Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi

Stéphane Hacquard1,*, Barbara Kracher1,*, Kei Hiruma1,1, Philipp C. Münch2,3,4, Ruben Garrido-Oter1,5,6, Michael R. Thon7, Aaron Weimann3,5, Ulrike Damm8,†, Jean-Félix Dallery9, Matthieu Hainaut10,11, Bernard Henrissat10,11,12, Olivier Lespinet13,14, Soledad Sacristán15, Emiel Ver Loren van Themaat1,*, Eric Kemen1,6, Alice C. McHardy3,5,6, Paul Schulze-Lefert1,6 & Richard J. O’Connell1,9

The sessile nature of plants forced them to evolve mechanisms to prioritize their responses to simultaneous stresses, including colonization by microbes or nutrient starvation. Here, we compare the genomes of a beneficial root endophyte, Colletotrichum tofieldiae and its pathogenic relative C. incanum, and examine the transcriptomes of both fungi and their plant host Arabidopsis during phosphate starvation. Although the two species diverged only 8.8 million years ago and have similar gene arsenals, we identify genomic signatures indicative of an evolutionary transition from pathogenic to beneficial lifestyles, including a narrowed repertoire of secreted effector proteins, expanded families of chitin-binding and secondary metabolism-related proteins, and limited activation of pathogenicity-related genes in planta.

We show that beneficial responses are prioritized in C. tofieldiae-colonized roots under phosphate-deficient conditions, whereas defense responses are activated under phosphate-sufficient conditions. These immune responses are retained in phosphate-starved roots colonized by pathogenic C. incanum, illustrating the ability of plants to maximize survival in response to conflicting stresses.
Fungal endophytes are a ubiquitous and phylogenetically diverse group of organisms that establish stable associations with living plants, but in most cases their ecophysiological significance is poorly understood. Species of the fungal genus *Colletotrichum* are best known as destructive pathogens on >3,000 species of dicot and monocot plants worldwide, causing anthracnose diseases and blights on leaves, stems, flowers and fruits. However, *Colletotrichum* species can also grow benignly as endophytes on symptomless plants, and although only a few pathogenic members of the genus attack plant roots, *Colletotrichum* endophytes are frequently isolated from the roots of healthy plants. Moreover, although the genome sequences and *in planta* transcriptomes were recently described for four species pathogenic on above-ground plant parts, such information is not available for any root-associated *Colletotrichum* pathogens or endophytes.

We found recently that *C. tofieldiae* (*Ci*) is an endophyte in natural populations of *Arabidopsis thaliana* growing in central Spain. The fungus initially penetrates the rhizoderm by means of natural populations of *Arabidopsis thaliana* (also pathogen isolated from diverse dicot and monocot hosts in Europe). *C. thaliana* isolate 0861, a root endophyte isolated from natural populations sequenced the genome of the plant growth-promoting fungus *Colletotrichum sp.*

**Results**

**Genome sequencing and evolution of *Ct* and *Ci* lifestyles.** We sequenced the genome of the plant growth-promoting fungus *Ct* isolate 0861, a root endophyte isolated from natural populations of *A. thaliana* in Spain, and those of four other *Ct* isolates isolated from diverse dicot and monocot hosts in Europe (Supplementary Note 1). We also sequenced the broad host-range pathogen *Ci*, isolated from radish (*Raphanus sativus*) leaves in Japan, that strongly impairs plant growth when inoculated onto *Arabidopsis* roots. (Supplementary Fig. 1, Supplementary Table 1 and Supplementary Note 1). Illumina short reads were used to build high-quality genome assemblies of similar size for all isolates, ranging from 52.8 to 53.6 Mb (Supplementary Table 2 and Supplementary Note 2). Molecular phylogeny, whole-genome alignment and divergence date estimates indicate that *Ci* and *Ct* are closely related taxa within the *Colletotrichum* species complex and diverged only ~8.8 million years ago (Fig. 1a, Supplementary Figs 2 and 3, Supplementary Table 3 and Supplementary Note 3). Our phylogenetic analysis suggests that evolution from pathogenic ancestors towards the beneficial endophytic lifestyle in *Ct* is a recent adaptation in *Colletotrichum* fungi.

**SNP distribution and reproductive mode of *Ct* isolates.** Although the five *Ct* isolates originate from widely separated geographical areas and distantly related plant hosts, they diverged only ~0.29 million years ago and the aligned fractions (>93%) of their genomes share >99% sequence identity (Fig. 1a and Supplementary Tables 1, 3 and 4). The overall frequency of single-nucleotide polymorphisms (SNPs) between isolates was similar (2.22–3.04 SNPs per kb) but the SNP distribution within each genome was uneven, with alternating tracts of low (0.22–0.32 SNPs per kb) and high (4.25–5.12 SNPs per kb) SNP density (Fig. 1b, Supplementary Fig. 4 and Supplementary Table 5). This peculiar SNP distribution, also visible in the genomes of other plant-interacting fungi, is consistent with chromosome recombination events. However, the SNP density profiles are remarkably similar between isolates and large haplotype blocks are conserved between all (21%), four (19%), three (18%) or two (17%) of them, with only 22% being isolate specific (Fig. 1b,c, Supplementary Fig. 4, Supplementary Table 6 and Supplementary Note 4). These conserved SNP signatures in the genomes of geographically distant isolates were likely generated by rare or ancestral sexual/parasexual reproduction and maintained by frequent clonal propagation.

**Evolutionary dynamics of multigene families in *Colletotrichum*.** Similar numbers of protein-coding genes were predicted in *Ct*0861 and *Ci* (~13,000; Supplementary Table 2), with >11,300 orthologous genes shared between both species. By clustering protein-coding sequences into sets of orthologous genes using OrthoMCL, we identified 7,297 gene families conserved across all six analysed *Colletotrichum* species and 10,519 shared between *Ct*0861 and *Ci* (Fig. 2a,b and Supplementary Note 5). Using a maximum-likelihood approach, we also reconstructed ancestral genomes for each *Colletotrichum* lineage and predicted the number of gene families that were likely gained or lost in each species compared with its corresponding ancestor (Supplementary Fig. 5 and Supplementary Note 6). We found significantly more gene families gained (1,009) than lost (198) on the branch leading to *Ct* compared with other branches of the tree (Fisher’s exact test, *P* = 3.98 × 10⁻¹³⁶; Supplementary Fig. 5 and Supplementary Data 1). Functional enrichment analysis among the 1,009 gene families gained (Supplementary Fig. 5) and the 1,486 *Ct*-specific gene families (Fig. 2b) revealed a significant enrichment for genes encoding secondary metabolite biosynthesis-related proteins in *Ct* (Fisher’s exact test, *P* = 5.89 × 10⁻³ and 3.31 × 10⁻⁶, respectively). This result contrasts with the very low number of secondary metabolite-related genes detected in the genomes of other root-associated fungal endophytes and mycorrhizal fungi, and suggests that either fungal secondary metabolites have roles in establishing a beneficial endophytic interaction with host plants or in limiting...
Figure 1 | Colletotrichum evolutionary divergence dates and SNP distribution in C. tofieldiae isolates. (a) Phylogeny of Colletotrichum species inferred from analysing 20 single-copy gene families using PhyML and r8s. Nodes 1–3 (green) are calibration points and nodes 4, 5 and 6 (red) represent estimated divergence dates (see Supplementary Note 3). (b) Circular visualization of the alignment of genome sequencing reads and SNP locations of four C. tofieldiae isolates with respect to the Ct0861 reference assembly. Tracks represent (from the outside) the five largest Ct0861 contigs (scale: kb); locations of predicted genes; locations of SNPs versus Ct0861 in CBS495, CBS130, CBS127, CBS168 (see Supplementary Table 1 for full culture IDs) and SNPs common to these four isolates; conserved regions with low SNP density between all the five isolates; mean read coverage (per 100 bases) for isolates CBS495, CBS130, CBS127 and CBS168. Coverage plot scales are 0 to 1,000 (CBS495) or 0 to 500 (CBS130, 127, 168). (c) SNP density (per 1 kb) in isolates CBS495, CBS130, CBS127 and CBS168 versus Ct0861, compared with gene density (per 10 kb) and GC content (%) on the five largest Ct0861 contigs.
the colonization of microbial competitors inside roots. Evaluation of the selective forces (dS/dd ratio) acting on all the protein families in the Ct genome revealed that genes involved in ‘signal transduction mechanisms’, ‘RNA processing and modification’ and ‘lipid transport and metabolism’ showed the strongest evidence of adaptive evolution (false discovery rate (FDR) < 0.05, Fisher’s test). This contrasts with pathogenic Colletotrichum species for which gene families belonging to the categories ‘defense mechanisms’, ‘cell wall/membrane/envelope biogenesis’ and ‘RNA processing and modification’ show the highest dS/dd ratios (Supplementary Figs 6 and 7, Supplementary Data 2 and Supplementary Note 7).

Genomic signatures of the pathogenic to beneficial transition. Ct encodes large repertoires of transporters, secreted proteins, proteases, carbohydrate-active enzymes (CAZymes) and secondary metabolism key enzymes, very similar to Ci and four other pathogenic Colletotrichum species (Supplementary Figs 8–12 and Supplementary Note 8). By comparing the Ct gene repertoires to those of five other plant-associated fungal endophytes from both ascomycete and basidiomycete lineages, we found no obvious common genomic signatures to indicate the convergent evolution of an endophyte ‘toolkit’ (Supplementary Figs 8–12). Furthermore, the convergent loss of decay mechanisms characteristic of ectomycorrhizal fungi15,16 is not a hallmark shared by the non-mycorrhizal root endophytes (Supplementary Fig. 12), suggesting that these fungi have followed different evolutionary trajectories to acquire the ability for intimate growth in living root tissues14,17.

Despite the overall similar secretome size of all analysed Colletotrichum species (13.3–15.9% of the total proteome), the proportion of genes encoding candidate secreted effector proteins (CSEPs), which may promote fungal infection18, varied considerably between species (6.6–15.8% of the total secretome; Fig. 3a and Supplementary Table 7). The smaller CSEP repertoire in C. tofieldiae (133 versus 189 in Ci) is largely explained by the reduction of species-specific CSEPs (34 versus 72 in Ci; Fig. 3a, Supplementary Fig. 13, Supplementary Table 7 and Supplementary Data 3). As expected, calculation of dS/dd ratios among 331 CSEP families derived from all the 10 analysed Colletotrichum genomes indicates they are under diversifying selection (median 0.35, interquartile range 0.21–0.49) relative to non-CSEP families (median 0.20, interquartile range 0.07–0.33; Fisher’s exact test, P < 2.2 × 10-16; Fig. 3b). Genomes from additional Ct isolates are now needed to determine whether there is differential host-selective pressure on the CSEP repertoires of endophytic Ct and pathogenic Ct that reflect their contrasting lifestyles. Similar to other Colletotrichum species2, CSEPs in Ct and Ci are not organized into large multigene families, possibly due to a low frequency of duplication events in their respective genomes (Fig. 3c,d and Supplementary Table 2).

Both Ct and Ci genomes encode a very broad range of CAZymes, including large arsenals of pectate lyases, carbohydrate esterases and glycoside hydrolases acting on all major plant cell wall constituents (Fig. 4a, Supplementary Fig. 12 and Supplementary Data 4). However, the number of predicted carbohydrate-binding modules is inflated in Ct compared with pathogenic Colletotrichum species, especially chitin-binding CBM18 (48 versus 28–40) and CBM50 (57 versus 30–54) modules (Fig. 4a, Supplementary Data 4), though few of the corresponding Ct genes were induced in planta (Supplementary Fig. 14). These two chitin-binding modules are similarly highly enriched in the genomes of two other non-mycorrhizal root symbionts19,20 (Piriformospora indica and Harpophora oryzae; Supplementary Data 4), suggesting this is a genomic signature common to independently evolving root-associated fungal endophytes.

Dual RNAseq of Arabidopsis roots and fungal partners. We report elsewhere that Ct promotes Arabidopsis growth under phosphate-deficient (− P) but not phosphate-sufficient (+ P) conditions and that transfer of radioactive32 P from Ct hyphae to host plants is strictly regulated by Pi (inorganic phosphate) availability3. To compare the transcriptional dynamics of beneficial Ct and pathogenic Ct during colonization of Arabidopsis roots and study the corresponding host responses, we extensively re-analysed the previously created RNA-seq data for the Ct-Arabidopsis interaction (6, 10, 16 and 24 days post inoculation (d.p.i.), + P: 625 μM, − P: 50 μM; ref. 8) and included new samples for the Ci-Arabidopsis interaction (10 and 24 d.p.i., − P: 50 μM) (Supplementary Figs 15 and 16). After mapping Illumina reads to their respective genomes, we obtained expression data for >20,000 Arabidopsis genes, 8,613 Ci genes and 6,693 Ct genes (Supplementary Fig. 17, Supplementary Table 8 and Supplementary Note 9). The expression data were validated using quantitative PCR with reverse transcription.
versus in vitro alignments, CSEP length (0–500 amino acids), species-specific (Sp, black) and genus-specific (Ge, white) CSEPs, normalized gene expression (Exp.) levels of CSEPs (secretomes of predicted secreted proteins (circles, violet sectors) and candidate secreted effector proteins CSEPs (circles, yellow sectors) in the proteomes and encoding secreted proteins and CSEPs compared with the remaining gene families (One-sided Fisher’s test, ***). Note 10).

Figure 3 | Conservation and expression of genes encoding candidate secreted effector proteins in *C. tofieldiae* and *C. incanum*. (a) Proportions of predicted secreted proteins (circles, violet sectors) and candidate secreted effector proteins CSEPs (circles, yellow sectors) in the proteomes and secretomes of *Colletotrichum* species, respectively. The number of genus- and species-specific CSEPs detected for each species is indicated in the barplot. (b) Boxplot with a rotated kernel density on each side showing \( |\log_2 FC| \) measured in the proteome, the secretome and the CSEP repertoires of 10 *Colletotrichum* isolates using the gene families defined by MCL clustering (see Fig. 2). The overall \( |\log_2 FC| \) ratio is significantly higher for gene families encoding secreted proteins and CSEPs compared with the remaining gene families (One-sided Fisher’s test, **P < 0.001**). (c,d) Expression and regulation of CSEPs in *C. tofieldiae* 0861 (c) and *C. incanum* (d). The circular plots show (from the inside): dendrograms of the CSEPs based on protein sequence alignments, CSEP length (0–500 amino acids), species-specific (Sp, black) and genus-specific (Ge, white) CSEPs, normalized gene expression (Exp.) levels in vitro (IV) and in planta (IP) at 10 and 24 days post inoculation, CSEPs significantly up- (violet) and downregulated (green) at 10 days post inoculation versus in vitro (10P/IV) and 24 days post inoculation versus in vitro (24IP/IV) (\( |\log_2 FC| \geq 1 \), FDR < 0.05).

Transcriptional shutdown of pathogenicity genes in *Ct*. Among the 3,885 *Ct* genes significantly regulated (moderated t-test, \( |\log_2 FC| \geq 1 \), FDR < 0.05), only few (61) were impacted by phosphate status (described in ref. 8) or the fungal developmental stage in planta (845; Supplementary Data 5 and Supplementary Fig. 19). In contrast, ~80% were induced upon host contact and particularly those encoding CAZymes, for which a dynamic expression pattern was observed (Fig. 4b and Supplementary Figs 19 and 20). A first wave of activation (6–16 d.p.i.) involved few plant cell wall-degrading enzymes (PCWDEs) acting mostly on hemicellulose, while a second wave (24 d.p.i.) involved induction of numerous PCWDEs acting on all major wall polymers, including cellulose, hemicellulose and pectin (Fig. 4b). Thus, at later infection stages, *Ct* displays significant saprotrophic capabilities. However, genes encoding CSEPs, secreted proteases, secondary metabolism key enzymes and transporters showed no clear activation (Supplementary Fig. 21), in contrast to the highly stage-specific deployment of such genes by *C. higginsianum* during infection of *Arabidopsis* leaves 2. Surprisingly, the activation of *Ct* CSEPs was almost non-existent in planta, with only 18/133 expressed during colonization, 8/133 induced in planta (\( \log_2 FC \geq 1 \)) and 4/133 ranking among the 1,000 most highly expressed genes (Fig. 3c). These few expressed CSEP genes showed similar \( dS/dk \) ratios compared with CSEPs that were silent in planta (Supplementary Data 3). The contracted repertoire and small number of CSEPs activated in planta suggests *Ct* requires extremely few effectors for host invasion and maintenance of the beneficial relationship.

Gene deployment in planta reflects fungal lifestyles. To uncover transcriptional adaptations associated with the evolutionary transition from the ancestral pathogenic lifestyle to beneficial endophytism, we compared the normalized expression levels of 6,804 *Ct* and *Ci* orthologous gene pairs that are expressed (RT–qPCR) with a subset of *Arabidopsis* and *Ct* genes (Supplementary Fig. 18, Supplementary Table 9 and Supplementary Note 10).
More than twice as many gene pairs were differentially expressed at 10 d.p.i. (621 up, 842 down) than at 24 d.p.i. (306 up, 273 down; moderated t-test, \(|\text{log}_2\text{FC}| > 1, \text{FDR} < 0.05\)), suggesting that early colonization events are critical for determining the outcome of the interaction. GO term enrichment analysis showed that processes related to melanin biosynthesis were significantly enriched in C. tofieldiae, consistent with the formation of melanized microsclerotia in C. tofieldiae but not C. higginsianum (Supplementary Table 10). We also found major differences between C. tofieldiae and C. incanum in the expression of gene categories typically associated with fungal pathogenicity. In planta activation of CSEPs was more pronounced in C. incanum compared with C. tofieldiae, with seven times more CSEPs highly expressed (top 1,000 expressed genes) and three times more upregulated in planta at 10 d.p.i. (Fig. 3c,d and Supplementary Data 5 and 7). Likewise genes encoding CAZymes and secondary metabolism enzymes displayed earlier and stronger transcriptional activation in planta and broader diversity in C. incanum (Fig. 4b,c and Supplementary Fig. 22). Consistent with this, we observed a reduced number of living cells and a depletion of beta-linked polysaccharides (including cellulose) from host cell walls in C. incanum-colonized roots at 10 d.p.i., but not in C. tofieldiae-colonized roots (Supplementary Fig. 1). This finding suggests that pathogenic C. incanum harvests carbon from plant cell walls more aggressively than C. tofieldiae. Thus, despite their phylogenetic proximity and similar gene arsenals, gene deployment during infection was strikingly different between C. tofieldiae and C. incanum. The in planta transcriptome of C. incanum resembles that of other pathogenic Colletotrichum species, whereas the less dynamic transcriptome...
of Ct might contribute to, or be a consequence of, the beneficial relationship. Overall, our results suggest that the recent transition from pathogenic to beneficial lifestyles might be partly controlled through transcriptional downregulation of pathogenicity-related genes in Ct.

Host responses to Ct are phosphate-status dependent. To disentangle how Pi-starved and non-starved Arabidopsis roots respond to Ct colonization over time, we compared Ct-colonized and mock-inoculated roots under +P and −P conditions. In total, 5,661 Arabidopsis genes were differentially expressed in at least one of the 16 pairwise comparisons (moderated t-test, log2FC ≥ 1, FDR < 0.05) and grouped into 20 major gene expression clusters (Fig. 5a and Supplementary Data 8). GO term enrichment analysis among these clusters indicated that the phosphate level used in our study (50 μM) was sufficient to provoke a phosphate starvation response in Arabidopsis roots (clusters 2 and 4; Fig. 5b). Furthermore, our analysis indicates that ‘response to stimulus’, ‘indole glucosinolate metabolic process’, ‘defense response’ and ‘ethylene metabolic process’ are activated in Ct-colonized roots under +P but not −P conditions (Fig. 5b and Supplementary Data 9). In contrast, the genes related to ‘root cell differentiation’ (cluster 8, Fig. 5b) and phosphate uptake (20) were preferentially activated in Pi-starved Arabidopsis roots during Ct colonization, similar to mycorrhizal symbiont–host interactions (21). To identify key regulatory genes (hub genes) that might orchestrate transcriptional reprogramming in the contrasting directions seen in clusters 8 and 9, we checked which of these genes are often co-regulated in other expression data sets using the ATTED-II gene co-expression database (Fig. 5c). Among the hub genes that showed high connectivity within cluster 8 (highlighted with black dots), many encode proteins involved in cell wall remodelling and root hair development. Particularly, genes encoding the root hair-specific proteins RHS8, RHS12, RHS13, RHS15 and RHS19 (ref. 22) are upregulated (moderated t-test, log2FC ≥ 1, FDR < 0.05) in Ct-colonized versus mock-treated roots under −P conditions, which was validated by RT–qPCR (Fig. 5d and Supplementary Fig. 18). This expression pattern suggests that Ct-dependent remodelling of root architecture might play a key role to enhance phosphate uptake during starvation (Supplementary Note 9). Similarly, we identified 27 hub genes within cluster 9 (Fig. 5c, black dots), encoding well-characterized defense-related proteins such as the transcription factors WRKY33 and WRKY40 (ref. 23), the ethylene-responsive factors ERF11 and ERF13 (ref. 24), as well as MYB51 (ref. 25), a transcription factor regulating Tryptophan (Trp)-derived indole glucosinolate metabolism. Four other genes involved in indole glucosinolate metabolism were also highly differentially regulated in cluster 9, including the myrosinase PEN2 and the P450 monoxygenase CYP81F2 required for the biosynthesis of 4-methoxy-indol-3-ylmethylglucosinolate, the substrate of PEN2 myrosinase (26,27) (Supplementary Data 9). The PEN2-dependent metabolism of Trp-derived indole glucosinolates in A. thaliana is activated upon perception of pathogen-associated molecular patterns (PAMPs) and induced by pathogen infection (28). As tightly connected GO terms are functionally linked, only the major enrichment terms were manually checked and compared to other expression data sets (Supplementary Note 9).

Figure 5 | Transcriptional reprogramming of Pi-starved and non-starved Arabidopsis roots in response to C. tofieldiae. (a) Transcript profiling of 5,561 Arabidopsis genes significantly regulated (moderated t-test, log2FC ≥ 1, FDR < 0.05) between colonized versus mock-treated roots and phosphate-starved (−P: [50 μM]) versus non-starved roots (+P: [625 μM]) at 6, 10, 16 and 24 days post inoculation. Overrepresented (yellow to red) and underrepresented transcripts (yellow to blue) are shown as log2 (fold changes) relative to the mean expression across all stages. Using k-means partitioning, the gene set was split into 20 major gene expression clusters. (b) Gene Ontology term enrichment network analysis among the 10 clusters highlighted in a. Each significantly enriched GO term (P < 0.05, hypergeometric test, Bonferroni step-down correction) is represented with a circle and the contribution (%) of each cluster to the overall GO term enrichment is represented using the same colour code as in a. As tightly connected GO terms are functionally linked, only the major host responses outputs are indicated (dotted line). (c) For cluster 8 and cluster 9, gene relationships based on co-regulation were assessed using other Arabidopsis expression data sets (see Supplementary Note 9). The genes within each cluster that show strong expression relationships in other expression data sets are likely to encode key regulatory hubs. Hub genes (cluster 9: ≥ 5 connections, *cluster 8: ≥ 10 connections) are highlighted in black. The corresponding characterized Arabidopsis genes are indicated below the co-expression networks. (d) Validation of the expression profiles of the hub genes RHS19 (cluster 8) and ERF13 (cluster 9) using RT–qPCR (see Supplementary Note 10). Error bars indicate standard error (n = 3 biological replicates), NA, data not available; REI, Relative Expression Index.
patterns by receptors of the innate immune system and is needed for broad-spectrum defence to restrict the growth of fungal pathogens\textsuperscript{26,27}. Notably, in Arabidopsis mutants that cannot activate PEN2-mediated antifungal defense, the promotion of plant growth by Ct is impaired, while the depletion of all Trp-derived secondary metabolites renders Ct a pathogen on Arabidopsis\textsuperscript{8}. These findings strongly suggest that the phosphate starvation response and Trp-derived indole glucosinolate metabolism are interconnected to control fungal colonization of Arabidopsis roots\textsuperscript{28}. Phosphate status-dependent activation of defense responses was also observed among the 411 expressed Arabidopsis genes annotated as ‘chitin-responsive’ (Supplementary Fig. 23), based on GO term enrichment among all significantly regulated genes (Supplementary Fig. 24) and this was validated by RT–qPCR (Fig. 5d and Supplementary Fig. 18). These data reveal a remarkable capacity of Arabidopsis roots to prioritize different transcriptional outputs in response to Ct, favouring either defense responses under + P conditions or root growth and phosphate metabolism under − P conditions.

Phosphate-starved roots activate defense responses to Ct. To clarify whether the reduced activation of defense responses observed in Ct-colonized roots under − P conditions is not simply due to phosphate deficiency, we compared the transcriptomes of Pi-starved Arabidopsis roots in response to either Ci or Ct at 10 d.p.i. In total, 2,009 differentially expressed genes were identified (moderated \( t \)-test, \( |\log_{2}\text{FC}| \geq 1 \), FDR < 0.05), including 988 genes induced in Ct-colonized roots (cluster 1) and 1,021 genes in Ci-colonized roots (cluster 2; Fig. 6a and Supplementary Data 10). GO term enrichment analysis revealed that ion transport and root cell differentiation mechanisms were activated in Ct-colonized roots, whereas strong defense responses were triggered in Ci-colonized roots (Fig. 6b). Thus, although Pi-starved Arabidopsis roots remain able to mount immune responses against pathogenic Ci, transport and root growth are instead prioritized during interaction with beneficial Ct.

Discussion

Deciphering the genetic basis of the transition from pathogenic to beneficial plant-fungal interactions is crucial for a better understanding of the evolutionary history of fungal lifestyles\textsuperscript{20,29}. It was recently shown that the ectomycorrhizal lifestyle arose independently multiple times during evolution and that the transition was associated with (1) convergent loss of genes encoding PCWDEs present in their saprotrophic ancestors and (2) the repeated evolution of lineage-specific ‘toolkits’ of mycorrhiza-induced genes\textsuperscript{15}. However in striking contrast with ectomycorrhizal fungi, this transition in Ct, P. indica and H. oryzae was not accompanied by contraction of their PCWDE repertoires\textsuperscript{19,20}. In our study, the close phylogenetic relatedness of beneficial Ct and pathogenic Ci, and their ability to infect the same plant host, allowed us to resolve both genomic and transcriptomic signatures associated with this evolutionary transition. The overall high genomic similarity between Ct and Ci suggests that this transition involved only subtle remodelling of the gene repertoire (that is, a reduced set of CSEPs and expansion of chitin-binding and secondary metabolism-related protein families). The retention of abundant pathogenicity- or saprophyt-related genes implies that they are still needed by Ct, perhaps for exploitation of other plant hosts or during plant senescence when Arabidopsis leaves are extensively colonized by Ct mycelium\textsuperscript{8}. Our results also suggest that changes in fungal...
gene expression patterns during host colonization, rather than extensive remodelling of the gene repertoire, provides an alternative and probably transient adaptation to a beneficial endophytic lifestyle. This may reflect the relatively recent transition from pathogenic to non-pathogenic lifestyles in Ct and, consequently, a latent capacity to revert to a pathogenic lifestyle.

During the last decade, the molecular mechanisms by which plants respond to colonization by pathogenic or mutualistic fungi have been extensively studied. However, it remains unclear how plants discriminate and respond appropriately to closely related fungal partners with different lifestyles. The sedentary nature of plants suggests they have evolved regulatory systems to integrate exposure to conflicting biotic and abiotic stresses and balance their resource allocation strategically to maximize growth and survival. A recent report showed that plant responses to multiple stresses are not cumulative and suggested that prioritization of stress responses does take place. For plant–mycorrhizal associations, an inverse correlation was observed between nutrition and defense. Whether phosphate stress–fungal interactions (Supplementary Fig. 25).

Methods

Genome sequencing and assembly.
C. incanum and the five C. tofieldiae isolates were grown in liquid Mathur’s medium (2.8 g glucose, 1.22 g MgSO4·7H2O, 2.72 g KH2PO4 and 2.18 g Oxoid mycological peptone in 1 l deionized water) supplemented with 100 µg ml–1 rifampicin and 125 µg ml–1 streptomycin. Genomic DNA was isolated using the DNeasy Plant Mini Kit (Qiagen) from 100 mg of fungal mycelium. Library construction, quality control and DNA sequencing for 454 FLXL+ or Illumina Hiseq2500, respectively. By mapping the genome to a reference genome, an hybrid assembly strategy was used combining 454 and Illumina data. Unpaired 454 reads were first assembled using MIRA 4.0 (ref. 33) and filtered for MIRA-contigs (>5,000 bp) were further used for scaffolding of Illumina paired-read assemblies from SPAdes 3.0 (ref. 34). The established SPAdes 3.0 pipeline was used in ‘careful’ mode providing 454 MIRA assembled contigs as untrusted-contigs for scaffolding only and a kmer scan using 21, 31, 41, 61, 75 and 81. All other assemblies were constructed only from Illumina data using a combination of VELVET 1.2.1 (ref. 35) and SPAdes36. Using BLASTN searches, contigs were identified that were missing from assembled contigs that present in VELVET assemblies. To integrate those contigs and extend further wherever possible, SPAdes was re-run as described above but in ‘trusted-contigs’ mode where trusted contigs were provided as fasta files with absent contigs only. The assemblies were generated using ‘careful’ mode in SPAdes to avoid mis-pairing of contigs by scaffolding and for further analyses, contigs <100 bp were removed. To identify and remove potential contaminating sequences, assemblages were aligned to the genomes of A. thaliana, H. sapiens and PhIX (sequencing spike-in control) using MUMmer37 with default parameter settings. Contigs that aligned with more than 50% of their sequence (coverage; COV) and at least 85% sequence identity (IDY) to any of the tested contaminants were removed from the assemblies. In addition, contigs that aligned with 75–85% identity (and >50% coverage) or with 10–50% coverage (and >85% identity) were also removed, if the judgment of the sequence being non-fungal was confirmed through BLASTN searches in the NCBI nr database (with default settings). For the C0861 assembly, RNA-seq data were used for further clean-up. Finally, assembly quality was assessed on the basis of the percentage of reads mapped to a single contig and scaffold (and 50/75/90 and >95% of the total reads with REAPR37 (version 1.0.16, default settings) and gene space coverage estimated with CEGMA38 (version 2.0, default settings).

Repetitive DNA analysis. We identified repetitive DNA in the genome assemblies using either de novo or homology approaches. For de novo searches, we used PILER and PALS39 to identify repetitive sequences and classify them into families. The resulting libraries of consensus sequences were then used to scan the genome sequences using RepeatMasker40 (version 4.0.3) to identify individual repetitive elements. For homology-based searches, we used RepeatMasker using a library of all fungal elements in the Repbase database41 (version 20140131). Phylogeny and divergence date estimation. All phylogenetic analyses performed in this study are described in the Supplementary Note 11. For evolutionary divergence date estimation, clustering, protein family selection and phylogenetic analyses were performed with scripts in the Mirlo package (https://github.com/mthonn/mirlo). The phylogeny was calibrated using the penalized-likelihood method implemented in r8s (ref. 42) using one primary and two secondary calibration points (Supplementary Note 11).

Short-read alignment and SNP analysis. To compare the genome sequences of Ct isolates, Illumina short reads of the four other isolates were mapped onto the genome of C. tofieldiae CBS0861 using Bowtie2 (ref. 43) (default settings for paired-end data). Subsequently, duplicate reads were removed using the rmdup function from the SAMtools toolkit44 (default settings). On the basis of the mapped genome sequence reads, single-nucleotide polymorphisms (SNPs) were identified using samtools mpileup function in SAMtools44 (version 0.1.18, with options ‘-f’). The obtained SNP sets were filtered using the bcftools script vcfutils.pl varFilter (SAMtools) with adjusted read depth settings according to the respective sequencing read coverage to -d 80 and -D 800 for CBS495 and to -d 40 and -D 400 for CBS130, CBS127 and CBS168. The SNP locations, read coverage for each isolate and locations of conserved regions were visualized using the Circos software package45 (version 0.62.1). In addition, we also calculated SNP densities (SNPs per kb) relative to C0861 for each isolate as a function of the genomic location on all C0861 contigs larger than 50 kb, using a 10-kb sliding window that moved 1 kb at each step. For visualization of the SNP densities, these windows were sorted in the increasing order by contig number and position on the contig. To identify windows with a low SNP density, that is, a common haplogroup, between isolates, we classified the SNP density in each window as either ‘low’ or ‘high’ using a two-state hidden Markov model (HMM). This HMM was created and fitted on the observed 10 kb SNP densities by the expectation-maximization algorithm using functions ‘de novo’ and ‘fit’ (R package depmixS4) and were evaluated with the posterior state sequence (with states ‘low’ and ‘high’), computed via the Viterbi algorithm, was extended with function ‘posterior’ (R package depmixS4).

Gene annotation. The prediction of Ct and Ci gene models was performed using the MAKER pipeline46 (version 2.2.8), which integrates different ab initio gene prediction tools together with evidence from EST and protein alignments. In a first step, for each genome, the pipeline was run using Augustus47 (with species model Fusarium graminearum) and GeneMark-ES48 for ab initio gene prediction together with transcript and protein alignment evidence. The resulting gene models from this first run were used as training set for a third ab initio prediction tool, SNAP49.
and subsequently the annotation pipeline was re-run, this time including all three prediction tools together with the transcript and protein alignment and subsequently the annotation pipeline was re-run, this time including all three prediction tools. Gene families and clusters of orthologous genes were inferred using OrthoMCL [version 2.0] with standard parameters and granularity 1.5 for the MCL clustering step. Functional enrichment and overrepresentation analyses were performed using a Fisher’s exact test, adjusting for FDR. For each gene family inferred with orthoMCL, a multiple sequence alignment of the protein sequences was obtained using Clustal Omega [version 201402] and an HMM model was generated with the hhmaccel program of the HHSuite toolkit. Sequences from the fungal database fuNOG were similarly aligned and HMM models generated. To annotate whole gene families, the hsearch program was used to obtain matches between the gene family and the funOG HMMs and only hits with a probability equal to or higher than 0.9 were considered. To annotate whole gene families, the hsearch program was used to obtain matches between the gene family and the fuNOG HMMs and only hits with a probability $\geq 0.99$ were considered.

**Ancestral genome reconstruction.** Gene families inferred with OrthoMCL were used to reconstruct the ancestral genomes of each *Colletotrichum* lineage. GLOOME [maximum-likelihood approach] was used to infer ancestral gene gains and losses (GGLs) and to reconstruct the ancestral GGLs of gene families on the species tree of *C. orbiculare* and the other five genomes available for this genus. Evolution of the GGLs along the branches of a phylogenetic tree was modelled as a continuous time Markov process using a binary character alphabet corresponding to gene family presence or absence. Default parameters were used, corresponding to a transition model that allows varying GGL rates across gene families. We approximated the total number of gene families that were gained or lost on a branch by summing up the individual posterior probabilities for each gene family to be gained or lost on that branch and rounding this number to the closest integer. The number of gene families either gained or lost (annotated with one specific category) was computed as the total number of gene families crossed the branch of the tree. The significance was assessed using Fisher’s exact test and FDR corrected.

**d$_g$/d$_l$ analysis.** A multiple-sequence alignment (MSA) of orthologous groups of coding sequences (CDSs) was created with Clearcut [version 2.0] using standard parameters and granularity 1.5 for the MCL clustering step. Functional enrichment and overrepresentation analyses were performed using a Fisher’s exact test, adjusting for FDR. For each gene family inferred with orthoMCL, a multiple sequence alignment of the protein sequences was obtained using Clustal Omega [version 201402] and an HMM model was generated with the hhmaccel program of the HHSuite toolkit. Sequences from the fungal database fuNOG were similarly aligned and HMM models generated. To annotate whole gene families, the hsearch program was used to obtain matches between the gene family and the funOG HMMs and only hits with a probability $\geq 0.99$ were considered. To annotate whole gene families, the hsearch program was used to obtain matches between the gene family and the fuNOG HMMs and only hits with a probability $\geq 0.99$ were considered.

**RCA sequencing.** The DNA-seq samples presented in Hiruma et al. and the other five genomes available for this genus. Evolution of the GGLs along the branches of a phylogenetic tree was modelled as a continuous time Markov process using a binary character alphabet corresponding to gene family presence or absence. Default parameters were used, corresponding to a transition model that allows varying GGL rates across gene families. We approximated the total number of gene families that were gained or lost on a branch by summing up the individual posterior probabilities for each gene family to be gained or lost on that branch and rounding this number to the closest integer. The number of gene families either gained or lost (annotated with one specific category) was computed as the total number of gene families crossed the branch of the tree. The significance was assessed using Fisher’s exact test and FDR corrected.

**RCA sequencing.** The DNA-seq samples presented in Hiruma et al. and the other five genomes available for this genus. Evolution of the GGLs along the branches of a phylogenetic tree was modelled as a continuous time Markov process using a binary character alphabet corresponding to gene family presence or absence. Default parameters were used, corresponding to a transition model that allows varying GGL rates across gene families. We approximated the total number of gene families that were gained or lost on a branch by summing up the individual posterior probabilities for each gene family to be gained or lost on that branch and rounding this number to the closest integer. The number of gene families either gained or lost (annotated with one specific category) was computed as the total number of gene families crossed the branch of the tree. The significance was assessed using Fisher’s exact test and FDR corrected.

**Statistical analysis of differential gene expression.** All statistical analyses of plant and fungal gene expression were performed in R (codes are available upon request). For the analyses of plant gene expression, genes with less than 100 mapped fragments in total (that is, across all the analysed samples) were ranked as ‘not expressed’ and therefore excluded. For analyses of fungal gene expression, we excluded genes that were not sufficiently expressed in the in planta samples, that is, genes with less than 100 (Ct, 24 samples) or less than 50 (Ci, 6 samples) mapped fragments across all the analysed samples. Subsequently, the count data for all expressed genes was TMN-normalized and log-transformed using the functions ‘calcNormFactors’ (R package EdgeR) and ‘voom’ (R package limma) to yield RNA-seq expression values. RNA-seq read counts were then transformed into a fragment count per gene per sample using the htseq-count script ($s = \text{reverse}, t = \text{exon}$) in the package HTSeq. The complete RNA-Seq data presented by Hiruma et al. and in this manuscript have been deposited under the GEO series accession number GSE70094.

**RT-qPCR analysis.** First-strand cDNA was synthesized from 1 µg DNase-treated total RNA using the iScript cDNA synthesis kit (Bio-Rad) and PCR amplification was performed using the iQ real-time PCR detection system (Bio-Rad). For each gene, specific primers were designed with the Primer 3 and AmpliFx programs. BLASTN searches against the *C. orbiculare* and *A. thaliana* genomes were performed to rule out cross-annealing artefacts. Gene expression levels were also co-expressed in other Arabidopsis expression data sets were identified using ATTED-II [http://atted.jp] and co-expression networks were generated using Cytoscape [version 3.1.1].
Microscopy methods. For cytology experiments, surface-sterilized A. thaliana Col-0 seeds were inoculated with either C5 or C1 conidia (5 x 10^5 spores ml^-1). The seeds were then transferred to half-strength Murashige and Skoog agar medium without sucrose and low-phosphate content (50 μM). Inoculated plants were grown at 22 °C with a 10-h photoperiod (80 μE m^-2 s^-1) for 1 to 24 days. The roots were either mounted in water for viewing GFP or first stained with Calcofluor white (0.01 %, Sigma) or fluorescein diacetate (10 μg ml^-1, Sigma). For visualization of GFP and FDA fluorescence, we used an Olympus FV1000 confocal microscope equipped with dry x 20 and x 40 objectives, using the 488-nm line of an Argon laser for excitation and fluorescence was collected at 490–520 nm. For imaging Calcofluor fluorescence, we used a Zeiss Axioskop epifluorescence microscope (filter set BP 365, FT 395, LP 397).

References

1. Rodríguez, R. J., White, J. F. J., Arnold, A. E. & Redman, R. S. Fungal endophytes: diversity and functional roles. New Phytol. 182, 314–330 (2009).

2. O’Connell, R. J. et al. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. Nat. Genet. 44, 1060–1065 (2012).

3. Hyde, K. D. et al. Colletotrichum—names in current use. Fungal Divers. 39, 147–189 (2009).

4. Sukno, S. A., Garcia, V. M., Shaw, B. D. & Thon, M. R. Root infection and mechanisms of symbiosis genes in mycorrhizal mutualists. Curr. Opin. Plant Biol. 17, 45–52 (2014).

5. Goetz, M. et al. Fungal endophytes in potato roots studied by traditional isolation and cultivation-independent DNA-based methods. FEMS Microbiol. Ecol. 58, 404–413 (2006).

6. Keim, J., Mishra, B., Sharma, R., Ploch, S. & Thines, M. Root-associated fungi of Arabidopsis thaliana and Microbispora perfusiformis. Fungal Divers. 66, 99–111 (2014).

7. Gan, P. et al. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of Colletotrichum fungi. New Phytol. 197, 1234–1249 (2012).

8. Hirma, K. et al. Root endophyte Colletotrichum tospofallae confers plant fitness benefits that are phosphate-status dependent. Cell 165, 1–11 (2016).

9. Delaux, P. M., Delmotte, M., Trouvelot, A., Hacquard, S. & Hance, I. Comparative genomics uncovers the impact of symbiotic associations on host genome evolution. PLoS Genet. 10, e1004487 (2014).

10. García, E., Alonso, A., Platas, G. & Saracín, S. The endophytic mycobiota of Arabidopsis thaliana. Fungal Divers. 60, 71–81 (2013).

11. Sato, T. et al. Anthracnose of Japanese radish caused by Colletotrichum tofieldiae. Curr. Opin. Plant Biol. 17, 357–359 (2012).

12. Stukenbrock, E. H., Christiansen, F. B., Hansen, T. T., Dutheil, J. Y. & Schierup, M. H. Fusion of two divergent fungal individuals led to the recent divergence of a nonpathogenic, endophytic mutualist. Genome Res. 23, 1648–1656 (2012).

13. Hacquard, S. et al. Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. Proc. Natl. Acad. Sci. USA 109, 10954–10959 (2012).

14. Li, H. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).

15. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 49 (2011).

16. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

17. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

18. Gruber, O. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676 (2005).

19. Powell, S. et al. eggNOG: an extended orthology inference approach. BMC Bioinformatics 6, 3 (2005).

20. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652 (2011).

21. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676 (2005).

22. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of orthologous groups for eukaryotic genomes. Genome Res. 13, 2178–2189 (2003).

23. Heberle-Bors, E. et al. The transcription factor PHR1 as a major regulator of metabolic changes during Arabidopsis thaliana growth of infection units. Plant Physiol. 178, 1783–1794 (2008).

24. Korf, I. Gene finding in novel genomes. Nucleic Acids Res. 37, 1639–1645 (2009).

25. Li, L. et al. D231–D239 (2014).
Acknowledgements

This work was primarily supported by the Max Planck Society (S.H., B.K., R.G.-O., E.V.Lv.T., E.K., P.S.-L. and R.J.O.), the Agence Nationale de la Recherche grant ANR-12-CHEX-0008-01 (R.J.O.), the European Research Council advanced grant (ROOTMICROBOTA) (P.S.-L.), the ‘Cluster of Excellence on Plant Sciences’ program (funded by the Deutsche Forschungsgemeinschaft; R.G.-O., E.K., A.C.M, P.S.-L.). Other funding sources included the German Center for Infection Research, DZIF (P.C.M.) and the Japan Society for the Promotion of Science (Post-doctoral Fellowship for Research Abroad; K.H.). We are grateful to Lisa Vaillancourt (University of Kentucky) and Pedro Crous (CBS-KNAW, Utrecht) for providing fungal cultures. We thank the staff of the Max Planck Genome Centre Cologne for their expertise in generating the sequencing data. We thank Anthony Levasseur (Aix-Marseille University) for assistance with annotation of CAZyme genes and Antonios Zampounis for bioinformatic assistance with secondary metabolism gene analysis.

Author contributions

P.S.-L., R.J.O. and K.H. initiated the project. S.H., R.J.O. and P.S.-L. coordinated the project. B.K., E.K. and A.C.M. coordinated the bioinformatics. K.H. performed the inoculation experiments and isolated the DNA/RNA. B.K, E.V.Lv.T. and E.K. performed the sequence assembly and annotation. B.K. and E.K. performed the gene predictions and managed the data. U.D., R.G.-O. and M.R.T. performed the phylogeny and evolution analyses. B.K. and R.G.-O. performed the comparative genomics and MCL analyses. M.R.T analysed the repeats. B.K. and S.H. analysed RNA-Seq data. S.H., B.K., R.J.O., J.-F.D., M.H., B.H. and O.L. annotated gene families. A.W. and A.C.M performed ancestral genome reconstruction. P.C.M. and A.C.M performed dN/dS analysis. B.K. performed SNP analysis. S.S. contributed C. tofieldiae isolate 0861 and performed soil analyses. S.H., B.K., A.W., P.M., U.D., O.L., J.-F.D., B.H., M.R.T., R.G.-O., A.C.M. and E.K. prepared the tables, figures and text. S.H, B.K, R.J.O and P.S.-L wrote and edited the paper.

Additional information

Accession codes: The genome assemblies have been deposited at DDBJ/EMBL/GenBank with accession numbers LFIV01000000 (C0), LFIV01000000 (C0861), LFHR01000000 (CBS127), LFHS01000000 (CBS130), LFHP01000000 (CBS495), LFHRQ01000000 (CBS168). The RNA-Seq data have been deposited in the NCBI Gene Expression Omnibus under GEO Series accession number GSE70094.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Hacquard, S. et al. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat. Commun. 7:11362 doi: 10.1038/ncomms11362 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
Erratum: Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi

Stéphane Hacquard, Barbara Kracher, Kei Hiruma, Philipp C. Münch, Ruben Garrido-Oter, Michael R. Thon, Aaron Weimann, Ulrike Damm, Jean-Félix Dallery, Matthieu Hainaut, Bernard Henrissat, Olivier Lespinet, Soledad Sacristán, Emiel Ver Loren van Themaat, Eric Kemen, Alice C. McHardy, Paul Schulze-Lefert & Richard J. O’Connell

Nature Communications 7:11362 doi: 10.1038/ncomms11362 (2016); Published 6 May 2016; Updated 29 Sep 2016

The HTML version of this Article previously published omitted Supplementary Data 8. This has now been corrected in the HTML version of the Article; the PDF was correct from the time of publication.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
© The Author(s) 2016