











# Smut infection of perennial hosts: the genome and the transcriptome of the Brassicaceae smut fungus *Thecaphora thlaspeos* reveal functionally conserved and novel effectors

Kaitlyn J. Courville<sup>1\*</sup>, Lamprinos Frantzeskakis<sup>1\*</sup> , Summia Gul<sup>1</sup>, Natalie Haeger<sup>1</sup>, Ronny Kellner<sup>1,2</sup> , Natascha Heßler<sup>1</sup>, Brad Day<sup>3</sup> , Björn Usadel<sup>4</sup> , Yogesh K. Gupta<sup>5</sup> , H. Peter van Esse<sup>5</sup> , Andreas Brachmann<sup>6</sup> , Eric Kemen<sup>2</sup> , Michael Feldbrügge<sup>1</sup>  and Vera Göhre<sup>1</sup> 

<sup>1</sup>Institute for Microbiology, Cluster of Excellence on Plant Sciences, Heinrich-Heine University, Building 26.12.01, Universitätsstr. 1, Düsseldorf 40225, Germany; <sup>2</sup>Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, Cologne 50829, Germany; <sup>3</sup>Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI 48824-6254, USA; <sup>4</sup>Unit of Botany and Molecular Genetics, Institute for Biology I, BioSC, RWTH Aachen University, 52074 Aachen, Germany; <sup>5</sup>The Sainsbury Laboratory, Norwich, NR4 7UH, UK; <sup>6</sup>Faculty of Biology, Genetics, Ludwig-Maximilians-Universität München, Großhaderner Str. 2-4, Planegg-Martinsried 82152, Germany

## Summary

Author for correspondence:

Vera Göhre

Tel: +49 211 81 11529

Email: vera.gohre@uni-duesseldorf.de

Received: 21 December 2018

Accepted: 9 January 2019

New Phytologist (2019) 222: 1474–1492

doi: 10.1111/nph.15692

**Key words:** effector, fungal endophyte, genome sequencing, infection, plant immune responses, RNA-seq.

- Biotrophic fungal plant pathogens can balance their virulence and form intricate relationships with their hosts. Sometimes, this leads to systemic host colonization over long time scales without macroscopic symptoms. However, how plant-pathogenic endophytes manage to establish their sustained systemic infection remains largely unknown.
- Here, we present a genomic and transcriptomic analysis of *Thecaphora thlaspeos*. This relative of the well studied grass smut *Ustilago maydis* is the only smut fungus adapted to Brassicaceae hosts. Its ability to overwinter with perennial hosts and its systemic plant infection including roots are unique characteristics among smut fungi.
- The *T. thlaspeos* genome was assembled to the chromosome level. It is a typical smut genome in terms of size and genome characteristics. *In silico* prediction of candidate effector genes revealed common smut effector proteins and unique members. For three candidates, we have functionally demonstrated effector activity. One of these, *TtTue1*, suggests a potential link to cold acclimation. On the plant side, we found evidence for a typical immune response as it is present in other infection systems, despite the absence of any macroscopic symptoms during infection.
- Our findings suggest that *T. thlaspeos* distinctly balances its virulence during biotrophic growth ultimately allowing for long-lived infection of its perennial hosts.

## Introduction

The *Thecaphora thlaspeos*-Brassicaceae pathosystem is a remarkable example of a sustained systemic plant–microbe interaction. *T. thlaspeos* establishes an infection of the entire plant, which can be maintained over several years (Vanky *et al.*, 2008; Frantzeskakis *et al.*, 2017). After penetration, intercellular hyphae of *T. thlaspeos* proliferate along the vasculature throughout the entire plant without visible impact on plant development. When the host plant develops siliques each year, fungal hyphae differentiate into spores that replace the developing seeds. In addition, fungal hyphae keep proliferating in the newly growing vegetative tissue. The capability of *T. thlaspeos* to overwinter with its perennial hosts and sustain the systemic infection within the entire plant is a unique characteristic among smut fungi studied to date.

*T. thlaspeos* is a relative of the well studied grass smut *Ustilago maydis*, which is adapted to Brassicaceae hosts (Vanky

*et al.*, 2008; Frantzeskakis *et al.*, 2017). Closely related sister species of *T. thlaspeos* comprise devastating crop pathogens such as *T. solani* on potato (up to 85% losses, Andrade *et al.*, 2004) or *T. frezii* on peanut (Andrade *et al.*, 2004; Conforto *et al.*, 2013). In addition to its *Arabidopsis* hosts, *T. thlaspeos* can colonize the model plant *Arabidopsis thaliana*. Therefore, the *T. thlaspeos*-Brassicaceae pathosystem benefits from the well developed resources of *A. thaliana* research that overcome experimental constraints of grass smuts due to the genetic complexity of their hosts (Frantzeskakis *et al.*, 2017). While plant–fungus interactions of pathogens and symbionts are well studied (Gutjahr & Parniske, 2013; Lo Presti *et al.*, 2015), the molecular mechanisms that enable *T. thlaspeos* to establish and maintain its remarkably long biotrophic interaction with Brassicaceae over years are completely unknown. A deeper understanding of this pathosystem therefore might unveil molecular processes related to the endophytic phase of fungal infections.

\*These authors contributed equally to this work.

Biotrophic pathogens have evolved distinct mechanisms to evade plant immunity and establish genetic interactions with their host (Brefort *et al.*, 2009). During invasion, plant cell wall-degrading enzymes are secreted, which promote fungal penetration of the plant cell (Choi *et al.*, 2013). Subsequently, fungal hyphae proliferate inside the apoplast and/or grow through host cells, establishing an intimate contact zone for the exchange of nutrients and proteins. Functional genomic analyses of the grass smut fungi *U. maydis*, *Sporisorium reilianum* and *U. hordei* have greatly contributed to our understanding of smut infection and the associated host responses (Kämper *et al.*, 2006; Brefort *et al.*, 2009; Ghareeb *et al.*, 2015; Lanver *et al.*, 2018). In short, these studies have revealed different repertoires of conserved and host-adapted effector proteins (Okmen & Doehleemann, 2014; Lanver *et al.*, 2017). In *U. maydis* and *S. reilianum*, effector-encoding genes are clustered as exemplified by a locus of 26 genes named 'Cluster 19A' (Kämper *et al.*, 2006; Schirawski *et al.*, 2010). When the entire cluster is deleted, tumor formation in maize is impaired and ultimately spore formation is defective (Kämper *et al.*, 2006). However, clustering of effector genes is not always conserved, as exemplified by *U. bromivora*, the false brome (*Brachypodium* sp.) smut (Rabe *et al.*, 2016). In addition, functional analysis in *U. maydis* confirmed the contribution of single effector proteins to fungal virulence (Lanver *et al.*, 2017). For example, Pep1, a protein essential for fungal penetration, was initially identified outside of the effector clusters and was characterized as an apoplastic peroxidase inhibitor (Doehleemann *et al.*, 2009; Hemetsberger *et al.*, 2012), which is conserved in several grass smut species as well as the dicot-infecting smut *Melanopsichium pennsylvanicum* (Hemetsberger *et al.*, 2015).

In response to fungal colonization, plants have evolved mechanisms to inhibit pathogen infection and proliferation (Dodds & Rathjen, 2010). To detect invading pathogen, plants deploy two major strategies. First, plasma membrane-located pattern-recognition receptors (PRR) recognize conserved microbial elicitors, called pathogen-associated molecular patterns (PAMPs), and induce PAMP-triggered immunity (PTI, Zipfel, 2014). Second, pathogen effector molecules are recognized by intracellular host nucleotide-binding leucine-rich repeat (NLR) immune receptors that induce effector-triggered immunity (ETI, Gassmann & Bhattacharjee, 2012; Bialas *et al.*, 2017). PTI and ETI involve similar immune responses including the activation of signaling cascades, massive transcriptional reprogramming, and the accumulation of the two defense hormones, salicylic acid and jasmonic acid (Thomma *et al.*, 1998; Tsuda & Somssich, 2015). In addition, and specific to R-protein activation, ETI induces a local programmed cell death response referred to as the hypersensitive response, as well as systemic acquired resistance (Giraldo & Valent, 2013; Lo Presti *et al.*, 2015).

While factors conferring resistance to smut infection are of agronomic importance, to date, the only known resistance gene is the maize wall-associated kinase *ZmWAK* which protects maize against the head smut *S. reilianum* (Zuo *et al.*, 2014). In addition, the barley smut *U. hordei* encodes three dominant avirulence genes, but the corresponding resistance genes remain undiscovered (Linning *et al.*, 2004). In the current study, we investigate

the systemic and long-lasting smut infection in Brassicaceae. Using a combination of genomic DNA and RNA sequencing of the recently described smut fungus *T. thlaspeos*, we present a functional characterization of its first effector candidates. These give a first insight into how *T. thlaspeos* balances its virulence during biotrophic growth and provide an inventory of effector candidates for future studies.

## Materials and Methods

### Cloning of expression vectors

Standard USER cloning procedures (NEB) were followed to generate the AvrRPS4-*TnNlp1* construct. *TnNlp1* was amplified from cDNA from start codon to stop codon excluding the signal peptide and inserted in frame after the AvrRPS4 leader sequence in the pEDV3 expression vector (Sohn *et al.*, 2007). Standard Golden Gate cloning (Engler *et al.*, 2014; Patron *et al.*, 2015) procedures were followed to generate binary expression vectors for *in planta* expression. From left border to right border, expression cassettes contained kanamycin resistance, an olesin *AtOLE1*-RFP protein fusion (Shimada *et al.*, 2010), and the *T. thlaspeos* effector gene controlled by the cauliflower mosaic virus (CaMV) 35S promoter and g7 terminator. Standard Gateway cloning procedures were used to generate pEarley Gate 103-*TnNlp1*-Gfp and pEarley Gate 103-*PsojNIP*-Gfp expression vectors (Karimi *et al.*, 2002; Qutob *et al.*, 2002; Earley *et al.*, 2006). Cloning of *Tipep1* into plasmid p123-pep1 (Aichinger *et al.*, 2003) and transformation into solo-pathogenic strain SG200 $\Delta$ pep1 was carried out according to Hemetsberger *et al.* (2015).

### Strains, transgenic *A. thaliana* lines, and infection assays

*Pseudomonas syringae* pv. *tomato* DC3000-LUX (*Pst-LUX*) was transformed with pEDV3-*TnNlp1* and pEDV3-empty (Katagiri *et al.*, 2002). Four-wk-old plants were spray inoculated with the bacterial strains as described in Fabro *et al.* (2011). At 3 d post infection, total photon counts – a measure of *Pst-LUX* growth – were quantified and normalized to the foliar area or leaf fresh weight.

Transgenic *A. thaliana* Col-0 effector-expressing lines were generated via the floral dipping method using the *A. tumefaciens* AGL1 strain expressing the effector constructs (Koncz & Schell, 1986). Primary *A. thaliana* transformants (T1) for two independently transformed lines per effector were selected based on RFP-marker-fluorescence of the seeds (Shimada *et al.*, 2010). Rosette areas were measured 4-wk post sowing. One leaf per plant was harvested for RNA extraction.

Transient expression of *TnNlp1*-Gfp and *PsojNIP*-Gfp in *Nicotiana benthamiana* was assessed using a Zeiss LSM780 confocal microscope (Bleckmann *et al.*, 2010) 2 d after infiltration of *A. tumefaciens* strain GV3101 (pMP90 RK) containing the respective effector. All bacterial strains in this study were grown overnight at 28°C in Luria-Bertani (LB) medium.

*U. maydis* growth, infection of maize, and microscopy of maize infection was performed as previously described (Hemetsberger *et al.*, 2015; Bösch *et al.*, 2016). *T. thlaspeos* infection of *Arabidopsis*

*hirsuta* was performed by co-germination of seeds and spores on soil (Frantzeskakis *et al.*, 2017).

### *T. thlaspeos* genome assembly, annotation and comparative genomics

For genomic DNA (gDNA) sequencing of the *T. thlaspeos*, high-molecular-weight gDNA was prepared from pure cultures using phenol extraction (Bösch *et al.*, 2016). LF1 gDNA was sequenced by PacBio long-read sequencing (P6-C4, Max Planck Genome Centre, Cologne, Germany) and by Illumina short-read sequencing (2 × 300 bp; Illumina MiSeq, v3 chemistry, Genomics Service Unit at the Biocenter of Ludwig-Maximilians University, Munich, Germany). Long reads were assembled with CANU v.1.3 (Koren *et al.*, 2017) and short reads trimmed with TRIMMOMATIC v.0.32 (Bolger *et al.*, 2014) were used with Pilon (Walker *et al.*, 2014) for error correction. LF2 gDNA was sequenced by short-read sequencing (2 × 150 bp; Illumina HiSeq, Biomedical Research Center, HHU). The LF2 short reads were assembled using SPAdes v.3.8.0 (Bankevich *et al.*, 2012). REPEATMASKER v.4.0.5 was subsequently used to report and mask repetitive regions in the genome (Jurka *et al.*, 2005; Tempel, 2012).

Annotation of both genomes was performed using MAKER2 (Holt & Yandell, 2011) as previously described (Campbell *et al.*, 2014). Briefly, for the LF1 genome, an annotation was generated providing as evidence to MAKER assembled transcripts of LF1 in nutrient-rich culture conditions (Complete Medium; Holliday, 1961), proteomes of several Ustilaginales species (Supporting Information Table S1), and data from the UniProt protein reference database. After two iterations, 397 gene models were manually curated and used to train AUGUSTUS v.3.0.3 (Stanke & Morgenstern, 2005) and SNAP v2006-07-28 (Korf, 2004). For assessing the completeness of the datasets BUSCO v1.1b1, was used (Simão *et al.*, 2015).

Functional annotation was carried out using INTERPROSCAN 5.19 (Jones *et al.*, 2014), dbCAN (Yin *et al.*, 2012) and ANTI-SMASH v.4.0 (Weber *et al.*, 2015) were used to mine the genome for CAZymes and secondary metabolism-related genes. Genome to genome alignments were performed using MUMMER v.3.23 (Delcher *et al.*, 2003) using default user settings and the results were processed using auxiliary scripts provided with the package (e.g. SHOW-COORDS, DNADIFF). Search for orthologues between the Ustilaginales genomes used here (Table S1) and the generation of a multilocus based phylogeny tree was done utilizing ORTHOFINDER v.1.1.2 (Emms & Kelly, 2015).

### Data availability

The data generated were deposited in ENA (PRJEB24478).

### Quantitative RNA sequencing

Samples from LF2, *T. thlaspeos* spore-infected *Ar. hirsuta* (spores and seed collected in Ronheim, Germany in 2015), and healthy *Ar. hirsuta* were snap frozen in liquid nitrogen. Leaves from *A. thaliana* Tuel lines were harvested following phenotyping.

Total RNA was extracted using the RNeasy Plant Mini kit (Qiagen) including a DNaseI treatment (NEB). cDNA for RT-PCRs was generated using the Protoscript II First Strand cDNA Synthesis kit (NEB) and cDNA libraries were generated using the TruSeq RNA Library Prep kit v2 (Illumina) and sequenced on an Illumina HiSeq 3000 platform (Biomedical Research Center, HHU).

RNA-seq data for *Ar. hirsuta* were assembled using Trinity (Grabherr *et al.*, 2011). The transcript models were added to the ones derived from the genome release. Subsequently, *Ar. hirsuta* RNA-seq data from infected plants were mapped against this transcriptome set using BOWTIE (Langmead *et al.*, 2009). Reads that did not map to *T. thlaspeos* were retained and merged with RNA-seq data from healthy *Arabidopsis* plants. This combined set was then used to generate a transcriptome using Trinity either using a relatively standard pipeline or correcting errors in the reads using Rcorrector (Song & Florea, 2015) and assembling the data using minimal coverage of 2. Both assemblies were filtered and analyzed using transrate (Smith-Unna *et al.*, 2016). As the standard approach yielded better transrate values, the resulting transrate filtered standard Trinity assembly was used in the subsequent analysis. These *Ar. hirsuta* gene models were pooled with those from the *T. thlaspeos* genome assembly and all RNA-seq data from healthy and infected *Ar. hirsuta* plants were mapped against this combined set using subread. The data were summarized using EXPRESS (Roberts *et al.*, 2011), but only uniquely mapped reads were extracted. Data were split and separately analyzed for the plants using EDGER (Robinson *et al.*, 2009). Gene ontology enrichment assessment was carried out using GORILLA (Eden *et al.*, 2009) and visualizations were generated with REVIGO (Supek *et al.*, 2011).

For the analysis of the fungal transcriptome short reads were mapped to the genome using STAR 2.5.2 (Dobin *et al.*, 2013), tables with raw read counts were parsed and analyzed with DESEQ2 (Love *et al.*, 2014). For the analysis of the Tuel line, short reads were mapped to the genome using HISAT2 (Kim *et al.*, 2015), and analyzed with STRINGTIE (Pertea *et al.*, 2016) and DESEQ2.

## Results

### Assembly and annotation of *T. thlaspeos* LF1 and LF2 genomes

To assemble the reference genome for *T. thlaspeos*, gDNA from the haploid strain LF1 of the mating type *alb1* (Frantzeskakis *et al.*, 2017) was sequenced using both long-read (PACBIO, *c.* 40 × coverage) and short-read (Illumina MiSeq, *c.* 53 × coverage) platforms. The two approaches resulted in 332 950 single long reads and 5433 377 paired short reads, respectively. PACBIO long reads were assembled into 33 scaffolds and further polished using short Illumina reads. The resulting assembly is of high continuity, reaching chromosome level. The mitochondrial genome was fully assembled in a single scaffold of 108.2 kb (Fig. S1). Here, 19 out of the 32 nuclear scaffolds have telomeric repeats (TTAGGG) at both ends, five have repeats at one end (Tables 1, S2). Hence,

*T. thlaspeos* has at least 22 chromosomes, similar to its distantly related sister species *U. maydis* and other grass smuts featuring 23 chromosomes (Kämper *et al.*, 2006; Schirawski *et al.*, 2010; Rabe *et al.*, 2016). In parallel, a draft genome of the compatible mating type LF2 (*a2b2*) was assembled from short-read data (7795 622 paired-end reads, *c.* 107× coverage; Table 1).

For the gene annotation of strain LF1, we combined *ab initio* prediction, homology-based modeling using 21 smut fungal proteomes, and transcriptomic data from *T. thlaspeos* in the MAKER2 pipeline. The resulting 6509 gene models were manually curated using Apollo (Lee *et al.*, 2013) removing unsupported gene calls (absence of expression or protein homology evidence), giving a final dataset of 6239 high-confidence gene models. We then used these curated models to annotate the second strain LF2 with the MAKER2 pipeline, and generated 6504 gene models. This number is slightly lower than for the sequenced grass smuts (Rabe *et al.*, 2016). Verification of completeness in both LF1 and LF2 using BUSCO showed that the genomes contain *c.* 93% and the annotations *c.* 97% complete single-copy BUSCOs (Table S3). Hence, despite the fragmentation of the LF2 dataset, high gene space completeness was achieved.

Subsequent functional annotation focused on the high-confidence gene models predicted from LF1. 5093 of the 6239 protein models (81%) contain known domains (Table S4), and 355 genes were found to encode putative secreted proteins (Table S5). Interestingly, this is one-third less than predicted for some of the grass smut fungi (Table S6). Here, 267 of these secreted candidate proteins have at least one orthologue in species of the genus *Ustilago*, *Sporisorium* or *Melanopsichium*, and 200 are shared between all of them. Based on APOPLASTP, 63 of the secreted proteins are predicted to localize in the apoplast. As expected, this group comprises several predicted cell wall-degrading enzymes (CAZymes, proteases). Prediction of effectors using EFFECTORP resulted in 29 candidates including the conserved effector *pep1*, (THTG\_03661) and *ccel* (Seitner *et al.*, 2018) (Table S5). Other well described effectors, such as the chorismate dismutase *Cmu1* (Djamei *et al.*, 2011), the seedling efficient effector required for tumor induction *See1* (Redkar *et al.*, 2015), and the cysteine protease inhibitor *Pit2* (Doehlemann *et al.*, 2011) are missing in the *T. thlaspeos* genome, suggesting a reduced overlap between *T. thlaspeos* and grass smut effectors.

**Table 1** Assembly statistics.

	LF1	LF2
Number of scaffolds	32	537
Minimum size (bp)	17 833	82
1 <sup>st</sup> quartile (bp)	259 599	135
Median (bp)	594 082	354
Mean (bp)	643 487	38 053
3 <sup>rd</sup> quartile (bp)	863 537	480
Max (bp)	1 714 324	1 222 775
Total (bp)	20 591 595	20 434 990
N50 (bp)	863 537	347 457
N90 (bp)	456 084	93 318
N95 (bp)	382 946	51 399
GC content	61%	61%

As expected for a biotrophic smut fungus, the repertoire of carbohydrate-active enzymes (CAZymes; Huang *et al.*, 2017) is small (Table S7). *T. thlaspeos* carries several genes encoding for pectin degradation enzymes (GH53, PL3, PL4), which are absent from grass smut fungi (Table S7). These might reflect an adaptation to the pectin-rich cell wall of dicot host plants. Furthermore, the genome of *T. thlaspeos* lacks the secondary metabolite clusters known from *U. maydis* important for the production of ustilagic acid, itaconic acid or MELs, as well as the flocculosin gene cluster encoded in the closest relative *Anthracozygia flocculosa* (syn. *Pseudozyma flocculosa*), an epiphytic biocontrol yeast. Despite the overall agreement in whole-genome alignments between these species (see next paragraphs; Fig. 1a), in this specific locus synteny is lost (Fig. 1b; Teichmann *et al.*, 2007, 2011). Furthermore, ANTI-SMASH predictions (Weber *et al.*, 2015) did not reveal any novel clusters for secondary metabolites (Table S8).

Taken together, assembly and annotation delivered a high-quality dataset comparable with the well established genomes of *U. maydis* and *S. reilianum*. Based on our results, *T. thlaspeos* has a typical smut genome, which is small in size (*c.* 20 Mbp), is organized in 22–24 chromosomes and has a low repeat content mostly comprised of dinucleotide repeats (Table S9).

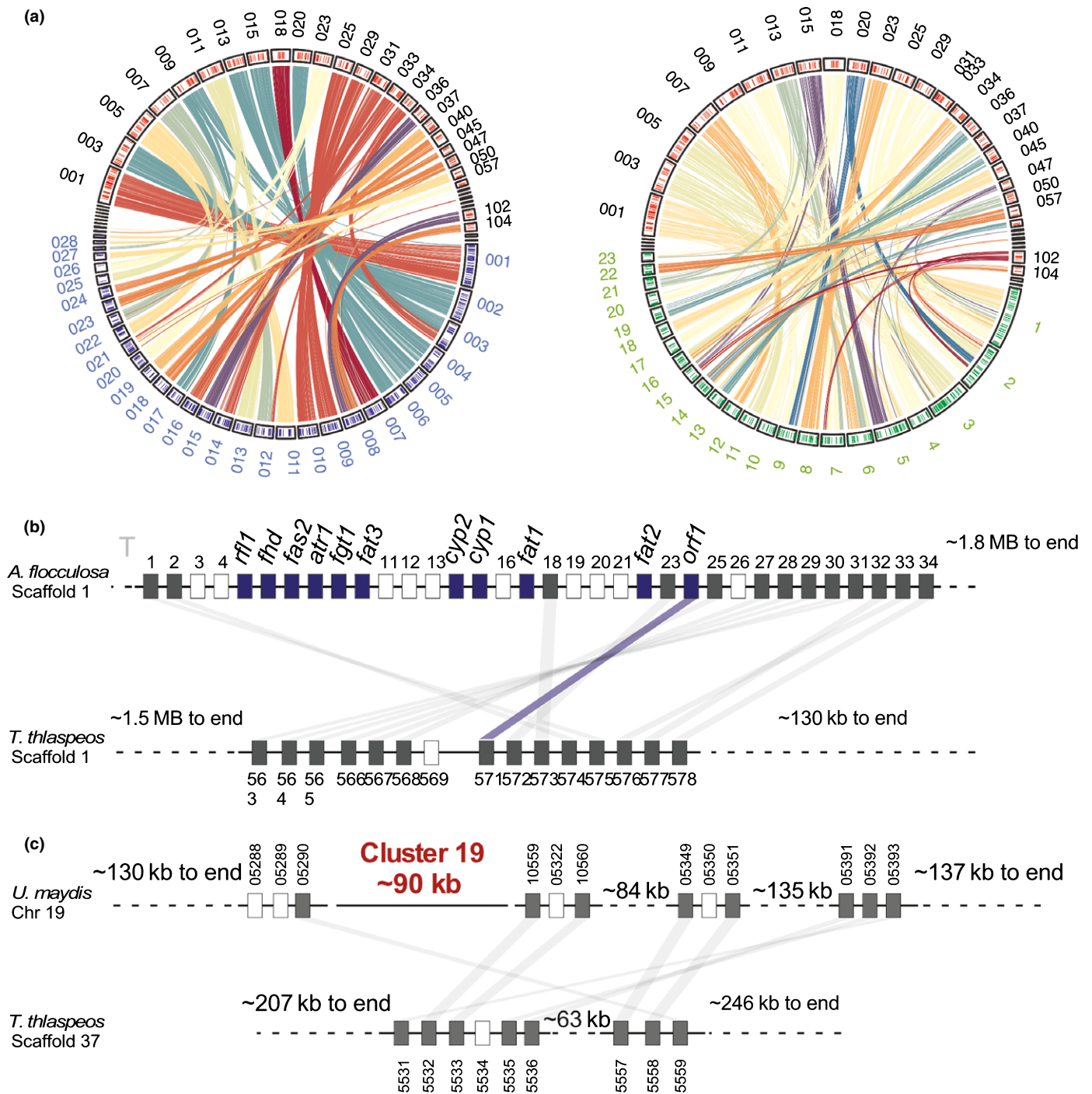
### Mating in *T. thlaspeos* relies on conserved molecular mechanisms

We previously described the presence of a mating system in *T. thlaspeos* (Frantzeskakis *et al.*, 2017). Now, the mating type loci *a1* and *a2* of LF1 and LF2 are assembled revealing massive rearrangements between *T. thlaspeos* (Fig. S2) and other smut fungi (Kellner *et al.*, 2011). In contrast with *U. maydis*, in *T. thlaspeos* the pheromone receptors *pra1* and *pra2* are not flanked by the border genes *lba* and *rba*. This rearrangement is conserved in *A. flocculosa* suggesting divergent evolution in the *Thecaphora* clade from the grass smuts. In grass smut fungi, the *a2* allele of the mating locus harbors the locus-specific genes *rga2* and *lga2*, which are important for uniparental inheritance of mitochondrial DNA (Fedler *et al.*, 2009). While *rga2* is present in the *a2* locus of *T. thlaspeos*, *lga2* is absent from the genome, similar to *Ustanciosporium gigantosporum*, the white beak-sedge smut (Kellner *et al.*, 2011).

Besides the mating locus, pheromone response elements (Urban *et al.*, 1996) (Fig. S2) and downstream signaling components of mating, such as the transcription factor *prf1* and genes involved in signaling via the cAMP pathway and the MAPK cascade, are conserved in *T. thlaspeos* (Table S10) (Feldbrügge *et al.*, 2004). Hence, the mating process of *T. thlaspeos* (Frantzeskakis *et al.*, 2017) appears to rely on the same molecular processes that are conserved in smut fungi.

### Intra- and interspecies comparison between *T. thlaspeos*, commensal and grass smut fungi

To scan genomic assemblies of 13 smut fungi species (< 100 scaffolds or N50 > 500 kb) for conserved synteny with *T. thlaspeos*



**Fig. 1** Synteny is higher between *Thecaphora thlaspeos* and *Anthracocystis flocculosa* than *T. thlaspeos* and *Ustilago maydis*. (a) Circos plots between *T. thlaspeos* and *A. flocculosa* (left) or *U. maydis* (right). Colored lines depict syntenic blocks larger than 2 kb. Outer ring depicts the location of secreted proteins in the corresponding scaffold or chromosome. Scaffolds of *A. flocculosa* are in blue, while scaffold of *U. maydis* are in green (b) Synteny of the flocculosin secondary metabolite cluster in *A. flocculosa* and *T. thlaspeos*. Blue boxes depict genes involved in flocculosin production, white boxes depict genes with no orthologues in the compared genome. (c) Synteny of the effector Cluster 19A between *U. maydis* and *T. thlaspeos*. White boxes depict genes with no orthologues in the compared genome.

we used whole-genome alignments (Fig. S3). Overall, *T. thlaspeos* scaffolds align best to *A. flocculosa* with an average alignment rate of 51.4% and an average similarity ranging from 74.2% to 78% (Fig. 1a; Table S11). By contrast, alignment rate and sequence similarity between *T. thlaspeos* and the model smut fungus

*U. maydis* drops to averages of 32.4% and 73.4%, respectively (Fig. 1a; Table S11).

Loss of synteny between genomes of fungal plant pathogens has been shown to increase with genetic distance and, moreover, involves genomic regions that are often enriched for virulence-

related genes (Raffaele & Kamoun, 2012). In the grass smuts, synteny breaks are almost exclusively found in so-called virulence gene clusters that encode effector genes with partly crucial virulence function (Schirawski *et al.*, 2010; Rabe *et al.*, 2016). In *T. thlaspeos*, except for Stp1 (UMAG\_02475) in cluster 5B (Schipper, 2009) and the nonvirulence-related Cluster 9A, we did not find any of the *U. maydis* virulence clusters (Table S12). In some cases, cluster-flanking genes were partially present and rearranged in *T. thlaspeos* as exemplarily shown for 'Cluster 19A' in Fig. 1(c).

To identify unique genes of *T. thlaspeos* which we hypothesize may functionally replace the missing effector clusters, we searched for shared orthologues. Out of the 145 061 genes included in the analysis, 93.4% (135 570) could be placed in 10 059 orthogroups. *T. thlaspeos* shares most orthogroups with *A. flocculosa* (Fig. 2a; Table S13). This close relationship is further supported by multitype locus phylogeny generated from 1307 single-copy orthologues (Fig. 2b), which clearly places *T. thlaspeos* and *A. flocculosa* separate from the grass smuts. Out of the 6239 *T. thlaspeos* predicted proteins, 233 have no orthologues in smut fungi. The majority of these genes (205 out of 233) encodes proteins of unknown function (Table S14); 44 of the unique proteins contain a predicted signal peptide (Fig. 3a), indicating that they might be involved in the interaction between *T. thlaspeos* and its host. Hence, we have generated a unique repertoire of *T. thlaspeos* specific candidate virulence-related genes. Indeed, two of the unique and secreted proteins carry a necrosis-inducing protein (NPP) domain, which is a ubiquitous effector protein of dicot plant pathogens (Oome & Van den Ackerveken, 2014).

Finally, genome comparison of the two *T. thlaspeos* strains LF1 and LF2 as expected showed overall a high degree of synteny, as well as 11 509 single nucleotide polymorphisms (SNPs). In total, we obtained 280 syntenic blocks with on average 75.3 kb and 99.8% identity. We only detected very few structural variations in the one-to-one alignments with insertions or gaps that are short in length (Table S15). One example is the mating type locus *a* (Fig. S2). In addition, 31 genes were found to have no orthologous sequences in one or the other isolate (Table S16). In particular, an effector candidate presented in this study, THTG\_04398, was identified in LF1 and not in LF2 (Fig. S4a). Interestingly, THTG\_04398 is also absent in isolates of *T. thlaspeos* collected in Hohe Leite, Germany suggesting that there might be population differences in the effector distribution (Fig. S4b). Additionally, THTG\_01646 is specific to LF1 and matches effector criteria (secreted, no functional annotation and orthology to other smut fungi), indicating that these candidate effectors could be isolate and/or mating-type specific.

### *In planta*-induced genes are enriched for unique, small, secreted proteins

To gain insight into the fully established biotrophic phase of *T. thlaspeos*, we conducted a whole transcriptome sequence experiment (RNA-seq) comparing *T. thlaspeos*-infected *Ar. hirsuta* rosette leaf tissue with axenic *T. thlaspeos* cultures and healthy

*Ar. hirsuta*. RNA of rosette leaves from 10-wk-old teliospore-infected and healthy plants as well as fungal culture was sequenced, resulting in *c.* 30 million reads for each sample (Fig. S5). The abundance of *T. thlaspeos* reads in infected samples was low (0.18–0.28%), which is in agreement with the early phase of maize infection by *U. maydis*, when the fungal hyphae have not started proliferating and the coverage is < 0.5% (Lanver *et al.*, 2018). *U. maydis* then proliferates massively at the local infection site, while *T. thlaspeos* grows along the vasculature, resulting in relatively low levels of fungal biomass. Despite the low coverage, we captured 988 genes expressed during infection (> 5 raw read count, averaged between four infection samples).

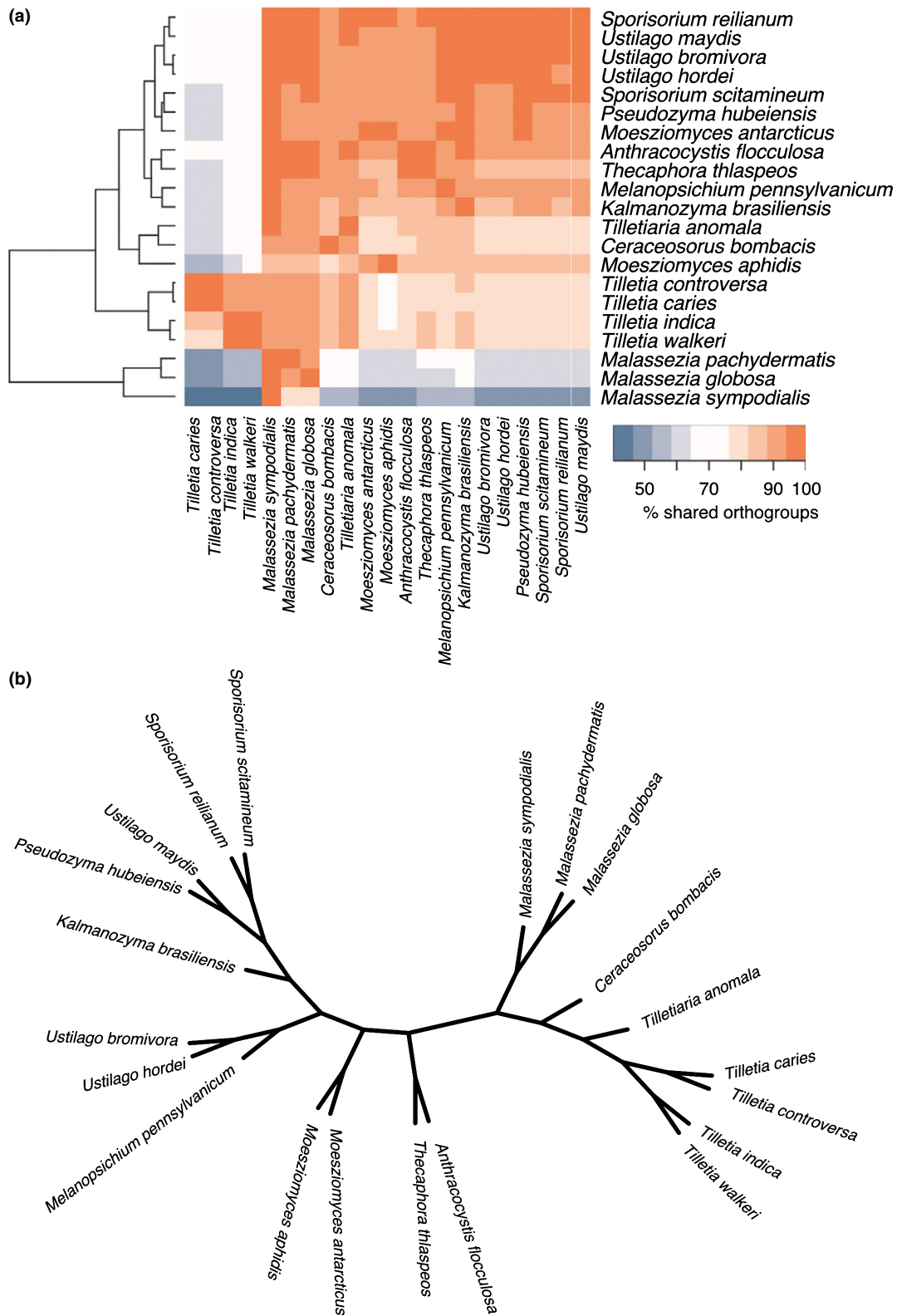
Here, 132 genes were found to be differentially expressed during infection ( $\log_2$  fold change >  $\pm 2$ , adjusted *P*-value < 0.05; Table S17). Among the induced genes, we identified several infection-related factors, including a necrosis and ethylene-inducing like protein (Nlp1, THTG\_00351), plant hydrolytic enzymes, a nutrition acquisition-related genes (e.g. an ammonium transporter; THTG\_03538), and a sugar transporter (THTG\_00350). However, more than half of the differentially expressed genes had no functional annotation. Of the induced genes, 51 were predicted to be secreted (Fig. 3b) and, of these, 19 are unique to *T. thlaspeos* (Table S18), which we named *Thecaphora*-unique effector candidates (Tue). Out of these 19 Tues, we confirmed the top 10 candidates based on upregulation during infection (Figs 3c, S4c) and further investigated them together with *Ttpep1* and *Ttnlp1* in heterologous expression systems.

### The *T. thlaspeos* orthologue of the conserved smut effector Pep1 is active in *U. maydis*

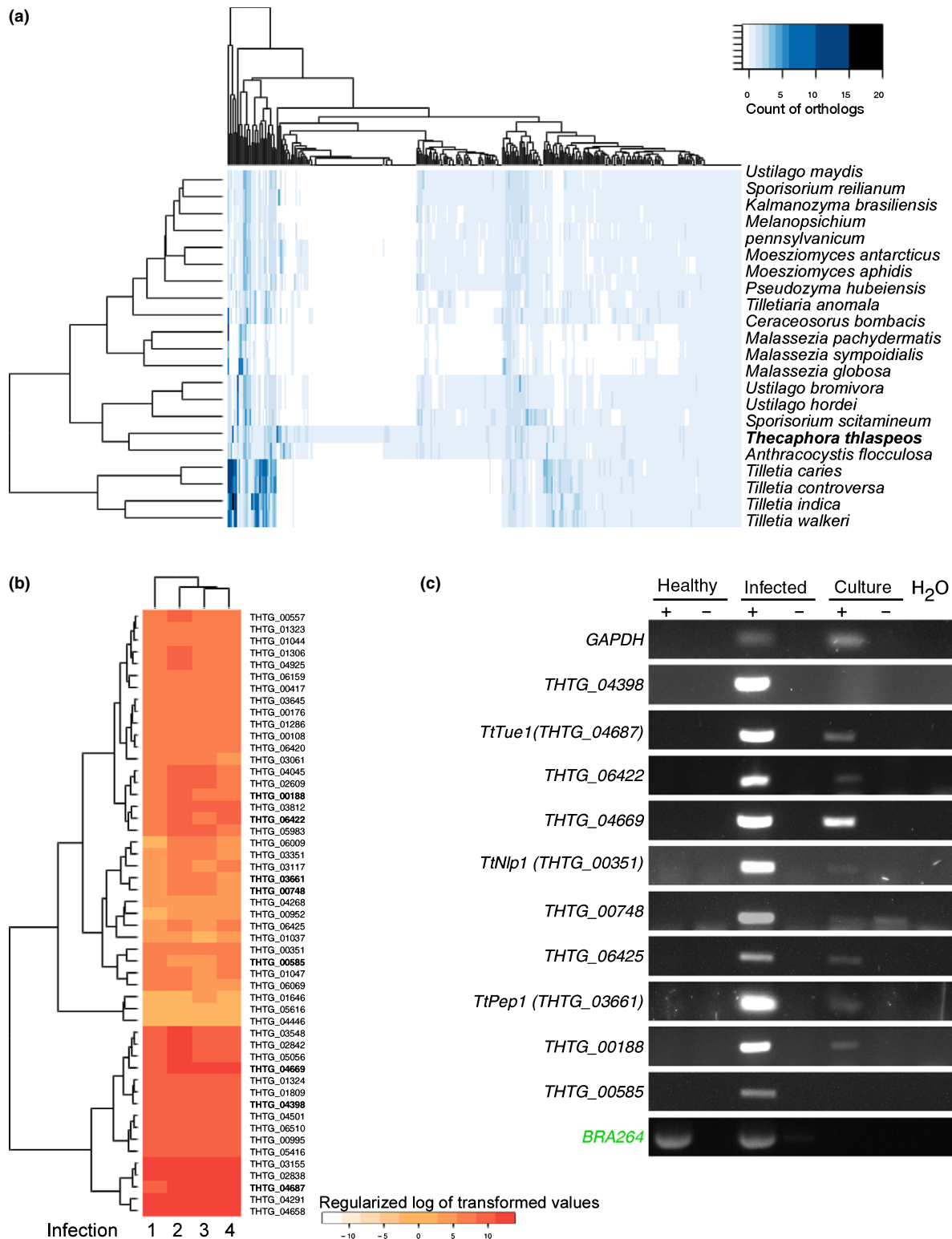
Pep1 is a peroxidase inhibitor that allows penetration by inhibiting apoplastic plant defense peroxidases (Doehlemann *et al.*, 2009; Hemetsberger *et al.*, 2012). As the function of this effector is conserved among grass smuts (Hemetsberger *et al.*, 2015; Rabe *et al.*, 2016), we tested whether also the *T. thlaspeos* orthologue can rescue virulence in *U. maydis*. Integration of *Ttpep1* into the *U. maydis* deletion strain SG200 $\Delta$ *pep1* partially complemented the infection phenotype in that tumors were formed in the leaves during seedling infection (Fig. 4a,b). Furthermore, *TtPep1*-mCherry is secreted into the apoplast (Fig. 4c) similar to *UmPep1*-mCherry (Doehlemann *et al.*, 2009). This suggests that *TtPep1* potentially targets an *Arabidopsis* peroxidase related to POX12 of *Z. mays* and thereby inhibits the apoplastic ROS burst also in Brassicaceae hosts.

### The *TtNlp1* is a noncytotoxic effector of *T. thlaspeos*

Nlp effectors are primarily found in genomes of pathogens that infect dicot plants, and therefore, the known grass smut fungi do not use such effectors. Their activity was mapped to a region comprising a highly conserved heptapeptide inducing necrosis (GHRHDWE; Schouten *et al.*, 2007; Ottmann *et al.*, 2009) and a 20 amino acid domain (nlp20) that induces immune responses such as ethylene production or ROS burst (Böhm *et al.*, 2014). A

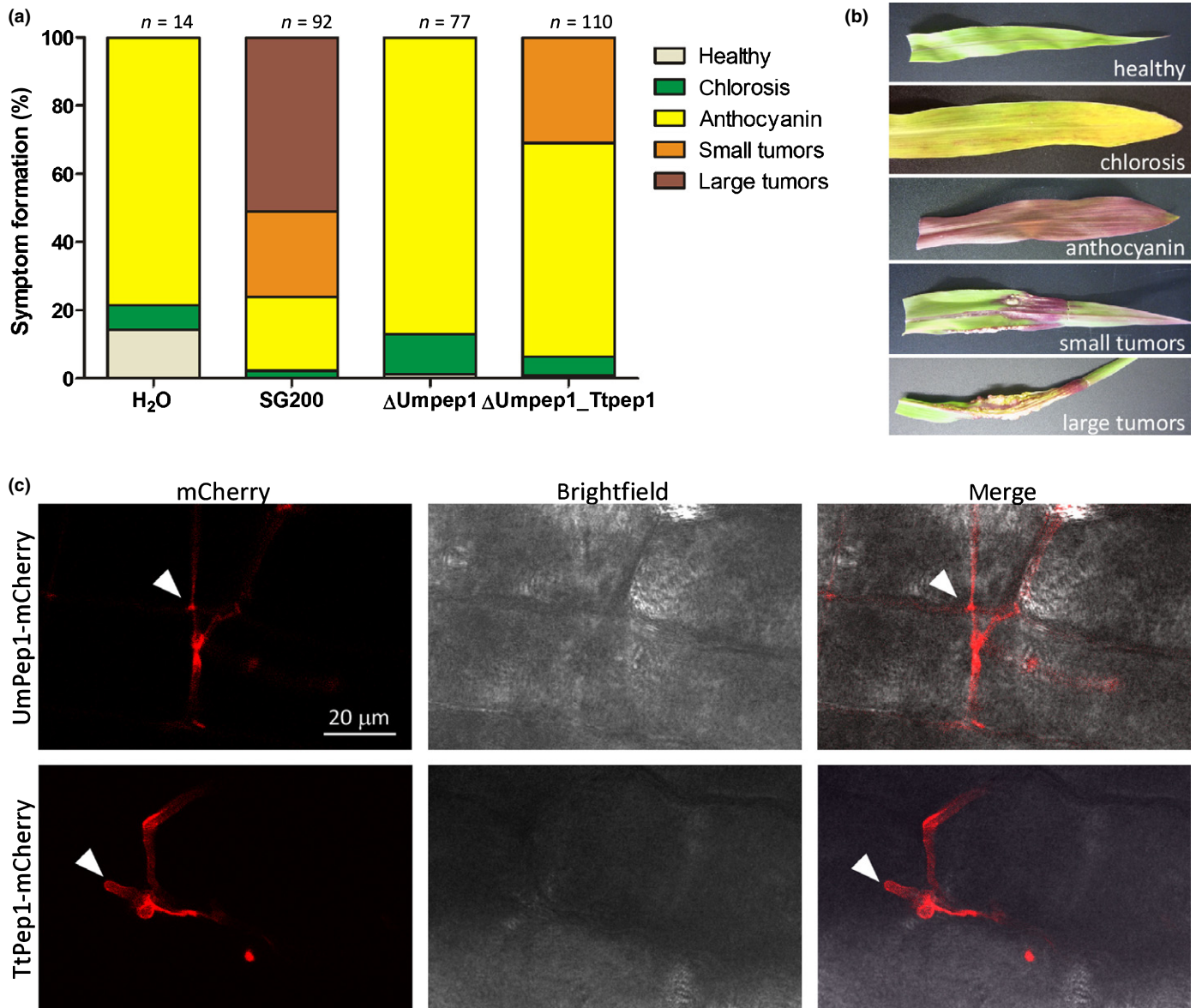


**Fig. 2** *Thecaphora thlaspeos* and *Anthracocystis flocculosa* are genetically separate from the grass smuts. (a) Orthology analysis of all *T. thlaspeos* predicted genes in comparison to the predicted genes of other Ustilaginales species. The heatmap depicts the % overlap of orthologous groups. Cladogram on the left is based on hierarchical clustering (Euclidean method). (b) Multilocus phylogeny of the Ustilaginales species used for the analysis based on 1307 single-copy orthologues.



**Fig. 3** Candidate effectors of *Thecaphora thlaspeos* are identified via differential expression analysis during infection and confirmed by RT-PCR. (a) Orthology analysis of the *T. thlaspeos* predicted secretome in comparison to the predicted secretomes of other Ustilaginales species. Cladogram on the left and on top is based on hierarchical clustering (Euclidean method). Color-coding depicts the amount of orthologues in other species for every *T. thlaspeos* secreted protein-coding gene. (b) Expression values of 51 differentially expressed secreted protein-coding genes during the infection of *Arabis hirsuta*. Each column represents a biological replicate. Cladogram on the left is based on hierarchical clustering (Euclidean method) and effector candidates verified by RT-PCR are highlighted in bold. Color code represents regularized log transformed values derived from the DESeq2 analysis. (c) Effector candidates have visibly higher mRNA accumulation during infection compared with in culture. Effector mRNA accumulation is normalized by that of *gapdh*. Plant marker *BRA264* (Stockenhuber *et al.*, 2015) was used to verify samples containing plant tissue cDNA. RT, Reverse transcriptase, reaction: 35 cycles.





**Fig. 4** *Ttpep1* partially complements deletion of *Umpep1* in maize infection. (a) Disease rating of 1-wk-old Early Golden Bantam maize plants 3 d post infection (dpi) with H<sub>2</sub>O (mock) and *Ustilago maydis* strains SG200\_Δ*pep1* (Δ*Umpep1*), SG200, and SG200\_Δ*pep1*\_Ttpep1 (Δ*Umpep1*\_Ttpep1). The values indicate the total number of plants infected in three independent experiments. (b) A representation of each disease category. (c) Confocal imaging of maize leaves infected with *UmPep1*-mCherry (SG200\_Δ*pep1*\_UmPep1-mCherry), top, and *TtPep1*-mCherry (SG200\_Δ*pep1*\_TtPep1-mCherry), bottom. Arrows indicate apoplastic regions into which *UmPep1* is secreted, as shown by Doehlemann *et al.* (2009), and therefore where *TtPep1* is also likely secreted.

second class of Nlp proteins are noncytotoxic and induce immune responses, but do not elicit HR-related cell death (Santhanam *et al.*, 2012; Oome *et al.*, 2014). To date, the role of the noncytotoxic class of Nlps during infection remains elusive.

In addition to the induced *Ttnlp1*, the genome of *T. thlaspeos* encodes two *nlp* genes with predicted NPP1-domains (Pfam accession PF05630): THTG\_00343 = *nlp2*, and THTG\_04815 = *nlp3*. *Ttnlp1* and *Ttnlp2* are located on scaffold 1 and have predicted signal peptides. This *nlp* locus contains additional genes with predicted signal peptides and, hence, might comprise the first *T. thlaspeos* effector gene cluster (Fig. 5a). *Ttnlp3*, located on scaffold 33, is substantially shortened at the N-terminus and

does not have a signal peptide. Amino acids in the necrosis-inducing heptapeptide are not conserved in the three *T. thlaspeos* proteins suggesting they are noncytotoxic (Fig. 5b). In line with these findings, transient expression assays in *N. benthamiana* confirmed that *TtNlp1* fails to cause necrosis (Fig. 5c). To test whether *TtNlp1* plays a role in virulence, we utilized *Pseudomonas syringae* pv. *tomato* DC3000-LUX (*Pst*-LUX) for pathogen-mediated delivery of *TtNlp1* into *A. thaliana*. As expected, this noncytotoxic protein does not cause HR. However, bacterial growth significantly increased in the presence of *TtNlp1*, suggesting a virulence function for this candidate effector (Fig. 5d,e). In the future, it will be important to confirm this

virulence function by generating deletion mutant strains of *T. thlaspeos*.

### *Thecaphora thlaspeos* unique effector 1 (*TtTue1*) is a novel virulence factor

To investigate effector function, stable expression in *A. thaliana* offers several advantages in that growth phenotypes and morphological alterations can be detected *in planta* (Germain *et al.*, 2017), and plant targets can be identified by interaction studies. Therefore, we successfully generated transgenic *A. thaliana* Col-0 lines for six of the top effector candidates as well as the bona-fide effector *TtNlp1* and monitored rosette size and color (Table 2). We did not observe growth or morphological alterations caused by the effector *TtNlp1*. This phenotype is similar to the noncytotoxic *H. arabidopsidis* *HpNlp1* (Oome *et al.*, 2014).

However, 4-wk-old rosettes of plants expressing the *T. thlaspeos* unique effector *TtTue1* (THTG\_04687) were significantly smaller than the control plants and displayed minor chlorosis (Fig. 6a,b). The other candidate effector lines overall resembled the control plants, even though the fungal effectors were expressed (Figs 6a,b, S6). To confirm virulence activity of *TtTue1*, we infected the transgenic lines with *PstLUX*. Bacterial proliferation was indeed increased in lines expressing *TtTue1* to the same level as in the *bak1-5* mutant (Chinchilla *et al.*, 2007; Fig. 6c) further supporting that this protein is an effector. The other lines permitted bacterial proliferation similar to wild-type levels.

To gain insight into the virulence function of *Tue1*, plant responses to *TtTue1* were detected by genome-wide transcriptome analysis of the *TtTue1* transgenic line. 105 genes were differentially expressed ( $\log_2$  fold change  $>+/- 2$ , FDR  $<0.05$ ; Table S19). The 93 induced genes fell mainly into Gene Ontology (GO) biological processes that comprise defense responses and responses to different stress stimuli (Fig. 6d). As expected, among these genes are negative regulators of defense such as *IDL7* (Vie *et al.*, 2017) or *SYP122* (Zhang *et al.*, 2007), but also genes involved in defense such as *RBOHD* and *ERF1*. Interestingly, only a few genes were downregulated, and four out of 12 repressed genes were related to cold acclimation, in particular the COR15 complex, which protects the chloroplast envelop against freezing (Thalhammer *et al.*, 2014). In summary, we could confirm *TtTue1* as the first novel virulence factor specific to *T. thlaspeos* with a novel link to cold acclimation, which we will functionally characterize in the future.

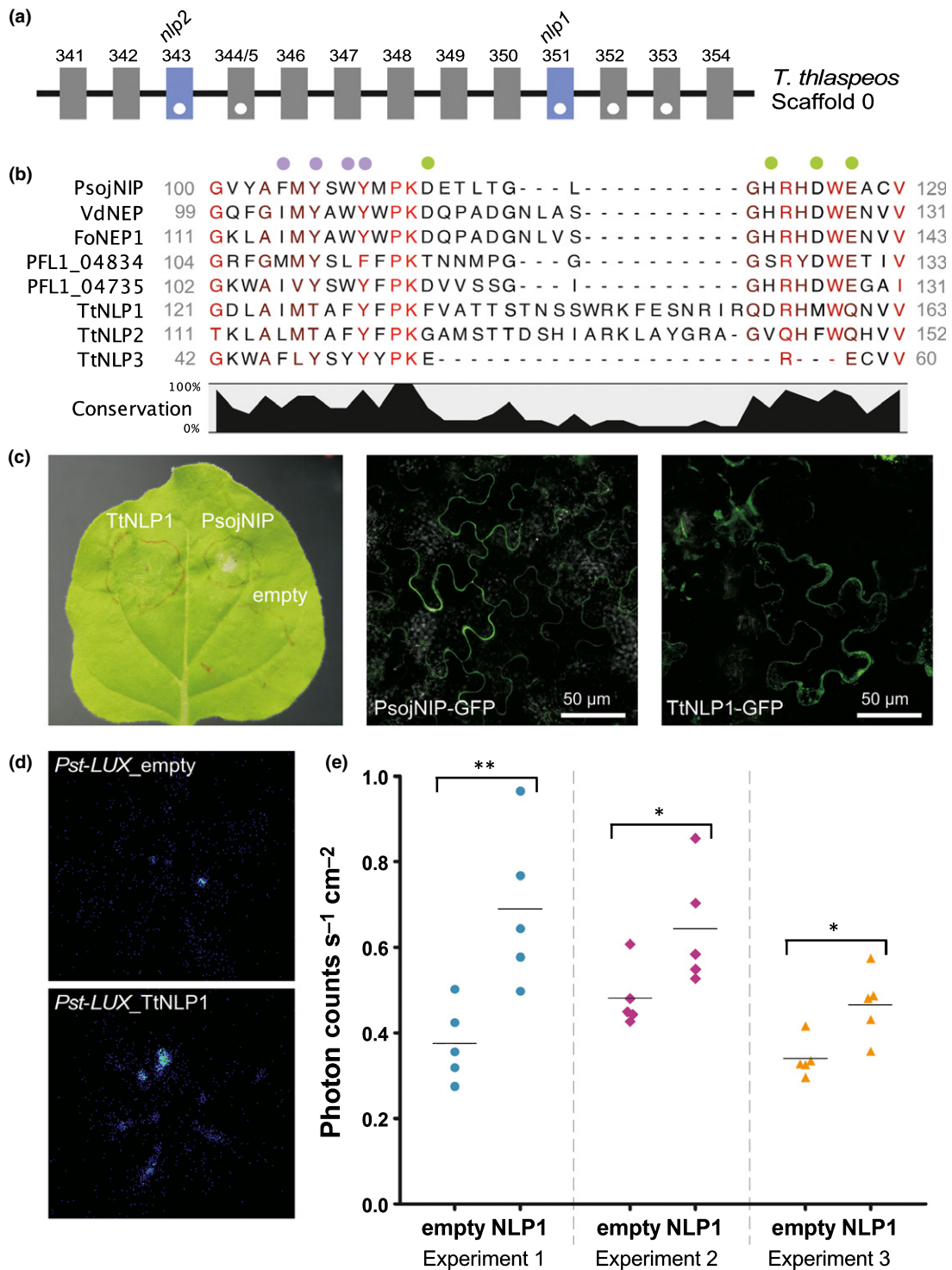
### *T. thlaspeos* infection induces plant defense responses

The lack of macroscopic infection symptoms during the biotrophic growth of *T. thlaspeos* suggests that host gene expression might only be mildly affected. Our expression dataset provides a snapshot of plant responses at the time of infection when biotrophy of the fungus is fully established and sporulation has not started. This status is representative for the major part of the interaction because *T. thlaspeos*

remains in the vegetative tissues for several months and overwinters in perennial host species. *De novo* transcript assembly of healthy and infected *Ar. hirsuta* plants resulted in 170 196 transcripts, out of which 110 864 have a homologous transcript in *A. thaliana* (Table S20).

Analysis of differentially expressed plant genes revealed that infected samples are enriched for functional categories related to biotic stress and defense responses (Fig. 7). This includes receptor-like proteins and kinases, peroxidases, chitinases and NLR domain carrying proteins. Among these are also wall-associated kinases (WAK), which have been shown recently to be involved in the interaction between smut fungi and their respective hosts (Zuo *et al.*, 2014). Salicylic acid (SA)-dependent signaling is a response associated with defense against biotrophic pathogens (Glazebrook, 2005; Huot *et al.*, 2014). Accordingly, the SA-marker gene *PR-1*, as well as *PR-2*, the SA-receptor and transcriptional co-activator *NPRI/NIMI* (Ryals *et al.*, 1997; Wu *et al.*, 2012), the pathogen responsive SA-extrusion exporter *EDS5* (Serrano *et al.*, 2013), the essential regulator of plant systemic acquired resistance *NPRI*, and the integrin *NDRI* (Knepper *et al.*, 2011) are induced during *T. thlaspeos* infection, while the TGA-transcription factors involved in *PR-1* induction (Knepper *et al.*, 2011) do not change (Table 3). Furthermore, *EDS1* and *PAD4* are induced, while *SAG101* is present in several isoforms, but not induced. This suggests that only the *EDS1-PAD4* heteromeric complex (Wagner *et al.*, 2013) might be responsive during infection.

When comparing our data to other systems, we saw similar responses. In maize, 2 d after infection with *U. maydis*, the fungal growth stage resembles most closely *T. thlaspeos* in that the mycelium proliferates and hyphae branch. Tumor development does not start until 4 d post infection (dpi) (Doehlemann *et al.*, 2008). In both smut infections, plant genes associated with biotic stress are induced and photosynthetic genes are repressed (Fig. 7b,c; Doehlemann *et al.*, 2008). However, upon tumor induction at 4 dpi, the plant response to *U. maydis* deviates strongly in contrast to *T. thlaspeos*, where plant morphology is not affected for the entire endophytic period. In comparison, hyphal distribution of the oomycete pathogen *Hyaloperonospora arabidopsidis* (*Hpa*) in *A. thaliana* leaves resembles to some extent the hyphal colonization by *T. thlaspeos*. Transcriptional responses in the compatible interaction with *Hpa* revealed an enrichment of disease resistance and SA-responsive genes including *PR-1* (Asai *et al.*, 2014) similar to *Ar. hirsuta* infected with *T. thlaspeos*, while the *PR-1* induction was not sustained in maize upon *U. maydis* infection (Doehlemann *et al.*, 2008). In addition, a homologue of the receptor-like protein RLP6 (At1g45616) and three of the five receptor-like kinases that were differently expressed in the *Plasmodiophora brassicae-A. thaliana* system (Irani *et al.*, 2018) are strongly induced by *T. thlaspeos* infection. Taken together, we find that *T. thlaspeos* infection induces a typical defense response at the transcriptional level that is comparable with various pathogens and endophytes. At present, due to lack of spatial resolution in



**Fig. 5** *TtNlp1* does not induce a hypersensitive response and increases bacterial luminescence on Col-0. (a) The Nlp locus in Scaffold 1. Genes encoding proteins with a signal peptide are marked with a white dot. (b) Alignment of the region including the ethylene-inducing domain and the heptapeptide sequence of *Thecaphora thlaspeos* Nlps and homologues from other plant pathogens (PsojNIP: *Phytophthora sojae* AAK01636.1, VdNEP: *Verticillium dahliae* AAS45247.1, FoNEP1: *Fusarium oxysporum* AAY88967.2, PFL1\_0434: *Anthracoctis flocculosa* XP\_007880553.1, PFL1\_04735: *A. flocculosa* XP\_007880454.1). Amino acids required for ethylene induction are marked with a purple dot and amino acids required for necrosis are marked with a green dot. (c) *Agrobacterium tumefaciens*-mediated transient expression of *TtNlp1*-Gfp in *Nicotiana benthamiana* along with the positive control *PsojNIP* and the empty vector pEG-103 (negative control). Agro-infiltration of Gfp-tagged *TtNlp1* and *PsojNIP* results in heterologous protein expression as detected by the Gfp signal in *N. benthamiana*. Necrosis is only visible upon infiltration of *PsojNIP* but not with *TtNlp1*. (d) Col-0 sprayed with *Pseudomonas syringae* pv. *tomato* DC3000-LUX (*Pst-LUX*) containing *TtNlp1* display increased luminescence compared with Col-0 sprayed with *Pst-LUX* containing the empty vector control. (e) *TtNlp1*-containing *Pst-LUX* significantly increases luminescence in Col-0 compared with the empty vector control strain. Statistical analysis was carried out using Student's *t*-test: \*\*,  $P < 0.01$ ; \*,  $P < 0.05$ .

**Table 2** Features of *Thecaphora thlaspeos* effector candidates investigated in *Arabidopsis thaliana*.

Candidate effector protein	Protein length (aa)	Signal peptide length (aa)	Upregulation in infection (RT-PCR)	Phenotype when expressed in <i>A. thaliana</i>
THTG_04398	235	27	Yes	–
THTG_04687 ( <i>TtTue1</i> )	294	22	Yes	Small rosettes
THTG_06422	162	18	Yes	–
THTG_04669	269	22	Yes	Few transformants in <i>A. thaliana</i>
THTG_00351 ( <i>TtNlp1</i> )	292	19	Yes	–
THTG_00748	137	25	Yes	–
THTG_06425	271	18	Yes	No effector mRNA accumulation
THTG_03661 ( <i>TtPep1</i> )	142	22	Yes	Few transformants in <i>A. thaliana</i>
THTG_00188	252	23	Yes	No <i>Agrobacterium tumefaciens</i> strains
THTG_00585	449	22	Yes	–

Phenotype refers to macroscopically detectable changes in growth or morphology 4 wk post sowing. Candidate effectors are listed from most upregulated during infection to least and those in gray were not further analyzed in phenotype screen.

our dataset, we cannot yet evaluate whether these responses contribute to limiting *T. thlaspeos* hyphae to the vasculature.

## Discussion

*T. thlaspeos* has a typical smut genome with unique effectors that suggest adaptation to dicot hosts

With a size of *c.* 20 Mb, a low repeat content, and 6239 predicted gene models, the genome of *T. thlaspeos* has the typical characteristics of most sequenced smut fungi. Despite the adaptation to a dicot host, its absolute gene content and predicted functional categories largely overlap with grass-infecting smut fungi (Sharma *et al.*, 2015; Dutheil *et al.*, 2016). However, two unique features stand out from the genome assembly and annotation. First, synteny between *T. thlaspeos* and the grass smuts is low and second, *T. thlaspeos* shares only few known effector candidate genes with its grass smut relatives. Hence, *T. thlaspeos* seems to deploy a different repertoire of effectors to establish and maintain its biotrophic lifestyle. Remarkably, *M. pennsylvanicum*, the only example of grass smuts that underwent a host jump from grasses to the dicot genus *Persicaria*, has maintained its typical grass smut effector repertoire and accordingly has a very low number of *T. thlaspeos* orthologues suggesting independent dicot adaptation in *Thecaphora* and *M. pennsylvanicum* (Sharma *et al.*, 2015; Fig. 2a; Table S13). For example, the Nlps are well known effectors that distinguish *T. thlaspeos* from other smut fungi. Notably, these are also absent in earlier diverging species of the Ustilaginales such as *Ceracesorus bombacis* (Sharma *et al.*, 2015), suggesting independent acquisition for example by horizontal gene transfer.

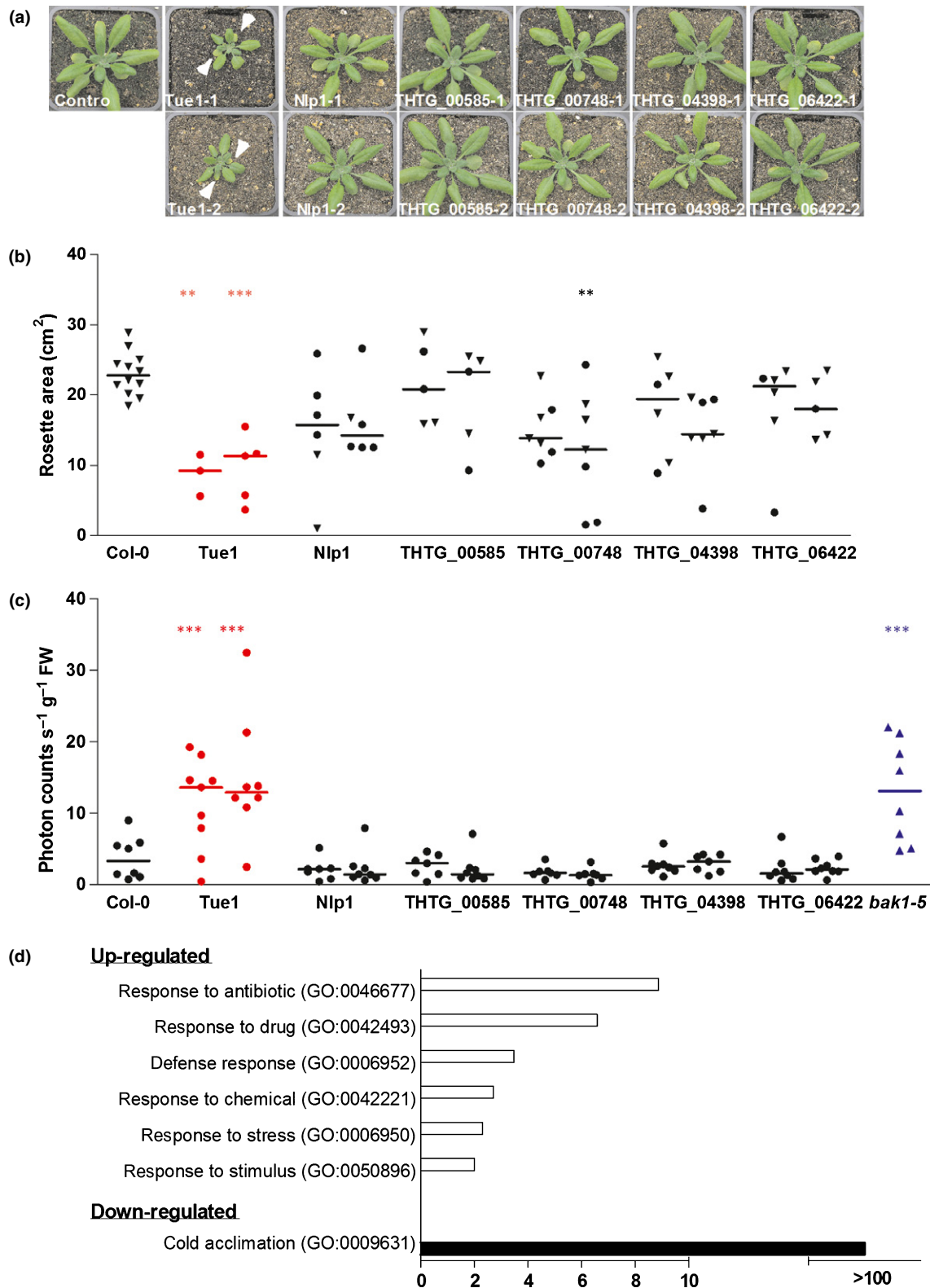
More closely related to *T. thlaspeos* is the epiphytic biocontrol agent *A. flocculosa* which has a significantly higher degree of synteny and a larger overlap in gene content including 17 candidate effector genes. Interestingly, *A. flocculosa* also carries Nlp domain encoding genes (Lefebvre *et al.*, 2013), yet these are nonorthologous to the *T. thlaspeos* Nlps. The close genetic distance to *T. thlaspeos*, along with the presence of these candidate effectors, supports the previously raised hypothesis that *A. flocculosa* besides

being a mycoparasite of powdery mildews (Laur *et al.*, 2017), could also be a yeast anamorph of a dicot-infecting smut species (Begerow *et al.*, 2014).

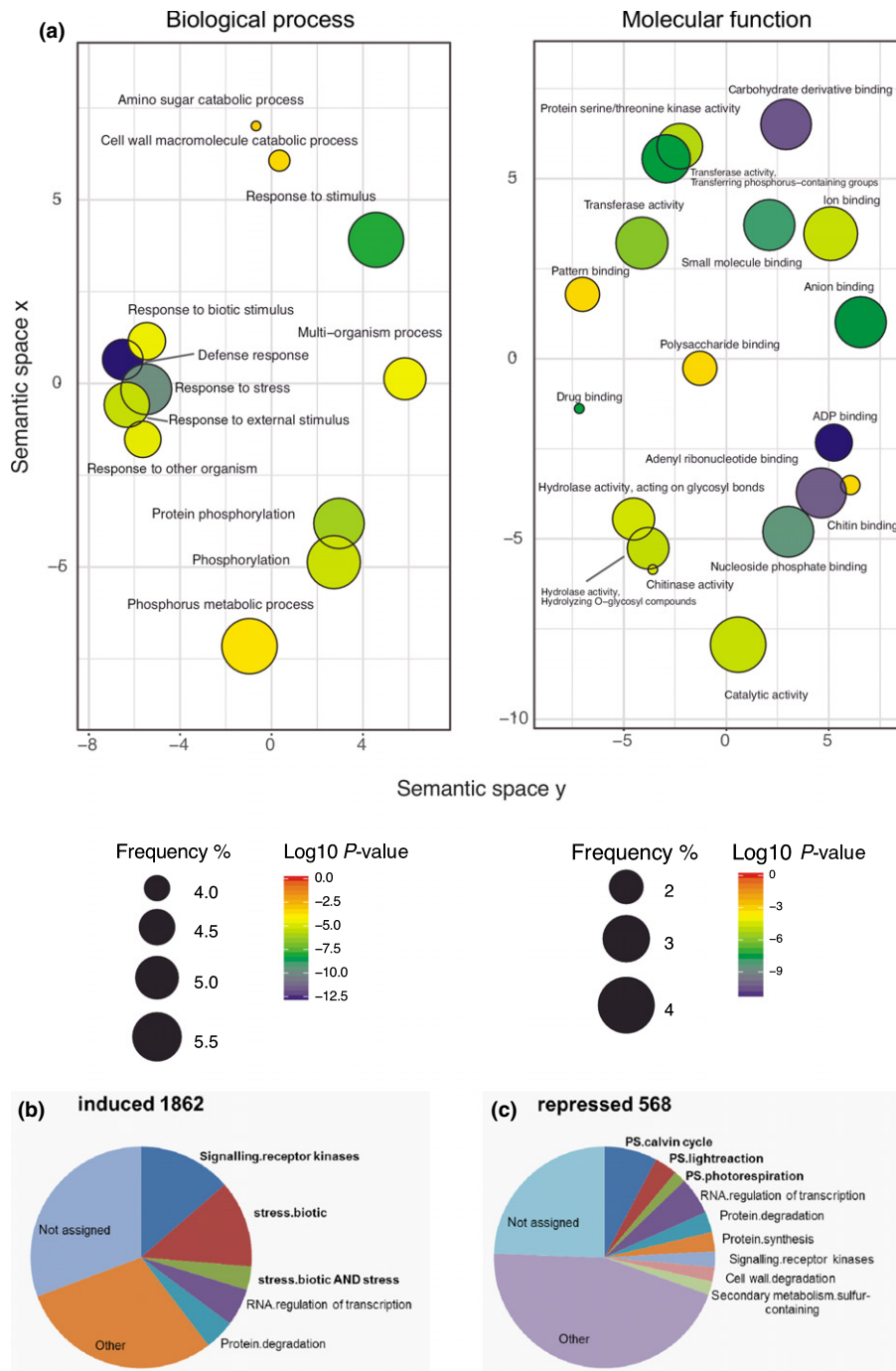
Comparing the two *T. thlaspeos* isolates, LF1 and LF2, revealed the first isolate-specific smut effectors. This is particularly interesting as the infectious form of smut fungi is a dikaryon. Considering that *T. thlaspeos* genetically contains the capacity for mating and that haploid isolates of opposite mating types form fusion hyphae (Frantzeskakis *et al.*, 2017), genetic exchange during mating in *T. thlaspeos* provides the potential to bring together certain virulence-related genes of the single strains. Hence, the combination of different mates could result in distinct fitness levels of the fungus due to alterations in effector dosage and/or content or due to complementation of effector gene losses. In the future, population genetics approaches can reveal distribution of effectors throughout populations and the stability of such populations over the years.

## *T. thlaspeos* infection strategy enables perennial biotrophy

In addition to pathogens, fungal endophytes also possess an astonishing diversity of host colonization strategies that independently evolved in several taxonomic groups (Rodriguez *et al.*, 2009; Brader *et al.*, 2017). Endophytes are microorganisms that colonize the inner plant tissues of macroscopically healthy host plants (Schulz & Boyle, 2005). Some fungal endophytes establish long-lasting interactions such as the generalist root endophyte *Piriformospora indica* or members of the grass endophyte genus *Epiclloë* that remain inside their host throughout the growing season (Rodriguez *et al.*, 2009; Franken, 2012). As for biotrophic pathogens, successful colonization requires complex molecular mechanisms. For example, the basidiomycete *P. indica* establishes biotrophy in *A. thaliana* and barley in a host species-dependent manner with distinct transcriptional responses (Lahrmann *et al.*, 2013). Despite a few well studied examples, it remains largely unknown how plant-pathogenic endophytes manage to establish and maintain such sustained systemic infections and what determines the type of interaction and host specificity.



**Fig. 6** Transgenic *Arabidopsis thaliana* accumulating mRNA of *Tttue1* are small with slightly chlorotic first true leaves. (a) Representative images of two independent plant lines (1 and 2) expressing the upregulated effector candidate. *Tttue1*-lines are clearly smaller and show signs of chlorosis (arrowheads). (b) Quantification of rosette area confirms that *Tttue1* plants are significantly smaller than the controls. Circles indicate plants whose effector mRNA accumulation was confirmed via RT-PCR (25 cycles), triangles indicate additional individuals. (c) *Pst-LUX* proliferation significantly increases in lines expressing *Tue1* and reaches the same level as in the susceptible *bak1-5* mutant. (d) GO term enrichments of differentially expressed genes (DEGs) in the *Tue1*-line reveals upregulation of categories related to response to stress and environment, while cold-stress acclimation seem reduced. (b, c) Individual replicates and median are shown, statistical analysis was carried out by one-way analysis of variance (ANOVA) followed by Bonferroni's post tests; \*\*\*,  $P < 0.001$ ; \*\*,  $P < 0.01$ .



**Fig. 7** Over-represented gene ontology (GO) terms and MapMan categories during the infection of *Arabis hirsuta* by *Thecaphora thlaspeos* show prevalence of biotic stress. (a) Two-dimensional semantic space scatterplots to describe over-represent GO terms were generated with GORILLA and REVIGO ( $P$ -value < 0.001). Circle sizes represent the frequency of the GO term in the Arabidopsis genome while the color indicates the  $P$ -value for the enrichment. (b) MapMan analysis of induced transcript revealed a strong overrepresentation of biotic stress-related categories (bold) in the induced transcripts. About one-third of the transcripts fall into signaling, receptor kinases, or stress/biotic. (c) MapMan analysis of repressed genes indicates a downregulation of photosynthetic genes.

Here, we characterized the biotrophic phase of *T. thlaspeos*. To maintain this ‘hidden’ growth without macroscopic symptoms, it employs typical effector proteins such as Pep1 and Nlps, but also unique effectors such as *TtTue1*. The latter, when overexpressed in *A. thaliana*, causes a growth defect reminiscent of the

phenotypes observed in several *HaNlps* expressed in *A. thaliana* as well as autoimmune mutants (Bowling *et al.*, 1994, 1997; Oome *et al.*, 2014), which transcriptionally activate host immune responses and suppress cold acclimation. Overall, the identification of these novel smut effectors opens the door to study the

**Table 3** Expression of salicylic acid (SA)-dependent defense signaling genes in healthy and infected *Arabis hirsuta* plants.

<i>A. thaliana</i> homologue	Trinity transcript ID	logFC	logCPM	P-value	FDR
PR-1 (At2g14610)*	TRINITY_DN13170_c0_g1_i1	<b>4.51</b>	<b>11.88</b>	<b>1.2E-49</b>	<b>5.1E-45</b>
PR-2 (At3g57260)	TRINITY_DN17095_c0_g1_i3	<b>1.07</b>	<b>6.28</b>	<b>1.0E-06</b>	<b>2.0E-04</b>
EDS1 (At3g48090)	TRINITY_DN27010_c2_g5_i1	<b>0.92</b>	<b>6.07</b>	<b>3.3E-06</b>	<b>5.6E-04</b>
	TRINITY_DN27010_c2_g5_i3	<b>0.71</b>	<b>5.02</b>	<b>3.4E-04</b>	<b>2.8E-02</b>
	TRINITY_DN27010_c2_g5_i6	<b>0.96</b>	<b>3.63</b>	<b>1.2E-05</b>	<b>1.7E-03</b>
	TRINITY_DN27010_c2_g5_i2	0.74	4.33	0.00088	0.06
	TRINITY_DN27010_c2_g7_i4	-0.40	1.14	0.335903	1
EDS5/SID1 (At4g39030)	TRINITY_DN25779_c1_g4_i6	<b>2.23</b>	<b>2.56</b>	<b>5.84E-09</b>	<b>1.9E-06</b>
	TRINITY_DN6765_c0_g1_i1	<b>2.12</b>	<b>1.89</b>	<b>5.8E-08</b>	<b>1.5E-05</b>
	TRINITY_DN25779_c1_g4_i9	-0.63	2.99	0.01	0.40
PAD4 (At3g52430)	TRINITY_DN20844_c0_g1_i1	<b>1.21</b>	<b>5.16</b>	<b>1.4E-07</b>	<b>3.3E-05</b>
	TRINITY_DN20844_c0_g1_i2	<b>0.74</b>	<b>4.58</b>	<b>0.00019</b>	<b>1.7E-02</b>
NDR1 (At3g20600)	TRINITY_DN15989_c0_g1_i1	<b>1.48</b>	<b>4.99</b>	<b>2.8E-09</b>	<b>9.7E-07</b>
	TRINITY_DN15989_c0_g1_i2	<b>1.44</b>	<b>4.33</b>	<b>1.8E-10</b>	<b>7.5E-08</b>
NPR1/NIM1 (At1g64280)	TRINITY_DN27057_c1_g2_i1	<b>1.02</b>	<b>3.73</b>	<b>4.4E-05</b>	<b>5.2E-03</b>
	TRINITY_DN27057_c1_g1_i11	0.54	5.81	5.8E-03	0.23
	TRINITY_DN27057_c1_g1_i18	0.93	1.38	7.4E-03	0.27
	TRINITY_DN27057_c1_g3_i1	0.30	3.82	0.15	1
	TRINITY_DN17408_c0_g3_i1	0.36	0.66	0.45	1

*Arabidopsis thaliana* genes with a role in SA-signaling were selected based on literature (see text). Homologues in the *Ar. hirsuta* transcript assembly were identified. For some genes, more than one transcript was assembled and those with good coverage (logCPM > 0) were considered. Significant FDR values are shown in bold and these transcripts were all induced during infection pointing at upregulation of SA-signalling. PR-5, SAG101, TGA1 and TGA2 were identified in several transcript isoforms, but none on these was induced.

\*For PR-1 several fungal reads from culture mapped to the transcript. These reads likely belong to TtPry1 (THTG\_03812), a putative sterol binding protein and member of the CAP protein superfamily, which is induced during infection.

specific activity of the *Thecaphora* clade effectorome and how, or when, it is utilized to manipulate the host's responses.

On the host side, transcriptional changes reflect a typical response to infection and register as a biotic stress event in the plant's transcriptome. This resembles the plant responses to a majority of microbes, which have effects on their host plant's transcriptomes to various degrees and establish a balance with the plant immune system leading to colonization and infection (Brader *et al.*, 2017). This observation also agrees with previous studies on smut fungi in which an overall upregulation of stress-related gene expression during early infection is kept at bay by effectors (Doehlemann *et al.*, 2008; Djamei & Kahmann, 2012). This effective balance and continuous interaction between *T. thlaspeos* and the host's immune system may therefore limit excessive fungal proliferation to the vasculature which could undermine plant fitness.

In summary, *T. thlaspeos* colonizes Brassicaceae hosts using a unique set of secreted proteins, different from both monocot infecting species but also the dicot-infecting *M. pennsylvanicum*. Excitingly, we find smut-typical effectors such as Pep1, dicot-typical effector genes such as the Nlps but also novel effector candidates such as *TfTue1* that seem to integrate abiotic stress factors, that is cold-stress response into the fungal infection mechanism. In addition, we show that the effector repertoire likely differs between *T. thlaspeos* isolates. Further studies on *T. thlaspeos* will elucidate whether the secreted protein-coding genes identified here present different expression patterns in various tissues or at different points during its long-term biotrophic stage. Finally, using the information and the resources provided

here, more extensive studies could address the pathogen-endophyte continuum using *T. thlaspeos* as a model organism.











## Acknowledgements

We warmly thank our colleagues at the Institute of Microbiology, The Sainsbury Laboratory (TSL), the John Innes Centre (JIC), and Michigan State University (MSU) for great discussions. We greatly thank the TSL plant transformation team for their work in generating our transgenic *A. thaliana* lines, the JIC horticultural team for excellent plant care services, and the TSL support team for media preparation. Special thanks to Simone Esch, Lesley Plücker and Merve Öztürk for technical assistance. Research in the laboratory of VG is funded by the Cluster of Excellence in Plant Sciences (CEPLAS, DFG EXC 1028) and the Bioeconomy Science Center (BioSC). The scientific activities of the BioSC were financially supported by the Ministry of Innovation, Science and Research within the framework of the NRW Strategieprojekt BioSC (No. 313/323-400-00213). RK and EK are supported by CEPLAS, LF and KJC were supported by a doctoral fellowship of the DFG International Research Training Group 1525 iGRADplant. Research performed at TSL was kindly supported by the 2Blades Foundation (2B3511-KK). Research in the laboratory of Brad Day is supported by a grant from the US National Science Foundation (IOS-1146128). Microscopy was carried out at the 'Center for Advanced Imaging', HHU. Computational support and infrastructure was provided by the 'Centre for Information and Media Technology', HHU and by the Institute for Cyber-Enabled Research, MSU.

## Author contributions

VG and MF planned and designed the research. LF and KJC designed and performed the experiments. SG, N Haeger and N Heßler provided infection figures. AB contributed genomic sequences. LF and BU performed the bioinformatic analysis. EK, YKG, HPvE, and BD contributed materials and supported the experimental design. KJC, LF, RK, and VG wrote the manuscript. KJC and LF contributed equally to this work.

## ORCID

Andreas Brachmann  <https://orcid.org/0000-0001-7980-8173>  
 Brad Day  <https://orcid.org/0000-0002-9880-4319>  
 Michael Feldbrügge  <https://orcid.org/0000-0003-0046-983X>  
 Lamprinos Frantzeskakis  <https://orcid.org/0000-0001-8947-6934>  
 Vera Göhre  <https://orcid.org/0000-0002-7118-1609>  
 Yogesh K. Gupta  <https://orcid.org/0000-0002-8696-9044>  
 Ronny Kellner  <https://orcid.org/0000-0002-4618-0110>  
 Eric Kemen  <https://orcid.org/0000-0002-7924-116X>  
 H. Peter van Esse  <https://orcid.org/0000-0002-3667-060X>  
 Björn Usadel  <https://orcid.org/0000-0003-0921-8041>

## References

- Aichinger C, Hansson K, Eichhorn H, Lessing F, Mannhaupt G, Mewes W, Kahmann R. 2003. Identification of plant-regulated genes in *Ustilago maydis* by enhancer-trapping mutagenesis. *Molecular Genetics and Genomics* 270: 303–314.
- Andrade O, Munoz G, Galdames R, Duran P, Honorato R. 2004. Characterization, *in vitro* culture, and molecular analysis of *Thecaphora solani*, the causal agent of potato smut. *Phytopathology* 94: 875–882.
- Asai S, Rallapalli G, Piquerez SJM, Caillaud MC, Furzer OJ, Ishaque N, Wirthmueller L, Fabro G, Shirasu K, Jones JDG. 2014. Expression profiling during Arabidopsis/downy mildew interaction reveals a highly-expressed effector that attenuates responses to salicylic acid. *PLoS Pathogens* 10: e1004443.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.
- Begerow D, Schaefer AM, Kellner R, Yurkov A, Kemler M, Oberwinkler F, Bauer R. 2014. 11 Ustilagomycotina. In: McLaughlin D, Spatafora J, eds. *Systematics and evolution. The mycota: systematics and evolution part A VII*. Berlin/Heidelberg, Germany: Springer.
- Białas A, Zess EK, De la Concepcion JC, Franceschetti M, Pennington HG, Yoshida K, Upson JL, Chanclud E, Wu C-H, Langner T *et al.* 2017. Lessons in effector and NLR biology of plant-microbe systems. *Molecular Plant–Microbe Interactions* 31: 34–45.
- Bleckmann A, Weidtkamp-Peters S, Seidel CAM, Simon R. 2010. Stem cell signaling in Arabidopsis requires CRN to localize CLV2 to the plasma membrane. *Plant Physiology* 152: 166–176.
- Böhm H, Albert I, Oome S, Raaymakers TM, Van den Ackerveken G, Nürnberger T. 2014. A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in Arabidopsis. *PLoS Pathogens* 10: e1004491.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bösch K, Frantzeskakis L, Vraneš M, Kämper J, Schipper K, Göhre V. 2016. Genetic manipulation of the plant pathogen *Ustilago maydis* to study fungal biology and plant-microbe interactions. *Journal of Visualized Experiments* 115: e54522.
- Bowling SA, Clarke JD, Liu Y, Klessig DF, Dong X. 1997. The *cpr5* mutant of Arabidopsis expresses both NPR1-dependent and NPR1-independent resistance. *Plant Cell* 9: 1573–1584.
- Bowling SA, Guo A, Cao H, Gordon AS, Klessig DF, Dong X. 1994. A mutation in Arabidopsis that leads to constitutive expression of systemic acquired resistance. *Plant Cell* 6: 1845–1857.
- Brader G, Compant S, Vescio K, Mitter B, Trognitz F, Ma L-J, Sessitsch A. 2017. Ecology and genomic insights into plant-pathogenic and plant-nonpathogenic endophytes. *Annual Review of Phytopathology* 55: 61–83.
- Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R. 2009. *Ustilago maydis* as a pathogen. *Annual Review of Phytopathology* 47: 423–445.
- Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics* 48: 4.11.11–14.11.39.
- Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nürnberger T, Jones JDG, Felix G, Boller T. 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* 448: 497–500.
- Choi J, Kim K-T, Jeon J, Lee Y-H. 2013. Fungal plant cell wall-degrading enzyme database: a platform for comparative and evolutionary genomics in fungi and Oomycetes. *BMC Genomics* 14: S7.
- Conforto C, Cazón I, Fernández FD, Marinelli A, Oddino C, Rago AM. 2013. Molecular sequence data of *Thecaphora frezii* affecting peanut crops in Argentina. *European Journal of Plant Pathology* 137: 663–666.
- Delcher AL, Salzberg SL, Phillippy AM. 2003. Using MUMmer to identify similar regions in large sequence sets. In: Baxevanis AD, Davison DB, eds. *Current protocols in bioinformatics, Chapter 10*. Hoboken, NJ, USA: John Wiley & Sons, 10.3.1–10.3.18.
- Djamei A, Kahmann R. 2012. *Ustilago maydis*: dissecting the molecular interface between pathogen and plant. *PLoS Pathogens* 8: e1002955.
- Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I *et al.* 2011. Metabolic priming by a secreted fungal effector. *Nature* 478: 395–398.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
- Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics* 11: 539–548.
- Doehlemann G, Reissmann S, Aßmann D, Fleckenstein M, Kahmann R. 2011. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Molecular Microbiology* 81: 751–766.
- Doehlemann G, van der Linde K, Assmann D, Schwambach D, Hof A, Mohanty A, Jackson D, Kahmann R. 2009. Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathogens* 5: e1000290.
- Doehlemann G, Wahl R, Horst RJ, Voll LM, Usadel B, Poree F, Stitt M, Pons-Kühnemann J, Sonnewald U, Kahmann R *et al.* 2008. Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. *The Plant Journal* 56: 181–195.
- Dutheil JY, Mannhaupt G, Schweizer G, M K Sieber C, Münsterkötter M, Güldener U, Schirawski J, Kahmann R. 2016. A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. *Genome Biology and Evolution* 8: 681–704.
- Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant Journal* 45: 616–629.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 1–7.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.



- Engler C, Youles M, Gruetner R, Ehnert TM, Werner S, Jones JDG, Patron NJ, Marillonnet S. 2014. A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology* 3: 839–843.
- Fabro G, Steinbrenner J, Coates M, Ishaque N, Baxter L, Studholme DJ, Körner E, Allen RL, Piquerez SJM, Rougon-Cardoso A *et al.* 2011. Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. *PLoS pathogens* 7: e1002348.
- Fedler M, Luh KS, Stelter K, Nieto-Jacobo F, Basse CW. 2009. The  $\alpha 2$  mating-type locos genes *lga2* and *rga2* direct uniparental mitochondrial DNA (mtDNA) inheritance and constrain mtDNA recombination during sexual development of *Ustilago maydis*. *Genetics* 181: 847–860.
- Feldbrügge M, Kämper J, Steinberg G, Kahmann R. 2004. Regulation of mating and pathogenic development in *Ustilago maydis*. *Current Opinion in Microbiology* 7: 666–672.
- Franken P. 2012. The plant strengthening root endophyte *Piriformospora indica*: potential application and the biology behind. *Applied Microbiology and Biotechnology* 96: 1455–1464.
- Frantzeskakis L, Courville KJ, Pluecker L, Kellner R, Kruse J, Brachmann A, Feldbrügge M, Göhre V. 2017. The plant-dependent life cycle of *Thecaphora thlaspeos*: a smut fungus adapted to Brassicaceae. *Molecular Plant–Microbe Interactions* 30: 271–282.
- Gassmann W, Bhattacharjee S. 2012. Effector-triggered immunity signaling: from gene-for-gene pathways to protein–protein interaction networks. *Molecular Plant–Microbe Interactions* 25: 862–868.
- Germain H, Joly DL, Mireault C, Plourde MB, Letanneur C, Stewart D, Morency MJ, Petre B, Duplessis S, Seguin A. 2017. Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology* 19: 191–200.
- Ghareeb H, Drechsler F, Löffke C, Teichmann T, Schirawski J. 2015. SUPPRESSOR OF APICAL DOMINANCE 1 of *Sporisorium reilianum* Modulates Inflorescence Branching Architecture in Maize and *Arabidopsis*. *Plant Physiology* 169: 2789–2804.
- Giraldo MC, Valent B. 2013. Filamentous plant pathogen effectors in action. *Nature Reviews Microbiology* 11: 800–814.
- Glazebrook J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology* 43: 205–227.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Gutjahr C, Parniske M. 2013. Cell and developmental biology of arbuscular mycorrhizal symbiosis. *Annual Review of Cell and Developmental Biology* 29: 593–617.
- Hemetsberger C, Herrberger C, Zechmann B, Hillmer M, Doehlemann G. 2012. The *Ustilago maydis* effector *Pep1* suppresses plant immunity by inhibition of host peroxidase activity. *PLoS Pathogens* 8: e1002684.
- Hemetsberger C, Mueller N, Matei A, Herrberger C, Doehlemann G. 2015. The fungal core effector *Pep1* is conserved across smuts of dicots and monocots. *New Phytologist* 206: 1116–1126.
- Holliday R. 1961. Induced mitotic crossing-over in *Ustilago maydis*. *Genetics Research* 2: 231–248.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12: 491.
- Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, Yi H, Yang Z, Yin Y. 2017. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Research* 46: D516–D521.
- Huot B, Yao J, Montgomery BL, He SY. 2014. Growth-defense tradeoffs in plants: a balancing act to optimize fitness. *Molecular Plant* 7: 1267–1287.
- Irani S, Trost B, Waldner M, Nayidu N, Tu J, Kusalik AJ, Todd CD, Wei Y, Bonham-Smith PC. 2018. Transcriptome analysis of response to *Plasmodiophora brassicae* infection in the *Arabidopsis* shoot and root. *BMC Genomics* 19: 23.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240.
- Jurka J, Kapitonov V, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110: 462–467.
- Kämper J, Kahmann R, Bölker M, Ma L-J, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Müller O *et al.* 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444: 97–101.
- Karimi M, Inze D, Depicker A. 2002. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* 7: 193–195.
- Katagiri F, Thilmony R, He SY. 2002. The *Arabidopsis thaliana*–*Pseudomonas syringae* interaction. *Arabidopsis Book* 1: e0039.
- Kellner R, Vollmeister E, Feldbrügge M, Begerow D. 2011. Interspecific sex in grass smuts and the genetic diversity of their pheromone-receptor system. *PLoS Genetics* 7: e1002436.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12: 357–360.
- Knepper C, Savory EA, Day B. 2011. The role of NDR1 in pathogen perception and plant defense signaling. *Plant Signaling and Behavior* 6: 1114–1116.
- Koncz C, Schell J. 1986. The promoter of T<sub>1</sub>-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Molecular and General Genetics* 204: 383–396.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27: 722–736.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5: 1–9.
- Lahrmann U, Ding Y, Banhara A, Rath M, Hajirezaei MR, Döhlemann S, von Wirén N, Parniske M, Zuccaro A. 2013. Host-related metabolic cues affect colonization strategies of a root endophyte. *Proceedings of the National Academy of Sciences, USA* 110: 13965–13970.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.
- Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitz M, Pellegrin C, Reissmann S, Altmüller J, Rensing SA *et al.* 2018. The biotrophic development of *Ustilago maydis* studied by RNAseq analysis. *Plant Cell* 30: 300–323.
- Lanver D, Tollot M, Schweizer G, Lo Presti L, Reissmann S, Ma LS, Schuster M, Tanaka S, Liang L, Ludwig N *et al.* 2017. *Ustilago maydis* effectors and their impact on virulence. *Nature Reviews Microbiology* 15: 409–421.
- Laur J, Ramakrishnan GB, Labbé C, Lefebvre F, Spanu PD, Bélanger RR. 2017. Effectors involved in fungal-fungal interaction lead to a rare phenomenon of hyperbiotrophy in the tritrophic system biocontrol agent-powdery mildew-plant. *New Phytologist* 217: 713–725.
- Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a web-based genomic annotation editing platform. *Genome Biology* 14: R93.
- Lefebvre F, Joly DL, Labbé C, Teichmann B, Linning R, Belzile F, Bakkeren G, Bélanger RR. 2013. The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis. *Plant Cell* 25: 1946–1959.
- Linning R, Lin D, Lee N, Abdennadher M, Gaudet D, Thomas P, Mills D, Kronstad JW, Bakkeren G. 2004. Marker-based cloning of the region containing the *UvAvr1* avirulence gene from the basidiomycete barley pathogen *Ustilago hordei*. *Genetics* 166: 99–111.
- Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal effectors and plant susceptibility. *Annual Review of Plant Biology* 66: 513–545.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 1–21.
- Okmen B, Doehlemann G. 2014. Inside plant: biotrophic strategies to modulate host immunity and metabolism. *Current Opinion in Plant Biology* 20: 19–25.
- Oome S, Raaymakers TM, Cabral A, Samwel S, Böhm H, Albert I, Nürnberger T, Van den Ackerveken G. 2014. Nep1-like proteins from three kingdoms of

- life act as a microbe-associated molecular pattern in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* 111: 16955–16960.
- Oome S, Van den Ackerveken G. 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. *Molecular Plant–Microbe Interactions* 27: 1–51.
- Ottmann C, Luberacki B, Kufner J, Koch W, Brunner F, Weyand M, Mattinen L, Pirhonen M, Anderlüh G, Seitz HU *et al.* 2009. A common toxin fold mediates microbial attack and plant defense. *Proceedings of the National Academy of Sciences, USA* 106: 10359–10364.
- Patron N, Orzaez D, Marillonnet S, Warzecha H, Matthewman C, Youles M, Raitskin O, Leveau A, Farre G, Rogers C *et al.* 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist* 208: 13–19.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11: 1650–1667.
- Qutob D, Kamoun S, Gijzen M. 2002. Expression of a *Phytophthora sojae* necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *The Plant Journal* 32: 361–373.
- Rabe F, Bosch J, Stirnberg A, Guse T, Bauer L, Seitner D, Rabanal FA, Czedik-Eysenberg A, Uhse S, Bindics J *et al.* 2016. A complete toolset for the study of *Ustilago bromivora* and *Brachypodium* sp. as a fungal-temperate grass pathosystem. *eLife* 5: 1–35.
- Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* 10: 417–430.
- Redkar A, Hoser R, Schilling L, Zechmann B, Krzymowska M, Walbot V, Doehlemann G. 2015. A secreted effector protein of *Ustilago maydis* guides maize leaf cells to form tumors. *Plant Cell* 27: 1332–1351.
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology* 12: R22.
- Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
- Rodriguez RJ, White JF, Arnold AE, Redman RS. 2009. Fungal endophytes: diversity and functional roles. *New Phytologist* 182: 314–330.
- Ryals J, Weymann K, Lawton K, Friedrich L, Ellis D, Steiner HY, Johnson J, Delaney TP, Jesse T, Vos P *et al.* 1997. The Arabidopsis NIM1 protein shows homology to the mammalian transcription factor inhibitor I kappa B. *Plant Cell* 9: 425–439.
- Santhanam P, van Esse HP, Albert I, Faino L, Nürnberger T, Thomma BPHJ. 2012. Evidence for functional diversification within a fungal NEP1-like protein family. *Molecular Plant–Microbe Interactions* 26: 278–286.
- Schipper K. 2009. *Charakterisierung eines Ustilago maydis Genclusters, das für drei neuartige sekretierte Effektoren kodiert*. PhD thesis, Philipps-Universität Marburg, Marburg, Germany. 1–160.
- Schirawski J, Mannhaupt G, Münch K, Brefort T, Schipper K, Doehlemann G, Di Stasio M, Rössel N, Mendoza-Mendoza A, Pester D *et al.* 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 330: 1546–1548.
- Schouten A, van Baarlen P, van Kan JAL. 2007. Phytotoxic Nep1-like proteins from the necrotrophic fungus *Botrytis cinerea* associate with membranes and the nucleus of plant cells. *New Phytologist* 177: 493–505.
- Schulz B, Boyle C. 2005. The endophytic continuum. *Mycological Research* 109: 661–686.
- Seitner D, Uhse S, Gallei M, Djamei A. 2018. The core effector Cce1 is required for early infection of maize by *Ustilago maydis*. *Molecular Plant Pathology* 19: 2277–2287.
- Serrano M, Wang B, Aryal B, Garcion C, Abou-Mansour E, Heck S, Geisler M, Mauch F, Nawrath C, Mettraux J-P. 2013. Export of salicylic acid from the chloroplast requires the multidrug and toxin extrusion-like transporter EDS5. *Plant Physiology* 162: 1815–1821.
- Sharma R, Xia X, Riess K, Bauer R, Thines M. 2015. Comparative genomics including the early-diverging smut fungus *Ceratosorus bombacis* reveals signatures of parallel evolution within plant and animal pathogens of fungi and oomycetes. *Genome Biology and Evolution* 7: 2781–2798.
- Shimada TL, Shimada T, Hara-Nishimura I. 2010. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *The Plant Journal* 61: 519–528.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Smith-Unna R, Bournnell C, Patro R, Hibberd JM, Kelly S. 2016. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Research* 26: 1134–1144.
- Sohn KH, Lei R, Nemri A, Jones JDG. 2007. The Downy Mildew effector proteins ATR1 and ATR13 promote disease susceptibility in *Arabidopsis thaliana*. *Plant Cell* 19: 4077–4090.
- Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4: 1–8.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33: 465–467.
- Stockenhuber R, Zoller S, Shimizu-inatsugi R, Gugerli F. 2015. Efficient detection of novel nuclear markers for Brassicaceae by transcriptome sequencing. *PLoS ONE* 10: e0128181.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6: e21800.
- Teichmann B, Labbé C, Lefebvre F, Bölker M, Linne U, Bélanger RR. 2011. Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa*. *Molecular Microbiology* 79: 1483–1495.
- Teichmann B, Linne U, Hewald S, Marahiel MA, Bölker M. 2007. A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Molecular Microbiology* 66: 525–533.
- Tempel S. 2012. Using and understanding RepeatMasker. *Methods in Molecular Biology* 859: 29–51.
- Thalhammer A, Bryant G, Sulpice R, Hinch DK. 2014. Disordered cold regulated 15 proteins protect chloroplast membranes during freezing through binding and folding, but do not stabilize chloroplast enzymes *in vivo*. *Plant Physiology* 166: 190–201.
- Thomma BPHJ, Eggermont K, Penninckx IAMA, Mauch-Mani B, Vogelsang R, Cammue BPA, Broekaert WF. 1998. Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *Proceedings of the National Academy of Sciences, USA* 95: 15107–15111.
- Tsuda K, Somssich IE. 2015. Transcriptional networks in plant immunity. *New Phytologist* 206: 932–947.
- Urban M, Kahmann R, Bölker M. 1996. Identification of the pheromone response element in *Ustilago maydis*. *Molecular and General Genetics* 251: 31–37.
- Vanky K, Lutz M, Bauer R. 2008. About the genus *Thecaphora* (Glomosporiaceae) and its new synonyms. *Mycological Progress* 7: 31–39.
- Vie AK, Najafi J, Winge P, Cattani E, Wrzaczek M, Kangasjärvi J, Miller G, Brembu T, Bones AM. 2017. The IDA-LIKE peptides IDL6 and IDL7 are negative modulators of stress responses in *Arabidopsis thaliana*. *Journal of Experimental Botany* 68: 3557–3571.
- Wagner S, Stuttmann J, Rietz S, Guerois R, Brunstein E, Bautor J, Niefind K, Parker JE. 2013. Structural basis for signaling by exclusive EDS1 heteromeric complexes with SAG101 or PAD4 in plant innate immunity. *Cell Host & Microbe* 14: 619–630.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9: e112963.
- Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W *et al.* 2015. antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* 43: 1–7.
- Wu Y, Zhang D, Chu JY, Boyle P, Wang Y, Brindle ID, De Luca V, Després C. 2012. The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Reports* 1: 639–647.

- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40: W445–W451.
- Zhang Z, Feechan A, Pedersen C, Newman MA, Qiu JL, Olesen KL, Thordal-Christensen H. 2007. A SNARE-protein has opposing functions in penetration resistance and defence signalling pathways. *The Plant Journal* 49: 302–312.
- Zipfel C. 2014. Plant pattern-recognition receptors. *Trends in Immunology* 35: 345–351.
- Zuo W, Chao Q, Zhang N, Ye J, Tan G, Li B, Xing Y, Zhang B, Liu H, Fengler KA *et al.* 2014. A maize wall-associated kinase confers quantitative resistance to head smut. *Nature Genetics* 47: 151–157.

## Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article:

**Fig. S1** Map of the mitochondrial genome of *T. thlaspeos* strain LF1.

**Fig. S2** The mating type locus *a* of *T. thlaspeos*.

**Fig. S3** Genome to genome alignments between *T. thlaspeos* and other Ustilaginomycotina species.

**Fig. S4** *T. thlaspeos* unique effector THTG\_04398 is absent in strain LF2 and in isolates from Hohe Leite, Germany, while present in LF1.

**Fig. S5** Principal component analysis (PCA) for the fungal RNA samples collected during axenic culture and infection of *Ar. hirsuta*.

**Fig. S6** Transgenic *A. thaliana* lines used in this study accumulate mRNA of each respective effector.

**Table S1** Datasets used in this study.

**Table S2** Scaffolds with telomeric repeats.

**Table S3** BUSCO analysis of the genomic assembly and annotation for the genomes of LF1 and LF2.

**Table S4** Functional annotation of the LF1 predicted proteome.

**Table S5** List of secreted proteins of *T. thlaspeos* LF1.

**Table S6** Count of predicted secreted proteins of other smut fungi species.

**Table S7** Number of carbohydrate-active enzyme coding genes in several smut fungi genomes including *T. thlaspeos*.

**Table S8** Genes related to secondary metabolism in *T. thlaspeos*, *U. maydis* and *A. flocculosa* as predicted by ANTIMASH 4.0.2.

**Table S9** Repetitive content of the *T. thlaspeos* LF1 genome.

**Table S10** *U. maydis* genes involved in mating and their putative orthologues in *T. thlaspeos*.

**Table S11** Whole-genome alignments between *T. thlaspeos* LF1 and other Ustilaginomycotina species.

**Table S12** Orthologue genes to *U. maydis* known effector clusters and effectors.

**Table S13** Orthology analysis between the *T. thlaspeos* LF1 and 20 Ustilaginomycotina proteomes.

**Table S14** *T. thlaspeos* LF1 genes with no homologues to other smut fungi.

**Table S15** Summary of differences between the genomes of *T. thlaspeos* LF1 and LF2.

**Table S16** *T. thlaspeos* gene models with no hits in the genomic assembly of the opposite isolate.

**Table S17** DESEQ2 analysis of the fungal transcriptome during infection of *Ar. hirsuta* with *T. thlaspeos* spores.

**Table S18** Fungal genes upregulated during infection of *Ar. hirsuta* with *T. thlaspeos* spores.

**Table S19** DESEQ2 analysis of the plant transcriptome of Tue1-expressing *A. thaliana*.

**Table S20** Transcriptome analysis of *Ar. hirsuta* plants infected with *T. thlaspeos* spores.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.