Mycoscience (2006) 47:55–64
DOI 10.1007/s10267-006-0279-7

© The Mycological Society of Japan and Springer-Verlag Tokyo 2006

REVIEW

**Irina S. Druzhinina · Alex G. Kopchinskiy
Christian P. Kubicek**

The first 100 *Trichoderma* species characterized by molecular data

Received: November 14, 2005 / Accepted: January 28, 2006

Abstract *Trichoderma* species are generally abundant on decaying wood and in soil because of their success in various heterotrophic interactions, including decomposition, parasitism, and even opportunistic endophytism. Many *Trichoderma* species or, precisely, many individual *Trichoderma* strains, have various important applications in industry and human life, which led to the inclusion of *Hypocrea jecorina* (*Trichoderma reesei*), the well-known producer of industrial enzymes, in the list of organisms whose genomes have been sequenced. *Trichoderma* species also have been adopted as agents of biological control of plant pathogenic fungi and as antibiotic producers. *Trichoderma longibrachiatum* is known as an opportunistic pathogen of immunocompromised mammals, including humans, and some species are common indoor contaminants. Given these properties, correct identification at the species level is highly desirable. However, within the past decade, the number of recognized *Trichoderma* species has tripled, reaching 100. Therefore, *Trichoderma* taxonomy and species identification is a difficult issue. The abundant homoplasy in phenetic characters is likely the reason, given that the number of morphologically distinct species is significantly lower than the number of phylogenetically distinct species recognized using methods of gene sequence analysis. In this review, we introduce to the scientific community the development of modern tools for *Trichoderma* species identification: the oligonucleotide barcode program *TrichO*KEY version 1.0, and *Tricho*BLAST, the multilocus database of vouchered sequences powered by a similarity search tool. We also discuss the application of the Genealogic Concordance Phylogenetic Species Recognition approach. In combination, these advances make it possible to identify all known *Trichoderma* species based on sequence analysis.

I.S. Druzhinina (✉) · A.G. Kopchinskiy · C.P. Kubicek
Research Area of Gene Technology and Applied Biochemistry, Institute of Chemical Engineering (ICE), Vienna University of Technology, Getreidemarkt 9/1665, A-1060 Vienna, Austria
Tel. +43-1-58801-17202; Fax +43-1-58801-17299
e-mail: druzhini@mail.zserv.tuwien.ac.at

Introduction

Fungi are different. In many botany courses they are still introduced in the sense of Linnaeus's *Cryptogams* (Systema Naturae; Linnaeus, 1703) as a group closely related to green plants. Moreover, many mycologists have a botanical education. However, recent phylogenetic reconstructions, such as the Tree of Life project based on nuclear small subunit ribosomal DNA (nSSU rDNA), put fungi and animals as sister clades (Heckman et al. 2001; Taylor et al. 2004). The confusion at such a high taxonomic level may exemplify the general difficulty in inferring an adequate fungal taxonomy based on phenetic data. All fungi are heterotrophs, but unlike animals, they live in their food. Probably evolutionary adaptations to such a specific nutritional strategy led to the development of abundant homoplasious morphs; in other words, fungi seem to be significantly more diverse genetically than can be observed solely from phenetic characters. The gradual introduction of molecular data into the systematics of fungi, which started in the 1980s and became a standard by the late 1990s, has thus boosted research on the diversity and speciation of those fungal groups that scientists have studied for centuries. In this review, we focus on a single genus of the ascomycetes (the mitosporic genus *Trichoderma*; Hypocreales, Ascomycota) to illustrate the current trends in fungal taxonomy based on molecular data. Our aim is to describe new methods of species identification in light of modern approaches to species recognition. We discuss their advantages, requirements, pitfalls, and further developments.

Why was the genome of *Trichoderma* sequenced?

The complete genome sequence of one species of *Trichoderma* (*Hypocrea jecorina/Trichoderma reesei*) was released to the public early in 2005 (http://gsphere.lanl.gov/trire1/trire1.home.html). This species was selected for genome sequencing because of its significance for the production of

Table 1. Distribution of *Hypocrea/Trichoderma* sequences in NCBI GenBank

	Molecularly characterized species	Unpublished species	Unaccepted species names	Tentatively identified (confer)	Undefined strains	Species complexes	Total number of records
Number of records in NCBI taxonomy browser/Number of core nucleotide sequences							
Trichoderma	23/929	7/14	7/62	10/20	246/303	1/144	300/1472
Hypocrea	62/842	8/16	2/5	11/13	19/27	–	103/903[a]

[a]This number does not include 554 core nucleotide sequences for *H. lixii* and 3836 for *H. jecorina*. For the latter species, 2407 sequences were obtained from the cellulase overproducing mutant NRRL 6156 (=QM 9414)

industrial enzymes and recombinant proteins (Kubicek and Penttilä 1998; Penttilä 1998). *Trichoderma* species are generally abundant in nature, frequently found on decaying wood and in soil (Samuels 1996; Klein and Eveleigh 1998), where its individual genet can comprise a major portion of the total fungal biomass (Danielson and Davey 1973; Nelson 1982; Widden and Abitbol 1980). This abundance results from its success in diverse heterotrophic interactions including decomposition, parasitism, and even opportunistic endophytism (Harman et al. 2004). Many *Trichoderma* species or, precisely, many individual *Trichoderma* strains have various important applications in industry and human life. Several of them have been adopted as agents of biological control of plant pathogenic fungi (i.e., *Trichoderma harzianum*, *T. atroviride*, *T. asperellum*; Hjeljord and Tronsmo 1998) and as antibiotic producers (Sivasithamparam and Ghisalberti 1998). *Trichoderma longibrachiatum* is known as an opportunistic pathogen of immunocompromised mammals including humans (Kredics et al. 2003), and some species are common indoor contaminants (Thrane et al. 2001). These diverse impacts of *Trichoderma* on human life and sustainable development of natural ecosystems are reflected in the fact that additional species of the genus (*T. virens, T. atroviride*) are in the pipeline for genome sequencing, which, through comparative genomics, will enable a general insight into this genus.

From 1 to 100 *Trichoderma* species

All these important properties make correct identification of unknown isolates of *Trichoderma* highly desirable. Printed scientific journals and monographs are the most reliable sources of information on fungal taxonomy. However, because the publishing process is still relatively time consuming, online interactive databases powered with search engines are ideally suited for gathering cumulative information from various origins. Several such sources attempt to maintain the current taxonomy of Ascomycota. The state of the art of higher ranks is presented electronically in MYCONET (www.umu.se/myconet/myconet6.html) edited by O.E. Ericsson. According to this source, *Trichoderma* combines anamorphic (mitosporic) fungi of genus *Hypocrea* belonging to the Hypocreaceae of the Hypocreales within the class Sordariomycetes. The ground-level taxonomy for the majority of published names, i.e., taxonomy at the level of species, is available in Index Fungorum (www.indexfungorum.org/Names/Names.asp). The search for *Hypocrea* and *Trichoderma* resulted in 427 and 110 species epithets, respectively (October 16, 2005). However, among the latter number, 40% (44 names) are either outdated (synonyms) or have an unknown identity. The NCBI Taxonomy browser contains the names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence. Unfortunately, this source also provides an unclear picture of *Trichoderma* taxonomy (Table 1). It is interesting to note that although the total number of sequences deposited for *Hypocrea* holomorphs is similar to the number of sequences for mitosporic *Trichoderma*, the latter group comprises only one-third as many species (23 compared to 62). At the same time, one can see that the number of unidentified *Trichoderma* records (*Trichoderma* spp.) is one order of magnitude higher than that for *Hypocrea*. It clearly exemplifies the difficulty in the identification of *Trichoderma* species, which leads to the fact that researchers prefer to submit sequences of certain strains as *Trichoderma* sp. where identifications are problematical. According to our preliminary estimates of the *Hypocrea* and *Trichoderma* records in GenBank, approximately 40% are either unidentified or misidentified at the species level. To understand the reasons for this, we briefly discuss the development of *Trichoderma* taxonomy.[1]

The generic name *Trichoderma* was introduced more than 200 years ago by Persoon (1794) on the basis of material collected in Germany (Fig. 1). He included four species in the genus, but only one, *Trichoderma viride*, actually proved to be *Trichoderma* based on subsequent investigations. Seventy years later Tulasne and Tulasne (1860) determined that *T. viride* is the asexual stage of *Hypocrea rufa*. The fact that this relationship is still true today renders their finding a milestone in the taxonomy of conidial fungi. Continuing into the early 20th century, additional species of *Trichoderma* were described, but because most of these isolates have not been maintained, their identity is doubtful. Many of those isolates that were kept turned out either to be identical to a described species or not to belong to *Trichoderma*. For example, Abbott (1927) recognized four

[1]The complete taxonomy of *Hypocrea* was not considered in this manuscript because it is still partly unresolved. The review is therefore dedicated to the molecular taxonomy of *Trichoderma*

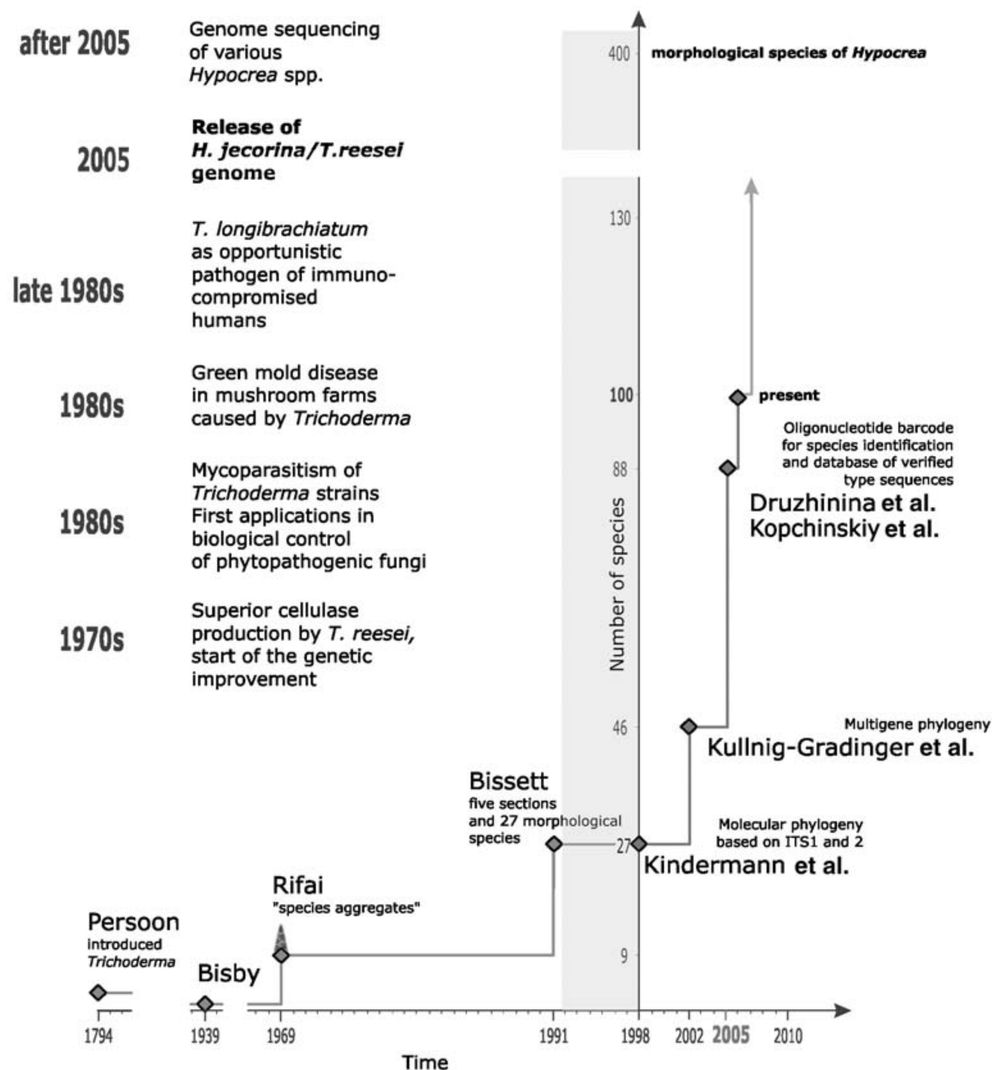

Fig. 1. Development of *Trichoderma* spp. taxonomy before and after methods of molecular phylogenetics were introduced. Only studies focusing on the entire genus are cited. *Gray line* indicates the anticipated growth in the number of recognized species in the near future

well-defined groups among the intergrading *Trichoderma* isolates from soil but distinguished only three species in a key; two of them are now considered to be synonyms of *T. viride.* Bisby (1939) therefore reduced all species of *Trichoderma* to the single species *T. viride*, a system that was followed until 1969. However, the high phenetic diversity of the one-species system prompted John Webster and Mein Rifai to review the taxonomy of *Trichoderma* and *Hypocrea* by examining life cycles of *Hypocrea* species (Rifai and Webster 1966a,b; Rifai 1969). They took the approach that the addition of characters from a teleomorph would help to define the *Trichoderma.* In a subsequent comprehensive monograph, Rifai (1969) recognized nine aggregate species, some of which were isolated from *Hypocrea* specimens. It is important to note that Rifai never considered his system to be complete, but emphasized that each of the "aggregate" species, particularly *T. hamatum*, could indeed contain different species which may be distinguishable once appropriate methods became available.

A first step toward this direction was done by Bissett (1984, 1991a–c, 1992). He viewed some of Rifai's aggregate species to be sections, and within those sections, Bissett recognized biological species. Indeed, many of Bissett's new species were based on cultures obtained as *T. hamatum*. He replaced the 9 aggregate species by formally recognizing four sections comprising 27 species (see Fig. 1).

With the availability of DNA analytics for fungal systematics, *Trichoderma* researchers quickly incorporated first restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD), and later on sequence analysis to the developing taxonomy of *Trichoderma* and *Hypocrea* (for review, see Lieckfeldt et al. 1998). Work conducted up to 2000, which included the identification of *T. reesei* as the anamorph of *H. jecorina*, the most important *Trichoderma* species industrially, and a revision of section *Longibrachiatum* (Kuhls et al. 1996, 1997), mainly used the internal transcribed spacer (ITS) regions of the ribosomal RNA-encoding genes, ITS1 and ITS2. Kindermann et al. (1998) attempted a first phylogeny of the whole genus based on ITS1 sequences. However, following the work on the *Gibberella fujikuroi* species complex (O'Donnell et al. 1998, 2000), researchers felt the need to put their phylogenies on the ground of analysis of several unlinked genes (Taylor et al. 2000). Kullnig-Gradinger et al.

(2002) pioneered with a phylogeny of all described *Trichoderma* species, based on sequence analysis of ITS1 and -2, the fifth intron of translation elongation factor 1-alpha (*tef1*), a partial exon of endochitinase 42 (*ech42*), and the small subunit of the mitochondrial rRNA-encoding gene. Their work raised the number of species distinguishable by molecular methods to 47. In the following years, the work of Chaverri and Samuels (Chaverri et al. 2003 a,b; Chaverri and Samuels 2004) on green-spored *Hypocrea* spp. and of Druzhinina and coworkers (Bissett et al. 2003; Kraus et al. 2004; Druzhinina et al. 2005; Jaklitsch et al. 2005) doubled the number of *Trichoderma* species.

Given that the majority of species were recognized in the "molecular era" of fungal taxonomy, the genus *Trichoderma* is exceptionally well documented by gene sequence data, as virtually every species is documented with diagnostic sequences from at least one to two genes (Fig. 2). When new species for which sequence data are deposited in GenBank but are not yet formally named are included in the enumeration of known diversity, and those of unaccepted species names are removed, 100 species of *Hypocrea/Trichoderma* have now been described. Still, we believe that the rapid rise in the number of phylogenetically distinct species is likely to continue because many of the new species were found in surveys of poorly sampled biogeographic areas (Siberia, South-East Asia, and Central and South America), whereas the investigation of areas such as Africa, Central Asia, and the Pacific are still missing. In addition to the significant number of "white spots" on the *Trichoderma* biogeographic map, the recent reconsideration of criteria for the recognition of species has led to the identification of new taxa. Many morphological species (e.g., sensu Bissett 1991a–c) have been shown to contain several cryptic phylogenetic species, and this will likely be the case for many other morphospecies not yet investigated in detail. However, it is important to note that the number of known *Trichoderma* "morphs" (morphological species) recognized by Bissett has not increased significantly, and they represent only one-third of the known phylogenetically distinct species. Therefore, it poses a challenge to those wishing to identify species of *Trichoderma*.

Identification of species based on the analysis of DNA sequences

The classical approach to identify fungi such as *Trichoderma* was based on differences in morphology and growth characters. However, because of the homoplasy of morphological characters, morphological species recognition is problematical even by specialists. Therefore, published studies on the ecology (Danielson and Davey 1973), enzyme production (Wey et al. 1994; Kovacs et al. 2004), biocontrol (Kullnig et al. 2000), human infection (Gautheret et al. 1995), and secondary metabolite formation (Cutler et al. 1999; Humphris et al. 2002) within *Trichoderma* are difficult to interpret. Consequently, almost all recent studies have used molecular data to characterize and identify species (Kullnig et al. 2000; Kubicek et al. 2003; Wuczkowski et al. 2003; Gherbawy et al. 2004; Chaverri and Samuels 2004). With the accumulation of sequence data in GenBank (see above; Fig. 2), using the most popular tool for this purpose, a NCBI similarity search tool (BLAST, Basic Local Alignment Search Tool; http://www.ncbi.nih.gov/BLAST), researchers are now theoretically able to identify all known species. Unfortunately this approach has several pitfalls because (a) the deposition of sequences within NCBI GenBank has not included a quality control of species identification; (b) some sequences are deposited under the name the species was originally obtained and not under the name it has been identified subsequently; (c) high similarity of sequences does not confirm species identity unless intraspecific variability of this sequence is known; and (d) even if it is known that a given species may show, e.g., 1% nucleotide (nt) variation, this may not apply to the entire sequence, and nts in some positions may be invariable.

To overcome these problems in the identification of *Trichoderma* by means of gene sequences, the subcommission for *Hypocrea/Trichoderma* taxonomy (ISTH) of the International Commission on the Taxonomy of Fungi (Mycological Division of IUMS) has initiated the development of automated methods for species identification in *Trichoderma* based on thoroughly revised and validated sequence data (www.isth.info). To this end, several tools for sequence analysis were developed and incorporated into the portal. The flowchart in Fig. 3 shows interconnections between various *Hypocrea/Trichoderma* sequence databases and programs for species identification. Below, we describe these tools individually, and explain the most optimal and up-to-date path through these tools, which should lead the user to a reliable identification of currently known *Hypocrea/Trichoderma* species.

*TrichO*Key: an oligonucleotide barcode for *Hypocrea* and *Trichoderma*

Druzhinina et al. (2005) developed the first fungal oligonucleotide barcode for the identification of *Hypocrea* and *Trichoderma* species, *TrichO*Key version 1.0. The program provides an online method for the quick molecular identification of an isolate at the genus, clade, and species levels based on a diagnostic combination of several oligonucleotides (hallmarks) specifically nested within the internal transcribed spacer 1 and 2 (ITS1 and -2) sequences of the rDNA repeat. The first version of the barcode was developed using 979 sequences of 88 vouchered species containing a total of 135 ITS1 and 2 haplotypes. Oligonucleotides, which were conserved in all known *Hypocrea*/*Trichoderma* ITS1 and ITS2 sequences but different in closely related fungal genera, were used to define genus-specific hallmarks. *TrichO*Key identifies most species unequivocally, although five species pairs and one triplet share identical ITS1 and ITS2 sequences and are therefore indistinguishable by the method.

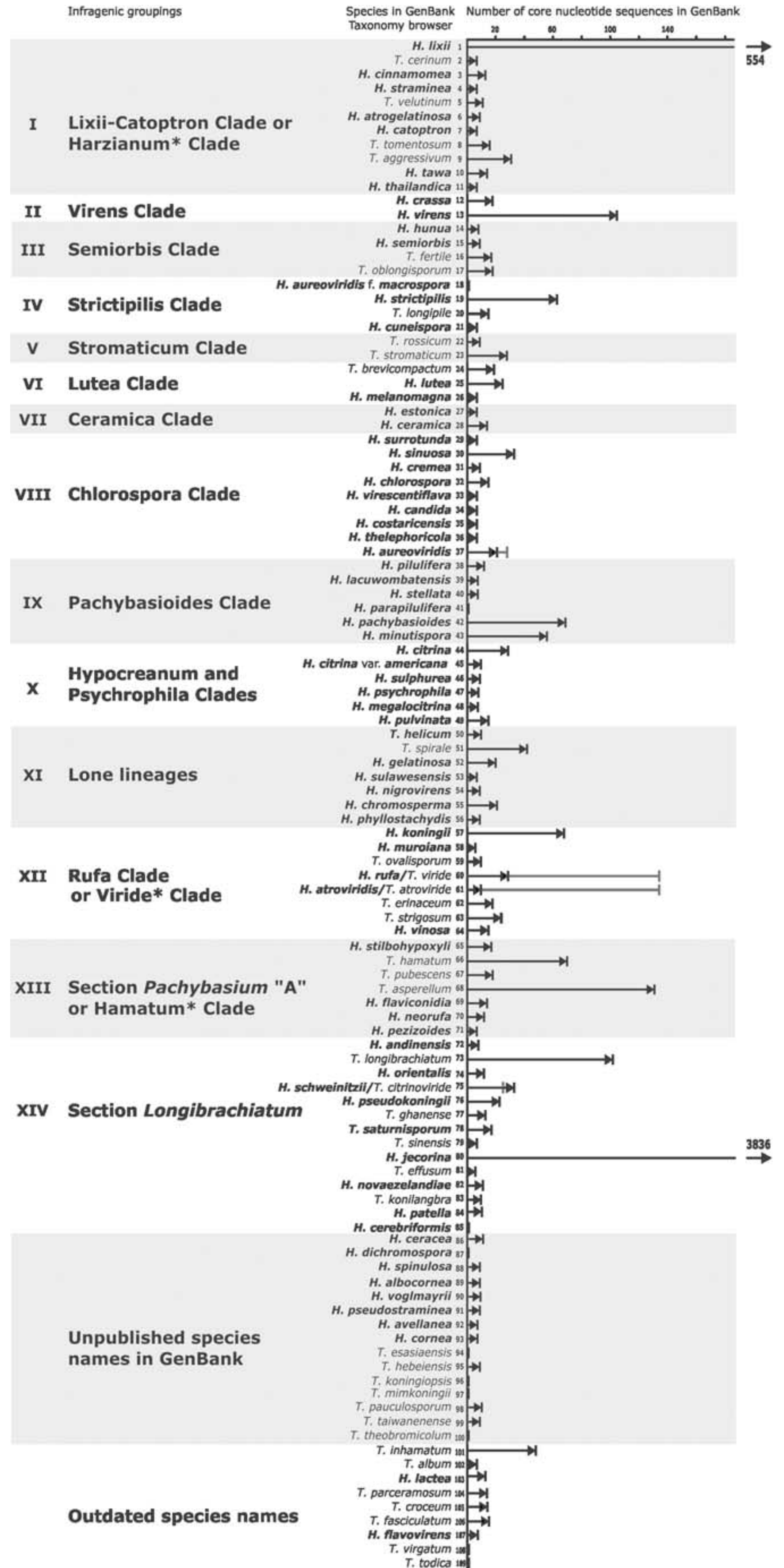

Fig. 2. *Hypocrea/Trichoderma* species names and number of their core gene sequences deposited in NCBI GenBank and Taxonomy browser. *, Updated clade names proposed by G.J. Samuels (personal communication)

*TrichO*Key version 1.0 also includes a library of species-, clade-, and genus-specific hallmarks, which enables the user to check the correctness of results of each of the steps: the detected hallmarks are displayed on both the query and the reference sequence by a barcode visualization module. The start page of the *TrichO*KEY version 1.0 user interface contains the link to the *Hypocrea/Trichoderma* biodiversity table, hallmark library, and the database of type sequences used for barcode development. The type sequence of the identified species can be retrieved from the database. Moreover, the ITS1 and ITS2 master alignment of *Hypocrea/Trichoderma* sequences is available on the same page.

The advantage of *TrichO*KEY version 1.0 is that it gives an unambiguous result, i.e., it can be applied even by researchers with only little experience in *Hypocrea/Trichoderma* taxonomy. The important point, which must be considered by the user, is the assigned level of identification reliability. For those species that are only known from a single strain, it is charged as "low." In contrast, barcodes that have been developed from the inspection of 20 or more sequences from a worldwide collection of isolates provide the most reliable identification (reliability level, "high"). However, a low reliability level does not necessarily negate the result. As an example, sequences for only three isolates of *T. oblongisporum* are present in our database, but one of these represents the *ex*-type (Siberia versus Canada; Kullnig et al. 2000). Moreover, the ITS sequences of all three strains share the same species diagnostic barcodes.

The lack of a reliable ITS barcode for *T. koningii/T. ovalisporum/H. muroiana* is a more serious shortcoming. Studies on the distribution of *Trichoderma* in different habitats (Kullnig et al. 2000; Kubicek et al. 2003; Wuczkowski et al. 2003; Druzhinina et al. 2005; Zhang et al. 2005) have frequently detected strains with the ITS sequence characteristic for these three species. Unfortunately, an accurate identification of species in the Rufa Clade (Viride Clade sensu Samuels et al., in manuscript) by other means (e.g., *tef1* analysis) is still hampered by the fact that species limits within this clade are under revision, and both *T. viride* and *T. koningii* appear to contain several cryptic species. Some of these species have the species-specific hallmarks in ITS1 and ITS2 sequences that are sufficient for barcode development, whereas others share the same allele of this locus. The observed diversity will be integrated in the subsequent version of *TrichO*KEY (G.J. Samuels and I. Druzhinina, unpublished data).

As already noted, the number of recognized *Trichoderma* species is constantly growing. Moreover, even the number of morphspecies within *Hypocrea* exceeds that for *Trichoderma* by fourfold. Therefore, it is reasonable to expect that new species or new ITS1 and ITS2 alleles of known taxa will be discovered. How does *Trich*OKEY handle unknown sequences? First, it is necessary to rule out that a lack of identification is not caused by a low sequence quality. To check this automatically, *TrichO*KEY version 1.0 was powered by a special module, which helps to differentiate between various reasons leading to the negative species identification (see Fig. 3). For instance, it distinguishes cases of truncated fragments when only separate ITS1 or ITS2 sequences were submitted. Moreover, if the query sequence is missing the most diagnostic hallmark area at the beginning of ITS1, the program gives the corresponding warning message and proposes several ways how to overcome the difficulty. In most cases, *TrichO*KEY version 1.0 also identifies the isolate at the subgenus level, i.e., assigns it to one of several defined clades within *Hypocrea/Trichoderma* (cf. Druzhinina and Kubicek 2005). Only if the unidentified query sequence has passed through all these filters, and it was assigned to *Hypocrea/Trichoderma*, then the hypothesis of an unknown species can be proposed and tested. In such a case, the next step in species identification should be the search for the next similar sequence and its comparison with the query. Therefore, *TrichO*KEY version 1.0 is supported by the link to the tool that implements all necessary options (see Fig. 3).

Tricho*BLAST: a sequence similarity search tool for *Hypocrea* and *Trichoderma

BLAST scripts (Altschul et al. 1997) available at NCBI are probably the most popular tools for identification of organisms based on sequence similarity. To eliminate the problems outlined above when using it for *Hypocrea/Trichoderma*, Kopchinskiy et al. (2005) developed *Tricho*BLAST, a publicly available database of vouchered sequences supported by sequence diagnosis and the similarity search tools, which includes all genetically characterized *Trichoderma* and *Hypocrea* species and contains almost complete sets of the five most frequently used phylogenetic markers: the internal transcribed spacers 1 and 2, ITS1 and ITS2; two introns [*tef1*_int4(large), *tef1*_int5(short)] and one exon [*tef1*_exon6(large)] of the gene encoding translation elongation factor 1-alpha (*tef1*), and a portion of the exon between the fifth and seventh eukaryotic conserved amino acid motives (Liu et al. 1999) of subunit 2 of the RNA polymerase gene (*rpb2*_exon). *Tricho*BLAST is also located on the ISTH website (www.isth.info), and is continuously updated by inclusion of new sequences of more species and/or loci and/or alleles, as they become available (see Fig. 3).

Because there is no consensus in the *Trichoderma* community about the primers to use for amplification and sequencing genes such as *tef1*, there is considerable variation in the length and location of the fragment within the gene of the sequences deposited in public databases under the same gene name. Consequently, the accuracy of a similarity search can be seriously corrupted: for example, if sequences of *tef1* containing both the short, highly variable intron and a long portion of the conserved exon are submitted to the similarity search, the "best hit" will be calculated based on the high score for exon–exon alignment while the intron–intron similarity will be neglected. To eliminate this obstacle *Tricho*BLAST is enforced by *Tricho*MARK (see Fig. 3), which enables the detection and retrieval of phylogenetic markers in query sequences and their subsequent indi-

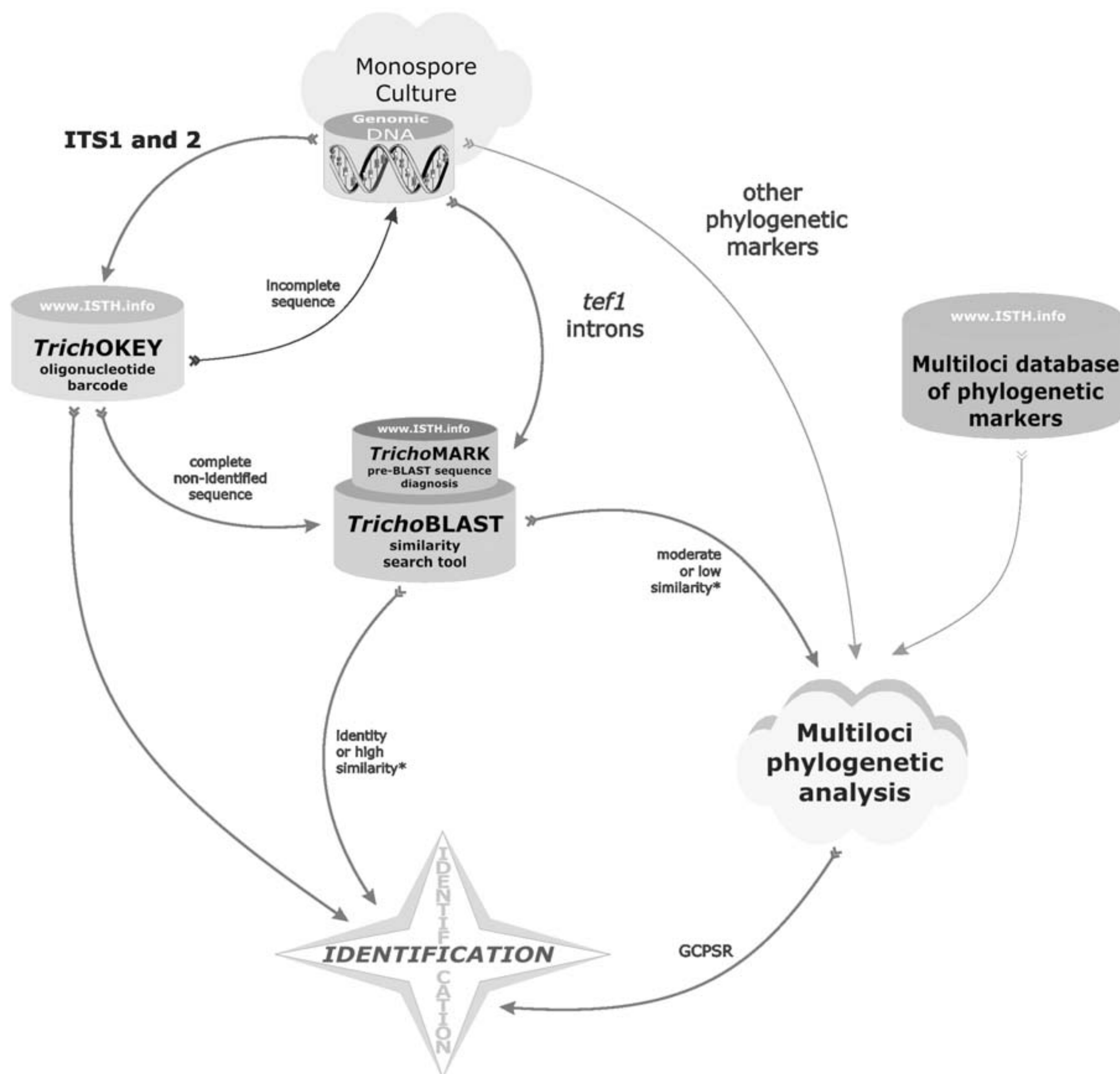

Fig. 3. Flowchart showing the optimal sequence of steps in the molecular identification of *Hypocrea/Trichoderma* species using tools available at the www.isth.info portal. The * indicates that the degree of similarity sufficient for species identification differs depending on the locus employed

vidual submission to the similarity search. The first version of *Tricho*MARK is able to diagnose ITS1 and ITS2 sequences as belonging to members of *Hypocrea/Trichoderma* based on genus-specific oligonucleotide sequences at the 5′- and 3′-end of this locus (Druzhinina et al. 2005) and retrieves the exact area of the ITS1 and ITS2 phylogenetic marker, excluding the flanking regions. In the case of the highly diagnostic *tef1* introns, *Tricho*MARK searches for conserved and genus-specific areas that flank the two introns, retrieves each intron individually for the similarity search, and also provides the results with a comparison of the actual and theoretically expected length of the detected phylogenetic marker. Similarly, the program scans for specific oligonucleotide stretches in the *tef1* and *rpb2* exons, respectively, to retrieve a fragment from the query sequence that exactly matches that of the corresponding phylogenetic marker in *Tricho*BLAST. Such pre-BLAST sequence diagnosis significantly increases the accuracy of the subsequent similarity search.

After this first step, the user can then automatically transmit the sequence to the similarity search and receive the results in the standard for BLAST way. However, a few general precautions with this approach must be stressed: first, similarity does not represent a sound measure of relatedness (de Queiroz 1992) and will depend on (i) which phylogenetic marker was used and (ii) whether *Tricho*BLAST contains the respective sequences for all known species. In this regard, we note that ITS1 and ITS2 are very diagnostic and are unique with the exception of the few cases mentioned earlier (all these exceptions are clearly noted in *Tricho*BLAST results). Cases where the ITS1 and ITS2 sequences do not exactly match any record in the

database, but differ by one or a few nucleotides, are either indicative of an unknown allele of a known species or very likely represent a new species. To confirm the species identification in such a case, a similarity search in *Tricho*BLAST should be done for other phylogenetic markers and/or the multiloci phylogenetic analysis. In the case of the two *tef1* introns, the situation is ambiguous, as sequence identities are rare because the high level of intraspecific variability within them (Chaverri et al. 2003a; Druzhinina et al. 2004).

Under the precautions given above and listed in Kopchinskiy et al. (2005), matches of lesser extent are indicative of relatedness at best; however, hypotheses of relationships should not be inferred directly from similarity indices in BLAST, but must be based on the application of phylogenetic inference methods instead (see following). To facilitate this, the third interactive module on www.isth.info was developed (see Fig. 3), i.e., a publicly available Multiloci Database of Phylogenetic Markers (MDPM; Kopchinskiy et al. 2005). This database contains only records exactly corresponding to five selected phylogenetic markers, which after manual retrieval and alignment with published sequences were trimmed to exactly fit the *Tricho*BLAST formats and thus can be used directly for phylogenetic analysis together with the sequence(s) in question.

Trichoderma species identification based on genealogical concordance

Despite the fact that the majority of *Trichoderma* isolates are easily identified by *TrichO*Key and *Tricho*BLAST, there will be cases where the sequence does not match that of an existing species, thus suggesting the presence of a putative newly discovered species. The most correct approach would then be to apply Genealogical Concordance Phylogenetic Species Recognition (GCPSR; Taylor et al. 2000) to investigate species limits. It requires the analysis of phylogenies of several unlinked genes, and further requires that the position of a phylogenetically distinct species is concordant in at least three of them and not contradicted in the others. GCPSR has not been stringently applied to *Hypocrea/Trichoderma*: Kullnig-Gradinger et al. (2002; see earlier) because of insufficient phylogenetic resolution in the genes that were used. Similar problems were faced by Chaverri et al. (2003a) and Chaverri and Samuels (2004) using *rpb2* and *tef1* exon sequences to distinguish phylogenetic species within the green-spored *Hypocrea* spp. This problem has recently been discussed in detail by Druzhinina and Kubicek (2005). Among 11 gene loci or fragments tested in *Hypocrea/Trichoderma*, the most promising ones appear to be the 4th and 5th introns of translation elongation factor 1-alpha (*tef1*, ~EF-1α,), and the coding portions of endochitinase 42 (*ech42*). Resolution of some clades can be obtained by the use of *rpb2* and the ITS1 and ITS2 diagnostic regions. Most of the other genes/loci tested provided only poor resolution, and thus the optimal combination of loci for all infrageneric groups of *Hypocrea/Trichoderma* allowing a straightforward application of GCPSR has not yet been found. A search for new phylogenetic markers in *Trichoderma* is therefore highly warranted.

Finally, we would like to outline our preferred phylogenetic approach to identify phylogenetic species. So far, most workers have employed the maximum-parsimony method to analyze sequence data, which does not employ a modeling of evolution and therefore poses several problems when investigating either very closely or very distantly related taxa (see Salemi and Vandamme 2003; and references therein). Maximum-likelihood methods are adequate but suffer from the long computation time they usually require. For this purpose, the Bayesian approach to phylogenetic inferences represents the most recent advance in phylogenetic analysis (Rannala and Yang 1996; Huelsenbeck and Ronquist 2001; Lutzoni et al. 2001). Bayesian approaches have been introduced in the analysis of phylogenies of other genera and have also recently been applied to *Hypocrea/Trichoderma* (Chaverri et al. 2003a; Druzhinina et al. 2004). When combined with rigorous testing of evolutionary models and the choice of an appropriate gene/locus, this method yields excellent resolution even for difficult to resolve clades (Samuels et al., manuscript submitted; Druzhinina et al., manuscript in preparation).

Trichoderma species identification using molecular data

As shown, *Trichoderma* taxonomy is a difficult issue. The abundant homoplasy in morphological and phenetic characters is likely the reason, which helps to explain why the number of morphologically distinct species has remained constant over time, whereas the number of phylogenetic species has rapidly reached 100, and it is expected to keep growing. In this review, we have attempted to introduce to the scientific community the recent effort in the development of modern tools for *Trichoderma* species identification. It has put these fungi in the privileged position that all its known species can be identified by the application of the very simple to perform polymerase chain reaction (PCR) technique, DNA sequencing, and user-friendly bioinformatics tools. This is a situation that is not yet true for most of the other fungal genera. Although online databases for identification of selected fungal groups such as ectomycorrhiza (UNITE, http://unite.zbi.ee for identification of ectomycorrhizal fungi) or *Fusarium* spp. (FUSARIUM-ID version 1.0; Geiser et al. 2004) are available, they only use sequences from a single locus, and they do not contain additional tools enabling the user to retrieve the respective sequences for subsequent phylogenetic analysis. With the tools described in this review, virtually every scientist working with *Trichoderma* can now reliably identify species using DNA sequence data. We believe that this will lead to the accumulation of a large amount of data on this genus, thereby advancing our understanding of its biology, ecology, and applied value in more detail.

Acknowledgments This work was supported in part by the Austrian Science Fund grants P-12748-MOB and FWF P-16601.

References

Abbott EV (1927) Taxonomic studies on soil fungi. Iowa State Coll J Sci 1:15–36
Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
Bisby GR (1939) *Trichoderma viride* Pers. ex Fries, and notes on *Hypocrea.* Trans. Br Mycol Soc 23:149–168
Bissett J (1984) A revision of the genus *Trichoderma*. I. Sect. *Longibrachiatum* sect. nov. Can J Bot 62:924–931
Bissett J (1991a) A revision of the genus *Trichoderma*. II. Infrageneric classification. Can J Bot 69:2357–2372
Bissett J (1991b) A revision of the genus *Trichoderma*. III. Sect. *Pachybasium*. Can J Bot 69:2373–2417
Bissett J (1991c) A revision of the genus *Trichoderma*. IV. Additional notes on section *Longibrachiatum*. Can J Bot 69:2418–2420
Bissett J (1992) *Trichoderma atroviride.* Can J Bot 70:639–641
Bissett J, Szakacs G, Nolan CA, Druzhinina IS, Kullnig-Gradinger CM, Kubicek CP (2003) Seven new taxa of *Trichoderma* from Asia. Can J Bot 81:570–586
Chaverri P, Castlebury LA, Samuels GJ, Geiser D (2003a) Multilocus phylogenetic structure within the *Trichoderma harzianum/Hypocrea lixii* complex. Mol Phylogenet Evol 27:302–313
Chaverri P, Castlebury LA, Overton BE, Samuels GJ (2003b) *Hypocrea/Trichoderma*: species with conidiophore elongations and green conidia. Mycologia 95:1100–1140
Chaverri P, Samuels GJ (2004) *Hypocrea/Trichoderma* (Ascomycota, Hypocreales, Hypocreaceae): species with green ascospores. Stud Mycol 48:1–116
Cutler HG, Cutler SJ, Ross SA, Sayed KE, Dugan FM, Bartlett MG, Hill AA, Hill RA, Parker SR (1999) Koninginin G, a new metabolite from *Trichoderma aureoviride*. J Nat Prod 62:137–139
Danielson R, Davey C (1973) The abundance of *Trichoderma* propagules and the distribution of species in forest soils. Soil Biol Biochem 5:485–494
de Queiroz K (1992) Phylogenetic definitions and taxonomic philosophy. Biol Philos 7:295–313
Druzhinina I, Chaverri P, Fallah P, Kubicek CP, Samuels GJ (2004) *Hypocrea flaviconidia*, a new species from Costa Rica with yellow conidia. Stud Mycol 50:401–407
Druzhinina I, Koptchinski A, Komon M, Bissett J, Szakacs G, Kubicek CP (2005) An oligonucleotide barcode for species identification in *Trichoderma* and *Hypocrea*. Fungal Genet Biol 42:813–828
Druzhinina I, Kubicek CP (2005) Species concepts and biodiversity in *Trichoderma* and *Hypocrea*: from aggregate species to species clusters? J Zhejiang Univ Sci 6:100–112
Gautheret A, Dromer F, Bourhis JH, Andremont A (1995) *Trichoderma pseudokoningii* as a cause of fatal infection in a bone marrow transplant recipient. Clin Infect Dis 20:1063–1064
Geiser DM, Jiménez-Gasco M, Kang S, Makalowska I, Veerarahavan N, Ward TJ, Zhang N, Kuldau GA, O'Donnell K (2004) FUSARIUM-ID v. 1.0: a DNA sequence database for identifying *Fusarium*. Eur J Plant Pathol 110:473–479
Gherbawy Y, Druzhinina I, Shaban GM, Wuczkowsky M, Yaser M, El-Naghy MA, Prillinger HJ, Kubicek CP (2004) *Trichoderma* populations from alkaline agricultural soil in the Nile valley, Egypt, consist of only two species. Mycol Prog 3:211–218
Harman GE, Howell CR, Viterbo A, Chet I, Lorito M (2004) *Trichoderma* species – opportunistic, avirulent plant symbionts. Nat Rev Microbiol 2:43–56
Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001) Molecular evidence for the early colonization of land by fungi and plants. Science 293:1129–1133
Hjeljord L, Tronsmo A (1998) *Trichoderma* and *Gliocladium* in biological control: an overview. In: Harman GE, Kubicek CP (eds) *Trichoderma and Gliocladium*, vol 2. Enzymes, biological control and commercial applications. Taylor & Francis, London, pp 131–151
Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755
Humphris SN, Bruce A, Buultjens E, Wheatley RE (2002) The effects of volatile microbial secondary metabolites on protein synthesis in *Serpula lacrymans*. FEMS Microbiol Lett 210:215–219
Jaklitsch WM, Komon M, Kubicek CP, Druzhinina IS (2006) *Hypocrea voglmayrii* sp. nov. from the Austrian Alps represents a new phylogenetic clade in *Hypocrea*/*Trichoderma*. Mycologia in press
Kindermann J, El-Ayouti Y, Samuels GJ, Kubicek CP (1998) Phylogeny of the genus *Trichoderma* based on sequence analysis of the internal transcribed spacer region 1 of the rDNA clade. Fungal Genet Biol 24:298–309
Klein D, Eveleigh DE (1998) Ecology of *Trichoderma*. In: Kubicek CP, Harman GE (eds) *Trichoderma* and *Gliocladium*, vol 1. Basic biology, taxonomy and genetics. Taylor & Francis, London, pp 57–73
Kopchinskiy A, Komon M, Kubicek CP, Druzhinina IS (2005) *Tricho*BLAST: a multiloci database of phylogenetic markers for *Trichoderma* and *Hypocrea* powered by sequence diagnosis and similarity search tools. Mycol Res 109:658–660
Kovacs K, Szakacs G, Pusztahelyi T, Pandey A (2004) Production of chitinolytic enzymes with *Trichoderma longibrachiatum* IMI 92027 in solid substrate fermentation. Appl Biochem Biotechnol 118:189–204
Kraus G, Druzhinina I, Bissett J, Prillinger HJ, Szakacs G, Koptchinski A, Gams W, Kubicek CP (2004) *Trichoderma brevicompactum* sp. nov. Mycologia 96:1059–1073
Kredics L, Antal Z, Doczi I, Manczinger L, Kevei F, Nagy E (2003) Clinical importance of the genus *Trichoderma*. A review. Acta Microbiol Immunol Hung 50:105–117
Kubicek CP, Bissett J, Druzhinina IS, Kullnig-Gradinger CM, Szakacs G (2003) Genetic and metabolic diversity of *Trichoderma*: a case study on South East Asian isolates. Fungal Genet Biol 38:310–319
Kubicek CP, Penttilä ME (1998) Regulation of production of plant polysaccharide degrading enzymes by *Trichoderma*. In: Harman GE, Kubicek CP (eds) *Trichoderma and Gliocladium*, vol 2. Enzymes, biological control and commercial applications. Taylor & Francis, London, pp 49–71
Kuhls K, Lieckfeldt E, Samuels GJ, Kovacs W, Meyer W, Petrini O, Gams W, Börner T, Kubicek CP (1996) Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*. Proc Natl Acad Sci USA 93:7755–7760
Kuhls K, Lieckfeldt E, Samuels GJ, Meyer W, Kubicek CP, Börner T (1997) Revision of *Trichoderma* section *Longibrachiatum* including related teleomorphs based on an analysis of ribosomal DNA internal transcribed spacer sequences. Mycologia 89:442–460
Kullnig CM, Szakacs G, Kubicek CP (2000) Molecular identification of *Trichoderma* species from Russia, Siberia and the Himalaya. Mycol Res 104:1117–1125
Kullnig-Gradinger CM, Szakacs G, Kubicek CP (2002) Phylogeny and evolution of the fungal genus *Trichoderma*: a multigene approach. Mycol Res 106:757–767
Lieckfeldt E, Kuhls K, Muthumeenakshi M (1998) Molecular taxonomy of *Trichoderma* and *Gliocladium* and their teleomorphs. In: Kubicek CP, Harman GE (eds) *Trichoderma* and *Gliocladium*, vol 1. Basic biology, taxonomy and genetics. Taylor & Francis, London, pp 35–74
Linnaeus C (1703) Systema naturae. Halae Magdesburgicae.
Lu B, Druzhinina I, Fallah P, Chaverri P, Gradinger C, Kubicek CP, Samuels GJ (2004) *Hypocrea/Trichoderma* species with *pachybasium*-like conidiophores: teleomorphs for *T. minutisporum* and *T. polysporum*, and their newly discovered relatives. Mycologia 96:310–342
Lutzoni F, Pagel M, Reeb V (2001) Major fungal lineages are derived from lichen symbiotic ancestors. Nature (Lond) 411:937–940
Nelson EE (1982) Occurrence of *Trichoderma* in a Douglas-fir soil. Mycologia 74:280–284
O'Donnell K, Cigelnik E, Nirenberg HI (1998) Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. Mycologia 90:465–493
O'Donnell K, Nirenberg HI, Aoki T, Cigelnik E (2000) A multigene phylogeny of the *Gibberella fujikuroi* species complex: detection of additional phylogenetically distinct species. Mycoscience 41:61–78
Penttilä ME (1998) Heterologous protein production in *Trichoderma*. In: Harman GE, Kubicek CP (eds) *Trichoderma* and *Gliocladium*,

vol 2. Enzymes, biological control and commercial applications. Taylor & Francis, London, pp 356–383

Persoon CH (1794) Disposita methodical fungorum. Römers Neues Mag Bot 1:81–128

Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic interference. J Mol Evol 43:304–311

Rifai MA (1969) A revision of the genus *Trichoderma.* Mycol Pap 116:1–56

Rifai MA, Webster J (1966a) Culture studies of *Hypocrea* and *Trichoderma* II. *H. aureoviridis* and *H. rufa* f. *sterilis* f. nov. Trans Br Mycol Soc 49:289–296

Rifai MA, Webster J (1966b) Culture studies on *Hypocrea* and *Trichoderma* III. *H. lactea* (=*H. citrina*) and *H. pulvinata.* Trans Br Mycol Soc 49:297–310

Salemi M, Vandamme A (2003) The phylogenetic handbook. A practical approach to DNA and protein phylogeny. Cambridge University Press, Cambridge

Samuels GJ (1996) *Trichoderma*: a review of biology and systematics of the genus. Mycol Res 100:923–935

Sivasithamparam K, Ghisalberti EL (1998) Secondary metabolism in *Trichoderma* and *Gliocladium*. In: Kubicek CP, Harman GE (eds) *Trichoderma* and *Gliocladium*, vol 1. Basic biology, taxonomy and genetics. Taylor & Francis, London, pp 139–191

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC (2000) Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31:21–32

Taylor JW, Spatafora J, O'Donnell K, Lutzoni F, James T, Hibbett DS, Geiser D, Bruns TD, Blackwell M (2004) The Fungi. In: Cracraft J, Donoghue MJ (eds) Assembling the tree of life. Oxford University Press, New York, pp 171–194

Thrane U, Poulsen SB, Nirenberg HI, Lieckfeldt E (2001) Identification of *Trichoderma* strains by image analysis of HPLC chromatograms. FEMS Microbiol Lett 203:249–255

Tulasne L, Tulasne R (1860) De quelques Sphéries fungicoles, à propos d'un mémoire de M. Antoine de Bary sur les *Nyctalis*. Ann Sci Nat Bot 13:5–19

Wey TT, Hseu TH, Huang L (1994) Molecular cloning and sequence analysis of the cellobiohydrolase I gene from *Trichoderma koningii* G-39. Curr Microbiol 28:31–39

Widden P, Abitbol JJ (1980) Seasonality of *Trichoderma* species in a spruce-forest soil. Mycologia 72:775–784

Wuczskowski M, Druzhinina IS, Gherbawy Y, Klug B, Prillinger HJ, Kubicek CP (2003) Taxon pattern and genetic diversity of *Trichoderma* in a mid-European, primeval floodplain-forest. Microbiol Res 158:125–134

Zhang CL, Druzhinina IS, Kubicek CP, Xu T (2005) Biodiversity of *Trichoderma* in China: evidence for a North to South difference of species distribution in East Asia. FEMS Microbiol Lett 251:251–257