

**Schriftliche Ausarbeitung**

**zum Seminar**

**Wie Zellen denken: Struktur und Funktion von Regulationsnetzen**

**WiSe 05/06**

**Christina Habermehl**

**und**

**Ulrike Schmidt**

**Referiertes Paper:**

**Conservation and evolvability in regulatory networks: The evolution of  
ribosomal regulation in yeast;**

**Amos Tanay, Aviv Regev, Ron Shamir**

**PNAS May 17, 2005 vol. 102 no. 20 7203–7208**

Einleitung und Methoden: Ulrike Schmidt

Ergebnisse und Diskussion: Christina Habermehl

## Einleitung

Durch die DNA-Microarray Technologie ist es heute möglich, die transkriptionale Antwort eines ganzen Genoms unter verschiedenen experimentellen Bedingungen zu untersuchen. Die im Folgenden vorgestellte Arbeit beschäftigt sich mit einer Methode zur Verbindung der Informationen aus Sequenz- und Genexpressionsdaten, um daraus mögliche Szenarien der Evolution rekonstruieren zu können.

## Transkriptionale Module

Um die transkriptionelle Regulation zu analysieren, müssen die co-regulierten Gene und die experimentellen Bedingungen, die die Co-Regulationen bedingen, bekannt sein. Diese Kombination aus einer Gruppe von Genen und Bedingungen ist hier definiert als ein transkriptionales Modul. Ein transkriptionales Modul ist eine Gruppe von Genen, die sich unter bestimmten Bedingungen gleich verhalten, die also co-reguliert sind.

Transkriptionale Module co-regulierter Gene spielen eine Schlüsselrolle in regulatorischen Netzwerken. Transkriptionale Module sind meist über die Spezies hinweg hoch konserviert und orthologe Module sind oft mit konservierten cis-Elementen assoziiert.

Ziel dieses Ansatzes

Mit dieser erweiterten Definition des Clusters, dem Prinzip der transkriptionalen Module, werden Gene zu kontextabhängigen transkriptionalen Modulen zusammengefasst. Diese können auch überlappen.

Expressionsdaten und bisher uncharakterisierte Sequenzdaten sollen so miteinander verbunden werden und dadurch eine Rekonstruktion der Evolution auf Promotorebene ermöglichen.

## Material und Methoden

Material

Es wurden die veröffentlichten Genomdaten für 17 Hefe-Spezies und die Genexpressionsdaten von zwei, phylogenetisch weit entfernten Hefen, der Bäckerhefe (*S.cerevisiae*) und der Spalthefe (*S.pombe*), verwendet. Auch der phylogenetische Stammbaum der 17 hier betrachteten Hefen ist aus der Literatur bekannt.

Methoden

### 1. Sequenzanalyse

Im ersten Schritt wurden die Promotoren identifiziert und einige Annotationsfehler in den Sequenzdaten der 17 Hefe-Spezies korrigiert.

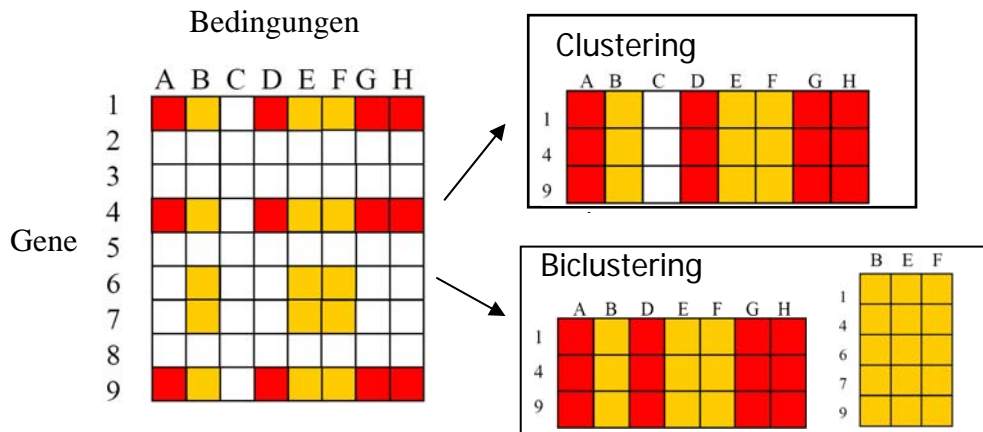
### 2. Identifizierung von Orthologen Transkriptionsmodulen

Zur Erkennung von transkriptionalen Modulen wurde der Biclustering-Algorithmus SAMBA (Statistical Algorithmic Method for Biclustering) auf die Expressionsdaten von *S.cerevisiae* und *S.pombe* angewendet.

Genexpressionsdaten werden üblicherweise als Matrix abgespeichert, wobei jede Reihe ein Gen und jede Spalte eine bestimmte experimentelle Bedingung. Jedes Element in der Expressionsmatrix repräsentiert den Expressionslevel eines bestimmten Gens unter einer bestimmten Bedingung und wird durch eine reelle Zahl dargestellt.

Es wurden viele Clustering-Ansätze entwickelt, um die Genexpressionsdaten aus Microarray-Experimenten zu analysieren, allerdings sind die Standard-Cluster-Methoden begrenzt, z.B. wenn es Bedingungen gibt unter denen die Genaktivität nicht korreliert und so kein Clustern möglich ist.

Anders bei Biclustering-Algorithmen, die auch Gruppen von Genen bestimmen können, die sich nur unter bestimmten experimentellen Bedingungen gleich verhalten. Dadurch können Teilcluster gefunden werden, die beim normalen Clustern verloren gegangen wären (Abb.1).



**Abb.1 Vergleich zw. Der Clustering- und Biclustering-Methode**

Bei den Standard-Clustering-Methoden werden die Gene, die sich unter allen Bedingungen gleich verhalten, zu einem Cluster zusammengefasst. Beim Biclustering werden auch Teilcluster erfasst, also Gengruppen, die sich nur unter bestimmten Bedingungen gleich verhalten.

#### Der SAMBA-Biclustering-Algorithmus

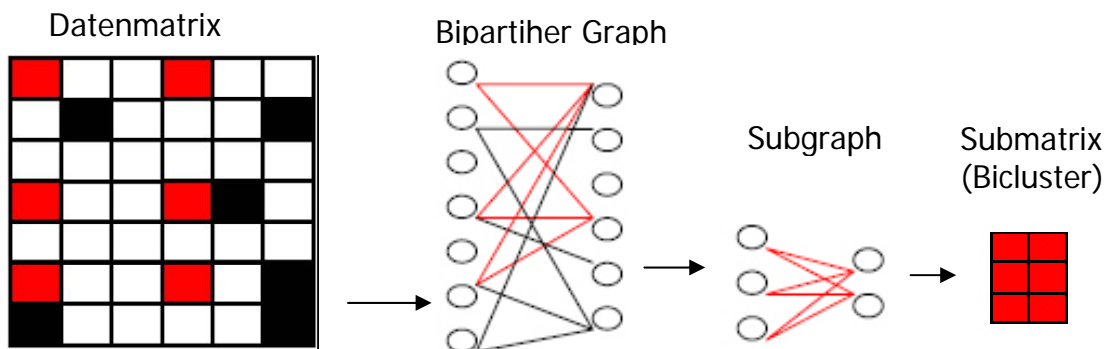
Bei dieser Biclustering-Methode wird die Expressionsmatrix in einen bipartiten Graphen konvertiert. Die beiden Teile entsprechen dabei den Genen und den Bedingungen.

Jede Kante steht für jede signifikante Expressionsänderung.

Das Finden eines Biclusters entspricht der Suche nach dem schwersten Subgraphen (Abb.2).

Es wird nach konsistenten Biclustern gesucht, die Auswahl der Spalten erfolgt so, dass alle den gleichen oder entgegengesetzten Effekt haben.

Der SAMBA-Algorithmus findet Bicluster mit kohärenten Evolutions.



**Abb.2 Der SAMBA-Algorithmus.**

Messen der Verwandtheitsgrade zwischen den Modulen

Ausgangspunkt sind die im vorherigen Schritt identifizierten transkriptionalen Module. Der Verwandtheitsgrad wird ermittelt, indem die Anzahl der Gene aus *S.cerevisiae*, die ein orthologes Gen in *S.pombe* haben bestimmt wird.

Der Ansatz hierfür ist die Hypergeometrische Verteilung zur Berechnung eines p-Wertes.

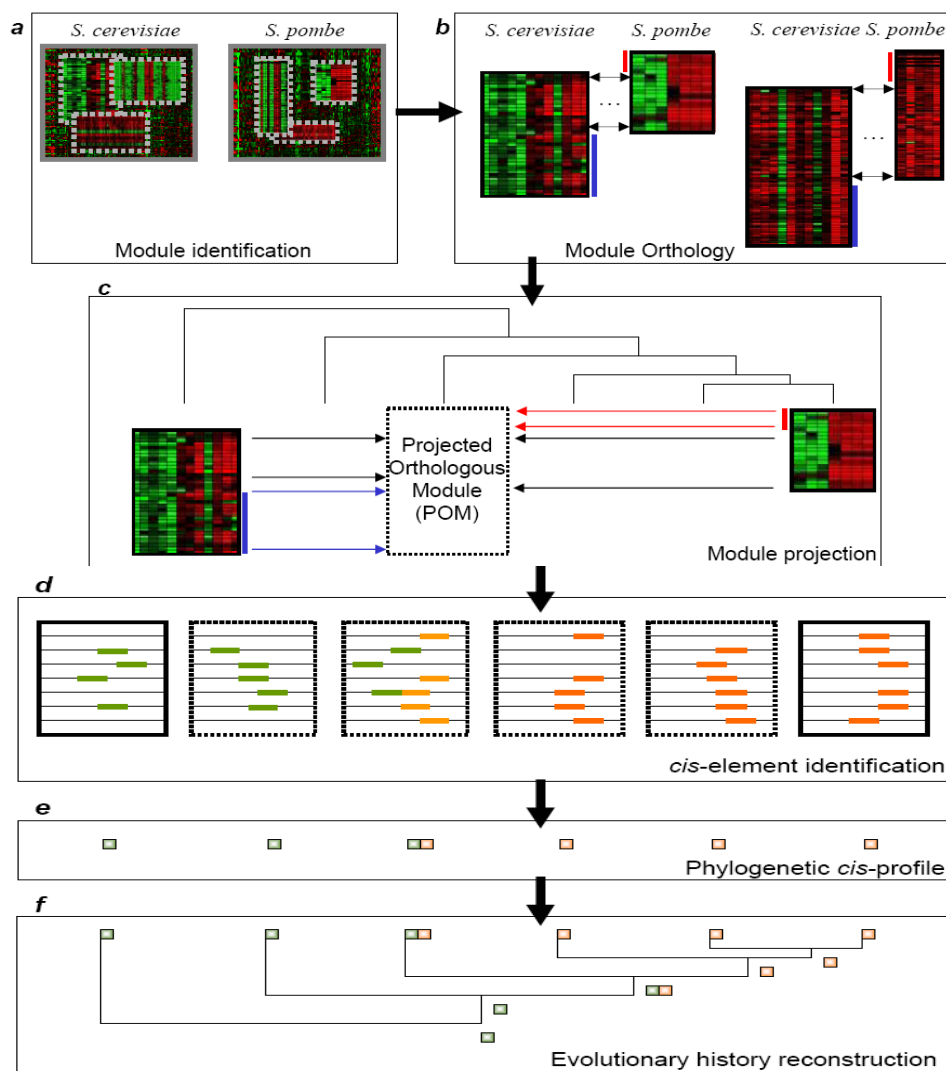
Die Identifizierung von orthologen Genen erfolgt mittels Standard-BLAST-Homologie-Suche von allen ORFs in einer Spezies gegen alle anderen, also paarweise Vergleiche.

Anschließend werden die gefundenen orthologen Module auf die anderen 15 Hefen projiziert. Man spricht dann von so genannten projizierten orthologen Modulen.

### 3. Identifizierung von cis-Elementen und phylogenetisches cis-Profilung

Zur Analyse der cis-Elemente, die mit den Genen eines Moduls verbunden sind, müssen diese zunächst identifiziert werden. Durch die Anwendung eines 2-Step-Motif-Finding and Optimization- Algorithmus werden alle cis-Elemente, die in den konservierten Modulen angereichert sind, gefunden und für jedes Modul das phylogenetische cis-Profil erstellt.

Abb.3 zeigt eine Übersicht über die einzelnen Schritte:



**Abb.3:** a) Identifizierung der transkriptionalen Module aus den Expressionsdaten aus *S.cerevisiae* und *S.pombe*. b) Bestimmung der orthologen Module. c) Projektion der Module auf die anderen 15 Hefe-Spezies, man spricht von den projizierten orthologen Modulen (POM). d) Identifizierung der cis-Elemente. e) Erstellen der phylogenetischen cis-Profile. f) Rekonstruktion der evolutionären Geschichte.

Ab hier: Christina Habermehl

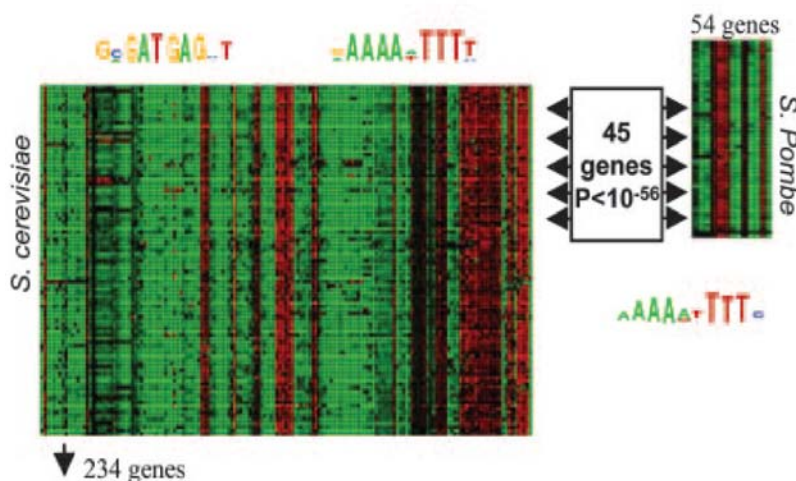
## Ergebnisse

Entsprechend der Reihenfolge der Methoden werden nachfolgend die Resultate der einzelnen Analyseschritte dargestellt.

Zuerst wurden konservierte transkriptionale Module mit Hilfe von Expressionsdaten zweier evolutionär weit entfernten Hefespezies identifiziert. Danach wurden vorhandene Sequenzdaten von 15 weiteren Hefen benutzt um orthologe Module herzuleiten und cis-regulatorische Elemente zu identifizieren, die mit jedem Modul in jeder Spezies verbunden sind. Als letzter Schritt wurde die Evolution dieser cis-Elemente rekonstruiert.

### Identifikation orthologer transkriptionaler Module

Die Anwendung des oben besprochenen SAMBA-Algorithmus in Kombination mit den Expressionsprofilen der Hefen *S.cerevisiae* und *S.pombe* führten zu einer Identifizierung transkriptionaler Module in diesen Spezies. Es konnten insgesamt sechs konservierte transkriptionale Module gefunden werden. Jedes dieser Module, hier das Beispiel des Ribosomalen-Biogenese Moduls (Abb. 4), mit einem bestimmten Profil stellt das Verhalten der einzelnen Gene (Reihen) unter verschiedenen experimentellen Bedingungen (die Spalten) dar.



**Abb. 4** Das ribosomale Biogenese Modul zusammen mit den assoziierten cis-Elementen. Die Reihen stellen die einzelnen Gene, die Spalten die verschiedenen Experimente dar. Rot bedeutet, das Gen ist induziert, Grün bedeutet Repression. Im Rechteck steht die Anzahl der orthologen Gene mit ihrem p-Wert.

Ebenfalls dargestellt sind die mit den Modulen jeweils verbundenen cis-Elemente, die in den Promotoren der Gene eines Moduls vermehrt angetroffen wurden. Diese wurden für *S.cerevisiae* und *S.pombe* separat ermittelt. Dabei wurden zum einen ähnliche cis-Elemente in den orthologen Modulen der beiden Hefen gefunden (Abb. 5), es wurden jedoch auch trotz phänotypischer Konservierung der Genantwort cis-Elemente gefunden, die stark divergierten (Abb. 6). Diese Divergenz trat auch in Modulen auf, die Schlüsselfunktionen einnehmen, so zum Beispiel das Modul zur Stress-Antwort und zur Ribosomalen Protein-Synthese. Daraus ergibt sich die Frage, wie eine solche Divergenz in cis-regulatorischen Elementen, die die Expression von hoch wichtigen und streng koordinierten Modulen regeln, ohne negative Auswirkungen stattfinden kann.

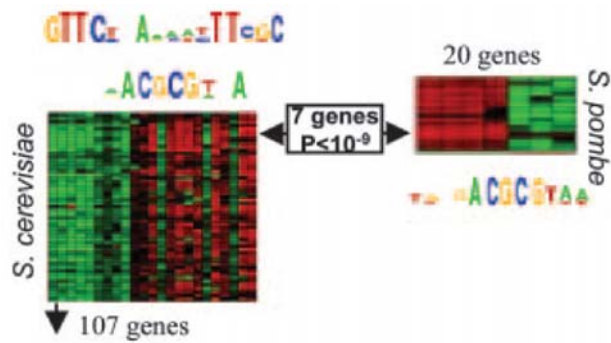


Abb. 5 s-Phase Modul als Beispiel für konservierte cis-Elemente (ACGCGT)

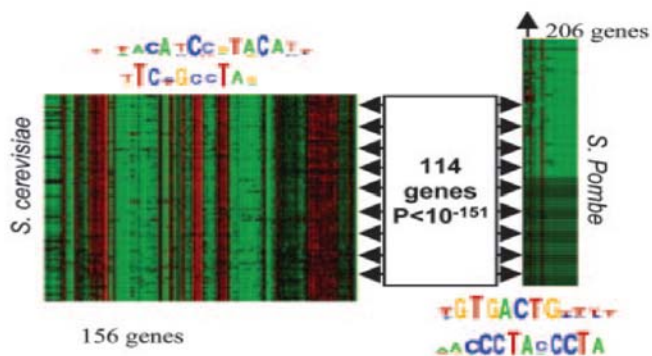


Abb. 6 Ribosomales Protein-Modul als Beispiel für divergente cis-Elemente. RAP1 – Site (TACATCCGTACAT) und IFHL- site (TCCGCTAG) finden sich in *S. cerevisiae* und eine Homol-D box (TGtGACTG) und eine Homol-E site (ACcCTAcCCTA) in *S. pombe*.

## Phylogenetische cis-Profile in 17 Hefen

Um dieser Frage nachzugehen, wurden cis-Elemente in weiteren 15 Hefen analysiert. Diese Hefen befinden sich evolutionär betrachtet zwischen *S. cerevisiae* und *S. pombe*. Aufgrund der Tatsache, dass zu diesen zusätzlichen Hefen nur Sequenzdaten vorlagen, und keine Expressionsdaten zugänglich waren, wurden projizierte orthologe Module (POM, siehe auch Methoden-Teil) erstellt.

Danach wurden angereicherte cis-Elemente in den Genen der POM ermittelt, und nach ihrem Auftreten in anderen Spezies gesucht. Das dadurch ermittelte phylogenetische cis-Profil verbindet jedes Modul mit einer Anzahl von cis-Elementen in jeder Spezies. Durch Zuordnung der cis-Profile für je ein Modul über alle Spezies in einem bekannten phylogenetischen Baum kann man Konvergenzen und Divergenzen übersichtlich darstellen (Abb. 7).

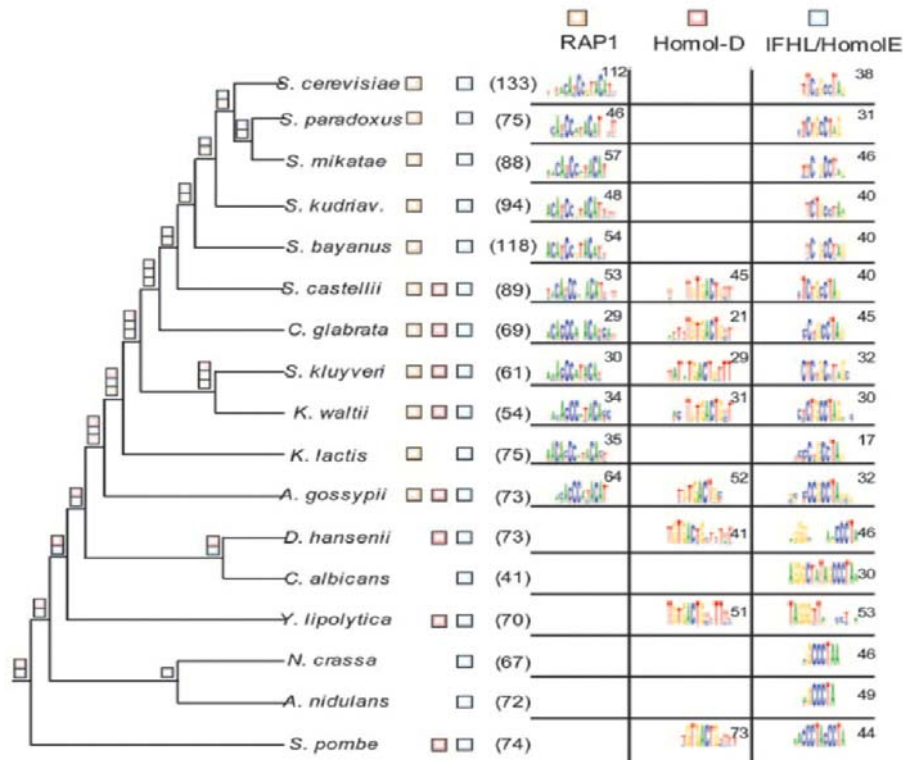


Abb. 7 Phylogenetische cis-Profile und Evolution der cis-Elemente im ribosomalen Protein-Modul. Die bekannte Phylogenie über 17 Hefen wird zusammen mit den assoziierten cis-Elementen der einzelnen Hefen dargestellt. Die angereicherten Elemente werden in RAP1-Site, Homol-D-Box und IFHL- und HomolE-site untergliedert und je nach Vorkommen den Hefen zugeordnet. Ein Auftreten der Homol-D-Box in *S. pombe* sowie das Vorhandensein von RAP1 in *S. cerevisiae* ist zu erkennen. In anderen Spezies, z.B. *A.gossypii* sind beide cis-Elemente angereichert.

## Rekonstruktion der Evolution des Ribosomalen Regulatorischen Programms

Durch Analyse der in Abb. dargestellten phylogenetischen cis-Profile des Ribosomalen Protein (RP) -Moduls können mögliche Szenarien, die Evolution der regulatorischen Elemente betreffend, erstellt werden. So fällt eine starke Divergenz zwischen *S. cerevisiae* und *S. pombe* auf: die Gene des RP-Moduls in *S. cerevisiae* haben in ihren regulatorischen Abschnitten die RAP-1-Bindungsstelle sowie die IFHL-site angereichert, wohingegen in *S. pombe* die Homol-D und die HomolE-Box in ihrem Auftreten dominieren.

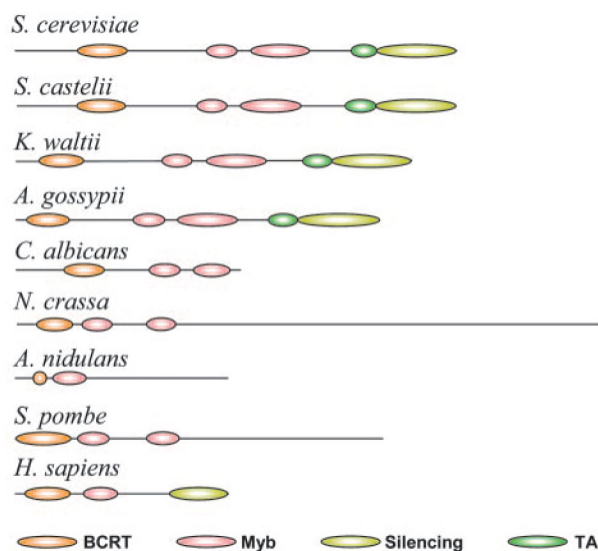
Die Profile von *S. castellii*, *C.glabrata*, *S.kluyveri*, *K.waltii* und *A.gossypii* enthalten sowohl die RAP-1-Bindungsstelle als auch die Homol-D-Box. Diese Redundanz von Bindungsstellen impliziert, dass ein Umschalten der regulatorischen Mechanismen im Laufe der Evolution stattgefunden haben muss. Eine Änderung der Bindungsstellen aufgrund von Veränderungen auf Sequenzebene, wie z.B. Gendrift liegt hier nicht vor. Auch ist es unwahrscheinlich, dass die Regulation einer Schlüsselfunktion einer sprunghaften Änderung ausgesetzt war.

Es ist davon auszugehen, dass eine alte Homol-D-Box die Schlüsselrolle in der Regulation des RP-Moduls einnahm. Vor der Entwicklung von *A.gossypii* erschien RAP-1 als zusätzlicher Regulator, ohne dass die Funktionalität von Homol-D verloren ging. Durch stetige Zunahme des Einflusses von RAP-1 verlor Homol-D im Laufe der Evolution sein Gewicht und ging in

einigen Hefen sogar verloren. Somit konnte durch „Einschleichen“ eines weiteren regulatorischen Elementes im Laufe der Zeit ein Umschalten der regulatorischen Mechanismen vonstatten gehen. Doch wie könnten die molekularen Mechanismen ausgesehen haben, die ein solches „Einschleichen“ und Umschalten ermöglichten?

Einige Punkte könnten dieses mögliche evolutionäre Szenario belegen. Der Transkriptionsfaktor Rap1p wird mit der Regulation der Telomerlänge in Verbindung gebracht. Dies ist eine alte und hochkonservierte Funktion.

In den Promotoren des RP-Modul ist ein Submotiv zu finden, das Rap1p binden kann. Somit kann angenommen werden, dass durch Entstehen einer Bindungsstelle im RP-Modul zu einem bereits existierenden Transkriptionsfaktor ein neuer regulatorischer Mechanismus eingeführt wurde. Weiterhin wurde das Eindringen von RAP1 in die Promotoren durch die Akquirierung einer neuen Transaktivierungsdomäne (TA) durch Rap1p vor der Entwicklung von *A.gossypii* ermöglicht (Abb. 8). Rap1p, das keine TA vorweisen kann (beispielsweise in *C.albicans*), ist nicht in der Lage Rap1p *mit* TA zu ersetzen, dieses Rap1p kann also nicht adäquat an RAP1 binden. Spezies, bei denen RAP1 als regulatorisches cis-Element gefunden wurde besitzen alle Rap1p *mit* TA, Hefen, in denen RAP1 keine regulatorische Rolle hat, haben Rap1p *ohne* TA.



**Abb. 8 Evolution der Rap1p – Sequenz in verschiedenen Spezies und dem Menschen. Schematische Darstellung der Protein-Sequenz. DNA-bindendes Myb ist in allen Sequenzen vorhanden. Die Transaktivierungsdomäne (TA) ist nicht in Spezies vorhanden, die keine RAP1 Bindungsstelle in ihrem RP-Modul haben (*C. albicans*, *N. crassa*, *A. nidulans* und *S.pombe*). Rap1p in Hefen, deren RP-Modul durch RAP1 reguliert wird, besitzt die TA (*S.cerevisiae*, *S. castellii*, *K. waltii* und *A. gossypii*).**

Dieses Erkenntnisse lassen den Schluss zu, dass RAP1 zum zusätzlichen Regulator für das RP-Modul wurde, indem ein alter Transkriptionsfaktor - Rap1p - eine neue TA erhielt und somit an die DNA im Promotorbereich des RP-Modul binden konnte. Durch diese zusätzliche Regulation verloren die Gene nicht ihre Funktionalität und im Laufe der Evolution konnte der alte Regulator, die Homol-D-Box, an Bedeutung verlieren. Das führte soweit, dass sie in einigen Hefen sogar eliminiert wurde.

Weitere Analysen des phylogenetischen cis-Profiles hinsichtlich der zweiten regulatorischen Sequenz, der IFHL- bzw. HomolE – Bindungsstelle deuten auf einen anderen Mechanismus

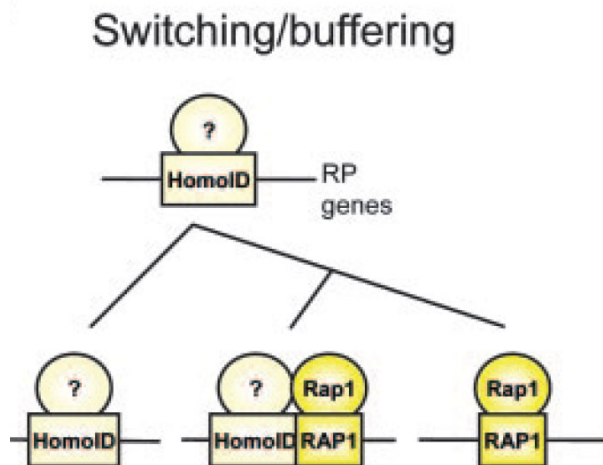


der zur Änderung in der Einflussnahme dieser führte. Hier deutet alles auf eine graduelle Divergenz in der DNA-Sequenz. Auch wenn die Sequenzen nicht immer übereinstimmen, so tragen doch viele Spezies Varianten des IFHL-Motives. Somit ist eher anzunehmen, dass die Divergenz der Bindungstellen kleineren Mutationen in der DNA-Sequenz oder einer Domänen-Duplikation bzw. Dimerisation im IFHL-Bindungsprotein zuzuschreiben ist (ohne Abb.).

## Diskussion

Die Analyse der Evolution transkriptionaler Module kann ein wichtiger Schritt sein, um regulatorische Netzwerke besser zu verstehen. Es wurde hier gezeigt, dass in vielen Fällen die Regulation wichtiger und konservierter transkriptionaler Module von ebenso konservierten cis-Elementen gesteuert wird. Andererseits wurde auch gezeigt, dass auch Divergenz bei den regulatorischen Mechanismen vorkommt und nicht unbedingt zu einem Verlust der Funktionalität des Moduls führt.

Die Arbeit stellt einen Teil der vielfältigen Möglichkeiten dar, wie die Evolution Mechanismen der Regulation transkriptionaler Module verändern kann. Diese Möglichkeiten beinhalten neben der Konservierung auch Divergenz in der Bindungsstelle aufgrund gradueller Sequenzänderungen, Erweiterung eines existierenden Programmes durch Einführung einer neuen Bindungsstelle, Bindungsstellenverlust, Umschalten des Transkriptionsfaktors ohne Änderung des cis-Elements oder, wie hier näher beschrieben das Umschalten der Regulation durch Einführung einer neuen Bindungsstelle bei gleichzeitiger Redundanz und allmählichem Verlust der Funktion des alten cis-Elements (Abb. 9).



**Abb. 9 Umschalten eines regulatorischen Programms durch Einführung einer neuen Bindungsstelle, zwischenzeitlicher Redundanz und Verlust von regulatorischen Elementen am Beispiel des RP-Moduls.**

Die Redundanz von cis-Elementen ermöglicht es dem Organismus flexibler auf Änderungen zu reagieren und erhöht so die Fähigkeit zur Entwicklung.

Es konnten uncharakterisierte Daten mit bekannten evolutionären Konzepten kombiniert werden. Mit Hilfe informationstechnischer Methoden konnten Sequenz- mit Expressionsdaten verknüpft werden, was zu einer Rekonstruktion der Evolution eines regulatorischen Netzwerkes führte.