Comparative genomics of three *Colletotrichum scovillei* strains and genetic analysis revealed genes involved in fungal growth and virulence on chili pepper

(Short Title: Genes involved in growth and virulence of *Colletotrichum scovillei*)

Dai-Keng Hsieh[1], Shu-Cheng Chuang[1,#a], Chun-Yi Chen[2], Ya-Ting Chao[2], Mei-Yeh Jade Lu[3], Ming-Che Shih[1,2,*] and Miin-Huey Lee[1,4,5,*]

[1] Ph.D. Program in Microbial Genomics, National Chung-Hsing University and Academia Sinica, Taichung, 40227, Taiwan

[2] Agricultural Biotechnology Research Center, Academia Sinica, Taipei 11529, Taiwan

[3] Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan

[4] Departmment of Plant Pathology, National Chung-Hsing University, Taichung 40277, Taiwan

[5] Advanced Plant Biotechnology Center, National Chung-Hsing University, Taichung 40277, Taiwan

[#a] Current Address: Department of Plant and Environment Protection Sciences, University of Hawaii at Manoa, Honolulu, HI, USA

*Corresponding authors: Miin-Huey Lee, e-mail: mhlee@nchu.edu.tw; Ming-Che Shih, e-mail: mcshih@gate.sinica.edu.tw

Abstract

Colletotrichum scovillei is a virulent pathogen and the dominant species causing anthracnose of chili pepper in many Asian countries. Three strains of this pathogen, Coll-524, Coll-153 and Coll-365, show varied virulence on chili pepper fruit. Among the three strains, Coll-365 showed significant defects in growth and virulence. To decipher the genetic variations among these strains and identify genes contributing to growth and virulence, in this study, comparative genomic analysis and gene transformation to verify gene function were applied. The genomes of the three strains were sequenced and Coll-524 had 1.3% and 1.5% more genes than Coll-153 and Coll-365, respectively. Compared to Coll-524 and Coll-153, Coll-365 had numerous gene losses including 33 effector genes that are distributed in different scaffolds and a cluster of 14 genes in a 34-kb genomic fragment. Through gene transformation, three genes in the 34-kb fragment were identified to have functions in growth and/or virulence of *C. scovillei.* Gene 15019 encoding a protein related to phospholipase A2-activating protein enhanced the growth of Coll-365. A combination of 15019 with one transcription factor gene 15022 and one C6 zinc finger domain-containing protein gene 15029 was found to enhance the pathogenicity of Coll-365. Introduction of gene 15215, which encodes a LysM domain-containing protein, into Coll-365 caused a reduction in the germination rate of Coll-365. In conclusion, the higher virulent strain Coll-524 had more genes and encoded more pathogenicity related proteins and transposable elements than the other two strains, which may contribute to the

high virulence of Coll-524. In addition, the absence of the 34-kb fragment plays a critical role in the defects of growth and virulence of strain Coll-365.

Author Summary

Colletotrichum scovillei is a highly virulent and dominant pathogen causing anthracnose of chili that leads to significant economic loss in chili production in many Asia countries. In this study we focus on finding the gene differences of three *C. scovillei* strains with different pathogenicity in chili pepper infection and verifying the function of some genes in the lowest virulence strain. We sequenced them and did gene annotation and genome comparison. We setup a simple mathematical method to identify gene variations between strong and weak virulence strains. Our results show that the lowest virulence strain has less pathogenicity-related genes. We also found that the absence of 14 genes in a compact genomic fragment was part of the reason of growth and virulence defect of the lowest virulence strain. We identified four genes that play roles on fungal growth and/or virulence on chili pepper. We also found a group of effector genes that specifically appear in species collected form infected chili in *C. acutatum* species complex. Our research provides detailed information for why the three strains have different virulence on chili pepper.

Introduction

Chili pepper (*Capsicum* spp.) is a globally significant spice crop. The cultivation of chili pepper is frequently threatened by the attacks of various pathogens, especially the anthracnose pathogens of the *Colletotrichum* species. *Colletotrichum* contains a large number of species members that have been classified into several species complexes [1]. *Colletotrichum* pathogens associated with *Capsicum* plants come from diverse species complexes including *C. acutatum*, *C. boninense*, *C. gloeosporioides*, *C. orchidearum* and *C. truncatum* complexes [2-6]. At least 28 species have been reported and most of them are from the acutatum, gloeosporioides and truncatum complexes [7, 8].

C. scovillei (formerly *C. acutatum*) is a highly virulent and dominant pathogen causing anthracnose of chili that leads to significant economic loss in chili production in many Asia countries [4, 8]. It can infect many different species of *Capsicum*, especially species that are mainly grown for human consumption such as *C. annuum*, *C. frutescens*, *C. chinense*, and *C. baccatum* [4, 8, 9]. *C. scovillei* is the most dominant species causing anthracnose of chili in Asia, and is found widely distributed in Indonesia, Malaysia, Thailand and Taiwan [8, 10]. Two pathotypes of *C. scovillei*, CA1 and CA2, have been identified in Taiwan by the AVRDC–World Vegetable Center [11, 12]. CA2 is more virulent than CA1 in most tested cultivars of *Capsicum* species [9, 11]. CA2 pathotype breaks down the resistance of *Capsicum chinense* PBC932-dervied lines, which are resistant to CA1 pathotype [9, 11, 12]. CA2 distribution was limited before the year 2000, but since then has become dominant and replaced CA1 in most of the chili pepper planting locations in Taiwan [11, 12]. According to amplified fragment length polymorphism (AFLP) analysis, CA2 members are

87 homogenous and mostly clonal, whereas CA1 are genetically diverse [11]. Strains Coll-153 and
88 Coll-365 are grouped in CA1 pathotype and strain Coll-524 is a member of the CA2 pathotype.
89 Understanding the variation among the three strains may provide insight about the high virulence of
90 the CA2 pathotype.

91 Whole genome sequencing has had a significant impact on understanding the biology, ecology,
92 genetics and evolution of various organisms. In fungi, comparative genomic study of genetic
93 variations has revealed mobile pathogenicity chromosomes in *Fusarium* [13], evolutionary
94 adaptations from a plant pathogenic to an animal pathogenic lifestyle in *Sporothix* [14], an influx of
95 transposable elements creating a genetically flexible landscape to respond to environmental changes
96 in *Pyrenophora tritici-repentis* [15], potential orthologs in adaptation to specific hosts or ecological
97 niches in *Botrytis*, *Colletotrichum, Fusarium*, *Parastagonospora* and *Verticillium* [16-21], the
98 evolution and diversity of putative pathogenicity genes in *Colletotrichum tanaceti* and other
99 *Colletotrichum* species [22, 23], core and specific genetic requirements for fungal endoparasitism of
100 nematodes in *Drechmeria coniospora* [24], host-specific and symptom-related virulence factors
101 horizontally transferred from *Fusarium* to *Verticillium* [25, 26]. Genomic comparisons with multiple
102 races within a species have also been performed recently in *Fusarium* and *Verticillium* and revealed
103 that secondary metabolites are crucial factors for stunting symptom development in *F. fujikori* and
104 defoliation symptom development in *V. dahliae* [26, 27]. In addition, a study on 18 isolates of *V.*
105 *dahliae* identified highly variable regions with race specific signatures [28]. However, most of these

106 studies were in-silica analyses. Only few included genetic studies to demonstrate the functions of the
107 potential factors selected by in-silica analysis, and these studies were on *Fusarium* and *Verticillium*
108 [18, 25-27].

109 In our previous study, we investigated the plant-pathogen interactions between the chili pepper
110 fruit and the three *C. scovillei* strains Coll-153, Coll-365 and Coll-524 and revealed the infection
111 process and several potential virulence factors determining the infection and colonization of host
112 tissues [9]. On the original chili pepper host, Coll-524 usually shows higher virulence than the other
113 two strains. Coll-365 shows significantly slower growth on agar medium and has the weakest
114 virulence on *Capsicum spp.*. Coll-365 produces less spores in axenic culture and *in planta*. Coll-365
115 was found to accumulate less turgor pressure in appressorium but produces higher levels of cutinase
116 activity than the other two strains. Examination of the infection process showed that the three strains
117 can form primary hyphae in the epidermal cells at 72 h post-inoculation (hpi), indicating no delay in
118 the penetration step for Coll-365 compared to Coll-524 and Coll-153. The three strains can grow into
119 the cuticle layer of chili pepper fruit at 24 hpi and form a highly branched structure (HBPS) within
120 the cuticle layer and penetrate epidermal cells at 72 hpi [29]. After infecting the epidermal cells of
121 chili pepper fruits, Coll-365 appeared to have less ability to colonize host tissues, formed limited
122 lesions on infected tissue and produced much fewer spores on the infected tissue. In addition, Coll-
123 524 was more resistant to host defense compound capsaicin than Coll-153 and Coll-365. In this

124 previous work, the genetic variations and genes contributing to the defects of growth and virulence
125 remained undeciphered.

126 In this study, to decipher the genetic variations of the three *C. scovillei* strains, we sequenced the
127 genomes of the three strains, investigated the differences in their genome compositions, identified
128 potential genes involved in the phenotypes, and verified genes involved in the defects of growth and
129 virulence of Coll-365. We setup a simple mathematical method to identify ORF variations and
130 successfully identified DNA fragments involved in the defects in growth and virulence of Coll-365.
131 Moreover, by gene transformation we have demonstrated four genes that function in germination,
132 growth and/or virulence of the chili pepper pathogen *C. scovillei*.

133

134 Results

135 Phylogenetic analysis revealed that Coll-524, Coll-153 and Coll-365 are
136 closely related to *C. scovillei*

137 To understand the phylogenetic relationships among the three *Colletotrichum* strains, multi-locus
138 phylogenetic analysis with 174 fungal strains was performed. The results indicated that Coll-153,
139 Coll-365 and Coll-524 all belong to the acutatum species complex **(S1 Fig)**. Further examination of
140 the original hosts of members in acutatum species complex revealed that the three strains formed a
141 small clade with other strains and most of the members were isolated from *Capsicum* plants (**S2 Fig**).
142 However, the phylogenetic tree generated from the 173 *Colletotrichum* strains could not distinguish

143 the phylogenetic relationship of the three strains with *C. scovillei* (CBS 126529, the holotype strain
144 of *C. scovillei* [2]) and C. *guajavae* (IMI 350839, the holotype strain of *C. guajavae* [2]) **(S2 Fig)**.
145 Another multi-locus phylogenetic analysis was conducted with 18 strains **(S1 Table)**, including 12
146 strains from **S2 Fig** and 6 strains (marked with #) that were closely related to *C. scovillei* CBS
147 126529 and *C. guajavae* IMI 350839 in the clade 2 of the *Colletotrichum acutatum* species complex
148 [2]. The results indicated that the three strains were closer to *C. scovillei* CBS 126529 than *C.*
149 *guajavae* IMI 350839 **(Fig 1A)**. In addition, Coll-524 was closer to *C. scovillei* TJNH1 isolated from
150 pepper in China and two *C. fioriniae* strains (HC89 and HC91) isolated from apple in the USA than
151 Coll-365 and Coll-153 **(Fig 1A)**. These results indicated that strains Coll-524, 153 and 365 all belong
152 to the *C. scovillei* species. In addition, the three strains showed significant differences in growth and
153 virulence but had a similar preference for temperature range (**Fig 1B-1D and** [9]).
154

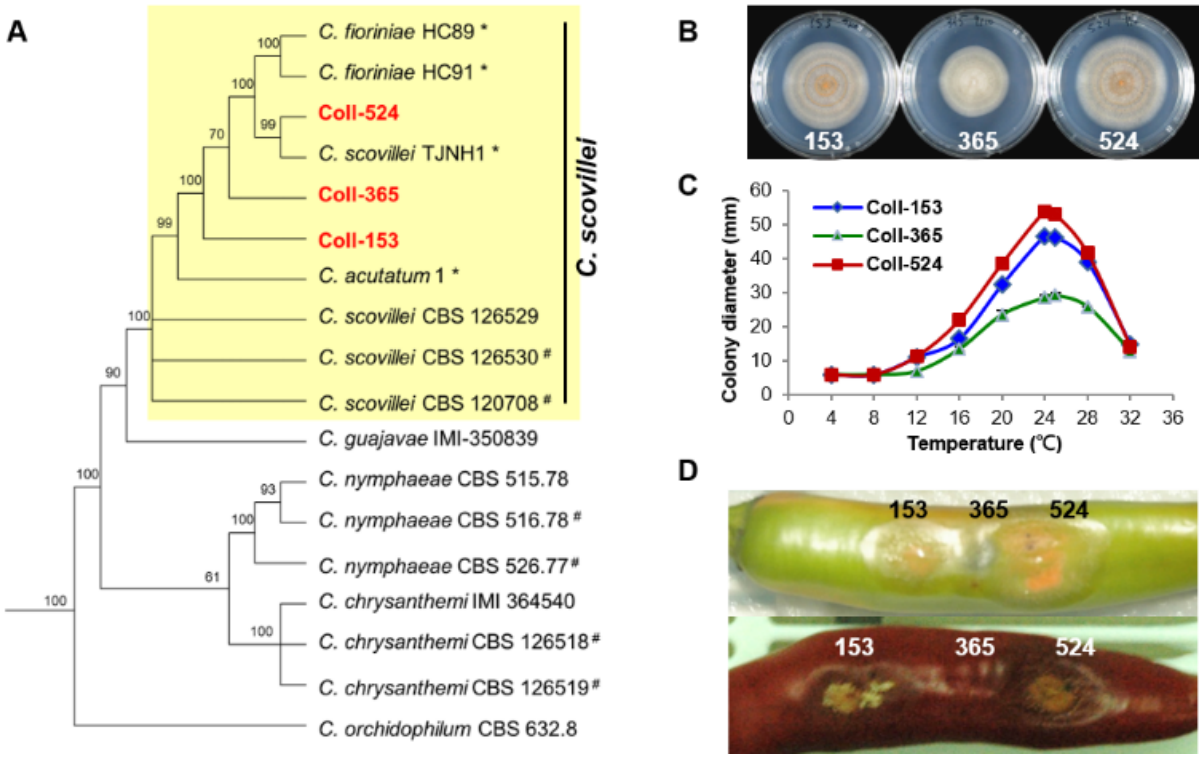


Figure 1. Phylogenic relationships (A), colony morphology (B), growth (C), and virulence (D) of strains Coll-524, Coll-153 and Coll-365. (A) Multi-locus phylogenetic tree of 18 *Colletotrichum* strains of acutatum species complex was constructed with five molecular markers (ACT1, CHS1, GAPDH, ITS and TUB2). Yellow square indicated that Coll-153, Coll-365 and Coll-524 were in the same clade with *C. scovillei* CBS 126529, the holotype strain of *C. scovillei*. Detail information for the 18 strains was in **S1 Table.** Posterior probability support values were indicated on branches. (B) Colony morphology of the three strains culturing on MS agar media for 6 days. (C) Growth of the three strains on various temperatures on PDA for 7 days. (D) Symptom of the three strains on green and red chili pepper fruits cv. Groupzest.

Genomic comparison revealed DNA deletion patterns among different scaffolds in the three strains

The genomic sequences of Coll-524 were generated from paired-end and mate-pair libraries and assembled and annotated using the ALLPATHS-LG [30] and MAKER [31] pipelines. Genomic sequences of Coll-153 and Coll-365 were generated from paired-end libraries and mapped to the assembled Coll-524 genome (for details, see Materials and Methods). The results of assembly and annotation are summarized in **Table 1**. The genome size of Coll-524 was 51. 491 MB and consisted of 54 scaffolds with an N50 of ~3.6 Mb. The genome size of Coll-153 was 50.114 MB and consisted of 59 scaffolds. The genome size of Coll-365 was 49.922 MB and consisted of 59 scaffolds. A total of 15,626, 15,432 and 15,387 genes were annotated in the Coll-524, Coll-153 and Coll-365 genomes, respectively **(Table 1).** BUSCO [32] analysis for genome completeness showed that 99.8% of the conserved proteins in sordariomycetes could be identified in all three strains.

Table 1. Genome features of *Colletotrichum scovillei* strains Coll-524, Coll-153 and Coll-365.

Genome features	**Coll-524**	**Coll-153**	**Coll-365**
Sequencing	Paired-end/mate pair	Paired-end	Paired-end
Assembly	De novo	Mapped to Coll-524	Mapped to Coll-524
Coverage	71.9	105	104
Fragment coverage	37.2	-	-
Jump coverage			

3K	24.4	-	-
8K	10.3	-	-
Number of scaffolds	54	59	59
Total scaffold length (Mb)	51.491	50.114	49.922
N50 scaffold size (Mb)	3.597	-	-
Max scaffold length (Mb)	8.040	7.995	7.989
Number of scaffolds (> 50 KB)	26	25	24
GC content (%)	51.98	52.03	52.05
BUSCO completeness (%)	99.80	99.80	99.80
Total genes	15,626	15,432	15,387

Coll-153 and Coll-365 had their own special genome fragments compared to Coll-524, and there were six scaffolds for Coll-153 and seven scaffolds for Coll-365. One (0.9-kb) and two (0.9-kb and 5.1-kb) of the 54 scaffolds in Coll-524 did not exist in Coll-153 and Coll-365, respectively, and no genes were encoded in the two scaffolds. A total of 29 and 30 genes were annotated from Coll-153 and Coll-365 specific genome sequences, respectively, and 23 of them are homologs in both Coll-153 and Coll-365. None of these genes had traits related to pathogenicity when blasting to the PHI database.

Genome mapping of Coll-153 and Coll-365 to Coll-524 was analyzed with CLC Genomic Workbench v.9.5.1 and provided the specified information of DNA insertion, deletion, removal, and miscellaneous differences (misc. differences, which included single nucleotide polymorphism (SNP)

and multiple nucleotide polymorphism (MNP) in this study; see Materials and Methods) in Coll-153 and Coll-365 **(Table 2)**. The results showed that Coll-153 and Coll-365 had similar patterns within the four types of variations, but Coll-365 had a slightly higher number of total variations than Coll-153. The removed sequence database (RSD; see Materials and Methods) showed that 991,585 and 1,179,624 bp were absent in Coll-153 and Coll-365, respectively, compared to Coll-524. The largest removed fragments were 28,776 bp in Coll-365 and 10,125 bp in Coll-153. Scaffolds 19, 20 and 22 in Coll-153 and Coll-365 showed high numbers of DNA sequence removal compared to the three scaffolds in Coll-524 **(Fig 2)**. The total number of sites of DNA sequence removal in scaffolds 19, 20 and 22 were 377, 386, and 256 in Coll-153, and 167, 114 and 76 in Coll-365, respectively. A notable difference in DNA sequence removal patterns between Coll-153 and Coll-365 was that DNA sequence removal occurred significantly in scaffolds 17 in Coll-365 but not in Coll-153 (**Fig 2**). Large DNA fragment removals occurred in Coll-365 more frequently than Coll-153 (**Table 3**). However, most of them were located at the short scaffolds 19, 20 and 22. Coll-365 had four removals with DNA fragments larger than 20 kb, but none of the large fragments was found in Coll-153. Regarding the removal of DNA fragments larger than 5 kb, a total of 66 removals appeared in Coll-365 but only 8 in Coll-153. Small DNA fragment removals (< 1 kb) were found most often in Coll-153.

Table 2. Genome mapping of strains Coll-153 and Coll-365 to Coll-524 analyzed with CLC to provide the datasets of DNA insertion, deletion, removed, and misc. difference in the genomes of Coll-153 and Coll-365.

Variation Datasets	No. of variation sites or nucleotides	
	Coll-153	**Coll-365**
Insertion		
Insertion sites	241	361
Insertion bps	338	506
Deletion		
Deletion sites	302	451
Deletion bps	417	618
Misc. difference[a]		
Misc. difference sites	7,411	9,833
Misc. difference in bps	9,179	12,218
Removed		
Removed sites (N[b]<20% of the removed region)	1,623	731
Removed bps (N[b]<20% of the removed region)	991,585	1,179,624

[a], the dataset of Misc. difference consisted of single nucleotide polymorphism (SNP) and polymorphism with more than one nucleotide which was named as multiple nucleotide polymorphism (MNP) in this study.

[b], N indicates gaps in scaffolds.

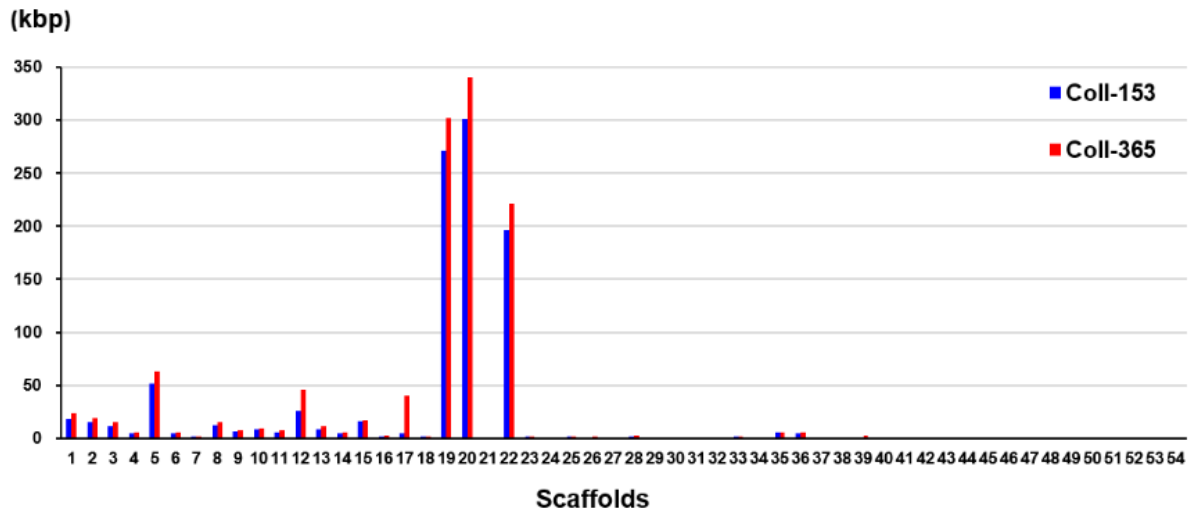


Figure 2. DNA sequence removal in strains Coll-153 and Coll-365 compared to strain Coll-524.

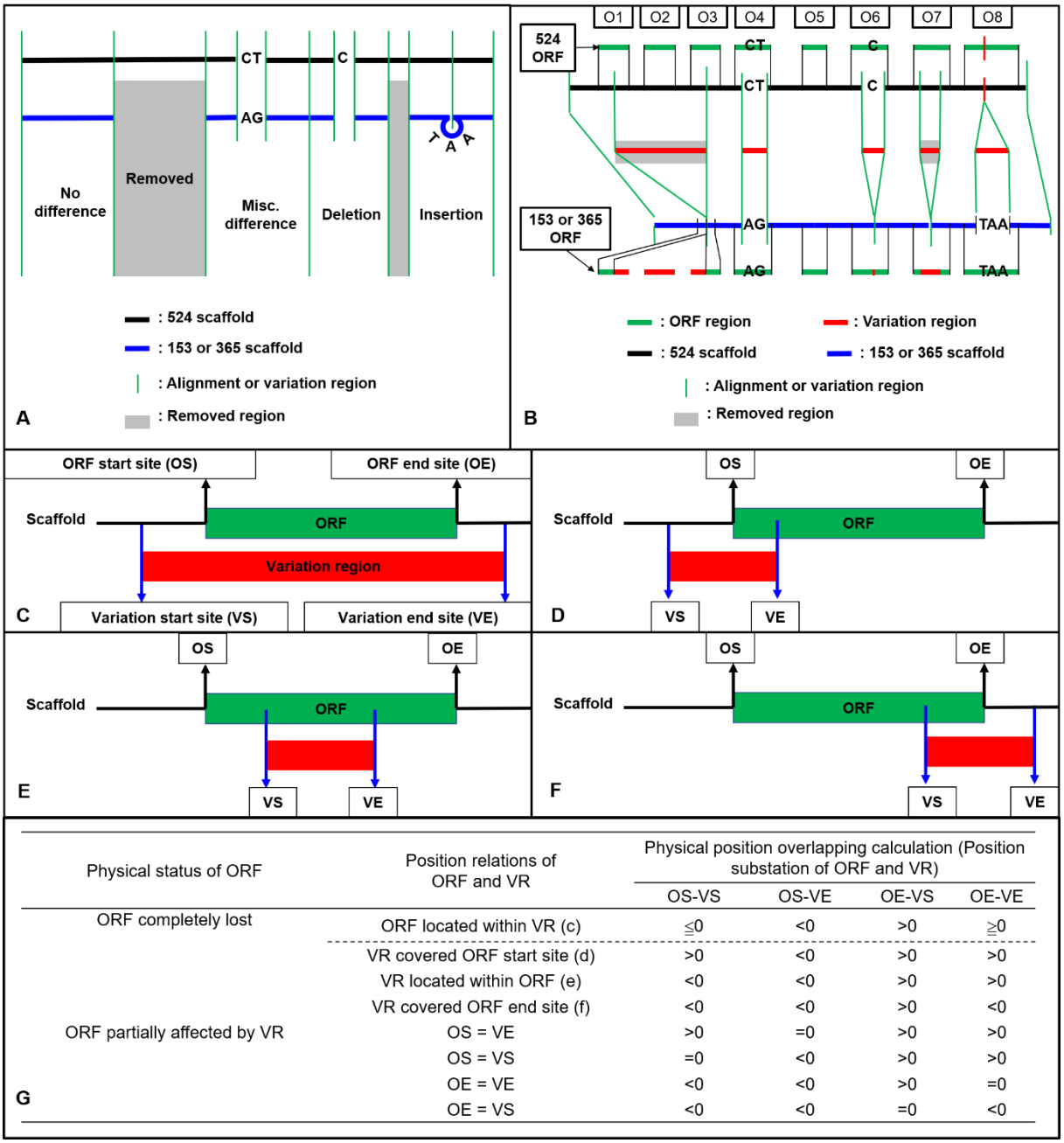
Total nucleotides removed in each scaffold were indicated as blue and red bars for Coll-153 and Coll365.

Table 3. Large DNA fragment removal statistic of strains Coll-153 and Coll-365.

DNA removal	Coll-153	Coll-365
≧20 kb	0	4
≧10 kb	1	21
≧5 kb	7	41
≧1 kb	325	175
<1 kb	1290	490
Sum	1623	731

ORF variation analysis identified ORF losses in strains Coll-153 and 365

DNA insertion, deletion, removal, and misc. differences (SNP and MNP) occurring in Coll-153 or Coll-365 may affect the open reading frames (ORFs) of their genes. To understand the influence on ORFs, we used a simple mathematical method to detect ORF variation (ORF-V) for all ORFs in Coll-153 and Coll-365 as illustrated in **Fig 3** and described in Materials and Methods. The results showed that most of the ORF-Vs occurred with SNP in Coll-153 and 365, and Coll-365 appeared to have significantly greater numbers of full ORF losses than Coll-153 **(Table 4)**. The total numbers of ORF-V of Coll-153 and Coll-365 compared to Coll-524 were 708 and 725, respectively, and 799 of them were singular. A total of 91 and 74 ORF-Vs were solely found in Coll-365 and Coll-153, respectively, but most of these ORF-Vs were SNP in Coll-153 or Coll-365, indicating that the two strains had similar genes with ORF-V to Coll-524. The major difference in the two strains with regard to ORF-V was found in scaffold 17 and was caused by ORF deletion as described below. The 799 singular ORFs were used for further analysis.



G

Physical status of ORF	Position relations of ORF and VR	Physical position overlapping calculation (Position substation of ORF and VR)			
		OS-VS	OS-VE	OE-VS	OE-VE
ORF completely lost	ORF located within VR (c)	≦0	<0	>0	≧0
	VR covered ORF start site (d)	>0	<0	>0	>0
	VR located within ORF (e)	<0	<0	>0	>0
	VR covered ORF end site (f)	<0	<0	>0	<0
ORF partially affected by VR	OS = VE	>0	=0	>0	>0
	OS = VS	=0	<0	>0	>0
	OE = VE	<0	<0	>0	=0
	OE = VS	<0	<0	=0	<0

Figure 3. Strategy illustration (A-B) and calculation method (C-G) used for ORF variation identification. (A) Variations were generated after reads mapping of Coll-153 or Coll-365 to Coll-524. The mapping could result in four types of variations (Removed, Misc. Difference, Deletion and Insertion). (B) ORF regions of Coll-524 were compared to variation regions to find the overlappings. An ORF with the overlapped region indicated that this ORF is varied between the two strains. (C-F)

248 Calculation methods used to identified all potential position relations between an ORF and variation
249 regions. (G) Calculation for position difference to identify ORF variation types, removed or partial
250 differences. VR, variation region; OS, ORF start site; OE, ORF end site; VS, variation start site; VE,
251 variation end site.

252

253 **Table 4. ORF variation statistics of strains Coll-153 and Coll-365 detected and calculated**
254 **based on the method shown in Fig 4.**

ORF variations	Variation types	No. of ORF with variations		ORF-V or ORF-Vr
		Coll-153	Coll-365	
ORF loss	Completely Removed	92	166	ORF-V/ORF-Vr
ORF partially affected	Partially Removed	126	84	ORF-V/ORF-Vr
	SNP	404	400	ORF-V
	MNP	4	4	ORF-V
	SNP, MNP, Deletion, Insertion and Removed	82	71	ORF-V
Sum		708	725	
Union of varying ORFs within Coll-153 and 365			799	

255 Note: SNP, single nucleotide polymorphism; MNP, polymorphism with more than one nucleotide.

256

257 ORFs affected by DNA removal (ORF-Vr) in each scaffold were further analyzed and the
258 results are displayed in **Fig 4**. ORF-Vr mostly occurred in scaffolds 19, 20 and 22 in Coll-153 and

365. There were a total of 260 and 274 ORF-Vr events in Coll-153 and Coll-365, respectively. All ORFs were ORF-Vr in scaffolds 19, 20, 22 and 44; however, scaffold 44 was a short scaffold with only one ORF (**Fig 3**). The distribution of ORF-Vr in Coll-153 and 365 was very similar except that 14 ORFs clustered in a 34-kb genomic fragment were totally or partially removed in Coll-365 compared to 1 ORF with SNP only in Coll-153 in scaffold 17 (**Fig 5**). The 14 ORFs encode 4 transcription regulation-related proteins, 1 GPI-anchored protein, 4 enzymes and enzyme-related proteins, and 5 hypothetical proteins. In addition, when compared with 62 genomes of other *Colletotrichum* strains that are available in the NCBI database, the 14 genes appeared in almost all members of the acutatum species complex, except Coll-365 (**Fig 6**). Moreover, genes 15003 and 15019 existed in nearly all assayed strains, indicating that the two genes are core genes of the *Colletotrichum* species. Five of these genes existed in almost all the strains in the acutatum species complex, suggesting the possibility of them being acutatum species complex-specific genes (**Fig 6**).

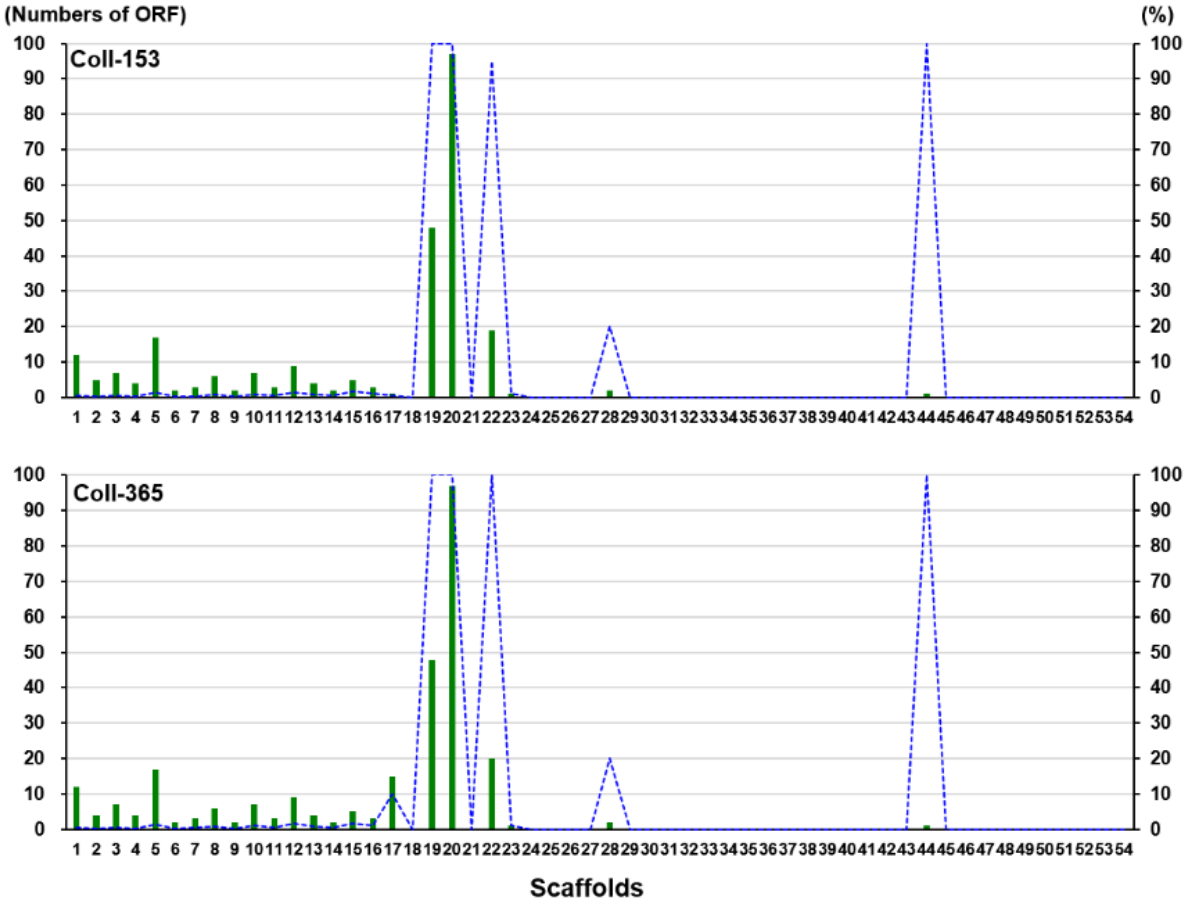
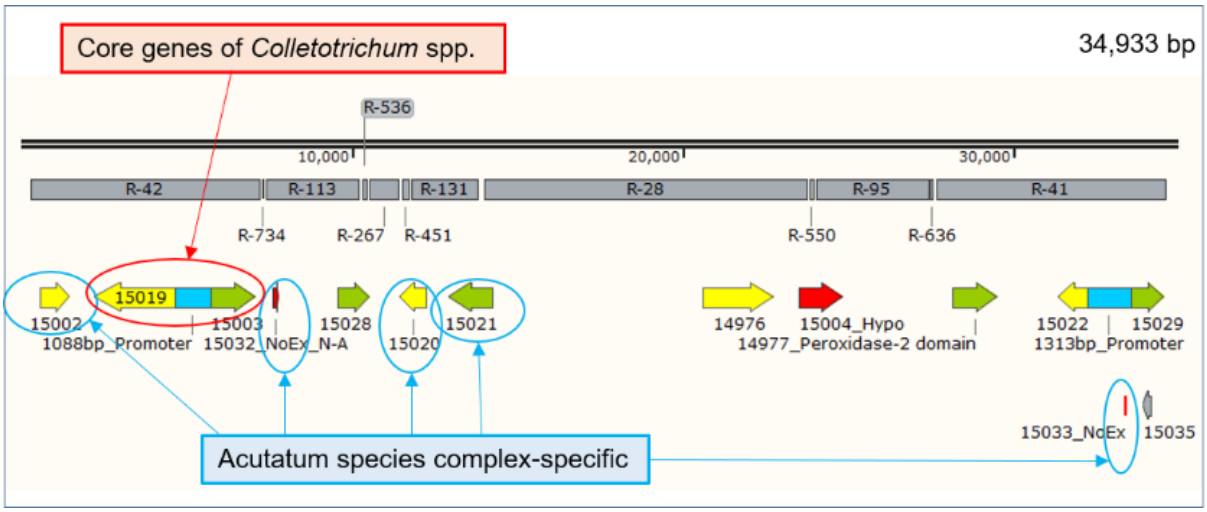


Figure 4. ORF-Vr in strains Coll-153 and Coll-365 compared to strain Coll-524 in each scaffold. Total ORF-Vr in each scaffold were indicated as green bars. The percentage of ORF-Vr in each scaffold by comparing with ORFs of Coll-524 in each scaffold was indicated with blue lines.



277

278 **Figure 5. Genetic map of the 14 removed genes in scaffold 17 of strain Coll-365.** The gray bars

279 with R-number indicated the removed fragments in Coll-365 compared to Coll-524. The 14 genes

280 located in the 34.9-kb fragment of scaffold 17 are genes 15002, 15019, 15003, 15032, 15028, 15020,

281 15021, 14976, 15004, 14977, 15022, 15029, 15033 and 15035. NoEx: Expression data are non-

282 available.

283

Species complex	Strain name	Host	15003	15019	14977	15029	15035	15022	14976	15028	15004	15021	15020	15002	15032	15033
Orbiculare	*C. orbiculare* MAFF 240422	*Cucumis sativus*	●	●		●	●		●							
	C. trifolii 543-2	*Medicago sativa*	●	●	●	●	●		●							
	C. sidae CBS 518-97	*Sida spinosa*	●	●		●	●		●	●						
	C. spinosum CBS-515-97	*Xanthium spinosum*	●	●	●	●	●									
	C. lindemuthianum 83.501	*Phaseolus vulgaris*	●	●	●		●									
	C. lindemuthianum 89 A2	*Phaseolus vulgaris*	●	●	●		●									
Acutatum	*C. godetiae* C184	*Olea europaea* (fruit)	●	●					●	●		●		●	●	
	C. acutatum C71	*Olea europaea* (olive endocarp)	●	●	●	●	●	●		●	●	●	●		●	●
	C. fioriniae PJ7	Strawberry in New Zealand	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	C. fioriniae HC89	*Malus* sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	C. fioriniae HC91	*Malus* sp.	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	C. scovillei TJNH1	*Capsicum annuum*	●	●	●		●	●	●	●	●	●	●	●	●	●
	Coll-524	*Capsicum annuum*	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Coll-153	*Capsicum annuum*	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	Coll-365	*Capsicum annuum*														
	C. acutatum 1	Pepper	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	C. nymphaeae SA-01	*Fragaria x ananassa*	●	●	●	●	●	●	●	●	●	●	●	●		●
	C. orchidophilum IMI 309357	*Phalaenopsis* sp.	●	●	●	●	●	●	●	●	●	●	●			
Graminicola	*C. graminicola* M1.001	*Zea mays*	●	●		●	●	●	●							
	C. graminicola M5.001	*Zea mays*	●	●		●	●	●	●							
	C. sublineola CgSl1	*Sorghum bicolor*	●	●	●	●	●	●	●							
	C. sublineola TX430BB	*Sorghum bicolor*	●	●	●	●	●	●	●							
	C. falcatum Cf671	*Saccharum officinarum*	●	●					●							
Spaethianum	*C. tofieldiae* 861	*Arabidopsis thaliana* (root)	●	●	●	●	●		●	●	●					
	C. tofieldiae CBS127615	*Agapanthus* sp.	●	●	●	●	●		●	●	●					
	C. tofieldiae CBS130851	*Semele androgyna*	●	●	●	●	●		●	●	●					
	C.tofieldiae CBS168.49	*Lupinus polyphyllus*	●	●	●	●	●		●	●	●					
	C. incanum MAFF 238704	*Raphanus sativus* L. Daikon Group	●	●	●	●	●	●	●							
	C. incanum MAFF 238712	*Raphanus sativus* var. *longipinnatus*	●	●	●	●	●	●	●							
Destructivum	*C. higginsianum* MAFF 305635-RFP	*Brassica rapa* var. *perviridis*	●	●	●				●	●						
	C. destructivum YC1	Tobacco (leaf)	●	●	●	●			●	●						
	C. shisoi PG-2018a	*Perilla frutescens* var. *crispa*	●	●	●				●							
	C. lentis CT-30	*Lentil*	●	●					●							
	C. tanaceti BRIP57314	*Tanacetum cinerariifolium*	●	●					●							
	C. coccodes NJ-RT1	Pepper fruit	●	●		●	●		●	●						
	C. coccodes RP180a	Pepper fruit	●	●		●	●		●	●						
	C. chlorophyti NTL11	*Solanum lycopersicum* (leaf)	●	●	●	●		●	●							
	*C.*sp. JS-367	*Sansevieria trifasciata*	●	●	●	●	●	●	●							
Gloeosporioides	*C. gloeosporioides* Cg01	*Huperzia serrata*	●	●	●	●	●	●		●						
	C. gloeosporioides ES026	*Huperzia serrata*	●	●	●	●	●	●		●						
	C. camelliae CcLH18	*Camellia oleifera* (leaf)	●	●	●	●	●	●		●						
	C. gloeosporioides CgLH19	*Camellia oleifera* (leaf)														
	C. sp. COLG25	*Citrus*	●	●	●	●	●	●		●						
	C. gloeosporioides SMCG1C	*Cunninghamia lanceolata* (leaf)	●	●	●	●	●	●								
	C. gloeosporioides Lc1	*Liriodendron chinense*	●	●	●	●	●	●	●							
	C. siamense CsLH17	*Camellia oleifera* (leaf)	●	●	●	●	●	●		●						
	C. siamense COLG-38	Citrus	●	●	●	●	●	●								
	C. siamense COLG-44	Citrus	●	●	●	●	●	●		●						
	C. siamense COLG-50	Citrus	●		●	●	●	●		●						
	C. siamense COLG-90	Citrus	●	●	●	●	●	●		●						
	C. fructicola CGMCC3-17371	Strawberry plant	●	●	●	●	●	●		●						
	C. fructicola Nara5	*Cucumis sativus*	●	●	●	●	●	●		●						
	C. fructicola 1104-7	From an apple leaf spot lesion	●	●	●	●	●	●		●						
	C. fructicola 15060	Mango cv. Tainong	●	●	●	●	●	●		●						
	C. gloeosporioides 30206	Apple	●	●	●	●	●	●		●						
	C.gloeosporioides COLG-95	Citrus	●	●	●	●										
	C. sp. COLG31	Citrus	●	●	●	●	●	●								
	C. gloeosporioides Cg 14	Avocado cv. Fuente	●	●	●	●		●								
	C. musae GM20	*Musa* sp.	●	●	●	●	●			●						
Truncatum	*C. capsici* KLC-C	*Capsicum annuum* (Fruit)	●	●	●	●		●	●	●						
	C. truncatum MTCC-3414	*Capsicum annuum* (Ripe fruit)	●	●	●	●		●	●	●						
	C. gloeosporioides TYU	*Taxus cuspidata*	●	●	●	●		●	●	●						
	C. truncatum KLC-C5	Pericarp of Hot pepper fruit	●	●	●	●		●	●	●						
Boninense	*C. karsti* CkLH20	*Camellia oleifera* (leaf)	●	●	●	●	●	●	●	●						
	C. sansevieriae Sa-1-2	*Sansevieria trifasciata*	●	●	●		●	●	●							

Figure 6. The distribution of the 14 genes of scaffold 17 in strains Coll-524, 153, and 365, and 62 *Colletotrichum* strains. The 14 genes were analyzed using nucleotide BLAST against the 62

287 genomes of *Colletotrichum* strains. Blue squares indicate the presence of genes by the blasting with
288 coverage >90% and identity >60%. The gene was ordered from left to right by distribution frequency
289 in the 62 genomes.

290

291 **Gene ortholog analysis showed that the three strains carry slightly**
292 **different ortholog compositions**

293 To understand the difference in orthology among the three strains, proteins were analyzed using
294 OrthoFinder [33]. Ortholog analysis showed a high similarity in gene compositions among the three
295 strains as shown in **Table 5**. More than 88% genes in the three strains belonged to single-copy
296 orthogroups, indicating that the three strains all carried the 13,779 single-copy genes (**Table 5**). A
297 total of 502 multiple-copy orthogroups with the same gene numbers containing 1,167 genes were all
298 identified in the three strains. A combination of single-copy orthogroups and multi-copy orthogroups
299 with the same gene copy number showed that there were a total of 14,946 genes in the three strains
300 and they were 95.7, 96.9 and 97.1% of the total genes in Coll-524, 153 and 365, respectively. This
301 suggests that the three strains had high similarity in gene compositions **(Table 5)**. The ortholog
302 differences within the three strains analyzed using the dataset of multi-copy orthogroups with
303 different gene number showed Coll-524 carried considerably greater numbers of genes than the other
304 two strains. The three strains all carried the 128 orthogroups but they had different gene numbers in
305 each of the orthogroups. Among the 128 orthogroups, Coll-524 had 97 and 101 more genes than

306 Coll-153 and Coll-365, respectively, and there were a total of 103 singular genes found from the 97
307 and 101 genes. Regarding the differences in orthogroups that only existed in two strains, 92, 33 and
308 94 orthogroups were identified in Coll-524 and Coll-153, Coll524 and Coll-365, and Coll-153 and
309 Coll-365, respectively (**Table 5**). Further analysis of the differences between pairs of two strains
310 showed that some orthogroups were only found in two strains, and Coll-524 and Coll-153 shared
311 more orthogroups (92 orthogroups, 94 genes) than Coll-524 and Coll-365 (33 orthogroups, 33
312 genes). Based on the data mentioned above, the combinations of the 103 genes, the 94 and 33 genes,
313 the 85 of strain-specific genes and the 196 unassigned genes of Coll-524 strain, 511 genes in total,
314 were used together with other criteria for further gene selections used in gene functional analysis.
315

316 **Table 5. Orthogroup types and distributions of *Colletotrichum scovillei* strains Coll-524, Coll-**
317 **153 and Coll-365.**

Numbers of strains in orthogroup	Descriptions of Orthogroups	Numbers of orthogroups (numbers of genes)		
		Coll-524	Coll-153	Coll-365
3	Single-copy orthogroups	13,779 (13,779)	13,779 (13,779)	13,779 (13,779)
	Multi-copy orthogroups with same numbers	502 (1,167)	502 (1,167)	502 (1,167)
	Multi-copy orthogroups with different numbers	128 (272)	128 (192)	128 (194)

2	Orthogroups in Coll-524 and Coll-153	92 (94)	92 (92)	-
	Orthogroups in Coll-524 and Coll-365	33 (33)	-	33 (33)
	Orthogroups in Coll-153 and Coll-365	-	94 (94)	94 (94)
1	Strain-specific orthogroups	17 (85)	4 (53)	4 (56)
Sum of orthogroups		14,551 (15,430)	14,599 (15,377)	14,540 (15,323)
Unassigned genes		(196)	(55)	(64)

Comparison analysis revealed variations in pathogenicity-related categories in the three strains

To understand the variations in pathogenicity-related genes among the three strains, analyses of pathogenicity-related functional categories were performed. Six categories were used, effectors, carbohydrate active enzymes (CAZymes), secondary metabolism clusters, transcription factor (TF), and enzymes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and pathogen-host interaction (PHI) related genes. The results showed the three strains had very similar numbers of genes in various categories **(S2 Table)**. When combining these data and the ortholog data in a further analysis, notable differences were observed **(Table 6 and S3 Table)**. With regard to strain-specific orthogroups and unassigned genes in ortholog analysis, Coll-524 had significantly greater gene

329 numbers of effectors and PHI than Coll-153 and 365 **(Table 6)**. With regard to multi-copy
330 orthogroups with different gene numbers, Coll-524 had more genes belonging to the six functional
331 categories than Coll-153 and 365 **(S3 Table)**. For orthologs only found in two strains, there were no
332 notable differences between any set of two-strains with regard to gene numbers among the six
333 functional categories.

334

335 **Table 6. Functional category statistics of strain-specific orthogroups and unassigned genes of**
336 **the three *Colletotrichum scovillei* genomes.**

Categories	Strain-specific orthogroup			Unassigned genes		
	Coll-524	Coll-153	Coll-365	Coll-524	Coll-153	Coll-365
Effector	7	0	0	38	6	12
CAZyme	0	0	0	0	0	0
SMURF	3	3	2	0	0	0
TF	0	0	0	1	2	0
KEGG	6	6	6	3	2	2
PHI						
Increased virulence	0	0	0	8	1	1
Reduced virulence	7	0	0	51	6	15
Loss of pathogenicity	0	0	0	8	3	1
Lethal	2	0	0	3	0	1
Mixed results	39	36	39	8	3	1
Mixed functions with reduced	2	2	2	1	0	0

virulence						
Mixed functions without function in PHI	0	0	0	1	1	2
Unmatched	19	6	7	74	31	29
Sum	85	53	56	196	55	64

337

338 The gene variations of the six functional categories in scaffolds 17, 19, 20 and 22 in the three
339 strains are summarized in **Table 7**. Coll-153 and Coll-365 had remarkably fewer genes than Coll-524
340 in PHI (114 genes) and effector (33 genes) categories (**Table 7**). The ORFs of the 33 effector genes
341 were further analyzed using nucleotide BLAST (blastn) against 62 genomes of *Colletotrichum*
342 strains with genomes available in the NCBI database. As shown in **Fig 7**, two strains of *C. acutatum*
343 species complex (Coll-524 and *C. scovillei* TJNH1) carried all the 33 effector genes, while one strain
344 of this complex, *C. acutatum* 1, had 32 of the effectors. However, two strains with phylogenetics
345 closely related to Coll-524, *C. fioriniae* HC89 and HC91, only carried two of the 33 effector genes.
346 Among the 62 *Colletotrichum* strains, 27 did not carry any of the 33 effectors and they were mainly
347 in the species complexes graminicola, spaethianum, and gloeosporioides. Interestingly, none of the
348 members of the graminicola species complex carried any of the 33 effectors. Moreover, a total of 24
349 of the 33 effectors were only found in the acutatum complex, and not in the other complexes.

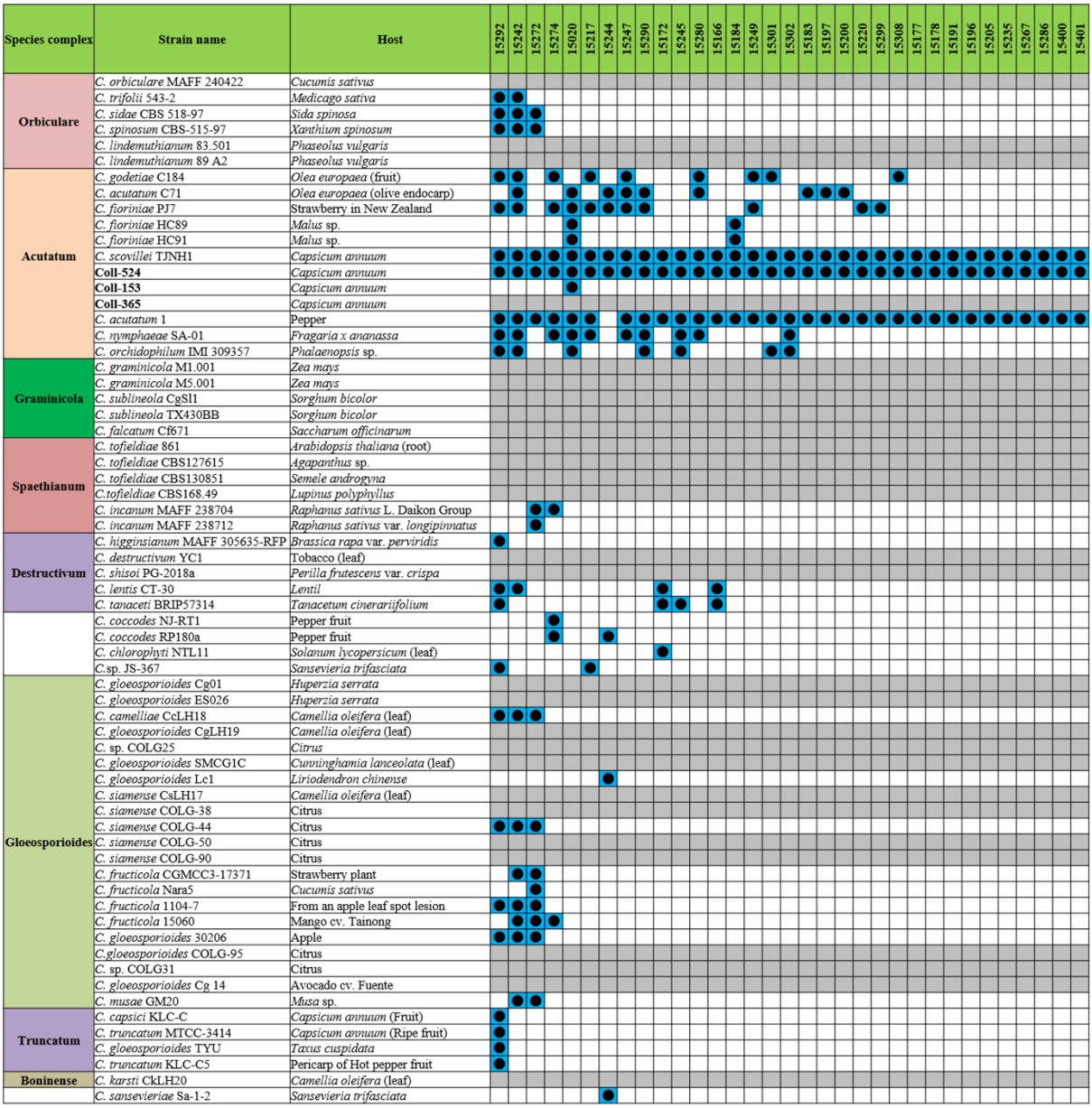


Figure 7. The distribution of the 33 effector genes in strains Coll-524, 153, and 365, and 62 *Colletotrichum* strains. The 33 effector genes were analyzed using nucleotide BLAST against the 62 genomes of *Colletotrichum* strains. Blue squares indicate the presence of genes by the blasting with coverage >90% and identity >60%. The gene was ordered from left to right by distribution frequency in the 62 genomes.

Table 7. Functional category statistics of genes located at scaffolds 17, 19, 20 and 22 of *Colletotrichum scovillei* strains Coll-524, Coll-153 and Coll-365

Categories	Coll-524/Coll-153/Coll365				Removed in Coll-153 and /or Coll-365
	S17	S19	S20	S22	S17+S19+S20+S22
Effector	20/19/19	11/0/0	19/0/1	2/0/0	33
CAZyme	8/8/7	0/0/0	0/0/0	0/0/0	1
SMURF	0/0/0	7/0/0	0/0/0	0/0/0	7
TF	6/4/4	1/0/0	3/0/0	1/1/0	7
KEGG	10/10/9	2/1/0	4/0/0	0/0/0	7
PHI					
Increased virulence	7/7/7	4/1/0	3/0/0	1/0/0	8
Reduced virulence	61/61/54	21/0/0	42/3/2	11/0/0	79
Loss of pathogenicity	8/7/6	6/0/0	5/1/0	0/0/0	12
Lethal	4/4/4	2/0/0	3/0/0	1/0/0	6
Mixed results	8/8/8	2/0/0	6/0/0	1/0/0	9

The variations of PHI genes among the three strains in scaffolds 17, 19, 20 and 22 were found to mainly occur in scaffolds 19, 20 and 22, especially genes required for fungal full virulence in the PHI database. A total of 79 genes with functions related to reducing virulence were found to be partially or fully removed in Coll-365 and 71 of them were not in Coll-153 (**Table 7**).

364 The variations of the three strains in KEGG pathways on scaffolds 17, 19, 20 and 22 were found
365 to mainly occur on scaffold 20, in which 100% of the KEGG genes were lost in Coll-153 and Coll-
366 365. The investigation of KEGG combined with ortholog analysis showed that gene number
367 variations within the three strains in scaffolds 17, 19, 20 and 22 were found in five metabolic
368 pathways. These pathways were metabolisms of drugs, purine, thiamine, phenylpropanoid, and N-
369 glycan biosynthesis. Coll-524 carried more genes than Coll-153 and/or Coll-365 in the five pathways
370 **(S3-7 Figs)**.
371

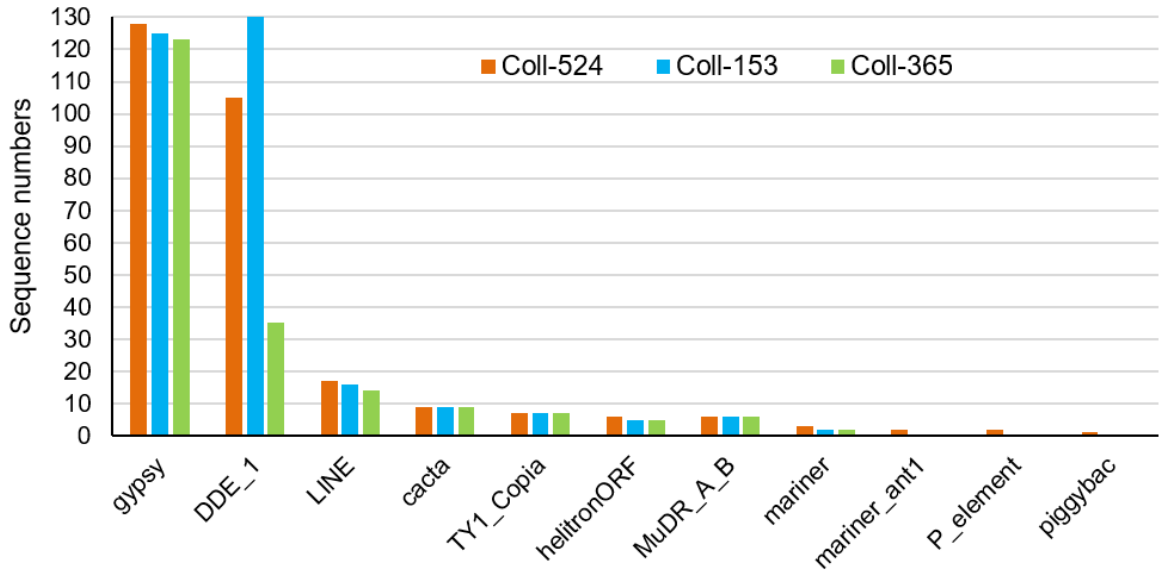
372 # Fot1 (TcMar-Fot1)-like transposon significantly decreased in strain
373 # Coll-365

374 Repeat sequences were analyzed with MIcroSAtellite (MISA) [34], TransposonPSI and
375 RepeatModeler [35]. Repeat sequence analysis with MISA showed the microsatellite compositions
376 were very similar in the three strains (**S4 Table**). TransposonPSI analysis revealed that 11 transposon
377 families were found in this species but the three strains carried different amounts of the transposon
378 families. Coll-524 contained all 11 families, but Coll-153 and Coll-365 only carried 9 families. In
379 addition, Coll-365 had significantly less DDE_1 domain-containing transposon than Coll-153 and
380 Coll-524 (**Fig 8**). Repeat sequence analysis with RepeatModeler showed that Coll-524, Coll-153 and
381 Coll-365 had 53, 67, and 44 repeat families, respectively. The DDE_1 domain identified by
382 TransposonPSI is from the TcMar-Fot1 family in the RepeatModeler database. The complete Coll-

383 Fot1 sequence was identified by combining sequences provided by the TransposonPSI and

384 RepeatModeler (**S8 Fig**).

385



386

387 **Figure 8. Distribution of repeat sequences among the three *Colletotrichum scovillei* strains.**

388

389 # Eight genes were selected for functional analysis

390 Orthogroup and ORF-V analysis revealed two sets of variations in gene composition of the three

391 strains. To obtain comprehensive analysis of variation, additional analysis was conducted by blastn

392 using Coll-524 genes against the genome sequences of Coll-153 and 365. A total of 219 genes from

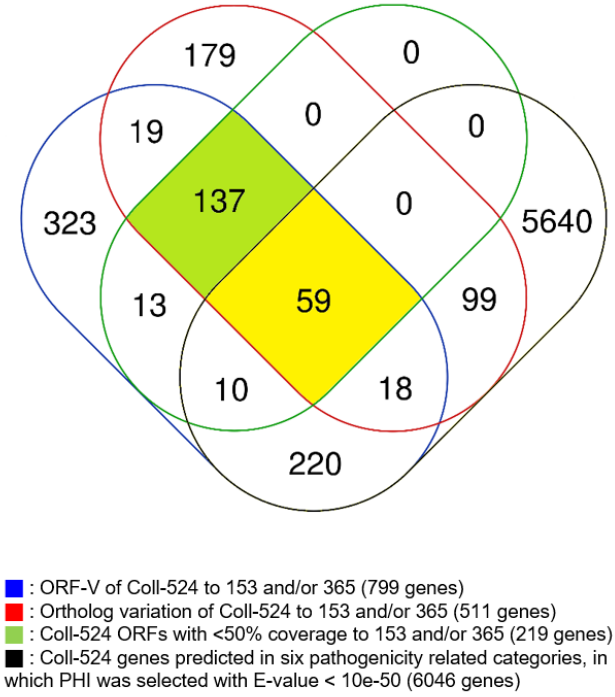
393 Coll-153 and/or 365 with less than 50% coverage to Coll-524 were identified and used with genes

394 selected from the other three groups, orthogroup analysis (511 genes), ORF-V analysis (799 genes)

395 and pathogenicity-related categories analysis, for further analysis (**Fig 9**).

396

A



B

Categories		Coll-153 Coll-365	Coll-365
Effector		26 (1)	8
CAZyme			
	GH31	1	0
	AA1	1 (1)	0
	GH17	0	1 (1)
TF		4	1 (1)
KEGG pathway		7	1
PKS cluster in scaffold 18 (including 2 effectors)		7	0
PHI (E-value < 10e-50)		2	1 (1)
Sum		47	12

Figure 9. Venn diagram of selected gene groups (A) and the clustered 59 genes distribution in six function categories (B). The four selected gene groups were indicated in the bottom of panel A. In panel B, the numbers of genes disappeared in both Coll-153 and Coll-365, and in Coll-365 only were presented. The number in parentheses indicates gene numbers selected for gene functional transformation assay.

The results are shown in **Fig 9A**. A total of 59 genes were found from the four groups. Among the 59 genes, 47 genes were lost in both Coll-153 and Coll-365, and 12 genes were only absent in Coll-365, including 8 effectors, one CAZyme, one TF, one KEGG pathway, and one PHI **(Fig 9B)**. Most of the 47 genes were located at scaffolds 19 and 20, while most of the 12 genes were located at the scaffolds 17 and 20 (**S5 Table**). A total of 137 genes were clustered from three gene groups

409 excluding the pathogenicity-related function categories (**Fig 9A**). These 137 genes contained 5
410 cytochrome P450 genes, 5 FAD binding domain-containing protein genes, one LysM domain-
411 containing protein gene (15215) as well as large numbers of unknown and hypothetical protein
412 genes. Eight genes, 5 from the 59 gene group and 3 from the 137 gene group, which were highly
413 expressed during infection, based on RNAseq data, were then selected for functional analyses. Five
414 genes from the 59 gene group are 11975, 14976, 14591, 15022, and 15019 (**Fig 9B**), and the 3 genes
415 from the 137 gene group are 15003, 15029 (**Fig 5**) and the LysM domain-containing protein gene
416 (15215). Genes encoding cytochrome P450 or FAD binding domain-containing protein were not
417 selected because they belong to large gene families in Coll-524. The eight genes were analyzed by
418 PCR to confirm their absence in the genome of Coll-153 and/or Coll-365 (**S9A Fig**).
419

420 The genes lost in a 34-kb fragment have a large effect on strain Coll-
421 365 morphology and pathogenicity

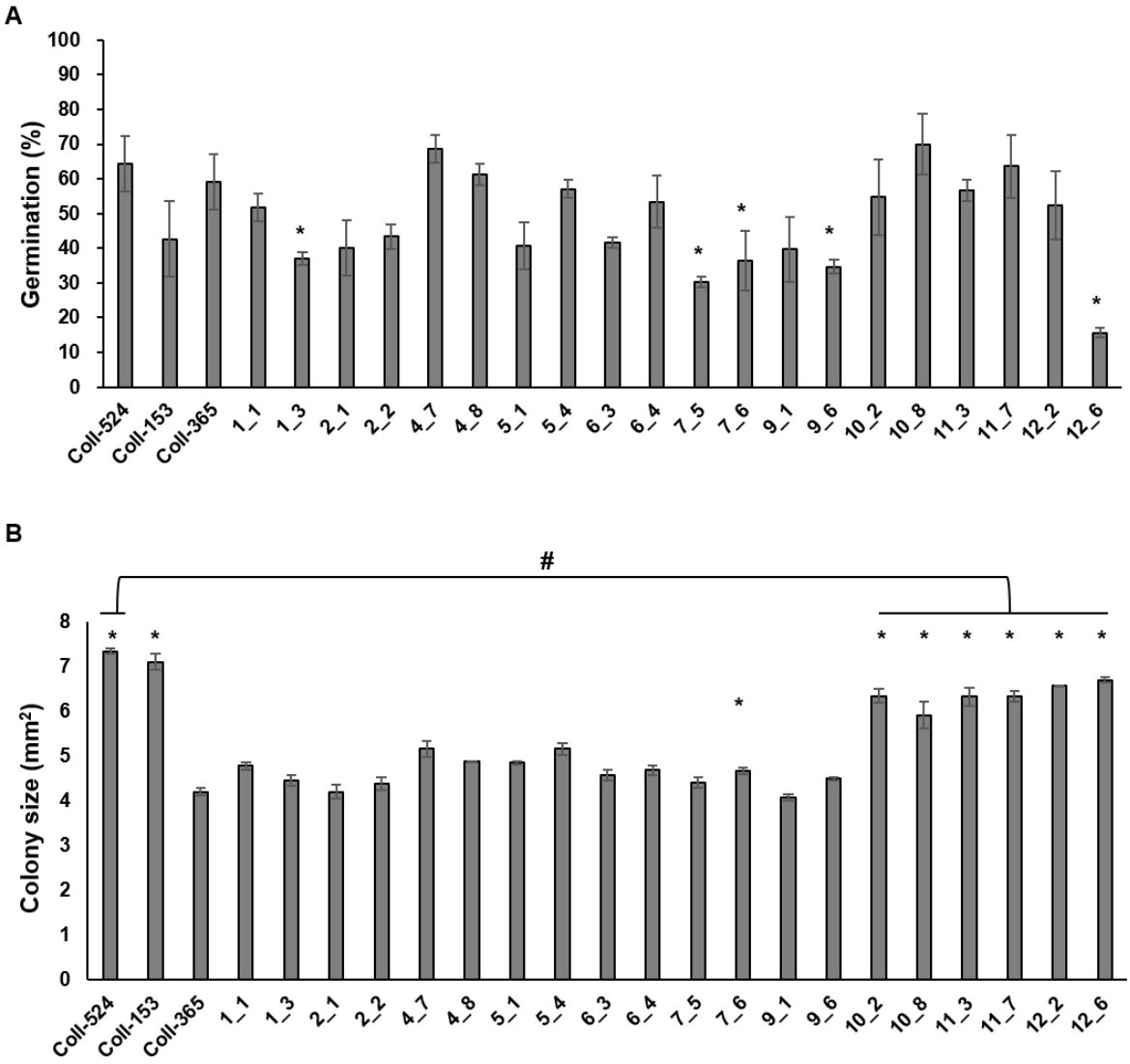
422 The eight selected genes were driven by their native promoters and transferred into Coll-365. Five
423 genes were located within a 34-kb fragment at scaffold 17 and two sets of two genes were closely
424 linked (15029 and 15022; 15003 and 15019). Therefore, for Coll-365 transgenic strains carrying two
425 closely linked genes, three or four of the four genes were generated (**Table 8**). All transgenic strains
426 were selected with PCR assays for the insertion fragment and/or further confirmed with RT-PCR for
427 their expression in Coll-365 (**S9B-C Fig**). The single spore-purified transgenic strains were used for

assays on spore germination, appressorium formation, growth, and pathogenicity assay on chili pepper fruits. Two independent transgenic strains of every gene transformation were used in all assays. Coll-524, Coll-153 and Coll-365 showed similar ability with regard to spore germination and appressorium formation (**Fig 10A**). Two independent transgenic strains of transformation IV (gene 15215, encoding a LysM domain containing protein; **Table 8**) showed a significantly lower germination rate than the wild-type Coll-365 in two independent experiments **(Fig 10A**). There were no significant differences in appressorium formation between the gene transformation strains and wild-type strain Coll-365. In the growth assay, Coll-365 had significantly slower growth than Coll-153 and Coll-524. Among the transgenic strains, two independent transgenic strains of transformation VIII (gene 15019), IX (carrying genes 15022, 15029 and 15019) and X (carrying genes 15022, 15029, 15019 and 15003) showed significantly enhanced growth on MS agar medium in comparison with their wild-type strain Coll-365, but still showed slightly slower growth than Coll-524 (**Fig 10B**). For pathogenicity, in a preliminary assay on chili pepper *Capsicum annuum* cv. Hero, Coll-365 was compared with transgenic strains carrying gene encoding effector, laccase, end-beta-1,3-glucasnase or LysM containing protein, and Coll-153 or Coll-524. Lesion size calculation showed that Coll-365 had significantly lower virulence than Coll-153 ($P = 0.023$) and Coll-524 ($P = 0.004$), and no notable difference from the tested transgenic strains. Further inoculation assays were conducted on *Capsicum annuum* cv. Fushimi-amanaga. Strains generated from transformation VII-X produced larger lesion sizes than wild-type Coll-365. The mean lesion sizes were increased between

447 49 and 400%. The lesion size increase level was similar for each transgenic strain in two independent

448 experiments **(Table 9)**. Transgenic strains of transformation VI and X were further inoculated on

449 *Capsicum annuum* cv. Groupzest and they caused significant lesion size increase as well (**S6 Table**).



450

451 **Figure 10. Spore germination (A) and mycelial growth of strains Coll-524, Coll-153 and Coll-**

452 **365 and the transgenic strains of Coll-365 (B).** The significant difference ($P < 0.5$) between Coll-

453 365 and other strains was indicated with a star when $P < 0.5$ appeared in two independent

454 experiments. The significant difference between Coll-524 and six transgenic strains of Coll-365 was

455 indicated with a hash in panel B.

456

457 **Table 8. Genetic transformation and phenotyping in strain Coll-365.**

Transformation	Gene(s)[a]	BLAST result	ORF length (bp)	Transgenic strains used for phenotyping	Functions
I	11975	hypothetical protein	683	1-1, 1-3	ND[b]
II	14591	probable laccase precursor	2264	2-1, 2-2	ND
III	14976	endo-β-1,3-glucanase	2202	6-3, 6-4	ND
IV	15215	LysM domain-containing protein	1599	7-5, 7-6	Inhibited germination
V	15022	bZIP transcription factor	871	4-7, 4-8	ND
VI	15029	C6 zinc finger domain-containing protein	1050	5-1, 5-4	ND
VII	15022 15029			9-1, 9-6	Enhanced virulence
VIII	15019	related to phospholipase a-2-activating protein	2425	10-2, 10-8	Enhanced growth and virulence
IX	15022 15029 15019			11-1, 11-7	Enhanced growth and virulence
X	15022 15029 15019			12-2, 12-6	Enhanced growth and virulence

	15003	WD domain containing protein	1387		

458 [a], Gene(s) of Coll-524 used to be transferred into Coll-365 in each transformation experiment.

459 [b], ND: non-detected; UD: undetermined.

460

461

462

463

464 **Table 9. Pathogenicity assay of gene transformation strains (transgenic) and wild-type strain**

465 **Coll-365 (WT) on fruits of *Capsicum annuum* cv. Fushimi-amanaga.**

Exp.	Transgenic strains	N[a]	Mean of lesion size (mm^2)		*P*	Lesion size increased (%)	
			WT	Transgenic		Individual Exp.	Means of Exp. 1 & 2
1	4_7	5	16.50	19.43	0.2467	18	2
	4_8	6	29.73	23.07	0.1429	-22	-27
	9_1	4	14.18	26.00	0.0102	83	284
	9_6	4	18.60	33.90	0.0916	82	71
	10_2	7	12.90	60.48	0.0000	369	274
	10_8	6	21.95	37.42	0.0814	70	49
	11_3	8	13.26	49.43	0.0000	273	221
	11_7	9	24.88	31.04	0.0364	25	54
	12_2	9	10.11	58.87	0.0000	482	400
	12_6	8	23.97	49.00	0.0000	104	123

	4_7	7	22.72	19.67	0.5799	-13	2
	4_8	6	33.73	23.20	0.1201	-31	-27
	9_1	5	5.63	32.88	0.0001	484	284
	9_6	5	19.00	30.40	0.0665	60	71
	10_2	7	23.60	65.87	0.0001	179	274
2	10_8	7	24.80	31.53	0.2146	27	49
	11_3	8	13.04	35.13	0.0000	169	221
	11_7	6	14.64	26.86	0.0269	83	54
	12_2	8	11.00	45.89	0.0000	317	400
	12_6	7	12.91	31.26	0.0015	142	123

[a], N indicates the numbers of fruits used in each inoculation.

Discussion

Colletotrichum species can cause great economic loss to various crops. Among the *Colletotrichum*, more than 30 species have been documented to cause chili anthracnose disease, constituting a major limitation to chili pepper production in tropical and subtropical regions [36]. In Taiwan and other Asia countries, *C. scovillei* attacks chili fruits [4], but its interactions with hosts at the molecular genetic level remain to be examined. In this study, we focused on genomic comparisons of the three *C. scovillei* strains and combined these with genetic approaches to identify genes involved in fungal growth and virulence. We have provided the genome sequences of the three strains with gene functional annotation for *C. scovillei*. We have setup a simple mathematical method to search for ORF variations between Coll-153 or Coll-365 and Coll-524 and successfully identified DNA fragments containing genes involved in the defects in growth and virulence of Coll-365. Moreover,

by genetic assay we have demonstrated four genes that have functions in germination, growth and/or virulence of *C. scovillei*.

Our data suggested that the three strains all belong to the acutatum complex and are members of *C. scovillei* because they were grouped in the same clade of *C. scovillei* CBS 126529, the holotype strain of *C. scovillei* [2]. The data were consistent with a previous study that showed that Coll-524 and Coll-153 are in the *C. scovillei* clade of acutatum species complex [4].

For closely related species, de-novo assembly by reference-guided or mapping through a well-sequenced species is efficient, and can often improve the completeness of the genome sequence [37]. Our data revealed the major variation of Coll-153 and Coll-365 was the result of sequence removal. Genetic variations can be caused by three different events, local nucleotide sequence changes, intragenomic rearrangement of DNA segments and the acquisition of a foreign DNA segment by horizontal gene transfer [38]. The transposable element is one of the major factors leading to variations in fungi. In *C. higginsianum*, two closely related strains carry large-scale rearrangements and strain-specific regions that are frequently associated with transposable elements [39]. In the two *C. higginsianum* strains, the gene-sparse regions are transposable element-dense regions that have more effector candidate genes, while gene-dense regions are transposable element-sparse regions harboring conserved genes [39]. *C. higginsianum* and other eukaryotic plant pathogens, such as *Phytophthora infestans* and *Leptosphaeria maculans* have been referred to as “two-speed genomes” as these genomes have a compartmentalized genome structure to protect housekeeping genes from

498 the deleterious effects of transposable elements and to provide rapid evolution of effector genes [40,
499 41]. In the genomes of Coll-524, 153 and 365, we did not find a relationship between the density of
500 transposable elements and effector genes or housekeeping genes. Coll-524 and *C. higginsianum* IMI
501 349063 have similar genome sizes but *C. higginsianum* IMI 349063 carries nearly three times the
502 number of transposable element genes to Coll-524 (S7 Table). Moreover, compared with *C.*
503 *higginsianum* IMI 349063, Coll-524 was found to be unlikely to have sequence structures like mini
504 chromosomes which harbor high density transposable elements with over 40% sequences encoding
505 transposable elements [42].

506 We designed a simple mathematical method to identify ORF-variations. Applying this method,
507 a 34-kb fragment (Fig 5) containing 14 genes that exist in the Coll-524 genome but are almost
508 completely lost in Coll-365 were identified. The loss of the 14 genes in Coll-365 is likely caused by
509 DNA deletion that might have resulted from DNA rearrangements. DNA rearrangement frequently
510 occurs in fungi via the parasexual cycle that leads to recombination and chromosome gain or loss
511 [43]. It is possible that Coll-524 gained part or all of these genes by DNA rearrangement during the
512 parasexual cycle in Coll-153 or Coll-365. Coll-365 and Coll-153 were collected 4 years earlier than
513 Coll-524 from the fields by the ACRDC-the World Vegetable Center [9]. Coll-153 and Coll-365 are
514 grouped in the CA1 pathotype and Coll-524 is a member of the CA2 pathotype. CA2 has higher
515 virulence than CA1 and has replaced CA1 to become dominant in Taiwan [11, 12]. Thus, Coll-524
516 might have arisen evolutionarily from the CA1 pathotype members such as Coll-153 and Coll-365 by

517 gaining genes horizontally or via gene arrangement. They might be evolutionarily developed from
518 different branches of *C. scovillei*.

519 Gene family expansions and contractions are related to the changing of host range and virulence
520 of plant pathogens [17, 44]. A set of 33 effectors was completely lost in Coll-365, but existed in two
521 closely related chili pepper pathogens *C. acutatum* strain 1 and *C. scovillei* strain TJNH1. This
522 suggests that the 33 effectors might play a role in the virulence of Coll-524 to chili pepper. However,
523 when one of the effectors was transferred to Coll-365, it did not promote the virulence of the
524 transgenic strains on chili pepper fruit, suggesting that multiple effectors may work together to affect
525 fungal virulence [45].

526 CAZyme genes are potentially involved in fungal growth and colonization of host tissues. Coll-
527 524, Coll-153 and Coll-365 have 661-663 CAZyme genes. In a study that compared the protein
528 family encoded by *Colletotrichum* species and others, CAZyme GH43 and S10 protease were found
529 to be duplicates with specific genes within the respective lineages in the wide host range pathogens
530 acutatum and gloeosporioides species complex [17]. Research on Magnaporthaceae specific clusters
531 showed CAZyme gene families may contribute to speciation [46]. Transgenic strains carrying endo-
532 beta-1,3-glucanase or laccase did not result in a gain of function of growth and virulence of chili
533 pepper. Endo-beta-1,3-glucanase is expected to play a key role in cell wall softening in ascomycetes.
534 The endo β-1,3-glucanase in *Schizosaccharomyces pombe, Candia albicans* and *Saccharomyces*
535 *cerevisiae* has been showed to be required for dissolution of the primary septum to allow cell

536 separation [47-49]. Fungal laccase can detoxify phenolic pollutants and has a role in detoxifying host
537 defense phenolic compounds, such as capsaicin in chili pepper [50]. Coll-524 has 5 copies of
538 laccase-related genes and two endo-beta-1,3-glucanase and one endo-beta-1,3(4)-glucanase. A single
539 laccase or endo-beta-1,3-glucase gene may not be able to contribute the function of growth or
540 virulence of Coll-365.

541 Eight genes were selected for gene transformation in Coll-365. Three genes 11975, 14591 and
542 14976 did not show a role in germination, growth and virulence of Coll-365 after transforming into
543 Coll-365. The effector gene 11975, which encodes a hypothetical protein, was expressed strongly
544 and specifically at the early infection stage, while laccase gene 14591 was expressed highly at all
545 infection stages but not in axenic cultures. Endo-beta-1,3-glucanase gene 14976 was highly
546 expressed in axenic cultures and at the late infection stage. Gene 15215 encoding a LysM domain-
547 containing protein was expressed at extremely high levels in axenic culture and at all infection
548 stages. The transformation of gene 15215 to Coll-365 showed no significant influence on fungal
549 growth and virulence but inhibited spore germination ability. LysM domain-containing protein can
550 bind to the fungal cell wall and protect from cell wall-degrading enzyme digestion and functions in
551 ecological niche competition and prevention from recognition by the host to trigger host immunity
552 [51]. The LysM domain-containing protein TAL6 of *Trichoderma atroviride* was shown to
553 specifically inhibit spore germination of *Trichoderma* species and was suggested to be involved in
554 the self-signaling process but not fungal-plant interactions [52]. *Colletotrichum* species can self-

555 inhibit spore germination by producing self-inhibitors [53]. Whether this LysM domain-containing
556 protein encoded by gene 15215 is involved in germination self-inhibition under high spore
557 concentration and pathogenicity in Coll-524 needs to be verified in the future.

558 Four genes (15022, 15029, 15019, 15003) located within the 34-kb fragment at scaffold 17 were
559 transferred into Coll-365 as individual genes or gene sets since they are closely linked. Gene 15022
560 had a high expression level in axenic cultures and all infection stages. The other three genes were
561 expressed in axenic cultures and the late infection stage, but genes 15019 and 15029 also had
562 relatively high expression in the cuticle infection stage. These four genes did not influence the
563 germination and appressorium formation of Coll-365. However, transgenic strains carrying one gene
564 15019, three genes 15022, 15029, and 15019, or four genes 15022, 15029, 15019, and 15003 have
565 similar colony sizes but are larger than the wild-type Coll-365, indicating that gene 15019
566 contributes to the growth enhancement of Coll-365. Gene 15022 is a bZIP transcription factor and
567 might be highly related to pathogenicity [54]. The opposite direction gene 15029, 1,313 bp away
568 from gene 15022 encodes a C6 zinc finger domain-containing protein. The C6 zinc finger proteins
569 are strictly fungal proteins and involved in fungal-host interactions [55]. Gene 15022 or 15029 alone
570 did not affect the virulence of Coll-524. However, transgenic strains carrying the two genes can
571 enhance the virulence of Coll-365, indicating that the co-existence of the two genes together might
572 contribute to the virulence enhancement of Coll-365.

573 Gene 15019 encodes a protein related to phospholipase A2-activation, while gene 15003 encodes
574 a WD domain-containing protein. Transgenic strains carrying one gene 15019, three genes 15022,
575 15029, and 15019, or four genes 15022, 15029, 15019, and 15003 enhance the virulence of Coll-365,
576 indicating gene 15019 might be the major contributor to the virulence enhancement of Coll-365.
577 Moreover, transgenic strains carrying genes 15022, 15029, and 15019, or genes 15022, 15029,
578 15019, and 15003 have a smaller *P* value than transgenic strains carrying gene 15019 only (Table 9),
579 which is consistent with the virulence contribution by gene 15022 and 15029 as mentioned above.
580 Gene 15019 encodes a protein related to phospholipase A2-activation. Phospholipase A2-activating
581 protein has been relatively intensively studied in humans and is involved in apoptosis and tumor
582 regression; however, there is only a single study related to its functional characterization in fungi
583 [56, 57]. The *Magnaporthe royzae Moplaa* gene encodes a phospholipase A2-activating protein and
584 the deletion of this gene results in the reduction of fungal growth and pathogenicity [56].
585 In conclusion, in this study, four genes were identified which function in germination, growth
586 and/or virulence of the chili pepper pathogen *C. scovillei.* Gene 15215, encoding a LysM domain-
587 containing protein, reduce the germination of Coll-365. Genes 15029 and 15022, encoding a C6 zinc
588 finger domain-containing protein and a bZIP transcription factor, together enhance the virulence of
589 Coll-365. Gene 15019, encoding a protein related to phospholipase A2-activation, enhances the
590 growth and virulence of Coll-365. In addition, the 34-kb fragment in scaffold 17 contributes a lot to
591 the defects in growth and virulence in Coll-365 because three genes 15019, 15022 and 15029 are

592 located on this fragment. Interestingly, we also identified 33 effectors lacking in Coll-153 and Coll-
593 365 which may be involved in the full virulence of Coll-524 on chili pepper. Future study will focus
594 on the 33 effectors and other effectors that may have led to Coll-524 becoming a dominant and
595 virulent pathogen.
596

597 Materials and Methods

598 Fungal strains and culture conditions

599 Three *Colletotrichum* strains Coll-524, Coll-153 and Coll-365 were obtained from the AVRDC-The
600 World Vegetable Center (Tainan, Taiwan) as mentioned in a previous study [9]. For sporulation,
601 fungal strains were cultured on MS agar medium (0.1% yeast extract, 0.1% peptone, 1% sucrose,
602 0.25% $MgSO_4 \cdot 7H_2O$, 0.27% KH_2PO_4 and 1.5% agar) for 6 days at 25°C with a 12 h light/dark cycle.
603 Spores were collected and used for general cultivation, assays of germination and appressorium
604 formation, morphological examination, growth, and pathogenicity assay.
605

606 Genomic DNA extraction and library preparation for sequencing

607 Spores were inoculated into a 500 mL flask containing 100 mL MS liquid medium (1×10^5
608 spores/ml) at 25°C with shaking at 200 rpm for 1 day. The young hyphae were collected by filtrating
609 with a layer of miracloth. The DNA was extracted from the young hyphae using the CTAB
610 extraction method [58]. DNA purity and concentration were determined by Nanodrop 2000 (Thermo

Fisher) and Qubit (Invitrogen) measurements. Libraries were generated from 2-5 μg of genomic DNA using the TruSeq 2 library preparation kit (Illumina, USA). Two types of libraries were prepared, short-fragment library (0.3 kb) for paired-end sequencing and long-fragment library (3 kb and 8 kb) for mate-pair sequencing. DNA fragmentation was made by shearing using an M220 Focused-Ultrasonicator (Covaris, USA) and the fragments with 296-300 bp, 3 kb or 8 kb were selected by gel-cutting. Without the 120 bp adaptor, the remaining 176-180 bp DNA fragments may have 20% overlap with 100 bp read length for paired-end sequencing. Long-fragment libraries, 3 kb and 8 kb, were prepared for Coll-524 only.

Genome sequencing and assembly

The Illumina HiSeq 2500 platform was used to sequence the genome of Coll-524 with 100 bp read length for paired-end and 3 kb mate pair libraries, and 125 bp read length for 8 kb mate pair library. The genomes of Coll-153 and Coll-365 were sequenced by Illumina Genome Analyzer with 150 bp read length. The raw reads were trimmed by Trimmomatic 0.39 to remove linkers and adaptors. Those refined reads of Coll-524 were further assembled by the ALLPATHS-LG *de novo* assembly program [30]. The genome sequences of Coll-524 have been deposited in NCBI with accession numbers JAESDM010000001-JAESDM010000054.

Genome assembly of Coll-153 and Coll-365 was achieved by using the "Map Reads to contigs" function in CLC Genomic Workbench v.9.5.1. All trimmed reads were used to map to the Coll-524 genome under default settings. The consensus mapped sequences were extracted and joined to form a scaffold by the "Extract Consensus Sequence" function. Unmapped reads were collected and assembled by using the "De Novo Assembly" function in CLC software. The assembled contigs with sizes larger than 5 kb were collected and counted as the genome sequences of Coll-153 or Coll-365. The genome sequences of Coll-153 and Coll-365 have been deposited in NCBI with accession numbers JAESDN010000001- JAESDN010000059 for Coll-153 and JAESDO010000001-JAESDO010000059 for Coll-365.The genome completeness of the three strains were assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO v.3.0) and the sordariomyceta_odb10 dataset [32, 59].

Multi-locus phylogenetic analysis

The DNA sequences of five marker genes (ACT1, CHS1, GAPDH, ITS and TUB2) were used for phylogenetic analyses [1]. The sequences of the five genes from the three strains, 170 other *Colletotrichum* strains and an outgroup species, *Monilochaetes infuscans*, were obtained and used in the assay. Among them, the five marker genes of 108 *Colletotrichum* species and *M. infuscans* were obtained from Q-bank database (https://qbank.eppo.int/fungi/) according to the GenBank accession numbers listed in a publication of Cannon *et al.* [1], while sequences from 62 *Colletotrichum* strains

were obtained by downloading whole genome sequences from NCBI FTP and then identifying the five gene sequences by using CLC Genomic Workbench v.9.5.1. The information for all the genes sequences is provided in **S1 Table**. All collected sequences were aligned by MAFFT v.7 online version [60] with default settings and then trimmed by trimAl 1.4.1 [61] with automated1 setting. Five marker genes were concatenated in each strain. Bayesian inference of all the concatenated sequences was analyzed by MrBayes 3.2.7 [62]. The Markov Chain Monte Carlo (MCMC) chains were set for 5,000,000 generations. Sample frequency was set for every 100 generations. Phylogenetic tree of the 174 strains were generated by FigTree v.1.4.3 [63].

Genome annotation

Gene annotation was conducted using MAKER (v.2.31.10) [31] pipeline with default settings (Augustus, PJ7, RNAseq reads). The genome of Coll-524 was used for gene annotation. The assembled contigs from Coll-524 RNAseq data was provided for EST evidence. The RNAseq data of Coll-524 included 10 sets of RNAseq data, which were infected purified cuticle layer, axenic 18 h hyphal culture, and 8 sets of infected chili pepper fruits at different infection time points (unpublished). The CDS and protein sequence of *Colletotrichum fioriniae* PJ7 (PRJNA244481) was provided to MAKER for the “EST evidence” and “Protein Homology Evidence” functions respectively. The gene annotation of Coll-153 and Coll-365 was conducted using the same method as

that for Coll-524. The CDS region of each gene was extracted from the genome sequence according to the GFF file which was originated from MAKER by GffRead [64].

Gene functional analysis, clustering and orthology analysis

The genes of Coll-524, Coll-153 and Coll-365 were searched with nr database by using BLASTx (BLAST 2.2.29+) to obtain the information about gene function. To find effector candidates, protein sequences of the three strains were analyzed by EffectorP 2.0 [65]. Carbohydrate active enzymes (CAZymes) candidates were searched by dbCAN2 meta server [66]. Candidate gene selection followed the recommendation setting provided by dbCAN2 meta server. Genes related to secondary metabolism were predicted using the Secondary Metabolite Unique Regions Finder (SMURF) database [67]. The Pathogen-Host Interactions database (PHI-base) v.4.10 was downloaded from PHI-base website. All genes of the three strains were used to blast against PHI-base using CLC Genomic Workbench [68]. Genes with E-values lower than 10 were selected according to the default setting of the PHIB-BLAST web server. The protein sequences of all annotated genes of Coll-153, Coll-365 and Coll-524 were used for ortholog analysis with the default settings of OrthoFinder [33]. After the Orthofinder analysis, single-copy orthogroups were generated directly. Other analyses of orthologues, including multi-copy orthogroups with the same gene numbers, multi-copy orthogroups with different gene numbers, orthogroups in every two strains, strain-specific orthologues, and unassigned genes, were further classified by EXCEL. These unassigned genes were considered as

685 "strain specific genes" together with genes in strain specific orthogroup. Genes of Coll-524 with
686 extra copy number in comparison with the other two strains in the orthogroup, and genes analyzed to
687 be Coll-524 strain specific genes were further considered as "Coll-524 extra genes".
688

689 # Identification of microsatellites, repeats and transposon elements

690 MISA 2.0 was used for microsatellite analysis [34] and transposon elements were analyzed by
691 Transposon PSI 2.2.26 (http://transposonpsi.sourceforge.net/). The transposon element sequences
692 were acquired according to the GFF files provided by Transposon PSI. These sequences were further
693 extracted to create a Transposon element database (TED). RepeatModeler was used for de novo
694 genome-wide repeat family analysis [35]. The sequences of TcMar-Fot1 were further blasted with
695 BLASTn to the NCBI database and the functional domains were analyzed using InterPro [69] to
696 determine if DDE was present [70].
697

698 # Identification of genome sequence variations and ORF variations of
699 # Coll-153 and Coll-365

700 During the mapping of the Coll-153 or Coll-365 genome to the Coll-524 genome, data containing the
701 variations between Coll-524 and the other 2 strains was generated with the "Extract Consensus
702 Sequence" function. The data included four datasets, Misc. difference (MD), Insertion, Deletion and
703 Removed, which are specified by CLC with default setting. Briefly, "Deletion" is for positions where

a gap is called while the reference has a non-gap; “Insertion” is for positions where a non-gap is called while the reference has a gap; “Removal” is for positions where no reads are mapped to the reference; “Misc. difference” is for every position where the consensus is different from the reference. The MD dataset consisted of single nucleotide polymorphism (SNP) and polymorphism with more than one nucleotide, which was named as multiple nucleotide polymorphism (MNP) in this study. The four datasets were combined to generate the genome sequence variations and used for genome comparisons to identify ORF variations between Coll-524 and the other two strains as described in **Fig 4**. To simplify the analysis of ORF variations, the Removed datasets of Coll-153 and Coll-365 were combined. Briefly, all the sequences which existed in Coll-524 but removed from Coll-153 or Coll-365 were extracted from Coll-524 genome to create a “removed sequence database” (RSD).

ORF variation analysis was based on the calculations of physical position overlapping between each ORF and the four datasets of variations (MD-MNP, Insertion, Deletion and Removed). The GFF file generated by MAKER could provide the physical positions of mRNA, exon, 5′UTR, CDS, 3′UTR for each gene. A gene usually contains multiple CDSs in eukaryotes and all CDSs together within the gene is the complete coding sequence to translate a protein encoded by the gene. To analyze ORF variation of Coll-153 and Coll-365, the ORF position of each Coll-524 gene was generated by setting the first nucleotide of the first CDS as the beginning position and the last nucleotide of the last CDS as the end position.

723 To analyze the variation of each ORF between Coll-524 and the other two strains, firstly, all
724 scaffolds of Coll-524 were added together one by one to form a single sequence. The physical
725 positions of each ORF were modified to fit into the single sequence. The four datasets of variations
726 (MD, Insertion, Deletion and Removed) generated by CLC were tagged with physical positions in
727 each scaffold of Coll-524. Therefore, the second step was to edit the physical positions of each
728 variation according to the single genome sequence of Coll-524. The position differences of each
729 variation to the ORF were compared and calculated as indicated in **Fig 4C-G**. The physical position
730 overlapping situation of each ORF and each variation was identified by the calculation formula listed
731 in **Fig 4G**. An ORF overlapped with a variation indicated that this ORF varied between Coll-524 and
732 the other strain carrying this variation. ORFs overlapped with variation were further called "ORF-
733 V".
734

735 Screening of genes potentially involved in the variations of the three
736 strains for functional verification in Coll-365

737 To select genes for genetic analysis to verify their functions on the defects of growth and virulence
738 of Coll-365, four selection criteria were set for gene selection. First were ORFs predicted to be ORF-
739 V. Second were "Coll-524 extra genes" according to the result of OrthoFinder analysis. Third were
740 genes analyzed to belonging to effector, CAZyme, TF, SMURF, enzyme or PHI (E-value < 10e-50).
741 Finally, were ORFs with <50% coverage to Coll-153 and/or 365 according to the blastn results.

Genes with the four criteria were analyzed by Venn Diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). Pathogen-host interaction related genes were selected with an E-value less than 10e-50. Genes with an E-value less than 10e-50 but greater than 10e-100 were considered to have almost identical sequences to the database according to the manual for CLC Genomics Workbench 9.5.

Plasmid construction

Plasmids BsHR and BsBR were constructed and used in this research to clone the selected target gene(s) for transformation into Coll-365. pBsHR was constructed by cloning hygromycin resistance cassette (HygR) of pBHt2 [71] into pBluescript SK(+) [72] with PCR amplification, restriction enzyme digestion (EcoRV and SwaI, designed in the primers), and DNA ligation. Plasmid BsBR was modified from pBsHR by replacing (HygR) with bleomycin resistance gene (BleoR) of pAN8-1 [73]. The BleoR cassette was PCR amplified with specific primers containing EcoRV and SwaI restriction sites and then cloned into EcoRV- and SwaI-digested pBsHR. To clone the target genes from Coll-524, a target gene containing an approximate 1-kb promoter and 0.5-kb terminator was amplified with a high-fidelity DNA polymerase KOD-plus-Neo (TOYOBO, Osaka, Japan) and ligated to EcoRV or SwaI digested and Shrimp Alkaline Phosphatase (rSAP) treated pBsHR. If double gene-transformation was needed, a second target gene was cloned in to pBsBR for protoplast transformation and phleomycin was used as the selection antibiotic [74]. For gene-transformation

761 VII, the two genes were closely linked and were amplified together with one primer set by PCR and
762 transferred to Coll-365 by single transformation. Gene-transformation IX was completed with two
763 transformations by transferring gene 15019 into a transgenic strain of gene-transformation VII. The
764 gene-transformation X was conducted by transferring a PCR fragment containing 15019 and 15003
765 into a transgenic strain of gene-transformation VII. The gene locations of 15019, 15003, 15022 and
766 15029 are indicated in **Fig 5**.
767

768 # Protoplast transformation

769 Polyethylene glycol (PEG)/Ca^{+2}-mediated protoplast transformation was used to transfer the target
770 gene to Coll-365. Spores collected from a 6-day MS agar medium were inoculated into a 500-ml
771 flask containing 400 ml MS liquid medium and a stirrer bar, and cultured for 16 h under 25°C at low
772 speed to prevent the attachment of spores to the flask surface. The young hyphae were collected by
773 filtering through a layer of miracloth and washed with 100 ml sterilized water and 100 ml wash
774 buffer (1 M NaCl and 10 mM $CaCl_2$) under suction. The washed hyphae (250 mg) were then
775 resuspended in a 50-ml flask containing 10 ml osmotic buffer (10 mM Na_2HPO_4, pH 5.8, 20 mM
776 $CaCl_2$, and 1.2 M NaCl) with 90 mg lysing enzymes (Sigma, L1412), and incubated in an orbital
777 shaker with 150 rpm at 30°C for 6 h. The undigested hyphae were removed by filtering through
778 miracloth and the protoplasts were collected by centrifugation at 1500 g, 4°C for 10 min. The
779 protoplasts were resuspended by adding 100 µl mixture of four parts of STC buffer (1.2 M sorbitol,

10 mM Tris-HCl, pH 7.5, and 10 mM $CaCl_2$) and one part of PEG [50% (w/v) polyethylene glycol 3350, 10 mM Tris-HCl, pH 7.5, and 10 mM $CaCl_2$]. For transformation, 20 µl of 20 µg plasmid DNA was added to a tube containing 100 µl protoplast suspension with gentle mixing and the mixture was placed in ice for 20 min. PEG was added into the DNA-protoplast solution four times with different volumes. The adding steps were performed by following a previous description [75] with slight modifications. Briefly, 20, 80, 300 and 600 µl PEG were added into protoplast suspension step by step. After each addition, the mixture was gently mixed and left to stand for 3 min, and then left to stand a further 20 min after the last addition at room temperature. After the 20 min incubation, 3 ml regenerate buffer [4 mM $Ca(NO_3)_2 \cdot 4H_20$, 1.5 mM KH_2PO_4, 1 mM $MgSO_4 \cdot 7H_2O$, 2.5 mM NaCl, 60 mM glucose, and 1 M sucrose] was added into the PEG-DNA-protoplast mixture and cultured at 25°C for 16 h with 100 rpm shaking. The protoplasts were collected with centrifugation at 25°C, 1800 g for 10 min. A total of 300 µl regenerated protoplasts were obtained and then evenly spread on three regeneration agar medium plates. After incubation for 6-8 days at 25°C, the transformants were isolated, single spore purification and PCR verification for the target gene insertion.

PCR analysis for gene loss in Coll-153 and Coll-365 and verification of gene transformation in Coll-365 transformants

798 Eight genes were selected for functional assay in Coll-365 as listed in **Table 8**. PCR assays were
799 used to confirm the loss of the eight genes in Coll-365 using specific primer sets as listed in **S8**
800 **Table**. Genomic DNA was extracted as described above and tubulin gene (Accession
801 number:MW073123) was used as the control for PCR assay. To verify gene transformation of
802 transgenic Coll-365 strains, regular PCR and RT-PCR assays were performed. Regular PCR using
803 genomic DNA as template to amplify the transgenic gene was used for all transgenic strains. For
804 transgenic strains carrying more than one transgene, RT-PCR assays were conducted in addition to
805 the regular PCR. Primers used in theses assays are listed in **Table S6**. For RNA extraction, spores
806 were inoculated into a 50 mL flask containing 20 mL MS liquid medium (1 x 10^5 spores/ml) at 25°C
807 with shaking at 150 rpm for 2 days. The hyphae were collected by filtrating with a layer of miracloth.
808 The RNA was extracted with Trizol reagent (Invitrogen). RNA purity and concentration were
809 determined and cDNA was synthesized using 5 μg RNA and the MMLV reverse transcription kit
810 (Invitrogen). PCR was performed using 20 μl reaction volume contained 1 × PCR Buffer, 0.2 mM
811 dNTP, 0.4 μM primers, 1 U Blend Taq plus (TOYOBO), and 1 μl of template DNA. PCR reaction
812 was performed as follows: 94°C for 2 min and 25 cycles of 94°C for 30 sec, 60°C for 30 sec, and
813 72°C for 3 min. The PCR products were analyzed by agarose gel electrophoresis.
814

815 # Fungal growth, spore germination and appressorium formation assays

816 Fungal growth was assayed by inoculating spore suspension on the center of three different agar
817 media. Spore germination and appressorium formation were detected in 96-well plates. Briefly,
818 spores were collected from MS agar medium and the concentration was adjusted to 2×10^5 spores/ml
819 by sterilized water. For growth assay, a 5-µl, spore suspension containing 500 spores was dropped on
820 the center of MS agar medium plate and incubated for 6 day at 25°C with a 12 h light/dark light
821 cycle. Colony sizes were measured by using ImageJ 1.53a [76]. To detect germination and
822 appressorium formation, 80 µl of spore suspension was added into each well of a 96-well plate and
823 incubated at 25°C with 12 h light/dark light cycle for 8 h. To count the numbers of germinated spores
824 and appressorium, the 96-well plate was placed upside down and examined using a light microscope.
825 All the experiments mentioned above were conducted at least two times with three replicates for
826 each strain at each time. The statistical significance of the data was determined by One-way ANOVA
827 at $P < 0.05$ using Statistical Package for the Social Sciences software, version 20 (IBM SPSS
828 software).
829

830 # Pathogenicity assay

831 Pepper fruits were used for pathogenicity assay, including *Capsicum annuum* cv. Hero, Fushimi-
832 amanaga and Groupzest. The fresh harvested fruits were washed with water and surface sterilized
833 with 0.5% bleach and left to dry overnight. Fungal spores were collected from MS agar medium and
834 the concentration was adjusted to 2×10^5 spores/ml. Spore suspension was dropped (5 µl/drop) on

the fruit surface with dual inoculation, in which the wild type strain was inoculated on one side of the fruit and a transgenic strain was inoculated on the other site of the same fruit. The fruits were incubated in a growth chamber at 25°C with 12 h light/dark light cycle. At least three fruits having 3-5 inoculation sites were used for each strain in each inoculation. Lesion sizes were measured with ImageJ and the statistical analysis was performed by paired *t*-test at $P < 0.05$ using Statistical Package for the Social Sciences software, version 20 (IBM SPSS software).

AUTHOR CONTRIBUTIONS

DKH, SCC, MHL, and MCS contributed to the design of the experiments. SCC designed and prepared the DNA and RNA materials for sequencing. DKH performed all bioinformatic analyses and gene transformation as well as functional characterizations. YTC and CYC assisted in sequence assembly and annotation. MYL performed NGS sequencing. MHL and MCS supervised the experiments. DKH, MHL, and MCS wrote the manuscript.

FUNDING

This research was funded by the Center for Sustainability Science of Academia Sinica in Taiwan (grant number AS-106-SS-A03), the Ministry of Science and Technology in Taiwan (grant number MOST 104-2313-B-005 -025 -MY3), and the Advanced Plant Biotechnology Center from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan.

ACKNOWLEDGMENTS

The authors would like to thank Academia Sinica High throughput Genomics Core for NGS sequencing services, and Ms. Miranda Loney for English editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at:

References

1. Cannon PF, Damm U, Johnston PR, Weir BS. Colletotrichum - current status and future directions. Stud Mycol. 2012;73(1):181-213. Epub 2012/11/09. doi: 10.3114/sim0014. PubMed PMID: 23136460; PubMed Central PMCID: PMCPMC3458418.
2. Damm U, Cannon PF, Woudenberg JH, Crous PW. The Colletotrichum acutatum species complex. Stud Mycol. 2012;73(1):37-113. Epub 2012/11/09. doi: 10.3114/sim0010. PubMed PMID: 23136458; PubMed Central PMCID: PMCPMC3458416.
3. Weir BS, Johnston PR, Damm U. The Colletotrichum gloeosporioides species complex. Stud Mycol. 2012;73(1):115-80. Epub 2012/11/09. doi: 10.3114/sim0011. PubMed PMID: 23136459; PubMed Central PMCID: PMCPMC3458417.
4. De Silva DD, Ades PK, Taylor PWJ. Pathogenicity of Colletotrichum species causing anthracnose of Capsicum in Asia. Plant Pathology. 2021. doi: 10.1111/ppa.13351.
5. Damm U, Sato T, Alizadeh A, Groenewald JZ, Crous PW. The Colletotrichum dracaenophilum, C. magnum and C. orchidearum species complexes. Stud Mycol. 2019;92:1-46. Epub 2018/07/13. doi: 10.1016/j.simyco.2018.04.001. PubMed PMID: 29997400; PubMed Central PMCID: PMCPMC6030544.
6. Damm U, Cannon PF, Woudenberg JH, Johnston PR, Weir BS, Tan YP, et al. The Colletotrichum boninense species complex. Stud Mycol. 2012;73(1):1-36. Epub 2012/11/09. doi: 10.3114/sim0002. PubMed PMID: 23136457; PubMed Central PMCID: PMCPMC3458415.
7. Mongkolporn O, Taylor PWJ. Chili anthracnose: Colletotrichum taxonomy and pathogenicity. Plant Pathology. 2018;67(6):1255-63. doi: 10.1111/ppa.12850.
8. de Silva DD, Groenewald JZ, Crous PW, Ades PK, Nasruddin A, Mongkolporn O, et al. Identification, prevalence and pathogenicity of Colletotrichum species causing anthracnose of Capsicum annuum in Asia. IMA Fungus. 2019;10:8. Epub 2020/05/02. doi: 10.1186/s43008-019-0001-y. PubMed PMID: 32355609; PubMed Central PMCID: PMCPMC7184891.
9. Liao C-Y, Chen M-Y, Chen Y-K, Wang T-C, Sheu Z-M, Kuo K-C, et al. Characterization of three Colletotrichum acutatum isolates from Capsicum spp. European Journal of Plant Pathology.

2012;133(3):599-608. doi: 10.1007/s10658-011-9935-7.
10. De Silva DD, Crous PW, Ades PK, Hyde KD, Taylor PWJ. Life styles of Colletotrichum species and implications for plant biosecurity. Fungal Biology Reviews. 2017;31(3):155-68. doi: 10.1016/j.fbr.2017.05.001.
11. Sheu Z-M, Chiu M-H, Chang R-J, Wang J-F, Kenyon L. Pathotype dynamics and genetic variability of Colletotrichum scovillei causing pepper anthracnose in Taiwan. World Vegetable Center, Taiwan, ROC. 2019;Poster Presentation 3.
12. Kumar S, Shieh HC, Lin SW, Schafleitner R, Kenyon L, Ramasamy S, et al. Peppers (Capsicum spp.): Domestication and Breeding for Global Use. 2018.
13. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010;464(7287):367-73. Epub 2010/03/20. doi: 10.1038/nature08850. PubMed PMID: 20237561; PubMed Central PMCID: PMCPMC3048781.
14. Teixeira MM, de Almeida LG, Kubitschek-Barreira P, Alves FL, Kioshima ES, Abadio AK, et al. Comparative genomics of the major fungal agents of human and animal Sporotrichosis: Sporothrix schenckii and Sporothrix brasiliensis. BMC Genomics. 2014;15:943. Epub 2014/10/30. doi: 10.1186/1471-2164-15-943. PubMed PMID: 25351875; PubMed Central PMCID: PMCPMC4226871.
15. Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, et al. Comparative genomics of a plant-pathogenic fungus, Pyrenophora tritici-repentis, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. G3 (Bethesda). 2013;3(1):41-63. Epub 2013/01/15. doi: 10.1534/g3.112.004044. PubMed PMID: 23316438; PubMed Central PMCID: PMCPMC3538342.
16. Buiate EAS, Xavier KV, Moore N, Torres MF, Farman ML, Schardl CL, et al. A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species Colletotrichum graminicola and Colletotrichum sublineola. BMC Genomics. 2017;18(1):67. Epub 2017/01/12. doi: 10.1186/s12864-016-3457-9. PubMed PMID: 28073340; PubMed Central PMCID: PMCPMC5225507.
17. Gan P, Narusaka M, Kumakura N, Tsushima A, Takano Y, Narusaka Y, 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. 2016;8(5):1467-81. Epub 2016/05/18. doi: 10.1093/gbe/evw089. PubMed PMID: 27189990; PubMed Central PMCID: PMCPMC4898803.
18. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BP, et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog. 2011;7(7):e1002137. Epub 2011/08/11. doi: 10.1371/journal.ppat.1002137. PubMed PMID: 21829347; PubMed Central PMCID: PMCPMC3145793.
19. Syme RA, Tan KC, Rybak K, Friesen TL, McDonald BA, Oliver RP, et al. Pan-Parastagonospora Comparative Genome Analysis-Effector Prediction and Genome Evolution. Genome Biol Evol.

2018;10(9):2443-57. Epub 2018/09/06. doi: 10.1093/gbe/evy192. PubMed PMID: 30184068; PubMed Central PMCID: PMCPMC6152946.

20. Valero-Jimenez CA, Veloso J, Staats M, van Kan JAL. Comparative genomics of plant pathogenic Botrytis species with distinct host specificity. BMC Genomics. 2019;20(1):203. Epub 2019/03/15. doi: 10.1186/s12864-019-5580-x. PubMed PMID: 30866801; PubMed Central PMCID: PMCPMC6417074.

21. van Dam P, de Sain M, Ter Horst A, van der Gragt M, Rep M. Use of Comparative Genomics-Based Markers for Discrimination of Host Specificity in Fusarium oxysporum. Appl Environ Microbiol. 2018;84(1). Epub 2017/10/17. doi: 10.1128/AEM.01868-17. PubMed PMID: 29030446; PubMed Central PMCID: PMCPMC5734036.

22. Lelwala RV, Korhonen PK, Young ND, Scott JB, Ades PK, Gasser RB, et al. Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in Colletotrichum tanaceti. PLoS One. 2019;14(5):e0212248. Epub 2019/06/01. doi: 10.1371/journal.pone.0212248. PubMed PMID: 31150449; PubMed Central PMCID: PMCPMC6544218.

23. Liu X, Li B, Yang Y, Cai J, Shi T, Zheng X, et al. Pathogenic Adaptations Revealed by Comparative Genome Analyses of Two Colletotrichum spp., the Causal Agent of Anthracnose in Rubber Tree. Front Microbiol. 2020;11:1484. Epub 2020/08/15. doi: 10.3389/fmicb.2020.01484. PubMed PMID: 32793128; PubMed Central PMCID: PMCPMC7385191.

24. Lebrigand K, He LD, Thakur N, Arguel MJ, Polanowska J, Henrissat B, et al. Comparative Genomic Analysis of Drechmeria coniospora Reveals Core and Specific Genetic Requirements for Fungal Endoparasitism of Nematodes. PLoS Genet. 2016;12(5):e1006017. Epub 2016/05/07. doi: 10.1371/journal.pgen.1006017. PubMed PMID: 27153332; PubMed Central PMCID: PMCPMC4859500.

25. Chen JY, Liu C, Gui YJ, Si KW, Zhang DD, Wang J, et al. Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in Verticillium dahliae and evidence of horizontal gene transfer from Fusarium. New Phytol. 2018;217(2):756-70. Epub 2017/10/31. doi: 10.1111/nph.14861. PubMed PMID: 29084346; PubMed Central PMCID: PMCPMC5765495.

26. Zhang DD, Wang J, Wang D, Kong ZQ, Zhou L, Zhang GY, et al. Population genomics demystifies the defoliation phenotype in the plant pathogen Verticillium dahliae. New Phytol. 2019;222(2):1012-29. Epub 2019/01/05. doi: 10.1111/nph.15672. PubMed PMID: 30609067; PubMed Central PMCID: PMCPMC6594092.

27. Niehaus EM, Kim HK, Munsterkotter M, Janevska S, Arndt B, Kalinina SA, et al. Comparative genomics of geographically distant Fusarium fujikuroi isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. PLoS Pathog. 2017;13(10):e1006670. Epub 2017/10/27. doi: 10.1371/journal.ppat.1006670. PubMed PMID: 29073267; PubMed Central PMCID: PMCPMC5675463.

28. Ingram TW, Oh Y, Adhikari TB, Louws FJ, Dean RA. Comparative Genome Analyses of 18 Verticillium dahliae Tomato Isolates Reveals Phylogenetic and Race Specific Signatures. Front

966 Microbiol. 2020;11:573755. Epub 2020/12/18. doi: 10.3389/fmicb.2020.573755. PubMed PMID:
967 33329432; PubMed Central PMCID: PMCPMC7734093.
968 29. Liao CY, Chen MY, Chen YK, Kuo KC, Chung KR, Lee MH. Formation of highly branched hyphae
969 by Colletotrichum acutatum within the fruit cuticles of Capsicum spp. Plant Pathology.
970 2012;61(2):262-70. doi: 10.1111/j.1365-3059.2011.02523.x.
971 30. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, et al. ALLPATHS: de
972 novo assembly of whole-genome shotgun microreads. Genome Res. 2008;18(5):810-20. Epub
973 2008/03/15. doi: 10.1101/gr.7337908. PubMed PMID: 18340039; PubMed Central PMCID:
974 PMCPMC2336810.
975 31. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool
976 for second-generation genome projects. BMC Bioinformatics. 2011;12:491. Epub 2011/12/24. doi:
977 10.1186/1471-2105-12-491. PubMed PMID: 22192575; PubMed Central PMCID: PMCPMC3280279.
978 32. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO
979 Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol.
980 2018;35(3):543-8. Epub 2017/12/09. doi: 10.1093/molbev/msx319. PubMed PMID: 29220515;
981 PubMed Central PMCID: PMCPMC5850278.
982 33. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics.
983 Genome Biol. 2019;20(1):238. Epub 2019/11/16. doi: 10.1186/s13059-019-1832-y. PubMed PMID:
984 31727128; PubMed Central PMCID: PMCPMC6857279.
985 34. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite
986 prediction. Bioinformatics. 2017;33(16):2583-5. Epub 2017/04/12. doi:
987 10.1093/bioinformatics/btx198. PubMed PMID: 28398459; PubMed Central PMCID:
988 PMCPMC5870701.
989 35. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for
990 automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A.
991 2020;117(17):9451-7. Epub 2020/04/18. doi: 10.1073/pnas.1921046117. PubMed PMID: 32300014;
992 PubMed Central PMCID: PMCPMC7196820.
993 36. Saxena A, Raghuwanshi R, Gupta VK, Singh HB. Chilli Anthracnose: The Epidemiology and
994 Management. Front Microbiol. 2016;7:1527. Epub 2016/10/18. doi: 10.3389/fmicb.2016.01527.
995 PubMed PMID: 27746765; PubMed Central PMCID: PMCPMC5044472.
996 37. Lischer HEL, Shimizu KK. Reference-guided de novo assembly approach improves genome
997 reconstruction for related species. BMC Bioinformatics. 2017;18(1):474. Epub 2017/11/12. doi:
998 10.1186/s12859-017-1911-6. PubMed PMID: 29126390; PubMed Central PMCID:
999 PMCPMC5681816.
1000 38. Arber W. Genetic engineering compared to natural genetic variations. N Biotechnol.
1001 2010;27(5):517-21. Epub 2010/05/18. doi: 10.1016/j.nbt.2010.05.007. PubMed PMID: 20472106.
1002 39. Tsushima A, Gan P, Kumakura N, Narusaka M, Takano Y, Narusaka Y, et al. Genomic Plasticity
1003 Mediated by Transposable Elements in the Plant Pathogenic Fungus Colletotrichum higginsianum.

Genome Biol Evol. 2019;11(5):1487-500. Epub 2019/04/28. doi: 10.1093/gbe/evz087. PubMed PMID: 31028389; PubMed Central PMCID: PMCPMC6535813.

40. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 2016;26(8):1091-100. Epub 2016/06/22. doi: 10.1101/gr.204974.116. PubMed PMID: 27325116; PubMed Central PMCID: PMCPMC4971763.

41. Grandaubert J, Lowe RGT, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, et al. Transposable element-assisted evolution and adaptation to host plant within the Leptosphaeria maculans-Leptosphaeria biglobosa species complex of fungal pathogens. BMC genomics. 2014;15:891-. doi: 10.1186/1471-2164-15-891. PubMed PMID: 25306241.

42. Plaumann PL, Koch C. The Many Questions about Mini Chromosomes in Colletotrichum spp. Plants (Basel). 2020;9(5). Epub 2020/05/23. doi: 10.3390/plants9050641. PubMed PMID: 32438596; PubMed Central PMCID: PMCPMC7284448.

43. Forche A, Alby K, Schaefer D, Johnson AD, Berman J, Bennett RJ. The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. PLoS Biol. 2008;6(5):e110. Epub 2008/05/09. doi: 10.1371/journal.pbio.0060110. PubMed PMID: 18462019; PubMed Central PMCID: PMCPMC2365976.

44. Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, et al. Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. BMC Genomics. 2016;17:555. Epub 2016/08/09. doi: 10.1186/s12864-016-2917-6. PubMed PMID: 27496087; PubMed Central PMCID: PMCPMC4974774.

45. Arroyo-Velez N, Gonzalez-Fuente M, Peeters N, Lauber E, Noel LD. From effectors to effectomes: Are functional studies of individual effectors enough to decipher plant pathogen infectious strategies? PLoS Pathog. 2020;16(12):e1009059. Epub 2020/12/04. doi: 10.1371/journal.ppat.1009059. PubMed PMID: 33270803; PubMed Central PMCID: PMCPMC7714205.

46. Okagaki LH, Sailsbery JK, Eyre AW, Dean RA. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. BMC Genomics. 2016;17:135. Epub 2016/02/26. doi: 10.1186/s12864-016-2491-y. PubMed PMID: 26911875; PubMed Central PMCID: PMCPMC4766678.

47. Martin-Cuadrado AB, Duenas E, Sipiczki M, Vazquez de Aldana CR, del Rey F. The endo-beta-1,3-glucanase eng1p is required for dissolution of the primary septum during cell separation in Schizosaccharomyces pombe. J Cell Sci. 2003;116(Pt 9):1689-98. Epub 2003/04/01. doi: 10.1242/jcs.00377. PubMed PMID: 12665550.

48. Esteban PF, Rios I, Garcia R, Duenas E, Pla J, Sanchez M, et al. Characterization of the CaENG1 gene encoding an endo-1,3-beta-glucanase involved in cell separation in Candida albicans. Curr Microbiol. 2005;51(6):385-92. Epub 2005/12/06. doi: 10.1007/s00284-005-0066-2. PubMed PMID: 16328626.

1042 49. Baladron V, Ufano S, Duenas E, Martin-Cuadrado AB, del Rey F, Vazquez de Aldana CR. Eng1p,
1043 an endo-1,3-beta-glucanase localized at the daughter side of the septum, is involved in cell
1044 separation in Saccharomyces cerevisiae. Eukaryot Cell. 2002;1(5):774-86. Epub 2002/11/29. doi:
1045 10.1128/EC.1.5.774-786.2002. PubMed PMID: 12455695; PubMed Central PMCID:
1046 PMCPMC126745.
1047 50. Strong PJ, Claus H. Laccase: A Review of Its Past and Its Future in Bioremediation. Critical
1048 Reviews in Environmental Science and Technology. 2011;41(4):373-434. doi:
1049 10.1080/10643380902945706.
1050 51. Kombrink A, Thomma BP. LysM effectors: secreted proteins supporting fungal life. PLoS
1051 Pathog. 2013;9(12):e1003769. Epub 2013/12/19. doi: 10.1371/journal.ppat.1003769. PubMed
1052 PMID: 24348247; PubMed Central PMCID: PMCPMC3861536.
1053 52. Seidl-Seiboth V, Zach S, Frischmann A, Spadiut O, Dietzsch C, Herwig C, et al. Spore
1054 germination of Trichoderma atroviride is inhibited by its LysM protein TAL6. FEBS J.
1055 2013;280(5):1226-36. Epub 2013/01/08. doi: 10.1111/febs.12113. PubMed PMID: 23289754.
1056 53. Inoue M, Takenaka H, Tsurushima T, Miyagawa H, Ueno T. Colletofragarones A1 and A2, novel
1057 germination self-inhibitors from the fungus Colletotrichum fragariae. Tetrahedron Letters.
1058 1996;37(32):5731-4. doi: https://doi.org/10.1016/0040-4039(96)01212-9.
1059 54. Yu PL, Wang CL, Chen PY, Lee MH. YAP1 homologue-mediated redox sensing is crucial for a
1060 successful infection by Monilinia fructicola. Mol Plant Pathol. 2017;18(6):783-97. Epub 2016/05/31.
1061 doi: 10.1111/mpp.12438. PubMed PMID: 27239957; PubMed Central PMCID: PMCPMC6638302.
1062 55. MacPherson S, Larochelle M, Turcotte B. A fungal family of transcriptional regulators: the zinc
1063 cluster proteins. Microbiol Mol Biol Rev. 2006;70(3):583-604. Epub 2006/09/09. doi:
1064 10.1128/MMBR.00015-06. PubMed PMID: 16959962; PubMed Central PMCID: PMCPMC1594591.
1065 56. Liu XH, Zhuang FL, Lu JP, Lin FC. Identification and molecular cloning Moplaa gene, a
1066 homologue of Homo sapiens PLAA, in Magnaporthe oryzae. Microbiol Res. 2011;167(1):8-13. Epub
1067 2011/04/13. doi: 10.1016/j.micres.2011.02.003. PubMed PMID: 21482087.
1068 57. Zhang F, Suarez G, Sha J, Sierra JC, Peterson JW, Chopra AK. Phospholipase A2-activating
1069 protein (PLAA) enhances cisplatin-induced apoptosis in HeLa cells. Cell Signal. 2009;21(7):1085-99.
1070 Epub 2009/03/05. doi: 10.1016/j.cellsig.2009.02.018. PubMed PMID: 19258036.
1071 58. Chao YT, Chen WC, Chen CY, Ho HY, Yeh CH, Kuo YT, et al. Chromosome-level assembly,
1072 genetic and physical mapping of Phalaenopsis aphrodite genome provides new insights into species
1073 adaptation and resources for orchid breeding. Plant Biotechnol J. 2018;16(12):2027-41. Epub
1074 2018/04/29. doi: 10.1111/pbi.12936. PubMed PMID: 29704444; PubMed Central PMCID:
1075 PMCPMC6230949.
1076 59. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing
1077 genome assembly and annotation completeness with single-copy orthologs. Bioinformatics.
1078 2015;31(19):3210-2. Epub 2015/06/11. doi: 10.1093/bioinformatics/btv351. PubMed PMID:
1079 26059717.

60. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160-6. Epub 2017/10/03. doi: 10.1093/bib/bbx108. PubMed PMID: 28968734; PubMed Central PMCID: PMCPMC6781576.

61. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp348. PubMed PMID: 19505945; PubMed Central PMCID: PMCPMC2712344.

62. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539-42. Epub 2012/02/24. doi: 10.1093/sysbio/sys029. PubMed PMID: 22357727; PubMed Central PMCID: PMCPMC3329765.

63. Rambaut A. FigTree, a graphical viewer of phylogenetic trees. (http://tree.bio.ed.ac.uk/software/figtree/) (20 December 2020, date last accessed). 2012.

64. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Res. 2020;9. Epub 2020/09/22. doi: 10.12688/f1000research.23297.2. PubMed PMID: 32489650; PubMed Central PMCID: PMCPMC7222033.2.

65. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol Plant Pathol. 2018;19(9):2094-110. Epub 2018/03/24. doi: 10.1111/mpp.12682. PubMed PMID: 29569316; PubMed Central PMCID: PMCPMC6638006.

66. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95-W101. Epub 2018/05/18. doi: 10.1093/nar/gky418. PubMed PMID: 29771380; PubMed Central PMCID: PMCPMC6031026.

67. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 2010;47(9):736-41. Epub 2010/06/18. doi: 10.1016/j.fgb.2010.06.003. PubMed PMID: 20554054; PubMed Central PMCID: PMCPMC2916752.

68. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base: the pathogen-host interactions database. Nucleic Acids Res. 2020;48(D1):D613-D20. Epub 2019/11/17. doi: 10.1093/nar/gkz904. PubMed PMID: 31733065; PubMed Central PMCID: PMCPMC7145647.

69. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49(D1):D344-D54. Epub 2020/11/07. doi: 10.1093/nar/gkaa977. PubMed PMID: 33156333; PubMed Central PMCID: PMCPMC7778928.

70. Braga RM, Santana MF, Veras da Costa R, Brommonschenkel SH, de Araujo EF, de Queiroz MV. Transposable elements belonging to the Tc1-Mariner superfamily are heavily mutated in

Colletotrichum graminicola. Mycologia. 2014;106(4):629-41. Epub 2014/06/05. doi: 10.3852/13-262. PubMed PMID: 24895425.

71. Mullins ED, Chen X, Romaine P, Raina R, Geiser DM, Kang S. Agrobacterium-Mediated Transformation of Fusarium oxysporum: An Efficient Tool for Insertional Mutagenesis and Gene Transfer. Phytopathology. 2001;91(2):173-80. Epub 2008/10/24. doi: 10.1094/phyto.2001.91.2.173. PubMed PMID: 18944391.

72. Short JM, Fernandez JM, Sorge JA, Huse WD. Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. Nucleic Acids Res. 1988;16(15):7583-600. Epub 1988/08/11. doi: 10.1093/nar/16.15.7583. PubMed PMID: 2970625; PubMed Central PMCID: PMCPMC338428.

73. Mullaney EJ, Hamer JE, Roberti KA, Yelton MM, Timberlake WE. Primary structure of the trpC gene from Aspergillus nidulans. Molecular and General Genetics MGG. 1985;199(1):37-45. doi: 10.1007/BF00327506.

74. Pilgeram AL, Henson JM. Transformation and cotransformation of Gaeumannomyces graminis to phleomycin resistance. Phytopathology. 1990;80(10):1124-9. doi: 10.1094/phyto-80-1124. PubMed PMID: IND90055602.

75. Liu Z, Friesen TL. Polyethylene glycol (PEG)-mediated transformation in filamentous fungal pathogens. Methods Mol Biol. 2012;835:365-75. Epub 2011/12/21. doi: 10.1007/978-1-61779-501-5_21. PubMed PMID: 22183664.

76. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5. Epub 2012/08/30. doi: 10.1038/nmeth.2089. PubMed PMID: 22930834; PubMed Central PMCID: PMCPMC5554542.

Supporting information

Supporting Tables

S1Table. Information on the five maker genes (ACT1, CHS1, GAPDH, ITS and TUB2) used in the phylogenetic analysis in Fig 1 and S1 and S2 Figs.

S2 Table. Functional category statistics of the three *Colletotrichum scovillei* genomes.

S3 Table. Functional category statistics of the multi copy orthogroups with different gene numbers for the three *Colletotrichum scovillei* genomes.

S4 Table. Repeat sequence analysis with MISA for microsatellite compositions in the three strains.

S5 Table. The distribution of the 59 genes clustered at different scaffolds in Coll-524. The numbers of genes absent in both strains Coll-153 and Coll-365, and in strain Coll-365 only, are presented.

S6 Table. Pathogenicity assay of transgenic strains and wild-type strain Coll-365 (WT) on fruits of *Capsicum annuum* cv. Groupzest by pair inoculation.

S7 Table. TransposonPSI analysis results of three *Colletotrichum scovillei* strains and *C. higginsianum* IMI 349063.

S8 Table. Primers used in this study.

Supporting Figures

S1 Fig. Phylogenetic tree of 173 *Colletotrichum* strains and one *Monilochaetes infuscans* strain showing the phylogenetic relationship of the three *Colletotrichum* strains (Coll-524, Coll-153 and Coll-365) and other *Colletotrichum* species. The phylogeny was constructed by using DNA sequences of five markers (ACT1, CHS1, GAPDH, ITS and TUB2) according to Cannon et al. (2012) with additional 62 *Colletotrichum* strains from NCBI as indicated in Table S1. The 62 strains from NCBI was marked with "*". Values at the nodes are Bayesian posterior probability values. Strains belong to various *Colletotrichum* species complex were indicated with different color boxes.

S2 Fig. The original host and phylogenetic relationship of strains in *Colletotrichum acutatum* species complex. This tree was part of the phylogenetic tree in S1 Fig. The original host of each strain was indicated in the right side behind each strain. The clade containing Coll-524, 153 and 365 was highlighted with yellow color.

S3 Fig. Drug metabolism – cytochrome P460 KEGG pathway (A) and the gene number variations of monooxygenase among the three strains (B). (A) The drug metabolism KEGG map shown here is a partial map of map00982. (B) The gene number differences of monooxygenase among the three strains were based on the investigation of scaffolds 17, 19, 20 and 22. The positions of monooxygenase in the drug metabolism – cytochrome P450 pathway are indicated with red circles in panel A.

S4 Fig. Purine metabolism KEGG pathway (A) and the gene number variations of adenylpyrophosphatase among the three strains (B). The position of the adenylpyrophosphatase in the purine metabolism pathway is indicated with a red circle in panel A.

S5 Fig. The thiamine biosynthesis KEGG pathway (A) and the gene number variations of phosphatase among the three strains (B). The position of the phosphatasease in the thiamine biosynthesis pathway is indicated with a red circle in panel A.

S6 Fig. Phenylpropanoid biosynthesis KEGG pathway (A) and the gene number variations of lactoperoxidase among the three strains (B). The positions of the lactoperoxidase in phenylpropanoid biosynthesis pathway are indicated with red circles in panel A.

S7 Fig. N-glycan biosynthesis KEGG pathways (A) and the gene number variations of alpha-1,6-mannosyltransferase among the three strains (B). (A) The N-glycan biosynthesis KEGG map shown here is a partial map of map00513. (B) The gene number differences of 1,6-mannosyltransferase among the three strains were based on the investigation of scaffolds 17, 19, 20 and 22. The positions of the 1,6-mannosyltransferase in the purine metabolism pathway are indicated with red boxes and lines in panel A.

S8 Fig. The complete Coll-Fot1 (1891 bp) of strain Coll-524 consists of a coding region and two inverted terminal repeats (ITRs). The DDE_1 containing region provided by TransposonPSI is highlighted with a green arrow. The coding region of 545 aa sequence was identified by SnapGene and is highlighted with a light blue arrow. The yellow bar indicates the DDE_1 domain resulted from InterPro analysis. Two ITRs with 55 bp in length are provided by RepeatModeler and is highlighted as a blue arrow.

S9 Fig. Verification of gene presence in strains Coll-524, 153 and 365 by PCR (A), and confirmation of transgenic strains with regular PCR (B) and RT-PCR (C) for target gene transformation in strain Coll-365.