Project title:	Carrot: An early warning system for risk of cavity spot in crops
Project number:	FV 448
Project leader:	Dr Nathalie Verhoef, NSure
Report:	Final report, November, 2016
Key staff:	Nathalie Verhoef (NSure, technical project leader) Theo Aanhane (NSure, general supervision) Chris Roelofsen (NSure, technician) Frank Hoeberichts (NSure, statistician) Lotte Terlouw-Langelaar (NSure, technician) Peter Balk (NSure, technical project supervision)
Location of project:	Lincolnshire, Yorkshire, Nottinghamshire and Wageningen (NL)
Industry Representative:	Ian Holmes, Strawson Limited, Featherstone House Farm, Bilsthorpe, Newark, Nottinghamshire, NG22 8RD James Bramley, M.H. Poskitt Ltd., Weeland Road, Kellington, Goole, N-Yorkshire, DN14 0FL
Date project commenced:	1 December 2015
Expected completion date:	1 December 2016

© Agriculture and Horticulture Development Board 2016. All rights reserved

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

© Agriculture and Horticulture Development Board 2016. No part of this publication may be reproduced in any material form (including by photocopy or storage in any medium by electronic mean) or any copy or adaptation stored, published or distributed (by physical, electronic or other means) without prior permission in writing of the Agriculture and Horticulture Development Board, other than by reproduction in an unmodified form for the sole purpose of use as an information resource when the Agriculture and Horticulture Development Board or AHDB Horticulture is clearly acknowledged as the source, or in accordance with the provisions of the Copyright, Designs and Patents Act 1988. All rights reserved.

All other trademarks, logos and brand names contained in this publication are the trademarks of their respective holders. No rights are granted without the prior written permission of the relevant owners.

The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.

© Agriculture and Horticulture Development Board 2016. All rights reserved

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

Dr. Nathalie Verhoef

Researchers's Representative

NSure

Signature ..... Date .....

Theo Aanhane

CEO

NSure

Signature ..... Date .....

Report authorised by:

[Name]

[Position]

[Organisation]

Signature ..... Date .....

[Name]

[Position]

[Organisation]

Signature ..... Date .....

© Agriculture and Horticulture Development Board 2016. All rights reserved

CONTENTS

© Agriculture and Horticulture Development Board 2016. All rights reserved

GROWER SUMMARY

Headline

A set of potential indicator genes for cavity spot has been identified. The aim is to prevent losses caused by cavity spot, by using a practical molecular test.

Background

Cavity spot is a major disease of carrots in the UK and is mainly caused by *Pythium violae*. Cavity spot reduces harvest quality. Visibly infected carrots are not acceptable for the fresh produce market or processing.

An early indication for cavity spot would be of great value, as it can be used as a decision support system. The test has to assess risk of cavity spot at two cost adding moments: before distribution of straw and before fields are covered. Selecting low risk fields will reduce losses and lead to reduced costs for labour and straw.

The aim of this project is to identify cavity spot specific indicator genes from the most prominent carrot cultivar in the UK 'Nairobi'. These genes will be used to develop a practical test that quantifies the expression of those genes, to determine the risk of cavity spot at an early stage. An early indication for cavity spot would be of great value, because it can be used as a decision support system. To find such indicator genes, NSure uses RNA sequencing (RNA-Seq). By using RNA-Seq, NSure can examine the expression of all genes present in any crop of interest and select those genes in which the expression is linked to a particular trait (Stattin et al., 2012; Kromwijk et.al, 2013).

The time projected to develop such a test requires two seasons of research (FV 448 and FV 448a). In the first season, samples were collected at different calendar dates from various fields in the area of Yorkshire and Nottinghamshire. In parallel, the fields were visually evaluated on the occurrence of cavity spot. Gene expression profiles on a selection of the collected samples were obtained using RNA-Seq and subsequently validated. A similar trial will be conducted in a second season to validate the functionality of the indicators selected in the first season. After defining the final set of indicators, NSure will define the decision criteria that determine whether a certain field shows a certain risk at developing cavity spot.

The project will result in a practical test to determine the risk on cavity spot. The test will consist of a simple sampling kit that enables the grower to collect a sample. The samples have to be sent to a lab either in the UK or to NSure's facilities in Wageningen. Within 48 hours after the samples have arrived, the grower will receive an indication of the risk of cavity

© Agriculture and Horticulture Development Board 2016. All rights reserved

spot. Samples can be taken throughout the season in order to support decision taking at several crucial time points during the season.

Summary

Collection of carrot samples with a varying occurrence of cavity spot

In collaboration with Strawsons Ltd, M H Poskitt Ltd and Elsoms Seeds, NSure received carrot samples from fields in Nottinghamshire and Yorkshire, in September and October 2015. All fields were evaluated for the incidence of cavity spot. The incidence of cavity spot ranged from 0% to 65%, the range required to start the search for cavity spot related genes, as both high and low risk samples were needed.

Processing of the RNA-Seq data

Based on the quality evaluation results, NSure selected two fields that showed a low occurrence of cavity spot and two fields with a high incidence (Table 1). Frozen samples collected from those fields in September and October were studied in detail by RNA-Seq. By using RNA-Seq, NSure is able to examine the expression of all genes present in a sample. RNA-Seq was carried without any technical problems and reliable gene expression values were obtained.

Table 1. Sample selection for RNA-Seq and field evaluation of the selected fields

Field	**Quality evaluation 1**		**Quality evaluation 2**	
	%	**Date**	**%**	**Date**
1	2	17-11-2015	7	1-2-2016
2	3	20-11-2015	Harvested	-
3	47	18-11-2015	Harvested	-
4	64	18-11-2015	Harvested	-

The identification and validation of potential cavity spot indicators

To get an impression about the global differences between the samples that were sequenced, a Principal Component Analysis (PCA) was performed. PCA is a statistical method in which a set of data points are re-expressed in terms of basic components that explain the most variance in the data. As a consequence, PCA allows you to separate different samples in a graph according to the global differences between the samples. Samples that are similar in overall gene expression will cluster together, whereas samples that differ appear separated in the PCA plot. The results from the PCA are displayed in figure 1. Looking at the second component, which is represented on the y-axis, a segregation of the samples with high and low incidence on cavity spot can be observed. This suggest the presence of cavity spot

© Agriculture and Horticulture Development Board 2016. All rights reserved

related genes which are differentially expressed in samples with a low or high incidence on cavity spot.

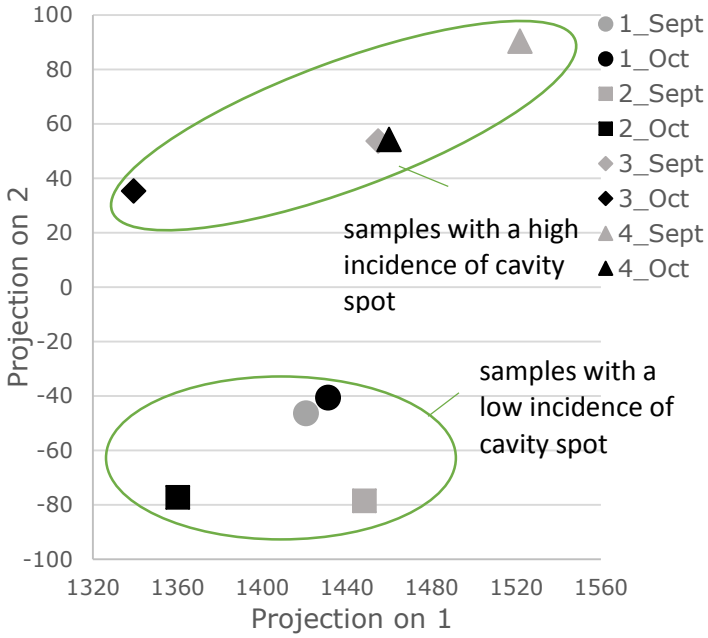


Figure 1. PCA plot using the observed expression values of all genes active in a certain carrot sample. The sequenced samples are indicated in different colours and shapes to indicate the field and timepoint. In total 2 samples per field (1-4) were sequenced, one sample collected in September and one collected in October. During the quality assessments performed at a later stage, fields 1 and 2 were characterized as fields with a low incidence on cavity spot, while fields 3 and 4 showed a high incidence. A separation of the low and high risk samples can be observed when looking at the second component which is projected on the y-axis. On the x-axis, three fields (2, 3 and 4) show a global difference in expression in time, meaning that other biological processes, which are specific for the carrots growing in that particular field, were triggered in time.

To obtain the best suitable cavity spot indicators, a stringent selection was performed based on pairwise comparisons between low and high risk samples. This resulted in approximately 70 candidate genes. As expected several of those genes are known to be involved in defence related processes. In figure 2, gene expression profiles are shown of 4 genes which are differentially regulated in samples collected from fields (3 and 4) characterized with a high incidence on cavity spot in comparison to fields (1 and 2) that showed a low incidence on cavity spot. Genes *30777* and *45235* are higher expressed in the high risk samples, while genes *6895* and *48839* are lower expressed. Further validation in samples collected from other fields led to a selection of potential carrot cavity spot indicator genes worthwhile to validate in a second season.

© Agriculture and Horticulture Development Board 2016. All rights reserved

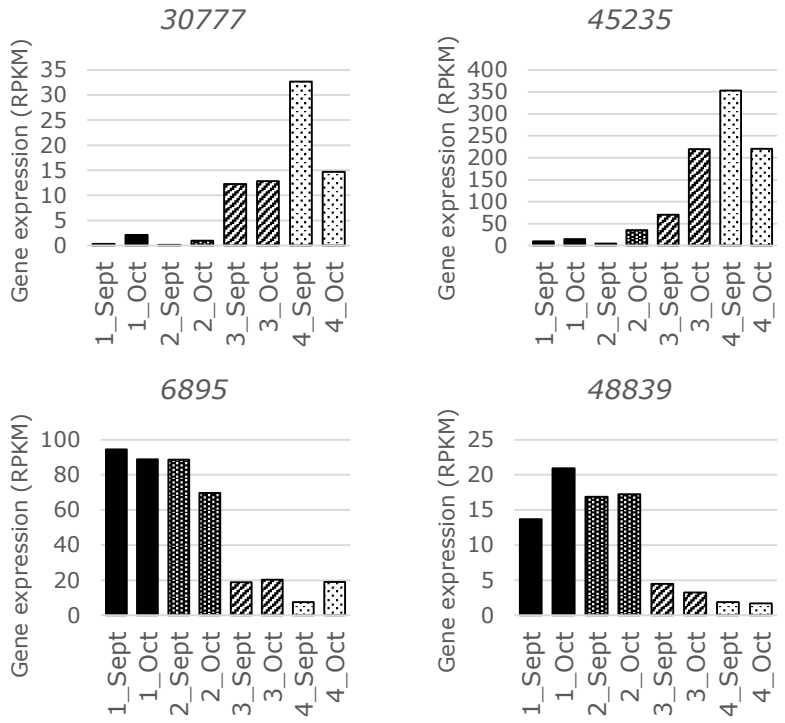


Figure 2. Gene expression profiles of 4 potential cavity spot related indicators. The RPKM value was used to indicate the level of gene expression, which is the standard quantification of gene expression when studying RNA-Seq data. Genes 30777 and 45235 are higher expressed in the high risk samples (3 & 4). The expression of genes *6895* and *48839* is down regulated in the high risk samples.

Next to frozen samples, which were used for the RNA-Seq study and indicator selection, juice samples were collected using the NSure sampling kit which is used for commercial tests (Figure 3). It is essential that similar results are obtained when analysing a frozen sample or a sample using the NSure sampling kit. For nearly all measured genes, similar results were obtained, showing that this sampling method can be used, once the putative indicator genes have been validated.

Financial Benefits

Carrot is one of the major crops in the UK. The total cultivated area exceeds 9,000 ha. 60% of the acreage, approx. 5,500 ha, is stored under straw. One hectare results on average in a gross income of £8,000. This means that the total turnover of covered carrots is approx. £44 million.

Losses due to cavity spot vary between years and geographical regions. Until recently, Scotland for example, had no occurrence of cavity spot. Other regions have more severe problems. In some fields the damage exceeds 40%. On average, cavity spot destroys 3% - 7% of the annual yield, resulting in a loss of between £1.25 and £3 million. However, this

© Agriculture and Horticulture Development Board 2016. All rights reserved

percentage seems to have increased over the years. Last season for example, the percentage was estimated to be between 5% and 10%, which almost doubled the losses.

The average cost of covering consist of straw (£3,000 per ha) and logistics (transport and covering). In total the costs for a carrot crop with straw covering are approx. £4,000 per ha.

It is clear that a high cavity spot occurrence means that a grower will not earn (instead: will lose) money on those batches. A predictive test that determines high risk fields, will support a grower to pick only low risk fields for covering. A reduction of cavity spot occurrence of only 1% on average, already makes a difference in turnover of £440,000 per year. This is exclusive of extra costs for covering, logistics and sorting. A predictive test would make UK carrot industry much more profitable.

Figure 3. Sampling procedure using the NSure sampling kit. 25 carrots are washed and from each carrot a vertical slice of 10 cm is collected with a potato peeler. The slices are transferred into the juice centrifuge and the extraction fluid is added. The juice is transferred in one half of the extraction bag which contains a sieve. The pipette is used to suck up some juice from the other half of the bag. Two droplets of juice are applied to the sampling card and the cards need to be dried for minimum of 2 hours.

Action Points

None for growers at this stage.

During this first season (FV 448), a set of potential indicator genes for cavity spot has been identified. In order to develop a reliable test, these genes need to be validated in a second season (FV 448a). Part of the work to be done involves the final description of the sampling protocol. This protocol will explain activities to be conducted by growers.

© Agriculture and Horticulture Development Board 2016. All rights reserved

SCIENCE SECTION

Introduction

Cavity spot is an important limiting factor in the carrot production worldwide. Multiple *Pythium* species, that differ from region to region, are involved in cavity spot. Cavity spot in the UK is mainly caused by *Pythium violae*. Cavity spot is characterised by small sunken elliptical lesions that appear on the tap root, while the aboveground plant parts do not show any visible symptoms. Currently, growers try to minimise the risk on this disease by using cultural practices such as avoiding fields with a history of cavity spot, growing carrots on raised plant beds and the use of fungicides.

Development of diagnostic methods enabling early detection of cavity spot would be of great importance in limiting the economic losses. ELISA (enzyme-linked immunosorbent assay) as well as PCR (polymerase chain reaction)-based tests have been developed to detect *Pythium* species in the soil or in carrot tissue (White et al., 1996; Klemsdal et al., 2008; Barbara D.J. et al., 2010). However, a positive result on *Pythium violae* in the soil or on carrots is no guarantee that cavity spot will occur as the health status of the carrot plays an important role. To predict whether a certain field of carrots will develop cavity spot, it is necessary to look at the crop status.

In 2013, NSure, launched a molecular test that can predict whether a batch of harvested carrots will develop black spots during cold storage (www.nsure.eu). This so called “black spot disease” can be caused by four different fungi: *Mycocentrospora acerina*, *Alternaria radicina*, *Chalaropsis thielavioides* and *Rhexocercosporidium carotae*. In addition, the test also predicts the risk of symptom development of *Phytophthora spp*. Months before symptoms become visible, NSure measures the expression of multiple disease related carrot genes that are altered upon an early infection with one of the black spot fungi. To find such indicator genes, NSure uses RNA sequencing (RNA-Seq). By using RNA-Seq, NSure can examine the expression of all genes present in any crop of interest and select those genes in which the expression is linked to a particular trait (Stattin et al., 2012; Kromwijk et.al., 2013).

The aim of this project is to identify cavity spot specific indicator genes from the most prominent carrot cultivar in the UK ‘Nairobi’. These genes will be used to develop a practical test that quantifies the expression of those genes to determine the risk of cavity spot at an early stage. An early indication for cavity spot would be of great value, as it can be used as a decision support system. The test has to assess risk of cavity spot at two cost adding moments: before distribution of straw and before fields are covered. Selecting low risk fields will reduce losses and leads to less costs for labour and straw.

© Agriculture and Horticulture Development Board 2016. All rights reserved

It is projected that it will take two seasons of research to develop the test. In the first season, samples will be collected at different calendar dates from various fields in the area of Yorkshire and Nottinghamshire. In parallel, the fields will be visually evaluated on the occurrence of cavity spot. Gene expression profiles of a selection of the collected samples will be obtained using RNA-Seq. Carrot genes that can be recognised as being correlated with the risk at occurrence of cavity spot will be selected and analysed by quantitative RT PCR (qPCR). A similar trial will be conducted in the second year to validate the functionality of the indicators selected in the first season. After defining the final set of indicators, NSure will define the decision criteria that determine whether a certain field shows a certain risk of developing cavity spot. Furthermore, a user-friendly sampling protocol will be developed for future customers.

The objectives and milestones to be achieved for the first year are depicted in table 1.

Table 1. Objectives and milestones first year

Objective 1. Collection of carrot samples from different fields in two different areas (Yorkshire and Nottinghamshire) from the largest commercial variety, Nairobi.	
Date	**Milestone**
01/11/2015	1.1 First collection of carrot samples (Nairobi) gathered from different sites.
01/05/2016	1.2 Quality evaluation of the sampled plots described for milestone 1.1.
Objective 2. Selection of a longlist of carrot genes that may determine the risk of development of cavity spot in the variety Nairobi.	
Date	**Milestone**
01/09/2016	2.1 A longlist of candidate genes identified by RNA-Seq.
Objective 3. Validation of carrot indicator genes in Nairobi.	
Date	**Milestone**
01/10/2016	3.1 qPCR validation on the carrot collection described for milestone 1.1.
Objective 4. Continuous and active two-way communication between growers and researchers.	
Date	**Milestone**
Continuous	4.1 Share results with growers and other researchers

© Agriculture and Horticulture Development Board 2016. All rights reserved

Materials and methods

NSure commissioned Strawson Limited as well as Poskitt Ltd to collect carrot samples from 15 different fields in September and October 2015. Per field, two replicate samples (A&B) of 25 carrots each had to be collected randomly from the field. The collected samples were transported to Elsoms Seeds where NSure performed the sampling.

After washing, a vertical slice was collected from each carrot using a potato peeler. The 25 slices were collected in a plastic bag, frozen in dry ice and stored in a -80 °C freezer for molecular analysis. Alongside frozen sampling, NSure took 40 samples by a different method using the NSure sampling kit (Figure 1). By this method, juice is extracted which is applied onto a special sampling card that fixates the genetic material.

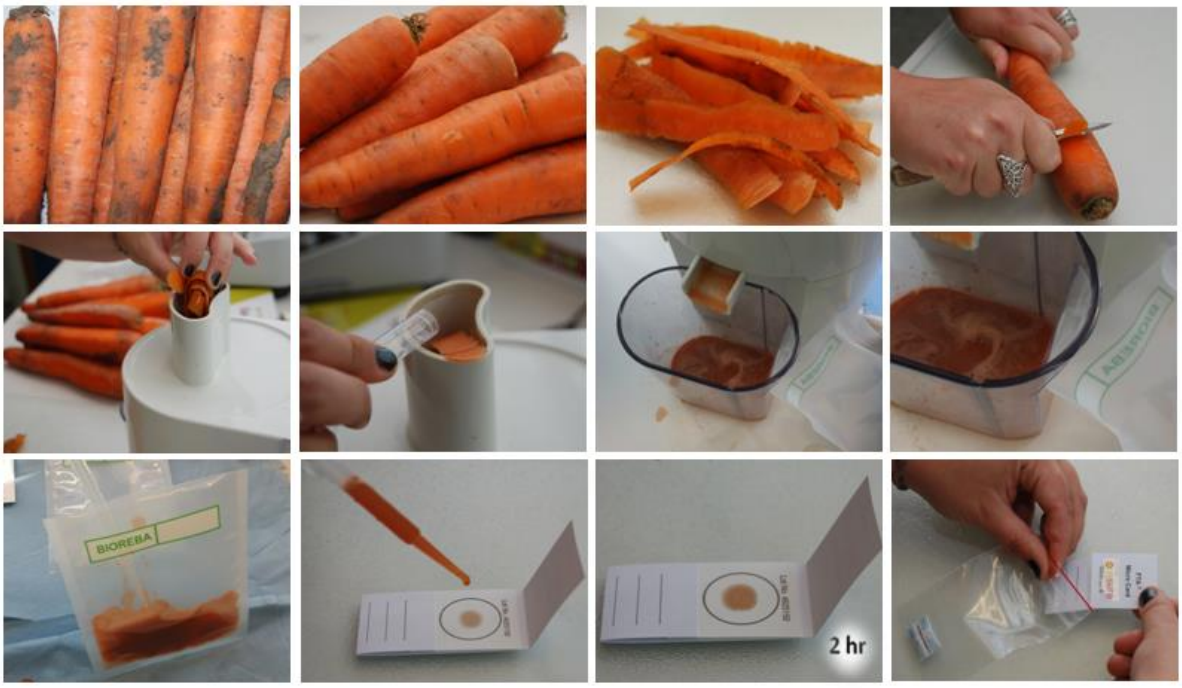


Figure 1. Sampling procedure using the NSure sampling kit. Twenty-five carrots were washed and from each carrot a vertical slice of 10 cm was collected with a potato peeler. The slices were transferred to the juice centrifuge and the extraction fluid was added. The juice was transferred in one half of the extraction bag which contains a sieve. The pipette was used to suck up some juice from the other half of the bag. Two droplets of juice were applied to the sampling card. The cards were dried for at least two hours.

© Agriculture and Horticulture Development Board 2016. All rights reserved

Results

Collection of carrot samples with a varying incidence of cavity spot

In collaboration with Strawson Limited, Poskitt Ltd and Elsoms Seeds, NSure received samples from the main carrot cultivar in the UK 'Nairobi' in September and October from various fields in the UK (Table 2 & 3). In Nottinghamshire, the sampling was conducted according to plan, but this did not apply for the sample collection in Yorkshire. Instead of collecting two replicate samples per field in September as well as in October, two samples were collected but one sample set from only five fields, instead of a total of four from all fields. From each sample received, NSure took a frozen sample. The frozen samples were used for RNA-Seq and validation of potential cavity spot indicators as a lot of genetic material is needed. In case of the samples collected from ten fields located in Nottinghamshire, an additional sampling method was performed by using the NSure sampling kit which is used for commercial tests. Only a limited amount of genetic material can be obtained from a sampling card which is sufficient once a test is developed, but not for research purposes. Although the cards cannot be used to identify cavity spot indicators, it is essential to check whether potential indicators show a similar gene expression pattern when using both sampling methods.

Table 2. Sample collection and field evaluation in Nottinghamshire

Field	Frozen sampling		Sampling card		Quality evaluation 1		Quality evaluation 2	
	Sept 2015	Oct 2015	Sept 2015	Oct 2015	%	Date	%	Date
1	A&B	A&B	A&B	A&B	2	17-11-2015	7	1-2-2016
2	A&B	A&B	A&B	A&B	3	20-11-2015	Harvested	-
3	A&B	A&B	A&B	A&B	47	18-11-2015	Harvested	-
4	A&B	A&B	A&B	A&B	64	18-11-2015	Harvested	-
5	A&B	A&B	-	-	2	20-11-2015	Harvested	-
6	A&B	A&B	-	-	13	17-11-2015	Harvested	-
7	A&B	A&B	-	-	3	18-11-2015	11	20-11-2016
8	A&B	A&B	-	-	10	17-11-2015	Harvested	-
9	A&B	A&B	-	-	14	14-10-2015	Harvested	-
10	A&B	A&B	A&B	A&B	2	19-11-2015	0	1-2-2016
11	A&B	A&B	A&B	A&B	2	17-11-2015	8	1-2-2016
12	A&B	A&B	A&B	A&B	0	20-11-2015	Harvested	-
13	A&B	A&B	A&B	A&B	5	20-11-2015	Harvested	-
14	A&B	A&B	A&B	A&B	3	17-11-2015	Harvested	-
15	A&B	A&B	A&B	A&B	46	18-11-2015	45	20-1-2016

© Agriculture and Horticulture Development Board 2016. All rights reserved

Strawson Limited and Poskitt Ltd evaluated the fields on the incidence of cavity spot by assessing 100 carrots randomly collected from the field. The incidence of cavity spot ranged from 0% to 65% (Table 2 & 3). Most of the fields in Nottinghamshire were already harvested when the second evaluation took place. This was due to difficult weather conditions. Where the conditions allowed, fields were harvested resulting in some of the fields which were planned to be harvested later moving forward. Strawson Limited also checked the fields on the occurrence of other diseases and those observations are listed in Appendix 1. Scab was observed in many fields, but also violet root rot, crown rot, carrot fly damages and necrotic lesions due to viruses. Poskitt Ltd did not record any other diseases.

Table 3. Sample collection and field evaluation in Yorkshire

Field	Frozen sampling		Sampling card		Quality evaluation 1		Quality evaluation 2	
	Sept 2015	Oct 2015	Sept 2015	Oct 2015	%	Date	%	Date
16	A&B	-	-	-	8	1-12-2015	37	28-1-2016
17	A	-	-	-	0	1-12-2015	0	28-1-2016
18	B	-	-	-	14	1-12-2015	20	28-1-2016
19	A&B	-	-	-	45	1-12-2015	Harvested	-
20	A&B	A	-	-	22	1-12-2015	Harvested	-
21	-	A	-	-	18	1-12-2015	22	-
22	-	A	-	-	0	1-12-2015	-	-
23	-	A	-	-	0	1-12-2015	-	-
24	-	A	-	-	0	1-12-2015	-	-
25	-	A	-	-	0	1-12-2015	5	28-1-2016
26	-	A	-	-	0	1-12-2015	0	28-1-2016
27	-	A	-	-	0	1-12-2015	0	28-1-2016
28	-	A	-	-	0	1-12-2015	0	28-1-2016
29	-	A	-	-	0	1-12-2015	0	28-1-2016
30	-	A	-	-	50	1-12-2015	5	28-1-2016
31	-	A	-	-	0	1-12-2015	10	28-1-2016
32	-	A	-	-	0	1-12-2015	5	28-1-2016
33	-	A	-	-	0	1-12-2015	0	28-1-2016
34	-	A&B	-	-	50	1-12-2015	Harvested	-

Sample selection RNA-Seq

By using RNA-Seq, NSure is able to examine the expression of all genes present in a certain sample. By comparing the complete set of active genes in low and high risk samples, cavity spot specific genes from carrot can be selected.

Based on the quality evaluation results on the incidence of cavity spot and other diseases, NSure selected two fields that showed a low incidence of cavity spot and two fields with a

© Agriculture and Horticulture Development Board 2016. All rights reserved

high incidence (Table 4). To identify cavity spot genes that showed a constitutive difference in gene expression between low and high risk samples, it was necessary to sequence the samples collected in September and October. At each time point two replicate samples (A&B) were collected. For RNA-Seq, the genetic material (RNA) isolated from replicate A and B were equally mixed into one sample. In total eight samples were send to Baseclear (The Netherlands) for single-read Illumina RNA sequencing.

Table 4. Samples selected for RNA sequencing

Field	County	Sampled in	Risk	RNA-Seq code
1	Nottinghamshire	Sept. 2015	Low	1_09
1	Nottinghamshire	Oct. 2015	Low	1_10
2	Nottinghamshire	Sept. 2015	Low	2_09
2	Nottinghamshire	Oct. 2015	Low	2_10
3	Nottinghamshire	Sept. 2015	High	3_09
3	Nottinghamshire	Oct. 2015	High	3_10
4	Nottinghamshire	Sept. 2015	High	4_09
4	Nottinghamshire	Oct. 2015	High	4_10

Processing of the RNA-Seq data

A schematic representation of RNA-Seq and subsequent data processing is depicted in figure 2. After isolation of the RNA (which contains the active genes), RNA is sequenced leading to the production of millions of short sequence reads that correspond to individual fragments. The amount of raw sequencing reads obtained per sample after sequencing is shown in table 5. Trimming, a procedure aimed at removing low quality or ambiguous reads, resulted in the removal of only a limited percentage of reads (<1%).

Table 5. Trimming and mapping of the reads

RNA-Seq code	No. of raw reads (million)	No. of reads after trimming (million)	% remaining after trimming	No. of reads mapped (million)	% mapped
1_09	25.3	25.2	99.6	18.6	73.9
1_10	27.2	27.0	99.6	19.7	72.7
2_09	30.0	29.9	99.6	21.7	72.6
2_10	21.8	21.7	99.8	15.7	72.6
3_09	27.8	27.7	99.6	20.3	73.3
3_10	19.1	19.0	99.7	14.0	73.4
4_09	35.9	35.7	99.6	26.1	73.1
4_10	29.8	29.8	99.8	21.8	73.1

© Agriculture and Horticulture Development Board 2016. All rights reserved

In all samples, >18 million high quality reads were available for downstream analysis which is more than sufficient. After trimming, the sequencing reads were aligned to a reference. This reference could be for example a publicly available genome or a self-created library of genes which is called a *de novo* assembly. In a former project, NSure already developed a *de novo* assembly for carrot based sequences of the cultivar Nerac. This *de novo* assembly consists out of 49.379 genes (or gene parts). Through mapping of the sequencing reads to this reference it is possible to quantify the level of gene expression of all genes present in a certain sample. A high percentage (>72%) of the obtained sequencing reads could be mapped to this reference (Table 5). This justifies the use of the *de novo* assembly for the cultivar Nairobi and implies that reliable gene expression values were obtained.

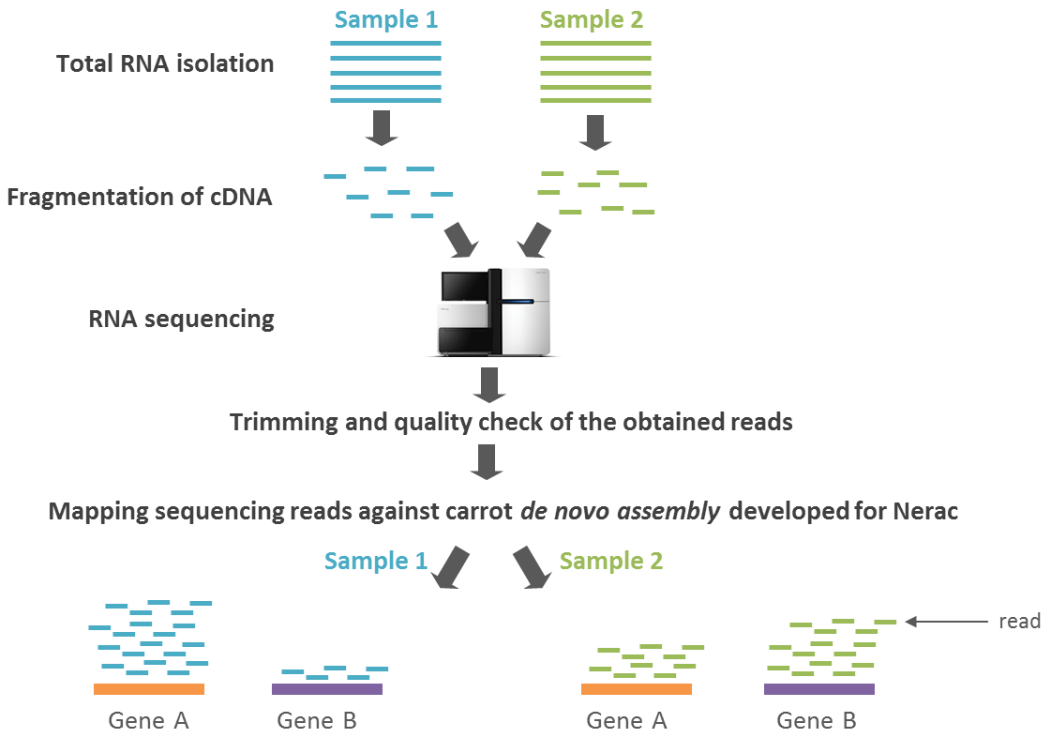


Figure 2. Schematic overview of RNA-Seq and subsequent data processing. RNA is isolated and converted into cDNA and fragmented. Prior to sequencing, the fragments are attached to adapters which are required to read the sequence. This collection of fragments is then sequenced which leads to the production of millions of short sequence reads. Next, trimming is performed, a procedure aimed at removing low quality or ambiguous reads. The high quality reads were aligned to a self-created library of genes (a *de novo* assembly) based on carrot sequences of the cultivar Nerac. Through mapping of the sequencing reads to this reference it is possible to quantify the level of gene expression of all genes present in a certain sample. Gene A is higher expressed in sample 1 as more reads map to reference gene A. Gene B is higher expressed in sample 2.

© Agriculture and Horticulture Development Board 2016. All rights reserved

The identification of potential cavity spot indicators by RNA-Seq

To get an idea about the global differences between the samples that were sequenced, a Principal Component Analysis (PCA) was performed. PCA is a statistical method in which a set of data points are re-expressed in terms of basic components that explain the most variance in the data. As a consequence, PCA allows you to separate different samples in a graph according to the global differences (variation) between the samples. Samples that are similar in overall gene expression (i.e. not showing much variation when looking at all expression values) will cluster together, whereas samples that differ appear separated in the PCA plot. The results from the PCA are displayed in figure 3. Looking at the first component, which is depicted on the x-axis (projection 1), it could be observed that the samples collected from fields 2, 3 and 4 moved in the same way (right to left) when looking at the position of the samples collected in September and October. The samples collected in September and October at field 1 do not show much variation as they cluster. When looking at the first component we cannot observe a segregation of the low and high risk samples. Interestingly, when looking at the second component which is represented on y-axis a separation of high and low risk samples can be observed, suggesting the presence of cavity spot related genes which are differentially expressed in low and high risk samples.

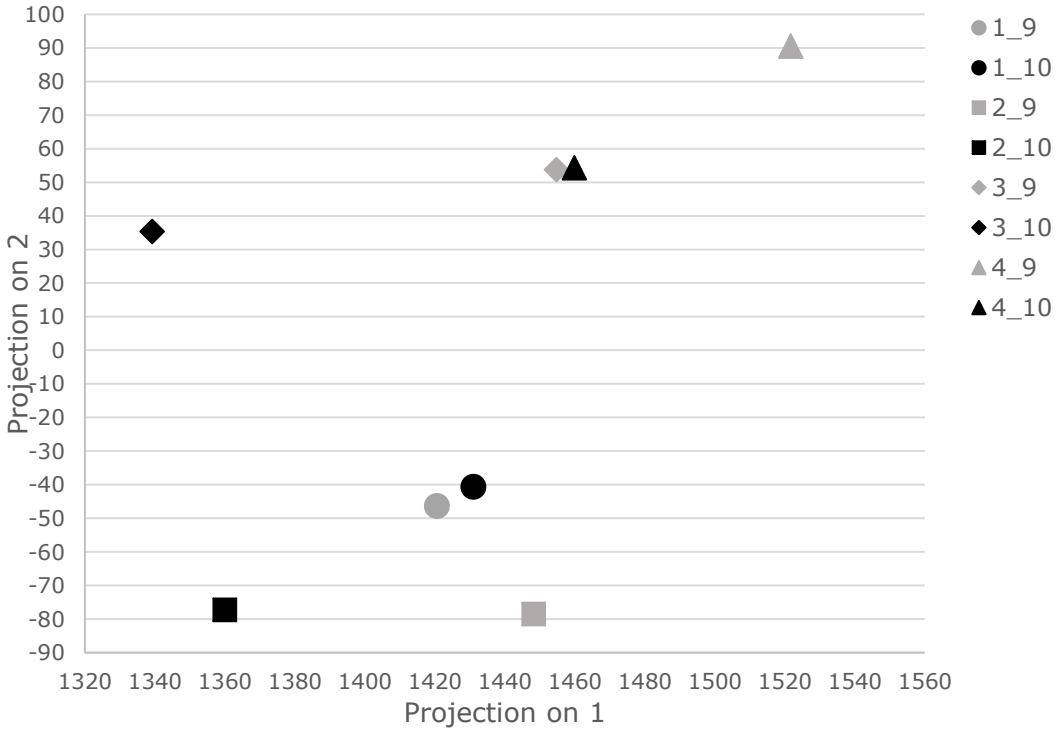


Figure 3. PCA plot using the observed log2 expression values of all 49.379 genes present in the *de novo* assembly. The sequenced samples are indicated in different colours and shapes. A separation of the low and high risk samples can be observed when looking at the second component which is projected on the y-axis.

© Agriculture and Horticulture Development Board 2016. All rights reserved

Besides global analysis of the data using PCA, pairwise comparisons were performed between the low and high risk samples to identify individual genes that showed differential expression upon infection with *Pythium violae*. The standard quantification of gene expression when studying RNA-Seq data is the RPKM (reads per kilobase per million) value. The higher the RPKM value, the higher the gene expression. A fold change ratio (FC) for every gene can be calculated by dividing the RPKM value of the high risk sample by the RPKM value of a low risk sample and vice versa (Figure 4). The FC is used to select differentially expressed genes (DEGs). This means that, for example, a gene is considered as being differentially expressed when its FC is bigger than 2 (FC>2). FC thresholds are arbitrary and can be adjusted, resulting in more or fewer differentially expressed genes (DEGs). As no replicates were sequenced, a more stringent selection was performed to identify the most promising indicator genes that show a consistent difference in expression over time. In total 16 pairwise comparisons were made between the 4 low and 4 high risk samples. For further analysis, a gene was considered to be differentially expressed, when the FC was 2,5 or higher in all comparisons. The number of DEGs which were up- or downregulated in response to an infection with *Pythium violae* with a FC>2,5 are depicted in table 6. As expected several of those genes are known to be involved in defence related processes.

Figure 4. Calculation of the fold change ratio. A fold change ratio for every gene can be calculated by dividing the RPKM value of a high risk sample by the RPKM value of a low risk sample. For example, 20 reads in the high risk sample mapped to gene A. In case of the low risk sample only 5 reads mapped to gene A. The fold change is therefore 4 (20/5). Gene A is 4 times higher expressed in the high risk sample. By calculating the fold change between low and high risk samples cavity spot indicator genes can be selected.

Table 6. Number of up- and downregulated genes at a FC>2,5 in high risk samples

Genes upregulated	Genes downregulated
37	30

© Agriculture and Horticulture Development Board 2016. All rights reserved

In figure 5, the gene expression profiles are shown of nine genes which are higher expressed in the high risk samples in comparison to the low risk samples. In figure 6, the gene expression profiles are shown of nine genes which are lower expressed in the high risk samples.

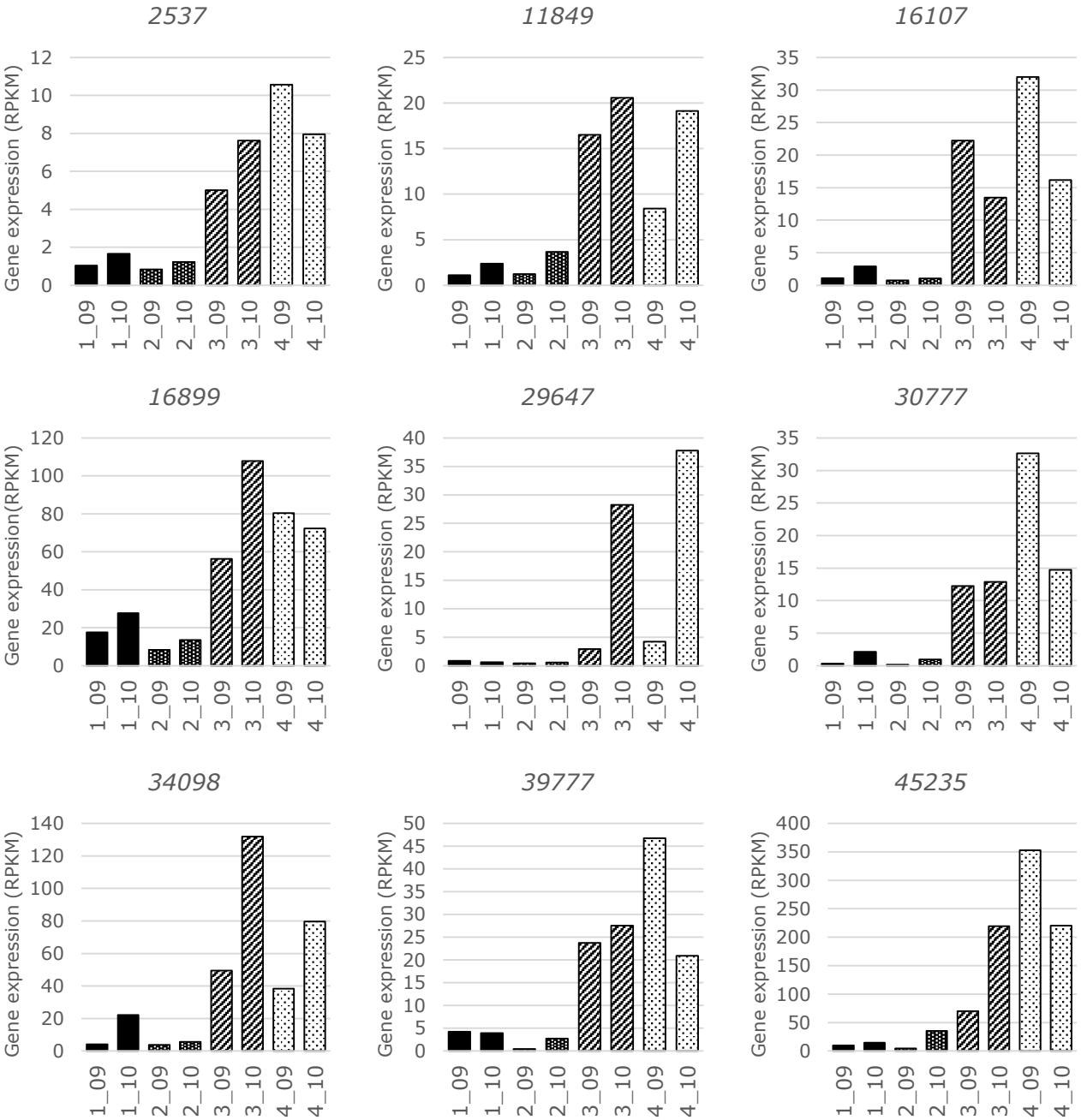


Figure 5. Gene expression profiles of 9 genes which are higher expressed in the high risk samples (3 & 4). The RPKM value was used to indicate the level of gene expression, which is the standard quantification of gene expression when studying RNA-Seq data.

© Agriculture and Horticulture Development Board 2016. All rights reserved

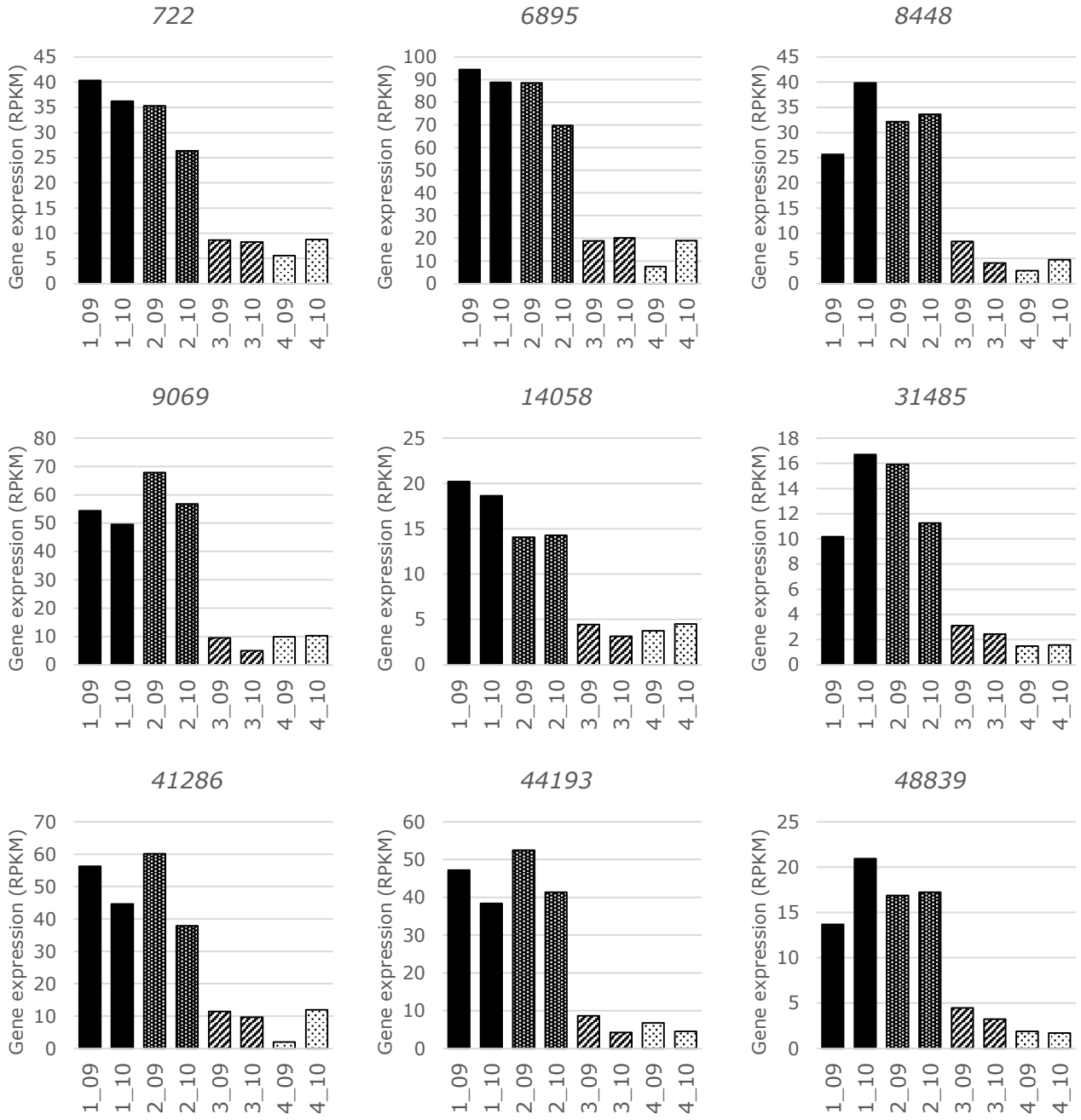


Figure 6. Gene expression profiles of 9 genes which are lower expressed in the high risk samples (3 & 4). The RPKM value was used to indicate the level of gene expression, which is the standard quantification of gene expression when studying RNA-Seq data.

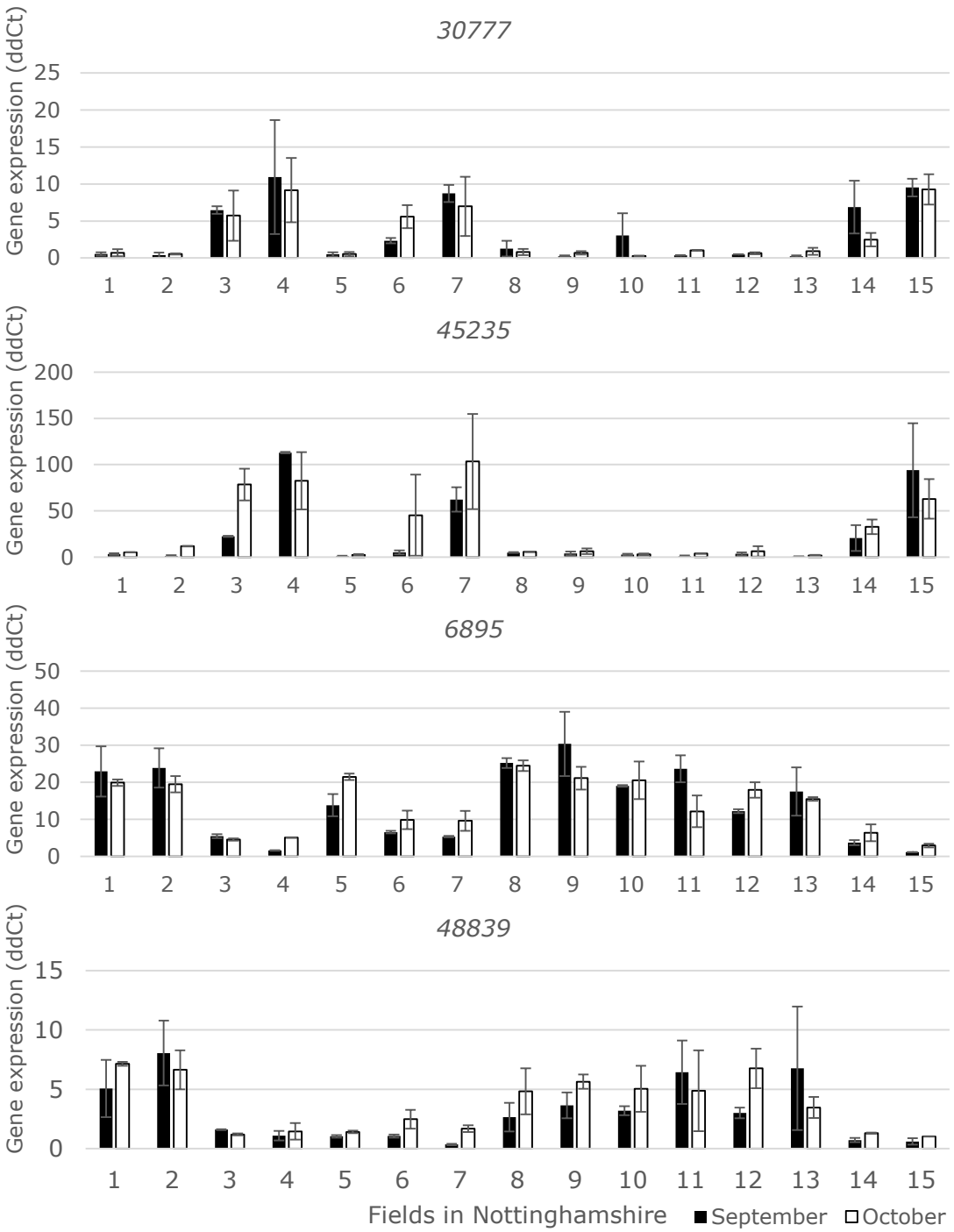
© Agriculture and Horticulture Development Board 2016. All rights reserved

Validation of the cavity spot indicators

Based on the longlist of potential cavity spot indicator genes, NSure selected 35 genes to study their expression via qPCR (CFX Connect Real-Time PCR Detection System, Bio-Rad). The gene expression profiles of those genes were first checked on the same samples that were selected for the RNA-Seq study. The gene expression profiles of the genes that were well measurable by qPCR and still showed a consistent difference between low and high risk samples were chosen to be measured in the remainder of the frozen samples collected from Nottinghamshire (Table 2). In figure 7, the expression profiles of four genes are shown that passed this selection. Genes *30777* and *45235* appeared to be induced in the high risk samples, while genes *6895* and *48* showed an opposite pattern. For several other genes similar patterns were observed (Appendix 2). According to the gene expression profiles obtained, fields 3, 4, 6, 7, 14 and 15 were considered to have a potential risk of cavity spot. The aim of farmers is to produce cavity spot free crops, but in reality they can work with crops with <15% incidence of cavity spot relatively well, whereas above this sort of level requires a greater level of staffing on the packing lines to remove the affected carrots. Because many of the fields were already harvested when the second evaluation took place, it was impossible to predict the progress of cavity spot. Quality evaluation on fields 3, 4 and 15 showed a high incidence on cavity spot which matched with the gene expression patterns observed. The estimated cavity spot incidence for fields 6 and 7 was below the <15%, but still substantial, which can explain the observed gene expression patterns. However, one might expect that similar patterns should be obtained for fields 8 and 9. According to the molecular analysis it was also assumed that field 14 would result in a field with high occurrence on cavity spot. Field 14 was harvested in December without covering with straw. As it was harvested early it only received one quality assessment. Although 3% of cavity spot incidence was recorded, it is possible that this field would have been a high risk if it had been left in the field for longer as the farm had issues with cavity spot previously. Moreover, this field was recorded as a potential high risk field and was harvested earlier to try and avoid the problem developing.

Sampling in Yorkshire was not performed according plan. Instead of collecting 2 replicate samples per field in September as well as in October, mostly 1 sample was collected instead a total of 4 (Table 3). The up- and downregulated genes that showed a consistent expression in the samples collected from Nottinghamshire were selected to study their expression in samples collected from Yorkshire. The gene expression of four of those genes is shown in figure 8.The consistent pattern of the up- and down regulated genes which was clearly visible for the Nottinghamshire samples, could not be observed in the analysed Yorkshire samples.

© Agriculture and Horticulture Development Board 2016. All rights reserved

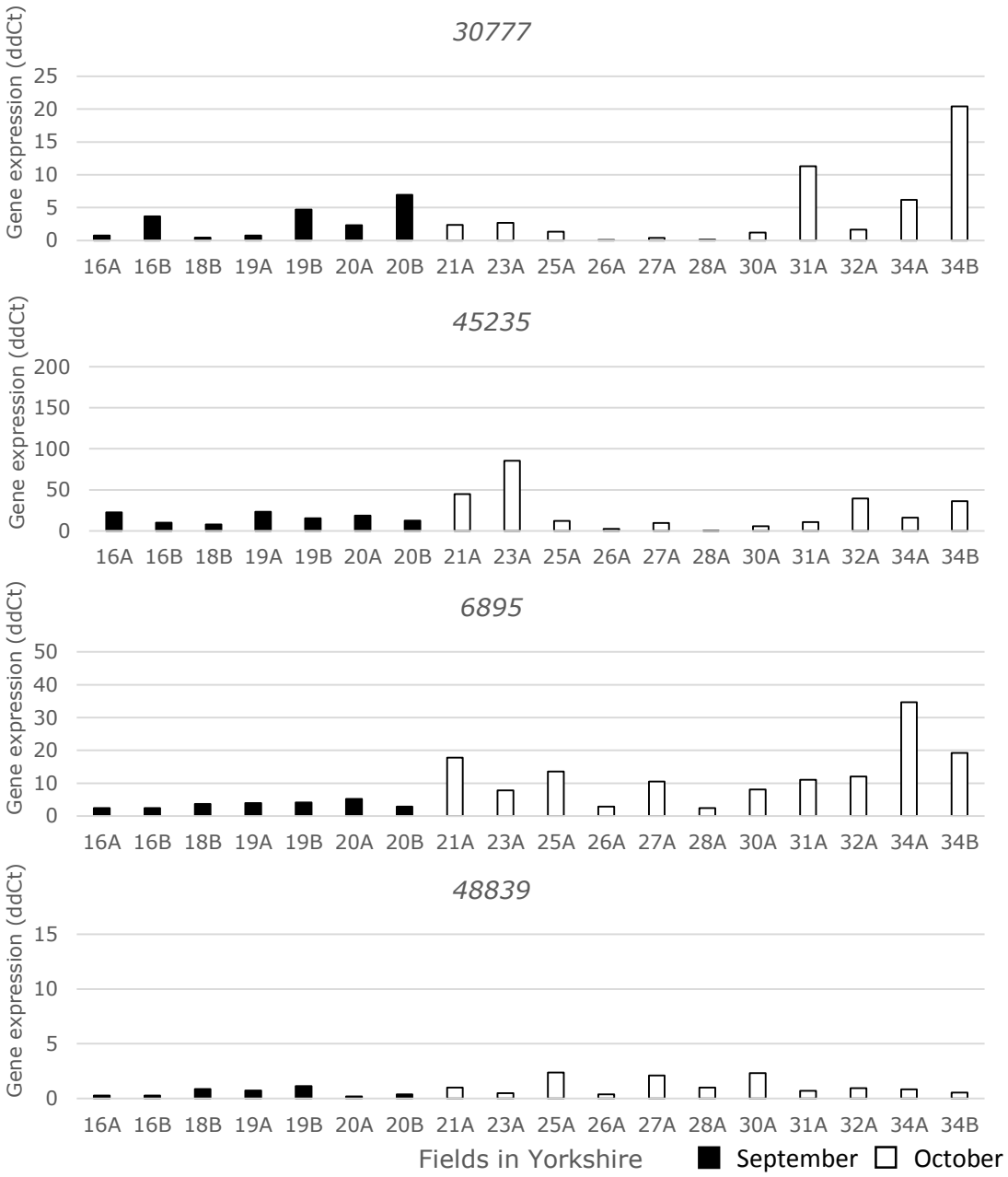


	Cavity spot %														
Field	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Evaluation 1	2	3	47	64	2	13	3	10	14	2	2	0	5	3	46
Evaluation 2	7	H*	H	H	H	H	11	H	H	0	8	H	H	H	45

* H = Harvested

Figure 7. Gene expression profiles of 4 genes measured in all collected samples from Nottinghamshire using qPCR. The level of gene expression was given as the ddCt, which is the relative Ct difference between the gene of interest and stable housekeeping genes. The error bars represent the standard deviation of the two replicates. Below, a table is presented containing the field evaluation results. Fields 1-4 were sequenced. Genes *30777* and *45235* are induced in the high risk samples (3 & 4) while genes *6895* and *48839* are down regulated.

© Agriculture and Horticulture Development Board 2016. All rights reserved



	Cavity spot %													
Field	**16**	**18**	**19**	**20**	**21**	**23**	**25**	**26**	**27**	**28**	**30**	**31**	**32**	**34**
Evaluation 1	8	14	45	22	18	0	0	0	0	0	50	0	0	50
Evaluation 2	37	20	H*	H*	22	-	5	0	0	0	5	10	5	H*

*H = Harvested

Figure 8. Gene expression profiles of 4 genes measured in collected samples from Yorkshire. Gene expression was measured using qPCR. The level of gene expression was given as the ddCt, which is the relative Ct difference between the gene of interest and stable housekeeping genes. Below, a table is presented containing the field evaluation results. Genes *30777* and *45235* should be induced in the high risk samples while genes *6895* and *48839* should be down regulated.

© Agriculture and Horticulture Development Board 2016. All rights reserved

Validation of the cavity spot indicators using the NSure sampling kit

Alongside frozen samples, juice samples were collected using the specific sampling card which is included in the NSure sampling kit used for commercial tests. In total 40 samples were taken from various fields in Nottinghamshire using this method (Table 2). Comparison of the gene expression results of the sampling card samples to the corresponding frozen samples, showed a high similarity for nearly all genes. Regression analysis performed on genes *30777*, *45235, 6895* and *48839* is depicted in figure 9. If the gene expression values obtained from frozen samples match to the genetic material which was isolated from the cards a R-squared value would be obtained of 1. The R-squared values obtained for these 4 genes were high considering that different sampling methods and RNA isolation protocols were used.

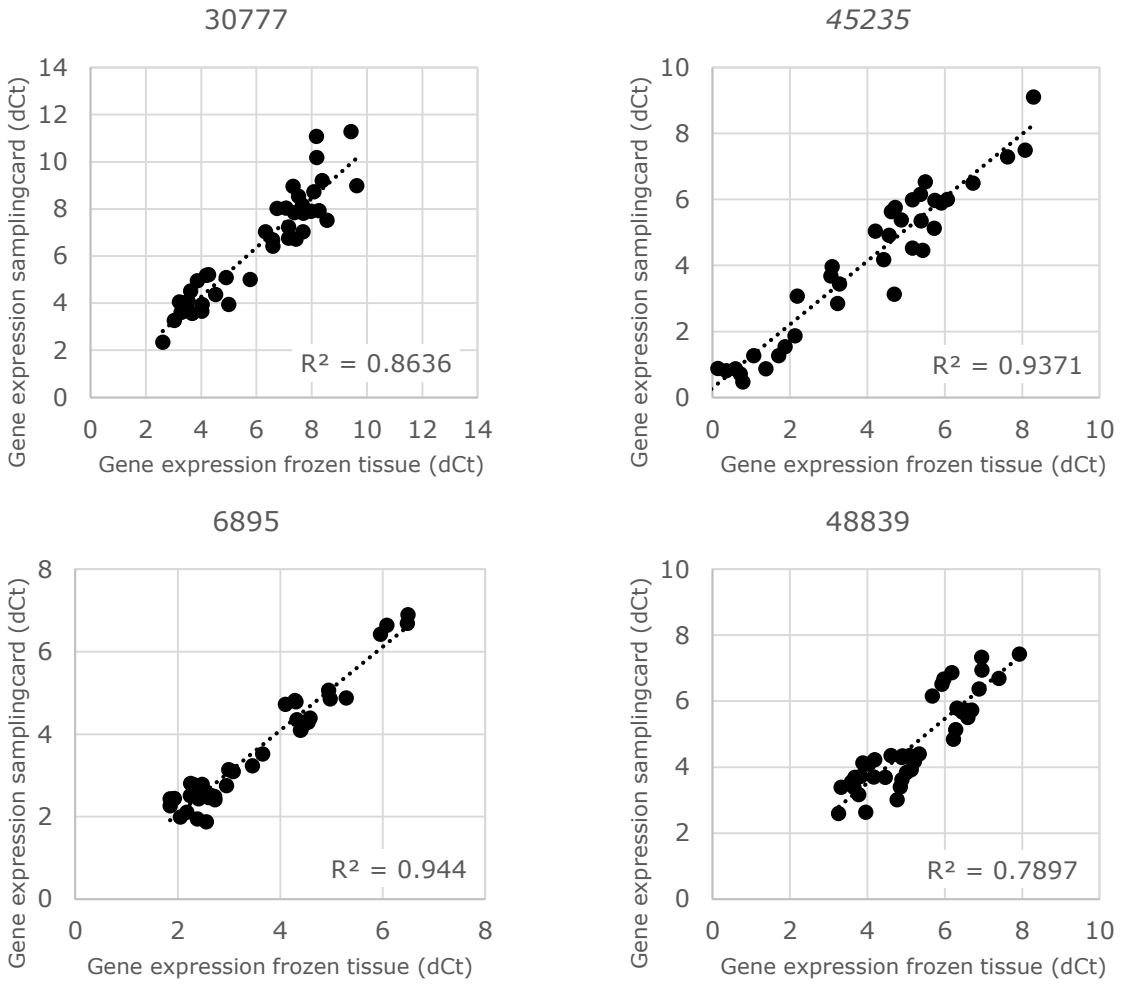


Figure 9. Regression analysis performed on 4 candidate genes. The level of gene expression was given as the dCt, which is the Ct difference between the gene of interest and stable housekeeping genes. R^2 stands for R-squared value. In total 40 comparisons were made between frozen samples and sampling card samples (Table 2).

© Agriculture and Horticulture Development Board 2016. All rights reserved

Discussion

The first objective, was to obtain a collection of carrot samples of the cultivar Nairobi with a wide range of cavity spot collected from 'Yorkshire' and 'Nottinghamshire'. The initial plan was to collect carrot samples from 15 different fields in Yorkshire as well as in Nottinghamshire in September and October. The plan was to collect two replicates per field, 25 carrots from each. In Nottinghamshire, the sampling was performed according to plan, but this was not the case in Yorkshire. The fields were visually assessed for the occurrence of cavity spot. Due to the weather conditions, most of the fields in Nottinghamshire were already harvested when the second assessment took place. The incidence of cavity spot ranged from 0% to 65% which was required to find cavity spot related genes as samples were needed with a low and high risk on cavity spot. The presence of other diseases made analysis difficult. Preferably, an RNA-Seq study would have been performed on carrots collected from fields which were only affected by cavity spot and carrots from disease free fields.

Based on the quality assessments and the sample availability, four fields were chosen to be studied in detail by RNA-Seq to identify cavity spot indicator genes. Two fields were selected that showed a low occurrence of cavity spot and two fields with a high level of cavity spot. To identify cavity spot indicator genes that showed a constitutive difference in expression between low and high risk samples, it was important to sequence samples that were collected in September and October. No technical issues occurred during processing of RNA-Seq data. Initially, it was assumed that the *de novo* assembly based on the sequences of the cultivar Nerac may not be suitable to map gene fragments of Nairobi. However, this was not the case. A high percentage of the obtained sequencing reads could be mapped to this reference and reliable gene expression values were obtained.

A Principal Component Analysis (PCA) was performed to obtain an impression of the global differences of the sequences samples. Based on this analysis it was observed that there was a segregation between the low and high risk categorized samples. The PCA analysis also demonstrated three fields had a global difference which seemed to be time related as the September and October samples were separated from each other in the PCA plot. Further study demonstrated that the genes differentially expressed in the three fields and between September and October only showed a little overlap (data not shown), meaning that in time diverse biological processes are triggered which are specific for the carrots growing in that particular field.

The next step, which was also depicted as an objective, was to generate a longlist of carrot cavity spot specific indicator genes. To obtain the best suitable candidate genes that showed a consistent difference in time between low and high risk samples, a stringent selection was

© Agriculture and Horticulture Development Board 2016. All rights reserved

performed based on pairwise comparisons between low and high risk samples. This resulted in approximately 70 genes. The fold change ratios between the low and high categorized samples were higher or equal to 2,5. No enormous fold changes were observed, but this finding is not unusual, as we were dealing with an early detection. In addition, it may be assumed that not every carrot within the sample was infected. Although the best suitable genes are genes that show a consistent difference in time between the low and high risk samples it could also be worthwhile to enlarge the longlist with indicators that change over time. Adding such genes to the test, can give an impression about the development of the disease. Out of the longlist, 35 genes were selected to validate their expression patterns via qPCR. For most of these genes the gene identity was known and their function could be related to plant defence related processes. Interestingly, two genes (Figure 5, genes *11849* and *29647*) were also incorporated in the StoreNSure carrot test which is used to detect black spot. These generic indicators could be worthwhile for a test, but indicator genes that only respond upon an infection with *Pythium violae* are the best candidates. The expression of the selected genes was initially checked on the same frozen samples that were selected for the RNA-Seq study. The gene expression profiles of the genes that were well measurable by qPCR and still showed a consistent difference between the low and high risk samples were selected to be measured in the remainder of the frozen samples. As sampling in Nottinghamshire was performed according plan, it was decided to study the gene expression patterns in those samples first. Several up- and down regulated genes showed a consistent pattern in the fields that were studied in September and October. Based on the gene expression profiles, six fields were considered as potential risk fields. Quality assessment showed for three fields a high level of cavity spot (>40%) which corresponds to the molecular data. Two other fields that were characterized as potential risk fields showed a cavity spot incidence of 10-15%. Similar results were expected for two other fields with similar levels of cavity spot incidence, but according to the gene expression patterns they were categorized as low risk fields. A possible reason for this could be, that a 10% infection rate is difficult to measure early in time as only a few carrots are infected. Increasing the sample size or collecting more replicates from the field could improve the detection of a low infection. In reality farmers can work with <15% incidence of cavity spot relatively well, suggesting that it is most important to pick out the fields that show a high incidence on cavity spot. The sixth field that was categorized as a potential risk field showed a low level of cavity spot during the first quality assessment. Although 3% of cavity spot incidence was recorded, it is very possible that this field would have been a high risk if it had been left in the field for longer as the farm had issues with cavity spot previously. This field was recorded as a potential high risk field and was harvested earlier to try and avoid the problem developing. The fields in

© Agriculture and Horticulture Development Board 2016. All rights reserved

Nottinghamshire were also affected by other pathogens, but the genes that passed the selection seem not to be influenced by the presence of other diseases.

The consistent pattern of the up- and down regulated genes that was observed in the samples collected from Nottinghamshire in September and October, could not be noticed in the analysed samples of Yorkshire. The limited sample collection makes it difficult to value the results. It could be questioned whether this difference is due to a certain difference between the regions, which seems unlikely, or is due to the way of sampling. Taken together, it would be worthwhile to study the functionality of the selected genes in Nottinghamshire and Yorkshire in a second season.

Alongside frozen samples, which were used for RNA-Seq and the indicator selection, juice samples were collected by using the NSure sampling kit used for commercial tests. It is essential that the gene expression patterns and levels obtained when using this sampling method is similar when the genetic material is isolated out of frozen tissue. The genes that showed a consistent pattern in the samples collected from Nottinghamshire were selected to study if similar gene expression patterns and values are obtained when using frozen tissue sampling or the NSure sampling kit. For nearly all genes similar results were obtained, showing that this method can be used, once the putative indicators have been validated.

Conclusions

The time projected to develop a molecular test to detect cavity spot at an early stage using carrot specific genes is two seasons of research. The milestones set for the first year of test development have been accomplished. Within this project a set of potential indicator genes for cavity spot have been identified which are valuable to validate in a second season. It is important that next season sampling and quality assessment goes according to plan to obtain a reliable sample collection. Some additional sequencing could be performed to possibly increase the number of cavity spot indicator genes.

Knowledge and Technology Transfer

Activity	To	Date
Summary of the progress	BCGA	22-03-2016
Oral presentation	Poskitt Ltd, Strawson Limited, Elsoms Seeds, AHDB	28-06-2016
Newsletter NSure	People subscribed to our newsletter. It can also be viewed on our website	06-09-2016
Summary of the progress	BCGA	13-10-2016

© Agriculture and Horticulture Development Board 2016. All rights reserved

Glossary

De *novo assembly*	- a self-created library of genes or gene parts
Down-regulated	- describes a gene which has been observed to have lower expression in one sample compared to another
Gene /expression	- the process by which a gene is turned on to produce the specific biological molecular (RNA) encoded by that gene. Lowering gene expression leads to less RNA of this particular gene, while induction leads to more RNA (and thus a higher gene expression)
Housekeeping gene	- a gene that is expressed at a relatively constant level across many known conditions
Principal component analysis (PCA)	- PCA allows you to separate different samples in a graph according to the global differences between the samples taken into account all available (gene expression) data for each one of the samples
Read	- piece of cDNA that is sequenced
RNA	- a nucleic acid that is the primary product of gene expression
RNA sequencing (RNA-Seq)	- method to reveal the presence and quantity of RNA (from the active genes) in a sample at a given moment in time
RPKM	- reads per kilobase per million reads in a dataset.
Quantitative RT PCR (qPCR)	- method to specific quantify the RNA you are interested in
Single-end sequencing	- a high-throughput sequencing technique that generates short reads of approximately 50-100 nucleotides in length

© Agriculture and Horticulture Development Board 2016. All rights reserved

Up-regulated	- describes a gene which has been observed to have higher expression in one sample compared to another

© Agriculture and Horticulture Development Board 2016. All rights reserved

References

Barbara D.J. (2010) Carrot cavity spot: (i) using quantitative PCR to predict disease in strawed crops: (ii) controlling moisture for optimum disease management. AHDB FV 353.

Klemsdal S.S., Herrero M.L., Wanner L.A., Lund G., Hermansen A (2008) PCR-based identification of *Pythium* spp. Causing cavity spot in carrots and sensitive detection in soil samples. Plant Pathology 57:877-886.

Kromwijk A., Verhoef N., van Noort F., Balk P., van Wordragen M (2013) Development of a molecular test to ensure good flowering of *Viburnum opulus* var. *roseum.* Acta Horticulturae 990:105-112.

Stattin E., Verhoef N., Balk P., van Wordragen M., Lindström A. (2012) Development of a molecular test to determine the vitality status of Norway spruce (*Picea abies*) seedlings during frozen storage. New Forests, DOI: 10.1007/s11056-012-9320-1.

White J.G., Lyons N.F., Petch G.M (1996) Development of a commercial diagnostic test for the pathogens which cause cavity spot on carrot. British Crop Protection Council Symposium Proceedings 65: 205-2010.

© Agriculture and Horticulture Development Board 2016. All rights reserved

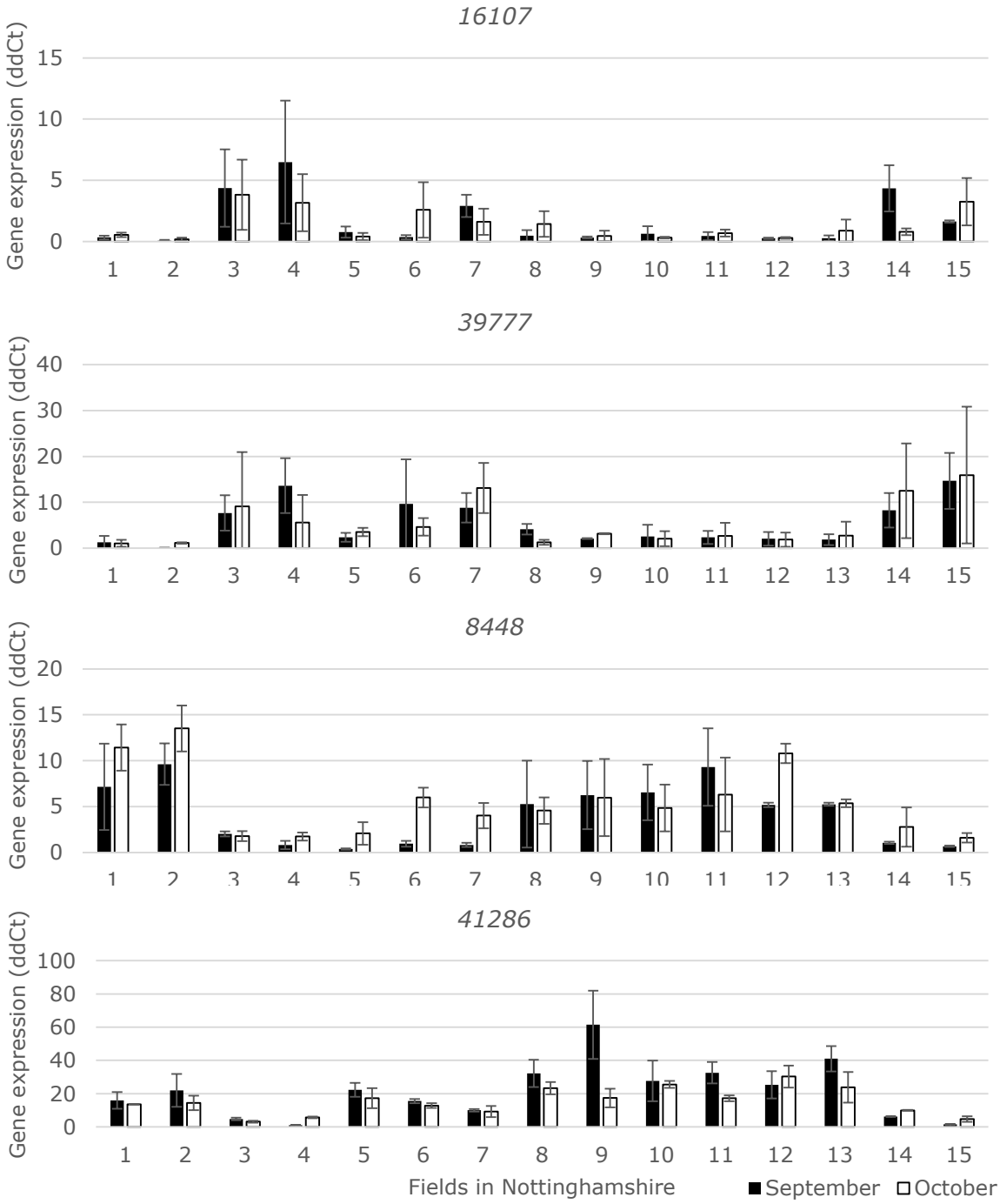
Appendix 1

Table 1. Field evaluation of fields in Nottinghamshire

	Quality evaluation 1		Quality evaluation 2	
Field	**Date**	**Number of other diseases**	**Date**	**Number of other diseases**
1	17-11-2015	scab (3)	1-2-2016	scab (5), violet root rot (1)
2	20-11-2015	scab (5) violet root rot (1)	-	-
3	18-11-2015	scab (4), virus (8), crown rot (2)	-	-
4	18-11-2015	scab (7), crown rot (2)	-	-
5	20-11-2015	scab (27)	-	-
6	17-11-2015	scab (6), crown rot (1), carrot fly damage (1)	-	-
7	18-11-2015	scab (8), virus (3), carrot fly damage (1), crown rot (1)	20-11-2016	scab (6), crown rot (4), virus (5), nematode (4), carrot fly damage (1)
8	17-11-2015	scab (6)	-	-
9	14-10-2015	scab (20, violet root rot (1)	-	-
10	19-11-2015	scab (16)	1-2-2016	scab (10), virus (4), crown rot (3)
11	17-11-2015	violet root rot (2)	1-2-2016	Scab (13), virus (3), carrot fly (1)
12	20-11-2015	scab (15), violet root rot (2)	-	-
13	20-11-2015	scab (20)	-	-
14	17-11-2015	scab (23), virus (2), carrot fly damage (2)	-	-
15	18-11-2015	scab (8), virus (2)	20-1-2016	scab (16)

© Agriculture and Horticulture Development Board 2016. All rights reserved

Appendix 2



Gene expression profiles of 4 genes measured in all collected samples from Nottinghamshire using qPCR. The level of gene expression was given as the ddCt. The error bars represent the standard deviation of the two replicates. Fields 1-4 were sequenced. Genes *16107* and 39777 are induced in the high risk samples while genes *8448* and *41286* are down regulated.

© Agriculture and Horticulture Development Board 2016. All rights reserved