

SCHWEIZERISCHE EIDGENOSSENSCHAFT
EIDGENÖSSISCHES INSTITUT FÜR GEISTIGES EIGENTUM

(11) **CH** **699 253 B1**

(51) Int. Cl.: **C12Q** **1/68** (2006.01)

Erfindungspatent für die Schweiz und Liechtenstein

Schweizerisch-liechtensteinischer Patentschutzvertrag vom 22. Dezember 1978

(12) PATENTSCHRIFT

(21) Anmeldenummer: 01806/00

(22) Anmeldedatum: 18.09.2000

(24) Patent erteilt: 15.02.2010

(45) Patentschrift veröffentlicht: 15.02.2010

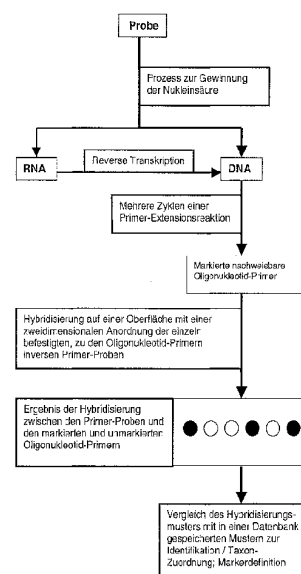
(73) Inhaber:
Eidgenössische Forschungsanstalt für Obst-, Wein- und
Gartenbau, Schloss
8820 Wädenswil (CH)

(72) Erfinder:
Jürg E. Frey, 8820 Wädenswil (CH)

(74) Vertreter:
E. Blum & Co. AG Patent- und Markenanwälte VSP,
Vorderberg 11
8044 Zürich (CH)

(54) Verfahren zur Charakterisierung und/oder Identifikation von Genomen.

(57) Die vorliegende Erfindung betrifft ein Verfahren zur Charakterisierung und/oder Identifikation von Genomen bzw. Zielorganismen, indem gleichzeitig die Präsenz oder Absenz einiger bis vieler Nukleinsäuresequenzen in einer Probe von einem biologischen Organismus bestimmt wird und das daraus resultierende Muster mit in einer elektronischen Datenbank gespeicherten Mustern unter Einsatz spezieller Clusteralgorithmen und statistischer Verfahren verglichen wird.



Beschreibung

[0001] Die vorliegende Erfindung betrifft ein Verfahren zur Charakterisierung und Identifikation von Genomen auf Nukleinsäurebasis. Das Verfahren ermöglicht die Identifikation von nukleinsäurehaltigen Organismen auf allen taxonomischen Ebenen.

[0002] Neben den etablierten Antikörper-basierenden diagnostischen Verfahren hat sich in den letzten Jahren die nukleinsäurebasierende Diagnose sowohl in der Medizin als auch in der landwirtschaftlichen Forschung zu einem neuen Standard entwickelt.

[0003] Die heute zur Verfügung stehenden Methoden der molekularen, nukleinsäurebasierenden Diagnostik arbeiten vorwiegend auf der Basis der Polymerase-Kettenreaktion (PCR; Saiki et al., 1986). Für eine zuverlässige und reproduzierbare Identifikation mit dieser Methode ist es notwendig, dass Informationen über kurze Abschnitte der Nukleinsäure jedes einzelnen zu identifizierenden Organismus vorhanden sind, die Primersequenzen. Da solche Primersequenzen für bisher genetisch wenig oder nicht untersuchte Organismen oft nicht bekannt sind, muss in der Regel für jeden Organismus eine Methodenoptimierung durchgeführt werden. Zur Charakterisierung von anonymen Genomen werden meistens zwei PCR-basierende Methoden eingesetzt, RAPD (random amplified Polymorphie DNA) und AFLP (amplified fragment length polymorphism). Die RAPD-Methode ist zwar sehr einfach durchzuführen, hat aber Probleme mit der Reproduzierbarkeit (Perez et al., 1998). AFLP ist demgegenüber sehr reproduzierbar, aber technisch anspruchsvoll, insbesondere, wenn nur kleine Mengen an DNA zur Verfügung stehen (Mueller and LaReesa Wolfenbarger, 1999).

[0004] Weitere diagnostische Methoden basieren auf der Microarray Technologie, bei der eine Vielzahl von einzelnen Analysen parallel auf einem zweidimensionalen Feld (Array) vorgenommen werden (Brown and Botstein, 1999). Die Technologie hat sich bewährt für Studien der Genexpression, für funktionelle Analysen und für das Genmapping (Lander, 1999). Diese Methode wird heute zur Charakterisierung von einzelnen Genotypen eingesetzt, wobei für diese Verwendung genau definierte DNA-Sequenzen der zu bestimmenden Genotypen verwendet werden.

[0005] In der landwirtschaftlichen Diagnostik steht der Verwendung der beschriebenen Methoden allerdings ein wichtiges Problem entgegen. Die grosse Vielfalt an gezüchteten Tierarten und -rassen und an schützenswerten Kulturpflanzen bringt eine unüberblickbare Zahl von zu identifizierenden Organismen (Taxa) mit sich, die unbekannte Genome besitzen (Frey and Frey, 1997) und mit den bestehenden Methoden nicht charakterisiert werden können.

[0006] Aufgabe der vorliegenden Erfindung war es deshalb, ein verbessertes Verfahren zur einfachen Charakterisierung und/oder Identifikation von Genomen zur Verfügung zu stellen.

[0007] Diese Aufgabe wird durch die Verfahren gemäss dem unabhängigen Anspruch gelöst.

[0008] Die Erfindung wird unten und in den Figuren näher beschrieben.

Fig. 1 zeigt ein exemplarisches Schema des Verfahrensablaufs.

Fig. 2 zeigt ein Clusterdiagramm als Ergebnis eines Versuches mit 12-mer Oligonukleotiden mit 70% G/C-Gehalt.

[0009] Die vorliegende Erfindung betrifft ein Verfahren zur Charakterisierung und/oder Identifikation von Genomen bzw. Zielorganismen, indem gleichzeitig die Präsenz oder Absenz einiger bis vieler Nukleinsäuresequenzen in einer Probe von einem biologischen Organismus bestimmt wird und das daraus resultierende Muster mit in einer elektronischen Datenbank gespeicherten Mustern unter Einsatz spezieller Clusteralgorithmen und statistischer Verfahren verglichen wird.

[0010] Dieses Verfahren hat mehrere Vorteile gegenüber existierenden Methoden zur genetischen Charakterisierung von unbekannten Genomen: Es sind keinerlei vorherige Kenntnisse über das Genom notwendig, und es wird nur wenig Ausgangsmaterial (DNA oder RNA) benötigt. Ausserdem ist sie einfach durchzuführen und kann zeitlich und finanziell ökonomisch organisiert werden.

[0011] Die vorliegende Erfindung ermöglicht es, die Präsenz oder Absenz einiger oder vieler verschiedener Polynucleotid-Teilsequenzen zu entdecken und dadurch ein zweidimensionales Muster zu generieren, das diagnostisch, also kennzeichnend ist für einen oder mehrere Organismen und durch den Vergleich mit in einer Datenbank abgelegten Mustern seine eindeutige Identifikation erlaubt.

[0012] Eine zweite Anwendung der Erfindung ist die Charakterisierung von genetischen Markern für phänotypisch nachweisbare Eigenschaften. Die grosse Zahl an anonymen Primern, mit welchen ein Genom gleichzeitig untersucht werden kann, erlaubt ein sehr effizientes Screening nach molekularen Markern für interessante Eigenschaften. Zum Beispiel können in Schadorganismen Marker für Gene gefunden werden, die eine hohe Resistenz gegen Pestizide verleihen. In Pflanzen können Marker für Resistenzgene gegen Schadorganismen oder für Qualitätsmerkmale gefunden werden.

[0013] Das Verfahren der vorliegenden Erfindung umfasst in einer beispielhaften Ausführungsform folgende Schritte:

[0014] Eine biologische Probe eines zu identifizierenden Organismus, zum Beispiel Blut oder eine Gewebeprobe, werden aufbereitet, um die Nukleinsäure der nachfolgenden chemischen Reaktion verfügbar zu machen. Dies geschieht zum Beispiel im Falle einer Gewebeprobe mit einer der etablierten Methoden zur mechanischen Zerstörung des Gewebes und

einer nachfolgenden Aufreinigung der Nukleinsäure. Falls die zu untersuchende Nukleinsäure eine RNA ist, wird diese zuerst in einer Reaktion mit einer reversen Transkriptase zu cDNA übersetzt.

[0015] Im nächsten Schritt werden mindestens ein Oligonukleotidprimer, bevorzugt bis zu einem Dutzend, bevorzugter bis zu 1000, noch bevorzugter bis 10 000 und am bevorzugtesten über 10 000 Oligonukleotidprimer zusammen mit einem Teil der aufgereinigten Nukleinsäure (jetzt DNA) der biologischen Probe in ein Reaktionsgemisch gegeben. Als Oligonukleotidprimer können Oligonukleotide mit einer zufälligen Folge von Basenpaaren und/oder Oligonukleotide mit einer Sequenz, die zu einer Zielsequenz der DNA in der Probe komplementär ist, verwendet werden.

[0016] Bevorzugt weisen alle Oligonukleotidprimer eine in gewissen Grenzen einheitliche Länge, einheitlichen G/C-Gehalt und einheitliche Schmelztemperatur auf, um Extension möglichst vieler Oligonukleotidprimer durch die Wahl geeigneter Bedingungen zu ermöglichen. Neben den für eine Oligonukleotid-Primer-Extensionsreaktion (auch Minisequencing Reaktion genannt) notwendigen Chemikalien enthält das Reaktionsgemisch ein oder mehrere markierte Didesoxynukleotidtriphosphate (ddNTPs). Falls mehrere unterschiedliche ddNTPs verwendet werden, zum Beispiel ddATP zusammen mit ddGTP, können die einzelnen ddNTPs mit unterschiedlichen Markern versehen werden. Falls alle 4 möglichen ddNTPs mit Fluoreszenzfarbstoffen markiert sind, vorzugsweise jedes einzelne ddNTP mit einem unterschiedlichen Fluoreszenzfarbstoff, kann die Methode zur Untersuchung von SNPs (Single Nucleotide Polymorphismus) verwendet werden. Als Alternative kann auch ein Gemisch aus ddNTPs und Desoxynukleotidtriphosphaten (dNTPs) verwendet werden, wobei entweder die ddNTPs und/oder die dNTPs markiert sind. In der Primer-Verlängerungsreaktion können auch Analoga der dNTPs und ddNTPs verwendet werden. Als Marker sind z.B. Chromophore, Fluorophore und radioaktive Stoffe geeignet. Vorteilhaft sind die ddNTPs oder dNTPs z.B. mit einem Fluoreszenzfarbstoff markiert.

[0017] Das resultierende Reaktionsgemisch wird auf eine Temperatur gebracht, bei der eine Hybridisierung der Oligonukleotid-Primer an komplementäre DNA-Abschnitte der Proben-DNA stattfinden kann. Diejenigen Oligonukleotid-Primer, für die eine komplementäre Zielsequenz auf der DNA vorhanden ist, hybridisieren mit der Zielsequenz. Sie dienen damit der ebenfalls im Reaktionsgemisch vorhandenen hitzestabilen Polymerase als Primer für eine Extensionsreaktion. In dieser Extensionsreaktion wird der Oligonukleotid-Primer mit einem markierten, bevorzugt fluoreszenzmarkierten Didesoxynukleotid verlängert, das komplementär zu dem auf der Zielsequenz der Oligonukleotid-Primer-Sequenz folgenden Nukleotid ist. Wenn ein Gemisch aus ddNTPs und dNTPs eingesetzt wurde, dann wird die Primer-Extensionsreaktion erst unterbrochen, wenn ein ddNTP in die sich verlängernde Primersequenz eingebaut wurde.

[0018] Durch Erhitzen werden die durch mindestens ein markiertes, bevorzugt fluoreszenzmarkiertes, Nukleotid verlängerten Oligonukleotid-Primer von der Zielsequenz gelöst. Durch erneutes Abkühlen auf die Hybridisierungstemperatur kann ein weiteres Set von Oligonukleotid-Primern an die entsprechenden komplementären Sequenzen auf der Ziel-DNA hybridisieren und die Polymerase kann auch an dieses Set das jeweils entsprechende markierte, bevorzugt fluoreszenzmarkierte, Didesoxynukleotid und/oder Desoxynukleotid anbauen. Dieser Vorgang kann mehrere Dutzend Male wiederholt werden, wodurch eine Signalverstärkung bewirkt wird, für jeden Primer mit einer passenden komplementären Zielsequenz nach der Regel (Anzahl Kopien der Zielsequenz mal Anzahl Zyklen). Wenn zum Beispiel in einer Proben-DNA für einen bestimmten Primer 1000 Kopien einer komplementären Zielsequenz existieren, dann werden durch die Extensionsreaktion nach 50 Zyklen circa 50 000 farbmarkierte Kopien des Primers erzeugt worden sein.

[0019] Nach der Markierungsreaktion wird für jeden Primer, der in die Reaktion gegeben wurde, untersucht, ob er in der Proben-DNA eine komplementäre Zielsequenz angetroffen hat und er deshalb eine Extension durch mindestens ein markiertes, bevorzugt fluoreszenzmarkiertes Desoxynukleotid und/oder Didesoxynukleotid erfahren hat. Dazu wird für jeden Primer ein seiner komplementären Sequenz entsprechendes Oligonukleotid, in der Folge Primer-Probe (PP) genannt, benötigt. Diese PP ist bevorzugt am 5'-Ende komplementär zum verwendeten Oligonukleotidprimer und weist an ihrem 3'-Ende eine bindungsvermittelnde Verlängerung auf, mit welcher sie an ein Substrat gebunden ist. Die bindungsvermittelnde Verlängerung ist oder enthält eine Verankerung, die z.B. ein Biotin-Molekül ist, welches eine stabile Bindung an ein Substrat ermöglicht. Bevorzugt ist zwischen der dem Oligonukleotidprimer komplementären Sequenz und der Verankerung z.B. ein Biotin-Molekül, ein Nukleotidschwanz vorhanden, um eine bessere Hybridisierung der PP mit dem entsprechenden Oligonukleotidprimer zu ermöglichen. Dieses Substrat kann zum Beispiel die Oberfläche eines Mikrotiterplatten-Well sein, die mit einer die Bindung ermöglichenden Substanz überzogen (gecoated) ist. Ein solches System ist zum Beispiel die Streptavidin-Biotin-Bindung. Ein an eine Oberfläche zu bindendes Oligonukleotid wird zum Beispiel an seinem 3'-Ende mit einem Biotin-Molekül versehen, das dann an eines an einer Oberfläche haftendes Streptavidinmolekül bindet (Affinitätsbindung), wodurch das Oligonukleotid fest mit der Oberfläche verbunden bleibt. Die PP für jeden in der Reaktion verwendeten Oligonukleotid-Primer kann z.B. in einem separaten Well der Mikrotiterplatte, an die Oberfläche eines Microarrays gebunden werden oder mit einer anderen Oberfläche fest verbunden werden. Vorteilhaft werden also genau so viele Oberflächen mit jeweils einem für eines der Primer-spezifischen PP bestückt, wie insgesamt unterschiedliche Oligonukleotid-Primer in der Reaktion waren. Wenn zum Beispiel 96 unterschiedliche Oligonukleotid-Primer in die Reaktion gegeben wurden, dann kann mit einer einzigen, Standard-96-Well-Mikrotiterplatte, die pro Well eine spezifische PP enthält, die Reaktion vollständig analysiert werden. Dazu wird z.B. in jedes der Wells eine Teilmenge der Reaktion gegeben, damit in einer Hybridisierungsreaktion die in diesem Well befestigte PP mit ihrem entsprechenden Oligonukleotid-Primer hybridisieren kann. Der Vorteil des Systems in Bezug auf die Hybridisierungsreaktion liegt darin, dass sich der Unterschied zwischen verschiedenen Oligonukleotid-Primern auf ein einziges Nukleotid (wenn in der Primerextensionsreaktion nur ddNTPs verwendet wurden) oder einige wenige Nukleotide (wenn ein Gemisch aus ddNTPs und dNTPs ver-

wendet wurden) beschränkt. Wenn in der Primerextensionsreaktion nur ddNTPs verwendet wurden, sind die markierten Oligonucleotid-Primer um ein einzelnes, markiertes Nukleotid verlängert worden. Wenn ein Gemisch aus ddNTPs und dNTPs verwendet wurde, sind die markierten Oligonucleotid-Primer um mindestens ein, üblicherweise mehrere Nukleotide verlängert worden, wobei eines oder mehrere davon markiert sein können. Durch dieses Vorgehen können die Bedingungen für die Hybridisierung sehr einheitlich gestaltet werden, und dadurch ist die Reproduzierbarkeit des Systems sehr hoch. Nach dieser Hybridisierungsreaktion werden die Substratoberflächen, z.B. Wells, gewaschen, um alle nicht hybridisierten Oligonucleotid-Primer zu entfernen.

[0020] Nach der Hybridisierung werden die mit der PP verbundenen Substrate einem Detektionstest unterworfen, um zu bestimmen, welche Primer in der Extensionsreaktion verlängert wurden. Falls die verwendeten ddNTPs und/oder dNTPs mit einem Fluoreszenzfarbstoff markiert waren und eine Mikrotiterplatte als Substrat verwendet wurde, kann z.B. mit einem Fluorometer gemessen werden, ob der in einem bestimmten Well hybridisierte Oligonucleotid-Primer ein fluoreszenzmarkiertes Extensionsprodukt enthält oder nicht. Wenn in der Extensionsreaktion für unterschiedliche Nukleotide verschiedene Fluoreszenzfarben verwendet wurden, kann mit einem Fluoreszenztest auch nachgewiesen werden, welches der vier möglichen Nukleotide beim jeweiligen Oligonucleotid-Primer eingebaut wurde. Nur diejenigen Wells zeigen Fluoreszenz, deren PP einen Oligonucleotid-Primer gebunden haben, der auf der Proben-DNA eine passende komplementäre Sequenz gefunden hatte und deshalb mittels der Primer-Extensionsreaktion ein fluoreszenzmarkiertes Didesoxynukleotid eingebaut haben. Wenn die verwendeten Wells in einer festen Anordnung zueinander stehen, wie zum Beispiel in einer Mikrotiterplatte, dann wird durch das Auftreten oder Fehlen von Fluoreszenz von Well zu Well ein Muster erzeugt. Dieses Muster ist diagnostisch für die Proben-DNA und kann daher zur Identifikation beigezogen werden.

[0021] Die Charakterisierung und/oder Identifikation der Proben-DNA ist der letzte Schritt des Verfahrens. Bei Verwendung einer Mikrotiterplatte und vielen Oligonucleotid-Primern ist es vorteilhaft zur Identifikation der Proben-DNA, eine Analyse der Ähnlichkeit des erzeugten Musters mit bekannten Mustern aus einer Datenbank zu vergleichen. Dazu können verschiedene Statistik-Programme verwendet werden, die Clusteralgorithmen enthalten.

[0022] Die Präzision der Identifikation kann z.B. dadurch gesteigert werden, dass in einer Auswahl-Prozedur die Muster von zufällig ausgewählten Untermengen der positiven Wells mit den entsprechenden Mustern in der Datenbank verglichen werden. Der Vorteil dieser Verfahren liegt darin, dass auch abweichende Muster richtig eingeordnet werden können. Zum Beispiel können so Abweichungen von den in der Datenbank enthaltenen Typenmustern, die auf Unterschieden zwischen verschiedenen Populationen basieren, kompensiert werden. Auch können so bisher unbekannte Taxa erkannt und mindestens deren grobe Zugehörigkeit zu bekannten Gruppen festgestellt werden.

[0023] Die vorliegende Erfindung wird nun anhand von Beispielen näher erläutert. Die Erfindung ist aber nicht auf die im experimentellen Teil beschriebenen Beispiele, resp. die darin explizit genannten Ausführungsformen, beschränkt.

1. Nachweis des Funktionsprinzips mit 10 000 Primern in Computersimulationen

[0024] Voraussetzungen – Die vollständigen Genomsequenzen von 22 Mikroorganismen wurden von Genbank heruntergeladen (s. Tab. 1). Basierend auf Literaturangaben zur genetischen Diversität bei *Escherichia coli* (Whittam and Ake, 1993) wurde dann das Genom dieser Art am Computer mutiert. Dabei wurden die folgenden Parameter zugrundegelegt: Nach Whittam and Ake (1993) beträgt der Anteil polymorpher Nukleotide bei *E.coli*, basierend auf einem Set von 11 Genen, durchschnittlich 7.4 Prozent. Damit ist durchschnittlich etwa eines von 14 Nukleotiden polymorph. Der Polymorphismus variiert zwischen verschiedenen Genen, von 1.3 bis 13.1 Prozent, also um den Faktor 10. Entsprechend wurden drei Gentypen designiert, PGL mit einem niedrigen Grad des DNA-Polymorphismus von 2.5 Prozent, PGM mit einem mittleren Polymorphismus-Grad von 7.0 Prozent, und PGH mit einem hohen Polymorphismus-Grad von 11.5 Prozent. Für die Grösse der Gene wurde mit 1200 Basenpaaren der ungefähre Mittelwert der von Whittam and Ake (1993) untersuchten Gene gewählt. Der Anteil jedes Gen-Typs wurde ebenfalls entsprechend den Ergebnissen von Whittam and Ake (1993) gewählt. Die Parameter sind in Tabelle 1 zusammengefasst.

Tabelle 1:

[0025] Liste der untersuchten Spezies mit der Accession-Nummer (Genbank) und Genomgrösse.

Species	Abkürzung	Genbank Accession Nr	Genom Grösse (bp)
<i>Aquifex aeolicus</i>	AQAEOL	AE000657	1551335
<i>Archaeoglobus fulgidus</i>	ARFULG	AE000782	2178400
<i>Bacillus subtilis</i>	BSU_ORI	AL009126	4214814
<i>Borrelia burgdorferi</i>	BOBURG	AE000783	910724
<i>Chlamydia pneumoniae</i>	CHPNEU	AE001363	1230230

CH 699 253 B1

Species	Abkürzung	Genbank Accession Nr	Genom Grösse (bp)
<i>Chlamydia trachomatis</i>	CHTRAC	AE001273	1042519
<i>Escherichia coli</i> K-12 MG1655	ECO_ORI	NC_000913	4639221
<i>Haemophilus influenzae</i>	HAINFL	L42023	1830138
<i>Helicobacter pylori</i>	HEPYLO	AE000511	1667867
<i>Methanobacterium thermo-autotrophicum</i>	METHER	AE000666	1751377
<i>Methanococcus jannaschii</i>	MEJANN	L77117	1664970
<i>Mycobacterium tuberculosis</i>	MYTUBE	AL123456	4411529
<i>Mycoplasma genitalium</i>	MYGENI	L43967	580074
<i>Mycoplasma pneumoniae</i>	MYPNEU	U00089	816394
<i>Pyrococcus abyssi</i>	PYABYS	AJ248283-7/AL096836	1500250
<i>Pyrococcus horikoshii</i>	PYHORI	Pyro_h	1738505
<i>Rickettsia prowazekii</i>	RIPROW	AJ235269	1111523
<i>Synechocystis</i> PCC6803	SYNESE	AB001339	3573470
<i>Thermotoga maritima</i>	THMARI	AE000512	1860725
<i>Treponema pallidum</i>	TRPALL	AE000520	1138011
<i>Saccharomyces cerevisiae</i>	SC_TOT	NC_001133-48	12069247

Tabelle 2:

[0026] Parameter für die Erzeugung der computergenerierten virtuellen Bakterienstämme. Die durchschnittliche Gengrösse ist 12 00 Basenpaare (bp). Für die Computeranalysen musste die Länge beider Genome leicht (<0.02%) verändert werden. Grad des Polymorphismus von Gentyp PGL: niedrig, PGM: mittel, PGH: hoch.

	N	Prozent	Mutationen pro Gen	Durchschnittlicher Prozentsatz polymorpher Nukleotidstellen
<i>E. coli</i>				
Gentyp	966	25,0	30	0,0250
PGL				
Gentyp	1611	41, 7	84	0,0700
PGM				
Gentyp	1289	33,3	138	0,1150
PGH				
Total	3866	100	88.5	0.0738
<i>B. subtilis</i>				
Gentyp	874	24, 9	30	0,0250
PGL				

	N	Prozent	Mutationen pro Gen	Durchschnittlicher Prozentsatz polymorpher Nukleotidstellen
Gentyp	1487	42, 3	84	0,0700
PGM				
Gentyp	1152	32, 8	138	0,1150
PGH				
Total	3513	100	88.3	0.0736

[0027] Computerprogramme – Für die folgenden Funktionsabläufe wurden entweder Programme selbst erstellt oder kommerziell erhältliche Software verwendet (Microsoft Excel; Microsoft Word):

[0028] 1) Herstellen aller möglichen Oligonucleotide einer bestimmten Länge und mit einem bestimmten G/C-Gehalt. Das Programm erstellt alle Sequenz-Kombinationen, die mit den gewählten Parametern möglich sind. Bei längeren Oligonucleotiden sind dies mehrere hundert Millionen Möglichkeiten.

[0029] 2) Herstellen einer Liste von 10 000 zufälligen Zahlen, die als Adressen aus allen Oligonucleotiden bestimmter Länge und G/C-Gehaltes eine zufällige Auswahl von 10 000 Kandidaten trifft.

[0030] 3) Herstellen der virtuellen Stämme von *E. coli* und *Bacillus subtilis* (*B. subtilis*) unter Verwendung der Parameter aus Tabelle 1. Dazu wurde eine Tabelle erstellt, die für jedes Gen die Zuordnung zur Polymorphiegruppe und zufällige Adressen enthält, die die zu mutierenden Nukleotide bezeichnen. Für jedes Genom der beiden Bakterienarten wurden anhand dieser Tabelle 6 virtuelle Stämme produziert, je 3 mit unterschiedlichen Zufalls-Adressen der zu verändernden Nukleotide, und je drei, bei welchen alle Gene einen niedrigen, mittleren beziehungsweise hohen Grad an Polymorphie aufweisen.

[0031] 4) Testen auf Präsenz/Absenz jedes der 10 000 Oligonucleotid-Kandidaten in jedem der Stämme von *E. coli* und *B. subtilis*, sowie in allen anderen Mikroorganismen-Genomen, die in die Analyse einbezogen wurden (Tabelle 2). Das Ergebnis ist eine Matrix von 1 (für vorhanden) beziehungsweise 0 (für nicht vorhanden) für jedes Oligonucleotid und alle getesteten Genome.

[0032] 5) Clusteranalyse mit der unter Punkt 4 hergestellten Matrix zur Bestimmung der Ähnlichkeit zwischen den verschiedenen Genomen.

[0033] Ergebnis – Wenn die richtigen Parameter gewählt werden (z.B. Länge und/oder G/C-Gehalt des Oligonucleotids), dann sollten alle virtuellen Stämme von *E. coli* mit der Originalsequenz eine Gruppe bilden, die sich klar von den anderen Genomen abhebt. Dasselbe gilt für die virtuellen Stämme von *B. subtilis*. Die Fig. 2 zeigt, dass dies tatsächlich eintritt, womit der Nutzen des Prinzips klar demonstriert ist. Ein ähnlich hoher Grad der Zuordnung von Stämmen zu einzelnen Arten kann auch mit anderen Sets von 10 000 zufallsmässig ermittelten Oligonucleotiden von 12 Basenpaaren Länge und darüber erreicht werden. Dies zeigt, dass die Methode sehr robust ist.

[0034] Fig. 2 zeigt ein Dendrogramm der Clusteranalyse der Datenmatrix (Präsenz/Absenz) für 10 000 zufallsmässig ausgewählte Oligonucleotide von 12 Basenpaaren Länge und einem Gehalt von 70% G/C. Alle computergenerierten Stämme von *E. coli* und *B. subtilis* werden der jeweils richtigen Gruppe zugeordnet. Die Ähnlichkeit zwischen den Stämmen wird klar offengelegt, indem bei beiden Arten die am leichtesten mutierten Stämme am nächsten beim Originalstamm zu liegen kommen und die am stärksten mutierten die stärkste Abweichung zeigen. Abkürzungen gemäss Tabelle 2; Stämme von *E. coli* (ECO) und *B. subtilis* (BSU): ST1-3, virtuelle Stämme 1 bis 3, jeder Stamm wurde mit einem anderen Set von Mutationsadressen erzeugt; Stämme PGL, PGM, PGH, alle Gene wurden mit derselben Anzahl Mutationen niedrig (30-PGL), mittel (84-PGM), beziehungsweise hoch (138-PGH) versehen.

2. Nachweis des Funktionsprinzips mit Proben im Mikrotiterplattenformat

[0035] Voraussetzungen – Alle zur Durchführung der Methode notwendigen Teilschritte sind in der Praxis bereits gut etabliert und haben sich bewährt. Zur Aufbereitung von DNA stehen heute Kits von vielen kommerziellen Anbietern zur Verfügung, die auch sehr problematische Templates extrahieren können (z.B. Dneasy Plant Mini Kit, Qiagen AG). Diejenigen Oligonucleotide beziehungsweise Oligonucleotid-Primer, für die auf der eingesetzten DNA eine Hybridisierungssequenz existiert, werden in einer Primer-Extensionsreaktion, auch Minisequencing-Reaktion genannt, markiert (z.B. Pastinen et al., 1997). Auch diese Methode ist bestens etabliert und es existieren Kits dafür (z.B. Snapshot, PEbiosystems AG). Nach der Minisequencing-Reaktion müssen die markierten Oligonucleotid-Primer nachgewiesen werden. Dazu wird das Reaktionsgemisch auf eine zweidimensionale Anordnung von Primer-Proben gegeben. Jede der Primer-Proben hat eine inverse Sequenz zu einem der im Test eingesetzten Oligonucleotid-Primer. Die Primer-Proben können zum Beispiel auf einem Microarray oder in einer Mikrotiterplatte aufgetragen und zum Beispiel durch Affinität fest an die Oberfläche gebunden sein,

beispielsweise mittels Biotin-Streptavidin-Bindung, wobei jedes der Microarray-Spots oder Mikrotiterplatten-Wells jeweils nur eine einzige Primer-Probe enthält. Diese Methode wird heute in der Microchip-Technologie breit angewendet und hat ihre Reproduzierbarkeit bewiesen (z.B. Hacia et al., 1998).

[0036] Durchführung – Im Folgenden wird die technische Durchführbarkeit des Prinzips der hier beschriebenen Methode getestet. Dazu muss eine genau bekannte Genomsequenz eingesetzt werden. Alle natürlichen Organismen, auch innerhalb nächster Verwandtschaftsgruppen, sind genetisch voneinander verschieden. Zudem können immer Mutationen auftreten, die bestimmte Gensequenzen verändern. Da die exakte Kenntnis der Basensequenz für den Test eine notwendige Voraussetzung ist, wurde dazu die genau beschriebene Sequenz des Cloning-Vektors pGEM-3Zf(+) gewählt (Accession Nr. X65306; IG0050). Die Sequenz ist 3199 Basenpaare lang und ist detailliert beschrieben. Je zwei Oligonucleotid-Primer mit (Match) und ohne (Mismatch) passende Hybridisierungssequenz auf der Template-DNA wurden gewählt (Orientierung 5'–3'; BIOT: biotinyliert am 3'-Ende): Match-Primer 1: CAGCGGGTGTG, Match-Probe 1: CAACACCCGCTG-BIOT; Match-Primer 2: GGAAGGGCGATC, Match-Probe 2: GATCGCCCTTCC-BIOT; Mismatch-Primer 1: CGTGACGTTGC, Mismatch-Probe 1: GCAACGTGCACG-BIOT; Mismatch-Primer 2: GCGCCTCATGAC, Mismatch-Probe 2: GTCATGAGGCGC-BIOT. In einer linearen Extensionsreaktion (Minisequencing) werden die Match-Primer markiert, indem ein fluoreszenzmarkiertes Dideoxynucleotid eingebaut wird, das komplementär zur der nächsten, auf die Match-Sequenz folgenden Base ist (unter Verwendung des Snapshot Kits von PE Biosystems). Die Mismatch-Primer finden keine passende Sequenz auf dem Template-Genom und werden daher nicht markiert. Im Folgenden wird mit Streptavidin-behandelten Mikrotiterplatten gearbeitet. Die biotinylierten Match- beziehungsweise Mismatch-Proben werden einzeln in vier Wells gegeben, also zum Beispiel Probe 1 in Well 1, Probe 2 in Well 2, und so fort, und an die Streptavidin-behandelte Oberfläche gebunden. Anschliessend an die Markierungsreaktion wird die Reaktionsmischung zu gleichen Teilen auf die vier Wells der Mikrotiterplatte mit den an die Oberfläche gebundenen Matchbeziehungsweise Mismatch-Proben gegeben. In einer Hybridisierungsreaktion binden diese gebundenen Match- beziehungsweise Mismatch-Proben die entsprechenden, zu ihrer Sequenz inversen Match- beziehungsweise Mismatch-Primer. In Well 1 bindet also die Match-Probe 1 den Match-Primer 1 und entsprechend in den nächsten 3 Wells. In einem Kontrollnachweis wird mit einem selektiv doppelsträngige DNA anfärbenden Medium (zum Beispiel CybrGold™) die Hybridisierungsspezifität getestet, indem alle Primer-Probe-Kombinationen analysiert werden. Das erwartete Ergebnis ist in Tabelle 3 dargestellt.

[0037] Aus den Streptavidin-behandelten Mikrotiterplatten werden dann die nicht gebundenen Primer ausgewaschen. Der Ablauf des letzten Schritts, des Fluoreszenznachweises mit dem Reaktionsgemisch, ist abhängig vom verwendeten Fluoreszenz-Nachweissystem. Die Mikrotiterplatte kann direkt in einem Fluoreszenz-Reader analysiert werden. Als Alternative kann die Mikrotiterplatte erhitzt oder mit denaturierenden Lösungen behandelt werden, um die hybridisierten und fluoreszenzmarkierten Match-Primer von den Match-Proben zu lösen. Die so frei gewordenen fluoreszenzmarkierten Match-Primer können dann aufgenommen und mit einem beliebigen Fluoreszenznachweisgerät nachgewiesen werden, zum Beispiel durch eine Kapillarelektrophorese auf einem ABI310 Genetic Analyzer (PE-Biosystems).

Tabelle 3:

[0038] Erwartete Ergebnisse der Analysen mit den ausgewählten Oligonukleotid-Primern und Primer-Proben bei einem Nachweis mit einem Fluoreszenzfarbstoff, der selektiv doppelsträngige DNA anfärbt; (+ = positives, – = negatives Signal).

Probe	Match-Probe 1	Match-Probe 2	Mismatch-Probe 1	Mismatch-Probe 2
Match-Primer 1	+	–	–	–
Match-Primer 2	–	+	–	–
Mismatch-Primer 1	–	–	+	–
Mismatch-Primer 2	–	–	–	+

Ergebnisse

[0039] Farbmarkierung der eingesetzten Primer: Mit allen vier eingesetzten Primern wurden je zwei unabhängige Primerextensionsreaktionen mit fluoreszenzmarkierten Dideoxynukleotiden durchgeführt. Dazu wurde das SNaPshot™ ddNTP Primer Extension Kit von PE Biosystems entsprechend den Herstellerangaben verwendet, und die Reaktion wurde anschliessend auf einem ABI310 Genetic Analyzer (PE Biosystems) analysiert. Nur die beiden Primer, die auf der verwendeten Template-DNA Sequenz (pGEM-3Zf(+)) eine entsprechende Hybridisierungssequenz vorfanden (beide Match-Primer), waren tatsächlich markiert. Die relativen Fluoreszenzwerte von je zwei unabhängigen Reaktionen waren für Match-Primer 1 1327 und 638 Einheiten und für Match-Primer 2 575 und 243 Einheiten. Für beide Reaktionen der Mismatch-Primer lagen die entsprechenden Werte im Bereich des Hintergrundrauschens (<20 Einheiten).

Hybridisierung der Primer an die immobilisierten Proben

[0040] Die biotinylierten Proben wurden in Streptavidin-impregnierten Mikrotiterplatten (Black Combiplate 8 Streptavidin Coated, Labsystems) immobilisiert. Jeweils 2 Ansätze von 20 µM biotinylierte Probe wurde in 50 µl Binde- und Waschpuffer (1M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA) für 30 Minuten bei Raumtemperatur unter schütteln (1000 rpm in Eppendorf Thermomixer Comfort) inkubiert, dann viermal mit 50 µl desselben Puffers gewaschen. Für die Hybridisierung wurden jeweils 20 µM Primer in 50 µl Hybridisierungsmischung (6xSSC:0.9M NaCl, 0.9M Natriumcitrat; 0.1% SDS; Denhard-Lösung: 1% Ficoll, 1% Polyvinylpyrrolidon, 1% Bovine Serum Albumin) dazugegeben und für 30 Minuten bei 40 Grad Celsius und 1000 rpm inkubiert. Dann wurde viermal mit 60 µl Hybridisierungs-Waschpuffer (0.1xSSC, 0.1% SDS) gewaschen. Zum Nachweis, dass die Proben nur ihre passenden inversen Primer gebunden haben, wurde 50 µl CYBR® Gold Nucleic Acid Gel Stain (Molecular Probes) dazugegeben und die relative Fluoreszenz in einem Fluorometer (Fluoroskan Ascent FL, Labsystems) gemessen. Die Messwerte zeigen, dass die immobilisierten Proben bevorzugt die passenden Primer binden (Tab. 4).

Tabelle 4:

[0041] Bevorzugte Hybridisierung der Primer mit den inversen Proben. Die Werte entsprechen dem Durchschnitt der relativen Fluoreszenzmessung aus zwei Replikationen (jeweils gemittelt über 8 Messwerte; Ausreisser mit mehr als einer Standardabweichung Differenz zum Mittelwert eliminiert).

	Match-Probe 1	Match-Probe 2	Mismatch-Probe 1	Mismatch-Probe 2
Match-Primer 1	1,27	0,49	1,23	0,99
Match-Primer 2	0,45	1,46	1,04	1,29
Mismatch-Primer 1	0,47	0,51	1,50	1,12
Mismatch-Primer 2	0,41	0,47	1,00	1,79

[0042] Nachweis der Hybridisierung markierter Primer. Um nachzuweisen, dass auch die markierten Primer hybridisieren, wurde zu einer der immobilisierten Proben (Match-Probe 2) 15 µl der Extensionsreaktion mit dem Match-Primer 2 gegeben und wie oben beschrieben hybridisiert und gewaschen. Danach wurde der gebundene Match-Primer 2 in 20 µl einer denaturierenden Lösung (0.125M NaOH, 0.1M NaCl) wieder von der Probe gelöst. Drei Mikroliter dieser Lösung wurden schliesslich auf dem ABI310 analysiert. Bei einem Hintergrundsignal von unter 4 Einheiten wurde ein Signal von 72 Einheiten gemessen, womit nachgewiesen ist, dass die markierten Primer tatsächlich an die immobilisierten Proben binden.

Sequenzprotokoll**[0043]**

<110>	Eidgenössische Forschungsanstalt für Obst-, Wein u
<120>	Verfahren zur Charakterisierung und/oder Identifikation von Genomen
<130>	Anonyme Primer hybridisierung
<140>	
<141>	
<160>	8
<170>	PatentIn Ver. 2.1
<210>	1
<211>	12
<212>	DNA
<213>	Künstliche Sequenz
<220>	
<223>	Beschreibung der künstlichen Sequenz: Primer Sequenz
<400>	1

CH 699 253 B1

cagcgggtgt tg		12
<210>	2	
<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Probe	
<400>	2	
caacacccgc tg		12
<210>	3	
<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Sequenz	
<400>	3	
ggaaggcga tc		12
<210>	4	
<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Probe	
<400>	4	
gatcgccctt cc		12
<210>	5	
<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Sequenz	
<400>	5	
cgtgcacgtt gc		12
<210>	6	
<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Probe	
<400>	6	
gcaacggca cg		12
<210>	7	

<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Sequenz	
<400>	7	
gcgcdtcacg ac		12
<210>	8	
<211>	12	
<212>	DNA	
<213>	Künstliche Sequenz	
<220>		
<223>	Beschreibung der künstlichen Sequenz: Primer Probe	
<400>	8	
gtcatgaggc gc		12

Referenzen

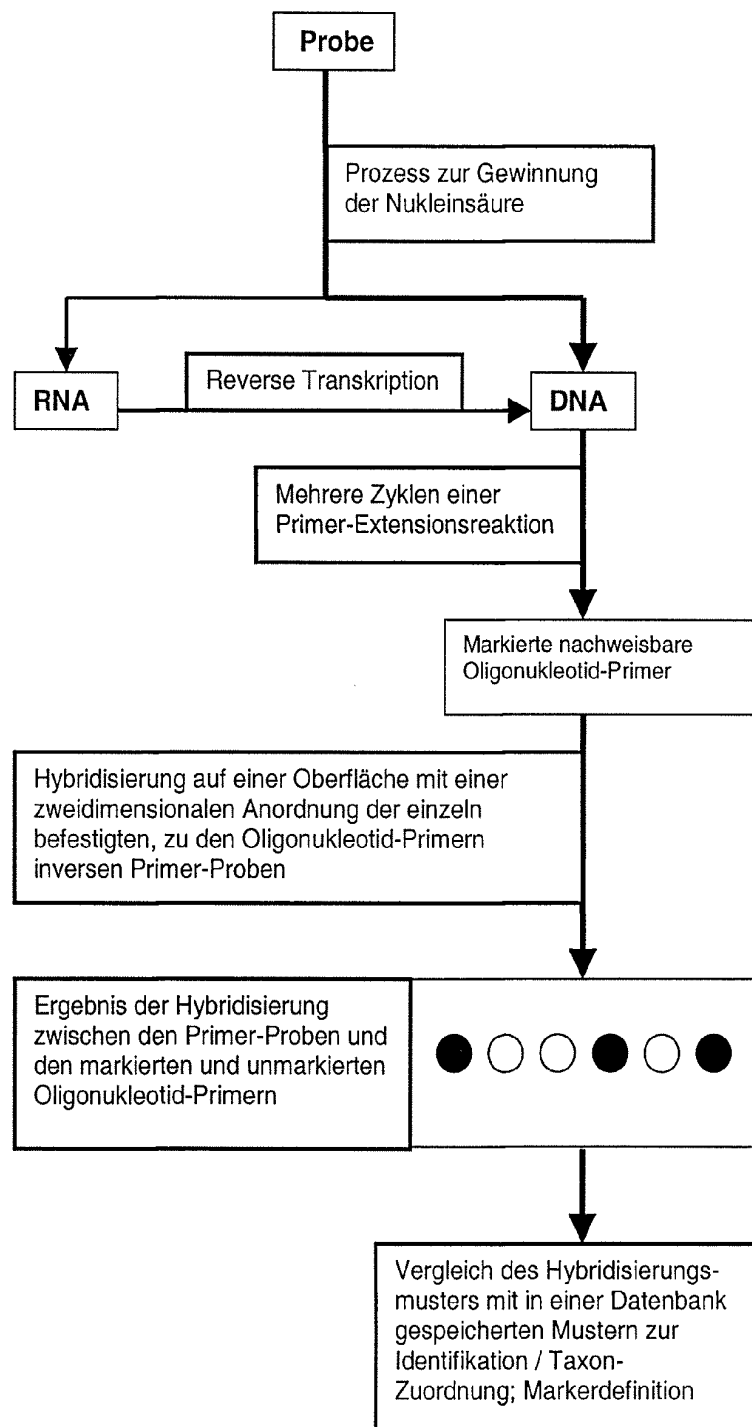
- [0044] Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genetics Supplement* 21: 33–37.
- [0045] Frey JE, Frey B (1997) PCR-based diagnostics of agricultural pests and diseases.
- [0046] In: H.-W. Dehne et al. (eds.), *Diagnosis and Identification of Plant Pathogens. Proceedings of the 4th International Symposium of the European Foundation for Plant Protection*, pp. 23–28.
- [0047] Hacia JG, Brody LC, Collins FS (1998) Applications of DNA chips for genomic analysis. *Molecular Psychiatry* 3: 483–492.
- [0048] Lander ES (1999) Array of hope. *Nature Genetics Supplement* 21: 3–4.
- [0049] Mueller UG, LaReesa Wolfenbarger L (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* 14: 389–394.
- [0050] Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvänen AC (1997) Minisequencing: A specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Research* 7: 606–614.
- [0051] Perez T, Albornoz J, Dominguez A (1998) An evaluation of RAPD fragment reproducibility and nature. *Molecular Ecology* 7: 1347–1357.
- [0052] Saiki RK, Gelfand DH, Stoffel S, Scharf J, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable Polymerase. *Science* 239: 487–491.
- [0053] Whittam TS, Ake SE (1993) Genetic polymorphisms and recombination in natural populations of *Escherichia coli*. In: *Mechanisms of molecular evolution*, Naoyuki Takahata, Andrew G. Clark (eds.), Sinauer Associates, Tokyo, pp. 223–245.

Patentansprüche

- Verfahren zur Charakterisierung und/oder Identifikation eines Genoms, welches folgende Schritte umfasst:
 - Hybridisierung mindestens eines Oligonukleotidprimers mit einer DNA-Probe des zu charakterisierenden Genoms
 - Extension des angelagerten Oligonukleotidprimers in einer Minisequencing Reaktion in Anwesenheit mindestens eines markierten Didesoxynukleotidtriphosphates und/oder mindestens eines markierten Desoxyukleotid triphosphates,
 - Hybridisierung der Extensionsreaktion mit einer Primer-Probe, die eine zum verwendeten Oligonukleotidprimer komplementäre Sequenz aufweist,
 - Nachweis eines gebundenen Extensionsproduktes und
 - Charakterisierung und/oder Identifikation des Genoms mittels eines Clusteralgorithmus-Programms.
- Verfahren gemäss Anspruch 1, dadurch gekennzeichnet, dass mehr als ein Oligonukleotidprimer verwendet wird, bevorzugt bis zu einem Dutzend, bevorzugter bis zu tausend, noch bevorzugter bis zu 10 000 und am bevorzugtesten über 10 000 Oligonukleotidprimer verwendet werden.

3. Verfahren gemäss Anspruch 1 oder 2, dadurch gekennzeichnet, dass die Primer eine zufällige Nukleotidsequenz aufweisen.
4. Verfahren gemäss Anspruch 1 oder 2, dadurch gekennzeichnet, dass die Primer eine Sequenz besitzen, die komplementär zu einer Zielsequenz des zu charakterisierenden Genoms ist.
5. Verfahren gemäss einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass ein Gemisch von Primern mit zufälliger Sequenz und Primern mit einer komplementären Sequenz zu einer Zielsequenz des zu charakterisierenden Genoms eingesetzt werden.
6. Verfahren gemäss einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass der mindestens eine verwendete Primer eine bestimmte Länge und/oder bestimmten G/C-Gehalt aufweist.
7. Verfahren gemäss einem der Ansprüche 1 oder 2, dadurch gekennzeichnet, dass der mindestens eine Primer eine bestimmte Schmelztemperatur besitzt.
8. Verfahren gemäss einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass das mindestens eine Didesoxynukleotidtriphosphat und/oder das mindestens eine Desoxynukleotidtriphosphat fluoreszenzmarkiert ist.
9. Verfahren gemäss einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass alle 4 ddNTPs fluoreszenzmarkiert sind, vorzugsweise jedes ddNTP mit einem verschiedenen Fluorophor.
10. Verfahren gemäss einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass die Primer-Probe am 5'-Ende einer zum verwendeten Oligonukleotidprimer komplementären Sequenz entspricht und am 3'-Ende eine bindungsvermittelnde Verlängerung aufweist, mit welcher sie an ein Substrat gebunden ist.
11. Verfahren gemäss Anspruch 10, dadurch gekennzeichnet, dass die bindungsvermittelnde Verlängerung eine Verankerung ist oder enthält.
12. Verfahren gemäss Anspruch 11, dadurch gekennzeichnet, dass die Verankerung ein Biotin-Molekül ist.
13. Verfahren gemäss einem der Ansprüche 10 bis 12, dadurch gekennzeichnet, dass zwischen der dem Oligonukleotidprimer komplementären Sequenz und der Verankerung ein Nukleotidschwanz vorhanden ist.
14. Verfahren gemäss einem der Ansprüche 10 bis 13, dadurch gekennzeichnet, dass das Substrat eine Oberfläche eines Mikrotiterplatten-Wells oder eines Microarrays ist.
15. Verfahren gemäss einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass die DNA-Probe des zu charakterisierenden Genoms mittels Reverse Transkription aus RNA hergestellt wurde.

Figur 1



Figur 2

