The allo-octoploid strawberry: simply complex

Thijs van Dijk

Thesis committee

Promotor
Prof. Dr R.G.F. Visser
Professor of Plant Breeding
Wageningen University

Co-promotor
Dr W.E. van de Weg
Senior Scientist, Wageningen UR Plant Breeding,
Wageningen University & Research

Other members
Prof. Dr Bart Thomma, Wageningen University
Prof. Dr Joost Keurentjes , University of Amsterdam, Wageningen University
Dr Bert Evenhuis, Wageningen Plant Research
Dr Cameron Peace, Washington State University, Pullman, USA

This research was conducted under the auspices of
the Graduate School Experimental Plant Sciences

The allo-octoploid strawberry: simply complex

Thijs van Dijk

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 11 November 2016
at 11 a.m. in the Aula.

Thijs van Dijk
The allo-octoploid strawberry: simply complex
186 pages.

PhD thesis, Wageningen University, Wageningen, NL (2016)
With references, with summary in English

ISBN 978-94-6257-963-7
DOI: 10.18174/392822

CONTENTS

Chapter 1

General Introduction

Thijs van Dijk

History

Strawberries... fragrant, sweet, juicy, fresh, mouth-watering… words that come to mind when people think of these delicious fruits. It is no wonder that the garden strawberry (*Fragaria x ananassa*) is the most important soft fruit species in the world. Nowadays, there are few sweet food items that do not come in strawberry flavour. Despite the present-day fame and importance of strawberry, the fruit knows humble beginnings.

The earliest mentions of strawberries in the Western world date back to Roman times [1], but its mentions are very sporadic and it is likely to have been a herb of little importance. Only by the 1500s did strawberry get mentioned more often as a cultivated fruit. In Europe only small fruited wild species of strawberry were known, such as the woodland strawberry (*Fragaria vesca*) and the musk strawberry (*Fragaria moschata*). In the early 1600s a new type of strawberry, named the Virginia strawberry made its way into Europe from North America. In Europe *Fragaria virginiana* was not regarded as an important fruit, and the woodland strawberry (*F.vesca*) continued to be the main cultivated species well into the 1700s. It was in 1714 that another type of new world strawberry (*Fragaria chiloensis*) made its way into Europe from South America. This new introduction proved to be the most important event in the history of strawberry cultivation and breeding [1], and the story of its discovery and introduction reads a bit like a James Bond novel (despite the topic being plants).

The story of its introduction is tied to the story of a certain French officer named Amédée Francois Frézier (Figure 1), whose surname and coat of arms, coincidentally, signify and feature strawberries [1]. In 1711, he was commissioned to sail to South America to spy on Spanish overseas fortifications as well as the strength and presence of Spanish influence there. In 1712 conditions were favourable enough for him to sail. Posing as a trader, aboard an armed merchant ship, he made his way to Chile. He quickly established himself there with Spanish officials, which allowed him to go about freely around their settlements. Being both an educated scientist and an army officer, Frézier was not only able to observe the military aspects, but made notifications on nearly all aspects of life including climate, geography, religion, customs and agriculture.

He found that local Indians in Concepcion, cultivated a type of strawberry which looked different from those cultivated in his native Europe, with rounder and thicker leaves, much larger fruits, and somewhat inferior taste (Figure 1). When Frézier left Concepcion in 1714, he took several plants with him and took care of them during the long journey home, which is difficult on a ship where potable water

is more precious than food. When Frézier arrived in France he had three plants, two others had been given to a person with whom he travelled on the ship. He gave one plant to a friend, who would plant it at King Louis XIV's royal garden, gave one plant to his boss, and kept one plant himself. From these three plants, interested botanists spread the *Fragaria chiloensis* over Europe. It would take another 50 years however, before people in Europe were able to cultivate strawberries of the same size as those observed in Chile by Frézier. This was due to mixed results with the cultivation of *Fragaria chiloensis*: it turned out that Frézier had only brought male-sterile plants, although this was not known to the growers at the time. Frézier continued working on his primary occupation in the study and construction of fortifications both overseas and domestic and lived to the ripe age of 91 [1]. Perhaps he even ate fruits of the plant he eventually helped create, but the discovery and description of the garden strawberry can be attributed to another Frenchman: Antoine Nicolas Duchesne.

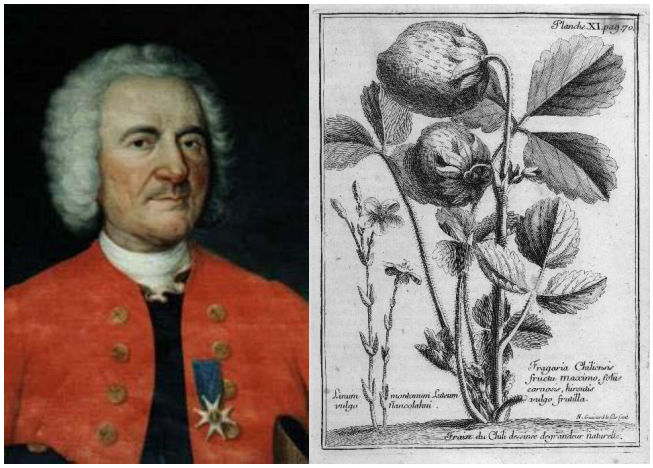


Figure 1. Left : Amédée Francois Frézier. Right : Drawing of the *Fragaria chiloensis* which Frézier made during his spying mission in Chile.

Antoine Nicolas Duchesne was 17 when he presented a pot of strawberries to King Louis XV in 1764. These fruits were of large size, because Duchesne had hand-pollinated *Fragaria chiloensis* with pollen from *Fragaria Moschata* [1]. Impressed, the King ordered him to document and collect all strawberries known in Europe. Duchesne contacted botanists and gardeners all over Europe for information about local strawberry plants and different specimens. He even contacted the famous Linnaeus and even Frézier, then 82, who told him about the habitat of *Fragaria chiloensis* and its introduction in France. Duchesne was able to identify that strawberry had different sexes, including male, female and hermaphrodite plants

among the different species of strawberry he collected. One of the plants he received was noted for its "monstrous" fruit, and although this specific plant had been known before Duchesne's description of it, Duchesne was the first one who recognized it for what it was: A hybrid between *Fragaria chiloensis* and *Fragaria virginiana*. He recognized characteristics from both species in this single plant, and found that the flowers were hermaphroditic in nature. He named the species *Fragaria x ananassa* due to its hybrid nature and distinct pineapple flavour [1]. Exactly where and when this hybridisation had occurred in Europe is unclear, but we know that the plant that Duchesne described must have been one of the earliest examples of the now widely cultivated garden strawberry, and probably an ancestor common to all modern day varieties.

Physiology and Morphology

Darrow, perhaps the foremost strawberry researcher and breeder of the 20th century, called the strawberry one of the most changeable of all crop plants [1]. This isn't surprising if you take into consideration its arsenal of survival and multiplication strategies. Strawberry is a low herbaceous perennial plant with both a vegetative as well as a sexual reproduction cycle (Figure 2). The shortened crown stem is the main part of the plant. From here, roots emerge from the base. From the axial buds of the shoot, different structures such as leaves, runners, inflorescences and branch crowns can emerge. A typical Northern-European growing season of a short-day strawberry consists out of different phases. In early spring, winter dormancy is broken and leaves start emerging. After several leaves are fully grown, inflorescences start emerging. Each inflorescence carries a number of flowers, the number and type of flower (male, female, hermaphrodite) can vary greatly depending on genotype. At the end of May, beginning of June the first fruits ripen and this continues for about three to four weeks, after which no more inflorescences emerge. The fruit consists of a fleshy part which is a false fruit (a swollen receptacle), upon which numerous true fruit are present (achenes), which look like seeds.

Starting at the harvest season, stolons (Figure 2) emerge from the plants which form runner plants at distances up to a few meters from the mother plant through successive runnering. This vegetative reproduction lasts throughout summer. When day length shortens to below 13-14 hours [1], the plant starts forming its generative meristems (inflorescences) for the following year. This initiation continues until the plant is fully dormant, which is usually around the end of November, depending on genotype, temperature as well as day-length. After winter, dormancy is broken in early spring and the cycle starts over. Branch crowns are formed throughout the

growing season, and are functionally the same as main crowns. Winter hardiness is dependent on the cultivar, those with strong dormancy generally performing better than those without. Some varieties can tolerate temperatures down to -20 degrees centigrade. In short, the garden strawberry is a highly adaptable plant because it is capable of vegetative (short distance) and seed propagation (long distance). There are strawberry varieties that can form flowers under short days, but also long days. It is capable of forming male, female and hermaphrodite flowers depending on genotype. And it has a wide range of dormancy types, depending on climate.

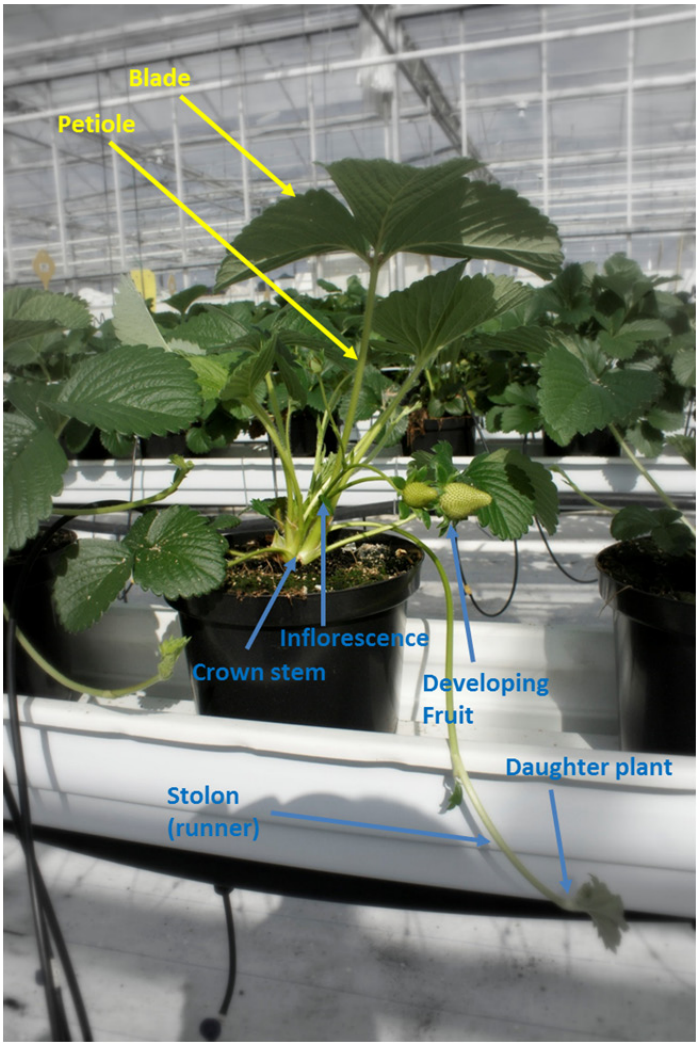


Figure 2. Strawberry plant structure.

Strawberry Cultivation

Strawberry is the most important soft fruit in both production volume as well as economic value (Hancock, 1999). Worldwide production in 2013 equalled 7.5 million tonnes, with Asia accounting for nearly 50% (mainly China), the Americas for 25% (mainly US) and Europe for 20% of production (FAOSTAT, 2013). Strawberry production in Europe can be divided up into Northern and Southern climate zones. In Northern zones, the main body of production occurs in Spring and Summer. Thanks to protected culture, the only production gap that remains is between mid-December and early March. This is when the main production occurs for Southern Climate zones, mainly in Spain, but also in Morocco, Egypt and Turkey. These regions are focused on export to Northern countries, to guarantee strawberry availability throughout the year.

Strawberries are generally grown on specialized farms, and require high economic inputs in terms of labour (manual picking) and plant material. Cultivation in Northern Europe is typified by a high diversity of production systems and cultivation techniques. Heated greenhouses are used for extending the season to Fall and early Spring. In the Netherlands, only 10% of the total acreage of strawberry consists of greenhouses, but they produce 50% of the total strawberry volume (Bert Meulenbroek, breeder at Fresh Forward, Eck and Wiel; personal communication). Open field production (Figure 3) in the Netherlands starts around mid-May and continues until late July. From there, tunnel and table top cultivation (semi-protected cultures, Figure 3) take over the main part of production, which in turn is succeeded by greenhouse production in Fall (Figure 3).

Plant material is produced by specialized nurseries in the preceding year. These nurseries produce a wide variety of plant types (e.g. tray plants, frigo plants, fresh plants), suited to the many different cultivation systems. Because plants can be cold-stored, growers enjoy a great flexibility in planning their plantings, and thereby enabling a continuous supply throughout the growing season. Short day (SD) plants remain the main flowering habit in use in Europe, because SD allows for intensive harvest seasons with relatively uniform quality and minimal pesticide usage, and because cold storage negates the disadvantages of SD flowering (lower per plant yield and short availability).

Figure 3. Different strawberry cultivation systems. Top Left: Open field production with straw mulch. Top Right: Tunnel cultivation (in Spain) with plastic mulch. Bottom Left : Tabletop cultivation on substrate with plastic cover. Bottom right : Greenhouse production on substrate.

Pests and diseases

There are many diseases and pests which attack strawberry. Only a limited number however, are of true economic importance due to their damage potential and their difficulty to control. *Botrytis* on mature fruit (grey mould) causes a lot of damage through losses both in the field as well as through limited shelf life. Crown rot (*Phytophthora cactorum)* can be extremely devastating when nursery plant material is infected and susceptible cultivars are grown, losses of up to 20% through plant death are not uncommon. The unpredictability of this disease makes it much feared by strawberry growers [3]. *Verticillium dahliae* is very important in open field production through reduced plant vigour and yield loss. Powdery mildew (*Podosphaera aphanis*), causes damage on both plants and fruits in both greenhouse as well as open field culture. Anthracnose (*Colletotrichum spp*.) can cause problems during warm and humid weather conditions. *Phytopthora fragariae* and *Xanthomonas fragariae* are mainly a problem for nurseries, as they are quarantine diseases. The detection of *Phytophthora fragariae* (a root pathogen) in fields makes them unavailable for future strawberry propagation. Whereas the detection of *Xanthomonas fragariae* can cause entire plant lots to be rejected. *X. fragariae* is a more recent disease in Europe,

whose importance may increase with global warming. The main insect problems are caused by the Western Flower Thrips, *Frankliniella occidentalis* (especially in greenhouses) and the relatively new plague insect *Drosophila suzukii*. Western flower thrips causes losses through superficial fruit damage, malformation and transmission of viruses to breeding stock. Infestations have proven very difficult to control. *Drosophila suzukii* is an exotic fruit fly from Asia, which has become a major problem in all soft fruit crops. The main problem lies in the ability of this fly to deposit eggs in unripe fruits, which can lead to infected fruits with living larvae ending up on supermarket shelves.

Abiotic stress

Abiotic stresses can cause serious damage to strawberry growers, mainly in unprotected cultures. Early in the season, freezing temperatures can cause serious damage to flowers and thereby affect yield. High temperatures cause water stress and loss of fruit size and quality, despite the use of irrigation. Heat stress in combination with *Verticillium* infection can have a synergistically destructive effect on fruit quality, leading to serious loss of fruit size, shine and taste through emergency ripening. Hail storms can have a devastating effect, albeit often occurring only locally. Heavy rain causes fruits to swell up and crack, affecting appearance, taste and allowing for severe *Botrytis* infections to occur.

In face of all these possible biotic and especially abiotic threats to strawberry production, the general trend is to move towards more protected cultures to assure a more constant fruit quality.

Strawberry Breeding

History

The 19th century saw the start of breeding by individual botanists and growers, leading to the first named varieties of strawberry [1]. At the end of the 19th century, ever bearing (flowering under long day) varieties were first described and introduced. In the 20th century, professional breeding started, first at universities and later at private companies as well.

Contrary to vegetable breeding, strawberry breeding is mainly performed at small private companies or at governmental institutes and universities. This relatively small scale makes many modern strawberry breeding programs comparatively old-fashioned, with regards to the application of for instance marker technologies. A

typical breeding cycle (Table 1) of strawberry is rather straightforward and is very comparable to potato breeding, as they are both clonally propagated and outbreeding.

Basically, crosses are made between varieties and interesting selections, seeds are sown and seedlings are selected the following year. These are then clonally propagated and evaluated over multiple years for their performance. Each year, production trials increase in size and the number of genotypes under evaluation diminishes. The last phase of the breeding cycle consists of multi-site testing, application of breeder's rights and multiplication of virus-free, certified plant material. Finally, after about 6 to 8 years of evaluation, a variety may be released.

Table 1. Typical breeding cycle for a specific market segment of strawberry (e.g. everbearing varieties). BT is Breeder Trial, FT is farmer trial (external).

Year	Stage	Nr of Genotypes	Trial size (n)
-1	Cross+ Sowing	10,000	1
0	Seedling	10,000	1
1	BT	300	6
2	BT	60	10
3	BT	20	20
4	BT&FT	15+5	40, 100-1k
5	FT	3	1k-10k
6	FT	2	100k+
7	Release	1	1m+

Breeding goals

Breeding goals in strawberry have shifted over the years. The diversification of production methods and intensivation of strawberry culture in general has resulted in a more fragmented market for strawberry varieties. This also results in more diversified breeding goals, where varieties will be increasingly tailored to suit a specific market. In general, breeding goals can be divided into two categories: fruit quality traits and agronomic traits. Important fruit quality traits include: taste, uniformity, shelf life, size, shape, colour and shine. Important agronomical traits include: Seasonality (early-late, ever bearing), yield, disease resistance, plant morphology. Many of the traits listed here are inter-connected, sometimes antagonistically (e.g. taste and yield + shelf life), thus making the breeding effort a balancing act which requires great skill. Current trends in breeding are consumer acceptance (better taste and appearance,

minimal pesticide usage) and farmer profitability (season extension, labour saving, high first class yields). This means that for traits which are under relatively simple genetic control and that are involved in one of these trends, massive gains could be made quickly in breeding through the use of marker assisted selection.

Strawberry taxonomy and genetics

The garden strawberry (*Fragaria x ananassa*), is a member of the Rosaceae family, a very important horticultural family that includes crops such as apple, pear, cherry, peach, almonds, plums, raspberry and rose. The genus *Fragaria* contains 20 species spread over mainly the Northern temperate climate zones [4]. The base chromosome number of *Fragaria* is 7, which is typical for a member of the Rosaceae family. However, over half the members of the *Fragaria* genus are polyploid, ranging from tetra- to decaploid [4]. The garden strawberry is also a polyploid (2n=8x=56), a characteristic which it shares with some of the most well-known crop species such as (bread)wheat, sugarcane, potato and cotton. Although strawberry is an octoploid, its genetics are simpler than for instance, the autotetraploid potato. This is due to the fact that strawberry is an allopolyploid, which means that the chromosomes are differentiated into subgroups (homoeologues or sub-genomes) between which no pairing occurs [5]. In practice this means that traits are inherited in a similar fashion as for diploid crops (disomically), but that performing genetic studies such as developing and analysing molecular markers, or performing sequencing projects, are complicated due to the high sequence homology between sub-genomes.

Strawberry Molecular research

Molecular genetic research in strawberry commenced in the mid-1990s with the publication of the first genetic maps in diploid strawberry (*Fragaria vesca*) [6, 7], and of the first predictive molecular markers for disease resistance in the octoploid strawberry [8-10]. Soon after, projects on the development of genome covering genetic linkage maps were set up for the cultivated octoploid strawberry [5, 11-14]. These early genetic maps mainly made use of microsatellite (SSR) or AFLP markers and eventually revealed that the inheritance of the octoploid strawberry was fully disomic. The first octoploid maps were plagued by several shortcomings however, such as splitted linkage groups, lack of sub-genome denotation and limited syntenic comparison between sub-genomes. Recently, much progress has been booked in terms of genomic tools available for strawberry. In 2011 the reference genome sequence of diploid strawberry *Fragaria vesca* was published [15]. Soon

after the first SNP-based marker maps were published [16], followed by the recent development of a genome wide 90k SNP array [17].

In recent years many QTLs and marker trait relations have been published and this trend seems to be accelerating every year as new tools become available. The range of traits for which marker associations have been published is now quite wide and includes : aroma [16, 18], sugar and acidity QTLs [19, 20], disease resistance genes and QTLs [8-10, 21-23] and flowering [14, 24, 25].

Another field which profited from the availability of new molecular tools is the study on the origins of strawberry's sub-genomes. Previous research had pointed towards the diploid species *Fragaria vesca* and *Fragaria innumae* as ancestral to the wild octoploids, with *F. vesca* being the maternal donor [26, 27]. Conclusive evidence for the contribution of these two species to the octoploid strawberry came in 2009 [28]. Only very recently however, did a genome wide study reveal the exact composition of all four sub-genomes in strawberry [29, 30]. These studies showed that *F. innumae* (or *F. innumae* like species) are ancestral to three out of four sub-genomes, with the maternal *F. vesca bracteata* making up only one. It is likely that tetraploid intermediaries were combined to create the octoploid current day strawberry [30].

Thesis outline

Within the context of enabling marker assisted selection for strawberry breeding programs, the goal of this thesis is twofold: First to provide new insights into and tools on strawberry genetics through the development of molecular marker tools and analysis methods for the allo-octoploid strawberry, second, to apply these tools in the detection of marker-trait associations.

In **Chapter 2** we developed a method we named MADCE, which stand for Microsatellite Allele Dose & Configuration Establishment. The MADCE method is a very thorough method that enables a researcher to establish the exact allelic configuration of all sub-genomes in an allopolyploid by making use of allele dosage and segregation. This allows for exact comparison of syntenic regions between the homoeologous chromosomes (sub-genomes) as well as for the establishment of haplotype information.

In **Chapter 3** a genome wide linkage map was created for the octoploid strawberry using the MADCE method and the Holiday x Korona (HxK) mapping population. The resultant HxK map was the first to provide genome wide haplotype and homozygosity information. Moreover, we uncovered a large inversion on LG 2D, possibly spanning almost an entire chromosome arm. Another innovation was

the first attempt at ordering the homoeologous linkage groups according to possible origin by their ability to amplify *F. vesca* based primers. The HxK map has been used in follow-up projects, including the development of the 90K Affymetrix SNP array [17].

In **Chapter 4** we performed a QTL analysis for *Phytophthora cactorum* resistance on two separate, but somewhat related mapping populations (Holiday x Korona and E1998-142 x Elsanta), using multiyear disease tests. One population (HxK) showed only weak QTLs with little promise. The other population (ExEls) showed a QTL of moderate effect on LG7D, which proved to be stable over different years.

In **Chapter 5** we investigated the genetics of perpetual flowering (PF) in strawberry. We found clear qualitative segregation of the PF phenotype in our two mapping populations and were able to fine map the trait to a region of 900kb on LG4A. This region contained two important flowering pathway candidate genes: *FaCDF2* and *FaFT2*. After extensive haplotyping, and sequencing the candidate gene *FaCDF2* in a genetically wide set of cultivars, we concluded that the *PF* locus is present in all perpetual flowering varieties, but that the locus also occurs in part of the seasonal flowering varieties (SF). A possible explanation for this is that a second locus is required for PF which was not segregating in our mapping populations. An alternative explanation is that we have been unable to find a (functional) polymorphism that distinguishes the PF-like haplotype in SF plants from that of PF plants.

Chapter 6 is the general discussion, in which I summarize the results of the experimental chapters and place them in a broader context that includes the fields of science in general, molecular genetic research and plant breeding in particular.

References

1. Darrow GM: **The strawberry; history, breeding, and physiology**, 1st edn. New York,: Holt; 1966.
2. Strand LL: **Integrated Pest Management for Strawberries**. Berkeley, California: Agriculture and Natural Resources, University of California; 1994.
3. Verstryinge J, Meurrens F, Morren H, Demeyere A: **Ziekten en plagen in aardbeien**. In. Edited by Visserij Le, 5 edn. Brussels, Belgium: Jules Van Liefferinge; 2015: 70.
4. Hummer K, Bassil N, Njuguna W: **Fragaria**. In: *Wild crop relatives : Genomic and Breeding Resources, Temperate Fruits* Edited by Kole C. Heidelberg: Springer-Verlag 2011: 247.
5. Sargent DJ, Fernandez-Fernandez F, Ruiz-Roja JJ, Sutherland BG, Passey A, Whitehouse AB, Simpson DW: **A genetic linkage map of the cultivated strawberry (*Fragaria x ananassa*) and its comparison to the diploid *Fragaria* reference map**. *Molecular Breeding* 2009, **24**(3):293-303.
6. Davis TM, Yu H: **A linkage map of the diploid strawberry, *Fragaria vesca***. *J Hered* 1997, **88**(3):215-221.
7. Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM *et al*: **An enhanced microsatellite map of diploid *Fragaria***. *Theor Appl Genet* 2006, **112**(7):1349-1359.
8. Haymes KM, Henken B, Davis TM, vandeWeg WE: **Identification of RAPD markers linked to a *Phytophthora fragariae* resistance gene (*Rpf1*) in the cultivated strawberry**. *Theor Appl Genet* 1997, **94**(8):1097-1101.
9. Haymes KM, Van de Weg WE, Arens P, Maas JL, Vosman B, Den Nijs APM: **Development of SCAR markers linked to a *Phytophthora fragariae* resistance gene and their assessment in European and North American strawberry genotypes**. *J Am Soc Hortic Sci* 2000, **125**(3):330-339.
10. Lerceteau-Kohler E, Guerin G, Denoyes-Rothan B: **Identification of SCAR markers linked to *Rca2* anthracnose resistance gene and their assessment in strawberry germplasm**. *Theor Appl Genet* 2005, **111**(5):862-870.
11. Ashley MV, Wilk JA, Styan SMN, Craft KJ, Jones KL, Feldheim KA, Lewers KS, Ashman TL: **High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana Mill. (Rosaceae)***. *Theor Appl Genet* 2003, **107**(7):1201-1207.
12. Rousseau-Gueutin M, Lerceteau-Kohler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arus P, Guerin G, Denoyes-Rothan B: **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* 2008, **179**(4):2045-2060.
13. Spigler RB, Lewers KS, Johnson AL, Ashman TL: **Comparative mapping reveals autosomal origin of sex chromosome in octoploid *Fragaria virginiana***. *J Hered*

2010, **101 Suppl 1**:S107-117.
14. Spigler RB, Lewers KS, Main DS, Ashman TL: **Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome**. *Heredity* 2008, **101**(6):507-517.
15. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP *et al*: **The genome of woodland strawberry (*Fragaria vesca*)**. *Nat Genet* 2011, **43**(2):109-U151.
16. Sanchez-Sevilla JF, Cruz-Rus E, Valpuesta V, Botella MA, Amaya I: **Deciphering gamma-decalactone biosynthesis in strawberry fruit using a combination of genetic mapping, RNA-Seq and eQTL analyses**. *BMC Genomics* 2014, **15**.
17. Bassil NV, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR *et al*: **Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa**. *BMC Genomics* 2015, **16**(1):1-30.
18. Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sanchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya I: **Genetic Analysis of Strawberry Fruit Aroma and Identification of O-Methyltransferase *FaOMT* as the Locus Controlling Natural Variation in Mesifurane Content**. *Plant Physiol* 2012, **159**(2):851-+.
19. Lerceteau-Kohler E, Moing A, Guerin G, Renaud C, Petit A, Rothan C, Denoyes B: **Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control**. *Theor Appl Genet* 2012, **124**(6):1059-1077.
20. Zorrilla-Fontanesi Y, Cabeza A, Dominguez P, Medina JJ, Valpuesta V, Denoyes-Rothan B, Sanchez-Sevilla JF, Amaya I: **Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria x ananassa*)**. *Theor Appl Genet* 2011, **123**(5):755-778.
21. Antanaviciute L, Surbanovski N, Harrison N, McLeary KJ, Simpson DW, Wilson F, Sargent DJ, Harrison RJ: **Mapping QTL associated with Verticillium dahliae resistance in the cultivated strawberry (Fragaria X ananassa)**. *Hortic Res-England* 2015, **2**.
22. Denoyes-Rothan B, Guerin G, Lerceteau-Kohler E, Risser G: **Inheritance of resistance to *Colletotrichum acutatum* in *Fragaria x ananassa***. *Phytopathology* 2005, **95**(4):405-412.
23. Denoyes-Rothan B, Lerceteau-Köhler E, Guérin G, Bosseur S, Bariac J, Martin E, Roudeillac P: **QTL analysis for resistance to *Colletotrichum acutatum* and *Phytophthora cactorum* in octoploid strawberry (*Fragaria x ananassa*).** . *Acta Hort* 2004, **663**:147-152.
24. Castro P, Bushakra JM, Stewart P, Weebadde CK, Wang D, Hancock JF, Finn CE, Luby JJ, Lewers KS: **Genetic mapping of day-neutrality in cultivated strawberry**. *Molecular Breeding* 2015, **35**(2).
25. Gaston A, Perrotte J, Lerceteau-Kohler E, Rousseau-Gueutin M, Petit A, Hernould

M, Rothan C, Denoyes B: ***PFRU*, a single dominant locus regulates the balance between sexual and asexual plant reproduction in cultivated strawberry**. *J Exp Bot* 2013, **64**(7):1837-1848.

26. Harrison RE, Luby JJ, Furnier GR: **Chloroplast DNA restriction fragment variation among strawberry (Fragaria spp) taxa**. *J Am Soc Hortic Sci* 1997, **122**(1):63-68.
27. Potter D, Luby JJ, Harrison RE: **Phylogenetic relationships among species of *Fragaria (Rosaceae)* inferred from non-coding nuclear and chloroplast DNA sequences**. *Syst Bot* 2000, **25**(2):337-348.
28. Rousseau-Gueutin M, Gaston A, Ainouche A, Ainouche ML, Olbricht K, Staudt G, Richard L, Denoyes-Rothan B: **Tracking the evolutionary history of polyploidy in *Fragaria L.* (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes**. *Mol Phylogenet Evol* 2009, **51**(3):515-530.
29. Govindarajulu R, Parks M, Tennessen JA, Liston A, Ashman TL: **Comparison of Nuclear, Plastid, and Mitochondrial Phylogenies and the Origin of Wild Octoploid Strawberry Species**. *American Journal of Botany* 2015, **102**(4):544-554.
30. Tennessen JA, Govindarajulu R, Ashman TL, Liston A: **Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes Revealed by Dense Targeted Capture Linkage Maps**. *Genome Biol Evol* 2014, **6**(12):3295-3313.

Chapter 2

Microsatellite Allele Dose and Configuration Establishment (MADCE): an integrated approach for genetic studies in allopolyploids.

Thijs van Dijk, Yolanda Noordijk, Tiphaine Dubos, Marco C.A.M. Bink, Bert J. Meulenbroek, Richard G F Visser, Eric van de Weg
BMC Plant Biol. 2012 Feb 17;12:25. doi: 10.1186/1471-2229-12-25.

Abstract

Genetic studies in allopolyploid plants are challenging because of the presence of similar sub-genomes, which leads to multiple alleles and complex segregation ratios. In this study, we describe a novel method for establishing the exact dose and configuration of microsatellite alleles for any accession of an allopolyploid plant species. The method, named Microsatellite Allele Dose and Configuration Establishment (MADCE), can be applied to mapping populations and pedigreed (breeding) germplasm in allopolyploids.

Two case studies are presented to demonstrate the power and robustness of the MADCE method. In the mapping case, five microsatellites were analysed. These microsatellites amplified 35 different alleles based on size. Using MADCE, we uncovered 30 highly informative segregating alleles. A conventional approach would have yielded only 19 fully informative and six partially informative alleles. Of the ten alleles that were present in all progeny (and thereby ignored or considered homozygous when using conventional approaches), six were found to segregate by dosage when analysed with MADCE. Moreover, the full allelic configuration of the mapping parents could be established, including null alleles, homozygous loci, and alleles that were present on multiple homoeologues. In the second case, 21 pedigreed cultivars were analysed using MADCE, resulting in the establishment of the full allelic configuration for all 21 cultivars and a tracing of allele flow over multiple generations. The procedure described in this study (MADCE) enhances the efficiency and information content of mapping studies in allopolyploids. More importantly, it is the first technique to allow the determination of the full allelic configuration in pedigreed breeding germplasm from allopolyploid plants. This enables pedigree-based marker-trait association studies the use of algorithms developed for diploid crops, and it may increase the effectiveness of LD-based association studies. The MADCE method therefore enables researchers to tackle many of the genotyping problems that arise when performing mapping, pedigree, and association studies in allopolyploids. We discuss the merits of MADCE in comparison to other marker systems in polyploids, including SNPs, and how MADCE could aid in the development of SNP markers in allopolyploids.

Background

Polyploidy is an integral part of the evolution of all plant species [1]. Several important crop species are polyploids, including bread wheat (*Triticum aestivum,* allohexaploid), cotton (*Gossypium spp.*, allotetraploid), potato (*Solanum tuberosum*, autotetraploid) and the very complex sugar cane (*Saccharum spp.*, auto-allo-polyploid). The success of polyploids can be ascribed to multiple factors, including

their ability to retain beneficial alleles while allowing the generation of novel variation in duplicated alleles and increased vigour through perpetual hybridity [2].

Autopolyploidy is the result of the combination or duplication of multiple genomes that are sufficiently similar to allow for random bivalent pairing and the formation of multivalents during meiosis. This random chromosomal pairing complicates genetic studies, especially mapping studies, and limits mapping studies to mostly simplex alleles that are in the coupling phase. In contrast, allopolyploids are derived either from the merging of differentiated genomes within one zygote [1] or through the gradual diploidisation of an autopolyploid [1,3,4]. These differentiated genomes behave essentially as meiotically independent sub-genomes (homoeologues) that almost exclusively form bivalents between chromosomes of the same sub-genome, resulting in essentially disomic inheritance. Genetic studies in allopolyploids are complex because alleles can be amplified from multiple homoeologues and because some of these alleles are identical (shared) between homoeologues. However, the disomic inheritance expressed by allopolyploids makes them amenable to diploid mapping procedures. In most genetic mapping studies, the pitfall of shared alleles is circumvented by evaluating only alleles for which one or both of the parents are heterozygous [5]. Once initial single parental maps have been created based on 1:1 segregating alleles, the maternal and paternal maps can be integrated using the 3:1 segregating alleles as bridge markers [6]. Although this approach makes genotyping in polyploids reliable and relatively simple, it is highly restrictive and a great deal of information is lost. Furthermore, this approach limits the number of homoeologous loci that can be mapped with individual microsatellite primer pairs when shared alleles are present. The use of microsatellite markers for genetic studies in allopolyploids has several benefits. Because these markers are multi-allelic, they can show as many different alleles as the ploidy level for an individual plant. This allows the simultaneous mapping of several homoeologous chromosomes and the subsequent evaluation of their macro-synteny. With the advent of fluorescent detection techniques for PCR products, it became possible to reliably quantify the abundance of an amplicon in a PCR reaction. The availability of this allele dose information is the first step in the establishment of allelic configurations in polyploids. Several studies have investigated the use of quantitative data to estimate allele dose in autotetraploids [7-12]. The MAC-PR method, which was developed by Esselink et al. [8], uses ratios between the alleles of a single locus within an accession and compares these ratios with those of other accessions in which the same alleles occur together. The presence of different ratios indicates variability in dose for at least one of the two alleles under investigation. Here we propose the MADCE method for dose estimation, which is based on improvements over the MAC-PR method. In MADCE, we also use disomic segregation patterns and virtual reference alleles to refine the dosage estimation. In this paper, we use the MADCE method to

determine allelic configurations in the allo-octoploid cultivated strawberry (*Fragaria x ananassa*) and to demonstrate its utility. This crop species has recently been studied extensively in an effort to create linkage maps [13-16]. The results from five microsatellite primer pairs analysed with MADCE are presented to demonstrate its effectiveness for the construction of genetic linkage maps and the determination of the allelic configuration of mapping parents. Extended methodologies are presented for the application of MADCE in pedigreed cultivars and breeding lines. Finally, we demonstrate the value of this method for examining the flow of alleles over multi-generation pedigrees through the Identity-By-Descent concept using the FlexQTL™ [17] and Pedimap [18] software packages.

Results

The MADCE procedure for mapping studies

The Microsatellite Allele Dose and Configuration Establishment (MADCE) method is composed of five successive phases. It starts with (I) a qualitative interpretation of microsatellite data, followed by (II) a quantitative assessment of allele doses, (III) an assessment of the initial allele configuration of the mapping parents, (IV) the generation of molecular marker linkage maps, and (V) the final characterisation of the parental haplotypes, including homozygous loci. Below, these steps are elaborated and exemplified through a step-by-step analysis of five microsatellite primer pairs (Table 1) used in a mapping population derived from a cross between the octoploid strawberry cultivars 'Holiday' (H) and 'Korona' (K), for which the original data are given in additional files 1 through 5. Two of these examples are presented in this document, and the other three are provided in additional files 6 to 8. It may be useful to keep a printout of Tables 2 and 3 at hand when going through the examples.

Table 1. Microsatellite names and sources

Name	Reference	ABI Platform	Study
ARSFL010	Lewers et al. 2005	ABI 3730	M,P
CO817823	Spigler et al. 2008	ABI 3730	M
CX661101	Spigler et al. 2008	ABI 3730	M,P
PSContig944	Spigler et al. 2008	ABI 3730	P
UAFv7500	Bassil et al. 2006	ABI 3730	M,P
UFFxa16H07	Sargent et al. 2006	ABI 3730	M,P

M=used in mapping study, P=used in pedigree study

Table 2. Allelic pairs in order of appearance and homoeologue assignment

Microsatellite	Homoeologue Nr	Allele 1 'Holiday'	Allele 2 'Holiday'	Allele 1 'Korona'	Allele 2 'Korona'	Homoeologue after mapping
ARSFL010						
	H1	257	234	286	246	A
	H2	269	269	269	269	C
	H3	0	248	0	0	B
	H4	0	0	0	0	D
Uffxa16H07						
	H1	298	269	269	306	B
	H2	262	262	262	262	D
	H3	266	266	266	268	A
	H4	279a	279b	273	287	C
UAFv7500						
	H1	336	345	330	345	C
	H2	342	348	348	342	A
	H3	342	342	342	333	B
	H4	348	348	348	348	D
CX661101						
	H1	212	212	212	212	D
	H2	223	221	223	224	A
	H3	223	220	223	223	B
	H4	204	204	0	0	C
CO817823						
	H1	199	203	199	193	B
	H2	216	195	203	195	C
	H3	236	207	209	236	A
	H4	207	207	207	207	D

Table 3. MADCE. analysis results of mapped microsatellites. Repulsion alleles are given if not homozygous

Microsatellite	Allele size (bp)	Holiday (H) or Korona (K)	Allele Presence Absence segregation	Actual Ratio Cluster segregation	Mean values of RCs (Ratio Clusters)	Single dose ratio	Holiday Ratio value
ARSFL010							
	234	H	**1:1**		0, 27	27	26
	246	K	**1:1**		0, 16	16	0
	248	H	**1:1**		0, 4	4	3
	257	H	**1:1**		0, 15	15	12
	269*	H,K	1:0		REF	REF	REF
	286	K	**1:1**		0, 4	4	0
UFFxa16H07							
	262	H,K	1:0		REF	REF	REF
	266	H,K	1:0	**1:1**	17, 39	17	42
	268	K	**1:1**		0, 13	13	0
	269	H,K	**3:1**	**1:2:1**	0, 9, 20	9	10
	273	K	**1:1**		0, 9	9	0
	279	H	1:0	**1:1**	10, 15	10, 15	22
	287	K	**1:1**		0, 6	6	0
	298	H	**1:1**		0, 3	3	3
	306	K	**1:1**		0, 2	2	0
UAFv7500							
	330	K	**1:1**		0, 6	6	0
	333	K	**1:1**		0, 3	3	0
	336	H	**1:1**		0, 6	6	6
	342	H,K	1:0	**1:3:3:1**	7,14,19,24	6	16
	345	H,K	**3:1**	**1:2:1**	0, 7, 13	7	7
	348	H,K	1:0	**1:2:1**	13,18,24	6	17
CO817823							
	193	H	**1:1**		0, 22	22	21
	195	H,K	**3:1**	**1:2:1**	0, 16,28	16	17
	199	H,K	**3:1**	**1:2:1**	0, 17,35	17	16
	203	H,K	**3:1**	**1:1:1:1**	0, 9,14,24	9(K1C)14(H1B)	14
	207	H,K	1:0	**1:1**	8, 13	4	13
	209	K	**1:1**		0, 8	8	0
	216	H	**1:1**		0, 6	6	6
	236	H,K	**3:1**	**1:2:1**	0, 4, 7	4	4
CX661101							
	204*	H	1:0		9	9	15
	212*	H,K	1:0		20	10	18
	220	H	**1:1**		0, 8	8	7
	221	H	**1:1**		0, 8	8	8
	223	H,K	1:0	**1:3:3:1**	8,16,23,31	7	12
	224	K	**1:1**		0, 9	9	0

Microsatellite	Korona Ratio value	Nr of doses present in Holiday	Nr of doses present in Korona	Holiday repulsion allele	Korona repulsion allele	Homoeologue
ARSFL010						
	0	1	0	257	-	A
	18	0	1	-	286	A
	0	1	0	0	-	B
	0	1	0	234	-	A
	REF	2	2	-	-	C\D
	4	0	1	-	246	A
UFFxa16H07						
	REF	2	2	-	-	D
	14	2	1	-	268	A
	13	0	1	-	266	A
	7	1	1	298	306	B
	8	0	1	-	287	C
	0	2	0	279	-	C
	5	0	1	-	273	C
	0	1	0	269	-	B
	2	0	1	-	269	B
UAFv7500						
	6	0	1	-	345	C
	3	0	1	-	342	B
	0	1	0	345	-	C
	13	3	2	348	333,348	AB
	7	1	1	336	330	C
	17	3	3	342(A)	342(A)	A D
CO817823						
	0	0	1	-	199	B
	14	1	1	216	203	C
	15	1	1	203	193	B
	9	1	1	199	195	BC
	8	3	2	236	-	A D
	8	0	1	-	236	A
	0	1	0	195	-	C
	4	1	1	207	209	A
CX661101						
	0	2	0	-	-	C\D
	18	2	2	-	-	D\C
	0	1	0	223	-	B
	0	1	0	223	-	A
	21	2	3	220(B)221(A)	224	AB
	11	0	1	-	223	A

* Homoeologue assignment can be switched. REF=used as reference allele

I. Qualitative interpretation of microsatellite data

As a first step, we identify all alleles that segregate in a qualitative fashion, i.e., presence/absence. By filtering for the presence and absence of these alleles, it is possible to identify homologous (repulsion) alleles, thereby forming allelic pairs.

Example: Microsatellite 1 - UFFxa16H07: Primer pair UFFxa16H07 amplifies nine different alleles (Table 2 & 3, additional file 2), and six of these segregate qualitatively. Thus, it is possible to establish that five of these six alleles are homologous: allele 273K is in repulsion to allele 287K, and 269HK is in repulsion to 298H and 306K. Qualitatively determined allelic pairs therefore include 273K-287K, 269H-298H and 269K-306K. No repulsion allele has yet been found for 268K**.**

II. Quantitative interpretation of microsatellite data

a. Identification of reference alleles

The ideal reference allele meets the following criteria: No variation in dose, present in all progeny and not influenced by stutters from other alleles. Stable references for all progeny come from alleles that are homozygously present in one or both parents (i.e., AA x aa or AA x AA, not AA x Aa). Usually, these alleles can only be distinguished through the use of other initial, less optimal reference alleles, such as simplex alleles that segregate 1:1 for presence and absence.

Example: Microsatellite 1 - UFFxa16H07: Allele 262HK is present in all progeny and in both parents. Ratio clusters with several 1:1 presence/absence-segregating alleles reveals that allele 262HK does not segregate by dose and is therefore homozygous in both parents. This allele can thus be used as a reference allele.

b. Ratio calculation, cluster identification

Allele doses are estimated by examining the peak area of an allele relative to the peak area of the reference allele. The resulting Ratio Values (RVs) are plotted in frequency distributions (see additional files 1–5). When the reference allele performs well, the RVs show an apparently normal frequency distribution around a certain mean RV. This cluster of ratio values is called a ratio-cluster (RC). The identification of more than one clear RC in the progeny implies the presence of dosage variation, which in

turn indicates segregation. Narrow, non-overlapping frequency distributions around RCs indicate that the reference allele performs well. The obtained RVs for each allele are multiplied by an empirically obtained factor to make the range of RVs similar for each allele and facilitate the generation of frequency distributions.

Example: Microsatellite 1 - UFFxa16H07: Alleles 266HK and 279H are always present in all progeny. 266HK has two RCs with mean values of 17 and 39 (Table 3) and cluster distributions that allow dosage quantification (Figure 1), reinforcing the suitability of 262HK as reference. The RCs show regular steps, with the higher mean approximately two times greater than the lower mean. 279H also shows two RCs with mean values of ten and 15 (Table 3) and thus also segregates in dose.

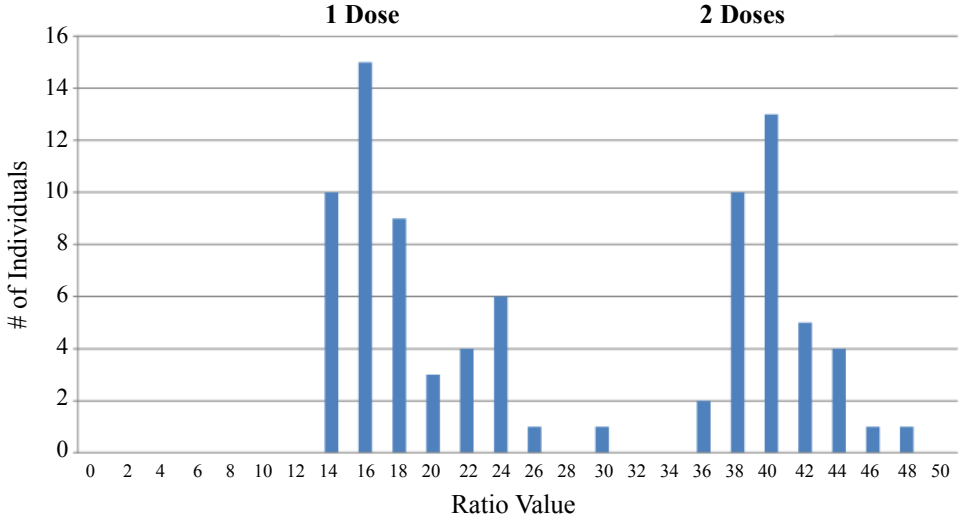


Figure 1. Example of the use of Ratio Value frequency distributions. Frequency distribution of ratio values for allele 266 of microsatellite UFFxa16H07 in the progeny from octoploid strawberry cross H x K

c. Segregation of Ratio Clusters

Segregation patterns of ratio clusters (RCs) are established by comparing the number of individuals in a certain RC to the numbers in other RCs. Common RC segregation ratios for presence/absence alleles are: 1:1 (Aa x aa), 1:2:1 (Aa x Aa) and 1:3:3:1 (AaBb x Aabb, where A and B are homoeologues). These ratios also occur for alleles that are always present due to homozygosity in at least one of the parents but that

still segregate in terms of dosage. For instance, the RC segregation ratio of 1:2:1 for an allele that is always present (and therefore must be present on more than one homoeologue) can be caused by several parental allelic configurations, such as AaBB x aaBb, AAbb x AaBb, AABb x aaBb and AABb x AABb. The allele doses of these parents are 3×1, 2×2, 3×1 and 3×3, respectively.

Example: Microsatellite 1 - UFFxa16H07: The two clusters of both 266HK (Figure 1) and 279H segregate in a 1:1 pattern (Table 3).

d. Identification of homologous alleles

Homologous allele pairs are identified in a manner similar to the qualitative method, but allele dose is considered instead of presence and absence. If an allele is present in its highest dose, this will automatically lead to a decrease in dose or absence of the homologous allele, and vice versa. If no such allele is found, it is likely that the repulsion allele is a null allele.

Example: Microsatellite 1 - UFFxa16H07: The highest dose of 266HK leads to the absence of simplex allele 268K, for which no repulsion allele could be found through qualitative analysis. This led us to establish 266K-268K as an allelic pair. Because 266HK is present in all progeny, our analysis on the two previous homologous 'Korona' alleles implies that 'Holiday' is 266-homozygous. We have now completed the allelic pairs for all four 'Korona' homoeologues and three out of the four 'Holiday' homoeologues. The only remaining allele to be analysed is 279H, which will be addressed in the next section.

When one or both parents have multiple doses of an allele, it is necessary to establish the dose of each allele to properly genotype individuals and identify allelic pairs. Allele doses are estimated by examining the peak area of an allele relative to the peak area of other allele(s) from the same PCR amplification, similar to the procedure published by Esselink et al. [8]. A large, consistent variation in the ratio between the peak areas of two different alleles indicates the segregation of allele dose for one or both alleles. The quantitative interpretation method is divided into four stages, which are described in the following sections.

III. Determining parental allele configuration and cross checking

The purpose of this step is to determine the allelic configuration of the mapping parents through their progeny. These configurations are subsequently validated using

the estimated allelic doses of the parents and vice versa. First, maternal allele pairs are matched with their paternal homologues using the alleles shared between parents. For example, if one parent (P1) has a homozygous allele of size (223,223) and the other parent (P2) has a heterozygous pairing between 223 and 225, (223,225), then the allele pairs P1(223,223) and P2(223,225) are considered homologous. Similarly, if an allele is homozygous in both parents (e.g., P1(229,229), P2(229,229)), these pairs are also considered homologous. When no alleles are shared between parents, matching can sometimes be achieved based on other criteria, such as differences in amplification efficiency between homoeologues or differences in the range of allele sizes. Additionally, when three sets of matching parental allele pairs have been identified for an allo-octoploid, the fourth set automatically consists of the last two remaining allele pairs. The process of matching parental allele pairs is continued until all sets are matched; these sets represent the homoeologues. Next, the number of obtained allelic sets is compared with the number of allele sets expected based on the ploidy level. If fewer sets are obtained, one of the homozygous alleles may actually be present in two homoeologues (A & B), e.g., P1(229A,229A)P2(229A,229A) +P1(229B,229B)P2(229B,229B). Alternatively, there may be a homoeologue that contains only null alleles, such as P1(229A,229A)P2(229A,229A) and P1(0B, 0B) P2(0B, 0B). Establishing whether a homoeologue is homozygous null or homozygous for a shared allele can be difficult. Examining the consistency of the allele doses inferred between the parents and progeny may be helpful, as long as the alleles have similar amplification efficiencies. For example, in an AAxAa cross, the RV of the mother should be approximately twice that of the father (2:1), and these RVs should be in agreement with the two RCs of the progeny (AA, Aa). If the RVs have a 4:3 relative value, this could indicate the presence of an additional homozygous set of allele pairs that had not been initially identified (i.e., AABBxAaBB). In some cases, the number of alleles can exceed the ploidy level. In such situations, indications for duplications or amplification efficiency differences should be examined.

Example: Microsatellite 1 - UFFxa16H07: Based on shared alleles (262HK, 266HK, 269HK) we can group the allele pairs of 'Holiday' and Korona into three sets of homoeologues (Table 3). The fourth set of homoeologues depends on the analysis of 279H. This allele shows two RCs and thus segregates by dose. However, the presence of a high dose of 279 is not associated with the absence of any other allele. Therefore, a logical hypothesis is that the homologous allele for 279 is a null allele and that 279 is homozygous on another homoeologue. However, this conclusion is problematic because it results in nine allele doses in a single parent ('Holiday')—six allele doses from the three allele pairs and three doses from

279—but only eight doses are possible in an allo-octoploid (when alleles are non-duplicated). An alternative is to assume the presence of two homologous alleles that differ in amplification efficiency (279a and 279b). The decrease in amplification efficiency leads to a low mean RC value (10) for one allele and a high mean RC value (15) for the other allele (Table 3). Because they are homologous, one of them is present in all progeny. Further evidence corroborating this hypothesis is that the RV of 279 in the 'Holiday' cultivar is much higher than the highest RC mean of the progeny; if the first hypothesis is correct, the 'Holiday' RV value should be similar to the highest RC mean of the progeny. In the alternative interpretation, the total dose in both 'Holiday' and 'Korona' produces eight doses per parent. Therefore, the 279 allele pair 279aH-279bH is joined with the last remaining 'Korona' allele pair, 273K-287K.

A virtual reference allele: Increasing throughput and power

Principle

The reference allele in the example above is based on a single allele. Using single allele based references can be a laborious procedure, especially if no monomorphic, homozygous loci that are not confounded by stutter bands can be found. Throughput can be increased considerably for most microsatellites through the use of a virtual reference allele that is based on the average of all, or part of, the allele peak areas for a primer pair in an individual. Automated calculation of averages and Ratio Values enables rapid dosage assessments. Because it is based on multiple alleles, a virtual reference allele diminishes the risks of inaccurate area estimation for individual alleles and reduces the impact of stutters influencing single peaks. Paradoxically, this makes dosage estimation better with increasing ploidy levels because the average is based on a larger number of observations.

Checking performance

Checking the performance of the virtual reference is best achieved by checking the narrowness of the RC distributions generated for alleles of known dose. If their width is narrow, the designed virtual reference is adequate. If they are wide, the interpretation of the data could be hampered. In that case, further optimisation of the virtual reference can usually be performed by accounting for putative interfering factors, such as segregating null alleles and large amplification differences between alleles. Interfering alleles can be identified by excluding suspect alleles from the average and then checking for improvement in the narrowness of the RC distributions.

Impact of differences in allele amplification efficiency

Efficiency of PCR amplification decreases with increasing allele size [10]. Furthermore, the presence of mutations in the primer sites can have a large influence on amplification efficiency. Consequently, homologous simplex alleles may have a many-fold difference in peak area (here, two- to three-fold, as shown in Table 3). This can have a significant influence on the reliability of dose estimation. As an illustration, assume a primer pair amplifies four different alleles over a tetraploid mapping population. Seven of the eight parental alleles amplify with equal efficiency, whereas the eighth allele has a three-fold higher efficiency. If this difference in amplification is not accounted for, the average of half of the tetraploid progeny represents six 'amplification units' ((3×1)+(1×3)), and the other half represents four units (4×1). The average-based reference in individuals carrying the efficient allele will thus be 50% higher than in individuals lacking this allele. This variability in the reference will greatly affect the width of the ratio clusters and may lead to false interpretations. In cases of too-wide RCs, using a subset of alleles with similar amplification efficiencies usually improves the width of the reference allele considerably. If this approach still does not suffice and if the involved markers are of great interest, a more sophisticated but also much more laborious approach may be considered: the introduction of allele-specific multipliers. Such multipliers are calculated for strongly deviating alleles based on their ratio to an initial reference allele with a known dosage. The inverse of the mean value of this ratio across the progeny can then be used as multiplier in the calculation of the virtual reference.

Example: Microsatellite 2 - UAFv7500: Primer pair UAFv7500 amplifies six different alleles, four of which segregate by presence/absence (Tables 2 & 3, additional file 3). Qualitative analysis reveals three 1:1 segregating alleles (330K, 333K, 336H) and one 3:1 segregating allele (345HK). 345HK is allelic with both 330K and 336H, indicating that they belong to the same homologous set (336H-345H, 330K-345K). No allelic pair can be found for 333K. Using the simplex alleles as references, all of the alleles that are always present segregate by dose. This makes it impossible to use a single peak as a reference for all the samples. Therefore, an average-based virtual reference allele must be constructed. Examination of the RCs of the simplex alleles showed narrow distributions (additional file 3), and thus, no further optimisation of the reference allele is required. Next, 345HK shows a 1:2:1 RC segregation, which confirms the qualitative analysis of 3:1 segregation. 342HK and 348HK are always present, indicating the presence of at least one homozygous locus for each. 342HK shows four RCs of regular increases (approximately 2×, 3× and 4× the ratio of 6) and a 1:3:3:1 segregation (Table 3, Figure 2). In allopolyploids, this segregation pattern

indicates triple heterozygosity, for which at least three heterozygous allele pairs must be involved in at least two homoeologues. Allele 348HK segregates 1:2:1 in RCs, indicating double heterozygosity. When samples are filtered for the highest RC of 342HK, the single-dose 333K allele is absent, and 348HK is at its lowest RC value. In contrast, when samples are filtered for the lowest RC of 342HK, 333K is always present and 348HK has its highest RC value. 342HK is thus fully complementary to 333K (333K-342K), and both of the other two segregating 342HK alleles are complementary to both of the segregating 348HK alleles. Next, the homozygous alleles must be assigned to their parents. To accomplish this, the ratio values of the parents are explored. For 342HK, these ratios indicate the presence of three doses in 'Holiday' and two in 'Korona' (Table 3). Because one of the two 342 alleles of 'Korona' is known to segregate due to its complementarity to 333K, the second 342K allele also has to segregate. The homozygous 342 locus and the third segregating 342 allele must thus originate from the 'Holiday' parent. The deduced allelic composition of the two involved homoeologues is thus (342H-342H, 333K-342K), (342H-348H, 342K-348K). This completes the analysis for 342HK.

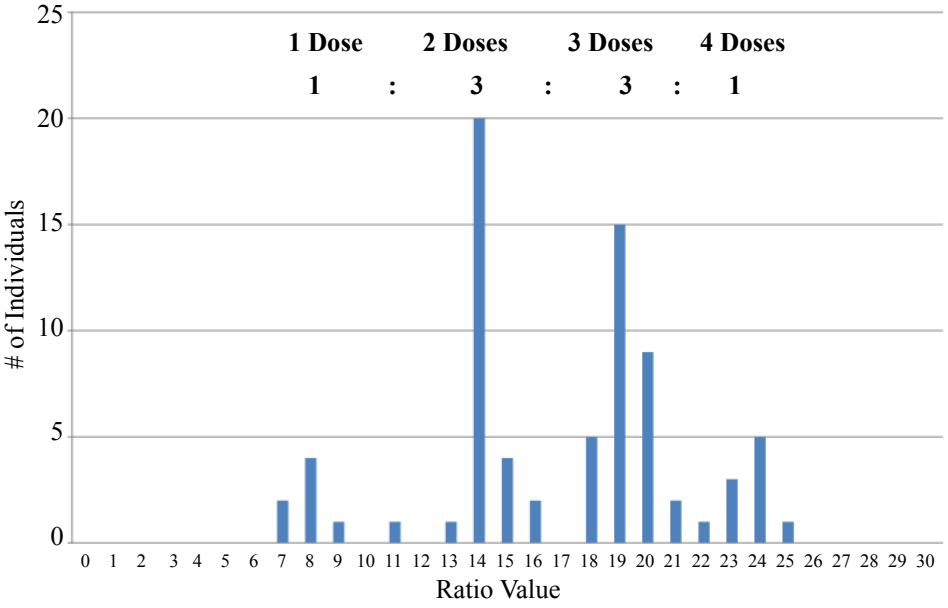


Figure 2. Example of the use of the Ratio Value frequency distributions. Frequency distribution of ratio values for allele 342 of microsatellite UAFv7500 in the progeny of octoploid strawberry cross H x K. The relative amplification ratio was calculated against a virtual average-based reference allele

At this point, the homozygous 348 allele has not yet been assigned and is assumed to be present on the 4th homoeologue. Because no other unassigned alleles are available, both parents may be 348-homozygous, or one may be homozygous null. The parental ratio values are helpful to distinguish these two options. They indicate the presence of three doses in each parent (Table 3), whereas only one segregating allele per parent has been assigned based on the segregation patterns of the progenies. This indicates that both parents are homozygous for 348 at the fourth homoeologue.

IV. Mapping and Validation

Having assigned alleles to homologous sets, we can now begin mapping. The mapping step serves five purposes: to validate and complete the assignment of allelic pairs, to group allelic pairs from multiple loci together into different homoeologous sub-genomes, to determine marker order and genetic distances, to establish an integrated map, and to determine the parental haplotypes. The mapping step is divided into four stages that are similar to the map integration procedure described by Barrett et al. [19]. The final validation of the MADCE-derived allele scores and homoeologue assignments is accomplished through the generation of linkage maps.

a. Creation of a priori integrated loci

The single parental allelic sets identified during the previous allele configuration step (step III) are combined into bi-parental sets and then translated into integrated loci. In our case, these loci were defined as Cross Pollinator (CP)-type loci in the software JoinMap (Kyazma B.V., Wageningen, NL). The integrated loci are constructed prior to mapping for two reasons. First, this serves as an additional check for scoring errors (e.g., individuals with three alleles for one locus). Second, it facilitates data-management; it is much more efficient to generate integrated loci first and use JoinMap 4.1 to convert them back to single parent loci when needed than to integrate loci at a later stage.

b. Creation of separate parental maps: additional error checks

To create separate parental maps, integrated loci are automatically converted into single-parent loci in JoinMap 4.1. After the maps are generated, a number of standard error checks are performed, such as comparison to a reference map (for strawberry FvxFb [20]) and a check for distorted loci and loci that create high tension. Finally,

marker genotypes are conditionally formatted in Excel using the phase information from JoinMap, which enables the easy identification of discover double crossover events.

c. Creation of separate parental maps: validation of integration and homoeologue assignment

Next, a new grouping is calculated and maps are drawn. The single parental maps are matched to each other based on allele sharing and can be used to validate integrated loci and create new ones from previously unintegrated loci, as described by Barrett et al. [19]. The pairs of matching parental maps are then randomly assigned a homoeologue letter (A, B, C, or D in the case of an octoploid).

d. Creation of integrated maps: final error check

The upgraded data from the previous step are loaded into JoinMap, and the final integrated maps are generated. The map and linkage phase information generated by JoinMap is combined with the allelic information from step III to establish the parental haplotypes.

Example: The final marker scores of the five example primer pairs (Table 1) for use in JoinMap are given in Additional files 1–5. The mapping results and final haplotypes are presented in Figure 3. The marker order is the same for all homoeologues, although some contain fewer or no segregating loci. Graphical analysis of the marker data shows the absence of any double recombination event over short distances, indicating that the quantitative interpretation of the microsatellite data was successful, as it led to unambiguous data. MADCE thus resulted in highly robust marker genotypes.

V. Completion of parental haplotypes for homozygous loci

If the presence of homozygous loci becomes evident in step III, the loci are manually added to the haplotype information to make the final haplotypes complete. Their genetic position is extrapolated or interpolated from the distances between markers on homoeologous genomes. These positions should not be considered biologically "true", but they can serve as a guideline to allow visualisation of homozygous stretches.

Example: In the current study, one of the homoeologues (1D) is shown to be completely homozygous (Figure 3).

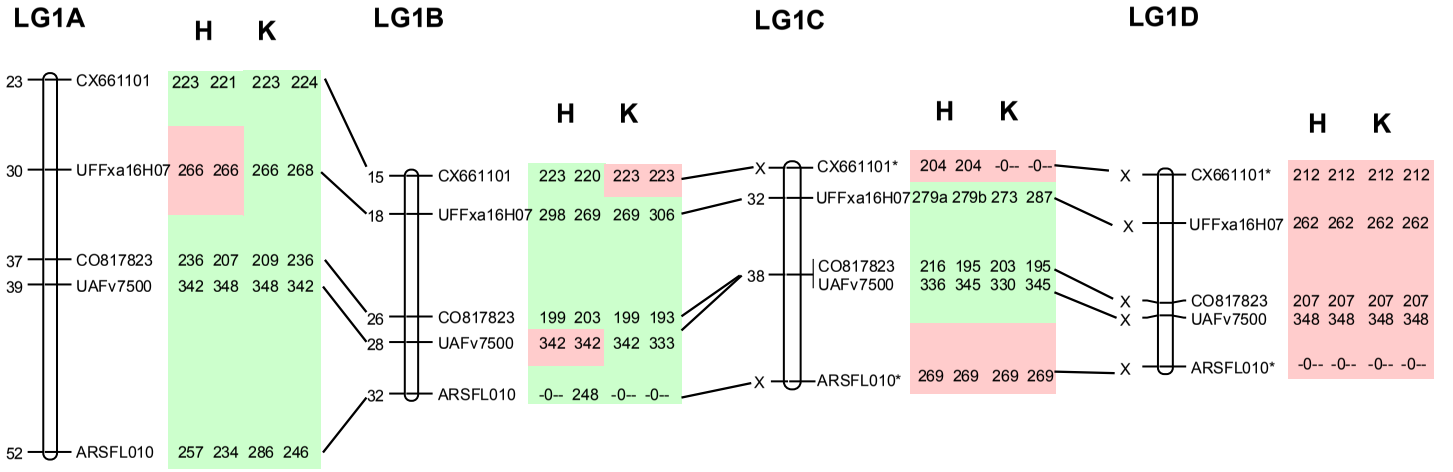


Figure 3. Linkage maps for the four homoeologous chromosome pairs of LG1 in the allo-octoploid strawberry.These maps are based on the five microsatellites that have been used to exemplify the MADCE approach for the assessment of allele doses and assignment of sub-genomes based on parent-specific allelic pairs. Positions (in cM) are shown on the left side of each linkage group (LG). Positions with an X instead of a value are homozygous in both parents and therefore cannot be used for determining map distances. A green background indicates heterozygosity, and a red background indicates homozygosity. H and K indicate the 'Holiday' and 'Korona' parents, respectively. * Allelic configurations are interchangeable between the related homoeologues. 0: inferred null allele

MADCE procedure for pedigreed cultivars

The MADCE procedure for mapping studies is not directly applicable to pedigreed cultivars because of their higher genetic complexity (e.g., a greater number of alleles per locus) and the very limited availability of segregation data (few progeny per individual). Here, we will describe the MADCE procedure as adapted for pedigreed cultivars and illustrate this procedure for microsatellite CX661101 (Table 4) using a set of 21 pedigreed cultivars (Figure 4).

Table 4. Dose assessment, Ratio Values and homoeologue assignment for CX661101 in a pedigreed set of cultivars

Allele		204	210	212	218	220	221	223	224	225	229	Dose Sum	Null Alleles
Homoeologue Assignment		C	C	D	D	B	A	AB	A	B	A		C
E-00188	D	**1**	**1**	**2**			**1**	**1**		**2**		**8**	
	RV	6	6	14			6	5		6			
E-03133	D	**1**		**2**				**3**		**1**		**7**	**1**
	RV	6		14				20		4			
E-93025	D		**1**	**2**				**3**	**1**			**7**	**1**
	RV		4	17				18	5				
'Elsanta'	D	**2**		**2**				**2**	**1**	**1**		**8**	
	RV	11		15				10	4	3			
'Fairland'	D	**2**		**1**	**1**			**3**	**1**			**8**	
	RV	9		7	6			16	5				
'Figaro'	D	**2**		**2**				**3**		**1**		**8**	
	RV	10		15				19					
'Gorella'	D	**2**		**2**				**2**	**1**	**1**		**8**	
	RV	9		14				12	5	3			
'Holiday'	D	**2**		**2**		**1**	**1**	**2**				**8**	
	RV	9		13		6	6	11					
'Induka'	D	**1**		**2**				**3**	**1**			**7**	**1**
	RV	6		16				15	7				
'Jerseybelle'	D	**2**		**2**			**1**	**3**				**8**	
	RV	9		14			6	15					
'Korona'	D			**2**				**3**	**1**			**6**	**2**
	RV			18				18	8				

Allele		204	210	212	218	220	221	223	224	225	229	Dose Sum	Null Alleles
'Pajaro'	D	**1**	**1**	**2**				**2**	**1**	**1**		**8**	
	RV	5	5	15				11	5	3			
'Polka'	D	**1**		**1**	**1**			**2**	**2**			**7**	**1**
	RV	7		7	7			12	11				
'Raritan'	D	**2**		**2**			**1**	**3**				**8**	
	RV	10		13			6	15					
'Redglow'	D	**1**		**2**				**3**	**1**			**7**	**1**
	RV	7		16				16	5				
'Senga S.'	D		**1**	**2**				**3**	**1**			**7**	**1**
	RV		5	15				17	6				
'Sivetta'	D	**2**		**1**	**1**			**1**	**1**	**1**	**1**	**8**	
	RV	11		8	6		1	4	5	3	6		
'Sonata'	D	**1**		**1**	**1**			**3**	**1**			**7**	**1**
	RV	6		8	7			18	5				
'Tago'	D	**1**		**2**				**3**			**1**	**7**	**1**
	RV	5		16				17			6		
'Talisman'	D	**1**		**2**				**3**			**1**	**7**	**1**
	RV	5		14				19			6		
'Tamella'	D	**1**		**2**				**4**				**7**	**1**
	RV	6		16				23					

D dose, *RV* ratio value. Homoeologue assignment of the underlined homoeologue letters was performed based on results from the HxK mapping population

I. Allele dose estimation requires a virtual reference

We recommended using virtual, average-based reference alleles because single alleles that fit the requirements for a reference are very rare in a large breeding germplasm. Ratio value (RV) calculation, frequency distribution and dose assignment are the same as for mapping populations. When an RV falls between two different clusters, it can be provisionally assigned the most likely dose and verified during step II. The occurrence of null alleles is more likely over a wide germplasm. Null allele doses are determined by subtracting the total dose estimated from the total dose expected (i.e., the same as the ploidy level). A separate column for null allele doses is added to the spreadsheet (Table 4). These null alleles can be assigned to homoeologues in a fashion similar to regular allele assignment, through the method described in the next section. For pedigreed cultivars, null alleles cannot be compensated for a priori.

Allele efficiency differences can only be compensated for with multipliers when they are very clearly present and can be quantified. These drawbacks make dosage estimation somewhat less reliable in cases where these pitfalls are present, but it is not impossible.

Example: For microsatellite CX661101, an average-based reference is created. The presence of distinct RV-clusters indicates the occurrence of different dose levels and a good performance of the virtual reference. The regular distances between RV clusters indicate the absence of very large differences in amplification efficiency between alleles (Table 4). For allele 204, we observe one to two doses with RCs, near RVs 5 and 10, respectively. Allele 212 is always present and also shows two different RCs (~8 and ~16), again suggesting a single vs. a double dose. Allele 223 has RVs ranging from 4 to 23, suggesting a more complicated situation. The RCs appear to increase and decrease in steps of approximately 5–6 (although the number of observations per cluster is quite low). This step size is therefore likely to represent the single dose value. From this we infer the dose levels of each cultivar, which ranges from 1–4 doses. Because we have mapped this microsatellite in 'Holiday' and 'Korona' (Table 3), we can use these cultivars as a check for correct dose estimation, and we find that the results are in agreement. Now, we can proceed to infer null allele doses, which we know to be present from the mapping study. The total dose of amplified alleles ranges from six to eight (Table 4), therefore, the total dose of null alleles ranges from two ('Korona') to zero. Not surprisingly, the ratio values for 212K and 224K are the highest of those in all cultivars and selections (after taking into account allele dose). This is due to the presence of the two null alleles, which led to a relatively low virtual reference value and therefore high ratio values.

II. Assigning alleles to homoeologues

a. Procedure: Assignment by total homoeologue allele dose

A very simple rule for disomically inheriting species is that the total dose of all alleles belonging to the same homoeologue in any given individual needs to be exactly two (null allele doses are also used in this calculation). If an allele of the same size is shared by multiple homoeologues, the total dose needs to be two times the number of homoeologues involved. These rules are a good starting point for situations in which no prior allele assignment from previous mapping studies is available. They are especially useful for large data sets and for alleles that occur at frequencies of>5%. We begin the allele assignment process by choosing the first allele and then

eliminating alleles that cannot be on the same homoeologue because the sum of their combined doses exceeds two. It is easiest to start with a high-frequency allele that varies in dose from zero to two. Alleles for which the summed dose never exceeds two are likely to be allelic to each other. This analysis is repeated starting with another unassigned allele to determine the composition of the other homoeologous series.

Example: We assume that there is no prior allele assignment information from mapping populations, except for the naming of the homoeologues. We begin the analysis with allele 204. This allele segregates by presence/absence and, when present, occurs at either a single or double dose (Table 4). Assuming that 204 occurs on a single homoeologue, we can, for each genotype, sum its dose with those of each of the other assumed homoeologue-specific alleles. Alleles that lead to a sum that is higher than two in any individual do not belong to the same homoeologue. This is the case for all alleles except 210 and the null allele. Summing the doses of these three alleles results in all genotypes having exactly two doses, indicating that allele assignment for this homoeologue is complete. This set of alleles is assigned homoeologue letter C based on the results of the mapping study.

We continue with a similar analysis of allele 212. The total dose exceeds two in combination with all alleles, except for 218. This allele is only present when 212 is at single dose, so 212 and 218 are likely to be homologous. The total dose sum for 212 and 218 is exactly two for all cultivars, and therefore, assignment for this homoeologue is complete. This set of alleles is assigned homoeologue letter D based on the results of the mapping study.

We observe that allele 223 is always present and occurs at up to four doses. It must therefore be present on at least two homoeologues. Because two homoeologues have been assigned already, 223 must be present on the other two (A&B). The remaining alleles must therefore be present on these two homoeologues as well. Of these, only 225 and 224 occur at a double dose. The assumption that double-dose (maximum) alleles are present on only one homoeologue eliminates the possibility that 221 is on the same homoeologue as 225 because they add up to three doses in selection E-00188. Allele 221 is present on homoeologue A in 'Holiday', so we assign 225 to homoeologue B. For double dose allele 224 ('Polka'), no eliminations can be achieved because 'Polka' carries only the shared allele 223. The remaining alleles (220, 224, 229) cannot be assigned through the dose procedure.

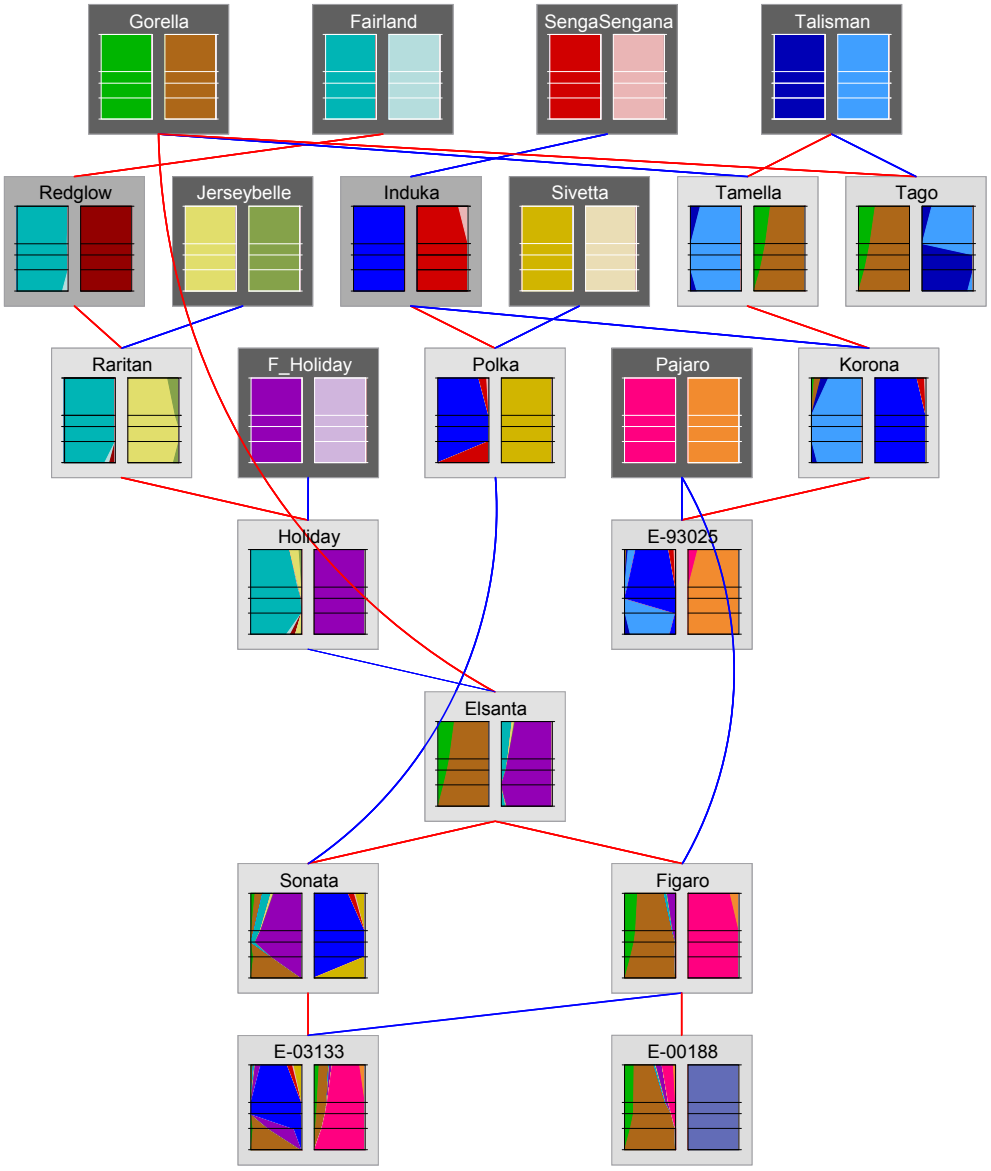


Figure 4. Genetic relationships among the 21 pedigreed strawberry cultivars and breeding lines, and a graphical representation of FlexQTL IBD (identity-by-descent) probabilities for LG1C. Dark grey, medium grey and light grey boxes represent cultivars for which none, one or both parents are included in the current survey. Haplotypes are represented by the coloured boxes, and each colour represents a different founder haplotype. Red lines indicate maternal parents and blue lines indicate paternal parents. Horizontal lines within a coloured box represent the positions of five different marker loci (from top to bottom: PSContig944, CX661101, UFFxa16H07, UAFv7500 and ARSFL010). The width of a colour at a particular height reflects the probability that the corresponding founder allele is present at that locus on the map (see Voorrips 2007)

b. Procedure: Assignment by transmission

To help in completing or validating assignments, we can consider allele transmission from parents to progeny. For alleles that are on the same homoeologue, only one allele should be transmitted. Simultaneous (non)transmission of alleles indicates that they belong to different homoeologues. By tracking the transmission of alleles over a large pedigree, it is possible to establish which alleles definitely do not belong to the same homoeologue and which alleles are likely to belong to the same homoeologue. The above principle does not apply to alleles that are shared between homoeologues.

Example: Using the pedigree shown in Figure 4 in combination with the data in Table 4, we observe that alleles 224 and 225 are both present in 'Gorella' and that neither of these alleles are transmitted from 'Gorella' to 'Tamella'. This indicates that these two alleles are present on different homoeologues. This is corroborated by the transmission from 'Gorella' to 'Elsanta', in which both are transmitted and neither could have come from 'Holiday', which is the other parent of 'Elsanta'. Because 225 was assigned to homoeologue B, we can now assign 224 to homoeologue A. Similarly, we find that 221 and 220 are both present in 'Holiday', but neither are transmitted to 'Elsanta'. This means that 220 is not homologous to 221 and therefore must be on homoeologue B. This leaves only 229 unassigned; this allele is present in 'Sivetta' along with 225 of homoeologue B. Because neither of these two alleles is transmitted to 'Polka', 229 must be present on homoeologue A. We thus obtained two allelic sets: 221-223-224-229 for homoeologue A and 220-223-225 for homoeologue B.

Once all alleles have been assigned, we can genotype the A and B homoeologues for the 223 allele. For instance, 'Polka' has 223 and 224 at a double dose. Because 224 occurs only on the A genome, 'Polka' must be homozygous for 224 on the A homoeologue and must thus be homozygous for 223 on the B homoeologue. 'Elsanta' has 223 at double dose and 224 and 225 at single doses. Alleles 224 and 225 belong to different homoeologues. Therefore, the double dose of 223 must come from the two different homoeologues (223A-224A, 223B-225B). Finally, we confirm whether the assignments are consistent with those of 'Holiday' and 'Korona' from the mapping population.

Results for four other example microsatellites:A similar analysis has been performed with five other microsatellites (Table 1). For each, a summary of the resulting allelic sets is presented in Table 5. For one microsatellite (CO817823), we encountered difficulties in dose estimation in a few cultivars, and it was not possible to complete the analysis of this microsatellite. Twelve alleles occurred at multiple

doses (seven at 2×, four at 3×, and one at 4×), all of which could be assigned to specific homoeologues.

With dose estimation completed, we can proceed to identify allelic pairs that belong to the same homoeologue (homologous alleles). This is best achieved by combining two methods, one based on allele dose and one on allele transmission.

Table 5. Microsatellite alleles observed in 21 cultivars of the allo-octoploid strawberry for six microsatellite markers and their assigned to the four homoeologues

Microsatellite	LG1A Diversity	LG1B Diversity	LG1C Diversity	LG1D Diversity	Max dose observed
ARSFL010	234,242,246,257, 286	null* (248 in 'Holiday')	244,**248**,252,259, 264,266,269	null	2× (234,257,269)
CO817823	205,207?,209,211?, 236	null?,193,199,**203**, 209?,211?	null?,195,**203**,216	null,205?, 207,210?	unknown
CX661101	221,**223**,224,229	220,**223**,225	null,204,210	212,218	4× (**223**)
PsContig944	null,**115**,152,156, **160**,161,162,169,173	**115?**,155,**160**,166,168, 179,181,184	**115**,137	152,158	3× (**115**)
UAFv7500	**342,348,351**	333,**342,345**	330,336,**345**	339,**345,348, 351**	3× (**342,345,348**)
UFFxa16H07	266,268	269,275,286,298,306	271,273,279,287	262	2× (262,266, 273,279)

Alleles in bold occur at multiple homoeologues. Alleles followed by a question mark have uncertain assignment. * Allele 248(HomB) was observed during mapping, but due to its low amplification it was not detected in the set of pedigreed cultivars.

III. Phase determination and assignment validation using FlexQTL™

The data are now used to generate linkage phase and Identity-By-Descent (IBD) estimations using the FlexQTL™ software [17]. FlexQTL™ also monitors the number of observed and expected single and double recombination events between successive markers, making this software an easy tool with which to quickly validate assignments and check for alternatives (erroneously assigned alleles lead to an increased number of apparent recombination events throughout the pedigree). Evidently, proper phase estimations can only be made in cases where a founder has sufficient offspring.

Example: Figure 4 presents the haplotypes obtained for the 21 studied cultivars. Figure 5 demonstrates the flow of individual alleles over the pedigree for a subset of these cultivars. The pedigree-derived haplotypes for 'Korona' are consistent with those determined by the mapping population (Figure3 and Figure5). This is also the case for the 'Holiday' cultivar (data not shown), thus delivering proof of concept of the suitability of MADCE for the genotyping of pedigreed germplasm in allopolyploid crops.

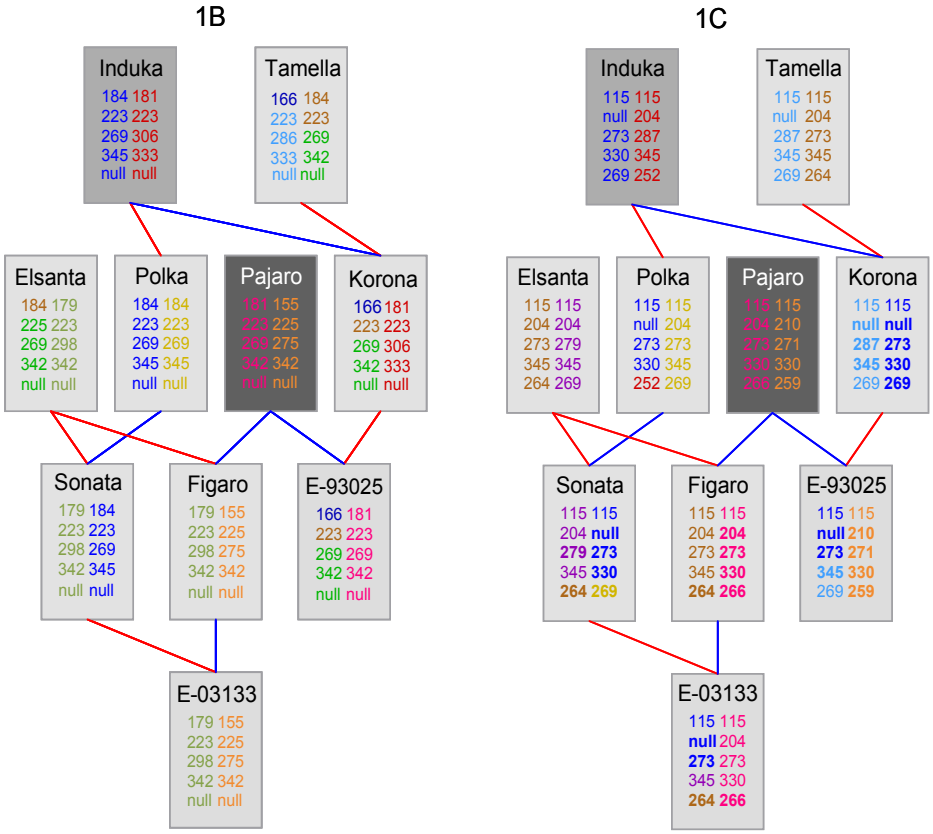


Figure 5. (left page) Subsection of the pedigree from Figure 3 showing the sizes and the most probable parental origins of alleles for the loci of LG1B (left) and LG1C (right). Red lines indicate maternal and blue lines paternal parents. Each horizontal pane describing allele sizes represents a different microsatellites (from top to bottom: PSContig944, CX661101, UFFxa16H07, UAFv7500 and ARSFL010). Different colours represent different founder haplotypes. Alleles in bold have an IBD value of >0.9. Dark grey, medium grey and light grey boxes represent cultivars for which none, one or both parents are included in the current survey

DISCUSSION

Methodological aspects of MADCE

A variety of allele dose estimation techniques have been examined in the past decade. The methods employed include visual evaluation of intensities [9,21,22], peak height [10] and peak area ratios (MAC-PR) [8,11]. MAC-PR [8] is likely to be the most accurate of these due to its use of an internal reference that allows compensation for PCR efficiency variation between samples. MADCE improves upon this method with the new concept of virtual (average-based) reference alleles and extends it with a pipeline for data analysis for the assessment of the full allelic configuration of allopolyploid genotypes. The robustness of MADCE is assured thanks to the presence of several internal feedback controls that monitor the accuracy and consistency of dosage information.

Virtual reference allele

The throughput and reliability of dose estimation is heavily dependent on the kind and quality of the reference allele and the "overall" quality of the microsatellite (stutters, peak shape, allele size proximity). The use of "virtual" average peak-area-based references is recommended for both mapping studies and pedigreed sets of cultivars because they are robust, are widely applicable and increase throughput. Virtual references are more robust because they are less sensitive to the quality of individual peaks and stutters. Virtual references increase throughput because a single fully informative reference can be used instead of a series of partially informative references, which would require more data handling. The performance of the virtual reference can be easily monitored by measuring the width of RC clusters, and the need for additional, performance improving measures can easily be detected. MADCE's use of a virtual reference provides a basis for semi-automated analyses including the development of dedicated software for dose estimation [12], thus enabling further increases in throughput in the future.

Homoeologous amplification

Within allopolyploids, primer pairs can potentially amplify products from one or all of the homoeologues and thus may yield a large number of alleles with complex banding and segregation patterns. This could be seen as a disadvantage because it hampers quick and easy data interpretation, especially in the assignment of alleles

to specific homoeologues. Alternatively, it could be seen as an advantage because it allows the examination of macro-synteny. In addition, fewer primer combinations are needed relative to methods that use homoeologue-specific primers.

The main disadvantage of homoeologous amplification is that alleles that are exactly the same size between multiple sub-genomes can occur. These shared alleles cause most of the complex segregation patterns in allopolyploids. Shared alleles could indicate the conservation of microsatellite repeat number between homoeologues but could also have arisen independently through different events (homoplasy) [23-25]. The case for homoplasy is corroborated by the fact that in our study, one of these shared alleles shows a difference in amplification efficiency (Table 3), indicating the presence of polymorphisms such as indels or SNPs at the primer annealing site, these types of polymorphism have lower mutation rates than repeats. Using MADCE, it is possible to tackle these obstacles by dissecting all alleles into simple Mendelian segregation patterns and assigning the alleles to their respective sub-genomes, despite their size similarities.

Improvement of mapping efficiency by MADCE

In our case study of the octoploid strawberry, five microsatellite primer combinations amplified 35 different alleles based on size, and 25 of these exhibited presence/absence segregation. Using MADCE, we uncovered 30 highly informative segregating alleles, whereas a conventional approach would have yielded only 19 fully informative and six partially informative alleles. Of the ten alleles that were present in all of the progeny and thereby ignored or considered homozygous when using conventional approaches, six were found to segregate by dosage. Additionally, the information content of the six 3:1 segregating markers increased because MADCE allows the discernment of homozygous and heterozygous progeny, thus doubling the proportion of informative meioses from 25% to 50%. By identifying the matching repulsion alleles, all four genotypic classes could be distinguished, thus increasing the meiotic information content to 100%. Moreover, one 3:1 segregating allele of one of these microsatellites could be correctly assigned to different homoeologues. If this allele had been used as a bridge marker for parental map integration, according to the protocol defined by Ripol et al. [6], this would have led to a false integration that would have linked homoeologous instead of homologous chromosomes.

Improvement of mapping quality by MADCE

The identification of homologous (repulsion) alleles and the creation of a single

locus for both alleles prior to mapping have been used previously [14]. This reduces the number of redundant loci, some of which might have looked like different loci at slightly different map positions due to inconsistencies in scoring or missing scores. The error removal steps of MADCE improve the final map quality. Falsely integrated parental maps that occur due to shared alleles are prevented by adequate allele assignment to the different homoeologues, as with the map integration method described in Barrett et al. [19]. The determination of the full allelic configuration of mapping parents by MADCE allows the easy identification of regions in the genome that are completely homozygous, as demonstrated for linkage group 1D. Information about the presence of large homozygous segments can reveal gaps and the occurrence of an excessive number of partial linkage groups when making linkage maps in outbreeding plant species. This information can thus prevent futile efforts to fill these gaps or complete these linkage groups by testing numerous additional primer combinations.

Enabling pedigree-based analysis

In allopolyploids, association studies are still in their infancy, despite the fact that many economically important crops, such as wheat and cotton, are allopolyploid. And despite that some of these plants have well-diverged sub-genomes [26-28] and propagate through inbreeding, which make them genetically less complicated. Over the last decade, there has been a shift from bi-parental QTL mapping studies towards studies on preferably unstructured plant germplasm through LD mapping [29]. Additionally, strategies have been developed for genetically structured breeding germplasm through a pedigree-based analysis (PBA) approach [17,30]. Both approaches offer perspectives for allopolyploids.

PBA is a QTL mapping approach for multiple pedigreed families, cultivars and selections. It allows the exploitation of known pedigree relationships and allows a relatively low marker density [30]. One informative microsatellite marker for every five cM is sufficient for PBA. PBA provides an understanding of the genetic structure of breeding germplasm and discovers the signature of breeders by pinpointing regions that are under high selection pressure without the need for familiarity with the involved traits. Proof-of-concept and statistical analyses of this methodology have been performed in apple through the EU-HiDRAS project [31]. Since then, PBA has been embraced in genetic research on diploid Rosaceae crops [32]. PBA requires advanced statistical software packages, such as FlexQTL™ [17] and Pedimap [18]. MADCE is able to deliver the data required by this software because it can be used to deduce the complete allelic configuration on all homoeologues for a

given microsatellite in a given plant, including null alleles and homozygous regions. MADCE thus enables the performance of PBA in complex allopolyploids for the first time. A precursor to PBA was provided by our test set of 21 pedigreed cultivars, in which we were able to follow the flow of marker alleles over pedigrees. MADCE is currently in use for the genotyping of an extensive set of breeding germplasm in strawberries.

Ability to distinguish disomic from polysomic inheritance

Knowledge regarding the type of polyploidy in a particular plant is critical because the mode of inheritance determines what types of methodology and software can be used in genetic studies. Conventional methods to assess the type of ploidy are based on cytogenetic studies of chromosome pairing behaviour during meiosis [33] and on segregation. The cytogenetic approach cannot provide absolute certainty because multivalents can also be observed in the early stages of meiosis in allopolyploids [34]. In segregation studies, ploidy types are distinguished based on i) segregation ratios for duplex alleles, ii) the occurrence of progeny-genotypes that can only arise due to double reduction, and iii) the ratio by which markers of linked loci are in coupling and repulsion phases [5]. None of these approaches can provide absolute certainty for ploidy type. Segregation ratios cannot provide certainty because it is difficult to distinguish segregation ratios that are greater than 3:1 from each other unless very large populations are used. Additionally, the occurrence of segregation distortion could interfere with these analyses. Double reductions are not reliable because false double reductions can be scored due to phenotyping errors, genotyping errors, outcrossing and DNA admixture. These types of errors generally occur at low frequency, but this is also true for actual double reduction events [35]. Finally, indications through linkage ratios led to the incorrect inference of mixed polysomy and disomy for the cultivated strawberry [13,36]. Moreover, conclusions about the mode of inheritance can only be made after linkage groups have been established, when it is already too late to take advantage of diploid methodology. Finally, this method uses coupling and repulsion phase linkages between loci rather than simply the repulsion allele within a locus. Therefore, as distance increases, the reliability of determining meiotic segregation patterns decreases. In contrast, MADCE offers a very effective approach for establishing the type polyploidy by confirming disomic inheritance prior to mapping through an examination of the presence of allelic pairs within a single locus. If such pairs are found for each of the different chromosomes or a sufficiently large set of random loci, fully functional disomy, and thus allopolyploidy, can be assumed.

MADCE and new high-throughput genomic tools

Rapid advances in the affordability and throughput of next generation sequencing technologies have sped up the development of high-throughput marker systems. Marker platforms other than microsatellites, including array-based SNP detection technologies [37], fluorescent SNP probes [12,38] and real time quantitative PCR [39], can also be used for dose estimation. These techniques are well developed, accurate and often high-throughput, whereas the use of microsatellites is relatively costly and labour intensive. These platforms are therefore likely to surpass SSR approaches in the near future. However, SNP arrays for allopolyploids are not frequently available because their development is quite complicated. Because sub-genomes are highly differentiated, most SNPs in allopolyploids are likely to be polymorphic between homoeologues but not within a homoeologue. In addition, for the relatively few SNPs that are polymorphic within a homoeologue and can therefore be used in assays, most will have interference from the other sub-genomes. To illustrate this, imagine how well an assay would need to perform to separate the clusters of signals from a SNP assay in an allo-octoploid (AATTTTTT vs. ATTTTTTT vs. TTTTTTTT). This interference would have to be dealt with, for instance by using adjacent sub-genome specific SNPs to make the assays sub-genome specific. The investment required for SNP development is therefore much steeper for higher-order polyploids than for diploids. Other argument for the continued use of microsatellites is that they are more suitable for application in genetically distinct germplasm due to their high transferability and level of polymorphism. This high level of polymorphism also makes them more likely to tag a haplotype (or trait) than random SNPs. Polyploidy hampers discovery of SNP haplotypes. MADCE could help in the discovery of haplotype tagging SNPs through their association with well-defined microsatellite alleles. Furthermore, information on homozygosity (for cross-pollinators) generated by MADCE enables the selection of the best lines for SNP development by complementing regions that are homozygous and heterozygous between these lines. Based on these perspectives, MADCE currently supports the efforts of the international RosBREED-Illumina consortium in testing SNP development strategies by helping to identify the most appropriate germplasm for use in SNP discovery.

CONCLUSION

The MADCE method for the quantitative interpretation of microsatellite data presented in this paper offers a novel tool for the genetic analysis of complex allopolyploid

plants. MADCE enhances the genotyping of allopolyploids by dealing with shared alleles between sub-genomes, null-alleles and homozygous loci. This can be used to establish the full allelic configuration of any individual allopolyploid plant. MADCE fosters genetic studies in allopolyploids by increasing the efficiency of generating molecular marker linkage maps and by enabling the fully informative genotyping of pedigreed breeding germplasm. MADCE thus enables the use of statistical and genetic software designed for diploid systems for allopolyploids. MADCE can also be used to aid SNP detection and SNP array development in complex polyploids.

Methods

Plant material

For the construction of a molecular marker linkage map, 82 seedlings from a cross between the strawberry cultivars 'Holiday' (H) and 'Korona' (K) were used. For the pedigree analysis, a set of 21 cultivars and breeding lines including 'Holiday' and 'Korona' was used (Figure 4). Leaves were sampled from the Fresh Forward Breeding germplasm collection or made available by the National Clonal Germplasm Repository at Corvallis, Oregon, US.

DNA isolation

Genomic DNA was extracted according to a modified version of the Fulton et al. [40] mini-prep protocol. Briefly, young, folded leaves were harvested. Leaves were freeze-dried and ground to powder in a 2 ml tube. To this tube, 700 µL of warm (65°C) 2% CTAB buffer was added, and the tube was mixed by vortexing and incubated for 10min. Next, 700 µL of chloroform:isoamyl alcohol (24:1) was added. The mix was centrifuged at room temperature at 10,000 × g for 2 min. Next, 600 µL of the top phase was transferred to a fresh tube. Isopropanol (480 µL) was added, and the sample was mixed and then centrifuged at 10,000×g for 2min at room temperature. The supernatant was discarded, and the pellet was washed with 500 µL of 70% ethanol, left for 2 min and then centrifuged at 10,000 x g for 2min. The supernatant was discarded by pipetting and the pellet was resuspended in 400 µL of TE. LiCl (135 µL, 8 M) was added to remove RNA and polysaccharides, and the mix was incubated for 30 min at -20°C. After incubation, the mix was centrifuged at room temperature at 10,000 × g for 2 min, and the supernatant was transferred to a fresh tube. Isopropanol (320 µL) was added and the mix was incubated at -20°C for 30 min. The mix was then centrifuged at 10,000 × g for 5 min and the supernatant

was discarded. The pellet was then washed and centrifuged twice with 500 µL of ethanol (70%), and then the dried pellet was dissolved in 50 µL of TE.

Microsatellites

Six microsatellite primer combinations known to be located on a single linkage group (LG1) were taken from a variety of sources (Table 1). Five were used both in the mapping and the pedigree analyses, whereas marker PScontig944 was only used for pedigree analysis.

PCR

PCRs were performed with indirect fluorescent labelling with an universal 17bp 5' end tail sequence (AACAGGTATGACCATGA) on the forward primer that matched a universal fluorescently labelled primer (6-FAM, HEX or ROX). This method was adapted from the protocol described in Schuelke [41]. Reverse primers had a GTTT tail added to them to reduce stutters, according to the protocol from Brownstein et al. [42]. The PCR mix was composed of 1X Goldstar PCR buffer, 0.5 µM unlabelled forward primer with tail, 2 µM of unlabelled reverse primer, 2 µM labelled universal tail primer, 0.3 U of GoldstarTaq polymerase (Eurogentec, The Netherlands) and 10 ng of DNA in a total reaction volume of 20 µL. The PCR conditions were one cycle at 94°C for 3 min, followed by 35 cycles at 94°C for 30 s, 50°C for 30 s and 72°C for 2 min, and a final extension cycle at 72°C for 10 min.

Microsatellite data preparation

Depending on amplicon intensity as observed from agarose gel, PCR products were diluted (on average approximately 300× in total) to prevent fluorescent intensity levels to exceed (or be below) the detection levels and thereby hamper dose estimation. Fluorescently labelled amplicons were separated and detected using an ABI capillary automated sequencing platform (ABI 3730, Perkin Elmer Biosystems, Foster City, California). Output from the ABI platform was analysed with Genemapper 4.0 software. Peaks corresponding to alleles were identified and their bin ranges, which are the window of allele sizes that are thought to represent a single allele, were defined. Next, for each sample, the software automatically identified the presence of alleles (peaks), their height and the area under the peak. Proper allele detection was checked manually and adjusted where necessary. The allelic data (size, area) for each individual was transferred to an Excel sheet (see additional files 1–5). Excel sheets

were formatted to show the area data for each individual in rows, and each column represents a different allele. These sheets were then used for further qualitative and quantitative analyses as described in the results section.

Construction of linkage maps

Linkage maps were created for each parent separately using JoinMap® 4.1 (Kyazma B.V., Wageningen, NL) and the Kosambi mapping function. Marker placement was determined using a minimum LOD threshold of 1, a recombination fraction threshold of 0.45, a ripple value of 1 and a jump threshold of 5. Comparisons of the separate parental linkage maps were used for the creation and validation of integrated loci and for error removal. After this data preparation step/upgrade, integrated maps were created using the same JoinMap settings as used for the separate parental maps. Positions for homozygous loci were estimated using interpolation and extrapolation of map distances of the segregating homoeologous loci. Drawings of the linkage maps were first created with the software packages Mapchart [43] and later finalised in Microsoft Powerpoint.

Pedigree analysis

The 21 cultivars from Figure 4 were genotyped using the six microsatellites in Table1. Dose and configuration of alleles were established according to the Microsatellite Allele Dose and Configuration Establishment (MADCE) method that was adjusted for pedigreed germplasm, as presented in this paper. Next, the inheritance of these cultivars over a pedigree was analysed using IBD estimates and the most likely linkage phases of alleles according to the software package FlexQTL™ [17]. To graphically represent the flow of alleles over the pedigree, IBD estimates were loaded into Pedimap[18].

Acknowledgements

We gratefully acknowledge Nahla Bassil of the National Clonal Germplasm Repository at Corvallis, Oregon, for supplying leaf material of cultivars ‘Fairland’, ‘Jerseybelle’, ‘Raritan’ and ‘Redglow’. We acknowledge Rene Smulders and Chris Maliepaard of Wageningen University department of Plant Breeding for critically reviewing this manuscript. Funding of this research was provided through the Technological Top Institute Green Genetics in collaboration with Fresh Forward Breeding B.V.

Additional files

Additional Files 1 through 8 are available online at:
http://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-12-25
doi: 10.1186/1471-2229-12-25

References

1. Soltis DE, Soltis PS, Bennett MD, Leitch IJ: **Evolution of genome size in the angiosperms.** American Journal of Botany 2003, 90(11):1596–1603.
2. Levin DA: Polyploidy and Novelty in Flowering Plants. Am Nat 1983, 122(1):1–25.
3. Gaut BS, Doebley JF: **DNA sequence evidence for the segmental allotetraploid origin of maize.** Proceedings of the National Academy of Sciences of the United States of America 1997, 94(13):6809–6814.
4. Soltis DE, Soltis PS: **Molecular-Data and the Dynamic Nature of Polyploidy**. Crit Rev Plant Sci 1993, 12(3):243–273.
5. Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD: **The Detection and Estimation of Linkage in Polyploids Using Single-Dose Restriction Fragments.** Theor Appl Genet 1992, 83(3):294–300.
6. Ripol MI, Churchill GA, da Silva JAG, Sorrells M: **Statistical aspects of genetic mapping in autopolyploids**. Gene 1999, 235(1–2):31–41.
7. D'hoop BB, Paulo MJ, Mank RA, van Eck HJ, van Eeuwijk FA: **Association mapping of quality traits in potato (Solanum tuberosum L.).** Euphytica 2008, 161(1–2):47–60.
8. Esselink GD, Nybom H, Vosman B**: Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method.** Theor Appl Genet 2004, 109(2):402–408.
9. Flajoulot S, Ronfort J, Baudouin P, Barre P, Huguet T, Huyghe C, Julier B: **Genetic diversity among alfalfa (Medicago sativa) cultivars coming from a breeding program, using SSR markers.** Theor Appl Genet 2005, 111(7):1420–1429.
10. Landergott U, Naciri Y, Schneller JJ, Holderegger R: **Allelic configuration and polysomic inheritance of highly variable microsatellites in tetraploid gynodioecious Thymus praecox agg.** Theor Appl Genet 2006, 113(3):453–465.
11. Nybom H, Esselink GD, Werlemark G, Vosman B: **Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dog-roses, Rosa L. Sect. Caninae DC**. Heredity 2004, 92(3):139–150.
12. Voorrips RE, Gort G, Vosman B: **Genotype calling in tetraploid species from bi-allelic marker data using mixture models.** BMC Bioinformatics 2011, 12(1):172.
13. Rousseau-Gueutin M, Lerceteau-Kohler E, Barrot L, Sargent DJ, Monfort A, Simpson

D, Arus P, Guerin G, Denoyes-Rothan B: **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 2008, 179(4):2045–2060.

14. Sargent DJ, Fernandez-Fernandez F, Ruiz-Roja JJ, Sutherland BG, Passey A, Whitehouse AB, Simpson DW: **A genetic linkage map of the cultivated strawberry (Fragaria x ananassa) and its comparison to the diploid Fragaria reference map.** Molecular Breeding 2009, 24(3):293–303.
15. Viruel MA: **An SSR and RFLP linkage map for the octoploid strawberry (Fragaria x ananassa).** In: Plant, Animal and Microbe genomes X conference: 2002; San Diego, CA; 2002: 660.
16. Sargent D, Passey T, Šurbanovski N, Lopez Girona E, Kuchta P, Davik J, Harrison R, Passey A, Whitehouse A, Simpson D: **A microsatellite linkage map for the cultivated strawberry Fragaria x ananassa suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection.** TAG Theoretical and Applied Genetics 2012. DOI 10.1007/s00122-011-1782-6
17. Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, de Weg WEV: **Bayesian analysis of complex traits in pedigreed plant populations.** Euphytica 2008, 161(1–2):85–96.
18. **Pedimap : Software for the visualization of genetic and phenotypic data in pedigrees** [http://www.plantbreeding.wur.nl/UK/software_pedimap.html]
19. Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G et al.: **A microsatellite map of white clover.** Theor Appl Genet 2004, 109(3):596–608.
20. Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM et al.: **An enhanced microsatellite map of diploid Fragaria**. Theor Appl Genet 2006, 112(7):1349–1359.
21. Truong C, Palme AE, Felber F, Naciri-Graven Y: **Isolation and characterization of microsatellite markers in the tetraploid birch, Betula pubescens ssp tortuosa.** Molecular Ecology Notes 2005, 5(1):96–98.
22. Wickramasinghe HAM, Miura H: **Gene dosage effect of the wheat Wx alleles and their interaction on amylose synthesis in the endosperm.** Euphytica 2003, 132(3):303–310.
23. Ellegren H: Microsatellites: **Simple sequences with complex evolution.** Nat Rev Genet 2004, 5(6):435–445.
24. Estoup A, Jarne P, Cornuet JM: **Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis**. Mol Ecol 2002, 11(9):1591–1604.
25. Levinson G, Gutman GA: **Slipped-Strand Mispairing - a Major Mechanism for DNA-Sequence Evolution.** Molecular Biology and Evolution 1987, 4(3):203–221.
26. Guo WZ, Cai CP, Wang CB, Han ZG, Song XL, Wang K, Niu XW, Wang C, Lu KY, Shi B et al.: **A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in gossypium.** Genetics 2007, 176(1):527–541.

27. Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW: **A microsatellite map of wheat.** Genetics 1998, 149(4):2007–2023.
28. Roder MS, Plaschke J, Konig SU, Borner A, Sorrells ME, Tanksley SD, Ganal MW: **Abundance, Variability and Chromosomal Location of Microsatellites in Wheat.** Mol Gen Genet 1995, 246(3):327–333.
29. Zhu C, Gore M, Buckler ES, Yu J: **Status and Prospects of Association Mapping in Plants.** Plant Gen 2008, 1(1):5–20.
30. Van De Weg WE, Voorrips RE, Finkers R, Kodde LP, Meulenbroek EJ, Jansen J, Bink MCAM: **Pedigree genotyping: A new pedigree-based approach of QTL identification and allele mining by exploiting breeding material.** In: Acta Hort. vol. 708; 2006: 483–488.
31. Patocchi A, Fernández-Fernández F, Evans K, Silfverberg-Dilworth E, Matasci CL, Gobbin D, Rezzonico F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M et al.: **Development of a set of apple ssrs markers spanning the apple genome, genotyping of HiDRAS Plant material and validation of genotypic data.** In: Acta Hort. vol. 814; 2009: 603–608.
32. Iezzoni A, Weebadde C, Luby J, Yue C, Van De Weg E, Fazio G, Main D, Peace CP, Bassil NV, McFerson J: RosBREED: **Enabling marker-assisted breeding in Rosaceae.** In: Acta Hort. vol. 859; 2010: 389–394.
33. Grant V**: Plant speciation**. New York, USA: Columbia university press; 1981.
34. Jenkins G, Rees H: **Strategies of Bivalent Formation in Allopolyploid Plants.** P Roy Soc Lond B Bio 1991, 243(1308):209–214.
35. Haynes KG, Douches DS: **Estimation of the Coefficient of Double Reduction in the Cultivated Tetraploid Potato.** Theor Appl Genet 1993, 85(6–7):857–862.
36. Lerceteau-Kohler E, Guerin G, Laigret F, Denoyes-Rothan B: **Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (Fragaria x ananassa) using AFLP mapping**. Theor Appl Genet 2003, 107(4):619–628.
37. Ishikawa S, Komura D, Tsuji S, Nishimura K, Yamamoto S, Panda B, Huang J, Fukayama M, Jones KW, Aburatani H: **Allelic dosage analysis with genotyping microarrays.** Biochem Bioph Res Co 2005, 333(4):1309–1314.
38. De Jong WS, De Jong DM, Bodis M: A fluorogenic 5′ nuclease (TaqMan) assay to assess dosage of a marker tightly linked to red skin color in autotetraploid potato. Theor Appl Genet 2003, 107(8):1384–1390.
39. Hosono N, Kato M, Kiyotani K, Mushiroda T, Takata S, Sato H, Amitani H, Tsuchiya Y, Yamazaki K, Tsunoda T et al: CYP2D6 **Genotyping for Functional-Gene Dosage Analysis by Allele Copy Number Detection.** Clinical Chemistry 2009, 55(8):1546–1554.
40. Fulton TM, Chunwongse J, Tanksley SD: **Microprep Protocol for Extraction of DNA from Tomato and Other Herbaceous Plants.** Plant Mol Biol Rep 1995, 13(3):207–209.
41. Schuelke M: **An economic method for the fluorescent labeling of PCR fragments.** Nat Biotechnol 2000, 18(2):233–234.

42. Brownstein MJ, Carpten JD, Smith JR: **Modulation of non-templated nucleotide addition by tag DNA polymerase: Primer modifications that facilitate genotyping.** Biotechniques 1996, 20(6):1004-&.
43. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.** J Hered 2002, 93(1):77–78.

Chapter 3

Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map.

Thijs van Dijk, Giulia Pagliarani, Anna Pikunova, Yolanda Noordijk, Hulya Yilmaz-Temel, Bert Meulenbroek, Richard GF Visser, Eric van de Weg
BMC Plant Biol. 2014; 14: 55. Published online 2014 Mar 1. doi: 10.1186/1471-2229-14-55

Abstract

Breeders in the allo-octoploid strawberry currently make little use of molecular marker tools. As a first step of a QTL discovery project on fruit quality traits and resistance to soil-borne pathogens such as *Phytophthora cactorum* and *Verticillium* we built a genome-wide SSR linkage map for the cross Holiday x Korona. We used the previously published MADCE method to obtain full haplotype information for both of the parental cultivars, facilitating in-depth studies on their genomic organisation.

The linkage map incorporates 508 segregating loci and represents each of the 28 chromosome pairs of octoploid strawberry, spanning an estimated length of 2050 cM. The sub-genomes are denoted according to their sequence divergence from *F.vesca* as revealed by marker performance. The map revealed high overall synteny between the sub-genomes, but also revealed two large inversions on LG2C and LG2D, of which the latter was confirmed using a separate mapping population. We discovered interesting breeding features within the parental cultivars by in-depth analysis of our haplotype data. The linkage map-derived homozygosity level of Holiday was similar to the pedigree-derived inbreeding level (33% and 29%, respectively). For Korona we found that the observed homozygosity level was over three times higher than expected from the pedigree (13% versus 3.6%). This could indicate selection pressure on genes that have favourable effects in homozygous states. The level of kinship between Holiday and Korona derived from our linkage map was 2.5 times higher than the pedigree-derived value. This large difference could be evidence of selection pressure enacted by strawberry breeders towards specific haplotypes.

The obtained SSR linkage map provides a good base for QTL discovery. It also provides the first biologically relevant basis for the discernment and notation of sub-genomes. For the first time, we revealed genomic rearrangements that were verified in a separate mapping population. We believe that haplotype information will become increasingly important in identifying marker-trait relationships and regions that are under selection pressure within breeding material. Our attempt at providing a biological basis for the discernment of sub-genomes warrants follow-up studies to streamline the naming of the sub-genomes among different octoploid strawberry maps.

Background

Cultivated strawberry (*Fragaria x ananassa*) is an important soft fruit species that is grown worldwide. Strawberry is a vegetatively propagated outbred species derived from the hybridisation of two new world species (*Fragaria chiloensis* and *Fragaria virginiana*) in the 18th century [1]. As a member of the Rosacaea family, it shares

ancestry with a variety of important food and ornamental crops such as apple, pear, peach and rose. Despite its economic importance and membership in a well-studied family, strawberry breeding to date rarely incorporates the use of molecular marker resources due to its complex, allo-octoploid genetic composition [2]. Because of this complexity, there are only a limited number of studies where clear marker-trait relationships for major genes/QTLs were identified [3-8].

The first comprehensive molecular genetic maps in strawberry were developed for the diploid wild species *Fragaria vesca* [9-12]. This effort culminated in the completion of the draft sequence of the diploid *Fragaria vesca* clone 'Hawai 4' in late 2010 [13], which provided the rosaceous community a highly valuable tool for further genomic research.

Soon after the first genetic map of diploid strawberry was published, similar studies were initiated for the octoploid strawberry, resulting in the completion of several (partial) genetic maps [6,14-22]. These mapping studies also conclusively revealed that the octoploid strawberry showed genome-wide disomic inheritance [1,2] and could therefore be classified as a full allopolyploid.

The origins of the homoeologues (or sub-genomes) in allo-octoploid strawberry have not been studied as extensively as those of other allopolyploid crops such as bread wheat and Cotton [23,24]. Molecular genetic studies revealed that the chloroplast DNA of octoploid strawberry is most closely related to that of the diploid species *Fragaria vesca* (subsp. bracteata) [25,26]. In another study on nuclear genes, it was confirmed that part of the genome was clearly related to *Fragaria vesca*, and another part was related to the wild diploid *Fragaria innumae*, leading to the hypothesis that the octoploid genome originated from the fusing of unreduced gametes of two tetraploid species from quite distinct genetic backgrounds. To date, no convention exists for naming homoeologues, and none of the octoploid mapping studies have incorporated information on the origins of the different homoelogues in the naming of the linkage groups. For this reason, the assignment of homoelogue letters is not consistent between the different octoploid maps.

The need to obtain complete haplotype information from microsatellites utilised in polyploids motivated Van Dijk et al. [27] to develop the MADCE (Microsatellite Allele Dose Configuration & Establishment) methodology for determining the allelic configuration of allopolyploid plant species [27]. This method essentially converts any allopolyploid genome into a diploid genome regarding the software and methodologies that can be employed for genetic analysis.

In this study, we created a highly comprehensive genetic SSR linkage map of the octoploid strawberry using MADCE. We used this map to differentiate homoeologues based on their efficiency in amplifying *F. vesca*-derived markers and to discover genomic rearrangements among the diploid sub-genomes (homoeologues). This map provides the genetic makeup of the two parental varieties and their levels

of homozygosity and haplotype sharing. Finally, we made comparisons of the cultivated strawberry genetic map to the physical reference map of the wild diploid *F. vesca*.

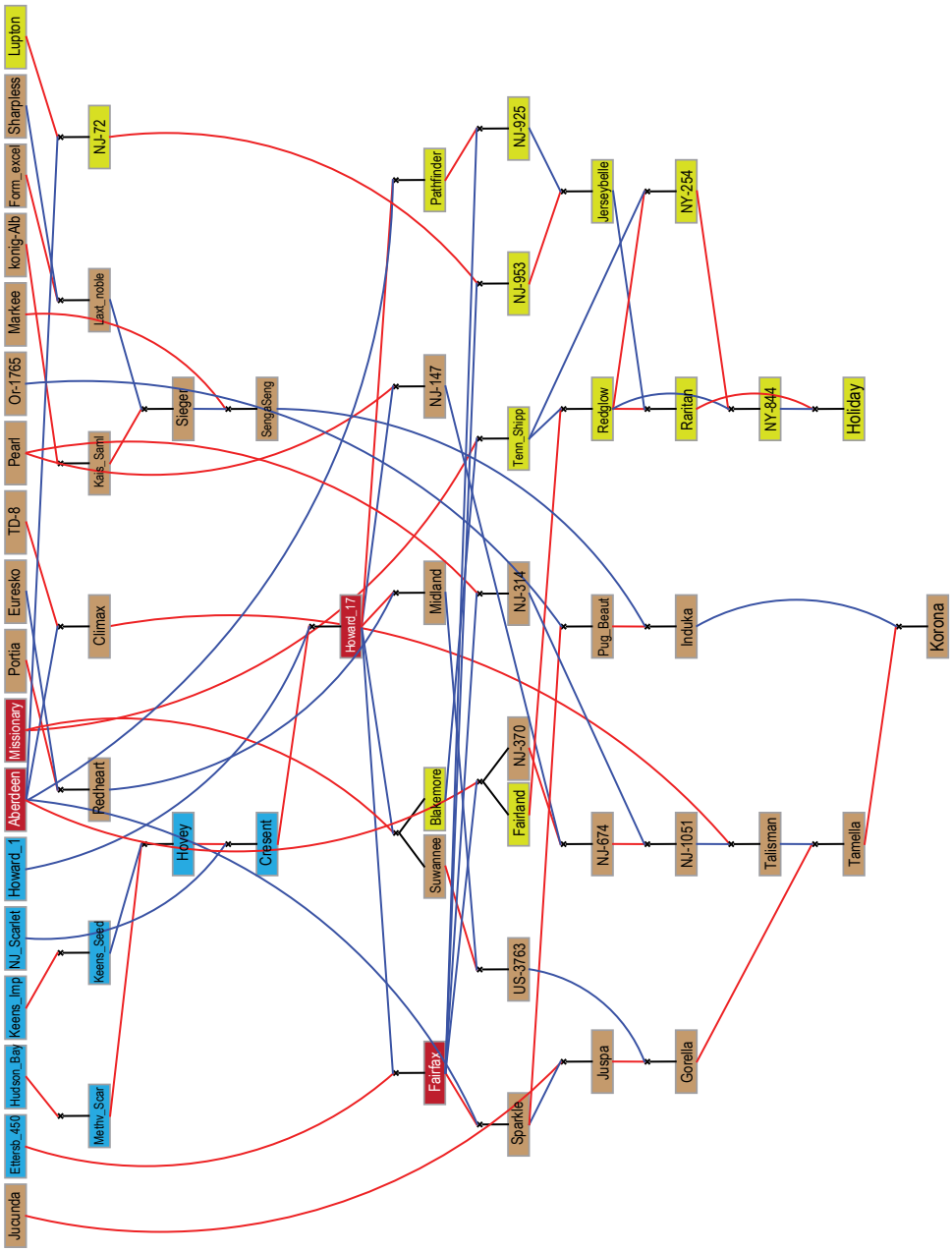


Figure 1. (left page) Pedigree of mapping parents Holiday and Korona. Red lines indicate maternal parents, and blue lines paternal parents. Yellow-green coloured parents are unique to Holiday, brown coloured parents are unique to Korona, and the red colour indicates the closest common ancestors for Holiday and Korona. Blue coloured individuals are ancestors or parents of the closest common ancestors of Holiday and Korona. This figure was drawn using Pedimap [28].

Methods

Plant materials

For the construction of a molecular marker linkage map, a subset of 92 seedlings from a cross between the strawberry cultivars 'Holiday' and 'Korona' was used. DNA admixture and possible outcrossings resulted in the removal of ten individuals, leaving a total of 82. The pedigree of Holiday and Korona is presented in Figure 1. Another F1 population of 133 individuals derived from a cross between 'Elsanta' and selection E1998-142 was used to confirm an inversion observed in Holiday x Korona. The mapping populations were created at and are maintained by the private breeding company Fresh Forward Breeding BV.

DNA isolation

Genomic DNA was extracted according to a modified version of the Fulton et al. [29] mini-prep protocol. Briefly, 1 g of young, folded leaves were harvested. The leaves were freeze-dried and ground to powder in a 2 ml tube. To this tube, 700 µL of warm (65°C) containing 2% CTAB buffer was added, and the contents were mixed by vortexing and incubated for 10 min. Next, 700 µL of chloroform:isoamyl alcohol (24:1) was added. The mixture was centrifuged at room temperature at 10,000 g for 2 min. Next, 600 µL of the top phase was transferred to a fresh tube. Isopropanol (480 µL) was added, and the sample was mixed and then centrifuged at 10,000 g for 2 min at room temperature. The supernatant was discarded, and the pellet was washed with 500 µL of 70% ethanol, left for 2 min and then centrifuged at 10,000 g for 2 min. The supernatant was discarded by pipetting, and the pellet was resuspended in 400 µL of Tris EDTA. LiCl (135 µL, 8 M) was added to remove RNA and polysaccharides, and the mixture was incubated for 30 min at -20°C. After incubation, the mixture was centrifuged at room temperature at 10,000 g for 2 min, and the supernatant was transferred to a fresh tube. Isopropanol (320 µL) was added, and the mixture was incubated at -20°C for 30 min. The mixture was then centrifuged at 10,000 g for 5 min, and the supernatant was discarded. The pellet was then washed and centrifuged twice with 500 µL of ethanol (70%), and then the dried pellet was dissolved in 50 µL of TE.

SSR markers

A total of 186 primer combinations from a variety of sources [6,10-12,15,18,22,30-43] were used for the construction of the linkage map. These primers were selected to obtain genome-wide coverage of 10-20 cM intervals with the least possible number of markers. The parameters considered were the length of the SSR repeat, the polymorphism level between our mapping parents and, when available, mapping information from previous publications. Complete information on these primer combinations can be found in Additional file 1: Table S1.

PCR

The first approximately 50 primer combinations used in this study were directly labelled with fluorescent dyes (6-FAM, NED or HEX). Subsequent PCR reactions were performed with indirect fluorescent labelling [44] using a universal 17 bp 5' end tail sequence (AACAGGTATGACCATGA) on the forward primer, which matched a universal fluorescently labelled primer (6-FAM, HEX or ROX) [44]. All reverse primers had a GTTT tail [45] on the 5' end to minimise stutter formation. The PCR mixture was composed of 1 X Goldstar PCR buffer, 0.05 μM unlabelled forward primer with a tail, 0.2 μM unlabelled reverse primer, 0.2 μM labelled universal primer, 0.3 U of Goldstar Taq polymerase (Eurogentec Nederland B.V., Maastricht, The Netherlands) and 10 ng of DNA in a total reaction volume of 20 μL. In the case of directly labelled primers, the same mixture was used except that the forward (directly) labelled primer and reverse primer were both present at 0.2 μM concentrations. The PCR conditions were one cycle at 94°C for 3 min followed by 35 cycles at 94°C for 30 s, 50°C for 30 s and 72°C for 2 min and a final extension cycle at 72°C for 10 min for both labelling methods.

Marker analysis

Fluorescently labelled amplicons were separated and detected using an ABI capillary automated sequencing platform (Initially ABI 3700 and later ABI 3730, Perkin Elmer Biosystems, Foster City, Calif.). The output from the ABI platform was analysed with either Genotyper 3.6 (ABI3700) or Genemapper 4.0 (ABI3730) software. Peaks corresponding to alleles were identified, and their bin ranges were defined. Next, for each sample, the software automatically identified the presence of alleles (peaks) and the area under the peak. Allele detection was checked manually and adjusted where necessary. The allelic data (size and area) for each individual

(parents and progeny) were transferred to an Excel spread sheet. The analysis of the data followed the MADCE procedure for establishing allelic configurations in allopolyploid populations [27], which allowed us to estimate allele dose and to identify pairs of homologous alleles for each of the sub-genomes.

Construction of linkage maps

The construction of linkage maps followed the same procedure as described by Van Dijk et al. [27]. Briefly, during data analysis, the alleles were first assigned to homologous groups on the assumption that alleles shared between parents are most likely to originate from the same sub-genome, unless the data indicated otherwise. This approach allowed the definition of so-called bridge markers that link the two pairs of parental homologs. These markers are of type <hkxhk> and <efxeg> (Annotation of JoinMap® 3.0 and following versions for Cross Pollinating systems). Early data consistency checks were performed using allelic pairs as described previously by Sargent et al. [17]. Next, linkage maps were created for each parent separately using JoinMap® 4.0 (Kyazma B.V.)[46] applying the regression approach and Kosambi mapping function. These separate parental maps were compared to each other to match the parental maps belonging to the same homoeologue based on the already-identified < hkxhk>, <efxeg > markers similar to the method of Barrett et al. [47]. This information was used for increasing the number of integrated loci by converting <lmxll> and <nnxnp> markers from the same primer pair into < abxcd > markers, as well as for the validation of the previously identified <hkxhk> and <efxeg> loci. After this data check, integrated maps were created when possible. JoinMap® output was imported into Excel to check for possible genotyping errors (double recombinants) through a graphical genotyping approach [48]. Putative double recombination events were checked up to the level of the original ABI output. The map was regarded as final when the latest corrections did not result in new putatively erroneous double recombination events, which typically required one or two rounds of corrections. The map positions of loci where both parents were homozygous were added later by imputing them from the relative positions of their homoeologous loci. Primer pairs that amplified heterologous chromosomes were never imputed and were only shown on Linkage Groups (LGs) to which their amplicons mapped. The phase information generated by JoinMap® was used to establish the parental haplotypes. Drawings of the linkage maps were first created with the software packages MapChart [49] and later finalised in Adobe Illustrator CS5 (Adobe Systems, San Jose, CA).

Denotation of sub-genomes

The assignment of a homoeologue letter (A, B, C or D) to a linkage group was based on the amplification efficiency of the *F. vesca*-derived primer pairs. The efficiency was expressed as the proportion of amplified alleles observed for all *F. vesca* primer pairs on a linkage group over the total numbers of alleles that were possible (amplified and null alleles).

Loci for which it was uncertain whether null alleles occurred (e.g., due to homozygosity on multiple LGs) were not included in the calculation for these LGs. Loci that amplified from heterologous chromosomes were only used for the efficiency calculations for the LGs on which they mapped.

Comparative mapping

Physical map locations of the microsatellites used in this study were obtained by blasting the SSR primer sequences to the *F. vesca* pseudo-chromosome assembly v 1.1 [13,50]. When no clear hits were found, we used full-length sequences of the marker, when available. In the visualisations, the physical positions of the microsatellites in mega-base pairs were multiplied by three to better fit the scale of the genetic maps. The octoploid genetic map was represented by the homoeologue that had a good density of segregating markers and showed few inconsistencies in marker order with the other homoeologues. The diploid genetic map of Sargent et al. [51] was chosen as it had most primer pairs in common with our map. The genetic positions of the CO–and CX-series of markers [18], were imputed using data from a recent diploid genetic map [50]. The maps were completed in Mapchart [49] and finalised in Adobe Illustrator CS5.

Estimation of homozygosity levels and haplotype sharing between parents

Holiday and Korona both have ancestors that occur multiple times in the known parts of their pedigree (Figure 1), due to which, part of their genomes are likely to be homozygous by descent. In addition, Holiday and Korona are likely to have shared haplotypes, as they have some ancestors in common. The theoretically expected level of homozygosity was derived from a numeric relationship matrix obtained with FlexQTL [52]. This matrix consists of doubled kinship coefficients. The observed levels of homozygosity and haplotype-sharing were estimated using our linkage map. For this map, we identified genetic regions that had multiple (3 or more) adjacent loci where the alleles were identical, within parents (for homozygosity estimation) or between parents (for haplotype sharing/kinship estimates). Because multiple

adjacent loci were used, these identical by state (IBS) regions were assumed to be identical by descent (IBD). The genetic length covered by such regions was assessed and totalled. To calculate the homozygosity levels, we divided the genetic size of the homozygous stretches by the genetic size of the genome. Briefly, the linkage map-derived kinship coefficients were calculated using Gillois identity states [53] for each genomic region (in cM) and their associated Jacquard condensed coefficients of identity [54]. On our linkage map, several identity states can be distinguished. First, there are areas where no haplotype is found in common; these areas have a kinship coefficient of 0 (Gillois identity state Δ9). Next, areas with one haplotype in common between the parents have a kinship coefficient of 0.25 (state Δ8). Areas with two different haplotypes in common between the parents have a kinship coefficient of 0.5 (state Δ7). Areas where a haplotype is homozygous in one parent and the same haplotype is heterozygous in the other parent have a kinship coefficient of 0.5 (state Δ3 and Δ5). Areas where a haplotype is homozygous in both parents have a kinship coefficient of 1 (state Δ1). As an example, when a chromosome of 60 cM has identity states Δ8, Δ3, Δ1 and Δ9 on areas of 15, 10, 5, and 30 cM, respectively, its total kinship coefficient amounts to 0.23 ((15 cM*0.25 + 10 cM*0.5 + 5 cM*1 + 30 cM*0)/60 cM).

Duplicated microsatellite analysis

To investigate the underlying causes of multi-locus targeting of microsatellites we performed a BLAST search of these sequences against the *Fragaria vesca* reference genome (cutoff value $1*E^{-10}$). We checked whether the reference genome annotation showed a transposable element identified by LTRHarvest [55] overlapping the location to which the marker was BLASTed. We then used 4 kb of flanking sequence from the most significant hit and performed a BLAST search against the nucleotide collection from NCBI to establish whether the microsatellite was present within a gene.

Results

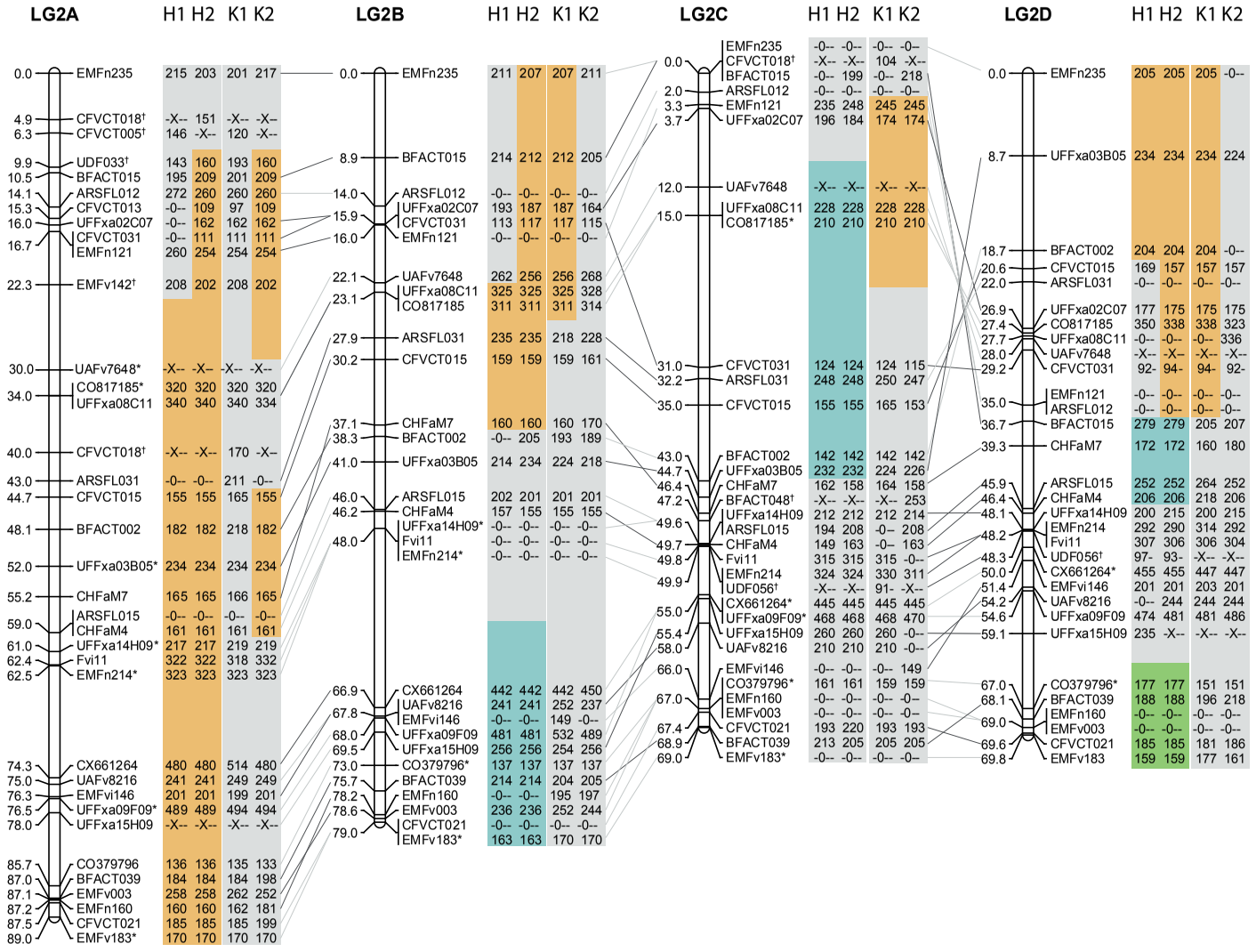
Global mapping results

A total of 186 SSR primer pairs were used to generate genetic linkage maps. They generated a total of 508 segregating loci, of which, 168 (35%) were bi-parental (<hkxhk>, <efxeg > and < abxcd > types). After splitting the bi-parental loci, the total number of loci segregating for Holiday amounted to 283 and for Korona to 393. The genetic map in its entirety is presented in Additional file 2: Figure S1. Linkage

groups 2 and 6 are presented as an example in Figure 2. All 28 chromosome pairs of the strawberry genome were recovered. For just one pair (3C), the single parental maps could not be merged into an integrated map due to large differences in the recombination rates between the two parents for shared marker loci. Two additional, small linkage groups segregating only for Korona could not be unambiguously connected to the main body of their respective linkage groups (LG3A and LG3B). Apparently, their genetic distance was too large to connect these bottom groups with the nearest informative locus from the main body of the linkage group, at least with the given family size. The total length of the integrated maps sums up to 1846 cM, making the average genetic length of a linkage group 66 cM and the average marker density one in every 3.6 cM. This total length does not include the distance between the two bottom fragments of chromosome 3 and their respective top segments, and it also excludes the segments on the extremities of a linkage group where both parents were homozygous. Using the homoeologous positions of these homozygous marker loci, the estimated total genetic length of this map extends approximately 200 cM to a total of approximately 2050 cM.

Denotation of sub-genomes

We denoted the four sub-genomes based on their level of sequence divergence from *F. vesca.* This divergence was determined by MADCE-derived genotype configurations using the proportion of amplified alleles over the total number of allowed alleles. The sub-genome with the highest efficiency (fewest null alleles) was assigned homoeologue letter A, and the sub-genome with the lowest efficiency (many null alleles) was assigned homoeologue letter D. The amplification efficiencies are shown in Table 1. LGs 1, 5 and 7 showed a stark contrast in *F. vesca* amplification efficiency between the first two homoeologues (A and B) versus the last two (C and D). In contrast, for LGs 3 and 6 and to a lesser extent LGs 2 and 4, the difference was mainly between homoeologue A and the other homoeologues. Another interesting phenomenon that was observed through the identification of null alleles was the presence of regions where several consecutive markers did not amplify any product. An example of this occurred in the centre of LG2B (Figure 2) where markers UFFxa14H09, Fvi11 and EMFn213, spanning more than one mega-base in physical distance, did not amplify any product. Other examples are observed on the distal parts of LG3B, -5B, -5C and 7C, as well as in the centre region of 6D (Additional file 2: Figure S1). These regions could constitute large deletions for specific homoeologues. Many other regions showed null alleles for one or two successive SSR loci. An increase in marker density for these regions may provide further evidence as to whether these results are indicative of true deletions or coincidental sequence divergences at the primer site(s).


LG2A
H1 H2 K1 K2
0.0 EMFn235 215 203 201 217
4.9 CFVCT018† -X-- 151 -X-- -X--
6.3 CFVCT005† 146 -X-- 120 -X--
9.9 UDF033† 143 160 193 160
10.5 BFACT015 195 209 201 209
14.1 ARSFL012 272 260 260 260
15.3 CFVCT013 -0-- 109 97 109
16.0 UFFxa02C07 -0-- 162 162 162
16.7 CFVCT031 -0-- 111 111 111
EMFn121 260 254 254 254
22.3 EMFv142† 208 202 208 202
30.0 UAFv7648* -X-- -X-- -X-- -X--
34.0 CO817185* 320 320 320 320
UFFxa08C11 340 340 340 334
40.0 CFVCT018† -X-- -X-- 170 -X--
43.0 ARSFL031 -0-- -0-- 211 -0--
44.7 CFVCT015 155 155 165 155
48.1 BFACT002 182 182 218 182
52.0 UFFxa03B05* 234 234 234 234
55.2 CHFaM7 165 165 166 165
59.0 ARSFL015 -0-- -0-- -0-- -0--
CHFaM4 161 161 161 161
61.0 UFFxa14H09* 217 217 219 219
62.4 Fvi11 322 322 318 332
62.5 EMFn214* 323 323 323 323
74.3 CX661264 480 480 514 480
75.0 UAFv8216 241 241 249 249
76.3 EMFvi146 201 201 199 201
76.5 UFFxa09F09* 489 489 494 494
78.0 UFFxa15H09 -X-- -X-- -X-- -X--
85.7 CO379796 136 136 135 133
87.0 BFACT039 184 184 184 198
87.1 EMFv003 258 258 262 252
87.2 EMFn160 160 160 162 181
87.5 CFVCT021 185 185 185 199
89.0 EMFv183* 170 170 170 170
LG2B
H1 H2 K1 K2
0.0 EMFn235 211 207 207 211
8.9 BFACT015 214 212 212 205
14.0 ARSFL012 -0-- -0-- -0-- -0--
15.9 UFFxa02C07 193 187 187 164
16.0 CFVCT031 113 117 117 115
EMFn121 -0-- -0-- -0-- -0--
22.1 UAFv7648 262 256 256 268
23.1 UFFxa08C11 325 325 325 328
CO817185 311 311 311 314
27.9 ARSFL031 235 235 218 228
30.2 CFVCT015 159 159 159 161
37.1 CHFaM7 160 160 160 170
38.3 BFACT002 -0-- 205 193 189
41.0 UFFxa03B05 214 234 224 218
46.0 ARSFL015 202 201 201 201
46.2 CHFaM4 157 155 155 155
UFFxa14H09* -0-- -0-- -0-- -0--
48.0 Fvi11 -0-- -0-- -0-- -0--
EMFn214* -0-- -0-- -0-- -0--
66.9 CX661264 442 442 442 450
67.8 UAFv8216 241 241 252 237
EMFvi146 -0-- -0-- 149 -0--
68.0 UFFxa09F09 481 481 532 489
69.5 UFFxa15H09 256 256 254 256
73.0 CO379796* 137 137 137 137
75.7 BFACT039 214 214 204 205
78.2 EMFn160 -0-- -0-- 195 197
78.6 EMFv003 236 236 252 244
79.0 CFVCT021 -0-- -0-- -0-- -0--
EMFv183* 163 163 170 170
LG2C
H1 H2 K1 K2
0.0 EMFn235 -0-- -0-- -0-- -0--
CFVCT018† -X-- -X-- 104 -X--
BFACT015 -0-- 199 -0-- 218
2.0 ARSFL012 -0-- -0-- -0-- -0--
3.3 EMFn121 235 248 245 245
3.7 UFFxa02C07 196 184 174 174
12.0 UAFv7648 -X-- -X-- -X-- -X--
15.0 UFFxa08C11 228 228 228 228
CO817185* 210 210 210 210
31.0 CFVCT031 124 124 124 115
32.2 ARSFL031 248 248 250 247
35.0 CFVCT015 155 155 165 153
43.0 BFACT002 142 142 142 142
44.7 UFFxa03B05 232 232 224 226
46.4 CHFaM7 162 158 164 158
47.2 BFACT048† -X-- -X-- -X-- 253
49.6 UFFxa14H09 212 212 212 214
ARSFL015 194 208 -0-- 208
49.7 CHFaM4 149 163 -0-- 163
49.8 Fvi11 315 315 315 -0--
49.9 EMFn214 324 324 330 311
UDF056† -X-- -X-- 91- -X--
55.0 CX661264* 445 445 445 445
55.4 UFFxa09F09* 468 468 468 470
58.0 UFFxa15H09 260 260 260 -0--
UAFv8216 210 210 210 -0--
66.0 EMFvi146 -0-- -0-- -0-- 149
67.0 CO379796* 161 161 159 159
EMFn160 -0-- -0-- -0-- -0--
EMFv003 -0-- -0-- -0-- -0--
67.4 CFVCT021 193 220 193 193
68.9 BFACT039 213 205 205 205
69.0 EMFv183* -0-- -0-- -0-- -0--
LG2D
H1 H2 K1 K2
0.0 EMFn235 205 205 205 -0--
8.7 UFFxa03B05 234 234 234 224
18.7 BFACT002 204 204 204 -0--
20.6 CFVCT015 169 157 157 157
22.0 ARSFL031 -0-- -0-- -0-- -0--
26.9 UFFxa02C07 177 175 175 175
27.4 CO817185 350 338 338 323
27.7 UFFxa08C11 -0-- -0-- -0-- 336
28.0 UAFv7648 -X-- -X-- -X-- -X--
29.2 CFVCT031 92- 94- 94- 92-
35.0 EMFn121 -0-- -0-- -0-- -0--
ARSFL012 -0-- -0-- -0-- -0--
36.7 BFACT015 279 279 205 207
39.3 CHFaM7 172 172 160 180
45.9 ARSFL015 252 252 264 252
46.4 CHFaM4 206 206 218 206
48.1 UFFxa14H09 200 215 200 215
48.2 EMFn214 292 290 314 292
Fvi11 307 306 306 304
48.3 UDF056† 97- 93- -X-- -X--
50.0 CX661264* 455 455 447 447
51.4 EMFvi146 201 201 203 201
54.2 UAFv8216 -0-- 244 244 244
54.6 UFFxa09F09 474 481 481 486
59.1 UFFxa15H09 235 -X-- -X-- -X--
67.0 CO379796* 177 177 151 151
68.1 BFACT039 188 188 196 218
EMFn160 -0-- -0-- -0-- -0--
69.0 EMFv003 -0-- -0-- -0-- -0--
69.6 CFVCT021 185 185 181 186
69.8 EMFv183 159 159 177 161

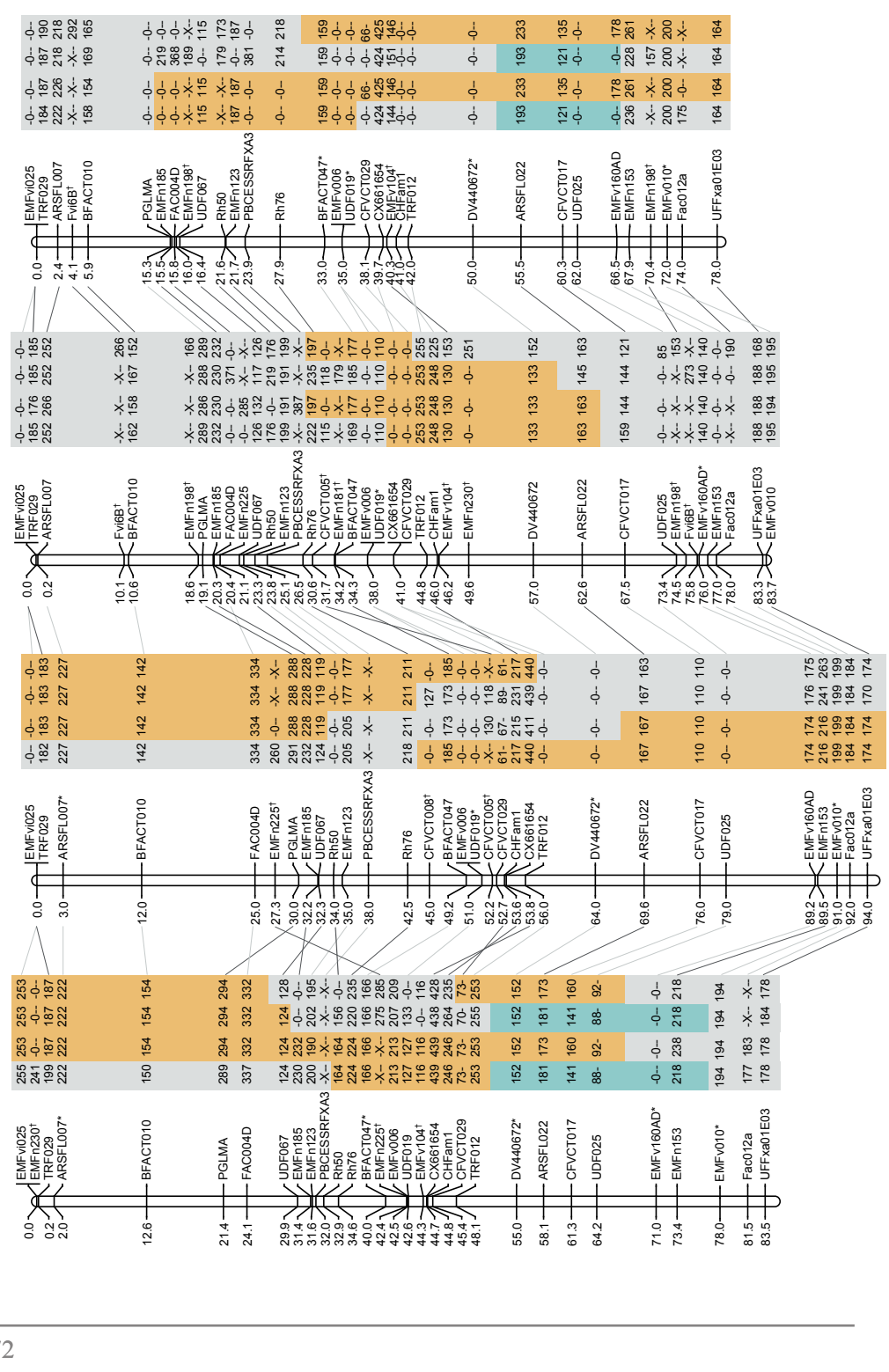

LG6A
LG6B
LG6C
LG6D
H1 H2
K1 K2
EMFvi025
EMFn230†
TRF029
ARSFL007*
BFACT010
PGLMA
FAC004D
UDF067
EMFn185
EMFn123
PBCESSRFXA3
Rh50
Rh76
BFACT047*
EMFn225†
EMFv006
UDF019*
EMFv104†
CX661654
CHFam1
CFVCT029
TRF012
DV440672*
ARSFL022
CFVCT017
UDF025
EMFv160AD*
EMFn153
EMFv010*
Fac012a
UFFxa01E03
EMFn198†
Fvi6B†
EMFn230†
EMFn225†
EMFn181†
CFVCT005†
CFVCT008†
EMFv160AD*

Figure 2. (Pages 74 and 75) Linkage maps for the 4 homoeologues of linkage groups 2 and 6 from the Holiday x Korona mapping population. Allele sizes are given in the boxes next to the names of the SSR primer pairs. "X" signifies that no allele could be assigned, as some of the observed alleles could not be reliably scored. In the figure, "0" stands for a null allele. H1 indicates Holiday haplotype 1, K1 indicates Korona haplotype 1, etc. Regions highlighted in the same colour (within a homoeologue) indicate identical haplotypes. Dark grey lines connect homoeologous loci that segregated for both neighbouring homoeologues. For light grey lines, one or both of the homoeologous loci had its position imputed. An asterisk (*) indicates that the allelic composition can be switched between homoeologues due to multiple occurrences of homozygosity. A dagger (†) indicates a primer pair that amplifies on multiple heterologous chromosomes. The minimum resolution that still represents a single recombination event is 0.6 cM for regions in which both parents segregate and 1.2 cM where only one parent segregates. Any unit that is smaller occurred due to technical issues such as missing values, uninformative individuals and integration between parents. All LGs are available in Additional file 2: Figure S1.

Table 1. Amplification efficiency of *Fragaria vesca*-derived SSR primer pairs

Linkage Group	nr of vesca derived primer pairs	amplified alleles/total alleles	total vesca efficiency in %
LG1A	13	48/52	92
LG1B	6	20/24	83
LG1C	5	10/20	50
LG1D	5	8/20	40
LG2A	11	42/44	95
LG2B	7	24/28	86
LG2C	8	27/32	84
LG2D	8	27/32	84
LG3A	13	46/52	88
LG3B	13	34/52	65
LG3C	12	27/48	56
LG3D	12	21/48	44
LG4A	6	23/24	96
LG4B	7	24/28	86
LG4C	4	13/16	81
LG4D	5	16/20	80
LG5A	7	28/28	100
LG5B	9	33/36	92
LG5C	8	22/32	69
LG5D	7	19/28	68
LG6A	10	38/40	95
LG6B	12	33/48	69
LG6C	11	30/44	68

Linkage Group	nr of vesca derived primer pairs	amplified alleles/total alleles	total vesca efficiency in %
LG6D	11	28/44	59
LG7A	6	19/24	79
LG7B	6	14/24	58
LG7C	4	5/16	31
LG7D	5	5/20	25

Differences in nr of primer pairs within a chromosome are caused by either heterologously amplifying primer pairs, or loci for which allele assignment was unclear (see footnote Figure 2).

Genomic organisation of homoeologues: collinearity and re-arrangements

Overall collinearity

The overall collinearity between the homoeologues of a chromosome was very high. There were many small-scale divergences of only 1-2 cM (Additional file 2: Figure S1), but these divergences are likely due to mapping or scoring errors that were overlooked in the error checking, missing values or the presence of less informative < hkxhk > markers. In some cases, the differences in marker order were caused by the integration of the two parental maps. An example of this discrepancy can be observed in the order of markers EMFn185 and UDF067 (LG6, Figure 2). For LG6A, UDF067 occurred before EMFn185, whereas for the other homoeologues, it did not. The nearest locus for which both parents segregated was EMFn123. The distance from EMFn185 to EMFn123 was based solely on recombination events within Holiday, and the distance from UDF067 to EMFn123 was based solely on recombination events within Korona. A difference in the recombination frequency between the parents for that small region is the likely cause of the altered marker order.

Rearrangement on chromosome 2

We identified a major rearrangement in the marker order for LG2D (Figures 2 and 3), which an inversion that spans 28 cM (from marker UFFxa03B05 at 9 cM to BFACT015 at 37 cM) (Figure 3). Because both parents show the same inversion and because multiple segregating loci are located within this region, we believe this to be a genuine inversion.

A second putative rearrangement was found on LG2C and occurs within the homoeologous region of the former inversion (Figure 2). Here, a large gap

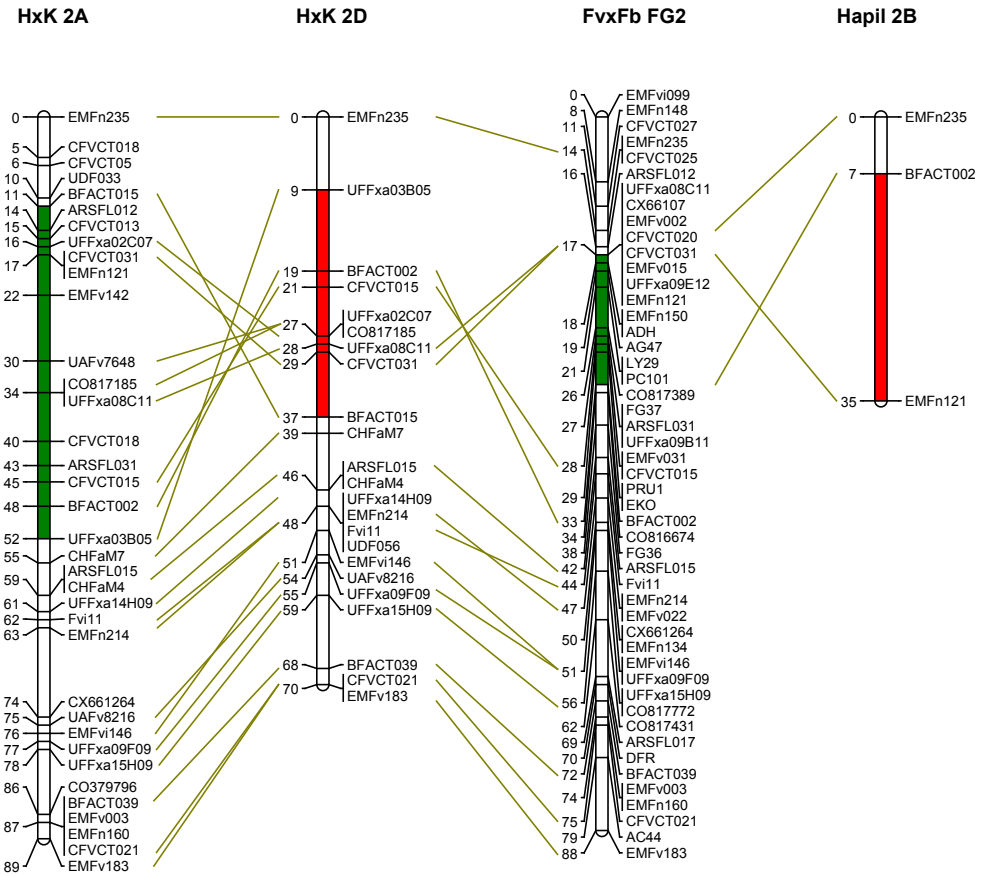


Figure 3. Linkage maps demonstrating the inversion of LG2D. On the left, LG2A of the octoploid Holiday x Korona is represented as a reference, and next to it is the LG2D of Holiday x Korona containing the putative inversion. To the right of LG2D is the diploid Fv x Fb map [51]. On the far right, LG2B of the Hapil parent from Sargent et al. [17] is shown. The filled chromosome segments indicate the regions of interest. The segments with the same colour have the same orientation. The lines were drawn from locus name to the position instead of from position to position to facilitate the traceability of locus names.

appeared between marker loci UFFxa02C07 (at 4 cM) and CFVCT031 (at 31 cM), and close linkage was observed between CFVCT031 and ARSFL031, whereas for linkage groups 2A and 2B, these three markers showed the opposite pattern (Figure 2). Unfortunately, it was not possible to verify whether this result occurred due to an inversion, a translocation or simply a large difference in the recombination rate because the markers that are normally located between CFVCT031 and ARSFL031 were not informative, being either homozygous or impossible to discern for LG2C. In any case, the size of the rearrangement is smaller than that for LG2D, due to the difference in the position of the homoeologous loci for UFFxa03B05.

The LG2 rearrangements were further examined in a second mapping population for which we had marker data available from a separate project, albeit at a lower marker density than that used for the Holiday x Korona map. The data confirmed the large inversion of LG2D in parent E1998-142 (Figure 4). Elsanta only had one marker segregating and could therefore not be used. For LG2C, we also found evidence for an inversion in E1998-142 (Figure 4). The evidence was not as strong as that for LG2D, however, because the two loci supporting the inversion were closely linked, and one of these loci (UFFxa02C07) was an < hkxhk > type marker, which are usually less accurately positioned due to hk progeny being uninformative for mapping. We re-examined previously published maps to support the existence of these rearrangements. The only indication for the occurrence of this inversion was in the octoploid map of Sargent et al. [16,17] for LG2B (which they later called LG2D) in cultivar 'Hapil' (Figure 3). It is likely that this linkage group matches our LG2D.

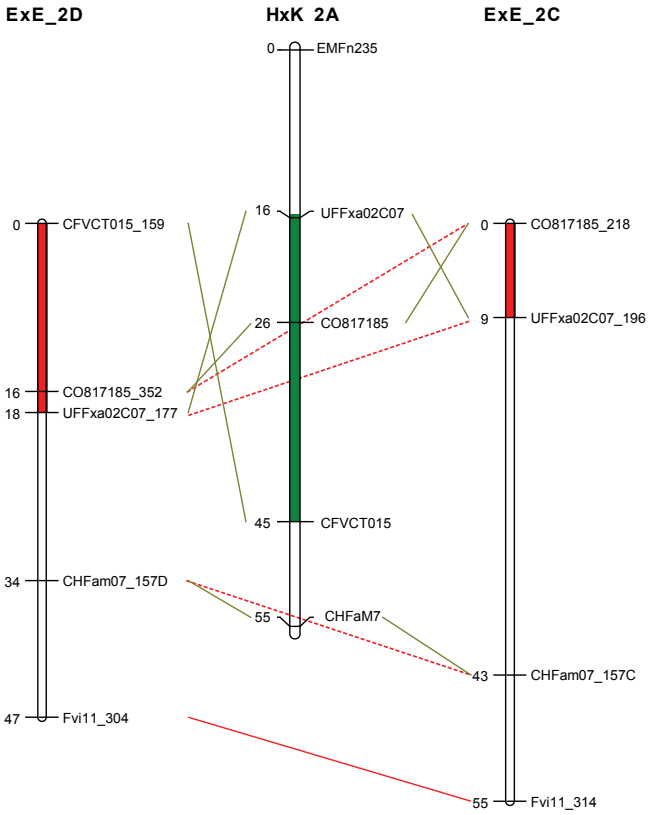


Figure 4. Linkage maps supporting inversions on LG2 in different population. Marker order of linkage groups LG2D and LG2C of mapping parent E1998-142 (from cross E1998-142 x Elsanta) and of the reference linkage group LG2A of Holiday x Korona.

Homozygosity and Heterozygosity

The level of observed homozygosity in the mapping parents is shown in Figure 5. The genome-wide level of homozygosity was almost three times higher in Holiday (33%) than in Korona (13%). This overall predominance of Holiday was also reflected in 14 linkage groups (2A-D, 3A, 3B, 3D, 4B, 4C, 5D, 6C, 6D, 7A and 7D) (Figure 5). However, one linkage group showed higher homozygosity for Korona (LG 5C). Additionally, 8 linkage groups were (nearly) completely heterozygous for both parents (1A, 1B, 3C, 4A, 4D, 5A, 5C and 7B). The overlap of homozygous regions between Holiday and Korona was 125 cM, which is close to the expected 88 cM. For Holiday, the observed level of homozygosity was similar to the theoretically expected 29% based on pedigree kinship coefficients, whereas for Korona, the observed level was more than three times higher than the expected 3.6%.

Linkage group	Homozygosity (%)					Kinship coefficient				
	A	B	C	D	total	A	B	C	D	total
Holiday LG1	0	0	74	84	39	0.08	0.14	0.36	0.70	0.32
Korona LG1	0	0	72	72	36					
Holiday LG2	71	33	43	44	49	0.10	0.08	0.00	0.20	0.10
Korona LG2	0	0	17	0	4					
Holiday LG3	30	36	10	55	30	0.13	0.01	0.00	0.20	0.08
Korona LG3	0	0	4	0	1					
Holiday LG4	0	92	46	0	40	0.00	0.00	0.00	0.06	0.01
Korona LG4	0	14	22	0	11					
Holiday LG5	0	0	0	80	20	0.16	0.00	0.17	0.14	0.12
Korona LG5	0	0	31	0	7					
Holiday LG6	18	30	21	21	22	0.30	0.22	0.16	0.20	0.22
Korona LG6	29	46	0	0	19					
Holiday LG7	39	0	63	22	31	0.51	0.16	0.00	0.43	0.27
Korona LG7	16	0	76	3	20					
Holiday Genome total %	22	27	37	44	33	0.18	0.09	0.10	0.27	0.16
Korona Genome total %	6	9	32	11	13					
Map total %	14	18	34	27	23					

Figure 5. Homozygosity and Kinship coefficients per linkage group. A,B,C and D stands for the different homoeologues of a chromosome.

Haplotype sharing (Kinship)

Holiday and Korona share four independent common ancestors, Aberdeen, Ettersburg 450, Howard 17 and Missionary (Figure 1), which are expected to contribute up to 49% and 21% of the Holiday and Korona genomes, respectively. This level of relatedness makes is likely that Holiday and Korona share marker haplotypes that are

identical by descent. The pedigree-derived kinship coefficient between Holiday and Korona was calculated as 0.06 (Table 2). This level of relatedness means that when we pick an allele at a locus in Holiday and then do the same for Korona, the chance that the two alleles are identical by descent is 6%. The actual kinship coefficient estimated from the linkage map was 2.5 times higher at 0.16 (Figure 5). Linkage groups in which both parents were homozygous generally also contained high kinship coefficients (e.g., 1D, 6A, 6B and 7A). A clear exception was homoeologue 7C in which no kinship was found even though this LG had very high levels of homozygosity for both parents. Conversely, on homoeologue 7D we found little shared homozygosity but a very high level of kinship for the heterozygous regions. Homoeologues 2C, 3B, 3C, 4A-C, 5B and 7C had very low kinship coefficients, indicating a high level of diversity between the cultivars.

Table 2. Pedigree based kinship coefficients for Holiday and Korona and their common ancestors

Variety	Holiday	Korona
Aberdeen	0.112	0.086
Ettersburg 450	0.071	0.036
Missionary	0.163	0.015
Howard 17	0.140	0.076
Holiday	0.645	0.060
Korona	0.060	0.518

The inbreeding values of Holiday and Korona can be calculated from this table using the kinship coefficient with self = 0.5* (1 + inbreeding value). This analysis amounts to inbreeding values of 0.29 and 0.036 for Holiday and Korona, respectively.

Duplicated microsatellites

A total of 19 SSR primer pairs yielded amplicons that mapped to more than one heterologous chromosome (Table 3). Of these primer pairs, six had previously been found to be multi-locus SSRs (CFVCT023, CFVCT032, EMFn181, EMFv104, EXP1A and Fvi6b [6,15-18]). For five primer pairs (BFACT048, CFVCT005, EMFv104, UDF033 and UDF056), we found that at least one of the primers was present in regions for which the LTRharvest algorithm found putative retro-transposons in the diploid reference genome (Table 3). For six primer pairs (BFACT041, BFACT048, CFVCT008, EMFv142, EXP1A and Fvi6b), the flanking sequence was located within a genic region that, at least in other species, had high homology to the sequences of large gene families.

Table 3. List of duplicated microsatellites in the octoploid Holiday x Korona map

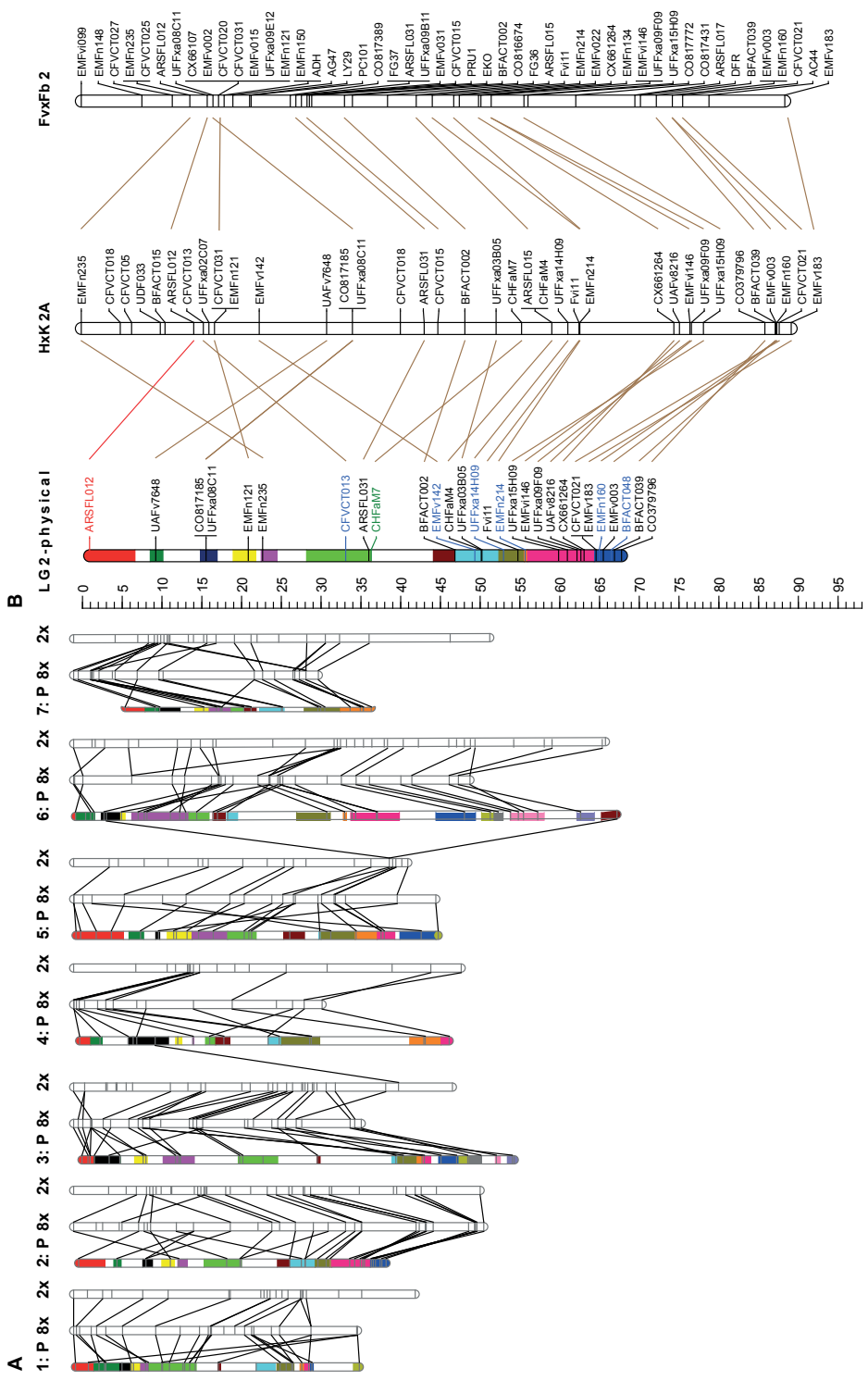
Marker	HxK nr of distinct alleles	Mapped LGs	Pseudomolecules_ v1.1_sign_hits	transpos-able element (LTRharvest)	Inside gene sequence	Large gene family	Similarity of best_scaffold 4 k region
BFACT041	21	1,4	0,1,2,3,4,5,6,7	n	y	y	Fragaria x ananassa beta-1,3-glucanase (BG2-2) gene
BFACT048	17	1,2,3,4,5,7	2,3,4,5,6,7	y	y	y	within intron of populus trichocarpa cytochrome P450 (CYP721)
CFVCT005	12	1,4,6	0,1,2,3,4,7	y	n	n	not within region with apparant function
CFVCT008	6	4,5,6,7	0,1,2,3,4,5,6,7	n	y	y	Fragaria x ananassa beta-1,3-glucanase (BG2-2) gene, complete cds
CFVCT018	24	1,2,7	0,1,2,3,4,5,6,7	n	n	n	not within region with apparant function
CFVCT023	12	5,7	7	n	n	n	not within region with apparant function
CFVCT032	12	3,4,7	0,1,2,3,4,5,6	n	n	n	not within region with apparant function
EMFn181	33	1,3,4,5,6,7	3,4,5,6	n	n	n	not within region with apparant function
EMFn198	12	1,3,4,6	no hits				no scaffold
EMFn225	8	5,6	0,1,2,3,4,5,6,7	n	n	n	not within region with apparant function
EMFn230	4	1,6	no hits				no scaffold
EMFv019	7	1,3	1	n	n	n	not within region with apparant function
EMFv104	11	3,6	3,6	y	y	n	within intron of ferredoxin-dependent glutamate synthase
EMFv142	10	2,4	3,4,7	n	y	y	within intron of adiponectin receptor protein 1-like (LOC100777701), mRNA
EXP1A	14	4,7	3,5,6,7	n	y	y	Expansin (EXP2)
Fvi6B	14	3,6	6	n	y	y	within intron of chromatin remodeling complex subunit (CHR923)
UDF004	7	3,5	3	n	y	n	within intron of molybdopterin cofactor sulfurase (ABA3)
UDF033	14	1,2,3	3,4,7	y	y	n	Pyrus pyrifolia genes for F-box proteins and S2-RNase, complete cds, haplotype: S2
UDF056	5	2,3	4,6	y	n	n	not within region with apparant function

Finally, for nine primer pairs (CFVCT018, CFVCT023, CFVCT032, EMFn181, EMFn198, EMFn225, EMFn230, EMFv019 and UDF004), we could not find any putative explanation for their targeting of heterologous loci. Two of these (EMFn198 and EMFn230) did not yield sufficiently specific hits in the reference genome to do further analysis. Another two (EMFn181 and EMFn225) corresponded to loci of varying positions among the four homoeologues of a chromosome and finally, two pairs (CFVCT018 and EMFn198) corresponded to loci of varying position within a homoeologue. This result strongly indicates that markers EMFn181, EMFn225 CFVCT018 and EMFn198 represent mobile elements, which is consistent with the lack of adjacent markers showing similar behaviour.

Comparison to the diploid genome

For a comparison of marker order between the pseudo-chromosomes of the diploid *F. vesca* reference genome (V 1.1) [13,50], the most representative homoeologues of our octoploid map and the diploid FvxFb map [51] are presented in Figure 6 and Additional file 3: Figure S2. The overall marker order conservation between the diploid physical and octoploid genetic map was found to be high, but nevertheless, it showed some discrepancies, which were classified into two types. Type I involved inversions in marker order over relatively small (scaffold size) distances. Two clear examples occur at the distal end of LG2 where the orientation of scaffolds seems to be inverted (Figure 6). The type II discrepancy involved mostly single loci that showed large differences in their position and order from the physical map to the octoploid genetic map. Examples include the marker loci EMFn235, EMFn121 and UFFxa08C11 for LG2 (Figure 6, Additional file 3: Figure S2). Overall, our genetic map and the diploid FvxFb genetic map were consistent with each other, especially in the case of type II discrepancies. This could indicate that there are still some mistakes in the orientation and position of a number of scaffolds in the diploid physical pseudo-chromosome maps. Our map of the octoploid strawberry may thus help to further improve the physical map.

Figure 6. (Right page) Comparative mapping octoploid vs diploid. A: An overview of comparative mapping between the physical reference genome, a representative octoploid homoeologue, and the diploid Fv xFb map İ51ı. Coloured bar segments represent scaffolds. B: A more detailed figure of LG2 including marker names and genetic positions. The ruler represents the position in cM for genetic maps, and the position in mega-bases for the physical map. The latter are multiplied by three in order to better fit the scale of the genetic maps. Blue font indicates a marker for which only 1 primer hit, but with 100% identity. Red font indicates a marker for which only 1 primer hit and the identity was not 100%. The lines for order comparison were drawn from locus name to position instead of position to position in order to facilitate traceability of locus names. The details of other LGs are presented in Additional file 3: Figure S2.


A
1: P 8x
2x
2: P 8x
2x
3: P 8x
2x
4: P 8x
2x
5: P 8x
2x
6: P 8x
2x
7: P 8x
2x
B
LG2-physical
ARSFL012
UAFv7648
CO817185
UFFxa08C11
EMFn121
EMFn235
CFVCT013
ARSFL031
CHFaM7
BFACT002
EMFv142
CHFaM4
UFFxa03B05
UFFxa14H09
Fvi11
EMFn214
UFFxa15H09
EMFvi146
UFFxa09F09
UAFv8216
CX661264
CFVCT021
EMFv183
EMFn160
EMFv003
BFACT048
BFACT039
CO379796
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
HxK 2A
EMFn235
CFVCT018
CFVCT05
UDF033
BFACT015
ARSFL012
CFVCT013
UFFxa02C07
CFVCT031
EMFn121
EMFv142
UAFv7648
CO817185
UFFxa08C11
CFVCT018
ARSFL031
CFVCT015
BFACT002
UFFxa03B05
CHFaM7
ARSFL015
CHFaM4
UFFxa14H09
Fvi11
EMFn214
CX661264
UAFv8216
EMFvi146
UFFxa09F09
UFFxa15H09
CO379796
BFACT039
EMFv003
EMFn160
CFVCT021
EMFv183
FvxFb 2
EMFvi099
EMFn148
CFVCT027
EMFn235
CFVCT025
ARSFL012
UFFxa08C11
CX66107
EMFv002
CFVCT020
CFVCT031
EMFv015
UFFxa09E12
EMFn121
EMFn150
ADH
AG47
LY29
PC101
CO817389
FG37
ARSFL031
UFFxa09B11
EMFv031
CFVCT015
PRU1
EKO
BFACT002
CO816674
FG36
ARSFL015
Fvi11
EMFn214
EMFv022
CX661264
EMFn134
EMFvi146
UFFxa09F09
UFFxa15H09
CO817772
CO817431
ARSFL017
DFR
BFACT039
EMFv003
EMFn160
CFVCT021
AC44
EMFv183

Discussion

In this study, we used the MADCE method [27] to develop an integrated genetic map of two octoploid strawberry cultivars that, for the first time for a polyploid plant species, included comprehensive haplotype information, even for homozygous regions and areas with null alleles. The benefits of having such extensive haplotype information are discussed in the following sections.

Map length and marker density

The map length of 2050 cM (corrected for homozygosity) is largely in line with the results of previous studies [6,15-17,22]. The marker density of one marker per every 3.6 cM does not provide improvements over some previously reported linkage maps in octoploid strawberry [16,22]. However, the use of MADCE allowed us to maximise the number of segregating loci per primer pair, allowing for better comparisons of marker order retention between homoeologues. In addition, the very precise pinpointing of homozygous regions showed that marker saturation in these areas for obtaining a better resolution would be futile, as was also indicated previously by Sargent et al. [16].

Denotation of sub-genomes

No convention exists regarding the differentiation and notation of sub-genomes. Here, we distinguished the sub-genomes based on their sequence divergence from *F. vesca* using molecular markers. This approach is in contrast to those of previous studies where technical features such as number of loci and map length were used to distinguish the homoeologues [6,14-22]. These parameters are affected by the level of homozygosity and may thus largely vary with the parents that were used for crossing. We believe that our approach, though rudimentary, is biologically more relevant and that in the near future the link between the octoploid sub-genomes and their diploid and tetraploid ancestors will become better specified, as has occurred for bread wheat [23].

Because strawberry is fully diploidized and for the ease of distinction, we denoted the sub-genomes with the letters A–D. However, if in the near future octoploid strawberries are shown to have originated from two tetraploid species of different origins, as suggested by Rousseau-Gueutin et al. [56], and as supported by our data on LGs 1, 5 and 7, then we may adopt an identical homoeologue naming convention.

Currently, the 'Holiday' x 'Korona' map is used for the mapping of thousands of SNP markers from the recently released Axiom® Strawberry Genotyping Array (also called International Strawberry 90 K SNP array or IStraw90), which was generated in a joint effort of the USDA-SCRI RosBREED project (http://www.rosbreed.org) and Affymetrix Ltd (Santa Clara, CA, USA). Now that large scale SNP arrays have come into use for strawberry, the use of a similar naming convention for octoploid maps should become straightforward due to the expected high number of common markers.

Homoeologue collinearity and re-arrangements

The marker order between homoeologous linkage groups was highly collinear, as observed in previous studies [6,15-17,22]. Inconsistencies spanning very small distances are more likely to be attributable to scoring errors, missing values, the use of markers that are not equally informative, and differences in recombination rates between the single parent maps than actual genomic rearrangements.

We described for the first time a large inversion in a linkage map of octoploid strawberry. This inversion on LG2D spanned almost 30 cM and was verified in an independent mapping population. Because we could not trace this inversion in the linkage maps of *Fragaria vesca* and *Fragaria bucharica* [10,11,42,51,57], it should derive from one of the other ancestors of *Fragaria x ananassa,* or, less likely, may have occurred after polyploidisation. Apart from the LG2D inversion, we also found evidence that LG2C may contain a rearrangement, although the evidence was less clear. It would be interesting to further investigate these rearrangements using different progenies, higher marker densities and fluorescent in-situ hybridisation (FISH, similar to Tang et al. [58]), as the rearrangements could reveal interesting insights into the relationship between the octoploid strawberry and its diploid relatives.

Breeding signatures in mapping parents

Homozygosity

The information on homozygosity and haplotype sharing generated by our genetic map revealed interesting features for both of the parental varieties, which allows us to hypothesise the possibility of breeding signatures. On a genome-wide level, the homozygosity in Holiday was found to be similar to what was theoretically expected (approximately 30%), whereas for Korona, the inbreeding level was much higher than

expected (13% vs. 3.6%). Holiday has a history of heavy inbreeding, and Korona does not. It is likely that normalisation occurs for homozygosity, where extreme levels of homozygosity and extreme levels of heterozygosity are not favoured during selection. This trend may be especially true for high levels of homozygosity, as such levels are known to lead to inbreeding depression in strawberry [59,60]. In crop species, certain traits of high agronomic value, such as plant size, adaptability and vigour, frequently favour heterozygous states [61]. Conversely, traits such as fruit firmness, shape and size are often inherited recessively or occur due to the additive nature of alleles and thereby favour homozygous states [62]. The differences in phenotype between the parents corroborate this hypothesis. The heavily inbred cultivar Holiday was purposefully bred for high fruit firmness and skin toughness, whereas Korona is a popular garden variety because of its adaptability and taste. However, it suffers from soft fruits and irregular fruit shape. Both varieties exhibit large fruit size and yield. There is a high probability that the differences and similarities in the distribution of homozygosity along the genome reflect the phenotype differences and similarities of the parental lines. Genes controlling fruit firmness, shape and skin vulnerability may therefore be located in areas where Holiday is homozygous and Korona is heterozygous, whereas genes controlling traits that favour agronomic performance are more likely to be located in regions where both parents are heterozygous.

IBD of haplotypes (kinship)

For Holiday and Korona, linkage map-derived kinship is more than twice as high as expected from their pedigrees. This result could indicate positive selection of these shared genetic regions. The two varieties with the theoretically largest contribution to the kinship between Holiday and Korona are Howard 17 and Aberdeen. These two varieties have been used extensively as parents in early 20th century strawberry breeding [63] and are therefore present in the pedigrees of many modern varieties. It is very likely that certain genomic regions of these founders are under positive selection, which would result in a higher than expected level of kinship in their descendants. Another explanation for the relatively high level of kinship could be the presence of close common ancestry among the founders of Holiday and Korona. The distribution of the shared haplotypes appears to be non-random. Certain chromosomes, such as 7A and 7D, were found to have almost three times the average haplotype-sharing, whereas all homoeologues of chromosome 4 had virtually no shared haplotypes. This result could be a coincidence but may also be due to positive or purifying selection for specific regions. Tracing the shared haplotypes between Holiday and Korona over a pedigree to their founders as well as their descendants

could further clarify which haplotypes are under strong positive or purifying selection. These could be interesting for marker assisted breeding, even without knowing the associated trait(s).

Duplications

Nineteen (10%) of the 186 microsatellite markers tested mapped to two or more heterologous chromosomes. Six of these had previously been reported as duplicated [6,15,16], and some were suggested to be remnants of a putative ancient chromosomal duplication event [16]. Our findings did not support the presence of ancient duplication events, as we could not find a clear pattern where the same heterologous chromosome segments were consistently being amplified by multiple duplicated microsatellites. The occurrence of several of these markers in known transposable elements and large gene families provides further evidence against the hypothesis of duplication events

Comparison to the diploid Fragaria genome

The comparative mapping revealed a generally high level of collinearity between the octoploid genetic map and the diploid physical and genetic map. This result is in line with that of previous studies [6,15-17,22]. However, we did find some discrepancies, which in most cases showed that the octoploid genetic map had better collinearity with the diploid Fv x Fb genetic map [51] than with the diploid *Fragaria vesca* pseudo chromosomes (v1.1). This result indicates that most of the divergence between our genetic maps and the physical map are due to limitations of the physical map. A possible explanation for this result could be that some of the scaffolds have been mapped and oriented with a BIN set comprised of a relatively low number of individuals. Knowledge of the identity of erroneously placed or erroneously oriented scaffolds as provided by this study may be of great help when fine-mapping genes of interest. However, it is not always possible to pinpoint which of the discrepancies are due to errors in the genetic maps or pseudo-chromosomes or due to real rearrangements. We could therefore not positively confirm nor reject the rearrangements observed by Sargent et al. [16] on LGs 1, 3 and 4.

Conclusion

The MADCE approach enabled the full assessment of marker haplotypes for sets of SSR-loci across the four homoeologues. It also enabled the identification of genomic

rearrangements, and the discernment of homoeologues based on their similarity to the *F. vesca* genome. Moreover, we were able to assess the level and distribution of homozygosity and haplotype-sharing, which could indicate breeding signatures. The availability of haplotype information is crucial to go from mapping population-derived QTLs to marker-assisted selection in breeding germplasm. Haplotype information will also prove to be a valuable tool in several other aspects of strawberry breeding, such as parent selection, the verification of pedigree information and IBD analysis. New technologies such as SNP arrays and Genotyping by Sequencing (GBS) could speed up the availability of such information in allopolyploids, if coupled with appropriate methodologies to discern the different sub-genomes. We hope that with this study we have provided a significant step towards the availability of such comprehensive genetic information in strawberry.

ACKNOWLEDGEMENTS

The Socrates-Erasmus Student Exchange Programme is acknowledged for the fellowship awarded to Hulya Yilmaz Temel. Tom Davis, University of New Hampshire, is acknowledged for sharing sequence information for the design of the primers TDFM1, -2, and -6, and Paul Arens, Wageningen UR Plant Breeding is acknowledged for the design of these specific primers. Funding for this research was provided through the Technological Top Institute Green Genetics, project number 1C004RP.

References

1. Hancock JF, Sjulin TM, Lobos GA: **Strawberries**. In *Temperate Fruit Crop Breeding.* Netherlands: Springer; 2008:393–437.
2. Whitaker VM: **Applications of molecular markers in strawberry.***J Berry Res* 2011, **1**(3)**:**115–127.
3. Lerceteau-Kohler E, Guerin G, Denoyes-Rothan B: **Identification of SCAR markers linked to*Rca2*anthracnose resistance gene and their assessment in strawberry germplasm.***Theor Appl Genet* 2005, **111**(5)**:**862–870.
4. Haymes KM, Henken B, Davis TM, Van de Weg WE: **Identification of RAPD markers linked to a*Phytophthora fragariae*resistance gene (*Rpf1*) in the cultivated strawberry.***Theor Appl Genet* 1997, **94**(8)**:**1097–1101.
5. Gaston A, Perrotte J, Lerceteau-Kohler E, Rousseau-Gueutin M, Petit A, Hernould M, Rothan C, Denoyes B: ***PFRU*, a single dominant locus regulates the balance between sexual and asexual plant reproduction in cultivated strawberry.***J Exp Bot* 2013, **64**(7)**:**1837–1848.
6. Spigler RB, Lewers KS, Johnson AL, Ashman TL: **Comparative mapping reveals autosomal origin of sex chromosome in octoploid*Fragaria virginiana*.***J Hered* 2010, **101:**S107–S117.
7. Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sanchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya I: **Genetic analysis of strawberry fruit aroma and identification of O-Methyltransferase*FaOMT*as the locus controlling natural variation in mesifurane content.***Plant Physiol* 2012, **159**(2)**:**851.
8. Haymes KM, Van de Weg WE, Arens P, Maas JL, Vosman B, Den Nijs APM: **Development of SCAR markers linked to a*Phytophthora fragariae*resistance gene and their assessment in European and North American strawberry genotypes.***J Am Soc Hortic Sci* 2000, **125**(3)**:**330–339.
9. Davis TM, Yu H: **A linkage map of the diploid strawberry,*Fragaria vesca*.***J Hered* 1997, **88**(3)**:**215–221.
10. Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH, Simpson DW: **A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid*Fragaria*.***Theor Appl Genet* 2004, **109**(7)**:**1385–1391.
11. Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH: **An enhanced microsatellite map of diploid*Fragaria*.***Theor Appl Genet* 2006, **112**(7)**:**1349–1359.
12. Cipriani G, Pinosa F, Bonoli M, Faedi W: **A new set of microsatellite markers for*Fragaria*species and their application in linkage analysis.***J Horticult Sci Biotechnol* 2006, **81**(4)**:**668–675.
13. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil NV, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael

TP, Setubal JC, Celton JM, Rees DJG, Williams KP, Holt SH, Rojas JJR, Chatterjee M, *et al*: **The genome of woodland strawberry (*Fragaria vesca*).***Nat Genet* 2011, **43**(2)**:**109–U151.

14. Lerceteau-Kohler E, Guerin G, Laigret F, Denoyes-Rothan B: **Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (*Fragaria x ananassa*) using AFLP mapping.***Theor Appl Genet* 2003, **107**(4)**:**619–628.
15. Rousseau-Gueutin M, Lerceteau-Kohler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arus P, Guerin G, Denoyes-Rothan B: **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* 2008, **179**(4)**:**2045–2060.
16. Sargent D, Passey T, Šurbanovski N, Lopez Girona E, Kuchta P, Davik J, Harrison R, Passey A, Whitehouse A, Simpson D: **A microsatellite linkage map for the cultivated strawberry*Fragaria x ananassa*suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection.***Theor Appl Genet* 2012, **124**(7)**:**1229–1240.
17. Sargent DJ, Fernandez-Fernandez F, Ruiz-Roja JJ, Sutherland BG, Passey A, Whitehouse AB, Simpson DW: **A genetic linkage map of the cultivated strawberry (*Fragaria x ananassa*) and its comparison to the diploid*Fragaria*reference map.***Mol Breed* 2009, **24**(3)**:**293–303.
18. Spigler RB, Lewers KS, Main DS, Ashman TL: **Genetic mapping of sex determination in a wild strawberry,*Fragaria virginiana*, reveals earliest form of sex chromosome.** *Heredity* 2008, **101**(6)**:**507–517.
19. Weebadde CK, Wang D, Finn CE, Lewers KS, Luby JJ, Bushakra J, Sjulin TM, Hancock JF: **Using a linkage mapping approach to identify QTL for day-neutrality in the octoploid strawberry.***Plant Breed* 2008, **127**(1)**:**94–101.
20. Viruel MA: **An SSR and RFLP linkage map for the octoploid strawberry (*Fragaria x ananassa*)**. In *Plant, Animal and Microbe genomes X conference.* San Diego, California, USA; 2002:660.
21. Zorrilla-Fontanesi Y, Cabeza A, Dominguez P, Medina JJ, Valpuesta V, Denoyes-Rothan B, Sanchez-Sevilla JF, Amaya I: **Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria x ananassa*).***Theor Appl Genet* 2011, **123**(5)**:**755–778.
22. Isobe SN, Hirakawa H, Sato S, Maeda F, Ishikawa M, Mori T, Yamamoto Y, Shirasawa K, Kimura M, Fukami M, Hashizume, F, Tsuji T, Sasamoto S, Kato M, Nanri K, Tsuruoka H, Minami C, Takahashi C, Wada T, Ono A, Kawashima K, Nakazaki N, Kishida Y, Kohara M, Nakayama S, Yamada M, Fujishiro T, Watanabe A, Tabata, S: **Construction of an integrated high density simple sequence repeat linkage map in cultivated strawberry*(Fragaria x ananassa)*and its applicability.***DNA Res* 2013, **20**(1)**:**79–92.
23. Kerby K, Kuspira J: **The phylogeny of the polyploid wheats*Triticum-Aestivum*(Bread Wheat) and*Triticum-Turgidum*(Macaroni Wheat).***Genome* 1987, **29**(5)**:**722–737.

24. Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH: **A detailed Rflp map of cotton,*Gossypium-Hirsutum X Gossypium Barbadense*–chromosome organization and evolution in a disomic polyploid genome.***Genetics* 1994, **138**(3)**:**829–847.
25. Njuguna W, Liston A, Cronn R, Ashman TL, Bassil N: **Insights into phylogeny, sex function and age of*Fragaria*based on whole chloroplast genome sequencing.***Mol Phylogenet Evol* 2013, **66**(1)**:**17–29.
26. Potter D, Luby JJ, Harrison RE: **Phylogenetic relationships among species of*Fragaria*(Rosaceae) inferred from non-coding nuclear and chloroplast DNA sequences.***Syst Bot* 2000, **25**(2)**:**337–348.
27. Van Dijk T, Noordijk Y, Dubos T, Bink MCAM, Meulenbroek EJ, Visser RGF, Van de Weg WE: **Microsatellite Allele Dose and Configuration Establishment (MADCE): an integrated approach for genetic studies in allopolyploids.***BMC Plant Biol* 2012, **12:**25.
28. Voorrips RE, Bink MCAM, van de Weg WE: **Pedimap: software for the visualization of genetic and phenotypic data in pedigrees.***J Hered* 2012, **103**(6)**:**903–907.
29. Fulton TM, Chunwongse J, Tanksley SD: **Microprep protocol for extraction of DNA from tomato and other herbaceous plants.***Plant Mol Biol Rep* 1995, **13**(3)**:**207–209.
30. Ashley MV, Wilk JA, Styan SMN, Craft KJ, Jones KL, Feldheim KA, Lewers KS, Ashman TL: **High variability and disomic segregation of microsatellites in the octoploid*Fragaria virginiana Mill.*(Rosaceae).***Theor Appl Genet* 2003, **107**(7)**:**1201–1207.
31. Bassil NV, Gunn M, Folta K, Lewers K: **Microsatellite markers for*Fragaria*from 'Strawberry Festival' expressed sequence tags.***Mol Ecol Notes* 2006, **6**(2)**:**473–476.
32. Bassil NV, Njuguna W, Slovin JP: **EST-SSR markers from Fragaria vesca L. cv. yellow wonder.***Mol Ecol Notes* 2006, **6**(3)**:**806–809.
33. Cipriani G, Testolin R: **Isolation and characterization of microsatellite loci in*Fragaria*.** *Mol Ecol Notes* 2004, **4**(3)**:**366–368.
34. Gil-Ariza DJ, Amaya I, Botella MA, Blanco JM, Caballero JL, Lopez-Aranda JM, Valpuesta V, Sanchez-Sevilla JF: **EST-derived polymorphic microsatellites from cultivated strawberry (*Fragaria x ananassa*) are useful for diversity studies and varietal identification among*Fragaria*species.***Mol Ecol Notes* 2006, **6**(4)**:**1195–1197.
35. Hadonou AM, Sargent DJ, Wilson F, James CM, Simpson DW: **Development of microsatellite markers in*Fragaria*, their use in genetic diversity analysis, and their potential for genetic linkage mapping.***Genome* 2004, **47**(3)**:**429–438.
36. Hadonou M, Sargent D, Walden R, Simpson D: **Characterisation of*Fragaria vesca*single sequence repeats (SSR) markers.***Acta Hort*(649)**:**99–102.
37. James CM, Wilson F, Hadonou AM, Tobutt KR: **Isolation and characterization of polymorphic microsatellites in diploid strawberry (*Fragaria vesca L.*) for mapping, diversity studies and clone identification.***Mol Ecol Notes* 2003, **3**(2)**:**171–173.
38. Keniry A, Hopkins CJ, Jewell E, Morrison B, Spangenberg GC, Edwards D, Batley

J: **Identification and characterization of simple sequence repeat (SSR) markers from*Fragaria x ananassa*expressed sequences.***Mol Ecol Notes* 2006, **6**(2)**:**319–322.

39. Lewers KS, Styan SMN, Hokanson SC, Bassil NV: **Strawberry GenBank-derived and genomic simple sequence repeat (SSR) markers and their utility with strawberry, blackberry, and red and black raspberry.***J Am Soc Hortic Sci* 2005, **130**(1)**:**102–115.
40. Monfort A, Vilanova S, Davis TM, Arus P: **A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid*Fragaria*species and to*Fragaria x ananassa*.***Mol Ecol Notes* 2006, **6**(1)**:**197–200.
41. Oyant LHS, Crespel L, Rajapakse S, Zhang L, Foucher F: **Genetic linkage maps of rose constructed with new microsatellite markers and locating QTL controlling flowering traits.***Tree Genetics Genomes* 2008, **4**(1)**:**11–23.
42. Sargent DJ, Rys A, Nier S, Simpson DW, Tobutt KR: **The development and mapping of functional markers in*Fragaria*and their transferability and potential for mapping in other genera.***Theor Appl Genet* 2007, **114**(2)**:**373–384.
43. Yan Z, Denneboom C, Hattendorf A, Dolstra O, Debener T, Stam P, Visser PB: **Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers.***Theor Appl Genet* 2005, **110**(4)**:**766–777.
44. Schuelke M: **An economic method for the fluorescent labeling of PCR fragments.***Nat Biotechnol* 2000, **18**(2)**:**233–234.
45. Brownstein MJ, Carpten JD, Smith JR: **Modulation of non-templated nucleotide addition by taq DNA polymerase: primer modifications that facilitate genotyping.** *Biotechniques* 1996, **20**(6)**:**1004.
46. Van Ooijen JW: *In Joinmap 4* ®. Wageningen, Netherlands: Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV; 2006.
47. Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D: **A microsatellite map of white clover.***Theor Appl Genet* 2004, **109**(3)**:**596–608.
48. Young ND, Tanksley SD: **Restriction fragment length polymorphism maps and the concept of graphical genotypes.***Theor Appl Genet* 1989, **77**(1)**:**95–101.
49. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.***J Hered* 2002, **93**(1)**:**77–78.
50. Sargent DJ, Kuchta P, Girona EL, Zhang H, Davis TM, Celton J-M, Marchese A, Korbin M, Folta KM, Shulaev V, *Simpson DW*: **Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence.***Plant Genome* 2011, **4**(3)**:**165–177.
51. Sargent DJ, Davis TM, Simpson DW, Folta KM, Gardiner SE: *Strawberry (Fragaria spp.) Structural Genomics Genetics and Genomics of Rosaceae.* 6th edition. New York: Springer; 2009:437–456.
52. Bink MCAM, Jansen J, Madduri M, Voorrips RE, Durel CE, Kouassi AB, Laurens F, Mathis F, Gessler C, Gobbin D, Rezzonico, F, Patocchi A, Kellerhals M, Boudichevskaia

A, Dunemann F, Peil A, Nowicka A, Lata B, Stankiewicz-Kosyl M, Jeziorek K, Pitera E, Soska A, Tomala K, Evans KM, Fernandez-Fernandez F, Guerra W, Korbin M, Keller S, Lewandowski M, Plocharski W, *et al*: **Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple.***Theor Appl Genet. Advance online publication*. doi:10.1007/s00122-014-2281-3.

53. Gillois M: **La relation d'identité genetique**. In *PhD thesis.* University of Paris: Faculté des Sciences; 1964.
54. Jacquard A: *The genetic structure of populations.* New York: Springer-Verlag; 1974.
55. Ellinghaus D, Kurtz S, Willhoeft U: **LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons.***BMC Bioinformatics* 2008, **9:**18.
56. Rousseau-Gueutin M, Gaston A, Ainouche A, Ainouche ML, Olbricht K, Staudt G, Richard L, Denoyes-Rothan B: **Tracking the evolutionary history of polyploidy in*Fragaria L.*(strawberry): new insights from phylogenetic analyses of low-copy nuclear genes.***Mol Phylogenet Evol* 2009, **51**(3)**:**515–530.
57. Sargent DJ, Cipriani G, Vilanova S, Gil-Ariza D, Arus P, Simpson DW, Tobutt KR, Monfort A: **The development of a bin mapping population and the selective mapping of 103 markers in the diploid*Fragaria*reference map.***Genome* 2008, **51**(2)**:**120–127.
58. Tang XM, Szinay D, Lang C, Ramanna MS, van der Vossen EAG, Datema E, Lankhorst RK, de Boer J, Peters SA, Bachem C, Stiekema W, Visser RGF, de Jong H, Bai YL: **Cross-species bacterial artificial chromosome-fluorescence in situ hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements.***Genetics* 2008, **180**(3)**:**1319–1328.
59. Rho IR, Woo JG, Jeong HJ, Jeon HY, Lee CH: **Characteristics of F-1 hybrids and inbred lines in octoploid strawberry (*Fragaria x ananassa Duchesne*).***Plant Breed* 2012, **131**(4)**:**550–554.
60. Shaw DV: **Comparison of ancestral and current generation inbreeding in an experimental strawberry breeding population.***Theor Appl Genet* 1995, **90**(2)**:**237–241.
61. Hochholdinger F, Hoecker N: **Towards the molecular basis of heterosis.***Trends Plant Sci* 2007, **12**(9)**:**427–432.
62. Bai YL, Lindhout P: **Domestication and breeding of tomatoes: what have we gained and what can we gain in the future?***Ann Bot* 2007, **100**(5)**:**1085–1094.
63. Darrow GM: *The strawberry; history, breeding, and physiology.* 1st edition. New York: Holt; 1966.

Additional files

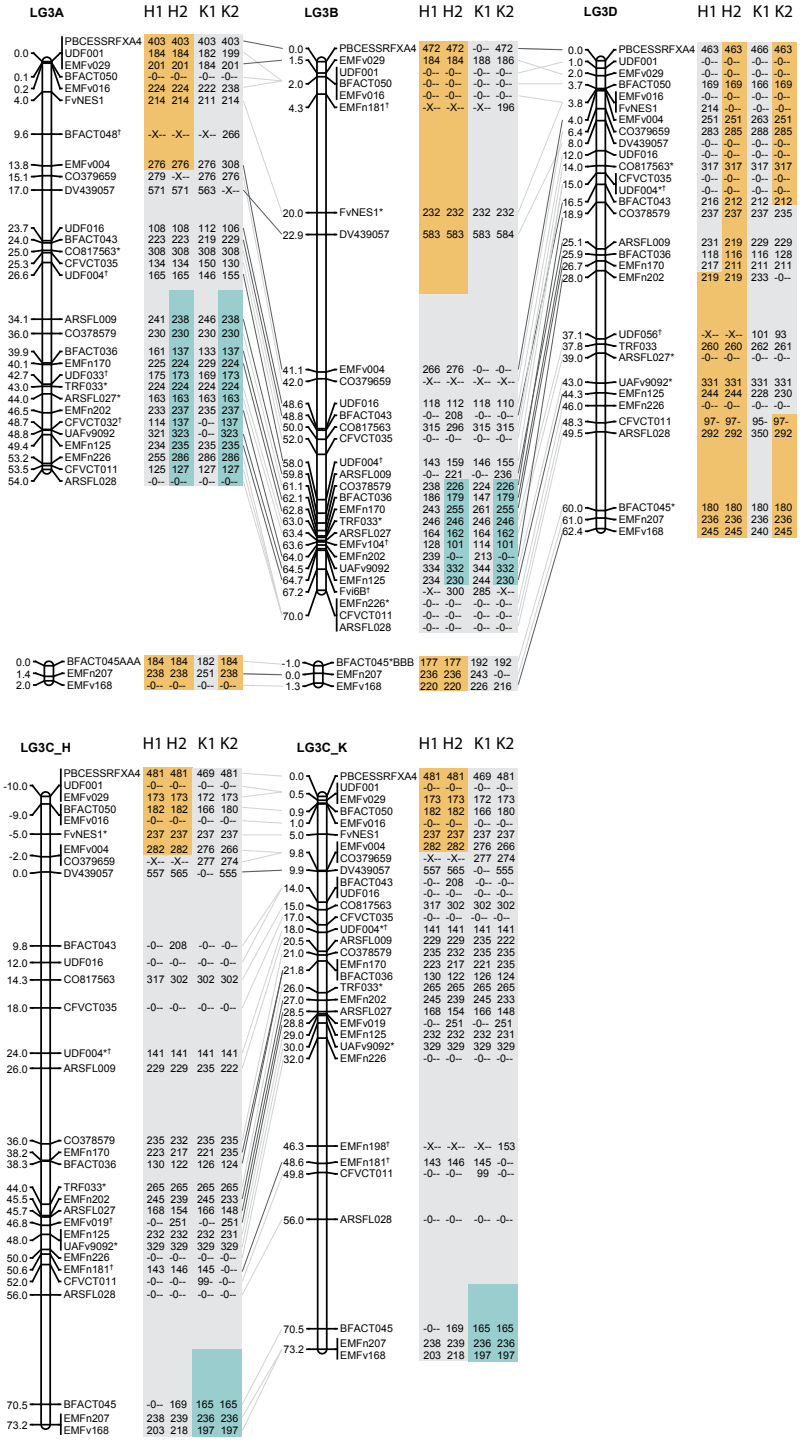
Additional_file_1 as XLSX
Additional file 1: Table S1. Primer pairs. Excel sheet with primer pairs used in this study. Available online at :
http://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-14-55
doi: 10.1186/1471-2229-14-55

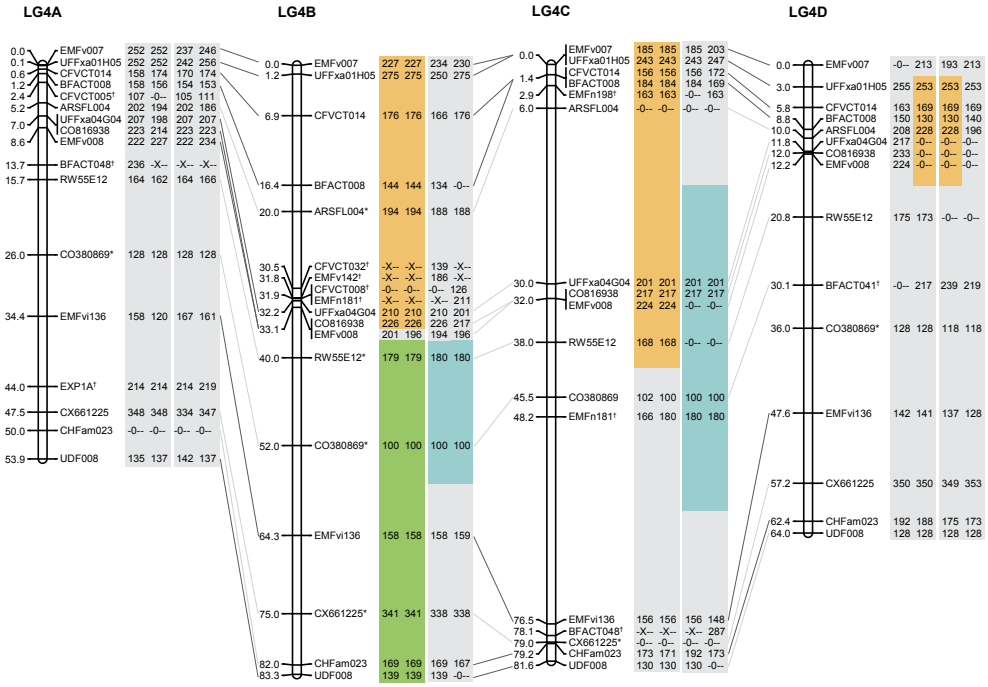
Additional_file_2 as PDF
Additional file 2: Figure S1Holiday x Korona All Maps. All linkage maps of this study. For further description see Figure 2 of manuscript. Available on the next few pages. Online at http://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-14-55
doi: 10.1186/1471-2229-14-55

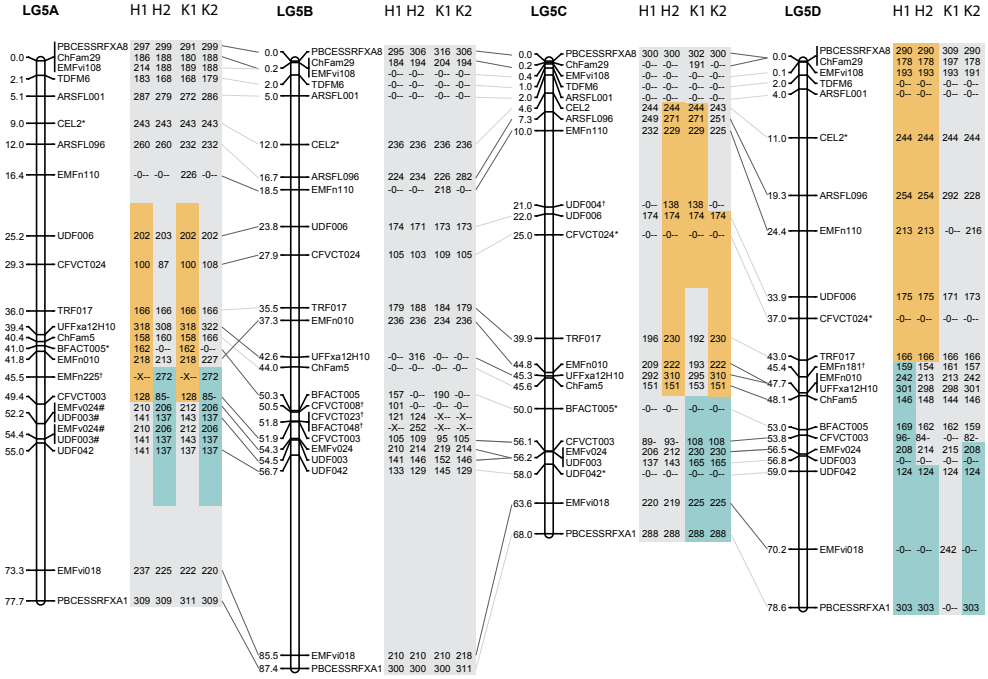
Additional_file_3 as PDF
Additional file 3: Figure S2 Comparative maps All Chromosomes. Comparison between our octoploid linkage maps and the diploid physical and genetic maps. For further description see Figure 6 this chapter. Available online at :
http://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-14-55
doi: 10.1186/1471-2229-14-55

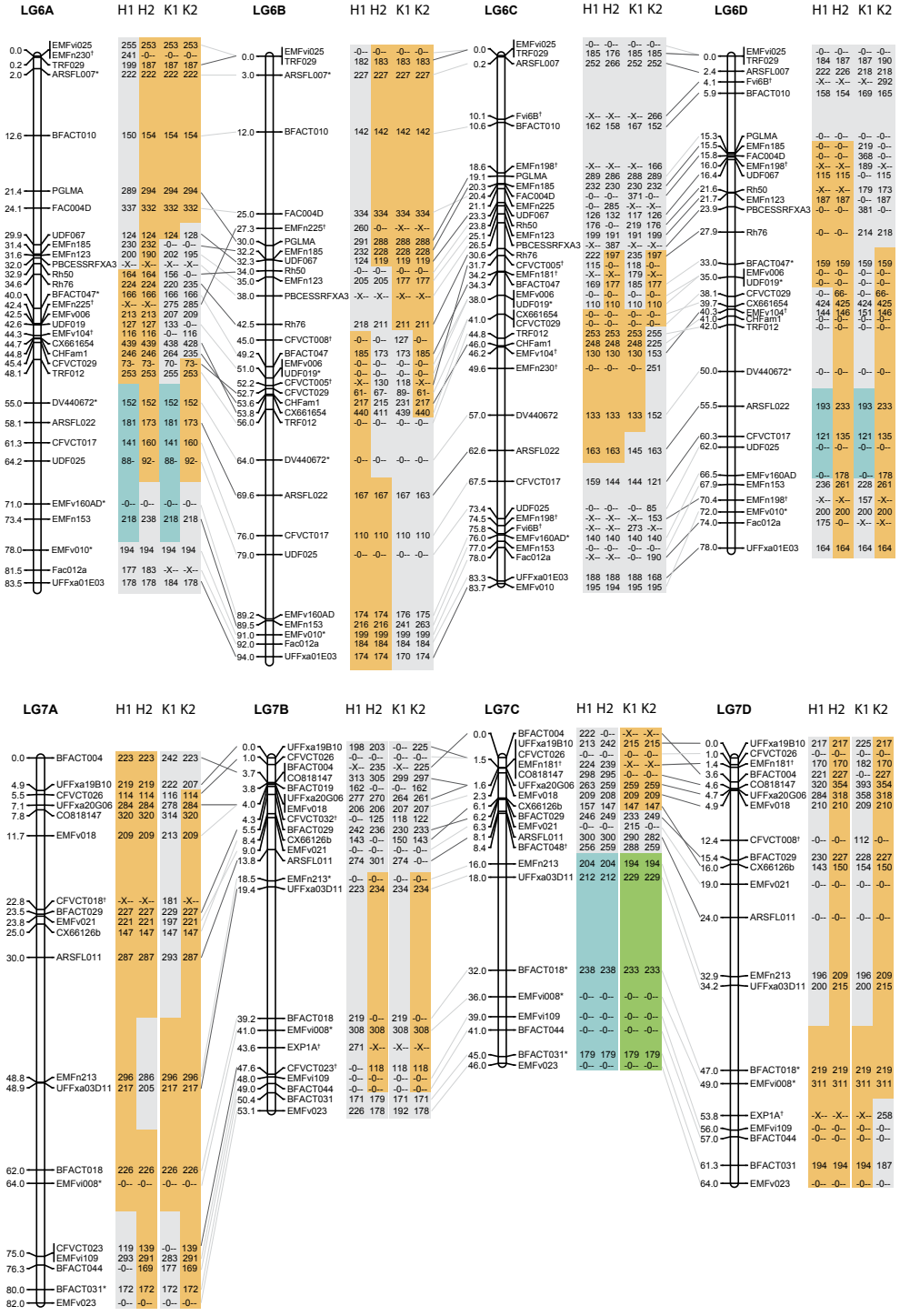
Additional file 2: Figure S1Holiday x Korona All Maps. For further description see Figure 2 of this chapter.


LG1A
LG1B
LG1C
LG1D
LG2A
LG2B
LG2C
LG2D


LG3A
H1 H2 K1 K2
0.0 PBCESSRFXA4 403 403 403 403
UDF001 184 184 182 199
0.1 EMFv029 201 201 184 201
0.2 BFACT050 -0-- -0-- -0-- -0--
4.0 EMFv016 224 224 222 238
FvNES1 214 214 211 214
9.6 BFACT048† -X-- -X-- -X-- 266
13.8 EMFv004 276 276 276 308
15.1 CO379659 279 -X-- 276 276
17.0 DV439057 571 571 563 -X--
23.7 UDF016 108 108 112 106
24.0 BFACT043 223 223 219 229
25.0 CO817563* 308 308 308 308
25.3 CFVCT035 134 134 150 130
26.6 UDF004† 165 165 146 155
34.1 ARSFL009 241 238 246 238
36.0 CO378579 230 230 230 230
39.9 BFACT036 161 137 133 137
40.1 EMFn170 225 224 229 224
42.7 UDF033† 175 173 169 173
43.0 TRF033* 224 224 224 224
44.0 ARSFL027* 163 163 163 163
46.5 EMFn202 233 237 235 237
48.7 CFVCT032† 114 137 -0-- 137
48.8 UAFv9092 321 323 -0-- 323
49.4 EMFn125 234 235 235 235
53.2 EMFn226 255 286 286 286
53.5 CFVCT011 125 127 127 127
54.0 ARSFL028 -0-- -0-- -0-- -0--
LG3B
H1 H2 K1 K2
0.0 PBCESSRFXA4 472 472 -0-- 472
1.5 EMFv029 184 184 188 186
UDF001 -0-- -0-- -0-- -0--
2.0 BFACT050 -0-- -0-- -0-- -0--
EMFv016 -0-- -0-- -0-- -0--
4.3 EMFn181† -X-- -X-- -X-- 196
20.0 FvNES1* 232 232 232 232
22.9 DV439057 583 583 583 584
41.1 EMFv004 266 276 -0-- -0--
42.0 CO379659 -X-- -X-- -X-- -X--
48.6 UDF016 118 112 118 110
48.8 BFACT043 -0-- 208 -0-- -0--
50.0 CO817563 315 296 315 315
52.0 CFVCT035 -0-- -0-- -0-- -0--
58.0 UDF004† 143 159 146 155
59.8 ARSFL009 -0-- 221 -0-- 236
61.1 CO378579 238 226 224 226
62.1 BFACT036 186 179 147 179
62.8 EMFn170 243 255 261 255
63.0 TRF033* 246 246 246 246
63.4 ARSFL027 164 162 164 162
63.6 EMFv104† 128 101 114 101
64.0 EMFn202 239 -0-- 213 -0--
64.5 UAFv9092 334 332 344 332
64.7 EMFn125 234 230 244 230
67.2 FvI6B† -X-- 300 285 -X--
EMFn226* -0-- -0-- -0-- -0--
70.0 CFVCT011 -0-- -0-- -0-- -0--
ARSFL028 -0-- -0-- -0-- -0--
LG3D
H1 H2 K1 K2
0.0 PBCESSRFXA4 463 463 466 463
1.0 UDF001 -0-- -0-- -0-- -0--
2.0 EMFv029 -0-- -0-- -0-- -0--
3.7 BFACT050 169 169 166 169
EMFv016 -0-- -0-- -0-- -0--
3.8 FvNES1 214 -0-- -0-- -0--
4.0 EMFv004 251 251 263 251
6.4 CO379659 283 285 288 285
8.0 DV439057 -0-- -0-- -0-- -0--
12.0 UDF016 -0-- -0-- -0-- -0--
14.0 CO817563* 317 317 317 317
15.0 CFVCT035 -0-- -0-- -0-- -0--
UDF004*† -0-- -0-- -0-- -0--
16.5 BFACT043 216 212 212 212
18.9 CO378579 237 237 237 235
25.1 ARSFL009 231 219 229 229
25.9 BFACT036 118 116 116 128
26.7 EMFn170 217 211 211 211
28.0 EMFn202 219 219 233 -0--
37.1 UDF056† -X-- -X-- 101 93
37.8 TRF033 260 260 262 261
39.0 ARSFL027* -0-- -0-- -0-- -0--
43.0 UAFv9092* 331 331 331 331
44.3 EMFn125 244 244 228 230
46.0 EMFn226 -0-- -0-- -0-- -0--
48.3 CFVCT011 97- 97- 95- 97-
49.5 ARSFL028 292 292 350 292
60.0 BFACT045* 180 180 180 180
61.0 EMFn207 236 236 236 236
62.4 EMFv168 245 245 240 245
0.0 BFACT045AAA 184 184 182 184
1.4 EMFn207 238 238 251 238
2.0 EMFv168 -0-- -0-- -0-- -0--
-1.0 BFACT045*BBB 177 177 192 192
0.0 EMFn207 236 236 243 -0--
1.3 EMFv168 220 220 226 216
LG3C_H
H1 H2 K1 K2
-10.0 PBCESSRFXA4 481 481 469 481
UDF001 -0-- -0-- -0-- -0--
-9.0 EMFv029 173 173 172 173
BFACT050 182 182 166 180
EMFv016 -0-- -0-- -0-- -0--
-5.0 FvNES1* 237 237 237 237
-2.0 EMFv004 282 282 276 266
CO379659 -X-- -X-- 277 274
0.0 DV439057 557 565 -0-- 555
9.8 BFACT043 -0-- 208 -0-- -0--
12.0 UDF016 -0-- -0-- -0-- -0--
14.3 CO817563 317 302 302 302
18.0 CFVCT035 -0-- -0-- -0-- -0--
24.0 UDF004*† 141 141 141 141
26.0 ARSFL009 229 229 235 222
36.0 CO378579 235 232 235 235
38.2 EMFn170 223 217 221 235
38.3 BFACT036 130 122 126 124
44.0 TRF033* 265 265 265 265
45.5 EMFn202 245 239 245 233
45.7 ARSFL027 168 154 166 148
46.8 EMFv019† -0-- 251 -0-- 251
48.0 EMFn125 232 232 232 231
UAFv9092* 329 329 329 329
50.0 EMFn226 -0-- -0-- -0-- -0--
50.6 EMFn181† 143 146 145 -0--
52.0 CFVCT011 -0-- -0-- 99- -0--
56.0 ARSFL028 -0-- -0-- -0-- -0--
70.5 BFACT045 -0-- 169 165 165
73.2 EMFn207 238 239 236 236
EMFv168 203 218 197 197
LG3C_K
H1 H2 K1 K2
0.0 PBCESSRFXA4 481 481 469 481
0.5 UDF001 -0-- -0-- -0-- -0--
EMFv029 173 173 172 173
0.9 BFACT050 182 182 166 180
1.0 EMFv016 -0-- -0-- -0-- -0--
5.0 FvNES1 237 237 237 237
9.8 EMFv004 282 282 276 266
CO379659 -X-- -X-- 277 274
9.9 DV439057 557 565 -0-- 555
BFACT043 -0-- 208 -0-- -0--
14.0 UDF016 -0-- -0-- -0-- -0--
15.0 CO817563 317 302 302 302
17.0 CFVCT035 -0-- -0-- -0-- -0--
18.0 UDF004*† 141 141 141 141
20.5 ARSFL009 229 229 235 222
21.0 CO378579 235 232 235 235
EMFn170 223 217 221 235
21.8 BFACT036 130 122 126 124
26.0 TRF033* 265 265 265 265
27.0 EMFn202 245 239 245 233
28.5 ARSFL027 168 154 166 148
28.8 EMFv019 -0-- 251 -0-- 251
29.0 EMFn125 232 232 232 231
30.0 UAFv9092* 329 329 329 329
32.0 EMFn226 -0-- -0-- -0-- -0--
46.3 EMFn198† -X-- -X-- -X-- 153
48.6 EMFn181† 143 146 145 -0--
49.8 CFVCT011 -0-- -0-- 99 -0--
56.0 ARSFL028 -0-- -0-- -0-- -0--
70.5 BFACT045 -0-- 169 165 165
73.2 EMFn207 238 239 236 236
EMFv168 203 218 197 197


LG4A
LG4B
LG4C
LG4D


LG5A
LG5B
LG5C
LG5D


LG6A
LG6B
LG6C
LG6D
LG7A
LG7B
LG7C
LG7D
H1 H2 K1 K2

Chapter 4

Mapping QTL for resistance against *Phytophtora cactorum* in strawberry.

Thijs van Dijk, Remmelt Groenwold, Herma Koehorst, Yolanda Noordijk, Richard G.F. Visser, Eric van de Weg

ABSTRACT

Phytophthora cactorum, the causal agent of crown rot, is a soil borne pathogen that can cause serious damage for both strawberry growers and propagators when susceptible varieties are grown. Resistance to *P. cactorum* is known to be highly quantitative. To help breeders increase the overall resistance level of strawberry varieties, we performed a QTL mapping study aimed to identify predictive markers that may be used in marker assisted breeding. For this we used two full-sib families with moderately resistant parents, and three years of phenotypic data from controlled plant disease tests.

Considerable differences were found in disease pressure between years, but not between mapping populations. Disease score correlations between years were low, indicating a high environmental component to resistance. QTL analysis revealed indications for minor QTLs in the Holiday x Korona (HxK) population, of which one was recovered when the average disease score over three years (AOTY) was used. In the other population, E1998-142 x Elsanta (ExEls), we identified two significant QTLs (on LG7C and LG7D). The LG7D QTL showed up in two years and, when using AOTY as a phenotype, it became more significant than in any of the three years individually, explaining 22% of the phenotypic variance.

The LG7D QTL appears stable over the years, and its resistant allele greatly reduces the chance of high susceptibility. However, for maximum effectivity in breeding, care needs to be taken that it is not crossed into highly susceptible backgrounds. For future QTL studies, it is recommended to use highly resistant times highly susceptible parental crosses.

INTRODUCTION

Phytophtora cactorum is a soil borne oomycete that causes leather rot (fruits) and crown rot (plant) in strawberries [1, 2]. *P. cactorum* has a broad host range spanning 200 species and 160 genera [3], and should therefore be considered a generalist pathogen. Leather rot develops when sporangia or zoospores from the soil are splashed onto the fruit by rain or irrigation. Crown rot occurs through the colonization of roots by zoospores and subsequent infection of the crown (rhizome) as well as direct infection of the crown by spores through splash water. When infection is severe, the entire crown attains a reddish brown discoloration, and water transport is blocked, which leads to wilting of the plant and eventual death [4]. In Europe, crown rot is the main cause of economic damage by *Phytophtora cactorum*in strawberry cultivation. Poorly drained soils or substrates and deep planting can greatly increase the risk

of infection in open field systems [5]. In protected culture, the recirculation of drainage water increases the risk of infection [6]. The most common source of initial infection in strawberry production fields is infected transplants. Once established in the soil, *Phytophtora cactorum* can remain viable many years, even in absence of host plants [7, 8]. Management practices are focused on preventive measures such as good drainage practices, soil fumigation and use of disease free transplants [9]. Strawberry cultivars can vary greatly in their susceptibility [4, 10-21]. The use of resistant varieties shows promise because resistance and susceptibility of varieties appears fairly consistent over different studies [4] indicating that *P. cactorum* has not evolved specific adaptation to different sources of resistance. However, selection for resistance is hampered by the highly quantitative nature thereof, which is confounded by a large influence of environment, plant age and physiology [4, 13]. For this reason, assessment of resistance often occurs only for advanced selections and only leads to dismissal in cases of extreme susceptibility (Bert Meulenbroek, breeder at Fresh Forward Breeding B.V., NL; personal communication, [4]).

Several recent varieties such as Elegance, Florida Radiance/Fortuna, Malling Pearl, Malling Centenary, Malwina and Sonata are highly susceptible to *P. cactorum* (Van Dijk, data not shown), necessitating very strict adherence to preventive measures in propagation and production fields of these varieties. To help breeders prevent future introduction of very susceptible varieties, we aimed to identify genetic regions underlying quantitative *Phytophthora cactorum* resistance. In addition we had a specific focus on identifying QTLs with stable expression over several years, as these are most useful for breeding. We used two genetically related full sib families made with moderately resistant varieties, which were phenotyped over three years and SSR genotyped using the MADCE method [22].

MATERIAL AND METHODS

Plant material

Two different progenies, Holiday x Korona (HxK) and E1998-142 x Elsanta (ExEls) each consisting of 133 offspring, were used for our QTL analysis. A partial pedigree of the mapping populations is shown in Figure 1. All four parents of these populations were found to be of intermediate resistance to *P. cactorum* in previous disease tests, as well as in practice. Reference cultivars such as the very resistant cultivar Senga Sengana and selection E1996-120, the intermediate cultivar Elsanta and the very susceptible cultivars Tamella [4] and Avanta were included in the disease tests. The mapping populations were created and are owned by the private breeding company

Fresh Forward Breeding B.V. (Eck en Wiel, The Netherlands). Runner plants were rooted in vermiculite for just over two weeks for the 2008 and 2010 disease tests and for four weeks for the 2011 test. At the day of inoculation the plants were cleaned of vermiculite and stored in tap water for a short period of time, awaiting inoculation.

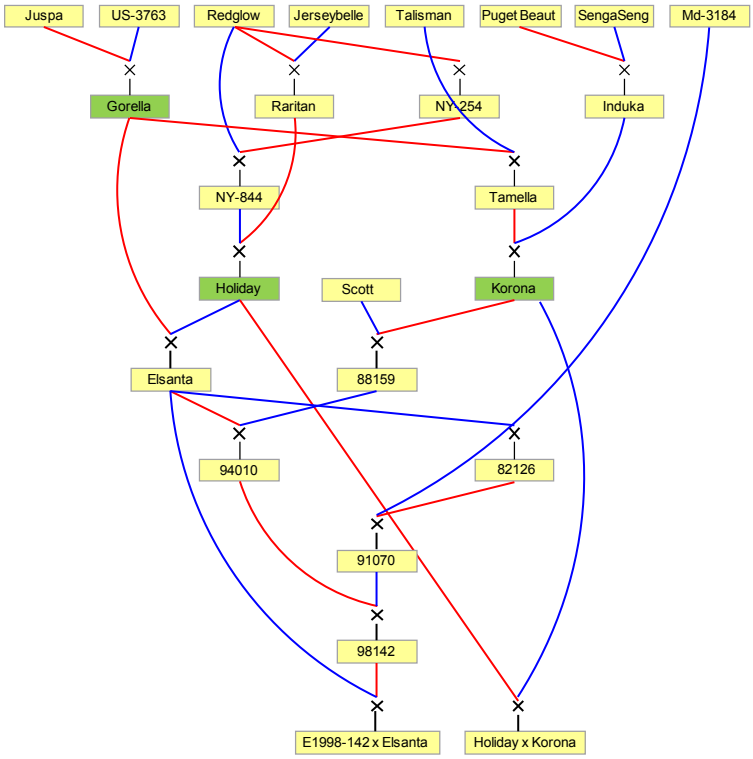


Figure 1. Partial pedigree of the mapping populations used in this study. Holiday, Korona and Gorella are highlighted in green because they represent the strongest hubs of kinship between the two mapping populations.

Inoculation dates

The mapping populations were phenotyped in three different years (2008, 2010 and 2011), however in 2008 they were not phenotyped at the same time. The summer 2008 disease test, in which the HxK population was tested, was inoculated on the 10th of July 2008. The Fall 2008 disease test, in which the ExEls population was tested, was inoculated on the 21st of August 2008. The Summer 2010 disease test was inoculated on the 19th of May 2010. The fall 2011 test was inoculated on the 31st of August 2011.

Experimental design

The disease tests of 2008 and 2010 were conducted in a non-randomized block design consisting of two blocks each with 5 replicates grouped per genotype. The rationale behind no randomization was mainly that it was more practical and it reduced the chance of observational errors. The two blocks were spread over three parallel tables in a greenhouse compartment with an ebb and flow watering system. Although the experiment was not randomized, the use of three tables for two blocks meant that the position along the length of the table for each genotype was different between the blocks.

The disease test of 2011 was performed in a what could be best described as "partially" randomized block design consisting of three blocks with three replicates each. Again practical considerations were favored over complete randomization. However we did try to reduce possible location bias by placing the replicates at different positions both within and between blocks.

Inoculation

A single spore isolate (2003-3), isolated from infected open field strawberry plants grown in the Netherlands was maintained on V8 medium. The isolate was revitalized twice yearly by infection of and re-isolation from apple fruits, and was used as a source of inoculum. Inoculum production was done by transferring small mycelial plugs to V8 plates. These plates were then incubated at 21°C under 24h light for about a week. After incubation the plates were starved by adding tap water, which was refreshed every day for five days and after 8 days the plates were dried in a either a laminar- or down-flow cabinet at room temperature. Two to three days after drying, the plates were evaluated for sporangia formation using a light microscope. At the day of inoculation, the plates with the best sporangia formation were selected for further use. In 2008, each plate was divided into four equally sized parts and these parts were divided over four successive plastic inoculation trays, such that each tray received a part from four different plates. The trays contained approximately 7l of cold (<5°C) water and ten bunches of plants, each bunch comprising all runner plants of a line. The next day, plants were gradually removed from these trays while potting them into plastic pots and dividing them over the experimental blocks.

In 2010 and 2011, the selected plates were shortly pulse-blended with tap water. This inoculum was then added to plastic trays with approximately 7l of cold (<5°C) water, such that each tray contained approximately 1.5 blended *Phytophtora cactorum*plates. The water was stirred before racks containing plants were

transferred, making sure that the entire root system was submerged in the inoculum. Each rack contained about 40 lines in bunches of 10 (2010) or 9 (2011) plants. After at least three hours, all plants were removed simultaneously from the inoculum and transferred to clean trays with water. The following morning the plants were potted and transferred to tables to remain there for the duration of the disease test.

Scoring

Plants were evaluated at the earliest eight days after inoculation (DAI), subsequent evaluations occurred every two to three days, until disease progression had slowed down considerably, which was typically after about 50 to 60 DAI. At each evaluation the onset of symptoms (wilting) and death in DAI for each individual plant was recorded. At the final evaluation all remaining living plants had their crown cut through, and were scored for internal browning. This was recorded on a scale of zero (no discoloration) to five (75-100% of crown diameter discolored). All plants that had died before the final cut received a score of five for their internal discoloration. For our QTL analysis we used four different final scores: The first consisted out of the average wilting day (WD, higher values are more resistant), the second out of the average cutting score (CS, higher values are more susceptible). Under high disease pressure WD is more indicative of resistance, whereas under low disease pressure CS becomes more indicative. To have a better comparison between years we created a third parameter which integrated and normalized CS and WD and was calculated as (square root((1/average wilting day) * average cutting score)))*10 (simply called COMBI, for combined, higher values are more susceptible). The square root transformation pushes extremes more towards the bulk of observations and therefore acts as normalization. The multiplication between WD and CS, integrates the two parameters. A multiplier of 10 was used to make the values of COMBI fall into an easy to use 1-10 range. The last parameter was the Average Over The Years (AOTY) in which a population was tested. This was calculated as the average of the COMBI scores expressed as a percentage of the susceptible control (Tamella) of each year, thus normalizing differences in disease intensity between experiments. For the HxK population 125 out of 133 offspring were phenotyped in all three years and for the ExEls population, 124 out of 133.

DNA isolation

Genomic DNA was extracted according to a modified version of the Fulton et al. [23] mini-prep protocol. This protocol is described in Van Dijk et al. [22].

Genotyping

For our genotyping we used a subset (Supplemental table 1) of the markers used in creating the previously published Holiday x Korona (HxK) SSR linkage map [24]. These were selected based on their position, ability to amplify multiple homoeologues and general ease of use. A few additional SSR markers were used that were not present in the HxK map, mainly to cover regions where no useful previous markers could be found.

Marker Analysis

PCR conditions for all markers were identical as those described by Van Dijk et al. [24]. We used the marker order and positions from the Holiday x Korona map [24] for both populations, rather than making de novo linkage maps. Marker analysis was similar to that described by Van Dijk et al. [24]. However, after identifying the allelic pairs for each locus (and thus sub-genome), the phase of alleles from each linkage group was determined automatically using FlexQTL [25]. For the few additional markers which were not on the HxK map we identified their position and phase manually. For this, we looked at co-segregation of markers which were supposed to be on the same linkage group according to literature, and then positioned them more precisely through graphical genotyping [26].

QTL analysis

QTL analysis was performed using the Genstat v 16.2 QTL module (VSN International Ltd). Single family single environment simple interval mapping was used to identify QTLs using integrated (CP population) genotypic data. The QTL detection threshold according to the Genstat derived Li-Ji values (α 0.05) were 3.377 and 3.428 for HxK and ExEls respectively. Nonetheless, we decided to report all QTLs above the threshold of 3.0, falling in range with what is suggested by Lander and Botstein [27]. The reason for this deviation is that through minimizing the chance of false positives (Type 1 error) you maximize the chance of false negatives (Type 2 error). This is especially the case for small effect QTLs[28], which are more prone to falling below the detection threshold. Cofactor analysis, using the most highly significant locus for each individual QTL, was performed up to three consecutive times to assess the reliability of previously identified QTLs and to identify additional QTLs. For the Cofactor analysis we used the same detection threshold as for simple interval mapping. Explained variances of the QTLs were calculated using the simple or multiple linear regression module of Genstat.

Results and Discussion

Phenotyping

In the test of Fall 2008, two sets of 8 and 9 sequential lines showed up for which no plant death was observed. This was highly remarkable since only few other lines had no plant death observed, and those that did were not sequential. We concluded that something must have gone wrong with either the plant rearing, inoculation or observation of these two groups of genotypes. These 17 lines were excluded from further analyses of the 2008 data.

Summary statistics of the *P. cactorum* disease assays on the full-sib families are shown in Table 1. The relatively low (≤50%) coefficients of variation for all disease assays, illustrate that major effect QTLs are unlikely to be found. The disease pressure in the tests of 2008 and 2010 was high, whereas for the disease test of 2011, the disease pressure was low.

The largest variance overall was observed in the Summer 2010 test. The differences between the two full-sib families for individual tests are largest in the Fall 2011 test, where the amount of variance in the ExEls population is much lower than that of HxK. Another point where both families differ is the cutting score (CS) of 2008, where HxK had a relatively low variance. Despite these differences, it can be concluded that the two mapping populations do not show large differences between them in disease response, especially with regards to mean disease score values. The three parameters (WD, CS and COMBI) were highly correlated within each year (0.88 on average), the only exception (0.6) was the correlation between WD and CS for the ExEls population in the Fall 2011 test. The correlation between the years was generally low. For the HxK population the correlation between traits between years averaged at 0.28 and ranged between 0.18 and 0.33). For ExEls the average value was 0.29 and ranged from 0.19-0.47). This indicates that between experiments there is low heritability and that the chance of finding QTLs that are stable over the different years is small. The average over the years (AOTY) is plotted for each population and several reference cultivars in Figure 2. The AOTY distribution shows that the ExEls population has slightly more extreme phenotypes, especially in resistance, but that otherwise the distributions are quite similar. There is clear transgressive segregation in the populations, as can be judged by the phenotypes of the parents.

Performance of reference cultivars

The resistant reference variety Senga Sengana showed a similar level of resistance as

the most resistant individuals of the two mapping populations (Figure 2). Selection E1996-120 outperformed Senga Sengana, being clearly more resistant than anything found in the mapping populations. The extremely susceptible variety Avanta is also located outside of the phenotypic distribution for both populations.

Table 1. Summary statistics on disease tests. COMBI is a normalized compound score of both cutting score (CS) and average wilting date (WD). For COMBI and CS, higher means more susceptible, whereas for WD this is opposite.

Trait	Pop	Year	#values	Mean	Min	Max	St Dev	Coeff. Var.%
WD	HxK	Sum 2008	133	31.14	14.6	60	12.37	39.7
WD	HxK	Sum 2010	125	27.49	9.6	60	13.40	48.7
WD	HxK	Fall 2011	133	51.92	18.1	60.6	9.14	17.6
CS	HxK	Sum 2008	133	4.67	3	5	0.41	8.8
CS	HxK	Sum 2010	125	3.91	0.4	5	1.15	29.4
CS	HxK	Fall 2011	133	2.26	0.4	4.9	1.00	44.1
COMBI	HxK	Sum 2008	133	4.21	2.4	5.9	0.98	23.3
COMBI	HxK	Sum 2010	125	4.04	0.8	6.8	1.49	36.9
COMBI	HxK	Fall 2011	133	2.06	0.9	4.5	0.66	32.2
WD	ExEls	Fall 2008	116	32.25	9.4	60	14.73	45.7
WD	ExEls	Sum 2010	125	31.18	8	60	15.90	51.0
WD	ExEls	Fall 2011	131	56.71	41.6	60.4	4.37	7.7
CS	ExEls	Fall 2008	116	4.20	1.4	5	0.85	20.1
CS	ExEls	Sum 2010	125	3.70	0.2	5	1.36	36.7
CS	ExEls	Fall 2011	131	2.45	1	4.8	0.66	27.0
COMBI	ExEls	Fall 2008	116	4.16	1.5	7.3	1.50	36.1
COMBI	ExEls	Sum 2010	125	3.81	0.6	7.1	1.70	44.8
COMBI	ExEls	Fall 2011	131	2.06	1.3	3.1	0.33	15.9

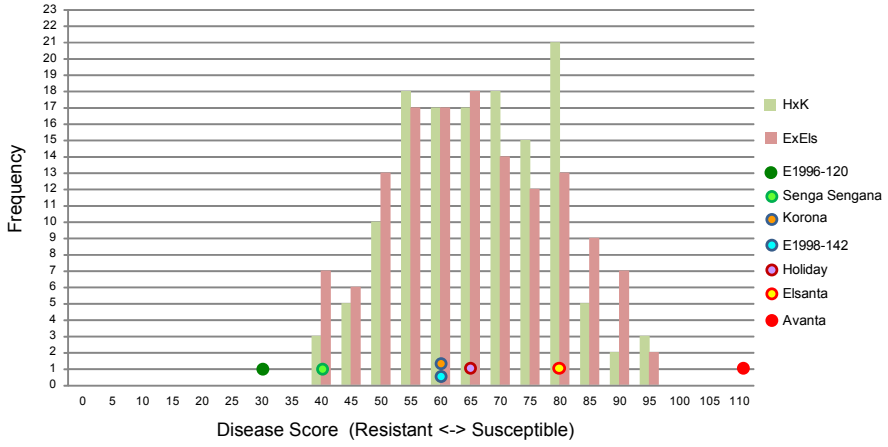


Figure 2. Phenotypic distribution of AOTY (average over the years) disease score for both mapping populations. Parents, as well as some control varieties, are denoted by different dots. resistance allele of LG7D and those that don't (susceptible allele).

Table 2. Interval mapping results. In the column QTL origin, the parent from which the QTL originates is given for additive QTLs. The term interaction signifies a QTL where a specific combination of alleles from both parents shows an effect on the phenotype. The column stable with cofactors indicates whether a QTL remained above detection threshold after one, or several rounds of cofactor analysis.

Pop	Test	Trait	Locus	LG	Pos cM	-10LOG (P)	QTL origin	% expl.	stable with cofactors*
HxK	combined	AOTY	UFFxa20G06	7C	1.6	3.84	interaction	11.8	nt
HxK	Sum 2008	COMBI	EMFn123	6A	31.6	3.2	Korona	5.7	no
HxK	Sum 2008	COMBI	UFFxa20G06	7C	1.6	3.7	interaction	12.1	no
HxK	Sum2008	COMBI	C3P58	3A	58.4	3.03	Korona	8.3	no
HxK	Fall 2011	WD	UFF09F09	2B	68	4.22	interaction	13.3	nt
HxK	Sum 2008	WD	C3P43	3A	43.5	3.41	Korona	8.7	yes
HxK	Sum 2008	WD	EMFn123	6A	31.6	3.05	Korona	6.3	no
HxK	Sum 2008	WD	UFFxa20G06	7C	1.6	3.03	interaction	10.4	no
ExEls	combined	AOTY	BFACT031	7D	61.3	8.2	Elsanta	22.2	nt
ExEls	Sum 2010	CS	C9P10	2B	9.9	3.8	E1998-142	9.6	no
ExEls	Sum 2010	CS	C21P3	7C	3.2	4.25	E1998-142	10.1	yes
ExEls	Sum 2010	CS	BFACT031	7D	61.3	6.73	Elsanta	19.4	yes
ExEls	Fall 2008	COMBI	BFACT031	7D	61.3	3.21	Elsanta	11.4	nt
ExEls	Sum 2010	COMBI	C9P10	2B	9.9	3.57	E1998-142	9.4	no
ExEls	Sum 2010	COMBI	UFFxa20G06	7C	1.6	7.39	E1998-142	14.7	yes
ExEls	Sum 2010	COMBI	BFACT031	7D	61.3	5.23	Elsanta	15.7	yes
ExEls	Sum 2010	WD	C9P10	2B	9.9	3	E1998-142	8.5	no
ExEls	Sum 2010	WD	UFFxa20G06	7C	1.6	5.77	E1998-142	12.1	yes
ExEls	Sum 2010	WD	BFACT031	7D	61.3	6.21	Elsanta	18.2	yes

*nt is not tested

QTL mapping

An overview of the QTL analysis results for the two mapping populations is shown in Table 2. The QTLs we found for the HxK population barely exceeded the significance threshold. The strongest HxK QTL (Fall 2011 WD) showed up on LG 2B as an interaction QTL and therefore dependent on specific allele combinations from both parents. However, further inspection of the location of this QTL revealed that it is suspicious because the surrounding area (>10cM) and the locus itself are homozygous for Holiday, and therefore there is scant information on which to base the dominance predictor. Of the four different putative QTLs observed in HxK, only the QTL on the top of LG 7C was above the threshold for the average over the years (AOTY). This QTL was an interaction QTL, indicating that a specific combination of alleles from both Korona and Holiday showed an effect on the phenotype. This makes the QTL less reliable, because marker UFFxa20G06 only segregates in Holiday and relies on imputed data from an adjacent marker of Korona.

The same locus was also detected in the ExEls population, albeit only in one year (summer 2010). For this population, the LG7C QTL has the highest significance level that was recorded for individual years (7.39). However, there are a few indications that this QTL might not be as solid as expected. Firstly, there is a rapid decline in -10LOGP value over a small genetic distance (it drops below detection threshold within 5cM). This indicates that only few recombinant individuals account for the high score. Secondly, the marker itself has a skewed segregation (2:1, favoring susceptible allele) and it is allele dose based, which makes scoring and especially the assignment of alleles more difficult.

QTL on LG7D

The most interesting QTL that was identified in this analysis was on LG 7D (marker BFACT031) in the ExEls population. This QTL reached the detection threshold in two separate years. In addition, it was most significant for the average over the years (AOTY) trait, with a -10LOGP of 8.2 and an explained variance of 22.2%. This demonstrates that this QTL is quite robust over the years and therefore could be of interest for breeding purposes. We decided to test if this locus had any effect in the 2011 test as well, and found that although the locus did not reach the detection threshold, it still showed an explained variance of 8.3%.

A distribution of the phenotype for AOTY for progeny with and without the resistant QTL allele on LG7D is shown in Figure 3. These distributions show that genotypes carrying the resistant allele were rarely very susceptible, but that

the presence of the resistant allele is not a guarantee for sufficient resistance. In addition, the two most resistant classes of individuals do not contain plants with a susceptible allele. This shows that presence of the resistant allele is required to attain the highest resistance levels. Even so, the main effect of the QTL appears to be a much reduced likelihood of high susceptibility, since the decline in frequency of the resistant allele into the susceptible phenotype classes is more abrupt than the decline of the susceptible allele into the resistant phenotype classes.

The resistance allele of the QTL is derived from the variety Elsanta, which in turn inherited it from Holiday. We were surprised that this QTL did not show up in the HxK population, which has Holiday as one of the parents, but found that Holiday is homozygous at the bottom of LG 7D for the markers that were used in this investigation. Recently we obtained high-density SNP data from the 90k Istraw SNP array [29] for part of the HxK population. We decided to check if any SNP markers could be found that segregate for Holiday at the bottom of LG7. We found that at the most distal end, close to marker BFACT031, Holiday switched from homozygous to heterozygous, and we used these markers for QTL analysis in Genstat. This resulted in elevated peaks at the bottom of LG7D for the data of summer 2010 and the average over the years, but unfortunately these peaks were below the threshold (at about ~2.5). It could be that the number of individuals (79) for which we had SNP data available was too few to reach this threshold.

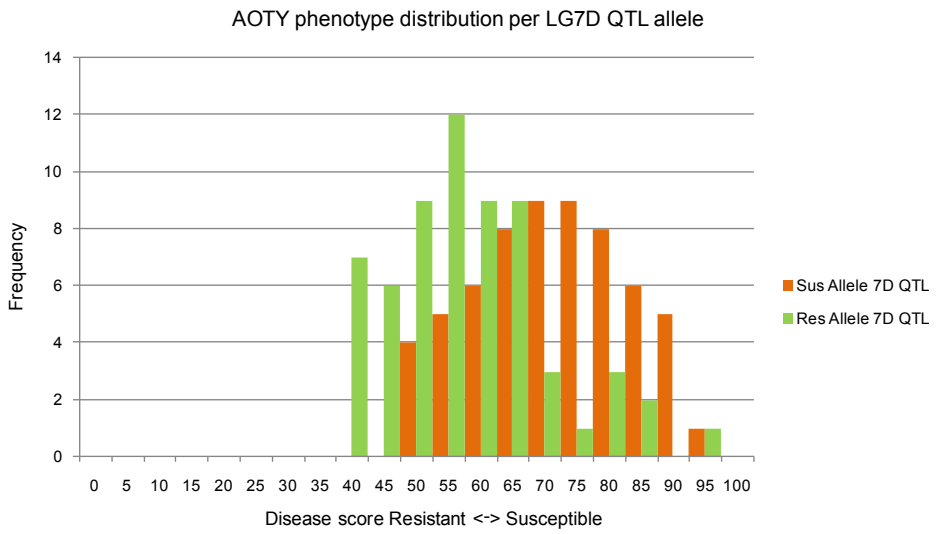


Figure 3. Frequency distribution of AOTY phenotype in ExEls progeny carrying the resistance allele of LG7D and those that don't (susceptible allele).

Combinatory effects of LG7C and LG7D in summer 2010

Since the QTLs on LG7C and LG7D were the only to remain stable in cofactor analysis for the same traits in one year (summer 2010), we decided to check their combined and interactive effect on resistance levels. The results of this analysis are shown in Table 3. It is clear that the added benefit of two resistance QTLs is only a slight improvement in disease resistance compared to presence of just one QTL. The explained variance of the two QTLs in an additive model was 28.1, 27.4 and 27 % for COMBI, WD and CS respectively. Adding the interaction term to the model increased the explained variance only by 1 to 2% depending on trait, which was negligible.

Table 3. Combined effect of resistant (+) and susceptible (-) alleles of QTL LG7C and LG7D on disease resistance per trait.

S2010_Combi Score			**S2010_Cutting Score**			**S2010_Wilting date**		
	7C			7C			7C	
7D	+	-	7D	+	-	7D	+	-
+	2.6	3.2	+	2.8	3.4	+	43	35
-	3.2	4.9	-	3.1	4.6	-	37	21

CONCLUSION

Disease Test

The low correlation between the disease tests showed that environment is an important factor in the expression of *Phytophtora cactorum* resistance. This is in line with previous research on this pathogen in strawberry, where it was found that seasonality, and physiology of the plant can greatly influence the onset of disease. The highly quantitative nature of *Phytopthora cactorum* resistance observed in this study has been observed in many previous studies as well [4, 5, 12-14, 18, 20, 21, 30]. For the disease tests of 2008 and 2010 we had such high disease pressures that it became difficult to separate the intermediate susceptible genotypes from the highly susceptible. For instance, in Fall 2008, Elsanta was found to be more susceptible than our susceptible control Tamella. The disease test of Fall 2011, showed a twice lower mean disease severity than previous tests, we think this was chiefly due to a longer rooting time, which was almost double that of 2008 and 2010. For the 2011 test, the lower disease pressure made it difficult to separate intermediate genotypes

from resistant genotypes. From literature it is also known that highly resistant and highly susceptible varieties react quite similarly between different tests, but that the intermediate types often show large ranges of disease scores between tests [12]. In the Fall 2011 test, the E1998-142 x Elsanta (ExEls) population had a twice lower variance than the Holiday x Korona (HxK) population, whereas the 2010 test showed more alike distributions. A possible explanation could be that the resistance factors that segregate in the ExEls population show up more clearly under high disease pressures.

Combined, the data from these three years could therefore offer better separation from resistant to intermediate to susceptible, than separately. For this reason we analysed the average over the years (AOTY) in our QTL analysis. Another reason was the assumption that a QTL that is still observed after averaging the three years is likely to work in multiple environments, and therefore more useful for breeding purposes.

QTL analysis

The QTL analysis showed several minor QTLs for the HxK population, similar in size and effect to those observed by Denoyes-Rothan et al. [30]. Most of the QTLs we found in the HxK population were of small effect and just above the threshold, and at least one (LG2B) is a likely artefact as it was located on a large chromosomal region for which no marker data were available for one of the parents. The only QTL from the HxK population that remained significant when the AOTY was used, was the QTL on the top of LG7C (marker UFFxa20G06). Interestingly, this same marker underlies a moderate QTL in the ExEls population for the disease test of 2010. However, also in this population the locus showed features that made its reliability doubtful. Another QTL (LG7D) that was observed in the ExEls population showed more promise. This QTL showed up as significant in two out of three disease tests. In addition, the QTL most significant for the AOTY trait, with a -10LOG(P) value of 8.2 and an explained variance of 22.2%. This is an unexpected result, considering the low correlation between disease tests. Phenotypic distributions of genotypes with and without the resistance allele showed that the QTL was required for the two highest resistance classes, and that it had a low frequency (<20%) in the most susceptible classes. This illustrates that in the ExEls cross the LG7D QTL is an important factor for attaining the maximum resistance level, but that the QTL alone cannot prevent high susceptibility in genetic backgrounds were no other factors of resistance are present. In fact, the donor of the resistance allele (Elsanta) is the most

susceptible of the four parents used in this study (Figure 2) and therefore likely to lack other resistance factors. This has consequences for the practical application of this QTL in breeding, where care needs to be taken that the other parent is not overly susceptible, lest the effect of the QTL be watered down through lack of other resistance factors in the progeny.

Disease Parameters in relation to QTL results

The different disease parameters (WD, CS and COMBI), were chosen based on our experience with *P. cactorum* disease trials over the years. Initially, a scoring similar to that of Bell et al. [10] was used, in which plants that die before the end of the trial receive the highest susceptibility scores depending on their time of death, and plants that survive receive the lowest scores depending on the interior discoloration of their crowns. In essence, the time component makes this analogous to an area under disease pressure score. Due to lack of crown discoloration in part of the wilted plants, and discolored crowns occurring in part of the healthy plants, we decided to separate these parameters to see if they lead to different QTLs. The correlation between the parameters (including COMBI) was very high in general. The QTL analysis revealed that separating the parameters had little effect, only one parameter specific QTL was found which was the doubtful LG2B QTL in the HxK population for WD in the Fall 2011 test. No QTLs for CS were found in the HxK population, whereas each of the ExEls QTLs involved all three parameters. Whether or not this lack of QTLs for CS in the HxK population is due to truly different behaviour of the CS trait, or a consequence of not finding any strong QTLs, remains debatable. A likely improvement on the current method of scoring would be to include vitality as a parameter in a similar fashion as performed by Shaw et al. [20, 21]. Currently, the only quantitative aspect of wilting is the date on which it is observed. There are however varying degrees of wilting, as well as varying degrees of plant health for non-wilted plants. This makes inclusion of vitality information necessary for more accurate as well as more agronomically relevant disease scores.

Final remarks

Strong *Phytophthora cactorum* resistance is not a necessity for professional strawberry production as long as resistance is not completely lacking and preventive measures are adequate. However, in the case of highly susceptible cultivars, unexpected losses can occur in both propagation and production fields despite such precautions. We identified several QTLs in this study, of which one proved to be of

moderate effect and stable over the years. This QTL shows potential for practical application in strawberry breeding, if care is taken that it is not introduced in overly susceptible backgrounds. The lack of strong QTLs supports the quantitative nature of resistance and the large influence of environment that was postulated by previous studies [4,20,21,30]. Alternatively, the use of intermediately resistant parents in this QTL discovery study could be the most important factor for the lack of major effect genes. For future genetic studies on *P. cactorum* resistance, the use of highly resistant and susceptible parents such as E1996-120 and Avanta, in the creation of mapping populations is recommended. Crosses made with such parents should provide better opportunities for the detection of major QTLs that are more easily applied in marker assisted breeding procedures for strawberry breeding programs.

References

1. Deutschmann F: **Eine Wurzelfaule an Erdbeeren, hervorgerufen durch *Phytophthora cactorum* (Leb. et Cohn) Schroet.***Nachrichtenblatt des Deutschen Pflanzenschutzdienstes* 1954(6):7-9.
2. Rose DH: **Leather rot of strawberries**. *J Agric Res* 1924(28):357-376.
3. ***Phytophthora cactorum* Pathogen Profile** [http://www.cals.ncsu.edu/course/pp728/cactorum/Pcactorum.html]
4. Eikemo H, Stensvand A, Davik J, Tronsmo AM: **Resistance to crown rot (Phytophthora cactorum) in strawberry cultivars and in offspring from crosses between cultivars differing in susceptibility to the disease**. *Ann Appl Biol* 2003, **142**(1):83-89.
5. Eikemo H, Brurberg MB, Davik J: **Resistance to Phytophthora cactorum in Diploid Fragaria Species**. *Hortscience* 2010, **45**(2):193-197.
6. Martinez F, Castillo S, Carmona E, Aviles M: **Dissemination of Phytophthora cactorum, cause of crown rot in strawberry, in open and closed soilless growing systems and the potential for control using slow sand filtration**. *Sci Hortic-Amsterdam* 2010, **125**(4):756-760.
7. Meyer DS, F.: **Untersuchungen uber die Entwicklung von *Phytophthora cactorum* (Leb. et Cohn) Schroet im Boden.** . *Pflanzenkrankheit* 1975(6):336-354.
8. Sneh BM, D. L.: **Recovery of Sporangia of Phytophthora-Cactorum from Soil**. *Canadian Journal of Botany-Revue Canadienne De Botanique* 1974, **52**(7):1777-1778.
9. Koike STB, G. T. Gordon, T. R.: **UC IPM Pest Management Guidelines : Strawberry**. In. Edited by Program USI. Davis: University of California; 2010: 99.
10. Bell JA, Simpson DW, Harris DC: **Development of a method for screening strawberry germplasm for resistance to *Phytophthora cactorum*.** . *Acta Hort* 1997, **439**:175-179.
11. Eikemo H, Stensvand A, Tronsmo AM: **Evaluation of methods of screening strawberry cultivars for resistance to crown rot caused by Phytophthora cactorum**. *Ann Appl*

Biol 2000, **137**(3):237-244.

12. Parikka P: **Susceptibility of strawberry varieties to crown rot (*Phytophthora cactorum*) in greenhouse tests.** . *Acta Hort* 2003(626):183-189.
13. Parikka P: **Screening for strawberry plant resistance to *Phytophthora cactorum* in nutrient film technique (NFT) system.** . *Acta Hort* 2006, **626**:183-189.
14. Parikka P: **Screening of strawberry resistance to *Phytophthora cactorum*.** . *Sodininkystė ir daržininkystė*2007, **26**(3):23-30.
15. Perez-Jimenez RM, De Cal A, Melgarejo P, Cubero J, Soria C, Zea-Bonilla T, Larena I: **Resistance of several strawberry cultivars against three different pathogens**. *Span J Agric Res* 2012, **10**(2):502-512.
16. Pitrat M, Risser G, Vaillen J, Ricard M: **Susceptibility of Strawberry to Phytophthora-Cactorum in Artificial Inoculation**. *Annales De L Amelioration Des Plantes* 1977, **27**(1):49-60.
17. Van Rijbroek PCL, Meulenbroek EJ, Van de Lindeloof CPJ: **Development of a screening method for resistance to *Phytophthora cactorum*.** . *Acta Hort* 1997, **439**:181-183.
18. Schafleitner S, Bonnet A, Pedeprat N, Rocca D, Chartier P, Denoyes B: **Genetic variation of resistance of the cultivated strawberry to crown rot caused by Phytophthora cactorum**. *J Berry Res* 2013, **3**(2):79-91.
19. Seemuller E: **Resistenzverhalten von Erdbeersorten gegen den Erreger der Rhizomfaule, *Phytophthora cactorum.*** . *Nachrichtenblatt des Deutschen Pflanzenschutzdienst (Braunschweig)* 1977(29):124-126.
20. Shaw DV, Hansen J, Browne GT: **Genotypic variation for resistance to Phytophthora cactorum in a California strawberry breeding population**. *J Am Soc Hortic Sci* 2006, **131**(5):687-690.
21. Shaw DV, Hansen J, Browne GT, Shaw SM: **Components of genetic variation for resistance of strawberry to Phytophthora cactorum estimated using segregating seedling populations and their parent genotypes**. *Plant Pathol* 2008, **57**(2):210-215.
22. Van Dijk T, Noordijk Y, Dubos T, Bink MCAM, Meulenbroek EJ, Visser RGF, Van de Weg WE: **Microsatellite Allele Dose and Configuration Establishment (MADCE): an integrated approach for genetic studies in allopolyploids.***BMC Plant Biology* 2012, **12**(25).
23. Fulton TM, Chunwongse J, Tanksley SD: **Microprep Protocol for Extraction of DNA from Tomato and Other Herbaceous Plants**. *Plant Mol Biol Report* 1995, **13**(3):207-209.
24. Van Dijk T, Pagliarani G, Pikunova A, Noordijk Y, Yilmaz-Temel H, Meulenbroek BJ, Visser RGF, Van De Weg WE: **Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map**. *BMC Plant Biology* 2014, **14**(55).
25. Bink MCAM, Jansen J, Madduri M, Voorrips RE, Durel CE, Kouassi AB, Laurens F, Mathis F, Gessler C, Gobbin D *et al*: **Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple**. 2014,

127(5):1073-1090.
26. Young ND, Tanksley SD: **Restriction Fragment Length Polymorphism Maps and the Concept of Graphical Genotypes**. *Theor Appl Genet* 1989, **77**(1):95-101.
27. Lander ES, Botstein D: **Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps**. *Genetics* 1989, **121**(1):185-199.
28. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao XY *et al*: **Mapping Quantitative Trait Loci Controlling Milk-Production in Dairy-Cattle by Exploiting Progeny Testing**. *Genetics* 1995, **139**(2):907-920.
29. Bassil NV, Davis TM, Zhang HL, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR *et al*: **Development and preliminary evaluation of a 90 K Axiom (R) SNP array for the allo-octoploid cultivated strawberry Fragaria x ananassa**. *BMC Genomics* 2015, **16**.
30. Denoyes-Rothan B, Lerceteau-Köhler E, Guérin G, Bosseur S, Bariac J, Martin E, Roudeillac P: **QTL analysis for resistance to *Colletotrichum acutatum* and *Phytophthora cactorum* in octoploid strawberry (*Fragaria x ananassa*).** . *Acta Hort* 2004, **663**:147-152.

Supplementary Files

Chapter_4_Supplemental Table 1 : Table of used primers in this study, is available under doi: 10.18174/387697. http://library.wur.nl/WebQuery/wurpubs/506541

Chapter 5

Fine mapping the perpetual flowering (PF) trait in cultivated strawberry.

Thijs van Dijk, Elmira Ziya Motalebipour, Yolanda Noordijk, Bert J. Meulenbroek, Richard G.F. Visser, Eric van de Weg

Abstract

In the octoploid cultivated strawberry (*Fragaria x ananassa*) two very distinct flowering habits are in use resulting in seasonal flowering (SF) and perpetual flowering (PF) cultivars. Recently it was shown that a single locus (*FaPFRU*) on the bottom of linkage group 4 is responsible for PF in strawberry for germplasm that carries the *F. virginiana* ssp. *glauca* source of PF.

In this study we confirmed the role and location of *FaPFRU* in *glauca* derived germplasm, and extended this role to earlier sources of PF (Pan American, Gloede's Seedling). We further fine mapped this gene to a region of 900kbp using two mapping populations and a series or PF cultivars and breeding selections. We identified two candidate genes (*FaCDF2* and *FaFT2*) in this region that are homologs to known key regulators of the flowering pathway in other plant species. Sequence analysis of *FaCDF2* showed just two allelic sequences for *FaPFRU* with polymorphisms occurring in the promoter region as well as non-synonymous mutations in the coding sequence.

Marker Haplotype results on genetically divergent germplasm revealed that the PF haplotype of *FaPFRU* is present in 99% of all PF varieties. However, the PF haplotype also occurred in a significant portion of SF varieties (~30%). We found that the PF haplotype was introduced more than a century ago, at the latest, in both PF and SF varieties. We came up with two scenarios that could explain the occurrence of the PF haplotype in SF varieties. The first scenario assumes that the PF haplotype in PF plants and the one in SF plants are functionally different. The second scenario assumes that there is no functional difference between the PF haplotype found in PF and SF plants, but that additional loci can act qualitatively on the switch to PF. So far we have been unable to proof which scenario holds true and further research will be needed to clarify the genetics of PF in strawberry.

Introduction

The optimal timing of a plant's reproduction is critical to the success of its offspring. To monitor the environment for the right conditions for reproduction, the flowering pathway of plants has developed into a complex regulatory network of genes which can integrate data on day length and temperature, and is under the influence of more environmental cues such as nutrient availability and water balance [1]. The domestication of a plant species for agricultural production often leads to selection pressure resulting in a rapid diversification of flowering habits in cultivated material which accomodates production over an extended period of time and different

environments. In the octoploid cultivated strawberry there are two very distinct flowering habits in use: seasonal flowering (SF), leading to so-called June-bearing cultivars, and perpetual flowering (PF) leading to so called everbearing, remontant, or day neutral (DN) cultivars. In SF cultivars, floral initiation is triggered in autumn during short daylengths (SD) (≤ 12h) and moderate temperatures (≤ 12ºCand ≥ 22ºC) [2-6].The plant subsequently enters a semi-dormant state and after winter vernalisation, the plant flowers and sets fruit in spring. As day length and temperature increase in late spring and summer, the SF plants start forming elongated primary stolons (runners) as a form of vegetative propagation [7, 8]. For PF cultivars two types are often distinguished in literature, "old" everbearing (EB) varieties which initiate flowers under long days and more modern day-neutral (DN) varieties derived from *F. virginiana* ssp. *glauca* which supposedly initiate flowers under any day length [7]. The "old" EB plants can be divided into varieties derived from an European source (Gloede's seedling) and varieties derived from an American source (Pan American) [9]. The DN varieties are derived from the *F.vriginiana glauca* DN source [10], which is present in most modern PF cultivars.

However, the ability of DN cultivars to initiate flowers under any day length is a matter of continuing debate among strawberry researchers and there is mounting evidence that there is no qualitative distinction between EB and DN cultivars in their response to day-length [7]. A contributing factor to the confusion is that flower initiation in both SF and PF cultivars is highly temperature dependent and varies between cultivars [4,7,11]. In addition, the use of vegetative material from the previous growing season instead of plants reared directly from seeds can influence the phenotyping [7].

A significant reduction in the ability to produce runners is often observed for PF cultivars compared to SF cultivars [11, 12]. This can be explained by the fact that both flowers and runners are developed from the same axillary meristems. As the signal for flowering is maintained in PF lines, there could be a negative effect on the number of runners that are produced during long days (LD) [13].

Molecular mechanisms in other crops

The molecular mechanisms underlying the flowering pathway have been studied extensively in the LD model plant *Arabidopsis thaliana*[14]. In Arabidopsis, a large number of genes have been identified that are involved in the observation of circadian rhythms. Two central genes that are involved in this are *CONSTANS* (*CO*) and *FLOWERING LOCUST* (*FT*) [15]. *CO* is essential for the observation of day length and follows a diurnal expression pattern. Under permissive conditions *CO* is

able to induce *FT* which encodes the mobile signal protein to the apical meristem where it switches on generative meristem identity genes [14-17].

Many of the genes of the photoperiodic pathway, and especially the central genes, are conserved throughout the plant kingdom [16]. The mode of action of orthologous genes in the pathway does vary between different crops. In rice, there are two *FT* homologs that promote flowering, however one promotes flowering under SD conditions and the other under LD conditions [18]. In addition, the *CO* ortholog of rice (*Hd1)* can switch from a promotor of *FT* into a repressor depending on day length. Flowering in sugar beet (LD) and tuberization in potato (SD), are under control of a dual paralogous *FT* system [19, 20], where one FT homolog suppresses the other.

Molecular mechanisms in diploid strawberry

In strawberry, research into the molecular control of the PF trait was initiated in the diploid woodland strawberry *Fragaria vesca*. Thetrait was mapped to linkage group 6, and found to be controlled by a *TFL1* (*TERMINAL FLOWERING 1*) gene, which belongs to the CETS family, which also includes *FT* [21-24]. *TFL1* is known to be a suppressor of flowering, with opposite action to *FT,* and in diploid *F. vesca* the PF trait is inherited recessively as a non-functional mutant [22, 23]. Although the role of *FvTFL1* is undisputed in the switch from SF to PF in *Fragaria vesca* [23], the role of *FT* is more enigmatic. For PF plants the role of *FvFT1* appears to be straightforward: under LD conditions, where flower initiation occurs, *FvFT1* is highly expressed. In addition, silencing of *FvFT1* in PF plants leads to a substantial delay of flowering. Extrapolating this to SF plants which flower under SD, *FvFT1* expression is expected to be high under SD. However under these conditions there is no expression of *FvFT1* in both SF and PF plants, whereas it is highly expressed under the non-inductive LD in SF plants. It appears that flowering under SD in SF plants uses an *FvFT1* independent pathway [23], or is perhaps even inhibited by *FvFT1*. The paralogous *FvFT2* could be responsible for this contrast. This gene was only examined for a spatial expression study in the SF (SD) *F. vesca,* and was found to be expressed exclusively in the flower bud [23]. The role of *CO* was not investigated in this study.

Mechanism of PF in the octoploid strawberry

An interesting contrast between the diploid and octoploid strawberry is that the inheritance of PF in the octoploid is dominant. This makes it likely that other loci

than those of diploid *F. vesca* are important for PF in the octoploid *F. x ananassa*. Research into the molecular control of the PF trait in cultivated strawberry has been very limited [10]. An expression study on the *CONSTANS* ortholog (*FrCO*), revealed that in SF plants the expression peaks at dawn under SD conditions, whereas PF plants display a weak peak in the middle of the day. Under LD conditions, SF plants show weak *FrCO* expression during the night before dawn, whereas PF plants show no expression of *FrCO* at all [25]. The number of loci involved in the PF trait for cultivated strawberry has long been a matter of debate. There are a number of studies suggesting monogenic inheritance, but polygenic inheritance has been found as well [9, 26-31]. A recent mapping study on a single full-sib family revealed a single locus (*FaPFRU*) on the bottom of one of the chromosome 4 homoeologues to be responsible for PF and reduced runnering in the octoploid strawberry [13]. This finding shows the presence of a major, qualitative switch from SF to PF which appears to be under monogenic control. However, no candidate genes have been mentioned, and the interval of the QTL is still quite large, despite additional studies on this locus [32]. Because of the importance of this trait in breeding, we initiated this study to develop a marker that could be used as a selection tool. For this, we fine mapped the PF trait using two separate full-sib families as mapping populations. Our results indicate that in our mapping populations the trait is under control of a single locus as found previously, and that several major flowering pathway genes are within the region of interest. In addition, we performed haplotyping studies on breeding germplasm and candidate gene sequencing in order to elucidate the mode of action of this locus.

Materials & Methods

Plant material

Two full-sib families, with a total number of 93 individuals, were used for mapping. The first family ("A") derived from the cross of cultivars Evie2 (PF) and Mara des Bois (PF) consisted of 46 individuals, which were selected from several hundred plants based on their flowering habit (23 PF and 23 SF). The PF trait from Evie 2 can be traced back to the variety "Everglade". Although we do not know the parentage of "Everglade" we believe it is derived from the modern *F. virginiana* ssp. *glauca* source for day neutrality (DN). Mara des Bois traces its PF trait to either Ostara (PF) or Hummi Gento (PF). Ostara is derived from the "old" European cultivar Uberreich, which was introduced before 1937 (http://193.205.128.6/agraria/ricerca/prog_ric/Wg1/GeneticResources/European%20Germplasm.htm) and whose parentage

is unknown. This source predates the introduction of the Utah *F. virginiana* ssp. *glauca.* Hummi Gento is derived from a German breeding program, and since it is an old cultivar with unknown genitors, it is more likely to be derived from an "old" European PF source.

The second family ("D") was derived from the cross between cultivars Rumba (SF) and Valor (PF) and consisted of 47 individuals which were also selected for their flowering habit, 24 PF and 23 SF. Valor is a modern PF variety derived from California and is likely to be of *F. virginiana* ssp. *glauca* origin. The F1 populations used in this study were created by the Dutch breeding company Fresh Forward Breeding B.V (Fresh Forward).

An additional set of 303 varieties and lines were included to further examine marker-trait associations that were discovered through the full-sib families. These included 127 PF and 164 SF individuals. Leaf tissue from these individuals came either from the germplasm collection of the USDA-ARS National Clonal Germplasm Repository, Corvallis, Oregon, or Fresh Forward Breeding B.V. For other individuals, DNA was already available through the DNA collection at the Plant Breeding department of Wageningen University & Research.

Phenotyping

The progenies were phenotyped as single potted seedlings in the open field at Fresh Forward Breeding B.V. The phenotyping was performed on the 30th of July 2012 after old trusses from the first fruiting had been removed approximately 6 weeks prior. PF and SF plants were differentiated based on the presence or absence of new flowers in the crown of the plant, ambiguous phenotypes were not included. For cultivars and breeding selections, information came from literature, or, in case of unnamed selections, from Fresh Forward Breeding B.V., these phenotypes were observed during yearly breeding trials over the past two decades.

Genotyping

DNA extraction was performed using a previously described protocol [33]. In this study we used 103 SSR primer pairs, of which 62 were previously published and 41 were developed in this study. Their origin and primer sequences are specified in Supplemental Table 1. Most PCR reactions were performed with indirect fluorescent labelling [44] using a universal 17 bp 5' end tail sequence (AACAGGTATGACCATGA) on the forward primer, which matched a universal fluorescently labelled primer (6-FAM, HEX or ROX) [33]. For the fine mapping

we used published microsatellite markers from the bottom of LG4 from the Holiday x Korona (HxK) map [36] and from the Isobe et al. [34] map. In addition we developed our own microsatellite markers using the *Fragaria vesca* v1.0 reference genome [35] and Tandem repeats finder (http://tandem.bu.edu/trf/trf.html). Notation of linkage groups followed the physical map of the diploid *F. vesca* 'Hawaii 4' [35]), and notation of sub-genomes (A-D) was according to Van Dijk et al. 2014 [36]. Microsatellites markers used in this study were run and analyzed using the MADCE method as described by Van Dijk et al. [33], which included assessment of marker allele doses.

QTL analysis

Because we phenotyped qualitatively based on presence and absence of the trait, we decided to forego traditional map creation and QTL analysis and instead we simply examined allele segregation patterns for their similarity to the segregation of the trait, by eye.

Candidate gene identification

To identify the presence and identity of candidate genes in strawberry, we blasted known flowering pathway genes from Arabidopsis and rice [14, 37, 38] against the *F. vesca* v1.0 LG4 pseudo-chromosome (http://www.rosaceae.org/).

Cloning & sequencing of candidate gene *FaCDF2*

To isolate the *FaCDF2* gene we needed to obtain sub-genome specific primers to avoid amplification of the same gene from other homoeologues. For this, sub-genomic SNPs were identified using resequencing data from SF cultivars Holiday, Korona, two progeny of Holiday and Korona and the PF breeding selection Cal.65.65-601 [9]. These sequences were aligned to the reference *Fragaria vesca* v1.0 genome [49] and further analyzed using the integrative genomics viewer (IGV) software from the Broad Institute [39]. SNPs flanking the *FaCDF2* gene that were present in all cultivars were likely to differentiate sub-genomes, and were incorporated in our primers on the 3' end. With these sub-genome specific primers we amplified *FaCDF2* alleles from eight different varieties (Table 1). To identify nucleotide incorporation errors we performed two independent PCRs per variety using the Multiplex PCR Kit (Qiagen). The PCR conditions were one cycle at 94°C for 15 min followed by 35 cycles at 94°C for 30 s, 58°C for 30 s and 72°C for 2 min, and a final extension

cycle at 72°C for 5 min. PCR fragments were cloned into PGEMT-easy (Promega corporation, Madison, WI, US), and plasmids were extracted using the QIAprep Spin Miniprep kit (Qiagen, Venlo, Limburg, Netherlands). Sequencing was performed using the LIGHTrun sequencing service at GATC-biotech AG (Konstanz, Germany) using a set of primers covering the total fragment (Supplement 2). Sequence analysis was performed using the Lasergene 10 core suite (DNASTAR, Madison, WI, USA).

Haplotyping FaPFRU & identification of additional loci putatively involved in PF

The *FaPFRU* region was haplotyped for a set of varieties and breeding lines of varying origins using three primer pairs (Supplemental Table 1). Several varieties were selected based on their haplotype or origin to receive more extensive haplotyping using additional primer pairs (Supplemental Table 1). Haplotype identification was done manually, using MADCE and subsequent identification of co-occurring alleles combined with inheritance information, in cases where pedigree information was available.

Results

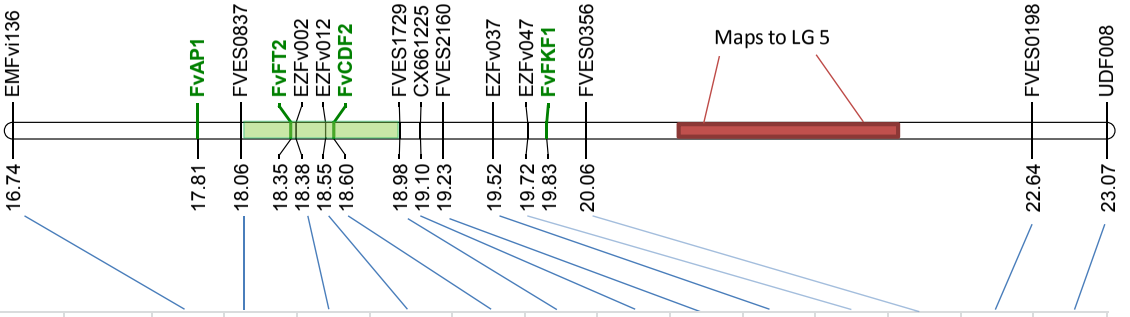
Initial screen for linked SSR markers

In the initial screen of 14 SSR markers (Supplemental Table 1), SSR-locus *CX661225* showed complete co-segregation with flowering habit in population "D" (Rumba (SF) x Valor (PF)). This locus was present at the bottom of LG4A on the HxK map. This is the same genetic region as where *FaPFRU* was mapped by Gaston et al. [13] and confirms their observation of monogenic control. CX661225 also showed co-segregation with the PF trait in population "A", but we could only track alleles from parent Mara des Bois, as Evie 2 was homozygous for this marker.

Identification of candidate genes in the vicinity of CX661225

Blasting the primer sequences of CX661225 against the diploid reference genome *Fragaria vesca* v1.1 (http://www.rosaceae.org/) yielded no hits. Blasts to v1.0 gave a clear hit on the bottom of LG4 at 19.10 Mbp. For this reason we only used v1.0 for identifying candidate genes in the *FaPFRU* region.

We identified four important flowering pathway genes within 3 Mbp of the *CX661225* locus: *FvAP1* (17.81 Mbp), *FvFT2* (18.35 Mbp*), FvCDF2* (18.60 Mbp), and *FvFKF1*(19.83 Mbp) whose closest Arabidopsis homologs are *AP1, FT*, *CDF2*and *FKF1*respectively(Figure 1).



genotype	parent	EMFvi136	FVES0837	EZFv002	EZFv012	CDF2_PSSR	FVES1729	CX661225	FVES2160	EZFv037	EZFv047	FVES0356	FVES0198	UDF008
A_E003	MDB	160	123	268	207	318	305	343	218	147	242	null	--	145
A_D005	MDB	160	115	270	208	357	318	--	220	148	242	null	--	145
A_D003	MDB	167	115	270	208	357	318	336	220	148	231	381		140
A_D007	MDB	167	115	270	208	357	318	--	220	148	242	null	--	140
A_D014	MDB	167	115	270	208	357	318	336	220	148	242	null	--	140
A_D023	MDB	167	115	270	208	357	318	336	220	148	242	null	--	140
A_D014	Evie2	--	123	268	207	318	318	--	220	148	242	null	191	145
A_D020	Evie2	--	123	270	208	357	318	--	220	148	242	null	191	145
A_D023	Evie2	--	115	270	208	357	318	--	220	148	242	null	191	140
A_D006	Evie2	--	115	270	208	357	318	--	220	148	242	384	197	140
A_D024	Evie2	--	115	270	208	357	305	--	225	150	193	384	197	140
A_E004	Evie2	--	115	268	207	318	305	--	225	150	193	--	197	140
D_E006	Valor	--	--	--	--	318	305	348	231	--	--	null	191	145
D_D016	Valor	--	--	--	--	357	318	336	220	--	--	null	191	140
	haplotype													
Mara des Bois	PF	167	115	270	208	357	318	336	220	148	242	null	191	145
	SF	160	123	268	207	318	305	343	216	147	231	381	191	140
Evie2	PF	167	115	270	208	357	318	336	220	148	242	null	191	145
	SF	167	123	268	207	318	305	336	225	150	193	384	197	140
Valor	PF	167	115	270	208	357	318	336	220	--	242	null	191	145
	SF	167	115	270	208	318	305	348	231	--	233	384	197	140

Markers do not recombine with phenotype

Figure 1. Physical positions (in Mbp, *F. vesca* reference v1.0) of markers used in this study compared to the genetic order of markers as observed by recombination events in our mapping populations. The region where the PF gene is located is highlighted in green on the physical map. Candidate genes are presented in **bold** and green color on the physical map. The red bar segment indicates a region for which the developed markers mapped on to LG5 in a mapping study by Isobe et al. [34]. For the recombination events, green represents the SF haplotype and yellow the PF haplotype, numbers indicate allele size in bp. When no allele size is given, the marker was either homozygous for that parent, not used or failed amplification.

Fine mapping the PF trait

The fine mapping results are presented in Figure 1. The candidate gene region was found to be delimited proximally by *FVES0837* and on the distal end by *EZFv047*. Both loci showed two recombinations with PF in the "A"-family. A set of seven markers spanning 800 kb showed no recombinations with the PF phenotype. Among these was a promoter repeat based microsatellite from the *CDF* gene (primer pair CDF_PSSR_1). The original primer combination (CDF_PSSR_1) only amplified the PF allele and was later replaced by CDF_PSSR_2 which also amplified SF alleles. Shown in Figure 1 are the allele sizes of CDF_PSSR_2. We did observe recombinations between these seven markers in two PF progeny (A-D014 and A-D024), however these recombinant lines also contained PF associated marker alleles from the other PF parent, and could therefore not be used directly for further fine mapping of the PF trait.

The number of flowering pathway genes that are within our area of interest has now been reduced to two (*FvFT2* and *FvCDF2*), provided that the physical positions of these candidate genes are correct. Of these two, we considered the *FvCDF2* gene to be more interesting for further investigation, as it is a known repressor of *CO*, which fits with the lack of *CO* expression in PF plants [25]. In addition, the presence of the CDF_PSSR microsatellite facilitates the identification of the right sub-genome for the isolation of *FaCDF2*.

Analysis of candidate gene FaCDF2

A number of sub-genome specific primer pairs were designed around the region of the *FvCDF2* promoter repeat. We tested these together with the previously found CDF_PSSR_1 marker on a subset of family "A". We identified that primer CDF_PSSR_S1_R4 (5'-ATCGTTTTAAGAAATGTTGGG-3') only amplified alleles from the sub-genome containing the PF trait (LG4A). This primer was tested with several sub-genome specific primers on the terminator region of the *FaCDF2* gene, where we observed that in combination with primer FvCDF_WG_2F_S2 (5-CACTATTCAGAATGGAGACAGC-3) clear single bands were amplified. The total fragment size amplified by this primer combination amounted to approximately 3600 bp, containing the entire 3210 bp ORF of *FaCDF2*and part of the promotor region. The CDF_PSSR_S1_R4 primer was also used to create the CDF_PSSR_2 sub-genome specific SSR primer pair (Supplemental Table 1), which, unlike the CDF_PSSR_1 primer pair, is able to amplify SF alleles as well. The resulting marker showed a length polymorphism where the PF allele was around 30bp larger than

the SF allele. We used this primer pair to screen a set of 96 cultivars and breeding selections to establish the specificity of the PF allele, the results of which are shown in Supplemental Table 3. We found that all PF varieties carried the PF allele in mainly heterozygous and sometimes homozygous form. The PF allele was also present in several SF varieties however. Because of these apparent false positives we decided to examine for the presence of additional polymorphisms and sequenced the *F. x ananassa FvCDF2* gene ortholog *FaCDF2* and its allelic variants from a total of eight cultivars including five varieties from three different sources of PF: *glauca* (Brighton, Evie 2), Uberreich (Ostara, possibly Mara des Bois), Pan American (Geneva), and three SF varieties showing a PF allele (Korona, Pajaro, Yamaska) (Table 1). Only two distinct *FaCDF2* alleles (FaCDF2_PF and FaCDF2_SF) were found in the eight varieties tested, with 16 polymorphisms between them (Supplement 4). We could not discover any sequence difference between the *FaCDF2* PF allele present in PF varieties and the one present in SF varieties. The FaCDF2 amino acid sequence alignment of the FaCDF2_PF to FaCFD_SF variants revealed two differences (Supplement 5): a Serine to Asparigine substitution at position 25 and a Proline to Arginine substitution at position 424, which is within the binding domain of the *CDF* repressor *GIGANTEA* (Supplement 6, Dr. Christian Bachem, Wageningen UR Plant Breeding personal communication) and Kloosterman et al. [19]). Although this putative functional polymorphism adds credit to *FaCDF2* as a candidate gene, the presence of the PF allele of *FaCDF2* in some SF cultivars suggests that *FaCDF2* as a single locus does not always confer the PF trait in cultivated strawberry.

Table 1. Varieties screened with marker CDF_PSSR_2 and subsequent selection for sequencing of *FaCDF2*.

name	phenotype	CDF_PSSR_2 alleles	FaCDF sequenced	DNA origin
Brighton	PF	SF,PF	yes	Fresh Forward
Evie_2	PF	SF,PF	yes	Fresh Forward
Geneva	PF	PF	yes	PI 551586
Mara Des Bois	PF	SF,PF	yes	Fresh Forward
Ostara	PF	SF,PF	yes	Fresh Forward
Korona	SF	SF,PF	yes	Fresh Forward
Pajaro	SF	SF,PF	yes	Fresh Forward
Yamaska	SF	PF	yes	Fresh Forward

Table 2. Haplotypes identified by screening three primer pairs within the *FaPFRU* region. Marker alleles from the PF haplotype are highlighted in yellow.

Haplotype NR	EZFv002	CDF_PSSR_2	FVES2160	Associated phenotype	occurrence in SF plant	occurrence in PF plant	occurrence in all plants
1	270	357	220	SF, PF	27	94	59
2	268	318	225	SF	37	44	40
3	285	321	218	SF	20	10	15
4	268	321	220	SF	17	6	12
5	270	357	225	SF, PF	11	9	10
6	268	318	218	SF	9	5	7
7	263	319	225	SF	14	0	7
8	270	317	231	SF	8	3	6
9	290	318	225	SF	5	0	3
10	270	318	231	SF	1	3	2
11	263	319	221	SF	4	0	2
12	268	317	218	SF	2	0	1
13	285	318	225	SF	1	1	1
14	294	354	?	SF	1	1	1
15	268	318	220	SF	1	1	1
16	268	318	226	SF	1	0	1
17	268	319	229	SF	1	0	1
18	268	318	229	SF	0	1	0
19	268	332	225	SF	0	1	0
20	268	321	217	SF	0	1	0
21	268	319	218	SF	0	1	0
22	268	321	217	SF	0	1	0
23	285	321	220	SF	0	1	0
24	263	319	224	SF	1	0	0
25	266	?	221	SF	1	0	0
26	266	319	221	SF	1	0	0
27	268	318	231	SF	1	0	0
28	268	317	229	SF	1	0	0
29	268	319	225	SF	1	0	0
30	268	321	229	SF	1	0	0
31	276	318	225	SF	1	0	0
32	288	364	?	SF	1	0	0
33	294	328	231	SF	1	0	0
34	294	317	?	SF	1	0	0
35	270	357	226	SF	1	0	0
36	274	321	225	SF	1	0	0
37	?	317	227	SF	1	0	0
38	294	354	218	SF	1	0	0
39	285	321	225	SF	1	0	0

SSR haplotyping a set of varieties.

In an effort to find out whether the PF candidate region contains a haplotype which is unique to PF plants, we genotyped a large set of SF and PF varieties using three

SSR loci (*EZFv002, CDF_PSSR_2, FVES2160*).The recombination fraction in our mapping populations between these loci corresponds to a genetic distance of 1.4cM. Despite the presence of recombinations, these three loci were shown to be fully linked to the phenotype in our fine-mapping study. The full haplotyping results are shown in Supplement 7 and an overview is given in Table 2. A total of 39 haplotypes were identified, of which 31had a frequency lower than 5% in the germplasm that was tested. The haplotype diversity in SF cultivars was higher than PF cultivars (33 vs 17 distinct haplotypes).Of the eight haplotypes which were not rare, two showed large contrasts in their distributions between SF and PF cultivars: Haplotype 1, which occurred in 94% and 27% of the PF and SF individuals respectively, and haplotype 7, which occurred in 14% of the SF individuals and was absent in PF individuals. The latter haplotype was specific to several European SF varieties (Table 2).

The PF haplotype distribution in PF and SF plants

At least one full copy of the PF haplotype, as observed in our mapping parents (Table 2), was present in 123 out of 129 PF individuals tested.

A further five PF individuals did not contain a full copy, but did have the first two marker alleles in common with the full PF haplotype. This partial PF haplotype probably is recombinant and is present in 12 out of 129 PF lines. Of these, 11 are the variety Pacific and its direct descendants. The only other PF line that carries the partial PF haplotype is Bolero from East Malling Research. With these recombinations we can discard marker FVES2160 as fully linked with the PF trait, which further reduces the candidate region in size. The only PF line which did not have a complete or partial (EZFv002 + CDF_PSSR_2) PF haplotype was ‘Ozark Beauty’, which had a FaPFRU haplotype that did not resemble that of any PF plant including those from the same donor (Pan American), it could be that this genotype is not true to type.

The haplotyping established that the PF haplotype is required for plants with the PF phenotype. However, just like the PF allele of FaCDF2, the QTL spanning PF haplotype is not unique to PF varieties. Out of the 139 SF cultivars and lines that were genotyped, 35 contained the full PF haplotype and 14the partial PF haplotype, that was also observed for some PF lines, and 2 contained both. The occurrence of the complete and partial PF haplotype in a considerable portion of SF cultivars makes it plausible that multiple loci are required for the PF trait, depending on genetic background. These results also showed that the PF haplotypes from three supposedly distinct sources of PF (*Fragaria virgininiana* ssp *glauca*, Pan American and European everbearers) [9] and even from SF cultivars are the same, which suggests a common point of origin.

Extensive haplotyping to unravel the origin of the PF haplotype

To add more weight to the hypothesis of a common point of origin, we haplotyped more extensively using a larger set of markers. For this we used a smaller set of cultivars including the oldest cultivars for which we had DNA available. We were able to complete the haplotype for an additional 4 microsatellites within, or in close proximity to the *FaPFRU* region. Our results showed that even with seven microsatellites the full PF haplotype in both SF and PF cultivars remains identical (Supplement 8), reinforcing the hypothesis of a common origin. The oldest PF line that was tested was the old European cultivar Liberation d'Orleans (1899, http://www.bordeaux.inra.fr/eustrawberrydb/accession/173).

All PF cultivars contained the full PF haplotype except for the previously identified recombinant lines Bolero and Pacific and the direct descendants thereof. We found that the partial PF haplotypes of Bolero and Pacific trace back to distinct recombination events (Supplement 8). The recombination events in Bolero and Pacific occurred somewhere between markers CDF2_PSSR_2 and FVES1729 and therefore makes the latter marker now the border of the *FaPFRU* region. This makes the maximum size of the *FaPFRU* region now 900kb, and the physical distance covered by markers that do not recombine with the phenotype just over 200kb (EZFv002 to CDF_PSSR_2).

The oldest SF cultivar for which we found a full PF haplotype was Dunlap (1890 [40]). We did not find evidence of a PF haplotype in some very old (semi) wild SF cultivars such as Scarlet, Little Scarlet, Vicomtesse Hericart de Thury and Eastern Pine. An effort to trace the source of both (full and partial) PF haplotypes in SF cultivars is shown in Figure 2. The pedigree information suggests that the full PF haplotype must have been present early in the breeding history of cultivated SF strawberry, as cultivars of relatively different origins such as Royal Sovereign, Dunlap and Senga Sengana all contained a full PF haplotype. The origin of the partial PF haplotype points back to Howard 17 and Missionary. Due to conflicts with our Howard 17 DNA and its offspring, we are not sure about the true to type-ness of our Howard 17 DNA, but we deduced that the original Howard 17 at least has a partial and possibly a full PF haplotype. The partial PF haplotype is also present in the Californian variety Wiltguard, for which we have no pedigree information tracing back to known cultivars, and Frau Mieze Schindler, which is an early 20th century German variety, from a distinct origin [40]. This suggests an early origin for the partial PF haplotype as well.

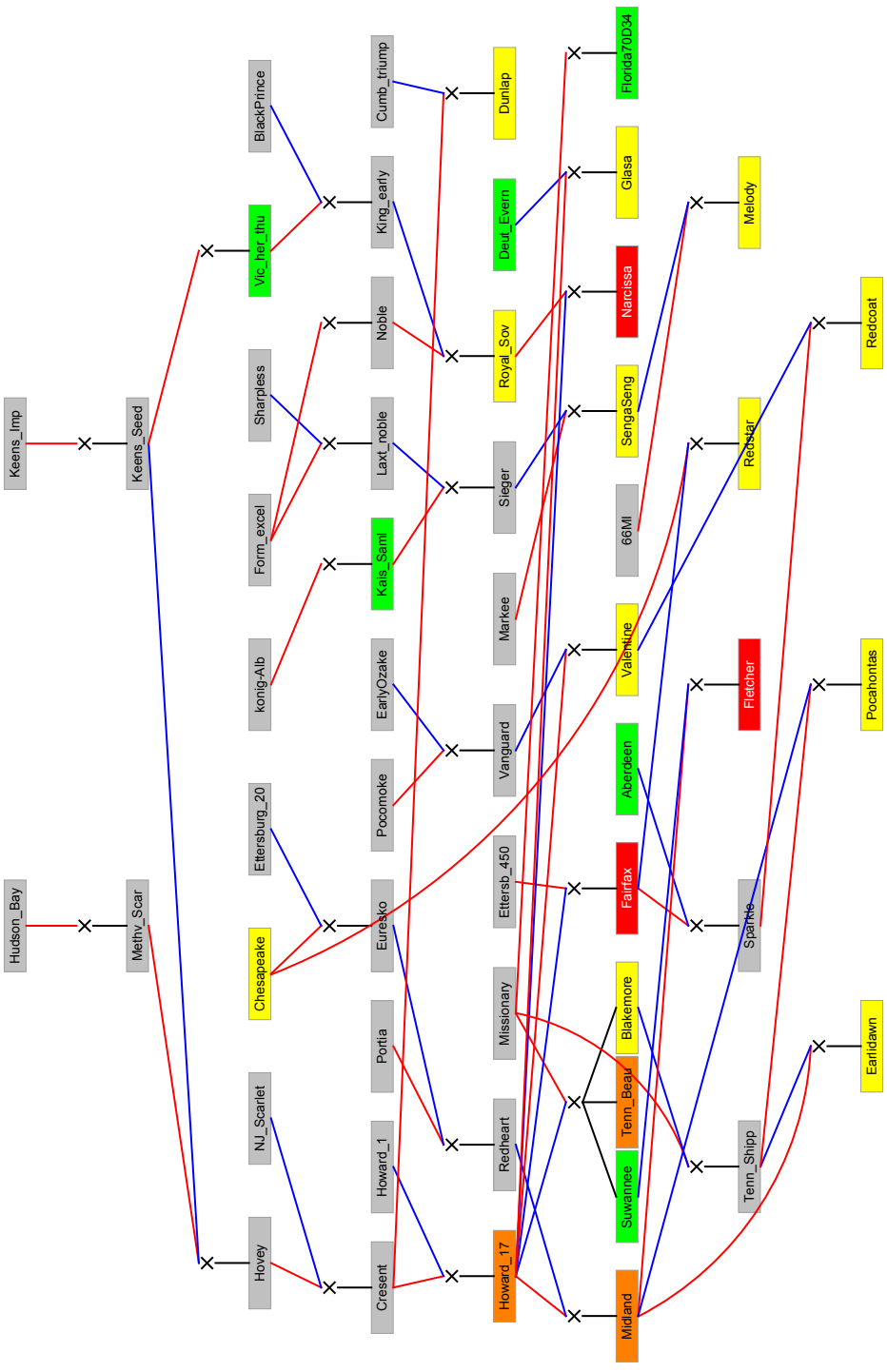


Figure 2. (Right page) Pedigree tree showing origin of PF haplotype in SF cultivars. In yellow are cultivars containing the full PF haplotype. In red are cultivars containing the Partial PF haplotype. In orange are cultivars containing both a full and a partial PF haplotype. Green cultivars only contain SF

Discussion

Fine-Mapping the PF trait

In this study we confirmed and further fine-mapped the *FaPFRU* major gene for perpetual flowering in strawberry which was previously identified and mapped by Gaston et al. [13]. Fine-mapping and haplotyping of cultivars delineated the region where the *FaPFRU* gene is located to an area of 900kb flanked by the newly developed SSR markers FVES0837 and FVES2160. For this *FaPFRU* region we developed seven SSR markers which showed two recombinations. However, these could not be used for fine-mapping because the recombinant individuals contained an intact PF-haplotype as well.

Flowering pathway genes in the FaPFRU region

We found two well-known flowering pathway genes (*FvCDF2* and *FvFT2*) within the *FaPFRU* genetic window on the *Fragaria vesca* reference genome v1.0 ((http://www.rosaceae.org/)). At first sight, the transcription factor cycling *DOF* (*CDF*) is an unlikely candidate as it is known to be a flowering repressor in Arabidopsis, where it represses the circadian oscillator *CONSTANS* (*CO*) and its downstream target *FT* [41]. Recently it was discovered that in potato *CDF* is able to promote tuberization indirectly through an *FT* homolog (*StSP6A*) by repressing a second antagonistic *FT* homolog *StSP5G* [19]. Antagonism between *FT* homologs has also been established for flowering in sugar beet [20]. The presence of multiple *FT* homologs in diploid strawberry from the clade orthologous to Arabidopsis *FT* [22], also fits with the possibility of antagonistic action between *FT* homologs in strawberry. *FT* homolog antagonism might explain the antagonism between *FaPFRU* presence and runnering observed by Gaston et al. [13] in cultivated strawberry. Further evidence suggesting a complex role for *FT* homologs in strawberry was given by Koskela et al. [23] where it was shown that PF plants under inductive LD conditions require *FvFT1* expression for flowering, but that SF plants under inductive SD conditions did not show any *FvFT1* expression. In fact SF plants only showed *FvFT1* expression under non-inductive LD conditions. A spatial expression experiment for both *FvFT1* and *FvFT2* from the same study revealed that in SF plants only *FvFT2* is highly expressed in flower buds, suggesting a role of this homolog in flowering as well. These findings from literature indicate that both *FaCDF2* and *FaFT2* could fit the mold of a dominantly inherited flowering promoter in cultivated strawberry. The possible mode of action of either gene is shown in Figure 3.

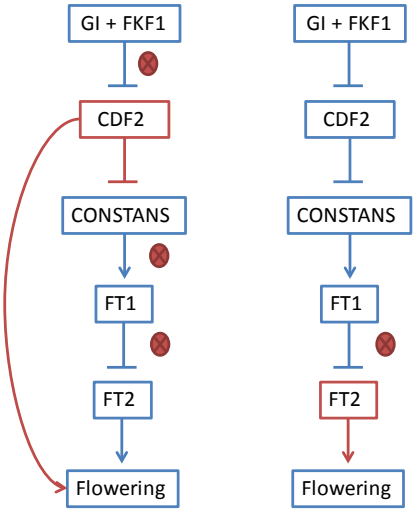


Figure 3. Possible mode of action of the two flowering pathway genes in the *FaPFRU* region. On the left the hypothetical mode of action of *CDF2*. Due to alteration of the *GI* binding site of *CDF2* the repression by the *GI+FKF1* complex no longer takes place (indicated by red cross sign). This subsequently disrupts the downstream components *CONSTANS* and *FT1*. As the repression of flowering signal *FT2* by *FT1* is lost, flowering can take place. The dashed arrow indicates the indirect activation of flowering by *CDF2*. On the right the hypothetical mode of action of *FT2*. An alteration in *FT2* disrupts its repression by *FT1*, which leads directly to flowering.

Sequence analysis of *FaCDF2*

Sequence analysis of the *FaCDF2* gene isolated from several unrelated SF and PF cultivars revealed that only two alleles were present, although we have to note that we did not sequence the full promoter region. The most significant differences between the two alleles were an amino acid substitution in the *GIGANTEA* (repressor) binding domain and large indel in the promoter sequence. We found that all PF varieties, but also several SF varieties carried a PF allele of *FaCDF2*, making it unlikely that *FaCDF2*, as a single gene, can provide PF flowering in all varieties. The other main candidate gene (*FaFT2*) has not yet been cloned and sequenced.

Haplotype analysis of *FaPFRU* region

Using a set of three microsatellite markers spanning the *FaPFRU* region, we found that 128 out of 129 PF lines contained a full (123) or partial (5) PF haplotype. The only exception, Ozark Beauty, had a *FaPFRU* haplotype that did not resemble that of any PF plant including those from the same donor (Pan American), and is therefore

possibly not true to type. These results showed that the *FaPFRU* region is a main switch that is always required for PF flowering. We identified eight different SSR haplotypes (from 3 SSR-loci) with a frequency of over 5% suggesting a moderate level of diversity. However we also found the complete and partial PF haplotype in 51 out of 139 SF cultivars. These include the SF cultivars for which the *FaCDF2* PF allele sequence was found. A second, more extensive, haplotyping did not reveal any differences between the full PF haplotype found in PF cultivars and that of SF cultivars, however it did reveal that the recombinant PF haplotype found in some PF plants must have arisen recently and independent of each other which further reduced the size of the candidate gene region for *FaPFRU*.

The origin of the PF trait

A major find of this study was the presence of the same PF haplotype in PF varieties that were previously believed to be derived from distinct PF sources. According to literature the first PF variety was introduced in 1866 in Europe and was known as Gloede's seedling or Ananas Perpetual [9, 31]. It is possible that this source is still present in the European PF varieties that were tested in this study such as Liberation d'Orleans, Ostara and Mara des Bois. At the end of the 19th century a chance PF seedling in New York was found in a field of Bismarck strawberries and named the 'Pan American' [9]. The Pan American source was used extensively in the US until the second half of the 20th century, when a *Fragaria viriginiana glauca* accession from the Wasatch mountains in Utah was used to introduce PF into a number of Californian varieties [9]. These *glauca* derived varieties were used extensively in breeding from the 1970s onward throughout the world and represent the vast majority of modern day PF cultivars [10]. For all these sources, the same haplotype is found in PF plants. This suggests that the initial source is likely to be the same or at least closely related, which, given the difference in space (Utah, New York and Europe) and time of introductions (mid to late 19th century and mid 20th century), makes it seemingly unlikely. The fact that an identical haplotype is found in several founder SF varieties from before the 20th century and a significant portion of their descendants complicates matters even further. It is clear that the PF haplotype in both SF and PF plants comes from the same ancestral source. In addition, given the high mutation rate of microsatellite markers [42], the diversity found at the *FaPFRU* locus and the fact that recombinations have occurred for the PF haplotype in recent breeding history, the differentiation -if there is any- between the PF haplotype in SF and PF plants must be relatively recent.

We propose that the PF conferring haplotype which is found in PF lines of

Fragaria x ananassa is solely derived from *Fragaria virginiana*and likely only the *glauca* subspecies, and, given that literature describes separate points of introduction (Europe, NY state and Utah [9]) for the PF trait, it was probably introduced through multiple events. This in line with the geographical distribution of this wild species, which ranges from the Western US along the Rocky mountains up to Alaska, and through Canada to the north-Eastern US [43]. The widespread use of the Utah *glauca* source in modern day breeding might be attributable to a quantitatively better day-length response than other *glauca* accessions as observed by Heide et al [44], or due to better fruit quality characteristics (firmness, uniformity, size) of the material in which it was introduced. The high frequency of the PF trait in all *Fragaria virginiana* spp *glauca* populations [45], points towards strong selection pressure for the PF trait in *glauca*, which allows for a rapid spread of a specific haplotype over a large area. This rapid spread is not surprising as perpetual flowering will give significant advantages in areas with a short growing season, which is mostly under LD. It is also not unthinkable that among the wild *Fragaria virginiana* plants brought to Europe [40, 46], a PF *Fragaria virginiana* (ssp *glauca*) was present as well, allowing for a separate European discovery of the PF trait.

We came up with two possible scenarios that could explain the occurrence of the PF haplotype in both SF and PF plants. In the first scenario, the qualitative switch to perpetual flowering is under monogenic control throughout the strawberry breeding germplasm and there is a functional differentiation between the PF haplotype found in SF plants and the PF haplotype found in PF plants. This differentiation must then have occurred relatively recently because both haplotypes are identical by state, despite the high mutation rate of microsatellite markers. In the second scenario more than one locus can qualitatively control the switch to perpetual flowering, and there is no functional difference between the PF haplotype found in SF and PF plants. The additional locus could be a suppressor that can counteract the effect of the *FaPFRU* locus, or an inducer that is required in addition to the *FaPFRU* locus.

The first scenario has more support from literature; many studies have observed monogenic inheritance for PF as a qualitative trait [9, 13, 29, 31], including the mapping populations that we used. There is a fly in the ointment with this scenario however. The difficulty lies in the question: where did the SF variant originate from? The mainly PF *glauca* and mainly SF *virginiana* subspecies of *Fragaria virginiana* are morphologically distinct, but genetically highly similar [47] and share overlapping habitats [43]. It is hard to imagine however, that a haplotype shared between two subspecies became functionally different but remained genetically (nearly) identical, and especially, did not recombine over many generations, especially when taking into account intercrossing. In this respect, the second (multi locus) scenario fits

better, as it allows for a single source of PF haplotype found in both SF and PF plants. However, the multi locus scenario also has its drawbacks, because a multi locus scenario would allow for SF plants to be crossed and yield fully PF plants. As far as we know this has not been observed [48] (Bert Meulenbroek, Fresh Forward breeder; personal communication), at least in North-Western Europe where climate allows for fairly accurate phenotyping of PF. With the data we have gathered so far, it is not possible to decide which scenario holds more merit. Fortunately, there are straightforward ways to find out which scenario fits best. In the short term, haplotyping several *Fragaria virginiana* ssp *virginiana* and ssp *glauca* accessions for the *FaPFRU* region could pinpoint whether the PF haplotype is unique to *glauca* or not. Extensive sequencing efforts for the entire *FaPFRU* region in SF and PF plants with a PF haplotype could reveal whether differences do occur. Additionally, by crossing PF plants with SF plants carrying a PF haplotype, and scoring them for flowering and *FaPFRU* marker segregation, we can confirm whether the PF haplotype in SF plants remains functional. Genome wide association analysis could reveal additional genomic locations that are required for PF cultivars and/or absent in SF cultivars. Finally, cloning and transformation of the main candidate genes *FaFT2* and *FaCDF2* (still a potential candidate gene under the multi locus scenario) into different SF backgrounds will also be able to provide answers on which scenario fits best.

Conclusion

It is clear that the *FaPFRU* locus acts as a dominant main switch for day-length dependent flowering in strawberry in our mapping populations. Flowering pathway genes *FaCDF2* and *FaFT2* are the most likely candidate genes for conferring the PF trait. Sequencing results on *FaCDF2* and haplotyping results on the *FaPFRU* region revealed that the PF allele is also present in SF varieties. This warrants further investigation to confirm whether a single or a multi locus scenario is applicable to the PF trait.

Acknowledgements

We would like to acknowledge TTI-Green genetics for the funding of this project. In addition we would like to thank the RosBREED consortium for allowing us early access to the *F. vesca* aligned resequencing data of several cultivars.

REFERENCES

1. Higgins JA, Bailey PC, Laurie DA: **Comparative Genomics of Flowering Time Pathways Using *Brachypodium distachyon* as a Model for the Temperate Grasses**. *Plos One* 2010, **5**(4).
2. Opstad N, Sonsteby A, Myrheim U, Heide OM: **Seasonal timing of floral initiation in strawberry: Effects of cultivar and geographic location**. *Sci Hortic-Amsterdam* 2011, **129**(1):127-134.
3. Sonsteby A, Heide OM: **Dormancy relations and flowering of the strawberry cultivars Korona and Elsanta as influenced by photoperiod and temperature**. *Sci Hortic-Amsterdam* 2006, **110**(1):57-67.
4. Sonsteby A, Heide OM: **Temperature responses, flowering and fruit yield of the June-bearing strawberry cultivars Florence, Frida and Korona**. *Sci Hortic-Amsterdam* 2008, **119**(1):49-54.
5. Verheul MJ, Sonsteby A, Grimstad SO: **Interactions of photoperiod, temperature, duration of short-day treatment and plant age on flowering of *Fragaria x ananassa Duch*. cv. Korona**. *Sci Hortic-Amsterdam* 2006, **107**(2):164-170.
6. Verheul MJ, Sonsteby A, Grimstad SO: **Influences of day and night temperatures on flowering of *Fragaria x ananassa Duch.*, cvs. Korona and Elsanta, at different photoperiods**. *Sci Hortic-Amsterdam* 2007, **112**(2):200-206.
7. Heide OM, Stavang JA, Sonsteby A: **Physiology and genetics of flowering in cultivated and wild strawberries - a review**. *Journal of Horticultural Science & Biotechnology* 2013, **88**(1):1-18.
8. Savini G, Giorgi V, Scarano E, Neri D: **Strawberry plant relationship through the stolon**. *Physiol Plant* 2008, **134**(3):421-429.
9. Ahmadi H, Bringhurst RS, Voth V: **Modes of Inheritance of Photoperiodism in *Fragaria***. *J Am Soc Hortic Sci* 1990, **115**(1):146-152.
10. Stewart PJ, Folta KM: **A Review of Photoperiodic Flowering Research in Strawberry (*Fragaria spp.*)**. *Crit Rev Plant Sci* 2010, **29**(1):1-13.
11. Bradford E, Hancock JF, Warner RM: **Interactions of Temperature and Photoperiod Determine Expression of Repeat Flowering in Strawberry**. *J Am Soc Hortic Sci* 2010, **135**(2):102-107.
12. Nicoll MF, Galletta GJ: **Variation in Growth and Flowering Habits of Junebearing and Everbearing Strawberries**. *J Am Soc Hortic Sci* 1987, **112**(5):872-880.
13. Gaston A, Perrotte J, Lerceteau-Kohler E, Rousseau-Gueutin M, Petit A, Hernould M, Rothan C, Denoyes B: ***PFRU*, a single dominant locus regulates the balance between sexual and asexual plant reproduction in cultivated strawberry**. *J Exp Bot* 2013, **64**(7):1837-1848.
14. Amasino R: **Seasonal and developmental timing of flowering**. *Plant J* 2010, **61**(6):1001-1013.
15. Zeevaart JAD: **Leaf-produced floral signals**. *Curr Opin Plant Biol* 2008, **11**(5):541-547.

16. Song YH, Ito S, Imaizumi T: **Similarities in the circadian clock and photoperiodism in plants**. *Curr Opin Plant Biol* 2010, **13**(5):594-603.
17. Turck F, Fornara F, Coupland G: **Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage**. *Annual Review of Plant Biology* 2008, **59**:573-594.
18. Tsuji H, Taoka K, Shimamoto K: **Florigen in rice: complex gene network for florigen transcription, florigen activation complex, and multiple functions**. *Curr Opin Plant Biol* 2013, **16**(2):228-235.
19. Kloosterman B, Abelenda JA, Gomez MDC, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S *et al*: **Naturally occurring allele diversity allows potato cultivation in northern latitudes**. *Nature* 2013, **495**(7440):246-250.
20. Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL, Nilsson O: **An Antagonistic Pair of *FT* Homologs Mediates the Control of Flowering Time in Sugar Beet**. *Science* 2010, **330**(6009):1397-1400.
21. Albani MC, Battey NH, Wilkinson MJ: **The development of ISSR-derived SCAR markers around the SEASONAL FLOWERING LOCUS (SFL) in *Fragaria vesca***. *Theor Appl Genet* 2004, **109**(3):571-579.
22. Iwata H, Gaston A, Remay A, Thouroude T, Jeauffre J, Kawamura K, Oyant LHS, Araki T, Denoyes B, Foucher F: **The *TFL1* homologue *KSN* is a regulator of continuous flowering in rose and strawberry**. *Plant J* 2012, **69**(1):116-125.
23. Koskela EA, Mouhu K, Albani MC, Kurokura T, Rantanen M, Sargent DJ, Battey NH, Coupland G, Elomaa P, Hytonen T: **Mutation in *TERMINAL FLOWER1* Reverses the Photoperiodic Requirement for Flowering in the Wild Strawberry *Fragaria vesca***. *Plant Physiol* 2012, **159**(3):1043-1054.
24. Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM *et al*: **An enhanced microsatellite map of diploid *Fragaria***. *Theor Appl Genet* 2006, **112**(7):1349-1359.
25. Stewart PJ, Winslow AR, Folta KM: **An initial characterization of fragaria *CONSTANS* mRNA levels in short-day and day-neutral cultivars.** . In: *North American Strawberry Symposium: 2007*; 2007: 100-104.
26. Powers L: **Inheritance of Period of Blooming in Progenies of Strawberries**. *P Am Soc Hortic Sci* 1954, **64**(Dec):293-298.
27. Serce S, Hancock JF: **Inheritance of day-neutrality in octoploid species of *Fragaria***. *J Am Soc Hortic Sci* 2005, **130**(4):580-584.
28. Shaw DV: **Heterogeneity of segregation ratios from selfed progenies demonstrate polygenic inheritance for day neutrality in strawberry (*Fragaria xananassa Duch.*)**. *J Am Soc Hortic Sci* 2003, **128**(4):504-507.
29. Sugimoto T, Tamaki K, Matsumoto J, Yamamoto Y, Shiwaku K, Watanabe K: **Detection of RAPD markers linked to the everbearing gene in Japanese cultivated strawberry**. *Plant Breed* 2005, **124**(5):498-501.
30. Weebadde CK, Wang D, Finn CE, Lewers KS, Luby JJ, Bushakra J, Sjulin TM, Hancock

JF: **Using a linkage mapping approach to identify QTL for day-neutrality in the octoploid strawberry**. *Plant Breed* 2008, **127**(1):94-101.

31. Richardson CW: **A preliminary note on the genetics of *Fragaria***. *J Genet* 1914, **3**(3):171-177.
32. Castro P, Bushakra JM, Stewart P, Weebadde CK, Wang D, Hancock JF, Finn CE, Luby JJ, Lewers KS: **Genetic mapping of day-neutrality in cultivated strawberry**. *Molecular Breeding* 2015, **35**(2).
33. van Dijk T, Noordijk Y, Dubos T, Bink MC, Meulenbroek BJ, Visser RG, van de Weg E: **Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids**. *BMC Plant Biol* 2012, **12**:25.
34. Isobe SN, Hirakawa H, Sato S, Maeda F, Ishikawa M, Mori T, Yamamoto Y, Shirasawa K, Kimura M, Fukami M *et al*: **Construction of an integrated high density simple sequence repeat linkage map in cultivated strawberry *(Fragaria x ananassa)* and its applicability**. *DNA Res* 2013, **20**(1):79-92.
35. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP *et al*: **The genome of woodland strawberry (*Fragaria vesca*)**. *Nat Genet* 2011, **43**(2):109-116.
36. Van Dijk T, Pagliarani G, Pikunova A, Noordijk Y, Yilmaz-Temel H, Meulenbroek BJ, Visser RGF, Van De Weg WE: **Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map**. *BMC Plant Biology* 2014, **14**(55).
37. Izawa T: **Daylength measurements by rice plants in photoperiodic short-day flowering**. *Int Rev Cytol* 2007, **256**:191-+.
38. Tsuji H, Taoka K, Shimamoto K: **Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation**. *Curr Opin Plant Biol* 2011, **14**(1):45-52.
39. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer**. 2011, **29**(1):24-26.
40. Darrow GM: **The strawberry; history, breeding, and physiology**, [1st edn. New York,: Holt; 1966.
41. Morris K, Thornber S, Codrai L, Richardson C, Craig A, Sadanandom A, Thomas B, Jackson S: ***DAY NEUTRAL FLOWERING* Represses *CONSTANS* to Prevent *Arabidopsis* Flowering Early in Short Days**. *Plant Cell* 2010, **22**(4):1118-1128.
42. Thuillet AC, Bru D, David J, Roumet P, Santomi S, Sourdille P, Bataillon T: **Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, Triticum turgidum (L.) Thell. ssp durum desf.***Mol Biol Evol* 2002, **19**(1):122-125.
43. Staudt G: **The species of Fragaria, their taxonomy and geographical distribution**. *Acta Hort* 1989, **265**:23-33.
44. Sonsteby A, Heide OM: **Flowering physiology of populations of Fragaria virginiana**. *Journal of Horticultural Science & Biotechnology* 2008, **83**(5):641-647.

45. Sakin M, Hancock JF, Luby JJ: **Identifying new sources of genes that determine cyclic flowering in Rocky Mountain populations of Fragaria virginiana ssp glauca Staudt**. *J Am Soc Hortic Sci* 1997, **122**(2):205-210.
46. Hancock JF: **Strawberries**: Cabi Pub.; 1999.
47. Harrison RE, Luby JJ, Furnier GR, Hancock JF: **Morphological and molecular variation among populations of octoploid Fragaria virginiana and F-chiloensis (Rosaceae) from North America**. *American Journal of Botany* 1997, **84**(5):612-620.
48. Simpson DW, Sharp DS: **The Inheritance of Fruit Yield and Stolon Production in Everbearing Strawberries**. *Euphytica* 1988, **38**(1):65-74.
49. Bassil NV, Davis TM, Zhang HL, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR *et al*: **Development and preliminary evaluation of a 90 K Axiom (R) SNP array for the allo-octoploid cultivated strawberry Fragaria x ananassa**. *BMC Genomics* 2015, **16**.

Supplementary Files

Chapter 5 Supplementary File 1, **Table of primer pairs used for mapping**
Chapter 5 Supplementary File 2, **Table of primer pairs used for walking the *FaCDF2* gene**
Chapter 5 Supplementary File 3, **Results of PF and SF lines tested with marker CDF_PSSR_2**
Chapter 5 Supplementary File 4, **DNA sequence alignment of *FaCDF2* PF and SF allele**
Chapter 5 Supplementary File 5, **Amino Acid alignment of FaCDF2 PF and SF allele**
Chapter 5 Supplementary File 6, **Amino Acid alignment of the strawberry CDFs with potato and Arabidopsis CDFs**
Chapter 5 Supplementary File 7, **FaPFRU small haplotyping on large varietal dataset**
Chapter 5 Supplementary File 8, **FaPFRU extensive haplotyping on varieties**

Are available under doi: 10.18174/387697 http://library.wur.nl/WebQuery/wurpubs/506541

CHAPTER 6

General discussion

Thijs van Dijk

This thesis can be broadly divided into two parts. In the first part, tools were developed to perform molecular genetic studies in strawberry, starting with the MADCE method for the use of SSR markers in allopolyploids (Chapter 2) and followed by the creation of the Holiday x Korona SSR genetic linkage map (Chapter 3).

For the second part, these tools were used to search for marker-trait relations that could be useful in a practical strawberry breeding programme. The first trait that was investigated was resistance against the soil pathogen *Phytophthora cactorum* (Chapter 4), an important and widespread strawberry pathogen. Lastly we investigated the perpetual flowering (PF) trait in strawberry (Chapter 5), a trait that can be very useful for extending the growing season of strawberry.

In this general discussion I will touch on the findings of the four experimental chapters by discussing them for their relevance, intermittently I will place the chapters in the broader context of science in general and plant breeding in particular.

Prologue: From H-index to 'molecular orphan crops'

Working in Academia is truly fascinating, not only do you get to discover things that no-one has seen before (or at least think that you do), but you also get to know more about the inner workings of academic research. One thing that sets Academia apart from commercial enterprises is that the value of its output is not expressed in monetary terms. Although this sounds like some kind of utopia, taken straight from the Star Trek universe, the reality is that this leads to a dilemma. The dilemma is that in business, turnover and profit are excellent objectives and quantifiable measures of worth. But how can we put an objective and quantifiable measure to the worth of a scientist's academic output?

Maybe, right now, you are thinking: Wait a minute, is it even possible, or better, desirable, to put a number to what science is worth? That doesn't sound right does it? But the truth is, that we already are measuring academic worth in a quantifiable manner, and that this quantification is becoming increasingly important. The impact factors of a scientist's publication list and his personal H-index are now commonly used in applications for personal grants and for moving to better positions. There are definite advantages to the use of these measures; a good and productive track record shows that an individual is able to prioritize and finish research lines, whereas an unimpressive one could indicate the opposite. In addition, human beings can make decisions much easier when a certain attribute is quantified, for example : Imagine you have to choose between two chickens, one: large, brown feathered, brown eggshell, clucks all day and a decent egg-layer, the other: medium sized,

black feathered, white eggshell, clucks a little and a decent egg layer. Choosing between the two is difficult, however, if I would have changed the description decent egg layer into: average 3.8 eggs/week for the brown hen and 4.0 eggs/week for the black hen, the choice not only becomes easier, but you also tend to forget the other non-quantitative attributes.

Another consequence of the increased importance of these quantifiable measures is that it leads to increased competition between scientists and scientific groups. This has the advantage that it increases the speed of- and total scientific output as a whole, and also the drive for quality, as high impact factor journals often have more thorough quality requirements. However, it also acts as a dual-edged sword, because when one's worth becomes dependent on such a measurement, it is human nature to find ways to improve it, even if these ways defeat the original purpose of the measurement, or even, are downright harmful. In science, this process is described as "impact factor mania" [1].

A striking example of behavior that defeats the purpose of a measurement is the recent scandal ("diesel dupe" or in Dutch "sjoemelsoftware") uncovered at Volkswagen, where the quantifiable variable "emission level" was purposefully defeated by cheating the measurement system.

Now, in science, purposeful cheating by tempering with data is fortunately rare, but it does occur and might even be on the increase as witnessed by websites such as retraction watch (http://retractionwatch.com/). The peer review process can be compromised by "gentleman's agreements" as well as purposeful obstruction as scientists' often narrow fields of research make anonymity virtually impossible. In addition, the impact factor and H-index are subject to inflation, which is driven by the growth of the scientific community as a whole, but less positively due to ever-increasing co-author lists [2].

Less obvious, but perhaps more damaging, are the possible long term consequences of over-reliance on publication and citation statistics. When both career perspectives as well as funding are better for scientific topics that are considered more important or are 'hot' and 'trending', this will lead to more scientists flocking towards such topics. This leads to more competition and activity around such topics, increasing time pressure. In turn, this could provide a feeding ground for less ethical behavior. On the other hand, less attractive topics –because of lower importance, inherent difficulty or risk, or because the possible outcomes of pursuing such topics are less likely to result in high impact science- will gradually become more desolate, and move to areas where less resources and scientific talent are available.

Even though this process is inherent to science, the rate at which it happens is increased when we attach too much importance to impact factors and citation indexes.

In the field of plant breeding and genetics, there are some striking examples of crops which have gotten little attention despite their relative importance in agriculture. One such crop is onion (*Allium cepa*), which is grown worldwide including industrialized countries, and is one of the most consumed crops in the world. Yet, despite its importance, molecular genetic resources in onion are scarce [3-5]. Few QTLs have been found so far and a reference genome is still under construction [6]. This makes onion a typical example of a crop where much scientific gains can be made, but these gains can only be made slowly because of a highly complex and large genome. This in turn means most discoveries will be relevant to the *Allium* community alone, as for most traits analogies will have been found previously in other plant species with relatively 'easy' genomes. For such highly complex genomes, the relatively slow progress and limited opportunities for 'new' discoveries are a double handicap in terms of reaching high impact factor journals, which could have a progressive slowing effect on the speed of scientific gains.

There are several economically important, but genetically highly complex plant species (onion, sugarcane, mango, ornamentals such as lily and tulip) that would fit this category of "molecular orphan crops", despite being anything but a true orphan crop. On the other hand, at least these species should be able to attract some interest for reasons of economic importance, and their scientific relevance for understanding the complexity of their genomes. Another category of crop species is even worse off because they miss the boat on a number of factors: e.g. being of moderate to low importance, or difficult to work with due to fairly complex genomes, long generation times, difficulties in propagation, lack of public appeal, or a combination of these factors.

One of the most striking examples of such a crop is spinach (*Spinacia oleracea*) : important but not a major food crop, a large but not exceptionally large or complex genome, difficult to work with practically due to its dioecious flowering and wind pollination, and perhaps still suffering from the "Eat your spinach" childhood experiences with regards to public appeal. When compared to other crops of similar importance, the genomic resources available to spinach are lacking [7, 8]. Recently genome sequencing has begun for this crop, whose cultivation is under threat from the extremely adaptable pathogen *Peronospora farinosa* (Downy Mildew) [7], but perhaps should have come much sooner. In this context (notwithstanding the necessity and importance of basic research on model plants) the nearly 800 documents found in scopus (www.scopus.com) over the past 5 years when using the words *Arabidopsis* and *Pseudomonas* compared to 10 when using the words spinach and *Peronospora*, offer a sobering view on where the priorities of the scientific community lie. Such figures show that attaining critical mass for research can be a

real problem for ‘molecular orphan crops’.

Strawberry

Now, finally, let’s get to the subject of this thesis: The garden strawberry (*Fragaria x ananassa*). When we consider strawberry in terms of attractiveness as a research target, it could be best characterized as a semi-orphan crop. It has moderate to high economic importance, it has an excellent public appeal due to its attractiveness and taste, but also has several practical and technical challenges for genetic studies such as being allo-octoploid, outbreeding and vegetatively propagated. An advantage is that one of the diploid ancestors (*Fragaria vesca*) of garden strawberry has features that make it an excellent model plant for many fields of study in plant science [9]. Too name a few : it is perennial, it has naturally occurringShort Day and Long Day flowering, it has a wide geographic distribution, it has both vegetative and seed propagation, it is outbreeding yet capable of selfing and becoming fully inbred, it has only 7 chromosomes and a small (~200mb) genome, it is part of a family with many economically important and diverse crops (Rosacaea),it forms accessory fruits, which are consumed, as well as true fruits (the achenes/seeds) and finally, it is fairly small and its generation time is 3-4 months. In short, there aren’t many plant species that can beat it for variation in survival strategies and versatility in plant physiology.

Coming back to the cultivated octoploid strawberry, with regards to genomic resources and scientific interest, it has partnered up with its diploid wild relative, *F. vesca* allowing for faster pioneering work such as the creation of the first linkage maps and the assembly of the FvH4 reference genome sequence [10-12] and the cultivated octoploid in turn giving credence for such research due to its economic importance. Right now, the genomic tools that are available to strawberry geneticists such as the diploid reference genome, a multitude of linkage maps and a 90k SNP array allow for the rapid discovery of trait loci that are interesting for breeders [12-20]. Now, this doesn’t mean that we are anywhere near the development level of more easy to use, economically important crops or model plant species, but it is clear that large strides have been made since the start of this PhD thesis in 2009.

Microsatellite Allele Dose and Configuration Establishment (MADCE)

The purpose of the first chapter was to develop a method to interpret the complex peak patterns generated by microsatellites in octoploid strawberry in such a way

that we ended up with as much information as one would expect to retrieve from a microsatellite in a diploid system. With this information we would then be able to generate high quality linkage maps for the different sub-genomes which could be easily aligned and compared. In the MADCE methodology, we first need to determine the dose of each allele present in an individual. The MADCE method makes use of all the fluorescent signal peaks in an SSR electropherogram as a reference for determining allele dose, the consequence of this is that higher ploidy levels actually improve the dosing of alleles. The next step is to establish the allelic configuration of each individual homoeologue (subgenome) within that individual.

Because of the disomic inheritance in strawberry, the repulsion of alleles from the same subgenome in offspring was the key to obtaining this information. We found that in single cross mapping populations, the deployment of our method was fairly easy due to the high number of offspring that can be used to establish repulsion between alleles. However, for pedigreed breeding germplasm this was another matter, as generally only a few offspring were available. Additional tricks had to be used to establish the allelic configuration, and, even with these tricks, some level of uncertainty or incompleteness would remain. Performing MADCE on breeding germplasm can best be compared to solving Sudokus, except that someone forgot to mention how many distinct numbers there are, and that sometimes, you might have to use the same number twice.

Usefulness of MADCE in the "next-gen" era.

Although MADCE is a tough method to master and could be considered the antithesis of "high throughput", it does offer an unprecedented insight into allopolyploid genomes, especially when several linked markers are considered. Regions of homozygosity, deletions and very distinctive haplotypes can be extracted, information which, ironically, is often not available to this extent in mapping studies of diploid crops. The MADCE method was developed specifically for microsatellite markers, which, in times where the lingo of geneticists is dominated by abbreviations such as "SNPs", "NGS", "GBS" and "GS", seems like preferring a zeppelin over a turbojet. Which is not a bad analogy, since a zeppelin might be slow, but does offer a good view of the landscape below. Similarly there are also advantages to the use of microsatellites. When it comes to tagging haplotypes, the multi-allelism and high mutability of microsatellites offer a distinct advantage over SNPs in that they can tag evolutionary events that happen in much shorter timeframes than SNPs. This makes microsatellites still very useful for haplotyping. However, looking into the "Next Gen" future of genetic studies, it does appear that the shelf-life of MADCE is

limited. Some new technologies such as genotyping by sequencing (GBS) offer both the numerical advantage of SNPs and the allele discrimination of microsatellites in one neat package, at least in theory.

There is one other aspect of MADCE however, that could give it a more lasting use. Over the years, we have had several students and guest-workers who worked on strawberry. Not everyone that tried to learn and apply MADCE succeeded, but generally those with a knack for it would do so very quickly. This is because MADCE requires both a proficiency with solving puzzles and a good understanding of allopolyploid genetics. In this respect MADCE could serve as a teaching tool for people starting to work on allopolyploids. This is because almost all the pitfalls that can be encountered in applying MADCE, are also encountered with other genotyping platforms in allopolyploids, with the exception that in MADCE these pitfalls become more obvious. In conclusion, we can state that for genome wide application, MADCE is not a viable method due to practical constraints, however MADCE can still be relevant as an addition to high throughput genotyping platforms or perhaps as a teaching tool for understanding allopolyploid genetics.

The Holiday x Korona genetic map

With the Holiday x Korona (HxK) genetic linkage map[19], we aimed to develop a reference genetic map for the strawberry community as a whole and for trait specific studies in this thesis. We used the MADCE methodology to attain maximum information for each microsatellite that was used. Since the use of MADCE in mapping populations is fairly straightforward, we managed to successfully obtain the full allelic configuration across all four subgenomes for most of the 186 primer pairs used in this study.

With 82 progeny, 186 primer pairs and a total of 508 segregating loci amounting to 28 integrated linkage groups spanning a total of 2050cM, the HxK map is not standing out in a numerical sense amongst earlier and contemporary genetic maps of the allo-octoploid strawberry. However, the quality of the HxK map is best conveyed by the features that we managed to uncover which had been missed by previous maps of strawberry. Firstly, the HxK map uncovered a large inversion on one of the subgenomes of LG2 (2D), which spanned almost 30cM. There are two things intriguing about this inversion: the reason it wasn't found in previous maps, and its implications for the origins of cultivated strawberry. The reason it wasn't found in previous maps with better, or equal, densities of microsatellites is simple: due to the use of MADCE we could find more segregating loci per primer pair and thereby better compare the synteny and allelic composition between subgenomes.

Intriguingly, if a high density SNP array had been used, we would have had more difficulty spotting the inversion. This is because only very few SNPs (if at all) segregate on more than one subgenome, and therefore syntenic relations would have to be inferred indirectly by comparing the physical positions of the mapped SNP markers between subgenomes, which can be further complicated by the quality of the physical reference map.

The presence of this inversion has consequences for theories on the diploid origins of the octoploid strawberry. Recent studies on this subject [21, 22], support that octoploid strawberry is derived from two distinct diploid genomes, which, probably through tetraploid intermediaries, ultimately became octoploid strawberry. The inversion on LG2D must have happened in one of these stages. The most logical would be the diploid stage, because the divergence between the diploid ancestors represents the largest evolutionary timescale. However, the presence of this particular inversion on just one of the four diploid sub-genomes suggests that it must have happened at a later stage (tetraploid or octoploid), unless the bi-diploid origin is wrong. Thus, by tracing the origin of this inversion, it is possible to obtain a better understanding on the evolutionary origin of octoploid strawberry [23].

A second feature that distinguishes the HxK map from previously published maps also involves the diploid origins of octoploid strawberry. By comparing amplification efficiencies of primer pairs for each sub-genome and relating this to the origin of the primer sequences, we were able to sort the sub-genomes based on their similarity to diploid ancestor *Fragaria vesca*. Although this was a very rudimentary method because of the limited number of markers of *F. vesca* origin, it was the first time that a parameter related to evolution was used for distinguishing the sub-genomes. We can expect that with re-sequencing data coming available and with the introduction of the strawberry SNP array [24], a definite naming convention for the sub-genomes for octoploid cultivars will be established in the near future.

Recently, a study by Tenessen et al. [23] uncovered that three distinct diploid genomes (of which two were closely related) contributed to the octoploid strawberry genome through two tetraploid intermediaries. The technique that they developed (POLiMAPS) enabled them to assign a diploid origin to each of the sub-genomes in a much more precise way than we did. Many of our findings from the HxK map such as the inversion on LG2D, the inconsistencies between genetic maps and the FvH4 physical reference map, are supported by their study.

Another distinguishing feature of the HxK map is that it shows the complete haplotype of each parental homolog, including homozygous regions which by definition would lack much information in traditional approaches. Although this haplotype map cannot be relied on for 100% due to the simultaneous occurrence of

homozygosity at multiple homoeologous regions, as well as the marker resolution, it does offer some interesting insights in the genetic makeup of the parental cultivars. The estimated homozygosity level of cultivar Holiday, was similar to that expected of its pedigree (33% and 29%). Whereas Korona, which showed little inbreeding in its pedigree (3.6%), had a much higher level of homozygousity (13%) estimated from its haplotype. Similarly, the level of shared haplotypes between the two cultivars, indicating identity by descent (IBD), was higher than expected.

The discrepancy between expected and estimated levels of homozygosity and IBD within and between the parents could be a sign of selection pressure enacted by breeders. An alternate theory could be a genetic bottleneck early on in the domestication of strawberry, leading to higher levels of IBD among all strawberry varieties. This genetic bottleneck did occur early on in breeding [25], but I think it is unlikely to explain a large portion of the discrepancy that we see between expected and estimated levels of homozygosity and even IBD. This is because wild octoploids are likely to have been crossed into the cultivated strawberry several times when breeding began in the 1800s [26] and more recently as well [25, 27]. In addition, in house marker data on a diverse set of cultivars suggests that diversity can be quite high for individual loci (Van Dijk, data not shown), and it also shows that diversity is higher in older cultivars than in modern cultivars. This indicates that the genetic bottleneck is not as high one would expect and that the influence of selection pressure enacted by breeders is more likely to explain the afore mentioned discrepancies.

In conclusion, the HxK map has revealed several new insights with regards to the origin and breeding history of strawberry. The impact of the HxK map in the strawberry research community has been significant. It has aided in the development of the now widely used 90k Axiom SNP array [13]. Even more recently, the HxK map provided the basis for the most dense strawberry genetic linkage map (14k SNP markers) (Koehorst et al. unpublished results). This highly dense linkage map is now being used in several ongoing projects such as: The USDA-ARS RosBREED project (www.rosbreed.org) for the discovery of QTL for resistance to *Xanthomonas fragariae*, *Phytophthora cactorum* and *Colletotrichum gloeosporioides* [28], as well as an improved version of the FvH4 diploid strawberry reference genome sequence (Koehorst et al. unpublished results), and the generation of an octoploid reference genome sequence (http://gtr.rcuk.ac.uk/projects?ref=BB/N006682/1).

Interlude

"The Next Big Thing"

In 2005, I had my first practical experience with molecular markers in plants, using restriction enzymes to create CAPS markers, one by one. At that time I had not yet heard of Next-Generation sequencing (NGS), although the first machine (GS-20, from 454/Roche) had already been released. After joining Monsanto in 2007, it became clear that NGS was a "Next Big Thing". At that time there were discussions about which platform to choose for sequencing projects, most aiming at development of SNP arrays for use in fast genotyping of mapping populations. Anyone who has ever made a genetic map using AFLP markers that are mostly manually scored will appreciate the ease of use, the reliability and speed of a SNP array.

The NGS revolution has proven to be the most important factor in making genome wide marker studies accessible for almost any crop species. It is easy to find scientific papers highlighting the successes of NGS, but studies where the use of NGS has failed to live up to its reputation are equally intriguing, although it is impossible to find publications where this is made evident. I am sure that every department that had ventures into NGS, especially in the early days, has had lackluster results in some projects, either due to experimental design, choosing the wrong technology, not having the infrastructure to handle the data, etc.

I believe that one of the reasons why NGS failed in certain projects has to do with the flocking behavior that is sometimes instilled upon scientists when faced with a novelty that is "hot" and increases the chance of scoring in high impact factor journals, or even more detrimentally, ticking the novelty boxes of a reviewing committee for grant proposals. Another reason might be inflated expectations due to excessive hyping of potential benefits without the nuance of mentioning potential hurdles.

For instance, earlier on, I hyped MADCE as offering an unprecedented insight into allopolyploid genomes, which sounds good, but should also be followed by the nuance: 'in the same manner as a 13th century monk copyist would have an unprecedented insight into the textual makeup of the bible.' Either way, many unsuspecting scientists started NGS projects, only to realize later the enormity of the data and the necessity of a good bioinformatics pipeline. In other cases a mismatch between NGS platform or plant material and the objective of the study could render a projects' output sub-optimal. What it all boils down to is actually quite simple: when you want to solve a problem, pick the right tool for the trade. You can break down a wall in seconds with explosives, but if you need to rebuild it afterwards, you might

opt for something requiring less glue.

So, coming back to the question, what will be the "Next Big Thing" for genome wide genetic studies in plants? Well that depends on the situation. For instance in QTL studies with full-sib mapping populations, the answer is quite simple, there is no need for a "Next Big Thing", *until* something can beat the ease of use, the resolution and the accuracy of a high density SNP platform. However for finding marker-trait associations over a wide germplasm there is definitely a need for a "Next Big Thing".

Genome Wide Association Studies (GWAS) is a method for finding marker-trait relations over a wide germplasm. If you make yourself familiar with the Hype Cycle of technology [29] (Figure 1) and GWAS, it is clear that the latter is an excellent illustration of the former. Several years ago GWAS seemed to become the "Next Big Thing" in establishing marker trait relations. The solution to all problems that are encountered in going from a single-cross based QTL to a usable (set of) marker(s) for breeding in a wide germplasm [30]. I guess that this was the start of inflating expectations, and that now the scientific community is starting to realize that there are many limitations to GWAS in plants [31], but also to refine the methodologies in order to get more useful results. It could be said that with regards to GWAS we have now entered the slope of enlightenment.

There are three pillars that define the success in finding reliable marker-trait relations using GWAS: phenotyping, population and genotyping. In a GWAS study, the weakest of these three pillars, determines the quality of the outcome. No amount of fiddling with complex biometrical models will rescue serious deficiencies in these three pillars. I will not go into great detail on how to optimize the population and phenotyping aspects, as the complexities of these pillars individually, are large enough to fill a book. In short, the population should be sufficient in size and genetic variation. Phenotypes of interest should be distributed over the population in sufficient frequency and variation, and extensive exchange of genetic material between structure groups of the population is better than, or at least complementary, to statistical correction for breaking down structure bias. Phenotyping should minimize the environmental influence by both replication and controlled growing conditions and the phenotype should be broken down into logical biological components in order to maximize the chance of finding causative loci.

As for genotyping, the requirements are quite simple, and by listing these requirements we immediately have a reference of what the "Next Big Thing" for genome wide genetic studies in plants will be. Ideally, the best genetic predictors are those that can discriminate all polymorphisms within the population under study in a consistent way, but since that is still very hard, if not impossible, we would settle for

genetic predictors that can discriminate all underlying haplotypes at a given genetic region in a population. Now, in theory, this is possible with genotyping by sequencing (GBS), but we should also consider two other important attributes to a genotyping system that influence their adoption. One is reliability and the other is practicality, which are the main features of a SNP platform. In reality GBS is still in its infancy (somewhere on the slope of inflated expectations) and it is not yet at the same level of reliability and practicality as a SNP platform. The practicality is presently low due to the enormous amount of data, which needs complex bioinformatics tools to be translated into discrete predictors for each genotype. The reliability is low because there are too many factors that can cause errors in assigning the right genotype, especially in crops with more complicated genetics such as outbreeders and polyploids, or worse, outbreeding polyploids such as potato and strawberry. In addition, allele dosing is much less reliable when using GBS, as it has to rely on read numbers.

So, finally, we might have to settle for the "Next Best Thing". This could be a version of GBS that is optimized for reliability and practicality, e.g. using high read-depth and long read length to extract haplotype based predictors. Another option is to improve the haplotype tagging ability of a SNP platform by making them ultra-high density and based on a highly diverse discovery set. This should result in reliable and practical predictors with improved chances of finding haplotype tagging SNPs and/or inferring discriminatory haplotypes computationally. The take home message is that the choice of your technology does not depend on how a technology compares to the other in terms of possibilities or quantity, but rather in terms of available (human) resources, data quality, your research question and the technical and genomic infrastructure and complexity of your plant species.

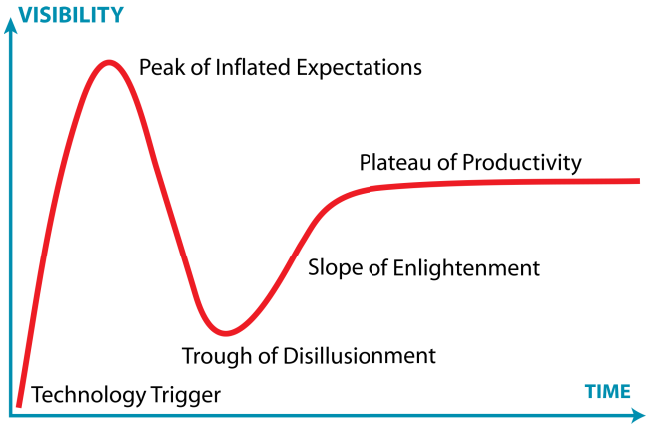


Fig. 1. The Gartner Hype Cycle (taken from https://commons.wikimedia.org/wiki/File:Gartner_Hype_Cycle.svg)

QTL mapping of *Phytophthora cactorum* disease resistance.

The objective of this chapter was to find, locate and characterize resistance QTLs against the soil-borne root pathogen *Phytophthora cactorum*, also known as crown rot. This generalist pathogen is present in soils worldwide and can become highly problematic under inductive conditions, especially when very susceptible strawberry cultivars are used. Resistance levels in strawberry varieties vary greatly, and it is generally believed that resistance is inherited quantitatively [32-35]. Two mapping populations were at our disposal, the Holiday x Korona population (HxK) that was used in the creation of our octoploid linkage map, and E1998-142 x Elsanta (ExEls), which was more specifically developed for mapping *P. cactorum* resistance. Both these populations were made by crossing two moderately resistant parents. The arguments we used for resorting to moderately resistant parents was that moderate resistance levels are manageable in practice, and additionally, previous experiences with the resistant cultivar (Senga Sengana) whose progenies showed too little variation (almost all offspring showed high proportions of resistance).

The HxK and ExEls populations were phenotyped over three years (2008, 2010 and 2011) under different seasons and conditions. The correlation between years was quite low for both populations (ranging from 0.18 to 0.47), indicating a large environmental effect on disease response. There were years with very high (2008, 2010) and very low (2011) disease symptoms, but we did not observe better correlations between the two years with similar levels. Results from the QTL analysis showed that most QTLs were small in effect and only just above the statistical significance threshold. This was particularly true for the HxK population, where most QTLs disappeared in cofactor analysis. For ExEls we uncovered two QTLs with relatively high significance levels. The first QTL was found on the top of LG7C, but only in the trial of 2010. The second QTL was found in both 2008 and 2010 on the bottom of LG7D, but was only highly significant in the 2010 trial.

Because of the high environmental influence, and the desire to have QTLs that are robust over environments, we used the average of all three years (AOTY) as an additional phenotype. The use of this variable limits the effect of (yearly) environmental variability and should therefore give a better indication which QTLs are reliable across environments. Such a QTL would prove more valuable for practical breeding purposes where reliability is required. Even though it might be scientifically interesting to explore genotype by environment interactions, for practical breeding purposes, a GxE QTL generally translates to performing Russian QTL roulette. This is not a good selling argument for a variety. When we used the

AOTY trait, the QTL on LG7D became stronger than for any of the individual years. Whereas for the LG7C QTL the significance dropped to just below threshold levels. These results indicated that removing environmental noise through averaging over experiments is a good way to uncover more reliable and therefore probably more valuable QTLs for a breeding program.

"Grey QTLs"

A recurring problem in QTL analysis is finding out which QTLs are "real" and which aren't. Sometimes an analysis is very straightforward, and yields strong major effect QTLs, in which case you might wonder why you went to the trouble of creating a linkage map. More often you end up with several QTLs, of which part are in a sort of grey area where you aren't sure whether they are useful or rubbish. There are a variety of methods described in literature for detecting reliable QTLs in mapping populations. Several of these methods, such as permutation testing, deal with finding optimum detection thresholds [36]. I must admit that I personally never use permutation testing, because once you get to QTLs that are around the detection threshold, it does not matter whether they are just above or below the threshold, but other factors, such as sharpness of the drop-off and the stability of the QTL in e.g. cofactor analysis become much more important. I therefore consider permutation testing as a typical case of over-engineering, with the detrimental effect of instilling a false sense of security. A good approach for finding reliable minor QTLs has to compensate for factors that cause false associations. These factors are legion, and every QTL study suffers from at least a few. Such factors include: the influence of extreme phenotypes (for which it usually is not clear which have a genetic basis and which an environmental), the influence of local environment (compensated by replication, randomization and blocking), phenotyping errors (sometimes caused by that same randomization), genotyping errors (e.g. switching individuals) and "lucky" correlation of markers with major effect genes (necessitating cofactor analysis). Some of these factors can be traced and removed from the dataset, but only in the more obvious cases.

A 'normal' QTL analysis performed on all individuals of a population always includes the aforementioned factors that potentially disturb the analysis. Statistically, there isn't much you can do to get a 'real' improvement if these factors cannot be dealt with in some way. One such way is through creating random subsets of individuals (subsample aggregation) [37, 38]. With this method some of the individuals that have a disturbing effect because of external (non-genetic) influences, have a chance of being excluded from the analysis. When this disturbing factor is removed, true QTLs should perform relatively better and spurious QTLs should perform relatively

worse. When a rather large subset is removed (e.g. 25-50%), the chance of having two or more disturbing factors removed in one analysis, becomes larger as well. After numerous cycles of intermittent in- and exclusion of disturbing factors, the 'true' QTL loci (even with minor effect) should have a slightly higher mean LOD aggregated over all analyses, as well as a more stable LOD compared to the spurious QTL loci. The stability of the LOD score per locus is a good indication on how much a limited number of genotypes affect whether a QTL becomes significant or not, and therefore its reliability. In theory, resampling methods appears to be well suited for getting the most out of a QTL analysis, however it is not yet complete, when it is only based on regressions for individual loci.

The use of multiple QTL models in QTL analysis is extremely important, especially because for many traits epistatic interactions are present, or, major QTLs (>30%EV) are found that overshadow the effect of true minor QTLs and can also cause false positives through spurious correlation to independent loci. The ultimate QTL detection method would therefore combine resampling with multiple QTL models. Fortunately, there are numerous model selection methods that make use of multiple QTL models, including Bayesian approaches [39]. Some multiple QTL models also make use of resampling methods [40, 41]. These methods have often been developed for complex situations such as for GWAS and Genomic Selection studies. The technicalities of these methods are beyond my mathematical comprehension, and I will therefore not discuss them. However, I do wish to point out, that none of them are unbiased, as parameter inclusion into a model is dependent on certain criteria. Unfortunately, it is computationally impossible to remove these criteria and explore all possible combinations of predictors in large datasets (not to mention to include resampling). For this reason, it can still be hard to find epistatic interactions, especially when effect size of individual QTLs is not large. Despite the impossibility to consider all combinations of predictors in an unbiased fashion, it is still possible to do this for just digenic interactions. Granted, the number of computations increases dramatically, but not to such an extent that it becomes impossible. Especially if care is taken in limiting the number of marker loci. By tracking the LOD for each combination of two loci over a number of random datasets it becomes possible to track down not only reliable QTLs, but also reliable interactions. In conclusion, I believe that a good way to make 'Grey QTL', more black and white is through using a detection method that combines resampling with (limited) multiple regression (e.g. totality of possible digenic interactions). This should in theory provide less biased QTL results. However, it is only necessary to perform such a computationally intensive analysis when no clear results are obtained with simpler methods and when the quality of the data is not in question.

Mapping the Perpetual Flowering trait

Two distinct types of flowering habit are used by strawberry growers world-wide. The most widely used flowering habit is seasonal flowering (SF), of which the varieties are commonly referred to as June-bearing. In temperate climates, these varieties initiate flowering under short day conditions in autumn and subsequently produce fruits in June. The other type is perpetual flowering (PF), of which the varieties are commonly referred to as day-neutrals or ever bearers. These varieties initiate flowering under long days, and can therefore produce fruit for a much longer period: throughout the summer and early fall. Apart from day-length dependence, flower initiation is also influenced by temperature [42, 43]. It is generally believed that in the cultivated strawberry PF is under dominant inheritance.

In this study we set out to map the main factor that causes PF in two mapping populations for which individuals had been selected and differentiated based on their flowering phenotype qualitatively. We quickly found a marker which co-segregated completely with the PF phenotype in both populations. This marker was located on the bottom of LG4D. At the moment of mapping, a paper was published by Gaston et al. [44] which mapped the same trait to the same location, albeit at a much less defined interval. We found that there were two very clear candidate genes within our QTL interval, *FaCDF2* and *FaFT2*, which were homologous to genes that are major factors in the flowering pathway of Arabidopsis and many other plant species [45, 46].

We subsequently performed further analyses, by sequencing one of the candidate genes (*FaCDF2*), and by investigating the haplotype diversity at the QTL interval. The sequencing of the *FaCDF2* gene indicated that only two allelic variants were present in the eight highly distinct varieties that were sequenced. The variant which was present in all PF varieties showed several polymorphisms, of which two SNPs that altered the amino acid sequence. However, this PF variant of *FaCDF2* was also present in several SF varieties. This indicated that either *FaCDF2* is not the causal gene, or that one or more other loci can have a qualitative effect on the switch from SF to PF. Our follow up haplotyping study was done on a large number of PF and SF varieties of varying origin, including three reputedly different and independent sources of PF [47].

The haplotyping revealed that all PF varieties of all origins carry the same haplotype in the PF QTL region, and that there weren't any recombinations between the candidate genes *FaCDF2* and *FaFT2*, which are 250kb apart on the physical genome. This makes it still undecided which of these two candidate genes are causal to the PF trait. Another interesting result from the haplotyping was that the complete PF haplotype

was present with moderate frequency in SF varieties as well. Not only does this suggest a common origin, it also complicates the establishment of a theory for the mechanisms behind perpetual flowering in cultivated strawberry. So far we have not been able to establish whether the PF haplotype in SF lines is functionally distinct from the PF haplotype in PF lines. All we know is that it does not confer perpetual flowering in these SF lines. Because the PF region was haplotyped quite extensively, and several distinct haplotypes have been uncovered, it is certain that the PF haplotype that is present in both PF and SF cultivars have a common origin. However, to prove whether there is functional differentiation, and whether multiple qualitative loci are involved in perpetual flowering, we should perform more experiments. A possible follow-up experiment is to perform a testcross between PF and SF lines which both contain a PF haplotype in heterozygous state and phenotype & genotype the progeny. Resolving this issue will also help to ease the deployment of marker assisted selection (MAS) for PF in a strawberry breeding program.

Finale: Marker assisted plant breeding

In this discussion I have touched upon many topics that deal with genetic research in general, for plants, and for strawberry in particular. In fact, it could be argued that so far, the discussion has lacked a sense of direction due to the wide variety of topics. In addition, I seem to have failed to bring issues from the prologue, such as molecular orphan crops and the flocking behavior around trending topics, to a satisfying closure. Finally, where do we see the practical plant breeder's perspective in all this?

Perhaps it is time to bridge the gap between academia and practical breeding...

Imagine that you are a practical plant breeder, to be more specific: a plant breeder in charge of a strawberry breeding program. Strawberry is an outbreeding and vegetatively propagated crop. Breeding strawberry involves the crossing of elite selections and screening the progeny for individuals with exceptional performance, which are then evaluated over several years. Typically, a strawberry breeding program screens tens of thousands of seedlings per year and is focused on several markets, which each have their own specific breeding goals. For instance, in Northern Europe we have open field production of early, mid and late varieties as well as ever bearers. Apart from open field, we also have tunnel and glasshouse production featuring different earliness types including ever bearers. Then there are Mediterranean varieties which differ considerably from Northern European varieties, and also make

use of different earliness types, and then there are the ever bearing varieties. On top of that there are strawberry varieties for industrial processing, and direct sales to consumers. This list is far from complete, but the general picture I am trying to paint here is quite clear: there are many production systems, all with different demands to what makes a good variety. Some varieties, such as Elsanta, are adaptable to various production systems, whereas others fit only in specific niches. The general trend however is that the time of broad purpose varieties is over, and that future varieties will have more specific adaptations to their respective production system.

This context is very important when you, as a breeder, want to introduce marker technologies into your breeding program. As a semi-molecular orphan crop, most practical breeding programs in strawberry, up until recently, did not make use of marker technology. The danger for a breeding program in this situation, is to get sucked into the whirlpool of currently "trending topics" in genetic research such as Genotyping By Sequencing (GBS), Genome Wide Association Studies (GWAS) and Genomic Selection (GS) without understanding or covering the basis first. This basis consists of important traits with simple genetics that are difficult to phenotype, and thereby priorities for marker development. This basis is otherwise known as the "low hanging fruit".

An example of such a low hanging fruit is the perpetual flowering trait of chapter 5, for which a marker could have considerable impact on a breeding program, enabling the breeder to throw away up to half of the seedlings just weeks after emergence as opposed to the following year with fully grown plants. Disease resistances and certain fruit quality traits are of the same category: spanning many market segments and of sufficient importance to warrant marker assisted selection. When sufficient markers that fit the low-hanging fruit category are available, the strawberry breeding program could routinely discard 50 to 80% of the totality of seedlings produced in a year. These percentages are rapidly achieved when an average of about three dominant traits are screened per crossing, and feasible from a cost perspective as well. This enables a breeding program to do more thorough phenotypic evaluations on the remaining seedlings, or, to increase the total number of seedlings screened. In strawberry, this basic set of markers covering the low hanging fruit can for a large part be discovered using simple QTL mapping in bi-parental crosses and when necessary, subsequent use of resequencing data to identify polymorphisms which tag a trait. GWAS and GBS could aid the discovery of (additional) markers suitable for MAS due to the wider genetic base and therefore increased potential of finding causal, or at least tagging, variants for traits. However, as suggested in the interlude, the employment of such methods should be considered only for traits where such problems are envisioned and extreme care should be taken in the choice of a suitable population and phenotyping method.

An example of a trending topic which could steer a small breeding program -on its maiden voyage across the ocean called molecular breeding- completely off course is Genomic Selection [48, 49]. Genomic selection (GS) entails complex statistical methodology which makes use of many different marker loci to predict the performance of an individual for a specific trait. Recently, there has been considerable attention for Genomic Selection in plant breeding. Several papers have been published describing the potential of GS in plant breeding [48, 50-52, 53, 54]. Genomic selection is heralded as the solution to accurate prediction of complex polygenic traits without phenotyping [48]. Indeed, the prediction accuracies of the models that are derived from GS seem very impressive, and its application in practice has met with success in cattle breeding [55] and shows promise in maize breeding [56]. However, there has also been some criticism on Genomic Selection for plants [57], but this appears to have gotten buried under the torrent of jubilant papers.

From my perspective, genomic selection has no place in most plant breeding programs, at least in its current state. There are numerous (even fundamental) reasons why Genomic Selection has limited applicability in plant breeding. Firstly, for most crops it is simply not economical. Seedling selection by GS makes no sense when phenotypic evaluation and marker assisted selection combined is only a fraction of the cost of GS. In later stages, with more advanced selections, the cost of phenotypic evaluation might in some cases exceed the costs of GS, but at this stage the number of genotypes under evaluation is generally much lower. The infrastructure required for GS (especially specialized personnel) is also problematic for any small breeding program. Another cost-related problem is the need for continuous updating of the prediction models as they lose accuracy over generations and between breeding groups and environments. This is directly in contradiction to one of the main purposes of GS, which is prediction of phenotypes without the hassle of doing the actual phenotyping. I assume that doing both the phenotyping and the GS on a recurrent basis does not constitute an improvement in efficiency. These downsides immediately rule out many small breeding programs for GS adoption. Parent selection is a more useful application of GS, and fits better with its original purpose in animal breeding. However, also for this purpose, arguments can be found against the use of GS.

A fundamental difference between plant breeding and animal breeding is that in animal breeding, GS is employed to ensure that the resultant progeny as a whole is genetically optimized for a certain trait, which makes sense, since you don't want your progeny to be useless clumps of meat (unless that is the trait under selection). In plant breeding however, we only have to find a specific genotype that is optimized, since we can replicate that genotype indefinitely. This means that at best, parent selection with GS makes the plant breeding program more efficient and

effective in terms of resultant population averages for a certain trait. While this is still a commendable achievement, we also have to consider other consequences and constraints in employing GS. When we go back to our strawberry breeding program, one such constraint can become quite clear. As mentioned at the start of this section, the success of a strawberry variety depends on a large number of traits, for which it needs to meet or exceed industry standards. Yield is an important trait, but is negatively correlated to e.g. earliness, taste and fruit size (more yield ~ more fruits ~smaller fruits). Many important traits affect other important traits in a complex manner, which will make a prediction model that optimizes one trait detrimental to others. On the other hand, a prediction model that needs to predict a compound trait such as "appearance", would lose accuracy even faster when predictions need to be made in other germplasm or environments. In plant breeding, the use of wild species or wide genetic backgrounds is quite common, the consequence is that new factors that influence a trait (either negatively or positively) are fairly common. As a consequence, even when loci truly have an effect in one population, that effect could easily be negated by other loci, enhanced, or even reversed (epistasis) despite that, strictly speaking, LD between marker and trait is maintained.

My final argument against using GS in a practical plant breeding program has to do with genetic diversity and recombination. When we consider the enormous parameter space from which GS prediction models are created, it is good to realize that this space also reflects the possible combinations of alleles (or the totality of possible unique genotypes) in a population. A genomic prediction model that enforces a strong selection pressure on specific alleles of hundreds of loci, greatly reduces the number of possible combinations in a population. When several beneficial or detrimental alleles are present on one homolog, strong selection pressure would essentially stop recombination for that part of the homolog. Although some flexibility is still allowed (not all top 2% individuals in a population will be near copies), the genome will become more static, especially when the prediction model is not updated over time, and definitely more static than under phenotypic selection alone. What it boils down to, is that DNA marker based selection pressure on many loci at once (genome wide) is effectively equal to background selection and will result in a loss of diversity, or at least reduce the possibilities of unique allelic combinations. This will have consequences on the long term for a breeding program and could mean the difference between success and failure. What it all boils down to, is that several of the factors that make GS a success in cattle breeding, are simply not present in most crop breeding programs.

The case of GS illustrates a widening gap between practical plant breeding and academic plant breeding research, specifically for "molecular orphan crops" such as

strawberry where research demands yesterday's technologies for yesteryear's topics. Despite that, I do not think the future looks bleak. Science inherently learns what paths are fruitful and which are not, even if it might sometimes take longer. As for the delayed introduction of marker technologies in strawberry there is at least one very positive aspect to it: it means that there is still a world waiting to be explored, and that for a molecular breeder of strawberry, these are exciting times indeed.

References

1. Alberts B: **Impact Factor Distortions**. *Science* 2013, **340**(6134):787-787.
2. Wren JD, Kozak KZ, Johnson KR, Deakyne SJ, Schilling LM, Dellavalle RP: **The write position - A survey of perceived contributions to papers based on byline position and number of authors**. *EMBO Rep* 2007, **8**(11):988-991.
3. Baldwin S, Revanna R, Pither-Joyce M, Shaw M, Wright K, Thomson S, Moya L, Lee R, Macknight R, McCallum J: **Genetic analyses of bolting in bulb onion (Allium cepa L.)**. *Theor Appl Genet* 2014, **127**(3):535-547.
4. Kim S, Kim M-S, Kim Y-M, Yeom S-I, Cheong K, Kim K-T, Jeon J, Kim S, Kim D-S, Sohn S-H *et al*: **Integrative structural annotation of de novo RNA-Seq provides an accurate reference gene set of the enormous genome of the onion (Allium cepa L.)**. *DNA Res* 2014.
5. Masamura N, McCallum J, Khrustaleva L, Kenel F, Pither-Joyce M, Shono J, Suzuki G, Mukai Y, Yamauchi N, Shigyo M: **Chromosomal Organization and Sequence Diversity of Genes Encoding Lachrymatory Factor Synthase in Allium cepa L.***G3-Genes Genom Genet* 2012, **2**(6):643-651.
6. Finkers R, van Workum W, van Kaauwen M, Huits H, Jungerius A, Vosman B, Scholten OE: **SEQUON-Sequencing the Onion Genome**. In: *Proceedings of the Plant & Animal Genome XXIII: 2015*; 2015.
7. Correll JC, Bluhm BH, Feng C, Lamour K, du Toit LJ, Koike ST: **Spinach: better management of downy mildew and white rust through genomics**. *Eur J Plant Pathol* 2011, **129**(2):193-205.
8. Feng C, Bluhm B, Correll J: **Construction of a Spinach Bacterial Artificial Chromosome (BAC) Library as a Resource for Gene Identification and Marker Development**. *Plant Mol Biol Report* 2015:1-10.
9. Zhang Q, Folta KM, Davis TM: **Somatic embryogenesis, tetraploidy, and variant leaf morphology in transgenic diploid strawberry (Fragaria vesca subspecies vesca 'Hawaii 4')**. *Bmc Plant Biology* 2014, **14**.
10. Davis TM, Yu H: **A linkage map of the diploid strawberry, *Fragaria vesca***. *J Hered* 1997, **88**(3):215-221.
11. Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM *et al*: **An enhanced microsatellite map of diploid**

Fragaria. *Theor Appl Genet* 2006, **112**(7):1349-1359.

12. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP *et al*: **The genome of woodland strawberry (*Fragaria vesca*)**. *Nat Genet* 2011, **43**(2):109-U151.
13. Bassil NV, Davis TM, Zhang HL, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR *et al*: **Development and preliminary evaluation of a 90 K Axiom (R) SNP array for the allo-octoploid cultivated strawberry Fragaria x ananassa**. *BMC Genomics* 2015, **16**.
14. Rousseau-Gueutin M, Lerceteau-Kohler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arus P, Guerin G, Denoyes-Rothan B: **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* 2008, **179**(4):2045-2060.
15. Sargent D, Passey T, Šurbanovski N, Lopez Girona E, Kuchta P, Davik J, Harrison R, Passey A, Whitehouse A, Simpson D: **A microsatellite linkage map for the cultivated strawberry *Fragaria x ananassa* suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection**. *Theor Appl Genet* 2012.
16. Sargent DJ, Fernandez-Fernandez F, Ruiz-Roja JJ, Sutherland BG, Passey A, Whitehouse AB, Simpson DW: **A genetic linkage map of the cultivated strawberry (*Fragaria x ananassa*) and its comparison to the diploid *Fragaria* reference map**. *Molecular Breeding* 2009, **24**(3):293-303.
17. Spigler RB, Lewers KS, Johnson AL, Ashman TL: **Comparative mapping reveals autosomal origin of sex chromosome in octoploid *Fragaria virginiana***. *J Hered* 2010, **101 Suppl 1**:S107-117.
18. Spigler RB, Lewers KS, Main DS, Ashman TL: **Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome**. *Heredity* 2008, **101**(6):507-517.
19. Van Dijk T, Pagliarani G, Pikunova A, Noordijk Y, Yilmaz-Temel H, Meulenbroek BJ, Visser RGF, Van De Weg WE: **Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map**. *BMC Plant Biology* 2014, **14**(55).
20. Zorrilla-Fontanesi Y, Cabeza A, Dominguez P, Medina JJ, Valpuesta V, Denoyes-Rothan B, Sanchez-Sevilla JF, Amaya I: **Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria x ananassa*)**. *Theor Appl Genet* 2011, **123**(5):755-778.
21. DiMeglio LM, Staudt G, Yu HR, Davis TM: **A Phylogenetic Analysis of the Genus Fragaria (Strawberry) Using Intron-Containing Sequence from the ADH-1 Gene**. *Plos One* 2014, **9**(7).
22. Rousseau-Gueutin M, Gaston A, Ainouche A, Ainouche ML, Olbricht K, Staudt G, Richard L, Denoyes-Rothan B: **Tracking the evolutionary history of polyploidy in**

***Fragaria L.* (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes**. *Mol Phylogenet Evol* 2009, **51**(3):515-530.

23. Tennessen JA, Govindarajulu R, Ashman TL, Liston A: **Evolutionary Origins and Dynamics of Octoploid Strawberry Subgenomes Revealed by Dense Targeted Capture Linkage Maps**. *Genome Biol Evol* 2014, **6**(12):3295-3313.
24. Bassil NV, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, Alperin ES, Rosyara UR *et al*: **Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa**. *BMC Genomics* 2015, **16**(1):1-30.
25. Darrow GM: **The strawberry; history, breeding, and physiology**, [1st edn. New York,: Holt; 1966].
26. Hancock JF, Luby JJ: **Genetic-Resources at Our Doorstep - the Wild Strawberries**. *Bioscience* 1993, **43**(3):141-147.
27. Maas JL, Gouin CC, Hokanson SC, Hartung JS: **Strawberry parent clones US 4808 and US 4809 resistant to bacterial angular leafspot disease caused by Xanthomonas fragariae**. *Hortscience* 2002, **37**(4):716-717.
28. Roach JA, Verma S, Peres NA, Jamieson AR, Weg WE, Bink MCAM, Bassil NV, Lee S, Whitaker VM: **FaRXf1: a locus conferring resistance to angular leaf spot caused by Xanthomonas fragariae in octoploid strawberry**. *Theor Appl Genet* 2016:1-11.
29. **The Gartner Hype Cycle** [http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp]
30. Jannink J-L, Walsh B: **Association mapping in plant populations**. *Quantitative genetics, genomics and plant breeding* 2002:59-68.
31. Korte A, Farlow A: **The advantages and limitations of trait analysis with GWAS: a review**. *Plant Methods* 2013, **9**.
32. Eikemo H, Stensvand A, Davik J, Tronsmo AM: **Resistance to crown rot (Phytophthora cactorum) in strawberry cultivars and in offspring from crosses between cultivars differing in susceptibility to the disease**. *Ann Appl Biol* 2003, **142**(1):83-89.
33. Schafleitner S, Bonnet A, Pedeprat N, Rocca D, Chartier P, Denoyes B: **Genetic variation of resistance of the cultivated strawberry to crown rot caused by Phytophthora cactorum**. *J Berry Res* 2013, **3**(2):79-91.
34. Shaw DV, Hansen J, Browne GT: **Genotypic variation for resistance to Phytophthora cactorum in a California strawberry breeding population**. *J Am Soc Hortic Sci* 2006, **131**(5):687-690.
35. Shaw DV, Hansen J, Browne GT, Shaw SM: **Components of genetic variation for resistance of strawberry to Phytophthora cactorum estimated using segregating seedling populations and their parent genotypes**. *Plant Pathol* 2008, **57**(2):210-215.
36. Churchill GA, Doerge RW: **Empirical Threshold Values for Quantitative Trait Mapping**. *Genetics* 1994, **138**(3):963-971.
37. Buhlmann P, Yu B: **Analyzing bagging**. 2002:927-961.
38. Valdar W, Sabourin J, Nobel A, Holmes CC: **Reprioritizing Genetic Associations in Hit**

Regions Using LASSO-Based Resample Model Averaging. *Genet Epidemiol* 2012, **36**(5):451-462.

39. Bink MCAM, Jansen J, Madduri M, Voorrips RE, Durel CE, Kouassi AB, Laurens F, Mathis F, Gessler C, Gobbin D *et al*: **Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple**. 2014, **127**(5):1073-1090.
40. Valdar W, Holmes CC, Mott R, Flint J: **Mapping in Structured Populations by Resample Model Averaging**. *Genetics* 2009, **182**(4):1263-1277.
41. Wurschum T, Liu W, Gowda M, Maurer HP, Fischer S, Schechert A, Reif JC: **Comparison of biometrical models for joint linkage association mapping**. 2012, **108**(3):332-340.
42. Heide OM, Stavang JA, Sonsteby A: **Physiology and genetics of flowering in cultivated and wild strawberries - a review**. *Journal of Horticultural Science & Biotechnology* 2013, **88**(1):1-18.
43. Weebadde CK, Wang D, Finn CE, Lewers KS, Luby JJ, Bushakra J, Sjulin TM, Hancock JF: **Using a linkage mapping approach to identify QTL for day-neutrality in the octoploid strawberry**. *Plant Breed* 2008, **127**(1):94-101.
44. Gaston A, Perrotte J, Lerceteau-Kohler E, Rousseau-Gueutin M, Petit A, Hernould M, Rothan C, Denoyes B: ***PFRU*, a single dominant locus regulates the balance between sexual and asexual plant reproduction in cultivated strawberry**. *J Exp Bot* 2013, **64**(7):1837-1848.
45. Amasino R: **Seasonal and developmental timing of flowering**. *Plant J* 2010, **61**(6):1001-1013.
46. Morris K, Thornber S, Codrai L, Richardson C, Craig A, Sadanandom A, Thomas B, Jackson S: ***DAY NEUTRAL FLOWERING* Represses *CONSTANS* to Prevent *Arabidopsis* Flowering Early in Short Days**. *Plant Cell* 2010, **22**(4):1118-1128.
47. Ahmadi H, Bringhurst RS, Voth V: **Modes of Inheritance of Photoperiodism in *Fragaria***. *J Am Soc Hortic Sci* 1990, **115**(1):146-152.
48. Jannink JL, Lorenz AJ, Iwata H: **Genomic selection in plant breeding: from theory to practice**. *Brief Funct Genomics* 2010, **9**(2):166-177.
49. Meuwissen THE, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps**. *Genetics* 2001, **157**(4):1819-1829.
50. Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J: **Genomic selection accuracies within and between environments and small breeding groups in white spruce**. *BMC Genomics* 2014, **15**.
51. Desta ZA, Ortiz R: **Genomic selection: genome-wide prediction in plant improvement**. *Trends Plant Sci* 2014, **19**(9):592-601.
52. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME: **Plant Breeding with Genomic Selection: Gain per Unit Time and Cost**. *Crop Sci* 2010, **50**(5):1681-1690.
53. Nakaya A, Isobe SN: **Will genomic selection be a practical method for plant breeding?***Annals of Botany* 2012, **110**(6):1303-1316.
54. Wurschum T, Reif JC, Kraft T, Janssen G, Zhao YS: **Genomic selection in sugar beet**

breeding populations. *Bmc Genetics* 2013, **14**.
55. Pryce JE, Daetwyler HD: **Designing dairy cattle breeding schemes under genomic selection: a review of international research**. *Anim Prod Sci* 2012, **52**(2-3):107-114.
56. Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE: **Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize**. *Genetics* 2014, **197**(4):1343-U1469.
57. Jonas E, de Koning DJ: **Does genomic selection have a future in plant breeding?** *Trends Biotechnol* 2013, **31**(9):497-504.

Summary

Summary

The garden strawberry (*Fragaria x ananassa*) is a fruit species that was developed through human intervention less than 300 years ago. Currently, it is the most important soft fruit in both production as well as value and renowned for its deliciousness. There are many challenges in growing such a delicate fruit, many of which have been overcome through improved cultivation techniques and breeding. The perishability of the product is, however, not the only challenge faced by strawberry breeders. In terms of genome composition, strawberry appears to have accumulated a wonderful array of obstacles to genetic studies. It is a vegetatively propagated allo-octoploid outbreeder, and only few crop species are worse of in this respect. Many of the molecular genetic ground work is therefore performed in its diploid ancestor, the woodland strawberry *Fragaria vesca*, which was sequenced in 2011. However, since nearly all strawberries that are eaten are octoploid, genetic research can't linger at the wild diploids forever. In this thesis we developed new tools and analysis methods for genetic studies in the allo-octoploid strawberry and subsequently applied these methods in the detection of marker-trait associations.

The purpose of **Chapter 2** was to develop a method to interpret the complex peak patterns generated by microsatellites in octoploid strawberry in such a way that we ended up with as much information as one would expect to retrieve from a microsatellite in a diploid system. This information could then be used to generate high quality linkage maps for the different sub-genomes and allow for easy alignment and comparison. We named the method MADCE, which stands for Microsatellite Allele Dose & Configuration Establishment. In the MADCE methodology, we first need to determine the dose of each allele present in an individual. For this we used the signal of fluorescent microsatellite peaks in relation to the total fluorescent signal generated by all peaks for that microsatellite. We then used the disomic inheritance of strawberry to establish the allelic configuration of each different homoeologue (subgenome). The repulsion of alleles from the same subgenome in offspring allowed us to form subgenomic pairs of alleles. We found that in single cross mapping populations, the deployment of our method was fairly easy due to the high number of offspring that can be used to establish repulsion between alleles. However, for pedigreed breeding germplasm this was another matter, as generally only few offspring were available. For this we added some additional tricks to the MADCE method, although some uncertainty about the configuration would remain for problematic lines and alleles.

In **Chapter 3** we used the MADCE method from Chapter 2 to generate a genome wide linkage map for the Holiday x Korona (HxK) mapping population.

This linkage map was to be used in subsequent experiments for QTL discovery as well as provide the strawberry community with a highly detailed map consisting not only of marker distances, but allele and haplotype configuration of the parents Holiday and Korona as well. The haplotype information revealed that inbreeding (homozygosity) levels in Holiday were similar to the levels expected from its pedigree, but that inbreeding levels of Korona were more than three times higher than expected, which could be resultant from selection pressure enacted by breeders. Selection pressure could also be causal to our discovery that the kinship between the two cultivars was twice as high as expected from their shared ancestry. Another discovery was a large inversion on one of the subgenomes of linkage group 2 (D). Up until the publication of our linkage map this inversion had not been reported in other linkage maps. Another innovation was our attempt at giving a biological or evolutionary meaning to the denomination of the linkage groups by arranging them according to similarity to the diploid ancestor *F. vesca*, based on *F. vesca* derived primer amplification efficiencies. The HxK map has been used in several (ongoing) research projects outside of our research group and has contributed to the development of the 90k Axiom SNP array for cultivated strawberry.

In **Chapter 4** we performed a QTL mapping study for disease resistance against the problematic pathogen *Phytophthora cactorum*, which causes crown rot in strawberry plants. In this study we used two different mapping populations: the Holiday x Korona (HxK) population from the previous chapter as well as E1998-142 x Elsanta (ExEls), developed more specifically for the purpose of finding resistance against *P. cactorum*. The HxK and ExEls populations were phenotyped over three years (2008, 2010 and 2011) under different seasons and conditions. The correlation between years for was quite low for both populations (ranging from 0.18 to 0.47), indicating a large environmental effect on disease pressure. Results from the QTL analysis showed that most QTLs were small in effect and only just above the statistical significance threshold. Only for ExEls we uncovered two QTLs with relatively high significance levels, but none were significant in all three years. Because of the high environmental influence, and the desire to have QTLs that are robust over environments, we used the average of all three years (AOTY) as an additional phenotype. When we used the AOTY trait, the QTL on LG7D became stronger than for any of the individual years. Whereas for the LG7C QTL the significance dropped to just below threshold levels. These results indicated that removing environmental noise through averaging over experiments is a good way to uncover the most reliable and therefore more valuable to a breeding program.

In **Chapter 5** we investigated the genetics behind two different flowering habits that are grown commercially worldwide: seasonal flowering habit (SF) and

perpetual flowering (PF) These varieties initiate flowering under long days, and can therefore produce fruit for a much longer period: throughout the summer and early fall. Evidence from literature and practical breeding suggested that PF is under dominant control. We decided to treat PF as a qualitative trait and divided two small mapping populations into PF and SF individuals. After screening several microsatellites, we found one locus that completely cosegregated with the PF trait at the bottom of LG4D. At the moment of mapping, a paper was published which mapped the same trait to the same location. We found that there were two very clear candidate genes within our QTL interval, *FaCDF2* and *FaFT2*, which were homologous to genes that are major factors in the flowering pathway of Arabidopsis and many other plant species. We then sequenced the *FaCDF2* gene from a number of distinct PF and SF cultivars. This resulted in the discovery of two quite distinct allelic variants, one of which was present in all PF cultivars. However this variant was also present in some of the SF cultivars, indicating that either *FaCDF2* is not the causal gene, or that other loci can have a qualitative effect on the switch from SF to PF. We then performed microsatellite haplotyping on hundreds of cultivars and this revealed that all PF varieties of all origins carry the same haplotype in the PF QTL region, and that there weren't any recombinations between the candidate genes *FaCDF2* and *FaFT2*, which are 250kb apart on the physical genome. This makes it still undecided which of these two candidate genes are causal to the PF trait. Another interesting result from the haplotyping was that the complete PF haplotype was present with moderate frequency in SF varieties as well. Not only does this suggest a common origin, it also complicates the establishment of a theory for the mechanisms behind perpetual flowering in cultivated strawberry. So far we have not been able to establish whether the PF haplotype that is present in SF cultivars is functionally distinct from the PF haplotype in PF cultivars. All we know is that it does not confer perpetual flowering in these SF cultivars, and further experiments would be needed to find out the exact mechanism behind perpetual flowering.

In the general discussion (**Chapter 6**), the results of this thesis were placed in the broader context of science in general and plant breeding in particular.

ACKNOWLEDGEMENTS

CURRICULUM VITAE

I have dreaded writing this part, since it has been so long since I started my PhD, in a few months it will have been exactly eight years. A lot can happen in eight years, people come, people go, a department becomes really dynamic when you compress time in your mind. Many of the PhD students that started before and around the same time as me, are long gone, many who started after me as well, for that matter. Still, many are still in the "neighbourhood" working for breeding companies or in academics. Another reason I dreaded this is because I am not known for being the most generous expresser of gratitude. I am sure that I will miss out some people by the end of these acknowledgements: if you are one of them, feel free to write your name below.

The first person I'd like to thank, is my supervisor Eric van de Weg. Eric, your mind is quite unique, there have been times that I felt I was communicating to an extra-terrestrial, and perhaps that feeling has sometimes been mutual. Nevertheless I have learned a lot from you, among which your attention to detail and perfectionism are the most obvious. Our styles are vastly different, I tend to feel around for the path of least resistance unafraid of leaving unfinished work, whereas you strive for completeness. Because of that, we both share the same handicap of never being finished with something. Another thing we share is an insatiable analytical mind and our discussions really helped in setting the path of this thesis.

Richard Visser, rational, pragmatic, objective, as the leader of the plant breeding department and promotor to more than a hundred PhDs, those qualities define you. Thanks for the critical analysis of my chapters and your fast response time and availability under such a busy schedule. I'd also like to thank you for gently nudging me into the direction of a completed thesis by simply stating whether I'd like to graduate in October or November.

Yolanda and Herma, Microsatellites are a pain to work with, especially in strawberry, and I applaud you for your tireless efforts in producing and helping to analyse the infinite amounts of fluorescent peaks that we generated. Remmelt, thanks for teaching me the practicalities of setting up large scale disease assays, as well as the production of inoculum. Together we attempted to murder thousands of plants, and I am proud to say we have, at times, been quite successful. Wendy and Doret, thanks for the help in DNA isolation and cloning respectively, and I'd also like to thank Fien for showing restraint in the confiscation of my freezer space.

I'd also like to thank the students and guest-workers what we had during my time at Eric's group, so here's to you Anna, Hulya, Sara, Giulia, Mario, Elmira, Madhu and Aziz, we had a great time together.

Bert, Karlijn, Rob, Kees, Jack, Geert, Frans, Pluk and my other colleagues at Fresh Forward, thanks for the help and support in supplying plants, facilities and labour during my thesis. Also, I am sorry in advance for any future impossible requests. Bert and Christian, thanks for allowing me to finish this thesis in part during working hours.

A special thanks to lunch buddies Peter Vos and Maarten Verlaan. Peter, we started here together in 2001, took separate paths through our studies, re-joined during our PhDs and now we will graduate within a week of each other. They say that men think of a certain topic several times each minute, and in your case Peter, that's potatoes. Maarten, you and I share a similar sense of humour, and I especially enjoyed your creative input on WhatsApp.

I'd like to thank my long-time office-mates Henk Schouten and Sabaz, for the discussions we had, both professionally as well as personally. Henk, we still share our office, and although I am not there as often as before anymore, we always have something to talk about.

I'd also like to thank Rene Smulders, we had discussions on various scientific topics and I have always been impressed by your knowledge and 'Vulcan'- like mastery of logic. Which brings me to Theo Borm, with whom discussions are always a fun mixture of technical knowhow flavoured with a heavy dose of scepticism. In that context I'd also like to thank Timothy Leary, for his words of wisdom.

Next, I'd like to thank my fellow victims, the (ex) PhD students whom I have met during my time here and with whom I have enjoyed company on occasions: Peter B., Johan W., Michiel, Geert, Arwa, Ram, Yusuf, Andres, Christos, Jan, Martijn, Diana, Dennis, Paul B., Tim, Ernest, Socrates, Peter Q, Manos, Marcela, Bjorn, Suxian, Marleen, Xuan and Erik. As well as the other colleagues of Plant Breeding such as Koen, Colette, Abe, Paul A., Chris, Sjaak, Yury, Arnaud, Patrick, Ronald, Herman, Roeland, Rients, Johan B., Jack, Ningwen, Greet, Nicole, Janneke, Letty, Annie and many others. They made for a great working environment and we often had great fun at the lab-trips and other social events.

I would also like to thank in advance my Paranymphs, Julian Verdonk and Emil Kraaikamp for their willingness to stare into emptiness during most of my defense.

A keen interest in biology and nature are what brought me here in the first place, and for that I would like to thank my family. First my parents Henk & Aly for raising me in an environment where you could always go out and explore nature, even if sometimes that meant setting fire to it, or stealing eggs from bird's nests. Also my older brothers and sister, Bert, Harry, Hanny and Alwin, for paving the way and teaching me how to steal bird's eggs and setting things on fire. For my interest in agriculture and growing things I'd like to thank oom Willem.

I'd like to thank my wife Emilie, for her support, I know it must have been hard for you to observe my slow pace and rather unworried approach to finishing this thesis. I can understand that, sometimes my unworriedness worries me as well. I guess the only consolation I can offer is that by finishing so late, I made the seven years it took you seem rather fast. My last thanks goes out to my two sons, Lute and Berend, for robbing me of all my free time without me minding it.

Thijs van Dijk

Curriculum vitae

Thijs van Dijk was born on the 29th of January 1983 in Dwingeloo, The Netherlands. After completing Atheneum in Meppel he enrolled at Wageningen University in 2001. There he studied Plant and Crop Sciences, with the specialisation plant breeding and genetic resources. His minor thesis was performed with Dr. Erin Bakker at the Laboratory of Nematology in 2005, where he fine-mapped the Grp1 resistance locus against the potato cyst nematode. His internship was performed in 2005 at Seminis Vegetable Seeds (now Monsanto Vegetable Seeds), where he fine-mapped QTLs against downy mildew (*Bremia lactucae*) in lettuce. In 2007 he completed his major thesis at the Department of Plant Breeding, at the research group of Dr. Rients Niks on the topic of a non-host resistance QTL against different rust species (*Puccinia spp.*) in Barley. In 2007 he graduated and started working at Monsanto Vegetable Seeds as a research associate at the Brassica and Leafy vegetable group, developing markers for use in Marker Assisted Selection (MAS). In January 2009 he started a PhD study in the Department of Plant Breeding with Dr. Eric van de Weg. The objective of this study was to develop molecular tools in the allo-octoploid strawberry and to apply them in QTL studies on traits of interest. Since 2014 he has been working at the breeding company Fresh Forward, as a molecular breeder of strawberries.

Publications

Bassil NV, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, Mahoney L, Wood D, van Dijk T, Rosyara UR *et al*: **Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria* × *ananassa***.*BMC Genomics* 2015, **16**(1):1-30.

Finkers-Tomczak A, Danan S, van Dijk T, Beyene A, Bouwman L, Overmars H, van Eck H, Goverse A, Bakker J, Bakker E: **A high-resolution map of the Grp1 locus on chromosome V of potato harbouring broad-spectrum resistance to the cyst nematode species *Globodera pallida* and *Globodera rostochiensis***. *Theor Appl Genet* 2009, **119**(1):165-173.

Khan SA, Chibon PY, de Vos RCH, Schipper BA, Walraven E, Beekwilder J, van Dijk T, Finkers R, Visser RG, van de Weg EW *et al*: **Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16**. *J Exp Bot* 2012, **63**(8):2895-2908.

Van Dijk T, Noordijk Y, Dubos T, Bink MCAM, Meulenbroek EJ, Visser RGF, Van de Weg WE: **Microsatellite Allele Dose and Configuration Establishment (MADCE): an integrated approach for genetic studies in allopolyploids.***BMC Plant Biology* 2012, **12**(25).

Van Dijk T, Pagliarani G, Pikunova A, Noordijk Y, Yilmaz-Temel H, Meulenbroek EJ, Visser RGF, Van De Weg WE: **Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map**. *BMC Plant Biology* 2014, **14**(55).

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: **Thijs van Dijk**
Date: **11 November 2016**
Group: **Laboratory of Plant Breeding**
University: **Wageningen University and Reserach**

1) Start-up phase	*date*
▸ **First presentation of your project**	
Title: 'Mapping in octoploid strawberry'	Jun 26, 2009
▸ **Writing or rewriting a project proposal**	
▸ **Writing a review or book chapter**	
▸ **MSc courses**	
▸ **Laboratory use of isotopes**	
Subtotal Start-up Phase	*1.5 credits**

2) Scientific Exposure	*date*
▸ **EPS PhD student days**	
EPS student day 2009 (Leiden)	Feb 26, 2009
EPS student day 2011 (Wageningen)	May 20, 2011
▸ **EPS theme symposia**	
EPS theme 4 symposium 'Genome Biology' (Wageningen)	Dec 10, 2010
EPS theme 1 symposium 'Developmental Biology of Plants' Wageningen)	Jan 21, 2016
▸ **Annual Meeting EPS and other National Platforms**	
Annual Meeting 'Experimental Plant Sciences' 2009 (Lunteren, NL)	Apr 06-07, 2009
Annual Meeting 'Experimental Plant Sciences' 2010 (Lunteren, NL)	Apr 19-20, 2010
Annual Meeting 'Experimental Plant Sciences' 2011 (Lunteren, NL)	Apr 04-05, 2011
Annual Meeting 'Experimental Plant Sciences' 2012 (Lunteren, NL)	Apr 02-03, 2012
Annual Meeting 'Experimental Plant Sciences' 2013 (Lunteren, NL)	Apr 22-23, 2013
Annual Meeting 'Experimental Plant Sciences' 2014 (Lunteren, NL)	Apr 14-15, 2014
▸ **Seminars (series), workshops and symposia**	
Plant Breeding Research day 2009 (Wageningen)	Mar 03, 2009
WorkShop At Fresh Forward Breeding	Mar 09, 2009
Linkage disequilibrium and association mapping – helping to overcome the paradox of modern plant breeding, by Dr Wallace	Jun 26, 2009
Plant science seminar Harro Bouwmeester & Ton Bisseling	Sep 08, 2009
Connecting genetics and genomics of pathogenicity and behavior in root-knot nematodes, Valerie Williamson – Dept of	Oct 23, 2009
Symposium "Photosynthesis: from femto to Peta and from nano to Global"□□	Nov 05, 2009
Plant science seminar Fred van Eeuwijk & Pierre de Wit	Nov 10, 2009
Plant science seminar Ken Giller & Richard Visser	Dec 08, 2009
Plant Breeding Research day 2010 (Wageningen)	Feb 08, 2010
Fresh Forward Lustrum Symposium	May 18, 2010
Seminar Professor Richard Michelmore at Keygene	Jun 03, 2010
Networking Event of TTI Green Genetics 2010	Sep 22, 2010
Seminar Régine Delourme Genetic and functional analysis of disease resistance in Brassica	Oct 05 2010
Plant Breeding Research day 2011 (Wageningen)	Mar 08, 2011
Networking Event of TTI Green Genetics 2011	Sep 21, 2011
Plant Breeding Research day 2012 (Wageningen)	Feb 28, 2012
Fascination of Plants day (Wageningen)	May 18, 2012
Networking Event of TTI Green Genetics 2012	Sep 19, 2012
"Role of the cytoplasmic genome in potato breeding"	Jul 17, 2014
Plant Breeding Research day 2014 (Wageningen)	Sep 23, 2014
▸ **International symposia and congresses**	
International conference FoodOmics (COST 863), Cesena, Italy	May 28-29, 2009
PAG XVIII Plant animal Genomen, San Diego, USA	Jan 09-13, 2010
International Horticulture Congress Lisbon	Aug 23-27, 2010
6th Rosaceae genomics conference, Trentino, Italy	Sep 30-Oct 04 2012
Next Generation Plant Breeding, Ede NL	Nov 11-14, 2012
▸ **Presentations**	
Talk: Fresh Forward Breeding company Eck&Wiel	Mar 09, 2009
Talk: Fresh Forward Breeding company Rhenen (lustrum)	May 18, 2009
Poster: International Berry Breeding Symposium Ancona Italy	May 28, 2009
Talk: Fresh Forward Breeding company Tiel	Oct 07, 2009
Talk: Plant Animal Genome conference (Fruit & nuts workshop)	Jan 09, 2010
Poster: IHC 2010 Lisbon, Portugal	Aug 23-24, 2010
Talk: Resistentieveredeling aardbei themamiddag aardbei DLV	Nov 15, 2010
Poster: TTI Green Genetics networking event 2011	Sep 21, 2011
Talk: Studiekring Plantenveredeling	Nov 18, 2011
Talk: Next Generation Plant Breeding Conference Ede	Nov 14, 2012
▸ **IAB interview**	
Interview with a member of the International Advisory Board of EPS	Feb 17, 2011
▸ **Excursions**	
Subtotal Scientific Exposure	*26.5 credits**

3) In-Depth Studies	*date*
▸ **EPS courses or other PhD courses**	
International Berry Summer School Ancona (COST 863)	May 25-27, 2009
Course 'Linear Models', Wageningen	May 30-Jun 01, 2011
Course 'Generalized Linear Models', Wageningen	Jun 09- Jun 10, 2011
Course 'Mixed Linear Models', Wageningen	Jun 20-Jun 21, 2011
Course 'Introduction to R', Wageningen	Oct 24-Oct 25, 2011
Course 'Bayesian Statistics', Wageningen	Oct 27-Oct 28, 2011
▸ **Journal club**	
Participation in a Literature Discussion Group Plant Breeding	2009-2011
▸ **Individual research training**	
Subtotal In-Depth Studies	*5.7 credits**

4) Personal development	*date*
▸ **Skill training courses**	
Advanced Course to scientific Artwork	Nov 04-05, 2010
Statistical learning methods for DNA based prediction of complex traits	Oct 17-21, 2011
▸ **Organisation of PhD students day, course or conference**	
Organizing The Plant Breeding Monday Seminar	Sep 2009-Apr 2012
Fascination of plants day stand : Aardbeien proef het verschil	May 18, 2012
▸ **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*3.8 credits**

TOTAL NUMBER OF CREDIT POINTS*	**37.5**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational

** A credit represents a normative study load of 28 hours of study.*

This work was performed at the Laboratory of Plant Breeding, Wageningen University, and financially supported by Technological Top institute Green Genetics (TTI), an initiative of the Dutch Ministry of Agriculture, Nature and Food Quality and the Dutch plant cultivation industry. Fresh Forward Breeding B.V. is acknowledged as the private participant for this TTI project.