Role of Rim101p in the pH response in Candida albicans

Michael Weyler

To cite this version:

Michael Weyler. Role of Rim101p in the pH response in Candida albicans. Biochemistry [q-bio.BM]. Université Paris Sud - Paris XI, 2007. English. NNT : . tel-00165802

HAL Id: tel-00165802

https://theses.hal.science/tel-00165802

Submitted on 27 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS XI
UFR SCIENTIFIQUE D'ORSAY

THESE

présentée par

Michael Weyler

pour obtenir le grade de

DOCTEUR EN SCIENCES
DE L'UNIVERSITE PARIS XI-ORSAY

LE RÔLE DE RIM101p DANS LA RÉPONSE AU pH

CHEZ *CANDIDA ALBICANS*

Soutenance prévue le 6 juillet 2007 devant le jury composé de:

Pr. Dr. H. Delacroix	**Président**
Dr. J-M. Camadro	**Rapporteur**
Pr. Dr. F. M. Klis	**Rapporteur**
Dr. G. Janbon	**Examinateur**
Dr. M. Lavie-Richard	**Examinateur**
Pr. Dr. C. Gaillardin	**Examinateur**

Remerciements

Tout d'abord je voudrais remercier vivement mon directeur de thèse, Prof. Claude Gaillardin, pour m'avoir permis d'effectuer ce travail au sein de son laboratoire, pour ses conseils et sa disponibilité malgré son calendrier bien remplis. Je lui remercie également pour m'avoir laissé beaucoup de liberté dans mon travail, et pour la possibilité de participer aux différents congrès au cours de ma formation de thèse.

Ensuite je tiens à remercier tous les autres membres du jury : à Dr Jean-Michel Camadro et Prof. Frans Klis d'avoir accepté d'être rapporteurs de cette thèse, à Prof. Hervé Delacroix pour la présidence, et à Dr. Guilhem Janbon et Dr. Mathias Lavie-Richard d'y participer en tant qu'examinateurs. A part Claude, j'y voudrais particulièrement mentionner deux personnes, qui m'ont soutenu tout au long de mon parcours de thésard : Frans en étant mon tuteur au sein du réseau européen « Galar FungailII » a qui j'ai toujours pu m'adresser pour des conseils indépendants, et Mathias, qui était pour moi la première adresse pour les grands et petits problèmes de toute sorte, en commençant par l'encadrement de mon travail pratique au labo, en passant par les problèmes administratives et en allant jusqu'à la correction orthographique de ma thèse.

Un grand merci également à Grégory et Frédérique qui ont débuté ce travail avant mon arrivée au laboratoire et qui m'ont permis de prendre le relais sans faire tomber le bâton. Je remercie toute l'équipe de « Signalisation et Virulence » pour l'accueil chaleureuse et leur sincérité qui m'ont permis de vite trouver ma place dans ce groupe malgré mes difficultés avec la langue française.

A ce propos, je voudrais également remercier Marie-Thérèse pour l'organisation de mes courses de français, et bien sûr Elisabeth et Colette pour l'enseignement. Sans eux, je crains que ces lignes soient complètement incompréhensibles. Je remercie aussi les membres du secrétariat, en particulier Caroline, Michèle et Anne pour l'aide avec toutes les démarches administratives qui étaient nécessaires tout au long de mon séjour au laboratoire.

Ensuite je tiens à remercier tout les autres membres du laboratoire de Microbiologie et génétique moléculaire de Grignon qui ont rendus mon séjour ici très agréable. Au delà du labo, j'avais la chance de trouver des amis et des bons conseilleurs au sein du réseau européenne de recherche « Galar FungailII ». J'ai beaucoup profité des multiples formations et rencontres organisé par les différents équipes, et je ressens beaucoup de gratitude envers tout ses participants.

Avant et à la fin de mon séjour à Grignon j'ai effectué quelques manips dans le laboratoire de Domenico Libri à Gif-sur-Yvette. Je tiens à remercier Domenico et toute son équipe pour l'accueil gentille qu'ils m'ont accordé à ces deux reprises, et pour leur aide dans des périodes critiques.

Finalement je voudrais dire un grand merci à ma famille, à mes parents et mes sœurs qui m'ont soutenu à longe distance, et surtout à mon épouse Raquel, pour son soutien dans la vie quotidienne tout au long de ma thèse et sa patience avec moi notamment dans cette dernière période stressante de la rédaction.

Résumé

Candida albicans est une levure pathogène opportuniste qui appartient à la classe phylogénétique des hémiascomycetes comme environ 200 autres espèces du genre « *Candida* ». Seules quelques-unes d'entre elles sont considérées comme pathogènes et encore un plus petit nombre encore est fréquemment retrouvé dans les isolats cliniques (*C. albicans*, *C. dubliniensis*, *C. parapsilosis*, *C. krusei*, *C. glabrata*) (Dujon 2006; Odds, Brown et al. 2006).

C. albicans fait partie de la flore orale et intestinale normale d'environ 50-60% des êtres humains sans pour autant déclencher des infections. Cette levure est donc d'une part un composant normal de notre flore microbienne commensale de différentes zones du corps humain, dont la peau, la bouche, le vagin et le tractus gastro-intestinal. Elle est d'autre part le plus important pathogène fongique chez l'homme et peut dans certains cas causer des maladies diverses.

Les infections diverses qui sont causé par *C. albicans* peuvent être classées en infections superficielles et en infections systémiques. Les infections superficielles, entraînant des lésions de la peau ou de la muqueuse orale ou vaginale, sont très communes, relativement faciles à diagnostiquer et sans danger pour la vie. Par contre, les infections systémiques sont plutôt rares, mais particulièrement dangereuses, et sont responsables d'une mortalité qui s'élève à environ 30 % chez les patients infectés. Une infection systémique passe généralement par plusieurs étapes. *C. albicans* doit d'abord traverser les barrières épithéliales, en franchissant la muqueuse intestinale par exemple. Dans tous les cas, les cellules de *C. albicans* s'attachent d'abord à un épithélium. Ensuite, elles peuvent pénétrer dans les tissus, gagner la voie sanguine et finalement disséminer dans tout l'organisme. Cette étape est facilitée pour les *C. albicans* qui peuvent coloniser un cathéter : ceux-ci constituent un facteur de risque majeur pour les candidoses disséminées. Après traversée des tuniques vasculaires, *C. albicans* va gagner accès à différents organes comme le rein ou le foie qu'elle colonisera.

Des multiples facteurs peuvent déclencher une infection avec *C. albicans*, mais une condition générale pour sa pathogenèse est que le patient possède un système immunitaire affaibli. En général le système immunitaire est capable d'empêcher *C. albicans* d'envahir les tissus ; néanmoins de simples déséquilibres hormonaux, du pH ou de la flore microbienne peuvent provoquer des infections bénignes (Singh 2001; Macphail, Taylor et al. 2002). Les infections avec *C. albicans* peuvent devenir particulièrement dangereuses lorsque le système

immunitaire humain est affaibli ou immature, comme par exemple dans le cas des patients atteints du SIDA, des nouveaux-nés, des patients en service de réanimation ou immunodéprimés suite à une transplantation d'organes ou une chimiothérapie (Wey, Mori et al. 1989; Bustamante 2005).

C'est aussi pour cette raison que *C. albicans* est devenu un organisme modèle pour la recherche médicale. Le génome de *C. albicans* a été le premier génome d'un champignon pathogène qui ait été entièrement séquencé (Jones, Federspiel et al. 2004). L'un des principaux objectifs de la recherche sur *C. albicans* est de comprendre les mécanismes moléculaires de sa pathogenèse. Des nombreux gènes ont déjà été identifiés comme nécessaires dans divers modèles animaux d'infections (Navarro-Garcia, Sanchez et al. 2001; Gow, Brown et al. 2002), et de multiples facteurs de transcription semblent être impliqués dans la régulation de la transcription de ces gènes. Par conséquent, le réseau contrôlant l'expression du pouvoir pathogène est très complexe (Ernst 2000; Liu 2001).

C. albicans est classé comme une levure, c'est-à-dire un champignon unicellulaire. Mais dans certaines conditions, d'autres phénotypes sont prédominants comme des les formes pseudohyphes ou vrais hyphes. La capacité d'alterner entre ces différents phénotypes est aujourd'hui considérée comme essentielle pour la pathogenèse : plusieurs travaux ont démontré que la plasticité morphologique est indispensable à la virulence de *C. albicans* (Odds, F. C., A. J. Brown, et al. 2006). La régulation de la transition entre la forme "levure" et la forme "hyphe" a donc intéressé la recherche depuis longtemps. L'une des raisons pour lesquelles les chercheurs s'intéressent à la réponse au pH de *C. albicans* concerne justement le rôle du pH dans cette transition morphologique. D'ailleurs, *C. albicans* est capable de pousser à des pH très différents, entre pH 2 et pH 10, et on peut imaginer que cette capacité à s'adapter au pH des différentes zones du corps humain peut être cruciale pour le succès de *C. albicans* comme commensal ou pathogène.

Rim101p est connu comme un régulateur majeur de la réponse au pH ambiant, et sa présence est nécessaire pour la transition morphologique pH-dépendante ainsi que pour la croissance à des pH très élevés. Il a été démontré qu'une souche délétée pour *RIM101* présente une virulence clairement diminuée dans un modèle de souris (Davis, Edwards et al. 2000), et que l'expression de plusieurs gènes connus pour être des cibles de Rim101p est nécessaire pour la pathogenèse dans des modèles animaux d'infections (Ghannoum, Spellberg et al. 1995) (De Bernardis, Muhlschlegel et al. 1998) (Soloviev, Fonzi et al. 2007). De plus, Rim101p est indispensable pour la formation des hyphes à pH alcalin (Davis, D. 2003).

La protéine Rim101p est uniquement active à pH alcalin bien que présente mais inactive à pH acide. L'activation à pH alcalin est régulé par une voie de signalisation très conservée chez les ascomycètes et nécessite un clivage C-terminal du précurseur pleine longueur présent à pH acide (Penalva and Arst 2004). En dehors de *C. albicans*, cette voie de signalisation a été étudiée en détail chez plusieurs autres organismes tels qu'*A. nidulans* ou *S. cerevisiae* et *Y. lypolytica* chez les levures.

Plusieurs gènes ont été démontrés comme étant nécessaires pour l'activation de Rim101p, trois d'entre eux codent pour des protéines à domaines transmembranaires, Rim9p, Rim21p et Dfg16p, d'autres codent pour des protéines cytosoliques, comme Rim8p, Rim20p, et la protéase Rim13p. De plus, un rôle important de la voie d'endocytose et des trois complexes ESCRT dans l'activation de Rim101p a été démontré récemment. On pense aujourd'hui qu'à pH alcalin les protéines transmembranaires pourraient transmettre le signal du pH externe au cytosol. En réponse, un complexe de plusieurs protéines s'organise, comprenant Rim101p, Rim20p, la protéase Rim13p et des protéines d'ESCRTIII (Snf7p et Vps20p). La formation de ce complexe est nécessaire pour le clivage de la partie C-terminale de Rim101p. La protéine tronquée peut alors entrer dans le noyau et induire ou réprimer les gènes de la réponse au pH.

Le sujet de ce travail est la description de la réponse transcriptionnelle sous contrôle de Rim101p. La majeure partie des résultats concerne l'identification de gènes cibles de Rim101p par une approche globale. Il existe des voies de signalisation indépendantes de Rim101p qui jouent également un rôle dans la réponse au pH de *C. albicans*. Pour nous affranchir de tout autre effet du pH, et pour nous focaliser uniquement sur la régulation Rim101p-dépendante, nous avons décidé de travailler à un pH constant et acide.

L'idée de cette approche était d'utiliser une forme tronquée de Rim101p, donc constitutivement active, et de placer l'allèle codant cette version nommé *Rim101SL* (<u>S</u>hort <u>L</u>ength) sous contrôle d'un promoteur fort et inductible dans une souche de *C. albicans* délétée pour les deux allèles de *RIM101*.

Ceci devait nous permettre, en théorie, de nous affranchir des effets pH non Rim101p dépendants, observés lors de l'activation physiologique de Rim101p (Davis, D., R. B. Wilson, et al. 2000). L'idée était donc de suivre les changements globaux de la transcription suite à l'expression de Rim101SLp sous contrôle du promoteur Met3 en utilisant des puces à ADN.

Plusieurs expériences ont confirmé la fonctionnalité de la construction dirigeant l'expression de Rim101SLp. Pour exprimer la version tronquée de Rim101p sous contrôle du promoteur Met3, il était important de confirmer que le promoteur était placé immédiatement

en amont de la séquence à transcrire. Deux codons initiateurs possibles, espacés de 174 paires de bases, étaient proposés dans les bases de données des séquences. En catographiant l'extrémité 5' des transcrits naturels par la technologie « RACE », nous avons pu confirmer que le premier codon initiateur était bien transcrit. C'est le codon qui a été utilisé pour la construction de *RIM101SL*.

Il a été également démontré que l'expression de Rim101SLp pouvait complémenter le défaut de croissance observé à pH alcalin sur une souche delétée pour *RIM101* et ce uniquement dans des conditions de carence à la méthionine et la cystéine, lorsque le promoteur Met3 était activé. Une expérience de quantification par PCR en temps réel a confirmé qu'en présence de ces acides aminés l'allèle *RIM101SL* n'était pas transcrit dans un milieu SC à pH 5,5, mais que sa transcription était fortement induite en l'absence de ces acides aminés. De plus nous avons pu détecter une forte augmentation des transcrits de *PHR1*, un gène cible connu de Rim101p normalement exprimé à pH alcalin (lorsque Rim101p est actif), indiquant que l'expression de l'allèle tronqué de Rim101p pouvait effectivement simuler la régulation Rim101p-dépendante d'une façon pH-indépendante. La même expérience a été utilisée pour définir les points clefs de la cinétique d'induction de la transcription de *RIM101SL* après application de la carence en acides aminés soufrés. Ces points ont été retenus pour les suivis cinétiques sur puces à ADN.

Deux cinétiques ont été effectuées : des échantillons ont été pris 0, 15, 30, 60 et 90 min après l'induction de la transcription de *RIM101SL*. Afin de disposer d'une référence commune du niveau d'expression pour toute la cinétique, un mélange des ARN extraits de tous les échantillons prélevés, lors d'une troisième cinétique, a été utilisé comme référence, ce qui permet d'utiliser le niveau d'expression moyen de chaque gène comme référence. Pour chaque point de temps de chaque cinétique, deux séries doubles de puces (soit 10 puces chacune) ont été produites. Pour la première série, l'échantillon a été marqué avec le chromophore Cy5 et la référence avec le chromophore Cy3 ; pour la deuxième série, le marquage a été inversé afin de compenser les biais de chromophores. De surcroît, ces 20 puces ont été complétées par deux séries de puces supplémentaires qui ont été hybridées avec un mélange de référence marquée par Cy5 et Cy3, ce qui nous permettait une correction additionnelles de ces effets (Dye Swap) en normalisant toutes les séries contre ces deux là.

Cette expérience nous a donné les profils d'expression des 6.000 gènes environ de *C. albicans* en réponse à l'activitation de Rim101SLp. Pour induire la transcription de *RIM101SL* sous contrôle du promoteur Met3, il nous a été nécessaire de transférer les cellules d'un milieu avec méthionine et cystéine à un milieu dépourvu de ces acides aminés. Par

conséquent, on s'attendait à voir varier l'expression non seulement des gènes dépendants de Rim101p, mais également des gènes qui jouent un rôle dans l'adaptation à ces nouvelles conditions, notamment des gènes codants pour des protéines du métabolisme de soufre. Effectivement, plusieurs gènes codant de telles fonctions se trouvaient parmi les gènes les plus fortement induits. Il était indispensable d'effectuer des expériences supplémentaires afin de pouvoir exclure ces gènes de l'analyse.

Pour différencier les gènes influencés par Rim101SLp de ceux qui étaient affectés par la carence en méthionine et cystéine, nous avons effectué des expériences de contrôle supplémentaires. Une souche témoin, isogénique, mais délétée pour *rim101-/-*, a été traitée de la même manière que la souche exprimant Rim101SLp, pour identifier ces gènes faux positifs. Pour cette expérience (dite CTRL), qui devait être utilisée uniquement pour identifier les faux positifs sans donner plus d'informations sur la régulation par Rim101SLp, nous avons analysé uniquement trois temps avec les puces à ADN : 0-15 min et 0-90 min, car les plus fortes régulations étaient apparentes à ces moments pour la souche exprimant Rim101SLp. En comparant les changements transcriptionels entre l'expériences CTRL et la cinétique complète, nous avons filtré les données : tous les gènes montrant une évolution similaire entre les expériences avec et sans expression de Rim101SLp ont été considérés comme faux positifs. Nous nous sommes ensuite focalisé sur 609 gènes qui étaient régulés dans la cinétique, mais qui n'étaient pas régulés ou régulés d'une façon différente dans l'expérience CTRL.

Les données normalisées de ces gènes ont ensuite été analysées avec le logiciel SAM (Significance Analysis of Microarray data). Ceci nous a permis d'identifier 133 gènes régulés par Rim101SLp qui présentaient une modification d'expression significative. Ces 133 gènes possédaient des profils différents d'expression. Nous avons ensuite tenté de regrouper les gènes avec profil d'expression similaire pour identifier des groupes des gènes corégulés par Rim101SLp.

Nous avons utilisé une classification hiérarchique de tous les profils d'expression pour effectuer ce tri. Nous avons ainsi pu regrouper les 133 gènes en cinq types de profils distincts: les gènes qui étaient immédiatement induits ou réprimés par Rim101p, les gènes induits ou réprimés progressivement, et un dernier group de gènes transitoirement induits. En tout, près des deux tiers des gènes étaient réprimés par Rim101p, ce qui pourrait indiquer que Rim101p agit majoritairement comme répresseur transcriptionnel.

Pour confirmer la fiabilité des résultats obtenus par puces à ADN, nous avons décidé de valider les résultats obtenus pour 20 gènes par PCR quantitative en temps réel. Ces gènes

ont été choisis selon plusieurs critères en fonction de leur régulation et de leurs caractéristiques de séquences. Nous avons choisi des gènes dont le promoteur portait ou non des motifs de liaison pour Rim101p et des gènes codant ou non pour des domaines transmembranaires ou des peptides signaux indiquant une localisation de la protéine à la surface cellulaire. Nous nous intéressions en effet particulièrement aux gènes susceptibles de coder des protéines de surface, cette localisation pouvant indiquer un rôle dans l'interaction levure-hôte et la pathogenèse de *C. albicans*. Nous avons également retenu quelques gènes sans fonction connue, et même si la plupart des gènes étaient réprimés au cours de la cinétique, nous avons sélectionné au moins un membre de chaque classe des gènes induits pour la confirmation. Pour ces gènes, les résultats obtenus par puce à ADN ont été confirmés en deux étapes.

Dans un premier temps, la qualité des résultats obtenus sur puces a été vérifiée en analysant par PCR quantitative en temps réel les échantillons des temps 0 min, 15 min et 90 min qui avaient été utilisés pour les puces à ADN. Les résultats de ces expériences montrent une bonne reproductibilité générale des observations faites sur puce à ADN, et indiquent très souvent même une régulation plus forte que celle observée par microarray.

Dans un second temps, et afin de savoir si les effets observés lors de la cinétique étaient physiologiquement représentatifs de l'activation de Rim101p, nous avons analysé les transcrits de ces vingt gènes dans une souche de référence portant le gène *RIM101* intact, en phase de croissance exponentielle à pH acide ou alcalin. Les niveaux de transcrits ont été mesurés par PCR quantitative. Malgré les différences de conditions expérimentales entre cette expérience et la cinétique, les résultats montrent une transcription pH-dépendante pour un grand nombre des gènes analysés : nous avons retrouvé la plupart des gènes induits par Rim101SLp parmi les gènes induits à pH alcalin, et la majorité des gènes réprimés par Rim101SLp dans la cinétique l'est également à pH alcalin. Bien qu'il y ait aussi quelques exceptions, des gènes qui sont non-régulés ou régulés inversement aux attentes, ces résultats indiquent que la majorité des gènes identifiés sur puce sont bien soumis à une régulation pH-dépendante par Rim101p.

La séquence CCAAG (avec souvent 3 A supplémentaires en 3') a été proposée pour être le motif de liaison reconnu par Rim101p sur les promoteurs *C. albicans* (Ramon and Fonzi 2003) . Cette séquence est légèrement différente de la séquence « GCCARG » reconnue par PacCp, l'orthologue de Rim101p chez *A. nidulans*. Nous avons analysé les promoteurs des gènes identifiés comme régulés par Rim101p pour la présence de ces séquences, et nous avons comparé ces résultats avec l'abondance de ces motifs dans les promoteurs du génome

entière. Ce motif et ses variantes étaient enrichis de façon significative dans les promoteurs des 133 gènes, ce qui suggère une présence importante de cibles directes de Rim101p parmi les gènes identifiés. Une analyse plus approfondie nous a permis de trouver une séquence élargie « GCCARGAA » qui est également surreprésentée dans les promoteurs des gènes Rim101SLp-régulés et qui semble mieux définir le motif reconnu par Rim101p, un résultat confirmé par une autre équipe (Baek, Martin et al. 2006).

L'une des questions les plus intéressantes concerne les fonctions cellulaires qui sont particulièrement affectées par l'activité de Rim101p. Nous avons comparé l'abondance des fonctions prédites pour les 133 gènes identifiés avec leur abondance globale dans le génome. Les résultats de cette recherche indiquent que Rim101p joue un rôle important dans la régulation de plusieurs classes fonctionnelles importantes : le métabolisme, la biogénèse des compartiments cellulaires, et les réponses cellulaires à divers stress. Tandis que la présence de beaucoup de gènes du métabolisme était plutôt inattendue, il est connu que Rim101p joue un rôle dans la réponse aux stress environnementaux. Parmi les gènes de biogenèse se trouvait un grand nombre de gènes impliqués dans l'assemblage de la paroi, ce qui est à mettre en relation avec le rôle de Rim101p comme modulateur de la morphogénèse.

Finalement, nous avons observé une forte influence de Rim101p sur l'expression d'une grande famille de gènes spécifiques de *C. albicans*, les gènes *ALS*, qui codent pour des protéines pariétales à ancre GPI (Glycosyl-Phosphatidyl-Inositol). Quatre des huit gènes *ALS* se retrouvaient parmi les 133 gènes régulés par Rim101p dans la cinétique suivie par puces à ADN. La deuxième partie du travail a consisté en une analyse spécifique de la régulation pH-dépendante de ces gènes et du rôle de Rim101p dans leur régulation.

Les protéines de paroi codées par ces gènes ont des fonctions importantes liées à la virulence. L'abréviation *ALS* signifie “Agglutinin-Like Sequence” et se réfère à la similarité des protéines Als avec l'alpha-agglutinine de *S. cerevisiae*. Cette fonction dans l'agglutination des cellules est conservée chez les protéines Als, mais celles-ci jouent de plus u rôle tout à fait intéressant dans l'adhésion des cellules à des surfaces différentes, telles que l'épithélium, l'endothélium ou le plastique (cathéters hospitaliers) (Sheppard. D. C. M. R. Yeaman. et al. 2004). Ces fonctions contribuent également à la capacité qu'a *C. albicans* de former des biofilms, et des mutants délétés pour *ALS1* ou *ALS3* sont très limités dans leur capacité à former des biofilms. Une dernière fonction importante pour la virulence de *C. albicans* a été démontrée récemment : deux protéines Als au moins sont nécessaires pour l'induction de l'endocytose, facilitant l'invasion des tissus (Phan, Myers et al. 2007).

La famille des gènes *ALS* comporte 8 membres, tous codant pour des protéines à ancre GPI. Une caractéristique de cette famille est la présence de larges séquences répétées de 108 paires de bases au milieu du gène. A cause de la forte ressemblance de ces gènes, les résultats obtenus par microarray n'étaient pas toujours spécifiques d'un seul gène *ALS* rendant l'interprétation difficile, voire impossible.

Nous avons obtenu des résultats gène-spécifiques pour l'expression des gènes *ALS* en fonction du pH par PCR quantitative en temps réel. Ces données nous indiquent que 6 sur 8 gènes *ALS* sont exprimés dans les conditions utilisées.

Les trois gènes les plus exprimés, *ALS1*, *ALS4* et *ALS9*, sont tous régulé par le pH, tandis que, parmi les gènes plus faiblement exprimés, seul *ALS2* montrait une légère répression à pH alcalin. *ALS1* est le seul gène *ALS* qui soit fortement induit à pH alcalin par rapport au pH acide, et les trois autres gènes étaient réprimés. Nous avons confirmé ces observations en montrant que l'expression d'*ALS1* était beaucoup plus faible à pH8 dans une souche délétée pour *RIM101* que dans une souche sauvage, et que symétriquement l'expression d'*ALS4*, réprimée dans la souche sauvage à pH alcalin, était complètement deréprimée à pH alcalin dans la souche *rim101*$^{-/-}$. Pour *ALS9* la situation apparaît plus compliquée, car bien que nous ayons pu observer une derépression à pH alcalin d'*ALS9* dans la souche *rim101*$^{-/-}$ par rapport à la souche sauvage, ce gène semble être encore réprimé à pH alcalin d'une façon Rim101p-indépendante. Ceci suggère que des voies de signalisation indépendantes de Rim101p seraient impliquées dans la régulation d'*ALS9* par le pH. Finalement, nous avons observé que la répression à pH alcalin d'*ALS2* était conservée dans la souche *rim101*$^{-/-}$, ce qui indique que ce gène est régulé par le pH indépendamment de Rim101p.

Nous avons ensuite essayé de mieux comprendre la régulation d'*ALS1* et *ALS4*. Pour cela des fusions de leurs promoteurs avec le gène rapporteur « LacZ » de *S. thermophilus* (optimisé pour le biais de codon de *C. albicans*) ont été construites et intégrése dans le génome de *C. albicans*. Dans un premier temps, nous avons placé un fragment du promoteur d'*ALS1* de 1000 paires de bases devant LacZ, mais nous n'avons pas pu détecter une activité beta-galactosidase à pH acide ou alcalin. Nous avons ensuite mis LacZ sous contrôle d'un fragment de 2000 paires de bases, et les clones obtenus ont montré une faible activité beta-galactosidase. L'activité beta-galactosidase sous contrôle de ce promoteur était environ 4x plus forte à pH alcalin qu'à pH acide, en accord avec nos observations faits via PCR quantitative pour ce gène. Ces résultats indiquent que la région nécessaire à l'expression et à la régulation d'*ALS1* est très éloignée du codon initiateur. De plus, plusieurs motifs de liaison

prédits pour Rim101p se trouvent dans cette région éloignée et ils sont plus complets que ceux qui sont plus proches.

Les résultats pour le promoteur d'*ALS4* indiquaient déjà une activité promotrice de 1000 paires de bases, mais néanmoins une activité plus forte a été obtenue avec la région de 2000 paires de bases. L'activité beta-galactosidase était beaucoup plus forte à pH4 qu'à pH8 pour les deux constructions, ce qui confirme qu'*ALS4* est un gène réprimé à pH alcalin. Ces données suggèrent que des régions de plus de 1000 paires de bases en amont de l'ATG sont également importantes pour le niveau d'expression d'*ALS4*, mais qu'elles ne sont pas indispensables pour la régulation en fonction du pH. Bien que ces résultats aient été bien en accord avec nos observations précédentes, nous avons obtenus des résultats beaucoup moins clairs avec les constructions contrôles utilisés pour tester l'expression de LacZ : les promoteurs des gènes *PHR1* et *PHR2* ne régulaient pas LacZ comme attendu. Le promoteur de *PHR1* semblait réprimer au lieu d'induire, et celui de *PHR2* n'était quasi pas régulé par de pH. En parallèle, une autre équipe a publié une étude très similaire sur la régulation de *PHR2* en utilisant un système rapporteur semblable. Elle a décrit des activités plus de 40 fois supérieures à pH acide qu'à pH alcalin, et a pu démontrer que les sites de liaison de Rim101p étaient indispensables pour cette régulation (Baek, Martin et al. 2006). Nos résultats avec *PHR1* et *PHR2* remettant en question les résultats obtenus avec les autres fusions, nous avons décidé d'arrêter ce projet et de mettre en place une approche différente pour analyser l'action de Rim101p sur ses promoteurs cibles.

Dans ce dernier projet, nous avons essayé de mettre en évidence par des approches d'immunoprécipitation de la chromatine (ChIP) que les promoteurs étient des cibles directes de Rim101p. Nous avons utilisé une souche qui exprimait une version de Rim101p étiquetée avec l'épitope V5. La chromatine couplée *in vivo* par Rim101p a été précipitée, purifiés et l'ADN a été quantifié en PCR en temps réel avec des oligonucléotides permettant une amplification des séquences autour des sites probables de Rim101p. Nous avons observé un plus grand nombre des promoteurs cibles dans les échantillons pris à pH alcalin que dans ceux pris à pH acides, conformément aux attentes pour des promoteurs régulés à pH alcalin par Rim101p. Toutefois, ces résultats n'étaient pas très solides, car la reproductibilité était faible et nous avons occasionnellement observé un enrichissement similaire pour des promoteurs non-régulés utilisés comme contrôles dans ces expériences.

Pour conclure nous avons étudié la contribution immédiate de Rim101p à la régulation de la réponse au pH régulé. Une liste de 133 gènes cibles probables de Rim101p a été proposée et leur régulation en fonction du temps après l'activation de Rim101p a été décrite.

Rim101p semble affecter particulièrement la transcription de plusieurs gènes *ALS* qui codent pour des protéines pariétales avec des fonctions importantes dans l'interaction entre *C. albicans* et son hôte.

Ce travail ouvre plusieurs pistes qui seraient intéressant à suivre :

L'un des compléments les plus immédiats à apporter serait de réaliser des expériences de retard sur gel pour valider au moins *in vitro* que les gènes *ALS* et d'autres gènes sont effectivement des cibles directes de Rim101p. Il serait également intéressant de regarder si un mutant deleté pour *RIM101* ou *ALS1* est affecté pour son adhérence dans des modèles d'adhérence à pH alcalin.

L'un des gènes régulés transitoirement par Rim101SLp dans notre cinétique était *EFG1*, un gène codant pour un facteur de transcription connu comme l'un des régulateurs clés de la morphogénèse chez *C. albicans*. Il a été postulé par d'autres groupes que Rim101p agissait en amont de *EFG1* dans la morphogénèse (El Barkani, Kurzai et al. 2000), mais à notre connaissance ceci est la première indication qu'*EFG1* serait directement régulé par Rim101p. Il serait donc envisageable de confirmer et de mieux caractériser cette régulation, par exemple par une autre cinétique suivie par PCR quantitative ou en construisant une fusion avec un gène rapporteur, qui pourrait servir également servir à étudier la liaison de Rim101p sur son promoteur.

Puisque les régulations détectées dans la cinétique effectuée indiquaient quelque fois une régulation différente de celles observées dans une souche sauvage cultivée à différents pH, il serait intéressant de savoir s'il y a des cofacteurs qui se lient à pH alcalin avec Rim101p pour former un complexe nécessaire à la régulation pH-dépendante de ces gènes. On pourrait isoler les complexes formés par Rim101p à pH alcalin par exemple par TAP-Tag, les séparer sur gel et identifier des interactants par spectroscopie de masse.

Finalement, il serait envisageable de connaître les gènes cibles des réponses au pH autres que celle contrôlée par Rim101p pour avoir une vue plus complète de cette régulation. Le produit du gène *MDS3* agit dans une réponse au pH Rim101p-indépendant (Davis. D. A.. V. M. Bruno. et al. 2002). Une approche similaire à celle présentée ici par puce à ADN pourrait être mise en place et et permettre une analyse comparative des deux voies de signalisation. Si on connaissait les gènes sous contrôle de cette réponse, on pourrait caractériser les contributions spécifiques de chacune de ces voies et éventuellement trouver des gènes clés de la réponse au pH qui sont contrôlés par les deux réponses.

Baek, Y. U., S. J. Martin, et al. (2006). "Evidence for novel pH-dependent regulation of Candida albicans Rim101, a direct transcriptional repressor of the cell wall beta-glycosidase Phr2." Eukaryot Cell **5**(9): 1550-9.

Bustamante, C. I. (2005). "Treatment of Candida infection: a view from the trenches!" Curr Opin Infect Dis **18**(6): 490-5.

Davis. D. (2003). "Adaptation to environmental pH in Candida albicans and its relation to pathogenesis." Curr Genet **44**(1): 1-7.

Davis, D., J. E. Edwards, Jr., et al. (2000). "Candida albicans RIM101 pH response pathway is required for host-pathogen interactions." Infect Immun **68**(10): 5953-9.

Davis. D. A.. V. M. Bruno. et al. (2002). "Candida albicans Mds3p. a conserved regulator of pH responses and virulence identified through insertional mutagenesis." Genetics **162**(4): 1573-81.

De Bernardis, F., F. A. Muhlschlegel, et al. (1998). "The pH of the host niche controls gene expression in and virulence of Candida albicans." Infect Immun **66**(7): 3317-25.

Dujon, B. (2006). "Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution." Trends Genet **22**(7): 375-87.

El Barkani, A., O. Kurzai, et al. (2000). "Dominant active alleles of RIM101 (PRR2) bypass the pH restriction on filamentation of Candida albicans." Mol Cell Biol **20**(13): 4635-47.

Ernst, J. F. (2000). "Transcription factors in Candida albicans - environmental control of morphogenesis." Microbiology **146 (Pt 8)**: 1763-74.

Ghannoum, M. A., B. Spellberg, et al. (1995). "Reduced virulence of Candida albicans PHR1 mutants." Infect Immun **63**(11): 4528-30.

Gow, N. A., A. J. Brown, et al. (2002). "Fungal morphogenesis and host invasion." Curr Opin Microbiol **5**(4): 366-71.

Jones, T., N. A. Federspiel, et al. (2004). "The diploid genome sequence of Candida albicans." Proc Natl Acad Sci U S A **101**(19): 7329-34.

Liu, H. (2001). "Transcriptional control of dimorphism in Candida albicans." Curr Opin Microbiol **4**(6): 728-35.

Macphail, G. L., G. D. Taylor, et al. (2002). "Epidemiology, treatment and outcome of candidemia: a five-year review at three Canadian hospitals." Mycoses **45**(5-6): 141-5.

Navarro-Garcia, F., M. Sanchez, et al. (2001). "Virulence genes in the pathogenic yeast Candida albicans." FEMS Microbiol Rev **25**(2): 245-68.

Odds, F. C., A. J. Brown, et al. (2006). "Toward a molecular understanding of Candida albicans virulence." Chapter 22 in "Molecular Principles of Fungal Pathogenesis" (Eds. Heitman, J., Filler, S., Mitchell, A.) ASM Press.

Penalva, M. A. and H. N. Arst, Jr. (2004). "Recent advances in the characterization of ambient pH regulation of gene expression in filamentous fungi and yeasts." Annu Rev Microbiol **58**: 425-51.

Phan, Q. T., C. L. Myers, et al. (2007). "Als3 Is a Candida albicans Invasin That Binds to Cadherins and Induces Endocytosis by Host Cells." PLoS Biol **5**(3): e64.

Ramon, A. M. and W. A. Fonzi (2003). "Diverged binding specificity of Rim101p, the Candida albicans ortholog of PacC." Eukaryot Cell **2**(4): 718-28.

Singh, N. (2001). "Changing spectrum of invasive candidiasis and its therapeutic implications." Clin Microbiol Infect **7 Suppl 2**: 1-7.

Soloviev, D. A., W. A. Fonzi, et al. (2007). "Identification of pH-regulated antigen 1 released from Candida albicans as the major ligand for leukocyte integrin alphaMbeta2." J Immunol **178**(4): 2038-46.

Abrevations

ALS: Agglutinin-Like Sequence

CGD: Candida Genome Database

ChIP: Chromatin Immune Precipitation

CTRL: ConTRoL

ESCRT: Endosomal Sorting Complex Required for Transport

GABA:γ-Amino-Butyric Acid

GFP: Green Fluorescent Protein

GlcNAc: N-Acetyl-Glucosamine

GO: Gene Ontology

GPI: Glycosyl-Phosphatidyl-Inositol

HVR: Hyper-Variable Region

IP: Immune Precipitation

LOWESS: LOcally WEighted linear regreSSion

MRS: Major Repeat Sequence

MVB: MultiVesicular Body:

NHS: N-Hydroxy-Succinylimidyl

ORF: Open Reading Frame

PHR : pH Responsive

qPCR: quantitative Polymerase Chain Reaction

RIM101SL: Rim101 Short Length

RIM: Regulator of Inducer of Meiosis

SAM: Significance Analysis of Microarray data

SDS : Sodium Dodecyl Sulfate

SOM: Self-Organizing Maps

TC: Time Course

Index

1 Introduction 1

1.1 General presentation of *C. albicans* 1

1.1.1 The target population and the different infection types 2
1.1.2 Antifungals and drug resistance 4

1.2 Molecular Biology of *C. albicans* 5

1.2.1 The genome of *C. albicans* 6
1.2.2 Molecular Genetics 9
1.2.2.1 Gene disruption 11
.... 12
1.2.2.2 Reporter genes 13
1.2.2.3 Global Transcription Analysis 14
- *Basic principles and different array types 15*
- *Fluorescense labeling and dye effects 17*
- *Experimental Setups 18*
- *Data normalization and analysis 20*
- *C. albicans arrays 23*

1.3 Virulence factors 25

1.3.1 Cell wall and host-pathogen interaction 26
1.3.1.1 Composition and architecture of cell wall 26
1.3.1.2 Cell wall proteins 27
1.3.1.3 Adhesion and recognition of host cell 28
- *The ALS gene family 28*
- *ALS mutant phenotypes 31*
- *Regulation of ALS gene expression 32*

1.3.2 Secreted Hydrolytic Enzymes 33
1.3.2.1 Secreted Aspartyl proteinases 34
1.3.2.2 Phospholipases 34
1.3.2.3 Lipases 35
1.3.3 Yeast-to-hyphae transition and its complex regulation 35
1.3.3.1 The role of hyphae formation in *C. albicans* virulence 35
1.3.3.2 The regulation of the Yeast - to - hyphae transition 36

1.4 The conserved fungal pH signaling pathway 38

1.4.1 *Aspergillus nidulans* 39
1.4.2 *Saccharomyces cerevisiae* 42
1.4.3 *Yarrowia lipolytica* 47
1.4.4 *Candida albicans* 49
1.4.4.1 A link between the fungal pH response and virulence 49
1.4.4.2 The role of Rim101p in *C. albicans* virulence 49
1.4.4.3 Regulation of Rim101p activity 50

1.4.4.4 Functionality of Rim101p 51
1.4.4.5 Genes under the control of Rim101p 54

1.5 Aims of the project 62

2 Materials and Methods 63

2.1 General 63

2.1.1 Strains used 63
2.1.2 Culture media and phenotypic tests 65
2.1.3 *ALS* primers selection 67
2.1.4 Primers used 68
2.1.5 Genomic DNA purification 69
2.1.6 Plasmid construction 72
2.1.7 Identification of the *RIM101* start codon by the RACE technique 74

2.2 Microarray time course 74

2.2.1 Genomic microarrays 74
2.2.2 Experimental Setup 75
2.2.3 RNA purification and cDNA labeling 75
2.2.4 Hybridisation and Scanning 77
2.2.5 Identification of Rim101p target genes 78
2.2.6 Real time quantitative PCR confirmation 79
2.2.7 Statistical tests 79

2.3 Chromatin Immunoprecipitation (ChIP) 80

2.3.1 Protein extraction 80
2.3.2 Western Blots 80
2.3.3 Chromatin extraction 80
2.3.4 Immunoprecipitation 81

2.4 β-galactosidase tests 82

3 Results and Discussion 83

3.1 Microarray to identify Rim101p target genes 83

3.1.1 Strain Construction and Microarray Time Course 83
3.1.2 Identification of the transcriptional start of the *RIM101* gene 85
3.1.3 Microarray Control Experiments and False Positive Filter 87
3.1.4 Identification and Clustering of Rim101p target genes 89
3.1.5 Sequence Analysis 91
3.1.5.1 Signal Peptides and Transmembrane Motifs 95
3.1.5.2 Rim101p Promoter Motifs 95
3.1.6 Confirmation of Microarray Results 97
3.1.6.1 Quantitative PCR on original cDNA 99

3.1.6.2 Quantitative PCR using wildtype samples from cultures grown at acidic and alkaline pH 101

3.2 Characterization of pH-dependent *ALS* gene regulation 103

3.2.1 *ALS* Sequence Analysis 103
3.2.2 Role of Rim101p in *ALS* gene regulation 105
3.2.3 β-Galactosidase reporter assays 107

3.3 Chromatin Immunoprecipitation to identify direct Rim101p targets ... 114

3.3.1 Experimental Setup ChIP 114
3.3.2 Results and Discussion ChIP 117
3.3.3 Conclusions ChIP 121

4 Conclusions and Perspectives 122

4.1 Conclusions 122

4.2 Perspectives 127

5 Appendix 129

6 References 165

1 Introduction

1.1 General presentation of *C. albicans*

The genus *Candida* comprises almost 200 yeasts species (Odds, Brown et al. 2006) and belongs to the class of hemiascomycetes, a phylogenetic group shared by distant neighbors such as *Saccharomyces cerevisiae*, *Candida glabrata*, *Kluyveromyces lactis* and *Yarrowia lypolitica* (Dujon 2006)(see also Figure 1).

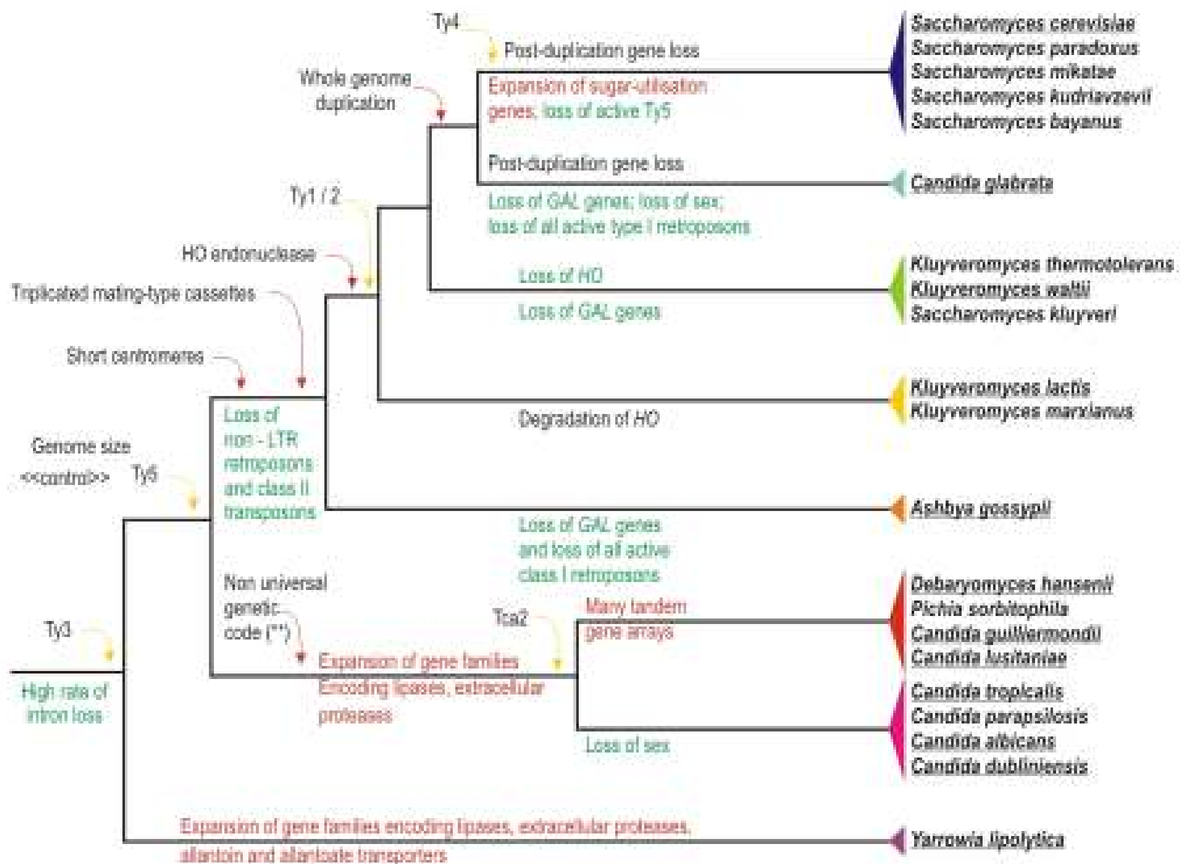


Figure 1: *Phylogenetic tree of hemiascomycetous yeasts adapted from Dujon et al. (Dujon 2006): Important evolutionary events are indicated at the origin of branches. Some important species of the different clades are noted on the right side of the diagram, species with available complete genome sequences are underlined.*

Among *Candida* species, about 65 % cannot grow at human body temperature (Schauer and Hanschke 1999), and only a handful is regularly encountered in clinical studies as commensal microorganisms with opportunistic pathogenic capacities: *C. dubliniensis*, *C. parapsilosis*, *C. krusei*, *C. glabrata* and *C. albicans* (Odds, Brown et al. 2006). *C. albicans* seems to be more sensitive to antifungal drugs than some other *Candida* species, in particular *C. krusei* and the more distantly related *C. glabrata* (Table 1). The extensive use of antifungal

drugs might also be the reason for a recent shift towards these species in clinical isolates (Trick, Fridkin et al. 2002; Wisplinghoff, Seifert et al. 2003), but *C. albicans* is still considered as one of the most important human fungal pathogens.

	Polyenes	Azoles		Echinocandines		
***Candida* species**	**Amphotericin B**	**Fluconazole**	**Voriconazole**	**Anidulafungin**	**Caspofungin**	**Micafungin**
C. albicans	0.25-1.0	0.5- 2.0	0.01-0.06	0.02-0.25	0.5	0.03
C. glabrata	0.50-2.0	0.5-64.0	0.50-2.00	0.06-0.50	1.0	0.06
C. krusei	0.50-2.0	32.0-64.0	0.50-1.00	0.03-1.00	2.0	0.25
C lusitania	0.50-1.0	0.5- 2.0	0.03-0.06	0.25-2.00	2.0	2.00
C. parapsilosis	0.50-1.0	0.5- 2.0	0.03-0.12	2.00-8.00	2.0	2.00
C. tropicalis	0.50-1.0	4.0-16.0	0.12-2.00	0.06-1.00	1.0	0.06

Table 1*: Susceptibility of various Candida species against different antifungal drugs. Minimum inhibitory concentrations are given as MIC_{90} values (minimal concentration necessary to inhibit growth to 90%) in µg/mL; data was taken from various studies summarized at http://www.formularyjournal.com/formulary/article/articleDetail.jsp?id=364693 . C. albicans is clearly less resistant against the different antifungal drugs than the other analyzed species, which might be one the reason of the decreasing relative abundance of C. albicans in clinical isolates.*

In addition to *C. albicans*, other important opportunistic fungal pathogens that are also frequently identified are *Aspergillus fumigatus*, *Histoplasma capsulatum, Cryptococcus neoformans* and *Coccidioides immitis*. *C. albicans* is a member of the normal commensal flora of skin, oral cavity, gastrointestinal tract and vagina of warm-blooded animals (Tanghe, Carbrey et al. 2005). Even if it has been occasionally isolated from different environments such as lemons (Newton-John, Wise et al. 1984), sand, sea water (Anderson 1979) and air (Calvo, Guarro et al. 1980; Wolf, Polacheck et al. 2000), it seems that, in contrast to other *Candida* species, its presence in such places is unusual (Tanghe, Carbrey et al. 2005).

1.1.1 The target population and the different infection types

As already mentioned, *C. albicans* is an opportunistic pathogen. It is found in the oral and gut mucosae in approximately 50-60 % of healthy humans (Glick and Siegel 1999) without developing a pathology, which corresponds to the definition of a commensal microorganism. Nevertheless *C. albicans* can also be the agent of different types of infections, reaching from relatively harmless superficial infections like vaginal candidosis or oral thrush of newborns to life-threatening blood stream infections. The human immune system is

normally able to limit the abundance of *C. albicans*, keeping a healthy equilibrium in the commensal flora of the human mucosa.

However, under certain conditions healthy individuals may become prone to irritating superficial infections. These occur generally as a consequence of imbalances in hormone levels, pH or in the normal microbiological flora (often following a treatment with antibiotics) (Singh 2001; Macphail, Taylor et al. 2002). The vast majority (ca. 75 %) of all women is infected by vaginal thrush at least once in their life, and between 5 and 10 % even suffer from recurrent vulvovaginal candidiasis (at least three infections per year;(Sobel 1992)). Other comparably harmless types of superficial infections are oral thrush and cutaneous lesions, which are associated with a weakened immune system and for example commonly encountered in neonates at the intensive care unit (Reef, Lasker et al. 1998).

Intriguingly, when the immune system is seriously weakened, *C. albicans* is able to become dominant in the mucosa, colonize different zones of the human body and cause severe infections. Persons at risk include newborns, HIV-positive patients, and patients in the intensive care unit, transplant recipients and patients that were subjected to chemotherapy or simply to broad-spectrum antibiotic treatments (Wey, Mori et al. 1989). Another possible risk factor are hospital devices like catheters where *C. albicans* can colonize as biofilms and gain a direct access to the human bloodstream (Ramage, Martinez et al. 2006). While superficial forms of *Candida* infections are in general early recognized and relatively harmless, the infections of deeper tissues (kidney, liver, spleen, heart, brain and lungs) in severely immunocompromised patients where *Candida* gains access to the bloodstream (known as Candidemia) are often difficult to diagnose and life-threatening (Wey, Mori et al. 1989; Schelenz and Gransden 2003). Although these bloodstream infections are relatively seldom compared to superficial infections, *Candida* species rank second (behind coagulase negative staphylococci) among the nosocomial bloodstream isolates (Bustamante 2005).

1.1.2 Antifungals and drug resistance

The antifungal drugs used in the treatment of *C. albicans* infections belong to four main functional classes: β-1,3-glucan synthase inhibitors (echinocandins), ergosterol ligands (polyenes like amphotericin B or nystatin), ergosterol biosynthesis inhibitors (thiocarbamates, morpholines, azoles and allylamines) and nucleic acid synthesis inhibitors (5-Flucocytosine). The most important drug target, ergosterol, is the main sterol in fungal membranes, in contrast to mammalian membranes, where cholesterol is predominant. The various antifungal drugs do not have the same mode of action. Echinocandins are directed immediately against targets in the fungal cell wall whereas polyenes bind to ergosterol and lead to cell membrane collapse: both function as fungicides at high concentration. The two other classes block important metabolic functions, but only have a fungistatic effect.

Besides the use of a single antifungal drug, the combination of two antifungals can enhance significantly the efficiency of the treatment. For instance, 5-FC is generally used in combination with amphotericin B or with azole drugs, because resistance towards 5-FC alone is easily selected *in vivo*. There are also examples of strains that gained a cross-resistance towards different azoles (White, Holleman et al. 2002). Resistance to amphotericin B has already been found in clinical isolates and laboratory strains (Sanglard, Ischer et al. 2003). Strains that are resistant to echinocandins have not been isolated so far, probably because these drugs have only recently become available for clinical use. Nevertheless, a mutant in β-1,3-glucan synthase from S. *cerevisiae* has been shown to be resistant against echinocandins, indicating that point mutations of the target could allow the development of resistance also in clinical strains (Douglas, Marrinan et al. 1994).

Taken together, until today only a handful of powerful antifungal drugs are available which are extensively used and directed against very few targets, mainly against ergosterol and its biosynthesis. The alarming multitude of (cross-) resistance mechanisms that have been observed in both clinical and laboratory strains (Hospenthal, Murray et al. 2004), only underscores the necessity to find new drug targets in fungal cells.

Although the clinical importance of *C. albicans* is decreasing, the relatively high sensibility of *C. albicans* against all common antifungal drugs (see Table 1) makes it a good model to discover other possible targets for new antifungal drugs, and the rapid progress in the development of powerful molecular biological techniques ((Magee, Gale et al. 2003); see 1.2) for *C. albicans* reinforces its current importance as a model organism for fungal pathogenesis.

1.2 Molecular Biology of *C. albicans*

The choice of *C. albicans* as a model organism was mainly driven by its importance as a human fungal pathogen, and not by its suitability for molecular genetics. Consequently, the research community had to solve some important problems because of the peculiarities of *C. albicans* biology.

One of the biggest challenges is the life cycle of *C. albicans*. Up to date, although there is evidence for mating, no stable haploid form of *C. albicans* has been observed (Soll, Lockhart et al. 2003; Bennett and Johnson 2005), so that researchers are obliged to analyze both alleles of the gene of interest. In addition, the elevated rate of sequence polymorphisms between the two alleles of a gene suggests that there might be functional differences between the two copies of a gene, which further complicates the interpretation of results.

Another particularity of *C. albicans* and some of its closest phylogenetic neighbors (including *Debaryomyces hansenii* and other *Candida* species; (Dujon, Sherman et al. 2004), see Figure 1) is a differential translation of the codon "CUG", which decodes serine rather than leucine (Santos and Tuite 1995). This codon is present in about two thirds of all ORFs (Open Reading Frames). The distinct translation of this codon becomes especially important when reporter genes of other species are used (Gauthier, Weber et al. 2003) or when heterologous expression of *C. albicans* genes is studied.

Furthermore, to perform genetical studies on the *C. albicans* model, it was essential to dispose of strains with auxotrophic markers that could permit an easy selection of mutant cells after transformation. Most of the strains used in research laboratories nowadays rely on a strain constructed by Fonzi *et al.* that is auxotrophic for uracil, CAI4 (Fonzi and Irwin 1993), which was derived from the sequenced strain SC5314 and which is the parental strain of many important strains used for research, for example the triple auxotrophic strain (*ura3/ura3, his1/his1 and arg4/arg4*) BWP17 (Enloe, Diamond et al. 2000). This has the big advantage that many studies become comparable because of their common genetical background. On the other hand, this might limit the relevance of laboratory results for the various clinical *C. albicans* isolates.

Finally it is important to mention that an auxotrophic marker can affect considerably the phenotype of a mutant. For instance, many *C. albicans* gene disruption methods make extensive use of the *URA3* marker (See also 1.2.2.1 Gene disruption). An important drawback of this marker is that its ectopic expression has been shown to affect both adhesion capacities (Bain, Stubberfield et al. 2001) and virulence (Sundstrom, Cutler et al. 2002) of *C. albicans*.

Thus, it is possible that the phenotype of a null mutant can be linked not to the gene knockout itself, but instead to the ectopic expression of the *URA3* marker. Additional experiments are then required to confirm that the observed phenotype is linked to the disrupted gene, like for example the complementation of the mutant with a wildtype allele or the construction of an independent disruption mutant for the gene using another marker.

1.2.1 The genome of *C. albicans*

C. albicans is one of the first eukaryotic pathogens that were selected for genome sequencing. Sequencing was performed on the whole genome using a shotgun method. The clinical isolate SC5314 was chosen for sequencing, mainly because it is the parental strain of many strains that are used in research laboratories for animal virulence assays and molecular biology (Fonzi and Irwin 1993) but also because of its high susceptibility against all clinically used antifungal agents (Odds, Brown et al. 2004)(See also Table 1), which makes it an ideal reference for drug resistant isolates.

A first preliminary Assembly (Assembly 4) of the genomic sequence with 5.4 x coverage of the genome was released in 1999. Some years later, a more complete assembly of the genome sequence with 10.9x coverage was published as Assembly 6 (Jones, Federspiel et al. 2004). It contained already most of the sequence information available today, but many sequences were misassembled. Indeed, the assembly software used (PHRAP) was well adapted for the assembly of haploid or homozygous diploid genomes, but not for diploid genomes with large allelic differences. Consequently, alleles were often annotated as distinct genes and the genome size exceeded by more than 20 % the haploid genome size of *C. albicans* (Jones, Federspiel et al. 2004).

With the official release of Assembly 19 in May 2004 (preliminary release already in 2002), most of these incoherencies had been fixed, mainly by using pairwise alignments of each possible pair of contigs from Assembly 6 to discover possible allelic sequences. Hence the diploid genome of *C. albicans* is accessible with sequences for both alleles for most of the genes, and a reference haploid genome with about 6400 ORF was generated that can be considered as relatively complete (Jones, Federspiel et al. 2004). The genome size and the electrophoretic karyotype now corresponded well to data from the physical map that was mainly derived from strain CBS5736 (Chibana, Magee et al. 1998).

Nevertheless the data were still erroneous for a large number of predicted genes, mainly due to absence of a complete physical map and to the difficulty to differentiate between polymorphic alleles of one gene and genes of recently diverged gene families. In addition many genes were fragmented into partial ORF's or contained large overlapping regions. Following a recent effort of the *Candida* research community, all available sequence data were manually reviewed using multiple bioinformatical tools. This effort led to the diploid assembly of 6354 genes, accessible at http://www.*Candida*genome.org/ and http://genolist.pasteur.fr/*Candida*DB/ (d'Enfert, Goyard et al. 2005). At the same time a more complete annotation was generated using a gene nomenclature homologous to that of *S. cerevisiae* genes and gene ontology (GO) terms (Braun, van Het Hoog et al. 2005). In the annotation process all coding sequences were blasted against ten different complete genomes, five fungal species (*S. cerevisiae*, *Schizosaccharomyces pombe*, *Aspergillus niger*, *Magnaporthe grisea* and *Neurospora crassa*) as well as five higher eukaryotic organisms (*Arabidopsis thaliana*, *Caenorabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*). One of the aims was to identify genes unique to *C. albicans* or fungal-specific genes that could eventually serve as targets for antifungal drugs. For 1218 genes, or 19.2 % of the genes, no significant similarity was found in other genomes, indicating a proportion of "unique" genes comparable to that of *S.cerevisiae* (Braun, van Het Hoog et al. 2005). In addition to these *Candida*-specific genes, 228 genes that are likely to be fungal-specific have been identified (Braun, van Het Hoog et al. 2005).

Another important feature of the genome of *C. albicans* is the high rate of polymorphic differences compared to other sequenced genomes. One in 234 bp differs between the two alleles (Nantel 2006) allelic differences appear in more than half of all genes (Odds, Brown et al. 2004). There are eleven highly polymorphic regions; the largest of them is the mating type locus of *C. albicans*. In addition, there are 82 large deletions or insertions, most of them being located in intergenic sequence regions. A possible explanation for this particularity is the high frequency of short tandem repeats, present in 41.5 % of all genes and that are likely to facilitate allelic rearrangements (Nantel 2006).

In a recent update of the *Candida* genome database, Assembly 20, many sequence gaps could be filled (Arnaud, Costanzo et al. 2007). But the most notable change is the attribution of a chromosomal location to each gene, thus combining the physical map data with sequence information (Chibana, Beckerman et al. 2000). The physical map provides information about size and structure of the genome of *C. albicans* (http://albicansmap.ahc.umn.edu/)(Figure 2). The genetic information of *C. albicans* is

distributed over its eight chromosomes, chromosomes 1-7 and R. However, there are huge variations in size and gene density between the different chromosomes.

Seven out of eight chromosomes contain at least one major repeat sequence (MRS), large sequence regions composed of multiple copies of three repetitive elements, RPS, RB2 and HOK. Differences in the MRS regions are the main reason for karyotypic variations between different *C. albicans* strains. As the RPS element contains the rare SfiI site, SfiI digests can be used to map different karyotypes. To date, besides enhancement of genetic variability (Lephart, Chibana et al. 2005), no concrete function could be attributed to these regions.

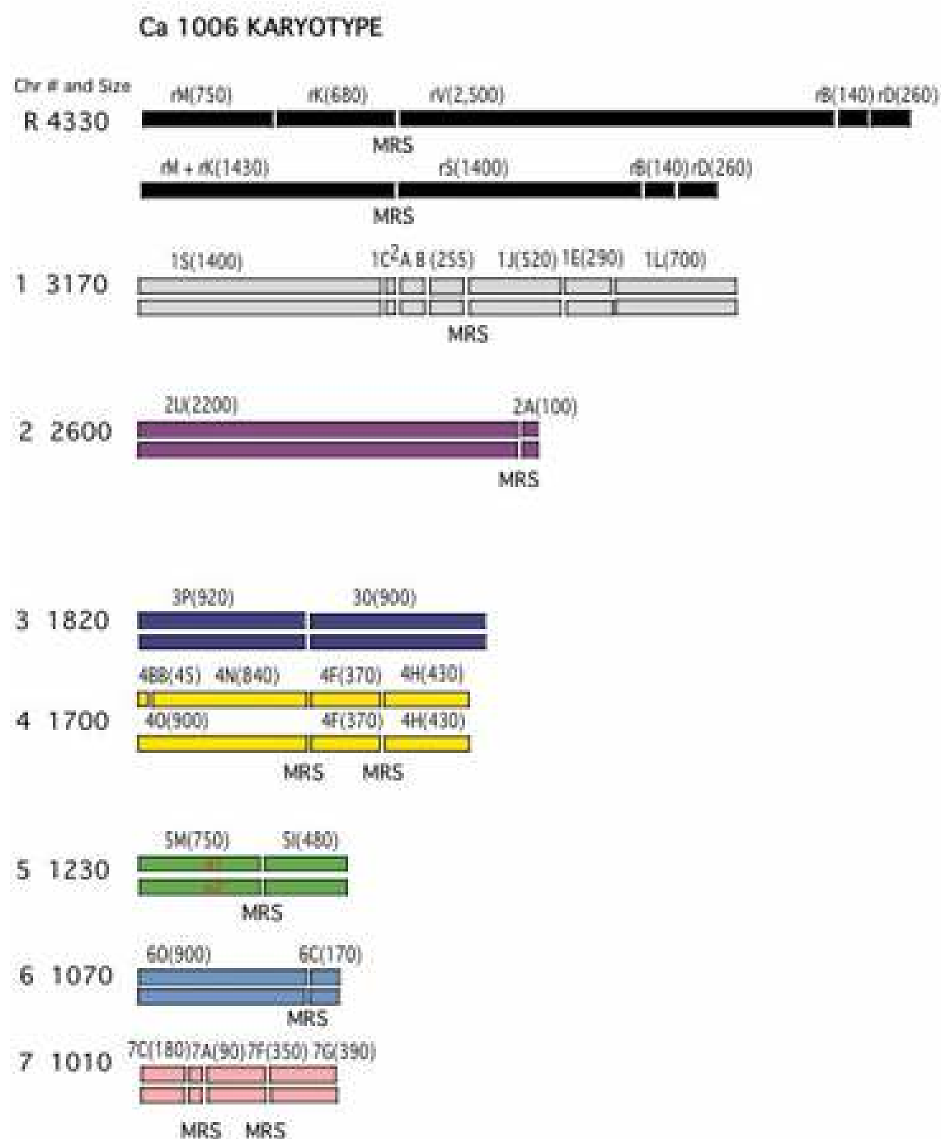


Figure 2: *Karyotypic map of C. albicans strain 1006 (same karyotype as the sequenced strain SC4314) borrowed from http://albicansmap.ahc.umn.edu/. Chromosome name and size (in kbp) are indicated on the left side. The interruptions indicate SfiI sites; restriction sites for this endonuclease mark all Major Repeat*

Sequences (MRS) and some other genomic locations. Most karyotypic changes between different strains are due to translocations that occur at or near to these MRS.

Although they provide important information about the physical location of different genes, the data of Assembly 20 have to be used with caution. For example, in contrast to Assembly 19, Assembly 20 is designed in haploid form and gives just information about one allele; for heterozygous polymorphic genes the use of Assembly 19 is still indispensable. Furthermore, the sequence traces used to fill gaps of Assembly 19 were derived from strain WO-1, and not from strain SC5314, as in previous Assemblies (http://www.*Candida*genome.org/help/Assembly20_Advisory.shtml). Nevertheless, this new Assembly complements the data already available, and the combination of the different sequence information gives already an improved picture of the genome of *C. albicans*. The nearest known phylogenetic neighbor of *C. albicans*, *Candida dubliniensis*, is currently sequenced (http://www.sanger.ac.uk/sequencing/*Candida*/dubliniensis/). A comparison between the two genomes might help to correct some of the remaining bias of the current version.

1.2.2 Molecular Genetics

With more and more genomic sequences available, the functional analysis of a gene often begins with a closer look at its sequence. Sequence similarities to well characterized genes of known function, either in the same organism (gene family) or in a phylogenetic neighbor can often give good predictions of the function its product. In particular, many genes of *C. albicans* have direct homologues in *S. cerevisiae* which have often already been characterized in detail. But even if the function is not known, sequence attributes like signal peptides, transmembrane motifs or DNA binding motifs can give important hints that help to identify the localization of the gene product and attribute a possible role to it. Nevertheless, to confirm that the described function is conserved, direct experimental proofs in *C. albicans* are still required.

This functional analysis of a gene product can be achieved in many different ways, but the first step generally targets the expression of the gene. For example, the disruption of a gene results in the complete abolition of its expression and might result in a phenotype different from that of the wild type strain which allows a first assumption about the function of the gene product. On the other hand, valuable indications about the gene function can also

be obtained using overexpression of the gene under the control of a promoter stronger than the native one. Heterologous expression of the gene can also be used, for instance the ability to complement a mutant of the homologous gene in another organism can be assessed. In some cases the phenotypic effect of heterologous expression can be directly conclusive (e.g. *ALS* gene expression in *S. cerevisiae*, chapter 1.3.1.3). To identify a genes function, it is as well useful to know under which conditions the gene is expressed and how it is regulated, as the conditions under which a gene is expressed are likely those under which it has a function. Gene transcription can be quantified by Northern Blot or RT-PCR, and also by the use of reporter genes under the control of the promoter of the gene of interest. Finally, some approaches allow a more global analysis of gene expression, including proteomic approaches (e.g. 2-D gel electrophoresis combined with mass spectroscopy methods; Figure 3) and global transcription analysis via microarray.

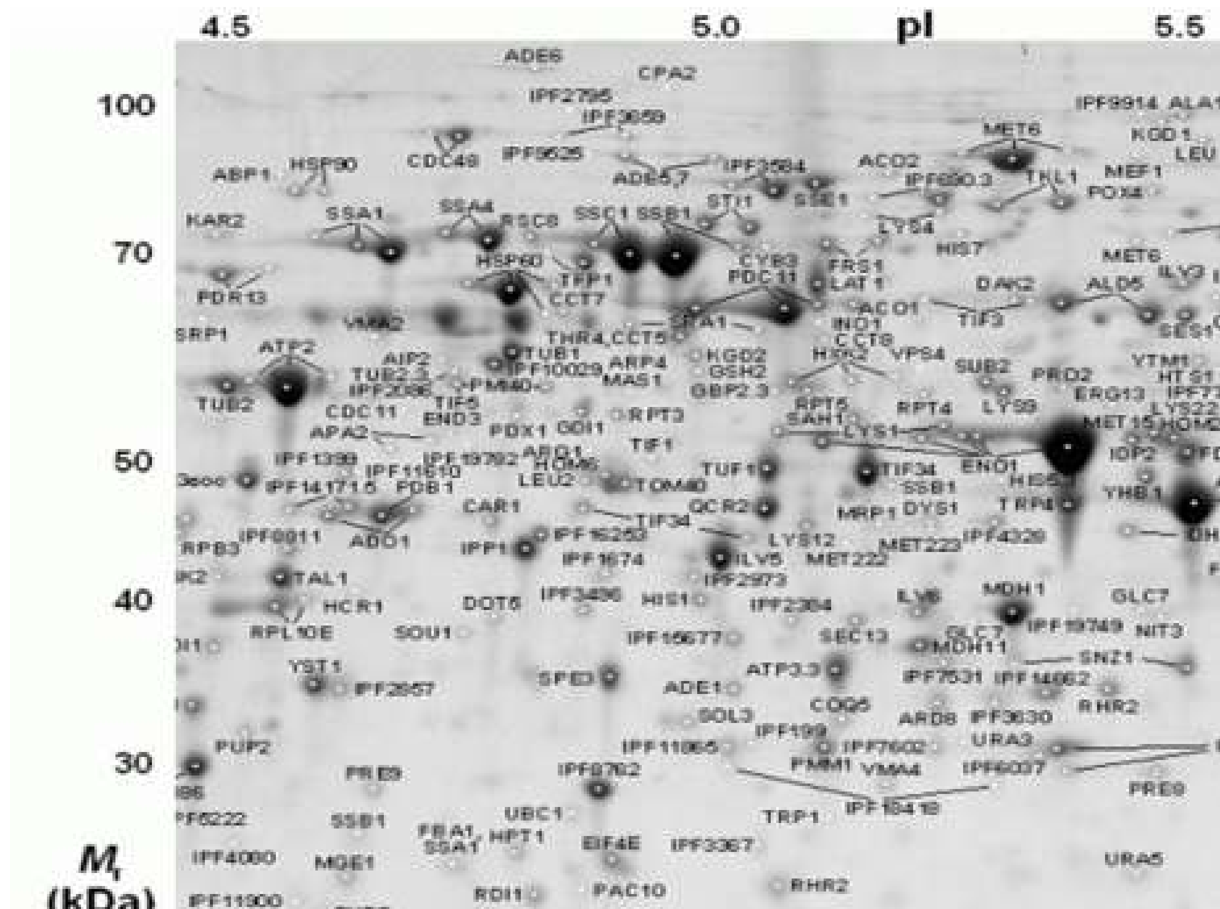


Figure 3: Section of a 2-D-gel of the proteome of C. albicans taken from http://www.abdn.ac.uk/cogeme . Samples are resolved in the first dimension on a pH gradient gel (pH 4 to 7) according to their isoelectric point, then in the second dimension on a SDS polyacrylamide gel depending on their mass. Sometimes multiple forms of one protein (for example differentially phosphorylated forms) coexist under the same experimental conditions which results in multiple signals that primarily differ in their pI.

These can be particularly enlightening if the activity of the gene of interest in the cellular network is located "upstream" of numerous other genes, as it is the case for transcription factors or components of signal transduction pathways (Enjalbert, Smith et al. 2006; Mulhern, Logue et al. 2006). Moreover, these global techniques can help to identify multiple genes which function under a condition of choice (Setiadi, Doedt et al. 2006; Sohn, Senyurek et al. 2006). In the following parts of this chapter, some of these techniques will be described more in detail.

1.2.2.1 Gene disruption

All disruption methods require at some point the introduction of genetic material into the cell. The first transformation of *C. albicans* was achieved in 1986 by Kurtz et al. (Kurtz, Cortelyou et al. 1986). Nowadays several transformation protocols for *C. albicans* using many different methods to facilitate DNA uptake have been established, including spheroplast (Kurtz, Cortelyou et al. 1986; Nakagawa, Kanbe et al. 2003), LiAc (Sanglard, Ischer et al. 1996) and electroporation (De Backer, Maes et al. 1999) based methods. The main challenge in the construction of knockout mutants of *C. albicans* is the necessity to disrupt both alleles of a given gene. Several different methods have been developed to achieve this goal (Figure 4). The most prominent and widely used of them is the so-called *URA*-blaster method, useful for Ura$^-$ strains derived from CAI-4. It is based on a disruption cassette carrying the *URA3* gene flanked by direct repeat regions (hisG) that facilitate excision by mitotic recombination. In a first transformation, it is integrated in one allele of the gene and *URA3*-positive clones are selected. After counterselection for subsequent excision of *URA3* using 5-FOA (5-Fluoroorotic Acid) plates, the now heterozygous *URA3*-negative strain can be retransformed with the same plasmid to disrupt the second allele, resulting in a null mutant strain for the selected gene which is heterozygous for *URA3* (Fonzi and Irwin 1993). An alternative method to recover the Ura- auxotrophy has been found by the Morschhäuser laboratory. They adapted the *FLP* recombinase gene from *S. cerevisiae* to fit *C. albicans* and cloned the Flp target sequence FRT on both sides of the *URA3* gene. With the right promoter in front of *FLP*, recombination can be induced and the second allele can be transformed using the same plasmid (Michel, Ushinsky et al. 2002). A more rapid PCR-mediated method has also been developed (Wilson, Davis et al. 1999). As two different markers are used, the excision of the marker by homologous recombination is not required, and the transformations can be performed without

counterselection. However, the method requires two auxotrophies in the parental strain, which strongly restricts the number of strains that can be used.

Figure 4: *Scheme of C. albicans gene disruption methods, as presented by Berman et al. (Berman and Sudbery 2002):*
A) The Ura-blaster method uses a recyclable URA3 cassette flanked by repeats (purple). After selection of URA3+ prototrophs a counterselection on 5-fluoro-orotic acid (5-FOA) identifies recombinant isolates that have lost the URA3 sequences through recombination between the repeats. The second allele of this heterozygous strain can be disrupted with the same disruption cassette.
B) PCR-mediated disruption: Two different selection markers are amplified using long primers carrying homology regions of the gene to disrupt. These cassettes replace the two alleles in two subsequent transformations.
C) Single transformation disruption using the UAU1 marker cassette Ura3Δ3'–ARG4–Ura3Δ5'. First Arg+ transformants are selected, subsequently these are screened for Arg+ Ura+ recombinants (step 3), which occur after mitotic gene conversion or break-induced replication.

All the above mentioned methods require two separate transformation steps. A newer method uses a so-called UAU1 cassette, basically two incomplete *URA3* genes interrupted by the *ARG4* gene (Enloe, Diamond et al. 2000). If this cassette is integrated in one allele of the gene of choice, it can be copied into the other allele in a rare event of mitotic crossing over or gene conversion. A long homologous region between both incomplete copies of *URA3* favours the excision of *ARG4*. The result is a strain that is URA^{+}/ARG^{+}, since one allele carries a copy of the complete UAU1-cassette (ARG^{+}) while the other one carries the *URA3* gene.

Recently some forward genetic approaches using transposon-mediated disruption have been developed. In one of them a library of randomly constructed heterozygous mutants was screened for haploinsuffiency in filamentation (Uhl, Biery et al. 2003), in the other one the above mentioned UAU1-cassette was used combined with transposon-mediated insertion to construct a library of homozygous mutants which was subsequently screened for defects in the pH response and biofilm formation of *C. albicans* (Davis, Bruno et al. 2002).

1.2.2.2 Reporter genes

Reporter genes are molecular tools that can provide valuable information about the biological role of another gene. If the reporter gene is expressed under the control of the promoter of a gene of interest, it can give quantitative information about the transcription of this gene under the conditions of choice, for instance the expression within a certain time frame, the response to environmental signals (such as temperature, pH or nutrients) or in a mutant compared to a reference strain. Reporter genes can also be used for the detection of transcription factor binding sites on a promoter sequence and the determination of their impact on the regulation of a gene. Finally, if an appropriate reporter gene is fused to the coding sequence of another gene, the detection of the resulting fusion protein can specify the cellular localization of the gene product. Most of the reporter genes commonly used in *C. albicans* are derived from reporter gene systems that were already described in other model organisms. However, as already mentioned, *C. albicans* uses a differential translation for the "CUG" codon and reporter genes from other organisms that possess such a codon often have to be codon-optimized for efficient use in *C. albicans*.

This could be seen for example in the use of common β-galactosidase reporters. A β-galactosidase gene, isolated from the conventionally coding yeast *Kluyveromyces lactis,* was

one of the first reporter genes tested in *C. albicans* (Magee, Gale et al. 2003). Its functionality could be validated only under the control of strong promoters in some of the transformants, and it might be that the presence of two CUG codons in the ORF limits its usefulness (Leuker, Hahn et al. 1992). A bacterial *lacZ* gene from *Streptococcus thermophilus* shows a much stronger activity and contains only one such codon, which has been optimized for use in *C. albicans* (Uhl and Johnson 2001). A luciferase reporter gene from *Renilla reniformis* that lacks any CUG codon can also be used for expression quantification (Srikantha, Klapach et al. 1996). Interestingly, in addition to its important function as an auxotrophic marker in many gene disruption methods, the *URA3* gene has also been successfully used to quantify gene expression (Myers, Sypherd et al. 1995).

Finally, about ten years ago GFP (Green Fluorescent Protein from *Aequorea victoria*) reporters became available to both quantify and localize gene expression in *C. albicans* (Cormack, Bertram et al. 1997; Morschhauser, Michel et al. 1998). Little later different wavelengths of fluorescence were obtained via site-directed mutagenesis, creating yellow- and cyano-fluorescent proteins (YFP and CFP) that can be used for simultaneous analysis of several genes (Gerami-Nejad, Berman et al. 2001).

1.2.2.3 Global Transcription Analysis

The transcriptional events that can be monitored with the different reporter fusion constructions always only represent the expression of a single or few genes. In addition, these methods are fairly labor-intensive, since the reporter gene constructs require manipulating DNA and transformation of cells with the resulting recombinant molecules. But thanks to the rapid technical development it is nowadays possible to get a simultaneous picture of the transcription of all *C. albicans* genes using microarray technology (Figure 5). In this chapter some of important technical aspects such as the experimental design, different labeling techniques and data normalization will be discussed.

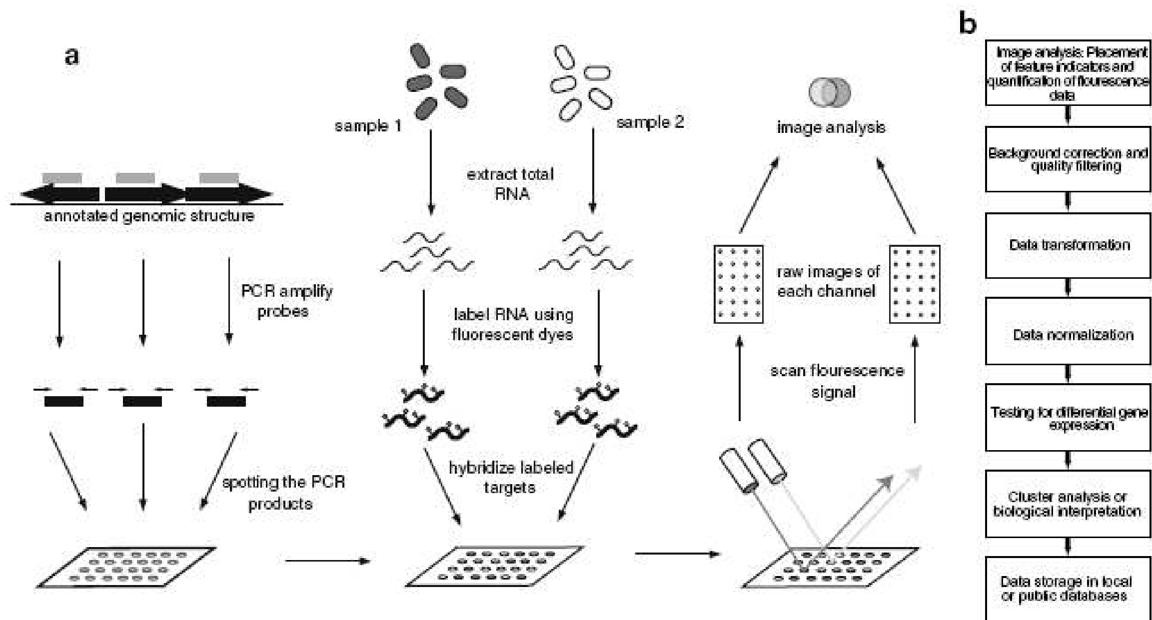


Figure 5: *The principal steps in a classical cDNA microarray experiment with two differentially labeled samples (taken from Ehrenreich (Ehrenreich 2006)); the procedure for single dye experiments is essentially the same. Specific probes are generated for all genes based on the available genome sequence. These can be spotted as denaturated PCR products as shown in the scheme or as oligonucleotides that are often synthesized directly on the array support. For target preparation, total RNA is extracted and labeled during reverse transcription. If two samples are hybridized to the same array slide, two different fluorescent dyes (usually Cy-5 and Cy-3) are used and the hybridization can be seen as a non-competitive binding reaction of two distinct cDNA populations to the probes. Fluorescence emission of each array spot is registered with a scan at the optimal wavelength for the dye(s) used, and images can be analyzed. Afterwards, the signal quality is checked for each spot and the data are transformed into quantitative values, normalized and analyzed.*

- ***Basic principles and different array types***

Microarrays have evolved as a highly developed large-scale version of earlier existing molecular tools for DNA quantification like Southern Blots (Southern 1975), dot blots (Kafatos, Jones et al. 1979) or macroarrays (Nguyen, Rocha et al. 1995). A microarray consists of a dense arrangement of several thousand short DNA fragments or oligonucleotides on a solid support, generally a glass slide or a nylon membrane. These DNA fragments are also called probes and their location defines the precise array position where the complementary cDNA strand will hybridize.

Principal differences between the various microarray types can be found in the way the probes are generated and presented to their targets for hybridization. Probes were initially generated by PCR from cDNA libraries and spotted as denaturated double-stranded probes on the slide.

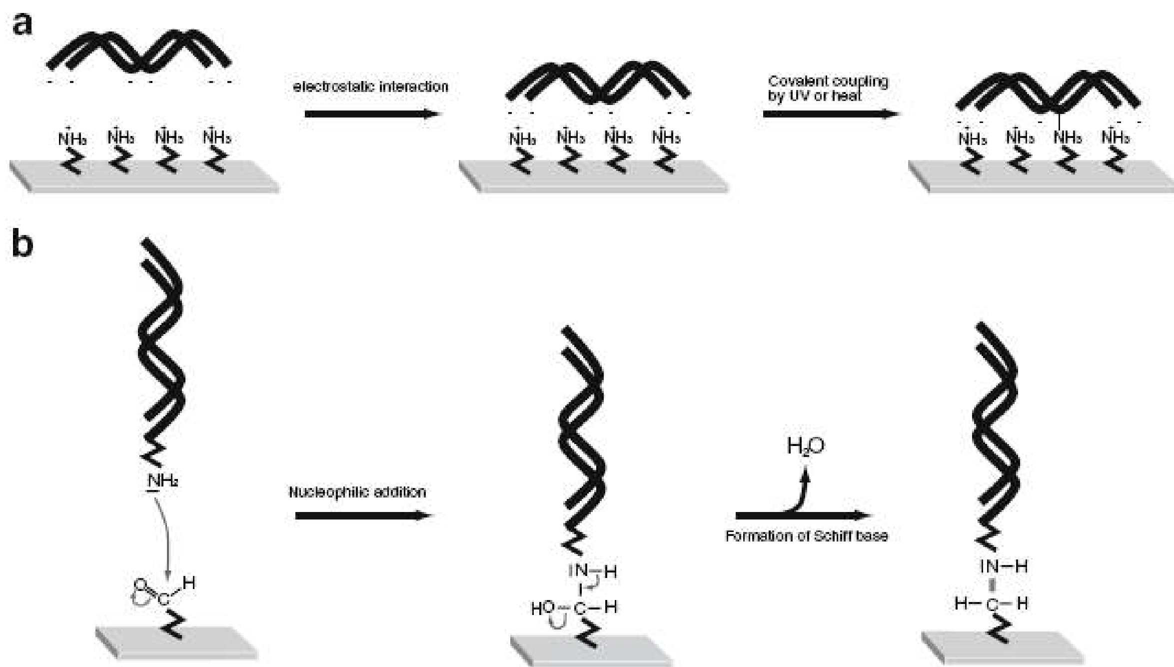


Figure 6: *Different modes of microarray probe fixation on the array support. a) In a first step DNA binds to the matrix due to the electrostatic interaction of its negative charge with the positive charge of ammonium ions on the support. In a second step the establishment of a covalent linking is favoured by UV or heat. b) Alternatively a 5'aminolinker can be attached to the probe to facilitate linkage to a surface with exposed aldehyde groups in a nucleophilic addition followed by the stabilizing formation of a Schiff base (with exclusion of a water molecule).*

It is more common nowadays to produce single-stranded gene-specific oligonucleotides. Affimetrix™, one of the pioneer companies in microarray technology and still the world leader on the high-density DNA microarray market (Gershon 2005), has developed a method that permits the *in situ* synthesis of oligonucleotides by photolithography (Fodor, Read et al. 1991) directly on the support (Lipshutz, Fodor et al. 1999). As the specificity of a probe of 25 nucleotides may not be high enough, each probe is accompanied by negative controls with a single differing base in the middle of the probe termed mismatch probe.

Global transcription is monitored indirectly through the quantification of the cDNA that corresponds to each transcript. During the reverse transcription the resulting cDNA strands are labeled to permit detection. Although radioactive labeling was initially used (Granjeaud, Bertucci et al. 1999), the currently used labeling strategies are generally based on fluorescent dyes. One huge advantage of fluorescence labeling is that two samples that are labeled with distinct fluorescence dyes can be analyzed on the same array after non-competitive hybridisation with the probes (admittedly in vast molar excess). The variance that might be generated during the preparation of distinct array slides (e.g. print quality, hybridisation and washing steps) is therefore minimized. *In situ* generated oligoncleotide

arrays even display a much higher reproducibility due to the different strategy used and to the numerous negative control probes. Since Affimetrix™ chips are hybridized sequentially with single preparations, they require a single dye (Ehrenreich 2006). This has the important advantage that variance due to labeling effects can be excluded.

- ***Fluorescense labeling and dye effects***

These dye effects deserve particular attention in the case of the widely used "direct labeling" procedure. In this case the two most commonly used dyes for microarrays, Cy-3 and Cy-5, are incorporated in the cDNA-strand during the reverse transcription process. This is reached by the use of a nucleotide triphosphate (usually cytosine) which carries the dye of choice. These labeled dCTP derivatives are added to an unbalanced dNTP mix with a lower dCTP concentration for the reverse transcription of mRNA samples (Khodursky, Bernstein et al. 2003), resulting in cDNA with randomly integrated fluorescence dyes. However, the slightly different chemical properties of Cy5-CTP and Cy3-CTP result in small differences in the incorporation frequency. This generates an artificial bias into the results which has to be corrected to obtain biologically relevant data.

This dye integration effect is circumvented when "indirect labeling" is used. In this case, both RNA preparations are reverse-transcribed to cDNA in the presence of an aminoallyl-modified dCTP. There are no differences in the reverse transcription reaction and thus no bias. In a second step, N-hydroxysuccinylimidyl (NHS) ester derivatives of Cy-3 or Cy-5 are coupled to the aminoallyl-modified cDNA molecules by a chemical reaction that is far less sensitive to the molecule size of the dye. However, although this labeling technique is bias-corrected, it is significantly more laborious and much of the advantages are compensated by a poorer yield due to additional purification steps and the sensitivity of NHS ester–modified dyes (Ehrenreich 2006).

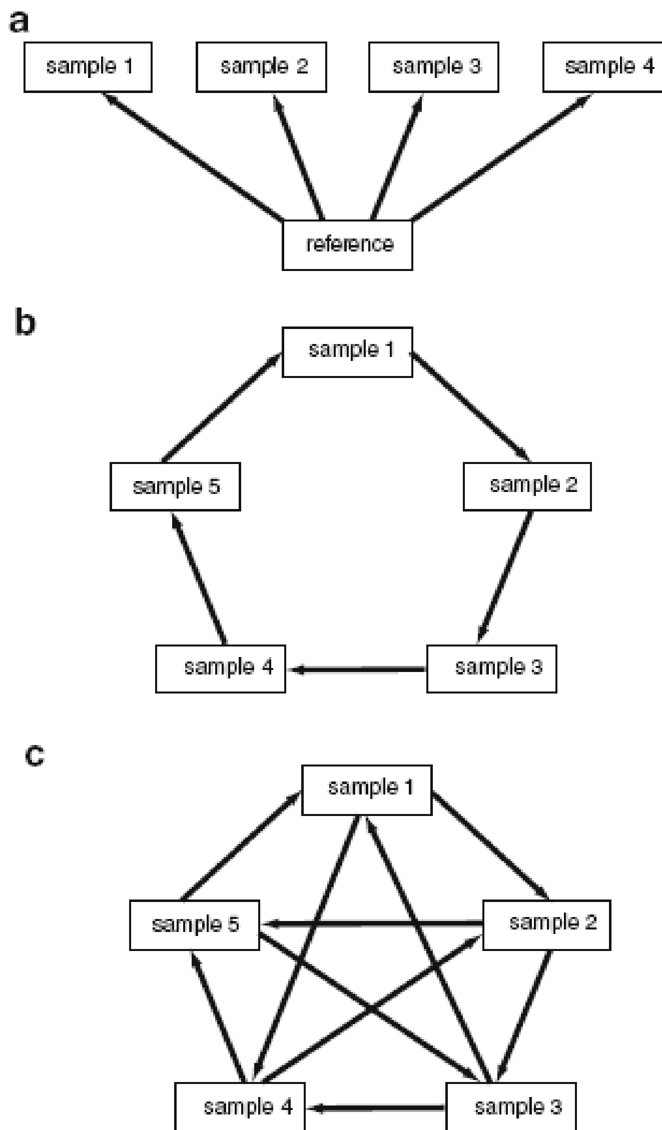
A slightly different method allows the labeling of genomic DNA for comparative genomic studies. In this case genomic DNA is fragmented into suitable sizes of 1 to 3 kb by restriction or sonication and Cy-3 or Cy-5 are integrated either by random priming with the Klenow fragment of DNA polymerase or directly in a nick translation (Ehrenreich 2006). Such approaches are useful for investigation of genetic rearrangements or intra-species diversity.

In case of the classical two-coloured arrays, equal amounts of the differently labeled cDNA samples need to be mixed together before hybridization with the array probes. The

hybridization might then be seen as the synchronous binding process of these two cDNA populations to their corresponding probes. As the probes are presented in excess, competitive binding only plays a minor role and both cDNA populations can be quantified in parallel. The choice of which sample is labeled with which dye is free. A common procedure is the so-called "dye swap", where a second array with identical samples, but inverted dye labeling, is prepared (Yang and Speed 2002). In this way dye-specific effects are detected; these data can then be used during the normalization procedure to correct for dye-related bias.

- ***Experimental Setups***

There are many different setups possible for an array experiment, but different biological questions favour distinct strategies. In a two-condition experiment, the samples can be directly hybridized against each other. In more complicated setups the use of a common reference might be advisable, because this allows an easier cross-comparison between multiple samples on different arrays. In this way, each sample is comparable to any other sample directly through the common reference (Figure 7 a). The standard error between any two samples should thus be principally the same, whereas for example in a "loop design" (Figure 7 b) the error in the comparison of samples depends on the number of arrays that are needed to "connect" them and increases with the number of arrays that form the loop. Another disadvantage of a loop design is that, because the loop is "closed", the addition of new arrays after the experiment becomes complicated, whereas this is possible without any problems when using a common reference. One drawback of the "common reference" model is that for the reference as much data are accumulated as for all the samples combined, although reference data are generally of minor interest. As a consequence, twice as many arrays are needed to obtain the same amount of meaningful data as in the corresponding loop design.



Array design	Advantages	Drawbacks
a) Reference design	- Easy extension with new samples - Distance and error between any two samples identical	- Double number of arrays necessary - Only indirect comparison
b) Loop design	- Direct comparison of adjacent samples possible - No reference data needed	- Difficult to extend - Unequal distance between sample pairs
c) All pair design	- Direct comparison of any two conditions possible - No reference data needed - Distance and error between any two samples identical	- Number of arrays increases rapidly when many conditions are examined

Figure 7: *Examples of some possible designs for a microarray experiment*

In the particular case of serial array experiments, for example in time course experiments (Kucho, Okamoto et al. 2005) or when the effects of serial dilutions of a chemical compound are tested, a common reference has the advantage that all conditions can be plotted against one, so that the results can be analyzed as "transcriptional patterns" arranged in the order of the different conditions analyzed. The choice of the common reference is another critical point. Often it makes sense to choose a special condition as this reference, for example the wildtype strain when mutants are analyzed, or an "untreated" towards a "treated" condition. However, it is not improbable that exactly under these reference conditions some genes are completely silent, which would result in infinite induction ratios for those genes if a signal is detectable in a sample, independently of the transcriptional strength. One strategy to avoid this problem would be to automatically define a "minimal base value" that is attributed to a reference gene when no transcription is detected. Other studies suggest to use a mixture of labeled oligonucleotides complementary to each probe (Dudley, Aach et al. 2002), labeled chromosomal DNA as reference (Belland, Zhong et al. 2003), or RNAs pooled from several sampling conditions (Laub, McAdams et al. 2000; Kucho, Okamoto et al. 2005). If the reference is a pool from RNA samples of all conditions that are analyzed, this has the advantage that the transcriptional patterns of each gene can be plotted against its average expression throughout the experiment, and each spot of an array of this experiment can give already an idea whether the gene is strongly or weakly expressed under the respective condition.

- ***Data normalization and analysis***

Once the practical part of a microarray experiment is finished, another crucial point is the normalization of the raw data. Data normalization is a process that aims to correct for all biases that might falsify the data, including dye effects like the different incorporation efficiencies mentioned before or different detection efficiencies of the dyes (Yang, Chen et al. 2002), but also dye-independent effects, for example global fluorescence intensity differences in different regions of an array. The process of normalization might be compared to the adjustment of Northern Blot or RT-PCR quantifications against a reference probe that is assumed to be constantly expressed. Many different approaches can be used to reach such an adjustment. One of them is the "total intensity normalization". The assumption here is that when comparing identical amounts of mRNA, the sum of the fluorescence signals measured from an array should be identical. Consequently, all the fluorescence values obtained for the

one dye are adjusted proportionally so that their sum matches exactly the sum of the signals measured on the other channel (Quackenbush 2002). Another normalization option is the adjustment of all data against those obtained for some selected genes (often housekeeping genes) that are assumed to be expressed stably throughout the experiment.

Many dye effects can be corrected through the normalization of all arrays against one special array. On this array, one and the same RNA sample is labeled with the different dyes and hybridised against itself, thus it should provide probe-specific information about the dye incorporation efficiency depending on each single sequence that can be applied to all other arrays. Usually the expression data of sample and reference are at some point transformed into a $\log_2$-ratio. Despite the practical usefulness of this measure, a drawback is that $\log_2$ values have a systematic dependency on intensity (Yang, Chen et al. 2002; Yang, Dudoit et al. 2002), which results in ambiguous data for low intensity spots. One quite commonly used normalization procedure to correct this is the so-called LOWESS (LOcally WEighted linear regreSSion) normalization (Cleveland 1979), which deemphasises the contributions of data that show abnormal variation (Figure 8).

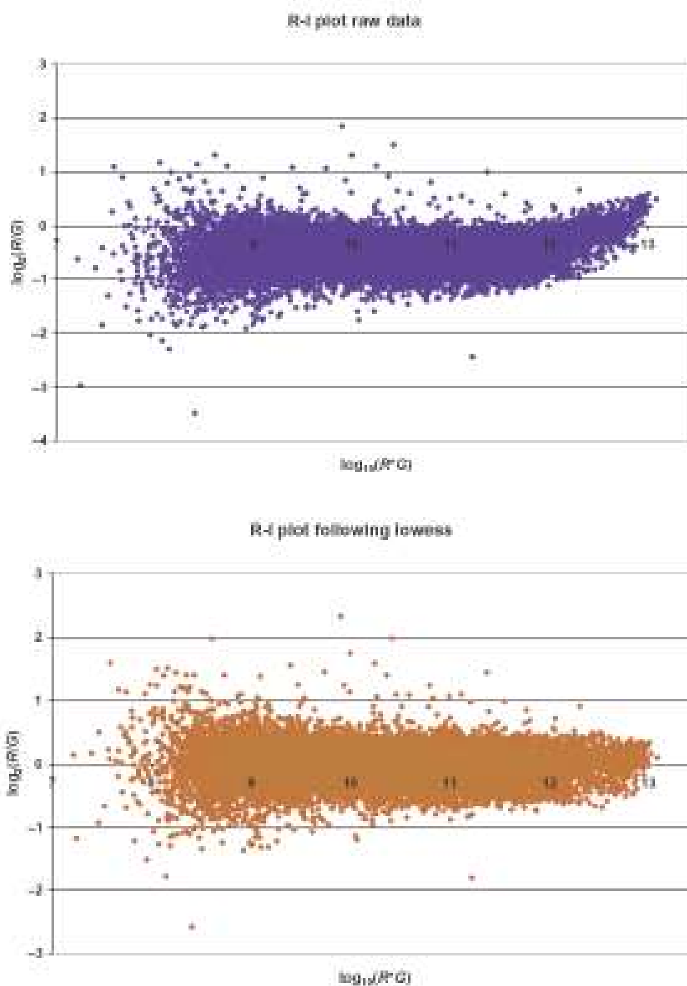


Figure 8: *The upper plot shows the log_2 of the ratio between the fluorescence signals obtained for each spot in a two-dye experiment against the log_2 of their product. This display reveals the typical "banana shape" which indicates the presence of systematic fluorescence intensity-dependent effects. These unwanted effects can be reduced during the normalization. The lower plot shows the same data set after applying a LOWESS (LOcally WEighted linear regreSSion) normalization. The signal ratio is now much less dependent on the intensity of the signal, and data points are distributed along the line of a 1:1 ratio ($log_2 = 0$) without an obvious dependency on the signal intensity. Both plots were taken from Quackenbush et al (Quackenbush 2002)*

Most of the different normalization strategies can be applied either on the complete array or locally. The latter possibility can help correcting for slight local differences in hybridization conditions across the array. However, it is important that the subregions of the array are large enough so that statistical distribution of fluorescence intensities can be assumed (Quackenbush 2002). It is also important to mention that most normalization procedures assume that global expression data follow a standard distribution. Although this might be approximately the case, this assumption is in some way opposed to the idea of active regulation of gene expression, because as a result a gene that is regulated will be considered as outlier of the statistical distribution rather than as a meaningful data point. As a consequence, the more normalization steps are undertaken, the more uniform the data set will

become, and it is up to the scientist to find the optimal balanced normalization that neutralizes the maximum of systemic biases, while maintaining the maximum of meaningful information.

After normalization, differentially transcribed genes can be identified using different statistical methods (e.g. SAM (Significance Analysis of Microarray data)(Tusher, Tibshirani et al. 2001)) that are best suited for the analysis (Cui and Churchill 2003). When multiple conditions are analyzed, it might be helpful to use different clustering methods such as Hierarchical Clustering, K-Means Clustering or Self-organizing maps (Quackenbush 2001) in order to identify genes that are similarly transcribed and could be co-regulated. Genes that are co-regulated are not unlikely to work in concert in the cell. It is thus sometimes possible to identify more easily by clustering regulated multiprotein complexes like ribosomal proteins or pathways (Stuart, Segal et al. 2003).

- ***C. albicans arrays***

Microarrays for *C. albicans* cDNAs were constructed by different groups shortly after the release of Assembly 4 and were quickly applied to a wide range of different investigation fields (Figure 9 (Garaizar, Brena et al. 2006)). Nowadays oligo-nucleotide arrays with 70mers (Cao, Lane et al. 2006) are available as well as arrays with spotted PCR-fragments from cDNA libraries (Fradin, Kretschmar et al. 2003; Barker, Crisp et al. 2004), and since recently also SNP (Single Nucleotide Polymorphism) arrays (Forche, May et al. 2005). The use of an oligonucleotide array for *C. albicans* that covers the whole genome including intergenic sequences has as well been reported (Srikantha, Borneman et al. 2006).

Reference	Source of gene information	Sequences included/ spots	Origin of spots	Microarray format	Used to study	Institution or company
Lane *et al*. (2001)	Stanford Genomic Center*	700	PCR fragments (ORFs)	Nylon membrane	Filamentation	University of California and Beckmann Coulter, Inc.
De Backer *et al*. (2001)	Stanford Genomic Center	6 600	PCR fragments (ORFs)	Glass slides	Response to antifungal	Incyte Genomics, Inc., Wilmington, Delaware
Rogers *et al*. (2002)	Stanford Genomic Center	6 600	PCR fragments (ORFs)	Glass slides	Response to antifungal	Incyte Genomics, Inc., Wilmington, Delaware
Bensen *et al*. (2004)	Assembly 6, Stanford Genomic Center	6 175	PCR fragments (ORFs)	Glass slides	pH adaptation	University of Minnesota
Barker *et al*. (2004)	Assembly 6, Stanford Genomic Center	6 039	PCR fragments (ORFs)	Glass slides	Response to antifungal	Eurogentec and European Galar Fungail Consortium
Kadosh & Johnson (2005)	Stanford Genomic Center	11 000	PCR fragments (ORFs)	Glass slides	Filamentation	University of California and Massachusetts Institute of Technology (MIT)
Cao *et al*. (2005)	Sequencing Library of *C. albicans*. National Center Biotechnology Information Unigene	3 102	PCR fragment (full-length and partial cDNA sequences)	Glass slides	Adhesion (quorum sensing)	United Gene Holdings, Ltd., Shanghai,
Enjalbert & Whiteway (2005)	Version 4, Stanford Genomic Center	6 333	PCR fragments (ORFs)	Glass slides	Filamentation (quorum sensing)	Biotechnology Research Institute (BRI)†, National Research Council, Montreal.
Andes *et al*. (2005)	Stanford Genomic Center	6 737	PCR fragments (ORFs)	Glass slides	Mouse infection	BRI, National Research Council, Montreal
Zhao *et al*. (2005)	Assembly 6, Stanford Genomic Center	6 392	PCR fragments (ORFs)	Glass slides	Adhesion	National Institute of Dental and Craniofacial Research, Bethesda, USA
Zhao *et al*. (2005)	Assembly 6, Stanford Genomic Center (ORFs) plus GeneBank (oligos)	6 266	70mer Oligos (5 948 ORFs plus 318 oligos)	Glass slides	Mating	Qiagen/Operon
Cao *et al*. (2006)	Assembly 6, Stanford Genomic Center	6 530	70 mer Oligos	384-Well plate format	Filamentation	Qiagen/Operon

*Candida Genome Database in Stanford Genomic Center of *C. albicans* SC5314 strain.
†Actually the BRI Microarray Facility produces and distributes the new *C. albicans* oligonucleotide array containing 70mer oligonucleotide probes for each of the 6354 genes.

Figure 9: *Overview about publications presenting studies with different microarrays for C. albicans as reported by Garaizar et al. (Garaizar, Brena et al. 2006). An additional study that was published prior to Lane et al. 2001 but that is not mentioned in this list, used a C. albicans macroarray with 2002 probes to find target genes of Nrg1 (Murad, d'Enfert et al. 2001).*

1.3 Virulence factors

We still poorly understand which are the mechanisms that underlie *Candida* pathogenesis and which are the molecular factors required. However, it is obvious that in order to cause disease an invasive pathogen must have particular traits which ensure that under certain conditions it is able to colonize a host, penetrate the surface, to cause tissue damage and to avoid the immune response (Hube and Naglik 2001). Although in the specific case of an opportunistic pathogen it is not always obvious to differentiate virulence factors from others that are important for survival of the organism in both commensal and pathogenous state (Navarro-Garcia, Sanchez et al. 2001), there are some extensively studied key elements that are widely recognized as absolutely required for *C. albicans* virulence (Gow, Brown et al. 2002). In this chapter we will discuss some of them; first we will evoke the particular importance of the fungal cell wall and its proteins in the interaction with the host with a special focus on adhesion, then we will discuss the role of secretion of hydrolytic enzymes, and finally the importance of yeast and hyphal growth forms and their regulation will be briefly discussed.

1.3.1 Cell wall and host-pathogen interaction

The cell wall has received a particular attention in *C. albicans* research for several reasons. As fungi are eukaryotic, the cells of fungal pathogens share much more attributes with mammalian cells than bacterial pathogens. Consequently, it is relatively difficult to develop antifungal drugs that do not affect mammalian cells. As the cell wall is a general cell attribute of fungi and plants, but not of animals, its composition is of particular interest for medical research. In addition, due to their exposed position components of the cell wall are easier to reach by drugs than cytoplasmic targets, since a cellular uptake of the drug is not required (e.g. polyenes as ligands of ergosterol). Nevertheless, different cytosolic cell wall synthesis enzymes have also been suggested as good targets for antifungal drugs. For example, echinocandins (see chapter 1.1.1.2) are inhibitors of β-1,3-glucan synthases, and also chitin synthesis is a center of antifungal drug research (nikkomycins, polyoxins;(Ruiz-Herrera and San-Blas 2003)).

The cell wall as the outer structure of fungal cells is responsible for the physical interaction with the host, including adhesion to and penetration (or induced endocytosis) into host tissues (Ruiz-Herrera, Elorza et al. 2006). Besides, as it is the first fungal structure to come in contact with the host, it carries important antigenic determinants of the fungus and is responsible for a possible cross-talk with the host (Poulain and Jouault 2004). Last but not least, the cell wall defines the cellular shape, which has been shown to be extremely dynamic in the case of *C. albicans* (Sudbery, Gow et al. 2004).

1.3.1.1 Composition and architecture of cell wall

The cell wall *C. albicans* of is a coherent and highly organized structure showing different layers in microscopy (Tokunaga, Kusamichi et al. 1986; Ruiz-Herrera, Elorza et al. 2006). It is composed of the same four classes of macromolecules as the *S. cerevisiae* cell wall: Mannoproteins, β-1,3-glucan, β-1,6-glucan and chitin. However, the relative proportions differ significantly, in particular the abundance of β-1,6-glucan is clearly higher in *C. albicans* than in *S. cerevisiae* cell walls (20 % compared to 5 % of dry weight) (Klis, de Groot et al. 2001).

Based on the extensive knowledge from the cell wall of *S. cerevisiae* (Kapteyn, Montijn et al. 1996; Kapteyn, Ram et al. 1997; Kollar, Reinhold et al. 1997; Kapteyn, Van Den Ende et al. 1999) and comparative studies (Kapteyn, Montijn et al. 1995; Kapteyn, Hoyer et al. 2000) it has been shown that *C. albicans* cell walls are composed by an inner chitin layer that is covered by a flexible network of β-1,3-glucan molecules that are linked to each other by hydrogen bonding. The other class of glucans, β-1,6-glucan, is generally linked either to a β-1,3-glucan chain or to a short β-1,3-glucan side chain of β-1,3-glucan. It has also been reported that β-1,6-glucan binds directly to chitin (Surarit, Gopal et al. 1988), a linkage that has not been shown for *S. cerevisiae*.

1.3.1.2 Cell wall proteins

Within this highly flexible network of glucan chains, two different classes of mannoproteins can be found: Pir proteins which are directly attached to the β-1,3-glucan layer and GPI (Glycosylphosphatidylinositol)-anchored proteins which are covalently bound through this anchor to the non-reducing ends of β-1,6-glucan (Van Der Vaart, te Biesebeke et al. 1996). While currently only two Pir proteins coded by non-identical alleles of the *PIR1* gene have been described in *C. albicans*, the class of GPI (glycosylphosphatidylinositol)-anchored proteins is much larger and comprises approximately 115 members (Richard and Plaine 2007). The function of about two thirds of predicted GPI proteins is unknown and, perhaps due to the difficulties in the detection and identification of proteins in cell wall and membrane (hydrophobicity, strong N- and/or O- glycosylation), the expression and localization of the vast majority of GPI proteins has yet to be proven (Richard and Plaine 2007).

Several mutants in genes that code for important cell wall structure proteins have been shown to be attenuated in different virulence models, including *ECM33* (Martinez-Lopez, Park et al. 2006), *UTR2* (implicated in cell wall regeneration;(Pardini, De Groot et al. 2006)) the β-1,3-glucanosyltransferases *PHR1* and *PHR2* (Ghannoum, Spellberg et al. 1995; De Bernardis, Muhlschlegel et al. 1998). The two aspartyl proteinases *SAP9* and *SAP10* of the *SAP* gene family are GPI anchored proteins that might influence indirectly *C. albicans* virulence via the proteolytic regulation of cell wall protein function (Richard, De Groot et al. 2002). Then there is evidence that *SOD5* mutants are more sensitive towards neutrophiles,

which could indicate that this superoxid dismutase increases the resistance to oxidative stress (Martchenko, Alarco et al. 2004; Fradin, De Groot et al. 2005). Furthermore, the Tup1-repressed gene *RBT1* has been suggested as a modulator of the host immune response (Braun, Head et al. 2000), although its exact function remains unclear. Finally, probably one of the best described functions of GPI-anchored proteins in virulence is the involvement in adhesion to different host tissues, which will be discussed in the following chapter.

1.3.1.3 Adhesion and recognition of host cell

A specific function in adhesion has been demonstrated for the products of several genes coding for GPI-anchored proteins, including *HWP1* and three members of the *ALS* gene family, *ALS1*, *ALS3* and *ALS5* (Sundstrom 2002; Phan, Myers et al. 2007). For Hwp1 it has been shown that it can ensure attachment of *C. albicans* to host cells by covalent binding (Staab, Bradway et al. 1999). This binding event is induced by the mammalian enzyme transglutaminase, which recognizes Hwp1p as a substrate and links it covalently to epithelial cells for example of oral mucosae (Staab, Bradway et al. 1999) (Staab, Ferrer et al. 1996) (Sundstrom, Cutler et al. 2002). Hwp1p is a hypha-specific protein that is found exclusively on the surface of germ tubes, but not in yeast or pseudohyphal cells. It has been suggested to be more important for infections of the oral mucosae than for systemic infections, as tissue invasion of a *hwp1* null mutant was deficient on lingual and oesophageal surfaces, but not in the gut (Balish, Warner et al. 2001).

- ***The ALS gene family***

The *ALS* gene family (Agglutinin-Like Sequence) consists of eight GPI-anchored genes: *ALS1-7* and *ALS9* (a gene initially described as *ALS8* has been shown to be identical with *ALS3* (Zhao, Oh et al. 2004)). They were named *ALS* genes because Als1p, the first identified member of the family, shares sequence similarities with *S. cerevisiae* α-agglutinin Agα1 (Figure 10)(Hoyer, Scherer et al. 1995). This observation already led to the speculation that *ALS1* and family members might be involved into adhesion to host tissues, a theory that quickly found experimental support (Gaur and Klotz 1997; Fu, Rieg et al. 1998). The other *ALS* genes were identified by crosshybridisation essays, functional screens (Hoyer, Payne et al. 1998; Hoyer, Payne et al. 1998) or from the genome sequence of strain SC5314 (Hoyer

and Hecht 2000; Hoyer and Hecht 2001). The most obvious attribute of an *ALS* gene is a conserved region of 108 bp tandem repeats that can vary significantly in the number of repeats (between two and 37 tandem repeats have been reported (Lott, Holloway et al. 1999; Hoyer and Hecht 2001)). These variations are not restricted to different *ALS* genes, considerable variations for the same gene in different strains (Hoyer and Hecht 2001) and even for the two alleles of one gene in the same strain have been reported (Hoyer, Scherer et al. 1995; Hoyer, Payne et al. 1998). Several studies indicate that the varying number of tandem repeats can significantly influence the function; in particular adhesion capacities seem to increase with the number of tandem repeats (Loza, Fu et al. 2004; Oh, Cheng et al. 2005).

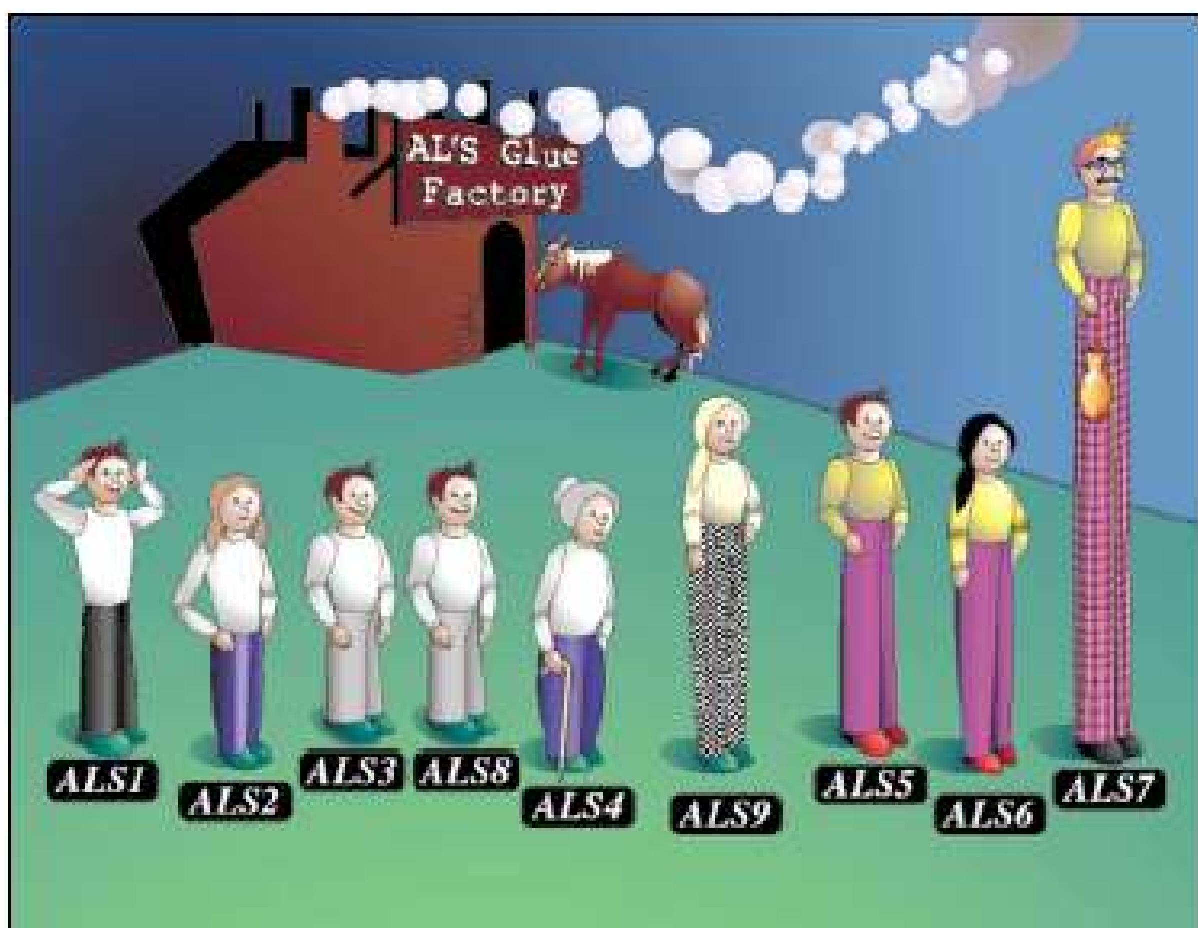


Figure 10: *The "ALS family portrait" as it was presented by Hoyer er al (Hoyer 2001): Each human figure represents an ALS gene, their heads represent the 5' domain, their pullovers the tandem repeat sequence, and their legs correspond to the 3' domain. Sequence similarities are shown by similar colours. ALS8 is shown as a twin of ALS3, as both are the same gene. There are two big subfamilies as far as the tandem repeat sequence is concerned, indicated by the white or yellow pullover colour. For the subfamily that comprises ALS5, ALS6 and ALS7 these sequence similarities extend to the 3' region.*

There are large sequence similarities between the different *ALS* genes, not only in the tandem repeat domains, but also within a 1300 bp region of the 5'domain, which is from 55 % to 90 % identical across the family (Hoyer 2001). The 3' domain is relatively variable both in

length and in sequence, but is always coding for a serine/threonine-rich peptide. Als7p shares only 55-60 % sequence identity with the other Als proteins and has a C-terminal tail of extraordinary length which in particular contains a large 137-147 amino acid region with unique tandemly arranged "V-A-S-E-S" (Valine-Alanine-Serine-Glutamate-Serine) repeats of currently unclear function (Hoyer 2001; Zhang, Harrex et al. 2003).

Although the N-terminal region is relatively conserved in length and sequence throughout the family (45–86 % amino acid identity), it has been demonstrated that there are seven hypervariable regions (HVR) which were suggested to mediate the divergent functions of Als proteins (Sheppard, Yeaman et al. 2004).

These HVR regions are flanked by eight conserved regions (CR) which are relatively invariable structural components (in particular anti-parallel β-sheets) that are characteristic for adhesins and invasins of the immunoglobulin superfamily (Figure 11)(Sheppard, Yeaman et al. 2004). Another particularity of the N-terminal regions is that they are poorly glycosylated, in contrast to the heavily N- and O-glycosylated C-terminal domain; only Als2p, Als6p and Als9p possess N-glycosylation sites within their amino-terminal domain (indicated by pony tails in Fig.10 (Hoyer 2001)). The absence of glycosylation results in a higher hydrophobicity of the N-terminal region. This might explain the increased adhesion to different surfaces of *S. cerevisiae* strains that express *C. albicans* Als proteins Als1p, Als3p or Als5p (not N-terminally glycosylated) compared to Als6p and Als9p (with an N-glycosylation site) (Sheppard, Yeaman et al. 2004). Accordingly, Als5p (formerly called Ala1p) has been originally isolated in a screen for adhesins to fibronectin-coated magnetic beads (Gaur and Klotz 1997), and Sheppard *et al.* could show with Als5p-Als6p chimers that the non-glycosylated N-terminal domain of Als5p is responsible for adhesive interaction (Sheppard, Yeaman et al. 2004). The different adhesion capacities of Als5p and Als6p despite their extremely similar amino acid sequences (> 80 % sequence identity) can possibly be explained by the observed differences in hydrophobicity (Sheppard, Yeaman et al. 2004).

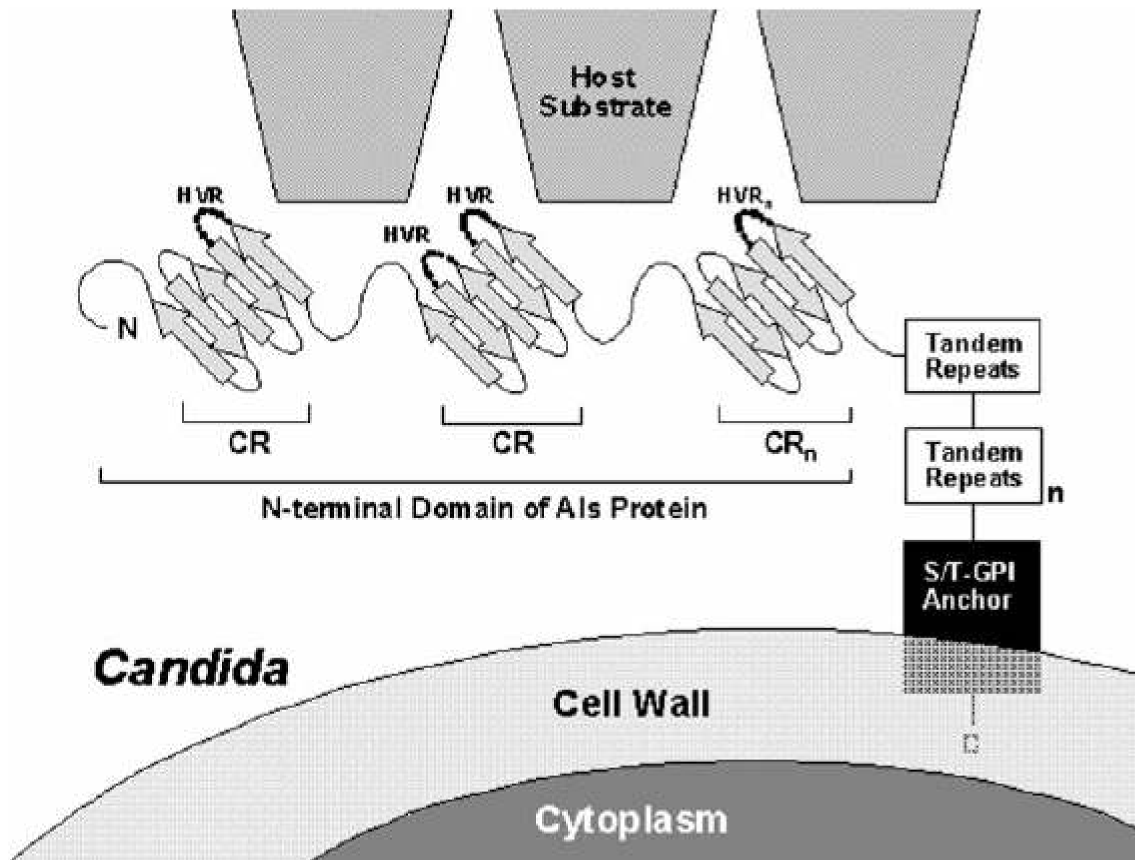


Figure 11: *According to the functional model proposed by (Sheppard, Yeaman et al. 2004) Als proteins interact via their N-terminal domain with host substrates, while the C-terminus is attached to the cell wall by a GPI anchor and the tandem repeats serve as a linker. The structure of the N-terminal domain is characterized by multiple anti-parallel β-sheet regions (CR) that are linked with extended span sequences. Embedded within exposed positions between the conserved β-sheets are hyper-variable regions (HVR) consisting of loop/coil structures which might govern the interaction with the substrates. These HVR might give distinct physicochemical properties to them which are thought to confer specific adhesive and invasive functions to each Als protein.*

- ***ALS mutant phenotypes***

Most of these results were obtained by heterologous expression of *ALS* genes in *S. cerevisiae*, but four of the eight *ALS* genes, *ALS1*, *ALS3*, *ALS4* and *ALS7* have already been disrupted in *C. albicans*. The phenotypes obtained for the mutants of *ALS1* and *ALS3* null mutants largely confirm the observations made in heterologous expression studies. Both mutants were shown to have an attenuated virulence in different cellular or animal models (Fu, Ibrahim et al. 2002; Kamai, Kubota et al. 2002; Zhao, Oh et al. 2004) and show defects in biofilm formation (Nobile, Andes et al. 2006; Zhao, Daniels et al. 2006), phenotypes that could be attributed to a defective adhesion. In addition, a recent study of the same group provided evidence that at least one *ALS* family member, *ALS3*, has an additional virulence-related function besides its contribution to the adhesion capacities of *C. albicans* (Phan,

Myers et al. 2007). They showed that the presence of Als3p is required for induced endocytosis as well in epithelial as in endothelial *ex vivo* cell models and suggested that Als3p works as an invasin that mimics the structure of human cadherins, binds to them and likewise induces endocytosis.

The phenotypes of the other available *ALS* mutants are far less impressive. The insertion mutant for *ALS7* was obtained by a large-scale transposon mediated random method and was not characterized in detail. It did not show a defective phenotype under the conditions tested (Nobile, Bruno et al. 2003). Although the absence of Als4p resulted in a germ tube formation defect under certain conditions and a significantly reduced adherence to vascular endothelial cell, the adhesion to and invasion of epithelial cells or oral RHE cells was not affected and biofilm formation was wild type-like (Zhao, Oh et al. 2005). In contrast, homozygous *ALS2* null mutants could not be obtained despite large efforts, suggesting that the gene might be essential. A heterozygous mutant with an intact allele placed under control of the *MAL2* promoter showed not only a germ tube formation defect and reduced adherence to epithelial and endothelial cells, but was in addition defective in biofilm formation (Zhao, Oh et al. 2005). In the same publication, *ALS2* was shown to be upregulated in the *ALS4* null mutant, and *ALS4* to be induced in the heterozygous *ALS2* mutant when the transcription from the intact allele was kept low. As *ALS2* and *ALS4* are virtually identical (95 % sequence identity) in the tandem repeat and the 3' region, this result was interpretated as an indication for possible compensatory function within the *ALS* gene family.

- ***Regulation of ALS gene expression***

Several other publications report differential expression of *ALS* genes. For example *ALS1* has been reported to be induced when a strain was transferred into fresh medium (Hoyer, Scherer et al. 1995), while the transcription of *ALS4* was induced later during *in vitro* growth phases (Hoyer, Payne et al. 1998). *ALS3* is expressed almost exclusively during filamentous growth and is thus known as a hypha-specific gene (Hoyer, Payne et al. 1998). Given the recently shown contributions of Als3p to induced endocytosis (Phan, Myers et al. 2007), its presence could explain why hyphal cells are taken up more efficiently than yeast cells (Phan, Belanger et al. 2000; Kumamoto and Vinces 2005). Little is known yet about the transcriptional regulation of *ALS* genes, but at least for *ALS3* it has been shown that several transcriptional regulators including Tup1p, Nrg1p and Rfg1p are involved in its repression, and Efg1p and Bcr1p are the main contributors to its activation (Argimon, Wishart et al.

2007). Efg1p (Fu, Ibrahim et al. 2002) as well as Bcr1p (Argimon, Wishart et al. 2007) seem to be also involved in the transcriptional regulation of *ALS1* during hyphal development or biofilm formation.

Several homologues of *ALS* genes were found in clinically important phylogenetic neighbors of *C. albicans*, including *C. dubliniensis* (Hoyer 2001), *C. tropicalis* and *C. parapsilosis*, which indicates that the role of this gene family in pathogenesis is probably not restricted to *C. albicans*. Recently a vaccine has been developed which uses the N-terminal domain of Als1p as a preventive agent (Ibrahim, Spellberg et al. 2005) against *C. albicans* colonization. This vaccine has been shown to reduce the fungal burden in both immunocompetent and immunocompromized mice (Spellberg, Ibrahim et al. 2005) and provides further evidence for the importance of *ALS* genes in *C. albicans* virulence.

Besides these GPI-protein coding genes, several other genes of *C. albicans* possess integrin-like sequence regions that could indicate a possible function in adhesion, including *ADH1* and *INT1* (Gale, Finkel et al. 1996; Gale, Bendel et al. 1998; Klotz, Pendrak et al. 2001). However, the colocalization of Int1p with septins between mother and daughter cells of yeast and pseudohyphae or in the filament rings of hyphae indicates a minor importance for adhesion. And although Adh1p has been identified with polyclonal antibodies against $\alpha5\beta1$ and $\alpha v\beta3$ integrins directed in a screen of an *in vitro* translated cDNA library, the surface expression of Adh1p and its implication into adhesion has still to be demonstrated. Thus, up to date the adhesive function has been proven only for members of the GPI anchored protein family.

1.3.2 Secreted Hydrolytic Enzymes

Another important aspect of the interaction of *C. albicans* is the secretion of enzymes with extracellular function, in particular the secretion of hydrolytic enzymes. There are three well described families, the secreted aspartyl proteinases (Sap), phospholipases (Plb) and lipases (Lip). Their function is assumed to be in the destruction of host surfaces to facilitate invasion, the destruction of host immune factors and nutrient acquisition (Hube and Naglik 2001).

1.3.2.1 Secreted Aspartyl proteinases

The *SAP* gene family includes ten genes (*SAP1-10*) and is probably the best studied of them. Two of them (*SAP9* and *SAP10*) encode GPI-anchored proteins, the others are thought to be secreted, where they get in direct contact with host tissues. The genes exhibit differential expression profiles under various conditions in *ex vivo* models (Schaller, Schafer et al. 1998) and in candidiasis patients (Naglik, Newport et al. 1999). While *SAP1-3* are obviously required in oral candidiasis (Schaller, Hube et al. 1999), *SAP4-6* are upregulated upon hyphae formation and involved in the invasion of pancreas and liver (Felk, Kretschmar et al. 2002). In addition, for *sap1* and *sap2* null mutants it has been demonstrated that they are defective for tissue damage in a RHVE model (Reconstituted human vaginal epithelium;(Schaller, Bein et al. 2003)). The activity of aspartyl proteinases is inhibitid by pepstatin A, and the use of this inhibitor remarkably reduces tissue damages (Kretschmar, Hube et al. 1999), which indicates the importance of *SAP* genes in *C. albicans* virulence.

1.3.2.2 Phospholipases

Similarly to aspartyl proteinases, phospholipases seem to play an important role in the pathogenesis of *C. albicans*. It has been demonstrated (Ibrahim, Mirbod et al. 1995; Kadir, Gumru et al. 2007) that the increased expression of phospholipases characterizes isolates from infections and not from commensal isolates, and that the mortality of mice is linked to the degree of phospholipase expression. In blood isolates, activity of both phospholipase B and lysophospholipase-transacylase could be detected. There seems to be a correlation between hyphal growth and phospholipase expression, as phospholipase activity is highest on hyphal tips (Pugh and Cawson 1977; Ghannoum 2000). Although phospholipases have been found in different *Candida* species, phospholipase activity seems to be particularly high in *C. albicans* strains (Samaranayake, Raeside et al. 1984; Kumar, Kumar et al. 2006). The activity of phospholipase A, B, C and D has been detected in *C. albicans* (Niewerth and Korting 2001). However, the three phospholipase C genes *PLC1-3* found in the genome of *C. albicans* do not carry signal peptides that indicate secretion of their products (Kunze, Melzer et al. 2005). In

contrast, disruption of genes coding for phospholipases B (Leidich, Ibrahim et al. 1998) and D (Dolan, Bell et al. 2004) resulted in a defect in tissue penetration and attenuated virulence in a mouse model. Genes with clear sequence similarity to mammalian or bacterial phospholipases A1 and A2 are lacking, but recently a phospholipase A2 function has been demonstrated for the *PLB5* gene, and a null mutant of this gene showed reduced tissue colonization in a mouse model.

1.3.2.3 Lipases

The third gene family coding for secreted hydrolases includes ten genes coding for lipases, *LIP1-10* (Hube, Stehr et al. 2000). Except Lip7p, all *C. albicans* lipases carry a putative N-terminal signal sequence suggesting a possible secretion (Roustan, Chu et al. 2005). Little is known about the involvement of lipases in fungal virulence, but a link of bacterial lipases to virulence has already been established (Konig, Jaeger et al. 1996). Stehr *et al.* (Stehr, Felk et al. 2004) propose that *C. albicans* might increase hydrophobic interactions by the release of free fatty acids through a high lipolytic activity similar to the black yeast *Hortaea werneckii* (Gottlich, de Hoog et al. 1995). In addition it has been shown in several RT-PCR studies that the expression of some *LIP* genes coincides with infections of different tissues (Kvaal, Lachke et al. 1999; Stehr, Felk et al. 2004; Schofield, Westwater et al. 2005). However, no null mutant of a *LIP* gene with a virulence-associated phenotype has yet been identified (www.candidagenome.org), probably due to the presence of paralogues and to potentially compensatory functions. It thus remains difficult to evaluate the concrete impact of lipases on *C. albicans* virulence.

1.3.3 Yeast-to-hyphae transition and its complex regulation

1.3.3.1 The role of hyphae formation in *C. albicans* virulence

There are two main lines of evidence suggesting that the ability to develop hyphae is important for *C. albicans* during an infection. The first is that hyphae are well adapted to penetrate epithelial surfaces, both actively by the use of mechanical forces (Scherwitz 1982)

and passively by the induction of endocytosis (Scherwitz 1982; Kumamoto and Vinces 2005). The second is the ability to damage endothelial cells and macrophages after their internalization (Lorenz, Bender et al. 2004). In addition to the direct contribution of the hyphal form to *C. albicans* virulence, there seems to be a set of genes which have a hyphae-independent effect on virulence, but which are expressed in response to the same regulatory mechanisms and therefore as well considered as hypha-associated genes (Kobayashi and Cutler 1998; Liu 2002). On the other hand it is important to emphasize that the hyphal form alone is not sufficient for *C. albicans* virulence, as indicated by the attenuated virulence of hyperfilamentous strains as *nrg1ΔΔ*, and it has been postulated that the yeast form might be required for dissemination of *C. albicans* in the bloodstream during systemic infections (Gow, Brown et al. 2002). Taken together, if one specific morphogenetic state is not sufficient to develop full virulence, it seems that the ability to switch between the yeast and the hyphal growth form is essential for pathogenesis of *C. albicans*, and hence huge efforts have been devoted to understand the complex regulation of the transition between these distinct morphogenic states of *C. albicans*.

1.3.3.2 The regulation of the Yeast - to - hyphae transition

Numerous different conditions have been described under which yeast cells are induced to form hyphae *in vitro*. For example, the addition of fetal calf serum (FCS) can rapidly induce blastospores to form hyphae in a rich medium as YPD. It is not completely clear which serum compounds are responsible for this effect, but it has been shown that proline and GlcNAc (N-Acetyl-Glucosamine) are inducers of hyphal formation, while albumine apparently does not play a role. Other environmental cues that induce hyphal formation include nitrogen or carbon starvation, oxygen availability and alkaline pH together with an elevated temperature (Ernst 2000). In contrary, hyphal formation is inhibited for example by the quorum sensing molecule farnesol, thus a high cell density can disfavor hyphal formation (Enjalbert and Whiteway 2005). These environmental signals are not always independent from each other: for example, in YPD medium buffered at alkaline pH no hyphal formation can be observed at 30 °C, and an elevated temperature is not sufficient for hyphal formation at acidic pH. Overexpression of an activated form of the pH-regulated transcription

factor Rim101 can partially bypass this temperature requirement, indicating that the effect of pH and temperature might be additive (El Barkani, Kurzai et al. 2000).

As a result of the crosstalk between different pathways, the regulation of many of their target genes is quite complex and highly variable dependent on the exact environmental conditions. The best example for this complexity might be the transcriptional repressor Efg1 paradigm. Null *efg1* mutants are defective in hyphae formation under most inducing conditions, including serum. Interestingly, this mutant seems to promote hyphae formation under microaerophilic/embedded conditions (Ernst 2000). Thus Efg1 seems to be able to induce as well as to repress filamentation, depending on the respective conditions. As Efg1 acts upstream of Tec1, another transcription factor that promotes filamentation in serum, the regulation under embedded conditions might depend on its crosstalk with another transcription factor, Czf1 (Liu 2001). Figure 12 gives a relatively compact summary of the regulatory network of *C. albicans* filamentation.

The Rim101 pathway and its role in the pH-response and induction of filamentation at alkaline pH will be characterized in detail in the next chapter of this introduction.

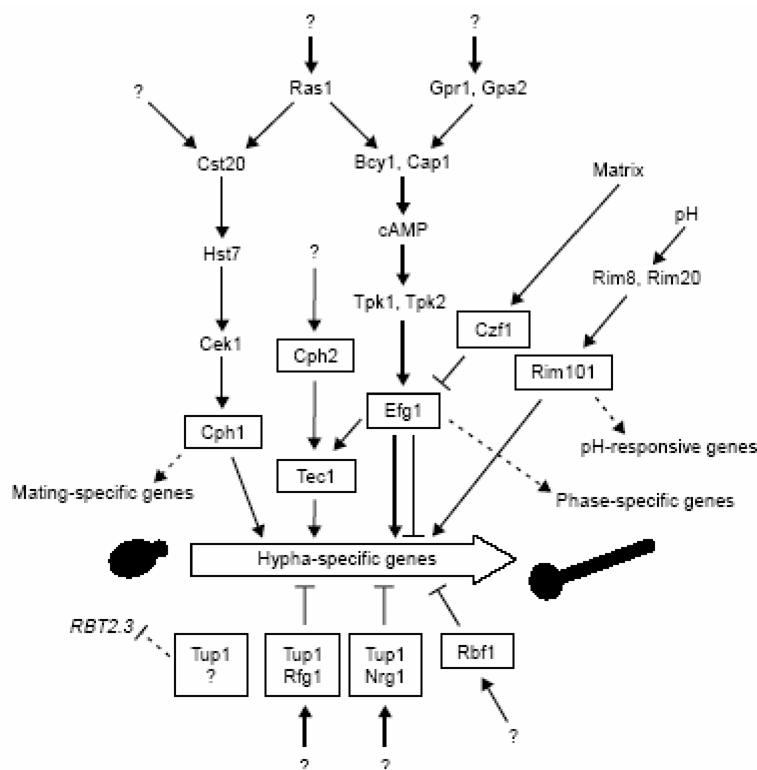


Figure 12: *Simplified regulatory model of the yeast-to-hyphae transition as presented by Liu et al. (Liu 2001). Transcription factors that positively regulate hyphal formation include Cph1, Tec1, Rim101 and Efg1. Efg1 is also involved in the repression of hyphal formation under embedded conditions. Important repressors of filamentation are Rbf1, Rfg1 and Nrg1; the latter two require the recruitment of Tup1 to be fully functional.*

Some of the pathways leading to their activation are already well described (cAMP pathway, MAPK pathway, Rim101 pathway), but their crosstalk is quite complex. For instance, it is not yet fully clear whether Rim101 and Czf1 act in parallel or through Efg1 or whether either ways are possible.

1.4 The conserved fungal pH signaling pathway

The adaptation to ambient pH plays an important role in the life of numerous microorganisms. This is probably even more the case for organisms which inhabit environments with different pH ranges or live in environments where variations in external pH can easily occur. To respond to these different conditions they had to develop and establish during their evolution an efficient system which allows them not only to sense the environmental pH, but also to transduce this signal to the nucleus in eukaryotes, where it induces an appropriate cellular reaction.

Consequently many fungal genes that code for products with an extracellular function or a function at the cell boundary are regulated by the ambient pH; among them we find for example permeases (Bailey, Penfold et al. 1979) and secreted enzymes (Madzak, Blanchin-Roland et al. 1999), but also intracellular proteins that are involved in post-translational modifications of secreted enzymes (Nozawa, May et al. 2003) or in the synthesis of important secreted molecules such as pH-modifying compounds or antibiotics (Espeso and Penalva 1996).

In the fungal kingdom a large number of these genes is apparently regulated by a conserved pH signaling pathway. This pathway has been well characterized in various ascomycetes including *A. nidulans* (Orejas, Espeso et al. 1995), *S. cerevisiae* (Denison, Negrete-Urtasun et al. 1998), *C. albicans* (El Barkani, Kurzai et al. 2000) and *Yarrowia lypolytica* (Lambert, Blanchin-Roland et al. 1997), but evidence exists that it is also conserved in at least one species belonging to the clade of basidiomycetes, *Ustilago maydis* (Arechiga-Carvajal and Ruiz-Herrera 2005).

The pH signaling pathway has been first identified in *A. nidulans* (Caddick, Brownlee et al. 1986), which is also the microorganism where it has been most extensively characterized.

1.4.1 *Aspergillus nidulans*

The history of the investigation of pH signaling in *A. nidulans* reaches back to the year 1965, when Gordon Brown isolated mutants that were defective for phosphatase production in a pH-dependent manner (Dorn 1965; Dorn 1965). He detected these mutants by staining colonies with a α-naphthyl-phosphate/diazonium salt mixture to monitor their phosphatase activity at different pH. He used the prefix "pal" to name mutants deficient for alkaline phosphatase production, and "pac" when the production of acidic phosphatase was affected. By these means he identified amongst others five genes known today as important components of the pH signaling pathway including *palA*, *palB*, *palC*, *palF*, and *pacC* (Dorn 1965; Dorn 1965).

However, it was only in 1986 when Caddick *et al.* revealed that *pal* mutants had several phenotypes that were not related to phosphatase expression, for instance an increased phosphodiesterase activity, a higher sensitivity towards molybdate (Arst and Cove 1970; Arst, MacDonald et al. 1970; Caddick, Brownlee et al. 1986), and an increased use of γ-amino-butyric acid (GABA) (Arst, Bailey et al. 1980; Caddick, Brownlee et al. 1986). These findings together with contradictory phenotypes observed for some *pacC* mutants (acidity- <u>and</u> alkalinity- mimicking *pacC* mutants were isolated) indicated a more general importance of these genes for the pH response of *A. nidulans* with a possible key role for PacC.

Mark Caddick was the first to suggest that *PacC* might code for a transcription factor and that the products of the different *pal* genes identified by Brown might be components of the pathway that results in activation of PacC (Caddick, Brownlee et al. 1986). Although several *pal* genes were later proven to be artifacts (*palE* is an allele of *palB* (Arst, Bignell et al. 1994)) or genes coding for alkaline phosphatases (*palD* and *palG*) rather than constituents of the signaling pathway (Caddick and Arst 1986), the mutant collection of Gordon Brown can be considered today as the key element that led to the discovery of the PacC fungal pH signaling pathway.

The discovery of two additional genes that code for putative membrane proteins with involvement in pH signaling, *palH* and *palI* (Arst, Bignell et al. 1994) completed the set of seven genes forming the PacC signaling pathway as it is known today.

According to the current common perception *palH* and *palI* code for transmembrane sensors that transmit the ambient pH signal into the cell (Denison, Negrete-Urtasun et al. 1998; Negrete-Urtasun, Reiter et al. 1999); PalI has four, PalH seven predicted transmembrane domains. While a *palH* null mutant is essential for growth at pH 8 like the

null mutants of the other *pal* genes, a *palI* null mutant is to some extent leaky and allows residual growth under these conditions. As some C-terminally truncated mutants of *palH* have a similar leaky phenotype, but an additional knockout of *palI* results in a phenotype identical to that of null mutants in *palA*, *palB*, *palC*, *palF* or *palH*, it has been suggested that function of both membrane proteins are additive (Negrete-Urtasun, Reiter et al. 1999).

The roles of PalC and PalF in pH signaling have been unclear for a long time. But recently it was shown that PalF is able to bind to the C-terminal domain of PalH (Herranz, Rodriguez et al. 2005). In addition, they provided evidence that PalF is phosphorylated and ubiquitinylated in a PalH-dependent manner and shares sequence homologies with the metazoan arrestin family. Consequently, they suggested that endocytosis of the PalH/PalF complex as a result of PalF phosphorylation and ubiquitinylation could be a key element in pH signaling.

Almost at the same time it was suggested that PalC function could be as well linked to endocytosis (Tilburn, Sanchez-Ferrero et al. 2005) and could possibly build a link to PalA. The authors found a region of PalC that is homologous to the Bro1 domain, thus this region might have a connection to the endocytosis pathway via AN4240 (the *A. nidulans* homologue of *S. cerevisiae* Snf7p/Vps32p), a component of the ESCRT-III (Endosomal Sorting Complex Required for Transport).

For PalA it has been shown that it interacts not only with this Snf7p-homologue, but that it binds at the same time to the transcription factor PacC, the final element of the pH signaling pathway. This connection is probably established through two short peptide motifs YPXL/I present in PacC, which are both recognized by PalA (Vincent, Rainbow et al. 2003). This binding event is essential to ensure an alkaline processing step of PacC (Vincent, Rainbow et al. 2003) which is necessary for its activation.

Cleavage of PacC takes place in two distinct processing steps. The first cleavage is ensured by PalB, a cystein protease with a catalytic region similar to calpains (Denison, Orejas et al. 1995). PalB is also called the “signaling protease”, because PalB-governed processing only takes place at alkaline pH and in presence of the other *pal* genes of the signaling pathway (Diez, Alvaro et al. 2002). This first cleavage is necessary for the second cleavage, which is probably carried out by a currently unknown protease, the “processing protease”. Indeed, this final PacC processing step becomes pH-independent in an alkalinity-mimicking *pacC* mutant that expresses a truncated version similar to the product of the first truncation (Diez, Alvaro et al. 2002).

A closer view on the different domains of PacC provides the information that leads to a widely accepted working model for the processing of PacC and its fungal homologues. The transcriptionally active region of PacC is a zinc finger region located close to the N-terminus. Three sequence motives susceptible to intramolecular interaction are located more downstream and maintain the protein preferentially in a "closed" formation to prevent processing (Mingot, Tilburn et al. 1999). They are interrupted by the "signaling protease box" where the pH-dependent cleavage takes place. The two flanking regions involve binding sites for PalA (Vincent, Rainbow et al. 2003). PalA is recruited to these sites only at alkaline pH when signaling takes place. Binding of PalA is necessary for the conversion of the closed form of PacC to an open form that permits the first processing by PalB resulting in the removal of a C-terminal interacting domain, which in turn allows the processing protease to access PacC and to cleave it to its active form (Penalva and Arst 2004). In fact, it seems that both closed and open form of full-length PacC coexist in an equilibrium at acidic and at alkaline pH; the crucial difference is that this equilibrium strongly favors the closed form at acidic pH, while at alkaline pH the equilibrium is shifted towards the open accessible form, a step which is most probably catalyzed by the binding of PalA (Espeso, Roncal et al. 2000).

Interestingly, the closed conformation of PacC prevents PacC processing, but not *in vitro* binding of PacC to its target genes, indicating that full-length PacC may be functional. The explanation for this apparent antagonism is that full-length PacC, although binding competent, is not be able to reach its targets because its localization is preferentially cytosolic, while truncated forms are mainly nuclear. While the closed conformation is distributed all over the cell, both truncated forms are located preferentially in the nucleus, as has been shown with help of PacC-GFP fusions by Mingot *et al.* (Mingot, Espeso et al. 2001). Furthermore, they provided evidence that truncated PacC contains a nuclear import signal that is probably hidden in the full-length form (Mingot, Espeso et al. 2001).

The DNA binding domain of PacC has already been characterized in detail. It is formed by three Cys_2His_2 zinc fingers which are able to recognize the specific sequence motif (T)GCCARG on PacC target promoters (Tilburn, Sarkar et al. 1995; Espeso, Tilburn et al. 1997). Furthermore it has been shown that the DNA interaction is restricted to finger 2 and 3, while finger 1 interacts with finger 2 rather than with promoter DNA (Espeso, Tilburn et al. 1997).

While its activity is restricted to neutral to alkaline pH values, PacC can act as an activator of alkaline-expressed genes as well as a repressor of acidic-expressed genes. Most of these genes code for proteins that are localized at the cellular surface or extracellular. Among

the alkaline-induced genes we find *acvA* and *ipnA*, two enzymes with an early function in penicillin biosynthesis; binding of PacC to their common promoter has already been characterized in detail (Espeso, Tilburn et al. 1997; Then Bergh and Brakhage 1998). PacC also significantly contributes to the increase of siderophore biosynthesis and uptake of siderophores (Eisendle, Oberegger et al. 2004). The alkaline protease PalD was identified as part of the *pal* mutant collection of Gordon Brown, its PacC-dependent regulation was shown by Caddick *et al.* (Caddick, Brownlee et al. 1986). The genes coding for the alkaline protease PrtA and the xylanase XlnA are also PacC-induced (de Graaff, van den Broeck et al. 1994; MacCabe, Orejas et al. 1998; Katz, Masoumi et al. 2000; vanKuyk, Cheetham et al. 2000). Finally, the *pacC* gene itself is preferentially expressed at alkaline pH and seems to be subjected to autoregulation to reinforce pH signaling (Tilburn, Sarkar et al. 1995). There is currently no evidence that other components of the pH signaling pathway in *A.nidulans* are pH-regulated.

Acidic genes which are repressed by PacC at alkaline pH include *pacA* (Caddick, Brownlee et al. 1986), which was already part of Brown's *pac* mutant collection, the acidic xylanase *xlnB* (MacCabe, Orejas et al. 1998; Perez-Gonzalez, van Peij et al. 1998), an acid phosphodiesterase (Caddick, Brownlee et al. 1986) and *abfB,* coding for an arabinofuranosidase (Gielkens, Gonzalez-Candelas et al. 1999). In addition, PacC-mediated repression of the GABA transporter has been characterized in detail (Hutchings, Stahmann et al. 1999; Espeso and Arst 2000) and there is evidence for a PacC-regulated molybdate permease expressed preferentially at acidic pH (Arst and Cove 1970; Arst, MacDonald et al. 1970).

1.4.2 *Saccharomyces cerevisiae*

As previously mentioned, the Pal pH signaling pathway is conserved in other fungal species and it has been well described in particular in yeasts. Homologues of all *Pal* genes have been identified in *S. cerevisiae*, with the exception of the *PalC* gene. With respect to their original isolation in a screen for genes involved in meiosis (Su and Mitchell 1993) yeast homologues of *Pal* genes are designated *RIM* (Regulator of Inducer of Meiosis) genes in *S. cerevisiae* and other yeasts, and the *Pal* pathway is consequently named *Rim* pathway (Figure 13: the transmembrane proteins homologous to PalI and PalH are denominated Rim9p and Rim21p respectively, the yeast orthologue of PalF is Rim8p, PalA corresponds to Rim20p,

Rim13p is the orthologue of the signaling *Aspergillus* protease PalB, and the zinc finger transcription factor PacC is represented under the name Rim101p in yeast proteomes.

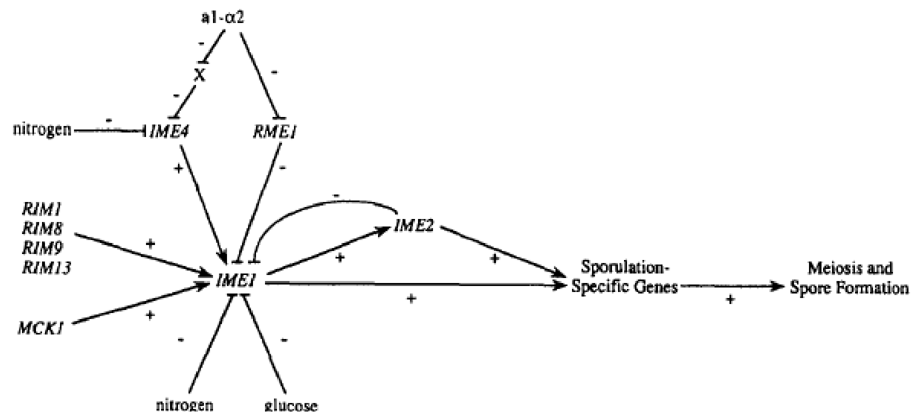


Figure 13: *The first characterization of the S. cerevisiae orthologue of AnPacC which lead to the current nomenclature of all yeast orthologues (Su and Mitchell 1993): ScRIM1 (= ScRIM101) and pathway components ScRIM8, ScRIM9 and ScRIM13 have been identified as indirect regulators of the meiosis via Ime1 (RIM = Regulator of Inducer of Meiosis).*

As the pH signaling pathway has been already described in detail for the *A. nidulans* model, only differences and additional information will be mentioned here.

Recently a third predicted integral transmembrane protein, Dfg16, has been shown to be required for the activation of Rim101p (Barwell, Boysen et al. 2005). The *A. nidulans* orthologue of *DFG16* is *PalH*, which is at the same time the closest orthologue of the previously identified membrane protein Rim21p (Barwell, Boysen et al. 2005). Both Rim21p and Dfg16p are predicted to have seven transmembrane domains, and Dfg16p has a long C-terminal tail and a signal peptide, which indicates a similarity to G-protein-coupled receptors (GPCR). Given that Rim8p is homologous to arrestin-like GPCR-interacting proteins, it was suggested that the pH signal might be received and transmitted by Dfg16p to a yet unknown G-protein that interacts with Rim8p, and the function of Rim21p might be to act in concert with Dfg16p in form of a heteromeric complex (Barwell, Boysen et al. 2005).

Another difference with *A. nidulans* seems to be the presence of a single processing step (Li and Mitchell 1997; Xu and Mitchell 2001). The final product of PacC differs by 400 aa residues from its full length form, while the activated Rim101p is only 70 residues shorter than its full length form (Li and Mitchell 1997). Whether this different processing reflects a distinct function of the active form of the transcription factor, is currently not sure, but it has been suggested that *S. cerevisiae* Rim101p might act primarily as a direct transcriptional repressor and execute its inducing function indirectly through the repression of the repressors Nrg1p and Smp1p (Lamb and Mitchell 2003).

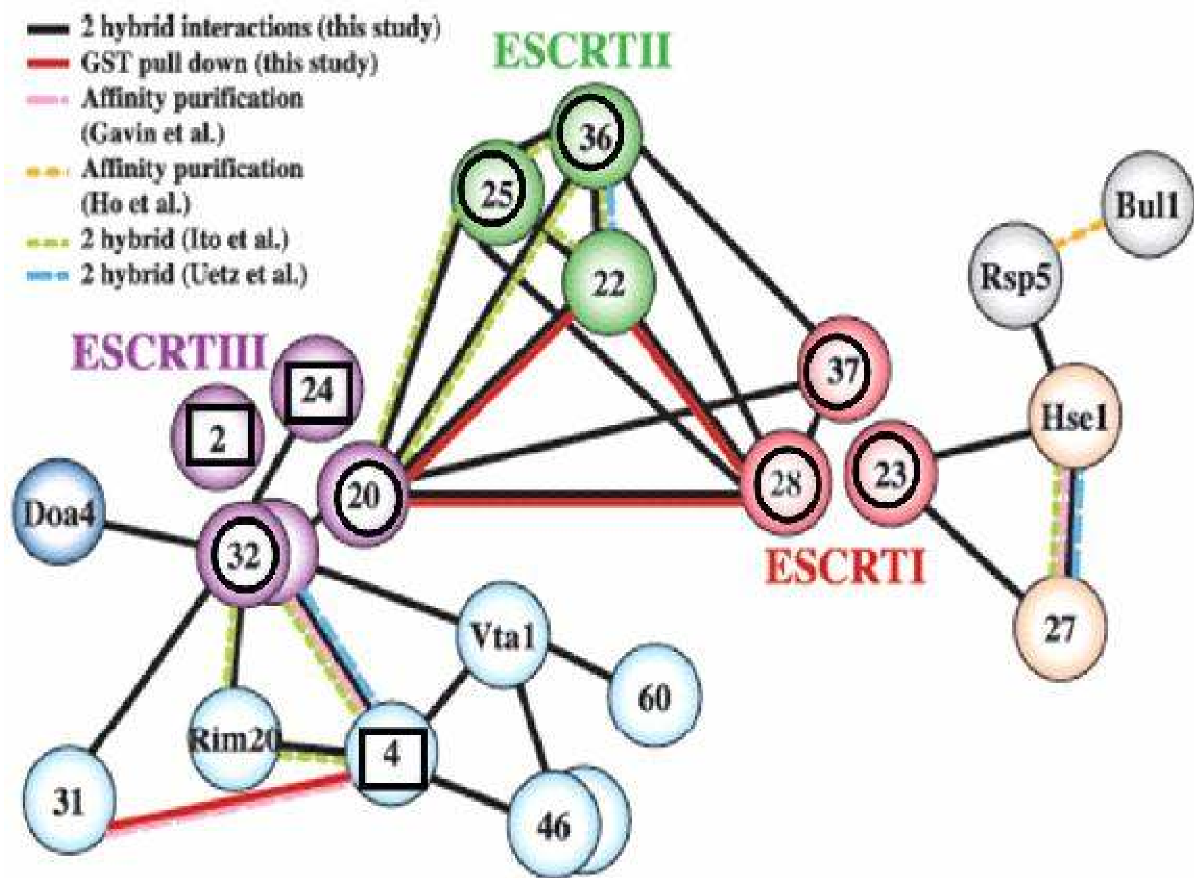


Figure 14: Overview about published two-hybrid interactions of ESCRT complex components: A clear link has been established between the ESCRT complexes and the Rim101 in S. cerevisiae pathway through several yeast-two hybrid studies. Rim20 interacts as well with Vps32/Snf7 as with Vps4, which are two components of the ESCRTIII complex (An additional interaction between Snf7 and Rim13 is not shown here). Mutants in genes of the three ESCRT complexes that are defective for alkaline Rim101 processing are marked with a black circle, mutants in genes of the ESCRTIII complex that show an alkanity-mimicking phenotype are marked with a black square. This figure has been adapted from Bowers et al. (Bowers, Lottridge et al. 2004)

S. cerevisiae is also the fungal model where the clearest link between endocytosis and the Rim101p pathway has been established. Yeast-two-hybrid interactions have been shown to exist between Rim20p and two proteins of the ESCRT-III (Figure 14), Vps4p and Vps32p (also known as Snf7p), as well as between the signaling proteinase Rim13p and Vps32p. Mutants in multiple genes of different ESCRT complexes have been shown to be defective for alkaline Rim101p processing; this is at least true for *VPS23*, *VPS28*, *SRN2/VPS37* (all ESCRT-I), *VPS36*, *VPS25* (both ESCRTII), *VPS20* and *SNF7/VPS32* (ESCRTIII) (Xu, Smith et al. 2004). On the contrary, another study recently showed that three other ESCRT-III proteins, *VPS2*, *VPS4* and *VPS24* are required to prevent processing of Rim101p under acidic conditions and can bypass the processing defects of mutations in some upstream components of the Rim101p signaling pathway, including *RIM9*, *RIM21* and *RIM8* (Hayashi, Fukuzawa et

al. 2005). Thus, there is clearly a link between the ESCRT-III and the Rim101p signaling pathway, and it seems that this interaction takes place at the signaling protease step, as direct interactions could be shown for Rim13p and Rim20p and as some mutations in upstream components of the pathway could be bypassed. However, the exact role of the late endosome in the pathway has yet to be unrevealed.

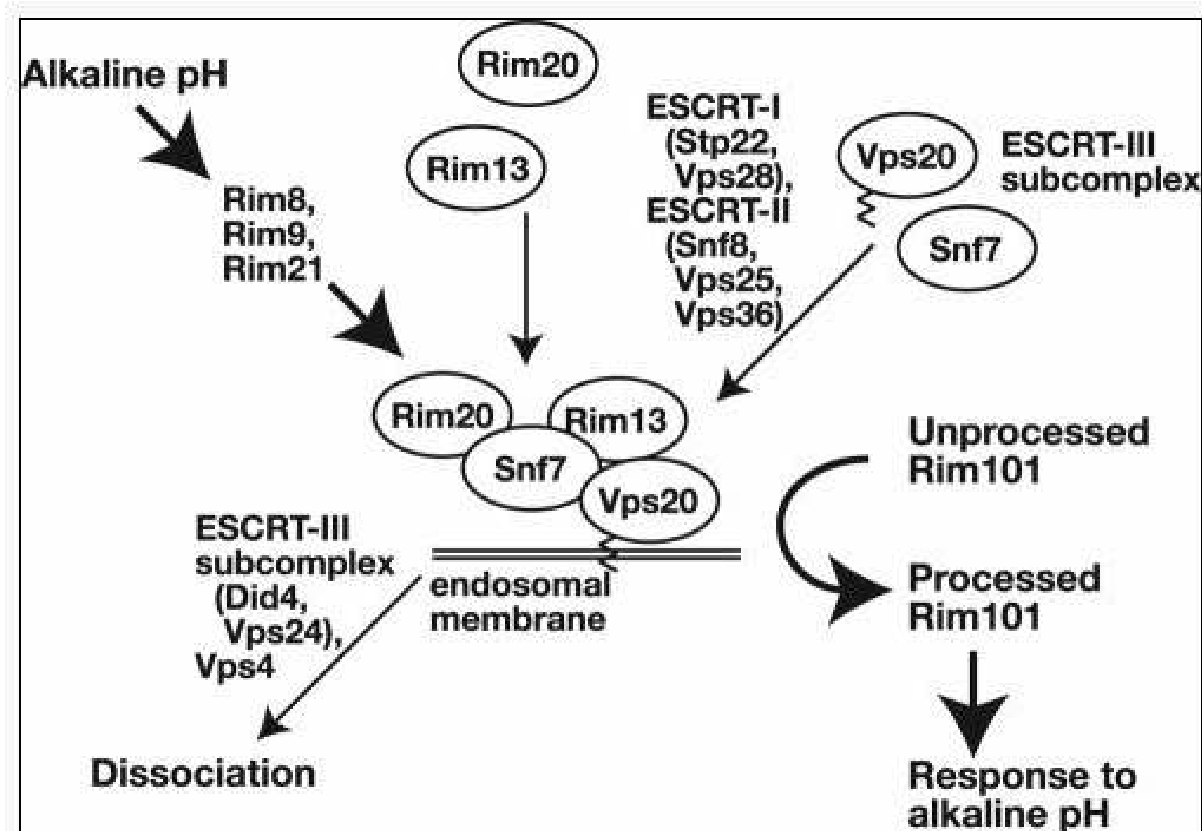


Figure 15: *Model for the role of the ESCRT complexes in the activation of of Rim101 as proposed by Hayashi et al. (Hayashi, Fukuzawa et al. 2005): Components of all three ESCRT complexes are required for the correct alkaline cleavage of Rim101. A central role is played by the association of Rim101/Rim20/Rim13 with the ESCRT-III components Snf7/Vps32 and Vps20 which is necessary for correct processing of Rim101. The ESCRT complexes I and II would then in some way be responsible for the correct assembly of this complex, while a subcomplex of ESCRT-III including Vps4, Vps2 and Vps24 might permit MVB sorting, but is not involved in the formation of the complex. Mutants in genes coding for ESCRT-I and II components as well as for Vps20 and Snf7 are thus expected to have an acidity-mimicking phenotype (correct complex formation is hampered), while mutants in VPS2, VPS4, VPS24 mimick alkalinity, because the dissociation of the activating complex is disturbed leading to constitutive activation of Rim101.*

Hayashi *et al.* proposed an interesting model (Figure 15) that includes a central processing event for which a protein complex Rim101p/Rim20p/Rim13p/Snf7p is required. The stabilization of this complex might be favored at alkaline pH, which could be explained

by a requirement for acidification of the endosomal lumen for MVB (MultiVesicular Body) sorting (Matsuo, Chevallier et al. 2004). This would explain the acidity-mimicking phenotype of mutations in some ESCRT-proteins, which would simply be required in some way for the formation of the "processing complex". Finally, the explanation for alkalinity-mimicking phenotype in mutants of *VPS4*, *VPS2* and *VPS24* could be that these proteins are required to form a subunit which is necessary for MVB sorting, but not for formation of the processing complex. These mutants might thus be blocked in MVB sorting after the formation of the "processing complex", resulting in the constitutive activation of Rim101p independently of the pH (Hayashi, Fukuzawa et al. 2005).

As already mentioned, *S. cerevisiae* Rim101p seems to act primarily as a repressor, in contrast to *A. nidulans* PacC. In a transcription profiling experiment, Lamb *et al.* compared a *rim101* mutant with a reference strain and identified 17 genes that were more than 2-fold upregulated in the mutant, indicating repression by Rim101p in the reference strain. In addition, they found 18 genes that were at least 2-fold repressed in the mutant, indicating their Rim101p-dependent upregulation in the reference strain. However, when they analyzed a 600 bp upstream region of their promoters, they found putative Rim101p binding sites in most promoters of Rim101p-repressed genes, but in none of the induced genes. Chromatin IP (Immuno Precipitation) experiments confirmed the capacity of Rim101p to bind several of these promoter regions. As the expression of the two transcriptional repressors, Nrg1 and Smp1 was found to be repressed by Rim101p, they suggested that Rim101p might act as an inducer only indirectly by causing a derepression of Nrg1p and Smp1p targets through the repression of these repressors (Lamb and Mitchell 2003). At least two of the genes that are induced in a Rim101p-dependent manner, *ENA1* and *ZPS1* (homologue to *C. albicans PRA1*), are known target genes of Nrg1p, another one, *CWP1*, is a target of Smp1p. Ena1p is a Na^+ efflux pump, which might explain the increased Na^+ sensitivity of *RIM* mutants, and Smp1p seems to be a repressor of invasive growth, which can explain the defect of a *rim101* null mutant to invade agar plates (Lamb and Mitchell 2003). Interestingly Rim101p is not only a repressor of *NRG1* transcription, but can also act as a co-repressor together with Nrg1p, as it has been shown for the common promoter region of *DIT1/2* genes (Rothfels, Tanny et al. 2005). Besides, many genes coding for proteins with a predicted function in the cell wall were regulated in a Rim101p-dependent way, as were some membrane proteins and iron transporters. The important function of the pH response in the yeast cell wall assembly is underlined by the hypersensitivity of *RIM* mutants against agents such as calcofluor white, caffeine and zymolyase, which can be reverted by a truncated constitutively active version of

Rim101p (Castrejon, Gomez et al. 2006). In addition Rim101p and Rim21p are synthetic lethal with Slt2p, a MAP kinase of the PKC signal transduction pathway with a similar mutant phenotype (Castrejon, Gomez et al. 2006). It is important to mention that Rim101p governs only a part of the pH response of *S. cerevisiae*, a good illustration of that is the complex regulation of the Na^+ pump Ena1p, in which also Snf1p and calcineurin signaling are involved in addition to Rim101p (Platara, Ruiz et al. 2006).

Finally, Rim101p seems to autoregulate itself in a more complex way than PacC. Like in *A. nidulans*, the transcription factor induces its own expression at alkaline pH, but additionally it represses the transcription of the upstream component *RIM8* (Lamb and Mitchell 2003). A possible explanation for this paradigm would be that Rim101p prevents the accumulation of its activated form.

1.4.3 *Yarrowia lipolytica*

Yarrowia lipolytica is a nonconventional dimorphic fungus that belongs to the ascomycetes. The genome of this preferentially haploid organism has been entirely sequenced during the "Genolevures 2" project (http://cbi.labri.fr/Genolevures/about/GL2_intro.php). Its secretion capacities have been studied extensively and led to some industrial applications, but it is as well used as a model to study the use of hydrophobic substrates, peroxisome biogenesis, mitochondrial complex I biogenesis, morphogenesis control, and also pH adaptation (Kerscher, Drose et al. 2002; Madzak, Gaillardin et al. 2004).

The Rim101p pH signaling pathway is conserved in *Yarrowia lipolytica*, and its Rim101p homologue has been identified together with some upstream components (Rim21p and Rim8p) in a screen for expression regulators for the secreted protease *XPR2* (Lambert, Blanchin-Roland et al. 1997). In further studies homologues of all other components of the Rim101p pathway have been identified. Finally there is also evidence for the presence of a PalC homologue in this yeast (Tilburn, Sanchez-Ferrero et al. 2005). *Y. lipolytica* separated early from the main hemiascomycete line (see figure 1; (Dujon, Sherman et al. 2004)) and might have conserved an ancestral PalC that has been consecutively lost in most other hemiascomycetes.

As in *S. cerevisiae*, it has been shown that components of the ESCRT-I machinery are clearly involved in the regulation of Rim101p activity (Blanchin-Roland, Da Costa et al. 2005). The same study shows that, in contrast to *S. cerevisiae*, Snf7p seems to be essential in

Yarrowia lipolytica, which might explain why the Snf7p orthologue was not isolated by mutagenesis in *A. nidulans*. The analysis of a *VPS4* mutant revealed that it is not required for Rim101p processing, but showed a slight alkalinity-mimicking phenotype at acidic pH, an observation also made by Hayashi *et al.* in *S. cerevisiae* (Hayashi, Fukuzawa et al. 2005); this included a growth defect at acidic pH similar to the one observed for a strain with a constitutively activated form of Rim101p and the induction of normally alkaline-expressed genes *XPR2* and *PHR2* at acidic pH compared to the wildtype reference.

Initially it had been shown that the pH signaling pathway played a role in mating and sporulation, similarly to Rim101p in *S. cerevisiae* (Lambert, Blanchin-Roland et al. 1997). On the other hand it has been demonstrated that in contrast to *C. albicans*, YlRim101p is not involved in pH-dependent filamentation (Gonzalez-Lopez, Ortiz-Castellanos et al. 2006). Little is known about *Y. lipolytica* genes that are under the control of Rim101p, as no global transcription profiling experiment has been performed so far. *XPR2* codes for an alkaline protease and its alkaline induction is clearly Rim101p-dependent. It has been shown by *in vivo* DMS (Di-Methyl Sulfate) footprinting that Rim101p is able to bind to a PacC binding motif on the *XPR2* promoter (Blanchin-Roland, Cordero Otero et al. 1994). Another alkaline-induced gene is the homologue of *C. albicans PHR1* (Gonzalez-Lopez, Szabo et al. 2002), which codes for a GPI-anchored surface protein with glycosidase function (Ghannoum, Spellberg et al. 1995). The regulation of genes with acidic expression seems to be more complex. *AXP1* codes for a secreted protease which is more transcribed at acidic than at alkaline pH. A deletion of components of the Rim101p pathway does not lead to derepression of *AXP1* transcription at alkaline pH, but surprisingly results in a weaker transcription at acidic pH, which might indicate that Rim101p acts as an inducer at acidic pH in *Y. lipolytica* in contrast to observations made in other organisms (Gonzalez-Lopez, Szabo et al. 2002).

1.4.4 *Candida albicans*

1.4.4.1 A link between the fungal pH response and virulence

The PacC/Rim101p-dependent pH response has important functions in the pathogenesis of several fungal organisms. This is true for different plant pathogens, including *Fusarium oxysporum* (Caracuel, Roncero et al. 2003), *Sclerotinia sclerotiorum* (Rollins 2003), and *Fusarium verticillioides* (Flaherty, Pirttila et al. 2003). Links between pH-dependent gene regulation and virulence have also been established for various other fungal plant pathogens (Prusky, McEvoy et al. 2001; Eshel, Miyara et al. 2002).

Intriguingly, for human fungal pathogens the conserved pH response also plays an important role in pathogenesis. Recently it has been shown that PacC is involved in the regulation of virulence factors in pulmonary aspergillosis (Bignell, Negrete-Urtasun et al. 2005), but the human fungal pathogen with the most extensively studied and best described pH response is probably *C. albicans*.

1.4.4.2 The role of Rim101p in *C. albicans* virulence

C. albicans is able to grow over a large pH range from pH2 to pH10 (Odds 1988) and can colonize host niches with very different pH (blood pH 7.4, oral cavity pH 6-6.5, vaginal and skin pH 5.5, gut pH 2-6). Depending on the environmental pH different genes are expressed which contribute to *C. albicans* virulence. For example, the presence of the pH-regulated glycosidases Phr1p and Phr2p has been shown to be critical for virulence in different models. While *PHR1* is essential for virulence in a hematogenously disseminated blood model (Ghannoum, Spellberg et al. 1995), thus under alkaline conditions, its functional homologue *PHR2* is required in vaginal and stomach models of infection, where an acidic pH is prevalent (De Bernardis, Muhlschlegel et al. 1998). Moreover, the ectopic expression of either of them can rescue the virulence defect of these mutants (De Bernardis, Muhlschlegel et al. 1998). Although these effects might be simply explained by the reduced growth rates of these mutants under conditions that require expression of the missing gene, it shows clearly that the capacity to adapt gene expression to different pH conditions is crucial for the success of *C. albicans* as a pathogen.

In another study it has been shown that the *C. albicans* homologue of *ScRIM101* is required for full pathogenesis (Davis, Edwards et al. 2000). This result cannot be explained by the influence of Rim101p on *PHR* genes regulation alone, because even if the alkaline transcription of *PHR1* is strongly reduced in a *rim101* null mutant, this should be compensated by the simultaneous derepression of its functional homologue *PHR2*. Thus, other genes under control of Rim101p might be responsible for the virulence defect.

1.4.4.3 Regulation of Rim101p activity

Rim101p was identified in 1999 by Fonzi *et al.* and originally named Prr2p (pH Response Regulator; (Ramon, Porta et al. 1999), but its name was soon changed to follow the *S. cerevisiae* nomenclature (Davis, Wilson et al. 2000).

C. albicans homologues of all *Pal* genes except *PalC* have also been isolated, and their function seems to be conserved within the different fungal species. Recently a second predicted transmembrane protein homologue to *S. cerevisiae* Dfg16p and *A. nidulans* PalH with conserved function in the pH response has been found. Moreover the link between endocytosis and pH signaling has been established similarly to *S. cerevisiae* (Xu, Smith et al. 2004), including upstream functions for ESCRT-I proteins Vps23p and Vps28p, ESCRT-II protein Vps36p and the ESCRT-III protein Snf7p. In contrary, the presence of ESCRT-III proteins Vps2p, Vps4p and Vps24p does not favor Rim101p processing. Thus, there are currently no indications for functional differences of ESCRT-components in *C. albicans* compared to *S. cerevisiae*, and it can be assumed that the interaction with the Rim101p signaling pathway is fully conserved in these species.

Consequently, deletions of ESCRT-components *CaVPS23*, *CaVPS28*, *CaVPS36* and *CaSNF7* (Xu, Smith et al. 2004) as well as Rim101p-pathway components *CaDFG16* (Barwell, Boysen et al. 2005), *CaRIM21* (Davis 2003), *CaRIM8*, *CaRIM20*, *CaRIM101* (Davis, Wilson et al. 2000) and *CaRIM13* (Li, Martin et al. 2004) lead to the expected acidity-mimicking phenotypes and growth defects, and the reinsertion of truncated versions of Rim101p that mimic the C-terminal processing event can restore both growth and pH-dependent filamentation of these mutants at alkaline pH (Davis, Wilson et al. 2000; Xu, Smith et al. 2004; Barwell, Boysen et al. 2005).

1.4.4.4 Functionality of Rim101p

The sequence of Rim101p includes 661 amino acid residues, but due to the presence of an internal methionine residue the sequence initially proposed by Davis *et al.* was 58 codons shorter at the N-terminus and led to a persistent confusion between the coding sequences available in the two web databases http://genolist.pasteur.fr/*Candida*DB/ and http://www.*Candida*genome.org/. The similarity between AnPacC, ScRim101p and CaRim101p is mainly concentrated on the tridactyl zinc finger regions and much less obvious in the rest of the sequence (Penalva and Arst 2002). Nevertheless it seems that the sequence binding motif recognized by CaRim101p is not identical to that of its homologues. It has been demonstrated that the promoter binding motif of PacC in *A. nidulans* is 5'-GCCARG-3' with a preference of an "A" in position 5 on both induced and repressed genes (Tilburn, Sarkar et al. 1995; Espeso, Tilburn et al. 1997). *In vitro* binding essays confirmed this result for several related species, including *Acremonium chrysogenum* (Schmitt, Kempken et al. 2001) and *Penicillium chrysogenum* (Suarez and Penalva 1996), and evidence for conservation of these sites also come from the Rim101p orthologues of *Y. lipolytica* (Madzak, Blanchin-Roland et al. 1999) and *S. cerevisiae* (Lamb and Mitchell 2003). However, the promoter of the *C. albicans PHR1* gene does not possess such an extended site, but it has been demonstrated that Rim101p recognizes a shorter sequence motif 5'-CCAAG-3' (with preferences for three additional A at the 3' end) more efficiently than a PacC motif (Ramon and Fonzi 2003). Recently it has been shown that Rim101p is able to bind to an extended motif 5'-GCCAAGAA-3' on the promoter of *PHR2*, which includes both previously suggested binding motifs. However, the concrete binding specifities seem to be promoter-dependent and do not necessarily always include the complete motif (Baek, Martin et al. 2006). The initial "G" was shown to be dispensable on one site on the *PHR2* promoter as already found for the *PHR1* promoter, but on the second binding site (which seems to be the most important for transcriptional regulation), mutation in this position resulted in significant reduced Rim101p binding capacities (Baek, Martin et al. 2006). The extended motif was also found to be prevalent within the promoters of Rim101p-dependent alkaline-induced genes identified by microarray analysis (Bensen, Martin et al. 2004), indicating that CaRim101p may act directly (and perhaps preferentially) as an inducer in contrast to the situation in *S. cerevisiae*, where Rim101p has been proposed to function primarily as a repressor. Since with *PHR2* at least one example for a directly Rim101p-repressed repressed gene exists, the function of CaRim101p seems to be closer to that of *A. nidulans* PacC than to that of ScRim101p.

Strain Name	Allele Name	Length [aa]	Mutant Background	Acidic pH Phenotype			Lit. Ref.
				***PHR* gene regulation**	**Growth**	**Hyphae**	
None	None	280	*phr2* null	?	No	No	1*
None	None	304	*phr2* null	?	No	No	1*
None	None	332	*phr2* null	?	No	No	1*
None	None	384	*phr2* null	?	Yes	No	1*
None	None	410	*phr2* null	?	Yes	No	1*
MC13	Rim101SL	414	*rim101* null	*PHR1* & *PHR2*	Yes	Yes	2*
Rim101-405	Rim101-405	461	*rim101* null	*PHR1* & *PHR2*	Yes	No	3*
None	None	462	*phr2* null	?	Yes	Yes	1*
CEM-1	RIM101-1426	475	*phr2* null	*PHR1*	Yes	Yes	4*
CAF3-16-1	RIM101-1426	475	wild type	*PHR1* & *PHR2*	Yes	Yes	4*
CAPR1-6	None	557	*rim8* null	*PHR1* & *PHR2*	Yes	Yes	5*
CAPR1-2	None	568	*rim8* null	*PHR1* & *PHR2*	Yes	Yes	5*
CAPR1-8	None	575	*rim8* null	*PHR1* & *PHR2*	Yes	Yes	5*
CAPR1-4	None	579/580	*rim8* null	*PHR1* & *PHR2*	Yes	Yes	5*
CEM-2	RIM101-1751	583	*phr2* null	*PHR1*	Yes	Yes	4*
CAPR1-1	None	584	*rim8* null	*PHR1* & *PHR2*	Yes	Yes	5*
CAPR1-3	None	586	*rim8* null	*PHR1* & *PHR2*	Yes	Yes	5*

1* Mühlschlegel et al., unpublished (Penalva and Arst 2002)

2* (Cornet, Bidard et al. 2005)

3* (Davis, Wilson et al. 2000)

4* (El Barkani, Kurzai et al. 2000)

5* (Porta, Wang et al. 2001)

Table 2: *Summary of C. albicans strains expressing C-terminally truncated versions of Rim101p. They are ranked by the length of the truncated protein they code for. The observed phenotypes suggest that a minimal length of about 410 amino acids acids is required to be functional.*

Several truncated versions of Rim101p have been described in the literature, a summary can be seen in Table 2. *RIM101-405* is a shortened *RIM101* allele (truncated after Asn462) that was constructed through mutagenesis by Davis *et al.* and was used to complement the alkaline filamentation defect in different pH signaling pathway mutants. However, for this truncated form no filamentation at acidic pH could be observed, indicating that its functionality might be restricted to alkaline pH values (Davis, Wilson et al. 2000; Barwell, Boysen et al. 2005).

Dominant alleles that bypass the pH restriction of filamentation have been isolated from *PHR2* mutants by El Barkani *et al.* (El Barkani, Kurzai et al. 2000). *PHR2* mutants are not able to grow under acidic conditions. However, in revertants of these mutants, the paralogue of Phr2p, Phr1p, was shown to be expressed as a result of truncating mutations in *RIM101* named *RIM1426* (truncated after Gln476) and *RIM1751* (truncated after Ser583). The

pH-dependent filamentation was demonstrated to be still dependent on both temperature 37° and Efg1p activity, indicating that *EFG1* might be a downstream target of Rim101p activity. However, the restrictive temperature could be lowered by multicopy insertions of *RIM1426* (still in a *phr2ΔΔ* background) to 29 °C (El Barkani, Kurzai et al. 2000).

In a similar way, Porta *et al.* could isolate revertants that rescued alkaline filamentation in a *rim8* null mutant. Their phenotype was indistinguishable from the revertants of El Barkani *et al.*, they filamented at acidic pH and constitutively expressed *PHR1* and repressed *PHR2*. All mutations were heterozygous dominant nonsense mutations resulting in truncations between Rim101p residues 557 and 586 (Porta, Wang et al. 2001).

Finally, another truncated version of Rim101p with only 415 residues was independently created based on hydrophobic clustering predictions. Integrated in the *RIM101* locus of a normally nonfilamentous *rim101* null mutant and several other mutants with impaired pH signaling, the resulting truncated protein could induce filamentation at alkaline and at acidic pH and restored at least partially the alkaline expression patterns of *PHR1* and *PHR2* (Cornet, Bidard et al. 2005).

Mühlschlegel *et al.* constructed a series of early C-terminally truncated versions of Rim101p between codons 280 and 463 and found that 384 residues were sufficient to repress *PHR2* transcription, but that up to 410 residues were not sufficient to promote filamentation at acidic pH. Given that the before cited 415-residue version was clearly able to do so, this truncation might be close to the minimal functional form of Rim101p (See Table 2).

The *in vivo* processing event has been assessed by Li *et al.* by integrating V5-tagged versions of Rim101p in the *HIS1*-locus of a *rim101* null mutant strain. These versions were tagged after residues 17, 348 and 436 of the coding region. Processed alkaline forms of Rim101p were detectable by Western Blotting for both the 17- and the 436-tagged form, indicating that the active wildtype form of Rim101p comprises at least 436 residues and is intact at the N-terminus. This is true provided that the tagged versions of Rim101p are cleaved like the wildtype form, which is probable given that they were not cleaved in *RIM* mutants but apparently fully functional after processing as they could restore both alkaline growth and filamentation in a *rim101* null mutant background (Li, Martin et al. 2004). However, a distinct processing event resulting in an even shorter form of Rim101p was reported to take place at acidic pH, which has not been reported for orthologues of CaRim101p. So far it could not been shown whether this shortened form is functional. There are few studies reporting phenomena that might be explained by a function of Rim101p at acidic pH: Nobile *et al.* found that an intact Rim101p pathway is necessary for chlamydospore

formation at acidic pH (Nobile, Bruno et al. 2003). The contribution of Rim101p to the resistance towards hygromycin and LiCl seems to be pH-independent (Li, Martin et al. 2004), and microarray studies also suggest a limited activity of Rim101p at acidic pH (Bensen, Martin et al. 2004). Finally, the *Y. lipolytica* Rim101p homologue was found to be responsible for acidic induction of *AXP1* (Gonzalez-Lopez, Szabo et al. 2002). However, the reports about acidic function of CaRim101p and its orthologues are very limited compared to the regulatory events reported at alkaline pH and they still lack experimental validation.

1.4.4.5 Genes under the control of Rim101p

Many studies have proven the importance of pH-dependent gene regulation in *C. albicans*. A search in the *Candida* genome data base (www.*Candida*genome.org) using the key words "Rim101, Rim, pH, acidic and alkaline" reveals 126 gene annotations that indicate a role of pH in their transcriptional regulation, including an almost equal number of alkaline induced (64) and repressed genes (62) (See Table 3 in the end of this chapter). Rim101p is reported as the transcriptional regulator for 24 of the induced and 19 of the repressed genes, thus roughly for a third of all pH-regulated genes. The majority of these genes (94) has been annotated as a consequence of the microarray results of Bensen *et al.* (Bensen, Martin et al. 2004).

In this work a whole genome array was used to compare global transcriptional events in a *rim101* knockout strain and a wild type strain at pH 4 and pH 8 after four hours of growth in M199 medium at 37 °C.

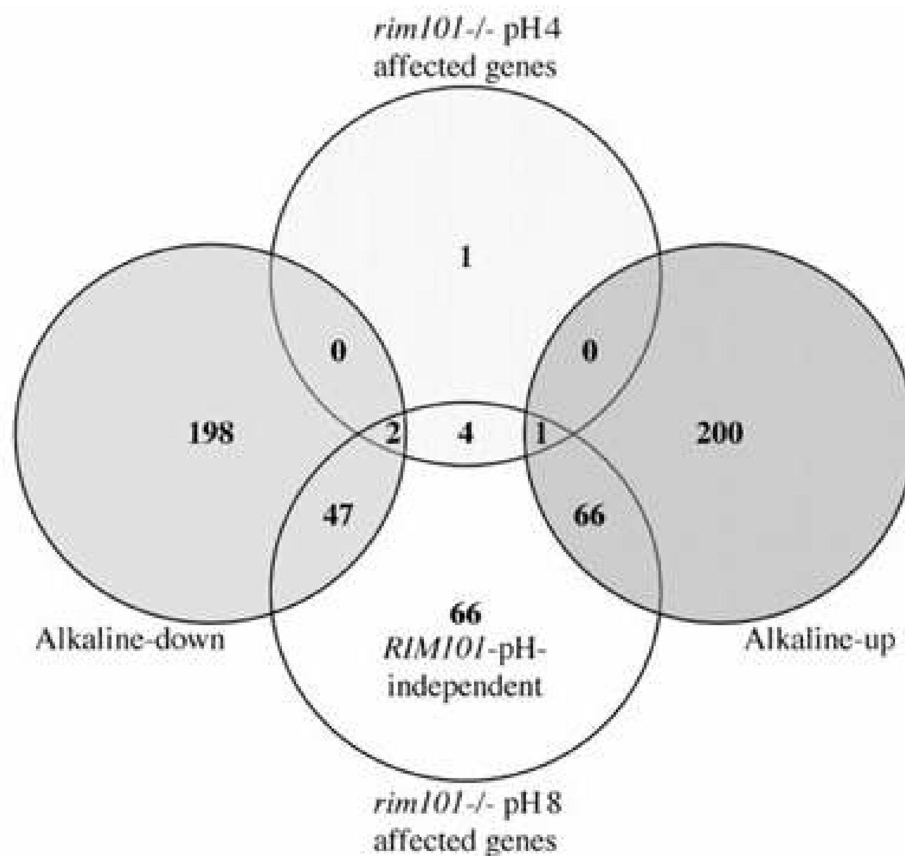


Figure 16: *Venn diagram summarizing the microarray results of Bensen et al. (Bensen, Martin et al. 2004). Rim101 seems to be responsible for only a subset of pH-dependent regulated genes, 200 genes remain induced and 198 remain repressed at alkaline pH when Rim101 is not present. A considerable impact of the presence or absence of Rim101 on gene transcription can only be seen at alkaline pH (186 genes regulated differently in a rim101 null mutant), while at acidic pH gene transcription is not considerably changed (only 8 regulated genes). For a total of 71 genes a pH-independent regulatory function of Rim101 was suggested.*

They identified 514 genes with more > 2-fold transcriptional change in the presence of Rim101p; 247 of them were down- and 267 were upregulated at alkaline pH (Figure 16). The global transcription of the *rim101* knockout strain was quite similar to the wild-type strain at acidic pH, with > 2-fold changes only in the transcription of 8 ORFs, which was comparable with the changes observed due to random variation between two wild type replicates (4 ORFs with > 2-fold change). Thus, at acidic pH, Rim101p appears to be dispensable for the transcriptional response in *C. albicans* (Bensen, Martin et al. 2004).

In contrast, at alkaline pH huge transcriptional differences were observed between the *rim101* knockout strain and the wild-type strain with > 2-fold changes for 186 genes, indicating that Rim101p contributes to the transcriptional regulation of many genes at alkaline pH. However, only 116 genes were also found either within the alkaline-induced or within the alkaline-repressed genes, for the remaining 70 genes no differential pH-dependent transcription was detected. This might be partially explained by missing or highly variable

data for some of these genes, but could also be explained by opposite effects of Rim101p to other pH-dependent regulatory events resulting in a relatively stable transcription.

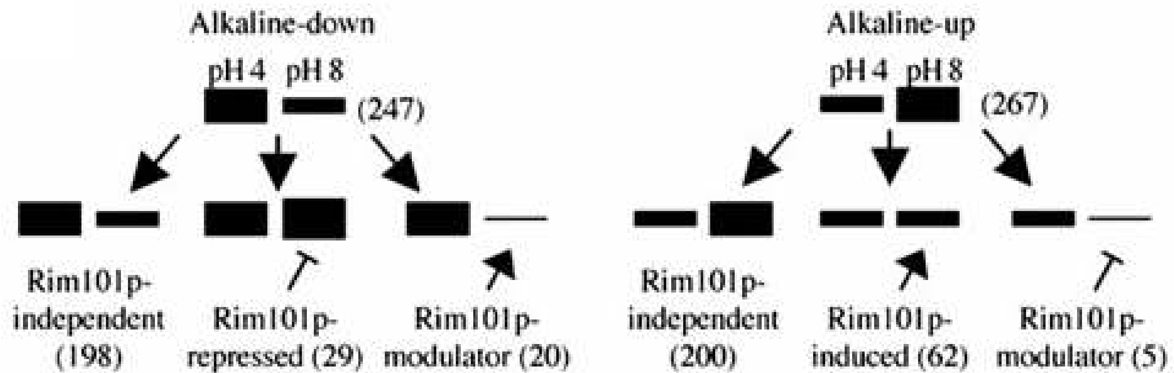


Figure 17: *These schemes were proposed by Bensen et al. (Bensen, Martin et al. 2004) in an attempt to describe the role of Rim101 in the transcriptional regulation. A large number of genes were found to be regulated at alkaline pH in a Rim101-independent manner. The set of genes that was differentially transcribed in the presence than in the absence of Rim101 at alkaline pH could be divided into two subgroups. For one group of genes the transcription in the rim101 null mutant was similar to that observed under acidic conditions in a wild type strain, thus normally alkaline-repressed genes were derepressed (29 genes left model middle group) and normally alkaline-induced genes were no more induced (62 genes right model middle group). The other group contains genes for which the absence of Rim101 even increased the regulation observed at alkaline pH in a wild type strain, thus normally alkaline-repressed genes were even more repressed in the absence of Rim101 (20 genes left model right group) and normally already pH-induced genes were even more strongly induced.*

The role of Rim101p in the transcriptional regulation of some genes seems to be quite complex as it is illustrated in the models proposed by Bensen *et al.* (Figure 17), in particular in the case of alkaline repression of genes. Of the 49 genes that were found to be downregulated at alkaline pH and differentially transcribed between the wild type and the *rim101* null mutant, only 29 appear to be repressed by Rim101p (Figure 17 A). Thus, 20 genes (40 %) were apparently even more strongly repressed in the *rim101* null mutant than in the wild type strain, which would indicate a role of Rim101p opposite to alkaline repression (Bensen, Martin et al. 2004).

A drawback of this experiment is that the experimental conditions favor the development of hyphae at alkaline pH in the wildtype strain, while in the *rim101* knockout strain this pH-dependent yeast-to-hyphae transition is impaired. These different phenotypes might contribute to the apparent large differences in gene expression between both strains at pH 8, and probably not all the differences observed can be attributed solely to the lack of Rim101p. For instance the adjustment of cell numbers and RNA quantity prior to the reverse transcription can be complicated by the distinct phenotypic properties. The pH-dependent filamentation has been shown to be dependent of Efg1p downstream of Rim101p (El Barkani, Kurzai et al. 2000), the observed Rim101p-dependent regulation of many hypha-specific

genes might thus be just indirect as a result of the downstream activity of Efg1p or other transcription factors.

They concluded that Rim101p is responsible for the regulation of only a subset of genes in the pH response of *C. albicans*, and that the other genes might be under control of Mds3p or other pH-dependent transcriptional responses. The pH-dependent regulation of many ion transporters and of the amino acid metabolism seems to be conserved between *S. cerevisiae* and *C. albicans*. On the other hand, there are specific differences, for example the role of ScRim101p in meiosis (Su and Mitchell 1993) and that of CaRim101p in filamentation that indicate a functional diversification (Ramon and Fonzi 2003). The transcription factor Nrg1p seems to function independently of Rim101p in *C. albicans* (Bensen, Martin et al. 2004). However, the impact of CaRim101p on the transcription of *NRG1* is not yet well understood, because another study reports that CaRim101p is able to repress *NRG1* transcription under certain conditions (Lotz, Sohn et al. 2004).

Finally, the data of Bensen *et al.* suggest that CaRim101p seems to be important for the regulation of genes involved in iron acquisition, which makes intuitively sense, as Fe^{2+} cations become oxidized at alkaline pH to the less soluble Fe^{3+} ions, thus the cells need to adapt to alkaline iron starvation. This function seems to be conserved in fungi, as ScRim101p also controls the expression of genes such as *ARN4*, an iron-siderophore binding protein (Lamb and Mitchell 2003).

Another large-scale approach was undertaken by Lotz *et al.* to identify cell surface proteins (Lotz, Sohn et al. 2004). They used a macroarray with 117 ORFs that are cell wall specific (Sohn, Urban et al. 2003). In a first part of the experiment they examined the effect of an overexpressed truncated and constitutively active version of Rim101p (El Barkani, Kurzai et al. 2000) on the transcription of these ORFs at acidic pH compared to a reference strain. In this way they identified nine cell wall genes that are upregulated by Rim101p. The second part of the experiment was more similar to that of Bensen *et al.*: they compared a *rim101* mutant with a reference strain at pH 7.4 and found 23 ORFs that were derepressed in the absence of Rim101p. A repression of these genes under the control of the truncated version of Rim101p at acidic pH or a silencing of Rim101p-induced genes in the *rim101* null mutant compared to the reference at alkaline pH was not reported. Different from Bensen *et al.*, these experiments were carried out at 30 °C in YPD medium, thus under conditions that do not promote filamentation in the reference strain at alkaline pH.

Interestingly, they found that the truncated form of Rim101p was able to induce the expression of hypha-associated genes such as *HWP1* and *RBT1*, although no hyphae could be

observed. However, the induction of these genes was much stronger at alkaline pH than at acidic pH, indicating that either the activity of the truncated form of Rim101p is pH-dependent or that other pH-dependent regulators are involved in the regulation of these genes. Three apparently *Candida*-specific possibly GPI-anchored proteins of unknown function that were found to be repressed by Rim101p were named Rbr1p, Rbr2p and Rbr3p and characterized further. *RBR1* seems to be positively regulated by Nrg1p (!) and repressed at alkaline pH by Rim101p through repression of Nrg1p; a possible requirement in filamentation under certain conditions was assumed, because filamentation of a *RBR1* mutant was impaired on M199 softagar plates at acidic pH.

Finally, several recent results highlight the role of Rim101p in the regulation of virulence-associated genes. In one report, Rim101p was found to be responsible for the regulation of several genes of the *SAP* family of secreted aspartyl phosphatases (Villar, Kashleva et al. 2007) during mucosal tissue invasion. This is the first evidence for a role of Rim101p in the regulation of these genes.

Another report demonstrated the importance of the long known pH-regulated gene *PRA1* (Sentandreu, Elorza et al. 1998) for the human neutrophile immune response (Soloviev, Fonzi et al. 2007). Although the expression of *PRA1* is apparently not under direct regulation of Rim101p, but strongly influenced by the presence or absence of the directly Rim101p-induced cell wall protein Phr1p (Choi, Yoo et al. 2003; Soloviev, Fonzi et al. 2007), this can be considered as another evidence that Rim101p triggers the expression of virulence factors.

Table 3: List of pH-regulated *C. albicans* genes

ORF19 Name	Gene Name	Alkaline induced	Alkaline repressed	Rim101 controlled*	Function
orf19.7114	CSA1			B	Surface antigen on elongating hyphae and buds
orf19.4887	ECM21				Similar to S. *cerevisiae* Ecm21p
orf19.1325	ECM38			B	Protein of unknown function
orf19.6070	ENA2			B	Sodium transporter
orf19.5634	FRP1				Ferric reductase
orf19.7112	FRP2			B	Ferric reductase
orf19.4304	GAP1			B	General amino acid permease; antigenic in human
orf19.385	GCV2			B	Glycine catabolism
orf19.4716	GDH3			B	NADP-glutamate dehydrogenase
orf19.4647	HAP3			B	CCAAT-binding transcription factor (respiration)
orf19.6073	HMX1				Heme oxygenase; acts in utilization of hemin iron
orf19.5760	IHD1			B	GPI-anchored protein of unknown function
orf19.7363	KRE6			L	Protein of beta-1,6-glucan synthesis
orf19.3981	MAL31			B	Protein of unknown function
orf19.5280	MUP1			B	Protein of unknown function
orf19.5674	PGA10				GPI anchored protein involved in heme-iron utilization
orf19.5635	PGA7				GPI-anchored precursor of a hyphal surface antigen
orf19.4599	PHO89			B	Putative phosphate permease
orf19.3829	PHR1			B	beta-1,3 Glycosidase
orf19.4025	PRE1			B	Protein of unknown function
orf19.5636	RBT5				GPI-anchored cell wall protein
orf19.7362	SKN1			L	Predicted role in beta-1,6-glucan synthesis
orf19.2770.1	SOD1				Cytosolic superoxide dismutase
orf19.7566				B	Protein of unknown function
orf19.4082	DDR48			B	Immunogenic stress-associated protein
orf19.4255	ECM331				Putative GPI-anchored protein
orf19.6489	FMP45			B	Mating process
orf19.1193	GNP1			B	Similar to asparagine and glutamine permease
orf19.4026	HIS1			B	Histidine biosynthesis
orf19.7447	JEN1			B	Lactate transporter
orf19.4279	MNN1			B	Putative alpha-1,3-mannosyltransferase
orf19.3117					Protein of unknown function
orf19.3740	PGA23			B	GPI-anchored protein of unknown function
orf19.3754	PHO111				Constitutive acid phosphatase
orf19.3727	PHO112				Constitutive acid phosphatase
orf19.6081	PHR2			L, B	beta-1,3 Glycosidase
orf19.6937	PTR2			B	Putative oligopeptide transporter
orf19.3765	RAX2			L	Protein of unknown function
orf19.7218	RBE1			L,B	Putative cell wall protein
orf19.5124	RBR3				Cell wall protein
orf19.5032	SIM1			L	DNA replication regulatory protein
orf19.6972	SMI1B			L	Protein of unknown function
orf19.7077					Protein of unknown function
orf19.7436	AAF1			B	Possible regulatory protein
orf19.3554	AAT1			B	Aspartate aminotransferase
orf19.1816	ALS3				Adhesin; ALS family
orf19.1170	ARO7			B	Putative chorismate mutase; fungal-specific
orf19.2098	ARO8			B	Protein of unknown function
orf19.3934	CAR1			B	Protein of unknown function
orf19.5641	CAR2			B	Protein of unknown function

ORF19 Name	Gene Name	Alkaline induced	Alkaline repressed	Rim101 controlled*	Function
orf19.3895	CHT2			B	Protein of unknown function
orf19.6402	CYS3			B	Sulfur amino acid biosynthesis
orf19.4536	CYS4			B	Predicted enzyme of sulfur amino acid biosynthesis
orf19.6139	FRE7			B	Ferric reductase
orf19.3538	FRE9			B	Ferric reductase
orf19.4802	FTH1			B	Putative high affinity iron transporter
orf19.3195	HIP1			B	Protein of unknown function, fungal-specific
orf19.5211	IDP1			B	Isocitrate dehydrogenase
orf19.4650	ILV6			B	Regulatory subunit of acetolacetate synthase
orf19.4225	LEU3				Predicted zinc-finger protein of unknown function
orf19.655	PHO84				High-affinity phosphate transporter
orf19.5650	PRO3			B	Mating process
orf19.7610	PTP3			B	Similar to S. *cerevisiae* tyrosine phosphatase
orf19.3974	PUT2			B	Protein of unknown function
orf19.6202	RBT4			L	Similar to plant pathogenesis-related proteins
orf19.2443	RGD1			B	GTPase activator
orf19.3911	SAH1			B	S-adenosyl-L-homocysteine hydrolase
orf19.657	SAM2			B	S-adenosylmethionine synthetase
orf19.386	SAM4			B	Protein of unknown function
orf19.3931	SFC1			B	Protein of unknown function
orf19.6763	SLK19				Plasma membrane protein
orf19.2270	SMF12			B	Manganese transporter
orf19.2069	SMF3			B	Vacuolar iron transporter
orf19.6190	SRB1				Essential GDP-mannose pyrophosphorylase
orf19.5908	TEC1			B	TEA/ATTS transcription factor
orf19.4265	UAP1			B	UDP-N-acetylglucosamine pyrophosphorylase
orf19.1822	UME6			B	Putative zinc cluster transcription factor
orf19.4197	YHM2			B	Protein of unknown function
orf19.1172				B	Phosphate transporter
orf19.2794					Protein of unknown function
orf19.4966				B	Protein of unknown function
orf19.5541				B	Similar to S. pombe Nrd1p
orf19.5761					Protein of unknown function
orf19.2762	AHP1			B	Alkyl hydroperoxide reductase
orf19.7469	ARG1			B	Argininosuccinate synthase
orf19.5610	ARG3			B	Protein of unknown function
orf19.6689	ARG4				Argininosuccinate lyase
orf19.4788	ARG5,6			B	Arginine biosynthetase
orf19.1847	ARO10			B	Pyruvate decarboxylase
orf19.6229	CAT1			B	Catalase
orf19.6948	CCC1			B	Manganese transporter
orf19.7517	CHT1			B	Chitinase
orf19.3656	COX15			B	Protein of unknown function
orf19.4630	CPA1			B	Protein of unknown function
orf19.5000	CYB2			B	Precursor protein of cytochrome b2
orf19.1770	CYC1			B	Cytochrome c
orf19.3527	CYT1			B	Cytochrome c1
orf19.4082	DDR48			B	Immunogenic stress-associated protein
orf19.5417	DOT5			B	Protein of unknown function
orf19.6656	DUR3			B	Protein of unknown function
orf19.6794	FESUR1			B	Ubiquinone reductase
orf19.6489	FMP45			B	Mating process

ORF19 Name	Gene Name	Alkaline induced	Alkaline repressed	Rim101 controlled*	Function
orf19.1153	GAD1			B	Protein of unknown function
orf19.4899	GCA1			B	Extracellular or plasma membrane glucoamylase
orf19.1979	GIT1			B	Glycerophosphoinositol permease
orf19.6257	GLT1			B	Protein of unknown function
orf19.1193	GNP1			B	Similar to asparagine and glutamine permease
orf19.1742	HEM3			B	Hydroxymethylbilane synthase
orf19.1744	HEM4			B	Protein described as uroporphyrinogen III synthase
orf19.4384	HXT5				Sugar transporter
orf19.5521	ISA1			B	Iron-sulfur cluster biogenesis
orf19.7498	LEU1			B	3-isopropylmalate dehydratase
orf19.3507	MCR1			B	NADH-cytochrome-b5 reductase
orf19.4495	NDH51			B	Subunit of NAD dehydrogenase complex I
orf19.6531	NUC2			B	NADH-ubiquinone oxidoreductase
orf19.5893	RIP1			B	Subunit of ubiquinol cytochrome c-reductase
orf19.974	ROT2				Alpha-glucosidase II subunit, cell wall synthesis
orf19.6595	RTA4				Fatty acid transport
orf19.2941	SCW4			B	Cell wall protein
orf19.3340	SOD2			B	Mitochondrial superoxide dismutase
orf19.6059	TTR1			B	Glutathione reductase
orf19.1585	ZRT2			B	Zinc transporter
orf19.251				B	Member of ThiJ/PfpI protein family; antigenic
orf19.1709				B	Protein of unknown function
orf19.3175				B	Protein of unknown function
orf19.4758				B	Protein of unknown function

Table 3: *List of the 126 C. albicans genes that are currently annotated as Rim101p- or pH-regulated according to www.candidagenome.org as revealed by a "text" search using the key words "Rim101, Rim, pH, acidic and alkaline". Genes are classified by the type of alkaline regulation which is indicated in colums 3 (induction) and 4 (repression), and by the importance of Rim101p in their regulation (column 5). A regulatory function of Rim101p is noted for about one third of these genes which are listed in the upper part of the table. Most of these genes were annotated based on the microarray studies of Bensen et al. (Bensen, Martin et al. 2004) or Lotz et al. (Lotz, Sohn et al. 2004); a B or L in column 5 indicates when this is the case.*

1.5 Aims of the project

The project of the thesis consisted in the identification and characterization of new target genes of Rim101p in the response of *C. albicans* to changes in ambient pH. At the start of this PhD thesis in April 2004 the idea of the project was already born, so the work presented here is the continuation and termination of an already initiated project.

When this work was begun, little was known about possible target genes of Rim101p, and no global transcriptional analysis had been performed to identify pH-regulated genes in *C. albicans*. However, it had been shown that Rim101p was not only involved in the pH-dependent regulation of filamentation, but that it also contributed to the pathogenesis of *C. albicans* at least in some virulence models (see introduction 1.4.4 The role of Rim101p in the pH response of *C. albicans*). Furthermore, it had been demonstrated that the *Pal/RIM* signaling pathway was highly conserved in *C. albicans*, and that its final effector was activated by N-terminal processing. Moreover, there was evidence that other pathways than the Rim101p pathway might be involved in pH-dependent gene regulation.

Thus, there was a need to understand the role of Rim101p in the pH response of *C. albicans* independently of other pH-dependent regulatory events, which led to the development of this work. The main tasks of this work might be defined as follows:

1) Completion of a microarray experiment in which a truncated and constitutively active form of Rim101p is used to evidence the specific contribution of Rim101p to the pH response of *C. albicans* in a pH-independent setup
2) Intensive analysis of the microarray data to identify new target genes under the control of Rim101p
3) Confirmation of microarray results for selected genes and evaluation of their relevance within the global pH-dependent regulation
4) Characterization of a particular subset of genes

2 Materials and Methods

2.1 General

2.1.1 Strains used

The bacterial strains used for transformation and amplification of recombinant DNA were *E. coli* TOP10 or JM109. Bacterial transformation were performed according to the protocol of Hanahan et al. (Hanahan, 1983).

The *C. albicans* strains used during this work are listed in Table 4 below. *C. albicans* sequence data were obtained from the *Candida*DB web site: [http://genolist.pasteur.fr/*Candida*DB/index.html](http://genolist.pasteur.fr/CandidaDB/index.html). All strains constructed for this study were derived from parental *rim101* disrupted strain DAY5 or reference strain DAY286. Strains DAY5, DAY25, DAY185 (Wilson, Davis et al. 1999), DAY286 (Davis, Bruno et al. 2002) and DAY492 (Li, Martin et al. 2004) were generous gifts from the laboratory of Dana Davis and Aaron Mitchell. The construction of strains FB1 and FB8 that were used in the microarray experiment was not part of my work and is described in detail in the attached publication. To construct the different β-galactosidase reporter strains, plasmids p*ALS1* (2kb)-LacZ, p*ALS1*(1kb)-LacZ, p*ALS4*(2kb)-LacZ, p*ALS4*(1kb)-LacZ, p*PHR1*-LacZ, p*PHR2*-LacZ, p*ADH1*-LacZ and p*Not1*-LacZ were digested with *NruI* to target them to the *HIS1* locus of strains DAY5 or DAY286.

C. albicans strains were transformed using a slightly modified form of the lithium acetate transformation protocol described by Kaiser et al. (Kaiser, 1994): to prepare competent cells, a 50 mL YPD culture with a start DO600 of 0.1-0.2 was grown for 4-5 hours at 30°C until an OD600 between 0.4 and 0.8. Cells were harvested, washed with TE buffer (0,1 M TrisHcl, 0,01 M EDTA, pH8) and resuspended in a final volume of 200 µL fresh LiAc solution (TE buffer containing 0.1M LiAc) before an overnight incubation at 4°C on ice. For transformation a mix of 5 µL carrier DNA (fish DNA 5 mg/mL) and 20 µL transformant DNA (about 1 µg) is prepared and incubated for 5 min on ice before carefully adding 50 µL competent cells and 300 µL of plate mix (fresh LiAc solution containing 40% PEG). After gently mixing by several inversions and incubating for 2-3 hours at 30°C without shaking a heat shock is carried out for 15 min at 44°C. Then 1 mL of YNB N5000 is added to dilute the PEG solution and cells are carefully spinned down for 20 sec at 3000 rpm. 1 mL of supernatant is discarded and cells are resuspended in the remaining volume before being spread on 3-4 SC His- plates (about 100 µL per plate). After incubation for 2-3 days at 30 °C

the appearing clones can be confirmed directly by PCR on the colonies. (For the description of the plasmids used the reader is referred to chapter 2.1.5. We screened for clones with the correct insertion event by PCR directly on the colonies using a reverse primer on LacZ together with a forward primer of the inserted promoter region (see 2.1.4). Afterwards we eliminated all clones with multiple insertions as detected by a PCR with a primer couple that permits the amplification of the junction region of two tandemly arranged integrated plasmids (p78insF/R, see 2.1.4).

Wilson *et al.*, 1999 (Wilson, Davis et al. 1999)

Table 4: Strains used in this study

***Candida albicans* strains**	***Genotype***				**Reference or source**
SC5314	*Clinical isolate*				Wilson *et al.*, 1999 (Wilson, Davis et al. 1999)
BWP17	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*his1 ::hisG* *his1 ::hisG*	*arg4 ::hisG* *arg4 ::hisG*		Wilson *et al.*, 1999 (Wilson, Davis et al. 1999)
DAY5	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*his1 ::hisG* *his1 ::hisG*	*arg4 ::hisG* *arg4 ::hisG*	*rim101 ::ARG4* *rim101 ::URA3*	Wilson *et al.*, 1999 (Wilson, Davis et al. 1999)
DAY25	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*pHIS1::his1::hisG his1::hisG*	*arg4::hisG* *arg4::hisG*	*rim101::ARG4* *rim101::URA3*	Wilson *et al.*, 1999 (Wilson, Davis et al. 1999)
DAY185	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*pHIS1::his1::hisG his1::hisG*	*pARG4::URA3::arg4:hisG* *arg4::hisG*		Wilson *et al.*, 1999 (Wilson, Davis et al. 1999)
DAY286	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*his1::hisG* *his1::hisG*	*pARG4::URA3::arg4:hisG* *arg4::hisG*		Davis *et al.*, 2002(Davis, Bruno et al. 2002)
DAY492	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*pHIS1::RIM101-V5-AgeI::his1::hisG* *his1::hisG*	*arg4::hisG* *arg4::hisG*	*rim101::ARG4* *rim101::URA3*	Li *et al.*, 2004(Li, Martin et al. 2004)
FB1	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*pHIS1::his1::hisG his1::hisG*	*arg4::hisG* *arg4::hisG*	*rim101::ARG4* *rim101::URA3*	This study
FB8	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*pRIM101SL::HIS1::hisG his1::hisG*	*arg4::hisG* *arg4::hisG*	*rim101::ARG4* *rim101::URA3*	This study
A1- 2kb	*ura3Δ ::λimm434* *ura3Δ ::λimm434*	*p2kbALS1-LacZ::HIS1::hisG* *his1::hisG*	*pARG4::URA3::arg4:hisG*		This study

			arg4::hisG	
A1- 1kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p1kbALS1-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
A4- 2kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p2kbALS4-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
A4- 1kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p1kbALS1-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
P1-1kb	ura3Δ ::λimm434 ura3Δ ::λimm434	pPHR1-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
P2-2kb	ura3Δ ::λimm434 ura3Δ ::λimm434	pPHR2-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
ADH1-bas	ura3Δ ::λimm434 ura3Δ ::λimm434	pADH1-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
wo-	ura3Δ ::λimm434 ura3Δ ::λimm434	pNotI-LacZ::HIS1::hisG his1::hisG	pARG4::URA3::arg4:hisG arg4::hisG	This study
A1-Rim 2kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p2kbALS1-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
A1-Rim 1kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p1kbALS1-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
A4-Rim 2kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p2kbALS4-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
A4-Rim 1kb	ura3Δ ::λimm434 ura3Δ ::λimm434	p1kbALS1-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
P1-Rim	ura3Δ ::λimm434 ura3Δ ::λimm434	pPHR1-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
P2-Rim	ura3Δ ::λimm434 ura3Δ ::λimm434	pPHR2-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
ADH1-Rim	ura3Δ ::λimm434 ura3Δ ::λimm434	pADH1-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study
wo- Rim	ura3Δ ::λimm434 ura3Δ ::λimm434	pNotI-LacZ::HIS1::hisG his1::hisG	arg4::hisG rim101::ARG4 arg4::hisG rim101::URA3	This study

Table 4: Strains used in this study and their genotype

2.1.2 Culture media and phenotypic tests

C. albicans strains were routinely grown at 30 °C. in YPD (2 % Bacto Peptone, 1 % yeast extract, 2 % dextrose, and 80 µg/mL of uridine, when needed). Defined SC medium for growth of *C. albicans* consisted of complete synthetic medium CSM without uracil-methionine-cysteine, 6.7 g/L yeast nitrogen base without amino acids (Difco) and 2 % glucose. 5 mM methionine and 2 mM cysteine were added when needed to repress *RIM101SL* transcription in strain FB8. For growth at a defined pH, the medium was buffered with 150

mM HEPES and the pH was adjusted with HCl or NaOH. For growth and hypersensitivity tests, droplets of serial dilutions of an exponential-phase culture in YPD medium were spotted onto SC or SC pH 10 media buffered with 50 mM glycine-NaOH (with or without 5 mM methionine and 2 mM cysteine); plates were incubated 4 days at 30 °C.

2.1.3 *ALS* primers selection

A set of gene-specific real time qPCR primers was recently published for eight *ALS* gene transcripts, *ALS1*-*ALS7* and *ALS9* (Green, Zhao et al. 2005). We intended to use these primers, but when testing the complete primer set on the Lightcycler® system, we unexpectedly encountered dimerization problems in PCR reactions with *ALS6* and *ALS9* primer couples. These problems are most probably due to the primer concentrations (500 nM) recommended for the use of this real time PCR system, which are higher than those Green *et al.* used (100 nM). We thus decided to design new specific primers for these two genes.

When we started our analysis, the sequence data from Assembly19 available at *Candida*DB (http://genolist.pasteur.fr/*Candida*DB/) and *Candida* Genome Database (CGD, http://www.*Candida*genome.org/) still suggested the presence of two additional *ALS* genes (*ALS10* and *ALS12*) in strain SC5314 which were not included in the analysis of Green *et al.* (Green, Zhao et al. 2005). The sequences of the two alleles of *ALS12* (orf19.2121/22/orf19.9669/70) were virtually identical to the sequence annotated as *ALS4* (orf19.4555/6) and the sequences of *ALS10* (orf 19.2355/orf19.9891) show high similarities to *ALS2* (orf19.1097, orf19.8699) and to *ALS3* (orf19.1816/orf19.9379) alleles. The *ALS3* primers of the above mentioned primer set did not distinguish *in silico* between *ALS3* (orf19.1816/orf19.9379) and *ALS10*, and those for *ALS4* also recognized *ALS12* with the same probability. The distinction of sequences designated as *ALS4* and *ALS12* seemed virtually impossible, so we decided to use the CGD sequence information to complete our primer set at least with new specific primer couples for *ALS3* and *ALS10*.

Recently, after completion of our real-time PCR quantifications using these new primers, *ALS10* and *ALS12* were found to be Assembly 19 artefacts and were removed from CGD in Assembly 20. Consequently there is no need for a distinction any more, but nevertheless our results for *ALS3* were accomplished using this new primer couple.

The list of newly designed primers used for *ALS3, -6* and *-9* gene transcript quantification is given in Table 5 below (2.1.4). To find suitable primers (Web primer tool http://seq.yeastgenome.org/cgi-bin/web-primer) only regions that were identical between the two alleles of a given gene, but different from the corresponding regions of all other *ALS* genes were used.

2.1.4 Primers used

Name	Sequence	Purpose
OFB16	TGTGACGACCATGTTGGTAGAAAGT	qPCR RIM101 cDNA
OFB17	CTTGAGGTCTCTTGAACGATTTGGG	qPCR RIM101 cDNA
OFB22	GCAGTGCTTCAATCAATAGCAAGGC	qPCR PHR1 cDNA
OFB23	AGAGCTTGAGCTGGACCCAGA	qPCR PHR1 cDNA
OFB32	AGTGTGACATGGATGTTAGAAAAGAATTATACGG	qPCR ACT1 cDNA
OFB33	ACAGAGTATTTTCTTTCTGGTGGAGCA	qPCR ACT1 cDNA
OFB40	ACACTGACGCTTCTGCTTTCG	qPCR PHR2 cDNA
OFB41	GCAGCTTCGTCTTCATCACCACA	qPCR PHR2 cDNA
F-Rim101	GACCTCGAGAATTACAACATTCATCCCG	Construction Rim101SL
R-Rim101SL	GTACCAAGCTTAGAAAGCAGTTATAGTTGG	Construction Rim101SL
RIMn697	CATGGTCGTCACACAAATGATCG	RACE Rim101 5’ end
RIMn433	GTTGGTAGCCATAAGTTGGTTGG	RACE Rim101 5’ end
A3newF	CCAAAACTTGTTCATCTAATGGTATCT	qPCR ALS3 cDNA
A3newR	TAGCATACGACAAGGTGTACGAAT	qPCR ALS3 cDNA
A6newF	TTTGATGATAAGTCGTCGGCA	qPCR ALS6 cDNA
A6newR2	GCGATAAATCCATTATTGGTTTCA	qPCR ALS6 cDNA
A9newF	ACCCTCATGGATCTGAGACTATTG	qPCR ALS9 cDNA
A9newR	ACCGAACCAGAACCATCGTAT	qPCR ALS9 cDNA
ALS1pF2	GCATTGGATAAAAACAGGTCATC	2kb ALS1 promoter, Verif. insertion
ALS1pF1	GCTAACTTTTATCGGCCTATTACTCC	1kb ALS1 promoter, Verif. insertion
ALS1pR	CTGATATTAACAATTGGTAGTTGTTTGAAC	ALS1 promoter
ALS4pF2	GGAAAAATTACTGAAGAAGCTTTGAGAA	2kb ALS4 promoter, Verif. insertion
ALS4pF1	CCTTCTCTTCATGTATTCGAAACAC	1kb ALS4 promoter, Verif. insertion
ALS4pR	GTTTTGGTGATAGATGGCTAATG	ALS4 promoter
PHR1pF	GATTACAAGTGGGATGCAAAAT	PHR1 promoter, Verif. insertion
PHR1pR	TTTTTTTTTGGCTTCAACCTGTAG	PHR1 promoter
PHR2pF2	AGTTTTTCCATGAATTTCTACGAATG	PHR2 promoter, Verif. insertion
PHR2pR	TAGCGATCGAATGTGTGTAGTTTC	PHR2 promoter
ADH1F	CCACCACGGCAAAGACATT	Verification pADH1-LacZ insertion
LACZearlyR	GAATTTTTTCAGTCATAGCCATGT	Verification insertion ALS1/4,PHR1/2
LACZF	AACATGGCTATGACTGAAAAAATTC	Verification p-LacZ insertion
LACZR	GTGGTTCAATCATGAAGCTTAATTG	Verification p-LacZ insertion
p78insF	TGTAGGTGGTGACGACACATG	Multiple insertion test
p78insR	TGGCGTTATTGGTGTTGATG	Multiple insertion test
NADAP	CCGGGGCGGCCGCC	NotI adapter
qA1pF	TTGCTAATCATCTTTGGAGATATTCG	qPCR ChIP ALS1 promoter
qA1pR	CTGTCTTACTTCTCCGTTTCATTAG	qPCR ChIP ALS1 promoter
qA4pF	TTGCTTACTGGAAATTTGCTCT	qPCR ChIP ALS4 promoter
qA4pR	CAATTGTTGTCCGAAATACCTG	qPCR ChIP ALS4 promoter
qPHR1pF	TTTCGTCTTACAGAGCACAACAAGAAC	qPCR ChIP PHR1 promoter
qPHR1pR	TTTCAAGGTGGAATGATTTGATCTAAGGAG	qPCR ChIP PHR1 promoter
qPHR2pF	TGGCTTTCCTCCCCTTAACTG	qPCR ChIP PHR2 promoter
qPHR2pR	TCGAATGTGTGTAGTTTCTTTGACGA	qPCR ChIP PHR2 promoter
qADH1pF	TAAATTATTACTTTCCGTGGCCAATCA	qPCR ChIP ADH1 promoter
qADH1pR	GGAAACTCTTTAGGCAATACTTGCT	qPCR ChIP ADH1 promoter
qPHO87F	TTGCATTAGGGAAAGCCGT	PHO87 cDNA qPCR
qPHO87R	CCAACTTCTTTAACCAAGGGGA	PHO87 cDNA qPCR
qCPA1F	ATGGAATCACCAAAAGTTCAATGT	CPA1 cDNA qPCR
qCPA1R	CCAATCGTAATTCTCCCCAATG	CPA1 cDNA qPCR
qIPF16514F	GGTGGCATTACTTCAGGTT	IPF16514 cDNA qPCR
qIPF16514R	AATTGGATGTGCTTGTGGT	IPF16514 cDNA qPCR
qCHO2F	AGAATAGTATTGGGGAGTGGAT	CHO2 cDNA qPCR

qCHO2R	TCCATAATAACCAACGTACCCT	CHO2 cDNA qPCR
qPGA52F	CTGGTATGGCTGCACCT	PGA52 cDNA qPCR
qPGA52R	AGTAACGGTAGCCAATGTAGT	PGA52 cDNA qPCR
qALS1F	ATCAAGCTTGACAACAGGC	ALS1F cDNA qPCR
qALS1R	GTTGAAGGTGAGGATGAGGTAA	ALS1F cDNA qPCR
qALS2/4F	TGTTTCACACACAGTGACCG	ALS2/4F cDNA qPCR
qALS2/4R	CTGTCGCAGTTGCAGAAG	ALS2/4F cDNA qPCR
qIPF6156F	GGGGAAAGATGCAGCTAGAG	IPF6156F cDNA qPCR
qIPF6156R	GTTCACAGGGAGCAGGT	IPF6156F cDNA qPCR
qEFG1F	ATTTCCAGGGTGGTGCT	EFG1F cDNA qPCR
qEFG1R	GGGTGATTGGTGCACAG	EFG1F cDNA qPCR
qIPF8762F	CGGTAGACCTAGAAAATATGCC	IPF8762F cDNA qPCR
qIPF8762R	TTGTCATCGAAACCAACAAAGT	IPF8762F cDNA qPCR
qQDR1F	TGACTATCGTGCCCTTAGC	QDR1F cDNA qPCR
qQDR1R	CAGATGCACCACTCTGTT	QDR1F cDNA qPCR
qPGA4F	CCCAAAGGTGCTTTGAAATACT	PGA4F cDNA qPCR
qPGA4R	AGTTGCAGATGAGCTGGAA	PGA4F cDNA qPCR
qIPF4580F	TCGCTCTATTGAACCGTCAAA	IPF4580F cDNA qPCR
qIPF4580R	TCCGGGCCACCATCTAA	IPF4580F cDNA qPCR
qIPF1372F	CGACTCCTTTACGCAAGA	IPF1372F cDNA qPCR
qIPF1372R	TGTAGGGTTTCGAGATGCC	IPF1372F cDNA qPCR
qWSC4F	TCCATACAGCAAGCAATCGT	WSC4F cDNA qPCR
qWSC4R	GGGAACATAGCTCCTCCATC	WSC4F cDNA qPCR
qKRE6F	AGACCAGGGTATTTGGGAT	KRE6F cDNA qPCR
qKRE6R	AATTTCAGGGGCACCTCTA	KRE6F cDNA qPCR
qIPF2280F	GTTTTCACTGCACTTCATGTTG	IPF2280F cDNA qPCR
qIPF2280R	TAGAGTGAGCAGCATCGG	IPF2280F cDNA qPCR
qHGH1F	GGAACCATTGAGTGGAATTCTT	HGH1F cDNA qPCR
qHGH1R	TCTCTTCAATATCAGCAGACTGT	HGH1F cDNA qPCR
qPHR1F	GCAGTGCTTCAATCAATAGCAAGGC	PHR1F cDNA qPCR
qPHR1R	AGAGCTTGAGCTGGACCCAGA	PHR1F cDNA qPCR
qPHR2F	ACACTGACGCTTCTGCTTTCG	PHR2F cDNA qPCR
qPHR2R	GCAGCTTCGTCTTCATCACCACA	PHR2F cDNA qPCR

Table 5: Primers used in this study and the purposethey were used for

2.1.5 Genomic DNA purification

A Genomic DNA has been extracted from C. albicans following a slightly modified version of the protocol elaborated by Querol et al (Querol, Barrio et al. 1992): the strain of interest was grown overnight in 5 mL YPD medium at 30 °C. After harvesting the cells by centrifugation the pellet is resusended in 500 µL of extraction buffer (1 M sorbitol, 0.1 M EDTA, pH8) containing 4 µg zymolase 100T (Seikagaku corporation, Tokio, Japan). After one hour of incubation at 37 °C, the mix is centrifugated and the pellet resuspended in a buffer containing 50 mM Tris/HCl with 20 mM EDTA at pH7.5. Afterwards 0.7 mL of isopropanol are added and the aqueous phase is extracted. Genomic DNA can now be precipitated by adding 40 µL of 2.5M NaAc and 1 mL absolute ethanol. Finally, the pellet is

resuspended in 100 μL H_2O, and RNA is degraded by the addition of 2 μl of 5 mg/ml RNase during an incubation of 15 min at 37°C prior to use.

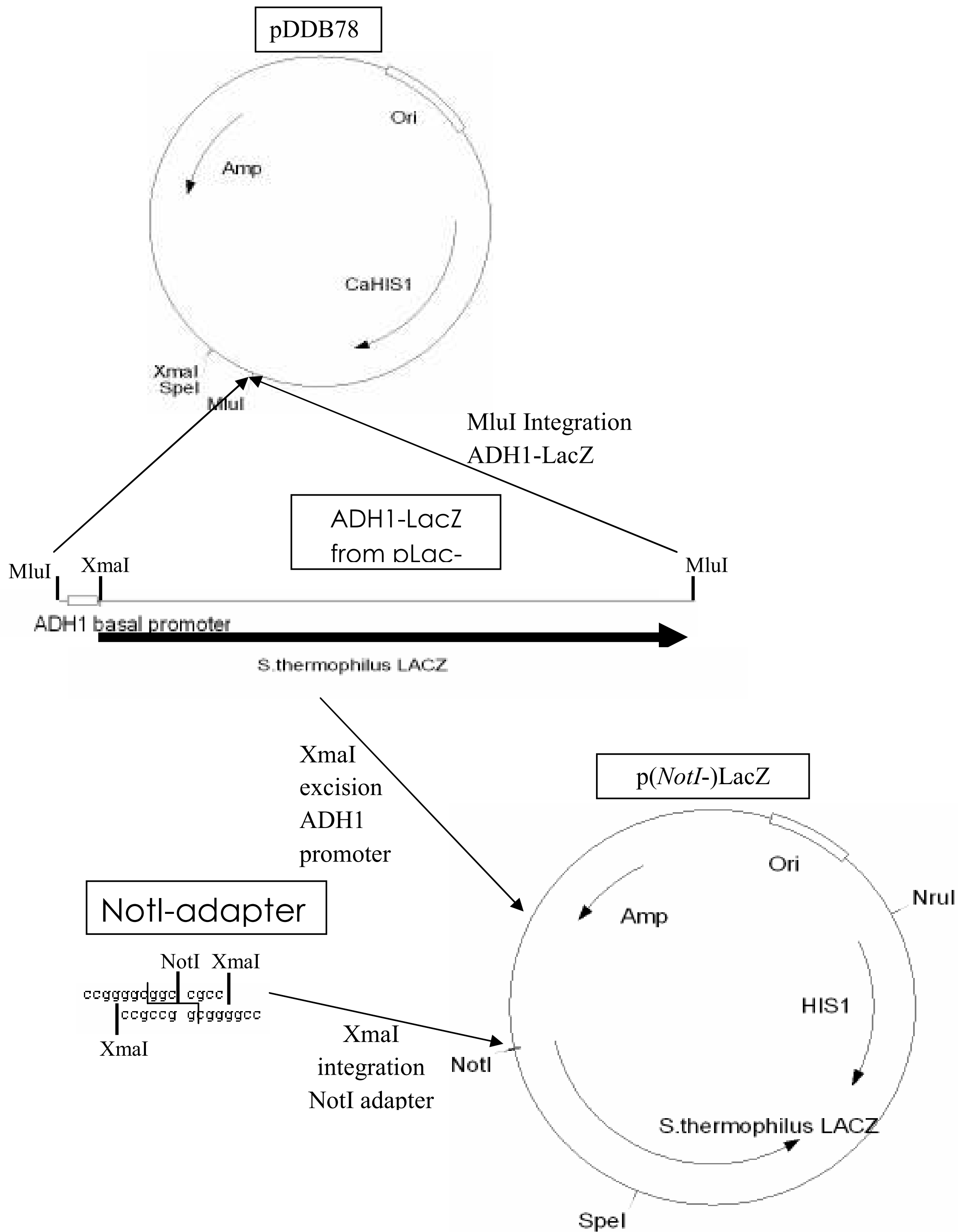
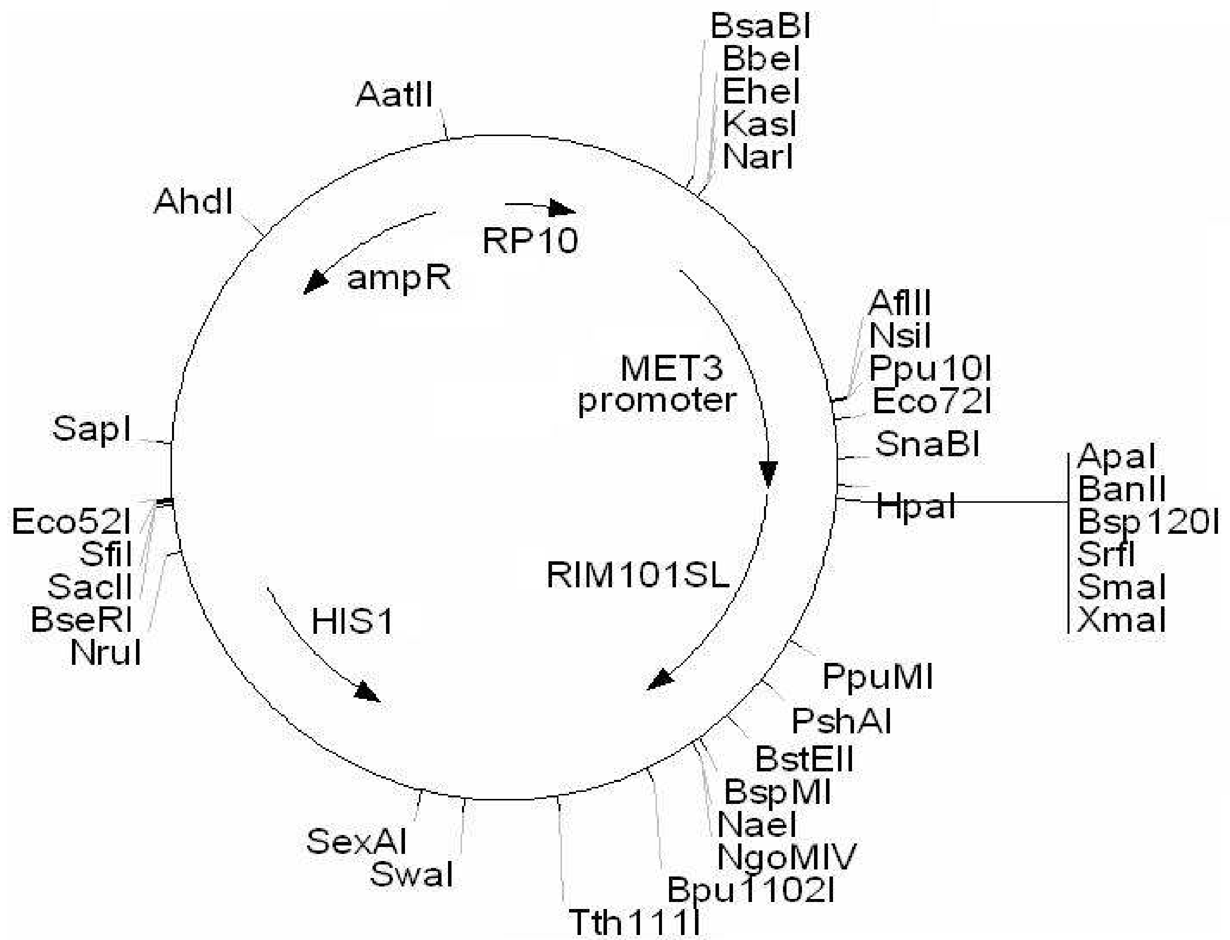


Figure 18: *Construction of plasmid p(NotI)-LacZ: A DNA fragment with MluI ends carrying a S. thermophilus LacZ gene under control of a basal ADH1 promoter was integrated into the MluI-digested plasmid pDDB78. Then the ADH1 promoter was excised using the XmaI sites present on ADH1-LacZ and pDDB78 and the plasmid was religated. We used the resulting single XmaI site in the plasmid to insert a NotI-Adapter (with XmaI ends), resulting in plasmid p(NotI)-LacZ.*

2.1.6 Plasmid construction



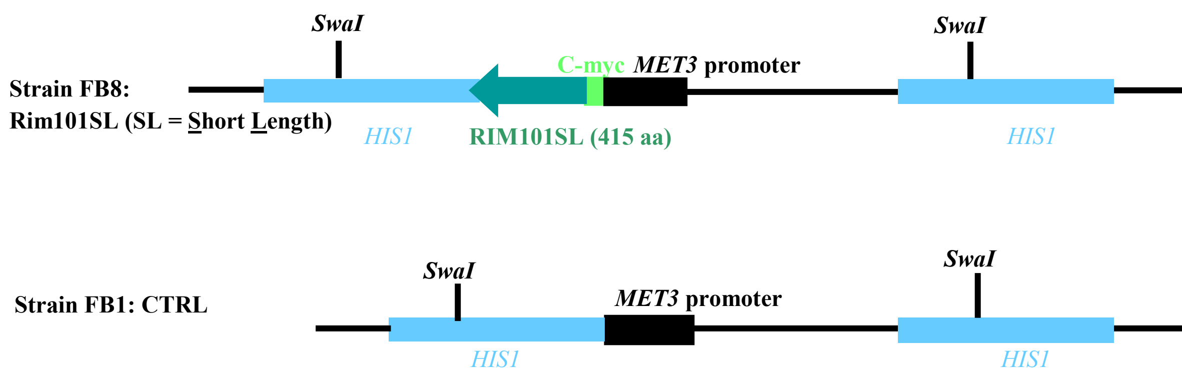


Figure 19: a) *Map of plasmid pINA1341 that carries the constitutively active truncated allele RIM101SL under the control of the MET3 promoter.* ***b)*** *Map of the HIS1 locus of strains FB8 (Rim101SL) and FB1 (CTRL) after integration of SwaI-digested pINA1341 (RIM101SL) and pINA1337 (pINA1341 without RIM101SL). Strain FB8 was used in the time course experiment, control strain FB1 was used to identify false positive genes that were regulated in MET3-inducing conditions.*

As the construction of the plasmids for the microarray study dates before my arrival in the lab, please refer to the attached publication for a detailed description (See plasmid map Results figure 19).

Plasmids for the β-galactosidase assays were constructed as follows (See also Figure 18): A basal *ADH1* promoter was *MluI*-excised together with the modified version of the *S. thermophilus* LacZ gene from plasmid plac-basal (provided by the group of A. Brown Aberdeen (Garcia-Sanchez, Mavor et al. 2005)) and integrated into *MluI*-linearized pDDB78 (Spreghini, Davis et al. 2003) for transformation of *C. albicans* His^- strains. Clones with an oriented integration of ADH1-LacZ downstream of the MCS of pDDB78 were selected (verification by *XmaI* digestion). This plasmid, p*ADH1*-LacZ, was later used as a control for a pH-independent LacZ reporter regulation. To generate plasmid pLacZ without any promoter governing LacZ expression, we *XmaI*-excised the *ADH1* promoter fragment and religated the remaining plasmid.

Initially we planned to directly integrate the PCR-amplified promoters of *ALS1*, *ALS4*, *PHR1* and *PHR2* into pLacZ (linearized with a *SmaI* digestion and dephosphorylated) via blunt end cloning, but despite multiple attempts we failed to obtain any plasmids with such insertions. We thus decided to change the strategy and to clone a *NotI* site right in front of LacZ to facilitate integration of the different promoters. Thus, we ordered a primer that creates by self-annealing (5 min 95 °C) a small DNA fragment with a *NotI* restriction site flanked by *XmaI* ends (primer "NADAP" see table 5 (2.1.4)) which could be integrated in *XmaI*-linearized pLacZ to create p*NotI*-LacZ (See Figure 18). This plasmid was used as a negative control (LacZ without any promoter) in the β-galactosidase tests and to integrate the promoter regions of *ALS1* (2 and 1 kb), *ALS4* (2 and 1 kb), *PHR1* (1 kb) and *PHR2* (2 kb). These promoter regions were PCR-amplified from genomic DNA of strain BWP17 with primer pairs ALS1pF2/ALS1pR, ALS1pF1/ALS1pR, ALS4pF2/ALS4pR, ALS4pF1/ALS4pR, PHR1pF/PHR1pR and PHR2pF2/PHR2pR respectively (see sequences in 2.1.4) and first integrated in $pGemT^{easy}$ (Promega). All integrated fragments were sequenced to check for possible mismatches. *NotI*-excised promoters were then introduced into *NotI*-linearized p*NotI*-LacZ. The resulting linker sequences between promoter end and LacZ start codon comprised 26 or 30 bp including the *NotI* site. The correct orientation of the integrated promoters was confirmed on agarose gels after *SpeI* (recognizes one restriction site

within pDDB78 and another one within the additionally from pGemT^{easy} imported base pairs between promoter and *NotI* end) digestion for all constructs.

2.1.7 Identification of the *RIM101* start codon by the RACE technique

RNA from strain DAY185 grown until OD_{600} = 1 in SC medium at pH 4 or 8 was purified as indicated below (see 2.2.2 Experimental Setup). The following steps were carried out according to manufacturer's instructions using the GeneRacer™ RACE cDNA kit (Invitrogen). De-capped *RIM101* mRNAs were reverse transcribed using the RIMn697 primer positioned on the 3' region of *RIM101* (Table 5). The cDNA products were then used as templates for a two- step PCR reaction. GeneRacer™ 5'1 primer and RIMn697 were used for a first PCR, and a nested PCR was then done on the first PCR product, with GeneRacer™ 5' Nested primer and RIMn433, a primer hybridising upstream from RIMn697 (Table 5). The procedure was repeated with RNA purified from cultures grown at either pH 4 or pH 8 to check for possible alternative start sites. PCR products were subsequently sequenced.

2.2 Microarray time course

2.2.1 Genomic microarrays

The microarrays used in this work were designed in a common effort of the European Galar Fungail I network and Eurogentec. They comprise 6038 spotted PCR sequences, corresponding to 6003 entries; for 35 *C. albicans* genes two probes exist as a control). Primer pairs were defined with a preference for the 600 bp 3'-terminal region of each gene using Primer3 software http://www-genome.wi.mit.edu/genome_software/other/primer3.html. Altogether the spotted probes correspond to 6003 entries, including probes for 5907 *C. albicans* genes and various control sequences. These control sequences are probes for *C. albicans* mitochondrial genes and intergenic regions, for *S. cerevisiae* and *S. pombe* genes, and the human glycerol-3-phosphate dehydrogenase and actin genes. Universal sequences of 15 bases were incorporated on the 5' of each specific forward and reverse primer to allow the generation of 5' amino-modified PCR products with an average length of 300 bp which could be attached covalently to the aldehyde-coated support (See Introduction Chapter 1.2.2.3 Global Transcription Analysis Figure 6B).

2.2.2 Experimental Setup

For the time course experiment (TC) with strain FB8, overnight precultures in YPD of strain FB8 were inoculated in liquid SC medium supplemented with uridine, methionine and cysteine at an OD_{600} of 0.2. Cultures were grown at 30 °C and 150 rpm until an OD_{600} of 0.8, before sample 0 was taken, then the rest of the culture was washed twice with SC medium without methionine and cysteine (SC+Uri-Met-Cys) and resuspended in the same volume of SC+Uri-Met-Cys to induce *MET3*-promoter activation. Additional samples were taken after 15, 30, 45 and 90 min of incubation. For each sample a volume that corresponds to $5*10^8$ cells was taken. Cells were then pelleted by centrifugation, flash-frozen in liquid nitrogen and stored at –80 °C.

Two independent biological replicates were used, resulting in two time points for each sample. Two arrays were hybridized for each sample, but for the second array sample and reference labeling was inversed (dye swap) to account for dye effects. The reference consisted of an inversed labeled pool of all time samples obtained from a third biological replicate, thus providing for each gene an average expression signal to which individual time point signals could be compared. Two additional arrays were done hybridizing the reference pool against itself labeled with Cy5 and Cy3. Results from these arrays were used later to normalize data from the other arrays to exclude labeling biases.

For the control experiments (CTRL) using strain FB1, growth conditions were identical, but samples were taken from the two independent biological replicates only at time points 0, 15 and 90. These time points were chosen after a first analysis of the time course results obtained with FB8. A pool of RNA extracted at the three time points was used as a reference as in the FB8 experiment.

2.2.3 RNA purification and cDNA labeling

For RNA preparation, frozen cells were broken in a 5 mL Teflon vessel of a Braun micro-dismembrator containing one 7 mm bead of tungsten carbide (Braun), both pre-cooled in liquid nitrogen. The closed flask was then shaken at 2.600 rpm for 2 min. RNAs were extracted using the RNeasy Midi Kit (Qiagen). Residual genomic DNA was removed using the RNase-Free DNase Set (Qiagen). RNA quantity and quality were controlled on a 1 % agarose gel and by measuring of the OD_{260} and the ratio OD_{260}/OD_{280} on a 1:500 dilution.

Reverse transcription and labeling reactions were carried out as follows:

20 µg of total RNA was mixed with 2 µL of a *C. albicans* specific primer mix (Eurogentec; 0.1 pmol/µL) and completed with RNAse-free water to a total volume of 19 µL. This mix was incubated for 5 min at 70 °C and then cooled down on ice to permit primer annealing. Then 8 µL of First strand buffer (Invitrogen), 4 µL 0.1M DTT, 1 µL of RNasin (Promega), 3 µL of a 20 mM dNTP mix (6.67 mM each dATP, dGTP, dTTP) and 1 µL of 1 mM dCTP were added to create an unequal dNTP mix. Finally, dependent on the desired labeling 2 µL of 1 mM Cy3 or Cy5 were added and the reaction was started by the addition of 1 µL Superscript II Reverse Transcriptase (Invitrogen; 200 u/µL). Samples were incubated for 2 h at 42 °C, after one hour 1 µL of Superscript II Reverse Transcriptase (200 u/µL) was added. To stop the reverse transcription reaction, 5 µL of 50 mM EDTA pH8 and 2 µL 10 M NaOH was added and incubated for 20 min at 65 °C, before a final neutralization of the mix using 4 µL of 5 M acetic acid. Labeled cDNA was purified using the Qia-quick PCR purification kit (Qiagen) with a final elution with 2 x 50 µL pre-warmed H_2O (42 °C).

The optical densities at 260 nm (cDNA concentration), 550 nm (Cy3 incorporation) and 650 nm (Cy5 incorporation) were measured in a spectrometer for the entire samples using 100 µL microcuvettes (Eppendorf). Amount and frequency of dye incorporation were calculated using the following formulas:

Frequency of dye incorporation [% labeled nucleotides] =
Dye incorporated [pmol] * 32.45 [% * ng / pmol]/ Amount of cDNA probe [ng]

with
Amount of cDNA probe [ng] = OD_{260} * 37 [ng / µL] * Total probe volume [µL]
Amount of incorporated Cy3 [pmol] = OD_{550} * Total probe volume [µL] / 0.15 [µL / pmol]
Amount of incorporated Cy5 [pmol] = OD_{650} * Total probe volume [µL] / 0.25 [µL / pmol]

Samples were volume-adjusted to have between 50 and 60 pmol of incorporated dye, and only samples with between 2 – 5 % of labeled nucleotides were used for hybridization.

2.2.4 Hybridisation and Scanning

Prior to hybridization the matched Cy5- and Cy3 samples were concentrated using microcon-30 filters (Amicon) to identical volumes of about 5 µL. Then both probes were mixed together and 5 µL of 10 mg/mL heat-denaturated salmon-sperm DNA was added. This mix was incubated for 2 min at 95 °C and afterwards chilled on ice. After the addition of 40 µL DIG easy hybridization buffer (Roche) a lifterslip (Erie Scientific Company) was placed on the printed microarray area, the probe was carefully injected under the lifterslip and the array was incubated overnight at 42 °C in a humid chamber.

The next day, the array was washed with freshly prepared array washing buffer (0.15 M NaCl, 15 mM Sodium citrate, pH 7.0) containing 0.1 % SDS at for 1 min, then again for 5 min. Finally residual SDS was washed off with array washing buffer in two additional washing steps (15 sec and 5 min) and arrays were dried by centrifugation (3 min at 1100 rpm in 50 mL Falcon tubes). Array slides were scanned using a Scanarray 4000 (Perkin Elmer LifeSciences), which used lasers to excite at 543 nm (570 nm emission filter) and 633 nm (670 nm emission filter) to measure the Cy3 and Cy5 fluorescence respectively. The photomultiplier tube (PMT) value was adjusted for each laser for each array to obtain a good range of intensities (the highest being less than saturated) and low background (values were always in a range of 63-83). Two pictures were obtained per slide for both dyes with a resolution of 5 µm. Pictures were analyzed using the software Quantarray (Packard BioChip Technologies). The median value of the signal detected for each spot at each wavelength and the local background were calculated. Low-quality spots were discarded including those with highly saturated signals, to avoid underestimation of the expression ratios.

Raw data were imported in form of a text file for the analysis of the results into GeneSpring (Silicon Genetics) software, were data were normalized and used to identify genes that were regulated only in the Rim101SLp time course but not in the control experiment, as described in the next paragraph.

2.2.5 Identification of Rim101p target genes

The exact selection procedure was the following: First we used the scatter plot view to arbitrarily choose a threshold of minimal regulation necessary to consider a gene as a potential Rim101SLp-regulated gene. In the scatter plot each gene is represented as a point within a 2-D diagram, and its coordinates are the expression values for the two different conditions compared, in our case two different time points of the TC experiment (see results Chapter 3.1.3 Figure 24). As in general the majority of genes are not regulated between two given conditions, the totality of genes is aligned as a "cloud" of points in the middle of the diagram close to the line marking a 1:1 ratio between the expressions measured under the two conditions. Genes that are far from this 1:1 line are the genes for which the expression differs between the two conditions. Genespring allows drawing parallel lines to this 1:1 ratio line which mark a foldchange ratio between the two conditions. Genes in the different sectors of such a divided diagram could then be selected. An arbitrarily chosen line marking a 1.4-fold induction or repression seemed to separate relatively well the potentially interesting outliers from the "cloud" of non-regulated genes (thus, lower foldchanges were considering as statistical artefacts). For this reason we decided in a first step to select only genes that showed an at least 1.4-fold induction or repression between timepoints 0 and 15 min and/or 0 and 90 minutes (this step yielded 1248 genes).

For these 1248 genes we observed now the transcriptional behaviour in the control experiments between the same time points. To be more stringent in the exclusion of possible false positive genes (genes with similar regulation in the CTRL experiment), we decided to divide the chosen genes in two groups, one containing genes with strong transcriptional changes (> 2-fold induction or repression between two conditions), the others with more modest regulations (between 1.4-fold and 2-fold regulation). This permitted us to define distinct selection thresholds for their transcriptional behaviour in the CTRL experiment: For strongly induced/repressed (> 2-fold) genes we allowed a maximal 1.4-fold induction/repression in the CTRL experiment, while we were much more rigorous with the modestly induced genes (less than 1.1-fold induction in the CTRL experiment was permitted). These criteria were applied for the transcriptional behaviour between both 0&15 and 0&90 min. The resulting 609 genes were subjected to a significance analysis using SAM (Significance Analysis of Microarrays (Tusher, Tibshirani et al. 2001) with a median "False

Significant Number" of 1.88 and a "False Discovery Rate" of 1.4). Normalised results for the identified 132 Rim101SLp-regulated genes are attached (Attachment 2).

2.2.6 Real time quantitative PCR confirmation

Gene expression was determined by real-time quantitative PCR using a LightCycler® (Roche Molecular Biochemicals). Suitable primers were chosen using the LightCycler® Probe Design Software 1.0, they are listed under 2.1.4. Total RNA was purified using the Qiagen MIDI kit as described above. The Superscript II RNase H-Reverse Transcriptase kit (Invitrogen) was used for reverse transcriptase assay from 1 µg of total RNA. For quantitative PCR cDNA samples were diluted 1:100. 20 µL PCR reactions contained 4 mM $MgCl_2$, 0.5 µM of each primer and 2 µL LightCycler® FastStart DNA Master SYBR Green I for 5 µL of template cDNA, and PCR cycles were started at 95 °C for 8 min, followed by 45 cycles at 95 °C for 10 s, 55 °C for 7 s, and 72 °C for 10 s. A negative control with sterile water was performed for each primer set. The threshold cycle was determined as the cycle above which the fluorescence signal reached a baseline level. All measurements were performed in duplicate (technical replicate) and on samples from two independent biological replicates. The expression levels of the genes were determined relative to the expression of the *ACT1* gene.

2.2.7 Statistical tests

The Chi^2 test – or the Fisher exact test when more appropriate – was used for data analysis. Data with $p < 0.05$ were considered as statistically different.

2.3 Chromatin Immunoprecipitation (ChIP)

2.3.1 Protein extraction

Cells from midlog cultures in SC medium (buffered with 150 mM HEPES at pH 4 or pH 7) were pelleted and stored at −80 °C prior to protein extraction. Cells were resuspended in assay buffer (5 % glycerol, 50 mM KCl, 100 μM NaCl, 2 mM EDTA, 0.1 % β-mercapto-ethanol 1 mM PMSF (phenylmethylsulfonyl fluoride) and protease inhibtors (Roche, 1 pill/50 mL) and transferred to 2 mL test tubes containing 1 mL of acid-washed glass beads. Cells were lysed by 5 cycles of 2 min vortexing at maximal speed, each followed by 2 min chilling on ice. Cell debris was removed by centrifugation, and supernatants were aliquoted and stored at −80 °C.

2.3.2 Western Blots

For Western blot assays, 20 μL of 2× sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer was added to 20 μL of each sample and the mix was boiled for 5 min. Samples were loaded on an SDS-8 % PAGE gel and run for 1.5 hours at 100 V. Proteins were transferred to nitrocellulose and blocked for 1.5 hours with 2 % nonfat milk in TBST buffer (10 mM Tris-HCl pH 8.0 , 150 mM NaCl, 0.05 % Tween 20). Then 1:5000 of Anti-V5-Antibody was added to the milk and incubated for 1.5 hours at room temperature. After three washing steps with TBST buffer to discard nonbound antiserum, the secondary horseradish peroxidase coupled antibody (Invitrogen) was added and incubated for more 1.5 hours before three final washing steps with TBST. We used autoradiographic films and the ECL kit (Amersham) for development (2-5 min).

2.3.3 Chromatin extraction

Overnight precultures were diluted to the start OD 0.2 with 100 mL fresh SC medium buffered with 150 mM HEPES at pH 4 or 7 and grown at 30 °C until midlog phase. Preparations of chromatin were performed essentially as described by Kuras *et al.* (Kuras and Struhl 1999) but with certain modifications. Briefly, 3 mL of 37 % formaline were carefully

mixed with 100 mL of culture before and incubated for 10 min at room temperature. After addition of 20 mL of 2.5M glycine to stop the reaction and a 5-min incubation, cells were chilled on ice and harvested by centrifugation, washed with cold 20 mM Tris pH 8 and resuspended with cold 500 µL FA buffer (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 % Triton X-100, 1 mM EDTA, 0.1 % sodium deoxycholate, 1 MM PMSF + protease inhibitors (Roche, 1 pill/50 mL)). Cells were then transferred to 2 mL test tubes containing 1 mL of acid-washed glass beads and lysed by 5 cycles of 2 min vortexing, each followed by 2 min chilling on ice. To separate the crude extract from the beads, holes were made in the bottom of the tube with a glowing 18-gauge needle, and crude extract was transferred into a new tube by centrifugation. Glass beads were washed with 1 mL FA buffer and the flowthrough was recovered in a second centrifugation. After resuspension and (optional) two hours of incubation at 4° C on a tube roller samples were centrifuged with 12.000 rpm for 20 min at 4 °C. Pellets were homogenized in 1.8 mL FA buffer. Finally, chromatin was sonicated in fragments of an average size of 500 bp (Branson sonifier 250 with 30-40 % output and 90 % duty cycle), each sample 7 times for 5 seconds with intermediate chilling on ice, and afterwards centrifuged with 10.000 rpm at 4 °C for 20 min. Different centrifugation speeds were tested without a strong influence on the result. The supernatant containing the chromatin fragments in the clarified lysate could be aliquoted and stored at -80 °C until use.

2.3.4 Immunoprecipitation

Two different IP approaches were validated by Western Blot and tested under varying conditions, the first being a direct immunoprecipitation with Anti-V5 Agarose affinity gel (Sigma-Aldrich), the second an indirect protocol using a free Anti-V5-Antibody (Sigma-Aldrich) with subsequent precipitation on a Protein A sepharose (Amersham) matrix. 250 µL of clarified lysate were used in IP reactions, and the whole procedure was performed at 4 °C or on ice. The indirect immunoprecipitation protocol was with or without an optional step of lysate preclearing by incubation for 2 hours with 80 µL of Protein A sepharose without Anti-V5-Antibody. The lysate was added to a new reaction tube and incubated for 30 min with 5 µL of Anti-V5-Antibody. Then, 80 µL of Protein A sepharose were added and the mix was incubated for 2-4 hours. Then the matrix was washed to discard unspecifically bound material. Generally two washing steps with FA buffer were followed by two or three additional washing steps with FA buffer supplemented with 500 mM NaCl, each washing

interrupted by 10-15 min of incubation. In some cases, one washing step with 10 mM Tris-HCl pH 8, 0.25 M LiCl, 10 mM EDTA, 10 % NP40; 0.5 % sodium deoxycholate and/or a final washing step with TE buffer (Tris-HCl pH 8 10 mM, EDTA 1 mM) were added to wash off DNA unspecifically bound to the sepharose matrix. For elution, 125 µL elution buffer (25 mM Tris- HCl pH 7.5, 5 mM EDTA, 0.5 % SDS) were added and samples were heated for 20 min to 65 °C. Elution was repeated with the same volume. Then the crosslinking reaction was reversed by overnight incubation at 65 °C; alternatively, decrosslinking was performed directly on the matrix beads together with the second elution (with similar results). In parallel, an aliquot of starting material lysate was also decrosslinked as an input reference for real-time PCR quantification. Finally, 1 µL of 20µg/µL Proteinase K was added and incubated for 30 min at 50 °C with the sample and DNA was purified using the Qiagen protocol. For promoter DNA quantification after chromatin immunoprecipitation, nondiluted eluate samples were used, while input samples were diluted 1:100 and served as a reference. Each real time qPCR experiment was done in a duplicate. Primers used to verify the expression of 20 particular genes are listed in 2.1.4, the localization of the amplified promoter fragments can be seen in Chapter 3.2.3 Figure 37.

2.4 β-galactosidase tests

YPD precultures were diluted to the start OD_{600} 0.2 with fresh YPD buffered with 150 mM HEPES at pH 4 or 8 and grown until midlog phase (OD600 = 1). Pellets were washed with water and resuspended in 450 µL Z buffer (60 mM Na_2HPO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercapto-ethanol) to an OD_{600} = 5. Cells were lysed by the addition of 0.2 % Sarcosyl and incubated for 30 min at 30 °C. Then 150 µL of 4 mg/mL ONPG (O-Nitrophenyl-β-D-galactopyranoside, suspended in Z buffer) was added and tubes were incubated until a yellow colour was visible (for a maximum of three hours); the reaction was stopped by the addition of 400 µL of Na_2CO_3 and the reaction time was noted. Assays were centrifuged after measurement of the OD420 of the supernatant β-galactosidase activity was determined:

β-galactosidase activity [Miller Unit] = 1000 * OD420/(t [min]* V [mL] * OD600)

with t = Reaction time; V = Volume of culture assayed in milliliters

3 Results and Discussion

3.1 Microarray to identify Rim101p target genes

3.1.1 Strain Construction and Microarray Time Course

The project had been started before my arrival in the laboratory. In the first paragraph, I will briefly summarize the state of the project when I began my work. A constitutively active version of the transcription factor Rim101p named Rim101SLp (Rim101p Short Length) had been constructed and placed under the control of the strong inducible *MET3* promoter. The processing site of *C. albicans* Rim101p was predicted through Hydrophobic Cluster Analysis (Gaboriaud, Bissery et al. 1987) assuming that the putative cleavage site of Rim101p shares the same environment as the known processing site of the *A. nidulans* homologue PacC (Also see 1.4.1). A plasmid carrying *RIM101SL* (pINA1341, Figure 19a) had been integrated in the *HIS1* locus of *rim101* disrupted strain DAY5 to create strain FB8. Thus, the expression and activity of Rim101SLp in this strain is not pH-dependent any more, but can instead be triggered by omitting methionine and cysteine from the medium. As a control, strain FB1 identical to FB8 but devoid of *RIM101SL* (See Figure 19b and attached publication) was constructed.

The functionality of Rim101SLp was then validated in two different ways. First, since the *rim101* null mutant exhibits a severe growth defect at pH10, we verified that Rim101SLp rescued in strain FB8 the *rim101* null mutant growth defect at pH10 under methionine/cysteine starvation, but not when these amino acids were present (Figure 20a).

Second, overexpression of *RIM101SL* when the medium was lacking sulphur amino acids was confirmed by quantitative real-time PCR. The resulting protein was able to regulate the transcription of a known Rim101p target, *PHR1*, in a similar way as the activated wild type version. Figure 20b shows the induction of *RIM101SL* transcription and of the Rim101p target gene *PHR1* at different time points after the change from repressing to inducing conditions in SC medium at pH 5.5 (where *PHR1* is normally not trancribed). Although *RIM101SL* is overexpressed approximately 10-fold compared to the wild type transcription at alkaline pH, the transcriptional levels measured for *PHR1* mRNA reflect well those observed in a reference strain at alkaline pH (Cornet, Bidard et al. 2005).

a)

Methionine/Cysteine

- +

DAY25 (*ΔΔrim101*)

DAY185 (WT)

FB8 (*RIM101SL*)

b)

Induction of Rim101 activity

% actin

1000
100
10
1
0

0 10 20 30 40 50 60 70 80 90 100

time [min]

RIM101SL transcripts

PHR1 transcripts

***Figure 20**: **a)** Validation of inducible Rim101 activity of strain FB8 on SC pH10 plates under MET3-inducing conditions in the Rim101SL strain (absence of methionine and cysteine) and under repressing conditions (presence of methionine and cysteine). Growth is compared to rim101-/- disruption strain DAY25 and to reference strain DAY185 **b)** Real time PCR transcript quantification for RIM101SL and the Rim101-induced gene PHR1 at 0, 15, 30, 45, 60 and 90 min growth in SC medium without Met/Cys*

Sequence 5' end of *RIM101* mRNA

-25 bp 0 174 bp

1st ATG

2nd ATG

Start codon
www.genolist.pasteur.fr/CandidaDB

5'UTR

Start codon www.candidagenome.org

***Figure 21:** Analysis of the RIM101 transcript revealed that both possible start codons are transcribed with a 25 bp UTR sequence*

Five time points (0, 15, 30, 60 and 90 minutes after *MET3* promoter induction) were selected for the microarray time course to reflect the early transcriptional events due to Rim101SLp activity. Finally, the global transcriptional pattern of all genes were monitored at these time points by hybridising samples taken from two independent biological replicates of strain FB8 against a reference pool of all time points together obtained from a third biological replicate. Two arrays were used for each single sample, but in the second the labeling of sample and reference were inversed to correct for unwanted dye effects (Dye swap; see also Introduction 1.2.2.3 paragraph “Fluorescense labeling and dye effects”). The microarray time course experiment led to the complete transcriptional profiling of 5889 *C. albicans* genes after induction of Rim101SLp activity. Please note that, to simplify the description of the microarray analysis, the term “gene” in “5889 *C. albicans* genes” for example is used here instead of “a microarray probe that recognizes *C. albicans* genes”, but it should be kept in mind that in some cases one gene is represented by several probes on the array, and that in few cases of highly similar genes one probe might recognize multiple genes.

3.1.2 Identification of the transcriptional start of the *RIM101* gene

As already mentioned in the introduction, there was a certain confusion concerning the start codon of *RIM101*, and in consequence also the correct size of the *RIM101* transcript and the position of the cleavage site for Rim101p activation was unclear (Introduction 1.4.4.4). Two alternative starts differing by 174 bp are still proposed in the sequence databases at http://genolist.pasteur.fr/*Candida*DB/ and http://www.*Candida*genome.org/. Our truncated version was constructed with the assumption that the first possible start codon was transcribed and might thus be the first translated one. Although we already had phenotypic proof that Rim101SLp was expressed and fully functional, we decided to analyze the *RIM101* transcript for evidence. We analyzed the 5’end of *RIM101* mRNA of a reference strain at both acidic and alkaline pH using RACE technology (2.2.1). Our results showed that the transcript starts independently of the pH with a 25 bp 5’ UTR sequence upstream of the first possible start codon (Figure 21).

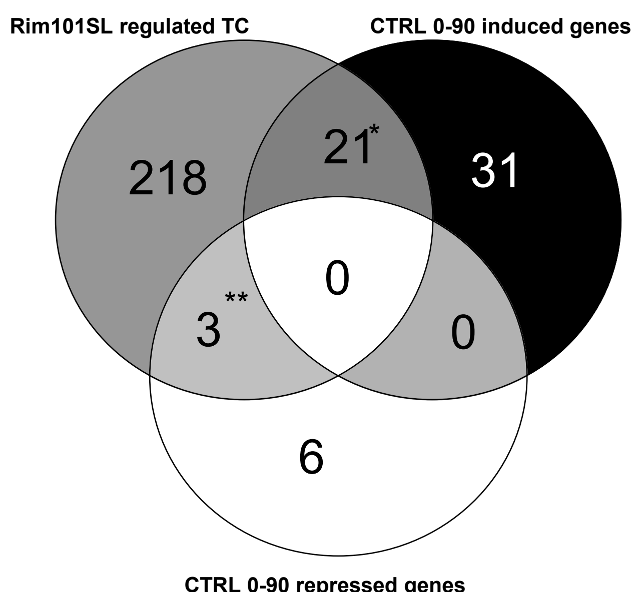


Figure 22: *Venn diagram of genes identified by applying the SAM software to the time course and the control experiment 0-90. The numbers of genes in common between the two experiments are indicated in the overlapping regions; common upregulated genes are listed in Table 6 (below).*

*** Common induced genes**			
orf19 name	**Name**	**Array**	**Predicted function**
orf19.541	IPF4290	CA4252	Unknown function
orf19.946	**MET14**	**CA5404**	**Adenylylsulfate kinase**
orf19.1159	**IPF7616**	**CA1569**	**Putative homoserine O-acetyltransferase**
orf19.1639	IFH1	CA0136	Dioxygenase
orf19.2028	**MXR1**	**CA0123**	**Methionine sulfoxide reductase**
orf19.2693	IPF7968	CA3260	Unknown function
orf19.2738	**SUL1**	**CA2698**	**High-affinity sulfate transport protein**
orf19.4076	**MET10**	**CA1620**	**Sulfite reductase flavin-binding subunit**
orf19.4099	**ECM17**	**CA4320**	**Putative sulfite reductase**
orf19.4536	**CYS4**	**CA4195**	**Cystathionine beta-synthase**
orf19.5025	**MET3**	**CA5238**	**ATP sulfurylase**
orf19.5059	**GSH1.exon2**	**CA0583**	**Gamma-glutamylcysteine synthetase, exon 2**
orf19.5060	**GSH1.exon1**	**CA0584**	**Gamma-glutamylcysteine synthetase, exon 1**
orf19.5280	**MUP1**	**CA4972**	**High affinity methionine permease**
orf19.5312	IPF8210	CA5480	Unknown function
orf19.5811	**MET1**	**CA4162**	**Siroheme synthase**
orf19.6398	IFH2	CA5130	Dioxygenase
orf19.6757	IPF3485	CA5940	Aldo/keto reductase
****Common repressed genes**			
orf19 name	**Name**	**Array**	**Predicted function**
orf19.492	ADE17	CA4513	Ribotide transformylase
orf19.3554	AAT1	CA2661	Aspartate aminotransferase
orf19.3707	YHB1	CA0943	Flavohemoglobin

Table 6: *List of genes that were found to be regulated in the time course (TC) and the control experiment (CTRL) 0-90: The majority of the induced genes had a predicted function which is linked to the metabolism of sulphur amino acids (marked in bold).*

3.1.3 Microarray Control Experiments and False Positive Filter

As a consequence of our choice for the use of the *MET3*-promoter to permit a pH-independent set-up, we expected to observe transcriptional changes not only due to Rim101SLp activity, but also as a response to the absence of methionine and cysteine in the medium. To account for these unwanted side effects, we performed an identical experiment with strain FB1 (isotrophic *rim101* null mutant strain) and used the end points (time 0 and 90 min) of the time course to identify such false positives.

A first evaluation of the results using SAM (Significance Analysis of Microarray Data) revealed that 242 genes showed significant transcriptional changes during the TC experiment, while 52 upregulated and 9 downregulated genes could be identified in the control experiment (CTRL). A comparison of TC and CTRL result revealed that a total of 24 (21 induced and 3 repressed) genes were identified as significantly regulated in both experiments (Figure 22; Table 6).

These common genes were almost exclusively genes that code for proteins with a probable function linked to sulphur amino acid starvation (Table 6), which clearly confirmed that our control experiment could be used to filter out such false positives. To more efficiently exclude false positive genes, we added a second control experiment with time point 15. We focused on the Rim101SLp-independent regulation immediately after the switch to inducing conditions, because globally in the periode between 0 and 15 minutes the most drastical transcriptional changes seem to occur (Figure 23).

Furthermore we opted for a less stringent method than SAM for the definition of a "false positive gene" to minimize the risk that such a gene could be later on considered as a "good candidate". We preselected genes that showed at least 1.4-fold transcriptional changes between the early or late phase of *RIM101SL* inducted and noninduced conditions (Figure 24), which was the case for 1248 genes. Then we compared the transcriptional behaviour of each gene between TC and CTRL experiments, and following strictly defined selection criteria we excluded any gene that appeared to be regulated in a similar way in both experiments (described in detail in Materials and Methods 2.2.3).

We ended up with 609 genes with transcriptional changes in the TC experiment that were not observed in any control experiment. These genes were finally subjected to a significance analysis with SAM. This led to the identification of 133 genes that all showed

Transcription profiles after *RIM101SL* induction

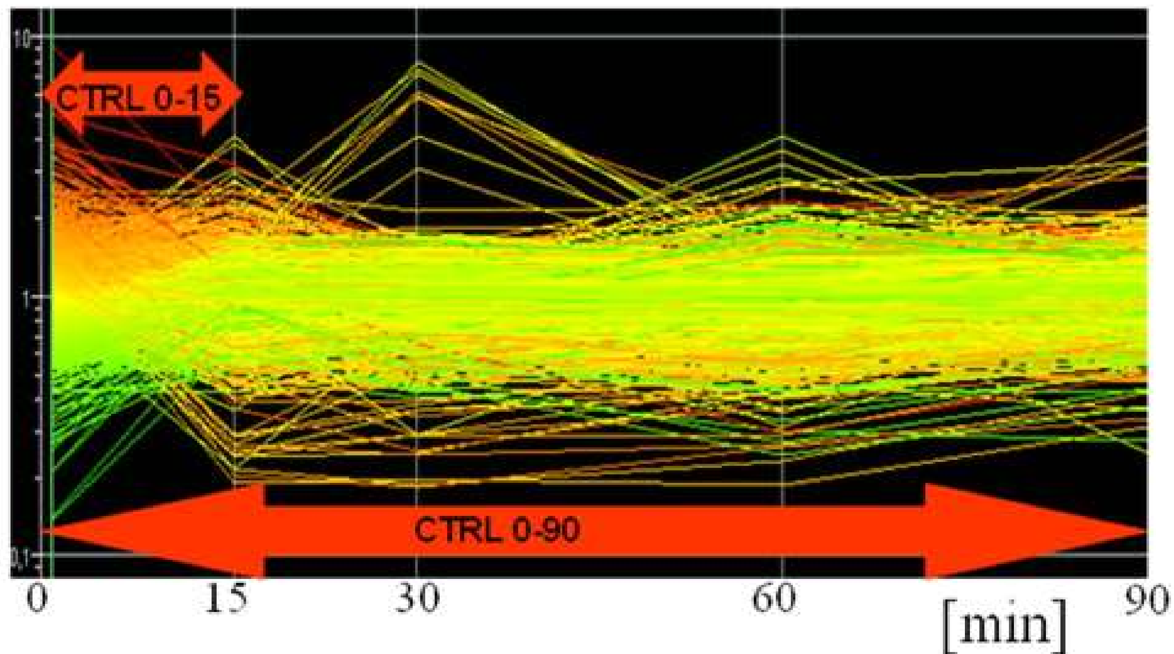


Figure 23: *Diagram with the various transcription patterns of all 5889 genes that were detected in the microarray time course after induction of RIM101SL. Two control experiments were carried out to identify "false positive" genes (CTRL0-15 and CTRL0-90)*

Scatter Plot of TC time points 0' and 15'

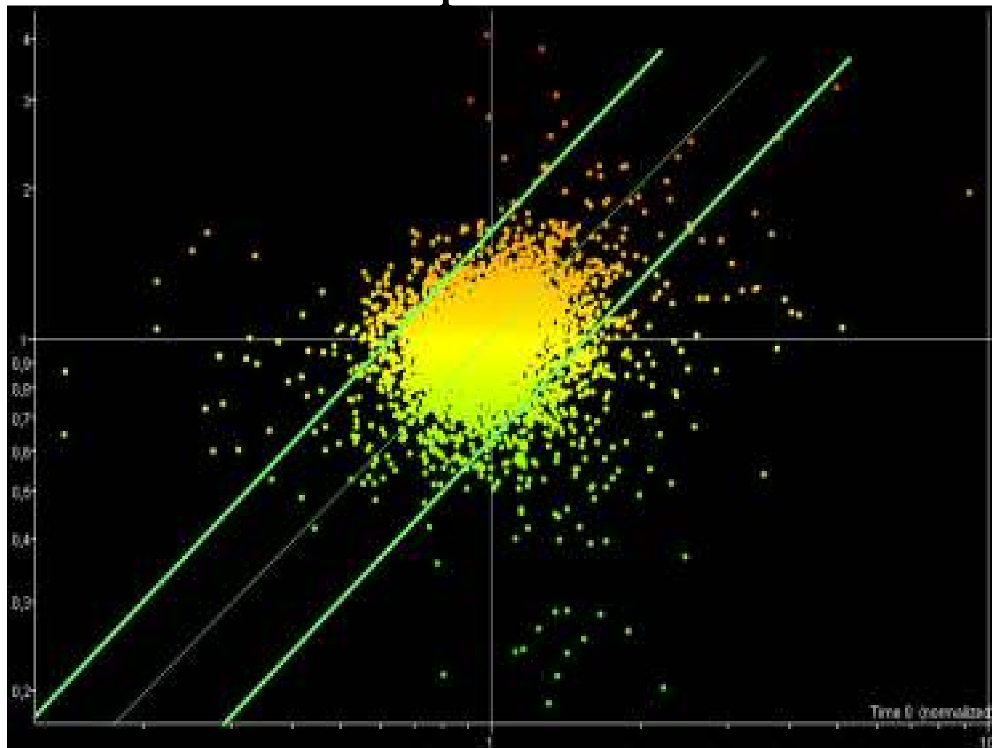


Figure 24: *Genespring* TM *Scatter Plot view of the transcriptional changes between time 0 and 15 minutes after the change to inducing conditions. The fine diagonal line in the middle marks a 1:1 transcription ratio, the parallel lines on its left and right represent a 2 fold induction or repression respectively. To identify false positive genes, genes with at least 1.4 fold transcriptional changes were selected and the corresponding Scatter Plot of the control experiment was used to detect genes that behave similarly (See Materials and Methods).*

significant transcriptional changes in response to Rim101SLp activity, even if they had very different patterns of regulation (Figure 25). Two of these 133 probes are CA4349 and CA4351, which recognize the 3' and 5' end of the gene *FUM12* which codes for a predicted fumarate hydratase gene, so that the effective number of identified genes is reduced to 132.

After 15 minutes 98 genes exhibited already at least 1.4-fold induction or repression compared to point 0. After 90 minutes only slightly more genes (107) were regulated. In spite of these similar numbers of regulated genes, there are 2.5 times more repressed than induced genes after 15 minutes (70 repressed compared to 28 induced genes), while after 90 minutes this ratio drops to 1.32 and is less impressive (61 repressed compared to 46 induced genes), suggesting that Rim101p may act primarily as a repressor.

Transcription profiles of identified Rim101SLp-regulated genes

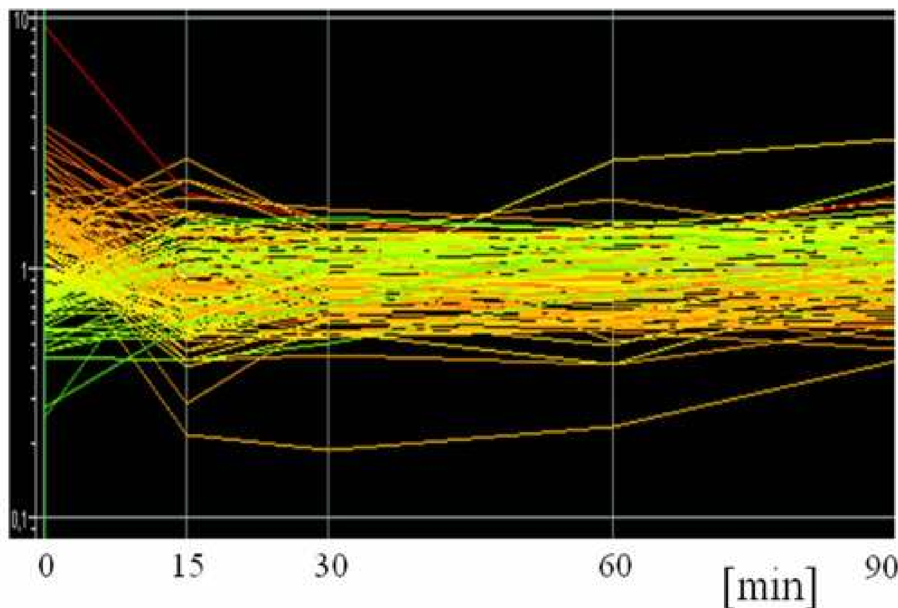


Figure 25: *Various transcriptional patterns of all 133 genes that were found to be regulated by Rim101SLp between time 0 and 15 minutes after the change to inducing conditions. The patterns of repressed genes are represented by orange lines, while induced genes are shown with green lines.*

3.1.4 Identification and Clustering of Rim101p target genes

Various clustering methods were tested to classify these 132 genes into groups of similar regulation patterns, including different forms of Self-Organizing Maps (SOM), K-mean clustering and Hierarchical Clustering.

We finally opted for a so-called "gene tree", a hierarchical clustering method available in the Genespring^{TM} software. Basically, Genespring^{TM} calculates the similarity coefficients for every possible combination of two transcriptional patterns. Then the couple of genes with

Classification of transcription patterns

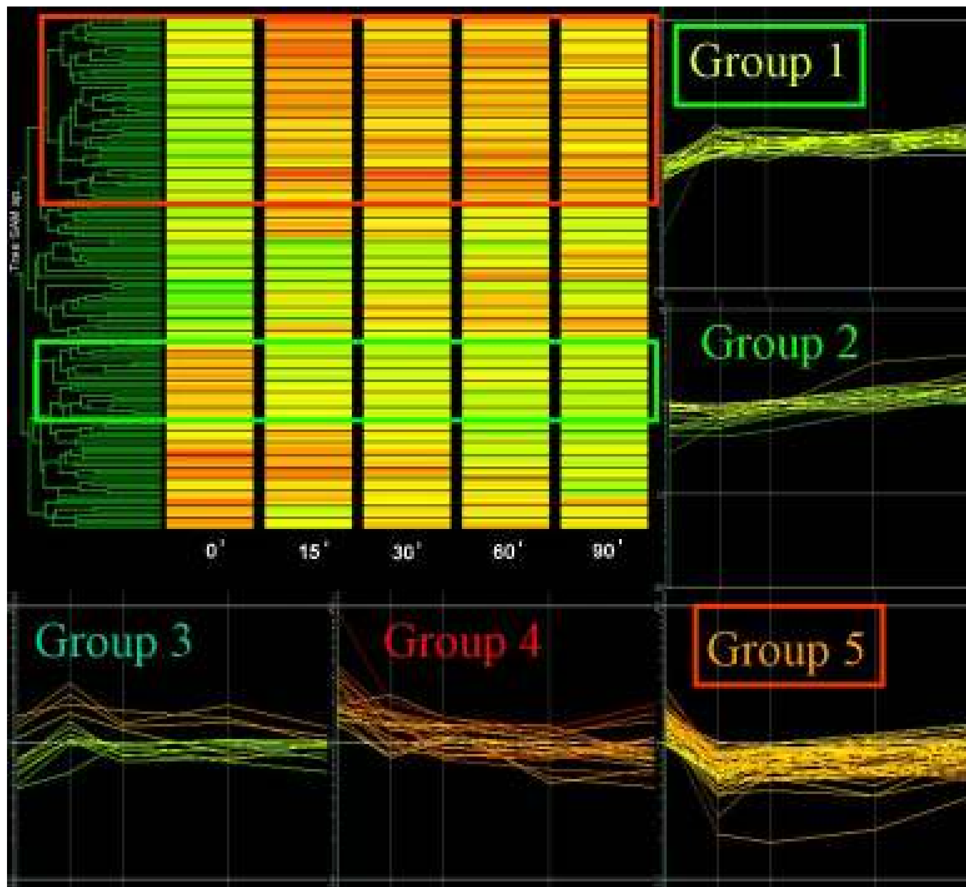


Figure 26: *Gene classification in the Genespring*^{TM} *tree view. Green colours indicate a strong expression, orange colours a weak expression at the corresponding time points. Genes with the most similar transcriptional patterns are neighboured. From this classification five distinct groups with similar expression patterns could be manually generated; the "branches" belonging to Group 1 and 5 (immediately induced and immediately repressed genes) are highlighted in the treeview.*

the highest similarity is branched together and its mean expression pattern is calculated. In the next iteration the correlation coefficients for this mean pattern with all other genes are calculated, and the next most similar patterns are branched together. This procedure is repeated until all genes are connected in a tree. The result is a dendrogram similar to that of phylogenetic trees, in which the direct link between any two genes provides information about how similar or distant their transcriptional patterns are (Figure 26). We then arbitrarily divided this tree into five "branches" that contain homogenous classes of highly similarly regulated genes: Class1/Class2 comprise genes which are immediately induced/repressed and remain so until the end of the experiment; Class3/Class4 regroups genes which are induced/repressed progressively throughout the experiment; Class5 genes undergo transient induction at early stages of the experiment (Figure 26, see attachments to find the genes belonging to each of the classes).

3.1.5 Sequence Analysis

To get an idea about the cellular processes affected by Rim101p activity, we had a look at the predicted function of the identified genes. Assuming functional conservation between *C. albicans* genes and its *S. cerevisiae* orthologues, we found that 109 out of 132 genes could be associated with a biological process according to the MIPS database (http://mips.gsf.de/proj/yeast/CYGD/hemi/).

Almost one third of the genes had a predicted function in metabolism, which is a relative high proportion compared to barely more than one fifth of the whole genome (31.1 % versus 18.4 %; $p < 0.0011$) throughout the genome (see Table 7). Among the different metabolic functions, carbohydrate metabolism (12.9 % vs. 6.8 %; $p < 0.02$), but also amino acid metabolism (8.3 % vs. 4.2 %; $p < 0.05$) were overrepresented compared to the whole genome.

Functional classification of identified target genes

MIPS Functional Class	Rim101SLp regulated genes	All *C. albicans* genes	Ratio 132 /All
Metabolism	**41 (31.1%)**	**1029 (18.4%)**	**1.7**
Energy	3 (2.3%)	314 (5.6%)	0.4
Cell cycle and DNA proc.	1 (0.8%)	506 (9.0%)	0.1
Transcription	3 (2.3%)	558 (10.0%)	0.2
Protein synthesis	9 (6.8%)	376 (6.7%)	1.0
Protein fate	15 (11.4%)	641 (11.4%)	1.0
Cellular transport	2 (1.5%)	734 (13.1%)	0.1
Signal transduction	1 (0.8%)	192 (3.4%)	0.2
Cell rescue	**13 (9.8%)**	**382 (6.8%)**	**1.4**
Transposable elements	1 (0.8%)	3 (0.1%)	14.2
Biogenesis of cell. comp.	**20 (15.2%)**	**493 (8.8%)**	**1.7**
Unclassified proteins	23 (17.4%)	379 (6.8%)	2.6
Total	132 (100.0%)	5607 (100.0%)	

Table 7: *Distribution of the identified genes in the different predicted functional classes compared to the global functional distribution. Three important functional classes that seem to be particularly affected by Rim101p activity are shown in grey.*

Another cellular process that was obviously affected was the biogenesis of cellular components with 15.2 % (20 genes) abundance in the selected genes versus 8.8 % throughout the genome ($p < 0.022$). Interestingly, 16 of these genes (or 12.1 % of the 132 selected) have a predicted function in cell wall organization and biogenesis, compared to only 2 % of the whole genome ($p < 0.0003$). This indicates that Rim101p participates in the modulation of the

the cellular surface structure in response to pH changes through the transcriptional regulation of involved genes. Rim101p is known to be required for the pH-dependent yeast-to-hyphae transition, but the fact that our experimental setup prevented the formation of hyphae (low pH, low temperature) suggests that Rim101p activity also affects the cell wall composition under conditions where the cellular morphology is not affected.

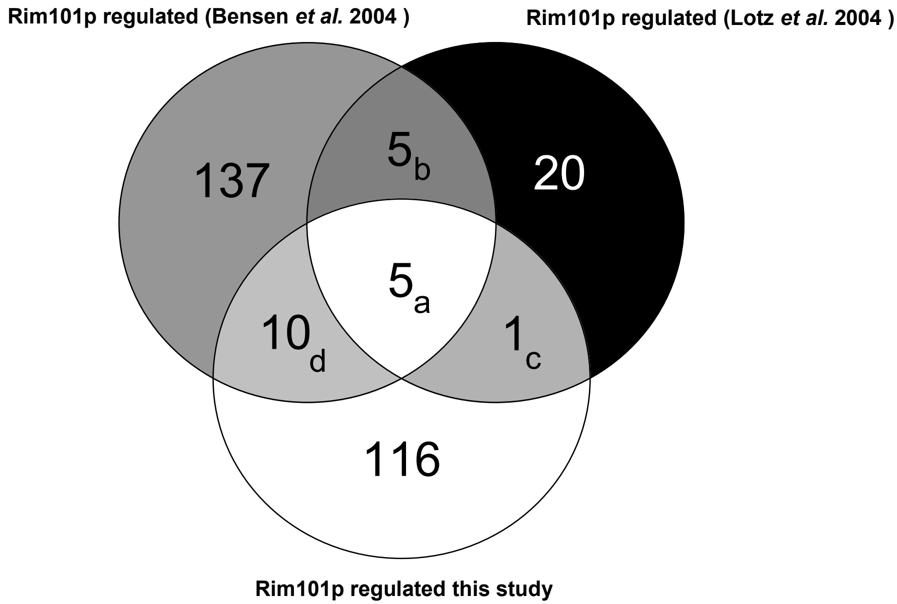
Finally, a third functional class that was slightly overrepresented within the identified Rim101p-regulated genes was involved in cell rescue; 9.8 % or 13 genes were predicted to have a related function (versus 6.8 % throughout the genome), interestingly 4 of these genes are involved in the cellular homeostasis. A role of Rim101p as a homeostasis regulator makes intuitively sense given the important impact of pH changes on the ion household.

In contrast, it might be expected that metabolism and in particular amino acid metabolism would be affected given that we needed to change to sulphur amino acid starvation conditions to activate *RIM101SL* transcription. However, the applied selection criteria were quite stringent (3.1.3 Microarray Control Experiments and False Positive Filter), and a closer look on the concerned group of genes indicated that the presence of sulphur amino acid metabolism genes is rather poor and that other amino acids (for example arginine metabolism) are strongly affected.

A research in the *Candida* genome database http://www.*Candida*genome.org/ revealed that nine of the 132 genes were already annotated as pH-regulated (see Table 3 (1.4.4.4)): *ARG1* (orf19.7469), *PHR1* (orf19.3829), *PHR2* (orf19.6081), *GCV2* (orf19.385), *KRE6* (orf19.7363), *COX15* (orf19.3656), *CPA1* (orf19.4630), *MNN1* (orf19.4279) and *ZRT2* (orf19.1585). While the role of Rim101p in the pH-dependent regulation of the β-1,3 Glycosidase genes *PHR1* and *PHR2* has already been well described, only three of the other genes have been described as Rim101p-regulated. These are *GCV2*, a gene involved in glycin catabolism, the putative α-1,3-mannosyltransferase *MNN1* and *KRE6* which is involved in β-1,6-glucan synthesis. Interestingly, *KRE6* is annotated as Rim101p-induced at alkaline pH according to the results of Bensen *et al.* (Bensen, Martin et al. 2004), but it was found to be Rim101p-repressed by Lotz *et al.*, a result that is confirmed by our experiments. Finally, the four other genes *ARG1* (Argininosuccinate synthase), *COX15* (Unknown function), *CPA1* (Unknown function) and *ZRT2* (Zinc transporter) are all described as “repressed at alkaline pH” after Bensen *et al.* (Bensen, Martin et al. 2004). Interestingly, no Rim101p-dependent regulation is annotated for these genes in the database, although Bensen *et al.* identified *ARG1* as an alkaline Rim101p-upregulated gene. We found a clear Rim101SLp-dependent repression for *ARG1*, *COX15* and *CPA1*, but identified *ZRT2* as a Rim101SLp-induced gene

in our time course experiment. 21 additional genes that are annotated as pH- and/or Rim101p-regulated were also found to be Rim101p-regulated in the time course experiment, but were considered as false positives because they were similarly regulated in the CTRL experiment (See 4.1 “Clonclusions” Table 12). Most of them were induced or repressed in the TC experiment consistently with their database annotation. The filtered data set of 609 genes contained not more than the previously mentioned nine annotated pH-regulated genes, indicating that the loss following the significance analysis might be less important than in the false positive rejection.

Finally, a direct comparison between the complete sets of Rim101p-regulated genes identified in the large scale experiments by Bensen *et al.*, Davis *et al.* and us yielded a relatively poor overlap (Figure 27). Furthermore, the results are partially contradictory for two of these five common genes, *ALS1* and *KRE6*.



b) Common regulated genes

a		b		c		d	
Gene name	**Orf number**	**Gene name**	**Orf number**	**Gene name**	**Orf number**	**Gene name**	**Orf number**
PHR1	orf19.3829	*PGA13*	orf19.6420	*PHO11*	orf19.2619	*IPF3485*	orf19.6757
ALS1	orf19.5741	*RBT1*	orf19.1327			*IPF407*	orf19.7504
KRE6	orf19.7363	*HWP1*	orf19.1321			*ARG1*	orf19.7469
PHR2	orf19.6081	*CRH11*	orf19.2706			*ASR2*	orf19.7284
PGA52	orf19.1911	*PGA6*	orf19.4765			*IPF8762*	orf19.822
						CIT1	orf19.4393
						GCV2	orf19.385
						MNN1	orf19.4279
						IPF19908	orf19.1344
						PHO87	orf19.2454

Figure 27: a) *Venn diagram comparing the results of our microarray study with those published by two other groups. Numbers in the overlapping regions indicate how many genes were common between the different studies, gene names can be found in the table below with help of the indices* ***b)*** *List of genes that were identified in more than one experiment.*

3.1.5.1 Signal Peptides and Transmembrane Motifs

As we were especially interested in genes coding for proteins that might be involved in direct host-pathogen interaction, we scanned the coding sequence of the selected genes for features that might indicate a localisation at the cellular surface. We used the "TMHMM" prediction server (http://www.cbs.dtu.dk/services/TMHMM/) to search for probable transmembrane domains, and the "SignalP 3.0 server" (http://www.cbs.dtu.dk/services/SignalP/) to find out whether the sequence possesses a signal peptide cleavage site.

For twenty genes a signal peptide or signal anchor was predicted, and 33 of the sequences are likely to include at least one transmembrane domain (See 6. Appendix Attachment 1: Clusters and raw data for Rim101SLp time course).

3.1.5.2 Rim101p Promoter Motifs

Rim101p has been shown to bind directly to different sites on the promoter regions of the *PHR1* and the *PHR2* gene. From the Rim101p orthologue PacC in *A. nidulans* it is known that it binds specifically to a short "5'-GCCARG-3'" sequence motif on the promoters of directly regulated genes. It has been proposed that the Rim101p recognition site might diverge from this site and be "5'-CCAAG-3'" with preference for three additional 3' "A", but recently it has been evidenced that the importance of the different positions in the sequence promoter-specific, so that sometimes variations of the above mentioned motif might still allow Rim101p binding and the regulation of the gene (Baek, Martin et al. 2006).

As the TC experiment was focused on the transcriptional events immediately after induction of Rim101SLp activity, we expected to identify direct target genes of Rim101p. Consequently, we checked whether the frequency of the above mentioned binding sites was significantly higher in the promoters of the selected genes than in the whole genome. This search was performed using "Regulatory Sequence Analysis Tools" web software (http://rsat.scmbb.ulb.ac.be/rsat/). The size of the promoter used for this analysis was limited to 1000 bp or to the start/stop of the adjacent ORF. A short sequence motif "CCAAG" was found in 63.2 % of the selected genes at least once, compared to a presence in 45.9 % of all promoters ($p < 0.04$). Moreover, the mean number of such motifs per promoter is almost 45 % higher in the selected genes than in the whole genome (1.08 motifs per promoter for the selected genes compared to 0.74 motifs per promoter for all genes). Similarly, the PacC

binding motif “GCCARG”, the motif “CCAAGAAA” and related motifs were significantly more abundant in the promoter of the selected genes than in the complete genomic promoter set (Figure 28).

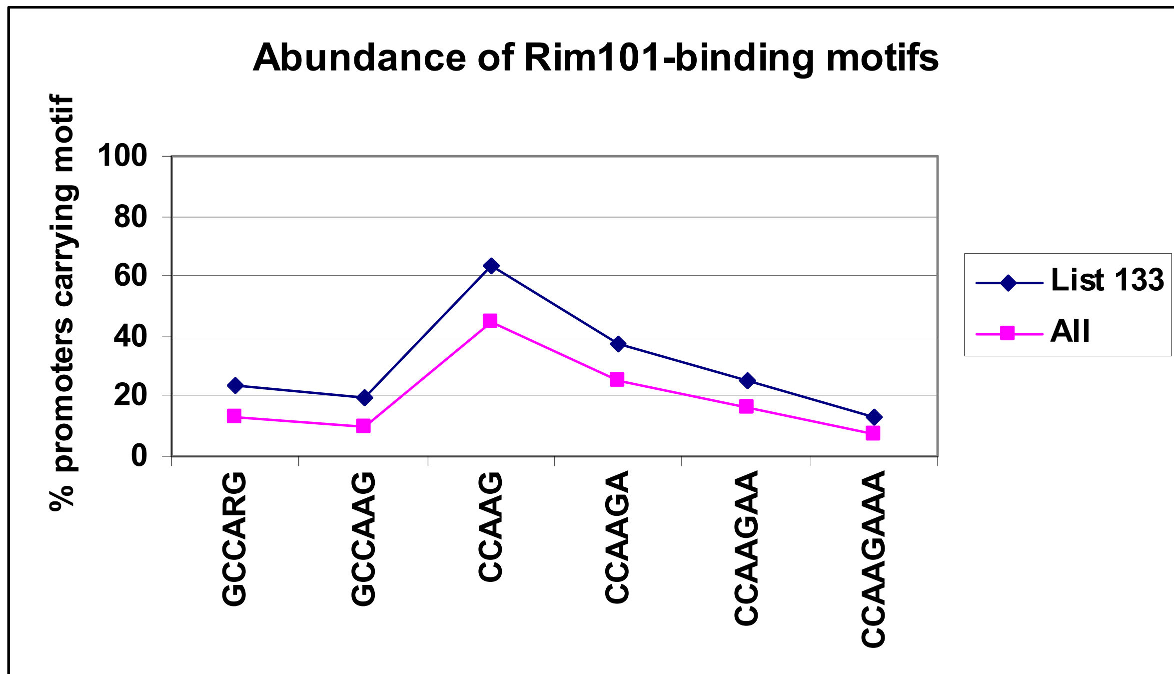


Figure 28: *In this diagram the abundance of different possible Rim101p binding motifs in the promoters of the 133 selected genes is compared to their abundance in all promoters. For each gene the 1000bp upstream of the start codon were considered as possible promoter; if an adjacent ORF was found in this upstream region, the promoter sequence was shortened until to the start/stop of this ORF. The presence of repeated motifs in a promoter is not considered here. All different motives searched were significantly (likelihood >95%) more abundant in the group of 133 genes than in all promoters of the genome.*

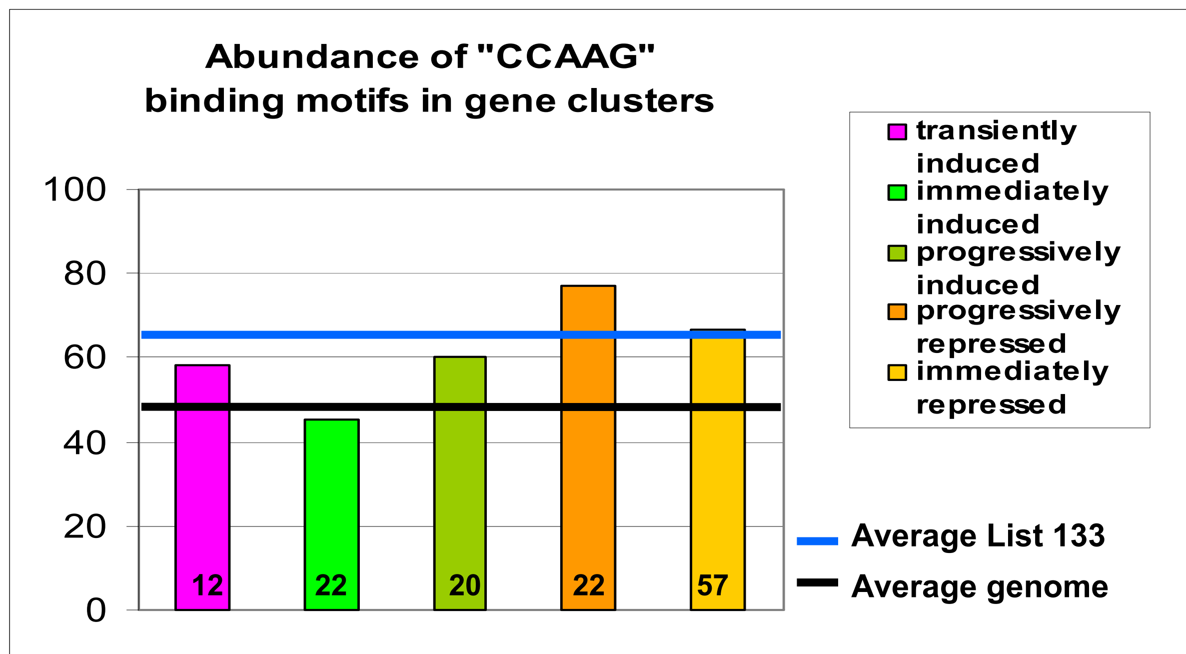


Figure 29: *In this diagram the percentage of promoters with the short CCAAG motif is shown for each of the five different clusters within the 133 selected genes. The number at the bottom of each column indicates how many genes belong to this cluster. The horizontal line indicates the global presence within all selected genes (blue) and within the complete genome of C. albicans (black).*

The distribution of the short putative binding motif "CCAAG" within the five clusters of differentially regulated genes was relatively homogenous, considering that the gene numbers in some of the clusters were relatively small. However, there was a higher abundance in the clusters of repressed genes than within the induced genes (69.6 % compared to 52.3 %); the highest abundance was observed within the promoters of the group of progressively repressed genes (77.8 %; group of *PHR2*), the lowest in the group of immediately induced genes (45.5 %) (Figure 29), supporting the idea of Rim101p acting primarily as a direct transcriptional repressor.

3.1.6 Confirmation of Microarray Results

Based on the analysis of the promoters and the coding sequences, we selected 20 genes to confirm the quality and the relevance of our microarray data by real-time qPCR (Table 8). This choice was made with a focus on one or several of the following criteria. Genes with:

- unknown function
- promoters that possess putative Rim101p binding sites
- predicted transmembrane domains

- predicted signal peptides or anchors

Note that their transcriptional patterns played only a minor role in the selection of these genes, hence the majority are members of the clusters of repressed genes (16 repressed and only 4 induced genes were selected). Eight genes from each group of repressed genes (immediately and progressively repressed) were among the selected genes, along with two progressively induced, one immediately induced and one transiently induced gene.

Genes selected for confirmation of the microarray TC results

Name	Gene cluster	CCAAG	TM domains	Sign. Pept.	ORF19	Array name	Predicted function
EFG1	immediately repressed	4	0		orf19.8243	CA2787	Enhanced filamentous growth factor
IPF1372	immediately repressed	1	5		orf19.6440	CA5100	Unknown function
IPF2280	immediately repressed	6	0		orf19.6658	CA4264	Unknown function
IPF6156	immediately repressed	1	1	p 0.6, a 0.3	orf19.1034	CA1625	Sim. to C.elegans LIM homeobox protein
IPF8762	immediately repressed	0	0		orf19.822	CA4220	Unknown function
PGA4	immediately repressed	4	0	p 0.6	orf19.4035	CA4800	GPI-anch. protein related to Phr1/2/3
QDR1	immediately repressed	4	10	a 0.7	orf19.508	CA4501	Putative antibiotic resistance proteins
WSC4	immediately repressed	1	1		orf19.7251	CA5369	Cell wall integrity by homology
PHR2	progressively repressed	4	1	p 1.0	orf19.13500	CA3867	GPI-anch. pH responsive glyc. transf.
ALS1	progressively repressed	3	0	p 1.0	orf19.5741	CA0316	Agglutinin-like protein
ALS4	progressively repressed	1	0	p 1.0	orf19.4556	CA1528	Agglutinin-like protein
CPA1	progressively repressed	0	0	p 0.9	orf19.12100	CA0874	Arg.-specific carb.phosphate synthase
IPF16514	progressively repressed	1	0		orf19.921	CA1388	Unknown function
KRE6	progressively repressed	3	1		orf19.7363	CA5661	Glucan synthase subunit
PGA52	progressively repressed	1	0	p 1.0	orf19.9467	CA0188	GPI-anchored protein
PHO87	progressively repressed	2	10		orf19.2454	CA0548	Member of the phosph. permease family
CHO2	immediately induced	0	10		orf19.169	CA1414	Phosphatidyleth.amine N-methyltransf.
IPF4580	progressively induced	1	8		orf19.6522	CA4955	Putative allantoate permease
PHR1	progressively induced	2	0	p 1.0	orf19.3829	CA4857	GPI-anch. pH responsive glyc. transf.
HGH1	transiently induced	1	0		orf19.4587	CA5149	Sim. to human HMG1/2 proteins

Table 8: *The twenty genes that were selected for confirmation of microarray results. They were chosen due to their sequence properties within the coding sequence (transmembrane domains (=TM), signal peptides) and the promoter (presence of CCAAG Rim101p binding motif). A preference was given to genes with no annotated function that carry sequence motives that indicate a possible function at the cellular surface.*

3.1.6.1 Quantitative PCR on original cDNA

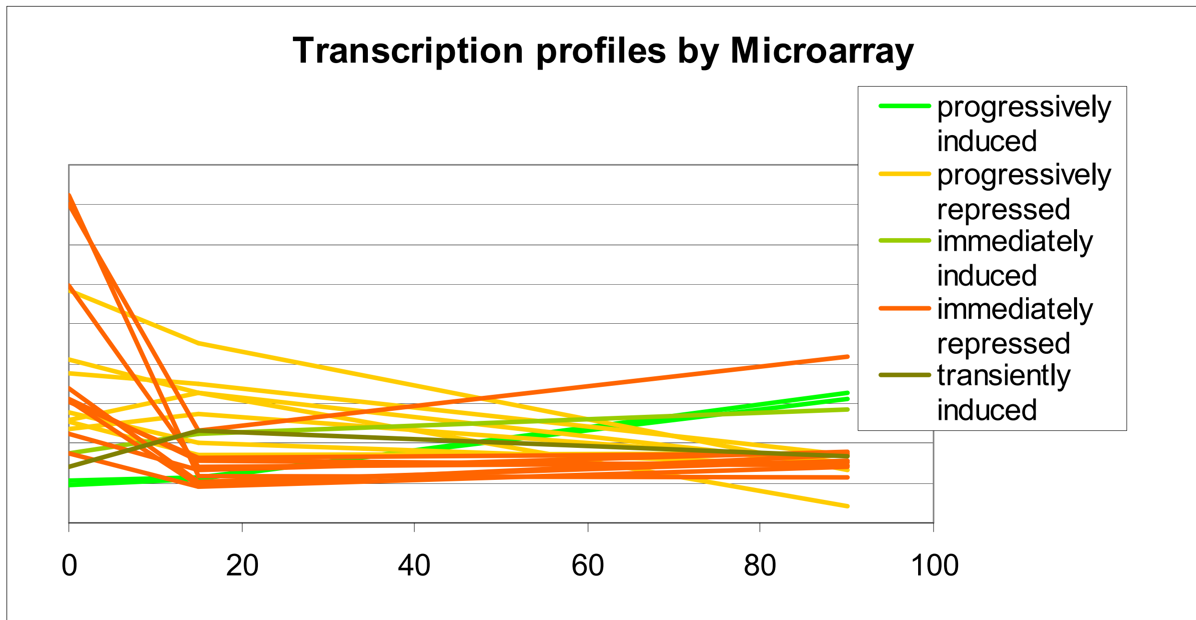
Our first aim was to test the reliability of our microarray data. This was done by performing independent reverse transcriptions on the original RNA samples used for the array. We decided to consider only the time points 0, 15 and 90 min, since they were the relevant ones for the selection procedure of the candidate genes. Their microarray transcription profiles can be seen in Figure 30a.

Figure 30b shows the corresponding transcription profiles as detected by realtime qPCR. The qPCR raw data of the transcriptional levels differed very much among the 20 genes. To permit a suitable graphical representation of all genes on the same scale the values for the distinct time points for each gene were normalized by dividing raw values by the mean of expression values at all time points. Since the reference in our microarray time course was a cDNA pool derived from a complete time course experiment, the normalization procedure we applied for the qPCR data corresponds to the signal ratio “Sample/Reference” that was determined for each time point on several microarray slides and permitted a direct comparison between the two data sets.

Genes that belonged to the same cluster in the microarray analysis clearly also form a homogenous group with a similar profile in the qPCR experiment. However, a notable difference is that most of the transcriptional profiles are more “pronounced”, so that the clusters differ more clearly from each other and show a stronger induction or repression than in the microarray data. This might indicate that real-time PCR is more sensitive for quantitative transcriptional changes than the microarrays. Furthermore, most genes of the cluster designated as “progressively repressed” in the microarray experiment appear to be slightly induced at the 15’ time point before getting clearly downregulated at 90’.

However, taken together most of the qualitative changes in gene expression detected by microarray analysis are clearly confirmed by the qPCR results, but slight quantitative differences exist.

a)



b)

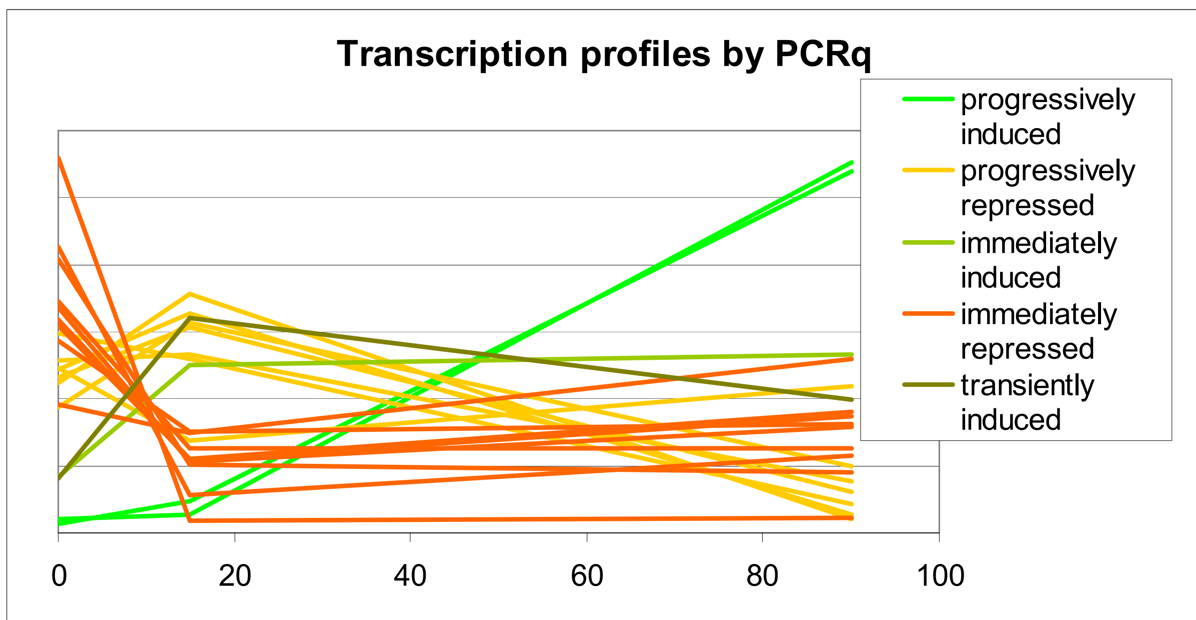


Figure 30: *Confirmation of microarray results by quantitative PCR:* ***a)*** *the microarray results 0, 15 and 90 min after RIM101SL induction are summarized for the 20 selected genes. Genes belonging to the same cluster are represented with the same colour.* ***b)*** *Results obtained for these genes by quantitative PCR. The result for each time point has been divided through the average expression of the gene for the three time points to permit the representation of all genes on the same scale. Again, data from genes of the same microarray cluster are depicted in the same colour. Raw quantitative data tables for each gene can be found in Table 10.*

3.1.6.2 Quantitative PCR using wildtype samples from cultures grown at acidic and alkaline pH

In a second confirmation experiment we wanted to correlate our results with the well documented function of Rim101p in the pH response. We thus checked whether similar transcriptional changes identified after induction of *RIM101SL* in the microarray time course experiment would also occur when the native full-length version of Rim101p was activated by pH-dependent processing. This experiment was also performed by qPCR, but this time on cDNA samples of reverse transcribed RNA extracted from the isogenic reference strain (DAY185) after 8 hours of growth in acidic or alkaline SC medium. Figure 31 displays the transcriptional changes between acidic and alkaline pH for the twenty selected genes. Again, expression values have been divided through their mean to be better presentable on the same scale (Figure 31).

Different green shades represent genes belonging to clusters of induced genes in the microarray time course, while orange and red colours indicate repressed genes as in Figure 30. Most of the analyzed genes were regulated similarly by the pH as observed in the Rim101SLp microarray TC. Three out of the four Rim101SLp-induced genes (*PHR1*, *CHO2* and *HGH1*) were also induced at alkaline pH compared to acidic pH, and ten Rim101SLp-repressed genes showed also a clear repression at alkaline pH (right half of Figure 31, *PHO87*-*PHR2*). Finally, five genes (*PGA4*, *IPF4580*, *IPF1372*, *EFG1* and *PGA52*) did not show any obvious pH-dependent regulation, and two genes were clearly induced although they had been characterized as Rim101SLp-repressed genes (after 90' of the TC experiment). This effect was particularly obvious for *ALS1*, which was the second most pH-induced gene in this experiment.

These last results are not very surprising if one considers the large differences between the experimental conditions of both assays. The microarray time course focuses on early transcriptional events immediately after Rim101SLp activation, while the above described experiment addresses the pH-dependent transcription of the same genes after adaptation for several hours to the ambient pH. Consequently Rim101p has been active under alkaline conditions for much longer than in the microarray TC, and gene transcription is affected by other pH-dependent regulators that might act in parallel to or together with Rim101p (Davis, Wilson et al. 2000; Davis, Bruno et al. 2002).

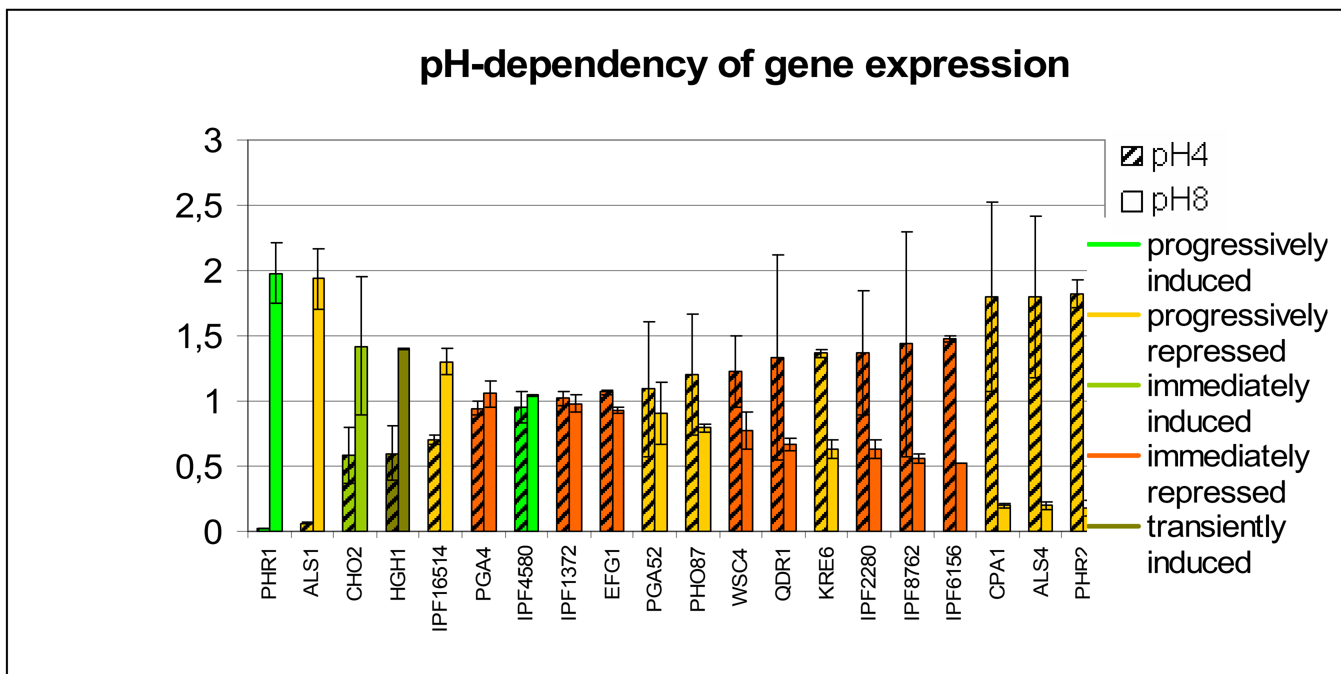


Figure 31: *The diagram shows the qPCR quantifications of the twenty selected transcripts of samples taken from SC cultures buffered at pH4 and pH8. Results for each gene are divided through their mean to permit presentation on the same scale for all the genes. Genes are sorted from the strongest alkaline induction to the strongest alkaline repression. Bars for genes that were induced in the microarray are coloured in different green tones and Rim101SLp-repressed genes are represented with yellow or orange tones according to the clustering results as in Figure 30.*

	Foldchange Repression		
Gene name	**Microarray 90'/0'**	**qPCR 90'/0'**	**qPCR pH4/pH8**
ALS4	**9.5**	**10.8**	**12.4**
IPF8762	**7.2**	**24.2**	**3.1**
PGA52	***4.4***	***7.1***	***1.2***
QDR1	**3.9**	**4.5**	**2.2**
IPF6156	**2.2**	**3.7**	**3.1**
PHR2	**2.2**	**8.9**	**10.3**
WSC4	**2.2**	**2.7**	**1.7**
PHO87	**1.9**	**3.3**	**1.9**
CPA1	**1.9**	**1.8**	**11.4**
EFG1	***1.8***	***1.7***	***1.3***
ALS1	**1.8**	**3.9**	***0.02***
IPF2280	**1.7**	**2.1**	**2.4**
KRE6	**1.5**	**1.9**	**2.4**
IPF16514	**1.4**	**1.1**	**0.5**
IPF1372	***1.3***	***1.8***	***1.1***
PGA4	***1.2***	***0.7***	***0.9***
	Foldchange Induction		
IPF4580	***3.4***	***27.0***	***1.0***
PHR1	**3.0**	**43.1**	**95.6**
CHO2	**1.6**	**3.2**	**2.6**
HGH1	**1.2**	**2.4**	**2.5**

Table 9: *Comparison of the results obtained by microarray and qPCR for the 20 selected genes.*

3.2 Characterization of pH-dependent *ALS* gene regulation

3.2.1 *ALS* Sequence Analysis

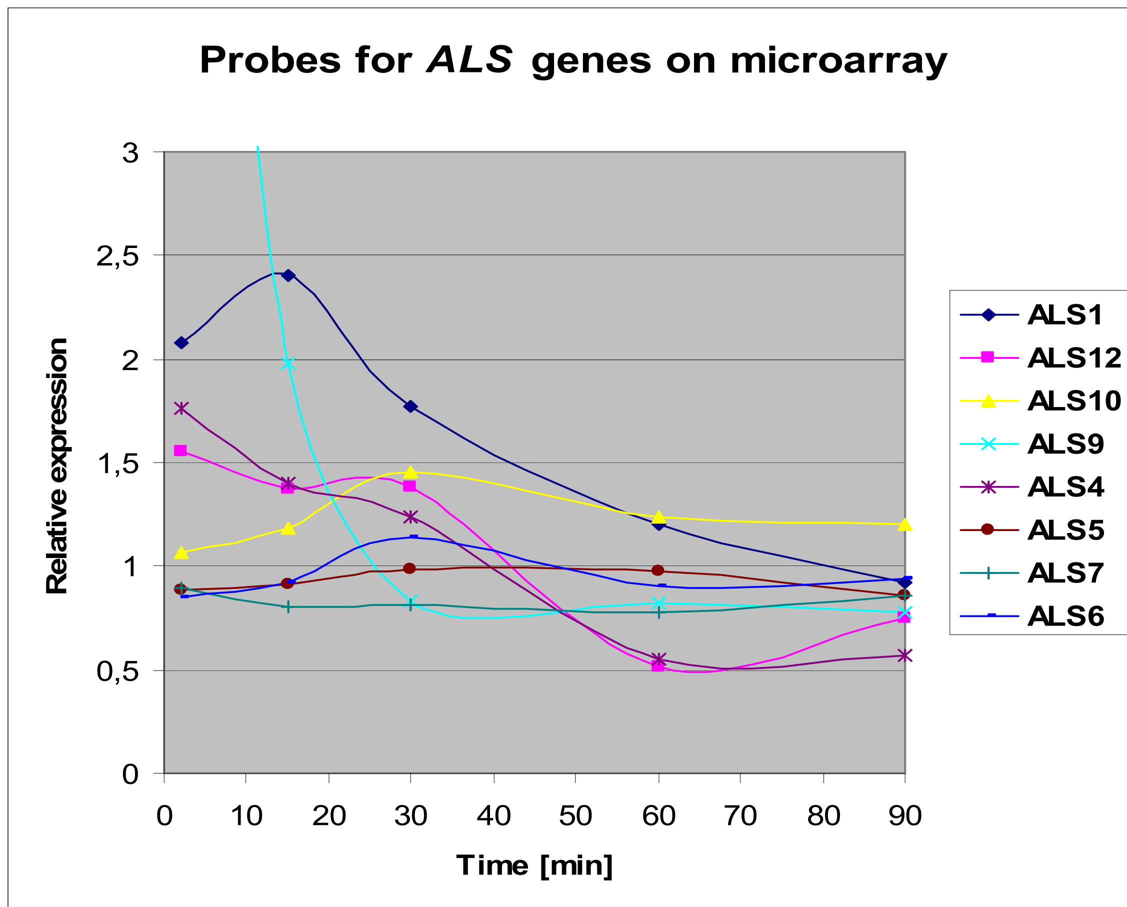
ALS genes belong to a large gene family and code for GPI-anchored cell wall proteins with important functions in biofilm formation, adhesion and endocytosis of *C. albicans* cells (See also introduction 1.3.1.3). We already analyzed the pH-dependency of *ALS1* and *ALS4* transcription, as both genes were among the twenty genes we selected for confirmation of the microarray results. The results of our microarray time course indicated an important impact of Rim101SLp on the transcription of *ALS* genes, because four of the eight detected *ALS* array probes were identified within the 133 probes (of almost 6000 detected) for significantly regulated genes: *ALS1*, *ALS4*, *ALS9* (annotated *ALS11* in Eurogenetec arrays) and *ALS12* (removed in Assembly20) were found to be progressively downregulated in the time course experiment (see Figure 32a and 6. Appendix Attachment 2).

Although the qPCR experiment with the original time course samples confirmed the microarray results for both *ALS1* and *ALS4*, the transcript quantifications at alkaline versus acidic pH clearly indicated that the transcription patterns of *ALS1* and *ALS4* were in fact opposite, and that *ALS1* was strongly induced (and not repressed) at alkaline pH, similarly to the welldescribed Rim101p target genes *PHR1* and *PHR2* that code for functional homologues.

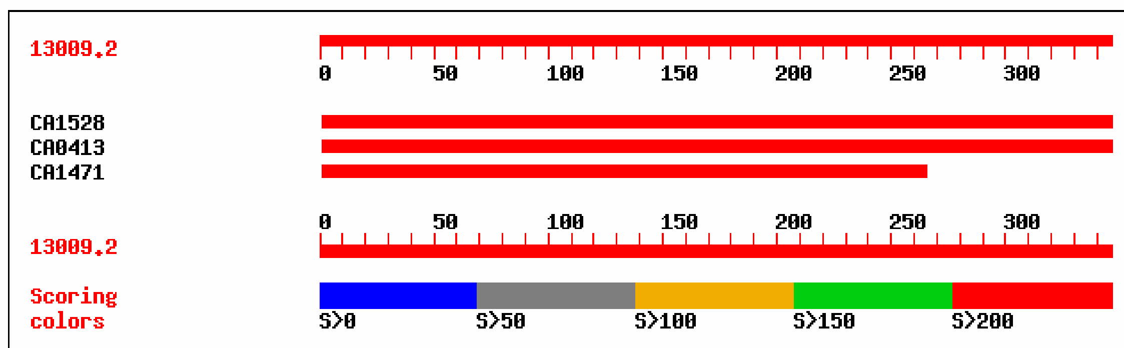
Functional homology might also be present within the *ALS* gene family, as the sequence similarities between different *ALS* genes are very strong in both promoter and coding region. The 108 bp tandem repeats are the most conserved sequence features, and it is the only part of their coding region where it is possible to find extended sequence traits that are identical between both alleles but specific for each *ALS* gene. Figure 33 shows a radial tree that resulted from an alignment of these homologous sequence regions. It clearly shows the high conservation of the tandem repeat region in *ALS1-4* and also of *ALS5*/6.

Due to the very strong sequence similarities between the different *ALS* genes some array probes do not match one specific *ALS* gene but are likely to recognize several *ALS* genes with a similar probability (Figure 34b).

a)



b)



CaALS4.3f agglutinin-like protein, 3-pri...	686	0.0
CaALS12.3f agglutinin-like protein, 3-pr...	587	e-168
CaALS2.3feoc agglutinin-like protein, 3-...	466	e-131

Figure 32: a) *The microarray time course data obtained with probes for ALS genes; four of these probes. Note that ALS1 which was found later to be induced at alkaline pH also seems to be slightly induced at time point 15 (confirmed by qPCR).*
b) *Due to high sequence similarities not all microarray probes for ALS genes recognize a single gene. The graphic shows the blast result for the ALS4 probe which indicates that the probe does not differentiate between ALS4 and ALS2.*

3.2.2 Role of Rim101p in *ALS* gene regulation

For these reasons we decided to characterize in detail the influence of pH changes and the impact of Rim101p activity on the transcriptional regulation of each of these genes using gene-specific primers and realtime qPCR (Figure 34 & Table 10). The use of *ALS*-specific primers for gene-specific real-time PCR quantification was pioneered by Green *et al.* (Green, Zhao et al. 2005). We decided to adopt their primer set to analyze the gene-specific regulation, but due to dimerization problems encountered with some of these primers and because of initial confusions in the sequence databases concerning the effective number of *ALS* genes we developed our own primers for some of the genes (See Materials and Methods 2.1.3).

Transcripts were detected for six of the eight known *ALS* genes; *ALS3* and *ALS7* were not transcribed under the experimental conditions used. At least for *ALS3* this is not surprising, as it is annotated as a hypha-specific gene, and no hyphae could be observed in our setup because of the restrictive temperature used (30 °C). Three other genes, *ALS2*, *ALS5* and *ALS6*, were poorly transcribed at both alkaline and acidic pH at levels below 1 % of actin transcription. *ALS5* and *ALS6* showed a constant pH-independent transcription. *ALS1* was confirmed as the only alkaline-induced *ALS* gene, while the transcription of *ALS2*, *ALS4* and *ALS9* was clearly repressed at pH 8.

In the second part of the experiment, we tested if Rim101p played a role in the pH-dependent regulation of these four *ALS* genes by comparing the results at acidic and alkaline pH of the reference strain DAY185 and for the *rim101* knockout strain DAY25 (Table 10). We could confirm that Rim101p was responsible for the repression of *ALS4*, as in DAY25 there was a complete derepression of its transcription. The induction of *ALS1* could also largely be attributed to Rim101p, although there was still a weak induction in a *rim101*$^{-/-}$ background. The transcription of *ALS9* was much less pH-dependent with only a 4.5-fold alkaline repression in the wild type strain compared to the almost 40-fold induction of *ALS1* and the 21-fold repression of *ALS4*. There seemed to be a partial derepression when Rim101p was not present. Finally, the alkaline repression of *ALS2* transcription seemed to be largely unaffected of the Rim101p status. However, *ALS2* transcription was very weak and the observed pH-dependent regulation was not very strong even in the wild type background under the conditions used, so it was difficult to come to a definite conclusion.

Figure 33: *Radial distance tree for the ALS gene family calculated with conserved sequence regions within the 108bp tandem repeat sequences in the middle of each ALS ORF.*

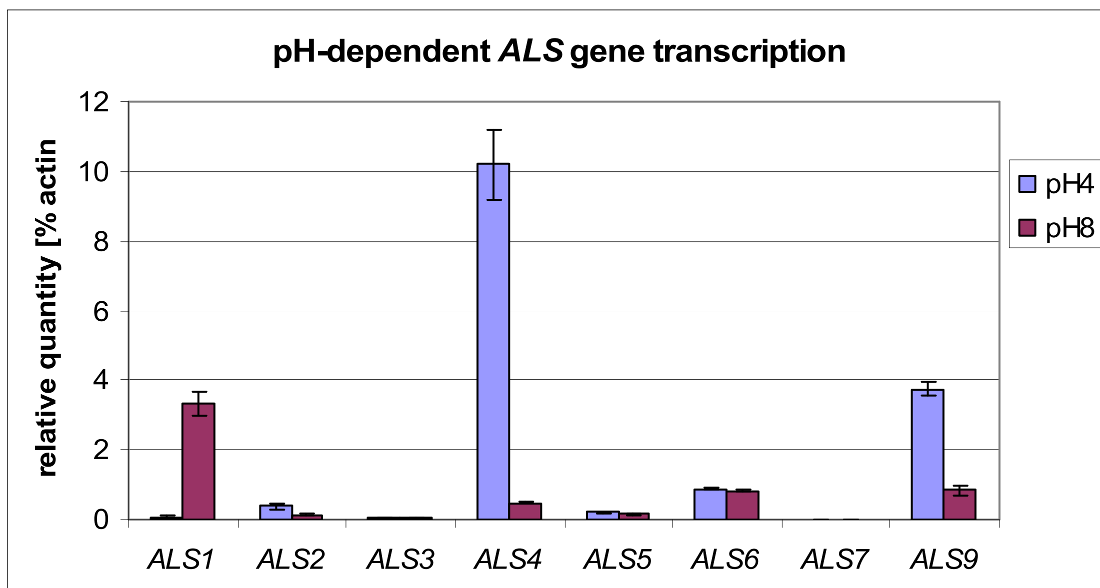


Figure 34: *The effect of ambient pH on the transcription of ALS genes.*

	Foldchange	
Induced	WT	ΔΔ*rim101*
ALS1	39.1	3.2
Repressed		
ALS2	-3.3	-2.5
ALS4	-20.9	-0.6
ALS9	-4.5	-1.9

Table 10: *Impact of Rim101p on the pH-dependent regulation of ALS gene transcription: The values in the second column represent the foldchange between pH4 and pH8 in a wild type background. The third column indicate alkaline induction or repression in a rim101 null mutant background.*

3.2.3 β-Galactosidase reporter assays

To further characterize the role of Rim101p in the regulation of *ALS1* and *ALS4* transcription, we wanted to focus on the promoter region of these *ALS* genes and we decided to construct strains that express a reporter gene under the control of these promoters. As already mentioned, these regions are also highly conserved across the gene family. Figure 35 depicts the similarity observed within the 1000 bp upstream to the start codon for each *ALS* gene. The most similar promoters are those of *ALS4* and *ALS2* (78.1 % identity), the *ALS1* promoter region strongly resembles both *ALS5* (69.6 %) and *ALS3* (59.5 % identity) promoter sequences. Most of the promoters are largely identical between the two alleles of the gene; exceptions are *ALS5* and *ALS9* promoters that have 12 and 24 mismatches in the 1000 bp upstream region respectively. As *ALS9* has also about 10 % of mismatches in the coding sequence, this might indicate that the two alleles have slightly different regulation and function. Due to the strong similarity of the different *ALS* promoters it was not easy to find promoter-specific primer couples for the promoters of *ALS1* and *ALS4*, in particular in the case of the *ALS1* promoter, where the 5' region is very similar to *ALS3* and the 3' region is almost identical to the *ALS5* promoter (Figure 36). Nevertheless we could directly amplify *ALS1* and *ALS4* promoter sequences (1kb and 2 kb upstream regions) from genomic DNA of *C. albicans* strain BWP17.

Initially we planned to use the fungal β-galacosidase gene *LAC4* from *Debaryomyces hansenii*, a close phylogenetic neighbour of *C. abicans*, as reporter gene. We chose this gene, because *D. hansenii* shares the nonconvential "CUG" codon use with *C. albicans* and because no orthologue is present in the genome of *C. albicans*. Besides, *D. hansenii* strains were available in the laboratory, so that we could directly amplify it from genomic DNA. However, despite the analysis of multiple clones no β-galactosidase activity could be detected when *DhLAC4* was integrated in strain DAY286 under the control of the strong constitutive *TEF1* promoter.

Several modified bacterial reporter genes are already available for *C. albicans* (see 1.2.2.2). A collaborating group (Alistar Brown, University of Aberdeen) could provide us with a plasmid carrying a functional version of the *Streptococcus thermophilus* LacZ gene under the control of a short constitutively active promoter fragment of *ADH1*. We thus decided to give up the *D. hansenii LAC4* project and to use instead this bacterial reporter gene

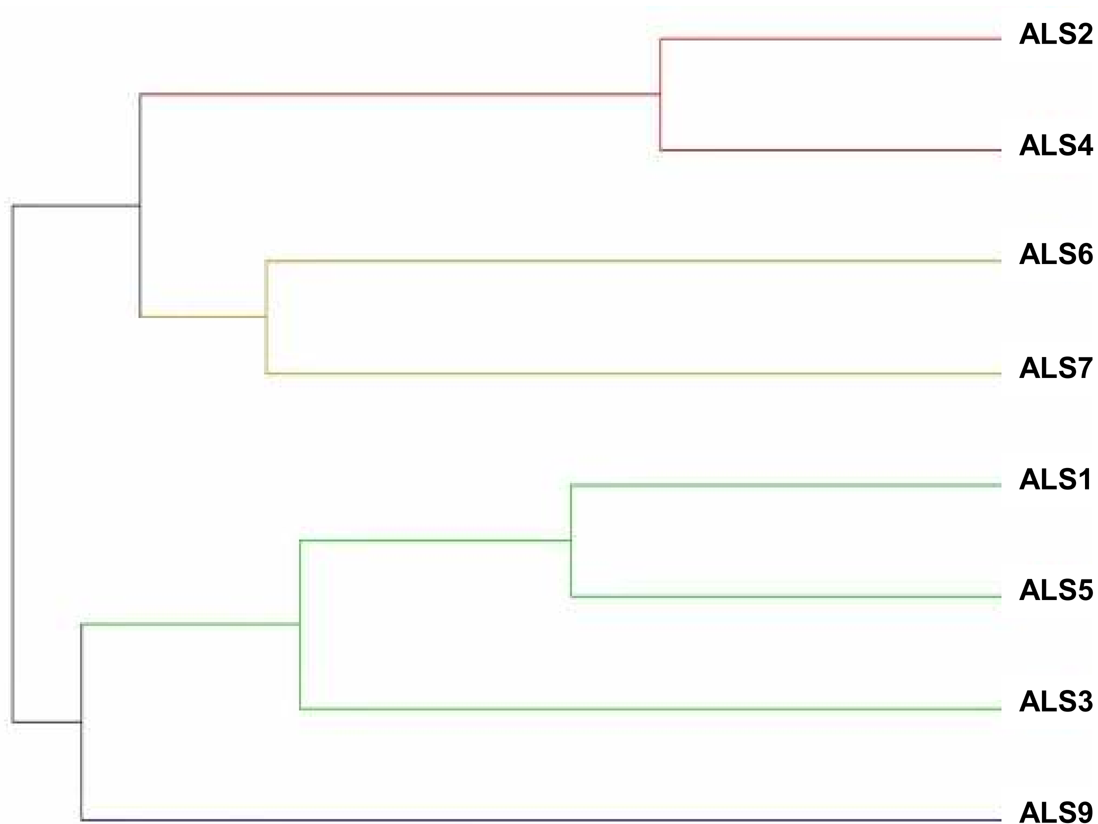


Figure 35: *Distance tree for the promoters of the ALS gene family calculated from the 1000bp upstream of the start codon. As in the distance tree for the coding regions, the promoter regions of ALS2 and ALS4 are very similar. The promoter of ALS1 is very similar to that of ALS5 and also shares important sequence similarities with the ALS3 promoter.*

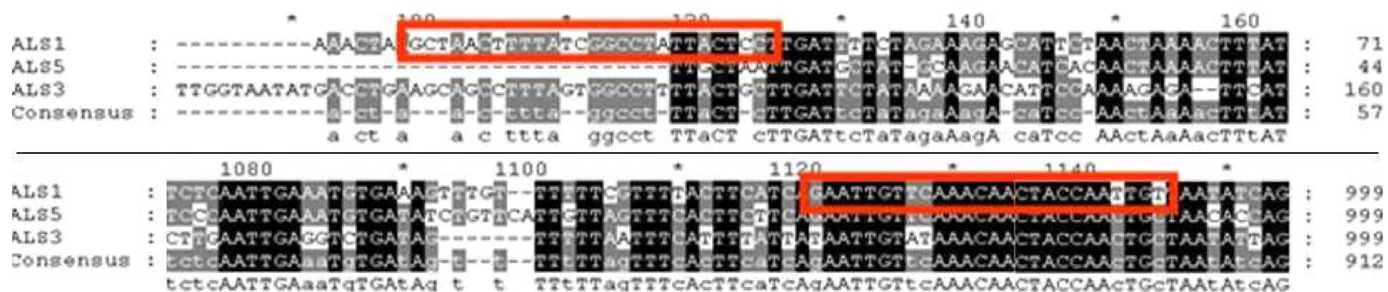
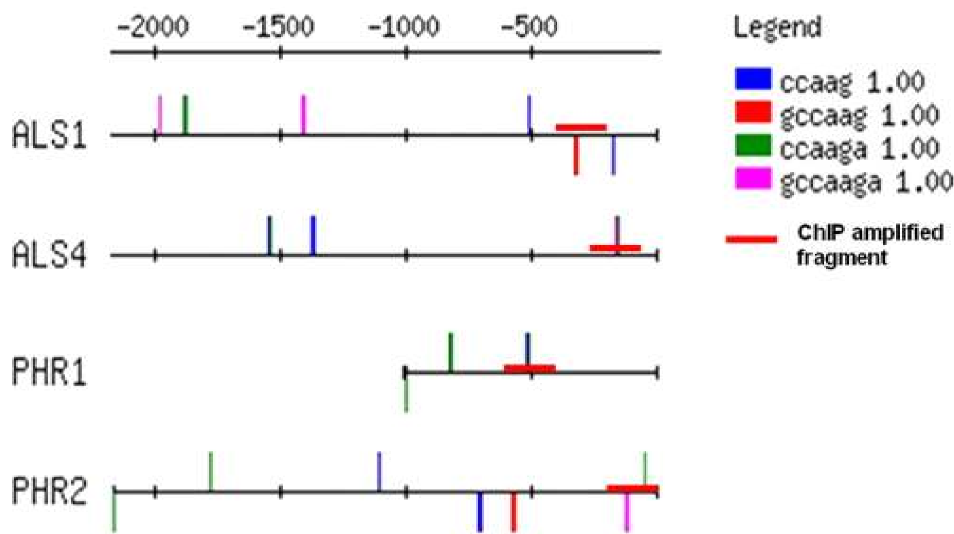


Figure 36: *Search for suitable primers to amplify the ALS1 promoter based on an alignment of the 1000bp upstream coding regions of ALS1, ALS3 and ALS5; red boxes indicate the sequences targeted by the primers used in this work.*

for the promoter analysis. We placed the LacZ gene in plasmid pDDB78 for transformation of *C. albicans* at the *HIS1* locus (See chapter 2.1.5 Figure 18).

Several possible Rim101p binding motifs are present in *ALS1* and *ALS4* promoter regions. (Figure 37). We decided to amplify a 1000 bp and an extended 2000 bp region of the promoter regions of both *ALS1* and *ALS4* from genomic DNA of strain BWP17. In addition, we constructed positive controls for both alkaline induced and repressed promoters using promoter regions of *PHR1* (1kb) and *PHR2* (2kb). LacZ under the control of the *ADH1* promoter fragment from the original plasmid was used as a control for a pH-independent promoter, and the plasmid without any promoter was used as a negative control. All these plasmids were integrated at the *HIS1*-locus of *C. albicans* strain DAY286 (wildtype *RIM101*) and in the *rim101* disrupted strain DAY5.

a)



b)

SeqID	Orientation	Pattern	Start	End	Sequence environment
ALS1	5' - 3'	GCCAAGAA	-1989	-1982	gaaaGCCAAGAAttcc
ALS1	5' - 3'	CCAAGAA	-1886	-1880	aaaaCCAAGAActtt
ALS1	5' - 3'	GCCAAGAAA	-1415	-1407	ctagGCCAAGAAAttcg
ALS1	5' - 3'	CCAAG	-512	-508	caacCCAAGttgt
ALS1	3' - 5'	GCCAAG	-325	-320	agcgGCCAAGggtt
ALS1	3' - 5'	CCAAG	-176	-172	agttCCAAGtcaa
ALS4	3' - 5'	GCCARG	-1678	-1673	taaaGCCAGGtcct
ALS4	5' - 3'	CCAAGAA	-1546	-1540	gattCCAAGAAggtc
ALS4	5' - 3'	CCAAG	-1374	-1370	cccaCCAAGgggt
ALS4	5' - 3'	GCCAAGAA	-164	-157	aataGCCAAGAAgaag

Figure 37: a) *Location of possible Rim101p binding sites in upstream regions of ALS1, ALS4, PHR1 and PHR2 promoter identified with the RSAT webtools. Both possible orientations are considered and antisense motifs are marked under the sequence. Note that the promoter region of PHR1 is shorter due to an adjacent ORF near -1kb. Horizontal red lines mark the position of the primers used in the ChIP experiment.* ***b)*** *Details about the positions of possible Rim101p binding sites on the promoters of ALS1 and ALS4 and their sequence environment can be seen in the table below. Note that the more complete motifs on the ALS1 promoter are only located in distant sequence regions.*

Not all constructs gave transformants with detectable β-galactosidase activity. While this was expected for the negative control, we were surprised not to find any activity in clones with the p*ALS1*(1kb)-LacZ fusion despite the test of 40 PCR-confirmed transformants. On the contrary, clones obtained after transformation with the p*ALS1*(2kb)-, the p*ALS4*(1kb)- and the p*PHR1*(1kb)-LacZ fusion showed a weak activity comparable to that obtained from the p*ADH1*-LacZ control. Finally, a stronger enzymatic activity could be detected in both the p*ALS4*(2kb)- and the p*PHR1*(2kb)-LacZ fusion. Initially important variations in LacZ activity were observed among clones of the same transformation, indicating the possibility of tandem insertion events. We screened the transformants by PCR with plasmid-specific primers to detect multiple insertions and excluded such clones. The remaining transformants possessed very similar activity levels.

Strain	Day286 (wild-type)		
Plasmid	Clones tested	LacZ activity	foldx between pH4 and pH8
ALS1-LacZ **(1 kb)**	40	0	-
ALS1-LacZ **(2 kb)**	10	10 - 60	4.1 x induced
ALS4-LacZ **(1 kb)**	4	10 - 100	6.8 x repressed
ALS4-LacZ **(2 kb)**	10	100 - 1000	6.8 x repressed
PHR1-LacZ **(1 kb)**	4	10 - 100	3.9 x repressed
PHR2-LacZ **(2 kb)**	10	500 - 1000	1.5 x repressed
ADH1-LacZ **(0.3 kb)**	4	50 - 100	1.2 x induced
LacZ **(0 kb)**	4	0	-

Table 11: *β-Galactosidase activity quantification for transformants of reference strain DAY286 expressing the different LacZ fusion constructs. Activities were generally weak with maximal values of 1000 U/hour measured at acidic pH for clones carrying 2kb upstream regions of alkaline-repressed genes PHR2 and ALS4. No activity could be measured for clones carrying the 1kb fusion of the ALS1 promoter despite testing 40 clones.*

As expected, activity under the control of the *ALS4*-promoter was repressed at alkaline pH under control of both 1 and 2 kb promoter fragments, while it was induced in the p*ALS1*(2kb)-LacZ fusion strains (Table 11). However, no clear pH-dependent regulation was observed in the transformants containing p*PHR1*- or p*PHR2*-LacZ, in spite of several independent experiments, and surprisingly we observed similar pH-dependent LacZ activity for all constructs in the corresponding *rim101* disrupted strain.

These last results were in stark contrast to all previous observations and publications about *PHR* gene regulation of other groups, so that there remains a profound doubt about the relevance of these data for the *in vivo* regulation of the examined promoters.

Recently another group (Baek, Martin et al. 2006) demonstrated the Rim101p-dependent regulation of *PHR2* with a very similar setup. They also used *S. thermophilus* LacZ as a reporter gene, and they basically used the same plasmid for the transformation of their *C. albicans* strains. With a 1 kb promoter fragment they measured similar β-Galactosidase activities as us at acidic pH (about 1200 compared to our 1000 Units/hour), but unlike us they observed a more than 20-fold Rim101p-dependent decrease in activity at alkaline pH. Thus, it should be generally possible to use the *PHR2* promoter as a control for an alkaline-repressed gene, and the explanation for the discrepancies between our and their results must lie in the details of plasmid construction, strains or experimental conditions used. The strains that were transformed are also quite similar, although not identical; while we used the prototrophic strains DAY286 and the *rim101* null mutant DAY5, they worked in a Ura3$^-$/Arg4$^-$ background (strains DAY1 and DAY432). We were wondering if the ectopic integration in the His1-locus could have perturbed the expression in our case, but Baek *et al.* transformed in the same genomic location, so that this possibility is unlikely. Moreover, in both cases media were buffered with 150 mM HEPES at pH 4 or pH 8, Baek *et al.* used M199 medium, while we used both SC and rich YPD medium (with similar results). Taken together, as the more extended *PHR2* promoter region we used in our study includes all three regulatory motifs that were identified in the study of Baek *et al.* (Baek, Martin et al. 2006)(see also Figure 37), we should observe similar regulatory effects.

However, the devil might be in the detail: we integrated the promoter regions as *NotI* restriction fragments after subcloning in another vector (pGemTeasy, see Materials and Methods). To permit this we cloned a *NotI* restriction site in front of the LacZ coding region (Figure 38; details in Figure 18 (Materials and Methods 2.1.5 Plasmid Construction)). Thus, all our promoters except the *ADH1* promoter are linked with this *NotI* site to the coding region of LacZ, whereas in the study of Baek *et al.* the *PHR2* promoter was directly linked to the start codon of LacZ. Consequently, it might be possible that this short sequence (including *NotI* site and few adjacent bases imported from the subcloning vector together with the *NotI*-excised promoter) between regulatory region and LacZ gene perturbs somehow the regulatory activity. If this is true, all plasmids would have to be reconstructed since the very beginning, and afterwards integrated into the different *C. albicans* strains to obtain quantitative data that really reflect the *in vivo* regulation under control of the different promoters. And only then could we start a more detailed analysis of the promoter sequences, for example by verifying the effect of deletions or mutations of Rim101p binding motifs of the promoter. After evaluating the time that might be necessary to arrive at the same point with a modified setup,

we decided to remain with these rather ambiguous results and to characterize the role of Rim101p in gene regulation using a completely different approach.

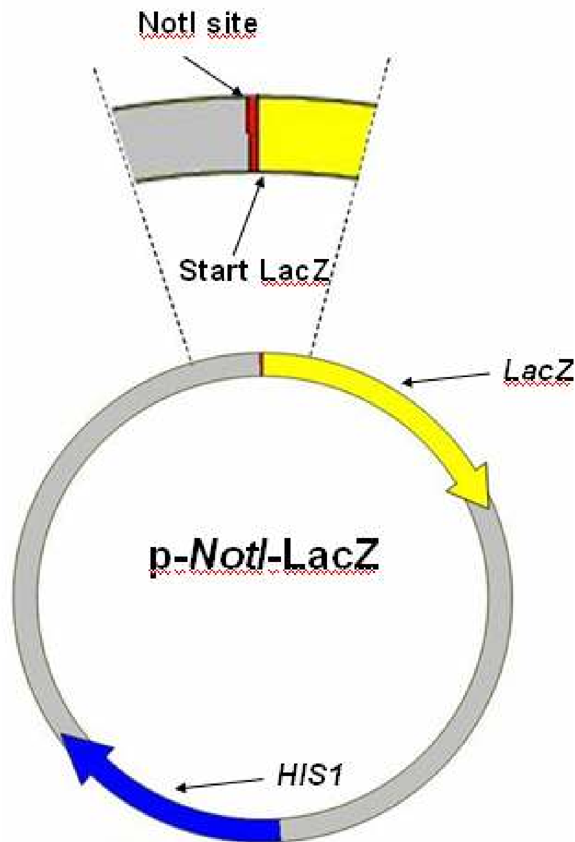


Figure 38: *Simplified presentation of the basic plasmid p-NotI-LacZ that was used to construct the promoter fusions with the S. thermophilus LacZ reporter gene for transformation of C. albicans His$^-$ strains. All promoter regions were amplified from genomic DNA of strain BWP17 and subcloned in pGemTeasy for sequence verification. To simplify the subsequent transfer of the DNA fragments into the transformation plasmid, we placed right a NotI restriction site in front of the start codon of LacZ that permits the easy integration of the different promoter fragments with adjacent NotI restriction sites. This linker region between the promoter and the LacZ gene might have influenced the regulation and perturbed the results for ALS1/4 and PHR1/2 promoters.*

3.3 Chromatin Immunoprecipitation to identify direct Rim101p targets

3.3.1 Experimental Setup ChIP

Our experimental setup in the microarray experiment was focused on the description of early transcriptional events under control of Rim101p activity. Although we found a high abundance of possible Rim101p binding sites in the promoter regions of the 132 identified genes, we did not provide any direct evidence that these motives are indeed recognized by Rim101p. To address this question we decided to try an *in vivo* chromatin immunoprecipitation technique using a strain that expresses a V5-tagged version of Rim101p (kindly provided by the group of Dana Davis). We expected then to monitor by real-time PCR quantification of promoter fragments the pH-dependent binding event of Rim101p to its target promoters (Figure 39); in addition, we would be in a good position for the identification of new direct Rim101p target genes.

Our assumption was that the relative proportion of Rim101p which is bound to the promoter of a directly regulated gene should be larger at alkaline than at acidic pH. In contrast, the interaction of Rim101p with the promoter of a non-regulated gene might be very limited and should be largely unaffected by changes in the ambient pH. Thus, in the ChIP experiment the relative quantity of immunoprecipitated promoter DNA should provide information about the presence or absence of Rim101p on the analyzed promoter sites and could help to show if a gene is directly Rim101p-regulated.

In preparative experiments, we first confirmed that the V5-tagged version of Rim101p is processed in a pH-dependent manner (Western Blot Figure 40a; (Li, Martin et al. 2004)). In addition, an independent study (Baek, Martin et al. 2006) showed by EMSA that V5-Rim101p binds *in vitro* to Rim101p binding sites on the *PHR2* promoter. These results suggest that the V5-tag does not strongly perturb the activation of Rim101p and its interaction with promoter DNA.

To prepare the chromatin for the immunoprecipitation experiment, we needed to shear the complexed genomic DNA into fragments of an average size of 500-1000 bp. We confirmed the sonication step by analyzing a decrosslinked fraction of the chromatin preparation on an agarose gel (Figure 40b).

ChIP with V5-Rim101p at acidic and alkaline pH

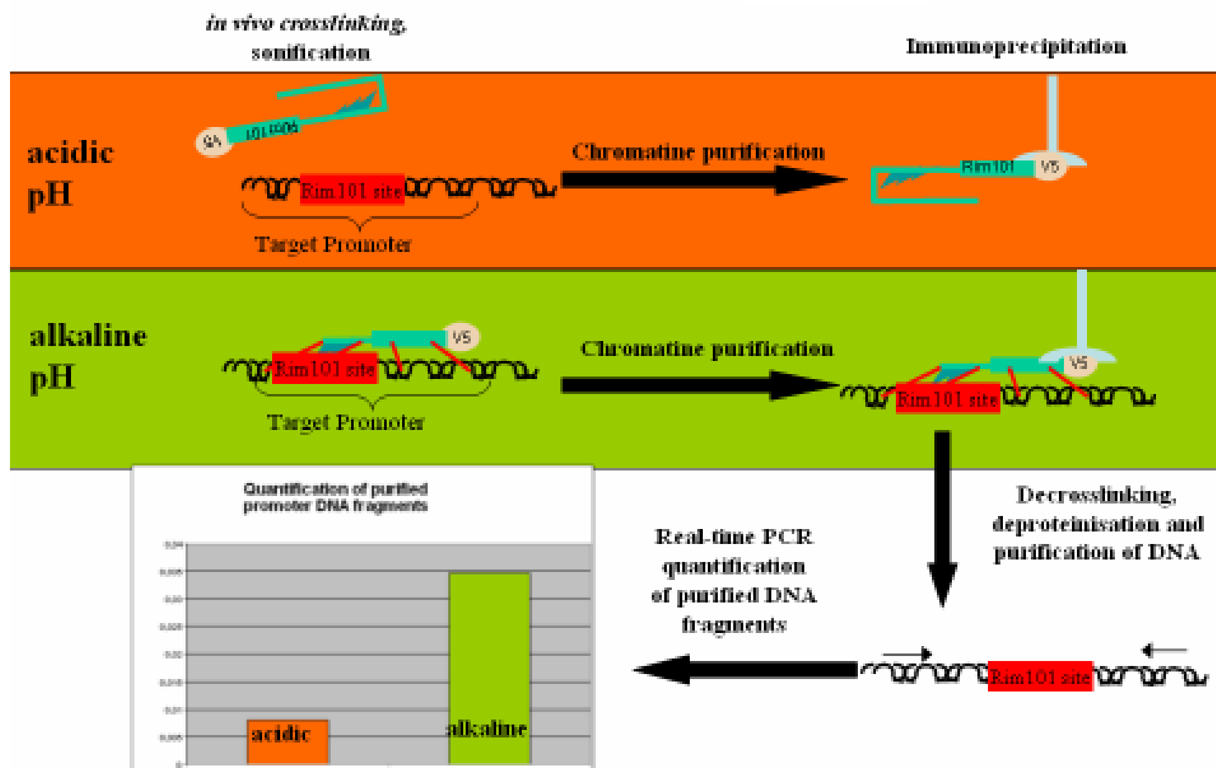


Figure 39: *Schematic description of the ChIP experiment: In a first step to preserve in vivo protein-DNA (and protein-protein) interactions covalent links are established by formaldehyde crosslinking. This should fix the alkaline active form of V5-Rim101p to its target promoters, while at acidic pH the inactive form of V5-Rim101p might not be proximal to these promoters and thus not crosslinked. During chromatin purification, the genomic DNA-protein complexes are fragmented into small 500-1000 bp units by sonication. In the following IP reaction V5-Rim101p and bound DNA fragments are isolated; while in the alkaline pH sample the crosslinked target promoters are expected to be precipitated together with V5-Rim101p, this is not expected at acidic pH. Then the crosslinking reaction is reversed and the precipitated promoter DNA is quantified by real-time qPCR to estimate indirectly the increased alkaline binding of Rim101p to its possible target sites.*

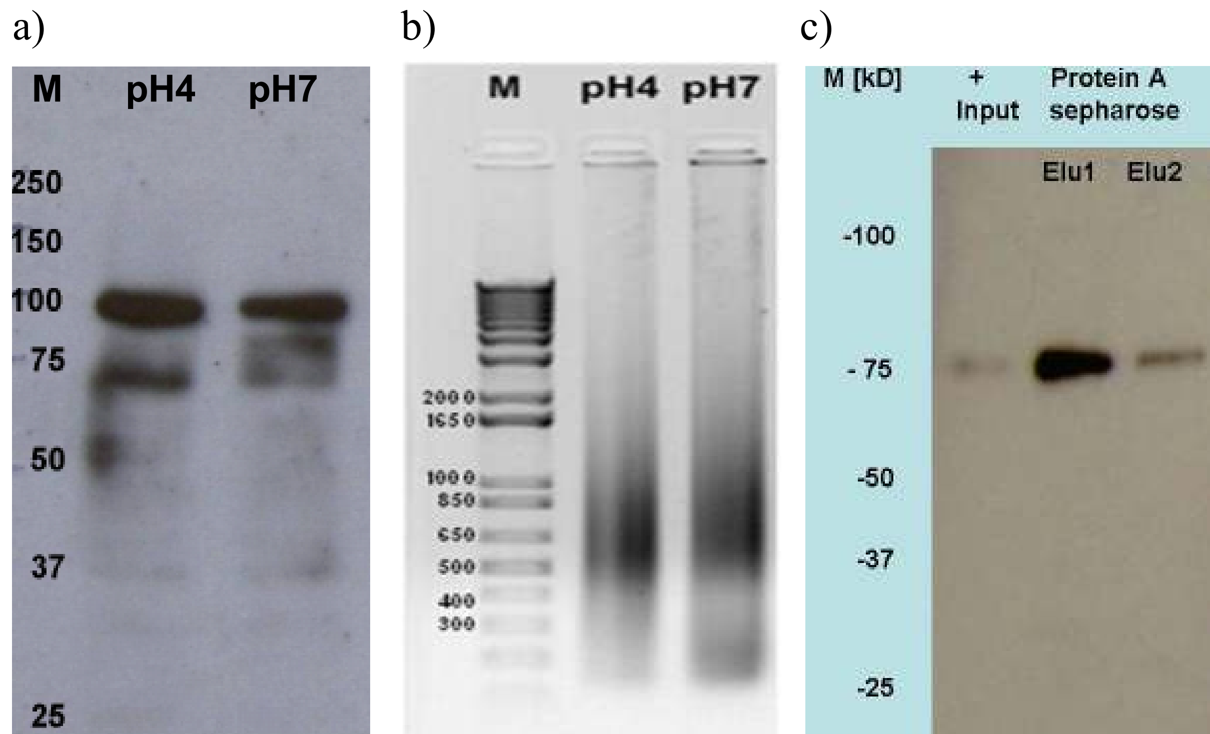


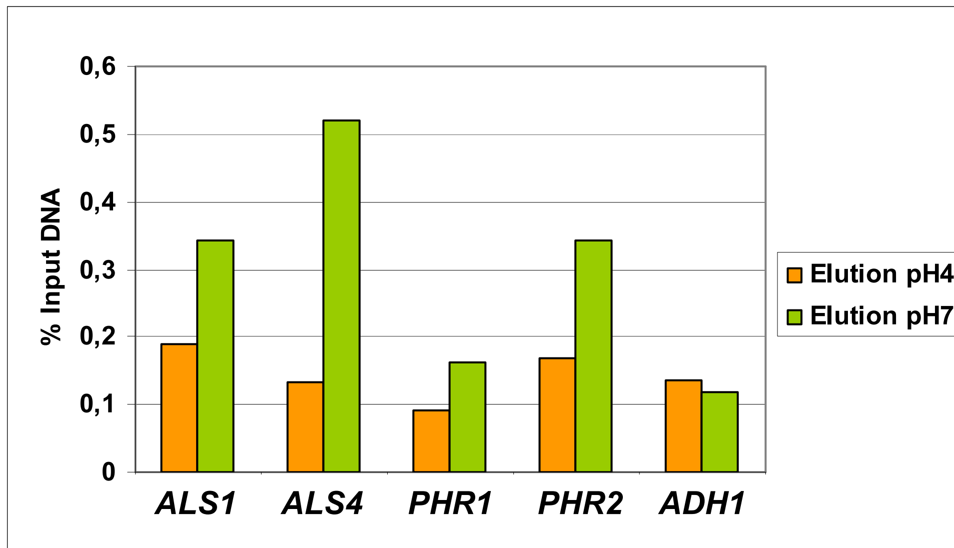
Figure 40: *Preliminary experiments for ChIP project:* ***a)*** *A Western Blot with crude protein extracts from samples taken at pH4 and pH7. The assumed active form of Rim101p migrates at approximately 75 kD and is present only in the pH7 sample.* ***b)*** *Agarose gel with DNA fragments purified from crosslinked chromatin after sonication. Similar DNA quantities were obtained from pH4 and pH7 samples, the average size being 500-600 bp, as expected.* ***c)*** *Western Blot where two elutions of an IP for the V5-epitope are loaded together with a pH7 chromatin extract prior to the IP. One clear band at about 75 kD appeared in all samples, which corresponds to the expected size of processed Rim101p.*

Following these control experiments, many different ChIP experiments were performed in order to optimize the protocol for our purpose. In particular we modified the different steps of the immunoprecipitation and DNA purification to optimize the promoter DNA yield. Various amounts of chromatin were used in the IP and the stringency of the different washing steps was modulated (see Materials and Methods 2.3) as well as the elution. Direct immunoprecipitation with V5-agarose beads was tested as well as an indirect protocol with the addition of V5-antibody prior to the precipitation using a Protein A-sepharose matrix (Figure 40c).

3.3.2 Results and Discussion ChIP

qPCR results for ChIP with V5-tagged Rim101p

a)



b)

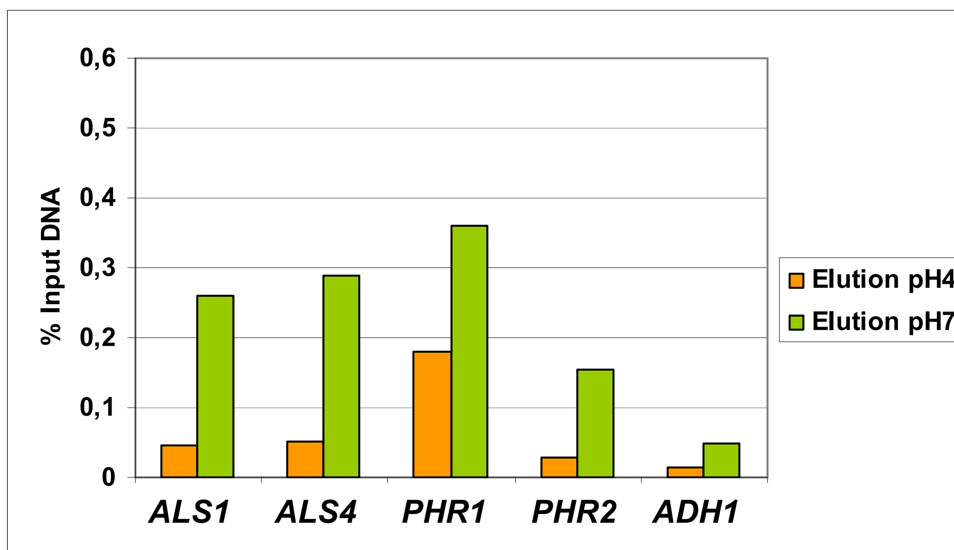


Figure 41: a&b) *Results of two independent ChIP experiments: samples were taken at OD1 from cultures in SC buffered with 150mM HEPES at pH4 or pH7, chromatin preparation was identical as described in Materials and Methods, and in both cases Anti-V5-agarose was used for ChIP, but in experiment* ***b)*** *it was blocked with fish DNA and washing was more stringent than in experiment* ***a)*** *to reduce background binding.*

Only some of the chromatin IP experiments gave similar results to what we expected (Figure 41). We almost always observed a relative increase of signals for all promoters at

alkaline compared to acidic pH. However, this increase was occasionally also observed for the control promoters of *ADH1* and *ACT1* in sequence regions that do not carry any clear Rim101p binding motifs. Moreover, the relative quantity of promoters detected at acidic pH was also larger than expected for conditions where Rim101p was not active.

This might be simply due to a possibly high level of nonspecific background binding of genomic DNA to the IP matrix. However, blocking the matrix with sonicated salmon sperm DNA prior to the IP reaction did not result in a clear decrease of background binding. In a control experiment with reference strain DAY286 which expresses the non-tagged wild type form of Rim101p, we recovered much less DNA material, which indicated that at least partially this "background" DNA is linked to Rim101p and cannot be explained solely by unspecific binding to the agarose matrix. It has been shown that the full length form of Rim101p has some affinity for the Rim101p binding site, and it has been suggested that the main reason why the truncated form of Rim101p is transcriptionally active might be that the truncated form is located preferentially in the nucleus, while full-length Rim101p is found primarily in the cytosol in analogy to PacC in *A. nidulans* (Mingot, Espeso et al. 2001).

It cannot be completely excluded that the detection of a relatively high level of Rim101p-controlled promoters at acidic pH is due to the fraction of full-length Rim101p that has accessed the nucleus or that during the chromatin purification process cytosolic Rim101p might get in contact and bind to promoter sites which are usually out of its reach. Rothfels *et al.* report that they abandoned their ChIP experiments, because they observed ScRim101p-recognized promoters in control experiments with non-crosslinked chromatin preparations, and they concluded that Rim101p binding events might have occured after the lysis of cells during the chromatin extraction which perturb the results (Rothfels, Tanny et al. 2005). The ChIP results published by Lamb *et al.* indicate that some promoters of ScRim101p-repressed genes are recognized by ScRim101p. However, they also do not see difference between the association of processed and full-length HA-tagged Rim101p (*rim13*$^{-/-}$strain) to the analyzed promoters, which they interpret as a proof for *in vivo* binding of fulllength Rim101p to these target promoters (Lamb and Mitchell 2003).

It is important to mention that the yield of recovered promoter DNA was always close to the detection limit of the real-time PCR system used (27-30 cycles were necessary to reach the fluorescence signal threshold), where the quantifications are less reliable. Consequently, the results we obtained proved to be quite difficult to reproduce. There are multiple reasons that might cause a low DNA yield. For example, little is known about the abundance of Rim101p in the nucleus. Transcription factors are generally not very abundant, and, although

RIM101 mRNA can be readily detected by different methods in particular at alkaline pH, little is known about its stability and about that of the different forms of Rim101p. Moreover, the V5-tag might affect the stability. However, it seems not likely that the instability of the protein is responsible for the poor yields, as V5-Rim101p could be readily detected by Western Blot at both acidic and alkaline pH, not only in the crude extracts, but also after the immunoprecipitation (Figure 40c). This result also demonstrated that an immunoprecipitation of V5-tagged Rim101p is technically possible, although the V5-epitope is not even complete in this construction. However, we failed to demonstrate that the ChIP with crosslinked Rim101p functions equally well, because the step (overnight 65° C) which was required to reverse the crosslink between Rim101p and its DNA targets resulted in the degradation of Rim101p. Nevertheless, it seems unlikely that crosslinked complexed DNA (or protein) partners could perturb the IP reaction, in particular because the V5-Tag apparently does not inhibit the processing or DNA binding activity of Rim101p.

Another possible explanation for the low DNA yield might be that the assumed Rim101p binding motifs are not recognized *in vivo* by Rim101p (or at least not by the V5-tagged form used in the experiment). Although possible Rim101p binding motifs are present on all analyzed promoters except the control sequences, it is not guaranteed that these sequences are indeed recognized and bound by Rim101p. For instance it has been shown for the promoters of *PHR1* and *PHR2* that not all predicted binding motifs are recognized with the same efficiency *in vitro* (Ramon and Fonzi 2003; Baek, Martin et al. 2006). The short promoter regions which were chosen for PCR verification all comprise at least one predicted Rim101p binding motive, but other possible binding motifs are present on these promoters that might be more efficiently recognized. In particular the results obtained with the *ALS1*-LacZ fusion construct, although ambiguous, suggested that the regulatory region of the *ALS1* gene is quite large and possesses several potential Rim101p binding sites. However, the *PHR2* promoter fragment amplified in this ChIP experiment comprises both the -124 and the -51 Rim101p binding site that were demonstrated to be bound by Rim101p and to be responsible for the regulation of this gene by (Baek, Martin et al. 2006). Moreover, chromatin complexes were sheared into random fragments of an average size of approximately 500 bp, but the size of the qPCR amplicons was chosen to be only about 100 bp. Consequently, the "detection range" covered with these primers is much larger and includes sometimes other motifs adjacent to the PCR-amplified fragment (even if the likelihood that the template sequence and the Rim101p binding site are present on the same DNA fragment decreases with their respective distance).

A possible concern of the experimental setup is the occurrence of unwanted side effects linked to the different pH conditions during the sample preparation. Formaldehyde is added for the crosslinking reaction in form of an aqueous 37 % solution called formalin, which contains short polymers of formaldehyde. As the depolymerisation of formaldehyde is catalyzed by hydroxide ions, the pH differences between the samples might to a certain extent influence its capacity to diffuse into the cells (see Figure 42). Since the crosslinking reaction itself takes place in the cytoplasm and nucleus, thus under physiological conditions, the pH should have no direct effect.

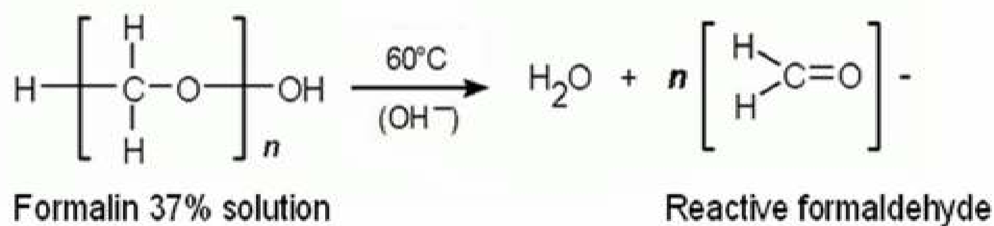


Figure 42: *Depolymerization reaction of formalin, adapted from http://publish.uwo.ca/~jkiernan/formglut.htm (Kiernan 2000)*

Another possible concern with regard to chromatin purification is the pH-dependent filamentation that might cause differences in the chromatin yield between alkaline and acidic samples because of possibly different efficiencies between hyphal and yeast cell forms in cell breakage and sonication. As we decided to grow cultures at 30 °C where no filamentation occurs due to the temperature restriction, such effects can be excluded. Besides, as the final ChIP results are expressed as a ratio output/input, the above mentioned biases are compensated and should have a minor effect on the final outcome.

3.3.3 Conclusions ChIP

Taken together, although the ChIP results indicate an alkaline enrichment of promoter sequences of the tested genes, they have to be treated with caution because of ambiguous results obtained with the control promoters, the globally weak PCR signals and the poor reproducibility of the experiments. This suggests that it is not possible to perform a trustful large scale analysis (by qPCR or "ChIP on chip") to identify direct target genes of Rim101p using under the tested conditions as originally planned. However, this might be still possible with a differently tagged strain and/or a modified IP protocol, if stronger signals and less background noise could be obtained.

Finally, it is important to mention that the experimental setup used here is based on important assumptions concerning the biological working mode of Rim101p which have not been validated yet. For example, it is not known if Rim101p binds the promoters of non-regulated genes if a suitable motif is present, and, if yes, which additional requirements are necessary to decide whether the gene is transcriptionally regulated by Rim101p or not. Consequently, only if the binding of Rim101p to a given promoter region means already that the corresponding gene is regulated by Rim101p, ChIP results (or gel shift experiments) could prove the direct regulation of a gene.

In addition, while the autoinduction of *PacC*/*RIM101* transcription at alkaline pH has been shown in *A. nidulans*, *Y. lipolytica*, *S. cerevisiae* and *C. albicans*, it has not yet been shown that in *C. albicans* Rim101p the pH-dependent processing event triggers its localization and thereby its activity. Even if this is probable with respect to the findings for PacC in *A. nidulans* (Mingot, Espeso et al. 2001), it has been proposed for *S. cerevisiae*, based on the ChIP results with HA-tagged Rim101p in a *rim13* null mutant background, that a pH-dependent, but processing-independent step might activate Rim101p (Lamb and Mitchell 2003). However, our results indicate that binding levels of Rim101p to target promoters are indeed pH-dependent, and the authors did not perform ChIP experiments under different pH conditions to prove their theory.

4 Conclusions and Perspectives

4.1 Conclusions

The main aim of the project was the identification of new Rim101p target genes at early stages of the pH response on the basis of a genome-wide experiment using microarray technology. The Rim101p-dependent pH response of *C. albicans* has received increased attention during the last few years, as it is also documented by two other publications on large-scale experiments in this field ((Bensen, Martin et al. 2004) (Lotz, Sohn et al. 2004); see introduction 1.4.4.5 "Genes under the control of Rim101p"). Although in both cases the authors emphasize the importance of the extent of Rim101p activity, of the possible direct or indirect interaction of Rim101p with other transcription factors and of parallel Rim101p-independent pathways in the regulation of many pH regulated genes, their analyses were restricted to an "on/off" scenario and gave likewise a rather static picture of the role of Rim101p in the pH response.

To our knowledge this work is the first attempt to obtain a more complete overview of the transcriptional events that take place when Rim101p becomes active. Considering the complexity of the possible regulatory events in the pH response of *C. albicans*, we opted for an experimental setup that permits to uncouple the activation of Rim101p from other pH-dependent regulatory events which might obscure the interpretation. By expressing a constitutively active version of Rim101p under the control of an inducible promoter we could simulate the transition from inactive to fully active Rim101p while excluding all other pH-dependent regulations. Instead of simply comparing the two extremes of the Rim101p status (completely inactive and fully active Rim101p), we decided to take "snapshots" at several different stages during its activation. In this way we hoped to get a new perspective on the transcriptional kinetics of each gene in function of Rim101p activity.

But did the results meet these expectations? After their analysis there are several aspects of the project that deserve reconsideration. An important one of them is the choice of the induction of *RIM101SL* under the control of the *MET3* promoter. The expression under the control of this promoter allowed to elegantly uncouple Rim101p activity from ambient pH conditions, but the transcriptional changes that were induced by the switch to methionine/cysteine lacking medium were more important than it might have been expected, and many genes of sulphur metabolism were among the strongest induced genes (Results

3.1.3, Table 6). As a consequence, the use of our control experiments to filter out "false positive" genes was extremely important to unequivocally identify genes with Rim101SLp-dependent regulation. We needed to find selection criteria that are stringent enough to exclude the maximal number of false positives on the one hand, but on the other hand not too rigorous so that only a minimum number of good candidates are lost.

We opted for very stringent control criteria so that the number of falsely identified genes would be minimal. In consequence of this decision it is expected that some of the excluded genes might be in reality Rim101p-regulated genes that are also affected by the sulphur amino acid starvation. For instance, among these rejected genes that were regulated similarly in both Rim101SLp time course and control experiment we found 22 genes that already had an annotation indicating pH- and/or Rim101p-dependent regulation (Table 12); and three of these 22 genes, *SAM2*, *CYS4*, *CYS3*, also have annotated functions linked to the sulphur amino acid metabolism. Another prominent victim of this selection process was *RIM8*, a gene involved in the activation process of Rim101p that has been shown to be itself repressed by Rim101p at alkaline pH in both *C. albicans* (Porta, Ramon et al. 1999) and *S. cerevisiae* (Lamb and Mitchell 2003). This gene was also repressed in our time course experiment as expected, but rejected due to its regulation in the control experiment.

In retrospective, the *MET3* promoter might be an unfortunate choice, because one of the cellular processes that were clearly affected by Rim101SLp activity was amino acid metabolism (See 3.1.5). This was particularly surprising, as sulphur amino acid metabolism genes were quite stringently excluded with help of the control experiments. However, with the current setup it is impossible to say whether the impact of Rim101p on the regulation of sulphur amino acid metabolism genes was not important, or whether we just could not see these Rim101p-dependent effects because such genes were excluded due to their sulphur amino acid starvation dependent regulation.

"Lost false positives"

Common induced genes				
Immediately induced genes			TC	CTRL
Orf19	CGD Annotation	Predicted function	15'/0'	15'/0'
orf19.5650	pH induced	Mating process	1.4	1.1
orf19.251	pH repressed	ThiJ/PfpI protein family	2.5	4.2
Progressively induced genes			TC	CTRL
Orf19	CGD Annotation	Predicted function	90'/0'	90'/0'
orf19.2770.1	Rim101 induced	Cytosolic superoxide dismutase	1.4	1.4
orf19.5280	Rim101 induced	Protein of unknown function	1.5	2.4
orf19.5541	pH induced	Similar to S. pombe Nrd1p	1.4	1.2
orf19.657	pH induced	S-adenosylmethionine synthetase	2.9	1.7
orf19.4536	pH induced	Sulphur amino acid biosynthesis	2.5	1.9
orf19.6402	pH induced	Sulphur amino acid biosynthesis	9.5	6.7
orf19.1770	pH repressed	Cytochrome c	1.6	1.1

Common repressed genes				
Immediately repressed genes			TC	CTRL
Orf19	CGD Annotation	Predicted function	0'/15'	0'/15'
orf19.4082	Rim101 repressed	Stress-associated protein	1.5	1.2
orf19.6937	Rim101 repressed	Putative oligopeptide transporter	1.7	1.6
orf19.7077	Rim101 repressed	Protein of unknown function	1.4	1.3
orf19.5610	pH repressed	Protein of unknown function	2.7	1.8
orf19.2762	pH repressed	Alkyl hydroperoxide reductase	4.3	1.4
orf19.3175	pH repressed	Protein of unknown function	1.5	1.2
orf19.5674	Rim101 induced	GPI protein heme-iron utilization	2.9	1.5
orf19.2098	pH induced	Protein of unknown function	2.8	2.0
orf19.3554	pH induced	Aspartate aminotransferase	1.8	1.3
orf19.6763	pH induced	Plasma membrane protein	1.6	1.3
Progressively repressed genes			TC	CTRL
Orf19	CGD Annotation	Predicted function	0'/90'	0'/90'
orf19.6229	pH repressed	Catalase	2.1	1.7
orf19.1153	pH repressed	Protein of unknown function	1.8	1.6
orf19.5000	pH repressed	Precursor of cytochrome b2	1.5	1.1

Table 12: *List of the 22 excluded genes that are currently annotated as Rim101p- or pH-regulated at the Candida Genome Data base http://www.candidagenome.org/; all of these genes showed an altered transcription in the time course experiment with Rim101SLp, but their transcription was also influenced in a similar way in the CTRL experiment. The vast majority of these genes were regulated in the TC experiment consistently with their annotation (marked in grey). Several of the induced genes have a predicted function in the metabolism of sulphur amino acids (marked in orange), which could explain their changed transcription under methionine/cysteine starvation..*

To reduce the impact of the experimental conditions, it might have been more prudent to choose another inducible promoter than the *MET3* promoter to modulate *RIM101SL* transcription. Several promoters have been successfully used to induce gene expression in *C. albicans*, but for our experiment we needed a strong and easily inducible promoter which limits the choice. For example, the *MAL2* (repressed by glucose and induced by maltose) (Zhao, Oh et al. 2005), *GAL1* (galactose-induced) (Srikantha, Klapach et al. 1996) and the *PCK1* (repressed by glucose and induced by succinate) (Leuker, Sonneborn et al. 1997) promoters all require changes in sugar composition of the medium, a condition that might potentially influence gene transcription to a similar extent as did the amino acid starvation in our setup. The same problem could be encountered with the *SAP2* promoter, which is induced in media containing proteins as the sole nitrogen source. The delay observed for the induction of this promoter (Staib, Michel et al. 2000) would be an additional inconvenience for our purposes.

It might thus be better to use a promoter that allows a medium-independent control of gene expression where simply an inducing or repressing substance is added which is sufficient to turn on or off the promoter, but which lets the cellular metabolism unaffected. One suitable possibility might be the use of the tetracycline-inducible promoter system adapted for use in *C. albicans* by Park *et al.* (Park and Morschhauser 2005). The addition of doxycycline should have less important consequences for the transcription of metabolism genes than a change in the medium composition, even if relatively high doses of this compound are necessary for full induction of the promoter which could have also side effects (Park *et al.* observed at higher doxycycline rates an inhibition of hyphae formation on solid medium). However, when this work has been started, this inducible promoter system was not yet available, and it still would have to be proven that the promoter is strong enough to quickly induce sufficient transcriptional levels of *RIM101SL*.

Nevertheless, even if some Rim101p-regulated genes might have been wrongly excluded due to our stringent selection criteria, the current setup allowed us to describe the transcriptional profiles of a significant number of Rim101p-dependent genes. Amongst others, the probably best described examples for Rim101p-regulated genes, *PHR1* and *PHR2* were clearly regulated as expected from what is known about their *in vivo* regulation. The microarray results indicated that *PHR1* was increasingly upregulated with the time after the expression of active Rim101SLp, while the transcription of *PHR2* decreased and reached its lowest level at the end of the experiment. These transcriptional profiles were shared by many other genes (see Attachment 2 Groups 3 and 4).

On the other hand the role of Rim101p in the transcriptional regulation of many genes seems to be rather complex and apparently cannot be described simply as proportional or inversely proportional to Rim101SLp expression. For instance, we found in the TC experiment a transient repression of *EFG1* transcription, a transcription factor that has been proposed by El Barkani *et al.* to act downstream of Rim101p (El Barkani, Kurzai et al. 2000). In contrast, we could not find a pH-dependent regulation of *EFG1* transcription when comparing midlog cultures of reference strain DAY185 at pH 4 and pH 8, and neither did Bensen *et al.* in their microarray screen for pH- and Rim101p-dependent genes. Although an induction of *EFG1* might have matched better with the expectations than a repression, a function of Rim101p in the transcriptional regulation of *EFG1* makes intuitively sense, and the transient character of the regulation could explain the lack of evidence in the literature.

Moreover, findings by other groups confirmed that it is often more difficult than for the *PHR* genes to define a clear regulatory role of Rim101p. For example, Bensen *et al.* find many genes where the Rim101p-dependent regulation seems to be opposed to the pH-dependent regulation. Other genes, like *KRE6* (beta-1,6-glucan synthesis), were found Rim101p-induced by Bensen *et al.*, but Rim101p-repressed by Lotz *et al.* and in our experiments. We also had the same type of experience with *ALS1*, which we originally identified as a Rim101p-repressed gene, but which turned out to be strongly induced in a Rim101p-dependent manner under alkaline conditions. Such controversial results might be partially explained with the different experimental conditions used by the various research groups, but also indicate that the regulatory function of Rim101p in a genes regulation might not always be easily described and is often dependent on the particular situation. A possible explanation might be the complex regulatory interaction of Rim101p with other transcription factors, similar to what is known from *S. cerevisiae*, where ScRim101p not only represses the transcription of *NRG1*, but also acts together with Nrg1p as a co-repressor of the *DIT1/2* genes (Rothfels, Tanny et al. 2005). We do not see an effect of Rim101p on the transcriptional regulation of *NRG1* (but Lotz *et al.* report it (Lotz, Sohn et al. 2004)), but the observed transient regulation of the important transcription factor *EFG1* suggests that a crosstalk between transcription factors might play an important role in the pH response.

The second part of this work consisted in the characterization of the Rim101p-dependent regulation of the *ALS* gene family. This family was chosen due to its importance in different pathogenesis-linked processes, in particular adherence and endocytosis, and because the microarray results already indicated an important impact of Rim101p on its regulation (see Introduction chapter 1.3.1.3 and Result Chapter 3.2). Although the specific analysis of these

genes is complicated by very strong sequence similarities within the gene family, we could define the role of Rim101p in the regulation of each of these genes. A more detailed characterization of the activity of Rim101p on *ALS* gene promoters was attempted by using both reporter gene assays and *in vivo* ChIP (Chapters 3.2.3 and 3.3). However, none of these methods gave clear results in our hands. In particular we could not demonstrate without doubt that Rim101p directly binds *in vivo* to the predicted promoter sites of the regulated genes *ALS1* and *ALS4*, and that the transcription of theses genes is modified as a consequence of this binding event. Possible reasons have already been discussed in the corresponding chapters. While technical problems in molecular cloning seem to be responsible for the ambiguous results in the reporter gene assays, there exist multiple possible explanations for the failure of the more complex ChIP project which have been discussed already in Chapter 3.3.

Nevertheless some of these results support the idea that Rim101p directly governs the induction of *ALS1* and the repression of *ALS4* at alkaline pH. The reporter gene assays indicated that an upstream promoter region of *ALS1* that contains several extended Rim101p binding motifs is important for the pH-regulated expression of *ALS1*, and also the alkaline repression of *ALS4* transcription was clearly observed in β-galactosidase assays. The ChIP results suggested that Rim101p binds *in vivo* to these promoters to a higher extent when it is activated, even if it could not be demonstrated that this binding event is specific to the promoter regions of pH-regulated genes that carry the predicted Rim101p motifs.

Taken together, our microarray results provide a picture of Rim101p-dependent regulation that is complementary to the ones existing already, because it approaches the function of Rim101p in a pH independent context and focuses on the immediate transcriptional reaction to Rim101p activation in function of the time. These results allowed us to unravel the important role of Rim101p as a regulator of *ALS* gene transcription.

4.2 Perspectives

On the basis of this work several directions could be chosen for the future. To further unravel the complex regulations of the pH response, it might be advisable to look at the contribution of other transcriptional regulators with a possible role. Mds3p has been identified as a regulator of a “weak acid response” (Davis, Bruno et al. 2002). It could be interesting to compare the transcriptional changes evoked by Mds3p with those of Rim101p to see how far these pathways act on separate targets, and how far coregulation of the two pathways plays a

role. This could be done either on a whole genome basis by microarray technology, or on a single gene level, for example for the *ALS* gene family. As we have been able to show, the *ALS* gene family comprises pH-regulated genes that are dependent on Rim101p and others that are not. It would be interesting to test whether Mds3p could be partially responsible for the alkaline repression of *ALS9* or *ALS2*.

In the same line, the importance of Efg1p for the pH-dependent regulation of *ALS* genes could be tested. It has been shown that Efg1p is involved in the regulation of *ALS3* and *ALS1* (Fu, Ibrahim et al. 2002; Argimon, Wishart et al. 2007). Is *ALS1* still alkaline-upregulated by Rim101p in a $efg1^{-/-}$ scenario? Does Rim101p play a role in the complex regulation of the preferentially hypha-transcribed gene *ALS3* (under hypha-inducing conditions: 37 °C, alkaline pH)?

It could be interesting to know how strong is the role of Rim101p-dependent gene regulation in adhesion to host tissues. The importance of Als1p in the adhesion to oral tissues has already been demonstrated *ex vivo* (Kamai, Kubota et al. 2002). The pH-dependent induction of *ALS1* transcription is linked to the presence of activated Rim101p, and the adhesion capacities were tested by Kamai *et al.* in PBS pH 7.4. Hence, Rim101p should be activated under their experimental conditions. Thus, a *rim101* null mutant might show an adhesion defect similar to that of the *als1* null mutant if Rim101p is responsible for regulation of *ALS1* gene expression under these conditions.

Finally, at a certain point one should perhaps come back to the list of 133 identified Rim101p-regulated genes. Among these genes there are certainly other candidates with important functions that might be worth a closer look. Furthermore, it could be worthwile to review our raw data in the context of all publically available results. In addition, microarray data exist for example for a null mutant of the transmembrane protein Dfg16p which is involved in the activation process of Rim101p (Bernhard Hube, personal communication). The use of all these raw data combined could help to identify interesting genes with a solid Rim101p- and pH-dependent regulation that might have passed just below the threshold of the present selection process.

5 Appendix

Attachment 1: Research Article

Titel:

Early transcriptional events triggered by activation of Rim101 in *C. albicans*

Status:

Submitted to Microbiology, currently in revision

Early transcriptional events triggered by activation of Rim101 in *C. albicans*

Michael Weyler, Frédérique Bidard-Michelot, Claude Gaillardin* and Mathias L. Richard

Laboratoire de Microbiologie et Génétique Moléculaire, AgroParisTech, UMR-INRA1238 UMR-CNRS2585, 78850 Thiverval-Grignon, France.

* Corresponding author: C. Gaillardin. Tel: +33-1-30 81 54 53; Fax: +33-1-30 81 54 57; e-mail: claude.gaillardin@grignon.inra.fr

Key Words: Transcript profiling, *MET3* promoter, quantitative PCR, *ALS* genes

Running title: Targets of Rim101 in *C. albicans*

SUMMARY

In order to identify targets of the transcriptional regulator Rim101 which controls pH adaptation and virulence in *C. albicans*, we monitored transcriptional profiles of a strain expressing an inducible, constitutively active form of Rim101 at constant pH. A set of 132 genes showing significant variation of their expression profile over a 90 min time course experiment were thus identified. They were clustered into five transcriptional classes. Out of these 132 genes, 92 underwent rapid changes of expression during the first 15 min of the time course, 26 being induced and 66 being repressed, suggesting direct action of Rim101 on these targets, either as a repressor or as an activator. Accordingly, the consensus Rim101 binding site (CCAAG) was significantly overrepresented in the promoters of this gene set. Results of the time course experiment were confirmed on a subset of these genes by quantitative PCR. Their expression was assessed at different pH in a wild type strain or in a strain deleted for *RIM101*. We thus confirmed that four out of the eight *ALS* genes are regulated by pH and suggest that three of them, *ALS1, ALS4* and *ALS9*, are directly regulated by Rim101.

INTRODUCTION

Candida albicans is a ubiquitous human commensal and a major opportunistic pathogen that causes superficial infections such as vaginal candidiasis or oropharyngeal candidiasis. Under specific clinical conditions, such as neutropenia, *C. albicans* can invade host tissues through the circulatory system and cause severe, often fatal, disseminated infections (Calderone 2002). *C. albicans* is thus able to adapt to highly diverse environments like the skin, the mouth, the digestive tract, the vagina or the blood where pH ranges from 2 to 7.7. Adaptability to these environments is considered as a virulence factor required for survival in the host and eventually invasion.

Adaptation to environmental pH in *C. albicans* involves several regulatory pathways (Davis, Bruno et al. 2002). The Rim101 signaling pathway has been intensively studied for two main reasons. First, in addition to pH adaptation, the Rim101 pathway is required for other functions linked to virulence like morphogenesis or iron uptake (Davis, Edwards et al. 2000; Bensen, Martin et al. 2004). Second, this pathway is not specific to *C. albicans* and orthologues of its components exist in several fungi like *Aspergillus nidulans*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica* or *Ustilago maydis* (Lamb, Xu et al. 2001; Penalva and Arst 2004; Arechiga-Carvajal and Ruiz-Herrera 2005; Blanchin-Roland, Da Costa et al. 2005). In all cases, the key transcriptional regulator Rim101 exists in at least two forms: (i) a full length form, which is present at acidic pH and is considered as transcriptionally inactive; (ii) a truncated and transcriptionally active form, which is formed at neutral to alkaline pH by proteolytic cleavage (Penalva and Arst 2004).

Direct Rim101 targets in *C. albicans* are however still elusive, apart from the *PHR1* and *PHR2* promoters that were shown to bind *in vitro* Rim101 (Ramon and Fonzi 2003; Baek, Martin et al. 2006). Global transcriptional profiles obtained with different experimental set-ups were used to identify candidate Rim101 target genes (Bensen, Martin et al. 2004; Lotz,

Sohn et al. 2004). H. Lotz and collaborators used a microarray specifically designed to monitor the transcriptional status of 117 genes encoding cell surface proteins (Lotz, Sohn et al. 2004). To identify Rim101-induced genes, they compared transcriptional profiles of the reference strain and of a strain overexpressing a constitutively active version of Rim101 (truncated at residue 475) after 6 hours of growth at pH 4.5 in YPD at 30 °C. The same culture set up was used at pH 7.4 for the identification of Rim101-repressed genes by comparing a *rim101*-/- knockout strain with a reference strain. On the other hand E.S. Bensen and collaborators used a whole genome array to compare the transcriptional profiles at pH 4 and pH 8 of a *rim101*-/- knockout and of a wild type strain, after 4 hours of growth in M199 at 37 °C (Bensen, Martin et al. 2004). A common problem with such studies is the distinction between direct and indirect targets of transcriptional regulators. In *S. cerevisiae*, Rim101 may act predominantly by repressing the expression of the repressors Nrg1 and Smp1 (Lamb and Mitchell 2003). In *C. albicans* on the contrary, Rim101 is not a repressor of *NRG1* (Ramon and Fonzi 2003; Bensen, Martin et al. 2004) and it may rather bind directly to its target promoters (Ramon and Fonzi 2003). Accordingly, promoter analysis evidenced that 14 out of 20 genes involved in the Rim101-dependent response carried one or more copies of the consensus Rim101 binding site (Bensen, Martin et al. 2004).

To identify direct targets of Rim101, we monitored in the present study transcriptional changes occurring just after activation of Rim101. We also chose to uncouple the Rim101 response from any other ambient pH effect by inducing overexpression at constant pH of a constitutively active form of Rim101, called Rim101SL. This approach is comparable to previously reported ones (Le Crom, Devaux et al. 2002), in which an active form of a transcriptional factor is conditionally expressed under the control of a regulated promoter. Transcriptional changes are then monitored along a time-course experiment. This enables identification of classes of genes immediately affected by expression of the regulator,

regardless of ambient conditions. Using this approach, we observed that expression of several *ALS* genes was modified upon induction of Rim101 and then showed that three of the six detected *ALS* transcripts (*ALS1*, *4* and *9*) are regulated by Rim101 and that *ALS2* is regulated by the pH independently of Rim101.

METHODS

Culture media and phenotypic tests.

All cultures were carried out at 30 °C. *C. albicans* was routinely grown in YPD plus uridine (2 % Bacto Peptone, 1 % yeast extract, 2 % dextrose, and 80 µg/mL of uridine). Defined SC medium for growth of *C. albicans* consisted of complete synthetic medium CSM without uracil-methionine-cysteine : 6.7g/L yeast nitrogen base without amino acid (Difco) and 2 % glucose. 5 mM methionine and 2 mM cysteine were added when needed to repress *RIM101SL*-transcription in strain FB8. For growth at pH 4 and pH 8, SC medium was buffered with 150 mM HEPES adjusted to pH 4 or pH 8 with HCl or NaOH, respectively. For growth and hypersensitivity tests, droplets of serial dilutions of an exponential-phase culture in YPD medium were spotted onto SC or SC pH10 media buffered with 50 mM glycine-NaOH (with or without 5 mM methionine and 2 mM cysteine); plates were incubated 4 days at 30 °C.

Plasmid construction.

In order to build pINA1337, pGEM-*HIS1* (Wilson, Davis et al. 1999) was digested with *Sal*I and *Sph*I and treated with Mung bean nuclease (New England Biolabs), the resulting *HIS1* containing fragment was cloned into *Hin*dIII-digested and T4 DNA polymerase treated pMET3-Flag (Umeyama, Nagai et al. 2002). To delete the supplementary *Bam*HI restriction site, the resulting plasmid was digested with *Sfi*I and *Bse*RI, treated by Mung bean nuclease and religated to create pINA1337.

Plasmid pINA1341 encoding the truncated, constitutively active form Rim101SL was constructed as follows. In a first step, a G to T substitution at position 1246 of *RIM101* (coordinates as for Swissprot accession number Q9UW14) was introduced to create an in

frame amber codon (D. Onesime, unpublished). This truncation site was chosen after HCA analysis of the *Aspergillus nidulans*, *Yarrowia lipolytica* and *C. albicans* Rim101 orthologs (Lambert, Blanchin-Roland et al. 1997). The 1249 bp *RIM101SL* DNA fragment was amplified by PCR from SC5314 genomic DNA using primers F-Rim101 and R-Rim101SL (Table 1) and checked by sequencing. The PCR fragment was digested with *Hin*dIII and *Xho*I and cloned into *Xho*I/*Hin*dIII-digested pESC-LEU (Stratagene) generating plasmid pDO1. Then, pDO1 was digested with *Bam*HI and *Nhe*I, the Rim101SL containing fragment was cloned into *Bam*HI/*Xba*I-digested pTZ19R (MBI Fermentas), generating pDO5. Plasmid pDO5 was digested with *Bam*HI and *Sph*I, the *RIM101SL* containing fragment was cloned into *Bam*HI/*Sph*I-digested pINA1337, generating plasmid pINA1341.

Strain construction and sequence data.

Bacterial strain used for transformation and amplification of recombinant DNA was *E. coli* DH5α. *C. albicans* sequences data were obtained from the CandidaDB web site: http://genolist.pasteur.fr/CandidaDB/index.html.

The *C. albicans* strains used in this study are listed in Table 2. FB1 and FB8 strains were created by targeting, at the *HIS1* locus of strain DAY5 (Wilson, Davis et al. 1999), the *Swa*I-digested plasmids pINA1337 and pINA1341 respectively. Transformants were selected on synthetic complete medium lacking histidine (SC-His) and single colonies were purified on SC-His. All integration events were confirmed by Southern blot analysis.

Identification of the *RIM101* start codon by the RACE technique.

RNA from strain DAY185 grown in SC medium at pH 4 or 8 was purified as indicated below (see Genomic microarray). The following steps were carried out according to manufacturer's instructions using the GeneRacer™ RACE cDNA kit (Invitrogen). De-capped

RIM101 mRNAs were reverse transcribed using the RIMn697 primer positioned on the 3' region of *RIM101* (Table 1). The cDNA products were then used as templates for a two-step PCR reaction. GeneRacer™ 5'1 primer and RIMn697 were used for a first PCR, and a nested PCR was then done on the first PCR product, with GeneRacer™ 5' Nested primer and RIMn433, a primer hybridising upstream from RIMn697 (Table 1). The procedure was repeated with RNA purified from cultures grown at either pH 4 or pH 8 to check for possible alternative start sites. PCR products were subsequently sequenced.

Genomic Microarray

Strains FB1 or FB8 were pregrown overnight in YPD and inoculated in liquid SC media supplemented with uridine, methionine and cysteine at an optical density at 600 nM (OD_{600}) of 0.2. Cultures were grown at 30 °C and 150 rpm until an OD_{600} of 0.8 was reached, sample 0 was taken, the rest of the culture was washed twice with SC medium without methionine and cysteine (SC+Uri-Met-Cys) and resuspended in the same volume of SC+Uri-Met-Cys to induce *MET3*-promoter activation. Additional samples were taken after 15, 30, 45 and 90 min of incubation. For each sample, a volume corresponding to $5x10^8$ cells was sampled. Cells were then pelleted by centrifugation, flash-frozen in liquid nitrogen and stored at –80 °C.

For RNA preparation, frozen cells were broken in a 5 mL Teflon vessel of a Braun micro-dismembrator containing one 7 mm bead of tungsten carbide (Braun), both pre-cooled in liquid nitrogen. The closed flask was shaken at 2.600 rpm for 2 min. RNAs were extracted using the RNeasy Midi Kit (Qiagen). Residual genomic DNA was removed using the RNase-Free DNase Set (Qiagen). RNA quantity and quality were controlled on a 1 % agarose gel and by measuring of the OD_{260} and the ratio OD_{260}/OD_{280} on a 1/500 dilution.

Cy3- and Cy5-labeled cDNAs were prepared from total RNA according to manufacturer's instruction (Eurogentec). The probes were hybridized to whole genome *C. albicans* microarrays containing 6359 genes spotted in duplicates along with 27 control spots (300 bp PCR products; Eurogentec).

For time course experiments with strain FB8, two independent biological replicates were made and labeled at each time point with Cy5 or Cy3 (dye swap). The reference consisted of a labeled pool of all time samples obtained from a third biological replicate, thus providing for each gene an average expression signal to which individual time point signals could be compared.

For the control experiments using strain FB1, samples were taken from two biological replicates at time points 0, 15 and 90. These time points were chosen after analysis of the time course results obtained with FB8. RNAs extracted at the three time points were pooled and used as a reference as in the FB8 experiment. Two additional arrays were done hybridizing the reference pool against itself labeled with Cy5 and Cy3. Results from these arrays were used later to normalize data from the other arrays to exclude labeling biases.

Slides were scanned using a Scanarray 4000 (Packard Biosciences). Two pictures were obtained per slide for both dyes with a resolution of 5 µm. Pictures were analyzed using the software Quantarray (Packard BioChip Technologies). The median value of the signal detected for each spot at each wavelength and the local background were calculated. Low-quality spots were discarded including those with saturated signals, to avoid underestimation of the expression ratios. GeneSpring (Silicon Genetics) software was used to normalize the data and to select genes that were regulated only in the Rim101SL time course, but not in the control experiment. The resulting genes were subjected to a significance analysis using SAM (Significance Analysis of Microarrays (Tusher, Tibshirani et al. 2001) with a median "False Significant Number" of 1.88 and a "False Discovery Rate" of 1.4. The Supplementary file S1

provides technical details and references of each experiment linked to each raw data file available for download.

Real time quantitative PCR experiments

Gene expression was determined by real time quantitative PCR (qPCR) using a LightCycler® (Roche Molecular Biochemicals). Suitable primers were chosen using the LightCycler® Probe Design Software 1.0. Oligonucleotides are listed in Table 1. Total RNA was purified using the Qiagen MIDI kit as described above. The Superscript II RNase H- Reverse Transcriptase kit (Invitrogen) was used for reverse transcription of 1 μg of total RNA. For quantitative PCR, cDNA samples were diluted 1:100. Twenty μL PCR reactions contained 4 mM $MgCl_2$, 0.5μM of each primer and 2 μL LightCycler® FastStart DNA Master SYBR Green I for 5 μL of cDNA template. PCR cycles were started at 95 °C for 8 min, followed by 45 cycles at 95 °C for 10 s, 55 °C for 7 s, and 72 °C for 10 s. A negative control with sterile water was performed for each primer set. The threshold cycle was determined as the cycle above which the fluorescence signal reached a baseline level. Gene expression levels were expressed as percentage of *ACT1* gene expression. Each experiment of real time qPCR was done in duplicate.

***ALS* primers**

A set of gene-specific real time qPCR primers was recently published for eight *ALS* gene transcripts, *ALS1*-*ALS7* and *ALS9* (Green, Zhao et al. 2005). We intended to use these primers, but unexpectedly encountered dimerization problems with *ALS6* and *ALS9* primer couples. This was most probably due to primer concentrations (500 nM) recommended for use of the Lightcycler® PCR system, which are higher than those Green *et al.* used (100 nM). To circumvent these problems, new primers were designed for these two genes.

When we started our analysis, the sequence data from Assembly19 available at CandidaDB (http://genolist.pasteur.fr/CandidaDB/) and Candida Genome Database (CGD, http://www.candidagenome.org/) suggested the presence of additional *ALS* genes (*ALS10 ALS11* and *ALS12*) in strain SC5314 which were not included in the analysis of (Green *et al.*, 2005). Since *ALS3* primers of the above mentioned primer set did not distinguish *in silico* between *ALS3* (orf19.1816/orf19.9379) and *ALS10*, a new primer couple was designed to solve this problem. After completion of our real-time PCR quantifications, *ALS10* (5' domain of *ALS2* and 3' domain of *ALS3*), *ALS11* (identical to *ALS9*) and *ALS12* (fragments of *ALS2* and/or *ALS4*) were recognized as assembly artefacts and removed from Assembly 20 of CGD. Consequently, results obtained for these genes are not reported below.

The list of primers used for *ALS3, ALS6* and *ALS9* gene transcript quantification is given on Table 1. All PCR products obtained with this set of primers were checked by sequencing for specificity.

RESULTS

Strain FB8 overexpresses an inducible, truncated and fully functional form of Rim101.

The transcriptional factor Rim101 is a major player of *C. albicans* pH response (Porta, Wang et al. 2001; Bensen, Martin et al. 2004). Our aim was to identify new targets of the Rim101 regulon by monitoring early transcriptional changes following activation of the Rim101 pathway. Since the pH response of *C. albicans* involves both Rim101-dependent and -independent pathways (Davis, Wilson et al. 2000), we wanted to dissociate the Rim101 response from other responses to ambient pH. To this end, we built the strain FB8 (see Methods), which conditionally expresses Rim101 under the control of the regulated promoter *MET3*, induced in media devoid of cysteine and methionine (Care, Trevethick et al. 1999).

In order to exchange *RIM101* and *MET3* promoters, we first identified the *RIM101* start codon. Two different translational starts, separated by 174 bp or 58 aa, are proposed in the databases (see GenBank accession numbers EAK96009 vs. AAD51714 or http://genolist.pasteur.fr/CandidaDB/ vs. http://www.candidagenome.org/). Identification of the *RIM101* mRNA start site (see Methods) indicated that both AUG codons were transcribed along with a 25 bp 5' UTR (data not shown), suggesting that the first AUG was used for translation. This yields a Rim101 protein of 661 amino acids as proposed initially (El Barkani, Kurzai et al. 2000). No difference in the starting site was observed when the culture was done either at pH 4 or pH 8.

Since the native *RIM101* product is deemed transcriptionnally inactive and requires proteolytic processing to be activated (Davis 2003), we designed a truncated, active form. The exact processing site of *C. albicans* Rim101 is unknown, but the final processing site of its homologue PacC has been pinpointed in the fungus *A. nidulans* (Penalva and Arst 2004). Through Hydrophobic Cluster Analysis (Gaboriaud, Bissery et al. 1987), we identified a

putative cleavage site of Rim101 sharing the same environment as the PacC site. Insertion of a stop codon after residue 415 (coordinates as in Swissprot accession number Q9UW14) of the native protein resulted in Rim101SL (Rim101 Short Length), a protein predicted to be constitutively active. *RIM101SL* was thus cloned under the control of the *MET3* promoter into the *rim101*-/- knockout strain DAY25, to yield strain FB8. A control strain FB1, identical to FB8 but devoid of *RIM101SL*, was also constructed (see Methods).

We then verified that the *RIM101SL* construct was fully functional in FB8. First, a phenotypic test showed that at pH10 on SC medium, FB8 growth was inhibited when the *MET3* promoter was repressed, mimicking a *RIM101* deletion (Davis, Bruno et al. 2002), whereas its growth became similar to the wild type when the *MET3* promoter was active (Figure 1A). Second, we quantified directly by real time qPCR the transcripts of *RIM101SL* and of its direct target *PHR1* (Ramon and Fonzi 2003) at different time points after the induction of the *MET3*-promoter (Figure 1B). Derepression of the *MET3* promoter led within 15 min to a rise of *RIM101* transcripts, from undetectable to levels close to those of actin. These levels were roughly 10 times higher than those reported for *RIM101* expressed under its own promoter at pH 7.5 in DAY185 (Cornet, Bidard et al. 2005). *PHR1* induction occurred more slowly than *RIM101SL* induction, as expected for a target of Rim101, and the transcriptional levels were in the same range as those reported previously (Cornet, Bidard et al. 2005). This shows that FB8 can be induced to overexpress a functional form of Rim101SL that regulates the expression of at least one of Rim101 targets.

Transcriptional changes following *RIM101SL* induction.

Based on this last experiment, we decided to monitor transcriptional patterns at time 0, 15, 30, 60 and 90 minutes after *MET3* promoter induction using microarrays. We compared the results obtained with strain FB8 to those of FB1 to exclude genes affected by the

induction conditions (sulphur amino acid starvation) independently of Rim101SL. This yielded 609 candidate genes, which were subjected to a Significance Analysis of Microarrays (Tusher, Tibshirani et al. 2001). By this mean, 132 genes were identified as potential Rim101 targets (see Supplementary data S2 and S3). This is likely an underestimate, since we used stringent cut-offs to avoid selecting spuriously induced genes. For instance, *RIM8* has been proposed to be repressed by Rim101 (Porta, Ramon et al. 1999). *RIM8* expression was actually repressed during the time course experiment, but the gene was eliminated during data analysis because a slight repression of *RIM8* was also observed in the control experiment.

Enrichment of the Rim101 binding site

A Rim101 recognition site CCAAG(AAA) related to the *A. nidulans* PacC site (GCCARG) has been proposed (Ramon and Fonzi 2003). We checked for the presence of this element using the Regulatory Sequence Analysis Tools (http://rsat.scmbb.ulb.ac.be/rsat/) in the upstream sequence of the 132 genes identified here, limiting the size of the upstream sequence to either 1000 bp or to the start/stop of an adjacent ORF (see S2 for data): 63.2 % of the selected genes contained at least one copy of the CCAAG motif in their promoter region, compared to 45.9 % ($p < 0.04$) for all upstream regions. An extended site RCCAAG motif was observed in 53 % of the cases. Altogether, these results suggest that our experimental design identified direct Rim101 targets. We hence conclude that in *C. albicans* Rim101 acts mainly directly on its targets contrary to Rim101 in *S. cerevisiae*.

We noticed however that expression of one global transcription factor, Efg1 (Stoldt, Sonneborn et al. 1997), was transiently repressed upon induction of Rim101SL, an observation confirmed by real time qPCR assays under the same conditions (see below).

Rim101 acts predominantly as a direct repressor

Most of the transcriptional changes occurred immediately after the induction of *RIM101SL*: 92 (nearly 70 %) out of the 132 genes were affected at time point 15 min and 66 out of these 92 genes were repressed. Globally, more cases of repression than of induction were observed (78 *vs.* 54, respectively).

Using the "Genetree clustering" option in Genespring (Agilent Technologies), the 132 genes were clustered into 5 classes (Figure 2 and Supplementary data S2). Class1 and Class2 represent genes that are immediately induced or repressed, respectively, and remain so until the end of the experiment. Class3 and Class4 group genes that are progressively induced or repressed, repectively, throughout the experiment. Class5 genes undergo transient induction at early stages of the experiment. The well documented *PHR1* and *PHR2* genes are in Classes 3 and 4 respectively, as expected from literature (Ramon and Fonzi 2003).

Functional classes of Rim101 regulated genes

The vast majority of the genes (109 of 132) had an annotated *S. cerevisiae* ortholog assigned to a functional class (see Supplementary data S3). Assuming conservation of functions of the homologues, we could compare representation of each functional class within the Rim101SL responsive set to its overall abundance in the genome. Three main functional classes, represented by 73 genes or 55 % of the total gene set, were affected by induction of Rim101SL. Genes involved in central metabolism appeared as the main target of Rim101SL (31.5 % *vs.* 20.4 % throughout the genome, $p < 10^{-7}$), especially those concerned with Carbohydrate Metabolism (12.9 % *vs.* 6.4 %, $p < 10^{-7}$), Amino Acid Metabolism (8.3 % *vs.* 3.6 %, $p < 10^{-4}$) and Vitamin Metabolism (3 % *vs.* 1.5 %, $p < 0.01$). The next functional class corresponded to biogenesis of cellular components (14.4 % *vs* 3.3 %), with 20 genes of which 16 correspond to genes involved in cell wall organisation and biogenesis. The last class

concerned 13 genes assigned to cell rescue (9.8 % vs. 4.3 %), mainly ion homeostasis and stress response (5 genes each).

Confirmation of Rim101 dependent regulation by real time qPCR and relation to pH.

In order to identify cell surface genes which may play a role in the pathogen life-style, we used the prediction servers "TMHMM Server v. 2.0" (http://www.cbs.dtu.dk/services/TMHMM/) and "SignalP 3.0 server" (http://www.cbs.dtu.dk/services/SignalP/) to search the gene products for transmembrane domains and/or signal peptide cleavage sites. For 33 of the sequences at least one transmembrane domain was predicted, and 20 sequences contained predicted signal peptides (Supplementary data S3). Twelve out of the 16 genes involved in cell wall organisation and biogenesis were thus predicted to encode exported proteins (Supplementary data S3).

This set of 20 genes was selected for confirmation of the microarray results by real time qPCR on the RNA samples used for hybridization (Table 3). We chose preferentially genes with CCAAG motifs in their promoter region and with a coding sequence indicating a probable localization at the cell surface. The real time qPCR results for time points 0, 15 and 90 validated in all cases the transcriptional profiles deduced from the microarray results. However, the real-time PCR experiments often yielded higher foldchange values than expected from micro-array data (Table 3, columns 5 and 6).

In order to correlate our results with the well documented function of Rim101 in the pH response, we checked whether similar transcriptional changes could be observed for these 20 genes in the isogenic wild-type strain DAY185 grown at different ambient pH. The fold change between pH 4 and pH 8 was then compared to the foldchange observed between time points 0 and 90 min of Rim101SL induction (Table 3, columns 6 and 7). Among the 16 genes that were repressed by Rim101SL in the microarray experiment, 10 were found to be

repressed at least 1.5-fold at pH 8 compared to pH 4 in DAY185. Similarly, three of the four Rim101SL-induced genes were also induced more than 1.5-fold at pH 8 compared to pH 4. However, the genes *ALS1* and *IPF16514* appeared to be induced at alkaline pH, although they were classified as repressed by Rim101SL in the microarray experiments. Finally, six genes were not significantly pH-responsive in DAY185 (*PGA52, IPF16514, EFG1, IPF1372, PGA4/GAS1,* and *IPF4580*), although they appeared Rim101SL-responsive according to the microarray data and carried putative Rim101 binding sites in their promoters. This suggests that some at least of the Rim101 targets are not directly linked to a sustained pH adaptation.

Effect of pH and Rim101 on *ALS* gene transcription

ALS genes expression is modulated during diverse biological processes like adhesion, biofilm formation and virulence (Kamai, Kubota et al. 2002; Sheppard, Yeaman et al. 2004; Zhao, Oh et al. 2005). Several *ALS* genes were shown to be pH- or Rim101-responsive : *ALS1, ALS3/8* and *ALS10* were reported to be upregulated under alkaline conditions (Bensen, Martin et al. 2004), whereas a truncated form of Rim101 was shown to induce *ALS1* and *ALS5* and to repress *ALS4* (Lotz, Sohn et al. 2004).

Among the 6039 probes present on the microarray, eight corresponded to *ALS* genes. Four of them appeared in the list of 132 genes classified as regulated by Rim101SL. All belonged to the group of gradually down-regulated genes (Class 4, Figure 2). When the array was designed, the correct assembly of *ALS* gene sequences and their nomenclature was still tentative, so that probes for misassembled *ALS* sequences like *ALS10*, *ALS11* and *ALS12* were present on the array but were not taken into account (see Methods). In addition, some microarray probes are predicted to bind several *ALS* genes with similar specificities, like *ALS5* and *ALS6* or *ALS2* and *ALS4*.

We confirmed the microarray results obtained for *ALS1* and *ALS2*/*ALS4* by quantitative PCR. In DAY185 grown at alkaline pH, *ALS2/4* expression was lower than at acidic pH, whereas *ALS1* levels were higher at alkaline than at acidic pH (table 3), as previously reported by others (Lotz, Sohn et al. 2004) but contrary to our time-course results.

To investigate the role of Rim101 in the specific regulation of the different *ALS* genes, we monitored more specifically the transcription of each *ALS* gene by real-time qPCR (see Methods) at pH 4 and pH 8 in strain DAY185 (Figure 3). No transcription of *ALS7* and *ALS3* could be detected under the conditions used here. *ALS3* is known to be hypha-specific (Hoyer, Payne et al. 1998; Green, Zhao et al. 2005), and was hence not expected to be expressed under our culture conditions. Previous studies with promoter fusion evidenced a transient expression of *ALS7* in a murine model of infection (Green, Zhao et al. 2005), which may suggest a host-dependent response. The highest transcriptional levels under our conditions were observed for *ALS1* at pH 8 and for *ALS4* at pH 4. The increased transcription of *ALS1* and the decreased transcription of *ALS4* at pH 8 compared to pH 4 were confirmed by qPCR. In addition, we observed that transcription of *ALS2* and *ALS9* was overtly lower at alkaline than at acidic pH in strain DAY185, whereas *ALS5* and *ALS6* transcription did not vary between pH 4 and pH 8. In summary, four out of the eight *ALS* genes were regulated in a pH dependent manner.

We approached the role of Rim101 in the regulation of *ALS* genes by comparing their expression in the reference strain DAY185 and in a *ΔΔrim101* strain (DAY25). The presence of Rim101 had a strong impact on the expression of *ALS1* and *ALS4* which became marginally pH responsive in the *rim101*-/- null mutant (Table 4). On the contrary, transcription of *ALS2* at alkaline pH and acidic pH was unaffected by the *RIM101* status of the strains. Finally, although *ALS9* transcription was repressed at alkaline pH in both DAY25 and DAY185, the repression in the *rim101*-/- null mutant was weaker than in the reference strain, indicating that Rim101 plays a minor role in the regulation of *ALS9* transcription.

CONCLUSION

To identify primary targets of the pH responsive Rim101 regulator, we relied on the induction of the artificially truncated form Rim101SL instead of using more physiological conditions like induction by an external pH shift. This strategy avoids monitoring pH adaptative responses not dependent on Rim101, but entails other drawbacks that should be kept in mind when analysing the array data. First, differences in regulatory activity between Rim101SL and Rim101 are expected, due to possible differences in mRNA stability or proteolytic processing of the truncated protein. Allele-specific variations in phenotype have been observed for PacC mutations in *A. nidulans* (Penalva & Arst, 2004) and for Rim101 in *C. albicans* : for instance, our truncated construct induces filamentation at acidic pH whereas others do or not (Cornet, Bidard et al. 2005). Second, our construct overexpresses *RIM101SL* transcripts as shown by qPCR comparative assays. We do not know if this resulted in non physiological levels of the Rim101 protein, but noticed that expression of one of its targets at least (*PHR1*) remained within physiological range. Third, induction of the *MET3* promoter required a shift from methionine-containing to methionine-free medium. Analysis of array data (see supplementary data S3) evidenced significant expression changes of genes involved in sulfur metabolism as well as of ribosomal protein encoding genes. To filter out these effects, we used data collected during a similar time course experiment from an isogenic strain devoid of *RIM101*. Finally, since time point experiments require numerous arrays, a low number of biological replicates was produced which may increase background noise. We notice however that qPCR assays confirmed array data on 20 genes analyzed at 3 time points, although fold changes were often underestimated on arrays. Taken together, we believe that our approach, although certainly biased in some cases, is largely robust.

The results we obtained suggest that Rim101 in *C. albicans* mainly acts directly on its targets and not through relay transcription factors like it does in *S. cerevisiae* with Nrg1 or Smp1 for instance (Lamb and Mitchell 2003). One question remains concerning the possible role of Efg1 as relay. We notice that *RIM101* has been suggested to act upstream of *EFG1* (El Barkani, Kurzai et al. 2000), and we observed a transient variation of *EFG1* expression following Rim101SL induction. The *EFG1* promoter is predicted to contain four Rim101 binding sites, but its expression under steady state conditions at pH 4 and pH 8 is constant (unpublished).

Comparison of our gene set with the previously published ones (Bensen, Martin et al. 2004; Lotz, Sohn et al. 2004) evidences very little overlap: five genes only were retrieved by the three experiments (*KRE6*, *ALS1*, *PHR1*, *PHR2* and *PGA52*; Figure 4). Two types of explanation can be considered. First, there is little overlap between previous studies, reflecting different experimental sets ups in terms of culture conditions (pH, temperature, medium and time before cell harvest) and probably also differences in data acquisition and analysis (cut offs, normalisation procedure, etc…). Opposite results can even be observed: for instance, *KRE6* was found induced by (Bensen *et al.*, 2004), while we and (Lotz *et al.*, 2004) found it repressed. Second, our set-up targets immediate responses, whereas other studies recorded steady state conditions where sustained transcriptional responses may reflect output of complex regulatory networks: a striking example is provided by *ALS1*, which was found was found to be immediately repressed upon induction of Rim101SL in our microarray assays, whereas we and other groups found it induced under steady state alkaline conditions: such discrepancies between immediate and steady state responses clearly deserve further studies.

Our aim was also to get a global view of changes in cell surface protein expression, in order to link Rim101 regulation and remodelling of the cell surface when *Candida* colonizes

environments of different pH (mouth, digestive tract or vagina). Thus, Als protein family, a well known cell surface protein family, was subjected to a preliminary analysis.

ALS1 and *ALS4* appear to be the main *ALS* genes which are differentially regulated by Rim101. *ALS1* expression is clearly modulated by other conditions than just pH, like biofilm formation (O'Connor *et al.*, 2005), addition of fresh medium or disseminated candidiasis. Intriguingly, in our experiments Rim101 seems to play a dual role in *ALS1* regulation: as a repressor immediately after its induction and as an activator during the sustained pH response. Since a transient repression of *EFG1* is observed immediately after Rim101SL induction, the apparent repressive effects of Rim101SL on *ALS1* expression may actually be due to Efg1, a known regulator of *ALS1* expression (Fu, Ibrahim et al. 2002). This point clearly deserves more studies. Concerning *ALS4*, recent data suggest that it may be turned on to compensate for an *ALS2* defect and that a mutant strain deleted for *ALS4* does not show a marked phenotype except for a slight decrease in adherence (Zhao, Oh et al. 2005). Our results suggest that it may be worth further analysing *ALS4* mutants under acidic conditions, mimicking e.g. vaginal infections.

Acknowledgments

We thank A.P. Mitchell for generously providing strains DAY5, DAY25 and DAY185 and to L. Hoyer for helpful comments on changes having occurred in the *ALS* nomenclature as well as on possible sources of primer specificity problems. We are indebted to Galar Fungail 2 consortium members for continuous discussion and support. MW and FBM were funded by the European Commission grants MRTN-CT-2003-504148 and QLK2CT-2000-00795, respectively.

TABLES

Table 1: *Primers used in this study*

Other primers for *ALS* gene amplification were defined by (Green *et al.*, 2005)

Name	Sequence	Target
OFB16	TGTGACGACCATGTTGGTAGAAAGT	RIM101
OFB17	CTTGAGGTCTCTTGAACGATTTGGG	Id.
OFB22	GCAGTGCTTCAATCAATAGCAAGGC	PHR1
OFB23	AGAGCTTGAGCTGGACCCAGA	Id.
OFB32	AGTGTGACATGGATGTTAGAAAAGAATTATACGG	ACT1
OFB33	ACAGAGTATTTTCTTTCTGGTGGAGCA	Id.
OFB40	ACACTGACGCTTCTGCTTTCG	PHR2
OFB41	GCAGCTTCGTCTTCATCACCACA	Id.
F-Rim101	GACCTCGAGAATTACAACATTCATCCCG	RIM101
R-Rim101SL	GTACCAAGCTTAGAAAGCAGTTATAGTTGG	Id.
RIMn697	CATGGTCGTCACACAAATGATCG	Id.
RIMn433	GTTGGTAGCCATAAGTTGGTTGG	Id.
A3newF	CCAAAACTTGTTCATCTAATGGTATCT	ALS3
A3newR	TAGCATACGACAAGGTGTACGAAT	Id.
A6newF	TTTGATGATAAGTCGTCGGCA	ALS6
A6newR2	GCGATAAATCCATTATTGGTTTCA	Id.
A9newF	ACCCTCATGGATCTGAGACTATTG	ALS9
A9newR	ACCGAACCAGAACCATCGTAT	Id.

Table 2: *Strains used in this study*

Candida albicans strains	*Genotype*	Reference or source
SC5314	*Clinical isolate*	(Fonzi and Irwin 1993)
DAY5	*ura3Δ ::λimm434 his1::hisG arg4::hisG rim101::ARG4* *ura3Δ ::λimm434 his1::hisG arg4::hisG rim101::URA3*	(Wilson, Davis et al. 1999)
DAY25	*ura3Δ ::λimm434 pHIS1::his1::hisG arg4::hisG rim101::ARG4* *ura3Δ ::λimm434 his1::hisG arg4::hisG rim101::URA3*	(Davis, Wilson et al. 2000)
DAY185	*ura3Δ ::λimm434 pHIS1::his1::hisG pARG4::URA3::arg4:hisG* *ura3Δ ::λimm434 his1::hisG arg4::hisG*	(Davis, Wilson et al. 2000)
FB1	*ura3Δ ::λimm434 pHIS1::his1::hisG arg4::hisG rim101::ARG4* *ura3Δ ::λimm434 his1::hisG arg4::hisG rim101::URA3*	This study
FB8	*ura3Δ ::λimm434 pRIM101SL::HIS1::hisG arg4::hisG rim101::ARG4* *ura3Δ ::λimm434 his1::hisG arg4::hisG rim101::URA3*	This study

Table 3: ***Confirmation of the array results***

Foldchange repression

Gene name	TM domains	Signal peptide	CCAAG motives	microarray 90'/0'	qPCR 90'/0'	qPCR pH4/pH8
ALS2/4	0	Yes	1	**9.5**	**10.8**	**12.4**
IPF8762	0	No	0	**7.2**	**24.2**	**3.1**
QDR1	10	Yes	4	**3.9**	**4.5**	**2.2**
IPF6156	1	Yes	1	**2.2**	**3.7**	**3.1**
PHR2	1	Yes	4	**2.2**	**8.9**	**10.3**
WSC4	1	No	1	**2.2**	**2.7**	**1.7**
PHO87	10	No	2	**1.9**	**3.3**	**1.9**
CPA1	0	Yes	0	**1.9**	**1.8**	**11.4**
IPF2280	0	No	6	**1.7**	**2.1**	**2.4**
KRE6	1	No	3	**1.5**	**1.9**	**2.4**
PGA52	0	No	1	*4.4*	*7.1*	*1.2*
EFG1	0	No	4	*1.8*	*1.7*	*1.3*
IPF1372	5	No	1	*1.3*	*1.8*	*1.1*
PGA4	0	Yes	4	*1.2*	*0.7*	*0.9*
ALS1	0	Yes	3	1.8	3.9	N/D
IPF16514	0	No	1	1.4	1.1	0.5

Foldchange induction

Gene name	TM domains	Signal peptide	CCAAG motives	microarray 90'/0'	qPCR 90'/0'	qPCR pH4/pH8
HGH1	**0**	No	**1**	**1.2**	**2.4**	**2.5**
CHO2	**10**	No	**0**	**1.6**	**3.2**	**2.6**
PHR1	**0**	Yes	**2**	**3.0**	**43.1**	**95.6**
IPF4580	*8*	No	*1*	*3.4*	*27.0*	*1.0*

Table 4: ***Rim101 and pH effects on ALS gene regulation***

Alkaline induction	WT (DAY185) pH8/pH4	***rim101-/-*** (DAY25) pH8/pH4
ALS1	39.1	3.2
Alkaline repression	WT (DAY185) pH4/pH8	***rim101-/-*** (DAY25) pH4/pH8
ALS2	3.3	2.5
ALS4	20.9	0.6
ALS9	4.5	1.9

FIGURE LEGENDS

Figure 1: *Phenotypes of a strain expressing Rim101SL*

(A) Cells from an overnight culture of WT (SC5314), ΔΔrim101 (DAY5) and Met3-RIM101SL (FB8) were spotted on SC and SC pH10, and photographed after 4 days of growth at 30 °C. (B) real-time qPCR on *PHR1* and *RIM101SL* transcripts were done 20, 40, 60, 80 and 100 min after induction of Rim101SL expression.

Figure 2: *Gene expression during the time course experiment after clustering*

Genetree clustering function in Genespring® software (Agilent Technologies) allows the clustering of gene with equivalent expression profile. 5 classes were created (see text).

Figure 3: *ALS gene regulation by pH*

Comparison of *ALS* gene expression was followed by real-time qPCR using 2 growth conditions: pH 4 and pH 8. Expression levels were compared to actin levels and expressed as % of actin mRNA.

Figure 4: *Comparison of the genes regulated by Rim101 in three different studies*

VENN diagram of the microarray results. Data from (Bensen, Martin et al. 2004; Lotz, Sohn et al. 2004) and from the present study are represented by a list of genes regulated by Rim101. Corresponding genes at the intersection of the 3 lists are presented in the table.

SUPPLEMENTARY DATA

S1 Array Technical Details

This file contains a description of the experimental set up (strains, growth conditions, labeling, microarray manufacturer as well as the informatics tools used). In addition, a table gathers all the information concerning the raw data: biological replicates and technical replicates.

S2 Genes Clustering

This file contains the results of a Genetree clustering using Genespring® software (Agilent). The clustering allowed the definition of 5 classes of genes regarding their regulation throughout the time course experiment.

S3 Functional Classes

This file contains a comprehensive dataset of the 132 Rim101 regulated genes according to their clustering class, their function using the predicted function of their *S. cerevisiae* orthologs as well as other information on their regulation provided by CGD website.

Supplementary Raw Data

This directory contains all the raw data used for the array analysis (see S1), available on request as text files, will be downloadable.

REFERENCES

Arechiga-Carvajal, E. T. & Ruiz-Herrera, J. (2005). The RIM101/pacC homologue from the basidiomycete Ustilago maydis is functional in multiple pH-sensitive phenomena. *Eukaryot Cell* **4**, 999-1008.

Baek, Y. U., Martin, S. J. & Davis, D. A. (2006). Evidence for novel pH-dependent regulation of Candida albicans Rim101, a direct transcriptional repressor of the cell wall beta-glycosidase Phr2. *Eukaryot Cell* **5**, 1550-1559.

Bensen, E. S., Martin, S. J., Li, M., Berman, J. & Davis, D. A. (2004). Transcriptional profiling in Candida albicans reveals new adaptive responses to extracellular pH and functions for Rim101p. *Mol Microbiol* **54**, 1335-1351.

Blanchin-Roland, S., Da Costa, G. & Gaillardin, C. (2005). ESCRT-I components of the endocytic machinery are required for Rim101-dependent ambient pH regulation in the yeast Yarrowia lipolytica. *Microbiology* **151**, 3627-3637.

Calderone, R. A. (2002). *Candida and Candidiasis*, Calderone, R.A. edn: ASM Press.

Care, R. S., Trevethick, J., Binley, K. M. & Sudbery, P. E. (1999). The MET3 promoter: a new tool for Candida albicans molecular genetics. *Mol Microbiol* **34**, 792-798.

Cornet, M., Bidard, F., Schwarz, P., Da Costa, G., Blanchin-Roland, S., Dromer, F. & Gaillardin, C. (2005). Deletions of endocytic components VPS28 and VPS32 affect growth at alkaline pH and virulence through both RIM101-dependent and RIM101-independent pathways in Candida albicans. *Infect Immun* **73**, 7977-7987.

Davis, D., Edwards, J. E., Jr., Mitchell, A. P. & Ibrahim, A. S. (2000a). Candida albicans RIM101 pH response pathway is required for host-pathogen interactions. *Infect Immun* **68**, 5953-5959.

Davis, D., Wilson, R. B. & Mitchell, A. P. (2000b). RIM101-dependent and-independent pathways govern pH responses in Candida albicans. *Mol Cell Biol* **20**, 971-978.

Davis, D. (2003). Adaptation to environmental pH in Candida albicans and its relation to pathogenesis. *Curr Genet* **44**, 1-7.

Davis, D. A., Bruno, V. M., Loza, L., Filler, S. G. & Mitchell, A. P. (2002). Candida albicans Mds3p, a conserved regulator of pH responses and virulence identified through insertional mutagenesis. *Genetics* **162**, 1573-1581.

El Barkani, A., Kurzai, O., Fonzi, W. A., Ramon, A., Porta, A., Frosch, M. & Muhlschlegel, F. A. (2000). Dominant active alleles of RIM101 (PRR2) bypass the pH restriction on filamentation of Candida albicans. *Mol Cell Biol* **20**, 4635-4647.

Fonzi, W. A. & Irwin, M. Y. (1993). Isogenic strain construction and gene mapping in *Candida albicans*. *Genetics* **134**, 717-728.

Fu, Y., Ibrahim, A. S., Sheppard, D. C., Chen, Y. C., French, S. W., Cutler, J. E., Filler, S. G. & Edwards, J. E., Jr. (2002). Candida albicans Als1p: an adhesin that is a downstream effector of the EFG1 filamentation pathway. *Mol Microbiol* **44**, 61-72.

Gaboriaud, C., Bissery, V., Benchetrit, T. & Mornon, J. P. (1987). Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett* **224**, 149-155.

Green, C. B., Zhao, X. & Hoyer, L. L. (2005a). Use of green fluorescent protein and reverse transcription-PCR to monitor Candida albicans agglutinin-like sequence gene expression in a murine model of disseminated candidiasis. *Infect Immun* **73**, 1852-1855.

Green, C. B., Zhao, X., Yeater, K. M. & Hoyer, L. L. (2005b). Construction and real-time RT-PCR validation of Candida albicans PALS-GFP reporter strains and their use in flow cytometry analysis of ALS gene expression in budding and filamenting cells. *Microbiology* **151**, 1051-1060.

Hoyer, L. L., Payne, T. L., Bell, M., Myers, A. M. & Scherer, S. (1998). Candida albicans ALS3 and insights into the nature of the ALS gene family. *Curr Genet* **33**, 451-459.

Hoyer, L. L., Fundyga, R., Hecht, J. E., Kapteyn, J. C., Klis, F. M. & Arnold, J. (2001). Characterization of agglutinin-like sequence genes from non-albicans Candida and phylogenetic analysis of the ALS family. *Genetics* **157**, 1555-1567.

Kamai, Y., Kubota, M., Hosokawa, T., Fukuoka, T. & Filler, S. G. (2002). Contribution of Candida albicans ALS1 to the pathogenesis of experimental oropharyngeal candidiasis. *Infect Immun* **70**, 5256-5258.

Lamb, T. M., Xu, W., Diamond, A. & Mitchell, A. P. (2001). Alkaline response genes of Saccharomyces cerevisiae and their relationship to the RIM101 pathway. *J Biol Chem* **276**, 1850-1856.

Lamb, T. M. & Mitchell, A. P. (2003). The transcription factor Rim101p governs ion tolerance and cell differentiation by direct repression of the regulatory genes NRG1 and SMP1 in Saccharomyces cerevisiae. *Mol Cell Biol* **23**, 677-686.

Lambert, M., Blanchin-Roland, S., Le Louedec, F., Lepingle, A. & Gaillardin, C. (1997). Genetic analysis of regulatory mutants affecting synthesis of extracellular proteinases in the yeast Yarrowia lipolytica: identification of a RIM101/pacC homolog. *Mol Cell Biol* **17**, 3966-3976.

Le Crom, S., Devaux, F., Marc, P., Zhang, X., Moye-Rowley, W. S. & Jacq, C. (2002). New insights into the pleiotropic drug resistance network from genome-wide characterization of the YRR1 transcription factor regulation system. *Mol Cell Biol* **22**, 2642-2649.

Lotz, H., Sohn, K., Brunner, H., Muhlschlegel, F. A. & Rupp, S. (2004). RBR1, a novel pH-regulated cell wall gene of Candida albicans, is repressed by RIM101 and activated by NRG1. *Eukaryot Cell* **3**, 776-784.

Penalva, M. A. & Arst, H. N., Jr. (2004). Recent advances in the characterization of ambient pH regulation of gene expression in filamentous fungi and yeasts. *Annu Rev Microbiol* **58**, 425-451.

Porta, A., Ramon, A. M. & Fonzi, W. A. (1999). PRR1, a homolog of Aspergillus nidulans palF, controls pH-dependent gene expression and filamentation in Candida albicans. *J Bacteriol* **181**, 7516-7523.

Porta, A., Wang, Z., Ramon, A., Muhlschlegel, F. A. & Fonzi, W. A. (2001). Spontaneous second-site suppressors of the filamentation defect of prr1Delta mutants define a critical domain of Rim101p in Candida albicans. *Mol Genet Genomics* **266**, 624-631.

Ramon, A. M. & Fonzi, W. A. (2003). Diverged binding specificity of Rim101p, the Candida albicans ortholog of PacC. *Eukaryot Cell* **2**, 718-728.

Sheppard, D. C., Yeaman, M. R., Welch, W. H. & other authors (2004). Functional and structural diversity in the Als protein family of Candida albicans. *J Biol Chem* **279**, 30480-30489.

Stoldt, V. R., Sonneborn, A., Leuker, C. E. & Ernst, J. F. (1997). Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. *Embo J* **16**, 1982-1991.

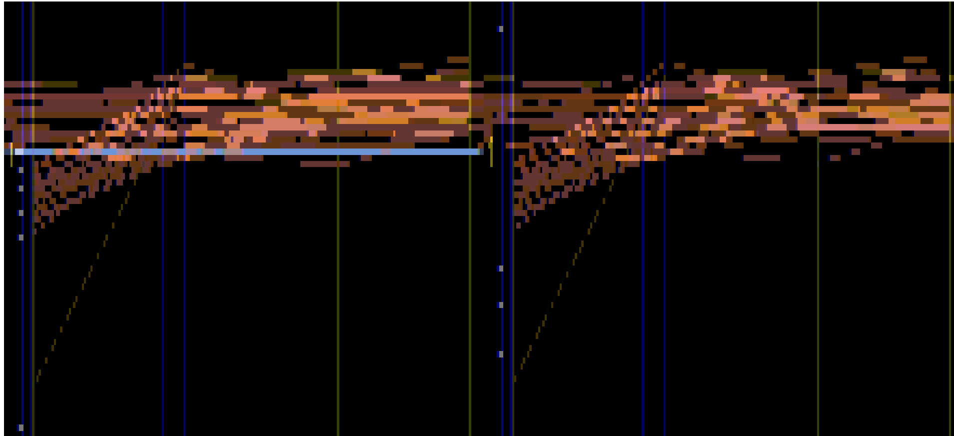
Tusher, V. G., Tibshirani, R. & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121.

Umeyama, T., Nagai, Y., Niimi, M. & Uehara, Y. (2002). Construction of FLAG tagging vectors for Candida albicans. *Yeast* **19**, 611-618.

Wilson, R. B., Davis, D. & Mitchell, A. P. (1999). Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. *J Bacteriol* **181**, 1868-1874.

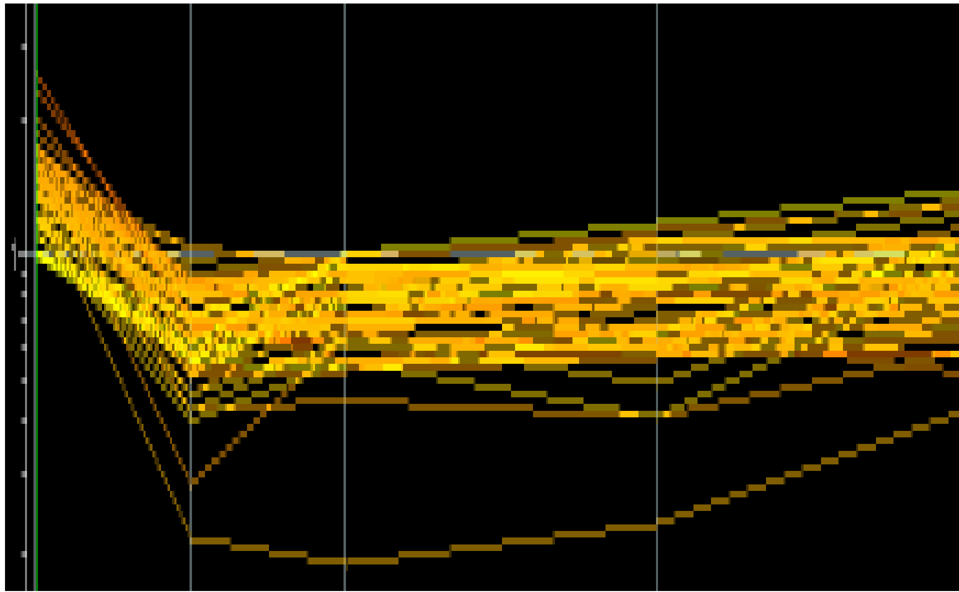
Zhao, X., Oh, S. H., Yeater, K. M. & Hoyer, L. L. (2005). Analysis of the Candida albicans Als2p and Als4p adhesins suggests the potential for compensatory function within the Als family. *Microbiology* **151**, 1619-1630.

Attachment 2: Clusters and raw data for Rim101SL time course



Group 1: Immediately induced genes										
Name	**0'**	**15'**	**30'**	**60'**	**90'**	**15'/0'**	**90'/0'**	**TM**	**Orf19 name**	**Array**
IPF11090	0.25	1.51	1.58	1.40	1.17	6.04	4.68	0	orf19.4623.2	CA5184
GCV2	0.69	1.65	1.24	1.06	1.31	2.37	1.89	0	orf19.385	CA3883
HSP90	0.61	1.23	1.14	0.99	1.29	2.01	2.11	0	orf19.6515	CA4959
RPL27A	0.80	1.46	1.25	1.29	1.21	1.83	1.52	0	orf19.5225.2	CA1972
RPS15.3	0.80	1.46	1.53	1.40	1.47	1.82	1.83	0	orf19.5927	CA6123
KAR2	0.69	1.24	1.24	1.25	1.15	1.81	1.67	1	orf19.2013	CA0915
SHM2	0.70	1.16	1.01	1.10	1.59	1.66	2.28	0	orf19.5750	CA0895
RPS3	0.91	1.49	1.34	1.38	1.44	1.63	1.57	0	orf19.6312	CA3278
CDC3	0.87	1.37	1.30	1.34	1.37	1.57	1.57	0	orf19.1055	CA0844
RPS19A.3	0.73	1.15	1.13	1.18	1.10	1.57	1.50	0	orf19.5996.1	CA6068
IPF8302	0.68	1.02	1.03	1.14	1.21	1.50	1.78	5	orf19.6007	CA6061
IPF5699	0.66	0.99	0.92	1.52	1.35	1.49	2.03	0	orf19.5824	CA3795
IPF470	0.79	1.14	1.13	1.09	1.12	1.44	1.43	0	orf19.7057	CA5638
SPE3	0.80	1.15	1.39	1.14	1.23	1.44	1.54	0	orf19.2250	CA3588
SER2	0.85	1.21	1.08	1.35	1.35	1.43	1.59	0	orf19.5838	CA3782
IPF10270	0.89	1.27	1.14	0.92	1.43	1.43	1.60	0	orf19.1272	CA0948
EGD2	0.75	1.02	1.30	1.38	1.35	1.36	1.81	0	orf19.5858	CA2956
IDH2	0.76	1.02	1.21	1.24	1.30	1.35	1.72	0	orf19.5791	CA4148
TOM20	0.82	1.09	1.17	1.25	1.39	1.33	1.68	1	orf19.2953	CA4179
IPF9582	0.80	1.02	1.11	1.35	1.57	1.28	1.97	0	orf19.688	CA1976
CHO2	0.78	0.98	1.24	1.34	1.33	1.25	1.70	10	orf19.169	CA1414
HSP10.3	0.76	0.95	1.10	1.34	1.72	1.25	2.26	0	orf19.7215.3	CA5341

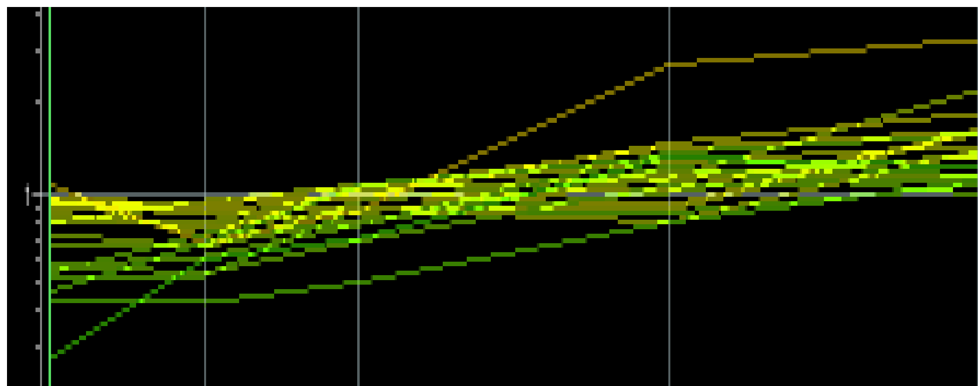
Data of the 22 immediately induced genes for all time points: Columns from left to right: Gene name. time points 0’. 15’. 30’. 60’ and 90’ followed by the foldchange induction between 0’ – 15’ and 0’-90’. In the final three columns you can see the number of predicted transmembrane domains. the ORF19 name and the name of the microarray probe. Genes that are underlied in grey carry a signal peptide or signal anchor ($p > 0.25$).



Group 2: Immediately repressed genes										
Name	**0'**	**15'**	**30'**	**60'**	**90'**	**0'/15'**	**0'/90'**	**TM**	**Orf19 name**	**Array**
CMK1	1.66	0.28	0.64	0.58	0.63	5.89	2.63	0	orf19.5911	CA6135
IPF946	1.56	0.70	0.65	0.77	0.79	2.23	1.98	0	orf19.7561	CA5968
IPF3485	3.41	1.12	1.27	0.85	0.85	3.05	4.03	0	orf19.6757	CA5940
IPF2471	1.32	0.81	0.87	0.99	1.01	1.63	1.31	0	orf19.7437	CA5728
IPF2431	0.98	0.60	0.98	1.15	1.37	1.62	0.71	1	orf19.7417	CA5714
IPF2857	1.61	0.84	0.90	0.61	0.68	1.90	2.35	0	orf19.7284	CA5526
TPS3.3	1.70	0.51	0.52	0.57	0.61	3.34	2.78	0	orf19.5348	CA5505
GSY1	1.51	0.42	0.45	0.41	0.58	3.57	2.57	0	orf19.3278	CA5467
WSC4	1.33	0.81	0.91	0.92	0.83	1.65	1.60	1	orf19.7251	CA5369
GPH1	1.88	0.85	0.70	0.81	0.79	2.21	2.38	0	orf19.7021	CA5206
FBA1	1.12	0.40	0.55	0.41	0.91	2.80	1.23	0	orf19.4618	CA5180
MAK3	1.52	0.61	0.81	0.65	1.04	2.47	1.46	0	orf19.4617	CA5179
MDH1	1.06	0.57	0.80	0.95	1.21	1.87	0.88	0	orf19.4602	CA5164
IPF1372	1.01	0.64	0.80	0.75	0.80	1.57	1.25	5	orf19.6440	CA5100
IPF15297	1.70	0.76	0.81	1.12	1.27	2.24	1.34	0	orf19.3053	CA5078
IPF8796	1.19	0.53	0.81	0.86	0.75	2.27	1.58	0	orf19.4035	CA4800
HSP78.5f	1.47	0.71	0.52	0.81	1.12	2.07	1.31	0	orf19.882	CA4684
NDH1	1.15	0.46	0.60	0.59	0.84	2.50	1.37	0	orf19.339	CA4633
PPM2	1.20	0.94	0.84	0.77	0.77	1.27	1.56	0	orf19.3303	CA4612
QDR1	2.57	0.67	0.59	0.64	0.77	3.84	3.35	10	orf19.508	CA4501
RNR22	1.36	0.21	0.19	0.23	0.42	6.39	3.24	1	orf19.1868	CA4492
PDA1	0.98	0.69	0.92	0.85	1.01	1.41	0.97	0	orf19.3097	CA4412
IPF2280	1.39	0.86	0.86	0.84	0.85	1.62	1.62	0	orf19.6658	CA4264
IPF8762	1.76	0.53	0.55	0.65	0.51	3.33	3.42	0	orf19.822	CA4220
PBI2	1.34	0.71	0.88	0.69	0.79	1.88	1.70	0	orf19.2769	CA4122
TPS1	1.60	0.78	0.77	0.75	1.02	2.04	1.57	0	orf19.6640	CA4084
CAF16	1.31	0.64	0.71	0.80	0.78	2.04	1.69	0	orf19.388	CA3880
IPF8884	1.35	0.68	0.79	0.72	0.71	1.99	1.91	1	orf19.3422	CA3756
PGI1	1.32	0.68	0.67	0.57	0.95	1.95	1.39	0	orf19.3888	CA3559
IPF13583	0.98	0.58	0.87	0.85	0.85	1.68	1.15	2	orf19.2334	CA3337
IPF19983	1.23	0.77	0.90	0.91	1.00	1.59	1.23	0	orf19.2335	CA3336
IPF12241	1.14	0.58	0.54	0.57	0.66	1.98	1.74	0	orf19.2132	CA3181
IPF2968	1.33	1.01	0.93	0.81	0.81	1.31	1.64	1	orf19.4286	CA3062
LPD1	1.14	0.78	0.86	1.05	1.30	1.46	0.88	0	orf19.6127	CA2998

EFG1	1.18	0.51	0.78	0.84	0.94	2.31	1.26	0	orf19.610	CA2787
Cirt4b	1.45	0.68	0.80	0.98	0.74	2.14	1.94	0	orf19.2839	CA2554
PDC11	1.18	0.42	0.63	0.49	0.93	2.81	1.26	0	orf19.2877	CA2474
GPD1	1.53	0.63	0.80	0.64	0.66	2.43	2.31	0	orf19.1756	CA2263
SUR2	2.04	0.71	0.55	0.58	0.89	2.89	2.28	2	orf19.5818	CA2225
IPF16901	1.42	0.83	0.88	0.90	0.83	1.70	1.71	0	orf19.842	CA2020
ECM42	1.50	1.01	0.89	0.82	0.83	1.49	1.81	0	orf19.6500	CA1732
COX15	1.00	0.55	0.52	0.64	1.02	1.81	0.98	8	orf19.3656	CA1688
IPF6156	1.42	0.46	0.76	0.61	0.73	3.10	1.94	1	orf19.1034	CA1625
IPF6342	1.54	0.58	0.85	1.01	1.02	2.68	1.51	0	orf19.1106	CA1458
IPF11858	1.16	0.76	0.62	0.65	0.67	1.52	1.75	0	orf19.1277	CA1411
MRF1	1.72	0.78	0.86	0.88	1.07	2.20	1.61	0	orf19.1149	CA1333
IPF10394	1.17	0.62	0.69	1.05	1.05	1.90	1.12	0	orf19.3364	CA1196
RIB3	1.46	0.73	0.69	0.84	0.86	2.00	1.69	0	orf19.5228	CA1111
IPF4905	1.38	0.67	0.63	0.57	0.71	2.07	1.95	0	orf19.411	CA0899
PHO11	1.07	0.57	0.69	0.66	0.56	1.88	1.90	0	orf19.2619	CA0616
UGP1	0.96	0.55	0.75	0.82	0.90	1.74	1.07	0	orf19.1738	CA0435
PRC1	1.72	0.85	0.97	1.07	1.07	2.02	1.61	0	orf19.1339	CA0430
IPF4065	1.46	0.71	0.60	0.60	0.76	2.05	1.93	0	orf19.1862	CA0386
PRB2	1.52	0.81	0.89	0.93	0.78	1.86	1.93	0	orf19.2242	CA0270
IPF20054	1.34	0.90	0.70	0.65	0.77	1.49	1.75	7	orf19.6117	CA0262
PRC3	1.20	0.69	0.69	0.88	0.96	1.74	1.25	0	orf19.2474	CA0035
IPF16843	1.48	0.86	0.62	0.89	0.96	1.71	1.54	0	orf19.2397.3	CA0026

Data of the 57 immediately repressed genes for all time points: Columns from left to right: Gene name. time points 0’. 15’. 30’. 60’ and 90’ followed by the foldchange repression between 0’ – 15’ and 0’-90’. In the final three columns you can see the number of predicted transmembrane domains. the ORF19 name and the name of the microarray probe. Genes that are underlied in grey carry a signal peptide or signal anchor ($p > 0.25$).

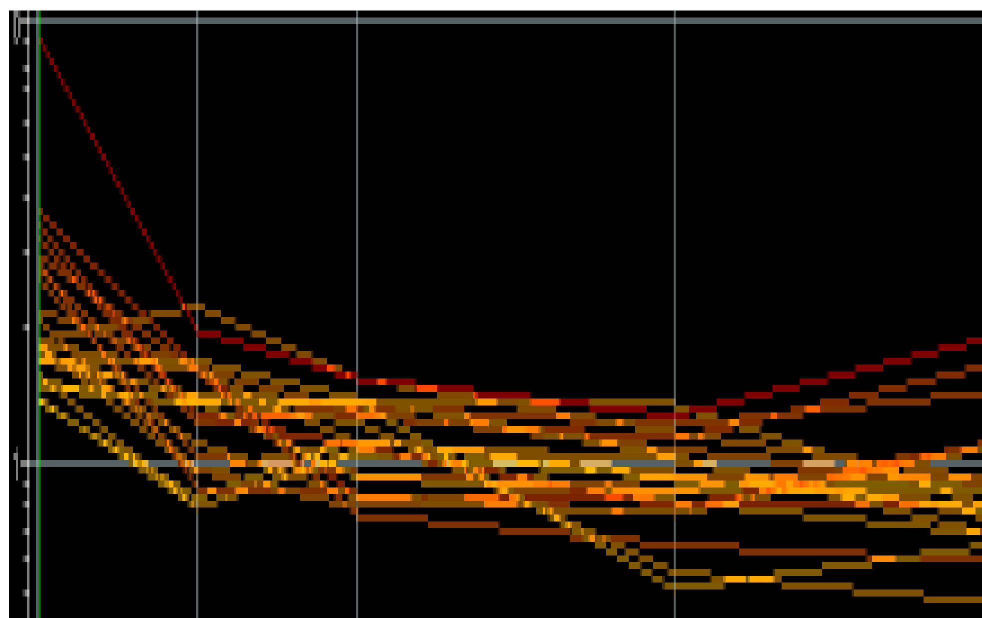


Group 3: Progressively induced genes

Name	0'	15'	30'	60'	90'	15'/0'	90'/0'	TM	Orf19 name	Array
SSS1	0.72	0.64	0.93	1.18	1.10	0.89	1.53	1	orf19.6828.1	CA5881
SLA2	0.27	0.60	0.70	1.31	1.22	2.18	4.43	0	orf19.7201	CA5327
MAM33	0.94	0.87	0.96	1.45	1.84	0.92	1.95	0	orf19.7187	CA5316
IPF4580	0.44	0.42	0.50	0.79	1.09	0.95	2.47	8	orf19.6522	CA4955
PHR1	0.53	0.72	0.77	1.36	1.57	1.35	2.95	0	orf19.3829	CA4857
MTD1	0.94	0.91	0.98	0.90	1.62	0.97	1.72	0	orf19.3810	CA4842
ADE17	0.84	0.75	0.88	0.79	1.37	0.88	1.63	0	orf19.492	CA4513
FUM12.3f	0.96	0.71	1.03	1.41	1.57	0.73	1.64	0	orf19.6725	CA4351
FUM12.5f	0.57	0.53	0.68	1.04	1.10	0.94	1.94	0	orf19.6724	CA4349

ENT3.3f	0.52	0.53	0.88	0.85	1.06	1.01	2.03	1	orf19.1553	CA3979
IPF6712.3f	0.56	0.60	0.77	1.21	0.98	1.07	1.73	0	orf19.1414.2	CA3956
CIT1.exon2	0.46	0.67	1.05	1.30	1.14	1.44	2.46	0	orf19.4393	CA3909
ZRT2	1.06	0.68	0.86	2.66	3.26	0.64	3.08	7	orf19.1585	CA3160
TRR1	0.90	0.92	1.03	1.26	1.43	1.02	1.58	0	orf19.4290	CA3059
AIP2	0.66	0.66	0.81	1.00	1.35	1.01	2.06	0	orf19.300	CA2406
IPF6679	0.91	0.87	0.96	1.01	1.55	0.95	1.70	0	orf19.1306	CA1326
CAN2	0.80	0.79	0.85	1.11	2.18	0.99	2.73	12	orf19.111	CA1191
ADE5.7	0.78	0.93	1.06	1.09	1.32	1.18	1.69	0	orf19.5062	CA0585
MEP3	0.90	0.84	0.88	0.86	1.60	0.93	1.77	11	orf19.1614	CA0302
MRPL3	0.81	0.79	1.01	1.12	1.30	0.98	1.60	0	orf19.5064	CA0089

Data of the 20 progressively induced genes for all time points: Columns from left to right: Gene name. time points 0'. 15'. 30'. 60' and 90' followed by the foldchange induction between 0' – 15' and 0'-90'. In the final three columns you can see the number of predicted transmembrane domains. the ORF19 name and the name of the microarray probe. Genes that are underlied in grey carry a signal peptide or signal anchor ($p > 0.25$).

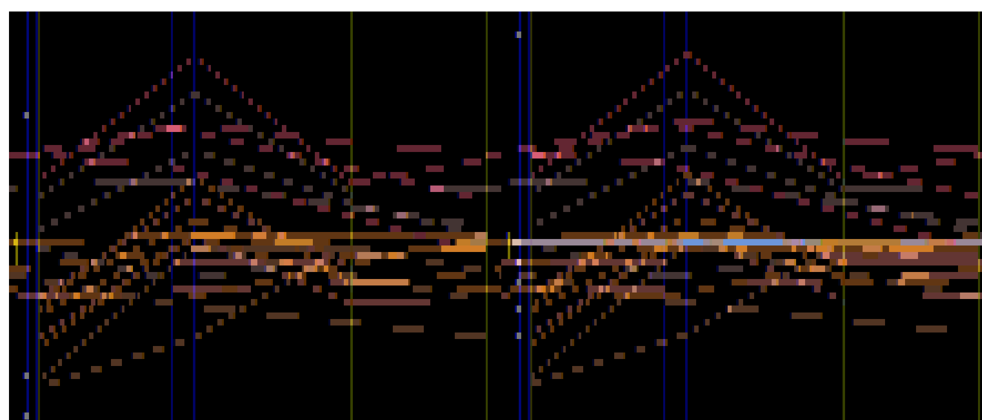


Group 4: Progressively repressed genes

Name	0'	15'	30'	60'	90'	0'/15'	0'/90'	TM	Orf19 name	Array
ARG1	9.19	1.96	1.54	1.27	1.89	4.68	4.86	0	orf19.7469	CA5818
KRE6	1.67	1.40	1.36	1.33	0.78	1.19	2.13	1	orf19.7363	CA5661
SOD22.3f	2.58	1.02	0.93	0.85	1.07	2.53	2.41	0	orf19.7111.1	CA5588
PHR2	1.70	1.69	1.38	1.02	0.68	1.01	2.52	1	orf19.6081	CA3867
IPF15925	2.84	0.87	0.84	0.75	1.10	3.26	2.59	0	orf19.2988	CA3437
POL21	1.55	0.78	1.04	0.80	0.67	1.98	2.31	0	orf19.2668	CA3332
IPF9211.3f	1.80	1.09	0.82	0.75	1.07	1.65	1.68	4	orf19.3712	CA3141
AQY1	1.46	1.36	0.91	0.83	0.87	1.07	1.68	6	orf19.2849	CA2873
CRD1	2.90	1.58	0.75	0.63	0.58	1.84	4.98	8	orf19.4784	CA2832
PST2	1.35	0.81	1.10	0.90	0.90	1.66	1.50	0	orf19.3612	CA1673
IPF8746	2.19	1.61	1.29	1.05	0.87	1.36	2.51	1	orf19.4279	CA1548
ALS4.3f	1.72	1.36	1.21	0.55	0.47	1.26	3.65	0	orf19.4556	CA1528
ALS11.3f	3.66	1.63	0.82	0.81	0.78	2.25	4.71	0	orf19.5745	CA1426
IPF16514	1.38	0.77	1.04	0.92	0.76	1.79	1.82	0	orf19.921	CA1388

ERG1	2.06	0.94	0.77	0.91	0.90	2.19	2.29	2	orf19.406	CA1353
ARG8	2.72	1.21	1.33	1.18	1.42	2.24	1.91	0	orf19.3771	CA1209
CPA1	3.19	1.25	1.26	1.12	1.65	2.55	1.94	0	orf19.4630	CA0874
CPA2	3.44	1.27	1.08	0.78	1.10	2.71	3.13	0	orf19.3221	CA0687
PHO87	1.85	0.93	1.11	0.93	0.77	1.98	2.38	10	orf19.2454	CA0548
ALS12.3f	1.50	1.34	1.35	0.50	0.63	1.12	2.37	0	orf19.2122	CA0413
ALS1.3eoc	1.87	2.22	1.52	1.14	0.90	0.84	2.08	0	orf19.5741	CA0316
IPF15442	1.84	1.60	1.09	0.99	0.74	1.15	2.49	0	orf19.1911	CA0188

Data of the 22 progressively repressed genes for all time points: Columns from left to right: Gene name. time points 0’. 15’. 30’. 60’ and 90’ followed by the foldchange repression between 0’ – 15’ and 0’-90’. In the final three columns you can see the number of predicted transmembrane domains. the ORF19 name and the name of the microarray probe. Genes that are underlied in grey carry a signal peptide or signal anchor (p > 0.25).



Group 5: Transiently induced genes										
Name	0'	15'	30'	60'	90'	15'/0'	90'/0'	TM	Orf19 name	Array
IPF89.3	0.46	0.60	0.99	0.99	0.74	1.30	1.59	0	orf19.5943.1	CA6109
IPF407	1.30	2.26	1.32	1.44	1.12	1.74	0.86	0	orf19.7504	CA5848
ACC1	0.70	1.43	0.80	0.96	0.95	2.05	1.36	0	orf19.7466	CA5816
HGH1	0.60	1.03	0.73	0.78	0.84	1.73	1.40	0	orf19.4587	CA5149
SDH42	0.75	1.12	0.93	1.06	1.05	1.51	1.41	2	orf19.4022	CA4788
CHA12	1.56	1.89	1.71	1.50	0.94	1.21	0.60	0	orf19.1996	CA3945
SNQ2	0.74	1.29	0.88	0.89	0.92	1.73	1.24	12	orf19.5759	CA3828
ARD8	0.69	1.04	0.82	0.73	1.03	1.51	1.50	0	orf19.6322	CA3288
RPL42.3	0.57	1.06	0.98	1.03	0.97	1.84	1.69	0	orf19.4909.1	CA2023
IPF19908	1.41	2.71	1.38	1.87	1.25	1.93	0.89	0	orf19.1344	CA1242
CAR1.3eoc	0.48	0.97	0.98	0.72	0.60	2.02	1.25	0	orf19.11416	CA0781
IPF12884	1.08	1.66	1.24	0.83	1.02	1.54	0.94	14	orf19.4779	CA0778

Data of the 12 transiently induced genes for all time points: Columns from left to right: Gene name. time points 0’. 15’. 30’. 60’ and 90’ followed by the foldchange induction between 0’ – 15’ and 0’-90’. In the final three columns you can see the number of predicted transmembrane domains. the ORF19 name and the name of the microarray probe. Genes that are underlied in grey carry a signal peptide or signal anchor (p > 0.25).

6 References

Anderson, J. H. (1979). "In vitro survival of human pathogenic fungi in seawater." Sabouraudia **17**(1): 1-12.
Arechiga-Carvajal, E. T. and J. Ruiz-Herrera (2005). "The RIM101/pacC homologue from the basidiomycete Ustilago maydis is functional in multiple pH-sensitive phenomena." Eukaryot Cell **4**(6): 999-1008.
Argimon, S., J. A. Wishart, et al. (2007). "Developmental regulation of an adhesin gene during cellular morphogenesis in the fungal pathogen Candida albicans." Eukaryot Cell.
Arnaud, M. B., M. C. Costanzo, et al. (2007). "Sequence resources at the Candida Genome Database." Nucleic Acids Res **35**(Database issue): D452-6.
Arst, H. N., Jr., C. R. Bailey, et al. (1980). "A possible role for acid phosphatase in gamma-amino-n-butyrate uptake in Aspergillus nidulans." Arch Microbiol **125**(1-2): 153-8.
Arst, H. N., Jr., E. Bignell, et al. (1994). "Two new genes involved in signalling ambient pH in Aspergillus nidulans." Mol Gen Genet **245**(6): 787-90.
Arst, H. N., Jr. and D. J. Cove (1970). "Molybdate metabolism in Aspergillus nidulans. II. Mutations affecting phosphatase activity or galactose utilization." Mol Gen Genet **108**(2): 146-53.
Arst, H. N., Jr., D. W. MacDonald, et al. (1970). "Molybdate metabolism in Aspergillus nidulans. I. Mutations affecting nitrate reductase and-or xanthine dehydrogenase." Mol Gen Genet **108**(2): 129-45.
Baek, Y. U., S. J. Martin, et al. (2006). "Evidence for novel pH-dependent regulation of Candida albicans Rim101, a direct transcriptional repressor of the cell wall beta-glycosidase Phr2." Eukaryot Cell **5**(9): 1550-9.
Bailey, C. R., H. A. Penfold, et al. (1979). "Cis-dominant regulatory mutations affecting the expression of GABA permease in Aspergillus nidulans." Mol Gen Genet **169**(1): 79-83.
Bain, J. M., C. Stubberfield, et al. (2001). "Ura-status-dependent adhesion of Candida albicans mutants." FEMS Microbiol Lett **204**(2): 323-8.
Balish, E., T. Warner, et al. (2001). "Oroesophageal candidiasis is lethal for transgenic mice with combined natural killer and T-cell defects." Med Mycol **39**(3): 261-8.
Barker, K. S., S. Crisp, et al. (2004). "Genome-wide expression profiling reveals genes associated with amphotericin B and fluconazole resistance in experimentally induced antifungal resistant isolates of Candida albicans." J Antimicrob Chemother **54**(2): 376-85.
Barwell, K. J., J. H. Boysen, et al. (2005). "Relationship of DFG16 to the Rim101p pH response pathway in Saccharomyces cerevisiae and Candida albicans." Eukaryot Cell **4**(5): 890-9.
Belland, R. J., G. Zhong, et al. (2003). "Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis." Proc Natl Acad Sci U S A **100**(14): 8478-83.
Bennett, R. J. and A. D. Johnson (2005). "Mating in Candida albicans and the search for a sexual cycle." Annu Rev Microbiol **59**: 233-55.
Bensen, E. S., S. J. Martin, et al. (2004). "Transcriptional profiling in Candida albicans reveals new adaptive responses to extracellular pH and functions for Rim101p." Mol Microbiol **54**(5): 1335-51.

Berman, J. and P. E. Sudbery (2002). "Candida Albicans: a molecular revolution built on lessons from budding yeast." Nat Rev Genet **3**(12): 918-30.
Bignell, E., S. Negrete-Urtasun, et al. (2005). "The Aspergillus pH-responsive transcription factor PacC regulates virulence." Mol Microbiol **55**(4): 1072-84.
Blanchin-Roland, S., R. R. Cordero Otero, et al. (1994). "Two upstream activation sequences control the expression of the XPR2 gene in the yeast Yarrowia lipolytica." Mol Cell Biol **14**(1): 327-38.
Blanchin-Roland, S., G. Da Costa, et al. (2005). "ESCRT-I components of the endocytic machinery are required for Rim101-dependent ambient pH regulation in the yeast Yarrowia lipolytica." Microbiology **151**(Pt 11): 3627-37.
Bowers, K., J. Lottridge, et al. (2004). "Protein-protein interactions of ESCRT complexes in the yeast Saccharomyces cerevisiae." Traffic **5**(3): 194-210.
Braun, B. R., W. S. Head, et al. (2000). "Identification and characterization of TUP1-regulated genes in Candida albicans." Genetics **156**(1): 31-44.
Braun, B. R., M. van Het Hoog, et al. (2005). "A human-curated annotation of the Candida albicans genome." PLoS Genet **1**(1): 36-57.
Bustamante, C. I. (2005). "Treatment of Candida infection: a view from the trenches!" Curr Opin Infect Dis **18**(6): 490-5.
Caddick, M. X. and H. N. Arst, Jr. (1986). "Structural genes for phosphatases in Aspergillus nidulans." Genet Res **47**(2): 83-91.
Caddick, M. X., A. G. Brownlee, et al. (1986). "Phosphatase regulation in Aspergillus nidulans: responses to nutritional starvation." Genet Res **47**(2): 93-102.
Caddick, M. X., A. G. Brownlee, et al. (1986). "Regulation of gene expression by pH of the growth medium in Aspergillus nidulans." Mol Gen Genet **203**(2): 346-53.
Calderone, R. A. (2002). Candida and Candidiasis, ASM Press.
Calvo, M. A., J. Guarro, et al. (1980). "Airborne fungi in the air of barcelona, Spain. V. The yeasts." Ann Allergy **45**(2): 115-6.
Cao, F., S. Lane, et al. (2006). "The Flo8 transcription factor is essential for hyphal development and virulence in Candida albicans." Mol Biol Cell **17**(1): 295-307.
Caracuel, Z., M. I. Roncero, et al. (2003). "The pH signalling transcription factor PacC controls virulence in the plant pathogen Fusarium oxysporum." Mol Microbiol **48**(3): 765-79.
Care, R. S., J. Trevethick, et al. (1999). "The MET3 promoter: a new tool for Candida albicans molecular genetics." Mol Microbiol **34**(4): 792-8.
Castrejon, F., A. Gomez, et al. (2006). "The RIM101 pathway contributes to yeast cell wall assembly and its function becomes essential in the absence of mitogen-activated protein kinase Slt2p." Eukaryot Cell **5**(3): 507-17.
Chibana, H., J. L. Beckerman, et al. (2000). "Fine-resolution physical mapping of genomic diversity in Candida albicans." Genome Res **10**(12): 1865-77.
Chibana, H., B. B. Magee, et al. (1998). "A physical map of chromosome 7 of Candida albicans." Genetics **149**(4): 1739-52.
Choi, W., Y. J. Yoo, et al. (2003). "Identification of proteins highly expressed in the hyphae of Candida albicans by two-dimensional electrophoresis." Yeast **20**(12): 1053-60.
Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." Journal of the American Statistical Association **Vol. 83**: 596-610.
Cormack, B. P., G. Bertram, et al. (1997). "Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans." Microbiology **143 (Pt 2)**: 303-11.
Cornet, M., F. Bidard, et al. (2005). "Deletions of endocytic components VPS28 and VPS32 affect growth at alkaline pH and virulence through both RIM101-

dependent and RIM101-independent pathways in Candida albicans." Infect Immun **73**(12): 7977-87.
Cui, X. and G. A. Churchill (2003). "Statistical tests for differential expression in cDNA microarray experiments." Genome Biol **4**(4): 210.
d'Enfert, C., S. Goyard, et al. (2005). "CandidaDB: a genome database for Candida albicans pathogenomics." Nucleic Acids Res **33**(Database issue): D353-7.
Davis, D. (2003). "Adaptation to environmental pH in Candida albicans and its relation to pathogenesis." Curr Genet **44**(1): 1-7.
Davis, D., J. E. Edwards, Jr., et al. (2000). "Candida albicans RIM101 pH response pathway is required for host-pathogen interactions." Infect Immun **68**(10): 5953-9.
Davis, D., R. B. Wilson, et al. (2000). "RIM101-dependent and-independent pathways govern pH responses in Candida albicans." Mol Cell Biol **20**(3): 971-8.
Davis, D. A., V. M. Bruno, et al. (2002). "Candida albicans Mds3p, a conserved regulator of pH responses and virulence identified through insertional mutagenesis." Genetics **162**(4): 1573-81.
De Backer, M. D., D. Maes, et al. (1999). "Transformation of Candida albicans by electroporation." Yeast **15**(15): 1609-18.
De Bernardis, F., F. A. Muhlschlegel, et al. (1998). "The pH of the host niche controls gene expression in and virulence of Candida albicans." Infect Immun **66**(7): 3317-25.
de Graaff, L. H., H. C. van den Broeck, et al. (1994). "Regulation of the xylanase-encoding xlnA gene of Aspergillus tubigensis." Mol Microbiol **12**(3): 479-90.
Denison, S. H., S. Negrete-Urtasun, et al. (1998). "Putative membrane components of signal transduction pathways for ambient pH regulation in Aspergillus and meiosis in saccharomyces are homologous." Mol Microbiol **30**(2): 259-64.
Denison, S. H., M. Orejas, et al. (1995). "Signaling of ambient pH in Aspergillus involves a cysteine protease." J Biol Chem **270**(48): 28519-22.
Diez, E., J. Alvaro, et al. (2002). "Activation of the Aspergillus PacC zinc finger transcription factor requires two proteolytic steps." Embo J **21**(6): 1350-9.
Dolan, J. W., A. C. Bell, et al. (2004). "Candida albicans PLD I activity is required for full virulence." Med Mycol **42**(5): 439-47.
Dorn, G. (1965). "Genetic Analysis of the Phosphatases in Aspergillus Nidulans." Genet Res **10**: 13-26.
Dorn, G. (1965). "Phosphatase mutants in Aspergillus nidulans." Science **150**(700): 1183-4.
Douglas, C. M., J. A. Marrinan, et al. (1994). "A Saccharomyces cerevisiae mutant with echinocandin-resistant 1,3-beta-D-glucan synthase." J Bacteriol **176**(18): 5686-96.
Dudley, A. M., J. Aach, et al. (2002). "Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range." Proc Natl Acad Sci U S A **99**(11): 7554-9.
Dujon, B. (2006). "Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution." Trends Genet **22**(7): 375-87.
Dujon, B., D. Sherman, et al. (2004). "Genome evolution in yeasts." Nature **430**(6995): 35-44.
Ehrenreich, A. (2006). "DNA microarray technology for the microbiologist: an overview." Appl Microbiol Biotechnol **73**(2): 255-73.
Eisendle, M., H. Oberegger, et al. (2004). "Biosynthesis and uptake of siderophores is controlled by the PacC-mediated ambient-pH Regulatory system in Aspergillus nidulans." Eukaryot Cell **3**(2): 561-3.

El Barkani, A., O. Kurzai, et al. (2000). "Dominant active alleles of RIM101 (PRR2) bypass the pH restriction on filamentation of Candida albicans." Mol Cell Biol **20**(13): 4635-47.
Enjalbert, B., D. A. Smith, et al. (2006). "Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen Candida albicans." Mol Biol Cell **17**(2): 1018-32.
Enjalbert, B. and M. Whiteway (2005). "Release from quorum-sensing molecules triggers hyphal formation during Candida albicans resumption of growth." Eukaryot Cell **4**(7): 1203-10.
Enloe, B., A. Diamond, et al. (2000). "A single-transformation gene function test in diploid Candida albicans." J Bacteriol **182**(20): 5730-6.
Ernst, J. F. (2000). "Transcription factors in Candida albicans - environmental control of morphogenesis." Microbiology **146 (Pt 8)**: 1763-74.
Eshel, D., I. Miyara, et al. (2002). "pH regulates endoglucanase expression and virulence of Alternaria alternata in persimmon fruit." Mol Plant Microbe Interact **15**(8): 774-9.
Espeso, E. A. and H. N. Arst, Jr. (2000). "On the mechanism by which alkaline pH prevents expression of an acid-expressed gene." Mol Cell Biol **20**(10): 3355-63.
Espeso, E. A. and M. A. Penalva (1996). "Three binding sites for the Aspergillus nidulans PacC zinc-finger transcription factor are necessary and sufficient for regulation by ambient pH of the isopenicillin N synthase gene promoter." J Biol Chem **271**(46): 28825-30.
Espeso, E. A., T. Roncal, et al. (2000). "On how a transcription factor can avoid its proteolytic activation in the absence of signal transduction." Embo J **19**(4): 719-28.
Espeso, E. A., J. Tilburn, et al. (1997). "Specific DNA recognition by the Aspergillus nidulans three zinc finger transcription factor PacC." J Mol Biol **274**(4): 466-80.
Felk, A., M. Kretschmar, et al. (2002). "Candida albicans hyphal formation and the expression of the Efg1-regulated proteinases Sap4 to Sap6 are required for the invasion of parenchymal organs." Infect Immun **70**(7): 3689-700.
Flaherty, J. E., A. M. Pirttila, et al. (2003). "PAC1, a pH-regulatory gene from Fusarium verticillioides." Appl Environ Microbiol **69**(9): 5222-7.
Fodor, S. P., J. L. Read, et al. (1991). "Light-directed, spatially addressable parallel chemical synthesis." Science **251**(4995): 767-73.
Fonzi, W. A. and M. Y. Irwin (1993). "Isogenic strain construction and gene mapping in *Candida albicans*." Genetics **134**(3): 717-28.
Forche, A., G. May, et al. (2005). "Demonstration of loss of heterozygosity by single-nucleotide polymorphism microarray analysis and alterations in strain morphology in Candida albicans strains during infection." Eukaryot Cell **4**(1): 156-65.
Fradin, C., P. De Groot, et al. (2005). "Granulocytes govern the transcriptional response, morphology and proliferation of Candida albicans in human blood." Mol Microbiol **56**(2): 397-415.
Fradin, C., M. Kretschmar, et al. (2003). "Stage-specific gene expression of Candida albicans in human blood." Mol Microbiol **47**(6): 1523-43.
Fu, Y., A. S. Ibrahim, et al. (2002). "Candida albicans Als1p: an adhesin that is a downstream effector of the EFG1 filamentation pathway." Mol Microbiol **44**(1): 61-72.
Fu, Y., G. Rieg, et al. (1998). "Expression of the Candida albicans gene ALS1 in Saccharomyces cerevisiae induces adherence to endothelial and epithelial cells." Infect Immun **66**(4): 1783-6.

Gaboriaud, C., V. Bissery, et al. (1987). "Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences." FEBS Lett **224**(1): 149-55.
Gale, C., D. Finkel, et al. (1996). "Cloning and expression of a gene encoding an integrin-like protein in Candida albicans." Proc Natl Acad Sci U S A **93**(1): 357-61.
Gale, C. A., C. M. Bendel, et al. (1998). "Linkage of adhesion, filamentous growth, and virulence in Candida albicans to a single gene, INT1." Science **279**(5355): 1355-8.
Garaizar, J., S. Brena, et al. (2006). "Use of DNA microarray technology and gene expression profiles to investigate the pathogenesis, cell biology, antifungal susceptibility and diagnosis of Candida albicans." FEMS Yeast Res **6**(7): 987-98.
Garcia-Sanchez, S., A. L. Mavor, et al. (2005). "Global roles of Ssn6 in Tup1- and Nrg1-dependent gene regulation in the fungal pathogen, Candida albicans." Mol Biol Cell **16**(6): 2913-25.
Gaur, N. K. and S. A. Klotz (1997). "Expression, cloning, and characterization of a Candida albicans gene, ALA1, that confers adherence properties upon Saccharomyces cerevisiae for extracellular matrix proteins." Infect Immun **65**(12): 5289-94.
Gauthier, C., S. Weber, et al. (2003). "Functional similarities and differences between Candida albicans Cdr1p and Cdr2p transporters." Antimicrob Agents Chemother **47**(5): 1543-54.
Gerami-Nejad, M., J. Berman, et al. (2001). "Cassettes for PCR-mediated construction of green, yellow, and cyan fluorescent protein fusions in Candida albicans." Yeast **18**(9): 859-64.
Gershon, D. (2005). "DNA microarrays: more than gene expression." Nature **437**(7062): 1195-8.
Ghannoum, M. A. (2000). "Potential role of phospholipases in virulence and fungal pathogenesis." Clin Microbiol Rev **13**(1): 122-43, table of contents.
Ghannoum, M. A., B. Spellberg, et al. (1995). "Reduced virulence of Candida albicans PHR1 mutants." Infect Immun **63**(11): 4528-30.
Gielkens, M., L. Gonzalez-Candelas, et al. (1999). "The abfB gene encoding the major alpha-L-arabinofuranosidase of Aspergillus nidulans: nucleotide sequence, regulation and construction of a disrupted strain." Microbiology **145 (Pt 3)**: 735-41.
Glick, M. and M. A. Siegel (1999). "Viral and fungal infections of the oral cavity in immunocompetent patients." Infect Dis Clin North Am **13**(4): 817-31, vi.
Gonzalez-Lopez, C. I., L. Ortiz-Castellanos, et al. (2006). "The ambient pH response Rim pathway in Yarrowia lipolytica: identification of YlRIM9 and characterization of its role in dimorphism." Curr Microbiol **53**(1): 8-12.
Gonzalez-Lopez, C. I., R. Szabo, et al. (2002). "Genetic control of extracellular protease synthesis in the yeast Yarrowia lipolytica." Genetics **160**(2): 417-27.
Gottlich, E., G. S. de Hoog, et al. (1995). "Cell-surface hydrophobicity and lipolysis as essential factors in human tinea nigra." Mycoses **38**(11-12): 489-94.
Gow, N. A., A. J. Brown, et al. (2002). "Fungal morphogenesis and host invasion." Curr Opin Microbiol **5**(4): 366-71.
Granjeaud, S., F. Bertucci, et al. (1999). "Expression profiling: DNA arrays in many guises." Bioessays **21**(9): 781-90.
Green, C. B., X. Zhao, et al. (2005). "Use of green fluorescent protein and reverse transcription-PCR to monitor Candida albicans agglutinin-like sequence gene

expression in a murine model of disseminated candidiasis." Infect Immun **73**(3): 1852-5.
Green, C. B., X. Zhao, et al. (2005). "Construction and real-time RT-PCR validation of Candida albicans PALS-GFP reporter strains and their use in flow cytometry analysis of ALS gene expression in budding and filamenting cells." Microbiology **151**(Pt 4): 1051-60.
Hayashi, M., T. Fukuzawa, et al. (2005). "Constitutive activation of the pH-responsive Rim101 pathway in yeast mutants defective in late steps of the MVB/ESCRT pathway." Mol Cell Biol **25**(21): 9478-90.
Herranz, S., J. M. Rodriguez, et al. (2005). "Arrestin-related proteins mediate pH signaling in fungi." Proc Natl Acad Sci U S A **102**(34): 12141-6.
Hospenthal, D. R., C. K. Murray, et al. (2004). "The role of antifungal susceptibility testing in the therapy of candidiasis." Diagn Microbiol Infect Dis **48**(3): 153-60.
Hoyer, L. L. (2001). "The ALS gene family of Candida albicans." Trends Microbiol **9**(4): 176-80.
Hoyer, L. L. and J. E. Hecht (2000). "The ALS6 and ALS7 genes of Candida albicans." Yeast **16**(9): 847-55.
Hoyer, L. L. and J. E. Hecht (2001). "The ALS5 gene of Candida albicans and analysis of the Als5p N-terminal domain." Yeast **18**(1): 49-60.
Hoyer, L. L., T. L. Payne, et al. (1998). "Candida albicans ALS3 and insights into the nature of the ALS gene family." Curr Genet **33**(6): 451-9.
Hoyer, L. L., T. L. Payne, et al. (1998). "Identification of Candida albicans ALS2 and ALS4 and localization of als proteins to the fungal cell surface." J Bacteriol **180**(20): 5334-43.
Hoyer, L. L., S. Scherer, et al. (1995). "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual agglutinin separated by a repeating motif." Mol Microbiol **15**(1): 39-54.
Hube, B. and J. Naglik (2001). "Candida albicans proteinases: resolving the mystery of a gene family." Microbiology **147**(Pt 8): 1997-2005.
Hube, B., F. Stehr, et al. (2000). "Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members." Arch Microbiol **174**(5): 362-74.
Hutchings, H., K. P. Stahmann, et al. (1999). "The multiply-regulated gabA gene encoding the GABA permease of Aspergillus nidulans: a score of exons." Mol Microbiol **32**(3): 557-68.
Ibrahim, A. S., F. Mirbod, et al. (1995). "Evidence implicating phospholipase as a virulence factor of Candida albicans." Infect Immun **63**(5): 1993-8.
Ibrahim, A. S., B. J. Spellberg, et al. (2005). "Vaccination with recombinant N-terminal domain of Als1p improves survival during murine disseminated candidiasis by enhancing cell-mediated, not humoral, immunity." Infect Immun **73**(2): 999-1005.
Jones, T., N. A. Federspiel, et al. (2004). "The diploid genome sequence of Candida albicans." Proc Natl Acad Sci U S A **101**(19): 7329-34.
Kadir, T., B. Gumru, et al. (2007). "Phospholipase activity of Candida albicans isolates from patients with denture stomatitis: The influence of chlorhexidine gluconate on phospholipase production." Arch Oral Biol.
Kafatos, F. C., C. W. Jones, et al. (1979). "Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure." Nucleic Acids Res **7**(6): 1541-52.

Kamai, Y., M. Kubota, et al. (2002). "Contribution of Candida albicans ALS1 to the pathogenesis of experimental oropharyngeal candidiasis." Infect Immun **70**(9): 5256-8.
Kamai, Y., M. Kubota, et al. (2002). "Contribution of Candida albicans ALS1 to the pathogenesis of experimental oropharyngeal candidiasis." Infect Immun **70**(9): 5256-8.
Kapteyn, J. C., L. L. Hoyer, et al. (2000). "The cell wall architecture of Candida albicans wild-type cells and cell wall-defective mutants." Mol Microbiol **35**(3): 601-11.

Kapteyn, J. C., R. C. Montijn, et al. (1995). "Covalent association of beta-1,3-glucan with beta-1,6-glucosylated mannoproteins in cell walls of Candida albicans." J Bacteriol **177**(13): 3788-92.

Kapteyn, J. C., R. C. Montijn, et al. (1996). "Retention of Saccharomyces cerevisiae cell wall proteins through a phosphodiester-linked beta-1,3-/beta-1,6-glucan heteropolymer." Glycobiology **6**(3): 337-45.

Kapteyn, J. C., A. F. Ram, et al. (1997). "Altered extent of cross-linking of beta1,6-glucosylated mannoproteins to chitin in Saccharomyces cerevisiae mutants with reduced cell wall beta1,3-glucan content." J Bacteriol **179**(20): 6279-84.

Kapteyn, J. C., H. Van Den Ende, et al. (1999). "The contribution of cell wall proteins to the organization of the yeast cell wall." Biochim Biophys Acta **1426**(2): 373-83.

Katz, M. E., A. Masoumi, et al. (2000). "The Aspergillus nidulans xprF gene encodes a hexokinase-like protein involved in the regulation of extracellular proteases." Genetics **156**(4): 1559-71.

Kerscher, S., S. Drose, et al. (2002). "Yarrowia lipolytica, a yeast genetic system to study mitochondrial complex I." Biochim Biophys Acta **1555**(1-3): 83-91.

Khodursky, A. B., J. A. Bernstein, et al. (2003). "Escherichia coli spotted double-strand DNA microarrays: RNA extraction, labeling, hybridization, quality control, and data management." Methods Mol Biol **224**: 61-78.

Kiernan, J. A. (2000). "Formaldehyde, formalin, paraformaldehyde and glutaraldehyde:
What they are and what they do. ." Microscopy Today (00-1): 8-12

Klis, F. M., P. de Groot, et al. (2001). "Molecular organization of the cell wall of Candida albicans." Med Mycol **39 Suppl 1**: 1-8.

Klotz, S. A., M. L. Pendrak, et al. (2001). "Antibodies to alpha5beta1 and alpha(v)beta3 integrins react with Candida albicans alcohol dehydrogenase." Microbiology **147**(Pt 11): 3159-64.

Kobayashi, S. D. and J. E. Cutler (1998). "Candida albicans hyphal formation and virulence: is there a clearly defined role?" Trends Microbiol **6**(3): 92-4.

Kollar, R., B. B. Reinhold, et al. (1997). "Architecture of the yeast cell wall. Beta(1-->6)-glucan interconnects mannoprotein, beta(1-->)3-glucan, and chitin." J Biol Chem **272**(28): 17762-75.

Konig, B., K. E. Jaeger, et al. (1996). "Role of Pseudomonas aeruginosa lipase in inflammatory mediator release from human inflammatory effector cells (platelets, granulocytes, and monocytes." Infect Immun **64**(8): 3252-8.

Kretschmar, M., B. Hube, et al. (1999). "Germ tubes and proteinase activity contribute to virulence of Candida albicans in murine peritonitis." Infect Immun **67**(12): 6637-42.

Kucho, K., K. Okamoto, et al. (2005). "Global analysis of circadian expression in the cyanobacterium Synechocystis sp. strain PCC 6803." J Bacteriol **187**(6): 2190-9.

Kumamoto, C. A. and M. D. Vinces (2005). "Contributions of hyphae and hypha-co-regulated genes to Candida albicans virulence." Cell Microbiol **7**(11): 1546-54.
Kumar, C. P., S. S. Kumar, et al. (2006). "Phospholipase and proteinase activities of clinical isolates of Candida from immunocompromised patients." Mycopathologia **161**(4): 213-8.
Kunze, D., I. Melzer, et al. (2005). "Functional analysis of the phospholipase C gene CaPLC1 and two unusual phospholipase C genes, CaPLC2 and CaPLC3, of Candida albicans." Microbiology **151**(Pt 10): 3381-94.
Kuras, L. and K. Struhl (1999). "Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme." Nature **399**(6736): 609-13.
Kurtz, M. B., M. W. Cortelyou, et al. (1986). "Integrative transformation of Candida albicans, using a cloned Candida ADE2 gene." Mol Cell Biol **6**(1): 142-9.
Kvaal, C., S. A. Lachke, et al. (1999). "Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of Candida albicans confers increased virulence in a mouse model of cutaneous infection." Infect Immun **67**(12): 6652-62.
Lamb, T. M. and A. P. Mitchell (2003). "The transcription factor Rim101p governs ion tolerance and cell differentiation by direct repression of the regulatory genes NRG1 and SMP1 in Saccharomyces cerevisiae." Mol Cell Biol **23**(2): 677-86.
Lamb, T. M., W. Xu, et al. (2001). "Alkaline response genes of Saccharomyces cerevisiae and their relationship to the RIM101 pathway." J Biol Chem **276**(3): 1850-6.
Lambert, M., S. Blanchin-Roland, et al. (1997). "Genetic analysis of regulatory mutants affecting synthesis of extracellular proteinases in the yeast Yarrowia lipolytica: identification of a RIM101/pacC homolog." Mol Cell Biol **17**(7): 3966-76.
Laub, M. T., H. H. McAdams, et al. (2000). "Global analysis of the genetic network controlling a bacterial cell cycle." Science **290**(5499): 2144-8.
Le Crom, S., F. Devaux, et al. (2002). "New insights into the pleiotropic drug resistance network from genome-wide characterization of the YRR1 transcription factor regulation system." Mol Cell Biol **22**(8): 2642-9.
Leidich, S. D., A. S. Ibrahim, et al. (1998). "Cloning and disruption of caPLB1, a phospholipase B gene involved in the pathogenicity of Candida albicans." J Biol Chem **273**(40): 26078-86.
Lephart, P. R., H. Chibana, et al. (2005). "Effect of the major repeat sequence on chromosome loss in Candida albicans." Eukaryot Cell **4**(4): 733-41.
Leuker, C. E., A. M. Hahn, et al. (1992). "beta-Galactosidase of Kluyveromyces lactis (Lac4p) as reporter of gene expression in Candida albicans and C. tropicalis." Mol Gen Genet **235**(2-3): 235-41.
Leuker, C. E., A. Sonneborn, et al. (1997). "Sequence and promoter regulation of the PCK1 gene encoding phosphoenolpyruvate carboxykinase of the fungal pathogen Candida albicans." Gene **192**(2): 235-40.
Li, M., S. J. Martin, et al. (2004). "Candida albicans Rim13p, a protease required for Rim101p processing at acidic and alkaline pHs." Eukaryot Cell **3**(3): 741-51.
Li, W. and A. P. Mitchell (1997). "Proteolytic activation of Rim1p, a positive regulator of yeast sporulation and invasive growth." Genetics **145**(1): 63-73.
Lipshutz, R. J., S. P. Fodor, et al. (1999). "High density synthetic oligonucleotide arrays." Nat Genet **21**(1 Suppl): 20-4.
Liu, H. (2001). "Transcriptional control of dimorphism in Candida albicans." Curr Opin Microbiol **4**(6): 728-35.

Liu, H. (2002). "Co-regulation of pathogenesis with dimorphism and phenotypic switching in Candida albicans, a commensal and a pathogen." Int J Med Microbiol **292**(5-6): 299-311.
Lorenz, M. C., J. A. Bender, et al. (2004). "Transcriptional response of Candida albicans upon internalization by macrophages." Eukaryot Cell **3**(5): 1076-87.
Lott, T. J., B. P. Holloway, et al. (1999). "Towards understanding the evolution of the human commensal yeast Candida albicans." Microbiology **145 (Pt 5)**: 1137-43.
Lotz, H., K. Sohn, et al. (2004). "RBR1, a novel pH-regulated cell wall gene of Candida albicans, is repressed by RIM101 and activated by NRG1." Eukaryot Cell **3**(3): 776-84.
Loza, L., Y. Fu, et al. (2004). "Functional analysis of the Candida albicans ALS1 gene product." Yeast **21**(6): 473-82.
MacCabe, A. P., M. Orejas, et al. (1998). "Opposite patterns of expression of two Aspergillus nidulans xylanase genes with respect to ambient pH." J Bacteriol **180**(5): 1331-3.
Macphail, G. L., G. D. Taylor, et al. (2002). "Epidemiology, treatment and outcome of candidemia: a five-year review at three Canadian hospitals." Mycoses **45**(5-6): 141-5.
Madzak, C., S. Blanchin-Roland, et al. (1999). "Functional analysis of upstream regulating regions from the Yarrowia lipolytica XPR2 promoter." Microbiology **145 (Pt 1)**: 75-87.
Madzak, C., C. Gaillardin, et al. (2004). "Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: a review." J Biotechnol **109**(1-2): 63-81.
Magee, P. T., C. Gale, et al. (2003). "Molecular genetic and genomic approaches to the study of medically important fungi." Infect Immun **71**(5): 2299-309.
Martchenko, M., A. M. Alarco, et al. (2004). "Superoxide dismutases in Candida albicans: transcriptional regulation and functional characterization of the hyphal-induced SOD5 gene." Mol Biol Cell **15**(2): 456-67.
Martinez-Lopez, R., H. Park, et al. (2006). "Candida albicans Ecm33p is important for normal cell wall architecture and interactions with host cells." Eukaryot Cell **5**(1): 140-7.
Matsuo, H., J. Chevallier, et al. (2004). "Role of LBPA and Alix in multivesicular liposome formation and endosome organization." Science **303**(5657): 531-4.
Michel, S., S. Ushinsky, et al. (2002). "Generation of conditional lethal Candida albicans mutants by inducible deletion of essential genes." Mol Microbiol **46**(1): 269-80.
Mingot, J. M., E. A. Espeso, et al. (2001). "Ambient pH signaling regulates nuclear localization of the Aspergillus nidulans PacC transcription factor." Mol Cell Biol **21**(5): 1688-99.
Mingot, J. M., J. Tilburn, et al. (1999). "Specificity determinants of proteolytic processing of Aspergillus PacC transcription factor are remote from the processing site, and processing occurs in yeast if pH signalling is bypassed." Mol Cell Biol **19**(2): 1390-400.
Morschhauser, J., S. Michel, et al. (1998). "Expression of a chromosomally integrated, single-copy GFP gene in Candida albicans, and its use as a reporter of gene regulation." Mol Gen Genet **257**(4): 412-20.
Mulhern, S. M., M. E. Logue, et al. (2006). "Candida albicans transcription factor Ace2 regulates metabolism and is required for filamentation in hypoxic conditions." Eukaryot Cell **5**(12): 2001-13.

Murad, A. M., C. d'Enfert, et al. (2001). "Transcript profiling in Candida albicans reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1." Mol Microbiol **42**(4): 981-93.
Myers, K. K., P. S. Sypherd, et al. (1995). "Use of URA3 as a reporter of gene expression in C. albicans." Curr Genet **27**(3): 243-8.
Naglik, J. R., G. Newport, et al. (1999). "In vivo analysis of secreted aspartyl proteinase expression in human oral candidiasis." Infect Immun **67**(5): 2482-90.
Nakagawa, Y., T. Kanbe, et al. (2003). "Disruption of the human pathogenic yeast Candida albicans catalase gene decreases survival in mouse-model infection and elevates susceptibility to higher temperature and to detergents." Microbiol Immunol **47**(6): 395-403.
Nantel, A. (2006). "The long hard road to a completed Candida albicans genome." Fungal Genet Biol **43**(5): 311-5.
Navarro-Garcia, F., M. Sanchez, et al. (2001). "Virulence genes in the pathogenic yeast Candida albicans." FEMS Microbiol Rev **25**(2): 245-68.
Negrete-Urtasun, S., W. Reiter, et al. (1999). "Ambient pH signal transduction in Aspergillus: completion of gene characterization." Mol Microbiol **33**(5): 994-1003.
Newton-John, H. F., K. Wise, et al. (1984). "Role of the lemon in disseminated candidiasis of heroin abusers." Med J Aust **140**(13): 780-1.
Nguyen, C., D. Rocha, et al. (1995). "Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones." Genomics **29**(1): 207-16.
Niewerth, M. and H. C. Korting (2001). "Phospholipases of Candida albicans." Mycoses **44**(9-10): 361-7.
Nobile, C. J., D. R. Andes, et al. (2006). "Critical role of Bcr1-dependent adhesins in C. albicans biofilm formation in vitro and in vivo." PLoS Pathog **2**(7): e63.
Nobile, C. J., V. M. Bruno, et al. (2003). "Genetic control of chlamydospore formation in Candida albicans." Microbiology **149**(Pt 12): 3629-37.
Nozawa, S. R., G. S. May, et al. (2003). "Mutation in a calpain-like protease affects the posttranslational mannosylation of phosphatases in Aspergillus nidulans." Fungal Genet Biol **38**(2): 220-7.
Odds, F. (1988). Candida and candidosis.
Odds, F. C., A. J. Brown, et al. (2004). "Candida albicans genome sequence: a platform for genomics in the absence of genetics." Genome Biol **5**(7): 230.
Odds, F. C., A. J. Brown, et al. (2006). "Toward a molecular understanding of Candida albicans virulence." Chapter 22 in "Molecular Principles of Fungal Pathogenesis" (Eds. Heitman, J., Filler, S., Mitchell, A.) ASM Press.
Oh, S. H., G. Cheng, et al. (2005). "Functional specificity of Candida albicans Als3p proteins and clade specificity of ALS3 alleles discriminated by the number of copies of the tandem repeat sequence in the central domain." Microbiology **151**(Pt 3): 673-81.
Orejas, M., E. A. Espeso, et al. (1995). "Activation of the Aspergillus PacC transcription factor in response to alkaline ambient pH requires proteolysis of the carboxy-terminal moiety." Genes Dev **9**(13): 1622-32.
Pardini, G., P. W. De Groot, et al. (2006). "The CRH family coding for cell wall glycosylphosphatidylinositol proteins with a predicted transglycosidase domain affects cell wall organization and virulence of Candida albicans." J Biol Chem **281**(52): 40399-411.
Park, Y. N. and J. Morschhauser (2005). "Tetracycline-inducible gene expression and gene deletion in Candida albicans." Eukaryot Cell **4**(8): 1328-42.

Penalva, M. A. and H. N. Arst, Jr. (2002). "Regulation of gene expression by ambient pH in filamentous fungi and yeasts." Microbiol Mol Biol Rev **66**(3): 426-46, table of contents.
Penalva, M. A. and H. N. Arst, Jr. (2004). "Recent advances in the characterization of ambient pH regulation of gene expression in filamentous fungi and yeasts." Annu Rev Microbiol **58**: 425-51.
Perez-Gonzalez, J. A., N. N. van Peij, et al. (1998). "Molecular cloning and transcriptional regulation of the Aspergillus nidulans xlnD gene encoding a beta-xylosidase." Appl Environ Microbiol **64**(4): 1412-9.
Phan, Q. T., P. H. Belanger, et al. (2000). "Role of hyphal formation in interactions of Candida albicans with endothelial cells." Infect Immun **68**(6): 3485-90.
Phan, Q. T., C. L. Myers, et al. (2007). "Als3 Is a Candida albicans Invasin That Binds to Cadherins and Induces Endocytosis by Host Cells." PLoS Biol **5**(3): e64.
Platara, M., A. Ruiz, et al. (2006). "The transcriptional response of the yeast Na(+)-ATPase ENA1 gene to alkaline stress involves three main signaling pathways." J Biol Chem **281**(48): 36632-42.
Porta, A., A. M. Ramon, et al. (1999). "PRR1, a homolog of Aspergillus nidulans palF, controls pH-dependent gene expression and filamentation in Candida albicans." J Bacteriol **181**(24): 7516-23.
Porta, A., Z. Wang, et al. (2001). "Spontaneous second-site suppressors of the filamentation defect of prr1Delta mutants define a critical domain of Rim101p in Candida albicans." Mol Genet Genomics **266**(4): 624-31.
Poulain, D. and T. Jouault (2004). "Candida albicans cell wall glycans, host receptors and responses: elements for a decisive crosstalk." Curr Opin Microbiol **7**(4): 342-9.
Prusky, D., J. L. McEvoy, et al. (2001). "Local modulation of host pH by Colletotrichum species as a mechanism to increase virulence." Mol Plant Microbe Interact **14**(9): 1105-13.
Pugh, D. and R. A. Cawson (1977). "The cytochemical localization of phospholipase in Candida albicans infecting the chick chorio-allantoic membrane." Sabouraudia **15**(1): 29-35.
Quackenbush, J. (2001). "Computational analysis of microarray data." Nat Rev Genet **2**(6): 418-27.
Quackenbush, J. (2002). "Microarray data normalization and transformation." Nat Genet **32 Suppl**: 496-501.
Querol, A., E. Barrio, et al. (1992). "Molecular Monitoring of Wine Fermentations Conducted by Active Dry Yeast Strains." Appl Environ Microbiol **58**(9): 2948-2953.
Ramage, G., J. P. Martinez, et al. (2006). "Candida biofilms on implanted biomaterials: a clinically significant problem." FEMS Yeast Res **6**(7): 979-86.
Ramon, A. M. and W. A. Fonzi (2003). "Diverged binding specificity of Rim101p, the Candida albicans ortholog of PacC." Eukaryot Cell **2**(4): 718-28.
Ramon, A. M., A. Porta, et al. (1999). "Effect of environmental pH on morphological development of Candida albicans is mediated via the PacC-related transcription factor encoded by PRR2." J Bacteriol **181**(24): 7524-30.
Reef, S. E., B. A. Lasker, et al. (1998). "Nonperinatal nosocomial transmission of Candida albicans in a neonatal intensive care unit: prospective study." J Clin Microbiol **36**(5): 1255-9.
Richard, M., P. De Groot, et al. (2002). "GPI7 affects cell-wall protein anchorage in Saccharomyces cerevisiae and Candida albicans." Microbiology **148**(Pt 7): 2125-33.

Richard, M. L. and A. Plaine (2007). "Comprehensive Analysis of Glycosylphosphatidylinositol-Anchored Proteins in Candida albicans." Eukaryot Cell **6**(2): 119-33.
Rollins, J. A. (2003). "The Sclerotinia sclerotiorum pac1 gene is required for sclerotial development and virulence." Mol Plant Microbe Interact **16**(9): 785-95.
Rothfels, K., J. C. Tanny, et al. (2005). "Components of the ESCRT pathway, DFG16, and YGR122w are required for Rim101 to act as a corepressor with Nrg1 at the negative regulatory element of the DIT1 gene of Saccharomyces cerevisiae." Mol Cell Biol **25**(15): 6772-88.
Roustan, J. L., A. R. Chu, et al. (2005). "A novel lipase/acyltransferase from the yeast Candida albicans: expression and characterisation of the recombinant enzyme." Appl Microbiol Biotechnol **68**(2): 203-12.
Ruiz-Herrera, J., M. V. Elorza, et al. (2006). "Molecular organization of the cell wall of Candida albicans and its relation to pathogenicity." FEMS Yeast Res **6**(1): 14-29.
Ruiz-Herrera, J. and G. San-Blas (2003). "Chitin synthesis as target for antifungal drugs." Curr Drug Targets Infect Disord **3**(1): 77-91.
Samaranayake, L. P., J. M. Raeside, et al. (1984). "Factors affecting the phospholipase activity of Candida species in vitro." Sabouraudia **22**(3): 201-7.
Sanglard, D., F. Ischer, et al. (1996). "Susceptibilities of Candida albicans multidrug transporter mutants to various antifungal agents and other metabolic inhibitors." Antimicrob Agents Chemother **40**(10): 2300-5.
Sanglard, D., F. Ischer, et al. (2003). "Candida albicans mutations in the ergosterol biosynthetic pathway and resistance to several antifungal agents." Antimicrob Agents Chemother **47**(8): 2404-12.
Santos, M. A. and M. F. Tuite (1995). "The CUG codon is decoded in vivo as serine and not leucine in Candida albicans." Nucleic Acids Res **23**(9): 1481-6.
Schaller, M., M. Bein, et al. (2003). "The secreted aspartyl proteinases Sap1 and Sap2 cause tissue damage in an in vitro model of vaginal candidiasis based on reconstituted human vaginal epithelium." Infect Immun **71**(6): 3227-34.
Schaller, M., B. Hube, et al. (1999). "In vivo expression and localization of Candida albicans secreted aspartyl proteinases during oral candidiasis in HIV-infected patients." J Invest Dermatol **112**(3): 383-6.
Schaller, M., W. Schafer, et al. (1998). "Differential expression of secreted aspartyl proteinases in a model of human oral candidosis and in patient samples from the oral cavity." Mol Microbiol **29**(2): 605-15.
Schauer, F. and R. Hanschke (1999). "[Taxonomy and ecology of the genus Candida]." Mycoses **42 Suppl 1**: 12-21.
Schelenz, S. and W. R. Gransden (2003). "Candidaemia in a London teaching hospital: analysis of 128 cases over a 7-year period." Mycoses **46**(9-10): 390-6.
Scherwitz, C. (1982). "Ultrastructure of human cutaneous candidosis." J Invest Dermatol **78**(3): 200-5.
Schmitt, E. K., R. Kempken, et al. (2001). "Functional analysis of promoter sequences of cephalosporin C biosynthesis genes from Acremonium chrysogenum: specific DNA-protein interactions and characterization of the transcription factor PACC." Mol Genet Genomics **265**(3): 508-18.
Schofield, D. A., C. Westwater, et al. (2005). "Differential Candida albicans lipase gene expression during alimentary tract colonization and infection." FEMS Microbiol Lett **244**(2): 359-65.

Sentandreu, M., M. V. Elorza, et al. (1998). "Cloning and characterization of PRA1, a gene encoding a novel pH-regulated antigen of Candida albicans." J Bacteriol **180**(2): 282-9.
Setiadi, E. R., T. Doedt, et al. (2006). "Transcriptional response of Candida albicans to hypoxia: linkage of oxygen sensing and Efg1p-regulatory networks." J Mol Biol **361**(3): 399-411.
Sheppard, D. C., M. R. Yeaman, et al. (2004). "Functional and structural diversity in the Als protein family of Candida albicans." J Biol Chem **279**(29): 30480-9.
Singh, N. (2001). "Changing spectrum of invasive candidiasis and its therapeutic implications." Clin Microbiol Infect **7 Suppl 2**: 1-7.
Sobel, J. D. (1992). "Vulvovaginitis." Dermatol Clin **10**(2): 339-59.
Sohn, K., I. Senyurek, et al. (2006). "An in vitro assay to study the transcriptional response during adherence of Candida albicans to different human epithelia." FEMS Yeast Res **6**(7): 1085-93.
Sohn, K., C. Urban, et al. (2003). "EFG1 is a major regulator of cell wall dynamics in Candida albicans as revealed by DNA microarrays." Mol Microbiol **47**(1): 89-102.
Soll, D. R., S. R. Lockhart, et al. (2003). "Relationship between switching and mating in Candida albicans." Eukaryot Cell **2**(3): 390-7.
Soloviev, D. A., W. A. Fonzi, et al. (2007). "Identification of pH-regulated antigen 1 released from Candida albicans as the major ligand for leukocyte integrin alphaMbeta2." J Immunol **178**(4): 2038-46.
Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." J Mol Biol **98**(3): 503-17.
Spellberg, B. J., A. S. Ibrahim, et al. (2005). "The anti-Candida albicans vaccine composed of the recombinant N terminus of Als1p reduces fungal burden and improves survival in both immunocompetent and immunocompromised mice." Infect Immun **73**(9): 6191-3.
Spreghini, E., D. A. Davis, et al. (2003). "Roles of Candida albicans Dfg5p and Dcw1p cell surface proteins in growth and hypha formation." Eukaryot Cell **2**(4): 746-55.
Srikantha, T., A. R. Borneman, et al. (2006). "TOS9 regulates white-opaque switching in Candida albicans." Eukaryot Cell **5**(10): 1674-87.
Srikantha, T., A. Klapach, et al. (1996). "The sea pansy Renilla reniformis luciferase serves as a sensitive bioluminescent reporter for differential gene expression in Candida albicans." J Bacteriol **178**(1): 121-9.
Staab, J. F., S. D. Bradway, et al. (1999). "Adhesive and mammalian transglutaminase substrate properties of Candida albicans Hwp1." Science **283**(5407): 1535-8.
Staab, J. F., C. A. Ferrer, et al. (1996). "Developmental expression of a tandemly repeated, proline-and glutamine-rich amino acid motif on hyphal surfaces on Candida albicans." J Biol Chem **271**(11): 6298-305.
Staib, P., S. Michel, et al. (2000). "A molecular genetic system for the pathogenic yeast Candida dubliniensis." Gene **242**(1-2): 393-8.
Stehr, F., A. Felk, et al. (2004). "Expression analysis of the Candida albicans lipase gene family during experimental infections and in patient samples." FEMS Yeast Res **4**(4-5): 401-8.
Stoldt, V. R., A. Sonneborn, et al. (1997). "Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi." Embo J **16**(8): 1982-91.

Stuart, J. M., E. Segal, et al. (2003). "A gene-coexpression network for global discovery of conserved genetic modules." Science **302**(5643): 249-55.
Su, S. S. and A. P. Mitchell (1993). "Identification of functionally related genes that stimulate early meiotic gene expression in yeast." Genetics **133**(1): 67-77.
Suarez, T. and M. A. Penalva (1996). "Characterization of a Penicillium chrysogenum gene encoding a PacC transcription factor and its binding sites in the divergent pcbAB-pcbC promoter of the penicillin biosynthetic cluster." Mol Microbiol **20**(3): 529-40.
Sudbery, P., N. Gow, et al. (2004). "The distinct morphogenic states of Candida albicans." Trends Microbiol **12**(7): 317-24.
Sundstrom, P. (2002). "Adhesion in Candida spp." Cell Microbiol **4**(8): 461-9.
Sundstrom, P., J. E. Cutler, et al. (2002). "Reevaluation of the role of HWP1 in systemic candidiasis by use of Candida albicans strains with selectable marker URA3 targeted to the ENO1 locus." Infect Immun **70**(6): 3281-3.
Surarit, R., P. K. Gopal, et al. (1988). "Evidence for a glycosidic linkage between chitin and glucan in the cell wall of Candida albicans." J Gen Microbiol **134**(6): 1723-30.
Tanghe, A., J. M. Carbrey, et al. (2005). "Aquaporin expression and freeze tolerance in Candida albicans." Appl Environ Microbiol **71**(10): 6434-7.
Then Bergh, K. and A. A. Brakhage (1998). "Regulation of the Aspergillus nidulans penicillin biosynthesis gene acvA (pcbAB) by amino acids: implication for involvement of transcription factor PACC." Appl Environ Microbiol **64**(3): 843-9.
Tilburn, J., J. C. Sanchez-Ferrero, et al. (2005). "Mutational analysis of the pH signal transduction component PalC of Aspergillus nidulans supports distant similarity to BRO1 domain family members." Genetics **171**(1): 393-401.
Tilburn, J., S. Sarkar, et al. (1995). "The Aspergillus PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH." Embo J **14**(4): 779-90.
Tokunaga, M., M. Kusamichi, et al. (1986). "Ultrastructure of outermost layer of cell wall in Candida albicans observed by rapid-freezing technique." J Electron Microsc (Tokyo) **35**(3): 237-46.
Trick, W. E., S. K. Fridkin, et al. (2002). "Secular trend of hospital-acquired candidemia among intensive care unit patients in the United States during 1989-1999." Clin Infect Dis **35**(5): 627-30.
Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-21.
Uhl, M. A., M. Biery, et al. (2003). "Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans." Embo J **22**(11): 2668-78.
Uhl, M. A. and A. D. Johnson (2001). "Development of Streptococcus thermophilus lacZ as a reporter gene for Candida albicans." Microbiology **147**(Pt 5): 1189-95.
Umeyama, T., Y. Nagai, et al. (2002). "Construction of FLAG tagging vectors for Candida albicans." Yeast **19**(7): 611-8.
Van Der Vaart, J. M., R. te Biesebeke, et al. (1996). "The beta-1, 6-glucan containing side-chain of cell wall proteins of Saccharomyces cerevisiae is bound to the glycan core of the GPI moiety." FEMS Microbiol Lett **145**(3): 401-7.
vanKuyk, P. A., B. F. Cheetham, et al. (2000). "Analysis of two Aspergillus nidulans genes encoding extracellular proteases." Fungal Genet Biol **29**(3): 201-10.

Villar, C. C., H. Kashleva, et al. (2007). "Mucosal tissue invasion by Candida albicans is associated with E-cadherin degradation, mediated by transcription factor Rim101p and protease Sap5p." Infect Immun.

Vincent, O., L. Rainbow, et al. (2003). "YPXL/I is a protein interaction motif recognized by aspergillus PalA and its human homologue, AIP1/Alix." Mol Cell Biol **23**(5): 1647-55.

Wey, S. B., M. Mori, et al. (1989). "Risk factors for hospital-acquired candidemia. A matched case-control study." Arch Intern Med **149**(10): 2349-53.

White, T. C., S. Holleman, et al. (2002). "Resistance mechanisms in clinical isolates of Candida albicans." Antimicrob Agents Chemother **46**(6): 1704-13.

Wilson, R. B., D. Davis, et al. (1999). "Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions." J Bacteriol **181**(6): 1868-74.

Wisplinghoff, H., H. Seifert, et al. (2003). "Current trends in the epidemiology of nosocomial bloodstream infections in patients with hematological malignancies and solid neoplasms in hospitals in the United States." Clin Infect Dis **36**(9): 1103-10.

Wolf, D. G., I. Polacheck, et al. (2000). "High rate of candidemia in patients sustaining injuries in a bomb blast at a marketplace: a possible environmental source." Clin Infect Dis **31**(3): 712-6.

Xu, W. and A. P. Mitchell (2001). "Yeast PalA/AIP1/Alix homolog Rim20p associates with a PEST-like region and is required for its proteolytic cleavage." J Bacteriol **183**(23): 6917-23.

Xu, W., F. J. Smith, Jr., et al. (2004). "Multivesicular body-ESCRT components function in pH response regulation in Saccharomyces cerevisiae and Candida albicans." Mol Biol Cell **15**(12): 5528-37.

Yang, I. V., E. Chen, et al. (2002). "Within the fold: assessing differential expression measures and reproducibility in microarray assays." Genome Biol **3**(11): research0062.

Yang, Y. H., S. Dudoit, et al. (2002). "Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation." Nucleic Acids Res **30**(4): e15.

Yang, Y. H. and T. Speed (2002). "Design issues for cDNA microarray experiments." Nat Rev Genet **3**(8): 579-88.

Zhang, N., A. L. Harrex, et al. (2003). "Sixty alleles of the ALS7 open reading frame in Candida albicans: ALS7 is a hypermutable contingency locus." Genome Res **13**(9): 2005-17.

Zhao, X., K. J. Daniels, et al. (2006). "Candida albicans Als3p is required for wild-type biofilm formation on silicone elastomer surfaces." Microbiology **152**(Pt 8): 2287-99.

Zhao, X., S. H. Oh, et al. (2004). "ALS3 and ALS8 represent a single locus that encodes a Candida albicans adhesin; functional comparisons between Als3p and Als1p." Microbiology **150**(Pt 7): 2415-28.

Zhao, X., S. H. Oh, et al. (2005). "Analysis of the Candida albicans Als2p and Als4p adhesins suggests the potential for compensatory function within the Als family." Microbiology **151**(Pt 5): 1619-30.

Summary

Rim101p is a conserved fungal transcription factor that becomes activated through C-terminal cleavage under neutral to alkaline conditions. The identification and analysis of Rim101p targets in *Candida albicans* was the main subject of the PhD thesis.

A constitutively active truncated version of Rim101p (Rim101SLp) was introduced under the control of the *MET3*-promotor into a *rim101* null mutant to monitor Rim101-dependent transcriptional changes independently of other pH-dependent regulatory events. Transcriptional changes were recorded using microarrays along a time course following induction of *RIM101SL* transcription.

After filtering the data, the transcriptional patterns of 133 selected genes was clustered into five distinct classes. Significantly more putative Rim101p binding sites were detected in the promoters of these genes than in a randomly chosen set of genes. Further analysis permitted to identify a putative extended Rim101p binding motif. Putative Rim101p targets were examined for predicted functions and amino acid landmarks like transmembrane domains and signal peptides that could indicate localization at the cell surface and thus a possible involvement in host interaction.

Microarray results were confirmed on 20 selected genes by quantitative real-time PCR. Furthermore, the relevance of the microarray data for the pH response of *C. albicans* was assessed by monitoring transcriptional changes of these genes in a wild type strain grown at pH 4 or pH 8. In spite of these experimental setup differences, a clear correlation of the results was observed for a large majority of the tested genes.

Microarray data suggested that Rim101p activity had a strong impact on the expression of genes of the *ALS* (Agglutinin-Like Sequence) gene family. The extremely high sequence conservation within this family hampered however a gene-specific analysis. Using a gene-specific primer set, transcription of each member of the *ALS* gene family was analyzed by real-time qPCR. Four *ALS* genes were shown to be transcribed in a pH-dependent manner, and Rim101p was found to be required for the alkaline induction of *ALS1* and the repression of *ALS4*. The two other genes, *ALS2* and *ALS9*, were also repressed at alkaline pH, but their regulation was at least partially independent of Rim101p.

Finally, the mechanism of *ALS1* and *ALS4* regulation was addressed by two different approaches. First, reporter strains that put a modified bacterial β-galactosidase gene under the control of *ALS* promoters were constructed in order to monitor more easily pH and Rim101p effects on *ALS1* and *ALS4* expression. Second, a tagged version of Rim101p was used to demonstrate *in vivo* binding of Rim101p to *ALS* promoters by Chromatin Immunoprecipitation (ChIP): however, no clear specific binding could be observed.

Résumé

Rim101p est un facteur de transcription qui est activé par cleavage N-terminale à pH alcalin. Il régule ainsi la réponse au pH et joue aussi un rôle majeur dans la pathogenèse de la levure *Candida albicans*. Mon projet est d'identifier et d'étudier des gènes de surface régulés par Rim101p qui ont une fonction dans l'interaction hôte-levure.

Une analyse du transcriptome suite à l'induction d'une forme tronquée et constitutivement active de Rim101p nous avait permis d'identifier et de classer 133 gènes qui semblent être des cibles de Rim101p. Les données des microarrays ont été confirmées pour 20 gènes par PCR quantitative, en utilisant des conditions plus physiologiques.

Plusieurs adhésines de la famille des gènes *ALS* (Agglutinin-Like-Sequence) qui comporte 8 gènes semblent être régulé par Rim101p. Certaines jouent un rôle important dans la formation des biofilms et ainsi dans la virulence de *C.albicans*. Malgré la grande ressemblance des gènes au niveau de leur séquence nous avons pu confirmer le rôle de Rim101p dans la régulation d'au moins deux gènes *ALS* (*ALS1* et *ALS4*) dans une analyse de la famille entière en PCR quantitative avec des oligos gène-spécifiques.

Pour analyser la régulation de ces gènes au niveau de leur promoteur, des fusions avec le gène rapporteur « LacZ » ont été intégré dans le génome de *C. albicans* et l'activité de la β-Galactosidase a été quantifiée en fonction du pH et de la présence ou absence de Rim101p. Nous avons abandonné ce projet car les résultats n'étaient pas reproductibles et cohérents avec les quantifications directes des transcripts.

Finalement nous avons utilisé une souche qui porte une version étiquetée de Rim101 avec l'étiquette V5 pour essayer de mettre en évidence par *in vivo* chromatine immunoprécipitation (ChIP) que les promoteurs sont des cibles directes de Rim101p. Nos résultats indiquaient un enrichissement

Dans un dernier projet, nous avons essayé de mettre en évidence par des approches d'immunoprécipitation de la chromatine (ChIP) que les promoteurs étient des cibles directes de Rim101p en utilisant une souche qui exprimait une version de Rim101p étiquetée avec l'épitope V5. Nous avons observé un plus grand nombre des promoteurs cibles dans les échantillons pris à pH alcalin que dans ceux pris à pH acides. Toutefois, ces résultats n'étaient pas très solides, car la reproductibilité était faible et nous avons occasionnellement observé un enrichissement similaire pour des promoteurs non-régulés utilisés comme contrôles dans ces expériences.