scientific reports

Check for updates

OPEN

Isolation, genome analysis and tissue localization of *Ceratobasidium theobromae*, a new encounter pathogen of cassava in Southeast Asia

Alejandra Gil-Ordóñez[1,2], Juan M. Pardo[1], Samar Sheat[3], Khamla Xaiyavong[4], Ana M. Leiva[1], Warren Arinaitwe[4], Stephan Winter[3], Jonathan Newby[4] & Wilmer J. Cuellar[1]✉

In Southeast Asia (SEA) fastidious fungi of the *Ceratobasidium* genus are associated with proliferation of sprouts and vascular necrosis in cacao and cassava, crops that were introduced from the tropical Americas to this region. Here, we report the isolation and in vitro culture of a *Ceratobasidium* sp. isolated from cassava with symptoms of witches' broom disease (CWBD), a devastating disease of this crop in SEA. The genome characterization using a hybrid assembly strategy identifies the fungus as an isolate of the species *C. theobromae*, the causal agent of vascular streak dieback of cacao in SEA. Both fungi have a genome size > 31 Mb (G+C content 49%), share > 98% nucleotide identity of the Internal Transcribed Spacer (ITS) and > 94% in genes used for species-level identification. Using RNAscope® we traced the pathogen and confirmed its irregular distribution in the xylem and epidermis along the cassava stem, which explains the obtention of healthy planting material from symptom-free parts of a diseased plant. These results are essential for understanding the epidemiology of CWBD, as a basis for disease management including measures to prevent further spread and minimize the risk of introducing *C. theobromae* via long-distance movement of cassava materials to Africa and the Americas.

Keywords Cassava witches' broom disease, *Ceratobasidium*, Fastidious fungus, *Manihot esculenta*, Re-emerging disease, *Theobroma cacao*, Vascular streak dieback, Xylem

Cassava (*Manihot esculenta* Crantz) has become an economically important crop for mainland Southeast Asia (SEA) which has led to a rapid expansion of its cropland as it provides a source of livelihood and attractive economic returns for millions of smallholder farmers in this region[1]. Being native to South America cassava was introduced to Asia in the 1800's and extensively cultivated in the twentieth century in the Philippines, India, and Indonesia from where it extended later to the rest of Asia. Today, it is estimated that more than 8 million farmers grow cassava in Asia on approximately 4.2 million ha[2,3] Along with intensification and range expansion, reports of pests and diseases affecting cassava in SEA have also increased[4–6]. The recent outbreak of cassava witches' broom disease (CWBD)[6,7] and the unexpected finding of *Ceratobasidium theobromae*, the causal agent of cacao vascular streak dieback (VSD) as the closest relative to the fungus found in cassava[8,9], prompted us to carry out a detailed characterization of this cassava disease and its associated pathogen.

Effective disease management requires an accurate diagnosis of the pathogen: from (1) its isolation, (2) accurate identification and eventually (3) confirmation of infectivity by transmitting the isolated pathogen back to its original host to fulfil Koch's postulates. In vegetative propagated crops such as cassava multiple pathogens can occur, and these mixed infections pose an extra challenge to assess the contribution of a singular organism to the

[1]Cassava Program, Crops for Nutrition and Health, International Center for Tropical Agriculture (CIAT), The Americas Hub, Km 17, Recta Cali-Palmira, 763537 Palmira, Colombia. [2]Departamento de Biología, Facultad de Ciencias Naturales y Exactas, Universidad del Valle, Calle 13 # 100-00, 760032 Cali, Colombia. [3]Plant Virus Department, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, 38124 Braunschweig, Germany. [4]Cassava Program Asia Office, Crops for Nutrition and Health, International Center for Tropical Agriculture (CIAT), P.O. Box 783, Vientiane, Lao PDR. ✉email: w.cuellar@cgiar.org

disease and particular symptoms. An incorrect identification of the pathogen can misdirect the whole disease management process[10]. Until recently, phytoplasmas were considered as the causing agent of CWBD based on early reports on detection and identification of 16S phytoplasma sequences in diseased plants however pending further confirmation with complementary analysis or epidemiological studies[6]. Indeed, for enigmatic pathogens Koch's postulates are very difficult to fulfil and it took very long to confirm *C. theobromae* as the causal agent of VSD of cacao[11] although, greenhouse experiments and epidemiological studies[8,12,13] were conducted to successfully guide the management of a fungal VSD and to effectively protect cacao against the disease[14].

Based on the morphology of basidiospores, the VSD fungus was initially classified as *Oncobasidium theobromae*[15]. Later, based on its binuclear hyphae, the formation of fruiting structures emerging from leaf scars of infected leaves and, the colonization of xylem tissue, the pathogen was reclassified as *Ceratobasidium theobromae.* It is a nearly obligate parasite (biotroph) with only minimal growth on standard nutrient media. Thus, *Ceratobasidium* fungi are fastidious basidiomycetes of the binucleate Rhizoctonia-like fungi that require specific conditions to enter sexual stages[8,12,13]. These unique biological characteristics of *C. theobromae* are also reflected in the conserved sequences of its Internal Transcribed Spacer region (ITS), the Large Subunit (LSU)[16] and other taxonomically informative genes, RNA polymerase II (*rpb2*), translation elongation factor 1 alpha (*tef1*), and ATP synthase subunit 6 (*atp6*)[17] classifying *C. theobromae* as a distinct fungal species.

Ceratobasidium sp. has a slow growth in vitro[8,12,16,18,19]. As a pure culture, the fungus soon dies out after very few transfers to new media which is a major constraint to obtaining good quality DNA for genomic and phylogenetic studies[9,18]. We isolated C*eratobasidium* sp. from petioles of CWBD-affected cassava plants and used a modified protocol[19,20] to obtain pure mycelia cultures from which we isolated genomic DNA for whole genome assembly and characterization using Illumina and Oxford nanopore Technologies. In situ hybridization localized the fungus to the xylem tissues and screenhouse tests confirmed fungal growth along the cassava stem. Our results confirm the identity of the fungus isolated from cassava with CWBD symptoms as *C. theobromae*, the causal agent of VSD and therefore a potential significant threat to cacao production in Africa and the Americas. This work also provides key information for early diagnostics of CWBD that can be used in screening for resistance and pre-emptive management of this disease in SEA and further cassava regions at risk.

Results

Fungus isolation and in vitro culture

Isolation of *Ceratobasidium* sp. (isolate LAO1) from diseased cassava plants using the protocol described here took an average of 23 days. Contamination with *Colletotrichum* sp. and bacteria was minimized by collecting petioles from the upper parts of mature cassava plants and by using benomyl and lactic acid. The use of PDA + CCM to stimulate growth of uniform mycelia allowed to collect sufficient fungal material for DNA extraction. The *Ceratobasidium* sp. LAO1 developed aerial hyphae with a yellowish coloration (Fig. 1). At × 40 magnification the right-angle septation pattern at branching points and binucleated hyphae[9,15] (Fig. 1) confirmed its morphological characteristics. We obtained on average 80 mg of *Ceratobasidium* mycelium per Petri dish and DNA extraction from these cultures produced ~ 300 ng of DNA per mg of fresh mycelium, or > 24 µg of total fungal DNA per Petri dish.

Genome characterization

The total length of *Ceratobasidium* sp. LAO1 was 33,386,600 bp (average coverage ~ 1160×), consisting of 547 contigs (N50 = 435,246 bp, L50 = 18) and 48.93% G+C content (Fig. 2A). The G+C content of all contigs ranged between 22.5 and 59.9% (Fig. 2B) and none were assigned by BLASTx search to bacteria (Supplementary Table 1, Fig. 2C). On the contrary, most of the classified sequences corresponded to Basidiomycota (434 contigs; 33 Mbp), mainly to *C. theobromae* (320 contigs; 31.4 Mbp); the remaining contigs were attributed to *Rhizoctonia* (80 contigs; 1.3 Mbp), *Ceratobasidium* (71 contigs; 1.5 Mbp) and other Basidiomycota genera (6 contigs; 172 kb) (Supplementary Table 1). Genome quality assessment performed with BUSCO indicated that *Ceratobasidium* sp. isolated from cassava contains 98% (94.5% complete, 3.5% fragmented) and 87% (83.6% complete, 3.6% fragmented) of the loci examined against the Eukaryote and Agaricomycetes lineages, respectively (Fig. 2D).

Ab initio analysis predicted 10,171 genes in the *Ceratobasidium* sp. LAO1 genome, with an average sequence length of 1714.26 (± 1349.47) bp. DIAMOND-BLASTp affiliated 6794 (66.8%; identity ≥ 90%) of predicted genes to the nr protein database, and functional characterization annotated 2302 (22.63%) KO identifiers with Blast-KOALA, and 9341 (91.84%) genes against at least one of the 17 databases inspected with InterProScan (Supplementary Table 2). Contig 60 was identified as mtDNA (130,767 bp; G+C content 36.2%) after evaluating the presence of the genes *nad1*, *cox2*, *rns*, *rnl*, *cox1*, *nad5*, *nad6*, *cob*, *cox3*, *nad2*, *rps3*, *atp6*, *nad4* and *atp9* by BLASTn. RNA-seq data from diseased cassava produced an average of 102.85 (± 13.16) M reads, of which 61.16 (± 11.83) M reads mapped against *Ceratobasidium* sp. LAO1 and further supported 81.08% of ab initio predicted genes by identifying 8,247 transcriptionally active genes (≥ 10 raw reads) (Supplementary Table 3).

According to gene cluster analysis based on predicted proteins of *Ceratobasidium* sp. LAO1, *C. theobromae* CT2 (GenBank: GCA_009078325.1) and *R. solani* AG-1 IA (GenBank: GCA_016906535.1) the fungal genomes share 5702 orthologous gene clusters with highest association among *Ceratobasidium* isolates (1646) and 86 singletons detected in isolate LAO1 (Fig. 3A). Phylogenetic analysis based on the substitution model revealed higher divergence between *R. solani* AG-1 IA and the *Ceratobasidium* isolate cluster (Fig. 3B). The nucleotide distance calculated by Mash between the genome assembly of *Ceratobasidium* sp. LAO1 and *C. theobromae* CT2 linked the genomes at the species level (0.0388, p-value < 0.01), while the distance of *R. solani* AG-1 IA from *Ceratobasidium* sp. LAO1 (0.2154, p value < 0.01) and *C. theobromae* CT2 (0.2156, p value < 0.01) assigned the genomes to different species. Moreover, the identity of the molecular markers evaluated was 94.89% for *atp6* of *C. theobromae* Gudang 4, Tawau Sabah (GenBank: GCA_012932095.1), and 97.53% for *tef1*, 98.18% for *rpb2* and

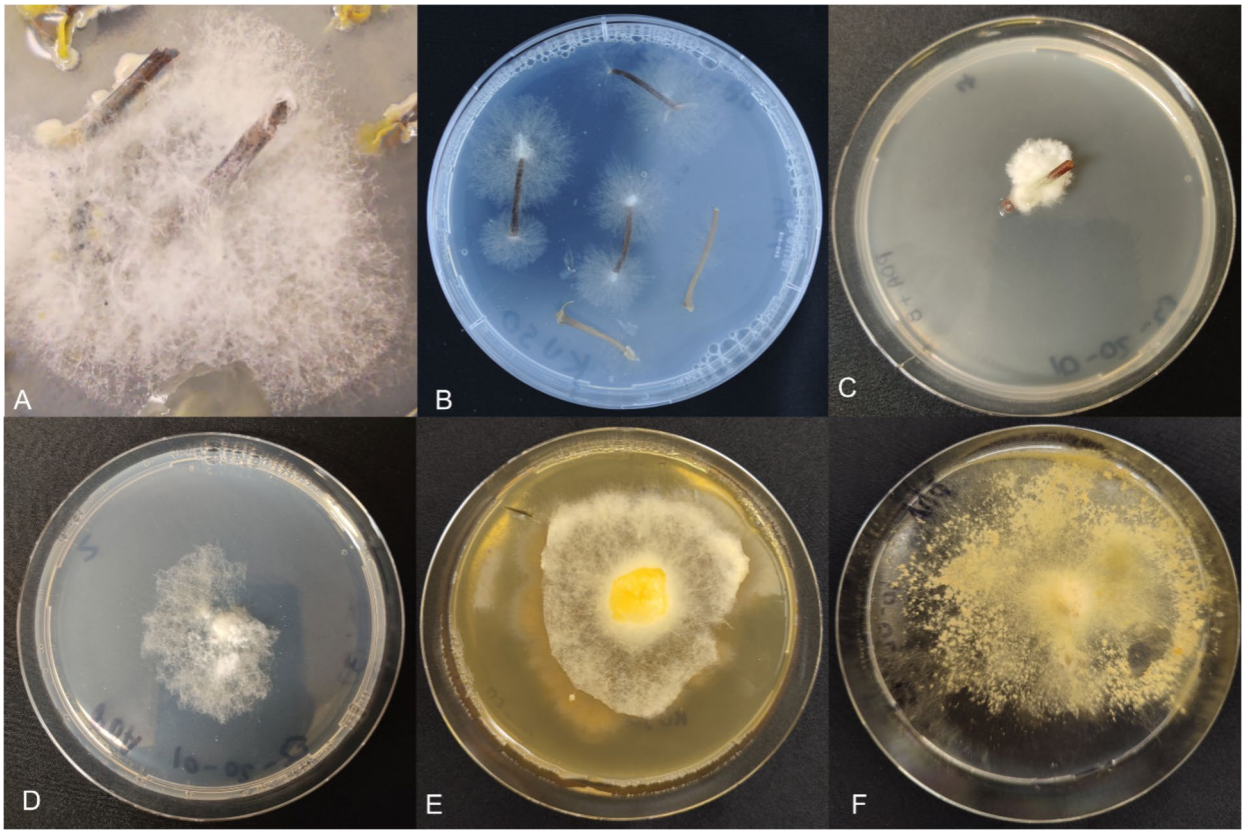

Figure 1. Isolation of *Ceratobasidium* sp. Isolate LAO1 from diseased petioles of cassava. (**A**) Bacterial and fungal contamination in PDA media. (**B**) 48 h mycelial growing on the tips of the petioles on water agar. (**C**) Cassava petiole with *Ceratobasidium* sp. LAO1 mycelial in a new petri dish with PDA + benomyl + lactic acid. (**D**) *Ceratobasidium* sp. LAO1 mycelial growing pure in a PDA + benomyl + lactic media. (**E**) Pure *Ceratobasidium* sp. LAO1 strain transferred from PDA + benomyl + lactic acid to PDA + CCM media to increase inoculum. The mycelia turned yellow after a few days. (**F**) *Ceratobasidium* sp. LAO1, a two-month-old colony growing on PDA + CCM media.

98.78% for ITS of *C. theobromae* CT2. In summary, our study provides molecular evidence that the fungus associated with CWBD of cassava is an isolate of *C. theobromae* and henceforth referred to as *C. theobromae* LAO1).

Secretome analysis predicted 244 putative effector proteins (PEP), of which 169 and 25 were predicted as apoplastic and cytoplasmic effectors, respectively. Comparison of PEPs validated by RNA-seq, of *C. theobromae* LAO1 (EffectorP 3.0: 112) and *C. theobromae* CT2 (EffectorP 2.0: 19), revealed that 101 were exclusive to LAO1, 8 to CT2 and 11 were shared between both isolates. According to DIAMOND-BLASTp, most of these PEPs showed homology with sequences from *C. theobromae* CT2 (not reported as PEPs) and *R. solani* however, the analysis identified 5 PEPs unique to *C. theobromae* LAO1 (Gene ID 1.g11, 22.g5906, 4.g2160, 5.g2603 and 59.g8215). Some of them were predicted as apoplastic (1.g11), cytoplasmic (4.g2160 and 5.g2603), or both apoplastic and cytoplasmic (22.g5906 and 59.g8215), most likely in the apoplast. Most PEPs were annotated as virulence factors such as pectate lyase (IPR002022) and catalytic enzymes such as glycoside hydrolases (PF00840), polysaccharide hydrolysis (SM00710) and pectin lyase (IPR012334) (Supplementary Tables 3, 4). A list of genome features of *C. theobromae* LAO1 identified in this work is presented in Table 1.

C. theobromae localization in cassava

Semi-thin cross-sections, between 10 and 15 µm thick, obtained from leaves, petioles, and stems, from infected cassava plant were subjected to the RNAscope® in situ localization method. Analyzing multiple sections from each tissue of infected cassava, we uncovered a localized distribution of *C. theobromae* within the infected tissues. Red dots or red clusters, predominantly occupying xylem cells were found in the vascular tissue of the leaf midrib, petioles, and stem tissues using the CAMK/CAMKL probe (Fig. 4A–C, red arrows). Notably, there was no background signal detected in leaf sections from healthy plants, highlighting the high specificity of the detection method (Fig. 4D). CAMK/CAMKL was used in a previous work as an efficient diagnostic marker for CWBD[9].

PEP4 probe signal was predominant in the xylem of the leaf midrib (Supplementary Fig. 1A, red arrow) and stems (Supplementary Fig. 1B,C, red arrows). There was no signal in tissue sections from healthy plant (Supplementary Fig. 1D). When a GADPH gene probe was used, highly intense signals in xylem tissue from the vascular tissue of the midrib (Supplementary Fig. 2A, red arrow) and infected stems (Supplementary Fig. 2B,C, red arrows) extending to the external phloem were found (Supplementary Fig. 2C, white arrow), while healthy

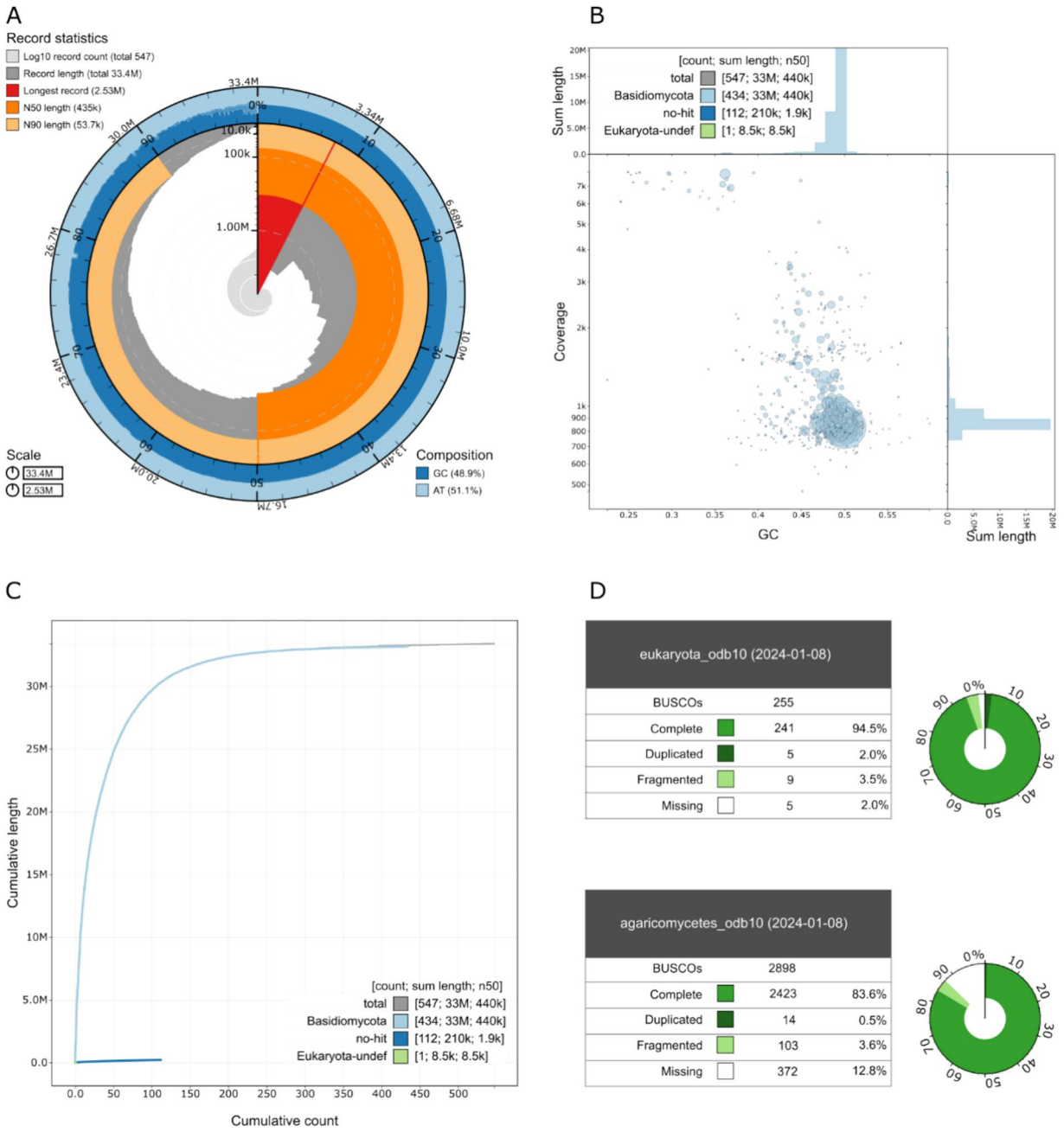

Figure 2. Genome statistics of *Ceratobasidium* sp. Isolate LAO1. (**A**) Snail plot describing the assembled genome of *Ceratobasidium* sp. LAO1. The main plot shows the contigs sorted by size around the circumference, with the length distribution of the contigs in dark gray and the radius of the plot scaled to the longest contig present in the assembly (2.53 Mbp) in red. A pale-gray spiral shows a logarithmic scale of the cumulative contig count, and scaled white lines indicate successive orders of magnitude. The lengths of the assembly metrics N50 and N90 are shown in orange and pale orange arcs, respectively, and the bands outside the main circumference indicate the G+C (blue) and A+T (pale blue) distribution. (**B**) Three-part plot showing histograms with G+C content (top) and coverage (right) weighted by the cumulative length of each contig. In the main panel, contigs are shown in circles colored by taxonomic assignment and distributed according to G+C content (x-axis) and coverage (y-axis), calculated by mapping the short reads against the assembled genome. (**C**) Rarefaction curve indicating taxonomic affiliation of *Ceratobasidium* sp. LAO1 contigs using DIAMOND-BLASTx against nr protein database (2023-07-28). (**D**) Summary of complete, fragmented, duplicated and missing BUSCO metrics based on orthologous genes reported in eukaryota_odb10 (2024-01-08) (top) and agaricomycetes_odb10 (2024-01-08) (bottom) dataset. Genome statistics were diagrammed using Blobtools v4.2.0.

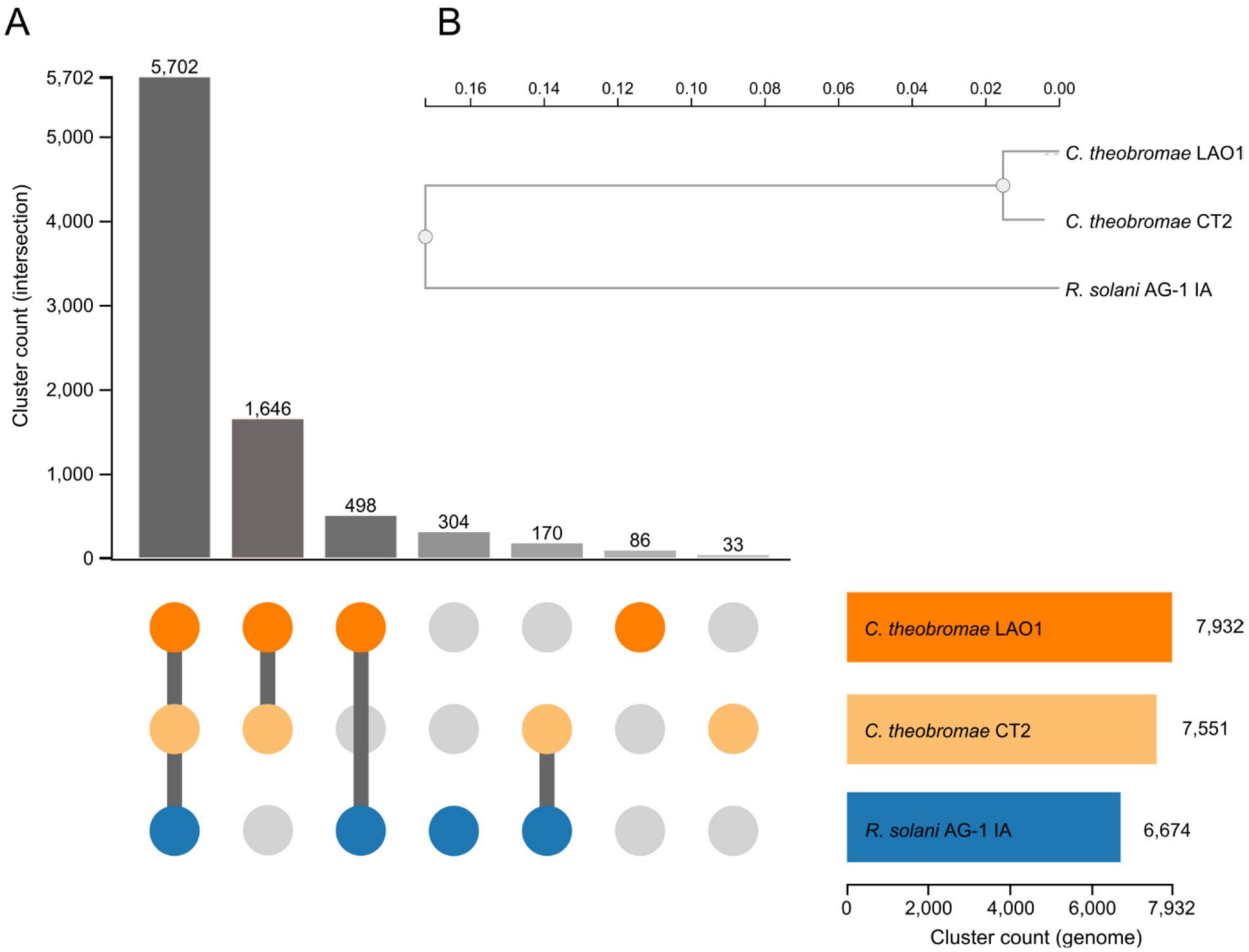

Figure 3. Orthologous cluster analysis of *C. theobromae* LAO1, *C. theobromae* CT2 and *R. solani* AG-1 IA. The UpSet table (**A**) illustrates both distinct and overlapping orthologous clusters among *C. theobromae* LAO1 (orange dots), *C. theobromae* CT2 (pale orange dots) and *R. solani* AG-1 IA (blue dots). The horizontal bar chart on the left illustrates the count of orthologous clusters of *C. theobromae* LAO1 (orange bar), *C. theobromae* CT2 (pale orange bar) and *R. solani* AG-1 IA (blue bar), while the vertical bar chart on the right (gray gradient) displays the number of orthologous clusters shared among species. Intersecting sets are represented by dark grey lines connecting the bars. (**B**) Evolutionary relationships and distances among *C. theobromae* LAO1, *C. theobromae* CT2 and *R. solani* AG-1 IA based on conserved single-copy genes. The phylogenetic tree was constructed using Maximum Likelihood method based on JTT + CAT evolutionary model with protein sequences. SH test was used to determine the reliability of each node. The horizontal bar chart on the right illustrates the count of orthologous clusters of *C. theobromae* LAO1 (orange bar), *C. theobromae* CT2 (pale orange bar) and *R. solani* AG-1 IA (blue bar).

sections were signal free (Supplementary Fig. 2D). All three probes used clearly demonstrate the xylem localization of *C. theobromae* in cassava.

To further elucidate the spatial distribution of the fungus within infected cassava plants, we carried on targeted analyses on various sections of infected plants. Our investigation unveiled intriguing patterns across different tissue samples within a single plant. Analyzing multiple sections from each tissue we observed an elusive signal, characterized by only one dot, in the upper part of the stem. Notably, this signal was distinctly localized within the xylem region (Fig. 5D, red arrow). Upon checking the middle part of the stem, the signal intensity increased, particularly within the xylem of the vascular tissue (Figs. 5A,E,F, red arrows). Additionally, within the symptomatic parts, we observed localized signals of heightened intensity in the epidermal cells. Intriguingly, our analysis showed no detectable signal in either the petiole (Fig. 5B) or the midrib of a leaf that did not display symptoms of the disease (Fig. 5C). In summary, *C. theobromae* present a distinctive distribution pattern within specific regions of the plant with higher signals in xylem and in the epidermal cells of the plant parts that show clear CWBD symptoms (sprout proliferation, short petioles, yellowing small leaves[6]).

Discussion

Precise identification of a pathogen causing a disease or of associated pathogens contributing to a syndrome is vital for disease control including containment and prevention of transboundary movements[10]. Ample evidence shows that introduction of a pathogen into a new region can cause new epidemics often with a dramatic impact

	C. theobromae LAO1
Total length (bp)	33,386,600
Contig number	547
GC content (%)	48.93
Contig or scaffold N50	435,246
Contig or scaffold L50	18
Min contig length (bp)	1000
Max contig length (bp)	2,528,383
Mean contig length (bp)	61,035.83
BUSCO Completeness (%)	94.50
Coverage (x) short reads	1150.31
Coverage (x) ONT reads	12.04
Predicted genes	10,171
Gene density[a]	0.52
Total gene length (bp)	17,435,730
Average gene length (bp)	1714.26
Refined putative effector proteins	244
Number of expressed genes[b]	8247

Table 1. Genome assembly and annotation metrics of *C. theobromae* LAO1. [a]CDS bases/total genome bases. [b]Only genes with sequence counts ≥ 10 detected in any of diseased cassava samples.

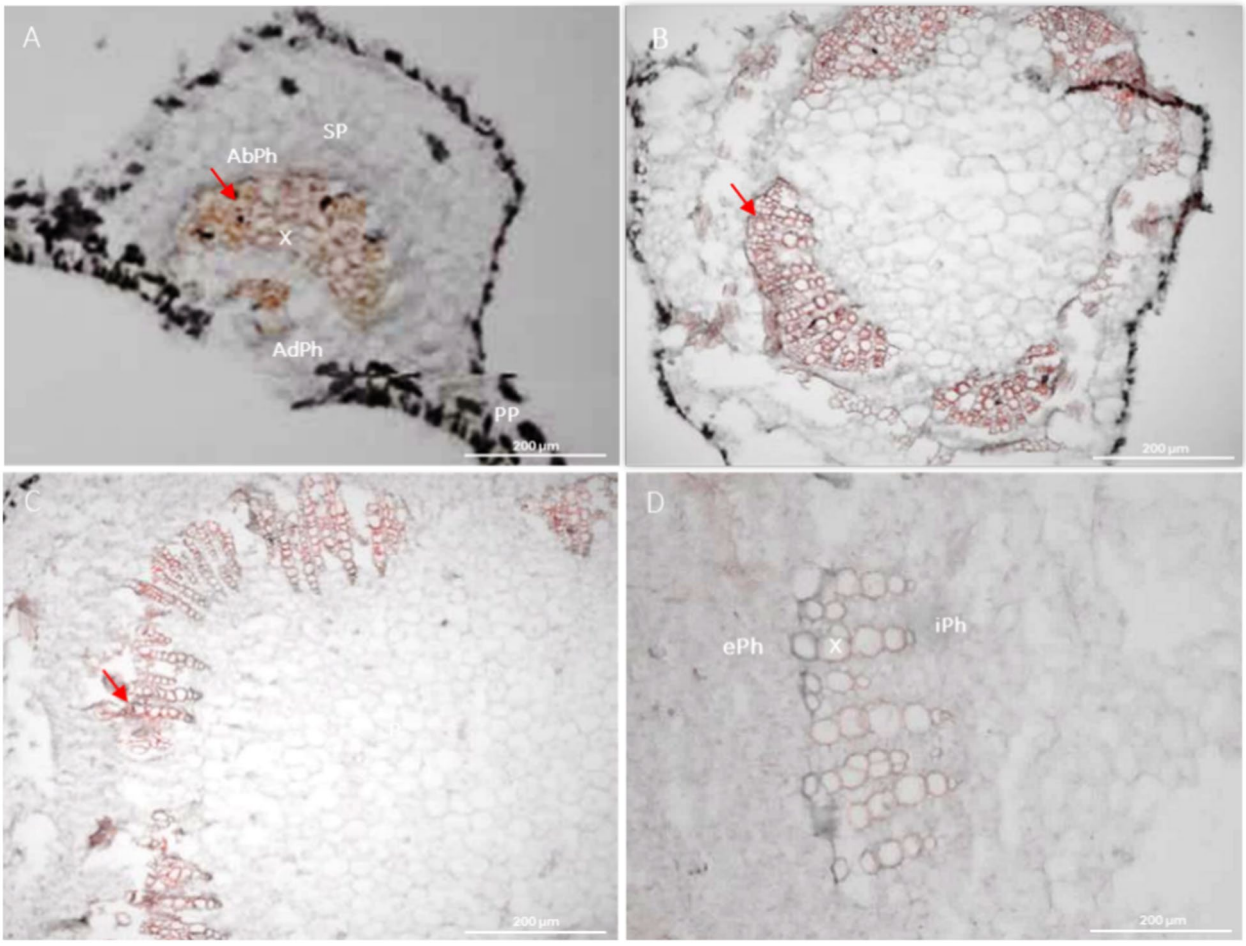

Figure 4. Distribution of *C. theobromae* in infected cassava tissues. Cassava cross sections from leaf mid rib (**A**), petiole (**B**) and stem (**C**) and healthy cassava stem (**D**) subjected to RNAscope® ISH using a CAMK/CAMKL probe. The red signals represent a specific hybridization to structures localized in xylem tissue (red arrows) in the vascular tissues in the infected materials. *X* xylem, *ePh* external phloem, *iPh* inner phloem, *PP* palisade parenchyma, *SP* spongy parenchyma, *AbPh* abaxial phloem, *AdPh* adaxial phloem.

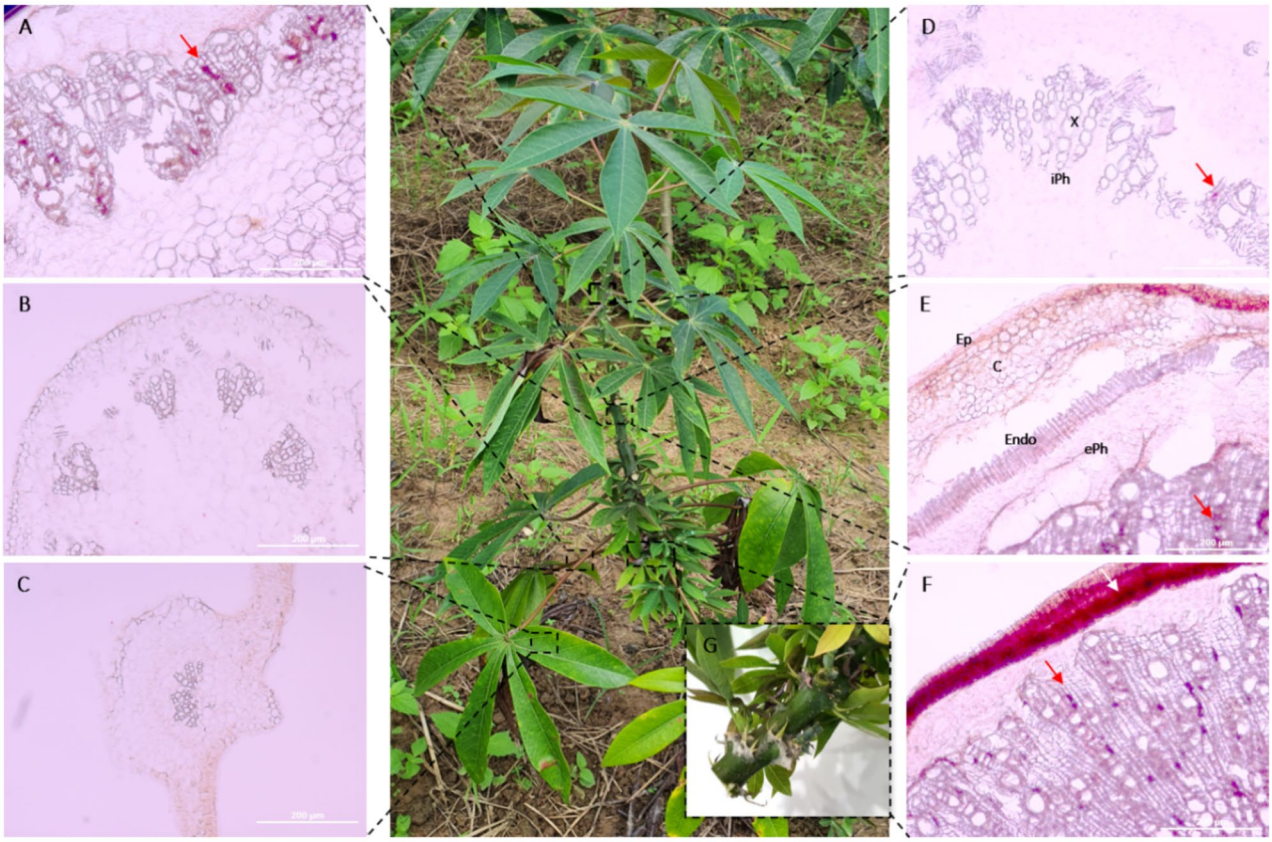

Figure 5. Distribution of *C. theobromae* in infected cassava plant at 4 months post planting. Cross sections of infected tissues from witches' broom diseased cassava from middle stems (**A**,**E**), petiole (**B**), midrib of leaf (**C**), upper stem (**D**) and lower stem (**F**) subjected to RNAscope® ISH using CAMK/CAMKL probe. The red signals are a hybridization to structures localized in xylem tissue (**A**,**D**,**E** and red arrows) in the vascular tissues in the infected stems and to the Epidermis tissues in the lower stem section (**F**, white arrow). No signal was found in the healthy petiole (**B**) and leaf (**C**) from the lower part of the plant. (**G**) Hyphae growth on the branching axillary buds of affected petiole. *X* xylem, *C* cortex, *ePh* external phloem, *iPh* inner phloem, *Ep* epidermis, *Endo* endodermis.

as shown with the European outbreak of *Xylella fastidiosa* causing a severe decline of olive trees in Southern Italy[21]. Vice versa, the introduction of new plant species to a new location where they are exposed to novel pathogens are the leading causes of major epidemics in animals and plants[22,23]. These alien plants may be favored by endemic pathogens expanding their host range causing a new disease with unpreceded impact. The disease caused by *C. theobromae* in cassava and cacao is a new encounter because these crops have been introduced to the SEA region to become exposed to a native fungus[24]. Interestingly, the majority of *Ceratobasidium* species have saprophytic lifestyles, or endophytes and mycorrhizal fungi with only a few members reported as facultative xylem pathogens[25].

Fungi belonging to the *Ceratobasidium* genus are classified as fastidious (i.e. requiring specific nutrients for culture in the laboratory)[8,12,13] which makes their maintenance in vitro a challenge and limits its genetic analysis[9,11] In general, the fastidious nature of *C. theobromae* favors the overgrowth of contaminants during the isolation and subculturing process. In our experience, one of the most frequent contaminants was *Colletotrichum* sp., which grows faster masking the typical structures of *Ceratobasidium* colonies. Furthermore, *Colletotrichum* sp. produces abundant conidia within a week, with the potential to contaminate further isolation steps. As shown this problem can be controlled with the use of benomyl in the culture media[26,27]. *Colletotrichum* sp. contamination is not unexpected; infection by this fungus in cassava is linked to leaf senescence which is another symptom associated with CWBD[6]. It is also known that the pathogen can remain in a quiescent stage within cassava petioles for extended periods and occasionally, small acervuli may be observed on the petiole surface[28]. Thus, the isolation protocol described here (Fig. 1) allowed us to control contamination and produce enough biomass and genetic material (> 25 µg of total fungal DNA per Petri dish) to proceed with the genome characterization of the fungus (Fig. 2). Furthermore, this protocol also served as a base for scaling up the production of inoculum, a necessary step for screening of germplasm to identify resistant cassava genotypes. Nevertheless, up to this moment we have failed to produce spores from in vitro cultures of *C. theobromae*.

The genome size of *C. theobromae* LAO1 (33.39 Mb) was comparable to that of the VSD-causing fungus *C. theobromae* CT2 draft genome (31.20 Mb; Ali et al.[18]). Furthermore, both have a near identical G+C content (LAO1: 48.93% vs CT2: 49.0%) (Fig. 2A). Comparing our assembly with *C. theobromae* CT2, we infer that we

have described an additional ~ 2 million bp to complete the genome of *C. theobromae*. Notably, none of the contigs produced were assigned to bacteria or any other DNA contaminant (Figure B,C), reaffirming the quality of the colonies grown using our isolation protocol. In addition, we validated with RNA-seq 81.08% of the predicted genes versus 38.83% that were validated in *C. theobromae* CT2 (Supplementary Table 3), likely due to differences in sequencing depth between CT2 (72–104 M reads[18]) and LAO1 (vs 82–120 M reads) assays. Identification and characterization of additional *C. theobromae* isolates including those from indigenous host(s) will help us understand the genomic diversity and population structure of this new encounter pathogen of cassava in SEA.

The fungal secretome has a clear role in plant infections by inhibiting pathogen-associated molecular patterns that trigger plant immunity[29]. The characterization of PEPs involved in virulence, suppression of plant defense and catalytic activity of cell walls and other cellular components, is important as they are promising candidates for future research directed to understand fungal virulence, symptom expression and severity as well as host range. In addition, PEP sequences constitute molecular targets for gene editing to control pathogenicity, and development of protein-based diagnostics for early detection in the field[30]. Interestingly, we identified PEPs that are highly specific for cassava and not found in *C. theobromae* CT2. Given that the xylem is an extracellular niche[31], it is expected that the largest proportion of PEPs found in *C. theobromae* LAO1 corresponds to apoplastic-located proteins. RNAscope® results confirmed the xylem localization of *C. theobromae* LAO1 but also revealed that at certain regions, the fungus can cross the endodermis to colonize epidermal tissues (Fig. 5F). In fact, under certain conditions, hyphae growth can be observed on the branching axillary buds of petioles (Fig. 5G)[6,9]. This plasticity is not exclusive of *C. theobromae* as there are no fungal pathogens that exclusively colonize xylem tissues[31] and require a repertoire of PEPs (both cytoplasmic and apoplastic) to complete their infection cycle.

The localization patterns of pathogens within their host plant are critical for understanding disease dynamics and for the design of effective management strategies. RNAscope® in situ hybridization, is a very sensitive method which can trace sequences present in extremely low concentrations and thus is particularly well suited for invasion studies requiring high resolution and sensitivity[32]. In this method background signals are almost absent giving a clear contrast to identify specific pathogen targets. The probes chosen for CAMK/CAMKL, PEP4, and GADPH, revealed slightly different signal strengths but all were specific and localized *C. theobromae* LAO1 almost exclusively in xylem cells of leaves, petioles and stems of infected plants (Fig. 4, Supplementary Figs. 1, 2). This preference likely arises from the difficulty of penetrating the high-osmotic-pressure of living cells in the phloem compared to the low-osmotic-pressure of dead tracheary elements in the xylem[33]. These results are consistent with studies on VSD of cacao in which *C. theobromae* CT2 was also exclusively found in xylem tissue[11]. In both hosts xylem colonization is reflected macroscopically, as a vascular necrosis observed in cross-sections of stem tissue[6,34].

Intriguingly, targeted analyses on various sections of infected cassava unveiled distinct distribution patterns of *C. theobromae* in different cassava tissues. More strikingly, there was a very tight correlation between symptoms and presence of the pathogen and a general failure to detect *C. theobromae* in non-symptomatic tissues (Fig. 5). This raises questions about disease development and progress. While CWBD has a devastating impact on young plants development, it appears that *C. theobromae* does not systemically colonize a fully grown cassava plant but rather remains localized in and around infection foci (Fig. 5). Thus, the distribution pattern of *C. theobromae* from cassava resembles that of the fungus in cacao: under high humidity conditions, basidiospores penetrate young soft leaves to colonize xylem vessels to eventually reach midrib and petioles, inducing the typical broom symptoms. In cacao, fungal colonization continues through the branch and in severe cases reaches the main stem of the tree and eventually the tree dies[11]. For cassava, it is not evident that the pathogen is effectively colonizing and systemically moving within its host. It rather appears that multiple infection foci from germinating basidiospores penetrating young tissues contribute to the witches' broom syndrome to cause a severe "local" disease (Fig. 5G[6]). The latter will explain previous observations (*not published*), where non-symptomatic portions of the stem of a diseased cassava plant (showing localized broom symptoms), produce apparently healthy plants while stem cutting from symptomatic tissue will soon die. As these observations were made before a molecular diagnosis tool was available[9], it remains to be confirmed whether the fungus detected, as shown with RNAscope®, in non-symptomatic tissue (Fig. 5D) is still viable to start a new infection.

Ceratobasidium spp. are also commonly found as endophytes in roots of weeds as well as in orchids. Thus, it is possible that the fungus originates as an endophyte that has changed its lifestyle to become a serious pathogen. Interestingly, a recent report indicates a likely role of climate change on the emergence of more severe VSD symptoms associated with *C. theobromae* in Indonesia[35]. These new climate events can induce changes in infection biology, new types of symptoms and altered host–pathogen interactions[36]. It is however premature to assume host range expansion of an endogenous *C. theobromae* (as cacao, avocado and cassava are introduced into SEA), rather than host jumping events on these different hosts[23]. However, there is no doubt that the disease continues spreading due to a range of factors, including intensification and expansion of the commercial crops to new regions of cultivation[6,7]. Exploring the diversity of PEPs in different isolates from cassava and cacao, in addition to pathogenicity tests to determine whether *C. theobromae* from cassava can infect cacao or vice versa, are needed to resolve this and other important questions.

We provide evidence that the fungus associated with CWBD in cassava belongs to the same species *C. theobromae*. It takes into account biological characters (xylem localization, binuclear hyphae, formation of fruiting structures emerging from exposed leaf/petiole scars, a nearly obligate parasite niche and minimal growth on standard nutrient media), whole genome analysis of orthologous clusters and the phylogenetic analysis based on single-copy genes (Fig. 3), and the nucleotide identity of specific marker genes (above 98% nucleotide identity for ITS and above 94% for *tef1*, *rpb2* and *atp6* gene sequences) for taxonomic assignment[17]). We have found no evidence for presence of phytoplasma sequences and hence while Koch's postulates are still pending, we can assume that *C. theobromae* is the causative agent of CWBD[9]. This notion has instant consequences for disease control because insecticides used to control insects-transmitting phytoplasmas and antibiotics to treat plants

infected by phytoplasmas[37] are useless to manage this disease. Lastly, it is important to determine the host-specificity of *C. theobromae* isolated from cassava to measure the risk it may pose to other crops. For example, beyond the serious threat to food security, it is unquestionable the negative economic and social impact that the introduction of *C. theobromae* could have on cacao production systems in Africa and the Americas due to its inadvertent introduction through contaminated cacao or cassava planting material. This is not a remote possibility given the ongoing commercialization and investment in these global value chains, including the formal and informal movement of plant genetic resources. Rapid implementation of management strategies based on the current knowledge we have on the identity of the pathogen in both crops systems is needed.

Materials and methods

Plant material, fungus isolation and in vitro culture

Plant material collections were carried out non-destructively and in accordance with relevant institutional guidelines and legislation, in coordination with our partners at the Rice and Cash Crops Research Centre (RCCRC), in Naphok, Vientiane, Lao PDR. To isolate *Ceratobasidium* sp. from cassava, we collected petioles (3–4 cm long) from the brooms formed in the middle part of an infected plants (8–10 months after planting) collected from CIAT experimental field located at RCCRC. We avoided petioles from the older parts of the plant, as they could have acervuli structures of *Colletotrichum* sp. (a common contaminant) that is present in senescent leaves. Petioles were disinfected using 1% hypochlorite solution followed by 75% ethanol. After rinsing with sterile distillate water, the petioles were dried, placed in a petri dish containing water agar (WA) and incubated at 25 °C for 48 h. After this time mycelial growth was observed at both ends of each petiole. Petri dishes were daily inspected under optical light microscopy for identification of fungal structures, and to check for fast growing endophytes and contaminants. When *Colletotrichum* sp. structures were observed, we transferred the samples to potato dextrose agar (PDA) containing 1 ml of benomyl solution (0.02% in water) and 10 ml of lactic acid (25% in water) per 1 L of PDA media[26,27]. The petioles with *Ceratobasidium* sp. mycelia were incubated in PDA + benomyl + lactic acid for 1 week and subcultured to the same media until a pure culture was observed. To increase the biomass of the pure isolated fungus for DNA extraction, we transferred the mycelia to PDA Corticium culture medium (PDA + CCM) (39 g PDA, 1.25 g KH_2PO_4, 0.625 g Mg_2SO_4, 6.25 g Maltose, 6.25 g Malt extract, 0.625 g peptone in 1000 ml of water)[38]. PDA + CCM provides maltose, peptone as nitrogen supply, and the essential inorganic salts dihydrogen potassium phosphate and magnesium sulphate that favor fungal growth. The CCM-supplemented media was developed to stimulate pycnidia formation in fungi from the order Sphaeropsidales[39] and was already used to cultivate *C. theobromae*[12,19,38]. We kept the isolated fungus growing in PDA + CCM media for 2–4 weeks before DNA extraction.

DNA extraction

Total DNA was extracted from a fungal colony grown for 2–4 weeks in PDA + CCM medium, by scraping out the mycelium (80 mg per petri dish) from agar medium and transferring it to 2 ml plastic tubes. The mycelium is then macerated with plastic pestles by hand in 200 µl of extraction buffer (15 mM NaCl, 50 mM Tris pH 8.0, 10 mM Na_2EDTA, 1% [w/v] SDS) plus 1 µl of Proteinase K (10 mg/ml), freshly added. The suspension was vortexed until we obtained a homogeneous solution and then incubated at 65 °C for 1 h before adding 200 µl of 3M sodium acetate and incubate the mix at room temperature (RT) for 10 min. After centrifugation at 13,400×*g* for 25 min at RT, the supernatants were transferred to new tubes with an equal volume (~ 400 µl) of chloroform:isoamyl alcohol (24:1), and mixed by inversion for 15 min. Afterwards, the solutions were centrifuged at 13,400×*g* for 5 min at 4 °C and the supernatant was recovered. The supernatant was then transferred to a new tube, and an equal volume of ice-cold isopropanol (~ 400 µl) was added. To favor precipitation the mixes were incubated at − 20 °C overnight and then centrifuged at 13,400×*g* for 10 min. The pellet was washed with 250 µl of ice cold 70% ethanol, gently shaking the tubes to detach the pellet from the bottom, then the tubes were centrifuged at 13,400×*g* for 5 min at 4 °C. The pellets were dried for at least 1 h at 37 °C in an incubator with the tubes placed upside down on paper towel. Finally, the pellets were resuspended in 30 µl of nuclease-free water and stored at 4 °C. Qubit™ dsDNA HS Assay Kit (Invitrogen, Life Technologies, USA) was used to quantify total DNA, obtaining an average of 850 ng/µl of total DNA.

DNA sequencing

For whole genome sequencing, we followed a hybrid strategy. Short read sequencing was conducted on an Illumina NovaSeq 6000 sequencer (Novogene, USA) using a 150 bp paired-end protocol. Sequence quality was inspected using FastQC[40] and low-quality sequences and adapters were discarded with Trimmomatic[41]. Long read sequencing was performed using Oxford Nanopore Technologies (ONT) MinION sequencer (Oxford Nanopore Technologies, UK). The sequencing library was prepared using Ligation Sequencing Kit SQK-LSK110 following the manufacturer's instructions and processed on a MinION R9.4 flow cell (FLO-MIN106D) for 72 h. Base calling of ONT reads (min quality score = 7) was performed with Guppy basecaller v6.4.2 using the high-accuracy algorithm (*hac*).

Genome assembly

A hybrid assembly was performed with 42 Gb of Illumina and ONT reads using Unicycler v0.4.8[42] (–min_fasta_length = 1000) and assembly metrics were calculated with BBMap v38.18[43]. DIAMOND v0.9.24[44] was used to compare contigs against the NCBI non-redundant (nr) protein database (2023-07-28) using BLASTx (e-value < 1e−25). This allowed us to detect and discard possible prokaryote sequences. Illumina and ONT reads were then aligned against our assembled genome with Minimap2[45], using default parameters. SAMtools v1.3.1[46] was used to convert the aligned (SAM) file to a binary (BAM) file using view -S and -b options and sorted and

indexed with the sort and index options. Coverage was calculated with Qualimap v2.2.2[47] and completeness was determined with BUSCO v5.5.0[48] using the agaricomycetes_odb10 (2024-01-08) and eukaryota_odb10 (2024-01-08) lineages as reference.

Ab initio genes prediction and comparative genomics

Gene prediction was performed with AUGUSTUS v4.4.0[49] trained with a closely related *Rhizoctonia solani* (GenBank: GCA_016906535.1) (*R. solani* AG-1 IA) gene model. A genome-wide inspection of orthologous clusters was performed with OrthoVenn3 (https://orthovenn3.bioinfotoolkits.net/home, accessed on 19 January 2024)[50] using OrthoMCL algorithm (e-value = 1e−10) by comparing *Ceratobasidium* sp. isolated from cassava (*Ceratobasidium* sp. LAO1) with *C. theobromae* isolated from cacao (GenBank: GCA_009078325.1) (*C. theobromae* CT2) and *R. solani* AG-1 IA. Mash v2.3[51] was used to calculate nucleotide distance between assembled genomes of *Ceratobasidium* sp. LAO1 and *C. theobromae* CT2 with -k 16 -s 100,000 specified options. Genomic distance threshold (k-mer size = 16; genomic distance < 0.04317) was used as a genetic criterion to determine identity at the species level of two genomes[52]. Additionally, nucleotide distance between *Ceratobasidium* sp. LAO1, *C. theobromae* CT2 and *R. solani* AG-1 IA was calculated to verify the ability of Mash to differentiate between closely related genera given these parameters. In addition, the nucleotide identity of 4 molecular markers (*atp6*, *tef1*, *rpb2* and ITS) used in the identification of fungi was evaluated by BLASTn (e-value < 1e−25). Finally, predicted proteins were annotated using BlastKOALA for individual gene KO identifier assignation[53], and InterProScan protein family's database[54]. Additionally, DIAMOND-BLASTp (e-value < 1e−25) analysis was conducted to identify sequence homology against nr protein database (2023-07-28).

Prediction and characterization of putative effectors proteins

PEPs were identified from the secretome of predicted protein sequences using similar criteria as FunSecKB2[55]. In the first screening, all sequences with a signal peptide predicted by SignalP v5.0[56] were included, and proteins with a transmembrane helix domain detected by TMHMM v2.0[57] were discarded. Subsequently, TargetP v2.0[58] was used to filter-out proteins with mitochondrial transit peptide (mTP), ScanProsite web server[59] to filter-out proteins with endoplasmic reticulum targeting sequence (PS00014) and NetGPI v1.1[60] to filter-out carrier proteins of glycophosphatidylinositol (GPI) anchoring motifs (likehood ≥ 0.90). This constitutes the predicted secretome. A list of apoplastic and cytoplasmic PEP was further identified from the predicted secretome using EffectorP v3.0[30] and characterized by DIAMOND-BLASTp (e-value < 1e−25) against nr protein database (2023-07-28).

Transcriptome sequencing and analysis

Transcriptomic analysis was conducted to validate predicted genes using petioles form three diseased cassava plants collected from CIAT experimental field mentioned above (Naphok, Vientiane, Lao PDR). RNA extraction was performed using 200–400 mg of petioles and CTAB (*N*-Cetyl-*N,N,N*-Trimethyl-Ammonium Bromide. Merck) as previously described[61] and quantified using Nanodrop 2000c spectrophotometer (ThermoFisher, USA). Quality was assessed with RNA OD260/280 ratio OD230/280 and visualized using 1% agarose gel electrophoresis. Total RNA was sequenced with Illumina NovaSeq 6000 platform (Novogene, USA) using a 150 bp paired-end whole transcriptome sequencing protocol. Quality control was performed using FastQC[40] and Trimmomatic[41] to discard adapter and low-quality score sequences. Trimmed reads were mapped against the genome of *Ceratobasidium* sp. LAO1 using STAR[62]. SAMtools[46] was used to convert alignment files (SAM) to binary format (BAM) as described above, and separate mapped from unmapped reads with –F option specified. Finally, featureCounts from *subRead* package (http://subread.sourceforge.net/) was used to determine transcripts as raw counts.

In situ localization using RNAscope® technique

Design and synthesis of RNAscope® probes

For *Ceratobasidium* sp. localization using RNAscope®, Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as housekeeping gene, Calcium (Ca2+)/calmodulin-dependent protein kinases (CAMK/CAMKL) which has been used for diagnostics[9] and PEP gene 4 (PEP4) with high sequence coverage, were chosen for specific detection and tissue localization. The probes, designed by ACD (Advanced Cell Diagnostics, Hayward, USA), are available in the ACD catalog as RNAscope™ Probe—F-Ceratobasidium-CaMK-C1 (Cat No. 1230851-C1) with targeting region 208–1150 and RNAscope™ Probe—F-Ceratobasidium-GAPDH-O1-C1 (Cat No. 1274121-C1) with targeting region 2–1017 and RNAscope™ Probe—F-PEP4-C1 (Cat No. 1280721-C1) with targeting region 2–879.

Fixation, embedding, and sectioning of plant tissue

Leaves, stems, and petioles samples with CWBD and from healthy cassava plants were collected from CIAT experimental field mentioned above (Naphok, Vientiane, Lao PDR). Small pieces of plant tissues (5 mm in length) were excised and fixed under vacuum in 10% neutral buffered formalin (Sigma-Aldrich, USA) for 45 min. After exchanging the fixing solution and an additional incubation step for 45 min at RT, the samples were transferred to fresh formalin and kept under vacuum for 16 h. Subsequently, the samples underwent two washes in Diethyl Pyrocarbonate (DEPC) treated PBS buffer (PBS, pH 7.4) for 15 min each, followed by dehydration washing in a series of increasing ethanol concentrations (30%, 50%, 70%) for 30 min at each concentration, and then stored at 4 °C. For formalin-fixed paraffin embedding (FFPE), the samples underwent two steps of ethanol infiltration at 95% and 100% ethanol for 30 min each before the substitution of ethanol with xylene. This substitution process involved stepwise incubation of the plant tissues at RT for 45 min in ethanol/xylene mixtures with decreasing portions of ethanol (2:1, 1:1, 1:2 (v/v)) and pure xylene. Subsequently, xylene was replaced with paraffin by incubating the samples in mixtures of xylene/paraffin (2:1, 1:1, 1:2 (v/v)) and pure paraffin at 60 °C

for 1 h each step. The samples embedded in pure paraffin were then transferred into Peel-A-Way molds and kept at RT. Before sectioning, the paraffin blocks were cooled at 4 °C, and sections of approximately 12 µm were cut using a Microm HM 355 rotary microtome (Thermo-Fisher Scientific, USA). These sections were obtained as paraffin ribbons from which two sections were separated, relaxed in a water bath at 37 °C, and then placed on Superfrost Plus slides. The sections were dried overnight at RT and fixed by baking for 1 h at 60 °C. Directly after baking, the sections were deparaffinized by incubating them twice in xylene for 5 min, followed by ethanol washing, and then stored in the dark.

In situ localization using RNAscope® for single target detection

RNAscope® in situ hybridization (ISH) was conducted essentially following the protocol outlined by Ref.[63] using the RNAscope® 2.5 HD Detection Reagent-RED kit (cat. no. 322360) from ACD. Initially, slides were baked at 60 °C for 30 min and allowed to cool for 20 min. Subsequently, a hydrophobic barrier was drawn around the sections using an ImmEdge hydrophobic barrier pen (Biozol diagnostic, Germany). To inhibit endogenous peroxidases, the sections were treated with hydrogen peroxide for 10 min at RT, followed by a single wash with distilled water. Next, the sections underwent incubation in target retrieval solution for 15 min (at 98 °C to 102 °C). This step aimed to disrupt crosslinking resulting from tissue fixation. The tissue sections were then air-dried overnight at RT. On the subsequent day, samples were treated with 'Protease Plus' solution at 40 °C for 15 min to render RNA accessible. This was followed by two washes in water. Hybridization was performed by incubating the specific probes at 40 °C for 2 h in the provided hybridization solution, followed by washing the sections for 15 min with two buffer exchanges. Serial amplification steps, as described by Munganyinka et al.[63], were then carried out to enhance the signal. Finally, the signal was developed using an alkaline phosphatase-labeled probe and Fast-Red as a substrate, resulting in a red signal. Slides were subsequently washed in water, counterstained with 50% hematoxylin (Gill's Hematoxylin, Sigma-Aldrich, USA) for 2 min, and rinsed several times in distilled water. Sections were baked at 60 °C for 45 min, briefly submerged in xylene, and covered with EcoMount mounting media (Biocare Medical, USA) and micro cover glasses (24 × 50 mm). The slides were then air-dried for at least 10 min at RT before examination of the sections.

Imaging of plant tissue sections

Leaf, stem and petiole sections from both healthy and diseased cassava were subjected to examination using bright-field microscopy. The Olympus SZX16 stereomicroscope (Olympus, Germany) was utilized for this purpose. Images were captured using 10× objective. The signal appeared as reddish points or clusters.

Data availability

The sequences reported in this study has been deposited in the NCBI database with BioProject PRJNA1076853, BioSample SAMN40039001 and SAMN40581416 and whole genome assembly of C. theobromae LAO1 JBBBZN000000000. The data presented in this study are available in the following Supplementary Materials section.

Received: 26 March 2024; Accepted: 31 July 2024
Published online: 05 August 2024

References

1. Newby, J., Smith, D., Cramb, R., Delaquis, E. & Yadav, L. Cassava value chains and livelihoods in Southeast Asia, a regional research symposium held at Pematang Siqantar, North Sumatra, Indonesia, 1–5 July 2019. In *ACIAR Proceedings Series No 148* (eds. Newby, J. *et al.*) 114 (Australian Centre for International Agricultural Research, 2020).
2. Howeler, R., Lutaladio, N. & Thomas, G. *Save and Grow: Cassava. A Guide to Sustainable Production Intensification* 87–97 (Food and Agriculture Organization of the United Nations, 2013).
3. Malik, A. I. *et al.* Cassava breeding and agronomy in Asia: 50 years of history and future directions. *Breed. Sci.* **70**(2), 145–166 (2020).
4. Graziosi, I. *et al.* Emerging pests and diseases of South-east Asian cassava: A comprehensive evaluation of geographic priorities, management options and research needs. *Pest Manag. Sci.* **72**(6), 1071–1089 (2016).
5. Siriwan, W. *et al.* Surveillance and diagnostics of the emergent Sri Lankan cassava mosaic virus (Fam. Geminiviridae) in Southeast Asia. *Virus Res.* **285**, 197959 (2020).
6. Pardo, J. M. *et al.* Cassava witches' broom disease in Southeast Asia: A review of its distribution and associated symptoms. *Plants* **12**(11), 2217 (2023).
7. Dolores, L. M. *et al.* Incidence, distribution, and genetic diversity of 'Candidatus Phytoplasma luffae'-related strain (16SrVIII) associated with the cassava witches' broom (CWB) disease in the Philippines. *Crop Prot.* **169**, 106244 (2023).
8. Keane, P. J., Flentje, N. T. & Lamb, K. P. Investigation of vascular-streak dieback of cocoa in Papua New Guinea. *Aust. J. Biol. Sci.* **25**(3), 553–564 (1972).
9. Leiva, A. M. *et al. Ceratobasidium* sp. is associated with cassava witches broom disease, a re-emergent threat to cassava cultivation in Southeast Asia. *Sci. Rep.* **13**, 22500 (2023).
10. Manawasinghe, I. S. *et al.* Defining a species in fungal plant pathology: Beyond the species level. *Fungal Divers.* **109**(1), 267–282 (2021).
11. Samuels, G. J. *et al.* Vascular streak dieback of cacao in Southeast Asia and Melanesia: In planta detection of the pathogen and a new taxonomy. *Fungal Biol.* **116**(1), 11–23 (2012).
12. Lam, C. H., Varghese, G. & Zainal Abidin, M. A. Z. In vitro production of *Oncobasidium theobromae* basidiospores. *Trans. Br. Mycol. Soc.* **90**(3), 505–507 (1988).
13. Uchida, J. Y., Aragaki, M. & Yahata, P. S. Basidiospore formation by *Ceratobasidium* sp. on agar. *Mycology* **78**(4), 587–592 (1986).
14. Holderness, M. Control of vascular-streak dieback of cocoa with triazole fungicides and the problem of phytotoxicity. *Plant Pathol.* **39**(2), 286–293 (1990).
15. Talbot, P. H. B. & Keane, P. J. *Oncobasidium*: A new genus of tulasnelloid fungi. *Aust. J. Bot.* **19**, 203–206 (1971).

16. Samuels, G. J. & Ismaiel, A. *Trichoderma evansii* and *T. lieckfeldtiae*: Two new *T. hamatum*-like species. *Mycology* **101**(1), 142–156 (2009).
17. Gonzalez, D. *et al.* Phylogenetic relationships of *Rhizoctonia* fungi within the Cantharellales. *Fungal Biol.* **120**(4), 603–619 (2016).
18. Ali, S. S. *et al.* Draft genome sequence of fastidious pathogen *Ceratobasidium theobromae*, which causes vascular-streak dieback in *Theobroma cacao*. *Fungal Biol. Biotechnol.* **6**, 1–10 (2019).
19. Junaid, M. & Guest, D. Modified culture assay to obtain a diversity of hyphal structures of *Ceratobasidium theobromae*-VSD pathogen on cocoa. *Biodivers. J. Biol. Div.* **22**(4), 1 (2021).
20. Pecchia, S. *et al.* Molecular detection of the seed-borne pathogen *Colletotrichum lupini* targeting the hyper-variable IGS region of the ribosomal cluster. *Plants* **8**(7), 222 (2019).
21. Rapicavoli, J., Ingel, B., Blanco-Ulate, B., Cantu, D. & Roper, C. *Xylella fastidiosa*: An examination of a re-emerging plant pathogen. *Mol. Plant Pathol.* **19**(4), 786–800 (2018).
22. Morse, S. S. Plagues and politics. In *Infectious Disease and International Policy* (ed. Price-Smith, A. T.) (Palgrave Macmillan, 2001).
23. Thines, M. An evolutionary framework for host shifts-jumping ships for survival. *New Phytol.* **224**, 605–617 (2019).
24. Guest, D. & Keane, P. Vascular-streak dieback: A new encounter disease of cacao in Papua New Guinea and Southeast Asia caused by the obligate basidiomycete *Oncobasidium theobromae*. *Phytopathology* **97**(12), 1654–1657 (2007).
25. Freestone, M. W. *et al.* Continental-scale distribution and diversity of *Ceratobasidium* orchid mycorrhizal fungi in Australia. *Ann. Bot.* **128**(3), 329–343 (2021).
26. Kobayashi, T. *et al.* Mushroom yield of cultivated shiitake (*Lentinula edodes*) and fungal communities in logs. *J. For. Res.* **25**(4), 269–275 (2020).
27. Lu, D. S. *et al.* Reticulate evolution and rapid development of reproductive barriers upon secondary contact pose challenges for species delineation in a forest fungus. *BioRxiv* **1**, 1 (2023).
28. Sangpueak, R., Phansak, P. & Buensanteai, N. Morphological and molecular identification of Colletotrichum species associated with cassava anthracnose in Thailand. *J. Phytopathol.* **166**(2), 129–142 (2018).
29. Lo Presti, L. *et al.* Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* **66**, 513–545 (2015).
30. Sperschneider, J. & Dodds, P. N. EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant-Microb. Interact.* **35**(2), 146–156 (2022).
31. De La Fuente, L., Merfa, M. V., Cobine, P. A. & Coleman, J. J. Pathogen adaptation to the xylem environment. *Annu. Rev. Phytopathol.* **60**, 161–186 (2022).
32. Sheat, S., Margaria, P. & Winter, S. Differential tropism in roots and shoots of resistant and susceptible cassava (*Manihot esculenta* Crantz) infected by Cassava brown streak viruses. *Cells* **10**(5), 1221 (2021).
33. Yadeta, K. & Thomma, B. The xylem as battleground for plant hosts and vascular wilt pathogens. *Front. Plant Sci.* **4**(97), 43568 (2013).
34. Keane, P. J. & Prior, C. *Vascular-Streak Dieback of Cocoa. Phytopathological Papers No. 33* (Commonwealth Mycological Institute, 1991).
35. Bryceson, S. R., Morgan, J. W., McMahon, P. J. & Keane, P. J. A sudden and widespread change in symptoms and incidence of vascular streak dieback of cocoa (*Theobroma cacao*) linked to environmental change in Sulawesi, Indonesia. *Agric. Ecosyst. Environ.* **350**, 108466 (2023).
36. Singh, B. K. *et al.* Climate change impacts on plant pathogens, food security and paths forward. *Nat. Rev. Microbiol.* **21**(10), 640–656 (2023).
37. Landicho, D. & Reyes, J. *Plant Quarantine Policy Formulation and Implementation for Cassava Witches' Broom (CWB) Disease Control in Bukidnon, Philippines* (University of the Philippines Los Banos, 2019).
38. Varghese, G., Abidin, Z. & Mainstone, B. J. Vascular streak dieback of cocoa in Malaysia. II. Isolation and culture techniques of causal pathogen. *Planter* **57**, 667 (1981).
39. Leonian, L. H. A study of factors promoting pycnidium-formation in some Sphaeropsidales. *Am. J. Bot.* **11**(1), 19–50 (1924).
40. Brown, J., Pirrung, M. & McCue, L. A. FQC dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **33**(19), 3137–3139 (2017).
41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15), 2114–2120 (2014).
42. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**(6), e1005595 (2017).
43. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner* (Lawrence Berkeley National Lab, 2014).
44. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**(1), 59–60 (2015).
45. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**(18), 3094–3100 (2018).
46. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).
47. García-Alcalde, F. *et al.* Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* **28**(20), 2678–2679 (2012).
48. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**(19), 3210–3212 (2015).
49. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, 309–312 (2004).
50. Sun, J. *et al.* OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* **51**(1), 397–403 (2023).
51. Ondov, B. D. *et al.* Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**(1), 1–14 (2016).
52. Gostinčar, C. Towards genomic criteria for delineating fungal species. *J. Fungi* **6**(4), 246 (2020).
53. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and BlastKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**(4), 726–731 (2016).
54. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**(9), 1236–1240 (2014).
55. Meinken, J. *et al.* FunSecKB2: A fungal protein subcellular location knowledgebase. *Comput. Mol. Biol.* **4**, 4 (2014).
56. Almagro-Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**(4), 420–423 (2019).
57. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**(3), 567–580 (2001).
58. Armenteros, J. J. A. *et al.* Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**, 5 (2019).
59. Gattiker, A., Gasteiger, E. & Bairoch, A. ScanProsite: A reference implementation of a PROSITE scanning tool. *Appl. Bioinform.* **1**(2), 107–108 (2002).
60. Gíslason, M. H., Nielsen, H., Armenteros, J. J. A. & Johansen, A. R. Prediction of GPI-anchored proteins with pointer neural networks. *Curr. Res. Biotechnol.* **3**, 6–13 (2021).
61. Jimenez, J., Leiva, A. M., Olaya, C., Acosta-Trujillo, D. & Cuellar, W. J. An optimized nucleic acid isolation protocol for virus diagnostics in cassava (*Manihot esculenta* Crantz.). *MethodsX* **8**, 101496 (2021).
62. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1), 15–21 (2013).

63. Munganyinka, E. *et al.* Localization of cassava brown streak virus in *Nicotiana rustica* and cassava *Manihot esculenta* (Crantz) using RNAscope® in situ hybridization. *Virol. J.* **15**, 1–11 (2018).

Acknowledgements
This research was funded by the Australian Centre for International Agricultural Research (ACIAR) through a Short Research Activity (SRA) Grant CROP/2023/157. We acknowledge financial support from the United States Agency for International Development (USAID) and CGIAR's Plant Health (PHI) and Accelerated Breeding (ABI) initiatives, and the ASEAN-CGIAR Program IP5 on Transboundary Diseases. We are thankful to all our colleagues from the Global Cassava Program for their help with laboratory and field activities. Thanks also to Viviana Dominguez for laboratory support and to Dr. Diana Lopez-Alvarez from Universidad Nacional (UNAL) in Colombia and Dr. Jorge Díaz Valderrama from Universidad Toribio Rodriguez de Mendoza (UNTRM) in Peru, for helpful advice with bioinformatic analysis and molecular characterization of fungi.

Author contributions
A.G.-O. designed and performed bioinformatic analyses. J.M.P. did microbiological isolation, and characterization of the fungus. S.S. carried out in situ localization. A.G.-O and A.M.L performed Nanopore sequencing. K.X., W.A. and W.J.C. collected plant material and isolated RNA for sequencing. A.G.-O., J.M.P. and S.S. prepared figures and tables. S.W., J.N. and W.J.C. overall supervision of the activities. A.G.-O., A.M.L and W.J.C. performed sequence data curation. W.J.C. designed research and wrote the main manuscript text. All authors reviewed, contributed and agreed with the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-69061-8.

Correspondence and requests for materials should be addressed to W.J.C.

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

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