ELSEVIER

FEMS Microbiology Letters 244 (2005) 139–148

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Detection of putative peptide synthetase genes in *Trichoderma* species: Application of this method to the cloning of a gene from *T. harzianum* CECT 2413

J.A. Vizcaíno [a], L. Sanz [a], R.E. Cardoza [a,b], E. Monte [a], S. Gutiérrez [b,*]

[a] *Spanish-Portuguese Center of Agricultural Research (CIALE), Departamento de Microbiología y Genética, Universidad de Salamanca, Edificio Departamental lab 208, Plaza Doctores de la Reina s/n, 37007 Salamanca, Spain*

[b] *Area de Microbiología, Escuela Superior y Técnica de Ingeniería Agraria, Universidad de León, Campus de Ponferrada, Avda. Astorga s/n, 24400 Ponferrada, Spain*

Received 9 November 2004; received in revised form 17 January 2005; accepted 20 January 2005

First published online 3 February 2005

Edited by L.F. Bisson

Abstract

Some of the secondary metabolites produced by *Trichoderma*, such as the peptaibols and other antibiotics, have a peptide structure and in their biosynthesis are involved proteins belonging to the Non-Ribosomal Peptide Synthetase family. In the present work, a PCR-mediated strategy was used to clone a region corresponding to an adenylation domain of a peptide synthetase (PS) gene from 10 different strains of *Trichoderma*. In addition, and using the fragment isolated by PCR from *T. harzianum* CECT 2413 as a probe, a fragment of 19.0 kb corresponding to a PS-encoding gene named *salps1*, including a 1.5 kb fragment of the promoter, was cloned and sequenced. The cloned region of *salps1* contains four complete, and a fifth incomplete, modules, in which are found the adenylation, thiolation and condensation domains, but also an additional epimerization domain at the C-terminal end of the first module. The analysis of the Salps1 protein sequence, taking into consideration published data, suggests that it is neither a peptaibol synthetase nor a protein involved in siderophore biosynthesis. The presence of two breaks in the open reading frame and the expression of this gene under nitrogen starvation conditions suggest that *salps1* could be a pseudogene.

Keywords: *Trichoderma*; Non-ribosomal peptide synthetase; Degenerated-PCR; Antibiosis; Pseudogene

1. Introduction

Trichoderma species are used as biological control agents of important plant-pathogenic fungi. Some species of this genus are active as mycoparasites and have been tested in field experiments and successfully shown to be effective against a range of economically important aerial and soil-borne plant pathogens [1].

Mycoparasitic strains of *Trichoderma* produce cell wall-degrading enzymes (CWDEs) and antibiotics. One of the main modes of action of *Trichoderma* is the production of a large variety of secondary metabolites: volatile (e.g. ethylene, hydrogen cyanide, alcohols, aldehydes and ketones up to C_4 chain-length), and non-volatile compounds, including peptide antibiotics (e.g. peptaibols) [2]. Peptaibols, a class of linear peptides with 5–20 residues, have three structural characteristics:

* Corresponding author. Tel.: +34 987 44 2060; fax: +34 987 44 2070.

E-mail address: degsgm@unileon.es (S. Gutiérrez).

0378-1097/$22.00
doi:10.1016/j.femsle.2005.01.036

(a) a high proportion of α,α′ dialkylated amino acids with an abundance of α-isobutyric acid (Aib), (b) a N-acyl terminus, usually acetyl, and (c) a C-terminal amino alcohol, such as phenyl-alaninol or leucinol [3,4]. Peptaibols generally exhibit antimicrobial activity against Gram-positive bacteria and fungi [5]. Peptaibols have been shown to act synergistically with CWDEs to inhibit the growth of fungal pathogens [6,7].

Another well-known mechanism of action of *Trichoderma* species is the competition for nutrients. Under iron starvation, as a potential means of competition in the soil, most fungi excrete at least one type of siderophore in order to solubilize the environmental iron. Fungal siderophores are typically short peptides containing also non-proteinogenic amino acids [8]. Iron-chelating siderophores have been described as biocontrol tools in some strains of *Trichoderma* [9], but they have not been characterized. In contrast, three types of hydroxamate siderophores have been described in *T. virens*: a monohydroxamate (*cis*- and *trans*-fusarinines), a dipeptide of *trans*-fusarinine (dimerum acid), and a trimer disdepsipeptide (copragen) [10].

Peptide antibiotics with an unusual amino acid content (like peptaibols) and siderophores, in bacteria and fungi, are often produced non-ribosomally by large multifunctional peptide synthetases (NRPSs). These large multifunctional enzymes assemble compounds from a wide range of precursors (including non-proteinogenic amino acids and hydroxy or carboxyl acids) [11].

NRPSs are organized into repetitive synthase units or modules, each of which has the functions required to complete a different single amino acid elongation step in the synthesis of the peptide product. Each module can be further partitioned into distinct adenylation (A), thiolation (T) and condensation (C) domains, which together represent a minimal repeating unit of NRPSs [12]. The functions of each unit include ATP-dependent activation to form a peptide bond (A), transfer of the acyl adenylates to specific thiols located in the enzyme-bound cofactors (4′-phosphopantetheine) (T), and condensation to form a peptide bond (C). The modification of the incorporated monomers (e.g. by epimerization or N-methylation) or the peptide backbone (e.g. by acylation, glycosilation or heterocyclization) can further functionalize the peptide product. These tailoring reactions are catalyzed by specialized domains or by fusion to polyketide synthase (PKS) modules [11,13–15]. In most NRPSs, the organization and order of the modules maps in a 1:1 manner to the amino acid sequence of the peptide products (co-linearity rule) [16].

In *Trichoderma* species, only one entire peptide synthetase gene (*tex1*) has been cloned and characterized (the largest NRPS so far), an 18-module peptaibol synthetase from *T. virens* [17]. A partial sequence (almost identical to *tex1*) from the final module of a putative siderophore synthetase has also been obtained from another *T. virens* strain [18].

The conservation of consensus sequence motifs in the domains provides a tool for detecting and cloning peptide synthetase genes in *Trichoderma* and other species using a degenerate PCR-based approach [17,18].

Analysis of the phenylalanine adenylation domain of the gramicidin synthetase, *Grs*A, has been used to determine the key residues responsible for A-domain specificity in substrate recognition. These have been called signature sequences [16,19]. It is hoped that a sufficiently large collection of verified signature sequences will provide a NRPS ''codon'' table allowing prediction of amino acids substrates based on the signature sequences in uncharacterized NRPSs.

In this article, we describe a method for the detection of peptide synthetase genes in different *Trichoderma* species through a PCR-based approach, and its application to the cloning of a partial peptide synthetase gene from the *T. harzianum* CECT 2413 (T34).

2. Materials and methods

2.1. Fungal isolates

Ten strains were used in this study: *Trichoderma asperellum* T3 (International Mycological Institute, Egham, UK, IMI 20179), *T. atroviride* T11 (IMI 352941), *T. harzianum* T14 (IMI 306222), *T. harzianum* T24 (IMI 352940), *T. asperellum* T25 (IMI 296237), *T. harzianum* T34 (Spanish Type Culture Collection, Valencia, Spain, CECT 2413), *T. atroviride* T35 (IMI 281112), *Trichoderma* sp. T37 (IMI 296235), *T. longibrachiatum* T44 (IMI 304058) and *T. longibrachiatum* T52 (NewBio-Technic, Sevilla, Spain, NBT52). Cultures were maintained on Potato Dextrose Agar (PDA, Difco) at 25 °C.

2.2. DNA and RNA manipulations

Mycelia for DNA extraction were grown in liquid cultures (200 rpm) at 28 °C in potato dextrose broth (PDB, Difco). Hyphae were collected by filtration, washed with distilled water, frozen and lyophilized. Fungal genomic DNA was isolated according to previously described protocols [20].

For Northern analysis, mycelia were grown in minimal medium (MM) [21] containing 2% glucose as carbon source (200 rpm) at 28 °C for 36–48 h. Then, the mycelia were harvested by filtration, washed with sterile water and placed into fresh MM containing different carbon sources: 2% glucose, 0% glucose for absence of carbon source, 1.5% chitin (*N*-acetylglucosamine polymer, Sigma), or 0.5% fungal cell walls from the strawberry pathogen *Colletotrichum acutatum*. Nitrogen starvation conditions corresponded to a 100-fold decrease in the

concentration of ammonium sulphate in the medium (50 mg l^{-1}) containing 2% glucose as carbon source. Mycelia were collected by filtration, thoroughly washed with sterile water, lyophilized and kept at −80 °C until RNA extraction. Fungal cell walls from the strawberry pathogen *C. acutatum* IMI 364856 were prepared as previously reported [22].

Fungal RNA was isolated using TRIZOL® reagent (Life Technologies), according to the manufacturer's instructions. Southern blotting and Northern blotting were performed by standard techniques [23]. RT-PCR was performed using Superscript™ RNAse H^- (Invitrogen), following the manufacturer's instructions. DNA sequencing was carried out by the Sanger method [24] using standard automated fluorescence techniques.

2.3. PCR conditions

Primers MTF2 (GCNGGYGGYGCNTAYGTNCC) and MTR (CCNCGDATYTTNACYTG) [25] were designed according to adenylation domain motifs A2 and A8 [11] (Fig. 1). PCR thermal cycling protocol included an initial denaturation at 94 °C for 2 min, followed by 35 cycles at 93 °C for 10 s, at 52 °C for 20 s, and at 72 °C for 1 min. Each PCR reaction contained 0.5 μl 10 mM dNTPs, 1 μl of each primer (100 μM), 2.0 mM $MgCl_2$ and 1.5 U DNA polymerase (Biotools) in a 25-μl reaction. Primers TRIPS1 (TCCRCGGATYTDACCTGSKTATCC) and TRIPS2 (TGCRGGYGGTGCITAYGTKCC), derived from the same motifs, were designed later from sequences obtained from the *Trichoderma* probes. PCR conditions were identical for both sets of primers. DNA bands were excised from agarose gels using the Gene-Clean® II system (Qbio Gene) according to the manufacturer's instructions, cloned in vector pGEM-T® Easy (Promega) and sequenced in both strands.

2.4. Sequence analysis

Sequences were analysed using the Lasergene package (DNASTAR). DNA-binding elements were found by looking for consensus sequences described elsewhere or with the help of the MatInspector ver. 6.2.1 program.

The nucleotide sequence of *salps1* was deposited in the GenBank database under the Accession No. AJ784403. The sequences of the probes were also deposited under the following Accession Nos.: strain T3: AJ784985, T11: AJ784986, T14: AJ784987, T24: AJ784988, T25: AJ785295, T35: AJ784990, T37: AJ784981, T44: AJ784989 and T52: AJ784984.

3. Results

3.1. Degenerate PCR

The degenerate primers MTF2 and MTR, useful in cyanobacteria [25], corresponding to portions of conserved motifs A2 and A8 from the adenylation domains (Fig. 1), were used to perform a PCR with genomic DNA from 10 different strains of *Trichoderma*. DNA fragments that were about 1000 bp long, which was consistent with the expected size, were cloned and sequenced. Probes with clear similarity with known peptide synthetase genes in the databases were found. Positive bands were obtained from the strains *T. asperellum* T3 (988 bp), *T. asperellum* T25 (1009 bp), *T. harzianum* T34 (1008 bp) and *T. longibrachiatum* T44 (996 bp), but many false-positive bands were also obtained for the other strains.

In order to obtain the probes more easily, the less degenerate primers TRIPS1 and TRIPS2, corresponding to the same conserved motifs (Fig. 1), were designed based exclusively in the sequences obtained previously from *T. asperellum* T3 and T25, *T. harzianum* T34 and *T. longibrachiatum* T44. The degeneracy of the primers was lowered from 512 (MTF2) to 48 (TRIPS1) and from 192 (MTR) to 64 (TRIPS2). Again, DNA fragments that were about 1000 bp long were cloned and sequenced. Positive bands with high similarity with peptide synthetases were obtained from the other studied strains: *T. atroviride* T11 (987 bp), *T. harzianum* T14 (1027 bp), *T. harzianum* T24 (989 bp), *T. atroviride* T35 (992 bp), *Trichoderma* sp. T37 (987 bp) and

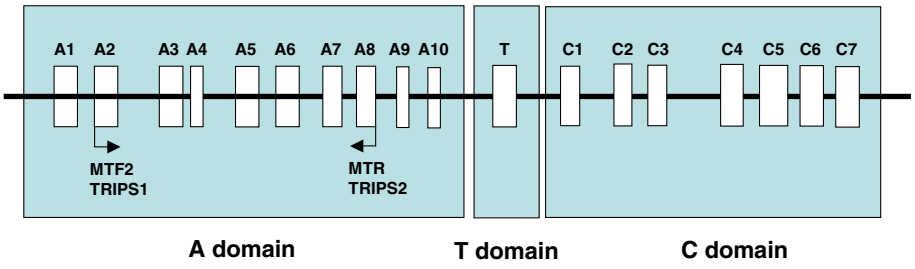

Fig. 1. Linear schematic structure of a peptide synthetase module, including the regions (Boxes) of similarity to consensus sequence motifs of peptide synthetases [11]. Arrows indicate the relative positions of the degenerate primers used in this study. A: Adenylation domain, T: Thiolation domain, C: Condensation domain.

T. longibrachiatum T52 (978 bp) (Fig. 2). No false-positive bands were obtained in this case.

The percentage of similarity within the amino acid sequences of the probes was from 32.6% (between T34 and T37) to 99.1% (between T11 and T35). The similarity between the bands obtained with the primers TRIPS1 and TRIPS2 was generally higher (e.g. between T11 and T35, as cited previously). Conserved motifs that are present in peptide synthetase adenylation domains are present in each of the modules [16] (Fig. 2).

In order to determine whether more than one module had been amplified in the same PCR reaction, 10 different cloned DNA fragments from *T. longibrachiatum* T52 were sequenced and three different fragments with high similarity to peptide synthetase modules were found (data not shown).

3.2. Cloning of salps1, a gene from T. harzianum CECT 2413 encoding a peptide synthetase

The 1008 bp band isolated from *T. harzianum* CECT 2413 (T34) was used as a probe to screen a lambda genomic DNA library. This strain was chosen because it has been considered as a reference organism in many studies [26–30]. A chromosome walking strategy was followed, and three overlapping bands from different positive phages were cloned in pBluescript® (Stratagene) and sequenced (Fig. 3): A *Bam*HI fragment that included 5.7 kb of the Open Reading Frame (ORF) and 1.5 kb upstream the gene, an *Eco*RI fragment of 8.9 kb, and a *Sal*I fragment of 5.7 kb. A total of 19.0 kb (17,534 bp of ORF and 1461 bp of the promoter) of DNA were sequenced in both strands. A partial protein sequence of 5815 amino acids, named Salps1, was deduced. The original DNA fragment obtained by PCR was included in the first adenylation domain of Salps1. The whole sequence, corresponding to the N-terminal end of Salps1 showed similarity to four complete and a fifth partial peptide synthetase modules.

Adenylation, thiolation and condensation domains are present in each module, although the cloned fragment ends before the thiolation domain of the fifth module. In the C terminus of the first module, an epimeration domain is also present (Fig. 4). The percentage similarity of the amino acid sequences in the A domains was between 33.0% (A3 and A5) and 47.4% (A1 and A2), in the C domains between 20.7% (C2 and C4) and 41.0% (C3 and C4), and in the T domains between 40.9% (T1 and T4) and 52.3% (T1 and T3). Conserved motifs from peptide synthetase domains are present in each of the modules [11], excepting C4–C7 from condensation domains (Fig. 5).

	A2	A3	A4	A5
T3	**AGGAYVPLD**	**L**YAVLF**TSGSTGK**P**KG**	**FD**IG	GL**YGPAE**
T11	**AGGAYVPLD**	**L**YAVLF**TSGSTGK**P**KG**	**FD**IG	GL**YGPAE**
T14	**AGGAYVPLD**	A**AY**VLF**TSGSTGA**P**KG**	**FD**AC	**N**CW**GPVE**
T24	**AGGAYVPLD**	**L**YAVLF**TSGSTGK**P**KG**	**FD**IG	GL**YGPAE**
T25	**AGGAYVP**AR	V**AY**LLF**TSGSTGV**P**KG**	**FDL**S	**N**GW**GPAE**
T34	**AGGAYVPLD**	A**AY**LIF**TSGSTGE**P**KG**	**FDV**S	VT**YGP**SR
T35	**AGGAYVPLD**	**L**YAVLF**TSGSTGK**P**KG**	**FD**IG	GL**YRPAE**
T37	**AGGAYVPLD**	A**AY**VLF**TSGSTGK**P**KG**	**FDV**S	**NGYGPTE**
T44	**AGGAYVPLD**	**L**YAVLF**TSGSTGK**P**KG**	**FD**IG	GL**YGPAE**
T52	**AGGAYVPLD**	ASVVLF**TSGSTGK**P**KG**	**FD**VG	GL**YGPAE**
Cons	AGxAYVPID	LAYxxYTSGTTGxPKG	FDxS	NxYGPTE
	L	S		

	A6	A7	A8
T3	**GEL**LIQSPI**LAYGYL**	**YRTGDL**	**GRKDNQVKIRG**
T11	**GEL**VIQSPI**LAYGYL**	**YLTGDL**	S**RKDNQVK**SA
T14	**GE**VLI**Q**G**P**TLF**RE**Y**L**	**YKTGDL**	**GRKDTQVK**SV
T24	**GEL**VIQSPI**LAYGYL**	**YLTGDL**	S**RKDTQVK**SVE
T25	**GE**IVM**Q**G**P**TLL**RE**Y**L**	**YKTGDL**	**GRKDTQVKIRG**
T34	**GEL**LI**E**G**P**IV**GR**GY**L**	**YKTGDL**	**GRKDTQVKIRG**
T35	**GEL**VIQSPI**LAYGYL**	**YLTGDL**	S**RKDK**Q**VK**SAE
T37	**GEL**VVQ**G**YTL**SR**GYI	**YLTGDL**	**GRKDNQV**RSA
T44	**GEL**LIQSPI**LAYGYL**	**YRTGDL**	**GRKDNQVKIRG**
T52	**GEL**LI**E**G**P**M**LARGYL**	**YRTGDL**	**GRKDTQV**RSA
Cons	GELxIxGxGVARGYL	YKTGDL	GRxDxQVKIRG
	L	R	

Fig. 2. Alignment of the adenylation domains obtained by degenerate PCR from the *Trichoderma* strains used in this work, and comparison with the highly conserved motifs defined by Marahiel et al. [11]. Identities with these consensus sequences are indicated in bold face.

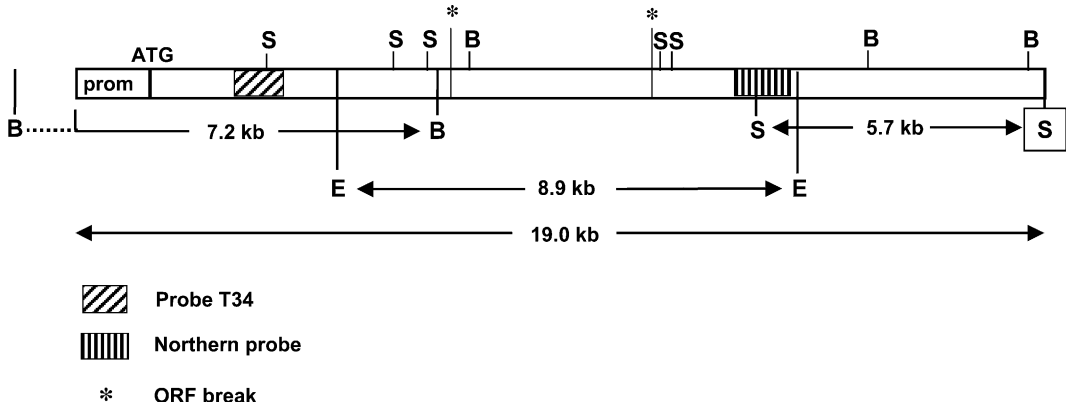

Fig. 3. Restriction map of the *salps1* cloned region. The *Bam*HI (B), *Eco*RI (E) and *Sal*I (S) fragments that were sequenced are shown. The locations of the initial 1008 bp probe obtained by degenerate PCR and the 1120 bp probe that was used in the Northern experiment are indicated in boxes. The initial ATG and the two breaks in the ORF are also indicated.

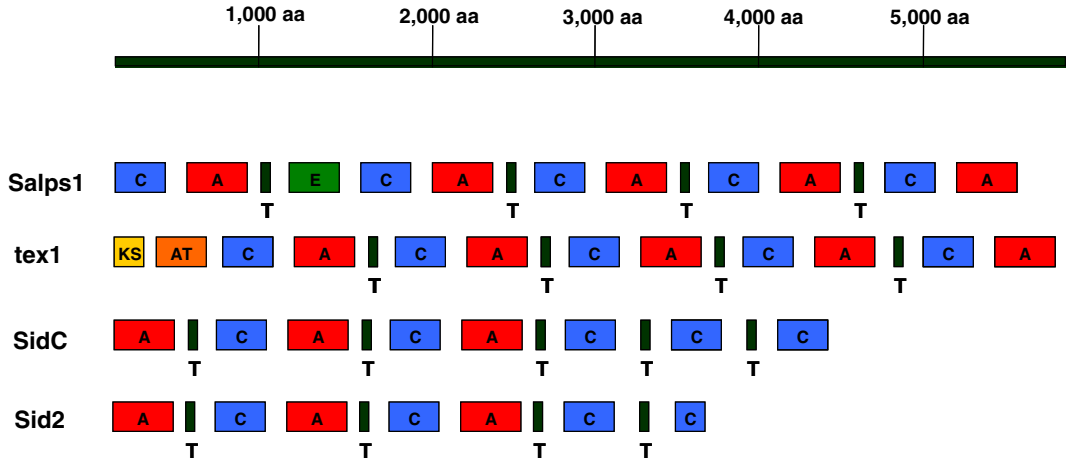

Fig. 4. Modular structure of the Salps1 protein, the first five modules of tex1 from *T. virens* [17], SidC from *Aspergillus nidulans* [34] and Sid2 from *Ustilago maydis* [35]. Domain names are indicated (C, condensation; A, adenylation; T, thiolation; E, epimerization; KS, ketoacyl synthase; AT, acyltransferase).

Two breaks in the ORF of *salps1* were indicated by the presence of premature stop codons (Fig. 3) at positions 5885 and 9781. In order to determine if the breaks were caused by the presence of introns, PCR primers anchored on presumed exons were designed and RT-PCR was performed, using mRNAs from nitrogen starvation conditions as template. Around the first putative stop codon, an 85 bp intron was found, from position 5767–5851. This intron was characterized by typical fungal 5′ splice (GTAAATC), 3′ splice (ATAG) and internal sequences (AACTGAC) [31]. Thus, the first break in the ORF appeared just after the intron at position 5885. No intron was found around the second break in the ORF (position 9781).

3.3. Sequence analysis of the salps1 promoter

About 1.5 kb of the gene promoter were sequenced and investigated for the occurrence of protein-binding motifs that may give hints as to its regulation. The *salps1* promoter contains two putative TATA-boxes at −360 and −299, and two putative CCAAT boxes at −1143 and −1010. In addition, two consensus motifs could be identified for the binding of the carbon catabolite repressor CreA/Cre1 (SYGGRG at −766 and −730) [32] and eight (four in each DNA strand) were also identified for the *A. nidulans* global nitrogen regulator AreA/Nit2 (HGATAR) [33].

3.4. Expression experiments

Owing to the presence of consensus sequences in the *salps1* promoter for potential binding of the carbon catabolite regulator CreA/Cre1 and the global nitrogen regulator AreA/Nit2, the regulation of *salps1* expression by carbon and/or nitrogen sources was investigated. This analysis was challenging due to the large size of the *salps1* transcript. Thus, samples of *T. harzianum*

Adenilation domains

	A1	A2	A3	A4	A5
A1	LTYRQL	LKAGGAFVPLD	AAYLIFTSGSTGEPKG	FDVS	VTYGPSR
A2	ATYAEL	LKSGGAFVPLD	AAYMIFTSGSTGEPKG	FDAS	NGYGPTE
A3	ISYKEL	LKLAAPLCHLI	AAYMLFTSGSTGEPKG	FDAC	NGYGPTE
A4	LTYAEL	LKTGAAFVMLE	TMYVIFTSGSTGKAKG	FDAS	VTYGPSE
A5	LTYAQL	MMAGGAIVPID	TAYIMFTSGSDGQPKG	LDVS	NAYRHTE
Cons	LTYxEL	LKAGxAYVPID	LAYxxYTSGTTGxPKG	FDxS	NxYGPTE
	S	L L	S		

	A6	A7	A8	A9	A10
A1	GELLIEGPIVGRGYL	YKTGDL	GRKDTQVKIRGQRVELGEVE	LPAYMIP	SGKTDR
A2	GELLIEGPIVGRGYH	YKTGDL	GRKDNQVKIRGQRAELGEIE	LPSYMVP	SGKIDR
A3	GELVVEGPIVGRGYL	YKTGDL	GRKDNQVKIRGQRIELGEVE	LPGYMMP	SGKIDR
A4	GELLLEGPIVGKGYL	YKTGDL	GRKDNQVKIRGQRVELGEVE	LPKYMDP	SGKTDR
A5	GELVIEGPVLAQGYL	YKTGDL	GRRDTRVTIRGQRVDLAEVE	LPAHMIP	SGMVDL
Cons	GELxIxGxGVARGYL	YKTGDL	GRxDxQVKIRGxRIELGEIE	LPxYMIP	NGKLDR
	L	R		V	V

Condensation domains

	C1	C2	C3
C1	TPIQQGLLSL	RMPILRTRI	IHHALYDGWS
C2	SPMQEGMLIS	RHGLLRALI	MNHAISDGYT
C3	TALQEGLMAL	SNPILRTRI	IHHALYDGWS
C4	TALQEGLVAL	MTPVLRTRI	IHHALYDGWS
C5	TPLQEGLLSL	MLPILRTRI	VHHVLYDGQF
Cons	SxAQxRLWxL	RHExLRTxF	MHHxISDGWS
	MY		V

Thiolation domains

T1	DSFFRLGGDSI
T2	DNFFLLGGQSI
T3	DSFFHLGGDSI
T4	DSFFHLGGDSI
Cons	DxFFxxGGHSI
	D L

Epimerization domain

	E1	E2	E3	E4	E5	E6	E7
E1	PIQKLY	HHLVVDLMS	EQLLTAGS	EGHGRE	RTVGWFTTVFPVQLA	SSNGWLYF	FNYVGA
Cons	PIQxWF	HHxISDGWS	DxLLxAxG	EGHGRE	RTVGWFTxxYPYPFE	PxxGxGYG	FNYLGR
		V			V		Q

Fig. 5. Alignment of Salps1 domains with the conserved core motifs defined by Marahiel et al. [11] for these domains. Identities with these consensus sequences are indicated in bold face. The domains are indicated as follows: A, adenylation; C, condensation; T, thiolation, E, epimeration. Each domain is numbered in order of appearance in the protein.

T34 mycelium grown at different times (4, 9 and 24 h) were collected for Northern analysis. A positive hybridization signal was observed under nitrogen starvation conditions at 9 and 24 h (Fig. 6). An estimation of the transcript size was performed and it was about 30 kb.

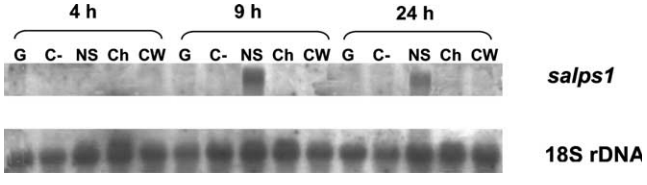

Fig. 6. Northern blot analysis of *salps1* gene expression. The experiment was carried out with total RNA (20 μg) extracted from mycelia of *T. harzianum* CECT 2413 grown in MM [21] under the following conditions: 2% glucose (G); absence of carbon source (0% glucose) (C); nitrogen starvation (ammonium sulphate 50 mg ml^{-1}); 1.5% chitin (Ch); 0.5% fungal cell walls (CW). Mycelia were cultivated for 4, 9 and 24 h. A 1120 bp PCR fragment was used as the probe (see Fig. 3). Radish 18S rDNA was used as loading control.

No transcript was detected in mycelia cultivated with 2% glucose or in the absence of carbon source (0% glucose).

It was also of interest to define whether the induction occurred in the presence of fungal cell walls from *C. acutatum* or chitin as carbon sources, in order to simulate a mycoparasitism experiment. As shown in Fig. 6, no signals were detected with either.

3.5. Signature sequences

The comparison of the signature sequence residues found in the adenylation domains included in the probes obtained by degenerate PCR and also in the remaining four adenylation domains of Salps1 (Table 1), showed that all the domains had the expected aspartate residue at position 235. All signature sequences obtained in this study were unique and did not exactly match other signature sequences found in other characterized NRPSs. Modules from *T. atroviride* T11, *T. harzianum* T24

Table 1
Signature sequences of putative adenylation domains of the *Trichoderma* fragments obtained by degenerate PCR and from Salps1 [19]

Module	Signature sequence position							
	235	236	239	278	299	301	322	330
T3	D	I	P	F	N	G	L	I
T11	D	I	L	I	C	A	L	I
T14	D	A	T	L	L	G	C	V
T24	D	I	L	I	C	A	L	I
T25	D	L	G	F	L	A	G	V
T34 (=Salps1 A1)	D	V	Q	L	V	G	T	H
T35	D	I	L	I	C	A	L	I
T37	D	V	A	T	V	A	G	M
T44	D	I	L	F	N	G	L	K
T52	D	V	T	F	N	G	L	I
Salps1 A2	D	A	M	F	I	G	N	I
Salps1 A3	D	A	M	F	I	G	L	E
Salps1 A4	D	A	E	D	I	G	T	P
Salps1 A5	D	V	L	I	I	A	A	M

Each adenylation domain is numbered in order of appearance in the protein.

and *T. atroviride* T35 had identical residues in all the positions, indicating that the fragment that had been cloned in those strains probably corresponded to an equivalent module of a PS gene.

4. Discussion

4.1. Degenerate PCR

We have developed a useful method to detect peptide synthetase genes in *Trichoderma* using a degenerate PCR based approach. Two sets of degenerate primers were used. However, as the degeneracy of the new set of primers (TRIPS1 and TRIPS2, based exclusively on *Trichoderma* sequences) was lower, the sequence similarity between these modules was higher, and it was easier to obtain the probes.

4.2. T. harzianum *CECT 2413 PS gene*

Four complete and one partial module from the N terminus of a PS gene (*salps1*) were cloned and sequenced. The known conserved motifs in the A, C and T domains were identified in each module and, interestingly, an epimeration domain appeared at the C terminus of the first module, which also contained the known conserved motifs [11] (Fig. 5). If the co-linearity rule module-amino acid is followed [16], the first amino acid of the final peptide product should be a D-amino acid.

The modular architecture of the N-terminal end of the Salps1 protein may suggest that this enzyme does not make peptaibols. This conclusion is based in the fact that in the N-terminal end of the peptaibol synthetase *tex1*, there have been found two typical domains of polyketide synthases, a ketoacyl synthase domain and an acyltransferase domain, which seem necessary for the typical N-terminal acetylation (acylation) of the peptaibols [17] (Fig. 4). Besides, an epimeration domain was found in the first module and to our knowledge, peptaibols with a D-amino acid in the N-terminus have been not described.

The modular architecture of Salps1 may suggest that this enzyme does not make siderophores either. Two PS genes that make fungal siderophores are known: *Sid2* from *Ustilago maydis* and *SidC* from *Aspergillus nidulans*. Both enzymes are predicted to comprise three complete modules. At the C-terminus, Sid2 contains one additional T domain and a partial C domain, while SidC includes two additional T–C domain units [34,35] (Fig. 4). The modular structures of Sid2 and SidC suggest that these enzymes synthesize tripeptides but they seem to be involved in ferrichromes-like siderophores synthesis and ferrichromes are made from six amino acids. It might be possible that these enzymes are responsible for the formation of the complete hexapeptides, via repeated use of one or more modules, as should be indicated by the presence of additional C-terminal domains [36]. The structure of Salps1 is quite different to Sid2 and SidC in the five known modules. Besides, to our knowledge, only siderophores up to three amino acids have been detected in *Trichoderma* species [10], and a gene of at least five modules would be too large for making siderophores derived from one to three amino acids. Although they have not been detected so far, ferrichromes could also be present in *Trichoderma* species, but again, the modular structure of Salps1 is quite different from Sid2 and SidC.

For the same reason, it seems that Salps1 is not involved in the synthesis of other secondary metabolites derived from amino acids found in *Trichoderma* species, such as diketopiperazines. These are compounds derived from cyclic dipeptides that arise by condensation of two α-amino acids, like the antibiotics gliotoxin or the more complex gliovirin [37].

Besides the "normal" positioning of the C domain between two modules mediating peptide-chain elongation, an extra C domain is found in the N-terminal end of Salps1. This is not very common and it has been found in several PS systems, such as at the amino terminus of cyclosporin synthetase [38] or at the carboxyl terminus of the enniatin [39] and HC-toxin [40] systems. According to the organization and the structure of the formed products, it can be concluded that these C-domains are probably involved in peptide-chain termination and cyclization [41]. In other bacterial PS systems, such as lichenysin [42], surfactin [43] or fengycin [44], the presence of an additional C domain located in the N-terminal end has been correlated with the fact that the first amino acid of the product peptide is acylated

with a fatty acid, indicating that the first C domain of each of these systems is responsible for the linking of a fatty acid instead of a peptidyl moiety to an amino acid. These findings make us think that the final peptide product of Salps1 could be acylated of cyclized.

In order to determine if an identical modular structure to Salps1 could be present in other fungal PS genes, we compared the modular structure of Salps1 with all fungal characterized PS present in the GenBank database. No identities were found. We also compared it with all the putative PS genes found in several fungal complete genomes, considering those genes that are at least as long as *salsp1* (17.5 kb). Hypothetical proteins from *Aspergillus nidulans* (Accession Nos.: EAA65335, EAA65835, EAA64650 and EAA59538), *Fusarium graminearum* (EAA69855, EAA69381, EAA69816 and EAA75314), and *Magnaphorte grisea* (EAA54366) were studied. Nevertheless, the lack of identity between the modular structures of Salps1 and these proteins suggest that the final product of *salps1* could be a peptide or a secondary metabolite only present in *Trichoderma*.

4.3. Signature sequences

Eight or 10 residues present in the active site of peptide synthetases have been proposed to play a major role in defining substrate specificity for incorporation of amino acids based on structural data [16,19]. These residues define the signature sequences specifying amino acid incorporation. The signature sequences [19] from all the modules do not exactly match the other signature sequences found in other characterized NRPSs (Table 1). All the modules have the expected aspartate residue at position 235 and these residues are invariant at this position for modules incorporating amino acids, differing only in modules that incorporate carboxylic acids [45]. As signature sequences are identical in probes from *T. atroviride* T11, *T. harzianum* T24 and *T. atroviride* T35, these domains seem to activate the same substrate. Further effort will have to be made in order to increase the non-ribosomal code available data, and much more if we consider the increasing amount of available sequence data from putative PS genes from different fungi or bacteria genome sequencing projects.

The genomes of several filamentous fungi have been fully sequenced and annotated. It is surprising for us the high number of putative PS genes that have been found in the fungal complete genomes (http://www.broad.mit.edu/annotation/fungi): e.g. 11 in *F. graminearum*, 15 in *A. nidulans* and 10 in *M. grisea*. Initially, our aim was to get probes from *Trichoderma* species to clone a peptide synthetase gene with a known activity (a peptaibol synthetase). If the number of PS genes was low, as we expected, degenerate PCR would be a good alternative, due to the high number of conserved motifs in PS genes. But, if it is expected to find a high number of putative PS genes, as it seems to be the case for *Trichoderma* and other filamentous fungi, the cloning of a PS encoding gene with a specific activity, using a degenerate PCR strategy, would be more complicated and could be simply a matter of luck.

4.4. Expression

salps1 has two breaks in the ORF, and thus it may be considered a pseudogene. Pseudogenes have been defined as non-functional sequences of genomic DNA originally derived from functional genes [46]. There is no available data about the pseudogene population in the fungal genomes. In *S. cerevisiae*, there are 183 disabled ORFs (dORFs), corresponding to less than 3% of the proteome [47]. In the same study, a microarray analysis showed that some dORFs are expressed even though they carry multiple disablements, as could be the case of *salps1*. It is intriguing that expression of these dORFs can be still detected, suggesting that these sequences at least possess functional promoters, and that their expression can be still detectable despite nonsense-mediated decay [48].

An 85 bp intron was found around the first break of the ORF. Introns have been found in other fungal peptide synthetases [49–51], but no introns were found in *tex1*, the largest NRPS so far, a 62.8 kb peptaibol synthetase from *T. virens* [17].

salps1 expression was triggered by incubation with low concentration of nitrogen sources. There are eight AreA/Nit2 consensus-binding sites (HGATAR) in the *salps1* promoter, and nitrogen-regulated genes usually contain several copies of this motif [52]. This observation is compatible with the regulation of *salps1* by nitrogen source. Up to our knowledge, regulation by nitrogen depletion has not yet been reported for a peptide synthetase. In fact, there are few available data about expression of fungal peptide synthetases. The expression of *AbrePsy1*, a 22.0 kb peptide synthetase gene with an unknown function from *Alternaria brassicae*, was investigated recently and its expression was not found to be inducible upon nitrogen starvation [49].

No *salps1* expression was found in 2% glucose or in absence of carbon source (0% glucose) (Fig. 6). The *salps1* promoter contains two CreA/Cre1 consensus-binding sites, organized by direct repeats [53,54]. Theoretically, this observation would be compatible with a regulation of *salps1* by glucose repression [55], but as no expression was found in absence of glucose either, this seems not to be the case.

No signals were found when the mycelia were grown in chitin or using the plant pathogen *C. acutatum* cell walls as carbon sources. *salps1* does not seem to be involved in the mycoparasitic process.

Here we report a method to detect peptide synthetase genes in *Trichoderma* species and the application of this

method to the partial cloning and initial characterization of a PS gene from the strain *T. harzianum* CECT 2413.

Acknowledgements

The authors acknowledge the financial support of the Spanish Foundations "Ramón Areces" (Madrid, Spain) and "Fundación Andaluza de I+D" (Seville, Spain). This research was also supported by the European Commission (project QLK3-CT-2002-02032). We recognize the work carried out by I. Chamorro, M. Sánchez and J.A. de Cote for their technical support.

References

[1] Elad, Y. (2000) Biological control of foliar pathogens by means of *Trichoderma harzianum* and potential modes of action. Crop Prot. 19, 709–714.
[2] Keszler, A., Forgacs, E., Kotai, L., Vizcaíno, J.A., Monte, E. and García-Acha, I. (2000) Separation and identification of volatile components in the fermentation broth of *Trichoderma atroviride* by solid-phase extraction and gas chromatography-mass spectrometry. J. Chromatogr. Sci. 38, 421–424.
[3] Rebuffat, S., Goulard, C., Bodo, B. and Roquebert, M.F. (1999) The peptaibol antibiotics from *Trichoderma* soil fungi; structural diversity and membrane properties. Recent Res. Devel: Org. & Bioorg. Chem. 3, 65–91.
[4] Chugh, J.K. and Wallace, B.A. (2001) Peptaibols: Models for ion channels. Biochem. Soc. Trans. 29, 565–570.
[5] Jen, W.C., Jones, G.A., Brewer, D., Parkinson, V.O. and Taylor, A. (1987) The antibacterial activity of alamethicins and zervamicins. J. Appl. Bacteriol. 63, 293–298.
[6] Schirmbock, M., Lorito, M., Wang, Y.L., Hayes, C.K., Arisan-Atac, I., Scala, F., Harman, G.E. and Kubicek, C.P. (1994) Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. Appl. Environ. Microbiol. 60, 4364–4370.
[7] Lorito, M., Woo, S.L., D'Ambrosio, M.D., Harman, G.E., Hayes, C.K., Kubicek, C.P. and Scala, F. (1996) Synergistic interaction between cell wall degrading enzymes and membrane affecting compounds. Mol. Plant–Microbe Interact. 9, 206–213.
[8] Leong, S.A. and Winkelmann, G. (1998) Molecular biology of iron transport in fungi. Met. Ions Biol. Syst. 35, 147–186.
[9] Chet, I. and Inbar, J. (1994) Biological control of fungal pathogens. Appl. Biochem. Biotechnol. 48, 37–43.
[10] Jalal, M.A., Love, S.K. and van der Helm, D. (1986) Siderophore mediated iron (III) uptake in *Gliocladium virens*. 1. Properties of *cis*-fusarinine, *trans*-fusarinine, dimerum acid, and their ferric complexes. J. Inorg. Biochem. 28, 417–430.
[11] Marahiel, M.A., Stachelhaus, T. and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. 97, 2651–2674.
[12] Kleinkauf, H. and von Dohren, H. (1996) A nonribosomal system of peptide biosynthesis. Eur. J. Biochem. 236, 335–351.
[13] von Dohren, H., Keller, U., Vater, J. and Zocher, R. (1997) Multifunctional peptide synthetases. Chem. Rev. 97, 2675–2706.
[14] Cane, D.E., Walsh, C.T. and Khosla, C. (1998) Harnessing the biosynthetic code: Combinations, permutations, and mutations. Science 282, 63–68.
[15] Walsh, C.T. (2004) Polyketide and nonribosomal peptide antibiotics: Modularity and versatility. Science 303, 1805–1810.
[16] Stachelhaus, T., Mootz, H.D. and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. 6, 493–505.
[17] Wiest, A., Grzegorski, D., Xu, B.W., Goulard, C., Rebuffat, S., Ebbole, D.J., Bodo, B. and Kenerley, C. (2002) Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase. J. Biol. Chem. 277, 20862–20868.
[18] Wilhite, S.E., Lumsden, R.D. and Straney, D.C. (2001) Peptide synthetase gene in *Trichoderma virens*. Appl. Environ. Microbiol. 67, 5055–5062.
[19] Challis, G.L., Ravel, J. and Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. 7, 211–224.
[20] Raeder, U. and Broda, P. (1985) Rapid preparation of DNA from filamentous fungi. Lett. Appl. Microbiol. 1, 17–20.
[21] Penttilä, M., Nevalainen, H., Ratto, M., Salminen, E. and Knowles, J. (1987) A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. Gene 61, 155–164.
[22] Fleet, G.H. and Phaff, H.J. (1974) Glucanases in *Schizosaccharomyces*. Isolation and properties of the cell wall-associated beta(1 leads to 3)-glucanases. J. Biol. Chem. 249, 1717–1728.
[23] Sambrook, J. and Russel, D.W. (2001) Molecular Cloning. Cold Spring Harbor Laboratory Press, New York.
[24] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[25] Neilan, B.A., Dittmann, E., Rouhiainen, L., Bass, R.A., Schaub, V., Sivonen, K. and Borner, T. (1999) Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. J. Bacteriol. 181, 4089–4097.
[26] Ait-Lahsen, H., Soler, A., Rey, M., de la Cruz, J., Monte, E. and Llobell, A. (2001) An antifungal exo-alpha-1,3-glucanase (AGN13.1) from the biocontrol fungus *Trichoderma harzianum*. Appl. Environ. Microbiol. 67, 5833–5839.
[27] de la Cruz, J. and Llobell, A. (1999) Purification and properties of a basic endo-β-1,6-glucanase (BGN16.1) from the antagonistic fungus *Trichoderma harzianum*. Eur. J. Biochem. 265, 145–151.
[28] Limón, M.C., Chacón, M.R., Mejías, R., Delgado-Jarana, J., Rincón, A.M., Codón, A.C. and Benítez, T. (2004) Increased antifungal and chitinase specific activities of *Trichoderma harzianum* CECT 2413 by addition of a cellulose binding domain. Appl. Microbiol. Biotechnol. 64, 675–685.
[29] Delgado-Jarana, J., Rincón, A.M. and Benítez, T. (2002) Aspartyl protease from *Trichoderma harzianum* CECT 2413: Cloning and characterization. Microbiology 148, 1305–1315.
[30] Dana, M.M., Limón, M.C., Mejías, R., Mach, R.L., Benítez, T., Pintor-Toro, J.A. and Kubicek, C.P. (2001) Regulation of chitinase 33 (*chit33*) gene expression in *Trichoderma harzianum*. Curr. Genet. 38, 335–342.
[31] Ballance, D.J. (1986) Sequences important for gene expression in filamentous fungi. Yeast 2, 229–236.
[32] Kulmburg, P., Mathieu, M., Dowzer, C., Kelly, J. and Felenbok, B. (1993) Specific binding sites in the alcR and alcA promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*. Mol. Microbiol. 7, 847–857.
[33] Ravagnani, A., Gorfinkiel, L., Langdon, T., Diallinas, G., Adjadj, E., Demais, S., Gorton, D., Arst Jr., H.N. and Scazzocchio, C. (1997) Subtle hydrophobic interactions between the seventh residue of the zinc finger loop and the first base of an HGATAR sequence determine promoter-specific recognition by the *Aspergillus nidulans* GATA factor AreA. EMBO J. 16, 3974–3986.
[34] Eisendle, M., Oberegger, H., Zadra, I. and Haas, H. (2003) The siderophore system is essential for viability of *Aspergillus nidulans*:

Functional analysis of two genes encoding L-ornithine N 5-monooxygenase (*sidA*) and a non-ribosomal peptide synthetase (*sidC*). Mol. Microbiol. 49, 359–375.
[35] Yuan, W.M., Gentil, G.D., Budde, A.D. and Leong, S.A. (2001) Characterization of the *Ustilago maydis sid2* gene, encoding a multidomain peptide synthetase in the ferrichrome biosynthetic gene cluster. J. Bacteriol. 183, 4040–4051.
[36] Haas, H. (2003) Molecular genetics of fungal siderophore biosynthesis and uptake: The role of siderophores in iron uptake and storage. Appl. Microbiol. Biotechnol. 62, 316–330.
[37] Sivasithamparam, K. and Ghisalberti, E.L. (1998) Secondary metabolism in *Trichoderma* and *Gliocladium*. In: *Trichoderma* and *Gliocladium* (Harman, G.E. and Kubicek, C.P., Eds.), vol. 1, pp. 139–191. Taylor & Francis, London.
[38] Lawen, A. and Traber, R. (1993) Substrate specificities of cyclosporin synthetase and peptolide SDZ 214-103 synthetase. Comparison of the substrate specificities of the related multifunctional polypeptides. J. Biol. Chem. 268, 20452–20465.
[39] Haese, A., Schubert, M., Herrmann, M. and Zocher, R. (1993) Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing *N*-methyldepsipeptide formation in *Fusarium scirpi*. Mol. Microbiol. 7, 905–914.
[40] Scott-Craig, J.S., Panaccione, D.G., Pocard, J.A. and Walton, J.D. (1992) The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. J. Biol. Chem. 267, 26044–26049.
[41] Konz, D. and Marahiel, M. (1999) Review: How do peptide synthetases generate structural diversity? Chem. Biol. 6, R39–R48.
[42] Konz, D., Doekel, S. and Marahiel, M.A. (1999) Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. J. Bacteriol. 181, 133–140.
[43] Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G. and van Sinderen, D. (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. Mol. Microbiol. 8, 821–831.
[44] Tognoni, A., Franchi, E., Magistrelli, C., Colombo, E., Cosmina, P. and Grandi, G. (1995) A putative new peptide synthase operon in *Bacillus subtilis*: Partial characterization. Microbiology 141 (Pt 3), 645–648.
[45] Weber, T. and Marahiel, M. (2001) Minireview: Exploring the domain structure of modular nonribosomal peptide synthetases. Structure 9, R3–R9.
[46] Balakirev, E.S. and Ayala, F.J. (2003) Pseudogenes: Are they junk or functional DNA?. Annu. Rev. Genet. 37, 123–151.
[47] Harrison, P., Kumar, A., Lan, N., Echols, N., Snyder, M. and Gerstein, M. (2002) A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. J. Mol. Biol. 316, 409–419.
[48] Lykke-Andersen, J. (2001) mRNA quality control: Marking the message for life or death. Curr. Biol. 11, R88–R91.
[49] Guillemette, T., Sellam, A. and Simoneau, P. (2004) Analysis of a nonribosomal peptide synthetase gene from *Alternaria brassicae* and flanking genomic sequences. Curr. Genet. 45, 214–224.
[50] Tudzynski, P., Hölter, K., Correia, T., Arntz, C., Grammel, N. and Keller, U. (1999) Evidence for an ergot alkaloid gene cluster in *Claviceps purpurea*. Mol. Gen. Genet. 261, 133–141.
[51] Bailey, A.M., Kershaw, M.J., Hunt, B.A., Paterson, I.C., Charnley, A.K., Reynolds, S.E. and Clarkson, J.M. (1996) Cloning and sequence analysis of an intron-containing domain from a peptide synthetase-encoding gene of the entomopathogenic fungus *Metarhizium anisopliae*. Gene 173 (2), 195–197.
[52] Muro-Pastor, M.I., González, R., Strauss, J., Narendja, F. and Scazzocchio, C. (1999) The GATA factor AreA is essential for chromatin remodelling in a eukaryotic bidirectional promoter. EMBO J. 18, 1584–1597.
[53] Cubero, B. and Scazzocchio, C. (1994) Two different, adjacent and divergent zinc finger binding sites are necessary for CREA-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans*. EMBO J. 13, 407–415.
[54] Mach, R.L., Strauss, J., Zeilinger, S., Schindler, M. and Kubicek, C.P. (1996) Carbon catabolite repression of xylanase I (*xyn1*) gene expression in *Trichoderma reesei*. Mol. Microbiol. 21, 1273–1281.
[55] Mach, R.L. and Zeilinger, S. (2003) Regulation of gene expression in industrial fungi: *Trichoderma*. Appl. Microbiol. Biotechnol. 60, 515–522.