BIOSYNTHETIC GENES AND HOST CELLS FOR THE SYNTHESIS OF POLYKETIDE ANTIBIOTICS AND METHOD OF USE

Inventors: Monique Royer, Montpellier (FR); Dean W. Gabriel, Gainesville, FL (US); Roger Frutos, Saint Georges d’Orques (FR); Philippe Rott, Clapiers (FR)

Assignees: Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD), Paris (FR); University of Florida Research Foundation, Inc., Gainesville, FL (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 400 days.

Appl. No.: 10/531,351
PCT Filed: Oct. 17, 2003
PCT No.: PCT/US03/33142
§ 371 (c)(1), (2), (4) Date: Oct. 3, 2005
PCT Pub. No.: WO2004/035760
PCT Pub. Date: Apr. 29, 2004

Prior Publication Data
US 06/0269988 A1 Nov. 30, 2006
Related U.S. Application
Provisional application No. 60/419,463, filed on Oct. 18, 2002.

References Cited
U.S. PATENT DOCUMENTS

FOREIGN PATENT DOCUMENTS

OTHER PUBLICATIONS

Primary Examiner—Hope A Robinson
Attorney, Agent, or Firm—Saliwanchik, Lloyd & Saliwanchik

ABSTRACT

Three gene clusters that together encode albicidin biosynthesis, the complete gene DNA sequences, and the deduced protein sequences for the enzymes and methods for using the DNA sequences are disclosed and discussed as well as methods for plant protection and creating new antibiotics. The novel Albicidin family of antibiotics is disclosed and their structure deduced.

8 Claims, 13 Drawing Sheets
OTHER PUBLICATIONS


* cited by examiner
FIG. 2
<table>
<thead>
<tr>
<th>Protein</th>
<th>motif I</th>
<th>motif II</th>
<th>motif III</th>
<th>motif IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sgl-TcmO</td>
<td>FVDLGGARG</td>
<td>PRADVFIV</td>
<td>ALTPGGAVLV</td>
<td></td>
</tr>
<tr>
<td>Sgl-TcmN</td>
<td>IADLGGDG</td>
<td>TGYDAYLF</td>
<td>IGDDDRALL</td>
<td></td>
</tr>
<tr>
<td>Smy-MdmC</td>
<td>VLEIGTFTG</td>
<td>GAFDIVFV</td>
<td>LVRPGGLVAI</td>
<td></td>
</tr>
<tr>
<td>Mxa-SafC</td>
<td>TLEVGVFTG</td>
<td>GTFDLAFI</td>
<td>LVRPGGLILL</td>
<td></td>
</tr>
<tr>
<td>Ser-EryG</td>
<td>VLDVGEFLG</td>
<td>ETRFDRVTS</td>
<td>VLPKGGLVAI</td>
<td></td>
</tr>
<tr>
<td>Spe-DauK</td>
<td>VLDVGGKG</td>
<td>RKADAILL</td>
<td>ALEPPGRILI</td>
<td></td>
</tr>
<tr>
<td>Sal-DmpM</td>
<td>VVDIGGADG</td>
<td>GGGLDYLV</td>
<td>AMPAHARLL</td>
<td></td>
</tr>
<tr>
<td>Shy-RapM</td>
<td>VLEVGCNMG</td>
<td>VQGDAEL</td>
<td>ALRRGGAALSH</td>
<td></td>
</tr>
<tr>
<td>Sav-AveD</td>
<td>VLDVGCAGS</td>
<td>GSFDAAWA</td>
<td>VLRPGGRLAV</td>
<td></td>
</tr>
<tr>
<td>Sar-CmE</td>
<td>VLDVACGHG</td>
<td>GPYDSLII</td>
<td>ATRPPGRRIGI</td>
<td></td>
</tr>
<tr>
<td>AlbII</td>
<td>VLDVAAGHG</td>
<td>SGYDVILL</td>
<td>ALNDDGVMVIT</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 3**

<table>
<thead>
<tr>
<th>Protein</th>
<th>motif I</th>
<th>motif II</th>
<th>motif III</th>
<th>motif IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sgl-tcmP</td>
<td>VVLHLAGLDSRAFMDVP109 DVDVPDVIELR</td>
<td>139 EDWLDTPV</td>
<td>150 PALVVAEGLTPY</td>
<td></td>
</tr>
<tr>
<td>Sme-PKS</td>
<td>TVLHLGCGLDSRIFRIDPGP109 ELDVDVPVISLR</td>
<td>139 RGWIERLP</td>
<td>150 PTVLVAEGVLPY</td>
<td></td>
</tr>
<tr>
<td>Pmu-tcmP</td>
<td>VVQGLAGLDRFERLGPQ111 DLLPEVINIR</td>
<td>141 TDWMKTVS</td>
<td>152 PULLILEGLVMFF</td>
<td></td>
</tr>
<tr>
<td>Mtu-Omt</td>
<td>TVALAAGLQTSFWRLDVAI113 TVDLPVIDLR</td>
<td>144 YSWMDSVD</td>
<td>155 GVFITAEGLLMY</td>
<td></td>
</tr>
<tr>
<td>Mlo-Hp</td>
<td>TVLHLGCGLDRTRVFRVDP109 DADYPQVIELR</td>
<td>139 PGWLAEPV</td>
<td>150 PANVVAEGTLPY</td>
<td></td>
</tr>
<tr>
<td>Mtu-Hp</td>
<td>QVAILASGLDSRAYLPWPT127 EIDQPKVMEFK</td>
<td>162 ADWPTALQ</td>
<td>178 PTAWLAEGLLYL</td>
<td></td>
</tr>
<tr>
<td>Mtu-Hp2</td>
<td>QVIAASGDLRAWRFLPWP129 ELDQPKVLMEFK</td>
<td>162 QDWPKALQ</td>
<td>178 PCAWLABGLVRL</td>
<td></td>
</tr>
<tr>
<td>Mtu-Hp3</td>
<td>QVIALAGDSRAYLPWP123 ELDRFQVLDFFK</td>
<td>156 DDDPOQALR</td>
<td>172 PASWIAEGLLLYL</td>
<td></td>
</tr>
<tr>
<td>Mtu-Hp4</td>
<td>QAVTVAAGLDCRAYLDPW126 EIDVPKVLEFK</td>
<td>161 TDWPTPLT</td>
<td>177 PASAWVEGLPYY</td>
<td></td>
</tr>
<tr>
<td>Sco-Hp</td>
<td>QVQLAGGMDSAFMDPEW118 EVDTPAPLEFK</td>
<td>153 EDWPSALA</td>
<td>169 PTAWIGEGLLLYL</td>
<td></td>
</tr>
<tr>
<td>AlbVI</td>
<td>QVIALAGGMDAYLDPW124 EIDHMDVLSDK</td>
<td>157 EDWPOALYK</td>
<td>173 ATLNLVEGLLCYL</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 4**
FIG. 5
### FIG. 6

<table>
<thead>
<tr>
<th>Strand</th>
<th>Description</th>
<th>Sequence</th>
<th>Score</th>
<th>Hit Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>XALB1</td>
<td>29 bp downstream from the TGA stop codon of <em>albXVI</em></td>
<td>ACCATTGTGAACGCCCCTTCCGCTTCATAGCGATTTTGATCGCCGC</td>
<td>4.30</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>400 bp downstream from the TAA stop codon of <em>albIV</em></td>
<td>CATGGCTGAGGCGTCGGTACGCTAGCGGACGACCTGCCC</td>
<td>4.13</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>62 bp, 170 bp and 560 bp downstream from the TAG stop codon of <em>albXVI</em></td>
<td>GGGGGGCAGTTGCCCAGCCGGTTTCTGTAAACGTTGGCTGTCCTGTA</td>
<td>3.95</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AACTCTAAAGAGATGATTGATTTAAATTCCTGCCTTTGTACGAGATA</td>
<td>4.42</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TACTTAATATAGATTGCCAGCTTGCTGGTAATGATTTTTCATAT</td>
<td>4.27</td>
<td>53</td>
</tr>
<tr>
<td>XALB3</td>
<td>8065=&gt; GCAAGAAAAGGCCCCTACGCGCCCTTTTTTCTTCA</td>
<td>4.78</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8072=&gt; AAAGCGGAAACGAAAAAGGCCCCTACGCGCCCTTTTTTCCTCCATCGTCGA</td>
<td>3.94</td>
<td>86</td>
<td></td>
</tr>
</tbody>
</table>
FIG. 7A
FIG. 7B
RifA-1  LGRVDVLQPACFAVMVGLAAVWESVGVRPDAVVGHSEQEI
RifA-2  LDQTMYTQGALFAVETALFRLFESWGVRPGLLAGHSIGEL
RifA-3  LDRVDVQQPASFAVMVGLAAVWTSLGVTDPDAVLGHSEQEI
RifB-1  LDRVDVQQPASFAVMVGLAAVWESVGVRPDAVVGHSEQEI
RifE-1  LNQTVFTGAGLFAVESALFLRAESWGVRFVDVVLGHSEQEI
B1mVIII ADDTRAQPALFAVEYALARTLMDWGVRPAAMLGHSLEGV

FIG. 10A

AlbXIII LEDRPRHRIVADTLTGLHAQFGPAIQAHNVAVIGHSVGGY
FenF    TRTMNAPAILTCSVIAYQYMQEIGIKPHFLAGHSLEYG
LipA    PDSRGRQLLAALDYLTGRSSVRGRIDSGRGLVGMGSGGG

FIG. 10B
Biosynthetic Genes and Host Cells for the Synthesis of Polyketide Antibiotics and Method of Use

Cross-Reference to Related Applications

This application is the U.S. national stage application of International Patent Application No. PCT/US2003/33142 filed Oct. 17, 2003, which claims the benefit of U.S. Provisional patent application with Ser. No. 60/419,463, filed Oct. 18, 2002 the disclosures of which are hereby incorporated by reference in their entirety, including all nuclear acid sequences, amino acid sequences, chemical formulations, tables and figures.

Technical Field

The Sequence Listing for this application is labeled “seq-list-replace.txt” which was created on Jun. 26, 2008 and is 323 KB. The entire contents of the sequence listing is incorporated herein by reference in its entirety.

The invention is in the field of genetic engineering, and in particular the isolation and expression of the biosynthetic genes that produce a family of antibiotics known generically as albicidins.

Background of the Invention

U.S. Pat. No. 4,525,354 to Birch and Patil described a “non-peptide” antibiotic of M. W. “about 842” called “albicidin.” Albicidin is described as produced by cultivating chlorosis-inducing strains of Xanthomonas albilineans isolated from diseased sugarcane, and mutants thereof. The antibiotic was isolated from the culture medium by adsorption on resin and was purified by gel filtration and High Performance Liquid Chromatography (HPLC). The chemical structure of this antibiotic was not determined and remained unknown, although the Birch and Patil patent disclosed spectral data for a fraction having antibiotic activity and the presence of approximately 38 carbon atoms and at least one COOH group.

Xanthomonas albilineans is a systemic, xylem-invading pathogen that causes leaf scald disease of sugarcane (interspecific hybrids of Saccharum species) (Ricaud and Ryan, 1989; Rott and Davis, 2000). Leaf scald symptoms include chlorosis, necrosis, rapid wilting, and plant death. Chlorosis-inducing strains of the pathogen produce several toxic compounds. The major toxic component, named albicidin, inhibits chloroplast DNA replication, resulting in blocked chloroplast differentiation and chlorotic leaf streaks that are characteristic of the plant disease (Birch and Patil, 1983, 1985b, 1987a and 1987b). Several studies established that albicidin plays a key role in pathogenesis and especially in the development of disease symptoms (Wall and Birch, 1997; Zhang and Birch, 1997; Zhang et al., 1999; Birch, 2001).

The prior art indicates that albicidin inhibits prokaryotic DNA replication and is bactericidal to a range of gram-positive and gram-negative bacteria (Birch and Patil, 1985a). Albicidin is therefore of interest as a potential clinical antibiotic (Birch and Patil, 1985a). However, low yield of toxin production in X. albilineans has slowed down studies into the chemical structure of albicidin and its therapeutic application (Zhang et al., 1998). The chemical structure of this albicidin remains unknown, however this albicidin has been partially characterized as a non-peptide antibiotic with a molecular weight of about 842 that contains approximately 38 carbon atoms with three or four aromatic rings, at least one COOH group, two OCH3 groups, a trisubstituted double bond and a CN linkage (Birch and Patil, 1985a; Huang et al., 2001).

Molecular cloning and characterization of the genes governing the biosynthesis of albicidin is of considerable interest because such information provides approaches to engineer overproduction of albicidin, to characterize its chemical structure, to allow therapeutic applications and to clarify the relationship between toxin production and the ability to colonize sugarcane. Two similar mutagenesis and complementation studies have been conducted to identify the genetic basis of albicidin production in X. albilineans strains isolated in two different geographical locations, Australia and Florida.

One study of X. albilineans strain LS155 from Australia revealed that genes for albicidin biosynthesis and resistance span at least 69 kb (Wall and Birch, 1997). Subsequently, three genes required for albicidin biosynthesis were identified, cloned and sequenced from two Australian strains of X. albilineans (LS155 and Xa13): xabA, xabB and xabC (Huang et al., 2001; Huang et al., 2000a; 2000b). The xabB gene encodes a large protein with a predicted size of 528.5 kDa with a modular architecture indicative of a multi functional polyketide synthase (PKS) linked to a nonribosomal peptide synthetase (NRPS) (Huang et al., 2001). The xabC gene, located immediately downstream from xabB, encodes an S-adenosyl-L-methionine (SAM)-dependent O-methyltransferase (Huang et al., 2000a). The xabA gene, located in another region of the genome, encodes a phosphopantetheinytransferase required for post-translational activation of PKS and NRPS enzymes (Huang et al., 2000b).

These first results demonstrated that the albicidin biosynthesis apparatus is a PKS and/or NRPS system. Such systems assemble simple acyl-coenzyme A or amino acid monomers to produce polyketides and/or nonribosomal peptides (Mahtiel et al., 1997; Cone, 1997; Cone and Walsh, 1999). These metabolites form very large classes of natural products that include numerous important pharmaceuticals, agrochemicals, and veterinary agents such as antibiotics, immunosuppressants, anti-cholesterolemic, as well as antitumor, antifungal and antiparasitic agents. Genetic studies of prokaryotic PKS and NRPS produced detailed information regarding the function and the organization of genes responsible for the biosynthesis of polyketides and nonribosomal peptides. Such knowledge, in turn, made it possible to produce combinations of PKS and NRPS genes from different microorganisms in order to produce novel antibiotics (McDaniel et al., 1999; Rodriguez and McDaniel, 2001; Peiffer et al., 2001). Investigating the complete albicidin biosynthesis apparatus is therefore of great interest because such results may contribute to the knowledge as to how PKS and NRPS interact and how they might be manipulated to engineer novel molecules.

A second study with X. albilineans strain Xa23R1 from Florida revealed that at least two gene clusters, one spanning more than 48 kb, are involved in albicidin production (Rott et al., 1996). This conclusion was based on the following data: (i) fifty Xa23R1 mutants defective in albicidin production were isolated; (ii) a Xa23R1 genomic library of 845 clones, designated pALB1 to pALB845, was constructed; (iii) two overlapping DNA inserts of approximately 47 kb and 41 kb, from clones pALB540 and pALB571 respectively, complemented forty-five mutants and were supposed to contain a major gene cluster involved in albicidin production; (iv) a DNA insert of approximately 36 kb, from clone pALB639, complemented four of the five remaining mutants not complemented by pALB540 and pALB571, and was supposed to contain a second region involved in albicidin pro-
The present invention describes and characterizes the family of antibiotics that is produced by the culture of X. albilineans. The present invention provides a method for producing the unique and previously uncharacterized family of antibiotics produced by X. albilineans and which includes a putative repressor of the putative antibiotic efflux pump, which was subjected to Genbank by Bostock and Birch (Accession No. AF403570).

SUMMARY OF THE INVENTION

The present invention describes and characterizes the family of antibiotics that is produced by culturing strains of X. albilineans and mutants thereof, together with the complete set of twenty biosynthetic genes capable of producing the unique and previously uncharacterized family of antibiotics produced by X. albilineans and previously lumped together as “albidins.” The set of twenty biosynthetic genes isolated, purified and cloned from a culture of X. albilineans revealed that this set of biosynthetic genes is capable of synthesizing products exhibiting a high level of variation among the products, indicating that albidins comprise a family of polyketide antibiotics. The albidins described in the present invention are synthesized by twenty genes, including one polyketide-peptide synthase, one polyketide synthase and two peptide synthases, but the substrates of the polyketide-peptide synthase and one peptide synthase are not α-amino acids. The biosynthetic enzymes represent a previously undescribed and unique polyketide antibiotic biosynthetic system.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a Physical Map and genetic organization of the DNA Region containing the major gene cluster XALB1 involved in the biosynthesis of Albidins.

FIG. 2 is an illustration of the organization of the four PKS modules and the seven NRPS modules identified in cluster XALB1 and comparison with the organization of the prior art material XabB.

FIG. 3 shows the sequence motifs in O-methyltransferases and C-methyltransferases involved in antibiotic biosynthesis in bacteria and in AlbII.

FIG. 4 shows the conserved sequence motifs in O-methyltransferases and in different temp-like hypothetical proteins and AlbVI.

FIG. 5 is an illustration of the alignment of the primary sequences between the conserved motifs A4 and A5 of Alb NPRSs and PKS-4 in Xanthomonas albilineans with the corresponding sequences of GnsA (Phe) accession number: P14607 (SEQ ID NO: 132 and DlmN NPRS-2 (G-As)- accession number AF210249 (SEQ ID NO: 133); Alb IV NPRS-1 (SEQ ID NO: 134); Alb NPRS-3 (SEQ ID NO: 135); AlbIX NPRS-6 (SEQ ID NO: 136); AlbIX NPRS-7 (SEQ ID NO: 137); AlbIV NPRS-5 (SEQ ID NO: 138); AlbVII PKS-4 (SEQ ID NO: 139); Alb NPRS-2 (SEQ ID NO: 140).

FIG. 6 shows Rho-independent transcription terminators identified in the intergenic regions of XALB1 and XALB3 clusters (SEQ ID NO: 141, XALB1 Strand+(29 bp downstream from the TGA stop codon of albXVII); SEQ ID NO: 142, XALB1 Strand+(400 bp downstream from the TAA stop codon of albIV); SEQ ID NOs: 143, 144 and 145, XALB1 Strand+(62 bp, 170 bp and 560 bp downstream from the TAG stop codon of albXVII); SEQ ID NOs: 146 and 147, XALB3 Start).
FIG. 12 illustrates subcloning of operons 3 and 4 (from pALB540), XALB2 (from pAC389.1) and XALB3 (from pEV639) into a single plasmid, pOp3-4/XALB2-3. A BamHI-PstI fragment from pALB540, corresponding to a portion of operon 4, was subcloned into pBSKS(+) yielding pS/Op4DXhol (step 2). The BcI fragment from pAC389.1 (XALB2) is then subcloned into pS/Op4DXhol, yielding pS/Op4DX/XALB2 (step 3). A BfrI fragment from pALB540 containing complete operon 3 and the beginning of operon 4 was subcloned into pS/Op4DX/XALB2, yielding pS/Op3-4/XALB2 (step 4). The Sall fragment from pEV639 (XALB3) was subcloned into pBSKS, yielding pBSKS/XALB3 (step 5). The Sall site located on the KpnI site of the polylinker was then destroyed and substituted by a Xhol restriction site, yielding pBSKs/XALB3Xhol (step 6). Finally, the Xhol cassette of pS/Op3-4/XALB2 was subcloned into the Sall restriction site of pBSKS/XALB3Xhol, yielding pBSKS/Op3-4/XALB2-3 (step 7). An Xhol site was added to the BamHI site of pLAfr3, yielding pLAfr3Xhol (step 8). The Xhol cassette of pBSKS/Op3-4/XALB2-3 was then cloned into pLAfr3Xhol, yielding pOp3-4/XALB2-3 (step 9).

**DETAILED DESCRIPTION OF THE INVENTION**

The invention results from the DNA sequencing of the complete major gene cluster XALB1, as well as from the noncontiguous fragments XALB2 and XALB3. XALB1 is present in the two overlapping DNA inserts of clones pALB540 and pALB571. Reading frame analysis and homology analyses allow one to predict the genetic organization of XALB1 and to assign a function to the genes potentially required for albicidin production. Based on the alignment of the different PKS and/or NRPS enzymes encoded by XALB1 we proposed a model for the albicidin backbone biosynthesis. However, the invention disclosed herein does not depend upon the accuracy of the proposed model. The invention includes the successful cloning and DNA sequencing of the second region of the genome (XALB2) involved in albicidin production and mutated in mutant AM37.

The invention includes the characterization of the third region of the genome (XALB3) involved in albicidin production present in clone pALB639. These results allowed the possibility to characterize all enzymes of the albicidin biosynthesis pathway including structural, resistance and regulatory elements and to engineer overproduction of albicidin.

The subject invention provides:

(a) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25;

(b) isolated, recombinant, and/or purified polynucleotide sequences comprising a polypeptide sequence encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47;

(c) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence that is complementary to a polynucleotide sequence selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25;

(d) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence that is complementary to a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47; or

(e) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide that is at least 70% homologous to: (1) a polynucleotide selected from the group consisting of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 25; (2) a polynucleotide sequence encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47; (3) a polynucleotide that is complementary to a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47; (f) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence encoding variant (e.g., a variant polypeptide) of a polypeptide selected from the group consisting of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47, wherein said variant has at least of the biological activities associated with the polypeptides of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47;

(g) isolated, recombinant, and/or purified polynucleotide sequences comprising polynucleotide sequence encoding a fragment of a polypeptide selected from the group consisting of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47 or a fragment of a variant polypeptide of SEQ ID NOs: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, and 47;

(h) isolated, recombinant, and/or purified polynucleotide sequences comprising a polynucleotide sequence encoding multimeric construct;

(i) a genetic construct comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

(j) a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

(k) a host cell comprising a vector a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

(l) a transformed plant cell comprising a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

(m) a transformed plant comprising a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

(n) a polynucleotide that hybridizes under low, intermediate or high stringency with a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), or (h).

"Polynucleotide sequence", "polynucleotide" or "nucleic acid" can be used interchangeably and are understood to mean, according to the present invention, either a double-stranded DNA, a single-stranded DNA or products of transcription of the said DNAs (e.g., RNA molecules). It should also be understood that the present invention does not relate to genomic polynucleotide sequences in their natural environment or natural state. The nucleic acid, polynucleotide, or polynucleotide sequences of the invention can be isolated, purified (or partially purified), by separation methods including, but
not limited to, ion-exchange chromatography, molecular size exclusion chromatography, or by genetic engineering methods such as amplification, subtractive hybridization, cloning, subcloning or chemical synthesis, or combinations of these genetic engineering methods.

A homologous polynucleotide or polypeptide sequence, for the purposes of the present invention, encompasses a sequence having a percentage identity with the polynucleotide or polypeptide sequences, set forth herein, of at least (or at least about) 70.00% to 99.99% (inclusive). The aforementioned range of percent identity is to be taken as including, and providing written description and support for, any fractional percentage, in intervals of 0.01%, between 20.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two nucleic acid sequences can be distributed randomly and over the entire sequence length. For example, homologous sequences can exhibit a percent identity of 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent with the sequences of the instant invention. Typically, the percent identity is calculated with reference to the polynucleotide of a particular SEQ ID NO.; the full-length of a selected polynucleotide, or the native (naturally occurring) polynucleotide. The terms “identical” or percent “identity”, in the context of two or more polynucleotide or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using a sequence comparison algorithm or by manual alignment and visual inspection.

A “complementary” polynucleotide sequence, as used herein, generally refers to a sequence arising from the hydrogen bonding between a particular purine and a particular pyrimidine in double-stranded nucleic acid molecules (DNA-DNA, DNA-RNA, or RNA-RNA). The major specific pairings are guanine with cytosine and adenine with thymine or uracil. A “complementary” polynucleotide sequence may also be referred to as an “antisense” polynucleotide sequence or an “antitRNA” sequence.

Sequence homology and sequence identity can also be determined by hybridization studies under high stringency, intermediate stringency, and/or low stringency. Various degrees of stringency of hybridization can be employed. The more severe the conditions, the greater the complementarity that is required for duplex formation. Severity of conditions can be controlled by temperature, probe concentration, probe length, ionic strength, time, and the like. Preferably, hybridization is conducted under medium, or high stringency conditions by techniques well known in the art, as described, for example, in Keller, G. H., M. M. Manak [1987] DNA Probes, Stockton Press, New York, N.Y., pp. 169-170.

It is also well known in the art that restriction enzymes can be used to obtain functional fragments of the subject DNA sequences. For example, BglI can be conveniently used for detection of the desired fragment of DNA (commonly referred to as “enzyme-a-base” procedures). See, for example, Maniatis et al. [1982] Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York; Wei et al. [1983] J. Biol. Chem. 258:13006-13512.

The present invention further comprises fragments of the polynucleotide sequences of the instant invention. Representative fragments of the polynucleotide sequences according to the invention will be understood to mean any nucleotide fragment having at least 5 successive nucleotides, preferably at least 12 successive nucleotides, and still more preferably at least 15 or at least 20 successive nucleotides of the sequence from which it is derived. The upper limit for such fragments is the total number of nucleotides found in the full-length sequence encoding a particular polypeptide (e.g., a polypeptide selected from the group consisting of SEQ ID NOs: 26-50). The term “successive” can be interchanged with the term “consecutive”. In some embodiments, a polynucleotide fragment may be referred to as “a contiguous span of at least X nucleotides, wherein X is an integer value beginning with 5. The upper limit for polynucleotide fragments of the subject invention is the total number of nucleotides found in the full-length sequence of a particular SEQ ID or the total number of nucleotides encoding a particular polypeptide (e.g., a particular SEQ ID NO).

In some embodiments, the subject invention includes those fragments capable of hybridizing under various conditions of stringency conditions (e.g., high or intermediate or low stringency) with a nucleotide sequence according to the invention; fragments that hybridize with a nucleotide sequence of the subject invention can be, optionally, labeled as set forth below.

Thus, the subject invention also provides detection probes (e.g., fragments of the disclosed polynucleotide sequences) for hybridization with a target sequence or the amplicon generated from the target sequence. Such a detection probe will comprise a contiguous/consecutive span of at least 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides. Labeled probes or primers are labeled with a radioactive compound or with another type of label as set forth above. Alternatively, non-labeled nucleotide sequences may be used directly as probes or primers; however, the sequences are generally labeled with a radioactive element (32P, 35S, 3H, 125I) or with a molecule such as biotin, acetylaminofluorene, digoxigenin, 5-bromo-deoxyuridine, or fluorescein to provide probes that can be used in numerous applications.

The subject invention also provides for modified nucleotide sequences. Modified nucleic acid sequences will be understood to mean any nucleotide sequence that has been modified, according to techniques well known to persons skilled in the art, and exhibiting modifications in relation to the native, naturally occurring nucleotide sequences.

The subject invention also provides genetic constructs comprising: a) a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NOs: 1-25; b) a polynucleotide sequence having at least about 70% to 99.99% identity to a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NO: 26-50, wherein said polynucleotide encodes a polypeptide having at least one of the biological activities of the polypeptides (e.g., a catalytic activity at set forth in Table 4); c) a polynucleotide sequence encoding a polypeptide selected from the group consisting of SEQ ID NO: 26-50, wherein said polypeptide has at least one of the biological activities of the polypeptides (e.g., a catalytic activity, or transport activity at set forth in Table 4); d) a polynucleotide sequence comprising SEQ ID NO: 1, 2, 3, or combinations thereof; e) a polynucleotide sequence encoding variant (e.g., a variant polypeptide) of a polypeptide selected from the group consisting of SEQ ID NO: 26-50, wherein said variant has at least one of the biological activities associated with the polypeptides (e.g., a catalytic or transport activity as set forth in Table 4); f) a polynucleotide sequence encoding a fragment of a variant polypeptide as set forth in (e); or g) a polynucleotide sequence encoding multimeric construct.
Genetic constructs of the subject invention can also contain additional regulatory elements such as promoters and enhancers and, optionally, selectable markers. Also within the scope of the subject instant invention are vectors or expression cassettes containing polynucleotides encoding the polypeptides, set forth supra, operably linked to regulatory elements. The vectors and expression cassettes may contain additional transcriptional control sequences as well. The vectors and expression cassettes may further comprise selectable markers. The expression cassette may contain at least one additional gene, operably linked to control elements, to be co-transformed into the organism. Alternatively, the additional gene(s) and control element(s) can be provided on multiple expression cassettes. Such expression cassettes are provided with a plurality of restriction sites for insertion of the sequences of the invention to be under the transcriptional regulation of the regulatory regions. The expression cassette(s) may additionally contain selectable marker genes operably linked to control elements.

In some embodiments, the expression cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of the invention, and a transcriptional and translational termination region functional in plants. The transcriptional initiation region, the promoter, may be native or analogous, or foreign or heterologous, to the plant host. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. By "foreign" is intended that the transcriptional initiation region is not found in the native plant into which the transcriptional initiation region is introduced. As used herein, a chimeric gene comprises a coding sequence operably linked to a transcriptional initiation region that is heterologous to the coding sequence.

The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of A. tumefaciens, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau et al. (1991) Mol. Gen. Genet. 262:141-144; Proudfoot (1991) Cell 64:671-674; Sanfieto et al. (1991) Genes Dev. 5:141-149; Mogen et al. (1990) Plant Cell 2:1261-1272; Munroe et al. (1990) Gene 91:151-158; Ballas et al. (1989) Nucleic Acids Res. 17:7891-7903; and Joshi et al. (1987) Nucleic Acid Res. 15:9627-9639.

Where appropriate, the polynucleotides encoding the polypeptides set forth supra can be optimized for expression in the transformed plant. That is, the genes can be synthesized using plant-preferred codons corresponding to the plant of interest. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Pat. Nos. 5,380, 831 and 5,436,391, and Murray et al. (1989) Nucleic Acids Res. 17:477-498, herein incorporated by reference.

The expression cassettes may additionally contain 5' leader sequences in the expression construct cassette. Such leader sequences can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region), Elroy-Stein et al. (1989) PNAS USA 86:6126-6130; potyvirus leaders, for example, TEV leader (Tobacco Etch Virus), Allison et al. (1986); MDMV Leader (Maize Dwarf Mosaic Virus), Virology 154:9-20; human immunoglobulin heavy-chain binding protein (Bip), Macek et al. (1991) Nature 353:90-94; untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), Jobling et al. (1987) Nature 325:622-625; tobacco mosaic virus leader (TMV), Gallie et al. (1989) in Molecular Biology of RNA, ed. Cech (Liss, New York), pp. 237-256; and maize chlorotic mottle virus leader (MCMV), Lommel et al. (1991) Virology 81:382-385. See also, Della-Cioppa et al. (1987) Plant Physiol. 84:965-968. Other methods known to enhance translation can also be utilized.

Also provided are transformed host cells, transformed plant cells and transgenic plants which contain one or more genetic constructs, vectors, or expression cassettes comprising polynucleotides of the subject invention, or biologically active fragments thereof, operably linked to control elements. As used herein, the term “plants” includes algae and higher plants. Thus, algae, monocots, and dicots may be transformed with genetic constructs of the invention, expression cassettes, or vectors according to the invention. In certain embodiments of the subject invention, the transformed cells or transgenic plants comprise at least one polynucleotide sequence selected from the group consisting of SEQ ID NOs: 1-25. In certain preferred embodiments, transformed cells or transgenic plants comprise at least one polynucleotide sequence comprising SEQ ID NOs: 1, 2, or 3. Optionally, the transformed cells or transgenic plants can comprise at least two or all three polynucleotide sequences selected from the group consisting of SEQ ID NOs: 1, 2, and 3.

Methods of transforming cells with genetic constructs, vectors, or expression cassettes comprising the novel polynucleotides of the invention are also provided. These methods comprise transforming a plant or plant cell with a polynucleotide according to the subject invention. Plants and plant cells may be transformed by electroporation, Agrobacterium transformation (including vacuum infiltration), engineered plant virus replication, electrophoresis, microinjection, microprojectile bombardment, vacuum infiltration of Agrobacterium, micro-LASER bean-induced perforation of cell wall, or simply by incubation with or without polyethylene glycol (PEG). Plants transformed with a genetic construct of the invention may be produced by standard techniques known in the art for the genetic manipulation of plants. DNA can be transformed into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by Agrobacterium exploiting its natural gene transferability. Agrobacterium transformation is used by those skilled in the art to transform algae and dicotyledonous species. Substantial progress has been made towards the routine production of stable, fertile transgenic plants in almost all economically relevant monocot plants. In particular, Agrobacterium mediated transformation has now emerged as a highly efficient transformation method in monocots. Microprojectile bombardment, electroporation, and direct DNA uptake are preferred where Agrobacterium is inefficient or ineffective. Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, e.g., bombardment with Agrobacterium-coated microprojectiles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with Agrobacterium (EP-A-486233).

Following transformation, a plant may be regenerated, e.g., from single cells, callus tissue or leaf discs, as is standard in the art. Almost any plant can be entirely regenerated from cells, tissues, and organs of the plant. Available techniques are reviewed in Vasil et al. (1984) in Cell Culture and Somatic Cell Genetics of Plants, Vols. I, II, and III, Laboratory Procedures and Their Applications (Academic press); and Weisbech et al. (1989) Methods for Plant Mol. Biol.

The transformed plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that expression of the
desired phenotypic characteristic is stably maintained and inherited, and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved.

The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practicing the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce nucleic acid into plant cells is not essential to or a limitation of the invention, nor is the choice of technique for plant regeneration.

Also according to the invention, there is provided a plant cell having the constructs of the invention. A further aspect of the present invention provides a method of making such a plant cell involving introduction of a vector including the construct into a plant cell. For integration of the construct into the plant genome, such introduction will be followed by recombination between the vector and the plant cell genome to introduce the sequence of nucleotides into the genome. RNA encoded by the introduced nucleic acid construct may then be transcribed in the cell and descendants thereof, including cells in plants regenerated from transformed material. A gene stably incorporated into the genome of a plant is passed from generation to generation to descendants of the plant, so such descendants should show the desired phenotype.

The present invention also provides a plant comprising a plant cell as disclosed. Transformed seeds and plant parts are also encompassed. As used herein, the expressions “cell,” “cell line,” and “cell culture” are used interchangeably and all such designations include progeny. Thus, the words “transformation” and “transformed cells” include the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to naturally occurring, deliberate, or inadvertent caused mutations. Mutant progeny that have the same function or biological activity as screened for in the originally transformed cell are included. Where distinct designations are intended, it will be clear from the context.

In addition to a plant, the present invention provides any clone of such a plant, seed, or hybrid descendants, and any part of any of these, such as cuttings or seed. The invention provides any plant propagule that is any part which may be used in reproduction or propagation, sexual or asexual, including cuttings, seed, and so on. Also encompassed by the invention is a plant which is a sexually or asexually propagated off-spring, clone, or descendant of such a plant; or any part or propagule of said plant, off-spring, clone, or descendant. Plant extracts and derivatives are also provided.

As is apparent to the routine in this technology, the disclosed methods allow for the expression of a gene of interest in any plant. The invention thus relates generally to methods for the production of transgenic plants (both monocots and dicots). As used herein, the term “transgenic plants” refers to plants (algae, monocots, or dicots), comprising plant cells in which homologous or heterologous polynucleotides are expressed as the result of manipulation by the hand of man. As is apparent to one of ordinary skill in the art, the peptides encoded by the disclosed herein may be encoded by multiple polynucleotide sequences because of the redundancy of the genetic code. It is well within the skill of a person trained in the art to create these alternative DNA sequences encoding the same, or essentially the same, amino acid sequences. These variant DNA sequences are within the scope of the subject invention.

The terms “purified” and “isolated”, when referring to a polynucleotide, nucleotide, or nucleic acid, indicate a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecules but is not flanked by both of the coding or non-coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs (e.g., DNA excised with a restriction enzyme); (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of (i) DNA molecules, (ii) transfected cells, and (iii) cell clones, e.g., as these occur in a DNA library such as a cDNA or genomic DNA library.

The term “polynucleotide” as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule and thus includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, “caps”, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications, such as those with uncharged linkages (e.g., methyl phosphonates, phosphothiotesters, phosphorimidates, carbamates, etc.) and with changed linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkyllators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

“Control elements” include both “transcriptional control elements” and “translational control elements.” “Transcriptional control elements” include “promoter”, “enhancer”, and “transcription termination” elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription [Maniatis et al. [1987] Science 236:1237]. Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in plants, yeast, insect and mammalian cells and viruses (analogous control elements, i.e., promoters, are also found in prokaryotes). The selection of a particular promoter and enhancer depends on what cell type is to be used to express the peptide of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types [for review see Voss et al. [1986] Trends Biochem. Sci. 11:287 and Maniatis et al. [1987] supra]. Transcriptional control elements suitable for use in plants are well known in the art. “Translational control elements” include translational initiation regions and translational termination regions functional in plants.

A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. Strong promoters may be used to produce high
levels of gene transcription. Alternatively, inducible promoters may be used to selectively activate gene transcription when the appropriate signal is provided. Constitutive promoters may be utilized to continuously drive gene transcription. Tissue-specific promoters may also be used in the practice of the invention in order to provide localized production of gene transcripts in a desired tissue. Developmental promoters may, likewise, be used to drive transcription of a gene during a particular developmental stage of the plant. Thus, a gene of interest can be combined with constitutive, tissue-specific, inducible, developmental, or other promoters for expression in plants depending upon the desired outcome.


markers, various restriction sites, a potential for high copy number and strong promoters.

By "vector" is meant any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, virus etc., which is capable of replication when associated with the proper control elements and which can transfer gene sequences between cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

During the preparation of the constructs, the various fragments of DNA will often be cloned in an appropriate cloning vector, which allows for amplification of the DNA, modification of the DNA or manipulation of the DNA by joining or removing sequences, linkers, or the like. Preferably, the vectors will be capable of replication to at least a relatively high copy number in E. coli. A number of vectors are readily available for cloning, including such vectors as pBR322, vectors of the PUC series, the M13 series vectors, and pBlueScript vectors (Stratagene; La Jolla, Calif.).

In order to provide a means of selecting transformed plants or plant cells, the vectors for transformation will typically contain selectable marker gene. Marker genes are expressible DNA sequences which express a polypeptide which resists a natural inhibition by, attenuates, or inactivates a selective substance. Examples of such substances include antibiotics and, in the case of plant cells, herbicides. Selectable markers for use in animal, bacterial, plant, fungal, yeast, and insect cells are well known in the art. Exemplary selectable markers include bacterial transposons Tn5 or Tn 601 (903) conferring resistance to aminoglycosides (selection for Geneticin-resistance (G418), mycophenolic acid resistance (MPAR) utilizing E. coli guanosine phosphoribosyl transferase (gpt) encoding the enzyme XGPR; selection is performed on medium containing MPA and xanthine), methotrexate resistance (MTXR), or cadmium-resistance which incorporates the mouse metallothionein gene (as cDNA cassette) on the vector which detoxifies heavy metal ions by chelating them.

Alternatively, a marker gene may be used as some visible indication of cell transformation. For example, it may cause a distinctive appearance or growth pattern relative to plants or plant cells not expressing the selectable marker gene in the presence of some substance, either as applied directly to the plant cells or as present in the plant or plant cell growth media. The use of such a marker for identification of plant cells containing a plastid construct has been described (Svab et al. [1993] supra). Numerous additional promoter regions may also be used to drive expression of the selectable marker gene, including various plant promoters and bacterial promoters which have been shown to function in plants.

A number of other markers have been developed for use with plant cells, such as resistance to chloramphenicol, the aminoglycoside G418, hygromycin, or the like. Other genes which encode a product involved in chloroplast metabolism may also be used as selectable markers. For example, genes which provide resistance to plant herbicides such as glyphosate, bromoxynil or imidazolinone may find particular use. Such genes have been reported (Stalker et al. [1985] J. Biol. Chem. 260:4724-4728 (glyphosate resistant EPSP); Stalker et al. [1985] J. Biol. Chem. 263:6310-6314 (bromoxynil resistant nitriace gene); and Sathasivan et al. [1990] Nucl. Acids Res. 18:2188 (AHAS imidazolinone resistance gene)).

Another aspect of the invention provides vectors for the cloning and/or the expression of a polynucleotide sequences taught herein in prokaryotic or animal cells. The subject invention also provides for the expression of a polypeptide, peptide, derivative, or variant encoded by a polynucleotide sequence disclosed herein comprising the culture of a procarotic or animal cell (a host cell) transformed with a polynucleotide of the subject invention under conditions that allow for the expression of a polypeptide, biologically active fragment, or multimeric construct encoded by said polynucleotide and, optionally, recovering the expressed polypeptide, peptide, derivative, or analog.

In this aspect of the invention, the polynucleotide sequences can be regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host cell transformed with the recombinant DNA molecule. For example, expression of a protein or peptide may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control expression include, but are not limited to, the CMV-3E promoter, the SV40 early promoter region (Benoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' terminal repeat of Rous sarcoma virus (Yamamoto, et al., 1980, Cell 22:787-797), the herpes simplex thymidine kinase promoter (Wagner et al., 1981, Proc. Natl. Acad. Sci. U.S. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster et al., 1982, Nature 296:39-42); prokaryotic vectors containing promoters such as the beta-lactamase promoter (Villa-Kumaroff, et al., 1978, Proc. Natl. Acad. Sci. U.S. 75:3727-3731, or the tac promoter (DeBoer, et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:21-25); see also “Useful proteins from recombinant bacteria” in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella et al., 1983, Nature 303: 209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, et al., 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose bisphosphate carboxylase (Herrera-Estrella et al., 1984, Nature 310:115-120); promoter elements from yeast or fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PCK (phosphoglycerate kinase) promoter, and/or the alkaline phosphatase promoter.

The vectors according to the invention are, for example, vectors of plasmid or viral origin. In a specific embodiment, a vector is used that comprises a promoter operably linked to a nucleic acid sequence encoding a polypeptide as disclosed herein, one or more origins of replication, and, optionally, one or more selectable markers (e.g., an antibiotic resistance gene). Expression vectors comprise regulatory sequences that control gene expression, including gene expression in a desired host cell. Exemplary vectors for the expression of the polypeptides of the invention include the pET-type plasmid vectors (Promega) or pBAD plasmid vectors (Invitrogen) or those provided in the examples below. Furthermore, the vectors according to the invention are useful for transforming host cells so as to clone or express the polynucleotide sequences of the invention.

The invention also encompasses the host cells transformed by a vector according to the invention. These cells may be obtained by introducing into host cells a nucleic acid sequence inserted into a vector as defined above, and then culturing the said cells under conditions allowing the replication and/or expression of the polynucleotide sequences of the subject invention.

The host cell may be chosen from eukaryotic or prokaryotic systems, such as for example bacterial cells, (Gram negative or Gram positive), yeast cells (for example, Saccharomyces cerevisiae or Pichia pastoris), animal cells (such as Chinese hamster ovary (CHO) cells), plant cells (e.g., algae), and/or insect cells using baculovirus vectors. In some embodiments, the host cells for expression of the polypeptides include, and are not limited to, those taught in U.S. Pat. No. 6,319,691, 6,277,375, 5,643,570, or 5,565,335, each of
which is incorporated by reference in its entirety, including all references cited within each respective patent.

Furthermore, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered polypeptide may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (e.g., glycosylation, phosphorylation) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can also to provide glycosylation of a protein.

The subject invention provides one or more isolated polypeptides comprising:

(a) SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47;

(b) a heterologous polypeptide sequence fused, in frame, to a polypeptide comprising SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47;

(c) a fragment of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47, wherein said fragment exhibits at least one biological function of the polypeptide of SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47; or

(d) a variant having at least 70% homology to a polypeptide comprising SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47, wherein said variant exhibits at least one biological function of the polypeptide comprising SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, or 47.

The term “peptide” may be used interchangeably with “oligopeptide” or “polypeptide” in the instant specification to designate a series of residues, typically L-amino acids, connected one to the other, typically by peptide bonds between the a-amino and carboxyl groups of adjacent amino acids. Linker elements can be joined to the polypeptides of the subject invention through peptide bonds or via chemical bonds (e.g., heterobifunctional chemical linker elements).

The subject invention encompasses polypeptide fragments of the full-length polypeptides disclosed herein. Polypeptide fragments, according to the subject invention, usually comprise a contiguous span of at least 5 consecutive (or contiguous) amino acids. The maximum length for a polypeptide fragment in the context of this invention is that is one amino acid less than the full length of a particular SEQ ID NO: from which the fragment was derived. In certain preferred embodiments, fragments of the polypeptides of the subject invention retain at least one biological activity/function of the full-length polypeptide from which they are derived (e.g., such similar or identical enzymatic activity or the ability to provide resistance to an antibiotic or transport an antibiotic out of a cell (see, for example, Table 4).

A “variant” polypeptide (or polypeptide variant) is to be understood to designate polypeptides exhibiting, in relation to the natural polypeptide, certain modifications. These modifications can include a deletion, addition, or substitution of at least one amino acid, a truncation, an extension, a chimeric fusion, a mutation, or polypeptides exhibiting post-translational modifications. Among the homologous polypeptides, those whose amino acid sequences exhibit between at least (or at least about) 70.00% to 99.99% (inclusive) identity to the full length, native, or naturally occurring polypeptide are another aspect of the invention. The aforementioned range of percent identity is to be taken as including, and providing written description and support for, any fractional percentage, in intervals of 0.01%, between 70.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two polypeptide sequences can be distributed randomly and over the entire sequence length. Thus, variant polypeptides can have 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent identity with the polypeptide sequences of the instant invention. In certain preferred embodiments, variants of the polypeptides of the subject invention retain at least one biological activity/function of the full-length polypeptide from which they are derived (e.g., such as similar or identical enzymatic activity or the ability to provide resistance to an antibiotic or transport an antibiotic out of a cell (see, for example, Table 4).

polypeptides to which the biotin element is attached (see, e.g., U.S. Pat. No. 5,478,925 for numerous methods of multimerization). Multimers of the invention may also be generated using chemical or genetic engineering techniques known in the art.

The invention, thus, provides a novel antibiotic family, Albidicins, produced by three novel biosynthetic gene clusters (XALB1, XALB2, and XALB3) contained within a host cell DNA in which one strand comprises non-contiguously SEQ ID No. 1, SEQ ID No. 2 and SEQ ID No. 3, and the cell expresses the DNA to provide peptides including those named AlbI (SEQ ID No. 26) (encoded by SEQ ID No. 20), AlbII (SEQ ID No. 27) (encoded by SEQ ID No. 21), AlbIII (SEQ ID No. 28) (encoded by SEQ ID No. 22), AlbIV (SEQ ID No. 29) (encoded by SEQ ID No. 23), AlbV (SEQ ID No. 31) (encoded by SEQ ID No. 18), AlbVI (SEQ ID No. 32) (encoded by SEQ ID No. 17), AlbVII (SEQ ID No. 33) (encoded by SEQ ID No. 16), AlbVII (SEQ ID No. 34) (encoded by SEQ ID No. 15), AlbX (SEQ ID No. 35) (encoded by SEQ ID No. 10), AlbXI (SEQ ID No. 36) (encoded by SEQ ID No. 9), AlbXII (SEQ ID No. 37) (encoded by SEQ ID No. 8), AlbXIII (SEQ ID No. 38) (encoded by SEQ ID No. 7), AlbXIV (SEQ ID No. 39) (encoded by SEQ ID No. 6), AlbXV (SEQ ID No. 40) (encoded by SEQ ID No. 5), AlbXVI (SEQ ID No. 42) (encoded by SEQ ID No. 11), AlbXVII (SEQ ID No. 43) (encoded by SEQ ID No. 12), AlbXIX (SEQ ID No. 44) (encoded by SEQ ID No. 13), AlbXX (SEQ ID No. 45) (encoded by SEQ ID No. 14), AlbXXI (SEQ ID No. 46) (encoded by SEQ ID No. 24), and AlbXXII (SEQ ID No. 47) (encoded by SEQ ID No. 25), that in turn interact within the host cell to produce one or more antibiotics as more fully illustrated in FIG. 11.

In one embodiment, the invention comprises a plurality of isolated and purified DNA strands which comprise nucleotide sequences selected from the group consisting of SEQ ID No. 1 to SEQ ID No. 25, each individual sequence, except the transposases AlbI (SEQ ID No. 30) (encoded by SEQ ID No. 19) and AlbXVI (SEQ ID No. 41) (encoded by SEQ ID No. 4) found in the XALB1 cluster, being necessary to the biosynthesis of the novel family of antibiotics, Albidicins.

The invention also includes the peptides or proteins encoded by the genes of the biosynthetic complex expressed by the combination of DNA with a strand having sequences SEQ ID Nos. 1 to 3. Proteins are named with roman numerals and the prefix Alb from AlbI to AlbXXII have the amino acid sequences of SEQ ID Nos. 26 to 47 (not in Roman numeral order but in the order of placement of the genes within sequences SEQ ID Nos. 1 to 3 that express each protein). Expression of the peptides having the amino acid sequences of SEQ ID Nos. 26 to 29, 31 to 40 and 42 to 47, have been found to be all required for the successful biosynthesis of Albidicins.

The invention further provides a method for producing Albidicins comprising providing a modified host cell with a heterologous DNA Albidicin Biosynthetic Gene Cluster or set of genes defined as DNA operably comprising DNA sequences substantially similar to SEQ ID Nos. 1 to 3. Substantially the same means DNA having sufficient homology to provide expressed proteins that function to provide an antibiotic material having the structural components identified herein. Preferably a given sequence will have at least 70 percent homology to one of SEQ ID Nos. 1 to 3, preferably 85% homology and most preferably at least 95% homology. The method includes the steps of modifying the DNA of the host cell to comprise an operable expression system for maintaining the modified host cell under conditions supporting biosynthesis of Albidicins and isolation of Albidicins from the host cell or its environment. The invention further provides a method of production of a group of novel antibiotic materials utilizing at least three of the Sequences selected from the group consisting of DNA SEQ ID No. 1 to SEQ ID No. 25 (excluding transposases encoded by SEQ ID Nos. 4 and 19) inclusive in combination with additional sequences to produce a modified Albidicin-like material.

More specifically, the invention provides DNA Sequences comprising at least about 68,498 base pairs and including about 55,839 bp region from the genome of X. abilinates designated as XALB1 (Albidicin Biosynthetic Gene Cluster 1; SEQ ID No. 1) an additional non-contiguous region of about 2,986 bp, XALB2 (Albidicin Biosynthetic Gene Cluster 2; SEQ ID No. 2), and a third region of about 9,673 bp, XALB3 (Albidicin Biosynthetic Gene Cluster 3; SEQ ID No. 3). Albidicin Biosynthetic Gene Clusters 1-3 may be referred to, collectively, as the Albidicin Biosynthetic Gene Clusters and these sequences were found to be required for biosynthesis of Albidicins. Homology analysis revealed the presence of (i) four large genes with a modular architecture characteristic of polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) potentially involved in albidicin precursor biosynthesis; (ii) four smaller genes potentially involved in albidicin substrate biosynthesis (iii) four modifying genes; (iv) one enzyme activating gene, (v) two regulatory genes, (vi) one chaperone gene, (vii) two genes of unknown function; and (viii) two resistance genes. These are named and discussed more fully below. Together these genes allow the successful operation of the biosynthetic pathway when cloned into suitable host cells.

Alignment of individual NRPS and PKS domains revealed an extraordinary biosynthetic apparatus believed to involve a trans-action of separate PKS and NRPS domains which could contribute to the production of multiple, structurally related albidicins by the same gene cluster. Furthermore, analysis of selectivity-conferring residues indicated that four NRPS modules of XALB1 specify an unusual substrate.

In an alternate embodiment the invention provides a method of producing a polyketide carrying para-aminobenzoic acid and/or carbamoyl benzoic acid by inserting at least one DNA fragment that encodes a PKS protein into a cell and causing the cell to express the encoded PKS protein under conditions such that the PKS protein functions to produce a polyketide carrying either a para-aminobenzoic acid or a carbamoyl benzoic acid or both. Another embodiment provides a method of producing polyketide/peptides carrying para-aminobenzoic acid and/or carbamoyl benzoic acid by inserting at least one DNA fragment that encodes a PKS protein into a cell and causing the cell to express the encoded PKS protein under conditions such that the PKS protein functions to produce a polyketide carrying either a para-aminobenzoic acid or a carbamoyl benzoic acid or both. Yet another embodiment, the invention provides a method of activating nonproteinogenic amino acids like paraminobenzoic acid and/or carbamoyl benzoic acid for incorporation into peptides or polyketides by inserting at least one DNA fragment that encodes a PKS protein into a cell and causing the cell to express the encoded PKS protein under conditions such that the PKS protein functions to produce a polyketide carrying either a para-aminobenzoic acid or a carbamoyl benzoic acid or both.

There are three regions of the X. abilinates genome specifying albidicin production. XALB2 and XALB3 regions each contain only one gene, both of which are required for post-translational activation and folding of albidicin PKS and NRPS enzymes. The XALB1, XALB2 and XALB3 gene clusters are characterized by an unusual hybrid NRPS-PKS
system, indicating that albichel biosynthesis may provide an excellent model for investigating the biosynthesis of hybrid polyketide-polyamide metabolites in bacteria. The availability of three genomic regions involved in albichel production, XAL1B1 and XAL1B2 and XAL1B3, also offers the ability to express individually the enzymes of the albichel family biosynthetic pathway including structural, resistance, regulatory and regulatory elements, and to engineer overproduction of albichel in mutared or modified host cells of the invention. The invention overcomes prior art limitations in albichel production due to low yields of toxin production in X. albi-
lineans and may also allow characterization of the chemical structure of albichel as well as application of this potent inhibitor of prokaryote DNA replication.

The invention results from a number of unpredictable results namely the number and complexity of the enzymes involved and the discovery of the complete sequence required for biosynthesis of Albichelines is previously unreported. The invention provides a novel process for production of molecules having a polyketide-polyamide backbone and the formula C₄₀H₁₃₀O₁₁N₁₀, a molecular weight of 839, and the structural elements shown in FIG. 11.

The invention further includes (a) the Albichel Family Biosynthetic Gene Cluster including (b) the structural and regulatory elements of the operons that encode c) the enzymes PKS-1, PKS-2, PKS-3, PKS4, NRPS-1, NRPS-2, NRPS-3, NRPS-4, NRPS-5, NRPS-6 and NRPS-7 as well as (d) the proteins Alb to AlbXXII, (e) the isolated enzymes, proteins, and active forms thereof, as well as mutants, fragments, and fusion proteins comprising any of the foregoing: (f) the uses of the enzymes or proteins encoded by the Albichelines Biosynthesis Gene Cluster or any one of its operons, (g) a host cell expressing one or more enzymes or proteins encoded by the Albichel Family Biosynthetic Gene Cluster; (h) use of host cells having the Albichel Biosynthesis Gene Cluster to produce an antibiotic; (i) methods of modifying the DNA sequences to produce members of a series of antibiotic compounds having structures related to Albichelines; (j) DNA sequences that encode the same proteins as any of SEQ. ID. Nos. 1 to 25 but differ in specific codons due to the multiplicity of codons that lead to expression of the same amino acid; (k) antibiotics produced by the process of expression of the Albichel Family Biosynthetic Genes in a genetically modified host cell sustained in a culture medium and thereafter separation of the antibiotic from the host cell and culture medium; (l) an isolated and purified antibiotic produced by a process that includes at least three proteins coded by DNA sequences selected for the group consisting of SEQ. ID. Nos. 1 to 25 in combination with additional enzymes that modify the product to provide a non-naturally occurring Albichel-like product having at least one of the useful properties reported for albichel. In certain embodiments, the antibiotic or antibiotics have at least one of the general structures illustrated in FIG. 11. In other embodiments, antibiotics of the subject invention have at least 4 of the structural elements illustrated in FIG. 11, and an elemental composition of C₄₀H₁₃₀O₁₁N₁₀.

The invention further provides a method of protecting a plant against damage from albichel which comprises applying an agent that blocks expression at least one gene in the Albichel Biosynthetic Gene Clusters to the plant to be protected. Additional inventions include a method of obtaining agents useful in blocking expression of albichel by screening materials against a modified host cell line that expresses the Albichel Biosynthesis Gene Clusters and selecting for materials that stop or decrease albichel production and a method of protecting a plant against pesticidal damage from an antibiotic that comprises inserting into the plant and operably expressing at least one resistance gene from the Albichel Biosynthesis Gene Clusters into the plant to be protected.

EXAMPLE 1

Materials and Methods

Bacterial strains and plasmids. The source of bacterial strains and their relevant characteristics are described in Table 1.

Media, antibiotics, and culture conditions. X. albinata strains were routinely cultured on modified Wilbrink's (MW) medium at 30°C without benomyl (Rott et al., 1994). For long-term storage, highly turbid distilled water suspensions of X. albinata were supplemented with glycerol to 15% (vol/vol) and frozen at B80°C. For X. albinata, MW medium was supplemented with the following antibiotics as required at the concentrations indicated: kanamycin, 10 or 25 μg/ml; rifampicin, 50 μg/ml. E. coli strains were grown on Luria-Bertani (LB) agar or in LB broth at 37°C and were maintained and stored according to standard protocols (Sambrook et al., 1989). For E. coli, LB medium was supplemented with the following antibiotics as required at the concentrations indicated: kanamycin, 50 μg/ml; ampicillin, 50 μg/ml.

Bacterial conjugation. DNA transfer between E. coli donor (DH5α “pAlb389 or pAC389; Table 1) and rifampicin-resistant X. albinata recipients (X. strains AM10, AM12, AM13, AM35 and AM37, Table 1) was accomplished by triparental conjugation with plasmid pBR327 as the helper as described previously (Rott et al., 1996).

Assay of albichel production. Albichel production was tested by a microbiological assay as described previously (Rott et al., 1996). Rifampicin and kanamycin exconjugants were spotted with sterile toothpicks (2-mm-diameter spots) onto plates of SPA medium (2% sucrose, 0.5% peptone, 1.5% agar) and incubated at 28°C for 2-5 days. The plates were then overlaid with a mixture of E. coli DH5α (10⁶ cells in 2 ml of distilled water) plus 2 ml of molten 1.5% (wt/vol) Noble agar (Difco) at ca. 65°C and examined for inhibition zones after 24 h at 37°C.

Nucleic acid manipulations. Standard molecular techniques were used to manipulate DNA (Sambrook et al., 1989) except for total genomic DNA preparation. Total genomic DNA for Southern blot hybridization was prepared as described by Gabriel and De Feyter (1992).
PCR Conditions. PCR amplifications were performed in an automated thermal cycler PTC-100™ (MJ Research, Inc.). The 25 μl PCR reaction mix consisted of 10 ng of genomic DNA or 1 ng of plasmid DNA, 2.5 μl of 10xPCR buffer without MgCl₂ (Eurobio), 80 μM dNTP mix, 2.5 units of EUробIOTAQUΗ® (Eurobio), 2.5 pmoles of each primer, 2.0 mM MgCl₂ (Eurobio) and sterilized distilled water to final volume. The PCR program was 95°C for 2 min, 25 cycles at 94°C for 1 min, 1 min for 1 min and 72°C for 1 min, with a final 72°C extension for 5 min. Each temperature was determined for each combination of primers and varied between 55°C and 60°C. A 5 μl aliquot of each amplified product was analyzed by electrophoresis through a 1% agarose gel. For sequencing, PCR products were cloned with the pGEM®-T Easy Vector System (Promega).

Oligonucleotide synthesis. Oligonucleotides were purchased from Genome Express (Grenoble or Montreuil, France).

DNA sequencing. Automated DNA sequencing was carried out on double-stranded DNA by the dideoxynucleotide chain termination (Sanger et al., 1977) using a Dye Terminator Cycle Sequencing kit and an ABI Perkin-Elmer sequencer according to the manufacturer’s procedure. Both DNA strands were sequenced with universal primers or with internal primers (20mers). This service was provided by Genome Express (Grenoble, France). Computer-aided sequence analyses were carried out using Sequence Navigator™ (Applied Biosystems, Inc.) and SeqMan (DNASTAR Inc.) programs.

Sequencing analysis. Nucleotide sequences were translated in all six reading frames using EditSeq (DNASTAR Inc.). Potential products of ORFs longer than 100 b were compared to protein databases by the PSI-BLAST program (Swiss-Prot and Genbank) on the NCBI site with ncbi.nlm.nih.gov/) using Altschul program (Altschul et al., 1997). The TERMINATOR program of the Genetics Computer Group was used to identify putative Rho-independent transcription terminators.

Procedures

EXAMPLE 2

Sequencing of the Double Strand Region of 55,839 Bp from X. albilineans Containing XALB1 SEQ ID NO. 1

In Fig. 1 is presented a physical map and genetic organization of XALB1. In the figure, E and K are restriction endonuclease sites for EcoRI and KpnI respectively. Rectangular boxes represent DNA fragments labeled A through N. The numbers below each rectangular box are the number of Tn5-gus insertion sites previously located in each DNA fragment (Rott et al., 1996). The DNA inserts carried by plasmids pALB571 and pALB540 are represented by bold bars above the physical map. The location and direction of putative ORFs identified in the XALB1 gene cluster are shown by arrows. Precise positions and proposed functions for individual ORFs are summarized in Tables 2 and 3, respectively. Position of insertional sites of eight albicin-defective mutants determined by sequencing are indicated by vertical arrows. The location and direction of putative ORFs identified in the XALB1 gene cluster are shown by arrow shapes. These twenty putative ORFs are potentially organized in four or five operons, as indicated at the bottom of the figure. Patterns indicate NRPS and PKS genes (diagonal crosshatch), methyl transferase and esterase genes (hollow rectangles), carbamoyl transferase gene (line crosshatch), benzoate-derived products biosynthesis genes (white), regulatory genes (vertical lined), resistance genes (diagonal lines) and other genes with function of unknown significance to albicin production (black), and three insertional sites of eight albicin-defective mutants determined by sequencing are indicated by vertical arrows. Dotted regions in the physical map and in ORFs represent the two internal duplicated DNA regions of XALB1.

The sequence illustrated in Fig. 1 was generated as follows. The sources of DNA are set out in Table 1. DNA fragments F, E, B, C, L and G, generated by the digestion of cosmid pALB571 (Rott et al., 1996) with EcoRI and/or KpnI, were subcloned into pBCKS (+) and were sequenced from the resulting subclones, pBC/F, pBC/E, pBC/B, pBC/C, pBC/L and pBC/G. DNA fragment D which corresponds to the part of fragment D present in cosmid pALB571 was sequenced from plasmid pUF1043/D following self-ligation of the complete EcoRI digested cosmid pALB571. DNA fragment H was sequenced from pAM45.1 (Rott et al., 1996), obtained following cloning into vector pBR325 of the 12 kb EcoRI fragment carrying Tn5 and flanking sequences from mutant strain XaAM45. DNA fragment A contains the part of fragment A present in cosmid pALB571 and was subcloned into vector pBCKS (+) and the resulting plasmid pBC/A was used for sequencing. The presence of a large internal duplication made alignment of sequence data obtained from pBC/A difficult. This difficulty was resolved using sequence data obtained from an additional plasmid, pAM4, obtained following cloning into vector pBluescript II KS (+) of the 12 kb EcoRI fragment carrying Tn5 and flanking sequences from mutant strain XaAM4, which contains only one copy of the large internal duplication. Sequence data from pBC/A were used to determine the first 1542 bp of fragment A between nucleotides C-19001 and G-20543. Sequence data from pAM4 and pBC/A were used to determine the last 4823 bp of fragment A between nucleotides G-21653 and G-26477. The overlapping region between nucleotides G-20469 and C-22159 was amplified by PCR from cosmid pALB571 using primers contig13-1160 (5’gcggacatgctggagattca3’) SEQ ID NO. 48 and pAM4-14 (5’gctgaagccgagatctg3’) SEQ ID NO. 49, and was sequenced. Resulting sequence data were used to complete sequencing of DNA fragment A. The junctions A/F, F/H, H/E, E/B, B/C, C/I, I/G, G/D between corresponding DNA fragments were sequenced directly from cosmid pALB571. EcoRI DNA fragment containing fragments A and F was subcloned from pALB540 into pBCKS (+), and the resulting plasmid pBC/AF was used to determine the part of DNA fragment A which was not present in cosmid pALB571 between nucleotides G-13682 and G-19001. EcoRI DNA fragments J, K, L, N were subcloned from pALB540 into pBC(+) and were sequenced from resulting plasmid pBC/J, pBC/K, pBC/L, and pBC/N. The junctions J/K, K/J, J/A between corresponding DNA fragments were sequenced directly from cosmid pALB540. DNA region between nucleotides G-7517 and T-8721 was amplified by PCR from cosmid pALB540 using primers E-14 (5’ccgacgttcagctggagtt3’) SEQ ID NO. 50 and F-14-380 (5’ccgaggatggcagacgt3’) SEQ ID NO. 51 and was sequenced. Resulting sequence data were used to determine the sequence of fragment M and of junctions N/M and M/L.
The nucleotide sequence of 55,839 bp containing the entire major gene cluster involved in Albicidin production was sequenced on both strands.

EXAMPLE 3

Analysis of the Large Internal Duplications in the DNA Sequence of XALB1

The sequence of the 55,839 bp genomic region (SEQ ID NO. 1) contains two large internal duplications as shown by the dotted regions in the physical map of FIG. 1. A direct duplication of 1736 bp was located in DNA fragment A between nucleotides G-19904 and G-21639 and between nucleotides G-23057 and G-24792. Another direct duplication of a 2727 bp was found in DNA fragments B and C between nucleotides C-40410 and G-43136 and between nucleotides C-46644 and C-49770. Comparison of the two copies of each duplication revealed that the two copies of the 1736 bp duplication are identical for one nucleotide at position 21058, and that the two copies of the 2727 bp duplication are 98.8% identical and differ by 30 nucleotides.

EXAMPLE 4

Comparison of XALB1 with the xabB EcoRI Fragment

Comparison of the DNA sequence of the 55,839 bp genomic region described in this study with the partial DNA sequence of 16,511 bp of the same region in Huang et al., 2001 (described by Huang et al. as an EcoRI fragment including full length xabB from X. albilineans strain Xa13 [GenBank accession No. AF239749]), revealed that the DNA sequence from strain Xa13 over 16,511 bp is identical to the sequence from strain Xa23R1, described herein, with the following exceptions: 1) five nucleotides are different at positions 42963, 42972, 42980, 43014 and 43071 of the XALB1 sequence, and 2) nucleotides from positions 43137 to 49770 are missing (internal to xabB; refer FIG. 1). Analysis of genomic DNA of seven strains isolated from different countries (Australia, Reunion Island, Kenya, Zimbabawe and USA), digested by KpnI and hybridized with the pB/C plasmid (Table 1) labeled with 32P, revealed that two DNA fragments corresponding to the XALB1 fragments B and C were present in all strains (data not shown). This result indicated that all studied strains contain xabB and not xabB because in xabB the pB/C plasmid probe hybridizes with the large internal duplication present in both DNA fragments B and C (FIG. 1). Based on this observation we postulated that the DNA sequence of XabB reported as full length by Birch in PCTWO/02/24736 A1 (Their seq. ID#1) appears to be incomplete and missing 6,234 bp of DNA sequence encoding 2,078 amino acids.

EXAMPLE 5

Reading Frame Analysis in XALB1

Analysis of the 55,839 bp double strand region for coding sequences revealed the presence of 20 open reading frames (ORFs) designated xabB to xabXX (Table 2 below) which are distributed in four groups of genes according to their position and their orientation in the XALB1 cluster (FIG. 1). Genes of each group may form part of the same operon as judged by their overlapping stop and start codons, or by the relatively short intergenic region which varies from 5 to 274 nucleotides. The 20 ORFs appear to be organized in four operons: xaboper 1 formed by xabI-xabV; xaboper 2 by xabV-xabIX; xaboper 3 by xabX-xabXI; and xaboper 4 by xabXVII-xabXX. The majority of the 16 ORFs are initiated with an ATG codon, except xabIX and xabXXIII which are initiated with a TGG codon, and xabIV and xabV which are initiated with a GTG start codon. In seven ORFs of XALB1, start codons are preceded by the consensus sequence GAGG which may correspond to the ribosome binding site. Other ORFs are preceded by a less conserved sequence which contain at least three nucleotides A or G and which may serve as a weak ribosome binding site.

EXAMPLE 6

Sequencing of the Tn5 Insertional Site of Eight Tox" Mutants Previously Located in XALB1

Eight of the 45 X. albilineans tox" mutants complemented by cosmid pLAB540 and/or cosmid pLAB571 and previously described (Rott et al., 1996) were further analyzed. All eight mutants contain a single Tn5 insertion and correspond to the following X. albilineans strains: XaAM7, XaAM15, XaAM45, and XaAM52 which are complemented by pLAB571 but not by pLAB540; XaAM4, XaAM29 and XaAM40 which are complemented by both cosmids; and XaAM1 which is complemented by pLAB540 but not by pLAB571. The Tn5 insertional site of each tox" mutant was sequenced from plasmids obtained following cloning in pBR325 or in phBluecript II KS (+) of the EcoRI fragments carrying Tn5 and flanking sequence using the sequencing primer GUSN (5'ggccttcgggctagatga'3') SEQ ID No. 52 that annealed 135 bp downstream from the insertional sequence IS50L of Tn5-gusA. The sequence of the Tn5 insertional site was compared with the 55,839 bp sequence containing XALB1 in order to determine the alb gene disrupted in each tox" mutant. xabB is disrupted by the Tn5 insertion in XaAM15 and XaAM45 at position 33443 and 34229, respectively (FIG. 1). xabIV is disrupted by the Tn5 insertion in XaAM7 and XaAM52 at position 53704 and 53915, respectively. xabIX is disrupted by the Tn5 insertion in XaAM4, XaAM29 and XaAM40 at position 21653, 23444 and 24376, respectively. and xabXI is disrupted by the Tn5 insertion in XaAM1 at position 13301. These results are in accordance with the previous characterization of tox" mutants using Southern blot hybridization (Rott et al., 1996), except for XaAM1. The Tn5-gusA insertion site of XaAM1 was previously located in DNA fragment A (Rott et al., 1996) but results of this study showed that this site is located in DNA fragment J (FIG. 1).

EXAMPLE 7

Homology Analysis of Proteins Potentially Encoded by XALB1

Preliminary functional assignments of individual ORFs were made by comparison of the deduced gene products with proteins of known functions in the Genbank database. The results are set out in Table 3 below. Among the ORF's identified from the sequenced XALB1 gene cluster, we found (i) four genes, xabI SEQ ID No. 20, xabIIV SEQ ID No. 23, xabIVI SEQ ID No. 17 and xabIX SEQ ID No. 15, encoding PKS and/or NRPS modules; (ii) one carbamoyl transferase gene, xabXV SEQ ID No. 5; (iii) two esterase genes, xabXI SEQ ID No. 9 and xabXIII SEQ ID No. 7; (iv) two methyltransferase genes, xabII SEQ ID No. 21 and xabVIII SEQ ID No. 18; (v) two benzoyl-derived products biosynthesis genes, xabXVII SEQ
The alb PKS and/or NRPS Genes

The potential product of alb, designated Alb1 SEQ ID No. 20, is a protein of 6879 aa with a predicted size of 755.9 kDa. This protein is very similar to the potential product of the xabB gene from X. albilineans strain Xa13 from Australia (Huang et al., 2001), but it differs in length and size (See Table 4 below). XabB is a protein of 4801 amino acids with a predicted size of 525.7 kDa. Comparison of Alb1 with XabB revealed that the N-terminal region from Met-1 to Ile-4325 of both proteins are identical except for five amino-acids which are Tyr-3941, Pro-3952, Ala-4054, Ala-4271 and Gln-4284 in Alb1 and His-3941, Ala-3952, Val-4054, Val-4271 and Glu-4284 in XabB. The same comparison revealed that the Alb1 C-terminal region from Arg-6404 to the stop codon is 100% identical to the XabB C-terminal region from Arg-4326 to the stop codon.

The N-terminal region (from Met-1 to Asp-3235) of Alb1 is 100% identical to the corresponding region in XabB which was previously described as similar to many microbial modular PKS (Huang et al., 2001). This PKS region may be divided into three modules (FIG. 2). Abbreviations used in the Figure are: A, adenylation; ACP, acyl carrier protein; AL, acyl-CoA ligase; C, condensation; KR, β-ketoacyl reductase; KS, β-ketoacyl synthase; NRPS, nonribosomal peptide synthase; PCP, peptidyl carrier protein; PKS, polyketide synthase; TE, thioesterase; HBCL, 4-hydroxybenzoate-CoA ligase. The question mark in the NRPS-2 domain indicates that this A domain is incomplete. The first module designated PKS-1 contains acyl-CoA ligase (AL) and acyl carrier protein (ACP) domains. The second module designated PKS-2 contains β-ketoacyl synthase (KS1) and β-ketoacyl reductase (KR) domains followed by two consecutive ACP domains (ACP2 and ACP3). The third module designated PKS-3 contains a KS domain (KS2) followed by a PCP domain (PCP1). Apart from their very high similarity with XabB, these three PKS modules exhibited the highest degree of overall similarity with polyketide synthases Sfil and PksM from Myxococcus xanthus and Bacillus subtilis, respectively (Table 4). The motifs characteristic of these domains are 100% identical to those of XabB which were previously aligned with those from other organisms (Huang et al., 2001). The AL domain contains the conserved adenylation core sequence (SSGS) and the ATPase motif (TGD). The three ACP domains contain a 4-phosphopantetheinyl-binding cofactor box GxDSx(L), except that A replaced C in ACP1. Both KS domains contain motif GPxxxxxxxGxxSL around the active site Cys, and two His residues downstream from the active site Cys, in motifs characteristic of these enzymes. The KR domain contains the NAD(P)H-binding site GxGxxG.

The PKS part of Alb is linked by the PCP1 domain to the four apparent nonribosomal peptide synthase modules designated NRPS-1, NRPS-2, NRPS-3 and NRPS-4 (FIG. 2). NRPS-1, NRPS-2 and NRPS-3 modules display the ordered condensation, adenylation (A) and PCP domains typical of such enzymes (Marahiel et al., 1997), and NRPS-4 consists of an extra C domain which may correspond to an incomplete NRPS module. Known conserved sequences, characteristic of the domains commonly found in peptide synthases (Marahiel et al., 1997), were compared to those from NRPS-1, NRPS-2, NRPS-3 and NRPS-4 (Tables 5, 6 and 7). Sequences characteristic of A, or PCP domains are conserved in these four NRPS, except in A domain of NRPS-2 module, suggesting that this latter A domain may not be functional. Comparison of the four NRPS modules among themselves revealed that NRPS-2, NRPS-3 and NRPS-4 modules were 93.7% to 94.4% and 47.5% similar to NRPS-1 module, respectively. Comparison of XabB revealed NRPS-2 and NRPS-3 modules were not present in XabB which contains only NRPS-1 and NRPS-4 modules (FIG. 2). The dotted box in FIG. 2 corresponds to the apparent deletion of the NRPS-2 and NRPS-3 modules in XabB as compared to Alb1. Apart their very high similarity with XabB, Alb1 NRPS modules exhibited the highest degree of overall similarity with non-ribosomal peptide synthases NosA and NosC from Nostoc sp.

alb IV potentially encodes a protein of 9413 aa (AlbIV) with a predicted size of 104.8 kDa. AlbIV is similar to several non-ribosomal peptide synthase such as the BA3 peptide synthase involved in bacitracin biosynthesis in Bacillus licheniformis (Table 4). AlbIV forms one NRPS module designated NRPS-5 that contains only an A domain and a PCP domain (FIG. 2). Sequences characteristic of the domains A and PCP commonly found in peptide synthases (Marahiel et al., 1997) are conserved in AlbIV (Tables 6 and 7). However the A domain present in AlbIV differs from A domains commonly found in peptide synthases: conserved sequences corresponding to cores A8 and A9 in AlbIV are separated by a very long peptide sequence of 390 amino-acids. This additional peptide sequence exhibits a significative similarity with the protein WbpG of 377 amino acids involved in the biosynthesis of a lipopolysaccharide in Pseudomonas aeruginosa (Table 4).

AlbVII potentially encodes a protein of 765 aa (AlbVII) with a predicted size of 83.0 kDa similar to the 4-hydroxybenzoate-CoA ligase from several bacteria and the closest protein (HbaA) was from Rhodospseudomonas palustris (Table 4). High similarity between AlbVII and HbaA suggests that AlbVII is a 4-hydroxybenzoate-CoA ligase and constitutes a fourth PKS module designated PKS-4. The size of HbaA is smaller (539 aa) and the similarity between the two proteins starts only at the residue 277 of AlbVII and at the residue 28 of HbaA. Comparison of AlbVII sequence located upstream from residue 277 produced no significant alignment AlbVII, like 4-hydroxybenzoate-CoA ligases, contains some conserved sequences characteristic of the A domain commonly found in peptide synthases (Table 6).

AlbIX encodes a protein of 1959 aa (AlbIX) with a predicted size of 218.4 kDa similar to non-ribosomal peptide synthases. Known conserved sequences, characteristic of the domains commonly found in peptide synthases (Marahiel et al., 1997), were compared with those from AlbIX which forms two NRPS modules designated NRPS-6 and NRPS-7 (Tables 5, 6 and 7). NRPS-6 contains only one A and one PCP domain. NRPS-7 contains the three domains characteristic of NRPS modules (A-C-PCP) followed by a TE domain (FIG.
The alb Carbamoyl Transferase Gene

albXXV potentially encodes a protein of 584 aa with a predicted size of 65.2 kDa. This protein, AlbXXV, is similar to BlmD, a carbamoyl transferase involved in bleomycin biosynthesis in _Streptomyces vertiluvis_ (Du et al., 2000, and to a probable carbamoyl transferase potentially expressed in _P. aeruginosa_ (Table 4). High similarity of AlbXXV with these proteins suggests that AlbXXV is a carbamoyl transferase.

EXAMPLE 10

The alb Esterase Genes

albXXV potentially encodes a protein of 315 aa with a predicted size of 35.9 kDa. This protein, AlbXI, exhibits low similarity to SyrC, a putative thioesterase involved in serine-glycine biosynthesis by _Pseudomonas syringae_ (Zhang et al., 1995), and to a potential hydrolyase encoded by _Streptomyces coelicolor_ (Table 4). Precise function of SyrC remains unknown but SyrC is similar to a number of thioesterases, including fatty acid thioesterases, haloperoxidases, and acyltransferases that contain a characteristic GxCxG motif. The corresponding SyrC domain GICAG is conserved in AlbXXV which contains the sequence GWCG, except that A replaces the last G, suggesting that AlbXXV may be an esterase despite its low overall similarity with SyrC.

albXIII potentially encodes a protein of 317 aa with a predicted size of 34.5 kDa. This protein, AlbXIII, is similar to hypothetical proteins with unknown function from several bacteria including _ Caulobacter crescentus_ (Table 4). AlbXIII and these hypothetical proteins contain a GxxG motif characteristic of serine esterases and thioesterases, the corresponding sequence in AlbXIII being GHSVG. In addition, AlbXIII presents a similarity with the 2-acetyl-1-alkylglycerophosphocholine esterase which hydrolyzes the platelet-activating factor in _Canis familiaris_ (Table 4), suggesting that AlbXIII is an esterase.

EXAMPLE 11

The alb Methyltransferase Genes

albII potentially encodes a protein of 343 aa (AlbII) with a predicted size of 37.7 kDa. AlbII is 100% identical to the xac cistron, previously described as encoding an O-methyltransferase downstream xacB (Huang et al., 2000a). This conclusion is based on the similarity of XacB with a family of methyltransferases that utilize S-adenosyl-L-methionine (SAM) as a co-substrate for O-methylation including TcmO protein from _Streptomyces glaucescens_ (Huang et al., 2000a). AlbII contains three highly conserved motifs of SAM-dependent methyltransferases, including the motif involved in SAM binding (FIG. 3). In the Figure, identical or similar amino acids (A=G; D=E; I=L=V) are shown in bold. Numbers indicate the position of the amino acid from the N-terminus of the protein. Abbreviations used in the Figure are: Sgl-TcmO (SEQ ID NOs: 55, 56 and 57) and Sgl-TcmN (SEQ ID NOs: 58, 59 and 60), multifunctional cyclase-hydrazide-O-Mutase and tetracanoylmethylpolyketide synthesis 8-O-Mutase of _Streptomyces glaucescens_, respectively (accession number: M80674); Smy-MdmC, midecamycin-O-Mutase of _Streptomyces mectarifaciens_ (accession number: M93938) (SEQ ID NOs: 61, 62 and 63); Mxa-SafC, Safarycin O-Mutase of _Myxococcus xanthus_ (accession number: U24657) (SEQ ID NOs: 64, 65 and 66); Ser-EryG, erythromycin biosynthesis O-Mutase of _Saccharopolyspora erythrea_ (accession number: S18533) (SEQ ID NOs: 67, 68 and 69); Spe-DauK, carminomycin 4-O-Mutase of _Streptomyces peucetius_ (accession number: L13453) (SEQ ID NOs: 70, 71 and 72); Sal-DmpM, O-demethylpolymycin-O-Mutase of _Streptomyces alboginer_ (accession number: M74560) (SEQ ID NOs: 73, 74 and 75); Shy-RapM, rapanycin O-Mutase of _Streptomyces hygroscopicus_ (accession number: X86780) (SEQ ID NOs: 76, 77 and 78); Sav-AvdE, avermectin B 5-O-Mutase of _Streptomyces avermitilis_ (accession number: G5921167) (SEQ ID NOs: 79, 80 and 81); Sar-Cm, mithramycin C-methyltransferase of _Streptomyces argillaceus_ (accession number: AF077869) (SEQ ID NOs: 82, 83 and 84).

AlbII, putative albacidin biosynthesis C-Methyltransferase of _Xanthomonas albilineans_ (SEQ ID NO: 27); identical to XacB, accession number: AF239749) (SEQ ID NOs: 85, 86 and 87).

Comparison of AlbII with the Genbank database revealed that AlbII, besides 100% identity to XacB, exhibited the highest degree of overall identity with MtmMII, a C-methyltransferase from _Streptomyces argillaceus_ (Table 4) involved in C-methylation of the polyketide chain for mithramycin biosynthesis, suggesting that AlbII is a C-methyltransferase. XacB was not compared with Birch and co-workers with MtmMII (Huang et al., 2000a) because the MtmMII sequence was not available until recently in the Genbank database. The three highly conserved motifs in SAM methyltransferases are also present in MtmMII (FIG. 3), suggesting that AlbII is a C-methyltransferase SAM-dependent.

albV potentially encodes a protein of 286 aa (AlbV) with a predicted size of 32.1 kDa similar to several hypothetical proteins from _Mycobacterium tuberculosis_ (Genbank accession Nos. AA46042, AA48238, AA44517, AA46218) and from _S. coelicolor_ (Genbank accession No. CAC03631). AlbV is also similar to the tetracanoylmethylene C synthesis protein (TcmP) of _Pseudomonas multocida_ (Table 4). Four highly conserved motifs in TcmP and other O-methyltransferases are also present in AlbV (FIG.4), suggesting that AlbV is an O-methyltransferase. In the Figure, identical or similar aa (A=G; D=E; I=L=V; K=R) are shown in bold. Numbers indicate the position of aa from the N-terminus of the protein. Abbreviations used in the Figure are: Sgl-tcmP, tetracanoylmethane C synthesis protein of _Streptomyces glaucescens_ (accession number: C47127) (SEQ ID NOs: 88, 89 and 91); Sme-PK, putative polyketide synthase of _Sinothizobium meliloti_ (accession number: AAK57534) (SEQ ID NOs: 92, 93, 94 and 95); Pnu-tcmP, tetracanoylmethane C synthesis protein of _Pseudomonas multocida_ (accession number: AAK03406) (SEQ ID NOs: 96, 97, 98 and 99); Mtu-Ont, putative O-methyltransferase of _Mycobacterium tuberculosis_ (accession number: AAK45444) (SEQ ID NOs: 100, 101, 102 and 103); Mlo-Hp, hypothetical protein containing similarity to O-methyltransferase of _Mesorhizobium loti_ (accession number: BAB50127) (SEQ ID NOs: 104, 105, 106 and 107); Mtu-Hp, hypothetical protein of _Mycobacterium tuberculosis_ (accession number: AA46042) (SEQ ID NOs: 108, 109, 110 and 111); Mtu-Hp2, hypothetical protein of _Mycobacterium tuberculosis_ (accession number: AAK48238) (SEQ ID NOs: 112, 113, 114 and 115); Mtu-Hp3, hypothetical protein of _Mycobacterium tuberculosis_ (accession number: AAK44517).
MtH-Hp4: hypothetical protein of *Mycobacterium tuberculosis* (accession number: AA466218) (SEQ ID NOs: 120, 121, 122 and 123); Sco-Hp: hypothetical protein of *Streptomyces coelicolor* (accession number: CAC03563) (SEQ ID NOs: 124, 125, 126 and 127); AlbV1, putative albicidin biosynthesis O-Methyltransferase of *Xanthomonas abutilineae* (this study) (SEQ ID NOs: 128, 129, 130 and 131).

**EXAMPLE 12**

The alb Derived-benzozate Products Biosynthesis Genes

albXVII potentially encodes a protein of 716 aa with a predicted size of 79.8 kDa. This protein, AlbXVI, is very similar to the para-aminobenzoate (PABA) synthase from *Streptomyces griseus* (Table 4). This enzyme is required for the production of the antibiotic candidin (Criado et al., 1993).

albXVII potentially encodes a protein of 137 aa with a predicted size of 15.0 kDa. This protein, AlbXIII, is similar to the 4-amino-4-deoxycytosamine lyase (ADCL) from *P. aeruginosa* (Table 4). The function of ADCL is to convert 4-amino-4-deoxycytosamine into PABA and pyrurate. The length of AlbXVIII is smaller (Table 4) than the length of ADCL and the similarity of AlbXVIII with this protein starts only at residue 161. AlbXVIII is preceded by a small ORF encoding a sequence of 59 amino acids similar to the first 42 amino acids of ADCL from *P. aeruginosa*. These data suggest that AlbXIII is probably a truncated form of AlbXVII and probably not functional. AlbXIII may, therefore, not be involved in albicidin biosynthesis. The region between albXVII and albXVIII was amplified by PCR from total DNA of *X. axonopodis* strain ID101 using primers ORF1 (5’aggagggagcgcttcag3’) and ORF2 (5’ccgctgttggcgagcagc3’) (Table 4). These ORFs were sequenced and the sequence data showed that the PCR generated product was identical to the sequence of pALBS540, indicating that the recombinant of albXVIII did not occur during cloning of the genomic fragment in pALBS540.

albXX potentially encodes a protein of 202 aa with a predicted size of 22.6 kDa. This protein AlbXX is similar to the 4-hydroxybenzozate synthase involved in ubiquinone biosynthesis by *Escherichia coli* (Siebert et al., 1992).

**EXAMPLE 13**

The alb Regulatory Genes

albIII potentially encodes a protein of 167 amino acids with a predicted size of 17.8 kDa that is similar to the transcription factors ComA of different bacteria such as *E. coli* and *B. licheniformis* (Table 4). ComA transcription factors appear to be involved in regulation of antibiotic production in bacteria. In *E. coli*, a gene similar to comA is present in the enterobacter biosynthesis gene cluster (Liu et al., 1989). In *B. subtilis*, ComAB was described as a positive activator of lichenysin synthetase transcription (Yakimov et al., 1998) and a gene similar to comA was shown to be essential for bacilysin biosynthesis (Yuzgan et al., 2001). These data suggest that AlbIII regulates transcription of genes involved in albicidin biosynthesis.

albVII potentially encodes a protein of 330 aa with a predicted size of 37.7 kDa. This protein, AlbVIII, is very similar to the SyrP like protein from *X. verticillus* and to SyrP protein from *P. syringae* (Table 4). SyrP participates in a phosphorylation cascade controlling syringomycin synthesis (Zhang et al., 1997) and the syrP like gene was described in the *X. verticillus* bleomycin biosynthetic gene cluster (Du et al., 2000). These data suggest that AlbVII regulates albicidin biosynthesis in *X. abutilineae*.

**EXAMPLE 14**

The alb Resistance Genes

albXIV potentially encodes a protein of 496 aa with a predicted size of 52.7 kDa. This protein, AlbXIV, is 100% identical to AlbF isolated from *X. abutilineae* strain Xa13 (GenBank Accession AF403709; direct submission by Bostock and Birch and described as “a putative albicidin efflux pump which confers resistance to albicidin in E. coli”). AlbXIV and AlbF are closely related to a family of transmembrane transporters involved in antibiotic export and antibiotic resistance in many antibiotic-producing organisms. AlbXIV and AlbF exhibited the highest degree of overall identity with the putative transmembrane efflux protein from *S. coelicolor* (Table 4). These data suggest that AlbXIV and AlbF may be involved in albicidin resistance by transporting the toxin out of the bacterial cells that produce it. Alternatively, AlbXIV and AlbF may simply play a role in antibiotic secretion and/or plant pathogenesis to effect the transport of albicidin outside of producing cells.

albXIX potentially encodes a protein of 200 aa with a predicted size of 22.8 kDa. This protein, AlbXIX, is similar to the MchG protein from *E. coli* (Table 4). In *Enterobacteriaceae*, the MchG protein, together with two other proteins (McbE and McbF), was shown to cause immunity to the peptide antibiotic microcin B17 which inhibits DNA replication by induction of the SOS repair system (Garrido et al., 1988). McbE and McbF proteins serve as a pump for the export of the active antibiotic from the cytoplasm, whereas a MchG alone also provides some protection: a well-characterized deficient-immunity phenotype is exhibited by microcin B17-producing cells in the absence of the immunity gene mchG (Garrido et al., 1988). The significant similarity between AlbXIX and MchG, together with the fact that albicidin also blocks DNA replication (Birch and Patil, 1985a) suggests that AlbXIX confers immunity against albicidin in *X. abutilineae*.

**EXAMPLE 15**

Transposition Proteins

albV is 100% identical to the thp gene described in a divergent position upstream from xobH (Huang et al., 2000a). The thp gene potentially encodes a protein of 239 aa displaying significant similarity to the IS21-like transposition helper proteins. In *X. abutilineae* strain LS155 from Australia, insertional mutagenesis of thp blocked albicidin production, but trans-complementation failed, indicating the involvement in albicidin production of a downstream gene in the thp operon (Huang et al., 2000a).

albXVI potentially encodes a protein of 88 aa with a predicted size of 9.8 kDa similar to the transposases from several bacteria such as *Xanthomonas axonopodis* or *Desulfovibrio vulgaris* (Table 4).
The presence of transposition proteins in the XALB1 cluster is probably a remnant from a past transposition event that may have contributed to the development of the albidicin XALB1 cluster.

**EXAMPLE 16**

**Unknown Functions**

AlbX potentially encodes a protein of 83 aa with a predicted size of 9.4 kDa. This protein, AlbX, is similar to an hypothetical protein from *P. aeruginosa* and to the MtbH protein from *Mycobacterium tuberculosis*. MtbH is a protein with unknown function found in the mycobacteric gene cluster (Quadri et al., 1998). A MtbH-like protein with unknown function was also described in the bleomycin biosynthetic gene cluster of *S. verticillus* (Du et al., 2000). These data suggest that AlbX is involved in albidicin biosynthesis but its function remains unknown.

AlbXI potentially encodes a protein of 451 aa with a predicted size of 51.6 kDa. This protein, AlbXII, is very similar to a protein of 55 kDa encoded by the boxB gene in *Azotobacter evansii* (Table 4). This protein is a component of a multicomponent enzyme system involved in the hydroxylation of benzoylCoA, a step of aminic benzoate metabolism in *Azotobacter evansii*, but its function remains unknown (Mohamed et al., 2001).

**EXAMPLE 17**

**Prediction of Amino Acid Specificity of Alb NRPS Modules**

In NRPSs, specificity is mainly controlled by A domains which select and load a particular amino- or carboxylic acid residue (Marahiel et al., 1997). The substrate-binding pocket of the phenylalanine adenylation (A) domain of the gramicidin S synthetase (GrsA) from *Brevibacillus brevis* was recently identified by crystal structure analysis as a stretch of about 100 amino acid residues at the N-terminus of the GrsA (Phe) substrate binding pocket, similar models have been published to predict the amino acid substrate which is recognized by an unknown NRPS A domain (Challis et al., 2000; Stachelhaus et al., 1999). These models postulate specificity-conferring domains for A domains from NRPS consisting of critical amino acid residues putatively involved in substrate specificity. The model proposed by Marahiel and co-workers (Stachelhaus et al., 1999) defined a signature sequence consisting of up to 10 amino acids lining with the C-terminal residues of a phenylalanine-binding pocket located at positions 235, 236, 239, 278, 299, 301, 322, 330, 331 and 516 in the GrsA (Phe) sequence (accession number: P14687). The model proposed by Townsend and co-workers (Challis et al., 2000) uses only the first eight of these critical residues.

Preliminary specificity assignments of albidicin synthase AlbI, AlbIV, AlbVII and AlbIX NRPS modules were made by comparison of complete sequences between conserved motifs A4 and A5 with sequences in the Genbank database. The corresponding sequence of the AlbIV NRPS-5 module is most related to domain 5 of bacitracin synthase (BAC3) from *B. licheniformis* that was suggested to activate Asn (Konz et al., 1997). Corresponding sequences of AlbI and AlbIX NRPS-1, NRPS-3, NRPS-6 and NRPS-7 modules, apart from their very high similarity with XalB, exhibited the highest degree of overall identity (39%) with the Blm NRPS2 module of the biosynthetic gene cluster for bleomycin from *S. verticillus* that specifies for β-Alanine (Du et al., 2000). The corresponding sequence of AlbVII PKS-4 produced the highest significant alignment with ace-3-CoA ligation from *Sulfobolus salifatricus* (Genbank accession number: AAK41550), ary-CoA ligase from *Comamonas testosteroni* (Genbank accession number: AAC38458) and 4-hydroxybenzoate-CoA ligase from *R. palustris*. The sequence between motifs A4 and A5 of the AlbI NRPS-2 could not be significantly aligned with any sequence present in the Genbank database. Comparison of this sequence with the corresponding sequence of GrsA (Phe) revealed that parts of the putative core and structural “anchor” sequences of AlbI NRPS-2 are deleted (Fig. 5), suggesting that the AlbI NRPS-2 substrate binding pocket is not functional. In the Figure, amino acids of the six Alb NRPSs and of Alb PKS4 that are identical or similar to GrsA or Blm modules (A-G, D-E, F-I, K-V) are shown in bold. Amino acids underlined in the GrsA sequence correspond to the phenylalanine-specific binding pocket. The positions of these amino acids in the GrsA primary sequence are indicated at the top of the figure. Amino acids underlined in other sequences correspond to putative constituents of binding pockets, aligned with the seven residues of the phenylalanine-specific binding pocket of GrsA. Shaded amino acids correspond to the putative core sequences and structural “anchors” based on comparison with the GrsA binding-pocket structure.

Alignment of the primary sequence between conserved motifs A4 and A5 of the AlbI, AlbIV, AlbVII and AlbIX NRPS-1, NRPS-3, NRPS-5, NRPS-6, NRPS-7 and PKS4 modules with the corresponding sequence of GrsA (Phe) (Fig. 5) revealed the putative constituents of binding pockets that constitute the codes as defined by Marahiel and co-workers (Stachelhaus et al., 1999). These codes were compared with those of proteins most related to the sequence between the A4 and A5 motifs (Table 8) and were analyzed with the model proposed by Townsend and co-workers (Challis et al., 2000; jnmunix.hcfjiu.edu/~ravel/nrps/). Using these codes, we were able to predict the Asparagine specificity of the AlbIV NRPS-5 module. The AlbIV NRPS-5 signature is 100% identical to BacC-M5 (Asn) and TyrC-M1 (Asn) codes identified in bacitracin synthetase 3 from *B. licheniformis* and in tyrocidine synthetase 3 from *B. brevis* (Table 8). The AlbIV NRPS-5 signature is also identical to the Asn code defined by Marahiel and co-workers (1997), except that Leu is replaced by L at position 299 (Table 8). The AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures did not match any of those defined by Marahiel and co-workers (1997). Similarly, convincing predictions using the model proposed by Townsend and co-workers were not obtained (Challis et al., 2000, jnmunix.hcfjiu.edu/~ravel/nrps/). The AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures diverged from all NRPS signatures previously described, except from the XabB signature that is identical to the AlbI NRPS-1 and 3 signatures. The signature most closely related to AlbI NRPS-1 and 3 specify Pro and the signature most closely related to AlbIX NRPS-6 and 7 specify Ser, but the degree of similarity in both cases is very weak (Table 8). The PKS-4 signature is similar to the AlbI NRPS-1 and NRPS-3 signatures at positions 235, 299 and 301.

Analysis of alignment of the primary sequence between conserved motifs A4 and A5 of the AlbI and AlbIX NRPS-1, NRPS-3, NRPS-5, NRPS-6 and NRPS-7 modules with the corresponding sequences of the bleomycin synthase (Bln) NRPS (β-Ala) and gramicidin S synthetase (GrsA) modules (Fig. 5) revealed that (i) sequences of AlbI NRPS-1 and AlbI NRPS-5 differ only at the level of two residues that are not involved in
substrate binding. (ii) sequences of AlbIX NRPS-6 and AlbIX NRPS-7 are 100% identical, (iii) sequences of AlbI NRPS-1 and AlbI NRPS-3 are very similar to sequences of AlbIX NRPS-6 and AlbIX NRPS-7 but differ at the level of five putative constituents of binding pocket, (iv) AlbI and AlbIX NRPS residues, which are similar to residues of Bln NRPS2 (β-Ala) or GnrA (Phe), are essentially located at the level of the putative core sequences and structural "anchor", and differ at the level of putative constituents of the binding pocket.

Binding-pocket constituents forming the NRPS signatures have been classified into three subgroups according to their variability among 160 specificity-conferring signature sequences (Stachelhaus et al., 1999): (i) invariant residues Asp235 and Lys517 that mediate key interactions with the α-amino and α-carboxylic group of the substrate, respectively; (ii) moderately variant residues in positions 236, 301 and 330 which correspond to aliphatic amino acids and which may modulate the catalytic activity and fine-tune the specificity of the corresponding domains; (iii) highly variant residues in positions 239, 278, 299, 322 and 331 which may facilitate substrate specificity. AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures are not totally in accordance with this classification. Invariant residue Lys517 is conserved in the four NRPS signatures, indicating the presence of an α-carboxylic group in the corresponding substrates. The Asp235/Ala alteration is not consistent with an α-amino acid substrate. Birch and co-workers (Huang et al., 2001) assumed that the initial alanine residue in the XabI signature was consistent with a nonproteinogenic hydroxy acid substrate by analogy with the initial glycine in the signature of the hydroxysolvanic acid (HVL) loading domain of enniatin synthetase. The presence of an initial Alanine in the AlbVII PKS-4 signature (Fig. 8) and in several 4-hydroxybenzoate-CoA ligase genes may confirm this hypothesis. However, the HVL loading domain of enniatin synthetase (Table 8) and AlbVII PKS4 are not preceded by a C domain and are not followed by a PCP domain, in contrast to the AlbI and AlbIX NRPS-1, 3, 6 and 7 modules. An Asp235Val alteration was recently described in the β-Ala specificity-conferring code (Du et al., 2000, Table 8), suggesting that the substrate of AlbI and AlbIX NRPS-1, 3, 6 and 7 modules may be different from α-amino acids but may contain an amino group. Residue 236 is an aliphatic residue (Val or Ile) in all AlbI and AlbIX NRPS-1, 3, 6 and 7 signatures. Residue 301 is an aliphatic residue (Ala) in the AlbI NRPS-1 and 3 codes, but it is a Ser in the AlbIX NRPS-6 and 7 signatures. Residue 330 is not an aliphatic residue in the four NRPS signatures but an Asp. Similar alterations are present in the β-Ala code: residue 236 is an Asp, residue 301 is a Ser and residue 330 is an aliphatic amino acid. Concerning highly variable residues, AlbI NRPS-1 and 3 signatures differ from AlbIX NRPS-6 and 7 signatures at residue positions 299, 322 and 331, confirming that both types of NRPS modules specify different substrates.

Table 8: Comparison of signature sequences, as defined by Marahiel and co-workers (Stachelhaus et al., 1999), derived from sequences between the A4 and A5 domains of the AlbI, AlbIV, and AlbIX NRPS modules with those of Tyr-M1 (Pro) (Tyr, Asp) synthetase 2 module 1 accession: AAC54929), VirS (Pro) (Virginiamycin S synthetase, accession number: CAA73210), HVL (hydroxysolvanic acid-CoA ligase, ACT1 enniatin synthetase, accession number: S39842), Emf-M1 (Ser) (Enniobactin synthetase, accession number: AAK92015), β-Ala code (β-Ala selectivity-conferring code defined by Du et al., 2000), BacC-M2 (Asn) (Bacitracin synthetase 3, accession number: AAC065834), Tyr-M1 (Asn) (Tyr, Asp) synthetase 3, accession number: AAC54930) and Asn code (Asn selectivity-conferring code defined by Marahiel and co-workers (Stachelhaus et al., 1999). Amino acids of AlbI NRPS-1 and NRPS-3 signatures identical or similar to Tyr-M1 (Pro), VirS (Pro) and HVL, signatures (A>G; D=E; I=L; V=R; K) are shown in bold. Amino acids of AlbIX NRPS-6 and NRPS-7 signatures identical or similar to Vir (Pro) and Bln (&Ala) signatures (A>G; D=E; I=L; V=R; K) are shown in bold. Variability: 0 indicates invariant residues, +/- moderately variant residues and ++ highly variant residues.

EXAMPLE 18

Identification of Putative Promoters and Putative Terminators in XALB1

Putative rho independent terminators were identified downstream of albIV and albXVI using the Terminator program (Brendel and Trifonov, 1986), run with the Wisconsin Package™ GCG software (Fig. 6). In the Figure, dashes indicate palindromic sequences. Symbols used in the Figure are: P Primary structure value of putative terminator (minimum threshold value of 3.5 represents 95 percent of known, factor-independent, prokaryotic terminators); S Secondary structure value of putative terminator. The presence of these terminators confirmed the proposed genetic organization of operons 1 and 3. A rho-independent terminator was identified in the intergenic region between albXVII and albXVIII, suggesting that the group of genes initially supposed to be organized in operon 4 may be in fact organized in two operons, operon 4 formed by albXVII and operon 5 by albXVIII B ألXX. No putative rho independent terminator was found downstream of albIX and from albX.

The 236 bp region between albI (operator 1) and albV (operator 2) is 100% identical to the sequence between xabB and the genes that is assumed to contain a bidirectional promoter (Huang et al., 2000a and 2001), suggesting that transcription of operon 1 and 2 is regulated by the same bidirectional promoter region (Huang et al., 2001).

The 412 bp region comprised between albX (operator 3) and albXVII (operator 4) also contains a putative bidirectional promoter (Fig. 7). In the Figure, the sequence of putative promoters are underlined, and putative ATG or TTAG start codons are in bold. The closest matches (TTCAGA-18X-TATAGT) to the consensus-35 (TTCAGA) and -10 (TATAAT) sequences for E. coli 350 promoters occurs 61 bp upstream from albX (operator 3). The closest matches (TTCAGA-19X-TATACA) to the consensus sequences for E. coli 350 promoters occur 320 bp upstream from albXVII (operator 4). The region between albXVII and albXVIII lacks any apparent E. coli 350 promoter. However, the sequence immediately upstream from albXIX, corresponding to the coding sequence of albXVIII, potentially contains an unidirectional promoter (Fig. 7). The closest match (TTCACG-19X-TATATT) to the consensus sequences for E. coli 350 promoters occurs 33 bp upstream from albX. The presence of a terminator downstream from albXVII and of a promoter upstream from albXIII suggests that albXVIII is not transcribed and that albXIX and albXX form operon 5.

EXAMPLE 19

Cloning of the XALB2 Gene Cluster

The 6 kb EcoR I fragment carrying Tn5 and flanking sequence from strain AM37 was cloned in pBR325 and the
obtained plasmid was designated pAM37 (Table 1). A 1.1 kb Hind III-Hind III DNA fragment from pAM37, named PR37 (Table 1), was labeled with 32P and used to probe the 845 clones from the genomic library of X. albilineans strain Xa23R1, previously described (Rott et al., 1996). Eight new cosmids hybridized to this probe and restored albicidin production in mutant AM37. One of these cosmids, pALB389, carrying an insert of about 37 kb (Table 1), was used for complementation studies of the five mutants not complemented by pALB540 and pALB571. Cosmid pALB389 complemented mutants AM10 and AM37. Mutant AM10 was initially thought to be complemented by pALB639 (Rott et al., 1996). However, further complementation studies showed that mutant AM10 was not complemented by pALB639 and that only three mutants (AM12, AM13 and AM36) were complemented by pALB639 containing the third genomic region XALB3 involved in albicidin production. A 3 kb EcoRI-EcoRI DNA fragment from pALB389 that hybridized with probe PR37 was sub-cloned in pUCR043 (Table 1). The resulting plasmid pAC389.1 complemented mutants AM10 and AM37, confirming that the second region involved in albicidin production, XALB2, was present in the 3 kb insert of pAC389.1.

EXAMPLE 20
Cloning of the XALB3 Gene Cluster

Cosmid PAlB639, carrying an insert of 36 kb (Rott et al., 1996; Table 1) was used as a probe to compare the EcoRI restriction profiles of X. albilineans strain Xa23R1 with those of mutants AM12, AM13 and AM36 which were supposed to be mutated in the XALB3 gene cluster. An 11 kb band which was found in strain Xa23R1 but not in the three mutants was supposed to contain the XALB3 gene cluster. A 9.7 kb EcoRI DNA fragment purified from cosmid pALB639 also used as a probe in Southern blot analysis revealed the same 11 kb band. This 9.7 kb EcoRI DNA fragment was sub-cloned in pUCR043 (Table 1) and the resulting plasmid pAlB639A complemented mutants AM12, AM13 and AM36. The third region involved in albicidin production, XALB3, was therefore present in the 9.7 kb insert of pAlB639A.

EXAMPLE 21
Sequencing of the Tn5 Insertional Site of tox'
Mutants Located in XALB2 and XALB3 and Sequencing of the Genomic Regions XALB2 and XALB3

In FIG. 8, E, H, Sa and S indicate restriction endonuclease cut sites for EcoRI, HindIII, SalI and Sau3AI, respectively. The DNA inserts carried by plasmids pAC389.1, pALB639A or pALV639 are represented by the bars at the top of the respective figures. Positions of the Tn5 insertional sites of mutants AM10, AM12, AM36 and AM37 were determined by sequencing and are indicated by vertical arrows. The DNA region corresponding to the Tn5 flanking regions in pAM10, pAM12.1, pAM36.2 and pAM37 and in the PR37 DNA fragment are represented by the bars at the bottom of the respective figures. The location and direction of albXXI and albXXII are indicated by thick black arrows. The location of other ORFs in XALB2 similar to those described by Huang et al. (2000b) are indicated by thick white arrows.

The 7 kb EcoRI fragment carrying Tn5 and flanking sequence from strain AM10 was cloned in pBluescript II KS (+), and the obtained plasmid was designated pAM10 (Table 1). The sequences between EcoRI sites and the Tn5 insertional site of mutants AM10 and AM37 were sequenced from the resulting plasmids pAM10 and pAM37, respectively. The complete double-strand nucleotide sequence of the 2,986 bp EcoRI B EcoRI insert of pAC389.1 was determined from sequencing results of plasmids pAC389.1, pAM10 and pAM37 (FIG. 8). The Tn5 insertional sites of mutants AM10 and AM37 were sequenced from plasmids pAM10 and pAM37 (Table 1), respectively, using the sequencing primer GUSN (5'gccaagcaggctgag3') that annealed 135 bp downstream from the insertional sequence IS30L of Tn5-gusA. The Tn5 insertional site of AM10 and AM37 was located at position 2107 and 1882, respectively.

The EcoRI fragments carrying Tn5 and the flanking sequences from mutants AM12 and AM36 were cloned in pBR325 (Rott et al., 1996; Table 1). The sequences between EcoRI site and the Tn5 insertional site of mutants AM12 and AM36 were sequenced from the resulting plasmids pAM12.1 and pAM36.2, respectively. The complete double-strand nucleotide sequence of the 9,673 bp EcoRI B Sau3AI insert of pALB639A was determined from the sequencing results of plasmids pAM12.1, pAM36.2 and pALB639A (FIG. 8). The Tn5 insertional site of mutants AM12 and AM36 was sequenced from plasmids pAM12.1, pAM36.2 using the sequencing primer GUSN (5'gccaagcaggctgag3') that annealed 135 bp downstream from the insertional sequence IS30L of Tn5-gusA. The Tn5 insertional site of AM12 and AM36 was located at position 6500 and 7232, respectively (FIG. 8).

EXAMPLE 22
Homology Analysis and Genetic Organization of XALB2 (FIG. 8)

The sequence of 2986 bp containing XALB2 is 99.4% identical to the sequence of 2989 bp containing xabA described in X. albilineans strain LS155 from Australia (Huang et al., 2000b; accession number AF191324). The Tn5 insertional site of mutant LS156 described in xabA is 15 bp upstream from the insertional site of AM37. The orf disrupted in AM37 and AM10, designated albXXI, is identical to xabA except a C which replaces a T at position 1642. albXXI potentially encodes a protein of 278 aa with a predicted size of 29.3 kDa which is 100% identical to the potential product of xabA, described as a phosphopantetheinyl transferase (Huang et al., 2000b). Region XALB2 contains three additional orfs (orf1, orf2, and orf3) similar to those described by Huang et al. (2000b; orf, rsp6 and asp1). orf2 and orf3 are 100% identical to rsp6 and asp1, respectively, and orf1 is similar to but smaller than orf. There are no close matches to the E. coli 70 promoter B10 (TAATA) and B35 (TTGACA) consensus sequence, and no putative RBS site upstream from the putative start codon ATG of albXXI. The putative factor-independent transcription site described at 42 bp downstream from the TGA stop codon of xabA (Huang et al., 2000b) is also present at the same position downstream from albXXI.

EXAMPLE 23
Homology Analysis and Genetic Organization of XALB3 (FIG. 8)

The orf disrupted in mutants AM12 and AM36 was located between nucleotide 6900 (ATG) and 8009 (TAA) and was designated albXXII. The first ATG at position 6900 is not preceded by a putative ribosome binding sequence, suggest-
ing that the start codon is the ATG at position 6105 which is preceded at position 90 by the putative ribosome binding site sequence GGAG. A putative ribo independent terminator was identified at position 8082, 73 b downstream from albXXII (FIG. 6). There are no close matches to E. coli σ7 promoter B10 (TTATAT) and B35 (TTGACA) consensus sequence upstream from the putative start codon. The Sall DNA fragment corresponding to DNA sequence from nucleotide 5510 to nucleotide 8124, which contains the 595 b upstream from the putative start codon, the orf albXXII and the putative ribo independent terminator, was sub-cloned in pUFRO43 in the opposite direction to LacZ (Table 1). The resulting plasmid pEV639 (Table 1) complemented mutants AM12, AM13 and AM36, confirming that (i) the third region involved in albicidin production, XALB3, was present in the insert of pEV639; (ii) albXXII is not transcribed as part of a larger operon; and (iii) the 595 b upstream the putative start codon contain a promoter.

The potential product of albXXII, designated AlbXXII, is a protein of 634 aa with a predicted size of 71.5 kDa. This protein is very similar to the heat shock protein HtpG from Pseudomonas aeruginosa (identities 82%) and from Escherichia coli (identities 60%) (Table 4). The methionine encoded by the putative start codon at position 6105 was aligned with the first aminoacid of the heat shock protein HtpG from Pseudomonas aeruginosa, confirming that albXXII initiates at position 6105.

Complementation of Tox Mutants with the albXXII Gene in Fusion with LacZ

A 1,948 b fragment corresponding to the entire 1,903 b orf of albXXII and flanking nucleotides was PCR amplified from cosmid pALB639 with the forward primer 5’tgtaggtcgcactacctcgttggtgg3’ and the reverse primer 5’ttggtagctggtgctacatcgcgegg3’. Convenient in frame EcoRI and BamHI restriction sites for further cloning were simultaneously introduced with forward and reverse PCR primers, respectively. The PCR fragment was cloned into pGEMT vector (Promega) and sequenced. Several clones of the resulting plasmid pGemT/albXXII were sequenced. Because several PCR derived point mutations were observed in all the sequenced clones, a 1,920 b BglII-Sall fragment from pEV639 (corresponding to the 1,809 5’ terminal nucleotides of albXXII orf plus 111 b downstream the stop codon) was cloned into a Pgel/ValbXXIIXI clone between the BglII site located at position 94 of the albXXII orf and the Sall site of the vector’s multiple cloning site. The resulting plasmid pGemT/albXXIIbIS was an intact albXXII orf that was then subcloned as an EcoRI-Sall fragment into pUFRO43 to generate pValbXXII. This construct of albXXII in fusion with LacZ was transferred by triparental conjugation into Xa23RI insertion mutants. pValbXXII complemented mutants AM12, AM13 and AM36 (see table 9). These results confirmed that (i) the third region involved in albicidin production, XALB3, was present in the insert of pValbXXII; and (ii) albXXII is not transcribed as a part of a larger operon.

Complementation of Tox Mutants with the htpG Gene from E. coli

A 2,343 b fragment corresponding to the htpG gene of E. coli plus 458 b downstream the stop codon was PCR amplified from purified DH15a genomic DNA with forward primer 5’ttgaggtcgcactacctcgtgggtg3’ and reverse primer 5’ggctgctgctgctgctgctgctgctgctgctgctg3’. A convenient in frame EcoRI restriction site was introduced with the forward PCR primer. The PCR fragment was cloned using the pGEMT vector system (Promega). Three resulting clones potentially containing plasmid pGemT/htpG were sequenced, and one clone containing the correct sequence was selected. The 2,343 b PCR insert was then subcloned as an EcoRI-Sall fragment into pUFRO43 to generate pEVHtpG, the Sall site corresponding to the site of the vector’s multiple cloning site. This HtpG gene, in fusion with the LacZ construct, was able to restore albicidin production after transfer by triparental conjugation into AM12, AM13 and AM36 Xa23RI mutants. This result is further evidence of the involvement of a molecular chaperone HtpG in the biosynthesis of albicidin (Table 9), i.e. the first report of the requirement of a molecular chaperone HtpG in NRPS and PKS metabolism.

EXAMPLE 24

Heterologous Production of Albicidin in Fast Growing Xanthomonas axonopodis pv. Vescitoria

This example illustrates the construction of a heterologous expression system harboring the three XALB regions, its transfer into a fast growing host, Xanthomonas axonopodis pv. vescitoria and the subsequent production of a potent toxin with an antibiotic activity similar to that of albicidin. This work is a milestone in the validation of the albicidin biosynthesis model because it gives experimental evidence that the entire biosynthetic machinery required for albicidin biosynthesis has been identified, cloned, sequenced and transferred into an heterologous host, driving the production of albicidin. Cosmid pALB571 which covers the complete sequences of operons 1 and 2 was used to transfer operons 1 and 2 (FIG. 1). Operons 3 and 4 (from pALB540, XALB2 (from pAC389.1) and XALB3 (from pEV639) were subcloned into a single plasmid, pOp3-4/XALB2-3 (see below). Plasmid pOp3-4/XALB2-3 derived from shuttle vector pAFLR3 that carries one selective gene for resistance to tet racyclin and that belongs to incompatibility group IncP (Table 1). Cosmid pALB571 derived from shuttle vector pUFRO43 that carries two selective genes for resistance to kanamycine and gentamycine and that belongs to incompatibility group IncW (Table 1).

Sub-cloning of Operons 3 and 4 and XALB2 and XALB3 Regions into a Single Plasmid (FIG. 12).

A 2,787 b BamHI-PstI fragment from pALB540, corresponding to a portion of operon 4, was subcloned into pBCKS (+), yielding pB/Op4A (step 1). A XhoI site was introduced into this vector immediately upstream from the BfII site by directed mutagenesis. Mutagenesis was performed with primers XholAlb anticodon 5’gccgactgctgctgctgggttg3’ and XholAlb codon 5’tgaggtcgcactacctcgtgggtg3’ and yielded plasmid pB/Op4A/Xhol (step 2). The 2,986 b EcoRI fragment from pAC389.1 (containing XALB2) was then subcloned into pB/Op4A/Xhol, yielding pB/Op4A/XALB2 (step 3). A 10,762 b BfII fragment from pALB540 and containing complete operon 3 and the beginning of operon 4 was subcloned into pB/Op4A/XALB2 yielding pB/Op3-4/XALB2 (step 4). The 2,615 b Sall fragment from pEV639 (containing XALB3) was subcloned into pBKS, yielding pBKS/XALB3 (step 5). The Sall site located on the KpnI side of the polylinker was then destroyed and substituted by a XhoI restriction site by directed mutagenesis. This mutagenesis was performed with primers XhoSalXa1HTPGR 5’aggtagggatcagggagagaggtctg3’ and XhoSalXa1HTPGF 5’aggtaggtcgcactacctcgtgggtg3’, yielding pBKS/XALB3Xhol (step 6). Finally, the XhoI cassette of pB/Op3-4/XALB2 was subcloned into the Sall restriction site of pBKS/XALB3Xhol, yielding pBKS/Op3-4/XALB2-3 (step 7). This construct harbors an XhoI
cassette containing complete operons 3 and 4 from XALB1, albXXI from XALB2 and albXXII from XALB3. An Xhol site was added to the BamHI site of the pLAFR3 shuttle vector polynclster using the adaptor AdApTBamHI/Xhol S9gategelgtagc3, yielding pLAFR3Xhol (step 8). The Xhol cassette from pRBK5/Op3/4/XALB2-3 was then cloned into pLAFR3Xhol yielding pOp3-4/XALB2-3 (step 9). This last construct was used, along with pLAB571 (operons 1 and 2), for heterologous expression of albicidin in *X. axonopodis pv. vesicatoria*.

**Albicidin Production Assays**

The four combinations of plasmids (i.e. pUF043-pLAFR3, pUF043-pOp3/4/XALB2-3, pAlb571-pLAFR3 and pAlb571-pOp3/4/XALB2-3) were transferred into *X. axonopodis pv. vesicatoria* strain Xcv 91-11B1R1 by triparental mating. Exconjugant clones resistant to tetracycline and kanamycin were isolated. Assays for albicidin production were performed with these exconjugants clones using the same method described in Example 1 except that tetracycline (12 mg/ml) and/or kanamycin (50 mg/ml) were added to SPA medium. Tetracycline and kanamycin resistant *E. coli* clones, DH5αKT and DH5αAlb′KT (Table 1), were used as tester strains to evaluate albicidin production to ensure that growth inhibition was not due to the presence of these two antibiotics in SPA medium. Both clones, DH5αKT and DH5αAlb′KT, are tetracycline and kanamycin resistant because they carry plasmids pLAFR3 and pUF043. The albicidin resistant DH5αAlb′KT clone derived from strain DH5αAlb′ (Table 1) which is a spontaneous albicidin resistant clone isolated in a growth inhibition zone produced by *X. albilineans* strain Xa23R1.

Without antibiotics in the SPA medium, growth of clones DH5αKT and DH5αAlb′KT was not inhibited in all assays performed with the different *X. axonopodis pv. vesicatoria* exconjugants. Surprisingly, when kanamycin was present in the SPA medium, growth of both DH5αKT and DH5αAlb′KT was inhibited in all assays performed with the *X. axonopodis pv. vesicatoria* exconjugants. These results suggested that, in the presence of kanamycin, all *X. axonopodis pv. vesicatoria* exconjugants produced an antibiotic inhibiting growth of *E. coli*. Because exconjugants containing only empty vectors (pUF043 and pLAFR3) induced inhibition of *E. coli*, this antibiotic did not result from the expression of XALB1, XALB2 and/or XALB3. Additionally, there was no cross resistance between this antibiotic and albicidin. When tetracycline was present in the bioassy medium, but not kanamycin, growth of the albicidin resistant clone (DH5αAlb′KT) was not inhibited by any of the exconjugants.

In contrast, growth of the albicidin susceptible *E. coli* strain (DH5αKT) was inhibited by the exconjugants harbouring pALB571 and pOp3/4/XALB2-3 plasmids, but not by exconjugants harbouring the other three combinations of plasmids (Table 10). This result suggested that expression of the XALB1, XALB2 and XALB3 regions in *X. axonopodis pv. vesicatoria* (harbouring pALB571 and pOp3/4/XALB2-3 plasmids) led to the production of an albicidin-like antibiotic. This product inhibited growth of an albicidin sensitive *E. coli* (DH5αKT) and had no effect on the growth of an albicidin resistant clone (DH5αAlb′KT).

Preliminary results indicated that pLAFR3 derived plasmids were relatively unstable in the absence of tetracycline in the culture medium, suggesting that genes carried by pOp3-4/XALB2-3 were not expressed when *X. axonopodis pv. vesicatoria* exconjugants pALB571/pOp3/4/XALB2-3 were grown without tetracycline. Consequently, these exconjugants did not produce the albicidin-like compound in absence of any antibiotic in the culture medium (Table 10). Preliminary results also indicated that pUF043 derived plasmids are relatively stable in *X. axonopodis pv. vesicatoria* in absence of antibiotic selection, suggesting that genes carried by pALB571 are expressed when *X. axonopodis pv. vesicatoria* exconjugants pALB571/pOp3/4/XALB2-3 were grown on medium without kanamycin. Consequently, these exconjugants produced the albicidin-like compound on SPA containing only tetracycline.

Two *E. coli* DH5αKT clones, that spontaneously grew within the growth inhibition zone of a *X. axonopodis pv. vesicatoria* pALB571-pOp3-4/XALB2-3 exconjugant on SPA-tetracycline medium, were isolated and tested for resistance to albicidin. No growth inhibition was observed when these clones were used as tester strains in an albicidin production assay performed with *X. albilineans* Xa23R1. These results showed that cross-resistance occurs between the albicidin-like product of *X. axonopodis pv. vesicatoria* and albicidin produced by *X. albilineans*, suggesting that both molecules are similar. Comparison of chemical characteristics of the two molecules will, however, be necessary to confirm that the two molecules are identical.

The invention includes the isolation and sequencing of a region of 55,839 bp from *X. albilineans* strain Xa23R1 containing the major gene cluster XALB1 involved in albicidin production. Analysis of this region allowed us to predict the genetic organization of the gene cluster XALB1 which contains 20 ORFs grouped in four or five operons (FIG. 1). Because albXVII is a truncated gene, XALB1 genes may be organized in five operons. Therefore, we will from now on consider albXVII as part of operon 4 and albXIX and albXX as part of operon 5. Similar operon-type organizations for antibiotic biosynthesis clusters are well known and have been postulated to facilitate cotranslation of genes within the operon to yield equimolar amounts of proteins for optimal interactions to form the biosynthesis complexes (Cane, 1997). Overlapping genes involved in the same process are also quite common in bacteria (Normark et al., 1983).

Previous results of transposon mutagenesis and complementation studies (Rott et al., 1996; Rott, unpublished results) are in accordance with the predicted genetic organization of XALB 1 described in this study, and allowed us to establish that operons 1, 2 and 3 are involved in albicidin biosynthesis: (i) Tox+ mutants with a Tn5-gusA insertion site located in DNA fragments B, C and D were complemented by cosmid pALB571 and not by cosmid pALB540, confirming that cosmid pALB571 potentially contains the entire operon 1; (ii) Tox- mutants with a Tn5-gusA insertion site located in DNA fragments A and H were complemented by both cosmids pALB540 and pALB571, confirming that both cosmids potentially contain the entire operon 2; (iii) mutant XaAM1 with a Tn5-gusA insertion site located in DNA fragment J is the only Tn5 Tox+ mutant complemented by cosmid pALB540 and not by cosmid pALB571, confirming that cosmid pALB540 potentially contains the entire operon 3. Our mutagenesis studies did not confirm that operons 4 and 5 are required for biosynthesis of albicidin. The para-amino-nobenzoato (PABA) is required for the growth of many bacteria probably including *X. albilineans*, suggesting that a mutation in albXVII may be lethal and explaining why we did not obtain any mutant disrupted in this gene.

Putative bidirectional promoters were identified between operons 1 and 2 (Huang et al., 2001) and between 3 and 4 (FIG. 7), confirming the prediction of genetic organization of XALB1. The region upstream from operon 1 is 100% identical to the region upstream from the xabB start codon which was described as a functional promoter during the phase of
Potential rho-independent transcription terminators were identified downstream from operons 1, 3, and 4 (FIG. 6) confirming prediction of the genetic organization of these three operons. Because operons 2 and 5 are convergent (FIG. 1) and separated by a very short region of 22 bp between albIX and albXX, stop codons may allow transcription termination in the absence of sequences corresponding to potential rho-independent transcription terminators downstream from these operons. It is quite possible that simultaneous transcription of operons 2 and 5 involving the presence of two RNA polymerases (one on each strand of DNA) may cause RNA polymerases to pause at the end of each operon because of steric interference between both polymerase complexes in the same short region.

The presence of putative RBS upstream of the ATG start codons of all ORFs, except for albXVIII, suggests that these ORFs are translated in X. albilineans. The absence of a canonical RBS upstream from albXVIII further indicates that this ORF is probably not expressed. GTG and TTG codons (usually valine and leucine codons) generally serve as pro-caryotic start codons when located near the 5' end of an RNA message, but GTG start codons were also described far from the 5' end of messenger RNA in the bacitracin biosynthesis cluster of B. licheniformis (Genbank Accession No. AF184956) or in the bleomycin biosynthetic gene cluster of S. verticillus (Genbank Accession No. AF210249). This is in accordance with the fact that the two potential TTG start codons in the albxI operon 1 and 4 of XAL1B and that the two potential GTG start codons initiate internal cistrons. The albl and albxVII genes, like albxI (Huang et al., 2001), use TTG as a start codon, which may impose post-transcriptional control of the rate of gene product formation (McCarthy and Guilerzi, 1990).

The predicted genetic organization of operons 1 and 2 presents similarities with the organization of the region involved in albicidin production in strain Xa13 of X. albilineans from Australia (Huang et al. 2000a, Huang et al., 2001). This latter region also contains two divergent operons involved in albicidin production, one comprising the xabB gene (similar to albl, but with a large deletion) and the xabC gene (100% identical to albl) and the other containing the xfp gene (100% identical to albxI). In addition, the sequence between the two operons in strain Xa13 is 100% identical to the sequence between operons 1 and 2, indicating that both clusters are controlled by the same bidirectional promoter. However, transposon mutagenesis studies of Xa13 showed no evidence of another cistron downstream of xabC that may be involved in albicidin production (Huang et al., 2000b), suggesting that the Xa13 xab operon differs from the Xa23R1 operon 1, which contains two additional genes downstream from albII that are potentially involved in albicidin production (albl and albxI; refer FIG. 1).

Homology analysis revealed that four NRPS and/or PKS genes are present in XALB1 (FIG. 2), and these genes may be involved in the biosynthesis of the albicidin polyketide-polypeptide backbone (albl, albxI, albxVII and albxIV). NRPS and PKS enzymes are generally organized into repeated functional units known as modules, each of which is responsible for a discrete stage of polyketide or polypeptide chain elongation (Cane and Walsh, 1999). Each PKS or NRPS module is made up of a set of three core domains, two of which are catalytic and one of which acts as a carrier, and together are responsible for the central chain-building reactions of polyketide or polypeptide biosynthesis. Both PKS and NRPS core domains utilize analogous acyl-chain elongation strategies in which the growing chain, tethered as an acyl-S-enzyme to the flexible 20 Å long phosphopantetheinyl arm of an acyl carrier protein (ACP) or peptidyl carrier protein (PCP) domain, acts as the electrophilic partner that undergoes attack by a nucleophilic chain-elongation unit, a malonyl- or aminoacyl-S-enzyme derivative, respectively, itself covalently bound to a downstream ACP/PCP domain. In the case of a PKS, the fundamental chain-elongation reaction, a C—C bond-forming step, is mediated by a ketosynthase (KS) domain that catalyzes the transfer of the polyketide acyl chain to an active-site cysteine of the KS domain, followed by condensation with the malonylmalonyl- or malonyl-S-ACP by a deacetylase/acyltransferase (AT) domain, which catalyzes the priming of the donor ACP side arm with the appropriate monomer substrate, usually malonylmalonyl- or malonyl-CoA. The comparable core domains of an NRPS biosynthetic module function in a chemically distinct but architecturally and mechanistically analogous fashion. In the latter case, the key chain-building reaction, a C—N bond-forming reaction, involves the generation of the characteristic peptide bond by nucleophilic attack of the amino group of an amino acyl-S-PCP donor on the acyl group of an upstream electrophilic acyl- or peptidyl acyl-S-PCP chain, catalyzed by a condensation (C) domain. In functional analogy to the PKS AT domain, the core of the NRPS module utilizes an adenylating (A) domain to activate the donor amino-acid monomer as an acyl-AMP intermediate, which is then loaded onto the downstream PCP side chain. Both the AT and A domains of the respective PKSs and NRPS modules act as important gatekeepers for polyketide or polypeptide biosynthesis, exhibiting strict or at least high specificity for their cognate malonyl-CoA, malonylmalonyl-CoA or amino acid substrates. In addition to the basic subset of core domains, each PKS or NRPS also has a special set of dedicated domains responsible both for the initiation of acyl-chain assembly by loading of a starter unit onto the first, furthest upstream PKS/ NRPS module, as well as a chain-terminating thioesterase (TE) domain, most often found fused to the last module, that is responsible for detachment of the most downstream covalent acyl enzyme intermediate and off-loading of the mature polyketide or polypeptide chain (Cane and Walsh, 1999).

XALB1 potentially encodes four PKS modules and seven NRPS modules. Most of the bacterial NRPS gene clusters described up to now are organized in operon-type structures, encoding multi modular NRPS proteins with individual modules organized along the chromosomes in a linear order that parallels the order of amino acids in the resultant peptide, following the “colinearity rule” for the NRPS-template assembly of peptides from amino acids (Cane, 1997; Cane et al., 1998; Cane and Walsh, 1999; von Dohren et al., 1999). PKS and NRPS modules are apparently not organized according to this “colinearity rule” for albicidin biosynthesis because of the following features: (i) NRPS and PKS genes are expressed in two divergent operons; (ii) no AT domain was identified in PKS-2 and PKS-3 domains, suggesting involvement of a separate enzyme; (iii) the A domain of NRPS-2 is not functional, suggesting the involvement of a trans-acting A domain; (iv) a single chain-terminating TE domain was identified in XALB1 which may be responsible...
of the release of the full length albicipin polyketide-polypeptide backbone from the enzyme complexes. Exception to the "collinear rule" has also been shown for the synergomycin synthetase of \textit{P. syringae} (Guerzoni et al., 1998), for the exochelin synthetase of \textit{Myxobacterium smegmatis} (Yu et al., 1998) and for the bleomycin synthetases of \textit{S. verticillus} (Du et al., 2000).

On the basis of the deduced functions of individual NRPS and PKS domains we have aligned the four PKS and the seven NRPS modules to suggest two different putative linear models for the synthesis of the albicipin polyketide-peptide backbone (FIG. 9). In the Figure, NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl-CoA ligase; AT, acyltransferase; C, condensation; HBC, hydroxybenzoyl-CoA ligase; KR, ketoreductase; KS, ketosynthase; PCP, peptidyl carrier protein. An designation asparagine. X1 and X2 indicate substrates incorporated by NRPS-1 and 3 and by NRPS-6 and 7, respectively. The crossed A domain in NRPS-2 indicates that this deleted domain may not be functional. In model 1, (FIG. 9A), (i) the PKS-1 module alone is responsible for the initiation of the acyl-chain assembly, (ii) PKS4 (HBC) interacts with PKS-2 and PKS-3 as an AT domain to allow acyl transfer and (iii) NRPS-5 interacts with only NRPS-2. In model 2 (FIG. 9B) two different modules, PKS-1 and PKS-4, are responsible for this initiation step. Model 2 leads to the biosynthesis of four different polyketide-poly peptide backbones; in this model (i) PKS-1 (AL) and PKS-4 (HBC) are in competition for initiation of albicipin precursors; (ii) a separate AT enzyme (potentially AlbXIII) interacts with PKS-2 and PKS-3 to allow acyl transfer; (iii) NRPS-5 interacts with NRPS-2; and (iv) NRPS-5 and NRPS-6 are in competition for interaction with NRPS4.

Both models are based on the fact that PKS-1 contains the AL and ACP1 domains, and PKS-4 shows homology with the hydroxybenzoyl-CoA ligases. In other PKS systems, an N-terminal AL domain is involved in the activation and incorporation of a 3,4-dihydroxycyclohexane carboxylic acid, a 3-amino-5-hydroxybenzoic acid or a long-chain fatty acid as a starter (Aparicio et al., 1996; Motamed and Shafiee, 1998; Tang et al., 1998; Duitman et al., 1999). PKS-4 may be also involved in the activation and incorporation of hydroxybenzoyl but this latter domain lacks any ACP or PCP domain, suggesting that PKS-4 is responsible for initiation of the acyl-chain assembly (FIG. 9B). One of the three ACP domains of Albl (ACP1, ACP2 or ACP3). The 277 amino acids preceding the PKS4 module in AlblVII may be necessary for the intercommunication between AlblVII and Albl. The presence of two different PKS modules potentially involved in the initiation of the acyl-chain assembly suggests a competition of these two modules for the initiation of two different albicipin polyketide-poly peptide backbones, and this could contribute to the production of multiple, structurally related albicipins by the same cluster XAL1B. Production of two different components, one initiated by PKS-4 containing an additional aromatic ring due to incorporation of hydroxybenzoyl, may explain why partial characterization of albicipin indicated the presence of a variable number (three or four) of aromatic rings (Hung et al., 2001).

In Albl, PKS-1 is followed by the PKS-2 module which contains a KS domain and a KR domain upstream from two ACP domains (ACP2 and ACP3) and it lacks any discernable AT domain. Tandem ACP domains are unusual within PKS modules but have been shown to occur in the biosynthesis of several fungal and bacterial polyketide synthases (Mayorga and Timberlake, 1992; Yu and Leonard, 1995; Takano et al., 1995; Albertini et al., 1995). However, the significance of the tandem ACP domains in these systems has not been solved yet. In our model 2, one of the tandem ACP (ACP2 or ACP3) may interact with PKS-4 for the initiation of an acyl-chain assembly (FIG. 9B). The absence of an AT domain in the PKS-2 module suggests that a separate AT domain is indispensable for the elongation of the acyl-chain initiated by this module. Separate AT enzymes encoded elsewhere in the genome were described in other systems for two PKS modules lacking AT domains: malonyl-CoA transacylase gene (fenF) located immediately upstream from the \textit{B. subtilis} PKS-NRPS mycA gene (Duitman et al., 1999) and an AT gene located 20 kb upstream from the \textit{M. xanthus} NRPS-PKS tal gene (Puitan et al., 1999). We have not identified an AT gene in the gene cluster XALB1 and in the two other genomic regions involved in albicipin production, XALB2 and XALB3, suggesting that the trans-acting AT gene may be encoded elsewhere in the genome. However, AlbXIII, which contains the motif G/H/S/o conserved in AT domains, may be potentially involved in the acyl transfer, but the similarity of AlbXIII with AT domains is not high enough to confirm this potential function of AlbXIII (FIG. 10). FIG. 10A describes alignment of the conserved motifs in AT domains from RisA-1, -2, -3, RilB-1, RilE-1 (Rifamyacin PKSs, August et al., 1998) and BlmVIII (Bleomycin PKS: Du et al., 2000). Identical amino acids are shown in bold. FIG. 10B describes alignment of AlbXIII (SEQ ID NO. 38). FenF (a malonyl-CoA transacylase located upstream from mycA, Duitman et al., 1999) and LipA (a lipase; Valdez et al., 1999); amino acids identical to conserved AT domains motifs are shown in bold.

AlbXIII contains only four of the eleven amino acids conserved in AT domains of rifamyacin PKSs (August et al., 1998) and Bleomycin PKS (Du et al., 2000), and the AlbXIII sequence appears to be more closely related to lipases such as LipA (Valdez et al., 1999) rather than to AT domains (FIG. 10). However, FenF, the trans-acting AT domain involved in mycosubtilin biosynthesis, contains only seven of the eleven amino acids conserved in AT domains (Duitman et al., 1999; FIG. 10). AlbVII, that contains a HBCL domain, may be another candidate for the acyl transfer in PKS-2 (FIG. 9A) because HBCL exhibits some similarity with A domains at the level of cores A1, A2, A3, A4, A5 and A6 (Table 6). However, no HBCL involved in such a function has been described in the PKSs characterized so far.

In Albl, PKS-2 is followed by the PKS-3 module which contains the KS2 and the PCP1 domains and it lacks any discernable AT or A domain. PKS-3 is located upstream from the NRPS modules and should therefore be involved in the linkage of polyketide and polypeptide moieties. The presence of a PCP domain in the PKS-3 module suggests the involvement of a trans-acting A domain rather than an AT domain. A putative candidate for this trans-acting A domain is the AlbIV NRPS-5 A domain because of the lack of a C domain in the AlbIV NRPS-5 module. However, by analogy with the BlmVIII PKS module, which is involved in the linkage of polyketide and polypeptide moieties of bleomycin and which contains an AT domain followed by a PCP domain (Du et al., 2000), the presence of a PCP is not incompatible with a possible interaction of the AlbII PKS-3 module with a separate AT domain. This latter trans-acting AT domain may be the same that interacts with the AlbI PKS-2 module, the AlbVII PKS-4 module, AlbXIII or an unidentified separate AT domain.

In Albl, the PKS-3 module is followed by four NRPS modules. The NRPS-1, 2 and 3 modules display the ordered A, C and PCP domains, suggesting that they are involved in the incorporation of three amino acid residues. The A domain of the NRPS-2 module exhibits poor consensus at A2, A3, A5,
A7, A8 A9 and A10 motifs and lacks completely the A6 motif (Table 6). In addition the NRPS-2 substrate binding pocket is partially deleted (Fig. 5). These features strongly suggest that the NRPS-2 A domain is inactive and that the loading of an amino-acid on the NRPS-2 PCP domain (PCP3) is possibly catalyzed by a trans-acting A domain as in HIPWP1 (Gehring et al., 1998) and HluIII (Du et al., 2000). A putative candidate for this trans-acting A domain is the NRPS-5 A domain present in AlbIV because of the lack of a C domain in NRPS-5 (Fig. 2). The additional sequence of 300 amino-acids present in the A domain of NRPS-5 may be necessary for the intercommunication between AlbI and AlbIV. As a consequence of the interaction between NRPS-2 and NRPS-5, a competition between PCP-3 and PCP-5 domains must occur to bind the amino acid activated by the NRPS-5 A domain. A similar competition between two PCP domains was described for syringomycin biosynthesis, during the interaction between SyrR, which contains A and PCP domains, and the last module of SyrE which contains C and PCP domains (Guenzi et al., 1998). The NRPS-4 module contains only a C domain which may transfer the intermediate products synthesized by AlbI to a PC domain present in another albicidin synthase. Similar transfers were described for mycosubtilin biosynthesis in which the MycA and MycB C-terminal C domains interact with the MycB and MycC N-terminal A domains, respectively (Dutman et al., 1999).

Two different PCP domains may be involved in the transfer of the intermediate products synthesized by AlbI: the PCP-5 and PCP-6 domains which are present in the AlbIV NRPS-5 and AlbIX NRPS-6 modules, respectively. This possible competition between the two NRPS modules that contain two different A domains could also contribute to the production of multiple, structurally related albidicins by the gene cluster XALB1 (Fig. 9B). Because of the absence of a C domain in the AlbIX NRPS-6 module, the intermediate product bound on the AlbIV PCP-5 domain would be necessarily transferred to the AlbIX PCP-7 domain, like the intermediate product bound on AlbIX PCP-6. AlbIX NRPS-7, which contains the single chain-terminating TE domain, may be responsible for detachment of the mature albidicin polypeptide-polypeptide backbone from the complex of enzymes.

The linear model 1 implies that operon 1 and operon 2 in X. Albilineans strain Xa23R1 from Florida potentially produce only one albidicin polypeptide-polypeptide backbone, with a competition at the level of ACP2/ACP3 and PCP3 and PCP5 which could explain the production by X. albilineans of compounds structurally related to albidicin (Fig. 9A). The linear model 2 implies that operon 1 and operon 2 in X. albilineans strain Xa23R1 from Florida potentially produce four different albidicin polypeptide-polypeptide backbones (Fig. 9D) because of (i) the competition of AL and HBCl domains for initiation of acyl chain assembly and (ii) the competition of AlbIV NRPS-5 and AlbIX NRPS-6 modules for the incorporation of the next to last amino acid of the albidicin backbone. These four albidicin backbones may lead to the production of four components structurally very different. The polypeptide moieties of the acyl chains initiated by the AlbI AL domain or by the AlbVII HBCl domain may be very different. The polypeptide moiety of acyl chains initiated by the AlbVII HBCl domain may be shorter and may contain an additional aromatic ring. The presence of four structurally different metabolites may explain the difficulty observed by Birch and Patil (1985a) to purify albidicin and to determine its chemical structure.

Homology analysis also revealed that AlbI NRPS-1 and 3 and AlbIX NRPS-6 and 7 specify unusual substrates which seem to contain an amino group and a carboxylate group but to be different from ß-amino acids and ß-alanine. Identification of several aromatic rings in albidicin (Huang et al., 2001) suggested that NRPS-1, -3, -6 and -7 are involved in incorporation of aromatic substrates. By analogy with the Asp235Val alteration in the ß-Ala specificity-conferring code (Du et al., 2000), the Asp235Ala alteration in the NRPS-1, -3, -6 and -7 signatures could be consistent with a large distance between the amino group and the carboxylate group in the substrate specified by these modules. Based on this hypothesis, we suggest that operons 3, 4 and 5 are involved in the biosynthesis of two aromatic substrates: the para-aminobenzoate potentially synthesized by AlbIXVII (para-aminobenzoate synthase), and the carbamoyl benzoate potentially synthesized by AlbXX (hydroxylbenzoate synthase) and AlbXV (carbamoyl transferase). Incorporation of these non-proteinogenic substrates may explain why albidicin is insensitive to proteases (Birch and Patil, 1985a).

According to biosynthesis model 1 leading to the biosynthesis of only one polypeptide-polypeptide albidicin backbone that may correspond to the major component produced by X.AlbI, we propose a model allowing prediction of the composition and the structure of albidicin (Fig. 11). In the Figure, NRPS and PKS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl-CoA ligase; C, condensation; KR, ß-ketoreductase; KS, ß-ketoacyl synthase; PCP, peptide carrier protein. C atoms of albidicin backbone are numbered 1 to 38. Bold methyl groups correspond to methylation of the albidicin backbone by AlbI or AlbVI. In this model, albidicin biosynthesis is initiated by loading of an acetyl-CoA by PKS-1 (step 1), and the chain product is elongated by incorporation of (I) malonyl-CoA by PKS-2 and PKS-3 (steps 2 and 3), (ii) para-aminobenzoate or carbamoyl benzoate by NRPS-1 and NRPS-3 (steps 4 and 6), (iii) aspartate by NRPS-2 coupled to NRPS-5 (step 5) and (iv) para-aminobenzoate or carbamoyl benzoate by NRPS-6 and NRPS-7 (steps 8 and 9). The presence of the KR domain in the PKS-2 module may lead to the formation of an hydroxyl group at the C14 atom of the albidicin backbone. This hydroxyl group might be methyalted by AlbVI (O-methyltransferase). The acyl chain may also be modified by AlbI (C-methyltransferase) at C13 or C14.

The chemical composition (C50H32O10N7S1) of the molecular weight (839), and the structure of the putative XALB1 product are in agreement with the partial characterization of albidicin published by Birch and Patil (1985a) which indicated that albidicin contains approximately 38 carbon atoms and a carboxylate group and that the molecular weight of albidicin was about 842. The presence of two ester linkages in our predicted albidicin structure is also in accordance with the fact that albidicin is detoxified by the AlbB esterase (Zhang and Birch, 1997). However, an unpublished albidicin analysis cited by Huang et al. (2001) indicated the presence of (I) two OCH3 groups and not one as in our predictive albidicin structure, (ii) one CN linkage and not eleven as in our predictive albidicin structure and (iii) a trisubstituted double bond that is not present in the putative XALB1 product.

In conclusion, homology analysis of XALB1 revealed unprecedented features for hybrid polypeptide-polypeptide biosynthesis in bacteria involving a trans-action of four PKS and seven NRPS separate modules which could contribute to the production of multiple, structurally related polypeptide-peptide compounds by the same gene cluster. Characterization of the full chemical structure of albidicin may be necessary to validate these models. Four NRPS modules seem to activate a very unusual substrate. Over-expression and purification of A domains from these four NRPS modules will be necessary to examine their substrate specificities. Substrate specificity of
each A domain will therefore be determined by analysis of the ATP-PPI exchange reaction with different substrate putatively incorporated into albicidin. Investigating albicidin backbone biosynthesis will be of great interest because such information adds to the limited knowledge as to how PKS and NRPS interact and how they might be manipulated to engineer novel molecules, and may explain how _X. albilineans_ produces several structurally related, toxic compounds.

Cloning and sequencing of XALB2 showed that the same phosphopantetheinyl transferase is required for albicidin production in an _X. albilineans_ strain from Florida and in an _X. albilineans_ strain from Australia (Huang et al., 2000b), explaining the precedent results showing that strain LS156 mutated in xabA (100% identical to albX) was not complemented by pALBS40, pALBS71 and pALB639 (Rott et al., 1996). Mutant LS156 was shown to be complemented by a construction containing the coding sequence of xabA in fusion with lacZ, revealing that xabA is required for albicidin production and that no other cistron downstream from xabA was involved in albicidin production (Huang et al., 2000b). However, this complementation study did not allow determination of whether xabA is transcribed as a part of a larger operon. Here we disclose the complementation of mutant AM37 with a 2986 bp insert from _X. albilineans_ containing albXXI (100% identical to xabA), confirming that albXXI is involved in albicidin biosynthesis and indicating that the promoter of albXXI is present in the 2986 bp insert and that albXXI is not expressed as part of a operon.

Cloning and sequencing of XALB3 showed that a heat shock protein HtpG was involved in albicidin production in _X. albilineans_. The heat shock protein HtpG is an _Escherichia coli_ homologue of eukaryotic HSP90 molecular chaperone. Hsp90 from eukaryotes has been demonstrated to possess chaperone activity (Jakob et al., 1995), acting as a non-ATP dependent “holdase,” and it also has an important role in signal transduction and the cell cycle. This protein is essential in both _drosophila_ and yeast (Borkovich et al., 1989; Cutforth and Rubin, 1994). In contrast, the HtpG gene can be deleted in _E. coli_ with no effect on the viability of the strain with the exception of decreased growth rate at high temperatures (Bardwell and Craig, 1988). The in vivo role of the HtpG protein remains unknown. However, preliminary results indicated that HtpG facilitates de novo protein folding in stressed _E. coli_ cells, presumably by expanding the ability of the DnaK-DnaJ-GrpE molecular chaperone system to interact with newly synthesized polypeptides (Thomas and Baneyx, 2000). Furthermore, HtpG was copurified in _E. coli_ with MccB17 synthetase, an enzyme involved in the biosynthesis of the peptide antibiotic microcin B17 which inhibits DNA replication by induction of the SOS repair system, suggesting the requirement of HtpG for production of the antibiotic (Li et al., 1996). However, when microcin B17 production by the _E. coli_ strain deleted for HtpG was compared to the one of the parental strain, there was no effect on microcin B17 production in vivo. This result implied that the copurification of HtpG with the MccB17 synthetase was potentially an artifact, or that another _E. coli_ chaperone could substitute for HtpG (Milne et al., 1999). To examine the effect of HtpG on the reconstitution of MccB17 synthetase in vitro, the chaperone was expressed and purified as a fusion to a hexahistidine (His_6_) tag. Addition of the His_6_-HtpG did not stimulate MccB17 synthetase reconstitution or heterocyclisation activity in vitro, suggesting that HtpG mediates complex assembly or stabilizes protein subunits prior to the hetero-oligomerisation (Milne et al., 1999). Based on these results, we suggest that the function of AlbXXII is to mediate complex assembly by facilitating de novo protein folding of PKS and NRPS enzymes (AlbI, AlbIV, AlbVII and AlbIX) involved in the albicidin backbone biosynthesis.

Characterization of the complete sequence of XALB1, XALB2 and XALB3 clusters enables one to characterize all enzymes of the albicidin biosynthesis pathway including structural, resistance, secretory and regulatory elements, and to engineer overproduction of albicidin. For example one may insert expression enhancing DNA into the genome of _X. albilineans_ in a position operable to enhance expression of the Albicidins Biosynthesis Gene Clusters. One may also modify naturally occurring Albicidins to obtain additional non-naturally occurring antibiotics by adding DNA encoding additional enzymes selected to produce a modified albicidin like molecule. This approach will allow (i) the purification of albicidin and the other compounds structurally related and potentially produced by the same biosynthesis apparatus; (ii) the characterization of chemical structure of albicidin; (iii) the investigation of mode of action of albicidin in the pathogenesis of _X. albilineans_ in sugarcane; and (iv) the characterization of the bactericidal activity of albicidin. For example one may also increase the resistance of plants to damage from _X. albilineans_ infection by inserting one or more of the resistance genes identified herein into the genome of the plant.

One may also provide materials to prevent damage by albicidin produced by _X. albilineans_ by applying an agent that blocks expression of the Albicidins Biosynthesis Gene Clusters to the plant to be protected. One may also use portions of the DNA of the Albicidins Biosynthesis Gene Clusters to obtain agents useful in blocking expression of albicidin by screening materials against a modified hest cell line that expresses the Albicidins Biosynthesis Gene Clusters and selecting for materials that stop or decrease albicidin production.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial strains and plasmids used in this study</strong></td>
</tr>
<tr>
<td><strong>Strains</strong></td>
</tr>
<tr>
<td>D150</td>
</tr>
<tr>
<td>D150 MCR</td>
</tr>
<tr>
<td>Xcv 91-11B</td>
</tr>
<tr>
<td>Xcv 91-11BR1</td>
</tr>
<tr>
<td>D150KT</td>
</tr>
<tr>
<td>Strains</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>DlStA</td>
</tr>
<tr>
<td>DlStA/VKT</td>
</tr>
</tbody>
</table>

**Plasmids**

<table>
<thead>
<tr>
<th>Strains</th>
<th>Relevant characteristics</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBR325</td>
<td>Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Gibco-BRL</td>
</tr>
<tr>
<td>pBKS (+)</td>
<td>Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Statagene</td>
</tr>
<tr>
<td>pBluescript II KS (+)</td>
<td>Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>PRK2073</td>
<td>PRK2013 derivative, Km&lt;sup&gt;+&lt;/sup&gt; (rpt::Tn7), Sp&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;, helper plasmid</td>
<td>Leong et al., 1982</td>
</tr>
<tr>
<td>pUFRO43</td>
<td>IncW Molce&lt;sup&gt;+&lt;/sup&gt; LacZ&lt;sub&gt;ts&lt;/sub&gt; Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;, Cos&lt;sup&gt;+&lt;/sup&gt;</td>
<td>De Feyter and Gabriél, 1991</td>
</tr>
<tr>
<td>pAb850</td>
<td>47 kb insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Rott et al., 1996</td>
</tr>
<tr>
<td>pAb571</td>
<td>36.8 kb insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAb539</td>
<td>36 kb insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM13.1</td>
<td>24 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM13 in pBR325, Km&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM40.2</td>
<td>11 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM40 in pBR325, Km&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM45.1</td>
<td>12 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM45 in pBR325, Km&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM12.1</td>
<td>13 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM12 in pBR325, Km&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td>A</td>
</tr>
<tr>
<td>pAM36.2</td>
<td>9 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM36 in pBR325, Km&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td>A</td>
</tr>
<tr>
<td>pAb850</td>
<td>37 kb insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pAc889.1</td>
<td>2.9 kb insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAb571</td>
<td>9.4 kb insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAC/N</td>
<td>2.9 kb Sal I insert from XbaI in pUFRO43, Gmr&lt;sup&gt;+&lt;/sup&gt;, Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pBCA'</td>
<td>7.5 kb Kpn I fragment carrying a part of fragment A from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/AF</td>
<td>15.2 kb EcoR I fragment carrying fragments A and F from pALBS40 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C</td>
<td>11.0 kb Kpn I fragment B from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C'</td>
<td>6.0 kb Kpn I fragment C from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/E</td>
<td>2.8 kb Kpn I fragment E from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C'</td>
<td>2.5 kb Kpn I-EcoR I fragment F from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/G</td>
<td>1.9 kb EcoR I fragment G from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C1</td>
<td>1.4 kb Kpn I-EcoR I fragment L from pAb571 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C'</td>
<td>0.6 kb EcoR I fragment J from pALBS40 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/K'</td>
<td>4.7 kb EcoR I fragment K from pALBS40 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C</td>
<td>0.4 kb EcoR I fragment L from pALBS40 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pB/C'</td>
<td>7.5 kb EcoR I fragment N from pALBS40 in pBKS (+), Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pUFRO43/D</td>
<td>2.2 kb EcoR I-BsaI 3A 1 fragment carrying a part of fragment D from pAb571 in pUFRO43</td>
<td></td>
</tr>
<tr>
<td>pAM1</td>
<td>5 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM1 in pBluescript II KS (+), Km&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM4</td>
<td>12 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM4 in pBluescript II KS (+), Km&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM7</td>
<td>6 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM7 in pBluescript II KS (+), Km&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM10</td>
<td>7 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM10 in pBluescript II KS (+), Km&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM10</td>
<td>10 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM10 in pBluescript II KS (+), Km&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM37</td>
<td>6 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM37 in pBR325, Km&lt;sup&gt;+&lt;/sup&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;, Cm&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pAM52</td>
<td>5 kb EcoR I fragment carrying Tn5 and flanking sequences of mutant AM52 in pBluescript II KS (+), Km&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pLA8R3</td>
<td>IncP, Mob&lt;sup&gt;+&lt;/sup&gt;, LacZ&lt;sub&gt;ts&lt;/sub&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;, cos&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Staskawicez et al., 1987</td>
</tr>
<tr>
<td>pLA8R3</td>
<td>pLA8R3 with a Xhol site added to the BamHI site using an adapter</td>
<td>This study</td>
</tr>
<tr>
<td>pBC/Op4A</td>
<td>BamHII-PstI fragment from pALBS40 cloned between BamHII and PstI sites of pBKS(+)</td>
<td></td>
</tr>
<tr>
<td>pBC/Op4A/Xhol</td>
<td>PBC/Op4A with a Xhol site created by directed mutagenesis upstream from the BfiI site</td>
<td></td>
</tr>
<tr>
<td>pBC/Op4A/XALB2</td>
<td>EcoRI DNA fragment from pAC389.1 cloned into the EcoRI site of pBC/Op4A/Xhol</td>
<td></td>
</tr>
<tr>
<td>pBC/Op4A/XALB2</td>
<td>BfiI DNA fragment from pALBS40 cloned into the BfiI site of pBC/Op4A/XALB2</td>
<td></td>
</tr>
<tr>
<td>pBKS/XALB3</td>
<td>Sall DNA fragment from pSV639 cloned into the SalI site of pBluescript II KS (+)</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 1-continued

Bacterial strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Relevant characteristics*</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBKS/XALB3Xhol</td>
<td>pBKS/XALB3 with a Xhol site created by directed mutagenesis to substitute the Sall site located on the KpnI site of the polyclinker</td>
<td>*</td>
</tr>
<tr>
<td>pBKS/Op3-4/</td>
<td>Xhol DNA fragment from pBCKO3-4/XALB2 cloned into the Xhol site of pBKS/XALB3Xhol</td>
<td>*</td>
</tr>
<tr>
<td>XALB2-3</td>
<td>Sall site of pBKS/XALB3Xhol</td>
<td>*</td>
</tr>
<tr>
<td>pOp3-4/XALB2-3</td>
<td>Xhol DNA fragment from pBKS/Op3-4/XALB2-3 cloned into the Xhol site of pLAFR3Xhol</td>
<td>*</td>
</tr>
<tr>
<td>pEVlbXXII</td>
<td>albXXII in fusion with LacZ in pUFRO43, Gm', Km'</td>
<td>*</td>
</tr>
<tr>
<td>pEVhpg</td>
<td>E. coli hpg in fusion with LacZ in pUFRO43, Gm', Km'</td>
<td>*</td>
</tr>
<tr>
<td>PGemT</td>
<td>ColE1 replicon, Ap', LacZ, single 3'-T overhangs at the insertion site</td>
<td>Promega</td>
</tr>
<tr>
<td>PGemT/albXXII</td>
<td>PCR fragment containing albXXII cloned into pGemT</td>
<td>This study</td>
</tr>
<tr>
<td>PGemT/albXXII bis</td>
<td>BglIII-SallDNA fragment from pBKS/XALD3 cloned between the BglIII and Sall sites of pGemT/albXXII</td>
<td>*</td>
</tr>
<tr>
<td>PGemT/Hpg</td>
<td>PCR fragment containing the E. coli hpg gene cloned into pGemT</td>
<td>*</td>
</tr>
</tbody>
</table>

**DNA Fragment**

| PR37 | 1.1 kb Hind III-Hind III from pAM37 | * |

*Ap', Cm', Gm', Km', Rif', Sp', Tc': resistant to ampicillin, chloramphenicol, gentamycin, kanamycin, rifampicin, spectinomycin, tetracycline, respectively.

Tos-, deficient in absicacid production.

Tn3-gae, Tn3-usdAl Km' Tc', forms transcriptional fusions.

Alb', Ap', Gm', Rif' and Tc': resistant to absicacid, ampicillin, gentamycin, rifampicin and tetracycline, respectively.

### TABLE 2

Analysis of putative translational signals and location of all putative orfs identified in the XALB1 gene cluster.

<table>
<thead>
<tr>
<th>Intergenic spacing between consecutive ORFs in each putative operon</th>
<th>ORF</th>
<th>Potential RBS* (distance from start codon)</th>
<th>Start codon (position)</th>
<th>Stop codon (position)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operon 1 (strand +)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>albI</td>
<td>GAGGG (5 b)</td>
<td>TAG (30166)</td>
<td>TAG (50805)</td>
</tr>
<tr>
<td></td>
<td>albII</td>
<td>GAGGG (5 b)</td>
<td>ATG (50851)</td>
<td>TAA (51882)</td>
</tr>
<tr>
<td></td>
<td>albIII</td>
<td>GAGGG (7 b)</td>
<td>ATG (51882)</td>
<td>TGA (52385)</td>
</tr>
<tr>
<td></td>
<td>GTG overlaps TAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>albIV</td>
<td>GAGG (7 b)</td>
<td>GTG (52382)</td>
<td>TAA (55207)</td>
</tr>
<tr>
<td>Operon 2 (strand −)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>albV</td>
<td>GGAGG (8 b)</td>
<td>ATG (29929)</td>
<td>TAA (29210)</td>
</tr>
<tr>
<td></td>
<td>albVI</td>
<td>AAGG (4 b)</td>
<td>GTG (29122)</td>
<td>TGA (28262)</td>
</tr>
<tr>
<td></td>
<td>albVII</td>
<td>GAG (4 b)</td>
<td>ATG (28200)</td>
<td>TAG (25930)</td>
</tr>
<tr>
<td></td>
<td>albVIII</td>
<td>AGGTTG (4 b)</td>
<td>ATG (25895)</td>
<td>TAA (24903)</td>
</tr>
<tr>
<td></td>
<td>20 b</td>
<td>AGTTG (3 b)</td>
<td>ATG (24882)</td>
<td>TAA (19033)</td>
</tr>
<tr>
<td>Operon 3 (strand −)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>albX</td>
<td>GGGGG (8 b)</td>
<td>ATG (14497)</td>
<td>TGA (14246)</td>
</tr>
<tr>
<td></td>
<td>albXI</td>
<td>AGGAAC (6 b)</td>
<td>ATG (14164)</td>
<td>TGA (13217)</td>
</tr>
<tr>
<td></td>
<td>5 b</td>
<td>GGCCTGA (5 b)</td>
<td>ATG (13211)</td>
<td>TAA (11856)</td>
</tr>
<tr>
<td></td>
<td>36 b</td>
<td>GGGG (3 b)</td>
<td>ATG (11819)</td>
<td>TAA (10866)</td>
</tr>
<tr>
<td></td>
<td>12 b</td>
<td>GGG (8 b)</td>
<td>ATG (10853)</td>
<td>TAG (9363)</td>
</tr>
<tr>
<td></td>
<td>41 b</td>
<td>GGA (6 b)</td>
<td>ATG (9321)</td>
<td>TAG (7567)</td>
</tr>
<tr>
<td></td>
<td>208 b</td>
<td>GGAGG (4 b)</td>
<td>ATG (7358)</td>
<td>TAG (7092)</td>
</tr>
<tr>
<td>Operon 4 (strand +)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>albXVII</td>
<td>GGAGG (5 b)</td>
<td>ATG (14009)</td>
<td>TGA (17059)</td>
</tr>
<tr>
<td></td>
<td>albXVIII</td>
<td>GCTCAG (8 b)</td>
<td>ATG (17334)</td>
<td>TGA (17747)</td>
</tr>
<tr>
<td></td>
<td>274 b</td>
<td>AGG (9 b)</td>
<td>ATG (17728)</td>
<td>TGA (18530)</td>
</tr>
<tr>
<td></td>
<td>41 b</td>
<td>AGAA (8 b)</td>
<td>ATG (18372)</td>
<td>TAG (18980)</td>
</tr>
</tbody>
</table>

*Ribosomal Binding Site
### TABLE 3

Deduced functions of the ORFs in the major albicidin biosynthetic cluster X-ALB1

<table>
<thead>
<tr>
<th>ORF</th>
<th>Number of amino acids</th>
<th>Sequence homolog*</th>
<th>Proposed function**</th>
<th>PKS modules</th>
<th>PKS domains</th>
<th>NRPS modules</th>
<th>NRPS domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operon 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbI</td>
<td>6879</td>
<td>XabB (AAK15074)</td>
<td>Polyketide — peptide synthase</td>
<td>PKS-1</td>
<td>AL ACP1</td>
<td>NRPS-1</td>
<td>C A PCP2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PKS-2</td>
<td>KS1 KR ACP2 ACP3</td>
<td>NRPS-2</td>
<td>C A PCP3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PKS-3</td>
<td>KS2 PCP1</td>
<td>NRPS-3</td>
<td>C A PCP4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NRPS-4</td>
<td>C</td>
</tr>
<tr>
<td>Albi</td>
<td>343</td>
<td>XabC (AAK15075)</td>
<td>C-methyltransferase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AlbIII</td>
<td>167</td>
<td>ComAB (CAA71583)</td>
<td>Activator of alb genes transcription</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AlbIV</td>
<td>941</td>
<td>MycA (T44806)</td>
<td>Peptide synthase</td>
<td>NRPS module</td>
<td>NRPS domains</td>
<td>NRPS-5</td>
<td>A PCP5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WbgP (E83253)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operon 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbV</td>
<td>239</td>
<td>Thp (AAK15074)</td>
<td>No function (transposition)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbVI</td>
<td>286</td>
<td>TenP (AAA67510)</td>
<td>O-methyltransferase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbVII</td>
<td>765</td>
<td>HbaA (AAS8338)</td>
<td>4-hydroxybenzoate CoA ligase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albVIII</td>
<td>330</td>
<td>SyrP (AAB63253)</td>
<td>Regulation</td>
<td>NRPS module</td>
<td>NRPS domains</td>
<td>NRPS-6</td>
<td>A PCP6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dlb1† (CAB04779)</td>
<td>Peptide synthase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operon 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbX</td>
<td>83</td>
<td>MbtH (O05821)</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbXI</td>
<td>315</td>
<td>SyrC (U25130)</td>
<td>Thioesterase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbXII</td>
<td>451</td>
<td>BoxB (AAK060010.1)</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albXIII</td>
<td>317</td>
<td>hph* (AAK25601)</td>
<td>Esterase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albXIV</td>
<td>496</td>
<td>AckII-2 (p64105)</td>
<td>Albicidin transporter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbXV</td>
<td>584</td>
<td>hph* (p6390)</td>
<td>Carbamoyl transferase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AbXVI</td>
<td>88</td>
<td>OrfA (AAC30166)</td>
<td>No function (transposition)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operon 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albXVII</td>
<td>716</td>
<td>PabAB (CAC22117)</td>
<td>Para-aminobenzoate synthase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operon 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albXVII</td>
<td>137</td>
<td>ADCL (AAC06352)</td>
<td>No function (not functional)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albXIX</td>
<td>200</td>
<td>MsrG (P65530)</td>
<td>Immunity against albicidin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>albXX</td>
<td>202</td>
<td>Ubc (Q25660)</td>
<td>4-hydroxybenzoate synthetase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Protein accession numbers in Genbank are given in parentheses.

**PKS and NRPS domains are abbreviated as follows: A, adenylation; ACP, acyl carrier protein; AL, acyl CoA ligase; KR, ketoreductase; KS, ketosynthase; PCP, peptidyl carrier protein.

*Underlined domains are likely inactive due to the lack of highly conserved motifs.

### TABLE 4

Summary of results obtained from BLAST analyses

<table>
<thead>
<tr>
<th>Putative Alb protein</th>
<th>No. of Aa residues</th>
<th>Protein homolog</th>
<th>Origin</th>
<th>Genbank accession #</th>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>AbI</td>
<td>6879</td>
<td>Xanthomonas albilineus</td>
<td></td>
<td>AAK15074</td>
<td>1352 bits (3498)</td>
<td>0.0</td>
<td>730/730 (100%)</td>
<td>23/532 (4%)</td>
<td></td>
</tr>
<tr>
<td>PKS-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SsaB (1770 aa)</td>
<td></td>
<td>Myxococcus xanthus</td>
<td></td>
<td>AAC44128</td>
<td>231 bits (589)</td>
<td>2e-59</td>
<td>175/532 (32%)</td>
<td>23/532 (4%)</td>
<td></td>
</tr>
<tr>
<td>Putative</td>
<td>No. of</td>
<td>Protein homolog</td>
<td>Origin</td>
<td>Genbank</td>
<td>Score</td>
<td>Expect</td>
<td>Identities</td>
<td>Positives</td>
<td>Gaps</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>-----------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>PKS-2</td>
<td>XabB (4801 aa)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>3464 bits (8983)</td>
<td>0.0</td>
<td>1882/1882 (100%)</td>
<td>1882/1882 (100%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PkSM (4273 na)</td>
<td><em>Bacillus subtilis</em></td>
<td>CAI13603</td>
<td>887 bits (2292)</td>
<td>0.0</td>
<td>626/1806 (33%)</td>
<td>938/1806 (49%)</td>
<td>140/1806 = 7%</td>
<td></td>
</tr>
<tr>
<td>PKS-3</td>
<td>XabB (4801 na)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>1274 bits (3296)</td>
<td>0.0</td>
<td>653/653 (100%)</td>
<td>653/653 (100%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PkSM (4273 na)</td>
<td><em>B. subtilis</em></td>
<td>CAI13603</td>
<td>577 bits (1486)</td>
<td>0.0</td>
<td>293/584 (50%)</td>
<td>391/584 (66%)</td>
<td>17/584 (2%)</td>
<td></td>
</tr>
<tr>
<td>NRPS-1</td>
<td>XabB (4801 na)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>1934 bits (5010)</td>
<td>0.0</td>
<td>1035/1046 (99%)</td>
<td>1039/1046 (99%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nosa (4379 na)</td>
<td><em>Nostoc sp.</em></td>
<td>AF204805</td>
<td>618 bits (1584)</td>
<td>0.0</td>
<td>319/1104 (36%)</td>
<td>586/1104 (53%)</td>
<td>8/1104</td>
<td></td>
</tr>
<tr>
<td>NRPS-2</td>
<td>Nosa (4379 na)</td>
<td><em>Nostoc sp.</em></td>
<td>AF204805</td>
<td>416 bits (1069)</td>
<td>0.0</td>
<td>337/1127 (29%)</td>
<td>496/1127 (43%)</td>
<td>128/1127 (11%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Peptide synthase (5060 aa)</td>
<td><em>Anabaena sp.</em></td>
<td>COC1604</td>
<td>402 bits (1034)</td>
<td>0.1</td>
<td>315/1073 (29%)</td>
<td>470/1073 (44%)</td>
<td>114/1073 (10%)</td>
<td></td>
</tr>
<tr>
<td>NRPS-3</td>
<td>XabB (4801 na)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>1847 bits (4784)</td>
<td>0.0</td>
<td>997/1044 (99%)</td>
<td>1077/1044 (99%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nosa (4379 na)</td>
<td><em>Nostoc sp.</em></td>
<td>AF204805</td>
<td>610 bits (1573)</td>
<td>0.0</td>
<td>392/1069 (36%)</td>
<td>571/1069 (52%)</td>
<td>86/1069 (8%)</td>
<td></td>
</tr>
<tr>
<td>NRPS-4</td>
<td>XabB (4801 na)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>889 bits (2297)</td>
<td>0.0</td>
<td>468/468 (100%)</td>
<td>468/468 (100%)</td>
<td>20/468 (4%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nosc (3317 na)</td>
<td><em>Nostoc sp.</em></td>
<td>AAF17280</td>
<td>240 bits (613)</td>
<td>0.0</td>
<td>156/438 (35%)</td>
<td>220/438 (51%)</td>
<td>95/438</td>
<td></td>
</tr>
<tr>
<td>AlbII</td>
<td>XabC (343 aa)</td>
<td><em>X. albilineans</em></td>
<td>AAK15075</td>
<td>633 bits (1633)</td>
<td>0.0</td>
<td>343/343 (100%)</td>
<td>343/343 (100%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MtROI (326 aa)</td>
<td><em>Streptomyces averillicus</em></td>
<td>AAD55584</td>
<td>144 bits (361)</td>
<td>0.0</td>
<td>98/323 (30%)</td>
<td>154/323 (47%)</td>
<td>4/323 (1%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TernO (339 aa)</td>
<td><em>S. glaucescens</em></td>
<td>P08666</td>
<td>81.7 bits (199)</td>
<td>0.0</td>
<td>79/314 (25%)</td>
<td>140/314 (44%)</td>
<td>12/314 (3%)</td>
<td></td>
</tr>
<tr>
<td>AlbIII</td>
<td>comaA operon protein 2 (136 na)</td>
<td><em>E. coli</em></td>
<td>AAC74756</td>
<td>133 bits (335)</td>
<td>0.0</td>
<td>68/135 (50%)</td>
<td>89/135 (65%)</td>
<td>3/135</td>
<td></td>
</tr>
<tr>
<td>AlbIV</td>
<td>BA3 (6359 na)</td>
<td><em>B. licheniformis</em></td>
<td>AAO6348</td>
<td>361 bits (926)</td>
<td>2e-99</td>
<td>19/441 (43%)</td>
<td>267/441 (60%)</td>
<td>14/441 (3%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Whbg (377 na)</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>EB5253</td>
<td>81.6 bits (200)</td>
<td>4e-15</td>
<td>44/119 (36%)</td>
<td>70/119 (57%)</td>
<td>4/119 (3%)</td>
<td></td>
</tr>
<tr>
<td>AlbV</td>
<td>Tmp (240 na)</td>
<td><em>X. albilineans</em></td>
<td>nd</td>
<td>nd</td>
<td>0.0</td>
<td>240/240 (100%)</td>
<td>240/240 (100%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IS trnspase (260 na)</td>
<td><em>Yersinia pestis</em></td>
<td>AAC82714</td>
<td>160 bits (404)</td>
<td>1e-38</td>
<td>87/183 (47%)</td>
<td>122/183 (66%)</td>
<td>2/183 (1%)</td>
<td></td>
</tr>
<tr>
<td>AlbVI</td>
<td>Hypothetical protein</td>
<td><em>TemP</em> (276 na)</td>
<td>AAK46402</td>
<td>138 bits (347)</td>
<td>6e-32</td>
<td>92/224 (41%)</td>
<td>125/224 (55%)</td>
<td>18/224 (8%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mycobacterium tuberculosis</td>
<td>Pasteurella multocida</td>
<td>AKE03406</td>
<td>36.6 bits (83)</td>
<td>0.24</td>
<td>32/132 (24%)</td>
<td>65/132 (49%)</td>
<td>29/132 (21%)</td>
<td></td>
</tr>
<tr>
<td>AlbVII</td>
<td>4-hydroxybenzoate-CoA ligase (539 aa)</td>
<td><em>Rhodopseudomonas palustris</em></td>
<td>AAA62694</td>
<td>203 bits (513)</td>
<td>5e-51</td>
<td>156/492 (31%)</td>
<td>242/492 (48%)</td>
<td>31/492 (6%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SyrP (353 aa)</td>
<td><em>Pseudomonas syringae</em></td>
<td>AAB62523</td>
<td>182 bits (458)</td>
<td>5e-45</td>
<td>106/306 (34%)</td>
<td>153/306 (50%)</td>
<td>4/306 (1%)</td>
<td></td>
</tr>
<tr>
<td>AlbIX</td>
<td>XabB (4801 na)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>481 bits (1239)</td>
<td>0.0</td>
<td>258/608 (47%)</td>
<td>378/608 (61%)</td>
<td>23/608 (3%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dbh (1278 na)</td>
<td><em>B. subtilis</em></td>
<td>CABI15186</td>
<td>354 bits (908)</td>
<td>1e-96</td>
<td>341/608 (55%)</td>
<td>21/608 (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRP-6</td>
<td>XabB (4801 na)</td>
<td><em>X. albilineans</em></td>
<td>AAK15074</td>
<td>874 bits (2258)</td>
<td>0.0</td>
<td>515/1110 (46%)</td>
<td>682/1110 (61%)</td>
<td>52/1110 (4%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nosa (4379 na)</td>
<td><em>Nostoc sp.</em></td>
<td>AF204805</td>
<td>551 bits (1420)</td>
<td>0.0</td>
<td>388/1148 (33%)</td>
<td>583/1148 (49%)</td>
<td>84/1148 (7%)</td>
<td></td>
</tr>
<tr>
<td>Albx</td>
<td>Hypothetical protein 2 (72 na)</td>
<td><em>P. aeruginosa</em></td>
<td>AAG05800</td>
<td>75.6 bits (185)</td>
<td>1e-13</td>
<td>34/61 (55%)</td>
<td>44/61 (71%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MtsH (71 na)</td>
<td><em>M. tuberculosis</em></td>
<td>CAIB8480</td>
<td>59 bits (142)</td>
<td>9e-09</td>
<td>25/55 (45%)</td>
<td>37/55 (66%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>AlbxI</td>
<td>SyrC (433 aa)</td>
<td><em>P. syringae</em></td>
<td>AAB85161</td>
<td>34.4 bits (78)</td>
<td>1.9</td>
<td>23/93 (24%)</td>
<td>40/93 (42%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hydrolase (261 na)</td>
<td><em>S. coelicolor</em></td>
<td>CAA16200</td>
<td>34 bits (77)</td>
<td>2.9</td>
<td>19/60 (31%)</td>
<td>30/60 (49%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>AlbxII</td>
<td>BoxB (473 na)</td>
<td><em>Azoarcus evansi</em></td>
<td>AAO0599</td>
<td>293 bits (751)</td>
<td>3e-78</td>
<td>174/448 (38%)</td>
<td>243/448 (53%)</td>
<td>12/448 (2%)</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 4-continued

<table>
<thead>
<tr>
<th>Putative Abl protein</th>
<th>No. of aa residues</th>
<th>Protein homolog</th>
<th>Origin</th>
<th>Genbank accession #</th>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>AblXIII</td>
<td>317</td>
<td>Hypothetical protein (335 aa)</td>
<td><em>Caulobacter crescentus</em></td>
<td>AAK25001</td>
<td>99.5 bits (247)</td>
<td>5e-200</td>
<td>88/296 (29%)</td>
<td>125/296 (41%)</td>
<td>5/296 (1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plasmodium PAF acetylhydrolase (444 aa)</td>
<td><em>C. familiaris</em></td>
<td>AAC48484</td>
<td>37.5 bits (86)</td>
<td>43/156 (28%)</td>
<td>56/156 (36%)</td>
<td>44/156 (28%)</td>
<td></td>
</tr>
<tr>
<td>AblXIV</td>
<td>496</td>
<td>Putative transmembrane efflux protein (505 aa)</td>
<td><em>S. coelicolor</em></td>
<td>CAB09083</td>
<td>225 bits (574)</td>
<td>0</td>
<td>154/465 (33%)</td>
<td>240/465 (51%)</td>
<td>8/465 (1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AblF, putative abscisic acid efflux pump (406 aa)</td>
<td><em>X. albilineans</em></td>
<td>AF403707</td>
<td>736 bits (1900)</td>
<td>496/496 (100%)</td>
<td>496/496 (100%)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>AblXV</td>
<td>584</td>
<td>Probable carboxyyl transferase (585 aa)</td>
<td><em>P. aeruginosa</em></td>
<td>AAG08590</td>
<td>201 bits (513)</td>
<td>1e-50</td>
<td>158/458 (34%)</td>
<td>222/458 (49%)</td>
<td>39/458 (8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BinED (545 aa)</td>
<td><em>S. verticillus</em></td>
<td>AAG02370</td>
<td>192 bits (506)</td>
<td>1e-47</td>
<td>149/441 (33%)</td>
<td>209/441 (46%)</td>
<td>33/441 (7%)</td>
</tr>
<tr>
<td>AblXVI</td>
<td>88</td>
<td>Transposase (933 aa)</td>
<td><em>X. axonopodis</em></td>
<td>AF215343</td>
<td>64.8 bits (157)</td>
<td>2e-10</td>
<td>27/45 (60%)</td>
<td>38/45 (84%)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transposase of A (88 aa)</td>
<td><em>Desulfotibrio vulgaris</em></td>
<td>AAC31666</td>
<td>61.0 bits (147)</td>
<td>3e-09</td>
<td>20/34 (59%)</td>
<td>38/34 (109%)</td>
<td>—</td>
</tr>
<tr>
<td>AblXVII</td>
<td>716</td>
<td>Pan-aminobenzoate synthase (723 aa)</td>
<td><em>Streptomyces griseus</em></td>
<td>CAC22117</td>
<td>503 bits (1295)</td>
<td>e-141</td>
<td>302/699 (43%)</td>
<td>409/699 (58%)</td>
<td>36/699 (5%)</td>
</tr>
<tr>
<td>AblXVIII</td>
<td>137</td>
<td>4-aminoo-4-deoxycytidine lyase (271 aa)</td>
<td><em>P. aeruginosa</em></td>
<td>AAG06352</td>
<td>81.4 bits (200)</td>
<td>4e-15</td>
<td>46/105 (43%)</td>
<td>65/105 (61%)</td>
<td>—</td>
</tr>
<tr>
<td>AblXIX</td>
<td>200</td>
<td>MchG (187 aa)</td>
<td><em>E. coli</em></td>
<td>CAA30724</td>
<td>60.5 bits (145)</td>
<td>9e-09</td>
<td>36/141 (25%)</td>
<td>58/141 (40%)</td>
<td>5/141 (3%)</td>
</tr>
<tr>
<td>AblX</td>
<td>202</td>
<td>4-hydroxybenzoate synthase (202 aa)</td>
<td><em>E. coli</em></td>
<td>AEC77009</td>
<td>45.6 bits (107)</td>
<td>5e-04</td>
<td>42/161 (26%)</td>
<td>21/161 (13%)</td>
<td>—</td>
</tr>
<tr>
<td>AblXXI</td>
<td>278</td>
<td>XabA (278 aa)</td>
<td><em>X. albilineans</em></td>
<td>AAG28384</td>
<td>430 bits (1106)</td>
<td>0</td>
<td>278/278 (100%)</td>
<td>278/278 (100%)</td>
<td>—</td>
</tr>
<tr>
<td>AblXXII</td>
<td>634</td>
<td>Heat shock protein IlpG (634)</td>
<td><em>P. aeruginosa</em></td>
<td>AAG04985</td>
<td>1051 bits (2688)</td>
<td>0</td>
<td>523/634 (82%)</td>
<td>588/634 (92%)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heat shock protein IlpG (624)</td>
<td><em>E. coli</em></td>
<td>AAC73575</td>
<td>743 bits (1899)</td>
<td>0</td>
<td>376/624 (60%)</td>
<td>476/624 (76%)</td>
<td>4/624 (0%)</td>
</tr>
</tbody>
</table>

---

TABLE 5

Comparison of conserved sequences in C domains of peptide synthetases and in putative C domains of the Abl modules

<table>
<thead>
<tr>
<th>Core synthetases*</th>
<th>Sequence</th>
<th>Abl module</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>SxQxQxR(L/M) (W/Y) 5L</td>
<td>TYAQERLWLV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>STAQERKPL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SYAQERLWLV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SLFQRERLWPV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SYQQRERLWPV</td>
</tr>
<tr>
<td>C2</td>
<td>RHEQLRTxF</td>
<td>RHVLRRTRP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RHAVLRTXP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RHEILRRTP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RHRILRRTP</td>
</tr>
<tr>
<td>C3</td>
<td>MHHEISDG(W/Y) S</td>
<td>IHHIISDGNS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IHHIVPDGNS</td>
</tr>
</tbody>
</table>

---

TABLE 5-continued

Comparison of conserved sequences in C domains of peptide synthetases and in putative C domains of the Abl modules

<table>
<thead>
<tr>
<th>Core synthetases*</th>
<th>Sequence</th>
<th>Abl module</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4</td>
<td>YMD(P/Y)AVQ</td>
<td>YADYAH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YADYAR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YADYAM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>YADYAN</td>
</tr>
<tr>
<td>C5</td>
<td>(1/Y)GxQVNT(Q/L)</td>
<td>IGFQPHILPLR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IGLFVPTLAVR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IGFQPHILPLR</td>
</tr>
</tbody>
</table>
### TABLE 5—continued

<table>
<thead>
<tr>
<th>Core synthetases*</th>
<th>Sequence</th>
<th>Alb module</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6</td>
<td>HQSVPFE</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>HQSVPFE</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>NQALPFE</td>
<td>NRPS-4</td>
</tr>
<tr>
<td></td>
<td>HRALPFE</td>
<td>NRPS-7</td>
</tr>
</tbody>
</table>

*Sources from Marahiel et al., 1997

### TABLE 6

<table>
<thead>
<tr>
<th>Core peptide synthetases*</th>
<th>Sequence</th>
<th>Alb module</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>L(T/E)YxEL</td>
<td>NSYQGQ</td>
</tr>
<tr>
<td></td>
<td>NSYQGQ</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>LSYQGQ</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>NQYQGQ</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>PSYQGQ</td>
<td>PKS-4</td>
</tr>
<tr>
<td></td>
<td>LSYQGQ</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A2</td>
<td>LKAGxYHL(V/L) P(L/I) D</td>
<td>FKAGACTVQID</td>
</tr>
<tr>
<td></td>
<td>FSGAASVLID</td>
<td>NSYQGQ</td>
</tr>
<tr>
<td></td>
<td>NSQAGAIVPQD</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td></td>
<td>LSLQGFLVAPID</td>
<td>PKS-4</td>
</tr>
<tr>
<td></td>
<td>LKAGGCTVQPD</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A3</td>
<td>LAYxYTSG(S/T) TQxPKG</td>
<td>LACVNMRTGSGTRPKG</td>
</tr>
<tr>
<td></td>
<td>?TRTMVEGSGLSSLEI?</td>
<td>PKS-4</td>
</tr>
<tr>
<td></td>
<td>PTVQICMTGSGTRPKG</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>PVNICTSGSTGPKA</td>
<td>PKS-4</td>
</tr>
<tr>
<td></td>
<td>LAYVNYTSGSTGRLPKG</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A4</td>
<td>FDxS</td>
<td>FAVS</td>
</tr>
<tr>
<td></td>
<td>FAVS</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>FDAE</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>FDCT</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>FAYG</td>
<td>PKS-4</td>
</tr>
<tr>
<td></td>
<td>FAIS</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A5</td>
<td>NXYGPTX</td>
<td>NNYGCTE</td>
</tr>
<tr>
<td></td>
<td>NNYGCTE</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>?AYGNAE?</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>NEYGPTX</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>DGYGCTE</td>
<td>PKS-4</td>
</tr>
<tr>
<td></td>
<td>YLYGCTE</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
</tbody>
</table>
### TABLE 6-continued

Comparison of conserved sequences in A domains of peptide synthetases and in putative A domains of the Alb modules

<table>
<thead>
<tr>
<th>Core peptide synthetases*</th>
<th>Sequence</th>
<th>Alb module</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6 GELxIxGxG(V/L)ARGYL</td>
<td>GELAHMYSGMARGMYW</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>np</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>GQIHGGAGAIGGYV</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>GSLVWVGRHTLTRGLYV</td>
<td>FK5-4</td>
</tr>
<tr>
<td></td>
<td>GEVHIESLGITHGYW</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A7 Y(R/K)TGDL</td>
<td>YKTGDM</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>?YKTIAL?</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>YASGDL</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>?PFTIAL?</td>
<td>FK5-4</td>
</tr>
<tr>
<td></td>
<td>YRTGDM</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A8 GRxDxQVKISxGTELGEIE</td>
<td>GRQDFEKVRGHRVDTQVE</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>?GSLDIVGQRRIDPRIDLCYVE?</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>GRKQ6xQKLRGTYELGEIE</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>?GRMGxAIKINGCHLSPLTEL?</td>
<td>FK5-4</td>
</tr>
<tr>
<td></td>
<td>GRMDVEVRHGRYRUVRQVE</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A9 LPxYM(I/V)P</td>
<td>LPTYMLP</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>?LPDYLPLP?</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>LPEYMLP</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>?LQHHTYP?</td>
<td>FK5-4</td>
</tr>
<tr>
<td></td>
<td>LPTYMLP</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
<tr>
<td>A10 NKxGx(1/L)DR</td>
<td>NKKLRD</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td>?GHRVDL?</td>
<td>NRPS-2</td>
</tr>
<tr>
<td></td>
<td>NKKVNR</td>
<td>NRPS-5</td>
</tr>
<tr>
<td></td>
<td>?GKVIIR?</td>
<td>FK5-4</td>
</tr>
<tr>
<td></td>
<td>NKEDT</td>
<td>NRPS-6 and NRPS-7</td>
</tr>
</tbody>
</table>

*Sourced from Marahiel et al., 1997
*Non conserved sequences
np: not present

### TABLE 7

Comparison of conserved sequences in PCP and TE domains of peptide synthetases and in putative PCP and TE domains of the Alb modules

<table>
<thead>
<tr>
<th>Sequences conserved in peptide domain synthetases*</th>
<th>Sequence</th>
<th>Alb module</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP DxFxkLDG(R/D)S(L/I)</td>
<td>D-FPAVGGSHSL</td>
<td>FK5-3 (PCP1)</td>
</tr>
<tr>
<td></td>
<td>DFSFALQGHSL</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
</tbody>
</table>

### TABLE 7-continued

Comparison of conserved sequences in PCP and TE domains of peptide synthetases and in putative PCP and TE domains of the Alb modules

<table>
<thead>
<tr>
<th>Sequences conserved in peptide domain synthetases*</th>
<th>Sequence</th>
<th>Alb module</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP DxFxkLDG(R/D)S(L/I)</td>
<td>D-FPAVGGSHSL</td>
<td>FK5-3 (PCP1)</td>
</tr>
<tr>
<td></td>
<td>DFSFALQGHSL</td>
<td>NRPS-1 and NRPS-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(PCP2 and PCP4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DFSFELQGHSL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(PCP3)</td>
</tr>
</tbody>
</table>
### TABLE 7-continued

<table>
<thead>
<tr>
<th>Sequences conserved in peptide main synthetases*</th>
<th>Alb module (domain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNPFEGGHGNS NRPS-5 (PCP5)</td>
<td></td>
</tr>
<tr>
<td>DNPFEGGHGSSL NRPS-6 and NRPS-7 (PCP6 and PCP7)</td>
<td></td>
</tr>
<tr>
<td>TE G(H/Y)SxG GWSG NRPS-7</td>
<td></td>
</tr>
</tbody>
</table>

*Sourced from Marahiel et al., 1997

### TABLE 8

<table>
<thead>
<tr>
<th>Position in GorA (Phe) and variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domains</td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Alb NRPS-1</td>
</tr>
<tr>
<td>Alb NRPS-2</td>
</tr>
<tr>
<td>TycE-M1 (Pro)</td>
</tr>
<tr>
<td>VirS (Pro)</td>
</tr>
<tr>
<td>HVCL</td>
</tr>
<tr>
<td>Alb NRPS-6</td>
</tr>
<tr>
<td>Alb NRPS-7</td>
</tr>
</tbody>
</table>

### TABLE 9

<table>
<thead>
<tr>
<th>Donor</th>
<th>AM12</th>
<th>AM13</th>
<th>AM36</th>
<th>AM10</th>
<th>AM15</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEV439</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pEVlaxXXII</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pEVHprG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pALB539</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pUFR043</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>none</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

+: restoration of albicin production by alb² mutant, - : no complementation.
All experiments were performed at least in duplicate with at least 2 exconjugants obtained from two independent tripatal conjugations.

### TABLE 10

<table>
<thead>
<tr>
<th>Bioassay medium containing</th>
<th>No antibiotic</th>
<th>Tetracycline</th>
<th>Kanamycin</th>
<th>Tetracycline + kanamycin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester strain</td>
<td>DH5αKT</td>
<td>DH5αAbKT</td>
<td>DH5αKT</td>
<td>DH5αAbKT</td>
</tr>
<tr>
<td>pUFR043 and pLAFR3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>and pOp3-4/XALB2-3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>pLH571 and pLAFR3</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+: presence of a growth inhibition zone
All experiments were performed at least in duplicate with at least 2 exconjugants obtained from two independent tripatal conjugations.
REFERENCES


---

**SEQUENCE LISTING**

```plaintext
<160> NUMBER OF SEQ ID NOS: 54
<210> SEQ ID NO 1
<211> LENGTH: 55939
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albidelines

<400> SEQUENCE: 1

gattcttttc gcacattgccc ggattgtatg acctgcctcg gagttcttg gagcctcttc
60
ggacctgtgc tcgtgtcgtc ctactgctgc gatgctatcg cgccgcaacc caacccgttg
120
gctactccgg tggaccgcat tggacgcatg ggcacagtgc ccgctagctc gcgcaagtgc
180
cgagctacgc gcgtgtcgct cgctcttttt tgcgaacgcg cggctgcgta gcggcggtga
240
cagcgccatga aagctccaa tggagggcct gacggcaggg cggccccggc tattcactcat
300
ggcgcttaca aacccgcgtg ggttccccag ccagggaccg ttttggaggg cggggttgcg
360
tgcctgtgc gcacgcgtgtc gcggtgctttc ctgcttttgg ccaacgcctg gtgtggtcgc
420
cctcggcacc cagctcgctc aacaaggtca gtaggcgtat gcctttcctag gtcaccagac
480
ttgcggcgca cggagctgcg cagctggcct ttaatccttg ggcagctgcc ggtaggctgg
540
ttcggcgctg cgctgcaacg tgactgctgta ctaatcactg gcggctgccgt ctgattagac
600
gccgcttaaa ttcggccgct ggcctatgct gcggccgacg ttgctctgg gctgcgctgg
660
ggcgctgccg ctggccggtgc cggagctgcc gccccttcct gccgctgctg cgtgctgctg
720
tgcgctgttg gcagctgctg cggagccgct cgggtttcct gccgctgctg cgctgtgtcg
780
ccgaaagttt atgctcgctg ggggcggctg ccggccgctg ccggtttgtt cggcgcgcct
840
cggagctgct gcggccgctg ctggccgctg ctgctctgg gctgcgctgg cggagctgcc
900
cagctgctg ccgagccgct gcggccgctg ccggttttct gcggccgctg gccgctgctg
960
cgcagataata ccgatctccg cgcctgctc ccaactaac cagctgtcct gcggctgccg
1020
ggtctgtccgc gcctgggctg gcgcgtgctga ggcggccgct gggaaatgtt ctggcccatt
1080
acgtgcggct ctggccgctg cgcctgctc cagctgctc gcgtgctgc gggtggtgtt
1140
tgcgctgctg ccgagccgct gcgtgctgc gcggccgctg ccggttttct gcggccgctg
1200
cggcgcggct gcggccgctg ccggttttct gcggccgctg ccggttttct gcggccgctg
1260
cgcagctgct gcggccgctg ccggttttct gcggccgctg ccggttttct gcggccgctg
1320
```
ttggcggcata ttccgcgcag gcacacaggg aaggtgcgcgc aagataaccaac cgggagagcct cggagtggcga 1380
tgaagctgga tttttacaag atcagcagagc ccccgccgagc aagcgcctgtg cttgctgatgc 1440
ccgctgggtct gtcgctgcct aggagcagtt tgaataaacc cttcaccgc tgcagagaaga 1500
aatagaaacc attgtaaatg tagtgtaaga taacagcggga taacaacaaac taacagcgca 1560
cctcaacgg caacgcagcag cgcgggtgta cagagtgatrac ataaaggaacctt 1620
cgagcagacgc aagccgcccgt cttgccggaac cgcagctgatt ttgtagctgct cctaaagcgc 1680
agattggtatt taactgcagcgc atcactataac ggaatggtatg ttgtggaaacg ataaaccaagtta 1740
tggcgctgct cgcgtgctgcgt gatgctgtgcgc gtcattggtgc gcaggtgcaag acgttgctgt 1800
cgctgctgtcag tcctgctgtgcgc gtcgctgctgc gcgatggtgcgc cagctgtggatc 1860
gtagctgggc gcggagccgg ccggttgttgct cgcagcgata tcccgcgtcgt cccaagtgttga 1920
tgcgctgctcag ctcggcggtgct ggtgtggagtt ttgaccagaa agaggttccca gaagtcagcgc 1980
tgcccgcacat ccgctggcgtcgt ggtgtggagtt ttgaccagaa agaggttccca gaagtcagcgc 1990
tgcgctgctgct gatcactgcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2040
gcagcgtcgc gcagacgcagaa cctgagcgtg ctcgtcagcag tatttccgg acgtcatgatgc 2100
ttcgaagctgt gcctgtccgag cgggaatttct gcgtcagctgc ggtgtggagttttcctgcagc 2160
ttcgaagctgt gcctgtccgag cgggaatttct gcgtcagctgc ggtgtggagttttcctgcagc 2220
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2280
gctgctgctcag ctcggcggtcgc gcgatggtgcgc cagctgtggatc 2340
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2400
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2460
gctgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2520
gcagcgtcgc gcctgtccgag cgggaatttct gcgtcagctgc ggtgtggagttttcctgcagc 2580
ttcgaagctgt gcctgtccgag cgggaatttct gcgtcagctgc ggtgtggagttttcctgcagc 2640
cggctggcagt gcgcagacttactgactgtatgc gctgtccgtagc 2700
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2760
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2820
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2880
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 2940
agatgggttgga caggtactgac aagctcgcgt cgcctcggctc acgtggcgag 3000
tgcgctgctcag tatttccgg acgtcatgatgc cgtctcagcag cggatcctttt 3060
atggcagcttg tccaactcactgt gcctagctgtgctgcc 3120
atgcagctgtg 3180
atgcagctgtg 3240
atgcagctgtg 3300
getcgcgcctg ctcgctgtcgc ctcgctgtcgc 3360
atgcagctgtg 3420
atgcagctgtg 3480
atgcagctgtg 3540
atgcagctgtg 3600
atgcagctgtg 3660
gegcatatca gctggccctt tacatggcgt gttttgcaaa cgggttgagg atgacccgac 3720
tctgttcttct ttcgacccat attacattgc cggtgacatt cggattggaa gcagaagctg 3780
cgggtcagag ggtgcgagct ccgacgtacat atgatggata ttcacgcagc gttttcc cac 3840
ttagcaccg aggtagggcg cttggcagcgc ggtgcgcgca cgcctgctgg caaatctaat 3900
cggtgaccag gttttcgcgt gaggcgctag gctatatcgc ttagctccgt gggccaaactc 3960
tgctgttccga tctccgagct tataattttg atgacccgac ggcacgacgt aacgactgtgc 4020
tggagtgccg tccagcgcgt ctatgtgcgt ttgagcgggt ctgcgaagaa aagctgtgatc 4080
acacgctgct gttagtcgcgc cctttacccc ttcgagccga cccacggtct actgtgtgac 4140
ccgccgaacc gcacactgtt gcgcgtcacaag gttggtgataa ccactcgagc ccgggccaca 4200
agttcgccgc gatgccacgc tctagggcg ccgacgcgac aacactggct gcagctggtat 4260
cggtgtctga aacgctgttgct tatacggcag tgcaggaagc aaggggaaat actttttttaac 4320
atatctaatg ctagtgctgc gttttacgcgt gcctttcagc agggctacgt tataattttg 4380
gcagtcttta ggcggcggttat tatacggagtag tgcaggaagc aagggggtaat actttttttaac 4440
cggtgtaagtc ggaggtgcacc aacaagcttt atccgctgct tctagcgcgg ctctgtgctg 4500
atgggctgca gtagcgtact tattgggtca aagaggtgca aagggggaag gattttataagc 4560
atgggctgca gtagcgtact tattgggtca aagaggtgca aagggggaag gattttataagc 4620
ccgctgcgtag ccgtagctgg ccggccgccc ggcagcgcg ccgagactgg ctctgtgctg 4680
caagagaggg gcagagaggg ctccgctgtg ttcctgagtc ttcagctcgg cccgctgctg 4740
ccgctgcggt cccgtcctgc aaggtggtcc gcccgtcgtg aaggtggtcc gcccgtcgtg 4800
tgctgcaatg gcggttaaggt cagggggtta ccttt tgtgtaagtt cggcagctgg 4860
cccattgcgca cagctcgtcct cccctccctg cccctcctcg cagagccgga agaatttggg 4920
acceptable aaccgaatgg cccacccctg ctcctcctcg aaggtggtcc gcccgtcgtg 4980
aatgacgatg cagactgaaat ggtgggtcag cctgggaatt aagggggaag gattttataagc 5040
ggacttgagc tttaaatgatg tgaagcgtggt gattctgtag aatataataa ttaccacgtag 5100
tgctcaaggg ataaatcttc cccctccctg cccctcctcg cagagccgga agaatttggg 5160
tgactotat ccagggagtc ggcgtcgcct tccacggctg cttgttgagtc cggcagatgc 5220
tgctgcgctt gcgcgtcgcct atgacgagtc accaagctgc tgcgcagctgc gtttttgcgt 5280
tggtcagata tttttcatacg cggacaagtt ttcacagggg aatataataa ttaccacgtag 5340
atgacgatgc cagctcctgg ggcgggtgat cttgccttcag cccctcctcg cagagccgga 5400
catatactgg gcagaatctgg tcggccgcag aaaaataattg aagggggaag gattttataagc 5460
tcagaagatt gccagaagtt ttcacagggg aatataataa ttaccacgtag 5520
aatatggcgc cgggttcgctct cgccagcagc gccatcttgc aatattttgc aatattttgc 5580
aatacctgcatt cgggtgcggt aatggggagc cggggtgggt aagggggaag gattttataagc 5640
ggtgttgggt cggtgttgggt cggggtgggt aagggggaag gattttataagc 5700
cttctctctg atgggtaaac ccgagagagc cgggtgcatg ttcacggtgg 5760
aatacctgcatt cgggtgcggt aatggggagc cggggtgggt aagggggaag gattttataagc 5820
aatacctgcatt cgggtgcggt aatggggagc cggggtgggt aagggggaag gattttataagc 5880
aatacctgcatt cgggtgcggt aatggggagc cggggtgggt aagggggaag gattttataagc 5940
aatacctgcatt cgggtgcggt aatggggagc cggggtgggt aagggggaag gattttataagc 6000
aatacctgcatt cgggtgcggt aatggggagc cggggtgggt aagggggaag gattttataagc 6060
gtgcaggaat tcaagagaga ccgtgtcagaa ctgacctgtg gcggcgatag ccaggttggc 9460
caatcctgac tgggtgcc gcctgcatacg ggcattgca tacacgtcgc gtcgcaactg 8520
gctgaaattgt gtgcggggcg ccaacatcag ctggatgccg tcagcgctga gcacactggc 8580
cacaaactcg tacgcttgcc gatactgtgt cctggtggcg gcaagcccca taacctttca 9640	
tttctgctgc gacaggggctc cgaagcggca aacotcgcc agoactacgt aagagaacac 9700
cagctgctgcc caacgtctgg tcctgctcttg ttcgatgcgc gtaagcttcc cctgttgcgt 9760
gttgtacgc ccagcagcat tttttttttcccc cattgocgtt acaatgctgca 9820
cctgcaaggg ctctgtcgacg agccgagccac cgcctgtgggt aagttgtgtct cgttaatcag 9880
gtgctgtggct gctgccagtg caggtctggcg cattctggtc aagctgtatt cagaccgttg 9940
tccggggtgc tccacacgcg cccgtgcgca aaccgtgcaag aagctcttggt ctgcatatg 9960
tcggcgtgac gcagctgccc gcctcgatcc gcggcttcag cctggttggc ttcgagcctgg 9980
gccggggggc agcgcgtgag aggctctcgg cgctctccaa tgcagctgag 10020
gggcaccagc gcgcgggtcg cagcatcagc tcggctggcg gcggcctggc aagagaaaat 10040
tccatcagta tcggctgcttc aagctgcaag gcggctcgg gtctccgggt gctgcagcg 10060
agccgcaagc aacgctgtgg cgccgagtgac ctaccaccc cggtccccgag aacgctgggc 10100
aaacggcgag cttcgcccttgc gcgcaagcct cagctgctgc ggctttctgg ccagttcgat 10120
tcggagccga gggtgcgctgg ccggcaccag cccaggtcgc gcggagttgg gttggcagcc 10160
ccggcgcgc gcgcgagcctt gttttcgcgt gcggaggagc gatagctggc ggcggtgcag 10180
agtcgacgcc cagcaacgcag tgcgatccg acggattggt ccgctcgtg cgcggtcagc 10220
agcgcggcct ccaagcattg gtttttttct gtttggcagg ggcggtgagc ccagctgctg 10260
ttcggtgctgg ccggagcacc cggccgcttc cagcagctgg gcggggggcc gccggcgggg 10300
agcgcggcgg cggcgggagg ggtgcgcggc cggcgcggg cggcggcggc cggcggcggc 10340
ggtactgctgg ccggcgg gccggagggg cggcgggagg cggcgggagg cggcgggagg 10380
taccacgcag aacgcgtcgc cgcggcgagc ttggagtcagc gacgctgcgg gcgggctgca 10420
tacaccgcttg gccggtgctgg gcggagttgc cgcggagttgc gcggagttgc gcggagttgc 10460
ttggctgacgc ccggcgcgcag cgccgcgtcg acggtgcatg ccgggcgcgc gcgggcgcgc 10500
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10540
gtgcgcatgc ccggagcggcc gggacacttc aagttaaac ccggcgcgcgc gcgggcgcgc 10580
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10620
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10660
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10700
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10740
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10780
agtcggaag gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg gcgcgagcctg 10820
cgggtggttg ccgtgcgcgg ccgcatccce agctgctcct tgggtttttg catgttgtgt   10860
gcctccattc cgctttaggc cgcttcggcc ggtgcgcatgg tgcaggttag aaaaatgcag  10920
atgtcggcct ccaacgcgcg ctcgaagcgc gtcggttcgc aagccacgag atcgcgtgcc   10980
gctgcgcca cgagcgtcgg tctgcctcag gggataaagc tgtagagcgc gtagtcgtgcg   11040
gatggggca cccgacgcgc ttccagctga cctacatgac ctcagcgctg ggcggtgcgc   11100
aacatacttt tcgcgcgcoca ttcagctccttt tcaacgacga tgacgacgac eggtaacctg   11160
acttggccga gggaattcctc gcgtcatatc aggtgcgaat cccggcacga ccggcaccacg   11220
gcgccagcgc gcgcacagcg tcggagcgcg ccgagcctcgta tcgggtactcg atttggtagc   11280
acccgcggcc tcgcggcgcc ttttgccgcgc tggtccggcc cagtcgcacaa atcgcacatg   11340
aaaccgcgta gcggccgcgc cccggccgatc gctaaaggcc tgcagcgcgc gcagggtagtg   11400
cgcacgcgct ctcagctgatc ggctgtaatg gcagggccga actgcgcctag gtcgcgtgac   11460
gtatcgatca ccggcgaggtg tggcgcggggtg cggttctcata gattctgata gctgtagcc   11520
agttgcagct gcggccggtgc tgcgcggcgca aagggcagat aacggctgtgc   11580
cgctgtagtt aagtgctgcgc cggcgagaaa aattggcgcgc cggccgctggt gcgcgcgtcg   11640
ataccgcctc ggcggtttcgt atccggcgc gcagccgacgc gcggcacgctgata   11700
taacgctcc caagggcagc cggccgctgc gcggcgcggca gatacagcgc ccgccgcggctgc   11760
accacccgctc ttgcatcacc atcgcttttc agtttttggc aacccacatg gctatcttag   11820
cgtacccgaa gaaaagctga aaaaagcctg gcgcttacgc ccggcgcggc cccggtgccga   11880
ctgccggcc gcgctgctgag ccaaccagcg gcgctgcctg ctgacctccg cctggtcctcc   11940
aaccgctgcgc gcgctgctgag atcctgcgc gcggcgcggca ccgctgagcg ccgctggccg   12000
tggtcagct gcgcgggtgtag gcacacctag tggacaggca cggcggcttg ccggccggga   12060
tgcgggttgga aaccgcgctgt cgccgctgctg gcgtgtcgcgc ccgggcacag ccggcgcgctg   12120
tgtgattcc gcgcggcgaac cggacgcgcgtg tgcagctcctc ggcggttatgc ccaaccacag   12180
 gcacgtggcgc gcgctctgcgc gcggctgtgcg gtcgtcgtgc gcggctgtgcg gtcgtcgtgc   12240
tcgagccgcc ccctcagctcg gcgggttcgc gcggcgcggc cggcgcggc cggcgcggc   12300
 tgggtcagcc cgccggccggcc ccggccggcc gcggcgcggc cggcgcggc cggcgcggc   12360
agccatatga taagtttttg tgcagctgcag gcggcgcgcgg ctcgcaatc gtcgctttttc   12420
gcggctgctg gcgctctgcgc gcggctgtgcg gtcgtcgtgc gcggctgtgcg gtcgtcgtgc   12480
tccagcggca atcattatag cagggttccc gcctacatca tggagcgcggc ggtggtgtgcg   12540
aaccgcggaa aagccgatcg gtcgacgcgc ttaaatctgcg tgcgcgatct ccgggttttcgcg   12600
aagcagaccc aatggtatgc gcggagcgctgc cggcctgttg gcgttagtgg aagggccgcgcg   12660
acggcggaggt ttcggtcagct gcgctctgcgc gcggctgtgcg gtcgtcgtgc gcggctgtgcg   12720
cggcgcggc gcgctctgcgc gcggctgtgcg gtcgtcgtgc gcggctgtgcg gtcgtcgtgc   12780
 tctgggtttcc cgcctgcgcgc gcgcggcgcggc gcggctgtgcg gtcgtcgtgc gcggctgtgcg   12840
 tggggtttcc cgcctgcgcgc gcgcggcgcggc gcggctgtgcg gtcgtcgtgc gcggctgtgcg   12900
 aacggagctgc gcggcgcggc gcgcggcgcggc gcggctgtgcg gtcgtcgtgc gcggctgtgcg   12960
 aaacacagcg gcggtgcagtc gggtggccatg aatgacatgg gcggctgtgcg gtcgtcgtgc   13020
cggcggcggc gcgctctgcgc gcggctgtgcg gtcgtcgtgc gcggctgtgcg gtcgtcgtgc   13080
 ttctgtcgcgc gcgctctgcgc gcggctgtgcg gtcgtcgtgc gcggctgtgcg gtcgtcgtgc   13140
acgggtagg aagtcaagtgg cctgaccagc ggcgctgtgtg gatacaaggtc gatattgtaa 13200
ggaatgtaca tgcggttcaac ggcgctgtgtg gcatctcgggt aacggccgcttc tgcgaaccgg 13260
tgacgccgat gaaagcgtag atcgcgtggtc ccaagccaccgc ggcgatttga aacagtggtc 13320
ggcgagaacgc atgcgtcagc ccggccac.gc gcacagcggg gtcgtggtcata gccaactgca 13380
cgccagggac gtcagggcgtg aagttgctt aacggctgtgg ccagatcagg gttagggtggtcc 13440
tgacccaggc cggccagcgc aacggccgaca gcagattttgc gtctccctccgc tgcggcacgc 13500
ttgtcgagga gaacaccaag ttgagctgtg gatgagtcag ctcgaggat cggctgagttt 13560
cgcggagacgc gtcgccgggt aacgcttgcc gcgcgttaattt gctcaggctg aagcggcagct 13620	ttggtcggct tgtcggcaacon gtcgctgonea cctggaagtaa gataggcaag gcgcaacgt 13680
cgagttcggc ttttacagag ctgcggagcgc gggcggcgcac gcgcacccaca gcctagtgcc 13740
tggttgtgac gcacagccagc gaacaggtccat gacggcccaaa ttgctgacgcc gcaacacacc 13800
cggaccaatgg gcactgggctg tacaagtgtg gttggcaggat gcgcacggag gtcgctggttccc 13860
ggcgagtaac atcgagacca gtcggtcgtt cgcgtgattg ccgctaggg aacgccagttg 13920
atttccagac gcacacccatg gcagcgtctg ccagcagcctggct gtcacagcaga 13980
agtagcagcgc gcagcgcagcct gcgtttgggtg cgtattggcca gcgcagtgcgc 14040
aatcagccg aacggtcctg ccagctgttg tcagattgtgg cgcacgctttt acccctgcaag 14100
gcgctgcgct cagggccgac gcattctctgt tttcgcgtata gcgcggcgaag cacacacggc 14160
ccatgttggc tctatacctccag ttcggagcac gcgcaacagc cggtggctcagc 14220
aaggctgtgg cggccccggc aacccctcagc cgcctagctgc ttcgagccacca ccctggtgtttg 14280
agasgctctgc gtcggccttaa ttcgttctgct cagatgctgc gcgcgatttt aaggcgctgtg 14340
cyggctgcgt ctcgcgtcag ctcgctcgttc acgtgtcctt gccgctcgcac gcgcctcctc 14400
tgacccggcc ggcggcgcaca tcgagcttg cgcgttatgc gtcagctgcgct 14460
tgatcagcgc cagataacgc tgcctctcgc aagctcatttg aagatacaac cgcggaaggtcc 14520
gtcgaagac ggagaaggtgtc tacatacgcga cagatgattgc gatgaggtcgc aaggaagcg 14580
aatcataccgc agtctactcgg gctgtttggg cgcctgttgcc atacaataata attacgattgc 14640
atgatgcttc gtcgctgccag cgcgttccagc tgcgctgccag ctcgccagcataa cggcgctttc 14700
attccagggt ctagctgatcc aggtgggcttg tcggctgtcagc gcgcaagattc 14760	ggtgtagctg caagagctgc ggcaggtcgg gggcgcaccc caagaggttt ttcctggggt 14820
gggtagaaaat tcagctgacgc ccgggtttgt gcgtggcttgc gcgcttttc 14880
gtctcaactcg gtcagccttg aggctatattg ggcggcctgtg gcctgtccatgc atccggcttg 14940
tgcacagcgc cagaacggat ggcgcagaaa ggcgcacact ggcagtatttt cctggcagtct 15000
gccagaacg gtagaactttg ggcacgaagtc gggcggatttc cgcgcttttt ttcgctgagttg 15060
cgggttaggc caggtcagctgc gggtgtggttt cggccgactc cgcctgggttt 15120
gtgcggtgac ggcggagttg cgggtggagt ggtgggattgg ccgtctgggttctgcagttgcag 15180
tgtgtctggtg tcggttcgcgt gcgctgggctg gtcggtcgtt ggcgcagtttc ttcctggggt 15240
cacccagcgc ggcgggctgtg ggcgcttttt gcgcgtctttgc gtcgcagagttt 15300
tcgctgtaat gtcgctgttgtt gcgcgcttttt gtcggtgcagc gcgccttggtg 15360
cgggtgacgc gggcgcgcttc gcgcgcttttt gcgcgcttttt gcgcgcttttt 15420
tgcgaggtgg cggggtgcagc gcgcgcttttt gcgcgcttttt gcgcgcttttt 15480
ggctggctgcc cagacggtgcagc gcgcgcttttt gcgcgcttttt gcgcgcttttt 15540
cgttcgcgc ccgcagatgc tgcaccgcgc ggtacgtcgc atgcgtgcagc ggaagatcaca 15600
gtgcggtgag ccggcggagag atgcgttctctgc gcgcctgttgc gcgcgtgacaa aggactctgtt 15660
cgcgtgtgc gcggcggacgg atgggctgcgcc tattggtgcct gcgggcaagct 15720
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 15780
cggcggggct gcgtggtgcgc gttgctgatgc gttgcgctgc gtagctgcgc 15840
cagcgcgcgca cgcctcgcggt ttcgccccgctg ccgctgtgcgct ctagagcgct 15900
cagcgcgcgca cgcctcgcggt ttcgccccgctg ccgctgtgcgct ctagagcgct 15960
cagcgcgcgca cgcctcgcggt ttcgccccgctg ccgctgtgcgct ctagagcgct 16020
cgataaggag gctctcgtgct gcggtcgcgtgc ccgtgacgacatggtactgc 16080
cagcgcgcgca cgcctcgcggt ttcgccccgctg ccgctgtgcgct ctagagcgct 16140
cgataaggag gctctcgtgct gcggtcgcgtgc ccgtgacgacatggtactgc 16200
tgcgcgacccca gctgtgcctgc ccgctgtgcgct ctagagcgct 16260
tgcgcgacccca gctgtgcctgc ccgctgtgcgct ctagagcgct 16320
tgcgcgacccca gctgtgcctgc ccgctgtgcgct ctagagcgct 16380
cagcgcgcgca cgcctcgcggt ttcgccccgctg ccgctgtgcgct ctagagcgct 16440
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16500
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16560
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16620
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16680
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16740
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16800
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16860
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16920
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 16980
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17040
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17100
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17160
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17220
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17280
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17340
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17400
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17460
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17520
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17580
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17640
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17700
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17760
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17820
gagcagagccc gagtgtagctgc gcgcattgcgct gcgcgcaggg agactattgtgc aggaggcagg 17880
cgcgaatgct ggttgtgtga ctgcaacctg agactctatca gcattcggca aaccagcttc 17940
atggaagtgc gttctctgga ctgcaagatg tctggtgtca actggaacag cgacaacttg 18000
catcgggtga agatgagggg ggctgtgctg ttccagcgct gcattctaat cgaactgctg 18060
ctactcggcc tatctctgga cgctgttggga gctactatca cagttgcaac 18120
tttaaagaa acaggtgaga gatggtgcag ttgtaagcata gaaactgctg 18180
tttcagcaaa ccagatgtcgc ggcggcgcag cttcagcgct 18240
atctccataa aggcatagtttt cttcagctgc uggagacgag cttcagctgc 19300
aadccgctgg atatcggacg gttcgggcttt gctgtcact gttcgcagag 19360
caacotggct ggtgatcctg cctcgggctg tgaacctggc gcagaagagc tggctacac 19420
gggaggcaca agacggtggc cttccgtgatt gctaccttt cggctggaacg 19480
tttcggggcccc caaacagctg ggtgatcctg gctgtcact gttcgcagag 19540
tggctgtgagc gttccggtgaga gatcggcaga gatcggcaga gaaactcatt gcctatcgcag 19600
gttcggcgagc ggcctggagc gttcggcgagc gttcggcgagc 19660
ctctcgtgactg ggcctggagc gttcggcgagc gttcggcgagc 19720
gggggccggccc gagccgagcct gttcggcgagc gttcggcgagc 19780
cgcagctgag cccgagctgag ctgctgggact accagcggcag gctgtcgttc gttcgcagag 19840
ctctcgtgacg gttccggtgagc gttccggtgagc gttccggtgagc 19900
cgcagctgag gttccggtgagc gttccggtgagc gttccggtgagc 19960
tgtgaggtcg gttccggtgagc gttccggtgagc gttccggtgagc 20020
caccgagcgg cggcctggatcg ggtggtgactgc cttcagcgct 190300
caccgagcgg cggcctggatcg ggtggtgactgc cttcagcgct 19090
aggagagagc ggtggtgactgc cttcagcgct 19150
aggagagagc ggtggtgactgc cttcagcgct 19200
aggagagagc ggtggtgactgc cttcagcgct 19260
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19320
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19380
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19440
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19500
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19560
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19620
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19680
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19740
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19800
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19860
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19920
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 19980
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 20040
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 20100
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 20160
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 20220
aggtggaggcc gactgtgcag ccagacggag aggcaagcactg ggtggtgactgc cttcagcgct 20280
gagtttcagg acaaccttcggc gtggcaccac gtaggcaacac agttcccgcc tgggcaagcg 20340
ttgccgcca aegatcagg gacgtcctggg gacgtcagtgg aaggtcttccc 20400
caacttgggac agttagctcg agttagctcg agttagctcg agttagctcg 20460
cagtccacgc tgaacctgtgt ccacagtggg gcgtcacttgt gcgtcactgtgc gcgtcactgtgc 20520
gccgggggct gcaagctgatt gctgctgactg gctgctgactg gctgctgactg gctgctgactg 20580
ataaggagctgc gatgaccgtcg ggcctccctgc cccatctgtgc cccatctgtgc cccatctgtgc 20640
cgctgtaatt ctgtgcatacgc gcacattggag cctgtatattt gcgtgctggcc gctgctgc 20700
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 20760
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 20820
aggttcctggactgcattggactgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 20880
cgtggtcaccgc tggagtctggtgc ggtggtctgtgc aaggtcttccc tgcagctgtgc gcgtcactgtgc 20940
tgtgctgctgc gatgaccgtcg ggcctccctgc cccatctgtgc cccatctgtgc cccatctgtgc 21000
gtctcattgag cctcattgag ctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 21060
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 21120
cgctgtaatt ctgtgcatacgc gcacattggag cctgtatattt gcgtgctggcc gctgctgc 21180
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 21240
gctgcgctgc tggagtctggtgc ggtggtctgtgc aaggtcttccc tgcagctgtgc gcgtcactgtgc 21300
gtctcattgag cctcattgag ctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 21360
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 21420
gctgcgctgc tggagtctggtgc ggtggtctgtgc aaggtcttccc tgcagctgtgc gcgtcactgtgc 21480
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 21540
gtctcattgag cctcattgag ctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 21600
cgctgtaatt ctgtgcatacgc gcacattggag cctgtatattt gcgtgctggcc gctgctgc 21660
gctgcgctgc tggagtctggtgc ggtggtctgtgc aaggtcttccc tgcagctgtgc gcgtcactgtgc 21720
gctgcgctgc tggagtctggtgc ggtggtctgtgc aaggtcttccc tgcagctgtgc gcgtcactgtgc 21780
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 21840
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 21900
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 21960
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 22020
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22080
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 22140
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22200
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22260
agggactggactccttgactgcg ggattcgtgca aaggtcttccc tgcagctgtgc gcgtcactgtgc 22320
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22380
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22440
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22500
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22560
gccagcaagct gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 22620
tacgccacg acgcataacct tgacgccggu ggcgtgggcca agatcgaata cgtgcgcgcu 22680
gtgcgcggc actaggcttg gcaaccttgcc ggggtgctcg tgtgcgcaa gcacttgggac 22740
cgcgcggtgc gtcggtcgag tcgtgttgcc ataactcgcc ccagcgccag caactatgcg 22800
ggggtgcaggt ctttcatgcg ccggcgccag cgcctgcggt tcgctggcag gctgtgacat 22860
cugaaagct gcaactgtga aatgcgcaggg gaaggggaca ttgtaactgg taactggtgct 22920
gggcagcgccc cccacgctca cctcgtgca cctgctgcag gaaacgaggg 22980
aactgcacag gcacaggggaag aagttggttg gggccggcgct ggagtgcgttt gttgcgcgcu 23040
tccgttgacg ctcgagcgaag aatccttccac cagcacgggg ctggacagcg ctgcgtgcag 23100
cacttcgatc tgaaactctc ggccgatacg cgaacacacag ttgggtgcca gaagcagatg 23160
acgcggcagc ctcgaagagg tgtgcgttgg gtcggcgctct aaggggcccgg cctgctgccgg 23220
gtcgggacagt tcctcgcacg gttgggcacag ggtggttgcag acctgctgcg gcggagccag 23280
gttgggtgcag gttggtgcttg cctcggtggc tgtggtgcag cctcggtggc 23340
atgggacgagcg ggcaggggag cggcgccgca ctcagacgcc tggcaaggg ccaggtgtgcc 23400
cagggcgccg ctcgagcagc gggagaggg cagggcagac ggctggagct gcctgtgctt 23460
caccaagccg cacaagtcgg ccctggggtgc gttggcgtgag cctcaggggc ccagggcgtcc 23520
ggcgttgtgg ctgcacacag ccagggagtg actgcgggct cgggtttttt cccacacgcgg cgggttgctg 23580
gcgaggttgcc accggtgtggg gggagcctgc cggagcgttgg aacaggagtcc 23640
gccggccacag cgcaccgcag ccgctggcagc gacagctgga ccaatcgcag tcgctgcagc 23700
gcgaggtgcttc ccaacgctgc cccagcggtctgc gcggagcagc gcgcgggtgg cttgcggtctc 23760
gatgggtgat ctcgaggtgc ccagagcgcc gcggaggtgct cgcgttgctt cggaggtgatg 23820
gacccctcgg cttgccggat ttggtgcgac ttcgctgacg cctcaggtgc cgcgtgcgac 23880
ggcggtaagc ccctcctgccg ctagtttgtgc ttcgctgacg cctcaggtgc cgcgtgcgac 24000
gccggccagcc cgcattggtca ccagaggtgct cgcgttgctt ctcgctgacg cctcaggtgc 24060
gccgcaagtgc cggagcgttc ccttcgtgc acgcgtggtg cggagcgttc cggagcgttc 24120
cctggtgacg ggaggggagc aagttgtacg ctcagacgccg ttcgctgacg cctcaggtgc 24180
aagttgtacg ggaggggagc aagttgtacg ctcagacgccg ttcgctgacg cctcaggtgc 24240
cacccctggt cccggaggtgc gcggaggtgct ctcgctgacg cctcaggtgc cgcgtgcgac 24300
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24360
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24420
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24480
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24540
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24600
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24660
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24720
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24780
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24840
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24900
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 24960
ctggctgatg ggacgaccacct cccctcctggg aacgccgttg gcggagggagc aacgcgtggtccgcggcgtgtgctg 25020
ggtctgcttc gttgcaaaac cagcataattg tcgctcagca gcagctgcttc cagatgccag 25080
tcasaatgga tcttgctctg ggcgaaaacct gtgcgcacact gcttgagcat gcggcgggtcg 25140
actgggctgta cattgacgagaa ataggggttg caggggcaag atgcgtcggc cagaagaaac 25200
agcataacagt tctgcgcagc tcgcctcagc gcgctgttaat gaaacaaggtg tgggghtggt 25260
aacoaacatc cattggcgcg ccggcagatt gaacggcagc ccggcgcagc ctgggcttggg 25320
gcagagcgcgt gcggcgtcaca tcgctctgctg atgtcgcctc ggccgbcasaata agctttctaat 25380
tcttcgctgt gtcggcggttat aaccaagcct tccacactgaca gtcgacaccc tgcacggtag 25440
ttcttgacgt agcgcacaccg ttgcgcgsca aagatttcggt gcaccctgcgg atcggatagcg 25500
gctgtaacact tcagcgatgtc agccacagcgc gtcgacgcgc gtcgcgttgcc ggggctgcaag 25560
caagggacga gcgtgtctcc ggcgcaagcgc cggctggtcag gtttctcgcga atgttctgcgt 25620
actggcagct gcggctgcgata tcggcgcggc gtcgtgaccat gcggcgcgcgc gtcgctgcgc 25680
gggcgtggac gataggtatca gcggctgcgc tcctagaagga aacagcgctga gatcgctggtc 25740
aagccaccgac ggtggcgcgaa gcggcgaacag ctagcgcgcc tgggtgctgcga tagacagggta 25800
ggccagctgcg tggcggcgcgt gcgcttcgcag tcgctggcgc gcggctgccgg 25860
 tgtatgcgca ggggaagagtc acgcttttgtg tgcacgcggg gacagctgcgg 25920
 gaagggaaca agtgctgggtt ggcgctgcac agttgtcctca aacaagcgcc aacaagccgg 25980
ccgggttacc ttcgctgcgct agtgcttggtt gcgctgcgctgc attcgctgcagc atagaggggg 26040
atagtgtcgc cttgcgtcgcg ttcgctgcagc aagagccgctg aggccggcgc gcggccgggc 26100
 tgcgatgatgc tgcgcgtgggt tcgctgcgatc cagggcgggt aagggccgctg atagccggcgc 26160
 ggcgcgtagc gcgcggcgcc gcgcgcgggt gcgcgtcggc gcgcgcggcgc gcgcgcggcgcc 26220
 cagcgcgcatc gcggcgggagc acaccctgtg gattttgacgc gagggagcgcc ttcgctgcagc 26280
 gttctggagag ggcgcgcgctgc gcggcgggctgc gcgcgctacgc gcgcgctgctc gcgcgcgctgc 26340
 gcggccagcgt tggccgcgttc gcggccggtggt ggcgcggcagc gtcgcgtgggg 26400
 cggcggcaaac cggccggcgcag tggccgcgcag tggccgcgaacc cggccggcgcag 26460
 cgtcgcgcag cggcgcggcag cggcgcggcag cggcgcggcag cggcgcggcag cggcgcggcag 26520
 atctggcagtt aaggggccgca gcgcgctgctg ccgcgcgctg ccgcgcgctg ccgcgcgctg 26580
 ccgaagcgatgc tggcgcgctgc gcggccgctg ccgcgcgctg ccgcgcgctg ccgcgcgctg 26640
 gcggccgacgc gcggccgctgc gcggccgctgc gcggccgctgc gcggccgctg gcggccgctg 26700
 gcggccgacgc gcggccgctgc gcggccgctgc gcggccgctgc gcggccgctg gcggccgctg 26760
 gcggccgacgc gcggccgctgc gcggccgctgc gcggccgctg gcggccgctg gcggccgctg 26820
 gcggccgacgc gcggccgctgc gcggccgctgc gcggccgctg gcggccgctg gcggccgctg 26880
 cggcgcgcgac gcgtgctgtgc agttggagatgc gcggccgctg ccgcgcgctg ccgcgcgctg 26940
 taccggctttg gcggccgctg ccgcgcgctg gcgcgtgctgt gcgcgcgctgc gcgcgcgctg 27000
 ggggcgcggc gcgcgcgcgc gcgcgcgctg gcgcgcgctg gcgcgcgctg gcgcgcgctg 27060
 tggcgcgcgcg gcgcgcgcgc gcgcgcgctg gcgcgcgctg gcgcgcgctg gcgcgcgctg 27120
 ccgcgcgcg acgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 27180
 gcggccgctg ccgcgcgctg gcgcgcgctg gcgcgcgctg gcgcgcgctg gcgcgcgctg 27240
 taccggctttg gcgcgcgcgc gcgcgcgctg gcgcgcgctg gcgcgcgctg gcgcgcgctg}
caacggatcg atcattgcgc cggctgaagc cagatgcgct ggggaaaaa tttgcggccg 27420
cacgtgtgac gtaaccagcg ggaggagcc gataggccga caggccaaca ccgccccggt 27480
tggtgccgac cagaagcgcga tggcctccat ggtgagcgtc gagaaccaga ccgccccggt 27540
atgctgcggcc atcgctaatc gttgaccttg cggagtccag aagcagcgtg tatacccgcc 27600
gatcaggtct tggccgaagg aagccgcagg ccaagagatc tgagccgagc tcgcagccac 27660
gggctcatac cctgcagcgcg gcaggtctct tgggctttgg ctcctcgactg atgcgcagag 27720
gcagctcgct ggtcagccgg caaaggttcc cagtcgcatc ccagggccgc ctgtccagccg 27780
gtgcggcttg ccgggtctct cggatccgga ccaattcagc gcgcaggggt tgtggccggg 27840
gggcgagttgc gacgcaagca ttggctgcgg atgggtaggg taaatgcaaa tcaaggccgg 27900
cacgtctgctt aacggtcgcc atctctgttg cagcggcctc agctgctctc aagggtaagg 27960
tggtgctcag tcaattgtta tgggctggcg gatggccccg ttcgctggaa ccgctgtcga 28020
gctgtgatga acggctggcg attgatcgag caggggttgg cattgaagag cattggcctc 28080
ggtgcgcctc ggccgctcgct gcactaacga ccagggccaa ttaaggtggt cgccggagct 28140
catgcgtcgg tttttcagtg gggtcatctc aagggctttc ggagtaaggc ctgcagcgct 28200
ccagggcgtg tggctgcctc gggggtcgcc gttgcggcagc gcagggagct gttgcggcagc 28260
ttcaggtttt cagccgctgg accagataac ttgtgcggcc gcctgctgttg ccgccccggt 28320
ccagggccgg cggcccaagc caatcattcg gcgtgtccgc gggcgcggcg gcctgctcgtc 28380
gccccacgcc cgggctgggt acaggtctttt ccagttcgtc aagctggccac ccgccccgga 28440
tgctgctgcag cactcgccgc gggtgcaggt tgcagttgga ggggtctggc ccagggccgg 28500
acacagcgtgc goccttgcccg ttcgtactgt caagctggcc cagggcggcc tttctcaggtc 28560
cgcccccaac gggctgtcccg tcaattgacg ttggtgctct ggtcagcctc gctgcagctc 28620
atccactccc cttaagtgcgc ttggtgccgc attgcaagct tggcgccatt cggccctgcg 28680
ctgcgcctgg grootgcctg ccaggggag ggattgcggtc ctcgctcagc gcgtgcgctc 28740
gtgctgattc ctggaccggc gttggctgcc gcagggggag acgataaagcc cgctgcgctc 28800	 tacccgccgc cggccggcag ccaggttgcc ctgctgcagc gcgtgctcagc gcgtgcgctc 28860
egtcgcgacg cgccgctgcag ccaggtcgtt ccagtgccag ccgctgcgctc gcgtgcgctc 28920
tgcggtgcccc ttcaagggct tccaaggtgg ccagggcccg cgcgtgcttc gcgtgcgctc 28980
ugcgctgcgt gcgcgtgcgg ccaggtgccc ggtgcggctc gcgtgcgctc gcgtgcgctc 29040
cgacttcagc ctggctgcgg ccagggcgcg ccaggtgcgg ccagggcgcg gcgtgcgctc 29100
tgctgatta ctcaggggag gtatggcagc gcgccccgca cagccatcagc ggtcagccgc gtttaaatc 29160
aaggggtgtg gtgctggtgcg ttggtctgct gcctgtgctc ggtgcggctc gcgtgcgctc 29220
teatttcaggc cggccgctgcag ccaggtgcgg ccaggtgccc ggtgcggctc gcgtgcgctc 29280
ttgccggcctc ctggctgcgg ccagggcgcg ccaggtgcgg ccaggtgccc ggtgcggctc 29340
ggcggggtg cggctgagcg ggaaggggcc tggccgcccc gcgcctgcgt gcgctgagcg 29400
gcgcgctggc cggcggcgtt ccggctgctc gcgcctgcgt gcgctgagcg ctcgctgcctc 29460
gattgagact gtcctctgct gcgcctgcgt gcgcctgcgt gcgctgagcg ctcgctgcctc 29520
cgcggcggtg gcgcggtccg ggtgcggctc ctcgctgcct gcgcctgcgt gcgctgagcg 29580
ggctctgtgg cggcttgcccg ccagggccgt cgccctgccc gcgcctgcgt gcgctgagcg 29640
cggcggcggc cggcggcgtg gcgcctgcgt gcgcctgcgt gcgctgagcg ctcgctgcctc 29700
cgcgcgcggc cggcgcggcgt gcgcctgcgt gcgcctgcgt gcgctgagcg ctcgctgcctc 29760
ctgtgtcggga aggtcttttt gcctgtgttg ggtgtcggcg ccgcgcatgc ccaatgtgcg 32160
cgtgcgaacctg ggcgtgctgta tgtcgctgca aagctgcagtct gtcgctgctc ccctgaagag 32220
tgctagcgggt gtcctctgca aaccgctatgt attgctgacttgagct gccggcaatc 32280
gccggaatagt ccgtcgtctg cggcgtgctg caagctgcgct cccgctgctg 32340
agcaacact gtcgaagaga acgcatgcct tttaagcttgagct gccggccgaag 32400
gcgggcaagaaaa cccgatgctct ttcgctctcag cggcgcctgag cggcgtgctg 32460
cgccatcagaa ccaaccggtc ggggggcaacct cggcggggtgg 32520
ggcaaacacg aggccccgcc tggggtggcttt ccggctggcttg gccggctgg 32580
cttgttgtggc ggtcgctacgt tgcctgctgttg gccctggctg cctggctttc 32640
gagaaggggt ggtggtggtgg ggtggtggtggtg gttttttttttt ggtggtggtgg 32700
cggcgcgctgc ggctggctggctggctggtgg gggggggtggt 32760
cggccaactg cgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 32820
cgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 32880
tggtctgggtg cggccgtggctgc ggggctggctgc ggggctggctgc ggggctggctgc 32940
gcggtcggctgc ggctggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33000
agsctggcgc ggctggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33060
ggctgtggctg ccggccggag ggtggtggtgg ggtggtggtgg ggtggtggtgg 33120
gggctggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33180
tccactctct ggtggtggtgg ggtggtggtgg ggtggtggtgg ggtggtggtgg 33240
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33300
gggcgcggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33360
ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33420
ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33480
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33540
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33600
ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33660
ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33720
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33780
tccactctct ggtggtggtgg ggtggtggtgg ggtggtggtgg ggtggtggtgg 33840
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33900
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 33960
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34020
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34080
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34140
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34200
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34260
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34320
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34380
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34440
ccgggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc ggggctggctgc 34500
gtacgtgccc ggtgctatac ggccgacccg cgttgtgatt tagcccaagt gccaaagct 34560
tgtgccgacg gctgtgtgtc ccggagaaac tgttatgcgc acttcaacgcct tgtcggattg 34620
cagctctgagg acgcgtcaaa atcgtgctga tcagatgctg cgcggccgca tggcggagggc 34680
gacccgatcg cattgggtgt ctctgcctct gcacacatca aagcttgagaag cagccagatgc 34740
gctctcctca gctctgcttg tcgtgcttgg cacgtcagcc ggcttgctca ggcgtcatgc 34800
gacccatagc ccctgcatata ccacgtgcaag cggatacagc tgtctgacgcc gatctcctcgc 34860
gaggtcctgcc gttgtgtctgcc tcaacgcctac gcacccagaca agtcctctgga catctgatccc 34920
cctgtatctcg aaggttaagct cttcgctcagc tctgcaaatc acacgagcccg tgacacgaaa 34980
gcgcctttccg cgtgcagggc gtcatcttttg cgcagcttcttg gacaagcttg cagccaaagc 35040
ggccggctat gcgcggcatcc ggcgcagaaa aatcgtctaggc ttgagagcgt ctcagctgga 35100
agctcaacct gcgcggagcg ggcgctcttg ccctggaggg cagcgccgtaa agtcacagcc 35160
acgcagatgg gcgcagaggg ccacgctatg cactctagg gcatgtgcgc gatacgcagag 35220
agcggcaacgc atatctgcaac gcagagctgc gcgtttctccg gctgtgtcaag 35280
gcgcgtcgtg gctgtgctgctt aagcagcgttct gcgcagtatg gactgtgtgc gatagcgcctg 35340
gcgcacctag gcacgacgctg ggcgcagcagc tggagcggagg tcaggtgggta 35400
gataatgcaaa aatggaaaaa cctgtctcctg ggcgcgcgatc ctagaaaaata ttggcgcgca 35460
tgggagcagat tcgctgcgct gaagtctggtgt gctgctctgcc gcggcagcccc gcgtgggctggg 35520
accttcgttc atcggtgcgcgt gtctgtgtcag tggagcggagg tcaggtgggta 35580
ccgctcagag tagagctgtct ctagcgtcgt atcgcggcgg ctggcggcgttc gcgggtaagta 35640
ttcgagcagc acactgctac gcgcgcagtgct ttcggtaggg gcggcgcgtgg 35700
tcgagcagc gactgtgcgc gcgcgcagctg cttgtgcgtgc ggcgcgcggctg 35760
tactacatccg gcgcggagag gcgcggctcagc gcggtggtggc ggcgagcataa tcgagctggt 35820
gcgcgtcgctgc ggcgcggccc tggagcagctc tcaactgcaactctctgacttc gcgtggtgaa 35880
gctgctctcca tagatccgtta gcggaggagaat gcgtctgctgc ggccgagctat gcagtcagca 35940
gtggtgcca ccacagccgt ggcgcgagctg ccgctgtcct cggctgcttgc ttttctcttct 36000
atccagccgc ctaccaagcgc gcgcgctgcaac gcgggtctgctgcctgcgat gcgtgtacgc 36060
gcggtgtcttg ggcgcgagcg ggcgcggcgtg gttggttggag taagaccaaat caatggattgc 36120
taccttcggt gcgcggagcg tgtcagcgcgt gcgggctcagc gcgcggctcagc ggcgagcataa 36180
ccccagtgtg gtggagtgggg tgcggcagcc aagcgttggct ttgctgtgcgt 36240
cyggcgtac gcggcggctct gcagagatgtc ggttggttgtg cgtgtctggtc 36300
ctgtgtggga caaaggggtt gcgcgagcatt tgggctggagact cgctggcagcc 36360
cagcgtcctgg ctgtctggtc gcggcgagctg ccgctgtcct cggctgcttgc 36420
ctggcgtgtgc gcgcgaggttc ctaggtgcttc gcggcgagctg ccgctgtcct cggctgcttgc 36480
tgcgacgttg ctacggcttg gcgcgaggttc ctaggtgcttc gcggcgagctg ccgctgtcct cggctgcttgc 36540
acacgacgct gcggagctgt gttacttctca ggcgcgagctg ccgctgtcct cggctgcttgc 36600
tcgagctttg ccgcagcagcgc gcgcgaggttc ctaggtgcttc gcggcgagctg ccgctgtcct cggctgcttgc 36660
ggcgcggctg gcggagctgt gttacttctca ggcgcgagctg ccgctgtcct cggctgcttgc 36720
ctgggctgctgcagcttggtcagcagcgc gcgcgaggttc ctaggtgcttc gcggcgagctg ccgctgtcct cggctgcttgc 36780
tcgagctttg ccgcagcagcgc gcgcgaggttc ctaggtgcttc gcggcgagctg ccgctgtcct cggctgcttgc 36840
-continued

cacctgtaa cgcgacgtgat tgcctacttg agacgacgac tagagtcgac gcctacgccg 36900
cggtctgaag tgcgctcgat cggggcgcggc aggggttgta ctaatgcgaa cggtgctaca 36960
cagtttcagg ccatatgtga gcatatctct atacgcaact gtcgctgctc 37020
ttcctgtcatt atgtagcggaa acaactacag cccagcgcgc ctattttgag gccgctgtgt 37080
ttcgcgtcag agcggcgcgac gacggcgcgac ggcgtggtat ctcggggcata cgaagtcgggtg 37140
atttccgcga acagctctga tgctacctgcgc gcacctgcgca aagcccttgcg caatgcgcaag 37200
gcaactctca aacctggggtct tgtgcctcttg ctcacagcag aagtgtgacgc cagcctgctc 37260
tgcacactga ttctcgtcct gctggagagc tgggtggtgc cccagcgaaca gatcctgtccg 37320
cctggaggct gtcgcgggtct ggtttgcgcac aacctgccga gcctgctgtga ggtgggttcct 37380
cttgggggcgg tcaagctggcga cagcgggggca cccagcgcag cggcagcggg gctctgtgc 37440	acagcagatgc ggtatttgctgc gcagcgcgtac agctgctggg tgtgcgcggga tgtgcctgctc 37500
acgccgctgt gcacgctggg ggaagcgccgg gataccggcg tgctgtcttg ccaaaaaactgc 37560
cctggagcgct ttgtggaaat gatccccgcgc acacgctgaaa gcaagcaggag tgggtgtgctc 37620	tagtggctgc attcagctgc ggctgatcgaa ctgtgtatgc cgtcgcgaga gctactcgcgg 37680
gttttgtag acagcggggtct gttggaaccg ccaattggat atacgctgtt ggtctctctctt 37740
gacgaggtct ctctctctgc taccctcca cggctgcgcgc aacacgcatc gctgcagcag 37800
cctgaaacc gcagcgggag ctacgctgtg ctcacagcag cccagcgcgc ccacagctgcag 37860
acgccgctgct gcacgctggg ggaagcgccgg gataccggcg tgctgtcttg ccaaaaaactgc 37920
gacgctcatgc cgccgtgctgc gcggatcgcag ctgtagtgaa gctttctgag cgacgctgctc 37980
gtgcggcact ggcgacgcga ggacgcgcgc gcctgctggg cagctgctct gcacgtgcagc 38040
gtcgtcattg gcctgtgggg atagctgtcgg cagctgctct gcacgtgcagc 38100
acgccgttgc ctacgtgcga aacgcgggcga aacgcggggc gaagctgggg ggaagcgccgg 38160
gacgaggtct gcacgctggg ggaagcgccgg gataccggcg tgctgtcttg ccaaaaaactgc 38220	tagtggctgc attcagctgc ggctgatcgaa ctgtgtatgc cgtcgcgaga gctactcgcgg 38280
tggcccgcaaca gcgtgctgct gcaatgcttg gcggctactc gcgtgctactc gcgtgctactc 38340	tagtggctgc attcagctgc ggctgatcgaa ctgtgtatgc cgtcgcgaga gctactcgcgg 38400
gaggtttctgg gtgaggtcttg tgcctgggag atagctgctgct gcgtgctgct ccaaaaaactgc 38460
gcggccgagtc gcacgcgacgc gctactcttg gcgtgctgct gcgtgctgct ccaaaaaactgc 38520
tgcgtgctgct gcgtgctgct gcgtgctgct gcgtgctgct gcgtgctgct ccaaaaaactgc 38580
gaggtgtcgcgc ggcgagcggg gctgcagcata gctgcagcata gctgcagcata gctgcagcata 38640
gaggtgtcgcgc ggcgagcggg gctgcagcata gctgcagcata gctgcagcata gctgcagcata 38700
gaggtgtcgcgc ggcgagcggg gctgcagcata gctgcagcata gctgcagcata gctgcagcata 38760
gaggtgtcgcgc ggcgagcggg gctgcagcata gctgcagcata gctgcagcata gctgcagcata 38820
acgagcgtgc ttctctcttc ttaacccggt gcggctgccg ggtctgtctg ggtctgtctg 38880
cctccggcgcc gcggctgccg ggtctgtctg ggtctgtctg 38940
ggggggtaaa atggtggtgc cggctgggat ggttgggccg cggctgggtgc cggctgggtgc 38990
ggggggtaaa atggtggtgc cggctgggat ggttgggccg cggctgggtgc cggctgggtgc 39050
cgagccccgg cgcagcgggg gctgcagcata gctgcagcata gctgcagcata gctgcagcata 39100
ggggggtaaa atggtggtgc cggctgggat ggttgggccg cggctgggtgc cggctgggtgc 39150
ggggggtaaa atggtggtgc cggctgggat ggttgggccg cggctgggtgc cggctgggtgc 39200
ggggggtaaa atggtggtgc cggctgggat ggttgggccg cggctgggtgc cggctgggtgc
tcggagctatgc gcgcagttgga ggccgcccgg aaccagctgg caaacggtcg gcgggggccag 41640
gggcgggcc ggacgctcgcc ggtgcgggtta cagagcctcgc gcacgcccgg aagtctgtatgc 41700
ggctgttcgg ggtatctttaa ggcgcgtctgc ttgcgtttgc ggcagcgccct gcgcgcccag 41760
ggcgcgcac ccgagcggtat cctggccgag tgcgcgctgt gcagtgcggt gcgcgcaagcgc 41820
gagcagtggcg tcggacgacgt gcaggcagtc acacaacctgc gttgcgcgca gcagacccgg 41880
cagcggctcg gcgcgtgaggg gcctccgggc gcgcgctctgg gcgcgtactgc gcgcgctgcag 41940
tgcgcgctgcc gcgcgccagg ggtgatgtgtg ccgcgtgtgc gcgcgacgtc ggcgtctgtgc 42000
gagcgtctggc agccctcctcc gcggagcggc ggaggagggc tgtgcagcag gcgcgctgcac 42060
ggcggctgc tgtgcgtaaa gcggagcgtca acggggttcgc gcgcggtgcag gcgcgacgcag 42120
tagtcgcggcg gcgcgcgtctg ggcgggtgcag cttggcggtac gcgtgtgcgg ggcgggtggtgc 42180
gggagctcgc gcgcgctgagc gcgcgctgcgc gcgcgctgctg gcgcgctgctg gcgcgctgctg 42240
tctgcgcgctgcc gcgcgcgtctgc gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42300
ggcgcgcgcct ggcagccacgc gcggcgcgcc gcgcgctgtgc gcgcgctgctg gcgcgctgctg 42360
ggcgcgcgcct ggcagccacgc gcggcgcgcc gcgcgctgtgc gcgcgctgctg gcgcgctgctg 42420
ttcgcgcgcgc gcgcgctgtgc gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42480
cggcgctgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42540
gggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42600
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42660
tgcggaggtgg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42720
cggcgctgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42780
cggcgctgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42840
cgcggccgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42900
gggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 42960
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43020
gggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43080
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43140
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43200
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43260
gggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43320
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43380
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43440
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43500
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43560
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43620
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43680
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43740
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43800
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43860
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43920
ggcgcgcgcct gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg gcgcgctgctg 43980
cagcatatg tgcattcttg gcagacacctg ggcgtcgatatt gtcgcgcgctg 44040
gecgccaga agcatttcatg cattgcgcagt cgggtcgcaca atgcgcaccc acggcagatgt 44100
cagctgtggca tcgggtgggcat gcgttcacctg gcgtgggtcatg ggagcgatact 44160
cagcagtgtgc gcgcagcttcg acaatggcata acaatggcata cgggtggcgcgtgc 44220
cgtgtgcagtc atcgcggcgattc cgcgccctggtc atcgcggcgattc gcgtgtgcagtc 44280
aaccgccgtg ggcagccatcc gctcaagtctc cggcgcaacagatc ggcacgctgc 44340
gatgtgcgctcg aagcgcacgtg gctctcctctg cccacgagca cggccacgtg ggcacgctgc 44400
tgcgcctgcag cgccgtgcatg ctcgtgcgcgctg gtcggtgcgcgctg 44460
ttgctgacgctgc gctgctgctgc gctgctgctgc gctgctgctgc gctgctgctgc 44520
tgcagcagcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 44580
gcgctgctgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 44640
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 44700
cagctcgcgctcg aagcgcacgtg gctctcctctg cccacgagca cggccacgtg ggcacgctgc 44760
aacatgtggc cttgcggttgc cttgcggttgc cttgcggttgc cttgcggttgc 44820
aagcgttgatgc cttgcggttgc cttgcggttgc cttgcggttgc cttgcggttgc 44880
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 44940
gggtgtccg ggcgcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45000
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45060
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45120
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45180
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45240
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45300
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45360
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45420
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45480
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45540
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45600
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45660
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45720
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45780
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45840
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45900
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 45960
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 46020
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 46080
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 46140
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 46200
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 46260
ggcggtgga tgtgagcgcgc gcgcgagcgcgc gcgcgagcgcgc gcgcgagcgcgc 46320
-continued

cctggcttg gcacgcgaac gcgctcgagc ccgctcggcc tcagcgcci tcagcgccct ggcagcataa 46380
ggcgcaatcg gcgtcctgga tcagccaggg cggatcctgc cggctgcaag gcgtacagcg 46440
gacgacgcaaa ccacagcaac catcactaac gcagcagatt tcagctcagct gcgtagcagct 46500
gtcggtcaag cggccggcgg cctgctggtgc ccgccggagc gcgctgcgtgc ccgccggcgg 46560
cgcttcacct gcggcggcgg cctgctggtgc gctggtcggc gcgctgcgtgc ccgccggcgg 46620
gtcggtcaag cggccggcgg cctgctggtgc gctggtcggc gcgctgcgtgc ccgccggcgg 46680
ggcgcaatcg gcgtcctgga tcagccaggg cggatcctgc cggctgcaag gcgtacagcg 46740
ggcgcaatcg gcgtcctgga tcagccaggg cggatcctgc cggctgcaag gcgtacagcg 46800
ggcgcaatcg gcgtcctgga tcagccaggg cggatcctgc cggctgcaag gcgtacagcg 46860
ggcgcaatcg gcgtcctgga tcagccaggg cggatcctgc cggctgcaag gcgtacagcg 46920
ggcgcaatcg gcgtcctgga tcagccaggg cggatcctgc cggctgcaag gcgtacagcg 46980
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47040
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47100
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47160
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47220
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47280
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47340
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47400
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47460
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47520
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47580
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47640
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47700
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47760
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47820
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47880
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 47940
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48000
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48060
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48120
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48180
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48240
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48300
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48360
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48420
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48480
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48540
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48600
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48660
tatggcagcc gcagcggcgg cctgctggtgc cagctctgag gcagagc cggatcctgc cggctgcaag gcgtacagcg 48720
-continued

tccgttttca taagcgcgca aagttgccttg ttcctacatag atcgcaacat ggccagtgtac s1120
tctggatgta agtggcggcg ctacotgggtg ggcagctaca atgtctctgcct tgtgcccctac s1180
atcattagcc cttcctcaggt ctgctgtgacgc ggtaaagcgc gaggcagacag taccagagac s1240
cggccgagcg tgggtggctag gcgtaccccg cacagggcgtt gattggacgc cagctgacac g1300
cggatgagcg ggcggctggtag ggtcgtggct gcacatattg ggctggtggcc ggtgatgagcg s1360
ccgctttggt tgtgagcctg ggcagcggcgc gaggcggctgtcg ggcgtcaggt tgtgcggccag s1420
cggcttgcag aggtctgagct gacctccgctt gcgtggcagc acgtaaagcg cgtacacgc g1480
gaaaaccgcc agggcgccca aagggcggcag cagccgcttg ttcgctgcag caagccagtc s1540
gactgtctgt aagggcggcag ctaagcagctg atcctggtca ccaacgcctg gcacactattc s1600
gattggctgct atcgccgagcg catcctgctgta aagggcggcag agtcgtcggta ggcagccgc s1660
atggtgtcaac ctacgtttagc atcgcgcgacg cagagagcgtg ccgttcgctg gtagagcgcg s1720
aatccgtggag tgtgtaaagcg ggcgctgcttg ggtgtgtagc attggccgacgc gcacacgtatgc s1780
ctggagagta ctggcgctggca tgcggctttcg gcagcagctg gagaaggccg gatgcccgtatgc s1840
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s1900
ctggagtgcg ccacacagcc aaggtgcctgt gcggagcagc acgtaaagcg cgtacacgc g1960
actaatctgt cttcctcagtgc tcggcgagat cggcgccgca gcacacgtatgc gtagagcgcg s2020
atggccgagc atcgccgagcg ctggcgctggca tgcggctttcg gcagcagctg gagaaggccg g2080
tgcggcgcgc ccacacagcc aaggtgcctgt gcggagcagc acgtaaagcg cgtacacgc g2140
ccagttgcgc atcgcagcagc gcagggtgctg gcggagcagc acgtaaagcg cgtacacgc g2200
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2260
sggagtagct gcacttgctg atcgggctgca gcacacgtatgc gtagagcgcg s2320
cggctgtacgc atcgcagcagc gcagggtgctg gcggagcagc acgtaaagcg cgtacacgc g2380
cggctgtacgc atcgcagcagc gcagggtgctg gcggagcagc acgtaaagcg cgtacacgc g2440
cggctgtacgc atcgcagcagc gcagggtgctg gcggagcagc acgtaaagcg cgtacacgc g2500
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2560
tgcgagagta ctggcgctggca tgcggctttcg gcagcagctg gagaaggccg gatgcccgtatgc s2620
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2680
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2740
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2800
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2860
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2920
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s2980
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3040
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3100
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3160
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3220
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3280
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3340
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3400
ccctctgcg gaaaaccgcc gcagtgccgc gcagcagctg gcacacgtatgc gtagagcgcg s3460
acggaggtgg ctcggaggg agtaagtgaa aatggtgttc
<210> SEQ ID NO: 2
<211> LENGTH: 2986
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albineans
<400> SEQUENCE: 2

gaattcagct aatgttgtctg cgggtgcccc cagctgagctt ctctagggcc cagaggtgtt
120
gcggcgttaa ccggcgtttg cggcagcgtt gctgcgcgtt ccggccaggt tgggtcttttt
240
cgctcagaga atcaatgcga tcggtctact cgttatctcg ttacgggcgg cctgccgcgca
360
cgcgcaagcg cgcggtggcg ttctgtcgcc gcagacaggg attcgccggt tcctgacgga
480
cgctcatcgtcg ccggcgccgg aatcagagcg tttcaggtgt cgggttctgc gccccttctcog
600
tctttctgt tctatggtcat caatggtcgct tgggtgcttcc cagctgtgag tctggcagctc
720
tgagcgcgca cggcgctcgt gaaagggggc cttccagggcc acctgcgacc cagcgccgctga
840
atcagtggc gcggggcagct ttcacggccg tgtatggtgc cggccgcggt cttccacggcc
960
gcgggaagct gtcttgagaat tctgctgttg tcttggccgg atgagctccg ggaaggccaa
1080
agctgttctctg ccaggcgcttg tatactgcag aatggctgca ggacagccgg agctggtgctc
1200
cggctcgctg atcggctcag cgccctcogg ccaggcccccc cagcgccccg cgaagccca
1320
cggcggttca ggccgatcgc tgcagcatc ggcctgtttt aatcagagcg tggatacgcc
1440
cgcgcaagcg cggcagagc gcttggtcac gcgccgagg tcggtccgcg ctggggcctc
1560
tctggcgctg tctagtccag caggtgtgcc gcggctgtgc aatcagagcg ggcgcgtacc
1680
ccttgcggtgg cacggcgccgg aatggctgca ggacagccgg agctggtgc ggccggttca
1800
cggcggttca ggccgatcgc tgcagcatc ggcctgtttt aatcagagcg ggcgcgtacc
1920
cgcgcaagcg cggcagagc gcttggtcac gcgccgagg tcggtccgcg ctggggcctc
2040
cggcggttca ggccgatcgc tgcagcatc ggcctgtttt aatcagagcg ggcgcgtacc
2160
ccactacgac tggtgagcac gggcagcac gaacccgcca acggcaacg gacctcaagc 2160
gtggcgtcgg cgcgtcgggt cgtgaaacct tcccccgcgc cgccacctgc ggctgcattg 2220
ggatggagga accgaggaat ctacacaaac agctgtgctgg caaggtcagtt tctctgcaga 2280
tgcaacatc gcacagcggc gcacgcaaca taatccgcca tctcttcctgg ctctgacacc 2340
tgcgtcgcg cgccgtcgcgct ggtgctgacc gaacctcaacc gacaaacactg 2400
gtgcctacca ccggacacg cctagccaga gcacccgagag ctcctgctcag gaaacgctgc 2460
acgtcagcga acgcgcgggg ggtctgccgag atgggctcagta gsaaccgacac ccaaacccctg 2520
cgatacgaca ttatacgacc gatggccgca acggctgcaag acgcgtacgc tcagcagcagt 2580
acacttgctc atctccatcgc gacccactgc ggcgagattt caagtcatgaa cggcgtgtag 2640
cgtgcaacat cgctggcgac ggaagatccgc cgctgagccgc gatactgccg tccgctggct 2700
ggctcgcggc cggcgctgac ccncaacgggg gctgctccttc gtcacactcgc gcggcgcctc 2760
tgctctcggc ccgactcggcgc tgcgctgccg gcggcgtgac ccctctggcag cctgcacacg 2820
agcggcgcct acggcgccgcgc ggcgctggct ggcgcgcgggg ccctctggcag ccctctgcag 2880
cagggcccc acgttcggac tgcagccggc cgcattcggag gggctggctgc ggtggcagac 2940
ccgggagcacc tccggatacc ccaccaacag gacatggctc gtaatc 2986

<210> SEQ ID NO 3
<211> LENGTH: 9673
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albinlines
<400> SEQUENCE: 3

gaaa tgggtgatgat ccctggacgc gcctggtgatg tcaatgcaac ggggttgagg 60
ctctggctgc gcctgcgtgc acgtggtcct cctggctgcgt gcgcctgcgt gacctgccct 120
tgcaacaccg gggggtggtgg accggtggct accggtggct accggtggct accggtggct 180
gtgcaacattc gcctggtgtgg ggtggttgtg tcctggtgtg tcctggtgtg tcctggtgtg 240
ggcatggacc gcctggtgttg gcctggtgttg gcctggtgttg gcctggtgttg gcctggtgttg 300
cacagcccc gcacagcccc gcacagcccc gcacagcccc gcacagcccc gcacagcccc 360
cacaccccac cctgcgggac gcctgcgggac gcctgcgggac gcctgcgggac gcctgcgggac 420
agggcagctg gcaccacttg ctcccagcttg gccggtggct accggtggct accggtggct 480
aggggcaag caggccgactc cccctccctc acccctgctc cccctccctc acccctgctc 540
acgccggtt ccctggctgct cgcctgggct cgcctgggct cgcctgggct cgcctgggct 600
tgcgggttc tgcgggttc tgcgggttc tgcgggttc tgcgggttc tgcgggttc 660
aaccggcagc aacatccagc atctccagc acgtggtgct acgtggtgct acgtggtgct 720
cggggggtc gcctggctgct gcctggctgct gcctggctgct gcctggctgct gcctggctgct 780
actgcgtggc gcgcctgctg ctggcggcac gcctggctgct gcctggctgct gcctggctgct 840
cggtgggct atctccagc cggcctggcct acgtggtgct acgtggtgct acgtggtgct 900
acctgggtgc ctcggctggc tggcggctggc tggcggctggc tggcggctggc tggcggctggc 960
ctggctgctg ctggctgctg ctggctgctg ctggctgctg ctggctgctg ctggctgctg 1020
gcggggtc gggggtgatgat ccctggacgc gcctggtgatg tcaatgcaac ggggttgagg 1080
gacccgtggc gcctggtgatg tcaatgcaac ggggttgagg 1140
gcctgggt gcctggtgatg tcaatgcaac ggggttgagg 1200
gaaacgctca cggggtatcg cttggctctgc ctggcgcgc gcgttgcgggt ggcaacacgc 1260
tagtgctgct gcggacgctc gcgtgctgaa gcgtcgcact atgcgtctgtg gtatgcggca 1320
tcgtgacgct cggcgaagct catgcgctgc ctgcgaccccg atgcgcgtcc atggcggcaa 1380
tcgcggcttg gcggagcgtgc gttgcgtctgg gcccacccca ccatatgcggc 1440
tgccaaagct gtctgagcgc cagggaacca ccataactgcg tgcctcaacgc ccagggcgcg 1500
tttcgcactgt tgacgcctgcg ctggtggcagc cyggttttgtg cttctgctgc aagcagtgcc 1560
agtcggtctgc ggtgtgcctgc acacaggcgtg cataacgctg tggcgcagatcg gctgcaagcc 1620
ggcggcggca tttgggctcg gtctgaacga ccaacgacac ctgctgtagt ggacacacga 1680
cagcagatgc tgcggcgttt cgaaggtcatc cgagcgtatgc ggcagacattc ctatogcag 1740
gagcacgcct ggctgtcgatg gcctatctcg ccgcccgcgc gcgtcgccga aataacgtgc 1800
cgcgcggcggca gcggccgcggc atggccagcg tgcgacgccc cagcccacag tcgagcgcgc 1860
acatcataata ccaacacaata cccacagact acgtctcagc tggcagccag acactgcacc 1920
tgacgcacat gcgcgtgctgc ggcgtcgccag gacgccacct gcaggcgcgt gcgcggcgcg 1980
cagcagatgc gcggccgcgc atggcagcgc tgcgacgcgg cagccagcgc gcaggacgagc 2040
gccgcgcttg gtcgctgcgc ccggtcgtgc aggagctctc gcggcagctc gacccctggtg 2100
cggggtcgc gttggggaaat tagacgcagc cggccgttgc gcctctccag gcctagttgg 2160
ctgcccagaa cattggtgagat ccgcggccgc cttcctggcgc ttgctcggcg cagccctggtg 2220
ggatgtgccg tgtgagttcg cctctgcaag ctaggtcagc ctctgctgtg cgtgctgcag 2280
tgcctagact cggcgcgcttc cttgagctgc ggccgcaggt gcggcggccg tgggtgcggg 2340
tcgagctgtgt cggcgcgggt gaggacgcgg actingtcggtc cggcaagatcg aagcctgctg 2400
ccggcggcgc gcggcgcgc gcgcgcgcgc gcgttgcggc gcgggcatac gcgagctgcg 2460
tgctggcgcgc gtcgctggcgc tttgggctgg cttggcgcgc gaagccatgc cggcggcgcg 2520
ccggcggcgc gcggcggcgc gcggcggcgc gcgttgcggc gcgggcatac gcgagctgcg 2580
ccctgggggata cgatccggga gcgcgcgacac agattgacgc ctcttgagttgc gcgcggcg 2640
ttgcccgcgg gccgtgctgtg gcgtgctgctg ctgcgcacgc gcgggtggtgct 2700
tgcgacgcac gcaggtgcgt gcgggtgggc ttgctgaggt tttgagcttg aaggacgcgc 2760
acgcatagt cgctgcttca gctaatgcgc cggcagcgcgc cgggttcttgg cgggttctggt 2820
aggaagagaag tgggcgtcag caggtggcgg gccgtcagctt ccagctgctt caggggtcggg 2880
tgcgcgccgc gcgtgctggt cggagcgttg tggggtgggc cggagcgtgc ccagagctgcg 2940
ggcagcagc cgcctgctgt gcgggtgggc gcgtgctggt cgcggttgccgc cggagcgtgc 3000
cggccggag cggggttctt gcgcgcgctgc gcgggtgtcgc cgcggttgccgc cggagcgtgc 3060
cggcggcggc gcgtgctgtgc gcgcgcgcgc gcgggtgtcgc cgcggttgccgc cggagcgtgc 3120
aacgcggtgt gcgcgcgccgc tgcgacgcgg gcggtgggctt gcggggttt 3180
ctgcgcgcga gccgcgcgccgc gcgcgcgccgcgc gcgtgctggtc gcgcgcgcgcgc gcgcgcgcgcgc 3340
atgcgcgcgccgc gcgcgcgcgcgc gcgcgcgcgcgc gcgtgctggtc gcgcgcgcgcgc gcgcgcgcgcgc 3500
acggtccccgc gcgcgcgcgcgc gcgcgcgcgcgc gcgtgctggtc gcgcgcgcgcgc gcgcgcgcgcgc 3660
acgcgacgac gcgcacgag cacgctggcga gcatctggcga gcgcgcctgcag 3660
gtgcgcgacg gtcgctggcga gcgcacgctg tcacgcgggga gcggcagcgctg 3720
gcgctgctgc acgagcaggg ctcgcgctgccag ctcgccgccgc gcgcgcctgcag 3780
agacgcacgag gcgcacgctg ccctgctgcag gcgcgccgctgc gcgcgcctgcag 3840
gtcgctgcag gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 3900
cgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 3960
acgcgcgcgc gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4020
ggcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag gcgcgcctgcag 4080
ggcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag gcgcgcctgcag 4140
acgcgcgcgc gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4200
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4260
cgcgcgcgc gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4320
ggcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag gcgcgcctgcag 4380
cgcgcgcgc gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4440
cgcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag gcgcgcctgcag 4500
cgcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag gcgcgcctgcag 4560
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4620
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4680
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4740
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4800
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4860
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4920
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 4980
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5040
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5100
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5160
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5220
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5280
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5340
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5400
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5460
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5520
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5580
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5640
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5700
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5760
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5820
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5880
tgcgtgcgat gcgcgcctgcag gcgcgcctgcag gcgcgccgctgc gcgcgcctgcag 5940
catcgaacgc aagcgttgcc gcacgagaacc agagatgtag acccaacccc tctgctggcgt 6000
atgcgtcgg cagcgcgga acggcagccc tggccgacc cgtgcccttg atgcctgcctg atagagcgcgc 6060
ggcgccccgc gcaagggccct cggcttccaa ttggagcacc actcagaaga ct ggtaaacc 6120
aacaagaac ac ctcgcttcat cagcgcgag ctcacccact gcgtgcacctg atgcattct 6180
cgttggacttc ccaccagagc atctttctct gcgcgcttgg atcctaatagc ctcrgcgcgt 6240
cgcagcaacc gcgtgtcgctc gcaagggaca aagcgcagtt ctcggggcgttc gcagcgcggc 6300
tgcgtcatcgc atggcgtcctgccagcggc agcggccctg cagcgcgcttg gcacgagggc 6360
tgcgtcatgc cctgagggag atcgcgctcgc accctgacc agctaccgaa tccgaccct 6420
cgcttttcct ccaagaacgc tcccgggctgtc agacagagac acagtaa gctgacccttg atgacgcatgt 6480
tgcgcggct ctcctacctgc ctggcgtcagc aagtcaggcgc tgcgcctgg agtgcgcctgc 6540
gcccgccgct gcagtggtcagg gcagggcctt actgcgctcc gcgcggtgcag gcagagtgcc 6600
aggtgccgctc gcagcgcgcc atacggtgac gcacgcgctg tcgcaggggc ttgacgagggc 6660
aacaagaagc ttcgccggccgc ctgggagcagc tcgcacgctgtcgagcctgc taacctgacc 6720
acgtcgcttc gcagctgctc ccataaggag ccacatcagt ccggagagac gcacgagggc 6780
atgggccccgc gtcggagcgt gcctacggt ccgcggcttg gcacgagggc cggcagcccg 6840
agattcggagc cagggagagc cacagagcct acgcagcctg tccgaggggc ggacgacgagc 6900
cgttcggcgct ccgctcctgcc aacgatggtg ccagcctgagc taacaagctgc ctgacgctgc 6960
tgcgctccgg tcgcgctcttc agagcgtgctgc gccgctcggc ccgcggccgg cagcaagtcg 7020
acgtcggctg cgtcttcgcc atgcgcgctgt cctcccctcgc ctggcggcgct tcggtcgcctg 7080
tcaacgctg ccctccgcggt tccgggctgtgc atgcgctgctg tggcctgctgc 7140
aacctggctcg ggtatcgagc gcgtgctccg ccagcagcgtc gcacgagggc agctgaagcg 7200
tcgagcgcgctg ccgcaagcgtc gcgtgctccg ccagcagcgtc gcacgagggc agctgaagcg 7260
aggtcggctc gcggcgcggt gctacgcgct gcgtgcgtgc ctgcggcggc ccggcggcgcc 7320
tgccggtcgc gtcgcggcgg gcagggcgggt gcggcggggt gcggcggggt gcggcggggt 7380
tgcggggtgc gtcggtgggt gcggcggggt gcggggtgc gctgctggcg gcgtgcgtgc 7440
aaattaagagc cagcagcagc cttgaggtcg tccgagggc ggctcagagc tgctgtgctg 7500
tcgagcggcct caggctcggc tggcgggtgc gctgctggcg gcgtgcgtgc 7560
tcgtcagctg cggcgcggt gcgtgctccg ccagcagcgtc gcacgagggc agctgaagcg 7620
agggcgcgcc aagagcgcggt agoacgcggt gcacagcgtc gcgtgcgtgc 7680
tcaagctgcg gcggcggcgc cttgaggtcg tccgagggc ggctcagagc tgctgtgctg 7740
ctggcctccg ccggcgcggt gcgtgcgtgc 7800
agaagctgcc ggcagggcct ccggcgggt gcgtgcgtgc 7860
aacgtgctcc gggacgctcc gcgtgcgtgc 7920
agggcgcgcc gcagcagcgtc gcgtgcgtgc 7980
acoagtgcg tggcgcgctg cggcgcggt gcgtgcgtgc 8040
agcagcttcag acgcagcagc cttgaggtcg tccgagggc ggctcagagc tgctgtgctg 8100
gccggggggt cttcggggt gcgcggtgc gcgtgcgtgc 8160
ctggcctccg gcgtgcgtgc 8220
accggcttcc aggcgagcag cttgaggtcg tccgagggc ggctcagagc tgctgtgctg 8280
agggcgcgcc gcagcagcgtc gcgtgcgtgc 8340
gtccagggtgg tccaagccca ctctggaagct cgggtgctcct ccatggacagc cgggtgtgctc ccatggacagc 8400
cctcagcctg tgataagcct agtacggtat ctggcgcgcc catcgccctg ccagtcatcct 9460
ggcggaacc caaatcggctg ctgctgctgct cgggtgctcct ccatggacagc cgggtgctcct 9520
gcgcagac gccagtgcct ccgtgcgtgtgc ccatggacagc ccagtctcctg ccagtctcctg 9580
cgagaaaatcg cgaggtcgtg ccgggtcctg cgggtgctcct ccatggacagc cgggtgctcct 9640
tagcggcgtga ctgagtcggcc gtcctcagcc cgggtgctcct ccatggacagc cgggtgctcct 9700
gacaagcct caggcgctct gcgggaaccc cgggtgtgctc ccatggacagc cgggtgctcct 9760
ccacactcgcc cgggtgctcct cgggtgtgctc ccatggacagc cgggtgctcct ccatggacagc 9820
gcagcagctag cgggtgctcct cgggtgtgctc ccatggacagc cgggtgctcct ccatggacagc 9880
tctggcgtgag cttggcgtggc cgggtgctcct ccatggacagc cgggtgctcct ccatggacagc 9940
cgggtgctcct cgggtgtgctc ccatggacagc cgggtgctcct ccatggacagc cgggtgctcct 9900
tcgagttggc agcgggtgcgg cgggtgctcct ccatggacagc cgggtgctcct ccatggacagc 9960
agtgggtggtc cgggtgctcct ccatggacagc cgggtgctcct ccatggacagc cgggtgctcct 9960
<210> SEQ ID NO 4
<211> LENGTH: 267
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 4
ctcgtcagcag cggcgcgccac ccagctgcat gccagcgcac gccagcgcac cggcgcgccac 60
cagctgtgag cggcgcgccac ccagctgcat gccagcgcac gccagcgcac cggcgcgccac 120
tctcccagcc agatgtgctggg cggcgcgccac ccagctgcat gccagcgcac gccagcgcac 180
ggagacagtgg cggcgcgccac ccagctgcat gccagcgcac gccagcgcac cggcgcgccac 240
tggtgggtggtc cggcgcgccac ccagctgcat gccagcgcac gccagcgcac cggcgcgccac 267
<210> SEQ ID NO 5
<211> LENGTH: 1755
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 5
tctcttttctg atccagtcatc cagcagctgc cggcgcgccac ccagctgcat gccagcgcac 60
cgcgacagtgg cggcgcgccac ccagctgcat gccagcgcac gccagcgcac cggcgcgccac 120
agacacagtgg cggcgcgccac ccagctgcat gccagcgcac gccagcgcac cggcgcgccac 180
-continued

tctctgttcgc agaaagatcga gaaacacccag gaaacagccgcc tgcacgacctc ataatgaacac 600
cacacggcag aagggccca acatgtgatgc gcagtggtgc attcaggtaag tcgcagcagc 660
ggcccggcag cgcacgctggt atggcagctg tgtcagcagc acaagacagcc ccaacagcgg 720
aataccggac gcaatgccgct ccaagcaceca cagctggccac cggctttctctgt gtctttctat 780
tatgctgcatca cagccggcag ggggctgcag tgggctgcgta cccgactcgt 840
atggattgcgc gcaatggcgcg gcacgtcggcc cccgcttggg cccgcttggg cccgcttggg 900
gagggcactc gcagcagttgc acagaaagatg gcagcagttgc acagaaagatg gcagcagttgc 960
gtggatctgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1020
agaacaagag gaggggctgcg gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1080
caagctttctgg cctggggccaa gcctctggcatt gcctggggcatt gcctggggcatt gcctggggcatt 1140
gtggatctgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1200
gcgcaacac gggctttctgg gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1260
cgcacaccgc gcacacgcctgt gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1320
gtggatctgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1380
atggattgcgc gcaatggcgcg gcacgtcggcc cccgcttggg cccgcttggg cccgcttggg 1440
gagggcactc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc gcagcagttgc 1491

<210> SEQ ID NO 8
<211> LENGTH: 956
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE:

ttatcagcgtt atggcgtgcg cagcgcgtgc cattggtgag cttgagaaat gcagagatgc 60
ggcattcaca tcctaacgcg aagggctgcg gcacgagccca gcacgacgctg cagccgcgctg 120
gccccgacag cggctattcg gcctgggtaa ccctgggcttg cgggcttggtt 180
gggcgaccag gcagcgttcg gcacaggact gcctggctgctg cgcagcgcctg 240
gctgttgctgc gcacgagctg cctccacccccc acacgactgc ggacactgccttt 300
gggcagcagc tcctagcagc ggacaccgctg gcagagccgc cccgccgctg 360
cacgcgagcgc cattgcgctg cagcgcgctg cggccagtgc gcacgagatgc cccgagccgc 420
cagcgcgagc gcgtgtcgag gcggagcgtgc gcacgagatgc cccgagccgc 480
gggacagtgc gcacagcgctg cttccagagttgc gcagagatgc cccgagccgc 540
cagcgcgagc gcgtgtcgag gcggagcgtgc gcacgagatgc cccgagccgc 600
gacgcgcggc cattgcgctg cttccagagttgc gcagagatgc cccgagccgc 660
tccttcagcgt gcacagcgctg cttccagagttgc gcagagatgc cccgagccgc 720
tccttcagcgt gcacagcgctg cttccagagttgc gcagagatgc cccgagccgc 780
tccttcagcgt gcacagcgctg cttccagagttgc gcagagatgc cccgagccgc 840
tccttcagcgt gcacagcgctg cttccagagttgc gcagagatgc cccgagccgc 900
tccttcagcgt gcacagcgctg cttccagagttgc gcagagatgc cccgagccgc 964
<400> SEQUENCE: 8

```
ttcgccggc atcagtttat ccagaatgct acgagccgcc gcagagcacc aagaggtacg
60
aacgatcggc acgtgccgtt ccgagcaac gcagctcag cctgcttcgc cgtgctacac
120
cagcgccggg gcagcgccgt gcagcgccgg gcagcgccgg gcagcgccgg gcagcgccgg
180
ggcaacccgg ttgcagcagc ccggaccggc ccggaccggc ccggaccggc ccggaccggc
240
aaaaaggggc ccagcggcgc tgtgctgttg tgtgcgcgagc cggactatag ccggagccc
300
gggtatgtat tgtgctggca cttaoacgat tattgaacag ggggagcgtt ccggagcgtt
360
ccccggcttc ttctgcacgc ccggaccacc ccggacccgc tgtacccgaga gttgtgcaac
420
gggggttgg tgtggctgag ccgggctggt ccgggctggt ccgggctggt ccgggctggt
480
ccggagacag ctaaagacttc tgcggtagcc atataaggtg tatttctgga gcacgtgcca
540
ccggagacag ctaaagacttc tgcggtagcc atataaggtg tatttctgga gcacgtgcca
600
tgggctgtgct gccagggcag acatctcctgc ctctgcggca gcctctgcggca gcctctgcggca
660
catgatgacg cttctcgggg ccggagaccgc gcgggttggg gcgggttggg gcgggttggg
720
cggggttgg gctggctgag ccgggctggt ccgggctggt ccgggctggt ccgggctggt
780
ctggctgttg ccagctcagc ccggacacag gcctctgcggca gcctctgcggca gcctctgcggca
840
cagccggggt cggggttggt ggtgctgtgc gcgtcaggtg ggtgctgtgc gcgtcaggtg
900
cagccggggt cggggttggt cggggttggt cggggttggt cggggttggt cggggttggt
960
atgggaatgc ctacgccagtc gttgcgcaca cccgctcagc aagcaagtagg acatgcctaa 780
ggccgcgatgc aaccaacagct tgggctgatgc ctgacgcctc aagccaaata gctcaagctg 840
cgtgcctcag tggatgcaag aatgcaacag cccaaccttc gcggaccgct cgcgcagagc 900
gacagacatt tgtttttcga gatagtcgcc caagcaagca cagcccat 948

<210> SEQ ID NO 10
<211> LENGTH: 252
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 10

tcgagctaaa ctgcatctca aaccaacctc gcgtggaagag ctgggctccc ctaaatctct 60
gtccttcagct gcgccggagt ataaagggggc catgtccgct caggtttcgt cgtataagct 120
catgcaccgct ttccttgccgc cagcataagc ttccttgccgc cgaggccagag gcagctcaag 180
atggcagcgc cacaagcagt atgcgctgcgt tgcgtgtgctg aggcacagat aagcctgctg 240
cctgaagctgc at 252

<210> SEQ ID NO 11
<211> LENGTH: 2151
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 11

ttcgcgctggc ttatcataaa taattacgat tgcgttcctt ggaatctcgc cgaatacgtta 60
ggcgtcagatat gcggggagc tctgctgggt gcgtccacac agaagctactc ctggcgcgag 120
cgcaccggc ggcggggtcgt ctgcgtcagc atctttccagc ggcgtccggt cgcggtgggtt 180
aatgtagggg atctttcact ctgctgctgac gcgtggctag gcgtggtgag aagaagcaggcccc 240
gcggttagtc gggtctcccg tagttgctag tgtctgctgag gtggcctgcc ctgcttgctg 300
cgcgtctgct gcctggcact ctgcctagcc gcgtgcctagc ggcgtctgctg gcgtgctgcc 360
gccgateccgc cgcgtgctagc gcgtgctgagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 420
tcgcgcgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 480
cgcggtgcag cgcggctgagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 540
ggcggtgctgagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 600
ggcggtgctgagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 660
ggcgtgctgagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 720
cgcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 780
cgcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 840
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 900
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 960

ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 1020
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 1080
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 1140
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 1200
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 1260
ggcggccgctgt gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc gcgtgcctagc 1320
gacctgttct cgttccagcc cgagctggaat ccatattggt cttatagcata cattgacgaa 1380
gggaacccac cgagctggtgg ggcattatttg ctagaaccctg gattgtgtcag cttattgact 1440
tcaacagag gttttttgta agtggacggc caagggcagca tcacacacaag gcaacacgag 1500
gggacctgcgc gcgctagcgg ggtcacaacaag cttggaaggctg gcgctagccgc 1560
gctcgggaa agacagcagc ggaaacactgg atagatgcgg caatgatgccc caaagacctgct 1620
agcggcctgag cgtggccgag cacgctacctt gcgtcgcgact gcgtcgcagt gcaagactc 1680
aagaacaggtc atccagatggtgcagcaggtggcagtcgctgagtgta 1740
gtcgaacgtc tcagggggtt gtcccccgcag gcggctgatac caaggccgagca gctgtgagc 1800
agtatggata ttatttggagt cccgggaactgc cggcggagtgc gcggtgtgagc cggcggactc 1860
ggtctacgtgg ctgcacatgc cgtcgccgacctt ctcacactgt cgagttcgcagctctcccttt 1920
gagcggcgcgg gagaaactgg gcgcgcgcgc gtctgccccctgctctgtcg gcgacccggg 1980
gagggatgtagc cggagagttg cgccagggagt cggacagctcg ctcagcggtcatgccggttttt 2040
cctactcgctgg cgaacacctcg ctgccacactg gcttggcagag agacagagctgctgtcgcgcc 2100
agacgacggtc ttcggcgagat gcggcgccgg cagggctttca tggagcactgctg 2151

<210> SEQ ID NO: 12
<211> LENGTH: 414
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albiniae

<400> SEQUENCE: 12
atgagggccc cagcggctcag ccgcaacacc gcgggtgcctg gcgtgtaaca ggcggccggg 60
gtcgtagag ccggcagcag caactctgtct ctggcgagaga ggcgcaacttt ggtggaagc 120
ggcggcgcgg cggcgcgcgc gcggcgccgg cggcgcgcgc ggtggtaacg atgggctttg 180
cggccggcgc gcggccggcgc gcggccggcgc ggtgctgtctgcc ggtgctgcttctctctct 240
ggtgctgtctgcc gccggtgcgcc gcggccggcgc gcggccggcgc gcggccggcgc 300
ggtgctgtctgcc gcggccggcgc gcggccggcgc gcggccggcgc gcggccggcgc 360
ttcgacgag gcgggcgcgc gcggccggcgc gcggccggcgc gcggccggcgc gcggccggcgc 414

<210> SEQ ID NO: 13
<211> LENGTH: 603
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albiniae

<400> SEQUENCE: 13
atgcgggccg aagcccttctga aagcaagtttatactgtgagg cgaagetttt gctcgcaagat 60
gctcgccgg gatgaggcag ggtgactgcga ttcgagggtag gcgatgctgcc gcagaacttt 120
tccacggag cgcgtgactgaa tgcggcagtttgcagccgc cggaggttctgcc gcgtgactgcc 180
tcgaggtcga tgcagctctgcc gctgagcccac cccgctgactgc gcggagctgtaa 240
gatcggatgtg gccgtgactgcc gctgagcccac cccgctgactgc gcggagctgtaa 300
tcggaggtcga tgcagctctgcc gctgagcccac cccgctgactgc gcggagctgtaa 360
cagcggccgg cggcgcgcgc gcggccggcgc gcggccggcgc gcggccggcgc gcggccggcgc 420
gcggcgcgcgc gcggccggcgc gcggccggcgc gcggccggcgc gcggccggcgc gcggccggcgc 480
gtcggaggtcga tgcagctctgcc gctgagcccac cccgctgactgc gcggagctgtaa 540
gtccagtcg cggcgcgcgc gcggccggcgc gcggccggcgc gcggccggcgc gcggccggcgc 600
tga 603
<210> SEQ ID NO 14
<211> LENGTH: 609
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 14

atgccatcgcgctgcgcgttg aacaogcag cagaagact gcggctgacacgc cgtgggtcttg 60
ttgaccgcgc acctgcgccct gtggggcgac aagctgacca gcggcagc aacaagcttg 120
gcagatgcct ggttgatgta atacgcgggtgt ctcggaactgc ccgcctgcgt gctttgcatg 180
gttcgtcagct ggttcgtgatc aagctgacca cgtcggagtgc ctcgaaactgc ctgcctgcatt 240
cgcgcgcgct gcgcgtgcacacgc gcggcgcccgc ctgcctgcacacgc gcggcgcgcccgc 300
gttcgtcagc ctggtcgttc gctgcgttcg cttcgtgatc ccgcctgccgc ctgcctgcatt 360
cgcgcgcgc gctgcgttcg cttcgtgatc ccgcctgccgc ctgcctgcatt 420
cgccgcgtcc tggcgcgtcgg ccgccgcgc gcggcgcgcccgc ctgcctgcatt 480
cgccgcgtcc tggcgcgtcgg ccgccgcgc gcggcgcgcccgc ctgcctgcatt 540
agctttggtt gccgcgcgccgcc gctgcgttcg cttcgtgatc ccgcctgcacacgc gcggcgcgcccgc 600
cgcgcgcgc gctgcgttcg cttcgtgatc ccgcctgcacacgc gcggcgcgcccgc ctgcctgcatt 609

<210> SEQ ID NO 15
<211> LENGTH: 5880
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 15

ttcggctagt gcagggcgttc gcggcgttgc gcgcgcgttc gcggcgttgc gcggcgttgc gcggcgttgc 60
cagcgtttgt gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 120
catgccgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 180
cgcgcgttgt gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 240
cgcgcgttgt gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 300
ccggcgcggtg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 360
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 420
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 480
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 540
ctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 600
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 660
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 720
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 780
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 840
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 900
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 960
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 1020
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 1080
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 1140
gctgcggttg gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc gcggcgttgc 1200
aagcgcgcg ccagcacttgga agcgcgcttg ggagtcagtg tagagccacc gtaacttgca gtttctggcag 1260
cagcctcgcg acgcctcgcg gtcgacaggcc accctcaacc ctgcacacaagt gcaccccagc gttcactgcgcg 1320
ttcggcggtg ccgccgacct gcgcacacc gagcaacgg aggaaactgg cccattggga 1380
ggagcgcggcg gttttctccca ctggcgcgcc cttggcgccag ctaaggggcgcc tttgactccgc 1440
gtgacgccag gcggcgagct gccgacgcg gactggcgtg agcagttgcta gctgagcgggc gcggcggtgg 1500
gtggccgtacc aggcgcagcgg ccgggagggc ccgggtggtgg cgagatgaga ccggcggtgg 1560
cggcgaggccc ttggcgcaca attgtcgctg aatgcgcgcgg tttggtatgtg gacagtccgctg 1620
cagctaatag gcgcgcgctgtgcgcagcctcctcgacacac acaatctgacgt ggcggcggtgg 1680
gatgagcgcgg gcggcgagcct gagcgccgctg cgcggccgcgt gcccggtggc gcggcgcggtg 1740
gtcgcgtcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 1800
gctcgtcgcgag ctggcgcgcg gccggttggtgc agtgatctgcca gtcggcgcgac 1860
cagcgcgcgcg gcgcgcgtcgcg cttcgtcgcgag ctcgagagtc agagctggatc ggcggcgtgg 1920
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 1990
gttgcgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2040
cgctgcctcgg acgcagcgggg gagtggcgtg ctcgagagcgatgatggacg caccaagagtc 2100
cgcgcggcgg gcgcgcgagcct gagcgccgctg cgcggccgcgt gcccggttggtgc agtgatctgcca 2160
cgcgtccggc ggcgagctgc ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2220
cgcgcggcgg gcgcgcgagcct gagcgccgctg cgcggccgcgt gcccggttggtgc agtgatctgcca 2280
ggtgcgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2340
tggctgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2400
cgcgtccggc ggcgagctgc ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2460
gcgagatcac gcgcggcgagcct gagcgccgctg cgcggccgcgt gcccggttggtgc agtgatctgcca 2520
gcccgtcggc ctggcgcgcgt gcgcggtggtgg cgagatgaga ccggcggtgg 2580
gacagctggc gcgcgcgagcct gagcgccgctg cgcggccgcgt gcccggttggtgc agtgatctgcca 2640
ggtgcgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2700
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 2760
gttgcgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2820
tggctgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2880
cgcgtccggc ggcgagctgc ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 2940
tggcgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 3000
tggctgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 3060
ggtgcgcgag ctcgagagcgatgatggacg caccaagagtc agacgcatcgg ggcggcgtgg 3120
aggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3180
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3240
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3300
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3360
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3420
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3480
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3540
gggagagatga taacagcggcg tgttcgcgag cgcggcgcgtgc aggaaaggtgg cggcgtgtgggc 3600
caagtgttaa tgctcgcata cgccgacgcac cgcataaccttg acgcagcgcc cgtggcgaag 3660
atgcaatacg tgctgcggtgt gtgcggagcact taagctcggttg tgcgcctggt 3720
gatgcagagc tattctgagc agttcgtggtg tcgcgcttgcctg tttttttttccc 3780
agctgcaatttg tttgctgatgtg agcgtcgtgact gcgcagcttg ggcagcgcgcgc 3840
cggcgcgcag tgcgacagcag aacccagccac cgtgcgtgctgc ggtcggggcgc 3900
gtaacttggtat cgtgtgtgcttc atgcacacag cacaggcgcctgc ctcgtgataa 3960
gtcgacgcga aaggaggccat cgctgtgaccgc aacgggagaaag atgtgtgtttg cgcaggctgg 4020
agcttgttgcag tgcgtgcggtg cgtgcggccttg ccgtgcacaccgc caggggatcc 4080
gaacacgctgt tcggccgcttg ctgcgctcgtc gcgtgtttgg tgcaggtccagc 4140
gtttggcagc agcgcagcctt ggcgcacgatc gatgattcgc aacgggagttgc 4200
agcggcgccgctc gtcgtcgcgttc atgccgtgcttg gcgtggaagcctgt gcgcgcggttc 4260
tgcgtggtgc aagttgatgtg tgtgcgtgatgc ggcgcctgctgg ggcgcgcgctg 4320
cacgtgttgat cgtgtgtgact tgggcagaagc ggcgcctgctgg ggcgcctgctgg 4380
tgcgcgagt tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 4440
caacgacggct gcgcgcggttc ggcgcctgctgg ggcgcctgctgg ggcgcctgctgg 4500
acgcgtcgcag ggcacagtgc tgggcgtgctgg ggcgcctgctgg ggcgcctgctgg 4560
caactgcgggt ggttcagctgc tagctgtccgca cgcgcagctggt gcgtgtgggtg 4620
cccgtcgcggt ggttcagctgc tagctgtccgca cgcgcagctggt gcgtgtgggtg 4680
cccgtcgcggt ggttcagctgc tagctgtccgca cgcgcagctggt gcgtgtgggtg 4740
cgcgcggt ggttcagctgc tagctgtccgca cgcgcagctggt gcgtgtgggtg 4800
tgggtgtggc tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 4860
tgcgcggtg tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 4920
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 4980
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5040
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5100
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5160
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5220
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5280
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5340
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5400
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5460
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5520
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5580
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5640
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5700
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5760
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5820
gtcgggtgta tgggtgctgg gcggagggctt ggcgcctgctgg ggcgcctgctgg 5880
<211> LENGTH: 993
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 16

tttaaactgct cgcttgagca tggtctgcc acggtctgac gcacgatgtt caggggaaca 60
tttctcgggc gaattaggt ctggccatca gcagaggata tcggtctgac cctctgtagg 120
tctcgcgtcgg tacggaacca caagattttt gcctcaacgc acggtctgac gcagaggat 180
aaaaaaaaaa tttctggatg gacaaacatg gacacactgg ccagacatgg caggggtgat 240
cggcgtgcca tccgggaaat gggccgtgcc cgccgagcgg ctgctgggca aacagcagcag 300
catcactcct tcgcgcagtc ctcgcagcgc agtgtaatga aacaggttgg ctctgtgga 360
cacaacattc tcgcgcagtc ccggctgagc cggcgacgct tcggtctgac cgttctgcag 420
caggccgctg cggcgtccat cggcgacgct tgcctggtgc ggcgaaataa aatcttacttc 480
tgcctgtgat ggcgtcttgaa aacagccttc ccttgcaggg tggcacccttc taacggtagt 540
cgcacgctg tgcgcagtc agtgtaatga aacaggttgg ctctctgcaat gccctgagc 600
tgcacccgtg tggcgcacgg cgcgacgct tgcctggtgc ggcgaaataa aatcttacttc 660
atgggagcag atgttcctcc gcggcgcagc ctgctgggca aacaggttgg ctctctgcaat 720
gacacgtcgc tggctgatatt cggcgtgatg tgcgcagtc agtgtaatga aacaggttgg 780
tgcgtggagc acgttcctac gcggggtcgc agtgtaatga aacaggttgg ctctctgcaat 840
gcggcgcagc gcggcgcagc gcggcgcagc gcggcgcagc gcggggtcgc gcggggtcgc 900
cggcgtgagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 960
gtgctgtgct gagaagagc acctggtgat gctctgcaatgccctgagc ggcgaaataa aatcttactt 993

<210> SEQ ID NO 17
<211> LENGTH: 2296
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 17

cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 60
tgcgtggagc atggtaatga aacaggttgg ctctctgcaat gccctgagc 120
tgcgtggagc atggtaatga aacaggttgg ctctctgcaat gccctgagc 180
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 240
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 300
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 360
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 420
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 480
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 540
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 600
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 660
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 720
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 780
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 840
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 900
cggcgcagc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc gcggggtcgc 960
<210>  SEQ ID NO: 18
<211>  LENGTH: 861
<212>  TYPE: DNA
<213>  ORGANISM: Xanthomonas albilineans

<400>  SEQUENCE: 18

tcagcttttc aggctggaga ccagaataat ttgcggcagc cgcttggtgc cgcgggtggc  60
gacgcggaac gcggcaacgc ccaattctgt gcagcggcgg cgagatggta caaccccttg  120
tctccagcccc agcgggtgca tcagcttttc cggtttctca gttggcgaacct gcggacgcc  180
cgcgtggcgc acctggccat ttggcctggt tgcgtagctg gacgagcacc gcggaccccc  240
cacgcgctcg cccggtgatc catgtagccg ctgagctgctg ccgccgaagcg ccggccttc  300
gggcgaagca tgtacgcttc tgtggcggg ggcgggctca cctgggcgctgg gggtggcctc  360
tctcaaccc caagcagcgt gcggcggcct ggcgggcgtt gcgcggtgat tgttggcgcc  420
tcggcgaacct ctgggggct gtgtctggct ccagcagcgt gcggcggggt ggtgtggcgc  480
tcgagctcg tagacggcgt ccaggggaga cgggtggctcccc cagcagcgt gcggcggggt  540
tcgcggcgcg ccgcgggat cttggtcctc cttggtgcat cttggtgctg gcggcggggt  600
tcggcgagcg gcggcgggg cttgccgcat cctggcgctag ctggcgtgctg gcggcggggt  660
tctcggcgtc tgcaccgaaat tgcaccgagc gcggcggggt gcggcggggt cgtggcggggt  720
-continued

cgctctgtc ccagcgcccc cgatgcgtc accacattaa gtttcttc 780
gatactccgc atgcataaag ctcggtctgt gcggcgtgtt gcggcgtgca trcatagat 840
genagactc atgctggaac c 861

<210> SEQ ID NO: 19
<211> LENGTH: 720
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 19

ttcgggtatg tgtctgtcgc gcacagcgac gagrgagataa acacactaag ctcgcaacgc 60
cgatataaat ttcgccccca ctcgcttggag gtgcatgc acacactaag ctcgcaacgc 120
cgtctcaga cggtgagtta gcggcgcaag ggacagcgggc tggcatggc 180
cgtctcaga cggtgagccag ggcggacctg agcgagcgggc tggcatggc 240
tgatctcaga cggtgactgta acgcacttcag ctcgcttggag gtgcatgc acacactaag ctcgcaacgc 300
cgtctcaga cggtgactgta acgcacttcag ctcgcttggag gtgcatgc acacactaag ctcgcaacgc 360
gcacagtgcc agggcgtatgc gcggcgttgtgc gcacagcgac gagrgagataa acacactaag ctcgcaacgc 420
cgtctcaga cggtgagccag ggcggacctg agcgagcgggc tggcatggc 480
cgtctcaga cggtgagccag ggcggacctg agcgagcgggc tggcatggc 540
cgtctcaga cggtgagccag ggcggacctg agcgagcgggc tggcatggc 600
cgtctcaga cggtgactgta acgcacttcag ctcgcttggag gtgcatgc acacactaag ctcgcaacgc 660
cgtctcaga cggtgactgta acgcacttcag ctcgcttggag gtgcatgc acacactaag ctcgcaacgc 720

<210> SEQ ID NO: 20
<211> LENGTH: 20640
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 20

ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 60
ggcggcgc gcctcgccgct gtaccccaat aagggggtta tggacgcggcc tggccgcggcc 120
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 180
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 240
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 300
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 360
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 420
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 480
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 540
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 600
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 660
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 720
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 780
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 840
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 900
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 960
ttgcccaaatg cgctctgtaa gataacttctt gcggcgccgg cgttgtgagt cgttattgta 1020
acgcttaagc atatcgtgtg tgggaggcag cgggtgccgg cctcgacacc gcaagcgttc 1080
ttcggagaaat tgcggcgaact cggggcgcgt acggagacct tcctgcgcac ttcggggttg 1140
tgtgaaacg tcggagcagc tgcgctgac aggregation aaacgacggct tttgaaacta 1200
gatgcggcagc ccttgaacaa aacgaagcgc gttgcggcag gggggcgcga ggcggcgttg 1260
aacagtgcata attgcggcgcg ctcgacacaa gatggcagat tcggctgacct ggtggag 1320
ggcgagactgt tgctgccagcc agatggagat ggcgaacaaa gggtaaagtc gcgtggagatc 1380
ggcgggtgtg agccgtttgca gaaaccggcc ggcagcggac ccgttcacctg cagcttgcgt 1440
cattacczag atagcggttta cttctgtacc ggcaaccttg ggttcaattgc cgatggcgtt 1500
cgttagatca cgaggaggtt aagaggttgt tcgtcatcag gggtaagaag taatcaccacc 1560
gccacatcg aagctctcgaat cgcgcgttac gcggtctctg gcggcatggcag cgggtgtg 1620
ltcagcagctgc agggggcggca gtaggagtct gcggctacgt ccgaaccgag 1680
tgcagcagcgg cagcagcgcgt cgcggcgcgc gcgggagcaag gccgggtgtg cggctaacc 1740
cgtaggtcgc ccgtgcggttc gcacgtccct gcggagcacc gcggggtgtg cgggaaacctt 1800
cggggccactg tgaagccgggc cttggtcgata gagacccatt tcaacgcggcata ctctggcgct 1860
ttgagcggat agggtaaagaa ccggcagttt cccagccggc gggcgccggc 1920
cacaaatggcgg cagcagctgg ttaatgctgg ttagggggtg gcggctcctg gggggcggcgt 1980
catgcccatt caagcgttgg cagggccgat ggggtcctttg gcgggctggt gcgggtgcgtg 2040
gggtggcagtc aagctgtggt gcgtggcgtgc ctgggctcaga cgggctcata cagcttggag 2100
acgtgaggcgg ctaagggtgct cgggtacgag gggggcggag gggggcggag gggggcggag 2160
cgtgcggcagc tggagcagcgc aacatcgtgca gggagccgct ggacagcttt ttgagctgac 2220
tgcgaagcgg cggaggggag cggagccact gcacggcttg gcgcttccct gcggagtgcag 2280
gtcagcagcgc tcgctggccgt cgaataaccg ccgcccggac hgcggacgcgc ggtgaaggcc 2340
tactcggcggc tcgagggcgtgc agcggcggagctctgccct gcgggtgtgat 2400
ctggagattgc gttgtcgcttg cttcggctgc ggggttgatt gcgggtgtgat 2460
cacggcgcagc gaggagttgg gcggctgcag gggggttgatt gcgggtgtgat 2520
ctctgggggc ggcggcgaggt gcgtgggtgtct gcgggctcgag tcggcagcttc atacgctggc 2580
ttggcggcgc ggcggccgca acgtgggcgt caatgtgtagg ttacgctcgcct gcgggtgtgat 2640
ctggagatag ctggcgcgctg ggggtcagat gcgtgggtgtc atcctccagg cccggcggcc 2700
gttctggcag gcggcgcggc gcgggttgtgt gcgtgggtgct gcgtgggtgtg 2760
cggcggggag gcagcttttt ggttcggagt gtctcgcggt gcgggtgtgat 2820
gtctgtgttc cccttggcag gcgggtgttc cttgctggcc aggctggctg gcgcgggagt 2880
acggcgcggc gcagcttgcgt gcggctgcag ggggtggtgt gcgggtgtgat 2940
ctcggcggcgg cggccgcgcgt gcgtggcttc cttcgggttc aggctggcgc gcgggtgtgat 3000
aaccagccgg ccggcgcgct gttctgcgcagg gtcggttgct gcgggtgtgat 3060
tgcagcagcgc tcggcgcggc gcgcggcgtg ctgctgggtgc ggttcggcgc gcgggtgtgat 3120
cggcgcggc gcgcggcgtg ctgctgggtgc ggttccgggcc gcgggtgtgat 3180
atgcgggagtg gcgcggcgtg ctgctgggtgc ggttccgggcc gcgggtgtgat 3240
cggcgcggc gcgcggcgtg ctgctgggtgc ggttccgggcc gcgggtgtgat 3300
acggcgcggc gcgcggcgtg ctgctgggtgc ggttccgggcc gcgggtgtgat 3360
-continued

cgagtgctg cattcaagtg ggtggccag cattccgcat ggccgtgcag cattccgcat cgtggccatc 3420
cagggccag ctcctcgccg tgcggccggt atccagctct tcacgtcctg cgtggccag atccagctct 3480
ggcaattgca ctgggtgcaac gcacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3540
ggcggtattct gctagccttg gttggtccac gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3600
ggctgcttgcct gctgtgcttgcct cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3660
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3720
ggcggtattct gctagccttg gttggtccac gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3780
catcatcgcc tgcgtgctgct gcacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3840
ggctgcttgcct gctgtgcttgcct cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3900
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 3960
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4020
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4080
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4140
ggcggtattct gctagccttg gttggtccac gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4200
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4260
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4320
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4380
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4440
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4500
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4560
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4620
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4680
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4740
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4800
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4860
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4920
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 4980
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5040
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5100
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5160
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5220
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5280
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5340
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5400
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5460
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5520
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5580
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5640
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5700
tacaccccttc aggtgaggtcg cgacagctgc gcacagctgc gcacagctgc gcacagctgc gcacagctgc 5760
aacactgacag tcaagtgtgtg tgcggcagac gtttgcgca ctggcccccct tgatgctctg 5820
cggtggctgt cttcactcaca gagaactacc aagcggccg ggaagcttaa ctaaggccgcc 5880
gggtctggct atggctgaag ttttgccgac atatatgcgg gacggggtttt gagggttaaag 5940
acatcaactt gggtgctaagggggacagtgg gcgtcgttag cggggcagag gggcgcggg 6000
cgcgtgggcccc aaacacacc tgcgtcgtat cggagcgUGC aagctgtaaga gttttgctgc 6060
ccggggtgtgt gcggctgtgc gctacgtcgc catctcacgta agatogaacgc tgaactacgc 6120
atcgggtgcct gggtggtagt agagggcagag agctgctgaa cttcctctgca cgacggcgcc 6180
gggcctccca gggggcagcc tggctcgcgtg ttcgaagtgtg cgggtgaatct gggcgcggg 6240
cggcaacggg aacactgcaac gggggcggtt cgggtcgtgag ttcgaagcct cgcaggcagc 6300
tctggaacct gggtggagca cgcggtggttg ttcgaagccca tgcgcggccc accggcgctg 6360
tacgtgagct ggcgtggaagca cgcggtgccg atgcgttggcc tgtatgggtta cccagcagc 6420
eaacggggggag cgcggtgagc gttggcgcac aacggggcag atgacgcacg ggtggtcttg 6480
aacagcgcttc gcgggagccag cgggtgcggtt cgggtcgtgat ggcggggttc ggggggtcc 6540
ccgggtcgca tggcggttggt cggggcaggt tcggggtgctg ggcggggtcc acgcggccta 6600
tcgttctggg cgggtgcggtt cgggtgcggtt cgggtgccaag aacgacccgc acgggtcttc 6660
gattacttca aacgacactc agtacagcag atgattgcatc actgaagcag aagctactag 6720
tgacgcgaact ctgggctgcct gacgtgcggt gccgaactgtgc gagacgcatg ccgggagccg 6780
gcagcgaacg gccggtggaat gttgacgtca tggaggaata tctcctcact 6840
gaccgtgtgc cttctcttctgc aacagcagatc tcgacccgac gacgtgactc 6900
tgcagccggct ctcagcgcgcc gggggagtgc gagaactcgc gcgggagcgc gggggagtgc 6960
gggggagccag gggggagcgc gggggagcgc gggggagcgc gggggagcgc gggggagtgc 7020
tggcgtacct ccccaaccct cgcggctctg cttgcttcaca cgaagtggc 7080
gagcggcagcc gtcggtctgc cccagccctc cttgctgcgt cgggctgctct 7140
gacaagcttc atgagcgtgtc cgggtgcggtt ggcgggtcggt cgggtgcttg 7200
tcggagggtg ccggtcgtgc cgggtgcttg ggcgggtcggt cgggtgcttg 7260
gcgcatctct gtggctggctg cgtggagata tggctggcag ccaagctctga cgaagtggc 7320
cggatgtccaa aagattcagc tgcgggttcag cggggggagc gagctggcttg 7380
tgctctcttc cactatttgcg gacgctctgct ctcgactttc ctcagtctct cgcagtctct 7440
caccctctgg gccggtgcct ctcgactttc ctcgactttc ctcgactttc cgcagctctc 7500
cggcggactc ccccaaccct cttgtgtggtt gtcggtcttg aacagcaga cgggatcttc 7560
tgggtgtggct cttgtgccac ggtctcttcct ctcgactttc ctcgactttc cgcagctctc 7620
gacatctgctgc aggagcgttc aaacagctgtg ggcgggtcggt tgcgactttc ctcgactttc 7680
atcgagccgctg gcagcgaactc ctcgactttc ctcgactttc ctcgactttc cgcagctctc 7740
gcctcgagat tggggtgcggt ggcgggtcggt tgcgactttc ctcgactttc cgcagctctc 7800
tcggggagaac gcggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg 7860
agaagcgaacc gcggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg 7920
ccggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg ccggggtggtg 7980
atggggaggg gggggagccg gggggagccg gggggagccg gggggagccg gggggagccg 8040
gggggagccg gggggagccg gggggagccg gggggagccg gggggagccg gggggagccg 8100
-continued

gcagcctctctgtaaatggcctcgacagcggctgagcgttgggaactcgatcggcagagaagggatatctgcagcagctctgacctgacaggcctcgtggttggactggtacgctgttggagttcagcgttgccttgagcttgccagcgtcctgctgccttctctgcaagtgcctggtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
-continued

tgccaacgcc aatggagacct ggcgcgttcgg cttggctact ggcgcgcggcg actgggagcc 10560
tccacgagcc gctggcactt gccctcaacgc cggcgcgcggg cgcgcacgcgc gcgcgtgcccgc 10620
gcaggggtgg ctaaactaacc catcagcggcg caactgggcgc aacaggtaggg gcgcactaagc 10680
cacagtaggc aacgcaagcgt ttctcatgaag cttgctgcagc gcgcgcggctt ggtgcgtggc 10740
cggcagcgcc atctcaaggg cgttgctgccag gcgcgcgcggc ctcggggcgg ccacagccgtg 10800
gacgcggaag agtcagaggcc cttttctcata aataattggcg cgcgtgcgagc gcacacgctac 10860

10920

gcgcacagct cgctgcgtgtt cagacacgtg ttgcagggcgc tcgcggctca gcgcgcagct 10960
agccagacct cgctgcgtcgc gcgcctgtcg gcagccacagag atctccgcgg acgcgcagat 11040
ggcgattgggc cggagggaggt gcgcgcgtcgc gcagcagggc tggggtcgga gcgcgcacacg 11100
cgcagcagc tcgccgatggc gttctacgcgc gacgcgcgtgc gcgtgcgagc gcagcctttga 11160
tacgcgcagcg aactcttctga cgcagcagag tcgcggcggag tcgagtgccgc ccacacggcag 11220
cgcctgggag cagttggcgg cccccagcagc ctggccgggt gcacagttggac cgacccctgg 11280

gcgcaagacgc gcgggtcgcttg tgcgcgtcga aataacccag gtaccgcagc gcacggggac 11340
agcctgcggcg acgcctctgca gcgcgcggcgc cggcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11400

gacgcgtgggt gcgtctgggag ctatgcgcagag ctgggggcgc gcgcgcgcgc gcgcgcgcgc 11460
ggcgccggag gcggggcccgc gcgcggacagc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11520
cgcgcagcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11580

gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11640
gtcggtgacgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11700
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11760

atgcgccttgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11820
tacacgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11880
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 11940
ggcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12000
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12060
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12120

ggcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12180
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12240
gtcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12300
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12360
ggcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12420

cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12480
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12540

cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12600
ggcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12660
cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12720

cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12780

cgcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 12840
ccacgcaacc aggcggagca acggctgagg gcactgttttg gcaaggtgct gcgggtggag
12900
cagttgagca tccacgaca cttttctgcc ttggtttggg acctgctgct tcgcatcgca
12960
cgtatctcg gcattttgac gcatctgtc gcagatgtca gcccttgcat gcgtttcag
13020
cgcacacagc tgcgggagtct tagcgaatcc ttcgacgagg atgcacgagg ccagggattc
13080
gattcgtccg ccgggaacgc cggaggagcg accatcgcag cttgcacagt gcagggcgc
13140
atgtggttcc gcacaaatgc cttggcggag acgcgattaca acacccaggag tctgcgttca
13200
ttcgaaggag aatctggtcag ttcggccttgg cagttagcag ttcgctgtgt getagaaaccg
13260
cagcggctgca tgctgatccca tttgctggaa accagggcag aatcggtcata ggttaattgc
13320
dgcagcagage ggttgttgcct gcagacttagg tgcattcgcag aacggagttga ttcgaattgc
13380
catttcgaca cagcggtcag cgaaccccttc gatattacag gcggctggcc attgccgccc
13440
ttcgtgtata gctgsgggaaga cccgggcaatt caactctgcaagt tggcctcag tcgcaatgc
13500
ttcgaggtct gcggacccct ctactctgtt tctgctgttg gcaggtcctg ccaccaaca atcgcgctagc
13560
cgcctgattgc aatccggcct ccgagcagcgg ccggcctgag gttttccccag
13620
tgcggagcctg cgcgggctga ccagggagac gcgtgctgca agcgggtcaata tggggaaaaag
13680
cagcgttccg atgcacccggc gctcgcgtttt cccagacgact atggcggcggcc ggtttcccccag
13740
aacattcaat ggcggcgtctg ggccgctccag atcgagcggc gcgtgctggca agcgctgcaag
13800
cgttttttcgg acgcaactaa cttcatttctg cagatcgcag tcgcaagcagc ctgagcgccc
13860
gtagttgctc gcacaaacag cttcgctctgc ctggcgtgtgc gccgatcggc
13920
gtcggcagcc agttcggcagc ttcggctggtc ggctgcttcat gcagccacag cggcagccgcc
13980
gatgtggagca ccaattggttt ggctgctgca atcgccagcgg cagatcgcag cttcggcagc
14040
cgcctgctaga cgggacggg gtcggtgggtgc gcgacttcat gcagccacag cggcagccgcc
14100
cgggatgca cggcgatcag gctgggagag atctggggcgc gatggcgctca ttcogcagcgg
14160
tctgggttctg cccctggtttg cgcgggttttc gcggtctcag ccaactgtag cctggcggggc
14220
aaagcggcgc acggcgggc gtcgggcttg cgcgggactg gttgtgggtc ccgctgtgca attccttcag
14280
tggcgcctgat cttcggctgc gcggctgagc gccgctgtcag tcgatcagc gcagctgtttg
14340
tgctgctgctg ctggtggctg gcggctgctg gcggctgctg gcggctgctg gcggctgctg
14400
aagcggcgc gcggctgttc gcggctgctg gcggctgctg gcggctgctg gcggctgctg
14460
ttacctgaggg cattctggtg gcctggccgt cggctggcctt cggctggcctt cggctggcctt
14520
gctgctgctg gcctggcctt cggctggcctt cggctggcctt cggctggcctt cggctggcctt
14580
gtcgggtctg gcggggacag gcggggacag gcggggacag gcggggacag gcggggacag
14640
gctgctgctg gcggggacag gcggggacag gcggggacag gcggggacag gcggggacag
14700
actggcagg gcctggctgg ccaacttttc gcgggggctt cggctggcttc cggctggctgg
14760
ttcgcattg cggcggttctg cggctggctgg cggctggctgg cggctggctgg cggctggctgg
14820
gtcgctgcag gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
14880
gctgctgctg gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
14940
ttcgctgctg gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
15000
gtcgctgctg gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
15060
gtcgctgctg gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
15120
gtcgctgctg gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
15180
gtcgctgctg gcggggggtc gcggggggtc gcggggggtc gcggggggtc gcggggggtc
15240
ctggctgtgc acggccgccc ttggcgcgat gcgcaccata ccggctctga taacgcttgct cgcctggttg 15300
cggagagtgg gcggctccgc acggcagtcg ggcggctcag ggagacgaca ggctggtgac 15360
gaatattgta gcgcggatct gcgcacccgc gccagctggc gcagtgtcgtg 15420
catcggcccg tgggtgcgcc ggagcgtcgg tggcaggtgc gcacgacacc 15480
ggcgcaagtgc cggcggcagt gggctgccgg ggtgtgctgg gcgcggctcgg 15540
tacaaaacgc agcgtatttt gcgcggctcgc gcacccgctg cacggggagt ggctggtgcc 15600
cggcagcgtgc gcacggtgtt gcgagctgatgc gcgtggtggtt gctggggctcc 15660
cgcctgttgg gcgtagatcg ccgcggcact gcgcgttccct ggcggcgtgtt gttatgctgc 15720
cttgctagct atctgctggc tcgcctggat gcgcgttca gcgagctgcgt gcgcgggtgc 15780
agatcgcgag cggcggcagt ggacatttgc cggctgctcg acgtctgctg 15840
cgcctgctgc actgctgagct cctacaccct gggcgagtttc accgtggcct cttggtgtca 15900
cggcttttgg atctgctggtg gacgtgtgtgg tacaccagcg caacggtggc gcggtggttg 16000
gttggcgttgt gc ggcctggtggt tgtctggcgc tggctggctcc gcggcggcctg 16060
aggtggcagc atcgcgtgtg acggcggcgt gcggctgtgg gcggcgcggcgc 16120
ccgctgctgc acgcggcagg gcggctgtgg gcggcgcggcgc 16180
gaggtgtggt gc gtcggctgtg gtttgggagt gacggcggcgc gacggtgttg 16240
tgttgtcgtgacctggcgtgc gggctggccg tggctgctgct gtttgggagt gtcggtggtg 16300
gccgcggcag cggcggcggc cggcggcggc cggcggcggc cggcggcggc 16360
ctgcgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 16420
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 16480
ttcgctttgc gcggctggttgc gtcggtggtg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 16540
gccgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 16600
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 16660
ttcgctttgc gcggctggttgc gtcggtggtg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 16720
gccgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 16780
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 16840
ttcgctttgc gcggctggttgc gtcggtggtg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 16900
gccgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 16960
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17020
ttcgctttgc gcggctggttgc gtcggtggtg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17080
gccgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 17140
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17200
ttcgctttgc gcggctggttgc gtcggtggtg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17260
gccgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 17320
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17380
ttcgctttgc gcggctggttgc gtcggtggtg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17440
gccgcgcgtg ccgctccgca cccctgggtt cttcggctgt ctgccggcgt gtcggtggtg 17500
aggtgtgcgt gc gcgcgtggt gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg gcgcggcgcgg 17560
-continued

ggcacgcagt tccaacgcca ggcacgcagc acgcacgcagg ccatacctag ctgtgacgcgt 17640
gggcgagtctt gcagctacac gacagtcgcg gacatgcgcc acacagctggcc acagcgcgtg 17700
cgtgggacgc gcctggcgcg gactggcgtgct gtcggctacct gacagcgcgac 17760
cgtcaattgg ccctggcgtcg gcgtacgtac gcgtacgtac gcgtacgtac 17820
gctacccgct cgcacgtcag gcagcgcaccc cgcacgtcag gcacgtcag gcacgtcag 17880
ggacacagcct gcagctacag gcagctacag gcagctacag gcagctacag gcagctacag 17940
cagcccaccg acgcgtcag gcacgtcag gcacgtcag gcacgtcag gcagctacag 18000
gactgcagct gcagcgcagg gcagcgcagg gcacgctacag gcagctacag gcagctacag 18060
tgcgttgacct gcgtgcacgt gcagcgcagc gcacgtcag gcacgtcag gcacgtcag 18120
gctgcacgt gcagcgcagc gcacgtcag gcacgtcag gcacgtcag gcacgtcag 18180
ggcagcgctg ggcacgctgg gcagcgcagc gcacgtcag gcacgtcag gcacgtcag 18240
gcgtacgctgc gcagcgcagc gcacgtcag gcacgtcag gcacgtcag gcacgtcag 18300
ggcacgctg ggcacgctgg gcagcgcagc gcacgtcag gcacgtcag gcacgtcag 18360
gcgtacgctgc gcagcgcagc gcacgtcag gcacgtcag gcacgtcag gcacgtcag 18420
gcgtacgctgc gcagcgcagc gcacgtcag gcacgtcag gcacgtcag gcacgtcag 18480
ccggggtacc gcgtggtgct gcacgtcaga gcgtggtgct gcacgtcaga gcgtggtgct 18540
cgcgtggtgct gcacgtcaga gcgtggtgct gcacgtcaga gcgtggtgct gcgtggtgct 18600
atgggcttgc gcacgtcaga gcgtggtgct gcacgtcaga gcgtggtgct gcgtggtgct 18660
tatctacgg ccctggcgtcg gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 18720
gacgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 18780
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 18840
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 18900
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 18960
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19020
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19080
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19140
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19200
tgcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc 19260
acgcacgcac gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc 19320
ggcacgcac gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc 19380
atgggcttgc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19440
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19500
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19560
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19620
gcgcttgcc gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac gcgtacgtac 19680
tgcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc 19740
tgcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc 19800
tgcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc gcgtgcgtgc 19860
atgcgcagct gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc 19920
ggtgcgctgc gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc gcacgcacgc 19980
cagcagcagc ctacacaca cacotatccg gctgaacttg tacaagccct ttggagcttc 20040
gtatcggcact atcagctgac cttgttctctt gcggctgttg ccagcttcgc gctcgttgtg 20100
acacacetgac gcaagccgag gatctgatgc atcagcttaca caacagcagc gagcagacagc 20160
cagtgcactg agaactccgatt gcttcttctc atacacactct tgcctttgtcg ctctgcgttg 20220
gagccgctgtc caagcagtcgt ccagcagccg gatctgagct gatcagcagc 20280
ttctgaaacc aggcgtatacc ttgagacacag ctgtctaaac ccctgcaaca gcacacgtgac 20340
acaccgcacggt ctcctctctag ttcggtattg atgagcttacg acaactccagc gagacagactc 20400
ggagcactgga ggcagtcggag cggctacgag gttgacccag gcttgtccgc tgcaccaccccg 20460
atgagatag tggcttaaat cttgggagccgct gtagcgggag ttcagcgatc aatgaaatgc 20520
gcgagggcagc tgtctctcagc aagcagcactc cgccgcatga cctcagcatca ccaactgcctg 20580
cgagctgagc tggctgcggcg ccatgaaggc gcacagctgcc cttggagatg tgcagactag 20640

<210> SEQ ID NO 21
<211> LENGTH: 1032
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albinelines

<400> SEQUENCE: 21
atggaattcgc ggttaaatgc atctgtcattt accttacgat atgtatagtag atctggagttgcagc 60
gccatactgac gccggtgcggc gatacgagct gctgttattct ggctgtgcttg cagcagcagc 120
ggagcagcagc gagagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 180
tggagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 240
tttcctctg atgagcagcagc gtagcagcagc gtagcagcagc gtagcagcagc 300
ggagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 360
gtggagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 420
cagtgcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 480
gcgagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 540
cagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 600
gtgagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 660
cgcagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 720
gtgagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 780
atgttctgagc caaccttctct gcacacacctc gtagctgctg atgtagcagc agtctgcttg 840
agaagcctgagc gcagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 900
gccgagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 960
ggcgagcagcagc gcagctgagct gcgtccagcagct gcctgtgcttg 1020
agggcccat aa 1082

<210> SEQ ID NO 22
<211> LENGTH: 504
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albinelines

<400> SEQUENCE: 22
atgatcgaagc ggctgctgagc cctgtgtggg gccaaacagc gactcgttgctg cagcagcagc 60
gagctctggtat caacccagc acatgtgtgtgc gcacagctgcagc 120
-continued

atgatcggaga tgctctggga ctatctgcgt tcctgcaagt ctggtggactg gcgtgtgcac
  180
cagcctttatg gggtatgccg tcggggcgcg tgcgggcacc tgcgggaggg tacgggacgc
  240
atggggcgcct ccatgtggct gcgctgccggg caaccgcttg gctggctctg caatctgcct
  300
aaccacagtc cggagctcgct gagtgccccatgc tggcgcggacgc gctggcaaacca
  360
gggggcattga cccagctgatc aagagtctgc aatgtgacgc cagagatcctg gcgcctgcct
  420
gtgcgcgggct ggcgctcgct ggtgctaccg gcaggctgcat gcggctgtgc acgcaggtgc
  480
gccagcagag gtcgctacag gtga
  504

<210> SEQ ID NO: 23
<211> LENGTH: 2826
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albiniae
<400> SEQUENCE: 23

gtgaacctgaa ctgcaactgt aaccaagcct acactcaagt ctgcaaggcc ggtataacg
  60
cacgctctcg tattcactgt gctgagggact gacctgacttc gacaccaagc tgcgccgct
  120
gctctctcg cggagctcgct aatccattgg ggcagctgcc gacactctggct cgcctcag
  180
gcacagtccg tgcgctcgct gcctgccgctggggcgcgct gctgctgggg gctgctgctg
  240
cgccagccag aatccattgg cggagctggct gcctgccgct gctgctgggg gctgctgctg
  300
cagactccag ctcctggccgt aacctcaagt ctggtctggt ggtgctcccc ccgctgcgct
  360
cgctccgct ggtgctgctg ctggtctggt ggtgctgctg ctggtctggt ggtgctgctg
  420
gatgctctcg tgcgctcgct gcctgccgct ggtgctgctg ctggtctggt ggtgctgctg
  480
tgcgctcctgc ctggtctggt ggtgctgctg ctggtctggt ggtgctgctg ctggtctggt
  540
gggtgctctgc tgcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
  600
gaacgctctgc tcgctggcctc cgccttctgc ctggtctggt ggtgctgctg ctggtctggt
  660
gacgactctgc acctgcgcctgc ggtgctgctg ctggtctggt ggtgctgctg ctggtctggt
  720
ggcggcgcggc gtcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
  780
acacocccgc aatccattgg cggagctggct gcctgccgct gcctgccgct gcctgccgct
  840
cgctccgct ggtgctgctg ctggtctggt ggtgctgctg ctggtctggt ggtgctgctg
  900
ggctggagtt gcggctgctg ctggtctggt ggtgctgctg ctggtctggt ggtgctgctg
  960
acacocccgc aatccattgg cggagctggct gcctgccgct gcctgccgct gcctgccgct
 1020
gtcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
 1080
ggcggcgcggc cgcggcgtctg cgcggcgtctg cgcggcgtctg cgcggcgtctg cgcggcgtctg
 1140
acacocccgc aatccattgg cggagctggct gcctgccgct gcctgccgct gcctgccgct
 1200
ggcggcgcggc cgcggcgtctg cgcggcgtctg cgcggcgtctg cgcggcgtctg cgcggcgtctg
 1260
acacocccgc aatccattgg cggagctggct gcctgccgct gcctgccgct gcctgccgct
 1320
gtcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
 1380
acacocccgc aatccattgg cggagctggct gcctgccgct gcctgccgct gcctgccgct
 1440
gtcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
 1500
gtcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
 1560
acacocccgc aatccattgg cggagctggct gcctgccgct gcctgccgct gcctgccgct
 1620
gtcgctcgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct gcctgccgct
 1680
tcggtgcctgc ttgagggact gtcgctcgct gcctgccgct gcctgccgct gcctgccgct
 1740
<210> SEQ ID NO 24
<211> LENGTH: 837
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 24

   1 aatccctggtca cttccacatt gaaacgccat ttcgggcaac ggcgctgcatgc ccacagcaac  
   180
gttgcaacg gcttgctcaca gacactatct aagctgggttag gcacagatgc  
   1860
   gggtgtaagc agatgtctcg ggggctgccc aagagcagc ggtcctgtag ccacagcaac  
   1920
   gacgtgcttc ggcgcagcact ttcccagacct cagattatatg tgaagagccc gagctgac  
   1980
gcgaagatct accattgcat gcagacagcgg ggcgccccgcc ggtcctgacac ctctggctgct  
   2040
   gcacagctac gttctgctgaa aagttagcgt ctcagcaacact cttacagcata ctcagctgac  
   2100
   agccagaagc acaaatagct cttctgctac gatggcggca ctctacatgca caagagaac  
   2160
   cttggtgtata cacaacatttc gttccctcata agttgggaaag tcggggtagag ccatatccca  
   2220
   gggaagagcg cagagcgcgaa gctggagcag aacacgcctga tagagagacgc caagttctcg  
   2280
   ggcgggctgca agcagatctgct ctatgactca aagctgcatgc atacccagcc ggggagctcg  
   2340
   cagctgatcg ctaactaagt ggccgagcag gaaaacgatc cggcgccatt gcgcgaatttt  
   2400
   tcggtcgtgc cttgccccgca gtacagcttgcc ctcgctttatt gtcgctgctgc gcacggaatg  
   2460
   cggtgagcgc gcaatgcccag ggtgaacccg ccagcactcg gcagggcgag gttgagcaag  
   2520
   aacccgctcg cagggcctac gggctgcagcg aagccgctag acagcacact gcgtgcaaatc  
   2580
   tggaagaggg tcggctgattc gcaaacggtc ggcgcttgagc cagacatctttc gagctgggcc  
   2640
   ggccactcgc tggcgccttg cagttgcctc tacaagctag ccagacgctca ccagaaatct  
   2700
   gtcagctac ccagcatttca ggttaaactc acaatcgacaggtgctgag gcacgctgctg  
   2760
   gcataa  
   2820

<210> SEQ ID NO 25
<211> LENGTH: 1905
<212> TYPE: DNA

<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 25

   1 atggcccaatg cgttacgat cgcagggcgc ggggacccct cgcagcgcca aagctgaagc  
   60
   cagaggtgct gacggcttcag cggcttgagc gcaggtcacag ctggctgacag gtcgctgcaag  
   120
   ggctagcaac gcagccctgc gggctgcagcc ctcggtcata gcgtgacagc ggccagcgtgc  
   180
   ctttacgctg tcggctgatcg ggcgcttgagc gggcttgctgc gcgaatttcc  
   240
   atggctcggc cgcagcacat cgcagcctgc ggcgcttgacag gtcgctgcaag gttccgctgc  
   300
   ggccggctgct gcgcgatcgc ggcgcttgacag gggctttgag ggcgctggag cgcagcagca  
   360
   gacattcgcag ggcggctggc gcagctcgcc gcagcttgctg gcggctggagc gcggctggagc  
   420
   tggcagcttc gcgctggcag ggcgcgagct gcagctcgcc gcagcttgctg gcggctggagc  
   480
   ggccgagctgc gcggctggagc gcggctggagc gcggctggagc gcggctggagc gcggctggagc  
   540
   atggctgctg gcagctcgcc gcagcttgctg gcggctggagc gcggctggagc gcggctggagc  
   600
   gtcgctgctg gcgacgcttgcc cggctgctg ccagctcgcc gcagcttgctg gcggctggagc  
   660
   gtcgctgctg gcgacgcttgcc cggctgctg ccagctcgcc gcagcttgctg gcggctggagc  
   720
   tggcagcttc gcgctggcag ggcgcgagct gcagctcgcc gcagcttgctg gcggctggagc  
   780
   ggccgagctgc gcggctggagc gcggctggagc gcggctggagc gcggctggagc gcggctggagc  
   837
ORGANISM: Xanthomonas albilineans

SEQUENCE: 25

atgagctgtg  aaaaaaacaa  ggaagctctg  gggattcaga  ctaggtcaca  aacactggtg  60
cagctgatga  ttcattggt  gatccgacaa  aaggaactct  ttcagcagcga  gctgactcgc  120
aaagctgctg  acggggcgca  caaaaactgct  tttcaggccag  tggctcaagc  ggacactttg  180
gccgctgctg  cgcaacgtcat  cacgctgcat  gttctgcgca  agaagcgcgg  cacccgctc  240
atcgcacgca  aacggtcgcg  gaggagctgg  cgcctcaccg  ggcgcgtc  300
gcacaactcg  gcacccgca  tttcttcaga  cactgtcgg  gcgtcagaa  gaagcctact  360
caacctgctg  gcacgtcgg  tgtggtcctg  tacatactct  gcacacttcg  egactcaagtg  420
gactgtgcac  gcgcgtgca  ccgctgtgac  gccaccgagc  tctgcactgc  gtctcctgcgt  480
gccggaagcg  agttccagct  gcgcaccact  gacaaagcgc  agcgggccc  cccgctgccg  540
tccgacactga  agacggaga  gaaaagtc  gcgccgctt  ggaagttcgg  cagagctcgt  600
cggcaacatc  ccgaccaact  cgcctgtgac  atgcagctaa  tcaaggaact  ctaacggccc  660
gcaccagca  agcggccagag  cccgggtggg  gagacagtca  atgcgcactg  cgcctgcttg  720
acggggtcgg  gcacccgcat  caagagcagc  gaaactacaa  acgtgtcacg  gcacactgctc  780
cacgacacg  aaaaaacggg  ggcgtgggac  ctaacaagag  tgcagggaga  actggaattc  840
accccrctgg  tgcacccgca  cgcggcgcac  gcttcgaca  tctgccagcg  gatgcctgct  900
cggcggtcga  acctctgctg  gaggcggcct  ctcctcaaa  aagcgcgca  aaacttcctg  960
cgctgtaccg  cagccctgtg  ccagctgccg  tcttcgagct  ggcgctgtgc  gcgcggctgt  1020
tgctggtaaaa  ttcgctcagc  tggctcgttct  atgcagctga  tgaatgctgg  gcgaccaag  1080
cgcgcactgc  acatactccagc  aaacccgact  cgcacacctta  cccgctgccg  cagctgtgcg  1140
tggaaagact  tggcgcagct  gctgagacgaa  gtytccgccc  aggaactgcagg  caacggcga  1200
aacatccgcc  gcgtggtcag  cttccgctc  accacacgcg  ggcacgcacg  cccggagaatg  1260
tgctggtcgg  acactctggc  gcggagttgg  gacagggaga  aagggcaggg  acaaggctagta  1320
gggcgaagct  acggcgccaat  caagggcagc  cggacactgg  aggtctccgg  caaagggcagc  1380
atgaggtgctg  tgcgcctcagc  gcacgcctac  gacagtgctg  tgaatgctgg  tcttcgtcgg  1440
ttcgccacgca  acctctcttg  gagaagccgg  cgcgggtca  tggacccgtg  caagtcgacgc  1500
agcgaagaa  aaaaaacggg  gcggagagaa  gcgcgcagag  ccaacgaca  gtcggtcgg  1560
cgcaccccg  aggtactcagc  gcacgggtctg  ggcggctcagc  cccggcagcgt  1620
gtggcgcgg  cgctctccgtc  cagggcagag  ggcgcagctg  ggcgcagatc  1680
tggacagcc  gcggagggaga  gtcgggctgcg  agaagcgcgg  tggctcagttc  caacgcgcgcg  1740
cacccgtgctg  gcgcagaact  gcgcacgtgc  atgcgtcctg  gccgggtgagg  1800
cggcggtcgtg  tggcgcctgcg  cgcgggcgagc  gcgggcagaa  gcgcgcagctg  1860
tgcgcgctgcg  gcgcggcagaa  gcgcgcagctg  1905

SEQ ID NO: 26
LENGTH: 6879
TYPE: PRT
ORGANISM: Xanthomonas albilineans

SEQUENCE: 26

Met Pro  Ala  Leu  Met  Gln  Ile  Thr  Leu  Val  Ala  Val  Gln  Phe  Ala
1  5  10  15
-continued

<table>
<thead>
<tr>
<th>Gln Pro Gln Leu Met Ala Gln Cys Gly Ala Tyr Ile Ser Gly Ser</th>
<th>865</th>
<th>870</th>
<th>875</th>
<th>880</th>
</tr>
</thead>
<tbody>
<tr>
<td>His Leu Thr Met Arg Ala Ser Arg Thr Pro Arg Asp Phe Thr</td>
<td>885</td>
<td>890</td>
<td>895</td>
<td></td>
</tr>
<tr>
<td>Gly Pro Ser Glu Val Ile Asn Ser Ala Cys Ser Ser Ser Leu Val Ala</td>
<td>900</td>
<td>905</td>
<td>910</td>
<td></td>
</tr>
<tr>
<td>Leu His Arg Ala Val Glu Ser Arg Ser Leu Val Ala</td>
<td>915</td>
<td>920</td>
<td>925</td>
<td></td>
</tr>
<tr>
<td>Leu Val Leu Glu Val Asn Leu Ile Leu Ala Pro Lys Val Leu Leu Ala</td>
<td>930</td>
<td>935</td>
<td>940</td>
<td></td>
</tr>
<tr>
<td>Ser Ala Ser Ala Gly Met Leu Ser Ser Pro Asp Gly Arg Cys Lys Thr Leu</td>
<td>945</td>
<td>950</td>
<td>955</td>
<td>960</td>
</tr>
<tr>
<td>Asp Ala Ala Ala Asp Gly Phe Val Arg Ser Glu Gly Ile Ala Gly Val</td>
<td>965</td>
<td>970</td>
<td>975</td>
<td></td>
</tr>
<tr>
<td>Ile Leu Lys Pro Leu Ala Glu Ala Leu Ala Asp Gly Asp Arg Val Tyr</td>
<td>980</td>
<td>985</td>
<td>990</td>
<td></td>
</tr>
<tr>
<td>Gly Leu Val Arg Gly Val Ala Val Asn His Gly Gly Arg Ser Asn Ser</td>
<td>995</td>
<td>1000</td>
<td>1005</td>
<td></td>
</tr>
<tr>
<td>Leu Arg Ala Pro Arg Val Asn Ala Gln Arg Gln Leu Leu Ile Arg</td>
<td>1010</td>
<td>1015</td>
<td>1020</td>
<td></td>
</tr>
<tr>
<td>Thr Tyr Gln Glu Ala Gly Val Gly Pro Ala Ser Val Gly Tyr Val</td>
<td>1025</td>
<td>1030</td>
<td>1035</td>
<td></td>
</tr>
<tr>
<td>Glu Leu His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Ile</td>
<td>1040</td>
<td>1045</td>
<td>1050</td>
<td></td>
</tr>
<tr>
<td>Gln Ala Leu Lys Glu Ala Phe Ile Ala Leu Gly Ala Gln Ala Ala</td>
<td>1055</td>
<td>1060</td>
<td>1065</td>
<td></td>
</tr>
<tr>
<td>Pro Ser Arg Cys Gly Ile Gly Ser Val Lys Ser Ala Leu Gly His</td>
<td>1070</td>
<td>1075</td>
<td>1080</td>
<td></td>
</tr>
<tr>
<td>Leu Glu Ala Ala Ala Gly Leu Thr Gly Leu Ile Lys Val Leu Leu</td>
<td>1085</td>
<td>1090</td>
<td>1095</td>
<td></td>
</tr>
<tr>
<td>Met Leu Lys His Gly Glu Gln Ala Gly Thr Arg His Phe Ser Thr</td>
<td>1100</td>
<td>1105</td>
<td>1110</td>
<td></td>
</tr>
<tr>
<td>Leu Asn Pro Leu Ile Asp Leu Arg Gly Thr Ser Phe Glu Val Val</td>
<td>1115</td>
<td>1120</td>
<td>1125</td>
<td></td>
</tr>
<tr>
<td>Ala Gln His Arg Ala Trp Ala Glu Val Gly Ile His Gly Thr</td>
<td>1130</td>
<td>1135</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>Leu Leu Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Gly Gly</td>
<td>1145</td>
<td>1150</td>
<td>1155</td>
<td></td>
</tr>
<tr>
<td>Ala Asn Ala His Ala Ile Val Glu Glu His Val Ile Ala Thr Pro</td>
<td>1160</td>
<td>1165</td>
<td>1170</td>
<td></td>
</tr>
<tr>
<td>Pro Ser Thr Ser Ser Ala Gly Gly Pro Val Gly Ile Val Leu Ser</td>
<td>1175</td>
<td>1180</td>
<td>1185</td>
<td></td>
</tr>
<tr>
<td>Ala Gly Ser Glu Ala Val Leu Arg Gln Glu Val Leu Ala Leu Ser</td>
<td>1190</td>
<td>1195</td>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>Ala Trp Leu Arg Gln Ser Pro Thr Pro Ala Glu Met Ile Asp</td>
<td>1205</td>
<td>1210</td>
<td>1215</td>
<td></td>
</tr>
<tr>
<td>Val Ala Tyr Thr Leu Glu Val Gly Arg Ala Ala Leu Ser His Arg</td>
<td>1220</td>
<td>1225</td>
<td>1230</td>
<td></td>
</tr>
<tr>
<td>Leu Ala Phe Ser Ala Thr Asp Ala Glu Gln Ala Leu Ala Arg Leu</td>
<td>1235</td>
<td>1240</td>
<td>1245</td>
<td></td>
</tr>
<tr>
<td>Glu Gly Arg Leu Ala Gly Val Met Asp Ala Glu Val His His Gly</td>
<td>1250</td>
<td>1255</td>
<td>1260</td>
<td></td>
</tr>
</tbody>
</table>
Val Val Asp Ala Ala Ala Thr Ala Pro Glu His Gly Arg Gln Thr 1265 1270 1276
Arg Gly Leu Ala Gly Leu Leu Arg Ala Trp Thr Gln Gly Val 1280 1285 1290
Arg Val Asp Trp Ser Ala Leu Tyr Gly Ile Gln Arg Pro Gln Arg 1295 1300 1305
Val Ser Leu Pro Val Tyr Pro Phe Ala Arg Glu Arg Tyr Trp Leu 1310 1315 1320
Pro Gly Gln Ala Met His Ala Ala Ala Asp Ala His Pro Met Leu 1325 1330 1335
Gln Leu Leu His Ala Asn Ala Lys Leu His Arg Tyr Ala Leu Arg 1340 1345 1350
Arg Ser Gly Cys Ser Ser Phe Leu Val Asp His Cys Val Asp Gly 1355 1360 1365
Arg Gly Val Leu Pro Ala Ala Val Gln Leu Glu Leu Val Arg Ala 1370 1375 1380
Val Ala Gln Arg Val Met Ala Gln Asp Gly Cys Ile Glu Leu 1385 1390 1395
 Ala Gln Val Ala Phe Leu His Pro Leu Met Met Glu Glu Thr Glu 1400 1405 1410
Leu Glu Val Glu Ile Glu Leu Ser Lys Ser Asp Gln Asp Glu Phe 1415 1420 1425
Asp Phe Gln Leu His Asp Ala His Arg Gln Glu Val Phe Ser Gln 1430 1435 1440
Gly His Val Arg Arg Arg Val Tyr Thr Ala Thr Pro Arg Leu Asp 1445 1450 1455
Leu Ala Gln Leu Gln Lys Leu Cys Ala Glu Arg Val Leu Ser Gly 1460 1465 1470
Glu Asp Cys Tyr Ala His Phe Thr Ala Cys Gly Leu Gln Leu Gly 1475 1480 1485
Asp Arg Leu Lys Ser Val Gln Ser Ile Gly Cys Gly Arg Asn Gly 1490 1495 1500
Glu Gly Glu Pro Ile Ala Leu Gly Val Leu Arg Leu Pro Pro Ser 1505 1510 1515
Ser Val Glu Asp Ser His Val Leu Pro Pro Ser Leu Leu Asp Gly 1520 1525 1530
A Asp Ser Ala Gly Ser Leu Gly Leu Gln Arg Asp Val Glu His Ile 1535 1540 1545
A Ala Met Pro Tyr Thr Leu Glu Arg Met Thr Val His Ala Pro Ile 1550 1555 1560
Pro Pro Glu Ala Trp Val Leu Leu Arg His Gly His Ala Ala Arg 1565 1570 1575
Gln Ser Leu Asp Ile Asp Leu Leu Asp Ser Glu Gly Arg Val Cys 1580 1585 1590
Val Ser Leu Gly Asn Tyr Thr Gly Arg Ala Pro Lys Ala Val Ser 1595 1600 1605
A Ala Val Arg Ala Leu Val Leu Ala Pro Val Thr Gln Ala Leu Thr 1610 1615 1620
Glu Thr Ala Pro Ala Trp Pro Asp Ala Glu Arg Ile Val Thr 1625 1630 1635
Val Gly Asp Ala Trp Arg Ser His Phe Gly Phe Asp Glu Pro 1640 1645 1650
A Ala Leu Ser Leu Glu Asp Ser Val Glu Val Ile Ala Thr Arg Leu
<table>
<thead>
<tr>
<th>1655</th>
<th>1660</th>
<th>1665</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>Gln</td>
<td>Ser</td>
</tr>
<tr>
<td>1670</td>
<td>1675</td>
<td>1680</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Ser</td>
</tr>
<tr>
<td>1685</td>
<td>1690</td>
<td>1695</td>
</tr>
<tr>
<td>Ala</td>
<td>Gly</td>
<td>Phe</td>
</tr>
<tr>
<td>1700</td>
<td>1705</td>
<td>1710</td>
</tr>
<tr>
<td>His</td>
<td>Arg</td>
<td>Pro</td>
</tr>
<tr>
<td>1715</td>
<td>1720</td>
<td>1725</td>
</tr>
<tr>
<td>Arg</td>
<td>Gin</td>
<td>Pro</td>
</tr>
<tr>
<td>1730</td>
<td>1735</td>
<td>1740</td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>1745</td>
<td>1750</td>
<td>1755</td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Lys</td>
</tr>
<tr>
<td>1760</td>
<td>1765</td>
<td>1770</td>
</tr>
<tr>
<td>Trp</td>
<td>His</td>
<td>Ala</td>
</tr>
<tr>
<td>1775</td>
<td>1780</td>
<td>1785</td>
</tr>
<tr>
<td>Tyr</td>
<td>Arg</td>
<td>Gin</td>
</tr>
<tr>
<td>1790</td>
<td>1795</td>
<td>1800</td>
</tr>
<tr>
<td>Pro</td>
<td>Pro</td>
<td>Pro</td>
</tr>
<tr>
<td>1805</td>
<td>1810</td>
<td>1815</td>
</tr>
<tr>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
</tr>
<tr>
<td>1820</td>
<td>1825</td>
<td>1830</td>
</tr>
<tr>
<td>Tyr</td>
<td>Arg</td>
<td>Ala</td>
</tr>
<tr>
<td>1835</td>
<td>1840</td>
<td>1845</td>
</tr>
<tr>
<td>Gly</td>
<td>Ile</td>
<td>Ala</td>
</tr>
<tr>
<td>1850</td>
<td>1855</td>
<td>1860</td>
</tr>
<tr>
<td>Pro</td>
<td>His</td>
<td>Tyr</td>
</tr>
<tr>
<td>1865</td>
<td>1870</td>
<td>1875</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>His</td>
</tr>
<tr>
<td>1880</td>
<td>1885</td>
<td>1890</td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>1895</td>
<td>1900</td>
<td>1905</td>
</tr>
<tr>
<td>Met</td>
<td>Glu</td>
<td>Ala</td>
</tr>
<tr>
<td>1910</td>
<td>1915</td>
<td>1920</td>
</tr>
<tr>
<td>Val</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>1925</td>
<td>1930</td>
<td>1935</td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>1940</td>
<td>1945</td>
<td>1950</td>
</tr>
<tr>
<td>Gly</td>
<td>Gln</td>
<td>Gly</td>
</tr>
<tr>
<td>1955</td>
<td>1960</td>
<td>1965</td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>Trp</td>
<td>Gly</td>
<td>Tyr</td>
</tr>
<tr>
<td>1985</td>
<td>1990</td>
<td>1995</td>
</tr>
<tr>
<td>Arg</td>
<td>Arg</td>
<td>Arg</td>
</tr>
<tr>
<td>1995</td>
<td>2000</td>
<td>2005</td>
</tr>
<tr>
<td>Glu</td>
<td>Ala</td>
<td>Met</td>
</tr>
<tr>
<td>2005</td>
<td>2010</td>
<td>2015</td>
</tr>
<tr>
<td>Arg</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td>2030</td>
<td>2035</td>
<td>2040</td>
</tr>
<tr>
<td>Leu</td>
<td>Gly</td>
<td>Val</td>
</tr>
<tr>
<td>2045</td>
<td>2050</td>
<td>2055</td>
</tr>
</tbody>
</table>
Pro Ala Glu Pro Pro Arg Gly Gln Pro Gly Pro Val Val Glu Leu
   2060  2065  2070
Ser Val Asn Leu Asp Ala Arg Arg Glu Arg Glu Thr Leu Leu Ala
   2075  2080  2085
Ala Trp Leu Leu Leu Leu Ile Glu Gln Leu Gly Gly Phe Pro Pro
   2090  2095  2100
Ala Ser Phe Asp Ile Ala Thr Leu Ala Gln Arg Leu His Ile Val
   2105  2110  2115
Pro Ala Tyr Arg Ser Trp Leu Glu His Ser Val Arg Met Leu Gly
   2120  2125  2130
Val Tyr Gly Tyr Leu Arg Ala Thr Gly Glu Ser Arg Phe Glu Leu
   2135  2140  2145
Ala Asp Lys Pro Pro Asp Ala Arg Gly Ala Trp Asn Ala His
   2150  2155  2160
Val His Glu Ala Ser Val Glu Ala Gly Gln Glu Arg Ala Gln Arg Arg
   2165  2170  2175
Leu Leu Asp Arg Cys Met Arg Ala Leu Pro Ala Val Leu Arg Gly
   2180  2185  2190
Glu Arg Lys Ala Thr Glu Leu Leu Phe Pro Glu Gly Ser Met Ala
   2195  2200  2205
Trp Val Glu Gly Ile Tyr Gln Asp Asn Pro Leu Ala Asp Tyr Phe
   2210  2215  2220
Asn Ala Gln Leu Val Thr Arg Leu Ile Ala Tyr Leu Arg Arg Arg
   2225  2230  2235
Leu Glu Ser Thr Pro Thr Ala Arg Leu Lys Leu Cys Glu Ile Gly
   2240  2245  2250
Ala Gly Ser Gly Gly Thr Ala Ser Val Leu Gln Gln Leu Gln
   2255  2260  2265
Ala Tyr Gly Glu His Ile Glu Gly Tyr Leu Tyr Thr Asp Leu Ser
   2270  2275  2280
Pro Val Phe Leu His His Ala Glu Lys His Tyr Gln Pro Arg Ala
   2285  2290  2295
Pro Tyr Leu Arg Thr Ala Cys Phe Asp Ala Ala Arg Ala Pro Thr
   2300  2305  2310
Ala Gln Ala Leu Glu Ser Gly Gly Tyr Asp Val Ile Ala Ala
   2315  2320  2325
Asn Val Leu His Ala Thr Arg Asp Ile Ala Lys Thr Leu Arg Asn
   2330  2335  2340
Ala Lys Ala Leu Leu Lys Pro Gly Gly Leu Leu Leu Leu Asn Glu
   2345  2350  2355
Val Ile Glu Arg Ser Leu Val Leu His Leu Thr Phe Gly Leu Leu
   2360  2365  2370
Glu Ser Trp Trp Leu Pro Gln Asp Lys Ile Leu Arg Leu Ala Gly
   2375  2380  2385
Ser Pro Leu Leu Ala Cys Ala Thr Trp Arg Ser Leu Leu Glu Ala
   2390  2395  2400
Glu Gly Phe Ala Gly Leu Ser Val His Arg Ala Gln Pro Asp Ala
   2405  2410  2415
Gly Gln Ala Ile Ile Cys Ala Tyr Ser Asp Gly Ile Val Arg Gln
   2420  2425  2430
Ala Ser Thr Ile Glu Val Ala Arg Asn Glu Lys Val Thr Val Pro
   2435  2440  2445
Ser Gln Pro Ala Glu Ala Gly Glu Ser Pro Leu Asp Leu Val Lys
2450
2455
2460
Lys Leu Leu Gly Arg Ile Leu Lys Met Asp Pro Ala Thr Leu Asp
2465
2470
2475
Thr Ser His Pro Leu Glu Tyr Tyr Gly Val Asp Ser Ile Val Ala
2480
2485
2490
Ile Glu Leu Ala Met Ala Leu Arg Glu Thr Phe Pro Gly Phe Glu
2495
2500
2505
Val Ser Glu Leu Phe Glu Thr Gln Ser Ile Asp Thr Leu Leu Gly
2510
2515
2520
Ser Leu Glu Gln Ala Pro Leu Leu Ala Thr Leu Thr Ala Pro Pro
2525
2530
2535
Gln Gln Asp Met Leu Gln Gln Leu Lys Gln Leu Leu Ala Arg Thr
2540
2545
2550
Leu Lys Leu Asp Ile Thr Gln Ile Asp Thr Ser Lys Thr Leu Glu
2555
2560
2565
Ser Tyr Gly Val Asp Ser Ile Val Ile Ile Glu Leu Ala Asn Ala
2570
2575
2580
Leu Arg Glu Arg Tyr Pro Ser Leu Asp Ala Ser Gln Leu Met Glu
2585
2590
2595
Thr Leu Ser Ile Asp Arg Leu Val Ala Gln Trp Gln Ala Thr Glu
2600
2605
2610
Pro Ala Val Pro Ala Glu Pro Thr Ala Glu Pro Pro Val Ala Asp
2615
2620
2625
Glu Asp Ala Ala Ala Ile Glu Gly Leu Ala Gly Arg Phe Pro Gly
2630
2635
2640
Ala Asp Thr Leu Glu Gln Phe Trp Asn Asn Leu Arg Asn Gly Gln
2645
2650
2655
Ser Ser Met Gly Glu Val Pro Gly Glu Arg Trp Asp His Gln His
2660
2665
2670
Tyr Phe Asp Ser Glu Arg Gln Ala Pro Gly Lys Thr Tyr Ser Arg
2675
2680
2685
Trp Gly Ala Phe Leu Arg Asp Ile Asp Gly Phe Asp Ala Ala Phe
2690
2695
2700
Phe Glu Trp Pro Asp Ser Val Ala Leu Glu Ser Asp Pro Gln Ala
2705
2710
2715
Arg Ile Phe Leu Glu Gln Ala Tyr Ala Gly Ile Glu Asp Ala Gly
2720
2725
2730
Tyr Thr Pro Gly Ser Leu Ser Lys Ser Gln Arg Val Gly Val Phe
2735
2740
2745
Val Gly Val Met Asn Gly Tyr Ser Gly Gly Ala Arg Phe Trp
2750
2755
2760
Gln Ile Ala Asn Arg Val Ser Tyr Gln Phe Asp Phe Arg Gly Pro
2765
2770
2775
Ser Leu Ala Val Asp Thr Ala Cys Ser Ala Ser Leu Thr Ala Ile
2780
2785
2790
His Leu Ala Leu Glu Ser Leu Arg Ser Gly Ser Cys Glu Val Ala
2795
2800
2805
Leu Ala Gly Gly Val Asn Leu Leu Val Asp Pro Gln Glu Tyr Leu
2810
2815
2820
Asn Leu Ala Gly Ala Ala Met Leu Ser Ala Gly Ala Ser Cys Arg
2825
2830
2835
Pro Phe Gly Glu Ala Ala Asp Gly Phe Val Ala Gly Glu Ala Cys
<table>
<thead>
<tr>
<th>2840</th>
<th>2845</th>
<th>2850</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly Val Val Leu Leu Lys Pro Leu Lys Gln Ala Arg Ala Asp Gly</td>
<td>2855</td>
<td>2860</td>
</tr>
<tr>
<td>Asp Val Ile His Ala Val Ile Arg Gly Ser Met Ile Asn Ala Gly</td>
<td>2870</td>
<td>2875</td>
</tr>
<tr>
<td>Gly His Thr Ser Ala Phe Ser Ser Pro Asn Pro Ala Ala Gln Ala</td>
<td>2890</td>
<td>2895</td>
</tr>
<tr>
<td>Ala Val Val Arg Gln Ala Leu Gln Arg Ala Gly Val Ala Pro Asp</td>
<td>2900</td>
<td>2905</td>
</tr>
<tr>
<td>Ser Ile Ser Tyr Ile Glu Ala His Gly Thr Gly Thr Val Leu Gly</td>
<td>2915</td>
<td>2920</td>
</tr>
<tr>
<td>Asp Ala Val Glu Leu Gly Ala Leu Asn Lys Val Phe Asp Lys Arg</td>
<td>2930</td>
<td>2935</td>
</tr>
<tr>
<td>Ala Ala Pro Cys Pro Ile Gly Ser Leu Lys Ala Asn Ile Gly His</td>
<td>2945</td>
<td>2950</td>
</tr>
<tr>
<td>Ala Gln Ser Ala Ala Gly Ile Ala Gly Leu Ala Lys Leu Val Leu</td>
<td>2960</td>
<td>2965</td>
</tr>
<tr>
<td>Gln Phe Arg His Gly Glu Leu Val Pro Ser Leu Asn Ala Phe Pro</td>
<td>2975</td>
<td>2980</td>
</tr>
<tr>
<td>Leu Asn Pro Tyr Ile Glu Phe Gly Arg Phe Gln Val Gln Gln Gln</td>
<td>2990</td>
<td>2995</td>
</tr>
<tr>
<td>Pro Ala Pro Trp Pro Arg Arg Gly Ala Gln Pro Arg Arg Ala Gly</td>
<td>3005</td>
<td>3010</td>
</tr>
<tr>
<td>Leu Ser Ala Phe Gly Ala Gly Gly Ser Asn Ala His Leu Val Val</td>
<td>3020</td>
<td>3025</td>
</tr>
<tr>
<td>Glu Glu Ala Pro Ala Met Ala Pro Gly Val Ser Ile Ser Ala Ser</td>
<td>3035</td>
<td>3040</td>
</tr>
<tr>
<td>Ser Pro Ala Leu Ile Val Leu Ser Ala Arg Thr Leu Pro Ala Leu</td>
<td>3050</td>
<td>3055</td>
</tr>
<tr>
<td>Gln Gln Arg Ala Arg Asp Leu Leu Val Trp Met Gln Ala Arg Gln</td>
<td>3065</td>
<td>3070</td>
</tr>
<tr>
<td>Val Asp Asp Val Met Leu Ala Asp Ala Tyr Thr Leu His Leu</td>
<td>3080</td>
<td>3085</td>
</tr>
<tr>
<td>Gly Arg Val Ala Met Glu Gln Arg Leu Ala Phe Thr Ala Gly Ser</td>
<td>3095</td>
<td>3100</td>
</tr>
<tr>
<td>Ala Ala Glu Leu Ser Glu Lys Leu Gln Ala Tyr Leu Gly His Ala</td>
<td>3110</td>
<td>3115</td>
</tr>
<tr>
<td>Ile Arg Ala Asp Ile Tyr Leu Ser Glu Asp Thr Pro Gly Lys Pro</td>
<td>3125</td>
<td>3130</td>
</tr>
<tr>
<td>Ala Gly Ala Pro Ile Val Ala Glu Glu Asp Leu Leu Thr Leu Met</td>
<td>3140</td>
<td>3145</td>
</tr>
<tr>
<td>Asp Ala Trp Ile Glu Lys Gly Gln Tyr Gly Arg Leu Leu Glu Tyr</td>
<td>3155</td>
<td>3160</td>
</tr>
<tr>
<td>Trp Thr Lys Gly Gln Pro Ile Asp Trp Asn Leu Tyr Trp Arg</td>
<td>3170</td>
<td>3175</td>
</tr>
<tr>
<td>Lys Leu Tyr Ala Asp Gly Arg Pro Arg Arg Ile Ser Leu Pro Thr</td>
<td>3185</td>
<td>3190</td>
</tr>
<tr>
<td>Tyr Pro Phe Glu His Arg Arg Tyr Trp Gln Thr Pro Val Pro Gly</td>
<td>3200</td>
<td>3205</td>
</tr>
<tr>
<td>Glu Arg Ser Leu His Ala Thr Ala Pro Ala Thr Arg Glu Thr Val</td>
<td>3215</td>
<td>3220</td>
</tr>
<tr>
<td>Ala Val Gly Ala Met Pro Asp Pro Ala Gly Ala Thr Val Gln Ala</td>
<td>3230</td>
<td>3235</td>
</tr>
<tr>
<td>Residue</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Arg Leu Cys Ala Leu Cys Gln Val Leu Leu Gly Lys</td>
<td>Pro Val Thr 3245 3250 3255</td>
<td></td>
</tr>
<tr>
<td>Ala Gln Met Asp Phe Phe Ala Val Gly Gly His Ser Val Leu Ala</td>
<td>3260 3265 3270</td>
<td></td>
</tr>
<tr>
<td>Ile Gln Leu Val Ser Arg Ile Arg Lys Ser Phe Gly Val Glu Tyr</td>
<td>3275 3280 3285</td>
<td></td>
</tr>
<tr>
<td>Pro Val Ser Ala Leu Phe Glu Ser Ala Leu Leu Ser Asp Met Ala</td>
<td>3290 3295 3300</td>
<td></td>
</tr>
<tr>
<td>Arg Gln Ile Glu Gln Leu Arg Val Asn Gly Val Ala Lys Arg Met</td>
<td>3305 3310 3315</td>
<td></td>
</tr>
<tr>
<td>Pro Ala Leu Leu Pro Ala Gly Arg Val Gly Ala Ile Pro Ala Thr</td>
<td>3320 3325 3330</td>
<td></td>
</tr>
<tr>
<td>Tyr Ala Gln Glu Arg Leu Trp Leu Val His Glu His Met Ser Glu</td>
<td>3335 3340 3345</td>
<td></td>
</tr>
<tr>
<td>Gln Arg Ser Ser Tyr Asn Ile Thr Phe Ala Met His Phe Arg Gly</td>
<td>3350 3355 3360</td>
<td></td>
</tr>
<tr>
<td>Val Asp Phe Arg Ala Glu Ala Met Arg Ala Ala Leu Asn Ala Leu</td>
<td>3365 3370 3375</td>
<td></td>
</tr>
<tr>
<td>Val Val Arg His Glu Val Leu Arg Thr Arg Phe Leu Ser Glu Asp</td>
<td>3380 3385 3390</td>
<td></td>
</tr>
<tr>
<td>Gly Gln Leu Gln Gln Val Ile Ala Ala Ser Leu Thr Leu Glu Val</td>
<td>3395 3400 3405</td>
<td></td>
</tr>
<tr>
<td>Pro Val Arg Glu Met Ser Val Glu Glu Val Asp Leu Leu Leu Ala</td>
<td>3410 3415 3420</td>
<td></td>
</tr>
<tr>
<td>Ala Ser Thr Arg Glu Thr Phe Asp Leu Arg Gln Gly Pro Leu Phe</td>
<td>3425 3430 3435</td>
<td></td>
</tr>
<tr>
<td>Lys Ala Arg Ile Leu Arg Val Ala Ala Asp His His Val Val Leu</td>
<td>3440 3445 3450</td>
<td></td>
</tr>
<tr>
<td>Ser Ser Ile His His Ile Ile Ser Asp Gly Trp Ser Leu Gly Val</td>
<td>3455 3460 3465</td>
<td></td>
</tr>
<tr>
<td>Phe Asn Arg Asp Leu His Gln Leu Tyr Glu Ala Cys Leu Arg Gly</td>
<td>3470 3475 3480</td>
<td></td>
</tr>
<tr>
<td>Thr Pro Pro Thr Leu Pro Thr Leu Ala Val Gln Tyr Ala Asp Tyr</td>
<td>3485 3490 3495</td>
<td></td>
</tr>
<tr>
<td>Ala Leu Trp Gln Arg Gln Trp Glu Leu Ala Ala Pro Leu Ser Tyr</td>
<td>3500 3505 3510</td>
<td></td>
</tr>
<tr>
<td>Trp Thr Arg Ala Leu Gln Gly Tyr Asp Asp Gly Leu Asp Leu Pro</td>
<td>3515 3520 3525</td>
<td></td>
</tr>
<tr>
<td>Tyr Asp Arg Pro Arg Gly Ala Thr Arg Ala Trp Arg Ala Gly Leu</td>
<td>3530 3535 3540</td>
<td></td>
</tr>
<tr>
<td>Val Lys His Arg Tyr Pro Pro Gln Leu Ala Gln Gln Leu Ala Ala</td>
<td>3545 3550 3555</td>
<td></td>
</tr>
<tr>
<td>Tyr Ser Gln Gln Tyr Gln Ala Thr Leu Phe Met Ser Leu Leu Ala</td>
<td>3560 3565 3570</td>
<td></td>
</tr>
<tr>
<td>Gly Leu Ala Leu Val Leu Gly Arg Tyr Ala Asp Arg Lys Asp Val</td>
<td>3575 3580 3585</td>
<td></td>
</tr>
<tr>
<td>Cys Ile Gly Ala Thr Val Ser Gly Arg Asp Gln Leu Glu Leu Glu</td>
<td>3590 3595 3600</td>
<td></td>
</tr>
<tr>
<td>Glu Leu Ile Gly Phe Phe Ile Asn Ile Leu Pro Leu Arg Val Asp</td>
<td>3605 3610 3615</td>
<td></td>
</tr>
<tr>
<td>Leu Ser Gly Asp Pro Cys Leu Glu Glu Val Leu Leu Arg Thr Arg</td>
<td>3620 3625 3630</td>
<td></td>
</tr>
<tr>
<td>Amino Acids</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>Gln Val Val Leu Asp Gly Phe Ala His Gln Ser Val Pro Phe Glu</td>
<td>3635</td>
<td>3645</td>
</tr>
<tr>
<td>His Val Leu Gln Ala Leu Arg Arg Gln Arg Asp Ser Ser Gln Ile</td>
<td>3650</td>
<td>3660</td>
</tr>
<tr>
<td>Pro Leu Val Pro Val Met Leu Arg His Gln Asn Phe Pro Thr Gln</td>
<td>3665</td>
<td>3670</td>
</tr>
<tr>
<td>Glu Ile Gly Asp Trp Pro Glu Gly Val Arg Leu Thr Gln Met Glu</td>
<td>3680</td>
<td>3690</td>
</tr>
<tr>
<td>Leu Gly Leu Asp Arg Ser Thr Pro Ser Glu Leu Asp Trp Gln Phe</td>
<td>3695</td>
<td>3700</td>
</tr>
<tr>
<td>Tyr Gly Asp Gly Ser Ser Leu Glu Leu Thr Leu Glu Tyr Ala Gln</td>
<td>3710</td>
<td>3715</td>
</tr>
<tr>
<td>Asp Leu Phe Asp Glu Ala Thr Val Arg Arg Met Ile Ala His His</td>
<td>3720</td>
<td>3725</td>
</tr>
<tr>
<td>Gln Gln Ala Leu Glu Ala Met Val Ser Arg Pro Gln Leu Arg Val</td>
<td>3740</td>
<td>3745</td>
</tr>
<tr>
<td>Gly Lys Trp Asp Met Leu Thr Ala Glu Glu Arg Arg Leu Phe Ala</td>
<td>3750</td>
<td>3755</td>
</tr>
<tr>
<td>Ala Leu Asp Ala Thr Gly Thr Pro Arg Glu Trp Pro Ser Leu Ala</td>
<td>3770</td>
<td>3775</td>
</tr>
<tr>
<td>Gln Gln Phe Glu Arg Gln Ala Gln Ala Thr Pro Gln Ala Ile Ile</td>
<td>3780</td>
<td>3785</td>
</tr>
<tr>
<td>Cys Val Ser Asp Gly Gln Ser Trp Ser Tyr Ala Gln Leu Glu Ala</td>
<td>3790</td>
<td>3795</td>
</tr>
<tr>
<td>Arg Ala Asp Gln Leu Ala Gln Ala Leu Arg Gln Gln Gly Ala Gly</td>
<td>3800</td>
<td>3805</td>
</tr>
<tr>
<td>Arg Asp Val Arg Val Ala Val Gln Ser Ala Arg Thr Pro Glu Leu</td>
<td>3810</td>
<td>3815</td>
</tr>
<tr>
<td>Leu Met Ala Leu Leu Ala Ile Phe Lys Ala Gly Ala Cys Tyr Val</td>
<td>3820</td>
<td>3825</td>
</tr>
<tr>
<td>Pro Ile Asp Pro Ala Tyr Pro Ala Ala Tyr Arg Glu Gln Ile Leu</td>
<td>3830</td>
<td>3835</td>
</tr>
<tr>
<td>Ala Glu Val Gln Val Ser Ile Val Leu Glu Gln Asp Glu Leu Ala</td>
<td>3840</td>
<td>3845</td>
</tr>
<tr>
<td>Leu Asp Glu Gln Gly Gln Phe His Asn Pro Arg Trp Arg Glu Gln</td>
<td>3850</td>
<td>3855</td>
</tr>
<tr>
<td>Ala Pro Thr Pro Leu Gly Leu Arg Glu His Pro Gly Asp Leu Ala</td>
<td>3860</td>
<td>3865</td>
</tr>
<tr>
<td>Cys Val Met Val Thr Ser Gly Ser Thr Gly Arg Pro Lys Gly Val</td>
<td>3870</td>
<td>3875</td>
</tr>
<tr>
<td>Met Val Pro Tyr Ala Gln Leu Tyr Asp Trp Leu His Ala Gly Trp</td>
<td>3880</td>
<td>3885</td>
</tr>
<tr>
<td>Gln Arg Ser Pro Phe Glu Ala Gly Glu Arg Val Leu Gln Lys Thr</td>
<td>3890</td>
<td>3895</td>
</tr>
<tr>
<td>Ser Ile Ala Phe Ala Val Ser Val Lys Glu Leu Leu Ser Gly Leu</td>
<td>3900</td>
<td>3905</td>
</tr>
<tr>
<td>Leu Ala Gly Val Glu Gln Val Met Leu Pro Asp Gln Glu Val Lys</td>
<td>3910</td>
<td>3915</td>
</tr>
<tr>
<td>Asp Ser Leu Ala Leu Ala Arg Ala Ile Glu Gln Trp Gln Val Thr</td>
<td>3920</td>
<td>3925</td>
</tr>
<tr>
<td>Arg Leu Tyr Leu Val Pro Ser His Leu Gln Ala Leu Leu Asp Ala</td>
<td>3930</td>
<td>3935</td>
</tr>
<tr>
<td>Thr Gln Gly Arg Asp Gly Leu Leu His Ser Leu Arg His Val Val</td>
<td>3940</td>
<td>3945</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Thr</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>4025</td>
<td>4030</td>
<td>4035</td>
</tr>
<tr>
<td>Ala</td>
<td>Arg</td>
<td>Leu</td>
</tr>
<tr>
<td>4040</td>
<td>4045</td>
<td>4050</td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Asn</td>
</tr>
<tr>
<td>4055</td>
<td>4060</td>
<td>4065</td>
</tr>
<tr>
<td>Gly</td>
<td>Thr</td>
<td>Phe</td>
</tr>
<tr>
<td>4070</td>
<td>4075</td>
<td>4080</td>
</tr>
<tr>
<td>Tyr</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>4080</td>
<td>4085</td>
<td>4090</td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>4095</td>
<td>4100</td>
<td>4105</td>
</tr>
<tr>
<td>Asn</td>
<td>Arg</td>
<td>Pro</td>
</tr>
<tr>
<td>4110</td>
<td>4115</td>
<td>4120</td>
</tr>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Pro</td>
</tr>
<tr>
<td>4125</td>
<td>4130</td>
<td>4135</td>
</tr>
<tr>
<td>Arg</td>
<td>Arg</td>
<td>Leu</td>
</tr>
<tr>
<td>4140</td>
<td>4145</td>
<td>4150</td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td>4155</td>
<td>4160</td>
<td>4165</td>
</tr>
<tr>
<td>Glu</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>4170</td>
<td>4175</td>
<td>4180</td>
</tr>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>His</td>
</tr>
<tr>
<td>4185</td>
<td>4190</td>
<td>4195</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>4200</td>
<td>4205</td>
<td>4210</td>
</tr>
<tr>
<td>Gin</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td>4215</td>
<td>4220</td>
<td>4225</td>
</tr>
<tr>
<td>Gin</td>
<td>Trp</td>
<td>Leu</td>
</tr>
<tr>
<td>4230</td>
<td>4235</td>
<td>4240</td>
</tr>
<tr>
<td>Arg</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td>4240</td>
<td>4245</td>
<td>4250</td>
</tr>
<tr>
<td>Val</td>
<td>Ala</td>
<td>Pro</td>
</tr>
<tr>
<td>4255</td>
<td>4260</td>
<td>4265</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td>4270</td>
<td>4275</td>
<td>4280</td>
</tr>
<tr>
<td>Phe</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>4285</td>
<td>4290</td>
<td>4295</td>
</tr>
<tr>
<td>Arg</td>
<td>Ile</td>
<td>Ala</td>
</tr>
<tr>
<td>4300</td>
<td>4305</td>
<td>4310</td>
</tr>
<tr>
<td>Phe</td>
<td>Glu</td>
<td>Pro</td>
</tr>
<tr>
<td>4315</td>
<td>4320</td>
<td>4325</td>
</tr>
<tr>
<td>His</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>4330</td>
<td>4335</td>
<td>4340</td>
</tr>
<tr>
<td>Glu</td>
<td>Ala</td>
<td>Thr</td>
</tr>
<tr>
<td>4345</td>
<td>4350</td>
<td>4355</td>
</tr>
<tr>
<td>Leu</td>
<td>His</td>
<td>Lys</td>
</tr>
<tr>
<td>4360</td>
<td>4365</td>
<td>4370</td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>4375</td>
<td>4380</td>
<td>4385</td>
</tr>
<tr>
<td>Phe</td>
<td>Arg</td>
<td>Cys</td>
</tr>
<tr>
<td>4390</td>
<td>4395</td>
<td>4400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Val Glu Thr Glu Gln Gln Cys Val Gln Val Ile Gly Ala Ala Glu
4430 4435 4440
Gln Phe Val Leu Gln Leu Arg Ser Ile Arg Asp Glu Ala Asp Leu
4445 4450 4455
His Gly Leu Leu His Thr Ala Val Ser Glu Pro Phe Asp Leu Glu
4460 4465 4470
Arg Glu Leu Pro Leu Arg Ala Leu Leu Tyr Arg Leu Asp Asp Arg
4475 4480 4485
Arg His Tyr Leu Ala Val Val Ile His His Ile Val Phe Asp Gly
4490 4495 4500
Trp Ser Thr Ser Ile Leu Phe Arg Glu Leu Ala Thr His Tyr Ala
4505 4510 4515
Ala Cys Arg His Gly Gln Ser Ala Pro Leu Pro Pro Leu Glu Leu
4520 4525 4530
Ser Tyr Ala Asp Tyr Ala Arg Trp Glu Arg Ala Arg Leu Asn Gln
4535 4540 4545
Glu Asp Ala Leu Arg Lys Leu Glu Tyr Trp Lys Thr Gln Leu Ala
4550 4555 4560
Asp Ala Pro Pro Leu Val Leu Pro Thr Thr Tyr Ala Arg Pro Val
4565 4570 4575
Phe Gln Asn Phe Asn Gly Ala Thr Val Ala Leu Gln Ile Glu Pro
4580 4585 4590
Pro Leu Leu Gln Arg Leu Gln Arg Phe Ala Asp Ala His Ser Phe
4595 4600 4605
Thr Leu Tyr Met Leu Leu Ala Ala Leu Gly Val Val Leu Ser
4610 4615 4620
Arg His Ala Arg Glu Lys His Phe Cys Ile Gly Ser Pro Val Ala
4625 4630 4635
Asn Arg Ala Arg Ala Glu Leu His Gly Leu Ile Gly Leu Phe Val
4640 4645 4650
Asn Thr Leu Ala Val Arg Leu Asp Leu Asp Gly Asn Pro Ser Val
4650 4660 4665
Arg Glu Leu Leu Glu Arg Ile His Cys Thr Thr Leu Ala Ala Tyr
4670 4675 4680
Glu His Gln Asp Val Pro Phe Glu Arg Ile Val Glu Ser Leu Lys
4685 4690 4695
Val Pro Arg Asp Thr Ala Arg Asn Pro Leu Gly Gln Val Met Leu
4700 4705 4710
Asn Phe Gln Asn Met Pro Met Ser Ala Phe Asp Leu Asp Gly Val
4715 4720 4725
Gln Val Gln Val Leu Pro Met His Asn Gly Thr Ala Lys Cys Glu
4730 4735 4740
Leu Thr Phe Asp Leu Leu Leu Asp Gly Ser Arg Leu Ser Gly Phe
4745 4750 4755
Val Glu Tyr Ala Thr Gly Leu Phe Ala Pro Glu Trp Val Glu Ala
4760 4765 4770
Leu Val Gln Gln Phe Lys Cys Val Leu Ala Ala Leu Val Glu Arg
4775 4780 4785
Pro Glu Ala Ser Leu Asn Arg Leu Pro Met Ala Pro Asn Glu Ala
4790 4795 4800
Gln Pro Ala Ser Pro Ala Leu Met Lys His Val Ala Pro Ser Leu
4805 4810 4815
-continued

Pro Asn Leu Leu Glu Ala Met Ala Ala Asn Asp Ala Ala Arg Leu 4820 4825 4830
Ala Leu Gln Ala Pro Glu Gly Ala Leu Ser Tyr Ala Gln Leu Ile 4835 4840 4846
Glu Ala Ala Asn Glu Phe Ala Trp Arg Leu Arg Cys Glu His Ala 4850 4855 4860
Gly Pro Asp Lys Val Val Ala Leu Cys Leu Ala Pro Cys Ser Ala 4865 4870 4875
Leu Val Val Ala Leu Leu Ala Ala Ser Leu Cys Gly Ala Ala Ser 4880 4885 4890
Val Leu Ile Asp Pro Thr Thr Thr Ala Glu Ala Gln Tyr Asp Gln 4895 4900 4905
Leu Phe Glu Thr Arg Ala Gly Ile Val Val Thr Cys Ser Ser Leu 4910 4915 4920
Leu Glu Lys Leu Pro Leu Asp Asp Gln Ala Val Val Leu Ile Asp 4925 4930 4935
Glu Gln Ala Ala Glu Ala Thr Pro Arg Leu Met His Phe Thr Asp 4940 4945 4950
Asp Pro Ala Leu Pro Ala Pro Leu Met Leu Tyr Cys Val Cys Asp Glu Lys 4955 4960 4965
Gly Arg Thr Arg Thr Ile Met Val Glu Ser Gln Ser Leu Ser Ser 4970 4975 4980
Arg Leu Leu Asp Ser Val Gln Arg Phe Ser Ser Leu Gln Arg Thr Asp 4985 4990 4995
Arg Phe Leu Leu Arg Ser Pro Leu Ser Ala Glu Leu Ala Asn Thr 5000 5005 5010
Glu Val Leu Gln Trp Leu Ala Ala Gly Gln Ser Leu Ser Ile Ala 5015 5020 5025
Pro Met His Gly Asp Phe Asp Ala Ala Ala Trp Leu Glu Thr Leu 5030 5035 5040
Ala Thr Tyr Ala Ile Thr Val Ala Tyr Leu Ala Gln Val Glu Leu 5045 5050 5055
Thr Glu Met Leu Ala His Leu Glu Asn His Pro Leu Glu Arg Asn 5060 5065 5070
Lys Leu Ala Gly Leu Arg Val Leu Val Val His Gly Ala Pro Leu 5075 5080 5085
Pro Ile Ala Pro Leu Met Arg Leu Asp Ala Trp Leu Arg Glu Val 5090 5095 5100
Gly Gly Ser Ala Arg Ile Phe Ala Ala Tyr Gly Asn Ala Glu Phe 5105 5110 5115
Gly Ala Glu Ile Leu Ser Glu Asp Val Ser Ala Ala Leu Gln Ala 5120 5125 5130
Gly Ile Gly Ala Glu Tyr Lys His Arg Arg Gly Leu Phe Pro Leu 5135 5140 5145
Gly Ala Asn Ser Met Cys His Val Val Gln Ser Asn Gly Arg Ile 5150 5155 5160
Ala Pro Asp Gly Met Val Gly Glu Leu Trp Ile Thr Glu Pro Ala 5165 5170 5175
Cys Leu Tyr Lys Thr Asp Ala Leu Val Arg Arg Leu Ala Asn Gly 5180 5185 5190
Gln Leu Glu Trp Leu Gly Ser Leu Asp Val Gln Ser Arg Ile Asp 5195 5200 5205
Asp Pro Arg Ile Asp Leu Cys Val Val Glu Ala Gln Leu Arg Leu
Cys Glu Asp Val Gly Glu Ala Val Val Leu Tyr Glu Pro Leu Lys
Arg Cys Leu Val Ala Tyr Leu Ser Ala Arg Ser Thr Ala Ala Ile
Met Thr Asp Glu Thr Leu Ala Arg Ile Arg Gln Ala Leu Ser Glu
Thr Leu Pro Asp Tyr Leu Leu Pro Ala Ile Trp Val Pro Leu Ala
His Trp Pro Arg Leu Pro His Gly Arg Val Asp Leu Gly Ala Leu
Pro Ala Pro Asp Phe Asp Leu Ala Arg His Glu Ser Tyr Ile Ala
Pro Arg Thr Ala Val Glu Gln Ala Val Ala Glu Ile Trp Gln Arg
Val Leu Lys Arg Thr Gln Val Gly Val His Arg Asn Phe Phe Glu
Leu Gly Gly His Ser Val Leu Ala Ile Gln Leu Val Ser Gly Leu
Arg Lys Ala Leu Ala Ile Glu Val Pro Val Thr Leu Val Phe Glu
 Ala Pro Ile Leu Gly Ala Leu Ala Arg Gln Ile Ala Pro Leu Leu
 Val Ser Glu Arg Arg Pro Arg Pro Pro Gly Leu Thr Arg Leu Glu
His Thr Gly Pro Ile Pro Ala Ser Tyr Ala Gln Glu Arg Leu Trp
Leu Val His Glu His Met Glu Glu Gln Arg Thr Ser Tyr Asn Ile
Ser Asn Ala Ala His Phe Ile Gly Ala Ala Phe Ser Val Glu Ala
Met Arg Ala Ala Leu Asn Ala Leu Val Ala Arg His Glu Val Leu
Arg Thr Arg Phe Leu Ser Glu Asp Gly Gln Leu Gln Gln Val Ile
 Ala Ala Ser Leu Thr Leu Glu Val Pro Val Arg Glu Val Ser Ala
Glu Glu Val Asp Leu Leu Ala Ala Ser Thr Arg Glu Thr Phe
Asp Leu Arg Gln Gly Pro Leu Phe Lys Ala Arg Ile Leu Arg Val
 Ala Ala Asp His His Val Val Leu Ser Ser Ile His His Ile Ile
Ser Asp Gly Trp Ser Leu Gly Val Phe Asn Arg Asp Leu His Gln
Leu Tyr Glu Ala Cys Leu Arg Gly Thr Pro Pro Thr Leu Pro Thr
Leu Ala Val Gln Tyr Ala Asp Tyr Ala Leu Trp Gln Arg Gln Trp
Glu Leu Ala Ala Pro Leu Ser Tyr Thr Thr Arg Ala Leu Glu Gly
Tyr Asp Asp Gly Leu Asp Leu Pro Tyr Asp Arg Pro Arg Gly Ala
<table>
<thead>
<tr>
<th>Thr</th>
<th>Arg</th>
<th>Ala</th>
<th>Trp</th>
<th>Arg</th>
<th>Ala</th>
<th>Gly</th>
<th>Leu</th>
<th>Val</th>
<th>Lys</th>
<th>His</th>
<th>Arg</th>
<th>Tyr</th>
<th>Pro</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>5615</td>
<td>5620</td>
<td>5625</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Leu</td>
<td>Ala</td>
<td>Gln</td>
<td>Gln</td>
<td>Leu</td>
<td>Ala</td>
<td>Ala</td>
<td>Tyr</td>
<td>Ser</td>
<td>Gln</td>
<td>Gln</td>
<td>Tyr</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>5630</td>
<td>5635</td>
<td>5640</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Leu</td>
<td>Phe</td>
<td>Met</td>
<td>Ser</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Val</td>
<td>Leu</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>5645</td>
<td>5650</td>
<td>5655</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Tyr</td>
<td>Ala</td>
<td>Asp</td>
<td>Arg</td>
<td>Lys</td>
<td>Asp</td>
<td>Val</td>
<td>Cys</td>
<td>Ile</td>
<td>Gly</td>
<td>Ala</td>
<td>Thr</td>
<td>Val</td>
<td>Ser</td>
</tr>
<tr>
<td>5660</td>
<td>5665</td>
<td>5670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Asp</td>
<td>Gln</td>
<td>Leu</td>
<td>Glu</td>
<td>Leu</td>
<td>Glu</td>
<td>Glu</td>
<td>Leu</td>
<td>Ile</td>
<td>Gly</td>
<td>Phe</td>
<td>Phe</td>
<td>Ile</td>
</tr>
<tr>
<td>5675</td>
<td>5680</td>
<td>5685</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Ile</td>
<td>Leu</td>
<td>Pro</td>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
<td>Ser</td>
<td>Gly</td>
<td>Asp</td>
<td>Pro</td>
<td>Cys</td>
<td>Leu</td>
</tr>
<tr>
<td>5690</td>
<td>5695</td>
<td>5700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Glu</td>
<td>Val</td>
<td>Leu</td>
<td>Leu</td>
<td>Arg</td>
<td>Thr</td>
<td>Arg</td>
<td>Gln</td>
<td>Val</td>
<td>Val</td>
<td>Leu</td>
<td>Asp</td>
<td>Gly</td>
<td>Phe</td>
</tr>
<tr>
<td>5705</td>
<td>5710</td>
<td>5715</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>His</td>
<td>Gln</td>
<td>Ser</td>
<td>Val</td>
<td>Pro</td>
<td>Phe</td>
<td>Glu</td>
<td>His</td>
<td>Val</td>
<td>Leu</td>
<td>Gln</td>
<td>Ala</td>
<td>Leu</td>
<td>Arg</td>
</tr>
<tr>
<td>5720</td>
<td>5725</td>
<td>5730</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Gln</td>
<td>Arg</td>
<td>Asp</td>
<td>Ser</td>
<td>Ser</td>
<td>Gln</td>
<td>Ile</td>
<td>Pro</td>
<td>Leu</td>
<td>Val</td>
<td>Pro</td>
<td>Val</td>
<td>Met</td>
<td>Leu</td>
</tr>
<tr>
<td>5735</td>
<td>5740</td>
<td>5745</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>His</td>
<td>Gln</td>
<td>Asn</td>
<td>Phe</td>
<td>Pro</td>
<td>Thr</td>
<td>Gln</td>
<td>Glu</td>
<td>Ile</td>
<td>Gly</td>
<td>Asp</td>
<td>Trp</td>
<td>Pro</td>
<td>Glu</td>
</tr>
<tr>
<td>5750</td>
<td>5755</td>
<td>5760</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Val</td>
<td>Arg</td>
<td>Leu</td>
<td>Thr</td>
<td>Gln</td>
<td>Met</td>
<td>Glu</td>
<td>Leu</td>
<td>Gly</td>
<td>Leu</td>
<td>Asp</td>
<td>Arg</td>
<td>Ser</td>
<td>Thr</td>
</tr>
<tr>
<td>5765</td>
<td>5770</td>
<td>5775</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Ser</td>
<td>Glu</td>
<td>Leu</td>
<td>Asp</td>
<td>Trp</td>
<td>Gln</td>
<td>Phe</td>
<td>Tyr</td>
<td>Gly</td>
<td>Asp</td>
<td>Gly</td>
<td>Ser</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td>5780</td>
<td>5785</td>
<td>5790</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Thr</td>
<td>Leu</td>
<td>Glu</td>
<td>Tyr</td>
<td>Ala</td>
<td>Gln</td>
<td>Asp</td>
<td>Leu</td>
<td>Phe</td>
<td>Asp</td>
<td>Glu</td>
<td>Ala</td>
<td>Thr</td>
</tr>
<tr>
<td>5795</td>
<td>5800</td>
<td>5805</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Arg</td>
<td>Arg</td>
<td>Met</td>
<td>Ile</td>
<td>Ala</td>
<td>His</td>
<td>Gln</td>
<td>Gln</td>
<td>Ala</td>
<td>Leu</td>
<td>Glu</td>
<td>Ala</td>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>5810</td>
<td>5815</td>
<td>5820</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Ser</td>
<td>Arg</td>
<td>Pro</td>
<td>Glu</td>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>Gly</td>
<td>Lys</td>
<td>Trp</td>
<td>Asp</td>
<td>Met</td>
<td>Leu</td>
<td>Thr</td>
</tr>
<tr>
<td>5825</td>
<td>5830</td>
<td>5835</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Glu</td>
<td>Arg</td>
<td>Arg</td>
<td>Leu</td>
<td>Phe</td>
<td>Ala</td>
<td>Ala</td>
<td>Leu</td>
<td>Asn</td>
<td>Ala</td>
<td>Thr</td>
<td>Gly</td>
<td>Thr</td>
</tr>
<tr>
<td>5840</td>
<td>5845</td>
<td>5850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Arg</td>
<td>Glu</td>
<td>Trp</td>
<td>Pro</td>
<td>Ser</td>
<td>Leu</td>
<td>Ala</td>
<td>Gln</td>
<td>Gln</td>
<td>Phe</td>
<td>Glu</td>
<td>Arg</td>
<td>Gln</td>
<td>Ala</td>
</tr>
<tr>
<td>5855</td>
<td>5860</td>
<td>5865</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
<td>Thr</td>
<td>Pro</td>
<td>Glu</td>
<td>Ala</td>
<td>Ile</td>
<td>Ala</td>
<td>Cys</td>
<td>Val</td>
<td>Ser</td>
<td>Asp</td>
<td>Gly</td>
<td>Gln</td>
<td>Ser</td>
</tr>
<tr>
<td>5870</td>
<td>5875</td>
<td>5880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td>Ser</td>
<td>Tyr</td>
<td>Ala</td>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
<td>Arg</td>
<td>Ala</td>
<td>Asn</td>
<td>Gln</td>
<td>Ala</td>
<td>Arg</td>
<td>Ala</td>
<td>Gln</td>
</tr>
<tr>
<td>5885</td>
<td>5890</td>
<td>5895</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Leu</td>
<td>Arg</td>
<td>Gly</td>
<td>Gln</td>
<td>Gly</td>
<td>Ala</td>
<td>Gly</td>
<td>Arg</td>
<td>Asp</td>
<td>Val</td>
<td>Arg</td>
<td>Val</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td>5900</td>
<td>5905</td>
<td>5910</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ser</td>
<td>Ala</td>
<td>Arg</td>
<td>Thr</td>
<td>Pro</td>
<td>Glu</td>
<td>Leu</td>
<td>Leu</td>
<td>Met</td>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>5915</td>
<td>5920</td>
<td>5925</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
<td>Cys</td>
<td>Tyr</td>
<td>Val</td>
<td>Pro</td>
<td>Ile</td>
<td>Asp</td>
<td>Pro</td>
<td>Ala</td>
<td>Tyr</td>
<td>Pro</td>
</tr>
<tr>
<td>5930</td>
<td>5935</td>
<td>5940</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Tyr</td>
<td>Arg</td>
<td>Glu</td>
<td>Ile</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Val</td>
<td>Gln</td>
<td>Val</td>
<td>Ser</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>5945</td>
<td>5950</td>
<td>5955</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Glu</td>
<td>Gln</td>
<td>Gly</td>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Asp</td>
<td>Glu</td>
<td>Gln</td>
<td>Gly</td>
<td>Gln</td>
<td>Phe</td>
</tr>
<tr>
<td>5960</td>
<td>5965</td>
<td>5970</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Asn</td>
<td>Arg</td>
<td>Arg</td>
<td>Trp</td>
<td>Arg</td>
<td>Glu</td>
<td>Ala</td>
<td>Pro</td>
<td>Thr</td>
<td>Pro</td>
<td>Leu</td>
<td>Gly</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>5975</td>
<td>5980</td>
<td>5985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>His</td>
<td>Pro</td>
<td>Gly</td>
<td>Asp</td>
<td>Leu</td>
<td>Ala</td>
<td>Cys</td>
<td>Val</td>
<td>Met</td>
<td>Val</td>
<td>Thr</td>
<td>Ser</td>
<td>Gly</td>
</tr>
<tr>
<td>5990</td>
<td>5995</td>
<td>6000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ser Thr Gly Arg Pro Lys Gly Val Met Val Pro Tyr Ala Gln Leu 0605 0610 0615
His Asn Trp Leu His Ala Gly Trp Gln Arg Ser Ala Phe Glu Ala 0620 0625 0630
Gly Glu Arg Val Leu Gln Lys Thr Ser Ile Ala Phe Ala Val Ser 0635 0640 0645
Val Lys Glu Leu Leu Ser Gly Leu Leu Ala Gly Val Gly Gln Val 0650 0655 0660
Met Leu Pro Asp Glu Gln Val Lys Asp Ser Leu Ala Leu Ala Arg 0665 0670 0675
 Ala Ile Glu Gln Trp Gln Val Thr Arg Leu Tyr Leu Val Pro Ser 0680 0685 0690
His Leu Gln Ala Leu Leu Asp Ala Thr Gln Gln Arg Asp Gly Leu 0695 0700 0705
Leu His Ser Leu Arg His Val Thr Ala Gln Glu Ala Leu Pro 0710 0715 0720
Ser Ala Val Gly Glu Ala Val Arg Val Arg Leu Pro Gln Val Gln 0725 0730 0735
Leu Trp Aem Asn Tyr Gly Cys Thr Glu Leu Asn Asp Ala Thr Tyr 0740 0745 0750
His Arg Ser Asp Thr Val Ala Pro Gly Thr Phe Val Pro Ile Gln 0755 0760 0765
 Ala Pro Ile Ala Asn Thr Glu Val Tyr Val Leu Asp Arg Gln Leu 0770 0775 0780
Arg Gln Val Pro Ile Gln Val Met Gly Glu Leu His Val His Ser 0785 0790 0795
Val Gly Met Ala Arg Gly Tyr Trp Asn Arg Pro Gln Leu Thr Ala 0800 0805 0810
Ser Arg Phe Ile Ala His Pro Tyr Ser Glu Glu Pro Gly Thr Arg 0815 0820 0825
Leu Tyr Lys Thr Gly Asp Met Val Arg Arg Leu Ala Asp Gly Thr 0830 0835 0840
Leu Glu Tyr Leu Gly Arg Gln Asp Phe Glu Val Lys Val Arg Gln 0845 0850 0855
His Arg Val Asp Thr Arg Gln Val Glu Ala Ala Leu Arg Ala Gln 0860 0865 0870
 Pro Ala Val Ala Gln Ala Val Ser Gly His Arg Val Asp Gly 0875 0880 0885
Asp Met Gln Leu Val Ala Tyr Val Val Ala Arg Glu Gly Gln Ala 0890 0895 0900
Pro Ser Ala Gly Glu Leu Lys Gln Gln Leu Ser Ala Gln Leu Pro 0905 0910 0915
Thr Tyr Met Leu Pro Thr Val Tyr Glu Trp Leu Gln Leu Pro 0920 0925 0930
Arg Leu Ser Asn Gly Lys Leu Asp Arg Leu Ala Leu Pro Ala Pro 0935 0940 0945
Gln Val Val His Ala Gln Glu Tyr Val Ala Pro Arg Asn Glu Ala 0950 0955 0960
Glu Gln Arg Leu Ala Ala Leu Phe Ala Glu Val Leu Arg Val Glu 0965 0970 0975
Gln Val Gly Ile His Asp Asn Phe Phe Ala Leu Gln Gly His Ser 0980 0985 0990
Leu Ser Ala Ser Gln Leu Ile Ser Arg Ile Arg Gln Ser Phe His
<table>
<thead>
<tr>
<th>Residue 1</th>
<th>Residue 2</th>
<th>Residue 3</th>
<th>Residue 4</th>
<th>Residue 5</th>
<th>Residue 6</th>
<th>Residue 7</th>
<th>Residue 8</th>
<th>Residue 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val</td>
<td>Arg</td>
<td>Asp</td>
<td>Leu</td>
<td>Pro</td>
<td>Val</td>
<td>Leu</td>
<td>Ser</td>
<td>Arg</td>
</tr>
<tr>
<td>6395</td>
<td>6400</td>
<td>6405</td>
<td>6415</td>
<td>6420</td>
<td>6425</td>
<td>6430</td>
<td>6435</td>
<td>6440</td>
</tr>
<tr>
<td>Gly</td>
<td>Leu</td>
<td>Val</td>
<td>Arg</td>
<td>Gln</td>
<td>Leu</td>
<td>Ala</td>
<td>Pro</td>
<td>Ser</td>
</tr>
<tr>
<td>6440</td>
<td>6445</td>
<td>6450</td>
<td>6455</td>
<td>6460</td>
<td>6465</td>
<td>6470</td>
<td>6475</td>
<td>6480</td>
</tr>
<tr>
<td>Ser</td>
<td>Ile</td>
<td>Ala</td>
<td>Arg</td>
<td>Val</td>
<td>Ala</td>
<td>Arg</td>
<td>Asp</td>
<td>Thr</td>
</tr>
<tr>
<td>6500</td>
<td>6505</td>
<td>6510</td>
<td>6515</td>
<td>6520</td>
<td>6525</td>
<td>6530</td>
<td>6535</td>
<td>6540</td>
</tr>
<tr>
<td>Arg</td>
<td>His</td>
<td>Glu</td>
<td>Ile</td>
<td>Leu</td>
<td>Arg</td>
<td>Thr</td>
<td>Arg</td>
<td>Phe</td>
</tr>
<tr>
<td>6545</td>
<td>6550</td>
<td>6555</td>
<td>6560</td>
<td>6565</td>
<td>6570</td>
<td>6575</td>
<td>6580</td>
<td>6585</td>
</tr>
<tr>
<td>Val</td>
<td>Met</td>
<td>Asn</td>
<td>Arg</td>
<td>Asp</td>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>Tyr</td>
</tr>
<tr>
<td>6590</td>
<td>6595</td>
<td>6600</td>
<td>6605</td>
<td>6610</td>
<td>6615</td>
<td>6620</td>
<td>6625</td>
<td>6630</td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
<td>Asp</td>
<td>Leu</td>
<td>Pro</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>6665</td>
<td>6670</td>
<td>6675</td>
<td>6680</td>
<td>6685</td>
<td>6690</td>
<td>6695</td>
<td>6700</td>
<td>6705</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Phe</td>
<td>Ala</td>
<td>Val</td>
<td>Leu</td>
<td>Asn</td>
<td>Tyr</td>
<td>Lys</td>
</tr>
<tr>
<td>6710</td>
<td>6715</td>
<td>6720</td>
<td>6725</td>
<td>6730</td>
<td>6735</td>
<td>6740</td>
<td>6745</td>
<td>6750</td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Ile</td>
<td>Gly</td>
<td>Thr</td>
<td>Thr</td>
<td>Ala</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td>6755</td>
<td>6760</td>
<td>6765</td>
<td>6770</td>
<td>6775</td>
<td>6780</td>
<td>6785</td>
<td>6790</td>
<td>6795</td>
</tr>
</tbody>
</table>
Thr Ile Gly Asp Trp Ser Asp Gly Ile Arg Thr Glu Val Ile Gln
6800 6805 6810
Arg Asp Leu Arg Ala Thr Pro Asn Glu Met Asp Leu Gln Phe Phe
6815 6820 6825
Gly Asp Gly Thr Gly Leu Ser Val Thr Val Glu Tyr Ala Ala Glu
6830 6835 6840
Leu Phe Ser Glu Ala Thr Ile Arg Arg Leu Ile His His His Gln
6845 6850 6855
Leu Val Leu Glu Gln Met Leu Ala Ala His Glu Ser Ala Thr Cys
6860 6865 6870
Pro Leu Asp Val Ala Asp
6875

<210> SEQ ID NO: 27
<211> LENGTH: 343
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 27

Met Asp Ser Ala Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr
1 5 10 15
Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile
20 25 30
Glu Leu Gly Leu Phe Asp Val Val Gly Gin Gin Gly Arg Thr Pro Ala
35 40 45
Ala Ile Ala Glu Ala Cys Gin Ala Ser Pro Arg Gly Ile Arg Ile Leu
50 55 60
Cys Tyr Tyr Leu Val Ser Ile Gly Phe Leu Arg Arg Asn Gly Gly Leu
65 70 75 80
Phe Tyr Ile Asp Arg Asn Met Ala Met Tyr Leu Asp Arg Ser Ser Pro
95 100 105 110
Gly Tyr Leu Gly Gly Ser Ile Lys Phe Leu Leu Ser Pro Tyr Ile Met
115 120 125
Ser Ala Phe Thr Asp Leu Thr Ala Val Val Arg Thr Gly Lys Ile Asn
130 135 140
Leu Ala Gin Asp Gly Val Ala Pro Asp His Pro Gin Trp Val Glu
145 150 155 160
Phe Ala Arg Ala Met Ala Pro Met Met Ala Leu Pro Ser Ala Leu Ile
165 170 175
Ala Asn Met Val Ser Leu Pro Ala Asp Arg Pro Ile Arg Val Leu Asp
180 185 190
Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala Gin Arg Phe
195 200 205
Arg Gin Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val Val Leu Asp Val
210 215 220
Ala Arg Glu Asn Ala Gin Ala Ala Lys Val Ala Glu Arg Ala Arg Phe
225 230 235 240
Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly Tyr Asp Val
245 250 255
Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val Asp Gly Glu
260 265 270
Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp Gly Met Val
275 280 285
Ile Thr Phe Glu Phe Ile Ala Asp Glu Glu Arg Ser Ser Pro Pro Leu
Ala Ala Thr Phe Ser Met Met Leu Gly Thr Thr Pro Ala Gly Glu
290 295 300
Ser Tyr Thr Tyr Ser Asp Leu Glu Arg Met Phe Arg His Ala Gly Phe
305 310 315 320
Gly His Val Glu Leu Lys Ser Ile Pro Pro Ala Leu Leu Lys Val Val
325 330 335
Val Ser Arg Lys Arg Ala Pro 340

SEQ ID NO 28
LENGTH: 167
TYPE: PRT
ORGANISM: Xanthomonas albineans

<400> SEQUENCE: 28
Met Ile Glu Ser Ala Thr Ser Pro Val Ala Lys Thr Glu Arg Ile Trp
1 5 10 15
Cys Thr Glu Leu Asp Leu Asp Ala Leu Asn Ala Met Ser Ala Asn Thr
20 25 30
Met Gln Ala Leu Leu Gly Ile Arg Met Ile Glu Ile Gly Ser Asp Tyr
35 40 45
Leu Val Ser Cys Met Ser Val Asp Trp Arg Cys His Gln Pro Tyr Gly
50 55 60
Val Leu His Gly Gly Ala Ser Val Thr Leu Ala Glu Ala Thr Gly Ser
65 70 75 80
Met Ala Ala Ser Met Cys Val Pro Ala Gly Gln Arg Cys Val Gly Leu
85 90 95
Asp Ile Asn Ala Asn His Ile Ala Ser Ile Ser Gly Glu Val Glu
100 105 110
Cys Ile Ala Arg Pro Leu His Ile Gly Ala Leu Thr Glu Trp Glu
115 120 125
Met Arg Ile Tyr Asp Glu Gly Asp Arg Thr Ile Cys Val Ser Arg Leu
130 135 140
Thr Met Ala Val Leu Ser Val His Val Arg Val Ser Pro Asn Pro
145 150 155 160
Ala Ser Ser Gly Val Glu Thr
165

SEQ ID NO 29
LENGTH: 941
TYPE: PRT
ORGANISM: Xanthomonas albineans

<400> SEQUENCE: 29
Met Asn Glu Thr Ala Thr Val Thr Lys Ala Thr Leu Ser Ser Ala Lys
1 5 10 15
Ala Ser Ile Thr Pro Ala Cys Val His Gln Trp Phe Glu Ala Gln Val
20 25 30
Ser Ser Thr Pro Asp Ala Pro Ala Ala Phe Leu Gly Glu Arg Arg Met
35 40 45
Ser Tyr Gly Glu Leu Asn Thr Arg Ala Asn Arg Leu Ala Arg Leu Leu
50 55 60
Gln Ser Gln Gly Val Gly Pro Gly Ala Arg Val Ala Val Trp Met Asn
65 70 75 80
Arg Ser Pro Glu Cys Leu Ala Ala Leu Leu Ala Val Met Lys Ala Gly
<table>
<thead>
<tr>
<th>85</th>
<th>90</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala Ala Tyr Val Pro Ile Asp Leu Ser Leu Pro Ile Arg Arg Val Gln</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td>Tyr Ile Leu Gln Asp Ser Gln Ala Arg Leu Val Leu Val Asp Asp Glu</td>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td>Gly Gln Gly Arg Leu Asp Glu Leu Leu Gly Ala Met Thr Ala Val</td>
<td>130</td>
<td>135</td>
</tr>
<tr>
<td>Asp Val Cys Gly Thr Leu Asp Gly Asp Glu Ala Asn Leu Asp Leu Pro</td>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td>Cys Asp Pro Ala Gln Pro Val Tyr Cys Ile Tyr Thr Ser Gly Ser Thr</td>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td>Gly Ser Pro Lys Gly Val Leu Val Arg His Ser Gly Leu Ala Asn Tyr</td>
<td>180</td>
<td>185</td>
</tr>
<tr>
<td>Val Ala Trp Ala Lys Arg Gln Tyr Val Thr Ala Asp Thr Thr Ser Phe</td>
<td>195</td>
<td>200</td>
</tr>
<tr>
<td>Ala Phe Tyr Ser Ser Leu Ser Phe Asp Leu Thr Val Thr Ser Ile Tyr</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td>Val Pro Leu Val Ala Gly Leu Cys Val His Val Tyr Pro Glu Gin Gly</td>
<td>225</td>
<td>230</td>
</tr>
<tr>
<td>Asp Asp Val Pro Val Ile Asn Arg Val Leu Asp Asp Asn Gin Val Asp</td>
<td>245</td>
<td>250</td>
</tr>
<tr>
<td>Val Ile Lys Leu Thr Pro Ser His Met Leu Met Leu Arg Asn Ala Ala</td>
<td>260</td>
<td>265</td>
</tr>
<tr>
<td>Leu Ala Thr Ser Arg Leu Lys Thr Leu Ile Val Gly Gly Glu Asp Leu</td>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td>Lys Ala Ala Val Ala Tyr Asp Ile His Gin Arg Phe Arg Arg Asp Val</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td>Ala Ile Tyr Asn Glu Tyr Gly Pro Thr Glu Thr Val Val Gly Cys Ala</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>Ile His Arg Tyr Asp Pro Ala Thr Glu Arg Glu Gly Ser Val Pro Ile</td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td>Gly Val Pro Ile Asp His Thr Ser Leu His Leu Leu Asp Glu Arg Leu</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td>Gin Pro Val Ala Pro Gly Glu Val Gly Gin Ile His Ile Gly Gly Ala</td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td>Gly Val Ala Ile Gly Tyr Val Asn Lys Pro Glu Ile Thr Asp Ala Gin</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td>Phe Ile Asp Asn Pro Phe Glu Gly Ser Gly Arg Leu Tyr Ala Ser Gly</td>
<td>395</td>
<td>390</td>
</tr>
<tr>
<td>Asp Leu Gly Arg Met Arg Ala Asp Gly Leu Leu Glu Phe Leu Gly Arg</td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>Lys Asp Ser Gin Ile Lys Leu Arg Gly Tyr Arg Ile Glu Leu Gly Glu</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>Ile Glu Asn Val Leu Leu Gly His Ala Ala Leu Arg Glu Cys Ile Val</td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>Asp Thr Thr Val Ala Pro Arg Arg Asp Tyr Ser Thr Val Ser Leu Arg</td>
<td>450</td>
<td>455</td>
</tr>
<tr>
<td>Tyr Cys Ala Arg Cys Gly Ile Ala Ser Asn Phe Pro Asn Thr Ser Phe</td>
<td>465</td>
<td>470</td>
</tr>
<tr>
<td>Asp Glu His Gly Val Cys Asn His Cys His Ala Tyr Asp Lys Tyr Arg</td>
<td>485</td>
<td>490</td>
</tr>
<tr>
<td>Asn Val Val Glu Asp Tyr Phe Arg Thr Glu Asp Glu Leu Arg Thr Ile</td>
<td>500</td>
<td>505</td>
</tr>
</tbody>
</table>
<210> SEQ ID NO 30
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>930</td>
<td>935</td>
<td>940</td>
</tr>
</tbody>
</table>

<400> SEQUENCE: 30
Met Asp Leu Gln Cys Ala Arg Ile Ala Ala Leu Cys Glu Gln Leu Lys  
1  5 10 15
Leu Ala Arg Leu Ser Ser Asp Trp Gln Ala Leu Ala Gln Ala Ala Ala  
20 25 30
Cys Glu Asp Ala Ser Tyr Phe Leu Glu Lys Val Leu Ala Ser Glu Gln  
35 40 45
Leu Ala Arg Glu Arg Lys Arg Thr Val Leu Thr Arg Leu Ala Arg  
50 55 60
Met Pro Ser Ile Lys Thr Leu Glu Gln Phe Asp Trp Ala Gln Ala Gly  
65 70 75 80
Gly Ala Ser Lys Ala Gin Ile Val Glu Leu Gly His Leu Thr Phe Val  
95 90 95
Glu Arg Ala Gin Val Val Met Leu Gly Pro Ser Gly Val Gly Gly Lys  
100 105 110
Thr His Ile Ala Leu Ala Leu Cys Gin Gin Ala Val Met Ala Gly His  
115 120 125
Lys Ala Arg Phe Ile Thr Ala Ala Asp Leu Met Gin Leu Ala Ala  
130 135 140
Val Lys Ala Gin Asn Arg Leu Lys Asp Tyr Phe Asn Arg Ala Val Leu  
145 150 155 160
Gly Pro Lys Leu Leu Val Val Asp Gln Ile Gly Tyr Leu Pro Phe Gly  
165 170 175
Arg Glu Pro Ala Gin Gly Cys Trp Ala Ala Thr Gly Phe Ala Leu Arg  
180 185 190
Ser Leu Ala Ala Arg Arg Trp Lys Thr Pro Gly Ser Asp Leu Leu  
195 200 205
Arg Arg Phe Lys Gly Lys Trp Val Lys Phe Lys Ser Ala Leu Thr Ala  
210 215 220
Asp Val Val Tyr Leu Ile Phe Arg Leu Arg Gin Val Ser Asp His Pro  
225 230 235

<210> SEQ ID NO 31
<211> LENGTH: 286
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

<400> SEQUENCE: 31
Met Pro Arg Ile Glu Tyr Cys Ile Ser Met Met His Arg Arg Lys Pro  
1  5 10 15
Thr Thr Asn Arg Ser Val Cys Met Arg Asp Ile Glu Arg Thr Ala Leu  
20 25 30
Trp Val Ala Gly Met Arg Ala Leu Glu Ser Glu Arg Gln Ala Leu  
35 40 45
Phe His Asp Pro Phe Ala Arg Arg Leu Ala Gly Asp Glu Phe Val Glu  
50 55 60
Glu Leu Arg Asp Asn Gin Asn Val Val Pro Met Pro Pro Ala Ile Glu  
65 70 75 80
---continued---

Val Arg Thr Arg Trp Leu Asp Asp Lys Ile Met Gln Ala Val Ser Glu 85 90 95
Gly Ile Gly Gln Val Val Ile Leu Ala Ala Gly Met Asp Ala Arg Ala 100 105 110
Tyr Arg Leu Pro Trp Pro Ser Asp Thr Arg Val Tyr Glu Ile Asp His 115 120 125
Met Asp Val Leu Ser Asp His Gln Lys Leu His Asp Ala Gln Pro 130 135 140
Val Cys Gln Arg Ile Ala Leu Pro Ile Asp Leu Arg Glu Asp Trp Pro 145 150 155 160
Gln Ala Leu Lys Glu Ser Gly Phe Val Gly Ser Ala Ala Thr Leu Trp 165 170 175
Leu Val Glu Gly Leu Leu Cys Tyr Leu Ser Ala Glu Ala Val Met Leu 180 185 190
Leu Phe Ala Arg Ile Asp Ala Leu Ser Ala Lys Gly Ser Ser Val Leu 195 200 205
Phe Asp Val Ile Gly Leu Ser Met Leu Aaa Ser Pro Aaa Ser Ala Arg Val 210 215 220
Leu His Ala Met Ala Arg Ile Gly Thr Asp Glu Pro Glu Ser Leu 225 230 235 240
Ile Gln Pro Leu Gly Trp Glu Pro Gly Val Leu Thr Ile Ala Ala Ala 245 250 255 260
Gly Gln Gln Met Gly Arg Trp Pro Phe Pro Val Ala Pro Arg Gly Thr 260 265 270
His Gly Val Pro Gln Ser Tyr Leu Val His Ala Leu Lys Arg 275 280 285

<210> SEQ ID NO 32
<211> LENGTH: 765
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 32

Met Arg Arg Ser Pro Tyr Pro Arg Thr Leu Met Asp Ser Pro Leu Thr 1 5 10 15
Aan Leu Pro Met His Ser Gly Thr Glu Leu Asp Leu Arg Trp Ser Val 20 25 30
Gly Gln Thr Arg Pro Gly Arg Asn Glu Ala Tyr Ala Arg Gln Trp Thr 35 40 45
Thr Leu Leu His Gln Trp Arg Arg Asp Tyr Pro Gly Leu Arg Ile Asp 50 55 60
Val Ser Asp Thr Pro Ile Gly Gln His Ile Thr Ile Asp Tyr Ala Ala 65 70 75 80
Pro Tyr Pro Cys Gly Ser Phe Gly Ser Leu Leu Arg Glu Tyr Ala Arg 85 90 95
Leu Gly Lys Leu Ala Gly Leu Ala Gly Asp Tyr Leu Lys His Arg His 100 105 110
Gln Ile Val Leu Ser Glu Ser Pro Pro Gly Ala Asn Thr Leu Ala Leu 115 120 125
Asp Leu Gly Arg Ile Glu Glu Pro Cys Gln Leu Leu Arg Asp Leu Gln Gly 130 135 140
Ala Leu Gly Met Ala Leu Glu Ala Ala Thr Arg Arg Ser Asp Gly 145 150 155 160
Leu Leu Leu Trp His Ala Asp His Arg Gln Arg Asn Leu Pro Asp Leu 165 170 175
Arg Asp Ser Ala Val Cys Gly Ser Ala Ala Ala Gin Ile Ser Leu Pro Ala 180 185 190
Leu Ser Cys Val Glu Asp Leu Ile Glu Val Asp Thr Ser Leu Leu Ala 195 200 205
Cys Asp His Gly Lys Leu Cys Gin Ile Ala Ser His Leu Pro Ala Ser 210 215 220
Trp Phe Ala Arg Ser Thr Asp Gly Pro Met Pro Ser Trp Ser Asp Ala 225 230 235 240
Ser Thr Ala Val Phe Ala Cys Ala Pro Ile Gly Phe Leu Pro Ser Val 245 250 255
Gln Val Asn Val Cys Ala Gin Ile Phe Ser Ala Ala His Leu Ala Ser 260 265 270
Thr Ala Gin Met Ile Asp Pro Leu Arg Gin Gin Ala Phe Ser Tyr Arg 275 280 285
Gln Leu Arg Ser Arg Ala Ala Thr Tyr Ala Arg His Leu Ser Leu Leu 290 295 300
Gly Leu Gin Ser Gly Asp Ala Val Ala Leu Ile Ala Ile Asp Ser Leu 305 310 315 320
Ala Gly Val Ala Leu Met Leu Ala Cys Leu Ala Ala Gly Gly Leu Val Phe 325 330 335
Ala Pro Ile Asp Leu Val Ser Leu Val His Phe Glu Thr Thr Leu 340 345 350
Lys Thr Ile Lys Pro Arg Leu Val Leu Ile Asp Ala Glu Leu Pro Pro 365 366 370
Ser His His Ala Ala Leu Arg His Leu Pro Thr Leu Glu Leu Thr Ser 375 380
Leu Met Pro Val Ile Glu Asn Asp Glu Leu Val Ala Pro Cys Ser 385 390 395 400
Ala Asp Ala Pro Ala Val Met Ile Cys Thr Ser Gly Ser Thr Gly Thr 405 410 415
Pro Lys Ala Val Thr His Ser His Ala Asp Phe Met His Cys His Leu 420 425 430
Asn Tyr Gin Gin Ala Val Leu Gly Leu Arg Ser Asp Asp Val Met Tyr 430 440 445
Thr Pro Ser Arg Leu Phe Phe Ala Tyr Gly Leu Asn Asn Leu Met Leu 450 455 460
Ser Leu Leu Ala Gly Val Ser His Val Ile Ala Ala Pro Leu Ser Val 465 470 475 480
Arg Gin Ile Ala Gin Thr Ile His Thr Tyr His Val Thr Val Leu Leu 485 490 495
Ala Val Pro Ala Val Phe Lys Leu Leu Ala Glu Ala Ala Ala Pro Asp 500 505 510
Ala Val Trp Pro Ala Leu Arg Leu Cys Ile Ser Ala Gly Gly Ser Leu 515 520 525
Pro Ala Arg Leu Gly His Ala Ile Ser Thr Arg Trp Gin Val Glu Val 530 535 540
Leu Asp Gly Ile Gly Cys Thr Glu Val Leu Ser Thr Phe Ile Ser Asn 545 550 555 560
Arg Pro Gly His Ala Leu Met Gly Cys Thr Gly Thr Pro Val Pro Gly 565 570 575
Phe Val Val Lys Leu Val Asn Lys Gin Gly Glu Ile Cys Arg Ile Gly 580 585 590
Glu Val Gly Ser Leu Trp Val Arg Gly Asn Thr Leu Thr Arg Gly Tyr
595
600
605
Val Gly Asp Pro Ile Leu Ser Ala Gln Leu Phe Val Asp Gly Trp Phe
610
615
620
Asp Thr Arg Asp Leu Phe Phe Ala Asp Ala Lys Gly Arg Phe His Asn
625
630
635
640
Leu Gly Arg Met Gly Ser Ala Ile Lys Ile Asn Gly Cys Trp Leu Ser
645
650
655
660
Glu Thr Leu Glu Ser Val Ile Gln Thr His Ala Cys Val Lys Glu
665
670
Cys Ala Ile Cys Leu Ile Glu Asp Glu Phe Gly Leu Pro Arg Pro Ala
675
680
685
Asp Leu Val Val Pro Val Asp Ala Ser Ile Asp Thr Gly Ala Leu Trp
690
695
700
 Ala Ala Leu Arg Ala Leu Cys Lys Asn Ala Leu Gly Lys His Tyr
705
710
715
720
Pro His Leu Phe Val Glu Val Ser Thr Ile Pro Arg Thr Cys Ser Gly
725
730
735
Lys Val Ile Arg Pro Ala Leu Glu Thr Leu Ala Ser Ala Lys His
740
745
750
Leu Gln Ser His Leu Phe Phe Val Gly His Ala Arg Thr
755
760
765

<210> SEQ ID NO 33
<211> LENGTH: 330
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 33

Met His Thr Asn Ala Asp Leu Pro Leu Thr Ile Lys Ala Asp Ser Ala
1
5
10
15
Glu Ala Thr Leu Thr Asp Trp Asn Ala Thr His Arg Ala Thr Trp Pro
20
25
30
Thr Leu Leu Trp Gln His Arg Ala Leu Leu Phe Arg Gly Phe Ala His
35
40
45
Pro Gly Gly Leu Gln Ile Ser Arg Cys Phe Phe Asp Glu Arg Leu
50
55
60
Ala Tyr Thr Tyr Arg Ser Thr Pro Arg Thr Asp Val Gly Gln His Val
65
70
75
80
Tyr Thr Ala Thr Glu Tyr Pro Arg Gln Leu Ser Ile Ala Gln His Cys
85
90
95
Gl Glu Asn Ala Tyr Gln Arg Val Trp Pro Met Lys Leu Leu Phe His Cys
100
105
110
Val Gln Pro Ala Ser Glu Gly Cys Thr Pro Leu Ala Asp Met Leu
115
120
125
Lys Val Thr Ala Ala Asp Pro Glu Val Arg Glu Ile Phe Ala Arg
130
135
140
Lys Gln Val Arg Tyr Val Arg Asn Tyr Arg Ala Gly Val Asp Leu Pro
145
150
155
160
Trp Glu Asp Val Phe Asn Thr Arg Asn Lys Gin Gln Glu Val Glu Ala Tyr
165
170
175
Cys Ala Arg Asn Asp Met Gin Cys Glu Trp Thr Gly Asp Gly Leu Arg
180
185
190
Thr Ser Gln Ile Cys Arg Ala Phe Ala Cys His Pro Ala Thr Gly Asp
195
200
205
Glu Val Trp Phe Asn Gln Ala His Leu Phe His Tyr Thr Ala Leu Glu
  210  215  220
  Ala Ala Ala Gln Lys Met Met Leu Ser Phe Phe Gly Glu Gln Gly Leu
  225  230  235  240
Pro Arg Asn Ala Tyr Phe Gly Asp Gly Thr Pro Ile Asp Pro Ala Met
  245  250  255
Leu Asp His Val Arg Thr Val Phe Ala Gin His Lys Ile His Phe Asp
  260  265  270
Trp His Arg Asp Asp Val Leu Leu Ile Asp Asn Met Leu Val Ser His
  275  280  285
Gly Arg Glu Pro Tyr Glu Gly Ser Arg Lys Ile Leu Val Cys Met Ala
  290  295  300
Glu Pro Tyr Ser Pro Glu Gin Ser Ser Pro Asp Ile Ala Ala Arg Ser
  305  310  315  320
Asp Gly Glu Ala Met Leu Lys Leu His Val
  325  330

<210> SEQ ID NO: 34
<211> LENGTH: 1959
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 34
Met Lys Leu Ser Ser Met Ser Leu Leu Asp Ala Glu Asp Val Ala Leu
  1  5  10  15
Thr Ala Ala Ser Pro Asp Thr Ala Leu Ala Leu Asp Trp Ser Arg Ser
  20  25  30
Val Leu Asp Leu Phe Asp Ala Gin Val Ala Leu His Ala Glu Glu Leu
  35  40  45
Ala Cys Ala Asp Gin His Arg Gin Leu Ser Tyr Ala Gin Leu Asp Gin
  50  55  60
His Ala Asn Arg Leu Ala His Cys Leu Ile Glu Arg Gly Leu Arg Pro
  65  70  75  80
Gln Glu Arg Val Ala Leu Trp Phe Gly Arg Ser Pro Asp Phe Leu Ile
  85  90  95
Ala Leu Leu Gly Val Leu Lys Ala Gly Gly Cys Tyr Val Pro Leu Asp
  100  105  110
Pro His Tyr Pro Thr Thr Tyr Ile Gin Gin Ile Leu Asp Asp Ala Gin
  115  120  125
Pro Arg Leu Leu Leu Cys Lys Gly Asp Ile Asp Gly Gin Leu Ile Gin
  130  135  140
Val Pro Arg Leu Arg Leu Asp Ala Ala Ile Ala Arg Gin Pro His
  145  150  155  160
Thr Pro Leu Pro His Ala Leu His Pro Ala Gin Leu Ala Tyr Val Met
  165  170  175
Tyr Thr Ser Gly Ser Thr Gly Arg Pro Lys Gin Met Val Pro His
  180  185  190
Arg Gin Ile Leu Asn Trp His Ala Leu Trp Ala Arg Ala Pro Phe
  195  200  205
Glu Ala Gly Glu Arg Val Ala Gin Lys Thr Ser Ile Ala Phe Ala Ile
  210  215  220
Ser Val Lys Glu Leu Leu Ala Gly Leu Leu Ala Gly Val Pro Gin Val
  225  230  235  240
Phe Ile Asp Glu Asp Thr Val Arg Asp Ile Pro Ala Phe Val Arg Ala
Val Leu Asp Leu Phe Asp Ala Gln Val Ala Leu His Ala Glu Glu
1085 1090 1096
Leu Ala Cys Ala Asp Gln His Arg Glu Leu Ser Tyr Ala Gln Leu
1100 1106 1110
Asp Gln His Ala Asn Arg Leu Ala His Cys Leu Ile Glu Arg Gly
1115 1120 1126
Leu Arg Pro Gln Glu Arg Val Ala Leu Thr Phe Gly Arg Ser Pro
1130 1135 1140
Asp Phe Leu Ile Ala Leu Leu Gly Val Leu Lys Ala Gly Gly Cys
1145 1150 1156
Tyr Val Pro Leu Asp Pro His Tyr Pro Thr Thr Tyr Ile Gln Gln
1160 1165 1170
Ile Leu Asp Asp Ala Glu Pro Arg Leu Leu Leu Cys Gly Lys Asp
1175 1180 1186
Ile Asp Gly Glu Leu Ile Glu Val Pro Arg Leu Arg Leu Asp Glu
1190 1196 1200
Ala Ala Ile Ala Arg Glu Pro His Thr Pro Leu Pro His Ala Leu
1205 1210 1216
His Pro Ala Gln Leu Ala Tyr Val Met Tyr Thr Ser Gly Ser Thr
1220 1225 1230
Gly Arg Pro Lys Gly Val Met Val Pro His Arg Gln Ile Leu Asn
1235 1240 1246
Trp Leu His Ala Leu Thr Ala Arg Ala Pro Phe Glu Ala Gly Lys
1250 1255 1260
Arg Val Ala Gln Lys Thr Ser Ile Ala Phe Ala Ile Ser Val Lys
1265 1270 1276
Glu Leu Leu Ala Gln Leu Ala Gln Val Pro Gln Val Phe Ile
1280 1285 1290
Asp Glu Asp Thr Val Arg Asp Ile Pro Ala Phe Val Arg Ala Leu
1295 1300 1306
Glu Thr Trp Gln Ile Thr Arg Leu Tyr Thr Phe Pro Ser Gln Leu
1310 1315 1320
Asn Ala Leu Leu Asp His Val Ala Glu Thr Pro Gln Arg Leu Ala
1325 1330 1336
Arg Leu Arg Gln Leu Phe Val Ser Ile Glu Pro Cys Pro Ala Glu
1340 1345 1350
Leu Leu Glu Arg Leu Arg Thr Leu Leu Pro Ala Cys Thr Ala Trp
1355 1360 1366
Tyr Ile Tyr Gly Cys Thr Glu Ile Asn Asp Met Thr Tyr Cys Asp
1370 1375 1380
Pro Ala Glu Gln His Ser Gly Ser Gly Phe Val Pro Val Gly Arg
1385 1390 1396
Pro Ile Ala Asn Thr Lys Val His Val Leu Asp Glu Gln Leu Arg
1400 1405 1410
Pro Leu Pro Pro Gly Ile Met Gly Glu Val His Ile Glu Ser Leu
1415 1420 1426
Gly Ile Thr His Gly Tyr Trp Arg Glu Gly Gly Leu Thr Ala Ala
1430 1435 1440
Arg Phe Ile Ala Asn Pro Tyr Gly Pro Pro Gly Ser Arg Leu Tyr
1445 1450 1456
Arg Thr Gly Asp Met Ala Arg Leu Leu Asp Asn Gly Thr Leu Glu
1460 1465 1470
Leu Leu Gly Arg Arg Asp Tyr Glu Val Lys Val Arg Gly Tyr Arg
Leu Leu Ala Gly Phe Glu Pro Gln Cys Leu Gln Pro Asn Ala His 1880 1885 1890
Leu Tyr Gln Ala Gin Thr Ala Val His Val Ser Tyr Ala Asp Met 1895 1900 1905
Ser Lys Pro Arg Gly Gly Ser Glu Val Leu Pro Asp Ile Thr Gly 1910 1915 1920
Tyr Val Pro Leu Ser Gin Ile Lys Ser Ala Ala Gly Asn His Tyr 1925 1930 1935
Ser Met Leu Gin Gly Asp Pro Leu Arg Glu Leu Ala Arg Met Leu 1940 1945 1950
Val Thr Asp Leu Asp Ala 1955

| <210> SEQ ID NO 35 |
|<211> LENGTH: 83 |
|<212> TYPE: PRT |
|<213> ORGANISM: Xanthomonas albelliana |
|<400> SEQUENCE: 35 |

Met Thr Phe Glu Glu Gln Ala Tyr Leu Val Leu Ile Aem Asp Glu Leu 1 5 10 15
Gln Tyr Ser Leu Trp Pro Ser Asp Leu Glu Val Pro Pro Gly Trp Arg 20 25 30
Lys Glu Gly Tyr Ala Gly Gly Lys Asp Glu Cys Met Ala Tyr Ile Asp 35 40 45
Glu Thr Trp Thr Asp Met Arg Pro Leu Ser Leu Arg Glu Leu Asp Asp 50 55 60
Lys Asn Leu Gly Asp Ser Ser Pro Asp Gly Ser Gly Phe Glu Ser 65 70 75 80
Ser Tyr Ser

| <210> SEQ ID NO 36 |
|<211> LENGTH: 315 |
|<212> TYPE: PRT |
|<213> ORGANISM: Xanthomonas albelliana |
|<400> SEQUENCE: 36 |

Met Gly Cys Ala Cys Leu Pro His Tyr Leu Glu Lys Gln Asp Leu Ser 1 5 10 15
Ala Leu Asp Asp Ala Leu Ala Gly Val Arg Leu Ser Gln Tyr Cys Thr 20 25 30
Thr Asp Gly Arg Gln Leu Glu Leu Tyr Trp Leu Gly Ala Gln Ala Ser 35 40 45
Pro Lys Leu Val Leu Leu Pro Pro Tyr Gly Met Ser Tyr Leu Leu Leu 50 55 60
Ser Arg Leu Ala Gin Arg Leu Ala Arg His Phe His Val Leu Cys Thr 65 70 75 80
Glu Ser Ile Gly Cys Pro Asn Ala Gin Thr Ser Val Thr Ala Glu Asp 95 99 100 104
Phe Asp Leu Asp Arg Gin Ala Ala Thr Leu Leu Gly Ile Leu His Gin 106 109 110
His Asp Tyr Ala Asp Cys His Phe Val Gly Trp Cys Gin Ala Ala Gin 115 120 125
Leu Ala Val His Ala Ile Ala Leu His Gly Phe Ala Pro Arg Ser Met 130 135 140
Ala Trp Val Ala Pro Ala Gly Leu Leu Pro Pro Ile Val Lys Ser Glu 145 150 155 160
Fhe Glu Arg Cys Ala Leu Pro Ile Tyr Leu Gin Ile Glu Arg His Gly 165 170 175
Leu Glu Gin Ala Lys Leu Ala Ala Ile Leu Asp Lys Tyr Arg Gly 180 185 190
Gln Pro Leu Arg Gly Asp Asp Leu Ala Glu Lys Leu Thr Met Leu His 195 200 205
Leu Ala Asp Pro Ala Ser Thr Leu Val Phe Ser Arg Tyr Met Arg Ala 210 215 220
Tyr Glu Glu Asn Lys Gin Ser Val Gin Ala Leu Leu Pro Thr Ala Leu 225 230 235 240
Gly Arg His Pro Thr Leu Ile Val His Cys Lys Asp Ser Phe Ser 245 250 255
His Tyr Ser Ala Ser Val Gin Leu Ala Arg His Asp Pro Ser Leu Arg 260 265 270
Leu Asp Leu Leu Asp His Gly His Leu Gin Leu Phe Asn Asp Pro 275 280 285
Gly Ala Val Ala Gln Arg Ile Ile Asp Phe Ile Gly Leu Thr Val Gly 290 295 300
Glu Val Ala Pro Thr Ser Met His Ser Ala Ala 305 310 315

<210> SEQ ID NO 37
<211> LENGTH: 451
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas abilinaneus

<400> SEQUENCE: 37
Met Tyr Ile Pro Asn Asn Ile Asp Leu Asp Pro His Ser Ala Leu Val 1 5 10 15
Arg Gin Leu Thr Ser Tyr Gin Val Arg Phe Leu Gin Trp Trp Arg Leu 20 25 30
Arg Gly Pro Ser Glu Phe His Arg Glu Met Asn Leu Arg Met Pro 35 40 45
Thr Gly Val Lys Gin Ser Glu Thr Trp Thr Arg Tyr His Arg Met Arg 50 55 60
Pro Ser Asp Tyr Arg Trp Gly Val Phe Met Met Pro Pro Asp Arg Asn 65 70 75 80
Thr Val Val Phe Gly Glu Arg Lys Gin Gin Val Ala Trp Ser Cys Val 95 90 95
Pro Glu Gly Tyr Arg Asp Leu Leu Asp His Val Thr Val Gin Gly 100 105 110
Asp Val Gly Asn Ala Ala Val Gin Ser His Gin Leu Thr Gin Met 115 120 125
Val Pro Ser Ala Ile Asp Leu Gin Thr Phe Gin Phe Phe Leu Glu 130 135 140
Glu Gly Arg His Thr Trp Ala Met Ser His Leu Ile Glu Tyr Phe 145 150 155 160
Gly Ser Asp Gly Ala Asp Ala Glu Gly Leu Leu Gin Arg Met Ser 165 170 175
Gly Asp Ala Gin Asn Pro Arg Leu Leu Asp Ala Phe Asn Tyr His Thr 180 185 190
Glu Asp Trp Leu Ser Gin Phe Met Thr Cys Phe Phe Leu Asp Arg Val 195 200 205
Gly Lys Tyr Gln Ile Gln Ala Val Thr Gln Ser Ala Phe Leu Pro Leu 210 215 220
Ala Arg Thr Ala Arg Phe Met Met Phe Glu Glu Pro Leu His Ile Lys 225 230 235 240
Phe Gly Val Asp Gly Leu Glu Arg Val Leu Tyr Arg Ser Ala Glu Ile 245 250 255
Thr Leu Arg Glu Asp Thr His Ala Ile Phe Asp Ala Gly Ala Ile Pro 260 265 270
Leu Pro Val Val Gln Lys Tyr Leu Asn Tyr Trp Leu Pro Lys Ile Phe 275 280 285
Asp Leu Phe Gly His Asp Val Ser Glu Arg Ser Arg Val Leu Tyr Glu 290 295 300
Ala Gly Ile Arg Ser Pro Arg Asp Asp Met Val Glu Gly Thr Glu 305 310 315 320
Val Ala Val Asp Val Arg Cys Glu Asp Arg Leu Val Ser Ser Thr Ala 325 330 335
Pro Ala Glu Leu Ala Ile Asn Ala Val Met Arg Arg Glu Tyr Ile Ala 340 345 350
Glu Val Gly Ala Ile Ile Gly Arg Trp Asn Gln Glu Leu Arg Arg Leu 355 360 365
Gly Leu Ala Phe Glu Leu Glu Leu Pro His Glu Arg Phe His Arg Asp 370 375 380
Phe Gly Pro Cys Lys Gly Leu Ala Phe Asp Leu Arg Gly Asn Pro Val 385 390 395 400
His Asp Ala Asp Gly Gln Arg Leu Ala Ala Leu Pro Thr Pro Gln 405 410 415
Asp Leu Ala Gly Val Arg Gly Gly Leu Glu Gly Glu Gly 420 425 430
Arg Thr Ala Val Trp Leu Ala Pro Ala Gly Ala Ser Leu Asp Lys Leu 435 440 445
Met Pro Ala
450

<210> SEQ ID NO 38
<211> LENGTH: 317
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 38

Met Asp Ser Tyr Val Gly Cys Glu Leu Glu Thr Asp Gly Asp Ala 1 5 10 15
Ser Arg Val Val Pro Met Trp Val Met Tyr Pro Thr Ala Thr Pro Ser 20 25 30
Arg Asp Thr Ala Met Gly Pro Tyr Thr Leu Asp Val Ala Leu Gly Ala 35 40 45
Pro Ile Glu Ala Gly Pro Phe Pro Leu Ala Val Ile Ser His Gly Thr 50 55 60
Arg Ser Ala Gly Leu Val Phe Arg Thr Leu Ala His Tyr Leu Ala Arg 65 70 75 80
His Gly Phe Ile Val Ala Leu Pro Glu His Pro Gly Asp Arg Leu Phe 85 90 95
Gln His Gin Leu Glu Tyr Ser Tyr Gin Asn Leu Glu Asp Arg Pro Arg 100 105 110
His Ile Arg Ala Val Ile Asp Thr Leu Thr Gly His Ala Gin Phe Gly
Pro Ala Ile Gin Ala His Asn Val Ala Val Ile Gly His Ser Val Gly
115 120 125
Gly Tyr Thr Ala Leu Ala Ile Ala Gly Glu Pro His Thr Gly Phe
130 135 140
Met Val Asp Phe Ala His Arg Pro Glu His Ala Glu Gln Pro Ala Trp
145 150 155 160
Thr Ala Leu Val Arg Gin Asn Arg Val Pro Ile Arg Ala Val Pro Val
165 170 175
Thr Ala Asp Pro Arg Val Arg Val Val Ala Leu Ala Pro Asp Phe
180 185 190 195 200 205
Ser Leu Tyr Met His Glu Asp Ala Leu Ala Lys Val Glu Val Pro Val
210 215 220
Leu Leu Ile Val Gly Lys Gin Trp Ala His Glu Thr Ile Val
225 230 235 240
 Ala Thr Arg Thr Ala Leu Gly Asp Gin Arg Leu Glu Ala Arg Val
245 250 255
Val Pro Asn Ala Glu His Tyr Ala Phe Ile Ser Val Phe Pro Glu Ala
260 265 270
Met Lys Ala Arg Val Gly Glu Ala Ala Ile Asp Gin Pro Gly Phe Asp
275 280 285
Arg Ser Ala Phe Gin Arg Glu Leu Glu Arg Asp Ile Leu His Phe Leu
290 295 300
Thr Val Thr Met Arg Pro Ala Glu Ala Ala Ile Ser Gly
305 310 315

<210> SEQ ID NO 39
<211> LENGTH: 496
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 39
Met Gln Lys Pro Lys Glu Ala Leu Gly Met Pro Pro Gly Met Ala Pro
1  5 10 15
Pro Gly Ala Gin Phe Asp Tyr Arg Trp Arg Trp Pro Ala Met Ile Val
20 25 30
Leu Leu Ser Ala Asn Phe Met Asn Leu Asp Val Gly Ile Val Asn
35 40 45
Val Ala Leu Pro Ser Ile Gin Lys Asn Leu Gly Ala Asp Gln Gln Gln
50 55 60
Leu Glu Trp Ile Val Ala Ile Tyr Ile Leu Leu Phe Ala Leu Gly Leu
65 70 75 80
Leu Pro Leu Gly Arg Leu Gly Asp Met Leu Gly Lys Arg Met Phe
85 90 95
Gly Thr Gly Val Ala Gly Phe Ile Leu Met Ser Ala Phe Cys Ala Ile
100 105 110
Ala Gly Asn Ile His Val Leu Ile Ile Ala Arg Ala Leu Gin Gly Leu
115 120 125
Ala Ala Ala Met Leu Ala Pro Gin Val Met Ala Ile Ala Gin Thr Met
130 135 140
Phe Ala Pro Lys Glu Arg Ala Ala Phe Ser Leu Phe Gly Leu Val
145 150 155 160
Ala Gly Leu Ala Ser Phe Ala Gly Pro Leu Val Ser Gly Leu Leu Ile
165 170 175
His Ile Asp Ala Phe Gly Val Gly Trp Arg Ala Ile Phe Leu Ile Arg
180 185 190
Val Pro Ile Gly Leu Val Thr Leu Leu Ala Ala Ala Ala Leu Trp Val Pro
195 200 205
Lys Val Pro Ala His Ala Gly His Asp Trp Val Gly Ile Ala
210 215 220
Leu Ala Ala Leu Ala Leu Leu Cys Leu Val Phe Pro Leu Ile Glu Gly
225 230 235 240
Arg Ala Tyr Gly Trp Pro Leu Trp Cys Phe Ala Ala Ile Ala Leu Gly
245 250 255
Ile Pro Leu Leu Val Ala Phe Val Ala Trp Gin Arg Arg Gin Ala His
260 265 270
Leu Ala Arg Pro Ala Leu Pro Ile Tyr Leu Met Ser His Arg Asp
275 280 285
Tyr Ile Leu Gly Ala Leu Ser Val Ser Val Phe Tyr Ser Ala Leu Gin
290 295 300
Gly Phe Phe Leu Val Phe Val Ile Phe Leu Gin Gin Leu Leu Tyr
305 310 315 320
Ser Ala Leu Glu Thr Gly Val Ala Thr Thr Pro Phe Pro Val Gly Val
325 330 335
Ala Ile Ala Ser Met Leu Ala Arg His Val Glu Ser Leu Arg Ala Lys
340 345 350
Ile Phe Ser Gly Ala Cys Leu Met Ile Ala Ser Tyr Leu Ala Leu Trp
355 360 365
Val Ile Ile Thr Arg Ser Glu Ser Leu Asp Pro Trp Thr Leu Thr
370 375 380
Leu Pro Leu Leu Ile Gly Gly Leu Gly Cys Gly Ile Thr Ile Ala Ser
385 390 395 400
Leu Phe Gln Thr Val Met Arg Thr Val Pro Leu Lys Asp Ala Gin Ala
405 410 415
Gly Ser Gly Ala Leu Gin Val Gin Val Gin Gly Gly Met Leu Gly
420 425 430
Ile Ala Leu Val Ser Glu Ile Phe Phe Ser Gly Leu His Gin His Leu
435 440 445
Gln Gly Pro Ala Gly Val Ala Leu Ala Phe Lys Gin Ala Phe Gly Ala
450 455 460
Thr Val Val Tyr Tyr Ile Ala Asn Ala Phe Val Ala Leu Ser Thr
465 470 475 480
Leu Gly Leu Gln Phe Leu Thr Gin Phe Ala Pro Gin Ser Ser Pro
485 490 495

<210> SEQ ID NO 40
<211> LENGTH: 584
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 40
Met Lys Arg Thr Tyr Ile Gly Leu Ala Asn Ser Phe His Asp Ser Ala
1  5 10 15
Ile Ala Ile Val Gly Asp Gly Gin Val Arg Phe Ala Glu Ala Thr
20 25 30
Glu Arg Tyr Leu Gln Tyr Lys Arg Ser Ile Gly Val Ala Pro Asp Val
35 40 45
Phe Gin Arg Ala Ile Lys Leu Val His Glu Tyr Gly Asp Pro Gly Ala
50 55 60
Glu Leu Val Val Ala Thr Ser Trp Ser Gly Gln Thr Pro Glu Leu Met 65 70 75 80
Arg Glu Gly Leu Gly Lys Thr Ala Gln Ala Val Asp Gln Tyr Arg Ser 95 90 95
Ala Phe Gly Asp Leu Pro Trp His Val Asn Lys Glu Phe Val Ala Gln 100 105 110
 Ser Phe Phe Tyr Arg Ser Gln Leu Ala Met Val Gln His Pro Gly His 115 120 125
Leu Leu Glu Tyr Asp Leu Ser His Met Ala Glu Pro Ala Phe Lys Pro 130 135 140
Pro Ser Tyr Arg His Tyr Glu His His Leu Thr His Ala Val Ala Gly 145 150 155 160
Cys Tyr Thr Ser Pro Phe Glu Glu Ala Val Cys Ala Val Leu Asp Gly 165 170 175
Met Gly Glu Lys Asn Ala Leu Ala Cys Tyr His Tyr Gln Gin Gin Gly Lys 180 185 190
Leu Thr Pro Ile His Gin Ser Glu Thr Ser Ser Ser Trp Ala Ser Leu Gly 195 200 205
Phe Phe Tyr Gly Met Ile Cys Glu Val Cys Gly Phe Gly Thr Leu Ser 210 215 220
Gly Glu Glu Trp Lys Val Met Gly Leu Ala Ala Tyr Gln Gly Gin His Asp 225 230 235 240
Arg Gin Leu Tyr Glu Leu Arg Gin Met Leu Arg Val Asp Gly Leu 245 250 255
Thr Leu Arg Phe Ala Pro Ala Ala Gin Phe Ser Gin Leu Gin Arg Thr 260 265 270
Leu Tyr Ala Met Arg Arg Cys Lys Gly Gin Pro Thr Ile Glu Leu Ala 275 280 285
Asn Leu Ala Tyr Ala Gly Gin Gin Val Phe Cys Asp Val Leu Phe Glu 290 295 300
Phe Leu His Asn Leu His Ala Leu Gly Leu Ser Asp His Leu Val Leu 305 310 315 320
Gly Gly Gly Cys Ala Leu Asn Ser Ser Ala Asn Gin Val Arg Leu Ala 325 330 335
Glu Thr Pro Phe Arg His Leu His Val Phe Ala Ala Pro Gly Asp Asp 340 345 350
Gly Asn Ala Val Gly Ala Leu Trp Ala His Ala Glu Asp His Pro 355 360 365
Glu Gln Thr Pro Pro Ala Arg Glu Gin Ser Pro Tyr Leu Gly Ser 370 375 380
Ser Met Ser Ala Glu Thr Leu His Asn Val Glu Arg Phe Gly Ala Leu 385 390 395 400
Ser Lys Phe Thr Arg Cys Leu Asp Ala Ala Glu Arg Ala Ala Arg 405 410 415
Leu Leu Thr Glu Gly Lys Ile Val Ala Trp Val Gin Gly Arg Ala Glu 420 425 430
Phe Gly Pro Arg Ala Leu Gly Asn Arg Ser Ile Leu Ala Asp Pro Arg 435 440 445
Ser Pro Ala Ile Lys Asp Ile Ile Asn Ala Arg Val Lys Phe Arg Glu 450 455 460
Glu Phe Arg Pro Phe Ala Pro Ser Ile Leu His Glu His Gly Ala Glu 465 470 475 480
Tyr Phe Glu Leu Tyr Gln Glu Ser Pro Tyr Met Glu Arg Thr Leu Lys
485 490 495
Phe Arg Ala Glu Ala Thr Arg Lys Val Pro Gly Val Val His His Asp
500 505 510
Gly Thr Gly Arg Leu Gln Thr Val Lys Gin His Trp Asn Pro Arg Tyr
515 520 525
His Ala Leu Ile Lys Glu Phe Tyr Arg Leu Thr Gly Ile Pro Leu Val
530 535 540
Leu Asn Thr Ser Phe Asn Val Met Gly Lys Pro Ile Ala His Ser Val
545 550 555 560
Glu Asp Ala Leu Ser Ile Phe Thr Ser Gly Leu Asp Ala Met Phe
565 570 575
Ile Asp Asp Val Leu Ile Glu Lys
580

<210> SEQ ID NO 41
<211> LENGTH: 88
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 41

Met Arg Thr Ser Lys Phe Asn Glu Thr Gin Ile Ile Ala Thr Leu Lys
1 5 10 15
Gln Ala Asp Ala Gly Val Pro Val Lys Asp Ile Cys Arg Gin Val Gly
20 25 30
Ile Ser Thr Ala Thr Tyr Gin Trp Lys Ser Lye Tyr Val Ala Ser
35 40 45
Glu Met Pro Ser Ser Arg His Thr Ser Leu Thr Trp Arg Pro Ser
50 55 60
Thr Cys Phe Ser Val Ala Thr Ile Trp Leu Ser Val Val Leu Leu Leu
65 70 75 80
Arg Ile Val Gly Arg Leu Gly Gly
85

<210> SEQ ID NO 42
<211> LENGTH: 716
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 42

Met Arg Cys Leu Ile Ile Asn Asn Tyr Asp Ser Phe Thr Trp Asn Leu
1 5 10 15
Ala Asp Tyr Val Ala Gln Ile Phe Gly Glu Asp Pro Leu Val Val His
20 25 30
Asn Asp Glu Tyr Ser Trp His Glu Leu Lys Asp Arg Gly Gly Phe Ser
35 40 45
Ser Ile Ile Val Ser Pro Gly Pro Gly Ser Val Val Asn Leu Ala Asp
50 55 60
Phe His Ile Ser Leu Gln Ala Leu Glu Gin Asn Glu Phe Pro Val Leu
65 70 75 80
Gly Val Cys Leu Gly Phe Gin Gly Leu Ala His Val Tyr Gly Gly Arg
85 90 95
Ile Leu His Ala Pro Val Pro Phe His Gly Arg Arg Ser Thr Val Ile
100 105 110
Asn Thr Gly Asp Gly Leu Phe Glu Gly Ile Pro Gin Arg Phe Glu Ala
115 120 125
Val Arg Tyr His Ser Leu Met Val Cys Gin Gin Ser Leu Pro Pro Val
130 131 135
Leu Lys Val Thr Ala Arg Thr Asp Cys Gly Val Val Met Gly Leu Gin
145 150 155 160
His Val Gin His Pro Lys Trp Gly Val Gin Phe His Pro Glu Ser Ile
165 170 175
Leu Thr Glu His Gly Lys Arg Ile Val Ala Asn Phe Ala Lys Leu Ala
180 185 190
Ala Arg His Ser Ala Pro Leu Leu Ala Gly Ser Gin Ala Gly Lys
195 200 205
Val Leu Ser Val Cys Ala Pro Glu Met Val Thr Pro Arg Val Arg Arg
210 220
Met Leu Ser Arg Lys Ile Lys Cys Arg Trp Gin Ala Asp Val Phe
225 230 235 240
Leu Ala Leu Phe Ala Asp Glu Lys His Cys Phe Trp Leu Asp Ser Gin
245 250 255
Leu Val Cys Ser Pro Met Ala Arg Tyr Ser Phe Met Gly Ala Val Asn
260 265 270
Glu Ser Glu Val Val Arg His Cys Val Arg Pro Gly Ser Met Val Gin
275 280 285
Glu Ala Gly Glu Arg Phe Leu Ala Glu Met Asp Arg Ala Leu Gin Ser
290 295 300
Val Leu Thr Glu Asp Val Ala Glu Arg Pro Pro Phe Ala Asp Phe Gly
305 310 315 320
Gly Tyr Val Gly Tyr Met Ser Tyr Glu Met Lys Ser Val Phe Gly Ala
325 330 335
Pro Ala Ser His Ala Asn Ala Ile Pro Asp Ala Leu Trp Met Arg Val
340 345 350
Glu Arg Phe Val Ala Phe Asp His Ala Thr Glu Val Glu Val Trp Leu Leu
355 360 365
Ala Leu Ala Asp Thr Glu Asp Leu Ser Ala Leu Ala Trp Leu Asp Ala
370 375 380
Ile Glu Gin Arg Ile His Ala Ile Gly Gin Ala Ala Pro Ala Cys Ile
385 390 395 400
Ser Leu Gly Leu Arg Ser Met Glu Ile Glu Leu Asn His Gly Arg Arg
405 410 415
Gly Tyr Leu Glu Ala Ile Glu Arg Cys Lys Gin Arg Ile Val Asp Gly
420 425 430
Glu Ser Tyr Glu Ile Cys Leu Thr Asp Leu Phe Ser Phe Gin Ala Glu
435 440 445
Leu Asp Pro Leu Met Leu Tyr Met Arg Arg Gly Asn Pro Ala
445 450 455 460
Pro Phe Gly Ala Tyr Leu Arg Asn Gly Ser Asp Cys Ile Leu Ser Thr
465 470 475 480
Ser Pro Glu Arg Phe Leu Glu Val Asp Gly His Gly Thr Ile Gin Thr
485 490 495
Lys Pro Ile Lys Gly Thr Cys Arg Ala Glu Asp Pro Gin Leu Asp
500 505 510
Arg Asn Leu Ala Met Arg Leu Ala Asp Glu Lys Arg Ala Glu
515 520 525
Asn Leu Met Ile Val Asp Leu Met Arg Asn Asp Leu Ser Arg Val Ala
530 535 540
Val Pro Gly Ser Val Thr Val Pro Lys Leu Met Asp Ile Glu Ser Tyr
Lys Thr Val His Gln Met Val Ser Thr Val Glu Ala Arg Leu Arg Ala
545 550 555 560
Asp Cys Ser Leu Val Asp Leu Leu Lys Ala Val Phe Pro Gly Gly Ser
565 570 575
Ile Thr Gly Ala Pro Lys Arg Leu Arg Ser Met Glu Ile Ile Asp Gly Leu
580 585 590
Glu Asn Ala Pro Arg Gly Val Tyr Cys Gly Ser Ile Gly Tyr Leu Gly
595 600 605
Tyr Asn Cys Val Ala Asp Leu Asn Ile Ala Ile Arg Ser Leu Ser Tyr
610 615
Asp Gly Gln Glu Ile Arg Phe Gly Ala Gly Gly Ala Ile Thr Phe Leu
620 625 630 635 640
Ser Asp Pro Gln Asp Glu Phe Asp Glu Val Leu Leu Lys Ala Glu Ala
640 645 650 655
Ile Leu Lys Pro Ile Trp His Tyr Leu His Ala Pro Asn Thr Pro Leu
660 665 670
His Tyr Glu Leu Arg Glu Asp Lys Leu Leu Ala Glu His Cys Val
675 680 685 690 695 700
Ser Glu Met Pro Ala Arg Gln Ala Phe Ile Glu Pro
705 710 715

<210> SEQ ID NO 43
<211> LENGTH: 137
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albineans

<400> SEQUENCE: 43
Met Arg Pro Pro Arg Leu Arg Ala Asn Gln Asp Gly Leu Leu Met Asp
1 5 10 15
Thr Ala Gly Arg Val Val Glu Gly Cys Thr Ser Asn Leu Phe Leu Val
20 25 30
Glu Asn Gly His Leu Val Thr Pro Asp Leu Gly Val Ala Gly Val Ser
35 40 45
Gly Ile Met Arg Gly Arg Val Ile Glu Tyr Gly Arg Gln His Gly Leu
50 55 60
Ala Cys Ala Val Lys His Leu Val Tyr Pro Asp Glu Leu Val Arg Ala Gin
65 70 75 80
Glu Val Phe Leu Thr Asn Ala Val Phe Gly Ile Leu Leu Val Arg Ser
85 90 95
Ile Asp Ala His Ser Tyr Arg Ile Asp Pro Val Thr Leu Arg Leu Leu
100 105 110
Asp Ala Leu Cys Gin Gly Val Tyr Phe Thr Glu Arg Ser Leu His Gin
115 120 125
Val Ser Thr His Ala Gly Gin Asp Pro
130 135

<210> SEQ ID NO 44
<211> LENGTH: 200
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albineans

<400> SEQUENCE: 44
Met Pro Ala Lys Thr Leu Glu Ser Lys Asp Tyr Cys Gly Glu Ser Phe
1 5 10 15
Val Ser Glu Asp Arg Ser Gln Ser Leu Glu Ser Ile Arg Phe Glu
<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>Cys</td>
<td>Thr</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Cys</td>
<td>Lys</td>
<td>Phe</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Ser</td>
<td>Ile</td>
<td>Pro</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>Met</td>
<td>Leu</td>
<td>Gly</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>Glu</td>
<td>Gly</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Tyr</td>
<td>Gly</td>
<td>Leu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Asn</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Ser</td>
<td>Asp</td>
<td>Leu</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Ile</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Ile</td>
<td>Lys</td>
<td>Arg</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Asp</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td>195</td>
<td>200</td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 45
<211> LENGTH: 202
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albilineans
<400> SEQUENCE: 45

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Met| His| Pro| Pro| Ser| Pro| Leu| Asn| Thr| Gln| Gln| Lys| Arg| Thr| Leu| Thr|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Arg| Gly| Gly| Ser| Leu| Thr| Ala| His| Leu| Arg| Leu| Leu| Gly| Gln| Val| Gln|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Val| Gln| Val| Gln| Arg| Gly| Lys| Ser| Met| Ala| Trp| Leu| Asp| Glu| Tyr|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Arg| Val| Leu| Gly| Leu| Ser| Arg| Cys| Leu| Val| Trp| Val| Arg| Glu| Val|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Val| Leu| Val| Val| Asp| Ala| Lys| Pro| Tyr| Val| Tyr| Ala| Arg| Ser| Leu| Thr|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Pro| Leu| Thr| Ala| Ser| Tyr| Asn| Ala| Trp| Gln| Ala| Val| Arg| Ser| Ile| Gly|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Ser| Arg| Pro| Leu| Ala| Asp| Leu| Phe| Arg| Asp| Arg| Ser| Val| Leu| Arg|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Ser| Ala| Leu| Ala| Ser| Arg| Arg| Ile| Thr| Ala| Gln| His| Pro| Leu| His| Arg|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Arg| Ala| Cys| Asn| Phe| Val| Ala| Gln| Ser| His| Ala| Thr| Gln| Ala| Leu| Leu|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Ala| Arg| Ser| Val| Phe| Thr| Arg| Gln| Gly| Ala| Pro| Leu| Leu| Ile| Thr|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Glu| Cys| Met| Leu| Pro| Ala| Leu| Thr| Leu| Glu| Pro| Val| Ala| Ala|    |    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| Pro| Arg| Gln| Ala| Ser| Leu| Ser| Ala| Asp| Gly| Pro| Cys| Arg| His| Ser| Ala|    |
|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
Gln Ile Val Ser Pro Glu Ser Met Leu Glu
195
200

<210> SEQ ID NO 46
<211> LENGTH: 278
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albineans

<400> SEQUENCE: 46
Met Pro Asn Ala Val Pro Met Glu Gly Ala Arg Gly Leu Pro Glu Pro
1 5 10 15
Gln Ala Met Asn Pro Gly Leu Pro Ser Val Gly Gln Ser Ala Gly
20 25 30
Gln Pro Leu Gln Leu Ser Ala Leu Pro Leu Gln Ala Ala Ala Arg
30 35 40 45
Ser Ala His Arg His Leu Leu Leu Asp Gly Thr Ala Leu Tyr Leu Leu
50 55 60
Ala Phe Asp Thr Ala Gln Phe Asp Pro Gly Ala Phe Ala Ala Met Ala
65 70 75 80
Ile Ala Arg Pro Asp Ser Ile Ala Arg Ser Val Arg Lys Arg Gln Ala
85 90 95
Glu Phe Leu Phe Gly Arg Leu Ala Ala Arg Leu Ala Leu Glu Glu Val
100 105 110
Leu Gly Pro Ala Gln Ala Ala Ala Asp Ile Ala Ile Gly Ala Thr Arg
115 120 125
Ala Pro Cys Trp Pro Ala Gly Ser Leu Gly Ser Ile Ser His Cys Glu
130 135 140
Asp Tyr Ala Ala Ala Ile Ala Met Ala Ala Gly Thr Arg His Gln Val
145 150 155 160
Gly Ile Asp Leu Gln Arg Pro Ile Thr Pro Ala Ala Arg Ala Ala Leu
165 170 175
Leu Ser Ile Ala Ile Asp Ala Asp Ala Ala Arg Leu Ala Lys Ala
180 185 190
Ala Asp Ala Glu Trp Pro Gln Asp Leu Leu Thr Ala Leu Phe Ser
195 200 205
Ala Lys Glu Ser Leu Phe Lys Ala Ala Tyr Ser Ala Val Gly Arg Tyr
210 215 220
Phe Asp Phe Ser Ala Arg Leu Cys Gly Ile Asp Leu Ala Arg Gln
225 230 235 240
Cys Leu His Leu Arg Leu Thr Glu Thr Leu Cys Ala Glu Phe Val Ala
245 250 255
Gly Gln Val Cys Glu Val Gly Phe Ala Arg Leu Pro Pro Asp Leu Val
260 265 270
Leu Thr His Tyr Ala Trp
275

<210> SEQ ID NO 47
<211> LENGTH: 634
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas albineans

<400> SEQUENCE: 47
Met Ser Val Glu Thr Gln Lys Gln Thr Leu Gly Phe Gln Thr Glu Val
1 5 10 15
Lys Gln Leu Leu Gln Leu Met Ile His Ser Leu Tyr Ser Asn Lys Glu
20 25 30
-continued

Ile Phe Leu Arg Glu Leu Ile Ser Asn Ala Ser Asp Ala Ala Asp Lys 35 40 45
Leu Arg Phe Glu Ala Leu Val Lys Pro Glu Leu Leu Asp Gly Asp Ala 50 55 60
Gln Leu Arg Ile Arg Ile Gly Phe Asp Lys Ala Gly Thr Val Thr 45 70 75 80
Ile Asp Asp Asn Gly Ile Gly Met Ser Arg Glu Ile Val Ala His 85 90 95
Leu Gly Thr Ile Ala Lys Ser Gly Thr Ser Asp Phe Leu Lys His Leu 100 105 110
Ser Gly Asp Glu Lys Asp Ser His Leu Ile Gly Gin Phe Gly Val 120 125
Gly Phe Tyr Ser Ala Phe Ile Val Ala Asp Gin Val Asp Val Tyr Ser 130 135 140
Arg Arg Ala Gly Leu Pro Ala Ser Asp Gly Val His Thr Ser Arg 145 150 155 160
Gly Glu Gly Glu Phe Glu Val Ala Thr Ile Asp Lys Pro Glu Arg Gly 165 170 175
Thr Arg Ile Val Leu His Leu Lys Glu Glu Glu Gly Phe Asp Ala 180 185 190
Gly Trp Lys Leu Arg Ser Ile Val Arg Lys Tyr Ser Asp His Ile Ala 195 200 205
Leu Pro Ile Glu Leu Ile Lys His Tyr Gly Glu Asp Lys Asp Lys 210 215 220
Pro Glu Thr Pro Glu Trp Gln Thr Val Asn Arg Ala Ser Ala Leu Trp 225 230 235 240
Thr Arg Pro Arg Thr Glu Ile Lys Asp Glu Glu Tyr Gln Glu Leu Tyr 245 250 255
Lys His Ile Ala His Asp His Glu Asn Pro Val Ala Trp Ser His Asn 260 265 270
Lys Val Glu Gly Lys Leu Glu Tyr Thr Ser Leu Leu Tyr Leu Pro Gly 275 280 285
Arg Ala Pro Phe Asp Leu Tyr Gin Arg Asp Ala Ser Arg Gly Leu Lys 290 295 300
Leu Tyr Val Gin Arg Val Phe Ile Met Asp Gin Ala Asp Gin Phe Leu 305 310 315 320
Pro Leu Tyr Leu Arg Phe Ile Lys Gly Ile Val Asp Ser Ser Asp Leu 325 330 335
Pro Leu Asn Val Ser Arg Glu Ile Leu Gin Ser Gly Pro Val Ile Asp 340 345 350
Ser Met Lys Ser Ala Leu Thr Lys Arg Ala Leu Asp Met Leu Glu Lys 345 350 355 360 365
Leu Ala Lys Asp Asp Pro Glu Arg Tyr Lys Gly Val Trp Lys Asn Phe 370 375 380
Gly Gin Val Leu Lys Glu Gly Pro Ala Gin Asp Phe Gly Asn Arg Glu 385 390 395 400
Lys Ile Ala Gly Leu Arg Phe Ala Ser Thr His Ser Gly Asp Asp 405 410 415
Ala Gin Asn Val Ser Leu Ala Asp Tyr Val Ala Arg Met Lys Asp Gly 420 425 430
Gln Asp Lys Leu Tyr Thr Leu Gly Glu Ser Tyr Ala Gin Ile Lys 435 440 445
Asp Ser Pro His Leu Glu Val Phe Arg Lys Lys Gly Ile Glu Val Leu
Leu Leu Thr Asp Arg Ile Asp Glu Trp Leu Met Ser Tyr Leu Thr Glu
450 455 460
Phe Asp Ser Lys Ser Phe Val Asp Val Ala Arg Gly Asp Leu Asp Leu
465 470 475 480
Gly Lys Leu Asp Ser Glu Glu Lys Gin Ala Gin Glu Glu Ala Ala
485 490 505 510
Lys Ala Lys Gin Gly Leu Ala Glu Arg Ile Gin Gin Val Leu Lys Asp
515 520 525
Glu Val Ala Glu Val Arg Val Ser His Arg Leu Thr Asp Ser Pro Ala
530 535 540
Ile Leu Ala Ile Gly Gin Gly Asp Met Gin Leu Gin Met Arg Gin Ile
545 550 555 560
Leu Glu Ala Ser Gin Gin Leu Pro Gin Gin Ser Gin Gin Lys Gin Arg
565 570 575
Phe Asn Pro Ala His Pro Leu Ile Gin Gin Leu Gin Leu Gin Gin Gin
580 585 590
Val Asp Arg Phe Gly Asp Leu Leu Val Leu Phe Asp Gin Glu Ala Ala
595 600 605
Leu Ala Ala Gly Asp Ser Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin
610 615 620
Leu Asn Lys Leu Leu Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin
625 630

<210> SEQ ID NO 48
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 48

gcctccagg gcctccagg 20

<210> SEQ ID NO 49
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 49

gctgaaacc ggaatctga 20

<210> SEQ ID NO 50
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 50

gcgcagcgg gcgcagcgg 20

<210> SEQ ID NO 51
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Xanthomonas albilineans

<400> SEQUENCE: 51

acgcagcgg gcgcagcgg 20

<210> SEQ ID NO 52
<211> LENGTH: 19
<212> TYPE: DNA
We claim:

1. A transformed isolated host cell that comprises one or more genetic constructs that comprises SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3, wherein said transformed host cell produces albicidin.

2. The transformed isolated host cell of claim 1 wherein said genetic construct contains a combination of SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3.

3. The transformed isolated host cell claim 1 wherein said host cell has been transformed with a first genetic construct comprising SEQ ID NO: 1, a second genetic construct comprising SEQ ID NO: 2 and a third genetic construct comprising SEQ ID NO: 3.

4. A transformed isolated host cell comprising one or more polynucleotides encoding a biosynthetic pathway for albicidin production, said polynucleotides encoding SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 39, 40, 41, 42, 43, 44, 45, 46 and 47 and wherein said host cell produces albicidin.

5. A combination of isolated polynucleotides that encodes SEQ ID NO: 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 39, 40, 41, 42, 43, 44, 45, 46 and 47.

6. The combination of isolated polynucleotides according to claim 5 wherein said combination of polynucleotides comprises SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3.

7. A method of making an antibiotic comprising culturing a transformed host cell according to claim 4 under conditions that allow for the production of said antibiotic.

8. The method according to claim 7, further comprising the isolation of said antibiotic.
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 7,510,852 B2
APPLICATION NO. : 10/531351
DATED : March 31, 2009
INVENTOR(S) : Monique Royer et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 4,
Line 23, “P14687 SEQ ID NO: 132” should read --P14687 (SEQ ID NO: 132)--.
Line 56, “RitE-1” should read --RifE-1--.

Column 10,

Column 16,

Column 22,
Line 21, “protecting a plant; against” should read --protecting a plant against--.
Line 40, “kanamycin” should read --kanamycin--.

Column 26,
Line 31, “(5'gccccagccgctgcgagg3')” should read --(5'tgccccacaggccgctgcgagg3')--.

Column 27,
Line 4, “SEQ D No: 6” should read --SEQ ID No: 6--.
Line 53, “J-ketoacyl synthase” should read --β-ketoacyl synthase--.

Column 29,
Line 22, “albXV” should read --albXI--.
Line 60, “motif involved” should read --motif I involved--.

Column 31,
Line 15, “AlbXVI” should read --AlbXVII--.
Line 20, “albXVII” should read --albXVIII--.
Line 21, “AlbXIII” should read --AlbXVIII--.
Lines 35-36, “(5'cgtgagagtgcagcgcgctg3')” should read --(5'cgtagagagtgcagcgcgctg3')--.

Column 34,
Line 7, “from Comamonas” should read --from Comamonas--.
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

**Column 36.**
Line 9, “BLm (& Ala)” should read --B1m (β-Ala)--.
Line 27, “confined” should read --confirmed--.
Line 58, “from albX” should read --from albXIX--.

**Column 40.**
Lines 63-64, “and XhoSalXaHT1PGF 5’cgatatgccttctgcagaggtgataage3’” should read --and XhoSalXaHTPGF 5’cgatatgccttctgcagaggtgataage3’--.

**Column 41.**

**Column 42.**
Line 29, “Because albXVII” should read --Because albXVIII--.
Line 46, “confining” should read --confirming--.

**Column 43.**
Line 51, “stain Xa13” should read --strain Xa13--.
Line 64, “and albIV)” should read --and albIX)--.

**Column 45.**
Line 47, “PKS4 module” should read --PKS4 module--.

**Column 47.**
Line 58, “AbVII HBCL” should read --AlbVII HBCL--.

**Column 49.**
Line 14, “identical to albX)” should read --identical to albXXI)--.

**Column 51.**
Table 1, row “pBC/f”, “2.5 kb Kpn I-EdoR I” should read --2.5 kb Kpn I-EcoR I--.

**Column 53.**
Table 1, row “pBKS/XALB3Xhol”, “pBKS/XALB3 with a Xhol site” should read --pBKS/XALB3 with a Xhol site--.
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 60.
Table 4, row “AlbXIII”, Columns “Identities”, “Positives”, “Gaps”,
    “43/156  56/156  44/156
    (28%)”
should read  --43/156  56/156  44/156 (28%)--.

Column 66.
Table 10, row “Tetracycline”,
    “DH5αKT       should read  --DH5αKT
    DHSαAlb’KT”       DH5αAlb’KT--.

Column 264.
Line 26, “36, 37, 39, 39, 40” should read --36, 37, 38, 39, 40--.
Lines 30-31, “36, 37, 39, 39, 40” should read --36, 37, 38, 39, 40--.

Signed and Sealed this
Twenty-third Day of June, 2009

John Doll
Acting Director of the United States Patent and Trademark Office