

Comunicado 64

Técnico

Dezembro, 2004
Campinas, SP



ISSN 1677-8464

Modelagem molecular da proteína *small heat shock* de *Xylella fastidiosa*

Paula Regina Kuser Falcão¹
Susely Ferraz de Siqueira Tada²
Anete Pereira Souza³
Michel Eduardo Beleza Yamaghishi⁴
Adauto Luiz Mancini⁵
Renato Fileto⁶
Roberto Hiroshi Higa⁷
Goran Neshich⁸

As proteínas chamadas *heat shock* (HSP) pertencem a classe das chaperonas que são definidas como proteínas celulares que mediam o enovelamento correto de outras proteínas. As chaperonas também podem exercer a função de se associarem com macromoléculas, evitando uma associação com proteínas ainda não corretamente enoveladas. As proteínas *heat shock* têm esta denominação porque são sintetizadas em grande quantidade quando a célula é submetida a altas temperaturas ($> 42^\circ\text{C}$) e situações de estresse metabólico, quando a maioria das proteínas celulares são desnaturadas. As HSPs são divididas em famílias de acordo com seu tamanho e sua sequência de aminoácidos. As proteínas *small heat shock* (smHSPs) pertencem a um grupo estruturalmente divergente dentro da superfamília das chaperonas, com peso molecular variando de 12 a 43kDa (Narberhaus, 2002; Crack et al., 2002; De Jong & Caspers, 1998). Sabe-se que estas proteínas contêm

uma região conservada ao longo da evolução chamada de domínio alfa-crystalino, localizada na região C-terminal. Este domínio conservado é importante para a associação das proteínas, para a termotolerância e para a atividade de chaperonas. A análise com espectroscopia de dicroísmo circular, que é um tipo de absorção espectrocópica que fornece informação sobre as estruturas secundárias das macromoléculas biológicas, indicou que sua estrutura secundária é composta principalmente de folhas beta e que seu terminal C é altamente flexível.

Apesar das propriedades interessantes da smHSP e de seu importante papel no mecanismo de enovelamento das proteínas, há pouca informação estrutural desta proteína. Este trabalho visa obter um modelo estrutural de uma nova smHSP do fitopatógeno *Xylella fastidiosa* (XF2234), expressa e purificada por pesquisadores do Centro de

¹ Ph.D. em Física Aplicada, Pesquisadora da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(e-mail: paula@cnptia.embrapa.br)

² Ph.D. em Pós-doc do Centro de Biologia Molecular e Engenharia Genética, Instituto de Biologia da Unicamp, Caixa Postal 6010 - 13083-970 - Campinas, SP.
(e-mail: susy_fs@yahoo.com.br)

³ Ph.D. em Biologia Molecular, Pesquisadora do Centro de Biologia Molecular e Engenharia Genética do Instituto de Biologia da Unicamp, Caixa Postal 6010 - 13083-970 - Campinas, SP. (anete@unicamp.br)

⁴ Ph.D. em Matemática, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(e-mail: michel@cnptia.embrapa.br)

⁵ Bacharel em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(e-mail: adauto@cnptia.embrapa.br)

⁶ Ph.D. em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(e-mail: fileto@cnptia.embrapa.br)

⁷ M.Sc. em Engenharia Elétrica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(e-mail: roberto@cnptia.embrapa)

⁸ Ph.D. em Biofísica, Pesquisador da Embrapa Informática Agropecuária, Caixa Postal 6041, Barão Geraldo - 13083-970 - Campinas, SP.
(e-mail: neshich@cnptia.embrapa.br)

Biologia Molecular e Genética do Instituto de Biologia da Universidade Estadual de Campinas (Azzoni et al., 2004). A *X. fastidiosa* é uma bactéria gram-negativa, responsável por uma variedade de doenças de importância econômica em diversas plantas. Uma das doenças causadas pela *X. fastidiosa* é a clorose variegada dos citros (CVC), que causa sérios prejuízos às culturas cítricas no Brasil (Frome et al., 2000; Lee et al., 1991; Laranjeira, 1997; Simpson et al., 2000). A clorose variegada dos citros é também conhecida pelo nome de "amarelinho", em virtude da clorose foliar generalizada que empresta um aspecto amarelado às plantas nos casos de ataques severos. O amarelinho afeta principalmente a qualidade dos frutos, tornando-os de tamanho reduzido para a comercialização "in natura" e mesmo para produção de suco cítrico concentrado (Fig. 1).

Este documento descreve o modelo molecular da proteína XF2234 e a análise preliminar da sua estrutura. O modelo foi construído por modelagem por homologia (ou modelagem comparativa) com a estrutura cristalográfica de uma proteína *small heat shock* de *Triticum aestivum* (trigo). A análise da estrutura tridimensional da proteína XF2234 tem o objetivo de contribuir para aumentar o conhecimento sobre o papel biológico das smHSPs, necessário para o combate à CVC.



FIG. 1. Frutos sadios ao lado de frutos de tamanho reduzido devido à CVC .
Fonte: Fundecitrus (2004).

Modelagem Molecular

É possível predizer a estrutura tridimensional de uma proteína através de uma técnica chamada modelagem por homologia, também conhecida como modelagem comparativa. Esta abordagem baseia-se em alguns fatos como:

- a) processo de evolução biológica das proteínas; homologia;
- b) entre seqüências de aminoácidos de proteínas, que implica em semelhança estrutural e funcional; e
- c) agrupamento das proteínas em um número limitado de famílias tridimensionais.

Quando se conhece a estrutura de pelo menos um representante de uma família, geralmente é possível modelar por homologia os demais membros da família. A modelagem por homologia implica basicamente em quatro passos sucessivos:

- a) identificação e seleção de proteínas molde;
- b) alinhamento da seqüência dos aminoácidos da proteína molde com a seqüência proteína alvo;

- c) construção das coordenadas da proteína alvo;
- d) validação e análise do modelo (Fig. 2).

A validação do modelo é feita através da verificação de sua estereoquímica e de comparações com os resultados bioquímicos e outros métodos como dicroísmo circular. Se o grau de identidade entre a seqüência da proteína molde e da proteína alvo for igual ou superior a cerca de 30%, quando o número de resíduos é superior a 80, existe grande probabilidade de que estas proteínas tenham estruturas tridimensionais semelhantes.

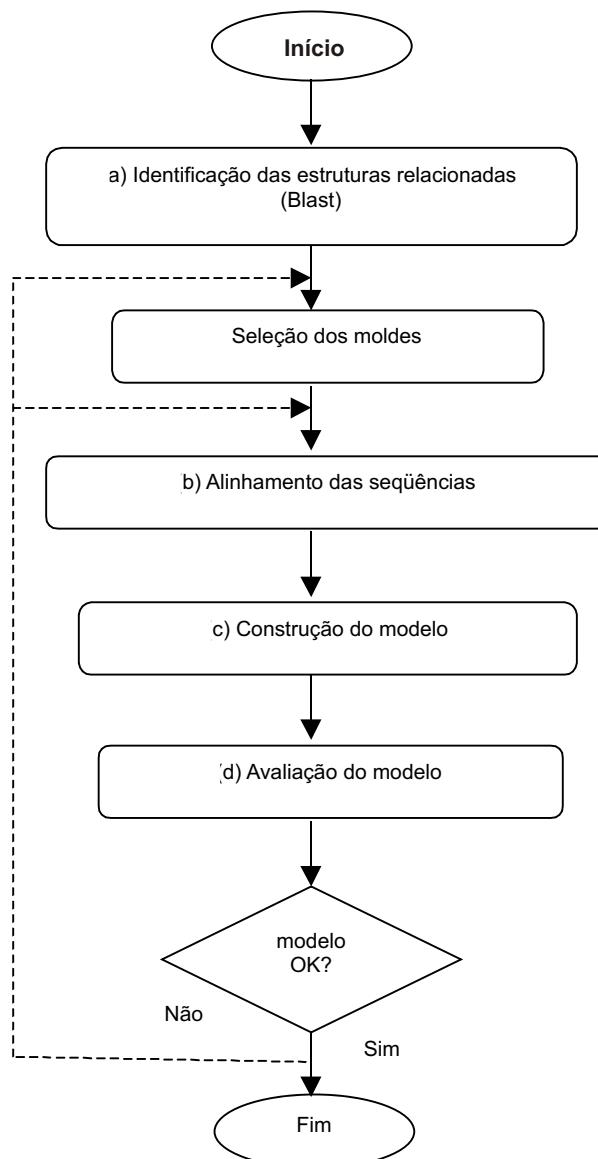


FIG. 2. Esquema geral da modelagem de proteínas por homologia.

A busca por uma proteína molde para a seqüência da XF2234 foi feita com o programa "*protein-protein Blast*" (National Center for Biotechnology Information, 2004; Altschul et al., 1990), utilizando-se o banco de dados de estruturas de proteínas PDB - Protein Data Bank (Berman et al., 2000). Nesta busca foram encontradas duas proteínas da família HSP que poderiam ser usadas como seqüências molde: a cadeia A da proteína com código pdb 1gme e a cadeia A da proteína 1shs, cujas identidades com a XF2234

são de 35% e 24%, respectivamente. A modelagem molecular da XF2234 foi feita com o programa Modeller (Sali & Blundell, 1993; Marti-Renom et al., 2000; Fiser et al., 2000), e a seqüência molde escolhida foi a 1gme, devido ao seu índice maior de identidade. Ambas as proteínas molde foram resolvidas por difração de raios X, com resolução 2.7 Angstrons para a 1gme, e 2.9 Angstrons para a 1shs.

O alinhamento é a entrada da etapa da modelagem juntamente com a estrutura da proteína molde. A saída do programa é um modelo tridimensional para a seqüência alvo que contém todos os átomos da cadeia principal e da cadeia lateral da proteína. No alinhamento das seqüências da proteína 1gme (modelo) e XF2234 (alvo) mostrado na Fig. 3, é possível verificar quais aminoácidos são idênticos e quais são conservados. As duas seqüências apresentam 35% de identidade e 58% de similaridade. Estes índices encontrados no alinhamento indicam que a estrutura cristalográfica de 1gme pode ser usada como estrutura molde.

O arquivo com o alinhamento das seqüências, juntamente com as coordenadas tridimensionais da proteína molde, são incluídos na próxima etapa do programa — a construção do modelo (geração das coordenadas cartesianas). O programa Modeller usa as coordenadas espaciais dos moldes para gerar um conjunto de restrições espaciais, que são aplicadas à seqüência alvo. As restrições são, por exemplo, a distância entre dois resíduos no modelo que está sendo construído, baseadas na distância entre dois resíduos homólogos do molde. As restrições também podem ser aplicadas aos ângulos de ligação, ângulos diedros e pares de diedros. Ao aplicar essas restrições espaciais, o programa Modeller limita efetivamente o número de configurações que o modelo pode assumir (Marti-Renom et al., 2000). A forma exata das restrições é baseada em uma análise estatística de diferenças entre pares de estruturas homólogas. As restrições espaciais baseadas em homologia não são as únicas restrições aplicadas ao modelo. Um campo de força para controlar a estereoquímica apropriada também é aplicado para que a estrutura do modelo não possa violar as regras da química para satisfazer as restrições espaciais derivadas do molde. Todas as restrições químicas e espaciais aplicadas ao modelo são combinadas em uma função (denominada função objetiva) que é otimizada no curso do processo de construção dos modelos.



FIG. 3. Alinhamento das seqüências da smHSP XP2234 e 1gme indicando os elementos de estrutura secundária. A seqüência de XP2234 tem 35% de identidade com a seqüência da cadeia A da proteína 1gme. O alinhamento foi obtido com o programa Modeller. Os aminoácidos que estão escritos em vermelho são aqueles que são idênticos nas duas seqüências (*), e aqueles que aparecem na cor azul são similares (:).

Desta modelagem inicial foram obtidos cinco modelos, que foram avaliados de acordo com a função objetiva e com os parâmetros estereoquímicos. O modelo preliminar obtido está ilustrado na Fig. 4, ao lado da proteína molde utilizada. É fácil observar que a estrutura tridimensional foi quase que totalmente preservada, com uma parte central constituída de uma folha com 3 fitas beta e outra folha com 4 fitas beta. Também foram preservadas duas hélices alfa na região N-terminal da proteína. No entanto, o final da parte C-terminal da estrutura, que corresponde aos últimos aminoácidos da seqüência e têm pouca similaridade com a seqüência molde (ver Fig. 4) está bastante diferente nas duas estruturas e precisa ser modelada com mais detalhes. Para este tipo de modelagem vai ser necessário pesquisar a existência de estruturas já resolvidas que contenham uma composição de aminoácidos similar à quella da XF2234: RAATTPRRIQVGNAVSNDVALQSM. As smHSPs têm em comum um domínio conservado na sua seqüência, chamado de domínio alfa-cristalino, que começa em geral após o aminoácido 40 e se prolonga até o aminoácido 140, ou ainda vai até o fim do C-terminal em algumas proteínas. É importante, portanto, fazer uma modelagem cuidadosa desta parte da proteína.

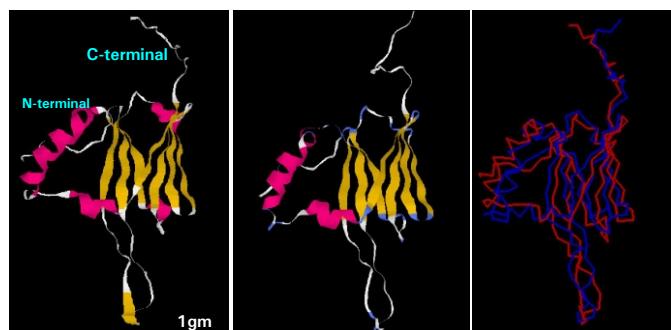


FIG. 4. Representações das estruturas tridimensionais da proteína molde - 1gme (esquerda) e da proteína alvo (centro), e superposição das duas estruturas (esquerda - 1gme em vermelho e XF2234 em azul).

Uma análise preliminar do modelo obtido, com o módulo Java Protein Dossier (^jPD, Neshich et al., 2004) do software Sting (Neshich et al., 2003) (Fig. 5), mostra que vários aminoácidos têm uma alta energia de ligação (indicados com seta) e podem ter um papel importante na funcionalidade de associação com outras proteínas. Uma análise completa da estrutura com o software Sting precisa ser feita, identificando os aminoácidos essenciais e comparando-os com outras estruturas da mesma família. Alguns desses aminoácidos (D-23 e E-27) encontram-se na primeira hélice do N-terminal, estão em áreas acessíveis da proteína e não fazem parte do domínio alfa-cristalino característico desta família de proteínas. Os aminoácidos D-65, D-68, R-84, E-87, R-100, H-106, R-108, D-116 e R-136 estão nas regiões próximas às fitas beta, na parte central da proteína e variam na questão da acessibilidade.

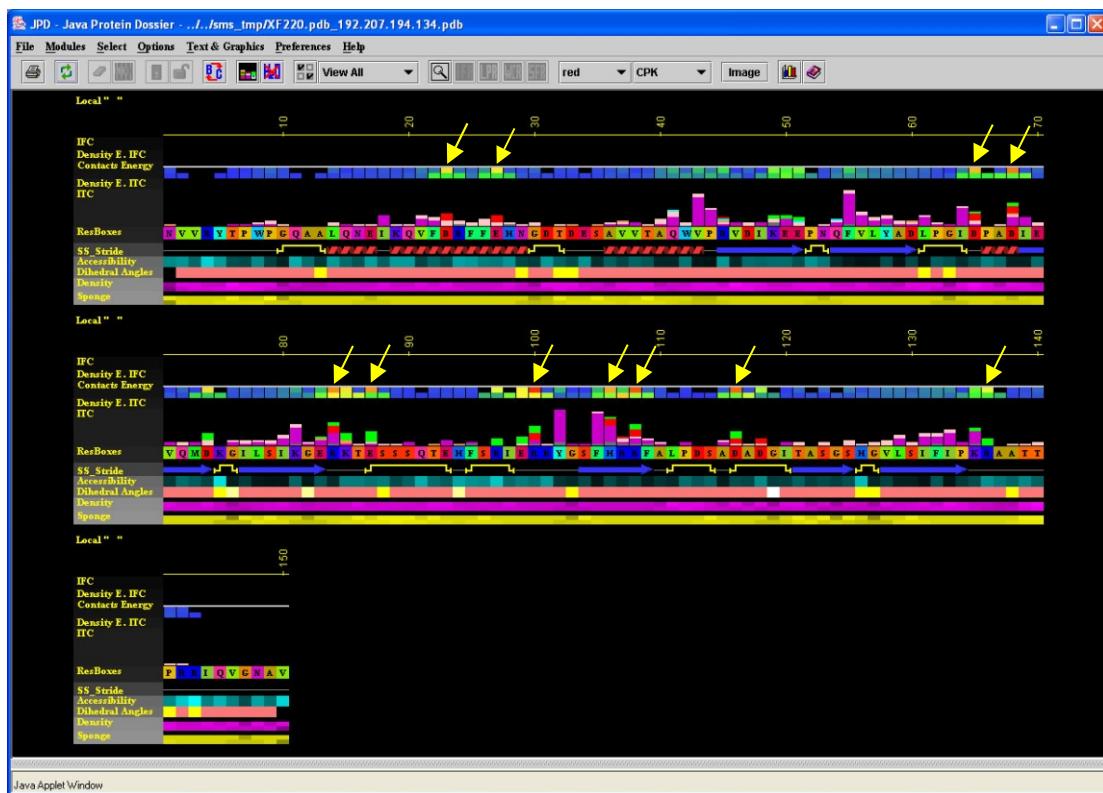


FIG. 5. JPD permite mostrar vários parâmetros para cada um dos aminoácidos da estrutura modelada. Os aminoácidos marcados com uma seta na linha "Contact Energy" podem ter importância para o funcionamento da proteína e devem ser analisados.

Conclusões

Este trabalho mostrou um modelo preliminar da estrutura tridimensional da proteína *small heat shock* de *Xylella fastidiosa*. Este modelo foi construído utilizando modelagem por homologia com a proteína smHSP 1gme. Como esperado para proteínas desta família, o modelo apresenta uma estrutura composta principalmente de fitas beta, com a presença de algumas hélices alfa. Como trabalho futuro deverá ser feita a modelagem das alças da proteína, uma dinâmica molecular e a validação do modelo. As diferenças funcionais entre membros de uma mesma família protéica são, em geral, consequência de diferenças estruturais na superfície externa das proteínas. Estas diferenças provêm de substituições, eliminações e inserções de resíduos nas cadeias de proteínas homólogas, principalmente nas alças, as regiões mais expostas da proteína. Por serem regiões estruturalmente variáveis, as alças geralmente determinam a especificidade das proteínas. Uma vez obtido o modelo completo será feita uma nova avaliação do modelo com o pacote *Sting*, inclusive com a comparação entre as estruturas para identificar quais aminoácidos são essenciais.

Referências Bibliográficas

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. *J. Mol. Biol.*, v. 215, p. 403-410, 1990.

AZZONI, A. R.; TADA, S. F.; ROSELLI, L. K.; PAULA, D. P.; CATANI, C. F.; SABINO, A. A.; BARBOSA, J. A.; GUIMARÃES, B. G.; EBERLIN, M. N.; MEDRANO, F. J.; SOUZA, A. P. Expression and purification of a small heat shock protein from the plant pathogen *Xylella fastidiosa*. *Protein Expr Purif.*, v. 33, n. 2, p. 297-303, 2004.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The Protein Data Bank. *Nucleic Acids Research*, v. 28, p. 235-242, 2000.

CRACK, J. A.; MANSOUR, M.; SUN, Y.; MACRAE, T. H. Functional analysis of a small heat shock/a-crystallin protein from *Artemia franciscana*. *Eur. J. Biochem.*, v. 269, p. 933-942, 2002.

DE JONG, W. W.; CASPERS, G. J. Genealogy of α -crystallinC small heat shock protein superfamily. *Int. J. Biol. Macromol.*, v. 22, p. 151-162, 1998.

FISER, A.; DO, R. K.; SALI, A. Modeling of loops in protein structures. *Protein Science*, v. 9, p. 1753-1773, 2000.

FROME, M.; CAMARGO, A.; HEBER, S.; CZINK, C.; SIMPSON, A. J. D.; HOHEISEL, J. D.; SOUZA, A. P. Mapping analysis of the *Xylella fastidiosa* genome. *Nucleic Acids Research*, v. 28, p. 3100-3104, 2000.

FUNDECITRUS. Clorose variegada dos citros – CVC. Disponível em:
<http://www.fundecitrus.com.br/doencas/cvc.html>.
 Acesso em: 20 dez. 2004.

LARANJEIRA, F. F. Ten years of citrus variegated chlorosis: what do we know? *Laranja*, v. 18, p. 123-141, 1997.

LEE, R. F.; DERRICK, K. S.; BERETTA, M. J. G.; CHAGAS, C. M.; ROSETTI, V. Citrus variegated chlorosis: a new destructive disease of citros in Brazil. *Citros Ind.*, p. 12-15, Oct. 1991.

MARTI-RENOM, M. A.; STUART, A.; FISER, A.; SÁNCHEZ, R.; MELO, F. E.; SALI, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, v. 29, p. 291-325, 2000.

NARBERHAUS, F. A-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microb. Mol. Biol. Rev.*, v. 66, p. 64-96, 2002.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **NCBI BLAST**. Disponível em: <<http://www.ncbi.nlm.nih.gov/BLAST/>>. Acesso em: 20 dez. 2004.

NESHICH G.; ROCCHIA, W.; MANCINI, A. L.; YAMAGISHI, M. E. B.; KUSER, P. R.; FILETO, R.; BAUDET, C.; PINTO, I. P.; MONTAGNER, A. J.; PALANDRANI, J. F.; KRAUCHENCO, J. N.; TORRES, R. C.; SOUZA, S.; TOGAWA, R. C.; HIGA, R. H. ^{Java} Protein Dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure. *Nucleic Acids Research*, v. 32 (Web Server issue), W595-W601, 2004. (DOI: 10.1093/nar/gkh480). Disponível em: <http://nar.oupjournals.org/cgi/content/full/32/suppl_2/W595>. Acesso em: 20 dez. 2004.

NESHICH, G.; TOGAWA, R. C.; MANCINI, A. L.; KUSER, P. R.; YAMAGISHI, M. E. B.; PAPPAS JUNIOR, G.; TORRES, W. V.; CAMPOS, T. F. e; FERREIRA, L. L.; LUNA, F. M.; OLIVEIRA, A. G.; MIURA, R. T.; INOUE, M. K.; HORITA, L. G.; SOUZA, D. F. de; DOMINIQINI, F.; ÁLVARO, A.; LIMA, C. S.; OGAWA, F. O.; GOMES, G. B.; PALANDRANI, J. F.; SANTOS, G. F. dos; FREITAS, E. M. de; MATTIUZ, A. R.; COSTA, I. C.; ALMEIDA, C. L. de; SOUZA, S.; BAUDET, C.; HIGA, R. H. STING Millennium: a Web based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. *Nucleic Acids Research*, v. 31, n. 13, p. 3386-3392, 2003.

SALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, v. 234, p. 779-815, 1993.

SIMPSON, A. J. G.; REINACH, F. C.; ARRUDA, P.; ABREU, F. A.; ACENCIO, M.; ALVARENGA, R.; ALVES, L. M. C.; ARAYA, J. E.; BAIA, G. S.; BAPTISTA, C. S.; BARROS, M. H.; BONACCORSI, E. D.; BORDIN, S.; BOVÉ, J. M.; BRIONES, M. R. S.; BUENO, M. R. P.; CAMARGO, A. A.; CAMARGO, L. E. A.; CARRARO, D. M.; CARRER, H.; COLAUTO, N. B.; COLOMBO, C.; COSTA, F. F.; COSTA, M. C. R.; COSTA-NETO, C. M.; COUTINHO, L. L.; CRISTOFANI, M.; DIAS-NETO, E.; DOCENA, C.; EL-DORRY, H.; FACINCANI, A. P.; FERREIRA, A. J. S.; FERREIRA, V. C. A.; FERRO, J. A.; FRAGA, J. S.; FRANÇA, S. C.; FRANCO, M. C.; FROHME, M.; FURLAN, L. R.; GARNIER, M.; GOLDMAN, G. H.; GOLDMAN, M. H. S.; GOMES, S. L.; GRUBER, A.; HO, P. L.; HOHEISEL, J. D.; JUNQUEIRA, M. L.; KEMPER, E. L.; KITAJIMA, J. P.; KRIEGER, J. E.; KURAMAE, E. E.; LAIGRET, F.; LAMBAIS, M. R.; LEITE, L. C. C.; LEMOS, E. G. M.; LEMOS, M. V. F.; LOPES, S. A.; LOPES, C. R.; MACHADO, J. A.; MACHADO, M. A.; MADEIRA, A. M. B. N.; MADEIRA, H. M. F.; MARINO, C. L.; MARQUES, M. V.; MARTINS, E. A. L.; MARTINS, E. M. F.; MATSUKUMA, A. Y.; MENCK, C. F. M.; MIRACCA, E. C.; MIYAKI, C. Y.; MONTEIRO-VITORELLO, C. B.; MOON, D. H.; NAGAI, M. A.; NASCIMENTO, A. L. T. O.; NETTO, L. E. S.; NHANI JUNIOR, A.; NOBREGA, F. G.; NUNES, L. R.; OLIVEIRA, M. A.; OLIVEIRA, M. C. de; OLIVEIRA, R. C. de; PALMIERI, D. A.; PARIS, A.; PEIXOTO, B. R.; PEREIRA, G. A. G.; PEREIRA JUNIOR, H. A.; PESQUERO, J. B.; QUAGGIO, R. B.; ROBERTO, P. G.; RODRIGUES, V.; ROSA, A. J. de M.; ROSA JUNIOR, V. E. de; SÁ, R. G. de; SANTELLI, R. V.; SAWASAKI, H. E.; SILVA, A. C. R. da; SILVA, A. M. da; SILVA, F. R. da; SILVA, W. A.; SILVEIRA, J. F. da; SILVESTRI, M. L. Z.; SIQUEIRA, W. J.; SOUZA, A. A. de; SOUZA, A. P. de; TERENZI, M. F.; TRUFFI, D.; TSAI, S. M.; TSUHAKO, M. H.; VALLADA, H.; VAN SLUYS, M. A.; VERJOVSKI-ALMEIDA, S.; VETTORE, A. L.; ZAGO, M. A.; ZATZ, M.; MEIDANIS, J.; SETUBAL, J. C. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, v. 406, p. 151-157, 13 July, 2000.

Comunicado Técnico, 64

Ministério da Agricultura,
Pecuária e Abastecimento



Embrapa Informática Agropecuária
Área de Comunicação e Negócios (ACN)
Endereço: Caixa Postal 6041 - Barão Geraldo
13083-970 - Campinas, SP
Fone: (19) 3789-5743
Fax: (19) 3289-9594
e-mail: sac@cnptia.embrapa.com.br

1^a edição on-line - 2004
Todos os direitos reservados.

Comitê de Publicações

Presidente: Marcos Lordello Chaim (presidente em exercício)
Membros Efetivos: Carla Geovana Macário, Ivanilde Dispato, José Ruy porto de Carvalho, Luciana Alvim Santos Romani, Marcia Isabel Fugisawa Souza, Suzilei Almeida Carneiro (secretária)
Suplentes: Carlos Alberto Alves Meira, Eduardo Delgado Assad, Maria Angelica Andrade Leite, Maria Fernanda Moura, Maria Goretti Gurgel Praxedes

Expediente

Supervisor editorial: Ivanilde Dispato
Normalização bibliográfica: Maria Goretti Gurgel Praxedes
Editoração eletrônica: Área de Comunicação e Negócios