cell Signal. 1998;10(10):699–711. https://doi.org/10.1016/s0898-6568(98)00042-4.
56. Hoekstra MF, Dhillon N, Carmel G, DeMaggio AJ, Lindberg RA, Hunter T, et al. Budding and fission yeast casein kinase I isoforms have dual-specificity protein kinase activity. Mol Biol Cell. 1994;5(8):877–86. https://doi.org/10.1091/mbc.5.8.877.
57. Castillo-Lluva S, Alvarez-Tabarés I, Weber I, Steinberg G, Pérez-Martín J. Sustained cell Polarity and virulence in the phytopathogenic fungus *Ustilago Maydis* depends on an essential cyclin-dependent kinase from the Cdk5/Pho85 family. J Cell Sci. 2007;120(9). https://doi.org/10.1242/jcs.005314.
58. García-Muse T, Steinberg G, Pérez-Martín J. Pheromone-Induced G_2Arrest in the phytopathogenic fungus *Ustilago Maydis*. Eukaryot Cell. 2003;2(3):494–500.

59. Mullins RD, Pollard TD. Structure and function of the Arp2/3 complex. Curr Opin Struct Biol. 1999;9(2):244–9. https://doi.org/10.1016/s0959-440x(99)80034-7.
60. Aparicio OM. Location, location, location: it's all in the timing for replication origins. Genes Dev. 2013;27(2):117–28. https://doi.org/10.1101/gad.209999.112.
61. Haber JE. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. Genetics. 2012;191(1):33–64. https://doi.org/10.1534/genetics.111.134577.
62. Douglas LM, Wang HX, Keppler-Ross S, Dean N, Konopka JB. Sur7 Promotes Plasma Membrane Organization and Is Needed for Resistance to Stressful Conditions and to the Invasive Growth and Virulence of *Candida albicans*. Berman J, editor. mBio. 2012;3(1). https://doi.org/10.1128/mBio.00254-11
63. Chen X, Ebbole DJ, Wang Z. The exocyst complex: delivery hub for morphogenesis and pathogenesis in filamentous fungi. Curr Opin Plant Biol. 2015;28. https://doi.org/10.1016/j.pbi.2015.09.003.
64. Giraldo MC, Dagdas YF, Gupta YK, Mentlak TA, Yi M, Martinez-Rocha AL, et al. Two distinct secretion systems facilitate tissue invasion by the rice blast fungus Magnaporthe oryzae. Nat Commun. 2013;4(1):1996. https://doi.org/10.1038/ncomms2996.
65. Subramoni S, Suaréz-Moreno ZR, Venturi V. Lipases as pathogenicity factors of plant pathogens. 2010. https://doi.org/10.1007/978-3-540-77587-4_248.
66. Lanver D, Berndt P, Tollot M, Naik V, Vranes M, Warmann T, et al. Plant surface cues prime *Ustilago Maydis* for biotrophic development. PLoS Pathog. 2014;10(7). https://doi.org/10.1371/journal.ppat.1004272.
67. Moreno-Sánchez I, Pejenaute-Ochoa MD, Navarrete B, Barrales RR, Ibeas JI. *Ustilago Maydis* secreted Endo-Xylanases are involved in fungal filamentation and proliferation on and inside plants. J Fungi. 2021;7(12):1081. https://doi.org/10.3390/jof7121081.
68. Doehlemann G, Reissmann S, Aßmann D, Fleckenstein M, Kahmann R. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. Mol Microbiol. 2011;81(3). https://doi.org/10.1111/j.1365-2958.2011.07728.x.
69. Narváez-Barragán DA, Tovar-Herrera OE, Segovia L, Serrano M, Martinez-Anaya C. Expansin-related proteins: biology, microbe-plant interactions and associated plant-defense responses. Microbiology. 2020;166(11). https://doi.org/10.1099/mic.0.000984.
70. Weiland P, Dempwolff F, Wieland Steinchen, Freibert S, Tian H, Glatter T, et al. Structural and functional analysis of the cerato-platanin-like protein Cpl1 suggests diverging functions in Smut fungi. Mol Plant Pathol. 2023;24(7):768–87. https://doi.org/10.1111/mpp.13349.
71. Marques JPR, Hoy JW, Appezzato-da-Glória B, Viveros AFG, Vieira MLC, Baisakh N. Sugarcane cell Wall-Associated defense responses to infection by *Sporisorium Scitamineum*. Front Plant Sci. 2018;9. https://doi.org/10.3389/fpls.2018.00698.
72. Ling H, Fu X, Huang N, Zhong Z, Liu T, Cui H, et al. A sugarcane Smut fungus effector hijacks plant vacuolar sorting Receptor-Mediated trafficking to evade host immune detection. Plant Cell Environ. 2025. https://doi.org/10.1111/pce.15500.
73. Horst RJ, Zeh C, Saur A, Sonnewald S, Uwe Sonnewald, Voll LM. The *Ustilago Maydis* Nit2 homolog regulates nitrogen utilization and is required for efficient induction of filamentous growth. Eukaryot Cell. 2012;11(3). https://doi.org/10.1128/EC.05191-11.
74. Lanver D, Müller AC, Happel P, Schweizer G, Haas F, Franitza M, et al. The biotrophic development of *Ustilago Maydis* studied by RNA-Seq analysis. Plant Cell. 2018;30(2):300–23. https://doi.org/10.1105/tpc.17.00764.
75. Wallen RM, Richardson K, Furnish M, Mendoza H, Dentinger A, Khanal S, et al. Hungry for sex: differential roles for *Ustilago Maydis* b locus components in haploid cells vis à vis nutritional availability. J Fungi. 2021;7(2). https://doi.org/10.3390/jof7020135.
76. Hewald S, Linne U, Scherer M, Marahiel MA, Kämper J, Bölker M. Identification of a gene cluster for biosynthesis of mannosylerythritol lipids in the basidiomycetous fungus *Ustilago Maydis*. Appl Environ Microbiol. 2006;72(8). https://doi.org/10.1128/AEM.00506-06.
77. Khanal S, Schroeder L, Nava-Mercado OA, Mendoza H, Perlin MH. Role for nitrate assimilatory genes in virulence of *Ustilago Maydis*. Fungal Biology. 2021;125(10). https://doi.org/10.1016/j.funbio.2021.04.010.
78. Teixeira PJPL, et al. High-Resolution transcript profiling of the atypical biotrophic interaction between theobroma Cacao and the fungal pathogen *Moniliophthora perniciosa*. Plant Cell. 2014;26(11):4245–69. https://doi.org/10.1105/tpc.114.130807.
79. Riquelme M, Aguirre J, Bartnicki-García S, Braus GH, Feldbrügge M, Fleig U, et al. Fungal Morphogenesis, from the polarized growth of hyphae to complex reproduction and infection structures. Microbiol Mol Biol Rev. 2018;82(2). https://doi.org/10.1128/MMBR.00068-17.
80. Gullner G, Komives T, Király L, Schröder P. Glutathione S-Transferase enzymes in Plant-Pathogen interactions. Front Plant Sci. 2018;9. https://doi.org/10.3389/fpls.2018.01836.
81. Falter C, Reumann S. The essential role of fungal peroxisomes in plant infection. Mol Plant Pathol. 2022. https://doi.org/10.1111/mpp.13180.
82. Richie DL, Hartl L, Aimanianda V, Winters MS, Fuller KK, Miley MD, et al. A role for the unfolded protein response (UPR) in virulence and antifungal susceptibility in *Aspergillus fumigatus*. PLoS Pathog. 2009;5(1):e1000258–8. https://doi.org/10.1371/journal.ppat.1000258.
83. Marín-Menguiano M, Moreno-Sánchez I, Barrales RR, Fernández-Álvarez A, Ibeas JI. N-glycosylation of the protein disulfide isomerase Pdi1 ensures full *Ustilago maydis* virulence. Xu JR, editor. PLoS Pathogens. 2019;15(11). https://doi.org/10.1371/journal.ppat.1007687.
84. Pinter N, Hach CA, Hampel M, Rekhter D, Zienkiewicz K, Feussner I, et al. Signal peptide peptidase activity connects the unfolded protein response to plant defense suppression by *Ustilago Maydis*. PLoS Pathog. 2019;15(4). https://doi.org/10.1371/journal.ppat.1007734.
85. Heimel K, Freitag J, Hampel M, Ast J, Bölker M, Kämper J. Crosstalk between the unfolded protein response and pathways that regulate pathogenic development in *Ustilago Maydis*. Plant Cell. 2013;25(10). https://doi.org/10.1105/tpc.113.115899.
86. Lanver D, Tollot M, Schweizer G, Presti LL, Reissmann S, Ma LS, et al. *Ustilago Maydis* effectors and their impact on virulence. Nat Rev Microbiol. 2017;15(7):409–21. https://doi.org/10.1038/nrmicro.2017.33.
87. Luo G, Wang T, Zhang J, Zhang P, Lu Y. *Candida albicans* requires iron to sustain hyphal growth. Biochem Biophys Res Commun. 2021;561:106–12. https://doi.org/10.1016/j.bbrc.2021.05.039.
88. Elías-Villalobos A, Barrales RR, Ibeas JI. Chromatin modification factors in plant pathogenic fungi: insights from *Ustilago Maydis*. Fungal Genet Biol. 2019;129:52–64. https://doi.org/10.1016/j.fgb.2019.04.006.
89. Ma LS, Wang L, Trippel C, Mendoza-Mendoza A, Ullmann S, Moretti M, et al. The *Ustilago Maydis* repetitive effector Rsp3 blocks the antifungal activity of mannose-binding maize proteins. Nat Commun. 2018;9(1). https://doi.org/10.1038/s41467-018-04149-0.
90. Djamei A, Kahmann R. *Ustilago maydis*: dissecting the molecular interface between pathogen and plant. PLoS Pathog. 2011;8(11). https://doi.org/10.1371/journal.ppat.1002955.
91. Baccelli I. Cerato-platanin family proteins: one function for multiple biological roles? Front Plant Sci. 2015;5. https://doi.org/10.3389/fpls.2014.00769.
92. Madden LV, Hughes G. Sampling for plant disease incidence. Phytopathology®. 1999;89(11):1088–103. https://doi.org/10.1094/PHYTO.1999.89.11.1088.
93. Shapiro SS, Wilk MB. An analysis of variance test for normality (Complete Samples). Biometrika. 1965;52:591–611. https://doi.org/10.2307/2333709.
94. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin. 1987;19:11–15.
95. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat Methods. 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2.
96. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36. https://doi.org/10.1101/gr.215087.116.
97. Taniguti LM, Vilanova PF, Kitajima JP, Monteiro-Vitorello CB. Updated chromosome-level genome assembly of *Sporisorium Scitamineum* with improved accuracy and completeness. bioRxiv. 2025. https://doi.org/10.1101/2025.05.23.649816.
98. Min B, Grigoriev IV, Choi IG. FunGAP: fungal genome annotation pipeline using evidence-based gene model evaluation. Bioinformatics. 2017;33(18). https://doi.org/10.1093/bioinformatics/btx353.
99. Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago Maydis*. Nature. 2006;444(7115):97–101. https://doi.org/10.1038/nature05248.
100. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19). https://doi.org/10.1093/bioinformatics/btv351.
101. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: Ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34. https://doi.org/10.1093/nar/gkl200.

102. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. 2017;110(10). https://doi.org/10.1007/s10482-017-0844-4.
103. The Galaxy Community. The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic Acids Res. 2024. https://doi.org/10.1093/nar/gkae410.
104. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14). https://doi.org/10.1093/bioinformatics/btp324.
105. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and samtools. Bioinformatics. 2009;25(16). https://doi.org/10.1093/bioinformatics/btp352.
106. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9). https://doi.org/10.1101/gr.107524.110.
107. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2012;6(2):80–92. https://doi.org/10.4161/fly.19695.
108. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar G, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2020;49. https://doi.org/10.1093/nar/gkaa913.
109. Jones DA, Bertazzoni S, Turo CJ, Syme RA, Hane JK. Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. Curr Opin Microbiol. 2018;46. https://doi.org/10.1016/j.mib.2018.01.017.
110. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast Genome-Wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8). https://doi.org/10.1093/molbev/msx148.
111. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2018;47. https://doi.org/10.1093/nar/gky1085.
112. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1). https://doi.org/10.1093/nar/gky418.
113. Fischer M. The lipase engineering database: a navigation and analysis tool for protein families. Nucleic Acids Res. 2003;31(1):319–21. https://doi.org/10.1093/nar/gkg015.
114. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2013;42(D1):D503–9. https://doi.org/10.1093/nar/gkt953.
115. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1). https://doi.org/10.1186/s13059-015-0721-2.
116. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92. https://doi.org/10.1093/bib/bbs017.
117. Gasteiger E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784–8. https://doi.org/10.1093/nar/gkg563.
118. Yoshida S, Ishida JK, Kamal NM, Ali AM, Namba S, Shirasu K. A full-length enriched cDNA library and expressed sequence Tag analysis of the parasitic weed, *Striga Hermonthica*. BMC Plant Biol. 2010;10(1). https://doi.org/10.1186/1471-2229-10-55.
119. Morlan JD, Qu K, Sinicropi DV. Selective Depletion of rRNA Enables Whole Transcriptome Profiling of Archival Fixed Tissue. Dadras SS, editor. PLoS One. 2012;7(8). https://doi.org/10.1371/journal.pone.0042882.
120. Miranda RP, Turrini PCG, Bonadio DT, Zerillo MM, Berselli AP, Creste S, et al. Genome organization of four Brazilian *Xanthomonas albilineans* strains does not correlate with aggressiveness. Microbiol Spectr. 2023;11(3). https://doi.org/10.1128/spectrum.02802-22.
121. FastQC:A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 11 Oct 2025.
122. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):10. https://doi.org/10.14806/ej.17.1.200.
123. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8). https://doi.org/10.1038/s41587-019-0201-4.
124. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013;30(7). https://doi.org/10.1093/bioinformatics/btt656.
125. Chen Y, Chen L, Lun AL, Baldoni P, Smyth G. EdgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Res. 2025;53(2). https://doi.org/10.1093/nar/gkaf018.
126. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1). https://doi.org/10.1093/bioinformatics/btp616.

Publisher's Note