ESPOO 2007 VTT PUBLICATIONS 654

Mikko Arvas

Comparative and functional genome analysis of fungi for development of the protein production host *Trichoderma reesei*

VTT PUBLICATIONS 654

Comparative and functional genome analysis of fungi for development of the protein production host *Trichoderma reesei*

Mikko Arvas

VTT

Faculty of Biosciences
Department of Biological and Environmental Sciences
Division of Genetics
University of Helsinki
Helsinki, Finland

To be presented with the permission of the Faculty of Biosciences of the University of Helsinki, for public criticism in the Auditorium 1041 at the Department of Biological and Environmental Sciences, Biocenter, Viikinkaari 5, Helsinki on October 19th, 2007 at 12 o'clock noon.

ISBN 978-951-38-7043-0 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7044-7 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 3, PL 2000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 3, PB 2000, 02044 VTT
tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 3, P.O. Box 2000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax + 358 20 722 4374

VTT, Tietotie 2, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 7071

VTT, Datavägen 2, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 7071

VTT Technical Research Centre of Finland, Tietotie 2, P.O. Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax +358 20 722 7071

Cover picture: Aleksi Salokannel

Edita Prima Oy, Helsinki 2007

Arvas, Mikko. Comparative and functional genome analysis of fungi for development of the protein production host *Trichoderma reesei*. Espoo 2007. VTT Publications 654. 100 p. + app. 105 p.

Keywords *Trichoderma reesei*, *Hypocrea jecorina*, *Saccharomyces cerevisiae*, comparative genomics, plant biomass degradation, secondary metabolism, transcriptome profiling, TRAC, cDNA-AFLP, cDNA subtraction library, protein secretion stress, Unfolded Protein Response, UPR, oxidative stress, histones, GCN4, CPC1, ATF4

Abstract

Filamentous fungi of the subphylum Pezizomycotina are well known as protein and secondary metabolite producers. Various industries take advantage of these capabilities. However, the molecular biology of yeasts, i.e. Saccharomycotina and especially that of *Saccharomyces cerevisiae*, the baker's yeast, is much better known. In an effort to explain fungal phenotypes through their genotypes we have compared protein coding gene contents of Pezizomycotina and Saccharomycotina. Only biomass degradation and secondary metabolism related protein families seem to have expanded recently in Pezizomycotina. Of the protein families clearly diverged between Pezizomycotina and Saccharomycotina, those related to mitochondrial functions emerge as the most prominent. However, the primary metabolism as described in *S. cerevisiae* is largely conserved in all fungi. Apart from the known secondary metabolism, Pezizomycotina have pathways that could link secondary metabolism to primary metabolism and a wealth of undescribed enzymes.

Previous studies of individual Pezizomycotina genomes have shown that regardless of the difference in production efficiency and diversity of secreted proteins, the content of the known secretion machinery genes in Pezizomycotina and Saccharomycotina appears very similar. Genome wide analysis of gene products is therefore needed to better understand the efficient secretion of Pezizomycotina. We have developed methods applicable to transcriptome analysis of non-sequenced organisms. TRAC (Transcriptional profiling with the aid of affinity capture) has been previously developed at VTT for fast, focused transcription analysis. We introduce a version of TRAC that allows more powerful signal amplification and multiplexing. We also present computational

optimisations of transcriptome analysis of non-sequenced organism and TRAC analysis in general.

Trichoderma reesei is one of the most commonly used Pezizomycotina in the protein production industry. In order to understand its secretion system better and find clues for improvement of its industrial performance, we have analysed its transcriptomic response to protein secretion stress conditions. In comparison to *S. cerevisiae*, the response of *T. reesei* appears different, but still impacts on the same cellular functions. We also discovered in *T. reesei* interesting similarities to mammalian protein secretion stress response. Together these findings highlight targets for more detailed studies.

Preface

This work was mainly done at the Technical Research Centre of Finland, VTT, in the Protein Production team. In addition work was done at the Group of Comparative Microbial Genomics (Center for Biological Sequence Analysis, Denmark), the School of Biological Sciences (University of Manchester, UK) and the Department of Plant Systems Biology (University of Ghent, Belgium). The work was financially supported by the Academy of Finland, the National Technology Agency of Finland and the European Union, which are acknowledged.

I am grateful to Executive Director Professor Juha Ahvenainen and Research Director Professor Hans Söderlund for the excellent working facilities. Research Manager Tiina Nakari-Setälä is thanked for her supportive and cheerful attitude. Also, I wish to thank professors Stephen Oliver and Dirk Inzé and David Ussery for outstanding visits to their laboratories.

My deepest gratitude goes to my supervisors Markku Saloheimo and Tiina Pakula and to my professor Merja Penttilä. All of them have contributed to make this period a personally enjoyable and fruitful experience, as well as given me very important guidance in times of trouble. I am particularly happy about the freedom, support and possibilities given to me to study novel skills. Tiina's determination and attention to detail will always serve as an example to me. Markku has impressed me with his foresight and leadership.

My most sincere thanks are due to my co-authors professor Merja Penttilä, professor Stephen Oliver, professor Esko Ukkonen, Markku Saloheimo, Kristiina Takkinen, David Ussery, Tiina Pakula, Teemu Kivioja, Kari Kataja, Karin Lanthaler, Geoff Robson, Reetta Satokari, Alex Mitchell, Tapani Suortti and Mari Valkonen. Especially, I am indebted to Teemu Kivioja. None of the bioinformatics in this work would have been possible without him.

I thank Professor Liisa Holm and Michael Ward for reviewing the manuscript for this thesis and for their useful comments.

Everybody working at Cell factory and other VTT Biotechnology labs are thanked for a good atmosphere and for their readiness to provide help. Even though I am not famous for my laboratory skills, the best of them were undoubtly learned from Aili. Also, I have especially enjoyed working with the sequencing team. In Bioinformatics problems the Quantitative Biology and Bioinformatics team has provided precious help. Especial thanks go to Marko, Gopal and Matej for always finding time for me. It has been a pleasure to work in the Protein production team. I thank you all for companionship and support. Eija and Virve have been wonderful roommates, thank you! Also I am in gratitude for my roommates in stays abroad, Alan Goossens, Balázs Papp, Luke Hakes, Thomas Grotjke and Kristoffer Kiil. In addition, I am especially in debt to mister Trichoderma, Diego Martinez. Without his help with the genome this work would have never been possible.

Finally I would like to thank my friends and family. My parents, Eeva and Ilkka, and my sister Paula have always been there for me. Allu, Saku, Santtu and Tupe have made my life a pleasure already for over twenty years. Heikki, Janne, Vesa and Tommi made studying biology great fun and have been good friends ever since. RPG sessions and other gaming ensured sanity when work endangered it and I thank everybody involved. Sampo and Mikko are thanked for many good discussions of life at large.

Most importantly, I would like to thank my wife Jenni for everything. Making a PhD surely was not the best way to make money for travelling with you. However, the promise of another voyage with you keeps me going better than anything else.

Contents

Appendices I–V of this publication are not included in the PDF version. Please order the printed version to get the complete publication (http://www.vtt.fi/publications/index.jsp)

List of symbols

aa	amino acid
BIP	Binding Protein
bp	base pair
bZIP	basic leucine-zipper
cDNA-AFLP	cDNA-Amplified Fragment Length Polymorphism
DTT	Dithiothreitol
ER	Endoplasmic Reticulum
ERAD	ER – Associated Degradation
EST	Expressed Sequence Tag
GC	Guanine-Cytosine
GSH	Glutathione
MFS	Major Facilitator Superfamily
NRPS	Non-Ribosomal Peptide Synthetase
nt	nucleotide
ORF	Open Reading Frame
PCA	Principal Component Analysis
PDI	Protein Disulphide Isomerase
PERK	PKR-like ER protein kinase
PKS	PolyKetide Synthetase
RESS	Repression under Secretion Stress
RIP	Repeat Induced Point mutation
ROS	Reactive Oxygen Species
RT-PCR	Reverse Transcriptase – PCR
TCA	Triacarboxylic Acid Cycle
tPA	tissue Plasminogen Activator

TRAC	TRanscript analysis with the aid of Affinity Capture
UPR	Unfolded Protein Response
UPRE	UPR Element

List of publications

I Arvas, M., Kivioja, T., Mitchell, A., Ussery, D., Saloheimo, M., Penttilä, M. and Oliver, S. 2007. Comparison of protein coding gene contents of the fungal phyla Pezizomycotina and Saccharomycotina. BMC Genomics, 8:325.

II Kataja, K., Satokari, R.M., Arvas, M., Takkinen, K. and Söderlund, H. 2006. A highly sensitive and multiplexed method for focused transcript analysis. Journal of Microbiol Methods, Vol. 67, No. 1, pp. 102–113.

III Kivioja, T., Arvas, M., Kataja, K., Penttilä, M., Söderlund, H. and Ukkonen, E. 2002. Assigning probes into a small number of pools separable by electrophoresis. Bioinformatics, Vol. 18, Suppl. 1, pp. S199–206.

IV Kivioja, T., Arvas, M., Saloheimo, M., Penttilä, M. and Ukkonen, E. 2005. Optimization of cDNA-AFLP experiments using genomic sequence data. Bioinformatics, Vol. 21, No. 11, pp. 2573–2579.

V Arvas, M., Pakula, T., Lanthaler, K., Saloheimo, M., Valkonen, M., Suortti, T., Robson, G. and Penttilä, M. 2006. Common features and interesting differences in transcriptional responses to secretion stress in the fungi *Trichoderma reesei* and *Saccharomyces cerevisiae*. BMC Genomics, 7:32.

Computational aspects of III and IV are outside the author's area of expertise and have been defended previously (Kivioja 2004).

1. Introduction

1.1 Pezizomycotina as industrial protein production systems

Fungi generally derive their nutrition from other organisms by externally digesting them with enzymes. This capacity has been used by man for centuries to modify organic materials. For example *Aspergillus oryzae*, a fungus of the subphylum Pezizomycotina, has been used in Sake brewing at least from the 3rd century A.D., however it is the yeasts of the subphylum Saccharomycotina that are commonly used for brewing. Unlike practically all other fungi, Saccharomycotina have lost their capacity to break down organic material and live instead on free sugars. Sake is produced from rice and to free the sugars, rice was originally chewed by people and subsequently fermented by Saccharomycotina. This step was later replaced by a process where *A. oryzae*, instead of human enzymes, liberates sugars from rice for yeast fermentation.

Today at least textile, food and feed, paper and pulp, and energy industries profit from the capacity of fungi to secrete enzymes that modify organic material. The industrial usefulness of fungi stems from the facts that they digest externally, they can digest practically all kinds of organic material and they can be grown easily. To retrieve the digesting enzymes for use one merely concentrates the used growth media in contrast to extracting them from inside of an organism. Modern biotechnology allows us to use fungi even for production of enzymes derived from other organisms. In an analogy to the Sake example, also cheese making used to require a mammalian enzyme, chymosin extracted from calf stomachs. Today calf chymosin is produced in genetically modified fungi like the Pezizomycotina *Aspergillus niger* (for review see (Berka et al. 1991)).

The fungus *Trichoderma reesei* (Reese et al. 1950) of the subphylum Pezizomycotina is one of the most commonly used industrial protein production organisms. It is an asexual form of the *Hypocrea jecorina* (Kuhls et al. 1996), a tropical soft rot fungus, and was discovered as the "jungle rot" that degraded soldier's tents and uniforms in the Pacific during World War II. *T. reesei* has been used extensively in the Finnish bioindustry.

Modern societies depend much on oil as a source of energy and plastics. In order to replace oil, renewable organic material must be modified and converted to different forms. Fungal enzymes provide natural and environmentally friendly means for this task. However, molecular biology is still in transition from a descriptive science to an engineering science. Systems biology strives for a holistic and quantative view of the complex interactions behind the phenotypic properties of organisms. This view would enable us to engineer for example fungi to meet the industrial needs of an oil free economy. The prerequisite of systems biology is to study and subsequently understand the constituents of an organism i.e. to analyse its genome.

1.2 Genome analysis of Pezizomycotina

The fungal kingdom is composed of phyla Chytridiomycota, Zygomycota, Glomeromycota, Ascomycota, Basidiomycota and Deuteromycota. Ascomycota is composed of subphyla Pezizomycotina, Saccharomycotina and Taphrinomycotina. So far genome sequences of six Pezizomycotina have been published: *Aspergillus fumigatus* (Nierman et al. 2005), *Aspergillus nidulans* (Galagan et al. 2005), *A. oryzae* (Machida et al. 2005), *Magnaporthe grisea* (Dean et al. 2005) and *Neurospora crassa* (Galagan et al. 2003). In addition, the genome of *Aspergillus flavus* has been published, but it is believed to be an ecotype of *A. oryzae* (Payne et al. 2006). Also, data from unpublished genome of *T. reesei* (D. Martinez, Joint Genome Institute, USA) is available (Table 1). Pezizomycotina live mostly in filamentous form and degrade biomass to free sugars to be used as source of carbon and energy. In contrast, Saccharomycotina, commonly referred to as yeasts, grow mostly in a unicellular form. They have generally no capability to degrade biomass and respectively live on free sugars. Undoubtly the most studied fungus and Saccharomycotina is *Saccharomyces cerevisiae*, the baker's yeast. Thus Saccharomycotina present a useful comparison to the Pezizomycotina. Far more Saccharomycotina genomes have been published and, as a reference to Pezizomycotina, for this work we have selected 6: *Ashbya gossypii* (Dietrich et al. 2004), *Candida glabrata, Debaryomyces hansenii, Kluyveromyces lactis* (Dujon et al. 2004), *Saccharomyces cerevisiae* (Goffeau et al. 1996) and *Yarrowia lipolytica* (Dujon et al. 2004). In addition three Basidiomycota genomes have been published: *Cryptococcus neoformans* (Loftus et al. 2005) *Phanerochaete chrysosporium* (Martinez et al. 2004) and *Ustilago maydis*

(Kamper et al. 2006). The phylogenetic relationships of the species discussed are presented in Figure 1.

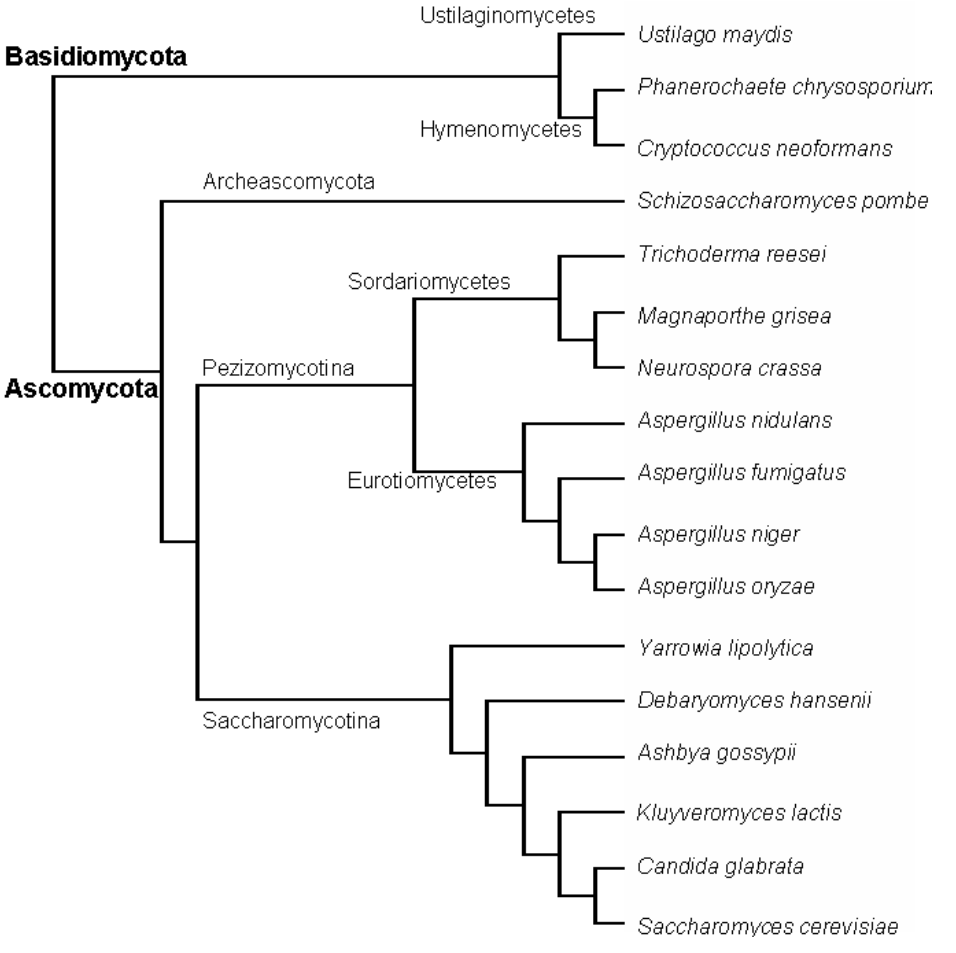


Figure 1. Cladogram of fungi with published genome sequences, adapted from (Kurtzman 2003; Diezmann et al. 2004; Pel et al. 2007)*. Branch lengths are arbitrary, subphyla and classes are shown for relevant cases.*

1.2.1 Genome structure

An average Pezizomycotina has about three times larger genome than a Saccharomycotina. However, there is only a two fold difference in counts of open reading frames (ORFs). ORFs in Pezizomycotina are slightly longer than in

Saccharomycotina, which is partly due to the fact that a Pezizomycotina ORF is typically composed of 2–3 exons, while generally only 5 percents of ORFs of a Saccharomycotina have introns. Also, Pezizomycotina genomes are less densely packed with ORFs. Another major difference between these two subdivisions is the difference in the GC percent. Saccharomycotina typically have a lower GC percent but, unlike in Pezizomycotina there is large variation. In light of these statistics, the Basidiomycota seem to resemble more Pezizomycotina than Saccharomycotina, except that their genome size and respectively ORF count varies a lot (Table 1). Hymenomycetes have a more complex ORF structure than the Ascomycota with 3–6 exons per ORF (Loftus et al. 2005), while only 70 percents of ORFs of the only sequenced Ustilaginomycetes, *U. maydis* have any introns.

Chromosomal regions containing repeated sequences are difficult to resolve with the commonly used whole genome shot gun (WGS) sequencing (Venter et al. 1998). Thus, the quality of information concerning repeats in published fungal genomes varies a lot. However, while approximately 40 percents of human and ~50 percents of the rice (Yu et al. 2002) genome consists of transposon sequences, in fungi they seem to be rare. Transposons are mobile genetic elements that can move from one location in the genome and reinsert at another site (for review (Kempken and Kuck 1998)). *A. gossypii* has no transposons (Dietrich et al. 2004) and other Saccharomycotina have variable, but small amounts of representatives of different classes (for review (Dujon 2006)). In Pezizomycotina, 10 percents of the genomes of *N. crassa* (Galagan et al. 2003) and *M. grisea* (Dean et al. 2005) are composed of repetitive DNA originating mostly from retrotransposons. Due to the Repeat Induced Point mutations (RIP) mechanism *N. crassa*, nevertheless, has no intact mobile elements. RIP specifically mutates duplications greater than 400 bp and more identical than 80 percents by introducing C:G to T:A mutations into both copies of the duplicated sequence. RIP takes place during the sexual cycle and probably involves methylation of the DNA prior to mutagenesis. The full mechanism has not yet been described. RIP effectively prevents gene duplications and keeps the genome extremely non-redundant (for review, see (Galagan and Selker 2004)). From 1 to 3 percents of genomes of the Aspergillus species consist of transposons (Galagan et al. 2005). There is evidence that RIP exists, although in a milder form, in *M. grisea* (Ikeda et al. 2002) and Aspergillus (Montiel et al. 2006) species. Only 1.1 percents of the *U. maydis* genome consists of mostly

non-functional transposon derived sequences (Kamper et al. 2006). *U. maydis* appears to have no RNA interference pathway nor RIP, which both could restrict transposons (Galagan and Selker 2004; Vastenhouw and Plasterk 2004).

Table 1. Descriptive statistics of selected sequenced fungal genomes. Numbers are from original publications referred to in text, except for "Calculated ORF density" which was calculated by dividing "Count of ORFs" with "Total length (Mbp)". "GC %" is the percent of guanine and cytosine nucleotides. "Coding %" is the percent of protein coding sequence. Statistics for S. cerevisiae are from (Dujon et al. 2004)*. T. reesei statistics are from the prepublication version (Diego Martinez, personal communication).*

Phylum	Subphylum / Class	Name	Chromosomes	Total length (Mbp)	GC %	Count of ORFs	Count of tRNAs	Coding %	Average ORF size (bp)	Calculated ORF density (bp)
Basidiomycota	Ustilaginomycetes	*Ustilago maydis*	23	20	54	6 902		64	1 935	2 869
	Hymenomycetes	*Phanerochaete chrysosporium*	< 10	30	57	11 777	200	45		2 539
		Cryptococcus neoformans	14	19	49	6 572	141	55	1 900	2 891
Ascomycota	Archeascomycota	*Schizosaccharomyces pombe*	3	14	36	4 824	174	58	1426	2 861
	Pezizomycotina	*Trichoderma reesei*	7	34	52	9 130		40	1 476	3 713
		Magnaporthe grisea	7	40	52	11 109	316	41	1 683	3 628
		Neurospora crassa	7	40	50	9 200	424	44	1 673	4 337
		Aspergillus nidulans	8	30	50	9 396	188	50	1 868	3 203
		Aspergillus fumigatus	8	28	49	9 009	179	49	1 644	3 108
		Aspergillus niger	8	34	50	13 238	269		1 572	2 561
		Aspergillus oryzae	8	37	48	12 074		45	1 414	3 073
	Pezizomycotina	**Average**	**8**	**35**	**50**	**10 451**	**275**	**45**	**1 619**	**3 375**
	Saccharomycotina	*Yarrowia lipolytica*	6	21	49	6 703	510	46	1 428	3 058
		Debaryomyces hansenii	7	12	36	6 906	205	79	1 167	1 767
		Ashbya gossypii	7	9	52	4 718	199			1 950
		Kluyveromyces lactis	6	11	39	5 329	162	72	1 383	1 989
		Candida glabrata	13	12	39	5 283	207	65	1 479	2 328
		Saccharomyces cerevisiae	16	12	38	5 807	274	70	1 455	2 084
	Saccharomycotina	**Average**	**9**	**13**	**42**	**5 791**	**260**	**66**	**1 382**	**2 196**

Transposons facilitate mutagenesis and horizontal gene transfer, thus allowing for example pathogenic bacteria to evolve fast (Wilson et al. 2002; Raskin et al. 2006/2/24). 5 percents of the genome of the human pathogen *C. neoformans* consists of transposons. They might drive its phenotypic instability and subsequently impair the treatment of infections (Loftus et al. 2005). Similarly,

M. grisea, the rice blast fungus, with the largest percent of intact transposon derived sequences among sequenced Ascomycota, is a difficult plant pathogen.

In conclusion, it is clear that known and yet undescribed mechanisms to restrict transposons exist in fungi. In addition, in Pezizomycotina RIP further prevents gene duplication and therefore stabilises genomes. This facilitates their use in research and industry.

1.2.2 Protein coding gene content of fungal genomes

The difference of life style between Pezizomycotina and Saccharomycotina is obvious from their protein coding gene content. While Saccharomycotina have few, Pezizomycotina genomes are abundant in ORFs related to the degradation of biomass such as glycoside hydrolases and proteinases. Glycosides are made from monosaccharides by glycosyltransferases and degraded by glycoside hydrolases. Derivates of glycosides, such as cellulose and hemicellulose, form the bulk of plant biomass.

The higher amounts of glycoside hydrolases are to be expected as a global correlation between genome size and the number of glycoside hydrolases and glycosyltransferases exists (Coutinho et al. 2003). Among Eurotiomycetes this correlation is weak. *A. niger* and *A. oryzae,* with the largest genomes, have the most extracellular enzymes. *A. niger* has the most glycoside hydrolases (Pel et al. 2007), while *A. oryzae* has the least (Machida et al. 2005). Instead, *A. oryzae* has the largest amount of secreted proteinases. In comparison to *S. cerevisiae, A. oryzae* has also more ORFs related to certain amino acid metabolism pathways and amino acid transportes. These features appear to make it well adapted to the traditional Japanese *koji* culture where it is often grown on the surface of solid ground soybean, a protein rich energy source (Machida et al. 2005).

Sordariomycetes, e.g. *T. reesei, N. crassa* and *M. grisea*, appear to have fever ORFs related to biomass degradation than Eurotiomycetes, e.g. Aspergilli, of the same genome size ((Pel et al. 2007), D. Martinez unpublished results). However, *M. grisea* has about twice the number of ORFs predicted to encode secreted proteins than *N. crassa*. The expanded protein families in *M. grisea* are related to plant cell wall degradation and many are transcriptionally induced in infection

related development (Dean et al. 2005). Consequently, this expansion appears to be directly related to the pathogenic life style of *M. grisea*.

Regardless of the difference in genome contents of ORFs of secreted proteins between Saccharomycotina and Pezizomycotina species, their genomes have revealed only few differences in secretion machinery related ORF content. Pezizomycotina species have for example additional paralogues of the *PDI1* (see chapter 1.4) and trafficking related GTPases (Borkovich et al. 2004; Pel et al. 2007; Diego Martinez, personal communication). This raises the questions, whether Pezizomycotina have some secretion mechanisms that has been missed by current homology based annotation, or whether the mechanism discovered in *S. cerevisiae* can be shaped into the needs of Pezizomycotina secretion with only minor modifications.

Table 2. Counts of major secondary metabolism enzyme type ORFs in fungi. Numbers are from original publications referred to in text. The number of PKS-NRPSs is not reported in (Payne et al. 2006) *and the count of A. oryzae NRPSs is 18 in* (Machida et al. 2005) *and 24 in* (Payne et al. 2006)*, T. reesei counts are based on the prepublication version (Scott Baker, personal communication).*

		PKS	NRPS	PKS-NRPS
Basidiomycota	*Phanerochaete chrysosporium*	9	7	0
Ascomycota	*Trichoderma reesei*	12	11	2
	Magnaporthe grisea	23	6	8
	Neurospora crassa	7	3	0
	Aspergillus nidulans	27	13	1
	Aspergillus fumigatus	14	13	1
	Aspergillus niger	34	17	7
	Aspergillus oryzae	30	24	0
	Aspergillus flavus	35	24	?

Another well known difference between Pezizomycotina and Saccharomycotina is the capability of Pezizomycotina to produce a wealth of secondary metabolites. Secondary metabolites are diverse bioactive low molecular weight compounds. Well known examples include penicillin produced by the Eurotiomycetes species *Penicillium chrysogenum* and aflatoxin produced by *A. flavus*. Hundreds of antibacterial, antifungal or antitumoral fungal secondary metabolites have been described. The two most common classes can be characterised by a defining enzyme in their biosynthesis pathway. Polyketides

(e.g. aflatoxin) are produced by polyketide synthetases (PKS) and non-ribosomal peptides (e.g. pencillin) are produced by non-ribosomal peptide synthetases (NRPS) (for review (Keller et al. 2005)). In addition there are PKS-NRPS fusion enzymes (Bohnert et al. 2004).

The counts of major secondary metabolism enzyme type ORFs seem to vary a lot inside phyla and even inside probable ecotypes of the same species such as *A. oryzae* and *A. flavus* (Payne et al. 2006) (Table 2). Interestingly, both *A. oryzae* and *A. flavus* genomes seem to contain the full aflatoxin biosynthesis pathway. Regardless, *A. oryzae* has the GRAS (Generally Regarded As Safe) status and has been used in food production for centuries.

PKSs and NRPSs often occur in a chromosomal cluster with ORFs likely related to the production of the secondary metabolite such as P450 monooxygenases, dehydrogenases, transporters and a transcription factor (Keller and Hohn 1997; Yu et al. 2004; Gardiner and Howlett 2005). These clusters are often found in subtelomeric regions (Nierman et al. 2005; Rehmeyer et al. 2006). Subtelomeric regions appear to have a potential for faster evolution than many other chromosomal regions (Naumov et al. 1996; Freitas-Junior et al. 2000; Daran-Lapujade et al. 2003; Rehmeyer et al. 2006), consequently subtelomeric positioning could provide a mechanism to explain the fast evolution of these clusters. Why secondary metabolite ORFs tend to cluster is not completely understood. It has been claimed that some clusters have been transferred from bacteria while others have not been. At least in *Aspergilli* an epigenetic transcription control through simultaneous methylation of the whole cluster has been shown (Bok and Keller 2004; Bok et al. 2005). This type of regulation could provide a selective advantage and a subsequent evolutionary pressure to keep the clusters intact (Keller et al. 2005). In Saccharomycotina species ORFs of an allantoin catabolism pathway are clustered in a subtelomeric region. They are also transcriptionally regulated in concert through a chromatin modification mechanism. However, an advantage conferred by transcriptional co-regulation is not believed to have caused the clustering of the allantoin pathway (Wolfe et al. 2000).

The genomes of *S. cerevisiae* and *C. glabrata* are characterised by a recent genome duplication (for review see (Dujon 2006)). The sequenced genomes have retained only ~11 percents of the postulated original gene pairs. However, in ~80 percents of the gene deletions *S. cerevisiae* and *C. glabrata* have retained

the same copy. Therefor the genes have been probably lost rapidly before speciation of the two species. In addition, genome duplications have been described in other eukaryotes such as rice, *Oryzea sativa* (Yu et al. 2005), *Arabidopsis thaliana* (Arabidopsis Genome Initiative 2000) and the puffer fish, *Tetradon negroviridis* (Jaillon et al. 2004). However, no traces of genome duplication have been found in published Pezizomycotina genomes. In addition fungi in general seem to be relatively poor in transposon derived sequences and the RIP mechanism seriously limits gene duplication in at least *N. crassa*. Since the development of RIP only six gene duplication events appear to have taken place, the genes escaping RIP due to a particular gene structure (Galagan and Selker 2004). In addition to RIP *N. crassa* has also post-transcriptional silencing based genome defense mechanisms that prevent activity of foreign or mutated genes. These mechanisms are likely to be conserved at least in Pezizomycotina. Each of the genome defence mechanisms is active in a unique stage of the *N. crassa* life cycle covering it entirely (for review see (Borkovich et al. 2004)).

In conclusion, it seems that many mechanisms exist to limit evolution in fungi. Is fungal evolution then particularly slow? The fast evolution of the secondary metabolism clusters shows that it is not, at least in all parts of the genome. Known plant biomass degradation and secondary metabolism related genes explain only partly why Pezizomycotina genomes have almost two times more genes than Saccharomycotina. Thus the protein coding gene content of Pezizomycotina and its evolutionary history remains largely unexplored.

1.2.3 Databases

In February 2007, the NCBI (National Center for Biotechnology Information) genome database had records of public sequencing projects of 68 fungal genomes. Of eukaryotes, fungi are the kingdom with most sequenced species. Sequenced genomes cover fungi broadly and include industrially, medically and agriculturally important species with very diverse genomes. Species of diverse evolutionary distances have been sequenced, such as strains of *C. neoformans* (Loftus et al. 2005), members of Saccharomyces *sensu stricto* (Kellis et al. 2003), ecotypes of *A. oryzae* (Payne et al. 2006) and species from different phyla of the kingdom. Thus, fungal genomes form an incomparable resource for eukaryotic comparative genomics research and method development.

To fully profit from these possibilities the fungal research community needs a comprehensive comparative genomic database. Extensive comparative genomic databases for some groups of eukaryotic species are available such as the Ensembl for mammals (Birney et al. 2006). Several different genome database types exists for fungi such as the *Saccharomyces* Genome Database (Dolinski et al. 2004) on which the databases of *A. gossypii* (Gattiker et al. 2007) and *C. albicans* (Arnaud et al. 2005) are based on. In addition the sites of the major sequencing centres TIGR (The Institute for Genomic Research), Sanger (The Welcome Trust Sanger Institute), JGI (The Joint Genome Institute) and Broad (The Broad Institute) exist. All existing database types use in-house software and databases incompatible with other sites and with little comparative capabilities. Several comparative databases exist for fungi such as MIPS (Mewes et al. 2004), CBS Genome Atlases (Pedersen et al. 2000), eFungi (Cornell et al. 2007), but none offers as detailed information as Ensembl. In conclusion, abundant resources exist to build and maintain genome databases for fungi, but little co-operation to pool resources into a common comparative database enabling full use of genome data exists.

1.3 Transcriptome analysis

As discussed above, many properties of genome sequences can be related to known phenotypes of an organism. However, our understanding of the relationships between genotype and phenotype is still poor. Undoubtedly these relationships are complex and, to decipher them, a common strategy is to study the intermediates between genotype and phenotype such as transcripts, proteins, lipids or metabolites. Of these, transcripts seem as the easiest target for genome wide studies. Unlike lipids or many proteins, transcripts i.e. RNAs are water soluble, making their laboratory manipulation easy. RNAs are also far more stable than many metabolites or lipids and thus easier to extract. The core of many current transcriptome profiling technologies (Table 3) is the property of single stranded nucleic acids to hybridise specifically with complementary nucleic acids to form double stranded molecules. This allows the recognition and at least relative quantification of the majority of RNA molecules in a sample simultaneously. Although, RNAs have little enzymatic activities compared to proteins, their quantities still give valuable information about cellular events. The fact that a gene is regulated in a certain way is a direct measure of decisions

leading to a regulation event, but unfortunately it gives us no direct information of the following events such as the activity of a protein coded by the gene.

Table 3. Common transcriptome profiling techniques. "Core technique" summarises whether the hybridisation of an RNA sample on probes or sequencing of ESTs form the core of the technique. See details in text.

Technique	Core technique	Genome sequence requirement	Practical limit of genome coverage
Oligonucleotide microarray	Hybridisation	Needed	Quality of genome sequence
cDNA microarray	Hybridisation	Not needed	Size of cDNA library
cDNA subtraction library	Hybridisation / Sequencing	Not needed	Size of cDNA library
SAGE	Sequencing	Due to short length of sequences in practice needed	Quality of genome sequence
MPSS	Sequencing	Due to short length of sequences in practice needed	Quality of genome sequence
EST sequencing	Sequencing	Not needed	Size of cDNA library
cDNA-AFLP	Sequencing	Not needed	Size of cDNA library

1.3.1 Transcriptome analysis of organisms with sequenced genomes

Currently the most important transcriptome profiling methods for sequenced organisms are cDNA (Schena et al. 1995) and oligonucleotide microarrays (Lockhart et al. 1996). In both the probes (either cDNA or oligonucleotide) covering most of the gene content of a genome, are attached to a solid support, typically a glass slide. Labelled cDNA sample derived from a cellular RNA sample is hybridised on the probes. Either of these methods can be applied effectively if at least a draft quality genome sequence is available. However, cDNA arrays are often constructed based just on EST (Expressed Sequence Tag) data. Commonly a single signal value is obtained for each predicted ORF of the genome, but the whole genome can also be tiled with overlapping probes, for a far more detailed measurement that is independent of ORF predictions (Yamada et al. 2003; David et al. 2006).

The selection of good oligonucleotide probes is a major challenge. A good probe should be sensitive, i.e. it should produce a clear signal when bound by the target molecule, and it should be specific, i.e. it should not bind any other transcripts in the sample. Unfortunately, the prediction of either property from sequence data reliably is still difficult (Kivioja 2004). In addition to the difficulties of

predicting thermodynamics of hybridisations of theoretical linear free nucleic acids, nucleic acids fold into secondary structures and in microarray technology the probe is bound to a solid surface further restricting its movement.

One of the major problems of the microarray methods is their lack of sensitivity. A relatively large sample is needed for hybridisations, typically several micrograms of total RNA and even then low abundance transcripts are difficult to detect (Evans et al. 2002). This renders microarrays unusable for thin needle biopsies, microdissected tissue slices, cytological samples in clinical research or specific parts of fungal hyphae. Also, it has been reported that fold changes of transcript levels are underestimated (Yuen et al. 2002; van de Peppel et al. 2003). This is due, at least partly, to the pre-processing of the signal values and thus possibly common to any genome wide transcriptome profiling technique. Comparative analyses between different microarray methods show that while large changes can be reliably measured, there is much room for improvement (Kuo et al. 2002; Jarvinen et al. 2004; Park et al. 2004).

A number of alternative hybridisation based methods have been proposed such as the Illumina bead arrays (Yeakley et al. 2002) and microelectronic arrays (Weidenhammer et al. 2002). In both, instead of binding probes to a common solid support, they are bound to beads that facilitate their handling.

Instead of hybridisation based methods an RNA sample can be transformed into short cDNA sequences i.e. ESTs (Expressed Sequence Tag) that are sequenced in large numbers. In theory, these methods could be applied to a non-sequenced organism. In practice ESTs are often too short to be mapped by similarity to other organisms and a genome sequence of at least a very close relative is needed. The advantage of these methods is that one gets a direct measurement of transcript abundance instead of a measurement of hybridisation signal. In hybridisation based techniques design of the hybridisation probe, hybridisation process, signal amplification methods used in hybridisation signal measurement etc. could all confound the original signal abundance. In SAGE (Serial Analysis of Gene Expression) cDNA made from transcripts are digested with class II restriction enzyme that cuts 10–20 bp away from its recognition site. The segment between recognition site and cut site from each transcript is retrieved and these are randomly ligated together into larger molecules suitable for sequencing. Sequences specific to particular genes are counted from sequencing

data for a measure of transcript abundance (Velculescu et al. 1997). The major difference between MPSS (Massively Parallel Serial Sequencing) and SAGE is that instead of a conventional sequencing device a specialised device capable of sequencing in parallel is used in MPSS for much higher throughput (Brenner et al. 2000). Though theoretically very interesting, due to high costs neither of these method has gained much popularity.

Although all the methods described above are considered fairly reliable, the expression profiles of candidate genes for further study are usually verified with an additional method. Northern hybridisations and RT-PCR (Reverse Transcriptase – PCR) based methods such as qRT-PCR (quantative Real Time – PCR) (Weis et al. 1992) are the most common.

1.3.2 Transcriptome analysis in fungi with sequenced genomes

Fungal transcriptome analysis boasts a hallmark paper where *S. cerevisiae* cultures were analysed in a wide variety of environmental perturbations (Gasch et al. 2000). This allowed the description of the environmental stress response (ESR) common to all stresses studied and formed an excellent reference for future studies. A number of Pezizomycotina species have been also studied with microarrays. However, in most cases the microarrays employed were cDNA arrays based on EST databases and thus cover the genome only partially.

Microarray transcriptome profiling of fungi was used for the first time in an analysis of the shift from fermentative to respirative metabolism, i.e. the diauxic shift, in *S. cerevisiae* (DeRisi et al. 1997). In the presence of oxygen and glucose *S. cerevisiae* prefers to ferment glucose to ethanol. The process does not require oxygen but, produces much less energy than aerobic, i.e. respirative, metabolism of glucose to CO_2. DeRisi and co-workers showed that in accordance with previous results, TCA (Triacarboxylic Acid Cycle) enzyme transcripts are down regulated in presence of glucose in *S. cerevisiae*. In contrast, *T. reesei* prefers respiration and a study of *T. reesei* in glucose rich and poor conditions showed that TCA enzyme transcripts are induced in the presence of glucose (Chambergo et al. 2002). A similar study in *A. oryzae* showed an induction in TCA transcripts in the presence of glucose (Maeda et al. 2004). Like *S. cerevisiae*, *N. crassa* ferments in aerobic and glucose rich conditions and its TCA transcripts are

repressed. These results suggest that gene regulation drives fungi to either aerobic or anaerobic metabolism in the presence of glucose. *N. crassa* is known for its high growth rate and Xie and co-workers propose that, similar to *S. cerevisiae*, it has a strategy of rapid removal of glucose from a glucose rich environment and production of ethanol for competition with other microbes (Xie et al. 2004). This makes *N. crassa* an attractive bioethanol production organism, as it could both degrade biomass and convert it to ethanol. After the genome duplication the *S. cerevisiae* ancestor rapidly lost thousands of ORFs (Dujon et al. 2004; Kellis et al. 2004) and at large Saccharomycotina genomes have less ORFs and higher ORF density then Pezizomycotina (Table 1). *N. crassa* appears to follow a similar evolutionary course of diminishing ORF content as Saccharomycotina with the development of RIP. This seems to correlate well with its mode of glucose metabolism.

T. reesei is used in the bioindustry mainly to produce cellulases. Thus cellulase gene regulation and cellular adaptation to cellulase production are of interest. Differences between cellulase induced and non-induced conditions and a mutant strain producing abundant cellulases compared to a parent strain have been studied (Foreman et al. 2003). The experiment showed that in the stage of fermentation where *T. reesei* actively produces cellulases, there is no corresponding induction of the secretion system (Foreman et al. 2003). To study specifically the adaptation of cells to conditions where protein secretion stresses cells, the effect of recombinant protein production was analysed in *A. nidulans*. The secretion stress induced by recombinant protein production was found to have milder transcriptomic effects than the one caused by the chemical dithiothreitol (DTT) (Sims et al. 2005).

Fungal development has been studied by transcriptome profiling in *N. crassa*. Novel clock controlled genes have been found (Correa et al. 2003) and a wealth of previously uncharacterised genes were associated to spore germination. Some of the gene expression events related to spore germination, were proposed to be conserved between *U. maydis* and *N. crassa* (Kasuga et al. 2005).

1.3.3 Transcriptome analysis of organisms with non-sequenced genomes

If no sequence data is available transcriptome profiling can be based on selective EST sequencing. A large EST sequencing project is in itself transcriptome profiling since it produces a catalogue of the most abundant transcripts of the cell. However this chapter will concentrate on methods where a preselection of transcripts prior to sequencing is used. Strategies of arraying unsequenced cDNAs, instead of selecting probes based on their sequence are also excluded.

The two major preselection methods are the cDNA subtraction library and differential display methods. In cDNA subtraction library cloning, cDNA from one sample (tester) is hybridised to excess of cDNA from another sample (driver) and non-hybridised tester molecules are separated from hybridised common molecules and non-hybridised driver sequences. Non-hybridised tester cDNA fragments represent transcripts in excess as compared to the driver sample. Typically, several cycles of hybridisation are carried out to further purify the tester sample. By inverting the roles of the driver and the tester samples both over and under represented cDNA fragments can be extracted. The final pool of unhybridised driver fragments is cloned to a vector and transformed to a host. The resulting subtracted cDNA library can then be sequenced.

The cDNA subtraction library cloning method has been in use for a long time (Hedrick et al. 1984) and numerous variations have been reported. The major improvements were the introduction of PCR to amplify the samples to improve sensitivity (Hara et al. 1991; Lisitsyn et al. 1993; Ausubel et al. 1994; Diatchenko et al. 1996) and oligo-dT coated beads to facilitate the removal of driver sample after hybridisation (Hara et al. 1991; Watson et al. 2000). Still the method has often a high rate of false positives. In a final subtraction library 10–50 percents of the clones can be false positives. However, low false positive rates of 5–10 percents have been reported with the Suppression Subtractive Hybridisation (SSH) method (von Stein et al. 1997; Qin et al. 2003). In SSH, 5' adaptors with terminal inverted repeats are added to the tester cDNA fragments. The inverted repeats will lead to stable secondary structures that prevent amplification of other than the true positive tester fragments in subsequent PCR amplifications.

A cDNA subtraction library analysis was recently applied to study a biotechnologically relevant problem in fungi, the protein secretion stress in *A. niger*. Among other responses, the work describes a novel DTT treatment specific induction of methyltransferases (Watson et al. 2000; MacKenzie et al. 2005). It has been shown previously that in protein secretion stress conditions in *A. niger* the transcription of *bip* and *pdi* (see chapter 1.4) are induced (Wiebe et al. 2001; Al-Sheikh et al. 2004). Also the authors themselves showed that *bip* and *pdi* are induced at least two-fold in their original RNA samples by a microarray analysis. However, the corresponding cDNA fragments were not found from the induced cDNA subtraction library.

Differential display methods have also been used for a long time. They are based on PCR amplification of cDNA derived from an RNA sample and subsequent separation of the individual amplified cDNAs by electrophoresis. The DDRT-PCR (Differential Display Reverse Transcriptase PCR) method uses random primers with for example two selective nucleotides to amplify a subset of cDNAs, i.e. a pool. Thus, for each primer pair only sequences whose ends match the selective primer extensions get amplified and form a pool. The cDNAs of a pool can be separated from each other in electrophoresis based on length. By comparing band intensities of electrophoresis runs of pools produced by the same set of primers from different RNA samples, differentially regulated transcripts can be recognised, hence differential display (Liang and Pardee 1992).

A major improvement in differential display techniques was the development of cDNA-AFLP (cDNA-Amplified Fragment Length Polymorphism) (Bachem et al. 1996; Breyne and Zabeau 2001; Fukumura et al. 2003) method. In this method, the cDNAs are digested with two different restriction enzymes, adapters are attached to the specific ends of the resulting fragments, and the fragments are amplified using primers matching the restriction sites and extended with additional selective nucleotides. This makes the pool production process more reproducible and in theory each pool should contain only one fragment from each transcript unlike in DDRT-PCR (Figure 2).

In practice the major problem of cDNA-AFLP is the high number of false positives. Neither digestions, ligations, affinity capture nor the selective PCR, work perfectly and even if they do, an electrophoresis band might still contain fragments of the same length derived from different transcripts. A common

strategy is to cut bands from electrophoresis gels and use them as probes in Northern hybridisations (Callard et al. 1994; Qin et al. 2000). Also, the single-stranded conformation polymorphism (SSCP) method has been used to separate fragments of the same length but different sequence (Mathieu-Daude et al. 1996; Gellatly et al. 2001). A cDNA-AFLP with SSCP was used to study the sirodesmin production of the Ascomycota *Leptosphaeria maculans* (Gellatly et al. 2001). Sirodesmin is a phytotoxin produced by an NRPS secondary metabolism cluster (Gardiner et al. 2005). 39 bands in the cDNA-AFLP electrophoresis gels were classified as differentially expressed. In SSCP analysis all were found to contain several fragments. However, in a subsequent Northern analysis of the most promising fragment from each cDNA-AFLP band, 28 were confirmed as differentially expressed. Variable transcript termination and polyadenylation presents additional problems. An extensive study of *S. cerevisiae* transcriptome with a modified cDNA-AFLP protocol suggested that on average up to seven mRNAs differing by their 3' ends could be produced from a single ORF (Fukumura et al. 2003).

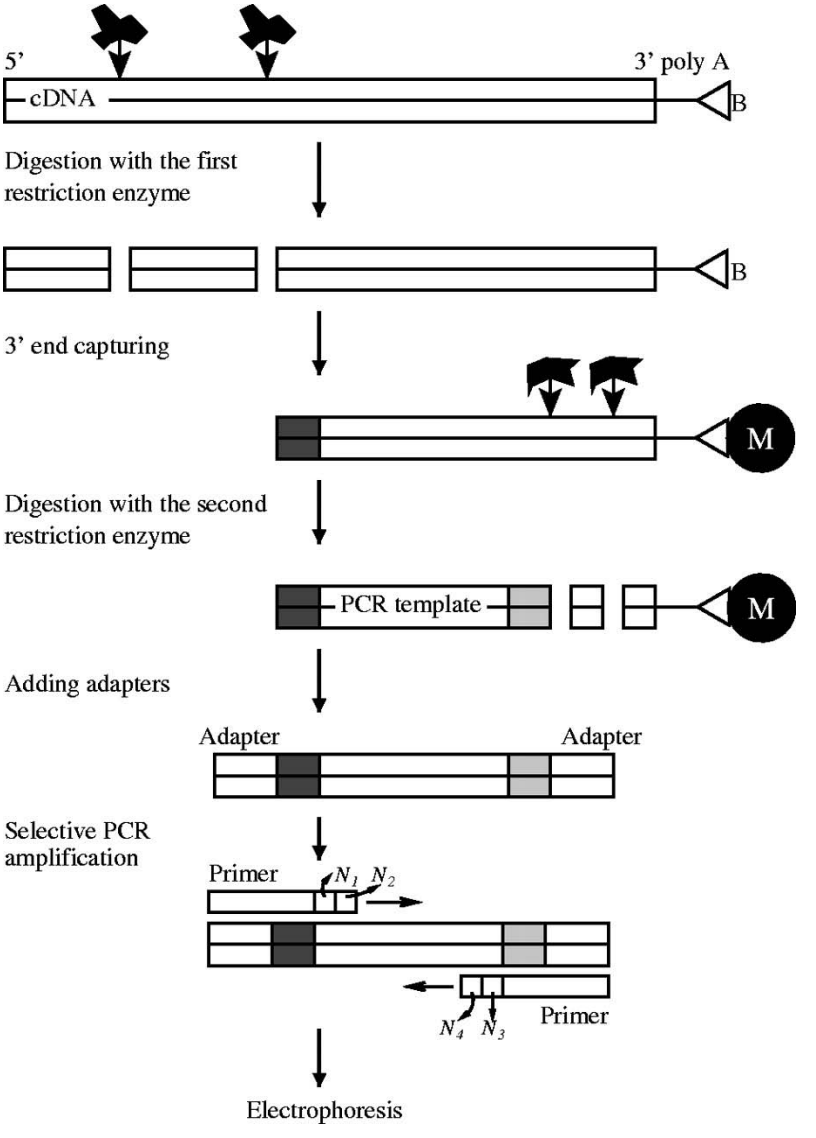


Figure 2. cDNA-AFLP with 3' end capture (Breyne and Zabeau 2001; Fukumura et al. 2003). *After the digestion with the first restriction enzyme, the biotin (B) labelled 3'-fragments are captured with the aid of streptavidin coated magnetic beads (M) followed by the digestion with the second restriction enzyme and the removal of the 3'-fragments. Adapters have to be attached to the specific ends (coloured) of the fragments to facilitate selective PCR (the selective nucleotides are marked as N_1, N_2, N_3 and N_4).*

Because of the PCR steps, cDNA-AFLP and cDNA subtraction cloning can be very sensitive methods. Theoretically, a successful PCR amplification can start from just one molecule. The efficiency of PCR is influenced by the length and sequence of the template. PCR amplifications of the whole sample have been also used in microarray techniques to improve their sensitivity. However, the variations in PCR efficiency limit this approach. For cDNA-AFLP, it has been claimed that simultaneous amplification of dozens of fragments is reproducible in one reaction when using a single, common, primer pair (Vos et al. 1995). Thus comparisons of identically produced pools should be reliable.

In comparison to the cDNA subtraction library method, cDNA-AFLP could have a lower false positive rate. As well as current microarray methods, cDNA subtraction libraries and cDNA-AFLP require a verification of the gene expression by an additional method. cDNA-AFLP allows comparisons of several samples simultaneously and the determination of quantative expression profiles (Breyne et al. 2003). However, the process of extracting the ESTs from electrophoresis gels is laborious. In contrast, the sequencing of cDNA subtraction libraries can be easily automated.

1.4 Protein secretion stress in filamentous fungi

Pezizomycotina are known to secrete abundantly a wide variety of proteins. Secretion is initiated in the nucleus by transcription, subsequently the transcripts are translated and proteins translocated to the endoplasmic reticulum (ER). The folding of secreted proteins is supported by chaperones, glycosylation enzymes and oxidoreductases. Of these BIP (binding protein), a chaperone, and PDI (protein disulphide isomerase), a disulphide bond formation catalyst, are the most commonly used indicators for ER status. Correct folding is monitored by the quality control (QC) system and proteins might go through repeated cycles of folding before destined for either the ERAD pathway (ER – Associated Degradation) or further transport (for review (Tsai et al. 2002; Ellgaard and Helenius 2003; Ma and Hendershot 2004)). The correctly folded proteins are transported to the Golgi compartment where further modification of the proteins takes place to modulate their function and stability (for review (de Graffenried and Bertozzi 2004)). After Golgi, proteins are transported in vesicles to the plasma membrane and released out of the cell.

In contrast to Saccharomycotina, Pezizomycotina live predominantly in a filamentous form and it has been shown that their secretion often occurs at the hyphal tip (Nykanen et al. 1997; Gordon et al. 2000). The vesicular transport of secreted proteins from other parts of the hyphae to the tip is believed to be mediated by microtubules, while short range transport between cell compartments is mediated by the actin filaments (Mourino-Perez et al. 2006). The role of actin and microtubules has been debated for long. Although actin and not microtubules are essential for hyphal growth, it is clear that microtubules are needed for fast growth (Fuchs et al. 2005; Horio and Oakley 2005; Sampson and Heath 2005; Mourino-Perez et al. 2006). For example, in *A. nidulans* a hyphal tip, possibly channelling the secretion capacity of the preceding cells, can grow up to five times faster than a germling (Horio and Oakley 2005). Thus, while microtubules are not essential in laboratory conditions they might be essential in nature by allowing the fungus to grow fast towards a source of carbon. At the tip of the growing hyphae is a structure unique to filamentous fungi, the Spizenkörper (SPK). It is believed to act as a vesicle supply centre (VSC) before final secretion out of the cell (Gierz and Bartnicki-Garcia 2001). Although SPK is rich in actin, a recent mathematical VSC model predicts that no active transport is needed between the SPK and the cell membrane (Tindemans et al. 2006).

1.4.1 Unfolded Protein Response

Accumulation of unfolded, misfolded or otherwise inefficiently secreted proteins or other impairment of secretion can cause stress to cells, i.e. secretion stress. It is often studied by exposing cells to various chemicals that inhibit protein folding or transport and induce strong, clearly measurable responses. The most commonly used chemicals are DTT and tunicamycin. DTT, a powerful reducing agent, inhibits protein folding by preventing the formation of disulphide bonds. Tunicamycin prevents N-linked glycosylation of proteins by blocking the first step of the glycosylation pathway.

The major eukaryotic response to secretion stress is the Unfolded Protein Response (UPR). In *S. cerevisiae*, UPR is believed to modify and enhance the activity of the secretion pathway affecting specifically genes involved in the ER functions. It is defined mainly through its transcriptional effects that are

controlled by the sensor Ire1p and the downstream transcription factor Hac1p. The chaperone BIP binds to Ire1p when not bound by unfolded proteins and keeps it in an inactive form (Bertolotti et al. 2000). However, Ire1p can sense secretion stress and subsequently activate UPR even without the capability to bind BIP (Kimata et al. 2004). In concert, it has been proposed that Ire1p could bind directly to unfolded proteins (Credle et al. 2005). Activated Ire1p promotes splicing of *HAC1* mRNA by an unconventional mechanism and only then Hac1p is actively translated and capable of activating its downstream genes (Cox and Walter 1996).

An upstream promoter element UPR Element (UPRE), bound by Hac1p has been shown to be necessary and sufficient for the UPR induction of *BIP*, *PDI1*, *EUG1*, *FKB2* and *LHS1* in *S. cerevisiae* (Mori et al. 1992; Kohno et al. 1993; Mori et al. 1998). Furthermore, the induction of 381 genes has been shown to depend on the *IRE1* and *HAC1* pathway. Of these 27 percents were involved in secretion related processes, such as translocation, glycosylation, protein folding, lipid and inositol metabolism, protein degradation, ER – Golgi transport, Golgi – ER retrieval, vacuolar targeting, distal secretion and cell wall biogenesis (Travers et al. 2000). The majority of these genes also require the Gcn4p for their induction (Patil et al. 2004). Gcn4p is a member of the bZIP (basic leucine-zipper) transcription factor family like Hac1p. It is the major controller of the amino acid starvation response (for review (Hinnebusch and Natarajan 2002)). Three elements all resembling UPRE and the binding site of Gcn4p as well have been found in the promoter regions of 36 percents of the UPR induced genes. Concomitant Gcn4p and Hac1p binding was shown for two element types that are found in 34 percents of the UPR induced genes. As bZIP transcription factors frequently form dimers, it could be that Hac1p and Gcn4p bind promoters as a heterodimer (Patil et al. 2004).

In conditions of extreme stress, such as combined DTT treatment and heat shock, an additional response, the S-UPR (Super-UPR), has been described in *S. cerevisiae*. Its activation requires Hac1p and an unrecognised transcription factor, UMF (UPR Modulating Factor) (Leber et al. 2004). Hac1p acts also as a negative regulator of transcription. Active Hac1p represses differentiation responses to nitrogen starvation, pseudohyphal growth and meiosis, apparently through a mechanism of tightening nucleosome structure. Interestingly, this mechanism links the ER load to *S. cerevisiae*'s differentiation programs (Schroder et al. 2000; Schroder et al. 2004).

The response to secretion stress in *T. reesei* has been previously shown to share several features with that in *S. cerevisiae*. Components of the UPR pathway have been isolated from *T. reesei* including the counterparts of the genes *IRE1* and *HAC1*, as well as UPR target genes such as *PDI1* (Saloheimo et al. 1999; Saloheimo et al. 2003; Valkonen et al. 2004). In *T. reesei* and *A. niger*, splicing of *hac1/hacA* mRNA and HAC1/HACA promoter binding activity has been shown. However, the process of mRNA splicing is different from *S. cerevisiae*. Instead of removing a 250 nt intron, the 5' flanking region is truncated and a 20 nt intron is removed (Saloheimo et al. 2003; Mulder et al. 2004).

The production of a heterologous protein is often stressful to cells, which can lead to reduced yields of proteins. For example, in *T. reesei* production of human antibody Fab fragment caused induction of *PDI1* (Saloheimo et al. 1999) and in *A. niger* expression of tPA (human tissue plasminogen activator) induced UPR (Wiebe et al. 2001). Vice-versa, constitutive induction of UPR can increase the production of heterologous proteins in fungi (Valkonen et al. 2003a; Valkonen et al. 2003b).

Even induction of massive secretion of hydrolytic enzymes by change of carbon source in *T. reesei* can induce the UPR. When a strain carrying a modified homologous egl1 (endoglucanase 1) cellulase was grown up with glucose and then transferred to lactose to initiate protein secretion, UPR was induced in the modified strain and, interestingly, also in its parental strain. The transcriptional UPR induction followed closely the transcriptional induction of the main cellulases (Collen et al. 2005). Hence, UPR might be a natural event in *T. reesei's* differentiation from a purely free sugar consuming state to an active secreting and thus degrading state. In parallel, UPR is required for differentiation of highly secreting mammalian cells. Like fungi, mammals have an IRE1 and bZIP transcription factor based UPR activation mechanism. The homologue of *HAC1* in mammals is XBP1. Its induction and subsequent UPR are a prerequisite of normal β-cell differentiation and UPR is probably relevant also for less studied types of secreting cells like hepatocytes and osteoblasts (for review (Wu and Kaufman 2006)). Combined, the transcription inductions and repressions of UPR appear as important regulators of eukaryotic differentiation.

1.4.2 Repression Under Secretion Stress

In addition to the Ire1p, PERK (PKR-like ER protein kinase), a similar sensor protein, resides on mammalian ER. PERK is activated by unfolded proteins in the ER. Once activated PERK phosphorylates the eIF2α (α subunit of the translation initiation factor 2). Phosphorylation of eIF2α leads to attenuation of general translation initiation and subsequently diminishes protein load in the ER (Harding et al. 2000a & 2000b).

Ascomycota do not appear to have a gene encoding PERK homologue in their genomes, but a different mechanism exists to relieve protein load inside ER. Under secretion stress conditions, *T. reesei* and *A. niger* exhibit a transcriptional down regulation of genes encoding secreted proteins (REpression under Secretion Stress, RESS). RESS commonly appears together with UPR, but not necessarily in similar scale. Even transcriptionally a very mild UPR response might coincide with a transcriptionally strong RESS. The mechanism of RESS is unknown, but it has been shown that shortening of the promoter of cbh1 (cellobiohydrolase 1), a major secreted cellulase, abolishes its RESS regulation in *T. reesei* (Pakula et al. 2003; Al-Sheikh et al. 2004). Thus the regulation is likely to occur at the level of transcription. A similar transcriptional down regulation of secreted proteins has been described in *A. thaliana* (Martinez and Chrispeels 2003). RESS has not been described in *S. cerevisiae*, but a recent microarray study suggested that a constitutively active Hac1p might down regulate proteins passing ER, such as transporters and proteins of periplasmic space and cell wall (Kimata et al. 2006). However, previous microarray work has also shown induction of some genes encoding proteins passing ER in UPR causing conditions (Travers et al. 2000).

1.4.3 Oxidative stress

Unfolded proteins appear to cause oxidative stress through a complicated and incompletely described process. Treatment of mouse cells by tunicamycin causes, beside UPR induction, oxidative stress (Harding et al. 2003). Along the same line, production of a mutated misfolding protein in *S. cerevisiae* ERAD deficient cells, kills cells by Reactive Oxygen Species (ROS). Both of these stresses can be alleviated by the addition of glutathione (GSH). The ROS

production requires the induction of UPR and specifically the induction of *ERO1* (ER oxidation 1 protein) (Haynes et al. 2004).

During disulphide bridge formation electrons are passed to oxygen by Ero1p producing approximately one hydrogen peroxide per 20 disulphides formed (Tu and Weissman 2002). In addition failure of any of the several electron transfer reactions required in disulphide bridge formation could cause additional oxidative damage. GSH can protect cells from oxidative damage, but large amounts of folding protein can deplete GSH storages (Figure 3). However, the exact role of glutathione in disulphide bridge formation and its transport between cytosol and ER remains unclear. The up regulation of *ERO1*, *PDI1* and other folding factors by UPR increases ER's folding capacity, but simultaneously seems to cause further oxidative stress (for review see (Tu and Weissman 2004; Chakravarthi et al. 2006)).

Both *S. cerevisiae* and mammals share a similar response to amino acid deprivation. Uncharged tRNAs activate the GCN2 kinase that phosphorylates eIF2α, which leads to attenuation of general translation initiation. Paradoxically, translation of some proteins, such as the mammalian transcription factor ATF4 and its *S. cerevisiae* homologue *GCN4*, is induced (for review (Hinnebusch and Natarajan 2002)). In mammalian secretion stress the activation of PERK by unfolded proteins leads to translation attenuation and ATF4 activation. However, the ATF4 mediated responses to secretion stress and amino acid deprivation appear to be distinct (Harding et al. 2000b; Harding et al. 2003; Averous et al. 2004). In secretion stress ATF4 transcriptionally up-regulates genes in amino acid biosynthesis related functions apparently to relieve oxidative stress by producing more glutathione (Harding et al., 2003).

Consistent with the fact that no PERK homologue has been found in Ascomycota neither tunicamycin nor DTT treatment appear to lead to translation attenuation in *S. cerevisiae* (Cherkasova and Hinnebusch 2003; Deloche et al. 2004; Rand and Grant 2006), nor does tunicamycin cause phosphorylation of eIF2α. However, the translation of *GCN4* is still induced (Patil et al. 2004). Whether this leads to increase in glutathione biosynthesis is not known.

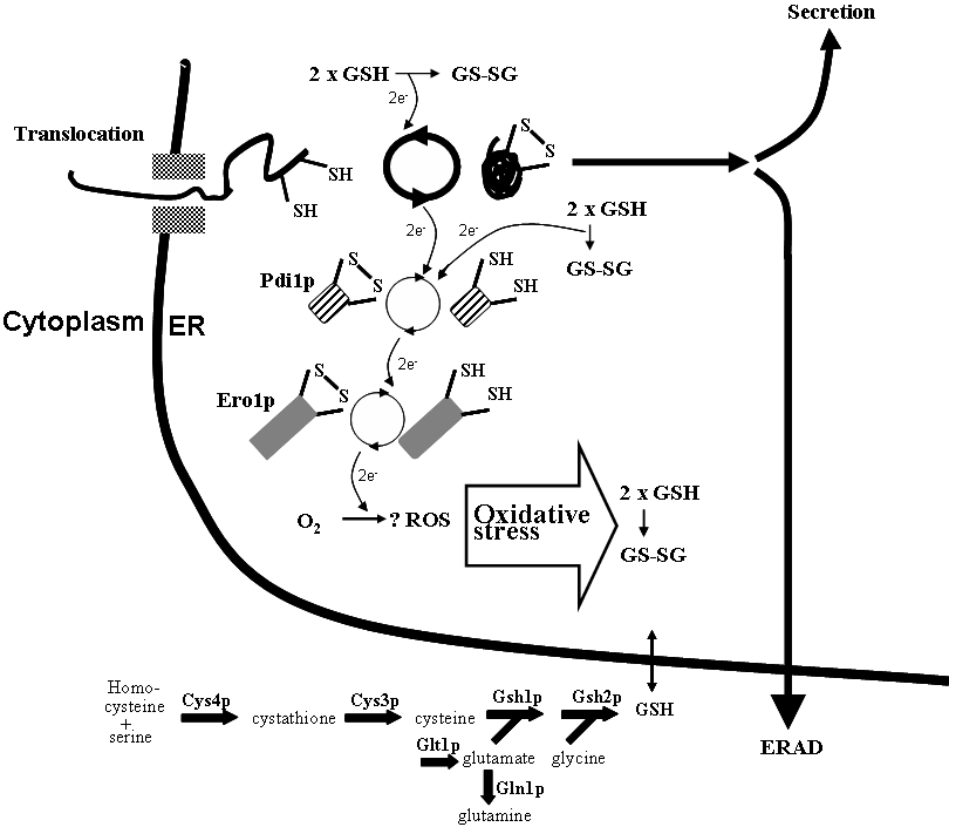


Figure 3. Major electron transfer reactions leading to oxidative stress and the role of glutathione inside ER. After translocation, formation of disulphide bonds (S-S) is catalysed primarily by Pdi1p. Correctly folded proteins are transported further in the secretion pathway, while incorrect S-S bonds are broken and proteins refolded or destined for degradation by ERAD. The S-S breakage mechanism probably involves oxidized glutathione (GS-SG) and possibly protein disulphide isomerases. The biosynthesis pathway leading to glutathione is well known, but the transport of GSSG and GSH between cytosol and ER, as well as the roles of ER glutathione reductases, are unclear. Adapted from (Tu and Weissman 2004; Chakravarthi et al. 2006).

1.5 Aims of the study

The cellular mechanisms involved in protein production have been studied in a number of Pezizomycotina fungi for a long time, with the aim of improving these as industrial production organisms. Both random mutagenesis and manipulation of individual genes implicated in functional studies in these fungi, as well as the yeast *S. cerevisiae*, have been successfully applied in strain engineering. The aim of this study was to develop and apply genome wide strategies to screen for novel characteristics of Pezizomycotina, whose manipulation could advance their performance in industry.

More specifically the aims of this study were:

- to test and improve previously developed transcriptome profiling methods for organisms with non-sequenced genomes in the analysis of *T. reesei*
- to analyse the response of *T. reesei* to secretion stress by transcriptional analysis
- with the availability of the genome of *T. reesei* and many other fungi, to use comparative genomics to improve our understanding of fungal genome structure and content

2. Materials and methods

Method	Used in
A. nidulans shake flask cultivation	V
cDNA subtraction library construction	V
cDNA-AFLP	V
cDNA-AFLP pool selection	IV
cDNA-AFLP primer selection	IV
Determination of intracellular amino acid concentrations by HPLC	V
DTT treatment of *A. nidulans*	V
DTT treatment of *T. reesei*	V
EST sequence analysis	V
EST sequencing	V
Northern hybridisation	V
Principal component analysis of InterPro annotations	I
Protein cluster annotation	I
Protein cluster enrichment analysis	I
Protein cluster species distribution clustering	I
Protein sequence and annotation data retrieval	I
Protein sequence clustering	I
Protein sequence InterPro analysis	I
Protein sequence Protfun analysis	I
qRT-PCR	II
S. cerevisiae batch bioreactor cultivation	II
S. cerevisiae mRNA extraction	II
S. cerevisiae transcriptome profiling data comparison	V

T. reesei and *A. nidulans* mRNA extraction	V
T. reesei batch bioreactor cultivation	V
T. reesei shake flask cultivation	V
T. reesei chemostat cultivation	V
TRAC PCR primer and probe design	II, III
TRAC PCR probe production	II
TRAC primer purification	II
TRAC probe set division to pools	III
TRAC with PCR probes and PCR sample amplification	II

3. Results and discussion

3.1 Comparative analysis of fungal genomes

Individual fungal genomes have been studied in detail and extensive comparative genomics of subphylum Saccharomycotina has been done (for review (Dujon 2006)). However, no comprehensive comparative genomic work of a wider phylogenetic group of fungi has been published. As the analysis and methodology used in studies with individual fungal genomes vary, it is hard to estimate the relevance of the obtained results in comparison to other fungi. In general the content of Saccharomycotina genomes is better known, while Pezizomycotina have twice as many ORFs (Table 1) and are less studied. In order to study the extent of previous findings and discover novel features we compared computationally Saccharomycotina and Pezizomycotina genomes with a dataset of 33 genomes. Of these 14 belong to Pezizomycotina and 13 to Saccharomycotina, in addition one Archeascomycota, four Basidiomycota and one Zygomycota were included as controls (I: Table 1).

To compare the genomes we used two complementary, but fundamentally different approaches, TRIBE-MCL protein clustering (Enright et al. 2002) and InterProScan analysis (Quevillon et al. 2005). Based on similarity scoring by an all-against-all comparison with the Blast program (Altschul et al. 1990) TRIBE-MCL divides genomes into protein clusters. InterPro is a database that merges several protein domain, family, superfamily, active-site etc. databases into one protein feature classification (Mulder et al. 2005). InterProScan detects these features, i.e. InterPro entries, from protein sequence. While TRIBE-MCL is comprehensive and uniform as it tries to divide all proteins into similar clusters, InterProScan is not. It detects only known features of widely different types, but it conveniently summarises current knowledge of protein features.

3.1.1 Overview of clustering results

To find out how well our analysis covers the genomes studied, we counted the percent of ORFs in orphan clusters and percent of ORFs without any InterPro entries for the species analysed with InterProScan (Figure 4). In bacterial

genomes the percent of ORFs with no homology in sequence databases varies from 1 to 50 percents and correlates with general sequence similarity of a genome to other sequenced genomes (Fukuchi and Nishikawa 2004). Very few Hymenomycetes species have been sequenced and accordingly *P. chrysosporium* and *Coprinus cinereus* have both high values in both counts. Closely related Eurotiomycetes species e.g. Aspergilli (Figure 1) have particularly low percents of unrecognised ORFs. Among Saccharomycotina, *Y. lipolytica* has the highest count of ORFs in orphan clusters. It also differs from other Saccharomycotina by a particularly large genome (Table 1) and it has diverged early from other Saccharomycotina (Figure 1). These characteristics show that it is particular among sequenced Saccharomycotina.

Interestingly, Sordariomycetes (*T. reesei*, *F. graminerum, C. globosum*, *M. grisea* and *N. crassa*) is divided to those with low unrecognised ORF values (*T. reesei* and *F. graminerum*) and those with higher values (*C. globosum*, *M. grisea* and *N. crassa*). *T. reesei* and *F. graminerum* have similar values than Eurotiomycetes. As *N. crassa*'s high values could be explained by RIP, this could apply also for its close relatives.

It has been proposed that 10–30 percents of predicted ORFs in microbial genomes might not be actual genes (Skovgaard et al. 2001). Although genome annotation has improved since, our analysis is likely to cover the studied Ascomycota genomes very well as orphans are more likely to be false positives than conserved ORFs.

The level of paralogy, i.e. how many ORFs of a genome have arisen through duplication, reflects major forces that have shaped a genome. Currently, there are no techniques to count the true level of paralogy. Instead, we counted genomic ORF redundancy. It was defined as the percent of ORFs in a genome found in clusters with more than one ORF from the same genome (Dujon et al. 2004) (Figure 5).

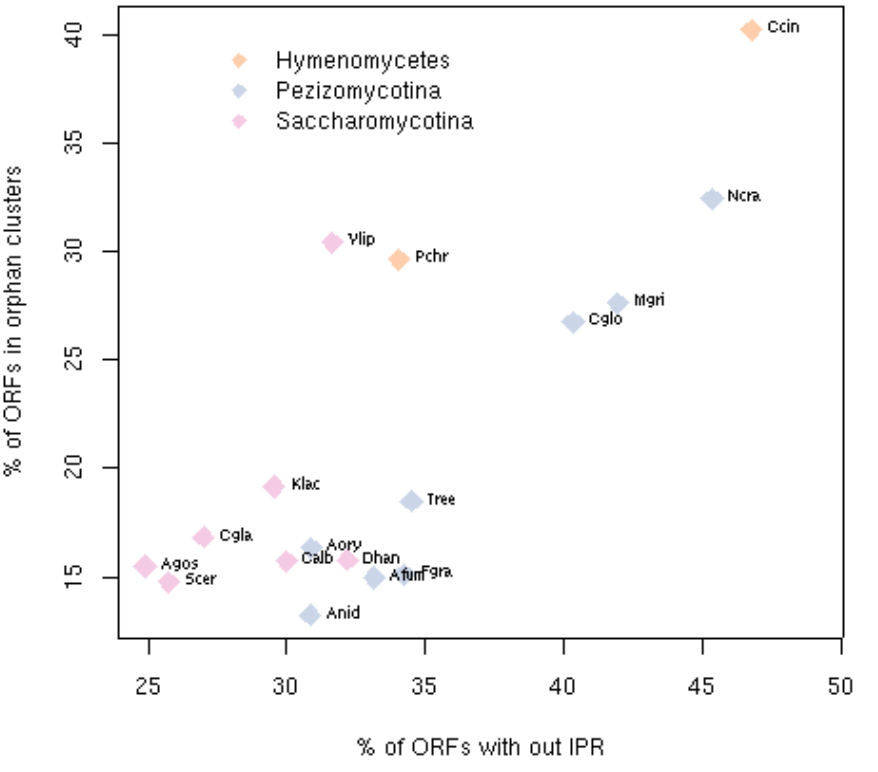


Figure 4. Unrecognised by InterPro or clustering. Percent of ORFs in orphan clusters (y-axis) versus the percent of ORFs unrecognised by InterPro (x-axis) for species which where analysed with InterProScan. Species are coloured by subphyla. Species name abbreviations: Spom, Schizosaccharomyces pombe; Afum, Aspergillus fumigatus; Anid, Aspergillus nidulans; Anig, Aspergillus niger; Aory, Aspergillus oryzae; Bcin, Botrytis cinerea; Cglo, Chaetomium globosum; Cimm, Coccidioides immitis; Fgra, Fusarium graminearum; Mgri, Magnaporthe grisea; Nhae, Nectria haematococca; Ncra, Neurospora crassa; Sscl, Sclerotinia sclerotiorum; Snod, Stagonospora nodorum; Tree, Trichoderma reesei; Agos, Ashbya gossypii; Calb, Candida albicans; Cgla, Candida glabrata; Cgui, Candida guilliermondii; Clus, Candida lusitaniae; Dhan, Debaryomyces hansenii; Klac, Kluyveromyces lactis; Ppas, Pichia pastoris; Psti, Pichia stipitis; Scas, Saccharomyces castellii; Scer, Saccharomyces cerevisiae; Sklu, Saccharomyces kluyveri; Ylip, Yarrowia lipolytica; Ccin, Coprinus cinereus; Cneo, Cryptococcus neoformans; Pchr, Phanerochaete chrysosporium; Umay, Ustilago maydis; Rory, Rhizopus oryzae.

Saccharomycotina species *S. cerevisiae* and *Saccharomyces castellii* have particularly high values of genomic ORF redundancy. They have undergone a recent genome duplication (Dujon et al. 2004). *Candida glabrata* belongs to the same duplication lineage, but has a more average value. *N. crassa*, due to its extreme RIP mechanism, has the lowest genomic ORF redundancy in relation to its genome size (Galagan et al. 2003). *N. crassa* has also the lowest counts of glycoside hydrolases and secondary metabolism related ORFs among Pezizomycotina, while Eurotiomycetes in general have more glycoside hydrolases than Sordariomycetes (Galagan et al. 2003; Dean et al. 2005; Galagan et al. 2005; Machida et al. 2005; Nierman et al. 2005; Pel et al. 2007). *Aspergillus* species and *M. grisea* have been proposed to have a milder RIP like mechanism (Ikeda et al. 2002; Montiel et al. 2006). Our analysis does not exclude this possibility. Whether the lower count of glycoside hydrolases in Sordariomycetes are due to a more active RIP, or is a specific adaptation to environment, remains to be seen. *N. crassa* and *S. cerevisiae* are probably the most commonly used fungal model organisms and thus it is notable that with regard to genomic ORF redundancy they have the most exceptional genomes.

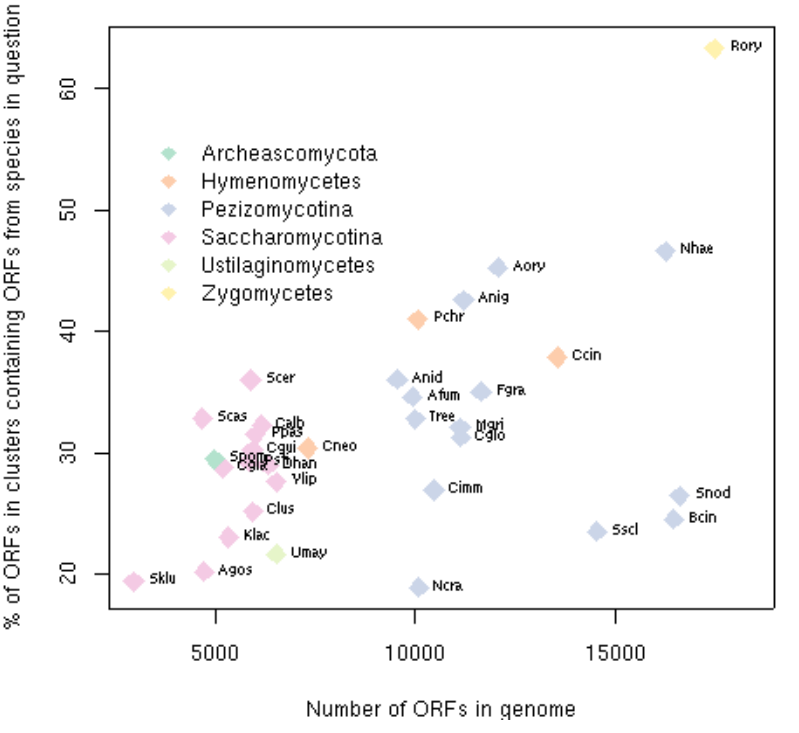


Figure 5. Genomic ORF redundancy. Percent of ORFs in clusters containing multiple ORFs from the species in question i.e. genomic ORF redundancy (y-axis) versus the size of the genome in ORFs (x-axis) for each species. See Figure 4 for further details.

3.1.2 Characterisation of clusters

In order to divide the clusters obtained to phylogenetically interesting groups, clusters containing at least 10 proteins were ordered by hierarchical clustering based on how many proteins of each species were included in each cluster (Figure 5). Subsequently four regions with interesting species distributions were selected for detailed analysis, A: "Pezizomycotina abundant" (I, Figure 6), B: "Pezizomycotina specific" (I: Figure 7), C: "Saccharomycotina absent" (I: Figure 8) and D: "Saccharomycotina unique" (I: Figure 9). Clusters of each region were analysed for enrichment of Funcat categories (I: Table 2 for A: "Pezizomycotina abundant" and I: Table 3 for others), InterPro entries (I: Table 4) and Protfun categories (I: Table 5) to characterise the content of regions. Funcat

is a protein function classification system based on extensive manual annotation (Ruepp et al. 2004). Protfun predicts protein function based on sequence, but does not use any sequence homology information (Jensen et al. 2003). In addition, we checked which clusters contained an *S. cerevisiae* ORF found in the metabolic model iMH805 (Herrgard et al. 2007). The model comprehensively describes the known primary metabolism of *S. cerevisiae*.

The 198 clusters in the region A: "Pezizomycotina abundant" have usually more ORFs in Pezizomycotina than in any other subphyla and none in Saccharomycotina (Figure 5, I: Figure 6). 20 percents of clusters were predicted to encode secreted proteins, while for 36 percents secretion status could not be assigned because of heterogeneous predictions. Because the software behind the predictions, TargetP (Emanuelsson et al. 2000), was trained with only a few known fungal secreted proteins, it might miss many of them. The clusters in the region are relatively well characterised as 63 percents have a Funcat assignment. The results of all three annotation analyses of these clusters agree well (I: Tables 2, 4 and 5). The region contains enzymes involved either in secondary metabolism or the secreted plant biomass degradation machinery. Funcat analysis found far more clusters involved in plant biomass degradation than InterPro analysis. However, InterPro describes glycoside hydrolases and glycosidetransferases rather poorly compared to for example the specialised Carbohydrate-active enzymes (CAZY) database (Coutinho and Henrissat 1999).

As the clusters of the region A: "Pezizomycotina abundant" typically have several ORFs in each Pezizomycotina species in contrast to B: "Pezizomycotina specific", it seems that these clusters represent the families that have recently expanded. Many of the enzymes in clusters of A: "Pezizomycotina abundant", such as cytochrome P450s, glycoside hydrolases and tyrosinases, are at the edge of metabolic networks with roles in synthesising final secreted products or degrading external sources of carbon. As discussed above, *N. crassa* has the lowest counts of plant biomass degradation and secondary metabolism related ORFs among Pezizomycotina. *N. crassa* has also on average the lowest counts of ORFs in A: "Pezizomycotina abundant" region among Pezizomycotina. It appears that the development of *N. crassa* RIP preceded the expansion of these enzymes at the edges of the metabolic network.

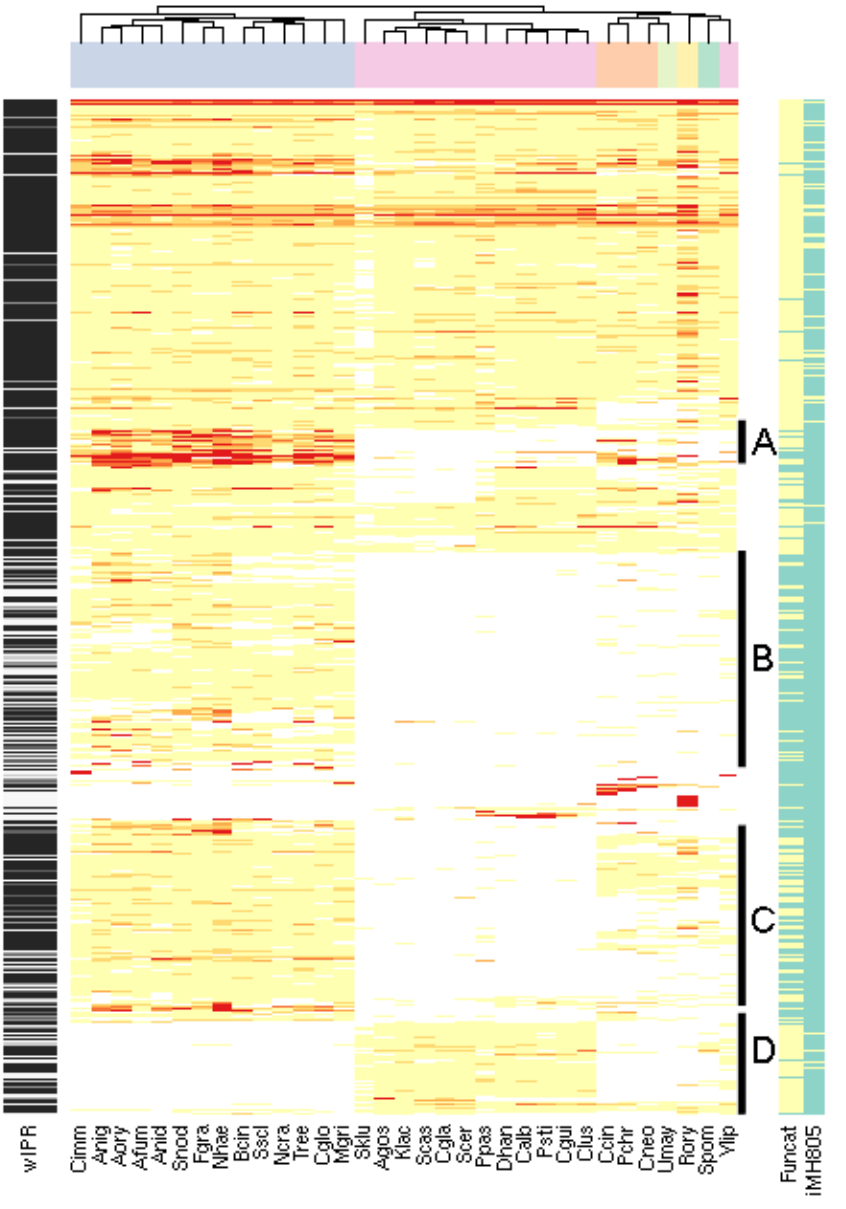
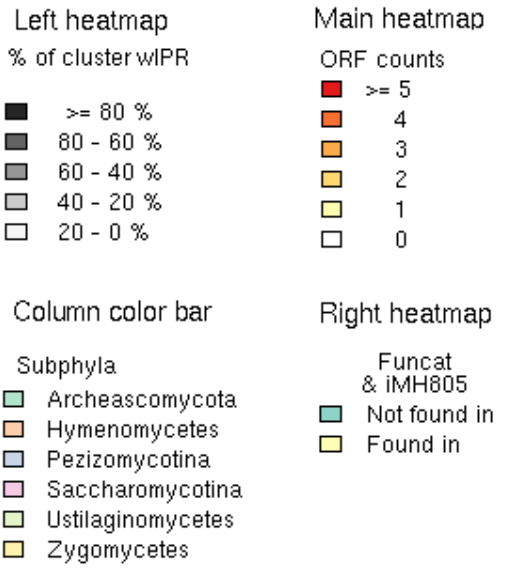


Figure 6. Protein clustering overview – A heatmap of clusters with at least ten member ORFs. In the main heatmap colour intensity of a cell shows the number of ORFs shown by clusters (rows) and by species (columns). Both rows and columns are ordered by hierarchical clustering to group similar rows or columns together. The dendrogram from hierarchical clustering is shown for columns and the phylum of species is indicated by a column colour bar between the

heatmap and the dendrogram. Under the heatmap each species is specified by an abbreviation explained in Figure 4. On the left side heatmap the percent of ORFs in a cluster that have an InterPro entry of all cluster's ORFs analysed with InterProScan ("wIPR") is shown. On the right side heatmap, whether or not the cluster has a Funcat classification ("Funcat") or has an ORF found in S. cerevisiae metabolic model iMH805 is shown ("iMH805").



The 1130 clusters in the region B: "Pezizomycotina specific" have usually just one ORF in each Pezizomycotina species and none in the other subphyla (I: Figure 7). In comparison to A: "Pezizomycotina abundant" clusters, far fewer member ORFs have any InterPro entries (Figure 6, column "wIPR"), only 20 percents have Funcat assignments and very few InterPro or Funcat annotations show enrichment (I: Tables 3 and 5). Most proteins in these clusters were predicted not to have transmembrane regions and 16 percents of them were predicted not to be enzymes. 11 percents of the clusters were predicted to have a secretion pathway signal sequence. Cellular role predictions "Cell envelope", "Regulatory functions", "Central intermediary metabolism" "Purines and pyrimidines" show significant enrichment. The enrichment of "Regulatory functions" and InterPro entries "IPR001138: Fungal transcriptional regulatory protein, N-terminal" and "IPR007219: Fungal specific transcription factor" coincide well. These two InterPro entries commonly occur in the same protein (see chapter 3.1.3).

The 1010 clusters in the region C: "Saccharomycotina absent" have usually just one ORF in each Pezizomycotina species and none in Saccharomycotina (I: Figure 8). According to Protfun predictions these are mostly enzymes and 43 percents have a Funcat assignment. Still both cluster InterPro entry and Funcat

category enrichment provide little detail, except for proteins involved in fatty acid β-oxidation in mitochondria. The capability for fatty acid β-oxidation in mitochondria has been shown experimentally for *A. nidulans* (Maggio-Hall and Keller 2004) and proposed by genome analysis for several other Pezizomycotina species (Maggio-Hall and Keller 2004; Galagan et al. 2005). However, Saccharomycotina can apparently only carry out fatty acid β-oxidation in peroxisomes (Kunau et al. 1988; Hiltunen et al. 1992; Kurihara et al. 1992; Smith et al. 2000). An analysis of the metabolic network around β-oxidation proposes a reason for other fungi than Saccharomycotina to have this functionality. It could provide precursors from the fatty acid and amino acid metabolism to the polyketide synthesis necessary for secondary metabolism (I: Figure 10).

The regions B: "Pezizomycotina specific" and C: "Saccharomycotina absent" contain protein families of lower internal sequence similarity that subsequently divide into small equally sized clusters in contrast to A: "Pezizomycotina abundant". They could contain proteins more connected in cellular metabolism and regulatory networks than clusters in the region A: "Pezizomycotina abundant" and thus be under more pressure not to expand to keep the topology of the networks conserved. This would allow the sequences to diverge during evolution without simultaneous expansion of the families and result into small clusters broadly conserved in Pezizomycotina.

Based on the distribution of the clusters with ORFs from a *S. cerevisiae* metabolic model, it is clear that over 90 percents of these ORFs are well conserved in all fungal species. These form the conserved core of a fungal metabolic network. B: "Pezizomycotina specific" and C: "Saccharomycotina absent" could then represent a mid layer around this core. Examples of these are the ORFs shown in I: Figure 10, that link primary metabolism to secondary metabolism and biomass degradation machinery. Biomass degradation machinery and secondary metabolism could then form the edges of the network. Being on the outer edge of the metabolic network, evolutionary pressure to maintain network topology would minimally limit their evolution and vice versa they could also be directly subjected to various evolutionary pressures created by changing environments and thus evolve fast.

Localisation of ORFs to subtelomeric regions could provide the necessary mechanism for fast ORF content evolution (Naumov et al. 1996; Freitas-Junior et al. 2000; Rehmeyer et al. 2006). Subtelomeric regions have been shown to contain secondary metabolism related genes, glycoside hydrolases and predicted secreted uncharacterised ORFs in *Magnaporthe oryzae* (Rehmeyer et al. 2006).

The 466 clusters in the region D: "Saccharomycotina unique" have usually just one ORF in each Saccharomycotina species and none in the other subphyla (Figure 9). Various Protfun cellular role predictions are enriched in these clusters, among them translation, transcription and mitochondrion related and these are also found also in Funcat enrichment analysis. Strikingly, clusters from Funcat category "42.16: Biogenesis of cellular components: Mitochondrion" correspond to at least 28 different components of mitochondrial ribosomal protein complexes. This implies that these complexes have diverged significantly between Pezizomycotina and Saccharomycotina as the corresponding ORFs are, however, found in Pezizomycotina. Given the differences in glucose metabolism between Ascomycota discussed in 1.2.1 and the Saccharomycotina strategy of fast growth and glucose consumption, it is not surprising that the mitochondria of Pezizomycotina and Saccharomycotina would be fundamentally different.

3.1.3 Principal Component Analysis results and their comparison to clustering

TRIBE-MCL clustering produced 77.278 clusters with only a single member however, 11.727 of these had InterPro entries. In addition the analysis of the clusters presented above dealt only with clusters of at least ten member ORFs. We used PCA (Principal Component Analysis) to find differences between counts of ORFs with InterPro entries independently of clustering to try to cover parts of the genomes omitted by cluster analysis. This analysis was only done for a representative subset of genomes (I: Table 1).

PCA finds principal components (PC), i.e. major sources of variation between the species and positions of the original samples and the loadings i.e. contributions of individual InterPro entries on the PCs. We found that the two most important PCs separated Pezizomycotina and Saccharomycotina and Basidiomycota and Ascomycota (I: Figure 11). Therefore, the InterPro entries

with highest loadings on these PCs have the largest differences between the respective subphyla. Less than a hundred InterPro entries were found to have a major contribution (I: Figure 12a) and these hundred entries were selected for further study (TOP100, I: Figure 13).

TRIBE-MCL clustering and PCA analysis of InterPro counts complemented each other well. The TOP100 InterPro entries contain interesting details not detected by clustering. We classified the entries into rough functional categories and found that the five most common, with their counts, were: "Macromolecule interaction, 19", "Secondary metabolism, 18", "Dubious, 9", "Protein modification, 8" and "Transporter, 7". InterPro entries assigned to "Dubious" are those that InterPro itself considers unreliable. Higher amounts of entries related to secondary metabolism and transporters in Pezizomycotina and higher amounts of macromolecule interaction related entries in Saccharomycotina contribute most to the differences between the subphyla. These transporter entries belong to the Major Facilitator Superfamily (MFS). They are transporters found for example in fungal secondary metabolite clusters for gliotoxin (Gardiner et al. 2005) or trichothecene (Alexander et al. 1999) synthesis. Small amounts of transposon related entries in Ascomycota, in contrast to Basidiomycota, seem as the subphyla's most important difference. This result coincides with the particular nature of fungal genome evolution discussed in chapter 1.2.2.

Enrichment of "IPR001138: Fungal transcriptional regulatory protein, N-terminal" was detected by both clustering and PCA. IPR001138 is a DNA binding N-terminal zinc binuclear cluster (Zn_2Cys_6) domain found in many fungal transcription factors such as *S. cerevisiae* GAL4 (Hashimoto et al. 1983), *A. niger* xlnR (van Peij et al. 1998) and *Aspergillus flavus* aflR (Woloshuk et al. 1994). On the average a Pezizomycotina has three times more ORFs with IPR001138 than a Saccharomycotina. The full importance of this expansion was only revealed by PCA, because there are 614 clusters with less than 10 members that have an IPR001138. As InterPro domain analysis uses searches optimised for individual protein families or domains, it can separate well very similar protein families such as MFS transporters and detect well very diverged protein families such as IPR001138, as long as the family has been previously well characterised.

The major difference between clustering and PCA results is that in clustering IPR001138 is found in B: “Pezizomycotina specific” and secondary metabolism related clusters fall into A: “Pezizomycotina abundant”, while PCA makes no such distinction. As secondary metabolism seems to be the major function connecting InterPro entries more abundant in Pezizomycotina, it could be expected that IPR001138 family has expanded to control secondary metabolism and that MFS transporters would have expanded to transport the produced metabolites. However, likely targets are also the many plant biomass degradation related clusters found in A: “Pezizomycotina abundant”, for example IPR001138 transcription factor *A. niger* xlnR (van Peij et al. 1998) is known to regulate cellulolytic genes while and *A. flavus* aflR (Woloshuk et al. 1994) regulates a secondary metabolism pathway. Furthermore, Pezizomycotina specific features like ascocarp formation or other yet incompletely described processes could involve these transcription factors.

3.1.4 Browser interface and further use of comparative genomics data

In this work we have concentrated on the analysis of the major trends in our dataset. The data could also be used in more detailed studies of protein families of special interest. In addition, automatic classifications of ORFs are not perfect and detailed sequence analysis is needed to understand the evolution of individual families. We have constructed a browser interface to our dataset that enables researches to profit of the complementarity of clustering and InterPro analysis and retrieve sequences easily (I: Figure 14). It was constructed using open source genome browser GBrowse (Stein et al. 2002) and can be expanded to explore other aspects of fungal genomics data such as transcriptome or other system wide measurements. It could also serve as a basis for analysis of novel sequence data from metagenomics or other targeted sequencing projects.

In order to classify protein clusters functionally, we have analysed the uniformity of InterProScan, ProtFun and Funcat annotation data in our clusters. Although sequence homology does not automatically imply functional similarity, it is practically the only existing basis of annotation for novel genomes. As automatic functional annotations are not perfect, it is likely that uniform annotation data of a multispecies cluster provides a more reliable source

of annotation than the analysis of a single ORF separately. Therefore, the cluster annotations produced could be effectively applied to system wide measurement data analysis such as transcriptome profiling data of any genome studied here.

3.2 Development of transcript profiling methods

3.2.1 Method for PCR fragment probe TRAC

TRAC (TRanscript analysis with the aid of Affinity Capture) is a novel multiplexed method for focused transcript analysis developed at VTT. It has several variations. Oligo TRAC uses short oligonucleotides as probes and it has been successfully applied to the study of *T. reesei* bioreactor cultivations (Rautio et al. 2006a; Rautio et al. 2006b) and the brewing process (Rautio et al. 2007). TRAC has also been adapted to quantification of species in mixed bacterial populations (Satokari et al. 2005; Maukonen et al. 2006). Here we describe a method where PCR fragments instead of oligonucleotides are used as probes (Figure 7).

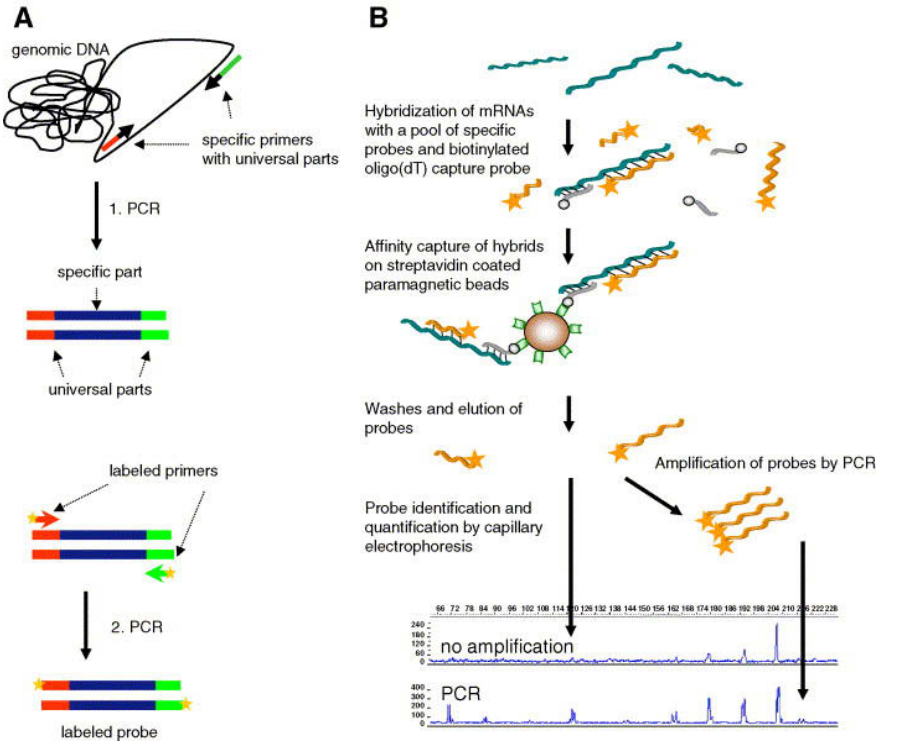


Figure 7. The TRAC technique. ***A:*** *Production of probes by PCR. Specific probe sequences are amplified from genomic DNA by using specific primers with additional universal linker parts. In the following second PCR reaction probes are labelled by using a labelled universal primer pair for all probes.* ***B:*** *The principle of the TRAC technique. A number of different specific probes (different lengths) and biotinylated oligo-dT capture probes are hybridized with the mRNA molecules. The formed hybrids are collected, washed and eluted and optionally amplified by PCR and finally analyzed by capillary electrophoresis. The distinct sizes of the probes identify the corresponding mRNAs in the sample and the signal intensity gives a quantative measure of their amounts.*

We selected 18 genes from *S. cerevisiae* as targets for method development based on an experiment that compares xylose and glucose as a carbon source (Salusjärvi et al., personal communication). Probes and primers to produce them were designed as described in III. In addition to specific parts, universal 5' PCR primer sequences were added to the probe production primers. This enables probe labelling by using labelled primers and optionally a PCR amplification of the probes after wash and elution steps to increase sensitivity. Currently the PCR

reaction used to produce probes from genomic DNA produces low molecular weight amplification products that need to be removed. This seriously limits the extension of the protocol to quantification of thousands of mRNAs.

Initially the quantitative nature of the assay was confirmed by a dilution series of sample mRNA (II: Figure 3a). The amount of input sample mRNA and the signal i.e. fluorescence intensity of the probe peaks, have a linear correlation of $r^2 \geq 0.96$ and approximately hundred fold linear range could be achieved. All probes could not be measured from smaller mRNA samples. Sensitivity can be improved by PCR amplification of probes. The maintenance of signal ratios between probes for different mRNAs was studied in order to confirm that differences in PCR amplification efficiency do not hamper simultaneous quantification of several probes (II: Figure 4). Similar results have been obtained for probe pools of a few dozen probes in cDNA-AFLP (Vos et al. 1995). With 16 cycles of PCR, 17 out of the 18 probes could be detected from 2.5 ng of mRNA. Identical RNA samples were analysed also with qRT-PCR and agreement with TRAC was shown (II: Table 2). Previously, good agreement of Northern hybridisation and oligo TRAC was demonstrated (Rautio et al. 2006a).

The major advantage in the protocol of TRAC to other transcript profiling methods is the lack of cDNA synthesis. cDNA synthesis is a complex and time consuming procedure that undoubtedly adds noise to any transcript measurement. In comparison to microarray methods where the probe is bound to solid surface, the hybridisation in TRAC is performed in solution. Solution hybridisation has been shown to have better reaction kinetics than solid phase hybridisations (Syvanen et al. 1986; Thompson and Gillespie 1987; Jungell-Nortamo et al. 1988). Solution hybridisation also facilitates automation. The whole protocol can be implemented with equipment compatible with standard 96-well plates. Additionally, oligo TRAC has been demonstrated to work with no RNA extraction step directly from cell lysate (Rautio et al. 2006a). PCR fragment TRAC offers other advantages. Optional PCR amplification adds sensitivity, without compromising quantification. In comparison to oligo TRAC a wider range of probe lengths can be used. This allows larger pool sizes. The maximum length of commercially available oligos limits the probe pool size in practise to about 15 probes (Jari Rautio, personal communication). In PCR fragment TRAC with current equipment at least 75 probes of different lengths can be fitted into one pool.

3.2.2 Computational optimisation of probe selection and pooling for TRAC

An efficient computational method to find and pool good probes for TRAC is a prerequisite for its efficient use. A good probe should be sensitive and specific and neither property can be explicitly predicted from sequence. However, by combining existing software, defining cut-offs based on experimental work and developing an algorithm for pooling, a satisfying solution can be found.

For PCR fragment TRAC the probe selection should fill the following criteria: probes must be unique enough to bind only one transcript, probes should be flanked by suitable primer sites, probe sizes must be suitable for detection with for example a standard capillary electrophoresis device (i.e. DNA sequencer) and probes should be divided to a minimal set of unique pools. The more pools are needed in the analysis of samples, the more laboratory work needs to be done.

In order to find specific probes the genome sequence of the organism studied must be available and coding sequences predicted. Studies on probe hybridisation between two nucleotide fragments suggest that if their overall sequence indentity is below 80 percents and there are no exact complementary subsequences of more than 15 bases their hybridisation is small (for review see (Kivioja 2004)). To find candidate probes for each ORF, all exact repeats of over 15 bases were filtered out with the program REPuter (Kurtz et al. 2000; Kurtz et al. 2001) from the combined ORFs. For each unfiltered area of an ORF the Primer3 program (Rozen and Skaletsky 2000) is run to find candidate probes that consist of the primer pair and the intervening fragment. For computational efficiency the number of candidates is limited to a suitable level, in the example below 20 candidates were selected. For each primer pair of a candidate probe it is subsequently checked that no alternative binding sites less than 5000 bp apart exist. An alternative site is defined as a site with more than 80 percents identity with the primer. They are searched with Myers' bit-parallel algorithm (Myers 1999). Finally, for each remaining candidate probe Myers' bit-parallel algorithm was used to check that no alternative target sequence with more than 80 percents identity exists. To pool the candidate probes, a polynomial time approximation algorithm was developed (III). The algorithm divides the probes to at most twice as many pools as compared to the optimal case.

To test the method we attempted to search probes for all predicted ORFs of *S. cerevisiae*. Probes for 87 percents of the 6132 ORFs could be found. Filtering of non-unique sequences lost 4 percents of all ORFs and 3 percents were lost in the primer selection phase. The final similarity checking phases lost 5 percents. Further analysis showed that theoretically often more than 80 probes could be fitted to a pool. Hence, 65 pools were found to be sufficient for the analysis of the 5334 ORFs (Kivioja 2004).

Recently, the software Tracfinder, the development of which was initiated in III, has been improved by two new features. These are the checking for GC percent and melting temperatures to find probes of uniform hybridisation properties and the RNA secondary structure predictions to discard probes likely to be affected by them. Also, probes for oligo TRAC can be searched and pooled by Tracfinder (Kivioja 2004). The software has been used for oligo TRAC probe design and pooling for studies of *Fusarium graminearum* (Jari Rautio, unpublished results), *Hordeum vulgare* (Annika Wilhelmson, unpublished results), *Pichia pastoris* ((Gasser et al. 2007)), *Saccharomyces pastorianus* (Jari Rautio, submitted), *S. cerevisiae* (Marilyn Wiebe, submitted) and *Trichoderma reesei* (Rautio et al. 2006a; Rautio et al. 2006b). Also, PCR fragment probes have been designed for *Homo sapiens* (Jari Rautio, unpublished results).

Given the advantages of TRAC, also oligo TRAC could be considered for genome wide transcriptional profiling. In oligo TRAC, no primer selection step is required. Consequently, based on the previous *S. cerevisiae* example, probes for at least 90 percents of the ORFs can be found. With 15 probes per pool, almost 400 pools would be needed. However, with the use of several fluorescent dyes, longer oligonucleotide probes and improvement of the performance of the detection device, the number of pools could be reduced drastically.

3.2.3 Computational optimisation of cDNA-AFLP experiments

cDNA-AFLP is one of the most attractive methods for transcriptome analysis of non-sequenced organisms (Figure 2). It is also a complex and tedious method and it is essential to try to decrease its workload. Successful adaptation of cDNA-AFLP for novel organisms depends on proper experimental design i.e. the selection of restriction enzymes and selective nucleotides. The coverage of

cDNA-AFLP can be defined as the percent of organism's ORFs that can be analysed i.e. covered with the selected design. In this study (IV) it was shown how cDNA-AFLP performance can be improved if some sequence data is available and that even sequence data from a fairly distantly related organism can be used to make cDNA-AFLP more efficient.

Previously, cDNA-AFLP has been simulated in a setting with one restriction enzyme pair and a set of input sequences (Qin et al. 2001; Rombauts et al. 2003; Bikandi et al. 2004). The coverage of an enzyme pair has been estimated by counting the number of fragments produced by the pair that lie in the length range of electrophoresis. The coverage of several enzyme pairs has been the union of individual pairs. The number of selective nucleotides has not been analysed, and subsequently the overlap of fragments of similar size has not been dealt with.

The selection of enzyme pairs and the number of selective nucleotides were treated in this study as one optimisation problem. Electrophoresis size separation was also taken into account and only fragments that are sufficiently separated from other fragments were consider as informative i.e. as covered ORFs. This allowed us to consider all combinations of enzyme pairs and subsets of selective nucleotides as possible experimental designs. We developed a heuristic algorithm based on an optimal algorithm for a case where only one enzyme was used (IV).

The algorithm, called Greedy, was tested with genomes of *S. cerevisiae*, *A. thaliana and H. sapiens*. Its performance was compared to a Fixed design. In Fixed design we selected the enzyme pairs that individually had the highest coverage with a given number of selective nucleotides and all possible combinations of selective nucleotides used. This is the way cDNA-AFLP experiments have been designed (for example, (Breyne et al. 2003; Fukumura et al. 2003)). However, we took into consideration the problem of similar sized fragments. The Greedy algorithm tries to select from all possible combinations of enzyme pairs and selective nucleotides, the combination with the highest coverage i.e. all possible combinations of selective nucleotides are not used. Because in practice the introduction of a new enzyme pair requires extra laboratory work, the maximum number of enzyme pairs was limited to four. As the enzymes of a pair are used in sequence (Figure 2), we use them in both orders as also has been done in the previous studies (Breyne et al. 2003; Fukumura et al. 2003).

The Greedy algorithm reaches the same coverage as the Fixed design with fewer pools, i.e. less laboratory work. For example when comparing a fixed design of 4 enzyme pairs and 512 pools with the greedy design, it reaches the same coverage with the following reductions in the number of pools: *S. cerevisiae* 48 percents, *A. thaliana* 27 percents and *H. sapiens* 36 percents (IV: Table 2). As cDNA-AFLP is a method primarily for non-sequenced genomes we compared the coverage of a greedy design to its coverage in another species. For example the coverages of *A. nidulans* with designs from different organisms are: *A. nidulans* 83 percents, *N. crassa* 81 percents, *S. cerevisiae* 76 percents and *H. sapiens* 74 percents. Eurotiomycetes and Sordariomycetes have been estimated to have diverged some 700 millions of years ago (Hedges 2002). Even inside Eurotiomycetes amino acid identity between orthologues has been estimated to be slightly less than 70 percents (Galagan et al. 2005). The use of hybridisation based transcriptome profiling techniques, for example microarrays, to profile another organism would not be possible across such evolutionary distances. Still a Sordariomycetes genome could serve as a base for designing cDNA-AFLP for Eurotiomycetes. As increasing numbers of genome sequences become available, finding a suitable genome for designing cDNA-AFLP for new organisms, unlikely to be sequenced in the near future, should always be possible.

3.2.4 Method for cDNA-subtraction library analysis

cDNA subtraction libraries have been used in the analysis of Unfolded Protein Response (UPR) of filamentous fungi (Watson et al. 2000; MacKenzie et al. 2005). We have applied it to study the effects of DTT and production of tPA (human tissue plasminogen activator) in T. reesei. Pre-experiments of a transformant strain producing tPA revealed that UPR is induced only mildly. Transcription of the UPR marker bip1 was induced 2-fold and pdi1 4-fold as compared to a non-transformed strain. The previously used protocol cannot detect less than 5-fold inductions of transcription according to authors (Watson et al. 2000). Therefore we combined two protocols in hope of achieving better sensitivity (Ausubel et al. 1994; Watson et al. 2000). The major difference to the studies of (Watson et al. 2000) was the addition of a digestion step of cDNAs before adaptor ligation to achieve a cDNA pool of a more uniform size distribution (V) as suggested by (Ausubel et al. 1994). Consequently, our cDNA subtraction libraries contained transcripts which were, according to Northern hybridisation, less

than 5-fold induced. However, previously a false positive rate of only 5 percents was reported (Watson et al. 2000). We found that according to Northern hybridisations, 12 of the 23 randomly selected library clones were at least two-fold induced. Therefore improved sensitivity was likely achieved by reducing specificity.

3.3 Transcriptome profiling of protein secretion stress

Profound understanding of secretion stress responses is essential for attempts to develop improved protein production organisms. As Pezizomycotina species secrete a wide variety of proteins in comparison to Saccharomycotina species, the responses to secretion stress in these fungal subphyla could have diverged. However, little work to systematically elucidate these differences has been done. Transcriptomic effects of secretion stress in *S. cerevisiae* have been analysed in several publications (Casagrande et al. 2000; Gasch et al. 2000; Hughes et al. 2000; Travers et al. 2000; Patil et al. 2004). In order to compare the responses of Pezizomycotina and Saccharomycotina we first combined the previously published data from *S. cerevisiae* to create a consensus model of its response. We then analysed the secretion stress responses of *T. reesei* with cDNA subtraction libraries and cDNA-AFLP and compared the results to the *S. cerevisiae* response. The genome of *T. reesei* was not sequenced at the time the experimental work was carried out. However, access to a pre-publication version of the genome greatly improved our data analysis.

3.3.1 Computational reanalysis of published *S. cerevisiae* transcriptome profiling data

S. cerevisiae has served as a major model organism in studies on protein secretion and on factors affecting the process. In addition to transcriptome profiling data of responses to secretion stress there is data on responses to a large variety of other types of stress conditions. We searched for genes induced or repressed in *IRE1* and *HAC1* dependent or independent manner from a dataset combined from literature (V). The dataset includes results from *S. cerevisiae* cultures treated with DTT or tunicamycin to inhibit protein folding and transport (Casagrande et al. 2000; Gasch et al. 2000; Hughes et al. 2000; Travers et al. 2000; Patil et al. 2004) and cultures of *S. cerevisiae* producing a secreted mouse

histocompatibility protein (Casagrande et al. 2000). Consequently, we used a larger and more varied data set in comparison to previous studies on transcriptomic effects of secretion stress. This enabled us to exclude effects specific to the inducer of stress (e.g. DTT, tunicamycin or heterologous protein), the specific amount of chemical or the strain used. Our computational reanalysis provided a robust dissection of transcriptional effects of secretion stress in *S. cerevisiae*, in order to compare it with our results of transcriptional effects of secretion stress in *T. reesei*.

The model used in the re-analysis divides genes to eight specific groups showing significant differential expression under the secretion stress conditions. Of these five were considered relevant and analysed further. Namely, the model selected a group of 46 strongly responding UPR-dependent genes ("coreUPR"), that were mainly up-regulated under the secretion stress conditions and not as a response to other type of stresses, and another group of 135 UPR dependent up regulated genes showing less stringent response to secretion stress (V: Figure 1A and C). Both groups, but especially the "coreUPR" group, had a higher content of secretion related genes as compared to the relative amount of secretion related genes in the genome. This is in accordance with previous information on the functional categories of UPR up regulated genes (Travers et al. 2000; Patil and Walter 2001; Kaufman 2002) (V: Table 1).

Previously, a group of 381 genes was selected as UPR (i.e. *HAC1* and *IREI*) dependently up regulated (V: "Travers", Figure 1B) (Travers et al. 2000). The "Travers" group was selected based on significant correlation to a group of UPR model genes. These genes react strongly also to other types of stress than UPR (V: "Travers", Figure 1B, shown in yellow). The "coreUPR" shows less response to other types of stresses than the "Travers" group, but interestingly, both show a modest response in diamide treatment data (Gasch et al. 2000) (V: Figure 1A and B). Diamide is a strong oxidasing agent, which acts particularly on thiol groups, for example oxidasing GSH to GSSG. The transcriptional response to diamide has been described as an average of responses to heat shock, H_2O_2 and menadione and DTT (Gasch et al. 2000). Of the "Travers" group only 86 genes were found to be UPR dependently induced in our reanalysis (35 were classified into "coreUPR" group, 51 in "UPR Dependent Up" group). In accordance, UPR promoter elements have only been found from 137 of the 381 "Travers" genes (Patil et al. 2004).

155 genes showed UPR dependence and were down regulated ("UPR Dependent Down", Fig. 1D). Some of the genes in the groups "UPR Dependent Up" and "UPR Dependent Down" genes showed responses also in other stress conditions, but these were on average much milder than the responses of UPR independent groups.

The analysis revealed also a large number of genes that were up regulated (592 genes, "UPR Independent Up") or down regulated (603 genes, "UPR Independent Down") under secretion stress conditions, but not in a UPR dependent manner that would require functional *IRE1* and *HAC1* genes. 60 genes of the "Travers" set were assigned to "UPR Independent Up". The genes in these groups were responsive also to other type of stress conditions (V: Figure 1E and 1F). In comparison to UPR dependent groups, responses of UPR independent groups to other stresses were mostly to opposite directions and on average more pronounced. For example, the median of "UPR Dependent Down" rises slightly in an experiment where *S. cerevisiae* was grown in stationary phase for 5 days. ("YPD 30C 5d", (Gasch et al. 2000)), while in "UPR Independent Down" it drops two fold. Consequently, the independent groups are likely to be regulated through general stress pathways. In UPR independently and dependently regulated groups, the up-regulated have a higher than genomic percent of protein secretion related genes while the down-regulated have a lower percent of them than the genomic average (V: Table 1).

The transcriptome profiling data from studies of *S. cerevisiae* secretion stress (Casagrande et al. 2000; Gasch et al. 2000; Hughes et al. 2000; Travers et al. 2000; Patil et al. 2004) and our computational analysis (V: Table 1) show that only a few genes of secreted proteins are down regulated under secretion stress in *S. cerevisiae*. As the transcriptomic responses of *A. nidulans* (Sims et al. 2005) and in *A. thaliana* (Martinez and Chrispeels 2003) appear to affect larger amounts of genes encoding secreted proteins, it is very likely that *S. cerevisiae* does not have the RESS response.

3.3.2 Analysis of the *T. reesei's* secretion stress EST collection

In order to identify genes differentially expressed in response to secretion stress in the *T. reesei* strain Rut-C30, a transformant expressing tPA was grown in

chemostat and Rut-C30 was treated with DTT in a batch bioreactor culture. Samples of the tPA producing transformant, DTT treatment and their respective untransformed and untreated references were analysed with cDNA-subtraction cloning and cDNA-AFLP.

DNA-subtraction libraries were prepared with consecutive hybridisations of sample and reference cDNA. Based on quality analysis of the libraries four of them were selected for sequencing. These were libraries derived from genes putatively up-regulated after 1h of DTT treatment or in the tPA producing transformant with respect to their references, and after one or five cycles of hybridisation. cDNA-AFLP gel electrophoresis bands that showed a clear up-regulation after 1h of DTT treatment and in the tPA producing transformant with respect to their references were selected for sequencing.

In total 2144 ESTs were sequenced. They were mapped to the prepublication genome sequence of *T. reesei* and found to correspond to 457 ORFs. In addition, 71 unique ESTs did not match to the genome. Based on the redundancy of the ESTs matching to the genome, these could correspond to roughly 20 ORFs. The ORFs with most corresponding ESTs included previously characterised UPR related genes *bip1* and *pdi1* (Pakula et al. 2003; Saloheimo et al. 2003), as well as ORFs assigned to UPR related functions in other organisms (V: Table 2 and Table 4). Also transcripts expected to be very abundant, such as those encoding translation elongation factor *tef1* (Nakari et al. 1993), actin *act1* (Velculescu et al. 1997), *hex1* (Bergquist et al. 2004), cellobiohydrolases (Penttila and Limon 2004) and ribosomal genes (Velculescu et al. 1997) were represented in the libraries. These are likely to be false positives. *T. reesei* ORFs corresponding to all sequenced ESTs were clearly enriched in secretion related genes as compared to the genome content of secretion related genes of *S. cerevisiae* (V: "Homologues of *T. reesei* ER stress ORFs" Table 1). However, only 10 percents of these ORFs belonged also to the "Travers" group and the overlap with our *S. cerevisiae* secretion stress regulated gene groups was even less.

Table 4. T. reesei ORFs from secretion stress EST library with homology to S. cerevisiae genes directly involved in protein secretion. Related processes such as protein degradation, cell wall biogenesis, vacuolar transport and translation were excluded. ESTs were mapped to T. reesei genome and found ORFs were matched to S. cerevisiae ORFs by blast (Altschul et al. 1990). *Gene Ontology (GO) annotations were retrieved from Saccharomyces Genome Database* (Dolinski et al. 2004).

T. reesei genome version 2.0 ORF identifier	Closest homologue in *S. cerevisiae* genome	UPR regulation verified by Northern	GO process	GO molecular function
tre120412	*age2*		ER to Golgi transport	ARF GTPase activator
tre55415	*gea2*		ER to Golgi transport	ARF guanyl-nucleotide exchange factor
tre80553	*arf1*		ER to Golgi transport	ARF small monomeric GTPase
tre119578	*erv29*	YES	ER to Golgi transport	molecular_function unknown
tre46639	*yip3*		ER to Golgi transport	molecular_function unknown
tre76172	*erv46*		ER to Golgi transport	molecular_function unknown
tre80898	*ypt1*		ER to Golgi transport	RAB small monomeric GTPase
tre45604	*arf2*		ER to Golgi transport	small monomeric GTPase
tre121223	*sec4*	YES	Golgi to plasma membrane transport	GTPase
tre123577	*ero1*	YES	protein folding	electron carrier
tre121997	*cpr3*		protein folding	peptidyl-prolyl cis-trans isomerase
tre121415	*pdi1*		protein folding	protein disulfide isomerase
tre35465	*lhs1*		protein transport	chaperone
tre121068	*sec53*	YES	protein-ER targeting	phosphomannomutase
tre120051	*srp68*		protein-ER targeting	signal sequence binding
tre119592	*pmr1*		secretory pathway	calcium-transporting ATPase
tre74462	*spc3*		signalpeptide processing	signal peptidase
tre121397	*sec61*		SRP-dependent cotranslational membrane targeting, translocation	protein transporter
tre122920	*kar2 / BIP*	YES	SRP-dependent cotranslational membrane targeting, translocation	unfolded protein binding
tre22560	*sil1*		SRP-dependent cotranslational membrane targeting, translocation	adenyl-nucleotide exchange factor
tre121754	*sec63*	YES	SRP-dependent cotranslational membrane targeting, translocation	protein transporter
tre68492	*YPL145c*		vesicle-mediated transport	oxysterol binding
tre60222	*sec16*		vesicle-mediated transport	structural molecule
tre120044	*gdi1*		vesicle-mediated transport	GDP-dissociation inhibitor
tre55318	*apl3*		vesicle-mediated transport	molecular_function unknown

Northern analysis was done to confirm the UPR-like regulation of genes found from the EST library. In addition to the samples described above a sample from a shake-flask cultivation of a transformant over-expressing *ire1* and its reference were included. Over-expression of *ire1* in *T. reesei* leads to a constitutive UPR (Valkonen et al. 2004). The responses induced by production of a heterologous protein and by the DTT treatment might also have features specific to each of

these two conditions and therefore we have focused on genes under differential expression in all culture conditions. A gene was classified as a UPR regulated gene if it was confirmed by Northern analysis to be up regulated after one-hour-treatment with DTT, in a tPA producing strain as well as in a strain over-expressing *ire1*. Northern analysis of a set of genes selected from the EST collection revealed UPR-like up regulation of a group of genes expected to be UPR up-regulated based on data obtained from other organisms, e.g. *bip1*, *pdi1*, *ero1, pmr1* and *sec61* or belonging to functional categories containing UPR regulated genes in other organisms e.g. *sec53*, *sec4*, *ste24*, *erv29* (Travers et al. 2000; Martinez and Chrispeels 2003; Sims et al. 2005) (V: Figure 2). Many of the genes analysed by Northern analysis were induced at a lower level in the tPA producing culture than in the DTT treated culture or in the strain over-expressing *ire1* (Table 4, V: Figure 2). This was reflected also in the abundance of ESTs corresponding to the differentially expressed genes in the different libraries, of which the abundantly expressed foldase/chaperon genes *pdi1* and *bip1* are good examples (V: Table 2). In addition, the analysis revealed induction of genes that have not been described as UPR induced in filamentous fungi or yeast, which are discussed in more detail below.

Based on the annotation of *T. reesei* ORFs represented in the library by homology to *S. cerevisiae* and Northern hybridisations, the combined EST library appears as a good source of secretion stress induced genes. However, the overlap with our or previous (Travers et al. 2000) definition of *S. cerevisiae* UPR regulated genes is surprisingly small. Partly, this might be explained by the use of a very simplistic homology mapping where the best blastp (Altschul et al. 1990) hit of a *T. reesei* ORF was accepted as the homologue in *S. cerevisiae*. However, based on genome analysis, the secretion machinery, as described in *S. cerevisiae*, is largely conserved, with the exception of expansion of some families, in Pezizomycotina species (see chapter 1.2.2). Therefore, even simple homologue detection should work acceptably.

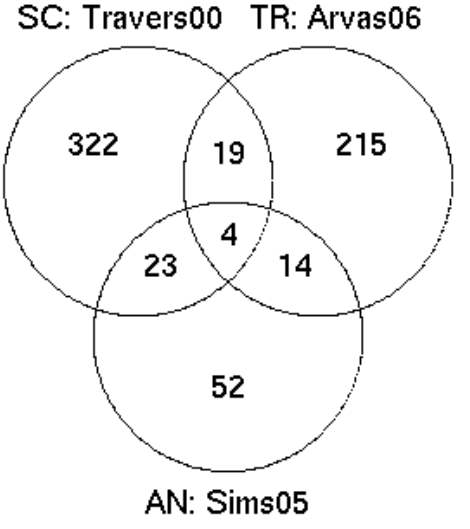


Figure 8. Counts of overlapping ORFs in secretion stress induced groups in three different fungi. Due to variable gene names exact counts vary slightly from original publications. Data from three species and different publications was used: SC (S. cerevisiae) (Travers et al. 2000) *, TR (T. reesei) (V) and AN (A. nidulans)* (Sims et al. 2005)*. S. cerevisiae homologues of T. reesei and A. nidulans ORFs were taken from the source publications.*

We compared secretion stress induced genes in Aspergillus nidulans (Figure 8). Unfortunately, the transcriptome profiling data available for such a comparison comes from very divergent sources. *S. cerevisiae* transcriptome profiling was done with whole genome microarrays, while the *T. reesei* profiling covers the transcriptomic response only partially as described above. *A. nidulans* data was done with cDNA microarrays covering approximately half of the genome. The group of *A. nidulans* ORFs shown in Figure 8 only covers ORFs described by authors to be somehow related to secretion and induced by secretion stress (Sims et al. 2005). Regardless of these lacks, it seems that in different fungi different genes are induced as a response to secretion stress. The responses share two major trends. The function of most of the genes is unknown and while several GO terms are abundant, in each organism different genes with the same term are induced. As expected secretion related terms are among the most abundant.

3.3.3 Nucleosomal gene induction as a response to secretion stress

ESTs corresponding to histone H2A were found from the secretion stress up-regulated EST library (V: Table 2). The UPR-like expression of *T. reesei* H2A and H4, and induction of *A. nidulans* H4 during DTT treatment was shown with Northern hybridisations (V: Figure 2 and 4).

Histones are generally thought to be strictly regulated with cell cycle, active synthesis taking place only during the S-phase (Meshi et al. 2000; Marzluff and Duronio 2002). *S. cerevisiae* transcriptome data collected for this study indicates that if some regulation of nucleosomal genes as a response to stress is taking place, it is primarily down regulation, and no up regulation is seen under secretion stress conditions. The induction of H2A and H4 appears to have no correlation with the growth rate of the cells. The comparison of the expression levels of the genes in the tPA producing strain and in its parental reference strain was carried out in well controlled chemostat cultures with the same specific growth rate. Growth activation could not explain the induction either, since the treatment of the cultures with DTT retards the growth to some extent, and also the *ire1* over expression does not lead to increased growth of the fungus.

In *S. cerevisiae* Hac1p represses expression of early meiotic genes through a mechanism of histone deacetylation and subsequent tightening of chromatin structure (Schroder et al. 2004). In human, XBP-1, the homologue of Hac1p, increases ERα (estrogen receptor) transcription through large scale chromatin unfolding (Fang et al. 2004). Therefore, a link between chromatin structure and UPR is not surprising, although the unexpected uncoupling of histone regulation from cell cycle control is.

3.3.4 *cpc1* induction as a response to secretion stress

ESTs corresponding to the *T. reesei* homologue of *N. crassa* cpc1 (cross pathway control) and *S. cerevisiae* GCN4 were found from the secretion stress up-regulated EST library (V: Table 2). The UPR-like expression of *T. reesei cpc1* and induction of the *A. nidulans* homologue (*cpcA*) during DTT treatment was shown with Northern hybridisations (V: Figure 2 and 4). The CPC proteins in filamentous fungi as well as the *S. cerevisiae* GCN4 have been shown to control amino acid biosynthesis (Paluh et al. 1988; Wanke et al. 1997). A notable difference in activation of the CPC homologues and GCN4 in response to amino acid deprivation is that *S. cerevisiae* GCN4 is mostly regulated on translational level (Albrecht et al. 1998) whereas the CPC-proteins in filamentous fungi are also under strong transcriptional control (Paluh et al. 1988; Wanke et al. 1997). In order to determine whether induction of *cpc1* in *T. reesei* could be due to amino acid starvation, we determined intracellular concentrations of 15 amino acids. No systematic lack of any of these amino acids was detected in any conditions.

The *S. cerevisiae* Gcn4p and the mouse homologue ATF4 have also been shown to be involved in secretion stress (Harding et al. 2000b; Harding et al. 2003; Patil et al. 2004). In *S. cerevisiae*, Gcn4p is required for induction of the majority of UPR induced genes under secretion stress (Patil et al. 2004). However, according to analysis of transcriptome data collected for this study amino acid biosynthesis genes are not induced in *S. cerevisiae* secretion stress (V: Figure 5). Mouse Atf4p appears to induce transcription of genes involved in glutathione biosynthesis under tunicamycin treatment. It has been proposed that the lack of reducing power and the need to up regulate glutathione biosynthesis were the major reasons for ATF4 induction in mouse under secretion stress conditions (Harding et al. 2003).

In order to study the effects of the *cpc1* induction, we carried out a Northern analysis in secretion stress conditions of a set of *T. reesei* genes whose homologues in *S. cerevisiae* have been shown to be Gcn4p targets based either on computational promoter analysis (Fondrat and Kalogeropoulos 1994; Kellis et al. 2003) or microarray experiments (Natarajan et al. 2001). As the role of CPC proteins in regulation of amino acid biosynthesis is very similar in *S. cerevisiae* and in filamentous fungi, it is likely that many of these selected genes are *cpc1* targets also in *T. reesei*. The genes *glt1* (glutamate synthase), *arg1* (arginosuccinate synthase) and *aro1* (pentafunctional arom protein, aromatic aminoacid biosynthesis) were shown by Northern analysis to be UPR induced genes in *T. reesei*. In addition, the gene *asn1* was induced almost 2-fold in the *ire1* over-expressing strain and *cys4* was induced more than 2.5-fold after one hour DTT treatment (V: Figure 2 and 3). In mouse a homologue of *asn1*, asparagine synthase (Asns) and a homologue of *cys3*, cystathione γ-lyase (Cth) which follows *cys4* in homocysteine and cysteine interconversion pathway (Figure 3), show ATF4 dependent regulation in secretion stress (Harding et al. 2003). The glutathione biosynthesis gene *gsh1*, thioredoxin genes (*trx1*, *trx2*) and glutathione reductase (*gsr1*) involved in glutathione related functions do not show up regulation in *T. reesei*, which is also the case for their homologues in mouse (Harding et al. 2003). GLNI that synthesizes glutamine from glutamate instead of glutathione (Figure 3) neither shows any induction. In addition, ESTs homologous to a human microsomal glutathione s-transferase 3 were found from the secretion stress library and UPR-like regulation was confirmed with Northerns (V: Figure 2). In mice, glutathione-s-transferases protect cells from oxidative stress (Chanas et al. 2002). At transcript level our results with *T. reesei*

are in accordance with the observations made in mouse in that not all the genes putatively under ATF4 regulation or involved in glutathione metabolism are affected under secretion stress conditions. Based on the assembled evidence we believe that *T. reesei* is manifesting a response similar to the ATF4 dependent secretion stress induced oxidative stress response of mouse.

4. Conclusions and future perspectives

Comparative genomics of eukaryotic microbes appears as a very fruitful strategy to improve our understanding of their biology and subsequently advance their industrial performance. Our comparative genomics work is one of the first of its kind and gives coarse grained, but interesting insight into Pezizomycotina genomes. Ascomycota probably have several mechanisms to slow down their evolution. However, our results show that at least some of the biomass degradation and secondary metabolism related protein families seem to have expanded recently in Pezizomycotina (I). It is notable that only these protein families seem to have such evolutionary history. Of the clearly diverged ORFs found both in Pezizomycotina and Saccharomycotina, those related to mitochondria emerge as the most prominent. However, the primary metabolism as described in *S. cerevisiae* is largely conserved in all fungi. Apart from the known secondary metabolism, Pezizomycotina have pathways that could link secondary metabolism to primary metabolism and a wealth of undescribed enzymes.

The next natural step in order to profit from and improve the capabilities of fungi is comparative modelling of cellular systems such as metabolism and protein secretion. In depth comparisons of not only topology, but also sequence divergence in context of such models would allow us to better explain phenotype with genotype. In addition, such models would greatly improve analysis of genome wide measurement data. However, they require improved homology detection and network construction algorithms.

With development of novel sequencing techniques such as pyrosequencing (Margulies et al. 2005), at least all currently used laboratory organism can be easily sequenced in the near future. However, there still remain niches for transcriptome profiling techniques for non-sequenced organisms. For example profiling of mixed natural populations, such as composts and gut flora could be envisaged. We have shown that only general knowledge of the subphyla under study would be required to select suitable genome sequences for efficient design of cDNA-AFLP experiments (IV). It has been proposed that cDNA-AFLP could be used as a way to produce PCR fragment TRAC probes (Kivioja 2004). The TRAC protocol is simpler and easier to automate than cDNA-AFLP. We have shown that TRAC can be done with PCR fragments instead of oligonucleotide

probes (II) and theoretically the PCR fragment TRAC can be used to cover whole transcriptomes (III). TRAC has also been shown to be sensitive and accurate. A cDNA-AFLP probe TRAC could allow very efficient profiling of complex microbial populations which have dynamic changes in relative proportions of species.

Saccharomycotina and Pezizomycotina have intriguingly similar genome content of protein secretion machinery related ORFs, given their difference in variety and volume of secreted proteins. To understand what kind of secretion system is required for a Pezizomycotina type secretion, instead of the currently well described Saccharomycotina type secretion, functional genomics studies are needed. We carried out one of the first transcriptome profiling studies of response to secretion stress in a Pezizomycotina (V). Our results suggest that although secretion stress responses regulate mostly secretion machinery related genes both in the Saccharomycotina S. cerevisiae and in the Pezizomycotina *T. reesei,* the actual genes regulated are different. The most notable differences are the up-regulation of the transcription factor *cpc1* and histone genes in response to secretion stress. Are these differences just noise due to mutations in the transcription regulation or is the secretion machinery regulated in some fundamentally different way for a fundamentally different function? The availability of genome sequences allows transcriptome profiling studies of much higher quality to refine these hypotheses. However, other high throughput techniques such as genome wide protein localisation and protein interaction studies are urgently needed.

References

Albrecht, G., Mosch, H.U., Hoffmann, B., Reusser, U. & Braus, G.H. 1998, "Monitoring the Gcn4 protein-mediated response in the yeast *Saccharomyces cerevisiae*", *J Biol Chem,* Vol. 273, No. 21, pp. 12696–12702.

Alexander, N.J., McCormick, S.P. & Hohn, T.M. 1999, "TRI12, a trichothecene efflux pump from Fusarium sporotrichioides: gene isolation and expression in yeast", *Molecular & general genetics: MGG,* Vol. 261, No. 6, pp. 977–984.

Al-Sheikh, H., Watson, A.J., Lacey, G.A., Punt, P.J., MacKenzie, D.A., Jeenes, D.J., Pakula, T., Penttila, M., Alcocer, M.J. & Archer, D.B. 2004, "Endoplasmic reticulum stress leads to the selective transcriptional downregulation of the glucoamylase gene in Aspergillus niger", *Mol Microbiol,* Vol. 53, No. 6, pp. 1731–1742.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990, "Basic local alignment search tool", *Journal of Molecular Biology,* Vol. 215, No. 3, pp. 403–410.

Arabidopsis Genome Initiative 2000, "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*", *Nature,* Vol. 408, No. 6814, pp. 796–815.

Arnaud, M.B., Costanzo, M.C., Skrzypek, M.S., Binkley, G., Lane, C., Miyasato, S.R. & Sherlock, G. 2005, "The Candida Genome Database (CGD), a community resource for *Candida albicans* gene and protein information", *Nucleic acids research,* Vol. 33, No. Database issue, pp. D358–63.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J. & Struhl, K. 1994, *Current protocols in Molecular Biology,* John Wiley & Sons, Inc.

Averous, J., Bruhat, A., Jousse, C., Carraro, V., Thiel, G. & Fafournoux, P. 2004, "Induction of CHOP expression by amino acid limitation requires both ATF4 expression and ATF2 phosphorylation", *J Biol Chem,* Vol. 279, No. 7, pp. 5288–5297.

Bachem, C.W., van der Hoeven, R.S., de Bruijn, S.M., Vreugdenhil, D., Zabeau, M. & Visser, R.G. 1996, "Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development", *The Plant Journal: for cell and molecular biology,* Vol. 9, No. 5, pp. 745–753.

Bergquist, P.L., Te'o, V.S., Gibbs, M.D., Curach, N.C. & Nevalainen, K.M. 2004, "Recombinant enzymes from thermophilic micro-organisms expressed in fungal hosts", *Biochem Soc Trans,* Vol. 32, No. Pt 2, pp. 293–297.

Berka, R.M., Kodama, K.H., Rey, M.W., Wilson, L.J. & Ward, M. 1991, "The development of *Aspergillus niger* var. *awamori* as a host for the expression and secretion of heterologous gene products", *Biochemical Society transactions,* Vol. 19, No. 3, pp. 681–685.

Bertolotti, A., Zhang, Y., Hendershot, L.M., Harding, H.P. & Ron, D. 2000, "Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response", *Nat Cell Biol,* Vol. 2, No. 6, pp. 326–332.

Bikandi, J., San Millan, R., Rementeria, A. & Garaizar, J. 2004, "In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction", *Bioinformatics (Oxford, England),* Vol. 20, No. 5, pp. 798–799.

Birney, E., Andrews, D., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cox, T., Cunningham, F., Curwen, V., Cutts, T., Down, T., Durbin, R., Fernandez-Suarez, X.M., Flicek, P., Graf, S., Hammond, M., Herrero, J., Howe, K., Iyer, V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Kokocinski, F., Kulesha, E., London, D., Longden, I., Melsopp, C., Meidl, P., Overduin, B., Parker, A., Proctor, G., Prlic, A., Rae, M., Rios, D., Redmond, S., Schuster, M., Sealy, I., Searle, S., Severin, J., Slater, G., Smedley, D., Smith, J., Stabenau, A., Stalker, J., Trevanion, S., Ureta-Vidal, A., Vogel, J., White, S., Woodwark, C. & Hubbard, T.J. 2006, "Ensembl 2006", *Nucleic acids research,* Vol. 34, No. Database issue, pp. D556–61.

Bohnert, H.U., Fudal, I., Dioh, W., Tharreau, D., Notteghem, J.L. & Lebrun, M.H. 2004, "A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice", *The Plant Cell,* Vol. 16, No. 9, pp. 2499–2513.

Bok, J.W., Balajee, S.A., Marr, K.A., Andes, D., Nielsen, K.F., Frisvad, J.C. & Keller, N.P. 2005, "LaeA, a regulator of morphogenetic fungal virulence factors", *Eukaryot Cell,* Vol. 4, No. 9, pp. 1574–1582.

Bok, J.W. & Keller, N.P. 2004, "LaeA, a regulator of secondary metabolism in *Aspergillus* spp", *Eukaryot Cell,* Vol. 3, No. 2, pp. 527–535.

Borkovich, K.A., Alex, L.A., Yarden, O., Freitag, M., Turner, G.E., Read, N.D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N., Plamann, M., Goodrich-Tanrikulu, M., Schulte, U., Mannhaupt, G., Nargang, F.E., Radford, A., Selitrennikoff, C., Galagan, J.E., Dunlap, J.C., Loros, J.J., Catcheside, D., Inoue, H., Aramayo, R., Polymenis, M., Selker, E.U., Sachs, M.S., Marzluf, G.A., Paulsen, I., Davis, R., Ebbole, D.J., Zelter, A., Kalkman, E.R., O'Rourke, R., Bowring, F., Yeadon, J., Ishii, C., Suzuki, K., Sakai, W. & Pratt, R. 2004, "Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism", *Microbiol Mol Biol Rev,* Vol. 68, No. 1, pp. 1–108, table of contents.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S.R., Moon, K., Burcham, T., Pallas, M., DuBridge, R.B., Kirchner, J., Fearon, K., Mao, J. & Corcoran, K. 2000, "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays", *Nature biotechnology,* Vol. 18, No. 6, pp. 630–634.

Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inze, D. & Zabeau, M. 2003, "Quantitative cDNA-AFLP analysis for genome-wide expression studies", *Mol Genet Genomics,* Vol. 269, No. 2, pp. 173–179.

Breyne, P. & Zabeau, M. 2001, "Genome-wide expression analysis of plant cell cycle modulated genes", *Current opinion in plant biology,* Vol. 4, No. 2, pp. 136–142.

Callard, D., Lescure, B. & Mazzolini, L. 1994, "A method for the elimination of false positives generated by the mRNA differential display technique", *BioTechniques,* Vol. 16, No. 6, pp. 1096–1097, 1100–1103.

Casagrande, R., Stern, P., Diehn, M., Shamu, C., Osario, M., Zuniga, M., Brown, P.O. & Ploegh, H. 2000, "Degradation of Proteins from the ER of *S. cerevisiae* Requires an Intact Unfolded Protein Response Pathway", *Molecular cell,* Vol. 5, No. 4, pp. 729–735.

Chakravarthi, S., Jessop, C.E. & Bulleid, N.J. 2006, "The role of glutathione in disulphide bond formation and endoplasmic-reticulum-generated oxidative stress", *EMBO reports,* Vol. 7, No. 3, pp. 271–275.

Chambergo, F.S., Bonaccorsi, E.D., Ferreira, A.J., Ramos, A.S., Ferreira Junior, J.R., Abrahao-Neto, J., Farah, J.P. & El-Dorry, H. 2002, "Elucidation of the metabolic fate of glucose in the filamentous fungus *Trichoderma reesei* using expressed sequence tag (EST) analysis and cDNA microarrays", *J Biol Chem,* Vol. 277, No. 16, pp. 13983–13988.

Chanas, S.A., Jiang, Q., McMahon, M., McWalter, G.K., McLellan, L.I., Elcombe, C.R., Henderson, C.J., Wolf, C.R., Moffat, G.J., Itoh, K., Yamamoto, M. & Hayes, J.D. 2002, "Loss of the Nrf2 transcription factor causes a marked reduction in constitutive and inducible expression of the glutathione S-transferase Gsta1, Gsta2, Gstm1, Gstm2, Gstm3 and Gstm4 genes in the livers of male and female mice", *Biochem J,* Vol. 365, No. Pt 2, pp. 405–416.

Cherkasova, V.A. & Hinnebusch, A.G. 2003, "Translational control by TOR and TAP42 through dephosphorylation of eIF2alpha kinase GCN2", *Genes Dev,* Vol. 17, No. 7, pp. 859–872.

Collen, A., Saloheimo, M., Bailey, M., Penttila, M. & Pakula, T.M. 2005, "Protein production and induction of the unfolded protein response in *Trichoderma reesei* strain Rut-C30 and its transformant expressing endoglucanase I with a hydrophobic tag", *Biotechnol Bioeng,* Vol. 89, No. 3, pp. 335–344.

Cornell, M., Alam, I., Soanes, D.M., Wong, H.M., Rattray, M., Hubbard, S., Talbot, N.J., Lings, B., Hoyle, D., Oliver, S.G. & Paton, N.W. "*e-Fungi*", 2007. Available: http://www.e-fungi.org.uk/index.html.

Correa, A., Lewis, Z.A., Greene, A.V., March, I.J., Gomer, R.H. & Bell-Pedersen, D. 2003, "Multiple oscillators regulate circadian gene expression in *Neurospora*", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 100, No. 23, pp. 13597–13602.

Coutinho, P.M. & Henrissat, B. 1999, "Carbohydrate-active enzymes: an integrated database approach" in *Recent Advances in Carbohydrate Bioengineering*, eds. H.J. Gilbert, G. Davies, B. Henrissat & B. Svensson, The Royal Society of Chemistry, Cambridge, pp. 3–12.

Coutinho, P.M., Stam, M., Blanc, E. & Henrissat, B. 2003, "Why are there so many carbohydrate-active enzyme-related genes in plants?", *Trends in plant science,* Vol. 8, No. 12, pp. 563–565.

Cox, J.S. & Walter, P. 1996, "A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response", *Cell,* Vol. 87, No. 3, pp. 391–404.

Credle, J.J., Finer-Moore, J.S., Papa, F.R., Stroud, R.M. & Walter, P. 2005, "On the mechanism of sensing unfolded protein in the endoplasmic reticulum", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 102, No. 52, pp. 18773–18784.

Daran-Lapujade, P., Daran, J.M., Kotter, P., Petit, T., Piper, M.D. & Pronk, J.T. 2003, "Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays", *FEMS yeast research,* Vol. 4, No. 3, pp. 259–269.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W. & Steinmetz, L.M. 2006, "A high-resolution map of transcription in the yeast genome", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 103, No. 14, pp. 5320–5325.

de Graffenried, C.L. & Bertozzi, C.R. 2004, "The roles of enzyme localisation and complex formation in glycan assembly within the Golgi apparatus", *Current opinion in cell biology,* Vol. 16, No. 4, pp. 356–363.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H., Read, N.D., Lee, Y.H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E. & Birren, B.W. 2005, "The genome sequence of the rice blast fungus *Magnaporthe grisea*", *Nature,* Vol. 434, No. 7036, pp. 980–986.

Deloche, O., de la Cruz, J., Kressler, D., Doere, M. & Linder, P. 2004, "A Membrane Transport Defect Leads to a Rapid Attenuation of Translation Initiation in *Saccharomyces cerevisiae*", *Molecular cell,* Vol. 13, No. 3, pp. 357–366.

DeRisi, J.L., Iyer, V.R. & Brown, P.O. 1997, "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science (New York, N.Y.),* Vol. 278, No. 5338, pp. 680–686.

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. & Siebert, P.D. 1996, "Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 93, No. 12, pp. 6025–6030.

Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R.A., Flavier, A., Gaffney, T.D. & Philippsen, P. 2004, "The *Ashbya gossypii* genome as a tool for mapping the ancient Saccharomyces cerevisiae genome", *Science,* Vol. 304, No. 5668, pp. 304–307.

Diezmann, S., Cox, C.J., Schonian, G., Vilgalys, R.J. & Mitchell, T.G. 2004, "Phylogeny and evolution of medical species of *Candida* and related taxa: a multigenic analysis", *Journal of clinical microbiology,* Vol. 42, No. 12, pp. 5624–5635.

Dolinski, K., Balakrishnan, R.: Christie, K. R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R., Oughtred, R., Theesfeld, C.L., Binkley, G., Lane, C., Schroeder, M., Sethuraman, A., Dong, S., Weng, S., Miyasato, S., Andrada, R., Botstein, D. & and Cherry, J.M. 2004, *Saccharomyces Genome Database.* Available: http://www.yeastgenome.org/.

Dujon, B. 2006, "Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution", *Trends in genetics: TIG,* Vol. 22, No. 7, pp. 375–387.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P. & Souciet, J.L. 2004, "Genome evolution in yeasts", *Nature,* Vol. 430, No. 6995, pp. 35–44.

Ellgaard, L. & Helenius, A. 2003, "Quality control in the endoplasmic reticulum", *Nat Rev Mol Cell Biol,* Vol. 4, No. 3, pp. 181–191.

Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. 2000, "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence", *Journal of Molecular Biology,* Vol. 300, No. 4, pp. 1005–1016.

Enright, A.J., Van Dongen, S. & Ouzounis, C.A. 2002, "An efficient algorithm for large-scale detection of protein families", *Nucleic Acids Res,* Vol. 30, No. 7, pp. 1575–1584.

Evans, S.J., Datson, N.A., Kabbaj, M., Thompson, R.C., Vreugdenhil, E., De Kloet, E.R., Watson, S.J. & Akil, H. 2002, "Evaluation of Affymetrix Gene Chip sensitivity in rat hippocampal tissue using SAGE analysis. Serial Analysis of Gene Expression", *The European journal of neuroscience,* Vol. 16, No. 3, pp. 409–413.

Fang, Y., Yan, J., Ding, L., Liu, Y., Zhu, J., Huang, C., Zhao, H., Lu, Q., Zhang, X., Yang, X. & Ye, Q. 2004, "XBP-1 increases ERalpha transcriptional activity through regulation of large-scale chromatin unfolding", *Biochemical and biophysical research communications,* Vol. 323, No. 1, pp. 269–274.

Fondrat, C. & Kalogeropoulos, A. 1994, "Approaching the function of new genes by detection of their potential upstream activation sequences in *Saccharomyces cerevisiae*: application to chromosome III", *Curr Genet,* Vol. 25, No. 5, pp. 396–406.

Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, J. & Ward, M. 2003, "Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*", *J Biol Chem,* Vol. 278, No. 34, pp. 31988–31997.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E. & Scherf, A. 2000, "Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*", *Nature,* Vol. 407, No. 6807, pp. 1018–1022.

Fuchs, U., Manns, I. & Steinberg, G. 2005, "Microtubules are dispensable for the initial pathogenic development but required for long-distance hyphal growth in the corn smut fungus *Ustilago maydis*", *Molecular biology of the cell,* Vol. 16, No. 6, pp. 2746–2758.

Fukuchi, S. & Nishikawa, K. 2004, "Estimation of the number of authentic orphan genes in bacterial genomes", *DNA research: an international journal for rapid publication of reports on genes and genomes,* Vol. 11, No. 4, pp. 219–31, 311–313.

Fukumura, R., Takahashi, H., Saito, T., Tsutsumi, Y., Fujimori, A., Sato, S., Tatsumi, K., Araki, R. & Abe, M. 2003, "A sensitive transcriptome analysis method that can detect unknown transcripts", *Nucleic acids research,* Vol. 31, No. 16, pp. e94.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C.B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M.A., Werner-Washburne, M., Selitrennikoff, C.P., Kinsey, J.A., Braun, E.L., Zelter, A., Schulte, U., Kothe, G.O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R.L., Perkins, D.D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R.J., Osmani, S.A., DeSouza, C.P., Glass, L., Orbach, M.J., Berglund, J.A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D.O., Alex, L.A., Mannhaupt, G., Ebbole, D.J., Freitag, M., Paulsen, I., Sachs, M.S., Lander, E.S., Nusbaum, C. & Birren, B. 2003, "The genome sequence of the filamentous fungus *Neurospora crassa*", *Nature,* Vol. 422, No. 6934, pp. 859–868.

Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.J., Wortman, J.R., Batzoglou, S., Lee, S.I., Basturkmen, M., Spevak, C.C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G.H., Draht, O., Busch, S., D'Enfert, C., Bouchier, C., Goldman, G.H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J.H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E.U., Archer, D.B., Penalva, M.A., Oakley, B.R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W.C., Denning, D.W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M.S., Osmani, S.A. & Birren, B.W. 2005, "Sequencing of *Aspergillus nidulans* and comparative analysis with *A.fumigatus* and *A. oryzae*", *Nature,* Vol. 438, No. 7071, pp. 1105–1115.

Galagan, J.E. & Selker, E.U. 2004, "RIP: the evolutionary cost of genome defense", *Trends in genetics: TIG,* Vol. 20, No. 9, pp. 417–423.

Gardiner, D.M. & Howlett, B.J. 2005, "Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*", *FEMS microbiology letters,* Vol. 248, No. 2, pp. 241–248.

Gardiner, D.M., Waring, P. & Howlett, B.J. 2005, "The epipolythiodioxopiperazine (ETP) class of fungal toxins: distribution, mode of action, functions and biosynthesis", *Microbiology (Reading, England),* Vol. 151, No. Pt 4, pp. 1021–1032.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D. & Brown, P.O. 2000, "Genomic expression programs in the response of yeast cells to environmental changes", *Mol Biol Cell,* Vol. 11, No. 12, pp. 4241–4257.

Gasser, B., Maurer, M., Rautio, J., Sauer, M., Bhattacharyya, A., Saloheimo, M., Penttila, M. & Mattanovich, D. 2007, "Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions", *BMC genomics,* Vol. 8, p. 179.

Gattiker, A., Rischatsch, R., Demougin, P., Voegeli, S., Dietrich, F.S., Philippsen, P. & Primig, M. 2007, "Ashbya Genome Database 3.0: a cross-species genome and transcriptome browser for yeast biologists", *BMC genomics,* Vol. 8, p. 9.

Gellatly, K.S., Ash, G.J. & Taylor, J.L. 2001, "Development of a method for mRNA differential display in filamentous fungi: comparison of mRNA differential display reverse transcription polymerase chain reaction and cDNA amplified fragment length polymorphism in Leptosphaeria maculans", *Canadian journal of microbiology,* Vol. 47, No. 10, pp. 955–960.

Gierz, G. & Bartnicki-Garcia, S. 2001, "A three-dimensional model of fungal morphogenesis based on the vesicle supply center concept", *Journal of theoretical biology,* Vol. 208, No. 2, pp. 151–164.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S.G. 1996, "Life with 6000 genes", *Science,* Vol. 274, No. 5287, pp. 546, 563–567.

Gordon, C.L., Archer, D.B., Jeenes, D.J., Doonan, J.H., Wells, B., Trinci, A.P.J. & Robson, G.D. 2000, "A glucoamylase::GFP gene fusion to study protein secretion by individual hyphae of *Aspergillus niger*", *Journal of microbiological methods,* Vol. 42, No. 1, pp. 39–48.

Hara, E., Kato, T., Nakada, S., Sekiya, S. & Oda, K. 1991, "Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells", *Nucleic acids research,* Vol. 19, No. 25, pp. 7097–7104.

Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M. & Ron, D. 2000a, "Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells", *Molecular cell,* Vol. 6, No. 5, pp. 1099–1108.

Harding, H.P., Zhang, Y., Bertolotti, A., Zeng, H. & Ron, D. 2000b, "Perk is essential for translational regulation and cell survival during the unfolded protein response", *Mol Cell,* Vol. 5, No. 5, pp. 897–904.

Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calfon, M., Sadri, N., Yun, C., Popko, B. & Paules, R. 2003, "An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress", *Molecular cell,* Vol. 11, No. 3, pp. 619–633.

Hashimoto, H., Kikuchi, Y., Nogi, Y. & Fukasawa, T. 1983, "Regulation of expression of the galactose gene cluster in *Saccharomyces cerevisiae*. Isolation and characterization of the regulatory gene GAL4", *Molecular & general genetics: MGG,* Vol. 191, No. 1, pp. 31–38.

Haynes, C.M., Titus, E.A. & Cooper, A.A. 2004, "Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death", *Molecular cell,* Vol. 15, No. 5, pp. 767–776.

Hedges, S.B. 2002, "The origin and evolution of model organisms", *Nature reviews.Genetics,* Vol. 3, No. 11, pp. 838–849.

Hedrick, S.M., Cohen, D.I., Nielsen, E.A. & Davis, M.M. 1984, "Isolation of cDNA clones encoding T cell-specific membrane-associated proteins", *Nature,* Vol. 308, No. 5955, pp. 149–153.

Herrgard, M.J., Lee, B.-S., Portnoy, V. & Palsson, B.O. "Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae", Genome Research, 2006* (accepted for publication).

Hiltunen, J.K., Wenzel, B., Beyer, A., Erdmann, R., Fossa, A. & Kunau, W.H. 1992, "Peroxisomal multifunctional beta-oxidation protein of *Saccharomyces cerevisiae*. Molecular analysis of the fox2 gene and gene product", *The Journal of biological chemistry,* Vol. 267, No. 10, pp. 6646–6653.

Hinnebusch, A.G. & Natarajan, K. 2002, "Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress", *Eukaryot Cell,* Vol. 1, No. 1, pp. 22–32.

Horio, T. & Oakley, B.R. 2005, "The role of microtubules in rapid hyphal tip growth of *Aspergillus nidulans*", *Molecular biology of the cell,* Vol. 16, No. 2, pp. 918–926.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M. & Friend, S.H. 2000, "Functional discovery via a compendium of expression profiles", *Cell,* Vol. 102, No. 1, pp. 109–126.

Ikeda, K., Nakayashiki, H., Kataoka, T., Tamba, H., Hashimoto, Y., Tosa, Y. & Mayama, S. 2002, "Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context", *Molecular microbiology,* Vol. 45, No. 5, pp. 1355–1364.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigo, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J. & Roest Crollius, H. 2004, "Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype", *Nature,* Vol. 431, No. 7011, pp. 946–957.

Jarvinen, A.K., Hautaniemi, S., Edgren, H., Auvinen, P., Saarela, J., Kallioniemi, O.P. & Monni, O. 2004, "Are data from different gene expression microarray platforms comparable?", *Genomics,* Vol. 83, No. 6, pp. 1164–1168.

Jensen, L.J., Ussery, D.W. & Brunak, S. 2003, "Functionality of system components: conservation of protein function in protein feature space", *Genome research,* Vol. 13, No. 11, pp. 2444–2449.

Jungell-Nortamo, A., Syvanen, A.C., Luoma, P. & Soderlund, H. 1988, "Nucleic acid sandwich hybridization: enhanced reaction rate with magnetic microparticles as carriers", *Molecular and cellular probes,* Vol. 2, No. 4, pp. 281–288.

Kamper, J., Kahmann, R., Bolker, M., Ma, L.J., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E., Muller, O., Perlin, M.H., Wosten, H.A., de Vries, R., Ruiz-Herrera, J., Reynaga-Pena, C.G., Snetselaar, K., McCann, M., Perez-Martin, J., Feldbrugge, M., Basse, C.W., Steinberg, G., Ibeas, J.I., Holloman, W., Guzman, P., Farman, M., Stajich, J.E., Sentandreu, R., Gonzalez-Prieto, J.M., Kennell, J.C., Molina, L., Schirawski, J., Mendoza-Mendoza, A., Greilinger, D., Munch, K., Rossel, N., Scherer, M., Vranes, M., Ladendorf, O., Vincon, V., Fuchs, U., Sandrock, B., Meng, S., Ho, E.C., Cahill, M.J., Boyce, K.J., Klose, J., Klosterman, S.J., Deelstra, H.J., Ortiz-Castellanos, L., Li, W., Sanchez-Alonso, P., Schreier, P.H., Hauser-Hahn, I., Vaupel, M., Koopmann, E., Friedrich, G., Voss, H., Schluter, T., Margolis, J., Platt, D., Swimmer, C.,

Gnirke, A., Chen, F., Vysotskaia, V., Mannhaupt, G., Guldener, U., Munsterkotter, M., Haase, D., Oesterheld, M., Mewes, H.W., Mauceli, E.W., DeCaprio, D., Wade, C.M., Butler, J., Young, S., Jaffe, D.B., Calvo, S., Nusbaum, C., Galagan, J. & Birren, B.W. 2006, "Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*", *Nature,* Vol. 444, No. 7115, pp. 97–101.

Kasuga, T., Townsend, J.P., Tian, C., Gilbert, L.B., Mannhaupt, G., Taylor, J.W. & Glass, N.L. 2005, "Long-oligomer microarray profiling in *Neurospora crassa* reveals the transcriptional program underlying biochemical and physiological events of conidial germination", *Nucleic acids research,* Vol. 33, No. 20, pp. 6469–6485.

Kaufman, R.J. 2002, "Orchestrating the unfolded protein response in health and disease", *J Clin Invest,* Vol. 110, No. 10, pp. 1389–1398.

Keller, N.P. & Hohn, T.M. 1997, "Metabolic Pathway Gene Clusters in Filamentous Fungi", *Fungal genetics and biology: FG & B,* Vol. 21, No. 1, pp. 17–29.

Keller, N.P., Turner, G. & Bennett, J.W. 2005, "Fungal secondary metabolism – from biochemistry to genomics", *Nature reviews.Microbiology,* Vol. 3, No. 12, pp. 937–947.

Kellis, M., Birren, B.W. & Lander, E.S. 2004, "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*", *Nature,* Vol. 428, No. 6983, pp. 617–624.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E.S. 2003, "Sequencing and comparison of yeast species to identify genes and regulatory elements", *Nature,* Vol. 423, No. 6937, pp. 241–254.

Kempken, F. & Kuck, U. 1998, "Transposons in filamentous fungi--facts and perspectives", *BioEssays: news and reviews in molecular, cellular and developmental biology,* Vol. 20, No. 8, pp. 652–659.

Kimata, Y., Ishiwata-Kimata, Y., Yamada, S. & Kohno, K. 2006, "Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins", *Genes to cells: devoted to molecular & cellular mechanisms,* Vol. 11, No. 1, pp. 59–69.

Kimata, Y., Oikawa, D., Shimizu, Y., Ishiwata-Kimata, Y. & Kohno, K. 2004, "A role for BiP as an adjustor for the endoplasmic reticulum stress-sensing protein Ire1", *The Journal of cell biology,* Vol. 167, No. 3, pp. 445–456.

Kivioja, T. 2004, *Computational Tools for a Novel Transcriptional Profiling Method*, University of Helsinki.

Kohno, K., Normington, K., Sambrook, J., Gething, M.J. & Mori, K. 1993, "The promoter region of the yeast KAR2 (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum", *Mol Cell Biol,* Vol. 13, No. 2, pp. 877–890.

Kuhls, K., Lieckfeldt, E., Samuels, G.J., Kovacs, W., Meyer, W., Petrini, O., Gams, W., Borner, T. & Kubicek, C.P. 1996, "Molecular evidence that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina*", *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 93, No. 15, pp. 7755–7760.

Kunau, W.H., Buhne, S., de la Garza, M., Kionka, C., Mateblowski, M., Schultz-Borchard, U. & Thieringer, R. 1988, "Comparative enzymology of beta-oxidation", *Biochemical Society transactions,* Vol. 16, No. 3, pp. 418–420.

Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L. & Kohane, I.S. 2002, "Analysis of matched mRNA measurements from two different microarray technologies", *Bioinformatics,* Vol. 18, No. 3, pp. 405–412.

Kurihara, T., Ueda, M., Okada, H., Kamasawa, N., Naito, N., Osumi, M. & Tanaka, A. 1992, "Beta-oxidation of butyrate, the short-chain-length fatty acid, occurs in peroxisomes in the yeast *Candida tropicalis*", *Journal of Biochemistry,* Vol. 111, No. 6, pp. 783–787.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J. & Giegerich, R. 2001, "REPuter: the manifold applications of repeat analysis on a genomic scale", *Nucleic acids research,* Vol. 29, No. 22, pp. 4633–4642.

Kurtz, S., Ohlebusch, E., Schleiermacher, C., Stoye, J. & Giegerich, R. 2000, "Computation and visualization of degenerate repeats in complete genomes", *Proceedings / ...International Conference on Intelligent Systems for Molecular Biology; ISMB.International Conference on Intelligent Systems for Molecular Biology,* Vol. 8, pp. 228–238.

Kurtzman, C.P. 2003, "Phylogenetic circumscription of *Saccharomyces, kluyveromyces* and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora", *FEMS yeast research,* Vol. 4, No. 3, pp. 233–245.

Leber, J.H., Bernales, S. & Walter, P. 2004, "IRE1-Independent Gain Control of the Unfolded Protein Response", *PLoS Biol,* Vol. 2, No. 8, pp. E235.

Liang, P. & Pardee, A.B. 1992, "Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction", *Science,* Vol. 257, No. 5072, pp. 967–971.

Lisitsyn, N., Lisitsyn, N. & Wigler, M. 1993, "Cloning the differences between two complex genomes", *Science,* Vol. 259, No. 5097, pp. 946–951.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. & Brown, E.L. 1996, "Expression monitoring by hybridization to high-density oligonucleotide arrays", *Nature biotechnology,* Vol. 14, No. 13, pp. 1675–1680.

Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I.J., Fraser, J.A., Allen, J.E., Bosdet, I.E., Brent, M.R., Chiu, R., Doering, T.L., Donlin, M.J., D'Souza, C.A., Fox, D.S., Grinberg, V., Fu, J., Fukushima, M., Haas, B.J., Huang, J.C., Janbon, G., Jones, S.J., Koo, H.L., Krzywinski, M.I., Kwon-Chung, J.K., Lengeler, K.B., Maiti, R., Marra, M.A., Marra, R.E., Mathewson, C.A., Mitchell, T.G., Pertea, M., Riggs, F.R., Salzberg, S.L., Schein, J.E., Shvartsbeyn, A., Shin, H., Shumway, M., Specht, C.A.,

Suh, B.B., Tenney, A., Utterback, T.R., Wickes, B.L., Wortman, J.R., Wye, N.H., Kronstad, J.W., Lodge, J.K., Heitman, J., Davis, R.W., Fraser, C.M. & Hyman, R.W. 2005, "The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*", *Science,* Vol. 307, No. 5713, pp. 1321–1324.

Ma, Y. & Hendershot, L.M. 2004, "ER chaperone functions during normal and stress conditions", *Journal of chemical neuroanatomy,* Vol. 28, No. 1–2, pp. 51–65.

Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K., Arima, T., Akita, O., Kashiwagi, Y., Abe, K., Gomi, K., Horiuchi, H., Kitamoto, K., Kobayashi, T., Takeuchi, M., Denning, D.W., Galagan, J.E., Nierman, W.C., Yu, J., Archer, D.B., Bennett, J.W., Bhatnagar, D., Cleveland, T.E., Fedorova, N.D., Gotoh, O., Horikawa, H., Hosoyama, A., Ichinomiya, M., Igarashi, R., Iwashita, K., Juvvadi, P.R., Kato, M., Kato, Y., Kin, T., Kokubun, A., Maeda, H., Maeyama, N., Maruyama, J., Nagasaki, H., Nakajima, T., Oda, K., Okada, K., Paulsen, I., Sakamoto, K., Sawano, T., Takahashi, M., Takase, K., Terabayashi, Y., Wortman, J.R., Yamada, O., Yamagata, Y., Anazawa, H., Hata, Y., Koide, Y., Komori, T., Koyama, Y., Minetoki, T., Suharnan, S., Tanaka, A., Isono, K., Kuhara, S., Ogasawara, N. & Kikuchi, H. 2005, "Genome sequencing and analysis of *Aspergillus oryzae*", *Nature,* Vol. 438, No. 7071, pp. 1157–1161.

MacKenzie, D.A., Guillemette, T., Al-Sheikh, H., Watson, A.J., Jeenes, D.J., Wongwathanarat, P., Dunn-Coleman, N.S., van Peij, N. & Archer, D.B. 2005, "UPR-independent dithiothreitol stress-induced genes in *Aspergillus niger*", *Molecular genetics and genomics: MGG,* Vol. 274, No. 4, pp. 410–418.

Maeda, H., Sano, M., Maruyama, Y., Tanno, T., Akao, T., Totsuka, Y., Endo, M., Sakurada, R., Yamagata, Y., Machida, M., Akita, O., Hasegawa, F., Abe, K., Gomi, K., Nakajima, T. & Iguchi, Y. 2004, "Transcriptional analysis of genes for energy catabolism and hydrolytic enzymes in the filamentous fungus *Aspergillus oryzae* using cDNA microarrays and expressed sequence tags", *Applied Microbiology and Biotechnology,* Vol. 65, No. 1, pp. 74–83.

Maggio-Hall, L.A. & Keller, N.P. 2004, "Mitochondrial beta-oxidation in *Aspergillus nidulans*", *Molecular microbiology,* Vol. 54, No. 5, pp. 1173–1185.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. 2005, "Genome sequencing in microfabricated high-density picolitre reactors", *Nature,* Vol. 437, No. 7057, pp. 376–380.

Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F., Coutinho, P.M., Henrissat, B., Berka, R., Cullen, D. & Rokhsar, D. 2004, "Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78", *Nature biotechnology,* Vol. 22, No. 6, pp. 695–700.

Martinez, I.M. & Chrispeels, M.J. 2003, "Genomic analysis of the unfolded protein response in Arabidopsis shows its connection to important cellular processes", *Plant Cell,* Vol. 15, No. 2, pp. 561–576.

Marzluff, W.F. & Duronio, R.J. 2002, "Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences", *Curr Opin Cell Biol,* Vol. 14, No. 6, pp. 692–699.

Mathieu-Daude, F., Cheng, R., Welsh, J. & McClelland, M. 1996, "Screening of differentially amplified cDNA products from RNA arbitrarily primed PCR fingerprints using single strand conformation polymorphism (SSCP) gels", *Nucleic acids research,* Vol. 24, No. 8, pp. 1504–1507.

Maukonen, J., Satokari, R., Matto, J., Soderlund, H., Mattila-Sandholm, T. & Saarela, M. 2006, "Prevalence and temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria", *Journal of medical microbiology,* Vol. 55, No. Pt 5, pp. 625–633.

Meshi, T., Taoka, K.I. & Iwabuchi, M. 2000, "Regulation of histone gene expression during the cell cycle", *Plant Mol Biol,* Vol. 43, No. 5–6, pp. 643–657.

Mewes, H.W., Amid, C., Arnold, R., Frishman, D., Guldener, U., Mannhaupt, G., Munsterkotter, M., Pagel, P., Strack, N., Stumpflen, V., Warfsmann, J. & Ruepp, A. 2004, "MIPS: analysis and annotation of proteins from whole genomes", *Nucleic Acids Res,* Vol. 32 Database issue, pp. D41–4.

Montiel, M.D., Lee, H.A. & Archer, D.B. 2006, "Evidence of RIP (repeat-induced point mutation) in transposase sequences of *Aspergillus oryzae*", *Fungal genetics and biology: FG & B,* Vol. 43, No. 6, pp. 439–445.

Mori, K., Ogawa, N., Kawahara, T., Yanagi, H. & Yura, T. 1998, "Palindrome with spacer of one nucleotide is characteristic of the cis-acting unfolded protein response element in *Saccharomyces cerevisiae*", *J Biol Chem,* Vol. 273, No. 16, pp. 9912–9920.

Mori, K., Sant, A., Kohno, K., Normington, K., Gething, M.J. & Sambrook, J.F. 1992, "A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins", *Embo J,* Vol. 11, No. 7, pp. 2583–2593.

Mourino-Perez, R.R., Roberson, R.W. & Bartnicki-Garcia, S. 2006, "Microtubule dynamics and organization during hyphal growth and branching in *Neurospora crassa*", *Fungal genetics and biology: FG & B,* Vol. 43, No. 6, pp. 389–400.

Mulder, H.J., Saloheimo, M., Penttila, M. & Madrid, S.M. 2004, "The transcription factor HACA mediates the unfolded protein response in *Aspergillus niger*, and up-regulates its own transcription", *Mol Genet Genomics,* Vol. 271, No. 2, pp. 130–140.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A.N., Orchard, S., Pagni, M.,

Ponting, C.P., Quevillon, E., Selengut, J., Sigrist, C.J., Silventoinen, V., Studholme, D.J., Vaughan, R. & Wu, C.H. 2005, "InterPro, progress and status in 2005", *Nucleic Acids Res,* Vol. 33, No. Database issue, pp. D201–5.

Myers 1999, "A fast bit-vector algorithm for approximate string matching based on dynamic programming", *Journal of the ACM,* Vol. 46, No. 3, pp. 395.

Nakari, T., Alatalo, E. & Penttila, M.E. 1993, "Isolation of *Trichoderma reesei* genes highly expressed on glucose-containing media: characterization of the tef1 gene encoding translation elongation factor 1 alpha", *Gene,* Vol. 136, No. 1–2, pp. 313–318.

Natarajan, K., Meyer, M.R., Jackson, B.M., Slade, D., Roberts, C., Hinnebusch, A.G. & Marton, M.J. 2001, "Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast", *Mol Cell Biol,* Vol. 21, No. 13, pp. 4347–4368.

Naumov, G.I., Naumova, E.S., Sancho, E.D. & Korhola, M.P. 1996, "Polymeric SUC genes in natural populations of *Saccharomyces cerevisiae*", *FEMS microbiology letters,* Vol. 135, No. 1, pp. 31–35.

Nierman, W.C., Pain, A., Anderson, M.J., Wortman, J.R., Kim, H.S., Arroyo, J., Berriman, M., Abe, K., Archer, D.B., Bermejo, C., Bennett, J., Bowyer, P., Chen, D., Collins, M., Coulsen, R., Davies, R., Dyer, P.S., Farman, M., Fedorova, N., Fedorova, N., Feldblyum, T.V., Fischer, R., Fosker, N., Fraser, A., Garcia, J.L., Garcia, M.J., Goble, A., Goldman, G.H., Gomi, K., Griffith-Jones, S., Gwilliam, R., Haas, B., Haas, H., Harris, D., Horiuchi, H., Huang, J., Humphray, S., Jimenez, J., Keller, N., Khouri, H., Kitamoto, K., Kobayashi, T., Konzack, S., Kulkarni, R., Kumagai, T., Lafon, A., Latge, J.P., Li, W., Lord, A., Lu, C., Majoros, W.H., May, G.S., Miller, B.L., Mohamoud, Y., Molina, M., Monod, M., Mouyna, I., Mulligan, S., Murphy, L., O'Neil, S., Paulsen, I., Penalva, M.A., Pertea, M., Price, C., Pritchard, B.L., Quail, M.A., Rabbinowitsch, E., Rawlins, N., Rajandream, M.A., Reichard, U., Renauld, H., Robson, G.D., Rodriguez de Cordoba, S., Rodriguez-Pena, J.M., Ronning, C.M., Rutter, S., Salzberg, S.L., Sanchez, M., Sanchez-Ferrero, J.C., Saunders, D., Seeger, K., Squares, R., Squares, S., Takeuchi, M., Tekaia, F., Turner, G., Vazquez de Aldana, C.R., Weidman, J., White, O., Woodward, J., Yu, J.H., Fraser, C., Galagan, J.E., Asai, K.,

Machida, M., Hall, N., Barrell, B. & Denning, D.W. 2005, "Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*", *Nature,* Vol. 438, No. 7071, pp. 1151–1156.

Nykanen, M., Saarelainen, R., Raudaskoski, M., Nevalainen, K. & Mikkonen, A. 1997, "Expression and Secretion of Barley Cysteine Endopeptidase B and Cellobiohydrolase I in *Trichoderma reesei*", *Applied and Environmental Microbiology,* Vol. 63, No. 12, pp. 4929–4937.

Pakula, T.M., Laxell, M., Huuskonen, A., Uusitalo, J., Saloheimo, M. & Penttila, M. 2003, "The effects of drugs inhibiting protein secretion in the filamentous fungus *Trichoderma reesei*. Evidence for down-regulation of genes that encode secreted proteins in the stressed cells", *J Biol Chem,* Vol. 278, No. 45, pp. 45011–45020.

Paluh, J.L., Orbach, M.J., Legerton, T.L. & Yanofsky, C. 1988, "The cross-pathway control gene of Neurospora crassa, cpc-1, encodes a protein similar to GCN4 of yeast and the DNA-binding domain of the oncogene v-jun-encoded protein", *Proc Natl Acad Sci U S A,* Vol. 85, No. 11, pp. 3728–3732.

Park, P.J., Cao, Y.A., Lee, S.Y., Kim, J.W., Chang, M.S., Hart, R. & Choi, S. 2004, "Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference", *Journal of Biotechnology,* Vol. 112, No. 3, pp. 225–245.

Patil, C. & Walter, P. 2001, "Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals", *Current opinion in cell biology,* Vol. 13, No. 3, pp. 349–355.

Patil, C.K., Li, H. & Walter, P. 2004, "Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response", *PLoS Biol,* Vol. 2, No. 8, pp. E246.

Payne, G.A., Nierman, W.C., Wortman, J.R., Pritchard, B.L., Brown, D., Dean, R.A., Bhatnagar, D., Cleveland, T.E., Machida, M. & Yu, J. 2006, "Whole genome comparison of *Aspergillus flavus* and *A. oryzae*", *Medical mycology: official publication of the International Society for Human and Animal Mycology,* Vol. 44 Suppl, pp. 9–11.

Pedersen, A.G., Jensen, L.J., Brunak, S., Staerfeldt, H.H. & Ussery, D.W. 2000, "A DNA structural atlas for *Escherichia coli*", *Journal of Molecular Biology,* Vol. 299, No. 4, pp. 907–930.

van Peij, N.N., Visser, J. & de Graaff, L.H. 1998, "Isolation and analysis of xlnR, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*", *Molecular microbiology,* Vol. 27, No. 1, pp. 131–142.

Pel, H.J., de Winde, J.H., Archer, D.B., Dyer, P.S., Hofmann, G., Schaap, P.J., Turner, G., de Vries, R.P., Albang, R., Albermann, K., Andersen, M.R., Bendtsen, J.D., Benen, J.A., van den Berg, M., Breestraat, S., Caddick, M.X., Contreras, R., Cornell, M., Coutinho, P.M., Danchin, E.G., Debets, A.J., Dekker, P., van Dijck, P.W., van Dijk, A., Dijkhuizen, L., Driessen, A.J., d'Enfert, C., Geysens, S., Goosen, C., Groot, G.S., de Groot, P.W., Guillemette, T., Henrissat, B., Herweijer, M., van den Hombergh, J.P., van den Hondel, C.A., van der Heijden, R.T., van der Kaaij, R.M., Klis, F.M., Kools, H.J., Kubicek, C.P., van Kuyk, P.A., Lauber, J., Lu, X., van der Maarel, M.J., Meulenberg, R., Menke, H., Mortimer, M.A., Nielsen, J., Oliver, S.G., Olsthoorn, M., Pal, K., van Peij, N.N., Ram, A.F., Rinas, U., Roubos, J.A., Sagt, C.M., Schmoll, M., Sun, J., Ussery, D., Varga, J., Vervecken, W., van de Vondervoort, P.J., Wedler, H., Wosten, H.A., Zeng, A.P., van Ooyen, A.J., Visser, J. & Stam, H. 2007, "Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88", *Nature biotechnology,* Vol. 25, No. 2, pp. 221–231.

Penttila, I.M. & Limon, C. 2004, "Molecular Biology of *Trichoderma* and Biotechnological Applications" in *Handbook of Fungal Biotechnology*, ed. D.K. Arora, Marcel Dekker, Inc, New York, pp. 413–427.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D. & Holstege, F.C. 2003, "Monitoring global messenger RNA changes in externally controlled microarray experiments", *EMBO reports,* Vol. 4, No. 4, pp. 387–393.

Qin, L., Overmars, H., Helder, J., Popeijus, H., van der Voort, J.R., Groenink, W., van Koert, P., Schots, A., Bakker, J. & Smant, G. 2000, "An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*", *Molecular plant-microbe interactions: MPMI,* Vol. 13, No. 8, pp. 830–836.

Qin, L., Prins, P., Jones, J.T., Popeijus, H., Smant, G., Bakker, J. & Helder, J. 2001, "GenEST, a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP", *Nucleic acids research,* Vol. 29, No. 7, pp. 1616–1622.

Qin, M., Zeng, Z., Zheng, J., Shah, P.K., Schwartz, S.M., Adams, L.D. & Sharifi, B.G. 2003, "Suppression subtractive hybridization identifies distinctive expression markers for coronary and internal mammary arteries", *Arteriosclerosis, Thrombosis, and Vascular Biology,* Vol. 23, No. 3, pp. 425–433.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. & Lopez, R. 2005, "InterProScan: protein domains identifier", *Nucleic Acids Res,* Vol. 33, No. Web Server issue, pp. W116–20.

Rand, J.D. & Grant, C.M. 2006, "The thioredoxin system protects ribosomes against stress-induced aggregation", *Molecular biology of the cell,* Vol. 17, No. 1, pp. 387–401.

Raskin, D.M., Seshadri, R., Pukatzki, S.U. & Mekalanos, J.J. 2006/2/24, "Bacterial Genomics and Pathogen Evolution", *Cell,* Vol. 124, No. 4, pp. 703–714.

Rautio, J.J., Huuskonen, A., Vuokko, H. and Londesborough, J. 2007. Monitoring yeast physiology during very high gravity wort fermentations by frequent analysis of gene expression. *Yeast*. 24:741–760.

Rautio, J.J., Kataja, K., Satokari, R., Penttila, M., Soderlund, H. & Saloheimo, M. 2006a, "Rapid and multiplexed transcript analysis of microbial cultures using capillary electophoresis-detectable oligonucleotide probe pools", *Journal of microbiological methods,* Vol. 65, No. 3, pp. 404–416.

Rautio, J.J., Smit, B.A., Wiebe, M., Penttila, M. & Saloheimo, M. 2006b, "Transcriptional monitoring of steady state and effects of anaerobic phases in chemostat cultures of the filamentous fungus *Trichoderma reesei*", *BMC genomics,* Vol. 7, pp. 247.

Reese, E., Lavinson, H., Downing, M. & White, L. 1950, "Quartermaster culture collection", *Farlowia,* Vol. 4, pp. 45–86.

Rehmeyer, C., Li, W., Kusaba, M., Kim, Y.S., Brown, D., Staben, C., Dean, R. & Farman, M. 2006, "Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*", *Nucleic acids research,* Vol. 34, No. 17, pp. 4685–4701.

Rombauts, S., Van De Peer, Y. & Rouze, P. 2003, "AFLPinSilico, simulating AFLP fingerprints", *Bioinformatics (Oxford, England),* Vol. 19, No. 6, pp. 776–777.

Rozen, S. & Skaletsky, H. 2000, "Primer3 on the WWW for general users and for biologist programmers", *Methods in molecular biology (Clifton, N.J.),* Vol. 132, pp. 365–386.

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M. & Mewes, H.W. 2004, "The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes", *Nucleic acids research,* Vol. 32, No. 18, pp. 5539–5545.

Saloheimo, M., Lund, M. & Penttila, M.E. 1999, "The protein disulphide isomerase gene of the fungus *Trichoderma reesei* is induced by endoplasmic reticulum stress and regulated by the carbon source", *Mol Gen Genet,* Vol. 262, No. 1, pp. 35–45.

Saloheimo, M., Valkonen, M. & Penttila, M. 2003, "Activation mechanisms of the HAC1-mediated unfolded protein response in filamentous fungi", *Mol Microbiol,* Vol. 47, No. 4, pp. 1149–1161.

Sampson, K. & Heath, I.B. 2005, "The dynamic behaviour of microtubules and their contributions to hyphal tip growth in *Aspergillus nidulans*", *Microbiology (Reading, England),* Vol. 151, No. Pt 5, pp. 1543–1555.

Satokari, R.M., Kataja, K. & Soderlund, H. 2005, "Multiplexed quantification of bacterial 16S rRNA by solution hybridization with oligonucleotide probes and affinity capture", *Microbial ecology,* Vol. 50, No. 1, pp. 120–127.

Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. 1995, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray", *Science,* Vol. 270, No. 5235, pp. 467–470.

Schroder, M., Chang, J.S. & Kaufman, R.J. 2000, "The unfolded protein response represses nitrogen-starvation induced developmental differentiation in yeast", *Genes & development,* Vol. 14, No. 23, pp. 2962–2975.

Schroder, M., Clark, R., Liu, C.Y. & Kaufman, R.J. 2004, "The unfolded protein response represses differentiation through the RPD3-SIN3 histone deacetylase", *The EMBO journal,* Vol. 23, No. 11, pp. 2281–2292.

Sims, A.H., Gent, M.E., Lanthaler, K., Dunn-Coleman, N.S., Oliver, S.G. & Robson, G.D. 2005, "Transcriptome analysis of recombinant protein secretion by *Aspergillus nidulans* and the unfolded-protein response in vivo", *Appl Environ Microbiol,* Vol. 71, No. 5, pp. 2737–2747.

Skovgaard, M., Jensen, L.J., Brunak, S., Ussery, D. & Krogh, A. 2001, "On the total number of genes and their length distribution in complete microbial genomes", *Trends in genetics: TIG,* Vol. 17, No. 8, pp. 425–428.

Smith, J.J., Brown, T.W., Eitzen, G.A. & Rachubinski, R.A. 2000, "Regulation of peroxisome size and number by fatty acid beta -oxidation in the yeast *Yarrowia lipolytica*", *The Journal of biological chemistry,* Vol. 275, No. 26, pp. 20168–20178.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A. & Lewis, S. 2002, "The generic genome browser: a building block for a model organism system database", *Genome research,* Vol. 12, No. 10, pp. 1599–1610.

Syvanen, A.C., Laaksonen, M. & Soderlund, H. 1986, "Fast quantification of nucleic acid hybrids by affinity-based hybrid collection", *Nucleic acids research,* Vol. 14, No. 12, pp. 5037–5048.

Thompson, J. & Gillespie, D. 1987, "Molecular hybridization with RNA probes in concentrated solutions of guanidine thiocyanate", *Analytical Biochemistry,* Vol. 163, No. 2, pp. 281–291.

Tindemans, S.H., Kern, N. & Mulder, B.M. 2006, "The diffusive vesicle supply center model for tip growth in fungal hyphae", *Journal of theoretical biology,* Vol. 238, No. 4, pp. 937–948.

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S. & Walter, P. 2000, "Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation", *Cell,* Vol. 101, No. 3, pp. 249–258.

Tsai, B., Ye, Y. & Rapoport, T.A. 2002, "Retro-translocation of proteins from the endoplasmic reticulum into the cytosol", *Nature reviews.Molecular cell biology,* Vol. 3, No. 4, pp. 246–255.

Tu, B.P. & Weissman, J.S. 2004, "Oxidative protein folding in eukaryotes: mechanisms and consequences", *The Journal of cell biology,* Vol. 164, No. 3, pp. 341–346.

Tu, B.P. & Weissman, J.S. 2002, "The FAD- and O2-Dependent Reaction Cycle of Ero1-Mediated Oxidative Protein Folding in the Endoplasmic Reticulum", *Molecular cell,* Vol. 10, No. 5, pp. 983–994.

Valkonen, M., Penttila, M. & Saloheimo, M. 2004, "The ire1 and ptc2 genes involved in the unfolded protein response pathway in the filamentous fungus *Trichoderma reesei*", *Mol Genet Genomics,* Vol. 272, No. 4, pp. 443–451.

Valkonen, M., Penttila, M. & Saloheimo, M. 2003a, "Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*", *Appl Environ Microbiol,* Vol. 69, No. 4, pp. 2065–2072.

Valkonen, M., Ward, M., Wang, H., Penttila, M. & Saloheimo, M. 2003b, "Improvement of foreign-protein production in *Aspergillus niger* var. *awamori* by constitutive induction of the unfolded-protein response", *Appl Environ Microbiol,* Vol. 69, No. 12, pp. 6979–6986.

Vastenhouw, N.L. & Plasterk, R.H. 2004, "RNAi protects the *Caenorhabditis elegans* germline against transposition", *Trends in genetics: TIG,* Vol. 20, No. 7, pp. 314–319.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E.,Jr, Hieter, P., Vogelstein, B. & Kinzler, K.W. 1997, "Characterization of the yeast transcriptome", *Cell,* Vol. 88, No. 2, pp. 243–251.

Venter, J.C., Adams, M.D., Sutton, G.G., Kerlavage, A.R., Smith, H.O. & Hunkapiller, M. 1998, "Shotgun sequencing of the human genome", *Science,* Vol. 280, No. 5369, pp. 1540–1542.

von Stein, O.D., Thies, W.G. & Hofmann, M. 1997, "A high throughput screening for rarely transcribed differentially expressed genes", *Nucleic acids research,* Vol. 25, No. 13, pp. 2598–2602.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M. 1995, "AFLP: a new technique for DNA fingerprinting", *Nucleic acids research,* Vol. 23, No. 21, pp. 4407–4414.
Woloshuk, C.P., Foutz, K.R., Brewer, J.F., Bhatnagar, D., Cleveland, T.E. & Payne, G.A. 1994, "Molecular characterization of aflR, a regulatory locus for aflatoxin biosynthesis", *Applied and Environmental Microbiology,* Vol. 60, No. 7, pp. 2408–2414.

Wanke, C., Eckert, S., Albrecht, G., van Hartingsveldt, W., Punt, P.J., van den Hondel, C.A. & Braus, G.H. 1997, "The *Aspergillus niger* GCN4 homologue, cpcA, is transcriptionally regulated and encodes an unusual leucine zipper", *Mol Microbiol,* Vol. 23, No. 1, pp. 23–33.

Watson, A.J., Worley, J., Elliott, R.M., Jeenes, D.J. & Archer, D.B. 2000, "Cloning Stress-Induced Genes from *Aspergillus niger* Using Polymerase Chain Reaction-Augmented Subtractive Hybridization", *Analytical Biochemistry,* Vol. 277, No. 1, pp. 162–165.

Weidenhammer, E.M., Kahl, B.F., Wang, L., Wang, L., Duhon, M., Jackson, J.A., Slater, M. & Xu, X. 2002, "Multiplexed, targeted gene expression profiling

and genetic analysis on electronic microarrays", *Clinical chemistry,* Vol. 48, No. 11, pp. 1873–1882.

Weis, J.H., Tan, S.S., Martin, B.K. & Wittwer, C.T. 1992, "Detection of rare mRNAs via quantitative RT-PCR", *Trends in genetics: TIG,* Vol. 8, No. 8, pp. 263–264.

Wiebe, M.G., Karandikar, A., Robson, G.D., Trinci, A.P., Candia, J.L., Trappe, S., Wallis, G., Rinas, U., Derkx, P.M., Madrid, S.M., Sisniega, H., Faus, I., Montijn, R., van den Hondel, C.A. & Punt, P.J. 2001, "Production of tissue plasminogen activator (t-PA) in *Aspergillus niger*", *Biotechnology and bioengineering,* Vol. 76, No. 2, pp. 164–174.

Wilson, J.W., Schurr, M.J., LeBlanc, C.L., Ramamurthy, R., Buchanan, K.L. & Nickerson, C.A. 2002, "Mechanisms of bacterial pathogenicity", *Postgraduate medical journal,* Vol. 78, No. 918, pp. 216–224.

Wolfe, S.A., Ramm, E.I. & Pabo, C.O. 2000, "Combining structure-based design with phage display to create new Cys2His2 zinc finger dimers", *Structure with Folding & design,* Vol. 8, No. 7, pp. 739–750.

Woloshuk, C.P., Foutz, K.R., Brewer, J.F., Bhatnagar, D., Cleveland, T.E. & Payne, G.A. 1994, "Molecular characterization of aflR, a regulatory locus for aflatoxin biosynthesis", *Applied and Environmental Microbiology,* Vol. 60, No. 7, pp. 2408–2414.

Wu, J. & Kaufman, R.J. 2006, "From acute ER stress to physiological roles of the Unfolded Protein Response", *Cell death and differentiation,* Vol. 13, No. 3, pp. 374–384.

Xie, X., Wilkinson, H.H., Correa, A., Lewis, Z.A., Bell-Pedersen, D. & Ebbole, D.J. 2004, "Transcriptional response to glucose starvation and functional analysis of a glucose transporter of *Neurospora crassa*", *Fungal genetics and biology: FG & B,* Vol. 41, No. 12, pp. 1104–1119.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G.,

Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. & Ecker, J.R. 2003, "Empirical analysis of transcriptional activity in the *Arabidopsis* genome", *Science,* Vol. 302, No. 5646, pp. 842–846.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S. & Fu, X.D. 2002, "Profiling alternative splicing on fiber-optic arrays", *Nature biotechnology,* Vol. 20, No. 4, pp. 353–358.

Yu, J., Chang, P.K., Ehrlich, K.C., Cary, J.W., Bhatnagar, D., Cleveland, T.E., Payne, G.A., Linz, J.E., Woloshuk, C.P. & Bennett, J.W. 2004, "Clustered pathway genes in aflatoxin biosynthesis", *Applied and Environmental Microbiology,* Vol. 70, No. 3, pp. 1253–1262.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L. & Yang, H. 2002, "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)", *Science,* Vol. 296, No. 5565, pp. 79–92.

Yu, J., Wang, J., Lin, W., Li, S., Li, H., Zhou, J., Ni, P., Dong, W., Hu, S., Zeng, C., Zhang, J., Zhang, Y., Li, R., Xu, Z., Li, S., Li, X., Zheng, H., Cong, L., Lin, L.,

Yin, J., Geng, J., Li, G., Shi, J., Liu, J., Lv, H., Li, J., Wang, J., Deng, Y., Ran, L., Shi, X., Wang, X., Wu, Q., Li, C., Ren, X., Wang, J., Wang, X., Li, D., Liu, D., Zhang, X., Ji, Z., Zhao, W., Sun, Y., Zhang, Z., Bao, J., Han, Y., Dong, L., Ji, J., Chen, P., Wu, S., Liu, J., Xiao, Y., Bu, D., Tan, J., Yang, L., Ye, C., Zhang, J., Xu, J., Zhou, Y., Yu, Y., Zhang, B., Zhuang, S., Wei, H., Liu, B., Lei, M., Yu, H., Li, Y., Xu, H., Wei, S., He, X., Fang, L., Zhang, Z., Zhang, Y., Huang, X., Su, Z., Tong, W., Li, J., Tong, Z., Li, S., Ye, J., Wang, L., Fang, L., Lei, T., Chen, C., Chen, H., Xu, Z., Li, H., Huang, H., Zhang, F., Xu, H., Li, N., Zhao, C., Li, S., Dong, L., Huang, Y., Li, L., Xi, Y., Qi, Q., Li, W., Zhang, B., Hu, W., Zhang, Y., Tian, X., Jiao, Y., Liang, X., Jin, J., Gao, L., Zheng, W., Hao, B., Liu, S., Wang, W., Yuan, L., Cao, M., McDermott, J., Samudrala, R., Wang, J., Wong, G.K. & Yang, H. 2005, "The Genomes of *Oryza sativa*: a history of duplications", *PLoS biology,* Vol. 3, No. 2, pp. e38.

Yuen, T., Wurmbach, E., Pfeffer, R.L., Ebersole, B.J. & Sealfon, S.C. 2002, "Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays", *Nucleic acids research,* Vol. 30, No. 10, pp. e48.

Appendices I–V of this publication are not included in the PDF version.
Please order the printed version to get the complete publication
(http://www.vtt.fi/publications/index.jsp)

Series title, number and
report code of publication

VTT Publications 654
VTT–PUBS–654

Author(s)
Arvas, Mikko

Title

Comparative and functional genome analysis of fungi for development of the protein production host *Trichoderma reesei*

Abstract

Filamentous fungi of the subphylum Pezizomycotina are well known as protein and secondary metabolite producers. Various industries take advantage of these capabilities. However, the molecular biology of yeasts, i.e. Saccharomycotina and especially that of *Saccharomyces cerevisiae*, the baker's yeast, is much better known. In an effort to explain fungal phenotypes through their genotypes we have compared protein coding gene contents of Pezizomycotina and Saccharomycotina. Only biomass degradation and secondary metabolism related protein families seem to have expanded recently in Pezizomycotina. Of the protein families clearly diverged between Pezizomycotina and Saccharomycotina, those related to mitochondrial functions emerge as the most prominent. However, the primary metabolism as described in *S. cerevisiae* is largely conserved in all fungi.

Previous studies of individual Pezizomycotina genomes have shown that regardless of the difference in production efficiency and diversity of secreted proteins, the content of the known secretion machinery genes in Pezizomycotina and Saccharomycotina appears very similar. Genome wide analysis of gene products is therefore needed to better understand the efficient secretion of Pezizomycotina. We have developed methods applicable to transcriptome analysis of non-sequenced organisms. TRAC (Transcriptional profiling with the aid of affinity capture) has been previously developed at VTT for fast, focused transcription analysis. We introduce a version of TRAC that allows more powerful signal amplification and multiplexing. We also present computational optimisations of transcriptome analysis of non-sequenced organism and TRAC analysis in general.

Trichoderma reesei is one of the most commonly used Pezizomycotina in the protein production industry. In order to understand its secretion system better and find clues for improvement of its industrial performance, we have analysed its transcriptomic response to protein secretion stress conditions. In comparison to *S. cerevisiae*, the response of *T. reesei* appears different, but still impacts on the same cellular functions. We also discovered in *T. reesei* interesting similarities to mammalian protein secretion stress response. Together these findings highlight targets for more detailed studies.

ISBN
978-951-38-7043-0 (soft back ed.)
978-951-38-7044-7 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN	Project number
VTT Publications 1235–0621 (soft back ed.) 1455–0849 (URL: http://www.vtt.fi/publications/index.jsp)	15643

Date	Language	Pages
September 2007	English	100 p. + app. 105 p.

Keywords

Trichoderma reesei, *Hypocrea jecorina*, *Saccharomyces cerevisiae*, comparative genomics, plant biomass degradation, secondary metabolism, transcriptome profiling, TRAC, cDNA-AFLP, cDNA subtraction library, protein secretion stress, Unfolded Protein Response, UPR, oxidative stress, histones, GCN4, CPC1, ATF4

Publisher

VTT Technical Research Centre of Finland
P.O. Box 1000, FI-02044 VTT, Finland
Phone internat. +358 20 722 4404
Fax +358 20 722 4374

VTT PUBLICATIONS

Julkaisu on saatavana	Publikationen distribueras av	This publication is available from
VTT PL 1000 02044 VTT Puh. 020 722 4404 Faksi 020 722 4374	VTT PB 1000 02044 VTT Tel. 020 722 4404 Fax 020 722 4374	VTT P.O. Box 1000 FI-02044 VTT, Finland Phone internat. + 358 20 722 4404 Fax + 358 20 722 4374

ISBN 978-951-38-7043-0 (soft back ed.) ISBN 978-951-38-7044-7 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1235-0621 (soft back ed.) ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)