Project title:	Next Generation Berries – Implementing Genome-wide Selection Approaches in Strawberry
Project number:	CP163
Project leader:	Richard Harrison, NIAB EMR Michael Shaw, University of Reading
Report:	Annual Report; September 2018
Previous report:	Annual Report, September 2017
Key staff:	Joe Q He, NIAB EMR
Location of project:	NIAB EMR
Industry Representative:	Tom Rogers, Soloberry Ltd., Rumwood Green Farm, Maidstone, UK, ME17 3ND
Date project commenced:	01/10/2016

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.

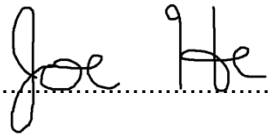
AUTHENTICATION

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 .....  Date .....

[Name]

[Position]

[Organisation]

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

Report authorised by:

[Name]

[Position]

[Organisation]

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

[Name]

[Position]

[Organisation]

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

CONTENTS

GROWER SUMMARY

Headline

Genomic Selection (GS) is likely to increase the genetic gain per unit time in strawberry breeding programmes.

Background

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, disease resistance and cropping season duration. Current goals in strawberry breeding include improvements in maintenance of post-harvest fruit quality, yield, texture and flavour.

Traditionally, crossing is conducted based on identification of desirable traits in parental germplasm material. Offspring from a cross are assessed throughout the growing season and scored on a weighted index of favourable traits. The highest scoring individuals are selected to progress onto further larger scale trials, where additional information, such as yield and picking speed are gathered, 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, making crosses to release of a novel cultivar may take between 7 and 10 years.

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 genetic variants controlling economically important traits tend to be co-inherited with the desirable genetic variant 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) 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. GS also potentially allows control of inbreeding and elimination of certain field experiments.

Summary

Deployment of genomic selection (GS) in strawberry breeding programmes is likely to increase genetic gain per unit time. Three areas for improvement in current GS approaches were identified to make GS more accessible for commercial breeders:

1. High-throughput Quantitative 3D Phenotyping of Strawberry Fruit. 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.

   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 £1000 to set up.

2. Cost-effective scalable genotyping. Currently, genotyping is prohibitively expensive for most commercial strawberry breeders to routinely deploy for GS. Two methods are proposed to reduce the cost of genotyping in strawberry and are being explored in this project. Firstly, in a typical breeding programme, a large population is to be assessed, with all individuals of the same species. Multiplex processing of this population is likely to reduce time, labour and reagents required and thus reduce cost per individual genotyped. Secondly, rational selection of a subset of variants likely to be informative for GS would allow the reduction of genotyping targets with little loss of informative data, thus reducing genotyping costs. A range of rational design parameters will be implemented. Genotyping-in-Thousands and Bait Capture Genotyping will be explored as multiplexing approaches, combined with barcoding and genotype identification by sequencing with Illumina technology.

3. Statistical techniques for Genomic Selection. There are different statistical models in the literature for GS, each with different assumptions, strengths and weaknesses. Phenotypic and genotypic data was collected over 4 years as part of this project and a previous project relating to a biparental mapping population. A range of GS models will be implemented on the data to determine optimal models and a custom model may be designed based on the allo-octoploid strawberry. The models will be assessed for advantages and disadvantages and the most suitable model for GS in strawberries will be determined.

On the basis of these three approaches, GS will be deployed in a commercial strawberry breeding population. Comparisons will be made to GS implemented using the "gold standard" SNP array and selections made by breeders based on conventional selection.

To date, significant progress has been made in developing 3D image capture and cost-effective scalable genotyping. This annual report will focus on cost-effective scalable genotyping; for information relating to 3D image capture, please refer to the previous annual report and the publication "A novel 3D imaging system for strawberry phenotyping"; *Plant Methods* (He, Harrison, and Li 2017).

Financial Benefits

The gold standard Affymetrix IStraw90 Axiom SNP array costs approximately £50 per sample, which is likely too high for commercial deployment for GS. Estimates of the proposed rational library design, coupled with sample multiplexing suggests reasonable power can be achieved at £5 - £10 per sample. This represents a significant saving and is likely to be commercially viable in breeding programmes.

Action Points

None to date regarding cost-effective genotyping. It is hoped that a library of rationally selected variants will be developed in strawberry, which will be optimised for GS by the end of the project. Additionally, a pipeline for development will be created so different libraries can be generated according to the specific goals in breeding programmes.

SCIENCE SECTION

Introduction

Strawberry Physiology

A fully-grown plant is short statured shrub, typically less than 30 cm in height. It comprises of a central dense, woody crown from which petioles, trusses and stolons emerge above ground and roots emerge below ground (Badenes and Byrne 2012).

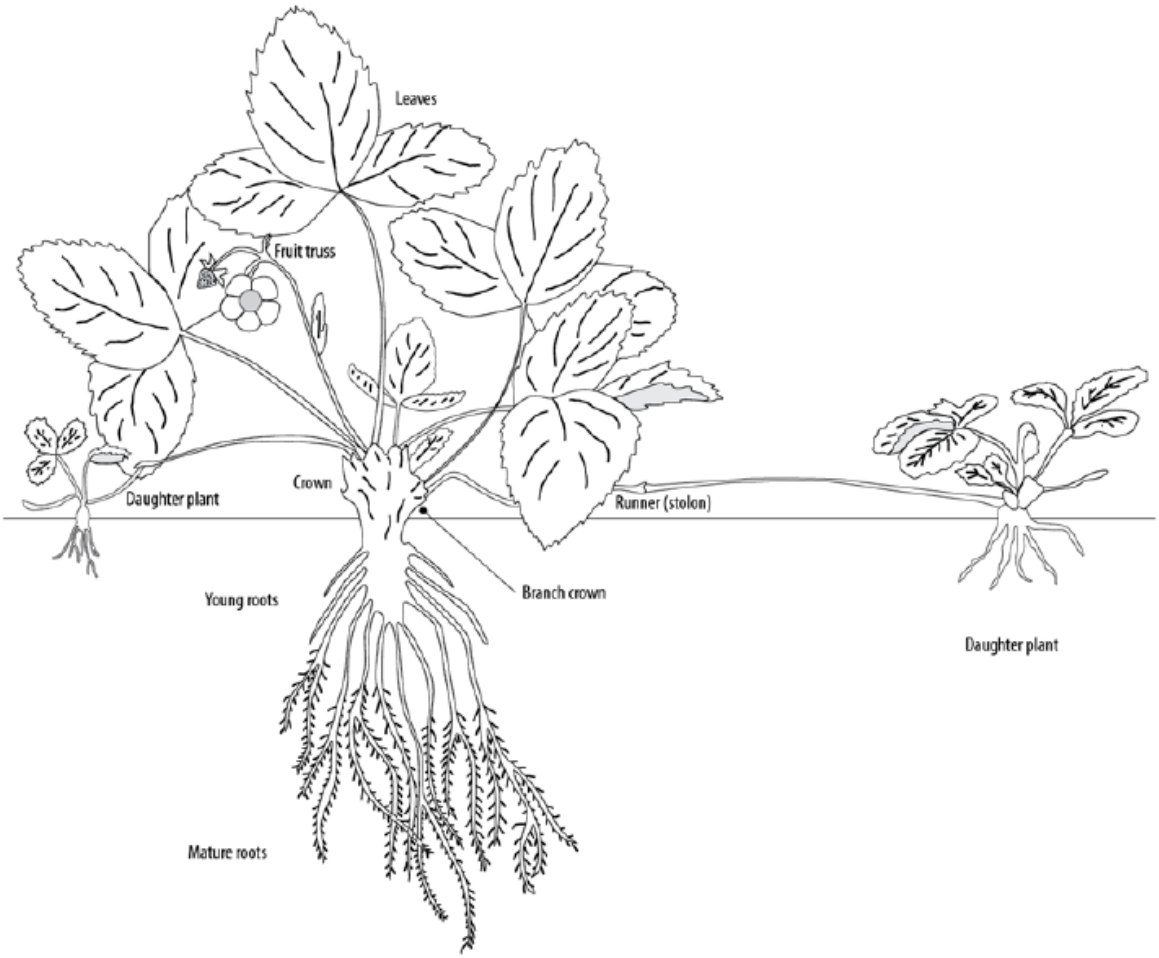


Figure 1. Strawberry morphology (Trejo-Téllez and Gómez-Merino 2014)

Flowers and fruit are borne in clusters at the end of inflorescences. The primary flower is the first to develop and mature, followed by secondary, tertiary, and additional flowers. The earlier developing flowers develops into fruit first and are the largest and most prized by growers (Kirsten 2014).

Flowers are typically hermaphrodites, though pistillate and staminate genotypes do exist. Strong selection has been applied to commercially grown varieties for hermaphrodites to allow ease of crossing of favourable varieties. Flowers comprise of five white petals, possibly

tinged with pink or purple, surrounding a 20 - 35 golden stamen. The stamens surround a conic receptacle, which is covered by up to 60 - 600 pistils (Kirsten 2014).

Fertilisation is achieved when pollen is brushed onto the receptacle and the ovules at the base of each pistil develops into achenes. Both self-fertilisation and cross-fertilisation can occur, with cross-fertilisation conducted mostly by insects, particularly bees. Upon fertilisation, the receptacle swells and enlarges, becoming green and then red as the fruit ripens. Biologically, the achenes are the true fruit of the strawberry, with the surrounding red fleshy part being a modified shoot tip (Shulaev et al. 2011).

Strawberries are considered to be non-climacteric as they do not respond to ethylene with an increased rate of ripening. This has implications with the time point of harvest as strawberries cannot be harvested unripe, stored until needed and artificially ripened (Symons et al. 2012).

The leaves are compound trifoliates and the primary site for photosynthesis and transpiration. They are evergreen in all known species of *Fragaria*, except *F. iinumae*. Older leaves typically die during the winter and younger leaves grow to replace them (Liston, Cronn, and Ashman 2014). Leaf growth is slowed and inhibited by high temperature.

The roots of strawberries are fairly shallow and serve to gather water and nutrients for the plant as well as acting as a physical anchor. Strawberries have two types of roots. Primary roots persist for years, whist feeder roots typically have a lifetime of days to weeks (Kirsten 2014).

Strawberry Utilisation

Strawberries are popularly consumed across the world both fresh and processed. Fresh, they are consumed by themselves, as part of cakes, salads, in breakfast cereals, and dipped in chocolate. Processed, they are consumed as jam, drinks, ice cream, yoghurts and sweets (Badenes and Byrne 2012; Siles et al. 2013). Strawberries are a healthy food choice, containing a range of beneficial chemicals. Firstly, its high vitamin C content (Halvorsen et al. 2002) may protect cells through antioxidant activity (Duarte, Cooke, and Jones 2009) as well as maintenance of bone mass (Gabbay et al. 2010) and support of collagen biosynthesis (Boyera, Galey, and Bernard 1998). Secondly, its dietary fibre content may contribute to reduction of blood sugar and calorie intake through satiation (Giampieri et al. 2012). Thirdly, strawberries contain a range of micronutrients, notably manganese, which improves bone mineral density (Bae and Kim 2008). Additionally, strawberries contain anthocyanins, ellagitannins and other phenolic compounds associated with antimicrobial, antiallergenic and antihypertensive properties (Giampieri et al. 2012).

Clinical and *in vivo* studies have further added evidence to the health benefits of strawberry consumption. Strawberry consumption increases the serum antioxidant concentration significantly and decreases oxidative stress and DNA damage in cells (Cao et al. 1998; Pajk et al. 2006). Berry consumption may also reduce blood pressure and risk of mortality from cardiovascular diseases (Hooper et al. 2008). Strawberry and strawberry extract have also been demonstrated to inhibit transformation of cancer cells (Xue et al. 2001), though other experimental systems suggest a more limited effect (Boivin et al. 2007). Extracts of strawberries induces apoptosis in leukaemia and breast cancer cells *ex vivo* and prolong lifespan of mice bearing breast adenocarcinoma (Somasagara et al. 2012).

Strawberry Taxonomy and Genetics

Strawberry (*Fragaria x ananassa* Duch.) is a eudicotyledon, one of approximately 90,000 species in the subclass *Rosidae*. There are 17 orders within *Rosidae*, including legumes (*Fabales*), brassicas (*Brassacales*) and several nitrogen fixing clades (Sun et al. 2016). The order *Rosales* comprises of 9 families including cannabis (*Cannabaceae*), nettles (*Urticaceae*) and buckthorn (*Rhamnaceae*) (Zhang et al. 2011). Within *Rosales*, *Rosaceae* is a diverse family of 10 tribes, over 90 genera and around 3000 species (Christenhusz and Byng 2016; Feng et al. 2017), including a range of nutritionally and economically important crops. Based mostly on fruit morphological features, four sub-families were originally recognised (Schulze-Menz 1964), but recent phylogenetic analysis suggests that only three monophyletic clades exist. Maximum parsimony and Bayesian analysis on 88 genera using sequence data from nuclear and chloroplastic genes suggest *Rosaceae* is divided into the *Dryadoideae*, *Spiraeoideae* (renamed to *Amygdaloideae* under recommendation from the International Code of Nomenclature (McNeill et al. 2012)) and *Rosoideae* sub-families (Potter et al. 2007). This finding is further supported by transcriptomic and genomic analysis of 124 *Rosaeceae* species covering nearly all the multi-species genera in the family (Xiang et al. 2016).

The subfamily *Dryadoideae* is the smallest of the three subfamilies within *Rosaceae* and believed to be the basal branch (Xiang et al. 2016). It comprises of approximately 10 species in 4 genera, solely consisting of diploids with nitrogen fixing root nodules (Dickinson, Lo, and Talent 2007). *Amygdaloideae* contains a range of widely consumed crops including plum, cherry, apricot, peach and almond. *Rosoideae* contains many important genera, notably raspberries and blackberries *Rubus*, and the ornamental rose (*Rosa*). The subfamily encompasses the tribe *Potentilleae*, which contains the subtribe *Fragariinae*. The *Fragaria*

genus includes over twenty species with a range of ploidies. The Dessert Strawberry thus has a scientific name of *Fragaria* x *ananassa*.

It is clear from early experiments in cytology that strawberries have a base chromosome number of 7, with the dessert strawberry being an octoploid (Ichijima 1926). Early studies into the origins of the polyploid were mostly based on observations of meiotic chromosome pairing. With the advent of novel molecular tools, 43 accessions representing 14 species were genotyped at the nuclear ITS region and the chloroplastic *trnL* region. Maximum parsimony was used to cluster the accessions, suggesting that the *F. vesca* and *F. nubicola* were likely donors of genomic material (Potter, Luby, and Harrison 2000). 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). Additional donors are unclear (DiMeglio et al. 2014; Tennessen et al. 2014; Vining et al. 2017).

For breeding strategies, it is desirable to understand the method of inheritance for strawberry. Low resolution genetic maps have provided some evidence of polysomic inheritance, but more sensitive experiments to date have only found evidence of disomic inheritance. Analysis of 4 microsatellites in *F. virginiana* was found to be consistent with disomic Mendelian inheritance (Ashley et al. 2003). A genetic map for *F. ananassa* was generated using 148 molecular markers. In the 42 linkage groups where markers in both coupling and repulsion phase was found, there was a 1:1 ratio of coupling and repulsion phase markers in resulting recombinant progeny, consistent with disomic inheritance (Rousseau-Gueutin et al. 2008).

Strawberry Breeding and Genetic Modification

Breeders have crossed plants exhibiting favourable characteristics for thousands of years. The resulting sexual recombination generates variation, upon which the breeder makes selections for individuals exhibiting combinations of desirable traits. Organisation of breeders into breeding programmes allows for sharing of resources and an increase of labour to achieve specific goals. Breeding programmes share broadly similar aims of improving fruit quality, pathogen resistance and productivity (Yue et al. 2012). Funding comes from a mixture of sources including governmental, private and royalties on intellectual property (Knight et al. 2005).

Over the past 200 years of breeding, a range of strawberry traits have been improved, including fruit size, marketable yield, pathogen resistance and production season length (Badenes and Byrne 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). Usually, a new cultivar takes 7 years to develop from breeding to commercial release, but may take up to 20 years. Breeding is based on crossing germplasm material with agronomically important traits, such as high yield, and selecting for the most favourable offspring. These offspring are then trialled over several years, usually under different environmental conditions to confirm these traits before release (Badenes and Byrne 2012).

The development of modern gene theory and transformation technologies has allowed for the insertion of genes from other domains into strawberries as a means to generate variation. This targeted insertion is particularly attractive for the pyramiding of monogenic traits (or traits with a small number of QTLs) rapidly and without compromising existing characteristics (Passey, Barrett, and James 2003).

Transformation of strawberries to resist a range of pests and diseases have seen success at research level and shows promise in reducing chemical controls required (Qin et al. 2008). Insertion of the cowpea trypsin inhibitor into strawberries resulted with plants having up to 362% greater root weight compared to the control when exposed to the vine weevil *Otiorhynchus sulcatus* (Graham, Gordon, and McNicol 1997). Transformation of a chitinase into strawberry resulted in plants that were significantly less susceptible to *Verticillium dahliae* (Chalavi, Tabaeizadeh, and Thibodeau 2003). More recently, strawberries transformed with the *Arabidopsis thaliana* NPR1 gene shows increased resistance to anthracnose, powdery mildew and angular leaf spot (Silva et al. 2015).

Strawberries have also been genetically engineered to be resistant to abiotic stresses. Using *Agrobacterium* mediated transformation, antifreeze protein from fish has been inserted into strawberry, though no experimental evidence of cold resistance was presented (Khammuang et al. 2005). Transformation of the wheat acidic dehydrin gene into strawberry resulted in plants that were able to resist ion leakage at temperatures 5°C lower than the untransformed control (Houde et al. 2004).

Despite the promise of utilising genetic modification in strawberry breeding, there are no known large scale commercially available genetically engineered strawberries. One major obstacle is the reluctance of the public to accept consumption of transgenic crops (Schaart et al. 2011). Additionally, there are issues with low efficiencies of transformation, difficulty identifying and isolating genes for transferring into strawberry and variable expression after transformation (Qin et al. 2008).

Genomic Selection

Genomic selection (GS) is an advanced breeding technique that utilises a densely genotyped and phenotyped training population, from which associations are made relating the magnitude and direction of quantitative trait loci (QTLs) associated with agronomically important traits (Meuwissen, Hayes, Goddard, et al. 2001). GS has been successfully deployed in a range of crops, including grape (Viana et al. 2016), wheat (Heffner, Jannink, and Sorrells 2011; Thavamanikumar, Dolferus, and Thumma 2015), maize (Shikha et al. 2017) and strawberry (Gezan et al. 2017). A statistical model has to be 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).

There are a number of benefits associated with GS. Firstly, assuming that there are sufficient markers available, GS has been demonstrated to generate greater prediction accuracy than conventional selection. This is largely due to the approach ignoring the variable and non-hereditary environment (De Los Campos et al. 2009). Secondly, GS allows the regression of a genotype onto multiple individuals, allowing increase in power of detection of small effects. Moreover, this allows selection on rare, expensive, or otherwise difficult to phenotype traits. For example, harsh drought in the summer is a rare event, expected to occur only every 10 years. It is possible that this does not occur during the entire breeding cycle of a novel cultivar, and thus no phenotypic information about crop performance can be gathered. Under conventional selection, no predictions can be made regarding plant performance under harsh drought, but under genomic selection, predictions can be made based on the genotype of the novel cultivar and data gathered from genotypes from all instances of harsh drought (Heffner et al. 2010; Peace et al. 2017). Perhaps most importantly, GS allows for the reduction of the duration of the breeding cycle as plants can be genotyped and selected for in the seedling phase. Additionally, GS allows for elimination of some field experiments and better planning of crosses by providing information on relatedness (Gezan et al. 2017). GS performs better than marker assisted breeding (MAS) because it simultaneously estimates the effect size for all markers, and thus is less likely to overestimate the "'significant" and underestimate the "insignificant" MAS markers.

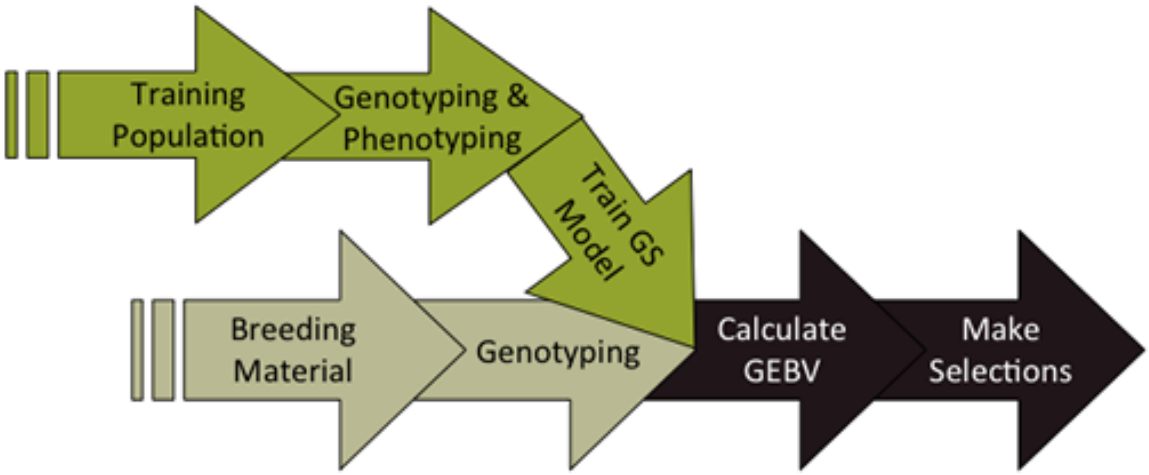


Figure 2. Schematic of genomic selection (Heffner et al. 2009)

In order for genomic selection to be viable for commercial implementation, it must be more cost effective than the currently employed method. Tools have been developed to perform cost-benefit analysis and to optimise resource allocation for implementation of genomic selection. Analysis with DeltaGen on a forage breeding population, for example, suggests that genotyping for genomic selection approximately doubles the cost, but also nearly doubles the increase in genetic gain per cycle when compared to selection without genotypic information (Jahufer and Luo 2018). Genomic selection has been experimentally implemented in strawberry utilising the IStraw90 Axiom SNP array (Bassil et al. 2015) to generate genotypic information. High prediction accuracies were observed for a range of agronomically important traits, but it was acknowledged that the cost of the SNP array was likely too high for commercial deployment (Gezan et al. 2017).

Cost Effective Genotyping

In order to minimise the cost of genotyping, two strategies will be employed: multiplexing of individuals for genotyping saves reagent used and time required; and rational selection of markers minimises the number of loci targeted, whilst maximising power to detect agronomically important QTLs. Detection of markers will utilise Illumina short sequencing technology as it is low cost per base sequenced and allows open scalability, enabling markers to be added or removed as resources allow.

One potential method of multiplexing large populations of individuals for a small (50 - 500) panels of SNPs Genotyping-in-Thousands (GT-seq). It utilises two PCR steps to add Illumina sequencing primer sites, Illumina capture sites and unique barcodes to each individual, whilst also amplifying the targeted region. Then the DNA from each individual is normalised and

pooled into a single test tube. Next, the sample is sequenced using Illumina short read technology to identify genotypes. Finally, a bioinformatics pipeline is employed to resolve the reads into individuals and ratios of genotypes taken to determine the genotype at each loci (Campbell, Harmon, and Narum 2014; **Figure 3**).

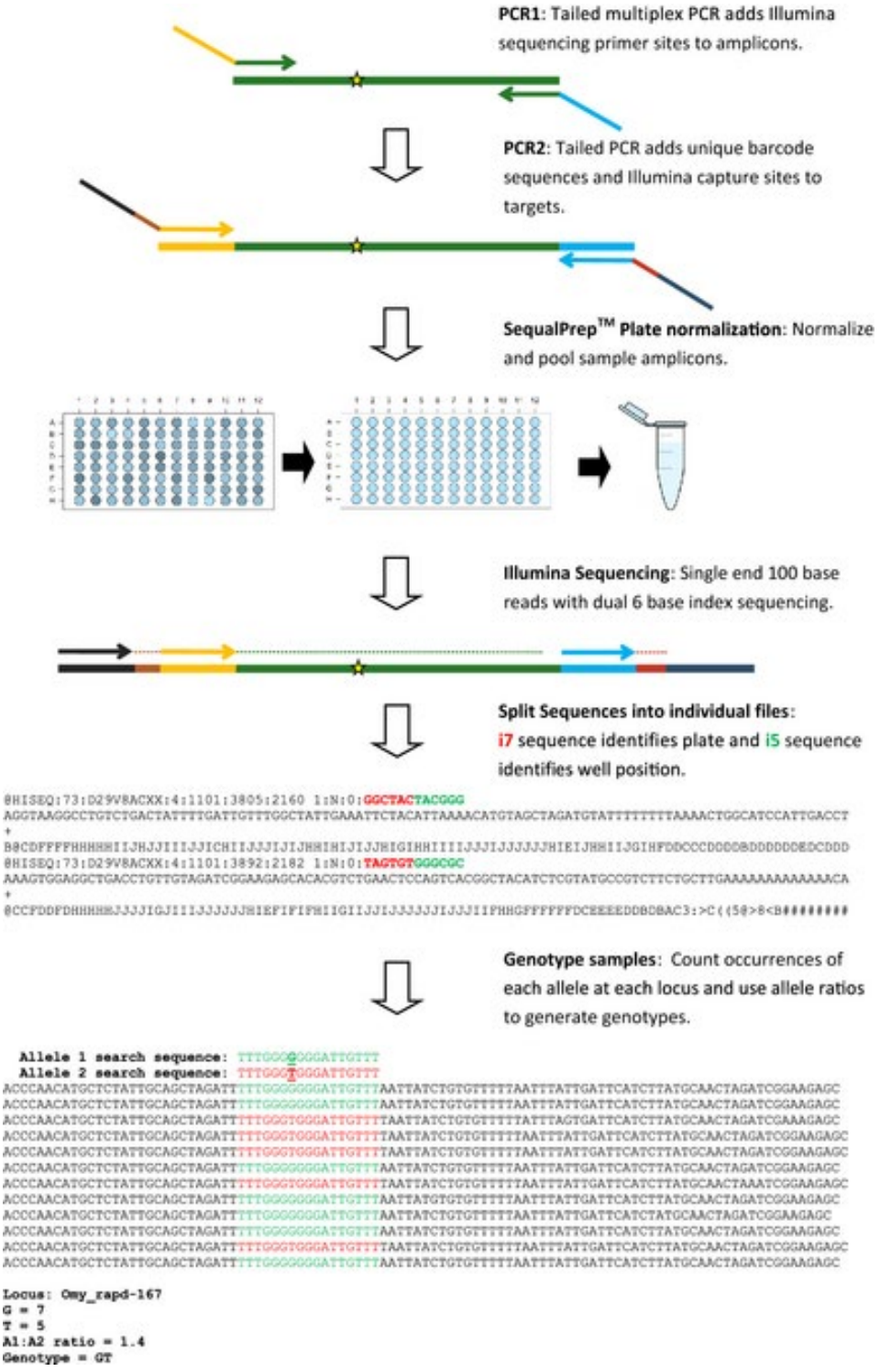


Figure 3. Genotyping-in-Thousands process (Campbell et al. 2014)

An alternative method of multiplexing individuals utilises targeted bait capture (Samorodnitsky et al. 2015). Using this approach, custom markers are utilised to enrich genomic DNA for target loci, before genotype resolution by sequencing (**Figure 4**). Both methods will be trialled to assess feasibility for use in genotyping for GS.

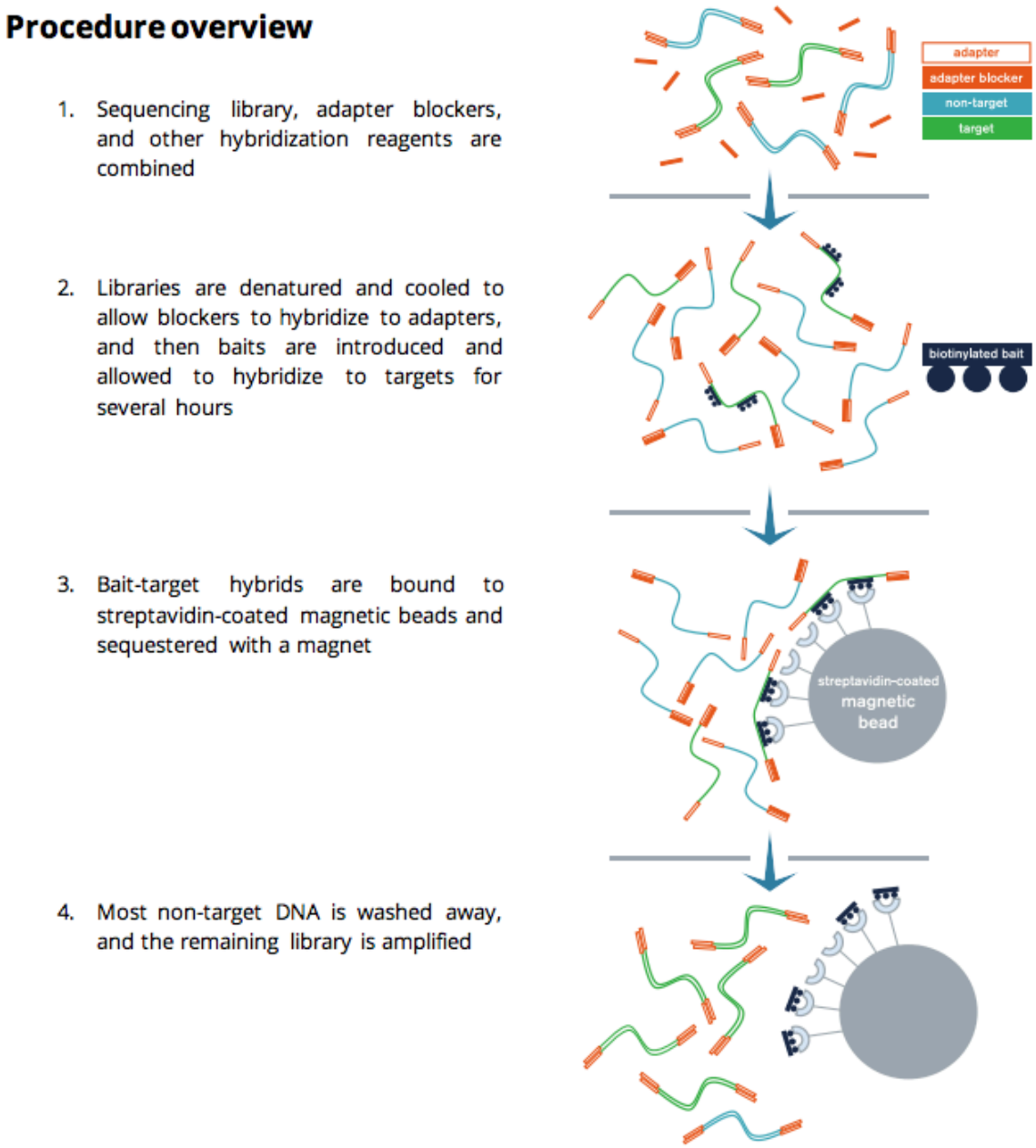


Figure 4. Bait Capture for enrichment of target loci (www.arborbiosci.com/mybaits-manual, Hybridization Capture for Targeted NGS V4.0.1 April 2018)

In order to identify a subset of the most informative markers, several criteria were identified for selection. Amplicons were designed to be approximately 450bp in length. The maximum

length to be sequenced using the Illumina Miseq Reagent Kit v2 (2 x 250) (Illumina, Cambridge, UK) is 500bp, but it is known that the ends of the paired-end reads typically suffers from poorer accuracy (Schirmer et al. 2015). The allowance of a 50bp overlap enables the region to be sequenced at a higher depth to compensate for this.

Firstly, it was reasoned that an amplicon with high numbers of SNPs, particularly if they span homeologous chromosomes is likely to be more informative. Sequencing of these amplicons would allow for information to be gathered from a greater number of sites, without increasing the cost. Data from within the same subgenome may be under strong linkage equilibrium, as they are physically close together, and is thus less valuable than data from across homeologous chromosomes. To reflect this, each additional SNP increased the probability of being selected. However, increasing numbers of markers within the same subgenome had increasing penalties to score gain, so that an amplicon with SNPs in more subgenomes always scored higher than amplicons with high numbers of SNPs in the same subgenome. Additionally, it was reasoned that markers that were physically close together were less likely to be informative as they were likely to be in strong linkage disequilibrium. Consequently, amplicons that were close together were penalised in the selection procedure.

Secondly, it is believed that purifying selection is disproportionately active in the coding regions of genomes. Retrotransposons in *Arabidopsis* are disproportionately found in the heterochromatic pericentromeric regions of the genome (Pereira 2004). There is also evidence that coding regions within strawberry is under purifying selection. Analysis of the bZIP family of genes within strawberries, peach and apples shows that the ratio of synonymous to non-synonymous mutations (K_a/K_s) is typically < 0.4 indicating strong purifying selection (Meng et al. 2015). Interestingly, the K_a/K_s ratio in strawberry chloroplasts vary between 0.1 and 0.2 in multiple regions when compared with other rosaceous crops, suggesting strong purifying selection is also present in the chloroplasts. Unknown variants are likely to affect binding of primers and thus affect amplification efficiency. Over 30 million potential variants were discovered in the design of the “gold standard” SNP array for strawberries (Bassil et al. 2015), indicating that many more polymorphisms are likely to be present than our pipeline is able to discover. In order to mitigate this impact, amplicons that were in coding regions were assigned a higher relative probability of selection.

Thirdly, Genome-wide Analysis Studies (GWAS) have been conducted on strawberry and have identified QTLs associated with agronomically important traits, including *Fusarium wilt* resistance (Pincot et al. 2018), and fruit aroma (Zorrilla-Fontanesi et al. 2012). It is expected that these QTLs are likely to be disproportionately informative if included within a GS marker panel as there is already evidence that they are associated with agronomically important

traits, so markers associated with some disease resistance markers were included in the amplicon library.

Fourthly, some models suggest that maximisation of the long term selection values requires an index which increases the value of rare beneficial alleles (Goddard 2009). This allows for the mean genetic value to increase more, if a rare allele is beneficial, by selecting for its proliferation in the population. However, in an experimental implementation of GS in strawberry, markers with a minor allele frequency (MAF) of > 0.05 were removed, presumably as confidence in them being true markers is lower (Gezan et al. 2017). For the bait capture library to be developed, both common and rare alleles will be included, to determine if they are reliable and if they contribute to larger genetic gains in the long term.

Materials and methods

Genomic Resources

For the design of Genotyping-in-Thousands sequencing (GT-seq), the *F. vesca* genome v1.1 (Shulaev et al. 2011) and the v2.0.1 annotation was utilised (Darwish et al. 2015); for the design of bait capture, the *F. vesca* genome v4.0 was utilised (Edger et al. 2018). The current best assembly of the dessert strawberry is being used as the reference genome for the design of bait capture. This assembly is highly fragmented, but it is likely to be useful in the design of the bait library as it represents contiguous regions within the target organism.

Ab initio gene annotation was performed on the best assembly of the dessert strawberry using AUGUSTUS (Stanke and Morgenstern 2005) with *A. thaliana* as the reference standard.

Variant Identification

Work relating to variant discovery is ongoing. A panel of European strawberry breeding companies were asked to provide samples of strawberries from their breeding programmes which they felt represented the genetic diversity of their programmes or were important donors of genetic material. 202 strawberries were sequenced using the Illumina sequencing platform for variant discovery. BWA (Li and Durbin 2009) will be utilised to align the sequences to the best available genome and GATK (Depristo et al. 2011) will be utilised to identify variants in the alignments.

Rational Design of Genotyping targets

No genome of the dessert strawberry was available when the design of GT-seq was conducted, so the *F. vesca* genome was used to aid design. The best available genome of strawberry was used in rational design for bait capture. All possible windows of 450 base pairs (bp) were selected where the first bp of the window was a variant.

For GT-seq: SNPs from the IStraw 90K Axiom SNP array were used as the pool of variants to select from. Windows were selected from the *F. vesca* genome and scored based on containing multiple SNPs (including across homeologues) and inclusion of 42 markers associated with Verticillum Wilt (Cockerton et al. 2018) and 12 markers associated with Crown Rot (Nellist *et al.*, in press). Sequential addition of amplicons in an "evolutionary" style algorithm was then implemented to identify multiplexes suitable for GT-seq, with the multiplex design program MPprimer (Shen et al. 2010). Briefly, if MPprimer successfully generates a twoplex associated with the amplicons, a third amplicon is added and the threeplex inputted into MPprimer; else the input amplicons are discarded and another twoplex is generated. If the threeplex is successful, then a fourth and subsequent primer pair is designed, until the desired level of multiplexing is achieved.

For bait capture: variants from the variant identification step will be used as the pool for selection. Selection will be based on coverage across homologous and homeologous chromosomes, incidence of nearby windows, tiling haplotypes for greater power, minor allele frequencies of markers, markers known to be associated with QTL, linkage disequilibrium estimations and gene location predictions.

Read Depth and Cost Simulation

In order to estimate the total number of reads necessary to have sufficient power to resolve genotypes, a stochastic *in silico* simulation of the sequencing stage of GT-seq was implemented to model the number of reads required per genotype. The simulation requires input population size, ploidy, number of amplicons per individual, read depth required, an acceptable failure rate, and a probability for achieving these conditions. The simulation assumes an approximately equimolar (no more than three-fold variation, Invitrogen SequelPrep Normalisation Plate manufacturer instructions, ThermoFisher, CA, USA) concentration of sample is to be sequenced. The variance of concentrations of each sample have a random flat distribution. It is assumed that the amount of DNA is large enough that sequestration of DNA for sequencing negligibly impacts the concentration of that sequence.

Sequencing is independent and random, proportional to the concentration of the given haplotype.

Plant Material

Young leaves from *F. ananassa* cultivars "Redgauntlet", "Hapil" and "Emily" were picked and stored in darkness overnight prior to DNA extraction to prevent build-up of metabolites, which may degrade DNA. Extraction was performed with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The final concentrations of DNA obtained was 507 ng μl^{-1}, 621 ng μl^{-1}, and 262 ng μl^{-1} respectively[1].

Thermocycle conditions

Primers were ordered from Integrated DNA Technologies (IDT, IA, USA) with standard desalting purification for all experiments. The final mix for thermocycling included forward and reverse primers at 1 μM each, genomic template at 1 ng μl^{-1} and Kapa Hifi HotStart ReadyMix (2x) Mix (Kapa Biosystems, Wilmington, MA, USA) according to the manufacturer's guidelines. Dilution in all cases was with DNAse free water.

The enzyme was heat activated at 95°C for 3 minutes. 30 thermocycles were them implemented: 95°C for 30 seconds; annealing temperature for 30 seconds; 72°C for 30 seconds. The reaction was held at 72°C for 5 minutes to ensure extension completed after the thermocycles and then chilled to 4°C until analysis. The touchdown procedure followed the same conditions, except the annealing temperature was decreased by 0.2°C per cycle.

Gel Electrophoresis and Imaging

A 2% agarose gel was prepared using standard methods, spiked with Gel Red. Each sample was mixed with DNA loading buffer according to the manufacturer's instructions and loaded into the wells. A DNA ladder (GeneRuler 1 kb Plus DNA Ladder, ThermoFisher, Waltham, MA, USA) was included as a known standard to evaluate sizes. Electrophoresis was conducted by immersion of the loaded gel in Tris, acetic acid and EDTA (TAE) buffer, maintaining potential difference across the gel at 200mV, until separation was achieved.

[1] Extraction performed by C. Nellist

Pictures of the gel was captured using an imager (Gel Doc XR+, BioRad, Hercules, CA, USA) and adjusted manually for contrast.

Results

Read Depth and Cost Estimation

Based on the stochastic model, assuming a 95% chance of achieving 10 fold coverage in 90% of samples in 192 amplicons in 2068 octoploids, 57123468 reads would be needed. Analysis took approximately 10 hours (MacBook Pro, Intel core i7 x 4 @ 3.1GHz, macOS High Sierra, Apple Inc, CA, USA). Preliminary estimation using the Genotyping-in-Thousands (GT-seq) analysis method would cost approximately £9 per sample.

A 6plex was designed for GT-seq

Based on the described method, a 6plex was designed for GT-seq (**Table 1**). The stringency of conditions is relaxed relative to those typically used for uniplexes to increase the rate of successful multiplexes generated. Illumina barcodes and amplification sequences were included according to manufacturer's instructions.

Table 1. 6Plex Primers for GT-seq

Name	ID	Sequence	Target Length
407734927	JH_01_F	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGATGGGCATGTTGGAGCAGTGGC	329
	JH_01_R	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCCGTGCAGCAGTTAAGCCAGCA	
222622454	JH_02_F	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGTGGAGCCCCAGCCTGAGAAGAG	163
	JH_02_R	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGTGGGCCAAAAGGGTCTGAGGGAA	
720416880	JH_03_F	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGCGGAACCGGTGGTAGCGAAAT	260
	JH_03_R	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGCAGCAGACCTGTGTTGCAGCGA	
520595999	JH_04_F	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGAGGCCCCTTCAACAAAGGCTCC	426
	JH_04_R	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGAAGGCTCTCCGCTCCAGCAAGT	
309523186	JH_05_F	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGGTTGAAGACCGTAGCCCTCGT	207
	JH_05_R	GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAGTTTTCGCCCAAGCCCTCTTAGC	
721334717	JH_06_F	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGTGAGCGCAGCAGCAGGAATGA	514
	JH_06_R	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGCCAAGCCGAAGGCATCAAGGT	

Multiplex primers interact unexpectedly

Uniplexes of each of the primer pairs generates the expected sized fragments (**Figure 5**, right) with higher temperatures having greater specificity. 64°C annealing temperature had

the highest specificity (results not shown) so that temperature was used in subsequent experiments.

When multiplexed, there is uneven amplification of different primers with one band dominating the reaction (likely JH06). Additionally, there appears to be a range of off target fusion products. In order to identify if any single primer pair in the reaction causes the formation of the fusion products, all combinations of 5plexes were implemented (**Figure 5**, left). Fusion bands are observed and missing bands do not correspond to the expected missing band.

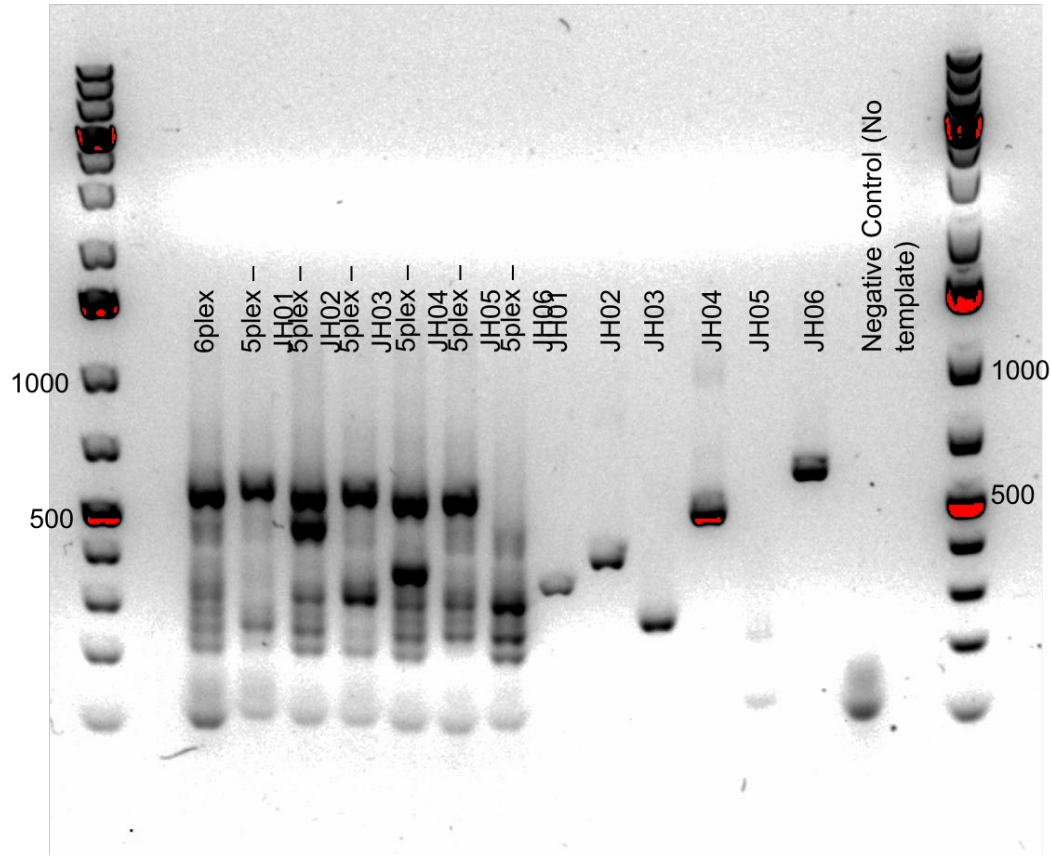


Figure 5. 6Plex products from GT-seq

Touchdown PCR does not reduce multiplex interactions

To investigate if the temperature used influenced the formation of side products in the multiplex reaction, a series of touchdown PCR reactions was conducted with a range of temperatures. Under all attempted conditions, side products appeared to form with a JH06 as a dominant band, with complex fusion fragments present (**Figure 6**).

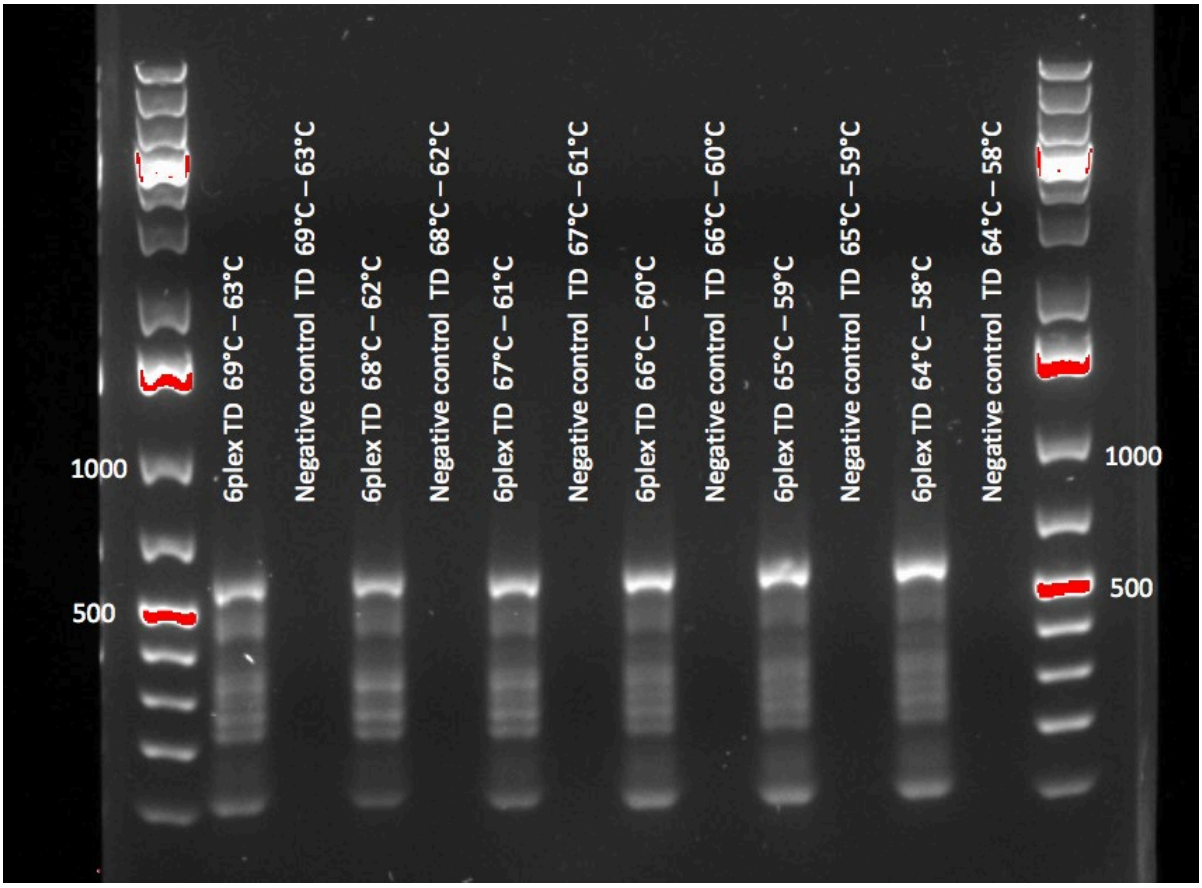


Figure 6. Touchdown PCR of 6plexes for GT-seq

Discussion

GT-seq Genomic Resources

A key assumption for GT-seq was that each of the subgenomes in the octoploid *F. ananassa* was similar to *F. vesca*. This was done as there was no reliable assembly of the octoploid genome and GT-seq required base pair resolution in order to maximise its value for GS. There is strong evidence that *F. vesca* contributed genetic material to one of the subgenomes, with two other subgenomes being perhaps derived from *F. innumae*.

This approach is likely to be biased in favour of discovering SNPs from the A subgenome (the subgenome most closely related to *F. vesca*). This may not be a significant problem for GS as GS models the plant as a black box with no models of mechanism of function inferred. Selection using markers biased toward one subgenome will generate a stronger response in that subgenome so if significant variation is present in other subgenomes, or if QTL for an agronomic trait(s) are primarily in other subgenomes, then GS would be ineffective for these traits.

GT-seq Multiplex Design

It is not necessary to perform rational design of primers for GT-seq, though in these cases, several amplicons may dominate the PCR reaction and removal of their primers may be necessary (Campbell et al. 2014; Onda et al. 2018). In this report, a novel rational design method is presented with the aim of generating genotypes suitable for GS in strawberry. Particular attention is paid to ensure that the genotyping is cost effective, and maximises the information gained per resolved genotype. It was found that the GT-seq method described was insufficient to produce high level multiplexes suitable for genotyping in GS. Efforts are underway to utilise bait capture as an alternative method to rationally select markers for GS.

Even amplification may not have been achieved due to several reasons. Firstly, the inclusion of Illumina barcodes and amplification sequences were not tested using a multiplex design system as they are essential for Illumina sequencing, and could affect the amplification efficiency. Secondly, the design process utilised *F. vesca*, which may be distantly related to *F. ananassa*, such that amplification of ectopic regions occurs. However, this is a less likely hypothesis, as the uniplexes all appeared to amplify single targets as expected. Finally, it is possible that MPprimer is unable to operate effectively with the *F. vesca* genome as it was designed for use with the human genome. It is noted that MPprimer had not tested it on as many samples as would be required of GS (Shen et al. 2010). Moreover, MPprimer utilises some outdated programs, which may contribute to reduction efficacy.

Conclusions

The rational design process ought to increase the power per locus genotyped for GS and thus reduce the costs of genotyping. Further work is needed to ensure the genotyping resolution is successful.

One advantage of this system is that it is scalable. In the future, if additional QTLs associated with genes of agronomic importance are discovered, they can be included into the model. Assuming that they are suitable for multiplexing with the existing library, primers can simply be added to the PCR mixture without additional steps, unlike a SNP array. Moreover, the scalability of this system allows customisation and optimisation to different breeding programmes. A program with less resources may utilise a smaller library.

It is significant that the cost model calculated here is significantly higher than reported of $3.98 (Campbell et al. 2014). This may be due to inflation and the conductance of the

experiment using reagents purchased in USA. Nonetheless, the cost is significantly lower than using the “gold standard” Istraw90 or Istraw35 SNP arrays, and is probably close to an acceptable cost for genotyping in a commercial deployment of GS.

Further work will be on utilisation of bait capture as a method for enrichment of markers for cost-effective genotyping. This method utilises complementary RNA probes to enrich a sample for markers that rationally are likely to be suitable for GS. This approach potentially allows for more loci to be targeted than GT-seq at similar costs.

Additionally, marker discovery will be performed on 202 strawberry accessions, representing the diversity of germplasm material in European breeding programmes. Based on this marker set and the current best assembly of the dessert strawberry, rational selection of markers for GS will be implemented and validated. The results of cost-effective GS will be compared with GS utilising the “gold standard” genotyping array and conventional selection approaches.

Knowledge and Technology Transfer

AHDB Student Industry Visit (July 2018)

Soft Fruit walk, Kent, UK (June 2018)

AHDB Studentship Conference, UK (November 2017) – Poster Presentation on Genotyping-in-Thousands as a cost-effective method of genotyping strawberry

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

Current and future applications of phenotyping for plant breeding, Novi Sad, Serbia (September 2017) – Oral and poster presentation on 3D strawberry imaging

References

Ashley, M. V. et al. 2003. “High Variability and Disomic Segregation of Microsatellites in the Octoploid Fragaria Virginiana Mill. (Rosaceae).” *Theoretical and Applied Genetics* 107(7):1201–7.

Badenes, Marisa Luisa and David H. Byrne. 2012. “Chapter 9: Strawberry.” Pp. 305–25 in *Fruit Breeding*. Retrieved (https://books.google.com/books?hl=en&lr=&id=ct489cJAUdIC&pgis=1).

Bae, Yun Jung and Mi Hyun Kim. 2008. “Manganese Supplementation Improves Mineral

Density of the Spine and Femur and Serum Osteocalcin in Rats." *Biological Trace Element Research* 124(1):28–34.

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.

Boivin, Dominique, Mélanie Blanchette, Stéphane Barrette, Albert Moghrabi, and Richard Béliveau. 2007. "Inhibition of Cancer Cell Proliferation and Suppression of TNF-Induced Activation of NFκB by Edible Berry Juice." *Anticancer Research* 27(2):937–48.

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. Retrieved (http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12357/abstract%5Cnhttp://onlinelibrary.wiley.com/store/10.1111/1755-0998.12357/asset/men12357.pdf?v=1&t=i3fllimc&s=86e9bd60f7bb64636eaeda023f52ca686fe60b46).

Cao, G., R. M. Russell, N. Lischner, and R. L. Prior. 1998. "Serum Antioxidant Capacity Is Increased by Consumption of Strawberries, Spinach, Red Wine or Vitamin C in Elderly Women." *The Journal of Nutrition* 128(12):2383–90.

Chalavi, V., Z. Tabaeizadeh, and P. Thibodeau. 2003. "Enhanced Resistance to Verticillium Dahliae in Transgenic Strawberry Plants Expressing a Lycopersicon Chilense Chitinase Gene." *Journal of the American Society for Horticultural Science* 128(5):747–53.

Christenhusz, Maarten J. M. and James W. Byng. 2016. "The Number of Known Plants Species in the World and Its Annual Increase." *Phytotaxa* 261(3):201–17.

Cockerton, Helen M. et al. 2018. "Identification of Powdery Mildew Resistance QTL in Strawberry (Fragaria × Ananassa)." *Theoretical and Applied Genetics*. Retrieved (http://link.springer.com/10.1007/s00122-018-3128-0).

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. Retrieved (http://www.biomedcentral.com/1471-2164/16/29/abstract%5Cnhttp://www.biomedcentral.com/1471-

2164/16/29%5Cnhttp://www.biomedcentral.com/content/pdf/s12864-015-1221-1.pdf).

Depristo, Mark A. et al. 2011. “A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data.” *Nature Genetics* 43(5):491–501.

Dickinson, T. A., E. Lo, and N. Talent. 2007. “Polyploidy, Reproductive Biology, and Rosaceae: Understanding Evolution and Making Classifications.” *Plant Systematics and Evolution* 266(1–2):59–78.

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.

Duarte, Tiago L., Marcus S. Cooke, and George D. D. Jones. 2009. “Gene Expression Profiling Reveals New Protective Roles for Vitamin C in Human Skin Cells.” *Free Radical Biology and Medicine* 46(1):78–87. Retrieved (http://dx.doi.org/10.1016/j.freeradbiomed.2008.09.028).

Edger, Patrick P. et al. 2018. “Single-Molecule Sequencing and Optical Mapping Yields an Improved Genome of Woodland Strawberry (Fragaria Vesca) with Chromosome-Scale Contiguity.” *GigaScience* 7(2):1–7.

Feng, Tao et al. 2017. “Phylogenetic Study of the Tribe Potentilleae (Rosaceae), with Further Insight into the Disintegration of Sibbaldia.” *Journal of Systematics and Evolution* 55(3):177–91.

Gabbay, Kenneth H. et al. 2010. “Ascorbate Synthesis Pathway: Dual Role of Ascorbate in Bone Homeostasis.” *Journal of Biological Chemistry* 285(25):19510–20.

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):1–9. Retrieved (http://www.nature.com/articles/hortres201670).

Giampieri, Francesca et al. 2012. “The Strawberry: Composition, Nutritional Quality, and Impact on Human Health.” *Nutrition* 28(1):9–19.

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. Retrieved (http://journal.ashspublications.org/content/134/3/337.full).

Goddard, Mike. 2009. “Genomic Selection: Prediction of Accuracy and Maximisation of Long Term Response.” *Genetica* 136(2):245–57.

Graham, J., S. C. Gordon, and R. J. McNicol. 1997. “The Effect of the CpTi Gene in Strawberry against Attack by Vine Weevil (Otiorhynchus Sulcatus F. Coleoptera: Curculionidae).” *Annals of Applied Biology* 131(1):133–39.

Halvorsen, Bente L. et al. 2002. “A Systematic Screening of Total Antioxidants in Dietary Plants.” *The Journal of Nutrition* 132(3):461–71.

He, Joe Q., Richard J. Harrison, and Bo Li. 2017. “A Novel 3D Imaging System for Strawberry Phenotyping.” *Plant Methods* 13(1):93. Retrieved (http://plantmethods.biomedcentral.com/articles/10.1186/s13007-017-0243-x).

Heffner, Elliot L., Jean-luc Jannink, and Mark E. Sorrells. 2011. “Genomic Selection Accuracy Using Multifamily Prediction Models in a Wheat Breeding Program.” *The Plant Genome* 4(1):65–75.

Heffner, Elliot L., Aaron J. Lorenz, Jean Luc Jannink, and Mark E. Sorrells. 2010. “Plant Breeding with Genomic Selection: Gain per Unit Time and Cost.” *Crop Science* 50(5):1681–90.

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

Hooper, Lee et al. 2008. “Flavonoids , Flavonoid-Rich Foods , and Cardiovascular Risk : A Meta-Analysis of Randomized Controlled Trials 1 , 2.” 38–50.

Houde, Mario, Sylvain-Dallaire, Daniel N’Dong, and Fathey Sarhan. 2004. “Overexpression of the Acidic Dehydrin WCOR410 Improves Freezing Tolerance in Transgenic Strawberry Leaves.” *Plant Biotechnology Journal* 2(5):381–87.

Ichijima, K. 1926. “Cytological and Genetic Studies on Fragaria.” *Genetics* 11(6):590–604. Retrieved (http://www.genetics.org).

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. Retrieved (http://www.biomedcentral.com/1471-2148/11/9).

Jahufer, M. Z. Z. and Dongwen Luo. 2018. “Deltagen: A Comprehensive Decision Support Tool for Plant Breeders.” *Crop Science* 58(3):1118–31.

Khammuang, Saranyu, Srisulak Dheeranupattana, Prasert Hanmuangjai, and Sasitorn Wongroung. 2005. “Agrobacterium-Mediated Transformation of Modified Antifreeze Protein Gene in Strawberry Agrobacterium -Mediated Transformation of Modified Antifreeze Protein Gene in Strawberry.” *Songklanakarin Journal of Science and Technology* 27(4):693–703.

Kirsten, Amanda. 2014. *The Mid-Atlantic Berry Guide*.

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.” (July).

Li, Heng and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25(14):1754–60.

Liston, Aaron, Richard Cronn, and Tia-Lynn Ashman. 2014. “Fragaria: A Genus with Deep Historical Roots and Ripe for Evolutionary and Ecological Insights.” *American Journal of Botany* 101(10):1686–99.

De Los Campos, Gustavo et al. 2009. “Predicting Quantitative Traits with Regression Models for Dense Molecular Markers and Pedigree.” *Genetics* 182(1):375–85.

McNeill, J. et al. 2012. *International Code of Nomenclature for Algae, Fungi, and Plants*.

Mee Kin, Chai and Tan Guan Huat. 2010. “Headspace Solid-Phase Microextraction for the Evaluation of Pesticide Residue Contents in Cucumber and Strawberry after Washing Treatment.” *Food Chemistry* 123(3):760–64. Retrieved (http://dx.doi.org/10.1016/j.foodchem.2010.05.038).

Meng, Fanjuan et al. 2015. “Genetic Diversity and Population Structure Analysis in Wild Strawberry (Fragaria Nubicola L.) from Motuo in Tibet Plateau Based on Simple Sequence Repeats (SSRs).” *Biochemical Systematics and Ecology* 63:113–18. Retrieved (http://dx.doi.org/10.1016/j.bse.2015.09.018).

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.

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

Onda, Yoshihiko, Kotaro Takahagi, Minami Shimizu, Komaki Inoue, and Keiichi Mochida. 2018. “Multiplex PCR Targeted Amplicon Sequencing (MTA-Seq): Simple, Flexible, and Versatile SNP Genotyping by Highly Multiplexed PCR Amplicon Sequencing.” *Frontiers in Plant Science* 9(March). Retrieved (http://journal.frontiersin.org/article/10.3389/fpls.2018.00201/full).

Pajk, Tanja, Vida Rezar, Alenka Levart, and Janez Salobir. 2006. “Efficiency of Apples, Strawberries, and Tomatoes for Reduction of Oxidative Stress in Pigs as a Model for Humans.” *Nutrition* 22(4):376–84.

Passey, A. J., K. J. Barrett, and D. J. James. 2003. “Adventitious Shoot Regeneration from Seven Commercial Strawberry Cultivars (Fragaria x Ananassa Duch.) Using a Range of Explant Types.” *Plant Cell Reports* 21(5):397–401.

Peace, Cameron P. et al. 2017. “DNA-Informed Breeding of Rosaceous Crops: Promises, Progress and Prospects.” *Horticulture Research* 4(December 2016):17006. Retrieved (http://www.nature.com/articles/hortres20176).

Pereira, Vini. 2004. “Insertion Bias and Purifying Selection of Retrotransposons in the Arabidopsis Thaliana Genome.” *Genome Biology* 5(10):R79. Retrieved (papers://361aed09-ebce-4b55-aeec-a0834473af1d/Paper/p156).

Pincot, Dominique D. A. et al. 2018. “Genome-Wide Association Mapping Uncovers *Fw1,* a Dominant Gene Conferring Resistance to Fusarium Wilt in Strawberry.” *G3: Genes|Genomes|Genetics* 8(May):g3.200129.2018. Retrieved (http://g3journal.org/lookup/doi/10.1534/g3.118.200129).

Potter, D. et al. 2007. “Phylogeny and Classification of Rosaceae.” *Plant Systematics and Evolution* 266(1–2):5–43.

Potter, Daniel, James J. Luby, and Richard E. Harrison. 2000. “Phylogenetic Relationships among Species of Fragaria (Rosaceae) Inferred from Non-Coding Nuclear and Chloroplast DNA Sequences.” *Systemic Botany* 25(2):337–48.

Qin, Yonghua, Jaime A. Teixeira da Silva, Lingxiao Zhang, and Shanglong Zhang. 2008. “Transgenic Strawberry: State of the Art for Improved Traits.” *Biotechnology Advances* 26(3):219–32.

Rousseau-Gueutin, Mathieu et al. 2008. “Comparative Genetic Mapping between Octoploid and Diploid Fragaria Species Reveals a High Level of Colinearity between Their Genomes and the Essentially Disomic Behavior of the Cultivated Octoploid Strawberry.” *Genetics* 179(4):2045–60.

Samorodnitsky, Eric et al. 2015. “Comparison of Custom Capture for Targeted Next-Generation DNA Sequencing.” *Journal of Molecular Diagnostics* 17(1):64–75.

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

Schirmer, Melanie et al. 2015. “Insight into Biases and Sequencing Errors for Amplicon Sequencing with the Illumina MiSeq Platform.” *Nucleic Acids Research* 43(6).

Schulze-Menz, G. K. 1964. *Engler’s Syllabus Der Pflanzenfamilien II. 12th Ed.*

Shen, Zhiyong et al. 2010. “MPprimer: A Program for Reliable Multiplex PCR Primer Design.” *BMC Bioinformatics* 11(1):143. Retrieved (http://www.biomedcentral.com/1471-2105/11/143%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/20298595%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2858037).

Shikha, Mittal et al. 2017. “Genomic Selection for Drought Tolerance Using Genome-Wide SNPs in Maize.” *Frontiers in Plant Science* 8(April):1–12. Retrieved (http://journal.frontiersin.org/article/10.3389/fpls.2017.00550/full).

Shulaev, Vladimir et al. 2011. “The Genome of Woodland Strawberry (Fragaria Vesca).” *Nature Genetics* 43(2):109–16. Retrieved October 31, 2013 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3326587&tool=pmcentrez&rendertype=abstract).

Siles, J. A., A. Serrano, A. Martín, and M. A. Martín. 2013. “Biomethanization of Waste Derived from Strawberry Processing: Advantages of Pretreatment.” *Journal of Cleaner Production* 42:190–97. Retrieved (http://dx.doi.org/10.1016/j.jclepro.2012.11.012).

Silva, Katchen Julliany P., Asha Brunings, Natalia A. Peres, Zhonglin Mou, and Kevin M. Folta. 2015. “The Arabidopsis NPR1 Gene Confers Broad-Spectrum Disease Resistance in Strawberry.” *Transgenic Research* 24(4):693–704. Retrieved (http://dx.doi.org/10.1007/s11248-015-9869-5).

Somasagara, Ranganatha R. et al. 2012. “Extracts of Strawberry Fruits Induce Intrinsic Pathway of Apoptosis in Breast Cancer Cells and Inhibits Tumor Progression in Mice.” *PLoS ONE* 7(10):1–11.

Stanke, Mario and Burkhard Morgenstern. 2005. “AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints.” *Nucleic Acids Research* 33(SUPPL. 2):465–67.

Sun, Miao et al. 2016. “Phylogeny of the Rosidae: A Dense Taxon Sampling Analysis.” *Journal of Systematics and Evolution* 54(4):363–91.

Symons, G. M. et al. 2012. “Hormonal Changes during Non-Climateric Ripening in Strawberry.” *Journal of Experimental Botany* 63(13):4741–50.

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.

Thavamanikumar, S., R. Dolferus, and B. R. Thumma. 2015. “Comparison of Genomic

Selection Models to Predict Flowering Time and Spike Grain Number in Two Hexaploid Wheat Doubled Haploid Populations." *G3: Genes|Genomes|Genetics* 5(October):1991–98. Retrieved (http://g3journal.org/cgi/doi/10.1534/g3.115.019745).

Trejo-Téllez, Libia Iris and Fernando C. Gómez-Merino. 2014. "Nutrient Management in Strawberry: Effects on Yield, Quality and Plant Health" in *Strawberries*.

Viana, Alexandre Pio, Marcos Deon Vilela de Resende, Summaira Riaz, and Michael Andrew Walker. 2016. "Genome Selection in Fruit Breeding: Application to Table Grapes." *Scientia Agricola* 73(2):142–49. Retrieved (http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162016000200142&lng=en&tlng=en).

Vining, Kelly J. et al. 2017. "Genotyping-by-Sequencing Enables Linkage Mapping in Three Octoploid Cultivated Strawberry Families." *PeerJ* 5:e3731. Retrieved (https://peerj.com/articles/3731).

Xiang, Yezi et al. 2016. "Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication." *Molecular Biology and Evolution* 34(2):262–81.

Xue, H. et al. 2001. "Inhibition of Cellular Transformation by Berry Extracts." *Carcinogenesis* 22(2):351–56. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/11181460).

Yue, Chengyan et al. 2012. "Technical and Socioeconomic Challenges to Setting and Implementing Priorities in North American Rosaceous Fruit Breeding Programs." *HortScience* 47(9):1320–27.

Zhang, Shu Dong, Douglas E. Soltis, Yang Yang, De Zhu Li, and Ting Shuang Yi. 2011. "Multi-Gene Analysis Provides a Well-Supported Phylogeny of Rosales." *Molecular Phylogenetics and Evolution* 60(1):21–28. Retrieved (http://dx.doi.org/10.1016/j.ympev.2011.04.008).

Zorrilla-Fontanesi, Y. et al. 2012. "Genetic Analysis of Strawberry Fruit Aroma and Identification of O-Methyltransferase FaOMT as the Locus Controlling Natural Variation in Mesifurane Content." *Plant Physiology* 159(2):851–70. Retrieved (http://www.plantphysiol.org/cgi/doi/10.1104/pp.111.188318).

Appendices

The **table 2** lists the input conditions for MPprimer for multiplex design. Parameters are less stringent than typical for uniplexes to increase the probability of finding multiplexes.

Table 2. MPprimer options for GT-seq primer design

MPprimer Design Parameter	Value
PRIMER OPT SIZE	22
PRIMER MAX SIZE	35
PRIMER MIN SIZE	16
PRIMER OPT TM	60.0
PRIMER MAX TM	70.0
PRIMER MIN TM	50.0
PRIMER OPT GC PERCENT	50.0
PRIMER MAX GC	65.0
PRIMER MIN GC	35.0
PRIMER MAX END STABILITY	20.0
PRIMER SELF ANY	10.0
PRIMER SELF END	3.0
PRIMER NUM NS ACCEPTED	100
PRIMER PRODUCT SIZE RANGE	[150, 600]
PRIMER SALT CONC	50.0
PRIMER DIVALENT CONC	0
PRIMER DNTP CONC	0
PRIMER DNA CONC	50.0
PRIMER TM SANTALUCIA	1
PRIMER SALT CORRECTIONS	1
PRIMER FILE FLAG	0
PRIMER PICK INTERNAL OLIGO	0
PRIMER EXPLAIN FLAG	1