

OPEN

Colletotrichum shiso sp. nov., an anthracnose pathogen of *Perilla frutescens* in Japan: molecular phylogenetic, morphological and genomic evidence

P. Gan¹, A. Tsushima^{1,2}, R. Hiroyama¹, M. Narusaka³, Y. Takano⁴, Y. Narusaka³, M. Kawaradani⁵, U. Damm⁶ & K. Shirasu^{1,2}

Species of the fungal genus *Colletotrichum* are among the most devastating pathogens of agricultural crops in the world. Based on DNA sequence data (ITS, *GAPDH*, *CHS-1*, *ACT*, *TUB2*) and morphology, we revealed *Colletotrichum* isolates infecting the oil crop *Perilla frutescens*, commonly known as shiso, to represent a previously unknown species of the *C. destructivum* species complex and described it as *C. shiso*. We found that *C. shiso* appears to be able to adopt a hemibiotrophic lifestyle, characterised by the formation of biotrophic hyphae followed by severe necrotic lesions on *P. frutescens*, but is less virulent on *Arabidopsis*, compared to its close relative *C. higginsianum* which also belongs to the *C. destructivum* species complex. The genome of *C. shiso* was sequenced, annotated and its predicted proteome compared with four other *Colletotrichum* species. The predicted proteomes of *C. shiso* and *C. higginsianum*, share many candidate effectors, which are small, secreted proteins that may contribute to infection. Interestingly, *C. destructivum* species complex-specific secreted proteins showed evidence of increased diversifying selection which may be related to their host specificities.

Perilla frutescens, or shiso, is a herbaceous plant belonging to the Lamiaceae family and was originally cultivated throughout East and South-East Asia as a culinary herb, oil source and as a traditional medicine^{1,2}. The plant produces a significant amount of volatile compounds³ that contribute to its distinctive odour, as well as a high amount of polyphenols with antioxidant activities⁴. Apart from direct consumption, *P. frutescens* is also used industrially, whereby oil from the seed may be used as a drying oil². In Korea, perilla oil has consistently been the third highest domestically produced vegetable oil with a volume of more than 30,000 metric tonnes produced in 2015/16⁵. In Japan, 6,708 tonnes of *P. frutescens* were reported to be grown for direct consumption, while a further 2,763 tonnes were grown for industrial uses in 2014⁶. In the West, perilla has been grown as an ornamental plant since the Victorian Era, where it is known as the “beefsteak plant”⁷.

Fukui (1925) first reported an anthracnose disease of perilla in Japan and described the causal organism as *Colletotrichum yoshinaoi* Fukui⁸. In a study from Korea, the causal agents of perilla anthracnose were identified as *C. gloeosporioides*, *C. coccodes*, *C. dematium* and *Glomerella cingulata*, which had previously been regarded as the sexual morph of *C. gloeosporioides*⁹. In the study, the authors hypothesised the species described by Fukui to be a synonym of *C. gloeosporioides*, since *C. gloeosporioides* was more frequently isolated from infected plants⁹. However, none of these reports was confirmed by molecular data and the systematic position of *C. yoshinaoi*, which lacks a living type strain, is unknown.

¹RIKEN Center for Sustainable Resource Sciences, Yokohama, Kanagawa, Japan. ²Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo, Japan. ³Research Institute for Biological Sciences, Okayama Prefectural Technology Center for Agriculture, Forestry, and Fisheries, Okayama, Japan. ⁴Graduate School of Agriculture, Kyoto University, Kyoto, Japan. ⁵Research Institute of Environment, Agriculture and Fisheries, Osaka, Japan. ⁶Senckenberg Museum of Natural History Görlitz, 02806, Görlitz, Germany. Correspondence and requests for materials should be addressed to K.S. (email: ken.shirasu@riken.jp)

Received: 11 January 2019
Accepted: 5 September 2019
Published online: 16 September 2019

More recently, *C. destructivum* was identified as being responsible for perilla anthracnose in Japan using ITS sequences¹⁰, while fungal strains causing anthracnose of *Lamium amplexicaule* (henbit), that also belongs to the Lamiaceae, were identified as *C. higginsianum* based on ITS sequences¹¹. However, all previous reports were carried out before the epitypification of the respective species and the treatments of the respective species complexes^{12–16} and were based on morphology or ITS only.

With the advent of affordable high throughput genome sequencing technologies, the genomes of multiple members of the *Colletotrichum* genus of plant pathogenic fungi have been sequenced and released^{17–27}. Sequenced genomes have included strains belonging to different species complexes, which comprise closely related species that are phylogenetically distinct from other members of the same genus. Members of the same species complex exhibit similarities in terms of their infection lifestyles and whole genome comparative analyses have revealed genomic adaptations that contribute to these differences^{17,18,28}. For example, members of the *Colletotrichum graminicola* species complex which specifically infect graminaceous hosts have reduced numbers of pectin-degrading enzymes^{17,18}. Comparisons between different members of the same species complex, such as between *C. sublineola* and *C. graminicola*, which infect sorghum and maize respectively, have also been performed; this led to the identification of genes that were not found to be conserved between different members of the same species complex, and which may contribute to adaptation to their specific host niches²⁹.

The aims of this study were to characterise one of the causal agents of anthracnose of *P. frutescens* in Japan based on multi-locus sequence data and morphology. Further, we aimed to characterise the species at the molecular level by sequencing and assembly of its genome. In addition, we wanted to identify what types of genes show different conservation patterns between shiso-infecting *Colletotrichum* and close relatives in the *Colletotrichum* genus. In particular, we aimed to analyse the conservation patterns of genes encoding small, secreted proteins, since these may contribute to differences in infection outcomes.

Results

Multi-locus phylogenetic analysis. An initial BLASTn search of the NCBI non-redundant nucleotide database using the internal transcribed spacers (ITS) sequence from *Colletotrichum* strain JCM 31818 from *P. frutescens* as a query was conducted, revealing that seven of the top ten hits, differing by 7–8 mismatches, belong to the *C. destructivum* species complex (Supplementary Table S1). As strain MAFF 240106 was also isolated from *P. frutescens* and was previously identified as *C. destructivum* on the basis of its ITS sequence¹⁰, sequences from both strains were compared and found to be identical. In order to identify these strains to the species level, a phylogenetic tree based on ITS, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), chitin synthase 1 (*CHS-1*), actin (*ACT*) and beta-tubulin (*TUB2*) sequences was calculated and used for comparison of the strains from *P. frutescens* with all currently accepted species in the *C. destructivum* species complex (Supplementary Table S2). DNA sequences obtained from the MAFF Genbank project of several strains isolated from *L. amplexicaule*, a host from the same family as *P. frutescens*, which had previously been identified as *C. higginsianum* based on ITS sequences¹¹, were also included (Supplementary Table S2).

In the multi-locus phylogenetic analysis, sequences were aligned, trimmed and then concatenated to generate a sequence alignment comprising 1,778 characters (gene boundaries ITS: 1–546, *GAPDH*: 547–740, *CHS-1*: 741–1,020, *ACT*: 1,021–1,275, *TUB2*: 1,276–1,778) from 94 isolates.

In maximum parsimony analyses, 1,318 characters were found to be constant, while 266 and 194 of the variable characters were found to be parsimony informative and uninformative respectively. The heuristic search yielded 64 equally most parsimonious trees (tree length: 659, consistency index (CI): 0.819, retention index (RI): 0.931, rescaled consistency index (RC): 0.763, homoplasy index (HI): 0.181). Analysis of the concatenated alignment as well as alignments of each individual gene indicated that the strains from *P. frutescens* are distinct from the other members of the *C. destructivum* species complex (Supplementary Figs 1–6) and may represent a separate species.

To confirm this, maximum likelihood and Bayesian phylogenetic analyses were carried out. The best model for phylogenetic analysis of *ACT*, *CHS-1*, *GAPDH*, ITS and *TUB2* was calculated as HKY + G, K80, HKY + I, K80 + I + G and K80 + I, respectively. The consensus tree obtained from Bayesian analysis of the multi-locus alignment showed the strains from *P. frutescens* form a distinct clade on a long branch with a Bayesian posterior probability value of 1.00 (Fig. 1), while the strains from *L. amplexicaule* (MAFF 244502, 244503) clustered with *C. higginsianum*, confirming their identities as *C. higginsianum* strains. In each consensus tree of individual loci generated by Bayesian analysis (Supplementary Figs 7–11), the strains obtained from *P. frutescens* formed a distinct clade within the *C. destructivum* species complex with Bayesian posterior probability values above 0.9. However, the position of this clade containing isolates from *P. frutescens* differed depending on the locus. The topologies of the ML trees calculated from the single and multi-locus alignments were consistent with the results from the Bayesian analyses (Supplementary Figs 12–17).

Taxonomy. Based on the DNA sequence data, the *Colletotrichum* species from *P. frutescens* was found to be distinct from other species in the *C. destructivum* species complex and therefore described as a new species below.

Colletotrichum shiso. P. Gan, A. Tsushima, M. Kawaradani, Damm & K. Shirasu sp. nov., Mycobank MB 828333, Fig. 2.

Etymology. Refers to the host from which the species was isolated, *Perilla frutescens* var. *crispa*, commonly known as shiso.

Sexual morph not observed. **Asexual morph on synthetic nutrient-poor agar (SNA).** Vegetative hyphae 3–5.5 μm , average (av.) \pm standard deviation (SD): 4.0 \pm 1.0 μm diam, hyaline, smooth-walled, septate and branched. **Chlamydospores** not observed. **Conidiomata** absent, conidiophores formed directly on conidiogenous cells.

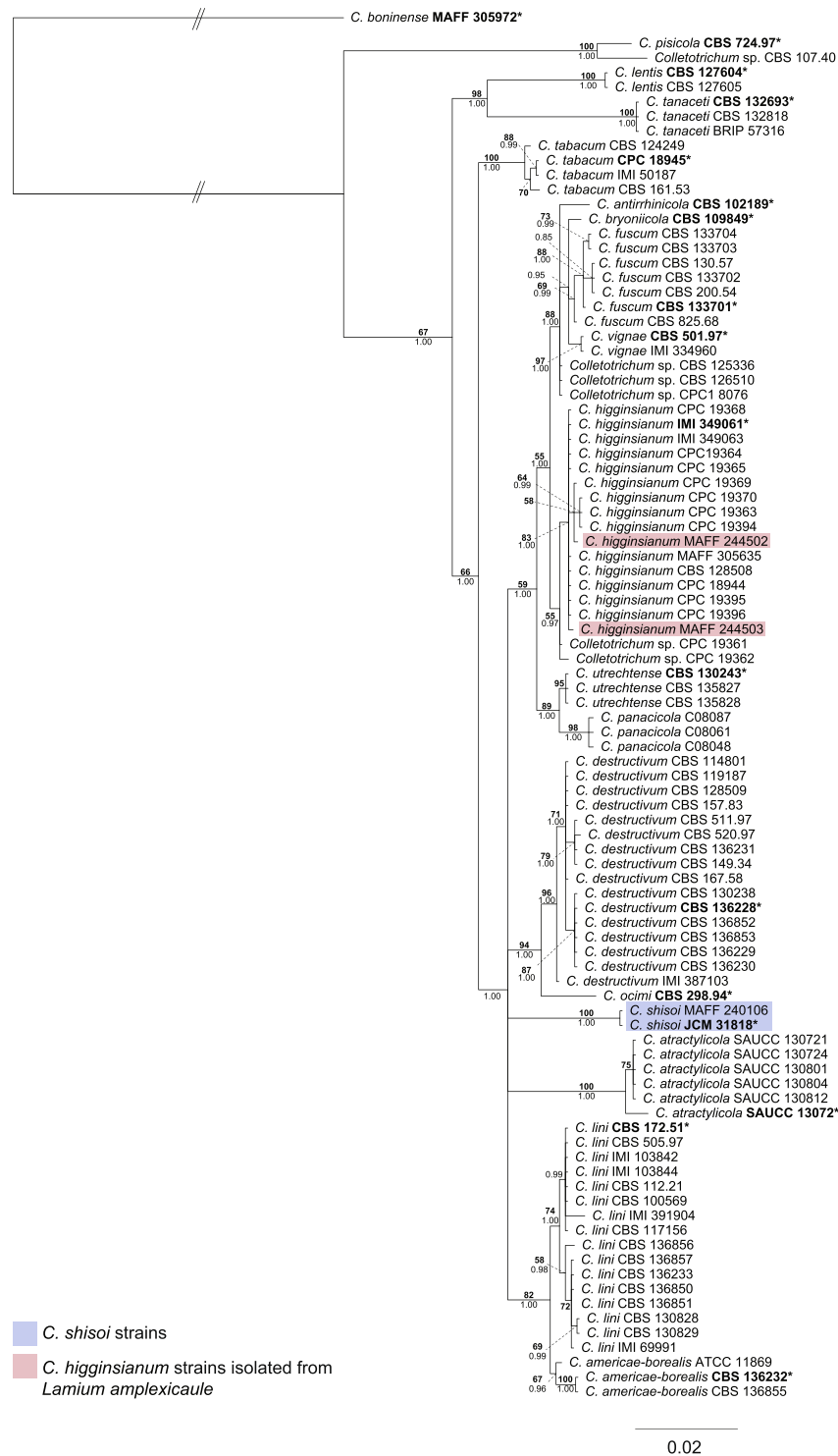


Figure 1. Multi-locus phylogenetic tree based on ITS, *GAPDH*, *CHS-1*, *ACT* and *TUB2* sequences of the *Colletotrichum destructivum* species complex using *Colletotrichum boninense* MAFF 305972 as an outgroup. Values at the nodes are Bayesian posterior probability values above 50%. *C. shisoii* strains are highlighted in blue. *C. higginsianum* strains isolated from *Lamium amplexicaule* are highlighted in red. Ex-type cultures are marked with an asterisk and in bold. Branches with double-bars are truncated two-fold.

Setae not observed *Conidiogenous cells* hyaline to pale brown, smooth-walled, septate, cylindrical. *Conidia* hyaline, smooth-walled, aseptate, cylindrical, straight, with both ends round or with one end slightly tapered, $13\text{--}22.5\ \mu\text{m} \times 3\text{--}5\ \mu\text{m}$, av. \pm SD: $16.5 \pm 2.5\ \mu\text{m}$ diam $\times 4.0 \pm 0.5\ \mu\text{m}$, L/W ratio: 4.5. *Appressoria* single, olivaceous

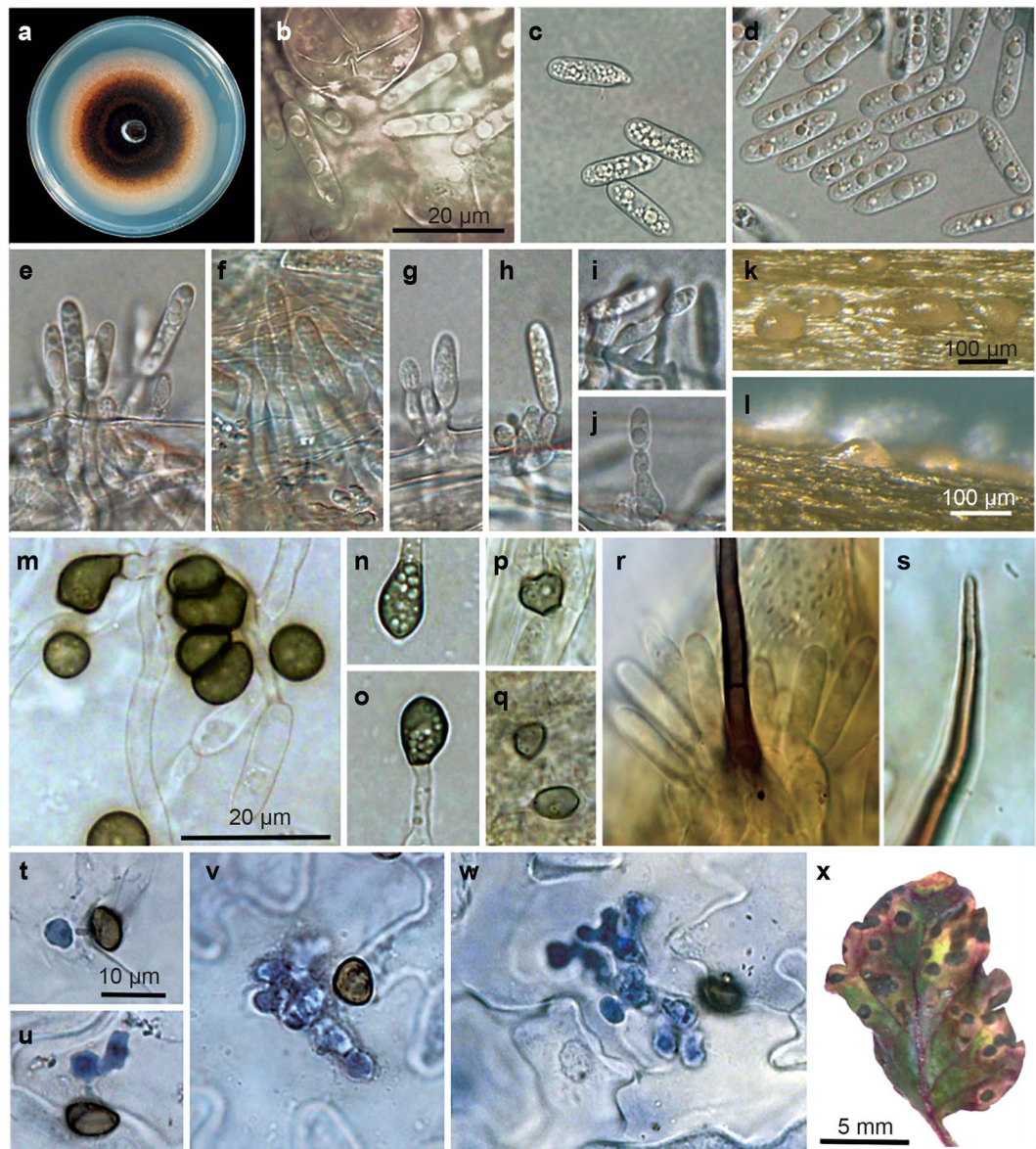


Figure 2. *Colletotrichum shioi* (ex-holotype strain JCM 31818). (a) 10-day old culture on PDA. (b–d) Conidia from *P. frutescens* leaves, PDA and SNA, respectively. (e–j) Conidiophores on *Anthriscus sylvestris* stems. (k–l) Conidiomata on *A. sylvestris* stems. (m). Appressoria formed on OA. (n–o) Appressoria formed on SNA. (p–q) Appressoria formed on *A. sylvestris* stems. (r–s) Setae formed on *P. frutescens* leaves. (t–u) Trypan blue stained *C. shioi* growing in perilla leaves at 40 hpi. (v–w) Biotrophic hyphae at 60 hpi in infected *P. frutescens* leaves. (x) Symptoms on *P. frutescens* leaves at two weeks after inoculation with *C. shioi*. Scale bar of (b) applies to (e–j); scale bar of (m) applies to (n–s) and scale bar for (t) applies to (u–w).

black, smooth-walled, obovoid to clavate with truncate base, $5.5\text{--}10\ \mu\text{m} \times 4\text{--}7.5\ \mu\text{m}$, av. \pm SD: $7.5 \pm 1.0\ \mu\text{m}$ diam $\times 6.0 \pm 1.0\ \mu\text{m}$, L/W ratio: 1.5.

Asexual morph on oatmeal agar (OA). *Vegetative hyphae* $1.5\text{--}4.0\ \mu\text{m}$ diam, hyaline, smooth-walled, septate and branched. av. \pm SD: $2.9 \pm 0.5\ \mu\text{m}$ diam. *Setae* unbranched, medium brown, smooth-walled, 1–2 septate, $29.5\text{--}76.0\ \mu\text{m}$ long, tip rounded. *Conidia* $12\text{--}23\ \mu\text{m} \times 3.5\text{--}5.5\ \mu\text{m}$, av. \pm SD: $16.0 \pm 3.0 \times 4.5 \pm 0.5\ \mu\text{m}$, L/W ratio: 3.5. *Appressoria* single or in groups, greenish olivaceous to olivaceous buff, smooth-walled, globose to subglobose, $5.5\text{--}11\ \mu\text{m} \times 4.5\text{--}10\ \mu\text{m}$, av. \pm SD: $8 \pm 1.0\ \mu\text{m}$ diam $\times 6.5 \pm 1.0\ \mu\text{m}$, L/W ratio: 1.0.

Asexual morph on PDA. *Vegetative hyphae* $2.0\text{--}7.0\ \mu\text{m}$ diam, hyaline, smooth-walled, septate and branched. av. \pm SD: $4.0 \pm 1.0\ \mu\text{m}$ diam. *Conidia* $15.0\text{--}27.0\ \mu\text{m} \times 3.0\text{--}5.0\ \mu\text{m}$, av. \pm SD: $17.0 \pm 2.0\ \mu\text{m} \times 4.0 \pm 0.0\ \mu\text{m}$, L/W ratio: 4.0. *Setae* unbranched, medium brown, smooth-walled, 1–2-septate, $23.5\text{--}78.0\ \mu\text{m}$ in length, tip rounded. *Appressoria* $7.0\text{--}11.0\ \mu\text{m} \times 5.0\text{--}8.0\ \mu\text{m}$, av. \pm SD: $9.0 \pm 1.0\ \mu\text{m} \times 7 \pm 1.0\ \mu\text{m}$, L/W ratio: 1.0.

Asexual morph on autoclaved Anthriscus stem. *Conidiophores* formed on hyaline cells. *Setae* not observed. *Conidiophores* hyaline, smooth-walled, simple. *Conidiogenous cells* hyaline, smooth-walled, straight with round ends or slightly tapered at one end. *Conidia* hyaline, smooth-walled, aseptate, cylindrical,

straight to slightly curved, with both ends straight or with one end tapered $15.5\text{--}28.5\ \mu\text{m} \times 3.0\text{--}5.0\ \mu\text{m}$, av. \pm SD: $19.0 \pm 3.0\ \mu\text{m} \times 4.0 \pm 0.5\ \mu\text{m}$, L/W ratio: 5.0. *Appressoria* single or in loose groups, globose to subglobose, occasionally with an irregular in shape, $3.0\text{--}11.0\ \mu\text{m} \times 3.5\text{--}7.5\ \mu\text{m}$, av. \pm SD: $6.5 \pm 2.0\ \mu\text{m} \times 5.0 \pm 1.0\ \mu\text{m}$, L/W ratio: 1.0.

Asexual morph, infection structures and symptoms in/on leaves of *Perilla frutescens*: Lesions on leaves small, elliptical or irregular, appearing on cotyledons and fully developed leaves, gradually enlarging and becoming dark brown. Acervuli observed forming on lesions under conditions of high humidity. *Setae* unbranched, medium brown, smooth-walled, 1–4-septate, $77.5\text{--}45.5\ \mu\text{m}$ long and $3\text{--}8.5\ \mu\text{m}$ diameter in the base, base ampulliform or cylindrical, tip rounded. *Conidia* hyaline, smooth-walled, aseptate, cylindrical, straight to very slightly curved, with ends round or with one end tapered, $11.0\text{--}21.5\ \mu\text{m} \times 3.0\text{--}5.0\ \mu\text{m}$, av. \pm SD: $17.0 \pm 2.0\ \mu\text{m} \times 4.0 \pm 0.5\ \mu\text{m}$, L/W ratio: 4.5. *Appressoria* single or in loose groups, $4.0\text{--}8.0\ \mu\text{m} \times 3.0\text{--}6.0\ \mu\text{m}$, av. \pm SD: $6.0 \pm 1.0\ \mu\text{m} \times 4.5 \pm 1.5\ \mu\text{m}$, L/W ratio: 1.5. **Intracellular hyphae in detached *Perilla frutescens* leaves:** bulbous, hyphae observed within penetrated cells from 40 hpi.

Culture characteristics. Colonies on SNA at 25°C flat with entire margin, hyaline, filter paper and *Anthriscus* stem partly covered with salmon and dark chestnut acervuli. Whitish aerial mycelia on medium. Reverse same colors. Growth rate $34\text{--}39\ \text{mm}$ in 7 d ($48\text{--}51.5\ \text{mm}$ 10 d). Colonies on OA flat, radially striate with lobate edge, reverse the same, Growth rate $36.5\text{--}37.5\ \text{mm}$ in 7 d ($51.5\text{--}53\ \text{mm}$ 10 d), olivaceous brown to brick, with white aerial hyphae at the edge, colonies of strain MAFF 240106 differ in forming radial crinkles, Colonies on PDA. Flat, olivaceous brown to light orange, hyaline at the edge. Growth rate $46.5\text{--}51\ \text{mm}$ in 7 d ($68\text{--}71\ \text{mm}$ 10 d). *Conidia* in mass saffron.

Materials examined. JAPAN, Osaka, Ibaraki City from lesions of cultivated *Perilla frutescens* var. *crispa* cv. Aka-shiso, collection date 1 August 2006, collected by M. Kawaradani (TNS-F-40462 holotype, culture ex-holotype JCM 31818); JAPAN, Osaka, Ibaraki City from lesions of cultivated *Perilla frutescens* var. *crispa* cv. Aka-shiso, collection date July 2006, collected by M. Kawaradani (MAFF 240106). MAFF 240106 was characterised as being able to infect green and red shiso; as well as egoma varieties of *Perilla frutescens* from Japan, with strongest symptoms on red shiso¹⁰.

Notes: *Colletotrichum shioi* is only known from *P. frutescens* plants in Japan. It belongs to the *C. destructivum* species complex and can be identified by its ITS, *ACT*, *CHS-1*, *GAPDH*, *TUB2* sequences. Fukui (1925) reported a new anthracnose disease of *P. frutescens* in Japan caused by *C. yoshinaoi*⁸. Kim *et al.* (2001) regarded the name *C. yoshinaoi* as invalid because both a Latin diagnosis and the indication of a type is lacking⁹. However, a Latin diagnosis was only required between 1 January 1935 and 31 December 2011, and an indication of a type is only mandatory from 1 January 1958 (Art. 37.1)³⁰; *C. yoshinaoi* is therefore validly described. *Conidia* of *C. yoshinaoi* were described as being oval with round ends and sometimes slightly curved, measuring $15\text{--}17 \times 4\text{--}5\ \mu\text{m}$ with an L/W ratio = 4, which is overlapping with *C. shioi*. However, setae of *C. yoshinaoi* measure $40\text{--}50 \times 3\ \mu\text{m}$ and are sometimes slightly curved and appressoria are round (corresponding to L/W ratio of 1) and about $6\ \mu\text{m}$ diam, while setae of *C. shioi* are larger, measuring $45.5\text{--}77.5 \times 3\text{--}8.5\ \mu\text{m}$ and are straight and appressoria of *C. shioi* on the host plant measure $4\text{--}8\ \mu\text{m} \times 3\text{--}6\ \mu\text{m}$ with a L/W ratio of 1.5. Moreover, *C. yoshinaoi* infects stems causing early defoliation and inhibits fruiting and was never observed on leaves⁸, whereas *C. shioi* infects cotyledons and fully developed leaves. Therefore, Kawaradani (2008) did not regard strain MAFF 240106 (included in this study as *C. shioi*) as *C. yoshinaoi* but identified it as *C. destructivum*. Consequently, we describe the species in the *C. destructivum* complex infecting perilla leaves as a new species, instead of epitypifying *C. yoshinaoi*.

Another *Colletotrichum* species, *C. perillae*, causes a similar disease as *C. yoshinaoi* on stems and pedicels of *P. ocyroides* in the Primorskaya and Ussurskaya Oblast, an area in Russia close to Japan. This species forms acervuli with straight, cylindrical conidia with rounded ends, measuring $18\text{--}22 \times 4.5\text{--}6\ \mu\text{m}$ and straight to flexuous, aseptate, olivaceous setae becoming paler towards the tip, measuring $43\text{--}48 \times 4\text{--}5\ \mu\text{m}$ ³¹. Apart from the fact that the disease caused is different, the conidia of *C. perillae* are larger than those of both *C. shioi* (and *C. yoshinaoi*) and setae are shorter and possibly also differ in septation from those of *C. shioi*. This species was also described without a Latin diagnosis, but before 1 January 1935; the name is therefore invalid (Art. 39.1)³².

Pathogenicity tests. Three-week-old intact *P. frutescens* plants spray-inoculated with *Colletotrichum shioi* JCM 31818 displayed typical symptoms of anthracnose lesions two weeks after inoculation while mock inoculated plants showed no symptoms (Fig. 3a). Symptoms were similar to symptoms of perilla anthracnose observed in nurseries of cultivated Aka-shiso *P. frutescens* plants previously reported by Kawaradani *et al.*¹⁰. Infected plants had smaller leaves than mock inoculated plants (Fig. 3a). These differences were reproduced in three independent experiments. The same fungus was consistently re-isolated from lesions of inoculated plants.

As *C. shioi* is closely related to the *Arabidopsis thaliana*-infecting species *C. higginsianum*, we tested if it can infect *A. thaliana*. *C. shioi* did not form lesions on *A. thaliana* ecotypes Bay-0 and Ws-0 but could form lesions on Ler-0, although to a lesser extent than *C. higginsianum* (Fig. 3b). The distributions of lesion areas were found to be significantly different between *C. shioi* and *C. higginsianum* with P-values < 0.01 according to Mann-Whitney *U* tests in all three ecotypes.

Genome sequence analysis. The genome size of *C. shioi* was estimated, according to k-mer analysis, to be 58.6 Mb and sequenced to $603 \times$ coverage. A total of 36,350 contigs were assembled from the 100 bp paired-end libraries with an N50 value of 7,997. These were then assembled into 20,745 scaffolds with N50 of 9,321 bp (Table 1). According to BUSCO analysis, 98.3% of 3,725 sordariomycete conserved proteins could be identified as complete sequences within the assembly, with an additional 0.8% found to be fragmented, indicating coverage of most of the gene coding space (Table 1). A total of 11,848 genes were predicted in the *C. shioi* genome. The number of genes encoded is consistent with the gene numbers predicted in other *Colletotrichum* species (Fig. 4a), whose numbers range from 16,287 genes (*C. gloeosporioides*) to 10,419 (*C. chlorophyti*).

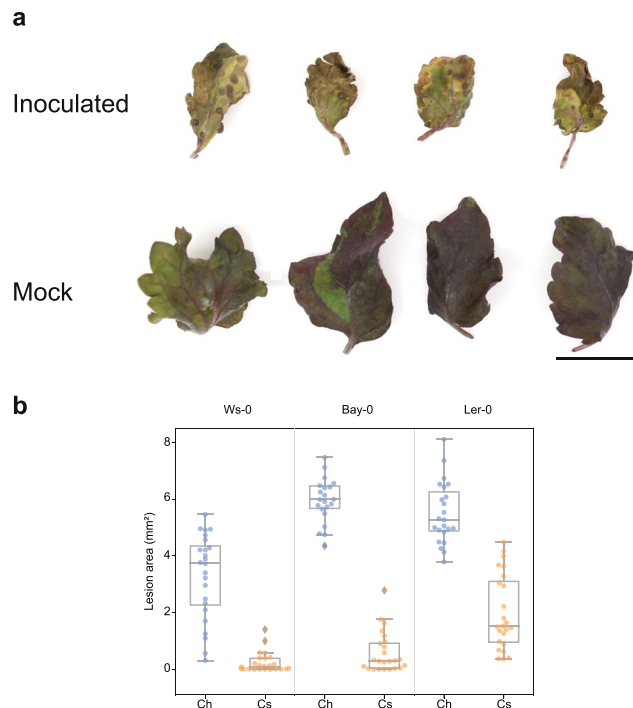


Figure 3. Pathogenicity of *Colletotrichum shisoii*. **(a)** *Perilla frutescens* var. *crispa* Aka-shiso leaves inoculated with *C. shisoii* (top) compared to mock inoculated leaves (bottom) at two weeks post inoculation. Scale bar = 1 cm. The same results were observed in three independent experiments. **(b)** Boxplots comparing area sizes of lesions formed on leaves of *Arabidopsis thaliana* ecotypes after inoculations with *C. shisoii* (Cs) and *C. higginsianum* (Ch). The distributions of lesion areas were found to be significantly different between *A. thaliana* leaves inoculated with *C. shisoii* compared to leaves inoculated with *C. higginsianum* in all three *A. thaliana* ecotypes according to Mann-Whitney *U* tests (*P*-values < 0.01). Significant differences were detected in two independent experiments.

Assembly	
Coverage	603×
Number of scaffolds (>=0 bp)	20,745
Number of scaffolds (>=1,000 bp)	9,321
Largest scaffold	153,238 bp
Total length	69.7 Mb
Total length of scaffolds (>=1,000 bp)	63.87 Mb
GC (%)	47.14
N50	15,836 bp
N75	4,863 bp
L50	1,051
L75	2,981
Number of N's per 100 kbp	943.26
Complete	98.3%
Fragmented	0.8%
Duplicated	0.1%

Table 1. Genome assembly statistics. Complete: Percentage of BUSCO sordariomycete_odb9 conserved single copy gene set that was present as a complete coding sequence in the assembly; Fragmented: percentage of BUSCO sordariomycete_odb9 conserved single copy gene set that were identified as partial coding sequences in the assembly; Duplicated: percentage of BUSCO sordariomycete_odb9 single copy gene set that was found to be present with more than one copy in the assembly.

Conservation of proteins amongst *Colletotrichum* species. The conservation of genes between *C. shisoii* and four closely related and sequenced *Colletotrichum* species was assessed (Figs 4 and 5). The four that were chosen for comparisons were *C. higginsianum*, which has a chromosome-level assembly, publicly available annotations and also belongs to the *C. destructivum* species complex, *C. incanum* and *C. tofieldiae*, from the *C. spaethianum* species

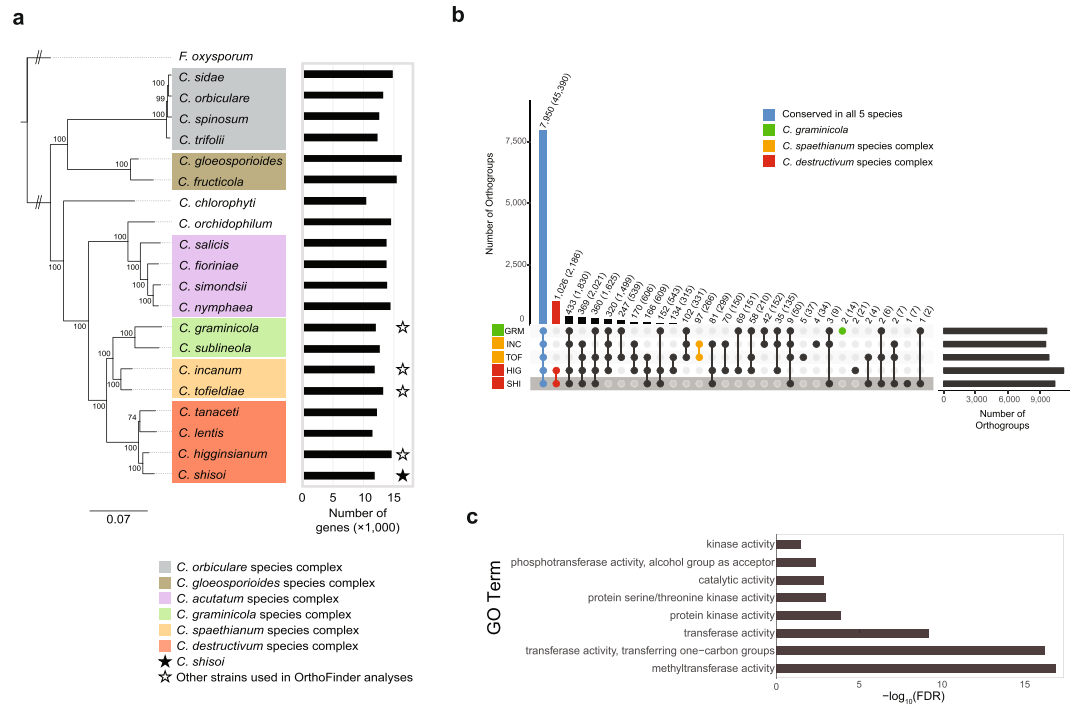


Figure 4. Conservation of all predicted proteins in selected *Colletotrichum* species. **(a)** Maximum likelihood phylogeny based on 128,141 characters from 254 eukaryotic conserved single copy protein orthologs identified by BUSCO analysis in different members of the *Colletotrichum* genus, showing the relationship between *C. shisoii* and other *Colletotrichum* species complex members. Predicted gene numbers are indicated by the bars on the right. *Fusarium oxysporum* was used as an outgroup. Numbers at the nodes represent percent support values out of 500 bootstrap replicates. **(b)** UpSet plot showing conservation of protein orthogroups in *C. graminicola* (GRM), *C. incanum* (INC), *C. tofieldiae* (TOF), *C. higginsianum* (HIG) and *C. shisoii* (SHI). Box plots show the distributions of orthogroup protein length means. Bars represent number of orthogroups showing specific conservation patterns with numbers also given above each bar. Numbers in brackets show number of proteins that belong to orthogroups according to their respective conservation pattern. **(c)** Bar chart showing GO terms significantly enriched among *C. shisoii* genes present only in the *C. destructivum* species complex (*C. shisoii* and *C. higginsianum*). FDR: False discovery rate of hypergeometric test after Benjamini-Hochberg multiple testing correction.

complex, and *C. graminicola* from the *C. graminicola* species complex. Members of the *C. spaethianum* and the *C. graminicola* species complexes were selected since they are closely related to the *C. destructivum* species complex (Fig. 4a). Among the five species assessed, *C. higginsianum* encodes the greatest number of predicted genes (14,651 genes). The number of predicted proteins for the other species were closer to the number of genes in *C. shisoii* with 11,436, 12,501 and 12,006 proteins predicted in *C. incanum*, *C. tofieldiae* and *C. graminicola*, respectively (Fig. 4a). A total of 11,914 orthogroups with two or more proteins were identified (Fig. 4b). Of these, 7,950 groups (74.0% proteins from *C. shisoii*, 63.0% from *C. higginsianum*, 78.3% from *C. incanum*, 74.2% from *C. tofieldiae* and 73.7% from *C. graminicola*) were conserved in all five species (Supplementary Tables S3 and S4). From this analysis, all *C. shisoii* genes could be classified into an orthogroup with a related sequence in one of the four other species or in the same genome. Only one orthogroup was predicted to be *C. shisoii*-specific. This orthogroup consisted of seven proteins annotated as MFS transporter proteins. Similarly, all *C. higginsianum* proteins were classified into an orthogroup with only two orthogroups found to be specific to *C. higginsianum*, one consisting of 13 ABC transporter genes and the second, consisting of 8 secondary metabolite regulator laeA protein-encoding genes. As expected from their close evolutionary relationship, *C. shisoii* and *C. higginsianum* were found to share an additional 2,585 orthogroups consisting of 23.4% proteins from *C. shisoii* and 20.8% proteins from *C. higginsianum*, including 1,026 orthogroups (8.7% proteins from *C. shisoii* and 7.4% proteins from *C. higginsianum*), which are present only in these two members of the *C. destructivum* clade. Proteins of *C. shisoii* from the *C. destructivum*-specific orthogroups were significantly enriched for Gene Ontology (GO) terms involved in methyltransferase and protein kinase activity (FDR < 0.05) (Fig. 4c). In contrast, *C. tofieldiae* and *C. incanum*, which both belong to the *Colletotrichum spaethianum* clade, share only 97 orthogroups, consisting of 1.3% proteins from *C. incanum* and 0.9% proteins from *C. tofieldiae*, which were specific to these two members of the *C. spaethianum* clade.

Conservation of secreted proteins. As secreted proteins are known to be important for infection, their conservation between the five species was also assessed (Fig. 5a, Supplementary Table S5). A total of 1,360 secreted protein orthogroups were identified. Of these, 540 orthogroups (39.7%) were found to be conserved in all

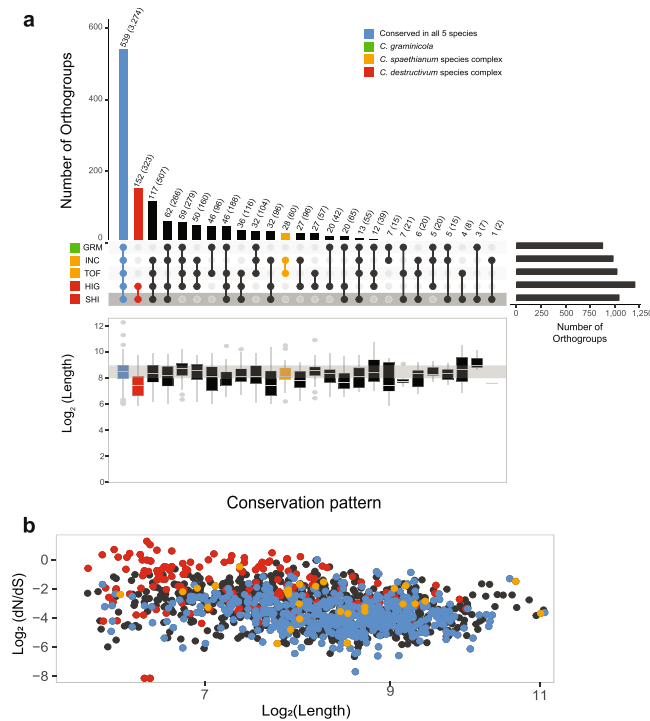


Figure 5. Conservation of secreted proteins in *Colletotrichum* species. **(a)** UpSet plot showing conservation of secreted protein orthogroups in *C. graminicola* (GRM), *C. incanum* (INC), *C. tofieldiae* (TOF), *C. higginsianum* (HIG) and *C. shiso* (SHI). Box plots show the distributions of orthogroup protein length means. Bars represent number of orthogroups showing specific conservation patterns with numbers also given above each bar. Numbers in brackets show number of proteins that belong to orthogroups according to their respective conservation pattern. **(b)** Scatterplot showing orthogroup protein length means relative to the mean rate of pairwise non-synonymous mutations (dN)/synonymous mutations (dS) in each orthogroup.

five species. These included 48.4% proteins out of the 1,274 predicted secreted proteins from *C. shiso* and 40.6% proteins of the 1,644 predicted secreted proteins from *C. higginsianum* (Supplementary Table S5). A further 154 secreted protein orthogroups, were identified as specific to the two *C. destructivum* clade members. In contrast, 28 secreted protein orthogroups were identified as being *C. spaethianum* clade-specific. No GO terms were found to be significantly associated with *C. shiso* secreted proteins that were in *C. destructivum* clade-specific orthogroups. Since effector proteins tend to be small, secreted proteins under positive selection, we plotted the average lengths of orthogroups consisting of secreted proteins (Fig. 5a) to investigate if there was a relationship between conservation pattern and orthogroup protein length. All 2,186 *C. destructivum* clade-specific proteins (Fig. 4b) were found to be shorter than proteins belonging to orthogroups that were conserved in the five *Colletotrichum* species. Further, of particular interest to this study, the *C. destructivum* clade-specific secreted proteins (Fig. 5a) were found to be also under higher rates of positive selection, with higher rates of non-synonymous to synonymous mutations (dN/dS), compared to secreted proteins that were conserved in all five tested *Colletotrichum* species (Fig. 5b). In contrast, this was not observed among *C. spaethianum* clade-specific secreted proteins (Fig. 5b). No species-specific groups were identified amongst the secreted protein orthogroups in all five species, indicating that species-specific sequences did not belong to multi-gene families. A total of 846 secreted proteins, consisting of 216 proteins from *C. graminicola*, 128 proteins from *C. incanum*, 112 proteins from *C. tofieldiae*, 225 proteins from *C. higginsianum* and 135 proteins from *C. shiso* were not assigned to any orthogroup, and were found to be species-specific (Supplementary Table S5).

Discussion

Colletotrichum species can infect a wide range of plants. In this study, we identified a new species in the *C. destructivum* clade that infects the commercially important oil crop *P. frutescens*. Previously, species of the *C. gloeosporioides* clade, *C. gloeosporioides*, *C. dematium* and *C. coccodes* were reported as pathogens of *P. frutescens* in Korea by morphological examination of isolates⁹, and *C. destructivum* was identified as a pathogen of *P. frutescens* in Japan based on ITS sequences¹⁰. In this study, a multi-locus phylogenetic analysis showed that strains from *P. frutescens* previously identified as *C. destructivum*, are genetically distinct from other known species of the *C. destructivum* species complex, and were thus described as a new species, *C. shiso*. Since well-studied species of the *C. destructivum* species complex have previously been confused with *C. coccodes*, *C. gloeosporioides* and *Glomerella cingulata*¹², the strains from *P. frutescens* in Korea⁹ identified as these species could also represent *C. shiso* and their re-examination may be warranted.

In order to characterise *C. shiso* and to allow comparisons to other members of the *Colletotrichum* genus at the molecular level, the genome of *C. shiso* was sequenced and assembled. The size of the *C. shiso* assembly is

69.7 Mb, which is larger than those of other sequenced members in the *C. destructivum* species complex, including *C. higginsianum*³³ (50.72 Mb), *C. lentis*²¹ (56.1 Mb) and, most recently, *C. tanacetii*²⁶ (57.9 Mb). These sizes are closer to the genome size of *C. shisoii* estimated by k-mer analysis (58.6 Mb). A genome assembly size may deviate from k-mer estimates due to high levels of repeats or heterozygosity³⁴. Members of the *C. destructivum* clade are known to be haploid pathogens that propagate asexually and thus heterozygosity is unlikely¹². Further, BUSCO analysis of conserved genes reveals that only 0.1% of the conserved coding sequences present in the genome are duplicated, indicating that the genome is not heterozygous and that the gene coding regions at least are likely not to be duplicated. This is consistent with the predicted number of genes (11,848) in *C. shisoii*, which is within the range of other sequenced members of the *C. destructivum* species complex including the recently published genomes of *C. lentis*²¹ (11,436), *C. tanacetii*²⁶ (12,172) and another isolate of *C. higginsianum*²⁵ (MAFF 305635-RFP), which has 12,915 protein coding genes. It is noted that the genome assembly of *C. shisoii* generated in this study was sequenced using only short reads and is highly fragmented, with more than half of the scaffolds (11,424 scaffolds totalling 5.73 Mb) being less or equal to 1 kb in length. An earlier version of the genome of its close relative, *C. higginsianum*, which was assembled using a combination of Illumina GAI, Roche 454 and Sanger Fosmid reads, also suffered from fragmentation (10,269 scaffolds)¹⁸, possibly due to the abundance of transposable element-rich genomic regions. Since then, a chromosome level assembly for *C. higginsianum* has been generated using a combination of PacBio long reads and optical mapping data³³. The *C. shisoii* genome may similarly be better resolved by adopting a similar sequencing strategy.

Comparison of the genomes of *C. shisoii* and *C. higginsianum* showed that the majority of *C. shisoii* genes (89%) have orthologous sequences in *C. higginsianum*. The latter species was previously reported¹¹ to infect *L. amplexicaule*, a plant belonging to the Lamiaceae family. Both pathogens appear to adopt similar infection strategies. *C. shisoii* was observed to form bulbous intracellular hyphae within infected epidermal cells in early infection of *P. frutescens* leaves. These hyphae are morphologically similar to the primary, biotrophic hyphae formed by *C. higginsianum* infecting *A. thaliana* leaves³⁵. Further, as in the case of *C. higginsianum*-infected *A. thaliana* plants, necrotic lesions formed later in infection. Taken together, these observations suggest that *C. shisoii* also adopts a hemibiotrophic infection strategy, as do other members of the *C. destructivum* complex³⁶.

Previously, genus-wide analyses including members from the *C. orbiculare*, *C. acutatum*, *C. graminicola*, *C. gloeosporioides* and *C. destructivum* species complexes of *Colletotrichum* revealed that *C. higginsianum* has the highest number of lineage-specific genes amongst the genomes tested². However, at the time, *C. higginsianum* was the only *C. destructivum* species complex member whose genome had been sequenced. Further, the assembly was highly fragmented, leading to the possibility that gene numbers were inflated. Since then, a chromosome level assembly for *C. higginsianum* has been published³³ and our results indicate that *C. higginsianum* and *C. shisoii* do indeed have a high number of orthogroups specific to these two *C. destructivum* species complex members. The *C. destructivum* clade-specific genes were significantly enriched in kinases, indicating the presence of *C. destructivum* clade-specific signalling pathways. It is interesting to note that *C. spaethianum* clade members, despite their close phylogenetic relationship to the *C. destructivum* species complex, do not exhibit the same expansion in lineage-specific genes.

Secreted proteins that were specific to the two members of the *C. destructivum* species complex analysed in this study were also found to be subject to higher rates of diversifying selection than secreted proteins that were identified as *C. spaethianum* clade-specific. *C. shisoii* is less virulent on *A. thaliana* compared to *C. higginsianum*, forming significantly smaller lesions or no lesions at all on the accessions tested. Interestingly, the more distantly related strains from the *C. spaethianum* species complex, *C. incanum* and *C. tofieldiae*, have both been shown to infect *A. thaliana* plants^{17,37}, indicating that components required for successful invasion of *A. thaliana* were possibly present in the ancestor of the *C. destructivum* and *C. spaethianum* clades. Given that small, secreted proteins known as effectors, which are important for manipulation of hosts, are often under diversifying selection to avoid recognition by specific host immune components³⁸, the *C. destructivum* clade-specific secreted proteins could be candidate effectors involved in infection of host plants and their diversification may have resulted in the differences in the observed infection outcomes of *C. higginsianum* and *C. shisoii*.

Finally, *Perilla frutescens* produces a range of antimicrobial compounds and has been characterised by transcriptomic and metabolomic analyses³⁹. The examination of the genome of its pathogen, *C. shisoii*, will provide insights into the mechanisms of this pathogen to overcome host defence and thus enable the development of better control strategies.

Materials and Methods

Isolates. The strains studied here originate from leaves of *Perilla frutescens* with anthracnose symptoms that had been collected in August 2006 and July 2006 from a perilla seedling bed in Ibaraki city, Osaka, Japan as previously described¹⁰. As described by Kawaradani *et al.* (2016), the seedling bed was located in the shaded part of a southwestern-facing mountain slope. Leaves showing symptoms were surface sterilised with sterile water and incubated on PDA plates containing 100 ppm streptomycin at 25 °C¹⁰. Isolates were isolated by hyphal tipping¹⁰. The holotype of the new species was deposited in the mycological herbarium of the National Museum of Nature and Science (TNS-F-40462), Tsukuba, Japan and the ex-type culture in the Japan Collection of Microorganisms (JCM 31818), Tsukuba, Japan. Isolates were stored as glycerol stocks at –80 °C and revived by incubation on PDA at 24 °C in the dark prior to experiments.

Phylogenetic analyses. Sequences of *ACT*, *CHS-1*, *GAPDH*, ITS and *TUB2* were identified from the JCM 31818 assembly by BLASTn searches with sequences from *C. higginsianum* IMI 349063 and selecting sequence regions with the lowest E-values. The ITS sequence was used to query the NCBI non-redundant nucleotide database using default BLASTn settings to identify closely related fungal species.

Sequences for MAFF 240106 ITS, *GAPDH*, *CHS-1*, *ACT* and *TUB2* were amplified using the primer pairs ITS-1F⁴⁰ + ITS-4⁴¹, GDF1 + GDR1⁴², CHS-354R + CHS-79F, ACT-512F + ACT-783R⁴³ and T1⁴⁴ + Bt-2b⁴⁵. PCR was carried out in a thermocycler using 2 × PCR Taq polymerase mix (Promega) at 95 °C for 3 min, followed by 35 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min and a final extension step at 72 °C for 5 min. Phylogenetic trees were calculated as previously described⁴⁶. Sequences of each locus (Supplementary Table S2) were aligned in MAFFT v7.215⁴⁷ using the auto setting and trimmed using trimAl v1.2rev59⁴⁸ using the automated1 setting. Maximum parsimony analyses were carried out with PAUP* (Phylogenetic Analysis Using Parsimony) version 4.0a (build 165)⁴⁹ using a heuristic search of 100 random sequence additions with tree bisection and reconstruction (TBR) as the branch-swapping algorithm. All sites were treated as unordered and equally weighted with gaps treated as missing data. A total of 1,000 bootstrap replicates using the same settings were carried out to determine support for the trees. To determine the best model for analyses, jModelTest2⁵⁰ was run on alignments with BIC criterion. For single locus trees, maximum likelihood trees were calculated using RAXML-ng using the specified jModelTest2 model with 1,000 bootstrap replicates. For Bayesian inference phylogenies based on single loci were calculated twice using MrBayes (v3.2.1) with 5 × 10⁶ generations, sampling every 1,000 generations. Under these settings, the average standard deviation of split frequencies was found to be 0.006037 for *ACT*, 0.007120 for *CHS-1*, 0.005246 for *GAPDH*, 0.007322 for ITS and 0.006178 for *TUB2*. For multi-locus sequence analysis, the trimmed alignments were concatenated and then a maximum likelihood tree was calculated with RAXML-ng⁵¹ using the specified jModelTest2 model for each partition with 1,000 bootstrap replicates. Bayesian inference of the concatenated alignment was calculated twice using MrBayes (v3.2.1)⁵² with 5 × 10⁶ generations, sampling every 1,000 generations. Under these settings, the standard deviation of split frequencies was 0.004981 and performance scale reduction factors were close to 1.000 for all tested parameters. The first 25% generations were discarded as burnin. Phylogenetic trees were generated for individual loci using the calculated jModelTest2 models as well as for the concatenated alignment using *C. boninense* as an outgroup.

For the genus-wide maximum likelihood tree of selected *Colletotrichum* species, BUSCO was run on the genome assemblies of *C. trifolii* (RYZW01000000), *C. sidae* (QAPF01000000), *C. orbiculare* (AMCV02000000) and *C. spinosum* (QAPG01000000)²⁴, *C. fruticola* (ANPB00000000.1)¹⁹, *C. gloeosporioides* (QFRH00000000)²⁷, *C. chlorophyti* (MPGH01000000)⁵³, *C. orchidophilum* (MJBS00000000.1)⁵⁴, *C. salicis* (JFFI00000000.1), *C. fioriniae* (JARH00000000.1), *C. simondsii* (JFBX00000000.1), *C. nymphaea* (JEMN00000000.1)²⁸, *C. sublineola* (JMSE00000000.1)²⁹, *C. graminicola* (ACOD00000000.1)¹⁸, *C. incanum* (JTLR01000000)¹⁷, *C. tofieldiae* (LFIV01000000)³⁷, *C. lentis* (NWB01000000)²⁶, *C. tanacetii* (PJEX00000000)²⁶, *C. higginsianum* (LTAN01000000)³³ and *Fusarium oxysporum* (GCF_000149955.1)⁵⁵ to identify highly conserved, single copy eukaryote genes (eukaryote_odb9). Sequences of orthologs from 254 single copy genes that were identified as non-duplicated in all the tested genomes were aligned using MAFFT and trimmed using trimAl as described above. Modeltest-ng v0.1.5 (<https://github.com/ddarriba/modeltest>) was run to determine the best model for amino acid substitutions under BIC criterion (Supplementary Table S6). Sequences for all 254 single copy genes were concatenated and RAXML-ng was used to estimate the maximum likelihood phylogeny using the modeltest-ng specified best model for each partition with 500 bootstrap replicates.

The generated trees were visualised in FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Morphological characterisation. Culture morphology was assessed on PDA (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan), oatmeal agar (OA, Difco) and synthetic nutrient-poor agar (SNA⁵⁶) plates. Autoclaved filter paper and double-autoclaved stems of *Anthriscus sylvestris* were placed on the surface of SNA plates. Plates were inoculated with 8-mm-diameter mycelia-grown agar plugs from the edge of actively growing cultures. Cultures were incubated at 24 °C under near UV light with 12 h photoperiod for 10 d. Colony colours were rated after 7 d using the Rayner colour chart⁵⁷. Growth rates were estimated after 7 and after 10 d. Structures were observed after 10 d. For formation of appressoria, conidial suspensions were incubated on SNA or OA coated glass slides with a plastic cover slip in 100% humidity at 24 °C in the dark for 24 h. Fungal structures were observed using a Leica M165FC dissecting microscope (DM) and an Olympus BX51 microscope with differential interference contrast (DIC) optics. Samples for DIC were mounted directly in lactic acid. Measurements were made for at least 30 structures each.

Pathogenicity tests and observation of the infection process. Perilla plants (Aka Chirimen Shiso, Takii & Co., Ltd.) were cultivated on sterile soil in 12 h white light/12 h dark at 24 °C for three weeks before infections. For observations of trypan blue-stained invasive hyphae, approximately 100 µL of a conidia suspension of strain JCM 31818 (1 × 10⁵ conidia/mL) was dropped onto the abaxial side of each detached perilla leaf. Then, pieces of 1.5 cm × 1 cm nylon mesh with 100 µm pores were placed on to conidial droplets to ensure even distribution of the conidial suspension on the surface of the leaf. Inoculated leaves were incubated in petri dishes for 40 and 60 hpi (hours post inoculation) at 24 °C under white light with 12 h light/12 h dark photoperiod and 100% humidity in a plastic dish with autoclaved filter paper moistened by sterilised water. For trypan blue staining, infected perilla leaves were boiled in 1 ml of alcoholic lactophenol (ethanol: phenol: glycerol: lactic acid: water (4:1:1:1:1, v/v/v/v/v)) containing 0.1 mg/ml trypan blue for 10 min at 95 °C and left overnight at room temperature. Boiled leaves were destained with chloral hydrate solution (500 g chloral hydrate + 200 ml water) overnight. Stained leaves were mounted in 60% glycerol solution and observed with DIC under a microscope. Further, for observation of setae, conidia and appressoria in/on *P. frutescens* leaves, intact three-week-old plants grown on sterile soil were sprayed with 5 × 10⁵ conidia/mL conidial solution. Leaves were detached just prior to observation and mounted directly in lactic acid before DIC imaging and measurements. For pathogenicity tests, intact three-week-old plants grown on sterile soil were sprayed with 5 × 10⁵ conidia/mL conidial solution. As negative controls, mock-treated plants were sprayed with water. Plants were maintained at 24 °C in 12 h white light/12 h

dark conditions in 100% relative humidity and assessed for anthracnose lesions at two weeks post-inoculation. At least four plants were tested for each treatment in three independent experiments.

For pathogenicity tests of *A. thaliana*, the first three fully expanded leaves from four-week-old plants grown under short day conditions (8 h light/16 h dark) at 21 °C were inoculated with 5×10^5 conidia/ml conidial suspensions of strain JCM 31818 from perilla or *C. higginsianum* strain MAFF 305635 from *Brassica rapa*. Each leaf was inoculated with one 5 µl droplet of prepared conidial suspension. Infected plants were maintained at 100% humidity under the same light and growth conditions as perilla plants. Images of infected leaves were captured 6 d after inoculation using a Canon EOS-M camera and lesion areas were determined using ImageJ⁵⁸. Experiments were repeated twice using eight plants per ecotype per experiment. Lesion area size distributions were tested for significant differences using Mann-Witney *U* tests. For both *P. frutescens* and *A. thaliana* pathogenicity tests, leaves were only detached from intact plants just prior to imaging.

DNA extraction and genome sequencing. For sequencing and assembly of the JCM 31818 genome, genomic DNA was extracted and sequenced as previously described¹⁷. In brief, PD broth (BD Biosciences, Franklin Lakes, NJ, USA) was inoculated with hyphae from a growing colony. After incubating for 3 d at 24 °C under dark conditions, the mycelium was harvested and ground in liquid nitrogen and then the genomic DNA was extracted using CTAB buffer and 100/G genomic tips (QIAGEN, Hilden, Germany) as previously described⁵³. DNA from MAFF 240106 was extracted using the QIAGEN genomic DNeasy kit according to the manufacturer's instructions. Two differently sized insert libraries, 150 bp and 500 bp, were prepared using the Illumina TruSeq PCR-free DNA sample prep kit (Illumina) and sequenced to generate 100 bp paired-end reads with an Illumina HiSeq 2000 sequencing system (RIKEN Omics Science Center, Yokohama, Japan).

Genome assembly and annotation. Low quality reads were trimmed using TrimGalore wrapper with cutadapt (v1.2.1) and fastqc (v0.11.7). Sequences were assembled using Megahit⁵⁹ followed by scaffolding using the SSPACE-Standard-3.0 scaffolder (Baseclear). The assembly was assessed using quast v4.5⁶⁰ and BUSCO v3.0 using the sordariomyceta_odb9 dataset⁶¹. The size of the genome was estimated by kmer analysis using jellyfish v1.14⁶² as previously described⁶³. Genes were predicted with the MAKER v 2 pipeline⁶⁴ after optimizing Augustus v 3.3⁶⁵ gene model parameters by running the BUSCO pipeline⁶¹ with the --long option to identify *C. shioi* homologs of 3,659 sordariomycete conserved proteins; training Genemark-ES (v3.51)⁶⁶ on the *C. shioi* genome using the option to run the program using the branch point model for fungal gene predictions; and including proteins from *C. higginsianum*⁶⁷ as additional evidence for gene model support.

Orthogroup identification. All predicted proteins from *C. graminicola*¹⁸, *C. higginsianum*⁶⁷, *C. incanum*¹⁷ and *C. tofieldiae*^{37,68} were analysed using OrthoFinder v 2.2.6⁶⁹ with the default settings. For identification of secreted protein orthogroups, DeepLoc v1.0 was utilised to predict the localisations of proteins from each species⁷⁰. Then, OrthoFinder was run on the predicted secreted proteins to identify orthogroups within the secreted protein sequences. For analysis of dN/dS values of secreted protein-encoding gene sequences, the genes of secreted proteins grouped together by OrthoFinder were aligned using PRANK⁷¹ to produce codon alignments using default settings. Codon alignments were then analysed using the yn00 model⁷² implemented in the PAML suite of programs⁷³. Conservation plots were drawn using the UpsetR package⁷⁴ in R⁷⁵. GO terms were assigned to *C. shioi* sequences using Trinotate⁷⁶ and enrichment of GO terms in selected groups was tested using the hypergeometric test in the GStats package⁷⁷ and applying the Benjamini-Hochberg multiple test correction on P-values using R.

Data Availability

C. shioi sequences used for phylogenetic analyses are deposited under GenBank accession numbers MH660928–MH660937. The whole genome shotgun sequences were deposited in DDBJ/ENA/GenBank under BioProject PRJNA431477 with accession number PUHP00000000. In this study, version PUHP01000000 is described.

References

- Igarashi, M. & Miyazaki, Y. A review on bioactivities of perilla: progress in research on the functions of perilla as medicine and food. *Evid.-Based Complement. Altern. Med. ECAM* **2013** (2013).
- Brenner, D. M. Perilla: Botany, uses and genetic resources. in *New crops* (eds Janick, J. & Simon, J. E.) 322–328 (Wiley, New York, 1993).
- Ghimire, B. K., Yoo, J. H., Yu, C. Y. & Chung, I.-M. GC–MS analysis of volatile compounds of *Perilla frutescens* Britton var. Japonica accessions: Morphological and seasonal variability. *Asian Pac. J. Trop. Med.* **10**, 643–651 (2017).
- Meng, L., Lozano, Y., Gaydou, E. & Li, B. Antioxidant activities of polyphenols extracted from *Perilla frutescens* Varieties. *Molecules* **14**, 133–140 (2008).
- Freeman, M. B., Choi, S. & Hinkle, A. Edible oils market brief-Palm and soy oils lead the way. (2018).
- MAFFstat. 平成24年地域特産野菜生産状況調査 (2014 survey of regional special vegetable production). Available at: <http://www.estat.go.jp/SG1/estat/List.do?lid=000001155203>. (Accessed: 25th July 2018) (2016).
- Weaver, W. W. The Landreth Seed Company: Testing ground for a new American cuisine. *Gastronomica* **11**, 24–28 (2011).
- Fukui, T. Two kinds of perilla diseases. *Byo-Chugai Zasshi* **12**, 572–573 (1925).
- Kim, W.-G., Lee, B.-D., Cho, W.-D. & Shin, D.-B. Anthracnose of perilla caused by *Colletotrichum* spp. and *Glomerella cingulata*. *Plant Pathol. J.* **17**, 236–241 (2001).
- Kawaradani, M. et al. *Colletotrichum destructivum* によるシソ炭疽病 (新称). *Jpn J Phytopathol* **74**, 335–339 (2008).
- Orihara, N. et al. Occurrence of anthracnose on *Lamium amplexicaule* L. and *Portulaca oleracea* L. Caused by *Colletotrichum higginsianum* and the pathogenicity on *Brassica rapa* var. *perviridis* (Komatsuna). 関東東山病害虫研究会報 **2012**, 47–50 (2012).
- Damm, U., O'Connell, R. J., Groenewald, J. Z. & Crous, P. W. The *Colletotrichum destructivum* species complex – hemibiotrophic pathogens of forage and field crops. *Stud. Mycol.* **79**, 49–84 (2014).
- Cannon, P. F., Buddie, A. G. & Bridge, P. D. The typification of *Colletotrichum gloeosporioides*. *Mycotaxon* **104**, 189–204 (2008).
- Damm, U., Woudenberg, J. H. C., Cannon, P. F. & Crous, P. W. *Colletotrichum* species with curved conidia from herbaceous hosts. *Fungal Divers.* **39**, 45 (2009).

15. Weir, B. S., Johnston, P. R. & Damm, U. The *Colletotrichum gloeosporioides* species complex. *Stud. Mycol.* **73**, 115–180 (2012).
16. Liu, F., Damm, U., Cai, L. & Crous, P. W. Species of the *Colletotrichum gloeosporioides* complex associated with anthracnose diseases of *Proteaceae*. *Fungal Divers.* **61**, 89–105 (2013).
17. Gan, P. *et al.* Genus-Wide Comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol. Evol.* **8**, 1467–1481 (2016).
18. O'Connell, R. J. *et al.* Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* **44**, 1060–1065 (2012).
19. Gan, P. *et al.* Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* **197**, 1236–1249 (2013).
20. Baroncelli, R., Sreenivasaprasad, S., Sukno, S. A., Thon, M. R. & Holub, E. Draft genome sequence of *Colletotrichum acutatum* Sensu Lato (*Colletotrichum fioriniae*). *Genome Announc.* **2**, e00112–14 (2014).
21. Bhadauria, V. *et al.* Genetic map-guided genome assembly reveals a virulence-governing minichromosome in the lentil anthracnose pathogen *Colletotrichum lentis*. *New Phytol.* **221**, 431–445 (2019).
22. Alkan, N. *et al.* Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. *Mol. Plant-Microbe Interact. MPMI* **26**, 1345–1358 (2013).
23. de Queiroz, C. B., Correia, H. L. N., Menicucci, R. P., Vidigal, P. M. P. & de Queiroz, M. V. Draft genome sequences of two isolates of *Colletotrichum lindemuthianum*, the causal agent of anthracnose in common beans. *Genome Announc.* **5** (2017).
24. Gan, P. *et al.* Genome sequence resources for four phytopathogenic fungi from the *Colletotrichum orbiculare* species complex. *Mol. Plant. Microbe Interact.*, <https://doi.org/10.1094/MPMI-12-18-0352-A> (2019).
25. Tsushima, A. *et al.* Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. *Genome Biol. Evol.* **11**, 1487–1500 (2019).
26. Lehwala, R. V. *et al.* Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in *Colletotrichum tanacetii*. *PLoS One* **14**, e0212248 (2019).
27. Huang, L. *et al.* A high-quality draft genome sequence of *Colletotrichum gloeosporioides* sensu stricto SMCG1#C, a causal agent of anthracnose on *Cunninghamia lanceolata* in China. *Mol. Plant. Microbe Interact.* **32**, 139–141 (2019).
28. Baroncelli, R. *et al.* Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics* **17**, 555 (2016).
29. Buiaite, E. A. S. *et al.* A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC Genomics* **18**, 67 (2017).
30. McNeill, J. *et al.* *International Code of Botanical Nomenclature (Vienna Code)*. **146** (2006).
31. Abramov. Bolezni sel'skokhozyaistvennykh rastenii na dal'nem vostokey. In 247 (1938).
32. McNeill, J. *et al.* *International Code of Nomenclature for algae, fungi and plants. Regnum Veg.* **154**, (2012).
33. Dallery, J.-F. *et al.* Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. *BMC Genomics* **18** (2017).
34. Binghang, L. *et al.* *Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects* (2013).
35. O'Connell, R. *et al.* A novel Arabidopsis-*Colletotrichum* pathosystem for the molecular dissection of plant-fungal interactions. *Mol. Plant. Microbe Interact.* **17**, 272–282 (2004).
36. Crouch, J. *et al.* The genomics of *Colletotrichum*. in *Genomics of Plant-Associated Fungi: Monocot Pathogens* (eds Dean, R. A., Lichens-Park, A. & Kole, C.) 69–102 (Springer Berlin Heidelberg, 2014).
37. Hacquard, S. *et al.* Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat. Commun.* **7**, 11362 (2016).
38. Stergiopoulos, I. & de Wit, P. J. G. M. Fungal effector proteins. *Annu. Rev. Phytopathol.* **47**, 233–263 (2009).
39. Fukushima, A., Nakamura, M., Suzuki, H., Saito, K. & Yamazaki, M. High-throughput sequencing and *de novo* assembly of red and green forms of the *Perilla frutescens* var. *crispa* transcriptome. *PLoS ONE* **10**, e0129154 (2015).
40. Gardes, M. & Bruns, T. D. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113–118 (1993).
41. White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc. Guide Methods Appl.* **18**, 315–322 (1990).
42. Guerber, J. C., Liu, B., Correll, J. C. & Johnston, P. R. Characterization of diversity in *Colletotrichum acutatum* sensu lato by sequence analysis of two gene introns, mtDNA and intron RFLPs, and mating compatibility. *Mycologia* **95**, 872–895 (2003).
43. Carbone, I. & Kohn, L. A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* **91** (1999).
44. O'Donnell, K. & Cigelnik, E. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* **7**, 103–116 (1997).
45. Glass, N. L. & Donaldson, G. C. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl. Environ. Microbiol.* **61**, 1323–1330 (1995).
46. Gan, P., Nakata, N., Suzuki, T. & Shirasu, K. Markers to differentiate species of anthracnose fungi identify *Colletotrichum fructicola* as the predominant virulent species in strawberry plants in Chiba Prefecture of Japan. *J. Gen. Plant Pathol.* 1–9, <https://doi.org/10.1007/s10327-016-0689-0> (2016).
47. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
48. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
49. L. Swofford, D. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4.0b10. **Version 4.0**, (2002).
50. Darrriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and high-performance computing. *Nat. Methods* **9**, 772 (2012).
51. Kozlov, A. amkozlov/raxml-ng: RAXML-NG v0.6.0 BETA (Version 0.6.0). *Zenodo* (2018).
52. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinforma. Oxf. Engl.* **19**, 1572–1574 (2003).
53. Gan, P. *et al.* Draft genome assembly of *Colletotrichum chlorophyti*, a pathogen of herbaceous plants. *Genome Announc.* **5**, e01733–16 (2017).
54. Baroncelli, R. *et al.* Whole-genome sequence of the orchid anthracnose pathogen *Colletotrichum orchidophilum*. *Mol. Plant. Microbe Interact.* **31**, 979–981 (2018).
55. Ma, L.-J. *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373 (2010).
56. Nirenberg, H. I. Untersuchungen über die morphologische und biologische Differenzierung in der *Fusarium*-Section Liseola. *Mitteilungen Biol. Bundesanst. Für Land- Forstwirtschaft.* **169**, 1–117 (1976).
57. Rayner, R. W. *A mycological colour chart*. (Commonwealth Mycological Institute, Kew, Surrey, 1970).
58. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671 (2012).
59. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
60. Gurevich, A. I., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29** (2013).

61. 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* **btv351**, <https://doi.org/10.1093/bioinformatics/btv351> (2015).
62. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
63. Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463** (2010).
64. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
65. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637 (2008).
66. Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998).
67. Zampounis, A. *et al.* Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease. *Genome Announc.* **4**, e00821–16 (2016).
68. Hiruma, K. *et al.* Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. *Cell* **165**, 464–474 (2016).
69. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
70. Kraus, O. Z. *et al.* Automated analysis of high-content microscopy data with deep learning. *Mol. Syst. Biol.* **13**, 924 (2017).
71. Löytynoja, A. & Goldman, N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635 (2008).
72. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
73. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
74. Conway, J. & Gehlenborg, N. UpSetR: A more scalable alternative to Venn and Euler diagrams for visualizing intersecting sets (2017).
75. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (2008).
76. Bryant, D. M. *et al.* A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**, 762–776 (2017).
77. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinforma. Oxf. Engl.* **23**, 257–258 (2007).

Acknowledgements

This work was supported in part by Grant-in-Aid for Scientific Research (KAKENHI) (17H06172 and 15H05959 to K.S.; 18H02204 to Y.T.; 19K15846 to P.G.), the Science and Technology Research Promotion Program for the Agriculture, Forestry, Fisheries, and Food industry to Y.N., Y.T. and K.S. and JSPS Grant-In-Aid for JSPS Research Fellow (17J02983) to A.T. We would like to thank Dr. Volker Otte for helpful translation of texts.

Author Contributions

P.G., A.T. and R.H. conducted experiments and analysed results, M.K. collected the *Colletotrichum* isolate and conducted experiments, P.G., K.S. and U.D. conceived and designed experiments, Y.T., Y.N., M.N. and K.S. conceived the project. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-50076-5>.

Competing Interests: The authors declare no competing interests.

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



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

© The Author(s) 2019