Integrated genomic and transcriptomic analysis reveals mycoparasitism as the ancestoral life style of *Trichoderma*

Christian P. Kubicek[1], Alfredo Herrera-Estrella[2], Verena Seidl[1], Stéphane Le Crom[3], Diego A. Martinez[4], Irina S. Druzhinina[1], Susanne Zeilinger[1], Sergio Casas-Flores[5], Benjamin A. Horwitz[6], Prasun K. Mukherjee[7], Mala Mukherjee[6], László Kredics[8], Luis David Alcaraz[2], Andrea Aerts[9], Zsuzsanna Antal[8], Lea Atanasova[1], Mayte Guadalupe Cervantes-Badillo[5], Jean Challacombe[9], Olga Chertkov[9], Kevin McCluskey[10], Fanny Coulpier[3], Nandan Deshpande[11], Hans von Döhren[12], Daniel J. Ebbole[13], Edgardo Ulises Esquivel-Naranjo[2], Erzsébet Fekete[14], Michel Flipphi[15], Fabian Glaser[6], Elida Yazmín Gómez-Rodríguez[5], Sabine Gruber[1], Cliff Han[9], Bernard Henrissat[16], Rosa Hermosa[17], Miguel Hernández-Oñate[2], Levente Karaffa[14], Idit Kosti[6], Erika Lindquist[9], Susan Lucas[9], Mette Lübeck[18], Peter Stephensen Lübeck[18], Antoine Margeot[19], Benjamin Metz[1], Monica Misra[9], Helena Nevalainen[11], Markus Omann[1], Nicolle Packer[11], Giancarlo Perrone[20], Edith Elena Uresti-Rivera[5], Asaf Salamov[9], Monika Schmoll[1], Bernhard Seiboth[1], Harris Shapiro[9], Serenella Sukno[17], Juan Antonio Tamayo-Ramos[15], Michael Thon[17], Doris Tisch[1], Aric Wiest[10], Heather H. Wilkinson[13], Michael Zhang[9], Pedro M. Coutinho[16], Charles M. Kenerley[13], Enrique Monte[17], Scott E. Baker[9, 21], Igor V. Grigoriev[9]

(1) Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering Vienna University of Technology, 1060 Vienna, Austria; (2) Laboratorio Nacional de Genómica para la Biodiversidad, Cinvestav Campus Guanajuato, 36821, Irapuato, Gto, Mexico; (3) Institut de Biologie de l'École normale supérieure (IBENS), Institut National de la Santé et de la Recherche Médicale U1024,

Centre National de la Recherche Scientifique UMR8197, 75005 Paris, France; (4) Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA; (5) División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), 78216 San Luis Potosí, Mexico; (6) Department of Biology, Technion, Haifa 32000, Israel; (7) Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Mumbai 400 085, India; (8) Department of Microbiology, Faculty of Science and Informatics, University of Szeged, Hungary; (9) DOE Joint Genome Institute, Walnut Creek, California 94598, USA; (10) Fungal Genetics Stock Center, School of Biological Sciences, University of Missouri- Kansas City, USA; (11) Chemistry and Biomolecular Sciences, Macqaurie University, Sydney, Australia; (12) TU Berlin, Institut für Chemie, FG Biochemie und Molekulare Biologie, 10587 Berlin, Germany; (13) Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas 77843, USA; (14) Department of Genetics and Applied Microbiology, Faculty of Science and Technology, University of Debrecen, H-4010, Debrecen, Hungary; (15) Instituto de Agroquimica y Tecnologia de Alimentos (IATA), Consejo Superior de Investigaciones Cientificas (CSIC), 46100 Burjassot, Valencia, Spain; (16) Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Universités Aix-Marseille I & II, 13288 Marseille, France; (17) Spanish-Portuguese Centre for Agricultural Research (CIALE), Department of Microbiology and Genetics, University of Salamanca, Salamanca, Spain; (18) Section for Sustainable Biotechnology, Department of Biotechnology, Chemistry and Environmental Engineering, Copenhagen Institute of Technology, Aalborg University, Denmark; (19) IFP, Biotechnology Department, Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France; (20) Institute of Sciences of Food Production-National Research Council (ISPA-CNR), 70126 Bari, Italy; (21) Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, Richland, WA 99352, USA

**To whom correspondence may be addressed. E-mail:* ckubicek@mail.zserv.tuwien.ac.at

29 April 2011

ACKNOWLEDGMENTS:

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

DISCLAIMER:

Integrated genomic and transcriptomic analysis reveals mycoparasitism as the ancestoral life style of *Trichoderma*

Christian P. Kubicek[1], Alfredo Herrera-Estrella[2], Verena Seidl[1], Stéphane Le Crom[3], Diego A. Martinez[4], Irina S. Druzhinina[1], Susanne Zeilinger[1], Sergio Casas-Flores[5], Benjamin A. Horwitz[6], Prasun K. Mukherjee[7], Mala Mukherjee[6], László Kredics[8], Luis David Alcaraz[2], Andrea Aerts[9], Zsuzsanna Antal[8], Lea Atanasova[1], Mayte Guadalupe Cervantes-Badillo[5], Jean Challacombe[9], Olga Chertkov[9], Kevin McCluskey[10], Fanny Coulpier[3], Nandan Deshpande[11], Hans von Döhren[12], Daniel J. Ebbole[13], Edgardo Ulises Esquivel-Naranjo[2], Erzsébet Fekete[14], Michel Flipphi[15], Fabian Glaser[6], Elida Yazmín Gómez-Rodríguez[5], Sabine Gruber[1], Cliff Han[9], Bernard Henrissat[16], Rosa Hermosa[17], Miguel Hernández-Oñate[2], Levente Karaffa[14], Idit Kosti[6], Erika Lindquist[9], Susan Lucas[9], Mette Lübeck[18], Peter Stephensen Lübeck[18], Antoine Margeot[19], Benjamin Metz[1], Monica Misra[9], Helena Nevalainen[11], Markus Omann[1], Nicolle Packer[11], Giancarlo Perrone[20], Edith Elena Uresti-Rivera[5], Asaf Salamov[9], Monika Schmoll[1], Bernhard Seiboth[1], Harris Shapiro[9], Serenella Sukno[17], Juan Antonio Tamayo-Ramos[15], Michael Thon[17], Doris Tisch[1], Aric Wiest[10], Heather H. Wilkinson[13], Michael Zhang[9], Pedro M. Coutinho[16], Charles M. Kenerley[13], Enrique Monte[17], Scott E. Baker[9, 21], Igor V. Grigoriev[9]

(1) Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering Vienna University of Technology, 1060 Vienna, Austria; (2) Laboratorio Nacional de Genómica para la Biodiversidad, Cinvestav Campus Guanajuato, 36821, Irapuato, Gto, Mexico; (3) Institut de Biologie de l'École normale supérieure (IBENS), Institut National de la Santé et de la Recherche Médicale U1024, Centre National de la Recherche Scientifique

UMR8197, 75005 Paris, France; (4) Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA; (5) División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), 78216 San Luis Potosí, Mexico; (6) Department of Biology, Technion, Haifa 32000, Israel; (7) Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Mumbai 400 085, India; (8) Department of Microbiology, Faculty of Science and Informatics, University of Szeged, Hungary; (9) DOE Joint Genome Institute, Walnut Creek, California 94598, USA; (10) Fungal Genetics Stock Center, School of Biological Sciences, University of Missouri- Kansas City, USA; (11) Chemistry and Biomolecular Sciences, Macqaurie University, Sydney, Australia; (12) TU Berlin, Institut für Chemie, FG Biochemie und Molekulare Biologie, 10587 Berlin, Germany; (13) Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas 77843, USA; (14) Department of Genetics and Applied Microbiology, Faculty of Science and Technology, University of Debrecen, H-4010, Debrecen, Hungary; (15) Instituto de Agroquimica y Tecnologia de Alimentos (IATA), Consejo Superior de Investigaciones Cientificas (CSIC), 46100 Burjassot, Valencia, Spain; (16) Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Universités Aix-Marseille I & II, 13288 Marseille, France; (17) Spanish-Portuguese Centre for Agricultural Research (CIALE), Department of Microbiology and Genetics, University of Salamanca, Salamanca, Spain; (18) Section for Sustainable Biotechnology, Department of Biotechnology, Chemistry and Environmental Engineering, Copenhagen Institute of Technology, Aalborg University, Denmark; (19) IFP, Biotechnology Department, Avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France; (20) Institute of Sciences of Food Production- National Research Council (ISPA-CNR), 70126 Bari, Italy; (21) Chemical and Biological Process Development Group, Pacific Northwest National Laboratory, Richland, WA 99352, USA.

Keywords: fungi – mycoparasitism – genomics – transcriptomics - biocontrol - *Hypocrea*

Abstract

Background

Mycoparasitism, a lifestyle where one fungus is parasitic on another fungus has special relevance when the prey is a plant pathogen, providing a strategy for biological control of pests for plant protection. Probably, the most studied biocontrol agents are species of the genus *Hypocrea/Trichoderma*.

Results

Here we report an analysis of the genome sequences for the two biocontrol species – *T. atroviride* (teleomorph *H. atroviridis*) and *T. virens* (formerly *Gliocladium virens,* teleomorph *H. virens*). Several gene families were expanded in the two mycoparasitic species relative to *T. reesei* or other ascomycetes (e.g. Zn(2)Cys(6) transcription factors, solute transporters of the major facilitator superfamily, short chain alcohol dehydrogenases, S8 peptidases, ankyrin proteins with CCHC zinc finger domains, proteins with WD40, HET and NACHT domains, fungal cell wall degrading enzymes, and non-ribosomal peptide synthetases), and were overrepresented in non syntenic genome regions. A phylogenetic analysis shows that *T. reesei* and *T. virens* are derived relative to *T. atroviride.* The mycoparasitism-specific genes thus arose in a common *Trichoderma* ancestor but were subsequently lost in *T. reesei.* To investigate mechanisms of *Trichoderma* mycoparasitism, we compared the transcriptional responses of *T. atroviride* and *T. virens* with those of a saprotrophic member of the genus *T. reesei* during confrontation with a'prey' fungus (*Rhizoctonia solani*). Striking differences occurred between *T. reesei* and the two mycoparasitic species, but also between the latter two suggesting they use alternative approaches to combat the same 'prey'.

Conclusions

The data offer a better understanding of mycoparasitism, and thus enforce the development of improved biocontrol strains for efficient and environmentally friendly protection of plants.

Background

Mycoparasitism is the phenomenon whereby one fungus is parasitic on another fungus, a lifestyle that can be dated at least 400 myo years back by fossil evidence (1). This has special relevance when the prey is a plant pathogen, providing a strategy for biological control of pests for plant protection ("biocontrol"). The movement toward environmentally friendly agricultural practices over the past two decades has thus accelerated research in the use of biocontrol fungi (2). Probably, the most studied biocontrol agents (BCAs) are species of the genus *Hypocrea/Trichoderma*, *T. atroviride (Ta)* and *T. virens (Tv,* teleomorphs *Hypocrea atroviridis* and *H. virens* resp.) being among the best mycoparasitic BCAs used in agriculture (3). The beneficial effects of *Trichoderma* spp. on plants comprise traits such as the ability to antagonize soil-borne pathogens by a combination of enzymatic lysis, secretion of antibiotics, and competition for space and substrates (4,5). In addition, it is now known that some *Trichoderma* biocontrol strains also interact intimately with plant roots, colonizing the outer epidermis layers, and acting as opportunistic, avirulent plant symbionts (6).

Science-based improvement of biocontrol agents for agricultural applications requires an understanding of the biological principle of their action. So far, some of the molecular aspects - such as the regulation and role of cell wall hydrolytic enzymes and antagonistic secondary metabolites - have been studied in *Trichoderma* (3-5). More comprehensive analyses (e.g. by the use of subtractive hybridization techniques, proteomics or expressed sequence tag (EST) approaches) have also been performed with different *Trichoderma* species, but the interpretation of the data obtained is complicated by the lack of genome sequence information for the species used (reviewed in [7]).

Recently, the genome of another *Trichoderma*, *T. reesei* (*Tr*, teleomorph *H. jecorina*), which has a saprotrophic lifestyle and is an industrial producer of biomass hydrolyzing

enzymes, has been sequenced and analysed (8). Here we report the genome sequencing and comparative analysis of two widely used biocontrol species of *Trichoderma*, i.e. *Ta* and *Tv.* These two were chosen because they are distantly related to *Tr* (9) and represent well defined phylogenetic species (10, 11) in contrast to *T. harzianum* sensu lato, which is also commonly used in biocontrol but constitutes a clade of several cryptic species (12). We also analyzed their transcriptome when confronted with the plant pathogen *Rhizoctonia solani* (teleomorph *Thanatephorus cucumeris*, Basidiomycota, Dikarya), and compared it to that of the saprotrophic *Tr* (13) cultivated under the same conditions.

RESULTS

Properties of the *T. atroviride* and *T. virens* genomes

The genomes of *Ta* IMI 206040 and *Tv* Gv29-8 were sequenced using a whole genome shotgun approach to approximately eight-fold coverage and further improved. Their genome sizes were 36.1 (*Ta*) and 38.8 Mbp (*Tv*), and thus larger than the 34 Mbp determined for the genome of *Tr* (8). Gene modeling, using a combination of homology and *ab initio* methods, yielded about 11865 gene models for *Ta* and 12518 gene models for *Tv* respectively (Table 1), both greater than the estimate for *Tr* (9143). As shown in Figure 1, the vast majority of the genes occur in all three *Trichoderma* species. Yet *Tv* and *Ta* contain about 2215 and 1856 genes, respectively, which have no true orthologue in any of the other species, whereas *Tr* has only 3 unique genes. Thus, *Tv* and *Ta* have practically all genes of *Tr*, which makes a comparison with it more straightforward. They share 1873 orthologues that are not present in *Tr*, which could thus be part of the factors that make *T. atroviride* and *T. virens* mycoparasites.

Genome Synteny

A comparison of the genomic organization of genes in *Ta, Tv* and *Tr* showed that most of the genes are in synteny: only 92 (1 %) genes of *Tr ,* whereas 686 (5.5 %) genes of *Tv* and 1012 (8.5 %) genes of *Ta* are located in non-syntenic regions (identified as a break in synteny by a series of three or more genes; Table 2). As observed for other fungal genomes (14-16), extensive rearrangements have occurred since the separation of these three fungi with the prevalence of small inversions in fungal genome evolution as observed previously between genes in other fungi (17). The numbers of the synteny blocks increased with their decreased size, compatible with the random breakage model (18) as in Aspergilli (16, 17).

The evolutionary distances (sequence identity between orthologous and syntenic proteins) were 70 (*Tr* vs. *Ta*), 78 (*Tr* vs. *Tv*), and 74 (*Tv* vs. *Ta*) %, respectively. These values are in the same range as those calculated for Aspergilli (e.g. *A. fumigatus* vs *A. niger* (sixty-nine) and *A. nidulans* (sixty-eight), and resemble those between fish and man (17). Given this evolutionary distance, the level of synteny (*vide supra*) is remarkable.

Transposons

A scan of the genome sequences with the *de novo* repeat finding program 'Piler' (19) – which can detect repetitive elements that are least 400 bp in length, have more than 92% identity and are present in at least 3 copies - unsuccessful at detecting repetitive elements. The lack of repetitive elements detected in this analysis is unusual in filamentous fungi and suggests that, like the *Tr* genome (8), but unlike most other filamentous fungi, the *Ta* and *Tv* genomes lack a significant repetitive DNA component of the genome.

We also scanned the genomes with RepeatMasker and RepeatProteinMask to

identify sequences with similarity to known transposable elements from other organisms. Thereby, sequences with significant similarity to known transposable elements (TEs) from other eukaryotes were identified (Table 3). In most cases, the TE families that we detected were fragmented and highly divergent from one another, suggesting that they did not arise from recent transposition events. Based on these results, we conclude that no extant, functional TEs exist in the *Trichoderma* genomes. The presence of ancient, degenerate TE copies suggests that *Trichoderma* spp. are occasionally subject to infection, or invasion by TEs, but that the TEs are rapidly rendered unable to replicate and rapidly accumulate mutations.

Paralogous gene expansion in *T. atroviride* and *T. virens*

We used MCL analysis (20) and included 9 additional ascomycete genomes present in the JGI genome database (including Eurotiomycetes, Sordariomycetes and Dothidiomycetes) to identify paralogous gene families that have become expanded either in all three *Trichoderma* spp. or only in the two mycoparasitic *Trichoderma* spp**.** Forty-six such families were identified for all three species, of which 26 were expanded only in *Ta* and *Tv*. The largest paralogous expansions in all three *Trichoderma* spp. have occurred with genes encoding Zn(2)Cys(6) transcription factors, solute transporters of the major facilitator superfamily, short chain alcohol dehydrogenases, S8 peptidases and proteins bearing ankyrin domains (Table 4). The most expanded protein sets, however, were those that were considerably smaller in *Tr* ($p < 0.05$). They included ankyrin proteins with CCHC zinc finger domains, proteins with WD40, HET and NACHT domains, NAD-dependent epimerases, and sugar transporters.

Carbohydrate active enzymes. Mycoparasitism depends on a combination of events that include lysis of the preys' cell wall (Figure 2). The necessity to degrade the carbohydrate armour of the prey hyphae is reflected in an increased abundance of chitinolytic enzymes (composing most of the CAZy glycoside hydrolase family GH18 fungal proteins along with more rare endo-β-N-acetylglucosaminidases) and β-1,3-glucanases (families GH17, GH55, GH64, and GH81) in *Trichoderma* relative to other fungi. Family GH18, containing enzymes involved in chitin degradation is also strongly expanded in *Trichoderma*, but particularly in *Tv* and *Ta* which contain the highest number of chitinolytic enzymes of all described fungi (Table 5). Chitin is a substantial component of fungal cell walls and therefore chitinases are an integral part of the mycoparasitic attack (3, 21). It is conspicuous that not only the number of chitinolytic enzymes was elevated, but that many of these additional proteins contain CBMs. Mycoparasitic *Trichoderma* spp. are particularly rich in subgroup B chitinases that contain CBM1 modules, historically described as cellulose binding modules, but binding to chitin has also been demonstrated (22). *Tv* and *Ta* have each a total of five CBM1-containing GH18 enzymes. Subgroup C chitinases possess CBM18 (chitin-binding) and CBM50 modules (described as peptidoglycan-binding modules). Interestingly, CBM50 modules in *Trichoderma* are not only found in chitinases, but also frequently as multiple copies in proteins containing a signal peptide, but with no identifiable hydrolase domain. In most cases the respective genes can be found adjacent to chitinases in the genome. Proteins containing CBM50 modules were recently classified into different groups by de Jonge and Thomma (23). Proteins that consist solely of CBM50 modules are type-A LysM proteins, and there is evidence for the role of LysM proteins as virulence factors in plant pathogenic fungi. However, high numbers of LysM proteins can also be found in saprophytic fungi, indicating other/additional roles for these proteins in fungal biology that are not understood yet. The

presence these repeat proteins is however compatible of an alternative role in protecting the fungal cell wall by masking chitin against the action of chitinases of both endogenous and exogenous origin, necessary mycoparasitism and for protection against plant defense chitinases, respectively.

Together with the expanded presence of chitinases, the number of GH75 chitosanases is also enriched in all three analyzed *Trichoderma* spp. Chitosan is a partially deacetylated derivative of chitin and it has been speculated that in mature fungal cell walls chitin is partially deacetylated and that fungi deacetylate chitin as a defense mechanism against e.g. plant chitinases (24, 25). Chitosan degradation may therefore also be a relevant aspect of mycoparasitism that has not been regarded yet, as some fungi associated with the rhizosphere might have a more enriched chitosan cell wall.

As with plant pathogenic fungi (26, 27), we have also observed an expansion of plant cell wall degrading enzyme gene families. Overall, the carbohydrate-active enzyme machinery is compatible with a saprophytic behaviour, but the set of enzymes involved in the degradation of "softer" plant cell-wall components such as pectin is reduced. A possible plant symbiotic relationship (3) would therefore rely on a mycoparasitic capacity along with a reduced specificity for pectin, thereby minimizing the plant defense reaction. A full account of all the carbohydrate active enzymes is presented in **Supplementary Tables S1–S6**. Additional details about the Trichoderma CAZome are given in **Supplementary Appendix Chapter 1**.

Secondary metabolite biosynthesis. With respect to gene families commonly associated with secondary metabolite biosynthetic pathways, the three *Trichoderma* spp. contained a varying assortment of non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) (Table 6; see also **Supplementary Tables S7-S8**). While *Tr* (10 NRPS, 11 PKS

and 2 NRPS/PKS fusion genes; [8]) ranked at the lower end when compared to other ascomycetes, Tv exhibited the highest number (fifty) of PKS, NRPS and PKS-NRPS fusion genes, mainly due to the abundance of NRPS genes: twenty eight, twice as much as in other fungi. A phylogenetic analysis showed that this was due to recent duplications of cyclodipeptide synthases, cyclosporin/enniatin synthase-like proteins, and NRPS-hybrid proteins (**Supplementary Figure S1**). Most of the secondary metabolite gene clusters present in *Tr* were also found in *Tv* and *Ta,* but about half of the genes remaining in the latter two are unique for the respective species, and occur within non-syntenic islands of the genome (see below). Within the NRPS, all three *Trichoderma* spp. contained two peptaibol synthases, one for short (10-14 aa) and one for long (18-25 aa) peptaibols. The long peptaibol synthetase encoding genes lack introns and produce an mRNA that is 60 to 80 kb long encoding ~25,000 amino acid proteins, the largest fungal proteins known.

Evolution of the non-syntenic regions

A search of overrepresentation of PFAM domains and GO terms in these non-syntenic regions revealed that all RHS-(retroposon hot spot) repeat domains (28) are found in the non-syntenic regions. In other eukaryotes, these regions are located in subtelomeric areas which exhibit a high recombination frequency (29). In addition, the genes for the protein families in *Tv* and *Ta* that were significantly more abundant compared to *Tr* were enriched in the non-syntenic areas („NS-genes"; Table 7). Also the number of paralogous genes was significantly increased in the NS regions. We considered three possible explanations for this: (i) the NS genes were present in the last common ancestor of all three *Trichoderma* species but were then selectively and independently lost; (ii) the non-syntenic areas arose from the core genome by duplication and divergence during evolution of the genus *Trichoderma*; and

(iii) the NS genes were acquired by horizontal transfer. To distinguish among the hypotheses for their origin, we compared the sequence characteristics of the genes in the NS regions to those present in the syntenic regions in *Trichoderma* and genes in other filamentous fungi. We found that the majority (> 78 %) of S as well as the NS-encoded proteins have their best BLAST hit to other ascomycete fungi, indicating that also the NS regions are of fungal origin. Also, a high number of proteins encoded in the NS regions of *Ta* and *Tv* have paralogs in the S region. Finally, codon usage tables and codon adaptation index (CAI) analysis (30) indicate that the NS-genes exhibit a similar codon usage (**Supplementary Figure S2**). Taken together, the most parsimonious explanation for the presence of the paralogous genes in *Ta* and *Tv* is that the NS genes arose by gene duplication within a *Trichoderma* ancestor, followed by gene loss in the three lineages which was much stronger in *Tr*.

Tr, Ta and *Tv* each occupy very diverse phylogenetic positions in the genus, as shown by a Bayesian *rpb2* tree of 110 *Trichoderma* taxa (Figure 3 A). In order to determine which of the three species more likely resembles the ancestral state of *Trichoderma*, we performed a Bayesian phylogenetic analysis (31) using a concatenated set of 100 proteins that contained orthologues in syntenic positions in the three *Trichoderma* species and also *Gibberella zeae* and *Chaetomium* (http://genome.jgi-psf.org/Chagl_1/Chagl_1.home.html). The result (Figure 3 B) shows that *Ta* occurs in a well-supported basal position to *Tv* and *Tr*. These data indicate that *Ta* resembles the more ancient state of *Trichoderma* and that both *Tv* and *Tr* evolved later. The lineage to *Tr* thus appears to have lost a significant number of genes present in *Ta* and maintained in *T. virens*. The long genetic distance of *Tr* further suggests that it was apparently evolving faster since the time of divergence.

Transcriptional response of *Trichoderma* to the presence of *R. solani*.

To relate the gene inventory in the three *Trichoderma* spp. to the molecular events accompanying attack of a prey, we performed microarray analysis of *Tv*, *Ta* and *Tr* at three stages: (1) prior to contact, (2) during initial physical contact, and (3) during overgrowth of the *R. solani* (see Figure 2) The latter is a plant pathogenic fungus with a wide host range and worldwide distribution, which causes death of seedlings in agriculture (32), and is frequently used as a model host for *Trichoderma* (3-5). The three *Trichoderma* spp. thereby revealed essential differences in their response to *R. solani (*Fig. 3 B*)*: *Tr* was unable to attack it, and built up a dense fence of hyphae at the border between itself and *R. solani* which the latter could not overcome. *Tv* completely arrested the growth of *R. solani* and then parasitized on it. *Ta,* in contrast did not stop growth of *R. solani*, but continued to overgrowth it and eventually finally also parasitized on it.

The *T. virens* tiling arrays detected 8007, 8082 and 7946 transcripts in mycelia (from before, at and after contact, respectively), equaling the expression of more than 70 % of the genes annotated in the genome. When the expression patterns were normalized to that of a *T. virens* culture confronted by itself (to compensate for nutrient gradients arising from fungal growth), 213, 225 and 77 genes (from before, at and after contact, respectively) were found to be differentially expressed with respect to the control. Four-fifths of these genes were downregulated, whereas only a fifth of them increased during one or more stages of confrontation (Figure 4). Only three genes were upregulated under all stages of interaction, i.e. a cyanide hydratase (Trive1:78142), the gliotoxin synthase GliP (Trive1:78708) and a hypothetical protein (Trive1:64933) bearing low similarity to an amidohydrolase of *Penicillium marneffei*. A full list of the genes we selected as upregulated is given in **Supplementary Table S9**. The majority (>70 %) of upregulated genes comprised those involved in gliotoxin biosynthesis, glutathione biosynthesis, response to stress, glycosyl

hydrolases and solute transporters. The latter comprised 10 of the 27 genes expressed already before contact. The strong upregulation of gliotoxin biosynthetic genes suggests that gliotoxin is the principle by which *Tv* arrests the growth of R. solani (*vide supra*).

In *T. atroviride*, the tiling arrays detected 7679, 8379 and 7661 transcripts in mycelia (from before, at and after contact, respectively), equaling the expression of 71.2 % of the genes annotated in the genome. When these expression patterns were normalized to those of a *Ta* culture confronted by itself (to compensate for nutrient gradients arising from fungal growth), 458, 571, and 12 genes (from before, at and after contact, respectively) were found to be differentially expressed with respect to the control. About a third of the genes were upregulated before contact, half at contact, and all but one during overgrowth (Figure 4). A full list of the genes we selected as upregulated is given in **Supplementary Table S10.** In contrast to *Tv*, unknown and unique genes comprised a significant portion of them. However, among genes for which a function or name could be attributed, overexpression of PTH11 receptors, C-type lectins, and small secreted cystein-rich proteins (SSCRPs) were observed. Also the abundance of GH16 ß-1,3-glucanases within the highly expressed CAZy genes was striking. However, proteases of various groups were the major hydrolytic enzymes secreted by *Ta*, accompanied by the overexpression of oligopeptide and amino acid permeases. As for secondary metabolites, one polyketide synthase (PKS) of the reducing (lovastatin/citrinin) clade I (Triat2:134224; see **Supplementary Figure S3**) was expressed at all stages of interaction, and peaked at the phase of contact. A second (Triat2:85006) from the reducing clade was expressed only at the stage of contact. Interestingly, *Ta* also expressed two KP4 killer-toxin like genes, antifungal proteins that have so far not been described for filamentous fungi. The overexpression of genes encoding oxidative enzymes such as cytochrome P450, dioxygenases, monooxygenases and FAD-linked extracellular

oxidase was also notable, and these genes may be involved either in the formation of unknown secondary metabolites or protection against metabolites secreted by *R. solani*.

In *Tr*, 6265, 6634 and 6660 transcripts were detected before, at and after contact, respectively, also equaling the expression of 71-72 % of the genes annotated in the genome. 250, 18 and 269 genes (from before, at and after contact, respectively) were found to be differentially expressed with respect to the control. The portion of upregulated genes among these was strongly different, ranging from >60 % (before contact) to 50 % (contact) down to 35 % (overgrowth) (Figure 4). A full list of the genes we selected as upregulated is given in **Supplementary Table S11**. The trancriptome of *Tr* before contact did not reveal any signs of sensing or attempting to attack the prey. In contrast, it showed a massive upregulation of cellulolytic and hemicellulolytic CAZys, ribosomal proteins and transporters for metal ions (especially iron). The upregulation of cellulases is likely due to the fact that the agar plates were covered with cellophane for easy removal of mycelia, which was partially metabolized by *T. reesei*. The phase of contact with *R. solani* is represented by a strong reduction in gene expression. Genes encoding a multidrug transporter, a stress-responsive protein (RDS1) and a polyketide synthase were strongly induced and may indicate a shift to defense against the confrontant. The phase of overgrowth is characterized again by a strong increase in the genes encoding ribosomal proteins, but not of cellulolytic and hemicellulolytic CAZys. *Tr* accumulates a few glucanases and transcription factors, but strikingly lacks expression of proteases and (with the exception of the above mentioned PKS) secondary metabolite genes.

Genome distribution of *Trichoderma* genes differentially expressed during confrontation with *R. solani*

Mapping the differentially regulated genes on their scaffolds revealed that 10 – 50 % of them occurred non-randomly distributed in the genome and displayed at least a four-fold increased gene density over the average (expressed genes/total number of genes; Figure 5). This non-random occurrence suggests an epigenetic component to regulation of their expression, although not necessarily specific only for mycoparasitism. This was most evident in *Ta*. About a third of co-expressed gene clusters were located in the NS-areas. This may have created a favorable environment for the rapid generation of novel genes needed for adaptation to new ecological niches (29).

Discussion

Comparison of the genomes and transcriptomes of two mycoparasitic and onesaprotrophic *Trichoderma* species revealed remarkable differences: in contrast to other multicellular ascomycetes genomes such as Aspergilli (16, 17), those of *Trichoderma* appear to be have the highest level of synteny (>90 % vs. 68-75 %), and most of the differences between *Ta* and *Tv* versus *Tr* or other ascomycetes occur in the NS areas. Nevertheless, at a molecular level the three species are distant from each other like Apes from *Pices* (fishes) or *Aves* (birds) (17), suggesting that their genomic synteny undergoes strong evolutionary pressure. Our data also suggest that the ancestral state of *Hypocrea /Trichoderma* was mycoparasitic. This supports an earlier speculation (33) that the ancestors of *Trichoderma* were mycoparasites on wood degrading basidiomycetes and acquired the saprotrophic ability to follow their hosts into their substrate.

Furthermore, the three *Trichoderma* spp. have the lowest number of transposons reported so far. This is unusual for filamentous fungi, as most species contain approximately 10%-15% repetitive DNA, primarily composed of transposable elements. A notable

exception is *F. graminearum* (26), which, like *Trichoderma* spp. contains less than 1% repetitive DNA (8). The paucity of repetitive DNA may be attributed to Repeat-Induced Point Mutation (RIP), which has been reported in both *F. graminearum* and *Tr*. It is likely that this process also contributes to prevent the accumulation of repetitive elements in the *Ta* and *Tv* genomes as well.

The strong expansion of genes for solute transport, oxidoreduction, ankyrins (a family of adaptor proteins that mediate the anchoring of ion channels or transporters in the plasma membrane; [34]), would render *Trichoderma* more compatible in this habitat (e.g. to successfully compete with the other saprotrophs for limiting substrates). In addition, the expansion of WD40 domains that act as hubs in cellular networks (35) could aid in more versatile metabolism or response to stimuli. These features correlate well with a saprotrophic life style that makes use of plant biomass that has been pre-degraded by earlier colonizers. The expansion of HET proteins (proteins involved in vegetative incompatibility specificity) further indicates that *Trichoderma* spp. may frequently encounter related yet genetically distinct individuals. In fact, the presence of several different *Trichoderma* species can be detected in a single soil sample (36).

Gene expansions in *Tv* and *Ta* that do not occur in *Tr* may comprise genes related to mycoparasitism. As a prominent example, proteases have expanded in *Ta* and *Tv*, supporting the hypothesis that the degradation of proteins is a major trait of mycoparasites (37). Also the increase in chitinolytic enzymes and ß-glucanases is remarkable and illustrates the importance of destruction of the prey´s cell-wall in mycoparasitism. Another outstanding property of particularly *Tv* is the enlarged number of genes for the synthesis of NRPS and the generally high number of PKS and NRPS in *Ta* and *Tv*. Among the latter, the transcriptome data suggest that gliotoxin - an important virulence factor of invasive aspergillosis (38, 39) –

is also the major antifungal metabolite of *Tv*, whereas the synthesizing gene cluster in *Ta* is absent. Gliotoxin is a epipolythiodioxopiperazine toxin, whose formation has been described as one of *Tv*'s antifungal principles (40-44). It is also a product from the filamentous fungus *Aspergillus fumigatus* (45) and contributes to the virulence of the latter in nonneutropenic mice and in fruit flies with functional phagocytes. It is synthesised from the amino acids phenylalanine and serine by a nonribosomal peptide synthase GliP. The resulting L-phenylalanyl-L-seryl compound then undergoes a series of oxidative enzymatic transformations before methylation to mature gliotoxin. Evidence for this reaction pathway is limited but comparison with sirodesmin biosynthesis in *Leptosphaeria maculans* has provided valuable insights (40). Patron et al. (46) identified a gliotoxin biosynthesis cluster in *Tv* and *Tr.* Interestingly, maximum likelihood phylogeny of six concatenated proteins from the cluster placed the proteins from the two *Trichoderma* spp. in different clades, that of *Tv* thereby clustering with *L. maculans* sirodesmin. Search of the *Trichoderma* genomes showed that indeed *Tv* contains three GliP-like proteins, *Tr* two, and *Ta* one. A neighbor joining analysis of the GliP-like proteins from *Trichoderma* with other closely similar proteins from other fungi shows that GliP 70742 from scaffold 40 forms (together with a protein from *Penicillium marneffei*) a basal branch to a clade consisting of only *Trichoderma* and *Talaromyces stipitatus* NRPS proteins which are not a member of a secondary metabolic cluster (**Supplementary Figure S4**). These branches are part of a larger, well supported bifurcating clade which we – due to the presence of *L. maculans* in it – name SirP clade. *Tv* GliP 78708 (not reported by Patron et al. [46] and *Tr* GliP 24586 are part of clade containing *A. fumigatus* GliP and we thus consider these two proteins as GliP orthologs. This conclusion is also supported by the microarray results that showed only the genes from the cluster on scaffold 54 – but not that on scaffold 40 – to be expressed during antagonism. Wilhite and

Straney (42) reported that the accumulation of gliotoxin occurs only during a short phase of cultivation. In contrast, the present microarray data show that most of the genes responsible for gliotoxin biosynthesis are induced already before contact, expressed throughout the period of antagonism, and only some of them declined during the phase of overgrowth (see **Supplementary Table S9**). Gliotoxin biosynthesis may be limited at the level of its precursor amino acids during the phase of overgrowth, as an enzyme of L-phenylalanine biosynthesis (phospho-2-dehydro-3-deoxyheptonate aldolase 84699) was upregulated throughout, and it was also conspicuous that genes related to glutathione (glutamate cysteine ligase 75334, γ-glutamyltransferase 46190) biosynthesis were upregulated as well. All in all, these data suggest that gliotoxin biosynthesis is a major event in *Tv* during antagonism of *R. solani*.

Besides PKS and NRPS, *Ta* and *Tv* have further augmented their antibiotic arsenal by genes for cytolytic peptides such as aegerolysins, pore forming cytolysins typically present in bacteria, fungi and plants, yeast-like killer toxins and cyanovirins (**Supplementary Appendix, Chapter 2**). In addition, we found two high molecular weight toxins in *Ta* and *Tv* that bear high similarity (E value 0 for 97 % coverage) to the Tc-toxins of *Photorhabdus luminescens*, a bacterium which is mutualistic with entomophagous nematodes ([47]; **Supplementary Table S12**). Yet there may be several more secondary metabolite genes to be detected: *Trichoderma* spp. contain expanded arrays of cytochrome P450 CYP4/CYP19/CYP26 subfamilies (**Supplementary Table S13**). There are also soluble epoxide hydrolases that could act on the epoxides produced by the latter (**Supplementary Figure S5**). These genes may define non-clustered secondary metabolite biosynthetic routes (48).

The life style of *Trichoderma* spp. as rhizosphere competent, sometimes endophytic mycoparasites and saprobes implies that they are able to recognize and interact with both fungi and plants. As for fungi, the present study demonstrates that *Ta* and *Tv,* but not *Tr*,

respond to the presence of the prey already before physical contact, suggesting a sensing mechanism. There is some evidence (37) that soluble decomposition products from the prey's cell wall may be perceived by G-protein coupled receptors (GPCRs). All three *Trichoderma* spp. possess an abundant number of members of GPCR PTH11 class ([49]; **Supplementary Table 14**). PTH11 encoding genes were also among those up-regulated in the mycoparasite *Coniothyrium minitans* during colonization of *Sclerotinia sclerotiorum* (50), thus rendering them candidates for a function in mycoparasitism. In addition, all three *Trichoderma* spp. contain an arsenal of C-type lectins that were particularly abundant and over expressed before and at the contact in *Ta* (**Supplementary Table 15**). We also identified several proteins consisting of carbohydrate binding (CBM13) modules that were first identified in plant lectins such as ricin (51). *Ta* also induced a gene encoding a cyanovirin gene, a mannose-binding lectin (52). Lectins have been suggested to be involved in coiling of *Trichoderma* mycoparasites around their prey (53), but they may also mediate the contact between *Trichoderma* and plants. Finally, we found an abundant assortment of small cysteine-rich proteins (SSCP), known to act as effector proteins which manipulate host cell signaling or to suppress defense pathways during infection in pathogenic species and ectomycorrhizae (54) (**Supplementary Appendix, Chapter 3**). The ability of members from one of their subgroups – the ceratoplatanin-like SSCPs – to stimulate plant defense has been demonstrated in *Tv* (55), but an effect on the prey is unknown. Upon phylogenetic analysis, the largest group of SSCPs - Cluster 1 (**Supplementary Table 16**), is organized (with two exceptions) into sets of homologs in each species (**Supplementary Figure S6**). All these predicted proteins are of unknown function, but some members are predicted to contain CFEM domains or consensus sequences for GPI anchors. Conservation of these genes suggests that they may encode cell surface proteins with important roles in all three species.

Some members of Cluster 1 are upregulated in *T. atroviride* during antagonistic interaction (Supplementary Table S10).

All three *Trichoderma* spp. also displayed an oxidative response in confrontation with the prey (expression of heat shock proteins, cytochrome C peroxidase, proline oxidase, and ER-bound glutathione-S-transferases). Likewise, genes for detoxification processes (ABC efflux transporters, the pleiotropic drug resistance (PDR) transporters and the multidrug resistance MDR-type transporters) were induced. *R. solani* has been shown to use radical oxygen species as signaling molecules during sclerotia formation (56), and excrete antifungal components (57), both of which may have elicited this response. An ABC-transporter from *Ta* (TAABC2) has already been shown to be involved in biocontrol of *R. solani* (58).

A nitrilase gene was the most strongly induced gene in *Tv* under all stages of antagonism, and its orthologue was up-regulated also in *Ta* (cf. **Supplementary Tables 9-11**) and *Tr*. A role of nitrilases in antagonism of *R. solani* or other fungal preys is not known. However, they detoxify cyanide when in contact with plant roots, and can convert the plant metabolite indole-3-acetonitrile (IAN) to the plant growth hormone indole-3-acetic acid (IAA; [59]), and they, therefore, may be a part of the genes expressed by *Trichoderma* during endophytism. Their induction during the confrontation with *R. solani* may thus suggest a linkage between sensing of a fungal prey and plant roots which would also explain the up-regulation of C-type lectins and SSCPs (*vide supra*).

Conclusions

Our comparative genome analysis of the three *Trichoderma* spp. now opens new opportunities for the development of improved and research-driven strategies to select and improve *Trichoderma* spp. as biocontrol agents. The availability of the genome sequences

published in this study, as well as of several pathogenic fungi and their potential host plants (cf. http://www.diark.org/diark/search) further provides a challenging opportunity to develop a deeper understanding of the underlying processes by which *Trichoderma* interacts with plant pathogens in the presence of living plants within their ecosystem.

Materials and Methods

Genome sequencing and assembly.

The genomes of *Trichoderma virens* and *Trichoderma atroviride* each were assembled from shotgun reads using the JGI (Joint Genome Institute, USA Dept. of Energy) assembler, Jazz (**Supplementary Table S17** for summary of assembly statistics). Each genome was annotated using the JGI Annotation pipeline, which combines several gene prediction, annotation and analysis tools. Genes were predicted using Fgenesh (60), Fgenesh+ (61), and Genewise programs (62). ESTs from each species (see **Supplementary Appendix Chapter 4**) were clustered and either assembled and converted into putative full-length (FL) genes directly mapped to genomic sequence or used to extend predicted gene models into FL genes by adding 5' and/or 3' UTRs to the models. From multiple gene models predicted at each locus, a single representative model was chosen based on homology and EST support and used for further analysis. Gene model characterisics and support are summarized in **Supplementary Tables S18 and S19**.

All predicted gene models were functionally annotated by homology to annotated genes from NCBI non-redundant set and classified according to Gene Ontology (63), eukaryotic orthologous groups (KOGs; [64]), and KEGG metabolic pathways (65). See **Supplementary Table S20 and S21** for a summary of the functional annotation. Automatically predicted genes and functions were further refined by user community-wide

manual curation efforts using web-based tools at www.jgi.doe.gov/trive1 and www.jgi.doe.gov/triat1 . The latest version gene set containing manually curated genes is called GeneCatalog

Assembly and annotation data for *T.virens* and *T.atroviride* are available through JGI Genome Portals homepage at http://www.jgi.doe.gov/trive and http://www.jgi.doe.gov/triat. The genome assemblies, predicted gene models, and annotations were deposited at DDBJ/EMBL/GenBank under the project accessions (ABDF00000000.1 and ABDG00000000.1, respectively). Automatically predicted genes and functions were further refined by user community-wide manual curation efforts using web-based tools at www.jgi.doe.gov/trive1 and www.jgi.doe.gov/triat1 . The latest version gene set containing manually curated genes is called GeneCatalog

Genome Similarity Analysis and Genomic Synteny

Orthologous genes, as originally defined, imply a reflection of the history of species. In recent years, many studies have examined the concordance between orthologous gene trees and species trees in bacteria. With the purpose of identifying all the orthologous gene pairs for the three *Trichoderma* spp., a Best Bidirectional Blast Hit (BBH) approach as described elsewhere (66, 67) was performed, using the predicted translated gene models for each of the three species as pairwise comparing set.

The areas of relationship known as syntenic regions or syntenic blocks are anchored with orthologs (calculated as mutual best hits or bi-directional best hits) between the two genomes in question, and are built by controlling for the minimum number of genes, minimum density, and maximum gap (genes not from same genome area) as compared with randomized data as described in (68). While this technique may cause artificial breaks, it

highlights regions that are dynamic and picking up a large number of insertions or duplications.

Orthologous and paralogous gene models were identified by first using BLAST to find all pairwise matches between the resulting proteins from the gene models. The pairwise matches from BLAST were then clustered into groups of paralogs by using MCL (http://micans.org/mcl/). In parallel we applied orthoMCL (69) to the same pairwise matches to identify the proteins that were orthologous in all of the three genomes. by subtracting all the proteins that were identified as orthologs from the groups of paralogs we were left with only the the protein products of gene models that have expanded since the most recent common ancestor (MRCA) of the three *Trichoderma* genomes. In addition, the MCL and orthoMCL results also identified the proteins that arose since the MRCA and appear to have no similarity on the other two genomes. We term these unique gene models singletons. We then calculated the p-value under the null hypothesis that the number of paralogs or singletons that are non-syntenic are less than the number of paralogs or singletons that are syntenic.

Identification of transposable elements

We scanned the *Trichoderma* genomes with the *de novo* repeat finding program Piler (19). Next, we searched for sequences with similarity to known repetitive elements from other eukaryotes with the program RepeatMasker (70) using all eukaryotic repetitive elements in the RepBase (version 13.09) database. After masking repetitive sequences that matched the DNA sequence of known repetitive elements, we scanned the masked genome sequences with RepeatProteinMask, (a component of the RepeatMasker application). This search located additional degenerate repetitive sequences with similarity to proteins encoded by

transposable elements in the RepBase database.

CAZome identification and analysis

All protein models from the *T. atroviride* and *T. virens* were compared against the set of libraries of modules derived from CAZy (Carbohydrate-Active Enzymes database, http://www.cazy.org) as described elsewhere (71). The identified proteins were subject to manual analysis for correction of protein model, for full modular annotation and for functional inference against a library of experimentally characterized enzymes. Comparative analysis was made by the enumeration of all modules identified in the three *Trichoderma* and 14 other published fungal genomes.

Transcriptome analysis

For mycoparasitism confrontation assays *T. virens* Gv29-8, *T. atroviride* IMI 206040 and *T. reesei* QM 6a were grown on potato dextrose agar plates (BD Dicfo, Franklin Lakes, NJ, USA), covered with cellophane, in constant light at 25°C and harvested when the mycelia were ca. 5 mm apart ("before contact"), at contact of the mycelia ("at contact") and after *Trichoderma* had overgrown the host fungus *Rhizoctonia solani* by ca. 5 mm ("after contact"). As control, the respective strain of *Trichoderma* was confronted with itself and harvested at contact. Peripheral hyphal zones from each confrontation stage were harvested and shock frozen in liquid nitrogen. Mycelia were ground to a fine powder under liquid nitrogen and total RNA was isolated using the guanidinium thiocyanate method (72). For cDNA synthesis, RNA was treated with DNase I (Fermentas, Burlington, Canada) and purified with the RNeasy MinElute Cleanup Kit (Qiagen, Valencia, CA, USA). 5 µg RNA/reaction were

reverse transcribed using the the SuperScript™ III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) and a mixture of random hexamer primer and oligo$(dT)_{18}$ primer.

We designed *T. virens*, *T. atroviride* and *T. reesei* tiling array with 60 mer oligonucleotides (oligo length) each 93 bp apart (oligo distance), using the unmasked fasta file of the genomes of *T. virens*, *T. atroviride* and *T. reesei*. The design was done with Teloenn (73) using a MAX_PREFIX_LENGTH set to 30 for the uniqueness calculation made by genome tools. Complexity is evaluated using the masked genome by counting the number of masked bases for each probe. Tm values are calculated using the nearest neighbour thermodynamic model. No filters were applied after probe parameter calculations. In order to obtain the probe quality score, the calculated parameters were weighted as follows: 0.4 for Tm, 0.3 for uniqueness, 0.2 for GC content, and 0.1 for complexity. To get final oligonucleotide scores, a weighting of 0.75 was assigned to quality scores, and a weighting of 0.25 was assigned to position scores. Teolenn software designed 415373, 387105 and 359089 final probes respectively for *T. virens*, *T. atroviride* and *T. reesei*. Oligonucleotides were loaded on eArray software and 2x 400k microarrays were obtained from Agilent.

The microarray data and related protocols are available at the GEO web site (www.ncbi.nlm.nih.gov/geo/) under accession number: GSE23438. Briefly, the RNAs of two independent biological replicates for each condition were labelled with Cy3 or Cy5 dye using the indirect labelling procedure. Possible dye bias was eliminated using dye switch labelling protocol. We then hybridized 1.5 µg of labelled cDNA with the 2x400k DNA chip (Agilent). The array was read using an Agilent G2505C DNA microarray scanner and the TIFF images extracted with the Agilent Feature Extraction software (version 10.5.1.1) using the 20 bit coding ability. Data pre-treatment was applied on each result file to discard flagged spots by Feature Extraction software. The data were normalized without background subtraction by

the global Lowess method performed with the Goulphar software (74). For each experimental condition, the two file results were merged together. For each probe the hybridization ratio was linked to genome annotation coming from the JGI website. The final ratio for each transcript was obtained by averaging the detected hybridization values from all probes located inside the coding sequence on the matching strand. Transcripts with no or only one probe marked as detected were discarded from further analysis. Finally we kept only transcript with a final hybridization ratio greater than $\log_2 = 1.5$ or lower than $\log_2 = -1.5$. Clustering analysis was done using the MultiExperiment Viewer software (75). For each gene list, an expression matrix was built gathering the three conditions, and a hierarchical clustering was performed using Euclidian distance and the average linkage method. Statistical treatment of the data was done by applying the linear modeling approach implemented by lmFit and the empirical Bayes statistics implemented by eBayes from the limma R package (76). Identified proteins were categorized according to the Functional Catalogue (77).

References

1. Taylor TN, Hass H, Kerp H, Krings M, Hanlin RT: **Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism.** *Mycologia* 2005, **97:** 269-285.
2. Vincent C, Goettel MS, Lazarovits G: **Biological control: a global perspective.** CAB International, Wallingfort, UK. 2007.
3. Harman GE, Howell CR, Viterbo A, Chet I, Lorito M: ***Trichoderma* species-opportunistic, avirulent plant symbionts.** *Nat Rev Microbiol* 2004, **2:** 43-56.
4. Howell CR: **Mechanisms employed by *Trichoderma* species in the biological control of plant diseases: the history and evolution of current concepts.** *Plant Disease* 2003, **87:** 4-10.
5. Harman GE: **Overview of mechanisms and uses of *Trichoderma* spp.** *Phytopathology* 2006, **96:** 190-194.
6. Shoresh M, Harman GE, Mastouri F: **Induced systemic resistance and plant responses to fungal biocontrol agents.** *Annu Rev Phytopathol* 2010, **48:** 21-43.
7. Lorito M, Woo SL, Harman GE, Monte E: **Translational research on *Trichoderma*: from 'omics to the field.** *Annu Rev Phytopathol* 2010, **48:** 395-417.
8. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, et al: **Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*).** *Nat Biotechnol* 2008, **26:** 553-560.
9. Druzhinina IS, Koptchinskiy A, Kubicek CP: **The first one hundred *Trichoderma* species characterized by molecular data.** *Mycoscience* 2006, **47:** 55–64.
10. Chaverri P, Samuels GJ, Stewart EL: ***Hypocrea virens* sp. nov., the teleomorph of *Trichoderma virens*.** *Mycologia* 2001, **93:** 1113-1124.
11. Dodd SL, Lieckfeldt E, Samuels GJ: ***Hypocrea atroviridis* sp. nov., the teleomorph of *Trichoderma atroviride*.** *Mycologia* 2003, **95:** 27-40.
12. Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J: **The *Trichoderma harzianum* demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages.** *BMC Evol Biol* 2010, **10:** 94.
13. Seidl V, Schmoll M, Scherm B, Balmas V, Seiboth B, Migheli Q, Kubicek CP: **Antagonism of *Pythium* blight of zucchini by *Hypocrea jecorina* does not require cellulase gene expression but is improved by carbon catabolite derepression.** *FEMS Microbiol Lett* 2006, **257:** 145-151.
14. Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury J-M, Ségurens B, Poulain J, et al: **The genome sequence of the model ascomycete fungus *Podospora anserina*.** *Genome Biol* 2008, **9:** R77.
15. Fischer G, Rocha EP, Brunet F, Vergassola M, Dujon B**: Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages.** *PLoS Genet* 2006, **2:** e32.

16. Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, Batzoglou S, Lee S-I, Bactürkmen M, Spevak CC, Clutterbuck J, et al*:* **Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*.** *Nature* 2005, **438:** 1105-1115.

17. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, Crabtree J, Silva JC, Badger JH, Albarraq A, et al: **Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*.** *PLoS Genet* 2008, **4:** e1000046.

18. Nadeau J, Taylor B: **Lengths of chromosomal segments conserved since divergence of man and mouse.** *Proc Natl Acad Sci* 1984, **81:** 814-818.

19. Edgar RC, Myers EW: **PILER: identification and classification of genomic repeats.** *Bioinformatics* 2005, **21:** Suppl 1**:**i152-i158.

20. Enright AJ, van Dongen S, Ouzounis CA: **An efficient algorithm for large-scale detection of protein families.** *Nucleic Acids Res* 2002, **30:** 1575-1584.

21. Seidl V: **Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions.** *Fungal Biol Rev* 2008, **22:** 36-42*.*

22. Limon MC, Chacón MR, Mejías R, Delgado-Jarana J, Rincón AM, Codón AC, Benítez T: **Increased antifungal and chitinase specific activities of *Trichoderma harzianum* CECT 2413 by addition of a cellulose binding domain.** *Appl Microbiol Biotechnol* 2004, **64:** 675 - 682.

23. de Jonge R, Thomma BP: **Fungal LysM effectors: extinguishers of host immunity?** *Trends Microbiol* 2009, **17:** 151 – 154.

24. Baker LG, Specht CA, Donlin MJ, Lodge JK. **Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*.** *Eukaryotic Cell* 2007, **6:** 855 - 862

25. El Gueddari NE, Rauchhaus U, Moerschbacher BM, Deising HB: **Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi.** *New Phytol* 2002, **156:** 103 -111.

26. Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, et al: **The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization.** *Science* 2007, **317:** 1400-1402.

27. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu J-R, Pan H, et al: **The genome sequence of the rice blast fungus *Magnaporthe grisea*.** *Nature* 2005, **434:** 980-986.

28. Wellinger RJ, Sen H: **The DNA structures at the ends of eukaryotic chromosomes.** *Eur J Cancer* 1997, **33:** 735-749.

29. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellems TE, Scherf A: **Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*.** *Nature* 2000, **407:** 1018-1022.

30. Sharp PM, Li W-H: **The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications.** *Nucl Acids Res* 1987, **15:** 1281-1295.

31. Yang Z, Rannala B: **Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method.** *Mol Biol Evol* 1997, **14:** 717-724.

32. Jones RK, Belmar SB: **Characterization and pathogenicity of *Rhizoctonia* spp. isolated from rice, soybean and other crops grown in rotation with rice in Texas.** *Plant Dis* 1989, **73:**1004-1010.

33. Rossmann AY, Samuels GJ, Rogerson CT, Lowen R: **Genera of *Bionectriaceae*, *Hypocreaceae* and *Nectriaceae* (*Hypocreales*, *Ascomycetes*).** *Stud Mycol* 1999, **42:** 1- 83

34. Bennett V, Baines AJ. **Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues.** *Physiol Rev* 2001, **81:** 1353-1392.

35. Stirnimann CU, Petsalaki E, Russell RB, Müller CW: **WD40 proteins propel cellular networks.** *Trends Biochem Sci* 2010, **35:** 565-574.

36. Migheli Q, Balmas V, Komoñ-Zelazowska M, Scherm B, Fiori S, Kopchinskiy AG, Kubicek CP, Druzhinina IS: **Soils of a Mediterranean hot spot of biodiversity and endemism (Sardinia, Tyrrhenian Islands) are inhabited by pan-European, invasive species of *Hypocrea/Trichoderma*.** *Environ Microbiol* 2009, **11:** 35-46.

37. Seidl V, Song L, Lindquist E, Gruber S, Koptchinskiy A, Zeilinger S, Schmoll M, Martínez P, Sun J, Grigoriev I, et al: **Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey.** *BMC Genomics* 2009, **10:** 567.

38. Bok JW, Chung D, Balajee SA, Marr KA, Andes D, Nielsen KF, Frisvad JC, Kirby KA, Keller NP: **GliZ, a transcriptional regulator of gliotoxin biosynthesis, contributes to *Aspergillus fumigatus* virulence.** *Infect Immun* 2006, **74:** 6761–6768.

39. Grovel O, Pouchus YF, Robiou du Pont T, Montagu M, Amzil Z, Verbist J: **Ion trap MS(n) for identification of gliotoxin as the cytotoxic factor of a marine strain of *Aspergillus fumigatus* Fresenius.** *J Microbio. Methods* 2002, **48:** 171-179.

40. Gardiner DM, Cozijnsen AJ, Wilson LM, Pedras MS, Howlett BJ: **The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans.*** *Mol Microbiol* 2004, **53:** 1307-1318.

41. Wilhite SE, Lumsden RD, Straney DC: **Mutational analysis of gliotoxin production by the biocontrol fungus *Gliocladium virens* in relation to suppression of *Pythium* damping-off.** *Phytopathology* 1994, **84:** 816-821

42. Wilhite SE, Straney DC: **Timing of gliotoxin biosynthesis in the fungal biological control agent *Gliocladium virens* (*Trichoderma virens*)**. *Appl Microbiol Biotechnol* 1996, **45**: 513-518.

43. Lumsden RD, Locke JC: **Biological control of damping-off caused by *Pythium ultimum* and *Rhizoctonia solani* with *Gliocladium virens* in soilless mix.** *Phytopathology* 1989, **79:** 361-366

44. Lumsden RD, Locke JC, Adkins ST, Walter JF, Ridout CJ: **Isolation and localization of the antibiotic gliotoxin produced by *Gliocladium virens* from alginate prill in soil and soilless media.** *Phytopathology* 1992, **82:** 230-235

45. Cramer RA Gamcsik MP, Brooking RM, Najvar LK, Kirkpatrick WR, Patterson TF, Balibar CJ, Graybill JR, Perfect JR, Abraham SN, Steinbach WJ: **Disruption of a nonribosomal peptide synthetase in *Aspergillus fumigatus* eliminates gliotoxin production.** *Eukaryot Cell* 2008, **5:** 972-980

46. Patron NJ Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, Howlett BJ: **Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes.** *BMC Evol Biol* 2007, **7:** 174

47. Münch A, Stingl L, Jung K, Heermann R: ***Photorhabdus luminescens* genes induced upon insect infection.** *BMC Genomics* 2008, **9:** 229.

48. Isin EM, Guengerich FP: **Complex reactions catalyzed by cytochrome P450 enzymes.** *Biochim Biophys Acta* 2007, **1770:** 314-329.

49. Kulkarni RD, Thon MR, Pan H, Dean RA: **Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea.*** *Genome Biol* 2005, **6:** R24.
50. Muthumeenakshi S, Sreenivasaprasad S, Rogers CW, Challen MP, Whipps JM: **Analysis of cDNA transcripts from *Coniothyrium minitans* reveals a diverse array of genes involved in key processes during sclerotial mycoparasitism.** *Fungal Genet Biol* 2007, **44:** 1262-1284.
51. Hartley MR, Lord JM: **Cytotoxic ribosome-inactivating lectins from plants.** *Biochim Biophys Acta* 2004, **1701:** 1-14.
52. Fujimoto YK, TerBush RN, Patsalo V, Green DF: **Computational models explain the oligosaccharide specificity of cyanovirin-N.** *Protein Sci* 2008, **17:** 2008-2014.
53. Inbar J, Chet I: **The role of lectins in recognition and adhesion of the mycoparasitic fungus *Trichoderma* spp. to its host.** *Adv Exp Med Biol* 2003, **408:** 229-231.
54. Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, et al: **The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis.** *Nature* 2008, **452:** 88-92.
55. Djonović S, Pozo MJ, Dangott LJ, Howell CR, Kenerley CM: **Sm1, a proteinaceous elicitor secreted by the biocontrol fungus *Trichoderma virens* induces plant defense responses and systemic resistance.** *Mol Plant Microbe Interact* 2006, **19:** 838-53.
56. Papapostolou I, Georgiou CD: **Superoxide radical induces sclerotial differentiation in filamentous phytopathogenic fungi: a superoxide dismutase mimetics study.** *Microbiology* 2010, **156:** 960-966.
57. Aliferis KA, Jabaji S. **Metabolite composition and bioactivity of *Rhizoctonia solani* sclerotial exudates.** *J Agric Food Chem* 2010, **58:** 7604-7615.
58. Ruocco M, Lanzuise S, Vinale F, Marra R, Turrà D, Woo SL, Lorito M: **Identification of a new biocontrol gene in *Trichoderma atroviride*: the role of an ABC transporter membrane pump in the interaction with different plant-pathogenic fungi.** *Mol Plant Microbe Interact* 2009, **22:** 291-301.
59. Piotrowski M, Volmer JJ: **Cyanide metabolism in higher plants: cyanoalanine hydratase is a NIT4 homolog.** *Plant Mol Biol* 2006, **61:** 111-122.
60. Salamov AA, Solovyev VV: ***Ab initio* gene finding in *Drosophila* genomic DNA.** *Genome Res* 2000, **10:** 516-522.
61. Birney E, Durbin R: **Using GeneWise in the *Drosophila* annotation experiment.** *Genome Res* 2000, **10:** 547-548.
62. Zdobnov EM, Apweiler R: **InterProScan--an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17:** 847-848.
63. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. **Gene ontology: tool for the unification of biology The Gene Ontology Consortium.** *Nat Genet* 2000, **25:** 25-29
64. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS et al: **A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes.** *Genome Biology* 2004, **5:** R7.
65. Alcaraz LD, Olmedo G, Bonilla G, Cerritos R, Hernández G, Cruz A, Ramírez E, Putonti C, Jiménez B, Martínez E et al: **The genome of *Bacillus coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment**. *Proc Natl Acad Sci U S A* 2008, **105:** 5803-5808

65. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucleic Acids Res* 2004, **32:** D277-280.

66. Castillo-Ramirez S, Gonzalez V. **Factors affecting the concordance between orthologous gene trees and species tree in bacteria.** *BMC Evolutionary Biology* 2008, **8:** 300.

68. Moreno-Hagelsieb C, Janga SC. **Operons and the effect of genome redundancy in deciphering functional relationships using phylogenetic profiles.** *Proteins* 2008, **70:** 344-352.

69. Li L, Stoeckert CJ Jr, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13:**2178-89.

70. Smit AFA, Hubley R, Green P: **RepeatMasker Open-3.0.1996.** 2010. http://www.repeatmasker.org

71. Cantarel BL Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics.** *Nucl Acids Res* 2009, **37** (Database issue): D233 (2009).

72. MacDonald RJ, Swift GH, Przybyla AE, Chirgwin JM: Isolation of RNA using guanidinium salts. *Methods Enzymol*. 1987, **152:**219-27.

73. Jourdren L, Duclos A, Brion C, Portnoy T, Mathis H, Margeot A, Le Crom S: **Teolenn: an efficient and customisable workflow to design high-quality probes for microarray experiments.** *Nucl Acids Research* 2010, 38: e117

74. Lemoine S ,Combes F, Servant N, Le Crom S: **Goulphar: rapid access and expertise for standard two-color microarray normalization methods**. *BMC Bioinformatics* 2006, **7:** 467.

75. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J: **TM4 microarray software suite.** *Meth Enzymol* 2006, **411:** 134-193.

76. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Statistical Applications in Genetics and Molecular Biology*, 2004, **3:** No 1, Article 3.

77. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Güldener U, Mannhaupt G, Münsterkötter M: **The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes.** *Nucleic Acids Res* 2004, **32:** 5539-5545.

Acknowledgements

Genome sequencing and analysis was conducted by the U.S. Department of Energy Joint Genome Institute and supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. M.G.C.-B., E.Y.G.-R., M. H-O, and E.E.U.-R are indebted to Conacyt for doctoral fellowships. S.L.C and F.C was supported by the Infrastrutures en Biologie Santé et Agronomie (IBISA). E.M. and R.H. work was supported by the grants Junta de Castilla y León GR67, MICINN AGL2008-0512/AGR and AGL2009-13431-C02. The work of I.S.D., L.A., B.S. , B.M., S.Z., M.S. and C.P.K. was supported by the Austrian Science Foundation (grants FWF P17895-B06, P18109-B12, P-19421, V139B20 and P-19340). P.M.C and B.H. work was supported by project No. AANR-07-BIOE-006 from the French national program PNRB. M.F. was the recipient of a postdoctoral contract Ramón y Cajal from the Spanish Ministry of Science and Innovation (MCINN: RYC-2004-003005). S.Z. acknowledges support from the Vienna Science and Technology Fund (WWTF LS09-036).

Authors contributions

C.P.K., I.V.G., B.H., E.M., S.E.B., C.M.K. and A.H.E. contributed equally to this work as senior authors. A.A., J.C., M.M., A.S., I.V.G. performed global annotation and analysis, M.Z. and H.S. did the assembly, O.C. and C.H. finished the assembly and E.L. and S.L. performed the genome and EST sequencing. S.E.B., A.H.-E., C.M.K. and C.P.K. designed the study, and coordinated and supervised the analysis; C.P.K. drafted and submitted the paper. All other authors contributed research (annotations and/or analyses).

Legends to Figures

Figure 1

Distribution of orthologues of *T. atroviride*, *T. virens* and *T. reesei*. The Venn diagram shows the distribution found for the three species of *Trichoderma sp*.

Figure 2

Schematic presentation of the different stages of the mycoparasitic attack and their phenotypes. Green circles in the back show the phenotype of plates when *Tv* (right) is confronted with *R. solani* (left). (A) represents the stage before contact, and shows the colony fronts between the two fungi, illustrating the chemotropical growth of *Tv* against the prey; (B) shows the phase of contact, eventual coiling, and invasion (EM, SEM); and (C) illustrates penetration and overgrowth.

Figure 3

(A) Position of the *Ta*, *Tv* and *Tr* within the genus *Hypocrea/Trichoderma*. The phylogram was constructed by Bayesian analysis of partial exon nucleotide sequences (824 total characers; 332 characters are parsimony-informative) of the *rpb2* gene (encoding RNA polymerase B II) from 110 *ex*-type strains, thereby spanning the biodiversity of the whole genus. The tree was obtained after 5 million MCMC generations sampled for every 100 trees, using burnin=1200 and applying the general time reversible model of nucleotide substitution. Nodes with significant posterior probabilities (>0.95) are indicated by black cycles. Some *Trichoderma* species are given by names, but most other taxa are given by their NCBI ENTREZ accession numbers listed below: 1 - HQ260620; 3 - DQ08724; 4 - **HM182969**; 5 -

HM182984; 6 - HM182965; 7 - AF545565; 8 - AF545517; 16 - FJ442769; 17 - AY391900; 18 - FJ179608; 19 - FJ442715; 20 - FJ442771; 21 - AY391945; 22 - EU498358; 23 - DQ834463; 24 - FJ442725; 25 - AF545508; 26 - AY391919; 27 - AF545557; 28 - AF545542; 29 - **FJ442738**; 30 - AF545550; 31 - AY391909; 32 - AF545516; 33 - AF545518; 34 - AF545512; 35 - AF545510; 36 - AF545514; 37 - AY391921; 38 - AF545513; 39 - AY391954; 40 - AY391944; 41 - AF545534; 42 - AY391899; 43 - AY391907; 44 - AF545511; 45 - AY391929; 46 - AF545540; 47 - AY391958; 48 - AY391924; 49 - AF545515; 50 - AY391957; 51 - AF545551; 52 - AF545522; 53 - FJ442714; 54 - AF545509; 55 - AY391959; 56 - DQ087239; 57 - AF545553; 58 - AF545545; 59 - DQ835518; 60 - DQ835521; 61 - DQ835462; 62 - DQ835465; 63 - DQ835522; 64 - AF545560; 65 - DQ835517; 66 - DQ345348; 67 - AF545520; 68 - DQ835455; 69 - AF545562; 70 - AF545563; 71 - DQ835453; 72 - FJ179617; 73 - DQ859031; 74 - EU341809; 75 - FJ179614; 76 - DQ087238; 77 - AF545564; 78 - FJ179601; 79 - FJ179606; 80 - FJ179612; 81 - FJ179616; 82 - EU264004; 83 - FJ150783; 84 - FJ150767; 85 - FJ150786; 86 - EU883559; 87 - FJ150785; 88 - EU248602; 89 - EU241505; 90 - FJ442762; 91 - FJ442741; 92 - FJ442783; 93 - EU341805; 94 - FJ442723; 95 - FJ442772; 96 - EU2415023; **97 - EU341801**; 98 - EU248600; 99 - EU341808; 100 - EU3418033; 101 - EU2485942; 102 - AF545519; 103 - EU248603; 104 - EU248607; 105 - EU341806; 106 - DQ086150; 107 - DQ834460; 108 - EU711362; 109 - EU883557; 110 - FJ150790. (B) Bayesian phylogram based on the analysis of amino acid sequences of 100 orthologous syntenic proteins (MCMC, 1 mil. generations, 10449 characters) in *Tr, Tv, Ta, Gibberella zeae* and *Chaetomium globosum*. 100% posterior probabilities and significant bootstrap coefficients are marked with a circle above the nodes. The numbers indicate the genome sizes and gene counts and % net gain regarding *Ta*. Photoplates to the right show the mycoparasitic reaction after the contact between *Trichoderma* species and *R. solani*. *Trichoderma* spp. are always on the left side, dashed line

indicates the advantage of *Trichoderma* overgrowth of *R. solani*. (C) Total number of up-regulated genes belonging to 12 most significantly regulated gene families for all three *Trichoderma* species at the same stages of mycoparasitism on *R. solani* (BC, before contact; CO, contact; AC, after contact). Digits near the bars indicate the number of total and unknown up-regulated genes detected at each stage respectively.

Figure 4.

Distribution of upregulated genes (> 1.5 $\log_2$-fold) in the three Trichoderma spp. and under the three conditions analysed.Light grey: precontact; dark grey: contact; empty: overgrowth. Numbers specify the numbers of transcripts found under any of this conditions or combinations thereof.

Figure 5

Gene distribution enrichment on individual scaffolds (number of expressed genes over the total number of genes =1). Blue bars indicate the phase before contact, orange bars the phase of contact, and green bars indicate the phase of overgrowth, respectively. Multiple clusterson the same scaffold are shown in different grades of color. The bars for *T. virens*/overgrowth are 1/10th of their actual size to fit into the graph. Scaffolds are only given up to the number beyond which no more clusters were observed.

Table 1. **Genome assembly and annotation statistics**

	T.atroviride	***T.virens***	***T. reesei***
Genome Size [Mbp]	36.1	38.8	34.1
Coverage	8.26x	8.05x	9.00x
Assembly gaps, Mbp	0.1 (0.16%)	0.2(0.4%)	0.05 (0.1%)
Number of scaffolds	50	135	89
Number of predicted genes	11865	12518	9143
Gene Length (bp)	1747.06	1710.05	1793,25
Protein length (aa)	471.54	478.69	492,27
Exons per gene	"2,93	"2,98	3,06
Exon Length (bp)	528.17	506.13	507,81
Intron Length (bp)	104.20	104.95	119,64
Supported by homology (NR)	10,219 (92%)	10,915 (94%)	8409 (92%)
Supported by homology (Swissprot)	8,367(75%)	8,773 (75%)	6763 (74%)
has Pfam domain	5,883 (53%)	6,267 (54%)	5096 (56%)

Table 2. Occurence of orthologues, paralogues and singletons in the genomes of the three *Trichoderma* spp.

Genome	Synteny	total Genes S/NS	paralogs	singletons	p-value*	p-value**
T. atroviride	syntenic	9012	700	390	**2.2e-16**	.28
	NS	**1012**	**365**	48		
T. virens	syntenic	8356	854	283	**2.2e-16**	.06
	NS	**686**	**296**	31		
T. reesei	syntenic	9053	346	211	**4.4e-12**	.635
	NS	**92**	**22**	2		

S, synteny; NS, non synteny.

* p-value: null hypothesis that the proportion of paralogs that are syntenic are smaller than the proportion of paralogs that are not syntenic

** p-value: null hypothesis that the proportion of singletons that are syntenic is smaller than the proportion of singletons that not syntenic

Table 3. The major classes of Transposable Elements found in the *Trichoderma* genomes.

	T. atroviride		***T. reesei***		***T. virens***	
Class	**C.No.**	**Total Length (bp)**	**C. No.**	**Total Length (bp)**	**C.No.**	**Total Length (bp)**
DNA	372	39899	446	50448	370	52358
LTR	533	64534	559	76482	541	67484
RC	40	9235	45	9962	34	8547
LINE	561	65202	530	54928	349	59414
TOTAL**		178,870 (0.49%)		191,820 (0.57%)		187,803 (0.48%)

*C.No., copy number

*Total in base pairs and percentage of genome of TEs found in the genomes.

Table 4. Major paralogous gene expansions in *Trichoderma

PFAM domain	*T. reesei*	*T. virens*	*T. atroviride*	other fungi*
unknown protein with ankyrin (PF00023), CCHC zinc finger (PF00098; C-X2-C-X4-H-X4-C)	19	38	45	
and purine nucleoside phosphorylase domain (01048)				4
Zn(II)Cys6 transcription factor (00172) cluster 1-5	20	43	42	5,1
peptidase S8 subtilisin cluster 1-4	10	33	36	9,6
unknownprotein with WD40, NACHT and HET domain	13	38	35	3,4
short chain alcohol dehydrogenase (PF00106) cluster 1 and 2	20	32	34	4,7
unknown protein family 1-4	12	25	28	5
NAD-dependent epimerase (PFAM 01370)	10	21	23	5,8
isoflavon reductase, plus PAPA-1 (INO80 complex subunit B), epimerase and Nmr1	9	18	19	
domain				6
ankyrin domain protein	10	17	19	8
sugar transporters	11	24	18	10,8
GH18 chitinases	6	11	16	2
Protein kinase (00069) plus TPR domain	2	24	15	4,7
unknown major facilitator subfamily (PF07690) domain	9	15	15	5,5
F-box domain protein	7	10	11	1,7
ankyrin domain protein with protein kinase domain	6	8	11	2,7
Amidase	4	11	11	2,8
epoxide hydrolase (PF06441) plus AB hydrolase_1 (PF00561)	5	14	11	3,2
FAD_binding_4, plus HET and berberine bridge enzymes (08031) domain	5	13	11	6,1
FMN oxidoreductases	2	8	10	2,5
unknown protein with DUF84 (NTPase) and NmrA domain	5	19	10	3,7
protein with GST_N and GST_C domains	6	12	10	4,6
class II hydrophobins class	6	8	9	1,1
proteins with LysM binding domains	6	7	9	1,2
unknown protein family with NmrA domain	2	11	8	0,2
Pro_CA	5	9	8	1,3
WD40 domain protein	5	11	8	2,2
C2H2 transcription factors	1	5	7	1,4
GFO_IDH_MocA (01408 and 02894) oxidoreductase	3	9	7	1,5
Protein kinase (00069)	4	6	6	0,7

nonribosomal peptide synthase	3	4	5	1
SSCP ceratoplatanin-family	3	4	5	1
GH75 chitosanase	3	5	5	1,1
SNF2, DEAD box helicase	3	5	5	1,3
Nitrilase	3	6	5	2,2
GH65 trehalose or maltose phosphorylase (PFAM 03632)	4	4	4	0,8
AAA-family ATPase (PF00004)	4	3	4	1
pyridoxal phosphate dependent decarboxylase (00282)	2	3	4	1,2
unknown protein	3	4	4	1,3

* results from MCL analysis of three Trichoderma spp. (Tr, Ta, Tv) and mean values from nine other ascomycetes whose genomes are present in the JGI database (http://genome.jgi-psf.org) Data selected from a total of 28919 clusters, average cluster number 5.8 [S.D. 15.73].

Table 5. GH families involved in chitin/chitosan and β-1,3 glucan hydrolysis that are strongly expanded in mycoparasitic *Trichoderma* species.

	Chitin/chitosan*		ß-glucan*			total ß-glucan	
GH family	**18**	**75**	**17**	**55**	**64**	**81**	**217**
Trichoderma atroviride	**29**	**5**	**5**	**8**	**3**	**2**	**18**
Trichoderma virens	**36**	**5**	**4**	**10**	**3**	**1**	**18**
Trichoderma reesei	**20**	**3**	**4**	**6**	**3**	**2**	**15**
Aspergillus nidulans	19	2	5	6	0	1	**12**
Aspergillus niger	14	2	5	3	0	1	**9**
Fusarium graminearum	19	1	6	3	2	1	**12**
Laccaria bicolor	10	0	4	2	0	0	**6**
Magnaporthe grisea	14	1	7	3	1	2	**13**
Nectria haematococca	28	2	6	5	2	1	**14**
Neurospora crassa	12	1	4	6	2	1	**13**
Penicillium chrysogenum	9	1	5	3	2	1	**11**
Phanerochaete chrysosporium	11	0	2	2	0	0	**4**
Podospora anserina	20	1	4	7	1	1	**13**
Postia placenta	20	0	4	6	0	0	**10**
Saccharomyces cerevisiae	2	0	4	0	0	2	**6**
Schizosaccharomyces pombe	1	0	1	0	0	1	**2**
Tuber melanosporum	5	1	4	2	0	3	**9**

* main substrates for the respective enzymes

Table 6. The number of polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) of *Trichoderma* compared to other fungi

Fungal species	**PKS**	**NRPS**	**PKS-NRPS NRPS-PKS**	**Total**
Trichoderma virens	**18**	**28**	**4**	**50**
Aspergillus oryzae	26	14	4	44
Aspergillus nidulans	26	13	1	40
Cochliobolus heterostrophus	23	11	2	36
Trichoderma atroviride	**18**	**16**	**1**	**35**
Magnaporthe oryzae	20	6	8	34
Fusarium graminearum	14	19	1	34
Gibberella moniliformis	12	16	3	31
Botryotinia fuckeliana	17	10	2	29
Aspergillus fumigatus	13	13	1	27
Nectria haematococca	12	12	1	25
Trichoderma reesei	**11**	**10**	**2**	**23**
Neurospora crassa	7	3	0	10

Table 7. Number of Pfam domains that are enriched in the non syntenic paralogous genes

	T. reesei	*T. virens*	*T. atroviride*
Zn2Cys6 transcription factors	9	**95**	69
WD40 domains	1	11	14
sugar transporters	**0**	18	13
proteases	2	28	23
cytochrome P450	7	**40**	15
NmrA-domains	2	19	21
major facilitator superfamily	7	52	60
HET domains	3	26	27
glycoside hydrolases	3	33	26
FAD-binding proteins	2	28	24
ankyrins	4	44	37
alcohol dehydrogenases	**4**	51	71
α/ß-fold hydrolases	2	26	15
ABC transporters	4	14	**3**
number of genes	50	485	418
total gene number in NS areas	92	686	1012

Boxed numbers are those that are significantly ($p < 0.05$) different from the two other species when related to the genome size.