Comparative genomics to explore phylogenetic relationship, cryptic sexual potential and host specificity of Rhynchosporium species on grasses. BMC Genomics 17 (1), 953. 10.1186/s12864-016-3299-5

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Comparative genomics to explore phylogenetic relationship, cryptic sexual potential and host specificity of Rhynchosporium species on grasses

Daniel Penselin¹, Martin Münsterkötter², Susanne Kirsten¹, Marius Felder³, Stefan Taudien³, Matthias Platzer³, Kevin Ashelford⁴, Konrad H. Paskiewicz², Richard J. Harrison⁵, David J. Hughes⁷, Thomas Wolf⁸, Ekaterina Shelest⁸, Jenny Graap¹, Jan Hoffmann¹, Claudia Wenzel¹, Nadine Wöltje¹, Kevin M. King⁹, Bruce D. L. Fitt¹⁰, Ulrich Güldener¹¹, Anna Avrova¹² and Wolfgang Knogge¹*

Abstract

Background: The Rhynchosporium species complex consists of hemibiotrophic fungal pathogens specialized to different sweet grass species including the cereal crops barley and rye. A sexual stage has not been described, but several lines of evidence suggest the occurrence of sexual reproduction. Therefore, a comparative genomics approach was carried out to disclose the evolutionary relationship of the species and to identify genes demonstrating the potential for a sexual cycle. Furthermore, due to the evolutionary very young age of the five species currently known, this genus appears to be well-suited to address the question at the molecular level of how pathogenic fungi adapt to their hosts.

Results: The genomes of the different Rhynchosporium species were sequenced, assembled and annotated using ab initio gene predictors trained on several fungal genomes as well as on Rhynchosporium expressed sequence tags. Structures of the rDNA regions and genome-wide single nucleotide polymorphisms provided a hypothesis for intra-genus evolution. Homology screening detected core meiotic genes along with most genes crucial for sexual recombination in ascomycete fungi. In addition, a large number of cell wall-degrading enzymes that is characteristic for hemibiotrophic and necrotrophic fungi infecting monocotyledonous hosts were found. Furthermore, the Rhynchosporium genomes carry a repertoire of genes coding for polyketide synthases and non-ribosomal peptide synthetases. Several of these genes are missing from the genome of the closest sequenced relative, the poplar pathogen Marssonina brunnea, and are possibly involved in adaptation to the grass hosts. Most importantly, six species-specific genes coding for protein effectors were identified in R. commune. Their deletion yielded mutants that grew more vigorously in planta than the wild type.

(Continued on next page)
**Conclusion:** Both cryptic sexuality and secondary metabolites may have contributed to host adaptation. Most importantly, however, the growth-retarding activity of the species-specific effectors suggests that host adaptation of *R. commune* aims at extending the biotrophic stage at the expense of the necrotrophic stage of pathogenesis. Like other apoplastic fungi *Rhynchosporium* colonizes the intercellular matrix of host leaves relatively slowly without causing symptoms, reminiscent of the development of endophytic fungi. *Rhynchosporium* may therefore become an object for studying the mutualism-parasitism transition.

**Keywords:** CAZymes, Effectors, Host specificity, Leotiomycetes, Non-ribosomal peptide synthetases, Phylogenetic evolution, Polyketide synthases, *Rhynchosporium*, Sex-related genes, Whole genome sequencing

---

**Background**

The genus *Rhynchosporium* (order: Leotiales) consists of haploid fungi that are pathogenic to grass species (*Poaceae*) including cereal crops. In an early review article [1], only two species were described, *R. secalis* and *R. orthosporum*, differing both in the shape of their two-celled conidia and in their host specificities. The name of the fungal genus originates from the beak (Greek: *rhynchos*) carried by one of the spore cells of *R. secalis*. In contrast, *R. orthosporum* produces cylindrical spores (Greek: *ortho*, straight). *R. secalis* was first isolated from rye (*Secale cereale*; [2]; cited in [1, 3]). Nevertheless, the major host worldwide is barley (*Hordeum vulgare*), on which the fungus causes the economically important disease leaf blotch or scald. *R. orthosporum* is a pathogen of orchard grass (cocksfoot, *Dactylis glomerata*).

It is usually assumed that plant pathogens originated on direct ancestors of their modern crop hosts and coevolved with them during domestication [4]. Therefore the center of origin of a pathogen is usually also the center of genetic diversity [5]. Gene diversity analysis of *Rhynchosporium* suggests, however, that the fungus did not originate from the Fertile Crescent [5], where its host plants were domesticated about 10,000 years ago. From there the fungus spread globally most likely due to anthropogenic dispersal of infected host material [9–11], reaching regions such as South Africa, California, Australia and New Zealand only during the last few hundred years [10].

The occurrence of host-specialized ‘races’ of *R. secalis* was reported early on [1]. Nevertheless, for a long time there were conflicting reports about the host range of *Rhynchosporium* isolates [12]. Recently, multigene-sequencing and phylogenetic analysis (genealogical concordance phylogenetic species recognition; [13, 14]) demonstrated that *R. secalis* was actually a complex of at least three closely related host-specialized species [15]. Given that the fungus was initially described on rye, the name *R. secalis* was retained for the species infecting rye and triticale (*x Triticosecale* Wittmack). The species colonizing cultivated barley and other *Hordeum* spp. as well as brome grass (*Bromus diandrus*) was named *R. commune*, whereas *R. agropyri* is pathogenic to couch grass (*Elymus repens*; syn. *Agropyron repens*). Finally, fungal isolates from perennial ryegrass (*Lolium perenne*) were found to represent yet another species, *R. lolii* [16]. The cylindrical shape of the spores as well as phylogenetic analysis based on nucleic acid sequences position the latter species closer to *R. orthosporum* (cylindrical conidia group, CCG) than to the other three species (beaked conidia group, BCG). Due to its economic importance the majority of studies has been carried out on the barley-infecting fungal species, *R. commune* (in the older literature called *R. secalis*).

Phylogenetic concordance analyses based on the sequences of four housekeeping genes (ITS1-5.8S-rRNA-ITS2, EF1-α, α-tubulin, β-tubulin) and neutral RFLP loci using several hundred fungal isolates suggested that the CCG and the BCG branches separated 14,000–35,000 years ago, hence long before the advent of agriculture, with *R. lolii* diverging from *R. orthosporum* 4300–7200 years ago [16]. In contrast, *R. commune*, *R. secalis* and *R. agropyri* originated from a common ancestor only 1200–3600 years ago [8, 9, 16]. Furthermore, reconstructions of demographic growth indicate that populations of the three BCG species evolved in parallel. After the host shift, which was soon followed by the split of the three fungal lineages, they experienced a bottleneck due to their reproductive isolation. Specific adaptation to their modern hosts [15] led to the recovery of the sizes of all 3 populations in particular during the last 250 years [10, 17]. Thus, worldwide dissemination and population increase of the pathogen coincide with intensification and global expansion of cereal production [17, 18].

In evolutionary terms, the *Rhynchosporium* BCG species are very young and closely related. Thus it can be expected that their genomes did not diverge substantially yet, except for those genes that are involved in...
adaption to the different hosts and/or environments. Hence, they represent an excellent basis to address the question of host specialization by comparative genomics. Genes under diversifying selection may be uncovered during a fully unbiased approach by scanning the genomes of several isolates from the different species. Alternatively, crucial processes associated with host adaption of a pathogen and epitomized in the concept of (quantitative) virulence may be defined a priori. This would allow the identification of patterns of adaptive evolution such as species-specifically occurring genes or of gene family expansion.

What are the factors that contribute to fungal virulence? Rhynchosporium development in planta starts with germination of conidia on the surface of host leaves and the penetration of fungal hyphae through the cuticle, frequently without forming obvious appressoria [19, 20]. Then, a long latent period of a few days up to several months [21, 22] without any visible disease symptoms follows. During this stage fungal mycelia are established between the cuticle and the outer layer of epidermis cells. Finally, after formation of a dense subcuticular fungal stroma necrotic lesions occur. New spores are formed both during the latent period and in the lesion areas. Penetration as well as subcuticular proliferation in the host tissue require the secretion of hydrolytic enzymes to degrade the cuticle and cell wall structures and to release carbohydrate nutrients form the plant cell wall polysaccharides. In addition, the physiology of the host needs to be redirected towards tolerating and accommodating the pathogen. For this purpose (hemibiotrophic) fungal pathogens secrete small molecules (secondary metabolites; [23]) as well as effector proteins. Therefore, the genes coding for cell wall-degrading enzymes, key secondary biosynthetic enzymes or protein effectors are prime targets of diversification/diversifying selection during adaptive evolution.

Finally, at the base of any considerations regarding evolutionary processes is the mating system of an organism. For Rhynchosporium, no teleomorph has been identified to date despite the presence of mating-type (MAT) genes [24]. Nevertheless, ‘the signature of sex’ [25], i.e., mating type equilibrium in most populations at least of the BCG species and evidence for MAT gene expression [5, 26, 27], high genotype diversity, gametic equilibrium among neutral markers and low levels of clonality, was found in fungal populations worldwide [5, 9, 18, 28, 29]. These results suggest that sexual recombination occurs cryptically but regularly.

This paper presents a comparative analysis of the genomes of isolates from four of the five Rhynchosporium species described to date. The intra-genus evolution and the broader phylogenetic relationship to other fungal species are re-analyzed. Furthermore, the possibility of sexual recombination is assessed using a genomics approach by screening for the presence of sex-related genes known to be involved in meiosis and recombination in other organisms. Finally, genes that may be involved in controlling host specificity, including those coding for key enzymes of fungal secondary metabolism or effector proteins are identified.

Results and discussion

Genome statistics

Three isolates of R. commune and one isolate each of R. secalis, R. agropyri, R. lolii and R. orthosporum were sequenced using a whole-genome shotgun strategy [30] with the Roche 454 GS FLX and the Illumina GAIIx platforms. De novo assembly resulted in a total sequence length of 52–58 Mb. Overall sequencing and assembly statistics are presented in Table 1. The assembled genome of R. commune, isolate UK7, consisting of only 163 scaffolds was established as the reference genome. Gene models for R. commune, isolate UK7, R. secalis and R. agropyri were generated by three de novo prediction programs with different matrices. Completeness of genomic predictions was assessed by BLASTp comparison to a set of 246 highly conserved fungal single-copy protein-coding homologs (SCPCH, [31]) and to a set of 248 core genes from the core Eukaryotic genes mapping approach (CEGMA, [32]; Table 2). The predicted coding sequences comprise about 30–33% of the genomic sequences, the estimated gene numbers range from 10,254 to 13,674 with about 50% coding for proteins of unknown function. Repetitive DNA elements were identified and categorized. The total coverage of repetitive sequences is relatively high as compared to other fungi reaching 31% in R. commune, but lower than in the close relative Marssonina brunnea fsp. ‘multigermtubi’ (see below; 42% [33]).

Phylogenetic relationship

Phylogenetic analysis using multilocus DNA sequence data (concatenated α-β-tubulin, β-tubulin and ITS sequences) identified R. secalis as the closest sister species to R. commune, whereas the CCG species R. orthosporum is more distantly related [15, 17]. This relationship was supported in the present study by comparing the Rhynchosporium rDNA gene structures (Fig. 1). The genes of R. commune and R. secalis are highly similar lacking any introns, whereas an intron is present in the 28S gene of R. agropyri. The CCG species R. orthosporum and the recently described R. lolii [16] carry two introns of the same size in the 18S gene and single introns differing in length by one nucleotide in the 28S gene. The latter are homologous to the 28S intron of R. agropyri and located at the same site, but their sequences are 99 bp and 87 bp shorter at the 5’ and the 3’
| Parameter          | R. commune (UK7) | R. (AU2) | R. commune (13-13) | R. secalis | R. agropyri | R. orthosporum |
|--------------------|------------------|----------|--------------------|------------|-------------|---------------|
| Primary seq. data  | Illumina PE (GAIIx) | 454      | Illumina PE (GAIIx) | 454        | Illumina PE (GAIIx) | 454          |
| No. of reads       | 18 Mio.          | 35 Mio.  | 6.5 Mio.           | 2.37 Mio.  | 32.1 Mio.   | 34 Mio.       |
| Read length        | 2 x 36 bp        | 2 x 151 bp, 270 bp dist. | 2 x 151 bp, 3 kb dist. | 2 x 101 bp, 76 bp dist. | 2 x 101 bp, 300 bp dist. | 2 x 101 bp, 270 bp dist.  |
| Total length       | 1.29 Gb          | 9.7 Gb   | 1.4 Gb             | 894 Mb     | 1.92 Gb     | 2.13 Gb       |
| Nominal seq. depth | 26x              | 194x     | 29x                | 18x        | 130x        | 156x          |
| Assembly software  | ALLPATHS-LG + SSPACE | Velvet | Velvet + Newbler + Minimus2 | Newbler | ALLPATHS-LG + SSPACE | Velvet |
| Total seq. length (LFI) | 55.59 Mb  | 54.32 Mb | 53.93 Mb          | 56.80 Mb   | 52.18 Mb    | 58.16 Mb      |

* 0.5 run 20 kb PE library, 1.5 runs 0.8 kb library
| Assembly statistics | R. commune UK7 | R. secalis | R. agropyri | R. ortho-sporum | M. brunnea |
|---------------------|---------------|------------|-------------|----------------|-----------|
| Total scaffold length (Mb) | 55.59 | 56.80 | 52.18 | 58.16 | 51.95 |
| N50 contig (kb) | 49.6 | 47.5 | 61.6 | | |
| N50 scaffold (kb) | 800.5 | 113.7 | 303.4 | 104.3 | |
| Scaffolds/Chromosomes | 163 | 2129 | 508 | 29183 | 89 |
| GC-content (%) | 42.33 | 43.91 | 42.9 | 41.1 | 42.72 |
| - coding (%) | 49.86 | 49.85 | 49.75 | 49.75 | 44.12 |
| - non-coding (%) | 39.09 | 40.87 | 39.52 | 38.12 | |
| Coding sequence | | | | | |
| Percent coding (%) | 30.06 | 33.84 | 33.08 | | 28.78 |
| Average gene size (bp) | 13677 | 1271.6 | 1269.4 | | 1491 |
| Average gene density (gene/kb) | 0.22 | 0.27 | 0.26 | | 0.19 |
| Protein-coding genes | 12212 | 13151 | 13674 | | 10027 |
| Exons | 38363 | 39806 | 41720 | | 33423 |
| Average exon size | 434.4 | 419.1 | 415.1 | | 394.6 |
| Exons/gene | 3.1 | 3 | 3.1 | | 3.9 |
| tRNA genes | 113 (29 mt*<sup>a</sup>) | 117 (30 mt*<sup>a</sup>) | 107 (29 mt*) | | nd |
| Validation of prediction | | | | | |
| SCPCH (246) | 246 | 246 | 246 | 246 | |
| CEGMAS (248) | 248 | 248 | 247 | 247 | |
| Secretome | | | | | |
| Secreted proteins ≥300aa | 711 | 690 | 724 | | 509 |
| Small secreted proteins <300aa | 457 | 574 | 655 | | 429 |
| Non-classical secreted proteins ≥300aa | 77 | 90 | 89 | | 61 |
| Small non-classical secr. proteins <300aa | 63 | 167 | 196 | | 63 |
| non-secreted protein | 10904 | 11630 | 12010 | | 8965 |
| All proteins | 12212 | 13151 | 13674 | | 10027 |
| Non-coding sequence | | | | | |
| Introns | 26151 | 26655 | 28046 | | 24747 |
| Introns/gene | 26 | 2.5 | 2.5 | | 2.9 |
| Average intron length (base) | 80.5 | 82.4 | 83.1 | | 118.5 |
| Average intergenic distance (bp) | 3013 | 2319 | 2398 | | 3398 |
| Repeat sequences | | | | | |
| DNA Transposon | 8.04% | 4.27% | 6.80% | | 7.64% |
| LINE | 0.39% | 0.97% | 0.44% | | 1.56% |
| SINE | 0.07% | 0.11% | 0.07% | | 0.15% |
| LTR Retrotransposon | 17.06% | 7.84% | 10.56% | | 27.41% |
| Unclassified nonLTR-Retrotransposon | 0.02% | 0.03% | 0.00% | | 0.07% |
| Unclassified Retrotransposon | 2.56% | 3.35% | 2.69% | | 2.54% |
| Unclassified | 0.22% | 0.60% | 0.42% | | 1.36% |
| Total TE class | 28.21% | 16.68% | 20.88% | | 40.33% |
| Simple sequence repeats | 0.90% | 1.03% | 0.94% | | 2.57% |
| Total excl. Tandem repeats | 29.16% | 17.79% | 21.89% | | 43.07% |
| Tandem repeats | 8.27% | 5.05% | 7.38% | | 11.46% |
| Total repeat coverage | 31.23% | 20.09% | 24.00% | | 45.64% |

*mitochondrial
side, respectively. In contrast to these results, phylogenetic analysis based on concatenated sequences of 13 core mitochondrial genes resulted in *R. agropyri* and *R. secalis* as being most closely related [30].

To better resolve the *Rhynchosporium* phylogeny a two-step analysis was carried out in the present study. The first step aimed at confirming the general position of the *Rhynchosporium* genus in the *Leotiomycetes* class of Ascomycota [34]. To this end, the sequences of three protein-coding genes (EF1-α, RNA polymerase II subunits RPB1 and RPB2) and of almost complete rDNA regions (18S rDNA, ITS1-5.8S rDNA-ITS2, 28S rDNA; [35]) from 12 members of the *Leotiomycetes* class were used in lack of complete genome sequences of these fungi other than *Botryotinia fuckeliana* [36]. The resulting phylogeny shows the *Rhynchosporium* BCG branch with *R. commune* and *R. agropyri* as the closest sister species alongside the CCG branch (Fig. 2a, Additional file 1: Figure S1). In a second step, to resolve the BCG evolution in more detail, single nucleotide polymorphisms (SNPs; Additional file 2: Table S1) were identified after aligning the genomes of BCG species and isolates (Fig. 2b). The phylogenetic tree inferred from the concatenated SNPs again confirms that *R. agropyri* branched off “shortly” before *R. commune* and *R. secalis* diverged. According to the published evolutionary time scale the former occurred 1200–3600 years ago [8, 9, 16], the latter about 500 years later. In contrast, the age of the 3 *R. commune* isolates is probably around 100 years or less.

Using the amino acid sequences of the three proteins EF1-α, RPB1 and RPB2 the relationship of *Rhynchosporium*
to more distantly related plant-interacting fungal species was established (Fig. 3). The closest relative of Rhynchosporium spp., the poplar pathogen \textit{M. brunnea} f.sp. ‘multigermtubi’ [33], belongs to the Dermataceae, a family of one of the most diverse ascomycete orders, the Helotiales. This confirms a previous ITS-based phylogenetic analysis, which located \textit{R. commune} and \textit{R. orthosporum} between the Dermataceae species \textit{Tapesia yallundae} (syn. \textit{Oculimacula yallundae}) and \textit{Pyrenopeziza brassicae}, pathogens of grasses and oil seed rape (\textit{Brassica napus}), respectively [34].

All host species of Rhynchosporium belong to the Pooidae subfamily of Poaceae [37, 38]. However, while the fungal BCG species, \textit{R. commune}, \textit{R. secalis} and \textit{R. agropyri}, infect closely related grass species from genera of the subtribe Hordeinae (supertribe Tritici; tribe Hordeae), the CCG members, \textit{R. lolii} and \textit{R. orthosporum}, are pathogenic to species from the more distantly related subtribes Loliinae and Dactylidinae (supertribe Pooodae; Fig. 4). Nevertheless there are exceptions to this rule. Firstly, from \textit{Bromus diandrus}, a grass species belonging to the Bromeae tribe of the Tritici; tribe Hordeae), a few isolates were collected that grouped into the \textit{R. commune} lineage [17]. Interestingly, in an early report isolates from brome grass were unable to infect other grass species tested including barley [1]. Secondly, triticale became susceptible to \textit{Rhynchosporium} about 30 years after its introduction as a crop [39]. This cereal species originated from hybridizing the non-host species wheat and the host species rye and was assigned to the Tritici; tribe Hordeae). \textit{Rhynchosporium} isolates collected from triticale grouped into the \textit{R. secalis} lineage [17]. Lastly, two strains of \textit{R. commune} were isolated from Italian ryegrass (\textit{Lolium multiformum}) that were also pathogenic to barley [16] and, thus, able to colonize hosts from both Tritici; tribe Hordeae) and Pooodae supertribes. Despite these exceptions, however, the relationship found between the \textit{Rhynchosporium} species corresponds largely to host systematics.

\textbf{Genes related to sexual reproduction}

Work since the 1930s has failed to demonstrate a sexual cycle in any \textit{Rhynchosporium} species. Availability of the \textit{Rhynchosporium} genomes allows now a genomic interrogation to assess whether the genetic machinery required for sexual reproduction is present or not. Such insights into the pathogen life cycle are important as this impacts on the possibility for recombination and evolution of the pathogen [40]. Despite the apparent absence of a teleomorph \textit{R. commune} isolates have been shown to possess the mating type loci \textit{MAT1-1} or \textit{MAT1-2}, which are typical for heterothallic ascomycetes [24]. \textit{MAT1-1} is characterized by genes coding for proteins with two different DNA-binding domains, an α-box motif in \textit{MAT1-1-1} and a high-mobility group (HMG) DNA-binding motif in \textit{MAT1-1-3}. Both loci were also found in the other BCG species, \textit{R. secalis} and \textit{R. agropyri}, whereas the CCG species \textit{R. orthosporum} and \textit{R. lolii} lack \textit{MAT1-2} [27]. In the present study, only the \textit{R. agropyri} isolate had the \textit{MAT1-2} genotype, whereas all others carried the \textit{MAT1-1} locus. Closer inspection of the \textit{MAT1-1} locus revealed the presence of an additional ORF coding for a protein with about 30% similarity to the \textit{MAT1-1-5} protein from \textit{B. cinerea} and several \textit{Sclerotinia} species (Fig. 5). An additional ORF with homology

\begin{figure}
\centering
\includegraphics[width=\textwidth]{integration_of_genus_rhynchosporium_into_the_fungal_systematics}
\caption{Integration of the genus \textit{Rhynchosporium} into the fungal systematics. The concatenated amino acid sequences of elongation factor EF1-a and of the polymerase II subunits RPB1 and RPB2 from 21 taxa were used to construct the phylogenetic tree. Numerals on the nodes represent the percentages from 500 bootstraps. Scale: number of substitutions per nucleotide.}
\end{figure}
to a hypothetical gene from the poplar pathogen *M. brunnea* f.sp. *multigermtubi* was detected at the MAT1-2 locus. However, this ORF codes for a protein with no homology to a functional protein from other organisms. Sequence comparison of the MAT loci demonstrated that the idiomorphic region comprises most of the MAT1-1-1 gene and the genes MAT1-1-5 and MAT1-1-3 in the MAT1-1 locus (c. 4150 bp). In MAT1-2 it spans from the truncated MAT1-1-1 gene to the 3’ end of the MAT1-2-1 gene (c. 3600 bp). Both MAT loci are flanked by the genes SLA2 and APN2 coding for a cytoskeletal protein and a DNA lyase, respectively. The positions of these genes are highly conserved across *Pezizomycotina* species [41–44].

Meiosis is “at the heart of sexual reproduction” of Eukaryotes [45]. Therefore, searching for genes involved in meiosis and sexual reproduction can identify the capacity for cryptic sexuality. To this end, the *Rhynchosporium* genomes were screened using a “meiosis detection tool kit” [46], which comprises a set of genes defining a “core meiotic recombination machinery” (Table 3, cf. Additional file 3: Table S2A) [47, 48]. The products of these genes are involved in generating meiotic double-strand breaks, crossing over and cohesion of sister chromatids and homologous chromosomes. Bidirectional BLASTp analysis detected 26 out of the 29 core meiotic genes in the *Rhynchosporium* (with the exception of *R. loli*) genomes with sequence similarities mostly exceeding 50%. For two additional proteins, HOP2 and MND1, the similarity to *Rhynchosporium* proteins is low (c. 30%), although the functional domains were identified. Many genes required for meiosis also have a role in mitosis. However, nine meiosis-specific genes are regarded to reliably predict the capacity of a sexual cycle [48]. Three of these genes that code for elements of the synaptonemal complex, HOP1, HOP2 and MND1, are missing in *Rhynchosporium* (HOP1) or their products show only low similarity with *Rhynchosporium* proteins (HOP2, MND1). Nevertheless this does not argue against sexual reproduction, because these genes are also absent from the genomes of several fungi with known sexuality (*Neurospora crassa*, *Gibberella zeae*, *Magnaporthe grisea*, *Ustilago maydis*) and of other sexual organisms such as *Drosophila melanogaster* [48]. Hence, meiosis appears to be possible without these gene products. Transcripts for all but two (*DMC1*, *MSH5*) of the 26 core meiotic genes have been identified in germinated conidia and/or cDNA libraries from infected barley.
leaves (Table 3). This confirms that the core meiotic genes are not only present in the *Rhynchosporium* genomes but are actively transcribed.

Of an additional set of 70 genes involved in different aspects of meiosis and recombination in *N. crassa*, *Saccharomyces cerevisiae* and *Podospora anserina*, 68 were found to be maintained in the genomes of *R. commune*, *R. agropyri* and *R. orthosporum* (Additional file 3: Table S2A). Two genes, KEX1 and RAM1, were missing only from the genome of *R. secalis*. Transcripts for 63 of these genes, including those coding for meiotic recombination protein REC4 and all pheromone proteins essential for mating, have been identified in germinated conidia and/or cDNA libraries from infected barley leaves (Table 3).

| No. | Gene annotation/putative protein function | Protein | *Rhynchosporium* spp. | Similarity [%] | Transcript 2 |
|-----|-------------------------------------------|---------|-----------------------|---------------|-------------|
| **Meiosis** | | | | | |
| **Double-strand DNA breaks and processing** | | | | | |
| 1 | Transsterase (REC12) | SPO11 | + | 53 | gC, 3 dpi |
| 2 | Meiotic recombination protein | DMC1 | + | 84 | - |
| 4 | DEAD/DDEAH box DNA helicase | MER3 | + | 58 | gC, 3 dpi |
| 7 | DNA repair protein (US6) | RAD50 | + | 61 | gC, 3 dpi |
| 8 | 3’-5’ dsDNA exonuclease, ssDNA endonuclease | MRE11 | + | 57 | gC, 3 dpi |
| **Single strand invasion** | | | | | |
| 9 | DNA repair protein (MEI3) | RAD51 | + | 97 | gC, 3 dpi |
| 10 | DNA repair and recombination protein (MUS11) | RAD52 | + | 58 | 3 dpi |
| **Proteins involved in crossing over** | | | | | |
| 19 | DNA mismatch repair protein MutS | MSH4 | + | 54 | gC, 3 dpi |
| 20 | DNA mismatch repair MutS family | MSH5 | + | 57 | - |
| **Synaptonemal complex** | | | | | |
| 26 | Structural maintenance of chromosome: SMC protein | SMC5 | + | 63 | 3 dpi |
| **Mismatch repair proteins** | | | | | |
| 31 | DNA mismatch repair protein | MSH2 | + | 89 | gC, 3 dpi |
| 32 | DNA mismatch repair protein | MSH6 | + | 75 | gC, 3 dpi |
| **Other** | | | | | |
| 47 | Synaptonemal complex protein | HOP1 | - | - | |
| 48 | Interhomolog meiotic recombination protein | HOP2 | (+) | 30 | (3 dpi) |
| 49 | Interhomolog meiotic recombination protein | MND1 | (+) | 28 | (gC) |
| **Cohesion** | | | | | |
| 51 | Cohesin complex subunit | SMC1 | + | 81 | gC, 3 dpi |
| 52 | Chromosome segregation protein SUDA | SMC3 | + | 75 | gC, 3 dpi |
| 53 | Cohesin complex subunit, sister chromatid cohesion | SCC3 | + | 60 | gC, 3 dpi |
| 54 | DSB repair protein (SCC1) | RAD21 | + | 72 | gC, 3 dpi |
| 55 | Cohesin complex subunit, sister chromatid cohesion | REC8 | + | 45 | gC, 3 dpi |
| **Condensins** | | | | | |
| 58 | Nuclear condensin complex subunit | SMC2 | + | 81 | gC, 3 dpi |
| 59 | Nuclear condensin complex subunit | SMC4 | + | 76 | gC, 3 dpi |
| **Expanded core meiotic recombination machinery** | | | | | |
| 93 | DNA mismatch repair protein (mul1) | MLH1 | + | 80 | gC, 3 dpi |
| 94 | DNA mismatch repair protein (mul12) | MLH2 | + | 63 | gC, 3 dpi |
| 95 | DNA mismatch repair protein | MLH3 | + | 49 | gC, 3 dpi |
| 96 | DNA mismatch repair protein | PMS1 | + | 64 | gC, 3 dpi |
| 97 | Postreplication repair E3 ubiquitin-protein ligase (RAD18) | SMG6 | + | 59 | gC, 3 dpi |
| 98 | Sister chromatid cohesion protein (Spo76) | PDS5 | + | 45 | gC, 3 dpi |
| 99 | DNA repair 3’-5’ exonuclease (MEI9) | RAD1 | + | 61 | gC, 3 dpi |

1 numbering refers to Table S3A
2 presence of transcript in germinated conidia (gC) and/or in barley leaves (3 dpi)
3 TRIP domain (pfam01706); 23% similarity in domain sequence
4 MND1 domain (pfam038962); *R. commune* and *R. secalis* proteins annotated with longer N-termini

Table 3: Core meiotic genes
leaves. Finally, 74 genes known to be involved in different aspects of sexual reproduction in the aspergilli such as sensing of environmental signals that control sexual cycle induction, intracellular signal transduction, transcription control and ascospore production were examined (Additional file 3: Table S2B; [49]). Outside the MAT genes and the genes coding for pheromones and their receptors, only 5 genes were not found in the *Rhynchosporium* genomes, whereas for 12 gene products a relatively weak similarity with the *A. nidulans* proteins (<40%, 4 with <30%) was observed. Transcripts for 64 of the 65 genes, including MAT1 and *esdC*, have been identified in germinated conidia and/or cDNA libraries from infected barley leaves. Taken together these results provide strong additional evidence for the presence and expression of the genetic machinery required for sexual reproduction to occur in the *Rhynchosporium* life cycle.

**Enzymes involved in cell wall degradation**

Plant cell walls are built up of polysaccharides (cellulose, hemicellulose, pectin), (glyco-) proteins and aromatic polymers (lignin). Phytopathogenic fungi need to degrade this barrier to penetrate and colonize host tissues as well as to release nutrients for colony growth. To this end, they produce many carbohydrate-active enzymes (CAZymes) that cleave or modify oligo- and polysaccharides as well as other glycoconjugates. A recent genome comparison comprising 103 fungal species from different classes revealed a great diversity in the cell wall-degrading capacity [50]. In phytopathogens, CAZyme numbers correlate by and large with the fungal life style. In both hemibiotrophic and necrotrophic fungi large numbers of CAZymes are found, whereas obligately biotrophic fungi such as *Blumeria graminis* have smaller numbers. Furthermore, polygalacturonases of the GH28 family have been found to be less abundant in monocot genomes, whereas for 12 gene products a relatively weak similarity with the *A. nidulans* proteins (<40%, 4 with <30%) was observed. Transcripts for 64 of the 65 genes, including MAT1 and *esdC*, have been identified in germinated conidia and/or cDNA libraries from infected barley leaves. Taken together these results provide strong additional evidence for the presence and expression of the genetic machinery required for sexual reproduction to occur in the *Rhynchosporium* life cycle.

| CAZy family | R. commune | R. secalis | R. agropyri |
|-------------|------------|------------|-------------|
| GH<sup>a</sup> | 182 | 53.8 | 176 | 53.7 | 183 | 54.3 |
| CE<sup>b</sup> | 66 | 19.5 | 62 | 18.9 | 64 | 19.0 |
| PL<sup>c</sup> | 9 | 2.7 | 9 | 2.7 | 9 | 2.7 |
| CBM<sup>d</sup> only | 20 | 5.9 | 20 | 6.1 | 20 | 5.9 |
| AA<sup>e</sup> | 61 | 18.0 | 61 | 18.6 | 61 | 18.1 |
| Σ | 338 | 99.9 | 328 | 100.0 | 337 | 100.0 |
| CWDE substrate | R. commune | R. secalis | R. agropyri |
| Cutin | 14 | 4.1 | 14 | 4.3 | 14 | 4.2 |
| Cellulose | 64 | 18.9 | 64 | 19.5 | 64 | 19.0 |
| Hemicellulose | 107 | 31.7 | 101 | 30.8 | 108 | 32.0 |
| Lignin | 18 | 5.3 | 18 | 5.5 | 18 | 5.3 |
| Pectin | 27 | 8.0 | 27 | 8.2 | 27 | 8.0 |
| FCW<sup>f</sup> | 70 | 20.7 | 68 | 20.7 | 70 | 20.8 |
| unknown | 38 | 11.2 | 36 | 11.0 | 36 | 10.7 |
| Σ | 338 | 99.9 | 328 | 100.0 | 337 | 100.0 |

<sup>a</sup>glycoside hydrolases  
<sup>b</sup>carbohydrate esterases  
<sup>c</sup>polysaccharide lyases  
<sup>d</sup>carbohydrate-binding modules  
<sup>e</sup>auxiliary activities  
<sup>f</sup>fungal cell wall

In addition to enzymes, proteins possessing one or more carbohydrate-binding modules but lacking domains for enzyme activity were found (CBM only, c. 6%). Among these the LysM motif (CBM50) has attracted most attention [56]. This module has general N-acetylglucosamine binding properties and is therefore able to bind to chitin or chitin-like carbohydrates [57]. In fungi, LysM domains are predominantly found in secreted LysM effector proteins and in subgroup C chitinases. In the *Rhynchosporium* genomes seven proteins are encoded that contain one or more LysM domains (LysM type A [57]; LysM1, LysM5 and LysM7 with one domain, LysM2 with two domains, LysM3 with three domains, LysM4 and LysM6 with four domains). In addition, two genes code for enzymes that contain LysM proteins that are encoded that contain one or more LysM domains (LysM type A [57]; LysM1, LysM5 and LysM7 with one domain, LysM2 with two domains, LysM3 with three domains, LysM4 and LysM6 with four domains). In addition, two genes code for enzymes that contain LysM proteins that are encoded that contain one or more LysM domains (LysM type A [57]; LysM1, LysM5 and LysM7 with one domain, LysM2 with two domains, LysM3 with three domains, LysM4 and LysM6 with four domains).
domain pairs (Additional file 6: Table S4A), a subgroup C chitinase (CAZy family GH18; LysM type B; see below), which also possesses a different chitin-binding motif (CBM18), and a putative peptidoglycan lytic transglycosidase (CAZy family GH128). Four of the LysM proteins are secreted through the classical pathway, whereas three of the proteins, LysM5, LysM6 and LysM7 and the GH23 enzyme may be secreted via the unconventional pathway. A central transmembrane domain predicted for LysM7 may cause this protein to reside in the plasma membrane with the LysM domain oriented outwards. Interestingly, in *M. brunnea* a family of 24 genes was identified that code for effector proteins of 142–151 aa with 1 or 2 LysM motifs [58]. This family is completely missing in *Rhynchosporium*. Two representative members of this gene family were shown to weaken the chitin-induced defense response when expressed in *Arabidopsis thaliana*. As proposed for the effector proteins ECP6 from *Cladosporium fulvum* [59], Mg3LysM from *Mycosphaerella graminicola* (Zymoseptoria tritici; [60]) and Slp1 from *Magnaporthe oryzae* [61], this effect is probably due to the sequestration of chitin oligosaccharides that occur as breakdown products of fungal cell walls during invasion and that are able to trigger host immunity. Other LysM effectors including Avr4 from *C. fulvum* [62], Mg1LysM and Slp1 prevented hyphal lysis by plant chitinases [60, 61].

When the *Rhynchosporium* CWDEs were assigned to their putative substrates (cellulose, hemicellulose, pectin, lignin, other polysaccharides), about 21% were found to act or are likely to act on the fungal cell wall, whereas c. 64% presumably target the host cell wall (Table 4, Additional file 5: Table S3, Fig. 6b). For about 11% of the enzymes their substrate is unknown or ambiguous. Fourteen enzymes (4%) are putative cutinases. These enzymes, albeit no proper CWDEs, are required at a very early stage of pathogenesis to allow or facilitate the fungal penetration of the plant surface.

The physiological role of the CAZymes targeting the fungal cell wall is thought to be in cell wall remodeling during fungal morphogenesis (septum formation, hyphal branching, spore germination). The cell walls of filamentous fungi are mainly built up of β-1,3-glucan (65–90%), glycoproteins (20–30%) and chitin (10–20%; [63]). Accordingly, glucan-degrading enzymes constitute the biggest group of hydrolytic enzymes encoded in the *Rhynchosporium* genomes with at least 22 members (CAZy families GH17, GH55, GH81, GH128, GH72, possibly supplemented by members of families GH3, GH16 and GH131 [64]). Eight enzymes presumably target glycoproteins (GH20, GH114, GH125; Additional file 6: Table S4). In addition, 17 chitinolytic enzymes were identified (GH18, GH75, GH89; CE4; Additional file 6: Table S4). Eight of the 14 chitinases (GH18) belong to subgroup A (40–50 kDa, no CBMs; [65]). The biggest of these chitinases is predicted for unconventional secretion. One member of this subgroup was found to lack a signal peptide and, hence, to be located intracellularly (cf. [66]). Three chitinases belong to subgroup B (30–90 kDa, CBMs), but two additional enzymes may fall into this group despite their high molecular mass (c. 120 kDa). Subgroup C chitinases were defined to have a high molecular mass (140–170 kDa) and to carry a chitin-binding domain (CBM18) and LysM motifs. These criteria are fulfilled by one *Rhynchosporium* enzyme of almost 150 kDa. This enzyme may not be fully released into the apoplast, but rather tethered to the outside of the plasma membrane as suggested by a predicted C-terminal transmembrane helix.

In *Trichoderma atroviride* the gene coding for subgroup C chitinase TAC6 is located next to the gene encoding the TAL6 protein, which contains 7 LysM motifs and was considered to inhibit spore germination [67]. TAC6 appears to be inactivated by a deletion in the catalytic domain and hence to be a chitin-binding protein. Interestingly, a tac6 deletion mutant was reported to
grow faster than the wild type, suggesting a growth de-
celerating function of the protein and thus a function in
the regulation of fungal development. Similar to T. atro-
viride, the subgroup C chitinase-encoding gene of R.
commune is located next to a gene coding for a protein
with 4 LysM motifs, LysM4. Furthermore, chitinase-
LysM gene pairs are also found for both high molecular
mass subgroup B chitinases. Their genes are located next
to the genes coding for LysM1 and for a protein that
lacks a LysM domain but has high similarity to a LysM
protein from Zymoseptoria brevis [68], respectively. This
chitinase-LysM gene arrangement may therefore point to
the participation of LysM proteins in a more wide-
spread mechanism to control fungal growth in planta.

Regrouping of the plant CWDEs according to the dif-
f erent types of cell wall substrates suggested that c. 32%
are involved in hemicellulose degradation and about
19% target cellulose (Fig. 6b). The enzymes acting on
pectin (8%) and lignin (5%) constitute smaller groups,
thus reflecting the low content of these materials in the
cell walls of grasses. Due to their large number and the
functional redundancy of enzyme activities involved in
cell wall degradation, the likelihood is generally pre-
sumed to be small that any one of these enzymes is re-
ponsible for host specificity. Recently, however, a com-
bination of comparative genomics and transcripto-
mics revealed that members of CWDE families from Z.
tritic i are differentially transcribed during pathogenesis
[69]. This suggested the specific occurrence of at least
some CWDEs during relevant stages of the fungal life
cycle rather than their general functional redundancy. In
addition, selection analysis yielded evidence that a few
cutinases, cellulases and hemicellulases may be involved
in host adaptation or evasion of host recognition.

Based on the genome sequences, a very coarse intra-
genus comparison was done for the Rhynchosporium
CAZY mases, which considered the presence or absence of
genes in a family. In several of the CAZY families (e.g.,
arabinose family GH93, poorly characterized family
GH89 [70]) single members are missing from one of the
three BCG species. It remains to be shown however
whether these differences reflect differences in the com-
position of the targeted cell walls and, thus, adaptations
to the host.

Secondary metabolism
Generally, secondary metabolites can be regarded as
tools that play a role in the interaction of the producing
organism with its environment. Fungal secondary prod-
ucts are not only notorious toxins, but also mediators of
chemical communication with other microbes and the
environment or defense agents [71]. Hence, they also
play a more or less prominent role in the interaction of
phytopathogenic fungi with their host plants. Most
fungal secondary metabolites belong to one of four
structural classes, polyketides, non-ribosomal peptides,
cylic terpenes or tryptophan-derived indole alkaloids
[72]. Among the fungal compounds with known effect
in plant-pathogen interactions are several of the host-
specific toxins of the Dothideomycetes (for review s.
[73]). T-toxin from Cochliobolus heterosporus and PM-
toxin from Mycosphaerella zeae-maydis are linear poly-
ketides. Victorin from Cochliobolus victoriae, AM-toxin
from Alternaria alternata and HC-toxin from Cochlio-
bolus carbonum are cyclic non-ribosomal peptides,
whereas HS-toxin from Bipolaris sacchari is a sesquer-
pene galactoside. Several fungi produce perylenequinone
toxins (e.g. cercosporin from Cercospora spp.). These are
cyclic polyketides with non-host-specific activity. Some
secondary metabolites have a less pronounced effect in
the interaction with host plants. Examples are the poly-
ketide botcinic acid and the sesquiterpene botrydial,
which play a redundant role in the virulence of Botrytis
cinerea on bean leaves [74]. Finally, many secondary
products have an impact on human and livestock health;
these include the ergot indole alkaloids with their
tryptophan-derived lysergic acid moiety [75] or the ses-
quiterpenoid trichotheccenes [76].

The only secondary metabolite described to date from a
Rhynchosporium species is orthosporin, a polyketide
from R. orthosporum [77]. This isocoumarin was also
isolated from Drechslera siccans, a pathogen of oats
(Avena sativa), perennial ryegrass (L. perenne) and Italian
ryegrass (L. multiflorum). It is phytotoxic causing necrosis
on leaves of several plants, but not on perennial ryegrass
[78], and it has an inhibitory effect on root growth of the
host plant, orchard grass (D. glomerata) and lettuce
(Lactuca sativa).

The four major fungal secondary metabolite classes
are synthesized via four characteristic key enzymes
[72, 79], polyketide synthases (PKS), non-ribosomal
peptide synthetases (NRPS), terpene cyclases (TC) or
dimethylallyl tryptophane synthases (DMATS). In
addition, several hybrid compounds are known, which
are either synthesized by PKS-NRPS hybrid enzymes
or by two separate enzymes that are encoded in a
single gene cluster. The Rhynchosporium genomes
were screened for the presence of these key enzyme
genomes to find clues for their possible involvement in
determining host specificity. In almost all isolates/species
three DMATS genes and one TC gene were identified,
whereas DMATS3 was missing only in R. orthosporum,
suggesting that a host-specific role of these enzymes is
unlikely.

PKS genes constitute the largest group of key enzyme
genomes of secondary metabolism. PKS catalyze the synthe-
sis of a large and structurally diverse group of chemical
substances from acyl-CoA precursors. Fungal PKS are
usually multidomain enzymes that elongate and modify their products iteratively (type I PKS; [80]). Three domains are essential for polyketide synthesis, ketoacyl CoA synthase (KS), acyltransferase (AT) and a phosphopantetheine attachment site (PP). Several other enzymatic motifs such as the three β-keto processing domains ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) and a methyl transferase domain (MT) are not found in all PKS. Lack of all or some of the reducing domains yields non-reducing or partially reducing PKS. The polyketide product is usually released from the enzyme through thioreductase (TR), thioesterase (TE) or transferase activities. Twelve PKS genes (PKS1-12) including 2 PKS-NRPS genes (PKS1, PKS11) occur in all BCG species (Table 5). In contrast, PKS13 was not found outside *R. commune* isolates and appears to be species-specific, whereas an additional gene, PKS14, was present only in the genome of *R. commune* isolate AU2. Orthologs of eight of the PKS genes were also identified in *R. orthosporum* along with four unique genes including two PKS-NRPS hybrid genes (PKS15, PKS16). In the closely related *M. brunnea, orthologs of only 6 Rhynchosporium* genes were identified along with two unique PKS genes. As the conclusion, the specific occurrence of several of the genes may suggest their involvement in the biosynthesis of polyketides with a role in interaction specificity at the species and isolate level.

The phylogeny of type I PKS from ascomycetes and bacteria as inferred from KS domain relationship revealed three main clades comprising fungal reducing PKSs, fungal non-reducing (NR) PKSs and bacterial PKSs [81]. The fungal clades are further subdivided into four reducing subclades (I-IV), three NR subclades (NR I-III) and a subclade basal to NR subclades I and II (NR b I + II). In addition, the bacterial PKS clade contains two subclades of fungal PKS. This phylogeny was first reconstructed based on the KS domain sequences from 86 ascomycete PKS (including eight from *M. brunnea* [33]), ten bacterial PKS and two animal fatty acid synthetases. Subsequently, it was complemented with the KS domain sequences of the 18 *Rhynchosporium* PKS (Fig. 7), which were predicted by BLASTp analysis [82] and comparison to the PFAM 28.0 database (Table 5; [83]). The archetypal PKS domain architecture KS-AT-DH-MT-ER-KR-PP [81] was found in only three *Rhynchosporium* PKS, RcPKS6, RcPKS7 and RoPKS17 (Table 5), which are members of reducing subclades I and IV, respectively. Both subclades contain additional *Rhynchosporium* PKS lacking the MT domain (I: PKS8, IV: PKS10, PKS18). Most *Rhynchosporium* PKS were in reducing subclade II, which lacks the ER domain and contains all PKS-NRPS hybrid enzymes. Notably, *R. commune*-specific PKS13 appears to be closely related to

### Table 5 Polyketide synthases

| Enzyme | *R. commune* | *R. secalis* | *R. agropyri* | *R. orthosporum* | *M. brunnea* | Domain architecture |
|--------|--------------|--------------|--------------|-----------------|--------------|---------------------|
| PKS1 (PKS-NRPS) | UK7 | 13-13 | AU2 | | | KS-AT-DH-MT-KR-PP-C-AP-PP-TR |
| PKS2 | | | | | | KS-AT-DH-ER-KR-PP |
| PKS3 | | | | | | SAT-KS-AT-DH-PP-PP-TE |
| PKS4 | | | | | | SAT-KS-AT-DH-PP-M-CE |
| PKS5 | | | | | | KS-AT-DH-MT-KR-PP |
| PKS6 | | | | | | KS-AT-DH-MT-ER-KR-PP-CAT |
| PKS7 | | | | | | KS-AT-DH-MT-ER-KR-PP |
| PKS8 | | | | | | KS-AT-DH-ER-KR-PP |
| PKS9 | | | | | | KS-AT-DH-ER-KR-PP |
| PKS10 | | | | | | KS-AT-DH-ER-KR-PP |
| PKS11 (PKS-NRPS) | | | | | | KS-AT-DH-ER-KR-PP |
| PKS12 | | | | | | KS-AT-DH-ER-KR-PP |
| PKS13 | | | | | | A-PP-KS-AT-DH-KR-PP-TE |
| PKS14 | | | | | | KS-AT-DH-MT-KR-PP |
| PKS15 (PKS-NRPS) | | | | | | KS-AT-DH-MT-KR-C-AP-PP-TR |
| PKS16 (PKS-NRPS) | | | | | | KS-AT-DH-MT-KR-C-AP-PP-TR |
| PKS17 | | | | | | KS-AT-DH-MT-ER-KR-PP |
| PKS18 | | | | | | KS-AT-DH-ER-KR-PP |

KS: ketosynthase, AT: acyltransferase, DH: dehydratase, MT: methyltransferase, ER: enoylreductase, KR: ketoreductase, PP: thiolation domain

1. two additional PKS similar to PKS3
2. gene in cluster with DMT173
3. gene in cluster with NRPS2
LovB, one of the enzymes involved in the biosynthesis of lovastatin [84] in Aspergillus terreus (47.3% identity, 63.5% similarity, 6.7% gaps), suggesting that both enzymes catalyze similar reactions. Lovastatin is an inhibitor of 3-hydroxy-3-methylglutaryl-CoA reductase, which catalyzes a crucial step in the biosynthetic pathway leading to ergosterol (antifungal properties) or cholesterol (hypolipidemic agent; [85]). Two Rhynchosporium enzymes, PKS3 and PKS4, occur in the non-reducing clades II and III, respectively. Like all other Rhynchosporium PKS, however, the latter contain the DH reduction domain and therefore should be called partially reducing PKSs. Finally, PKS12 groups with C. heterostrophus PKS24 in the bacterial PKS clade, both being NRPS-PKS hybrids (cf. also NRPS below) and suspected of being horizontally transmitted from bacteria to fungi [86].

Among the key genes of secondary biosynthesis in Rhynchosporium those coding for NRPS and NRPS-like proteins represent the second largest group. These enzymes typically consist of one or more modules comprising an amino acid-adenylation domain (A), a thiolation domain (PP) and a condensation domain (C). Monomodular NRPS4, tri-modular NRPS3 and tetra-modular NRPS2 were present in all species, whereas hexa-modular NRPS1 was missing in R. orthosporum (Table 6). Remarkably, R. commune isolate AU2 was again unique by carrying a specific gene coding for the most complex enzyme, the octa-modular NRPS5. Orthologs of the eight genes coding for NRPS-like (NPSL) proteins were found in all Rhynchosporium species with the exception of NPLS4, which was missing in R. orthosporum. In comparison to Rhynchosporium, closely related M. brunnea had orthologs of only six of the NRPS and NPSL genes. Only one NRPS gene shows isolate-specific occurrence indicating a possible role of its product in host specificity at the isolate level. In addition, several genes may be involved in BCG specificity.

Neither structures nor functions of the NRPS products in Rhynchosporium spp. are known. Therefore, the domain structures and, in particular, the A domains of the Rhynchosporium NRPS and NRPS-like proteins were compared to those of other fungi and bacteria. In a
previous phylogenomic analysis of fungal NRPS and NRPS-like (NPSL) proteins [86] two main enzyme groups had been identified, one group containing seven subfamilies of primarily mono- and bi-modular fungal enzymes along with the bacterial clade (Fig. 8), the other group containing two subfamilies of primarily multi-modular and exclusively fungal (Eusacymycetes, EAS; Fig. 9) enzymes. Members of the latter are presumed to be of more recent origin, most of them having less conserved structures suggesting the biosynthesis of products with more specific functions in niche adaptation including virulence. In contrast, members of the former group are presumably of more ancient origin, exhibiting more conserved domain architectures suggestive of more basic functions.

Following the previous analysis [86], the A domains of NRPS from Rhynchosporium were compared to those of other fungi including the closely related species M. brunnea and bacteria. Most NRPS subfamilies contain at least one member from the Rhynchosporium genus. Two proteins, tri-modular RcNRPS3 and mono-modular RcNRPS4, are predicted to be involved in the biosynthesis of two types of Fe$^{3+}$-chelating hydroxamate-type siderophores required for iron uptake and storage. RcNRPS3 and its M. brunnea ortholog (XP_007288653) group with type-IV siderophore synthetases [87] (Fig. 8). Each of its three A domains joins a sub-branch with one of the A domains of C. heterostrophus NPS2, a ferrichrome synthetase. Ferrichromes, which are involved in intracellular iron storage, are typically cyclic hexapeptides built up of three $N^5$-acetyl-$N^5$-hydroxy-L-ornithines (AHO) and three small amino acids, an obligatory Gly and one to two Ser, Gly or Ala [87]. RcNRPS3 and its M. brunnea ortholog share the domain architecture with C. heterostrophus NPS2 including the C-terminal C-PP-C-PP module, which is predicted to function in ring closure of the nascent peptide [88]. In addition, the amino acid residues suggested to determine the substrate specificity of the A domains are almost identical in both enzymes and very similar to those of the A domains of C. heterostrophus NPS2 (Table 7). The NRPS2 product, ferricrocin, is an intracellular siderophore involved in iron storage, oxidative stress resistance, germination and sexual development in A. nidulans and other fungi [89]. This molecule is made up of two glycines and one serine in addition to the three AHO [90]. According to the 10-aa code for amino acid binding, the A1 domain of the NRPS from the three fungi binds

| Enzyme       | R. commune | R. secalis | R. agropyri | R. orthosporum | M. brunnea | Domain architecture                                      |
|--------------|------------|------------|-------------|----------------|------------|--------------------------------------------------------|
| NRPS1 (n=6)  | UK7        | 13-13      | AU2         |                |            | [A-PP-C]_4                                              |
| NRPS2 (n=4)  |            |            |             |                |            | [A-PP-C]_4                                              |
| NRPS3 (n=3)  |            |            |             |                |            | (A-PP-C) dA-PP-C                                       |
| NRPS4 (n=1)  |            |            |             |                |            | (A-PP-C) dA-PP-C                                       |
| NRPS5 (n=8)  |            |            |             |                |            | (A-PP-C) rC-A-PP-C                                       |
| NRPS-like 1  |            |            |             |                |            | A-PP-TR                                               |
| NRPS-like 2  |            |            |             |                |            | A-PP-TR                                               |
| NRPS-like 3  |            |            |             |                |            | A-PP-TR-DH                                             |
| NRPS-like 4  |            |            |             |                |            | A-PP-TR                                               |
| NRPS-like 5  |            |            |             |                |            | A-PP-TR                                               |
| NRPS-like 6  |            |            |             |                |            | A-PP-TR                                               |
| NRPS-like 7  |            |            |             |                |            | A-PP-MAT-like                                           |
| NRPS-like 8  |            |            |             |                |            | A-FAD/NAD(P)-binding-site                              |
| DMATS1        |            |            |             |                |            | n.d.                                                  |
| DMATS2        |            |            |             |                |            | n.d.                                                  |
| DMATS3        |            |            |             |                |            | n.d.                                                  |
| TC1          |            |            |             |                |            | n.d.                                                  |

1 gene in cluster with PKS12
2 Dimethylallyl tryptophan synthase
3 gene in cluster with PKS4
4 terpene cyclase

**Table 6** Nonribosomal peptide synthetases

- A: adenylation, PP: thiolation, C: condensation
glycine and the A2 domain binds serine [87, 91]. In the A3 domains of RcNRPS3 and its M. brunnea ortholog an L→M exchange has occurred compared to the A4 domain of C. heterostrophus NPS2. Nevertheless, the domain appears to be responsible for AHO binding (10-aa code: 9/10, 13-aa code: 11/13, 17-aa key positions: 14/17). Therefore, despite the presence of the second presumably glycine-binding domain A3 in C. heterostrophus NPS2, the siderophore product of RcNRPS3 and its M. brunnea ortholog is predicted to be identical with the NPS2 product, ferricrocin.

RcNRPS4 was in the most conserved clade of the EAS group exemplified by C. heterostrophus NPS6 [86] coding for a different type of siderophore synthetases (Fig. 9). No M. brunnea homolog was identified. RcNRPS4 has the typical domain structure of this enzyme group, a single A-PP-C module followed by a module with a degenerate A domain (dA-PP-C). NPS6 of C. heterostrophus was shown to be responsible for the synthesis of extracellular siderophores of the coprogen family [92]. Deletion of the NPS6 gene and its orthologs in different plant pathogenic fungi resulted in reduced virulence and hypersensitivity to H2O2.

Mono-modular RcNPSL3 and its M. brunnea ortholog belong to the NPS10 subfamily, which also has a conserved domain structure. The A-PP domain is followed by a thioester reductase (TR)-like domain and a dehydrogenase domain. RcNPSL4 has a single A-PP-TR
module and groups together with mono-modular enzymes from *A. nidulans* (AN8105) and *A. fumigatus* (Afu8g01640) with the cyclosporin synthetases. RcNPSL6 is in a clade with high bootstrap support that comprises α-aminoadipate reductases (AAR) such as Lys2 from *S. cerevisiae* and Lys1+ from *S. pombe*. AAR are evolutionary related to but not identical with NRPS and catalyze a step in the fungi-specific biosynthesis of lysine [93]. Neither RcNPSL7 nor *B. cinerea* protein BC1G_11613 fits into one of the clades described. Both proteins have the same domain architecture. The A and PP domains are followed by a domain that makes up the C-terminal half of the protein, is predicted to contain three pairs of transmembrane domains (TMDs) and is exclusively found in NRPS.

RcNPSL8 is in the same clade as *C. heterostrophus* NPS12 and *Alternaria brassicicola* TmpL (with high bootstrap support). These proteins lack thiolation (PP) and condensation (C) domains and, hence, are no true NRPS proteins. Instead, they are characterized by an A domain and a C-terminal FAD/NAD(P)-binding domain, which are separated by putative TMDs. The number of TMDs is predicted to be between 5 and 7 (depending on the exclusion or inclusion of two TMDs with low probability following TMD1). Therefore, membrane orientation of these proteins needs to be verified experimentally. TmpL was shown to be involved in virulence through regulation of intracellular ROS concentrations and tolerance to external ROS [94]. The biochemical function of the TmpL-like proteins is not known, but they are

---

**Fig. 9** NRPS phylogenetic tree 2. The amino acid sequences of the A domain from 151 mostly oligo-modular specifically fungal NRPS were used to construct the phylogenetic tree. Numerals at the nodes represent the percentages from 500 bootstraps. Numerals <50 were omitted. Colored backgrounds indicate enzyme groups: blue, NPS8-like; pink, NPS6-like, orange, ergot alkaloid synthetases, green, peptaibol synthetase TEX1; grey, AM-toxin synthetase. Red and blue dots mark the modules of HC-toxin and peramine synthetase, respectively. Coloured boxes indicate modules of the 4 Rhynchosporium NRPS.
suggested to possess an enzymatic function involving FAD/NAD(P) binding like other flavoenzymes such as enzymes from the ferric reductase group.

The basic architecture of NPSL1-6 is very similar with A- and PP-domains followed by a TR domain. However, the A domains of RcNPSL3, RcNPSL4 and RcNPL6 are structurally similar to the A domains of RcNRPS1-5 (NCBI PSSM-ID CD05930), but deviate from those of RcNPSL1 (CD04433), RcNPSL2 (CD05907) and RcNPSL5 (CD05936). As a consequence, the latter along with RcPKS12 (NRPS-PKS hybrid, CD04433) form a phylogeny outgroup. Other outgroup A domains comprise the Rhynchosporium ortholog of CPS1, a protein consisting of two A domains (A1: CD05931, A2: CD05905), which was shown to be a virulence factor of C. heterostrophus and Gibberella zeae on maize and of C. victoriae on oats [95]. Further, A domains from bacterial proteins, B. subtilis 2,3-dihydroxybenzoate-AMP ligase (DbhE, CD5920) and P. aeruginosa pyoverdine synthetase (PvdL A1, CD05931), as well as an A. nidulans NPSL protein, CBF82791 (CD05906), were used as outgroup.

In the mostly multi-modular EAS group (Fig. 9), little information can be deduced from the phylogenetic position of the A domains regarding the function of NRPS products. However, occurrence of the A domains of tri-modular AM-toxin synthetase from A. alternata or 14 of the 19 modules of peptaibol synthetase TEx1 from Trichoderma virens in a single subclade with high bootstrap support is suggestive of an evolutionary origin of the enzymes by duplication events. A similar mechanism may have played a role in the evolution of Rhynchosporium NRPS enzymes. Three of the A domains of hexa-modular RcNRPS1 occur in one subclade, two in another subclade and one in a third subclade. Similarly, the A domains of octa-modular RcNRPS5 are in three subclades with four, three and one members, respectively. In contrast, the A domains of tetra-modular RcNRPS2 and its M. brunnea ortholog are spread over four different subclades indicating a recombination-type mode of molecular evolution. Finally, the four PKS-NRPS (PKS1, PKS11, PKS15, PKS16) from Rhynchosporium and the PKS1 ortholog from M. brunnea, which all belong to (partially) reducing PKS clade II (cf. PKS above), are found in the PKS-NRPS subfamily of the phylogenetic tree (Fig. 8). They all have the same domain architecture except for RcPKS16, which lacks the C-terminal PP domain (for review s. [96, 97]).

The basic structures formed by the key secondary biosynthesis enzymes are usually modified by decorating

Table 7 Amino acid code of siderophore synthases

|   | 226 | 229 | 230 | 235 | 236 | 239 | 240 | 243 | 276 | 278 | 280 | 299 | 301 | 312 | 322 | 326 | 330 | 331 | 517 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| R. commune Sid_A3 | Y | F | A | D | V | M | E | W | H | D | T | I | G | Y | G | S | I | G | K |
| M. brunnea Sid_A3 | F | G | A | D | V | M | E | W | H | D | T | I | G | Y | G | T | I | G | K |
| C. heterostrophus NPS2_A4 | W | F | A | D | V | L | E | W | H | D | T | I | G | Y | G | T | I | G | K |
| F. graminearum NPS2_A3 | W | F | A | D | V | L | E | W | H | D | T | I | G | Y | A | T | I | G | K |
| A. nidulans SidC_A3 | L | T | G | D | P | E | E | H | S | V | T | G | V | A | T | I | G | K |
| R. commune Sid_A1 | F | F | A | D | V | F | F | E | F | A | E | T | L | I | Y | M | T | I | H | K |
| M. brunnea Sid_A1 | F | F | A | D | V | F | F | E | F | A | E | T | L | I | Y | M | T | I | H | K |
| C. heterostrophus NPS2_A3 | F | F | A | D | V | F | F | E | F | A | E | T | L | I | W | M | T | I | H | K |
| F. graminearum NPS2_A1 | F | F | A | D | V | F | F | E | A | E | T | L | I | W | M | T | I | H | K |
| A. nidulans SidC_A1 | W | F | A | D | P | M | E | V | V | M | T | W | M | V | A | T | I | N | K |
| R. commune Sid_A2 | F | S | C | D | V | F | F | H | S | T | T | A | W | G | S | N | I | K |
| M. brunnea Sid_A2 | F | S | C | G | V | F | E | F | H | S | T | T | A | W | G | S | N | I | K |
| C. heterostrophus NPS2_A2 | L | A | C | D | V | F | E | F | H | S | T | T | A | W | G | S | N | I | K |
| F. graminearum NPS2_A2 | L | A | C | D | V | F | E | F | Y | H | S | T | V | A | W | G | S | N | I | K |
| A. nidulans SidC_A2 | C | F | S | D | V | Q | D | Y | H | T | T | T | F | T | A | V | V | K |

10-aa code
13-aa code

Gly
Ser

N⁵-acetyl-N⁵-hydroxy-L-ornithine

Ferricrocin (NPS2)
enzymes such as oxidoreductases, glycosyl or prenyl transferases. Their genes are frequently found in contiguous clusters, i.e., most or all genes of a biosynthetic pathway, possibly along with genes responsible for regulation and transport, are co-localized and co-regulated [98]. For instance, the cluster for the biosynthetic pathway of the polyketide asperfurane in A. nidulans comprises genes coding for two fungal PKSs, five additional enzymes and a transcriptional activator [99]. Similarly, putative PKS gene clusters were detected in the Rhynchosporium BCG genomes (Additional file 7: Table S5). Two clusters contain genes coding for the key biosynthetic enzymes of two different secondary metabolite classes: the PKS4 cluster includes the DMATS3 gene (Fig. 10) and the NRPS2 cluster harbors the PKS12 gene. DMATS3 might be a transferase required for the prenylation of a PKS4-derived polyketide. Alternatively, the gene pairs might be involved in the biosynthesis of mixed compounds through separate assembly lines.

**Small secreted proteins: necrosis-inducing proteins**

Secreted effector molecules have been recognized as crucial for host colonization and pathogenicity [100]. In both pathogens and mutualists they function by deregulating the host immune response. In addition, they have a role in self-defense and competition with other microbes [101, 102]. Furthermore, they influence the physiology of the host in favor of the pathogen’s needs, thus enhancing virulence [103]. Finally, they contribute to host adaptation and to establishing host specificity, thereby driving microbial speciation [104].

In *R. commune*, the genes NIP1, NIP2 and NIP3 code for effector proteins that contribute quantitatively to the virulence of *R. commune* depending on the host cultivar [105]. Therefore, the sequenced genomes of all Rhynchosporium species were screened for the presence of orthologous genes using the NIP gene sequences from *R. commune* isolate UK7 as probes. In this context, only those *R. lolii* genes that have homologs in *R. orthosporum* could be included. NIP1 and NIP3 occur as single genes in the genomes. NIP3 orthologs coding for highly similar proteins are present in the three BCG species, whereas NIP1 orthologs are found in *R. commune*, albeit not in all isolates (cf. [106]), and in the CCG species *R. orthosporum* (Fig. 11). NIP2 is also a single gene in the two CCG species. In contrast, NIP2 families of 7-10 members are present in the BCG species (NIP2.1- NIP2.11).

Comparison of the deduced NIP2 amino acid sequences using Multiple Expectation Maximization for Motif Elicitation (MEME, version 4.9.1; [107]) revealed three conserved protein motifs, a 40-amino acid stretch spanning the N-terminal half of the protein, followed by a 15-amino acid domain around an absolutely conserved CRS motif, and near the C-terminus of about 50% of the proteins another 15-amino acid sequence. In addition, BLASTp analysis identified genes coding for proteins with similarity to NIP2 in the BCG genomes. These NIP2-like proteins (NLPs) differ from NIP2 by a smaller degree of domain conservation, but most strikingly by the absence of the CRS motif, which is absolutely conserved in NIP2 (Fig. 12). All NIP2 and NLP carry six cysteine residues at homologous positions. Interestingly, all proteins except NIP2.2 have a cysteine at the C-terminus. This structural feature was also found in NIP1. The phylogenetic tree of the mature NIP2 proteins demonstrates that the different isoforms are more similar between fungal species than they are to each other. This implies that duplication of the NIP2 gene occurred in a common ancestor before the three BCG species diverged, possibly to allow or facilitate host jump and colonization of new cereal crop species.

To date no biochemical function could be assigned to NIP2. In contrast, NIP1 and NIP3 both stimulate the host plant plasma membrane H⁺-ATPase [108]. This enzyme generates a cellular membrane potential, which may affect nutrient accumulation and fungal membrane transport processes. In addition, acidification of the apoplast may create more convenient growth conditions for the fungus [109]. The originally reported induction of necrosis by the purified proteins [110] may therefore be a consequence of over-stimulation of the H⁺-ATPase due to severe overloading of the host leaf tissue [23]. An amino acid sequence comparison of NIP1 and NIP3 revealed that a stretch of six amino acids, -G<sub>45</sub>EGGN<sub>50</sub>-, which is present in exposed loop III near the C-terminus of RcNIP1 [111], but absent from RoNIP1 (50.8% similarity, 39.7% identity with RcNIP1), is almost identically
found near the C-terminus of NIP3 as -GEGEG-. Remarkably, G45 is essential for NIP1 activity [7, 112], which suggests that this amino acid motif could be critical for the common function of NIP1 and NIP3.

NIP1 from *R. commune* is targeted by the barley resistance gene *Rrs1* [7], thereby triggering defense reactions in the host [113]. To be able to infect *Rrs1*-barley the fungus therefore evolved two strategies to prevent recognition of NIP1, namely introduction of mutations into the *NIP1* gene and its complete deletion [106, 112]. As the consequence, *NIP1* was only found in about half of the *R. commune* isolates analyzed and inactive NIP1 isoforms were identified [7, 112]. Nevertheless, the importance of NIP1 as a virulence factor of *R. commune* was shown in a recent study on the quantitative virulence of 14 fungal isolates from each of nine global field populations worldwide. As the result, “on average isolates carrying a functional NIP1 protein had significantly higher virulence than isolates where the protein was missing or non-functional” [114]. In contrast to NIP1, NIP3 was found in almost all *R. commune* isolates studied [106] as well as in *R. secalis* and *R. agropyri*. Hence, at least one of the H⁺-ATPase stimulators, NIP1 or NIP3, appears to be present in a given fungal species/isolate.

To give a résumé, NIP1 and NIP3 appear to be specific for *R. commune* and the BCG species, respectively. The presence of an – albeit possibly “inactive” – NIP1 in *R. orthosporum* indicates that the gene may have been lost in the other species. In contrast, each of the five *Rhynchosporium* species harbors at least one NIP2 gene. However, evolution of the NIP2 gene family by gene duplication and sequence diversification [115] suggests an involvement of this protein in the adaptation of the BCG species to their hosts.

**Small secreted proteins: specific occurrence of effector candidates in *R. commune***

The NIP effectors had been identified originally through their necrosis-inducing activity [110]. With the genome sequences available, additional candidate effector genes were obtained through an *in silico* approach. To this end, several general criteria based on existing knowledge about protein effectors from other fungal pathogens were established that ought to be fulfilled by candidate effector proteins (cf. [116, 117]). Effectors are secreted, i.e. most of them carry a signal peptide while lacking an additional transmembrane domain. Furthermore, they are usually of small size and cysteine-rich. In addition, they should be specific for the genus *Rhynchosporium* and expressed during pathogenesis, probably during the stage of fungal establishment *in planta*. Of particular interest are those effector genes that are present exclusively in the genome of a single species and, hence, may be involved in determining host specificity of this fungal species.

The genome of *R. commune*, isolate UK7, was screened for candidate genes coding for proteins that match the general criteria (Fig. 13). The subgroup containing ≥2% cysteine residues contained 139 proteins. The occurrence of these genes was probed (BLASTn) in the genomes of the other *Rhynchosporium* species (with the exception of *R. lolii*) to identify candidates, which may be crucial for host specificity. In *R. commune*, seven gene models (*RcSP1, RcSP2, RcSP3, RcSP4, RcSP5, RcSP6, RcSP9*) were found to be specifically present, among them the NIP2.6 paralog (*RcSP6*). When the OrthoMCL algorithm [118] was applied to the
Rhynchosporium genomes, 29 *R. commune* proteins were found to lack orthologs in the other species. BLASTn analysis identified the seven *RcSP* genes and two additional genes in the <2%-cysteine fraction (*RcSP7*, *RcSP8*). No candidate proteins were identified in the 201–350 amino acid size fraction.

The gene models as predicted by the annotation algorithms were verified using EST libraries from *R. commune* isolate UK7 grown in liquid culture (A. Navarro-Quezada and W. Knogge, unpublished), from germinated conidia of British *R. commune* isolate 214 and from epidermal strips of barley leaves 3 dpi with British *R. commune* isolate 2LA (A. Avrova, unpublished). In addition, RT-PCR was carried out using gene-specific primers and RNA from *ex planta* grown *R. commune* isolate UK7 as a template. Sanger sequencing of the amplification products confirmed the structures of seven of the *R. commune* genes (*RcSP1*, *RcSP2*, *RcSP3*, *RcSP5*, *RcSP6*, *RcSP7*, *RcSP9*). In contrast, *RcSP4* had been annotated incorrectly, while *RcSP8* was not expressed. More detailed BLASTn and OrthoMCL analysis detected paralogs of *RcSP1*, *RcSP2*, *RcSP3* and *RcSP5* in *R. commune* and homologs in other *Rhynchosporium* species (Table 8, Additional file 8: Table S6). The original genes were renamed accordingly.

Expression of the seven *R. commune* candidate effector genes during pathogenesis was quantified using RNA
extracted from inoculated leaves of susceptible barley cultivar ‘Ingrid’ at different time-points post inoculation with \( R. \ commune \) isolate UK7 [105]. Transcript amounts of the target genes were determined relative to the transcript amounts of a constitutively expressed fungal gene, \( GPD \), by qRT-PCR [119]. As observed for the \( NIP \) effector genes [105], the \( RcSP \) genes were all transiently expressed early during fungal infection of susceptible host leaves. Between 10 and 14 dpi, when disease symptom occur, \( RcSP \) mRNAs virtually can no longer be detected (Fig. 14). \( RcSP2.1, RcSP5 \) and \( RcSP9 \) showed a maximum transcript abundance at 3 dpi, whereas \( RcSP1.1 \) and \( RcSP3.1 \) mRNA levels reached a maximum at 6 dpi. However, \( RcSP1.1 \) mRNA abundance accounted for only about 4\% of that of \( RcSP3.1. \ RcSP6 \) and \( RcSP7 \) mRNA levels were even lower, peaking at 3–6 dpi. When the infection process of the hemibiotroph \( Colletotrichum \ higginsianum \) on \( Arabidopsis \ thaliana \) was studied successive waves of effector gene expression could be observed during the consecutive developmental stages from unpenetrated spores through penetrated biotrophic hyphae to the biotrophy to necrotrophy switch [120]. The different stages of \( Rhynchosporium \) development cannot be discerned as clearly as in \( C. \ higginsianum \). Nevertheless, maximum transcript abundance for some genes (\( RcSP2.1, RcSP5, RcSP9, NIP1 \)) occur at 1–3 dpi, for others (\( RcSP1.1, RcSP3.1, NIP2, NIP3 \)) at 6 dpi suggesting a similar sequence of gene expression waves.

To assess the role of the \( R. \ commune \)–specific effector genes during pathogenesis, deletion mutants were generated. Using split-marker recombination [121] single deletion mutants were obtained for \( RcSP1.1 \) and \( RcSP6 \), two independent mutants for \( RcSP5.1 \) and three for \( RcSP2.1, RcSP3.1 \) and \( RcSP9 \). After inoculation of susceptible barley cultivar ‘Ingrid’ none of the deletion mutants caused a phenotype deviating substantially from that caused by wild type isolate UK7 at 14 dpi (not shown). However, when fungal biomass in planta was quantified by qPCR 10 of the 13 individual \( RcSP \) deletion mutants had grown to a significantly greater amount (1.7–4.5-fold) than the wild type UK7 (Fig. 15, Additional file 9: Figure S3), whereas deviations from wild type did not occur during fungal growth on agar plates (not shown). Their overlapping 95\% confidence intervals allowed to combine the data of the independent deletion mutants of single \( RcSP \) genes (Fig. 15). The weakest growth response occurred with the \( \Delta RcSP5 \) mutant, whereas the largest confidence interval was found with the \( \Delta RcSP6 \) mutant. The latter finding may be due to the fact that \( RcSP6 \) is identical with \( NIP2.6 \), which is part of a gene family with seven additional members in \( R. \ commune \) isolate UK7. The observed growth increase of the deletion mutants is in contrast to the \( NIP \) gene deletion mutants, which did not differ from wild type (\( \Delta NIP1, \Delta NIP2 \)) or grew substantially slower (\( \Delta NIP3 \)) than the wild type on highly susceptible cultivar ‘Ingrid’. (All three mutants showed strong growth reduction on moderately susceptible barley cultivar ‘Atlas’; [105]). When growth of the three \( RcSP9 \) deletion mutants and that of wild type were compared in more detail during pathogenesis, mutant biomasses deviated significantly from that of isolate UK7 from about 6 dpi on (Fig. 16, Additional file 10: Figure S4). In addition, close inspection of disease phenotype development revealed that the first symptoms became visible already at 8–10 dpi with the deletion mutants, whereas wild type symptoms did

| Table 8 RcSP homologs |
|-----------------------|
| \( R. \ commune \) protein | Identity [%] | Orthologs | \( R. \ secalis \) | \( R. \ agropyri \) | \( R. \ orthosporum \) |
| RcSP1.1 | 100 |  |  |  |
| RcSP1.2 | 50.0 |  |  |  |
| RcSP1.3 | 27.6 | RsSP1.3 (44.9) |  |  |
| RcSP2.1 | 100 |  |  |  |
| RcSP2.2 | 44.4 |  | RasP2.2 (82.1) |  |
| RcSP3.1 | 100 |  | RasP3.1 (63.4) |  |
| RcSP3.2 | 46.5 | RsSP3.2 (45.5) | RasP3.2 (46.6) |  RoSP3.1 (47.5) |
| RcSP3.3 | 57.4 | RasP3.3 (98.1) |  |  RoSP3.3 (47.0) |
| RcSP3.4 | 100 |  |  |  |
| RcSP5.1 | 28.0 |  | RasP5.2a (98.2) RasP5.2b (51.0) RaSP5.2c (55.8) |  |
| RcSP5.2 | 55.3 | RasP5.3 (52.5) |  |  RoSP5.3 (37.2) |
| RcSP5.3 | 26.7 | RsSP5.4 (78.9) | RasP5.4 (75.0) |  |

Numbers in brackets indicate the amino acid identity (%) with the respective \( R. \ commune \) protein.
not occur before 10 dpi (Fig. 17; Additional file 11: Figure S5). At 14 dpi the mutant phenotype appeared more severe than the wild type phenotype; almost the entire upper blade of the mutant-inoculated leaves was necrotic, whereas green tissue remained between the lesions of wild type-infected leaves.

In summary, significant quantitative effects on fungal growth and symptom development became apparent upon deletion of the six *R. commune*-specific genes. Notably, the deletion-caused growth acceleration suggested that the fungus is not adapted for the most rapid development in planta. This is similar to mutants of the grass endophyte *Epichloë festucae*, which had switched from mutualistic to pathogenic growth including a dramatically increased fungal biomass on perennial ryegrass. Different types of non-effector proteins were inactivated in the mutants; a MAP kinase (SakA, [122]), an NADPH oxidase (NoxA, [123]), which is regulated by NoxR [124] through interaction with small GTP binding protein RacA [125], and a siderophore-synthesizing NRPS (SidN, [126]). Therefore, it was postulated that functional stress signaling, regulated ROS synthesis and siderophore-mediated iron homeostasis are crucial for restricting fungal growth as prerequisite for maintaining the mutualistic interaction with the host. It is tempting to speculate that the *R. commune* RcSPs fulfil an analogous role in affecting hyphal growth, probably by interacting with plant compounds, thereby establishing an endophyte-like growth in host leaves during the early symptomless period of pathogenesis.

It remains to be shown whether the effects caused by the RcSPs contribute to host specialization. However, the occurrence of specific effector genes in the genome of one *Rhynchosporium* species and their absence from others species represents the simplest model to explain host specificity. Alternatively, effector genes, albeit present in several *Rhynchosporium* genomes, may be specifically expressed only in the interaction of one of them with its host plant and not in the specific interactions of other species. Finally, host specificity may
originate from structural differences of effector proteins rather than from presence or expression of specific genes. In addition, small molecules such as host-specific tosins originating from fungal secondary metabolism may play a role in colonizing a particular host species.

**Conclusions**

Phylogenetic analysis confirmed that the *Rhynchosporium* genus belongs to the *Leotiomycetes* clade of fungal systematics, most likely to the *Dermataceae* family, the BCG and CCG species clearly forming separate branches of the *Rhynchosporium* genus. Genome-wide SNP analysis along with the structure of the rDNA regions suggested that during BCG evolution *R. agropyri* branched off shortly before *R. commune* and *R. secalis* diverged. At least in the BCG species, complete mating type loci exist. In addition, core meiotic genes are present and expressed along with most genes required for meiosis and sexual recombination in other (ascomycete) fungi, providing molecular evidence for the existence of a cryptic sexual cycle.

Bioinformatic analysis revealed a large number and types of CWDEs, which are characteristic for hemibiotrophic and necrotrophic fungi infecting monocotyledonous hosts, but failed to provide evidence for their involvement in host specialization. In contrast, the occurrence of genus-, species- and isolate-specific PKS and NRPS genes suggests a possible role in host adaptation in particular of those genes, which are missing from the genome of the closest sequenced relative, the poplar pathogen *M. brunnea*. The most important finding was, however, that six *R. commune*-specific effector proteins contribute to decelerating fungal development *in planta*. Host adaptation of this hemibiotrophic fungus therefore appears to aim at stabilizing the biotrophic growth stage in favor of the necrotrophic destructive stage. Unlike haustoria-forming fungi, apoplastic fungi like *R. commune* colonize the intercellular matrix of their hosts relatively slowly, usually without triggering a fast hypersensitive cell death response. The long biotrophic stage of *Rhynchosporium* development resembles endophytic fungal growth, rendering this fungus a possible object for studying the mutualism-parasitism transition.

**Methods**

**Fungal strains**

The following fungal isolates were subjected to genome sequencing; for *R. commune* UK isolates UK7 and 13-13 and Australian isolate AU2 [21], for *R. secalis* Swiss isolate 02CH4-6a.1 [17], for *R. agropyri* Swiss isolate 04CH-RAC-A.6.1 (B. McDonald, ETH Zurich), for *R. lolii* UK isolate 15lp11 [16], and for *R. orthosporum* Swiss isolate 04CH-BAR-A.1.1.3 (B. McDonald, ETH Zurich).
Fungal cultures, DNA extraction, genome sequencing and assembly

Fungal culture conditions, DNA extraction and genome sequencing were described previously [30]. Briefly, the genomes of R. commune, isolates UK7 and 13-13 and R. secalis were sequenced using the Roche 454 Genome Sequencer FLX with GS FLX Titanium series reagents, fosmids and the Illumina Genome Analyzer IIx (paired-end and mate-pair libraries). Paired-end libraries of R. commune, isolate AU2, R. agropyri (along with a mate-pair library) and R. orthosporum were sequenced using Illumina GAIIx and HiSeq. To establish the genome of R. commune, isolate UK7, as the reference genome, contigs were assembled using the ALLPATHS-LG software package [127] and pre-assembled contigs were combined into scaffolds using the SSPACE algorithm [128]. The same strategy was followed to assemble the genome of R. agropyri, whereas genome assembly for R. commune, isolate AU2 and for R. orthosporum was carried out using the Velvet algorithm package [129], and for R. secalis using the Newbler software package. For R. commune isolate 13-13, 454 and Illumina GAIIx reads were assembled independently using the Newbler and Velvet algorithm packages, respectively, followed by combined assembly using Minimus2 [130]. The genome of isolate 15lp11 from the fifth species, R. loli [16], was sequenced by Richard Harrison (Illumina MiSeq, East Malling Research, East Malling, UK) and a draft assembly using the VELVET algorithms [129] was obtained by David Hughes (University of Hertfordshire, Hatfield, UK). Sequence reads of this isolate were mapped to the R. orthosporum genomic contigs. Hence, differential sequence information could be identified only for R. loli genes that are orthologous to R. orthosporum genes.

Structural annotation

Gene models for R. commune UK7, R. secalis and R. agropyri were generated by three de novo prediction programs: 1) Fgenes [131] with different matrices (trained on Aspergillus nidulans, Neurospora crassa and a mixed matrix based on different species), 2) GeneMark-ES [132] and 3) Augustus [133] with R. commune ESTs as training sets. Annotation was aided by exonerate [134] hits of protein sequences from B. cinerea T4, B. cinerea B05.10, S. sclerotinia, R. commune UK7, R. secalis and R. agropyri, respectively, to uncover gene annotation gaps. EST transcripts were assembled and mapped on the genome using Blat, a final cleaning was done with Perl Scripts. The different gene structures and evidences (exonerate mapping and EST transcripts) were visualized in GBrowse [135], allowing manual validation of coding sequences. The best fitting model per locus was selected manually and gene structures were adjusted by splitting or fusion of gene models or redefining exon-intron boundaries if necessary. The final call sets comprises 12,212 (R. commune UK7), 13,151 (R. secalis) and 13,674 (R. agropyri) protein coding genes. In addition, 91 (+22 mitochondrial) tRNA-encoding genes are predicted for R. commune UK7, 95 (+22 mitochondrial) for R. secalis and 107 for R. agropyri using tRNAscan-SE [136]. The predicted protein sets were searched for highly conserved single (low) copy genes to assess the completeness of the genomic sequences and gene predictions. Orthologous genes to all 246 single copy genes were identified for all three proteomes by Blastp comparisons (eVal: 10⁻³) against the single-copy families from all 21 species available from the FUNYBASE [137]. In addition, all of the 248 core genes commonly present in higher euukaryotes (CEGs) could be identified by Blastp comparisons (eVal: 10⁻³) for two proteomes, whereas only one protein was missing in the R. agropyri proteome [138].

Identification of repetitive DNA elements

Determination of repeat sequences involved first the calculation of repeat families de novo followed by assessment of known repeat elements in a second step. Families of previously unknown interspersed repeat elements were identified by RepeatScout [139]. Repeat families were included when they comprised more than 10 repeats and when their consensus sequence length was longer than 50 bp. In addition, low complexity and simple sequence repeats were determined with the tools NSEG [140] and Tandem Repeats Finder [141], which are part of the RepeatScout procedure and remove them from the interspersed repeat library.

The RepBase database [142] was used to detect previously published families of transposable elements, pseudogenes and retroviruses. In order to determine the exact locations of the repetitive elements on the genome we used the RepBase library and the calculated library of interspersed repeat families as input for RepeatMasker [143]. RepeatMasker was also used to find and mask genomic regions of low complexity. We applied the automated classification tool TEmask [144] to categorize the predicted repeat sequences into the four main transposable element categories DNA transposon, long interspersed nuclear element (LINE), short interspersed nuclear element (SINE) and retrotransposon with long terminal repeats (LTRs).

Functional annotation of predicted open reading frames and data repositories

The protein coding genes were analyzed and functionally annotated using the PEDANT system [145]. Data sets are accessible at http://pedant.helmholtz-muenchen.de/genomes.jsp?category=fungal. The genome and annotation data were submitted to the European Nucleotide Archive (ENA, R. commune, isolate UK7: http://www.ebi.ac.uk/
Phylogenetic analyses

Sequence data for phylogenetic estimations of fungal relationships were generated using standard PCR protocols and primers that were (a) designed according to the 18S rDNA sequence from *R. commune*, isolate 788 (GenBank accession no. AY038583.1, [34]; Additional file 12: Table S7) or (b) described in a previous study [35]. After amplification and cloning into pJET (Thermo Scientific, Schwerte, Germany) Sanger sequencing (LG Genomics, Berlin) was done on the nearly full-length rDNA region (18S rDNA-ITS1-5.8 S rDNA-ITS2-28S rDNA) using five (*R. commune*, *R. secalis*, six (*R. agropyri*) or seven (*R. orthosporum*) overlapping primer pair combinations starting on the 5′ side with primer 18SrRNA1f and ending on the 3′ side with primer LR7 (Additional file 12: Table S7). The sequences coding for elongation factor 1-α (EF1-α) and the RNA polymerase II subunit RPB1 and RPB2 were extracted from the genome database.

To integrate the *Rhynchosporium* genus into the estimated fungal phylogeny [35], sequences coding for 18S rRNA, 28S rRNA, the ITS region including the 5.8S rDNA, *elongation factor 1-α (EF1-α)* and the RNA polymerase II subunit RPB1 and RPB2 were obtained from GenBank (genome projects, AFTOL data) for 12 taxa of the *Leotiomycetes* class (Additional file 13: Table S8). Sequences for each gene were aligned using the ClustalW algorithm in MEGA (v6.06). Alignments that included more distantly related fungal species. In a second phylogenetic estimation, a tree was generated that includes more distantly related fungal species. To this end, the amino acid sequences of EF1-α, RBP1 and RBP2 from *R. commune* were submitted to BLASTp searches [148] to identify the homologous genes in 17 different taxa (Additional file 13: Table S8). An RBP1 sequence from *B. fuckeliana* was not found. The combined and concatenated sequence data were fitted into a single EF1-α/RBP1/RPB2 matrix. No trimming was performed due to the availability of complete coding sequence for all proteins investigated. Amino acid sequences were aligned using the ClustalW algorithm in MEGA (v6.06). Thus, phylogenetic estimation was carried out under the LG substitution model [149] utilizing gamma distributed rates with invariant sites (G + I, four discrete gamma categories). Maximum Likelihood analyses took 500 bootstrap replications of heuristic searches into account (MLBS). Partial deletion was estimated for gaps and missing data and tree improvement were accomplished using NNI. The MEGA-estimated phylogeny was verified by MLBS estimation in PHYML using the LG method with four substitution rate classes. 250 bootstraps were performed and tree improvement used NNI and SPR.

To construct the phylogenetic tree of the NIPs the amino acid sequences of the mature proteins were used. For the phylogenetic trees of polyketide synthases and non-ribosomal peptide synthases, amino acid sequences from other mostly fungal taxa were obtained from the NCBI databases. Enzymatic domains were identified by BLASTp analysis using NCBI’s Conserved Domain Database [82] and aligned with MUSCLE [152]. Evolutionary analyses were done in MEGA (v6.06). The evolutionary model with gamma distribution in four rate classes. 250 bootstraps were done and tree improvement used NNI and subtree pruning and regrafting (SPR).

**In a second phylogenetic estimation**, a tree was generated that includes more distantly related fungal species. To this end, the amino acid sequences of EF1-α, RBP1 and RBP2 from *R. commune* were submitted to BLASTp searches [148] to identify the homologous genes in 17 different taxa (Additional file 13: Table S8). An RBP1 sequence from *B. fuckeliana* was not found. The combined and concatenated sequence data were fitted into a single EF1-α/RBP1/RPB2 matrix. No trimming was performed due to the availability of complete coding sequence for all proteins investigated. Amino acid sequences were aligned using the ClustalW algorithm in MEGA (v6.06). Thus, phylogenetic estimation was carried out under the LG substitution model [149] utilizing gamma distributed rates with invariant sites (G + I, four discrete gamma categories). Maximum Likelihood analyses took 500 bootstrap replications of heuristic searches into account (MLBS). Partial deletion was estimated for gaps and missing data and tree improvement were accomplished using NNI. The MEGA-estimated phylogeny was verified by MLBS estimation in PHYML using the LG method with four substitution rate classes. 250 bootstraps were performed and tree improvement used NNI and SPR.

To integrate the *Rhynchosporium* genus into the estimated fungal phylogeny [35], sequences coding for 18S rRNA, 28S rRNA, the ITS region including the 5.8S rDNA, *elongation factor 1-α (EF1-α)* and the RNA polymerase II subunit RPB1 and RPB2 were obtained from GenBank (genome projects, AFTOL data) for 12 taxa of the *Leotiomycetes* class (Additional file 13: Table S8). Sequences for each gene were aligned using the ClustalW algorithm in MEGA (v6.06). Alignments that included more distantly related fungal species. In a second phylogenetic estimation, a tree was generated that includes more distantly related fungal species. To this end, the amino acid sequences of EF1-α, RBP1 and RBP2 from *R. commune* were submitted to BLASTp searches [148] to identify the homologous genes in 17 different taxa (Additional file 13: Table S8). An RBP1 sequence from *B. fuckeliana* was not found. The combined and concatenated sequence data were fitted into a single EF1-α/RBP1/RPB2 matrix. No trimming was performed due to the availability of complete coding sequence for all proteins investigated. Amino acid sequences were aligned using the ClustalW algorithm in MEGA (v6.06). Thus, phylogenetic estimation was carried out under the LG substitution model [149] utilizing gamma distributed rates with invariant sites (G + I, four discrete gamma categories). Maximum Likelihood analyses took 500 bootstrap replications of heuristic searches into account (MLBS). Partial deletion was estimated for gaps and missing data and tree improvement were accomplished using NNI. The MEGA-estimated phylogeny was verified by MLBS estimation in PHYML using the LG method with four substitution rate classes. 250 bootstraps were performed and tree improvement used NNI and SPR.
history was inferred by using the Maximum Likelihood method based on the LG substitution model [149]. In each case the tree with the highest log likelihood is shown. Initial trees for the heuristic search were obtained by applying the Neighbor-Joining-Interchange (NNI) method to a matrix of pairwise distances estimated using a JTT model. To model evolutionary rate differences among sites Gamma distribution (four discrete categories) with Invariant sites (G + I) was used. All positions with <95% site coverage were eliminated, i.e., <5% alignment gaps, missing data and ambiguous amino acids were allowed at any position. 500 bootstraps were done to assess the robustness of the phylogeny.

Identification of CWDEs

Genes coding for enzymes with carbohydrate substrates were identified using the CAZy database [51, 52]. CWDEs belong to CAZy families GH, CE, PL and AA. Therefore, all members of these families were subjected to secretion analysis by using the SecretomeP algorithm [55]. A neural network score of ≥0.6 was taken to predict secreted CAZymes, i.e. CWDEs.

Identification of gene clusters

To predict putative secondary metabolism gene clusters in silico two different algorithms were used, the Secondary Metabolite Unique Regions Finder (SMURF, [153]) and the Motif Density Method (MDM, [154]). With SMURF predictions are based on PFAM and TIGRFAM domain content and chromosomal position of a gene. MDM uses the density of occurrences of cluster-specific regulatory motifs in the promoter regions as an additional criterion to distinguish cluster from non-cluster genes. Application of the two algorithms to the Rhynchosporium BCG genomes yielded putative PKS gene clusters with varying degrees of overlap.

Identification of effector candidates

Candidate effector genes were identified by screening the genome for gene models whose products fulfill the criteria (1) secretion, (2) small size, (3) high cysteine content [117]. The TargetP 1.1 [155] and WoLF PSORT [156] protein localization algorithms (NN-based SP score >0.5, prediction confidence from RC = 1 to RC = 3) and transmembrane helix prediction (TMHMM 2.0, TMH = 0; [157]) were used to predict secreted proteins. Sorting of the candidates obtained according to length and cysteine content was done to yield proteins with ≤200 amino acids and ≥2% cysteine residues. The nucleotide sequences of these genes were used as queries in a BLASTn analysis to identify candidates, for which homologous sequences do not exist in the genomes of the other species. In addition, OrthoMCL analysis [118] was carried out on the three BCG genomes for orthology detection. Those proteins that lack any orthologs and paralogs were compared to the ≤200-amino acid protein fraction. Finally, using the single-gene sequences coding for small, cysteine-rich proteins as a query, BLASTn search of the other genomes was done to identify genes that are unique for a given genome.

EST libraries

Three EST libraries were generated from R. commune mycelia, from germinated conidia and from epidermal strips of inoculated barley leaves. Mycelia of isolate UK7 were grown for 14 d in liquid culture, subsequently filtered, air-dried and ground to a fine powder using glass beads (acid washed, 150–212 μm, SIGMA) to efficiently disrupt fungal hyphae [30]. Total RNA was extracted from mycelia using the Trizol method [158], then treated with DNase (Roche), quality-controlled on a 1.2% agarose gel and shipped for cDNA synthesis using the SMART technology [159]; Evrogen Technologies, Moscow, Russia). After normalization using the duplex-specific nuclease technology [160] the cDNA was 454-sequenced, assembled and trimmed (GATC Biotech, Konstanz, Germany).

Conidia of isolate 214 were incubated in sterile distilled water for 24 h and mRNA was extracted from germinated conidia using Dynabeads Oligo (dT)25 (Invitrogen Dynal AS, Oslo, Norway). RNA yield was measured using a NanoDrop Micro Photometer (NanoDrop Technologies, Inc., Rockland, USA). Prior to cDNA synthesis, RNA samples were treated with DNasel using the Ambion DNA-free kit following the manufacturer’s protocol. First strand cDNA for sequencing was synthesized from 0.5 μg of mRNA and PCR-amplified for 11 cycles using the SMART PCR cDNA synthesis kit (Clontech Laboratories/Takara Bio Europe SAS, Saint-Germain-en-Laye, France) following the manufacturer’s protocol. The non-normalized cDNA library was 454-sequenced, trimmed and assembled at the University of Liverpool, UK.

Barley plants were inoculated with spores of UK isolate 2LA. At 3 dpi epidermal strips were prepared and total RNA was extracted using a RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the protocol supplied by the manufacturer. After testing RNA integrity by agarose gel electrophoresis, total RNA was sent to the University of Exeter, UK, for mRNA isolation, illumina sequencing, trimming and assembling. Finally, all ESTs were mapped on the R. commune genome.

Quantitative PCR

Relative fungal biomass and fungal gene expression during pathogenesis were quantified by qPCR with DNA as template and by qRT-PCR with cDNA as template using the efficiency calibrated model [119] as described
mold disease on many plant species (B. fuckeliana) and the Phacidiopycnis post-harvest fruit rot of pear (P. pyri), respectively. Numerals on the nodes represent the percentages from 500 bootstraps. Scale: number of substitutions per nucleotide (PPTX 50 kb)

Additional file 2: Table S1. SNPs in organism pairs. (XLSX 11 kb)

Additional file 3: Table S2. (A) Core meiotic recombination machinery genes. (B) Genes involved in sexual reproduction in the aspergilli. (XLSX 43 kb)

Additional file 4: Figure S2. CAZyme numbers in different fungal species. GH, glycosyl hydrolases; CE, carbohydrate esterases; PL, pectate lyases; GT, glycosyl transferases; CBM, carbohydrate binding motifs (modified after [50]). (PPTX 72 kb)

Additional file 5: Table S3. CAZymes. (XLSX 24 kb)

Additional file 6: Table S4. (A) Cell wall degrading enzymes; (B) CWDE numbers according to their substrates. (XLSX 72 kb)

Additional file 7: Table S5. Secondary metabolism gene clusters. (XLSX 43 kb)

Additional file 8: Table S6. RcSP homologs (gene IDs). (XLSX 12 kb)

Additional file 9: Figure S3. Growth acceleration of single deletion mutants. Relative biomass of fungal deletion mutants and wild-type isolate UK7 were determined by qPCR at 14 dpi on barley cv. 'Ingrid'. Wild-type references are associated to the mutants to the right. Bars represent the 95% confidence intervals. n-values are given above bars. (PDF 14 kb)

Additional file 10: Figure S4. Growth of DRcSP9 mutants. Primary leaves of barley cv. 'Ingrid' were inoculated with spores of wild-type isolate UK7 or of the mutants and photos were taken at indicated times post inoculation. C, mock inoculation. (PPTX 145 kb)

Additional file 11: Figure S5. Disease phenotype of DRcSP9 mutants. Bars represent 95% confidence intervals. n

Additional file 12: Table S7. Primers used in this study. (DOCX 17 kb)

Additional file 13: Table S8. Genbank and internal numbers (Rhynchosporum) for taxa used in this study. (XLSX 13 kb)

Acknowledgements
The authors would like to thank B. McDonald, ETH Zurich, for providing the Swiss fungal isolates and for critically reading the manuscript.

Funding
This work was funded by the Leibniz Association in the framework of the Competition Committee’s competitive procedure and by the Scottish Government Rural and Environment Science and Analytical Services (RESAS).

Availability of data and materials
Datasets used for phylogenetic analyses are available from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.qm130). Sequence data sets are accessible at http://pedant.helmholtz-muenchen.de/genomes.jsp?category=fungal. The genome and annotation data were submitted to the European Nucleotide Archive (ENA) R. commune UK7: http://www.ebi.ac.uk/ena/data/view/FJUX01000001-FJUX01000164; R. sectalis: http://www.ebi.ac.uk/ena/data/view/FXC01000001-FXC01002130; R. agropyri: http://www.ebi.ac.uk/ena/data/view/FJUX01000001-FJUX01005058.

Abbreviations
BCG: Beaked conidia group; BLAST: Basic local alignment search tool; CCAZyme: Carbohydrate-active enzyme; CBM: Carbohydrate-binding motif; CCG: Cylindrical conidia group; CEGMA: Core Eukaryotic genes mapping approach; CWDE: Cell wall degrading enzymes; DMATs: Dimethylallyl tryptophane synthase; EAS: Eusaucomesites; EST: Expressed sequence tag; ETF1-a: Eukaryotic translation termination factor 1-a; GPD: Glyceraldehyde-3-phosphate dehydrogenase; MAT: Mating type; NIP: Necrosis inducing protein; NLP: NIP2-like protein; NPSL: NRPS-like protein; NRPS: Non-ribosomal peptide synthetase; PSSK: Polyketide synthetase; SCPCh: Single-copy protein-coding homolog; TC: Terpene cyclase

Additional file 1: Figure S1. The Leotiomycetes class of Ascomycetes. The concatenated nucleotide sequences of 185 rDNA, 285 rDNA, ITS region, elongation factor EF-1-a and RNA polymerase II subunits RPB1 and RPB2 from Rhynchosporum and twelve Leotiomycetes species were used to construct the phylogenetic tree. An enlargement of the Rhynchosporum subtree is shown in Fig. 1. Most of the species live a saprobic lifestyle (S). Three are plant pathogens (P) causing brown rot blossom blight disease on stone fruit and occasionally on pome fruit trees (M. laxa), grey mold disease on many plant species (B. fuckeliana) and the Phacidiopycnis post-harvest fruit rot of pear (P. pyri), respectively. Numerals on the nodes represent the percentages from 500 bootstraps. Scale: number of substitutions per nucleotide (PPTX 50 kb)

Additional file 2: Table S1. SNPs in organism pairs. (XLSX 11 kb)

Additional file 3: Table S2. (A) Core meiotic recombination machinery genes. (B) Genes involved in sexual reproduction in the aspergilli. (XLSX 43 kb)

Additional file 4: Figure S2. CAZyme numbers in different fungal species. GH, glycosyl hydrolases; CE, carbohydrate esterases; PL, pectate lyases; GT, glycosyl transferases; CBM, carbohydrate binding motifs (modified after [50]). (PPTX 72 kb)

Additional file 5: Table S3. CAZymes. (XLSX 24 kb)

Additional file 6: Table S4. (A) Cell wall degrading enzymes; (B) CWDE numbers according to their substrates. (XLSX 72 kb)

Additional file 7: Table S5. Secondary metabolism gene clusters. (XLSX 43 kb)

Additional file 8: Table S6. RcSP homologs (gene IDs). (XLSX 12 kb)

Additional file 9: Figure S3. Growth acceleration of single deletion mutants. Relative biomass of fungal deletion mutants and wild-type isolate UK7 were determined by qPCR at 14 dpi on barley cv. 'Ingrid'. Wild-type references are associated to the mutants to the right. Bars represent the 95% confidence intervals. n-values are given above bars. (PDF 14 kb)

Additional file 10: Figure S4. Growth of DRcSP9 mutants. Primary leaves of barley cv. 'Ingrid' were inoculated with spores of wild-type isolate UK7 or of the mutants and photos were taken at indicated times post inoculation. C, mock inoculation. (PPTX 145 kb)

Additional file 11: Figure S5. Disease phenotype of DRcSP9 mutants. Bars represent 95% confidence intervals. n

Additional file 12: Table S7. Primers used in this study. (DOCX 17 kb)

Additional file 13: Table S8. Genbank and internal numbers (Rhynchosporum) for taxa used in this study. (XLSX 13 kb)

Acknowledgements
The authors would like to thank B. McDonald, ETH Zurich, for providing the Swiss fungal isolates and for critically reading the manuscript.

Funding
This work was funded by the Leibniz Association in the framework of the Competition Committee’s competitive procedure and by the Scottish Government Rural and Environment Science and Analytical Services (RESAS).

Availability of data and materials
Datasets used for phylogenetic analyses are available from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.qm130). Sequence data sets are accessible at http://pedant.helmholtz-muenchen.de/genomes.jsp?category=fungal. The genome and annotation data were submitted to the European Nucleotide Archive (ENA) R. commune UK7: http://www.ebi.ac.uk/ena/data/view/FJUX01000001-FJUX01000164; R. sectalis: http://www.ebi.ac.uk/ena/data/view/FXC01000001-FXC01002130; R. agropyri: http://www.ebi.ac.uk/ena/data/view/FJUX01000001-FJUX01005058.

Abbreviations
BCG: Beaked conidia group; BLAST: Basic local alignment search tool; CCAZyme: Carbohydrate-active enzyme; CBM: Carbohydrate-binding motif; CCG: Cylindrical conidia group; CEGMA: Core Eukaryotic genes mapping approach; CWDE: Cell wall degrading enzymes; DMATs: Dimethylallyl tryptophane synthase; EAS: Eusaucomesites; EST: Expressed sequence tag; ETF1-a: Eukaryotic translation termination factor 1-a; GPD: Glyceraldehyde-3-phosphate dehydrogenase; MAT: Mating type; NIP: Necrosis inducing protein; NLP: NIP2-like protein; NPSL: NRPS-like protein; NRPS: Non-ribosomal peptide synthetase; PSSK: Polyketide synthetase; RcSP: Rhynchosporum; SCPCh: Single-copy protein-coding homolog; TC: Terpene cyclase

Additional file 1: Figure S1. The Leotiomycetes class of Ascomycetes. The concatenated nucleotide sequences of 185 rDNA, 285 rDNA, ITS region, elongation factor EF-1-a and RNA polymerase II subunits RPB1 and RPB2 from Rhynchosporum and twelve Leotiomycetes species were used to construct the phylogenetic tree. An enlargement of the Rhynchosporum subtree is shown in Fig. 1. Most of the species live a saprobic lifestyle (S). Three are plant pathogens (P) causing brown rot blossom blight disease on stone fruit and occasionally on pome fruit trees (M. laxa), grey
Authors’ contribution
MF, ST and MP performed the genome sequencing and assembly except for isolate 13-13. KP, KA and AA sequenced and assembled the 13-13 genome and transcriptome and made these data available. MM and UG conducted the genome annotation and all large-scale genome analyses. DP and SK carried out general lab work, inoculation experiments and phylogeny analyses and provided the R-cSP gene expression data. In addition, DP participated in gene annotation and collected the NP2 gene data. CW and WK carried out the bioinformatics analyses of secondary metabolism enzymes. TW and ES carried out the SMURF- and MDM-based gene cluster analyses. JG, JH, NW and SK generated and characterized the R-cSP deletion mutants. KMK, RH, DJH and BDLF made the R. loli data available. DP was involved in drafting the manuscript, which was mainly written by WK with contributions from AA and MM. All authors provided intellectual input, all principal investigators read and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval
Not applicable.

Author details
1Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Halle/Saale, Germany. 2Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany. 3Genomic Analysis, Leibniz Institute on Ageing, Fritz Lipmann Institute, Jena, Germany. 4Institute of Medical Genetics, Cardiff University, Cardiff, UK. 5Exeter Sequencing Service, Biosciences, University of Exeter, Exeter, UK. 6NIAB EMR, East Malling, UK. 7Systems Biology and Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Jena, Germany. 8Biological Chemistry and Crop Protection, Rothamsted Research, Harpenden, Hertfordshire, UK. 9Systems Biology and Bioinformatics, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Jena, Germany. 10Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland. 11Present address: Food Quality and Nutrition, Agroscope, Bern, Switzerland.

Received: 7 July 2016 Accepted: 15 November 2016

Published online: 22 November 2016

References
1. Caldwell RM. Rhynchosporium scald of barley, rye, and other grasses. J Agric. Res. 1937;55(3):175–98.
2. Oudemans CAJA. Observations mycologiques. Verslag Wis- en Natuurk Afd K Acad Wetensch Amsterdam. 1897;68:86–92.
3. Frank AB. Über die Zerstörung der Gerste durch einen neuen Getreidepilz. 1907;82:233–52.
4. Rehman MA, Gierlich A, Hermann H, Hahn M, Schmidt B, Rosahl S, Knogge W. The race-specific elicitor, NIP1, from the barley pathogen, Rhynchosporium secalis, determines avirulence on host plants of the Rc1 resistance genotype. EMBO J. 1995;14(17):4168–77.
5. Brunner PC, Schürch S, McDonald BA. The origin and colonization history of the barley scald pathogen Rhynchosporium secalis. J Evol Biol. 2007;20(4):1311–21.
Ramesh MA, Malik SB, Logsdon JM. A phylogenomic inventory of meiotic genes encoding sex determining properties and discovering C4 origins. New Phytol. 2012;193(2):304–12.

Weltly RE, Metzger RJ. First report of scald of triticale caused by Rhynchosporium secalis in North America. Plant Dis. 1996;80(11):1220–3.

McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol. 2002;40:349–79.

Debuchi R, Turgeon BG. Maturing-type structure, evolution, and function in Eusarcymycetes. In: Kues U, Fischer R, editors. The Mycota I: growth, differentiation and sexuality. Berlin: Springer; 2006. p. 293–323.

Rydholm C, Dyer PS, Lutzeni F. DNA sequence characterization and molecular evolution of MAT1 and MAT2 mating-type loci of the self-compatible ascomycete mold Neosartorya fischeri. Eukaryot Cell. 2007;6(5):868–76.

Poegeleir S, O’Gorman CM, Hoff B, Kueck U. Molecular organization of the mating-type loci in the homothallic Ascomycete Eupenicillium crustaceum. Fungal Biol. 2011;115:7615–24.

Hughes TJ, O’Donnell K, Sirk S, Rooney AP, Mercedes Scandlani M, Luque A, Bhattacharyya MK, Huang X. Genetic architecture and evolution of the mating type locus in fusaria that cause soybean sudden death syndrome. Mycologia. 2014;106(4):686–704.

Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomes. Nucleic Acids Res. 2009;37:D233–8.

Lombard V, Ramulu HG, Druela E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(D1):D490–5.

Nickel W, Seedorf M. Unconventional mechanisms of protein transport to the cell surface of Eukaryotic cells. Annu Rev Cell Dev Biol. 2008;24:287–308.

Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel. 2004;17(9):349–56.

Akcapinar GB, Kappel L, Sezerman OJ, Seidl-Seiboth V. Molecular diversity of LysM carbohydrate-binding motifs in fungi. Curr Genet. 2015;62(2):103–13.

de Jonge R, Thomma BPHJ. Fungal LysM effectors: exchangers of host immunity? Trends Microbiol. 2009;17(4):151–7.

Jiang C, He B, Huang R, Huang M, Xu L. Identification and functional analysis of LysM effectors from Marssonina brunnea. Australas Plant Pathol. 2014;43(6):615–22.

de Jonge R, van Esse HP, Kombrik A, Shinya T, Desaki Y, Bours R, van der Krool S, Shibuya N, Joosten MJAH, Thomma BPHJ. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science. 2010;329(5994):953–5.
85. Keller NP. Translating biosynthetic gene clusters into fungal armor and weaponry. Nat Chem Biol. 2015;11(9):671–7.
86. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. BMC Evol Biol. 2010;10:26.
87. Bushley KE, Ripoll DR, Turgeon BG. Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis. BMC Evol Biol. 2008;8:328.
88. Lee BN, Kroken S, Chou DYT, Robberst B, Yoder OC, Turgeon BG. Functional analysis of all nonribosomal peptide synthetases in Cochliobolus hetero-/sphericus reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. Eukaryot Cell. 2005;4(3):545–55.
89. Eisendle M, Schertel M, Krogf C, Muller D, Ilmer P, Haas H. The intracellular siderophore ferriconin is involved in iron storage, oxidative-stress resistance, germination, and sexual development in Aspergillus nidulans. Eukaryot Cell. 2006;5(10):1596–603.
90. Oide S, Krasnoff SB, Gibson DM, Turgeon BG. Intracellular siderophores are essential for ascomycete sexual development in heterothallic Cochliobolus heterostrophus and homothallic Gibberella zeae. Eukaryot Cell. 2007;6(8):1339–53.
91. Schrêdl M, Willger SD, Park S-W, Puttikomonkul S, Grahl N, Cho Y, Roberts B, Yoder OC, Turgeon BG. Module evolution and substrate specificity of fungal nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycete. Plant Cell. 2006;18(10):2836–53.
92. Zabriskie TM, Jackson MD. Lysine biosynthesis and metabolism in fungi. Nat Prod Rep. 2000;17(1):85–97.
93. Xue H, Li G, Willger SD, Park S-W, Puttikomonkul S, Grahl N, Cho Y, Roberts B, Yoder OC, Turgeon BG. Module evolution and substrate specificity of fungal nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycete. Plant Cell. 2006;18(10):2836–53.
94. Fisch KM. Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS hybrids. ChemBioChem. 2013;14(11):28–42.
95. Fisch KM. Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS. RCS Adv. 2013;3(40):18228–47.
96. Keller NP, Hohn TM. Metabolic pathway genes in filamentous fungi. Fungal Genet Biol. 1997;21(1):17–29.
97. Chiang Y-M, Szewczyk E, Davidson AD, Keller N, Oakley BR, Wang CCG. A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranein, in Aspergillus nidulans. J Am Chem Soc. 2009;131(11):2965–70.
98. Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. Emerging concepts in effector biology of plant-associated organisms. Mol Plant-Microbe Interact. 2009;22(1):115–22.
99. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
100. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
101. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
102. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
103. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
104. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
105. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
106. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
107. Raffil M, Jelonek L, Akum NF, Zhang F, Kogel KH. Effector candidates in the secretome of Peronospora indica, a ubiquitous plant-associated fungus. Front Plant Sci. 2013;4:228.
