

Microarray-basierende molekulare Diagnostik in der Agronomie

Jürg E. Frey und Monika Pfunder

Agroscope FAW Wädenswil, Eidg. Forschungsanstalt für Obst-, Wein- und Gartenbau,
Fachbereich Pflanzenschutz, Labor für Molekulare Diagnostik, 8820 Wädenswil

Zusammenfassung

Die neueste Literatur sowie unsere eigenen Erfahrungen zeigen, dass die Microarray Technologie ein grosses Potential in der molekularen Diagnostik hat. Durch die Möglichkeit, hochredundante Abfragen zu designen, ist die Technologie sehr robust, präzise und effizient. Eine ähnlich hohe Sicherheit der Typisierung könnte nur mittels Sequenzierung von mehreren Genen/Genabschnitten pro Taxon erreicht werden, was sowohl technisch als auch bezüglich Effizienz deutlich schwieriger durchzuführen wäre.

In unserem inzwischen durchwegs autonomen Microarray-Labor und in Zusammenarbeit mit WissenschaftlerInnen aus insgesamt 21 Ländern (COST Action 853, www.cost853.ch) evaluieren wir das Potential der Technologie und verfolgen zwei grundsätzlich verschiedene Ansätze, um sowohl bekannte wie auch unbekannte Organismen zu diagnostizieren.

Mit der Entwicklung von kostengünstigen „Gene Chips“ können wir Organismen diagnostizieren, von denen bereits spezies- oder typusspezifische Sequenzinformationen besitzen. Diese Methode haben wir in verschiedenen Projekten für Säugetiere, Viren und Bakterien validiert. Parallel dazu sind wir an der Entwicklung eines „Genome Chips“, welcher es ermöglichen wird, unbekannte Organismen zu mit der Hilfe von zufälligen „Treffern“ mit kurzen Oligonukleotiden auf dem Genom zu charakterisieren. Ein Experiment mit 412 13-mer Oligonukleotiden wurde durchgeführt, um das Verhalten der Hybridisation auf der Oberfläche eines Microarray-Chips zu studieren. Dabei wurden empirische Werte mit den erhältlichen theoretischen Algorithmen verglichen. Aufgrund dieser Resultate einigten wir uns auf die Selektion-on-chip, da die Voraussagen sehr spekulativ sind. Ausserdem etablierten wir ein sehr vielversprechendes Probe-Design für die Fänger-moleküle auf dem Chip. Wir verbesserten unsere Labelling-Technik der genomischen DNA und konnten die Sensitivität des Systems bereits um ein Mehrfaches erhöhen, erwarten und benötigen aber noch weitere Verbesserungen in diesem Bereich.

Inhaltsverzeichnis

1 –	Einleitung	1
2 –	Entwicklung des Genome Chips	2
	2.1 Oligo-Design	2
	2.2 Probestem-Design	3
	2.3 Probe-Orientierung	3
	2.4 Spotting	4
	2.5 Labelling	5
	2.6 Voramplifikation geringer DNA Mengen	6
	2.7 Schmelztemperatur-Bestimmung	7
	2.8 Hybridisierung	8
3 –	Entwicklung des Gene Chips.....	11
4 –	Diskussion der vorläufigen Ergebnisse	11
5 –	Ueberblick über den Stand der Entwicklung von diagnostischen Microarrays für Microorganismen (Stand Anfang 2004)	11
	Beilage	14

1 – Einleitung

Um Mikroorganismen mit Hilfe von Microarrays zu diagnostizieren, verfolgen wir generell zwei Ansätze, den Genome Chip und den Gene Chip:

Genome Chip: Der Ansatz des Genome Chips macht sich zunutze, dass eine zufällige Folge von 13 Basen im Durchschnitt zwischen 0 und 1 Treffer auf einem Mikroorganismen-Genom findet. Durch die Variabilität des Genoms verschiedener Organismen variiert die Trefferquote des einzelnen Oligonucleotide (Oligos). Dadurch entsteht für verschiedene Organismen bei Benutzung der gleichen Oligos ein unterschiedliches Muster an Treffern und „Nieten“.

Der Vorteil dieser Methode besteht darin, dass die Gensequenz des zu untersuchenden Organismus nicht bekannt sein muss. Es wird ein bestimmtes Muster produziert und dieses muss innerhalb desselben Organismus reproduzierbar sein. Wie in einer Verbrecherdatei werden Fingerabdrücke von Unbekannten in einer Datei abgelegt. Für bereits Bekannte kann man dem Fingerabdruck ein Gesicht zuordnen, resp. dem Muster einen Mikroorganismus. Im Gegensatz zum Fingerabdruck kann man aber auch Hinweise auf die Verwandtschaft eines unbekanntem Mikroorganismus erhalten, weil sein Muster demjenigen eines näher verwandten bekannten Organismus ähnlicher ist, als jenem eines entfernteren Verwandten. Dies gilt allerdings nur, wenn die Gesamtzahl der Oligos hoch genug ist.

Gene Chip: Der Gene Chip funktioniert wie der Genome Chip mit Oligos, die einen Treffer auf der Ziel-DNA haben. Im Unterschied zum Genome Chip dagegen wird das Oligo schon gezielt so entworfen, dass es nur bei einem bestimmten Organismus einen Treffer haben wird. Um einen Satz von artspezifischen Oligos zu entwickeln, muss also die DNA aller gesuchten Organismen bekannt sein. Zusätzlich ist es von Vorteil, von möglichst allen im zu untersuchenden System vorkommenden Organismengruppen mindestens einen Vertreter in der Datenbank zu haben. Da nicht das gesamte Genom aller, für eine bestimmte Fragestellung interessanter Organismen sequenziert werden kann, beschränkt man sich hier auf bestimmte DNA-Fragmente. Bevor ein Array hergestellt werden kann, müssen zuerst die entsprechenden DNA-Daten vorhanden sein, also z.B. die 16S-Gensequenz aller involvierten Mikroorganismen.

2 – Entwicklung des Genome Chips

2.1 Oligo-Design

Das Design der Oligos, die für die Abfrage über Präsenz/Absenz im Genom eingesetzt werden, ist einer der kritischsten Faktoren in der Entwicklung des Genome Chips. Einer der wichtigsten zu berücksichtigenden Faktoren ist, dass die Oligos nicht miteinander hybridisieren dürfen, dass sie nicht selbst-komplementär sind, dass sie alle in etwa ein ähnliches Schmelzverhalten aufweisen und dass sie keine homobasische Sequenz aufweisen. Um diese Faktoren zu vereinen, wurde ein Computerprogramm entwickelt, das zur Zeit weiter optimiert wird. Zum Beispiel haben wir in einer ersten Phase ein Set von 412 Oligos nach folgenden Kriterien am Computer modelliert:

- Oligo Länge 13 Basen
- Schmelztemperatur (NearestNeighbor, Sugimoto) = 57 ± 0.5 °C
- GC Gehalt höher als 50% und max 55%
- Nicht mehr als 4 nacheinander folgende gleiche Basen (PolyNs)
- Nicht mehr als 3 nacheinander folgende Basen nur aus G oder C bestehend (GC-clamps)

- Keine Hairpins mit mehr als 4 Basen
- Keine Komplementarität zu den anderen Random Oligos
- Mindestens 5 Basen Unterschied zu den anderen Random Oligos
- Keine Selbstkomplementarität

Durch dieses strenge Ausleseverfahren wurden aus ca. 64 Mio möglichen Kandidaten insgesamt 824 Oligos berechnet. Für unsere Versuche wurde davon die Hälfte, selektiv jedes zweite Oligo, ausgewählt und zusammen mit seinem Komplement (als Probe auf den Chip) bestellt (Microsynth GmbH). Mit diesen Oligos wurden Versuche zum Schmelzverhalten (Tm) durchgeführt (Pfunder and Frey, in press; Preprint beiliegend).

2.2 Probestem-Design

Im Rahmen von Untersuchungen zur Optimierung der Signalstärke der Spots auf den Chips stellten wir fest, dass die Signalintensität stark abhängig ist vom Design der Probe, also dem „Fängermolekül“, das auf dem Chip mittels Aminolinkern befestigt wird. Eine Probe kann in der Regel nur ihr optimales Signal realisieren, wenn sie einen mindestens 12 Basen langen „Spacer“ (12 T) aufweist.

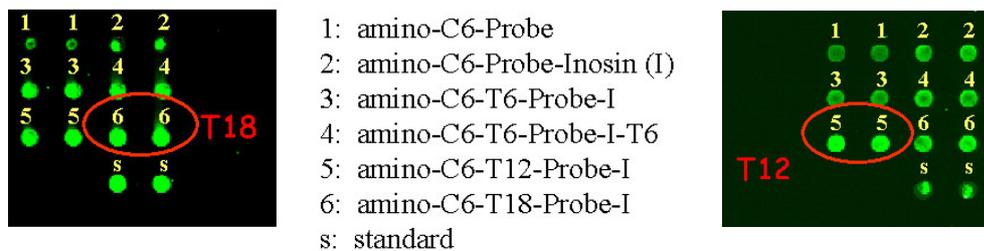


Abb. 1: Test von Oligonucleotiden mit verschiedenen „Spacer“-Längen.
 Links: Test mit Proben für *Cydia pomonella*; Rechts: Test mit Proben für *Delia platura*

2.3 Probe-Orientiation

Eine weitere wichtige Beobachtung in diesen Versuchen war, dass die Orientierung der Probe einen starken Einfluss auf die Signalstärke haben kann. Bei vielen Proben ist das Signal signifikant stärker, wenn sie am 3'-Ende an der Chip-Oberfläche befestigt ist als wenn sie am 5'-Ende befestigt ist.

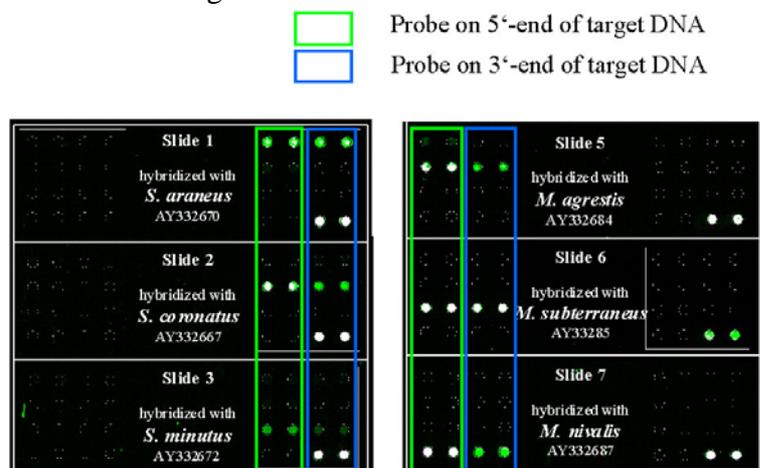


Abb. 2: Abhängigkeit der Signal-Intensitäten von der Position der Probe auf der DNA (hier 21- 28mer Oligonukleotide mit DNA Fragmenten von ca. 800bp Länge).

2.4 Spotting

Beim Spotting werden Fänger-Oligos (Probes) auf einen Chip gebracht. In unserem Labor verwenden wir zur Zeit Glasslides mit einer Aldehyd besetzten Oberfläche (Genetix). Die Fänger-Oligos werden in diesem Fall durch die an den Oligos befestigten Aminolinker kovalent an die Oberfläche gebunden. Die Probes werden mit nadelförmigen Pins auf die Oberfläche gebracht. Durch nachfolgendes Waschen wird die nicht durch Oligos besetzte Oberfläche blockiert und überschüssige Oligos gewegewaschen.

Ursprünglich haben wir einen Handspotter verwendet, bei dem 8 Pins gleichzeitig in einer geführten Schiene auf das Glas gebracht werden. Für kleine Versuche (bis max. 98 Spots) ist diese billige Version einigermaßen geeignet, allerdings zeitaufwändig und anfällig auf Fehler.

Nun spotten wir mit einem Roboter (QArray Mini von Genetix). In einer ersten Phase wurde mit „solid pins“ gearbeitet. Obwohl sehr zuverlässig, müssen diese pins aber zwischen jedem Spotting-Schritt zurück zur Quell-Platte, wodurch ein hoher Durchsatz sehr viel Zeit braucht. Jetzt verwenden wir „split pins“, die ein rasches Spotten mit regelmässiger Verteilung der Proben zulassen, und deshalb für high-density Arrays optimal sind.

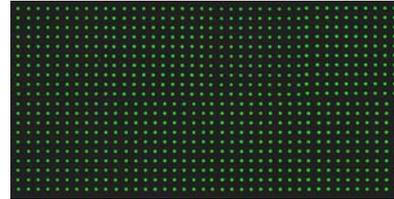


Abb. 3: 800 Cy-3 markierte Probes gespottet mit 8 Splitpins auf Aldehyd-behandelten Glasslides

Im Verlauf der Evaluierung der Split-Pins haben wir festgestellt, dass eine Veränderung des Waschprotokolls (in Abweichung zum Protokoll des Slide-Herstellers) zu einer Erhöhung der Sensitivität um den Faktor 5 führt. Diese Entwicklung könnte massgeblich dazu beitragen, dass wir die Sensitivitätsprobleme beim direkten Labelling lösen könnten (siehe nächstes Kapitel).

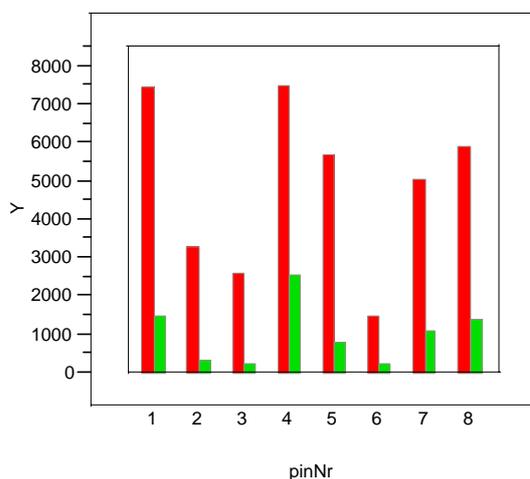


Abb. 4: Erhöhung der Sensitivität durch eine Veränderung des Waschprotokolls. Grüne Balken: Originalprotokoll; rote Balken: Neues Protokoll.

2.5 Labelling

Für den Genome Chip basiert die von uns bisher verwendete Markierung auf der Methode der Single-Base-Extension, dasselbe Prinzip, das auch bei der Markierung von einzelsträngigen PCR Produkten in der Sequenzierreaktion sowie bei SNP-Analysen angewendet wird.

Alle im Array beteiligten Oligos werden mit Template-DNA in einem PCR Schritt vereint, wobei diejenigen Oligos, die auf dem Template einen Hit haben, durch eine Base (das Komplement der nächsten Base auf der Template-DNA) verlängert werden. Die Basen, die dafür in der Reaktion zur Verfügung stehen, sind mit einem Fluorophor (Cy-3) markiert. An den Oligos mit einem Hit wird also nach 60-99 Zyklen ein Fluorophor hängen, während diejenigen Oligos, welche kein Komplement auf der DNA finden, ohne Basenverlängerung und deshalb ohne Markierung bleiben.

Während der Hybridisierung auf dem Chip werden sowohl markierte als auch nicht markierte Oligos mit den komplementären Probes zu einem Doppelstrang verbunden. Aber nur die markierten Oligos werden im Lesegerät sichtbar sein und dadurch ein spezifisches Muster produzieren. Um das Labelling unabhängig von der Hybridisierung auf dem Chip zu testen, werten wir Resultate (Lichtintensität der markierten Oligos) teilweise auf dem ABI Prism 3100 Genetic Analyser aus.

Für den Genome Chip Ansatz müssen sehr kurze Oligos (9-13mer) verwendet werden, damit auf der Template-DNA die richtige Anzahl Hits gefunden werden kann. Um die Methode der Single-Base-Extension für unser Labelling zu testen, haben wir uns eines künstlichen Systems bedient. Wir bestellten vier künstliche DNA-Template („fake template“) von 23 bp Länge, welche an ihrem 3'-Ende mit je einem der Random-Oligos ein Komplement bilden. In einem Experiment gaben wir je eines der vier künstlichen Templates (FT) zu je einem von vier zufällig gewählten Oligos. Wir konnten zeigen, dass gezielt nur dann eine Extension stattfindet, wenn das Oligo mit seinem spezifischen Template hybridisiert. Allerdings sind die physikalischen Bedingungen in der Mixtur für die Labelling-Reaktion abhängig von den einzelnen Komponenten wie Template-DNA-Menge, Anzahl Oligos, Anzahl ddNTPs usw.

Der Schritt von kurzen Templates zu einem gesamten Genom eines Mikroorganismus ist uns noch nicht gelungen. Wir haben genomische reine *E. coli* DNA verwendet, sie unter anderem mit DNaseI verdaut, um sie zu verkleinern, aber bisher noch kein Signal erhalten. Das Verhältnis der einzelnen Komponenten muss noch optimiert werden. Wir sind überzeugt, dass das Labeling inzwischen funktioniert, dass aber die Sensitivität des Systems noch zu tief ist.

Ein Problem, das sich während unserer Versuche herauskristallisierte, ist die grosse Variabilität im Verhalten der einzelnen Oligos (siehe unten, Unterkapitel Schmelztemperatur Bestimmung). Wir sind heute der Überzeugung, dass wir uns nicht auf die berechneten Schmelztemperatur-Werte von allgemein verwendeten Algorithmen verlassen können, um Oligos mit ähnlichem Verhalten zu selektionieren. Die Schmelztemperatur bestimmt die Temperatur, unter welcher das Annealing von den Oligos zur Template-DNA stattfindet. Wenn diese Temperatur zwischen Oligos stark schwankt, ist eine bestimmte Temperatur für einige der Oligos zu hoch, was ein Annealing nicht zulässt und ein falschnegatives Resultat produziert. Für andere Oligos ist die Temperatur zu tief, was dazu führt, dass ein Annealing auch bei Mismatches ermöglicht wird und dadurch ein falschpositives Resultat entsteht.

2.6 Voramplifikation geringer DNA Mengen

Die Single-Base-Extension Methode, die wir verwenden, um unsere gesuchten DNA Fragmente zu markieren, ist dieselbe, die auch bei Sequenzierreaktionen und SNP-Analysen zum Tragen kommt. Beim Sequenzieren wird von einer Mindestmenge von 2-3 µg Template-DNA ausgegangen (ABI Prism BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems)). Um auch kleine Mengen DNA aus einzelnen Organismen analysieren zu können, ist deshalb auch für unsere Labelling-Reaktion in vielen Fällen eine vorgängige Amplifikation nötig. Hierfür werden in der Literatur verschiedene Methoden vorgestellt. Der wohl einfachste und spezifischste Ansatz ist das PCR, die gezielte Amplifikation einzelner oder mehrere Gensequenzen, einzeln oder im Multiplex. Verschiedene Arbeiten beschäftigen ausserdem mit Whole-Genome-Amplifications (WGA), d.h. dem Versuch, die gesamte DNA gleichermassen zu multiplizieren. Wir haben die folgenden Methoden ausgewählt, um einige Versuche zu machen:

DOP-PCR: Degenerate oligonucleotide primed polymerase chain reaction. Cheung and Nelson 1996. Ein stark degenerierter Primer wird zur genomischen DNA zugegeben und sollte, gleichzeitig als forward und reverse Primer dienend, das gesamte Genom, aufgeteilt in kleine Stücke, amplifizieren.

TGA: Target Genome Amplification. Statt nur einen DOP primer zu verwenden, amplifizieren wir mit einem TGA-Primer und unseren Random-Primern. So sollten wir spezifisch nur diejenige DNA amplifizieren, für die unsere Random Oligos ein Target gefunden haben.

Die beiden Methoden haben insofern funktioniert, als dass wir aus einer Startmenge von 20 ng genomischer DNA nach 60 Amplifikationszyklen rund 100 ng/ul aufgereinigte DNA (ungereinigt rund 500-700ng) amplifizieren konnten (Abb. 1). Für eine Sequenzierreaktion mit BigDye (Applied Biosystems, siehe oben) werden bei PCR Produkten mit >2000bp mindestens 40-100ng vorausgesetzt.

Bei allen WGA-Methoden besteht immer die Gefahr, dass einzelne Gen-Abschnitte in grösserer Zahl amplifiziert werden als andere und somit eine spezifische Selektion nicht zu vermeiden ist. Ausserdem liegen einzelne DNA Fragmente schon in genomischer DNA in höherer Anzahl vor (rDNA, mtDNA). Eine exponentielle Amplifikation wird dieses Verhältnis noch verstärken.

Die TGA-Methode ist deshalb bestechend, weil durch Verwendung der Oligos aus dem Array nur (oder grösstenteils) diejenigen Target-Sequenzen amplifiziert werden, für die der Random-Oligo ein Komplement findet. Durch die hohe Anzahl an Kopien wird die anschliessende Labelling-Reaktion, die mit denselben Oligos stattfindet, gezielt gefördert.

Leider gelang es uns noch nicht, TGA Produkte zu hybridisieren. In der TGA wird zusätzlich zu den nach strengen Kriterien ausgewählten Oligos ein weiteres stark degeneriertes Oligo verwendet (TGA15: 5'-NNNNNNNNNNNNNGTA-3'), das mit einer höheren Wahrscheinlichkeit auf dem Genom einen Treffer hat und dadurch die Amplifikationsprodukte relativ klein hält. Unsere Tests haben ergeben, dass TGA-Oligos offenbar untereinander und mit den Random Oligos komplementäre Verbindungen eingehen können. Während der

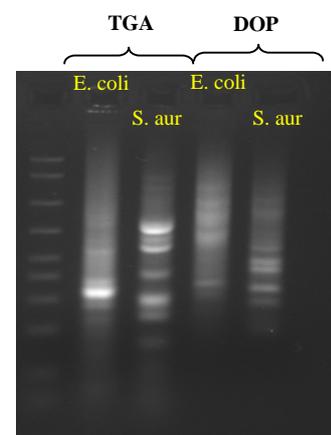
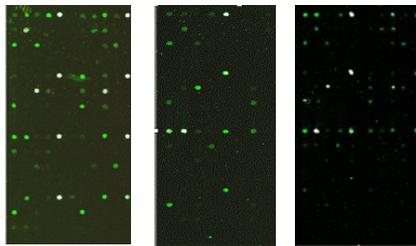


Abb. 5: TGA und DOP amplifizierte DNA von *Escherichia coli* und *Staphylococcus aureus*

anschliessenden Labelling-Reaktion wird somit eine ungewollte und ungezielte Single-Base-Extension ermöglicht, unabhängig von der Target DNA und sogar in reinem Wasser. Dadurch erhalten wir, unabhängig vom Target, immer sehr ähnliche Muster (Abb. 2).



E. coli **S. aureus** **H2O**

Abb. 6: Gleiches Muster für die TGA Amplifikationen von *E. coli*, *S. aureus* und reinem Wasser.

Diese Arbeiten wurden durch die rasche Entwicklung am Markt eingeholt. Es wurde im Sommer 2004 ein neues Kit auf den Markt gebracht, GenomiPhi (Amersham), das eine gleichmässige Amplifikation des gesamten Genoms ermöglicht. Die Technologie basiert auf einer isothermischen Reaktion, bei welcher die DNA-Stränge fortlaufend kopiert werden (rolling circle amplification). Das Kit produziert in ca. 16h aus ca. 50ng DNA 4-5ug DNA, was für unsere Zwecke ausreichend ist. Die Eignung dieses Verfahrens für unseren Zweck wird nächstes Jahr eingehend getestet.

2.7 Schmelztemperatur-Bestimmung

Jedes Oligo ist aus einer anderen Basensequenz aufgebaut. Entsprechend ändert sich sein physikalisches Verhalten, wenn es sich mit einem Komplement zu einem Doppelstrang verbindet. Wichtige Faktoren für dieses Verhalten sind aber auch die Umgebung, in der das Oligo hybridisieren soll: die Temperatur, die Ionenkonzentration, die Primerkonzentration usw. Einer der wichtigsten Werte, sowohl für die Hybridisierung auf dem Chip als auch für das Labelling durch Single-Base-Extension, ist die Schmelztemperatur T_m . Der T_m -Wert beschreibt die Temperatur, bei welcher in gegebener Umgebung die eine Hälfte aller Oligos als Einzelstränge und die andere Hälfte als Doppelstränge vorliegen.

Für uns galt der T_m -Wert als eines der Grundkriterien für die Selektion von Oligos innerhalb eines Microarrays, da jeweils ein Array nur mit einer Temperatur hybridisiert werden kann. Entsprechend basierten wir unser Oligo-Design auf einem theoretischen Algorithmus (Sugimoto), und wählten nur Oligos aus, die eine theoretische Schmelztemperatur von $57 \pm 0.5^\circ\text{C}$ besaßen.

Es bestehen heute verschiedenste Algorithmen, die den T_m -Wert theoretisch berechnen. Wir beziehen uns in unserer Arbeit auf zwei Ansätze:

- a) Hyther (<http://ozone2.chem.wayne.edu/Hyther/hythermain.html>), John Santa Lucia Lab, Wayne State University
- b) Nearest-Neighbour Berechnung nach Breslauer et al., Proc. Nat. Acad. Sci. 83, 3746-50 mit den Werten von Sugimoto *et al.*, Nucl. Acids Res. **24**, 4501-4505, 1996

Bei Verlaufe unserer ersten Versuche (siehe Unterkapitel Labelling) kam uns der Verdacht, dass die theoretischen Werte für so kurze Oligos nicht stimmen könnten. Wir entschieden uns für einen empirischen Versuch, um dieser Beobachtung nachzugehen. Für alle 412 Oligos, die uns zur Verfügung standen, evaluierten wir mit Hilfe von Real-time PCR und dem Marker SybrGreen gezielt den T_m -Wert jedes Primers. Die Hybridisierung fand zwischen dem O-Oligo (primer) und dem P-Oligo (probe, normalerweise auf dem Chip) sowohl in PCR Buffer

(AmpliTag) als auch in 1xSSC Buffer in Lösung statt. Für dieselben Buffer kann der theoretische Wert berechnet werden.

Die Daten bestätigen unseren Verdacht, dass für kurze Oligos wie 13mer die berechneten Werte nicht mehr gültig sind (Pfundner and Frey, in press; Preprint beigelegt): Der T_m -Wert der empirischen Daten liegt im Durchschnitt 8 Grad höher als der berechnete Wert von Hyther (Abb. 7) und zeigt eine nicht-lineare Variabilität zwischen den verschiedenen Primer/Probe Paare.

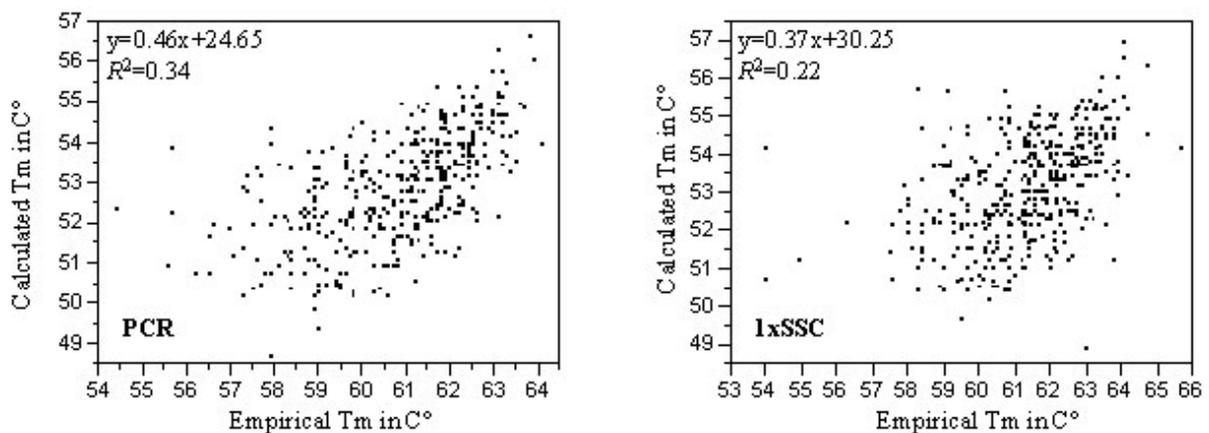


Abb. 7: In PCR Buffer (links) und 1xSSC (rechts) hybridisierte Paare von 13mer Oligonukleotiden mit ihren Probes zeigen grosse Abweichungen von denjenigen Daten, die mit Hilfe des Programms Hyther theoretisch berechnet wurden.

2.8 Hybridisierung

Bei der Hybridisierung der Oligos auf dem Chip wird die Spreu vom Weizen getrennt. Alle in der Labellingreaktion enthaltenen Oligos sollten jetzt ihren Partner (Probe) auf dem Chip finden und bei den gegebenen Bedingungen hybridisieren. Somit wird der Mix „demultiplexed“ und einzelne Spots werden von markierten und andere von nicht-markierten Oligos besetzt.

In einem Versuch mit 40 13mer Oligos testeten wir die Spezifität sowie die Vorhersagbarkeit des Hybridisierungsverhaltens auf dem Chip gemäss den Daten, die wir mit Hilfe von Dissociationskurven evaluiert hatten (siehe oben). Die Hybridisierungen waren sehr spezifisch und nur gerade 2 von allen 40 getesteten Oligos zeigten Kreuzhybridisierungen mit anderen Probes. Dabei blieb die Intensität immer weit unter den korrekten Signalen (max. 30%).

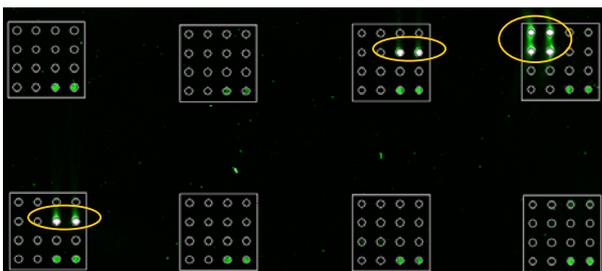


Abb. 8: Hybridisierung von vier 13mer Oligos. Rechts unten in den Rechtecken befinden sich die positiven Kontrollen, die gelben Kreise zeigen die hoch spezifischen Hybridisationen.

In einem weiteren Schritt testeten wir das Verhalten der 40 13mer Oligos auf dem Chip im Vergleich zu den vorhergesagten Schmelztemperaturen.

Parameter	Theoretisch berechnete Werte			Experimentelle Werte	
	HyTher™ PCR	HyTher™ 1xSSC	Sugimoto	In PCR	In 1xSSC
ρ^2	0.024	0.024	0.040	0.174	0.063
Prob> $ \rho $	0.343	0.343	0.217	0.008	0.117

Tabelle 1: Spearman's Rank Korrelation der Signal Intensität von 40 primer/probe Paaren gegenüber den theoretisch errechneten (HyTher, Sugimoto) und den von uns ermittelten empirischen Schmelztemperatur-Vorhersagen.

Die Resultate lassen deutlich werden, wie wenig Möglichkeiten wir zur Vorhersage des Verhaltens von kurzen Oligonukleotiden auf dem Chip haben. Maximal 17% der Variabilität der Signalstärke kann auf die Schmelztemperatur zurück geführt werden. Die restlichen 83% beinhalten Faktoren wie Sterische Hinderung oder Intraspot Interferenzen.

Aus diesen Resultaten schliessen wir, dass in Zukunft eine Selektion auf dem Chip stattfinden sollte. Dabei rechnen wir aufgrund der zufriedenstellenden Resultate der Spezifität nicht mit mehr als 10% Ausfall, was unseren Erachtens finanziell vertretbar und günstiger als das Dissoziations-Screening ist.

Interessanterweise fanden wir, dass die Oligos mit tieferen Schmelztemperaturen signifikant schwächere Signale zeigten als diejenigen mit höheren Schmelztemperaturen. Und dies, obwohl die Hybridisierungstemperatur weit unter dem Schmelzwert lagen.

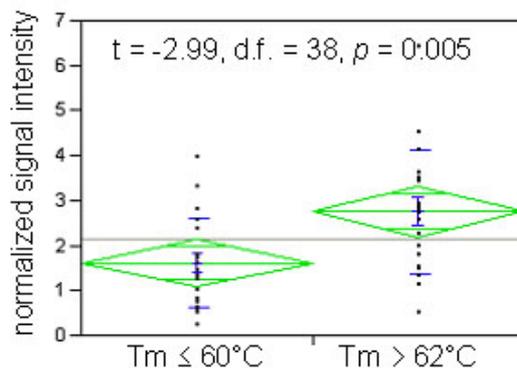


Abb. 9: Die mittlere Signalintensität lag bei den Oligos mit tieferem Schmelzpunkt (T_m) signifikant tiefer (1.63 ± 1.0 SD) als bei denjenigen mit einem hohen T_m (2.77 ± 1.39 SD).

3 – Entwicklung des Gene Chips

Parallel zu den Entwicklungen am Genome Chip haben wir eine Methode entwickelt, die auf der „klassischen“ Strategie für den Einsatz von Microarrays basiert. Hier werden bestimmte Genabschnitte amplifiziert und sequenziert. Auf der Basis dieser Information werden dann gezielt Proben designed, die die entsprechenden zu differenzierenden Taxa identifizieren können. Der Prozess besteht im Wesentlichen aus drei Schritten: 1) Amplifikation und Sequenzierung von Gen-Fragmenten, 2) Gruppenbildung (zB durch Clusteranalyse) und entsprechendes Sequenz-Alignment, 3) Design von Art- bzw. Taxon- spezifischen Proben. Wir haben durch eine Kombination aus publizierten Verfahren eine einfache und robuste Methode entwickelt, die mit sehr preisgünstigen Glasslides ausgezeichnete Resultate liefert. Wir haben diese Methode bei verschiedenen Organismen eingesetzt und teilweise publiziert, In einer Arbeit wurde sie zur Identifikation verschiedener, anhand morphologischer Merkmale nur sehr schwierig zu unterscheidenden Kleinsäugetern (Mäusen und Spitzmäusen) eingesetzt (Abb. 10; Pfunder et al., 2004; Reprint beigelegt).

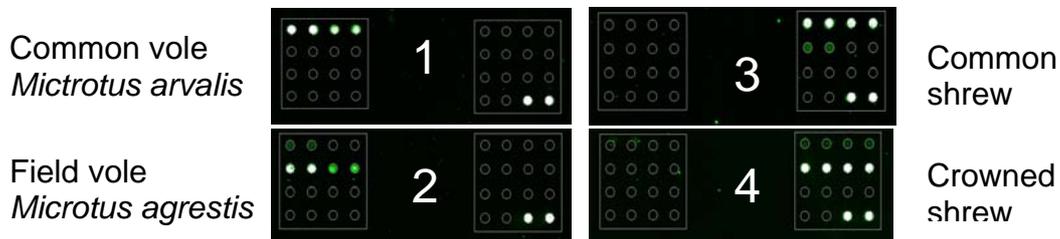


Abb. 10: Beispiele des Identifikationsmusters von Microarrays zur Artenidentifikation von verschiedenen Kleinsäugetern.

In einer weiteren „proof-of-principle“ Studie wurde geprüft, ob diese Methode für die Differenzierung auf der Stufe der Serogruppen bzw. –Untergruppen dieselbe Auflösung bietet wie die heute als Standard geltende ELISA Methode. Die Resultate zeigen, dass der Microarray ausgezeichnet geeignet ist, solche Diagnosen durchzuführen (Abb. 11; Zhang et al., 2004; Preprint beigelegt). Schliesslich konnten wir auch erfolgreich zeigen, dass humanpathogene Mikroorganismen (Mycobakterien) mit derselben Methode sehr gut identifiziert werden können.

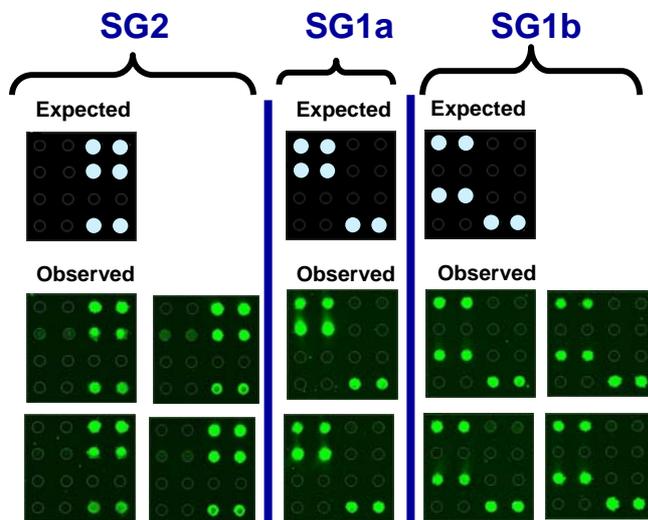


Abb. 11: Identifikationsmuster von Microarrays zur Zuordnung von 10 verschiedenen Stämmen des Gurkenmosaikvirus zu den Serogruppen 1a (SG1a), 1b (SG1b), und 2 (SG2).

4 – Diskussion der vorläufigen Ergebnisse

Mit Hilfe des Computers wurden 412 13mer Random-Oligos entworfen, die theoretisch sehr ähnliche physikalische Eigenschaften besitzen sollten. Aber sowohl bei den Labelling-Experimenten, als auch beim Hybridisieren stellten wir fest, dass die Theorie der Praxis nicht standhält. Grosse Unterschiede zwischen den einzelnen Oligos lassen eine Harmonisierung innerhalb eines Arrays nicht zu, weshalb wir uns entschieden haben, die Selektion in Zukunft auf dem Chip durchzuführen. Es besteht auch die Möglichkeit, basierend auf unseren Hybridisierungsdaten einem Algorithmus zu suchen, der präzisere Ergebnisse liefert als die zur Zeit zur Verfügung stehenden.

Das neue Verfahren zur Amplifikation ganzer Genome (rolling circle amplification) löst eines unserer Probleme. Ein anderes grosses Problem, die Hybridisierung zwischen den vielen einzusetzenden Primern, konnte mit einer neu entwickelten Strategie gelöst werden. Dabei werden die Primer mit einem 5'-„Tag“ (einer Verlängerung aus spezifischen Basen) versehen. Die Labeling-Reaktion findet in einem mit Streptavidin ausgelegten Tube statt. Dieses wurde mit einem „Anker“, einem biotinylierten Oligo mit zum 5'-Tag inverser Basensequenz, belegt. So werden die der Reaktion hinzugegebenen Primer gebunden, wodurch eine Kreuzhybridisation ausgeschlossen wird. Wir konnten zeigen, dass diese Strategie funktioniert und wir werden in Zukunft damit arbeiten.

Eine weitere Limitation stellt die Sensitivität dar. Nebst anderen Ansätzen werden wir von Fluoreszenz-Labels auf Gold umstellen und versprechen uns davon eine Verbesserung um mindestens das 10-Fache. Zusammen mit den Verbesserungen durch das veränderte Waschprotokoll sollten wir also mindestens 15-fach verbesserte Signalstärken erhalten. Diese Evaluierung dieser Neuerungen und weitere Versuche zur Steigerung der Sensitivität des Detektionssystems werden uns im Jahr 2005 beschäftigen.

5 – Ueberblick über den Stand der Entwicklung von diagnostischen Mikroarrays für Mikroorganismen (Stand Anfang 2004)

Microarrays werden bereits in vielen Bereichen der mikrobiellen Forschung eingesetzt. Ye et al. (2001) und Cook and Sayler (2003) geben einen generellen Ueberblick über die Technologie und die aktuellen Einsatzbereiche. Die Problematik der Detektion von bakteriellen Pathogenen wird von Call et al. (2003) zusammengefasst. Speziell auf die Detektion von Viren geht Wang et al. (2002) ein, wobei sie eine interessante Strategie für den Einsatz von Oligonucleotiden auf Microarrays entwickeln.

Diese Arbeiten und die darin zitierte Literatur sowie die diesem Bericht angefügte Bibliographie zeigen, dass im Bereich der Entwicklung von molekularen Methoden der Identifikation von Mikroorganismen eine immense Dynamik herrscht, die von den politischen Entwicklungen der letzten Jahre noch zusätzlich beschleunigt wurde. Das Ziel dieser Entwicklungen ist es, in kurzer Zeit und mit hoher Zuverlässigkeit einzelne Arten von Mikroorganismen, die generell als Mischpopulationen vorliegen, zu identifizieren und deren pathogene Eigenschaften zu charakterisieren. Diese beiden Aufgaben, die Identifikation und die Charakterisierung, werden auch heute noch in der Regel in voneinander unabhängigen Analysen durchgeführt (Call et al., 2003).

Identifikation von Mikroorganismen

Eine Vielzahl an Methoden wurden für PCR-basierende Methoden (RFLP, AFLP, Sequenzierung, SNP) beschrieben und sind zur Zeit in Entwicklung. Diese können im Wesentlichen in zwei Gruppen eingeteilt werden, solche, die die genetische Variation über das gesamte Genom charakterisieren (Genom Analysen), und solche, die diese in einem bestimmten Bereich eines Gens untersuchen (Genspezifische Analysen).

a) Genom Analysen

Der Vorteil dieser Methoden ist, dass eine sehr feine Auflösung bis hin zur eindeutigen Stammzugehörigkeit möglich ist. Auch können mit diesen Methoden bisher unbekannte Organismen charakterisiert werden. Der Nachteil der einfacheren unter diesen Methoden besteht allerdings darin, dass nur artenreine Proben zu einer Identifikation führen. Die beiden am häufigsten eingesetzten Techniken sind RAPD (random amplification of polymorphic DNA) und AFLP (amplification fragment length polymorphism). Während die RAPD Technologie sehr einfach durchgeführt werden kann, sind zur Sicherstellung der Reproduzierbarkeit der Daten einige Massnahmen notwendig (zB aufgereinigte DNA, exakte Quantifizierung, redundante Probenverarbeitung). Die AFLP Technik ist anspruchsvoller, man erzielt damit aber leichter reproduzierbare Ergebnisse. Die Analyse auf einem hochauflösenden Gel oder in einem Kapillarelektrophoresegerät ist komplex und sollte mit (teurer) Spezialsoftware durchgeführt werden (www.keygene-genetics.com). Bisher ist uns nur eine Arbeit bekannt, die diese Methoden mit der Microarray-Technologie zusammenführen, obwohl hierzu ein interessantes Potential besteht (Salazar and Caetano-Anollés, 1996). Im Grundsatz entsprechen die Daten, die mit diesen Methoden erzeugt werden, denjenigen, die vom Genom Chip erwartet werden können. Allerdings hat die Genom Chip Strategie den Vorteil, dass mit einer einzigen Reaktion viel mehr Marker erzeugt werden können.

Eine aufwändige, jedoch sehr erfolgreiche Methode zur Beschreibung von unbekanntem Mikroorganismen in komplexen Systemen ist die heute weit verbreitete Praxis des Klonierens von rRNA von Mischpopulationen von Mikroorganismen, gefolgt von DNA Sequenzierung (Kurzer Ueberblick in DeLong, 2002). Der Vorteil der Methode liegt darin, dass mit Ausnahme sehr seltener (mit sehr geringer Abundanz) Mikroorganismen-Arten, alle in der Probe vorkommende Arten, also auch unkultivierbare, charakterisiert werden können. Allerdings ist der dazu notwendige Aufwand sehr hoch. Die so ermittelten Sequenzen werden heute immer häufiger zur Entwicklung von Microarrays zugezogen.

b) Genspezifische Analysen

Bei diesen Analysen wird die Variation in einem oder mehreren spezifischen Gen(en) zur Identifikation herangezogen. Einer der häufigsten und zugleich zuverlässigsten Marker zur Identifikation von Mikroorganismen ist die 16S rDNA. Obwohl bekannt ist, dass dieses Gen insgesamt nicht genügend Variation aufweist, um alle Mikroorganismen-Arten identifizieren zu können (Fox et al., 1992), hat es sich beim Einsatz in bestimmten Nischen sehr bewährt. So müssen zur Diagnose von Humanpathogenen keine Nicht-Humanpathogene berücksichtigt werden. Durch die Vielfalt an Mikroorganismen ist die Wahrscheinlichkeit eines falsch-positiven Signals sehr gering. Allerdings kann dieses Gen nur selten Auskunft über die Stammzugehörigkeit eines Zielorganismus geben, da die Variation innerhalb einer Art zu gering ist (Dahllöf, 2002). Ein weiteres Problem besteht im Polymorphismus, der von einem Multi-Copy Gen wie 16S innerhalb eines Bakterienstammes verursacht wird (Dahllöf et al., 2000). Alternative Gene werden daher evaluiert (zB IGS, ITS, einige single copy genes wie *rpoB*), werden aber über längere Zeit nur in einigen Nischen eine Bedeutung haben, da vergleichende Daten fehlen. Alle auf spezifischen Genen basierenden Identifikationsmethoden können auf Microarrays umgesetzt werden. Dazu sind allerdings

spezifische Algorithmen notwendig, die aus den vergleichenden Sequenzdaten optimierte Microarray-Probes extrahieren können (Rouillard et al., 2003; Kaderali and Schliep, 2002; Borneman et al., 2001; Herwig et al., 2000). Wie unsere Erfahrungen zeigen, ist dies nicht trivial (s. Resultate).

Charakterisierung der Pathogenizität von Mikroorganismen

Die Identifikation eines Pathogens sagt oft noch nichts über seine Pathogenizität aus. Es ist daher notwendig, in weiteren Analysen verschiedene Gene zu untersuchen, von welchen bekannt ist, dass sie für die Pathogenizität eines Organismus verantwortlich sind. Die Basis dazu bildet eine Genfragment-spezifische PCR-Reaktion. Die PCR Produkte werden dann mit den notwendigen Nachweismethoden untersucht (Agarosegel-, Polyacrylamid- oder Kapillarelektrophorese zur Fragmentanalyse; Sequenzierung). Zu solchen Analysen existiert eine wichtige Literatur, die hier nicht wiedergegeben wird.

Die Kombination der Sequenzinformation zu Pathogenizitätsgenen mit Microarrays bedeutet einen wichtigen Fortschritt, da die genetische Information von mehreren Genen gleichzeitig untersucht werden kann. Ein ausgezeichnetes Beispiel eines robusten und spezifischen Microarray-basierenden Systems zur Aufklärung der Pathogenizität von Mikroorganismen ist die Arbeit von Chizhikov et al. (2001). Sie verwendeten eine Multiplex-PCR Reaktion zur Amplifikation von sechs Genen, die unterschiedliche bakterielle Antigen-Determinanten und Virulenzfaktoren codieren (*rfbE*, *fliC*, *eaeA*, *slt-I*, *slt-II*, *ipaH*). Mit Sequenz-spezifischen Probes konnten sie die Präsenz dieser Faktoren in allen 15 getesteten Mikroorganismen (*Salmonella*, *Shigella*, und *E.coli* Stämme) nachweisen. Die Methode war den anderen Nachweismethoden überlegen (Hybridisations-Test und Agarose-basierende PCR Produkt Analyse). Auch zu diesen Microarray-basierenden Methoden findet man bereits eine Vielzahl von Publikationen, einige der Wichtigsten sind im Literaturverzeichnis wiedergegeben. Viele dieser Arbeiten beschäftigen sich zur Zeit mit Optimierungen, die insbesondere die Hybridisierung betreffen, da hier offenbar die heikelsten Probleme zu lösen sind (zB Spezifität, Cross-Hybridisierung; Small et al., 2001; Peplies et al., 2003; Loy et al., 2002).

Beilagen

- Deyong Z, Willingmann P, Heinze C, Günter A, Pfunder M, Frey B, Frey JE (2004) Differentiation of cucumber mosaic virus isolates by hybridisation to oligonucleotides in a microarray format. *Journal of Virological Methods*, in press.
- Pfunder, M, Frey, JE (2005) Dissociation analysis in PCR- and 1xSSC-buffer as a prerequisite for selection of 13mer microarray probe sets with uniform hybridization behaviour. *Molecular Biotechnology*, in press.
- Pfunder M, Holzgang O, Frey JE (2004) Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome b as genetic markers. *Molecular Ecology* 13: 1277-1286.