



(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2008 054 660.7**

(22) Anmeldetag: **15.12.2008**

(43) Offenlegungstag: **21.10.2010**

(51) Int Cl.⁸: **C12N 15/31** (2006.01)

C07K 14/195 (2006.01)

C12N 1/00 (2006.01)

(71) Anmelder:

**Rheinische Friedrich-Wilhelms-Universität Bonn,
53113 Bonn, DE**

(74) Vertreter:

**Michalski Hüttermann & Partner Patentanwälte,
40221 Düsseldorf**

(72) Erfinder:

**Piel, Jörn, 53173 Bonn, DE; Fisch, Katja, 53121
Bonn, DE; Gurgui, Christian, 53127 Bonn, DE**

(56) Für die Beurteilung der Patentfähigkeit in Betracht
zu ziehende Druckschriften:

US 68 33 447 B1

**SCHIRMER, A., et al.: Metagenomic analysis
reveals diverse polyketide synthase gene
clusters in microorganisms associated with
marine sponge, 2005. In: Appl. and
Environmental Microbiology, Vol. 71,
S. 4740-4849 Zusammenf., ges. Dokument**

**Internet-Recherche am 05.11.2009.
www.ncbi.nlm.nih.gov. Uncultured bacterium**

**clone spF-KS44i modular polyketide synthase
gene Acc.Nr. AY897143 Zusammenf.,
ges. Dokument**

Internet-Recherche am 05.11.2009:

**www.ncbi.nlm.nih.gov. Sequenzvergleich der
Sequenz SEQ ID NO: 4 mit AY897143
Zusammenf., ges. Dokument**

Internet-Recherche am 05.11.2009:

**www.ncbi.nlm.nih.gov. Myxococcus xanthus
DK 1622, ges. Geno., Acc.Nr.: CP000113
Zusammenf., ges. Dokument**

Internet-Recherche am 05.11.2009:

**www.ncbi.nlm.nih.gov. Sequenzvergleich der
Sequenz SEQ ID NO: 23 mit CP000113
Zusammenf., ges. Dokument**

Internet-Recherche am 09.11.2009:

**www.ncbi.nlm.nih.gov. Sequenz 12902 aus dem
Patent US 6833447 GI-Nr.: 59743305
Zusammenf., ges. Dokument**

Internet-Recherche am 09.11.2009:

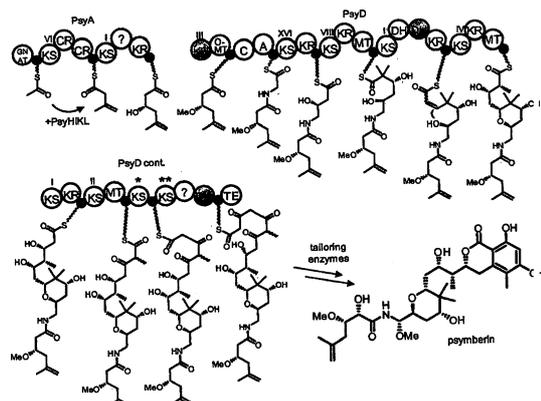
**www.ncbi.nlm.nih.gov. Sequenzvergleich der
Sequenz SEQ ID NO: 23 mit 59743305
Zusammenf., ges. Dokument**

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Rechercheantrag gemäß § 43 Abs. 1 Satz 1 PatG ist gestellt.

(54) Bezeichnung: **Polyketid Biosynthesegene**

(57) Zusammenfassung: Die vorliegende Erfindung stellt ein neues Verfahren zur gezielten Isolierung von Polyketidbiosynthesegenen aus Metagenomen bereit. Weiterhin werden erstmalig die Nuklein- und Aminosäuresequenzen des PKS-Clusters von Psymberin, sowie Expressionssysteme zur Herstellung dieser Polyketids offenbart.



Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Isolierung von Gensequenzen, sowie die mittels dieses Verfahrens isolierte Gensequenz und Fragmente derselben, die für Peptide zur Biosynthese von Polyketiden codiert.

[0002] Schwämme (Porifera) stellen die wichtigste Quelle mariner Produkte mit therapeutischem Potential dar (Blunt et al., Marine natural products. Nat. Prod. Rep. 25, 35–94; 2008). Viele der aus Schwämmen isolierten Wirkstoffkandidaten sind komplexe Polyketide (PK), die auf metabolischer Ebene eine enge Beziehung zu natürlichen Produkten aus Bakterien aufweisen (Sipkema et al., Marine sponges as pharmacy. Mar. Biotechnol. 7, 142–144; 2005). Diese enge Beziehung führte zu der Vermutung, dass die Polyketide nicht von den Schwämmen selbst, sondern von prokaryotischen Symbionten erzeugt werden. Untersuchungen am Schwamm *Theonella swinhoei* deuten darauf hin, dass diese Vermutung richtig ist (fiel et al., Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc. Natl. Acad. Sci. U.S.A. 101, 16222–16227; 2004). Ein Schwamm kann mehrere hundert Arten prokaryotischer Symbionten beherbergen, von denen jede eigene Polyketidsynthase (PKS) Gene für verschiedene Biosynthesewege bereitstellt. Das so entstehende Metagenom eines Schwamms kann daher hunderte von individuellen (prokaryotischen) Genomen enthalten, von denen jedes seine eigenen PKS Gene bereitstellt (Kim et al., Diversity of polyketide synthase genes from bacteria associated with the marine sponge *Pseudoceratina clavata*: culture-dependent and culture-independent approaches. Environ. Microbiol. 8, 1460–1470; 2006).

[0003] Eine direkte Kultivierung einzelner Symbionten führte in den meisten Fällen nicht zum Erfolg, womit die Isolierung eines bestimmten, interessierenden PKS Genclusters für die Biosynthese von Polyketiden aus einem solchen Metagenom ein großes Problem darstellt. Unter anderem wird nämlich z. B. die Herstellung und Durchmusterung von extrem großen DNA-Bibliotheken erforderlich und die Identifizierung des korrekten PKS Genclusters stellt eine große Herausforderung dar.

[0004] Die vorliegende Erfindung stellt eine Lösung zu diesem Problem bereit und offenbart einen neuen Ansatz zur spezifischen Amplifikation von PKS-Genregionen, der auf der Substratspezifität von Kethosynthasedomänen beruht.

[0005] In einem ersten Aspekt bezieht sich die vorliegende Erfindung auf eine Nukleinsäure umfassend zumindest ein Biosynthesegen für ein Polyketid ausgewählt aus der Gruppe bestehend aus Psymberin oder Mycalamid. In einer spezifischen Ausführungsform umfasst die Nukleinsäure eine Sequenz umfasst die ausgewählt ist aus der Gruppe bestehend aus SEQ ID NO 1–19, 39, 75, einem Homolog dieser Sequenzen und dem Komplement der Sequenzen oder des Homologs, wobei das Homolog zumindest 75% Sequenzidentität aufweist, oder unter Hochsalzbedingungen mit der Nukleinsäure hybridisiert.

[0006] In einem weiteren Aspekt bezieht sich die vorliegende Erfindung auf ein Peptid für das eine der erfindungsgemäßen Nukleinsäuren codiert. In einer spezifischen Ausführungsform weist das Peptid eine Aminosäuresequenz auf, die ausgewählt ist aus der Gruppe bestehend aus SEQ ID NO 20–38 oder einem Homolog dieser Sequenzen, wobei das Homolog zumindest eine Sequenzidentität Identität zu SEQ ID NO 20 von 36%, zu SEQ ID NO 21 von 54%, zu SEQ ID NO 22 von 44%, zu SEQ ID NO 23 von 41%, zu SEQ ID NO 24 von 52%, zu SEQ ID NO 25 von 42%, zu SEQ ID NO 28 von 44%, zu SEQ ID NO 29 von 94%, zu SEQ ID NO 30 von 96%, zu SEQ ID NO 31 von 96%, zu SEQ ID NO 32 von 94%, zu SEQ ID NO 33 von 98%, zu SEQ ID NO 34 von 95%, zu SEQ ID NO 35 von 97%, zu SEQ ID NO 36 von 98%, zu SEQ ID NO 37 von 96% oder zu SEQ ID NO 38 von 96% aufweist.

[0007] In einem weiteren Aspekt bezieht sich die vorliegende Erfindung auf einen die erfindungsgemäßen Nukleinsäuren umfassenden Vektor sowie eine rekombinante Wirtszelle, die die erfindungsgemäßen Nukleinsäuren oder den erfindungsgemäßen Vektor umfasst.

[0008] In einem weiteren Aspekt bezieht sich die vorliegende Erfindung auf ein Verfahren zur Herstellung eines Polyketids umfassend den Schritt der Kultivierung einer erfindungsgemäßen rekombinanten Wirtszelle unter Bedingungen, die die Expression der Biosynthesegene und Produktion des Polyketids durch die Wirtszelle erlauben.

[0009] In einem weiteren Aspekt bezieht sich die vorliegende Erfindung auf ein Verfahren zur Isolierung zumindest eines Polyketid-Biosynthesegens, umfassend die Schritte: a) Bereitstellen von zumindest einem Primerpaar, das auf der Substratspezifität einer Kethosynthasedomäne einer trans-AT Polyketidsynthase basiert; und b) Isolierung von zumindest einem Polyketid-Biosynthesegen unter Verwendung des zumindest einen Pri-

merpaars aus Schritt a). In einer spezifischen Ausführungsform umfasst Schritt b) die Amplifikation der vom Primerpaar eingeschlossenen Nukleinsäureregion. In weiteren spezifischen Ausführungsformen ist das zumindest eine Primerpaar spezifisch für ein Motiv das bei phylogenetisch benachbarten Kethosynthesedomänen von trans-AT Polyketidsynthasen konserviert ist, oder ist zumindest ein Primer des Primerpaars spezifisch für zumindest 4 zusammenhängende Aminosäurereste eines der in [Fig. 3](#), [Fig. 4](#) oder SEQ ID NO: 40–71 definierten Aminosäuremotive. In einer weiteren spezifischen Ausführungsform der Erfindung ist das isolierte Polyketid-Biosynthesegen in die Biosynthese von Psymberin oder Mycalamid involviert.

[0010] [Fig. 1](#) zeigt eine Übersicht über einige beispielhaften und mit Ausnahme von Pederin aus Schwämmen isolierte komplexe Polyketide, die von Symbionten synthetisiert werden oder von denen vermutet wird, dass sie von Symbionten synthetisiert werden. 1 Psymberin (Irciniastatin A), 2 Onnamide A, 3 Pederin, 4 Mycalamid A, 5 Laulimalid (Fijianolid B), 6 Latrunculin A, 7 Mycothiazol, 8 Pateamin A, 9 Spongistatin 1 (Altohyrtin A). Das Sternzeichen (*) gibt die Position der β -Verzweigungen an.

[0011] [Fig. 2](#) ist ein bayesisches Kladogramm von Vollängen-KS-Domänen aus trans-AT-PKS. Die KS-Nummerierung bezieht sich auf die Position innerhalb des Genclusters, wobei vom 7'tream-Ende her gestartet wird. Beispielsweise ist LkcKS3 die dritte Lankacidin KS Domäne. Die cis-AT KS4 der Erythromycin PKS wurde als Outgroup verwendet. Wahrscheinlichkeitswerte > 0.6 werden an den Knotenpunkten angegeben. Kladentypen werden in römischer Nummerierung zusammen mit dem Haupt-Substrattyp angegeben. Bae, Bacillaene; BT, nicht charakterisierte PKS aus *B. thailandensis*; Chi, Chivosazol; Dif, Difficidin; Dsz, Disorazol; GU, nicht charakterisierte PKS aus *Geobacter uraniumreducens*; Lkc, Lankacidin; Lnm, Leinamycin; Mln, Macrolactin; Mmp, Mupirocin; Onn, Onnamide; Ped, Pederin; Ta, Myxovirescin.

[0012] [Fig. 3](#) zeigt das Sequenzalignment von KS Protein Sequenzen einzelner Kladen. Die Nummerierung bezieht sich auf die Aminosäurepositionen von Ped1, welche PedKS1 beinhaltet. Die Pfeile zeigen Regionen an, zu denen erfindungsgemäß Primer hergestellt wurden, die zur Amplifikation der PKS Genfragmente genutzt wurden. Kladen-spezifische Aminosäurereste innerhalb dieser Regionen sind in Fettschrift wiedergegeben. Das Sternzeichen markiert das konservierte Cystein der active-site. V, Sequenzen der Klade V; VI, Sequenzen der Klade VIb; S, Sequenzen von SupA; BP17KS1 stammt aus *Burkholderia pseudomallei* 1710b (BURPS1710b_A2618, erste KS Domäne); KAKS1 stammt aus *Kordia algicida* (KAOT1_04270, erste KS); SupA.Ts1 stammt aus *Theonella swinhoei* (GenBank accession ABE03935, erste KS); SupA.Aa1 stammt aus *Aplysina aerophoba* (ABE03915, erste KS); SupA.Aa2 stammt aus *A. aeophoba* (ABE03895, erste KS); SupA.Pc1 stammt aus *Pseudoceratina clavata* (ABB73286, erste KS); SupA.PKSA stammt aus *Discodermia dissoluta* (SA1_PKSA, zweite KS).

[0013] [Fig. 4A–Fig. 4I](#) zeigt ein Sequenzalignment der kladenspezifischen Motive, die für ein erfindungsgemäßes gruppenspezifisches Primerdesign geeignet sind. Die gruppenspezifischen Aminosäuren sind als weiße Buchstaben vor schwarzem Hintergrund dargestellt; allgemein konservierte Aminosäuren sind grau unterlegt (80% threshold) und Aminosäuren, die in einigen, jedoch nicht allen Gruppen konserviert sind, sind in Fettdruck dargestellt. Die Aufteilung in die Kladen basiert auf einer phylogenetischen Analyse nach Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C. and Piel J. *Nat. Biotechnol.*, 2008, 26, 225–233 (siehe [Fig. 2](#)). supA1: ABK01346.1, supA2: ABK01347.1 und supA4: ABK01355.1 wurden ebenfalls zu dem Sequenzalignment hinzugefügt, um ubiquitäre PKS Sequenzen aus Schwämmen zu vermeiden.

[0014] [Fig. 5A](#) zeigt die genomische Organisation des isolierten PKS-Genclusters für die Psymberinbiosynthese. [Fig. 5B](#) stellt schematisch die Struktur der identifizierten PsyA und PsyD Proteine dar, wobei die einzelnen Module sowie die daran ablaufenden biochemischen Umwandlungen verdeutlicht werden. [Fig. 5C](#) stellt vergleichend die Strukturen der entsprechenden Proteine der PKS für die Pederin- und Onnamidbiosynthese gegenüber.

[0015] Die vorliegende Erfindung stellt ein neues Verfahren zur gezielten Isolierung von Polyketidbiosynthesegenen aus Metagenomen bereit, wobei sich die Substratspezifität der Polyketidketten-verlängernden Ketosynthase (KS) Domänen zunutze gemacht wird.

Definitionen

[0016] Unter dem Term „Nukleinsäure“ im Sinne der vorliegenden Erfindung wird insbesondere – aber nicht darauf beschränkt – natürliche, vorzugsweise lineare, verzweigte oder zirkuläre Nukleinsäuren wie RNA, insbesondere mRNA, einzelsträngige und doppelsträngige virale RNA, siRNA, miRNA, snRNA, tRNA, hnRNA

oder Ribozyme, genomische, bakterielle oder virale DNA (einzelsträngig und doppelsträngig), chromosomale und episomale DNA, frei zirkulierende Nukleinsäure und dergleichen, synthetische oder modifizierte Nukleinsäuren, beispielsweise Plasmide oder Oligonukleotide, insbesondere für die PCR verwendete Primer, Sonden oder Standards, mit Digoxigenin, Biotin oder Fluoreszenzfarbstoffen markierte Nukleinsäuren oder sogenannte PNAs („peptide nucleic acids“) verstanden. Eine Nukleinsäure besteht häufig aus zwei komplementären Strängen. Daher soll bei einer durch diese Erfindung beanspruchten Nukleinsäure generell auch immer das reverse Komplement dieser Nukleinsäure als mitbeansprucht gelten.

[0017] Unter dem Term „Hochsalzpuffer“ wird insbesondere – aber nicht darauf beschränkt – ein Puffer aufweisend eine hohe Salzkonzentration (bevorzugt chaotrope Substanzen), bevorzugt ≥ 100 mM, bevorzugter ≥ 500 mM und noch bevorzugter ≥ 1 M, verstanden.

[0018] Unter dem Term „Hochsalzbedingungen“ wird im Folgenden ein Milieu verstanden, das einen Hochsalzpuffer verwendet, bevorzugt einen Hochsalzpuffer enthaltend chaotrope Salze. Durch Hochsalz, bevorzugt aufweisend chaotrope Salze, wird die Löslichkeit von Nukleinsäuren in Wasser herabgesetzt. Grund ist das Aufbrechen von Wasserstoffbrücken und damit verbunden eine Verringerung der Stabilisierung von Sekundär- und Tertiär-Strukturen der Nukleinsäuren in Wasser. Wird nun eine polare Oberfläche als Wasserstoffbrückendonator angeboten, binden die Nukleinsäuren an dieser Oberfläche, da sie dort eine bessere Stabilisierung erfahren als in Wasser. Wird die Salzkonzentration verringert, wird Wasser wieder ein besserer Wasserstoffbrückendonator als die polare Oberfläche, und die Nukleinsäuren lassen sich wieder von der Oberfläche ablösen.

[0019] Unter dem Term „chaotrope Substanzen“ bzw. „chaotrope Salze“ werden insbesondere – aber nicht darauf beschränkt – Substanzen, die die Sekundär-, Tertiär- und/oder Quartärstruktur von Proteinen und/oder Nukleinsäuren ändern und zumindest die Primärstruktur intakt lassen, die Löslichkeit polarer Substanzen in Wasser verringern und/oder hydrophobe Wechselwirkungen verstärken, verstanden. Bevorzugte chaotrope Substanzen sind Guanidinhydrochlorid, Guanidinium(iso)thiocyanat, Natriumiodid, Natriumperchlorat, Kaliumiodid, Natrium(iso)thiocyanat und/oder Harnstoff.

[0020] Die Begriffe „Protein“, „Peptid“, „Polypeptid“ und „Enzym“ werden synonym verwendet, sofern nicht Gegenteiliges beschrieben wird.

[0021] Unter dem Begriff „Amplifikation“ oder „Amplifikationsreaktion“ wird ein Verfahren verstanden, das es ermöglicht, die Konzentration eines oder mehrerer Analyte – bevorzugt Nukleinsäuren – mindestens zu verdoppeln. Man unterscheidet hier zwischen isothermen und thermocyclischen Amplifikationsreaktionen. Bei ersteren bleibt die Temperatur während des gesamten Verfahren stets gleich, während bei letzterer thermische Zyklen durchlaufen werden, mit deren Hilfe die Reaktion und die Amplifikation gesteuert wird.

[0022] Bevorzugte isotherme Amplifikationsreaktionen sind z. B.

- Loop mediated isothermal amplification (LAMP),
- Nucleic Acid Sequence Based Amplification (NASBA),
- Rolling Circle Chain Reaction (RCCR), oder Rolling Circle Amplification (RCA), und/oder
- Transcription mediated amplification (TMA) Bevorzugte thermocyclische Amplifikationsreaktionen sind z. B.
- Ligase-Kettenreaktion (LCR), und/oder
- Polymerase Kettenreaktion (PCR)

[0023] Unter dem Term „Polymerase Kettenreaktion“ (PCR) wird ein Verfahren zur in vitro-Vervielfältigung von Nukleinsäuren verstanden, wie es z. B. in Bartlett & Stirling (2003) beschrieben ist.

[0024] Unter dem Term „Ligase-Kettenreaktion“ (LCR) wird ein Nachweisverfahren für geringste Mengen von Nukleinsäuren verstanden, das ähnlich wie die Polymerase-Kettenreaktion funktioniert, nur unter Verwendung eines anderen Enzyms (anstatt der Polymerase eine Ligase). Zwei Proben pro DNA-Strang werden zu einer Probe ligiert. Die entstehenden, oft nur 30–50 bp langen Amplifikate eines Zyklus dienen in den folgenden Zyklen selbst wieder als Ansatzpunkt für die supplementierten Primer.

[0025] Unter dem Term „Loop-mediated Isothermal Amplification“ (LAMP) wird eine Methode zur isothermalen Nukleinsäureamplifikation verstanden, bei welcher 6 verschiedene Primer eingesetzt werden, die bestimmte Regionen auf der Target-Sequenz erkennen und daran binden. LAMP nutzt eine DNA-Polymerase mit Strand-displacement Aktivität und läuft bei einer konstanten Temperatur von etwa 65°C ab. Amplifikation und Detektion der Target-Sequenz findet in einem einzigen Schritt statt.

[0026] Unter dem Term „Nucleic Acid Sequence Based Amplification“ (NASBA), wird ein Verfahren zur Amplifikation von RNA verstanden (Compton 1991). Dabei wird eine RNA-Matrix zu einem Reaktionsgemisch gegeben, und ein erster Primer bindet an die komplementäre Sequenz im Bereich des 3'-Endes der Matrix. Anschließend wird mit einer Reversen Transkriptase der zur Matrix komplementäre DNA-Strang polymerisiert. Mit Hilfe von RNase H wird dann die RNA-Matrix verdaut (RNase H verdaut nur RNA in RNA-DNA Hybriden, nicht single-stranded RNA). Anschließend wird ein zweiter Primer an das 5' Ende des DNA-Stranges gebunden. Dieser wird von der T7 RNA Polymerase als Startpunkt für die Synthese eines zum DNA-Strang komplementären RNA-Moleküls verwendet, welches dann wieder als Ausgangsmatrix verwendet werden kann. NASBA wird bei einer konstanten Temperatur von normalerweise 41°C. durchgeführt und liefert unter bestimmten Umständen schnellere und bessere Resultate als PCR.

[0027] Unter dem Term „Transcription Mediated Amplification“ (TMA) wird ein von der US-Firma Gen-Probe entwickeltes isothermes Amplifikationsverfahren verstanden, das der NASBA ähnlich ist und bei dem ebenfalls RNA Polymerase und Reverse Transkriptase verwendet werden (Hill, 2001)

[0028] Der Term „Rolling Circle Chain Reaction“ (RCCR) oder „Rolling circle Amplification“ (RCA) betrifft ein Amplifikationsverfahren, das die allgemeine Nukleinsäure-Replikation nach dem rolling-circle-Prinzip imitiert und unter anderem in der US 5854033 beschrieben ist.

[0029] Unter dem Begriff „Nested PCR“ wird ein Verfahren verstanden, bei dem ein bereits vervielfältigtes DNA-Fragment (oder ein Teil davon) ein weiteres Mal amplifiziert wird; dieser Vorgang erfolgt mit einem zweiten Primerpaar, das innerhalb des in der ersten Reaktion verwendeten Primerpaares angeordnet ist.

[0030] Unter dem Begriff „Promotor“ wird vorliegend eine Nukleinsäuresequenz verstanden, die die regulierte Expression eines Gens ermöglicht. Sie liegt upstream, d. h. 5' zum RNA-kodierenden Bereich des Gens. Eine wichtige Eigenschaft eines Promotors ist die spezifische Wechselwirkung mit bestimmten DNA-bindenden Proteinen, welche den Start der Transkription des Gens durch die RNA-Polymerase vermitteln und als Transkriptionsfaktoren bezeichnet werden. Die wichtigsten prokaryotischen Transkriptionsfaktoren sind die Sigma-Faktoren. Die mit Abstand meisten Promotoren bei z. B. E. coli werden über den Faktor Sigma-70 erkannt und weisen die folgenden Promotorelemente auf: 1) ein AT-basenpaarreiches UP-Element oberhalb der -35-Region (was auch direkt mit der α -Untereinheit der bakteriellen RNA-Polymerase interagieren kann), 2) die -35-Region, mit der Konsensussequenz: 5'-TTGACA-3', 3) die -10-Region (Pribnow-Box), mit der Konsensussequenz: 5'-TATAAT-3'. Weiterhin sind starke Promotoren direkt vor dem Startpunkt der Transkription reich an AT-Basenpaaren. Die Konsensussequenzen geben nur einen ungefähren Anhaltspunkt für den Aufbau eines Promotors. Ein bestimmter Sigma-70-abhängiger Promotor kann an mehreren Stellen von diesen Sequenzen abweichen. In einer bevorzugten Ausführungsform ist der Promotor ein induzierbarer Promotor. Ein Beispiel für einen induzierbaren Promotor ist der künstlich erzeugte tac-Promoter, der zur Erhöhung des Genexpressionslevels in Bakterienzellen (z. B. E. coli) und somit zur Überexpression rekombinanter Proteine genutzt werden kann. Dieser synthetische Promotor besteht aus der -35 Region des starken trp-Promotor aus E. coli und der -10 Region des IPTG und Laktose regulierbaren/induzierbaren lac-Promotors. Der induzierbare tac-Promoter ist ca. 3-mal stärker als der trp-Promoter und ca. 10-mal stärker als der lac-Promoter.

[0031] Unter dem Begriff „Biosynthese“ wird vorliegend die enzymatische Herstellung eines Polyketids verstanden, bevorzugt mittels einer oder mehrerer Polyketidsynthasen, bevorzugt einer trans-AT-Polyketidsynthase. Der Begriff „Biosynthesegen“ umfasst damit jedes Gen, welches für ein Protein codiert, das Teil einer solchen Polyketidsynthase ist.

[0032] Weiterhin umfasst der Begriff „Biosynthesegen“ auch Fragmente eines Biosynthesegens, die für ein trunkiertes Protein einer Polyketidsynthase codieren, das die gleiche Funktion aufweist, wie das Vollängenprotein. Dem Fachmann ist hinlänglich bekannt, dass durch Trunkierung (d. h. Deletion einzelner Bereiche, bevorzugt der Enden) eines Proteins eine gesteigerte oder verringerte Enzymaktivität herbeigeführt werden kann. Unter gleicher Funktion wird hier die prinzipiell gleiche biochemische Funktion verstanden, unabhängig davon, ob diese mit einer erhöhten oder erniedrigten Effektivität oder Geschwindigkeit ausgeführt wird. Als Beispiel für ein Biosynthesegen sei ein Gen für eine KS-Domäne genannt, deren biochemische Funktion es ist, $\alpha\beta$ -gesättigte Intermediate zu verlängern (siehe beispielsweise die KS der Klade V), oder ein Gen, dessen Genprodukt in einer PKS die biochemische Funktion ausübt, einen Methylrest zu transferieren (z. B. psyB). Jede trunkierte Version solcher Gene, die zu trunkierten Genprodukten führen, die noch dieselbe biochemische Funktion ausführen fallen somit ebenfalls unter den Begriff „Biosynthesegen“. Weitere Beispiele für Biosynthesegene und spezifische Ausführungsformen der Erfindung sind die in Tabelle 1 aufgelisteten Gene psyA-psyN.

[0033] Unter dem Begriff „Metagenom“ wird vorliegend eine Menge von zumindest zwei Genomen verschiedener Spezies verstanden. Ein Metagenom umfasst damit die gesamte genomische Information dieser zumindest zweier Spezies. In einer spezifischen Ausführungsform umfasst das Metagenom das Genom eines Schwamms und das von zumindest einem Prokaryoten. In einer weiteren Ausführungsform lebt der Prokaryot in Symbiose mit dem Schwamm.

[0034] Ein Primer wird im Sinne der vorliegenden Anmeldung dann als „spezifisch“ für eine gegebene Aminosäuresequenz angesehen, wenn er – bevorzugt unter stringenten Bedingungen – mit dem sense- oder antisense-Strang einer Nukleinsäure hybridisiert oder zu diesem identisch ist, die für die gegebene Aminosäuresequenz oder einen Teil davon kodiert. Da der Aminosäurecode degeneriert ist, können für eine gegebene Aminosäuresequenz mehrere geeignete Primer verwendet werden. Dem Fachmann ist bekannt, welche Basentriplets für welche Aminosäurereste codieren. Der Primer sollte eine minimale Länge aufweisen, die es ihm ermöglicht, mit einer Nukleinsäure zu hybridisieren, die für zumindest 2, zumindest 3, oder zumindest 4 zusammenhängende Aminosäurereste der gegebenen Aminosäuresequenz kodiert. In einer spezifischen Ausführungsform hybridisiert der Primer mit einer Nukleinsäure, die für zumindest 5, 6, 7 oder mehr als 7 zusammenhängende Aminosäurereste der gegebenen Aminosäuresequenz kodiert.

[0035] Ein Primer wird im Sinne der Erfindung auch dann als spezifisch für die definierten Aminosäuremotive angesehen, wenn er für eine Aminosäuresequenz spezifisch ist, die zumindest 2, zumindest 3, oder zumindest 4 zusammenhängende Aminosäurereste eines der in [Fig. 3](#) und [Fig. 4](#) oder SEQ ID NOs: 40–71 definierten Aminosäuremotive umfasst und zusätzlich am 3'- oder 5'-Ende weitere Nukleotide für nicht in den definierten Aminosäuremotiven enthaltene Aminosäurereste aufweist. Hierbei kann die Anzahl der Nukleotide des Primers, die für Aminosäurereste spezifisch sind, die nicht im definierten Aminosäuremotiv enthalten sind, bis zu 30% der Gesamtanzahl der Nukleotide des Primers betragen. Als Beispiel kann ein Primer, der spezifisch ist für das in [Fig. 3](#) definierte Aminosäuremotiv „YYQAGML“ auch so hergestellt werden, dass er für die Aminosäuresequenz „YYQAGML_A“ spezifisch ist, wobei die zusätzlichen Nukleotide für das Alanin, welches nicht im definierten Aminosäuremotiv enthalten ist, 12,5% der gesamten Primersequenz ausmachen.

[0036] Polyketidsynthasen lassen sich über ihre Struktur und Funktion in drei Klassen einteilen. Typ I Enzyme sind multifunktionelle, häufig modular aufgebaute Proteine. Bakterielle modulare Typ I PKS erzeugen eine große Menge therapeutisch wichtiger Naturstoffe, wie z. B. Erythromycin, Epothilone und FK506 (Fischbach & Walsh. *Assembly-line enzymology for poliketide and nonribosomal peptide antibiotics: logic, machinery and mechanisms.* Chem. Rev. 106, 3468; 2006). [Fig. 1](#) zeigt einige beispielhafte Polyketide, die, soweit deren Gencluster bekannt sind, von Typ I PKS gebildet werden.

[0037] Jedes dieser Enzyme umfasst wenigstens eine β -Ketosynthase (KS), eine Acyltransferase (AT) sowie ein Acyl-Carrier-Protein (ACP), die den nächsten Baustein zum Aufbau des Polyketids auswählen, aktivieren und unter Bildung eines β -Ketoacyl-S-ACP Intermediats, eine decarboxylative Claisen-Kondensation zwischen dem nächsten Baustein und der wachsenden Polyketidkette katalysieren. Während des Syntheseprozesses überträgt eine AT Domäne den korrespondierenden Acyl-CoA Baustein auf eine ACP Domäne und eine KS Domäne verlängert die Polyketidkette mit dieser Acyl-Einheit, wobei zusätzliche Domänen das resultierende β -oxo Intermediat weiter modifizieren können. Die Reihenfolge der Module in den PKS gibt hierbei die Reihenfolge der Biosyntheseschritte zwingend vor. Somit kann durch Variation der Domänen/Module eine hohe strukturelle Diversität in den entstehenden Polyketiden erzeugt werden.

[0038] Eine besondere und evolutionär abgegrenzte Enzymfamilie der PKS bilden die trans-AT-PKS, die freistehende AT verwenden, um die zur Herstellung des Polyketids benötigten Bausteine in trans zu selektieren (Cheng et al., *Type I polyketide synthase requiring a discrete acetyltransferase for polyketide biosynthesis.* Proc. Natl. Acad. Sci. U.S.A. 100, 3149–3154; 2003). Demgegenüber weisen cis-AT PKS integrierte AT Domänen auf. Die bisher veröffentlichten Studien über komplexe Polyketide haben gezeigt, dass alle bisher aus Symbionten geklonten PKS-Cluster zu dieser Familie der trans-AT-PKS gehören.

[0039] Die vorliegende Erfindung basiert auf der überraschenden Entdeckung, dass die KS von trans-AT-PKS ein einzigartiges phylogenetisches Muster aufweisen, bei dem die nächstverwandten KS-Domänen für dieselben oder ähnliche Substrate spezifisch sind. Demgegenüber werden unterschiedliche Substrate, z. B. β -hydroxylierte und olefinische, von evolutionär klar unterscheidbaren, d. h. nicht engverwandten, KS prozessiert. Die phylogenetische Analyse eines oder mehrerer KS liefert somit nützliche Informationen über die Struktur der involvierten Intermediate.

[0040] Erfindungsgemäß wird der umgekehrte Weg eingeschlagen, bei dem aufgrund der chemischen Struk-

tur des Polyketids Rückschlüsse auf die involvierten KS gezogen werden können. Es zeigte sich überraschender Weise, dass dieser Ansatz die Identifikation von konservierten Sequenzmustern/Motiven innerhalb dieser substratspezifischen KS Domänen ermöglicht. Aufgrund dieser konservierten Motive können dann erfindungsgemäß Primer generiert werden, die spezifisch sind für KS Domänen mit einer bestimmten, gewünschten Substratspezifität. Mithilfe dieser Primer ist dann die Amplifikation solcher KS Domänen möglich, die für ein bestimmtes Substrat spezifisch sind. Diese Amplikons können dann zur Identifizierung und Isolierung eines oder mehrerer Polyketid-Biosynthesegene, z. B. in Genclustern verwendet werden, z. B. durch PCR, Screening von Fosmid-Bibliotheken oder Genome Walking.

[0041] Einer der Hauptvorteile der vorliegenden Erfindung liegt somit darin, dass ausgehend vom Polyketid gezielt eine in dessen Biosynthese involvierte KS Domäne einer trans-AT-PKS identifiziert werden kann, um in einem nächsten Schritt deren Nukleotidsequenz, sowie die Nukleotidsequenz des sie umfassenden Polyketid-Biosynthese-Genclusters zu identifizieren und zu isolieren. Der vorliegende Erfindung liegt demnach die überraschende Entdeckung zugrunde, dass ausgehend von der chemischen Struktur, d. h. den funktionellen Gruppen eines Polyketids, konservierte Motive innerhalb von KS Domänen identifiziert werden können, die in die Biosynthese dieses Polyketids involviert sind. Basierend auf diesen konservierten Motiven lassen sich dann Primer ableiten die zur Amplifikation und Isolierung der in die Biosynthese dieses Polyketids involvierten Gene genutzt werden können.

Nukleinsäuren und Peptide der vorliegenden Erfindung

[0042] Ein Aspekt der vorliegenden Erfindung ist eine Nukleinsäure, die zumindest ein Gen umfasst, das für ein Polypeptid codiert welches an der Biosynthese der Polyketide Psymberin (auch bekannt als Irciniastatin A), oder Mycalamid beteiligt ist. In einer Ausführungsform liegt die Nukleinsäure in isolierter Form vor. In einer weiteren Ausführungsform ist die Nukleinsäure das Komplement einer Nukleinsäure, die zumindest ein Gen umfasst, das für ein Polypeptid codiert welches an der Biosynthese der Polyketide Psymberin oder Mycalamid beteiligt ist. In einer weiteren Ausführungsform ist das Gen ein prokaryotisches Gen und/oder Teil eines polycistronischen Genabschnitts oder Gen-Clusters das/der für zumindest ein Protein codiert, das an der Biosynthese eines Polyketids beteiligt ist. In einer Ausführungsform gehört das Gen zu den Typ I PKS; in einer bevorzugten Ausführungsform gehört das Gen zu der evolutionär abgegrenzten Enzymfamilie der trans-AT PKS. Trans-AT PKS verwenden freistehende Acetyltransferasen (AT), um die zur Herstellung des Polyketids benötigten Bausteine in trans zu selektieren (Cheng et al., Type I polyketide synthase requiring a discrete acetyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 100, 3149–3154; 2003), während cis-AT PKS integrierte AT Domänen aufweisen.

[0043] In einer bevorzugten Ausführungsform wurde die erfindungsgemäße Nukleinsäure aus dem Metagenom eines Schwamms amplifiziert. Eukaryotische Schwämme (Porifera) gehen mit diversen Prokaryoten Symbiosen ein, wobei diese Prokaryoten vielfach noch nicht isoliert kultiviert werden konnten. Das dadurch aus Schwamm und Prokaryoten entstehende Metagenom umfasst das Genom des eukaryotischen Schwamms sowie die prokaryotischen Genome der einzelnen Symbionten. In einer Ausführungsform gehört der Schwamm zu den Demospongiae, den Dictyoceratida oder den Irciniidae. In einer bevorzugten Ausführungsform gehört der Schwamm zu Psammocinia sp.; noch bevorzugter ist der Schwamm Psammocinia aff. bulbosa. In einer bevorzugten Ausführungsform gehört der Schwamm zu den Demospongiae. In weiteren bevorzugten Ausführungsformen gehört der Schwamm zu der Familie der Spongiidae, der Thorectidae, der Halichondriidae, der Callyspongiidae, der Chalinidae, der Mycalidae, der Coelosphaeridae, der Tedaniidae, der Plakinidae, oder der Triaenosina. Im Sinne dieser Erfindung bevorzugte Vertreter dieser Familien sind Spongia spp., Cacospongia spp., C. mycofijiensis, Halichondria spp., Callyspongia spp., Haliclona spp., Mycale spp., M. hentscheli, Lysodendoryx spp., Tedania spp., Plakortis spp., Theonella swinhoei, und Discodermia spp.

[0044] Die Amplifikation der Nukleinsäure aus dem Metagenom kann mittels eines beliebigen, dem Fachmann bekannten, Verfahrens zur Nukleinsäureamplifikation durchgeführt werden. In einer Ausführungsform erfolgt die Amplifikation der erfindungsgemäßen Nukleinsäure mittels einer isothermen oder einer thermocyclischen Amplifikationsreaktion. Bevorzugt erfolgt die Amplifikation mittels Ligase-Kettenreaktion (LCR), und/oder Polymerase Ketttenreaktion (PCR); noch bevorzugter ist die PCR eine nested PCR.

[0045] In einer Ausführungsform umfasste die erfindungsgemäße Nukleinsäure zumindest eine Sequenz, die ausgewählt ist aus der Gruppe bestehend aus SEQ ID NO 1–19, 39, einem Homolog dieser Sequenzen und dem Komplement der Sequenzen oder des Homologs. In einer bevorzugten Ausführungsform umfasst die erfindungsgemäße Nukleinsäure zumindest eine Sequenz, die ausgewählt ist aus der Gruppe bestehend aus SEQ ID NO 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 (psyA, psyB, psyC, psyD, psyE, psyF, psyG, psyH,

psyl, psyJ, psyK, psyL, psyM und psyN), einem Homolog dieser Sequenzen und/oder dem Komplement der Sequenzen oder des Homologs. In einer weiteren Ausführungsform weist das Homolog zumindest 75%, 80%, 85%, 87%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, bevorzugt 97%, 98%, 99%, 99,5%, 99,7%, 99,9% Sequenzidentität mit der verglichenen Nukleinsäure auf, und/oder hybridisiert unter Hochsalzbedingungen mit der Nukleinsäure.

[0046] In einer weiteren Ausführungsform codiert die erfindungsgemäße Nukleinsäure für ein Protein, das zumindest eine Identität zu SEQ ID NO 23 von 41%, zu SEQ ID NO 24 von 52%, zu SEQ ID NO 25 von 42%, zu SEQ ID NO 28 von 44%, zu SEQ ID NO 29 von 94%, zu SEQ ID NO 30 von 96%, zu SEQ ID NO 31 von 96%, zu SEQ ID NO 32 von 94%, zu SEQ ID NO 33 von 98%, zu SEQ ID NO 34 von 95%, zu SEQ ID NO 35 von 97%, zu SEQ ID NO 36 von 98%, zu SEQ ID NO 37 von 96% oder zu SEQ ID NO 38 von 96% aufweist. In einer bevorzugten Ausführungsform codiert die erfindungsgemäße Nukleinsäure für ein Protein, das zumindest eine Identität zu SEQ ID NO 23 von 50%, zu SEQ ID NO 24 von 55%, zu SEQ ID NO 25 von 50% oder zu SEQ ID NO 28 von 50% aufweist.

[0047] Ein weiterer Aspekt der vorliegenden Erfindung ist ein Peptid, Polypeptid, Protein oder Enzym, für das eine der erfindungsgemäßen Nukleinsäuren codiert.

[0048] In einer Ausführungsform weist das Peptid eine Aminosäuresequenz auf, die ausgewählt ist aus der Gruppe bestehend aus SEQ ID NO 20–38.

[0049] In einer weiteren Ausführungsform weist das Peptid zumindest eine Identität zu SEQ ID NO 20 von 36%, zu SEQ ID NO 21 von 54%, zu SEQ ID NO 22 von 44%, zu SEQ ID NO 23 von 41%, zu SEQ ID NO 24 von 52%, zu SEQ ID NO 25 von 42%, zu SEQ ID NO 28 von 44%, zu SEQ ID NO 29 von 94%, zu SEQ ID NO 30 von 96%, zu SEQ ID NO 31 von 96%, zu SEQ ID NO 32 von 94%, zu SEQ ID NO 33 von 98%, zu SEQ ID NO 34 von 95%, zu SEQ ID NO 35 von 97%, zu SEQ ID NO 36 von 98%, zu SEQ ID NO 37 von 96% oder zu SEQ ID NO 38 von 96% auf. In einer bevorzugten Ausführungsform weist das Peptid zumindest eine Identität zu SEQ ID NO 23 von 50%, zu SEQ ID NO 24 von 55%, zu SEQ ID NO 25 von 50% oder zu SEQ ID NO 28 von 50% auf. In einer weiteren bevorzugten Ausführungsform weist das Peptid eine Sequenzidentität von mindestens 60%, 70%, 75%, 80%, 90%, 92%, 95%, 97%, 98%, 99% oder 99,5% auf.

[0050] In einer weiteren Ausführungsform betrifft die Erfindung einen Vektor, der die erfindungsgemäße Nukleinsäure umfasst. Bevorzugt ist dieser Vektor ein Cosmid, ein YAC, BAC oder ein Virus, noch bevorzugter ist der Vektor ein Plasmid. In einer Ausführungsform umfasst der Vektor weiterhin einen Promotor unter dessen Kontrolle die Expression der erfindungsgemäßen Nukleinsäure steht. Bevorzugt ist dieser Promotor ein induzierbarer Promotor. In einer weiteren Ausführungsform steht die erfindungsgemäße Nukleinsäure unter der Kontrolle eines induzierbaren Expressionssystems. In einer spezifischen Ausführungsform ist das induzierbare Expressionssystem das Tetracyclin Operator/Repressor (TetO2/TetR) System.

[0051] In einer weiteren Ausführungsform betrifft die Erfindung eine rekombinante Wirtszelle, die eine der erfindungsgemäßen Nukleinsäuren oder Vektoren umfasst. Damit stellt die vorliegende Erfindung auch ein Expressions-/und Produktionssystem für Polyketide bereit. Vorzugsweise ist diese rekombinante Wirtszelle nach der Transformation/Transfektion in der Lage Polyketide zu synthetisieren. In einer Ausführungsform ist die die rekombinante Wirtszelle eine E. coli Zelle, in weiteren Ausführungsformen ist die rekombinante Wirtszelle eine Wirtszelle aus *Pseudomonas* spp., *Bacillus* spp., *Streptomyces* spp. oder *Acinetobacter baylyi*.

Erfindungsgemäße Verfahren zur Herstellung von Polyketiden

[0052] Ein weiterer Aspekt der Erfindung betrifft die Verwendung der erfindungsgemäßen Nukleinsäuren, Vektoren oder rekombinanten Wirtszellen in einem Verfahren zur Herstellung eines Polyketids. Hierbei wird ein Expressionssystem verwendet, welches die Expression der erfindungsgemäßen Nukleinsäuren erlaubt.

[0053] In einer Ausführungsform ist das Expressionssystem eine rekombinante Wirtszelle, die die erfindungsgemäßen Nukleinsäuren oder Vektoren umfasst und zur Expression der Genprodukte der erfindungsgemäßen Nukleinsäuren in der Lage ist. Die exprimierten Genprodukte können dann zur Synthese des Polyketids *in vivo* oder *in vitro* genutzt werden. In einer bevorzugten Ausführungsform ist das hergestellte Polyketid Psymberein oder Mycalamid. In einer weiteren bevorzugten Ausführungsform umfasst das Verfahren oder das Expressionssystem die Verwendung von zumindest einer Nukleinsäure umfassend eine Sequenz ausgewählt aus der Gruppe bestehend aus SEQ ID NO 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 (psy A, psyB, psyC, psyD, psyE, psyF, psyG, psyH, Asyl, psyJ, psyK, psyL, psyM und psyN) oder einem Homolog dieser Sequenzen.

[0054] In einer spezifischen Ausführungsform findet die Kultivierung einer erfindungsgemäßen rekombinanten Wirtszelle unter Bedingungen statt, die die Expression der Biosynthesegene und Produktion des Polyketids durch die Wirtszelle erlauben. Solche Bedingungen sind dem Fachmann hinlänglich bekannt oder lassen sich im Wege simpler Routineexperimente für ein gegebenes Polyketid erhalten. In einer spezifischen Ausführungsform ist die rekombinante Wirtszelle eine Wirtszelle aus *Pseudomonas* spp., *Bacillus* spp., *Streptomyces* spp., *Corynebacterium* spp., *Acinetobacter baylyi*, *Pseudomonas stutzeri*, *Bacillus subtilis* oder *Bacillus amyloliquefaciens*. Falls das zumindest eine erfindungsgemäße Gen, welches sich in der Wirtszelle befindet unter der Kontrolle eines induzierbaren Expressionssystem (z. B. des Tetracyclin Operator/Repressor (TetO2/TetR) Systems) steht, kann durch Zugabe der jeweiligen induzierenden Substanz die Expression des Genprodukts gezielt induziert werden.

[0055] Zur Herstellung des Polyketids werden die Gene des Genclusters in die Genome der Expressionswirte integriert. Die Integration erfolgt dabei sukzessive durch homologe Rekombination kürzerer (in etwa bis 40 kb) Abschnitte des Genclusters. Um die Gene effizient zu transferieren, können bevorzugt transformierbare Bakterien wie *Pseudomonas stutzeri*, *Acinetobacter baylyi*, *Bacillus subtilis* oder *Bacillus amyloliquefaciens* verwendet werden, die linearisierte DNA aufnehmen können. Im Bedarfsfall werden die natürlichen Promotoren des Genclusters gegen wirtseigene Promotoren ersetzt (z. B. den *baeB*-Promotor bei *B. amyloliquefaciens*). Alternativ zur Integration in das Genom können Gene auf einem oder verteilt auf mehreren kompatiblen Plasmiden durch Konjugation, Protoplastentransformation oder Elektroporation in die Expressionswirte transferiert werden. Die Expression erfolgt dann von frei replizierenden Plasmiden oder wiederum nach Integration in das Genom.

[0056] In einer weiteren Ausführungsform umfasst das Verfahren weiterhin den Schritt der Isolierung des produzierten Polyketids. In einer Ausführungsform erfolgt diese Isolierung mittels HPLC und/oder präparativer Säulenchromatographie.

Erfindungsgemäße Verfahren zur Isolierung von Polyketid-Biosynthesegenen

[0057] Die Isolierung des zumindest einen Polyketid-Biosynthesegens unter Verwendung des zumindest einen Primerpaars umfasst in einer Ausführungsform die Amplifikation der vom jeweiligen Primerpaar eingeschlossenen Nukleinsäureregion mittels des Primerpaars. Somit wird eine Nukleinsäureregion der Kethosynthasedomäne amplifiziert. In einer bevorzugten Ausführungsform wird diese Amplifikation mittels PCR durchgeführt, noch bevorzugter mittels einer nested PCR.

[0058] In einem weiteren Aspekt ist die vorliegende Erfindung auf ein Verfahren zur Isolierung von zumindest einem Polyketid-Biosynthesegen gerichtet und umfasst die Schritte des Bereitstellens von zumindest einem Primerpaar, das aufgrund der Substratspezifität einer Kethosynthasedomäne einer trans-AT Polyketidsynthase erstellt worden ist, und die Isolierung von zumindest einem Polyketid-Biosynthesegen unter Verwendung dieses zumindest einen Primerpaars.

[0059] Die erfindungsgemäße Bereitstellung, d. h. das „Design“ der Primer basiert auf der überraschenden Entdeckung, dass die KS von trans-AT PKS ein einzigartiges phylogenetisches Muster aufweisen, bei dem die nächstverwandten KS Domänen für dieselben oder ähnliche Substrate spezifisch sind. Somit kann, ausgehend von den funktionellen Gruppen eines Polyketids, ein konserviertes Motiv innerhalb von KS Domänen identifiziert werden, auf dessen Basis Primer hergestellt werden können, die für ein solches konserviertes Motiv – und damit für eine KS Domäne, die in die Biosynthese des Polyketids involviert ist – spezifisch sind. Die Amplifikation und Isolierung der KS Domäne und damit der die KS Domäne umfassenden PKS sowie weiteren in die Biosynthese des Polyketids involvierten Genen, die häufig in einem Gencluster vorliegen, wird somit erfindungsgemäß ermöglicht.

[0060] Ob eine gegebene KS oder KS Domäne zu einer zweiten KS oder KS Domäne ähnlich oder phylogenetisch benachbart ist, kann zum Beispiel durch Sequenzvergleiche bestimmt werden, bei denen zueinander ähnlichere Sequenzen phylogenetisch enger benachbarten KS oder KS Domänen entsprechen. In einer bevorzugten Ausführungsform wird hierfür eine phylogenetische Analyse eingesetzt. In einer spezifischen Ausführungsform wird für die phylogenetische Analyse ein Kladogramm erstellt. In einer weiteren spezifischen Ausführungsform wird das Kladogramm mit der Bayesischen Methode und wie in Beispiel 3 beschrieben erstellt.

[0061] In einer weiteren Ausführungsform wird nach der phylogenetischen Analyse für die KS die Domänenarchitektur des N-terminal benachbarten Moduls analysiert und aus der Polyketidstruktur das Substrat der

KS-Domäne abgeleitet. Unter der Domänenarchitektur versteht man hierbei die Abfolge von Domänen in einem Modul, wie z. B. KS DH KR ACP.

[0062] In einer bevorzugten Ausführungsform ist das erfindungsgemäße Verfahren auf die gerichtete Isolation zumindest eines in die Polyketid-Biosynthese involvierten Gens ausgerichtet, d. h. es wird/werden ausgehend von den funktionellen Gruppen eines Polyketids das/die zur Herstellung dieses Polyketids nötige(n) Gen(e) isoliert. In einer beispielhaften, jedoch bevorzugten Ausführungsform ist das Polyketid-Biosynthesegen in die Biosynthese von Psymberin involviert. Psymberin ([Fig. 1](#), 1) trägt keinen vollständig reduzierten Rest, sondern eine Acetylableitete Startereinheit, die im Regelfall von einer ganz bestimmten Art von KS verlängert wird. In einer phylogenetischen Analyse ([Fig. 2](#)) zeigte sich, dass KS mit einer solchen Substratspezifität zu Klade VI gehören. Klade VI lässt sich weiterhin differenzieren in KS Domänen, die einem Modul mit einer GCN5-related N-Acetyltransferase (GNAT) Domäne folgen (Klade VIa) und solchen, die keine GNAT Domäne aufweisen (Klade VIb). Ein Sequenzalignment der in Klade VIa enthaltenen KS Domänen zusammen mit allen bekannten Sequenzen der Klade VI ergab zwei konservierte Motive: das EDAGY Motiv, sowie das Klade VIa Motiv YYQ/KAGML, welches einem Modul mit einer GCN5-related N-Acetyltransferase (GNAT) Domäne folgte ([Fig. 3](#)). Diese konservierten Motive traten nicht in den Sequenzen der anderen Kladen auf. Weiterhin wurde im EDAGY Motiv ein Tyrosinrest identifiziert, der in 84% der trans-AT KS konserviert ist, jedoch in den ubiquitären sup KS nicht auftritt. Aufgrund der identifizierten konservierten Motive können daher nun zur Isolation eines in die Psymberinbiosynthese involvierten Gens Primer generiert werden, die in der forward-Orientierung spezifisch sind für das EDAGY-Motiv und in der reverse-Orientierung spezifisch sind für das YYQ/KAGML-Motiv.

[0063] Die unter Verwendung dieser Primer erzeugten Amplifikate können dann zur Isolierung der Nukleinsäuresequenz für die KS Domäne, der Nukleinsäuresequenz für die die KS Domäne umfassende PKS, sowie der Nukleinsäuresequenz des gesamten zur Polyketidsynthese nötigen Genclusters verwendet werden. Verfahren zur Isolierung solcher Nukleinsäuresequenzen sind dem Fachmann wohlbekannt und schließen beispielsweise Genome-Walking oder das Screening von Cosmid- oder Fosmid-Bibliotheken ein. Die isolierten Nukleinsäuren können dann mittels Sequenzierung bestimmt/identifiziert werden.

[0064] In einer spezifischen Ausführungsform ist zumindest ein Primer spezifisch für 2, 3, 4, 5, 6, oder mehr als 6 zusammenhängende Aminosäurereste eines der in [Fig. 3](#) oder [Fig. 4](#) definierten kladenspezifischen Aminosäuremotive, die ebenfalls in SEQ ID NO 40–71 wiedergegeben sind. In einer weiteren spezifischen Ausführungsform ist der zumindest eine Primer des Primerpaars am 3' Ende des Primers spezifisch für die besagten Aminosäurereste.

[0065] In einer weiteren spezifischen Ausführungsform sind beide Primer des Primerpaars spezifisch für zumindest 2, 3, 4, 5, 6 oder mehr als 6 zusammenhängende Aminosäurereste der in [Fig. 3](#) oder [Fig. 4](#) definierten Aminosäuremotive einer Klade, wobei die beiden Primer bezogen auf die Nukleinsäure, mit der sie hybridisieren, einen ausreichenden Abstand zueinander einhalten, um eine Amplifikation zu ermöglichen.

[0066] In einer anderen spezifischen Ausführungsform hat das verwendete Primerpaar die Nukleotidsequenz: 5'-GCN HTN GAR GAY GCN GGN TAY GC-3' (EDAGY; SEQ ID NO: 72) und 5'-C NAR CAT NCC NGC YTK RTA RTA-3' (YYQAGMLA; SEQ ID NO: 73).

[0067] In einer weiteren Ausführungsform wird für die Amplifikation der vom zumindest einen Primerpaar eingeschlossenen Nukleinsäureregion der KS Domäne neben dem zumindest einen, erfindungsgemäßen Primerpaar ein zweites Primerpaar verwendet, wobei das erste und das zweite Primerpaar verschachtelt sind, d. h. das eine Primerpaar vollständig innerhalb des vom anderen Primerpaar zu amplifizierenden Nukleinsäurebereichs liegt. Das zweite Primerpaar ist in einer bevorzugten Ausführungsform ein erfindungsgemäß hergestelltes Primerpaar und basiert damit ebenfalls auf der Substratspezifität einer Kethosynthesedomäne einer trans-AT Polyketidsynthase. In einer anderen Ausführungsform ist das zweite Primerpaar ein degeneriertes Primerpaar. In einer besonders bevorzugten Ausführungsform ist dieses Primerpaar spezifisch für die Aminosäuresequenz KSDPQQF (forward) bzw. KSHGTGR (reverse). In einer noch bevorzugteren Ausführungsform hat das zweite Primerpaar die Nukleotidsequenz 5'-MGN GAR GCN NWN SMN ATG GAY CCN CAR CAN MG-3' bzw. 5'-GGR TCN CCN ARN SWN GTN CCN GTN CCR TG-3'.

[0068] In spezifischen Ausführungsformen wird die Amplifikation mit den zwei Primerpaaren gleichzeitig oder zweistufig durchgeführt.

[0069] Bei der zweistufigen Amplifikation wird in der ersten Amplifikationsrunde das außenliegende Primer-

paar verwendet. Danach werden in einer zweiten Amplifikationsrunde die Amplifikate aus der ersten Amplifikationsrunde mit den innenliegenden, verschachtelten Primern amplifiziert. In einer weiteren bevorzugten Ausführungsform ist die mit den zumindest zwei Primerpaaren durchzuführende Amplifikation eine nested PCR.

[0070] In einer weiteren Ausführungsform erfolgt die Isolation des zumindest einen Polyketid-Biosynthesegens aus einem Metagenom. In einer bevorzugten Ausführungsform umfasst das Metagenom das Genom zumindest eines Eukaryoten und zumindest eines Prokaryoten. Bevorzugt befinden sich Eukaryot und Prokaryot in einer Symbiose. In einer weiteren Ausführungsform ist der Eukaryot ein Schwamm, bevorzugt gehört dieser Schwamm zu den Demospongiae, den Dictyoceratida oder den Irciniidae. In einer bevorzugten Ausführungsform gehört der Schwamm zu *Psammocinia* sp.; noch bevorzugter ist der Schwamm *Psammocinia* aff. *bulbosa*.

[0071] In einer weiteren bevorzugten Ausführungsform ist das zu isolierende Polyketid-Biosynthesegen in die Biosynthese von Psymberin, den Mycalamiden, den Onnamiden, den Theopederinen, den Icadamiden, Pederin, den Laulimaliden (Fijianoliden), den Latrunculinen, Mycothiazol, Pateamin A, den Spongistatinen (Altohyrtinen), Discodermolid, Dictyostatin, Tedanolid und/oder Calyculin involviert, d. h. Teil einer Polyketidsynthese, die eines dieser Polyketide synthetisiert.

[0072] In einer weiteren Ausführungsform umfasst das erfindungsgemäße Verfahren weiterhin den Schritt der Isolierung des gesamten für die Polyketidsynthese benötigten Genbereichs basierend auf den Sequenzinformationen des isolierten zumindest einen Polyketid-Biosynthesegens. In einer spezifischen Ausführungsform erfolgt diese Isolierung des gesamten Genbereichs unter Verwendung einer Library. In weiteren spezifischen Ausführungsformen ist die Library eine Cosmid-, YAC-, BAC- oder Fosmid-Library. In einer weiteren spezifischen Ausführungsform wurde die Library aus einem Metagenom erzeugt. In weiteren spezifischen Ausführungsformen der Erfindung ist das Metagenom das Metagenom eines Schwammes.

BEISPIELE

Beispiel 1

[0073] Die Aminosäuresequenzen von 138 bekannten trans-AT-PKS-Modulen wurden aus der GenBank Datenbank entnommen und unter manueller Kontrolle mittels ClustalX geordnet und ausgerichtet („Alignment“). Hierbei wurden Deletionen und nicht ausrichtbare Sequenzbereiche ausgeschlossen.

[0074] Hiernach wurde eine phylogenetische Analyse der Vollängen-KS-Domänen der bearbeiteten Aminosäuresequenzen mittels der MrBayes Software (v3) unter Verwendung des „WAG replacement models“ und einer Gammaverteilung mit vier Kategorien durchgeführt. Die Markovketten-Monte Carlo Analyse lief für 2,5 Millionen Zyklen, wobei alle 100 Generationen ein Sample entnommen wurde. Die Konvergenz wurde mittels Plots der Maximum-Likelihood-Werte und einer Standardabweichung von $< 0,05$ bewertet. Der Consensus-Baum und die späteren Kladenwahrscheinlichkeiten wurden nach Verwerfen derjenigen Bäume berechnet, die vor dem Erreichen der Konvergenz als Sample entnommen worden waren.

[0075] Die auf dem Neighborhood-Joining-Verfahren beruhende phylogenetische Analyse wurde mittels der Programme Seqboot, Protdist, Neighbor und Consens der PHYLIP-Software durchgeführt. Die Bootstrap-Analyse wurde mit 1000 Pseudoreplikatsequenzen durchgeführt. Für die Maximum-Parsimony-Analyse wurde die heuristische Suchoption der Software PAUP*4.0b10 verwendet.

[0076] Das in [Fig. 2](#) abgebildete abgeleitete Kladogramm zeigt deutlich, dass einzelne Kladen KS von verschiedenen Bakterien enthalten. Anschließend wurde für alle KS die Domänenarchitektur des N-terminal benachbarten Moduls analysiert und daraus zusammen mit biosynthetischen Überlegungen (vorausgesetzt aus der Polyketidstruktur) die Substrate aller KS-Domänen abgeleitet. Es ergaben sich phylogenetisch klar abgegrenzte Gruppen von KS mit gleichen Substraten, also eine Verteilung der Domänen der individuellen Biosynthesewege auf einzelne Kladen.

[0077] Alle in diesem Kladogramm gegebenen Kladen wurden weiterhin eingehend getrennt analysiert, um Sequenzmuster aufzuspüren, welche die Bereitstellung von für diese Motive spezifische Primer erlauben.

[0078] [Fig. 3](#) und [Fig. 4](#) zeigen die so detektierten Konsensussequenzen, aus denen sich erfindungsgemäß Primer ableiten lassen, die jeweils nur für diejenige KS-Domäne spezifisch sind, die in den Biosyntheseweg einer bestimmten in einem Polyketid gefundenen Gruppe involviert sind.

[0079] Weiterhin wurde ein Tyrosinrest in einem EDAGY-Motiv identifiziert, der in 84% aller trans-AT-KS konserviert ist, jedoch in den ubiquitären sup KS nicht auftritt.

Beispiel 2: Isolierung der Metagenom-DNA und Bereitstellung einer Library

Isolierung

[0080] Proben der Schwämme *P. aff. bulbos* und *M. hentscheli* wurden gesammelt und bei -80°C eingefroren. Das Metagenom aus *P. aff. bulbos* wurde zur Isolierung des Psymberin-PKS-Genclusters verwendet, während das Metagenom aus *M. hentscheli* die Quelle für die Mycalamid-Biosynthese-Gencluster bereitstellte.

[0081] Die Gesamt-DNA aus den Proben wurde mittels eines dem Fachmann bekannten Verfahrens isoliert, um den hohen Gehalt an Polysacchariden zu entfernen (Murray, M. G. and W. F. Thompson (1980). Rapid isolation of high-molecular-weight plant DNA. *Nuc. Acids Res.* 8: 4321–4325). Hierbei wurde das gefrorene Schwammmaterial aufgetaut und ein kleines Stück von ca. 1 g wurde abgeschnitten, welches sowohl Schwammoberfläche als auch Innenleben des Schwamms enthielt. Das Gewebe wurde unter flüssigem Stickstoff unter Verwendung von vorgekühlten Mörser und Pistill zerrieben. Das resultierende Pulver wurde direkt im Anschluß in eine 50 ml Falcon-Tube überführt, die 10 ml Lysepuffer (8 M Harnstoff, 2% Sarkosyl, 1 M NaCl, 50 mM EDTA und 50 mM Tris-HCl pH 7,5) enthielt. Die Lysereaktion wurde für 15 min bei 60°C durchgeführt, wobei der Ansatz alle 5 min vorsichtig durchmischt wurde.

[0082] Die isolierte DNA wurde danach zweimal mit 10 ml Phenol/ CHCl_3 /Isoamylalkohol (25:24:1; v:v:v) und einmal mit 10 ml CHCl_3 extrahiert. Jeder Extraktion folgte ein Zentrifugationsschritt bei 11.000 rpm für 5 min.

[0083] Um Polysaccharide zu entfernen, wurde die obere wässrige Phase, die die DNA enthielt, mit einer 10%igen (w/v) wässrigen Cetyltrimethylammoniumbromid (CTAB) Lösung mit einer Endkonzentration von 0,5% behandelt. Die Mischung wurde für 15 min bei 60°C inkubiert und das Präzipitat wurde mittels Zentrifugation bei 11.000 rpm für 15 min entfernt.

[0084] Die Nukleinsäuren wurden mit 2 Volumina vorgekühlten Ethanols (98%) und 1/10 Volumen einer 3 M Natriumacetatlösung (pH 7) gefällt. Das nach einem 30 minütigen Zentrifugationsschritt bei 11.000 rpm bei 4°C erhaltene Präzipitat wurde zweimal mit kaltem 70%igen Ethanol gewaschen. Die crude DNA (ca. 5 μg) wurde für 10–15 min an der Luft getrocknet und in Tris-Puffer (10 mM, pH 8,5) resuspendiert.

Library

[0085] Zur Herstellung der Library wurde das CopyControl Fosmid Library Production Kit (Epicentre Biotechnologies) verwendet, wobei das Protokoll des Herstellers modifiziert wurde. Die DNA wurde in einem 1% (w/v) low-melting point Agarosegel aufgetrennt und Fragmente mit einer Größe von ca. 35 kb wurden mittels Gelverdaus mit GELase isoliert. Hierbei wurde 1 U der GELase Enzymlösung zu jeweils 300 μl geschmolzener Agarose hinzugegeben, gefolgt von einer Inkubation von 2 h bei 45°C . Nach Inaktivierung des Enzyms (70°C , 10 min) und Entfernung von unverdauter Agarose durch Zentrifugation für 20 min bei 11.000 rpm wurde die DNA bei Raumtemperatur mit 2,5 Volumina von 98%igem Ethanol und 1/10 Volumen von 3 M Natriumacetat (pH 5,2) gefällt. Nach einem Zentrifugationsschritt für 30 min bei 11.000 rpm und zwei Waschschritten mit 70%igem Ethanol bei 4°C wurde die pelletierte DNA in 50 μl Tris-Puffer (10 mM, pH 8,5) aufgenommen. Diese metagenomische DNA wurde mit Isopropanol und 3 M Natriumacetat (pH 5,2) gefällt.

[0086] Ungefähr 250 ng von blunt-ended DNA (mit Tris-Puffer eluiert; 10 mM, pH 8,5) wurden in einem Ligationsschritt mit dem pCC1FOS Vektor über Nacht und bei 16°C nach Herstellerangaben verwendet, wobei jedoch keine T4 DNA Ligase eingesetzt wurde. Die DNA wurde dann in den λ phagen gepackt und zur Transfektion den EPI300-T1^R E. coli Zellen gemäß den Angaben des Herstellers verwendet.

[0087] Für die Herstellung von Libraries von großer Größe wurden alle Schritte von der DNA-Präparation bis zum Ausplattieren der Library an einem Tag (und ohne Einfrieren der Proben) durchgeführt.

Beispiel 3: Isolierung eines Psymberin KS Genfragments

[0088] In einem ersten Schritt wurde zunächst die Komplexität der DNA-Probe mittels einer ersten PCR-Amplifikation erniedrigt. Hierbei wurden degenerierte Primer eingesetzt, die bereits zuvor für die Isolierung der Gene von Pederin und Omnamid genutzt wurden (fiel et al., Antitumor polyketide biosynthesis by an unculti-

vated bacterial symbiont of the marine sponge *Theonella swinhoei*. PNAS 101, 16222–16227; 2004; Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. PNAS 99, 14002–14007; 2002).

[0089] Diese verwendeten Primer waren für die Aminosäuremotive KSDPQQF bzw. KSHGTGR spezifisch. Der forward-Primer hatte die Sequenz 5'-MGNGARGCNNWNSMNATG GAYCCNCARCANMG-3'; der reverse-Primer hatte die Sequenz 5'-GGRTCNCNA RNSWNGTNCNGTNC CRTG-3'.

[0090] Die PCR fand unter folgenden Bedingungen statt: 96°C für 45 s, 55°C für 45 s, 72°C für 1 min (40 x), dann 72°C für 10 min. Verwendet wurde Taq Polymerase (NEB).

[0091] Das Amplifikat wurde mittels Agarosegel aufgetrennt und die bei 700 bp laufende DNA mittels Gelverdau wie oben beschrieben aufgereinigt. 1 µl dieses aufgereinigten PCR Amplifikats wurde als Template für eine zweite PCR-Runde verwendet.

[0092] Für die zweite PCR-Runde wurde als Ziel die Acetyl-abgeleitete Startereinheit ausgewählt, die für gewöhnlich durch KS verlängert wird, die zu Klade VI gehören. In einem zusätzlich durchgeführten Alignment, das alle veröffentlichten Klade VI Sequenzen enthielt, wurde das einzigartige Sequenzmotiv YYQ/KAGML (Klade VIa-Motiv, [Fig. 3](#)) in allen KS gefunden, die einem Modul mit einer GCN5-related N-acetyltransferase (GNAT) Domäne folgen. Dieses Motiv wurde für das Primerdesign ausgewählt, da gezeigt werden konnte, dass GNATs über Decarboxylierung von Malonyl-CoA Acetyl-Startereinheiten spezifisch in Polyketide einbauen.

[0093] Die verwendeten Primer waren spezifisch für das EDGAY-Motiv, sowie für das YYQAGML-Motiv. Der forward-Primer hatte die Sequenz 5'-GCNHTNGARGAYG CNGGNTAYGC-3' (SEQ ID NO: 72); der reverse Primer hatte die Sequenz 5'-CANRCATNCCNGCY TKRTARTA-3' (SEQ ID NO: 73).

[0094] Die Reaktionsbedingungen waren: 95°C für 300 s, 95°C für 60 s, 52°C für 60 s und 74°C für 100 s (35 X), 74°C für 10 min. Es wurde Taq Polymerase (NEB) verwendet.

[0095] Die Amplifikation ergab ein einzelnes PCR-Produkt mit der erwarteten Größe, dessen Sequenzierung ergab, dass es in der Tat in die Acetyl-spezifische Klade VIa gehörte.

[0096] Diese Resultate belegen, dass mittels des erfindungsgemäßen Verfahrens einzelne PKS-Gene in komplexen Metagenomen mit einer außergewöhnlichen Genauigkeit ausgewählt und amplifiziert werden können.

Beispiel 4: Isolierung des Psymberin PKS-Genclusters

[0097] Eine Fosmid-Library aus Gesamt-DNA des Psymberin-positiven Schwamms *P. aff. bulbosa* wurde erzeugt (410.000 Klone) und nach einer vorbekannten Strategie (Hrvatín & Piel, Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. J. Microbiol. Methods 68, 434–436; 2007) auf positive Klone hin durchmustert.

[0098] Die Durchmusterung wurde mittels PCR und unter Verwendung von spezifischen Primern durchgeführt, die aus dem Amplikon der Klade VI abgeleitet wurden.

[0099] Innerhalb von drei Runden des „Primer Walkings“ konnten so fünf positive Fosmide isoliert werden. **Fig. 5A** und **Tabelle 1** geben das Ergebnis der Sequenzierung (SEQ ID NO: 39) dieser fünf Fosmide wieder. Der isolierte Gencluster zeigte eine typisch bakterielle Architektur: es wurden keinerlei Introns detektiert, Shine-Dalgarno-like Sequenzen befanden sich vor den einzelnen Genen und die enge Nachbarschaft der Gene lässt auf eine polycistronische mRNA schließen.

Tabelle 1:

ORF	Proteingröße	Vorgeschlagene Funktion	SEQ ID NO
ORF1	302	Zn-dependent hydrolase	1,20
ORF2	426	Adenylosuccinate synthetase	2,21
ORF3	367	Transposase	3,22
psyA	3297	PKS	4,23
psyB	331	Methyltransferase	5,24
psyC	343	PedK-like	6,25
ORF4	367	Transposase	7,26
ORF5	491	Transposase	8,27
psyD	12644	PKS-NRPS	9,28
psyE	571	Phosphoenolpyruvate synthase $\beta + \gamma$ subunit	10,29
psyF	497	Phosphoesterase-like	11,30
psyG	309	Phosphoenolpyruvate synthase subunit	12,31
psyH	369	Acyltransferase	13,32
psyl	440	HMG-CoA-synthase	14,33
psyJ	254	Crotonase superfamily	15,34
psyK	367	Flavin-dependent oxygenase	16,35
psyL	81	ACP	17,36
psyM	429	3-Oxoacyl ACP synthase	18,37
psyN	709	Cation transport ATPase	19,38
ORF6		Unbekannt	
(teilweise)			

[0100] Wie aus z. B. **Fig. 5A** zu entnehmen ist, umfasst die isolierte Genregion (SEQ ID NO: 39) in ihrer zentralen Region einen psy-Gencluster (psyA bis psyN) von 62238 bp Länge. Die großen Gene psyA und psyD codieren für PKS mit drei bzw. zehn Modulen, wobei letzterer mit einer Thioesterasedomäne abschließt (siehe **Fig. 5B**). Damit wird zum ersten Mal die vollständige Sequenzinformation eines Polyketidbiosyntheseclusters offenbart.

[0101] Hinsichtlich der Architektur der PKS-Domänen sind PsyA und die ersten fünf Module von PsyD nahezu identisch zu den Omnamid- (OmnB und OnnI) und Pederin- (PedI und PedF) Biosynthesewegen (**Fig. 5B, C**).

Beispiel 5: Isolierung eines Mycalamid KS Genfragments

[0102] Erneut wurde in einem ersten Schritt zunächst die Komplexität der DNA-Probe mittels einer ersten PCR-Amplifikation erniedrigt, wobei die verwendeten Primer ebenfalls die Sequenzen 5'-MGNGARGCNN-WNSMNATGGAYCCNCARCANMG-3' (forward) und 5'-GGRTCNC CNARNSWNGTNCNGTNC CRTG-3' (reverse) hatten.

[0103] Die erste PCR-Runde fand unter folgenden Bedingungen statt: 96°C für 45 s, 55°C für 45 s, 72°C für 1 min (40 x), dann 72°C für 10 min. Verwendet wurde Taq Polymerase (NEB).

[0104] Das Amplifikat wurde mittels Agarosegel aufgetrennt und die bei 700 bp laufende DNA mittels Gelverdau wie oben beschrieben aufgereinigt. 1 μ l dieses aufgereinigten PCR Amplifikats wurde als Template für eine zweite PCR-Runde verwendet.

[0105] Für die zweite PCR-Runde wurde als Ziel ein Motiv verwendet, welches in KS detektiert wurde, die $\alpha\beta$ -gesättigte Intermediate verlängert (Klade V). Das identifizierte Motiv hat die Consensussequenz EPI/VE/DTAC (s. **Fig. 4A**).

[0106] Die verwendeten Primer waren spezifisch für das EDGAY-Motiv, sowie für das EPI/VE/DTAC-Motiv. Der forward-Primer hatte die Sequenz 5'-GCNHTNGARGAYG CNGGNTAYGC-3' (SEQ ID NO: 72); der reverse Primer hatte die Sequenz 5'-RCANGCNGTNTCDATNGGTTTC-3' (SEQ ID NO: 74).

[0107] Die Reaktionsbedingungen waren: 95°C für 105 s, 95°C für 60 s, 51°C für 60 s und 74°C für 100 s (35 X), 74°C für 10 min. Es wurde Hot Start Taq Polymerase (Jena Biosciences) verwendet.

[0108] Die Amplifikation ergab erneut ein einzelnes PCR-Produkt mit der erwarteten Größe von 211 bp. Weiterhin zeigte eine anschließend durchgeführte Sequenzierung ebenfalls, dass lediglich ein einzelnes KS-Genfragment amplifiziert worden war, welches perfekt in die vorausgesagte Klade passte (SEQ ID NO: 75).

[0109] Diese Resultate belegen erneut, dass mittels des erfindungsgemäßen Verfahrens einzelne PKS-Gene in komplexen Metagenomen mit einer außergewöhnlichen Genauigkeit ausgewählt und amplifiziert werden können.

Beispiel 6: Isolierung des Mycalamid PKS-Genclusters

[0110] Das oben unter Beispiel 4 beschriebene Durchmusterungsverfahren einer Fosmid-Library aus Mycalamid-positiven Proben des Schwammes *M. hentscheli* wird mit Primern wiederholt, die für das in Beispiel 5 erhaltene KS-Genfragment spezifisch sind.

[0111] Aus den detektierten positiven Fosmiden wird die Gensequenz des gesamten PKS-Genclusters bestimmt, der für die Mycalamidbiosynthese benötigt wird.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Patentliteratur

- US 5854033 [\[0028\]](#)

Zitierte Nicht-Patentliteratur

- Blunt et al., Marine natural products. Nat. Prod. Rep. 25, 35–94; 2008 [\[0002\]](#)
- Sipkema et al., Marine sponges as pharmacy. Mar. Biotechnol. 7, 142–144; 2005 [\[0002\]](#)
- Fiel et al., Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. Proc. Natl. Acad. Sci. U.S.A. 101, 16222–16227; 2004 [\[0002\]](#)
- Kim et al., Diversity of polyketide synthase genes from bacteria associated with the marine sponge Pseudoceratina clavata: culture-dependent and culture-independent approaches. Environ. Microbiol. 8, 1460–1470; 2006 [\[0002\]](#)
- Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C. and Piel J. Nat. Biotechnol., 2008, 26, 225–233 [\[0013\]](#)
- Fischbach & Walsh. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery and mechanisms. Chem. Rev. 106, 3468; 2006 [\[0036\]](#)
- Cheng et al., Type I polyketide synthase requiring a discrete acetyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 100, 3149–3154; 2003 [\[0038\]](#)
- Cheng et al., Type I polyketide synthase requiring a discrete acetyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. U.S.A. 100, 3149–3154; 2003 [\[0042\]](#)
- Murray, M. G. and W. F. Thompson (1980). Rapid isolation of high-molecular-weight plant DNA. Nuc. Acids Res. 8: 4321–4325 [\[0081\]](#)
- Fiel et al., Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. PNAS 101, 16222–16227; 2004 [\[0088\]](#)
- Piel, A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. PNAS 99, 14002–14007; 2002 [\[0088\]](#)
- Hrvatin & Piel, Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. J. Microbiol. Methods 68, 434–436; 2007 [\[0097\]](#)

Patentansprüche

1. Nukleinsäure umfassend zumindest ein Biosynthesegen für ein Polyketid ausgewählt aus der Gruppe bestehend aus Psymberin oder Mycalamid.

2. Nukleinsäure gemäß Anspruch 1, wobei die Nukleinsäure eine Sequenz umfasst die ausgewählt ist aus der Gruppe bestehend aus SEQ ID NO 1–19, 39, 75, einem Homolog dieser Sequenzen und/oder dem Komplement der Sequenzen oder des Homologs, wobei das Homolog zumindest 75% Sequenzidentität aufweist, oder unter Hochsalzbedingungen mit der Nukleinsäure hybridisiert.

3. Nukleinsäure gemäß Anspruch 1–2, wobei die Nukleinsäure für ein Protein codiert, das zumindest eine Identität zu SEQ ID NO 23 von 41%, zu SEQ ID NO 24 von 52%, zu SEQ ID NO 25 von 42%, zu SEQ ID NO 28 von 44%, zu SEQ ID NO 29 von 94%, zu SEQ ID NO 30 von 96%, zu SEQ ID NO 31 von 96%, zu SEQ ID NO 32 von 94%, zu SEQ ID NO 33 von 98%, zu SEQ ID NO 34 von 95%, zu SEQ ID NO 35 von 97%, zu SEQ ID NO 36 von 98%, zu SEQ ID NO 37 von 96% oder zu SEQ ID NO 38 von 96% aufweist.

4. Peptid für das eine der Nukleinsäuren der vorstehenden Ansprüche kodiert.

5. Peptid ausgewählt aus der Gruppe bestehend aus SEQ ID NO 20–38 oder einem Homolog dieser Sequenzen, wobei das Homolog zumindest eine Sequenzidentität Identität zu SEQ ID NO 20 von 36%, zu SEQ ID NO 21 von 54%, zu SEQ ID NO 22 von 44%, zu SEQ ID NO 23 von 41%, zu SEQ ID NO 24 von 52%, zu SEQ ID NO 25 von 42%, zu SEQ ID NO 28 von 44%, zu SEQ ID NO 29 von 94%, zu SEQ ID NO 30 von 96%, zu SEQ ID NO 31 von 96%, zu SEQ ID NO 32 von 94%, zu SEQ ID NO 33 von 98%, zu SEQ ID NO 34 von 95%, zu SEQ ID NO 35 von 97%, zu SEQ ID NO 36 von 98%, zu SEQ ID NO 37 von 96% oder zu SEQ ID NO 38 von 96% aufweist.

6. Vektor umfassend eine Nukleinsäure gemäß Anspruch 1–3.

7. Rekombinante Wirtszelle umfassend eine Nukleinsäure gemäß Anspruch 1–3 oder einen Vektor nach Anspruch 6.

8. Verfahren zur Herstellung eines Polyketids umfassend den Schritt: Kultivierung einer rekombinanten Wirtszelle gemäß Anspruch 7 unter Bedingungen, die die Expression der Biosynthesegene und Produktion des Polyketids durch die Wirtszelle erlauben.

9. Verfahren gemäß Anspruch 8, weiterhin umfassend den Schritt der Isolierung des produzierten Polyketids.

10. Verfahren zur Isolierung zumindest eines Polyketid-Biosynthesegens, umfassend die Schritte:

- a) Bereitstellen von zumindest einem Primerpaar, das auf der Substratspezifität einer Kethosynthesedomäne einer trans-AT Polyketidsynthase basiert; und
- b) Isolierung von zumindest einem Polyketid-Biosynthesegen unter Verwendung des zumindest einen Primerpaars aus Schritt a).

11. Verfahren nach Anspruch 10, wobei Schritt b) die Amplifikation der vom Primerpaar eingeschlossenen Nukleinsäureregion umfasst.

12. Verfahren nach Anspruch 10 oder 11, wobei das zumindest eine Primerpaar spezifisch ist für ein Motiv das bei phylogenetisch benachbarten Kethosynthesedomänen von trans-AT Polyketidsynthasen konserviert ist.

13. Verfahren nach Anspruch 10–12, wobei zumindest ein Primer des Primerpaars spezifisch ist für zumindest 4 zusammenhängende Aminosäurereste eines der in [Fig. 3](#), [Fig. 4](#) oder SEQ ID NO: 40–71 definierten Aminosäuremotive.

14. Verfahren nach Anspruch 11–13, wobei, für die Amplifikation der vom Primerpaar eingeschlossenen Nukleinsäureregion neben dem zumindest einen Primerpaar aus Schritt a) ein weiteres Primerpaar verwendet wird, wobei die zwei Primerpaare verschachtelt sind.

15. Verfahren nach Anspruch 10–14, wobei das Polyketid-Biosynthesegen aus einem Metagenom isoliert

wird.

16. Verfahren nach Anspruch 10–15, wobei das Polyketid-Biosynthesegen in die Biosynthese von Psymberin, den Mycalamiden, den Onnamiden, den Theopederinene A, den Icadamiden, Pederin, den Laulimaliden, den Latrunculinen A, Mycothiazol, Pateamin A, den Spongistatinen 1, Discodermolid, Dictyostatin, Tedanolid und/oder Calyculin involviert ist.

17. Verfahren nach den Ansprüchen 10–16, weiterhin umfassend den Schritt der Isolierung des gesamten für die Polyketidsynthese benötigten Genbereichs basierend auf den Sequenzinformationen des in Schritt b) isolierten zumindest einen Polyketid-Biosynthesegens.

Es folgen 13 Blatt Zeichnungen

Anhängende Zeichnungen

FIG. 1

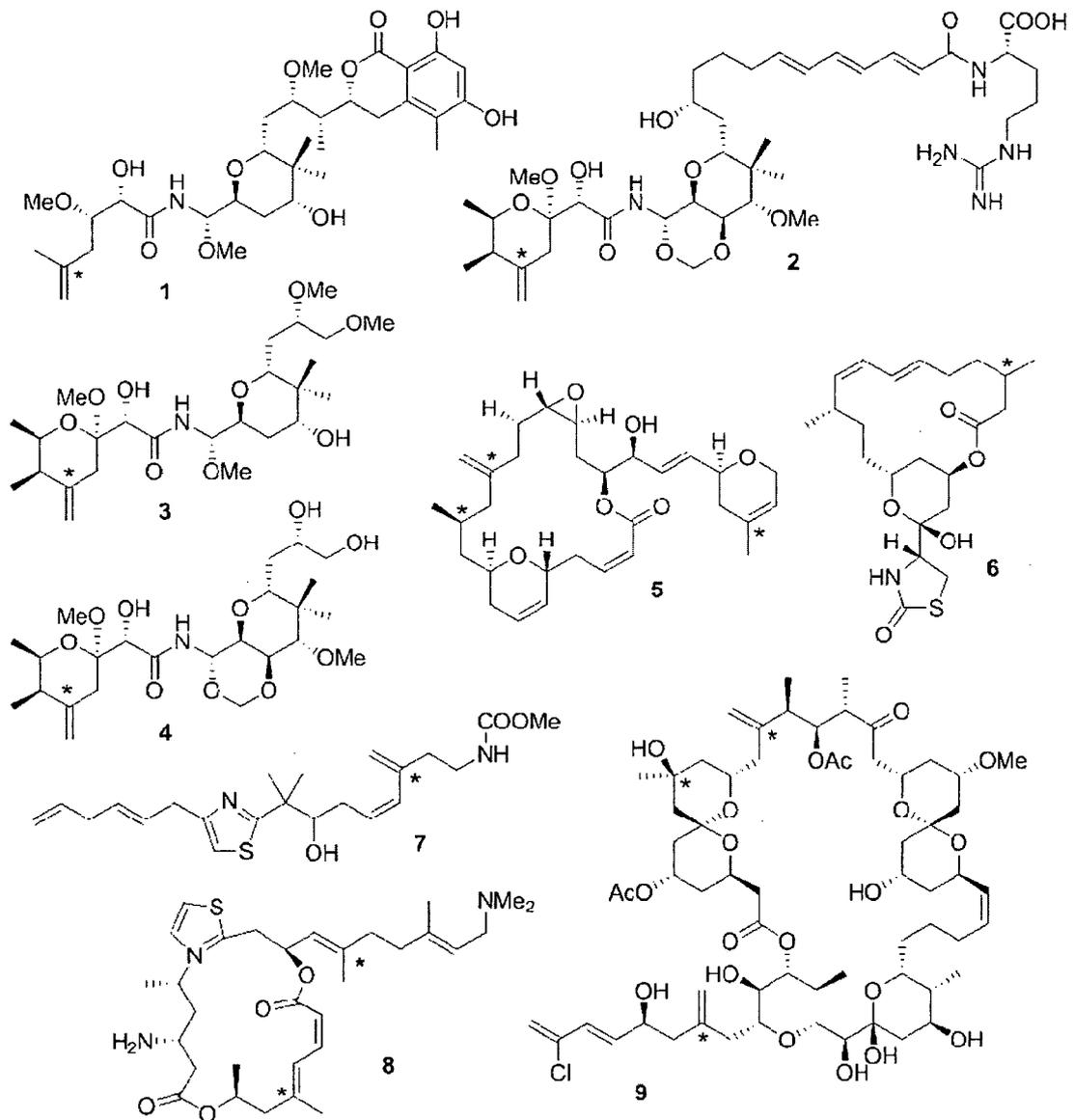


FIG. 2

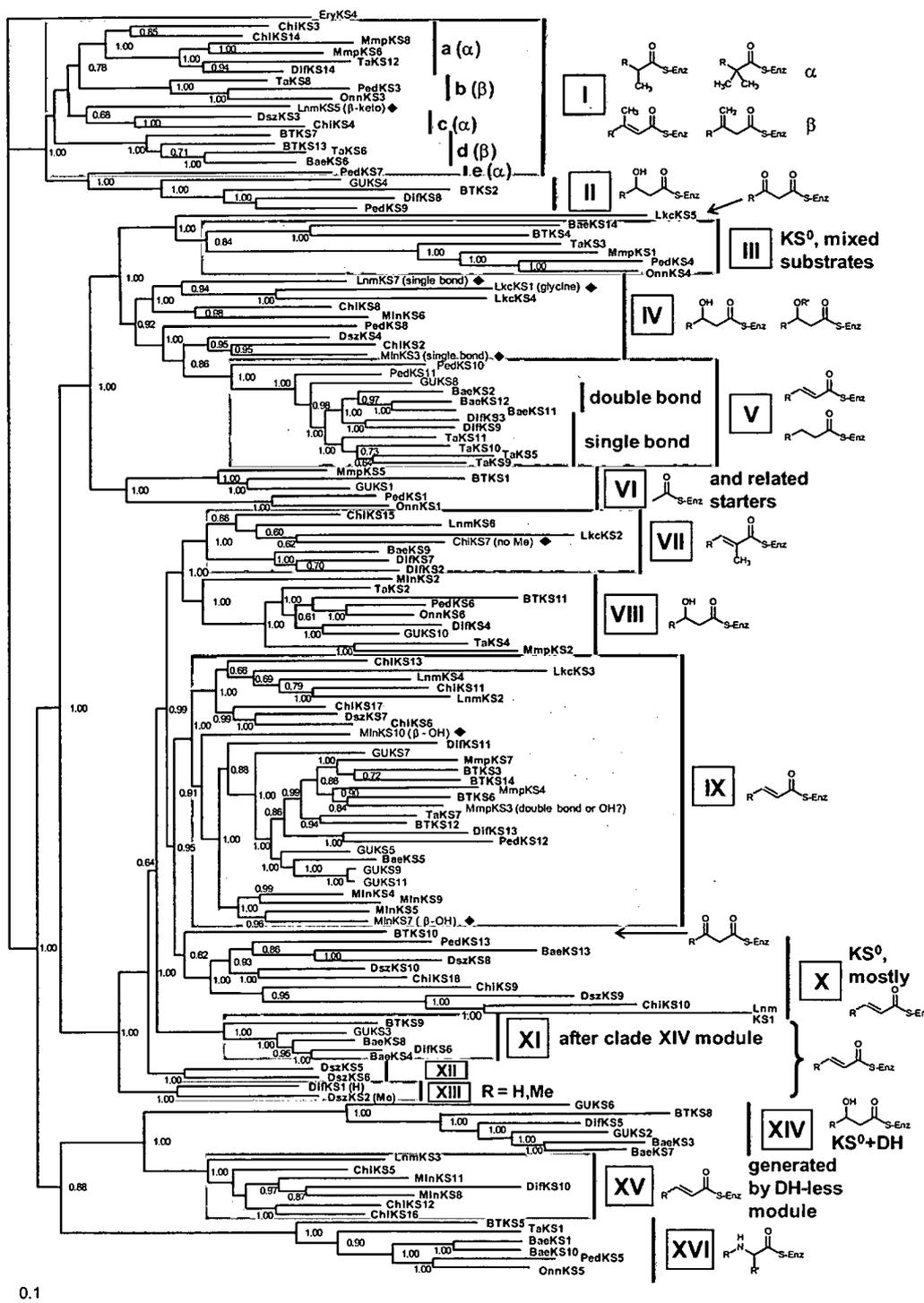


FIG. 3

		450	467	512	*	533	559	576				
V	PedKS11	YVWKA	VEDAGYA	PASLSG	FLNLHGP	SEP	VETACSSSLV	PQTHIS	FNKAGML	CSDGR
	DifKS3	HVWKA	IEDAGCS	AKSLSG	FLNIHGP	SEPI	ETACSSSLV	PDMHIS	FNKAGML	SEDGR
	DifKS9	YVWKA	IEDAGYS	AKRIAG	FLDIHGP	SEP	VETACSSSLV	PEGHIG	FQKAGML	SEDGR
	TaKS11	HVWRA	IEDAGYA	PHLSLG	LLNLHGP	SEP	VETACSSSLV	PGGHIA	FDKAGML	SEDGR
	TaKS10	YVWLAM	EDAGYS	ASSLSG	FLNIKGP	SEPI	ETACSSSLV	PAGHIG	FDKAGML	SPDGR
Vla	PedKS1	LCWQT	IEHAGYA	PDALAG	FYDFTGP	SLL	LDTACSSSLV	PANSIA	YYKAGML	AKDGL
	OnnKS1	LSWQV	LEDAGYA	PKDLKK	FYDFYGP	SLV	VDTACSSSLV	PANSLA	YYQAGML	AKDGL
	BP17KS1	LAWEA	IERAGHC	ADAVAG	TYDLRGP	SIQ	VDTACSSSLV	PGNTIA	YYKAGML	SPQGR
	KAKS2	LAWACL	EDAGIP	AKKLIG	YFDLSGP	SLL	VDTACSSSLV	PDL	SIAYHKAGML	SPEAK
	BryKS9	LSWQV	IEHAGY	KPSVFSG	FYNFKGP	SVA	IDTACSSVLV	PSDTV	SYQAGML	SKSGK
	PsyKS1	LAWSAF	EDAGYS	KDAVEG	FFDLQGP	SLL	IDTACSSSLV	PAMTLA	YYKAGML	SPDGR
S	SupA.Ts1	TSWQAL	EDAGMD	PDRLKG	VLGLMGP	PAK	AVDAACASSLV	GRVFEL	RADSMML	SPDGQ
	SupA.Aa1	TCWRA	LEDAGID	PDNLKG	ALGLEGP	PALA	LDTACSSSLV	GRLLEM	RANAGML	SPDGR
	SupA.Aa2	TSWQAL	EDAGID	PDVLKG	ALGLQGP	PAIA	IDTACSSSLV	GRLLEL	RANAGML	APDGR
	SupA.Pc1	TSWEAL	EDAGID	PDGLRE	VLGLRGP	PAK	AVDAACASSMV	GRIYEL	RADAMML	SPDGQ
	SupA.PKSA	TSWQAL	EDAGIA	PASLGG	TEGLEGP	AMA	VDTACSSSLV	PTVTES	FASGGML	APDGR

Fig. 4A

REGION 1:**Klade V****Einzel- oder Doppelbindungen**

```

PedKS11  -SSTGVVGSVGPNRMSYFLNHHGPESEPVETACSSSIVAIHRALSAM
baeKS2   AATGSMIPSTIGPNRLSYLLNHHGPESEPIETACSSSIVAIHRAVSAI
baeKS12  -SAANMSPSAGPNRVSYMLNHHGPESEPIDTACSSSIVAIHHAVCAI
baeKS11  -SAANTSPSVGPNRVSYTLNHHGPESEPIDTACSSSIVAIHHAVSSI
difKS3   -SAAGTVPSIGPNRMSYFLNHHGPESEPIETACSSSIVAVHRAVRAI
difKS9   -TMTGIVPSVGPNRMSYFLDIHGPESEPVETACSSSIVAIHRAVKAI
TaKS11   -SSTAVVPSIGPNRMSYLLNHHGPESEPVETACSSSIVALHRAMEAI
TaKS10   -ASTAAVPSVGPNRMSYFLNIKGPSEPIETACSSSIVAVHKAMLAI
TaKS5    -TSAGLVPSMGPNRLSYLLDHHGPESEPVETACSSSIVALHRAMTSL
TaKS9    -TATGSVPSVGPNRLSYLLDHHGPESEPIETACSSSIVAVHRALLAI
Motiv      ---TGxVPSVGPN-----
alt. Motiv ---.....I.....
Motiv2     -----EPIeTAC-----

```

Klade VIa**Acetyl-Starter mit GNAT Domäne im upstream Modul**

```

Ped KS1   HYGTGSSMAVLANRISYFYDFTGPESLLLDTACSSSIVAVHKAVQSI
Onn KS1   HFGTGSAMAVLANRISYFYDFYGPSLVVDTACSSSIVAVHKAVQSL
BP17 KS1  HVATGASMAVIANRISYTYDIRGPESIQVDTACSSSIVALHQAVQAL
KA KS2    HHAVGSSIALILANRLSYFYDISGPESLLVDTACSSSIVAIHSAVQSL
Bry KS9   YLMTGSMLSFLPNRISYFYNFKGPESVAIDTACSSVIVAVDQAVYAI
Psy KS1   VMGTGTAVAVLPNRISYFYDIQGPESLLIDTACSSSIVAIHEAVQAL

```

Klade VIb**Häufig nicht-Acetylstarter, GNAT Domäne abwesend**

```

Mmp KS5   HHSTGNAAAVIANRISHFYDHHGPESVLVDTACSGSLSAIHHAVQSL
GU KS1    YHSTGTAAAVIANRISHYHHKGPSLAIDTACSSSINAIHSAAQSL
Etn KS1   YGSTGTSGISLNRVSFFDFRGPESVTVDTACSSSIVAVHLAARAI
Tai KS1   YYASGVAASVANRLSYWFDRGPSVTVDTACSGSIYALHLAREAL
Motiv      ---TGxAXAVIAN-----
alt. Motiv ---.....L.....

```

Fig. 4B

Klade XIV**Doppelbindungen**

GUKS7	NPILGIGQNYLATNISRFFNITGPESLVLDTACSSGLTAISMASDTL
BTKS8	HPIVAVGQNYLAANLSRHFDEINGACALVDTACSSATVAMHSAVLAL
difKS5	NPIVAAGPNYLAANLSOFFDITKGEISLVLDTACSSAIVGMNMAIQAL
GUKS2	RNPTLVGQNFMAANI SHFFDITRGEISVVIDTACSSAILAMNMGIQAL
baeKS3	NPIVALGQNYLAANLSQYFDMRGEISVVLDTACSSAIVGMNMAVQAL
baeKS7	NPIVAGGQNYLAANVSOFFDITRGEISVLDTACSSAITGMNMAVQAL
Motiv	-----VAXGQNYLAAN-----

Klade XV**Nicht-kanonische Erzeugung von Doppelbindungen**

LnmKS3	LVGMGSDQNEIAARIAHFDITRGENLVVDSAGSSSIVLALQACRSL
ChiKS5	SVMASTGQNEIAAHI SHVFDITRGEISLLVDTACSSSIVAIHLACHSL
MlnKS11	SHIVGLSQNEIASHVSHFFNFHGEISMVSDTACSSSIVSIHLACQSI
difKS10	DAIIGAGQNEITAHMAHIYNEITGENLVVDTACSSSITAIHLAVQAI
MlnKS8	NTIVGIGQNEIGAHI SHFFNITRGENLVVDTACSSSIVSVHLACQSL
ChiKS12	ASIVSGQNEIAAHLSHALNITRGEISLVVDTACSSAIVGISLACQSL
ChiKS16	NGIPGVAQNEVAAALLAHFLDITRGENLVVDTACSSSIVGIHLACQSL
Motiv	-----GXGQNEIAAH-----

Klade XVI**Aminosäuren**

BTKS5	VGWILEQGGTIPALISYKLGITRGEISLYVHTNCSSSI AALYA AWQTI
TaKS1	VLWVLAQAGSIPTMVSYKLGITRGEISLFVHTNCSSSI SALYVAQQAI
baeKS1	VSWVLAQSGTIPTMISHKLGITRGEISYFVHANGSSSI IGLHSAFQSL
baeKS10	VSWVLAQSGTIPTMVSHKLGITRGEISYFVHSNCSSSI VGLYSAYKSI
PedKS5	AAWIAGQGGTIPTMISYQLGTRGEISFAVHSNCSSSI VGLYLASQCL
OnnKS5	AAWIASQSGTIPTMISYQLGTRGEISAFVHTNCSSSI SGLYFAVQSL
Motiv	-----QSGTIPTM-----

Fig. 4C

Vergleich mit der sup Klade und anderen Kladen

supA1	YALSGTNLNGTSGRVSFVLGIMGPAKAVDAAGASSMVSVHDAVADL	} sup KS
supA2	YALSGTNLNGASGRVSFVLGIRGPAKAVDAAGASSMVSVHDAVADL	
supA4	YTVSGTSYNTAIGRVAFALGIRGPAIAIDTACSSSIIVAIHQAVTGL	
PedKS3	QGLLGESVAMLPARVAVYLNMQGPECLAIDTACSSSIIVALASACDSL	} Klade I
OnnKS3	QSLGCVVSVILPARISYFLNMQGPECLAVDTACSSSIITAIANACDSL	
MmpKS8	QAFWGNASSIIPARIAMHLDMQGPAAIVDTACSSSIIVAIHLACQAL	
difKS14	QAFWGNHNSVTPARIAMHLNMQGPAITVDTACSSSIITAIHLACQGL	
difKS8	-NVTSHHNGILAAFLAMHLNMQGPMVAINTACSSSIVSVHQAVLSL	} Klade II
GU KS4	-LINSNQATLSARIAMMLDMKGPENLALTAAGSSSIVAVHQACQAL	
PedKS3	-SLTSMHNGILASRLAMFLNMQGPMVAINTACSSSIVAVHQACQSL	
BT KS4	FSMTSMAPALMPNRISYALDMKGPSEMDDTACSSSIIVALHRAVRSL	} Klade III
MmpKS1	GEGLDQAASMLANRVSHFYDMKGPSERIDAMGAGGAVALHHAVTAL	
Onn KS4	GEGFGHAS-MIANRIAMFFDFRGPSEFVDAQGAGAAVALYRAVSTL	
LkcKS4	YTVTGNAHAMLANRVSHLLDFTGSPQTVDTACSSSIVALNRGVLSL	} Klade IV
PedKS8	QTPSEFPFMIANRISYWFNFGPSEAITACSSSIIVAVHKAIESL	
DszKS4	YALTGTPHSILANRISFLNMQGPESEPINTACSSSIVAVHRAAETL	
ChiKS15	-TGAQSAWYSIANRVSEFLNMQGPESLAVDTACSSSMTAIHLACESL	} Klade VII
LkcKS2	-QPPTSAFWSMANRASYHFDWHGPMVAVDTACSSSIITALHLACQSL	
difKS7	---TGAHYHFIPNRVSVVDFHGPSEIAVDTACSSSIITAIHLAAESI	
TaKS2	--RPYTSFASVANRVSYLLDMKGPSMPVDTMCSSSIITAVHMACEAL	} Klade VIII
BT KS11	--NPQTSFSALVNRVSYVLDLHGPSVPIDTMCSSSIITAVHEACEHI	
Ped KS6	--FPHTSFSSVANRVSYCLNMQGPMPIDTMCSSSIITAIHEACQHL	
ChiKS13	GPTLGSLYASVANRVSYFFNFAGPSVSLDTMCSSSIITAIHLACQAI	} Klade IX
GUKS8	TALVGSP-ASIANRVSYSFNFHGPSMAVDSMCSSSIITALHLACL CI	
PedKS12	-----HASVANRVSYVLDLHGPSMTVDTMCSSSIITALHLACQDL	
BTKS10	QFAAPSQIFSI SNRVSYVMNFTGPESLTVDTACSSSIYAIHLACESL	} Klade X
ChiKS18	-QLSSFPGSAIANRVSHYFDLMKGPSMLVDTACSSSIITAIYMACESL	
PedKS13	-LVTMSSHSSIANRVSYFFDFQGPSVAVDTMCSSSIIVAVHMACE SL	
BTKS9	PTPLAVNRAPVANRVSYVCFDFHGPSVAIDTACSSSIIVAVHLAAQSL	} Klade XI
GUKS3	VFPLSLSCALIANRVSYFCNFHGPSIAVDTVCSSSIIVAVHLALESI	
difKS6	VFPYPLNYAPIANRVSYFCNFHGPSMAVDTVCSSSIITAVHMAAESL	

Fig. 4E

Klade XI

Doppelbindungen

BTKS9	ASATNHVGAVA-GIMAPSPAAQAAALIGECLRRANVDARTIGYVEAHGTGT
GUKS3	GSAINHVGAVS-GVTVESPIAHADMIETCVEKTGIDPRTISYVEAHGTGT
baeKS8	GSAINHVGKVS-GITVESPAAQAEVIKACLKAGISPRTVSYVEAHGTGT
difKS6	GSAVNHVGTVS-GITVESPVAQGGDMIAECLKAGIDPRTISYVEAHGTGT
baeKS4	GSAVNHVGTVS-GISVESPVAQADLIEECLDKAGIDPRTVSYVEAHGTGT
Motiv	----NHVGtVS-GItvPSP-----

Klade XIV

Doppelbindungen

GUKS6	AIAVNNDGRTL-GPGSPNIDAQKEVMKEALELSRKKPEEVDYIEVNGGGS
BTKS8	GLAVNNDGRTA-GPSSNFAAQQAVMRRALAQSGLRADDVRHVEANGSGS
DifKS5	GISVNNDGRTA-GPATENIQAQKEVMKALAKSGRKAEDISYLEANGSGS
GUKS2	AVAVNNDGRTA-GPATENLQAQKEVMRTALERSGRGAEIISYLEANGSGT
baeKS3	AASVNNDGRTA-GPATPSLEAOKSVMKTALEKSGKQPEDITHIEANGSGT
baeKS3	AAAMNNDGRTA-GPSAPNMQAQKDVMSALFKSGKKPEDISYLEANGSGS
Motiv	----NNDGRTA-GPatPN-----

Klade XV

Nicht-kanonische Erzeugung von Doppelbindungen

LnmKS3	AVAVNNDGRTM-GLITPNPAAQAKVRRALAAAGRRADDEVGLIEAHGTGT
ChiKS5	ATAVNNDGRTM-GITTPNPDAAQARVVSAAALRRAGISARAITYVEVTHGTGT
MlnKS11	ASAINNDGATM-GITTPNPKLQEQVIQDAFEKADITPGSVSYVEVTHGTGT
difKS10	GSAINNDGSTM-GVTTPNPRAQODLIEQAVQSAEIEPDTMTYVEVTHGTGT
MlnKS8	SSAVNNDGQTM-GMTTPNPKAQSAVIKKALKKGGIDADTLSYVEVTHGTGT
ChiKS12	GIAVNNDGRTM-GISTPNPEAQSAVIEEALS RAGAAAETLGYVEAHGTGT
ChiKS16	GSAVNNDGRTM-GVTTPNPDAAQGVIEQALARAGAGADTVSYVEAHGTGT
motif	-----TM-GiTPNP-----

Fig. 4F

Klade XVI**Aminosäuren**

BTKS5	DVALNNDGAAKAGFYAPSVRCQ	QAQVIDALLRRTGVRAADIVY	VEAHGTGT
TaKS1	KVGINNDGODKVGIFYAPSATCQ	AEVIRRLFDRTGIDPASIGY	VEAHGTGT
BaeKS1	GIGVNNDGADKVGIFYAPSVKQ	AEVIQKVIDQTGIHPETIAM	VEAHGTGT
BaeKS10	GIGLNNDGADKVGIFYAPSVKQ	TDVIQHVLDSTNIHPETISY	VEAHGTGT
PedKS5	GVAVNNDGSDKVGIFYAPSVNCG	AAVIQKALDITGVDPQSVAY	VEAHGTGT
OnnKS5	GISLNNDGSDKTFIFYAPSVKQ	SEVIGKVLRAATNVDPSSISY	VEAHGTGT
Motiv	-----DKvGFYAPSVxG-----		

Fig. 4G

Vergleich mit der sup Klade und anderen Kladen

supA1	GSAVNHGGGA-SIGLTVPHTPALEQVIETALSQAGVPASDVDVLEAHGTGT	} sup KS
supA2	GAAVNHGGGA-SVGLTVPNTPALEEVMETALS DAGISPLEMDVLEAHGTGT	
supA4	GSALNODGA-STGLTVPEEEAQEQVILDALRRSGVSPSQVDVVEAHGTGT	
ChiKS3	GSGVNODGRTN-GISAPSTLSQTELEVSVYERSGVSPATVGYVEAHGTGT	} Klade I
MmpKS6	GSGINODGASN-GITAPSSALSQORLHESVYARYGIDPATVQLVEAHGTGT	
PedKS7	ASGINODGASN-GITAPSGIAQQQLITDVYRRYAIDPRRITYVEAHGTGT	
BTKS2	ASGVNYDGRTN-GMTAPSSGLAQARLLTDAWRTAGIDPAELGYVVAHGTGT	} Klade II
DifKS8	ASGINYDGKTN-GITAPSGVAQTRLLTSVYDRGRINPEDIEYIVTHGTGT	
PedKS9	GSGINYDGKTN-GITAPSGASQTELELGXYRQCALQPODISYIVTHGTGT	
LkcKS4	AVAENHDGRSN-SLTA PNPDAQVALLNRVYTPELAER--VSHIEAHGTGT	} Klade IV
ChiKS8	ATAVNHGGRAA-SLTA PNAQAQADLLVDAYARAGVSPETVTLIETHGTGT	
PedKS8	GSGENHGGRSA-SPTAPNGNAQKRLLVDIYSRADIDPRTISYIEAHGTGT	
PedKS10	GSAENHGGMST-SLTA PNPKAQASLIVEAHRKAKVDPRSIGYIECHGTGT	} Klade V
GUKS9	GTAENHGGGRAN-SLTA PNPVAQADLLTSAYTTAGIDPRTVGYIEAHGTGT	
baeKS2	GTAENHGGGRAN-TLTS PNPKLAQADLLVKAYRKAGVDPSTVTYIEAHGTGT	
MmpKS5	GSAVNHCGKTH-TLTY PPSAAQAQVIEQALGDAHIPASSVSYIEAHGTGT	} Klade VI
BTKS1	GSAVNHSGRTY-SLTY PDADAQARVIRAAFDAAGVSPAQVSYVEAHGTGT	
OnnKS1	GTSCNHGGLAS-GLTV PNPPEQAALLOQAWRDARISPLELSYIEAHGTGT	
MlnKS2	GTAINHGGKTN-GYTV PNPAAQADLIQGALKKAAGWSARSVSFIEAHGTGT	} Klade VIII
TaKS2	STSVNHGGKTN-GYTV PNPQAQAAALIRSALDKAGLSARDVSYIEAHGTGT	
BTKS11	ATGINHGGKVN-GFTV PNPNAQRELIRATLERAGVRARRVSYVEAHGTGT	
ChiKS13	GTAVNHGGGRAS-GFTV PNPSPKAQGDCEIEALRKAGVAARSISYIEAHGTGT	} Klade IX
GUKS8	GSMINSGGKTS-GYTV PNPAAQGEVIREAIERAGFNARDISYIEAHGTGT	
PedKS12	SSVLNHGGKTH-GFSV PNPKAQOHLISRALREAEVDPRAITYVEAHGTGT	
BTKS10	GSGVSHGGRTN-GFAV PNPSPDAQALAIRRAVAQAGVAPRSVGYVEAHGTGT	} Klade X
ChiKS18	GGFVNHGGRTH-GATV PNPSPAQADLIVEAFRRAGVRPDAVSYIEVAANGS	
PedKS13	STTINHSGQSN-GYFV PNPAAQTELMVSNFTKAGIDPRTLSYVESAAANGS	

Fig. 4H

Region 3:

Klade II

β-hydroxyl groups

GUKS4	GLVPAEAVAAVVVLRPLSKAVADKDRHYGCDKASGVNYDGKTNGLIATPLSQAEL
BTKS2	GFVPGEAVVVVALKRLADALADGDATHGVIIRASGVNYDGRITNGMTAPSSGLAQARL
DifKS8	GMVPGEAAAAVVLRPLREAEKDGDPHYGVIRASGINYDGKTNGLITAPSSGVAQTRL
PedKS9	GLVPGEAVAAVVVLRPLSKALADGDPIKALIRGSGINYDGKTNGLITAPSSGASQTEL
Motiv	-----AvavVvLK-----

Vergleich mit der sup Klade und anderen Kladen

SupA1	GYVRGEGCVVILKRLISEAEADGNRIWGVIRGSAVNHGGASI-GLTVPH	} sup KS
SupA2	GYVRGEGCVVVVLRKRLISEAEADGDRIWAVIRGAAVNHGGASV-GLTVPT	
SupA4	GYVRGEGCGIIVLKRRLREAEADGDRIWGVIRGSALNODGAST-GLTVPE	
ChiKS3	GFVPGEVGVVVVLRPLHAALRDGDRILGIIRKSGVNOGDRTN-GISAPST	
MmpKS6	GFVPGEAGAVVLRKRLADALADRDHTHGLIRGSGINODGASN-GITAPSA	} Klade I
PedKS7	GMVMSEGVGMVVLRKRLDQALSDDHTHYGVIRASGINODGASN-GITAPSG	
BaeKS14	GFVRGEGAAALVLRPLAAARRDGDFFVYACIKGTAVHHGGRGA-ALTA	} Klade III
MmpKS1	GYLRGEGVCSLLKPLSKAEADGDHTYGLIRNTAVNYNGGDAASIAAPSV	
PedKS4	GHLRAEGVCSLLKPLTKALADGDPIYASIKHSAVNFNGGGASIAAPNV	
LkcKS1	GYVRGEGYGTFLVKRLADALADGDRIAVVESVAENHGGGRAN-SLTA	} Klade IV
ChiKS8	GFVKGEGVGVVLRKPLERAVLDGDRILAVIRATAVNHGGGRAA-SLTA	
PedKS8	GYVRSEGVAILLKPLRKAIAADNRIHGLIRGSGENHGGRSA-SPTAPXG	

Region 4:

Klade Ib

β-branches

TaKS8	MDPQQRFLAACWHCIEDAAIRPSSLSLTR-----CGVFEVCGPTDYGRDL
PedKS3	MDPQQRFLFLNSWHCIEDAGYDPTRLSG-----SLCGIFEVCGAASDYSQLA
OnnKS3	MDPQQRFLAQACWSCIEDAGYNPKTLSG-----SSCGVFEVCGDMGDYGRSV
Motiv	-----CGvFVGC-----

Klade XI

Doppelbindungen

BTKS9	MDPQERVFLTAWAAIEDAGHTPDTLARRAPGGGDAGAPVGVFVGMQSEYALIQ
GUKS3	LDPQERLFLFETCWETIEDAGYTPSSLVAS-RGRHKRRD-VGVFAGVMHKDYTLLE
baeKS8	MDPQERLFLFETCWETIEDAGYTPETLGN--KG-EKQH-PIGVFAGVMHKDYSLIG
difKS6	KDPQERLFLFETCWETIEDAGYTPETIVQ-PKGNKRMN-VGVFAGVMHMDYSLAG
baeKS4	MDPQERLFLFETCWETIEDAGYTPKSLAE-ARGRNKRQR-VGVFAGVMHKDYTLVG
Motiv	-----FLETcwETIEDAGYTP-----

Fig. 4I

Klade XVI

Aminosäuren

BTKS5	MDPQARQLLQHAWLAFEDAGYVPADAPD-----	TAVEVSAHSRYAAKQ
TaKS1	MDPQFRLLLMHAWKAVEDAATTPERLGP-----	CGVFM T ASNSFYMHOGS
BaeKS1	MDPQLRMLLHLHSWKAIEDAGYISKEIPE-----	TSVYMSAS TNS YRSLL
BaeKS10	MDPQLRLLLHLHSWKAIEDAGYVSKEIPK-----	TSVYMSAS NNS YRSLL
PedKS5	MDPQYRVLLQQA WQA IEDAGYVAQDIPE-----	TAVE MS AS NNF YK TLL
OnnKS5	MDPQLRLLLTHSWQAMEDAGYCARDIPD-----	TAVE MS AC NSF YK TLL
motif	-----IPx-----	TxV F MSA-----

sup Klade

supA1	MDPCHRMML ETT WQALE DAG IEPDR LK GSL-----	TGVY TGI S NDE Y RML V
supA2	MDPQORMML EMS WEALEDAGIDPEGLKESR-----	TGVY TGI S NDE Y RML V
supA4	-----RLM LET SWQALE DAG IDP DAL K GSR -----	TGVY GGV S NNE Y RGL I

Fig. 5

