UPPSALA
UNIVERSITET

Improved Phylogenetic Classification of Ascomycetous Fungi Using Partial Beta-Tubulin Sequences as Molecular Markers

Tajamul Hussain

Degree Project in Biology 45hp, 2013

Fungal Genetics and Genomics Unit, Department of Applied Genetics and Cell Biology (DAGZ), University of Natural Resources and Life Sciences, (BOKU), Vienna, Austria.

Supervisor: Prof. Dr. Joseph Strauss (Head of the unit)

Co-Supervisor: Dr. Markus Gorfer (Senior-Scientist)

Table of Contents

Abbreviations

P	Penicillum
rDNA	Ribosomal DNA
ITS	Internal Transcribed spacer regions
SSU	Small ribosomal subunit
LSU	Large ribosomal subunit
IGS	Intergenic spacer regions
PCR	Polymerase chain reaction
CLC	Control logic centre (Bio-tool)
S	Svedberg Sedimentations unit
BLAST	Basic Local Alignment Search Tool
NCBI	National Center for Biotechnology Information
DD	Double distilled
Pmol	Pico-molar
BSA	Bovine-serum albumin
NJ	Neighbours Joining
Bt	Beta-tubulin
DMSO	Dimethyl sulfoxide
TAE	Tris-Acetic acid EDTA
EDTA	Ethylenediaminetetraacetic acid
MEA	Malt-extract Agar
mRNA	Messenger Ribonucleic Acid
µg	Micro-gram
µl	Micro-liter
ml	Milli-liter
M	Marker

Summary

Ribosomal RNA (rRNA) genes are the most commonly used loci in molecular systematics studies of fungi [1]. However, the rRNA sequences exist as multiple copies and can be variable even within a single spore. This variation compromises their usage for species identification [2, 3]. In many cases these markers provide good phylogenetic resolution from the phylum to the genus level and the internal transcribed spacer regions (ITS) has recently been accepted as the official universal DNA barcode for fungi [4], but the species identification in some cases still is not possible. The most prominent group, where ITS sequences do not always allow correct species discriminations are the well-known genera *Aspergillus* and *Penicillium*. Additional marker gene sequences are therefore necessary for correct species delineation.

Beta-tubulin genes are found reliably in all eukaryotes and have been used for phylogenetic analysis in fungi from the entire kingdom to the species level [5-7]. The beta-tubulin gene sequences contain 3.5-fold more phylogenetic information than the small sub-unit (SSU) rRNA gene, thus it has been reported that the beta-tubulin gene is an ideal marker for analysis of deep-level phylogenies and for complex species groups [8].

The present study focused on the highly conserved gene coding for beta-tubulin in order to use it as an alternative to the frequently used phylogenetic marker ITS region (rDNA) for discriminating the most widespread and common *Penicillium* (Genus of Ascomycetous fungi) species. This is relevant especially for the *Penicillium chrysgenum*, the major industrial source of the beta-lactam antibiotic penicillin, with annual worldwide sale of about US $8 billion [9,10].

The tree generated by control logic centre (CLC, Denmark) - the world's leading bioinformatics analysis software – used in our study was clearly able to distinguish the two distinct species of *Penicillium chrysogenum* [10-12], which the ITS based tree failed to highlight.

PCR based amplification of the beta-tubulin gene and its subsequent in-silico sequencing analyses in the current study demonstrates that beta-tubulin gene sequences provided higher resolution than ITS in *Pencillium*, particularly in the *Penicillium chrysogenum* and *Penicillium spinulosum*.

Key words: ITS, beta-tubulin, CLC, rRNA, phylogenetic marker, PCR

1. Central idea behind the study

The major goal behind this study was to assess the utility of the beta-tubulin gene as a secondary phylogenetic marker in order to improve the existing ITS-based classification of the largest phylum Ascsomycota of the Kingdom fungi.

As the Fungal Genetics and Genomics Unit at the hosting laboratory holds its own collection of fungal strain from different habitats and locations. The strain collection currently holds ~480 ascomycetes that still lack beta-tubulin sequence information. These strains are used as a valuable resource for a range of projects, e.g. for the study of nitrate assimilation by fungi in soil or the identification of new bioactive compounds. For these aspects a correct and reliable taxonomic classification is indispensable. Closely related species can differ markedly in their secondary metabolite profiles. The main focus in this study was on the genus *Penicillium;* however, in future studies *Aspergillus* and other important strains of *Ascomycetous* will be studied as well.

2. General Introduction

Fungi are the largest kingdom of eukaryotic life after animalia [4] consisting of around 100,000 known and potentially 1.5-5.1 million un-described species [13]. They play key roles mainly as decomposers, mutualists, and parasites in every biome on the earth [13], and they are a big source of antibiotics. Fungi have a role in the production of bread and beer but are also responsible for many horrors in their sister kingdoms, so it has correctly been quoted by Hauke Harms (the famous mycologist) that “A world without fungi would be totally boring, not only boring, but maybe even unthinkable“. Moreover, fungi have been found to be closely related to the animals from an evolutionary perspective, thus, it is highly crucial to comprehensively understand the diversity and ecology of this important kingdom.

The kingdom fungi or Eumycota is composed of four phyla, Ascomycota, Basidiomycota, Zygomycota- and Chytidimycota. The Ascomycota, or sac fungi, are monophyletic and account for approximately 75% of all described fungi containing more than 32000 named species. This large group of fungi is significant both economically and medicinally as this is the phylum which contains the most influential members of fungi i.e. edible forms, producers of antibiotics and other drugs, cheese formers, wood rotter and a few of them are even pathogenic.

Some of the most famous fungi among Ascomycota are *Saccharomyces cerevisiae* - the yeast of commerce and the foundation of the baking and brewing industries, *Penicillium chrysogenum* producer of penicillin, *Morchella esculentum*, the edible morel, and *Neurospora crassa*, the "one-gene-one-enzyme" organism. Others are known for their harmful nature like the notorious *Aspergillus flavus* (Aflatoxin producer), which is one the most potent known natural carcinogens and a contaminant of nuts, grains etcetera. *Aspergillus, fumigatus, flavus, terreus* [14] and *Talaromyces (*Syn=*Penicillium*) *marneffei* [15] are recognized as opportunistic pathogens of humans, especially those with weakened immune systems, yet others are *Candida albicans*, the major cause of thrush, diaper rash and vaginitis and *Cryphonectria parasitica*, responsible for the demise of 4 billion chestnut trees in the eastern USA [16].

Notwithstanding the immense importance of Ascomycota, the systematics or taxonomy of these fungi still is problematic. There are, for example, difficulties in resolving species boundaries as traditional taxonomic studies rely primarily on phenotypic analysis (e.g. [17]) however, morphology alone is inadequate as structural differences are often small and therefore other characteristics are needed to recognize or verify species complexes. In order to resolve such complexes at the species level, many investigations have been done through sequence based analysis, particularly after the invention of PCR and automated sequencing techniques, using different potential phylogenetic markers. These markers are either coding or non-coding fragments of DNA used in phylogenetic reconstructions. Such fragments of the DNA are known to have no or few predictable variations within a given species, and are available for most or all species of a genus. So far the markers used frequently for phylogenetic classification of fungi are rRNA sequences, mitochondrial and chloroplast based sequences plus some potential conserved protein coding genes.

Ribosomal RNA genes are the most commonly used loci in molecular systematic studies of fungi [1]. The majority of fungal phylogeny is based on one or two of these

loci [1] and they have been widely used for more than 20 years for fungal diagnostic and phylogenetics [4]. The eukaryotic rRNA cistron consist of the 18S, 5.8S, and 28S rRNA genes transcribed as a unit by RNA polymerase 1, and separated by non-transcribed intergenic spacers (IGS). Posttranscriptional processes split the cistron and remove two internal transcribed spacers. These two spacers including the 5.8S gene are usually referred as the ITS region [4]. Nowadays, the ITS region is the most widely sequenced DNA region in fungi [18]. The ITS region has a higher degree of variation, so it is being used frequently in molecular systematic and taxonomy especially in resolving species complex.

As the rRNA sequences exist in multiple copies and can be variable even within a single spore - this variation compromises their usage for species identification [2, 3]. In many cases these markers provide good phylogenetic resolution from the phylum to the genus level, and the ITS has been officially accepted recently as a universal DNA barcode for fungi by a multinational, multi-laboratory consortium [4]. Nevertheless, the species identification remains problematic, especially in genus like *Penicillium* and *Aspergillus.* Support for alternative identifying markers in this context is therefore necessary for correct species delineation. Therefore in this study an effort was made to discover the potential of the highly conserved and functionally significant beta-tubulin as an alternative phylogenetic marker gene to correctly identify and discriminate the genus *Penicillium* of the phylum Ascomycota at the species level. Beta-tubulin, a crucial member of the globular protein tubulin, which is present in almost all eukaryotic cells and even in prokaryotes its homologue, FtsZ, has been discovered. Tubulin represents the elementary subunits of microtubules, which in turn represent the major components of the cytoskeleton (including mitotic spindles) and eukaryotic flagellae [19]. The tubulins constitute a protein family involving α(alpha)-(beta)-(gamma)-(delta)-(epsilon)-(zeta)- and (eta)-tubulins [20]. Of the seven varieties, α and β-tubulins are the most abundant in the eukaryotic cell, as their heterodimers are the primary constituents of the microtubules. The gamma subunit plays a role in the nucleation of microtubule assembly during the establishment of microtubule organizing centers, and the remaining four tubulins varieties are functionally associated with the centriole or basal body of eukaryotic cells and organisms [20, 21]. Alpha and beta tubulins, especially their N-terminal peptides, are remarkably conserved [22]; therefore genes for these tubulins, especially for beta-tubulin, are receiving increasing attention in the investigation of evolutionary relationships at all levels: (i) in kingdom-level phylogenetic analyses [23,24], and (ii) in studies of complex species groups within protists, animals, fungi and plants [25-27]. The beta tubulin gene sequences contain 3.5-fold more phylogenetic information than the SSU rRNA gene, thus it has been reported that beta-tubulin gene is an ideal marker for analysis of deep-level phylogenies and for complex species groups [8]. The beta-tubulin gene may be ideal for studying relationships among Ascomycota for a number of reasons, as beta-tubulin genes are found reliably in all eukaryotes and have been used for phylogenetic analysis in fungi from the kingdom to species level [5-7]. In addition, protein coding genes, unlike, rDNA genes, accumulate fewer mutations and are less variable among the few copies that most genomes contain [19]. The beta tubulin gene has highly conserved sequences, thus facilitating the design of primers. The present study checked the efficiency of beta-tubulin via comparing this with the ITS in order to evaluate the utility of beta-tubulin as an alternative phylogenetic marker for the genus *Penicillium*.

To achieve the designed goal, the following tasks were performed sequentially:

- DNA isolation from fungi (from which no DNA was stored in the collection).
- PCR amplification of the beta-tubulin gene with primers that were previously shown to work on most (but not all) members from the Ascomycetes.
- Preparation of PCR products for sequencing. Sequencing was performed by a specialised company.
- Editing and assembling the raw sequence reads.
- Phylogenetic placement of fungal strains based on all available gene markers (ITS and beta-tubulin).
- Construction of the tree for the comparative analysis of the two mentioned markers.

3. Material/Methods and Results

3.1 Workflow diagram of the project

Figure 1 depicts the overall workflow of the project including cultivation of the strains, extraction, purification and amplification of the DNA, analyses of the amplification through agarose gel electrophoresis, sequencing and *in silico* analysis.

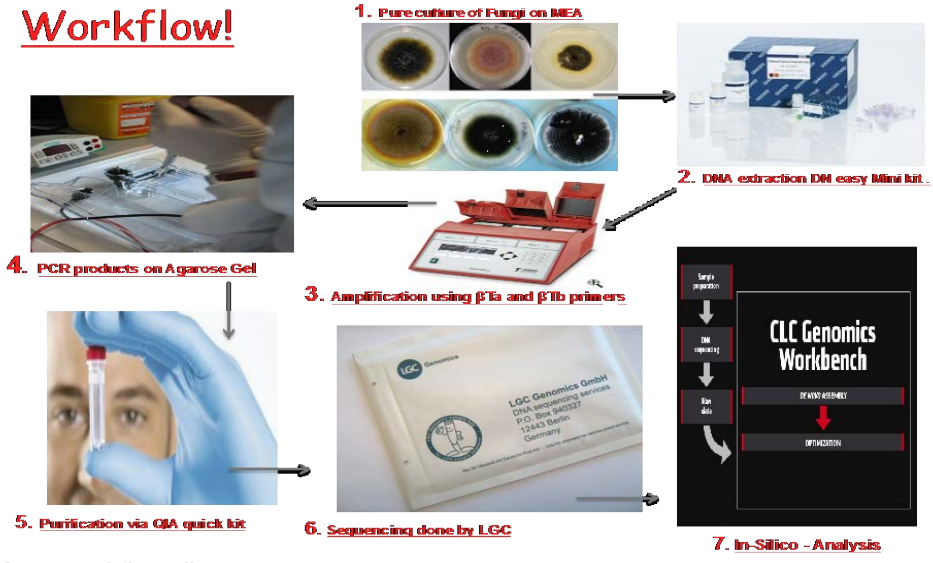

Fig.1 Workflow diagram

3.2 Penicillium strains

The *Penicillium* strains and isolates used in this study are listed in table 1. These are stored in the strain collection at fungal genetics and genomics unit in the department of applied genetics and cell biology (DAGZ) (University of Natural Resources and Life-sciences, Vienna). The collection currently holds ~480 Ascomycetes, which lack beta-tubulin sequence information. Ten out of approximately 66 *Penicillium* strains

had no DNA samples, 56 had their DNA already isolated and were stored at -20ºC. The 10 strains (for details see table 2) from which the DNA was extracted, were collected from the - 80ºC refrigerator and were grown and incubated on MEA (Composition g/Litre: Malt extract: 30.0, peptone from soymeal: 3.0; Agar-agar:15.0) plates for 7 days at 25ºC.

Table 1 List of the *Penicillium* strains used in the study

No	ID	Code	DNA	Genus	Species
1	27	RSF-Q205	+	*Penicillium*	*decaturense*
2	154	Hbs-K08	+	*Penicillium*	*commune*
3	349	PRF-18	+	*Penicillium*	*restrictum*
4	681	KO55	+	*Penicillium*	*corylophilum*
5	758	NG_p02	+	*Penicillium*	*canescens*
6	777	NG_p22	+	*Penicillium*	*canescens*
7	778	NG_p23	+	*Penicillium*	*islandicum/rugulosum*
8	788	NG_p34	+	*Penicillium*	*sp. Related*
9	796	NG_p43	+	*Penicillium*	*islandicum*
10	812	NG_06	+	*Penicillium*	*canescens*
11	817	NG_11	+	*Penicillium*	*n.s.*
12	820	NG_17	+	*Penicillium*	*n.s.*
13	821	NG_18	+	*Penicillium*	*n.s.*
14	822	NG_20	+	*Penicillium*	*sp.*
15	824	NG_23	+	*Penicillium*	*janthinellum*
16	825	NG_24	+	*Penicillium*	*glandicola*
17	826	NG_25	+	*Penicillium*	*ochrochloron*
18	827	NG_26	+	*Penicillium*	*janthinellum*
19	850	GabP	+	*Penicillium*	*biourgeianum*
20	856	SC01	+	*Penicillium*	*verruculosum-related*
21	859	SC04	+	*Penicillium*	*Piceum*
22	861	SC06	+	*Penicillium*	*Brevicompactum*
23	869	D_D04	+	*Penicillium*	*Piceum*
24	886	Li0102II	+	*Penicillium*	*glabrum/thomii*
25	887	Li0102III	+	*Penicillium*	*glabrum/thomii*
26	888	Li0102IV	+	*Penicillium*	*glabrum/thomii*
27	889	Li0102V	+	*Penicillium*	*Spinulosum*
28	890	Li0102VI	+	*Penicillium*	*glabrum/thomii*
29	893	Li0102IX	+	*Penicillium*	*Spinulosum*
30	895	Li0102XI	+	*Penicillium*	*glabrum/thomii*
31	896	Li0102XII	+	*Penicillium*	*Lividum*

No	ID	Code	DNA	Genus	Species
32	933	D_D12	+	*Penicillium*	*Chrysogenum*
33	938	D_D22(P)	+	*Penicillium*	*chrysogenum*
34	941	D_D25	+	*Penicillium*	*chrysogenum*
35	942	D_D26	+	*Penicillium*	*chrysogenum*
36	943	D_D27	+	*Penicillium*	*crustosum*
37	944	D_D28	+	*Penicillium*	*glabra*
38	958	A04	+	*Penicillium*	*miczynskii*
39	961	W121	+	*Penicillium*	*echinulatum*
40	1009	D_D34	+	*Penicillium*	*brevicompactum*
41	1021	D_D38	+	*Penicillium*	*olsonii*
42	1035	D_D42	-	*Penicillium*	*corylophilum*
43	1043	D_D43	-	*Penicillium*	*erythromellis*
44	1045	D_D47	-	*Penicillium*	*polonicum*
45	1127	D_D53	-	*Penicillium*	*chrysogenum*
46	1129	E13a	-	*Penicillium*	*vulpinum*
47	1132	SAN1075-3	-	*Penicillium*	*chrysogenum complex*
48	1135	E17	-	*Penicillium*	*brevicompactum/biourgeianum*
49	1139	D_D60	-	*Penicillium*	*decumbens*
50	1209	D_D70	-	*Penicillium*	*steckii-related*
51	1255	STC1101-4	-	*Penicillium*	*clavigerum*
52	1408	NG_33	-	*Penicillium*	*chrysogenum*
53	149	Hbs-K10	+	*Penicillium*	*alberechii*
54	156	Hbs-K13	+	*Penicillium*	*brevicompactum*
55	150	Hbs-K14	+	*Penicillium*	*alberechii*
56	136	Hbs-SZ5	+	*Penicillium*	*Chrsogenum*
57	175	MC-A12	+	*Penicillium*	*Chrsogenum*
58	178	MC-A32	+	*Penicillium*	*Chrsogenum*
59	155	MC-B11	+	*Penicillium*	*Commune*
60	17	RSF-P404	+	*Penicillium*	*canescens*
61	24	RSF-Q202	+	*Penicillium*	*sopii*
62	7	RSF-P101	+	*Penicillium*	*arenicola*
63	23	RSF-Q201	+	*Penicillium*	*soppii*
64	137	MX-C1	+	*Penicillium*	*pinophilum*
65	138	MX-C2	+	*Penicillium*	*crustosum*
66	1075	KF1016_M1	+	*Penicillium*	*chrysogenum*

n.s.; not specified

3.3 Genomic DNA preparation

The genomic DNA from the 10 strains of *Penicillium* listed in table 2 were extracted from the grown mycellium in their respective plates using the DNeasy plant mini kit (Qiagen, Valenica, CA, USA, CAT No/ 69104) according to the manufacturers' guidelines. The DNA concentration was checked via the agarose gel electrophoresis (2%) and 1:10 diluted DNA was used (45µl PCR water-deionised plus 5µl of the concentrated DNA) for amplification.

Table 2: Depiction of the 10 strains and their results of DNA extraction and amplification

Lane No	***ID***	***Code***	***DNA***	***PCR***	***Genus***	***Species***
1	*1021*	*D_D42*	**+**	**+**	*Penicillium*	*olsonii*
2	*1035*	*D_D43*	**+**	**+**	*Penicillium*	*corylophilum*
3	*1043*	*D_D47*	**+**	**+**	*Penicillium*	*erythromellis*
4	*1045*	*D_D53*	**+**	**+**	*Penicillium*	*polonicum*
5	*1127*	*E13a*	**+**	**+**	*Penicillium*	*chrysogenum*
6	*1129*	*SAN1075-3*	**+**	**+**	*Penicillium*	*vulpinum*
7	*1132*	*E17*	**+**	**+**	*Penicillium*	*chrysogenum complex*
8	*1135*	*D_D60*	**+**	**+**	*Penicillium*	*brevicompactum/biourgeianum*
9	*1139*	*D_D70*	**+**	**+**	*Penicillium*	*decumbens*
10	*1209*	*STC1101-4*	**+**	**+**	*Penicillium*	*steckii-related*

3.4 PCR amplification of the target gene

Amplification of the beta-tubulin gene in the 66 strains of the Penicillum was executed using the primers Bt2a = (5'-GGTAACCAAATCGGTGCTGCTTTC-3') and Bt2b= (5'-ACCCTCAGTGTAGTGACCCTTGGC -3') [28].
PCR reactions were performed in 50µl reaction mixture containing:

- ✓ 25 µl -2x PCR mix (green master mix).
- ✓ 13 µl- DD PCR water (double distilled)
- ✓ 5 µl –Bt2a (10pmol/ µl - forward)
- ✓ 5 µl –Bt2b (10pmol/ µl - reverse)
- ✓ 1 µl – BSA
- ✓ 1 µl – DNA (1:10)

Amplifications were carried out in a thermocylcer (T3000-Thermocylcer 48, Biometra -Serial No. 2509344). The amplification mode was set up as follows:
Initial strands were denatured at 94ºC for 5 cycles of 1 min, followed by primer annealing for 90s at 68ºC, and extension for 2 min. at 72ºC with a decrease of annealing temperature of 1º/cycle, followed by 25 cycles of denaturation at 94ºC for 1 min, followed by primer annealing for 9s at 64ºC, extension for 2 min. at 72ºC and a final 10 min. elongation step at 72ºC.

After complete amplification, the PCR products were analyzed by 2% agarose gel (2g of agarose in 100ml of 1x TAE buffer (1000ml DD water plus 20ml TAE from 50 x TAE) with gel-red nucleic acid staining (10.000x in DMSO, room temperature, no UV-light).

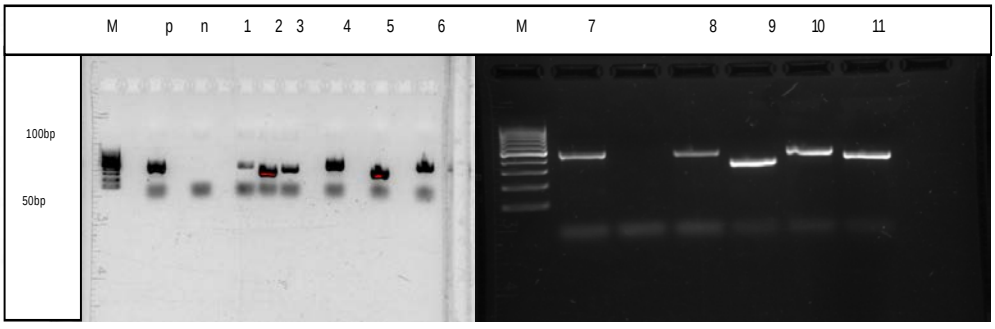

Fig.2 PCR products of the above enlisted 10 strains of *Penicillium.* M: 100bp DNA ladder marker

Out of the 66, 56 samples were amplified in the same amplification procedure (for details see figure 3). However, 7 of them failed to show a proper amplification result. Detail results of these strains are depicted in the table 3

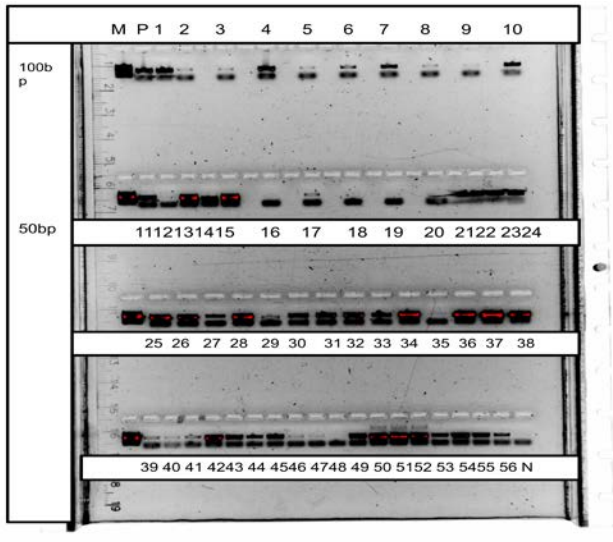

Fig.3 PCR amplification results of the beta-tubulin genes from the 56 strains of the remaining *Penicillium*. M: DNA ladder of 100bp, P: positive control N: negative control.

3.5 DNA purification and sequencing

In order to make the sequencing reaction, the extracted DNA samples were purified using the QIA PCR purification kit protocol with the help of micro-centrifuge following the manufacturer's instructions. After the purification the samples were analysed for

its final concentration via running over 1% electrophoresis using 6x LD Thermo-Scientific DNA loading dye. Afterwards, the purified DNA of around 50µl was stored in the 1.5ml microfuge tube at minus 20ºC for future use.
Figure 4 depicts the purification results. Enough DNA for sequencing and *in silico* analyses was obtained from all of the strains.

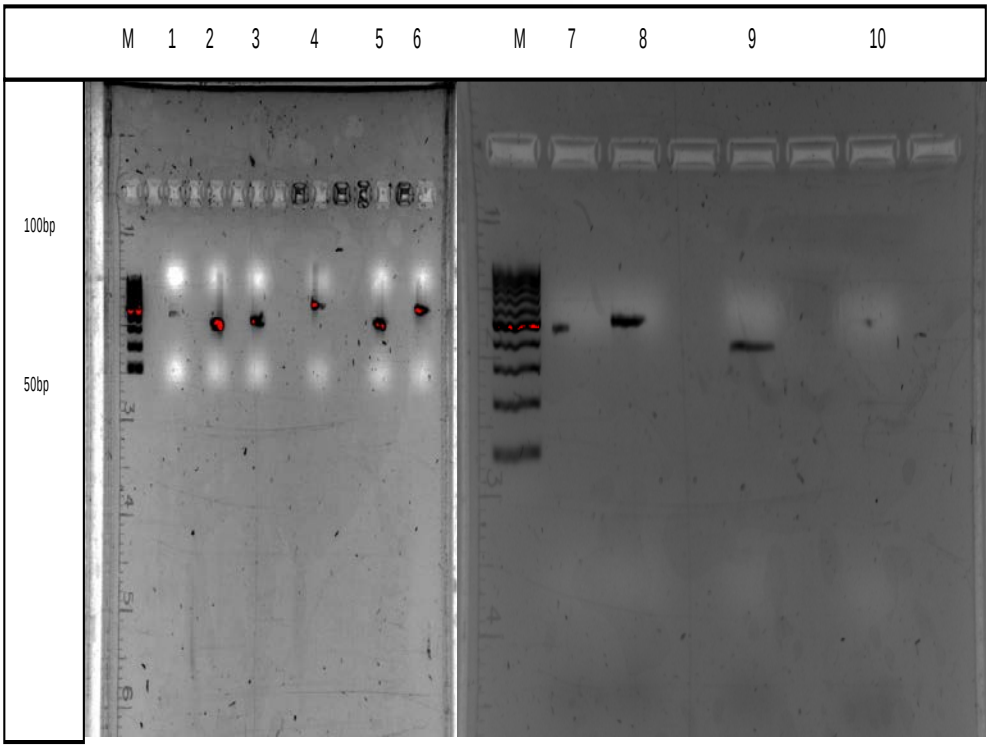

Fig.4 The purified DNA of the above listed 10 strains of *Penicillium*.M: Gene ruler (100mb)

The 28µl of the sequencing reaction mixture was made as follows:
4µl of each primer (10pmol/µl), 8µl of purified DNA and 16µl of PCR water and the samples were then dispatched to LGC- Genomics unit in Berlin for an accurate sequencing, after 3 or 4 days, the sequenced data were received via email.
LGC is an internally renowned leader in laboratory services and recently it has been ranked NO.1 in DNA Sequencing in the world. Sequencing orders have been done online via LGC Genomics-website: http://shop.lgcgenomics.com/

3.6 Phylogenetic data analysis

The in-silico analysis of the sequenced data was done mainly via the Danish based authentic bio-informatics software package called CLC (Genomics workbench- a desktop application with a graphical user interface for faster and advanced DNA, RNA and proteins analysis). This program is user friendly and evolves continuously to keep it up to date.

This software tool connects analytical activities step by step:

- ✓ Assembling the raw sequenced data (manual adjustment was carried out to maximize homology; i.e. trimming the primer regions and resolving the gap (if any) by manual adjustment to improve the alignment).
- ✓ Alignments of the assembled sequences were done via BLAST programme (NCBI).
- ✓ The aligned sequences were then used for the construction of the final tree based on neighbour-joining algorithm by CLC.
- ✓ The trees were analysed visually for the comparative analysis.

4. Summary of the overall results

From 66 beta-tubulin gene sequences from *Penicillium* (see table 3), for 53 the final beta-tubulin based tree was generated (for details see figure 5). Seven *Penicillium* strains failed to amplify in PCR and six could not be sequenced therefore further *in silico* analysis could not be performed. Out of 53, 24 *Penicillum* strains have shown different genus and species ID from the ITS based IDs after the BLAST. While, 29 strains had the same ITS based ID. These differences can be seen in overall topology and branch support of the trees inferred under NJ algorithm using CLC bio-tool (depicted in figure 5 and 6).

Table 3 Overall summary of all results obtained during the course of this study[1]

ID	Code	Box	Genus	Species	DNA	PCR	Sequence	Assembly	ID bTub
27	RSF-Q205	G2	Penicillium	Decaturense	positive	positive	positive	positive	Penicillium pancosmium
154	Hbs-K08	D1	Penicillium	Commune	positive	positive	negative	negative	
349	PRF-18	F2	Penicillium	Restrictum	positive	positive	positive	positive	same
681	KO55	F2	Penicillium	Corylophilum	positive	positive	positive	positive	same
758	NG_p02	J4	Penicillium	Canescens	positive	positive	positive	positive	same
777	NG_p22	J4	Penicillium	Canescens	positive	positive	positive	positive	same
778	NG_p23	J3	Penicillium	islandicum/rugulosum	positive	negative			
788	NG_p34	J5	Penicillium	sp. Related	positive	positive	positive	positive	Penicillium piscarium
796	NG_p43	J5	Penicillium	Islandicum	positive	positive	positive	positive	Penicillium sp.
812	NG_06	J1	Penicillium	Canescens	positive	positive	positive	positive	Penicillium sp.
817	NG_11	J1	Penicillium	n.s.	positive	positive	positive	postive	same
820	NG_17	J1	Penicillium	n.s.	positive	positive	positive	positive	same
821	NG_18	J2	Penicillium	n.s.	positive	neagtive			
822	NG_20	J2	Penicillium	sp.	positive	negative			
824	NG_23	J2	Penicillium	Janthinellum	positive	negative			
825	NG_24	J2	Penicillium	Glandicola	positive	positive	negative		
826	NG_25	J2	Penicillium	Ochrochloron	positive	positive	negative		
827	NG_26	J2	Penicillium	Janthinellum	positive	negative			
850	GabP	K3	Penicillium	Biourgeianum	positive	positive	positive	positive	Penicillium bialowiezense
856	SC01	K1	Penicillium	verruculosum-related	positive	positive	positive	positive	Talaromyces macrosporus
859	SC04	K1	Penicillium	Piceum	positive	positive	negative	negative	
861	SC06	K1	Penicillium	brevicompactum	positive	positive	positive	positive	same
869	D_D04	K3	Penicillium	Piceum	positive	positive	negative	negative	

ID	Code	Box	Genus	Species	DNA	PCR	Sequence	Assembly	ID bTub
886	Li0102II	K4	Penicillium	glabrum/thomii	positive	positive	positive	positive	Penicillium spinulosum
887	Li0102III	K4	Penicillium	glabrum/thomii	positive	positive	positive	positive	Penicillium spinulosum
888	Li0102IV	K4	Penicillium	glabrum/thomii	positive	positive	positive	positive	Penicillium spinulosum
889	Li0102V	K4	Penicillium	Spinulosum	positive	positive	positive	positive	same
890	Li0102VI	K4	Penicillium	glabrum/thomii	positive	positive	positive	positive	Penicillium spinulosum
893	Li0102IX	K4	Penicillium	Spinulosum	positive	positive	positive	positive	same
895	Li0102XI	K4	Penicillium	glabrum/thomii	positive	positive	positive	positive	Penicillium spinulosum
896	Li0102XII	K4	Penicillium	Lividum	positive	positive	positive	positive	same
930	D_D20	K5	Penicillium	roseopurpureum-related	positive	negative	negative	negative	
933	D_D12	K5	Penicillium	Chrysogenum	positive	positive	positive	positive	same
938	D_D22(P)	L1	Penicillium	Chrysogenum	positive	positive	positive	positive	same
941	D_D25	L1	Penicillium	Chrysogenum	positive	positive	positive	positive	same
942	D_D26	L1	Penicillium	Chrysogenum	positive	positive	positive	positive	same
943	D_D27	L1	Penicillium	Crustosum	positive	positive	positive	positive	same
944	D_D28	L1	Penicillium	Glabra	positive	positive	positive	positive	Penicillium glabrum
958	A04	L1	Penicillium	Miczynskii	positive	positive	positive	positive	Fusarium oxysporum f. sp.
961	W121	L2	Penicillium	Echinulatum	positive	positive	positive	positive	Penicillium solitum
1009	D_D34	L4	Penicillium	brevicompactum	positive	positive	positive	positive	same
1021	D_D38	L4	Penicillium	Olsonii	positive	positive	positive	positive	same
1035	D_D42	L2	Penicillium	Corylophilum	positive	positive	positive	positive	same
1043	D_D43	L5	Penicillium	Erythromellis	positive	positive	positive	positive	Penicillium sp
1045	D_D47	L3	Penicillium	Polonicum	positive	positive	positive	positive	same
1075	KF1016_M1	K2	Penicillium	Chrysogenum	positive	positive	positive	positive	Chaetomium cruentum
1127	D_D53	M3	Penicillium	Chrysogenum	positive	positive	positive	positive	same
1128	D_D54	M3	Penicillium	Decumbens	positive	positive	negative	negative	
1129	E13a	M3	Penicillium	Vulpinum	positive	positive	positive	positive	same
1132	SAN1075-3	M3	Penicillium	chrysogenum complex	positive	positive	positive	positive	Penicillium

ID	Code	Box	Genus	Species	DNA	PCR	Sequence	Assembly	ID bTub
									chrysogenum
1135	*E17*	*M4*	*Penicillium*	*brevicompactum/biourgeianum*	*positive*	*positive*	*positive*	*positive*	*Penicillium biourgeianum*
1139	*D_D60*	*M4*	*Penicillium*	*Decumbens*	*positive*	*positive*	*positive*	*positive*	*same*
1209	*D_D70*	*N5*	*Penicillium*	*steckii-related*	*positive*	*positive*	*positive*	*positive*	*Penicillium copticola*
1255	*STC1101-4*	*O2*	*Penicillium*	*Clavigerum*	*positive*	*positive*	*positive*	*positive*	*same*
1408	*NG_33*	*J3*	*Penicillium*	*Chrysogenum*	*positive*	*negative*	*negative*	*negative*	
149	*Hbs-K10*	*D1*	*Penicillium*	*Alberechii*	*positive*	*positive*	*positive*	*positive*	*penicillium bialowiezense*
156	*Hbs-K13*	*D1*	*Penicillium*	*brevicompactum*	*positive*	*positive*	*positive*	*positive*	*same*
150	*Hbs-K14*	*A1*	*Penicillium*	*Alberechii*	*positive*	*positive*	*positive*	*positive*	*Penicillium bialowiezense*
136	*Hbs-SZ5*	*D1*	*Penicillium*	*Chrsogenum*	*positive*	*positive*	*positive*	*positive*	*Penicillium palitans*
175	*MC-A12*	*D2*	*Penicillium*	*Chrsogenum*	*positive*	*positive*	*positive*	*positive*	*same*
178	*MC-A32*	*D2*	*Penicillium*	*Chrsogenum*	*positive*	*positive*	*positive*	*positive*	*same*
155	*MC-B11*	*D1*	*Penicillium*	*Commune*	*positive*	*positive*	*positive*	*positive*	*Penicillium chrysogenum*
17	*RSF-P404*	*G1*	*Penicillium*	*Canescens*	*positive*	*positive*	*positive*	*positive*	*same*
24	*RSF-Q202*	*G2*	*Penicillium*	*Sopii*	*positive*	*positive*	*positive*	*positive*	*Penicillium canescens*
7	*RSF-P101*	*G1*	*Penicillium*	*Arenicola*	*positive*	*positive*	*positive*	*positive*	*same*
65	*MX-C2*	*M3*	*Penicillium*	*Crustosum*	*positive*	*positive*	*positive*	*positive*	*same*

n.s., not specified; [1]negative results are presented in red colour

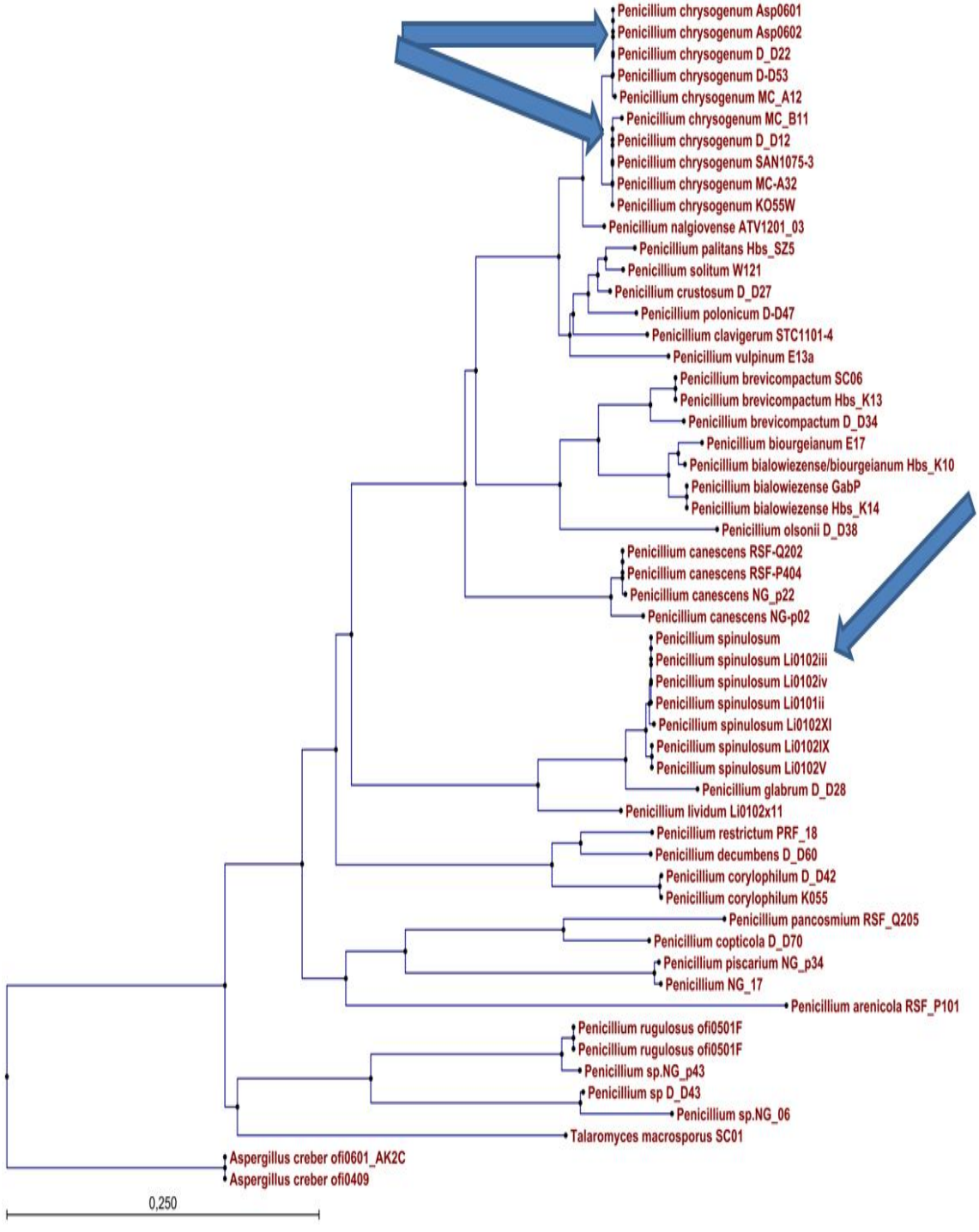

Fig.5 Beta-tubulin based phylogenetic tree of 53 *Penicillium* strains. The tree was constructed with the neighbor-joining method using CLC workbench. *Aspergillus versicolor* was included as out-group.

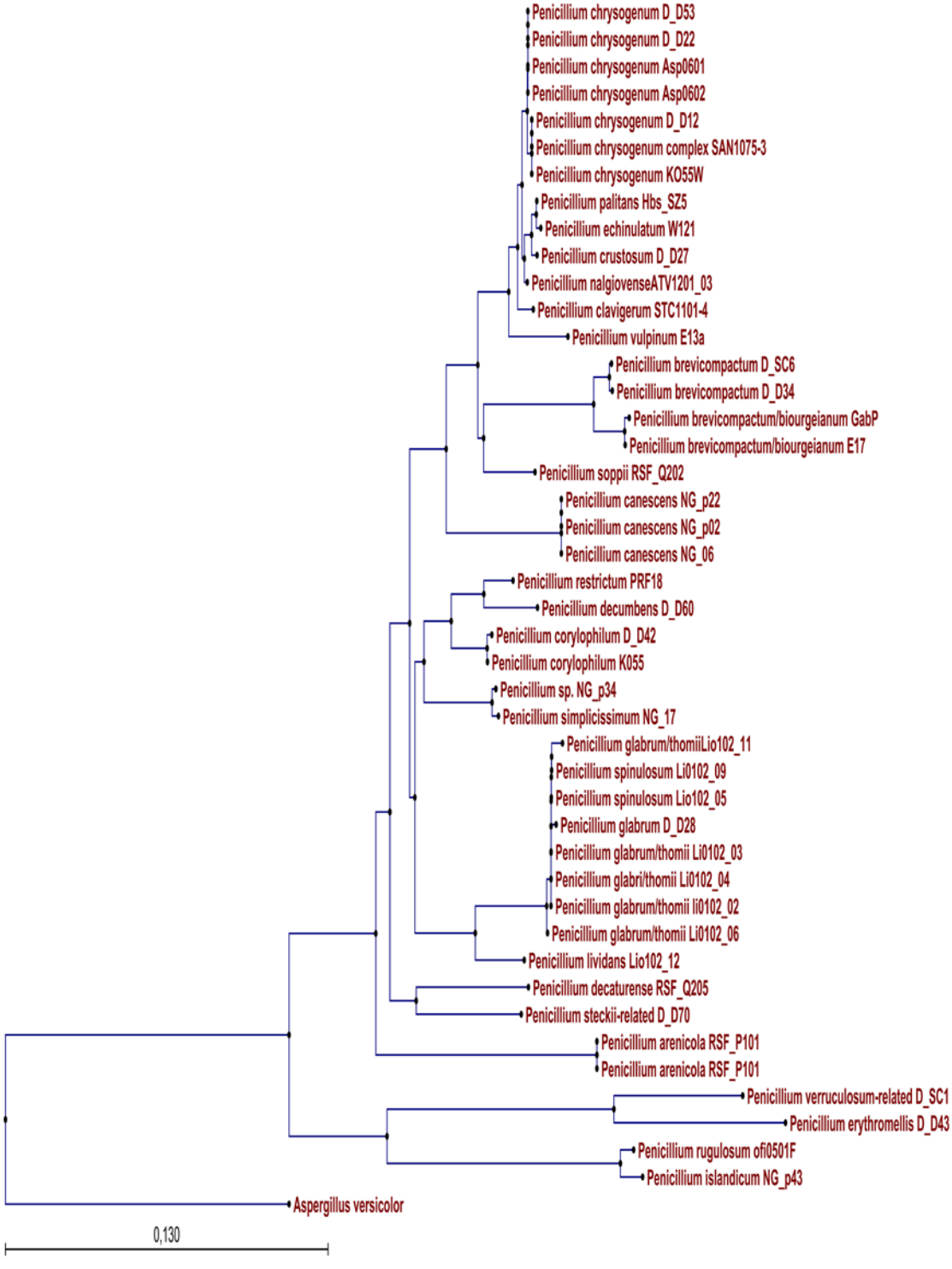

Fig.6 ITS-based phylogenetic tree of 53 *Penicillium* strains. The tree was constructed with the neighbor-joining method using CLC workbench. *Aspergillus versicolor* was included as the out-group.

5. Discussion and Conclusion

DNA-DNA hybridization protocols which permits the approximation of genomic similarities between two species [29] have long been applied in microbial identification. However, these are expensive, labor-intensive and only allow the detection of a few clusters [30]. Likewise, other molecular assays seem to have similar shortcomings.

A revolution in molecular microbial identification, especially in bacterial and fungal molecular systematics, started in the early 90s with the advent of PCR based amplification of RNA genes and sequencing [31, 32]. Thus, RNA genes, peculiarly the ribosomal RNA genes, started receiving a huge attention in phylogenetic studies of plants, animals and fungi [33,34]. Later studies identified certain intrinsic drawbacks in the rRNA genes e.g. its repetitive nature and, consequently, its high intra-specific mutation rate among multiple copies due to weak selective constraints [35] for example in *Penicillium*. Skoube et al. [36] showed that ribosomal internal transcribed spacer (ITS) sequences were relatively invariant in subgenus *Penicillium*, with only 29 variable sites providing a poorly resolved phylogram. Since then, more attention has been paid to certain conserved protein coding genes as they accumulate fewer mutations and are less variable among the few copies that most genomes contain [19]. Among these the one encoding for tubulins, especially for beta-tubulin, is receiving increasing attention in the investigation of evolutionary relationships at all levels: (i) in kingdom level phylogenetic analyses [23,24] and (ii) in studies of complex species groups within protists, animals, fungi and plants [25,27]. Since its 1st use in phylogentetic studies of *Epichloe* [37], beta-tubulin sequences have been used as putative phylogentic species marker in a variety of Ascomycetes and Hyphomycete genera [38]. The inclusion of organisms in phylogenetic analyses depend on reliable amplification of beta-tubulin genes [19].

As beta-tubulin genes are found in all eukaryotes, which makes this gene ideally suited for phylogenetic analysis, especially for the estimation of deep-level phylogenies [19]. In this study, beta-tubulin sequences of around 53 strains belonging to the genus *Penicillium* was determined to estimate the applicability of the beta-tubulin gene as a potential phylogentic marker.

The results of the study further substantiate the aforementioned conclusive evidences from multiple authors regarding the potentiality of beta-tubulin gene as a promising phylogentic marker.

Recent phylogentic-oriented studies have confirmed that *P. chrysogenum* – the source of the antibiotic penicillin - is composed of at least two distinct sub-species [10-12]. The tree inferred by this study supports this outcome. The thorough visual comparison of the two trees based on ITS and beta-tubulin shown in the figure 5 and 6 vividly distinguish the clear differences between the two markers. The beta-tubulin inferred tree seems to be more conclusive and well resolved as it elaborates the division of *P. chrysogenum* more clearly than the ITS based tree. Moreover, beta-tubulin based tree shows also a clear and well resolved classification of the *Penicillium spinulosum* than the ITS based tree. Taking these results into consideration it can be concluded that the beta-tubulin target gene is a potential phylogenetic marker for an accurate and rapid discrimination of *Penicillium* especially for *P. chrysogenum* and *P. spinulosum* using PCR based amplification and sequencing assay.

6. Future perspectives

The remaining 480 Ascomycota strains from the fungal genetics and genomics unit's collection will be classified phylogenetically using the beta-tubulin gene as marker of interest. This study will be then able to establish an improved classification of these important strains for their better use in the wide ranges of ongoing major projects at the fungal genetics research station in Tulln (Austria).

7. Acknowledgments

I want to thank my supervisors, Prof. Dr. Joseph Strauss (group leader) and Dr. Markus Gorfer (senior scientist) of the Fungal Genetics and Genomics Unit, BOKU Austria. Furthermore, I thank Dragana and Elisabeth (technicians) for their kind support in the laboratory work.

Special thanks to Madam Eva Dam, International Students Coordinator at the Biology Education Centre, Uppsala University, Sweden, for her administrative assistance.

I have to thank the whole Fungal Genomics Research Group for being so friendly and cooperative during the entire course of my stay.

Finally, special thanks to my family especially to Vicki and Stefan for their love and support.

8. References

[1] F. Lutzoni, F. Kauff, C.J. Cox, D. McLaughlin, G. Celio, B. Dentinger, M. Padamsee, D. Hibbett, T.Y. James, E. Baloch, M. Grube, V. Reeb, V. Hofstetter, C. Schoch, A.E. Arnold, J. Miadlikowska, J. Spatafora, D. Johnson, S. Hambleton, M. Crockett, R. Shoemaker, G.H. Sung, R. Lucking, T. Lumbsch, K. O'Donnell, M. Binder, P. Diederich, D. Ertz, C. Gueidan, K. Hansen, R.C. Harris, K. Hosaka, Y.W. Lim, B. Matheny, H. Nishida, D. Pfister, J. Rogers, A. Rossman, I. Schmitt, H. Sipman, J. Stone, J. Sugiyama, R. Yahr and R. Vilgalys Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits, Am J Bot 91 (2004) 1446-1480.

[2] J.P. Clapp, A.H. Fitter and J.P. Young Ribosomal small subunit sequence variation within spores of an arbuscular mycorrhizal fungus, Scutellospora sp, Mol Ecol 8 (1999) 915-921.

[3] T. Helgason, A.H. Fitter and J.P. Young Molecular diversity of arbuscular mycorrhizal fungi colonisin Hyacinthoides non-scripta (bluebell) in a seinatural woodland., Molec. Ecol. 8 (1999) 659-666.

[4] C.L. Schoch, K.A. Seifert, S. Huhndorf, V. Robert, J.L. Spouge, C.A. Levesque and W. Chen Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi, Proc Natl Acad Sci U S A 109 (2012) 6241-6246.

[5] S.L. Baldauf and J.D. Palmer Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins, Proc Natl Acad Sci U S A 90 (1993) 11558-11562.

[6] K. O´Donnell, E. Cigelnik and G.L. Benny Phylogenetic relationships among the Harpellales and Kickxellales, Mycologia 90 (1998) 624-639.

[7] M.R. Thon and D.J. Royse Evidence for two independent lineages of shiitake of the americas (Lentinula boryana) based on rDNA and beta-tubulin gene sequences, Mol Phylogenet Evol 13 (1999) 520-524.

[8] D. Begerow, B. John and F. Oberwinkler Evolutionary relationships among beta-tubulin gene sequences of basidiomycetous fungi, Mycol Res 108 (2004) 1257-1263.

[9] M.S. Barber, U. Giesecke, A. Reichert and W. Minas Industrial enzymatic production of cephalosporin-based beta-lactams, Adv Biochem Eng Biotechnol 88 (2004) 179-215.

[10] J. Böhm, B. Hoff, C.M. O'Gormana, S. Wolfers, V. Klix, D. Bingera, I. Zadrad, H. Kürnsteiner, S. Pöggeler, P.S. Dyer and U. Kück Sexual reproduction and mating-type–mediated strain development in the penicillin-producing fungus Penicillium chrysogenum, Proc Natl Acad Sci U S A 110 (2012) 1476-1481.

[11] D.A. Henk, C.E. Eagle, K. Brown, V.D.B. MA, P.S. Dyer, S.W. Peterson and M.C. Fisher Speciation despite globally overlapping distributions in Penicillium chrysogenum: the population genetics of Alexander Fleming's lucky fungus, Mol Ecol (2011).
[12] J. Houbraken, J.C. Frisvad and R.A. Samson Fleming's penicillin producing strain is not Penicillium chrysogenum but P. rubens, IMA Fungus 2 (2011) 87-95.
[13] H. Toju, A.S. Tanabe, S. Yamamoto and H. Sato High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples, PLoS One 7 (2012) e40863.
[14] T.J. Walsh, M.C. Wissel, K.J. Grantham, R. Petraitiene, V. Petraitis, M. Kasai, A. Francesconi, M.P. Cotton, J.E. Hughes, L. Greene, J.D. Bacher, P. Manna, M. Salomoni, S.B. Kleiboeker and S.K. Reddy Molecular detection and species-specific identification of medically important Aspergillus species by real-time PCR in experimental invasive pulmonary aspergillosis, J Clin Microbiol 49 (2011) 4150-4157.
[15] T. Sudjaritruk, T. Sirisanthana and V. Sirisanthana Immune reconstitution inflammatory syndrome from Penicillium marneffei in an HIV-infected child: a case report and review of literature, BMC Infect Dis 12 (2012) 28.
[16] C.J. Alexopoulos, C.W. Mims and M. Blackwell. Introductory Mycology. John Wiley and Sons, New York. 868p, (1996).
[17] K.B. Raper and D.I. Fennell The genus *Aspergillus*. Williams and Wilkins, Baltimore, (1965).
[18] K.G. Peay, P.G. Kennedy and T.D. Bruns Fungal community ecology: a hybrid beast with a molecular master., Bioscience 58:799–810 58 (2008) 799-810.
[19] S. Einax, Voigt, K. Oligonucleotide primers for te niversal amplification of beta-tubulin genes facilitate phylogenetic analyses in the regnum fungi, Org Divers Evol 3 (2003) 185-194.
[20] P.G. McKean, S. Vaughan and K. Gull The extended tubulin superfamily, J Cell Sci 114 (2001) 2723-2733.
[21] B.R. Oakley gamma-Tubulin, Curr Top Dev Biol 49 (2000) 27-54.
[22] M. Little, E. Krauhs and H. Ponstingl Tubulin sequence conservation, Biosystems 14 (1981) 239-246.
[23] P.J. Keeling and W.F. Doolittle Alphα-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. , Mol. Biol. Evol. 13 (1996) 1297-1305.
[24] S.L. Baldauf, A.J. Roger, I. Wenk-Siefert and W.F. Doolittle A kingdom-level phylogeny of eukaryotes based on combined protein data, Science 290 (2000) 972-977.

[25] M.A. Ayliffe, P.N. Dodds and G.J. Lawrence Characterisation of a beta-tubulin gene from Melampsora lini and comparison of fungal beta-tubulin genes. , Mycol. Res. 105 (2001) 818-826.
[26] P.J. Keeling, Luker, M. A. & Palmer, J. D. (2000): Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi., Mol Biol Evol 17 (2000).
[27] W. Mages, B. Cresnar, J.F. Harper, M. Brüderlein and R. Schmitt Volvox carteri alpha-2-tubulin-encoding and beta-2-tubulin-encoding genes: regulatory signals and transcription., Gene 160 (1995) 47-54.
[28] N.L. Glass and G.C. Donaldson Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes, Appl Environ Microbiol 61 (1995) 1323-1330.
[29] C.P. Kurztzman Yeast species recognition from gene sequence analysis and other molecular methods., Mycoscience 47 (2006) 65-71.
[30] D.S. Pontes, C.I. Lima-Bittencourt, E. Chartone-Souza and A.M. Amaral Nascimento Molecular approaches: advantages and artifacts in assessing bacterial diversity, J Ind Microbiol Biotechnol 34 (2007) 463-473.
[31] C.A. Petti Detection and identification of microorganisms by gene amplification and sequencing, Clin Infect Dis 44 (2007) 1108-1114.
[32] L.T. Wang, F.L. Lee, C.J. Tai and H. Kasai Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the Bacillus subtilis group, Int J Syst Evol Microbiol 57 (2007) 1846-1850.
[33] U. Englisch and S. Koenemann Preliminary phylogenetic analysis of selected subterranean amphipod crustaceans, using small subunit rDNA gene sequences. , Org. Divers. Evol. 1 (2001) 139–145.
[34] H.T. Lumbsch, I. Schmitt and M.I. Messuti Utility of nuclear SSU and LSU rDNA data sets to discover the ordinal placement of the Coccotremataceae (Ascomycota). , Org. Divers. Evol. 1 (2001) 99-112.
[35] K. Voigt, A. Matthäi and J. Wöstemeyer Phylogeny of zygomycetes: a molecular approach towards systematics of Mucorales. , Cour. Forsch.-Inst. Senckenberg 215 (1999) 207-213.
[36] P. Skouboe, M. Boysen, L.H. Pedersen, J.C. Frisvad and L. Rossen Identification of Penicillum species using the internal transcribed spacer (ITS) regions. In: Fungal identification techniques (Rossen L., Rubio V., Dawson M.T. Frisvad J.C, eds.). European Commission, Brussels, Belgium: 160-164, (1996).
[37] C.L. Schardl, A. Leuchtmann, K.R. Chung and D.S. Penny, M.R. Coevolution by common descent of fungal symbionts (Epichloe spp.) annd grass hosts. , Molecular Biology and Evolution 14 (1997) 133-143.

[38] R.A. Samson, K.A. Seifert, A.F.A. Kuijpers, J.A.M.P. Houbraken and J.C. Frisvad Phylogenetic analysis of *Penicillium* subgenus *Penicillium* using partial ß-tubulin sequences, Stud Mycol 49 (2004) 175-200.

9. Appendix

The appendix includes the materials used and their details.

1. DNeasy Plant mini kit(50) Cat No:69104, store at room temperature
2. QIA Quick PCR purification kit (250) Cat No: 28106 ,store at room temperature
3. CLC Genomic workbench (The latest version)
4. SIGMA lab mini Centrifuge (SN: 76986/Max speed=14000rpm/min) made in Germany.
5. Thermomixer-5437 (220 V, 50Hz, 107W, 0.5A) SN: 5437 00892, made in Germnany.
6. Clean Air. Laminar Hood
7. HYBRAID Ribolyser (Model No: FP120HY-230,SN: FP120HY-9L520163-4C)
8. Vortex GENE 2 (Scientific Industries,INC,Model No:G550E,50HZ,0.50 A) made in USA.
9. BIO RAD molecular image Chemi Doc with image lab software installed,(Model No: Universal Hood11,SN: 72113R04575) made in USA.Gel tank and related items from BIORAD as well. Gel red Stain (Lot:10G1112)
10. GALAXY MINI CENTRIFUGE.(VWR-International,SN:09040108) made in Korea, PCR tube Shaker-QUALITRON,INC(SN:01080659,230V,50HZ,0.5A) made in Korea.
11. Thermo-Scientific AB gene –PCR plates(48 wells-6x8) and Doomed Cap strips(AB-0602
12. Normal Autoclave Machine