UNIVERSIDADE ESTADUAL DE CAMPINAS

doutorado

INSTITUTO DE BIOLOGIA

UNIVERSIDADE ESTADUAL DE CAMPINAS

Marcos Antonio de Oliveira

CLONAGEM E CARACTERIZAÇÃO DE FATORES DE TRANSCRIÇÃO DE *Hansenula polymorpha* ENVOLVIDOS NA REGULAÇÃO POR GLICOSE DO GENE *MOX*.

Este exemplar corresponde à redação final da tese defendida pelo(a) candidato (a) Marcos Antonio de Oliveira.

e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de Biologia para obtenção do título de Doutor em Genética e Biologia Molecular na área de Genética de Microorganismos

Orientador: Prof. Gonçalo Amarante Guimarães Pereira

Campinas - S.P.

2003

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

Oliveira, Marcos Antonio de

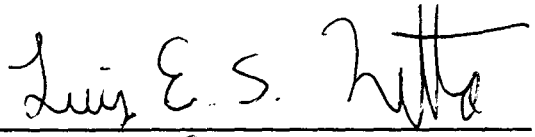
OL4c Clonagem e caracterização de fatores de transcrição de *Hansenula polymorpha* envolvidos na regulação por glicose do gene *MOX* / Marcos Antonio de Oliveira.--
Campinas, SP: [s.n.], 2002.

Orientador: Gonçalo Amarante Guimarães Pereira
Tese (Doutorado) – Universidade Estadual de Campinas . Instituto de Biologia.

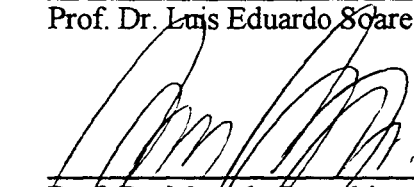
1.Transcrição genética. 2.Levedura. 3.Biotecnologia. I. Pereira, Gonçalo Amarante Guimarães. II. Universidade Estadual de Campinas. Instituto de Biologia. III.Título.

BANCA EXAMINADORA:

Prof. Dr. Gonçalo Amarante Guimarães Pereira

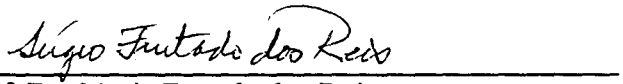


Prof. Dr. Luis Eduardo Soares Netto



Prof. Dr. Marcelo Brocchi

Prof. Dr. Marcelo Menossi Teixeira



Prof. Dr. Sérgio Furtado dos Reis

Aos meus pais (*in memoriam*), os quais me ensinaram o sentido das palavras caráter e perseverança e, apesar do breve tempo, me mostraram que o sentido de tudo se encontra dentro de nós mesmos.

Ao meu irmão, meu melhor amigo e meu maior exemplo. Sempre presente nos momentos importantes de minha vida, tornando-a sempre melhor. Muito obrigado de todo o coração!!!

Ao Morfeus (†) e Schultz: grandes amigões desta longa caminhada.

You can't always get what you want
But if you try sometimes,
You might find....

(M.J & K.R)

AGRADECIMENTOS

À Carla, pessoa fundamental na minha vida. Pelo grande amor, carinho e paciência em todos os momentos, principalmente os mais difíceis.

Ao Prof. Dr. Gonçalo (mande bala), por todo o apoio, amizade e paciência demonstrada ao longo destes anos. Aos sonhos que se realizaram e aqueles que irão se concretizar.

Ao Prof. Dr. Sérgio Furtado dos Reis (irmão), uma pessoa extraordinária que acima de um pesquisador de imensas qualidades, é um amigo de qualidades inenarráveis.

Ao Prof. Dr. Marcelo Brocchi, grande amigo, que sempre foi excepcional e exemplar desde os tempos da *Xylella* (que ralação!!!). Tenho grande honra em tê-lo tanto na minha banca quanto na pré-banca.

Ao Prof. Dr. Luis Eduardo Soares Netto. Exemplo de cientista sério e competente, como também amigo, o qual me proporcionou grandes oportunidades e para tanto demonstrou sua grande paciência.

Às Profas. Dras. Fernanda Gadelha e Maria Isabel Cano pela valioso auxílio na pré-banca e grande amizade demonstrados ao longo destes anos.

Ao Prof Dr. Flávio Henrique Silva, meu orientador de mestrado o qual sempre foi e será um grande exemplo.

À Mariana Lyra, representante discente na CPG, com certeza a grande responsável pelo término da novela... Prova que a inteligência e o bom senso são essenciais para a solução de todas as coisas. Muito obrigado!!!

Aos grandes amigos de bancada que foram cultivados ao longo destes anos: Anderson (Kiwi) felizmente meu primeiro amigo no laboratório o qual admiro pela garra, honestidade e seriedade (hum...). Dentre as suas inúmeras qualidades posso ressaltar duas: indescritível habilidade para ciência e grande competência para juiz de sinuca... Ao Victor (Vô), grande ariano, pelos grandes

momentos em que trabalhamos juntos (olha os tubinhos boiando...). Muitas coisas deram erradas (RNA bizarro, actina mutante, seqüênciamentos sem fim...), mas nos fizeram aprender bastante.

Anita (orgulho), sinônimo de garra, inteligência, franqueza e honestidade. Acredito que acompanhei os primeiros passos de uma grande carreira.

Às malditas magricelas: Ana (Dropout) pela inenarrável amizade, a qual foi se fortalecendo com o tempo, dentre as suas inúmeras qualidades, ressalto: o senso de quantidade e a capacidade de guardar segredos... Carlinha (Honeybee girl), grande amiga, que sabe tudo de alemão, além de ser um exemplo de grande perseverança na elaboração de géis de poliacrilamida e sequenciamentos de DNA...

À galera da bioinformática: Buda, Marcelo Calrazzolle, Marcelo Marchi, Luciano e Marcos Renato pelo auxílio no tratamento dos dados de seqüênciamento de *Hansenula polymorpha* (trimagem, blastagem, contigagem e todas as outras bioinformaticagem).

À todo o povo do laboratório destacando: Oodaalys (a cubana mais incrível do mundo), Naiara (senhorita sardinha), Anders (senhor pinto), Bárbara (Olá pessoax), Baiky (rainha do bambu), Andréa (por todas as prezadas brigas do seqüenciador), Fernandinha (lixo no meio das pernas...), Eliane (pessoa de fundamental importância no laboratório tanto profissional quanto pessoalmente) Diana (três cabeças), Raquel (Fof's), Camila (chuchu), Vitão (cuidado nas próximas Bioartes...), Amanda (grande baixinha) e Lilica (magnífica autoclavadora de frascos).

Ao *protehomem* Marquinhos, grande amigo e um cientista nato. Por todos os papos científicos, ou não, travados nas cervejinhas de final de tarde.

Aos amigos de sempre: Akira, Bruno, Catatau, E.T, Guigas, Helio, Jarbão, Jairo, Marcelo (Alf), Negrão, Pompeu(s), Udo, Zuba e todos os meus grandes amigos extralaboratoriais.

À todas as pessoas que de forma direta ou indireta auxiliaram na realização deste trabalho.

À FAPESP pelo auxílio financeiro.

Ao tudo que possui vários nomes e só um significado, Deus .

RESUMO

O controle da expressão gênica é tema de grande interesse da biologia molecular atual. O entendimento dos processos envolvidos no principal ponto de regulação, a iniciação da transcrição, deve fornecer nos próximos anos, ferramentas úteis para o controle e tratamento de doenças de caráter genético bem como a manipulação da expressão de genes visando a produção de proteínas de interesse farmacêutico, alimentar e bioquímico industrial em microrganismos, animais e plantas.

Um dos sistemas de regulação gênica mais bem estudos é o da regulação catabólica (regulação por glicose). Este se caracteriza pela repressão de transcrição de genes envolvidos na degradação de fontes alternativas de carbono quando a glicose está presente no meio. Um interessante modelo para o estudo de regulação por glicose e encontrado em um pequeno grupo de leveduras todos pertencentes aos gêneros *Candida*, *Pichia*, *Hansenula* e *Torulopsis*, capazes de degradar metanol como única fonte de carbono, dentre elas a levedura *Hansenula polymorpha.*
A primeira reação bioquímica necessária pela degradação do metanol em *H. polymorpha* ocorre nos peroxissomos, sendo catalisada pela enzima Metanol Oxidase (Moxp).

O gene responsável pela produção da enzima (*Mox*) é fortemente regulado em nível transcricional. Quando esta levedura cresce em meio fermentativo, como glicose ou etanol, não é detectado nenhum transcrito de *Mox*, entretanto quando esta levedura cresce em meio contendo metanol como única fonte de carbono a transcrição é fortemente ativada e seu produto (Moxp) pode representar até 30% de todas as proteínas solúveis da célula. Graças a estas características o promotor *MOX* é largamente utilizado para a produção de proteínas heterólogas de grande importância na indústria farmacêutica. Entretanto, apesar de sua grande importância, nenhum fator responsável pelo controle de sua expressão se encontra identificado.

Em *Saccharomyces cerevisiae*, organismo modelo para estudos genéticos e bioquímicos, diversos fatores que controlam a transcrição de genes reprimidos por glicose, de forma direta ou indireta, já se encontram caracterizados. Alguns destes fatores agem de forma pontual sobre determinados genes de uma via catabólica, enquanto outros agem de forma global e são altamente conservados desde eucariotos inferiores até os superiores como plantas e mamíferos. Dentre os elementos descritos mais importantes, podemos destacar: os complexos SWI/SNF e SNF1 os quais respondem por eventos de ativação de transcrição e o complexo formado por Mig1p/Ssn6p/Tup1p, o qual é responsável pela repressão da transcrição de diversos genes sujeitos a repressão catabólica.

Neste trabalho foi possível identificar em *H. polymorpha* fatores homólogos a *Swi2, Tup1* e *Snf1* descritos em *S. cerevisiae* e avaliar o envolvimento destes fatores no metabolismo de fontes alternativas de carbono, em especial o metanol o qual investigamos a transcrição de *Mox* e genes corregulados através de disrupção gênica. A disrupção do gene *Tup1* não alterou as características de utilização das fontes de carbonos testadas. A disrupção de *Snf1*, levou a uma queda nos transcritos de *Mox* e *Das*. Ao passo que a disrupção *Swi2* ou *Swi3* levou a diminuição na utilização das diversas fontes de carbono testadas, bem como um efeito inesperado: a restauração da via uma via catabólica quiescente.

Também foi objetivo deste trabalho a criação de banco de dados genéticos bem como de ferramentas de bioinformática para a análise genômica o qual foi utilizado para a análise de RSTs (Random Sequence Tags) de *H. polymorpha*.com a finalidade de identificar fatores adicionais que podem influenciar na expressão de *Mox* e biosíntese de peroxissomos, bem como elementos adicionais relevantes presentes no genoma desta levedura. Para tanto foi realizado um projeto de seqüênciamento de RSTs que gerou 5.784 "*reads*" os quais se encontram processados, anotados e categorizados em banco de dados de nosso laboratório.

SUMMARY

The control of the gene expression is subject of great interest of current molecular biology. The understanding of the involved processes in the main point of regulation, the transcription initiation, must supply, in the next years, useful tools for the control and treatment of illnesses of genetic character as well as the manipulation of the expression of genes, aiming for the production of important proteins in microorganisms, animals and plants, used in the pharmaceutical, food and biochemical industries.

One of the most studied systems of genic regulation is the catabolic regulation (glucose regulation). This is characterized by the transcription repression of genes involved in the degradation of alternative carbon sources when the glucose is present in the environment. An interesting model for the study of glucose regulation is found in a small group of yeasts, all belonging to the genera *Candida*, *Pichia*, *Hansenula* and *Torulopsis*, capable of degrading methanol as the only carbon source, amongst them *Hansenula polymorpha*. The first biochemical reaction required for the methanol degradation in *H. polymorpha* occurs in the peroxisomes, being catalyzed by the enzyme Methanol Oxidase (Moxp).

The responsible gene for the production of the enzyme (*Mox*) is strongly regulated at the transcriptional level. When this yeast grows in fermentative medium, as glucose or ethanol, no *Mox* transcript is detected. However, when *H. polymorpha* grows in methanol-supplied medium as the only carbon source, the transcription is strongly activated and its product (Moxp) can represent up to 30% of all soluble proteins of the cell. Due to these characteristics the *Mox* promoter is widely used for the heterologous proteins production, which are very important for the pharmaceutical industry. However, despite its great importance, no responsible factor for the control of its expression has been identified.

In *Saccharomyces cerevisiae*, a model organism for genetic and biochemical studies, several factors that control the transcription of genes repressed by glucose, directly or indirectly, are characterized. Some of these factors act punctually on specific genes of a catabolic pathway, whereas others act globally in the transcription of several genes, and are highly conserved, from lower eukaryotes to higher ones, as plants and mammals. Amongst the most important elements described, we can highlight: SWI/SNF and SNF1 complexes, which respond for events of transcription activation and the complex formed by Mig1p/Ssn6p/Tup1p, responsible for the transcription repression of several genes subjected to the catabolic repression.

In this work we have identified in *H. polymorpha* homologous factors to *Swi2*, *Tup1* and *Snf1* described in *S. cerevisiae* and evaluated the involvement of these factors in the metabolism

of alternative carbon sources, specially methanol, where we investigate the transcription of *MOX* and genes co-regulated through gene disruption. The *Tup1* gene disruption did not modify the characteristics in the use of the carbon sources tested. The *Snf1* disruption led to a decrease in *MOX* and *DAS* transcripts. The *Swi2* or *Swi3* disruptions led to the reduction in the utilization of several carbon sources, as well as an unexpected effect: the restoration of quiescent pathway.

The creation of genetic database of *H. polymorpha* was also the objective of this work with the purpose to identify additional targets that can influence in the *MOX* gene expression and peroxisomes biosynthesis, as well as relevant additional elements present in the genome of this yeast. To achieve this goal a sequencing project of RSTs (Random Sequence Tags) was performed, generating 5,784 reads which have been processed, annotated and categorized in our laboratory database.

ABREVIATURAS

Adh2: Acohool dehidrogenase 2

Adr1: Adh2 regulator

AMP: adenosina monofosfato

AMPK: AMP- activated Protein Kinase

ATP: adenosina trifosfato

CPI: Complexo de Pré-Iniciação de transcrição

Das: Dihidroxiacetona sintase

Cat: Catalase

CTD: Carboxy Terminal Domain

GTF: General Transcription Factor

HAT: Histone Acetil Transferase

mRNA: RNA mensageiro

Mig1: Multicopy inhibitor of GAL expression

Mox: Metanol oxidase

MS: Mass Spectrometry

SANT: SWI-ADA -N-CoR-TFIIIB

Snf1: Sucrose non fermenting

Ssn6: Supressor of snf (6)

Swi2: mating type Switching (2)

Swi3: mating type Switching (3)

TAF: Transcription Associated Factors

TBP: Tata binding protein

TIC: Taf Iniciator dependent Cofactors

Tup1: Timidine uptake

TPR: TetratricoPeptide Repeat

UAS: Upstream Activation Sequence

URS: Upstream Repressing Sequence

1. REVISÃO DA LITERATURA

1.1 Transcrição em eucariotos

A regulação gênica foi, em última instância, o que possibilitou o surgimento e manutenção da vida. As complexas interações bioquímicas que regem a regulação podem se dar desde a transcrição e tradução ou até mesmo por meio de mecanismos pós-traducionais. Entretanto, o ponto crucial de regulação é o de iniciação da transcrição.

Enquanto que em organismos procariotos a seleção da transcrição é determinada pelo fator σ, nos eucariotos a distinção do gene que necessita ser transcrito é muito mais complexa. Os genes eucarióticos são transcritos por três RNA polimerases distintas: RNA polimerase I (RNApol I) que responde pela transcrição dos RNAs ribossomais; a RNA polimerase III (RNApol III) a qual transcreve os genes de RNAs transportadores e outras pequenas partículas de RNA e a RNA polimerase II (RNApol II) responsável pela transcrição dos genes que codificam os RNAs mensageiros, que por sua vez irão dar origem às proteínas [1-3].

Para que a RNA pol II transcreva os genes de forma adequada é necessário que ela forme um complexo com mais de 30 proteínas distintas, os Fatores Gerais de Transcrição (GTFs), na região regulatória (promotora) dos genes que transcreve. Este complexo é denominado Complexo de Pré-Iniciação (PIC), o qual é responsável pela sustentação de um nível basal de transcrição, que pode ser modulada de forma positiva ou negativa por meio de ativadores ou repressores transcricionais, que podem ser globais ou gene/classe específicos [2, 4, 5].

1.1.2 RNA polimerase II

Em procariotos o cerne da RNApol é um tetrâmero formado pelas proteínas β, β' e $\alpha 2$, ao passo que nos eucariotos ela é formada por 10–12 proteínas. A RNApol II mais bem caracterizada é a de *Saccharomyces cerevisiae* [6,7]. É um complexo formado por 12 polipeptídeos (Rpb1-Rpb12) e possui uma massa de aproximadamente 500 KD. Esta enzima apresenta um grau de conservação notável entre os organismos. Por exemplo, as seqüências de todos os genes constituintes da RNA pol II humana possuem elevada similaridade com seus equivalentes de *S. cerevisiae*, ainda, pelo menos 10 Rpbs podem ser substituídas pelos seus correspondentes em levedura [6,7].

As proteínas da RNA pol II eucarióticas também apresentam grande similaridade com seus representantes de procariotos. As subunidades Rpb1 e Rbp2, respectivamente, apresentam

grande similaridade com a as subunidades β e β' bacterianas, e as subunidades Rpb3 e Rpb11 compartilham uma baixa homologia com a subunidade α procariótica [6].

A estrutura da RNA Pol II de *S. cerevisiae*, foi recentemente resolvida por Cramer e colaboradores (2001) através de cristalografia de raios-X em conjunto com Ressonância Magnética Nuclear (RMN) e os resultados são altamente relevantes para o entendimento da função de cada subunidade. As duas maiores subunidades do complexo Rpb1 e Rbp2 formam uma fenda rica em poli-alanina que contém o sítio ativo, identificado através de substituição de íons Mg^{2+} por Zn^{2+}, Mn^{2+} ou Pb^{2+}. A região inferior da fenda constituída por Rpb1-Rpb2 contém dois poros que podem estar relacionados com canais para a entrada de nucleotídeos ou saída da cadeia nascente de RNA e fatores que afetam na clivagem da cadeia de RNA [7, 8].

A subunidade Rpb5 juntamente com Rpb1 e Rpb9 constituem uma estrutura em forma de "mandíbula" que devem auxiliar no posicionamento do complexo proteico na fita de DNA. Rpb1-Rpb2 juntamente com Rpb9 formam uma estrutura em forma de grampo que, aparentemente, possuem a função de se ligar ao DNA e estabilizar o complexo DNA-proteína. Rpb3 e Rpb11 apresentam uma estrutura similar ao α dímero bacteriano e são parte de um subcomplexo juntamente com Rpb10 e Rpb12 [8].

As subunidades Rpb4 e Rbp7 formam um subcomplexo, que pode se dissociar do complexo principal, e está associado com resposta à estresse e iniciação de transcrição [9, 10]. Quando *S. cerevisiae* está em fase logarítmica de crescimento, os montantes de Rpb4 e Rpb7 são encontradas em quantidades subestequiométricas em relação às demais Rpbs. Ao passo que na fase estacionária, Rpb4 e Rpb7 são encontradas em quantidades equivalentes às demais subunidades [10].

A subunidade Rpb1 possui um domínio carboxiterminal essencial na regulação da atividade enzimática da RNApol II, a cauda CTD (Carboxi Terminal Domain). Este domínio é constituído por repetições adjacentes *(in tandem)* do heptapeptídeo Y-S-P-T-S-P-S, sendo altamente conservado entre os eucariotos. O número de repetições pode variar de 26 em *S. cerevisiae* até 52 em *H. sapiens*. A cauda CTD existe em pelo menos dois estados de fosforilação os quais variam de acordo com o momento no processo de transcrição (iniciação ou elongação) [11]. Por exemplo, a RNApol II encontrada em complexos de iniciação de transcrição apresenta a CTD defosforilada, enquanto que as presentes na etapa de elongação possuem a cauda altamente fosforilada [6]. Os elementos que promovem a fosforilação pertencem a outras classes de fatores (GTFs e complexo Mediador) e serão apresentados a seguir.

1.1.3 Fatores Gerais de Transcrição (GTFs)

Para que a RNApol II promova a transcrição é necessário que ela forme o complexo de pré-iniciação de transcrição com os GTFs. Seis fatores gerais de transcrição se encontram bem caracterizados: TFIID, TFIIA, TFIIB, TFIIF, TFIIH e TFIIE, cada um com funções específicas nos eventos de transcrição [1, 6, 11].

TFIID: É composto pela TBP (Tata Binding Protein), TFAs (TBP Associated Factors) e TICs (TAF/Iniciator dependent cofactors). A TBP é responsável pelo reconhecimento do elemento "TATA-BOX" (seqüência de DNA rica em AT com capacidade de interagir com proteínas regulatórias), presente na maioria dos promotores eucariotos e passo inicial para a montagem do complexo basal de transcrição. A TBP é suficiente para direcionar sozinha o restante da maquinaria basal de transcrição para promotores que possuem TATA-BOX. Em promotores que não possuem TATA-BOX (TATA-Less promotor) os TAFs são essenciais. Em alguns promotores que contém o elemento Inr (Iniciator) são necessários também os TICs (TAF Iniciator dependent Cofactors). Em promotores que possuem tanto o TATA-BOX quanto o Inr estas proteínas agem de forma sinérgica podendo tanto inibir quanto aumentar a seletividade pelos promotores.

TFIIA: Inicialmente creditava-se ao TFIIA somente a capacidade de se ligar ao TFIID e estabilizar o complexo formado entre ele e a cadeia de DNA. Entretanto, foi constado que o TFIIA age também como um modulador positivo de transcrição pois é alvo de diversos ativadores de transcrição.

TFIIB: Está envolvido com a seleção do ponto de iniciação de transcrição (TSP-Transcription Starting Point) possivelmente regulando a distância entre os promotores e os TSPs. Mutações no TFIIB causa mudança no ponto de iniciação de transcrição e perda da interação entre TFIIB e a RNApol II. Estudos estruturais indicam que a distância entre o TFIIB e o sítio catalítico da RNApol II é de aproximadamente 32 bases, a mesma distância encontrada entre o TATA-BOX e o TSP. O TDFIIB também é alvo de um grande número de ativadores de transcrição.

TFIIF: É responsável pela estabilização do complexo de pré iniciação de transcrição, bem como, supressão de ligações não específicas à fita de DNA. Possui também atividade quinase que pode estar envolvida em eventos de fosforilação da cauda CTD de forma direta ou indireta.

TFIIE: O fator TFIIE tem atividade quinase e também fosforila a cauda CTD da RNAPol II, adicionalmente também possui a habilidade de manter as fitas de DNA separadas. Foi constatado que TFIIE possui um motivo do tipo "*zinc ribbon*" que deve estar relacionado com a separação e/ou manutenção da abertura das fitas de DNA na região promotora.

TFIIH: Age como helicase que possibilita a abertura das fitas de DNA na região promotora, essencial para o início da transcrição. Possui uma subunidade a qual apresenta atividade de ciclina quinase e fosforila resíduos de serinas da cauda CTD da RNApol II. O TFIIH pode ser separado em dois subcomplexos, um com atividade de helicase, denominado de cerne TFIIH e o segundo subcomplexo com atividade ciclina quinase. O cerne TFIIH também é encontrado no complexo de reparo de excisão de nucleotídeos. Mutações nas helicases destes complexos levam a doenças genéticas humanas como síndrome de Cocayne e xeroderma pigmentosum.

1.1.4 Complexo de Pré-Iniciação (CPI)

A RNApol II associada com os GTFs compreende o Complexo de Pré-Iniciação (CPI), o qual é capaz de sustentar uma transcrição denominada basal. Existem duas vertentes que explicam a ligação do complexo basal de transcrição à região promotora [1, 12] Na primeira o complexo é formado de maneira ordenada e seqüencial. Inicialmente o TFIID se liga a região promotora seguido pela ligação do TFIIA que estabiliza este complexo. Após estes eventos a RNApol II é recrutada. O próximo fator a se ligar é o TFIIB, o qual interage com seqüências localizadas acima e abaixo ao TATA Box regulando a distância entre o complexo e o ponto de iniciação de transcrição. A este complexo se associa, então, o TFIIF seguida da associação, em seqüência, dos fatores TFIIE e TFIIH [1]. A segunda hipótese postula a existência de um complexo multiproteico, previamente estabelecido, denominado de holoenzima RNApol II, o qual se caracteriza por estar associada com diversos fatores gerais de transcrição (GTFs) bem como outras proteínas de caráter regulatório independente de sua associação com as regiões promotoras dos genes a qual transcreve [12]. Existem grandes divergências quanto a composição protéica da holoenzima RNApol II [5, 10, 12-14].

1.2 Elementos Globais Envolvidos em Ativação/Repressão de Transcrição

1.2.1 Complexo Srb (Mediador)

Na tentativa de se purificar um grupo de proteínas, denominadas SRBs (Supressor of RNA polymerase B), que tinham a capacidade de suprimir mutações na cauda CTD da RNApol II, foi isolado um grande complexo proteico que possuía a RNA pol II e os GTFs. O complexo holoenzimático purificado tinha a capacidade de iniciar a transcrição e responder à estímulos de ativadores de transcrição. Em contraste, frações altamente purificadas de RNApol II e GTFs não respondiam aos ativadores. Posteriormente, em um experimento onde a holoenzima foi separada

em seus três constituintes (RNApol II, GTFs e complexo SRB) com um anticorpo anti-CTD, foi demonstrado que a adição do complexo SRB em frações purificadas de RNApol II e GTFs, conferia ao complexo a capacidade de responder à ativadores de transcrição. Este complexo foi denominado de mediador de ativação ou complexo mediador [6, 14].

Muitos componentes do complexo mediador são essenciais para transcrição de diversos genes, provavelmente por fornecer alvos para ativadores de transcrição e neste processo fornecer sinais regulatórios para a RNA pol II e outros fatores envolvidos nos eventos de iniciação de transcrição. O complexo mediador é composto por até 26 proteínas dentre elas: Srb2, Srb4, Srb7, Srb9, Srb10, Srb11, Med1, Med2, Med4, Med6, Med7, Med11, Cse2, Gal11, Pgd1, Rgr1, Rox3 e Sin4. Estas proteínas são capazes de formar subcomplexos os quais modulam a transcrição de forma positiva (subcomplexo Srb2, Srb4, Srb5, Srb6, Med6 e Rox3) ou negativa (subcomplexo Srb8, Srb9, Srb10 e Srb11) [12].

1.2.2 Cromatina: elemento regulador global da atividade gênica

Em eucariotos a cromatina, o empacotamento fisiológico de DNA e histonas, é elemento de fundamental importância em eventos de ativação e repressão de transcrição. Já foi observado que promotores de genes induzíveis se encontram em regiões com alta densidade de nucleossomos, enquanto que promotores de genes constitutivos possuem um número reduzido de nucleossomos [15]. Ainda, deleções nas unidades nucleossomais levam à uma expressão constitutiva de genes fortemente regulados [16, 17]. Apesar das evidências descritas acima, somente estudos cristalográficos recentes da cromatina, permitiram um maior entendimento dos mecanismos de regulação. A unidade nucleossomal é um heteroctâmero composto por um tetrâmero H3/H4 e dois dímeros H2A-H2B envoltos por 146 nucleotídeos [18]. Estas proteínas formam uma estrutura cilíndrica composta de dois heterodímeros das histonas H3 e H4 flanqueados por dois dímeros das histonas H2A e H2B [19,20].

Os resultados cristalográficos permitiram o entendimento da organização dinâmica entre as histonas e o DNA, a qual esta relacionada com a função biológica da cromatina. Primeiro, o DNA se enovela fortemente em torno do heteroctâmero por meio de múltiplas interações entre as moléculas de fosfato e as histonas. A organização do DNA em volta do octâmero é flexível e permite interações entre as caudas N-terminais das histonas e a unidade nucleossomal como também de nucleossomos adjacentes o que permite uma flexibilidade no arranjo da cromatina e esta maleabilidade está diretamente envolvida em eventos de expressão gênica [21-24].

As caudas N-terminais das histonas são alvos altamente acessíveis no ambiente nuclear e podem ser atrativos para enzimas envolvidas na transdução de sinal. A estrutura dos nucleossomos pode ser modula por diversas modificações covalentes como: metilação, fosforilação, acetilação, ubiquitinação e adenilação, sendo que as três principais modificações covalentes das histonas, relacionadas com regulação gênica, são a metilação, a fosforilação e a acetilação [25-27]. Além das modificações covalentes descritas acima existem também as modificações não-covalentes, realizadas por complexos denominados "remodeladores de cromatina", os quais possuem a capacidade de remodelar a estrutura do nucleossomo às custas de ATP [28-32]. Esta modificação da partícula nucleossomal permite a interação de fatores de transcrição ao DNA, modulando assim, a atividade transcricional. Os principais elementos dos processos descritos acima são apresentados a seguir.

1.2.3 Modificações covalentes

Metilação de Histonas

A metilação de histonas leva, geralmente, a repressão de transcrição. Recentemente, foi demonstrado que a metilação da histona H3 leva a eventos de repressão da transcrição [27]. Apesar de já terem sido identificados um grande número de proteínas envolvidas nos processos de acetilação e fosforilação de histonas (eventos que levam ativação de transcrição), pouco se sabe sobre os fatores envolvidos na metilação de histonas. As primeiras histona metil transferases (HMT) foram descritas em triagens genéticas de *D. melanogaster* (*Su(var)3-9*) [33] e atualmente já foram identificados homólogos funcionais em *Schizosaccharomyces pombe* (*clr4*), ratos (*SUV39h1* e *SUV39h2*) e humanos (*SUV39H1*) [33,34].

Experimentos "*in vitro*" demonstram que a proteína Clr4p de *S. pombe* é capaz de metilar especificamente o resíduo 9 de serina nas histonas H3 e que a deleção deste gene leva a um padrão anômalo da organização da cromatina e defeitos de segregação na divisão celular [27]. Em ratos, a disrupção dos genes *SUV39h1* em somente um "*locus*" não impede sua viabilidade, entretanto, a disrupção do gene em ambos os alelos leva ao nascimento de somente 20-25% da prole e os indivíduos restantes apresentam um crescimento lento, demonstrando graves anormalidades cromossômicas em células germinativas, indicando sua importância também no desenvolvimento [35]. Apesar de exaustivos estudos, não foram descritos fatores com capacidade de histona demetilase (HDMase). Um possível mecanismo apontado para a remoção dos grupamentos metil é a proteólise da cauda N-terminal. Já foi demonstrado que a região amino

terminal das histonas H3 ubiquitinadas são suscetíveis à proteólise e são removidas em determinado momento do ciclo celular ou em estadios específicos do desenvolvimento [36,37].

Fosforilação de histonas

Um dos exemplos melhores caracterizados de fosforilação de histonas e ativação da transcrição foi descrito em células de mamíferos expostos a agentes mitogênicos ou estresse. Sob estas condições Mahadevan e colaboradores [38] demonstraram que a histona H3 é rapidamente fosforilada na serina 10 dos genes pp15 e pp16 (envolvidos na divisão celular) e esta modificação está associada à uma elevada ativação da transcrição.

Posteriormente foi demonstrado que a fosforilação da histona H3 está fortemente relacionada com a cascata efetivada pelas ERKs. De forma simples, quando expostos a agentes de crescimento as ERKs são ativadas (p42/ERK1 e P44/ERK2), através da via Ras/Raf, e induz proliferação como resposta [39]. Recentemente foi visto que a rsk2 está diretamente envolvida com a fosforilação da serina 10 das histonas H3 "*in vivo*". Células deficientes em rsk2 (Coffin-lowry Sindrome-CLS) são incapazes de ativar o gene c-fos e fosforilar a histona H3. É interessante frisar que a fosforilação ocorre normalmente durante a mitose, mesmo nas células deficientes (CLS e rsk2) indicando que existem outras quinases envolvidas nas modificações nucleossomais [25].

O mecanismo pelo qual a fosforilação de histonas influencia na ativação de transcrição ainda não é bem conhecido. Entretanto acredita-se que a adição de grupamentos fosfatos na porção N-terminal das histonas possa disromper as interações eletrostáticas entre as caudas básicas da H3 e o esqueleto de DNA. Outra explicação é que as modificações nas histonas servem como sinal para o recrutamento de fatores de transcrição e complexos regulatórios [25]. Entretanto, nenhum fator nuclear identificado é capaz se ligar exclusivamente em moléculas de H3 fosforiladas. Recentemente foi demonstrado em *S. cerevisiae* que durante mitose ocorre a fosforilação da histona H3 no resíduo de serina 10 e que esta modificação é efetuada pela quinase Ip1p, sendo que a defosforilação logo após o processo mitótico, é realizada pela fosfatase Glc7p [40].

Até recentemente acreditava-se que a porção C terminal das histonas estava fortemente associada com a manutenção da estrutura dos nucleossomos. Entretanto, Rogakou e colaboradores (1999), demonstraram que a fosforilação da serina 129, presente na região C-terminal da histona H2A *de S. cerevisiae* é um sinalizador de danos ao DNA [41]. Posteriormente foi constatado que na variante humana H2A.X a fosforilação da serina 129 coincide com o início da fragmentação do DNA [42]. Trabalhos recentes também correlacionam células humanas em

estado apoptótico e a fosforilação da porção C-terminal das histonas H2B, H3 e H1 [43,44], contudo ainda não foram identificadas quinases responsáveis pela fosforilação das serinas C-terminais destas histonas. A identificação destas quinases seria de grande valia nos estudos de morte celular programada.

Acetilação de Histonas

Dentre os processos de modificação da cromatina o mais estudado é o promovido pela acetilação de histonas, o qual está largamente envolvido na ativação gênica. As caudas N-terminais das histonas são ricas em resíduos de lisinas e, de uma forma geral, diversos trabalhos apontam que histonas hiperacetiladas estão associadas com alta atividade transcricional, por outro lado histonas hipoacetiladas são encontradas em regiões com diminutos níveis de transcrição [45]. É postulado que, assim como na fosforilação de serinas, a acetilação de lisinas pode levar a uma diminuição da interação entre DNA e histonas e/ou agir como sinal para o recrutamento de fatores de transcrição [46].

Peças fundamentais no processo de acetilação são representadas por um grupo de proteínas denominadas HATs (Histone Acetil Transferases). Dentre os diversos representantes desta família podemos destacar o fator Gcn5p de *Saccharomyces cerevisiae* [47,48]. Estudos "*in vitro*" demonstram que Gcn5p é capaz de acetilar histonas, mas não nucleossomos. Para que Gcn5p seja capaz de atuar sobre nucleossomos é necessário que ele forme um complexo, juntamente com os fatores Spt (Spt3p, Spt7p, Spt8p e Spt20p) e Ada (Ada2p e Ada3p), denominado SAGA (Spt, Ada e Gcn5 Acetiltransferases components). Atualmente já foram descritas proteínas de diversas espécies que guardam grande semelhança com Gcn5p e constituem a superfamília denominada de GNAT (Gcn5-related N AcetilTransferases) as quais se caracterizam por transferir o grupo Acetil da coenzima-A para porções N-terminais de proteínas [49]. Até o momento já foram identificados seis complexos protéicos com esta característica em *S. cerevisiae*, os quais são agrupados em duas classes: Tipo A: HATs que se localizam no núcleo e possuem afinidade por fatores nucleares (ADA, SAGA, NuA3 e NuA4); Tipo B: HATs citoplasmáticas. Acredita-se que os membros dessa classe possuam como função a acetilação de histonas recém sintetizadas (Hat1 e yTFIID). É importante salientar que cada complexo, mesmo pertencendo ao mesmo tipo, possui características peculiares quanto ao substrato. Por exemplo: o complexo SAGA possui maior afinidade pelas histonas H2B e H3 ao passo que o complexo NuA4 (Nucleosome Acetiltranferase 4) possui maior afinidade pelas histonas H2A e H4 [21].

A acetilação de lisinas é um processo reversível e já foram identificadas diversas proteínas denominadas HDAC (Histone DeACetilases) que são capazes de realizar o processo

inverso [6]. A primeira histona deacetilase identificada foi HDAC1 e por meio de fracionamento de extratos proteicos de *S. cerevisiae* foi possível identificar duas porções da HDAC1 com atividades distintas de deacetilase: HAD e HDB. No qual a unidade catalítica de cada complexo foi identificada como Hda1p (HAD) e Rpd3p (HDAB) [50]. Mutações nestas subunidades resultam em hiperacetilação dos resíduos de lisina das histonas H3 e H4. Assim como as HATs, as HDAC formam complexos multiproteicos, dentre eles Sin3 e NuRD [51]. Atualmente diversas proteínas capazes de desacetilar a cauda das histonas já foram identificadas. Recentemente demonstrou-se a proteína Sir2p, largamente envolvida em processos de silenciamento de transcrição, é uma deacetilase dependente de NAD, o que revela a existência de uma grande variedade de histonas deacetilases como também liga os processos de silenciamento e envelhecimento com a deacetilação [52].

É importante salientar que estas modificações transcricionais agem em concerto e servem como base de uma nova hipótese denominada código histônico [26]. Neste contexto as diferentes combinações nas modificações das histonas levam à intricandos mecanismos de regulação gênica e os diferentes estados podem levar a eventos interconectados para a modulação da atividade celular. Estudos demonstram que a afinidade das HATs e extremamente aumentada pelas histonas H3 quando estas apresentam o resíduo de serina 10 fosforilado. Já foi constatado que a fosforilação da serina 9 de H3 além de impedir a metilação da lisina 9 de H4, facilita a acetilação das lisinas 9 e/ou 14 de H4, o que indica que estes processos são cooperativos e que as diferentes possibilidades de combinações das modificações covalentes nas oito caudas terminais histônicas presentes nas unidades nucleossomais levam a uma complexa modulação nos eventos de ativação e repressão gênica [26, 53].

1.2.4 Modificações não-covalentes

Adicionalmente aos processos de modificações covalentes da cromatina, existe um segundo grupo de complexos proteicos capazes de modificar a estrutura da cromatina de forma não covalente. Estes possuem a capacidade de deslocar os nucleossomos de forma progressiva, as custas de ATP, de modo a enfraquecer ou mesmo romper a interação entre DNA e histonas. Já foi demonstrado "*in vitro*" que estes complexos remodeladores de cromatina podem deslocar octâmero histônico para a mesma molécula de DNA ("*cis*") ou até mesmo para outra molécula de DNA adjacente ("*trans*") [54].

O primeiro complexo remodelador identificado foi o SWI/SNF de *S. cerevisiae*, inicialmente descrito em mutantes para troca de "mating type" e incapazes de metabolizar sacarose (mating type SWItching/ Sucrose Non Fermenting). É composto por 11 diferentes

fatores, incluindo Swi1/Adr6, Snf2/Swi2, Swi3, Swi5, Swi6, Swi11, Swp73/Snf12 e Tfg3/TAF30/Anc1 [32, 55]. O complexo SWI/SNF age promovendo a disrupção da estrutura do nucleossomo, na presença de ATP, por meio da subunidade Swi2p/Snf2p, uma ATPase estimulada por DNA [56, 57].

Já foram descritos diversos complexos remodeladores de cromatina que podem ser divididos em duas classes: I) Membros pertencentes à família SWI/SNF dos quais podemos destacar SWI/SNF e RSC de levedura, dSWI/SNF e Complexo Brahma de *Drosophila*, hSWI/SNF, hBrahma e NURD de *Homo sapiens*, os quais se caracterizam por possuir tanto afinidade por DNA livre de nucleossomos (*naked DNA*) como por arranjos nucleossomais de DNA. II) A família ISWI (Imitation of SWI) onde se classificam os complexos ISWI e ISWI2 de leveduras, os complexos NURF, ACF e CHRAC de *Drosophila* e o complexo RSF de humanos que apresentam como propriedade a afinidade somente por DNA nucleossomal [24]. Com exceção do complexo NURD, o qual possui uma deacetilase e está envolvido em eventos de repressão. De forma geral todos estes complexos são capazes de estimular a transcrição.

Este processo torna acessíveis os elementos "*cis*" presentes nos promotores aos fatores de transcrição, que podem agir ativando ou reprimindo a transcrição [54]. Foi demonstrado através de experimentos de "DNA microarray", contendo 5626 ORFs da levedura *Saccharomyces cerevisiae,* que cerca de 126 genes são regulados de forma positiva por SWI/SNF ao passo que 203 são regulados de forma negativa [58]. Apesar de agir modulando a transcrição de um reduzido número de genes (menos de 10% das ORFs analisadas), possui um papel fundamental na regulação da transcrição de genes sujeitos a repressão catabólica visto que a sua disrupção leva a diminuição da transcrição de genes envolvidos na degradação de formas alternativas de carbono como sacarose, rafinose ou galactose [58].

Através de comparações de similaridades entre seqüências de DNA de diferentes organismos presentes em bancos de dados genéticos, é possível constatar a presença de diversas seqüências que possuem alta similaridade com muitas proteínas do complexo SWI/SNF, sendo hoje considerada uma superfamília de complexos protéicos característicos de eucariotos os quais possuem como principal função remodelar a cromatina de forma permitir uma acurada regulação da transcrição [32, 54, 57].

O direcionamento dos complexos remodeladores aos promotores é alvo de intenso estudo. Dois modelos, provavelmente complementares, mostram que o complexo SWI/SNF pode ser recrutado às regiões promotoras por dois mecanismos: a) por ativadores ou repressores transcricionais que interagem com proteínas distintas do complexo SWI/SNF1 [59]; b) pelo próprio estado de acetilação da partícula nucleossomal. Já foi demonstrado que o complexo

SWI/SNF possui grande afinidade por histonas acetiladas e que o elemento responsável por este reconhecimento é a proteína Snf2p por meio de uma região denominada bromodomínio [60]. Em complementaridade com estas observações já foi determinado que eventos de acetilação e remodelamento parecem estar conectados, em alguns casos, sugerindo uma íntima relação entre os complexos acetiladores e remodeladores de cromatina [48, 60]. Uma observação importante reside no fato de que algumas proteínas identificadas primeiramente como pertencentes ao complexo SWI/SNF, são na realidade componentes de complexos acetiladores como ADA, SAGA e Gcn5 [61, 62]. Ainda, estudos demonstram que em diversas situações ambos os complexos (acetiladores e remodeladores) estão associado holoenzima RNApol II [14,63], evidenciando uma ligação direta entre acetilação, remodelamento de cromatina e regulação da transcrição [64].

1.2.5 Elementos modificadores agindo como ativadores e repressores de transcrição pontuais ou classe específicos

Além dos elementos descritos anteriormente existem fatores de transcrição que modulam o processo transcricional de forma positiva ou negativa. Estes podem atuar somente em um determinado gene ou mesmo sobre classes e famílias gênicas e são denominados de ativadores ou repressores de transcrição. A forma de ação destes reguladores de transcrição envolve uma interação direta com elementos da maquinaria basal de transcrição (RNApol, GTFs), o complexo mediador, complexos que alteram a estrutura da cromatina (SWI/SNF – STH1 e SAGA) e a própria cromatina. Alguns elementos serão apresentados a seguir, tomando como exemplo um dos sistemas de transcrição mais bem caracterizados: a regulação por glicose.

1.3 Regulação por Glicose

Em leveduras diversos sistemas gênicos são controlados em nível de iniciação da transcrição. Dentre eles a regulação por glicose ou catabólica, é um dos mais bem descritos sistemas gênicos controlados em nível transcricional [65].

Em *Saccharomyces cerevisiae*, organismo modelo para estudos genéticos e bioquímicos, diversos fatores que controlam a transcrição de genes reprimidos por glicose, de forma direta ou indireta, já se encontram caracterizados. Alguns destes fatores agem de forma pontual sobre determinados genes de uma via catabólica, enquanto outros atuam de modo global sobre diversos conjuntos de genes distintos. Ainda, diversos fatores são altamente conservados desde eucariotos inferiores até os superiores, como plantas e mamíferos, o que evidencia uma notável importância

biológica sobre determinadas vias de regulação gênica. Dentre os elementos descritos mais importantes, podemos destacar os complexos SWI/SNF e SNF1 os quais respondem por eventos de ativação de transcrição e o complexo formado por Mig1p/Ssn6p/Tup1p, o qual é responsável pela repressão da transcrição de diversos genes sujeitos a repressão catabólica e serão abordados a seguir.

1.3.1 Complexo SNF1: Ativador de transcrição de genes controlados por glicose

O complexo protéico SNF1 de *S. cerevisiae* é considerado um regulador global do metabolismo de carboidratos. Sua estrutura e mecanismo molecular de ação são extremamente conservados entre os eucariotos e já foram descritos representantes homólogos em diversos organismos como fungos, plantas e mamíferos [66,67] e atualmente estes elementos são agrupados em uma família de proteínas denominada Snf1-AMPK [68,69].

Os complexos enzimáticos SNF1-AMPK são heterotrímeros que compreendem uma subunidade catalítica α, e duas subunidades regulatórias denominadas β e γ. Em *S. cerevisiae* é um heterotrímero sendo que a unidade catalítica é representada por Snf1p (α), uma serina-treonina quinase largamente conservada em eucariotos inferiores e superiores [66, 70], Snf4p que representa a unidade γ e se caracteriza por ser uma proteína que possui um domínio carboxiterminal denominado ASC, o qual interage com diversos adaptadores que caracterizam as subunidades β (Sip1p, Sip2p ou Gal83p), e formam diferentes subcomplexos que modulam a atividade e especificidade de SNF1 [65, 71-74]. O complexo SNF1 é capaz de neutralizar a repressão de vários "*loci*" controlados por glicose [75]. SNF1 é capaz de ativar a transcrição gênica de diversas formas: i) Através da ativação de fatores de transcrição por meio de fosforilação, como é o exemplo de Sip4 e Cat8 os quais estão envolvidos na utilização de fontes não fermentáveis de carbono [76-78]; ii) Interagindo com elementos do complexo mediador da holoenzima RNApol II [79]; iii) Modificando a cromatina através de fosforilação da Serina 10 de histonas H3 [80] iv): Fosforilando a proteína repressora Mig1p em resíduos de serina, o que anula a afinidade pelos promotores aos quais se liga e desfaz o complexo Mig1p-Ssn6p-Tup1p [65, 81, 82].

1.3.2 Complexo repressor Mig1p-Tup1p-Ssn6p

Um dos complexos protéicos de repressão catabólica mais bem caracterizados em leveduras é o formado pelos fatores Mig1p, Tup1p e Ssn6p. O primeiro elemento reconhecido foi

Mig1 em trabalhos visando reconhecer genes que controlassem a expressão do gene *Gal1* [71], posteriormente foi também isolado como supressor de mutações dos genes snf1 e snf4 [83, 84].

Mig1

Em *Saccharomyces cerevisiae* o fator Mig1p possui um papel fundamental na repressão de genes por glicose como: *Gal1* e *Gal4* (galactose), *Mel1* (melibiose); *Suc2* (sacarose), *Mal63* (maltose) e *Cat8* (ativação de enzimas gliconeogênicas) [65, 71, 85, 86]. Mig1p age através de interação direta com regiões ricas em GC, dos promotores aos quais reprime, por meio de dois motivos "*zinc finger*" do tipo C_2H_2, localizados próximos à porção N-terminal [87]. Entretanto a afinidade de Mig1p pelas regiões "*cis*" atuantes dos promotores está diretamente ligada ao estado de fosforilação no qual se encontra. Já foi demonstrado que em situações de repressão, Mig1p se apresenta defosforilado e se localiza no núcleo ao passo que em ausência de glicose Mig1p se encontra fosforilado e sua localização é citoplasmática [88].

Estudos recentes indicam que Mig1p é fosforilado por SNF1, em ao menos três resíduos de serina (Ser-278, Ser-311 e Ser-381) o que leva a dois eventos simultâneos: perda da afinidade de ligação ao DNA e mudança de sua localização celular do núcleo para o citoplasma [81]. Os eventos existentes neste processo não estão bem caracterizados, entretanto há evidências que a exportina Msn5p, esteja envolvida na translocação núcleo-citoplasma de Mig1p [88]. Trabalhos recentes demonstram que a proteína fosfatase 1 (PF1) a qual é composta por duas subunidades, Reg1p, a qual representa a unidade regulatória, e Glc7p a unidade catalítica, como a responsável pela defosforilação e restabelecimento de suas funções como repressor [65, 79]. Genes similares a *Mig1* já foram clonados nas leveduras *K. lactis*, *K. marxianus*, *S. pombe* e *C. albicans* [66, 89, 90].

Ssn6-Tup1

Para que haja a inibição da transcrição, é necessário que Mig1p recrute o subcomplexo Ssn6p-Tup1p [72]. Neste contexto, Mig1p é responsável pelo reconhecimento do sítio de ligação com o DNA, a proteína Ssn6p desempenha o papel de adaptador entre Mig1p e Tup1p, sendo Tup1p responsável pela interação de repressão [91, 92]. É importante salientar que a proteína Mig1p está intrinsecamente ligada com a repressão catabólica ao passo que o complexo corepressor formado pelas proteínas Ssn6p-Tup1p está relacionado com a repressão de diversos grupos genes que participam de numerosos processos celulares. Calcula-se que o subcomplexo Ssn6-Tup1 seja responsável pela repressão de até 3% dos genes de *S. cerevisiae* [84]. Graças à habilidade do sistema Tup1p-Ssn6p de controlar a expressão de vários genes, mutações nestes fatores levam a efeitos pleiotrópicos [84, 93-95]. Estudos estequiométricos demonstram que o sub-complexo Ssn6p-Tup1p constitui um heteropentâmero formado por uma molécula de Ssn6p e

quatro subunidades de Tup1p [96]. Este complexo apresenta diversos domínios proteicos intrinsecamente ligados ao seu papel biológico [13, 97-99].

Ssn6p de *Saccharomyces cerevisiae* possui dez repetições adjacentes de um motivo denominado de TPR (TetratricoPeptide Repeat) de trinta e quatro aminoácidos, rico em resíduos hidrofóbicos, descrito primeiramente na proteína reguladora do ciclo celular Cdc23p e posteriormente identificados em diversas proteínas [100, 101]. A função destas repetições está associada com interações proteína-proteína, sendo que a associação preferencial se dá com motivos do tipo WD. Já foi demonstrado que as repetições TPR1-3 são essenciais para a interação entre Ssn6p e Tup1p [102], ao passo que combinações dos TPRs 4-10, mediam interações com diferentes proteínas repressoras, específicas para cada família gênica regulada pelo complexo Ssn6p-Tup1p [97].

A proteína Tup1p, pertence à uma família de proteínas que se caracteriza por possuir domínios denominados WD40, nome que reflete os aminoácidos comumente encontrados no motivo e o tamanho da unidade repetitiva [99]. Elementos semelhantes à *Tup1* de *S. cerevisiae* já foram descritos desde leveduras até humanos sendo que o número de unidades WD pode variar 5-7 repetições [103]. Já foi demonstrado em *S. cerevisiae* que a integridade dos sete motivos WD é necessária para a manutenção da atividade biológica de Tup1p e que a deleção de qualquer uma das repetições WD leva a disrupção da repressão dos genes mediada pelo complexo Ssn6p-Tup1p [92, 97].

O modo de ação do complexo Ssn6p-Tup1p já foi muito estudado e já foram descritos três mecanismos distintos, não excludentes:

i) Interferência com ativadores de transcrição: O complexo Mig1p-Ssn6p-Tup1p inibe a ativação de transcrição dos genes necessários para o metabolismo de galactose ocluindo regiões UAS reconhecidas pelo ativador Gal4p [104, 105].

ii) Interações diretas com a cromatina: Já foi demonstrado que complexos repressores que possuem Ssn6p-Tup1p promovem o remodelamento de cromatina na região promotora de duas classes de genes reprimidos por Ssn6p-Tup1p (reparo de DNA e tipo sexual). Ainda, experimentos "*in vitro*" apontam que Tup1p é capaz de interagir fisicamente com as caudas N-terminais das histonas H3 e H4 e que deleções ou mutações nestas regiões, em alguns casos, promovem a desrepressão de genes sujeitos à ação promovida por Ssn6p-Tup1p [22, 98, 105].

iii) Interagindo com componentes Holoenzima RNApol II: As evidências que suportam esta linha mostram que mutações em subunidades do complexo mediador da holoenzima RNApol II (Srb8, Srb9, Srb10, Srb11, Sin4, Rgr1, Rox3 e Hrs1) enfraquecem o efeito do complexo

repressor [13]. Experimentos "*in vitro*" apontam que Ssn6p-Tup1p, impede a ligação do TPB aos promotores [79]. A figura 1 ilustra os processos citados acima.

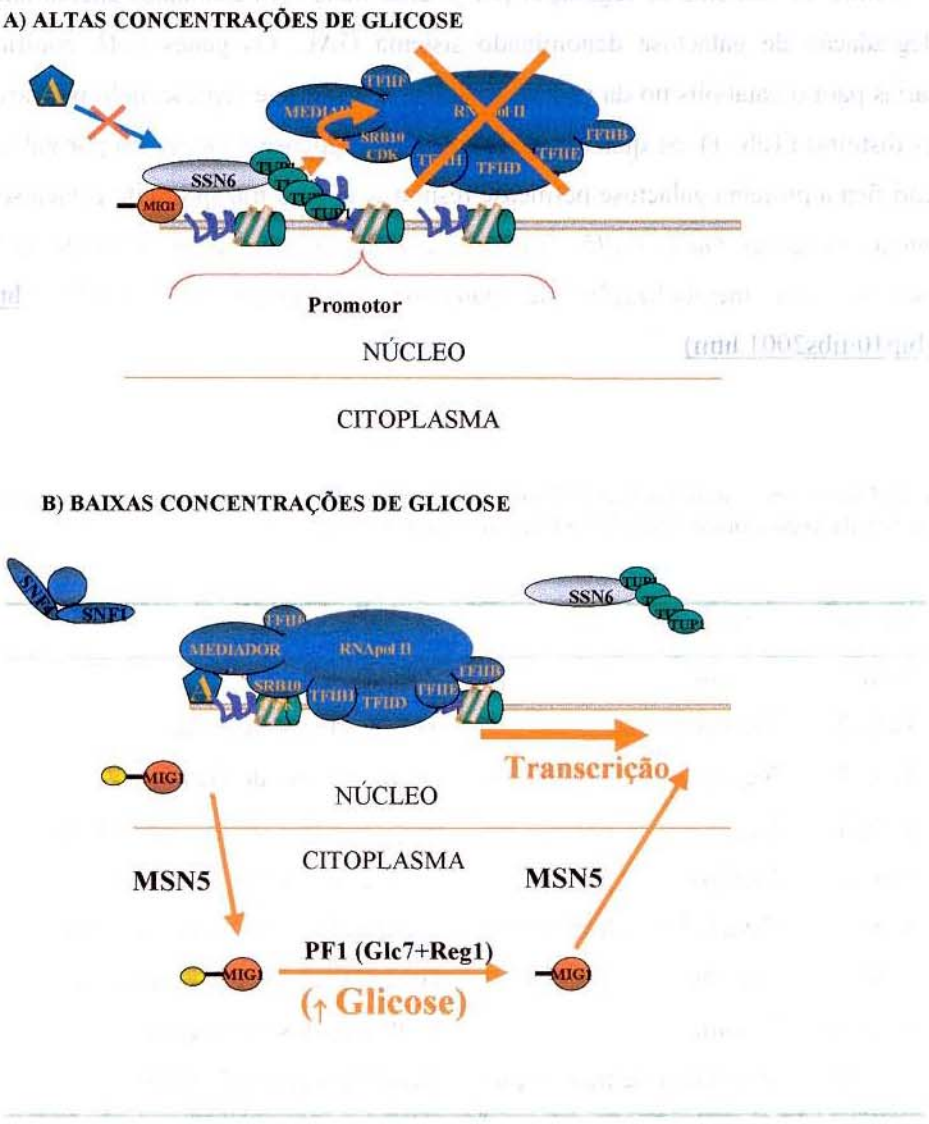


Figura 1: Diagrama de regulação por glicose envolvendo os complexos Mig1p-Ssn6p-Tup1p e SNF1. A) Em altas concentrações de glicose a proteína "*zinc finger*" Mig1 se encontra defosforilada e complexada com os fatores Tup1p-Ssn6p podendo agir de três formas: ocluindo regiões necessárias para o recrutamento de ativadores de transcrição, interagindo com histonas H3 e H4 e com elementos da holoenzima RNA polimerase. B) Em baixas concentrações de glicose Mig1p é fosforilado por SNF1, levando a dissociação do complexo Mig1p-Tup1p-Ssn6p o que permite o recrutamento de ativadores e da holoenzima RNA polimerase e consequente ativação da transcrição. O fator Mig1p fosofrilado é transportado para o citoplasma por MSN5 e quando a concentração de glicose se torna elevada Mig1p é defosforilado pelo complexo PF1 e retorna para o núcleo reestabelecendo a repressão.

1.4 Sistemas de repressão por glicose: elementos gerais, classe específicos e pontuais.

1.4.1 O sistema *GAL*

Dentre os modelos de regulação por glicose mais bem estudados encontramos o sistema para degradacão de galactose denominado sistema GAL. Os genes *GAL* codificam enzimas necessárias para o catabolismo da galactose, em *S. cerevisiae* é representado por nove genes com funções distintas (Tab. 1), os quais são reprimidos por glicose e induzidos por galactose. O gene *Gal2* codifica a proteína galactose permease responsável pelo transporte de galactose para o meio intracelular, os genes *Gal1*, *Gal5*, *Gal7* e *Gal10* codificam as enzimas do ciclo de Leloir responsáveis pela metabolização de galactose em glicose-6-P (Fig2). (http://bip.cnrs-mrs.fr/bip10/tibs2001.htm).

Tabela 1: Genes envolvidos no metabolismo de galactose em *S. cerevisiae* e seu respectivo produto e função (compilado de Cornish-Bowden e Cardeñas, 2000) [106].

GENE	PRODUTO	FUNÇÃO
GAL1	Enzima	Galactoquinase
GAL2	Transportador	Transportador de galactose
GAL3	Regulador de transcrição	Anula o efeito de Gal80p
GAL4	Regulador de transcrição	Ativa a expressão dos genes GAL
GAL5	Enzima	Fosfoglicomutase
GAL6	Regulador de transcrição	Age na via de resistência as drogas
GAL7	Enzima	Gal1-P Glc1-P uridiltransferase
GAL10	Enzima	UDP galactose epimerase
GAL80	Regulador de transcrição	Reprime o efeito de Gal4p

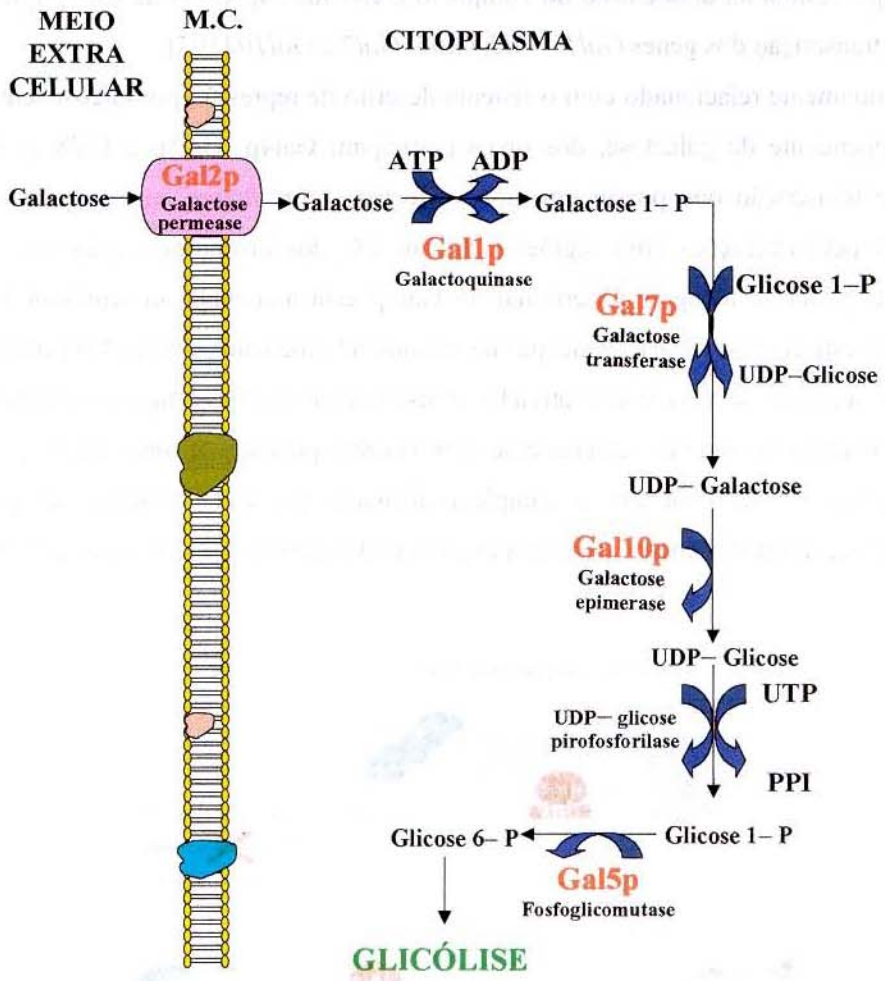


Figura 2: Metabolismo de galactose (ciclo de Leloir). As proteínas necessárias para a degradacão de galactose são codificadas pelos genes *Gal1*, *Gal2*, *Gal5*, *Gal7* e *Gal10*.

Os genes que codificam as proteínas necessárias para a degradacão de galactose estão sob controle de um rígido sistema de ativação/repressão da transcrição nos quais estão envolvidos os fatores de transcrição específicos para os genes *GAL*: Gal3p, Gal4p e Gal80p bem como os complexos Mig1p-Ssn6p-Tup1p e SNF1.

1.4.3 Controle da transcrição dos genes *GAL*

A transcrição dos genes *GAL* esta sujeito a um duplo controle de expressão: são reprimidos por glicose e induzidos por galactose [65]. O fator Gal4p é um ativador de transcrição capaz de ativar a sua própria transcrição bem como de todos os genes *GAL*. Quando a glicose está presente no meio o complexo Mig1p-Ssn6p-Tup1p se liga a uma região URS de *Gal4* diminuindo sua transcrição, ao passo que em baixos níveis de glicose o complexo SNF1 fosforila a proteína

Mig1p, o que resulta na dissociação do complexo e elevada expressão de Gal4p que por sua vez irá ativar a transcrição dos genes *Gal1*, *Gal2*, *Gal5*, *Gal7* e *Gal10* [107].

Intimamente relacionado com o sistema descrito de repressão por glicose temos o sistema ativador dependente de galactose, dos quais participam Gal4p, Gal3p e Gal80p. Gal4p é um ativador de transcrição que possui um motivo do tipo "*zinc finger*" na sua porção N-terminal responsável pela interações com regiões ricas em CG dos promotores com que interage. Na ausência de galactose a região C-terminal de Gal4p está associada ao repressor transcricional Gal80p. Sob este contexto Gal4p é incapaz de recrutar a holoenzima RNApol II [108].

Na presença de galactose o ativador transcricional Gal3p se liga ao complexo formado por Gal4p-Gal80p, interagindo diretamente com Gal80p provocando uma mudança na estrutura deste complexo. Neste contexto, o complexo formado por Gal4p-Gal80p-Gal3p é capaz de recrutar a holoenzima RNApol II e ativar a expressão dos genes *GAL* em cerca de 1000 vezes [58, 65].

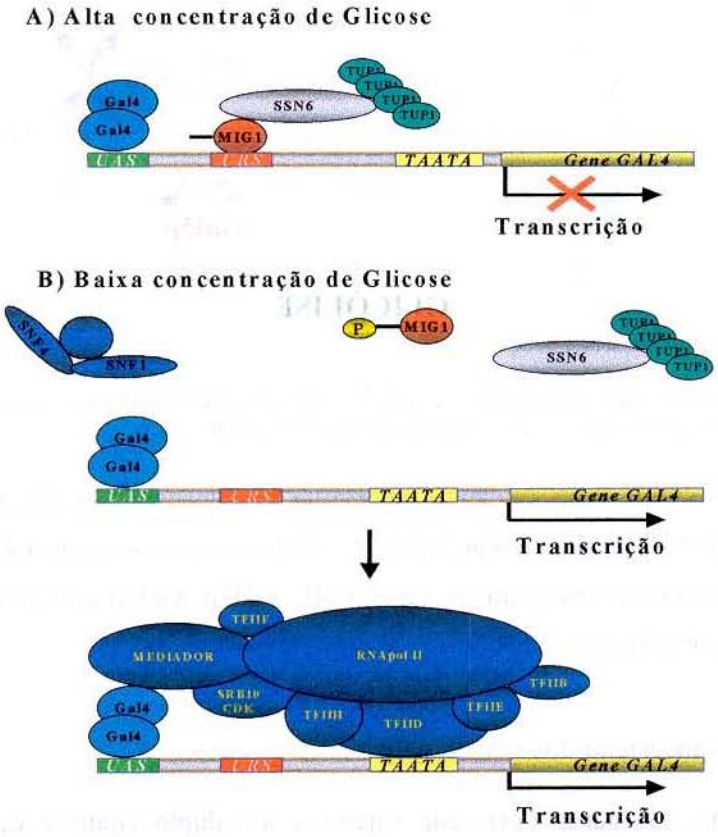


Figura 3: Regulação por glicose do gene *Gal4*. A) Na presença de glicose o produto do gene *Gal4* é transcrito em pequenas quantidades graças a repressão efetuada pelo sistema Mig1p-Tup1p-Ssn6p. B) Em baixos níveis de glicose o complexo SNF1 fosforila Mig1p o que leva a dissociação do complexo repressor e conseqüente transcrição através do recrutamento da holoenzima RNApol II pelo ativador de transcrição Gal4p.

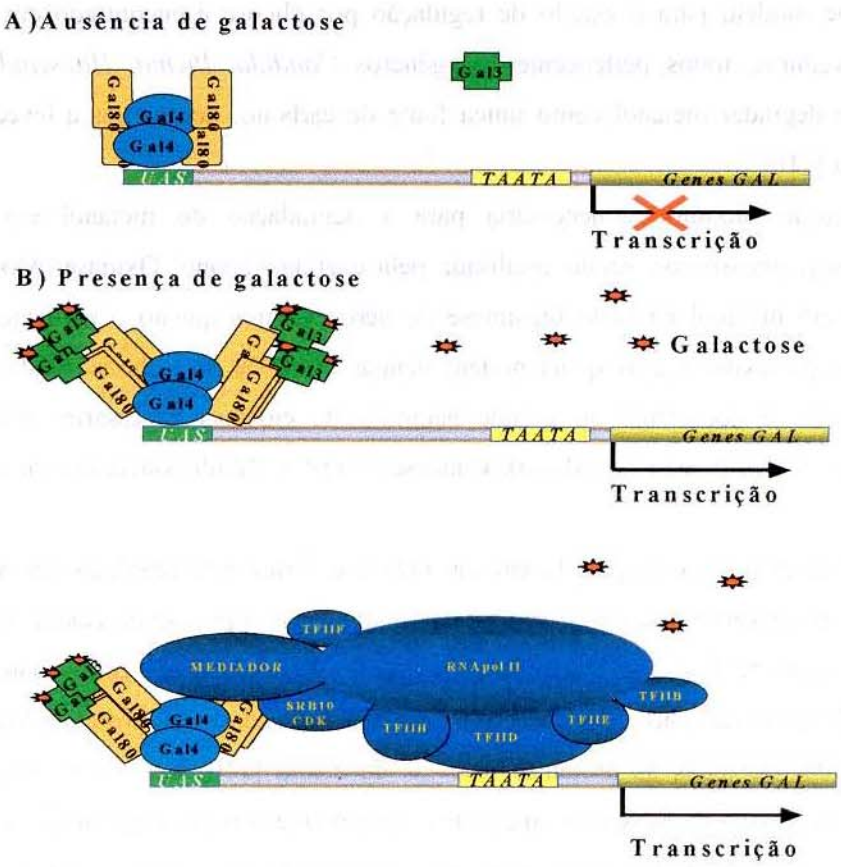


Figura 4: Indução de transcrição dos genes *GAL* por galactose. A) Na ausência de galactose a atividade de Gal4p é bloqueada pelo fator Gal80p. B) Em presença de galactose o fator Gal3p se liga a Gal80p o que leva à uma mudança de estrutura do complexo Gal4p-Gal80p, permitindo o recrutamento da holoenzima RNApol II e conseqüente transcrição.

Adicionalmente, já foi demonstrado que quando Gal4p ativa a transcrição ele se encontra fosforilado na serina 699 e que esta fosforilação é essencial para a completa indução dos genes *GAL* pois estabiliza a conformação ativa do complexo Gal4p-Gal80p-Gal3p [109]. Estudos realizados por Rohde e colaboradores [108] demonstraram que o fator que responde pela fosforilação da serina 699 de Gal4p é o componente do complexo mediador Srb10p, indicando uma íntima relação entre o ativador transcricional Gal4p e a holoenzima RNApol II.

1.5 Controle de transcrição do gene *Mox* de *Hansenula polymorpha*

Um interessante modelo para o estudo de regulação por glicose é encontrado em um pequeno grupo de leveduras, todos pertencentes ao gêneros *Candida*, *Pichia*, *Hansenula* e *Torulopsis*, capazes de degradar metanol como única fonte de carbono, dentre elas a levedura *Hansenula polymorpha* [110].

A primeira reação bioquímica necessária para a degradação do metanol em *H. polymorpha* ocorre nos peroxissomos, sendo catalisada pela enzima Metanol Oxidase (Moxp). Em células crescendo em metanol há tanto biosíntese de peroxissomos quanto o aumento do tamanho das organelas pré-existentes as quais podem ocupar ate 80% do volume celular. O aumento destas organelas é decorrente do grande acúmulo de enzimas necessárias para a degradação do metanol: Metanol oxidase (Moxp), Catalase (Catp) e Dihidroxiacetona sintetase (Dasp) [110,111].

O gene responsável pela produção da enzima (*Mox*) é fortemente regulado em nível transcricional. Quando esta levedura cresce em meio fermentativo, como glicose ou etanol, não é detectado nenhum transcrito de *Mox*. Entretanto, quando esta levedura cresce em meio contendo metanol como única fonte de carbono a transcrição é fortemente ativada e seu produto (Moxp) pode representar até 30% de todas as proteínas solúveis da célula [110,112]. Outro aspecto importante é o fato da expressão de Moxp ser co-regulada juntamente com a expressão de outras proteínas peroxissomais e também por apresentar uma íntima relação com a biogênese de peroxissomos [110].

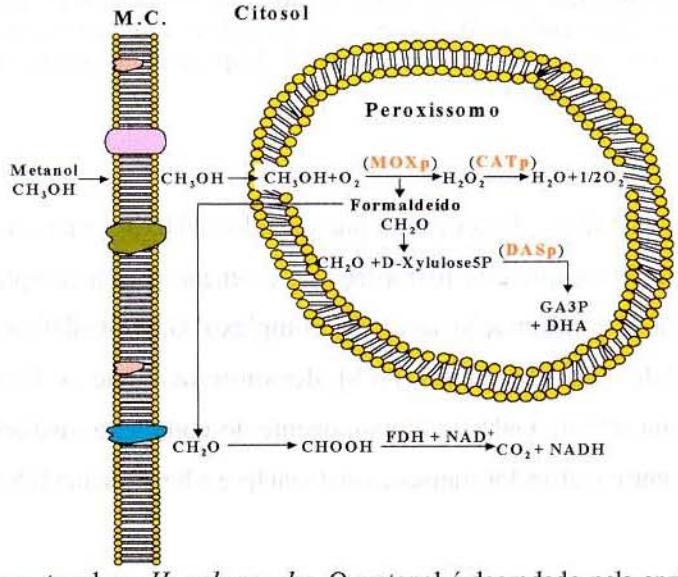


Figura 5: Degradação de metanol em *H. polymorpha*. O metanol é degradado pela enzima Moxp gerando H_2O_2, o qual é degradado pela enzima Catp, e formaldeído que é condensado juntamente com D-xilulose5P pela enzima Dasp gerando como produto GA3P e DHA que serão metabolizados na glicólise.

A estrutura da região promotora responsável pelo controle da transcrição de *Mox* foi bastante estudada e importantes informações já foram descritas [113, 114]. A região promotora *Mox* pode ser divida em três partes funcionais (*Mox A*, *Mox B* e *Mox C*) cada uma contendo elementos *cis* essenciais para um acurado mecanismo de transcrição. Dentre os elementos identificados podemos destacar: dois sítios de ligação para elementos envolvidos na repressão de transcrição (URS1, URS2); dois sítios de ativação da transcrição (UAS1 e UAS2) e dois pontos de iniciação de transcrição (TSP1 e TSP2) (Fig. 6A). Um mecanismo de regulação de *Mox* ainda pouco entendido concerne aos seus dois pontos de iniciação de transcrição (TSPs). O primeiro TSP se localiza –425, na região *Mox A*, sendo transcrito em células crescendo em glicose e em quantidades ínfimas. O segundo se localiza a –25 (*Mox C*) sendo transcrito em altas quantidades em células crescendo em metanol e irá dar origem a Moxp.

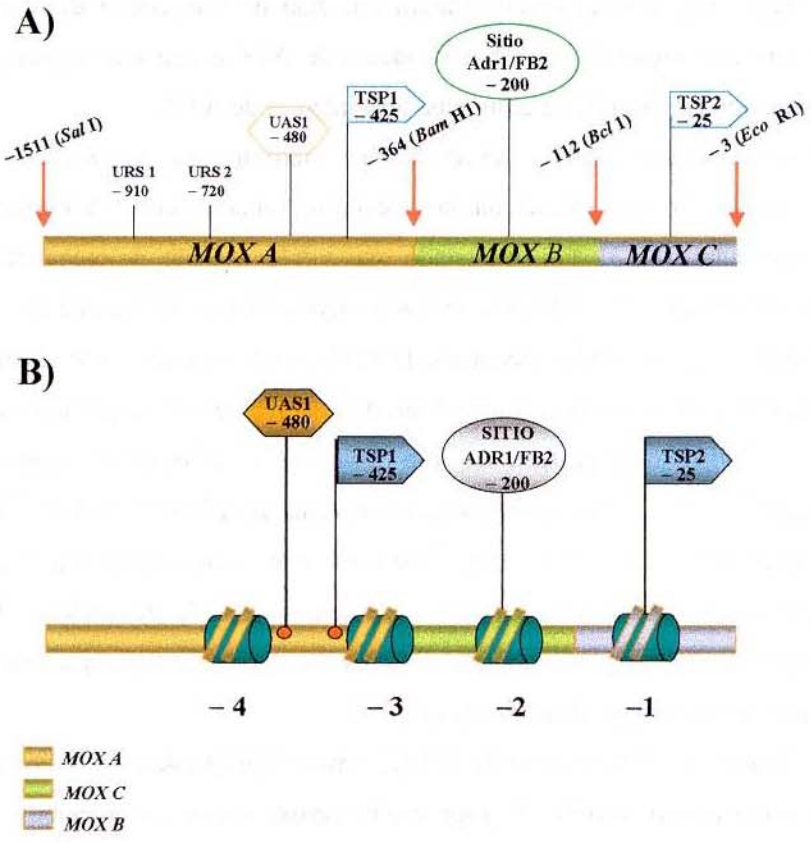


Figura 6: Estrutura do promotor *Mox*. A) O promotor *Mox*: A região que responde pelo controle da transcrição do gene *Mox* compreende 1511 pb "*upstream*" ao ponto de iniciação de transcrição (TSP1) e pode ser dividido em três regiões, as quais possuem elementos cis atuantes distintos. B) Posicionamento de nucleossomos no promotor *Mox*: Em células crescendo em glicose o posicionamento de nucleossomos oclui o sitio de ligação para o fator Adr1p /FB2 e o TSP2 e deixa exposto TSP1 podendo assim permitir a transcrição de pré *Mox*.

Este mecanismo foi analisado em detalhes e gerou as seguintes conclusões: O transcrito longo não é traduzido e não possui função aparente. A explicação para sua existência reside na estrutura do promotor *Mox*. O TATA box que controla sua síntese se encontra em uma região livre de nucleossomos. Ao passo que o TSP2 gera um transcrito curto que por sua vez irá dar origem a Moxp, se encontra em região onde há posicionamento de nucleossomos (Fig.6B) [111, 115].

A análise da estrutura da cromatina através de "*footprinting in vivo*" do promotor *Mox* demonstrou que há modificação no arranjo de nucleossomos quando este passa de um estado reprimido para ativado [111, 115]. Ainda, experimentos de análise de expressão do gene *Mox* por "*northern blotting*" no qual as células foram induzidas em meios contendo DNP e antimicina A (inibidores da cadeia respiratória) demonstraram que não há transcrição de *Mox* [116]. Estes dados sugerem que a ativação de *Mox* é dependente de ATP e que sua transcrição pode estar sujeita a complexos remodeladores de cromatina dependentes de ATP.

Trabalhos realizados com a levedura não metilotrófica *Saccharomyces cerevisiae* mostraram que o promotor *Mox* é funcional nesta espécie, sendo sujeito a repressão por glicose e ativação pelo fator Adr1p [113]. Foi constatado que Adr1p é capaz de reconhecer um sítio de ligação na região promotora *Mox B* [113] entre as posições -361 e -112, e que este sítio é alvo de um suposto fator de *H. polymorpha*, denominado FB2. Sendo este um forte candidato a ser um homólogo ao fator Adr1p. A proteína Adr1p é um dos fatores de transcrição mais bem estudados em levedura [117-119]. É uma proteína do tipo "*zinc finger*" e foi primeiramente descrita por ativar a transcrição do gene *ADH2*, que codifica a proteína citoplasmática Álcool Desidrogenase 2, a qual é regulada por glicose [118, 119]. Postula-se que a atuação de Adr1p seja através de interações com diversos fatores de transcrição e elementos da CPI, incluindo o TFIIB E TFIID, por meio de seus múltiplos sítios de ativação e repressão, o que conduz a um remodelamento da cromatina e conseqüente ativação da transcrição [120].

] Além de atuar na desrepressão de *ADH2*, o fator Adr1p está envolvido na regulação da expressão de enzimas peroxissomais, na biogênese de peroxissomos e no metabolismo de glicerol [119, 121]. Esses fatos levaram à sugestão de que o fator Adr1p e sua função no metabolismo peroxissomal estejam conservados em outras espécies [113].

Outro aspecto relevante é o fato do promotor *Mox* largamente utilizado para a produção de proteínas heterólogas de interesse farmacológico e industrial [122-124] e informações sobre fatores que controlam sua transcrição podem auxiliar na otimização dos processos de produção.

2. OBJETIVOS

Identificação de possíveis fatores de transcrição envolvidos na ativação ou repressão catabólica do gene *Mox*. Para tanto, foram escolhidos como candidatos fatores de transcrição *Snf1*, *Swi2*, *Adr1*, *Ssn6*, *Mig1* e *Tup1* bem estudados em *S. cerevisiae* e largamente envolvidos em processos de repressão catabólica.

Análise do envolvimento dos fatores identificados na manutenção dos mecanismos de repressão catabólica presentes no metabolismo de metanol por meio de disrupção gênica dos genes identificados e investigação dos níveis de transcrição de genes envolvidos no metabolismo de metanol.

Criação de banco de dados genéticos e ferramentas de bioinformática para a identificação de elementos relevantes presentes em genomas. Aplicação desta tecnologia em RSTs de *H. polymorpha*.

3. JUSTIFICATIVA

O controle da transcrição do gene *Mox* é fortemente regulado pela fonte de carbono disponível. O esclarecimento dos mecanismos que regem ou influenciam em sua regulação podem contribuir para o entendimento do controle da transcrição de genes regulados por glicose. Ainda, esta levedura é largamente utilizada para a produção de proteínas heterólogas de interesse comercial e informações sobre fatores que controlam sua transcrição e de novos elementos presentes em seu genoma podem auxiliar na otimização dos processos de produção, como também em um maior entendimento de sua biologia.

4. MATERIAIS E MÉTODOS

4.1 Reagentes

Agarose, brometo de etídeo, SDS, X-gal, poliacrilamida, TEMED (Invitrogen); antibióticos, tritom X-100, peptona, triptona, extrato de levedura, etilenoglicol, DMF, etanol, metanol (Merck); membrana de nylon (Pharmacia); oligonucleotídeos (Operon); nucleotídeos, enzimas de restrição, *Taq* DNA polimerase, *Pfu* DNA polimerase, Klenow, T4 DNA ligase foram adquiridas junto a Invitrogen, Promega ou Pharmacia.

4.2 Linhagens e plasmídeos

Bactérias

DH5αF' - F' (Z80dlacZ_(lacZ)M15)_(lacZYA-argF)U169 recA1 endA1 hsdR17(r_k^-,m_k^+).

DH10B – F⁻ mcrA Δ (mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araΔ139 D (ara, leu) 7697 galK1⁻ rpsL nupG.

Levedura

H. polymorpha WT-ATCC34438 (WT)

Plasmídeos

YEp352

pUC18

pGEM-T (PROMEGA)

pSORT (INVITROGEN)

4.3 Meios de cultura

LB: 1%Triptona / 1%Nacl / 0,5%NaCl

YEPD: 2% glicose/2%peptona/1%extrato de levedura

YNB: 0,67%YNE; 0,1% extrato de levedura; 5% glicose ou 1% glicerol ou 1% metanol..

→ Os meios sólidos foram preparados com a adição de 2% de ágar.

→ Meio seletivo de LB: 150μg/ml de ampicilina e 50μl de uma solução de X-gal 20mg/ml.

4.4 Extração de DNA genômico de levedura

Após crescimento em 50 ml de YEPD 16hs/37°C/250 rpm, as células foram coletadas por centrifugação e rompidas com 2,0 ml de "breaking buffer" (2% Triton X-100/1% SDS/100mM de NaCl/20mM Tris-HCl) seguido pela adição de 2,0 ml de "glass beads" e submetidas ao vortex por 2 min. A essa mistura foram adicionados 2,0 ml de fenol/clorofórmio/álcool isoamílico (25:24:1) e submetida a centrifugação 4°C/30min/14.000 rpm. A fase aquosa foi recuperada e o DNA precipitado com 2V de Etanol 100% e centrifugação 4°C/30min/14.000 rpm, seguido por lavagem em 10ml de Etanol 70% e ressuspensão em 2ml de TE (10mM Tris-HCl/1mM EDTA).

4.5 Extração de RNA de levedura

O RNA foi extraído por meio de Trizol (Invitrogen) ou pelo método do fenol ácido [125].

4.6 Preparação de bactérias competentes (CaCl2) e eletrocompetentes

Os procedimentos para a preparação de bactérias competentes e eletrocompetentes foram realizados de acordo com os protocolos presentes em Ausubel, 1998 [125].

4.7 Transformação de bactérias

Eletroporação: Foi adicionado 2µl da solução contendo o DNA plasmidial a 50µl de células *E.coli* eletrocompetentes, esta mistura foi acondicionada em cubeta de eletroporação e aplicado pulso de 2,5KV, 200ohms, foi adicionado 1ml de SOC seguido por incubação de 37°C por 1h e plaqueamento em LB/Amp/X-gal.

Choque térmico: Foram adicionados 5µl da solução contendo o DNA plasmidial a 100µl de células *E.coli* CsCl - competentes, estas foram submetidas a choque térmico (42°C 2min), foi adicionado 1ml de SOC seguido por incubação de 37°C por 1h e plaqueamento em LB/Amp/X-gal.

4.8 Preparação de leveduras eletrocompetentes

Este protocolo foi adaptado de Cregg e Russell [126]. Foi feito um pré inóculo células de *H. polymorpha* ATCC34438 em 10ml de YPD e crescidas 16hs/37°C/300 rpm. Seguido este período as células foram inoculadas em 500 ml 37°C/300 rpm até atingir $\lambda 600 = 0.6$, as células foram decantadas por centrifugação (5000rpm/5min/4°C), o sobrenadante foi descartado e as

células ressuspendidas com YEPD/HEPES (100ml de YEPD / 20 ml HEPES 1M pH 8.0) e 1,25ml de DTT 1M, e incubadas à 37°C/15 min. Decorrido este tempo foi adicionado 250 ml de sorbitol 1M gelado, e centrifugadas por 5000rpm/5min/4°C. O sobrenadante foi descartado e as células ressuspendidas em 10 ml de sorbitol 0,2M gelado e centrifugadas (5000rpm/5min/4°C). O sobrenadante foi descartado e as células foram ressuspendidas em 500µl Sorbitol 1M gelado, aliquotadas em tubos eppendorf de 1,5 ml (40µl) e mantidas em gelo por 48hs.

4.9 Transformação de leveduras

Foi adicionado ao tubo contendo as leveduras 10µg do cassete de disrupção, a mistura foi transferida para cubetas apropriadas e aplicado um pulso de 2,5KV/300ohms, transferidas para tubo eppendorf contendo 1ml de YPD/200mM sorbitol e incubadas por 2hs/37°C). Após este período, as leveduras foram plaqueadas em YPD / 200mM sorbitol/200mg de G418 e incubadas por 48-72hs/37°C.

4.10 Extração de plasmídeos em mini/micro escala

Foi utilizado o método da lise alcalina [125].

4.11 Isolamento de DNA de Gel de Agarose e ligação dos fragmentos de DNA

O DNA proveniente da PCR ou digestão enzimática foi extraído de gel de agarose por meio do kit CONCERT (Invitrogen) seguindo as orientações do fabricante. Os fragmentos isolados foram ligados, em plasmídeos apropriados, por 16 horas a 15°C.

4.12 Biblioteca genômica de *H. polymorpha* (insertos de 5-7Kb)

Biblioteca genômica de *Hansenula polymorpha* de 30.000 clones com insertos de 5-7 kb clonados em YEp352, gentilmente cedida pela Doutora Dirce Carraro (Esalq-USP).

4.13 Biblioteca de cDNA de *H. polymorpha*

Para a preparação de RNA inicialmente utilizou o reagente Trizol (Invitrogen), entretanto as quantidades de RNA obtidas foram diminutas. Optou-se pela extração por Fenol Ácido. Para a purificação de RNA mensageiro foi utilizado o mRNA purification Kit (Amershan-pharmacia). Para a preparação da biblioteca usamos Superscript plasmid system for cDNA library (Invitrogen).

4.14 Oligonucleotídeos

Tabela 2: Oligonucleotídeos utilizados nesse trabalho.

Oligonucleotídeos	Seqüência
Degenerados	
DSnf1.1	5'GGRRTTTCAATTRYRTCNGTRTCRAAGC3'
DSnf1.2	5'GCYTGNARRTGYAARTTYGCNTGRTGNGG3'
DSwi2.1	5' ATGATHRTHGAYGAYGG 3'
DSwi2.2	5' TGMRMNMGNDSNTTYG 3'
Dswi3.1	5' TTGGGMYTNWWHAAYTAYCA 3'
Dswi3.2	5' GTNSWNARNACNGGTTYTC 3'
DTup1.1	5' GGGGTACCYTTCCADATNCKNGCYTTRCARTCNCC 3'
DTup1.2	5' GGCTGCAGGGCNAYGARCARGAYATHTAYTC 3'
Específicos	
Dasf	5' ACTTCAACAGAGACAGGTTTG 3'
Dasr	5' AAATTGACAGACACAGACAGGTC 3'
MoxProF	5'AACTGAGGGGGTTCTGC 3'
MoxProR	5' GCTGCACGGTAGCATAAG 3'
SSnf1.1	5' TCACCATCAAGGAGATCATGGAGG 3'
SSnf1.2	5' GGTTGTATTTGGTTCCATTATTCG 3'
Sswi2.1	5' AGTTACAAAGGATCTCCGCAAATG 3'
Sswi2.2	5' GCGTCTTCTTCTTTCAGATAGCTC 3'
SSwi3.1	5' GCCATAAGCGGACACGTTGG 3'
Sswi3.2	5' GCGAGCGCCTCATTGTCTTG 3'
STup1.1	5' CTTCGACTCAGCCAATGAAAGCG 3'
STup1.2	5' TCTGATAACAGAGTTTCTGTGTCC 3'
Disrupção	
CDSnf1.1	5' TCACCATCAAGGAGATCATGGAGGACGAGT AGCTTGCCTCGTCCCCGCCGGGTC 3'
CDSnf1.2	5' GGTTGTATTTGGTTCCATTATTCGTCGACAG CAGTATAGCGACCAGC3'
CDSwi2.1	5' TGAAAAAGATCAGTTACAAAGGATCTCCGC AAATGCCTCGTCCCCGCCGGGTC 3'
CDSwi2.2	5' ACAAGAAGCGTCTTCTTCTTTCAGATAGCTC TCGACAGCAGTATAGCGACCAGC 3'
CDSwi3.1	5' GCGAGCGCCTCATTGTCTTGCTCCAACATCT GCTTTTGCCTCGTCCCCGCCGGGTC 3'
CDSwi3.2	5' CCATAAGCGGACACGTTGGAACCCGGACCA AGGGAATCGACAGCAGTATAGCGACCAGC 3'
CDTup1.1	5' TTTCTGGCTCAGAAGCCAGCCCAGTAAGATT TTGGAGCTCGCTCGTCCCCGCCGGGTC 3'
CDTup1.2	5' CCCGACACCTCTCTTCTGGGTCATAACCGGT TCGACAGCAGTATAGCGACCAGC 3'
Seqüênciamento	
M13-21 Forward	5' GTAAAACGACGGCCAGT 3'
M13-28 Reverse	5' GGAAACAGCTATGACCATG 3'
T7 Promoter	5' ATTTAGGTGACACTATAG 3'
SP6 Promoter	5' TAATACGACTCACTATAGGG 3'

4.15 Condições de PCR utilizadas para oligos degenerados e específicos

As PCRs consistiram de 2µl de tampão sem magnésio, 1,5 µl de $MgCl_2$ 50mM, 45pmol de cada primer, quando utilizados oligos degenerados, ou 15pmol de cada primer, quando específico, 4µl de DNTPs 1,25mM e 2u de *Taq* DNA polimerase.

Utilizando os oligonucleotídeos descritos anteriormente para cada gene, foram realizadas PCRs que consistiram em 30 ciclos 94°C 1min, variando a temperatura de anelamento de 40°C/60°C - 30seg. para os oligos degenerados para que se pudesse caracterizar a temperatura ótima, onde apresentasse o menor número de bandas espúrias possíveis para cada gene. Para os oligonucleotídeos específicos a temperatura de anelamento foi baseada na TM média dos oligos utilizados para a PCR, e extensão 72°C 1min.

4.16 Condições de PCR utilizadas para construção dos modulos de disrupção

As PCRs continham: 5µl de tampão sem magnésio, 2,5 µl de $MgCl_2$ 50mM, 50 pmol de cada primer, 8µl de DNTPs 1,25mM e 5u de *Taq* DNA polimerase. O programa utilizado foi: 1 ciclo inicial de denaturação 96 °C seguido de 40 ciclos de 95°C 1 min, 60°C 30 seg 72°C 1 min.

4.17 Reações de seqüênciamento

As reações de seqüênciamento foram realizadas em volume final de 10µl, sendo: 3µl de tampão (200mM de Tris/HCl), 0,8µl de Bigdye terminator Mix (Applied Biosystems) 15 pmol de primer quando degenerado ou 4 pmol de primer específico. O molde quando PCR foi de 10ng de DNA para cada 100 bases de produto amplificado, quando em pUC18 ou pGEM-T 400ng de plasmídeo e em Yep352 e pSPORT 5 µg. Completando-se para volume final com H_2O milliQ. Os ciclos de seqüênciamento e a precipitação das amostras seguiram as orientações do fabricante.

4.18 Análise dos dados de seqüênciamento

Os resultados do seqüênciamento foram analisados por meio dos programas BlastN e BlastX (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST ou http://www.lge.ibi.unicamp.br) para análise de identidade com fatores já conhecidos. O mapeamento de restrição e ORFs foi efetuado por meio do programa Generunner. O processamento dos "*reads*" (análise de qualidade da seqüência, trimagem e blastagem) relativos ao projeto de seqüênciamento e criação de banco de dados genéticos de *Hansenula polymorpha* foram realizados de forma automática no Centro de

Bioinformática de nosso laboratório (http://www.lge.ibi.unicamp.br). As seqüências se encontram depositas no "*website*" http://www.lge.ibi.unicamp.br/hansenula.

4.19 "Southern blotting", "northern blotting" hibridação e marcação radioativa de sonda de DNA

Os procedimentos de transferência de fragmentos de DNA para suporte sólido, hibridação e marcação radioativa de DNA (sonda) pelo método RPS (Random Primed Synthesis) foram realizados de acordo com os protocolos presentes em Ausubel *et. al.*, 1998.

4.20 Ensaios de viabilidade

Os ensaios de viabilidade foram conduzidos de acordo com os procedimentos descritos na literatura [127]. A porcentagem da fonte de carbono adicionada às placas de YNB+ágar foi de 3%, exceto para metanol (1%).

4.21 Curvas de crescimento

Foi realizado um pré-inóculo das linhagens estudas e crescidas por 16hs. Após este período as culturas foram diluídas para OD_{600}= 0,01 em YNB sem carbono e inoculadas em meios YNB contendo a fonte de carbono apropriada (metanol 1%, glicose 3%, glicerol 1% ou galactose 3%) e o crescimento foi acompanhado por 24 hs por espectrofotometria em λ=600. Durante o crescimento as culturas foram analisadas ao microscópio e ao final do experimento foram inoculas (riscadas) em placa YEPD/ágar para se evitar resultados incoerentes devido a contaminação.

5. RESULTADOS E DISCUSSÃO

Dos genes propostos inicialmente no trabalho, não foi possível a identificação satisfatória dos fatores *Adr1*, *Ssn6* e *Mig1* em *H. polymorpha*. Estes resultados são provavelmente devidos a natureza dos fatores que possuem pequenas regiões conservadas somente nos motivos essenciais para o seu papel biológico, sendo que o restante do gene não apresenta grande conservação entre as espécies, o que dificulta a sua identificação pelas técnicas utilizadas. Entretanto foi possível a caracterização satisfatória dos fatores *Snf1*, *Swi2* e *Tup1*. Os resultados são apresentados a seguir.

5.1 Identificação e disrupção dos genes *SNF1* e *TUP1* de *Hansenula polymorpha*

TUP1

Utilizando a técnica de oligonucleotídeos degenerados foi possível clonar extensas regiões dos genes propostos neste trabalho. Para o gene *Tup1* foi possível identificar uma porção gênica de 700pb, que foi posteriormente estendia para ≈1200pb com alta similaridade à *Tup1* de *C. albicans* (2e-89).

As proteínas Tup1p já identificadas em leveduras e fungos apresentam como características marcantes um domínio de poli-glicina, localizado nas proximidades da região amino terminal e as repetições WD40 (W=triptofano e D= ácido aspártico) que ocorrem aproximadamente a cada 40 bases e se estendem por toda a proteína [91,96]. As repetições WD foram observadas primeiramente na β transducina, uma das três subunidades das proteínas G que agem como intermediários na transdução de sinal gerado pelos receptores transmembrana (por conta deste fato recebem também o nome de repetições do tipo β transducina). Em leveduras, o número de domínios WD nas proteínas Tup1 pode variar de 5-7 repetições WD dependendo da espécie [99]. A tradução da seqüência de DNA obtida de Tup1 *H.polymorpha* revela cinco repetições WD adjacentes à porção carboxiterminal que se alinha fortemente com a região carboxiterminal de *C. albicans* (Manuscrito 1-em anexo). O módulo utilizado para disrupção de *HpTUP1* elimina todas as repetições WD presentes no gene identificado.

SNF1

A porção identificada em nosso trabalho corresponde, como em *Tup1*, a região carboxi-terminal e se estende para além da região central da enzima. As proteínas da classe de Snf1 possuem como característica uma região altamente conservada com a função de fosforilar resíduos de serina e treonina das proteínas com as quais interage. Estruturalmente possuem uma região repetitiva de polihistidina localizada próximo a região aminoterminal (16 A.A), um domínio quinase (≈ 250 A.A.), um sítio de ativação com um resíduo de ácido aspártico altamente conservado, importante para a atividade da enzima. Também é encontrada uma serina na porção

mediana do sítio catalítico que pode servir de alvo para autofosforilação e portanto regular a atividade de Snf1 [128]. A porção gênica identificada de *Snf1* em *H. polymorpha* deve representar cerca de ≈ 70% da proteína (≈1300pb), uma vez que possui uma notável similaridade com *Snf1* de *C. albicans*, espécie em que o gene *Snf1* possui 1820 pb (Manuscrito 1). O módulo para a disrupção de *HpSNF1* deletou ≈ 1200pb, incluindo cerca 200 bases do domínio com função quinase.

5.2 Efeito das disrupções na utilização de fontes alternativas de carbono

Apesar do interesse central deste trabalho visar a identificação de genes envolvidos no metabolismo de metanol, foram testadas várias fontes de carbono com o intuito de analisar mais detalhadamente os efeitos das disrupções dos genes *HpTup1* e *HpSnf1*. Para tanto foi utilizado como controle células crescidas em glicose. Para a análise de efeitos fenotípicos na utilização de metanol foram testadas as habilidades das células *tup1* e *snf1* crescer em meios indutores (glicerol e metanol). Adicionalmente, foram testadas outras fontes de carbono (sacarose, rafinose e galactose) com intuito de comparação com fenótipos descritos em *S. cerevisiae* [83]. As disrupções de *HpSnf1* ou *HpTup1* não levaram à redução no crescimento em nenhuma das fontes de carbono testadas, sugerindo não haver envolvimento do produto destes genes nos eventos de regulação por glicose em *H. polymorpha*. Para confirmar estes resultados as leveduras foram crescidas em YPD líquido suplementado com glicose 3% ou metanol 1% e as taxas de crescimento foram monitoradas por 22 horas. Também foram realizados experimentos de "*northern blot*" para análise da transcrição dos genes *Mox, Cat* e *Das* os quais estão diretamente envolvidos no metabolismo de metanol. Os resultados são descritos a seguir no manuscrito 1.

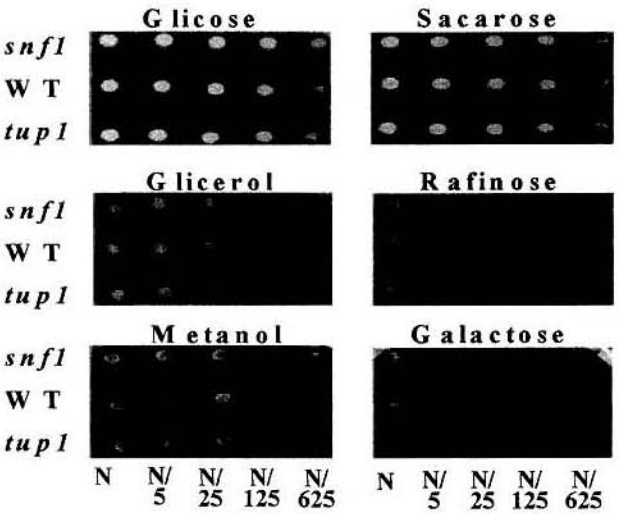


Figura 7: Efeito das disrupções dos genes *HpSnf1* e *HpTup1* na utilização de diversas fontes de carbono. As placas foram fotografadas após 36 horas de incubação a 37 °C. Com exceção das placas contendo glicerol (2%) e metanol (1%) as concentrações finais das demais fontes de carbono foram de 3%.

The general regulators *TUP1* and *SNF1* are not essential for regulation by glucose of the major genes involved in methanol metabolism from *Hansenula polymorpha*

5.3 MANUSCRITO 1: Submetido ao periódico: *Genetics and molecular biology*

Marcos A. Oliveira[1], Victor Genu[1], Anita P. T. Salmazo[1], Dirce M. Carraro[2] and Gonçalo A. G. Pereira[1,#]

[1]Laboratório de Genômica e Expressão - Departamento de Genética e Evolução - Instituto de Biologia - Universidade Estadual de Campinas - Campinas, SP, Brazil.

[2]Ludwig Institute for Cancer Research- São Paulo, SP, Brazil

Running title: TUP1 and *SNF1* analysis in *H. polymorpha*

Keywords: *TUP1*, *SNF1*, *Hansenula polymorpha*, glucose repression, *MOX*.

[#]Send correspondence to Gonçalo A. G. Pereira, Laboratório de Genômica e Expressão (LGE), Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, CEP: 13083-970. Campinas - SP - Brazil. E-mail: goncalo@unicamp.br

Phone: 00 55 19 3788 6237

Fax: 00 55 19 3788 6235

ABSTRACT

Hansenula polymorpha is a methylotrophic yeast employed in biotechnology as a "protein factory". Most promoters used for heterologous expression belong to genes involved in the peroxisomal methanol metabolism, which are under strict glucose repression. Understanding this regulation may be useful to maximize fermentation processes. Interestingly, the *H. polymorpha* methanol oxidase promoter (*MOX*) is under glucose regulation in *Saccharomyces cerevisiae*, a non-methylotrophic yeast where the phenomenon is well studied. In this species, the factors Tup1p and Snf1p play crucial roles in, respectively, repressing and derepressing genes in response to glucose availability. In this work, to test whether the roles of these genes were conserved in *H. polymorpha*, *HpTUP1* and *HpSNF1* were partially cloned and used to disrupt their genomic copies. Accordingly, the deletion of *HpSNF1* led to significant reduction in the derepression of major methanol metabolism genes. In contrast, deletion of *HpTUP1* did not affect glucose repression on these genes. In conclusion, these data indicate that, though conserved, *HpTUP1* does not seem to be involved in general glucose repression in *H. polymorpha* and the effect of *HpSNF1* in activation may be through a *HpTUP1*-independent circuit.

INTRODUCTION

Hansenula polymorpha (syn. *Pichia angusta*) belongs to a limited group of yeast able to metabolize methanol. The enzymes necessary for the first steps of methanol utilization, namely methanol oxidase (Moxp), catalase (Catp) and dihydroxiacetone synthase (Dasp), are located in the peroxisomes. Their expressions are transcriptionally controlled and, at least for *MOX* and *DAS*, seem to be co-ordinately regulated. Moxp can represent up to 30% of the soluble proteins when the cells grown in methanol as sole carbon source and peroxisomes can take up over 80% of the cell volume under this condition (Veenhuis *et al.*, 1983; van der Klei *et al.*, 1991). In view of that, the *MOX* promoter has been widely employed for heterologous protein expression and *H. polymorpha* has been considered a "protein factory" (Gellissen, 2000; van Dijk *et al.*, 2000). However, *MOX* is also under strict glucose repression. Thus, although glucose is a very suitable carbon source for general industrial fermentation, this sugar can not be usually employed for protein expression using *H. polymorpha*. It would therefore be a significant technological improvement if the *MOX* promoter could work in glucose-containing media.

The *MOX* gene promoter has been studied in detail. Promoter mapping revealed four *cis*-acting regulatory elements (Godecke *et al.*, 1994), but no *H. polymorpha* factor responsible for *MOX* regulation has been identified to date. Remarkably, *MOX* promoter can drive glucose repressible expression under a reporter gene in *Saccharomyces cerevisiae*, a non-methylotrophic yeast (Pereira and Hollenberg, 1996). In this species, *MOX* promoter activation is dependent on Adr1p, a transcription factor that activates a number of genes encoding peroxisomal enzymes (Simon *et al.*, 1991, Simon *et al.*, 1995;

Gurvitz *et al.*, 2001), the class to which Moxp belongs. This finding indicates that mechanisms governing regulation by glucose may be conserved in both species.

In *S. cerevisiae*, a number of genes from different carbon source utilization pathways, like sucrose or galactose, are repressed by the complex Mig1p-Ssn6p-Tup1p when cells grow in presence of glucose (Bu and Schmidt, 1998; Rohde *et al.*, 2000). When glucose is exhausted, the very genes need the protein kinase Snf1p to be derepressed (Ostling *et al.*, 1996; Gancedo, 1998). Both the Mig1p-Ssn6p-Tup1p and the Snf1 complexes are highly conserved among several species and they seem to be involved not only in glucose regulation (Yamashiro *et al.*, 1996; Cassart *et al.*, 1997; Zaragoza *et al.*, 2000; Lo; *et al.*, 2001). For example, the deletion of a *TUP1* homologue in *Candida albicans* causes failure in hypha development (Braun and Johnson, 1997).

In this study, *H. polymorpha* genes homologous to *TUP1* and *SNF1* were identified and analyzed for their possible role in the regulation of genes involved in the methanol metabolism. The results show that, although highly conserved, Hp*TUP1* and Hp*SNF1* do not seem to play an essential role in regulation by glucose in *H. polymorpha*.

MATERIALS AND METHODS

Strains, plasmids and oligonucleotides

The *H. polymorpha* strain used in this work was the CBS 4732 (wild type - WT). *Escherichia coli* strain DH10B (Invitrogen) was used for all DNA manipulations. The plasmid pF6A-kanmx4 (Wach *et al.*, 1994) was used for construction of the disruption modules. pGEM-T (Promega) and pUC18 were used for cloning of PCR products. The oligonucleotides used in this study are listed in Table 1.

Media

YPD (1% yeast extract, 2% peptone, 2% glucose) was used as rich medium. SD medium (0,17% YNB; 0,5% ammonium sulfate) was used as synthetic medium and was supplemented with 5% glucose or 1% methanol.

Culture and induction experiments

H. polymorpha cells were grown in 250 ml flasks with 100ml of YPD for 16hs at 37°C and 280 rpm using orbital shaker. After this period, cells were harvested by centrifugation for 10 min at 5.000 rpm and inoculated in appropriated medium for 6hs at 37°C and 280 rpm.

Cloning of *H. polymorpha* genes

Strategy of cloning was based on degenerated oligonucleotides. We designed degenerated primers after conserved regions of the *TUP1* and *SNF1* gene sequences from *S. cerevisiae*, *Schizosaccharomyces pombe* and *C. albicans* (Tab. 1). PCRs were carried out using *H. polymorpha* genomic DNA as template. Sequences containing 600bp and 800bp were amplified using DTup1 and DSnf1 primers, respectively. These DNA fragments were cloned into pGEM-T (Promega) and sequenced. The DNA sequences obtained were analyzed by Blast X (www.ncbi.nlm.nih.gov/blast) and presented high similarities with

C. albicans TUP1 (4e-84) *and SNF1* (5e-54). The sequence of each fragment's extremity was used to design the specific primers shown in Tab. 1 (Stup1F, Stup1R; Ssnf1F, Ssnf1R). Each primer was used in combination with T7 or SP6 custom primers to amplify cDNA fragments from a *H. polymorpha* cDNA library, constructed using the vector pSPORT (Invitrogen). By this procedure the putative Hp*TUP1* and Hp*SNF1* gene fragments were expanded. The predicted amino acid sequences of the *H. polymorpha TUP1* and *SNF1* genes were compared with known homologues of other yeast species (see Fig. 1) using the Multialin tool (www.prodes.toulouse.inra.fr/multalin/multalin.html).

Disruption and screening of *Hptup1 and Hpsnf1* strains

The *H. polymorpha TUP1* and *SNF1* DNA sequences described above were also used to design chimerical disruption oligonucleotides. These oligonucleotides (see Table 1 and Fig. 2) present ~30bp of the genes to be disrupted in their 5' extremities; the 3' extremities are composed of ~25bp complementary to the Multiple Cloning Site of the pF6A plasmid where the kanmx4 disruption module was integrated (Wach *et al.*, 1994). Specific *TUP1* and *SNF1* kanmx4 modules were constructed with these primers (CDTup1.1+CDTup1.2 and CDSnf1.1+CDSnf1.2, respectively) by standard PCRs using the pF6A-kanmx4 plasmid as template. These linear disruption modules bear, at both 5' and 3' ends, sequences homologous to *TUP1* and *SNF1* genes and were used to transform *H. polymorpha* WT electro-competent cells. The kanmx4 module confers resistance to the drug G418, which was used for preliminary screening. Positive transformants growing in the presence of G418 (Amersham Pharmacia) were confirmed by colony PCR and also by PCRs with genomic DNAs, using the specific primers STup1.1+STup1.2; SSnf1.1+SSnf1.2 (Tab. 1).

PCR Conditions

All PCRs were made in a final volume of 50 µl using kits (Promega) and following instructions provided by the manufacturer. The reactions were carried out on a 9700 thermocycler (Applied Biosystems) using the following programs. For fragment amplification: 96°C for 4min, followed by 30 cycles at 94°C for 30 sec, 40°C (degenerated primers) or 50°C (specific primers) for 30 sec and 72°C for 1 min. For disruption module construction: 96°C for 4min, followed by 30 cycles: 94°C for 30 sec, 70°C for 30 sec and 72°C for 1.5 min.

Growth characterization

The strains were grown till log-phase in YPD, harvested and resuspended in the appropriate media in an OD_{600}= 0.01. Aliquots were collected each hour for 24 hours and the absorbance at λ= 600nm was measured. The results presented are the average of three independent experiments.

Solid growth assay

The assays were performed as described elsewhere (Verdoucq *et al.*, 1999; Demasi *et al.*, 2001).

Northern blot analysis

Cells were grown in SD plus 0.1% yeast extract under the conditions described in *Results*. Aliquots of cells were collected and total RNAs were extracted by the hot acid phenol method (Ausubel *et al.*, 1998). Samples were resolved by formaldehyde gel electrophoresis, blotted onto a positively charged nylon membrane (Hybond N^+-Amersham Pharmacia) and fixed to it following standard protocols. The probes used were the following DNA fragments (numbers refer to position relative to ORF start codon): *MOX*- a 2,3kb *Eco*RV fragment (+21 to +2323), *CAT*- a 1,7kb *Eco*RI/*Eco*RV fragment

(+172 to +1695) and *DAS*, a 1.0kb (+197/+1196) PCR fragment obtained with the primers Dasf and Dasr. All probes were labeled by random primer extension with α-dATP^{32} (Amersham Pharmacia). Hybridization was carried out at 42°C following standard protocols (Ausubel *et al.*, 1998).

RESULTS

Identification of *H. polymorpha TUP1* and *SNF1* homologues

H. polymorpha DNA fragments were obtained by PCR using degenerated primers for important regions of the *TUP1* and *SNF1* genes (Table 1 and Fig. 1). The putative Hp*TUP1* encompassed six WD repeats (Fig.1), a typical motif of this protein family, which is involved in protein-protein interaction and signalling (Smith *et al.*, 1999). The putative Hp*SNF1* gene fragment bears a region with high similarity to related proteins, including part of the serine/threonine domain presented in all *SNF1*-like kinases (Hanks and Hunter, 1995) (Fig.1). Both genes presented the highest similarities with homologues from the *Candida genera*, which is phlylogenetically closer to *Hansenula*. From the sequence comparison analysis, it was concluded that the DNA fragments obtained encompass the most important functional regions of *H. polymorpha TUP1* and *SNF1* genes.

Construction of *TUP1* and *SNF1 H. polymorpha* mutants by PCR

The Hp*TUP1* and Hp*SNF1* fragments were used to construct disruption modules as described in *Materials and Methods*. *H. polymorpha* wild type cells were transformed and positive colonies growing on G418 were selected. About 150 transformants for each gene were screened for the presence of the disruption module by PCR amplification using specific primers for the Hp*TUP1* and Hp*SNF1* gene fragments. Fig. 2 shows the PCR products of correct integrations, observed for 7% of the transformants, in comparison to amplification of the wild type fragments. Growth abilities of *tup1* and *snf1* cells on plates containing G418 were re-checked in comparison to wild type cells (not shown).

Expression analysis of genes involved in methanol utilization in *tup1* and *snf1* mutants

The expression of the genes *MOX*, *CAT* and *DAS*, which encode peroxisomal enzymes involved in methanol metabolism, were analysed by Northern Blot in cells from the strains *tup1* and *snf1* in comparison to wild type cells. In *H. polymorpha tup1* cells, the transcripts level of the three genes were very similar to those presented by wild type cells on media glucose or methanol (Fig. 3). This indicates that *MOX*, *CAT* and *DAS* regulation by glucose are *TUP1*-independent in *H. polymorpha*. Interestingly, *snf1* cells presented a significant reduction on *MOX* and *DAS* transcription when grown on media containing methanol (Fig. 3), suggesting that *SNF1* might be needed for activated levels of these genes. Conversely, *CAT* transcription was unaffected under this condition (Fig. 3).

Disruption effects of the Hp*TUP1* and Hp*SNF1* on carbon utilization

In *S. cerevisiae*, disruption of *TUP1* or *SNF1* led to several pleiotropic effects in addition to glucose regulation. For example, *TUP1* disruption led to cell flocculation (Lipke and Hull-Pillsbury, 1984) and *SNF1* disruption to a general slow grow (Schuller and Entian, 1987). To examine whether disruption of Hp*TUP1* or Hp*SNF1* would lead to general growth defects or specifically affect the methanol metabolism, the growth rates of mutant cells were compared to the wild type counterparts in solid and liquid media (Fig. 4). In contrast to the phenotypes observed in *S. cerevisiae*, no significant difference in growth was observed in either carbon sources.

DISCUSSION

In *H. polymorpha* the *MOX* promoter has been investigated due to its importance for industrial expression of heterologous protein. The understanding of the factors governing its regulation may be useful to expand its expression profile, allowing utilization of *MOX* promoter in conditions under which it is ineffective (e.g. in glucose-containing medium). In addition, *MOX* regulation is tightly connected with the peroxisomal metabolism, so that the comprehension of the factors affecting its expression may help to understand the factors involved in peroxisomal regulation.

Remarkably, the *MOX* promoter can drive glucose repressible expression of a reporter gene in *S. cerevisiae* (Pereira and Hollenberg, 1996), a non-methylotrophic yeast that has been widely employed as model organism. This finding indicates that *S. cerevisiae* could conserve the molecular mechanisms that regulate *MOX* in *H. polymorpha.* Consistent with this hypothesis, Adr1p, a transcription factor that activates *ADH2* (alcohol dehydrogenase 2) and a number of genes encoding peroxisomal proteins (Simon *et al.*, 1995; Gurvitz *et al.*, 2001; Young *et al.*, 2002) regulates also the *MOX* promoter in *S. cerevisiae.* Thus, *H. polymorpha* genes homologous to *S. cerevisiae* factors that regulate genes involved in carbon source utilization may be responsible for the transcriptional regulation of the methanol metabolism genes. For that reason, *H. polymorpha* homologues of the genes Sc*TUP1* and Sc*SNF1*, which play key roles in the regulation of several genes subjected to glucose repression in *S. cerevisiae*, were investigated in this work.

TUP1 mutations eliminate or reduce glucose repression of many repressible *S. cerevisiae* genes and induce other phenotypes, including flocculation, failure to sporulate,

and sterility of MAT alpha cells (Williams and Trumbly, 1990). However, in *H. polymorpha* no evident phenotype was observed for the *tup1* deletion. Mutant cells were indistinguishable from wild type under microscopic analysis and no variation in growth behavior or in expression of major genes involved in methanol metabolism could be observed. It is important to note that although the deletion does not encompass the complete ORF, the removed region bears the WD40 repeats, which are essential and sufficient for *TUP1* repressing function in *S. cerevisiae* (Williams and Trumbly, 1990; Sprague *et al.*, 2000). This result is surprising since clear effects of *TUP1* mutations have been demonstrated in other yeast species closer to *H. polymorpha*. For example, *Candida albicans TUP1*, which works in a complex similar to that identified is *S. cerevisiae*, is involved in general regulation of metabolism, cellular morphogenesis and stress responses (Murad *et al.*, 2001). Although *TUP1* is not essential in this species, its deletions led to female sterility, reduced growth rates and failure to turn on filamentous growth (Yamashiro *et al.*, 1996; Braun and Johnson, 1997).

Concerning *SNF1*, the product of this gene in *S. cerevisiae* is required to derepress many glucose-repressible genes, such as, *SUC2* (Carlson *et al.*, 1981), *GAL1* (Flick and Johnston, 1990) and *ADH2* (Denis and Audino, 1991), in response to conditions of low external glucose, interfering in the utilization of sucrose, galactose and maltose (Celenza and Carlson, 1984). The *snf1* mutation led to pleiotropic effects, such as a petite like phenotype and slow growth, having the mutant cells a "sick" aspect (Carlson *et al.*, 1981). One of the main roles of Snf1p is to relieve repression by the Mig1p-Ssn6p-Tup1p complex (Ostling *et al.*, 1996), but it is also involved in the operation of transcription factors such as Adr1p (Young *et al.*, 2002), and possibly other factors that are still unidentified [65]. Indeed, the protein kinase Snf1p is shared by various circuits and is

therefore a central element in several regulatory processes [65]. Disruption of the *Candida glabrata SNF1* homolog resulted in the loss of the ability to utilize trehalose, indicating that even in an organism with such a limited carbon utilization spectrum, the regulatory mechanism governing catabolite repression is preserved (Petter and Kwon-Chung, 1996).

Consistent with the possibility of playing a role in derepression in *H. polymorpha*, the deletion of *SNF1* led to significant reduction of the *MOX* and *DAS* expression in cells growing in methanol. This effect seems to be specific for some genes, since *CAT* expression remained unaffected by the deletion. These results reinforce the view that *MOX* and *DAS* are tightly co-regulated and that possibly Snf1p plays a role in their regulatory circuits. However, this role may not be due the inactivation of the Mig1p-Ssn6p-Tup1p complex, as has been reported for several *S. cerevisiae* genes [65]. As described above, Hp*TUP1* does not seem to repress either *MOX* or *DAS* expression. More probably, Hp*SNF1* could exert its effect by interacting with a putative HpAdr1p that is believed to regulate *MOX* (Pereira and Hollenberg, 1996). In *S. cerevisiae*, several genes, including some that encode peroxisomal proteins, are regulated by both factors (Denis and Audino, 1991; Simon *et al.*, 1992; Navarro and Igual, 1994). Moreover, Snf1p has been recently demonstrated to promote binding of Adr1p to chromatin (Young *et al.*, 2002). These facts indicate that both factors are connected to the regulation of some genes and that this connection may be conserved in other organisms.

Nevertheless, despite the reduction in *MOX* and *DAS* expression, deletion of *SNF1* did not produce any obvious external phenotype. The general aspect of the *snf1* cells was normal under microscopy, indistinguishable from wild type. Also the growth rate was

normal, even in methanol, indicating that the reduction in *MOX* expression was not sufficient to create a bottleneck in peroxisomal methanol metabolism.

The slight effect of *SNF1* deletion in *H. polymorpha* is quite surprisingly. In *C. tropicalis*, a species closer to *H. polymorpha*, the *SNF1* homologue (Ct*SNF1*) seems to be an essential gene and to affect expression of several carbon utilization genes (Kanai *et al.*, 1999). Similar situation is observed for *C. albicans* (Petter *et al.*, 1997).

In summary, our work indicates that, in contrast to the observations made for other yeast species, the genes *TUP1* and *SNF1* do not play essential roles in carbon source regulation in *H. polymorpha*. Nevertheless, Snf1p may participate on the activation of genes repressed by glucose, but at least in case of the genes *MOX* and *DAS*, its mechanism of action is diverse from counteract the repression of the Tup1p containing complex.

ACKNOWLEDGEMENTS

We thank Thiago C. Lopes for supplying the pF6A-kanmx4 plasmid; Anderson F. Cunha, Ana P. D. Demasi, Carla C. J. Maria and Sérgio F. Reis for critical reading of the manuscript.This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Brazil) to M.A.O. and G.A.G.P.

REFERENCES

Ausubel M F, Brent R, Kingston RE, Moore DD, Seidman GJ, Smith JA and Struhl K (1998) Current Protocols in Molecular Biology. New York, John Wiley & Sons, Inc.

Braun, BR and Johnson AD (1997) Control of filament formation in *Candida albicans* by the transcriptional repressor *TUP1*. Science 277:105-109.

Bu, Y and Schmidt MC (1998) Identification of cis-acting elements in the *SUC2* promoter of *Saccharomyces cerevisiae* required for activation of transcription. Nucleic Acids Res. 26:1002-1009.

Carlson M, Osmond BC and Botstein D (1981) Mutants of yeast defective in sucrose utilization. Genetics 98:25-40.

Cassart JP, Ostling J, Ronne H and Vandenhaute J (1997) Comparative analysis in three fungi reveals structurally and functionally conserved regions in the Mig1 repressor. Mol. Gen. Genet. 255:9-18.

Celenza JL and Carlson M (1984) Structure and expression of the *SNF1* gene of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 4:54-60.

Demasi AP, Pereira GG, Netto LES (2001) Cytosolic thioredoxin peroxidase I is essential for the antioxidant defense of the yeast with dysfunctional mitochondria. FEBS Lett. 509:430-434.

Denis CL and Audino DC (1991) The CCR1 (SNF1) and SCH9 protein kinases act independently of cAMP-dependent protein kinase and the transcriptional activator ADR1 in controlling yeast *ADH2* expression. Mol. Gen. Genet. 229:395-399.

Flick JS and Johnston M (1990) Two systems of glucose repression of the *GAL1* promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 10:4757-69.

Gancedo JM (1998) Yeast carbon catabolite repression. Microbiol. Mol. Biol. Rev. 62:334-361.

Gellissen G (2000) Heterologous protein production in methylotrophic yeasts. Appl. Microbiol. Biotechnol. 54:741-750.

Godecke S, Eckart M, Janowicz ZA and Hollenberg CP (1994) Identification of sequences responsible for transcriptional regulation of the strongly expressed methanol oxidase-encoding gene in *Hansenula polymorpha*. Gene 139:35-42.

Gurvitz A, Hiltunen JK, Erdmann R, Hamilton B, Hartig A, Ruis H and Rottensteiner H (2001) *Saccharomyces cerevisiae* Adr1p governs fatty acid beta-oxidation and peroxisome proliferation by regulating *POX1* and *PEX11*. J. Biol. Chem. 276:31825-31830.

Hanks SK and Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. Faseb J. 9:576-596.

Kanai T, Ogawa K, Ueda M and Tanaka A (1999) Expression of the *SNF1* gene from *Candida tropicalis* is required for growth on various carbon sources, including glucose. Arch. Microbiol. 172:256-263.

Lipke PN and Hull-Pillsbury C (1984) Flocculation of *Saccharomyces cerevisiae tup1* mutants. J. Bacteriol. 159:797-799.

Lo WS, Duggan L, Tolga NC, Belotserkovskya R, Lane WS, Shiekhattar R and Berger SL (2001) Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. Science 293:1142-1146.

Murad A M, d'Enfert C, Gaillardin C, Tournu H, Tekaia F, Talibi D, Marechal D, Marchais V, Cottin J and Brown AJ (2001) Transcript profiling in *Candida albicans* reveals new cellular functions for the transcriptional repressors *CaTup1*, *CaMig1* and *CaNrg1*. Mol. Microbiol. 42:981-993.

Navarro B and Igual JC (1994) ADR1 and SNF1 mediate different mechanisms in transcriptional regulation of yeast *POT1* gene. Biochem. Biophys. Res. Commun. 202:960-966.

Ostling J, Carlberg M and Ronne H (1996) Functional domains in the Mig1 repressor. Mol. Cell. Biol. 16:753-761.

Pereira GG and Hollenberg CP (1996) Conserved regulation of the *Hansenula polymorpha MOX* promoter in *Saccharomyces cerevisiae* reveals insights in the transcriptional activation by Adr1p. Eur. J. Biochem. 238:181-191.

Petter R, Chang YC and Kwon-Chung KJ (1997) A gene homologous to *Saccharomyces cerevisiae SNF1* appears to be essential for the viability of *Candida albicans*. Infect. Immun. 65:4909-4917.

Petter R and Kwon-Chung KJ (1996) Disruption of the *SNF1* gene abolishes trehalose utilization in the pathogenic yeast *Candida glabrata*. Infect. Immun. 64:5269-5273.

Rohde JR, Trinh J and Sadowski I (2000) Multiple signals regulate *GAL* transcription in yeast. Mol. Cell. Biol. 20:3880-3886.

Schuller HJ and Entian KD (1987) Isolation and expression analysis of two yeast regulatory genes involved in the derepression of glucose-repressible enzymes. Mol. Gen. Genet. 209:366-373.

Simon M, Adam G, Rapatz W, Spevak W and Ruis H (1991) The *Saccharomyces cerevisiae ADR1* gene is a positive regulator of transcription of genes encoding peroxisomal proteins. Mol. Cell. Biol. 11:699-704.

Simon M, Binder M, Adam G, Hartig A and Ruis H (1992) Control of peroxisome proliferation in Saccharomyces cerevisiae by *ADR1, SNF1* (*CAT1*, *CCR1*) and *SNF4* (*CAT3*). Yeast 8:303-309.

Simon MM, Pavlik P, Binder M, Ruis H, Cook WJ, Denis CL and Schanz B (1995) A C-terminal region of the *Saccharomyces cerevisiae* transcription factor ADR1 plays an important role in the regulation of peroxisome proliferation by fatty acids. Mol. Gen. Genet. 249:289-296.

Smith TF, Gaitatzes C, Saxena K and Neer EJ (1999) The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci. 24:181-185.

Sprague ER, Redd MJ, Johnson AD and Wolberger C (2000) Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. Embo J. 19:3016-3027.

van der Klei IJ, Harder W and Veenhuis M (1991) Biosynthesis and assembly of alcohol oxidase, a peroxisomal matrix protein in methylotrophic yeasts: a review. Yeast 7:195-209.

van Dijk R, Faber KN, Kiel JA, Veenhuis M and van der Klei I (2000) The methylotrophic yeast *Hansenula polymorpha*: a versatile cell factory. Enzyme Microb. Technol. 26:793-800.

Veenhuis M, Van Dijken JP and Harder W (1983) The significance of peroxisomes in the metabolism of one-carbon compounds in yeasts. Adv. Microb. Physiol. 24:1-82.

Verdoucq L, Vignols F, Jacquot JP, Chartier Y and Meyer Y (1999) In vivo characterization of a thioredoxin h target protein defines a new peroxiredoxin family. J. Biol. Chem. 274:19714-19722.

Wach A, Brachat A, Pohlmann R and Philippsen P (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast 10:1793-1808.

Williams FE and Trumbly RJ (1990) Characterization of *TUP1*, a mediator of glucose repression in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 10:6500-6511.

Yamashiro CT, Ebbole DJ, Lee BU, Brown RE, Bourland C, Madi L, and Yanofsky C (1996) Characterization of *rco-1* of *Neurospora crassa*, a pleiotropic gene affecting growth and development that encodes a homolog of *Tup1* of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 16:6218-6228.

Young ET, Kacherovsky N and Riper KV (2002) Snf1 protein kinase regulates Adr1 binding to chromatin but not transcription activation. J. Biol. Chem.

Zaragoza O, Rodriguez C and Gancedo C (2000) Isolation of the *MIG1* gene from *Candida albicans* and effects of its disruption on catabolite repression. J. Bacteriol. 182:320-326.

LEGENDS

Figure 1: Identification of Hp*TUP1* and Hp*SNF1* gene fragments. Alignments of the predicted HpTup1p (A) and HpSnf1p (B) proteins with other known homologues. (HP = *Hansenula polymorpha*, CA = *Candida albicans*, CG = *Candida glabrata*, CT = *Candida tropicalis*, EM = *Emericella nidulans*, KL = *Kluyveromyces lactis*, NC = *Neurospora crassa*, Sc = *Saccharomyces cerevisiae*, Sp = *Schizosaccharomyces pombe*, YL= *Yarrowia lipolytica*). The known conserved domains identified in *C. albicans* proteins are schematically shown on top of the alignments (see text for details). Alignments were done with the Multalin tool (www.prodes.toulouse.inra.fr/multalin/multalin.html) and the DNA sequences were deposited in the NCBI GenBank (Hp*TUP1*: AY145087 and Hp*SNF1*: AY155202). The amino acids in gray shade are identical among the species. The amino acids in black shade differ in just one species.

Figure 2. Disruption strategy and screening of *H. polymorpha tup1* and *snf1* mutants. The specific kanmx4 disruption modules were constructed by PCR using chimeric oligonucleotides (see *Materials and Methods* for details). Screening for positive mutants growing in G418 was made by PCR using specific primers for the Hp*TUP1* and Hp*SNF1* genes. These primers (STup1.1, STup1.2; SSnf1.1, SSnf1.2) amplify respectively 500bp and 300bp of the wild type genomic copies of each gene. Mutants in which the genes were efficiently disrupted present a single 1.5kb band as a result of the PCRs.

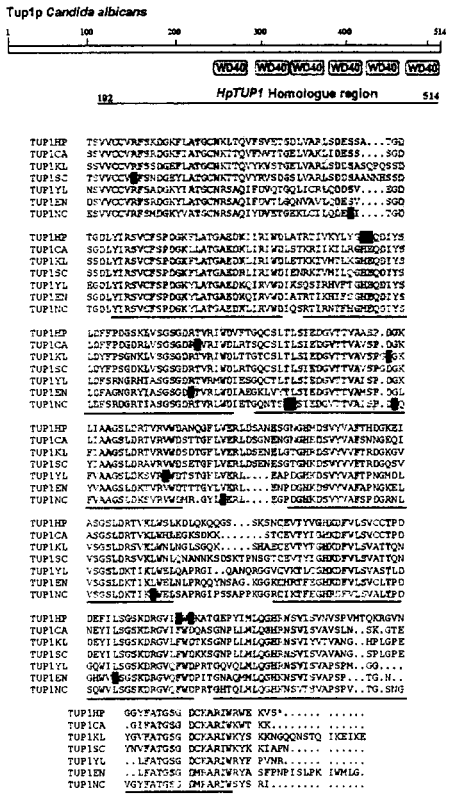
Figure 3: Analyses of *MOX*, *CAT* and *DAS* expression in WT, *tup1* and *snf1* cells. Northern blotting of samples collected after shifting cells from repressing (cells grown overnight in glucose 5% - time 0) to derepressing conditions (methanol 1%) for 6 hours: G – glucose; M – methanol. rRNA bands of samples used in each filter are shown in agarose-formaldehyde gels stained with ethidium bromide prior to blotting. The probes used are described in Materials and Methods.

Figure 4: Effects of *tup1* and *snf1* deletions on growth in glucose and methanol. A: Growth assays were performed with wild type, *swi2* and *swi3* cells in agar plates containing 5% glucose or 1% methanol as described elsewhere (Verdoucq *et al.*, 1999; Demasi *et al.*, 2001). Plates were photographed after 48 hours of growth at 37°C. **B:** Growth rates of cells from wild type, *tup1* and *snf1* strains on liquid media containing 5% glucose or 1% methanol were monitored for up to 24 hours.

Figures

Fig. 1

A) *HpTUP1* alignment



A) *HpSNF1* alignment

Fig. 2

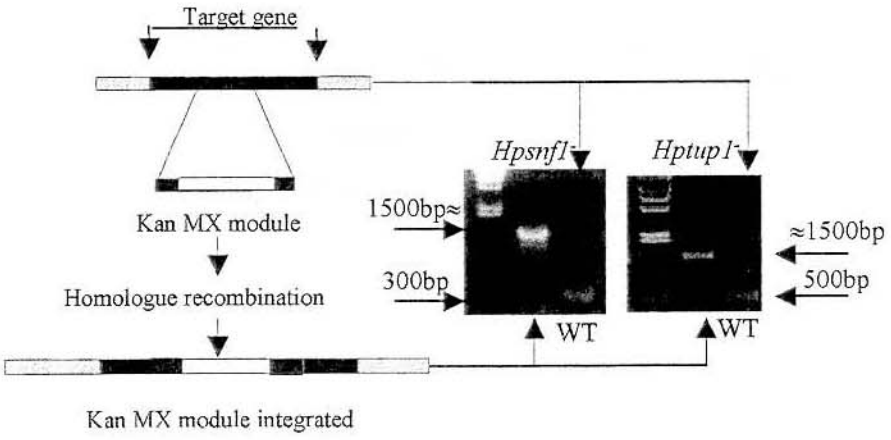

Target gene
Kan MX module
Homologue recombination
Kan MX module integrated
Hpsnf1-
Hptup1-
1500bp≈
300bp
≈1500bp
500bp
WT
WT

Fig. 3

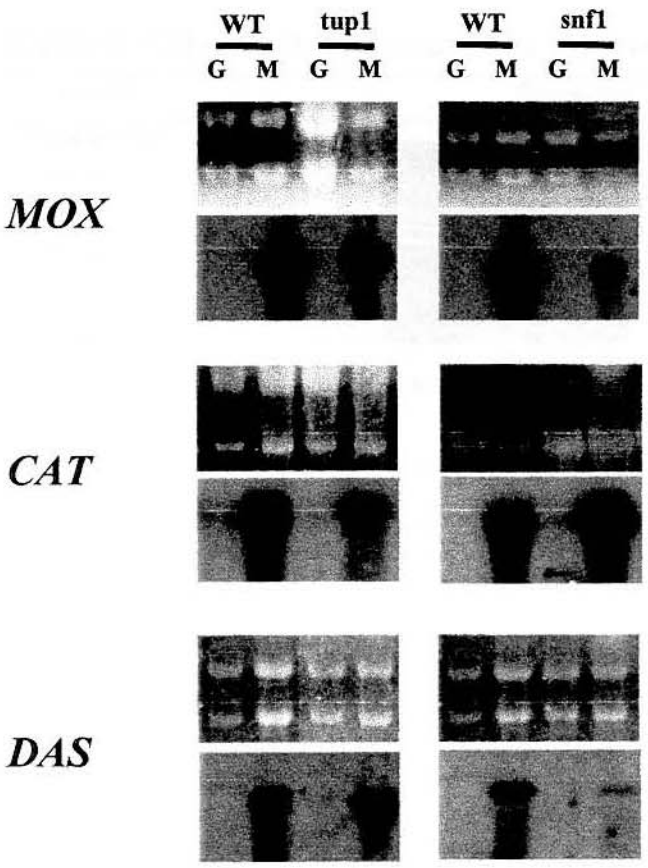


Fig. 4

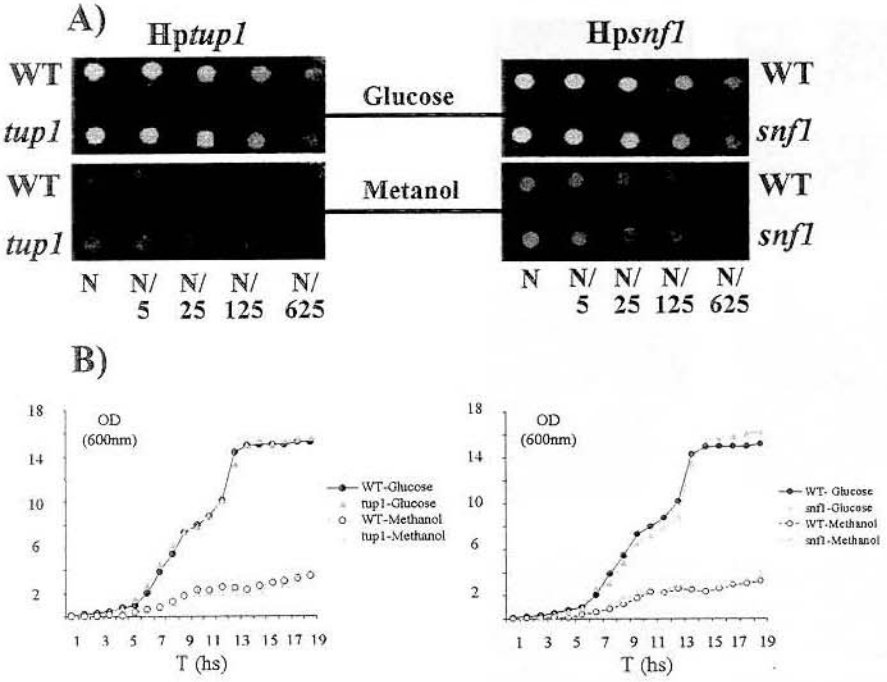

A)
Hptup1
Hpsnf1
WT
tup1
Glucose
WT
snf1
WT
tup1
Metanol
WT
snf1
N
N/5
N/25
N/125
N/625
B)
OD (600nm)
WT-Glucose
tup1-Glucose
WT-Methanol
tup1-Methanol
T (hs)
OD (600nm)
WT- Glucose
snf1-Glucose
WT-Methanol
snf1-Methanol
T (hs)

Table 1. Oligonucleotides used in this work

Primer	Sequence	Source
Degenerated		
DTup1.1	5'GGGGTACCYTTCCADATNCKNGCYTTRCARTCNCC3'	Braun and Johson, 1997.
DTup1.2	5'GGCTGCAGGGCNAYGARCARGAYATHTAYTC3'	Braun and Johson, 1997.
DSnf1.1	5'GGRRTTTCAATTRYRTCNGTRTCRAAGC3'	This study
DSnf1.2	5'GCYTGNARRTGYAARTTYGCNTGRTGNGG3'	This study
Disruption		
CDTup1.1	5'TTTCTGGCTCAGAAGCCAGCCCAGTAAGATTTTGGAGCTCGC TCGTCCCCGCCGGGTC3'	This study
CDTup1.2	5'CCCGACACCTCTCTTCTGGGTCATAACCGGTTCGACAGCAGT ATAGCGACCAGC3'	This study
CDSnf1.1	5'TCACCATCAAGGAGATCATGGAGGACGAGTAGCTTGCCTCG TCCCCGCCGGGTC3'	This study
CDSnf1.2	5'GGTTGTATTTGGTTCCATTATTCGTCGACAGCAGTATAGCGA CCAGC3'	This study
Specific and Screening		
STup1.1	5' CTTCGACTCAGCCAATGAAAGCG 3'	This study
STup1.2	5'TCTGATAACAGAGTTTCTGTGTCC3'	This study
SSnf1.1	5'TCACCATCAAGGAGATCATGGAGG3'	This study
SSnf1.2	5'GGTTGTATTTGGTTCCATTATTCG3'	This study
Dasf	5'ACTTCAACAGAGACAGGTTTG3'	This study
Dasr	5'AAATTGACAGACACAGACAGGTC3'	This study

5.4 Identificação e disrupção dos genes *SIW2* e *SWI3* de *Hansenula polymorpha* [1]

Para os genes *Swi2* e *Swi3* foi possível por meio da técnica de oligonucleotídeos degenerados identificar segmentos gênicos com alta similaridade aos correspondentes em outros organismos. As proteínas pertencentes à família de helicases do tipo Swi2p, constituintes dos complexos remodeladores de cromatina, possuem três segmentos altamente conservados: bromodomínio, DEGH e um domínio de helicase dependente de ATP. O segmento gênico identificado como *HpSWI2* codifica o domínio DEGH e grande parte do domínio de helicase, os quais acredita-se serem essênciais para as funções do complexo SWI/SNF no remodelamento da estrutura nucleossomal [56].

Fatores similares a Swi3p possuem como característica fundamental, a existência de um domínio denominado SANT (SWI-ADA -N-CoR-TFIIIB), o qual confere a habilidade ao complexo SWI/SNF de se ligar ao DNA e reconhece a sequüência consenso YAAC(G/T)G [75,129]. O fragmento identificado como homólogo a *Swi3* em *H. polymorpha* codifica grande parte do domínio SANT. As disrupções de *Swi2* e *Swi3* de *H. polymorpha* eliminaram grandes segmentos dos domínios essenciais para a função biológica destas proteínas (Manuscrito 2). A extinção dos transcritos de Hp*swi2* e Hp*swi3* foram confirmados por RT PCR (Manuscrito 2).

5.5 Efeito das disrupções na utilização de fontes alternativas de carbono

5.5.1 Crescimento em meio sólido

Assim como para os mutantes *tup1* e *snf1*, foram testadas várias fontes de carbono com a finalidade de delinear os efeitos das deleções gênicas parciais de *swi2* e *swi3*. Os resultados obtidos para os mutantes *swi2* e *swi3* estão de acordo com os descritos para *S. cerevisiae* como crescimento lento e tendência à floculação. Ainda, os efeitos das disrupções são extremamente semelhantes para ambos os mutantes, reforçando que as mutações afetaram o mesmo complexo proteico. É interessante salientar que os mutantes *swi2* e *swi3* possuem a habilidade de crescer em todas as fontes de carbono testadas. Entretanto, nos meios contendo glicerol ou metanol como única fonte de carbono é possível constatar uma diferença notável de crescimento entre as células

[1] A identificação, disrupção e caracterização de *Swi2* de *H. polymorpha* foi realizada conjuntamente com o Aluno Victor Genu o qual possui grande interesse em posicionamento de nucleossomos no promotor *Mox*. Os resultados apresentados nesta seção incluem parte do trabalho de iniciação científica da aluna Anita Paula Testa Salmazo a qual trabalhou na clonagem e caracterização do gene *Swi3* de *Hansenula polymorpha*. Os resultados são apresentados em conjunto por se complemetarem e fazerem parte de um mesmo manuscrito submetido à apreciação do corpo editorial do periódico *FEMS letters* o qual faz parte dos resultados da tese.

mutantes e as selvagens, sugerindo o envolvimento do complexo SWI/SNF no metabolismo de metanol (Fig.8).

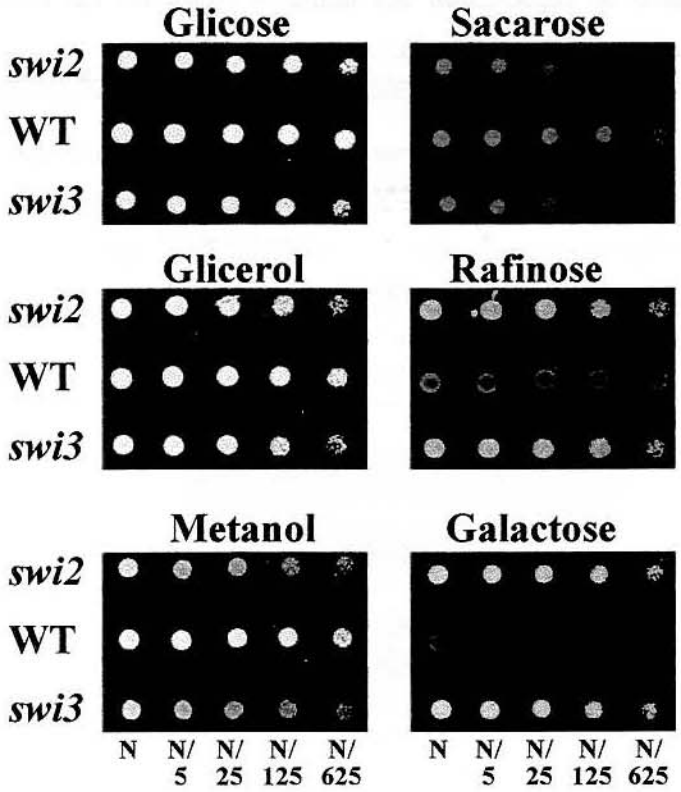


Figura 8: Efeito das disrupções dos genes *HpSwi2* e *HpSwi3* na utilização de diversas fontes de carbono. As placas foram fotografadas após 48 horas de incubação a 37 °C. Com exceção das placas contendo glicerol (2%) e metanol (1%) as concentrações finais das demais fontes de carbono foram de 3%.

O crescimento das células *swi2 e swi3* em sacarose se apresentou bastante prejudicado (Fig.8). Em *S. cerevisiae* já foi descrito que a deleção destes genes causa grande perturbação na estrutura do promotor de *SUC2* (gene que codifica a invertase) causando uma diminuição nos níveis de transcritos e conseqüente redução na utilização desta fonte de carbono [6].

Fenômeno bastante interessante é o exibido pelos mutantes quando a fonte de carbono é galactose e com pouco menos intensidade em rafinose. De acordo com os dados da literatura (Centraalbureau voor Schimmelcultures-http://www.cbs.knaw.nl/) a linhagem ATCC34438 (CBS4732) não é capaz de metabolizar galactose. Entretanto constatamos que esta linhagem possui um metabolismo basal para galactose, ainda, nas linhagens mutantes este metabolismo se apresenta extremamente acentuado (Fig.8). Os resultados são abordados no manuscrito2.

5.5.3 Curvas de crescimento em meio líquido

Adicionalmente as observações efetudas em meio sólido para a análise do aproveitamento dos carboidratos foram analizadas as capacidades dos mutantes crescerem em meio líquido contendo glicose (controle), glicerol ou metanol (Fig.9).

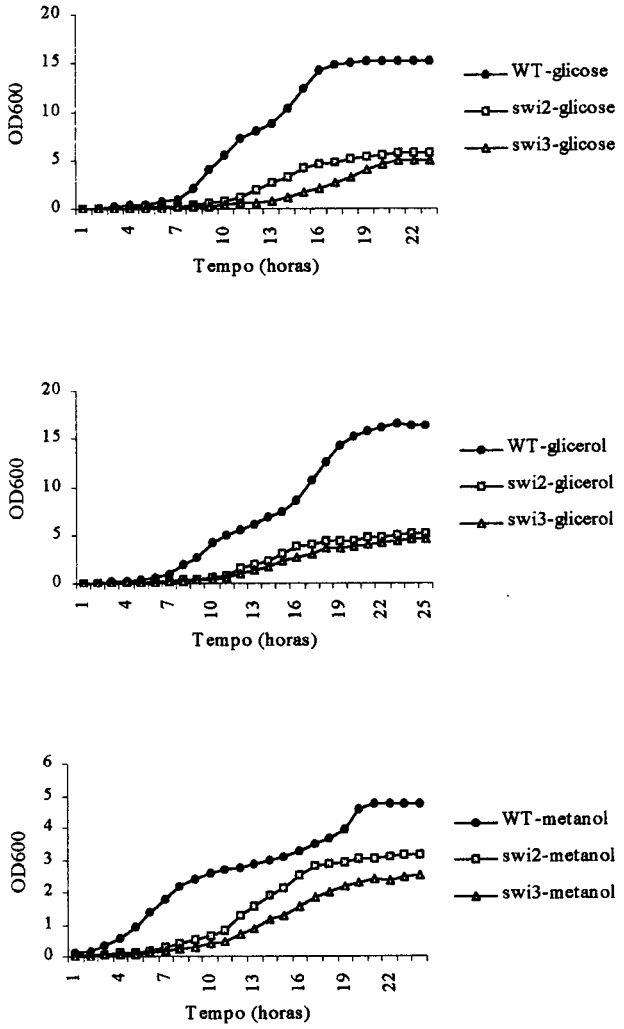


Figura 9: Curvas de crescimento de células das linhagens WT, *swi2* e *swi3* em meios líquidos contendo 3% glicose, 2% glicerol ou 1% metanol. As culturas foram monitoradas medindo as OD_{600} a cada hora por até 30 horas. Os resultados apresentados são a média de três experimentos independentes.

As leveduras *swi2* e *swi3* apresentaram um compromentimento no crescimento em todos os meios testados. Em glicose os mutantes *swi2* e *swi3* apresentaram um crescimento inferior ao apresentado pelo tipo selvagem, entrando em fase log somente após 14 horas de crescimento atingindo fase estacionária em $OD_{600} \approx 5$. Em contraste, a levedura selvagem atinge a fase

logarítmica em 7 horas e entra em fase estacionária $OD_{600}\approx15$. Em glicerol e metanol as leveduras também apresentaram uma fase lag exacerbada e atingiram ODs muito inferiores a levedura do tipo selvagem (Fig.9).

5.5.4 Expressão do gene *Mox* nos mutantes *swi2* e *swi3*

Os resultados obtidos nos experimentos de crescimento dos mutantes em meio sólido e liquido indicaram um comprometimento na utilização de metanol como fonte de carbono. Para avaliar o efeitodas disrupções estavam influenciando na transcrição do principal gene envolvido na degradação de metanol (*Mox*) foram conduzidos experimentos de "*northern blot*" (Fig.10).

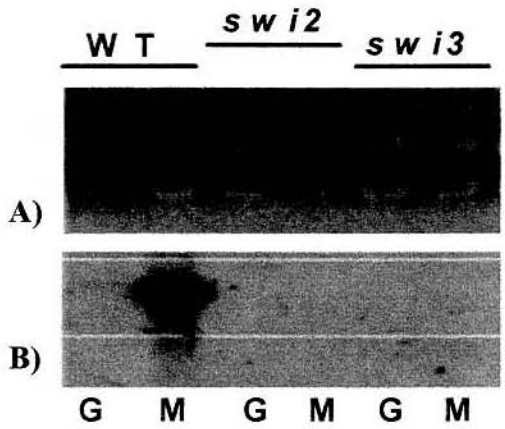


Figura 10: Análise da expressão do gene *MOX* em linhagens selvagens (WT) e mutantes (*swi2* e *swi3*). As células foram crescidas por 16 horas 37°C em YNB + 3% de glicose. Após este período as células foram transferidas ou para meio repressor (**G**-Glicose 5%) ou submetidas à indução de *Mox* (**M**- Metanol 1%) por um período de 6 horas. **A)** Gel de agarose/MOPS/formamida mostrando o RNA ribossomal de células selvagens e das células *swi2* e swi3. **B)** Resultado de northern blotting e hibridação utilizando como sonda fragmento do gene *Mox* marcado com P^{32}.

A levedura selvagem apresentou forte indução na transcrição de *Mox* sob as condições testadas, entretanto, não foi possível a detecção de transcritos *Mox* nos mutantes *swi2* e *swi3*. Os experimentos de northern blot foram repetidos sob diversas condições (dados não apresentados). Este resultado é particularmente curioso, uma vez que as células mutantes conseguem crescer em meio contendo metanol como única fonte de carbono. Para se confirmar se as células eram realmente de *H. polymorpha* examinos suas características ao microscópio, amplificamos um segmento da região regulatória do gene *Mox* e seqüenciamos os produtos de amplificação. Ambos os experimentos comprovaram que os mutantes *swi2* e *swi3* eram *H. polymorpha*. Estes resultados sugerem que a disrupção de Hp*Swi2* ou Hp*Swi3* levam a uma dramática redução nos níveis de transcrição do gene *Mox*.

Disruption of homologues of the *SWI2* and *SWI3* genes restore a quiescent galactose metabolism in *Hansenula polymorpha*

5.6 MANUSCRITO 2: Submetido ao periódico: *FEMS Letters*

Marcos A. OLIVEIRA [1, #]; Victor GENU [1, 2, #]; Anita P. T. SALMAZO [1]; Bárbara C. R. MONTE-MOR [1] and Gonçalo A. G. PEREIRA [1, *]

1 - Laboratório de Genômica e Expressão - Departamento de Genética e Evolução - Instituto de Biologia - Universidade Estadual de Campinas - Campinas, SP, Brazil.

2 - Departamento de Bioquímica - Instituto de Biologia - Universidade Estadual de Campinas - Campinas, SP, Brazil.

# - These authors contributed equally to this work

* - Corresponding author:

Gonçalo Pereira - Laboratório de Genômica e Expressão (LGE), Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas - Campinas, SP, Brazil. CEP: 13083-970.

E-mail: goncalo@unicamp.br

Tel/Fax: 00 55 19 3788 6237 / 00 55 19 3788 6235

Web site: www.lge.ibi.unicamp.br

ABSTRACT

Hansenula polymorpha belongs to the specialized group of methylotrophic yeasts and, like most species of this group, is unable to grow in galactose as a sole carbon source. In this work, mutants harboring partial deletions of homologues of the *S. cerevisiae SWI2(SNF2)* and *SWI3* were produced to understand possible contributions of the SWI-SNF complex in *H. polymorpha* carbohydrate metabolism. Surprisingly, the deletions rendered cells able to utilize galactose, in contrast to the original wild type phenotype. These observations imply that *H. polymorpha* has the genes for the galactose metabolism, which were however turned off during evolution. Our results suggest that SWI-SNF may play an important role in the silencing of *H. polymorpha GAL* genes and that its inactivation restore a quiescent pathway of galactose utilization.

Keywords: SWI-SNF, *Hansenula polymorpha*, *GAL*, silencing, galactose, glucose repression.

INTRODUCTION

The yeast *Hansenula polymorpha* (syn. *Pichia angusta*) has been studied for several years due to its ability to grow in methanol as a sole carbon and energy source. Most genes involved in this metabolism are repressed by glucose or ethanol and activated at different levels by growth in other unrelated carbon sources (e.g. glycerol < sorbitol < methanol) (Veenhuis et al., 1983). This kind of activation can be regarded as derepression because it is triggered by the absence of a compound responsible for repression. In the model yeast *Saccharomyces cerevisiae*, this process has been found for several genes involved in respiration or in the utilization of alternative carbon sources. For example, the genes encoding invertase (*SUC2*) or alcohol dehydrogenase 2 (*ADH2*) are repressed by glucose and derepressed by the absence of this compound [65]. No specific inducer is necessary. In contrast, other genes, like the *GAL* genes, need both the absence of glucose and the presence of an inducer (galactose) to be fully activated (reviewed in Lohr et al., 1995). Absence of glucose alone allows only low expression of the *GAL* genes. Under this condition, they are considered derepressed.

In order to understand the components of these regulatory systems, in earlier experiments different *S. cerevisiae* mutants have been isolated due to their inability to derepress the *SUC2* gene and therefore to ferment sucrose (*sucrose non fermenter - snf* mutants). The *snf* mutants presented pleiotropic effects, affecting the derepression of several other genes involved in carbon utilization, like *GAL1*, *GAL10*, and *ADH2* (Abrams et al., 1986; Peterson and Herskowitz, 1992). Using similar strategies, mutants have been isolated for other independent cellular processes, like mating type switching (*swi* mutants) and Ty-mediated transcription (*tye* mutants). Remarkably, it has been demonstrated that regulatory genes affecting these processes were identical: *SWI2= SNF2*

and *TYE3*; *SNF5* = *SWI10* and *TYE4* (Ciriacy et al., 1991) and *SWI3=TYE2* (Lohning et al., 1993). These genes encoded subunits of an ATP-dependent chromatin remodeling complex, SWI/SNF (Cairns et al., 1994; Peterson et al., 1994), which has counterparts in several organisms, from yeast to human (reviewed in Vignali et al., 2000).

It has been demonstrated that *H. polymorpha* methanol oxidase (*MOX*) gene is properly regulated in *S. cerevisiae*, a non methylotrophic yeast (Pereira and Hollenberg, 1996). The *MOX* promoter is repressed by glucose and derepressed by the absence of this compound. Thus, it is likely that *H. polymorpha* and *S. cerevisiae* present similar regulatory systems to control sets of genes involved in carbohydrate utilization. We have partially cloned and disrupted *H. polymorpha* genes homologous to *MIG1* and *TUP1*, involved in glucose repression in *S. cerevisiae*, and the genes *SNF1*, *SWI2* (*SNF2*) and *SWI3*, generally regarded as involved in derepression.

In this work, we present the analysis of Hp*SWI2* and Hp*SWI3* cloning and disruption. In general, *swi2* or *swi3* mutations led to similar effects, indicating that they also belong to a multiprotein complex in *H. polymorpha*. The cells presented significant reduction in growth on several carbon sources, like glucose, sucrose, glycerol and methanol. Surprisingly, in contrast to *H. polymorpha* wild type cells, that are unable to utilize galactose as carbon source (www.cbs.knaw.nl), the mutants acquired the ability to grow on galactose. This fact indicates that *H. polymorpha* has a quiescent galactose pathway, whose genes are kept silent by factors somehow connected with the SWI/SNF complex. The relevance of these findings will be discussed.

MATERIALS AND METHODS

Yeast strains and growth conditions – The *H. polymorpha* strain used to disrupt homologues of the *SWI2* and *SWI3* genes was isogenic to strain CBS4732 (wild type - WT). YPD (1% yeast extract, 2% peptone, 2% glucose) was used as rich medium. SD medium (0.17% YNB; 0.5% ammonium sulfate) was used as a synthetic medium and was supplemented with 5% glucose, 3% sucrose, 1% methanol, 2% glycerol, 1% glucosamine, or 3% galactose. When used without any carbon source, SD medium is referred to as SD^0. Cells were grown at 37°C.

Bacterial strains, plasmids and oligonucleotides – *Escherichia coli* strain DH10B (Invitrogen) was used for all DNA manipulations. Plasmid pF6A-kanmx4 (Wach et al., 1994) was used for construction of the disruption modules. pGEM-T (Promega) and pUC18 were used for cloning of PCR products. The oligonucleotides used in this study are listed in Table 1.

Cloning and disruption of *H. polymorpha* homologues of *SWI2* and *SWI3* genes - Strategy of cloning was based on degenerated oligonucleotides (see Table 1). We designed degenerated primers for *SWI2* (Dswi2F and Dswi2R) and *SWI3* (Dswi3F and Dswi3R) gene sequences from other species (*S. cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*). These primers were used to conduct PCRs with genomic DNA from *H. polymorpha* (WT strain) as the template. Selected PCR products were cloned into pGEM-T (Promega) and sequenced. The DNA sequences obtained were analyzed by BlastX. Sequences presenting high similarities with *SWI2* and *SWI3* (AY145088 and AY142211, respectively; see Fig. 1) were used to design chimerical disruption oligonucleotides and specific primers for subsequent screening. Disruption oligonucleotides (see Table 1 and Fig. 2) present ~30bp of the genes to be disrupted in their 5' extremities; the 3' extremities

are composed of ~25bp complementary to the Multiple Cloning Site of the pF6A plasmid where the kanmx4 disruption module was integrated (Wach et al., 1994). Specific *SWI2* and *SWI3* kanmx4 modules were constructed with these primers (Cdswi2F+Cdswi2R and Cdswi3F+Cdswi3R, respectively) by standard PCRs using the pF6A-kanmx4 plasmid as a template. These linear disruption modules bear, at both 5' and 3' ends, sequences homologous to *SWI2* and *SWI3* genes and were used to transform *H. polymorpha* WT electro-competent cells. The kanmx4 module confers resistance to the drug G418 (Wach et al., 1994), which was used for preliminary screening. Positive transformants growing in the presence of G418 (Amersham Pharmacia) were confirmed by colony PCR and also by PCRs with genomic DNAs.

PCR Conditions – All PCRs were made in a final volume of 50 µl using kits (Promega) and following instructions provided by the manufacturer. All reactions were carried out on a 9700 thermocycler (Applied Biosystems). The programs used for degenerated or specific reactions (respectively) were: 96°C for 4min, followed by 30 cycles at the subsequent temperatures and times: 94°C for 30 sec, 40°C or 50°C for 30 sec and 72°C for 1 min. For construction of the disruption modules, the program used was: 96°C for 4min, followed by 30 cycles: 94°C for 30 sec, 70°C for 45 sec and 72°C for 1.5 min.

RT-PCR - Total RNA samples were extracted (conditions described in *Results*) by the hot acid phenol method (Ausubel et al., 1998) and then treated with DNase I (Invitrogen). Sample quantification was done by measuring its A_{260} in a spectrophotometer. RT reactions were performed with Superscript II Reverse Transcriptase (Invitrogen) following instructions provided by the manufacturer. Primers Sswi2R and Sswi3R were

used for first strand synthesis; these were combined with primers Sswi2F and Sswi3F (respectively) for the standard subsequent PCRs.

Growth assays on solid media – Yeast cells were grown in liquid SD media plus 5% glucose till they reached an $OD_{600} = 0.6$, harvested and resuspended in the same media in an $OD_{600} = 0.2$. When they reached $OD_{600} = 0.6$ again, they were re-diluted to $OD_{600} = 0.2$ with SD^0 media. From this point on, serial dilutions were made with SD^0: N ($OD_{600} = 0.2$), N/5, N/25, N/125 and N/625. Finally, 12µl of each dilution were dropped on agar plates supplied with different carbon sources (indicated in *Results*). The plates were incubated at 37°C for up to three days and photographed. These procedures were based on Verdoucq et al., (1997) and Demasi et al., (2002).

Growth characterization – Cells from the three strains were grown till log-phase in YPD, harvested and resuspended in the appropriate media (indicated in *Results*) at an OD_{600}= 0.01. Aliquots were collected each hour for 29 hours and the absorbance at λ= 600nm was measured. The results presented are the average of three independent experiments.

RESULTS

Identification of *H. polymorpha SWI2* and *SWI3* homologues

H. polymorpha DNA fragments were obtained by PCR using degenerated primers for important regions of the *SWI2* and *SWI3* genes (Table 1 and Fig. 1). The putative Hp*SWI2* fragment encompassed a large predicted helicase domain, which includes the DEGH box (Fig. 1A). These regions are believed to be important for Swi2p function *in vivo* (Du et al., 1998). The putative Hp*SWI3* gene fragment bears a region with high similarity to related proteins, including part of the SANT domain (Fig. 1B) responsible for interacting with DNA (Aasland et al., 1996). Therefore, it was concluded that the DNA fragments encompass important functional regions of the *H. polymorpha SWI2* and *SWI3* genes.

Construction of *swi2* and *swi3 H. polymorpha* mutants by PCR

The Hp*SWI2* and Hp*SWI3* fragments were used to construct disruption modules as described in the *Materials and Methods*. *H. polymorpha* wild type cells were transformed and positive colonies growing on G418 were selected. About 150 transformants for each gene were screened for the presence of the disruption module by PCR amplification using specific primers for the Hp*SWI2* and Hp*SWI3* gene fragments. Fig. 2A shows the PCR products of correct integrations, observed for 5% of the transformants, in comparison to amplification of the wild type fragments. Growth abilities of *swi2* and *swi3* cells on plates containing G418 were re-checked in comparison to wild type cells (Fig. 2B).

Transcription of the Hp*SWI2* and Hp*SWI3* genes were analyzed by RT-PCR using total RNA extracted from WT, *swi2* and *swi3* cells growing in 2% glycerol and the same set of specific primers from Fig. 2A. Fig. 2C shows that both genes are transcribed when

grown in glycerol in wild type cells, but not in the mutants, confirming that the deletions abolished transcription. To make certain that the selected colonies were actually *H. polymorpha* cells, we amplified a specific 270bp segment of the *MOX* promoter region (Ledeboer et al., 1985) from genomic DNAs (Fig. 2D) and sequenced the resulting PCR fragments (data not shown).

Effects of *swi2* and *swi3* mutations in *H. polymorpha* on carbohydrate utilization

Wild type, *swi2* and *swi3* single colonies were inoculated from glucose containing SD plates to SD liquid media containing fermentative (5% glucose or 3% sucrose), non-fermentative (1% methanol or 2% glycerol) or nonmetabolizable (1% glucosamine or 3% galactose) carbon sources. Visual inspection indicated that *swi2* and *swi3* cells presented similar growth in all tested media, however growth was significantly reduced in comparison to wild type cells on glucose, sucrose, methanol and glycerol. As expected, none of the cell types could grow on glucosamine. Surprisingly, in contrast to wild type, both *swi2* and *swi3* mutants could efficiently grow in galactose. *H. polymorpha* is normally characterized as a galactose minus species (www.cbs.knaw.nl).

Growth was also tested on plates. Cells were grown in liquid media to the same log phase OD_{600}, then serially diluted and dropped on plates (described in *Materials and Methods*). Thus, drops of the same dilution have comparable amounts of cells and the density of the drop area through time will be proportional to the growth rate. The results are shown in Fig. 3A. The strains presented comparable growth on glucose. On plates containing glycerol or sucrose, *swi2* and *swi3* cells displayed reduced growth in comparison to wild type cells. On galactose plates, only the mutants were able to grow efficiently.

To further confirm these results, we monitored the growth abilities of the cells in liquid media containing glucose or galactose (Fig. 3B). Consistent with the previous data, all strains were able to grow on glucose, but *swi2* and *swi3* cells presented a longer lag phase and reached stationary phase at lower ODs. Wild type cells entered log phase after 6 hours of growth and reached stationary phase at an $OD_{600} = 15$; in contrast, *swi2* and *swi3* entered log phase after 11 hours and reached stationary phase at $OD_{600} = 7$ and 6, respectively. On galactose, wild type and mutant cells presented similar features at lag phase, which lasted about 11 hours. After that, however, the mutants entered a strong log phase, reaching stationary phase in an $OD_{600} \sim 11$ after 25 hours of growth. In contrast, wild type cells failed to enter log phase and the culture maintained a stable $OD_{600} = 1$ all through the experiment.

DISCUSSION

In this work, we cloned functional important regions of the *H. polymorpha SWI2* and *SWI3* genes and showed the effect of the disruption of either gene in this yeast.

In *S. cerevisiae, SWI2* and *SWI3* encode components of a multiprotein complex, named SWI/SNF (Cairns et al., 1994; Peterson et al., 1994), implicated in the transcriptional activation of several differently regulated promoters (Abrams et al., 1986; Peterson and Herskowitz, 1992; Laurent and Carlson, 1992). Genetic and biochemical evidence indicates that the SWI/SNF complex relieves chromatin-mediated repression at specific promoters in an ATP-dependent manner, facilitating interactions between activators and chromatin (Hirschhorn et al., 1992; Coté et al., 1994; reviewed in Vignali et al., 2000). Specificity seems to be achieved by the interaction of SWI/SNF with transactivation domains of given factors, which recruits the complex to particular promoters (Yudkovsky et al., 1999). Loss of components of this system lead to failure in the activation of several genes, like *SUC2* (Matallana et al., 1992; Wu and Winston, 1997) and *GAL1-10* (Peterson and Herskowitz, 1992). Consequently, cells become defective for growth on sucrose and galactose (Neigeborn and Carlson, 1984).

In *H. polymorpha,* deletion of either *SWI2* or *SWI3* led to a general reduction in growth on all tested fermentative and non-fermentative carbon sources. Growth on glucose was less affected in comparison to other sources, like sucrose and glycerol, which requires the derepression of additional genes encoding enzymes like invertase (Carlson and Botstein, 1982) and glycerol kinase (Pavlik et al., 1993). Thus, this phenotype is compatible with the existence of a SWI/SNF complex in *H. polymorpha* with functions similar to those described for *S. cerevisiae*; i.e. facilitate the expression of a number of genes (Sudarsanam and Winston, 2000). However, the mutations lead additionally to a

fully unpredicted result. In contrast to *S. cerevisiae*, which need SWI/SNF to grow on galactose (Peterson and Herskowitz, 1992), the *H. polymorpha swi/snf* cells acquired the ability to efficiently utilize this compound, which is not considered as a carbon source for this species (www.cbs.knaw.nl).

Though puzzling, these facts led to the following conclusions: first, *H. polymorpha* has the minimal set of genes necessary for galactose utilization; second, these genes are silenced in wild type cells; third, the SWI/SNF complex is somehow involved in this silencing.

Galactose utilization involves basically two biochemical processes: uptake of galactose by a hexose transporter and conversion of galactose into glucose-1-phosphate (Lohr et al., 1995). Specialized enzymes and genes for this process are: galactokinase (Sc*GAL1*); galactose-1-phosphate uridylyltransferase (Sc*GAL7*) and UDP-glucose 4-epimerase (Sc*GAL10*) (Frey, 1996). No alternative pathway has yet been described. To date, no *H. polymorpha* sequence with significant similarity to these specific genes have been reported. In our laboratory, sequencing and analyses of ~ 7000 random cDNA and genomic shot-gun fragments failed to identify homologous to the *GAL* genes (data not shown). However, our results indicate that they exist and are potentially functional, but are rendered silent.

During evolution, selection forces drive the organisms for efficiency. Consequently, over time organisms gain or lose functions according to the environment, which they live. Silencing of metabolic pathways may have the function of saving energy and avoiding the expression of genes whose products are no longer necessary. It is often achieved by loss-of-function of gene regulatory or coding regions, either by the accumulation of point mutations (Funchain et al., 2000), deletions (Cooper et al., 2001)

or by silencing through chromatin packaging, like heterochromatin (Donze and Kamakaka, 2002). In the two initial cases, the restoration of biochemical pathways can only be accomplished by re-introduction of the missing or malfunctioning fragments of the genome; and in the last case, by relieving nucleosomes or higher order structures of regulatory regions. However, if a gene is kept silent for long periods, it will no longer be under selection. Therefore, the cells will have no disadvantage if it is rendered worthless by the accumulation of mutations.

This does not seem to be the case with the putative *H. polymorpha GAL* genes. While currently unable to grow on galactose, *H. polymorpha* sequences seem to be functional. That means that silencing of this pathway may be a recent event and this view is consistent with the fact that closely related methylotrophic yeast species, like *Pichia ofunaensis* for example, can use this carbon source. Furthermore, remarkable cases of maintenance of a biochemical pathway through long evolutionary periods have been demonstrated in higher eukaryotes. For example, the development of teeth in birds can be re-activated once given the appropriate stimulus (Chen et al., 2000). This situation implies that, although silent, the genes involved in teeth formation in birds have been kept potentially functional for at least 60 million years.

Finally, the most interesting point is the fact that the galactose utilization pathway in *H. polymorpha* is kept silent by SWI/SNF, a complex that is generally characterized as a derepressor of the *GAL* genes in *S. cerevisiae* (Peterson and Herskowitz, 1992). Indeed, it has been demonstrated that homologous regulators can play distinct roles in different organisms (Guarente and Bermingham-McDonogh, 1992). Here, two scenarios appear possible to explain this kind of connection. First, genome-wide expression of a *S. cerevisiae swi2* strain has shown that the SWI-SNF complex works also in the repression

of a subset of genes (Holstege et al., 1998; Sudarsanam et al., 2000). It has also been suggested that the RSC chromatin remodeling complex, which is highly similar to SWI-SNF (Cairns et al., 1996), may participate in nucleosome assembly on the TATA box of the *CHA1* promoter, leading to repression of the *CHA1* gene (Moreira and Holmberg, 1999). Accordingly, it is possible that the SWI-SNF complex could exert a similar repressive role on any of the putative *H. polymorpha GAL* promoters. Alternatively, the SWI-SNF complex could act as a transcriptional co-repressor, enabling an unknown repressor to gain constitutive access to the *H. polymorpha GAL* gene promoter. Second, the effect on galactose metabolism observed in *H. polymorpha swi2* and *swi3* mutants could be indirect. It has been reported that *swi2* mutations cause global chromatin perturbations in *S. cerevisiae* (Peterson and Herskowitz, 1992). Hence, if the *H. polymorpha GAL* genes are silenced due to specific nucleosome positioning of their promoters, these global chromatin modifications could affect the accessibility in these regions, allowing transcription.

Regardless of the SWI/SNF action mechanism, the fundamental point of this work is the existence of quiescent galactose metabolism in *H. polymorpha* that can be reactivated by the depletion of a regulatory complex generally accepted as an activator in closely related species. These facts show how evolution can rearrange both functional and regulatory genes to create new metabolic patterns.

ACKNOWLEDGMENTS

We thank Thiago C. Lopes for the kind gift of the pF6A-kanmx4 plasmid and Carla C. J. Maria, Ana P. D. Demasi, Gisele Monteiro and Anderson F. Cunha for helpful discussion; we are indebted to Sérgio F. Reis for his invaluable help on the manuscript and to Lyndel W. Meinhardt for critically reviewing the manuscript. This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - Brazil) to M. A. O., V. G. and G. G. P.

REFERENCES

1. Aasland R, Stewart AF, Gibson T (1996) The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. Trends Biochem Sci 21: 87-88
2. Abrams E, Neigeborn L, Carlson M (1986) Molecular analysis of SNF2 and SNF5, genes required for expression of glucose-repressible genes in Saccharomyces cerevisiae. Mol Cell Biol 6: 3643-3651
3. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1998) Current protocols in molecular biology. John Wiley & Sons Inc., New York, USA
4. Cairns BR, Kim YJ, Sayre MH, Laurent BC, Kornberg RD (1994) A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5 and SNF6 gene products isolated from yeast. Proc Natl Acad Sci USA 91: 1950-1954
5. Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis, L, Erdjument-Bromage H, Tempst P, Du J, Laurent B, Kornberg RD (1996) RSC, an essential, abundant chromatin-remodeling complex. Cell 87: 1249-1260
6. Carlson M, Botstein D (1982) Two differentially regulated mRNAs with different 5'ends encode secreted with intracellular forms of yeast invertase. Cell 28: 145-154
7. Chen Y, Zhang Y, Jiang T, Barlow AJ, St Amand TR, Hu Y, Heaney S, Francis-West P, Chuong CM, Maas R (2000) Conservation of early odontogenic signaling pathways in *Aves*. Proc Natl Acad Sci USA 97: 10044-10049
8. Ciriacy M, Freidel K, Lohning C (1991) Characterization of trans-acting mutations affecting Ty and Ty-mediated transcription in Saccharomyces cerevisiae. Curr Genet 20: 441-448

9. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183: 2834-2841

10. Côté J, Quinn J, Workman JL, Peterson CL (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF protein complex. Science 265: 53-60

11. Demasi AP, Pereira GAG, Netto LES (2001) Cytosolic thioredoxin peroxidase I is essential for the antioxidant defense of the yeast with dysfunctional mitochondria. FEBS Lett 509: 430-434

12. Donze D, Kamakaka RT (2002) Braking the silence: how heterochromatin gene repression is stopped in its tracks. Bioessays 24: 344-349

13. Du J, Nasir I, Benton BK, Kladde MP, Laurent BC (1998) Sth1p, a *Saccharomyces cerevisiae* Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins. Genetics 150: 987-1005

14. Frey PA (1996) The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. FASEB J 10: 461-470

15. Funchain P, Yeung A, Stewart JL, Lin R, Slupska MM, Miller J.H. (2000) The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics 154: 959-970

16. Gancedo JM (1998) Yeast carbon catabolite repression. Microbiol Mol Biol Rev 62: 334-361

17. Guarente L, Bermingham-McDonogh O (1992) Conservation and evolution of transcriptional mechanisms in eukaryotes. Trends Genet 8: 27-32

18. Hirschhorn JN, Brown SA, Clark CD, Winston F (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev 6: 2288-2298

19. Holstege FCP, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA (1998) Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95: 717-728

20. Laurent BC, Carlson M (1992) Yeast SNF2/SWI2, SNF5, and SNF6 proteins function coordinately with the gene-specific transcriptional activators GAL4 and Bicoid. Genes Dev 6: 1707-1715

21. Ledeboer AM, Edens C, Visser C, Bos JW, Verrips CT, Janowicz Z, Eckart M, Roggenkamp R, Hollenberg, CP (1985) Molecular cloning and characterization of a gene coding for methanol oxidase in *Hansenula polymorpha*. Nucleic Acids Res 13: 3063-3083

22. Lohning C, Rosenbaum C, Ciriacy M (1993) Isolation of the TYE2 gene reveals its identity to SWI3 encoding a general transcription factor in Saccharomyces cerevisiae. Curr Genet 24: 193-199

23. Lohr D, Venkov P, Zlatanova J (1995) Transcriptional regulation in the yeast *GAL* gene family: a complex genetic network. FASEB J 9: 777-787

24. Matallana E, Franco, L, Perez-Ortin JE (1992) Chromatin structure of the yeast SUC2 promoter in regulatory mutants. Mol Gen Genet 231: 395-400

25. Moreira JMA, Holmberg S (1999) Transcriptional repression of the yeast *CHA1* gene requires the chromatin-remodeling complex RSC. EMBO J 18: 2836-2844

26. Neigeborn L, Carlson M (1984) Genes affecting the regulation of SUC2 expression by glucose repression in Saccharomyces cerevisiae. Genetics 108: 845-858

27. Pavlik P, Simon M, Schuster T, Ruis H (1993) The glycerol kinase (GUT1) gene of Saccharomyces cerevisiae: cloning and characterization. Curr Genet 24: 21-25

28. Pereira GG, Hollenberg CP (1996) Conserved regulation of the *Hansenula polymorpha MOX* promoter in *Saccharomyces cerevisiae* reveals insights in the transcriptional regulation by Adr1p. Eur J Biochem 238: 181-191

29. Peterson CL, Dingwall A, Scott MP (1994) Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. Proc Natl Acad Sci USA 91: 2905-2908

30. Peterson CL, Herskowitz I (1992) Characterization of the yeast *SWI1*, *SWI2*, and *SWI3* genes, which encode a global activator of transcription. Cell 3: 573-583

31. Sudarsanam P, Iyer VR, Brown PO, Winston F (2000) Whole-genome expression analysis of snf/swi mutants of Saccharomyces cerevisiae. Proc Natl Acad Sci USA 97: 3364-3369

32. Sudarsanam P, Winston F (2000) The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet 16: 345-350

33. Veenhuis M, Van Dijken JP, Harder W (1983) The significance of peroxisomes in the metabolism of one carbon compound is yeast. Advances in Microbial Physiology 24: 1-82

34. Verdoucq L, Vignols F, Jacquot JP, Chartier Y, Meyer Y (1999) *In vivo* characterization of a Thioredoxin h target protein defines a new Peroxiredoxin family. J Biol Chem 274: 19714-19722

35. Vignali M, Hassan AH, Neely KE, Workman JL (2000) ATP-dependent chromatin-remodeling complexes. Mol Cell Biol 20: 1899-1910

36. Wach A, Brachat A, Pohlmann R, Philippsen P (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast 13: 1793-1808

37. Wu L, Winston F (1997) Evidence that Snf-Swi controls chromatin structure over both TATA and UAS regions of the SUC2 promoter in Saccharomyces cerevisiae. Nucleic Acids Res 25: 4230-4234

38. Yudkovsky N, Logie C, Hahn S, Peterson CL (1999) Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. Genes Dev 13: 2369-2374

LEGENDS

Figure 1. Identification of Hp*SWI2* and Hp*SWI3* gene fragments. Alignments of the predicted HpSwi2p (**A**) and HpSwi3p (**B**) proteins with other known homologues: H.P.- *H. polymorpha*; S.C.- *S. cerevisiae*; G.G.- *Gallus gallus*; H.S.- *Homo sapiens*; D.M.- *Drosophila melanogaster*; C.E.- *Caenorhabditis elegans*; S.P.- *Schizosaccharomyces pombe*; M.M.- *Mus musculus*. The known conserved domains identified in *S. cerevisiae* proteins are schematically shown on top of the alignments (see text for details). Alignments were done with the MultAlin tool (www.prodes.toulouse.inra.fr/multalin/multalin.html). The amino acids in gray shade are identical among the species. The amino acids in black shade differ in just one species.

Figure 2. Disruption strategy and screening of *H. polymorpha swi2* and *swi3* mutants. A: The specific kanmx4 disruption modules were constructed by PCR using chimeric oligonucleotides (see *Materials and Methods* for details). Screening for positive mutants growing in G418 was made by PCR using specific primers for the Hp*SWI2* and Hp*SWI3* genes (positions indicated by arrows). These primers (Swi2F, Swi2R; Swi3F, Swi3R) amplify respectively 400 and 300bp of the wild type genomic copies of each gene. Mutants in which the genes were efficiently disrupted present a single 1.5kb band as a result of the PCRs. **B:** Growth of wild type, *swi2* and *swi3* cells on plates containing G418. **C:** RT-PCR with RNA samples extracted from cells growing in 2% glycerol for 12hs. The primers Swi2R and Swi3R were used for first strand synthesis; these were combined with the primers Swi2F and Swi3F for the subsequent PCRs. **D:** *swi2* and *swi3* cells were confirmed to be *H. polymorpha* mutants by amplifying a 270bp specific fragment of the *MOX* promoter (Ledeboer *et al.*, 1985). The same fragment amplified from WT genomic DNA is shown as a control.

Figure 3. Effects of different carbon sources on growth. A: Growth assays were performed with wild type, *swi2* and *swi3* cells in agar plates containing 5% glucose, 3% sucrose, 2% glycerol, or 3% galactose as described in the *Materials and Methods*. Plates containing glucose, glycerol and sucrose were photographed after 48 hours of growth; plates containing galactose were photographed after 72 hours of growth. **B:** Growth rates

of cells from wild type, *swi2* and *swi3* strains on liquid media containing 5% glucose or 3% galactose were monitored for up to 29 hours.

Figures

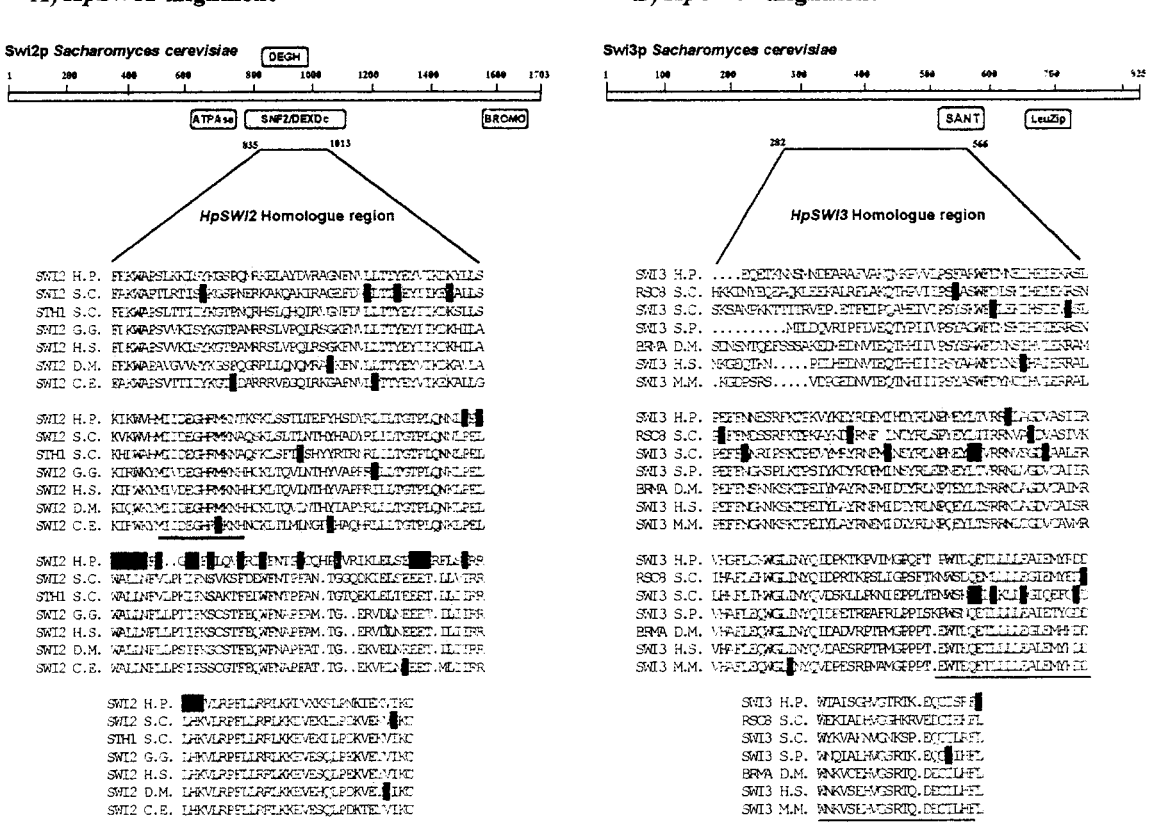


Fig. 1

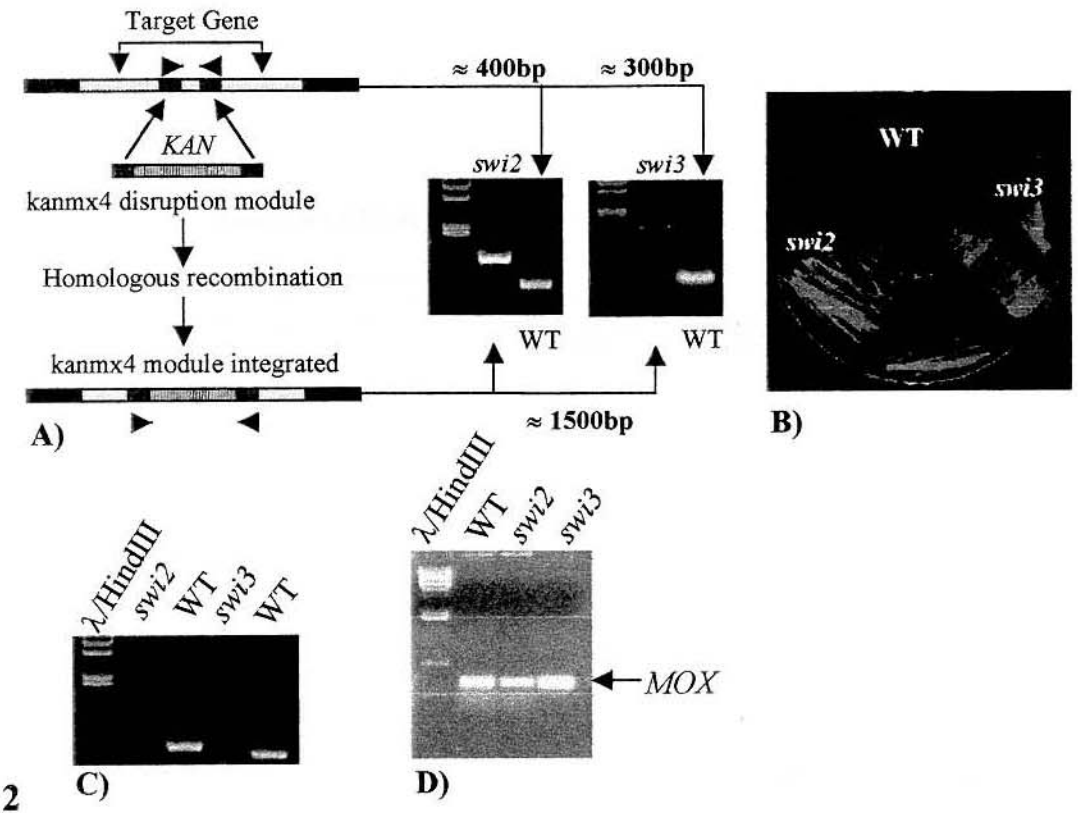
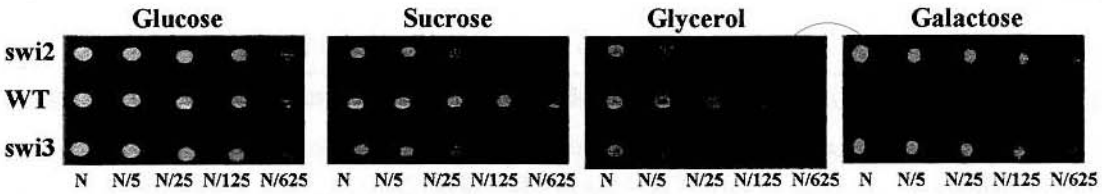

Target Gene
≈ 400bp
≈ 300bp
KAN
swi2
swi3
kanmx4 disruption module
Homologous recombination
kanmx4 module integrated
WT
WT
≈ 1500bp
A)
WT
swi3
swi2
B)
λ/HindIII
swi2
WT
swi3
WT
C)
λ/HindIII
WT
swi2
swi3
MOX
D)

Fig. 2

A)


Glucose
Sucrose
Glycerol
Galactose
swi2
WT
swi3
N N/5 N/25 N/125 N/625
N N/5 N/25 N/125 N/625
N N/5 N/25 N/125 N/625
N N/5 N/25 N/125 N/625

B)

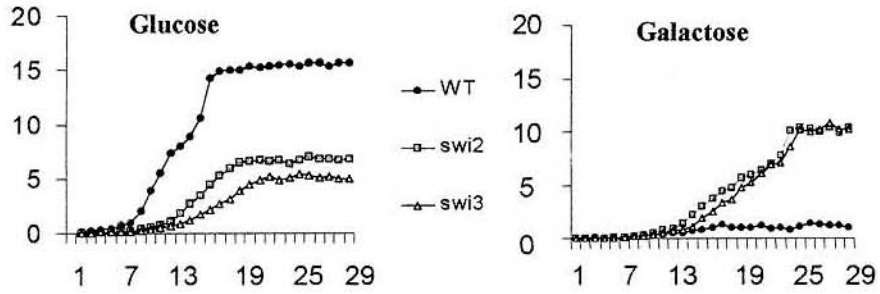

Glucose
Galactose
WT
swi2
swi3
20
15
10
5
0
1 7 13 19 25 29

Fig. 3

Table 1.

Primers	Sequence
Degenerated	
Dswi2F	5'CCNGARATHACNATHCCNAARCC 3'
Dswi2R	5'CCATDATRTCCATNACYTGNGTC 3'
Dswi3F	5'TTGGGMYTNWWHAAYTAYCA3'
Dswi3R	5'GTNSWNARNACNGGTTYTC3'
Disruption	
Cdswi2F	5'TGAAAAAGATCAGTTACAAAGGATCTCCGCAAAT GCCTCGTCCCCGCCGGGTC3'
Cdswi2R	5' ACAAGAAGCGTCTTCTTCTTTCAGATAGCTCT CGACAGCAGTATAGCGACCAGC3'
Cdswi3F	5'GCGAGCGCCTCATTGTCTTGCTCCAACATCTGCTT TTGCCTCGTCCCCGCCGGGTC3'
Cdswi3R	5'CCATAAGCGGACACGTTGGAACCCGGACCAAGG GAATCGACAGCAGTATAGCGACCAGC3'
Screening and confirmation	
Sswi2F	5'CTTCGACTCAGCCAATGAAAGCGG3'
Sswi2R	5'TCTGATAACAGAGTTTCTGTGTCC3'
Sswi3F	5'GCCATAAGCGGACACGTTGG3'
Sswi3R	5'GCGAGCGCCTCATTGTCTTG3'
MoxProF (-419/-403)	5'AACTGAGGGGGTTCTGC3'
MoxProR (-148/-165)	5'GCTGCACGGTAGCATAAG3'

5.6 Criação de banco de dados genético

Durante o doutoramento participamos do projeto de seqüênciamento da bactéria *Xylella fastidiosa*. Este organismo é um fitopatôgeno reponsável por diversas doenças em plantas economicamente importantes como: clorose variegada do citrus (amarelinho) em laranja, doença de pierce em videiras e escaldadura de folhas de café e ameixa [130]. No projeto de seqüênciamento genômico do fitopatôgeno *Xylella fastidiosa* (artigo em anexo) construímos um banco de dados genético com a finalidade de processamanto (análise de qualidade, montagem de "*contigs*" etc....), armazenamento e classificação das sequências obtidas. O banco de dados foi rearranjado diversas vezes de forma a disponiblizar ferramentas e organizar os dados de acordo com as sugestões dos pesquisadores envolvidos no projeto. Durante todo o andamento do projeto era possível avaliar o estatus de sequenciamento de cada cosmídeo e a montagem do genoma bem como efetuar "downloads" das sequencias de plasmídeos e cosmídeos.

Inicialmente o banco de dados foi desenvolvido utilizando-se o software MySQL, configurado adequadamente com tabelas relacionais que armazenam todos os dados do projeto. Este banco de dados caracteriza-se por ser de fácil manutenção e também ser de domínio público, de acordo com as normas descritas pela GNU/GPL. As seqüências geradas por cada grupo de sequenciamento foram submetidas pelos grupos de seqüênciamento para o LBI (laboratório de bioinformática) sediado na Unicamp. Foi desenvolvido um sistema de submissão eletrônica dos eletroferogramas (cromatogramas) gerados. Cada seqüência de DNA era extraída pelo phred/phrap/consed gerando os arquivos: cromatogramas: o próprio cromatograma, tal qual foi submetido; Gráficos de qualidade: uma figura identificando a qualidade daquele cromatograma base a base. Logo após a submissão um relatório era enviado ao remetente via e-mail contendo as informações: número de bases da seqüência, total de bases com vetor e a data de submissão.

A anotação das ORFs foi realizada de forma manual em um esforço conjunto dos grupos de pesquisa constituintes do ONSA (Organization for Nucleotide Sequence and Analysis). Atualmente o banco se encontra totalmente anotado sendo possível o acesso através do endereço http://aeg.lbi.ic.unicamp.br/xf, sendo possível realizar pesquisas por palavras chave, nome da ORF bem como obter informações de função e a tradução de cada ORF.

5.7 Seqüêncimento e criação de banco de dados genético de *Hansenula polymorpha*

Com o seqüênciamento total do genoma da levedura *S. cerevisiae* foram descobertos um grande número de ORFs (Open Reading Frames) até então totalmente desconhecidas da comunidade científica [131]. Atualmente graças à potentes ferramentas de bioinformática, bem

como experimentos de transcrição e expressão em larga escala já foi possível estabelecer a função de diversas destas ORFs [132-135]. Atualmente já foram identificados genes similares aos descritos primeiramente em *S. cerevisiae* em diversos eucariotos inferiores ou superiores, evidenciando uma notável conservação genética entre os organismos [136]. A levedura *H. polymorpha* é altamente relacionada com *S. cerevisiae*, ambos são Hemiascomicetos, o que sugeria que um grande número de ORFs, de funções já descritas, pudessem ser facilmente identificadas em *H polymorpha*. Ainda, o sistema de banco de dados e as ferramentas de bioinformática criadas no projeto genoma *X. fastidiosa* permitiria que este sistema fosse muito fácil de ser implementado.

Com base nas informações descritas acima foi realizado um estudo genômico em pequena escala para se identificar genes de *H. polymorpha* com o objetivo de fornecer novos elementos para futuros trabalhos desenvolvidos em nosso laboratório.

Dada às características dos genes de leveduras apresentarem em pequena quantidade genes que possuem introns ≈4% [131] a técnica escolhida foi a RST (Random Sequence Tags) a qual se baseia no seqüênciamento direto de biblioteca genômica de DNA (gentilmente cedida pela Dra. Dirce Carraro). Foram geradas 5.784 seqüências de aproximadamente 400pb, o que cobre cerca de 25% do genoma de *Hansenula polymorpha* (estimado em 9 Mb) [136].

5.7.1 Processamento dos dados de seqüênciamento

Os dados de seqüênciamento foram processados localmente em nosso laboratório de bioinformática (http://www.lge.ibi.unicamp.br). As seqüências foram primeiramente classificadas de acordo com a qualidade da seqüência, regiões de baixa qualidade ou seqüências de vetor ou *E. coli* foram excluídas e após este processamento as seqüências foram armazenadas no banco de dados genéticos (Fig.11).

Complementando o banco de dados estabelecido para *X. fastidiosa* implementamos novos mecanismos de "*data mining*" dentre eles, blastagem e categorização automática. No primeiro processo os resultados do BLAST são apresentados na forma de arquivos arquivos de texto daquele cromatograma contra o banco "*nr*" (um compêndio que acumula vários bancos de dados mundiais para anotação de dados genômicos que está disponível via internet em ftp://ftp.ncbi.nlm.nih.gov/blast/db), Este procedimento foi efetuado a cada vez que ocorria uma submissão e é refeito mensalmente sobre todas as submissões quando da atualização do banco de dados "*nr*". Os resultados de homologia foram utilizados para a categorização dos produtos de seqüênciamento de acordo com o banco de dados Go (http://www.godatabase.cgi-bin) (Fig.11).

Várias ferramantas de bioinformática como também de "*scripts*" para automação de ferramentas já existentes foram desenvolvidos pelos alunos: Fernando Tsukumo, Marcos Renato Rodrigues Araújo, Marcelo Marchi, Luciano Digiampietri e Marcelo Falsarella Carazzolle.

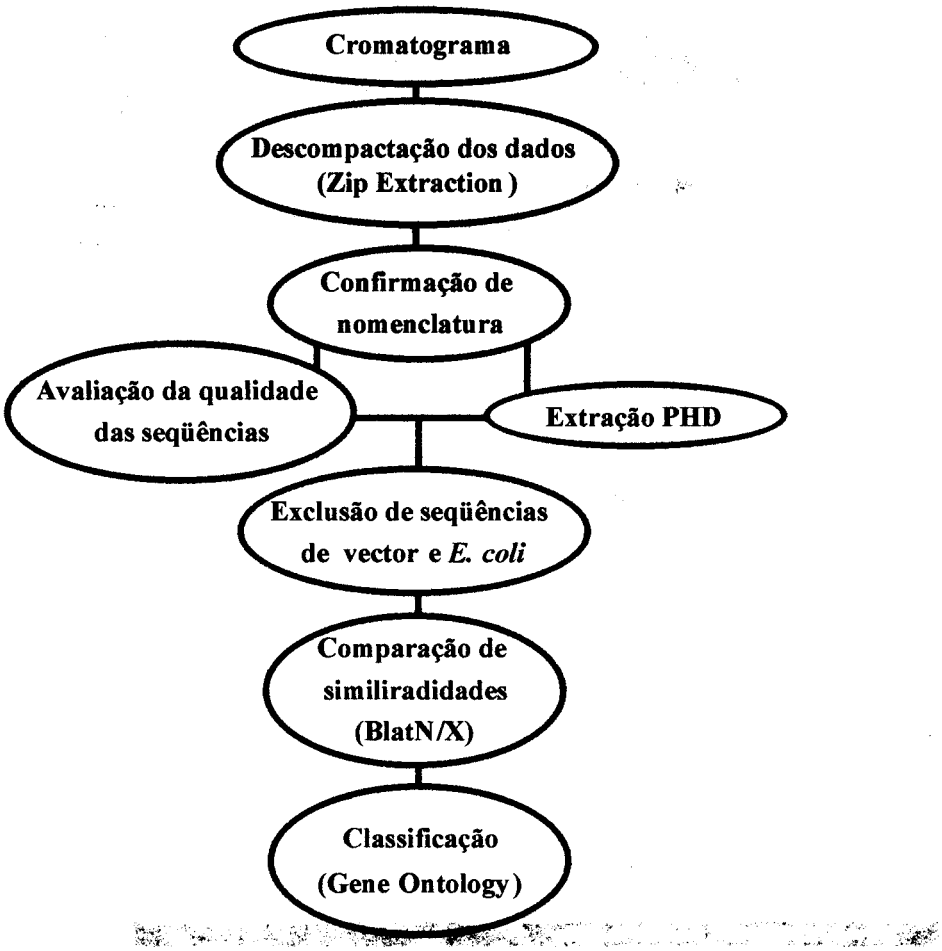


Figura 11: Fluxograma representando as etapas de geração e tratamantos de dados do banco de dados genéticos de *H. polymorpha*.

A página de *H. polymorpha* disponibiliza diversas ferramentas computacionais para a análise das seqüências bem como serviços de download dos cromatogramas, arquivos FASTA e resultados de busca de homologia (Blasts). Dentre as ferramentas computacionais podemos destacar: procura por palavras chave, busca de similaridades com seqüências de *H.polymorpha* contidos em nosso banco de dados por meio dos programas BlastN/X e busca de motivos conservados já descritos. A Figura 12 apresenta a página inicial da internet do projeto de seqüênciamento *H.polymorpha* e um exemplo da utilização da ferramenta de busca por palavras-chave.

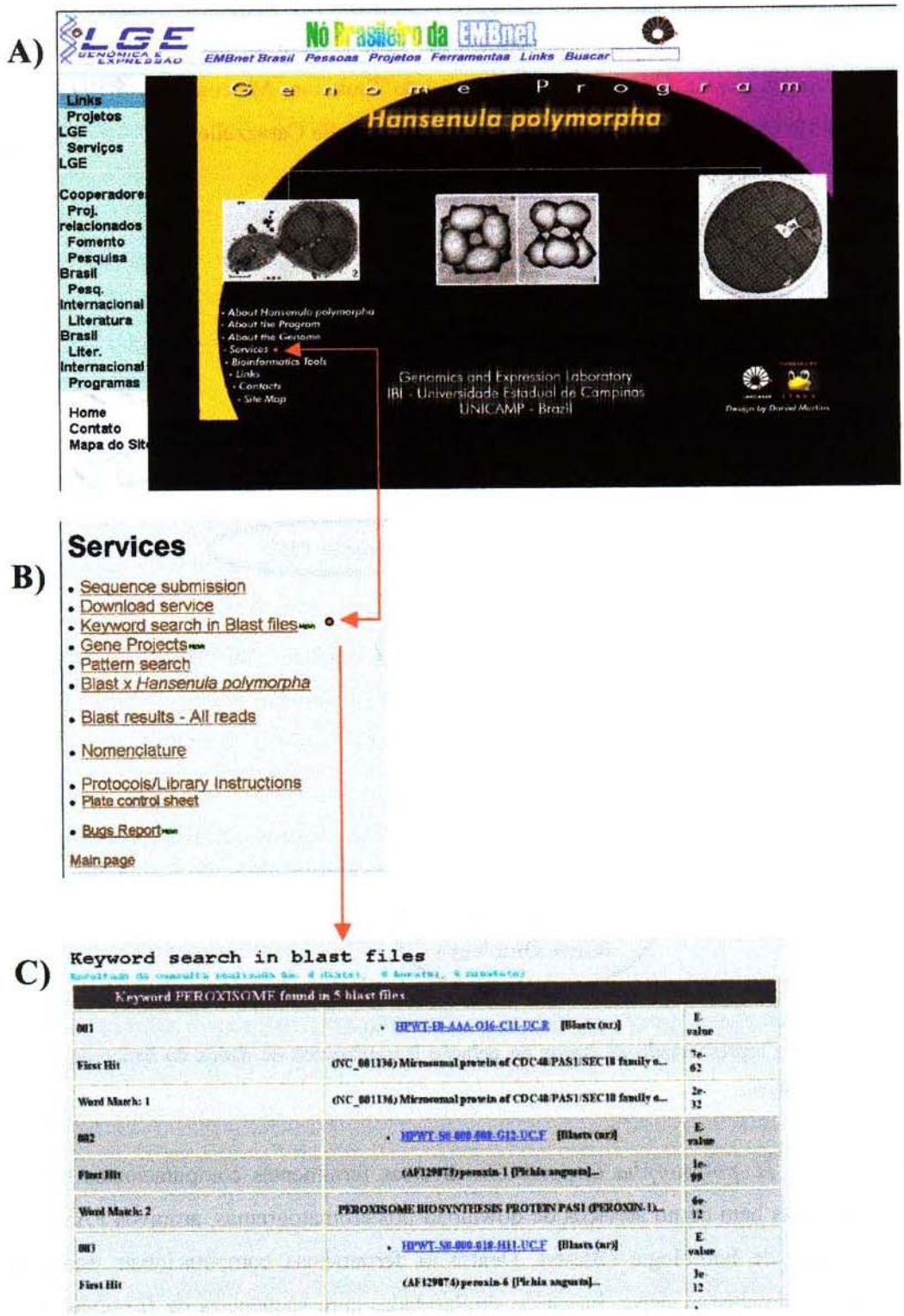


Figura 12: Acesso e funcionamento do banco de dados genéticos de *H. polymorpha*. A) Página inicial do banco de dados de *H. polymorpha* (http://www.lge.ibi.unicamp.br/hansenula). B) Ferramentas disponíveis para análise do banco de dados. C) Resultado de pesquisa usando a ferramenta de busca por palavras-chave (peroxisome).

5.7.2 Estimativa da freqüência dos codons em *H. polymorpha*

A grande maioria dos aminoácidos possui mais de um codon, entretanto, diversos microorganismos utilizam preferencialmente determinados codons para um aminoácido específico. Este tendência é denominada de Codon Bias (CB) e já foi demonstrado que proteínas altamente expressas possuem um valor de CB > 0,2 [154,155]. Como citado anteriormente, a levedura *H. polymorpha* é utilizada para a produção de proteínas de interesse comercial de diversos organismos [122,124] e o estudo da freqüência dos codons pode auxiliar na melhoria da expressão destas proteínas.

Para se calcular a freqüência dos codons (codon bias) em *H. polymorpha* foram escolhidos 150 reads, de forma aleatória, que apresentavam o resultado de Blast X um e value abaixo de –50 com seqüências protéicas presentes em genebank. As seqüências utilizadas possuíam um número de nucleotídeos entre 350 e 2500 pb e o número total de codons utilizados para o cálculo de freqüências foi de ≈ 25.000. Para os cálculos não foram computados os codons de terminação, uma vez que a maioria das seqüências representam segmentos gênicos e não genes inteiros. Os resultados são apresentados na tabela 3. Os cálculos de freqüência dos codons foram realizados com o auxílio dos programas Generunner e Codon W (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html).

Tabela 3: Freqüência dos codons em *H. polymorpha*.

Codon	Freqüência	N°	Codon	Freqüência	N°	Codon	Freqüência	N°
TTT	19.6	419	TCT	18.5	394	TAT	9.5	203
TTC	27.3	582	TCC	17.1	365	TAC	24.6	524
TTA	4.2	90	TCA	6.5	138	TAA	0.0	0
TTG	26.3	562	TCG	13.4	285	TAG	0.0	0
CTT	16.0	342	CCT	13.1	280	CAT	6.7	143
CTC	13.7	293	CCC	5.2	111	CAC	14.4	308
CTA	5.1	108	CCA	20.1	428	CAA	15.2	325
CTG	24.5	523	CCG	7.5	161	CAG	19.0	405
ATT	26.6	568	ACT	17.2	367	AAT	14.2	302
ATC	28.3	604	ACC	21.8	464	AAC	29.8	635
ATA	3.4	73	ACA	9.4	201	AAA	23.6	503
ATG	19.5	417	ACG	10.2	218	AAG	44.3	946
GTT	27.4	585	GCT	21.8	466	GAT	20.6	440
GTC	21.8	465	GCC	28.3	604	GAC	37.2	793
GTA	2.0	42	GCA	12.8	272	GAA	20.0	427
GTG	22.7	485	GCG	6.1	130	GAG	42.4	905
TGT	7.8	167	CGT	5.2	110	AGT	4.1	88
TGC	7.8	166	CGC	4.0	85	AGC	7.4	157
TGG	12.6	269	CGA	2.4	51	AGA	29.0	618
GGT	27.2	580	CGG	3.3	70	AGG	4.6	99
GGA	20.0	426	GGC	19.2	410	GGG	6.4	136

Utilizando os dados de freqüência foram estimados os CBIs (Codon Bias Index) de proteínas envolvidas no metabolismo de metanol. Os CBIs para *Mox* e *Cat* foram idênticos

(0,343), o gene que codifica a enzima dihidroxiacetona sintetase apresentou um CBI de 0,334. Os dados indicam uma alta utilização dos codons mais freqüentes e está de acordo com os dados presentes na literatura para outros organismos [154].

5.7.3 Clusterização de RSTs

As seqüências de DNA foram clusterizadas com o auxílio do programa "phrap" [156]. No processo de clusterização incorporamos ao nosso banco genético 5082 seqüências de *H. polymorpha* presentes no Gene Bank [137], totalizando 10.886 "*reads*". A clusterização resultou em 2876 contigs com fragmentos de DNA que apresentaram tamanhos variados. Dentre eles podemos destacar um cluster de 6678 pb contendo diversos genes mitocôndriais e outro de 8696 pb representa um cluster de 170 RSTs contendo as seqüências dos RNAs ribossomais 25s, 18S, 5.8S e 5S. O conteúdo GC nos "contigs" foi de 44,2%, bastante semelhante ao encontrado em leveduras deste gênero [137, 138]

5.7.4 Categorização das seqüências de DNA.

As seqüências já devidamente tratadas foram categorizadas usando como referências as classes definidas pelo gene ontology. Este banco de dados subdivide as seqüências primeiramente em três grandes classes: i) Processos biológicos ii) Componente celular e iii) Função molecular (Fig.13). Dentro classes as ORFs são ordenadas em 45 categorias, que por sua vez, são subdivididas sequencialmente totalizando 11.444 categorias.

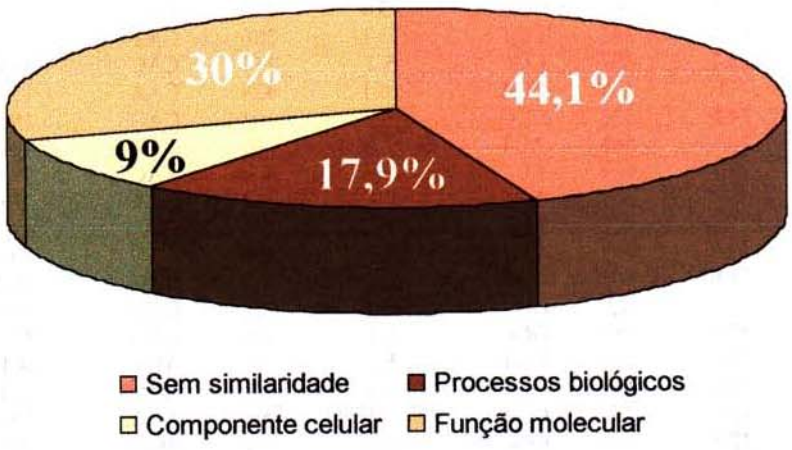


Figura 13: Categorização das seqüências de *H. polymorpha* presentes em nosso banco de dados genéticos. Os segmentos de DNA que não apresentam similaridade com ORFs já descritas representam 44,1% do banco de dados. As seqüências que se enquadram na categoria Função molecular representam 30% do banco de dados. Segmentos que apresentam similaridades com ORFs que respondem por elementos envolvidos em Processos biológicos e Componente celular representam 17,9 % e 9% respectivamente.

Diversas seqüências de DNA não apresentaram similaridade com nenhuma ORF categorizada no Godatabase. Em iniciativas semelhantes com leveduras, o número de seqüências que não apresentaram similaridade com ORFs já descritas alcançaram níveis semelhantes aos encontrados em nosso estudo [137-139]. A explicação para este fato pode ser dada por diversas causas não excludentes: existência de seqüências divergentes que podem fazer parte de regiões pouco conservadas existentes nas ORFs, seqüência de DNA de promotores, regiões inter ou intragênicas ou até mesmo serem genes específicos do gênero ou espécie.

Entretanto, grande parte das seqüências possui correspondentes com ORFs categorizadas dentre as quais se destacam as categorias de enzimas, proteínas ligantes e proteínas transportadoras. A Figura 14 apresenta as classes de ORFs que apresentaram maior representatividade. O padrão proporcional de distribuição categórica de ORFs se mostrou similar ao apresentado no geneontology para outros organismos (http://www.godatabase.cgi-bin).

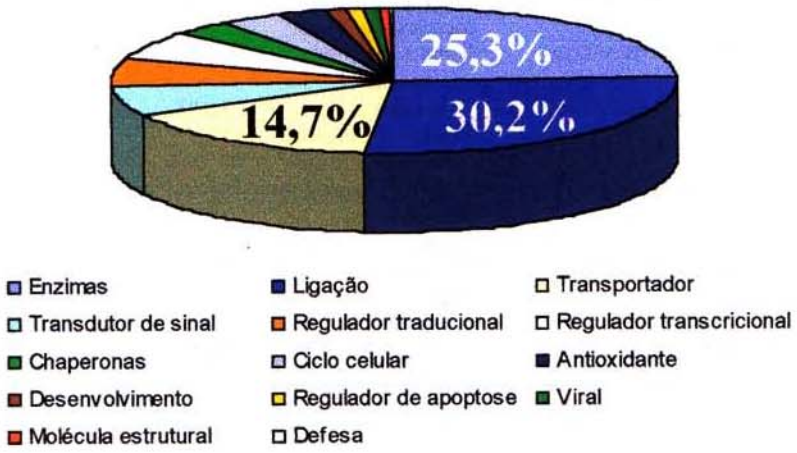


Figura 14: Gráfico de categorização de ORFs apresentando as classes mais representativas, que apresentaram similaridade com ORFs já identificadas em outros organismos, encontradas no banco de dados de *H. polymorpha*. Sendo que as classes enzimas (30,2%), proteínas ligantes (25,3%) e proteínas transportadoras (14,7%) representam mais de 70% das ORFs categorizadas.

Após estes procedimentos foi possível definir ORFs que apresentam similaridade com elementos importantes para o entendimento da regulação por glicose em *H. polymorpha* e biogênese de peroxissomos. Na Tabela 3 estão representadas 20 ORFs de grande importância investigativa para futuros trabalhos em *H.polymorpha*. Dentre estes elementos podemos destacar: um cofator de ativação de SNF1 (*SIP3*) [140]; componentes dos complexos remodeladores de cromatina SWI/SNF (*SWI1*) e RSC (*STH1*) [56]; um fator do complexo mediador (*SRB9*) o qual interage com o complexo Ssn6p-Tup1p [92]; reguladores negativos de transcrição (*RGR1* e *NGR2*) [65] e um ativador de transcrição do sistema Hap2/3/4/5 (*HAP2*). Além de ORFs com

similiridade com genes do sistema *GAL* (*Gal2 e Gal 6*) e diversas proteínas do metabolismo peroxisomal, ainda não descritas. Apesar da anotação automática ser um instrumento de grande valia para a organização dos dados é necessária uma acurada confirmação manual dos dados obtidos. É importante salientar que as anotações das ORFs apresentadas na Tabela 4 foram conferidas manualmente.

Tabela 4: Lista de ORFs de *H polymorpha* que são alvos primários para estudos a serem realizados em nosso laboratório. As ORFs comparadas que não apresentam referência são seguidas pelo código de identidade genética do genebank (http://www.ncbi.nlm.nih.gov).

ORF-*Hansenula polymorpha*	**ORF Similar/Organismo**	**Função**	**e-value**	**Frame**	**Referência**
HPWT-S0-000-005-C09-UC.F	SIP3/*S. cerevisiae*	Co-ativador Interage com Snf1p	7e-21	+3	[140]
HPWT-E0-SWT-001-B07-UC.F	STH1/*S. cerevisiae*	Helicase dependente de ATP	6e-44	+3	[56]
HPWT-S0-000-005-D12-UC.F	SRB9/*S. cerevisiae*	Componente do complexo mediador	1e-06	+2	[14]
HPWT-S0-000-051-H12-UC.F	RGR1/*S. cerevisiae*	Membro do complexo mediador	1e-10	-3	[13]
HPWT-S0-000-101-A02-UC.F	HAP1/*S. cerevisiae*	Ativador genes envolvidos no combate ao estresse oxidativo	3e-07	-3	[141]
HPWT-S0-000-160-G01-UC.F	HAP2/*K. lactis*	Ativador de transcrição de genes sujeitos a regulação por glicose	4e-19	+2	[65]
HPWT-E0-AAA-007-B07-UC.F	GCN4/*C. albicans*	Ativador de transcrição de genes envolvidos em biosíntese de A.A	2e-18	-3	[142]
HPWT-E0-AAA-009-D04-UC.R	NRG2/*S. cerevisiae*	Repressor transcricional catabólico	2e-16	+2	[143]
HPWT-S0-MIG-115-D02-UC.R	SWI1(ADR6)/*S. cerevisiae*	Membro do complexo SWI/SNF	2e-7	-1	[6]
HPWT-E0-AAA-033-B03-UC.F	GAL2/*S. cerevisiae*	Transportador de galactose	6e-18	+1	[144]
HPWT-E1-AAA-001-E02-UC.F	GAL6/*S. cerevisiae*	Bleomicina hidrolase	9e-25	+1	[144]
HPWT-E0-AAA-007-D01-UC.F	BRP1/*S. pombe*	Ativador global de transcrição	3e-12	-2	[145]
HPWT-E0-AAA-009-H12-UC.R	KPYK/Yarrowia lipolytica	Piruvato quinase	5e-75	+2	[146]
HPWT-E0-AAA-016-F06-UC.R	SOD/*C. albicans*	Remoção de peróxidos	7e-68	+1	[147]
HPWT-E0-AAA-013-H07-UC.F	CISY/Candida tropicalis	Citrato sintase	2e-71	-1	gi:2493724
HPWT-E0-AAA-009-A04-UC.F	CAT2/*S. cerevisiae*	Carnitina O-acetiltransferase	3e-23	+1	[65]
HPWT-E0-AAA-007-H09-UC	POX18 (FOX2)/*S. cerevisiae*	Oleato sintase peroxisomal	2e-7	+2	[149]
HPWT-E0-AAA-009-E03-UC.R	ADH2/Pichia stipitis	Alcool desidrogenase 2	7e-72	+2	[150]
HPWT-S0-000-052-B09-UC.F	ACS/*Zygosaccharomyces bailii*	Acetil-CoA sntetase	1e-32	-3	gi:14331097
HPWT-S0-000-114-B03-UC.F	GUT1/*S. cerevisiae*	Glicerol quinase	4e-29	+1	[151]

6. CONCLUSÕES E CONSIDERAÇÕES FINAIS

6.1 Identificação de fatores que influenciam na transcrição do gene *Mox*

Dada a proposta inicial do trabalho, a qual era identificar fatores envolvidos na regulação do gene *Mox*, é possível afirmar que ao menos dois fatores propostos neste trabalho influenciam em sua transcrição (*Snf1* e *Swi2*).

O fator *Tup1*, que possui extensa homologia com fatores semelhantes já caracterizados em outros organismos que respondem por eventos de repressão de transcrição de genes regulados por glicose atuando não somente neste processo, mas também na repressão de genes de resposta a choque térmico, "*mating type*", estresse oxidativo, floculação entre outros processos celulares. A disrupção deste gene, graças a sua participação em diversos processos celulares, ocasiona em outras leveduras efeitos pleiotrópicos [84, 93, 94]. Em *H. polymorpha* não foi detectada nenhuma mudança no fenótipo da levedura, nem tão pouco do padrão de expressão de genes da via de degradação do metanol.

As células *snf1* também apresentaram taxas de crescimento similares à linhagem selvagem em todas as fontes de carbono testadas (Fig.7 e Manuscrito1-Fig. 4A e B), entretanto o "*northern blotting*" revela uma sensível redução nos transcritos de Metanol Oxidase e Dihidroxiacetona sintetase (Manuscrito 1-Fig. 3). Estes resultados demonstram claramente que a disrupção afeta não somente a transcrição de *Mox* como também de *Das*. O que demonstra que estes produtos são transcritos em níveis superiores aos necessários para desempenhar seu papel biológico, uma vez que a redução da transcrição destes genes não reduziu as taxas de crescimento em metanol (Manuscrito1-Fig.4A e B). Recentemente, foi demonstrado "*in vitro*" que SNF1 responde pela fosforilação de Adr1p, conferindo a este fator uma maior capacidade de interagir com a cromatina e recrutar os elementos necessários para a formação do PIC [152].

Como apresentado anteriormente, a região promotora de *Mox* possui duas UAS (Figura.7) e o fator Adr1p de *S. cerevisiae* é capaz de se ligar à UAS2 e ativar a transcrição de *Mox*. Nesta mesma região foi possível detectar a ligação de um fator próprio de *H. polymorpha* (FB2), o qual deve ser um ativador transcricional de *Mox* [113]. Um dos objetivos iniciais foi a identificação do fator FB2, infelizmente não foi possível identificar com as técnicas empregadas neste trabalho. Deste modo, é possível que o fator FB2 exerça função majoritária na transcrição da via metabólica do metanol em *H. polymorpha*, assim como Adr1p na degradação de etanol em *S. cerevisiae*. Apesar de não termos identificado um fator similar à *Adr1*, é notável o decréscimo nos

níveis de transcrição de *Mox* e *Das* nas células *snf1* demonstrando o envolvimento do produto deste gene na via de degradação do metanol.

No entanto, não existe redução considerável nos níveis de transcritos *Cat.* É importante lembrar que o H_2O_2 é extremamente tóxica para os sistemas biológicos. Se existe uma expressão de Moxp, mesmo que reduzida como é o caso das células *snf1*, é fundamental a presença da catalase para permitir a viabilidade celular. É importante salientar que o peróxido de hidrogênio não é produzido somente pela degradação de metanol e é coerente supor que o sistema da ativação de transcrição deste gene não dependa somente de um ativador, mas sim, de vários fatores que devem agir em concerto regulando os níveis de sua transcrição.

Com relação aos mutantes *swi2* os efeitos da disrupção são notáveis e estão em acordo com os apresentados pelos mutantes *swi3*. Os mutantes apresentam um crescimento muito lento e tendem a flocular. Nos ensaios de crescimento em meio líquido ou sólido as taxas de crescimento dos mutantes nos meios contendo glicerol ou metanol como fonte de carbono é acentuadamente lento (Figuras 8 e 9).

Fato intrigante é a não detecção dos transcritos de *Mox* nas linhagens *swi2* e *swi3* nas condições analisadas. As explicações para estes resultados podem ser fornecidas por dois mecanismos: i) A existência de via alternativa para a degradação do metanol; ii) As taxas de transcrição são ínfimas, não permitindo a detecção pela técnica empregada. Dentro destes dois cenários a primeira hipótese parece pouco provável, uma vez que nenhuma via alternativa foi descrita em leveduras metilotróficas. A favor do segundo mecanismo proposto, temos as observações dos mutantes *snf1*. Conforme discutido anteriormente, apesar de haver uma sensível redução dos transcritos de *Mox*, os mutantes *snf1* apresentam um crescimento semelhante ao selvagem, o que sugere uma transcrição deste gene muito acima da necessária. Realmente as taxas de expressão de Moxp em células crescendo em metanol por longos períodos de tempo levam a uma elevada biogênese de peroxissomos e grande acúmulo de metanol oxidase nos peroxissomos, levando a formação de estruturas cristalinas de Moxp [112], o que indica uma produção excessiva desta proteína. Ainda, a meia vida desta proteína é de aproximadamente 20hs (http://www.expasy.org/tools/protparam.html), portanto é razoável propor que apesar da metodologia utilizada não permitir a detecção de *Mox*, existe um nível de transcrição mínimo não detectável, mas que o acúmulo da proteína (devido a grande meia vida) permite o crescimento desta levedura em metanol.

Em conjunto, esses resultados apontam que a transcrição de *Mox* é independente do fator Tup1p. Entretanto, aponta que a transcrição deste gene (como também do gene *Das*) é

parcialmente influenciada pelo complexo SNF1 e altamente dependente do complexo SWI/SNF, o que caracteriza estes complexos como os primeiros ativadores de transcrição de *Mox*.

6.2 Restauração do metabolismo de galactose

Fenômeno inesperado, mas bastante importante, é a habilidade exibida pelos mutantes *swi2* e *swi3* quando a fonte de carbono é galactose e com um pouco menos de intensidade em rafinose. O crescimento acentuado dos mutantes em meio contendo rafinose (Fig.9) também pode ser devido a este alta capacidade de degradar galactose, tendo em vista que a rafinose é um polissacarídeo composto por uma molécula de glicose, uma de frutose e uma de galactose e é coerente sugerir que este efeito se deve ao melhor aproveitamento desta fonte de carbono pelos mutantes.

Em *S.cerevisiae* para metabolizar a galactose é necessário o produto de pelo menos oito genes, cada um deles com funções específicas para o metabolismo [106]. No banco de dados da biblioteca genética de *H. polymorpha*, foi possível identificar ORFs que apresentam grande similaridade com o gene *Gal2*, porém, este dado não auxiliou muito o entendimento do fenômeno uma vez que o transportador Gal2p pode ser substituído por outras hexoses [65]. Entretanto, em projeto de seqüênciamento genômico de *H. polymorpha* realizado por iniciativa privada (ainda não concluído) foi possível identificar ORFs similares aos genes necessários para o metabolismo de galactose, com exceção de *Gal1* (Hollemberg, C.P., comunicação pessoal). No entanto, a existência ou não deste gene pode não influenciar no metabolismo de galactose em *H. polymorpha*. Já foi demonstrado que a levedura *K. lactis* é capaz de degradar eficientemente a galactose apesar de não possuir o gene *Gal1*. A explicação para este fato é dada pela bifuncionalidade do produto do gene *Gal3*, o qual possui atividade de quinase sendo capaz de fosforilar a molécula de galactose, passo essencial para a degradação deste carboidrato [137].

Portanto, os resultados apontam que *H. polymorpha* apresenta em seu genoma genes que codificam todas as proteínas essenciais para a degradação de galactose, mas que não são expressas ou são expressas em quantidades ínfimas que não sustentam sua viabilidade por um longo período de tempo (Manuscrito 2- Fig.3).

Uma suposição pertinente, neste caso, é que a via de degradação da galactose estava silenciada devido à especialização da espécie *H. polymorpha* para sobreviver em um meio hostil, como o ambiente contendo metanol, e graças a esta especialização, mutações em regiões promotoras de genes *GAL* não ocasionariam pressão seletiva no ambiente ao qual esta se especializou. Curiosamente, o resultado de busca em banco de dados de culturas (Centraal bureau voor Schimmelcultures-http://www.cbs.knaw.nl/) aponta que os gêneros *Pichia* e *Hansenula*

(*Hansenula polymorpha* = *Pichia angusta*) possuem um reduzido número de espécies capazes de metabolizar galactose. Entretanto, nossos dados não permitem nenhuma inferência evolutiva.

É fato que em *H. polymorpha* este quadro é revertido com a disrupção dos fatores *Swi2* ou *Swi3*, o que indica que a participação do complexo SWI/SNF neste processo. Adicionalmente, estudos de linhagens de *S. cerevisiae* que foram classificadas como Gal$^-$, mostraram que estas possuem defeitos na expressão de pelo menos um dos genes do sistema *GAL* (comunicação pessoal do Dr. Dennis Lohr).

Já foi demonstrado em *S.cerevisiae* que SWI/SNF é capaz de atuar de forma negativa na transcrição de diversos genes [58]. No caso de *H. polymorpha* é pertinente supor que o complexo SWI/SNF estaria atuando de forma repressora sobre os genes *GAL* posicionando os nucleossomos em regiões essenciais para ativação de transcrição. Exatamente este mesmo mecanismo tem sido proposto para explicar a repressão exercida pelo complexo remodelador RSC sobre o promotor *CHA1* [153].

Uma outra possibilidade, levando em conta que o complexo SWI/SNF é um regulador geral de transcrição, é que as disrupções levaram a uma desorganização global da cromatina, o que favoreceria a transcrição de genes que estivessem sob controle negativo do complexo SWI/SNF. Ambas as propostas são pertinentes, entretanto a segunda explicaria melhor as características apresentadas pelos mutantes. O crescimento em todas as fontes de carbono testadas foi lento, ainda, em meio líquido é possível observar que as células entram em fase log após 13hs de crescimento, independente se a fonte de carbono utilizada for glicose, glicerol ou metanol (Fig.9). Este fato se repete mesmo quando a fonte de carbono é galactose (Manuscrito 2-Fig.3.B). O que indica um "atraso" na expressão em todas as fontes de carbono testadas, o que sugere um mecanismo mais amplo, como o proposto na segunda hipótese.

Independe das hipóteses de como o complexo SWI/SNF estaria atuando na repressão dos genes *GAL*, este trabalho demonstrou a restauração de uma via metabólica quiescente através da inativação de um complexo remodelador de cromatina e abre perspectivas para estudos fundamentais, ainda não abordados, do silenciamento de vias metabólicas.

6.3 Banco de dados genético de *Hansenula polymorpha*

A iniciativa da construção do banco de dados de *H. polymorpha* permitiu a geração de um grande volume de dados que deve auxiliar a compreender melhor a biologia desta levedura. Alguns elementos já foram apontados como candidatos para estudos mais aprofundados no metabolismo de metanol (Tab.4). Entretanto, os dados armazenados necessitam de uma anotação

mais rigorosa de forma a permitir um melhor aproveitamento dos dados, fornecendo novos alvos não só para o entendimento do controle da expressão de proteínas envolvidas no metabolismo do metanol como também para outros processos biológicos.

Estas informações também podem ser utilizadas para estudos taxonômicos e evolutivos desta levedura. Como descrito anteriormente existe grande controvérsia quanto à classificação das leveduras pertencentes aos gêneros *Pichia* e *Hansenula*, portanto seria interessante averiguar os genes ortólogos. Finalmente, os dados de "Codon Bias" podem auxiliar em estudos mais aplicados e permitir um melhor rendimento na expressão de proteínas heterólogas utilizando *H.polymorpha* como hospedeiro [154, 155].

7. PERSPECTIVAS

Ensaios da atividade de Moxp nos mutantes *swi2* e *swi3*.

Análise do posicionamento de nucleossomos na região promotora de *Mox* e *Das* nos mutantes *swi2 swi3* e *snf1*.

Investigação da expressão de genes necessários para a degradação de galactose nas linhagens *swi2* e *swi3* através de "northern blotting" e RT PCR.

Análise do posicionamento de nucleossomos na região promotora de genes envolvidos no metabolismo de galactose nos mutantes *swi2* e *swi3*.

Anotação detalhada das seqüências de DNA do banco de dados genético de *H. polymorpha* para uma classificação mais acurada das seqüências de DNA e verificação de introns e genes ortólogos.

Utilização dos clones na elaboração de microarrays de DNA para a investigação de transcrição visando a proliferação de peroxissomos e outros processos celulares em *H. polymorpha*.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Roeder, R.G., *The role of general initiation factors in transcription by RNA polymerase II.* Trends Biochem Sci, 1996. **21**(9): p. 327-35.
2. Verrijzer, C.P. and R. Tjian, *TAFs mediate transcriptional activation and promoter selectivity.* Trends Biochem Sci, 1996. **21**(9): p. 338-42.
3. Albright, S.R. and R. Tjian, *TAFs revisited: more data reveal new twists and confirm old ideas.* Gene, 2000. **242**(1-2): p. 1-13.
4. Struhl, K., *Yeast transcriptional regulatory mechanisms.* Annu Rev Genet, 1995. **29**: p. 651-74.
5. Bjorklund, S. and Y.J. Kim, *Mediator of transcriptional regulation.* Trends Biochem Sci, 1996. **21**(9): p. 335-7.
6. Lee, T.I. and R.A. Young, *Transcription of eukaryotic protein-coding genes.* Annu Rev Genet, 2000. **34**: p. 77-137.
7. Cramer, P., et al., *Architecture of RNA polymerase II and implications for the transcription mechanism.* Science, 2000. **288**(5466): p. 640-9.
8. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution.* Science, 2001. **292**(5523): p. 1863-76.
9. Choder, M. and R.A. Young, *A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival.* Mol Cell Biol, 1993. **13**(11): p. 6984-91.
10. Edwards, A.M., et al., *Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro.* J Biol Chem, 1991. **266**(1): p. 71-5.
11. Dahmus, M.E., *Reversible phosphorylation of the C-terminal domain of RNA polymerase II.* J Biol Chem, 1996. **271**(32): p. 19009-12.
12. Greenblatt, J., *RNA polymerase II holoenzyme and transcriptional regulation.* Curr Opin Cell Biol, 1997. **9**(3): p. 310-9.
13. Lee, M., S. Chatterjee, and K. Struhl, *Genetic analysis of the role of Pol II holoenzyme components in repression by the Cyc8-Tup1 corepressor in yeast.* Genetics, 2000. **155**(4): p. 1535-42.
14. Myers, L.C. and R.D. Kornberg, *Mediator of transcriptional regulation.* Annu Rev Biochem, 2000. **69**: p. 729-49.
15. Lewin, B., *Genes V.* 6th ed. 1994, Oxford, USA.: Oxford University Press.
16. Han, M. and M. Grunstein, *Nucleosome loss activates yeast downstream promoters in vivo.* Cell, 1988. **55**(6): p. 1137-45.
17. Kim, U.J., et al., *Effects of histone H4 depletion on the cell cycle and transcription of Saccharomyces cerevisiae.* Embo J, 1988. **7**(7): p. 2211-9.
18. Kornberg, R.D. and J.O. Thomas, *Chromatin structure; oligomers of the histones.* Science, 1974. **184**(139): p. 865-8.

19. Paranjape, S.M., R.T. Kamakaka, and J.T. Kadonaga, *Role of chromatin structure in the regulation of transcription by RNA polymerase II.* Annu Rev Biochem, 1994. **63**: p. 265-97.
20. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
21. Brownell, J.E. and C.D. Allis, *Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation.* Curr Opin Genet Dev, 1996. **6**(2): p. 176-84.
22. Davie, J.K., R.J. Trumbly, and S.Y. Dent, *Histone-dependent association of Tup1-Ssn6 with repressed genes in vivo.* Mol Cell Biol, 2002. **22**(3): p. 693-703.
23. Hagmann, M., *How chromatin changes its shape.* Science, 1999. **285**(5431): p. 1200-1, 1203.
24. Morales, V., et al., *Chromatin structure and dynamics: functional implications.* Biochimie, 2001. **83**(11-12): p. 1029-39.
25. Cheung, P., C.D. Allis, and P. Sassone-Corsi, *Signaling to chromatin through histone modifications.* Cell, 2000. **103**(2): p. 263-71.
26. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
27. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.
28. Cairns, B.R., et al., *RSC, an essential, abundant chromatin-remodeling complex.* Cell, 1996. **87**(7): p. 1249-60.
29. Crosby, M.A., et al., *The trithorax group gene moira encodes a brahma-associated putative chromatin-remodeling factor in Drosophila melanogaster.* Mol Cell Biol, 1999. **19**(2): p. 1159-70.
30. Imbalzano, A.N., *SWI/SNF complexes and facilitation of TATA binding protein:nucleosome interactions.* Methods, 1998. **15**(4): p. 303-14.
31. Kingston, R.E. and G.J. Narlikar, *ATP-dependent remodeling and acetylation as regulators of chromatin fluidity.* Genes Dev, 1999. **13**(18): p. 2339-52.
32. Peterson, C.L. and J.L. Workman, *Promoter targeting and chromatin remodeling by the SWI/SNF complex.* Curr Opin Genet Dev, 2000. **10**(2): p. 187-92.
33. Aagaard, L., et al., *Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31.* Embo J, 1999. **18**(7): p. 1923-38.
34. Paro, R., *Imprinting a determined state into the chromatin of Drosophila.* Trends Genet, 1990. **6**(12): p. 416-21.
35. Peters, A.H., et al., *Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability.* Cell, 2001. **107**(3): p. 323-37.
36. Allis, C.D., et al., *Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in Tetrahymena micronuclei.* Cell, 1980. **20**(1): p. 55-64.
37. Baarends, W.M., et al., *Histone ubiquitination and chromatin remodeling in mouse spermatogenesis.* Dev Biol, 1999. **207**(2): p. 322-33.

38. Mahadevan, L.C., A.C. Willis, and M.J. Barratt, *Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors.* Cell, 1991. **65**(5): p. 775-83.
39. Thomson, S., et al., *The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase.* Embo J, 1999. **18**(17): p. 4779-93.
40. Hsu, J.Y., et al., *Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes.* Cell, 2000. **102**(3): p. 279-91.
41. Rogakou, E.P., et al., *Megabase chromatin domains involved in DNA double-strand breaks in vivo.* J Cell Biol, 1999. **146**(5): p. 905-16.
42. Rogakou, E.P., et al., *Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139.* J Biol Chem, 2000. **275**(13): p. 9390-5.
43. Ajiro, K., *Histone H2B phosphorylation in mammalian apoptotic cells. An association with DNA fragmentation.* J Biol Chem, 2000. **275**(1): p. 439-43.
44. Enomoto, R., et al., *Phosphorylation of histones triggers DNA fragmentation in thymocyte undergoing apoptosis induced by protein phosphatase inhibitors.* Mol Cell Biol Res Commun, 2001. **4**(5): p. 276-81.
45. Kuo, M.H. and C.D. Allis, *Roles of histone acetyltransferases and deacetylases in gene regulation.* Bioessays, 1998. **20**(8): p. 615-26.
46. Grunstein, M., *Histone acetylation in chromatin structure and transcription.* Nature, 1997. **389**(6649): p. 349-52.
47. Georgakopoulos, T. and G. Thireos, *Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription.* Embo J, 1992. **11**(11): p. 4145-52.
48. Syntichaki, P., I. Topalidou, and G. Thireos, *The Gcn5 bromodomain co-ordinates nucleosome remodelling.* Nature, 2000. **404**(6776): p. 414-7.
49. Dyda, F., D.C. Klein, and A.B. Hickman, *GCN5-related N-acetyltransferases: a structural overview.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 81-103.
50. Rundlett, S.E., et al., *HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14503-8.
51. Ahringer, J., *NuRD and SIN3 histone deacetylase complexes in development.* Trends Genet, 2000. **16**(8): p. 351-6.
52. Imai, S., et al., *Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase.* Nature, 2000. **403**(6771): p. 795-800.
53. Berger, S.L., *Histone modifications in transcriptional regulation.* Curr Opin Genet Dev, 2002. **12**(2): p. 142-8.
54. Whitehouse, I., et al., *Nucleosome mobilization catalysed by the yeast SWI/SNF complex.* Nature, 1999. **400**(6746): p. 784-7.

55. Dingwall, A.K., et al., *The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex.* Mol Biol Cell, 1995. **6**(7): p. 777-91.
56. Du, J., et al., *Sth1p, a Saccharomyces cerevisiae Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins.* Genetics, 1998. **150**(3): p. 987-1005.
57. Shen, X., et al., *A chromatin remodelling complex involved in transcription and DNA processing.* Nature, 2000. **406**(6795): p. 541-4.
58. Holstege, F.C., et al., *Dissecting the regulatory circuitry of a eukaryotic genome.* Cell, 1998. **95**(5): p. 717-28.
59. Ryan, M.P., R. Jones, and R.H. Morse, *SWI-SNF complex participation in transcriptional activation at a step subsequent to activator binding.* Mol Cell Biol, 1998. **18**(4): p. 1774-82.
60. Hassan, A.H., et al., *Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes.* Cell, 2002. **111**(3): p. 369-79.
61. Pollard, K.J. and C.L. Peterson, *Role for ADA/GCN5 products in antagonizing chromatin-mediated transcriptional repression.* Mol Cell Biol, 1997. **17**(11): p. 6212-22.
62. Pollard, K.J., et al., *Functional interaction between GCN5 and polyamines: a new role for core histone acetylation.* Embo J, 1999. **18**(20): p. 5622-33.
63. Wilson, C.J., et al., *RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling.* Cell, 1996. **84**(2): p. 235-44.
64. Agalioti, T., G. Chen, and D. Thanos, *Deciphering the transcriptional histone acetylation code for a human gene.* Cell, 2002. **111**(3): p. 381-92.
65. Gancedo, J.M., *Yeast carbon catabolite repression.* Microbiol Mol Biol Rev, 1998. **62**(2): p. 334-61.
66. Tanaka, N., et al., *Isolation and characterization of an invertase and its repressor genes from Schizosaccharomyces pombe.* Biochem Biophys Res Commun, 1998. **245**(1): p. 246-53.
67. Halford, N.G. and D.G. Hardie, *SNF1-related protein kinases: global regulators of carbon metabolism in plants?* Plant Mol Biol, 1998. **37**(5): p. 735-48.
68. Carling, D., et al., *Mammalian AMP-activated protein kinase is homologous to yeast and plant protein kinases involved in the regulation of carbon metabolism.* J Biol Chem, 1994. **269**(15): p. 11442-8.
69. Hardie, D.G., D. Carling, and M. Carlson, *The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell?* Annu Rev Biochem, 1998. **67**: p. 821-55.
70. Wu, J. and R.J. Trumbly, *Multiple regulatory proteins mediate repression and activation by interaction with the yeast Mig1 binding site.* Yeast, 1998. **14**(11): p. 985-1000.

71. Nehlin, J.O. and H. Ronne, *Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins.* Embo J, 1990. **9**(9): p. 2891-8.
72. Treitel, M.A. and M. Carlson, *Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein.* Proc Natl Acad Sci U S A, 1995. **92**(8): p. 3132-6.
73. Jiang, R. and M. Carlson, *The Snf1 protein kinase and its activating subunit, Snf4, interact with distinct domains of the Sip1/Sip2/Gal83 component in the kinase complex.* Mol Cell Biol, 1997. **17**(4): p. 2099-106.
74. Ostling, J. and H. Ronne, *Negative control of the Mig1p repressor by Snf1p-dependent phosphorylation in the absence of glucose.* Eur J Biochem, 1998. **252**(1): p. 162-8.
75. Tamkun, J.W., et al., *brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2.* Cell, 1992. **68**(3): p. 561-72.
76. Randez-Gil, F., et al., *Glucose derepression of gluconeogenic enzymes in Saccharomyces cerevisiae correlates with phosphorylation of the gene activator Cat8p.* Mol Cell Biol, 1997. **17**(5): p. 2502-10.
77. Vincent, O., et al., *Interaction of the Srb10 kinase with Sip4, a transcriptional activator of gluconeogenic genes in Saccharomyces cerevisiae.* Mol Cell Biol, 2001. **21**(17): p. 5790-6.
78. Liu, Y., et al., *Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes.* J Biol Chem, 2001. **276**(10): p. 7169-75.
79. Kuchin, S., I. Treich, and M. Carlson, *A regulatory shortcut between the Snf1 protein kinase and RNA polymerase II holoenzyme.* Proc Natl Acad Sci U S A, 2000. **97**(14): p. 7916-20.
80. Lo, W.S., et al., *Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription.* Science, 2001. **293**(5532): p. 1142-6.
81. Smith, F.C., et al., *The SNF1 kinase complex from Saccharomyces cerevisiae phosphorylates the transcriptional repressor protein Mig1p in vitro at four sites within or near regulatory domain 1.* FEBS Lett, 1999. **453**(1-2): p. 219-23.
82. Carlson, M., *Glucose repression in yeast.* Curr Opin Microbiol, 1999. **2**(2): p. 202-7.
83. Schuller, H.J. and K.D. Entian, *Extragenic suppressors of yeast glucose derepression mutants leading to constitutive synthesis of several glucose-repressible enzymes.* J Bacteriol, 1991. **173**(6): p. 2045-52.
84. Vallier, L.G. and M. Carlson, *Synergistic release from glucose repression by mig1 and ssn mutations in Saccharomyces cerevisiae.* Genetics, 1994. **137**(1): p. 49-54.
85. Nehlin, J.O., M. Carlberg, and H. Ronne, *Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response.* Embo J, 1991. **10**(11): p. 3373-7.
86. Hu, Z., et al., *MIG1-dependent and MIG1-independent glucose regulation of MAL gene expression in Saccharomyces cerevisiae.* Curr Genet, 1995. **28**(3): p. 258-66.

87. Ostling, J., M. Carlberg, and H. Ronne, *Functional domains in the Mig1 repressor.* Mol Cell Biol, 1996. **16**(3): p. 753-61.
88. DeVit, M.J. and M. Johnston, *The nuclear exportin Msn5 is required for nuclear export of the Mig1 glucose repressor of Saccharomyces cerevisiae.* Curr Biol, 1999. **9**(21): p. 1231-41.
89. Cassart, J.P., et al., *Comparative analysis in three fungi reveals structurally and functionally conserved regions in the Mig1 repressor.* Mol Gen Genet, 1997. **255**(1): p. 9-18.
90. Zaragoza, O., C. Rodriguez, and C. Gancedo, *Isolation of the MIG1 gene from Candida albicans and effects of its disruption on catabolite repression.* J Bacteriol, 2000. **182**(2): p. 320-6.
91. Trumbly, R.J., *Cloning and characterization of the CYC8 gene mediating glucose repression in yeast.* Gene, 1988. **73**(1): p. 97-111.
92. Williams, F.E. and R.J. Trumbly, *Characterization of TUP1, a mediator of glucose repression in Saccharomyces cerevisiae.* Mol Cell Biol, 1990. **10**(12): p. 6500-11.
93. Zitomer, R.S., et al., *Approaches to the study of Rox1 repression of the hypoxic genes in the yeast Saccharomyces cerevisiae.* Methods, 1997. **11**(3): p. 279-88.
94. Ozcan, S. and M. Johnston, *Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose.* Mol Cell Biol, 1995. **15**(3): p. 1564-72.
95. Fujiwara, D., et al., *Molecular mechanism of the multiple regulation of the Saccharomyces cerevisiae ATF1 gene encoding alcohol acetyltransferase.* Yeast, 1999. **15**(12): p. 1183-97.
96. Varanasi, U.S., et al., *The Cyc8 (Ssn6)-Tup1 corepressor complex is composed of one Cyc8 and four Tup1 subunits.* Mol Cell Biol, 1996. **16**(12): p. 6707-14.
97. Tzamarias, D. and K. Struhl, *Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8-Tup1 corepressor complex to differentially regulated promoters.* Genes Dev, 1995. **9**(7): p. 821-31.
98. Edmondson, D.G., M.M. Smith, and S.Y. Roth, *Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4.* Genes Dev, 1996. **10**(10): p. 1247-59.
99. Smith, T.F., et al., *The WD repeat: a common architecture for diverse functions.* Trends Biochem Sci, 1999. **24**(5): p. 181-5.
100. Lamb, J.R., S. Tugendreich, and P. Hieter, *Tetratrico peptide repeat interactions: to TPR or not to TPR?* Trends Biochem Sci, 1995. **20**(7): p. 257-9.
101. Das, A.K., P.W. Cohen, and D. Barford, *The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions.* Embo J, 1998. **17**(5): p. 1192-9.
102. Gounalaki, N., D. Tzamarias, and M. Vlassi, *Identification of residues in the TPR domain of Ssn6 responsible for interaction with the Tup1 protein.* FEBS Lett, 2000. **473**(1): p. 37-41.

103. Braun, B.R. and A.D. Johnson, *Control of filament formation in Candida albicans by the transcriptional repressor TUP1.* Science, 1997. **277**(5322): p. 105-9.
104. Griggs, D.W. and M. Johnston, *Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression.* Proc Natl Acad Sci U S A, 1991. **88**(19): p. 8597-601.
105. Smith, R.L. and A.D. Johnson, *Turning genes off by Ssn6-Tup1: a conserved system of transcriptional repression in eukaryotes.* Trends Biochem Sci, 2000. **25**(7): p. 325-30.
106. Cornish-Bowden, A. and M.L. Cardenas, *Complex networks of interactions connect genes to phenotypes.* Trends Biochem Sci, 2001. **26**(8): p. 463-5.
107. Johnston, M., J.S. Flick, and T. Pexton, *Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae.* Mol Cell Biol, 1994. **14**(6): p. 3834-41.
108. Rohde, J.R., J. Trinh, and I. Sadowski, *Multiple signals regulate GAL transcription in yeast.* Mol Cell Biol, 2000. **20**(11): p. 3880-6.
109. Sadowski, I., C. Costa, and R. Dhanawansa, *Phosphorylation of Gal4p at a single C-terminal residue is necessary for galactose-inducible transcription.* Mol Cell Biol, 1996. **16**(9): p. 4879-87.
110. van der Klei, I.J., W. Harder, and M. Veenhuis, *Biosynthesis and assembly of alcohol oxidase, a peroxisomal matrix protein in methylotrophic yeasts: a review.* Yeast, 1991. **7**(3): p. 195-209.
111. Pereira, G.G., *Analysis of the transcriptional regulation of the MOX gene encoding peroxisomal methanol oxidase from Hansenula polymorpha.* 1994, Universität Düsseldorf, German.
112. Veenhuis, M., et al., *Substructure of crystalline peroxisomes in methanol-grown Hansenula polymorpha: evidence for an in vivo crystal of alcohol oxidase.* Mol Cell Biol, 1981. **1**(10): p. 949-57.
113. Pereira, G.G. and C.P. Hollenberg, *Conserved regulation of the Hansenula polymorpha MOX promoter in Saccharomyces cerevisiae reveals insights in the transcriptional activation by Adr1p.* Eur J Biochem, 1996. **238**(1): p. 181-91.
114. Godecke, S., et al., *Identification of sequences responsible for transcriptional regulation of the strongly expressed methanol oxidase-encoding gene in Hansenula polymorpha.* Gene, 1994. **139**(1): p. 35-42.
115. Costanzo, G., et al., *Multiple overlapping positions of nucleosomes with single in vivo rotational setting in the Hansenula polymorpha RNA polymerase II MOX promoter.* J Biol Chem, 1995. **270**(19): p. 11091-7.
116. Genu, V., C. Hollenberg, and G.G. Pereira, *Induction of Hansenula polymorpha MOX gene is dependent on respiration and involves a switch between 2 alternative transcription starting points.* Submitted to Eur J Biochem, 2002.
117. Cheng, C., et al., *Identification of potential target genes for Adr1p through characterization of essential nucleotides in UAS1.* Mol Cell Biol, 1994. **14**(6): p. 3842-52.

118. Denis, C.L., M. Ciriacy, and E.T. Young, *A positive regulatory gene is required for accumulation of the functional messenger RNA for the glucose-repressible alcohol dehydrogenase from Saccharomyces cerevisiae.* J Mol Biol, 1981. **148**(4): p. 355-68.
119. Simon, M., et al., *The Saccharomyces cerevisiae ADR1 gene is a positive regulator of transcription of genes encoding peroxisomal proteins.* Mol Cell Biol, 1991. **11**(2): p. 699-704.
120. Komarnitsky, P.B., et al., *ADR1-mediated transcriptional activation requires the presence of an intact TFIID complex.* Mol Cell Biol, 1998. **18**(10): p. 5861-7.
121. Simon, M.M., et al., *A C-terminal region of the Saccharomyces cerevisiae transcription factor ADR1 plays an important role in the regulation of peroxisome proliferation by fatty acids.* Mol Gen Genet, 1995. **249**(3): p. 289-96.
122. Hollenberg, C.P. and G. Gellissen, *Production of recombinant proteins by methylotrophic yeasts.* Curr Opin Biotechnol, 1997. **8**(5): p. 554-60.
123. Sakai, Y., Y. Tani, and N. Kato, *Biotechnological application of cellular functions of the methylotrophic yeast.* Journ Mol Cat, 1999. **6**(6): p. 161-73.
124. Dominguez, A., et al., *Non-conventional yeasts as hosts for heterologous protein production.* Int Microbiol, 1998. **1**(2): p. 131-42.
125. Ausubel, F.M., et al., *Current Protocols in Molecular Biology.* 1998, New York, USA.: John Wiley & Sons Inc.
126. Cregg, J.M. and K.A. Russell, *Transformation.* Methods Mol Biol, 1998. **103**: p. 27-39.
127. Verdoucq, L., et al., *In vivo characterization of a thioredoxin h target protein defines a new peroxiredoxin family.* J Biol Chem, 1999. **274**(28): p. 19714-22.
128. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* Faseb J, 1995. **9**(8): p. 576-96.
129. Aasland, R., A.F. Stewart, and T. Gibson, *The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB.* Trends Biochem Sci, 1996. **21**(3): p. 87-8.
130. Bhattacharyya, A., et al., *Whole-genome comparative analysis of three phytopathogenic Xylella fastidiosa strains.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12403-8.
131. Goffeau, A., et al., *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.
132. Velculescu, V.E., et al., *Characterization of the yeast transcriptome.* Cell, 1997. **88**(2): p. 243-51.
133. Becerra, M., et al., *The yeast transcriptome in aerobic and hypoxic conditions: effects of hap1, rox1, rox3 and srb10 deletions.* Mol Microbiol, 2002. **43**(3): p. 545-55.
134. de Nobel, H., et al., *Parallel and comparative analysis of the proteome and transcriptome of sorbic acid-stressed Saccharomyces cerevisiae.* Yeast, 2001. **18**(15): p. 1413-28.

135. Devaux, F., P. Marc, and C. Jacq, *Transcriptomes, transcription activators and microarrays.* FEBS Lett, 2001. **498**(2-3): p. 140-4.
136. Malpertuy, A., et al., *Genomic exploration of the hemiascomycetous yeasts: 19. Ascomycetes-specific genes.* FEBS Lett, 2000. **487**(1): p. 113-21.
137. Blandin, G., et al., *Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta.* FEBS Lett, 2000. **487**(1): p. 76-81.
138. Blandin, G., et al., *Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis.* FEBS Lett, 2000. **487**(1): p. 91-4.
139. Bolotin-Fukuhara, M., et al., *Genomic exploration of the hemiascomycetous yeasts: 11. Kluyveromyces lactis.* FEBS Lett, 2000. **487**(1): p. 66-70.
140. Lesage, P., X. Yang, and M. Carlson, *Analysis of the SIP3 protein identified in a two-hybrid screen for interaction with the SNF1 protein kinase.* Nucleic Acids Res, 1994. **22**(4): p. 597-603.
141. Zitomer, R.S. and C.V. Lowry, *Regulation of gene expression by oxygen in Saccharomyces cerevisiae.* Microbiol Rev, 1992. **56**(1): p. 1-11.
142. Hinnebusch, A.G., *Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome.* J Biol Chem, 1997. **272**(35): p. 21661-4.
143. Vyas, V.K., S. Kuchin, and M. Carlson, *Interaction of the repressors Nrg1 and Nrg2 with the Snf1 protein kinase in Saccharomyces cerevisiae.* Genetics, 2001. **158**(2): p. 563-72.
144. Lohr, D., P. Venkov, and J. Zlatanova, *Transcriptional regulation in the yeast GAL gene family: a complex genetic network.* Faseb J, 1995. **9**(9): p. 777-87.
145. Aves, S.J., et al., *A fission yeast gene mapping close to suc1 encodes a protein containing two bromodomains.* Mol Gen Genet, 1995. **248**(4): p. 491-8.
146. Strick, C.A., et al., *The isolation and characterization of the pyruvate kinase-encoding gene from the yeast Yarrowia lipolytica.* Gene, 1992. **118**(1): p. 65-72.
147. Rhie, G.E., et al., *Manganese-containing superoxide dismutase and its gene from Candida albicans.* Biochim Biophys Acta, 1999. **1426**(3): p. 409-19.
148. Liao, X.S., et al., *Intramitochondrial functions regulate nonmitochondrial citrate synthase (CIT2) expression in Saccharomyces cerevisiae.* Mol Cell Biol, 1991. **11**(1): p. 38-46.
149. Szabo, L.J., G.M. Small, and P.B. Lazarow, *The nucleotide sequence of POX18, a gene encoding a small oleate-inducible peroxisomal protein from Candida tropicalis.* Gene, 1989. **75**(1): p. 119-26.
150. Cho, J.Y. and T.W. Jeffries, *Pichia stipitis genes for alcohol dehydrogenase with fermentative and respiratory functions.* Appl Environ Microbiol, 1998. **64**(4): p. 1350-8.
151. Pavlik, P., et al., *The glycerol kinase (GUT1) gene of Saccharomyces cerevisiae: cloning and characterization.* Curr Genet, 1993. **24**(1-2): p. 21-5.
152. Young, E.T., N. Kacherovsky, and K. Van Riper, *Snf1 protein kinase regulates Adr1 binding to chromatin but not transcription activation.* J Biol Chem, 2002. **277**(41): p. 38095-103.

153. Moreira, J.M. and S. Holmberg, *Nucleosome structure of the yeast CHA1 promoter: analysis of activation-dependent chromatin remodeling of an RNA-polymerase-II-transcribed gene in TBP and RNA pol II mutants defective in vivo in response to acidic activators.* Embo J, 1998. **17**(20): p. 6028-38.
154. Kurland, C.G., *Codon bias and gene expression.* FEBS Lett, 1991. **285**(2): p. 165-9.
155. Rodriguez-Belmonte, E., et al., *PICDI, a simple program for codon bias calculation.* Mol Biotechnol, 1996. **5**(3): p. 191-5.
156. Gordon, D.Abajian, C.Green, P. Consed: a graphical tool for sequence finishing. Genome Res,1998. **8**(3): p.195-02.

APÊNDICE

The genome sequence of the plant pathogen *Xylella fastidiosa*

The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis*, São Paulo, Brazil

* *A full list of authors appears at the end of this paper*

***Xylella fastidiosa* is a fastidious, xylem-limited bacterium that causes a range of economically important plant diseases. Here we report the complete genome sequence of *X. fastidiosa* clone 9a5c, which causes citrus variegated chlorosis—a serious disease of orange trees. The genome comprises a 52.7% GC-rich 2,679,305-base-pair (bp) circular chromosome and two plasmids of 51,158 bp and 1,285 bp. We can assign putative functions to 47% of the 2,904 predicted coding regions. Efficient metabolic functions are predicted, with sugars as the principal energy and carbon source, supporting existence in the nutrient-poor xylem sap. The mechanisms associated with pathogenicity and virulence involve toxins, antibiotics and ion sequestration systems, as well as bacterium–bacterium and bacterium–host interactions mediated by a range of proteins. Orthologues of some of these proteins have only been identified in animal and human pathogens; their presence in *X. fastidiosa* indicates that the molecular basis for bacterial pathogenicity is both conserved and independent of host. At least 83 genes are bacteriophage-derived and include virulence-associated genes from other bacteria, providing direct evidence of phage-mediated horizontal gene transfer.**

Citrus variegated chlorosis (CVC), which was first recorded in Brazil in 1987, affects all commercial sweet orange varieties[1]. Symptoms include conspicuous variegations on older leaves, with chlorotic areas on the upper side and corresponding light brown lesions, with gum-like material on the lower side. Affected fruits are small, hardened and of no commercial value. A strain of *Xylella fastidiosa* was first identified as the causal bacterium in 1993 (ref. 2) and found to be transmitted by sharpshooter leafhoppers in 1996 (ref. 3). CVC control is at present limited to removing infected shoots by pruning, the application of insecticides and the use of healthy plants for new orchards. In addition to CVC, other strains of *X. fastidiosa* cause a range of economically important plant diseases including Pierce's disease of grapevine, alfalfa dwarf, phony peach disease, periwinkle wilt and leaf scorch of plum, and are also associated with diseases in mulberry, pear, almond, elm, sycamore, oak, maple, pecan and coffee[4]. The triply cloned *X. fastidiosa* 9a5c, sequenced here, was derived from the pathogenic culture 8.1b obtained in 1992 in Bordeaux (France) from CVC-affected Valencia sweet orange twigs collected in Macaubal (São Paulo, Brazil) on May 21, 1992 (ref. 2). Strain 9a5c produces typical CVC symptoms on inoculation into experimental citrus plants[5], and into *Nicotiana tabacum* (S. A. Lopes, personal communication) and *Catharantus roseus* (P. Brant-Monteiro, personal communication)—two novel experimental hosts.

General features of the genome

The basic features of the genome are listed in Table 1 and a detailed map is shown in Fig. 1. The conserved origin of replication of the large chromosome has been identified in a region between the putative 50S ribosomal protein L34 and *gyrB* genes containing *dnaA*, *dnaN* and *recF*[6]. The *Escherichia coli* DnaA box consensus sequence TTATCCACA is found on both DNA strands close to *dnaA*. In addition, there are typical 13-nucleotide (ACCACCAC-CACCA) and 9-nucleotide (two TTTCATTGG and two TTTTA-TATT) sequences in other intergenic sequences of this region. This region is coincident with the calculated GC-skew signal inversion[7]. We have designated base 1 of the *X. fastidiosa* genome as the first T of the only TTTTAT sequence found between the ribosomal protein L34 gene and *dnaA*.

The overall percentage of open reading frames (ORFs) for which a putative biological function could be assigned (47%) was slightly below that for other sequenced genomes such as *Thermotoga maritima*[8] (54%), *Deinococcus radiodurans*[9] (52.5%) and *Neisseria meningitidis*[10] (53.7%). This may reflect the lack of previous complete genome sequences from phytopathogenic bacteria. Plasmid pXF1.3 contains only two ORFs, one of which encodes a replication-associated protein. Plasmid pXF51 contains 64 ORFs, of which 5 encode proteins involved in replication or plasmid stability and 20 encode proteins potentially involved in conjugative transfer. One ORF encodes a protein similar to the virulence-associated protein D (VapD), found in many other bacterial pathogens[11]. Four regions of pXF51 present significant DNA similarity to parts of transposons found in plasmids from other bacteria, suggesting interspecific horizontal exchange of genetic material.

The principal paralogous families are summarized in Table 2. The complete list of ORFs with assigned function is shown in Table 3. Seventy-five proteins present in the 21 completely sequenced genomes in the COG database[12] (as of 15th March 2000) were also found in *X. fastidiosa*. Each of these sequences was used to

Table 1 General features of the *Xylella fastidiosa* 9a5c genome

Main chromosome	
Length (bp)	2,679,305
G+C ratio	52.7%
Open reading frames (ORFs)	2,782
Coding region (% of chromosome size)	88.0%
Average ORF length (bp)	799
ORFs with functional assignment	1,283
ORFs with matches to conserved hypothetical proteins	310
ORFs without significant data base match	1,083
Ribosomal RNA operons	2 (16SrRNA-Ala-TGC-tRNA-Ile-GAT-tRNA-23SrRNA-5SrRNA)
tRNAs	49 (46 different sequences corresponding to all 20 amino acids)
tmRNA	1
Plasmid pXF51	
Length (bp)	51,158
G+C ratio	49.6%
Open reading frames (ORFs)	64
Protein coding region (% of plasmid size)	86.9%
ORFs with functional assignment	30
ORFs with matches to conserved hypothetical proteins	8
ORFs without significant data base match	24
Plasmid pXF1.3	
Length (bp)	1,285
G+C ratio	55.6%
Open reading frames (ORFs)	2
ORFs with functional assignment	1

Table 2 Largest families of paralogous genes

Family (total number of families = 312)	Number of genes (total number of genes = 853)
ATP-binding subunits of ABC transporters	23
Reductases/dehydrogenases	12
Two-component system, regulatory proteins	12
Hypothetical proteins	10
Transcriptional regulators	9
Fimbrial proteins	9
Two-component system, sensor proteins	9

generate a phylogenetic tree of the 22 organisms. In 69% of such trees, *X. fastidiosa* was grouped with *Haemophilus influenzae* and *E. coli*, consistent with a phylogenetic analysis undertaken with the 16S rRNA gene[13].

One ORF, a cytosine methyltransferase (XF1774), is interrupted by a Group II intron. The intron was identified on the basis of the presence of a reverse transcriptase-like gene (as in other Group II introns), conserved splice sites, conserved sequence in structure V and conserved elements of secondary structure[14]. Group II introns are rare in prokaryotes, but have been found in different evolutive lineages including *E. coli*, cyanobacteria and proteobacteria[15].

Transcription, translation and repair

The basic transcriptional and translational machinery of *X. fastidiosa* is similar to that of *E. coli*[16]. Recombinational repair, nucleotide and base-excision repair, and transcription-coupled repair are present with some noteworthy features. For example, no photolyase was found, indicating exclusively dark repair. Although the main genes of the SOS pathway, *recA* and *lexA*, are present, ORFs corresponding to the three DNA polymerases induced by SOS in *E. coli* (DNA polymerases II, IV and V)[17] are missing, indicating that the mutational pathway itself may be distinct.

Energy metabolism

Even though *X. fastidiosa* is, as its name suggests, a fastidious organism, energy production is apparently efficient. In addition to all the genes for the glycolytic pathway, all genes for the tricarboxylic acid cycle and oxidative and electron transport chains are present. ATP synthesis is driven by the resulting chemiosmotic proton gradient and occurs by an F-type ATP synthase. Fructose, mannose and glycerol can be utilized in addition to glucose in the glycolytic pathway. There is a complete pathway for hydrolysis of cellulose to glucose, consisting of 1,4-β-cellobiosidase, endo-1,4-β-glucanase and β-glucosidase, suggesting that cellulose breakdown may supplement the often low concentrations of monosaccharides in the xylem[18]. Two lipases are encoded in the genome, but there is no β-oxidation pathway for the hydrolysis of fatty acids, presumably precluding their utilization as an alternative carbon and energy source. Likewise, although enzymes required for the breakdown of threonine, serine, glycine, alanine, aspartate and glutamate are present, pathways for the catabolism of the other naturally occurring amino acids are incomplete or absent.

The gluconeogenesis pathway appears to be incomplete. Phosphoenolpyruvate carboxykinase and the gluconeogenic enzyme fructose-1,6-bisphosphatase, which are required to bypass the irreversible step in glycolysis, are not present. The absence of the first is compensated by the presence of phosphoenolpyruvate synthase and malate oxidoreductase, which together can generate phosphoenolpyruvate from malate. There appears, however, to be no known compensating pathway for the absence of fructose-1,6-bisphosphatase. It is possible that among the large number of unidentified *X. fastidiosa* genes there are non-homologous genes that compensate for steps in such critical pathways. Barring this possibility, however, the absence of a functional gluconeogenesis pathway implies a strict dependence on carbohydrates both as a source of energy and anabolic precursors. The glyoxylate cycle is absent and the pentose phosphate pathway is incomplete. In the latter pathway, genes for neither 6-phosphogluconic dehydrogenase nor transaldolase were identified.

Small molecule metabolism

X. fastidiosa exhibits extensive biosynthetic capabilities, presumably an absolute requirement for a xylem-dwelling bacterium. Most of the genes found in *E. coli* necessary for the synthesis of all amino acids from chorismate, pyruvate, 3-phosphoglycerate, glutamate and oxaloacetic acid[16] were identified. However, some genes in *X. fastidiosa* are bi-functional, such as phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase (XF2213), aspartokinase/homoserine dehydrogenase I (XF2225), imidazoleglycerolphosphate dehydratase/histidinol-phosphate phosphatase (XF2217) and a new diaminopimelate decarboxylase/aspartate kinase (XF1116) that would catalyse the first and the last steps of lysine biosynthesis. In addition, the gene for acetylglutamate kinase (XF1001) has an acetyltransferase domain at its carboxy-terminal end that would compensate for the missing acetyltransferase in the arginine biosynthesis pathway. Other missing genes include phosphoserine phosphatase, cystathionine β-lyase, homoserine *O*-succinyltransferase and 2,4,5-methyltetrahydrofolate-homocysteine methyltransferase. The first two enzymes are also absent in the *Bacillus subtilis* genome, the third is absent in *Haemophilus influenzae* and the fourth is missing in both genomes[12]. We thus presume that alternative, unidentified enzymes complete the biosynthetic pathways in these organisms and in *X. fastidiosa*.

The pathways for the synthesis of purines, pyrimidines and nucleotides are all complete. *X. fastidiosa* is also apparently capable of both synthesizing and elongating fatty acids from acetate. Again, however, some *E. coli* enzymes were not found, such as holo acyl-carrier-protein synthase (also absent in *Synechocystis* sp., *H. influenzae* and *Mycoplasma genitalium*) and enoyl-ACP reductase (NADPH) (*FabI*) (also absent from *M. genitalium*, *Borrelia burgdorferi* and *Treponema pallidum*)[12].

X. fastidiosa appears to be capable of synthesizing an extensive variety of enzyme cofactors and prosthetic groups, including biotin, folic acid, pantothenate and coenzyme A, ubiquinone, glutathione, thioredoxin, glutaredoxin, riboflavin, FMN, FAD, pyrimidine nucleotides, porphyrin, thiamin, pyridoxal 5′-phosphate and lipoate. In a number of the synthetic pathways, one or more of the enzymes present in *E. coli* are absent, but this is also true for at least one other sequenced Gram-negative bacterial genome in each case[12]. We therefore again infer that the missing enzymes are either not essential or replaced by unknown proteins with novel structures.

Transport-related proteins

A total of 140 genes encoding transport-related proteins were identified, representing 4.8% of all ORFs. For comparison, *E. coli*, *B. subtilis* and *M. genitalium* have around 10% of genes encoding transport proteins, whereas *Helicobacter pylori*, *Synechocystis* sp. and *Methanococcus jannaschii* have 3.5–5.4% (ref. 19). Transport systems are central components of the host–pathogen relationship (Fig. 2). There are a number of ion transporters and transporters for the uptake of carbohydrates, amino acids, peptides, nitrate/nitrite, sulphate, phosphate and vitamin B12. Many different transport

Figure 1 Linear representation of the main chromosome and plasmids pXF51 and pXF1.3 of the *Xylella fastidiosa* genome. Genes are coloured according to their biological role. Arrows indicate the direction of transcription. Genes with frameshift and point mutations are indicated with an X. Ribosomal RNA genes, the tmRNA, the principal repeats, prophages and the group II intron are indicated by coloured lines. Transfer RNAs are identified by a single letter identifying the amino acid. Pie chart represents the distribution of the number of genes according to biological role. The numbers below protein-producing genes correspond to gene IDs. ▶

families are represented and include both small and large mechanosensitive conductance ion channels, a monovalent cation:proton antiporter (CAP-2) and a glycerol facilitator belonging to the major intrinsic protein (MIP) family. In addition, 23 ABC transport systems comprising 41 genes can be identified. *X. fastidiosa* appears to possess a phosphotransferase system (PTS) that typically mediates small carbohydrate uptake. There are both the enzyme I and HPr components of this system, as well as a gene supposedly involved in its regulation (*pstK* or *hprK*); however, there is no PTS permease—an essential component of the phosphotransferase complex. The functionality of the system therefore remains in question.

There are five outer membrane receptors, including siderophores, ferrichrome-iron and haemin receptors, which are all associated with iron transport. The energizing complexes, TonB–ExbB–ExbD and the paralogous TolA–TolR–TolQ, essential for the functioning of the outer membrane receptors, are also present. In all, 67 genes encode proteins involved in iron metabolism. We propose that in *X. fastidiosa* the uptake of iron and possibly of other transition metal ions such as manganese causes a reduction in essential micronutrients in the plant xylem, contributing to the typical symptoms of leaf variegation.

The *X. fastidiosa* genome encodes a battery of proteins that mediate drug inactivation and detoxification, alteration of potential drug targets, prevention of drug entry and active extrusion of drugs and toxins. These include ABC transporters and transport processes driven by a proton gradient. Of the latter, eight belong to the hydrophobe/amphiphile efflux-1 (HAE1) family, which act as multidrug resistance factors.

Adhesion

X. fastidiosa is characteristically observed embedded in an extracellular translucent matrix *in planta*[20]. Clumps of bacteria form within the xylem vessels leading to their blockage and symptoms of the disease such as water-stress leaf curling. We deduce, from our analysis of the complete genome sequence, that the matrix is composed of extracellular polysaccharides (EPSs) synthesized by enzymes closely related to those of *Xanthomonas campestris* pv *campestris* (Xcc) that produce what is commercially known as xanthan gum. In comparison with Xcc, however, we did not find *gumI* (encoding glycosyltransferase V, which incorporates the terminal mannose), *gumL* (encoding ketalase which adds pyruvate to the polymer) or *gumG* (encoding acetyltransferase which adds acetate), suggesting that *Xylella* gum may be less viscous than its *Xanthomonas* counterpart.

Positive regulation of the synthesis of extracelullar enzymes and EPS in *Xanthomomas* is effected by proteins coded by the *rpf* (regulation of pathogenicity factors) gene cluster[21]. Mutations in any of these genes in *Xanthomomas* results in failure to synthesize the EPS. In consequence, the strain becomes non-pathogenic[21]. *X. fastidiosa* contains genes that encode RpfA, RpfB, RpfC and RpfF, suggesting that both bacteria may regulate the synthesis of pathogenic EPS factors through similar mechanisms.

Fimbria-like structures are readily apparent upon electron microscopical observation of *X. fastidiosa* within both its plant and insect hosts[22]. Because of the high velocity of xylem sap passing through narrow portions of the insect foregut, fimbria-mediated attachment may be essential for insect colonization. Indeed, in the insect mouthparts the bacteria are attached in ordered arrays, indicating specific and polarized adhesion[23]. In addition, fimbriae are thought to be involved in both plant–bacterium and bacterium–bacterium interactions during colonization of the xylem itself. We identified 26 genes encoding proteins responsible for the biogenesis and function of Type 4 fimbria filaments. This type of fimbria is found at the poles of a wide range of bacterial pathogens where they act to mediate adhesion and translocation along epithelial surfaces[24]. The genes include *pilS* and *pilR* homologues, which encode a two-component system controlling transcription of fimbrial subunits, presumably in response to host cues, and *pilG, H, I, J* and *chpA*, which encode a chemotactic system transducing environmental signals to the pilus machinery.

In addition to the EPS and fimbriae, which are likely to have central roles in the clumping of bacteria and in adhesion to the xylem walls, we also identified outer membrane protein homologues for afimbrial adhesins. Although fimbrial adhesins are well characterized as crucial virulence factors in both plant and human pathogens[25], afimbrial adhesins, which are directly associated with the bacterial cell surface, have been hitherto associated only with human and animal pathogens, where they promote adherence to epithelial tissue. Of the three putative adhesins of this kind identified in *X. fastidiosa*, two exhibit significant similarity to each other (XF1981, XF1529) and to the *hsf* and *hia* gene products of *H. influenzae*[26]. The third (XF1516) is similar to the *uspA1* gene product of *Moraxella catarrhalis*[27]. All these proteins share the common C-terminal domain of the autotransporter family[28]. Direct experimentation will be required to establish whether these adhesins promote binding to plant cell structures or components of the insect vector foregut, or both. Nevertheless, their presence in the *X. fastidiosa* genome adds to the increasing evidence for the generality of mechanisms of bacterial pathogenicity, irrespective of the host organism[29].

We also identified three different haemagglutinin-like genes. Again, similar genes have not previously been identified in plant pathogens. These genes (XF2775, XF2196, XF0889) are the largest in the genome and exhibit highest similarity to a *Neisseria meningitidis* putative secreted protein[10].

Intervessel migration

Movement between individual xylem vessels is crucial for effective colonization by *X. fastidiosa*. For this to occur, degradation of the pit membrane of the xylem vessel is required. Of the known pectolytic enzymes capable of this function, a polygalacturonase precursor and a cellulase were identified, although the former contains an authentic frameshift. These genes exhibited highest similarity to orthologues in *Ralstonia solanacearum*—which causes wilt disease in tomatoes—where the polygalacturonase genes are required for wild-type virulence.

Toxicity

We identified five haemolysin-like genes: haemolysin III (XF0175), which belongs to an uncharacterized protein family, and four others (XF0668, XF1011, XF2407, XF2759) which belong to the RTX toxin family that contains tandemly repeated glycine-rich nonapeptide motifs at the C-terminal domain. One of these ORFs is closely related to bacteriocin, an RTX toxin also found in the plant bacterium *Rhizobium leguminosarum*[30]. RTX or RTX-like proteins are important virulence factors widely distributed among Gram-negative pathogenic bacteria[31].

There are two Colicin-V-like precursor proteins. Colicin V is an antibacterial polypeptide toxin produced by *E. coli*, which acts against closely related sensitive bacteria[32]. The precursors consist of 102-amino-acid peptides (XF0262, XF0263) that have the typical conserved leader 15-amino-acid motif, and have some similarity with Colicin V from *E. coli* at the remaining C-terminal portion. The necessary apparatus for Colicin biosynthesis and secretion is also present. Interestingly, in *E. coli* most of the genes necessary for biogenesis and export of Colicin V are in a gene cluster present in a plasmid, whereas in *X. fastidiosa* these genes are dispersed in the chromosome.

We found four genes that may function in polyketide biogenesis: polyketide synthase (PKS), pteridine-dependent deoxygenase, daunorubicin C-13 ketoreductase and a NonF-related protein. These genes belong to the synthesis pathways of frenolicin, rapamycin, daunorubicin and nonactin, respectively. These pathways

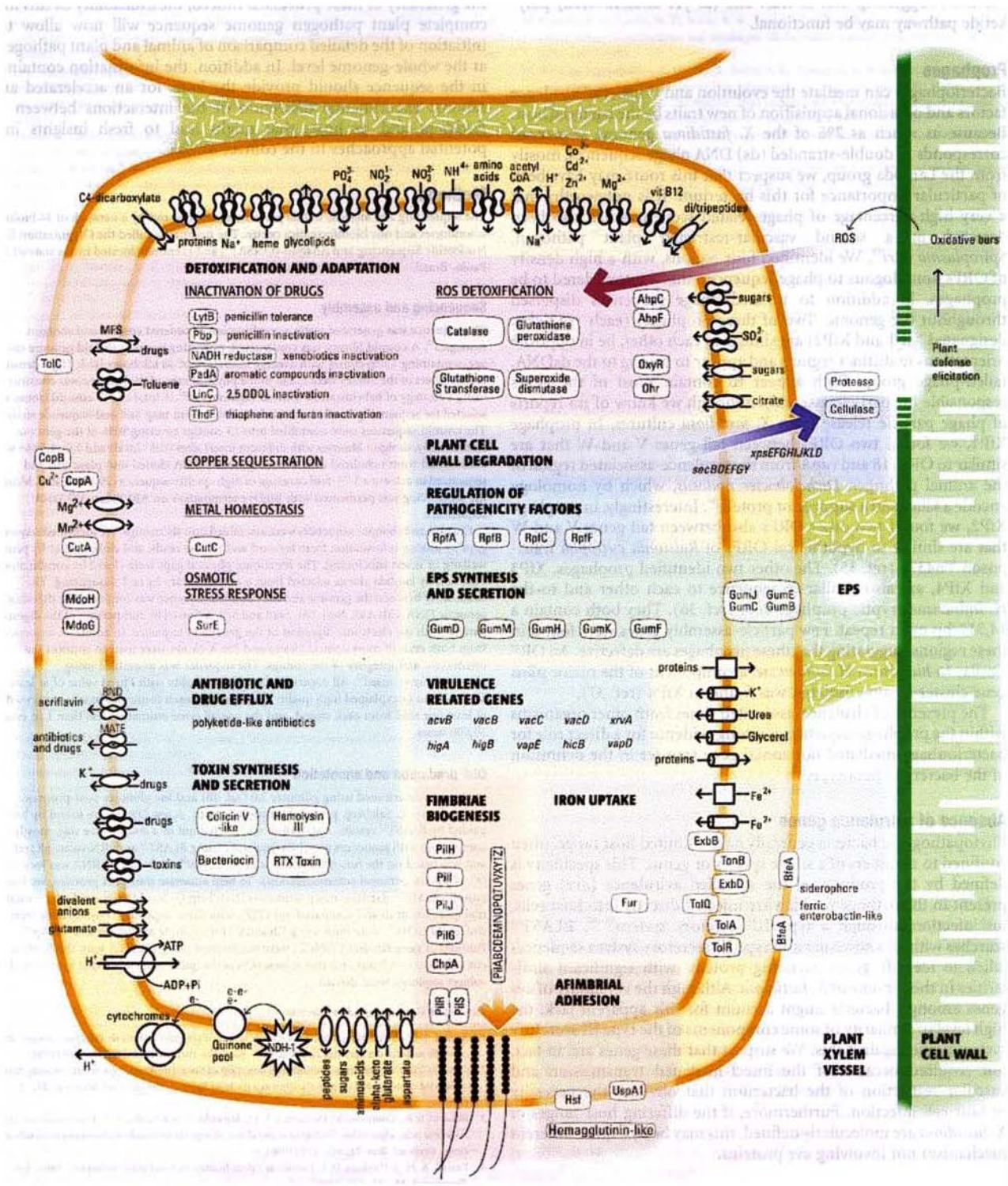


Figure 2 A comprehensive view of the biochemical processes involved in *Xylella fastidiosa* pathogenicity and survival in the host xylem. The principal functional categories are shown in bold, and the bacterial genes and gene products related to that function are arranged within the coloured section containing the bold heading. Transporters are indicated as follows: cylinders, channels; ovals, secondary carriers, including the MFS family; paired dumbbells, secondary carriers for drug extrusion; triple dumbbells, ABC transporters; bulb-like icon, F-type ATP synthase; squares, other transporters. Icons with two arrows represent symporters and antiporters (H^+ or Na^+ porters, unless noted otherwise). 2,5DDOL, 2,5-dichloro-2,5-cyclohexadiene-1,4-diol; EPS, exopolysaccharides; MATE, multi-antimicrobial extrusion family of transporters multidrug efflux gene (XF2686); MFS, major facilitator superfamily of transporters; Pbp, β-lactamase-like penicillin-binding protein (XF1621); RND, resistance-nodulation-cell division superfamily of transporters; ROS, reactive oxygen species.

include many more enzymes, which we did not find; however, some of the genes listed lie close to ORFs without significant database matches, suggesting that at least one (as yet undiscovered) polyketide pathway may be functional.

Prophages

Bacteriophages can mediate the evolution and transfer of virulence factors and occasional acquisition of new traits by the bacterial host. Because as much as 7% of the *X. fastidiosa* genome sequenced corresponds to double-stranded (ds) DNA phage sequences, mostly from the Lambda group, we suspect that this route may have been of particular importance for this bacterium. It is noteworthy that a very high percentage of phage-related sequences has also been detected in a second vascular-restricted plant pathogen, *Spiroplasma citri*[33]. We identified four regions, with a high density of ORFs homologous to phage sequences, that we considered to be prophages, in addition to isolated phage sequences dispersed throughout the genome. Two of these prophages (each ~42 kbp, designated XfP1 and XfP2) are similar to each other, lie in opposite orientations in distinct regions and appear to belong to the dsDNA, tailed-phage group. Both appear to contain most of the genes responsible for particle assembly, although we know of no reports of phage particle release from *X. fastidiosa* cultures. In prophage XfP1, we found two ORFs between tail genes V and W that are similar to ORF118 and *vapA* from the virulence-associated region of the animal pathogen *Dichelobacter nodosus*, which by homology encode a killer and a suppressor protein[34]. Interestingly, in prophage XfP2, we found two other ORFs also between tail genes V and W that are similar to hypothetical ORFs of *Ralstonia eutropha* transposon Tn4371 (ref. 35). The other two identified prophages, XfP3 and XfP4, are also similar in sequence to each other and to the *H. influenzae* cryptic prophage ϕflu (ref. 36). They both contain a 14,317-bp exact repeat. Few particle-assembly genes were found in these regions, suggesting that these prophages are defective. An ORF similar to *hicB* from *H. influenzae*, a component of the major pilus gene cluster in some isolates, was found in XfP4 (ref. 37).

The presence of virulence-associated genes from other organisms within the prophage sequences is strong evidence for a direct role for bacteriophage-mediated horizontal gene transfer in the definition of the bacterial phenotype.

Absence of avirulence genes

Phytopathogenic bacteria generally have a limited host range, often confined to members of a single species or genus. This specificity is defined by the products of the so-called avirulence (*avr*) genes present in the pathogen, which are injected directly into host cells, on infection, through a type III secretory system[38–40]. BLAST[41] searches with all known *avr* and type III secretory system sequences failed to identify genes encoding proteins with significant similarities in the genome of *X. fastidiosa*. Although the variability of *avr* genes amongst bacteria might account for this apparent lack, the high level of similarity of some components of the type III secretory system argues against this. We suspect that these genes are, in fact, not required because of the insect-mediated transmission and vascular restriction of the bacterium that obviates the necessity of host cell infection. Furthermore, if the differing host ranges of *X. fastidiosa* are molecularly defined, this may be by a quite different mechanism not involving avr proteins.

Conclusions

Before the elucidation of its complete genome sequence, very little was known of the molecular mechanisms of *X. fastidiosa* pathogenicity. Indeed, this bacterium was probably the least characterized of all organisms that have been fully sequenced. Our complete genetic analysis has determined not only the basic metabolic and replicative characteristics of the bacterium, but also a number of potential pathogenicity mechanisms. Some of these have not previously been postulated to occur in phytopathogens, providing new insights into the generality of these processes. Indeed, the availability of this first complete plant pathogen genome sequence will now allow the initiation of the detailed comparison of animal and plant pathogens at the whole-genome level. In addition, the information contained in the sequence should provide the basis for an accelerated and rational experimental dissection of the interactions between *X. fastidiosa* and its hosts that might lead to fresh insights into potential approaches to the control of CVC. □

Methods

The sequencing and analysis in this project were carried out by a network of 34 biology laboratories and one bioinformatics centre. The network is called the Organization for Nucleotide Sequencing and Analysis (ONSA)[42], and is entirely located in the state of São Paulo, Brazil.

Sequencing and assembly

The sequence was generated using a combination of ordered cosmid and shotgun strategies[43]. A cosmid library was constructed, providing roughly 15-fold genome coverage, containing 1,056 clones with average insert size of 40 kilobases (kb). High-density colony filters of the library were made, and a physical map of the genome was constructed using a strategy of hybridization without replacement[44]. A total of 113 cosmid clones was selected for sequencing on the basis of the hybridization map and end-sequence analysis. The cosmid sequences were assembled into 15 contigs covering 90% of the genome. Additionally, shotgun libraries with different insert sizes (0.8–2.0 kb and 2.0–4.5 kb) were constructed from nebulized or restricted genomic DNA cloned into plasmids, and sequenced to achieve a 3.74-fold coverage of high-quality sequence (29,140 reads). Most of the sequencing was performed with BigDye terminators on ABI Prism 377 DNA sequencers.

Cosmid and shotgun sequences were assembled into six contigs. We identified sequence gaps by linking information from forward and reverse reads, and closed either by primer walking or insert subcloning. The remaining physical gaps were closed by combinatorial PCR and by lambda clones selected from a λDash library by end-sequencing. The collinearity between the genome and the obtained sequence was confirmed by digestion of genomic DNA with *AscI*, *NotI*, *SfiI*, *SmiI* and *SrfI*, followed by comparison of the digestion pattern with the electronic digestion of the generated sequence. In addition, sequences from both ends of most cosmid clones and 236 λ clones were used to confirm the orientation and integrity of the contigs. The sequence was assembled using phred+phrap+consed[45]. All consensus bases have quality with Phred value of at least 20. There are no unexplained high quality discrepancies, each consensus base is confirmed by at least one read from each strand, and the overall error estimate is less than 1 in every 10,000 bases.

ORF prediction and annotation

ORFs were determined using glimmer 2.0 (ref. 46) and the glimmer post-processor RBSfinder (S. L. Salzberg, personal communication). A few ORFs were found by hand guided by BLAST[41] results. Annotation was carried out in a cooperative way, mostly by comparison with sequences in public databases, using BLAST[41] and tRNAscan-SE (ref. 47) and was based on the functional categories for *E. coli*[48]. Only one tmRNA was located (K. Williams, personal communication). To help annotate transport proteins, we built a custom BLAST[41] database using sequences from http://www-biology.ucsd.edu/~msaier/transport/toc.html and compared our ORFs with these sequences. Phylogenetic trees for conserved COGs[12] were built using ClustalX[49] for multiple alignment and Phylip[50]. Paralogous gene families (Table 2) were determined using BLASTX with the E-value cut-off equal to e-5 and such that at least 60% of the query sequence and at least 30% of the subject sequence were aligned.

Received 24 March; accepted 24 May 2000.

1. Rosseti, V. *et al.* Présence de bactéries dans le xylème d'orangers atteints de chlorose variégée, une nouvelle maladie des agrumes au Brésil. *C. R. Acad. Sci. Paris série III*, **310**, 345–349 (1990).
2. Chang, C. J. *et al.* Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Curr. Microbiol.* **27**, 137–142 (1993).
3. Roberto, S. R., Coutinho, A., De Lima, J. E. O., Miranda, V. S. & Carlos, E. F. Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Dilobopterus costalimai*, *Acrogonia terminalis* e *Oncometopia facialis* em citros. *Fitopatol. Bras.* **21**, 517–518 (1996).
4. Purcell, A. H. & Hopkins, D. L. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* **34**, 131–151 (1996).
5. Li, W. B. *et al.* A triply cloned strain of *Xylella fastidiosa* multiplies and induces symptoms of citrus variegated chlorosis in sweet orange. *Curr. Microbiol.* **39**, 106–108 (1999).
6. Ye, F., Renaudin, J., Bové, J. M. & Laigret, F. Cloning and sequencing of the replication origin (*oriC*) of the *Spiroplasma citri* chromosome and construction of autonomously replicating artificial plasmids. *Curr. Microbiol.* **29**, 23–29 (1994).
7. Francino, M. P. & Ochman, H. Strand asymmetries in DNA evolution. *Trends Genet.* **13**, 240–245 (1997).
8. Nelson, K. E. *et al.* Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 (1999).

9. White, O. *et al.* Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**, 1571–1577 (1999).
10. Tettelin, H. *et al.* Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809–1815 (2000).
11. Katz, M. F., Strugnell, R. A. & Rood, J. I. Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*. *Infect. Immun.* **60**, 4586–4592 (1992).
12. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
13. Preston, G. M., Haubold, B. & Rainey, P. B. Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. *Curr. Opin. Microbiol.* **1**, 589–597 (1998).
14. Knoop, V., Kloska, S. & Brennicke, A. On the identification of group II introns in nucleotide sequence data. *J. Mol. Biol.* **242**, 389–396 (1994).
15. Ferat, J. L. & Michel, F. Group II self-splicing introns in bacteria. *Nature* **364**, 358–361 (1993).
16. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
17. Bridges, B. A. DNA repair: Polymerases for passing lesions. *Curr. Biol.* **9**, R475–R477 (1999).
18. Brodbeck, B. V., Andersen, P. C. & Mizell, R. F. Effects of total dietary nitrogen and nitrogen form on the development of xylophagous leafhoppers. *Arch. Insect Biochem. Physiol.* **42**, 37–50 (1999).
19. Paulsen, I. T., Sliwinski, M. K. & Saier, M. H. J. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *J. Mol. Biol.* **277**, 573–592 (1998).
20. Chagas, C. M., Rossetti, V. & Beretta, M. J. G. Electron-microscopy studies of a xylem-limited bacterium in sweet orange affected with citrus variegated chlorosis disease in Brazil. *J. Phytopathol.* **134**, 306–312 (1992).
21. Tang, J. L. *et al.* Genetic and molecular analysis of a cluster of *rpf* genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *campestris*. *Mol. Gen. Genet.* **226**, 409–417 (1991).
22. Raju, C. B. & Wells, J. M. Diseases caused by fastidious xylem-limited bacteria. *Plant Disease* **70**, 182–186 (1986).
23. Brlansky, R. H., Timmer, L. W., French, W. J. & McCoy, R. E. Colonization of the sharpshooter vectors, *Oncometopia igricans* and *Homalodisca coagulata*, by xylem-limited bacteria. *Phytopathology* **73**, 530–535 (1983).
24. Fernandez, L. A. & Berenguer, J. Secretion and assembly of regular surface structures in Gram-negative bacteria. *FEMS Microbiol. Rev.* **24**, 21–44 (2000).
25. Soto, G. E. & Hultgren, S. J. Bacterial adhesins: common themes and variations in architecture and assembly. *J. Bacteriol.* **181**, 1059–1071 (1999).
26. Geme, J. W. Molecular determinants of the interaction between *Haemophilus influenzae* and human cells. *Am. J. Respir. Crit. Care Med.* **154**, S192–S196 (1996).
27. Cope, L. D. *et al.* Characterization of the *Moraxella catarrhalis uspA1* and *uspA2* genes and their encoded products. *J. Bacteriol.* **181**, 4026–4034 (1999).
28. Henderson, I. R., Navarro-Garcia, F. & Nataro, J. P. The great escape: structure and function of the autotransporter proteins. *Trends Microbiol.* **6**, 370–378 (1998).
29. Rahme, L. G. *et al.* Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268**, 1899–1902 (1995).
30. Oresnik, I. J., Twelker, S. & Hynes, M. F. Cloning and characterization of a *Rhizobium leguminosarum* gene encoding a bacteriocin with similarities to RTX toxins. *Appl. Environ. Microbiol.* **65**, 2833–2840 (1999).
31. Lally, E. T., Hill, R. B., Kieba, I. R. & Korostoff, J. The interaction between RTX toxins and target cells. *Trends Microbiol.* **7**, 356–361 (1999).
32. Havarstein, L. S., Holo, H. & Nes, I. F. The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. *Microbiology* **140**, 2383–2389 (1994).
33. Ye, F. *et al.* A physical and genetic map of the *Spiroplasma citri* genome. *Nucleic Acids Res.* **20**, 1559–1565 (1992).
34. Billington, S. J., Johnston, J. L. & Rood, J. I. Virulence regions and virulence factors of the ovine footrot pathogen, *Dichelobacter nodosus*. *FEMS Microbiol. Lett.* **145**, 147–156 (1996).
35. Merlin, C., Springael, D. & Toussaint, A. Tn4371: A modular structure encoding a phage-like integrase, a Pseudomonas-like catabolic pathway, and RP4/Ti-like transfer functions. *Plasmid* **41**, 40–54 (1999).
36. Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl Acad. Sci. USA* **96**, 2192–2197 (1999).
37. Mhlanga-Mutangadura, T., Morlin, G., Smith, A. L., Eisenstark, A. & Golomb, M. Evolution of the major pilus gene cluster of *Haemophilus influenzae*. *J. Bacteriol.* **180**, 4693–4703 (1998).
38. Alfano, J. R. & Collmer, A. The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J. Bacteriol.* **179**, 5655–5662 (1997).
39. Galan, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322–1328 (1999).
40. Young, G. M., Schmiel, D. H. & Miller, V. L. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proc. Natl Acad. Sci. USA* **96**, 6456–6461 (1999).
41. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
42. Simpson, A. J. & Perez, J. F. ONSA, the São Paulo Virtual Genomics Institute. *Nature Biotechnol.* **16**, 795–796 (1998).
43. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
44. Hoheisel, J. D. *et al.* High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* **73**, 109–120 (1993).
45. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
46. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636–4641 (1999).
47. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res.* **25**, 955–964 (1997).
48. Riley, M. Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* **57**, 862–952 (1993).
49. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
50. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3. 2). *Cladistics* **5**, 164–166 (1989).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com), or on the author's World-Wide Web site (http://www.lbi.ic.unicamp.br/xf) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

The consortium is indebted to J. F. Perez, Scientific Director of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), for his strategic vision in creating and nurturing this project as well as C. A. de Pian and Juçara Parra for their administrative coordination. We thank our Steering Committee: S. Oliver, A. Goffeau, J. Sgouros, A. C. M. Paiva and J. L. Azevedo for their critical accompaniment of the work. We also thank R. Fulton and P. Minx for their timely contribution and advice. Project funding was from FAPESP, the RHAE programme of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundecitrus. For the full list of individuals who contributed to the completion of this project see (http://www.lbi. ic.unicamp.br/xf).

Correspondence and requests for materials should be addressed to J.C.S. (e-mail: setubal@ic.unicamp.br).

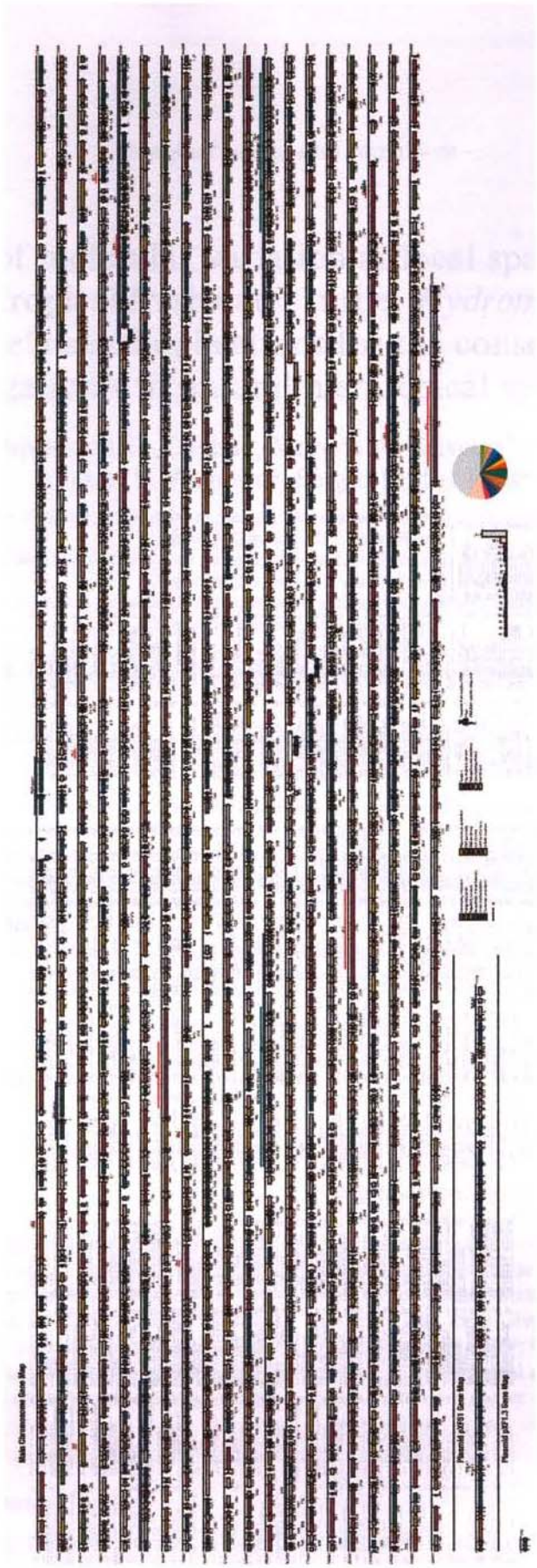
The sequence has been deposited in GenBank with accession numbers AE003849 (chromosome), AE003850 (pXF1.3) and AE003851 (pXF51).

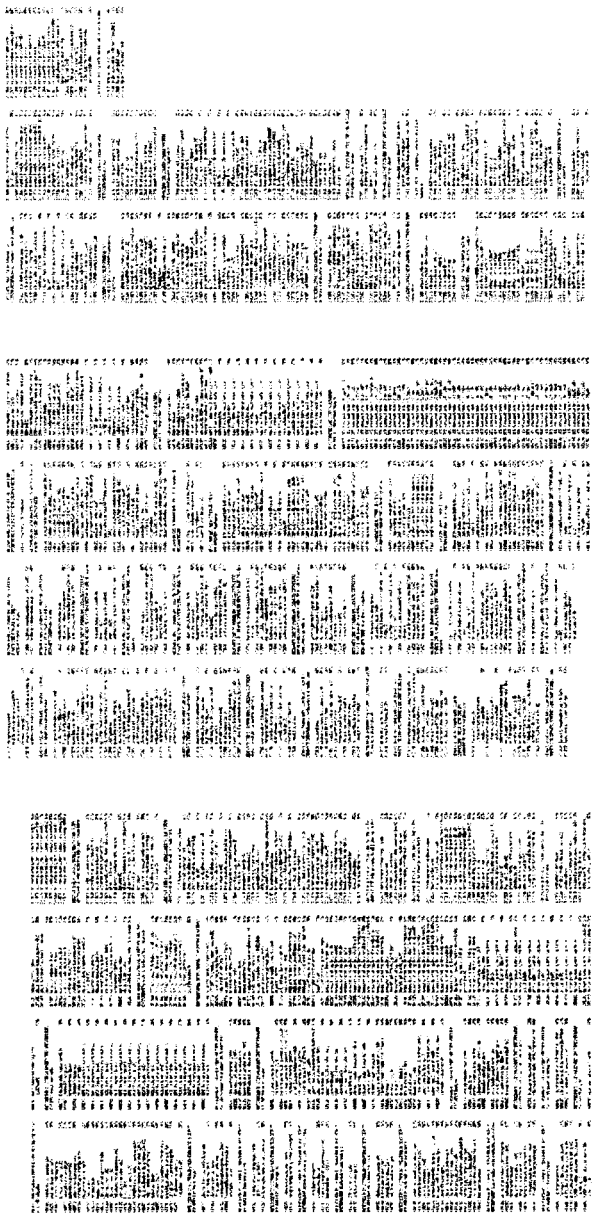
Authors: A. J. G. Simpson[1], F.C. Reinach[2], P. Arruda[3], F. A. Abreu[4], M. Acencio[5], R. Alvarenga[2], L. M. C. Alves[6], J. E. Araya[7], G. S. Baia[2], C. S. Baptista[8], M. H. Barros[8], E. D. Bonaccorsi[2], S. Bordin[9], J. M. Bové[10], M. R. S. Briones[7], M. R. P. Bueno[11], A. A. Camargo[1], L. E. A. Camargo[12], D. M. Carraro[12], H. Carrer[12], N. B. Colauto[13], C. Colombo[14], F. F. Costa[9], M. C. R. Costa[15], C. M. Costa-Neto[16], L. L. Coutinho[12], M. Cristofani[17], E. Dias-Neto[1], C. Docena[2], H. El-Dorry[2], A. P. Facincani[6], A. J. S. Ferreira[2], V. C. A. Ferreira[18], J. A. Ferro[6], J. S. Fraga[4], S. C. França[19], M. C. Franco[20], M. Frohme[21], L. R. Furlan[22], M. Garnier[10], G. H. Goldman[23], M. H. S. Goldman[24], S. L. Gomes[2], A. Gruber[4], P. L. Ho[25], J. D. Hoheisel[21], M. L. Junqueira[26], E. L. Kemper[3], J. P. Kitajima[27], J. E. Krieger[26], E. E. Kuramae[28], F. Laigret[10], M. R. Lambais[12], L. C. C. Leite[25], E. G. M. Lemos[6], M. V. F. Lemos[29], S. A. Lopes[19], C. R. Lopes[13], J. A. Machado[30]†, M. A. Machado[17], A. M. B. N. Madeira[4], H. M. F. Madeira[12]†, C. L. Marino[13], M. V. Marques[8], E. A. L. Martins[25], E. M. F. Martins[18], A. Y. Matsukuma[2], C. F. M. Menck[8], E. C. Miracca[5], C. Y. Miyaki[11], C. B. Monteiro-Vitorello[12], D. H. Moon[20], M. A. Nagai[5], A. L. T. O. Nascimento[25], L. E. S. Netto[11], A. Nhani Jr[6], F. G. Nobrega[8]†, L. R. Nunes[31], M. A. Oliveira[32], M. C. de Oliveira[33], R. C. de Oliveira[31], D. A. Palmieri[13], A. Paris[13], B. R. Peixoto[2], G. A. G. Pereira[32], H. A. Pereira Jr[6], J. B. Pesquero[18], R. B. Quaggio[2], P. G. Roberto[19], V. Rodrigues[34], A. J. de M. Rosa[34], V. E. de Rosa Jr[28], R. G. de Sá[34], R. V. Santelli[2], H. E. Sawasaki[14], A. C. R. da Silva[2], A. M. da Silva[2], F. R. da Silva[3,27], W. A. Silva Jr[15], J. F. da Silveira[7], M. L. Z. Silvestri[2], W. J. Siqueira[14], A. A. de Souza[17], A. P. de Souza[3], M. F. Terenzi[23], D. Truffi[12], S. M. Tsai[20], M. H. Tsuhako[18], H. Vallada[35], M. A. Van Sluys[33], S. Verjovski-Almeida[2], A. L. Vettore[3], M. A. Zago[15], M. Zatz[11], J. Meidanis[27] & J. C. Setubal[27].

Addresses: 1, Instituto Ludwig de Pesquisa sobre o Câncer, Rua Prof. Antonio Prudente, 109 - 4° andar, 01509-010, São Paulo - SP, Brazil; 2, Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes, 748, 05508-900, São Paulo - SP, Brazil; 3, Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Caixa Postal 6010, 13083-970, Campinas - SP, Brazil; 4, Laboratório de Biologia Molecular, Departamento de Patologia, Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, Av. Prof. Dr. Orlando Marques de Paiva, 87, 05508-000, São Paulo - SP, Brazil; 5, Disciplina de Oncologia, Departamento de Radiologia, Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo, 455, 01296-903, São Paulo - SP, Brazil; 6, Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, Universidade Estadual Paulista, Via

de Acesso Prof. Paulo D. Castellane s/n, Km 5, 14884-900, Jaboticabal - SP, Brazil; 7, Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Botucatu, 862, 04023-062, São Paulo - SP, Brazil; 8, Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Av. Prof. Lineu Prestes, 1374, 05508-900, São Paulo - SP, Brazil; 9, Hemocentro, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, 13083-97, Campinas - SP, Brazil; 10, Institut National de la Recherche Agronomique et Université Victor Ségalen Bordeaux 2, 71 Avenue Edouard Bourleaux, Laboratoire de Biologie Cellulaire et Moléculaire, 33883 Vilenave d'Ornon Cedex, France; 11, Departamento de Biologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-900, São Paulo - SP, Brazil; 12, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Av. Pádua Dias, 11, 13418-900, Piracicaba - SP, Brazil; 13, Departamento de Genética, Instituto de Biociências, Universidade Estadual Paulista, Distrito de Rubião Junior, 18618-000, Botucatu - SP, Brazil; 14, Centro de Genética, Biologia Molecular e Fitoquímica, Instituto Agronômico de Campinas, Av. Barão de Itapura, 1481, Caixa Postal 28, 13001-970, Campinas - SP, Brazil; 15, Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14049-900, Ribeirão Preto - SP, Brazil; 16, Departamento de Biofísica, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Botucatu, 862, 04023-062, São Paulo - SP, Brazil; 17, Centro de Citricultura Sylvio Moreira, Instituto Agronômico de Campinas, Caixa Postal 04, 13490-970, Cordeirópolis - SP, Brazil; 18, Laboratório de Bioquímica Fitopatológica e Laboratório de Imunologia, Instituto Biológico, Av. Cons. Rodrigues Alves, 1252, 04014-002, São Paulo - SP, Brazil; 19, Departamento de Biotecnologia de Plantas Medicinais, Universidade de Ribeirão Preto, Av. Costábile Romano, 2201, 14096-380, Ribeirão Preto - SP, Brazil; 20, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Av. Centenário, 303, Caixa Postal 96, 13400-970, Piracicaba - SP, Brazil; 21, Funktionelle Genomanalyse, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120, Heidelberg, Germany; 22, Departamento de Melhoramento e Nutrição Animal, Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Fazenda Lageado, Caixa Postal 560, 18600-000, Botucatu - SP, Brazil; 23, Departamento de Ciências Farmacêuticas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Av. do Café s/n, 14040-903, Ribeirão Preto - SP, Brazil; 24, Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14040-901, Ribeirão Preto - SP, Brazil; 25, Centro de Biotecnologia, Instituto Butantan, Av. Vital Brasil, 1500, 05503-900, São Paulo - SP, Brazil; 26, Laboratório de Genética e Cardiologia Molecular/LIM 13, Instituto do Coração (InCor), Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Enéas de Carvalho Aguiar, 44, 05403-000, São Paulo - SP, Brazil; 27, Instituto de Computação, Universidade Estadual de Campinas, Caixa Postal 6176, 13083-970, Campinas - SP, Brazil; 28, Departamento de Produção Vegetal, Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Fazenda Lageado, Caixa Postal 237, 18603-970, Botucatu - SP, Brazil; 29, Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, Universidade Estadual Paulista, Via de Acesso Prof. Paulo Donato Castellane, 14884-900, Jaboticabal - SP, Brazil; 30, Hospital do Câncer - A.C. Camargo, R. Antonio Prudente, 211, 01509-010, São Paulo - SP, Brazil; 31, Núcleo Integrado de Biotecnologia, Universidade de Mogi das Cruzes, Av. Dr. Cândido Xavier de Almeida Souza, 200, 08780-911, Mogi das Cruzes - SP, Brazil; 32, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6010, 13083-970, Campinas - SP, Brazil; 33, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-900, São Paulo - SP, Brazil; 34, Departamento de Parasitologia, Microbiologia e Imunologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14049-900, Ribeirão Preto - SP, Brazil; 35, Departamento de Psiquiatria, Instituto de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo, Rua Dr. Ovídeo Pires de Campos s/n, Sala 4051 (3º. andar), 05403-010, São Paulo - SP, Brazil.

† Present addresses: Novartis Seeds LTDA, Av. Prof. Vicente Rao, 90, 04706-900, São Paulo – SP, Brazil (J. A. Machado); Centro de Ciências Agrárias e Ambientais, Pontifícia Universidade Católica do Paraná, BR-376, Km 14, Caixa Postal 129, 83010-500, São José dos Pinhais – PR, Brazil (H. M. F. Madeira); Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba, Av. Shishimi Hifumi, 2911, 12244-000, São José dos Campos – SP, Brazil (F. G. Nobrega).





X

ELSEVIER

Biological Conservation 104 (2002) 119–126

BIOLOGICAL CONSERVATION

www.elsevier.com/locate/biocon

Partitioning of molecular variation at local spatial scales in the vulnerable neotropical freshwater turtle, *Hydromedusa maximiliani* (Testudines, Chelidae): implications for the conservation of aquatic organisms in natural hierarchical systems

Franco L. Souza[a], Anderson F. Cunha[b], Marcos A. Oliveira[b], Gonçalo A.G. Pereira[b], Hildete P. Pinheiro[c], Sérgio F. dos Reis[a,*]

[a]*Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6109, 13083-970 Campinas, São Paulo, Brazil*

[b]*Departamento de Genética, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6109, 13083-970 Campinas, São Paulo, Brazil*

[c]*Departamento de Estatística, Instituto de Matemática e Computação Científica, Universidade Estadual de Campinas, Caixa Postal 6065, 13083-970 Campinas, São Paulo, Brazil*

Received 15 November 2000; received in revised form 26 April 2001; accepted 26 June 2001

Abstract

Hydromedusa maximiliani is a vulnerable freshwater turtle endemic to mountainous regions of the Atlantic rainforest in south-eastern Brazil. Random amplified polymorphic DNA (RAPD) were surveyed with the purpose of assessing the genetic structure and determining the partitioning of molecular variation in *H. maximiliani* across the natural spatial hierarchical scale of its habitat. The goal of the study was to integrate ecological data with estimates of molecular genetics diversity to develop strategies for the conservation of this freshwater turtle. Specimens were sampled from rivers and streams across three drainages. Nine of the 80 primers used generated 27 scoreable bands of which 10 (37%) were polymorphic and produced 16 RAPD phenotypes. Significant heterogeneity was found in the distribution of RAPD molecular phenotypes across the three drainages. Analysis of molecular variance for molecular phenotypes showed that the heterogeneity had a spatial structure since a significant amount (22%) of the total variance was attributable to variation among rivers and streams. Since the genetic variation of this turtle seems to be structured according to the natural hierarchical system of rivers and streams within drainages, it is suggested that local populations should be considered as separate management units.

Keywords: Conservation; Habitat fragmentation; *Hydromedusa maximiliani*; Molecular variation; Random amplified polymorphic DNA; Spatial scales; Turtle

1. Introduction

The degradation and fragmentation of natural ecosystems has produced landscape islands, with the populations in such areas becoming small because of the reduction in the species' original distribution (Primack, 1998). In species with small populations resulting from habitat fragmentation, stochastic factors (demographic and environmental) assume considerable importance in the population dynamics and in the species survival (Templeton et al., 1990; Lacy, 1993; Primack, 1998). Understanding the ecological and demographic patterns which drive population dynamics is therefore fundamental for conservation efforts (Caughley, 1994), particularly for long lived organisms such as turtles (Dunham et al., 1989; Congdon et al., 1994). Because of their limited dispersal capabilities (exception for marine and some freshwater turtles; Pritchard and Trebbau, 1984; Nichols et al., 2000; Valenzuela, 2001) and their often specific habitat requirements (Pritchard and Trebbau, 1984; Ernst and Barbour, 1989; Cabrera, 1998), turtles are highly vulnerable to becoming

* Corresponding author. Tel.: +55-019-788-7651; fax: +55-019-289-3124.

E-mail address: sfreis@unicamp.br (S.F. dos Reis).

0006-3207/02/$ - see front matter
PII: S0006-3207(01)00160-4

restricted to habitat-isolated populations with little or no gene flow between them.

Hydromedusa maximiliani is endemic to the Atlantic forest of the coastal region of eastern Brazil and ranges from the State of Espírito Santo to the State of São Paulo (Ernst and Barbour, 1989; Iverson, 1992). The Atlantic forest habitat of *H. maximiliani* has been severely disturbed by humans since the sixteenth century, and this has resulted in extensive fragmentation of the tropical mountain rain forest (Dean, 1996; Chiarello, 1999; Tabarelli et al., 1999). Throughout its range the species distribution is disjunct, occurring in isolation at elevations above 600 m. Locally, its habitat is topologically complex, with sequences of ridges and valleys each drained by river and stream systems (Souza, 1995).

The natural disjunct distribution of *H. maximiliani* across its range and the increasing fragmentation of its natural habitat, associated with a low reproductive rate and highly seasonal dynamics (Souza, 1995), makes this species particularly susceptible to demographic and environmental stochastic phenomena that may interact to cause local population extinction (McCarthy, 1996; Hanski, 1998; Legendre et al., 1999). Capture–recapture studies have shown that these turtles have a limited dispersal ability, with a mean daily displacement of 2 m. Thus, streams are occupied by the same turtles for a long time (Souza, 1995; Souza and Abe, 1997a). The dispersion or migratory behavior is limited to the rainy season, when temporary or intermittent water systems are connected with the main water courses (Souza and Abe, 1997a). This limited dispersion suggests that each population of *H. maximiliani* inhabits a specific river within a given drainage system, and therefore exists as a metapopulation (sensu Hanski and Simberloff, 1997). The population dynamics of *H. maximiliani* associated with its habitat characteristics led Souza and Abe (1997a) to define the conservation status of this species as Vulnerable.

Except for the studies of population ecology reviewed above, very little is known on any other aspects of the biology of this species. In this paper, we used RAPD markers (Williams et al., 1993) to investigate the genetic variability in *H. maximiliani* sampled according to the drainage systems typical of the species habitat. Molecular markers targeted by RAPD have been used increasingly to determine the patterns of genetic variation within populations and to partition genetic variation among populations of vertebrate species (Gibbs et al., 1994; Haig et al., 1994; Gibbs, 1998; Mockford et al., 1999; Cooper, 2000; Vucetich et al., 2001). As emphasized by Haig et al. (1994), appropriately sampled RAPD molecular phenotypes can be conveniently tested for homogeneity across populations, and molecular variation can be partitioned into hierarchical levels to yield information on variation and population structure, with important implications for evolutionary and conservation biology. Here, measures of genetic subdivision were used to determine the partitioning of variation in *H. maximiliani* across the natural spatial hierarchical scale of its habitat, which involved drainages and rivers and streams within drainages. The goal of the study was to integrate ecological records with estimates of molecular genetics diversity to develop strategies for the conservation of *H. maximiliani*.

2. Methods

2.1. Sample collection

Field work was conducted from November 1998 to November 1999 at the Parque Estadual de Carlos Botelho (PECB), state of São Paulo, southeastern Brazil (24°00′–24°15′ S, 47°45′–48°10′ W). The PECB is a protected reserve that encompasses over 37,000 ha of intact tropical mountain rain forest typical of southeastern Brazil (Whitmore, 1990; Veloso et al., 1991). The region is topologically complex, with ridges and valleys drained by numerous rivers and streams (Pfeifer et al., 1986; Souza and Abe, 1998).

During this study an area of approximately 2700 ha containing three drainages was sampled based on the natural spatial hierarchy formed by the rivers and streams. The three drainages are referred to as I, II, and III (Fig. 1). Within each drainage, specimens of *H. maximiliani* were hand-caught in shallow rivers and streams (Fig. 1). In this Figure, each dot corresponds to the collecting site of each individual. The sampling effort was not equal in the three drainages because of difficulties in reaching rivers and streams in drainages II and III. The number of individuals obtained for this study in each drainage were: drainage I ($n=25$), drainage II ($n=8$), and drainage III ($n=11$).

From each individual, 200–300 μl of blood was drawn from the scapula vein/brachial artery (Avery and Vitt, 1984) using a 26-gauge needle and a 1-ml syringe. The blood samples were immediately preserved in plastic vials containing 1 ml of absolute ethanol (Miyaki et al., 1998) and stored at room temperature. All turtles not already individually identified by marginal scute notches used in earlier studies (Souza, 1995; Souza and Abe, 1995, 1997a,b, 1998) were marked at the moment of blood sampling as part of a long-term study. The turtles were released at the point of capture after blood sample collection.

2.2. DNA extraction, primer selection, and RAPD-PCR amplifications

Genomic DNA was extracted from the blood samples by two successive organic extractions with phenol:-chloroform using the protocol outlined by Bruford et al.

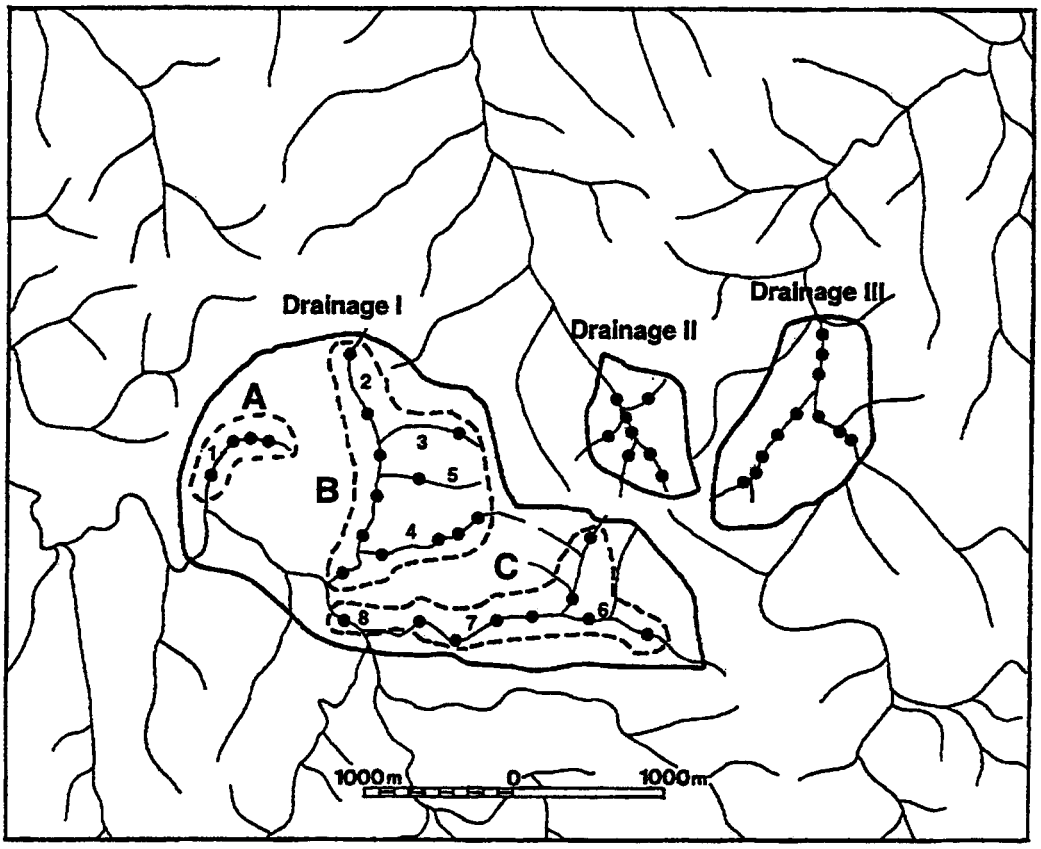


Fig. 1. Map of the study site at Parque Estadual de Carlos Botelho, state of São Paulo, southeastern Brazil, showing the three drainages and rivers sampled for *Hydromedusa maximiliani*. Dots indicate the collecting site of each individual. For drainage I, numbers 1–8 indicate rivers surveyed while dashed lines indicate the three samples sites (A, B, and C) of samples pooled for AMOVA (see text).

(1992) and Miyaki et al. (1998), and then precipitated with 1/10 volume of 1 M sodium acetate (pH 5.3) and two volumes of 100% ethanol. Approximately 100 μl of blood/ethanol solution from each turtle were placed in Eppendorf tubes containing 300 μl of 1X TNE (50 mM Tris–HCl, 100 mM NaCl, 6.3 mM EDTA, pH 7.5), 30 μl 1 M Tris–HCl pH 7.5, 20 μl of proteinase K (25 mg/ml) and 8 μl of 25% SDS, and kept at 37 °C for 12 h. The DNA pellet was washed in 70% ethanol and allowed to dry for 5 min before redissolving in 100–150 μl of sterile. The quality of the extracted DNA was evaluated in agarose gels (0.7%) stained with ethidium bromide and quantified by comparison with DNA standards run in the same gel. DNA was diluted to a working concentration of approximately 0.6 ng/μl.

The primers (Operon Technologies, Inc.; Alameda, California, EUA) were selected using a two-step procedure. Initially, 80 primers (Kits A, C, G, N; 20 primers per kit) were surveyed in side-by-side duplicate reactions for two turtles to identify primers that yielded a reproducible and scoreable banding pattern. Only consistently well amplified bands were screened and variation in intensity was ignored. This first approach revealed nine primers meeting these criteria. These nine primers were used to analyze the full set of DNA samples from the 44 individuals captured in the studied area.

The polymerase chain reaction (PCR) was done in a Perkin-Elmer GeneAmp™ PCR system 9700 with a total volume of 12.5 μl containing 10 mM Tris HCl (pH 8.4), 50 mM KCl, 3.5 mM $MgCl_2$, 1 μl of each dNTP, 2 μl of primer, 0.3 μl *Taq* polymerase, 1.2 ng of genomic DNA and sterile water. Negative controls in which DNA was substituted for water were run to check for the possibility of contamination. Reproducibility was assessed by comparing duplicate reactions, which were usually adjacent to one another in the thermocycler, and products were run side-by-side on the same gel. The reactions condition involved: initial denaturation of DNA for 2 min at 94 °C, 39 cycles of 1 min denaturation at 94 °C, 1 min annealing at 40 °C, 2 min extension at 72 °C, and one 5 min cycle at 72 °C for final extension. The amplification products were separated on 1.3% agarose gels stained with ethidium bromide, run in 1X TAE buffer at 80 V for 5 h. Monochrome photographic negatives were taken of the gels and the

individual profiles were scored by two of us (FLS and AFC) for the presence/absence of fragments for each primer.

2.3. Statistical analyses

Statistical analyses were done based on the spatial hierarchy (rivers and streams within drainages, and among drainages) of sampling defined by the topographic characteristics of the *H. maximiliani* habitat. The null hypothesis that the distribution of RAPD molecular phenotypes was homogeneous across the three drainages was tested with Fisher's exact test because of the small number of individuals in some cell frequencies in the contingency tables (Sokal and Rohlf, 1981). The heterogeneity in RAPD phenotype distribution among rivers and streams within each drainage was also verified with Fisher's exact test and, whenever appropriate, sequential Bonferroni adjustment for multiple comparisons was applied (Sokal and Rohlf, 1981). The partitioning of molecular variation at the hierarchical levels for which significant departures from the null hypothesis were detected was done by the analysis of molecular variation (AMOVA; Excoffier et al., 1992). The theoretical foundation for AMOVA as derived by Excoffier et al. (1992) is framed in the context of *U*-statistics (Hoeffding, 1948; Pinheiro, 1997), and is based on the fact that the sum of squares in conventional analysis of variance can be written as the sum of squared distances between pairs of phenotypes. The distance between individual RAPD molecular phenotypes used here was a simple Euclidean metric and was calculated by summing the squared differences of pairwise vectors of zeros (band absent) and ones (band present) over all polymorphic loci (Excoffier et al., 1992). The phenotypic distances were partitioned into components of variance in the AMOVA framework and the Φ statistics (Excoffier et al., 1992) was calculated from the variance component representing variation at the hierarchical levels defined by the sampling design for this study. Fisher's exact test was computed with PROC FREQ of SAS (SAS/STAT, 1989) and the AMOVA was computed with Arlequin ver 1.1 (Schneider et al., 1997).

3. Results

Nine of the 80 primers used yielded band patterns that were clear and could be scored with confidence (Table 1). These primers produced 27 potentially scoreable bands (range 1–6; mean 3), of which 10 bands (37%) were polymorphic and produced 16 RAPD phenotypes. Twelve of the 16 phenotypes were represented in the 25 individuals sampled from drainage I, and eight phenotypes were found only in the rivers and streams of this drainage. Five phenotypes were represented in the eight individuals sampled from drainage II, of which one phenotype occurred only in this drainage. Five phenotypes were represented in the 11 individuals sampled from drainage III, of which three phenotypes were found only in this drainage. Three phenotypes were common to all three drainages.

The null hypothesis that RAPD molecular phenotypes were distributed homogeneously across drainages was rejected by Fisher's exact test for comparison of drainages I, II, and III simultaneously ($P < 0.000656$). To determine which of the three drainages differed significantly in the distribution of RAPD phenotypes, pairwise comparisons between the three drainages were done with Fisher's exact test. The appropriate level of significance that controls for Type I error was obtained through sequential Bonferroni adjustment for multiple comparisons by dividing our α level of 0.05 by the number of comparisons (3), which yielded an adjusted significance level of $P < 0.016$. Comparison of drainages I and II, and of drainages II and III resulted in non-significant differences (P-values of 0.095 and 0.263, respectively), whereas drainages I and III differed significantly ($P < 0.000195$) in the distribution of RAPD molecular phenotypes.

The organization of variation in molecular phenotypes at the level of rivers and streams within drainages was also evaluated with Fisher's exact test. A deviation from the null hypothesis of homogeneous distribution of phenotypes among rivers and streams was found for drainage I ($P < 0.015$), whereas for drainages II and III the results were not significant ($P = 0.871$ and $P = 0.558$, respectively).

Drainage I, which gave a statistically significant result in Fisher's exact test, was investigated with AMOVA in order to provide insight into the partitioning of RAPD phenotypic variation among rivers and streams in drainage I. Since some rivers in drainage I were sampled for only a few individuals, samples were pooled according to the spatial hierarchy of the main rivers and their tributaries, resulting in three samples sites, as follows: sample A including river 1, sample B including rivers 2, 3, 4, and 5, and sample C including rivers 6, 7, and 8 (Fig. 1). For this sampling design, the phenotypic AMOVA analysis was used to derive variance compo-

Table 1
RAPD primers, their sequences, and the size (molecular weight) of the polymorphic marker bands generated for *Hydromedusa maximiliani*

Primer	Sequence (5′–3′)	Polymorphic bands
OPA07	GAAACGGGTG	625
OPA09	GGGTAACGCC	425
OPA10	GTGATCGCAG	425, 475
OPA20	GTTGCGATCC	450
OPN01	CTCACGTTGG	725, 750
OPN09	TGCCGGCTTG	550, 850, 950

nents and to calculate the value of Φ_{ST} that quantifies the level of variation among sample areas A, B, and C. The level of Φ_{ST} was statistically significant ($P < 0.00684$) and accounted for 22% of the total variance in the molecular phenotypes among rivers and streams within drainage I (Table 2).

To gain further insight into the pattern of molecular variation in drainage I the relationship between pairwise RAPD phenotype distances, computed as Euclidean distances squared (Rohlf, 1994) and spatial distance, measured in river distance (measurements taken with a swivel handle map measurer), was assessed with Mantel's (1967) test using NTSYS-pc (Rohlf, 1994). The correlation between spatial and molecular phenotype distances ($r = -0.125$) for all individuals sampled in drainage I was not significant ($t = -1.228$; $P = 0.110$). Mantel's test was also applied to all individuals samples in the three drainages. The pairwise RAPD phenotype distances were calculated as above, whereas the geographic distances were calculated as great circle (or straight-line) distances because rivers in different drainages are not connected. The correlation between spatial and molecular phenotype distances ($r = 0.068$) for all individuals sampled in drainages I, II, and III was also not significant ($t = 1.201$; $P = 0.885$).

4. Discussion

The level of polymorphism of *H. maximiliani* (37%), based on the RAPD molecular phenotypes observed in this study, was eightfold higher than that reported for Blanding's turtle, *Emydoidea blandingii* (4.5%), the only turtle species for which RAPD data are available (Mockford et al., 1999). Mockford et al. (1999) suggested that the low polymorphism detected in *E. blandingii* was not surprising since early studies with turtles had shown low rates of mutation in both mitochondrial (Avise et al., 1992; Bowen et al., 1993) and nuclear DNA (FitzSimmons et al., 1995). Mockford et al. (1999) also regarded the low levels of polymorphism in *E. blandingii* to be characteristic of Testudines. Estimates of polymorphism based on RAPD markers are sensitive to a certain level of subjectivity during the scoring of polymorphic bands (Grosberg et al., 1996), and this limits the effectiveness of comparisons across studies. Nevertheless, the data obtained here for *H. maximiliani* indicated that the levels of polymorphism in Testudines, as measured by RAPD molecular phenotypes, may not necessarily be as low as currently thought (see also Seddon et al., 1998). This finding is all the more remarkable considering the fine spatial scale of sampling of *H. maximiliani* which involved distances of about 5 linear (airline) km, whereas for *E. blandingii* the sampling scale ranged over 1000 km.

Table 2
Analysis of molecular variance (AMOVA) for 25 individuals of *Hydromedusa maximiliani* for drainage I in the Parque Estadual de Carlos Botelho, state of São Paulo, southeastern Brazil[a]

Source of variation	d.f.	SSD	Variance	% of total variance	*P*
Among sites	2	4.308	0.19184	21.99	0.00684
Within sites	22	14.972	0.68056	78.01	0.0001

[a] Populations within this drainage were pooled into three samples sites according to the spatial hierarchy of the main rivers and streams (see text). d.f., degrees of freedom; SSD, sums of squared deviations; *P*, probability of obtaining a larger variance by chance under the null hypothesis of zero variance (estimated from 1000 sampling permutations).

There were significant differences in the distribution of RAPD phenotypes in *H. maximiliani* between drainages and among rivers and streams within drainages. At the spatial scale of drainages there was also significant heterogeneity in the distribution of phenotypes between drainages I and III, which were located further apart (Fig. 1). The distribution of RAPD molecular phenotypes deviated from the null hypothesis of homogeneity at the scale of rivers and streams within a drainage. This result was found only for drainage I and may reflect the fact that this drainage had the largest sample size, which would increase the power of the statistical test. The AMOVA results for the three sampling areas within drainage I produced a significant estimate of Φ_{ST}, providing genetic evidence for population substructuring in *H. maximiliani* on very local spatial scales; that is, among rivers and streams within a drainage.

The evidence obtained for substructuring derived from molecular markers mirrors the evidence available from ecological studies. Direct estimates of dispersal for *H. maximiliani* based on mark-recapture data gathered during a 1993–1994 study (Souza and Abe, 1997a) showed that the movement of individuals was fairly restricted, with a mean daily displacement of 2 m, suggesting that turtles from each river and stream within a drainage could be structured as metapopulations (Hanski and Simberloff, 1997). In fact, individuals from which blood samples were taken for the present study in 1999 were found only a few meters from the site where they had been marked in 1993–1994.

For naturally subdivided populations such as aquatic organisms, microevolutionary processes are expected to drive the organization of genetic variability, and this may sometimes result in genetic differentiation on a local scale (Johnson & Black, 1991; Perault et al., 1997; Arnaud et al., 1999; Shaffer et al., 2000). For organisms with a sedentary nature and a low dispersal ability, such as snails, the genetic structuring of populations on small geographic scales is frequent (Johnson and Black, 1991; Arnauld et al., 1999). Although many factors, including population history and departures from the equilibrium

between drift and mutation rate, may confound the expected relationship between population structure and dispersal ability (see review in Bohonak, 1999), the general picture emerging from our study is that turtle behavior (sedentary behavior) is closely associated with habitat characteristics (a natural fragmented habitat, with mountain ridges, drainages, and rivers and streams within drainages). This relationship may limit gene flow and results in genetic structuring of the populations on a small spatial scale.

The data obtained from the AMOVA of molecular RADP markers presented here, combined with the ecological data and estimates of dispersal indicates that *H. maximiliani* populations are highly structured in close association with the topographic characteristics of this species habitat. Also it suggests that human-induced habitat fragmentation may not severely impact this species. Nevertheless fragmentation and loss of habitats certainly implies the extinction of local populations, and an understanding of patterns of population substructure combined with knowledge of ecological and natural history data is of fundamental importance for developing management strategies for insuring the long-term persistence of *H. maximiliani*.

The molecular and ecological findings reported here indicate that significant demographic and genetic processes are operating in *H. maximiliani* at the scale of rivers and streams within drainages. These results suggest that such local populations may be considered separate management units (MUs), in the sense that they harbor characteristic demographic processes and genetic variability (Moritz, 1994, 1999; Shaffer et al., 2000). The component populations of MUs are the natural units for population monitoring and demographic study and are, therefore, the target of short-term management (Moritz, 1994; Shaffer et al., 2000). Local population extinction can be counteracted by reintroductions from component populations of MUs (Moritz, 1994; Shaffer et al., 2000). In the case of the samples and geographic area we examined in this study, there was no evidence for isolation by distance in *H. maximiliani* and, therefore, the stocks for re-establishment efforts could be derived from any neighboring population.

Most research on turtles has examined molecular variation and population structure on very large spatial scales (reviewed in Walker and Avise, 1998; Avise, 2000). To our knowledge, this is the first study to document molecular variation on local spatial scales in a neotropical freshwater turtle. The pattern of population substructuring observed seems to be closely associated with the complex topography of the landscape matrix, typical of the habitat of the freshwater turtle *H. maximiliani*. Population phenomena and processes, such as the substructuring of populations on local spatial scales, are now regarded as important for process-oriented conservation approaches including species diversity, ecosystems and landscapes (Moritz, 1999; Poiani et al., 2000; Shaffer et al., 2000), and the results described here could contribute to the goal of conservation of *H. maximiliani*. The scale of sampling available for this study was limited both in geographic area and sample sizes, and most of the inference regarding conservation efforts was based on one drainage for which sample size was larger. Nevertheless, the data obtained provided evidence for potential MUs within *H. maximiliani*. Additional sampling should include not only more drainages, and rivers and streams within drainages, but also the entire geographic range of *H. maximiliani*; and mitochondrial and nuclear sequence markers should also be targeted. Such information will allow the investigation of genetic structure and genealogy at larger spatial scales of sampling to verify the validity of assertions made in this study regarding current population structure, and also to search for evidence of genealogical (i.e. historical) structure. The combination of evidence from both scales, that is, population and genealogy, should provide ecological and evolutionary perspectives for conservation efforts in *H. maximilliani*.

Acknowledgements

The authors thank E.L.A. Monteiro-Filho and L.C. Duarte for helpful comments and criticisms on the manuscript and S. Hyslop for reviewing the English of the manuscript. We thank one anonymous reviewer for a careful review of the manuscript. H.B. Shaffer provided thoughtful comments that greatly contributed to improve the clarity of the manuscript. We also are indebted to C.Y. Miyaki for assistance at the early stages of this work. We thank Instituto Florestal for logistic support and P.P. Soares and A. John for assistance in the field. J.R. Somera helped with the line drawings. We are deeply indebted to the Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (FAPESP: grant 00/00805-9) and Fundo de Apoio ao Ensino e à Pesquisa, Universidade Estadual de Campinas, São Paulo, Brazil (grant 0023/00) for funding our research. F.L.S. is the recipient of a post-doctoral fellowship from FAPESP (contract number 99/02761-0). S.F.R. is partially supported by a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

References

Arnaud, J.-F., Madec, L., Bellido, A., Guiller, A., 1999. Microspatial genetic structure in the land snail *Helix aspersa* (Gastropoda: Helicidae). Heredity 83, 110–119.

Avery, H.W., Vitt, L.J., 1984. How to get blood from a turtle. Copeia 209–210.

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge.

Avise, J.C., Bowen, B.W., Lamb, T., Meylan, A.B., Bermingham, E., 1992. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. Molecular Biology and Evolution 9, 457–473.

Bohonak, A.J., 1999. Dispersal, gene flow, and population structure. The Quarterly Reveiw of Biology 74, 21–45.

Bowen, B.W., Nelson, W.S., Avise, J.C., 1993. A molecular phylogeny for marine turtles: trait mapping, rate assessment, and conservation relevance. Proceedings of the National Academy of Sciences of the USA 90, 5574–5577.

Bruford, M.W., Hanotte, O., Brookfield, J.F.Y., Burke, T., 1992. Single-locus and multilocus DNA fingerprinting. In: Hoelzel, A.R. (Ed.), Molecular Genetic Analysis in Conservation. IRL Press, Oxford, pp. 225–270.

Cabrera, M., 1998. Las Tortugas Continentales de Sudamérica Austral. BR Cópias, Córdoba, Rep, Argentina.

Caughley, G., 1994. Directions in conservation biology. Journal of Animal Ecology 63, 215–244.

Chiarello, A.G., 1999. Effects of fragmentation of the Atlantic forest on mammal communities in south-eastern Brazil. Biological Conservation 89, 71–82.

Congdon, J.D., Dunham, A.E., van Loben Sels, R.C., 1994. Demographics of common snapping turtles (*Chelydra serpentina*): implications for conservation and management of long-lived organisms. American Zoologist 34, 397–408.

Cooper, M.L., 2000. Random amplified polymorphic DNA analysis of southern brown bandicoot (*Isoodon obesulus*) populations in Western Australia reveals genetic differentiation related to environmental variables. Molecular Ecology 9, 469–479.

Dean, W., 1996. A Ferro e Fogo: a História e a Devastação da Mata Atlântica Brasileira. Cia. das Letras, São Paulo.

Dunham, A.E., Grant, B.W., Overall, K.L., 1989. The interface between biophysical ecology and population ecology of terrestrial vertebrate ectotherms. Physiological Zoology 62, 335–355.

Ernst, C.H., Barbour, R.W., 1989. Turtles of the World. Smithsonian Institution Press, Washington, DC.

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479–491.

FitzSimmons, N.N., Moritz, C., Moore, S.S., 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Molecular Biology and Evolution 12, 432–440.

Gibbs, H.L., Prior, K.A., Weatherhead, P.J., 1994. Genetic analysis of populations of threatened snake species using RAPD markers. Molecular Ecology 3, 329–337.

Gibbs, J.P., 1998. Genetic structure of redback salamander *Plethodon cinereus* populations in continuous and fragmented forests. Biological Conservation 86, 77–81.

Grosberg, R.K., Levitan, D.R., Cameron, B.D., 1996. Characterization of genetic structure and genealogies using RAPD-PCR markers: a random primer for the novice and nervous. In: Ferraris, J.D., Palumbi, S.R. (Eds.), Molecular Zoology: Advances, Strategies and Protocols. Wiley-Liss, New York, pp. 67–100.

Haig, S.M., Rhymer, J.M., Heckel, D.G., 1994. Population differentiation in randomly amplified polymorphic DNA of red-cockaded woodpeckers *Picoides borealis*. Molecular Ecology 3, 581–595.

Hanski, I., 1998. Connecting the parameters of local extinction and metapopulation dynamics. Oikos 83, 390–396.

Hanski, I.A., Simberloff, D., 1997. The metapopulation approach, its history, conceptual domain, and application to conservation. In: I.A. Hanski, M.E. Gilpin (Eds.), Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, pp. 5–26.

Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. Annals of Mathematical Statistics 19, 293–325.

Iverson, J.B., 1992. A revised Checklist with Distribution Maps of the Turtles of the World. Privately printed, Richmond, Indiana, USA.

Johnson, M.S., Black, R., 1991. Genetic subdivision of the intertidal snail *Bembicium vittatum* (Gastropoda: Littorinidae) varies with habitat in the Houtman Abrolhos Islands, Western Australia. Heredity 67, 205–213.

Lacy, R.C., 1993. Vortex: a computer simulation model for population viability analysis. Wildlife Research 20, 45–65.

Legende, S., Clobert, J., Møller, A.P., Sorci, G., 1999. Demographic stochasticity and social mating in the process of extinction of small populations: the case of passerines introduced to New Zeland. The American Naturalist 153, 449–463.

Mantel, N.A., 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27, 209–220.

McCarthy, M.A., 1996. Extinction dynamics of the helmeted honeyeater: effects of demography, stochasticity, inbreeding and spatial structure. Ecological Modelling 85, 151–163.

Miyaki, C.Y., Matioli, S.R., Burke, T., Wajntal, A., 1998. Parrot evolution and paleogeographical events: Mitochondrial DNA evidence. Molecular Biology and Evolution 15, 544–551.

Mockford, S.W., Snyder, M., Herman, T.B., 1999. A preliminary examination of genetic variation in a peripheral population of Blanding's turtle, *Emydoidea blandingii*. Molecular Ecology 8, 323–327.

Moritz, C., 1994. Defining 'evolutionary significant units' for conservation. Trends in Ecology and Evolution 9, 373–375.

Moritz, C., 1999. Conservation units and translocations: strategies for conserving evolutionary processes. Heredity 130, 217–228.

Nichols, W.A., Resendiz, A., Seminoff, J.A., Resendiz, B., 2000. Transpacific migration of a loggerhead turtle monitored by satellite telemetry. Bulletin of Marine Science 67, 937–947.

Perault, D.R., Wolf, P.G., Edwards Jr, T.C., 1997. Hierarchical analysis of genetic partitioning by *Tamias minimus* and *T. umbrinus*. Journal of Mammalogy 78, 134–145.

Pfeifer, R.M., Carvalho, W.A., Silva, D.A., Rossi, M., Mendicino, L.F., 1986. Levantamento semi-detalhado dos solos do Parque Estadual de Carlos Botelho, SP. Boletim Técnico do Instituto Florestal, São Paulo 40, 75–109.

Pinheiro, H.P., 1997. Modelling variability in the HIV genome. PhD Thesis, University of North Carolina.

Poiani, K.A., Richter, B.D., Anderson, M.G., Richter, H.E., 2000. Biodiversity conservation at multiple scales: functional sites, landscapes, and networks. BioScience 50, 133–146.

Primack, R.B., 1998. Essentials of Conservation Biology. Sinauer, Sunderland.

Pritchard, P.C.H., Trebbau, P., 1984. The Turtles of Venezuela. Society for the Study of Amphibians and Reptiles, Athens, Ohio.

Rohlf, F.J., 1994. NTSYS-pc—Numerical Taxonomy and Multivariate Analysis System, version 1.80. Exeter Software, Setauket, NY.

SAS/STAT., 1989. User's Guide, version 6, Fourth Edition, Vol. 1, Cary, NC: SAS Institute Inc.

Schneider, S., Kueffer, J.M., Roessli, D., Excoffier, L., 1997. Arlequin ver. 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneve, Switzerland.

Seddon, J.M., Baverstock, P.R., Georges, A., 1998. The rate of mitochondrial 12S RNA gene evolution is similar in freshwater turtles and marsupials. Journal of Molecular Evolution 46, 460–464.

Shaffer, H.B., Fellers, G.M., Magee, A., Voss, S.R., 2000. The genetics of amphibian declines: population substructure and molecular differentiation in the Yosemite Toad, *Bufo canorus* (Anura, Bufonidae) based on single-strand conformation polymorphism analysis (SSCP) and mitochondrial DNA sequence data. Molecular Ecology 9, 245–257.

Sokal, R.R., Rohlf, F.J., 1981. Biometry: Principle and Practice of Statistics in Biological Research, 2nd edn. W. H. Freeman, New York.

Souza, F.L., 1995. História natural do cágado*Hydromedusa maximiliani* (Mikan 1820) no Parque Estadual de Carlos Botelho, SP, região de Mata Atlântica (Reptilia, Testudines, Chelidae). MS Thesis, Universidade Estadual Paulista, Rio Claro, SP, Brasil.

Souza, F.L., Abe, A.S., 1995. Observations on feeding habits of *Hydromedusa maximiliani* (Testudines: Chelidae) in Southeastern Brazil. Chelonian Conservation and Biology 1, 320–322.

Souza, F.L., Abe, A.S., 1997a. Population structure, activity, and conservation of the Neotropical freshwater turtle, *Hydromedusa maximiliani*, in Brazil. Chelonian Conservation and Biology 2, 521–525.

Souza, F.L., Abe, A.S., 1997b. Seasonal variation in the feeding habits of *Hydromedusa maximiliani* (Testudines, Chelidae). Boletin de la Asociacíon Herpetológica Española 8, 17–20.

Souza, F.L., Abe, A.S., 1998. Resource partitioning by the neotropical freshwater turtle, *Hydromedusa maximiliani*. Journal of Herpetology 32, 106–112.

Tabarelli, M., Mantovani, W., Peres, C.A., 1999. Effect of habitat fragmentation on plant guild structure in the montane Atlantic forest of southeastern Brazil. Biological Conservation 91, 119–127.

Templeton, A.R., Shaw, K., Routman, E., Davis, S.K., 1990. The genetic consequences of habitat fragmentation. Annals of the Missouri Botanical Garden 77, 13–27.

Valenzuela, N., 2001. Genetic differentiation among nesting beaches in the highly migratory giant river turtle (*Podocnemis expansa*) from Colombia. Herpetologica 57, 48–57.

Veloso, H.P., Rangel-Filho, A.L.R., Lima, J.C.A., 1991. Classificação da Vegetação Brasileira Adptada a um Sistema Universal. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro.

Vucetich, L.M., Vucetich, J.A., Joshi, C.P., Waite, T.A., Peterson, R.O., 2001. Genetic (RAPD) diversity in *Peromyscus maniculatus* populations in a naturally fragmented landscape. Molecular Ecology 10, 35–40.

Walker, D., Avise, J.C., 1998. Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. Annual Review of Ecology and Systematics 29, 23–58.

Whitmore, T.C., 1990. An Introduction to Tropical Rain Forests. Oxford University Press, New York.

Williams, J.G.K., Hanafey, M.K., Rafalski, J.A., Tingey, S.V., 1993. Genetic analysis using random amplified polymorphic DNA markers. Methods in Enzymology 218, 704–740.

Genetics and Molecular Biology, 25, 2, 151-155 (2002)

www.sbg.org.br

Estimating dispersal and gene flow in the neotropical freshwater turtle *Hydromedusa maximiliani* (Chelidae) by combining ecological and genetic methods

Franco L. Souza[1], Anderson F. Cunha[2], Marcos A. Oliveira[2], Gonçalo A.G. Pereira[2] and Sérgio F. dos Reis[3]

[1]*Departamento de Biologia, Centro de Ciências Biológicas e da Saúde, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brazil.*
[2]*Departamento de Genética, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil.*
[3]*Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil.*

Abstract

Hydromedusa maximiliani is a vulnerable neotropical freshwater turtle endemic to mountainous regions of the Atlantic rainforest in southeastern Brazil. Random amplified polymorphic DNA (RAPD) was used to estimate the gene flow and dispersal for individuals inhabiting rivers and streams within a drainage. Nine primers generated 27 scoreable bands, of which 9 (33%) were polymorphic and produced 12 RAPD phenotypes. The gene flow estimates (*Nm*) among turtles inhabiting different rivers and streams were variable, ranging from 0.09 to 3.00 (mean: 0.60). For some loci, the rates of gene flow could offset population differentiation ($Nm > 1$), whereas for others random genetic drift could result in population divergence ($Nm < 1$). Since the genetic variation of this turtle seems to be structured according to the natural hierarchical system of rivers and streams within drainages, management programs involving translocations between different regions across the geographical range of *H. maximiliani* should be viewed with caution.

Key words: Hydromedusa maximiliani, dispersal, gene flow, conservation.

Received: April 2, 2002; accepted: June 5, 2002.

Introduction

The neotropical freshwater turtle *Hydromedusa maximiliani* (Chelidae) is endemic to the Atlantic forest of the coastal region of southeastern Brazil, with a geographical distribution ranging from the State of Espírito Santo to the State of São Paulo (Ernst and Barbour, 1989; Iverson 1992). Throughout its range, the distribution of *H. maximiliani* is disjunct, occurring in isolation at elevations above 600 m. Locally, its habitat is topologically complex, with sequences of ridges and valleys each drained by river and stream systems (Souza, 1995).

Capture-recapture studies have shown that these turtles have limited dispersal ability, with a mean daily displacement of 2 m (Souza 1995; Souza and Abe, 1997a). Dispersal over longer distances is apparently limited to the rainy season, when temporary or intermittent water systems are connected with the main watercourses (Souza and Abe, 1997a). The observed limited dispersal suggests that each *H. maximiliani* population inhabits a specific river within a given drainage system, and this hypothesis was addressed in a preliminary work on the population genetic structure of the species, which indicated a substantial partitioning of molecular variation across rivers and streams inhabited by this turtle (Souza *et al.*, 2002). These findings of population structure on a small spatial scale defined by the complex topographical features of the species' habitat, a region topologically complex, with ridges and valleys drained by numerous rivers and streams (Pfeifer *et al.*, 1986; Souza and Abe 1998), could imply that there is little or no gene flow among turtles inhabiting different water courses or distinct drainage. In this study, we used random amplified polymorphic DNA (RAPD) markers (Williams *et al.*, 1993) to assess levels of gene flow among local *H. maximiliani* populations, and briefly discuss the conservation and management implications of our findings.

Send correspondence to Sérgio Furtado dos Reis. Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6109, 13083-970 Campinas, São Paulo, Brazil. E-mail: sfreis@unicamp.br.

Material and Methods

Field work was conducted from November 1998 to November 1999 at the Parque Estadual de Carlos Botelho (PECB), state of São Paulo, southeastern Brazil (24°00'-24°15'S, 47°45'-48°10'W). The PECB is a protected reserve that encompasses over 37,000 ha of intact tropical montane rain forest typical of southeastern Brazil (Whitmore, 1990; Veloso *et al.*, 1991). In this study, an area of approximately 500 ha containing eight rivers and streams (1-8) was surveyed (Figure 1). This 500 ha area was the same as used in previous studies of this species' natural history [see Souza and Abe (1995, 1997a,b, 1998) for a more detailed description of the area]. Three sites were defined within these drainage based on the spatial hierarchy of the main rivers and their tributaries: site A (which included river 1), site B (which included rivers 2, 3, 4, and 5), and site C (which included rivers 6, 7, and 8) (Figure 1).

Turtles ($n = 25$) were hand-caught and 200-300 µL of blood was drawn from the scapula vein/brachial artery (Avery and Vitt, 1984) using a 26-gauge needle and a 1 mL syringe. The blood samples were immediately preserved in 1 mL of absolute ethanol (Miyaki *et al.*, 1997) in plastic vials and stored at room temperature. The turtles were released at the point of capture after blood sample collection.

Genomic DNA was extracted from the blood samples by two successive organic extractions with phenol:chloroform:isoamyl alcohol as outlined by Bruford *et al.* (1992) and Miyaki *et al.* (1998), and then precipitated with 1/10 volume of 1 M sodium acetate (pH 5.3) and two volumes of 100% ethanol. The quality of the extracted DNA was evaluated in agarose gels (0.7%) stained with ethidium bromide and quantified by comparison with DNA standards run in the same gel. DNA was diluted to a working concentration of approximately 0.6 ng/µL.

Eighty primers (Operon Technologies, Inc.; Alameda, California, EUA) were initially screened for consistently reproducible and scoreable amplified bands. Nine primers that met these criteria were used to analyze the DNA samples from the 25 individuals captured in the studied area. The polymerase chain reaction (PCR) was performed in a Perkin-Elmer GeneAmp™ PCR system 9700 with a total volume of 12.5 µL containing 10 mM Tris HCl (pH 8.4), 50 mM KCl, 3.5 mM $MgCl_2$, 1 µL of each dNTP, 2 µL of primer, 0.3 µL *Taq* polymerase, 1.2 ng of genomic DNA and sterile water. Negative controls in which water was substituted for DNA were run to check for the possibility of contamination. Reproducibility was gauged by comparing duplicate reactions, which were usually adjacent to one another in the thermocycler, and products were run side-by-side on the same gel. The reaction conditions involved initial denaturation of DNA for 2 min at 94 °C, 39 cycles of 1 min denaturation at 94 °C, 1 min annealing at 40 °C, 2 min extension at 72 °C, and one 5 min cycle at 72 °C for final extension. The amplification products were

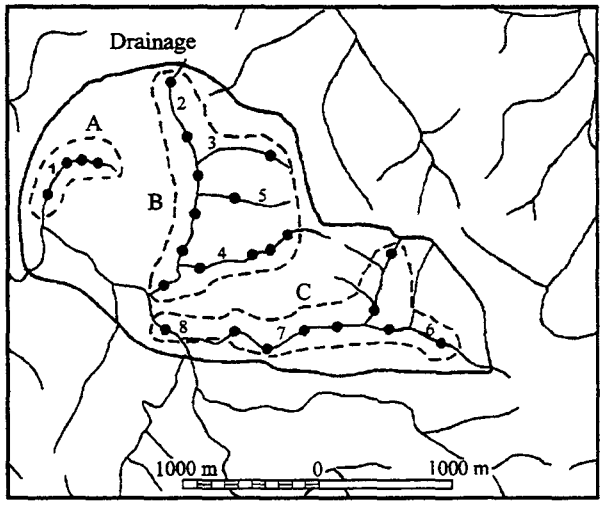


Figure 1 - Map of the study site at Parque Estadual de Carlos Botelho, state of São Paulo, southeastern Brazil. The sample drainage for *Hydromedusa maximiliani* is showed as a heavy line with numbers 1-8 indicating rivers and streams surveyed. Dotted lines show the three sample sites (A, B, C) defined according to the spatial hierarchy of the main rivers and their tributaries Dots indicate individual collection sites.

separated on 1.3% agarose gels stained with ethidium bromide, run in buffer 1X TAE at a constant voltage of 80 V for 5 h. Monochrome photographic negatives were taken of the gels and the individual profiles were scored by two of the authors (FLS and AFC) for the presence/absence of fragments for each primer (see Souza *et al.*, 2002, for a full description of these methods).

Allele frequencies were estimated for standard genetic analysis of population structure. Most (~90%) alleles amplified by arbitrarily primed PCR segregate as dominant markers. Since these RAPD alleles are revealed as the presence or absence of a band, it is not generally possible to distinguish heterozygous individuals from those homozygous for the dominant allele at such loci because both have the "band present" phenotype (Ferreira and Grattapaglia, 1998). We assumed that all loci considered in our analyses met this criterion and that the genotypes were in Hardy-Weinberg equilibrium. Allele frequencies were obtained using the asymptotically unbiased estimator, $\hat{q}$, derived by Lynch and Milligan (1994), as follows

$$\hat{q} = \sqrt{\hat{x}}\left(1 - \frac{\mathrm{Var}(\hat{x})}{8\hat{x}^2}\right)$$

where $\hat{x}$ is the frequency of the null allele calculated as the square root of the frequency of the null phenotype (*i.e.*, absent band).

The analysis of molecular variance, AMOVA (Excoffier *et al.*, 1992) was used to measure the variation in allelic frequencies for each locus among populations of turtles inhabiting different rivers and streams from the sampled drainage. Gene flow (*Nm*) was estimated from *F*-statistics, F_{ST}, as a measure of the genetic interaction

among populations, indicating the number of immigrants per population per generation (Slatkin, 1985, 1987). We used the formula $F_{ST} = 1/(4Nm + 1)$, where N is the local population size and m is the average rate of immigration. For this estimate, we assumed neutrality, negligible mutation and a stepping-stone population structure model (Kimura and Weiss, 1964). This assumption was crucial because the relation of F_{ST} to underlying microevolutionary parameters changes with different models of population structure (Slatkin, 1985, 1987). Contingency chi-square values were calculated to determine whether F_{ST} estimates varied from zero (significant population differentiation), using the formula $\chi^2 = 2N F_{ST}(k\text{-}1)$, where N is the total sample size, and k is the number of alleles; df = $(k\text{-}1)(m\text{-}1)$, where m is the number of samples (in this case, the number of sites within sampled drainage) (Johnson and Black, 1991).

Table II - Among population differentiation (F_{ST}) and gene flow estimates (Nm) of RAPD allele frequencies for nine variable loci (L) from *Hydromedusa maximiliani* specimens inhabiting a drainage in the Parque Estadual de Carlos Botelho, state of São Paulo, southeastern Brazil. d.f.: degrees of freedom.

Locus	F_{ST}	χ^2	d.f.	p	Nm
L1	0.1411	7.06	2	0.05	1.52
L2	0.4620	23.10	2	0.001	0.29
L3	0.3858	19.29	2	0.001	0.39
L4	0.7428	37.14	2	0.001	0.09
L5	0.1864	9.32	2	0.01	1.09
L7	0.0770	3.85	2	0.05	3.00
L8	0.1864	9.32	2	0.01	1.09
L9	0.1864	9.32	2	0.01	1.09
L10	0.2726	13.63	2	0.01	0.67
Total	0.2934	132.03	18	0.0001	0.60

Results

Nine of the 80 primers used yielded band patterns that were clear and could be scored with confidence. These primers produced 27 potentially scoreable bands (range 1-6; mean 3), of which nine bands (33%) were polymorphic and produced 12 RAPD phenotypes. The estimated allele frequencies at RAPD-encoding loci from these nine polymorphic bands varied among rivers and streams, indicating that most loci were spatially structured (Table I) and endemic to some rivers and streams. Locus 2 was found only for site A (river 1), which was the river located farthest west in this drainage. Two unique alleles were found for turtles inhabiting river 4 at site B (loci 5 and 9), while one unique allele (locus 8) was found for individuals from river 2 (site B). One allele, locus 1, was the most widespread allele and was detected at all three sites, including rivers 2, 4, 7, and 8 (Table I).

The estimated F_{ST} for each locus ranged from 0.077 to 0.743 (Table II), and only one locus (locus 7) showed no significant difference among populations ($p > 0.05$). The overall differentiation ($F_{ST} = 0.293$) among populations of turtles inhabiting different rivers and streams from the sampled drainage was significant ($\chi^2_{18} = 132.03$; $p < 0.0001$), indicating variation among populations. The average Nm estimate was 0.60 (range = 0.09-3.00).

Discussion

The combination of ecological and biogeographical processes is an important mechanism for shaping the genetic structure of populations. Thus, life-traits (mating systems, dispersal ability), historical events (fragmentation, range expansion, colonization), and landscape matrix (mountain ridges, watersheds) may provide the background for understanding the geographic structure of genetic variation among populations (Templeton *et al.*, 1995).

Nm estimates of more than one immigrant into the average deme per generation, as calculated for five loci (loci 1, 5, 7, 8, and 9), indicated that gene flow among populations inhabiting different rivers and streams was sufficient to deter population differentiation if they were at equilibrium between migration and random genetic drift (Wright,

Table I - Band frequencies (locus: L) among nine oligonucleotide primers for *H. maximiliani* sampled from eight rivers and streams which formed the three sites (A, B, and C) of a drainage in the Parque Estadual de Carlos Botelho. The values indicate the presence frequency of variable bands (n = sample size).

Site	River	L1	L2	L3	L4	L5	L7	L8	L9	L10
A ($n = 4$)	1 ($n = 4$)	1.00	0.75	1.00	1.00	1.00	0.00	1.00	1.00	1.00
B ($n = 12$)	2 ($n = 6$)	0.83	1.00	1.00	1.00	1.00	0.00	0.83	1.00	1.00
	3 ($n = 1$)	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00
	4 ($n = 4$)	0.50	1.00	1.00	1.00	0.75	0.00	1.00	0.75	0.75
	5 ($n = 1$)	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00
C ($n = 9$)	6 ($n = 2$)	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00
	7 ($n = 6$)	0.67	1.00	0.67	0.17	1.00	0.17	1.00	1.00	1.00
	8 ($n = 1$)	0.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00

1931; Slatkin, 1985, 1987). In contrast, *Nm* values for four loci (loci 2, 3, 4, and 10) were less than one, implying that the differentiation among populations could have occurred through random genetic drift alone (Slatkin, 1985, 1987). Overall, the *Nm* estimate of 0.589 indicated that the *H. maximiliani* populations inhabit different rivers.

Population structure bears an inverse relationship with the dispersal capability of organisms (Hartl and Clark, 1997), and the observed pattern of *H. maximiliani* population structure on a local spatial scale (Souza *et al.*, 2002) is consistent with the known dispersal behavior of the species. Direct and indirect methods for estimating gene flow (Slatkin, 1987) in this species yielded similar results. Estimates of dispersal for *H. maximiliani* based on mark-recapture data (direct method) gathered during a 1993-1994 study (Souza and Abe, 1997b) showed that the movement of individuals was fairly restricted, with a mean daily displacement of two meters, suggesting that turtles from each river and stream within a drainage could be structured in as a metapopulation (Hanski and Simberloff, 1997). Indeed, individuals from which blood samples were taken for the present study in 1999 were found only a few meters from the site where they had been marked in 1993-1994.

The hypothesis of a metapopulation structure was confirmed by indirect (genetic) methods. The fact that several alleles were unique for rivers and streams within a drainage indicated that, because of the turtles' limited dispersal, mating must be restricted to related individuals, which would facilitate the development of a local genetic structure. Furthermore, the molecular markers produced high F_{ST} values and low *Nm* values, indicating low gene flow among turtles inhabiting different rivers and streams. Templeton *et al.* (1990) found that mitochondrial DNA haplotypes in strictly aquatic salamanders were partitioned according to rivers within drainage, in a similar manner to the partitioning of RAPD variation for *H. maximiliani*. For organisms with a sedentary nature and low dispersal ability, such as snails, the genetic structuring of populations on small geographic scales is frequently detected (Johnson and Black, 1991; Arnaud *et al.*, 1999, 2001), and could have originated from limited gene flow among populations.

Additional sampling of drainages and rivers and streams within drainage, as well as the use of mitochondrial and nuclear sequence markers will be necessary to estimate effective population sizes, population genealogies and other aspects fundamental to *H. maximiliani* conservation. These genetic analyses are important for the protection and management of this turtle species because such information could be useful in determining options for translocations. The genetic variation in this turtle species is structured according to the natural hierarchical system of rivers and streams within drainage, and limited gene flow is detected among populations from different rivers and streams. Thus, each watershed may harbor endemic populations, which suggests a strong degree of genetic structure and differentiation in the geographical range of the species.

Acknowledgements

The authors thank E.L.A. Monteiro-Filho and L.C. Duarte for helpful comments and criticisms on the manuscript, S. Hyslop for reviewing the English of the manuscript, and C.Y. Miyaki for assistance in the early stages of this study. The Instituto Florestal provided logistic support and P.P. Soares and A. John provided assistance in the field. J.R. Somera helped with the line drawings. This study was funded by the Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (FAPESP, grant 00/00805-9) and Fundo de Apoio ao Ensino e à Pesquisa, Universidade Estadual de Campinas, São Paulo, Brazil (FAEP, grant 0023/00). FLS was the recipient of a post-doctoral fellowship from FAPESP (contract number 99/02761-0). SFR is partially supported by a research fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

References

Arnaud JF, Madec L, Bellido A and Guiller A (1999) Microspatial genetic structure in the land snail *Helix aspersa* (Gastropoda: Helicidae). Heredity 83:110-119.

Arnaud JF, Madec L, Guiller A and Bellido A (2001) Spatial analysis of allozyme and microsatellite DNA polymorphism in the land snail *Helix aspersa* (Gastropoda: Helicidae). Mol. Ecol. 10:1563-1576.

Avery HW and Vitt LJ (1984) How to get blood from a turtle. Copeia 1984:209-210.

Bruford MW, Hanotte O, Brookfield JFY and Burke T (1992) Single-locus and multilocus DNA fingerprinting. In: Hoelzel AR (ed) Molecular Genetic Analysis in Conservation, Oxford, pp 225-270.

Ernst CH and Barbour RW (1989) Turtles of the World. Smithsonian Institution Press, Washington, D.C., 313pp.

Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Ferreira ME and Grattapaglia D (1998) Introdução ao Uso de Marcadores Moleculares em Análise Genética. Embrapa, Brasília, 220 pp.

Hanski I and Simberloff D (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. In: Hanski IA and Gilpin ME (eds) Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, New York, pp 5-26.

Hartl D and Clark AG (1997) Principles of Population Genetics. Sinauer, Sunderland, 389 pp.

Iverson JB (1992) A Revised Checklist with Distribution Maps of the Turtles of the World. Privately printed, Richmond, 363pp.

Johnson MS and Black R (1991) Genetic subdivision of the intertidal snail *Bembicium vittatum* (Gastropoda: Littorinidae) varies with habitat in the Houtman Abrolhos Islands, Western Australia. Heredity 67:205-213.

Kimura M and Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561-576.

Lynch M and Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Mol. Ecol. 3:91-99.

Myiaki CY, Pereira SL, Biasia I and Wajntal A (1997) DNA fingerprint applied to parrot captive breeding programs. Ararajuba 5:127-133.

Miyaki CY, Matioli SR, Burke T and Wajntal A (1998) Parrot evolution and paleogeographical events: mitochondrial DNA evidence. Mol. Biol. Evol. 15:544-551.

Pfeifer RM, Carvalho WA, Silva DA, Rossi M and Mendicino LF (1986) Levantamento semi-detalhado dos solos do Parque Estadual de Carlos Botelho, SP. Bol. Tec. I. F., São Paulo, 40:75-109.

Slatkin M (1985) Gene flow in natural populations. Ann. Rev. Ecol. Syst. 16:393-430.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787-792.

Souza FL (1995) História natural do cágado *Hydromedusa maximiliani* (Mikan 1820) no Parque Estadual de Carlos Botelho, SP, região de Mata Atlântica (Reptilia, Testudines, Chelidae). MS Thesis, Universidade Estadual Paulista, Rio Claro.

Souza FL and Abe AS (1997a) Population structure, activity, and conservation of the neotropical freshwater turtle, *Hydromedusa maximiliani*, in Brazil. Chel. Conserv. Biol. 2:521-525.

Souza FL and Abe AS (1997b) Seasonal variation in the feeding habits of *Hydromedusa maximiliani* (Testudines, Chelidae). Bol. Asoc. Herpetol. Esp. 8:17-20.

Souza FL and Abe AS (1998) Resource partitioning by the neotropical freshwater turtle, *Hydromedusa maximiliani*. J. Herpetol. 32:106-112.

Souza FL, Cunha AF, Oliveira MA, Pereira GAG, Pinheiro HP and dos Reis SF (2002) Partitioning of molecular variation at local spatial scales in the vulnerable neotropical freshwater turtle, *Hydromedusa maximiliani* (Testudines, Chelidae): implications for conservation of aquatic organisms in natural hierarchical systems. Biol. Conserv. 104:119-126.

Templeton AR, Shaw K, Routman E and Davis SK (1990) The genetic consequences of habitat fragmentation. Ann. Miss. Bot. Gard. 77:13-27.

Templeton AR, Routman E and Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics 140:767-782.

Veloso HP, Rangel-Filho ALR and Lima JCA (1991) Classificação da Vegetação Brasileira Adaptada a um Sistema Universal. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 123 pp.

Whitmore TC (1990) An Introduction to Tropical Rain Forests. Oxford University Press, Oxford, 296 pp.

Williams JGK, Hanafey MK, Rafalski JA and Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. Met. Enzymol. 218:704-740.

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97-159.

JOURNAL OF HERPETOLOGY

Preliminary phylogeographic analysis of the neotropical freshwater turtle *Hydromedusa maximiliani* (Chelidae)

Franco L. Souza[1,3], Anderson F. Cunha[2], Marcos A. Oliveira[2], Gonçalo A. G. Pereira[2], and Sérgio F. dos Reis[1]

[1] Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6109, 13083-970 Campinas, São Paulo, Brazil

[2] Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6109, 13083-970 Campinas, São Paulo, Brazil

Key words: Chelidae, control region, *Hydromedusa maximiliani*, phylogeography, turtle

[3] Present Address: Departamento de Biologia, Centro de Ciências Biológicas e da Saúde, Universidade Federal de Mato Grosso do Sul, 79070-900, Campo Grande, Mato Grosso do Sul, Brazil

ABSTRACT.– Phylogeography of *Hydromedusa maximiliani* was assessed by sequencing a 416 bp fragment from the mtDNA control region. Based on fragment polymorphism, 18 haplotypes were detected among individuals from four populations throughout southeastern Brazil. Two major geographical groups could be distinguished, corresponding to a western population, represented by individuals from São Paulo state and an eastern population, represented by individuals from Rio de Janeiro and Minas Gerais states. Divergence dates of 8-16 million years ago for northern and southern *H. maximiliani* populations the orographical activities resulting in Serra do Mar and the Serra da Mantiqueira mountain uplifts during Pliocene and Miocene.

Hydromedusa maximiliani is endemic to southeastern Brazil. Its typical habitat is cold and clear waters of shallow small rivers and streams found throughout pristine mountainous regions of Atlantic rainforest. This habitat is becoming fragmented by extensive human activity (Dean, 1996; Chiarello, 1999; Tabarelli et al., 1999) and this fragmentation may affect gene flow among turtle populations. Natural history records from capture-recapture studies and genetic surveys have shown that population structure in this species can be detected on a small scale defined by drainages in the complex landscape suitable matrix of habitat (Souza, 1995; Souza and Abe, 1997, 1998; Souza et al., 2002a,b), and local populations can be considered management units (Souza et al., 2002a,b) that harbor local demographic processes and genetic variability (Moritz, 1994, 1999). Since genetic variation in this turtle is structured according to the natural hierarchy

of rivers and streams within drainages (Souza et al., 2002a,b), each watershed may harbor endemic populations.

Turtles have limited dispersal capability (except for marine and some freshwater species; Pritchard and Trebbau, 1994; Sites et al., 1999; Nichols et al., 2000; Valenzuela, 2001), and physical (i.e., large rivers, mountains, valleys) or climatic (i.e., temperature) barriers represent important factors in determining the population genetic structure of freshwater and terrestrial chelonians (Lamb et al., 1989; Schroth et al., 1996; Walker and Avise, 1998; Walker et al., 1997; Roman et al., 1999; Avise, 2000). The geographical range of *Hydromedusa maximiliani* encompasses two of the most notable mountain complexes in southeastern Brazil, the Serra do Mar and the Serra da Mantiqueira. These geographical components play an important role in shaping population genetic structure of plants and animals in this region (Vanzolini, 1973; Dixon, 1979; Simpson, 1979; Cardoso et al., 2000; Ditchfield, 2000; Lara and Patton, 2000; Moritz et al., 2000). Because metapopulation structure over a small spatial scale has been suggested by direct (mark-recapture) and indirect (genetic) methods for *H. maximiliani* (Souza, 1995; Souza and Abe, 1997, 1998; Souza et al., 2002a,b), the complex landscape found throughout the species' distribution, together with behavioral characteristics (e.g. sedentary behavior, habitat requirements), may favor maintenance of a phylogeographic pattern in the genetic structure of this species.

We present preliminary results regarding phylogeographic structure in *Hydromedusa maximiliani* based on partial sequencing of the mtDNA control region from specimens sampled throughout the species' geographical range. Genetic data on *H. maximiliani* populations from different geographic areas are useful to investigations of

genetic structure and genealogy on a larger spatial scale. Combined evidence from population and genetic studies provides ecological and evolutionary perspective to conservation efforts in *H. maximiliani*.

MATERIAL AND METHODS

Blood samples (200–300 μL; Avery and Vitt, 1984) were taken from turtles at four localities: Parque Estadual de Carlos Botelho, São Paulo state (PECB; 24°00' to 24°15'S, 47°45 to 48°10'W; $n = 48$), Nova Friburgo, Rio de Janeiro state (NF; 22°19'S, 42°02'W; $n = 4$), Teresópolis, Rio de Janeiro state (TERES; 22°27'S, 42°28'W; $n = 1$), and Juiz de Fora, Minas Gerais state (JF; 21°42'S, 43°21'W; $n = 1$) (Fig. 1). These localities represent Serra do Mar (PECB, NF, and TERES) and Serra da Mantiqueira (JF) populations. Samples of *Hydromedusa maximiliani* were obtained under permit number 41.224/99 issued by Secretaria do Meio Ambiente, Brazil.

Genomic DNA was initially extracted from blood following procedures outlined by Bruford et al. (1992) and Miyaki et al. (1998). Quality of extracted DNA was evaluated in agarose gels (0.7%) stained with ethidium bromide, and quantified by comparison with DNA standards run in the same gel. Purified DNA was used in an initial polymerase chain reaction (PCR) for the mtDNA control region under the following thermal conditions: 98 °C for 3 min for initial denature, 36 cycles of 94 °C for 45 s for denature, 53 °C annealing for 45 s, and 72 °C extension for 90 s, followed by 72 °C for 90 s for final extension. PCR was conducted (GeneAmp™ PCR System 9700; Perkin Elmer) in a final volume of 12.5 μL containing 20 ng DNA, 1.5 μL 10X PCR buffer, 0.5 μL 50 mM $MgCl_2$, 1.25 mM dNTPs, 0.9 μM primers, and 0.175 μL *Taq* polymerase (Platinum™ *Taq*; GibcoBRL®).

Primers were developed based on sequence alignments of the mitochondrial genome from *Podocnemis expansa*, *Chelus fimbriata*, *Pelusios williamsi*, *Chrysemys picta*, *Chelonia mydas* and *Dermochelys coriacea*. The forward primer (*cytb* 5'-CTACACMYMTCMAAACAAC-3') and the reverse primer (*12S1* 5'-GGACCAAATCTTTGTGTTTTTG-3') amplified an approximately 1,400 bp fragment of the mitochondrial region encompassing cytochrome *b*, 12S, and Thr-proline genes, as well as the D loop region. The size of the amplified fragment was checked by electrophoresis in 0.7% agarose minigels stained with ethidium bromide. PCR product was run on a 1.2% agarose 1X TAE gel and the target DNA fragment was then excised. To purify double stranded product, a PCR purification kit (Concert™ Gel Extraction Systems; GibcoBRL®) was used and, depending on yield, 2 – 5 μL of purified product served as template for single-stranded PCR with the Prism cycle sequencing kit (Applied Biosystems) and *12S1* primer under the following conditions: 96 °C for 3 min for initial denature, 40 cycles of 94 °C for 40 s for denature, 53 °C annealing for 30 s, and 60 °C extension for 4 min. Labeled extension was cleaned by two isopropanol and two ethanol precipitations and then analyzed with a 377 Automated DNA sequencer. Fragments were aligned with Clustal X (Thompson et al., 1997) and edited manually.

Estimates of population genetic variation were obtained in the form of haplotype (h) and nucleotide (π) diversities (Nei, 1987; equations 8.4 and 10.5 respectively), using the software DnaSp (Rozas and Rozas, 1999). Estimates of nucleotide sequence divergence between mtDNA haplotypes were calculated using the Kimura two-parameter model (Kimura, 1980) according to algorithms implemented in Arlequin version 2.0 (Schneider et al., 2000). Estimates of the proportion of gene diversity among populations

were calculated by Φ_{ST} values of AMOVA (Excoffier et al., 1992) after 2,000 permutations of original data matrices, also with Arlequin. Genetic distances among sequences were clustered by the neighbor-joining method (Saitou and Nei, 1987) using MEGA version 2.0 (Kumar et al., 2001). Maximum parsimony analysis using branch-and-bound searches was also performed with MEGA version 2.0 and consensus trees (rooted by *Hydromedusa tectifera*) for both clustering methods were evaluated by means of bootstrap values (Felsenstein, 1985) based on 1,000 replicates.

RESULTS

Amplification resulted in a 416 bp fragment from the 5' end of the control region, which was sequenced in all 54 *Hydromedusa maximiliani* samples. Thirty-one polymorphic sites were found (19 transitions, 11 transversions, and one site with a transition/transversion), 12 of which were parsimony informative (Table 1). Based on polymorphism in control region fragments, 18 (I-XVIII) distinct haplotypes were detected among the 54 individuals. Reference sequences for representative haplotypes have been deposited in GenBank (AF448833-AF448837).

Haplotype distribution among localities had both strong geographic and endemic structure: haplotypes I-XIII were found only in PECB, haplotypes XIV-XVI occurred only in NF and haplotypes XVII and XVIII were endemic to TERES and JF, respectively (Table 1). Haplotype diversity (mean ± sd) among localities was $h = 0.583 \pm 0.082$ and nucleotide diversity (mean ± sd) was $\pi = 0.006 \pm 0.004$. Analysis of molecular variance (AMOVA) indicated a significant percentage (83.2%; $P << 0.001$) of total variance could

be attributed to variation in haplotype distribution among localities. Genetic distances between haplotypes ranged from 0.002 to 0.032 (mean = 0.0152).

Four *Hydromedusa maximiliani* populations showed similar topologies for both neighbor-joining and maximum parsimony consensus trees (Fig. 2). Although topologies demonstrated no clear local genetic partitioning, two well-defined geographic groups could be distinguished: a cluster corresponding to a western population represented by PECB haplotypes (supported by bootstrap values >90%), and a cluster corresponding to an eastern population represented by NF, TERES, and JF haplotypes (supported by bootstrap values >55%). Eastern group individuals were clearly distinguished from western group individuals at 16 polymorphic sites resulting from transitions and transversion events in control region sequences between sites 102 and 395 (Table 1).

DISCUSSION

Mitochondrial DNA control region sequences of *Hydromedusa maximiliani* revealed two major populations represented by specimens from eastern and western parts of the species' range. While our limited number of samples (6) from the eastern region may not be representative, natural history records, together with previous population genetic studies, provide support for the phylogeographic pattern we observed.

As a consequence of aquatic habit, limited dispersal ability, and complex habitat topology, genetic partitioning over a small spatial scale has been detected in some *Hydromedusa maximiliani* populations (Souza et al., 2002a,b). This localization of genetic structure may be present in other populations and could result in increase of population differentiation across the species' range. Colonization events among regions

must be extremely rare (a 0.832 Φ_{ST} value indicates such episodes) and must have occurred a long time ago when there was a panmitic population. Nucleotide divergences of up to 3.2% among sequences from individuals in these regions are suggestive of the background evolutionary scenario which shaped the structure of *H. maximiliani* populations. One could argue that unbalanced samples from four localities could result in such estimates of sequence divergence. To compensate for this bias, a sub-sample of just 6 random individuals from PECB sample was taken to compare with the other six samples from eastern population. This new analysis, when compared with previous results including all 48 individuals from PECB, resulted in a similar value of genetic distances between haplotypes (mean = 0.015; range: 0.003 to 0.031) and the same tree topology as depicted in Figure 2. If an estimate of DNA sequence divergence of 0.2%–0.4% per million years for the control region of turtle mtDNA (Avise et al., 1992; Bowen et al., 1993; Encalada et al., 1996; but see Seddon et al., 1998) is applied to the mtDNA control region sequences of *H. maximiliani*, eastern and western groups could have diverged at least 8-16 million years (Myr) ago, during Pliocene and Miocene.

Although molecular clocks must be viewed with caution (Rodríguez-Trelles et al., 2001), our results are concordant with geological events in the region, the fossil record for the South American Chelidae in general, and the genus *Hydromedusa* in particular. During Pliocene and Miocene, southeastern Brazil was subjected to intense orographical activity resulting in Serra do Mar and Serra da Mantiqueira mountain uplifts (Almeida, 1976; Suguio and Martin, 1976). These uplifts presently provide habitat for *H. maximiliani*. Fossil records for South American Chelidae turtles date from the Cretaceous and early Paleocene, whereas *H. tectifera* fossils date from Eocene, 56 Myr ago (Wood

and Moody, 1976; de la Fuente et al., 2001; de la Fuente and Bona, 2002). In southeastern Brazil, turtle fossils are recorded from Oligocene (34 Myr ago; Almeida, 1976) in the Vale do Paraíba (Paraíba Valley), a valley formed by the Paraíba River, located between the two mountain ridges (see Fig. 1). A panmitic *H. maximiliani* population may have been present in the region before mountain uplift. After geologic disruption, populations may have experienced periods of isolation by ridges and valleys typical of present day relief. Thus, divergence dates estimated from mitochondrial control region sequences for eastern and western *H. maximiliani* could reflect distinct evolutionary histories shaped by local phenomena.

Increased dryness in South America during Oligocene and Miocene allowed expansion of savanna-like vegetation, evidenced by paleofloral records from later Oligocene (Safford, 1999). This period of high temperature and dry climate had marked biogeographic and evolutionary implications. For South America in general, and southeastern Brazil in particular, periods of forest contraction forced populations of plant and animal species adapted to a wet climate to become isolated in wet areas, mainly in valleys (see revision in Safford, 1999). Since *Hydromedusa maximiliani* is presently found primarily in shallow rivers and streams in mountainous regions of Atlantic rainforest, some populations may have become isolated in propitious habitats of fragmented forest during critical periods of dryness. Local differentiation could then have ensued. Both vicariant and climatic events during Oligocene and Miocene could have resulted in the phylogeographical pattern observed for *H. maximiliani*. Although samples from Rio de Janeiro (Nova Friburgo and Teresópolis) and Minas Gerais (Juiz de Fora) states were from distinct mountain ranges (Serra do Mar and Serra da Mantiqueira,

respectively), these areas are geographically separated by only 80 km. Such geographical proximity could have placed these regions in the same phylogeographical group (although small sample size must be considered).

Eastern and western *Hydromedusa maximiliani* populations sampled were separated by approximately 1,200 km. Since a significant population genetic structure has been detected in this species over a small spatial scale (Souza et al., 2002a,b), similar genetic partitioning may occur elsewhere. At least one evolutionarily significant unit can be identified (eastern population), within which distinct drainages should be considered as conservation management units (Souza et al., 2002a,b). Studies of population biology and sampling of additional localities must be considered future steps in development of species management programs.

Acknowledgements. – We thank S. Hyslop for reviewing the English text and an anonymous reviewer for comments on the manuscript. We also thank C. Y. Miyaki for help in early stages of this work and Andréia de Assis for laboratory assistance. The Instituto Florestal provided logistic support, P. P. Soares and A. John gave assistance in the field, and J. C. Pombal Jr. and R. Pineschi provided blood samples. This work was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (FAPESP; grants 95/09449-0 and 00/00805-9) and Fundo de Apoio ao Ensino e à Pesquisa, Universidade Estadual de Campinas, São Paulo, Brazil (FAEP-UNICAMP; grant 0023/00). FLS is the recipient of a post-doctoral fellowship from FAPESP (contract number 99/02761-0). SFR is partially supported by a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

LITERATURE CITED

ALMEIDA, F. F. M. DE. 1976. The system of continental rifts bordering the Santos basin, Brazil. Anais da Academia Brasileira de Ciências 48 (Suplemento):15-26.

AVERY, H. W., AND L. J. VITT. 1984. How to get blood from a turtle. Copeia 1984:209-210.

AVISE, J. C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge.

AVISE, J. C., B. W. BOWEN, T. LAMB, A. B. MEYLAN, AND E. BERMINGHAM. 1992. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the testudines. Molecular Biology and Evolution 9:457-473.

BOWEN, B. W., W. S. NELSON, AND J. C. AVISE. 1993. A molecular phylogeny for marine turtles: trait mapping, rate assessment, and conservation relevance. Proceedings of the National Academy of Sciences of the USA, 90:5574-5577.

BRUFFORD, M. W., O. HANOTTE, J. F. Y. BROOKFIELD, AND T. BURKE. 1992. Single-locus and multilocus DNA fingerprint.. In A. R. Hoelzel (ed.). The South American Herpetofauna: Its Origin, Evolution, and Dispersal, pp. 225-270. Molecular Genetic Analysis in Conservation. IRL Press, Oxford.

CARDOSO, S. R. S., N. B. ELOY, J. PROVAN, M. A. CARDOSO, AND P. C. G. FERREIRA. 2000. Genetic differentiation of *Euterpe edulis* Mart. populations estimated by AFLP analysis. Molecular Ecology 9:1753-1760.

CHIARELLO, A. G. 1999. Effects of fragmentation of the Atlantic forest on mammal communities in south-eastern Brazil. Biological Conservation 89:71-82.

DE LA FUENTE, M., F. D. DE BROIN, AND T. M. DE BIANCO. 2001. The oldest and first nearly complete skeleton of a chelid, of the *Hydromedusa* sub-group (Chelidae, Pleurodira), from the upper Cretaceous of Patagonia. Bulletin de la Societe Geologique de France 172:237-244.

———————, AND P. BONA. 2002. Una nueva especie del género *Hydromedusa* Wagler (Pleurodira, Chelidae) del Paleógeno da Patagonia. Ameghiniana 39:77-83.

DEAN, W. 1996. A Ferro e Fogo: a História e a Devastação da Mata Atlântica Brasileira. Cia. das Letras, São Paulo.

DITCHFIELD, A. D. 2000. The comparative phylogeography of Neotropical mammals: patterns of intraspecific mitochondrial DNA variation among bats contrasted to nonvolant small mammals. Molecular Ecology 9:1307-1318.

DIXON, J. R. 1979. Origin and distribution of reptiles in lowland tropical rainforests of South America. In W. E. Duellman (ed.), The South American Herpetofauna: Its Origin, Evolution, and Dispersal, pp. 217-240. Monograph of the Museum of Natural History, The University of Kansas, 7.

ENCALADA, S. E., P. N. LAHANAS, K. A. BJORNDAL, A. B. BOLTEN, M. M. MIYAMOTO, AND A. B. BOWEN. 1996. Phylogeography and populations structure of the Atlantic and Mediterrean green turtle *Chelonia mydas*: a mitochondrial DNA control region sequence assessment. Molecular Ecology 5:473-483.

EXCOFFIER, L., P. E. SMOUSE, AND J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, AND M. NEI. 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Bioinformatics (submitted).

LAMB, T., J. C. AVISE, AND J. W. GIBBONS. 1989. Phylogeographic patterns in mitochondrial DNA of the desert tortoise (*Xerobates agassizi*), and evolutionary relationships among the North American gopher tortoises. Evolution 43:76-87.

LARA, M. C., AND J. L. PATTON. 2000. Evolutionary diversification of spiny rats (genus *Trinomys*, Rodentia: Echimyidae) in the Atlantic Forest of Brazil. Zoological Journal

of the Linnean Society 130:661-686.

MIYAKI, C. Y., S. R. MATIOLI, T. BURKE, AND A. WAJNTAL. 1998. Parrot evolution and paleogeographical events: Mitochondrial DNA evidence. Molecular Biology and Evolution 15:544-551.

MORITZ, C. 1994. Defining 'Evolutionary Significant Units' for conservation. Trends in Ecology and Evolution 9:373-375.

————. 1999. Conservation units and translocations: strategies for conserving evolutionary processes. Heredity 130:217-228.

MORITZ, C., J. L. PATTON, C. J. SCHNEIDER, AND T. B. SMITH. 2000. Diversification of rainforests faunas: an integrated molecular approach. Annual Review of Ecology and Systematics 31:533-563.

NEI, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

NICHOLS, W. A., A. RESENDIZ, J. A. SEMINOFF, AND B. RESENDIZ. 2000. Transpacific migration of a loggerhead turtle monitored by satellite telemetry. Bulletin of Marine Science 67:937-947.

PRITCHARD, P. C. H., AND P. TREBBAU. 1984. The Turtles of Venezuela. Society for the Study of Amphibians and Reptiles, Athens, Ohio, USA.

RODRÍGUEZ-TRELLES, F., R. TARRÍO, AND F. J. AYALA. 2001. Erratic overdispersion of three molecular clocks: GPDH, SOD, and XHD. Proceedings of the National Academy of Sciences USA 25:11405-11410.

ROMAN, J., S. D. SANTHUFF, P. E. MOLER, AND B. W. BOWEN. 1999. Population structure and cryptic evolutionary units in the alligator snapping turtle. Conservation Biology 13:135-142.

ROZAS, J., AND R. ROZAS. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174-175.

SAFFORD, H. D. 1999. Brazilian Páramos I. An introduction to the physical environment and vegetation of the *campos de altitude*. Journal of Biogeography 26:693-712.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: a new method for reconstruction phylogenetic trees. Molecular Biology and Evolution 4:406-425.

SCHNEIDER, S., J. M. KUEFFER, D. ROESSLI, AND L. EXCOFFIER. 2000. Arlequin ver. 2.0: A Software for Population Genetic Data Analysis. Genetics and Biometry Laboratory, University of Geneve, Switzerland.

SCHROTH, W., B. STREIT, AND B. SCHIERWATER. 1996. Evolutionary handicap for turtles. Nature 384:521-522.

SEDDON, J. M., P. R. BAVERSTOCK, AND A. GEORGES. 1998. The rate of mitochondrial 12S RNA gene evolution is similar in freshwater turtles and marsupials. Journal of Molecular Evolution 46: 460-464.

SIMPSON, B. B. 1979. Quaternary biogeography of the high montane regions of South America. In W. E. Duellman (ed.), The South American Herpetofauna: Its Origin, Evolution, and Dispersal, pp. 157-188. Monograph of the Museum of Natural history, The University of Kansas, 7.

SITES, J. W., JR., N. N. FITZSIMMONS, N. J. SILVA JR., AND V. H. CANTARELLI. 1999. Conservation genetics of the Giant Amazon river turtle (*Podocnemis expansa*; Pelomedusidae) – Inferences from two classes of molecular markers. Chelonian Conservation and Biology 3:454-463.

SOUZA, F. L. 1995. História natural do cágado Hydromedusa maximiliani (Mikan 1820)

no Parque Estadual de Carlos Botelho, SP, região de Mata Atlântica (Reptilia, Testudines, Chelidae). Unpubl. MS Thesis, Universidade Estadual Paulista, Rio Claro, São Paulo state, Brasil.

SOUZA, F. L., AND A. S. ABE. 1997. Population structure, activity, and conservation of the Neotropical freshwater turtle, *Hydromedusa maximiliani*, in Brazil. Chelonian Conservation and Biology 2:521-525.

————, AND ————. 1998. Resource partitioning by the neotropical freshwater turtle, *Hydromedusa maximiliani*. Journal of Herpetology 32:106-112.

SOUZA, F. L., A. F. CUNHA, M. A. OLIVEIRA, G. A. G. PEREIRA, H. P. PINHEIRO, AND DOS REIS, S. F. 2002a. Partitioning of molecular variation at local spatial scales in the vulnerable neotropical freshwater turtle, *Hydromedusa maximiliani* (Testudines, Chelidae): implications for the conservation of aquatic organisms in natural hierarchical systems. Biological Conservation 104:119-126.

————, ————, ————, ————, AND ————. 2002b. Estimating dispersal and gene flow in the neotropical freshwater turtle *Hydromedusa maximiliani* (Chelidae) by combining ecological and genetic methods. Genetics and Molecular Biology 25:151-155.

SUGUIO, K, AND L. MARTIN. 1976. Brazilian coastline quaternary formations – the states of São Paulo and Bahia littoral zone evolutive schemes. Anais da Academia Brasileira de Ciências 48 (Suplemento):325-334.

TABARELLI, M., W. MANTOVANI, AND C. A. PERES. 1999. Effect of habitat fragmentation on plant guild structure in the montane Atlantic forest of southeastern Brazil. Biological Conservation 91:119-127.

THOMPSON, J. D., T. J. GIBSON, F. PLEUNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acid Research 24:4876-4882.

VALENZUELA, N. 2001. Genetic differentiation among nesting beaches in the highly migratory giant river turtle (*Podocnemis expansa*) from Colombia. Herpetologica 57:48-57.

VANZOLINI, P. E. 1973. Distribution and differentiation of animals along the coast and in continental islands of the state of S. Paulo, Brasil. I. Introduction to the area and problems. Papéis Avulsos de Zoologia, São Paulo 26:281-294.

WALKER, D., AND J. C. AVISE. 1998. Principles of phylogeography as illustrated by freshwater and terrestrial turtles in the southeastern United States. Annual Review of Ecology and Systematics 29:23-58.

WALKER, D., W. S. NELSON, K. A. BUHLMANN, AND J. C. AVISE. 1997. Mitochondrial DNA phylogeography and subspecies issues in the monotypic freshwater turtle *Sternotherus odoratus*. Copeia 1997:16-21.

WOOD, R. C., AND R. T. J. MOODY. 1976. Unique arrangement of carapace bones in the South American chelid turtle *Hydromedusa maximiliani* (Mikan). Zoological Journal of Linnean Society of London, B 59:69-78.

Table 1. Polymorphic loci related to the 18 mtDNA control region haplotypes from 54 *Hydromedusa maximiliani* from throughout the species' distribution in southeastern Brazil: PECB (Parque Estadual de Carlos Botelho, SP), NF (Nova Friburgo, RJ), TERES (Teresópolis, RJ), and JF (Juiz de Fora, MG).

	position																											sample region			
								1	1	1	1	1	1	1	3	3	3	3	3	3	3	4	4	4	4	4	4				
		3	4	4	4	5	5	5	0	2	2	3	4	5	5	8	1	2	6	6	8	9	9	9	0	0	0	1	1	1	
Haplotypes	1	3	9	0	1	9	1	2	3	2	0	2	8	2	1	9	6	0	6	4	7	7	5	6	9	5	7	8	0	2	3	PECB	NF	TERES	JF
I	C	A	A	C	C	T	T	A	T	T	A	A	T	A	G	T	A	A	T	A	A	C	A	T	T	C	T	C	G	C	T	35			
II	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	1			
III	.	.	G	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1			
IV	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	C	T	.			1			
V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.			1			
VI	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C			2			
VII	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.			1			
VIII	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.			1			
IX	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.			1			
X	.	.	.	.	.	C	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.			1			
XI	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.			1			
XII	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	A	.	C	.	.		1			
XIII	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.	.			1			
XIV	.	.	.	T	.	.	.	.	C	.	.	.	.	A	C	.	.	A	.	.	.	G	.	.	.	.	.	.	.				2		
XV	.	.	.	.	.	.	.	.	C	.	.	C	T	A	C	.	.	A	.	G	.	G	.	.	.	.	.	.	.				1		
XVI	.	.	.	.	T	.	.	.	C	G	G	C	.	A	C	.	.	A	.	.	.	G	.	.	.	.	.	.	.				1		
XVII	.	.	.	.	.	.	.	.	C	G	G	C	.	A	C	.	.	A	.	.	.	G	.	.	.	.	.	.	.					1	
XVIII	.	.	.	.	.	.	.	.	C	G	G	C	.	A	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.						1

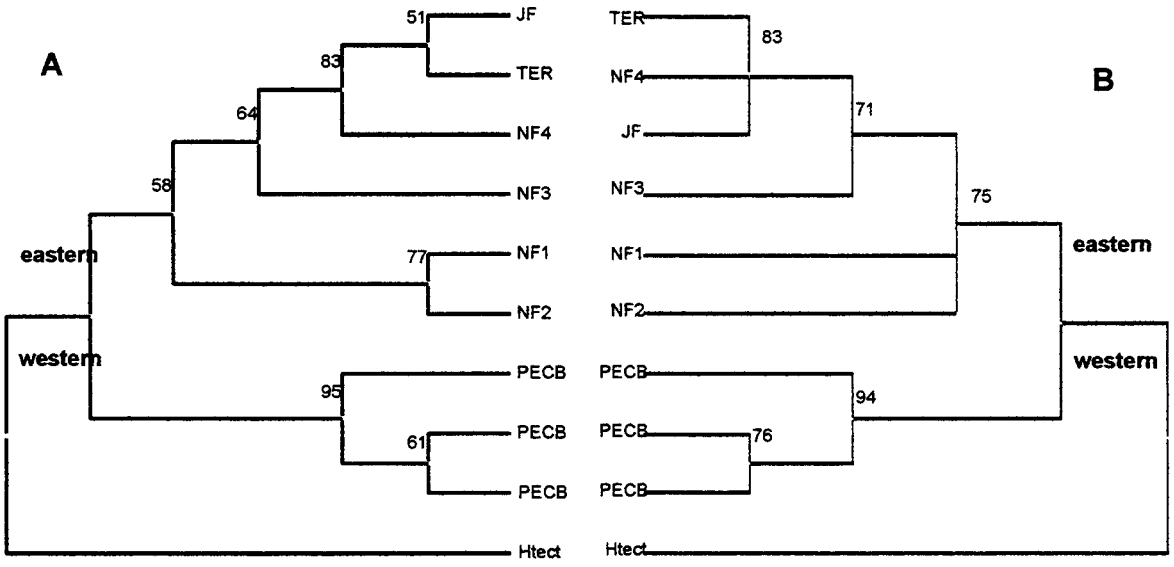

A
B
51
JF
83
TER
64
NF4
58
NF3
eastern
77
NF1
NF2
western
95
PECB
61
PECB
PECB
Htect
TER
83
NF4
71
JF
NF3
75
NF1
eastern
NF2
western
PECB
94
PECB
76
PECB
Htect

Figure legends

Figure 1. Localities sampled. S. Miguel Arcanjo corresponds to Parque Estadual de Carlos Botelho (PECB).

Figure 2. Neighbor-joining (A) and maximum parsimony (B) consensus trees for haplotypes of the partial mtDNA control region of *Hydromedusa maximiliani* from Juiz de Fora, MG (JF), Nova Friburgo, RJ (NF1-NF4), Teresópolis, RJ (TER), and Parque Estadual de Carlos Botelho, SP (PECB). Values above branches represent >50% bootstrap. Trees were rooted by *Hydromedusa tectifera* (Htect). For simplification, only three (of 48) individuals from PECB are shown. For the maximum parsimony: tree length = 120; CI = 0.97, and RI = 0.85.

-Participação em cursos e disciplinas

Cristalização de proteínas-módulo II - Laboratório Nacional de Luz Síncrontron (2002).

Treinamento prático de microarray - LGE/ Perkin Elmer-Unicamp (2001).

Métodos computacionais em biologia - Laboratório Nacional de Computação Ciêntifica (2000).

Bioquímica Experimental I - NB325 (2001).

Introdução à Engenharia Genética - NG224 (2000).

Seminários do Departamento de Genética - NG120 (1998).

Comunicações em congressos

"Clonning And Disruption Of The SWI2 And SWI3 Genes Encoding Subunits Of The SWI-SNF Complex in The Yeast *Hansenula Polymorpha*". V. Genu, A.P.T. Salmazo, M.A. Oliveira & G.A.G. Pereira.
XXXI Reunião Anual da SBBq (2002).

"The knockout of *TUP1*, *MIG1* and *SNF1* genes, enhances the growth and methanol Metabolism In The Yeast *Hansenula polymorpha*". M.A. Oliveira & G.A.G. Pereira.
XXXI Reunião Anual da SBBq (2002).

"Clonning the gene similar to SRB10 of the transcription mediator complex in *Hansenula polymorpha.*" B.Monte-Mor, M.A. Oliveira & G.A.G. Pereira
XXXI Reunião Anual da SBBq (2002).

"Disruption of Transcriptional Factors that Control the *MOX* Gene Expression in *H. polymorpha".* M.A. Oliveira & G.A.G. Pereira.
XXXI Reunião Anual da SBBq (2001).

"Cloning and Characterization of *SWI3* Transcription Factor in Yeast *H. polymorpha".* A.P.T. Salmazo, M.A. Oliveira & G.A.G. Pereira.
XXX Reunião Anual da SBBq (2001).

"Identification of a *Lilliputian* Gene (*Drosophila melanogaster*) Homologue Sequence in *Melipona quadrifasciata* (Stingless Bee)." C. M. Judice , M.A. Oliveira & G.A.G. Pereira.
XXX Reunião Anual da SBBq (2001).

"The cDNA Cloning of BJCUL, a Lectin from the venom of *Bothrops Jararacussu*". B.K. Hussein, M.A. Oliveira, G.A.G. Pereira & C. Novello.
XXX Reunião Anual da SBBq (2001).

"Universal Genome Experience and Expression Laboratory (LGE) Bioinformatics : The Chicken Est Genome" . F. Tsukumo, M.A. Oliveira & G.A.G. Pereira.
I Brazilian International Genome Conference(2000)

"Cloning and Characterization of Transcription Factors that Control the *MOX* Gene Expression in *H. polymorpha"* M.A. Oliveira & G.A.G. Pereira.
XXIX Reunião Anual SBBq (2000).

"Screening of DNA libraries by PCR techinique". A. Salmazo, M.A. Oliveira & G.A.G. Pereira.
XXIX Reunião Anual SBBq (2000).

"New system for industrial production of lipase". E. Bogo, M.A. Oliveira, G.M. Pastore & G.A.G. Pereira.
XXIX Reunião Anual SBBq (2000).

"A comparison between PCR and hybridization to perform paternity analysis in a double incest case" A. Medaglia , E. Matheucci Jr, D. Hansen, M.A. Oliveira & F. Henrique – Silva.
XXVIII Reunião Anual SBBq (1999).

"Identification of transcription factors: *Snf1*, *Mig1*, *Tup1* and *Ssn6.* to investigate the involving in the control of *MOX* gene expression in *H polymorpha.*" M.A. Oliveira & G.A.G. Pereira.
XXX Reunião Anual da SBBq (1999).