



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial

(21) **PI0706193-5 A2**



* B R P I 0 7 0 6 1 9 3 A 2 *

(22) Data de Depósito: 11/12/2007

(43) Data da Publicação: 25/08/2009

(RPI 2016)

(51) *Int.Cl.:*

C12N 15/29 (2009.01)

C12N 15/67 (2009.01)

C12N 15/62 (2009.01)

C12N 15/52 (2009.01)

C12N 15/63 (2009.01)

C12N 1/00 (2009.01)

(54) Título: **COMPOSIÇÕES E MÉTODOS PARA DIRECIONAR A EXPRESSÃO DE GENES USANDO O PROMOTOR DO GENE DA FAMÍLIA DAS PEROXIDASES DE PLANTAS DE CAFÉ**

(73) Titular(es): Embrapa - Empresa Brasileira de Pesquisa Agropecuária, Instituto Agrônômico, Universidade Estadual Paulista Julio de Mesquita Filho

(72) Inventor(es): Ivan de Godoy Maia, Marcos Brandalise, Mirian Perez Maluf, Wallace Gonçalves

(57) Resumo: COMPOSIÇÕES E MÉTODOS PARA DIRECIONAR A EXPRESSÃO DE GENES USANDO O PROMOTOR DO GENE DA FAMÍLIA DAS PEROXIDASES DE PLANTAS DE CAFÉ. A presente invenção compreende a descrição de uma nova seqÜência regulatória provida para direcionar a expressão de uma seqÜência de nucleotídeo, tal como genes estruturais, em raízes de plantas, incluindo monocotiledÔneas e dicotiledÔneas. Mais especificamente a invenção refere-se a seqÜências regulatórias de polinucleotídeos isoladas de plantas de café que são capazes de iniciar e acionar a transcrição de polinucleotídeos, e ao uso destas seqÜências regulatórias no direcionamento de transcrição de polinucleotídeos endógenos e/ou heterólogos para tecidos radiculares e produção de polipeptídeos. A invenção descreve ainda construções de DNA que contém o promotor de um gene que codifica uma proteína da família das peroxidases de plantas de café que está operacionalmente ligado a um gene heterólogo e/ou endógeno. Além disso, a invenção diz respeito ao uso destas construções na forma de vetores de expressão, vetores recombinantes e em plantas, células vegetais ou protoplastos transgênicos. A invenção ainda descreve um método utilizando tais construções contendo o promotor de um gene que codifica uma proteína da família das peroxidases de plantas de café para produção de plantas, células vegetais ou protoplastos transgênicos.



PI0706193-5

COMPOSIÇÕES E MÉTODOS PARA DIRECIONAR A EXPRESSÃO DE GENES USANDO O PROMOTOR DO GENE DA FAMÍLIA DAS PEROXIDASES DE PLANTAS DE CAFÉ

CAMPO DA INVENÇÃO

5 A presente invenção está relacionada a um novo promotor para expressão específica de genes em raízes de plantas. Mais especificamente a invenção refere-se a seqüências regulatórias de polinucleotídeos isoladas de plantas de café que são capazes de iniciar e acionar a transcrição de polinucleotídeos, e ao uso destas seqüências regulatórias na modificação de transcrição de polinucleotídeos endógenos e/ou
10 heterólogos e produção de polipeptídeos no tecido radicular. A invenção descreve ainda construções de DNA que contém o promotor de um gene da família das peroxidases de plantas de café que está operacionalmente ligado a um gene heterólogo e/ou endógeno. Além disso, a invenção diz respeito ao uso destas construções na forma de vetores de expressão, vetores recombinantes e em plantas, células vegetais ou protoplastos
15 transgênicos. A invenção ainda descreve um método utilizando tais construções contendo o promotor do gene da família das peroxidases de plantas de café para produção de plantas, células vegetais ou protoplastos transgênicos.

FUNDAMENTOS DA INVENÇÃO

20 A cafeicultura desempenha atualmente um importante papel na economia nacional gerando divisas na ordem de 2 bilhões de dólares anuais com 26 milhões de sacas exportadas ao ano (Companhia Nacional do Abastecimento – Conab). Com um mercado ainda em expansão, o agronegócio do café gera, em todo mundo, recursos na ordem de 91 bilhões de dólares ao comercializar em média 115 milhões de sacas. O Brasil ocupa uma posição dominante no mercado internacional, se destacando como
25 maior produtor (contribuição de 30% da produção mundial nas últimas safras) e exportador mundial de café, com participação média de 25%. O agronegócio do café representa mais de 2,5 % do valor total das exportações brasileiras além de gerar oito

milhões de empregos diretos e indiretos, destacando-se como o setor que mais emprega no Brasil. Apesar do ótimo desempenho, a expansão da produção de café no Brasil é limitada, dentre outros fatores, por pragas e doenças que atacam a cultura. Dentre as principais pragas do café, destaca-se, afetando as raízes, os nematóides do gênero

5 *Meloidogyne* sp.

A crescente demanda mundial por alimentos de melhor qualidade, principalmente livres de defensivos agrícolas, tem promovido um avanço significativo em diversas áreas da produção vegetal. Dentre estas se destaca o melhoramento genético vegetal.

10 O melhoramento genético de plantas visando melhorias agronômicas vem sendo realizado pelo homem há milhares de anos utilizando-se de procedimentos clássicos, funcionais, de cruzamentos controlados entre espécies. Entretanto com o advento da biologia molecular nos últimos anos, somado ao desenvolvimento de técnicas biotecnológicas avançadas, o melhoramento genético de plantas via engenharia genética

15 tornou-se uma ferramenta importante na obtenção de plantas genotipicamente superiores em um curto espaço de tempo, sendo, portanto de grande interesse para o setor agrícola.

Métodos avançados de melhoramento genético vegetal têm se destacado principalmente devido à possibilidade de obtenção de resultados uniformes e direcionados para características desejáveis. Dentre os inúmeros métodos

20 biotecnológicos mundialmente utilizados para o melhoramento, destaca-se como um dos mais importantes, a produção de plantas geneticamente modificadas. A transgenia tem permitido o melhoramento de genótipos previamente selecionados por métodos convencionais através da introdução de um ou poucos genes que, em muitos casos, são encontrados em espécies distantes e que não poderiam ser transferidos via

25 recombinação, o que acontece obrigatoriamente entre indivíduos sexualmente compatíveis. Por outro lado, se fossem assim transferidos levar-se-ia anos para obtenção do genótipo desejado. Em se tratando de melhoramento de espécies perenes, mais precisamente da espécie *Coffea arabica*, pode-se esperar um mínimo de 25 anos entre as primeiras hibridações até a obtenção de novos cultivares.

30 O melhoramento de espécies perenes através da obtenção de organismos geneticamente modificados tem sido apontado como uma alternativa ao melhoramento

clássico. Nesse contexto, a introdução via transgenia de características desejáveis como a resistência a fatores bióticos e abióticos tende a ser facilitada, uma vez que um grande número de genes induzidos por tais fatores em diferentes espécies tem sido identificado e caracterizado. Muitos destes genes têm sido identificados em projetos onde genomas
5 inteiros foram seqüenciados, como nos casos de *Arabidopsis thaliana* e do arroz, ou em projetos nos quais foram seqüenciados milhares de cDNAs oriundos de bibliotecas construídas a partir de diferentes tecidos e em diferentes condições de estresse, também conhecidos como projetos ESTs (expressed sequence tags). Como exemplo do último caso, na área vegetal, destaca-se o projeto Genoma Café.

10 Os projetos de seqüenciamento acima citados são de suma importância, entretanto sabe-se que a simples identificação de um gene não é garantia única para a sua utilização na obtenção de transgênicos. A produção eficiente de uma proteína heteróloga em uma planta depende, por exemplo, da obtenção de altos níveis de transcrição do gene introduzido e para tal, promotores altamente ativos são necessários.

15 A expressão de um gene é regulada, em parte, pelos processos celulares envolvidos na transcrição. Durante a transcrição, um RNA de filamento único, complementar à seqüência de DNA a ser transcrita é formado pela ação de polimerases de RNA e fatores de transcrição. A iniciação da transcrição em células eucarióticas é regulada por interações complexas entre motivos de DNA *cis*-atuantes, localizados
20 dentro do gene a ser transcrito, e fatores protéicos *trans*-atuantes. Dentre as regiões regulatórias *cis*-atuantes estão as seqüências de DNA, chamadas promotores, às quais a RNA polimerase é primeiramente ligada, direta ou indiretamente. Para efeitos da presente invenção, o termo “promotor” se refere a seqüências específicas de DNA, usualmente localizadas à montante (5’) da região codificadora de um gene, que controla
25 a expressão desta região codificadora devido a sua capacidade de prover um sítio de reconhecimento para a RNA polimerase e/ou outros fatores necessários para o início da transcrição, e para que esta se inicie no local correto do gene.

Tipicamente, um promotor possui um “TATA box” e uma região de ativação à montante. Este “TATA box” é o responsável pela localização do sítio de início da
30 transcrição, aproximadamente 25 pares de bases na direção do início da seqüência codificadora do gene (3’).

Existem dois tipos básicos de promotores, os induzíveis e os não induzíveis, também chamados constitutivos. Os promotores constitutivos são capazes de dirigir a expressão de seqüências de DNA durante todo o desenvolvimento de uma planta, além de não possuir uma especificidade quanto ao local de expressão desta seqüência, sendo esta, então, expressa em uma grande variedade de células e tecidos. Apesar disso, o termo “constitutivo” não implica que a seqüência é expressa nos mesmos níveis em todas as células vegetais. O promotor induzível, por sua vez, é um promotor capaz de ativar (direta ou indiretamente) a transcrição de uma ou mais seqüências de DNA, ou genes, em resposta a um determinado indutor. Na falta deste indutor, as seqüências de DNA ou genes não serão transcritos. O indutor pode ser um componente químico (descrito, por exemplo, no documento de patente WO9519443), um estresse de origem fisiológica (como no caso de ferimentos, que é descrito, por exemplo, no documento de patente US6677505), ou um composto endógeno gerado em resposta a mudanças no desenvolvimento da planta.

Os promotores constituem, portanto, uma ferramenta chave em processos biotecnológicos a fim de garantir que a expressão de um gene de interesse seja efetiva e regulada espacial e temporalmente. Dentre os promotores usualmente empregados na produção de plantas geneticamente modificadas destacam-se o promotor 35S do Vírus do Mosaico da Couve Flor (CaMV 35S), os promotores dos genes da nopalina e octopina de *Agrobacterium tumefaciens* e o promotor do gene que codifica a ubiquitina de milho. Apesar dos grandes avanços obtidos com o emprego desses promotores, os padrões de expressão dos transgenes submetidos à regulação dos mesmos são variados e baixos em alguns casos, não havendo garantia de expressão no órgão/tecido adequado. O uso de promotores ubíquos como os citados, determina a expressão do produto gênico em todos os órgãos da planta, o que nem sempre é desejável. Tal característica, desnecessária e indesejável em plantas geneticamente modificadas destinadas a alimentação humana, tende a diminuir a aceitação do produto final, ou a de seus derivados, pelos consumidores resultando em resistência aos produtos geneticamente modificados. Nesse caso, promotores tecido-específicos, que garantam uma regulação espacial do transgene, seriam de grande importância e vários estudos estão sendo empreendidos com esse propósito.

Existem inúmeros promotores tecido-específicos descritos para plantas, como é o caso daqueles que apresentam expressão específica em semente (WO8903887), tubérculo (como mencionado no pedido de patente US20030175783, Keil et al., 1989 EMBO J. 8: 1323:1330), folhas (como mencionado no pedido de patente
 5 US20030175783, Hudspeth et al., 1989, Plant Mol Biol 12:579-589), fruta (Edwards and Coruzzi, 1990, Annu.Rev.Genet. 24: 275-303 e US5753475), caule (como mencionado no pedido de patente US20030175783, Keller et al., 1988, EMBO J. 7: 3625-3633), tecidos vasculares (como mencionado no pedido de patente US20030175783, Peleman et al., 1989, Gene 84: 359-369 e Schmölling et al., 1989,
 10 Plant Cell 1: 665-670), raiz (US20060143735 e como mencionado no pedido de patente US20030175783, Keller et al., 1989 Genes Devel. 3:1639-1646), estames (WO8910396, WO9213956) e promotores específicos da zona de deiscência (WO9713865) e meristema (Ito et al., 1994, Plant Molecular Biology, 24: 863-878).

Dentre os estudos conduzidos com promotores específicos de tecido radicular podemos citar o realizado por Mazarei et al. (Mazarei, M., Lennon, K.A., Puthoff, D.P., Rodermel, S.R. and Baum, T.J. 2004. Homologous soybean and Arabidopsis genes share responsiveness to cyst nematode infection. Mol. Plant Pathology, 5: 409 -423) que
 15 isolaram em soja, um promotor raiz-específico induzido pelo nematóide cístico *Heterodera glycines*. Promotores com tais características permitem o correto direcionamento da expressão dos genes de interesse, evitando principalmente os efeitos indesejáveis derivados da expressão constitutiva na planta como mencionado.
 20

Outros exemplos de promotores que controlam a expressão de genes especificamente em raízes e que são importantes para resistência a estresses bióticos (como nematóides) e abióticos podem ser visualizados nos documentos de patentes:
 25 US6384207 (promotor de milho específico para raiz), US 6649748 (promotor da peroxidase de batata induzível por estresse), US6703541 (promotor de milho induzível por nematóides), US6395963 (promotores de *Arabidopsis thaliana* induzidos por nematóides), US6262344 (promotor de *Arabidopsis thaliana* induzível por nematóide), US5750386 (promotor de *Nicotiana tabacum* induzido por patógeno de planta),
 30 US2005010974 (promotores de milho específicos de raiz), US20010047525 (promotores de milho específicos de raiz), US7196247 (promotor isolado da isoflavona

sintetase de soja para dirigir a expressão para raiz), US2006230470 (promotor de arroz para induzir expressão nos ápices de raiz e caule com o interesse de expressar substâncias contra nematóides), US6927320 (promotores de *Arabidopsis thaliana* específicos de raiz), US2006260013 (promotores de *M. truncatula* específicos de raiz),
 5 US20060143735 (promotores de *Arabidopsis thaliana* ativos em raiz), US7193135 (promotor do gene que codifica a oxigenase 2-oxoglutarate dependente de milho, induzível e específico de raiz), US2005097633 (um promotor do gene *CRWAQ81* de milho específico de raiz), US20060107347 (promotor do gene *Cr1Bio* de milho específico de raiz), US2004248304 (promotor de soja Pyk20 regulado por nematóides),
 10 US2005070697 (promotor de soja regulado por nematóides), US2003167507 (promotor da isoflavona sintetase I de soja específico de raiz), US2004019934 (promotor de coníferas específico de raiz), US7193136 (promotor de soja induzível e expresso em raiz), US7119254 (promotor induzido por nematóides isolado de *Arabidopsis thaliana*), US6906241 (promotor induzido por nematóides isolado de *Nicotiana* - Ntce17),
 15 US677428 (promotor do gene da metalotienina de milho específico para raiz), US6518483 (promotor do gene *Knox1* de milho específico de raiz), US5837876 (promotor de tabaco que dirige a expressão de um DNA heterólogo para o córtex da raiz), US5633363 (promotor de milho específico para raiz), US5459252 (seqüência de promotor de *Nicotiana tabacum* denominada RB7 capaz de dirigir a transcrição específica para raiz),
 20 US5401836 (promotor de *Brassica* específico para raiz), US4886753 (promotor de soja específico para nódulos de raiz), US7041873 (promotores de soja regulados por nematóides), US6271437 (promotores isolados de genes de soja induzíveis por nematóides), US20070050862 (promotor isolado de *Arabidopsis thaliana* (AtCe15) que é expresso exclusivamente na raiz) e
 25 US20020032917 (promotor do gene *SCARECROW* de *Arabidopsis thaliana* específico para raiz). Outros promotores específicos e induzíveis por estresses bióticos podem ser encontrados nos trabalhos de Park et al. (2000) (Park, K.S., Kloepper, J. Activation of PR-1a promoter by rhizobacteria that induce systemic resistance in tobacco against *Pseudomonas syringae* pv. Tabaci. Biological Control, 18: 2-9, 2000); Song et al. (2002)
 30 (Song, F., Goodman, R.M. Cloning and identification of the promoter of the tobacco Sar8.2b gene, a gene involved in systemic acquired resistance. Gene, 290: 115-124,

2002); Wang et al. (2002) (Wang, S-J., Lan, Y-C., Chen, S.F., Chen, Y.M., Yeh, K.-W. Wound-response regulation of the sweet potato sporamin gene promoter region. *Plant Molecular Biology*, 48: 223-231, 2002); e Mazarei et al. (2004) (Mazarei, M., Lennon, K.A., Puthoff, D.P., Rodermeil, S.R. and Baum, T.J. Homologous soybean and Arabidopsis genes share responsiveness to cyst nematode infection. *Mol. Plant Pathology*, 5: 409-423, 2004). Trabalhos com promotores sinteticamente construídos para este fim também podem ser encontrados em Rushton et al. (2002) (Rushton, P.J., Reinstadler, A., Volker, L., Lippok, B. and Somssich, I.E. Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen and wound induced signaling. *The Plant Cell*, 14: 749-762, 2002). Em café, entretanto, a disponibilidade de promotores com padrões de expressão conhecidos e restritos a determinados órgãos/tecidos é praticamente inexistente e ainda não há cultivares de *Coffea arabica* resistentes a nematóides.

Apesar de existir um grande número de promotores específicos para os tecidos radiculares, nenhum deles é específico para as plantas de café e para nematóides das espécies *Meloidogyne*. Além do mais, o promotor da presente invenção não apresenta nenhuma similaridade com os promotores já descritos no estado da técnica (vide exemplo 1). Esses fatos demonstram a necessidade de identificação de novos promotores para uso em plantas dicotiledôneas, especialmente café, e a importância do promotor da presente invenção para a produção de plantas geneticamente modificadas.

Um motivo de preocupação relativo ao uso de plantas geneticamente modificadas refere-se à expressão de proteínas tóxicas como as utilizadas para o controle de determinados insetos praga. Muitos trabalhos relatam a eficiente utilização da proteína Bt (de *Bacillus thuringiensis*) em plantas transgênicas visando o controle de lagartas (Shu, Q.Y., Ye, G.Y., Cui, H.R., Cheng, X., Xiang, Y., Wu, D., et al.. Transgenic rice plants with a synthetic cry1Ab gene from *Bacillus thuringiensis* were highly resistant to eight lepidopteran rice pest species. *Molecular Breed*, 6: 433-439, 2000; Leroy, T., Henry, A.M., Royer, M., Altosaar, I., Frutos, R., Duris, D., Philippe, R.. Genetically modified coffee plants expressing the *Bacillus thuringiensis* cry1Ac gene for resistance to leaf miner. *Plant Molecular reports*, 19: 382-389, 2000; Nester, E.W., Thomashow, L.S., Matthew, M. and Gordon, M. 2002. 100 Years of *Bacillus*

thuringiensis: a critical scientific assessment. American Academy of Microbiology, 1752 N Street, NW Washington, DC, 16 pp). Entretanto, tal metodologia pode fazer com que os insetos alvo adquiriram resistência quando mantidos sob constante pressão de seleção (Meng, F., Shen, J., Zhou, W., and Cen, H. Long-term selection for resistance to transgenic cotton expressing *Bacillus thuringiensis* toxin in *Helicoverpa armigera* (Hubner) (Lepidoptera:Noctuidae). Pest Manage Sci., 60: 167-172, 2004), o que normalmente acontece quando promotores constitutivos são utilizados para expressão das proteínas heterólogas. Esses efeitos indesejados poderiam ser minimizados através do uso de promotores órgão/tecido-específicos que, se possível, fossem responsivos a determinados estímulos como, por exemplo, a um determinado estresse biótico. Neste caso, a expressão do transgene seria limitada ao período de duração do estímulo e principalmente direcionada para o local onde o estímulo foi produzido, ressaltando mais uma vez a importância do promotor da presente invenção.

A obtenção de plantas geneticamente modificadas com genes de interesse agrônomo (resistência a pragas, doenças, seca, geada, etc...) isolados de espécies selvagens de *Coffea* representa um avanço significativo, pois permite acelerar drasticamente o processo de obtenção de novos cultivares. Nesse contexto, o direcionamento da expressão dos genes de interesse para determinados órgãos específicos da planta, utilizando-se de promotores tecido-específicos provenientes da própria espécie, constitui uma importante ferramenta biotecnológica. Tal inovação viabiliza não só a construção de cassetes de expressão compostos de subunidades (promotor e gene) pertencentes ao gênero *Coffea*, mas também a obtenção de um padrão de expressão tecido-específico altamente desejável tanto do ponto de vista de biossegurança como de opinião pública. A busca por promotores com tais características tem sido uma prioridade em programas biotecnológicos que objetivam a produção e liberação comercial de organismos geneticamente modificados.

SUMÁRIO DA INVENÇÃO

A presente invenção compreende a descrição de uma nova sequência regulatória (promotora) capaz de direcionar a expressão de uma sequência de nucleotídeo em tecidos radiculares, tal como genes estruturais, em plantas, incluindo monocotiledôneas e dicotiledôneas. De acordo com a presente invenção é descrito um promotor tecido-específico isolado de plantas de café (*Coffea arabica*), denominado **CaPEROX**, junto com um método para uso dessa região regulatória de polinucleotídeo para modificação de expressão de polinucleotídeos endógenos e/ou heterólogos em plantas transgênicas.

Em um primeiro aspecto, a presente invenção provê uma sequência de polinucleotídeo isolada de plantas de café (*Coffea arabica*) que possua 40%, preferencialmente 60%, mais preferencialmente 75%, mais preferencialmente ainda 90% de identidade com a SEQ ID NO1; sequência reversa de SEQ ID NO1; sondas e iniciadores correspondendo a SEQ ID NO1.

Em outro aspecto, a presente invenção provê genes quiméricos compreendendo o polinucleotídeo da presente invenção, ou sozinho, ou em combinação com um ou mais polinucleotídeos conhecidos, junto com células e organismos compreendendo estes genes quiméricos.

Em um aspecto relacionado, a presente invenção provê vetores recombinantes compreendendo, na direção 5'-3', uma sequência de polinucleotídeo do promotor da presente invenção, um polinucleotídeo a ser transcrito, e uma sequência de terminação da transcrição. O polinucleotídeo a ser transcrito pode compreender uma região de leitura aberta de um polinucleotídeo que codifica um polipeptídeo de interesse, ou pode ser uma região de não codificação, ou não traduzida, de um polinucleotídeo de interesse. A matriz de leitura aberta pode ser orientada em uma direção "sense" ou "antisense". Preferivelmente, a sequência de terminação da transcrição é funcional em plantas. Mais preferivelmente, a sequência de terminação é a do gene de interesse, mas podendo ser outras descritas no estado da técnica (Griffiths, A.J.F., Gelbart, W.M., Miller, J.H., Lewontin, R.C. 2000. Regulation of gene transcription In: Modern Genetic Analysis. 3rd edition, W.H. Freeman and Company, New York, NY.) como o terminador da nopalina sintetase de *A. tumefaciens*. Os vetores recombinantes podem ainda incluir um marcador para a identificação de células transformadas.

Em uma outra concretização da invenção são providos microorganismos transformados compreendendo o vetor recombinante da presente invenção.

Em outro aspecto, as células de plantas transgênicas compreendendo o vetor recombinante da presente invenção são providas, junto com organismos, como plantas, compreendendo estas células transgênicas, e frutos, sementes e outros produtos, derivados, ou progênie destas plantas. Os propágulos das plantas transgênicas inventivas são incluídos na presente invenção.

Em outro aspecto da presente invenção é provido um método para modificar a expressão de genes em um organismo, como uma planta, incluindo a incorporação estável no genoma do organismo contendo o vetor recombinante da presente invenção.

Em outro aspecto da presente invenção é provido um método para produzir um organismo transformado, como uma planta, tendo a expressão de um polipeptídeo dirigida para tecidos radiculares. Esse método compreende transformar uma célula de planta com o vetor recombinante da presente invenção para prover uma célula transgênica de condições que conduzem à regeneração e crescimento da planta madura.

Ainda em outro aspecto da presente invenção é provido um método para identificação de um gene responsável por uma função ou fenótipo desejado. O método compreende: 1) a transformação de uma célula de planta contendo um vetor recombinante compreendendo uma sequência de promotor de polinucleotídeos da presente invenção ligada operacionalmente a um polinucleotídeo a ser testado, 2) cultivar a célula de planta sob condições que conduzem a regeneração e crescimento da planta madura de modo a prover uma planta transgênica, e 3) comparar o fenótipo da planta transgênica com o fenótipo de plantas não transformadas, ou de tipo selvagem.

Os aspectos acima mencionados e adicionais da presente invenção e o modo de obter os mesmos se tornarão evidentes, e a invenção será melhor compreendida por referência na “Descrição Detalhada da Invenção”.

BREVE DESCRIÇÃO DAS FIGURAS

Figura 1 – Representação esquemática do protocolo adotado para amplificação (por *genome walking*) e clonagem do promotor **CaPEROX** a partir de DNA genômico de café.

Figura 2 - Esquema de localização dos elementos regulatórios encontrados no promotor **CaPEROX** por análise comparativa com banco de dados do PLACE (Higo et al. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27: 297-300, 1999).

5 Figura 3 - Representação esquemática do vetor pRT103-Gus (~5 Kb) que apresenta o gene repórter *uidA* (GUS) inserido entre os sítios *Bam*HI e *Nco*I sob o controle do promotor constitutivo CaMV35S.

Figura 4 - Representação esquemática do gene quimérico desenvolvido para a análise funcional da região promotora amplificada via *genome walking* (o promotor não se encontra em escala). Promotor **CaPEROX** fusionado ao gene repórter *uidA* que codifica a enzima β -glucoronidase (GUS) de *Escherichia coli*.

Figura 5 - Representação esquemática do plasmídeo pCAMBIA-1381 (Cambia) utilizado na construção do gene quimérico visando a posterior transformação estável de tabaco empregando *Agrobacterium tumefaciens*.

15 DESCRIÇÃO DETALHADA DA INVENÇÃO

No contexto dessa descrição, inúmeros termos serão utilizados e por isso serão devidamente detalhados a seguir:

Um “gene quimérico” é um gene compreendendo um promotor e uma região codificadora de diferentes origens. No caso da presente invenção, o gene quimérico compreende o polinucleotídeo da presente invenção ligado a regiões codificadoras de genes endógenos e/ou exógenos.

Uma “seqüência consenso” é uma seqüência artificial na qual a base em cada posição representa a base freqüentemente mais encontrada na seqüência atual comparando-se diferentes alelos, genes ou organismos.

25 Um “promotor” é aquela porção do DNA localizada acima da região codificadora que contém sítios de ligação para os fatores de transcrição e RNA polimerases para iniciar a transcrição do DNA.

“Expressão” é a transcrição e/ou tradução de um gene estrutural, endógeno ou heterólogo.

30 “GC box” é um elemento comum no promotor que pode aumentar a atividade do mesmo.

“TATA box” é um elemento no promotor, localizado aproximadamente 30 bases acima do sítio de início da transcrição. O “TATA Box” está associado com fatores de transcrição em geral, incluindo a RNA polimerase.

5 O termo “gene” significa uma unidade física e funcional da hereditariedade, representada por um segmento de DNA que codifica uma proteína funcional ou molécula de RNA.

Um “gene endógeno” é um gene próprio da célula ou do organismo.

10 Um “gene heterólogo” é um gene isolado de um organismo doador e recombinação no organismo receptor transformado. É um gene que não é próprio da célula ou do organismo.

Um “gene repórter” é uma unidade codificadora cujo produto é facilmente testado, por exemplo, genes *CAT*, *GUS*, *GAL*, *LUC* e *GFP*. A expressão de um gene repórter pode ser usada para testar a função de um promotor ligado a esse gene repórter.

15 O termo “EST” refere-se a etiquetas de seqüências expressas (Expressed Sequence Tags), ou seja, são curtas seqüências das extremidades de uma seqüência de nucleotídeo transcrita (codificando uma proteína ou não). Essas seqüências, derivadas diretamente dos genes ativos, podem ser usadas na identificação em larga escala de genes e na determinação do padrão de expressão dos mesmos.

20 O termo “contig” conforme usado na invenção diz respeito a uma montagem virtual de coleção de ESTs correspondentes a um mesmo gene.

O termo “propágulo” como usado na presente invenção significa qualquer parte de uma planta que pode ser usada na reprodução ou propagação, sexual ou assexual, incluindo as mudas.

25 “Sense” significa que a seqüência de polinucleotídeo está na mesma orientação 5’-3’ com relação ao promotor.

“Antisense” significa que a seqüência de polinucleotídeo está na orientação contrária com relação à orientação 5’-3’ do promotor.

30 Como usado aqui, o termo “x-mero”, como referência a um valor específico de “x”, refere-se a uma seqüência compreendendo pelo menos um número específico (“x”) de resíduos do polinucleotídeo identificado como SEQ ID NO1. De acordo com formas de realização preferidas, o valor de x é preferivelmente pelo menos 20, mais

preferivelmente pelo menos 40, mais preferivelmente ainda pelo menos 60 e mais preferivelmente pelo menos 80. Assim, polinucleotídeos da presente invenção compreendem um polinucleotídeo de 20 meros, 40 meros, 60 meros, 80 meros, 100 meros, 120 meros, 150 meros, 180 meros, 220 meros, 250 meros, 300 meros, 400 meros, 500 meros ou 600 meros identificado como SEQ ID NO1 e variantes dos mesmos.

O termo “polinucleotídeo (s)” como usado aqui, significa um polímero de filamento único ou duplo de bases de deoxiribonucleotídeos ou ribonucleotídeos e inclui moléculas correspondentes de RNA e DNA, incluindo moléculas de HnRNA e mRNA, de filamentos tanto “sense” como “antisense”, e compreende cDNA, DNA genômico, e DNA recombinante, assim como polinucleotídeos completamente ou parcialmente sintetizados. Uma molécula de HnRNA contém íntrons e corresponde a uma molécula de DNA considerando uma complementaridade de modo um a um. Uma molécula de RNAm corresponde a uma molécula de DNA e HnRNA da qual os íntrons foram removidos. Um polinucleotídeo pode consistir de um gene completo, ou qualquer porção do mesmo. Os polinucleotídeos “antisense” operáveis podem compreender um fragmento do polinucleotídeo correspondente, e a definição de “polinucleotídeo” inclui assim todos esses fragmentos antisense operáveis. Os polinucleotídeos antisense, e técnicas envolvendo polinucleotídeos antisense, são bem conhecidos no estado da técnica (Sambrook, J et al. Molecular cloning. A laboratory manual, 2nd Ed., Cold Spring Harbor Laboratory Press, 1989)

Os polinucleotídeos descritos na presente invenção são preferencialmente cerca de 80% puros, mais preferencialmente pelo menos cerca de 90% puros, e mais preferencialmente pelo menos cerca de 99% puro.

O termo “oligonucleotídeo” refere-se a um segmento relativamente curto de uma sequência de polinucleotídeo, geralmente compreendendo entre 6 a 60 nucleotídeos. Esses oligonucleotídeos podem ser usados como sondas ou iniciadores (“primers”), onde as sondas podem ser usadas para uso em testes de hibridização e iniciadores para uso na amplificação de DNA por reação em cadeia da polimerase.

O termo “sonda” utilizado na presente invenção refere-se a um oligonucleotídeo, polinucleotídeo ou ácido nucléico, sendo RNA ou DNA, se ocorrendo naturalmente

como em uma digestão de enzima de restrição purificada ou produzida sinteticamente, que seja capaz de se parear com, ou especificamente hibridizando com, um ácido nucléico contendo seqüências complementares à sonda. Uma sonda pode ser ainda de cadeia simples ou cadeia dupla. O comprimento exato da sonda dependerá de muitos
5 fatores, incluindo temperatura, origem da sonda e uso do método. Por exemplo, dependendo da complexidade da seqüência alvo, a sonda de oligonucleotídeo tipicamente contém 15-25 ou mais nucleotídeos, embora ela possa conter menos nucleotídeos. As sondas aqui são selecionadas para serem complementares para diferenciar cadeias de uma seqüência de um ácido nucléico particular. Isto significa que
10 a sonda pode ser suficientemente complementar para ser capaz de “hibridizar especificamente” ou parear com suas respectivas cadeias-alvo sob uma série de condições pré-determinadas. Conseqüentemente, a seqüência da sonda não necessita refletir exatamente a seqüência complementar do alvo. Por exemplo, um fragmento de nucleotídeo não complementar pode ser ligado ao final 5’ ou 3’ da sonda, com o restante
15 da seqüência da sonda sendo complementar à cadeia alvo. Alternativamente, bases não complementares ou seqüências longas podem estar intercaladas dentro da sonda se esta apresentar complementaridade suficiente com a seqüência do ácido nucléico alvo para se associar especificamente a ele.

O termo “primer” como utilizado aqui se refere a um oligonucleotídeo, sendo
20 RNA ou DNA, cadeia simples ou cadeia dupla, derivado de um sistema biológico, gerado através de digestão com enzimas de restrição, ou produzido sinteticamente que, quando colocado em um ambiente próprio, é capaz de agir funcionalmente como um iniciador de uma síntese de ácido nucléico dependente de molde. Quando apresentado com um molde de ácido nucléico apropriado, nucleotídeos trifosfatos adequados
25 precursores de ácidos nucléicos, uma enzima DNA polimerase, cofatores adequados e condições tais como temperatura e pH adequados, o “primer” pode ser estendido em seu terminal 3’ pela adição de nucleotídeos pela ação de uma polimerase ou com atividade similar para produzir uma primeira extensão do produto. O ‘primer’ pode variar em comprimento dependendo de condições particulares e requerimentos para aplicação. Por
30 exemplo, em aplicações de diagnósticos, o ‘primer’ oligonucleotídeo possui tipicamente de 15-25 ou mais nucleotídeos em comprimento. O ‘primer’ deve ter

complementaridade suficiente com o molde desejado para iniciar a síntese da extensão do produto desejado. Isso não significa que a sequência do 'primer' deva representar um complemento exato do molde desejado. Por exemplo, uma sequência de nucleotídeo não complementar pode ser ligada ao final 5' de um 'primer' complementar.

5 Alternativamente, bases não complementares podem estar intercaladas dentro da sequência de oligonucleotídeo do 'primer', desde que o 'primer' tenha complementaridade suficiente com a sequência da cadeia molde desejada para prover funcionalmente um complexo molde-primer para a síntese da extensão do produto.¶

O termo "hibridizando especificamente" diz respeito à associação entre duas

10 moléculas de ácidos nucléicos de cadeia simples que possuam sequências suficientemente complementares para permitir tal hibridização sob condições pré-determinadas geralmente descritas no estado da técnica (Anderson, M.L.M. Nucleic Acid Hybridization. Springer, Bios Scientific Publishers, New York, 1999, 240 p; Innis, Michael A.; Gelfand, David H.; Sninsky, John J.; White, Thomas J. PCR protocols. A

15 guide to methods and applications. Academic Press. San Diego, CA, USA. 1990. 482 p.).

Em particular, o termo se refere à hibridização de um oligonucleotídeo com uma sequência substancialmente complementar contendo uma molécula de DNA ou RNA de cadeia simples da presente invenção. Condições apropriadas necessárias para realização

20 da hibridização específica entre moléculas de ácidos nucléicos de cadeia simples de complementaridade variada estão bem descritas no estado da técnica (Anderson, M.L.M. Nucleic Acid Hybridization. Springer, Bios Scientific Publishers, New York, 1999, 240 p; Innis, Michael A.; Gelfand, David H.; Sninsky, John J.; White, Thomas J. PCR protocols. A guide to methods and applications. Academic Press. San Diego, CA,

25 USA. 1990, 482 p). Uma fórmula comum para se calcular as condições de estringência requeridas para se ter uma hibridização entre moléculas de ácido nucléico segue abaixo (Sambrook et al. "Molecular Cloning, A Laboratory Manual" 2nd ed. (1989), Cold Spring Harbor Laboratory Press):

$$T_m = 81,5^{\circ}\text{C} + 16,6 \log [Na^{+}] + 0,41 (\%G+C) - 0,63 (\% \text{ formamida}) - 600/\text{pb no duplex (sonda)}$$

Como ilustração da fórmula acima, usando $[Na^+] = [0,368]$ e 50% de formamida, com conteúdo de GC de 42% e um tamanho de sonda médio de 200 bases, a T_m será de 57°C.

5 Sondas ou primers são descritos como correspondendo ao polinucleotídeo da presente invenção identificado como SEQ ID NO1 ou uma variante do mesmo, se a sonda de oligonucleotídeo ou primer, ou seu complemento, estiver contido dentro da sequência especificada como SEQ ID NO1, ou uma variante desta.

10 O termo “oligonucleotídeo” é referido aqui como ‘primers’ e ‘sondas’ da presente invenção, e é definido como uma molécula de ácido nucléico compreendendo de dois ou mais ribo ou deoxiribonucleotídeos, preferencialmente mais do que três. O tamanho exato dos oligonucleotídeos dependerá de vários fatores e na aplicação particular e uso dos mesmos. Oligonucleotídeos preferidos compreendem 15-50 pares de base consecutivas complementares à SEQ ID NO 1. As sondas podem ser facilmente selecionadas usando procedimentos bem descritos no estado da técnica (Sambrook et al.
15 “Molecular Cloning, a laboratory manual”, CSHL Press, Cold Spring Harbor, NY, 1989), levando em consideração estringências de hibridização DNA-DNA, recombinação e temperaturas de fusão, e potencial para formação de laços e outros fatores, que são conhecidos no estado da técnica.

20 A definição dos termos “complemento”, “complemento reverso” e “sequência reversa”, como usados aqui, é ilustrada pelo seguinte exemplo. Para a sequência 5’AGTGAAGT3’, o complemento é 3’TCACTTCA5’, o complemento reverso é 3’ACTTCACT5’ e a sequência reversa é 5’TGAAGTGA3’.

25 Como usado aqui, o termo “variante” ou “substancialmente similar” compreende sequências de aminoácidos ou nucleotídeos diferentes de sequências especificamente identificadas, em que um ou mais nucleotídeos ou resíduos de aminoácidos é deletado, substituído ou adicionado. As variantes podem ser variantes alélicas, de ocorrência natural, ou variantes de ocorrência não natural. As sequências variantes ou substancialmente similares dizem respeito a fragmentos de ácidos nucléicos que podem ser caracterizados pela porcentagem de similaridade de suas sequências de nucleotídeos
30 com a sequência de nucleotídeo descrita aqui (SEQ ID NO 1), como determinada por algoritmos comuns empregados no estado da técnica. Os fragmentos de ácidos nucléicos

preferidos são aqueles cujas seqüências de nucleotídeos têm pelo menos cerca de 40 ou 45% de identidade de seqüência, preferencialmente cerca de 50% ou 55% de identidade de seqüência, mais preferencialmente cerca de 60% ou 65% de identidade de seqüência, mais preferencialmente cerca de 70% ou 75% de identidade de seqüência, mais preferencialmente cerca de 80% ou 85% de identidade de seqüência, mais preferencialmente ainda com cerca de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de seqüência quando comparada com a seqüência de referência. A identidade percentual é determinada por alinhamento de duas seqüências a serem comparadas, determinando o número de resíduos idênticos na porção alinhada, dividindo este número pelo número total de resíduos na seqüência pesquisada e multiplicando o resultado por 100. Esse alinhamento pode ser feito através de softwares existentes na internet, sendo um deles o BLASTN, que está disponível na página do *National Center for Biotechnology Information/NCBI* (www.ncbi.nlm.nih.gov).

O termo “vetor” diz respeito a um replicon, tal como plasmídeo, cosmídeo, bacmídeo, fago ou vírus, no qual outras seqüências genéticas ou elementos (seja DNA ou RNA) são ligadas e replicadas quando associadas a ele. Preferencialmente o vetor derivado de vírus é selecionado entre os bacteriófagos, vaccínias, retrovírus ou vírus do papiloma bovino. O “vetor recombinante” é resultante da combinação de um vetor comercial com genes quiméricos, ou o polinucleotídeo da presente invenção operacionalmente ligado a um polinucleotídeo endógeno e/ou heterólogo de interesse que por sua vez está operacionalmente ligado a um sinal de terminação. Tais vetores podem ser obtidos comercialmente, incluindo Clontech Laboratories, Inc (Palo Alto, Calif.), Stratagene (La Jolla, Calif), Invitrogen (Carlsbad, Calif.), New England Biolabs (Beverly, Mass.) e Promega (Madison, Wis.). Alguns exemplos de vetores utilizados na presente invenção, mas não limitados a, são os vetores pGEM® -T Easy (Promega), pRT103 e pCAMBIA-1381 (Cambia).

O termo “clone” como usado na presente invenção corresponde a fragmentos de DNA inseridos no interior de uma bactéria e propagados de forma assexuada. Esses fragmentos de DNA estão normalmente ligados a um vetor de clonagem.

O termo “seqüências acentuadoras de expressão” corresponde aos amplificadores (“enhancers”) que potenciam a taxa de transcrição e que podem estar muito distantes do

promotor (antes ou depois, “upstream” ou “downstream”). Estes amplificadores não são específicos e potenciam a transcrição de qualquer promotor que estiver na sua vizinhança. A eficiência da expressão de um gene em um tecido específico depende da adequada combinação e integração dos amplificadores, dos promotores e das seqüências adjacentes.

O primeiro intensificador descoberto que estimulava a transcrição de genes eucariotos foi o SV40 (presente no genoma do Vírus 40 de Símio). Após a descoberta desse intensificador, outros diversos foram identificados como HSV-1, AMV, TMV, HPV-16, em outros genomas virais e no DNA de células eucarióticas.

O termo “operacionalmente ligado” significa que as seqüências regulatórias necessárias para expressão da seqüência codificadora são colocadas na molécula de DNA em posições apropriadas relativa à seqüência codificadora para efeito de expressão dessa. Essa mesma definição é às vezes aplicada para o arranjo de seqüências codificadoras e elementos controladores da transcrição (por exemplo, promotores, auxiliares ou “enhancers” e elementos ou seqüências de terminação) no vetor de expressão. Uma região codificadora exógena é tipicamente flanqueada por regiões regulatórias operacionalmente ligadas que regulam a expressão dessa região codificadora exógena em uma célula transformada (podendo ser microorganismo, vegetal ou animal). Uma região regulatória típica operacionalmente ligada a uma região codificadora exógena inclui um promotor, isto é, um fragmento de ácido nucléico que pode viabilizar a transcrição de regiões codificadoras exógenas, e portanto posicionado em sua porção 5'. No caso da presente invenção a região regulatória diz respeito às regiões substancialmente similares a SEQ ID NO 1. Para auxiliar no aumento da transcrição de um determinado polinucleotídeo, a seqüência promotora da presente invenção poderá estar ligada a outras seqüências regulatórias já descritas, tais como: ATATT (elemento de forte expressão na raiz), AACAAAC e GCCACCTCAT (elementos relativos a expressão específica em sementes), CACGTG e CCTACC (ambas seqüências podem ser estimuladas por um fator de estresse), entre outros [Ai-Min Wu et al. Isolation of a cotton reversibly glycosylated polypeptide (GhRGP1) promoter and its expression activity in transgenic tobacco. *Journal of Plant Physiology*, 163: 426-435, 2006].

Uma “seqüência de terminação” é uma seqüência de DNA que sinaliza o final da transcrição. São exemplos de seqüências de terminação, mas não estão limitados a, o sinal de terminação de SV40, sinal de adenilação de HSV TK, sinal de terminação do gene da nopalina sintetase (NOS) ou da octopina sintetase de *Agrobacterium tumefaciens*, sinal de terminação do gene 19S e 35S do CaMV, sinal de terminação do gene da álcool desidrogenase de milho, sinal de terminação do gene da manopina sintetase, sinal de terminação do gene da beta-faseolina, sinal de terminação do gene da ssRUBISCO, sinal de terminação do gene da sucrose sintetase, sinal de terminação do vírus que ataca o *Trifolium subterranean* (SCSV), sinal de terminação do gene *trpC* de *Aspergillus nidulans* e outros semelhantes. A presente invenção provê uma região regulatória de polinucleotídeos isolados que podem ser empregadas na manipulação de fenótipos de plantas, junto com polinucleotídeos isolados compreendendo estas regiões regulatórias. Mais especificamente a presente invenção está relacionada ao promotor ou seqüência regulatória que ocorre naturalmente em plantas de café (*Coffea arabica*), responsável pela expressão de um gene que codifica uma proteína da família das peroxidases nessa espécie vegetal. O referido promotor foi denominado na presente invenção de **CaPEROX** (SEQ ID NO1).

O polipeptídeo de interesse específico pode ser expresso especificamente em tecidos radiculares de planta através da incorporação de genes, ou seqüências de codificação, codificando o polipeptídeo, ligado operacionalmente a seqüência de promotor da presente invenção (SEQ ID NO 1), no genoma de um organismo, como uma planta. Uma diminuição na quantidade do polipeptídeo pode ser obtida por transformação de planta com cópias antisense destes genes.

O polinucleotídeo da presente invenção foi isolado de plantas de café, mais especificamente de *Coffea arabica*, mas ele pode ser alternativamente sintetizado usando técnicas de síntese convencionais. Especificamente o polinucleotídeo isolado da presente invenção inclui a seqüência identificada como SEQ ID NO1; o complemento reverso da seqüência identificada como SEQ ID NO1; complemento reverso da seqüência identificada como SEQ ID NO1.

Estudos da atividade do promotor da presente invenção são detalhados nos exemplos deste relatório.

O polinucleotídeo da presente invenção pode ser identificado em seqüências de DNA genômico de plantas para as quais informações genômicas encontram-se disponíveis ao público, ou isolado de várias bibliotecas de polinucleotídeos, ou pode ser sintetizado usando técnicas que são bem conhecidas no estado da técnica (Sambrook et al. "Molecular Cloning, a laboratory manual", CSHL Press, Cold Spring Harbor, NY, 1989) O polinucleotídeo pode ser sintetizado, por exemplo, usando sintetizadores automatizados de oligonucleotídeos (por exemplo, sintetizador de DNA OLIGO 1000M Beckman) para obter segmentos de polinucleotídeos de até 50 ou mais pares de bases. Diversos destes segmentos de polinucleotídeos podem ser então ligados usando técnicas de manipulação de DNA padrões que são bem conhecidas no estado da técnica (Sambrook et al. "Molecular Cloning, a laboratory manual", CSHL Press, Cold Spring Harbor, NY, 1989). Uma técnica de síntese de polinucleotídeo convencional e exemplar envolve a síntese de um segmento de polinucleotídeo de filamento único, tendo, por exemplo, 80 pares de bases, e hibridizando este segmento a um segmento de 85 pares de bases complementar, sintetizados, para produzir um 'overhang' de 5 nucleotídeos. O próximo segmento pode ser então sintetizado de modo similar, como um 'overhang' de 5 nucleotídeos no filamento oposto. As extremidades "pegajosas" ou coesivas asseguram uma ligação apropriada quando as duas porções são hibridizadas. Deste modo, o polinucleotídeo desta invenção pode ser sintetizado completamente *in vitro*.

Como observado acima, a seqüência de promotor da presente invenção pode ser empregada em vetores recombinantes e/ou de expressão para acionar a transcrição e/ou expressão de um polinucleotídeo de interesse em tecidos radiculares. O polinucleotídeo de interesse pode ser endógeno ou heterólogo para um organismo, por exemplo, uma planta, a ser transformada. Os vetores recombinantes e/ou de expressão da presente invenção podem ser assim empregados para direcionar a transcrição e/ou expressão de um polinucleotídeo, por exemplo, um gene que está presente na planta de tipo selvagem, ou pode ser empregado para prover transcrição e/ou expressão de uma seqüência de DNA que não é encontrada na planta de tipo selvagem, incluindo, por exemplo, um gene que codifica um gene repórter, como GUS.

Em algumas formas de realização da presente invenção, o polinucleotídeo de interesse compreende uma matriz de leitura aberta que codifica um polipeptídeo de

interesse. A matriz de leitura aberta é inserida no vetor em uma orientação sense e a transformação com essa construção genética/vetor recombinante irá geralmente resultar em uma expressão do polipeptídeo selecionado na raiz de plantas. O polipeptídeo de interesse, que será regulado pelo promotor da presente invenção, poderá ser inserido no

5 vetor na orientação sense, antisense ou em ambas as direções. A transformação com um vetor recombinante e/ou de expressão contendo o promotor da invenção regulando a expressão do polinucleotídeo de interesse na orientação antisense ou em ambas as orientações (sense e antisense) irá geralmente resultar na expressão reduzida do polipeptídeo selecionado nos tecidos radiculares.

10 O polinucleotídeo de interesse, como uma sequência de codificação, é ligado de modo operacional em uma sequência do promotor de polinucleotídeo da presente invenção de modo que uma célula hospedeira é capaz de transcrever um RNA acionado pela sequência do promotor ligada ao polinucleotídeo de interesse. A sequência do promotor de polinucleotídeo está geralmente posicionada na extremidade 5' do

15 polinucleotídeo a ser transcrito. O uso de promotor tecido-específico, como a sequência do promotor do gene que codifica uma peroxidase de *Coffea arabica* identificada como SEQ ID NO1, irá direcionar a transcrição do polinucleotídeo de interesse apenas em tecidos radiculares da planta transformada.

O vetor recombinante ou vetor de expressão da presente invenção também pode

20 conter um marcador de seleção que é eficaz em células do organismo, como uma planta, para permitir a detecção de células transformadas contendo o vetor recombinante inventivo. Estes marcadores, que são bem conhecidos, tipicamente conferem resistência a uma ou mais toxinas. Um exemplo deste marcador é o gene *nptII*, cuja expressão resulta em resistência a canamicina ou neomicina, antibióticos que são geralmente

25 tóxicos para células de plantas em uma concentração moderada. As células transformadas podem ser assim identificadas por sua capacidade de crescer em meio contendo o antibiótico em questão. Outros marcadores que podem ser utilizados para construir vetores recombinantes e/ou de expressão contendo o polinucleotídeo da presente invenção podem ser, mas não estão limitados a: gene *hpt* confere resistência ao

30 antibiótico higromicina, gene *manA* e o gene *bar*.

O sistema que usa o gene *manA* (que codifica a enzima PMI - phosphomannose isomerase) de *Escherichia coli* (Miles e Guest. Complete nucleotide sequence of the fumarase gene *fumA*, of *E. coli*. Nucleic Acids Res., 12(8): 3631–3642, 1984), tendo a manose como agente seletivo, é um dos novos sistemas sugeridos como alternativos aos

5 dois primeiros descritos acima (Joersbo et al. Parameters interacting with mannose selection employed for the production of transgenic sugar beet, *Physiologia Plantarum*, 105: 109, 1999 doi:10.1034/j.1399-3054.1999.105117.x). As espécies vegetais que não metabolizam manose sofrem severa inibição de crescimento quando esta é oferecida como única fonte de carbono em meio de cultura. Os efeitos adversos e

10 inibitórios do uso da manose são conseqüências do acúmulo de manose-6-fosfato, produto da fosforilação da manose por uma hexoquinase. PMI promove a interconversão de manose-6-fosfato e frutose-6-fosfato, permitindo assim que a primeira possa ser catabolizada na via glicolítica (Ferguson e Street, 1958. Análise de sistemas gene marcador/ agente seletivo alternativos para seleção positiva de embriões

15 somáticos transgênicos de mamoeiro, Rev. Bras. Fisiol. Veg., 13: 365-372, 2001. ISSN 0103-3131; Malca et al. Advances in the selection of transgenic plants using non-antibiotic marker genes, *Physiologia Plantarum*, 111: 269, 2001; doi:10.1034/j.1399- 3054.2001.1110301.x). O gene *bar* (que codifica a enzima PAT – phosphinothricin-N-acetyltransferase) de *Streptomyces hygroscopicus* (Murakani et al.

20 The bialaphos biosynthetic genes of *Streptomyces hygroscopicus*: molecular cloning and characterization of the gene cluster. Molecular and General Genetics, 205: 42-50, 1986), tendo o glufosinato de amônio (PPT) como agente seletivo, é, dentre os sistemas tipo *gene de tolerância a herbicida*, um dos mais amplamente empregados pela engenharia genética no desenvolvimento de OGMs vegetais. PAT inativa herbicidas que

25 apresentam o PPT como composto ativo mediante detoxificação deste. A detoxificação, que é resultante da acetilação do grupamento amino livre presente no PPT, torna este incapaz de competir de forma inibitória com a glutamina sintetase (GS), possibilitando assim a remoção da amônia tóxica da célula vegetal pela conversão de glutamato em glutamina, reação esta catalizada pela GS (Lindsey, 1992 Molecular cloning of ICAM-3,

30 a third ligand for LFA-1, constitutively expressed on resting leukocytes. Nature, 360: 481 – 484, 1992; doi:10.1038/360481a0).

Alternativamente, a presença do gene quimérico em células transformadas pode ser determinada por meio de outras técnicas conhecidas no estado da técnica (Sambrook et al “Molecular Cloning, a laboratory manual”, CSHL Press, Cold Spring Harbor, NY, 1989), como Southern e PCR.

5 As técnicas para ligar de modo operacional os componentes dos vetores recombinantes ou de expressão inventivos são bem conhecidas no estado da técnica e incluem o uso de ligadores sintéticos contendo um ou mais sítios de endonuclease de restrição, como descrito, por exemplo, em Sambrook et al (“Molecular Cloning, a laboratory manual”, CSHL Press, Cold Spring Harbor, NY, 1989). Genes quiméricos da
10 presente invenção podem ser ligados a um vetor tendo, pelo menos um sistema de replicação, por exemplo, o de *E. coli*, assim após cada manipulação, as construções resultantes podem ser clonadas e seqüenciadas.

Vetores recombinantes e/ou de expressão da presente invenção podem ser usados para transformar uma variedade de organismos incluindo, mas não limitando a plantas.
15 As plantas que podem ser transformadas usando vetores recombinantes e/ou de expressão da presente invenção incluem angiospermas monocotiledôneas (por exemplo gramíneas, milho, grãos, aveia, trigo e cevada...), angiospermas dicotiledôneas (por exemplo *Coffea arabica*, *Arabidopsis*, tabaco, legumes, alfafa, aveia, eucalipto, bordo...), e gimnospermas (por exemplo pinheiro, espruce branco, lariço...). Os
20 protocolos de transformação de plantas já são bem conhecidos no estado da técnica (Manual de transformação genética de plantas. Brasília: EMBRAPA-SPI/EMBRAPA-CENARGEM, Capítulo 3 e 7, 1998). Em uma forma de realização preferida, os vetores recombinantes e/ou de expressão da presente invenção são empregados para transformar plantas dicotiledôneas. Preferivelmente, a planta é selecionada da família das
25 Rubiaceae, mais preferivelmente da espécie *Coffea arabica*. Outras plantas que podem ser transformadas de modo útil com o vetor recombinante e/ou de expressão da presente invenção incluem, mas não são limitadas a: *Anacardium*, *Anona*, *Arachis*, *Artocarpus*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Carica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*,
30 *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoseyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*,

Panicum, Pannisetum, Passiflora, Persea, Phaseolus, Pistachia, Pisum, Pyrus, Prunus, Psidium, Raphanus, Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, e Zea.

Os vetores recombinantes da presente invenção também podem ser inseridos em microorganismos. “Microorganismos” referem-se a organismos microscópicos tais como bactérias, vírus, fungos e protozoários. Preferencialmente a presente invenção tem como microorganismos os organismos escolhidos do grupo das bactérias e fungos. “Bactérias” referem-se a organismos procariotos, com exceção das cianofíceas. Os ácidos nucleicos da invenção podem ser utilizados para conferir tratos desejados em essencialmente qualquer bactéria. Então, a invenção possui uso sobre várias espécies de bacterias, que incluem mas não estão limitadas ao grupo das Actinobacteria, Aquificae, Bacteroidetes/grupo Chlorobi, Chlamydiae/ grupo Verrucomicrobia, Chloroflexi, Chrysiogenetes, Cyanobacteria, Gloeobacteria, Nostocales, Oscillatoriales, Pleurocapsales, Prochlorales, Stigonematales, Deferribacteres, Deinococcus-Thermus, Dictyoglomi, Fibrobacteres/grupo Acidobacteria, Firmicutes, Fusobacteria, Gemmatimonadetes, Nitrospirae, Planctomycetes, Proteobacteria, Spirochaetes, Thermodesulfobacteria, Thermotogae. “Fungos” referem-se a organismos do reino Fungi que podem ser unicelulares ou pluricelulares. Os ácidos nucleicos da invenção podem ser utilizados para conferir tratos desejados em essencialmente qualquer fungo. Então, a invenção possui uso sobre várias espécies de fungos, que incluem mas não estão limitadas ao grupo dos Ascomycota e Basidiomycota. Preferencialmente a presente invenção utiliza espécies de microorganismos selecionadas dos gêneros *Aspergillus*, *Bacillus*, *Escherichia*, *Pichia*, *Saccharomyces* ou *Streptomyces*.

Uma “célula transfectada” ou uma “célula transformada” significa um célula dentro da qual tem sido introduzida, através de técnicas do DNA recombinante, uma molécula de ácido nucleico da presente invenção. As células podem ser de um organismo hospedeiro que incluem, mas não estão limitados a, células bacterianas, células de fungos, células de insetos, célula de plantas e células de animais. Preferencialmente, a célula é uma célula eucariótica de um organismo pluricelular (pex plantas).

O sinal de terminação da transcrição e a região de poliadenilação da presente invenção incluem, mas não está limitado a, sinal de terminação de SV40, sinal de adenilação de HSV TK, sinal de terminação do gene da nopalina sintetase de *A. tumefaciens* (nos), sinal de terminação do gene *RNA 35S* do CaMV, sinal de terminação do vírus que ataca o *Trifolium subterranean* (SCSV), sinal de terminação do gene *trpC* de *Aspergillus nidulans*, e outros semelhantes. Preferivelmente o terminador utilizado na presente invenção é o terminador do gene que codifica para a proteína peroxidase de café.

Os vetores recombinantes e/ou de expressão da invenção podem ser introduzidos dentro do genoma da planta hospedeira desejada através de uma variedade de técnicas convencionais. Por exemplo, introdução mediada por *A. tumefaciens*; eletroporação; fusão de protoplasto; injeção em órgãos reprodutivos; injeção em embriões imaturos; microinjeção de protoplastos de células de plantas; utilizando-se métodos balísticos, tais como bombardeamento de partículas recobertas com DNA e outros. A escolha da técnica irá depender da planta a ser transformada. Por exemplo, plantas dicotiledôneas e algumas monocotiledôneas e gimnospermas podem ser transformadas por tecnologia de plasmídeo Ti de *Agrobacterium*. Os vetores recombinantes e/ou de expressão podem ser combinados com regiões flanqueadoras de T-DNA apropriadas e introduzidas dentro de vetor convencional do hospedeiro *A. tumefaciens*. A função de virulência do hospedeiro *A. tumefaciens* direcionará a inserção das construções gênicas e marcador adjacente dentro do DNA da célula vegetal quando a mesma é infectada pela bactéria. Técnicas de transformação mediadas por *A. tumefaciens*, incluindo desarmamento e o uso de vetores binários, são bem descritas na literatura científica (como mencionado no pedido de patente US 20020152501; Horsch et al. Science 233:496-498, 1984; e Fraley et al. Proc. Natl. Acad. Sci. USA 80:4803, 1983).

Técnicas de microinjeção são conhecidas no estado da técnica e bem descritas em literatura científica e patentária. A introdução de vetores recombinantes e/ou de expressão utilizando-se precipitações de polietileno glicol é descrita em Paszkowski et al. (1984) (Embo J. 3:2717-2722 como mencionado no pedido de patente US20020152501). Técnicas de eletroporação são descritas em From et al. (1985) (Proc. Natl. Acad. Sci. USA 82:5824, como mencionado no pedido de patente

US20020152501). Técnicas de transformações balísticas são descritas em Klein et al. (Nature 327:70-73, 1987, como mencionado no pedido de patente US20020152501). A introdução dos vetores recombinantes e/ou de expressão da presente invenção pode ser feita em tecidos, como tecido de folha, células dissociadas, protoplastos, sementes, embriões, regiões meristemáticas, cotilédones, hipocotilédones, e outros. Preferencialmente a presente invenção utiliza a transformação através da introdução mediada por *A. tumefaciens* utilizando a *A. thaliana* com planta modelo (Clough et al. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *A. thaliana*. Plant J., 16:735-43, 1998). No entanto outros métodos de transformação podem ser utilizados para inserir vetores recombinantes e/ou de expressão da presente invenção, tais como a biobalística, que consiste em uma técnica de transformação direta do DNA que utiliza microprojéteis impulsionados em alta velocidade para carrear o DNA para dentro das células [Rech, E.L; Aragão, F.J.L. Biobalística. In: Manual de Transformação Genética de Plantas (Brasileiro, A.C.M. & Carneiro, V.T.C. eds.), EMBRAPA Serviço de Produção de Informações –SPI. 1998, 106pp], e via tubo polínico. O método de transformação via tubo polínico foi divulgado pela primeira vez por Zhou et al. (1983) (Zhou, G., Wang, J., Zeng, Y., Huang, J., Qian, S., and Liu, G. Introduction of exogenous DNA into cotton embryos. Methods in Enzymol. 101:433-448, 1983) e consiste na aplicação de uma solução de DNA na parte superior da maçã jovem do algodoeiro após a polinização. Utilizando esta técnica, o DNA exógeno pode alcançar o ovário através da passagem deixada pelo tubo polínico e integrar as células zigóticas já fertilizadas, porém não divididas.

Uma vez que as células foram transformadas, por qualquer uma das técnicas mencionadas acima, as células tendo o vetor recombinante e/ou de expressão da presente invenção incorporado em seu genoma podem ser selecionadas por meio de um marcador, como o marcador de resistência a higromicina ou a canamicina. As células de plantas transformadas podem então ser cultivadas para regenerar uma planta inteira que possua o genótipo transformado e, finalmente, o fenótipo desejado. Tais técnicas de regeneração contam com a manipulação de certos fitohormônios em meio de crescimento de cultura de tecidos, tipicamente contendo um marcador biocida e/ou herbicida, que deve ser introduzido junto com a sequência de nucleotídeos desejada.

Regeneração de plantas a partir de cultura de protoplastos é descrita em Evans et al. (1983) (Evans et al. Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176, MacMillan Publishing Company, New York, 1983; e Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985, como mencionado no pedido de patente US20020152501). A regeneração pode ser também obtida através de calos de planta, explantes, órgãos, ou parte da mesma. Tais técnicas de regeneração são bem descritas no estado da técnica, tais como em Leelavathi et al. (2004) [Leelavathi et al, A simple and rapid *Agrobacterium*-mediated transformation protocol for cotton (*G. hirsutum* L.): Embryogenic calli as a source to generate large numbers of transgenic plants, Plant Cell Rep, 22:465–470, 2004]. A transformação de café é realizada eficientemente através da técnica de bombardeamento de micropartículas, como descrito em Ribas et al. (2005) (Ribas, A.F., Kobayashi, A.K., Pereira, L.F.P, Vieira, L.G.E. Genetic transformation of *Coffea canephora* by particle bombardment. Biol. Plantarum 49: 493-497, 2005).

A metodologia para manutenção e crescimento de culturas de microorganismos (bactérias, fungos, leveduras) é bem conhecida por técnicos no assunto. A descrição de tais técnicas também pode ser encontrada em manuais da área como Gerhardt et al., 1994 (Gerhardt, P.; R.G.E. Murray; R.N. Costilow; E. W. Nester; W. A. Wood; N.R. Krieg & G.B. Phillips eds. Manual of Methods for General Bacteriology. American Society for Microbiology, Washington, D.C) ou Brock, 1989 (Brock, T.D. 1989. Biotechnology: A Textbook of Industrial Microbiology. Second edition, Sinauer Associates, Inc., Sunderland, Mass).

A produção de RNA em células pode ser controlada por escolha da sequência do promotor, por seleção do número de cópias funcionais ou através do sítio de integração de polinucleotídeos incorporados no genoma hospedeiro. Um organismo pode ser transformado utilizando um vetor recombinante e/ou de expressão da presente invenção contendo mais de uma matriz de leitura aberta de codificação para um polipeptídeo de interesse.

O polinucleotídeo isolado da presente invenção também tem utilidade no mapeamento do genoma, em mapeamento físico e em clonagem posicional de genes. A sequência identificada como SEQ ID NO1 e suas variantes podem ser usadas para

projetar sondas e iniciadores de oligonucleotídeos. As sondas de oligonucleotídeos projetadas usando o polinucleotídeo da presente invenção podem ser usadas para detectar a presença do promotor do gene que codifica a proteína peroxidase em qualquer organismo tendo seqüências de DNA suficientemente similares em suas células usando técnicas bem conhecidas no estado da técnica, como as técnicas de hibridização de DNA e dot blot (Anderson, M.L.M. Nucleic Acid Hybridization. Springer, Bios Scientific Publishers, New York, 1999, 240 p; Sambrook et al. Molecular cloning a laboratory manual. 2nd edition. New York: Cold Spring Harbor Laboratory Press, 1989).

Os iniciadores de oligonucleotídeos projetados usando o polinucleotídeo da presente invenção podem ser usados para amplificações por PCR. O polinucleotídeo da presente invenção também pode usado para etiquetar ou identificar um organismo ou material reprodutivo do mesmo. Esta etiqueta pode ser obtida, por exemplo, por introdução estável de um identificador de polinucleotídeo heterólogo, não funcional, não disruptivo, em um organismo, sob o controle do polinucleotídeo da presente invenção.

EXEMPLOS

A presente invenção é ainda definida nos seguintes exemplos. Deve ser entendido que esses Exemplos, enquanto indicam parte da invenção, são colocados como forma de ilustração somente, não tendo, portanto, qualquer cunho limitante do escopo das presentes invenções.

Técnicas usuais de biologia molecular tais como transformação de bactérias e eletroforese em gel de agarose de ácidos nucléicos são referidos através de termos comuns para descrevê-los. Detalhes da prática dessas técnicas são bem conhecidos no estado da técnica e descritos em Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed., 1989, Cold Spring Harbor Laboratory Press). Várias soluções utilizadas nas manipulações experimentais são referidas por seus nomes comuns tais como “agarose”, “TBE”, “miniprep”, etc. As composições dessas soluções podem ser encontradas na referência Sambrook, et al. (supracitada).

Exemplo 1

Comparação da seqüência do promotor da presente invenção com as seqüências de promotores descritos no estado da técnica

O promotor da presente invenção foi comparado com seqüências de promotores descritas em documentos de patentes, tais como: US6384207 (promotor de milho específico para raiz), US 6649748 (promotor da peroxidase de batata induzível por estresse), US6703541 (promotor de milho induzível por nematóides), US6395963 (promotores de *Arabidopsis thaliana* induzidos por nematóides), US6262344 (promotor de *Arabidopsis thaliana* induzível por nematóide), US5750386 (promotor de *Nicotiana tabacum* induzido por patógeno de planta), US2005010974 (promotores de milho específicos de raiz), US20010047525 (promotores de milho específicos de raiz), US7196247 (promotor isolado da isoflavona sintetase de soja para dirigir a expressão para raiz), US2006230470 (promotor de arroz para induzir a expressão nos ápices de raiz e caule com o interesse de expressar substâncias contra nematóides), US6927320 (promotores de *Arabidopsis thaliana* específicos de raiz), US2006260013 (promotores de *Medicago truncatula* específicos de raiz), US20060143735 (promotores de *Arabidopsis thaliana* ativos em raiz), US7193135 (promotor do gene codificando para oxigenase 2-oxoglutarate dependente de milho, induzível e específico de raiz), US2005097633 (promotor do gene *CRWAQ81* de milho específico de raiz), US20060107347 (promotor do gene *Cr1Bio* de milho específico de raiz), US2004248304 (promotor de soja Pyk20 regulado por nematóides), US2005070697 (promotor de soja regulado por nematóides), US2003167507 (promotor da isoflavona sintetase I de soja específico de raiz), US2004019934 (promotor de coníferas específico de raiz), US7193136 (promotor induzível de soja expresso em raiz), US7119254 (promotor induzido por nematóides isolado de *Arabidopsis thaliana*), US6906241 (promotor induzido por nematóides isolado de *Nicotiana* - Ntce17), US677428 (promotor do gene que codifica a metalotienina de milho específico para raiz), US6518483 (promotor do gene *Knox1* de milho específico de raiz), US5837876 (promotor de tabaco que dirige a expressão de um DNA heterólogo para o córtex da raiz), US5633363 (promotor de milho específico para raiz), US5459252 (seqüência de promotor de *Nicotiana tabacum* denominada RB7 capaz de dirigir a transcrição específica para raiz), US5401836 (promotor de *Brassica* específico para raiz), US4886753 (promotor de soja específico para nódulos de raiz), US7041873 (promotores de soja regulados por nematóides), US6271437 (promotores isolados de

genes de soja induzíveis por nematóides), US20070050862 (promotor isolado de *Arabidopsis thaliana* (AtCe15) que é expresso exclusivamente na raiz) e US20020032917 (promotor do gene *SCARECROW* de *Arabidopsis thaliana* específico para raiz). As seqüências foram analisadas através do alinhamento feito com o uso

5 aplicativo Clustal X 1.83 com os parâmetros default (Anexo 1). O resultado do alinhamento foi analisado pelo programa Genedoc, para cálculo da % de identidade. A Tabela 1 e o alinhamento demonstrado na Anexo 1 apresentam os resultados obtidos e mostram que o promotor da presente invenção é significativamente diferente dos promotores descritos no estado da técnica.

10 Tabela 1. Porcentagem de identidade entre os promotores descritos no estado da técnica e o promotor da presente invenção.

NÚMERO PATENTE	IDENTIDADE
US6262344	30%
US2003167507-17	30%
US6395963	30%
US6927320	29%
US7196247	29%
US5837876-1	29%
US5837876-2	29%
US2005010974-2	27%
US5633363-1	26%
US2004248304	25%
US6384207	25%
US2005097633	24%
US2005097633-2	24%
US6518483	24%
US7041873-2	24%
US2006260013-3	22%
US6906241-9	21%
US20060107347	21%
US2006260013-1	21%
US5837876-3	21%
US7119254	20%
US5837876-4	20%
US5689056-3	20%
US20070050862	19%
US6271437-1	19%
US5837876-6	19%
US6703541	18%
US2005010974-1	18%
US2006260013-2	18%
US6649748	17%
US2003167507-6	17%
US7041873-3	17%

US20060230470	17%
US20060143735-7	16%
US5837876-5	16%
US7193136	15%
US6271437-2	15%
US5689056-1	15%
US5689056-2	15%
US20060143735-2	15%
US20060143735-5	15%
US20060143735-4	14%
US7193135	14%
US2003167507-9	13%
US2005010974-3	12%
US5459252-9	11%
US20060143735-6	11%
US2004019934-5	10%
US7193136-1	10%
US20060143735-1	10%
US2005010974-4	9%
US20060107347-2	8%
US6774282	8%
US5837876-8	8%
US5689056-5	7%
US7193136-2	6%
US20060143735-3	6%
US5689056-4	6%
US5750386	5%
US7041873-1	5%
US2004019934-7	4%
US7193136-3	4%
US5837876-9	3%
US2004019934-2	2%
US20010047525-2	1%
US20010047525-3	1%
US20010047525-1	0%
US20010047525-4	0%
US20010047525-5	0%
US20010047525-7	0%
US20010047525-8	0%
US20010047525-6	0%

Exemplo 2 - Isolamento da sequência de DNA promotora do gene que codifica a proteína peroxidase

Para identificação da região promotora da presente invenção foram selecionadas seqüências expressas ou ESTs que apresentassem padrão de expressão em raiz. Estes ESTs foram selecionados *in silico* no banco de dados

do Genoma Café (<http://www.lge.ibi.unicamp.br/cafe/>). Para tal foram utilizadas as ferramentas de *northern* eletrônico disponíveis no site. A identificação dos ESTs expressos em raiz foi feita comparando-se os contigs presentes em dada biblioteca com os contigs presentes nas demais bibliotecas. Os contigs formados por seqüências provenientes unicamente da biblioteca de raiz foram considerados tecido-específicos. Tais contigs foram então analisados quanto à presença dos mesmos em bibliotecas oriundas de raízes submetidas a determinados estresses (raízes infectadas com nematóides e raízes tratadas com acibenzolar-S-methyl). Todos os ESTs selecionados e tidos como específicos de raiz foram submetidos a análises de Blast no banco do NCBI (<http://www.ncbi.nlm.nih.gov.br>) visando confirmação da identidade dos mesmos. No caso do EST (CA00-XX-RT8-009-H02-JM.F) usado na clonagem do polinucleotídeo da presente invenção, análises de tblastx no NCBI indicaram que a sua seqüência apresenta alta similaridade a uma peroxidase (secretory peroxidase) de *Arabidopsis thaliana*, pertencente à classe III da superfamília das peroxidases.

A especificidade de expressão dos ESTs selecionados nas análises *in silico* foi validada por RT-PCR empregando-se RNA total extraído de diferentes órgãos/tecidos de café. Para tal, pares de oligonucleotídeos específicos para cada EST identificado no banco foram sintetizados e utilizados nas reações de amplificação. Nesse caso, a seqüência de nucleotídeos do EST mais representativo do contig selecionado foi utilizada para desenho dos oligos.

No caso do EST usado na clonagem do polinucleotídeo da presente invenção, os oligos desenhados para a validação via PCR foram: 5' TCGAACACCAAGTCTTGAC 3' (forward) e 5' TGGATCAGTGTCACCTTTGC 3' (reverse).

As reações de PCR foram realizadas como descrito a seguir: 0.5 µl de cDNA, 1X PCR Buffer; 1,5 mM MgCl₂; 0.2 mM dNTP mix; 1 µM dos oligos forward e reverso; e 5 u de Taq DNA polimerase. As condições de ciclagem foram: desnaturação inicial a 94°C por 5 minutos, seguida de 30 ciclos de 94°C por 45 segundos, 52°C por 40 segundos e polimerização a 72°C por 1 minuto.

A região do DNA localizada a montante (5') do EST validado por RT-PCR como apresentando expressão específica em raiz (CA00-XX-RT8-009-H02-JM.F), correspondente portanto à região regulatória do gene em análise, foi amplificada empregando a técnica de *genome walking* (Devic et al. Efficient PCR walking on plant genomic DNA. Plant Physiol Biochem, 35: 1-9, 1997) e DNA genômico extraído de plantas de café. As regiões promotoras putativas foram amplificadas utilizando-se o kit *Universal GenomeWalker* (Clontech; Cat. nº K1807-1), seguindo-se a risca o protocolo fornecido pelo fabricante e sucintamente esquematizado na Figura 1: A) Isolamento de DNA genômico; digestão deste com enzimas de restrição *DraI*, *StuI*, *EcoR V* e *Pvu II*; ligação de adaptadores. B) Utilização do fragmento de restrição ligado aos adaptadores como molde de amplificação via PCR, utilizando os oligonucleotídeos AP e GSP. C) Clonagem dos fragmentos amplificados em vetor pGEM-Teasy e inserção em *E. coli* cepa DH5 α .

Para tal, dois oligonucleotídeos gene-específicos reversos (GSP1 5'-CAATGGCATTGTCCAAGTCTCTTTCTC-3' e GPS2 5'-CCACGCGTCCGATTGGTTCGTAGCA-3'), dispostos em sequência, foram sintetizados com base na sequência alvo (read depositado no banco - CA00-XX-RT8-009-H02-JM.F). Esses oligos (GSPs) foram posicionados o mais próximo da região 5' do EST validado e cuja sequência estava disponível no banco.

Para a construção das bibliotecas para a realização do *genome walking*, quatro alíquotas de DNA genômico de café (20 μ g/cada) foram digeridas separadamente com as enzimas de restrição *DraI* e *StuI*, durante 18 horas, sendo a digestão monitorada periodicamente em gel de agarose. Após digestão do DNA genômico com as enzimas *DraI* e *StuI*, adaptadores foram então ligados à porção terminal dos fragmentos originados para a montagem das bibliotecas, exatamente como recomendado no protocolo descrito no *Universal GenomeWalker Kit User Manual* (Clontech; Protocol PT3042-1) (etapa A da Figura 1). Visando a amplificação das regiões promotoras, alíquotas de DNA das referidas bibliotecas (1 μ l) foram submetidas a sessões de amplificação por PCR (de acordo com o protocolo descrito no *Universal GenomeWalker Kit User*

Manual) utilizando-se dos oligos GSP1/GPS2 descritos acima e oligonucleotídeos complementares aos adaptadores fornecidos pelo kit (AP1-5'-GTAATACGACTCACTATAGGGC-3' e AP2-5'-ACTATAGGGCACGCGTGG-3') (etapa B da Figura 1).

5 O fragmento único de aproximadamente 2 kb amplificado na etapa de *genome walking* (como representado na Figura 1) foi retirado do gel de agarose com auxílio de uma lâmina descartável, e purificado utilizando-se o kit *QIAEX II Gel Extraction* (Quiagen) de acordo com as especificações fornecidas pelo fabricante. O DNA foi eluído da coluna utilizando-se água ultrapura a 70°C.

10 O produto de amplificação purificado foi em seguida inserido no vetor pGEM® -T Easy (Promega) especificamente usado para inserção de produtos de PCR que possuam um par de nucleotídeos de adenina em suas extremidades (overhang) (etapa C da Figura 1). Aliquotas de aproximadamente 150 ng de DNA foram dispostas em microtubo de 1 ml juntamente com 50% de Tampão de Ligação 2X, 50 ng de vetor e 5 u de T4 DNA ligase. A reação de ligação
15 permaneceu por 16 horas à 14°C. Após esse período, uma alíquota do produto de ligação foi inserida em *Escherichia coli* cepa DH5α por eletroporação como descrito (Sambrook et al. Molecular Cloning, a laboratory manual. CSHL Press, Cold Spring Harbor, NY, 1989).

20 Imediatamente após o pulso (2.200-volts), as células foram ressuspensas em 1 ml de meio Luria-Bertani (LB) líquido (10 g triptona; 5 g extrato de levedura e 10 g NaCl), transferidas para tubo de microcentrifuga de 1.5 ml e mantidas sob agitação constante de 300 rpm a 37°C durante 1 hora. Em seguida, 300 µl da cultura foram plaqueados em placas de Petri contendo meio
25 LB sólido acrescido do antibiótico apropriado, no caso ampicilina (100 mg/l) acrescido de IPTG (24 mg/l) (isopropyl-β-D-thiogalactopyranoside) e X-gal (20 mg/l). As placas foram mantidas em estufa a 37°C durante 16 horas.

Foram selecionados três clones de cada evento de transformação para validação quanto à presença do plasmídeo recombinante. Prováveis colônias
30 recombinantes (brancas) foram repicadas em 3 ml de meio LB líquido acrescido de ampicilina 100 mg/l e incubadas sob agitação constante de 300 rpm a 37°C

durante 16 horas. Após o período de incubação as amostras foram centrifugadas e sucessivas minipreparações de DNA plasmidial foram realizadas, utilizando-se o *QIAGEN Kit Miniprep* (Quiagen), conforme descrito pelo fabricante.

Os plasmídeos foram analisados via padrão de digestão com *EcoRI* e visualização em gel de agarose 1% corado com brometo de etídeo. As digestões foram realizadas a 37°C na presença de tampão de reação 1X, 30 ng de DNA plasmidial e água. Confirmada a presença do inserto de tamanho esperado (2 kb), o DNA plasmidial (200 ng) foi submetido a uma reação de seqüenciamento empregando o *BigDye™ Terminator Version 3.1 Ready Reaction Cycle Sequencing Kit* (Applied Biosystems). Foram seqüenciados 3 clones de cada evento utilizando um equipamento automático modelo ABI PRISM 377 (Applied Biosystems).

As seqüências foram analisadas utilizando-se o programa Chromas 2.30 e um consenso a partir dos três clones seqüenciados foi obtido. Um PCR de validação foi realizado a fim de confirmar a sobreposição da região amplificada em relação ao seu respectivo EST. Para tal foram utilizados dois oligos (TGAGGCTTCGCTGTCTTTG e CTCGGTCTCACCATACAAAC) desenhados com base na seqüência promotora amplificada e os oligos reversos gene-específicos GSP1 e GSP2 utilizados nas reações de *genome walking* descritos anteriormente.

As seqüências obtidas foram submetidas a análises comparativas com seqüências depositadas no banco de dados do NCBI e Genoma Café usando a ferramenta Blastn. Essas análises revelaram a ausência de qualquer identidade ou similaridade significativa com as seqüências depositadas nesses bancos, indicando que a seqüência correspondente ao promotor **CaPEROX** identificada na presente patente (SEQ ID NO1) era até então desconhecida. Ao mesmo tempo foram realizadas buscas por motivos regulatórios normalmente presentes em regiões promotoras de genes de plantas utilizando o banco do PLACE database (<http://www.dna.affrc.go.jp/htdocs/PLACE/>) (Higo et al. Plant cis-acting regulatory DNA elements (PLACE) database:1999. Nucleic Acids Research, 27: 297-300, 1999). Os elementos regulatórios encontrados ao longo da seqüência de

nucleotídeos da região promotora da presente invenção estão listados na Figura 2: TGACG, TGTCA, CATATG, AAAGAT, CTCTT, GACAAT, GAGAC, CAATTG, GATA, CANNTG, TATA Box.

5 Exemplo 3 - Clonagem da sequência promotora do gene que codifica uma peroxidase de café em vetor de transformação de plantas

O fragmento amplificado (2 kb) contendo o promotor obtido por *genome walking*, como descrito anteriormente, foi clonado no vetor pRT103-Gus (Figura 3) (pRT103 modificado; Topfer et al. A set of plant expression vectors for transcriptional and translational fusions. Nucleic Acids Res., 15:5890, 1987) visando ensaios expressão transiente em café. Para tal, oligonucleotídeos específicos (5'-CTTAAGCTTGCTGGTCCTAGAAAGTAAAC-3' e 5'-CTAGCCATGGCAGTTGAAGAAAGATAAGAG-3') foram desenhados e usados na amplificação completa da sequência regulatória em estudo.

Sítios de reconhecimento para as enzimas de restrição *Hind*III na região 5' e *Nco*I na região 3' (sublinhados acima) foram adicionados aos referidos oligos para viabilizar a clonagem do produto amplificado no vetor pRT103-Gus igualmente digerido (reação de digestão: 1 µg de DNA; tampão de enzima diluído para 1X, 1 u de enzima de restrição e água deionizada para um volume final de reação de 20 µl – a reação foi incubada a 37°C por 1h). Nesse caso, uma digestão parcial do referido vetor com a enzima *Hind*III teve que ser realizada uma vez que o mesmo apresenta dois sítios de restrição para tal enzima em sua sequência de bases. As digestões foram planejadas de maneira a permitir que o promotor 35S presente no vetor pRT103-Gus fosse substituído pela sequência promotora da presente invenção, sendo a mesma inserida em fusão transcricional com o gene repórter *uidA* (GUS).

As ligações foram realizadas na proporção de 3:1 (inserto : vetor) durante 18 horas a 16°C, sendo os produtos de ligação posteriormente utilizados para transformação de *E. coli* eletrocompetente, cepa DH5α. As transformações foram realizadas em eletroporador (BioRad) como descrito no exemplo 2.

30 Cinco clones recombinantes foram então repicados e colocados para

crescer em meio LB líquido seletivo a 37°C durante 16 horas sob agitação constante (200 rpm). Minipreparações de DNA plasmidial dos clones obtidos e ensaios de digestão com enzimas de restrição foram realizados para confirmar a correta inserção do fragmento promotor no vetor. A seguinte construção foi então obtida: gene quimérico contendo o promotor **CaPEROX** da presente invenção fusionado transcricionalmente ao gene *uidA* (Figura 4).

O gene quimérico descrito acima foi também transferido para o vetor pCAMBIA-1381 (Cambia) (Figura 5) visando os ensaios de transformação estável em tabaco. Para tal, o conjunto promotor/gene repórter foi liberado do vetor pRT103 por digestão com as enzimas *HindIII* e *NcoI* (vide reação de digestão descrita anteriormente) e inserido no vetor de expressão de plantas pCAMBIA-1381 igualmente digerido.

As ligações foram realizadas na proporção de 3:1 (inserto:vetor) durante 18 horas a 16°C, sendo os produtos de ligação posteriormente utilizados para transformação de *E. coli* eletrocompetente, cepa DH5α (vide exemplo 2). As bactérias foram então plaqueadas em meio LB seletivo contendo antibiótico apropriado (canamicina 100 µg/ml) e incubadas em estufa a 37°C durante 18 horas.

Três clones foram repicados em meio LB líquido seletivo e colocados para crescer a 37°C durante 16 horas sob agitação constante (200 rpm). Após a obtenção do DNA plasmidial através de miniprep empregando o *QIAGEN Kit Miniprep* (Quiagen), estes foram seqüenciados em seqüenciador automático como descrito anteriormente, visando confirmar a correta inserção do promotor no vetor.

Exemplo 4 - Expressão transiente em raiz de *Coffea arabica* via biobalística

Para os ensaios de biobalística foi utilizado o gene quimérico obtido no exemplo 3 e como controle positivo o vetor pRT103-Gus fechado, isto é, contendo o gene repórter *uidA* sob controle do promotor 35S. Os tratamentos usados no ensaio foram:

Gene quimérico 1 – Promotor 35S + gene repórter GUS (Controle positivo; Figura 3)

Gene quimérico 2 - Promotor *CaPEROX* da presente invenção + gene repórter GUS (Figura 4)

5 As versões dos genes quiméricos acima descritas foram bombardeadas tanto na região hipocotiledonar como na hepilotiledonar adotando os procedimentos descritos no protocolo do Manual de Transformação Genética de Plantas, produzido pela Embrapa Milho e Sorgo - Sete Lagoas, com algumas modificações. Nesse caso, foram realizados 8 disparos por planta, sendo quatro
10 na região hipocotiledonar (pressão de 1100 psi), dois disparos de cada lado da plântula e quatro na região hepilotiledonar (pressão de 600 psi), sendo repetida as inversões.

Após os ensaios de biobalística, o material vegetal foi mantido em papel de germinação umedecido com água Milli-Q em câmara de crescimento a 30°C
15 durante 48 horas, tempo após o qual foi iniciado o ensaio enzimático de revelação da atividade GUS.

O material bombardeado foi submerso em solução de 50 mM de fosfato de sódio pH 7.5, 10 mM Na₂EDTA, 0.1% Sarcosil, 0.1% Triton X-100, 5 mM Ferrocianato de potássio, 5 mM Ferrocianeto de potássio, 10 mM β-mercaptopetanol e 1 mM do substrato X-Gluc. As amostras foram mantidas no
20 escuro a temperatura de 30°C durante 16 horas. Após esse período o material foi lavado com etanol 70%, visando o bloqueio da atividade GUS e a completa retirada da clorofila, e então fixado em glicerol 50%. O anexo 2 mostra a expressão transiente do gene repórter GUS em raiz de *Coffea arabica*
25 transformada via biobalística com o gene quimérico contendo o promotor *CaPEROX*. A detecção da atividade do gene repórter GUS (pontos azuis) em raízes transformadas via biobalística revela que a região isolada via *genome walking* apresenta atividade transcricional específica. A confirmação da transformação pontual das células por análise de microscopia óptica também
30 pode ser vista no anexo 2. Nos painéis A –F do anexo 2 são observadas regiões

com coloração azul, resultante da atividade do gene GUS, em células de raiz, visualizadas em diferentes níveis de magnitude. Nesse ensaio foram utilizadas partículas de ouro BioRad 1.0 micron sob pressão de disparo de 1100 psi.

Exemplo 5 - Obtenção de plantas transgênicas de tabaco contendo o promotor ***CaPEROX***

Clones de *Agrobacterium tumefaciens* cepa LBA4404 transformadas com o gene quimérico contendo o promotor ***CaPEROX*** em pCAMBIA-1381 (Figura 5) foram incubadas em 5 ml de meio YEP líquido seletivo (10 g Triptona; 10 g Extrato de Levedura; 10 g NaCl contendo 100 µg/ml de canamicina e 100 µg de rifampicina) com agitação entre 100-150 rpm a 28°C, até atingirem uma absorbância (600 nm) entre 0.5 e 1. Posteriormente, 1 ml do inoculo de cada construção foi transferido para um tubo de centrífuga de 50 ml, ao qual foi adicionado 4 ml de meio YEP líquido. Tais amostras foram mantidas à temperatura ambiente até a utilização.

Plantas de tabaco (*Nicotiana tabacum* SR1) cultivadas *in vitro* foram utilizadas no processo de transformação. Discos foliares de 1 cm² foram retirados com auxílio de um vazador de ferro previamente esterilizado, e mantidos em placa de Petri até o processo de transformação. Para a transformação foram utilizadas cinco placas de Petri com dez discos foliares cada. Os discos foram imersos por 15 minutos no meio YEP contendo as agrobactérias sendo posteriormente secos em papel de filtro estéril e transferidos para placas contendo meio MS sólido (Murashige, T. e Skoog, F. A revised medium for rapid growth and bio-assays with tobacco tissue cultures. *Physiol. Plant.*, 15: 473-497, 1962). Essas placas foram mantidas no escuro durante 48 horas em sala de crescimento a 28°C. Após esse período, os discos foram transferidos para placas de Petri contendo meio MS sólido adicionado de 100 µg/ml de higromicina, 500 µg/ml de cefotaxima, 1 µg/ml de 6-benzilaminopurina (BAP) e 100 µg/ml ácido naftalenoacético (ANA). As placas foram mantidas por aproximadamente 30 dias em sala de crescimento sob fotoperíodo de 16 horas de luz até o surgimento de calos. Como controle

positivo foram utilizados discos foliares não transformados mantidos no meio seletivo acima citado. Posteriormente as pequenas plântulas formadas a partir desses calos foram individualizadas em potes de vidro contendo meio de enraizamento (meio MS descrito acima, porém desprovido do hormônio 6-benzilaminopurina) para indução da formação de raízes.

Foram regeneradas 27 plantas transformadas com o gene quimérico contendo o promotor específico de raiz. A fim de confirmar a inserção do gene quimérico, DNA genômico dessas plantas foi extraído a partir de tecido foliar, utilizando protocolo previamente descrito (Konieczny, A. e Ausubel, F.M. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J., 4:403-410, 1993).

A inserção do gene quimérico nas plantas regeneradas foi confirmada em reações de PCR empregando 2 µl de DNA genômico em tampão de reação 1X, contendo 1,5 mM MgCl₂, 0,2 mM de cada dNTP, 1 µM dos oligonucleotídeos específicos à região promotora (5'-CTTAAGCTTGCTGGTCCTAGAAAGTAAAC-3' e 5'-CTAGCCATGGCAGTTGAAGAAAGATAAGAG-3') e 5 u de Taq DNA polimerase (Invitrogen). Os produtos de amplificação foram analisados em gel de agarose 1% corado com brometo de etídeo. As plantas tidas como positivas na reação de PCR, que apresentavam, portanto o fragmento de tamanho esperado (2 kb), foram submetidas a ensaio histoquímico de GUS (anexo 3) como descrito no exemplo 4.

Para a análise da expressão do gene repórter *uidA* em resposta ao estresse biótico, plantas transgênicas transformadas com gene quimérico contendo o promotor da peroxidase foram infectadas com nematóides da espécie *Meloidogyne javanica* reproduzidos e isolados de tomateiros. Aproximadamente 300 gramas de tecido radicular infectado com *M. javanica* foram triturados durante 30 segundos, com auxílio de um liquidificador comercial Walita, em 200 ml de hipoclorito de sódio a 0,05%. A solução foi então peneirada em telas de nylon de 250 e 25 µM, respectivamente. O material depositado na peneira de 250 µM foi descartado (restos de tecido radicular) e o retido na peneira de 25

μM (ovos e nematóides em estágio juvenil) foi enxaguado com água em abundância, visando a completa retirada do hipoclorito de sódio. O material foi posteriormente coletado, diluído em 200 ml de água potável e transferido para um Becker de vidro. Esta solução foi mantida sob aeração constante, com auxílio de uma bomba de aquário, e mantida em banho-maria a 28°C durante 72 horas no escuro. Esta etapa é de extrema importância a fim de obter um inóculo homogêneo, com grande quantidade de nematóides em estágio juvenil. Após esse período, uma amostra de 1 ml da solução foi analisada em microscópio óptico em aumento de 20X e uma contagem estimada do número de nematóides foi realizada. Foram inoculadas 60 plantas (12 plantas de 5 diferentes eventos de transformação) todas positivas em ensaio de PCR, utilizando-se para tal, aproximadamente 4.000 nematóides em estágio juvenil/planta. As plantas inoculadas foram submetidas a ensaio histoquímico de GUS, como descrito no exemplo 4, após 24, 72 e 336 (14 dias) horas de tratamento.

O anexo 3 mostra o ensaio histoquímico da atividade da β-glucuronidase em raízes de tabacos transgênicos transformados com o cassete de expressão contendo o gene *uidA* sob controle do promotor ***CaPEROX***, em resposta ao ataque de nematóides. Os painéis A, B, C, D, E, F, G, H e I representam o padrão de expressão de GUS observado em todas as raízes das plantas transgênicas submetidas ao estresse biótico após 14 dias da infecção. A expressão do gene repórter (coloração azul) é observada especificamente nas regiões parasitadas pelos nematóides, denominadas galhas, ficando a mesma restrita a região central da raiz (cilindro vascular), local onde os nematóides iniciam a atividade parasitária.

CaPEROX	: CCTAGAAAGTAAACTAAGTATTTCAGATTGAGTTTTTCATTCTTCAGGCTTAGAAGGTAATTAAGCATTTCATATTG	: 80
US7119254	: -----	: -
US20070050	: -----	: -
US6649748	: -----	: -
US6927320	: TGTATCGAGACAAACGAGAAATCATGATGAGCCTATGCATTAAGTGCCGTTGGTTAATTAGAGGTTCCGATATACATAA	: 80
US20042483	: -----	: -
US6906241	: -----	: -
US5689056	: -----	: -
US6703541	: -----	: -
US6262344	: ACTTTTAAATCACTACATCATCACCAGATATGGATGAATATTTATATAATATTCAGTGCCAATTAGTCTTTTAAACATA	: 80
US20050109	: -----	: -
US20060143	: -----	: -
US6384207	: GCGAACGGGTTTATAGCCCGTGTTCCATTATGAGGACATGAACGGTTTAAACAAAGTTACATATCATCCCAGCTAGCTAC	: 80
US20050109	: -----	: -
US20040199	: -----	: -
US20040199	: -----	: -
US20040199	: -----	: -
US20031675	: TATTATTATTATTATTATTATTATTATTATATACAAAAATCAATGATAGTGGTGTATTATGAAATTAATGTAAGATGG	: 80
US20031675	: TATTATTATTATTATTATTATTATTATTATATACAAAAATCAATGATAGTGGTGTATTATGAAATTAATGTAAGATGG	: 80
US7196247	: TATTATTATTATTATTATTATTATTATTATATACAAAAATCAATGATAGTGGT-----TATGAAATTAATGTAAGATGG	: 75
US20031675	: -----	: -
US7193136	: -----	: -
US7193136	: -----	: -
US7193136	: -----	: -
US7193136	: -----	: -
US20050976	: G-CCACTCATTCGAATGATAATAAATATAGACATACATAACAACATATTTATAGGTTAATTAATAAATATATATTTAGT	: 79
US20050976	: GTCCACTCATTCGAATGATAATAAATATAGACATACATAACAACATATTTATAGGTTAATTAATAAATATATATTTAGT	: 80
US20060107	: -----	: -
US20060107	: -----	: -
US20062600	: -----	: -
US20062600	: -----	: -
US20060143	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US20010047	: -----	: -
US6518483	: GTCTGTTGGCCACTGTCTCTGTGTGTTTGTGTTTGAAGCCATGCATATATGGCTCGCCGCGGAGGCGTCATATATGATGT	: 80
US20050109	: -----	: -
US6774282	: -----	: -
US20060143	: -----	: -
US5689056	: -----	: -
US6271437	: -----	: -
US6271437	: -----	: -
US5689056	: -----	: -
US5689056	: -----	: -
US5750386	: -----	: -
US5459252	: -----	: -
US5837876	: AATTTGTGTTTCTTTTGCCTAATTTATTGTGTTATCCCCTTTATCCTATTTTGTTCCTTACTTATTTATTGCTTCTA	: 80
US5837876	: AATTTGTGTTTCTTTTGCCTAATTTATTGTGTTATCCCCTTTATCCTATTTTGTTCCTTACTTATTTATTGCTTCTA	: 80
US5837876	: -----	: -
US5837876	: -----	: -
US5837876	: -----	: -
US5837876	: -----	: -
US5837876	: -----	: -
US7041873	: -----	: -
US7041873	: GGCCAGCCTGTTAAACGCCGTCGACGAGTCTAACGGACACCAACCAGCGAACCAGCTCGCGTCGGGCCAAGCGAAGCAGA	: 80
US5633363	: GATATATGTGATGTATATATGTGATGTATATGTGACATATGTGATGTATATGTGGTATCTTTTGTGTTGTTGGATGGAAT	: 80
US20050109	: TAAACGTCTCTAGTGGCAGCCATACATATTGTCGGTTCCTGTCCAAAAAECTCTACCACCTTTCAATAGCTCTTTTGGG	: 80
US7193135	: -----	: -
US20060230	: -----	: -
US20060143	: -----	: -
US20060143	: -----	: -
US5689056	: -----	: -
US6395963	: TTGGTAATTTTGTGGAATTGACATGTGATACATGAATATATCTATGTAGAGTGAAAGAAAAATTAATGATGGTTAGCATAA	: 80
US20062600	: -----	: -
US20060143	: -----	: -
US20060143	: -----	: -

		*	100	*	120	*	140	*	160	
CaPEROX	:	ATTAGTCTTAATTCACTAAATATCCAACCATGATGATTTGTCAACAAGTTGCTAATATTGGGGTGCAAAATTATCAATAT								: 160
US7119254	:	-----								: -
US20070050	:	-----								: -
US6649748	:	-----								: -
US6927320	:	ACCAGTAGACATATGGATAAATATGAACACACACACCAAAAAAGTGGGAAATCTAAATAAGTGTAGAGAATAATAAGTCC								: 160
US20042483	:	-----								: -
US6906241	:	-----								: -
US5689056	:	-----								: -
US6703541	:	-----								: -
US6262344	:	TATATGCTGCTTGACATATAGGTCATCCAAATTTTATAGGTTCAAAACAAAACCAAAAGAAACAGAAAAGATCTGATAA								: 160
US20050109	:	-----								: -

		*	180	*	200	*	220	*	240	
CaPEROX	:	AAAAGTTTGACGTACAAAATCATGAAAAAAAAAATTAGAACTCTCAAGTGTAATTTATTCAAATTTATTGCTCGGTCTCAC	:	240						
US7119254	:	-----	:	-						
US20070050	:	-----	:	-						
US6649748	:	-----	:	-						
US6927320	:	TCAGGTGGGAGATTCAAAGAGAGGACAATGAAGGGTATATAGACTCTAAACAAAAATGGCATGACTTAGTGGAGAGGGTT	:	240						
US20042483	:	-----	:	-						
US6906241	:	-----	:	-						
US5689056	:	-----	:	-						
US6703541	:	-----	:	-						
US6262344	:	AAAGTCTTCATTTTGTAGACGAGGGATCAAACCTTATCTAGTGCATCTGAATGAAAAAAAAATGATCTTAACACTGCAGGTGAA	:	240						
US20050109	:	-----	:	-						
US20060143	:	-----	:	-						
US6384207	:	ATATCACACGTCAGCTTATATTACGTAAAGTGGGGTTAGTTTTCAAGAAGCGTGGGACCAGTCACCTCTGCAGTCTGACC	:	240						
US20050109	:	-----	:	-						
US20040199	:	-----	:	-						
US20040199	:	-----	:	-						
US20040199	:	-----	:	-						
US20031675	:	GATAATTTAAATTCCTCCTACTATTTTTTATTTATTTAAAAAAGAAATCATACTTAAGAAATTACTATATGAAATACTTT	:	240						
US20031675	:	GATAATTTAAATTCCTCCTACTATTTTTTATTTATTTAAAAAAGAAATCATACTTAAGAAATTACTATATGAAATACTTT	:	240						
US7196247	:	GATAATTTAAAT-----CCTACTATTTTTTATTTATTTAAAAAAGAAATCATACTTAAGAAATTACTATATGAA-----TT	:	221						
US20031675	:	-----	:	-						
US7193136	:	-----	:	-						

```

US7193136- : ----- : -
US7193136- : ----- : -
US7193136- : ----- : -
US20050976 : CGTTGTTGTACCAGACCGCACCTAAATCAAACCTGTTTTCAGAGATGGCCATTCTATTATTGTAGATTTGTGA-----T- : 232
US20050976 : CGTTGTTGTACCAGACCGCACCTAAATCAAACCTGTTTTCAGAGATGGCCATTCTATTATTGTAGATTTGTGA-----TA : 234
US20060107 : ----- : -
US20060107 : ----- : -
US20062600 : ----- : -
US20062600 : ----- : -
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : AGTGCAGGGAGGGCGAGGGCCGCGCCGCTATGGGTGTACAGCAGGGTCCGTCGCCCCGCCCGCCGCGCGTGGCGTGCC : 240
US20050109 : ----- : -
US6774282 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6271437- : ----- : -
US6271437- : ----- : -
US5689056- : ----- : -
US5689056- : ----- : -
US5689056- : ----- : -
US5750386 : ----- : -
US5459252- : ----- : -
US5837876- : AACTTTTTAATTGTAGATATTTCTGATGATTTTATCTCTTACTACCAATATAACGCTTGAATTGACGAAAATTTGTGTCC : 240
US5837876- : AACTTTTTAATTGTAGATATTTCTGATGATTTTATCTCTTACTACCAATATAACGCTTGAATTGACGAAAATTTGTGTCC : 240
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US7041873- : ----- : -
US7041873- : ----- : -
US7041873- : ----- : -
US7041873- : AATTGCGGAGCGGCAGACGTGAGCCGGCACGGCAGGCGGCTCCTCCTCTCACGGCACCCTACGGGGGATTCTCTTTC : 240
US5633363- : GGGCCATAGCACTCGGTAGAGAACCAATACTTAGGCACCGGTAAGCTTTTTGGCGAGTGTGTGGCCCTGGCACTCAG : 240
US20050109 : CTCCGATCTATTAGTTCTTAGACATCTAAGTGCTTTAAGGTTACTCGGTGATTGCGTGTAGGTGTTTTGCAAAGTACTTAA : 240
US7193135 : ----- : -
US20060230 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6395963 : ACCTTTCCTCTGTGATTCACAGTTGGGTCAAGCCTGATTGTTTCTCCTTCCCTAACTTACTCCATCCATTAGGTAATCTA : 240
US20062600 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -

```

```

*          260          *          280          *          300          *          320
CaPEROX : CATACAAACACAATATACAAGTCAAAACCAATCATCTCATTTCAATTCTCATATTACAGTAACGTGTGATTTGCATGTCA : 320
US7119254 : ----- : -
US20070050 : ----- : -
US6649748 : ----- : -
US6927320 : TTAAATTGAAACAAGTAGGATTGAAGAACAAGAAAACAAGAAGCATGCCCTAGATTTCTGAGATAATAATTACACATTG : 320
US20042483 : -----CCTCTTTTAGTGCTATTTAATGTCCATTAAT : 31
US6906241- : ----- : -
US5689056- : ----- : -
US6703541 : ----- : -
US6262344 : GGCTGGCTCAATCTTTGACAATATATTGATCTGCGATGACCCGGCATATGCAAGAAGCATCGTGGATGACTATTTTGCCC : 320
US20050109 : ----- : -
US20060143 : ----- : -
US6384207 : TTGGCTTCAGCTTCGACAGCAAACAGTCATCTCTTGAAGCTAAGGACAGTCTCCAACAGTCAACAAAGCAGCGGTCTGC : 320
US20050109 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : ATCAGCTACGGATAAAAAATAAGCTAATAAAATTCGTATGTAATAGAGAATTAATTGACTTTGGTGGGTGCAATACAGT : 320
US20031675 : ATCAGCTACGGATAAAAAATAAGCTAATAAAATTCGTATGTAATAGAGAATTAATTGACTTTGGTGGGTGCAATACAGT : 320
US7196247 : ATCAGCTACGGATAAAAAATAAGCTAATAAAATTCGTATGTAATAGAGAATTAATT----TGCTTGGGTGCAATACAGT : 296
US20031675 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : CGATGTACTTTTTTATCCATAAAAATACCATACCATATGATATGGATATCTTGATGAGAGGGACTCATTATCTCTCTCTC : 312
US20050976 : CGATGTACTTTTTTATCCATAAAAATACCATACCATATGATATGGATATCTTGATGAGAGGGACTCATTATCTCTCTCTC : 314
US20060107 : -----TATAGTAGATGTAGATCACAACCTTTTATTATT--TTAAGCTGC : 43
US20060107 : ----- : -
US20062600 : -----GCTGGAGTTCGAATCCCGAAC--AC : 23
US20062600 : -----GTGATCTGCAGGTGCAGCTTGACAACCGACTAGGTTAGAATATCACATTGAC : 52
US20060143 : ----- : -
US20010047 : ----- : -

```

```

US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : GTGCGGCGGCACCGACAGGCCGCGGTGCGAACAGCTGTGGGCTGTGGGCGTGGCGGGGGCTGGCGCGCACCGCCTCGCTG : 320
US20050109 : ----- : -
US6774282 : ----- : -
US20060143 : ----- : -
US5689056 : ----- : -
US6271437 : ----- : -
US6271437 : ----- : -
US5689056 : ----- : -
US5689056 : ----- : -
US5689056 : ----- : -
US5750386 : ----- : -
US5459252 : ----- : -
US5837876 : AAATATCTAGCAAAAAGGTATCCAATGAAATATATCATATGTGATCTTCAAATCTTGTGCTTATGCAAGATTGATACT : 320
US5837876 : AAATATCTAGCAAAAAGGTATCCAATGAAATATATCATATGTGATCTTCAAATCTTGTGCTTATGCAAGATTGATACT : 320
US5837876 : ----- : -
US5837876 : ----- : -
US5837876 : ----- : -
US5837876 : ----- : -
US5837876 : ----- : -
US5837876 : ----- : -
US7041873 : ----- : -
US7041873 : ----- : -
US7041873 : CCACCGCTCCTACTAGAGATAATGAGCATTGCATGTCTATAAAAAATTACCACATATTTTTTTTGTCACTTGTGTA : 320
US5633363 : CTTTGCCGAGTGCCTCACAGAGCACTCGACAAAGAACCTGACAAATGGACCGCTGGTAAATCCTTTACCGAGTGCAGGT : 320
US20050109 : GTCAGTTAGACCGCTCACTATAACACTTGCTCTAGGTTATGCCTAGTTGAGCTAGACACCTCAAACGGTCAAACCCCTTGA : 320
US7193135 : ----- : -
US20060230 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -
US5689056 : ----- : -
US6395963 : CATTCTGTTTCATTTTTTACGCGTGAAAGTATCTAAGAAAAAGGTATCTCAAAAATATATTTATCATTAGGTTTATGTAA : 320
US20062600 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -

```

```

          *          340          *          360          *          380          *          400
CaPEROX : TCAAGTTCGGATGTATAGAGTTAAACCATAAAAAGATTCCATGCTCGAACATATTTTATCTGAGGCTTCGCTGTCTTTGTT : 400
US7119254 : ----- : -
US20070050 : ----- : -
US6649748 : ----- : -
US6927320 : CTGTTTATATAAGGTAAGAGAATATGACACATTGGTTGGTTTCTTACGGGTAAATGTGAAGAAAAAAATAGTAATATT : 400
US20042483 : GGTTTAAACTATATACCAATACGGGTTTCTTCTGTATACTATTTATGAAAACGTATTATAAGATTAAATATTATAAGAT : 111
US6906241 : ----- : -
US5689056 : ----- : -
US6703541 : ----- : -
US6262344 : AACACAGGGAGGTAGATTTCCAAGTCTTTGTATATCTTTTTGCTTCTTTTGGATAAAATCAAAGAAGTTTTTGTATCTT : 400
US20050109 : ----- : -
US20060143 : ----- : -
US6384207 : TTGTAGTTCTCCCTTGACAGACCAGCTATATCTAGCATCATAACAACGGTAAGATCATCTCTAGCAGACAACTTAGTT : 400
US20050109 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : TAAATGTCAAGATCTAATGTGATACTTTTATATCAATGTTTATTAATAAAAAAGTTACCAATAAATTCATCCTCGTAA : 400
US20031675 : TAAATGTCAAGATCTAATGTGATACTTTTATATCAATGTTTATTAATAAAAAAGTTACCAATAAATTCATCCTCGTAA : 400
US7196247 : TAAATGTCAAGATCTAATGTGATACTTTTATA-----GTTTATTAATAAAAAAGTTACCAATAAATTCATCCTCGTAA : 371
US20031675 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : TATATATATAAAC-CCTATATATCAACAGGCATAAAGAAAAACAGATGATTTTTT-CTGAAGTAGAGTGACAGAAGC : 390
US20050976 : TATATATATAAACACCTATATATCAACAGGCATAAAGAAAAACAGATGATTTTTTCTGAAGTAGAGTGACAGAAGC : 394
US20060107 : ATGAGGTTCAAGTTGTAGTATAGTAAATAAATAAATATGCGATCCAATCATCTCTGCTATCCGATCGGCCGGTCTCATAA : 123
US20060107 : ----- : -
US20062600 : CCCACTTCTC-----CACATTTAAATGTGTGAACCTCAA-GTCACTAT-----GTTATTTGACCAAAAAACAAATGA : 91
US20062600 : CTTATCACTTATATAGTGACCTAAGATAAATAAATGCGACATTACTGTTACTTAATATCCAACATATAAACAGGGTGG : 132
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : TCGCGGAATCCAACGCGGCCAACCCACCCCAACGCGTGGCGGGGGCTCAAGCTGAACACCGCGCGGGCTGCCGAC : 400
US20050109 : ----- : -
US6774282 : ----- : -
US20060143 : ----- : -

```

```

US5689056- : ----- : -
US6271437- : ----- : -
US6271437- : ----- : -
US5689056- : ----- : -
US5689056- : ----- : -
US5750386 : ----- : -
US5459252- : ----- : -
US5837876- : TTGTTCAATGGAAGAGATTGTGTGCATATTTTAAATTTTATTAGTAATAAGATTCTATATAGCTGTTATAGAGGGA : 400
US5837876- : TTGTTCAATGGAAGAGATTGTGTGCATATTTTAAATTTTATTAGTAATAAGATTCTATATAGCTGTTATAGAGGGA : 400
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US7041873- : ----- : -
US7041873- : ----- : -
US7041873- : GTGCAGTTTATCTTTATACATATATTTAACTTTACTCTACGAATAATATAATCTATAGTACTACAATCAGTGTTTTAGA : 400
US5633363- : CAGTAGACACTCGGCAAAGGTAACCTCTTTGCCGAGTGCCGCTTAGAACATTTGACAAAGGGTCATCTCCGTACCCGGT : 400
US20050109 : TGCTTGTGCGCACCATTATATGCATCGACACTTGCGCGTGTTCGTAGAGTTGTACCAGGAAGGGTGTATCTTGTGAGAT : 400
US7193135 : ----- : -
US20060230 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6395963 : TTAAGTTTAATTTGTAATTGTGTTTGTAAATCGTCTAGTAGGTATACGGGACAAGGGCATATGCTCAGTCAAAGTTGGCC : 400
US20062600 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -

```

```

          *      420      *      440      *      460      *      480
CaPEROX : AGTTCCTCTCTCAGTCTAAGATTCCAATAGCTAAGTATTCCTTAAAGAATAATTTATCAATTAGTCCTTTGAC : 480
US7119254 : ----- : -
US20070050 : ----- : -
US6649748 : ----- : -
US6927320 : TGAGAAATCTAAATAGTAAAGAGGTATATATGGAGAAGAAGAGAGAAAGGAAAAATAGTGGCAGAGAATGGAGAGA : 480
US20042483 : TACAATATTATTTGATTAAATATACCTCTTTATCTGTGCAGATATATCATCGATATCTTCTCCTCATCTAGCGAGTC : 191
US6906241- : ----- : -
US5689056- : ----- : -
US6703541 : ----- : -
US6262344 : GCAAGGTGTAGTAATTGCAATGGATTTTCTGCATGCTATTATATACGAAATGTCTTATTAGTGAATTTGATATGCTAT : 480
US20050109 : ----- : -
US20060143 : ----- : -
US6384207 : TAATTAATTATGTCTAATCCGTTGTTGTGTAGCTTAACTTTCTAGCCTCCTATGCTAAGAGAGTTCTCTAGTTCTACTCA : 480
US20050109 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : GATTATTTTAAATCTTTGCATTCACTTAAATATTTATATAAAAAAATTGAAACATTATTACAAAGATATTAAATAAGAT : 480
US20031675 : GATTATTTTAAATCTTTGCATTCACTTAAATATTTATATAAAAAAATTGAAACATTATTACAAAGATATTAAATAAGAT : 480
US7196247 : GATTATTTTAAATC-----ATTCACTTAAATATTTATATAAAAAAATTGAAACATTATTACAAAGATAT----TAAGAT : 442
US20031675 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : AGAAGTGTGAGTCTTTTGTTCATTTTATAATGTGTAAGAAAAATGACGCCAATGAAATATGTGTCTGGGCTGACGTG : 470
US20050976 : AGAAGTGTGAGTCTTTTGTTCATTTTATAATGTGTAAGAAAAATGACGCCAATGAAATATGTGTCTGGGCTGACGTG : 474
US20060107 : AGGCAGCCCATACCTGTTGGTC--GTTGTGTCGTCTGATTAAATTCAGGCCGCGTATCGGTTTTCCTTCCCACCAAC : 201
US20060107 : ----- : -
US20062600 : GAAAAA-----AAAGACCAAGCGGGGAAAGTACCCATTACCAAGTCGATACAAATAC : 146
US20062600 : ATTCGAATCCTAACCGCTGCTT--ACTTTTCTGCAACCAAAAAATAGGTAACCTTGCTCTGTTGCAATCAGATAAAAAATG : 210
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : TGC GCGTGTGGACCAGCCAGCCAGCCAGCCAAACCAACCAAGCAAGGCATGGCAACCACCGACCCGGCTGCTCTACT : 480
US20050109 : ----- : -
US6774282 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6271437- : ----- : -
US6271437- : ----- : -
US5689056- : ----- : -
US5689056- : ----- : -
US5750386 : ----- : -
US5459252- : ----- : -
US5837876- : TAATTTTACAAAGAACTATAAATATGATTGTTGTTGTTAGGGTGTCAATGGTTCGGTTCGACTGGTTATTTATAAAA : 480
US5837876- : TAATTTTACAAAGAACTATAAATATGATTGTTGTTGTTAGGGTGTCAATGGTTCGGTTCGACTGGTTATTTATAAAA : 480
US5837876- : ----- : -
US5837876- : ----- : -

```

```

US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US7041873- : ----- : -
US7041873- : ----- : -
US7041873- : GAATCATATAAATGAACAGTTAGACATGGTCTAAAGGACAATTGTATTTTGACAACAGGACTCTACAGTTTATCTTTT : 480
US5633363- : GTCGTGACGGCCGCTTTCTTTGCCGAGTGCCTGATAGAAAGTACTCGGCAAGAAGTCGTTGCCAATGTATTGTTTCGCT : 480
US20050109 : CACACTAACCACGTTTATGCTGCGGCCTCCACCGTACACTAGAGCGAAGGAGTTCATGAAGTTTCAGTCAGAAGTTCGAT : 480
US7193135 : ----- : -
US20060230 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6395963 : ACAATTTTACATGCCAAGCCTTGAGCAAGCAATTAAAGGTATATTTCAAATTAATAATTACATCTTCTTTTTTGT : 480
US20062600 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -

```

```

*          500          *          520          *          540          *          560
CaPEROX : ATATATTGCAGTTGAAACTTGGGGAAGAATTACATTAGTGTCTGCGAAACTCAGCTTTAGTTTTCCTACTGGAACCTTT : 560
US7119254 : -----CAGGTCAACGGATCACATGCATCAGCACTATTACAAACAATCCTTTAGGGTATAT : 55
US20070050 : ----- : -
US6649748 : ----- : -
US6927320 : GGTTAGGAGGCAAAAGGCAAAATGTGGAGCTTTGATGATGTTGATGCACGCCGTGAGCTTTTCTTCACGCCCTGCTCCCACTC : 560
US20042483 : CTCAACATATGAATTAGATATTATGTAATCGTCATTGTGAGAATCTTCATCTAAATTTTGCAAATCTTTGAACCTTTTC : 271
US6906241- : -----AT : 2
US5689056- : ----- : -
US6703541 : ----- : -
US6262344 : ATACTTGGCCATATGCACCAGTCTGAGAAGGAGTTATTTGCGGAAGCAGAGAAAGAAAGAAAGCTAGAGAAGATGAGGT : 560
US20050109 : ----- : -
US20060143 : ----- : -
US6384207 : GGTGGATTGATATATAAATTTGGGAATCTTCTAGCGCTCACAAAGGTATGGTACACATCAATCAATGAACGGACAAAGCAAC : 560
US20050109 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : TTAATTTTCTCTCTCATGTATCCTCCTTCAAGAATAACAGCGGATATTCAATTAAATTAATCATTTTATCGCATAATATT : 560
US20031675 : TTAATTTTCTCTCTCATGTATCCTCCTTCAAGAATAACAGCGGATATTCAATTAAATTAATCATTTTATCGCATAATATT : 560
US7196247 : TTAATTTTCTCTCTCATGTATCCTCCTTCAAGAATAACAGCGGATATTCAATT----AATCATTTTATCGCATAATATT : 517
US20031675 : ----- : -
US7193136 : ----- : -
US7193136- : ----- : -
US7193136- : ----- : -
US7193136- : ----- : -
US20050976 : TTGTTTGGTGAAAGCCAATTTTGTGTATATAGGGGGGCCAGAGCCAGTTGTATTTGTTGCCGGGACTGGCGCAAAAAA : 550
US20050976 : TTGTTTGGTGAAAGCCAATTTTGTGTATATAGGGGGGCCAGAGCCAGTTGTATTTGTTGCCGGGACTGGCGCAAAAAA : 554
US20060107 : AAATTTACTATCCGATTTTTTTTTTGGTGCATGCTCGGTAGGGCATTTCCCAACGAAAGTTTCGTACTCTCAACAAAAATTT : 281
US20060107 : ----- : -
US20062600 : TAA----TTGACTTTTATTTTTTGGTT-----AAAAATAA-TAATTGACTTTTTTTAGGGGAAATTAATAATAA : 209
US20062600 : CAAGACATCTAACCTTATATTTTATATCGTCAACTTACAAAAATAAGTAACAAGAATTGTTTGACGAAACGACTTTCC : 290
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : CGGTGCCGCCGCCGACCGCGCTGCCCGCCTGGGGCCTTCTCGCCTCCACAGCGTTTGTACTTTGCGGATTCGGTTC : 560
US20050109 : -----GAACATGACGGATGAAGAATCACTATT : 27
US6774282 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6271437- : ----- : -
US6271437- : ----- : -
US5689056- : ----- : -
US5689056- : ----- : -
US5750386 : ----- : -
US5459252- : ----- : -
US5837876- : TTTGTACCATACCATTTTTTCGATATCTATTTGTATAACCAAAATTAGACTTTTCGAAATCGTCCCAATCATGTCGG : 560
US5837876- : TTTGTACCATACCATTTTTTCGATATCTATTTGTATAACCAAAATTAGACTTTTCGAAATCGTCCCAATCATGTCGG : 560
US5837876- : ----- : -
US5837876- : -----TCATGTCGG : 9
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US7041873- : ----- : -
US7041873- : AGTGTGCATGTGTTCTCCTTTTTTTGCAAAAGCTTACCTATATAATACTTCATCCATTTTATAGTACATCCATTTAG : 560
US5633363- : GAGGTCTCTTTGTCAAGTATTACACTCGGCAAGAGCTGTGCCGAGTGTTCAGACTTTGCCGAGTGGTTTAAGCACTC : 560
US20050109 : AGTGAAGACAACCAAGAGCATCCGAGAGAGGCCAAATACGGAGCGCCACTTATGAGTGGAGAAGGCTCGTGACTATCTGC : 560
US7193135 : ----- : -
US20060230 : ----- : -

```


US20060143	:	-----	:	-
US20060143	:	-----	:	-
US5689056	:	-----	:	-
US6395963	:	AATTTTTTTCACGTAACTAATAATCAGAAATATTATCACTGAATGTGCTTACAGGACAGAAATGCAAACGTGACGATA	:	560
US20062600	:	-----	:	-
US20060143	:	-----	:	-
US20060143	:	-----	:	-

	*	580	*	600	*	620	*	640
CaPEROX :	GAGTTTTGGTTGGAGGAACATGCACA	AATGTGACTAATAAAAAAA	AAAAAAAAATCTTCCACC	GCAAAATTGACTACTTTAAC :	640			
US7119254 :	TCAACCCCGTAACACCATTCGTACCC	ATTAAATCATGAACATTTTCGCAA	AGTTTTCCC	AAAAACGGCGTCGGATAAGGGTTT :	135			
US20070050 :	-----GTTCCAGAAAAGAATCATTGATTACTCGTTGTACATAATATCTCCTTCTAAATTG :				55			
US6649748 :	-----				-			
US6927320 :	ACTCACACCTATGAACATTCTCTCT	TATTTATAATTATATTCACATGTCT	CATGTTACTATGTA	AATGGTGACCACT :	640			
US20042483 :	ATTCAATTGCCAGCTATATTATTC	CCACATGATAATAGCGAAATTTGTC	CAGTTCGATGCAGAAGC	ACCGACAACATCTACT :	351			
US6906241- :	GCCCCATTCATCAAGAATCTAGAT	CATGAAGTGCATAACTATT	TATGCTAAGGAGTGGACT	AAAAGTTATTGTCAATTATA :	82			
US5689056- :	-----				-			
US6703541 :	-----ACTCACACGCTAGCTTCTCTCTATAGTCTATAGAGTGCCGGTTC :				44			
US6262344 :	TTGTAGTTCACAAAAAAGTTCTT	GTCTCTTTTCAAGTCTTCTCT	GTATATCCTAGTTAAC	GAGCATGCCCATACAAATC :	640			
US20050109 :	-----				-			
US20060143 :	-----				-			
US6384207 :	GGTAAGATCCGACCCAGTAA	AAGTAATAGCGTTAGGGCATGT	ACAACCTAGACACTGATGC	ACAGTACTCCAAGTATAAG :	640			
US20050109 :	-----				-			
US20040199 :	-----				-			
US20040199 :	-----				-			
US20040199 :	-----				-			
US20031675 :	AGTATACTAATAATACAGTA	ATAGTTAAATTTATTAAGT	ACATCTATATAAAAA	CATTTTACATGAATTA	AACCACTTAT : 640			
US20031675 :	AGTATACTAATAATACAGTA	ATAGTTAAATTTATTAAGT	ACATCTATATAAAAA	CATTTTACATGAATTA	AACCACTTAT : 640			
US7196247 :	AGTATACTAATAATACAGTA	ATAGTTAAATTTA--AAGT	ACATCTATATAAAAA	CATTTTACATGAATTA	AACCACTTAT : 595			
US20031675 :	-----				-			
US7193136 :	-----				-			
US7193136- :	-----				-			
US7193136- :	-----				-			
US7193136- :	-----				-			
US20050976 :	AAATCCCCTTGGATATGACT	AGATTTCTTTATAACGA	ACTAGGAAC TTCATGTT	AAGGTCGATTAGGATTTAGG	ACTCCA : 630			
US20050976 :	AAATCCCCTTGGATATGACT	AGATTTCTTTATAACGA	ACTAGGAAC TTCATGTT	AAGGTCGATTAGGATTTAGG	ACTCCA : 634			
US20060107 :	CATACTCATTATATACGAT	GATACACCATCATACAT	AGTGACA-----TGACAT	ACAATTA	AAAGCAGAGATATA : 351			
US20062600 :	TT-----GACATTTTTTTTT	TAATTTTTTACAAAAA	ATTGACTTTATGTACCG	GATTATAATTATCATAA	CCCCAC : 281			
US20062600 :	TTTTGGATACGATACTCT	ACTTTATCAGTTTATTACT	TGTTAAACGATTGGT	TACATGGCGACATTCG--	ACCATCATAG : 368			
US20060143 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US20010047 :	-----				-			
US6518483 :	GGTCGCTCGCTCGCTTCT	GGTCCGGTACACGTCAGG	ACTACCTGGATTGCTCT	GTCTCATCAGCCTCGGCT	CCGGCCGTC : 640			
US20050109 :	GTTTTCTTCTTTCGGTAGA	AGATAAATACTCCTCCT	GTCCCTAAATTAATATT	TTGTTTAAACTTTTTT	TATGGATTTCATGTAA : 107			
US6774282 :	-----				-			
US20060143 :	-----				-			
US5689056- :	-----				-			
US6271437- :	-----				-			
US6271437- :	-----				-			
US5689056- :	-----				-			
US5689056- :	-----				-			
US5689056- :	-----				-			
US5750386 :	-----				-			
US5459252- :	-----				-			
US5837876- :	TTTCACTTCGGTATCGGT	ACCGTTCGGTTAATTTTC	ATTTTTTTTTAAATGT	CATTAA	AATTCAC	TAGTAAAAATAGAAT : 640		
US5837876- :	TTTCACTTCGGTATCGGT	ACCGTTCGGTTAATTTTC	ATTTTTTTTTAAATGT	CATTAA	AATTCAC	TAGTAAAAATAGAAT : 640		
US5837876- :	-----AAAAATAGAAT				: 11			
US5837876- :	TTTCACTTCGGTATCGGT	ACCGTTCGGTTAATTTTC	ATTTTTTTTTAAATGT	CATTAA	AATTCAC	TAGTAAAAATAGAAT : 89		
US5837876- :	-----				-			
US5837876- :	-----GTACAGTTATAGATTTATATA				AAAAATCTACGG : 32			
US5837876- :	-----				-			
US7041873- :	-----ACATGGTGGAGCACGA :				16			
US7041873- :	-----CGTCAACATGGTGGAGCACGA :				21			
US7041873- :	TTTAGGGTTAATGGTTTTT	ATAGACTAATTTTTTTAGT	ACATCFATTTTATTC	TATTAGCCTCTAAATTA	AGAAAACTAA : 640			
US5633363- :	AGCAAAAGCGCTCGATTTC	GGTAGTGACGGTTGTTTG	GCAATAGFAAAATC	AGCCCTCTCCCGTGGG	AAAAAACTGGTA : 640			
US20050109 :	GGAGTTACCTGGCTAGG	AGCTTGGCCCTCGCGT	GGGCTACCCCTTTCG	TAGGGGTTCCGAC	GAGGATTAGGAAGA	AGCTT : 640		
US7193135 :	-----				-			
US20060230 :	-----CTCAGGTCACGGATCTTATTTT				TCCA : 28			
US20060143 :	-----				-			
US20060143 :	-----				-			
US5689056- :	-----AAATTGCTGAGATACCTC :				18			
US6395963 :	AATGCAGTCCATCCAGGA	ACTGTTAAACTGGAATTAT	CAGAGCACACAAGG	GCTTTTTACAGGTA	AAAAAACTTTTT : 640			
US20062600 :	-----				-			
US20060143 :	-----				-			
US20060143 :	-----				-			

CaPEROX : TACACAAATGGCAGCTGGTACGTGTGAATTCGGTTGAGCCCTCAATTATTTAGGCCTAAGCACAATTTTTTAAATAAGA : 720

```

US7119254 : TTGGCATTGTTGTTCTTTTTTGTGTGCATATAATTTTCATTTTAAACCGTACTATTGGAAGATTTTAAATTTGGATAAA : 215
US20070050 : ATGATTTATAGTTATGTATAACATGTTTGGCAAGAGATTTGAAATCTTCTATGTATCAATTCATGCTGTACCTCTTT : 135
US6649748 : -----CAAGGCCATGATCAGATCGATAATACCAATGGTACCACCTAACTAGGTGATAT : 54
US6927320 : TAAGTATTTATATATCATGTATATATCTTATAGGTATCATACAAAATGGTCAATGAACTTTTGCATTTCAATCTACTTG : 720
US20042483 : TCAACGCACAATTCAGGACACCCATTTCTTGTGTGTGTTGAAACTTTCAATCATTTGTTCAACATCTTCATCGTCACA : 431
US6906241 : TTGACACATGTTTGTATCAATGTAAAAAATAGGAGTTTGCAGAAATTCATTTATATAGGCAGGTGTCATATGCTTTTG : 162
US5689056 : ----- : -
US6703541 : ATACCAATTTTTATGGCACCCGTCGTGTGGCCCTATCATATATAATACATATTTGCATATATTCCTTCTTAGCTACC : 124
US6262344 : AACGTTTACAGGAAGCTCGGATAGCACGGGAAGAAGGTGAACGCAGGCGGAAGAAAGGGACCACCGGTATGGAGACAGG : 720
US20050109 : ----- : -
US20060143 : ----- : -
US6384207 : ACACAATAAAACACAACATAATAATACAGTGGTTATATCTAAACATGTGTCTTACCATATTCATTTGACCAATTAGAA : 720
US20050109 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : ATTGTTTTTATACAAAGGGGAAATGCATTTGCCATTGGGCAACTCTCAAAATAAACTTGGCCAGTCCACACCAATTATCA : 720
US20031675 : ATTGTTTTTATACAAAGGGGAAATGCATTTGCCATTGGGCAACTCTCAAAATAAACTTGGCCAGTCCACACCAATTATCA : 720
US7196247 : ATTGTTTTTATAC--AGGGGAAATGCATTTGCCATTGGGCAACTCTCAAAATAAACTTGGCCAGTCCACACCA--TATCA : 671
US20031675 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : TCGGCAATCTATTCTACGGGCTATATGAGGAGGATAGGGTCCCTAAGTGACATTCACACATCTCATCATAGAAGTAA : 710
US20050976 : TCGGCAATCTATTCTACGGGCTATATGAGGAGGATAGGGTCCCTAAGTGACATTCACACCACTCATCATAGAAGTAA : 714
US20060107 : GAAAGAGCTTATGGGAGATGGTAGAGTTTCATAGAGATAAAATCTATATATACAATTACCTAGTTTAAATATGGTGTGA : 431
US20060107 : ----- : -
US20062600 : ATAATATAACGTCATAGTTTAAATTGACAGTTGGTCTGAAACAAAATATAGTT--TAAGTTGTTAGTTTTTTAGAGGAGT- : 359
US20062600 : A-GCTCTAGCGTC--GACACGATTATGAGTTAGCCTCAATCTGTTTTCGCTATGTTTGTAGTTTCTGTTTATTTGGA : 445
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : CGTGCTCCTCCCATCCCTTGCTGCTGCTGCTGCTGCGTTGCATTGCATTGCAGCTGTACATACAGGACCACACAT

```

GCGAC : 720

```

US20050109 : TAATTAATGTATGTGTTTATATATATGTCTAGATTTATAATTATTCATATGAATATGGACATAAAAACTAGGGCTAAA : 187
US6774282 : ----- : -
US20060143 : ----- : -
US5689056 : ----- : -
US6271437 : -----CGTCTTCTCCAACCTTTCTTAGTGTCATTTTGTGACTAGAT : 40
US6271437 : -----CGTCTTCTCCAACCTTTCTTAGTGTCATTTTGTGACTAGAT : 40
US5689056 : -----AAAGTCCAGCGCGGCAACCGGGTCC--TCTATAATACATTTCTCT : 44
US5689056 : -----AAATGTAGCGCGCCACCGCTTTTCACTCTATAAATACGTTTCTT : 46
US5750386 : ----- : -
US5459252 : ----- : -
US5837876 : GCAATAACATACGTTCTTTTATAGGACTTAGCAAAAGCTCTCTAGACATTTTACTGTTTAAAGGATAATGA--ATTAA : 717
US5837876 : GCAATAACATACGTTCTTTTATAGGACTTAGCAAAAGCTCTCTAGACATTTTACTGTTTAAAGGATAATGA--ATTAA : 717
US5837876 : GCAATAACATACGTTCTTTTATAGGACTTAGCAAAAGCTCTCTAGACATTTTACTGTTTAAAGGATAATGA--ATTAA : 88
US5837876 : GCAATAACATACGTTCTTTTATAGGACTTAGCAAAAGCTCTCTAGACATTTTACTGTTTAAAGGATAATGA--ATTAA : 166
US5837876 : ----- : -
US5837876 : ----- : -
US5837876 : TTCTTCAGAGAAACCTAAAAATCGGTTCCGTTGCGGACGGTTTCGATCGGTTTAGTCGATTTTCAAATATTCA---TTGA : 108
US5837876 : ----- : -
US5837876 : ----- : -
US7041873 : CACTCTCGTCTACTCC-----AAGAATATCAA--AGATACAGTCTCACCAAAGGGCTA---TTGAG---ACTTT : 77
US7041873 : CACTCTCGTCTACTCC-----AAGAATATCAA--AGATACAGTCTCACCAAAGGGCTA---TTGAG---ACTTT : 82
US7041873 : AACTCTATTTTAGTTTTTTTTATTTAATAATTTAGATAAAATAGAATAAAATAAAGTGACTAAAAATTAACAAATACCCCT : 720
US5633363 : GGATCTGGCTCGTGGCTAAGATTCTTTCTTCCCTTTGTAAAAAAGAGAGAAAAAAGGAGCTGTCACGGTGCCT : 720
US20050109 : ACATATTTCTTGATACCTTAGTAAAAAATATCAGCGTGTCCACATGACTTTATATCTCTATCTCATTTAAGCTTTCGCAT : 720
US7193135 : ----- : -
US20060230 : TTATAATATATATAATAAATAAATATATGTTTACTTTTATTATAGTACTTTAAAGATAAATCTATATATGTTGTTCTAG : 108
US20060143 : ----- : -
US20060143 : ----- : -
US5689056 : CTTTGTCTTTTTTCCACTGTGGTCTCAGTTGACTATAAGAAATGTTGCTAACTATTTTAAATGTCCAAACCTCAAATATC : 98
US6395963 : TTTCCATACATATATATAGAAATGCAAAAGGAATCTAATAAGTAAGAAAAAATCCTAAAACCAATTGATGTGGTTTTT : 720
US20062600 : -----CCACTTGTCTTTATTTTAAAGGGCA : 26
US20060143 : ----- : -
US20060143 : ----- : -

```

```

*          740          *          760          *          780          *          800
CaPEROX : AAGACAAACAAACTTTATTAATGAGATTGTTGGAATTTGTCTAGCACAAATGTTTGATATAATTTAGAGGAAAAATTC : 800
US7119254 : GATGATTCATACATAGTCGCTTTGTTGTTACTAGTGATAAATTCATGTTAATGATTCTATGATTGCCAGCTATCTCATTA : 295
US20070050 : TTTATCCCTCGAGTTCCTAAGTACTTAGTCCCTAGTGATTAAACCAATTAGCCAGCATATTTCTTTTAGAACATATATC : 215
US6649748 : ATATTATGTATGTCAATTTTAAACTGTATTACAAGACTATTTTTCATTAATTTGGTACAAAGAAAAATTAACAGAA : 134
US6927320 : TTCATTGTAGATGCTAGCTTTTCACATGTTTGGAAATTTAGTCTGGATCTGAAATTTCTTAATTAGCATTTGTTTGG : 800
US20042483 : AATTGCAATGCAATATATTTACTTGCACATAAAATCGGCAAGTCATAGTAGATATACTTTCAATTTTTATCTAATTTTA : 511
US6906241 : ATCATTTACTATCTCGGCTTTCAACATTCACCTAATAACACATGAGGAATACGAATGTGCAATTAGCTTAAGTTGGGAA : 242
US5689056 : ----- : -

```

[illegible][illegible]

[illegible]

	*	900	*	920	*	940	*	960	
CaPEROX :	CTAACATCAATACAATGTTGATCTAATAAAACTAGATGCACATCCTCTATAATCGGACCTAAACCAGACAGAATGTCTTC	:							960
US7119254 :	CTTTTTTTTGGTTTAATATTCTATTATATCCGAATTTTGTTATATACTAAAATGAACATTTGTCTCTGATTACTAGTTAAT	:							455
US20070050 :	AATGTTTGTATTACTATTTCCTTTTAGTATTACCAAAGACTTCGCAGGGGGTGAAAAAAGGATCAACCTGATGTTATC	:							375
US6649748 :	TTTTCAGATGTAATAATGAGTGGACGAAGAGGTGTAGCAATATTGTGATTAACTCTGCTACGATAATTATGATCCAC	:							294
US6927320 :	ATTAAGAGAAAAATAGAGTTTTCATGCACCAGTGTGATAGTAACGTAGTCGCGGAATGTCTAAAACGATTATGAGTTTG	:							960
US20042483 :	CTAATGCCGTCGTCACTTTCAAATATTCTTCCATTCAAATAAACCGAACAAACAATCAAAGAAGACTCATTTTTATAATAA	:							671
US6906241- :	GATATACAATGACAACTACATGTTCTTTGTACTGTGCAGCTTAATTTAACATG-TGATATTAAAGTTAATTACGTATAA	-:							401
US5689056- :	-----	-:							-
US6703541 :	ATATATATAGTAATATATTTCGATTACTAGATAGTTATAGACTAGTTTAAAGCTAAAAAGGAGATATATAATACTCACAA	:							364
US6262344 :	TACCATGTAAGTGTTCCTTCTATCTCAACCACCTTTAAAAAGAATGGTTTATGCATTTTAGTACTGAACTCATCTTAACTGT	:							960
US20050109 :	-----	-:							-
US20060143 :	TGCATTTCGATATTTGCAGGATTTGCTTCAACACTATAACATGGCCTGCAAGTTTGTATAGATTCCAAGAAATTTGTTTCA	:							91
US6384207 :	ACCTAAGTTATCATGGATGACCACGGCTGGCAATTAaaaaaaTaATTTTTGCCCTAAAACCTCTTCTTAATTGGTCTCT	:							960
US20050109 :	-----GGCTGCGGGTATGGGTAGTGGTAGAACAAACCATGCCACGTAACCGTTGGGTTTCCATTT	-:							62
US20040199 :	-----	-:							-
US20040199 :	-----	-:							-
US20040199 :	-----	-:							-
US20031675 :	AATGCACCTTGCCG-CATCGTGATGGATATTT	:							911
US20031675 :	AATGCACCTTGCCG-CATCGTGATGGATATTTAAATTTCTTCTTCACAAAAA-GCCAAGAGAAATGCTTGTAAAAGGTC	:							957
US7196247 :	AATGCACCTTGCCG--ATCGTGATGGATATTTAAATTTCTTCTTCACAAAAA-GCCAAGAGAAATGCTTGTAAA--TC	:							900
US20031675 :	AATGCACCTTGCCG-CATCGTGATGGATATTTAAATTTCTTCTTCACAAAAA-GCCAAGAGAAATGCTTGTAAAAGGTC	:							98
US7193136 :	TAAATTTTTTTATC-CTTTATTATAAAATTTTATTTTATCTTCTTGTGCAAT-TATTTTGAATAAAATACTTTGAATTA	:							139
US7193136- :	-----	-:							-
US7193136- :	-----	-:							-
US7193136- :	-----	-:							-
US20050976 :	CACCTCATTTGGAATGATAATAAATATAGATATACATACAACTATATTCATAGG-TTAATTAAATAAATGTATATTTAGTCT	:							941
US20050976 :	CACCTCATTTGGAATGATAATAAATATAGATATACATACAACTATATTCATAGG-TTAATTAAATAAATGTATATTTAGTCT	:							947
US20060107 :	CACCTAAGAAATGACCTTACAAAACCTGAAACAAAAGATGTCACCTTGTCTTAAAG-CTCACCACCTGAGAAAGGCCTTA--CA	:							657
US20060107 :	-----	-:							-
US20062600 :	AAACTAAAAATGTTTA-CTTAATGGTAGAGTTATAAACCTCGGGTCCGCGTAAG-CATAGCTCAGTTGGCAG-AGACATGC	:							592

```

US20062600 : ACATGACTGCTTTTATCTCTAATGTTGC--TTGAAAGGGTTAGTTGCGTAGATG-CCCCGTGTGATGGTAGTAGTCGT-- : 673
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : TCCGTGCGGTAGCAGTAGCAGCTCCGATCCGTCCAGCAGCAACCACACGTA-CGCTCGCTCGGTTTGCATGTGTTTGC : 958
US20050109 : TTTAAGCTCCCCAACTGCTAGTGATTCTGTAAACAAATAATGTTTATTGTGAT-TTTAACAACAGCCGTCTAGCTCAGT : 425
US6774282 : ----- : -
US20060143 : TTGTTTCTGATAAATTTGAAAGATATGATTTATTTTCTGTCTTGG-TAAGGTC-AGTTATCACATTAAAGACTAAA--T : 215
US5689056 : -----CG-AATCAACTAGTTTATACTCTCT----C : 24
US6271437- : AAAGGGGAAAGAGAGAGGAAACAAGAAAGAGGATAACAGGA-CGAGATT-GGCGACGACGATGACGATTGGGAGAA : 277
US6271437- : -----TT-GGCGACGACGATGACGATTGGGAGAA : 220
US5689056 : TGTTTATCCATCTAAGGTCCTTTGCCGCCGATGAAAACTCTCAAACCCACACA-AGAAACAACAACAACAAGAGGAC : 280
US5689056 : TTTATGCACATCAAAAGATGCTTCCGCCGGCG--AACCTCT-----GAAACAACAACA : 254
US5750386 : ----- : -
US5459252 : ----- : -
US5837876 : ATTTAAATAATATGAAAGGAAACATATTCAATACATTGTAGTTTGTACTACTCAT-AATCGCTAGAATACTT-TGTGCCTTG : 954
US5837876 : ATTTAAATAATATGAAAGGAAACATATTCAATACATTGTAGTTTGTACTACTCAT-AATCGCTAGAATACTT-TGTGCCTTG : 954
US5837876 : ATTTAAATAATATGAAAGGAAACATATTCAATACATTGTAGTTTGTACTACTCAT-AATCGCTAGAATACTT-TGTGCCTTG : 325
US5837876 : ATTTAAATAATATGAAAGGAAACATATTCAATACATTGTAGTTTGTACTACTCAT-AATCGCTAGAATACTT-TGTGCCTTG : 403
US5837876 : -----GGAACATATTCAATACATTGTAGTTTGTACTACTCAT-AATCGCTAGAATACTT-TGTGCCTTG : 61
US5837876 : CTTGACGGTGTATGTCACATATTTATTTAAACTAGAAAAACAGCGTCAA-ACTAGCAAAAATC-----CAACGGA : 337
US5837876 : ----- : -
US7041873- : ATGGACCCCCACCCACGAGGAGCATCGTGAAAAAGAGACGTTCCAACCTCTT-CAAAGCAAGTGGATTGATGTGATGAT : 309
US7041873- : ATGGACCCCCACCCACGAGGAGCATCGTGAAAAAGAGACGTTCCAACCTCTT-CAAAGCAAGTGGATTGATGTGATGAT : 314
US7041873- : -TGGACCCCTCTC----GAGAGTTCCGCTCCACCGTTACTTGTCTCCG-CTGTC-GGCATCCAG-AAATTG-CGTGGCGGA : 946
US5633363- : ACATCATCTGATCTGATCTGATTTGGTCTGATATGATCTGGGTAGTCATTCTT-GCAATGAGCATCTATCAGGTGAACCA : 958
US20050109 : GTTTTGCTTAAGAGATTCAATTAGAAGCCTAAATTAGCGTGAATAGTTGTTCT-GATTTCATCCCTCTCTTAGATGCCAC : 958
US7193135 : ATTTGACTCTGTTTCGTGAAAGATGCACATATGTGTAGAAACCTAGCCGGCCTT-CACATGAAATCATAAACTTCTACGG : 112
US20060230 : TGAACACACCGTGATTCTTTGCCACTTGCATAATATTCTAGCACTGCTATACT-ACACGACGACTGACGGCGACGTCAGT : 346
US20060143 : ATTTTGATCATTTGGAATAAAATTTATAAAGGAAACGAAAGCGCCTTCTCACG-GGTCCCATCCATTGAAATATATTCTC : 134
US20060143 : GATTTTGTATTGCTATTCAAATCATTAAGATAATTGGTCTTAATAATC-TTCTAAACTTTATATGATCTGATTGA : 88
US5689056 : TACTTATGAAATTTGTAGGTTTAAATTTGTGTAGGCTGGGATAGAAAATTGAATTATCAGATCGGATGGACAGTTTT : 338
US6395963 : TAGATTCCTTTTGTGTTTCTGTTTGTATCCCAATCAAGGTTTGTATCATCAATATTATGATGACATCATATACA : 960
US20062600 : CTCTAACTACCAGTATAGTTGAATATTGGGTTGCAACAGATACAAGGGGTATTAATATATTGAGCATATATCTGTTTCC : 265
US20060143 : ----- : -
US20060143 : ----- : -

*          980          *          1000          *          1020          *          1040
CaPEROX : TTTATCAATACTAGGCACTAAGCACA--TAATATTCATACTTATAACATACATATGAATATTTTTTGCACATATTAT : 1037
US7119254 : TAACCTTTAGCTTCACTAATAAAAAATCTCATGATTTTGATACTTCAAAACATATTAATAACAAATTAGCAGCTTTTAA : 535
US20070050 : ATATTACACGTGAATCTCCAGATCGAGATTTGCTACAAGCAATCAAGAAATATTATGTAGGATTCTGTTGGAACGAG : 455
US6649748 : TCACCTAGTCATTTTTCAGACCAAGCAACTAGCTTTGAGTTTATTTGATGTGTTGTTGTTGTTGTTGTTGTTGTTGTT : 374
US6927320 : TGTTTGTGTTGGTTAGATTGGTATTAGTAGGACATCTAACTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT : 1040
US20042483 : ACCTACAGAAATAATTTTCATTAAACAATAACATAAAAAAATCCTAAATTAATAAAAAATTAATAAATATATTACTG : 751
US6906241 : ATTATTAATAATTAATTTTATGATTAGTGACTGATTGTGTTAATAATTTTAAACTGTCAGTGAATATATCTTTAAAT : 481
US5689056 : ----- : -
US6703541 : TTCATATGTCGCCCTCAATCGAACACAGATCACAAATTCATCACTTTTAGTTTCATCCAACACAAGTACATGCTCCAGATCA : 444
US6262344 : TCTAAAAATGTAAGTTTGTATTGATTCTGAATTTCTGTTAGGACGAGCTATGAGGCGCAGAGTGGTGGGCATTGGCAAA : 1040
US20050109 : ----- : -
US20060143 : TTTTCATCGCTGATTTTGCAGTTTGTCTAGAAAGTTGGAGATATCGAGAAAAACACAACCCCTTGCTTCTTCTAAGGG : 171
US6384207 : TGCTTGCAAAATCACCAGCGAACCCATATGAAGGATGCTCAAAATCTGGCCACCGCATCAGGTTGGTGAATGCAACGTA : 1040
US20050109 : GAACCCATTAAACAACTCATGGGTACAGAAATTGACCCAAACCCATACCTTAATAGAGCAAAAACCCACCGGTTTCGGG : 142
US20040199 : -----AATTCATAGAAGTGTGGTGATTACTCCAACAAATCTAATCCATCCATAAAAAATATACACA : 62
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : ----- : -
US20031675 : GTACTAGTTCAGATACAAAATACAACCTTGCCAGATGCCCTCGACTTACTCCAGGCACAACATTTTTTTTAAACCGGAG : 1037
US7196247 : GTACTAGTTCAGATACAAAATACAACCTTGCCAGATGCCCTCGACTTACTCCAGGCA--ACATTTTTTTTAAACCGGAG : 978
US20031675 : GTACTAGTTCAGATACAAAATACAACCTTGCCAGATGCCCTCGACTTACTCCAGGCACAACATTTTTTTTAAACCGGAG : 178
US7193136 : TCTCTTTTTTCCCTAAAAAGATCGAAGATGAGTATTTCAATAAAGTTGAAGATATATTTCAAAAGTATTATAAATTTA : 219
US7193136 : ----- : -
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : AAAATGAAATATATTT-ACCCATCGTATTCCTTATGCATGAAATGTTGATCTACTTGTCTGATGAAAAATACTACGAG : 1020
US20050976 : AAAATGAAATATATTTTACCCATCGTATTCCTTATGCATGAAATGTTGATCTACTTGTCTGATGAAAAATACTACGAG : 1027
US20060107 : AAACATAAACAAAAAATATGTCACATGTTCTGAAGCTCACCCTGAGATGGCCTTACAAACTGAAACAAAAATATG : 737
US20060107 : ----- : -
US20062600 : ATTATATATGATAGGGCTGGGGTTCAACCCCGGACACCCCACTTATTCATCTTGAAAAAT-ATGAATCTAACCACTA : 671
US20062600 : -TCATCGTACCTAAAGTATTTATCCA--CCGTGGTTCTTTTGTATCATTGGTGGCAATGATATATTATATTGTTT : 750
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : CGTGACCGTGAAGGACAGCCGGTTAGTTCTGGCCGCCCTCCGCCGCCGCCGCCGCCGCTTCTGCTGCACGTCGCCA : 1038

```

```

US20050109 : CGGTAGAGCGCAAGGCTCTTAACCTTGTGGTCGTGGGTTTCGAGCCCCACGGTGGGCGCTTTGAGATGTTTATTTTTGTT : 505
US6774282 : ----- : -
US20060143 : AGTAATAATTTTATGTTTTTTTG-TTTCTGAATTATTTTCAGAAATTTCTGTTTTATTTCATTTAATAT--TTA-ATAATATA : 290
US5689056 : ACTTATTTTTGTATTGTGACA-CTTTGTCTTACGT-GGTGCATTCTGTGTACTCGATTGAGGC---GCATGCGAGATA : 99
US6271437 : ACCACAGTCGCAATGTGTTTCGACTTCCGCGATGGAGCGTTGATTTTTAAATAATTCGACAGGGGAGGTCGTGGCGCTTC : 357
US6271437 : ACCACAGTCGCAATGTGTTTCGACTTCCGCGATGGAGCGTTGATTTTTAAATAATTCGACAGGGGAGGTCGTGGCGCTTC : 300
US5689056 : AAGAATACCTTCTCATTGATGCTCCGATGCTCTCCACTTCCCTTTGTATCTCACGAATGGCTTGTTTTTACCACATGTT : 360
US5689056 : -----GTTTCTAGTCCTAAAGCATCCGACGCGCTTCCACTCCCATTTGATCTAACCAATGGGTTGTTTTACCACATGTT : 328
US5750386 : ----- : -
US5459252 : ----- : -
US5837876 : CTAATAAAGATACTTGAATAGCTTAGTTTAAATATA-AATAGCATAATAGATTTTAGGAATTAGTATTTTGAGTTTAAAT : 1033
US5837876 : CTAATAAAGATACTTGAATAGCTTAGTTTAAATATA-AATAGCATAATAGATTTTAGGAATTAGTATTTTGAGTTTAAAT : 1033
US5837876 : CTAATAAAGATACTTGAATAGCTTAGTTTAAATATA-AATAGCATAATAGATTTTAGGAATTAGTATTTTGAGTTTAAAT : 404
US5837876 : CTAATAAAGATACTTGAATAGCTTAGTTTAAATATA-AATAGCATAATAGATTTTAGGAATTAGTATTTTGAGTTTAAAT : 482
US5837876 : CTAATAAAGATACTTGAATAGCTTAGTTTAAATATA-AATAGCATAATAGATTTTAGGAATTAGTATTTTGAGTTTAAAT : 140
US5837876 : CAAAAAATCGGCTGAATTTGATTTGGTTCCAACATTTAAAAAAGTTTCAGTGAGAAAGATCGGTGACTGTTGATGATA : 417
US5837876 : ----- : -
US5837876 : ----- : -
US7041873 : CCTATGCGTATGGTATGACGTGTGTTTGATGACTTCAAACCTACCTATGACGTATGGTATGACGTGTGTCGACTGATGAC : 389
US7041873 : CCTATGCGTATGGTATGACGTGTGTTTGATGACTTCAAACCTACCTATGACGTATGGTATGACGTGTGTCGACTGATGAC : 394
US7041873 : GCGGCAGACGTGAGCCGCGC---GCAGGCGGCTCCTCCTCTCAGCGGACCGGCAGCTACGGGGG---ATTCTTTTC : 1020
US5633363 : ATTAATATTGATGACATTATGAGTTCGAAGATATACTCTAAAGTGTATCTAAATACAGAGACATTCGTTCTGTTCTTTG : 1038
US20050109 : GTTCCTTCTAGAGACACTTGATCCTGCTTGACGGATGAAGGGGTAGTGAGTGAGTGACACACTTGCATGGGTGCTCTA : 1038
US7193135 : TCGGCTAGGTTTGTATACACAACACTCGACAAGAAGGCATATCTGAAACCGATGCTTTTTTTATAAGTGTGTAGTC : 192
US20060230 : TCAGTTTGTGTTGCGCGATCCATCGCGAAGGCTACTCTACCCATCCCATTTTTTTTTTAAAAAATACTATATAAATCTAA : 426
US20060143 : TCTTTTTGCTCTATATAATAATAACGCGTACTAAATTTGTAGTATATATTATACAAAGTCGATATTGATTGTTTGTGTA : 214
US20060143 : ATCTAAACCGTTCAATAAAATTAATTCCTTCAATCTAAACCGTTCAACAAAAATAAGTGAATAATTTTATTTTTTTAT : 168
US5689056 : TTGATATTGTTGTTTCGGTAATTAAGTATTTAATATGATATTGGAAGTACAATTTAGAAATCATAATATCTGAGTTTG : 418
US6395963 : TACATGCCACTAAGAATCAAAATTTACGTTGAATGATGTACACTTTAGATGATTTAATCGCCTCTTCTTGCAATTTCT : 1040
US20062600 : ACAAGTGAATCAATGTCAAATCTAAAAAATATCTCTTAATACCTTAAGTGAATGGATGCTCTAGATTTTAAATGGT : 345
US20060143 : -----AACTCCTAGCCTAAACCTAAA : 21
US20060143 : ----- : -

```

```

* 1060 * 1080 * 1100 * 1120
CaPEROX : TGTTCGAATTGTTAACTTGTTTATTGTGTTTTCCCTTAGATCGAAATCCGAGATGCAAAGAAATGATCCAAAGAGTTAA : 1117
US7119254 : TCGATAATGTGCTTAGATGAACGTTTCGATGAGAACTCTTTTGTTCGAATGCATGTTAAGAACAAGAAGTCGATAGCA : 615
US20070050 : AGAAAGTGTGTTTTGTTTTCCTTAACCTTAAATCTCAGTTTTCTTATAAATTTTATAGCCGCTCTTTATGCAAAACCC : 535
US6649748 : AAAAATAAAAGTTGCATCATTATATATGTTAGATTAAAGTAATGATCAATCAACGTTTAAATTTGCAATTTATCGGCAAG : 454
US6927320 : C--TTTTTATTTTACACCACTTGAGACTTGGGATCGATAGTACTTGAACACTTGGTTGGTTTCATGATTTTGGCCTATA : 1118
US20042483 : ATTTTCACCAACAAAAAGAAATATCAACAAAGTTGTAGTATAGAGAGAAAGAAATTTGG-AGAGATGCGAAATGACTATA : 830
US6906241 : TCATTAGTAACTTGCTAATCGTTAAATAAAAAATATAAAAGGCCATTGAAAAATCAA-ATGGTAGAATACTTCTCTGA : 560
US5689056 : ----- : -
US6703541 : CAGTTCATCACTTTTTAGTTCATCCAACCAAGTACATGCTTTGTTTGTGACAAATGCATGATTGCTTGTTCGAGCCA : 524
US6262344 : GCATTGGGGGAAATATCTATATTTTGCTTTGAATGTGACCTGTTGTAATTTTCATAATTTGTAACTTTTGATTTCA : 1120
US20050109 : ----- : -
US20060143 : CTAGTTTGTCTGTTTACCTAATAACAATTGTCTATATTCATTTGATGATTCGGTTACCTTTAACACTTTGCGCTGGAATAA : 251
US6384207 : AAAAATAATGCATAAATCAGCTCTCTGATCAGTTATATAATCGTGCTTTTAAATATTTCATGCCAGCTTTATCTGACTCA : 1120
US20050109 : TACCGGGTACCCATTGCCATCTCTACGAGTGACCGACAGAAAGCACTCGGCAAGAGCTGGATTTCGGGTAGCGCGGGGTT : 222
US20040199 : A-ACCTAATTTTCATTAAATCTTTTCTGCCCATTTAATAAAAGATCAAAATTCCTGGC--AGCCGAATATTGGTCTTTT : 139
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : ----- : -
US20031675 : T-ATCTTTTCAGTAAAGAAATCAAGATTATTTAATCTCTGGAATATTAATATCTTTTAA--TATTTTGGATGAAAAA--- : 1111
US7196247 : T-ATCTTTTCAGTAAAGAAATCAAGATTATTTAATCTCT--AAATATTAATATCTTTTAA--TATTTTGGATGAAAAA--- : 1050
US20031675 : T-ATCTTTTCAGTAAAGAAATCAAGATTATTTAATCTCTGGAATATTAATATCTTTTAA--TATTTTGGATGAAAAA--- : 252
US7193136 : A-TACCTTAAATTAATTAATTTCTTAATTAATAAACATAAATTAATCAATATATAT--TA-TAGGAATAAAAAA--- : 292
US7193136 : -----ATAA--T--TAGGAATAAAAAA--- : 19
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : T--TGTTGTACACAGCCGACCTAAATCAAACCTGTTTTAGAGGTGGCCATTCTATTA--TTGTAGATTGTGATAC--- : 1093
US20050976 : T--TGTTGTACACAGCCGACCTAAATCAAACCTGTTTTAGAGATGGCCATTCTATTA--TTGTAGATTGTGATAC--- : 1100
US20060107 : T--CATTTGTTTAGCTTGTCACTCTACTTTAG-GAAAAACAAAATCATCGATATGTTT--TTCTTGATGCCTGCTCG--- : 809
US20060107 : ----- : -
US20062600 : A---ATTACTTGA---CCAAAAAACCACCAATCTCTGATG---TAGGAGATCCTC--TGATTAATGTATAATTT--- : 734
US20062600 : G-TTCATCACTCGATGTGCGATGAATTTCAAGATCACTACGACTCTGTGAATTCAGT--TAGTAGTGGTTGATTTCGC : 827
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : T-TATTTCTGCTCAGAAAGGTCCCCGCCTCTCTCGTATCAGGTGCGGCTATTATCTCTT--ATCTAGCCGTACCGCACACT : 1115
US20050109 : G-ACTTTTGCTTGGTTGCGTGGGATTTCCGAGAAAAAAGGTACTACAGGTGTTTTA--AACTGGCAACCTTTGGAGCG : 582
US6774282 : ----- : -
US20060143 : TGTAAATGATACTATTACACTTTAGAAAAATGTGCTTATTCTATCGTACCAAACTT--CTCTTTTTTACCAGAACTTG : 368
US5689056 : T-TTGGATGCCA--AGCCTAAGTCAAAAGGTGCAAGAAATACAAAAAATAA---TTAGAGNGTAAAT-AGGACG : 173
US6271437 : CCGTGCCGCGGAGAGGACAATGCCAAGGGAGTCGCCGTCGCAATTAGAAAAAAT--AAAAAATAATACAGATATA : 435
US6271437 : CCGTGCCGCGGAGAGGACAATGCCAAGGGAGTCGCCGTCGCAATTAGAAAAAAT--AAAAAATAATACAGATATA : 377
US5689056 : T-TTCTCTGTATGATTTTCTCTATCAAGGTGGCGTGAGAAATCAGGAATCCAC--TCCTTTACATGTGTTACG : 437
US5689056 : T-TTCTCTGTATGATTTTCTCTCTGAAGGTGGCGTGAGAAATCCGTAATCTAT--TCCTCTCATGTGGTTACCC : 405
US5750386 : ----- : -
US5459252 : -----CCCATATGTCCTACCAATGTGAATTTGA--ATTAGTTTGGTCATACG--G : 47
US5837876 : TACTTATTGACTTGTAAACAGTTTTTATAATTCGAAGGCCATGAAAAATTTAATGCTT--TATTAGTTTAACTTA--C : 1109

```

```

US5837876- : TACTTATTGACTTGTAACAGTTTTATAATCCAAGGCCCATGAAAAATTTAATGCTT--TATTAGTTTAACTTA--C : 1109
US5837876- : TACTTATTGACTTGTAACAGTTTTATAATCCAAGGCCCATGAAAAATTTAATGCTT--TATTAGTTTAACTTA--C : 480
US5837876- : TACTTATTGACTTGTAACAGTTTTATAATCCAAGGCCCATGAAAAATTTAATGCTT--TATTAGTTTAACTTA--C : 558
US5837876- : TACTTATTGACTTGTAACAGTTTTATAATCCAAGGCCCATGAAAAATTTAATGCTT--TATTAGTTTAACTTA--C : 216
US5837876- : TAAACAAAGGGCAGTATGGTCAATAACCATAAAAAATATATGACAGCTACAGTTGGT--AGCATGTGCTCAGCT-A--T : 492
US5837876- : ----- : -
US5837876- : ----- : -
US7041873- : TCCACTCG-AGCGGCTATAAA-----TACGTACCTACGC-ACCCTGCGCTACCATCC-----CTAGAG----- : 445
US7041873- : TCCACTCG-ACTAGAGATAATGAGCATTGCATGTCTAAGTTATAAAAAATTAACCAT-----TTTGTGCACACTTG--T : 466
US7041873- : CCAGCTCCTACTAGAGATAATGAGCATTGCATGTCTAAGTTATAAAAAATTAACCAT-----ATTTTGTGCACACTTG--T : 1096
US5633363- : CCTATAACTCTAAAGGCTTGTAACACCCCTCATTCATCCCTCTATATACGAAGACTCTC--TCCTATCATTTTATCGATT : 1116
US20050109 : ACAAGGACCACTGAGGAGTGTGACTCTCTGATATCTCGACAAAACGTCGTCATGTTA--ATAATCCTTCTTTACTTTGA : 1116
US7193135 : CTGACACTCGACAAAGAGCTCCATTGCGCATCCAGCAGCACTTGGCAAAGTGACTG--GCAAAGGGGCCCACTATTGC : 270
US20060230 : ATATCTTACATTAGATTGTATATTTTAAAGCAAAGAGAATAATATAGATATAAGT--ATGTACCTACTCGCTCGA-- : 502
US20060143 : ACCTTGATATATTAATTTCTTGATGATGACAAAAAAGCATAGAAAGTAACGTGTG--AACATAGCATTAAACAAATA : 293
US20060143 : CTTCACTTTAAAGACATATAATTATAAATAACATAATTCATATGGTCTTTATTTTGCCA--AAAAAGTGATAACCTGTCA : 247
US5689056 : GTATGAAATATTAGTACCATTAAATATTTTCGTATATACCGAATTATTTATGAAGGCATG--TATCGACATGTCCATTATTC : 497
US56395963 : GTTTCCTCACTTCTTTTAGCTCCTTCTTGCTCTATTGCTTTTCCAAATTCATGTTCCA--CTCCAGATCCCTAGAAGT-- : 1117
US20062600 : TATGTATCTGTATTAGGATATATTAATCGGAAAGATAAATAGCATGTGAACATTA--TAGTTCCTTAAAAAAG : 424
US20060143 : ACACAGTTGATGACTAATTTGGTGGAGAACCATTAATAACCACTCTGGTCTCTGTTATCATCTGAAG--ATTCTTGTTTTACCTTCCA : 100
US20060143 : -----AACTACAAGGGAGACATAATATCACCATCTGGTCTCTGTTATCATCTGAAG--ATTCTTGTTTTACCTTCCA : 71

```

```

*          1140          *          1160          *          1180          *          1200
CaPEROX : CTCGCCAATAGTTGACTTGCCCTTTTAACTTGGCCCTAACTCCTAATACGACAACTTGGTCTACTTCTAATTTTTC : 1197
US7119254 : CCAATGCTTTTATGCTTAATGCC--TCATTAGAGATATAAACTGAGATTGACTGTGAATCATATAAATAAGGCACACAAA : 694
US20070050 : TTGAGTTTTTCATGCCTTAACTTGTGAATCGTGATAGAGTAAGAAGATTATGGATATTGTTAAGTGTAAATTAAGATCAA : 615
US6649748 : TGGAGGTTTCAACTTCCAGTCGAAGCTTAGAGAGTTCATTTGGAGACCTTGACCAGTTAACTAGCGGTGTCGAAAAACCTGCAC : 534
US6927320 : TATTAACCAACATCGTAATTATATACGGATTTTTCGGAATTTTACGCCATATCTGTAAGTATATATAACATGCATGTC : 1198
US20042483 : GAAAGTGTGTCTGAATGTATAGGCATGTATTGCGTCGATAGTAAATGGCGCAATACATTGTACTTAAGTAATTTGACACCA : 910
US6906241 : TTTAATAAATATATGGGAAAGAAAGTAATTAAGAAAGCTATTAATCTTCAACACCACAAGTCAGATAAAGATGTCAT : 640
US5689056 : -----GAATTCGAAGCAACTTTCCTGACTTGTGTTGGTCATGCTTATAGGTTCCACCCAGACTTCGCA : 61
US6703541 : ACAAGCTGGCTCATTAAACAAACCGAGCCAGATGCTACGTCAGCTCGTGATAAAATTCAAACGAACCTGATCAGATATAAG : 604
US6262344 : TATTCCTATAATGTATTTTGGCATGAAACCTTGACTTGTATTTTTCCTTCCAATACAAAAATCTAAAATTGGCAAGA : 1200
US20050109 : ---ACGATATGAACCTGGTATGTGTTTGTGGTTAGTTGCGTTAGCAGCTGATGCTATGTCCAGGTTTTCCTTTAATTAGT : 77
US20060143 : ACTACTCATTATGCTTCATGGTTTCCATTACCTCTCTGCCAGTGAATCTGACTTAAAGCTATTAACTCTCGTTTCATTAAT : 331
US6384207 : CGAAATCATTGATAAATATTTCTCAGCTGATTAGAAAGAGCAGTGTGTTTAACTTGGAAAGTGATGTGGAAGCGTGT : 1200
US20050109 : ATTTTGGAGTCTAAATTTGAAGTTGATATGCCCTTCGGAACGAATGACCCAAAATTTGGTGTCTGAAATGTTCTTTGA : 302
US20040199 : GAACCTAACGAATATATATATATATCCACTGAAAAGTCATTTTGAATATATGCATTCCACGGTAGCGATGACACGGCGT : 219
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : ----- : -
US20031675 : ATTTTATTTAACTTTTAAATTTAGATTAGTTAAGATTTTTCCTTGATATATTGA--AAGATTAAAGAAAAATACT : 1190
US7196247 : ATTTTATTTAACTTTTAAATTTGA--TTTAGATTAGTTAAGATTTTTCCTTGATATATTGA--AAGATTAAAGAAAAATACT : 1127
US20031675 : ATTTTATTTAACTTTTAAATTTAGATTAGTTAAGATTTTTCCTTGATATATTGA--AAGATTAAAGAAAAATACT : 331
US7193136 : TATTTATCTATTTTCAACCATCATTAATCATTCGGTAATTCCTATAAGTAATGAATTAA--TAGA--GAATAAGAAATGTT : 369
US7193136 : TATTTATCTATTTTCAACCATCATTAATCATTCGGTAATTAATTAAGTAATGAATTAA--TAGA--GAATAAGAAATGTT : 96
US7193136 : ----- : -
US7193136 : ----- : -
US20050976 : GTACGATGTACTTTTTTATCCA--TAAAAACCGTACCATTATGATATGGATATCTTGA--TGAGAGGGGACTCATTATCT : 1169
US20050976 : GTACGATGTACTTTTTTATCCA--TAAAAACCGTACCATTATGATATGGATATCTTGA--TGAGAGGGGACTCATTATCT : 1176
US20060107 : ATATGGTGTGTATATATATATA--TATATATATATATATATATATATATATATATATATATATATATATATATATAT : 884
US20060107 : ----- : -
US20062600 : TTGAAAGAAATGTATAACATATAGTTGAATATTAGTTTGCACAGATACAAGGGGTATTAATATATTGAGCATATCCTCA : 814
US20062600 : TCGAAACCGTGTTTATGGCAAA--TGACCAATACATCTATTTAAATGCTTAGAATATT--CATTTTGGCTTTTTTTTCG : 902
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : GACGCAGCACAACTGCTGGCTTCGGTCTTTTATTACTCCCTCTGTTTCTTTTAAATTTGTCGCTGGATTATGTAA : 1195
US20050109 : TCGAGGCGACGTGCATTTGCGTAGTTGAAGCTTACAAAGTTGCATATTCGATATGAGATGATTGCCGACATGAAGCGG : 662
US6774282 : -----CGACGGGCATTTGCGTAGTTGAAGCTTACAAAGT-----TGATATGAGATGAGTGCCGACATGAAGCGG : 66
US20060143 : T-----TACTGTTGTTTGTGAACGATTTTGT--TATAAAACATTAAGTCTGTTGATATAATTTAACAC--ATT-- : 437
US5689056 : T-----TA--GTTTATTTAAAGTGTGTTTAA--AATTTCAATTATAATCTACAAAGATACTTCTGCATT--ATA-- : 238
US6271437 : GAGAAACAACGACAAAGATTAGTTTGTTCCTAATAAAATACAAAGAACTTGTGTTGTTATCTTATGAGTA--ATTATC : 514
US6271437 : GAGAAACAACGACAAAGATTAGTTTGTTCCTAATAAAATACAAAGAACTTGTGTTGTTATCTTATGAGTA--ATTATC : 456
US5689056 : TTTCTGAATTTGGGTGCTATTGTTTCGTTAATGCTCTGTCATTATCTTCTTGGTTTCTTTGGGATTGGGTTGTTTCAG : 517
US5689056 : TTTCTGAATTTGTTAGCTATGGTGTCAATTGATGCTTCCTGTTTATATATCTATTGGGTTTCTTTGGGATTGGGTTGTTTCAG : 485
US5750386 : ----- : -
US5459252 : TATATCATATGATTATAAATAAAAA--AAATTAGCAAAAGATAAATTTATTAAT--ATTTTACACCATACCAACACCA : 124
US5837876 : TATATAAATTTTTCATATGTAATAATTTAATCGGTATAGTTTCGATATTTTTCATTTATTTTATAAATAAAAAACTTA : 1189
US5837876 : TATATAAATTTTTCATATGTAATAATTTAATCGGTATAGTTTCGATATTTTTCATTTATTTTATAAATAAAAAACTTA : 1189
US5837876 : TATATAAATTTTTCATATGTAATAATTTAATCGGTATAGTTTCGATATTTTTCATTTATTTTATAAATAAAAAACTTA : 560
US5837876 : TATATAAATTTTTCATATGTAATAATTTAATCGGTATAGTTTCGATATTTTTCATTTATTTTATAAATAAAAAACTTA : 638
US5837876 : TATATAAATTTTTCATATGTAATAATTTAATCGGTATAGTTTCGATATTTTTCATTTATTTTATAAATAAAAAACTTA : 296
US5837876 : TGAACAAATCTAAGAAGGTACATCTGTAACCGGAACCACTTAAATGACTAAATACCCTCATCAGAAGCAGATGGA : 572
US5837876 : ----- : -
US5837876 : ----- : -
US7041873 : ----- : -
US7041873 : TTGAAGTGCAGTTTATCTATCTTTATACATATATTTAACT---CTACGAATAATATAATCT--ATAGTACTACAATAATA : 542
US7041873 : TTGAAGTGCAGTTTATCTATCTTTATACATATATTTAACTTTTCTACGAATAATATAATCT--ATAGTACTACAATAATA : 1175
US5633363 : TATTTTTTTTATATTTAGACAATGGAATTAATAGAACTAAAAATATATAAGATGATATCTGAGGACCCGAGATGGTAA : 1196

```


		*	1220	*	1240	*	1260	*	1280	
CaPEROX	:		GAGAGTTAAAGTTTACAAACTGCTTTAAAGAGATGGAGACTAGGATCAATATAGTTGTTGAGGTTTCATTCTACCCA	:						1277
US7119254	:		GAACAGAACAGGAAATACTTAGCAATAGGTTTCCAATAAAAGTGAAGAAGAATACAAATAAACTTTATAAAAAAAGGT	:						774
US20070050	:		TTTCACTCTTACGAGATGGTATTCTGCTACGTTTCCAATCAAATAACTTCATTTTACCATTGTTATTATTATAAGAATA	:						695
US6649748	:		AACTTGAGATTAAATTGCATACCTTTTATATATGACCGGTTTATTTTTCCTGAGAAAAATTTGGAAGAAAAATAA	:						614
US6927320	:		GTTTTCAAATTCATATGATGAACGATCCAGTAAGTGCTACTACTCTCAACATTTGCATGAGAGAGATTATTTATA	:						1278
US20042483	:		GGGTTGCTCCAACCTGACATGACAAATACATTTTTCAGCTGGTGACATGCTGCCGAAAGCAGTCAATACTGACAGTTAACA	:						990
US6906241-	:		CTGCAGTAGAGGGTAAATATCACATCGCATTTGGTACTCCAAAAAGAGGAGGAAAGCAAGGTGATCGTTTTGGCCCTC	:						720
US5689056-	:		GCTCATTTGTAATGGAAGACAACCTTTGTGAACATGTCATGATAGTTTCTCCCTCCCTCATTTTGAAGTTCCTATA	:						141
US6703541	:		CCACGAGTAAATTTGTGAGCCCTAGCTCATGATGTTATCCAGACACCTAACGTATATTGTATCACATGTGGGCTAGTGTG	:						684
US6262344	:		ACGACTTACTACCATGCAAGTGAATTTGTGAAGTTTGTAGTGGTGGTAATTTTAATTGTTTCACCACAGAAAAATTTCTCTA	:						1280
US20050109	:		CATGTTACTCATGCATAGTTTCTCGGTCTAAATTCGGTTGTAACGCGGCTATATATTTTCCTTAATTTTCATTGCTAAGT	:						157
US20060143	:		GTTTCATGATTGAAATTTGAATCACTATTGGTGAATTATGTTTAGAAATTTGAATCAGACCTTGTTTTCTCTAACATCTAA	:						411
US6384207	:		GATTGCGGTTGAGCTTGTATAGGAGTAAAAATGAGGAACAGTAGGAAAAATAATTTTTCGGATTAAAAACCGGTTGTTTGGGA	:						1280
US20050109	:		ATGCTGTTCCCTAAAAAACCGTTTCGATGCTAAACAAAGCATAGGAAATATATAGTTTTCGTTGGTATACCTGCACCATG	:						382
US20040199	:		TGGTGAATGTAGAAGCCTTCGTAAGCGATTTCAGGCGCTCTTTGACATTTGTGTATCTCTCTAGATGGGTTAAACAACAG	:						299
US20040199	:		-----	:						-
US20040199	:		-----	:						-
US20031675	:		-----	:						-
US20031675	:		TTATTTACATGTATAAGCTTAAGATTAACACCTAATTATATATAAGATAGAATTATCTGTAACTATTATATTATACCA	:						1270
US7196247	:		TTA--TACATGTATAAGCTTAAGATTAACACCTAATTATATATAAGATAGAATTATCTGTTA--TATTATATTATACCA	:						1203
US20031675	:		TTATTTACATGTATAAGCTTAAGATTAACACCTAATTATATATAAGATAGAATTATCTGTAACTATTATATTATACCA	:						411
US7193136	:		TTAATAAAATTTATTTCAA-TAATT-ATTATATTATTTTAAATATGTGTAATAAACTCT--TAAATATCAATCATTT--T	:						443
US7193136-	:		TTAATAAAATTTATTTCAAATAAATTTATTTATTTATTTTAAATATGTGTAATAAACTCT--TAAATATCAATCATGAAAT	:						174
US7193136-	:		-----	:						-
US7193136-	:		-----	:						-
US20050976	:		CTCTCTATATATATAAACACCTATATATCAAAACAGGCATCAAGAAAAATAGATGATTTTTTCTTCTGAAGTAGAGTGACA	:						1249
US20050976	:		CTCTCTATATATATAAACACCTATATATCAAAACAGGCATCAAGAAAAATAGATGATTTTTTCTGGAAGTAGAGTGACA	:						1256
US20060107	:		ATATATACCGTTTCATAAATATATGACATCGCTGACTTTTTAAAAAATCTTAAATCATTTGCTCTATTTTAA--AAAAATAG	:						962
US20060107	:		-----GACATCGCTGACTTTTTAAAAAATCTTAAATCATTTGCTCTATTTTAA--AAAAATAG	:						55
US20062600	:		AGTGAATCAATGTCAATCTGAAATATCGTTTAAATTTCT--TAAACGGATGTCCTA--TATTTTCTAGTTGGTTAT-	:						886
US20062600	:		AATTCCTGTTTCGCCACATATGTAGGTTCCGCAGATTTCCATGCTAGTGGTGTCTAGAGTTAGTTCCTTTTATACC	:						982
US20060143	:		-----TGACTTTGTAAGGACAATTACCAAAAGAGATTTAAAA--GGACACAACAACG	:						51
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US20010047	:		-----	:						-
US6518483	:		ATTGCACCTATCCAGCGACAAATAAAAAAGAAATGGAGGGAGTACAATAAATAGCGACCTTCGACAAA-AAAAAACAAAAA	:						1274
US20050109	:		ATAACGTTTTAAACCTGGCAACCAATATCTTGCTGTTTCAAATTCAGGCGTGGGAAGTACGCG--GCCCTGGACGGCG	:						736
US6774282	:		ATAACGTTTTAAACCTGGCAACCAATATCTTGCTGTTTCAAATTCAGGCGTGGGAAGTACGCGTACGCGCTGGACGGCG	:						146
US20060143	:		---TTTGCATAAAAAATGTATCAAT--CATATACA-AATGAACATTACATTTTTAG-TCATTTTTAGTCAATTAAGTAGA	:						509

		1300	1320	1340	1360	
CaPEROX	:	TATGGAACATGTAAAGGGTGACATATGCCCTTCTCACCAGCATTTCGTTGACAATTATAGCTATATAATTTTCAGTA	:	1357		
US7119254	:	ATAATTTACACTCGAATCAACCAATGTAAGATGTCTTGTCCATTACACATCACAATAGTGGATTACAGATTGCAATT	:	854		
US20070050	:	TGTATATATATAAGAAAGTGAGTTTCATATTTTCATTTGTATGTATATAATGTAAAGATTATACAATAAGACAATAATA	:	775		
US6649748	:	GAATATGTATTCTGTAAAGCTAGGCCAAACGAATGTCTTTTCGTCGTTTTCGTTAAAGGTTAGATCATATTTTCATCT	:	694		
US6927320	:	AATTTTA-TTTTGAAGAAGAAATAAGAGGGAAGGTTACTTGGGTGGATCGATGTGAAAACAAAGAAAGAAAGCGCAAC	:	1357		
US20042483	:	CTACATA-CGCTAACATAGTCTGTCCAAAAGTGATTGTGCCACCACAGCTGTAGCAACCCATTGCGAAAACTTTTGACA	:	1069		
US6906241-	:	CCATTTA-TAGGGATTTCGAGACGGCCAAATGTGAAT-TTACAAGAGTTAATTTCCCTCGAATGCCATGATAGCCTTATAGA	:	798		
US5689056-	:	-TCGTGA-GGTGAGCATGTCAATCTTGATCTTTGACTTGTTCAGTTTCTTCATGTGTAGTCAACAAAGCAATCCAGAT	:	219		
US6703541	:	CGTGTGGCTACCTGTGTACATGCATGCATGACGAAATGGTTATTGCCGGCTTTCAGAGTCTAATTAATAAATTAGCA	:	764		
US6262344	:	TATCCTGAAGAAGATAGCTGAGTTGAACTGAGAGGTTGGCGTTTCTTAGTGAAAAACAAAAATAGAAATCTTTAGCTA	:	1360		
US20050109	:	GAATCGGGCGTGAGCTCCTAGCTCTTTTTCGAAGGTTCTTCCAAAGGAAGATACATAGCACATGTAGCAATGGATCCCAA	:	237		
US20060143	:	GGCTACTTTGTTCCTTTTGTATACATTTATCATCTTAAATTTGGTAATCTCGTTACTTTAGCACTTCTGCCTACAAA	:	491		
US6384207	:	CTGCGGCAGATACAATTATAGAGATAAAAAACCCGTAGAAGTATTAGAAGCCGATAAAGATTAAACCCAAATGAACGAA	:	1360		
US20050109	:	CATGCGTCCGCGCTATTCCTCTGAGGCGCTTGGTCTTGTGCGTACATGCCTATGGATGATCAGGACGCCCATTT	:	462		
US20040199	:	GACTGAC-AAAACCGCGGTGGTTAGGGAAGTTGGAACCATATTCATTGAAGGATCAGCCTGTAATAAATAAATAAACA	:	378		
US20040199	:	-----	:	-		
US20031675	:	-----	:	-		
US20031675	:	TATCAGT-TGTCACAACATGTGTTTCTGGGGTTATTGCCTCTTGAGTTCAATTGCAACTTGTTAAGCAAAATCGGCCGGCT	:	1349		
US7196247	:	TATCAGT-TGTCACAACATGTGTTTCTGGGGTTATTGCCTCTT-TTCAATTGCAACTTGTTAAGCAAAATCGGCCGGCT	:	1280		
US20031675	:	TATCAGT-TGTCACAACATGTGTTTCTGGGGTTATTGCCTCTTGAGTTCAATTGCAACTTGTTAAGCAAAATCGGCCGGCT	:	490		
US7193136	:	GGGAAA-TGGGGGAGTATATGTAGATTGTGAATCAATT-----GTTAAATTACAACCTTCAGGAGTGTGAGTCCATT	:	517		
US7193136-	:	GGGAAA-TGGGGGAGTATATGTAGATTGTGAATCAATT-----GTTAAATTACAACCTTTTGGGTAGGTGAGTCCATT	:	248		
US7193136-	:	-----	:	-		
US7193136-	:	-----	:	-		
US20050976	:	GAAGCAG-CTGAAGTGCGAGTCTTTTGTGTTCAATT-TTATAATGTGTAAGAAAAATGACGCC-AATGAAATATGTGTCT	:	1326		
US20050976	:	GAAGCAG-CTGAAGTGCGAGTCTTTTGTGTTCAATT-TTATAATGTGTAAGAAAAATGACGCC-AATGAAATATGTGTCT	:	1333		
US20060107	:	-----AG-TTGTCAATTA-----TTTTTGTGTGTTTGTGTTTATCACTTAAGGTAGTTTGTGCTTAATTAATAATTTATAC	:	1033		
US20060107	:	-----AG-TTGTCAATTA-----TTTTTGTGTGTTTGTGTTTATCACTTAAGGTAGTTTGTGCTTAATTAATAATTTATAC	:	126		
US20062600	:	-GTCTAT-GTATTAAGAATATATTACTTAAACATATAA-ATTAATT--AGAGTCCAAC-TTAAAAATTTATTAAC-GT	:	957		
US20062600	:	-ACTTTT-GTACCGACTTTAT-TTACGGATTAATTATCATATAAACCCACATAATATAATGTCATAGTTTGTGTTGTCAT	:	1059		
US20060143	:	AAGATTT-GGTCAACTTTAGGTTCAATCATATGCCACCAACAAAAATCTTTCACATTTTGT--GAAATTTGAAAA	:	128		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US20010047	:	-----	:	-		
US6518483	:	CAAAAA-GAACTCCCATTTATTTCAATTTCAAAGGGTACGGATTTTAAATACATTTCTCCATGTA--TGTAATCAGACA	:	1351		
US20050109	:	TGTAAAG-AGCCAGCATCGGCATCATTTGTCAAACGATCGACAAGGGCAAGAAATCCAAATATATTTTATTAATCCAAAA	:	815		
US6774282	:	TGTAAAG-AGCCAGCATCGGCATCATTTGTCAAACGATCGACAAGGGCAAGAAATCCAAATATATTT--ATTAATAAAAAA	:	223		
US20060143	:	-C-----G-TGTGTA-----TCATGATTAG--TTGATTTAATCATAATTATGATATTTAAATAAACAATAGTTTA-AT	:	572		
US5689056-	:	-C-----A-GGTACA-----TTG-GATT-----TTAGCTTGTGTTGGTGTGCCAGCAACAGCNAAGGAGTAGT---AT	:	368		
US6271437-	:	-CCCAAA-GGAATC-----TTAGAATC-----TTAATATCTTGTGTTTGTGATTAAACAAATAAATGTGTGGATT---T	:	657		
US6271437-	:	-CCCAAA-GGAATC-----TTAGAATC-----TTAATATCTTGTGTTTGTGATTAAACAAATAAATGTGTGGATT---T	:	599		
US5689056-	:	GTGTTG-TGGGCC-----TGCAACTACTCTTGTGCTGTCTGCTGTCACCACTGCTCGACAAATGCCCAAT	:	657		
US5689056-	:	GCGCGC-GGACCATGTGCTGCTGCAACTACTCTGGCTGCGTTGTGCTCCACCACCTGTTGCAAAAAATGCCCAAT	:	643		
US5750386	:	-----	:	-		
US5459252-	:	ATAGTTA-TTCTAAACAAGAAAGAAATTTGTTCTTGACATCAGACATCTAGTATTAT-----AACTCTAGTGGAGC	:	274		
US5837876-	:	GACGGTT-CGATCGGTTTAGTCGATTTTCAATATTCATTGACACTCCTAGTTGTTGTTATAGGTAAGGAGCTTACAG	:	1347		
US5837876-	:	GACGGTT-CGATCGGTTTAGTCGATTTTCAATATTCATTGACACTCCTAGTTGTTGTTATAGGTAAGGAGCTTACAG	:	1347		
US5837876-	:	GACGGTT-CGATCGGTTTAGTCGATTTTCAATATTCATTGACACTCCTAGTTGTTGTTATAGGTAAGGAGCTTACAG	:	718		
US5837876-	:	GACGGTT-CGATCGGTTTAGTCGATTTTCAATATTCATTGACACTCCTAGTTGTTGTTATAGGTAAGGAGCTTACAG	:	796		
US5837876-	:	GACGGTT-CGATCGGTTTAGTCGATTTTCAATATTCATTGACACTCCTAGTTGTTGTTATAGGTAAGGAGCTTACAG	:	454		
US5837876-	:	TGTAAGC-ACCTCAGCCATGTTAATGGAGTCTATTGCCTGTTAACTCTCACTTATA-----AAATAGTAGTAGAAU	:	723		
US5837876-	:	-----	:	-		
US5837876-	:	-----	:	-		
US7041873-	:	TTTATCT-TTTTAGTGTGCATGTGTTCTCCTTTTTTTTGCAAATAGCTTCACCTATA--CTTCATCCATTTTATTAGT	:	693		
US7041873-	:	TTTATCT-TTTTAGTGTGCATGTGTTCTCCTTTTTTTTGCAAATAGCTTCACCTATATTACTTCATCCATTTTATTAGT	:	1334		
US5633363-	:	AATTTGG-TCCCCATGTGCCGTCGCCGCAACTTCTCTGATCTAAATTAGTCTATTTCCATGTTAAACTATACTAAAA	:	1355		
US20050109	:	ATAAGAA-GCATTATTTAATATAATTACCTTCTTTTGGACATCTTAGTCTTGTACTAACTATTTTGAACCTTGTGC	:	1355		
US7193135	:	AGGACAC-TCGACATATGTCACCATCACTTGTGCTGTGCGGCAATTTTCTTGTCCGAGCACTAGGGGAGGTAAGAC	:	509		
US20060230	:	CATTTGG-GATGCCGCGCGGATTCTGATTTCAGCTCTCTGGCTTCTTTTAAACAAAAACGAAAAAGCAGTGATCC	:	741		
US20060143	:	ACTCT-G-TTTTATAACGATTGATTTTAAATATTAGTTTTTGTGTAAGAGATCATTTATATATGTCTGTAATATTTT	:	531		
US20060143	:	CAATTAT-TATTTAATTATGTTTGATTAAAAAATTATATTATAAATTTAAATATCAAAATTTATTCAAATTTTAGTA	:	485		
US5689056-	:	TATTATT-TTTATAATACA-----ATCCGATATAATTACAATACGATATTACCGAATATTATCTAAATCAAAATTTAATT	:	732		
US6395963	:	TTTTGTG-TTCTAATATTAAATCTATATCATTTAATCTTATAAGTTTTTTTGGGTTAATTTAAGTGAATTTAATATT	:	1356		
US20062600	:	ATTTTGA-ATCGTTGAACAACATAATTGTTTACTTAAAGTCTTAAGGTGTTTAAATCTCTTAAAGTCAACATCCCGAAAT	:	663		
US20060143	:	AATTGAA-TGTGGCGT---TGCTTCTCTGCTCTAATATAAAATTTGGATAATCATGAATTTCTCGATGATAGTGATA	:	335		
US20060143	:	ATCTGTT-CTTATGTCFAA-ATTTTGACTTCATTCACTTTTCTTCTGCTTTTAAAGAAAGCTTCCACAATCT---AAC	:	306		
		1380	1400	1420	1440	
CaPEROX	:	TCACACCAAACCTTTG--GCTTGCTTTTGCAATTTCTCATCTTGCAAC-ATGTATTATTAATAAATAAAAAAATTTACA	:	1434		
US7119254	:	GATGAAATCTGGATCT----TAGCTAAAAATTTATCTATATACATCG-AGTTTAAAGATGTTATATCAACACCAACA	:	929		
US20070050	:	AAGAGATTAAAGGAATC----ATTTACGTGTTAATCACAATGACTTGAG-AACTCTTTTAAATGAAGTCCCTCTATAG	:	850		
US6649748	:	GGTCCAACACTCAAACT--TGATAAATGGACGAATTTATGTCATTTTA-GACCTACCGGCTAGCGGACTTTTTGTGTT	:	771		
US6927320	:	CCACTAAGCCATTACAT--GATATCGACCTTCTTATCTTTTCTCTTT-ATTTATTTTCTCAGGACTTTTTCTACT	:	1434		
US20042483	:	CAACAACTCAATGTCA----CATGTGAGTTTCCCATGGAATGCGCCCT-GGCCTTTAGC	:			

GCAATGTAATAAAAAACAA : 1144

US6906241- : AGTTAGTAGTTGTGAA---GAAAGCTCAAAGATATTTGGTGAGATGCA-AGTGGAAATATCGGATTTAACATTTAAGCG : 873

US5689056- : TTCTTTAGCAGACTCA---CAGGCTGACACTCTATTGTACTCATCAGG-TCCTATCCCACAGACCATAAGAGTTTACC : 294

US6703541 : TGGACGTACTAATAAT---TCATAAGCTTGACGTCCGGTTTGATATA-TGTATTTTCTGATCCATAGCTAGCTAGTA : 839

US6262344 : GAAAGTGTGGTGTGGA---CCGAGCTGATGGTAACCATGTTTCATTGG-AGGAACTAATGTGAATATTAGCTAAAAGCA : 1435

US20050109 : GAGATCTTTTGACACT---CATCTCAATATCTCATAACTATTTTCAGAA-CTGTGCGTGTGACCCATCGACAGGTGTCC : 312

US20060143 : TAACTACTCATTATGC---TGAATTTGTTTGTGTTCCCTTACCAAAACC-AATTAAGATACTGTTTATAGTAAAGAAAA : 568

US6384207 : CAGGCTAAACAAATCCG---GCGCCTCAAAGTCAAGAGCAGGTACTGGGCTGTCTGCACACGTGCGCTTTTGTCTCCCC : 1438

US20050109 : CATTCTTAATCACTACTTAATTGACATCTAATGCCCTCTGCTCAACTTGGCAGCTACGCACCGTATGATACGCCAAGA : 542

US20040199 : CAAATATCCATTTC---TGCTTATCTTGTATCTGGTCGTATCTA-CTCTTTGTCAATTAATCGTCTCTGTAAGTGG : 453

US20040199 : ----- : -

US20040199 : ----- : -

US20031675 : ----- : -

US20031675 : TAAGACCTAAGCAACA---CAAGCAAGGCTTTAGGTTTTCAAAAAA-GGGTTCCAATTTTTTTAATATTATATAT : 1424

US7196247 : TAAGACCTAAGCAACA---CAAGCAAG--CTTTAGGTTTTCAAAAAA-GGGTTCCAATTTTTTTAATATTATATAT : 1353

US20031675 : TAAGACCTAAGCAACA---CAAGCAAGGCTTTAGGTTTTCAAAAAA-GGGTTCCAATTTTTTTAATATTATATATAT : 565

US7193136 : TGAATAAATAATTTATA---CTTTTAAT--TATGATTTTATACCTCGA-AAATGGAA---CTCAAACCTTTAAGTAG : 584

US7193136 : TGAATAAATAATTTATA---CTTTTAAT--TATGATTTTCCAAATCGA-AAATGGAA---CTCAAACCTTTAAGTAG : 315

US7193136 : -----ATTTTATACCCAAATCGA-AAATGGAA---CTCAAACCTTTAAGTAG : 43

US7193136 : ----- : -

US20050976 : GGGCTGACGTGTTGTT---TGGTGAAAGCCAATTTTGTGTATATAGG-GGGGCCAGAG---CCAGTTGTATTTGTTG : 1398

US20050976 : GGGCTGACGTGTTGTT---TGGTGAAAGCCAATTTTGTGTATATAGG-GGGGCCAGAG---CCAGTTGTATTTGTTG : 1405

US20060107 : TTTTGAATAAGATAAA---TGGTCAAAGT---TTTTTAAAAAATCAA-CATGTCA-----TATATCTGTGA : 1093

US20060107 : TTTTGAATAAGATAAA---TGGTCAAAGT---TTTTTAAAAAATCAA-CATGTCA-----TATATCTGTGA : 186

US20062600 : TACAAAAATAGTGCT---TGTGAAGCGTTGACAT-----TGGTGAA-TGCTTATGATTAAT-TAGGCATATACCCCA : 1025

US20062600 : TGCTCGATATGGCGCA---TGCACAA-GTTGATATACTATTGACGAG-TGTT--TGGTTGGTACTGTACGTTTGTGA : 1131

US20060143 : GCCAATATAAACATTT---CCTGTTTTA-AATTTTTTTATGAAATAAA-ACTCATAATG-----TTAACGAAATGAAT : 196

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US20010047 : ----- : -

US6518483 : TCATGTATATCTATGT---ATATAGAAA-AACTAAAATAGCTTATGTA-ATTTAGAATG-----AAAAGAGTATAACC : 1419

US20050109 : GAAGGCACAAATAGT---TTGGTTTT-AGTATGTGTGGCGGAGGAA-ATCTTGAGGA-----CGAACGTATCAAAG : 883

US6774282 : GAAGGCACAAATAGT---TTGGTTTT-AGTATGTGTGGCGGAGGAA-ATTTTGAGAA-----CGAACGTATCAAAG : 292

US20060143 : GTTATTATGTTGGTTT---TTAATATTT--TTAATTAAC--GTCAA-CCT--ATTT--TTTTTTATTTTTTTT : 635

US5689056 : TTTAATTAATAAGTAA---TCAATTTTGGG-TGAGAAAT---GCTGA-GAT--ACCT--CCTTTGTCTTTTTTTA : 433

US6271437 : GTTGAAAAATATGTTT---CCCAAATTTGC-TTGGCTCCCTAAGTACA-GCTGGAATTT---CTAACTAATATTTTCTA : 728

US6271437 : GTTGAAAAATATGTTT---CCCAAATTTGC-TTGGCTCCCTAAGTACA-GCTGGAATTT---CTAACTAATATTTTCTA : 670

US5689056 : GGCACACCTCAACCT---TCTA-----TGCTATGTTGAGAGAA-ACCTGCACCG--TTGATAA-CATCAGCTT : 719

US5689056 : GGTTCACACGCAACCT---GCCAAGGCAGCTTTGTCCCAACGGAGAA-GCCTGCGCCA---ATAATTA-TGCCAGCAT : 714

US5750386 : -----CCCATATAGCT : 11

US5459252 : TTACCTTTTCTTTTCC---TTCTTTTTTCTTCTTAAAAAATATC-ACTTTTTAAATCTTGTATATTAGTTAAGCT : 349

US5837876 : AGAGGTAAAAATAAAC---TTAAAAAATCAGTCTTAAGGAAAAAATTGA-CTTTTATAGTAAATGACTGTTATATAAGGA : 1422

US5837876 : AGAGGTAAAAATAAAC---TTAAAAAATCAGTCTTAAGGAAAAAATTGA-CTTTTATAGTAAATGACTGTTATATAAGGA : 1422

US5837876 : AGAGGTAAAAATAAAC---TTAAAAAATCAGTCTTAAGGAAAAAATTGA-CTTTTATAGTAAATGACTGTTATATAAGGA : 793

US5837876 : AGAGGTAAAAATAAAC---TTAAAAAATCAGTCTTAAGGAAAAAATTGA-CTTTTATAGTAAATGACTGTTATATAAGGA : 871

US5837876 : AGAGGTAAAAATAAAC---TTAAAAAATCAGTCTTAAGGAAAAAATTGA-CTTTTATAGTAAATGACTGTTATATAAGGA : 529

US5837876 : S-AGGTAAAAATAAAC---TTAAAAAATCAGTCTTAAGGAAAAAATTGA-CTTTTATAGTAAATGACTGTTATATAAGGA : 797

US5837876 : -----GTAATGACTGTTATATAAGGA : 22

US5837876 : ----- : -

US7041873 : ----- : -

US7041873 : ACATCCATTTAGGGTT---TAGGGTTAATGGTTTTTCTA---ATTTTT-TTAGTACATCTATTTTATTTCTATTTTAGCC : 765

US7041873 : ACATCCATTTAGGGTT---TAGGGTTAATGGTTTTTATAGTAATTTTT-TTAGTACATCTATTTTATTTCTATTTTAGCC : 1409

US5633363 : ATTTAATACACAGTCT---ATTATAAATAGCAACTAAATTTCTAAAG-TTGATGCATCTTGTAAATTTAATCTCGGTT : 1430

US20050109 : ACGTCATTTATGTTTC---TTTACTTGGAGCTTGCCAAATGTTTCT-GGGCTTTGTGAAGTGCATGAAGCACTCGGTT : 1430

US7193135 : TTTGCCGAGTGTCCGA---CATAATGCTATCGGTAAGAGGTCGTGCTG-TTACAGTTCACTGAGTCCCTTTTACTAAGT : 584

US20060230 : GATCGATCAGCATGAG---CGAGCTAGTAAGCTCCAAACAAATAGA-GTACGTACGTATAATCCTAGAGTCCGGATA : 816

US20060143 : ATACTGAGTTACATGA---TATTTAGTTATTATAGCGTAATTAACATAA-GATAAGAAATTAACATAAGTGATATTCTGA : 606

US20060143 : TAAAAACCCCAATATA---CCGATTGTCGCGGTTTACCAGCGATTAA-ATCCTAGTTATTATTATGCGATTTCCCTTT : 560

US5689056 : TATCATATCAAAATTAT---TAACTGATATTTCAAATTTAATTTA-ATATCTACTTTCAACTATTATACCAATT : 807

US6395963 : TCACATGTAACAAGAA---CTCCATTTTATTCTGTCCGGTTCTTGAA-TAAGATGTGAAAGTATGTATCAATTGATG : 1431

US20062600 : TATGCTTAAAGCATGTT---TAAATTTGTGATTAATGAAATTACTTTGG-TGATTAGCTAGTGAATTTCAACTCTTGATA : 738

US20060143 : TACCGTTGACACTGAG---TAGTTCATTGCAATAAGATGAAAGTCTTGT-TTTGCTAAAAAGATTCAATCATATTTTGAA : 410

US20060143 : TGTTTCGATGTGAAAC---TGAGATTCGAGTAAGAAATGTGAACGTG-TTATACTGTTTTTTAATTAGATAATTTAG : 381

* 1460 * 1480 * 1500 * 1520

CaPEROX : AATTGCTAAAGCAAGTTGTAATATCCCATCAATTTAATCAGTTGGAAGTACAAACTTGTATATAAATATGCCACAC : 1514

US7119254 : ACCACAAGTTTGAAAAAAGAAACAGAGTAATAATATATCAAAATAAAATTTTCATGGCTGATGGAACTCTTTCTAATTGTA : 1009

US20070050 : ACCCGTCGCCGTTATCTTGGCGCCATTTACGTACCCTCCTTTTTTCTAAATCATATCCCAAAATCCCTAACCTAACTA : 930

US6649748 : TCCATAAAGATTTCGATAATGTCATGGCCAGATGCAAGTTTGAAATTTAATGTTTGCCAAATCCTATCATACCCACAAC : 851

US6927320 : TAATGAAACCTCCAAACTATCTAACTAATACACTCCCATGTAGAAATAAGAAAAATTTATATAGATATTGTTGATATTTG : 1514

US20042483 : CCCTCAATAAATAGTTTGAAACAGTCATCTTTTAGTAAATAATTTAGAAAGGACTCCTAG---TATTTTGTCTTAA : 1220

US6906241 : TATACAAATACTGATAGTGAAACTTTTACATCATCACTATAGTTTAAATTTATATAGCAAA---TCACGTACTATTTA : 949

US5689056 : TTTGAAGCCCTTTCTATCTTTTCTGTGAGCATCATCATATTTCTGCTTGGGCTTTGGAA---CAGTGATGGTCTTT : 370

US6703541 : GTGACATGCATGCATGCATGATGAAGGCATATAACATTACGGGCCCTAGCTAGCTAATAAA---TATGCATGTAGTGT : 915

US6262344 : TATTTGTTGAGTGTGACAAAAAGACAACAGATAAATCGTCAAATCTACTCCACCTAGCTAA---CCTATCTCCCTAGT : 388

US20050109 : ACTGATTGCACCTCAGAGGATCCCACTACGCATCAATTAATTTAAAGGCGATCGACTCACTGAA---GATGATGATG : 648

US20060143 : GTGATTGAACAGTCTGACTTTCTTGTAAATTTGTAGTCGGAATCATGCGATGTTTATGGTTCCACCAACTATGCACT : 1517

US6384207 : CTGGCCCTGGGTGAGAGTAGTAGGGATGCTAAAGTTTGTCTCTTTTGGGCACTGTGATAG-GCTCTTGTAGTTG : 621

US20050109 : TCCCAAGCTAAA-ATAACACCAATCGTCATATAAAACAAAGTGTGAGTGCAGGCCAGCCCATGTCATGCGCAGCGATCTT : 528

US20040199 : GAACCGGAACCG-GAATCGTCTCTTGTGTTAGATGAGAAGAGAAATAACAAATCACCATTGG---GATCTAGAAACATC : 19

US20040199 : -----CATGG---GATCTAGAAACATC : 19

```

US20040199 : ----- : -
US20031675 : ----- : -
US20031675 : CTCAAAA--AAA--TTATTGAAAAATTATATTTGAAA--ATAAGTTTAAATTAATAATATATATA--ACTAACCGTTAATC : 1496
US7196247 : CTCAAAA--AA--TATTGAAAAATTATATTTGAAA--ATAAGTTTAAATTAATAATATATATA--ACTAACCGTTA--C : 1421
US20031675 : CTCAAAA--AAA--TTATTGAAAAATTATATTTGAAA--ATAAGTTTAAATTAATAATATATATA--ACTAACCGTTAATC : 637
US7193136 : CCCATAA--GCC--AAATTAGATGTGTATTGGGTTACATTGCCCTTTTCAAAGACATGTGC--TTTAACGTTTGACC : 657
US7193136 : CCCATAA--GCC--AAATTAGATGTGTATTGGGTTACATTGCCCTTTTCAAAGACATGTGC--TTTAACGTTTGACC : 388
US7193136 : CCCATAA--GCC--AAATTAGTTGACTATTTGGGTTACATTGCCCTTTTCAAAGACATGTGC--TTTAACGTTTGACA : 116
US7193136 : ----- : -
US20050976 : CCCG-GA--CTG-GCGCCAAAAAATAATCCGGATAGTACTATTCGCTAACTGTGTACAC--CTTTATCTAAAAAT : 1470
US20050976 : CCCG-GA--CTG-GCGCCAAAAAATAAT--TCCGGATAGTACTATTCGCTAACTGTGTACAC--CTTTATCTAAAAAT : 1475
US20060107 : ACGG-AG--GTT-GTATTACAGAA--TGTGCGACGTAC--ACGCTACCCA--ATAAA--ATACAACAAACATT : 1154
US20060107 : ACGG-AG--GTT-GTATTACAGAA--TGTGCGACGTAC--ACGCTACCCA--ATAAA--ATACAACAAACATT : 247
US20062600 : TTGATGCT-AAA--TTGATTCCTTTTGACTTTGGATGCATT--CTGATTTGGC--TAGGATT--GTTACCTATATATA : 1095
US20062600 : CTTGTGAT-ATA--TTATTGTGAATAGTTTATGAAAAATTT--TATGGTTGACATTCCTCACT--AAAAATTTACTTA : 1203
US20060143 : TATTAAATGATT-GAGACCATAAATGGACCGCATAGTAGTGCCTTCCACACATGTAA--CGTAATCTCTCTCT : 271
US20010047 : -----TGAGATCTG-GA--TCCGTTTCGACAAAA : 25
US20010047 : -----TGAGATCTG-GA--TCCGTTTCGACAAAA : 25
US20010047 : -----TGAGATCTG-GA--TCCGTTTCGCGGAG : 25
US20010047 : -----TGAGATCTG-GA--TCCGTTTCATGTAC : 25
US20010047 : -----TGAGATCTG-GA--TCCGTTTCATGCGGT : 25
US20010047 : -----TGAGATCTG-GA--TCCGTTTCAACAGT : 25
US20010047 : -----TGAGATCTG-GA--TCCGTTTCAGA-ATA : 24
US20010047 : -----TGAGATCTG-GA--TCCGTTTCGG--TA : 22
US6518483 : ATGGTGAGTGTA-TGCTCTTAGAATAGTGCTCGGTCAAATATTTTATTTTTTTTATA-AA--TCAGTTTACATACA : 1493
US20050109 : GAGGCACAAGAT-AATAGTTTGACGCGCGGGTAGAAGTTGCAG-AAGACAGTGGGTACGG--CTTTATCCTAATAA : 957
US6774282 : AAGGCACAAGAC-GATA--TAGATTGACGCGGCTAGAAGTTGCAGCAAGACAGTGGGTACGG--TCTTATA-TATCCT : 364
US20060143 : CATATAACTATG-TTGTATGGAAGTC--ACTATTAAATTAACCTG--ATAAAATATACAA--TAAATAGAAAAAA : 704
US5689056 : TATATATATATA-TATATATATATATA--TATATATATATAT--ATATATTAICS--TCTAAGTCAAAAGAA : 483
US6271437 : CGAACCATTAGT-TACGGAAGAGATCACGGTACTAGTGTGGCTGTGGGTAACAGATGATA--TCTAAGTCAAAAGAA : 803
US6271437 : CGAACCATTAGT-TACGGAAGAGATCACGGTACTAGTGTGGCTGTGGGTAACAGATGATA--TCTAAGTCAAAAGAA : 745
US5689056 : CGTCTGGGGAAG-ACGAAGAGATAAATTAATCCGTGGTGCAGGGGA--AAATACCATCATA--TCCATTGGAATCCA : 791
US5689056 : TATCGGAAGATG-ACGAGGAGATAATACAATCTGTTTTCAGGGTA--AAACACCATTACATA--TTCCTTGGAAATCAA : 786
US5750386 : TATC--TAAACA-AAGTTT--TAAATTCATTCTTAAAC--GTCCATTACAATGTA--ATA--TAACCTAGTCG--T : 75
US5459252 : TATC--TAAACA-AAGTTT--TAAATTCATTCTTAAAC--GTCCATTACAATGTA--ATA--TAACCTAGTCG--T : 413
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 1493
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 1493
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 864
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 942
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 600
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 868
US5837876 : TGTGTGTACAGA-GAGGTA--TGAGTGTAGTTGGTAAATATGTTCTTGACCGGTGATGTCA--CATATTATTTA--T : 93
US5837876 : ----- : -
US7041873 : ----- : -
US7041873 : TCTAAATTAAGA-AAACCT--CTATTTTAGTTTTTTTTATTAATAATTTA-GATATAAAATA--GAATAAAATAAAGT : 837
US7041873 : TCTAAATTAAGA-AACTAAACTATTTTAGTTTTTTTTATTAATAATTTA-GATATAAAATA--GAATAAAATAAAGT : 1483
US5633363 : TGTCAAGTTAT-ATTCATTGATATAATAAATTTGAATTTGACTCTTAATATCGTATTTTT--TCCTAACGGGACG : 1505
US20050109 : AAAACAGCTATC-TCCGGTGGTGAAGAGTTGGGGTAAATGTATAGTATTAGTACACTCTTT--TTCTAAACCAAGCG : 1505
US7193135 : GTCACACTCGAC-AAAGAGTTTGTGAGTTTAGGCTTTGCGGACGCTTGCACACTCGGCA--AAGCAGTATGTTTT : 659
US20060230 : ATAATAATCCGT-TTGGTTCGCGTTAAAAAGTCTTATCTCCTCGTGATCCCTTTTTTTTGGG--TCGATCCATGTTG : 891
US20060143 : TTATTATTATTT-TTGTAGGACACGTACGTGGAAGAACTAAACAC--TATAGGTTACAAAA--CGGTATAATAAACT : 679
US20060143 : CCCCATGCTCTG-AAATTTGTTAATCATTTTGTCCCTATTTCTGCTTAACCACTTACATTTCA--TAATTTTGGAGTTT : 635
US5689056 : ATCAAAATGCAAA-ATGTAT--GAGTTATTTTCAATAGCCAGTTCGTATCCAAATATTTTA--CACTTGACCAGTCA : 880
US6395963 : ATGATGAATGCA-TGTGTGCTAACCATTTACGTTATCTTTTTTACGATAAAATAAACGAAA--GTTGTTTGGTATGT : 1506
US20062600 : TATCCAATAGCAATTGCGCAAAAGTTGTATAAAATTTAGTTGTACAAATATCATTTATCTAT--AATACATTTTTCAT : 814
US20060143 : CACTTGCACCCACCACCTTCCAAAACCTCCACAATTCATAGATATAGTAATACACAGGTTA--CTTGACAAAAGGAA : 486
US20060143 : ATTGCACTCAGA-TAAATTAATAACATTCTCGAATACTTTTATGTGATTGGATATATTAGG--TATATCTGCCAACC : 456

```

```

* 1540 * 1560 * 1580 * 1600
CaPEROX : TGTTCAGAGTTTACGCAATTGATACCGAAATGTTATTAAGGTCAAATATCAATTTTCAACAAGCAATGATAAATAATAA : 1594
US7119254 : GGTCCAAAAAAGCCTAAATTAATGGGGAACA--AAACCAAATTCATAAATTTTACTAAATATGCTCTGGTTAAATA : 1087
US20070050 : AACTAAATACTACCTAACTAACTACCTTATTAAGATTTAAGACATGTTTATCTATCCCCTAATCGTACGTATTACG : 1010
US6649748 : ACATGCTCTCAGGGCCAGTGGGCACCAAGCAACATCTCTGTCATAATTAATTTTTTAAATGAGAGGAGAACTCACAGC : 931
US6927320 : TAACTAGAAAAATATATTGCTCTGTAATTTTT-CGTAAGTTAAATCAACATTTTTTCAAGTAAACAAATATTACTGCAAA : 1593
US20042483 : TTATACTTAAAACTCTATTACTCTTCTTAGA--ATATAGTTTGCAGAAATAACTCAGATCATTTCTTATAAATTAAT : 1298
US6906241 : TTGTAGGTTATGAATTAATATTTATTAATAA--AATTATTTGTAATTATCT--TTTAAATGATTGATTGTATAAATAT : 1025
US5689056 : CTCATCTTTACTTTCATCATTGGACAAGAGTCA--CTAGTATAATATCCCATAACTCGCTATCTTCAGCCATAATCGTGCA : 448
US6703541 : ATAGCATCGCCTTGACACACACCATATCACA--AGCAAGCGCAACCATCGAGCAGAAAGAAGATCGTCGAGATCGAGC : 993
US6262344 : TAATGTTAGAAAGCCACCTATTTGCATCCGTA--ATGATAAAACTAAAAAATATTAGATTATTAGAGTGATACATTTT : 1589
US20050109 : AGGTTAGGCATAACAAAGTGTAGCGCCTCAA--TGCCACCACAAAAAGGATCAAAAGCATGAGAATCCAGTGGTCTG : 466
US20060143 : GTCAGGTGTCACACCACCAATACACTATACTA-GAGACTTAACCATTAGAAATTTGGTCTTCAAATTGCCATATAAAT : 727
US6384207 : CTAGGGCTATGTTTATTAATATTTTGCGCTTTTACCTATGTACGTAAGAACCAGGATGGAATAATGCTATGCAGGAACCAAT : 1597
US20050109 : GGCCATTTCGCGGAGCCAAACCAAGCCGAGCAGCAAAATATGCGATACCGTATGAGGGAACCACTAGTTATTATACG : 701
US20040199 : CATCCATTCGGTTTATTCAAAGTCAGCAC-AC--ACAGTGGGGTCCGGGGGATCGTTGTCTTTAGGACTTATCGAAACC : 96
US20040199 : ----- : -
US20031675 : ----- : -
US20031675 : TTTTATTGGTATTATTAATAATAATCAA-TG--AGCAACAATCTTCCACCGACATCATATCTTTGGTTTTAAAAAATA : 1573
US7196247 : TTTTATTGGTATTATTAATAATAATCAA-TG--AGCAACAATCTTCCACCGACATCATATCTTTGGTTTTAAAAAATA : 1496
US20031675 : TTTTATTGGTATTATTAATAATAATCAA-TG--AGCAACAATCTTCCACCGACATCATATCTTTGGTTTTAAAAAATA : 714
US7193136 : --TCAAACTGCTGCGTATGCTGCAGTAAA-TA--ATATCCAAAAGTTGAATGGT-CAATAATTTGGGCT--GAAAGTGTT : 729
US7193136 : --TCAAACTGCTGCGTATGCTGCAGTAAA-TA--ATATCCAAAAGTTGAATGGT-CAATAATTTGGT--GAAAGTGTT : 460
US7193136 : AACTGCCACATGCGTATGCTGCAGTAAA-TA--ATATCCAAAAGTTGAATGGT-CAATAAGTGGCTTTTGAAAGTGTT : 192
US7193136 : -----A-TAATTTGGGCTGTTTGGAAAGTGTT : 26
US20050976 : A-GTCATCCAAATTAAGAACTAACCTTA-GA--TACAAAAAATAAACAAGATGACAAGT-TAGGT--AGCAAACTA : 1543
US20050976 : A-GTCATCCAAATTAAGAACTAACCTTA-GA--TACAAAAAATAAACAAGATGACAAGT-TAGGT--AGCAAACTA : 1548

```

[illegible]

```

US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : CGATAAAAAATCTGTTCTTTGCAGATGCAAAATTCGCTTAAACCATAATAATGAAGATTGTTTGGGTAATCAATGGTTTT : 1649
US20050109 : GCAGAGCAGCCTTATTAGCTTGCTTATTGTATCGTGGTGGTGTGACTTGTAAATTACTGGCAGCATTATCAACAAT : 1115
US6774282 : GTAGAGCAGCCTTATTAGCTTGCTTATTT--ATCGTGGTGGTG--TACTTGTAAATTACTGGCAGCATTATCAACAAC : 517
US20060143 : GATAGTGACTGAAGTGAAGAGCCAGCCACTATACGACATACGTATGTAATTACGTTTAAATAACCTGATGGCAAAATAT : 862
US5689056 : ----- : -
US6271437 : TACTTTCGGTATAATTTTAAATTCATATTTCAAACCACCGGACTTTCCAATCCCTCGTTCCTCATGCCCCCCAACCCAA : 961
US6271437 : TACTTTCGGTATAATTTTAAATTCATATTTCAAACCACCGGACTTTCCAATCCCTCGTTCCTCATGCCCCCCAACCCAA : 903
US5689056 : TACCATTGGAAGGATTAACTATGAATCTATTCTTGGGCAGTGTGTGAGATGCCAATGGGTACGTGCAGATACCAAGT : 949
US5689056 : TCCCATTTGGAAGGATTGACTATGAGTCTATTCTTGGACAATGCTGTGAGATGCCGTGAGGATATGTGCAAAATACCGGTT : 944
US5750386 : -----GGCGCCAATCA-----TTTGTCTTAGTCCACTC-----AAATAAGGCCCATGGTCGGCAAAACCAACACAAAAT : 197
US5459252 : -----GGCGCCAATCA-----TTTGTCTTAGTCCACTC-----AAATAAGGCCCATGGTCGGCAAAACCAACACAAAAT : 535
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 1628
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 1628
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 999
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 1077
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 735
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 1003
US5837876 : TTGATTTGGTTCCAACAT-----TTAAAAAAGTTTCAGTG-----AGAAAGAATCGGTGACTGTTGATGATATAAACAAAGG : 2028
US5837876 : ----- : -
US7041873 : ----- : -
US7041873 : GTTAAACGCCGTCGACGAG---TCTAACGGACACCAACGCGAACCCAGCAGCGTCGCGTCGGGCCAAGCGAAGCAGACGCG : 989
US7041873 : GTTAAACGCCGTCGACGAG---TCTAACGGACACCAACC---AGCAGCAGCGTCGCGTCGGGCCAAGCGAAGCAGACGCG : 1635
US5633363 : GATGGGCGAGAGAAATTTCTCCCTGCGGGGACGGGGGAGCCATATCCTCGGTGGAGAAATTTCCCATATCATCCTTTATT : 1663
US20050109 : AAGCCTCTAACTTTGAATGTTCCAACCTGAACATAATAAATGTTGACCCGATCAGGGACAGAGCTAAAGCAAAATTTAGTG : 1661
US7193135 : CCTGTTTCGGTTGTGACTTGGTATTGTTAGGCTGTGTATCCAATGCTATCATTAGATATAAACTAGATAACCCGGTAT : 817
US20060230 : CCATTAATAAACAACACGAGCGTGGACGAAGAAGAGGAGGCGCGCAAGAAGCATACTAGCAGCTACGAAACCCCTTT : 1049
US20060143 : ATTCCTTAAATAAATTTATTTGGTTTCGCATGAAATATGAAGGAGAGAAATTTATTATTGTCACCTTATATATATATA : 837
US20060143 : TCATTACGATACAAAATCATAGTCCAATTTATTTCTTCTTACTGACTTTTTTCGGCTCTTTTACAAAAGCTAATATCTT : 784
US5689056 : AAAATATAAAATGAAAAAGTCCAGCGCGGCAACCGGGTTCCTCTATAAAATACATTTCCCTACATCTTCTCTCTCTCAC : 1038
US6395963 : TATTTGGTATGTGATCTTCTATATATAAAAAAAATCCATCGTCGAAACTAAGATTTTGTTTTCTATATATCTTTTAA : 1664
US20062600 : TTACAATGATGATGTGTTTCTGTTAATGTTGAATGCTTAGGGTATCCAAGTTTAATCATTAAATTAACATAACTAGCATT : 972
US20060143 : AAAAAGAAAAAGAACAAAAAAGATAATATTGTTTTCCCGACGACTAATCCACATTGATTTTTATAAAATAACTG : 643
US20060143 : AATTCATATAGACTAAATTTGCTGGAGGAGCATCGGACGTGGGTACAGACCTTTCCAAGCACAGAGTAGTGCTTAG : 614

```

```

* 1700 * 1720 * 1740 * 1760
CaPEROX : AAAATTATGTCCTAGAAAGAACACACATTCTAAATCTATTGATCAAAATTTCCATTCTTTAAGAAAGCTGCCTCTAA : 1754
US7119254 : GATCACAGTAAAAAATTCACCTTTAATTTTTGTCTTTCTTCATATTCATCTCACAGATTTCACACCTCTTTTGGATTCTC : 1247
US20070050 : AATTACCAATTTGTGTCGGTGAACCTCATATAGACAAACCTATGTGGTAATCAAACTATTAGTTAAAAAAAATTTGAACAT : 1170
US6649748 : AAGTTCAAATCACCTCACCGGGTTGGCTCAGATTGGTTTTCATACAAGAGGGGGTGTGAACCTCCCGTCCGCGACCTCT : 1091
US6927320 : CAAAAATAAAGAGAGAGAGAATATTGTTCAAAAGTGGTCTCTCTCTCTAATTATGTTTCTCAATCAACCAATTAG : 1753
US20042483 : GTGGAATGGATAAAGAAAAGAAAAGGAAAGCGAACATCTGGGCAAGAAAATCTAATCATGTGTAGAGATATGTCGTAGT : 1456
US6906241 : TCCATTTTCTCTATAGTCTTCAAACACTTTGATTTGCTTTTAGTCTTTTGTTCACCTCATCTCGAAATTTCTTGGTCAACCT : 1185
US5689056 : ----- : -
US6703541 : CGGCGCGCTCGGCGCAGCTGTGTCGTCGAGTCTACGACAGGTGATGCCCCAACCGGCTGTCCACCATTAGGAGCGGCGTG : 1153
US6262344 : GACCTGGTGTGTCCTTTTCCCACTTAAATGTACGGCTGATAATCACATCAGTGGCGACTTTAGGAAATAGAAAATTCGCA : 1749
US20050109 : TCCCTGCAGGCCAATCGCTCGCACAGACTTGCCTGATTCAGTTGACGAGAAGATGCCATATAAAGAAAGGAAAATTTGGCC : 624
US20060143 : AATAATTCGACACAGATCTACCAGCAATTTTCGACCGTTGCAACAAAGCCATGATTCAAAAAGTTTGTCTTTTAAATCTT : 887
US6384207 : TGAATACAGTAATCATCCAAGCACGTAACCACTCGCTAGACGTTTATGCCGTGTAGCCTGCTGGGTGTGTAGACTGGAGT : 1757
US20050109 : GGTTACATGAACGCTCCCGTGCTAGTTGCACCGATTTCCTAATCCCGAGCATGGAATATAAAATCCCTTGGTAACACC : 861
US20040199 : CATCGGTGAGTCCAAATAACAAATAGTCACTTTTCGAGTGGCTTCTCATGCCATTACCTATGCGCTCTCTTTAATGTACAG : 764
US20040199 : CATCGGTGAGTCCAAATAACAAATAGTCACTTTTCGAGTGGCTTCTCATGCCATTACCTATGCGCTCTCTTTAATGTACAG : 255
US20040199 : -----AGTGGCTTCTCATGCCATTACCTATGCGCTCTCTTTAATGTACAG : 45
US20031675 : ----- : -
US20031675 : AGTCCATGATACAGGATGAGATATTGACTATTTGACTAGAAACGTAGTAGGTGATATATGGACATTTCTGGTTTATTT : 1733
US7196247 : AGTCCATGATACAGGATGAG---TTTGACTATTTGACTAGAAACGTAGTAGGTGATATATGGACATTTCTGGTTTATTT : 1650
US20031675 : AGTCCATGATACAGGATGAGATATTGACTATTTGACTAGAAACGTAGTAGGTGATATATGGACATTTCTGGTTTATTT : 874
US7193136 : TATCTTGGACTCACAATT---AATTAATTCATAATCCAACCTACGCGGACGAACGCTAGGAATCACTTGTGCTGTT : 877
US7193136 : TATCTTGGACTCACAATT---AATTAATTTTAAATCCAACCTACGCGGACGAACGCTAGGAATCACTTGTGCTGTT : 608
US7193136 : TATCTTGGACTCACAATT---AATTCATTTTAAATCCAACCTACGCGGACGAACGCTAGGAATCACTTGTGCTGTT : 340
US7193136 : TATCTTGGACTCACAATTT---AATTCATTTTAAATCCAACCTACGCGGACGAACGCTAGGAATCACTTGTGCTTTT : 175
US20050976 : -ACAGTTAACGAGCGGGTATGGGTTGTTTTCTTGGAGCACTGTTGTTCTCTAGAATCTCTGAATCTCTCTCTGCTTGAT : 1692
US20050976 : -ACAGTTAACGAGCGGGTATGGGTTGTTTCTTGGAGCACTGTTGTTCTCTAGAATCTCTGAATCTCTCTCTGCTTGAT : 1697
US20060107 : -ATAATTAACGAGCGGGTAAAGGTTAGTTTTCGTTGAGCACTGTTGCGGTTTGAATCGCTGGACCTGC---GTGTTTAT : 1355
US20060107 : -ATAATTAACGAGCGGGTAAAGGTTAGTTTTCGTTGAGCACTGTTGCGGTTTGAATCGCTGGACCTGC---GTGTTTAT : 448
US20062600 : TGTT-----GTACATTTTTTTTGGCAAACTTTGGTGCATATACAACTGTAGAGATTA-ATGGCTCGAGATAACTAA : 1310
US20062600 : ATTTCCCTAAGTAGAATGGATGTCCTAGATTTTAAATTG-GTCATGTCTATGATTAAGTATATATTACTTAAAAAGTAT : 1436
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : CATGCCCTAACGAACATAATAATGAACATTTTTTCGCCACCTTAGACGGAGGAGGAAGAGAAACCAATAATATATACT : 1729
US20050109 : GCAGTTCTGGATGTGAATCTAGACAAACATTTGCTAGGTTCCGCACGTATAATTTTTTTCTTTTTTTTT : 1186
US6774282 : GCAGTTCTGGATGTGAATCTAGACAAACATTTGCTAGGTTCCGCACGTATAGTTTTTCTCTTTTTTTTGGGGGGG : 597
US20060143 : AATATATTTTCGTCAAAAATAAATGGTCGTATCTTTGCTAAAACCAAAAT--GTCATGATGTGATTACTCAATTTTCT : 940
US5689056 : ----- : -
US6271437 : CATTCACGTCGACACGCTCTTCTACTCCTTAATTTCTCTCTCTTTTCAAACCTTGACAAAGCCACAACCTCTCTCTCATCT : 1041
US6271437 : CATTCACGTCGACACGCTCTTCTACTCCTTAATTTCTCTCTCTTTTCAAACCTTGACAAAGCCACAACCTCTCTCTCATCT : 983
US5689056 : GGAATAGCAGGGCCATTGTTGCTTAACGGAAGAGAGTTTCGGTGCCCATGGCAACCAAGAGGATGTTTAGTGGCTAG : 1029
US5689056 : GGTATTGACGGGCTTTGTTGCTTGATGGGAGGACTCATGTCGAATGCAACTACAGAGGGGTGTTTAGTGGCTAG : 1024

```

US5750386	: GTGTTATTTTAAATTTTTTCCTCTTTATTGTTAAAGTT-----GCAAAATGTGTTATTTTGGTAAGACCCTATGGA	: 270
US5459252	: GTGTTATTTTAAATTTTTTCCTCTTTATTGTTAAAGTT-----GCAAAATGTGTTATTTTGGTAAGACCCTATGGA	: 608
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 1699
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 1699
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 1070
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 1148
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 806
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 1074
US5837876	: GCACATTGGTCAATAAACCATAAAAAAATTATATGACAGCT-----ACAGTTGGTAGCATGT--GCTCAGCTATTGAACA	: 299
US5837876	: -----	: -
US7041873	: -----	: -
US7041873	: ACGGCATCTCTGTCGCCCTCTGGACCCCTCTCGAGAGTTCGCTCCACCGTTGGACTTGCTCCGCTGTCGGCATAGAAAT	: 1069
US7041873	: ACGGCATCTCTGTCGCCCTGCC---ACCCCTCTCGAGAGTTCGCTCCACCGTTGGACTTGCTCCGCTGTCGGCATCCAG--	: 1710
US5633363	: TGTGGTACATATATATGCATAATCTTTTTTTTGACTGACATGGGAAAGATATCCCATCTCAATAGTAGAAAAATCTTG	: 1743
US20050109	: ATTGCAATTGCCCTCTGAGCCGATTCAGTTTTCAGTAGGAGATACGTTTAGAGTAAAAATTTGATCATTTCTTCTACCA	: 1741
US7193135	: ACAGTGTTTTATTCAAACTCTACCCTCTCCCTATGAAACTGTCT--CTATATCTTTCCCTCTTAATTACAGTCACGCACAT	: 894
US20060230	: CTCTTCGTCGCCCAAAATGTCATACAAAAGGCGCCCTCTTCTCTCTCGTCCCTTATCACCCAAATCCGATCCCTC	: 1129
US20060143	: TCGTAATCATTTTTCGTGAATAATCTCTCTCCCATTCATTAATTTCTCAGTATCTCTCTTCTTCTTCTTCTTACTTTATT	: 917
US20060143	: TATCTACAGTTGATCATATTTTAAATTTTTCACCAAAAGTATAAATTAACAGAGAAGCATTAATCATATCAGTGATCAAGGTG	: 864
US5689056	: ATTCCTACACCTCTCTTTTAAACAATATATCTTGCAATCATCAATCCCAACAAACACATTTTCTCTCCTCTTTCTCT	: 1118
US6395963	: AATGTAATAAATATTAGGTAAATATATGATTTTCTCATGTATTTTTTTTCAATAACCTTACTATTTTTAAATCTATCT	: 1744
US20062600	: CAGATTTAAGAAACAAGAAAATTAATTAATAAAAAAATAGTGCTTTATAAAGCGTTTACATTGATGAATGCTGTGGAGTGC	: 1052
US20060143	: AACTCAAGTGTCTGATCAAAACAAGTACCCTTAATGTCTCTGTCTATATAAATGTACACTCTTCCCAACAAACCA	: 723
US20060143	: CCATGTCATGCTAACATACACCAATTTGGTTCATACAAAATCCAATCAAAATTAATTTTAAATCTTTTGACACAGCTCT	: 694

		*	1780	*	1800	*	1820	*	1840		
CaPeROX	:	A	A	C	C	C	G	A	A	A	1834
US7119254	:	T	C	A	C	C	G	T	A	C	1327
US20070050	:	A	T	T	A	A	G	A	A	T	1250
US6649748	:	T	T	T	G	A	G	G	A	C	1171
US6927320	:	A	T	T	C	A	A	A	G	A	1833
US20042483	:	G	G	G	G	T	C	C	A	A	1536
US6906241-	:	A	A	A	C	T	A	G	T	A	1265
US5689056-	:										-
US6703541	:	A	A	C	C	G	G	T	A	G	1233
US6262344	:	C	A	A	T	G	C	A	T	A	1829
US20050109	:	A	A	G	C	T	A	A	G	C	687
US20060143	:	A	C	C	G	T	G	G	C	T	967
US6384207	:	A	C	T	G	G	A	C	A	A	1837
US20050109	:	G	T	G	A	T	C	A	A	G	941
US20040199	:	A	T	T	T	A	A	G	G	T	838
US20040199	:	A	T	T	T	A	A	G	G	T	329
US20040199	:	A	T	T	T	A	A	G	G	T	119
US20031675	:										-
US20031675	:	T	A	T	A	T	T	C	T	A	1809
US7196247	:										1718
US20031675	:	T	A	T	A	T	T	C	T	A	950
US7193136	:	T	-	-	-	-	-	-	-	-	950
US7193136	:	T	-	-	-	-	-	-	-	-	678
US7193136	:	-	-	-	-	-	-	-	-	-	410
US7193136	:	T	A	G	C	A	T	G	C	A	247
US20050976	:	G	A	C	C	G	A	G	C	G	1763
US20050976	:	G	A	C	C	G	A	G	C	G	1768
US20060107	:	G	A	C	C	A	G	C	G	G	1423
US20060107	:	G	A	C	C	A	G	C	G	G	516
US20062600	:	G	A	T	A	T	T	A	A	A	1375
US20062600	:	A	A	A	T	A	A	T	A	A	1512
US20060143	:										-
US20010047	:										-
US20010047	:										-
US20010047	:										-
US20010047	:										-
US20010047	:										-
US20010047	:										-
US20010047	:										-
US20010047	:										-
US6518483	:	A	G	C	T	C	A	T	A	G	1805
US20050109	:										1253
US6774282	:	G	G	T	G	G	G	G	G	G	671
US20060143	:	A	C	T	A	A	A	T	A	A	1000
US5689056-	:										-
US6271437-	:	C	A	T	A	A	A	T	A	A	1117
US6271437-	:										


```

US7041873- : TCGGTGGCGGAG----CGGCAGACGTGAGCCGGCAGCGGCGCCTCCTCCTCCTCACGGCACCGGCAGCTACGGGGG : 1145
US7041873- : -GCGTGGCGGAG----CGGCAGACGTGAGCCGGCAGCGGCGCCTCCTCCTCCTC--CACCGGCAGCTACGGGGG : 1783
US5633363- : GGAACGGTAGGATCGAACACAAAGATCAGCTAGCTTGTAAATCACCGAGCCATATAGCTAGAGGGTAATAGATCATGAATC : 1823
US20050109 : TGCAGGGTGCACCCGTACCCACCGCTATATGCATAGCTCTGCGCTTGGATCCCGTGAAGTGCAGAAAAAATACTTCAACT : 1821
US7193135 : GATGCTGTATCACCCGCATGGAGAGAGAGAAAAATAGATTGGTCAGGGTTGTATATGCAAGGCCCTACGCTTCTATTTA : 974
US20060230 : TTCTCTTCTCTCTCTCTCTCCACATCCAGTTCGATTCTCATCTCTCCACACAAATCACGCCATGGATCCCTACAA : 1209
US20060143 : GTTGCTTTTAAACCTTCAATTGCTCATAAACCAATATATATATCAAAACAAACAAACAAAAATCAGAAATCCCCCTA : 997
US20060143 : GGCCAACATGGGCCC-ATACAGCCCAAAAGAACCCAAACCCAAAGTCCAAAGATATCTTTAAGCTCAGTTTATAAATT : 943
US5689056- : CACCGCGGCAGACTTACCGGTGAAAAATGGACGTTCCGCCGAGATCTGAAGAGCCTGTTTATCCATTAAGGTCCTTTG : 1198
US6395963 : GCAATATATATTTTATTTGAAAGGTAAAAAATTTTGGATTTAGTATTTGTATATGATATCTGATTATGTTTTCAGGAG : 1824
US20062600 : TGATTAGGTTAGGCATATACCATTTGATGCTGAATTTGATTCCTTTGACCTAGCAAGGTTTCATGTTACCTATATATATGCAT : 1132
US20060143 : TATCTCAAAAT----- : 734
US20060143 : TTGAAAAACACCTCTCATACTATAGCTACGGAAGCTTCAATTTCAAGGTTTGTCTAAAAGCTAACGATT----- : 763

```

```

CaPEROX : ATTTGCCTTACGATGTACTAATACTGAAACCAATAATTAATCTCATATCCAGCTATGTGTTTTTGGAGTGAAACAT : 1914
US7119254 : CCCACCCGGTCCAACCGGTTACAAACCCGGTCAATCCAAATTCGCCTCGGAATCCAAACCTCCATACCAATCTAACAG : 1407
US20070050 : TGAAGTAACTAAGAACAGTTTTTTTAGCAGTTAAGCTGAATTTCAAAAGCCATAATCTTCGCTCTGAATCTCATAGCC : 1330
US6649748 : CAAATAATTAATCAAAAAATAAAAAAGACCCCAATTAAAGTCCCAACCACTCTCAAATATTCTATTTAAGGGAACCT : 1251
US6927320 : TTATTCTCTAGTCTTTTAAATTAATAAAAACTCACAAATCCACCAACCCATTCTCTACAACCTACCTTCATCTAGA : 1913
US20042483 : TCCCGTACTTTTCTTTCTGCTGCCAATCTTTTCTTCACTCACTCTCTCTCTTATTCTTCTTCCACTTCTCCATC : 1616
US6906241- : CTCAAATGGAATTAATTAATAATTTTTTCTAAAGGGAATGACAAATAAGAAAAAGACACTTGATTGGAAGCCTTAAT : 1345
US5689056- : ----- : -
US6703541 : TGAAGTCTCCATCATCTTGCATTCTGCTTTCTTTCTAGAAAAACAAACACCAGCAAAATTTAGGGCATTGCTTTAAAAAC : 1313
US6262344 : CGTCTTTATAATAACATAAGATAAAATTTTACTTAATTAGCTACGGAACCTACATTAGTATTCAATTGATATACTAATGG : 1909
US20050109 : ----- : -
US20060143 : ACTATCGTTCTTGATCTTAAAAGCTCAACAAAA----- : 1000
US6384207 : GTAAGCCACTAGCTCACGTTCTCCATGAGCTCTTCTCTGCTGTTTCTTCTCTGCTAAGTGGGTTATGATATGACGTC : 1917
US20050109 : GCGAGGTAGCTGGCTGCTTGTGTGATCGAGC----- : 973
US20040199 : AT-AATATTGAAGAAATATAAGTATTGTGTAGTTGCGAGAGAGTTGAAAAATG----- : 889
US20040199 : AT-AATATTGAAGAAATATAAGTATTGTGTAGTTGCGAGAGAGTTGAAAAATG----- : 380
US20040199 : AT-AATATTGAAGAAATATAAGTATTGTGTAGTTGCGAGAGAGTTGAAAAATG----- : 170
US20031675 : ----- : -
US20031675 : GA-ATTCCTTCGCTCTTTTATGAGTGTAATAA-CAAAACATTGAATTAGGAACAATTATTATCACGTTACTTAAAAATAAATA : 1887
US7196247 : GA-ATTCCTTCGCTCTTTTATGAGTGTAATAA-CAAAACATTGAATTAG-----ATTATTATCACGTTACTTAAAAATAAATA : 1791
US20031675 : GA-ATTCCTTCGCTCTTTTATGAGTGTAATAA-CAAAACATTGAATTAGGAACAATTATTATCACGTTACTTAAAAATAAATA : 1028
US7193136 : TA-GTGCTCTATCTTCTCATGGAGTAGCATCACTACCTCAGTATATACTCAAAATGGACAACCTTCTCCTACCTCC--- : 1026
US7193136- : TA-GTGCTCTATCTTCTCATGGAGTAGCATCACTACCTCAGTATA---TCTAATTTGGACAACCTTCTCCTACCTCC--- : 751
US7193136- : TA-GTGCTCTATCTTCTCATGGAGTAGCATCACTACCT---CAGACTCACTAATTGGACAACCTTCTCCTACCTCC--- : 483
US7193136- : TA-GTGCTCTATCTTCTCATGGAGTAGCATCACTACCTA---TATACTCACTAATTGGACAACCTTCTCCTACCTCC--- : 321
US20050976 : AC-AACGTGAACCTCATGCAGCTTAAA-ATACAGCCA-----GTTGTGATCCATCTGCCACAGCTGTTTCTACCTCA--GA : 1835
US20050976 : AC-AACGTGAACCTCATGCAGCTTAAA-ATACAGCCA-----GTTGTGATCCATCTGCCACAGCTGTTTCTACCTCA--GA : 1840
US20060107 : CC-AATATTGTTCCATTGCTGCATGCACATTATCTAT---ACCACGACGACACAACGTGAACCTCGTGCAGCTTTTTAA : 1499
US20060107 : CC-AATATTGTTCCATTGCTGCATGCACATTATCTAT---ACCACGACGACACAACGTGAACCTCGTGCAGCTTTTTAA : 592
US20062600 : T--ATTAGGCACAATTTAGGTAGGTTACTTTATAGCTTTT---GAGCAGTTTATCCATTCTTACCTCAGCTCACAATAGT : 1451
US20062600 : TG-AATGTTTATGATTAAATGAGGCATATACCCCATTTGAT--GCTAAATTTGATTC-TTTTAACTCATGAGGCATACCTGG : 1588
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : CA-GGACTGATTATGGCCACTGTACGACCGACTGTAGGGGATGAAAGCCTCGTTGGGTTTGTTCAGGTGGCCGTTTCAGCA : 1884
US20050109 : TA-GATTACGGTGACCTTGATGCACCACCGCGTGCTATAAAAAACCAAGTGTCCCGAGTCTACTCATCAACCAATCCATAA : 1332
US6774282 : TA-GATTACGGTGACCTTGATGCACCACCGCGTGCTATAAAAAACCAAGTGTCCCGAGTCTACTCATCAACCAATCCA--- : 747
US20060143 : ----- : -
US5689056- : ----- : -
US6271437- : TT-CTCCGAAATGGTTGATTACATTGGAATCAAGATGCCAAGTTCGGACATGCCTTCCAAACCTCTAAAACCTCTCT : 1196
US6271437- : TT-CTCCGAAAT----- : 1069
US5689056- : GC-ACCATGTGTGAGGTTTCGGCACAGCCAAAGGGCAGCAGAGTTGAAGTTCTTTGTTGAAGATCCCATAAAAATTTGAGT : 1184
US5689056- : GC-TCCGGTTGTCCGGTTTCAACACCGCCAAAGAGCCGAGAGTTGAAATTTCTTCGTTAGGATCCCCCTTAACCTTTGAGA : 1179
US5750386 : ---GCTTCTTTTGGGGCATTTTTTC--TAGTGAGAA----- : 375
US5459252- : ---GCTTCTTTTGGGGCATTTTTTC--TAGTGAGAA----- : 713
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 1852
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 1852
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 1223
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 1301
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 959
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 1227
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 452
US5837876- : GT-GCTACAAATAACACACTATTCAACAACCATATAATAAAACGTTGTCAGTACTAAAAACAAATATAAAATAAATCTATGT : 149
US7041873- : ----- : -
US7041873- : AT-TCCTTTCCACCGCTCCTACTAGAACTAGTGGCCTATGCGTATGGTATGACGTGTGTTCAAGATGATGACTTCAAAAC : 1224
US7041873- : AT-TCCTTTCCACCGCTCCTACTAGAACTAGTGGCCTATGCGTATGGTATGACGTGTGTTCAAGATGATGACTTCAAAAC : 1859
US5633363- : AA-ATGTTTTTTTCAATAAATTAAGGCTCTAAATTTATTTTAAATTTAAAAATAAATAAATAAATATAGTTCGATTCTTAC : 1902
US20050109 : GT-TCAACAACTTTGGGCGTCCACCGTCCATGCATCACCATGATACCTCTGCTGATGTTTAGCATCGATCAGCTTACTAG : 1900
US7193135 : AA-GACGTAATCGTGGTTGAGGCGAGACACAAGCTAGCAATTGAAAAAATAAAGGCA----- : 1034
US20060230 : GG-TGCCGCTCTTTTCTGATTTCTTTCTTCTAGATCGATCGTCGATTTGGTTTGGTTTGGTTTCTGATGCGCTCATC : 1288
US20060143 : AT-A----- : 1000
US20060143 : GC-ATTCCTTTCTCTCTCTACAGGCCAACCCAAACATCAACGACTTATTCGAGAGA----- : 999
US5689056- : CC-GCCGATGAAAAACCTCTCAACCCCAACAAGAACCAACAACAACAAGAGGACAAGAATACCTTCTCATTTGATGC : 1277
US6395963 : CCGATTAGATGTCACCTTTTATGTCATTTAGCTTTTATGCTTTTCAACGCTTTTGAATTTGAACCTCAAAATATTTTGGCCT : 1904
US20062600 : ATTAATAGTTTAGCTGCTCAAAGTGCAGGCACCTAGCTAGCTCCTCATACAACTTGACTTCTCACAAGGAGGCTTCTCT : 1212

```

```
US20060143 : ----- : -
US20060143 : ----- : -
```

	*	1940	*	1960	*	1980	*	2000
CaPEROX :	TCAAAATGGCTCTA	ACTATTTTAAAGGGTGGC	AATCTGTTTTGTCTTC	CAGAGATTAGAATA	GAAAGAAATACATTG	GGTTC :	1994	
US7119254 :	AAAAACCTCCAAT	CACAACCTCCACG	TGGTGATCACTCAT	TGGCTCTTATCTG	GACCAAGAGGACCT	TTTTAGTATAA :	1487	
US200700050 :	CCGGTCTCCTTCT	AGCTATATATACGT	GTGTGAATAAGTCG	TACCCTCAAGCC	AACACCAATAACA	AGAATG----- :	1405	
US6649748 :	TAGAGGCAATTCA	TGCATCCTCAACCC	CTTCTTCTCATTTT	CTTAATCTTACAT	TTTCTCTTGACC----- :	1318		
US6927320 :	TTTACCCACTCCC	ACCGAGAAAACA	CAGAAAAAATATAC	ATATATAAAATAT	ACAAGACAACATG	ATGCTGATGCAA :	1993	
US20042483 :	TTAGATTTTTTCT	TTCAGATTTCGATT	TCAGATTAGGTTTT	CTTCTTCTTCTT	CTCTCTCTCTCT	CTCTGTCGAAAT	TGAGA : 1696	
US6906241- :	TTGCGTTGCTGT	GATTGGATTCCCC	AAGCCCTATATAA	AGGGGTCAATTTG	TCAATGTTTAAAT	CAACGCCATAAC	TAAC : 1425	
US5689056- :	-----	-----	-----	-----	-----	-----	- :	
US6703541 :	TCTAACCCATAAT	TTCTGCAAGGCTG	CGACGCTCCCTT	CTGCTGAAC-----	-----	-----	: 1363	
US6262344 :	TAATTACTAATTA	ATTGCGGAAGCCG	AGAGAGGTTGAT	GTTGCGACGGTGC	ATGTGAAGAGCTTT	TGATACGTAAGT	TGGAG : 1989	
US20050109 :	-----	-----	-----	-----	-----	-----	- :	
US20060143 :	-----	-----	-----	-----	-----	-----	- :	
US6384207 :	GTATAAATAATCT	CAATACTTCCTT	ATTTTCAGCATGG	CTCTTTATGTTT	ATTATTAACAGT	AGCAACCAACG	CCGCT : 1997	
US20050109 :	-----	-----	-----	-----	-----	-----	- :	
US20040199 :	-----	-----	-----	-----	-----	-----	- :	
US20040199 :	-----	-----	-----	-----	-----	-----	- :	
US20031675 :	-----	-----	-----	-----	-----	-----	- :	
US20031675 :	TACTAAAACCGTT	GAAATGAAATCTT	CATATTGATAAGT	GTAGGTAGACC	CCACAACCAACAT	TGAATAGAATAA	TAATTT : 1967	
US7196247 :	TACTAAAACCGTT	GAAATGAAATCTT	C-----	TGATAAGTGTAGG	TAGACCCACAAC	CAACATTTGAAT	AGAATAAATTT : 1866	
US20031675 :	TACTAAAACCGTT	GAAATGAAATCTT	CATATTGATAAGT	GTAGGTAGACC	CCACAACCAACAT	TGAATAGAATAA	TAATTT : 1108	
US7193136 :	-----	-----	-----	-----	-----	-----	- :	
US7193136- :	-----	-----	-----	-----	-----	-----	- :	
US7193136- :	-----	-----	-----	-----	-----	-----	- :	
US7193136- :	-----	-----	-----	-----	-----	-----	- :	
US20050976 :	TGTGCTACACAGT	GTATTACCTGTTT	CTACCTCGCAGAT	GTGCTACA-----	CAGTT-GCTTAT	GACTGCCTATAA	AAATG : 1909	
US20050976 :	TGTGCTACACAGT	GTATTACCTGTTT	CTACCTCGCAGAT	GTGCTACA-----	CAGTT-GCTTAT	GACTGCCTATAA	AAATG : 1914	
US20060107 :	AATACAGCCAGTT	GTGATCCATCTAC	CTGTCTGTGACAG	CGTGCTACAGCC	TACAGTTTAGT	GACTGCTGCCT	TATAAAATG : 1579	
US20060107 :	AATACAGCCAGTT	GTGATCCATCTAC	CTGTCTGTGACAG	CGTGCTACAGCC	TACAGTTTAGT	GACTGCTGCCT	TATAAAATG : 672	
US20062600 :	CAAGTTAGTTTTT	TTTTTTGTTTACT	AATATTTAATGT--	GTCTGTG-CTTA-	ATTATGACGGG	AA-ACTAGGTA-AC	: 1525	
US20062600 :	CTTGGCTGGGATT	GTTACGTATATAT	ATAGCTGAGTTT	TAATTATCTG-CT	CACAGTGTGCATA	AAACCTAACTCCT	TATAC : 1667	
US20060143 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US20010047 :	-----	-----	-----	-----	-----	-----	- :	
US6518483 :	GTCAATTCTCCT	CGTCCATATTTCT	TCCATAGAAACA	ACCAGGAACCGAG	TGGAGGAGAGGG	AGGGAGGATGGA	AAGGAG : 1964	
US20050109 :	CTCGAAACCTTTT	CTGTGCTCTGTCT	GTGTGTTTCCA	AAGCAACGAAAG	AGCTCGAGG-----	-----	: 1392	
US6774282 :	-----	-----	-----	-----	-----	-----	- :	
US20060143 :	-----	-----	-----	-----	-----	-----	- :	
US5689056- :	-----	-----	-----	-----	-----	-----	- :	
US6271437- :	CTCTCCGACAACA	AGTCCTTACCCT	CTTTGCAACTACC	CTTCCGACCACAG	ATATCTCTCACG	CCGCACCTTCTG	TTTG : 1276	
US6271437- :	-----	-----	-----	-----	-----	-----	- :	
US5689056- :	CACTTGCTAACGT	TTTTCAACCAGTA	AGT-----	-----	-----	-----	: 1212	
US5689056- :	TTCTTTCCCTT	TATGTTCAACAAG	TGAGT-----	-----	-----	-----	: 1207	
US5750386 :	-----	-----	-----	-----	-----	-----	- :	
US5459252- :	-----	-----	-----	-----	-----	-----	- :	
US5837876 :	TTGTAAGCACTCC	AGCCATGTTAAT	GAGTGTCTATTG	CGCTGTTAACTCT	CACCTTATAAAAT			

	*	2020	*	2040	*	
CaPEROX :		GTAGCAAAACAATGGCATTGTCCAAGTCTCTTCTCTTATCTTTCTTCAACTGTT				: 2050
US7119254 :		AGAGCCCCCTCGTTGGTCCTATCACCTTC-----				: 1516
US20070050 :		-----				: -
US6649748 :		-----				: -
US6927320 :		TATACACAACAAAGTATTAAATCTTAGATATTGTGGGTCTCCCTTTCTTCTATTCA				: 2049
US20042483 :		TAAATATAAGAGAAAATTGGGATAATAGGATG-----				: 1728


```

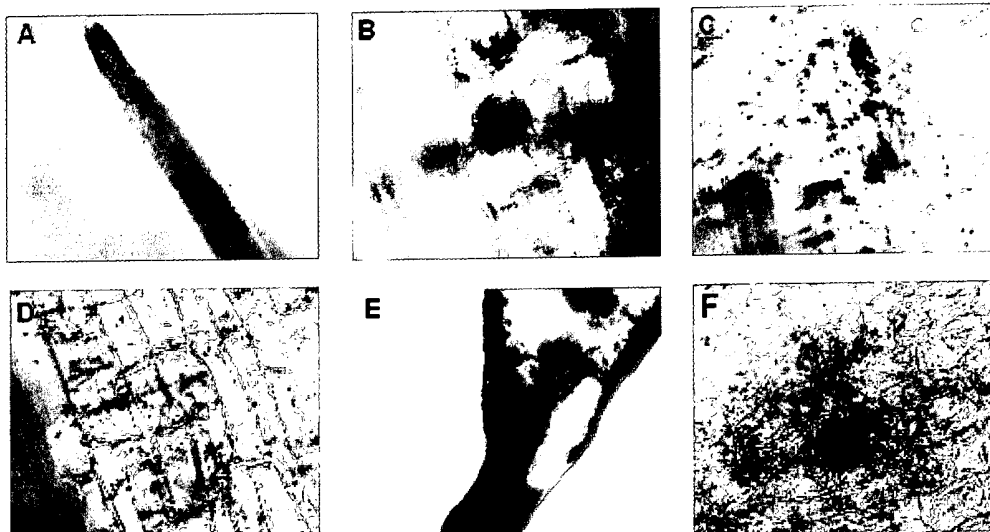
US6906241- : TCTAAATAAGCCATTTTATACTTTGAGCAAATCAAAAAAGTATA----- : 1470
US5689056- : ----- : -
US6703541 : ----- : -
US6262344 : CACTCATGATAAGCGAAGTTGTCTATTTATAAAGTTTAATTTACTGTGCTTTTAT : 2045
US20050109 : ----- : -
US20060143 : ----- : -
US6384207 : CGATGTTTCCTTCAAGAAACGGCCACTCACTATGTGGTGTGCAGAAGAACAAATGT : 2053
US20050109 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20040199 : ----- : -
US20031675 : ----- : -
US20031675 : CCCCGTACAGTGTCTGCTCCACTATGTGGCTATAAAATGGAAGCATTGAAGGTTGTTT : 2023
US7196247 : CCCCGT---TGTCGTCCACTATGTGGCTATAAAATGGAAGCATTGAAGGTTGTTT : 1918
US20031675 : CCCCGTACAGTGTCTGCTCCACTATGTGGCTATAAAATGGAAGCATTGAAGGTTGTTT : 1164
US7193136 : ----- : -
US7193136- : ----- : -
US7193136- : ----- : -
US7193136- : ----- : -
US20050976 : GCCGGGATCGGTGAGGCTGCTGGAACCAAGGAGAGAGAGCATATATATC----- : 1958
US20050976 : GCCGGGATCGGTGAGGCTGCTGGAACCAAGGAGAGAGAGCATATATATC----- : 1963
US20060107 : GCTGGCTGCTG-GAGCAAAGCCAAACCAATCAGATAGAGAGC----- : 1620
US20060107 : GCTGGCTGCTG-GAGCAAAGCCAAACCAATCAGATAGAGAGCATG----- : 716
US20062600 : AAATTCAG----- : 1533
US20062600 : CACTTTACTTCTTTATCAACTTGGTTTCTTGCAAGGAG----- : 1705
US20060143 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US20010047 : ----- : -
US6518483 : AGAGAGCCGTTGCTGGGGCGTCAGGTGAAATCGAGTGGCCTCCGGTTGCCGGGGAC : 2020
US20050109 : ----- : -
US6774282 : ----- : -
US20060143 : ----- : -
US5689056- : ----- : -
US6271437- : CGCCACTTACGACTACTATCTCCGCTCTTCCTCAACTCAGAAAGCTTTGGAACCTCT : 1332
US6271437- : ----- : -
US5689056- : ----- : -
US5689056- : ----- : -
US5750386 : ----- : -
US5459252- : ----- : -
US5837876- : ACCAAAAACAAC----- : 1945
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US5837876- : ----- : -
US7041873- : ----- : -
US7041873- : CGCTACCATCCCTAGAGCTGCA----- : 1326
US7041873- : CGCTACCATCCCTAGAGCTGCA----- : -

```

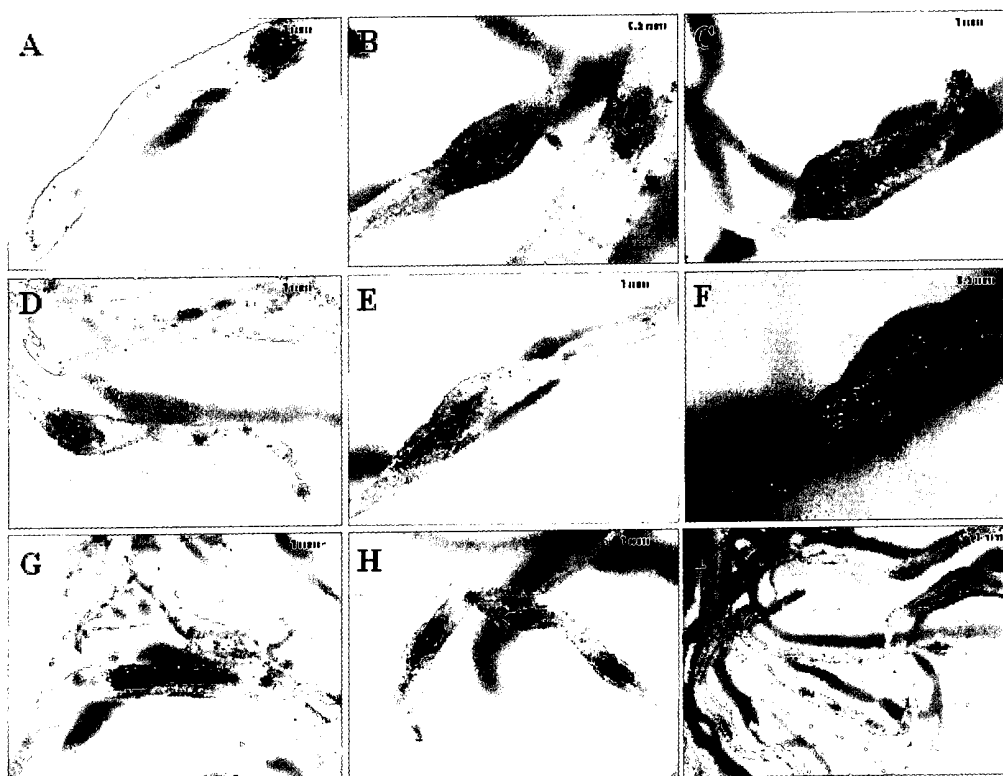
```

----- : 1955
US5633363- : TGAGATATAGAATAAGGTACCCCTATAGAATAAAGAACTTTCTCTGAAAAGTCTGA : 2038
US20050109 : CATCCACCAGTGGAGCAATCTTGGCCACTTGCCTGTTGAACTTGGTTCTGGAGT : 2036
US7193135 : ----- : -
US20060230 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -
US5689056- : TATCAAGGTGGCGTGAGAAAATCAGGAATTC----- : 1388
US6395963 : ATACTATTTGTCAAAAGAGGACTAAAATGCTAAATTCAGAAAGAAAATGAAGGAA : 2040
US20062600 : ----- : -
US20060143 : ----- : -
US20060143 : ----- : -

```



ANEXO 2



ANEXO 3

REIVINDICAÇÕES

1. Um polinucleotídeo com expressão tecido-específica caracterizado por compreender uma sequência selecionada dentre o grupo consistindo de:

a) sequências que possuam pelo menos 40% ou 45% de identidade de sequência, preferencialmente cerca de pelo menos 50% ou 55% de identidade de sequência, mais preferencialmente cerca de pelo menos 60% ou 65% de identidade de sequência, mais preferencialmente cerca de pelo menos 70% ou 75% de identidade de sequência, mais preferencialmente ainda cerca de pelo menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de sequência quando comparada com a SEQ ID NO1;

b) complementos da sequência descrita em SEQ ID NO1;

c) complementos reversos da sequência descrita em SEQ ID NO1;

d) sequências reversas da sequência descrita em SEQ ID NO1.

2. Um polinucleotídeo de acordo com a reivindicação 1 caracterizado pelo fato da expressão tecido-específica ser na raiz.

3. Um gene quimérico caracterizado por compreender:

a) um polinucleotídeo cuja sequência possua pelo menos 40% ou 45% de identidade de sequência, preferencialmente cerca de pelo menos 50% ou 55% de identidade de sequência, mais preferencialmente cerca de pelo menos 60% ou 65% de identidade de sequência, mais preferencialmente cerca de pelo menos 70% ou 75% de identidade de sequência, mais preferencialmente ainda cerca de pelo menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de sequência quando comparada com a SEQ ID NO1, opcionalmente ligado a sequências acentuadoras de expressão ou promotores de interesse; operacionalmente ligados a

b) uma sequência de polinucleotídeo de interesse.

4. Um gene quimérico de acordo com a reivindicação 3 caracterizado pelo fato da sequência de polinucleotídeo de interesse poder ser uma região de codificação ou uma região de não codificação.

5. Um gene quimérico de acordo com a reivindicação 4 caracterizado pelo fato da região de codificação ser isolada de um gene endógeno ou heterólogo.
6. Um gene quimérico de acordo com a reivindicação 3 caracterizado pelo fato da sequência de polinucleotídeo de interesse poder estar na orientação sense ou antisense.
7. Um gene quimérico de acordo com a reivindicação 3 caracterizado pelo fato das sequências acentuadoras de expressão serem selecionadas do grupo consistindo de SV40, HSV-1, AMV, HPV-16, TMV, entre outros.
8. Um vetor recombinante caracterizado por conter um gene quimérico de acordo com a reivindicação 3.
9. Um vetor recombinante caracterizado por compreender:
 - a) um polinucleotídeo cuja sequência possua pelo menos 40% ou 45% de identidade de sequência, preferencialmente cerca de pelo menos 50% ou 55% de identidade de sequência, mais preferencialmente cerca de pelo menos 60% ou 65% de identidade de sequência, mais preferencialmente cerca de pelo menos 70% ou 75% de identidade de sequência, mais preferencialmente ainda cerca de pelo menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de sequência quando comparada com a SEQ ID NO1, opcionalmente ligado a sequências acentuadoras de expressão ou promotores de interesse; operacionalmente ligados a
 - b) uma sequência de polinucleotídeo de interesse; e
 - c) uma sequência de terminação.
10. Um vetor recombinante de acordo com a reivindicação 9 caracterizado pelo fato da sequência de polinucleotídeo de interesse poder ser uma região de codificação ou uma região de não codificação.
11. Um vetor recombinante de acordo com a reivindicação 9 caracterizado pelo fato da sequência de polinucleotídeo de interesse ser isolada de um gene endógeno ou heterólogo.
12. Um vetor recombinante de acordo com a reivindicação 9 caracterizado pelo fato da sequência de terminação ser selecionada do grupo consistindo de sinal de terminação de SV40, sinal de adenilação de HSV TK, sinal de terminação do

gene da nopalina sintetase de *Agrobacterium tumefaciens* (NOS), sinal de terminação do gene da octopina sintetase, sinal de terminação do gene 19S e 35S do CaMV, sinal de terminação do gene da álcool desidrogenase de milho, sinal de terminação do gene da manopina sintetase, sinal de terminação do gene da beta-faseolina, sinal de terminação do gene da ssRUBISCO, sinal de terminação do gene da sucrose sintetase, sinal de terminação do vírus que ataca o *Trifolium subterranean* (SCSV), sinal de terminação do gene *trpC* de *Aspergillus nidulans* e outros semelhantes.

5

10

13. Um vetor recombinante de acordo com a reivindicação 9 caracterizado pelo fato das seqüências acentuadoras de expressão serem selecionadas do grupo consistindo de SV40, HSV-1, AMV, HPV-16, TMV, entre outros.

14. Uma célula transformada caracterizada pelo fato de conter um vetor recombinante de acordo com qualquer uma das reivindicações 8 a 13.

15

15. Um microorganismo transformado caracterizado pelo fato de conter um vetor recombinante de acordo com qualquer uma das reivindicações 8 a 13.

16. Uma planta, ou uma parte, ou um propágulo ou progênie da mesma caracterizada por compreender um vetor recombinante de acordo com qualquer uma das reivindicações 8 a 13.

20

17. Método para modificar a expressão de genes em um organismo caracterizado por incorporar estavelmente no genoma do organismo um vetor recombinante de acordo com qualquer uma das reivindicações 8 a 13 ou um gene quimérico de acordo com qualquer uma das reivindicações 3 a 7.

18. Método de acordo com reivindicação 17 caracterizado pelo fato do organismo ser uma planta.

25

19. Método para produzir uma planta tendo a expressão de um gene modificada caracterizada pelo fato de compreender as seguintes etapas:

a) transformar uma célula de planta, tecido, órgão ou embrião com um vetor recombinante de acordo com qualquer uma das reivindicações 8 a 13 ou um gene quimérico de acordo com qualquer uma das reivindicações 3 a 7;

30

b) selecionar células transformadas, calos de células, embriões ou sementes;

- c) regenerar plantas maduras de células transformadas, calos de células, embriões ou sementes selecionados na etapa (b);
- d) selecionar plantas maduras da etapa (c) tendo a expressão do gene modificada quando comparada com uma planta não transformada.

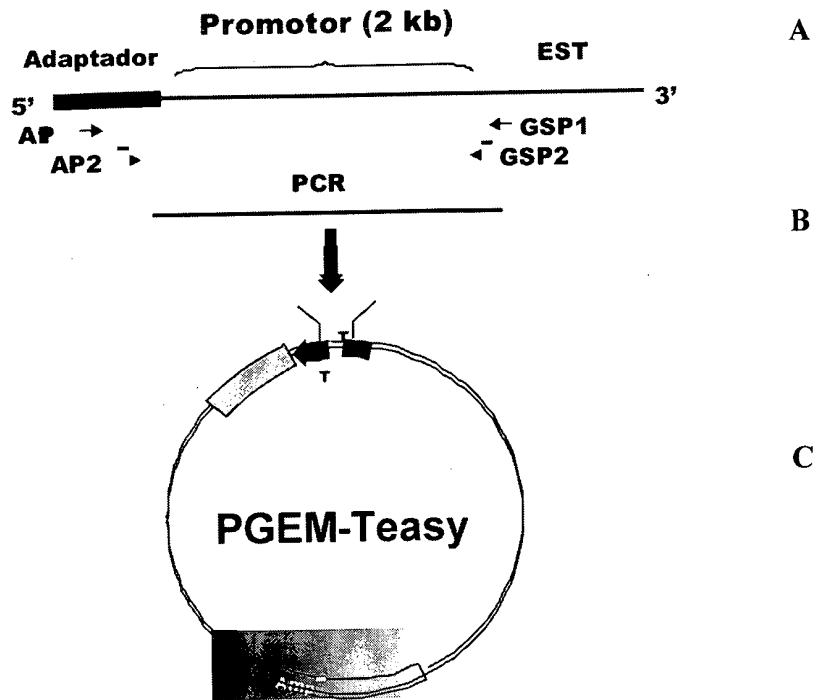


FIGURA 1

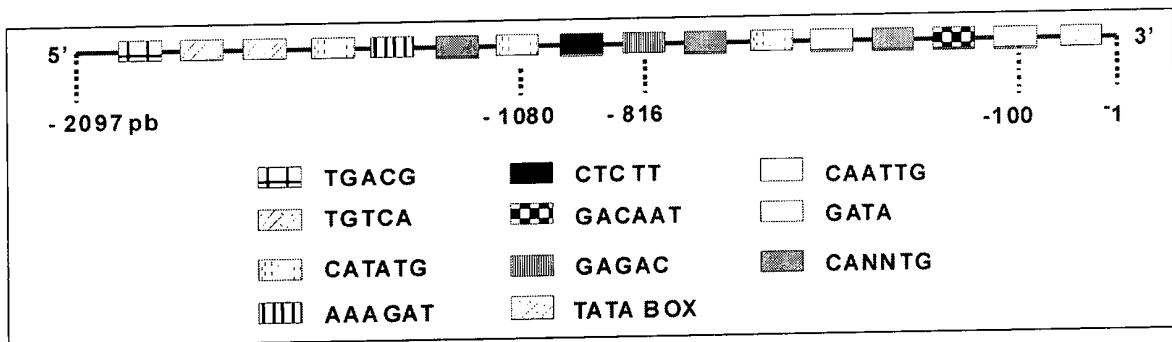


FIGURA 2

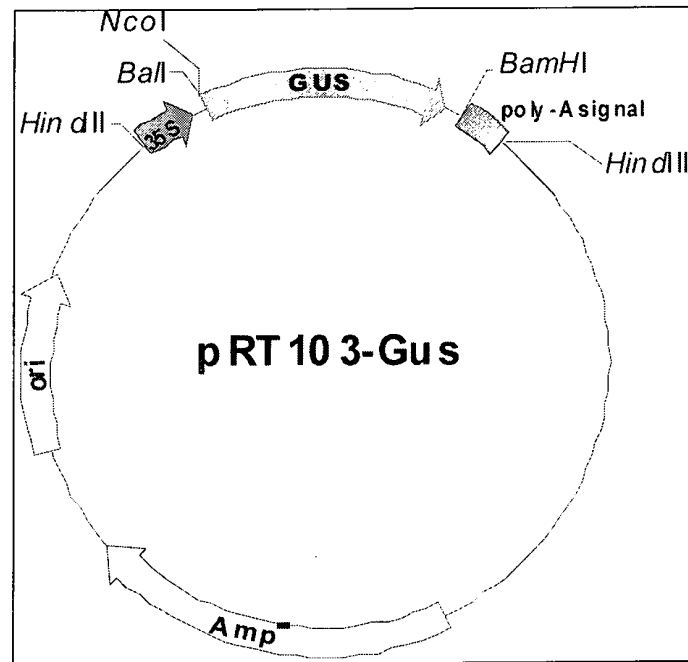


FIGURA 3

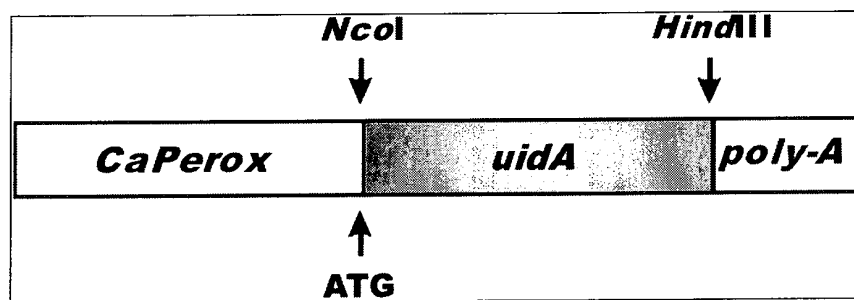


FIGURA 4

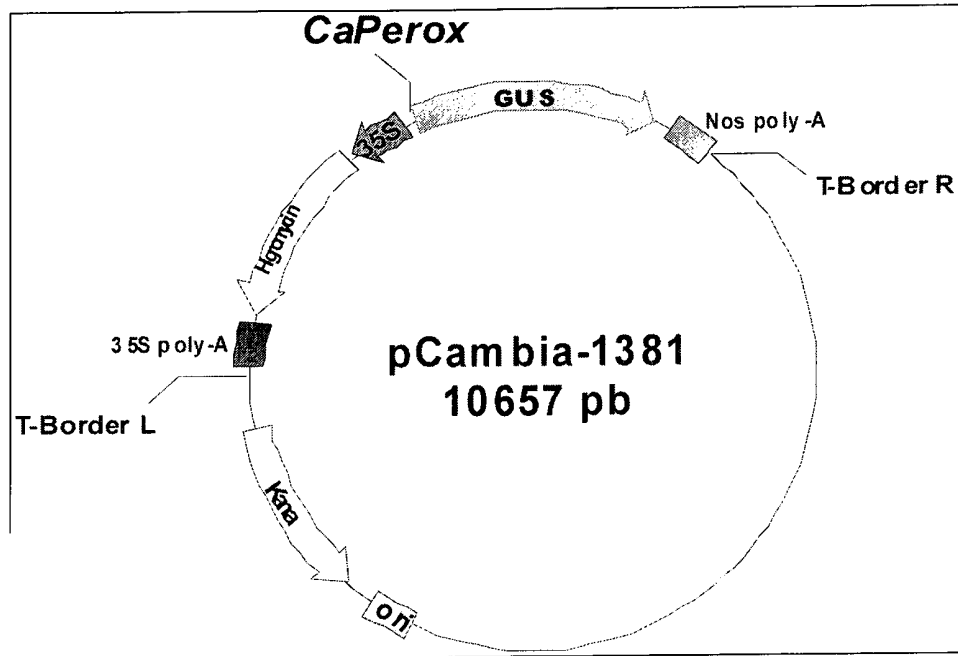


FIGURA 5

RESUMO**COMPOSIÇÕES E MÉTODOS PARA DIRECIONAR A EXPRESSÃO DE GENES USANDO O PROMOTOR DO GENE DA FAMÍLIA DAS PEROXIDASES DE PLANTAS DE CAFÉ**

5 A presente invenção compreende a descrição de uma nova sequência regulatória provida para direcionar a expressão de uma sequência de nucleotídeo, tal como genes estruturais, em raízes de plantas, incluindo monocotiledôneas e dicotiledôneas. Mais especificamente a invenção refere-se a sequências regulatórias de polinucleotídeos isoladas de plantas de café que são capazes de iniciar e acionar a transcrição de polinucleotídeos, e

10 ao uso destas sequências regulatórias no direcionamento de transcrição de polinucleotídeos endógenos e/ou heterólogos para tecidos radiculares e produção de polipeptídeos. A invenção descreve ainda construções de DNA que contém o promotor de um gene que codifica uma proteína da família das peroxidases de plantas de café que está operacionalmente ligado a um gene heterólogo e/ou endógeno. Além disso, a invenção diz

15 respeito ao uso destas construções na forma de vetores de expressão, vetores recombinantes e em plantas, células vegetais ou protoplastos transgênicos. A invenção ainda descreve um método utilizando tais construções contendo o promotor de um gene que codifica uma proteína da família das peroxidases de plantas de café para produção de plantas, células vegetais ou protoplastos transgênicos.