

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Bacharelado em Biotecnologia – Bioinformática

Fábio Carrer Andreis

Caracterização Filogenética de Ureasas

Dr. Hugo Verli
Orientador

Dra. Célia Regina Ribeiro da Silva Carlini
Co-orientadora

Me. Rodrigo Ligabue Braun
Co-orientador

Porto Alegre

2013

Fábio Carrer Andreis

Caracterização Filogenética de Ureases

Monografia apresentada como
requisito parcial para obtenção do
grau de Bacharel em
Biotecnologia - Bioinformática
Universidade Federal do Rio
Grande do Sul
Instituto de Biociências; Centro de
Biotecnologia do Rio Grande do
Sul

Dr. Hugo Verli
Orientador

Dra. Célia Regina Ribeiro da Silva Carlini
Co-orientadora

Me. Rodrigo Ligabue Braun
Co-orientador

Porto Alegre

2013

Agradecimentos

Aos meus pais, pelo apoio constante, apesar de opiniões divergentes.

À Ethel, pelo companheirismo, confiança e apoio a todo instante.

Ao Rodrigo, pelo suporte, paciência e amizade dentro e fora do laboratório.

Aos meus orientadores, pela oportunidade e por toda a ajuda na minha formação.

Ao Lado Negro, pelos momentos descontraídos e conselhos, sem os quais essa jornada seria consideravelmente mais difícil.

Aos não citados, porém não esquecidos, também a minha gratidão.

Nota: Todas as figuras presentes neste trabalho estão disponíveis para download em:

<http://goo.gl/VYTvR>

Sumário

Resumo	vi
Abstract.....	vii
Índice de Figuras	viii
1. Introdução.....	1
1.1. Ureases	1
1.2. Análise filogenética	2
1.3. Justificativa	6
2. Objetivos	7
3. Métodos.....	8
4. Resultados e Discussão	10
5. Conclusões e perspectivas.....	22
6. Referências	23
APÊNDICES	29
APÊNDICE A: Identificadores GenInfo das sequências de ureases utilizadas.....	30
APÊNDICE B: Análise filogenética molecular de regiões altamente variáveis das sequências de ureases pelo método de Máxima Verossimilhança.....	33
APÊNDICE C: Enraizamento por ponto médio matemático da árvore de sequências completas de ureases pelo método de Máxima Verossimilhança.....	34
APÊNDICE D: Enraizamento por ponto médio matemático da árvore de sequências completas de ureases pelo método de Inferência Bayesiana.....	35
APÊNDICE E: Enraizamento por ponto médio matemático da árvore de regiões conservadas de ureases pelo método de Máxima Verossimilhança.....	36
APÊNDICE F: Enraizamento por ponto médio matemático da árvore de regiões conservadas de ureases pelo método de Inferência Bayesiana.....	37
APÊNDICE G: Publicação de artigo científico.....	38

Resumo

Ureases são enzimas níquel-dependentes que catalisam a hidrólise da uréia em amônia e carbamato. Apesar da aparente riqueza de informação sobre as ureases, vários aspectos cruciais acerca dessas enzimas permanecem desconhecidos, ou são objeto de debates correntes. Um desses é, certamente, a sua organização estrutural: ureases de plantas e fungos são unitárias, enquanto as de archaea e bactérias possuem estruturas de duas ou três cadeias. Entretanto, o estado primitivo dessas proteínas – uma ou três cadeias – é desconhecido, apesar de muitos esforços para tanto. Através de inferência filogenética utilizando três conjuntos distintos de dados e dois algoritmos diferentes, pudemos observar a transição no número de cadeias na forma 3-para-1. Nossos resultados sugerem que o estado ancestral para as ureases é a organização em três cadeias, sendo as ureases de uma cadeia derivadas delas. As variantes de duas cadeias não se mostraram como intermediários evolutivos. Um processo de fusão distinto dos já descritos pode explicar essa transição estrutural.

Palavras-chave: urease; árvore filogenética; transição estrutural; evolução; fusão gênica; fissão gênica

Abstract

Ureasases are nickel-dependent enzymes which catalyze the hydrolysis of urea to ammonia and carbamate. Despite the apparent wealth of data on ureasases, many crucial aspects regarding these enzymes are still unknown, or constitute matter for ongoing debates. One of these is most certainly their structural organization: ureasases from plants and fungi have a single unit, while bacterial and archaean ones have two- or three-chained structures. However, the primitive state of these proteins - single- or three-chained - is yet unknown, despite many efforts in the field. Through phylogenetic inference using three different datasets and two different algorithms, we were able to observe chain number transitions displayed in a 3-to-1 fashion. Our results imply that the ancestral state for ureasases is the three-chained organization, with single-chained ureasases deriving from them. The two-chained variants are not evolutionary intermediates. A fusion process, different from those already described, may explain this structural transition.

Keywords: urease; phylogenetic tree; structural transition; evolution; gene fusion; gene disruption

Índice de Figuras

Figura 1: Esquema da reação de hidrólise da uréia..	1
Figura 2: Exemplos de mecanismos de alternância funcional.....	3
Figura 3: Organização estrutural das ureases.	4
Figura 4: Esquema de árvore filogenética.....	5
Figura 5: Testes de confiabilidade.	6
Figura 6: Fluxograma geral de trabalho.	9
Figura 7: Alinhamento das sequências de ureases utilizadas.	11
Figura 8: Análise filogenética molecular das sequências de ureases completas pelo método de máxima verossimilhança.....	12
Figura 9: Análise filogenética molecular das sequências de ureases completas pelo método de Inferência Bayesiana.....	13
Figura 10: Análise filogenética molecular das regiões conservadas de sequências de ureases pelo método de Máxima Verossimilhança.....	14
Figura 11: Análise filogenética molecular das regiões conservadas de sequências de ureases pelo método de inferência Bayesiana.	15
Figura 12: Organização dos operons ure das ureases microbianas estudadas.	16
Figura 13: Regiões de variabilidade identificadas com SimPlot.....	17
Figura 14: Esquema do processo de fusão para imunoglobulinas.....	18
Figura 15: Esquema do processo de exon shuffling.....	19
Figura 16: Transferência horizontal de genes.....	20

1. Introdução

1.1. Ureasas

As ureases (uréia amidoidrolases, EC 3.5.1.5) são metaloenzimas níquel-dependentes, responsáveis pela catálise da hidrólise da uréia em amônia e carbamato que, por sua vez, sofre hidrólise espontânea, formando ácido carbônico e uma segunda molécula de amônia (Figura 1) ¹. A uréia e a urease representaram papel significativo no desenvolvimento das ciências naturais (conforme revisado por Krajewska, 2009 ²). Após sua descoberta na urina humana em 1773, por Hillaire Rouelle, a uréia se tornou o primeiro composto orgânico a ser sintetizado a partir de materiais inorgânicos³. Embora a noção de que a amônia derivada de urina provém de uréia date de 1798, somente em 1890 a enzima ureolítica foi isolada. A urease de feijão-de-porco (*Canavalia ensiformis*) foi a primeira enzima a ser cristalizada⁴, provando que enzimas são proteínas e que são passíveis de cristalização. De igual importância, essa enzima foi a primeira na qual se demonstrou a presença de níquel em seu sítio ativo, essencial para sua atividade ⁵.

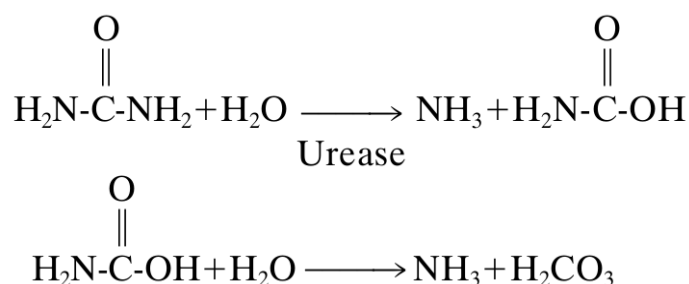


Figura 1: Esquema da reação de hidrólise da uréia. Acima, reação de hidrólise da uréia em amônia e carbamato, representando a catálise por urease. Abaixo, reação de hidrólise espontânea do carbamato, formando amônia e ácido carbônico. Adaptado de Mobley et al., 1995.

Em plantas, essas enzimas tomam parte no metabolismo de compostos nitrogenados, principalmente disponibilizando a uréia como fonte de nitrogênio, assumindo papel vital no seu desenvolvimento. Adicionalmente, a urease é sintetizada em fungos, algas, bactérias e invertebrados, além de existir no solo. Dada sua ampla distribuição entre os reinos da vida., as ureases assumem um grande papel no metabolismo de nitrogênio na natureza ².

Além da atividade ureolítica, a urease desempenha um leque de funções não associadas ao processo catalítico. Após a digestão da enzima por catepsinas (no trato digestório de besouros e percevejos), um peptídeo entomotóxico de 10 kDa é liberado, conferindo à urease propriedades inseticidas ^{6,7}. Além desta propriedade, em soja e feijão-de-porco, a urease atua inibindo o crescimento de diversos tipos de fungos ⁸. Outras funções independentes de catálise incluem a ativação plaquetária ⁷ e ligação a glicoconjugados ⁹. A presença de múltiplas funções não relacionadas entre si e ao processo catalítico permite classificar a urease como uma proteína

*moonlighting*¹⁰.

A ideia de que um gene codifica uma proteína que, por sua vez, desempenha uma função vem sendo subjugada pelo crescente número de proteínas descobertas que desempenham mais de uma função - fenômeno conhecido como *moonlighting*. Essa multiplicidade pode ser decorrente de mudanças na localização e/ou tipo celular, estado oligomérico, ou concentração celular de determinado ligante, substrato, cofator ou produto (Figura 2)^{10,11}. Uma grande variedade de proteínas, incluindo a enolase (em fungos, exerce atividades glicolítica e de importe de tRNA na mitocôndria), citocromo c (em animais, metabolismo energético e apoptose) e proteínas do cristalino ocular (também em animais, função estrutural - lente - e atividades enzimáticas diversas)¹², desenvolveram funções adicionais ao curso da evolução, e não há, aparentemente, características estruturais comuns a essas proteínas^{10,11}.

Um dos aspectos mais marcantes das ureases tange a sua organização estrutural (Figura 3). As ureases de plantas e fungos assumem estruturas homo-oligoméricas (trímeros α_3 ou hexâmeros α_6 , tipicamente), com aproximadamente 840 resíduos de aminoácidos¹. Diferentemente das enzimas vegetais, ureases microbianas são multímeros compostos de duas ou três subunidades (α , β e γ - ~570, ~120 e ~100 resíduos, respectivamente), normalmente formando trímeros $(\alpha\beta\gamma)_3$ ^{1,13}. Bactérias do gênero *Helicobacter*, diferentemente de todas as demais bactérias, apresentam enzimas constituídas por duas cadeias polipeptídicas, de tamanho equivalente¹. Apesar dos distintos estados oligoméricos, as sequências de aminoácidos das subunidades menores das ureases microbianas são altamente similares às das regiões correspondentes nas unidades simples das ureases eucarióticas².

Essa alta similaridade, observada entre ureases de diferentes reinos, sugere que todas derivam de uma única proteína ancestral e, provavelmente, possuem estruturas terciárias e mecanismos catalíticos similares^{1,13-15}. Entretanto, não há divisão filogenética clara entre as sequências de bactérias, ou até mesmo entre bactérias, plantas e fungos¹.

1.2. Análise filogenética

O conceito de filogenia surgiu com Darwin, e parte da ideia de que as espécies se originaram de espécies ancestrais. Através da construção de árvores filogenéticas, que podem ser baseadas em caracteres morfológicos, funcionais ou moleculares, pretende-se representar a história evolutiva dos organismos nela incluídos (chamados de unidades taxonômicas operacionais; OTUs, do inglês, *operational taxonomic units*). Essas árvores consistem, graficamente, de pontos (ou nós) ligados por linhas (ou ramos), cuja disposição define a ancestralidade entre as OTUs. As mesmas são representadas nos nós terminais, unidas por ramos

cujo nó interno representa o ancestral comum mais recente desses *taxa*¹⁶ (Figura 4).

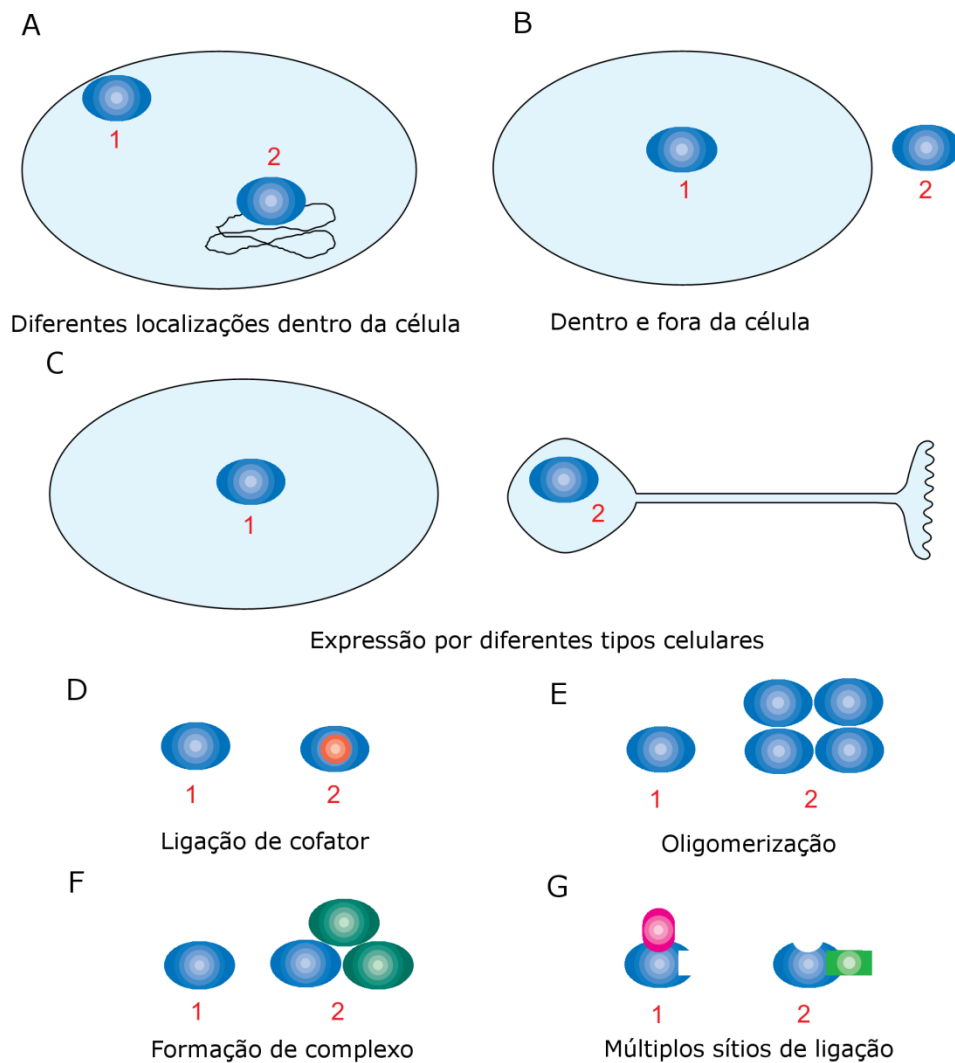


Figura 2: Exemplos de mecanismos de alternância funcional. Adaptado de Jeffery, 1999¹¹.

A análise evolutiva por meio de técnicas computacionais requer a escolha de algoritmos de reconstrução filogenética. Para aplicação em inferências baseadas em dados moleculares, destacam-se quatro técnicas principais: agrupamento de vizinhos, máxima parcimônia, máxima verossimilhança e inferência Bayesiana¹⁷.

O método de agrupamento de vizinhos (do inglês, *neighbor-joining*; NJ) pertence à classe dos métodos de distância (ou geométricos). Seu procedimento consiste na aplicação de matrizes de distâncias, construídas computando as diferenças entre todas as sequências-alvo (nesse contexto, sequências mais distantes possuem maior número de diferenças). A grande vantagem desse método é a independência de taxas de substituição constantes (o chamado relógio molecular)¹⁸. Outros algoritmos geométricos incluem UPGMA (do inglês, *unweighted pair group method with arithmetic means*) e evolução mínima¹⁹.

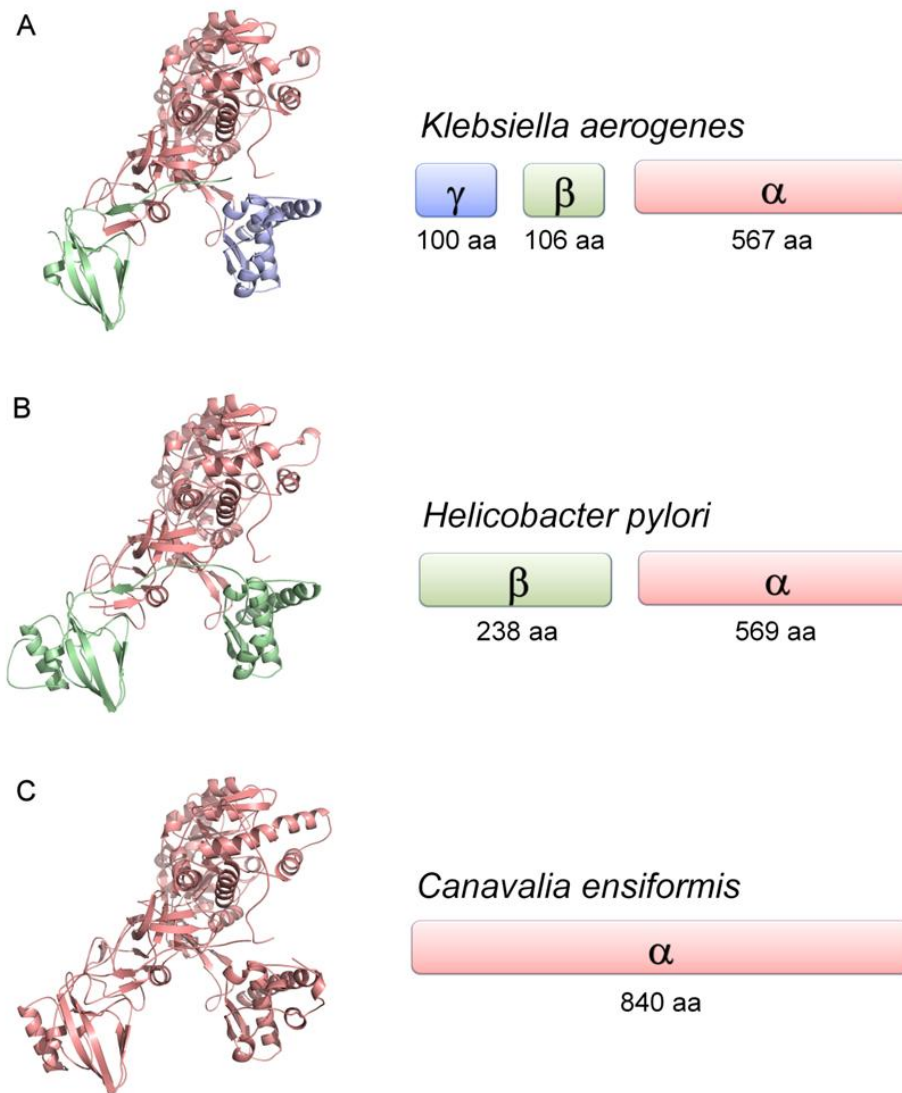


Figura 3: Organização estrutural das ureases. Representações tridimensionais e esquemáticas da organização das subunidades das ureases. **A)** Organização típica de ureases microbianas, composta de três cadeias (estrutura cristalográfica de PDB ID 1FWJ). **B)** Urease de Helicobacteraceae, composta de duas cadeias (estrutura cristalográfica de PDB ID 1E9Z). **C)** Urease eucariótica típica, composta por cadeia simples (estrutura cristalográfica de PDB ID 3LA4).

A máxima parcimônia (MP) é uma técnica baseada fortemente no princípio da navalha de Ockham, formulado pelo filósofo inglês William de Ockham. Em suma, o princípio dita que se há várias explicações aplicáveis a determinado fenômeno, a mais simples (que requer o menor número de pressupostos) é escolhida. No contexto filogenético molecular, refere-se ao número de substituições (ou mutações) de nucleotídeos, para DNA, e aminoácidos, para proteínas. Um dos primeiros registros da utilização do princípio de MP para reconstrução filogenética de macromoléculas data de 1966, por Eck e Dayhoff^{20, 16}.

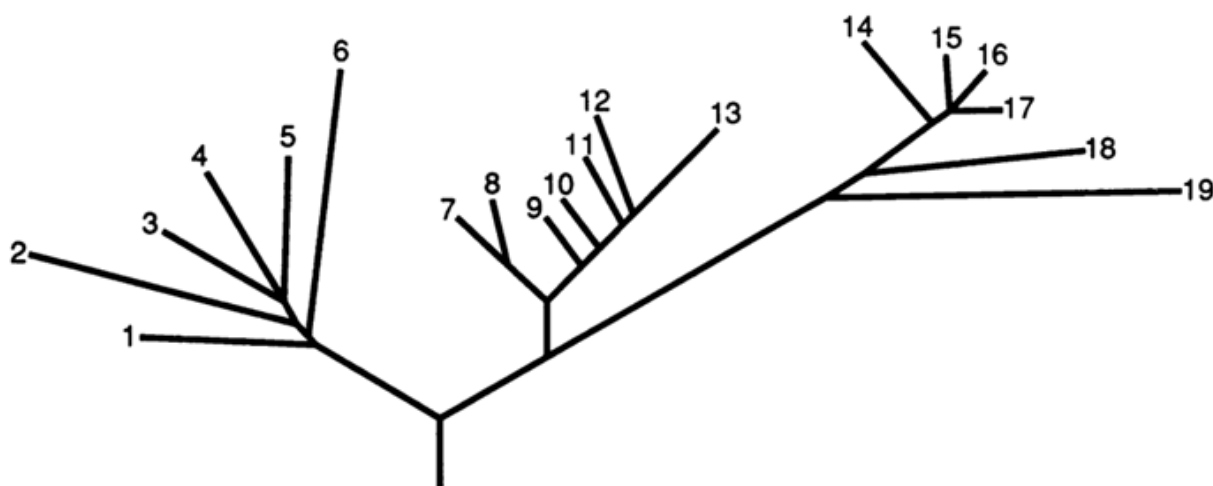


Figura 4: Esquema de árvore filogenética. OTUs representadas por números, posicionados nos nós terminais. Adaptado de Woese, 1990²¹.

Na classe de métodos probabilísticos, temos o método de máxima verossimilhança (ML; do inglês, *maximum likelihood*). A técnica tem base em um modelo pré-estabelecido de evolução de caracteres, podendo considerar parâmetros como taxa de substituição entre pares de nucleotídeos, frequência de bases, proporção de sítios invariáveis, e heterogeneidade de taxas de substituição entre sítios²². São computadas todas as árvores possíveis, sendo escolhida aquela com maior probabilidade de que os resultados tenham origem de acordo com o modelo evolutivo selecionado, baseado em um determinado alinhamento¹⁸.

Para avaliar a confiabilidade dos métodos NJ, MP e ML, emprega-se o teste de *bootstrap* (Figura 5A). Esse método consiste em reamostrar os dados, com reposição pseudo-aleatória. São construídas diversas árvores-réplica (geralmente 500 ou 1000, baseadas na primeira construída), cujas topologias são comparadas com a árvore original. Ao final do teste é fornecida a porcentagem de vezes em que a topologia original foi gerada nas árvores-réplica, para cada ramificação. Por exemplo, um *bootstrap* de 90% indica que a ramificação em questão foi observada em 90% das topologias geradas no processo¹⁹.

O método de inferência Bayesiana (BI; do inglês, *Bayesian inference*), matematicamente relacionado ao método de ML, baseia-se na estatística bayesiana desenvolvida pelo reverendo Thomas Bayes, em 1763. Ele propôs um método formal que incorpora os valores de uma variável obtidos *a priori* para determinar a probabilidade de ocorrência de um dado evento. No contexto filogenético, em suma, a aplicação do teorema de Bayes fornece a probabilidade de que determinada ramificação na árvore esteja correta (denominada probabilidade posterior), baseada em diversas amostragens (no contexto, são denominadas gerações)²³, sumarizadas ao final da análise e contabilizadas as suas frequências. O processo está representado na Figura 5B.

Em teoria, todos os métodos devem produzir topologias similares ou idênticas²², embora alguns métodos sejam mais robustos²⁴. Para fins de validação, tanto dos dados quanto dos parâmetros escolhidos, análises comparativas entre os resultados de cada método podem ser efetuadas.

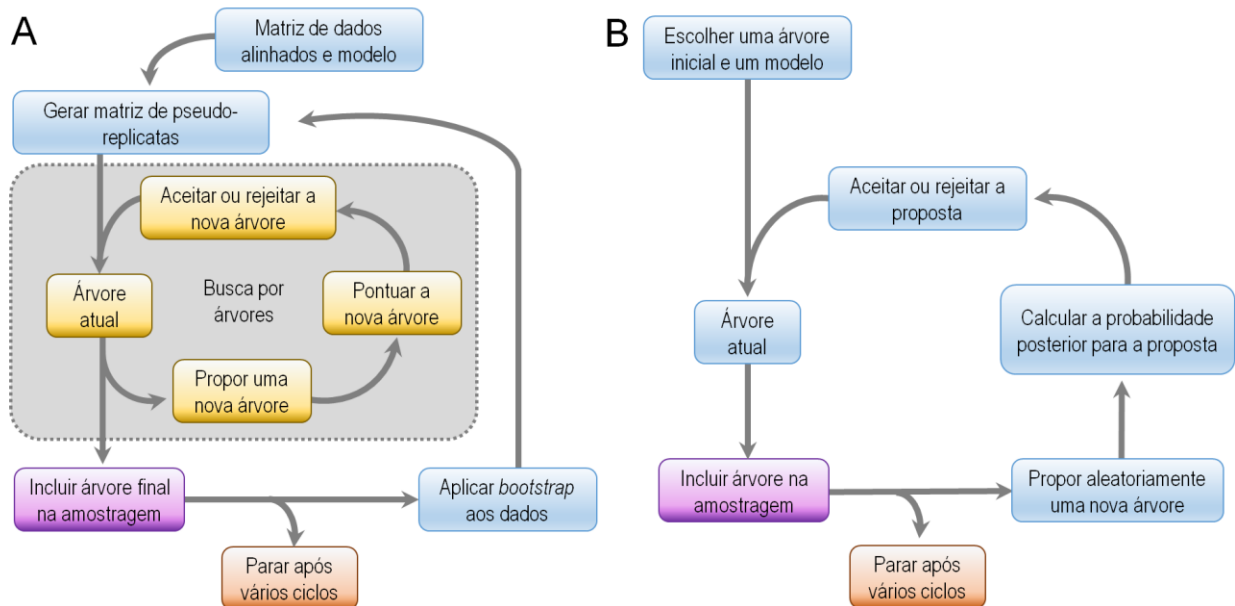


Figura 5: Testes de confiabilidade. A) Fluxograma esquemático do teste de *bootstrap*. B) Fluxograma esquemático do algoritmo bayesiano. Adaptado de Holder & Lewis, 2003²⁴.

1.3. Justificativa

Apesar da aparente riqueza de informação sobre as ureases, vários aspectos cruciais acerca dessas enzimas permanecem desconhecidos, ou são objeto de debates correntes. Tais tópicos incluem a natureza da ureólise²⁵⁻²⁷ e os mecanismos subjacentes da ativação da urease^{28,29}.

Permanece sem resposta, ainda, outro tópico central da biologia estrutural das ureases: qual o estado estrutural primitivo dessas enzimas? No momento, duas possibilidades surgem, conforme apontadas anteriormente por Hausinger³⁰: “O gene codificando a enzima de uma subunidade de plantas sofreu ruptura para produzir os genes múltiplos codificando as duas ou três subunidades bacterianas? Ou os genes bacterianos se fundiram para formar o gene codificante da subunidade de plantas?”. Nesse contexto, uma ampla análise das sequências de aminoácidos de ureases, buscando padrões, similaridades e diferenças possibilitaria elucidar o estado primitivo dessas enzimas e, possivelmente, auxiliaria na análise e predição de resultados experimentais.

2. Objetivos

O presente trabalho tem por objetivo geral a caracterização filogenética das relações evolutivas de ureases de diferentes organismos. Dessa forma, pretende-se contribuir com o entendimento das alterações no número de cadeias polipeptídicas dessas enzimas entre diferentes *taxa*.

Especificamente, objetiva-se evidenciar o estado primitivo da urease, um questionamento de longa data. Ademais, o presente trabalho visa avaliar o papel evolutivo das ureases de duas cadeias, caso exclusivo do gênero *Helicobacter*: se são etapas intermediárias do processo, ou uma ocorrência independente.

3. Métodos

Inicialmente, realizou-se uma prospecção por sequências de aminoácidos resultantes da busca pela palavra-chave “urease” do *National Center for Biotechnology Information*³¹. De modo a obter somente sequências completas de ureases, foram realizadas filtragens manuais, onde sequências incompletas (de tamanho muito inferior a 830 aminoácidos), erroneamente rotuladas (verificadas uma a uma através de seu número de acesso) ou relacionadas às proteínas acessórias da urease foram removidas.

Todas as sequências relacionadas a bactérias e archaea nos dados resultantes foram organizadas, sequencialmente, da seguinte maneira:

- Por reino;
- Por gênero;
- Por espécie.

Estas foram submetidas a alinhamentos múltiplos (por espécie e gênero, sequencialmente), visando encontrar similaridades. O algoritmo ClustalW³² foi utilizado em todos os alinhamentos, de modo que sequências com identidade igual ou superior a 95% foram consideradas idênticas e agrupadas como sequência única. Do conjunto original de 14.221 sequências, 32 delas provinham de fontes eucarióticas e não precisaram de filtragens adicionais em razão de seu menor número e ausência de duplicidades. Através dos alinhamentos por espécie e gênero, o número total foi reduzido a 162 sequências microbianas representativas.

As sequências dessas ureases microbianas foram alinhadas às ureases de plantas e fungos, e suas subunidades foram unidas manualmente em uma só sequência com base nesses alinhamentos (formando uma só sequência $\gamma\beta\alpha$, ou equivalente). O número de sequências microbianas foi reduzido ainda mais, visando diminuir a carga computacional. Baseado na árvore-guia de alinhamento gerada pelo ClustalW, sequências altamente similares (80% de identidade ou superior) foram consideradas idênticas, e somente um representante foi escolhido. O conjunto microbiano final foi montado e alinhado com as sequências eucarióticas, totalizando 124 sequências na lista final de ureases (Apêndice A). Regiões com longos *gaps* foram removidas manualmente dos alinhamentos.

Para inspecionar as sequências alinhadas em busca de regiões altamente variáveis, foi empregada a ferramenta SimPlot, que compara o percentual de identidade entre grupos de sequências para uma janela de cobertura específica, sobre toda a extensão das sequências³³. Para este procedimento, foi utilizada uma janela de 20 aminoácidos, deslocando-se a cada resíduo. Com base nesses resultados, foi possível identificar regiões de baixa similaridade geral

possibilitando a ramificação do conjunto inicial em três subconjuntos distintos: sequências completas de urease (sem longos gaps), regiões altamente variáveis, e regiões conservadas.

A avaliação das sequências de aminoácidos pelos programas ProtTest³⁴ e MEGA5³⁵ apontaram como mais adequado o modelo de substituição de aminoácidos de Whelan e Goldman (WAG) com distribuição Gamma discreta para contabilizar diferenças na taxa evolutiva entre sítios (+G), considerando alguns deles como evolutivamente invariáveis (+I)³⁶. Árvores filogenéticas foram calculadas por ML utilizando o MEGA5 com 1.000 replicatas de *bootstrap*, e árvores calculadas por BI utilizando o programa MrBayes³⁷ por $7,5 \times 10^6$ gerações, amostradas a cada 10 gerações. Todas as árvores foram enraizadas no ramo contendo o maior número de sequências de ureases pertencentes a archaea. As árvores obtidas foram visualizadas e editadas por meio do programa FigTree³⁸.

A Figura 6 retrata de forma sucinta a metodologia empregada.

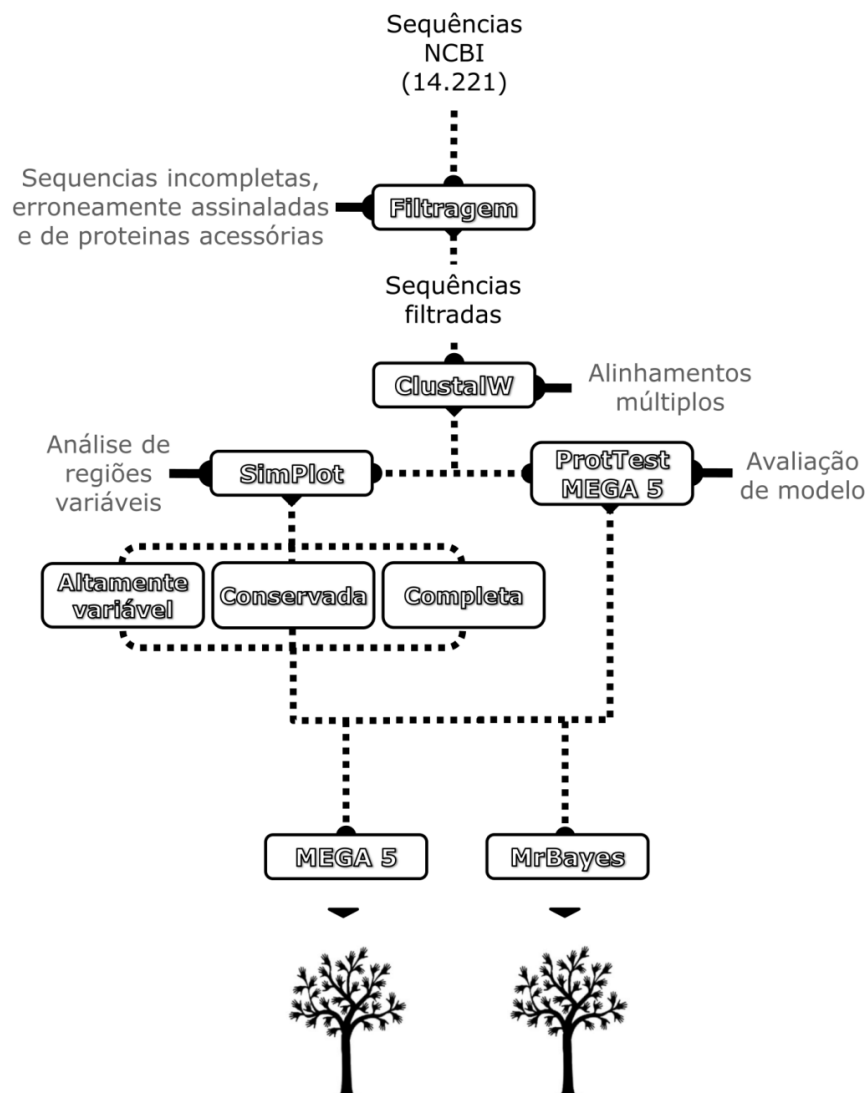


Figura 6: Fluxograma geral de trabalho.

4. Resultados e Discussão

A urease tem sido considerada uma enzima ancestral, relacionada ao ciclo peptídico primordial putativo³⁹. Essa enzima é encontrada nos três domínios da vida, sendo sintetizada em archaea, bactérias, fungos e plantas¹⁵. Há incerteza acerca da presença de ureases em animais, e, enquanto algumas descobertas indicam sua presença em certos invertebrados⁴⁰, outras indicam que a enzima é oriunda de fontes externas⁴¹ e que todos os animais perderam a urease em sua história evolutiva⁴². Por essa razão, as ureases de fontes animais putativas não foram incluídas em nossos conjuntos de dados.

Usando ML e BI contra nossos alinhamentos (Figura 7), pudemos construir duas árvores filogenéticas para o conjunto de dados de ureases completas (Figura 8 e Figura 9) e duas para o conjunto de dados de regiões conservadas (Figura 10 e Figura 11). Para as informações sobre regiões altamente variáveis, apenas uma árvore foi gerada por ML (Apêndice B), com valores de *bootstrap* muito baixos. Em virtude da baixa confiabilidade dessa árvore, a mesma não foi utilizada nas análises subsequentes. Durante os cálculos por BI, não foi possível atingir convergência, sendo que nenhum resultado considerável pôde ser obtido das respectivas regiões.

Todas as árvores apresentaram similar convergência geral em relação às suas ramificações, com pequenas diferenças. Observa-se que, na maioria dos casos, sequências pertencentes a organismos de mesmo filo agruparam-se. Todas as inferências mostraram padrões de distribuição afins em relação a ureases de cinco grupos microbianos compostos de Euryarchaeota, Firmicutes, Actinobacteria, Proteobacteria e Cyanobacteria. Além disso, todas as inferências filogenéticas sugerem que o número de cadeias das ureases evoluiu de maneira 3-para-1: ureases de três cadeias (presentes na maioria dos micróbios) existiram primeiramente, ocorrendo posterior unificação estrutural originando ureases de uma cadeia (presentes em plantas e fungos). Ureases de duas cadeias, pertencentes a Helicobacteraceae, apresentam-se como situações especiais dentre as enzimas de três cadeias, em oposição à hipótese que as considerava intermediários entre ureases de uma e três cadeias.

Anteriormente, as tentativas publicadas de estabelecer a história evolutiva das ureases basearam-se somente na subunidade catalítica de ureases microbianas⁴³, a enzima eucariótica completa⁴⁴, ou associações de filogenias em nível de organismo com características de enzimas⁴⁵. Neste trabalho, pudemos reconstruir a possível rota evolutiva seguida pelas ureases na sua transição estrutural de múltiplas cadeias para cadeias simples, de acordo com a “hipótese da fusão”³⁰. Pode-se argumentar que essa hipótese já é apoiada pela aplicação da navalha de Ockham⁴⁶, visto que as ureases mais abundantes e primitivas aparentam possuir três cadeias, e, portanto, menos propensas a terem todas sofrido fissão. Esta observação, somente, é muito

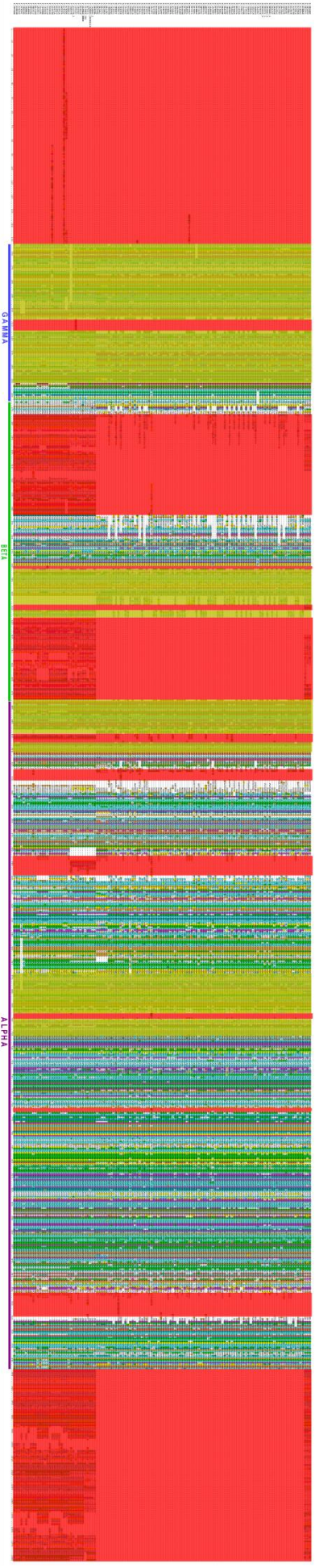


Figura 7: Alinhamento das sequências de ureases utilizadas. Regiões marcadas em vermelho foram removidas por conter gaps e regiões marcadas em amarelo foram removidas após análises de variabilidade com SimPlot. As regiões correspondentes às subunidades γ , β , e α estão marcadas de acordo.

simplicista para conferir validade ao pressuposto. Os resultados convergentes de quatro filogenias, construídas com base em dois métodos distintos sobre dois conjuntos de dados diferentes, fornecem suporte à conclusão parcimoniosa.

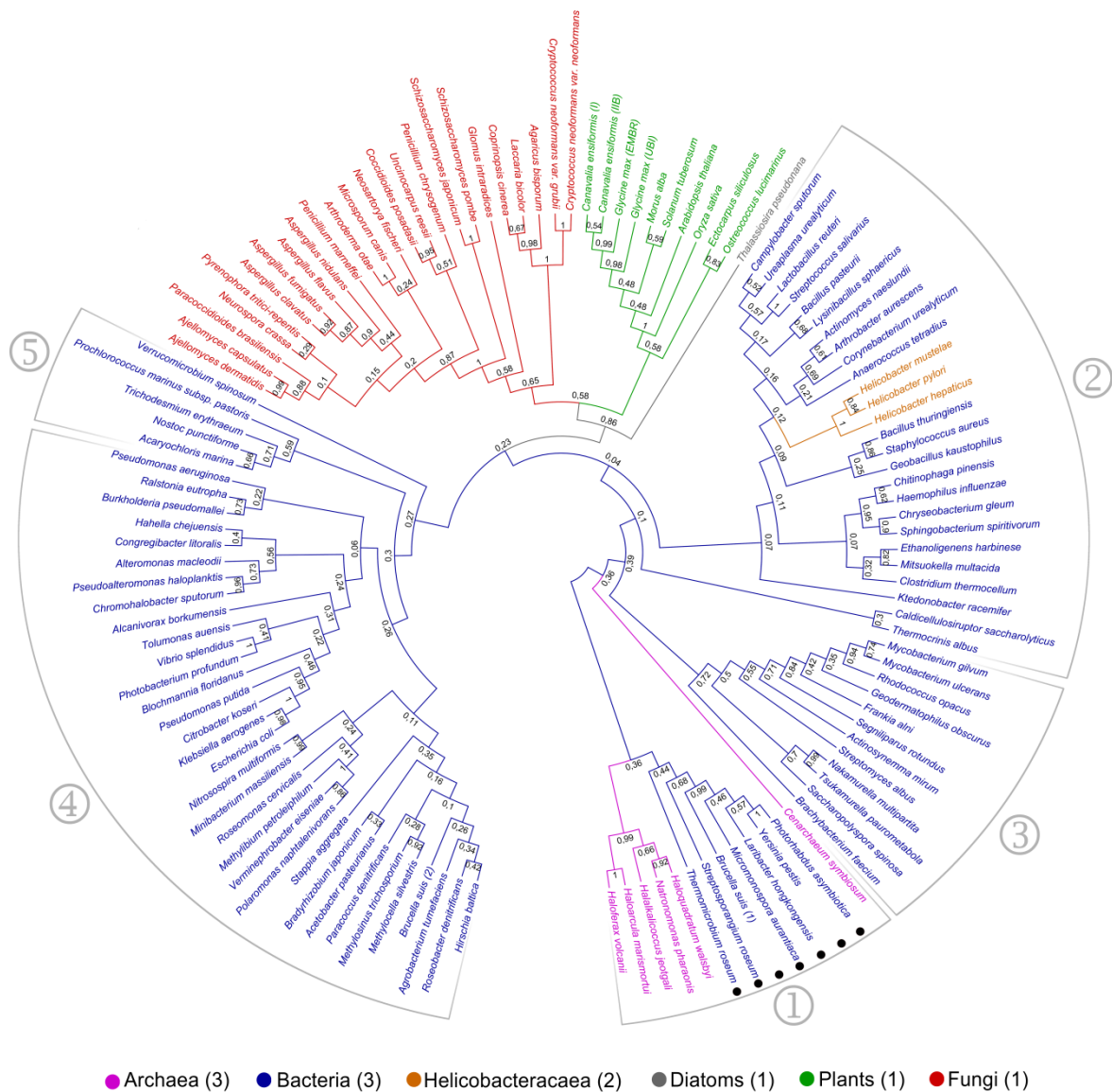


Figura 8: Análise filogenética molecular das seqüências de ureases completas pelo método de máxima verossimilhança. A história evolutiva foi inferida usando o método ML baseado no modelo WAG+G+I. O número de cadeias que compõem as ureases de diferentes grupos é apresentado entre parênteses. As separações gerais de filos microbianos estão marcadas em cinza (1 Euryarchaeota, 2 Firmicutes, 3 Actinobacteria, 4 Proteobacteria, 5 Cyanobacteria). Exceções nos grupos estão marcadas com pontos pretos.

Adicionalmente, essas árvores estão de acordo com análises prévias da organização do operon da urease em genomas microbianos completamente sequenciados²⁹. Os clados obtidos

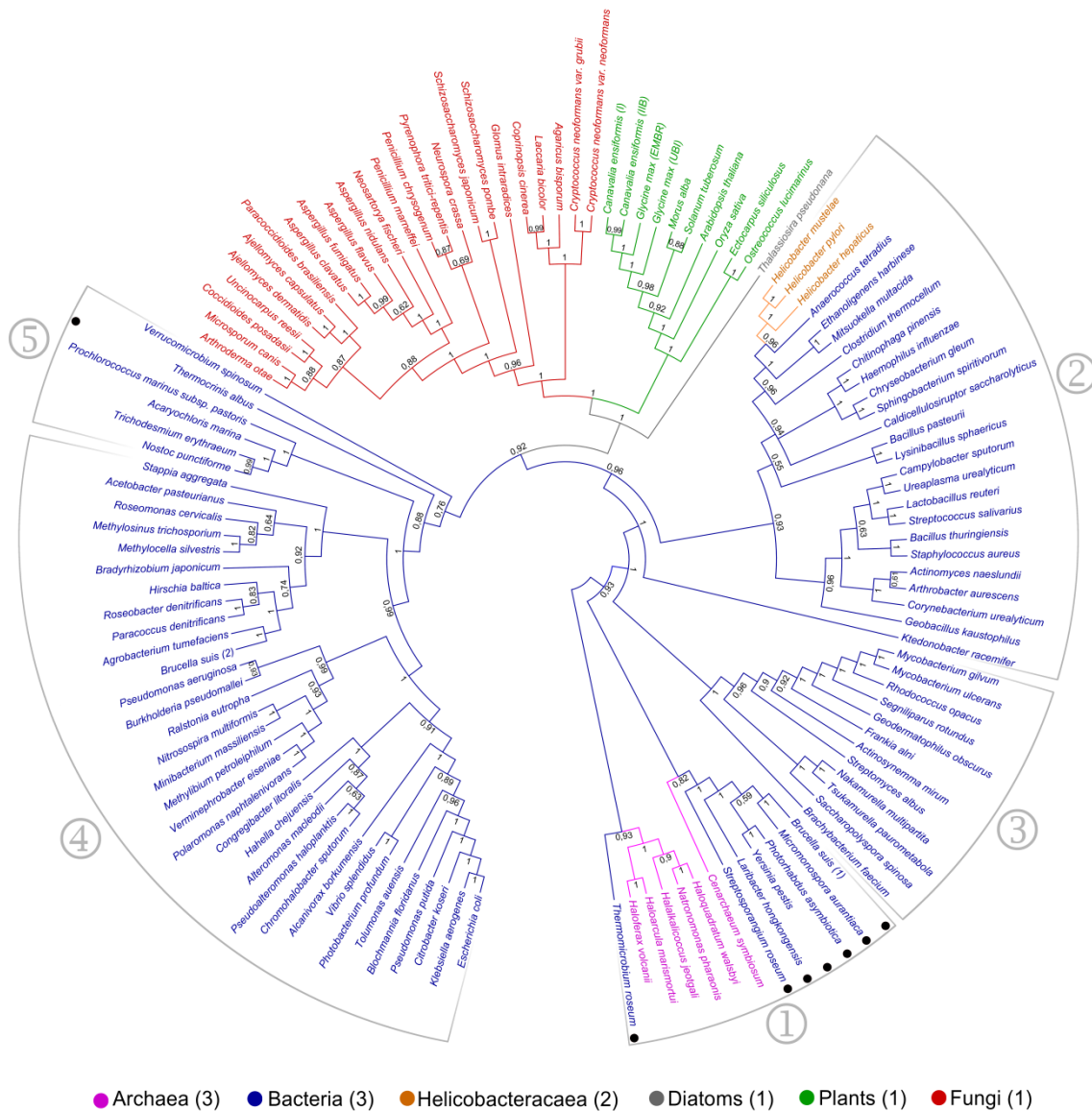


Figura 9: Análise filogenética molecular das seqüências de ureases completas pelo método de Inferência Bayesiana. A história evolutiva foi inferida usando o método Bayesiano baseado no modelo WAG+G+I. O número de cadeias que compõem as ureases de diferentes grupos é apresentado entre parênteses. As separações gerais de filos microbianos estão marcadas em cinza (1 Euryarchaeota, 2 Firmicutes, 3 Actinobacteria, 4 Proteobacteria, 5 Cyanobacteria). Exceções nos grupos estão marcadas com pontos pretos.

neste trabalho possuem operons *ure* com estruturas distintas: os representantes do Clado 1 (*H. marismortui*, *N. pharaonis*, *H. walsbyi*) estão organizados como UreBCAGDEF; os do Clado 2 (*C. thermocellum*, *H. influenzae*, *G. kaustophilus*, *S. aureus*, *C. urealyticum*, *A. aurescens*, *L. sphaericus*, *U. urealyticum*) estão organizados como UreABCEFGD, com a exceção de *H. pylori* e *H. hepaticus*, que estão organizados como Ure(AB)C-(gene não relacionado)-UreEFGD; o Clado 3 só possui um representante analisado (*M. gilvum*), organizado como UreABCGDEF; o

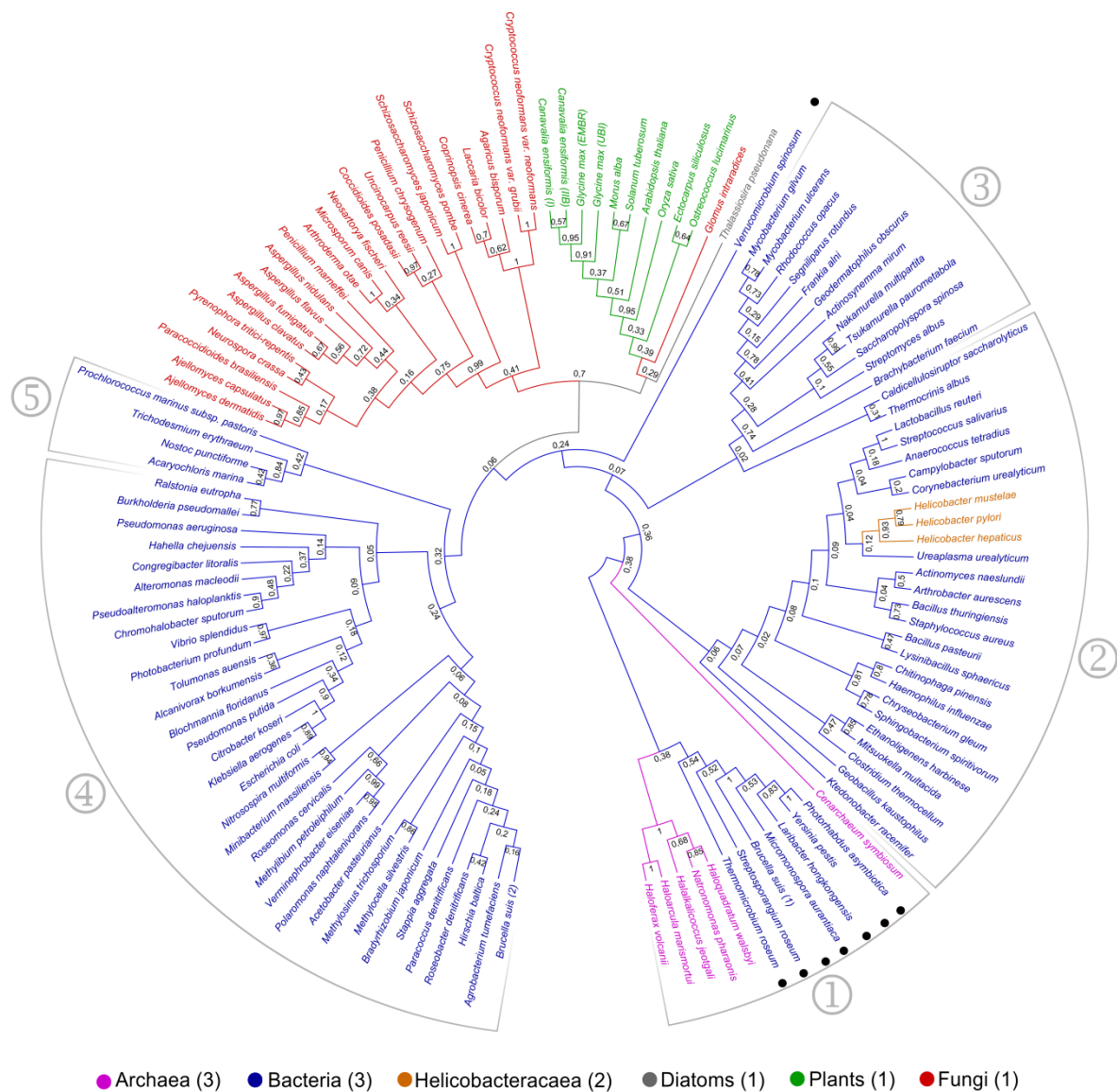


Figura 10: Análise filogenética molecular das regiões conservadas de seqüências de ureases pelo método de Máxima Verossimilhança. A história evolutiva foi inferida usando o método ML baseado no modelo WAG+G+I. O número de cadeias que compõem as ureases de diferentes grupos é apresentado entre parênteses. As separações gerais de filos microbianos estão marcadas em cinza (1 Euryarchaeota, 2 Firmicutes, 3 Actinobacteria, 4 Proteobacteria, 5 Cyanobacteria). Exceções nos grupos estão marcadas com pontos pretos.

Clado 4 tem montagens mistas, incluindo genes de urease não organizados em operons reconhecíveis (*A. marina*); e o Clado 5 (*P. aeruginosa*, *R. eutropha*, *B. pseudomallei*, *H. chejuensis*, *A. macleodii*, *P. haloplanktis*, *A. borkumensis*, *P. putida*, *E. coli*, *N. multiformis*, *P. naphthalivorans*, *P. denitrificans*, *M. silvestris*, *R. denitrificans*) está organizado tanto como UreDABCEFG, quanto UreDA-(gene não relacionado)-UreBCEFG (Figura 12).

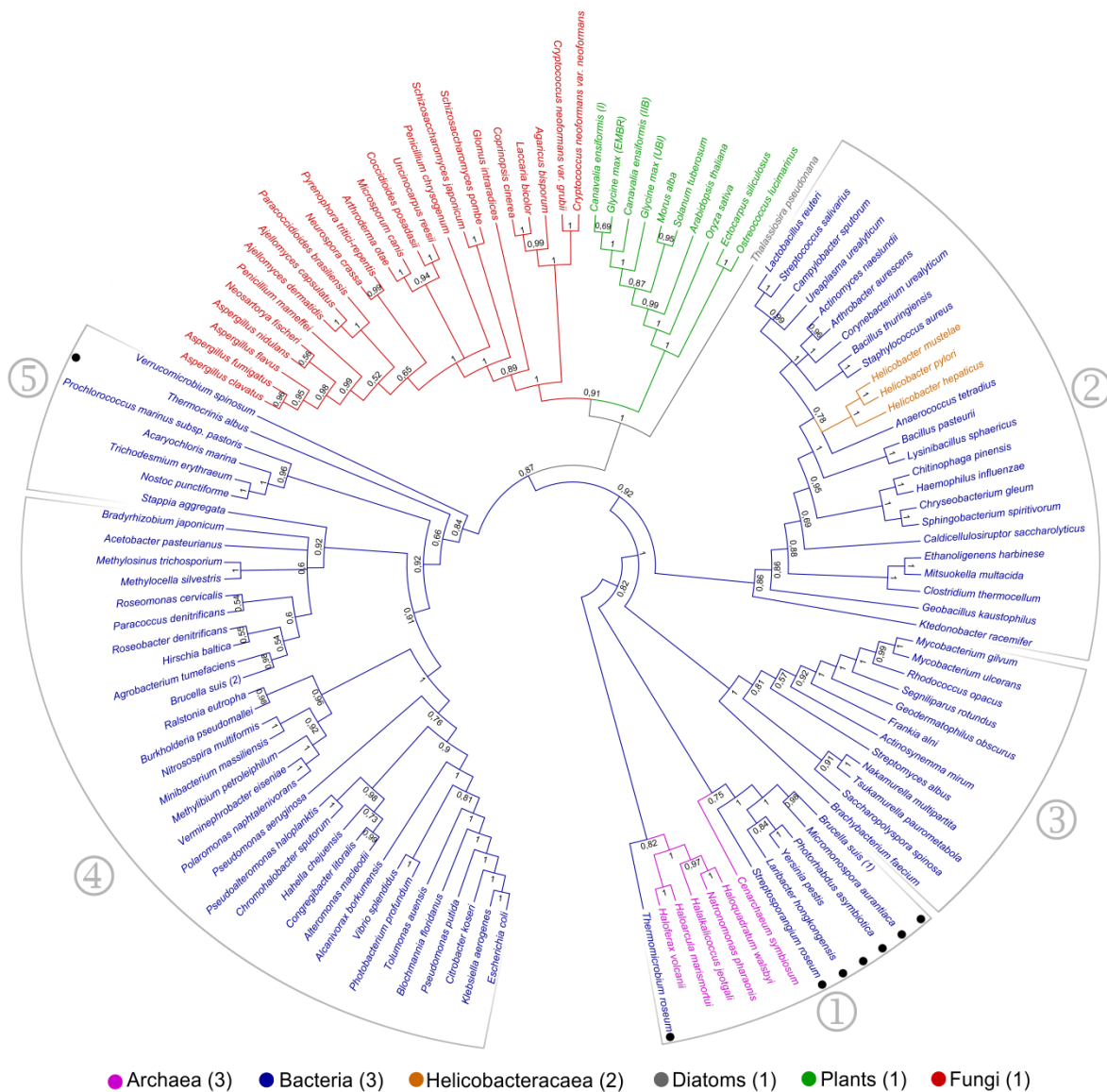


Figura 11: Análise filogenética molecular das regiões conservadas de seqüências de ureases pelo método de inferência Bayesiana. A história evolutiva foi inferida usando o método Bayesiano baseado no modelo WAG+G+I. O número de cadeias que compõem as ureases de diferentes grupos é apresentado entre parênteses. As separações gerais de filos microbianos estão marcadas em cinza (1 Euryarchaeota, 2 Firmicutes, 3 Actinobacteria, 4 Proteobacteria, 5 Cyanobacteria). Exceções nos grupos estão marcadas com pontos pretos.

A enzima dihidroorotase é considerada ancestral a todas as amidohidrolases relacionadas à urease⁴⁷ e seria a seqüência ideal para enraizar as filogenias de urease. Porém, a dihidroorotase é muito mais curta que a urease, e alinhamentos incluindo essa seqüência interfeririam nas árvores resultantes. Por essa razão, uma raiz teve de ser escolhida dentre os clados nas árvores de urease. Esse enraizamento foi realizado selecionando manualmente o clado contendo Archaea como raiz de ponto médio, considerando a posição filogenética intermediária das archaea na

árvore da vida²¹, como forma de permitir comparações entre as diferentes árvores.

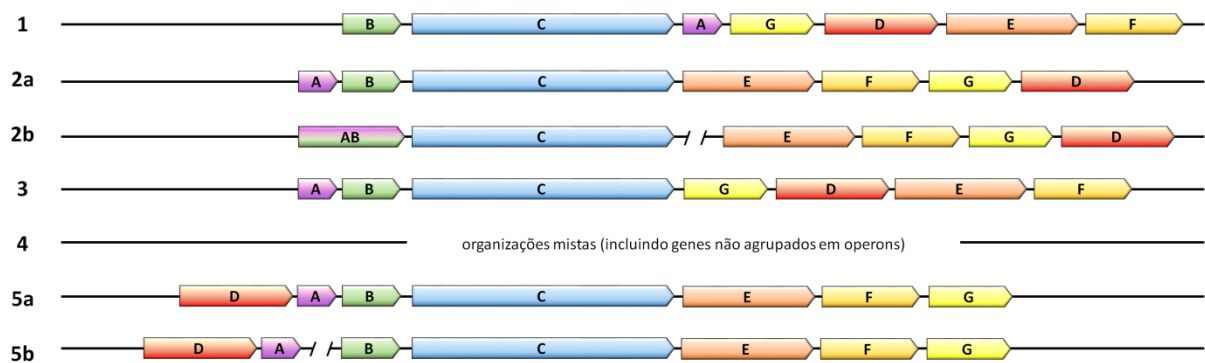


Figura 12: Organização dos operons ure das ureases microbianas estudadas. 1) Arranjo dos representantes do clado 1. 2a) Arranjo geral dos membros do clado 2, exceto 2b) para *H. hepaticus* e *H. pylori*. 3) Arranjo geral dos representantes do clado 3. 4) Organização dos membros do clado 4. 5a) e 5b) Arranjos observados para representantes do clado 5.

Quando enraizadas por ponto médio matemático⁴⁸, as árvores não apresentaram mudanças significativas em relação às enraizadas nas ureases de Archaea (Apêndices C,D,E e F). A inclusão das regiões altamente variáveis entre espécies nos alinhamentos se mostrou como uma fonte de “ruído molecular” nas análises filogenéticas. Tentativas de obter informação filogenética das regiões altamente variáveis, representadas na Figura 13, não foram estatisticamente confiáveis, enquanto a remoção dessas regiões causou pouco impacto nas árvores obtidas. Essas observações apontam para vantagens de se trabalhar com regiões menos variáveis dos alinhamentos, uma abordagem empregada para filogenias taxonômicas de fungos⁴⁹, artrópodes^{50,51}, cianobactérias⁵², e vírus⁵³.

A ocorrência de ureases de três cadeias em dois dos três domínios da vida (i.e. Bacteria e Archaea) indica que essa organização estrutural é muito difundida e pode ser considerada primitiva em relação a ureases de uma cadeia. As enzimas de duas cadeias, que poderiam ser tomadas com intermediários entre ureases de uma e três cadeias, aparentam ter surgido das cadeias triplas bacterianas, em um processo não correlato à origem das enzimas de uma cadeia. Bactérias do gênero *Helicobacter* spp. são conhecidamente sujeitas a distintas pressões seletivas relacionadas à sobrevivência no ambiente gástrico⁵⁴ e a organização molecular dodecamérica observada em ureases de *H. pylori* e *H. mustelae* aparenta ser uma resposta adaptativa a tais condições severas^{55,56}. A região conectora (ligando os equivalentes às subunidades γ e β) de *Helicobacter* spp., altamente divergente em relação à mesma região em eucariotos⁵⁵, pode tomar parte nessa organização diferencial.

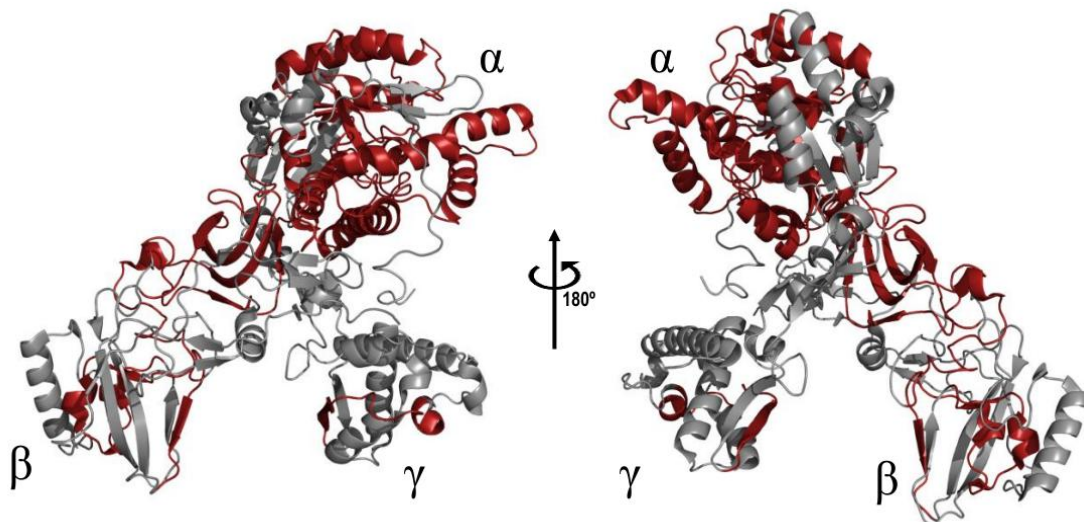


Figura 13: Regiões de variabilidade identificadas com SimPlot. Em vermelho, regiões conservadas nos alinhamentos. Em cinza, regiões altamente variáveis. Ressalta-se que a região do sítio ativo (subunidade α) mostrou-se altamente preservada em todos os casos. Os dados foram transpostos à estrutura tridimensional da urease de *Canavalia Ensiformis* (PDB ID 3LA4).

Há vantagens associadas à ligação de subunidades. Por meios de fusão genética artificial, foi mostrado que algumas proteínas virais oligoméricas são beneficiadas com taxa de enovelamento e estabilidade estrutural aprimoradas, e tolerância aumentada a inserções de outros segmentos^{57,58}. Quando artificialmente fundidas, as subunidades genéticas da citocromo ubiquinol oxidase de *E. coli* geraram uma enzima ativa, similar à de *Thermus thermophilus*⁵⁸. Diferentemente de fusões genéticas naturais em *Drosophila*⁵⁹, que resultaram em novos genes com novas funções, a junção de subunidades da urease em nível genético, incorporando segmentos conectores, manteve a atividade ureolítica original. Entretanto, as ureases são reconhecidas como proteínas multifuncionais (ou *moonlighting*). Elas possuem vários efeitos independente de catálise, incluindo neurotoxicidade para mamíferos, atividades inseticidas contra coleópteros e hemípteros, propriedades fungistáticas, papéis pró-inflamatórios, propriedades ligadoras de glicoconjugados, habilidade de ativação de plaquetas, e ação de comunicação interespecífica (revisado por Carlini & Polacco¹⁵ e Stanisquaski & Carlini⁶⁰). A maioria dessas propriedades ainda não foram mapeadas a regiões particulares dessas proteínas, sendo exceção um “domínio entomotóxico” contendo o(s) peptídeo(s) inceticida(s) liberado pelas ureases de *Canavalia ensiformis* frente à digestão por insetos⁶¹⁻⁶³. Este “domínio” está localizado na região intersubunidades β - α , e pode estar sujeito a taxas de divergência mais rápidas, visto que não está envolvida em catálise ou associação de subunidades⁶⁴. Também é intrigante a necessidade de subunidades/domínios não catalíticos na urease, considerando que somente o subdomínio barril-TIM da subunidade α é responsável pela catálise⁶⁵. Enquanto a

subunidade β foi associada ao processo de ativação da urease⁵⁶, desconhece-se alguma função específica da subunidade γ .

Revisando a literatura, não foi possível localizar outras proteínas que sofreram processo similar de fusão natural de subunidades. Os eventos observados para as ureases aparentam ser únicos, e não relacionados à fusão genética de imunoglobulinas⁶⁶ ou *exon shuffling*⁶⁷.

Linfócitos podem gerar receptores antigênicos de variedades praticamente infinitas, dentro do contexto de um genoma finito. A essa grande variabilidade se atribui o fenômeno denominado recombinação V(D)J⁶⁸. Nesse processo, as diversas versões dos segmentos V (*variable*), D (*diversity*) e J (*joining*), presentes em multiplicidade nas diversas partes codificantes da imunoglobulina (essas, por sua vez, encontram-se espalhadas em cromossomos distintos), são fundidas de modo a formar o gene final (Figura 14). No processo de rearranjo do DNA, apenas uma cópia de cada segmento é incorporada^{66,68}.

O processo de *exon shuffling* (Figura 15) caracteriza-se pela criação de novas estruturas exon-intron, através da junção ectópica de dois ou mais exons provenientes de diferentes genes, ou até por meio de duplicação de um segmento gênico⁶⁹. Há evidências de que esse processo, também conhecido como *domain shuffling*, está frequentemente associado à criação de proteínas mosaico através de recombinação de sequências codificantes de domínios proteicos⁷⁰.

A principal diferença desse processo é que, para a maioria das ureases microbianas estudadas, as subunidades γ , β , e α são ordenadas e geneticamente adjacentes no mesmo operon²⁹, e não dispersas ao longo do genoma, como é o caso para precursores de muitos genes fundidos. É esperado que, dos diferentes mecanismos que podem levar a genes quiméricos, ambas as regiões inter e intragênicas possam ser igualmente afetadas por inserções de segmentos⁶⁷. Isso não parece ter ocorrido para as ureases, visto que somente regiões intergênicas foram incorporadas.

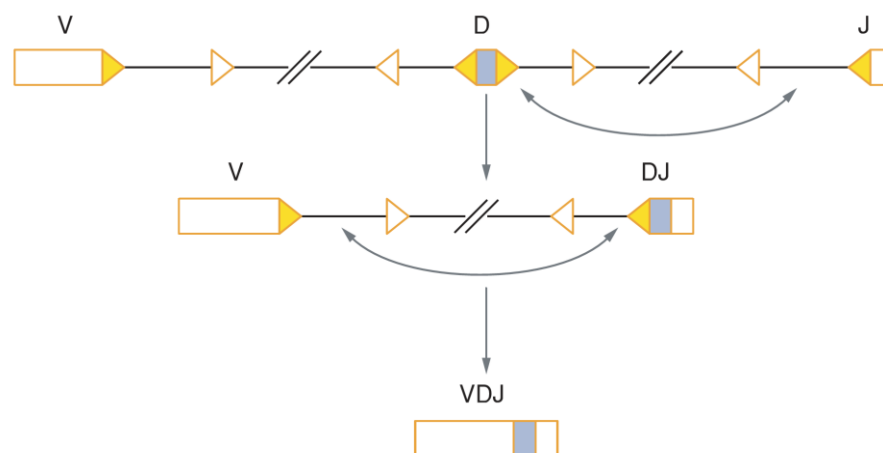


Figura 14: Esquema do processo de fusão para imunoglobulinas. Triângulos representam sequências de sinal de recombinação. Adaptado de Jung et al.,2006⁶⁸.

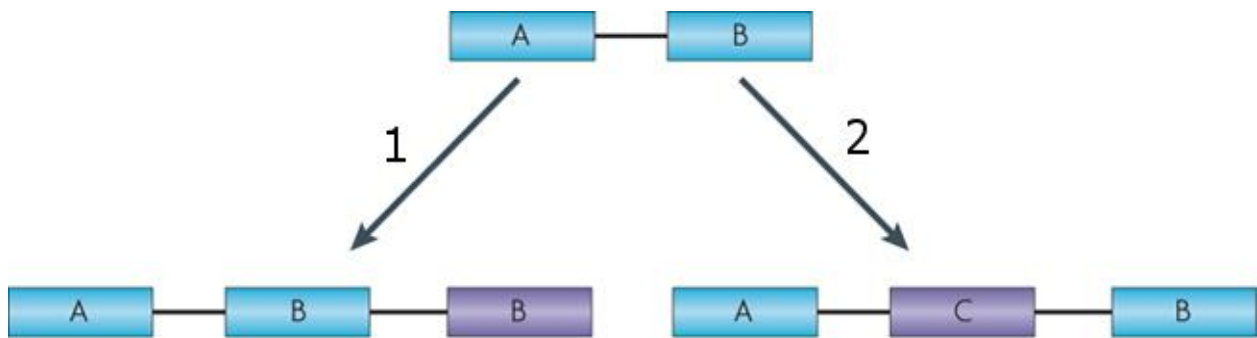


Figura 15: Esquema do processo de *exon shuffling*. Exons constitutivos representados em azul, regiões de *splicing* alternativo, em roxo, e introns representados pelas linhas sólidas. **1** Duplicação de exon dentro do mesmo gene. **2** Inserção exônica em um gene existente. Modificado de Keren et al., 2010⁷¹.

Das filogenias, infere-se que a transição de três subunidades para uma unidade ocorreu em evento único. Um mecanismo que pode se o responsável para tal resultado seria a “sobreleitura” (do inglês, *readthrough*) de transcrição ou tradução, a qual ignora códons de parada, incorporando regiões intergênicas como sequências codificantes. A ignora de stop codons da tradução é bem documentada em leveduras, onde toma parte em mecanismos regulatórios complexos ⁷², enquanto a sobreleitura transcricional vem sendo associada em desordens genéticas humanas ⁷³, respostas de plantas a stress ⁷⁴, e adenocarcinoma prostático, onde muitas quimeras induzidas por transcrição foram encontradas ⁷⁵. Para as ureases, contudo, somente eventos de sobreleitura não explicariam como os genes fundidos acabaram por ser incorporados nos genomas de seus organismos de origem, sendo necessário(s) outro(s) processo(s) subsequente(s) a nível traducional e transcricional. Nós especulamos que diferentes códigos genéticos seriam responsáveis por códons de parada não serem reconhecidos de acordo, permitindo transcrição contínua do complexo $\gamma\beta\alpha$ em uma única cadeia. Tal candidato seria o código genético mitocondrial da Chlorophyceae *Scenedesmus obliquus*, que interpreta UCA como códon de parada ao invés de codificar serina, conforme o código genético padrão ⁷⁶. Nesse cenário, para os códigos intergênicos serem traduzidos, seria necessário que os genes da urease fossem transferidos do genoma mitocondrial para o genoma nuclear, onde códons de parada não seriam reconhecidos como tais.

Há, entretanto, algumas dificuldades seguindo essa hipótese. Nenhum códon UCA é encontrado na posição terminal da região γ das enzimas de plantas e fungos, e nenhuma serina evidente é encontrada no terminal da cadeia β das mesmas ureases. Acerca dos códons de serina no terminal da cadeia γ , pode-se argumentar que uma serina seria benéfica naquela posição, logo, permitindo somente mutações conservativas do códon originalmente mal-interpretado. A respeito

da localização celular, a urease é geralmente considerada uma proteína citoplasmática, mas também foi encontrada em frações de membrana e parede celular de plantas ⁷⁷. Estudos proteômicos também indicam que a urease de *C. ensiformis* ou é ligada à mitocôndria, ou espacialmente relacionada a proteínas mitocondriais ⁷⁸.

Uma transferência horizontal inter-específica (Figura 16) também pode ser responsável pela sobreleitura hipotética. Essas transferências são consideradas uma grande ferramenta de modelagem do genoma, quando envolvendo elementos transponíveis ⁷⁹, e muitos casos de transferência de genes de bactérias para eucariotos foram documentados, incluindo transferências organela-para-núcleo, tais como o inserto mitocondrial transferido para o cromossomo 2 de *A. thaliana* ⁸⁰. Poucos dados estão disponíveis sobre transferências genéticas interespecíficas envolvendo ureases, e o único caso descrito até o momento é o segundo *cluster* de genes *ure* de *Brucella suis* ⁴³.

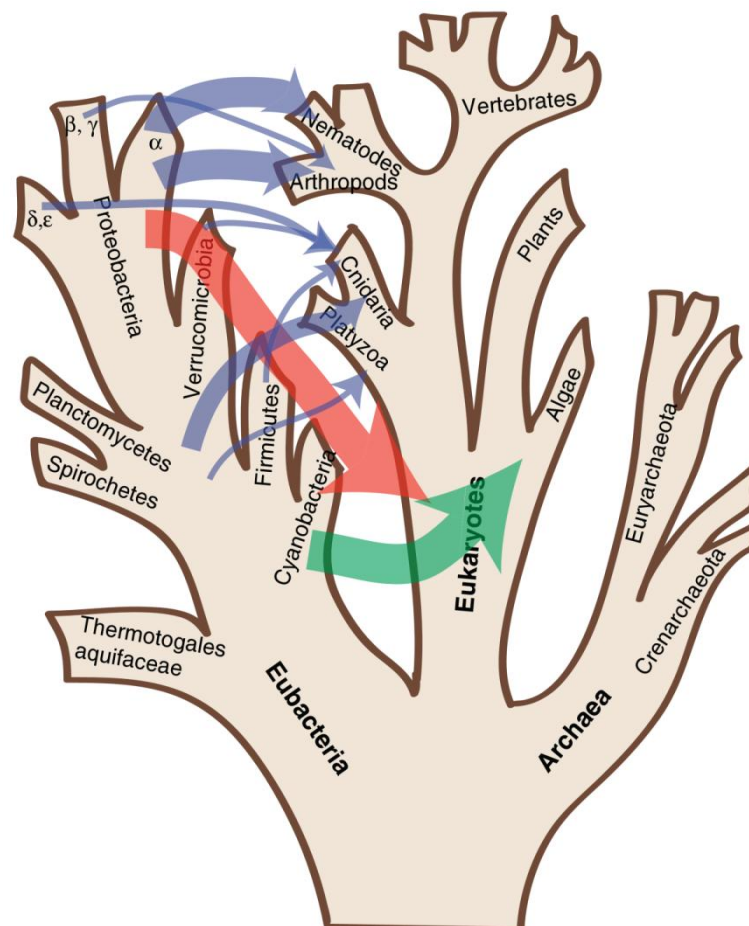


Figura 16: Transferência horizontal de genes. Eventos de transferência de bactérias para animais representados por setas azuis, cujas larguras representam a abundância de transferências. Transferências organelares (verde para cloroplastos e vermelho para mitocôndrias) estão mostradas na base da linhagem apropriada. A árvore em si serve a fins de representação, não ilustrando uma filogenia verdadeira. Modificado de Hotopp, 2011 ⁸⁰.

A hipótese de sobreleitura de código genético também demandaria regiões intergênicas de tamanho suficiente para compensar os segmentos incorporados. Operons de urease provenientes de algumas bactérias relacionadas ao Clado 2, tais como *Staphylococcus saprophyticus*, *Streptococcus thermophilus*, e *Corynebacterium glutamicum* (Gene IDs 3616069, 3167116, 1021080, respectivamente), possuem regiões intercistrônicas que satisfariam essa necessidade. Em outros casos, exemplificados pela urease de *Proteus mirabilis*, não há códon suficientes nessas regiões. Pelo contrário, há, ainda, sobreposição do último códon da subunidade β com o códon inicial da subunidade α ⁸¹. Quando a organização gênica eucariótica é levada em consideração, o quadro se torna ainda mais complexo. Uma inspeção preliminar de genomas anotados no banco de dados Ensembl⁸², revelou que todos os genes de urease de plantas depositados, i.e. *Arabidopsis thaliana*, *Brassica rapa*, *Glycine max* (ambas as isoformas), *Oryza sativa*, *Physcomitrella patens*, *Populus trichocarpa*, *Sorghum bicolor*, e *Vitis vinifera*, tem suas regiões codificantes organizadas em 18 exons. O mesmo número de exons é observado para *Solanum tuberosum*, o primeiro (e, até agora, único) caso de *splicing* alternativo em ureases⁸³. Para genes fúngicos de ureases depositados no Ensembl, o número de éxons é variável, abrangendo desde uma única sequência codificante sem íntrons (*Magnaporthe poae*, *Ustilago maydis*, *Schizosaccharomyces pombe*), até um número variável de exons: 3 (*Fusarium oxysporum*, *Gaeumannomyces graminis*, *Gibberella moniliformis*, *Magnaporthe oryzae*), 4 (*Aspergillus fumigatus*, *Neurospora crassa*), 5 (*A. fumigatus*, *A. niger*, *Nectria haematococca*, *Gibberella zeae*), 6 (*A. flavus*, *A. nidulans*), 7 (*Neosartorya fischeri*, *Phaeosphaeria nodorum*), 8 (*Aspergillus terreus*), 9 (*Fusarium oxysporum*), 13 (*Puccinia graministritici*), 14 (*Gibberella zeae*) ou 15 (*P. graministritici*). A diferença entre o número de éxons de plantas e fungos pode ser reflexo da rigorosa conservação estrutural em plantas. Até que mais genomas sejam decifrados, permitindo uma análise mais profunda de seus genes codificantes de urease, essa interpretação permanece especulativa.

5. Conclusões e perspectivas

Através de análises filogenéticas em grande escala, pudemos rastrear a transição estrutural de três para uma subunidade em ureases, transição esta que não envolveu intermediários de duas cadeias.

Também especula-se que tal transição ocorreu em evento único, e hipotetiza-se um mecanismo que resultaria na urease fundida. No entanto, muitas questões permanecem sem resposta. De certa maneira, a descoberta das rotas evolutivas da ureases (o “que”) começa a ser estabelecida, enquanto os mecanismos subjacentes as transições estruturais da urease (o “como”) ainda aguardam investigação.

Espera-se que os grandes conjuntos de dados e múltiplas abordagens empregadas nesse trabalho contribuam para elevar a compreensão do caso único da fusão de subunidades da urease, encorajando pesquisas futuras acerca do tema.

Em virtude do complexo mecanismo de ativação da urease, necessitando da ligação de um grande número de proteínas acessórias⁸⁴, pretende-se dar continuidade ao presente estudo, expandindo as comparações ao maior número possível dessas proteínas. Com isso, busca-se avaliar como se sucedeu o processo evolutivo para essas proteínas em relação ao processo ímpar das ureases avaliado neste trabalho.

6. Referências

1. Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiological reviews* **59**, 451–80 (1995).
2. Krajewska, B. Ureases I. Functional, catalytic and kinetic properties: A review. *Journal of Molecular Catalysis B: Enzymatic* **59**, 9–21 (2009).
3. Wohler, F. On the artificial production of urea. *Annalen der Physik und Chemie* **88**, (1828).
4. Sumner, J. B. The Isolation and Crystallization of the Enzyme Urease. *The Journal of Biological Chemistry* **69**, 435–441 (1926).
5. Dixon, N. E., Gazzola, C., Blakeley, R. L. & Zerner, B. Jack bean urease (EC 3.5.1.5). Metalloenzyme. Simple biological role for nickel. *Journal of the American Chemical Society* **97**, 4131–4133 (1975).
6. Follmer, C., Wassermann, G. E. & Carlini, C. R. Separation of jack bean (*Canavalia ensiformis*) urease isoforms by immobilized metal affinity chromatography and characterization of insecticidal properties unrelated to ureolytic activity. *Plant Science* **167**, 241–246 (2004).
7. Follmer, C., Real-Guerra, R., Wasserman, G. E., Olivera-Severo, D. & Carlini, C. R. Jackbean, soybean and *Bacillus pasteurii* ureases: biological effects unrelated to ureolytic activity. *European journal of biochemistry / FEBS* **271**, 1357–63 (2004).
8. Becker-Ritt, A. B. *et al.* Antifungal activity of plant and bacterial ureases. *Toxicon: official journal of the International Society on Toxinology* **50**, 971–83 (2007).
9. Follmer, C. *et al.* Canatoxin, a toxic protein from jack beans (*Canavalia ensiformis*), is a variant form of urease (EC 3.5.1.5): biological effects of urease independent of its ureolytic activity. *The Biochemical journal* **360**, 217–24 (2001).
10. Jeffery, C. J. Moonlighting proteins--an update. *Molecular bioSystems* **5**, 345–50 (2009).
11. Jeffery, C. J. Moonlighting proteins. *Trends in Biochemical Sciences* **24**, 8–11 (1999).
12. Huberts, D. H. E. W. & van der Klei, I. J. Moonlighting proteins: an intriguing mode of multitasking. *Biochimica et biophysica acta* **1803**, 520–5 (2010).
13. Jabri, E., Carr, M. B., Hausinger, R. P. & Karplus, P. A. The crystal structure of urease from *Klebsiella aerogenes*. *Science (New York, N.Y.)* **268**, 998–1004 (1995).
14. Sirko, a & Brodzik, R. Plant ureases: roles and regulation. *Acta biochimica Polonica* **47**, 1189–95 (2000).
15. Carlini, C. R. & Polacco, J. C. Toxic Properties of Urease. *Crop Science* **48**, 1665 (2008).
16. Miyaki, C. Y., Russo, C. A. de M. & Pereira, S. L. in *Biologia Molecular e Evolução* 113–122 (Holos Editora, 2012).

17. Matioli, S. R. & Fernandes, F. M. de C. *Biologia Molecular e Evolução*. 256 (Holos Editora, 2012).
18. Pereira, S. L., Miyaki, C. Y. & Russo, C. A. de M. in *Biologia Molecular e Evolução* 133–145 (Holos Editora, 2012).
19. Russo, C. A. de M., Miyaki, C. Y. & Pereira, S. L. in *Biologia Molecular e Evolução* 123–131 (Holos Editora, 2012).
20. Eck, R. V. & Dayhoff, M. O. Atlas of Protein Sequence and Structure. *National Biomedical Research Foundation* (1966).
21. Woese, C. R. Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* **87**, 4576–4579 (1990).
22. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics*. (Oxford University Press, 2000).
23. Pereira, S. L. in *Biologia Molecular e Evolução* 147–156 (Holos Editora, 2012).
24. Holder, M. & Lewis, P. O. Phylogeny estimation: traditional and Bayesian approaches. *Nature reviews. Genetics* **4**, 275–84 (2003).
25. Karplus, P. A., Pearson, M. A. & Hausinger, R. P. 70 Years of Crystalline Urease: What Have We Learned? *Accounts of Chemical Research* **30**, 330–337 (1997).
26. Benini, S. *et al.* A new proposal for urease mechanism based on the crystal structures of the native and inhibited enzyme from *Bacillus pasteurii*: why urea hydrolysis costs two nickels. *Structure* **7**, 205–216 (1999).
27. Estiu, G. & Merz, K. M. Competitive hydrolytic and elimination mechanisms in the urease catalyzed decomposition of urea. *The journal of physical chemistry. B* **111**, 10263–74 (2007).
28. Carter, E. L., Flugga, N., Boer, J. L., Mulrooney, S. B. & Hausinger, R. P. Interplay of metal ions and urease. *Metallomics : integrated biometal science* **1**, 207–21 (2009).
29. Zambelli, B., Musiani, F., Benini, S. & Ciurli, S. Chemistry of Ni²⁺ in urease: sensing, trafficking, and catalysis. *Accounts of chemical research* **44**, 520–30 (2011).
30. Hausinger, R. P. *Biochemistry of Nickel*. 23–57 (Plenum Press, 1993).
31. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic acids research* **40**, D13–25 (2012).
32. Larkin, M. a *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* **23**, 2947–8 (2007).
33. Lole, K. S. *et al.* Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J. Virol.* **73**, 152–160 (1999).

34. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)* **21**, 2104–5 (2005).
35. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731–9 (2011).
36. Whelan, S. & Goldman, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution* **18**, 691–699 (2001).
37. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**, 754–5 (2001).
38. Rambaut, A. FigTree. (2012). at <<http://tree.bio.ed.ac.uk/software/figtree/>>
39. Huber, C., Eisenreich, W., Hecht, S. & Wächtershäuser, G. A possible primordial peptide cycle. *Science (New York, N.Y.)* **301**, 938–40 (2003).
40. Pedrozo, H. A. *et al.* A mechanism of adaptation to hypergravity in the statocyst of *Aplysia californica*. *Hearing Research* **102**, 51–62 (1996).
41. Hirayama, C., Sugimura, M., Saito, H. & Nakamura, M. Host plant urease in the hemolymph of the silkworm, *Bombyx mori*. *Journal of Insect Physiology* **46**, 1415–1421 (2000).
42. Fujiwara, S. & Noguchi, T. Degradation of purines: only ureidoglycollate lyase out of four allantoin-degrading enzymes is present in mammals. *The Biochemical journal* **312** (Pt 1, 315–8 (1995).
43. Contreras-Rodriguez, A. *et al.* Enzymatic, immunological and phylogenetic characterization of *Brucella suis* urease. *BMC Microbiology* **8**, 121 (2008).
44. Mulinari, F. *et al.* Characterization of JBURE-IIb isoform of *Canavalia ensiformis* (L.) DC urease. *Biochimica et biophysica acta* **1814**, 1758–68 (2011).
45. Navarathna, D. H. M. L. P., Harris, S. D., Roberts, D. D. & Nickerson, K. W. Evolutionary aspects of urea utilization by fungi. *FEMS Yeast Research* **10**, 209–213 (2010).
46. Gernert, D. Ockham's Razor and Its Improper Use. *Journal of Scientific Exploration* **21**, 135–140 (2007).
47. Holm, L. & Sander, C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins* **28**, 72–82 (1997).
48. Hess, P. N. & De Moraes Russo, C. A. An empirical test of the midpoint rooting method. *Biological Journal of the Linnean Society* **92**, 669–674 (2007).
49. Camara, M. P. S., Palm, M. E., van Berkum, P. & O'Neill, N. R. Molecular phylogeny of *Leptosphaeria* and *Phaeosphaeria*. *Mycologia* **94**, 630–640 (2002).

50. Arango, C. P. Molecular approach to the phylogenetics of sea spiders (Arthropoda: Pycnogonida) using partial sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution* **28**, 588–600 (2003).
51. Hunt, T. & Vogler, A. P. A protocol for large-scale rRNA sequence analysis: towards a detailed phylogeny of Coleoptera. *Molecular phylogenetics and evolution* **47**, 289–301 (2008).
52. Gaylarde, P. M., Crispim, C. A., Neilan, B. A. & Gaylarde, C. C. Cyanobacteria from Brazilian building walls are distant relatives of aquatic genera. *Omics: a journal of integrative biology* **9**, 30–42 (2005).
53. Korber, B. Evolutionary and immunological implications of contemporary HIV-1 variation. *British Medical Bulletin* **58**, 19–42 (2001).
54. Gueneau, P. & Loiseaux De Goer, S. Helicobacter: molecular phylogeny and the origin of gastric colonization in the genus. *Infection, Genetics and Evolution* **1**, 215–223 (2002).
55. Ha, N. C. *et al.* Supramolecular assembly and acid resistance of Helicobacter pylori urease. *Nature structural biology* **8**, 505–9 (2001).
56. Carter, E. L. *et al.* Function of UreB in Klebsiella aerogenes urease. *Biochemistry* **50**, 9296–308 (2011).
57. Liang, H. Genetic Fusion of Subunits of a Dimeric Protein Substantially Enhances its Stability and Rate of Folding. *Proceedings of the National Academy of Sciences* **90**, 7010–7014 (1993).
58. Ma, J., Lemieux, L. & Gennis, R. B. Genetic fusion of subunits I, II, and III of the cytochrome bo ubiquinol oxidase from Escherichia coli results in a fully assembled and active enzyme. *Biochemistry* **32**, 7692–7697 (1993).
59. Jones, C. D. & Begun, D. J. Parallel evolution of chimeric fusion genes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 11373–8 (2005).
60. Stanisçuaski, F. & Carlini, C. R. Plant ureases and related peptides: understanding their entomotoxic properties. *Toxins* **4**, 55–67 (2012).
61. Ferreira-DaSilva, C. T., Gombarovits, M. E., Masuda, H., Oliveira, C. M. & Carlini, C. R. Proteolytic activation of canatoxin, a plant toxic protein, by insect cathepsin-like enzymes. *Archives of insect biochemistry and physiology* **44**, 162–71 (2000).
62. Piovesan, A. R. *et al.* Stage-specific gut proteinases of the cotton stainer bug Dysdercus peruvianus: role in the release of entomotoxic peptides from Canavalia ensiformis urease. *Insect biochemistry and molecular biology* **38**, 1023–32 (2008).
63. Defferrari, M. S., Demartini, D. R., Marcelino, T. B., Pinto, P. M. & Carlini, C. R. Insecticidal effect of Canavalia ensiformis major urease on nymphs of the milkweed bug Oncopeltus fasciatus and characterization of digestive peptidases. *Insect biochemistry and molecular biology* **41**, 388–99 (2011).

64. Mulinari, F. *et al.* Jaburetox-2Ec: an insecticidal peptide derived from an isoform of urease from the plant *Canavalia ensiformis*. *Peptides* **28**, 2042–50 (2007).
65. Balasubramanian, A. & Ponnuraj, K. Crystal structure of the first plant urease from jack bean: 83 years of journey from its first crystal to molecular structure. *Journal of molecular biology* **400**, 274–83 (2010).
66. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
67. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome research* **20**, 1313–26 (2010).
68. Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual review of immunology* **24**, 541–70 (2006).
69. Long, M., Betrán, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the young and old. *Nature reviews. Genetics* **4**, 865–75 (2003).
70. Kaessmann, H., Zöllner, S., Nekrutenko, A. & Li, W.-H. Signatures of domain shuffling in the human genome. *Genome research* **12**, 1642–50 (2002).
71. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nature reviews. Genetics* **11**, 345–55 (2010).
72. Von der Haar, T. & Tuite, M. F. Regulated translational bypass of stop codons in yeast. *Trends in microbiology* **15**, 78–86 (2007).
73. Du, L. *et al.* Nonaminoglycoside compounds induce readthrough of nonsense mutations. *The Journal of experimental medicine* **206**, 2285–97 (2009).
74. Hernández-Pinzón, I., de Jesús, E., Santiago, N. & Casacuberta, J. M. The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. *Journal of molecular evolution* **68**, 269–78 (2009).
75. Nacu, S. *et al.* Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC medical genomics* **4**, 11 (2011).
76. Nedelcu, A. M. The Complete Mitochondrial DNA Sequence of *Scenedesmus obliquus* Reflects an Intermediate Stage in the Evolution of the Green Algal Mitochondrial Genome. *Genome Research* **10**, 819–831 (2000).
77. Aguetoni Cambuí, C., Gaspar, M. & Mercier, H. Detection of urease in the cell wall and membranes from leaf tissues of bromeliad species. *Physiologia plantarum* **136**, 86–93 (2009).
78. Demartini, D. R., Carlini, C. R. & Thelen, J. J. Global and targeted proteomics in developing jack bean (*Canavalia ensiformis*) seedlings: an investigation of urease isoforms mobilization in early stages of development. *Plant molecular biology* **75**, 53–65 (2011).

79. Schaack, S., Gilbert, C. & Feschotte, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in ecology & evolution* **25**, 537–46 (2010).
80. Dunning Hotopp, J. C. Horizontal gene transfer between bacteria and animals. *Trends in genetics : TIG* **27**, 157–63 (2011).
81. Jones, B. D. & Mobley, H. L. *Proteus mirabilis* urease: nucleotide sequence determination and comparison with jack bean urease. *J. Bacteriol.* **171**, 6414–6422 (1989).
82. Flicek, P. *et al.* Ensembl 2012. *Nucleic acids research* **40**, D84–90 (2012).
83. Witte, C.-P., Tiller, S., Isidore, E., Davies, H. V & Taylor, M. A. Analysis of two alleles of the urease gene from potato: polymorphisms, expression, and extensive alternative splicing of the corresponding mRNA. *Journal of experimental botany* **56**, 91–9 (2005).
84. Ligabue-Braun, R., Real-Guerra, R., Carlini, C. R. & Verli, H. Evidence-based docking of the urease activation complex. *Journal of biomolecular structure & dynamics* (2012). doi:10.1080/07391102.2012.713782

APÊNDICES

APÊNDICE A: Identificadores GenInfo das sequências de ureases utilizadas.

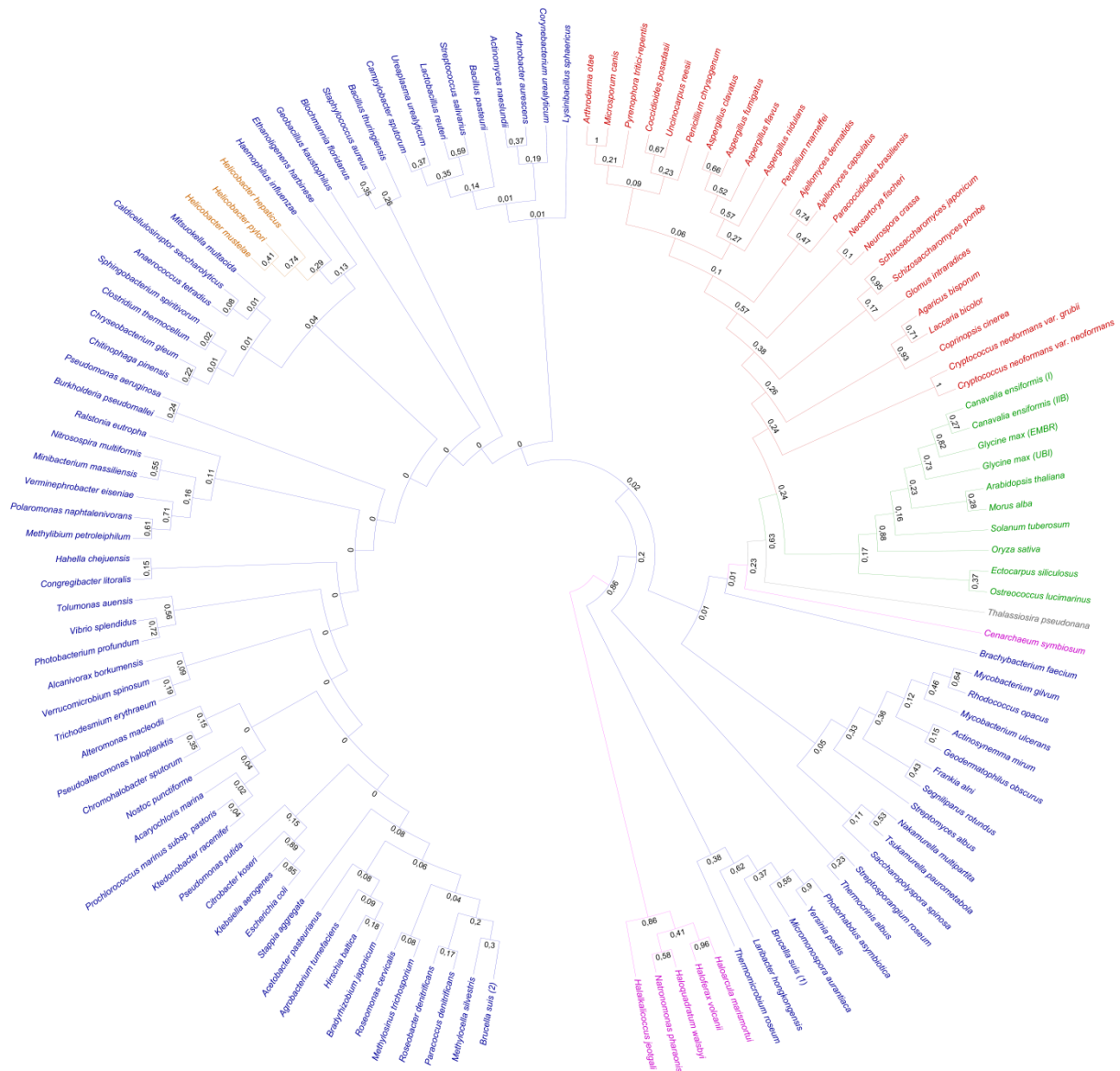
Organismo	GI		
	Alpha	Beta	Gamma
<i>Acaryochloris marina</i>	158338215	158338212	158338210
<i>Acetobacter pasteurianus</i>	258543403	258543404	258543405
<i>Actinomyces naeslundii</i>	4249601	4249599	326774128
<i>Actinosynnema mirum</i>	256374659	256374660	256374661
<i>Agaricus bisporus</i>	108859293		
<i>Agrobacterium tumefaciens</i>	15889672	159185200	159185201
<i>Ajellomyces capsulatus</i>	154281373		
<i>Ajellomyces dermatitidis</i>	261194902		
<i>Alcanivorax borkumensis</i>	110835580	110835581	110835582
<i>Alteromonas macleodii</i>	332142250	332142251	332142252
<i>Anaerococcus tetradius</i>	227500923	257066223	227484813
<i>Arabidopsis thaliana</i>	15220459		
<i>Arthrobacter aurescens</i>	119962058	119961176	119961081
<i>Arthroderma otae</i>	296818321		
<i>Aspergillus clavatus</i>	121702879		
<i>Aspergillus flavus</i>	238486420		
<i>Aspergillus fumigatus</i>	70990710		
<i>Aspergillus nidulans</i>	259489323		
<i>Bacillus pasteurii</i>	4557957	4557956	4557955
<i>Bacillus thuringiensis</i>	229061206	228909453	42782714
<i>Blochmannia floridanus</i>	33519971	33519972	33519973
<i>Brachybacterium faecium</i>	257067541	257067542	257067543
<i>Bradyrhizobium japonicum</i>	27376568	27376566	27376565
<i>Brucella suis</i>	225627811	237815759	225627809
<i>Brucella suis</i>	23501177	17987936	62289265
<i>Burkholderia pseudomallei</i>	53726283	53720268	53720267
<i>Caldicellulosiruptor saccharolyticus</i>	146297464	146297465	146297467
<i>Campylobacter sputorum</i>	260162332	260162331	260162330
<i>Canavalia ensiformis</i>	465008		
<i>Canavalia ensiformis</i>	219391588		
<i>Cenarchaeum symbiosum</i>	118575651	118575650	118575649
<i>Chitinophaga pinensis</i>	256420519	256420518	256420517
<i>Chromohalobacter salexigens</i>	92114426	92114427	92114428
<i>Chryseobacterium gleum</i>	300776841	300776840	300776839
<i>Citrobacter koseri</i>	157148630	157148629	237729986
<i>Clostridium thermocellum</i>	256003460	125974321	125974322
<i>Coccidioides posadasii</i>	303322831		
<i>Congregibacter litoralis</i>	88703327	88703326	88703325
<i>Coprinopsis cinerea</i>	299750141		
<i>Corynebacterium urealyticum</i>	172041452	172041451	172041450
<i>Cryptococcus neoformans var. grubii</i>	23822289		
<i>Cryptococcus neoformans var. neoformans</i>	58270418		
<i>Ectocarpus siliculosus</i>	299469707		

<i>Escherichia coli</i>	187775871	15800665	15800664
<i>Ethanoligenens harbinense</i>	317133580	317133581	317133582
<i>Frankia alni</i>	111220900	111220901	111220902
<i>Geobacillus kaustophilus</i>	56420465	56420466	56420467
<i>Geodermatophilus obscurus</i>	284988814	284988813	284988812
<i>Glomus intraradices</i>	297185890		
<i>Glycine max</i>	351722261		
<i>Glycine max</i>	351724331		
<i>Haemophilus influenzae</i>	260582757	16272484	16272485
<i>Hahella chejuensis</i>	83647213	83647212	83647211
<i>Halalkalicoccus jeotgali</i>	300710018	300710019	300710017
<i>Haloarcula marismortui</i>	55376690	55376689	55376691
<i>Haloferax volcanii</i>	292654327	292654326	292654328
<i>Haloquadratum walsbyi</i>	110669477	110669476	110669478
<i>Helicobacter hepaticus</i>	32265907	32265906	
<i>Helicobacter mustelae</i>	291277506	291276538	
<i>Helicobacter pylori</i>	208434038	15644703	
<i>Hirschia baltica</i>	254295077	254295078	254295080
<i>Klebsiella aerogenes</i>	10835900	3212375	152971982
<i>Ktedonobacter racemifer</i>	298249513	298249512	298249511
<i>Laccaria bicolor</i>	170098735		
<i>Lactobacillus reuteri</i>	194467060	194467061	194467062
<i>Laribacter hongkongensis</i>	226939965	226939964	226939963
<i>Lysinibacillus sphaericus</i>	169828213	169828212	169828211
<i>Methylibium petroleiphilum</i>	124265860	124265861	124265862
<i>Methylocella silvestris</i>	217978168	217978169	217978170
<i>Methylosinus trichosporium</i>	296447068	296447069	296447070
<i>Micromonospora aurantiaca</i>	315504551	302868979	302868980
<i>Microsporum canis</i>	238839950		
<i>Minibacterium massiliensis</i>	152980166	152980572	152980409
<i>Mitsuokella multacida</i>	255658020	260881031	255658018
<i>Morus alba</i>	222143560		
<i>Mycobacterium gilvum</i>	145223938	145223939	145223940
<i>Mycobacterium ulcerans</i>	183982730	118618421	183982728
<i>Nakamurella multipartita</i>	258650533	258650532	258650531
<i>Natronomonas pharaonis</i>	76801650	76801649	76801651
<i>Neosartorya fischeri</i>	119477773		
<i>Neurospora crassa</i>	85116050		
<i>Nitrosospora multiformis</i>	32966209	82702369	82702368
<i>Nostoc punctiforme</i>	186681311	186681310	186681309
<i>Oryza sativa</i>	17402589		
<i>Ostreococcus lucimarinus</i>	145343758		
<i>Paracoccidioides brasiliensis</i>	295673098		
<i>Paracoccus denitrificans</i>	119383953	119383954	119383956
<i>Penicillium chrysogenum</i>	255945993		
<i>Penicillium marneffeii</i>	212540166		

<i>Photobacterium profundum</i>	90414537	90414538	90414539
<i>Photorhabdus asymbiotica</i>	253989859	253989860	253989861
<i>Polaromonas naphthalenivorans</i>	121603883	121603884	121603886
<i>Prochlorococcus marinus subsp. pastoris</i>	33861519	33861520	33861521
<i>Pseudoalteromonas haloplanktis</i>	77360699	77360698	77360697
<i>Pseudomonas aeruginosa</i>	15600061	15600060	15600058
<i>Pseudomonas putida</i>	167033934	167033935	170721596
<i>Pyrenophora tritici-repentis</i>	189197501		
<i>Ralstonia eutropha</i>	113867104	113867103	73540695
<i>Rhodococcus opacus</i>	226365149	226365148	226365147
<i>Roseobacter denitrificans</i>	339502215	339502218	339502219
<i>Roseomonas cervicalis</i>	296536422	296536421	296536420
<i>Saccharopolyspora spinosa</i>	348172145	348172144	41350155
<i>Schizosaccharomyces japonicus</i>	213406373		
<i>Schizosaccharomyces pombe</i>	19115725		
<i>Segniliparus rotundus</i>	296394185	296394186	296394187
<i>Solanum tuberosum</i>	14599413		
<i>Sphingobacterium spiritivorum</i>	300772489	227537815	227537814
<i>Staphylococcus aureus</i>	15925280	257423771	15925278
<i>Stappia aggregata</i>	118591406	118591403	118591402
<i>Streptococcus salivarius</i>	2507522	2501628	2501633
<i>Streptomyces albus</i>	291450084	291450085	291450086
<i>Streptosporangium roseum</i>	271964227	271964226	271964224
<i>Thalassiosira pseudonana</i>	224014054		
<i>Thermocrinis albus</i>	289548494	289548493	289548481
<i>Thermomicrobium roseum</i>	221632556	221632557	221632558
<i>Tolumonas auensis</i>	237808376	237808377	237808378
<i>Trichodesmium erythraeum</i>	113474593	113474588	113474587
<i>Tsukamurella paurometabola</i>	296141089	296141090	296141091
<i>Uncinocarpus reesii</i>	258565811		
<i>Ureaplasma urealyticum</i>	171920953	185178985	167972857
<i>Verminephrobacter eiseniae</i>	121608691	121608692	121608694
<i>Verrucomicrobium spinosum</i>	171911815	171911816	171911817
<i>Vibrio splendidus</i>	84387642	84387641	84387640
<i>Yersinia pestis</i>	45442252	108808387	22125138

APÊNDICE B: Análise filogenética molecular de regiões altamente variáveis das sequências de ureases pelo método de Máxima Verossimilhança.

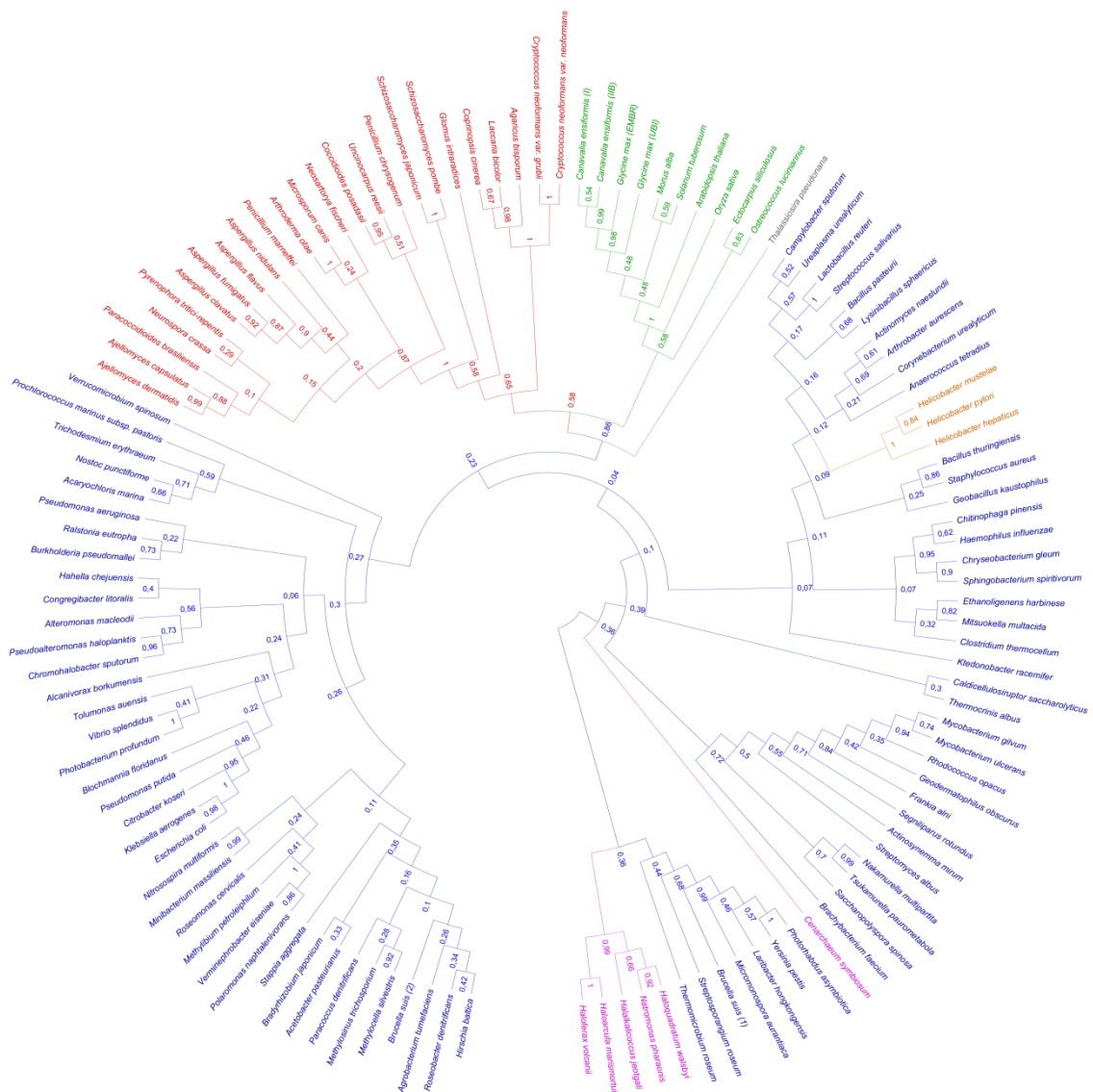
A história evolutiva foi inferida utilizando o método ML baseado no modelo WAG+G+I. O número de cadeias que compõem ureases de diferentes grupos estão em parênteses.



- Archaea (3)
- Bacteria (3)
- Helicobacteraceae (2)
- Diatoms (1)
- Plants (1)
- Fungi (1)

APÊNDICE C: Enraizamento por ponto médio matemático da árvore de seqüências completas de ureases pelo método de Máxima Verossimilhança.

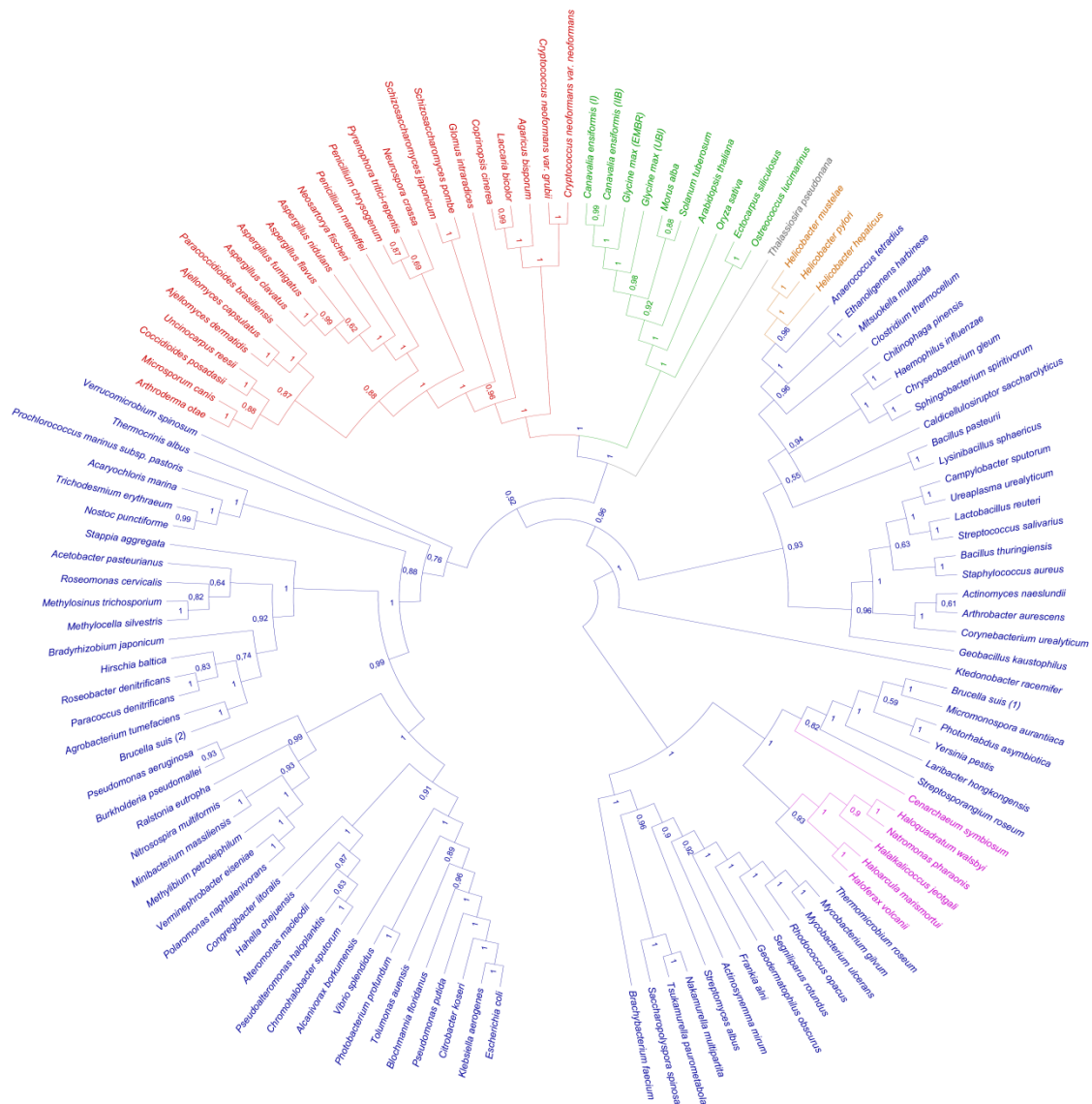
A história evolutiva foi inferida utilizando o método ML baseado no modelo WAG+G+I. O ponto médio matemático foi calculado com FigTree. O número de cadeias que compõem ureases de diferentes grupos estão em parênteses.



● Archaea (3) ● Bacteria (3) ● Helicobacteraceae (2) ● Diatoms (1) ● Plants (1) ● Fungi (1)

APÊNDICE D: Enraizamento por ponto médio matemático da árvore de seqüências completas de ureases pelo método de Inferência Bayesiana.

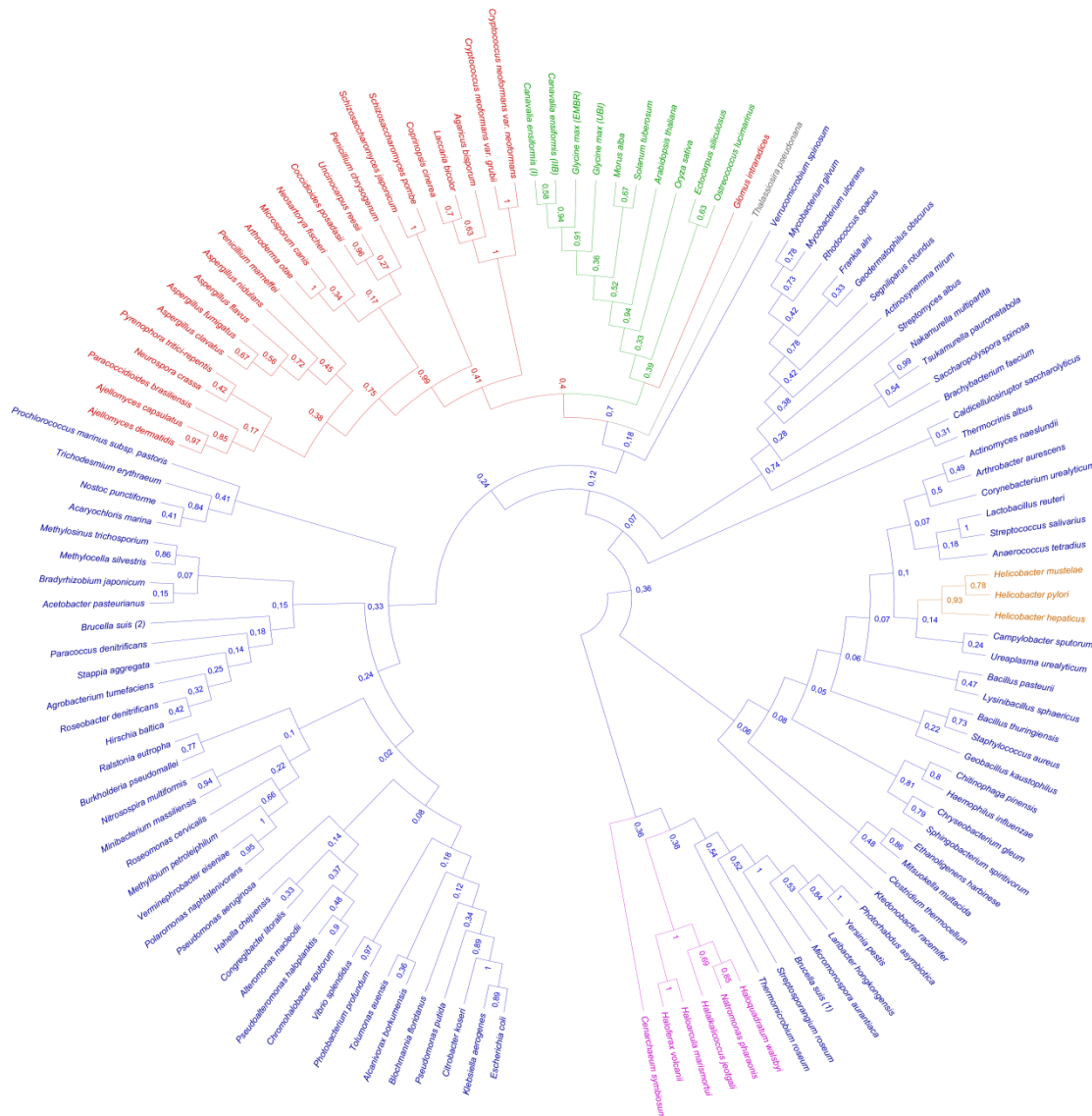
A história evolutiva foi inferida utilizando o método Bayesiano baseado no modelo WAG+G+I. O ponto médio matemático foi calculado com o FigTree. O número de cadeias que compõem ureases de diferentes grupos estão em parênteses.



● Archaea (3) ● Bacteria (3) ● Helicobacteraceae (2) ● Diatoms (1) ● Plants (1) ● Fungi (1)

APÊNDICE E: Enraizamento por ponto médio matemático da árvore de regiões conservadas de ureases pelo método de Máxima Verossimilhança.

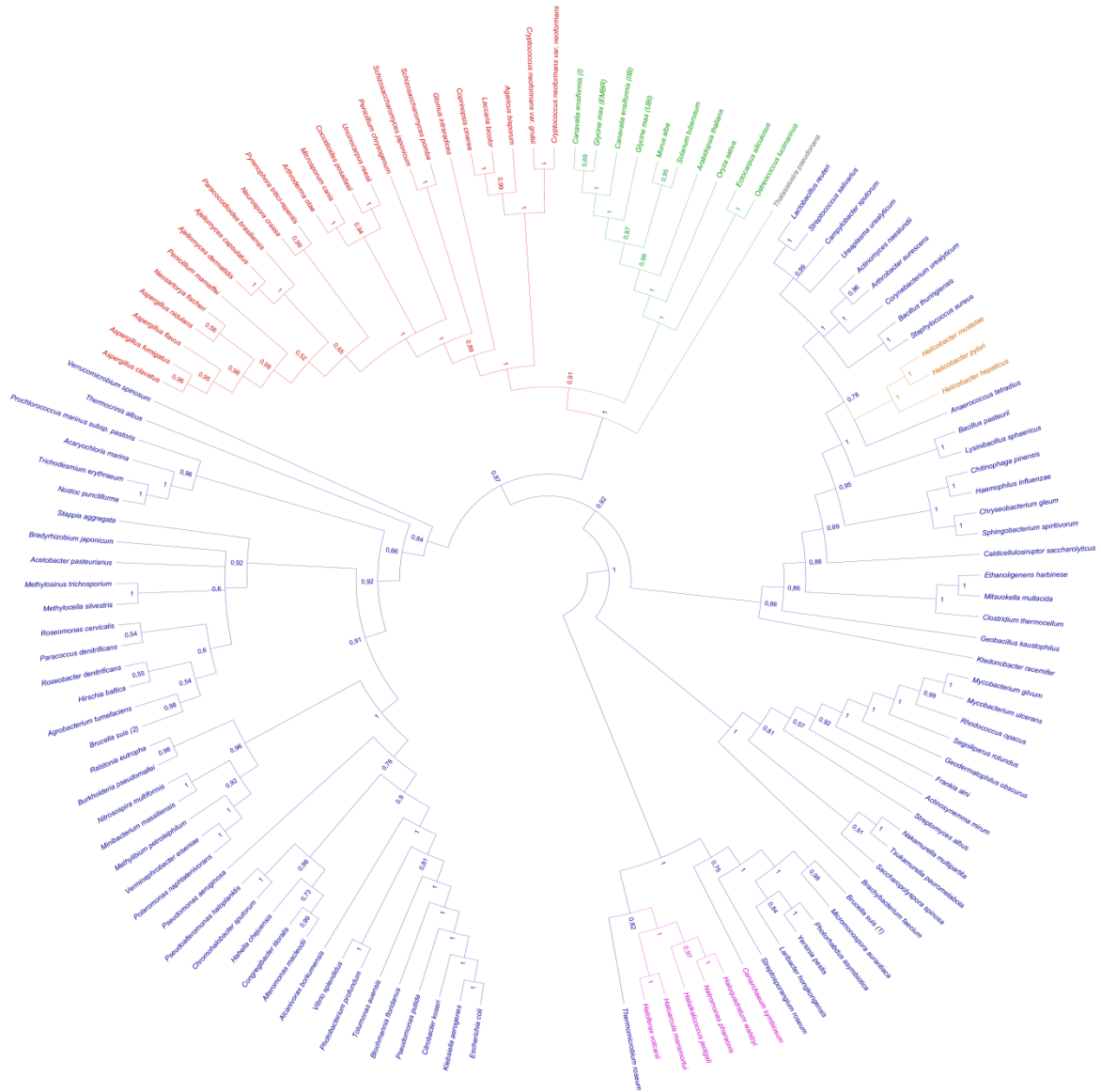
A história evolutiva foi inferida utilizando o método ML baseado no modelo WAG+G+I. O ponto médio matemático foi calculado com FigTree. O número de cadeias que compõem ureases de diferentes grupos estão em parênteses.



● Archaea (3) ● Bacteria (3) ● Helicobacteraceae (2) ● Diatoms (1) ● Plants (1) ● Fungi (1)

APÊNDICE F: Enraizamento por ponto médio matemático da árvore de regiões conservadas de ureases pelo método de Inferência Bayesiana.

A história evolutiva foi inferida utilizando o método Bayesiano baseado no modelo WAG+G+I. O ponto médio matemático foi calculado com FigTree. O número de cadeias que compõem ureases de diferentes grupos estão em parênteses.



● Archaea (3) ● Bacteria (3) ● Helicobacteraceae (2) ● Diatoms (1) ● Plants (1) ● Fungi (1)

APÊNDICE G: Publicação de artigo científico.

Os resultados contidos neste trabalho foram publicados sob o título de

3-to-1: unraveling structural transitions in ureases

no periódico *Naturwissenschaften* (Springer), fator de impacto 2,144 (2012). DOI 10.1007/s001 14-013-1045-2.

Autores: Rodrigo Ligabue-Braun, Fábio Carrer Andreis, Hugo Verli, Célia Regina Carlini.