



OKEE

Optimalisatie van Keten Efficiëntie via Expressieprofielering

Voortgangsrapportage Juni 2004 – Januari 2005

2003T1269 / April 2005

Vertrouwelijk

Consortium

Agrotechnology and Food Innovations BV

Plant Research International BV

KeyGene NV

Productschap Tuinbouw

Report 423



OKEE

Optimalisatie van Keten Efficiëntie via Expressieprofielering

Voortgangsrapportage Juni 2004 – Januari 2005

2003T1269 / April 2005

Vertrouwelijk

Consortium

Agrotechnology and Food Innovations BV

Plant Research International BV

KeyGene NV

Productschap Tuinbouw

Report 423

22410571

Colofon



Titel	OKEE Optimalisatie van Keten Efficiëntie via Expressieprofilering
Auteur(s)	Consortium
A&F nummer	423
ISBN-nummer	-
Publicatiedatum	April 2005
Vertrouwelijk	Vertrouwelijk
Project code.	1330002000

Agrotechnology & Food Innovations B.V.
P.O. Box 17
NL-6700 AA Wageningen
Tel: +31 (0)317 475 024
E-mail: info.agrotechnologyandfood@wur.nl
Internet: www.agrotechnologyandfood.wur.nl

© Agrotechnology & Food Innovations B.V.

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, hetzij mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever. De uitgever aanvaardt geen aansprakelijkheid voor eventuele fouten of onvolkomenheden.

All right reserved. No part of this publication may be reproduced, stored in a retrieval system of any nature, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher. The publisher does not accept any liability for the inaccuracies in this report.

Dit rapport is goedgekeurd door: Herman Peppelenbos



Het kwaliteitsmanagementsysteem van Agrotechnology & Food Innovations B.V. is gecertificeerd door SGS International Certification Services EESV op basis van ISO 9001:2000.

Abstract

In this report, the progress of the EET project named OKEE is explained. The three collaborating research partners have made progress and are basically on schedule as planned. The tomato quality project continued with the microarray hybridization experiments. After optimization of the hybridization protocol, good results were obtained which made it possible to compare batches of different storage time and batches of different quality on the basis of gene expression. This analysis resulted in the observation that gene expression changed most between days 0 to 3 and less after this period. By comparisons between different quality batches, sampled at day 0, candidate indicator genes could be identified.

The same holds through for the rose-botrytis part of the project. The planned sample range was analysed by microarray hybridization and after extended data analysis a selection of 30 candidate indicator genes could be selected for further experiments.

The Multiplex Quality Quantifier assay is making very good progress. This period the method was used to test the expression of rose genes, both reference genes with constant expression for normalization and test genes with known differences in expression among different developmental stages of the flower with or without fungal infection. The QQ-assay could quantify relative differences in gene expression and therefore appears to be very promising for applying on the candidate genes of tomato and rose in the coming period.

Content	
Abstract	3
1 Project gegevens	5
2 Voortgang in relatie tot de planning	7
2.1 Voortgang per taak over het afgelopen halfjaar	8
2.1.1 Ketenganalyse	9
2.1.2 Ontwikkelen kwaliteitsverloopmodellen	10
2.1.3 Analyse en definitie expressieprofielen	10
2.1.4 Ontwikkeling multiplex toetsmethode	11
2.1.5 Integratie en validatie	11
2.2 Kosten overzicht	11
2.3 Mijlpalen komend halfjaar	12
3 Resultaten	13
3.1 Overzicht onderzoeksresultaten	13
3.1.1 Botrytis kandidaatgenen selectie	13
3.1.2 Tomaat kandidaatgenen selectie	15
3.1.3 Multiplex Quality Quantifier toets	19
3.2 Knelpunten en oplossingen	23
3.3 Octrooiaanvragen	23
3.4 Interne rapportages	23
3.5 Openbare publicaties	24
4 Conclusies	25

1 Project gegevens

Projectcategorie: **EET project**

Titel: **Optimalisatie van Ketens Efficiëntie via Expressieprofilering (OKEE)**

Reductie van energieverbruik en uitval in de distributieketens van verse producten door innovatieve toepassing van genomics technologie.

Projectnummer: EETK01122

Verslagperiode: 1 juli 2004 tot 1 januari 2005

Penvoerder: Dr. J.J. Mes, Agrotechnology and Food Innovations BV

2 Voortgang in relatie tot de planning

Het doel in de drie jaar van het project is het leveren van een 'proof of principle' voor het concept dat complexe genexpressie profielen te koppelen zijn aan mathematische kwaliteitsverloopmodellen. Deze methodiek resulteert in een beslissingsondersteunend instrument. Dit instrument zal de keten helpen in optimalisatie van de efficiëntie en het verbeteren van de duurzaamheid van verse producten in de distributieketens.

De grove planning van het project was dusdanig opgebouwd dat je zou kunnen spreken van 3 fases, die samenvallen met de projectperiode van 3 jaar. Het eerste jaar zou gebruikt worden voor het maken van de microarrays van zowel tomaat en roos, het bemonsteren van partijen uit de praktijk en het verkennen van de verschillende mogelijkheden voor het snel detecteren van genexpressie via een multiplex ligatie en amplificatie systeem. Het tweede jaar zou dan gebruikt worden voor microarray analyse, microarray data verwerking in combinatie met kwaliteitseigenschappen en het testen van indicator genen met andere methoden voor het detecteren van genexpressie. Het multiplex detectiesysteem zou in deze periode verder ontwikkeld worden, daar mogelijk met de eerste indicator genen uit de microarray analyse. In het derde jaar zal dan verificatie en betrouwbaarheid van de indicator genen getest worden, zo mogelijk met de ontwikkelde multiplex methode, en zal de link onderzocht worden met de kwaliteitsmodellen en bewaarvoorspellingen.

Om dit te realiseren is het eerste jaar een analyse gemaakt van de meest voorkomende distributieketens voor de voorbeeldproducten roos en tomaat. Daaruit zijn de keuzes voor tomaat in relatie tot stevigheid en bij roos de botrytisaantasting naar voren gekomen als belangrijke knelpunten bij het opzetten van productgebaseerde ketens. Vervolgens zijn er proefopzetten ontwikkeld waarin relaties gelegd kunnen worden met de intrinsieke kwaliteit van het product. De resultaten van het onderzoek naar het kwaliteitsverloop van regulier geteelde producten bevestigde het onvoorspelbare karakter van de kwaliteitseigenschappen en daarmee het nut die het hier beoogde instrument in de praktijk zal hebben. Gebruikmakend van de eerste variabele partijen zijn cDNA banken gemaakt en daaruit vervolgens de microarray samengesteld. De eerste helft van het tweede jaar is vervolgens gebruikt om de microarray hybridisatie protocollen voor roos en tomaat te optimaliseren. Na eerste verkennende experimenten op ontwikkelingsreeksen van tomaat en roos zijn vervolgens de batches geanalyseerd waarbij genexpressie op dag 0 gecorreleerd moet worden met kwaliteitsdata van dezelfde producten. Bij tomaat gaat het hier stevigheidwaarnemingen uitgedrukt in het aantal dagen na de pluk tot stevigheidsgrens 5 terwijl bij roos het om Botrytis besmetting van de bloemen gaat uitgedrukt in % besmette bloemen of een andere ziekte index.

De resultaten van de microarray analyse bleken minder eenvoudig te interpreteren dan gehoopt. Door gebruik te maken van verschillende statistische bewerkingen en analysemethoden zijn we echter in staat gebleken genen te selecteren die mogelijk gebruikt kunnen worden als kwaliteitvoorspellers. Ook hebben de gevonden genen geleid tot speculaties en nieuwe hypotheses over het hoe en waarom batches variëren in kwaliteit. Voordat er echter van echte

indicatoren sprake is dienen genexpressies eerst met andere methoden geverifieerd te worden en zullen ook meer partijen geanalyseerd dienen te worden dan nu gebruikt voor microarray analyse. Nu de eerste potentiële indicator genen geïdentificeerd zijn en met nog een jaar aan verificatie en validatie periode voor de boeg lijken we volledig aan de planning te voldoen.

Dit geldt tevens voor de multiplex toetsontwikkeling. Vanuit de eerste verkennende experimenten met bekende sequenties van tomaat, optimalisatie van de OLA-PCR probes en detectie op cDNA en RNA is de methode doorontwikkeld tot een betrouwbaar systeem waarbij op basis van een snelle mRNA isolatie, first-strand cDNA synthese, OLA hybridisatie en PCR genexpressie gedetecteerd kan worden. Op basis van de expressie van het RIN gen van tomaat is aangetoond dat de methode veel potentie heeft, zeker in combinatie met een bufferende reeks aan constante genen. Aan dit laatste is dan ook het meest gewerkt de afgelopen periode waarbij direct met echte roosgenen en roosmateriaal is gewerkt. Een volgende periode zal gericht zijn op de verificatie en analyse van potentiële kandidaat-genen en het vergelijken van multiplex resultaten met microarray en Real-Time PCR data.

De haalbaarheidsindicatoren hoeven dus niet bijgesteld te worden, het project draait en maakt voortgang zoals gepland.

2.1 Voortgang per taak over het afgelopen halfjaar

De uitvoering van dit project valt in een aantal taken uiteen die deels parallel uitgevoerd worden. Sommige van deze taken worden uitgevoerd over een langer tijdbestek (zie timechart, figuur 1).

TAAK 1 - Ketenanalyse

Een constante interactie met de praktijk zal onderhouden worden om mogelijke veranderingen in de knelpunten en logistieke problemen die invloed hebben op dit project in een vroegtijdig stadium op te merken en mee te kunnen nemen in besluitvorming en richting geven aan het onderzoek. Resultaten van de kwaliteitsverloopproeven zullen teruggekoppeld worden met de praktijk om de knelpunten te evalueren en te toetsen of we op het goede spoor blijven.

TAAK 2 - Ontwikkelen kwaliteitsverloop modellen (KVM)

Kwaliteitsverloopproeven zullen (statistisch) geanalyseerd worden met als doel de producten te groeperen in kwaliteitsklassen. Deze klassenindeling zal gebruikt worden om tot een keuze van de monsters voor genexpressie analyse te komen. Voor tomaat zijn hiervoor in het afgelopen halfjaar nieuwe partijen geanalyseerd. Dit heeft extra informatie opgeleverd over de kwaliteit van de producten. Deze resultaten zullen teruggekoppeld worden met de praktijk en meegenomen worden bij de ketenanalyse en modelvorming.

TAAK 3 - Analyse en definitie expressieprofielen

Vele microarray hybridisaties zijn uitgevoerd gericht op het bepalen van gen expressie in het product aan het begin van de keten en die te correleren aan kwaliteit tijdens de naoogst fase. Uitgebreide data analyse met verschillende statistische programma's hebben geleid tot kandidaat indicatorgenen.

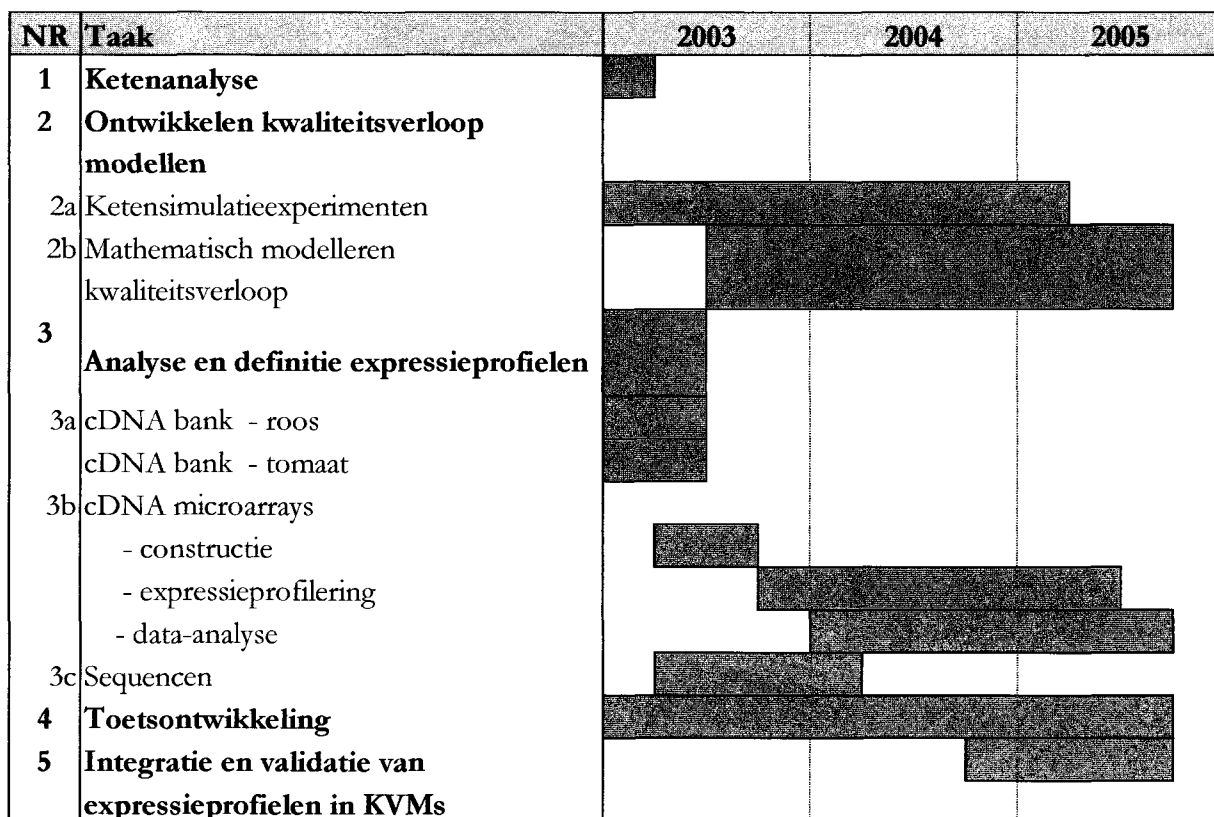
TAAK 4 - Ontwikkeling multiplex toetsmethode

De eerste expressie data van roos zijn gebruikt om vier indicatorgenen en twaalf constitutief tot expressie komende referentiegenen te selecteren. Een Quality Quantifier multiplex toets is ontwikkeld en ingezet om een eerste indruk te krijgen van de mogelijkheden om genexpressie verschillen kwantitatief te detecteren. De expressie data zijn met een nog in ontwikkeling zijnde methode geanalyseerd.

TAAK 5 - Integratie en validatie van KVMs en gen-expressie-profielen voor roos en tomaat

Een eerste aanzet is gemaakt om gen expressie en kwaliteitsverloop modellen aan elkaar te koppelen.

Een fasering per taak met tijdsplanning is weergegeven in de Gantt-chart. De voortgang per taak zal hieronder besproken worden.



Figuur 1 Timechart van het project verdeeld over de taken.

2.1.1 Ketenanalyse

Door een constante interactie met de praktijk zullen veranderingen in de knelpunten en logistieke problemen die invloed hebben op dit project meegenomen worden bij het bepalen van de richting van onderzoek. Tot op heden hebben zich geen veranderingen in de ketens zoals opgesteld in een eerdere fase van het project voorgedaan en is er dus geen aanleiding het project zoals ingeslagen bij te sturen.

2.1.2 *Ontwikkelen kwaliteitsverloopmodellen*

Op basis van kennis en expertise van het product is een proefopzet ontwikkeld voor zowel roos als tomaat waarin het kwaliteitsverloop efficiënt is te bepalen en te meten. Het afgelopen halfjaar is het aantal monsters voor tomaat uitgebreid wat geen nieuwe conclusies tot gevolg heeft gehad. Hetzelfde beeld blijft dus overeind dat er veel variatie is in kwaliteit bij de hier gebruikte voorbeeld producten, roos en tomaat.

Bij roos zijn we tot een indeling gekomen van drie kwaliteitsklassen: goed, matig en slecht. Op basis van deze indeling, die ook statistisch onderbouwd is, zijn er uit elke kwaliteitsklasse 4 partijen geselecteerd voor microarray analyse. Daarbij is naast de variatie in botrytis aantasting die de kwaliteitsklasse bepaalt ook gekeken naar een evenwichtige verdeling van andere parameters per klasse zoals herkomst, jaargetij van experiment en bloemstadium van de bloemen op dag van bemonsteren.

Voor tomaat zijn ook meerdere partijen, alle op dezelfde manier, geanalyseerd op verlies van stevigheid. Dit bleek een succesvolle proefopzet die een zeer duidelijk beeld van de invloeden op kwaliteit hebben aangetoond. De kwaliteit van de verschillende batches tomaat lijkt in vier kwaliteitsklassen indeelbaar: slecht, matig, goed en super. Een selectie van samples is gebruikt voor microarray analyse waaronder tijdreeksen en dag0 samples van uit verschillende kwaliteitsklassen.

De voortgang voldoet hiermee aan de planning.

2.1.3 *Analyse en definitie expressieprofielen*

In 2003 zijn de voorbereidingen uitgevoerd voor het maken van de microarray. Daarvoor zijn eerst banken gemaakt, de DNA volgorde van de klonen bepaald, een selectie gemaakt van klonen die gespot zouden worden waarna deze vermenigvuldigd zijn en opgezuiverd. Begin 2004 zijn de microarrays gespot zodat het begin van dit jaar met microarray hybridisatie begonnen kon worden. Na het opzetten van betrouwbare en reproduceerbare protocollen zijn de eerste hybridisaties uitgevoerd op ontwikkelingsreeksen (bloem en Botrytis van roos en rijpings reeks van tomaat).

Het afgelopen halfjaar heeft het onderzoek zich geconcentreerd rond de kwaliteitsklassen en de voorspelling van deze klassen. Microarray analyses zijn uitgevoerd met als doel het vinden van kandidaatsindicatoren. Voor roos zijn daarvoor 36 microarrays uitgevoerd waarbij 4 monsters uit elke kwaliteitsklasse in drievoud is geanalyseerd. De gegenereerde data zijn met diverse programma's geanalyseerd op potentiële kandidaatgenen die een correlatie geven met de kwantiteit van een dergelijk monster tijdens de bewaring. Deze analyse heeft geleid tot 35 potentiële kandidaatgenen die in eerste instantie gevalideerd zullen worden. Echter, het is mogelijk nog meer kandidaatgenen aan te wijzen indien nodig.

In het afgelopen halfjaar is ook voor tomaat een groot aantal microarrays gehybridiseerd en geanalyseerd. Hierbij is vooral gekeken naar monsters die in de kwaliteitsverloopprouwen het duidelijkst van elkaar verschilden. Uit die vergelijkingen bleek dat de resultaten van een vroeg oogsttijdstip zich onderscheiden van alle andere getoetste tijdstippen. Aldus zijn, na een analyse van de correlatie tussen genexpressie en kwaliteitsverloop, twee sets van indicatorgenen naar

voren gekomen, die deels overlappen (totaal: 16-20). Deze kandidaatgenen zullen in eerste instantie verder gevalideerd worden. Analyse van nog voorhanden zijnde monsters en van nieuwe monsters uit 2005 kunnen deze validatie verder versterken, en eventueel nog nieuwe indicatorgenen opleveren.

Met de identificatie van potentiële indicator genen aan het begin van het laatste jaar verloopt ook dit onderdeel volgens planning.

2.1.4 Ontwikkeling multiplex toetsmethode

Basis voor de te ontwikkelen multiplex toets is de door Keygene geïntroduceerde Quality Quantifier technologie. Hoogspecifieke DNA target herkenning op basis van oligo ligatie assays (OLA) en daaropvolgende amplificatie, gecombineerd met detectie van vooraf bepaalde lengtes van de specifieke ligatie producten wordt toegepast voor bepaling van mRNA (gen expressie) niveaus. De analyses uitgevoerd op een voorbeeld gen RIN van tomaat toonde aan dat deze methode zeer bruikbaar lijkt voor genexpressie vergelijking met meerdere genen tegelijk. Voortbouwend op de resultaten met het RIN gen is een voorlopige set genen van roos geselecteerd bestaande uit een viertal indicatorgenen voor botrytis infectie en twaalf referentie genen. Hierbij is een uitsplitsing gemaakt naar drie oplopende expressie niveaus, zoals gedetecteerd op de microarray. Onderzocht is of deze indicator en referentie genexpressie niveaus en verschillen tussen bloeminfectie stadia gedetecteerd konden worden in multiplextoetsen van alle genen of subsets hiervan. OLA-expressie data zijn naar intensiteit gekwantificeerd en onderzocht op de daadwerkelijke detectie van differentiële expressie van potentiële indicatorgenen ten opzichte van de gekozen (constante) referentiegenen. De data zijn met een (in ontwikkeling zijnde) kwantificatieprotocol geanalyseerd.

2.1.5 Integratie en validatie

Met het beschikbaar komen van genexpressie data is een eerste aanzet gemaakt tot het koppelen van kwaliteitsverloop met genexpressie. Aangezien genen nog gevalideerd moeten worden en deze validatieproeven mogelijk betrouwbaardere gen expressie profielen opleveren zal gewacht worden met uitgebreidere analyses tot meer gegevens beschikbaar zijn.

2.2 Kosten overzicht

De gemaakte kosten over de afgelopen periode bedragen voor A&F 22% van de totaal toegekende subsidie en komen daarmee op een uitputting van bijna 75% van de totale project toezegging. Met nog 1 jaar te gaan en de grootste inspanning op moleculaire analyse achter ons, liggen we hiermee goed op schema. Plant Research International heeft over deze periode 7,3% gedeclareerd en komt daarmee op 74 % uitputting op het toegezegde totaal bedrag. KeyGene heeft 8,6% gedeclareerd en komt op een totale uitputting van 59%. KeyGene heeft het geld voor het ontwikkelen dus goed verdeeld over de jaren met nog een groot gedeelte van het totaal bedrag over voor de grootschalige analyse van de kandidaatgenen over verschillende praktijk samples voor verificatiestudies. Net als in de vorige rapportageperioden heeft Productschap

Tuinbouw nog geen declaratie aangeleverd en is inmiddels afgesproken met PT deze declaratie helemaal aan het eind van het project pas te doen.

De projectfinanciering loopt dus volgens planning. Er zijn geen grote verschuivingen van de financiële inzet geconstateerd en worden ook niet verwacht.

2.3 Mijlpalen komend halfjaar

TAAK 1 - Ketenanalyse

Resultaten van de kwaliteitsverloop-proeven zullen teruggekoppeld worden met de praktijk om de knelpunten te evalueren en te toetsen of we op het goede spoor blijven. De komende periode verwachten we een start te kunnen maken met marktverkenning voor de in dit project beoogde kwaliteitdiagnostische toetsen.

TAAK 2 - Ontwikkelen kwaliteitsverloop modellen (KVM)

Afhankelijk van de validatie experimenten zal overwogen worden of het noodzakelijk is het komende halfjaar extra partijen van tomaat en roos te analyseren. Het kwaliteitverloop van deze partijen zal gecorrigeerd worden aan genexpressie data verkregen met Real Time PCR of multiplex toets assay resultaten.

TAAK 3 - Analyse en definitie expressieprofielen

Komend half jaar zullen de genexpressie profielen verkregen met microarray analyse gevalideerd worden met Real Time PCR en de Multiplex toets. Indien noodzakelijk zullen sequentie reacties uitgevoerd worden en primers ontworpen en geoptimaliseerd zodat betrouwbare amplificaties reacties ontstaan die de expressie nauwkeuriger kunnen bepalen dan de microarray hybridisatie experimenten. Tevens is het mogelijk dat nieuwe indicatorgenen geselecteerd en gevalideerd zullen worden op basis van goed werkende amplificaties of kwaliteitseigenschappen.

TAAK 4 - Ontwikkeling multiplex toetsmethode

In het komend half jaar zullen de geselecteerde en gevalideerde indicatorgenen van roos en tomaat getest worden in de multiplex toets in combinatie met een set van referentie-huishoud genen. De multiplex toets zal gevalideerd worden met de RealTime analysis bij de andere partners eventueel aangevuld met cDNA AFLP. Indien betrouwbaar gebleken zal de multiplex toets ingezet worden om meer monsters te screenen en de indicatoren te toetsen op hun inzetbaarheid, betrouwbaarheid en voorspellende waarde.

TAAK 5 - Integratie en validatie van KVMs en genexpressie profielen voor roos en tomaat

Op basis van gevalideerde gen expressie zullen kwaliteitverloop modellen en genexpressie aan elkaar gekoppeld worden om tot kwaliteitsvoorspellingen te komen en de marges van deze voorspellingen.

3 Resultaten

3.1 Overzicht onderzoeksresultaten

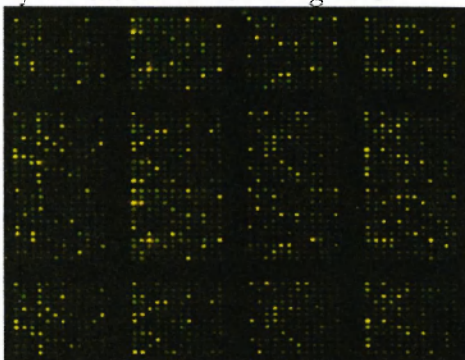
3.1.1 *Botrytis kandidaatgenen selectie*

Op basis van de reeds eerder getoonde kwaliteitsverschillen tussen diverse praktijk samples van roos en de kwaliteitsverschillen in Botrytis aantasting tijdens de uitbloeiperiode is er een selectie gemaakt van batches die verschillen in Botrytis aantasting na 7 dagen uitstal (en nog niet zichtbaar op dag 1) maar niet varieerde in andere eigenschappen zoals gemiddeld bloemstadium. In figuur 2 een overzicht van de geselecteerde partijen.

	Ieler	Botrytis	Gem botrytis	Bloemontw	Gem bloemontw
Goed	7	0.14	0.25	4.1	3.6
	21	0.26		3.0	
	26	0.30		4.3	
	12	0.30		3.0	
Midden	6	0.73	0.70	3.9	3.8
	3	0.64		3.8	
	16	0.70		4.1	
	15	0.72		3.6	
Slecht	5	1.14	1.27	3.4	3.6
	1	1.22		3.5	
	9	1.42		3.7	
	19	1.30		3.8	

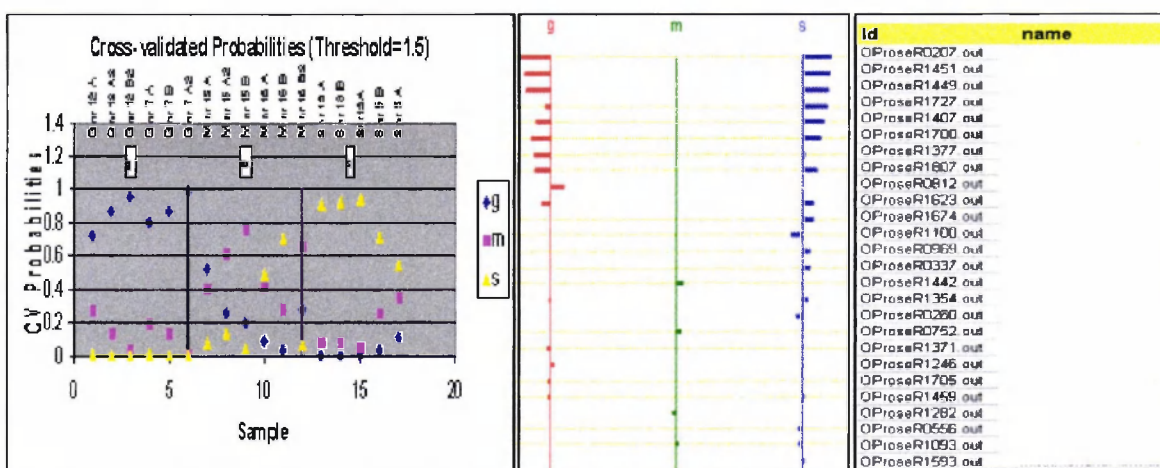
Figuur 2 Overzicht van de Botrytis aantasting (gewogen gemiddelde) en bloemstadium van de voor de microarray geselecteerde partijen.

Van de monsters die op dag 1 (dag van binnenkomst bij A&F, die zou overeenkomen met de dag van aanvoer op de veiling) zijn genomen en ingevroren is RNA geïsoleerd. Dit RNA is na kwantificeren opgedeeld in een sample gedeelte en een referent gedeelte. Het referent materiaal van alle monsters is vervolgens gepoold en telkens groen fluorescerend gelabeld. Het RNA van de monsters is vervolgens rood gelabeld waarna de microarrays gehybridiseerd zijn met een rood RNA monster en groen referent label. Omdat het referentmateriaal voor alle arrays gelijk is kunnen monsters nu direct onderling worden vergeleken. Een voorbeeld van een dergelijke hybridisatie is te zien in figuur 3.



Figuur 3 Voorbeeld van een microarray hybridisatie signaal compositie.

Na verschillende processingstappen van de fluorescerende spots die ontstaan op de plaats waar DNA fragmenten gespot zaten kan de genexpressie van de genen in de verschillende samples geanalyseerd worden. Men neemt waar of er meer of minder van een bepaald genproduct aanwezig is en dus of genen 'hoog' of 'laag' aan staan. De arrays zijn vervolgens op verschillende manieren geanalyseerd: 1) met een t-test om significant verschillende genexpressies te identificeren tussen monsters van verschillende kwaliteitsklassen, 2) met profiel correlaties bijvoorbeeld door te zoeken naar genen die hetzelfde expressie patroon hebben als de genormaliseerde *Botrytis* gegevens 3) met Predictie Analysis Microarray (PAM), een software programma dat genen selecteert die specifiek zijn voor een bepaalde klasse. Omdat het aantal mogelijke combinaties van analyses en het aantal genen per analyse methode te groot zijn om hier weer te geven zal in dit verslag alleen de PAM geïllustreerd worden (figuur 4).



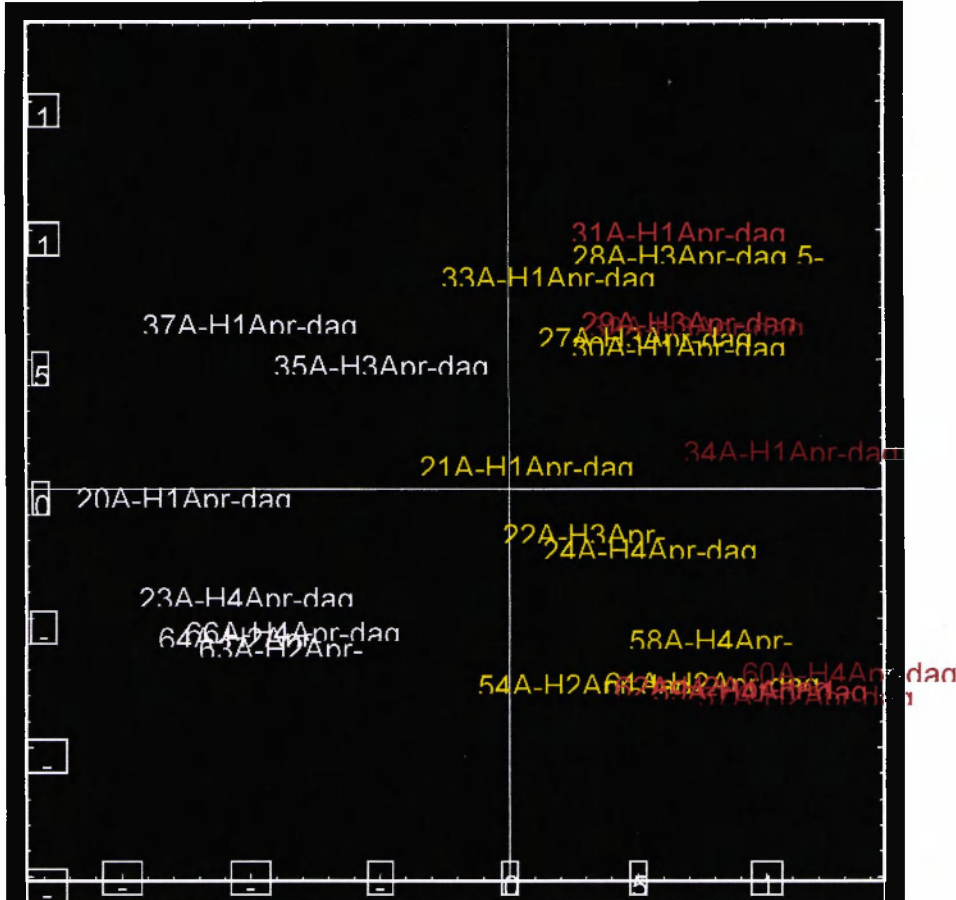
Figuur 4 PAM analyse van microarrays van roos afkomstig van 3 kwaliteitsklassen. Linker panel geeft de waarschijnlijkheid weer (bij threshold 1.5 = ongeveer 30 genen) dat de microarrays in de juiste kwaliteitsklasse worden ingedeeld. Midden panel geeft de genexpressie weer per gen specifiek voor de bepaalde klasse. Linker panel zijn de bijbehorende gen sequenties en mogelijke biologische functie.

Het programma wordt eerst getraind met microarrays waarvan de kwaliteitsklasse vrij zeker van bepaald is. Op basis van de genexpressie per gen zoekt het programma vervolgens naar genen die het meest indicatief zijn voor die bepaalde klasse van monsters. Het programma zoekt uit hoe met zo weinig genen als mogelijk zoveel mogelijk wordt voorspeld voor de aangegeven klassen. Door de threshold zo te kiezen kunnen wij vervolgens een keuze maken hoeveel genen en welke waarschijnlijkheid van predictie we acceptabel vinden. Gebruikmakend van dezelfde geselecteerde genen kan het programma vervolgens ook genexpressie van onbekende monsters gaan classificeren. Door deze analyse met verschillende aannames te starten en monsters in klassen in te delen, en gevonden genexpressie te evalueren kunnen verschillende lijstjes met kandidaat indicatoren gegenereerd worden. Van sommige kandidaatgenen weten we een mogelijke biologische functie waardoor we soms ook een hypothese rond de gevonden potentiële

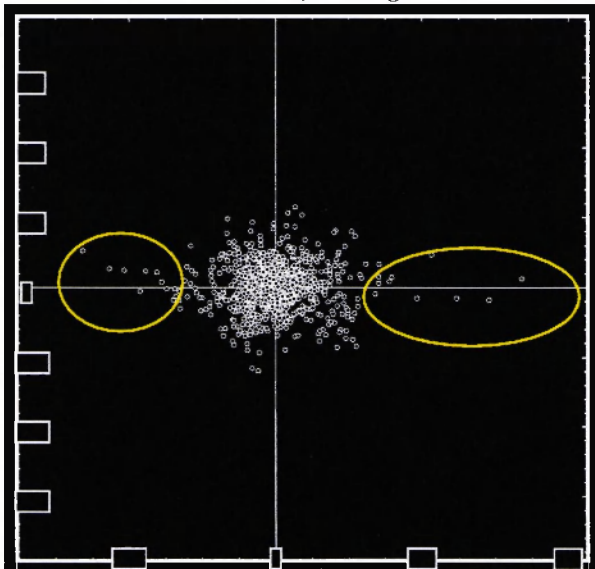
indicator genen kunnen ontwikkelen. Zo zien we een grote groep van de kandidaat indicatorgenen betrokken bij oxidatieprocessen. Mogelijk dat vatbaarheid van rozen voor botrytis dus te maken heeft met oxidatie gebaseerde afweerprocessen. Gebruikmakend van de verschillende analysemethoden zijn we tot een selectie van 35 genen gekomen die nu geanalyseerd zullen worden met andere genexpressie methoden als Real Time PCR.

3.1.2 *Tomaat kandidaatgenen selectie*

Na de aanvankelijke moeilijkheden met de kwaliteit van de hybridisaties van de tomaten microarray experimenten zijn in de tweede helft van 2005 een groot aantal hybridisaties van goede kwaliteit gedaan, en de resultaten ervan geanalyseerd. Technische uitvoering, aard van de metingen, en de presentatie ervan lijken veel op die voor roos/botrytis (hierboven beschreven). Alle hybridisaties zijn gedaan met een gemeenschappelijk referentie sample, een geamplificeerd mengsel van veel verschillende monsters, waarmee steeds 1 experimenteel sample vergeleken wordt. Voor het eerste oogstmoment, april 2004, zijn zoals al eerder gerapporteerd samples van verschillende tijden na de oogst (tijdens de kwaliteitsverloop experimenten) genomen. Dit diende om het systeem te testen (kunnen we veranderingen in genexpressie waarnemen?) en om te zien van welke genen de expressie verandert tijdens bewaren. De vraag was of partijen van verschillende herkomsten verschillen in expressie vertonen en of die gecorreleerd kunnen worden aan het verschillend in kwaliteitsverloop. Als dat laatste inderdaad zo is, dan kunnen diezelfde genen dus als indicatorgenen dienen. Anderzijds is het ook mogelijk dat genexpressieveranderingen tijdens opslag voor alle herkomsten gelijk zijn en dat potentiële indicatorgenen tot een andere set behoren. Al eerder werd gerapporteerd dat PCA (principle component analysis) een duidelijk effect van opslagtijd na oogst op de genexpressie kon onderscheiden. Figuur 5 laat zien hoe de verschillend experimenten door de PCA analyse gepositioneerd worden langs twee denkbeeldige assen, de meest invloedrijke variabele (X-as) en de een-na-meest invloedrijke variabele (Y-as) voorstellen. Monsters van dag 0 tot dag 7 zijn weergegeven (van wit naar felrood). Duidelijk is te zien dat monsters van dag 0 zich onderscheiden van de rest in hun positie langs de X-as. Er is verder nog een klein verschil tussen dag 3 en dag 5. De X-as kan hier tussen worden beschouwd als de variabele "tijd". PCA kan vervolgens de positie van alle datapunten (spots op de array) langs de twee assen weergegeven. Hoe groter de afstand van het datapunt tot het 0 punt van de as, hoe groter de rol van dat datapunt in die variatie. Door de punten te selecteren die het verst van de 0 afliggen, aan beide kanten (Figuur 6), worden de spots geselecteerd die het meest variëren met de tijd.

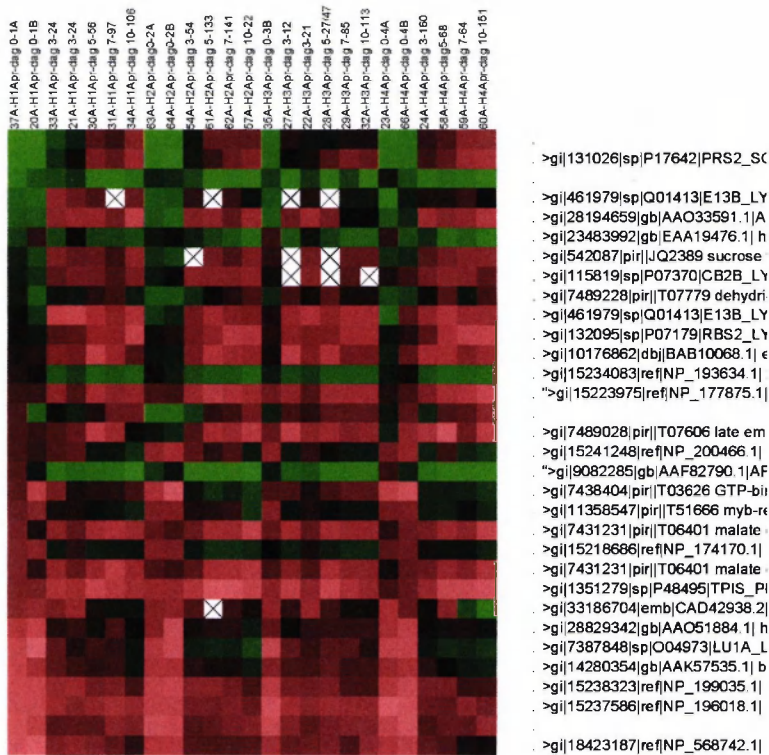


Figuur 5 PCA van de experimenten met monsters van April 2004, op dagen 0, 3, 5 en 7 (wit naar fel rood) na oogst



Figuur 6 De zelfde PCA als in Figuur 5, nu met de positie van de individuele datapunten (microarray spots) weergegeven. De spots die het meest bijdragen aan de variatie langs de X-as (variabele "tijd") zijn geselecteerd voor verdere analyse.

Vervolgens is het expressiepatroon van deze spots (ratio t.o.v. de referentie sample) weergegeven in een groen/rood patroon (groen=lager dan de ref; zwart=neutraal; rood=hoger dan de referentie), in Figuur 7.

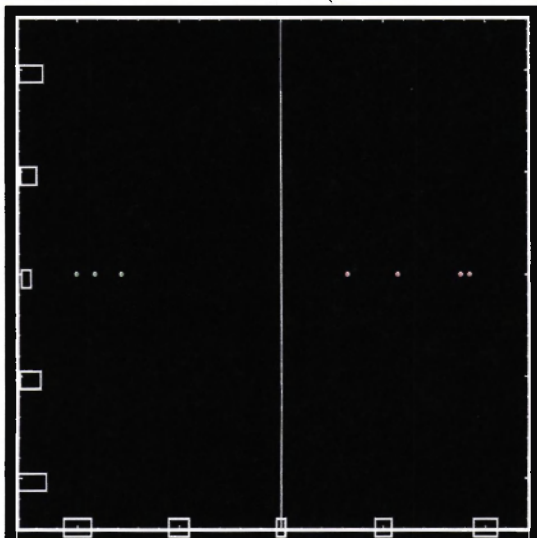


Figuur 7 Expressieratio's voor individuele spots (rijen), geselecteerd in Figuur 6, voor de verschillende herkomsten en tijden na oogst (kolommen). Kolommen (bovenaan gelabeld, H, herkomst, dag) zijn per herkomst gegroepeerd. Een deel van de annotatie van de spots is rechts getoond.

Ook in die weergave is duidelijk te zien dat de resultaten van dag 0 sterk afwijken van die van andere dagen, terwijl de verandering na dag 3 aanmerkelijk kleiner is. Hieruit concluderen we dat de meeste veranderingen in genexpressie plaatsvinden tussen dagen 0 en 3 na de oogst; daarna verandert veel minder.

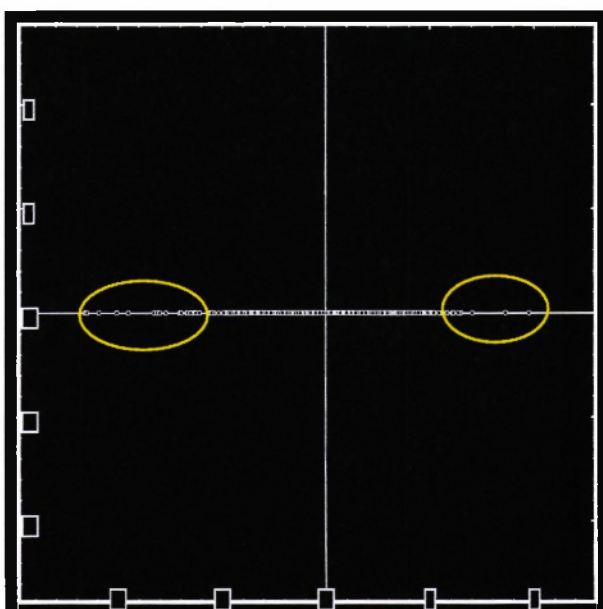
Vervolgens hebben we gekeken of tussen monsters van verschillende herkomst maar hetzelfde oogstmoment, met uiteenlopend kwaliteitsverloop, al op dag 0 alleen duidelijke verschillen in genexpressie waar te nemen zijn. Deze verschillen moeten dan leiden tot kandidaat indicatorgenen voor een toets. Hierbij zijn dezelfde technieken gebruikt. Daarbij zijn twee vergelijkingen gemaakt voor verschillend oogstmomenten: april 2004 en een combinatie van drie oogstmomenten daarna. Voor elk van deze twee momenten zijn de meeste extreme herkomsten

(in de zin van verschillend kwaliteitsverloop) gekozen voor analyse om de kans op significante verschillen te vergroten. Linear Discriminant Analysis (LDA) is een vorm van PCA waarbij vantevoren gekozen klassen datasets (hier: snelle kwaliteitsafname versus langzame afname) zo goed mogelijk onderscheiden worden op basis van de expressedata. Een voorbeeld van het scheiden van twee klassen (2 herkomsten elk) in April 2004 is te zien in Figuur 8.



Figuur 8 Linear Discriminant Analysis (LDA) van resultaten van April 2004.

Hierin is te zien hoe de experimenten in twee klassen gegroepeerd worden langs de X-as. Net als bij PCA kan ook bij LDA de positie van de individuele datapunten (spots) worden weergegeven, waar wederom de afstand tot het 0-punt een maat is voor het belang van die spot, maar nu bij het onderscheid tussen twee klassen (figuur 9).



Figuur 9 LDA van april 2004, nu met individuele datapunten weergegeven.

Ook hierbij kan weer een selectie van spots gemaakt worden, waarbij de meest onderscheidende spots voor de twee gedefinieerde klassen geselecteerd kunnen worden.

Voor beide vergelijken werden de 4% meest onderscheidende spots geselecteerd voor verdere studie. Tussen de twee geselecteerde groepen was maar gedeeltelijke overlap. Die overlap was groter met 8% van de eerste set, wat tot de conclusie leidde dat in de eerste set (april 2004) waarschijnlijk een aantal genen sterk varieerde op een niet-reproduceerbare wijze.

3.1.3 *Multiplex Quality Quantifier toets*

De resultaten met het Tomaat RIN gen dienen als uitgangspunt voor een eerste werkelijke QQ toets met roos indicator en referentie genen. Uit de microarray data zijn een aantal indicator en referentie genen geselecteerd voor ontwikkeling van OLA assays. De gekozen genen, onderverdeeld naar doel en waargenomen expressiesterkte, zijn samengevat in tabel 1.

Tabel 1 Selectie van genen voor de multiplex Quality Quantifier toets

Indicator	Expressie niveau	Constitutief-Referentie	Expressie niveau
MAL d1	sterk	Ankyrine	sterk
P450	sterk	Calmoduline	sterk
Glycosyl hydrolase	zwak	Tubuline beta 2 chain	sterk
Xylosidase	zwak	UDPglc-4-epimerase	sterk
		Glutamine synthase	middel
		Arginine decarboxylase	middel
		Beta amylase	middel
		Ubiquitine	middel
		Quinone reductase	zwak
		AAA type ATPase	zwak
		ATP synthase	zwak
		Actine 2	zwak

Vanuit de gensequenties zijn oligonucleotiden voor OLA ligatie ontworpen. Door inbouw van selectieve nucleotiden is het in de op de OLA ligatie volgende amplificatie nu ook mogelijk alleen een subset van de multiplex toets zichtbaar te maken. Hiermee kan de invloed van de intensiteit van fragmenten op de totale assay bestudeerd worden. Een schematische weergave van het toetsontwerp wordt getoond in figuur 10.

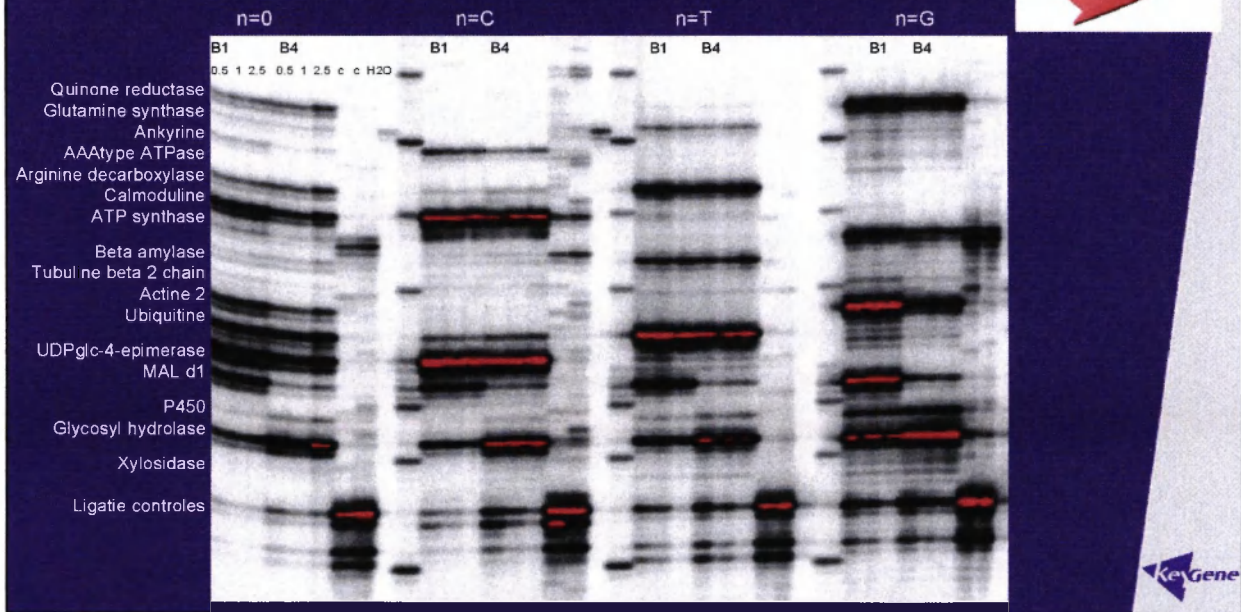
	totaal			sterk			medium		zwak	
132	G								132	G
129	T					129	T			
126	C		126	C						
123	G								123	G
120	T					120	T			
117	C		117	C						
114	G								114	G
111	T					111	T			
108	C		108	C						
105	G								105	G
102	T					102	T			
99	C		99	C						
96	N		96	N		96	N		96	N
93	N		93	N		93	N		93	N
90	N		90	N		90	N		90	N
87	N		87	N		87	N		87	N
84	N		84	N		84	N		84	N
82	N		82	N		82	N		82	N
79	N		79	N		79	N		79	N

Figuur 10 Schematische weergave van de selectieve mogelijkheden van de QQ toets. Door toepassing van selectieve amplificatie kunnen sets van genen zichtbaar gemaakt worden. Het eerste panel toont de volledige set van controles, indicatoren en referentiegenen. Ter verduidelijking tonen de verdere panels een opsplitsing gericht op scheiding van verwachte sterkte van expressie.

Uitgaande van het in figuur 1 getoond ontwerp is een QQ toets uitgevoerd ter vergelijking van twee bloem-botrytis infectie stadia, namelijk B1: bloem stadium 1/ Botrytis 0.41 en B4: bloem stadium 3/ Botrytis 0.81. Van bloemweefsel is mRNA geïsoleerd en cDNA gesynthetiseerd volgens het eerder ontwikkelde protocol. Verschillende combinaties van selectieve amplificatie zijn getest, en de resultaten van volledige (niet-selectieve) en selectieve amplificatie naar sterke, middel en zwakke expressie worden getoond in figuur 11.

Quality Quantifier

QQ detectie roos stadium B1 en B4

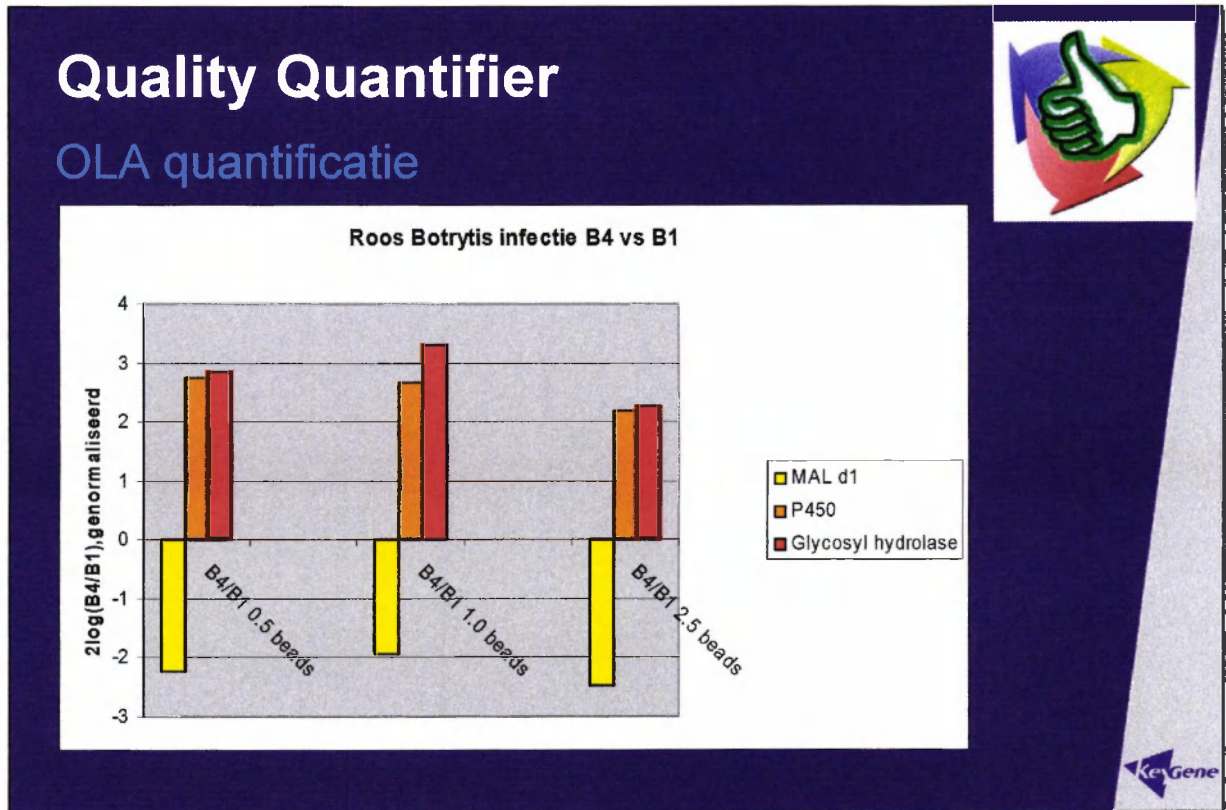


Figuur 11 Quality Quantifier toets voor Roos Botrytis infectie.

Panel n=0: toets van in principe alle OLA ligatie producten van indicator en referentie genen zonder selectieve amplificatie. Panel n=C: selectieve toets op alle indicator en de sterke referentie genen. Panel n=T: selectieve toets op alle indicator en de middelsterke referentie genen. Panel n=G: selectieve toets op alle indicator en de zwakke referentie genen. Voor ieder panel worden reacties getoond van de bloem/infectie stadia B1 (bloem1; botrytis 0.41) en B4 (bloem 3; botrytis 0.81), waarbij drie olopende hoeveelheden cDNA template zijn gebruikt (0.5, 1, 2.5 μ l)

Figuur 11 laat zien dat 3 van de vier indicatorgenen, en elf van twaalf referentiegenen een signaal geven (xylosidase en AAAtype ATPase geven geen signaal). De indicatorgenen laten signaal (expressie) verschillen zien terwijl deze bij de referentiegenen niet of veel minder aanwezig zijn. Duidelijk is ook het effect van de selectieve amplificatie zichtbaar overeenkomend met de schets in figuur 1. Opvallend is verder dat de intensiteit van de signalen op zich niet correleert met de expressieniveaus zoals vastgesteld op de microarray. De beoogde uitsplitsing naar signaalsterkte kan dus niet worden bewerkstelligd, en blijkt ook niet noodzakelijk daar alle signalen ook in het niet-selectieve experiment voldoende zichtbaar zijn. Voor kwantificatie in de QQ toets zijn niet de aparte afzonderlijke, maar juist de onderling verschillende signalen van belang. De intensiteit van de signalen is gemeten en de expressie profielen van de indicatorgenen zijn ten opzichte van de referentiegenen bepaald. Hierbij is genormaliseerd voor experimentele variatie van de

referentiegenen in de B1 en B4 stadia. De resultaten van deze eerste kwantificatie is te zien in figuur 12.



Figuur 12 Genormaliseerde indicatorgen expressie. De verschillen in expressie van de indicatorgenen is berekend en genormaliseerd aan de hand van de referentiegenen. De verschillen in expressie zijn weergegeven als 2log waarden. Drie hoeveelheden cDNA input in de OLA reactie zijn bekeken.

Figuur 12 laat zien dat voor de indicatorgenen P450 en glycosyl hydrolase een duidelijke up-regulatie wordt gedetecteerd terwijl MAL d1 een down-regulatie heeft. Het profiel van P450 en glycosyl hydrolase is in overeenstemming met de microarray data. De MAL d1 expressie is echter niet vergelijkbaar (omgekeerd) met de microarray data, en zou verklaard kunnen worden doordat een andere dan beoogd transcript (homoloog met een andere expressie) wordt gedetecteerd of dat het mRNA wordt veranderd (bijvoorbeeld door alternatieve splicing) waardoor de target sequentie voor de OLA probe bij hoge expressie en turnover niet meer aanwezig is. Al met al laten deze experimenten samen met de eerdere experimenten met het RIN tomaat gen zien dat het Quality Quantifier concept in staat is relatieve expressieverschillen aan te tonen.

Conclusies uit deze reeks experimenten met roos zijn:

1. de QQ toets kan relatieve expressieverschillen aantonen.
2. het concept van het gebruik van referentiegenen lijkt te werken, er wordt voor deze genen een constante expressie gemeten.
3. een genormaliseerd expressieverschil (2log) voor de indicatorgenen kan worden bepaald t.o.v de referentiegenen.
4. twee kandidaat indicatoren worden met de QQ toets bevestigd, MAL d1 geeft echter een afwijkend tegengesteld profiel.
5. de toets is niet geschikt om absolute genexpressie niveaus te detecteren, de signalen op zich correleren niet met de microarray expressedata.
6. de gevonden expressieverschillen moeten nu met een onafhankelijk platform worden bevestigd.

3.2 Knelpunten en oplossingen

Tot op heden hebben zich nog geen knelpunten voor gedaan. Ook voor het komende halfjaar verwachten we geen onoverkomelijke hindernissen op de weg te vinden. Het niet correleren van Real Time PCR met genexpressie kan vele redenen hebben, zoals betrouwbaarheid van de sequentie, aanwezigheid van genen in het genoom met vergelijkbare sequenties maar andere expressie profielen, processing van RNA moleculen etc. Deze mogelijke hindernissen om van microarray gebaseerde kandidaats genen te komen tot Real Time PCR of multiplex toetsen zullen systematisch geëlimineerd worden om tot betrouwbare merkers te komen.

3.3 Octrooiaanvragen

Was nog niet van toepassing maar als kandidaats indicatorgenen gevalideerd zijn en goede betrouwbare voorspellingsindicatoren blijken te zijn zal dat zeker overwogen worden.

3.4 Interne rapportages

Interne verslagen, notulen en presentaties zijn opvraagbaar bij de projectleider J. Mes (A&F).

Notulen:

- OKEE vergadering met begeleidingscommissie, 14 september 2004

Documenten:

- Eerste halfjaarverslag EET, 2004
- Eerste halfjaarverslag PT, 2004

Presentaties:

- Bijeenkomst met begeleidingscommissie, 14 september 2004:

- OKEE - Voortgang Tomaat
- OKEE - Voortgang Roos
- OKEE - Voortgang multiplex toetsmethode

3.5 Openbare publicaties

Jurriaan J. Mes, Renata M. Ariens, H. Martijntje Vollebregt, Eric P. Boer, Monique F. van Wordragen. 2004. Op weg naar een toets voor Botrytis aantasting in roos. KNPV werkgroep Botrytis gehouden op 18 mei, te Lisse. Abstract in Gewasbescherming jaargang 35(6): 16-17.

Jurriaan J. Mes, Renata M. Ariens, H. Elodie Brans, Eric P. Boer, Monique F. van Wordragen. 2004. Microarray analysis related to Botrytis infection in rose. Abstract in Second EPSO conference, 10-14 October 2004.

4 Conclusies

Op het moment dat dit project 'Optimalisatie van Keten Efficiëntie door Expressieprofieling' het laatste jaar ingaat lijken alle gestelde doelen voor dit project nog haalbaar. De planning was dat op dit moment een eerste selectie van indicatoren gemaakt kon worden die aanwijsbaar een voorspelling kunnen geven over de kwaliteit van het product. In beide onderzoeksmodellen roos en tomaat bleek het maken van deze selectie mogelijk, echter, we moeten wel rekening houden met het aantal gebruikte samples en analyse voor deze selecties. Het is zeer goed denkbaar dat er meerdere oorzaken van kwaliteitverlies zijn waarbij sommige nu niet geanalyseerd zijn. In de praktijk kan het er dus op neerkomen dat we met de nu geselecteerde genen niet alle partijen even goed kunnen voorspellen. Daarom zal een breed onderzoek nodig zijn waarbij de genexpressie van deze kandidaatgenen gecorreleerd wordt aan het kwaliteitsverloop van deze monsters. Alleen op die manier is de betrouwbaarheid en toepassing van deze techniek te valideren. De ontwikkelde toetsmethode lijkt daarvoor op het juiste moment gebruiksklaar. Als na enige validatiestudies blijkt dat de QQtest betrouwbaar genexpressie kan bepalen zoals de eerste proeven hebben uitgewezen dan is de weg vrij om deze methode direct in te zetten om de gevonden kandidaatgenen op een brede set van partijen te toetsen op betrouwbaarheid en inzetbaarheid. Koppelingen tussen genexpressie, kwaliteitsdata en -verloopmodellen zullen verder moeten uitwijzen hoe deze zich tot elkaar verhouden en hoe uit genexpressie de beste conclusies zijn te trekken in relatie tot de voorspelkwaliteit. Dit zal een belangrijke opdracht worden voor het laatste jaar van dit project.