Project title:	Next Generation Berries – Implementing Genome-wide Selection Approaches in Strawberry
Project number:	CP 163
Project leader:	Richard Harrison, NIAB EMR Michael Shaw, University of Reading Alison Bentley, NIAB
Report:	September 2017
Previous report:	Not Applicable
Key staff:	Joe Q He, NIAB EMR
Location of project:	NIAB EMR, New Road, East Malling, UK, ME19 6BJ
Industry Representative:	Tom Rogers, CPM Retail Ltd., Oakdene Farm, Leeds Road, Langley, UK,

	ME17 3LT
Date project commenced:	01/10/2016
Date project completed **(or expected completion date):**	Expected Early 2020

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.

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.

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.

Joe Q He

PhD Student

NIAB EMR

Signature ....Joe He.................... Date ....30/08/2017..................

[Name]

[Position]

[Organisation]

Signature .... Date ....

Report authorised by:

[Name]

[Position]

[Organisation]

Signature .... Date ....

[Name]

[Position]

[Organisation]

Signature .... Date ....

CONTENTS

GROWER SUMMARY

Headline

- Phenotyping of seven external strawberry fruit quality traits can be identified quantitatively with greater precision using a novel 3D image capture and analysis platform and this will facilitate the use of more powerful statistical models to aid breeders in selecting for these seven traits.

Background and expected deliverables

Strawberry breeders aim to generate novel genotypes that express traits suitable for the industry in their target region. Over the past 200 years, significant progress has been made in traits such as flavour, berry size, yield, and cropping season duration. Current goals in strawberry breeding include improvements in maintenance of post-harvest fruit quality, yield, texture and flavour.

In traditional breeding, crossing programmes have been based on identification of desirable traits in parental germplasm material. Offspring are assessed throughout the growing and harvest season and scored against a weighted index of favourable traits. The highest scoring individuals are selected to progress onto further trials. Larger scale trials are conducted to gather additional information, such as yield and picking speed, and to confirm the presence of the favourable traits. Additionally, the selected genotypes are assessed for suitability across a range of environmental conditions, with particular focus on the target region. Overall, the time taken from making crosses to release of a novel cultivar may be between 7 and 20 years.

Strawberry breeding programmes take place all over the world. They are funded by a mixture of private sources, governmental grants and returns on royalties from released cultivars. A typical breeding programme will employ fewer than five full time equivalent workers, cross tens to hundreds of parents and select from thousands to tens of thousands of offspring per year.

Genetic markers are detectable features within the genome of a plant that may differ between individuals of the same species. Markers that are physically close to genes controlling economically important traits tend to be co-inherited with the gene when the plant produces offspring, making some markers reliable proxies for these genes. Over the past 20 years, the number of known markers has dramatically increased and the cost of identifying them has greatly decreased. It is now possible to incorporate genomic information in the breeding process to aid breeders in selection of the optimal individuals.

Genomic selection (GS) is an advanced breeding technique which integrates genomic and trait data to make predictions on the breeding values of individuals. GS requires a “training population” which is genotyped for as many markers as possible and assessed for all relevant traits. A statistical model is generated which quantifies the effect of every marker on every trait. Subsequently, individuals from the “breeding population” are genotyped. Solely on the basis of the statistical model and genotypes, predictions can be made about the performance of each individual within the “breeding population”, and thus selections can be made.

GS offers a range of benefits relative to conventional breeding approaches. Firstly, it allows for greater selection accuracy as the confounding environmental effects on a trait can be eliminated. Secondly, it allows for strong selection on traits that are expensive or difficult to assess or selection on traits that are apparent only under rare environmental conditions. Thirdly, as multiple traits can be assessed, GS potentially allows selection at the juvenile stage, reducing the duration of the breeding cycle. Moreover, GS is particularly suitable for identification of traits that are controlled by many genes (polygenic traits) as its simultaneous regression of all markers on all traits reduces the likelihood of over/underestimation of effect size. As many economically important traits are expected to be polygenic, GS would allow further improvement in selection accuracy.

The ultimate aim of this project is improvement in commercial deployment of genomic selection (GS) in plant breeding. GS is likely to improve the accuracy of selection, increase genetic gain per unit time and also reduce the duration of the breeding cycle. Strawberry (*Fragaria x ananassa*) is used as a model organism as a popular, economically important and amenable species.

Summary of the project and main conclusions

Deployment of GS in strawberry breeding populations is likely to deliver increased genetic gain per unit time and reduced duration of the breeding cycle. Three major areas were identified for improvement in current GS approaches to optimise the process for octoploid strawberry:

1. **High throughput quantitative phenotyping** - The most powerful models for GS require quantitative inputs to generate quantitative predictions of breeding value. Currently, there are a range of highly precise and quantitative techniques such as mass spectrometry, liquid chromatography and diode arrays. However, these techniques are costly to implement, have low throughput and importantly, cannot assess many of the traits of interest, such as berry morphology and colour.

   Currently utilised approaches for assessing these traits mostly rely on the human eye, which potentially results in bias and typically generate less useful ordinal data. An imaging platform was developed using a camera and computational algorithms to capture data in 3D and quantify seven external fruit quality traits. Analysis of 100 fruit shows good concordance with manually measured traits and greater precision. Moreover, the novel method required approximately five-fold less labour and required less than £1,000 to set up.

2. **Cost effective scalable genotyping** - For GS to be commercially viable, the economic benefit of deploying it must outweigh the costs associated with its implementation. Once the initial statistical model is generated, GS requires little phenotyping to be conducted, whereas traditional assessment requires phenotyping of all individuals. In contrast, GS requires that every plant in the "breeding population" is genotyped. As GS is expected to offer a range of benefits, if genotyping costs are made comparable to phenotyping costs, then GS would likely be viable for commercial deployment. The current estimate of phenotyping is approximately £5 per individual, whereas the gold standard Affymetrix IStraw90 Axiom SNP array costs approximately £50 per sample.

   Genotyping-in-thousands (GT-seq) multiplexes hundreds to thousands of samples (commonly found within breeding populations) within a single Illumina sequencing run, with primers targeting several hundred specific loci to achieve cost reduction. Estimates suggest that this approach may reduce costs to less than £10 per sample, whilst maintaining sufficient power to detect most QTLs that the SNP array is predicted to identify. Moreover, GT-seq is scalable, allowing the incorporation of new, informative markers as they are discovered, or adjustments to the power of the selection procedure as resources allow.

3. **Statistical techniques for octoploid strawberry** - Currently, there remains no consensus for the optimal GS models to utilise. Different reports suggest different models generate optimal results for different species, or under specific theoretical conditions. As an allo-octoploid that suffers from inbreeding depression, strawberry is unlike any of the other species for which GS has been experimentally deployed.

   Phenotypic and genotypic data was generated over the past four years as part of this project and the project of a previous PhD student relating to a bi-parental "Redgauntlet" x "Hapil" mapping population. A range of GS models will be implemented on the data to identify the optimal model.

To date, significant progress has been made on high-throughput quantitative phenotyping and cost-effective scalable genotyping.

Financial benefits

Phenotyping of seven external strawberry fruit quality traits can be sped up five-fold, reducing labour costs, using a novel automated 3D image capture and analysis platform. The system is currently optimised for strawberry breeding, but may be adapted for quality control in production pipelines.

Ultimately, successful deployment of GS is likely to benefit breeders as greater genetic gain per unit time is achieved with reduced effort.

Action points for growers

- There are currently no action points for growers arising from this research.

SCIENCE SECTION

Introduction

Genetics and Breeding

Fragaria ananassa (strawberry) is an allo-octoploid (Ichijima 1926) formed from a chance hybridisation event between *F. virginiana* and *F. chiloensis* in Brittany, France. It was first identified and characterised by French botanist Antioine Nicolas Duchesne in 1766 (Darrow 1966). Strawberry has a basic chromosome number of 7 and an estimated genome size of 698 – 720 Mbp (Hirakawa et al. 2014).

Despite advances, the contribution of genetic material to *F. ananassa* remains unclear due to difficulties distinguishing between homeologous chromosomes. It is generally accepted that *F. vesca* contributed at least one subgenome (DiMeglio et al. 2014; Illa et al. 2011) with suggestions of another two subgenomes being related to *F. iinumae* (Tennessen et al. 2014). The remaining subgenome has no clear pattern of phylogeny (Sargent et al. 2015), making a working genomic structure of AABBCCDD the most appropriate assumption, utilized in this study.

The goal of strawberry breeding is to generate novel genotypes with attributes suited for the industry of its target region (Mathey et al. 2013). Generation of novel genotypes relies on crossing germplasm material exhibiting agronomically valuable traits, such as high yield or disease resistance. As the genotype of the plant cannot be directly observed, traditional breeding selects based on a weighted index of phonotypes. These offspring are then trialled over several years, usually under different environmental conditions to confirm these traits before release. Usually, a new cultivar takes 7 years to develop from breeding to commercial release, but may take up to 20 years (Chandler et al. 2012).

Strawberry breeding programmes are present across the world, which share broadly similar aims of improving fruit quality, pathogen resistance and productivity (Karina Gallardo et al. 2012). Funding comes from a mixture of sources including governmental, private and royalties on intellectual property. A programme typically has less than 5 full time equivalent workers, performs tens or low hundreds of crosses and screens tens of thousands of plants per year (Knight et al. 2005).

Over the past 200 years of breeding, a range of traits have been improved upon in strawberry, including fruit size, marketable yield, pathogen resistance and production season length (Chandler et al. 2012). However, this may have come at a cost to genetic diversity in the germplasm material, perhaps reducing potential improvements in these traits in the future

(Gil-Ariza et al. 2009). Whilst it has been possible to transform strawberries for many years (Nehra et al. 1990), there are no known plans to mass release genetically engineered strawberries for human consumption due to public perceptions against genetically modified foods (Schaart et al. 2011).

Genomic Selection

Genomic selection (GS) is an advanced breeding technique that integrates genotypic and phenotypic information to make performance predictions on a panel of agronomically important traits. Deployment requires a training population, which is densely genotyped and phenotyped for the agronomically relevant traits. A statistical model is developed, which associates the genotype and phenotype. Solely on the basis of the genotype and statistical model, breeding values for breeding material is estimated and selections are made (Heffner, Sorrells, and Jannink 2009; Meuwissen, Hayes, and Goddard 2001).

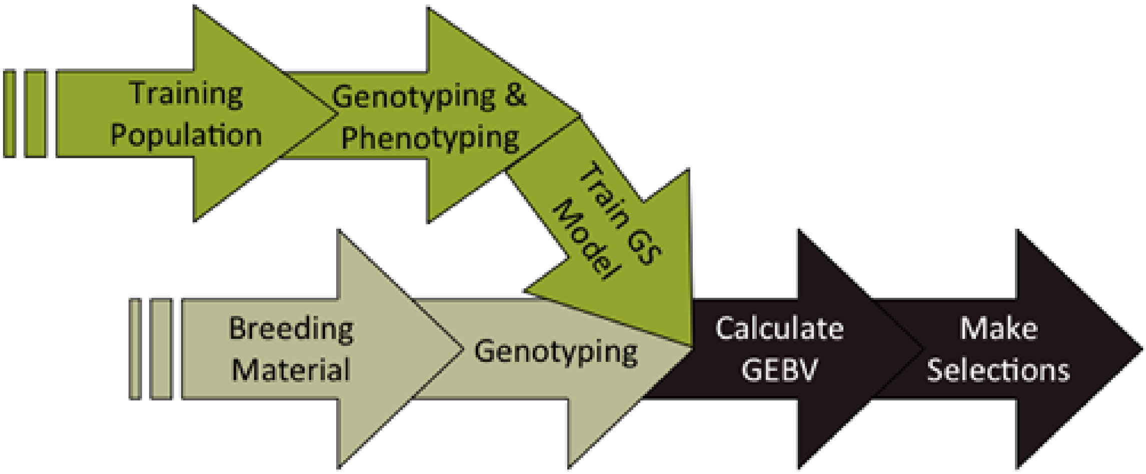


Figure 1. Schematic of GS. (Heffner et al. 2009)

Deployment of GS offers a range of potential benefits. Firstly, genotypic information allows for strong selection on traits that are difficult or expensive to phenotype, or whose expression depends on specific environmental conditions (Xu and Crouch 2008). Secondly, it potentially allows reduction of the duration of breeding cycle as selections can be made at the juvenile stage (Meuwissen et al. 2001). Thirdly, it allows for reduction of testing effort by reducing or eliminating some field experiments (Gezan et al. 2017). Strength of selection in strawberry is limited by the natural genetic variation present within the population. GS would also allow estimation of variability within the germplasm material to potentially control the reduction of genetic diversity to ensure future breeding remains effective (Daetwyler et al. 2007). The overall aim of this PhD project is to experimentally deploy GS in a commercial octoploid strawberry breeding population.

Project Outline

In order to deploy GS in strawberries, three major areas were identified for improvement from what is currently available. These three topics, along with experimental implementation and validation of GS in strawberry are likely to constitute the four results chapters in the thesis to be produced as part of this PhD project (outlined below). At present, significant progress has been made on high-throughput quantitative phenotyping and cost effective scalable genotyping.

1. High-throughput Quantitative Phenotyping

There are currently a range of methods to phenotype strawberries. Analysis of anti-oxidative properties of food frequently utilises chromatography separation followed by diode arrays to measure reduction potential (Aaby, Ekeberg, and Skrede 2007; Määttä-Riihinen, Kamal-Eldin, and Törrönen 2004). Investigation of the metabolome typically uses mass spectrometry (MS), or variants of this technology. MS allows for the analysis of hundreds of compounds in a single run, potentially allowing components of flavour and aroma to be quantified (Hanhineva 2011). In order to measure other important physical characteristics of strawberry, such as flower-related traits, plant characteristics and fruit characteristics, breeders usually rely on assessment by eye (Mathey et al. 2013).

Limitations remain on these assessment methods. MS, chromatography and diode arrays are costly and likely to be uneconomical to implement in a strawberry breeding programme. Moreover, despite efforts to improve throughput (Walker et al. 2006), these techniques require a large amount of time. These techniques are also unable to assess a range of agronomically important traits such as fruit colour or dimensions. Evaluation by eye typically scores traits on an ordinal 9-point scale (Mathey et al. 2013), making it unsuitable for the most powerful quantitative GS models. Furthermore, human scoring is likely to introduce biases and still requires significant time to train assessors and to make assessments.

2D bio-imaging techniques have been investigated as a method of overcoming these limitations. Colour analysis has been successfully conducted in apple (Throop et al. 2005), citrus (Blasco, Aleixos, and Moltó 2007), mango (Kang, East, and Trujillo 2008) and banana (Mendoza and Aguilera 2004). Strawberries could be graded by considering their colour and shape as captured using an imaging system (Liming and Yanchao 2010; Nagata et al. 2000). However, 2D systems are not always reliable for fruit phenotyping due to uneven colour distribution and occlusion of morphology from different viewing angles (Paulus et al. 2014).

Recently, 3D imaging has been explored as a method to overcome the limitation of occlusion, as the cost of hardware decreases and reconstruction algorithms improve. Multi-view stereovision (MVS) is a promising technique which captures images as the target is rotated (Rose, Paulus, and Kuhlmann 2015). Reconstruction of the 3D object can be achieved using the Structure from Motion (SfM) algorithm, which detects points of interest across the range of images and matches them to the same points as viewed from different angles (Fonstad et al. 2013).

2. Cost-effective Scalable Genotyping

Molecular markers are distinguishable features at specific loci that separate individuals from each other, with many known markers present as different alleles within the genome. As some markers are likely to be in close linkage with genes, genetic markers are potentially useful in identifying associations between traits of economic importance and genotype (Semagn, Bjørnstad, and Ndjiondjop 2006). A range of molecular markers have been developed, usually identified through their mechanism of detection. Commonly utilised markers include restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), inter-simple sequence repeats (ISSR), microsatellites, and single nucleotide polymorphisms (SNP) (Semagn et al. 2006).

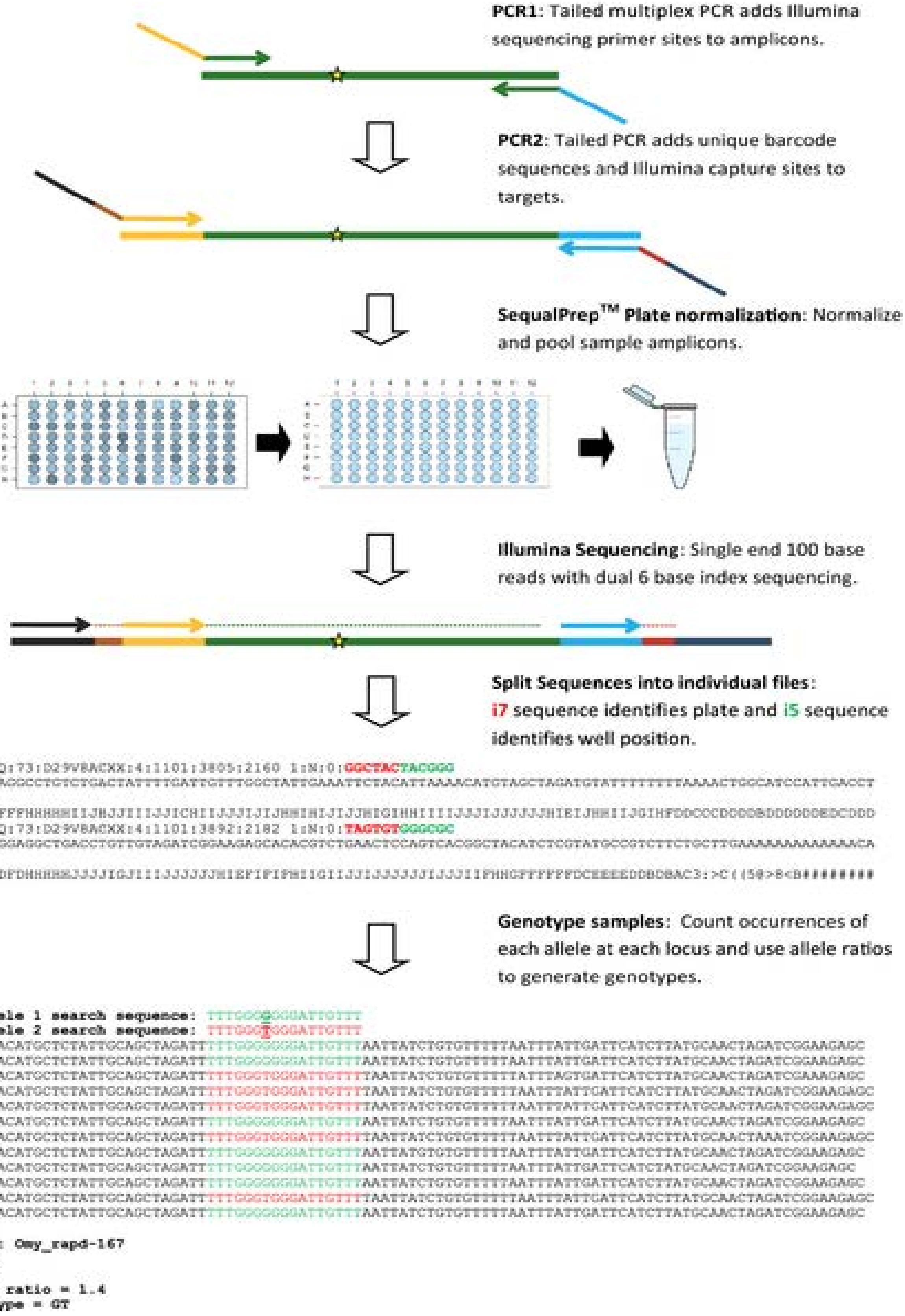
Of these markers types, only SNPs cannot be discriminated using gel electrophoresis, as a SNP marker is just a single base change in a DNA sequence. However, their abundance and fairly even distribution within genomes makes them highly attractive polymorphisms for marker assisted breeding efforts. In the case of strawberries, analysis of 384 individuals of 20 octoploid varieties identified over 36 million potential variants when mapped to the “Hawaii 4” diploid genome. On average, the Affymetrix IStraw90 Axiom SNP array contains 1 marker per 0.5cM (Bassil et al. 2015).

Current efforts in SNP detection for strawberry genotyping is largely based on microarrays. Microarrays are small glass chips encased in plastic that have thousands of microdots of cDNA flanking known SNPs. When the array is flooded with sample DNA, the presence of a particular allele will result in specific hybridisation to the appropriate microdot. Fluorescence can then be used to detect the hybridisation and thus determine the genotype of the sample (https://www.genome.gov/10000533/dna-microarray-technology/).

Although the density of markers using microarray based genotyping is sufficient to deploy GS, the total cost per individual is high compared to conventional breeding (deployment of the Affymetrix Istraw90 Axiom SNP array currently costs approximately £50 per sample, compared to approximately £5 per sample for conventional breeding). It is clear that the benefits of GS do not justify the costs of genotyping (Gezan et al. 2017). Moreover, targeted sequences in microarrays are not scalable as novel markers are detected, annotated or markers identified as ineffective; A new SNP array must be designed (Bassil et al. 2015; Verma et al. 2017).

Genotyping-in-Thousands (GT-seq) multiplexes hundreds to thousands of individuals in two normalised PCR reactions to generate targeted and uniquely bar-coded amplicons (Figure 2). After normalisation, the amplicons are pooled into a single tube for analysis with Illumina sequencing. The bar codes allow the identification of the individual to resolve alleles. One advantage of this system is its scalability. As GT-seq utilises a user defined set of primer pairs, markers known to associate with QTLs can be targeted and the total number of targets can be adjusted as resources allow. Another advantage of this approach is that a single amplicon can potentially target multiple homeologous chromosomes and thus generate information across multiple subgenomes with a single primer pair.

GT-seq will be explored as part of this PhD to cost effectively generate marker data for GS. Amplicons will be optimised for octoploid strawberry to maximise information gained, whilst minimising amplicons targeted.



PCR1: Tailed multiplex PCR adds Illumina sequencing primer sites to amplicons.

PCR2: Tailed PCR adds unique barcode sequences and Illumina capture sites to targets.

SequalPrep^{TM} Plate normalization: Normalize and pool sample amplicons.

Illumina Sequencing: Single end 100 base reads with dual 6 base index sequencing.

Split Sequences into individual files: i7 sequence identifies plate and i5 sequence identifies well position.

```
@HISEQ:73:D29VBACXX:4:1101:3805:2160 1:N:0:GGCTACTACGGG
AGGTAAGGCCTGTCTGACTATTTTGATTGTTTGGCTATTGAAATTCTACATTAAAACATGTAGCTAGATGTATTTTTTTTTAAAACTGGCATCCATTGACCT
+
B@CDFFFFHHHHHIIJHJJIIIJJICHIIJJJIJIJHHIHIJIJJHIGIHHIIIIJJJIJJJJJHIEIJHHIIJGIHFDDCCCDDDDBDDDDDDEDCDDD
@HISEQ:73:D29VBACXX:4:1101:3892:2182 1:N:0:TAGTGTGGGCGC
AAAGTGGAGGCTGACCTGTTGTAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTTCTGCTTGAAAAAAAAAAAAAACA
+
@CCFDDFDHHHHHJJJJIGJIIIJJJJJJJHIEFIFIFHIIGIIJJIJJJJJJJIJJJIIFHHGFFFFFFDCEEEEDDBDBAC3:>C((5@>8<B#########
```

Genotype samples: Count occurrences of each allele at each locus and use allele ratios to generate genotypes.

```
  Allele 1 search sequence: TTTGGGGGGGATTGTTT
  Allele 2 search sequence: TTTGGGTGGGATTGTTT
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGTGGGATTGTTTTAATTATCTGTGTTTTTTATTTAGTGATTCATCTTATGCAACTAGATCGAAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGTGGGATTGTTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGTGGGATTGTTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGTGGGATTGTTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAAATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATGTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGTGGGATTGTTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC
ACCCAACATGCTCTATTGCAGCTAGATTTTTGGGGGGGATTGTTTAATTATCTGTGTTTTTTAATTTATTGATTCATCTTATGCAACTAGATCGGAAGAGC

Locus: Omy_rapd-167
G = 7
T = 5
A1:A2 ratio = 1.4
Genotype = GT
```

Figure 2. Schematic of Genotyping-in-thousands (Campbell, Harmon, and Narum 2014)

3. Statistical Techniques for Octoploid Strawberry

There are two major statistical methods to perform GS: linear regression models (Ogutu, Schulz-Streeck, and Piepho 2012), and reproducing kernel Hilbert Spaces (Gianola and Van Kaam 2008). The former tends to be more widely implemented and can include a range of shrinkage methods including ridge regression, LASSO, elastic nets and Bayesian approaches to fit a range of assumptions. Currently, it is unknown which model is optimal for GS and experimental validation of GS tend to utilise different models to identify the highest accuracy (Gezan et al. 2017).

Phenotypic and genotypic information has been gathered from a “Redgauntlet” x “Hapil” mapping population for the past 4 years as part of this project and a previous PhD project. A range of models will be tested on the data to identify the most suitable techniques for GS in octoploid strawberry.

4. Deployment and Validation of Genomic Selection

Utilising results from the previous sections, GS will be implemented on a commercial octoploid strawberry breeding population based at NIAB EMR. Prediction accuracy will be assessed the following year. Simultaneously, conventional breeding will be applied to the breeding population and GS and conventional selection will be compared.

Materials and methods

High-throughput Quantitative Phenotyping

Fruit Material. 100 strawberries were purchased from local supermarkets, including 10 different varieties, to represent the diverse range of commercially available strawberry phenotypes. All fruit were assessed before their “best before” dates. Fruit would likely have been subjected to chilling to 4°C within 4 hours of harvest and kept at that temperature throughout the supply chain until sale. Fruit was maintained at 4°C until assessment.

Manual Assessment. In order to validate the results of the 3D analysis, phenotypic data was collected manually immediately after imaging. Measurements of dimensions were performed using a pair of digital callipers and measurement of volume was performed using an overflow can and a measuring cylinder.

Table 1. Manual Scoring metric for seven external strawberry fruit traits

External quality parameter	Scoring metric
Achene Number	Number of achenes visible, without disturbing calyx
Calyx size	Maximum Euclidean distance between any pair of points on the calyx
Colour	Scale 1 – 8 (Strawberry colour chart for experimental ends, Ctifl, France)
Height	Dimension of fruit from centre of calyx to tip of nose
Length	Greatest dimension of fruit orthogonal to the height
Width	Greatest dimension of the fruit orthogonal to both height and length
Volume	Volume of displaced water when fruit was completely submerged

Image Capture. The sample was pinned onto a dark blue holder (38mm × 19mm × 19mm) placed in the middle of a turntable and rotated at 0.02Hz. A single lens reflex (SLR) camera (Canon EOS 1200D, Canon Inc., Tokyo, Japan) was placed facing the sample with focal length 55mm. The distance between the lens and the sample was 50 cm with a viewing angle of 35° to horizontal. The relative positions of the camera and holder was fixed for all samples. The sample was illuminated with two white LED light sources against a white background. 146 images were captured per sample over 50 seconds with constant frequency.

3D Point Cloud Reconstruction. The point cloud reconstruction was implemented with commercial software (Agisoft Photoscan, Agisoft, LLC, St. Petersburg, Russia) utilising the Structure from Motion (SfM) algorithm (Zhang et al. 2016). Due to the high resolution (5184x3456) of each image and high overlap between adjacent images, pre-processing software was developed to automatically crop and rescaled each image to the resolution of 1000x1450, and reduced the number of images by discarding three frames from every four.

3D Image Analysis. The automated point cloud analysis software was developed in C++ with Point Cloud Library (PCL) (Rusu and Cousins 2011). The software is programmed to automatically load all point cloud files in order and process them in a batch by implementing the point cloud segmentation and external quality attributes measurement algorithms.

Point Cloud Segmentation. Each point cloud was first converted from Red Green Blue (RGB) space to Hue Saturation Value (HSV) space. Using arbitrary thresholds on the hue channel, which is defined as the attribute of a visual sensation to one of the perceived colours (Wu and Sun 2013), the point cloud was segmented into calyx, body, achenes and holder.

Orienting Bounding Box (OBB) Fitting. The OBBs was fitted to the segments of holder and the combination the holder and fruit body for the size measurement. The major eigenvectors of the covariance matrix of points in a point cloud define the major axis of its OBB (Ding, Mannan, and Poo 2004). The second axis was determined by calculating the maximum Euclidean distance of the points in the point cloud orthogonal to the major axis. The final axis was orthogonal to both other axes.

Height, Width and Length. An OBB was fitted to the point cloud of the combination of the fruit body and holder segments. The OBB was not fitted directly to the body as its irregular shape often resulted in misidentification of the vertical axis. The height of the combination of fruit body and holder was always the largest dimension, so the magnitude of the OBB major axis was assumed to be equivalent to the height the fruit body and holder model. As the fruit body was always longer and wider than the holder, the second and third dimensions of the OBB represented length and width respectively. The height of the holder was estimated by fitting an OBB to its point cloud and the difference in height between it and the combination of fruit body and holder OBB was assumed to be the height of the fruit. Ratios between the three fruit body dimensions and the height of the holder were multiplied by the true height of the holder to derive berry height, width and length.

Volume. The mesh of the strawberry body was constructed from the point cloud using Poisson Surface Reconstruction (Kazhdan, Bolitho, and Hoppe 2006), which produces an enclosed mesh without any edges or large holes. The mesh volume was calculated by summing the volume of every triangle based pyramid formed from each face of the mesh and the origin of the point cloud (Zhang and Chen 2001).

Calyx Size. The edges of the calyx segment were identified by applying convex hull (Cupec, Nyarko, and Filko 2011), enabling rapid calculation of the maximum Euclidean distance. The ratio of the calyx maximum distance and the height of the holder OBB was multiplied by the true height of the holder to estimate calyx size.

Achene Number. The segmentation of achenes from the point cloud was based on identifying points in the body segment with an arbitrary range in the hue channel of HSV space. A clustering algorithm based the Euclidean distances between points was implemented to group points corresponding to the same achene (Dixon and Brereton 2009) and the number of clusters was counted automatically.

Colour. As hue value in HSV space represents visual sensation of perceived colour (Wu and Sun 2013), the mean intensity of the hue channel of every point in the body segment was calculated for the assessment of berry colour.

Statistics. The concordance correlation coefficient (CCC) (I-Kuei Lin 1989) was used to measure the concordance between manually derived and 3D image derived external fruit quality traits. Additionally, the coefficient of determination (r^2) was calculated to estimate correlation between the sets of values. Statistical analysis was performed using R (R Core Team 2017). Linear models and associated coefficients were derived using the “lm” function, the root mean square error (RMSE) was derived using the “Metrics” package (Hamner 2012) and the CCC was derived using the “Agreement” package (Yu and Lawrence 2012).

Cost-effective Scalable Genotyping

Total Read Count Estimation. A stochastic model was developed to estimate the total number of reads needed to have sufficient coverage of each amplicon such that homologous and homeologous chromosomes could be distinguished. It was assumed that all homologous and homeologous amplicons were distinct and contained at least one marker. It was assumed that all targeted amplicons were present and differed in concentration no more than three-fold (Invitrogen 2008) with a uniform distribution. It was further assumed that the sequencing platform was flooded with an unlimited amount of multiplexed PCR product and that the genotyped amplicon was selected at random with a probability proportional to its concentration within the PCR mix.

This model was implemented until a defined proportion of homologous and homeologous amplicons achieved a minimum read depth, at which point the total number of reads was recorded. This was repeated and the 95th percentile total read count was used as an estimate of the total read count needed to achieve a given read depth per homologous and homeologous amplicon.

Amplicon Design. Amplicon design was heuristically optimised based on arbitrary scores of a number of local and global parameters. The highest scoring sets of amplicons will be utilised in GT-seq as a prediction of the most suitable sets. Markers of known agronomically important traits, such as disease resistance will be included, wherever possible.

Table 2. Factors of importance when deciding amplicons to include in GT-seq

Amplicon Factor	**Scoring criteria**
Amplification across multiple homologous sub-genomes	1 per homologous sub-genome targeted
Multiple SNP targets within the same homologous amplicon	0.5 * (n-1) where n is the nth occurrence of a SNP within the same homologous amplicon
Presence within genes	1 per amplicon start/end predicted to be within CDS of *F. vesca* (Darwish et al. 2015)
Specificity and interaction with other primers	Presence/absence as predicted by MPprimer (Shen et al. 2010)

Results

High-throughput Quantitative Phenotyping

In order to evaluate the measurement of seven external strawberry fruit quality parameters using 3D imaging (hereafter referred to as automated assessment), 100 berries were automatically and manually assessed. Reliable reconstruction could be achieved by taking a minimum of 37 images per berry with 100% successful reconstruction, though the nose of the fruit was often missing due to occlusion from the shooting angle. With the described setup, data capture took approximately 60 seconds, including 10 seconds of operator action per sample. Model reconstruction took approximately 15 minutes and parameter derivation took approximately 50 seconds. Both these operations were fully automated.

In order to validate the 3D reconstruction, the point cloud of the holder segment was manually measured in Meshlab (Cignoni et al. 2008), an open source software for 3D mesh visualisation. Although their absolute sizes were inconsistent, ratios among the height, width and length were the similar to true ratios. As there was no evidence of distortion, the absolute height of the holder was used as a standard for fruit dimension measurements. Moreover, incorporation of the holder point cloud ensured that the vertical axis was always greater than any other axis, allowing the major eigenvector of the point cloud covariance matrix to consistently define the vertical axis.

To validate the measurements, the seven traits were measured on a sample of 100 fruit using both manual and automated assessment (Figure 3). Concordance and correlation was assessed using CCC and r^2 respectively. Good concordance (CCC > 0.9) and correlation (r^2 > 0.9) were found between measurements of fruit dimensions and volume. Weaker concordance (CCC = 0.86) and correlation (r^2 = 0.87) was found between measurements of calyx size, which was possibly due to the soft calyx being moved during assessment. Weak concordance (CCC = 0.67) and correlation (r^2 = 0.77) was found between measurements of achene number, which is possibly due to lack of information gathered regarding the nose of the fruit. Weak correlation (r^2 = 0.68) was found between measurements of colour, with high variance in the manual scores. This was likely due to the variability of colour on each fruit and the subjective nature of the score.

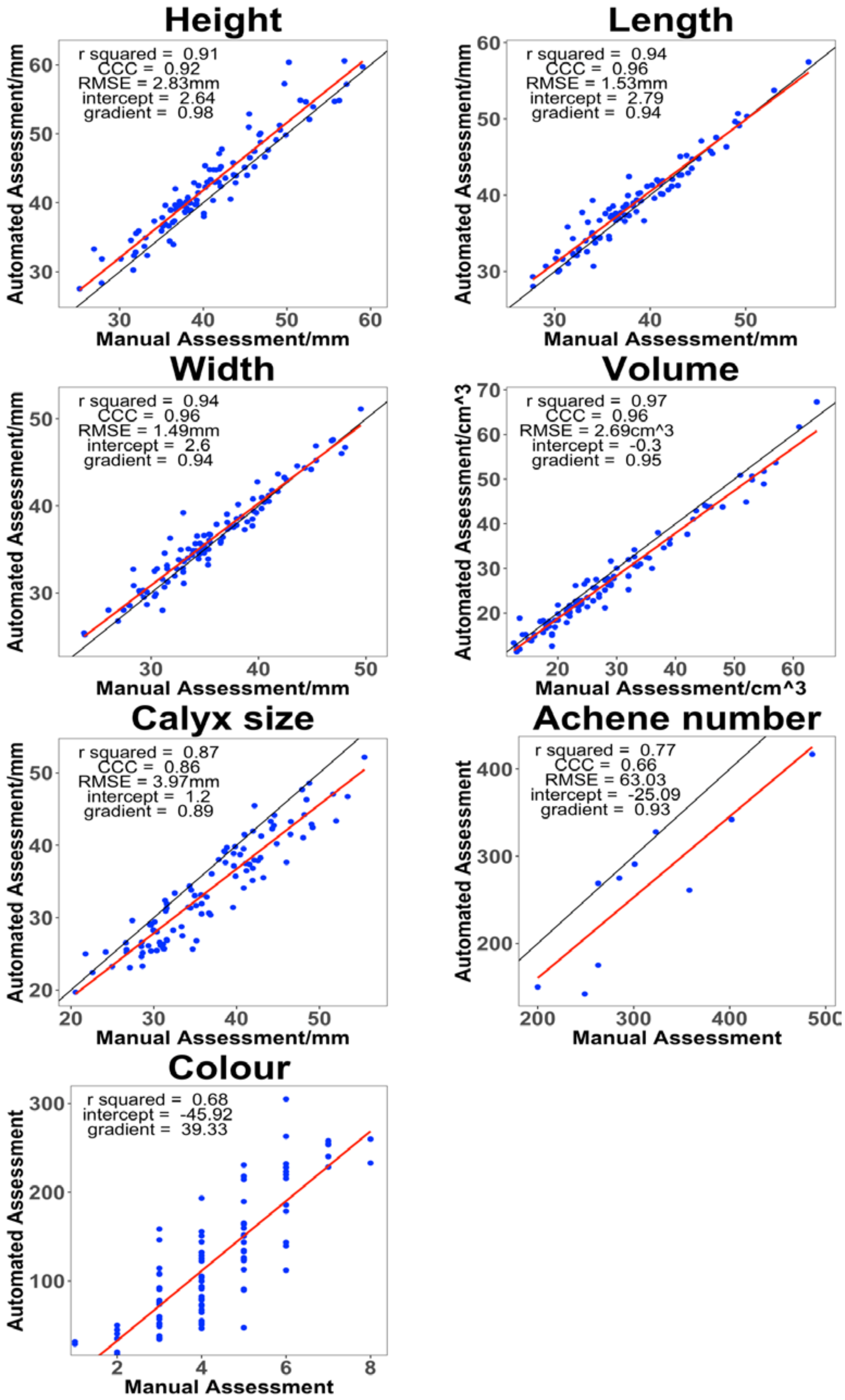


Figure 3. Regression analysis for 7 traits as measured by automated assessment and manual assessment. Sample size = 100 for all measurements, except Achene Number, where sample size = 10. Red lines are least squares linear regression curves and black lines are idealized regression curves (y = x).

Cost-effective Scalable Genotyping

Haploblock estimation suggests that there are approximately 1400 haplolocks within the wider germplasm material. The total read counts required, as estimated using the stochastic model, for 7 amplicons is 620 000 for 95% of homologous and homeologous amplicons achieving 30-fold coverage and 1 200 000 for the equivalent calculation for 60-fold coverage. Design is ongoing for a pilot experiment for 7 multiplexed amplicons.

Discussion

High-throughput Quantitative Phenotyping

Good concordance between manual and automated measurement of calyx size, height, length, width and volume, and promising results for achene number and colour were achieved. It is suggested that qualitative traits of strawberry currently used in breeding can be understood in terms of the measurements generated from this study. For instance, a “long conic” (Mathey et al. 2013) fruit has a large ratio of height to width and measurement of “Cap size” (Mathey et al. 2013) can be defined by the ratio of calyx size to fruit width and length.

With further development, automated assessment could be suitable for integration into existing strawberry breeding programmes, bringing a range of advantages. Firstly, the quantitative, accurate and unbiased measurements would increase the accuracy of selection in strawberry breeding. The precise measurements would be particularly suitable for input into models of genomic selection, which attempt to quantify small effect quantitative trait loci (QTLs) associated with polygenic traits (Gezan et al. 2017; Meuwissen et al. 2001). Secondly, automated assessment has the potential to improve the speed of assessment. The described setup requires approximately ten seconds of human operator time per sample, approximately 20-fold faster than making the equivalent manual measurements. Thirdly, the low cost and wide availability of hardware means that this approach can be easily introduced into existing breeding programmes with minimal capital expenditure.

Measurement error may have arisen from a range of sources. During manual assessment, the axis of measurement was determined by eye, potentially resulting in non-maximal distances or non-orthogonal axes. As the calyx is soft, errors may have been induced in the operation of the callipers. Correlation between measurements of colour may be weak as manual assessment is subjective and it is difficult to assess fruit with uneven colour distributions.

The imaging system can be developed to reduce the duration for data capture by the use of alternative imagers such as scientific cameras or webcams with programmable shutter speeds and resolutions. Use of multiple calibrated cameras to capture information from different viewpoints simultaneously could also be explored to further improve the data quality, particularly from the nose of the fruit and the data capture speed.

As both fruit body and achenes can take a range of colours, our current algorithm of arbitrary hue thresholding is unlikely to be reliable in identifying achenes from a range of cultivars. More sophisticated adaptive or texture based thresholding algorithms would likely improve the cluster identification.

It is believed that more traits could be derived from the gathered dataset. Firstly, algorithms exist that can calculate the surface area of a 3D mesh (Zhang and Chen 2001), which together with reliable achene counts could be used to quantify achene density. Secondly, rotational symmetry could be quantified by considering the distribution of the Euclidean distance of points to the principal axis in 2D slices of the point cloud orthogonal to the principle axis.

Cost-effective Scalable Genotyping

Work is currently being conducted to implement a pilot experiment of GT-seq. This approach has the potential to reduce the cost of genotyping to levels suitable for deployment in GS in commercial strawberry breeding programmes. Additionally, the scalable nature of the system would allow the incorporation of newly identified markers or adjustment of the number of markers genotyped according to the resources available to the programme.

Estimation of the haploblock locations within the breeding population would aid the targeting of amplicons to maximise the efficiency of identification of QTLs associated with economically important phenotypes. As a haploblock is, by definition, co-inherited, a single marker per haploblock is necessary and sufficient to identify the allele(s) of genes associated with the marker.

Currently a parallel project is ongoing at NIAB EMR to generate the sequence of “Redgauntlet”. As part of this project, contigs have been assembled of the octoploid strawberry. *In silico* GT-seq will be conducted using identified optimal primer sets to investigate further problems potentially associated with the technique, such as unexpected repetitive regions or indels. This would also potentially allow for additional SNPs to be identified from the strawberry.

Conclusions

The overall aim of this PhD project is the improvement of GS in plants, with a focus on strawberry as a model organism. To achieve this, three areas were identified which require improvement. The development of a cost-effective genotyping platform is likely to be the most important as the current resources dedicated to breeding programmes suggest that use of microarrays to genotype is prohibitively expensive. Estimates suggest that the cost for this approach will be less than £10 per sample, comparable to conventional selection. A pilot experiment is ongoing to demonstrate the efficacy of this approach.

Of secondary importance is the deployment of a high-throughput, quantitative phenotyping platform. GS is optimal for the detection of polygenic traits where each QTL has a small effect. In order for this to be sensitive and valid, the precision and accuracy of the measured phenotype must be high. Experimental results demonstrate that, with minimal expenditure, an automated phenotyping platform can be implemented to generate highly precise data with good concordance to manual assessment techniques in seven external strawberry phenotypes.

Knowledge and Technology Transfer

Upcoming

Current and future applications of phenotyping for plant breeding, Novi Sad, Serbia (September 2017) – Poster and oral presentation (TBC)

Crops Group Student Symposium, Reading, UK (Nov 2017)

AHDB Studentship Conference, UK (November 2017) – Oral presentation (TBC)

NIAB Student Outreach Event, Histon, UK (November 2017) – Oral and poster presentation on 3D strawberry phenotyping

Attended

4th International Horticultural Conference, East Malling, UK (July 2017) – Oral Presentation on 3D imaging in strawberry; poster presentation on cost-effective genotyping for strawberry breeding

Plant and Animal Genome XXV, San Diego, USA (January 2017) – Received Travel Award from AHDB and GCRI to attend conference

Tuscon Plant Breeding Institute, Tuscon, USA (January 2017) - Received Travel Award from AHDB and GCRI to attend course

AHDB studentship Conference, Stratford, UK (November 2016) – Oral presentation on PhD overview

Soft Fruit Day, East Malling, UK (November 2016) – Poster presentation on PhD overview

Grand Challenges in Plant Pathology, Oxford, UK (September 2016)

Software Carpentry, Norwich, UK (June 2016)

References

Aaby, Kjersti, Dag Ekeberg, and Grete Skrede. 2007. “Characterization of Phenolic Compounds in Strawberry (Fragaria X Ananassa) Fruits by Different HPLC Detectors and Contribution of Individual Compounds to Total Antioxidant Capacity.” *Journal of Agricultural and Food Chemistry* 55(11):4395–4406.

Bassil, Nahla V. et al. 2015. “Development and Preliminary Evaluation of a 90 K Axiom® SNP Array for the Allo-Octoploid Cultivated Strawberry Fragaria × Ananassa.” *BMC Genomics* 16(1):155.

Blasco, J., N. Aleixos, and E. Moltó. 2007. “Computer Vision Detection of Peel Defects in Citrus by Means of a Region Oriented Segmentation Algorithm.” *Journal of Food Engineering* 81(3):535–43.

Boyera, N., I. Galey, and B. a Bernard. 1998. “Effect of Vitamin C and Its Derivatives on Collagen Synthesis and Cross-Linking by Normal Human Fibroblasts.” *International Journal of Cosmetic Science* 20(3):151–58.

Campbell, Nathan R., Stephanie Harmon, and Shawn R. Narum. 2014. “Genotyping-in-Thousands by Sequencing (GT-Seq): A Cost Effective SNP Genotyping Method Based on Custom Amplicon Sequencing.” *Molecular Ecology Resources* 15(4):855–67.

Chandler, Craig, Kevin Folta, Adam Dale, Vance Whitacker, and Mark Herrington. 2012. “Chapter 9 - Strawberry.” Pp. 305–25 in *Fruit Breeding*. Retrieved (https://books.google.com/books?hl=en&lr=&id=ct489cJAUdIC&pgis=1).

Cignoni, P. et al. 2008. “MeshLab: An Open-Source Mesh Processing Tool.” *Sixth Eurographics Italian Chapter Conference* 129–36.

Cupec, Robert, EK Nyarko, and Damir Filko. 2011. “Fast 2.5D Mesh Segmentation to Approximately Convex Surfaces.” *5th European Conference on Mobile Robots* 3–8.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. “Inbreeding in Genome-Wide Selection.” *Journal of Animal Breeding and Genetics* 124(6):369–76. bmed/18076474).

Darrow, George M. 1966. *The Strawberry*. New England Institute for Medical Research. Retrieved (https://specialcollections.nal.usda.gov/speccoll/collectionsguide/darrow/Darrow_TheStrawberry.pdf).

Darwish, Omar, Rachel Shahan, Zhongchi Liu, Janet P. Slovin, and Nadim W. Alkharouf. 2015. “Re-Annotation of the Woodland Strawberry (Fragaria Vesca) Genome.” *BMC Genomics* 16(1):29.

DiMeglio, Laura M., Günter Staudt, Hongrun Yu, and Thomas M. Davis. 2014. “A Phylogenetic Analysis of the Genus Fragaria (Strawberry) Using Intron-Containing Sequence from the ADH-1 Gene.” *PLoS ONE* 9(7):1–12.

Ding, S., M. A. Mannan, and A. N. Poo. 2004. “Oriented Bounding Box and Octree Based Global Interference Detection in 5-Axis Machining of Free-Form Surfaces.” *CAD Computer Aided Design* 36(13):1281–94.

Dixon, Sarah J. and Richard G. Brereton. 2009. “Comparison of Performance of Five Common Classifiers Represented as Boundary Methods: Euclidean Distance to Centroids, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Learning Vector Quantization and Support Vector Machines, as Dependent on.” *Chemometrics and Intelligent Laboratory Systems* 95(1):1–17.

Fonstad, Mark A., James T. Dietrich, Brittany C. Courville, Jennifer L. Jensen, and Patrice E. Carbonneau. 2013. “Topographic Structure from Motion: A New Development in Photogrammetric Measurement.” *Earth Surface Processes and Landforms* 38(4):421–30.

Gezan, Salvador A., Luis F. Osorio, Sujeet Verma, and Vance M. Whitaker. 2017. “An Experimental Validation of Genomic Selection in Octoploid Strawberry.” *Horticulture Research* 4(October 2016):16070.

Gianola, Daniel and Johannes B. C. H. M. Van Kaam. 2008. “Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits.” *Genetics* 178(4):2289–2303.

Gil-Ariza, David Jesus et al. 2009. “Impact of Plant Breeding on the Genetic Diversity of Cultivated Strawberry as Revealed by Expressed Sequence Tag-Derived Simple Sequence Repeat Markers.” *J. Amer. Soc. Hort. Sci.* 134(3):337–47.

Hamner, Ben. 2012. “Metrics: Evaluation Metrics for Machine Learning. R Package Version 0.1.1.”

Hanhineva, K. 2011. “Recent Advances in Strawberry Metabolomics.” *Genes, Genomes and Metabolomics.* Retrieved (http://www.weizmann.ac.il/plants/aharoni/PDFs/b1.pdf).

Heffner, Elliot L., Mark E. Sorrells, and Jean-luc Jannink. 2009. “Genomic Selection for Crop Improvement.” *Crop Science* 49(February):1–12.

Hirakawa, Hideki et al. 2014. “Dissection of the Octoploid Strawberry Genome by Deep Sequencing of the Genomes of Fragaria Species.” *DNA Research* 21(2):169–81.

I-Kuei Lin, Lawrence. 1989. “A Concordance Correlation Coefficient to Evaluate Reproducibility.” *BIOMETRICS* 45:255–68.

Ichijima, K. 1926. “Cytological and Genetic Studies on Fragaria.” *Genetics* 11(6):590–604.

Illa, Eudald et al. 2011. “Comparative Analysis of Rosaceous Genomes and the Reconstruction of a Putative Ancestral Genome for the Family.” *BMC Evolutionary Biology* 11(1):9.

Invitrogen. 2008. “SequalPrep ™ Normalization Plate (96) Kit.” 1(May):4. Retrieved (https://tools.thermofisher.com/content/sfs/manuals/sequalprep_platekit_man.pdf).

Kang, S. P., A. R. East, and F. J. Trujillo. 2008. “Colour Vision System Evaluation of Bicolour Fruit: A Case Study with ‘B74’ Mango.” *Postharvest Biology and Technology* 49(1):77–85.

Karina Gallardo, R. et al. 2012. “An Investigation of Trait Prioritization in Rosaceous Fruit Breeding Programs.” *HortScience* 47(6):771–76.

Kazhdan, Michael, Matthew Bolitho, and Hugues Hoppe. 2006. “Poisson Surface Reconstruction.” *Proceedings of the Symposium on Geometry Processing* 61–70.

Knight, V. H., K. M. Evans, D. W. Simpson, and K. R. Tobutt. 2005. “Report on a Desktop Study to Investigate the Current World Resources in Rosaceous Fruit Breeding Programmes". randd.defra.gov.uk/Document.aspx?Document=HH3817SX_3360_FRP.doc

Liming, Xu and Zhao Yanchao. 2010. “Automated Strawberry Grading System Based on Image Processing.” *Computers and Electronics in Agriculture* 71(SUPPL. 1):32–39.

Määttä-Riihinen, Kaisu R., Afaf Kamal-Eldin, and A.Riitta Törrönen. 2004. “Identification and Quantification of Phenolic Compounds in Berries of Fragaria and Rubus Species (Family Rosaceae).” *Journal of Agricultural and Food Chemistry* 52(20):6178–87.

Mathey, Megan M. et al. 2013. “Large-Scale Standardized Phenotyping of Strawberry in RosBREED.” *Journal of the AAmerican Pomological Society* 67(4):205–16.

Mendoza, F. and J. M. Aguilera. 2004. “Application of Image Analysis for Classification of Ripening Bananas.” *Food Engineering and Physical Properties* 69(9):E471–77.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.” *Genetics* 157(4):1819–29.

Nagata, M., P. M. Bato, M. Mitarai, Cao Qixin, and T. Kitahara. 2000. “Study on Sorting System for Strawberry Using Machine Vision (Part 1). Development of Software for Determining the Direction of Strawberry (Akihime Variety).” *Journal of the Japanese Society of Agricultural Machinery* 62(1):100–110.

Nehra, Narender S. et al. 1990. “Genetic Transformation of Strawberry by Agrobacterium Tumefaciens Using a Leaf Disk Regeneration System.” *Plant Cell Reports* 9(6):293–98.

Ogutu, Joseph O., Torben Schulz-Streeck, and Hans-Peter Piepho. 2012. “Genomic Selection Using Regularized Linear Regression Models: Ridge Regression, Lasso, Elastic Net and Their Extensions.” *BMC Proceedings* 6(Suppl 2):S10.

Paulus, Stefan, Jan Behmann, Anne Katrin Mahlein, Lutz Plümer, and Heiner Kuhlmann. 2014. “Low-Cost 3D Systems: Suitable Tools for Plant Phenotyping.” *Sensors (Switzerland)* 14(2):3001–18.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*.

Rose, Johann C.hristian, Stefan Paulus, and Heiner Kuhlmann. 2015. “Accuracy Analysis of a Multi-View Stereo Approach for Phenotyping of Tomato Plants at the Organ Level.” *Sensors (Basel, Switzerland)* 15(5):9651–65.

Rusu, Radu Bogdan and S. Cousins. 2011. “3D Is Here: Point Cloud Library.” *IEEE International Conference on Robotics and Automation* 1–4.

Sargent, D. J. et al. 2015. “HaploSNP Affinities and Linkage Map Positions Illuminate Subgenome Composition in the Octoploid, Cultivated Strawberry (Fragaria x ananassa).” *Plant Science* 242:140–50.

Schaart, Jan G., Trygve D. Kjellsen, Lisbeth Mehli, and Reidun Heggem. 2011. “Towards the Production of Genetically Modified Strawberries Which Are Acceptable to Consumers Towards the Production of Genetically Modified Strawberries Which Are Acceptable to Consumers.” *Genes, Genomes and Genomics* (September 2016).

Semagn, K., Å. Bjørnstad, and M. N. Ndjiondjop. 2006. “An Overview of Molecular Marker Methods for Plants.” 5(25):2540–68.

Shen, Zhiyong et al. 2010. “MPprimer: A Program for Reliable Multiplex PCR Primer Design.” *BMC Bioinformatics* 11(1):143.

Tennessen, Jacob A., Rajanikanth Govindarajulu, Tia Lynn Ashman, and Aaron Liston. 2014. “Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes Revealed by Dense Targeted Capture Linkage Maps.” *Genome Biology and Evolution* 6(12):3295–3313.

Throop, J. A., D. J. Aneshansley, W. C. Anger, and D. L. Peterson. 2005. “Quality Evaluation of Apples Based on Surface Defects: Development of an Automated Inspection System.” *Postharvest Biology and Technology* 36(3):281–90.

Verma, S. et al. 2017. “Development and Evaluation of the Axiom ® IStraw35 384HT Array for the Allo-Octoploid Cultivated Strawberry *Fragaria* × *Ananassa*.” *Acta Horticulturae* (1156):75–82.

Walker, Paul G., Sandra L. Gordon, Rex M. Brennan, and Robert D. Hancock. 2006. “A High-Throughput Monolithic HPLC Method for Rapid Vitamin C Phenotyping of Berry Fruit.” *Phytochemical Analysis* 17(5):284–90.

Wu, Di and Da Wen Sun. 2013. “Colour Measurements by Computer Vision for Food Quality Control - A Review.” *Trends in Food Science and Technology* 29(1):5–20.

Xu, Yunbi and Jonathan H. Crouch. 2008. “Marker-Assisted Selection in Plant Breeding: From Publications to Practice.” *Crop Science* 48(2):391–407.

Yu, Yue and Lin Lawrence. 2012. *Agreement: Statistical Tools for Measuring Agreement. R Package Version 0.8-1*.

Zhang, Cha and Tsuhan Chen. 2001. “Efficient Feature Extraction for 2D/3D Objects in Mesh Representation.” *Virtual Reality* 1–4.

Zhang, Yu, Poching Teng, Yo Shimizu, Fumiki Hosoi, and Kenji Omasa. 2016. “Estimating 3D Leaf and Stem Shape of Nurserypaprika Plants by a Novel Multi-Camera Photography System.” *Sensors (Switzerland)* 16(6).