UNIVERSITÀ DEGLI STUDI DI UDINE

Dottorato di Ricerca in Scienze e Biotecnologie Agrarie
Ciclo XXIX
Coordinatore: prof. Giuseppe Firrao

TESI DI DOTTORATO DI RICERCA

MAPPING TRAITS OF AGRONOMIC INTEREST IN KIWIFRUIT AND DEVELOPMENT OF MARKERS FOR MARKER-ASSISTED SELECTION (MAS)

DOTTORANDA

Catalina Pinto Palacios

SUPERVISORE

Prof. Raffaele Testolin

CO-SUPERVISORE
Prof. Rodrigo Infante, Universidad de Chile

ANNO ACCADEMICO 2016/2017

"No importa que tan lento vayas
mientras no te detengas"
Confusio

ACKNOWLEDGEMENTS

First, a special thank goes to my family for their constant love and support, despite the distance. Furthermore, I would like to express my sincere gratitude to my advisor Prof. Raffaele Testolin for the continuous support of my Ph.D study. His guidance helped me during my research and my stay in Italy. My sincere thank also goes to Prof. Rodrigo Infante, who gives me the opportunity to participate to this project. Beside my advisors, I would like to thank Prof. Guido Cipriani and Giorgio Comuzzo for their support, insightful comments and encouragement. I also take this opportunity to thank all of you who have contributed to this research. Finally, I thank my friends for all the fun and smiles that we have had in the last three years. In particular, I am grateful to Gloria De Mori, Serena Foria, Laura Zanon and Federica Arrigoni that helped me in the field, laboratory, and computer work, but above all, for sharing my life in Italy.

This research was developed within the program Chile-Italia D09I1136 "Mejoramiento Genético del Kiwi apoyado en la Selección Asistida por Marcadores" supported by the Chilean Government- FONDEF, and with the funding of BECAS CHILE - 72160494 scholarship.

INDEX OF CONTENT

SUMMARY

In kiwifruit, dioecy and intraspecific polyploidy are great challenges in breeding and cultivar improvement. Marker-assisted selection (MAS) can enhance the selection efficiency and precision in kiwifruit breeding. The prerequisite to MAS is the construction of a high-density linkage map allowing linkage association between marker loci and quantitative trait loci (QTLs) of particular traits or genes of interest. Consequently, the production of dense maps in kiwifruit and mapping of traits, in particular, gender determinant and fruit quality traits was the aim of this thesis. Thus, a mapping population was generated from a cross between 'C8' (female) and 'A54.19' (male), both diploid kiwifruit (*A. chinensis*), genetically and geographically unrelated. The work of thesis is composed of five chapters.

In the first chapter, a literature review that supports the main aspects in kiwifruit breeding is presented.

In the second chapter, we report the construction of genetic maps by merging ddRAD SNPs and SSRs markers. Results showed that even in the presence of a reference genome (Huang et al. 2013), our genetic maps help to improve the genome assembly. Furthermore, the saturated maps help to restrict the chromosomal region where genes are located and to find genetic loci that control Mendelian traits like the gender. Thus, we have isolated four SSR markers by screening the contigs of sex-determining region. These markers were able to discriminate between male and female progeny, but they cannot be used to screen germplasm collection where the parentage of accessions is not known in advance.

In the third chapter, quantitative trait loci (QTLs) analysis traits related to fruit quality were carried out using fruit from mapping population. Many QTLs had large effects in terms of explained variance, however, they showed poor stability probably due to the limited size of the progeny.

In the fourth chapter, estimation of genetic variance and selection of superior male parents for their capability of transferring traits of interest to progeny were performed using best linear unbiased prediction (BLUP) and REML in a controlled mating design, involving 35 small cross families. This analysis include the estimation of genetic

component of variance, heritability, the combining ability of cross parents and breeding value for several vegetative and fruit quality traits in kiwifruit.

In the last chapter, we describe the sequencing of male parent of our mapping population and the future steps towards the exploration of the sex-determining region.

CHAPTER 1

OVERALL INTRODUCTION

Kiwifruit are amongst the best examples of the successful domestication and commercialization of a new crop (Ferguson 2016a). Seeds of *Actinidia deliciosa* [(A. Chev.) C.F. Liang et A.R. Ferguson] were introduced from China, the homeland of the genus *Actinidia*, to New Zealand at the beginning of the 20th century, and a small number of male and female plants selected from those seedlings gave rise to the few cultivars upon which the world-wide kiwifruit industry currently relies (Ferguson et al. 1996). Today, kiwifruit are considered an important fruit crop of temperate regions and are planted world-wide.

The total of world production, currently about three million tons of fruit per year, gives the kiwifruit industry a prominent place in the fresh fruit market. China produces about half of the total world production, but mostly for local consumption. On the other hand, Italy, New Zealand, Chile and Greece produce kiwifruit mainly for export. The predominant cultivar grown is *A. chinensis* var. *deliciosa* 'Hayward', a green-fleshed kiwifruit, which make up about two-thirds of the world production (Ferguson 2016a).

1.1 Motivation of this research

In Southern Hemisphere, New Zealand and Chile compete for Northern Hemisphere markets, but Chile has lower returns in comparison to New Zealand because of quality of fruit. Furthermore, productivity of Chilean orchards is lower than other countries. These two problems, in large part, are due to the fact that kiwifruit cultivation and world trade is based on a single cultivar, 'Hayward', as previously mentioned (Ferguson 2016a). This cultivar was developed for agroclimatic conditions different from those of Chile, and therefore its potential can be limited in Chilean conditions.

For these reasons, in order to generate new varieties with high sensory quality and improve the postharvest life, a kiwifruit breeding program using marker-assisted selection (MAS) was developed in collaboration with Italy.

This thesis is part of the program "Mejoramiento genético del kiwi apoyado en la selección asistida por marcadores" supported by Chilean Government, which includes a cooperation of the University of Chile with the University of Udine and the Institute of Applied Genomics (IGA), a non-for-profit Institute of Research linked to the University of Udine. Saturated genetic linkage maps in *Actinidia* would permit to provide tools for quantitative trait mapping, and to investigate the evolution and function of genetic control mechanisms. Consequently, the aim of this thesis was the production of saturated maps in kiwifruit and mapping of traits of agronomic interest, in particular, gender determinant and fruit quality related traits.

1.2 Literature review

Although kiwifruit breeding is recent, advances in molecular genetics have already revolutionized the approach to objectives and strategies in this species. The following overview of the literature covers current knowledge and research opportunities in kiwifruit breeding.

1.2.1 Kiwifruit breeding

Kiwifruit belongs to the genus *Actinidia*, which includes more than 60 species of climbing plants originating mainly in southern China (Liang 1984, Huang et al. 2014). According to Ferguson et al. (1996), the genetic base of the kiwifruit industry is extraordinarily narrow in comparison with the vastness of the wild genetic resources. Currently cultivars from three species are grown commercially: the green-fleshed kiwifruit, *Actinidia deliciosa*, the closely related yellow-fleshed *A. chinensis* Planch, and the kiwiberry, *A. arguta* (Sieb. et Zucc.) Planch. ex Miq.

Most of the kiwifruit cultivars grown commercially are selections of chance seedlings and there has been little systematic breeding (Ferguson and Huang 2007, Crowhurst et al. 2008). Huang (2016) mentioned that most species of genus *Actinidia* produce edible fruit and there is tremendous genetic variation at both the interspecific and the infraspecific level. Consequently, there are still many characteristics within the genus that could be incorporated into commercial cultivars (Ferguson 2007, Ferguson and Huang 2007).

As just said, in the past kiwifruit received very little attention from geneticists and breeders (Testolin 2011). One reason was, as previously mentioned, kiwifruit is a relatively new crop. Another were ploidy variation and dioecism, which make genetic studies difficult (Testolin et al. 2001; Marsh et al. 2003).

Actinidia has the basic chromosome number (x = 29) and there is also a very wide variation in ploidy ranging from diploid to hexaploid, with even occasional octoploid, decaploid or dodecaploid individuals (Ferguson and Huang 2007). In particular, within the genus, no assumptions should be made as to the ploidy of any particular genotype. Many taxa contain ploidy races and the ploidy of any genotype used in breeding programs should therefore be checked (Ferguson and Huang 2016). Ploidy races have been most fully studied in the *A. chinensis* species complex. This complex currently contains three varieties, *A. chinensis*, *A. deliciosa*, and *A. setosa*, which encompasses diploid, tetraploid and hexaploid races with morphological traits so similar that they have been treated, until recently, as a single species by botanists (Liang 1984). These polyploid races of the complex may be derived from diploid *A. chinensis* genotypes without there being a contribution from any other *Actinidia* species (Testolin and Ferguson 1997). Testolin et al. (2001) considered that if genetic studies are carried out on diploid races of *A. chinensis* the information obtained will be relevant to related polyploids such as tetraploid *A. chinensis* or hexaploid *A. deliciosa*.

One of the most characteristics features of the genus is dioecy. The great majority of flowering plants are hermaphrodites, however, a small portion (5 %) are dioecious with separate male and female plants (Charlesworth and Guttman, 1999). According to Ferguson (2016b), all *Actinidia* species are apparently dioecious, but functional dioecy has been recently confirmed in only a few species (Kawagoe and Suzuki 2004). Male vines carry staminate flowers with numerous stamens producing viable pollen, and rudimentary ovaries, which lack styles and do not form ovules. Female vines carry pistillate flowers with both well-developed stamens, but the pollen they produce is sterile (Testolin et al. 1999). It has been proposed that stamens of pistillate flowers aid reproduction by attracting pollinating insect (Kawagoe and Suzuki 2004). Dioecism is not absolute, and male vines bearing small fruit have been described (McNeilage 1991, Testolin et al. 1999), but being such incipient hermaphroditism erratic, breeding cannot be based on such genotypes. Consequently, due to the dioecism, kiwifruit has an inevitable disadvantage in breeding; paternal parents are selected with unknown fruit quality since male plants cannot bear fruit (Marsh et al. 2003; Testolin et al. 2011). In a typical cross population, dioecism results in a final sex ratio of 1:1 for female to male

vines (Testolin et al. 1995). Therefore, these male plants can represent a part of the progeny that are a waste of land and resources (Zhang et al. 2015).

Additionally, Marsh et al. (2003) indicated that another major limitation to progress in kiwifruit breeding is the long generation cycle. According to Van Nocker and Gardinier (2014), development and introduction of improved cultivars by plant breeders may require many breeding cycles and dozens of years. This cost of operating and maintaining of relatively large populations over a number of years is a limiting factor in any woody breeding program, and this is particularly so for kiwifruit where expensive support structures are necessaries (Marsh et al. 2003). Nowadays, recent advances in biotechnologies and genomics have the potential to accelerate cultivar development greatly in all crops. There would be considerable benefits in kiwifruit as well if the number and the duration of breeding cycles for tree crops are reduced (Van Nocker and Gardiner 2014).

Interestingly, recently publication of the genome sequence of *Actinidia chinensis* var. *chinensis* 'Hongyang' by a Chinese team (Huang et al. 2013) provided the kiwifruit scientific community with an assembled and annotated genome and offered great opportunities for advances in many areas of scientific research, breeding included.

1.2.2 Marker-assisted selection (MAS)

The development of techniques for the analysis of DNA polymorphisms and the continuous reduction in cost of molecular marker production generated the so-called marker-assisted selection (MAS), or marker-assisted breeding (MAB), which is a method whereby a phenotype is selected for the trait-associated marker (Collard et al. 2005; Testolin 2011). This approach using the presence or absence of a marker as a substitute for or to assist in phenotypic selection, in a way which may make it more efficient, effective, reliable and cost-effective compared to the more conventional plant breeding methodology (Collard et al. 2005). The steps required for the development of markers for use in MAS includes: the isolation/identification of molecular markers, the construction of linkage maps, and subsequently, the identification of genomic regions associated with traits of interest, that could be either Mendelian or quantitative traits, the latter being known as QTL (McCough and Doerge 1995; Mohan et al. 1997). DNA markers that are tightly linked to important genes may be used as molecular tools for MAS in plant breeding (Ribaut and Hoisington 1998).

1.2.2.1 Genetic maps

According to Testolin et al. (2001), unfortunately, dioecy maintains high levels of heterozygosity, confirmed by the difficulties encountered by the Chinese scientist in assembling the kiwifruit genome (Huang et al. 2013), and precludes the possibility of producing inbred lines, which are the genotypes of choice for most genetic analyses, including association analysis. Therefore, the construction of a genetic map in an obligate outbreeding species, such as *A. chinensis*, is more complex (Fraser et al. 2009). However, maps in outbreeding species have been developed, by adopting the so-called two-way pseudo-test cross, which permits the estimation of recombination frequencies from populations obtained by crossing heterozygous parents (Ritter et al. 1990; Grattapaglia and Sederoff 1994). A cross between two non-identical plants of an outbreeding species may segregate for up to four alleles per locus, and this may vary between loci, while the linkage phases usually are unknown (Maliepaard et al. 1997). However, the recombination frequencies can be separately estimated for each parent so that two maps are developed, and these maps can be integrated using markers that are heterozygous in both parents and therefore can bridge the two maps, identifying the homologous chromosomes (Fraser et al. 2009).

This strategy of two-way pseudo-test cross and the abundance of genomics and EST-derived SSR markers (Simple Sequence Repeats) in kiwifruit (Huang et al. 1998; Crowhurst et al. 2008; Fraser et al. 2009) allowed the production of the first genetic maps with low saturation in kiwifruit (Testolin et al. 2001; Fraser et al. 2009).

SNPs (Single Nucleotide Polymorphism) are the most abundant type of genetic marker and their high density makes them ideal for studying the inheritance of genomic regions (Berger et al. 2001). Currently, next-generation sequencing (NGS) technology provides an effective tool for MAS through generating huge number of DNA markers within a short period. The emergence of massively-parallel, NGS platforms capable of producing millions of short (50-250 bp) DNA sequence reads has reduced the costs of DNA sequencing and offers the tantalizing possibility of making affordable the genotyping-by sequencing (GBS) technology (Metzker 2010). Therefore, NGS-based marker systems allow highly efficient marker development for MAS in plant breeding. One such system is restriction site-associated DNA sequencing (RADseq), which detects polymorphic variants neighboring particular restriction enzyme recognition sites (Baird et al. 2008). RADseq has been used to detect SNPs in a variety of plant species, with or without an available reference genome (Chutimanitsakun et al. 2011). Compared to

existing RADseq approaches, in particular, a modified protocol called double digest RAD sequencing (ddRADseq), established by Peterson et al. (2012), permits greater flexibility and robustness in region recovery, and a substantial decrease in cost, required genomic material from samples and researcher time investment. Thus, all these advances lead to the construction of highly saturated maps to enable fine-scale genetic mapping and the anchoring of physical maps (Klein et al. 2000).

1.2.2.2 Mapping of Mendelian and quantitative traits

The rapid development of SNP markers through the NGS technologies makes possible the adoption of association studied based on linkage maps. Hence, once markers tightly linked to the traits being studied are found, they can be adopted for the rapid screening of progeny in MAS (Testolin and Cipriani 2016). In this thesis, it is in our interest to investigate traits associated to gender determination and fruit quality in kiwifruit.

Gender determination

Sex markers have been given particular emphasis because kiwifruit breeders are seldom interested in selecting males in their breeding programs and if male seedlings could be discarded at a very early stage through marker-assisted screening, breeders would save labour and space in the orchard (Testolin 2011). In kiwifruit sex-determining genes are localized in a pair of chromosomes, caryologically indistinguishable, that function like an XX/XY system with female homogametic (XX) and male heterogametic (XY) (Testolin et al 1995, Harvey et al. 1997). Gender is controlled by a single Mendelian determinant at any ploidy level (Testolin et al. 1995; Fraser and McNeilage 2016).

In the literature, there are several gender markers reported: SmX, Ke225, UDK096, and SmY (Gill et al. 1998; Fraser et al. 2009). All of them are close to the sex-determining region, which was mapped initially to the subtelomere of LG17 (Fraser et al. 2009). Ke225 is routinely used in breeding programs in New Zealand (http://www.plantandfood.co.nz) to discard male seedlings, but the primers have not been published. The others sex-linked markers were used in the past, but they were not robust and sometimes lose their polymorphism which limited their utility for breeding (Harvey et al. 1997; Gill et al. 1998). Testolin (2011) mentioned that it is difficult to find markers close to the sex determinant and it is also difficult to establish the order and the distances that separate markers, especially when mapping populations are small. According to Charlesworth (2013), sex determination genes often reside on sex

chromosomes, where the haplotypic region do not undergo recombination and this make the identification of genes that control the gender rather difficult to identify. For this reason, the saturation of existing genetic maps would make it possible to find markers with no apparent recombination with the sex determinant and this should be sufficient for more effective use of those markers in early screening for gender. Moreover, if sex-linked markers are successfully identified, they could also facilitate the dissection of the sex-determination system.

Fruit quality traits

Fruit ripening is a crucial physiological process for plants, since it represents the terminal stage of reproduction. During late ripening, physicochemical, biochemical and aromatic modifications occur. These changes influence the color, texture and flavor of the fruit (Beever and Hopkirk 1990). Ripe kiwifruit have a soft melting texture that, combined with distinctive flavors, aromas, and color, has made the fruit an international success. Understanding the genetic changes during ripening in kiwifruit is essential to the development of new kiwifruit cultivars with good postharvest characteristics, as long-storing fruit are relatively rare in *A. chinensis* var. *chinensis* and *A. chinensis* var. *deliciosa* germplasm material, and most other *Actinidia* species are characterizes by rapid softening and limited shelf life (Atkinson and Schröder 2016).

Large-scale expressed sequence tags (EST) sequencing programs have allowed identification of many genes expressed during fruit ripening (Crowhurst et al. 2008). Most recently, kiwifruit genome sequence has started reveal the allelic variation related to fruit development and ripening, fruit nutrient metabolism, and others (Huang et al. 2013).

However, despite the above milestones, few examples of markers associated with fruit quality are available in the literature. McNeilage et al. (2011) and Scaglione et al. (2015) reported some putative quantitative trait loci (QTL) associated with fruit quality, but, in comparison to other species, there is still a lot to do.

1.3 Outline of thesis

This work of thesis has been organized as follows:

In Chapter 2 genetic maps were produced using SSR from literature and SNPs markers obtained through a ddRAD approach of genotyping by sequencing (GBS) using Next Generation Sequencing (NGS) technologies. Additionally, new sex-linked SSR markers were isolated and tested in several crosses of the *A. chinensis/ A. deliciosa* complex at different ploidy level.

In Chapter 3 we report an exploratory search for QTLs in kiwifruit (*Actinidia chinensis*) using a small female progeny produced from the controlled cross 'C8' x 'A54.19', used for the production of the saturated genetic map. QTLs analysis were conducted for quality traits, such as ripening time, flesh and skin absorbance chlorophyll, fruit size, dry matter content, flesh firmness, pH, sugar and acid content.

In Chapter 4 best linear unbiased prediction (BLUP) and REML were computed for parent selection in a controlled mating design, involving 35 small cross families. This analysis include the estimation of genetic component of variance, heritability, the combining ability of cross parents and breeding value for several vegetative and fruit quality traits in kiwifruit.

In Chapter 5 male parent ('A54.19') of our mapping population was sequenced. Exploration of sex control region has started. Currently, this work is in process. We describe the metrics of the reads produced, a tentative automatic assembly and the future prospective for sex-region determination.

1.4 References

Atkinson RG, Schröder R (2016) Genetics of fruit softening. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 205–217

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:1–7. doi: 10.1371/journal.pone.0003376

Beever D, Hopkirk G (1990) Fruit development and fruit physiology. Kiwifruit Sci. Manage 97-126

Berger J, Suzuki T, Senti K-A, et al (2001) Genetic mapping with SNP markers in *Drosophila*. Nat Genet 29:475–481. doi: 10.1038/ng773

Charlesworth D (2013) Plant sex chromosome evolution. J Exp Bot 64:405–420. doi: 10.1093/jxb/ers322

Charlesworth D, Guttman DS (1999) Evolution of dioecy. In: Ainsworth CC (ed) Sex determination in plants. BIOS Scientific Publisher Limited, Oxford, pp 25-44

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, et al (2011) Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. BMC Genomics 12:1–13. doi: 10.1186/1471-2164-12-4

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196. doi: 10.1007/s10681-005-1681-5

Crowhurst RN, Gleave AP, MacRae E A, et al (2008) Analysis of expressed sequence tags from *Actinidia*: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. BMC Genomics 9:351. doi: 10.1186/1471-2164-9-351

Ferguson AR (2016a) World economic importance. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 37–42

Ferguson AR (2016b) Botanical description. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 1–13

Ferguson AR, Huang H (2016) Cytology, ploidy and ploidy manipulation. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 55–63

Ferguson AR (2007) The need for characterization and evaluation of germplasm: kiwifruit as an example. Euphytica 154(3):371-382.

Ferguson AR, Huang HW (2007) Genetic resources of kiwifruit: domestication and breeding. Hort Rev 33:1-121.

Ferguson AR, Seal AG, McNeilage MA, et al (1996) Kiwifruit. In: Janick J, Moore JN (eds) Fruit breeding. vol 2. Vine and small Fruits crops. J Wiley & Sons, New York, pp 371–417

Fraser LG, McNeilage MA (2016) Reproductive biology. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 65–84

Fraser LG, Tsang GK, Datson PM, et al (2009) A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X / Y sex-determining chromosomes. BMC Genomics 15:1–15. doi: 10.1186/1471-2164-10-102

Gill GP, Harvey CF, Gardner RC, Fraser LG (1998) Development of sex-linked PCR markers for gender identification in *Actinidia*. Theor Appl Genet 97:439–445. doi: 10.1007/s001220050914

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics 137:1121–1137. doi: 10.1007/s11033-010-0612-2

Harvey CF, Gill GP, Fraser LG, McNeilage MA (1997) Sex determination in *Actinidia*. 1. Sex-linked markers and progeny sex ratio in diploid *A. chinensis*. Sex Plant Reprod 10:149–154. doi: 10.1007/s004970050082

Huang H (2016) Genetic resources. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 15–36

Huang H-W, Zhong C-H, Jiang Z-W et al (2014) The genus *Actinidia*, a world monograph. Science Press, Beijing 317 pp

Huang S, Ding J, Deng D, et al (2013) Draft genome of the kiwifruit *Actinidia chinensis*. Nat Commun 4:1–9. doi: 10.1038/ncomms3640

Huang WG, Cipriani G, Morgante M, Testolin R (1998) Microsatellite DNA in *Actinidia chinensis*: isolation , characterisation , and homology in related species. Theor Appl Genet 97:1269–1278.

Kawagoe T, Suzuki N (2004) Cryptic dioecy in *Actinidia polygama*: a test of the pollinator attraction hypothesis. Can J Bot 82:214–218. doi: 10.1139/b03-150

Klein PE, Klein RR, Cartinhour SW, et al (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Res 10:789–807. doi: 10.1101/gr.10.6.789

Liang CF (1984) *Actinidia*. In: Feng KM (ed) Flora reipublicae popularis sinicae, vol 49. Science Press, Beijing, pp 196–208

Maliepaard C, Jansen J, Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet Res 70:237–250. doi: 10.1017/S0016672397003005

Marsh HD, Paterson T, Seal AG, Mcneilage MA (2003) Heritability estimates in kiwifruit. Acta Hortic 622:221–229.

McCough SR, Doerge RW (1995) QTL mapping in rice. Trends Genet 11:482–487. doi: 10.1016/S0168-9525(00)89157-X

McNeilage M, Fraser LG, Tsang P, et al (2011) Molecular genetics and genomics and kiwifruit breeding. Acta Hortic 913:63–70.

McNeilage M (1991) Gender variation in *Actinidia deliciosa*, the kiwifruit. Sex Plant Reprod 4: 267-273.

Metzker ML (2010) Sequencing technologies - the next generation. Nature 11:31–46. doi: 10.1038/nrg2626

Mohan M, Nair S, Bhagwat A, et al (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3:87–103. doi: 10.1023/A:1009651919792

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:1–11. doi: 10.1371/journal.pone.0037135

Ribaut J-M, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3:236–239. doi: 10.1016/S1360-1385(98)01240-0

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125:645–654.

Scaglione D, Fornasiero A, Cattonaro F, et al (2015) A RAD-based linkage map of kiwifruit (*Actinidia chinensis* Pl.) and map position of genetic determinants of traits of interest to breeders. Acta Hortic 1096:191–198.

Testolin R (2011) Kiwifruit breeding: from the phenotypic analysis of parents to the genomic estimation of their breeding value (GEBV). Acta Hortic 913:123–130.

Testolin R, Huang HW, Ferguson AR (eds) (2016) The kiwifruit genome. Series Compendium of Plant Genomes. Springer International Publishing AG 269 pp [ISBN 978-3-319-32272-8, DOI 10.1007/978-3-319-32274-2]

Testolin R, Cipriani G (2016) Markers, maps, and marker-assisted selection. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 85–99

Testolin R, Ferguson R (1997) Isozyme polymorphism in the genus Actinidia and the origin of the kiwifruit genome. Syst. Bot. 22:685-700.

Testolin R, Cipriani G, Costa G (1995) Sex segregation ratio and gender expression in the genus *Actinidia*. Sex Plant Reprod. doi: 10.1007/BF00242255

Testolin R, Cipriani G, Fiori G, et al (2011) Evaluation of parental genotypes as a tool for planning the mating design in kiwifruit breeding programs. Acta Hortic 913:213–220.

Testolin R, Cipriani G and Messina R (1999) Sex control in *Actinidia*. In: Ainsworth CC (ed) Sex determination in plants. BIOS Scientific Publisher Limited, Oxford, pp 137-146

Testolin R, Huang WG, Lain O, et al (2001) A kiwifruit (*Actinidia* spp.) linkage map based on microsatellites and integrated with AFLP markers. Theor Appl Genet 103:30–36.

Van Nocker S, Gardiner SE (2014) Breeding better cultivars, faster: applications of new technologies for the rapid deployment of superior horticultural tree crops. Hortic Res 1:14022. doi: 10.1038/hortres.2014.22

Zhang Q, Liu C, Liu Y, et al (2015) High-density interspecific genetic maps of kiwifruit and the identification of sex-specific markers. DNA Res 22:367–375. doi: 10.1093/dnares/dsv019

Manuscript published

by

Davide Scaglione[1], Alice Fornasiero[1], Catalina Pinto[2], Federica Cattonaro[1], Alessandro Spadotto[1], Rodrigo Infante[2], Claudio Meneses[3], Rachele Messina[4], Orietta Lain[4], Guido Cipriani[4], Raffaele Testolin[4]

in

Tree Genetics & Genomes (2015) 11:115

[1] IGA Technology Services, Parco Scientifico e Tecnologico, Udine, Italy

[2] Departamento de Producción Agrícola, Universidad de Chile, Santiago, Chile

[3] Centro de Biotecnología Vegetal, Universidad Andrés Bello, Santiago, Chile

[4] Dipartimento di Scienze AgroAlimentari, Ambientali e Animali, University of Udine, Udine, Italy

The online version of this article (doi:10.1007/s11295-015-0941-3) contains supplementary material

CHAPTER 2

A RAD-BASED LINKAGE MAP OF KIWIFRUIT (*ACTINIDIA CHINENSIS* PL.) AS A TOOL TO IMPROVE GENOME ASSEMBLY AND TO SCAN THE GENDER-CONTROLLING REGION

2.1 Abstract

Kiwifruit breeding still largely relies on phenotypic observation of cross progeny grown in the field to fruiting maturity, without any selection prior to the juvenility being overcome. Developing markers for the selection of traits of interest would greatly help breeders to rapidly screen breeding populations. With the aim of mapping several traits of interest in kiwifruit, a F1 population of diploid (2n=58) *Actinidia chinensis* was produced by combining parents with contrasting phenotypic traits. Ninety-four individuals were preliminarily analyzed to obtain a saturated genetic map based on 167 SSRs from the literature and 12,586 segregating restriction-site-associated DNA (RAD) loci obtained through an approach known as genotyping-by-sequencing (GBS) based on haplotype calling of SNP markers identified by a modified double digest restriction-associated DNA sequencing (ddRADseq) protocol as proposed by Peterson et al. (2012). To improve the accuracy of genotype calling, restriction-site-associated reads were aligned to the scaffolds of the recently published kiwifruit genome (Huang et al. 2013). This strategy provided genetic anchoring to 557 Mbp (90 %) of the assembly, helping also to anchor some 120 unmapped Mbp and to identify some mis-joined scaffolds. The analysis of the region controlling the dioecy in kiwifruit, spanning 16 scaffolds in the pseudomolecule 25 of the genome assembly (approximately 4.9 Mbp), with RAD markers that co-segregated with the gender determinant, allowed to sort out markers suitable for marker-assisted selection for the gender in the mapping population with successful extension to further controlled crosses having parents at different ploidy level and belonging to the *A. chinensis/A. deliciosa* complex.

Keywords: Genotyping-by-sequencing, Next-generation sequencing, single-nucleotide polymorphism, genetic map, marker-assisted selection

2.2 Introduction

Kiwifruit breeding is still based on phenotypic observation of traits in cross populations, but it should take great advantage from a selection based on molecular markers associated with traits of interest. Such an approach, popularly known as marker-assisted selection (or MAS), could speed up the progeny screening. Kiwifruit (*Actinidia chinensis*, diploid, 2n=58 and related diploid and polyploidy *Actinidia* species) is dioecious, and males, which represent 50 % of the progeny in any species and at any ploidy level, could be easily discarded by the analysis of a gender-associated molecular marker without waiting until the juvenility of seedlings is over (Testolin et al. 1995; Fraser et al. 2009; McNeilage et al. 2012). In recent papers, four molecular markers have been reported being suit- able for gender screening (Fraser et al 2009; McNeilage et al. 2012), but either primers were not reported or markers where monomorphic or did not produce PCR products in the population we tested.

Not only the gender but also any trait, no matter it is Mendelian or polygenic, can be early selected before it becomes phenotypically evident by the analysis of associated markers. Moreover, large marker collections well scattered along the genome would pave the way to the genomic estimation of breeding value (GEBV) that represents the frontier in the so-called genomic selection (Meuwissen et al. 2001; Kumar et al. 2012; McNeilage et al. 2012; Testolin 2012, 2013).

SNPs are the markers of choice for massive genotyping, given their ease of detection and low cost of screening, while simple sequence repeats (SSR) markers are ideal markers for the generation of consensus maps, relying on a much higher level of polymorphism in conserved loci, even in genotypes far related from those from which markers were isolated. Hundreds of kiwifruit SSR markers can be retrieved from the literature and EST database collections (Huang et al. 1998; Crowhurst et al. 2008; Fraser et al. 2009), while large-scale SNP detection requires massive sequencing and requires also a reference sequence to which to map the reads. Alternatively, SNP can be produced in the absence of a reference genome and, more interestingly, simultaneously discovered and mapped, leveraging next-generation sequencing (NGS) technologies coupled with reduced representation of individual genomes with restriction enzymes. These methods permit a fine-tuning over the abundance of fragments obtained by the digestion (Peterson et al. 2012) and thus providing adequate multiplexing for cost-effectiveness. A SNP-based genetic map of 4301 markers has been

recently produced through the SLAF-Seq approach with the aim to anchor the scaffolds in a kiwifruit genome sequencing project (Huang et al. 2013).

We adopted such an approach to saturate a SSR-based framework linkage map of diploid kiwifruit. As a side product, we also obtained valuable data points which can be used to further improve the current genome assembly (Huang et al. 2013), both in terms of anchoring-ordering and within-scaffold miss-assembly correction. Moreover, we were able to explore the region of the chromosome 25 where the gender-controlling locus maps and to identify markers suitable for marker-assisted selection for the gender in kiwifruit.

2.3 Material and methods

2.3.1 Plant material and sequencing

A F1 population ('C8' x 'A54.19') was produced using a pair of parents genetically and geographically unrelated of diploid (2n=58) *A. chinensis*. The female parent, coded 'C8', is a selection of Fruit & Tea Institute, Hubei province, China, collected in the Fang County in 1980 and carrying large fruits compared with most diploid accessions; the male parent, coded 'A54.19', is a male seedling introduced from the Beijing Botanical Garden. Ninety-four individuals grown in the open field to the maturity, when gender was recorded, were selected: 41 were female and 53 were male.

Parents and the 94 offsprings were used to produce a linkage according to the procedure previously described (Huang et al. 1998), whereas SNPs were produced adopting a custom double digest restriction-associated DNA sequencing (ddRADseq) protocol based on the method proposed by Peterson et al. (2012) using the *SphI* and *MboI* restriction enzymes to produce the DNA libraries of the two parents and the offsprings. DNA was extracted following a modified CTAB-chloroform as described by Cipriani and Morgante (1993), followed by a purification with Ampure XP beads (Agencourt, Beverly, MA, USA) with 1.8 volumes.

Briefly, for each sample, 200 ng of DNA was processed with 2U of both enzymes (NEB, Ipswich, MA, USA) at 37 °C for 2 h, followed by inactivation at 65 °C for 20 min, and purified with 1.5 volumes of Ampure XP beads (Beckman Coulter, Brea, CA, USA). Common (P2) and barcoded (P1) adapters (see ESM) were added to a ligation reaction, 3 and 1 pmol respectively, with 200 U of T4 DNA ligase and incubated at 23 °C for 2

h and 65 °C for 20 min. Samples were then pooled in 24-plex, concentrated, and run on low-melting 1.5 % agarose gel. Fragments were selected in the range of 300–450 bp and recovered with QIAquick Gel Extraction kit (Qiagen, Venlo, Netherlands). Enrichment PCR was performed with PCR primers to incorporate Illumina hybridization and sequencing sites along with index sequences for combinatorial multiplexing (ESM). Cycling parameter were set as follows: 95 °C (5 min), 10 cycles with 95 °C (30 s), 60 °C (30 s), and 72 °C (45 s), with final extension at 72 °C (2 min). Quality, quantity, and reproducibility of libraries were assessed on a Bioanalyzer instrument (DNA High Sensitivity chip). Sequencing was carried out on Illumina HiSeq 2500 platform (Illumina Inc, San Diego, CA, USA) following the manufacturer's protocol.

2.3.2 RAD sequences analysis

Raw Illumina reads were processed in order to de-multiplex samples on the basis of Illumina TruSeq index and custom inline barcodes. After removal of variable-length inline barcode sequence, all reads were trimmed to 85 bp. Alignment to the reference genome (Huang et al. 2013) was carried out using Bowtie2 aligner (Langmead and Salzberg 2012) with default parameters and processed with Stacks software (Catchen et al. 2011). Alleles where retained with a minimum of three reads of depth while intra-sample polymorphism was called using a bounded SNP model with maximum error likelihood of 0.5. A catalog of candidate segregating loci was populated by comparing parental haplotypes on the basis of their genomic coordinates. Reconstructed loci for each progeny individual were then matched against the former catalog in order to score segregating haplotypes. A minimum coverage of six reads was applied to generate homozygous calls. By leveraging a priori knowledge of SNP sites expected to segregate from parents, a correction to low-coverage alleles (i.e., below the minimum of three reads depth) was used to recover heterozygous calls in progeny individuals as implemented in the stacks software, while imposing a minimum allele frequency of 0.10. Haplotype calls were retained when at least 65 individuals out of the 94 in the progeny were genotyped and used to generate the final segregation scoring matrix.

2.3.3 Linkage analysis

The "two-ways pseudo-test cross" analysis allowed to produce a linkage map for each cross parent by managing the two parental meioses separately. A first analysis was conducted to obtain linkage groups for each parent and data cleanup. RAD markers were grouped using R/QTL package with limit of detection (LOD) score >10 and rec <0.20. Heat maps of LOD scores along with recombination fractions were manually inspected to determine *trans* states across groups and to switch markers accordingly, while singlets due to repetitive sequences or miss-alignments were removed. Markers ordering and correction of genotyping errors were achieved with two iterations of MSTmap software (Wu et al. 2008) with default parameters (using Kosambi function for distance estimation) followed by a Perl implementation of the SMOOTH program as described in Van Os et al. (2005). For each genetic bin, a single RAD marker was selected as representative and integrated to the SSR dataset. A final ordering state was calculated with the JoinMap (v4.0) program (Van Ooijen 2006; Stam 1993) using the maximum likelihood (ML) method. This last step was taken to reduce the conflicts in assigning the distances between markers due to the small size of the mapping population.

2.3.4 Analysis of the gender-controlling region and identification of new markers for the MAS

The 16 scaffolds of the pseudomolecule number 25 of the genome assembly (Huang et al. 2013), whose RAD markers co-segregated with the sex determinant in the mapping population, were analyzed, and many SSR-containing sequences were extracted using a modified version of the software Sputnik (Morgante et al. 2001). Primers were designed on the flanking regions using Primer3 (Rozen and Skaletsky 2000). Markers were preliminarily amplified in the parents and a few offsprings, and four of them that provided adequate amplificates and clear bands were then tested in several controlled crosses of the *A. chinensis/A. deliciosa* complex, with different ploidy levels (2n, 4n, and 6n) in the attempt of identifying gender-linked alleles suitable for the marker-assisted selection.

2.4 Results and discussion

2.4.1 RAD and SSR genotyping

A total of 540 million reads were produced on Illumina HiSeq2500 corresponding to some 270 M pairs. About 92 % of raw reads contained the expected restriction site over- hang on both pairs (*SphI* on forward read and *MboI* on reverse reads) along with discriminating inline barcodes. The average number of successfully de-multiplexed reads per sample was 2.8 M, with a standard deviation of 1.2 M, excluding parents. These latter ones obtained 7.8 and 16.3 M reads for the female and the male parent, respectively, by increasing their load to guarantee the coverage saturation of loci (Figure 1). Reads were all trimmed to a final length of 85 bp and aligned to the reference genome of *A. chinensis* recently published (Huang et al. 2013). As much as 93.9 and 95.7 % reads from female and male parent, respectively, were successfully mapped to the reference genome. An in silico prediction of double-digested fragments in the range of 220–370 bp on the available assembly was used to simulate a gel selection of 300–450 bp (having some 80 bp introduced by pre-PCR adapters ligation). This exercise led to the estimation of some 30,000 fragments, which would result in 60,000 RAD loci given that single-end read-stacks are analyzed independently. Figure 1 depicts an excess of loci collected on each individual sample, with parental lines counting up to 123,618 and 113,398. This phenomenon finds explanation in the fact that each individual can stochastically yield loci that are out of the target. This effect is also visible in progeny samples, where the yield of more reads directly reflects in the counting of more loci. However, by simply considering loci that are matching across a minimum set of 65 samples (i.e., they are truly enriched in the given size selection), we obtained some 56,700, which agrees with our prediction. Moreover, by analyzing the coverage distribution of the over-sequenced female sample, we calculate a weighted median coverage of 150× while all genomic coordinates with a coverage above 30× and 75× totalized for the 92.6 % and the 81.4 % of all the available reads. This indicated that our complexity reduction strategy was effective in sampling a well-defined set of genomic loci ensuring cost-effectiveness of the sequencing output and providing the possibility of accurate heterozygous versus homozygous calling.

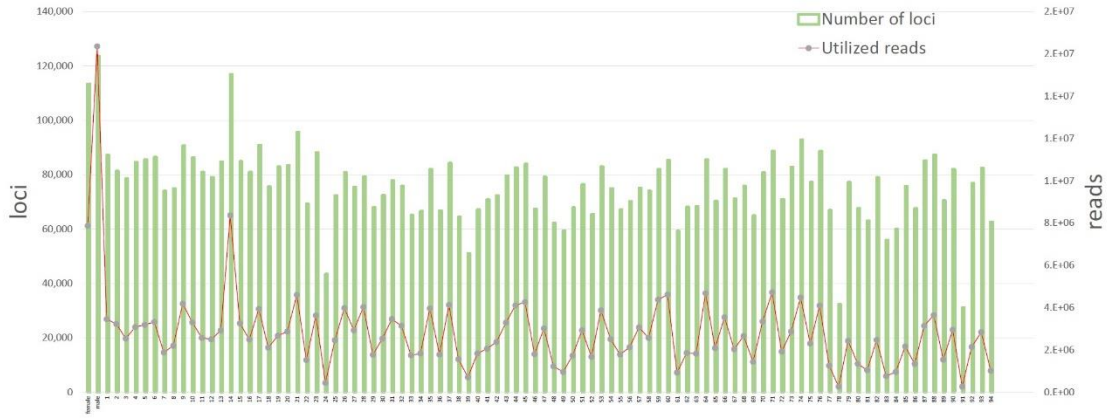


Figure 1. Distribution of sequenced reads across samples (red line, right axis) along with number of matched loci against the catalogue of polymorphic RAD sites as detected by parents analysis (green bar, left axis). Parents are the first two entries of the bar chart (female first).

A total of 73,993 unfiltered SNP sites were obtained with the analysis of parental genotypes over a set of 152,966 candidate genomic loci with 33,573 and 28,069 SNPs being identified in the female and male parents, respectively (Table 1). This was achieved by adopting filtering criteria such as minimum coverage and bounded SNP probability model (see Materials and methods). A catalog of candidate segregating RAD loci was generated for the two parents by means of called haplotypes (rather than each single SNP site), given the possible presence of more than one SNP site per RAD site. By matching offspring haplotypes against the populated catalog of polymorphic RAD sites and applying a minimum threshold of 65 successfully genotyped individuals, 12,586 RAD loci were found to be polymorphic and to segregate in a test cross fashion. Along with a set of 167 genotyped SSR markers, RAD loci were split according to the two-way pseudo-test cross design, collecting 6,347 and 6,470 informative markers for the female parent and the male one, respectively.

Table 1. Summary of parental SNP detection and statistics of *Actinidia* female and male linkage maps.

Parameters	unit	Female Parent	Male Parent
RAD fragment size	bp	220-370	220-370
Reads length	bp	85	85
Reads mapped to the reference genome	%	93.9	95.7
SNP identified	*n*	33,573	28,069
Candidate segregating loci	*n*	6,356	6,379
of which RAD haplotypes	*n*	6,214	6,365
of which SSR	*n*	142	114
Mapped loci	*n*	6,244	6,371
of which RAD haplotypes	*n*	6,112	6,262
of which SSR	*n*	132	108
Linkage groups	*n*	29	29
No. of markers in the map representation	*n*	1,390	1,429
of which bin RAD	*n*	1,260	1,322
of which SSR	*n*	130	107
Total map length	cM	3,614	3,276
Mean distance between genetic bins	cM	2.78	2.40
Mean marker density	cM	0.58	0.51

RAD restriction-site-associated DNA, SNP single-nucleotide polymorphism, SSR simple sequence repeats

2.5.2 Linkage analysis

Grouping of RAD markers with the R/QTL package provided 29 linkage groups for the female parent, totaling 6,112 markers; by side, the male map included 6,262 markers, with the same number of groups, corresponding to the haploid series of chromosomes of the diploid *A. chinensis*. The ordering procedure conducted with MSTmap provided a first framework to apply the correction/imputation routine on the large data frame. After correction and final ordering of bins along SSR markers (see Materials and methods), the female map consisted of 1,300 genetic bins, comprising 6,112 RAD loci, and 132 SSRs; the male map had 1,364 genetic bins, containing 6,262 RAD loci and 109 SSR markers (Table 1). SSR markers segregating from both parents along with shared polymorphic RAD loci were able to align the 29 linkage groups of the two parents.

Total map length was 3,614 cM for the female parent and 3,276 cM for the male one, with a mean distance between genetic bins of 2.78 and 2.40 cM for the female and male maps, while raw map densities of 0.58 and 0.51 cM, were reached, respectively (Table 1). As expected, the map length exceeded by far the length of previously published ones, which contained a more limited number of markers. For instance, the female and male maps published in 2009 by Fraser and coworkers (2009) contained 464 and 365 markers and had a length of 2,266 and 2,078 cM, respectively. Authors reported a theoretical estimate of 3,090 and 2,782 cM, by means of the method 3 as proposed by Chakravarti et al. (1991). Shorter linkage maps as compared to those reported in this paper are likely due to irregular marker distribution and large gaps which lead to underestimate the genetic distances. Vice versa, the map inflation at the increasing of markers number is known being mainly due to missing values and/or genotyping errors that reduce the proportion of correctly ordered markers and provide a less precise estimation of recombination distances (Hackett and Broadfoot 2003; Van Os et al. 2005). The map length inflation could be exacerbated if markers with no recombination are not removed from the set and when a consensus map is preferred to the individual parental maps (Ronin et al. 2012). These are likely the reasons for which the linkage map used to anchor scaffolds in the paper of kiwifruit genome assembly was 5,504 cM long (Huang et al. 2013). Obviously, in the last map, the correct calculation of genetic distances was sacrificed in favor of a better scaffold ordering.

Female and male maps, aligned through common SSR markers and scaffold coordinates of RAD markers, showed similar marker saturation with very small arrangements in the marker order. An exception, of which an explanation has not been found, is represented by chromosome 10, where the male linkage group is by far shorter than the female one with only 11 markers and a total length of 42.0 cM against the 52 markers and the 124.3 cM of the female linkage group. Skewed segregations were observed for a number of markers, but they were not analyzed in details, considering that they little affect the accuracy of maps.

SSR marker sequences together with scaffold coordinates of RAD markers allowed anchoring of all linkage groups to the chromosomal pseudomolecules of the assembly by Huang et al. (2013). This information was exploited as a proxy to assess reliability of ordering across the two parental maps by reciprocal alignment (Figure 2 and Appendix A). Genetic placement of scaffolds sequences across the two parental maps was able to anchor some 90 % of the current genome assembly as produced by Huang et al. (2013), corresponding to 557 Mbp. This resulted in the ability to anchor some

120 Mbp of previously unplaced scaffold sequence, raising the total amount of mapped sequence to 571 Mbp (92.6 % of the current assembly) from the 333.6 Mbp of the current release. Moreover, independent segregation analysis of the two parents ensured unbiased detection of false scaffold joining in the current version of the genome pseudomolecules. These misplaced scaffolds appear in our maps conflicting with their position in the genome assembly by means of several independent RAD segregations. Figure 3 depicts the male linkage group 16, corresponding to the pseudomecule Chr16 in Huang et al. (2013) where most of misplacements were found: The discrepancy that we recorded is that of several scaffolds, currently attributed to Chr10 unambiguously mapped to Chr16, for a total of 4.5 Mbp. Other events of scaffold misplacement were detected in Chr6, Chr16, Chr18, Chr19, Chr20, and Chr21 (Table 2).

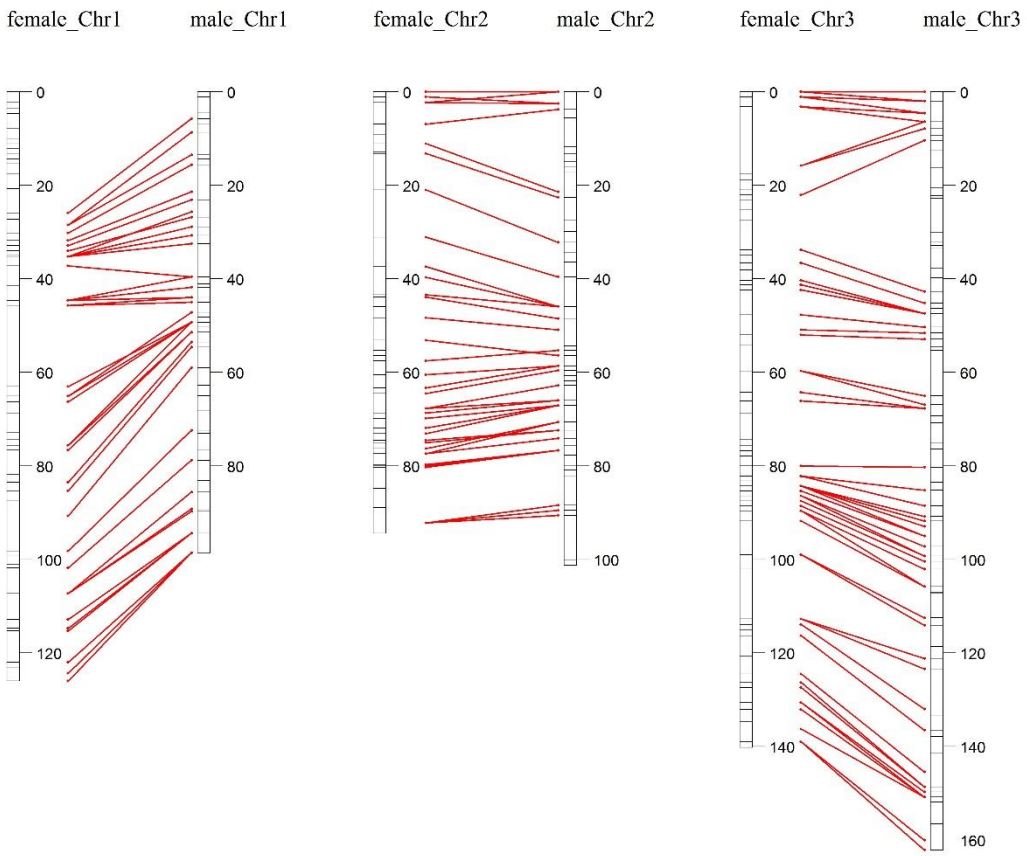


Figure 2. Alignment of parental maps for the first three chromosomes. RAD coordinates have been rounded to 10^4 bp to generate pseudo-loci for matching.

Table 2. Genome anchoring improvements for each chromosome.

LG name as from Huang *et al.* 2013	Detected scaffold* with misplacement	Previously unmapped sequence (bp)	Previously unmapped scaffolds
Chr1		2,184,434	20
Chr2		2,218,279	12
Chr3		4,233,834	30
Chr4		1,604,561	9
Chr5		3,555,725	29
Chr6	*From Chr23*: 212 (p)	2,660,972	20
Chr7		1,797,529	19
Chr8		4,028,411	15
Chr9		4,938,913	17
Chr10		10,836,296	36
Chr11		942,358	7
Chr12		5,605,708	22
Chr13		792,000	6
Chr14		6,156,229	20
Chr15		358,128	10
Chr16	*From Chr10*: 206, 438, 477, 803, 21, 1147, 252, 285	8,665,036	39
Chr17		3,323,865	21
Chr18	*From Chr11*: 54	5,978,979	38
Chr19	*From Chr6*: 418, 307	11,323,227	37
Chr20	*From Chr11*: 22 (p)	4,296,753	17
Chr21	*From Chr5*: 36 (p)	1,571,490	13
Chr22		7,419,277	20
Chr23		1,002,392	9
Chr24		688,414	6
Chr25		5,369,140	27
Chr26		3,133,093	25
Chr27		7,474,929	29
Chr28		1,260,193	10
Chr29		5,732,178	29
		119,152,343	592

*Scaffolds showing evidences of misplacement only for some RAD loci are reported with "p" as "partial" (i.e., possibly caused by erroneous scaffolding)

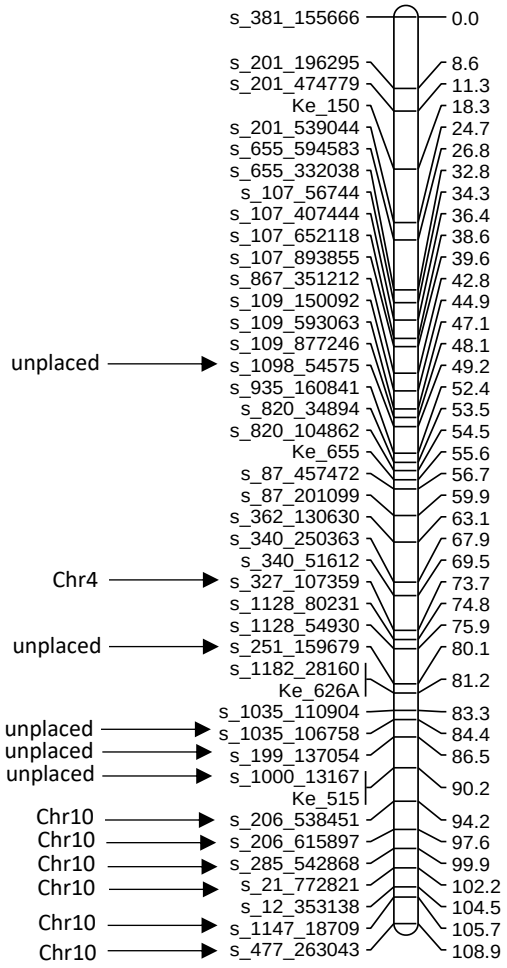


Figure 3. Major miss-assemblies were found in pseudomolecule "Chr 16". The picture depicts how RAD-based linkage map provided evidences of miss-anchored scaffolds in the current genome assembly as well as placement for un-positioned ones. On the left side of the LG the comments is highlighted the unplaced or miss-anchored scaffolds by means of a representative markers for each genetic bin. For miss-anchored scaffolds, the chromosomal pseudomolecule to which they were wrongly assigned is reported.

2.5.3 Mapping the gender determinant and development of associated molecular markers

Gender was recorded as heterozygous in male parent (XY) and homozygous in the female parent (XX), considering the model of genetic control of sex reported in the literature that is similar to that of *Silene*, with the male the heterogametic gender (Testolin et al. 1995). This Mendelian control of sex is maintained at any ploidy level (Testolin et al. 1995). Gender mapped in the linkage group 25 of the male parent, 2.166 cM from the top (Appendix B).

There are four markers suggested in the literature being associated to the gender, SmX, Ke225, UDK096, and SmY1, that span 4 cM in the region where the male phenotype maps (Fraser et al. 2009; McNeilage et al. 2012). We tested SmX, SmY1, and UDK096, and they were monomorphic or did not produce PCR products, although different annealing temperatures were assayed. Ke255 was not tested because primers were not published.

The analysis of the four SSR markers isolated from the scaffolds 514 and 776, whose RAD markers co-segregated with the sex determinant in the mapping population (Table 3), revealed a rather complex scenario (Table 4). Alleles in coupling with the sex determinant were identified for all four markers in the mapping population (C8 x A54.19). No recombinant was scored and, therefore, all four markers appeared suitable to screen the gender in the progeny, even in the cases in which the male-linked allele was a null allele (e.g., marker s_776B) or the same allele was present also in the female parent, e.g., marker s_776T (Table 4). In the latter case, the heterozygosity of the marker in the male parent helped to discriminate between male and female profiles in the progeny.

The analysis of marker segregation in crosses of higher ploidy level and in crosses of species that are different from which markers had been isolated increased the difficulty in scoring the profiles. Nevertheless, the alleles linked to the sex determinant were easily identified on the male parents and offsprings, except in the cases in which the sex-linked allele was not scorable because of the lack of amplification or in the case in which the allele likely coupled with the sex in the male parent had the same size of an allele carried by the female parent (Table 4).

Table 3. SSR markers extracted from the scaffolds 514 and 776 of the chromosome 25 of kiwifruit genome sequence (Huang et al. 2013), whose RAD markers co-segregated with the sex determinant.

Marker[a]	Core repeat	Primers	Primer 1st base in the scaffold	T anneal (°C)
s_514T	(GAA)13	F-ctggatcagcttctggact	65,435	55
		R-ggcaaaagatgaaaagagtg		
s_514B	(GTC)7	F-tctttgcgatcctattttgt	331,926	55
		R- tgcaaacagaaaaacaatga		
s_776T	(CTT)11	F-caatttgacccaagtaccac	15,320	55
		R-atggcaatcaatcactcaat		
s_776B	(AAT)7	F-tgatttgcttgctttatgaat	210,402	55
		R-tcggttttgtctgttttag		

[a] s-xxx stay for scaffold number, T=means toward the top, B=means towards the bottom of the scaffold

Table 4. Analysis of gender-association of SSR markers of Table 3 analysed in different kiwifruit cross populations

Marker	Species	Cross parents [a]	Parent allelic profile (female parent first)	Female progeny allelic profiles	Male progeny allelic profiles	Gender-linked allele	Recombinants/ no. offsprings
s_514T	*A. chinensis (2n)*	C8 x A54.19	220/220 x 212/217	212/220	217/220	217	0/94
	A. chinensis (4n)	C3 x male pool [b]	203/211/214 x [b]	203/211/214, 203/214, 211/214	203/209/214, 203/209/211, 209/211/214	209	0/30
	A. deliciosa (6n)	Koryoku x Matua	203/206/208 x 203/206/208/229	203/206/208, 203/206, 203/208	203/206/208/229, 203/206/229, 206/208/229	229	0/31
	A. deliciosa (6n)	Koryoku x C10	203/206/209 x 203/209/217/220	203/206/209/217/220, 203/209/217/220, 206/209/217/220, 203/209/220,	203/206/209/217/220, 203/209/217/220, 203/206/209/220, 203/209/217, 203/209/220	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	203/206/220 x 203/206/208/229	203/206/208, 203/206, 203/206/220	203/206/208/229, 203/206/229, 203/206/208/220/229	229	0/16
s_514B	*A. chinensis (2n)*	C8 x A54.19	167/188 x 217/229	167/229, 188/229	167/217, 188/217	217	0/94
	A. chinensis (4n)	C3 x male pool [c]	192/217/223 x [c]	Not scorable	Not scorable	223? [l]	- [l]
	A. deliciosa (6n)	Koryoku x Matua	210/226 x 210/2014/226	210/226	210/214/226, 214/226	214	0/30
	A. deliciosa (6n)	Koryoku x C10	210/226 x 217/223/226	209/217/222/225, 209/217/225, 217/222/225, 217/225	209/217/222/225, 209/217/225, 209/217/222, 209/217, 217/225	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	210 x 210/214	210	210/214	214	0/16

Marker	Species	Cross parents [a]	Parent allelic profile (female parent first)	Female progeny allelic profiles	Male progeny allelic profiles	Gender-linked allele	Recombinants /no. offsprings
s_776T	*A. chinensis (2n)*	C8 x A54.19	203/203 x 203/206	203/206	203/203	203 [f]	0/94
	A. chinensis (4n)	C3 x male pool [d]	188/196/199/202 x [d]	many	many	205	0/30
	A. deliciosa (6n)	Koryoku x Matua	182/188/194 x 176/188/194	182/188/194, 182/188	176/182/188/194, 176/182/188, 176/182	176	0/27
	A. deliciosa (6n)	Koryoku x C10	182/188/194 x 176/182/188/194	176/182/188, 176/188/194, 182/188/194, 182/188	176/182/188/194, 176/182/194, 182/188/194	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	182/194 x 176/182/194	182/194	176/182/194	176	0/16
s_776B	*A. chinensis (2n)*	C8 x A54.19	233/233 x 231/null	231/233	233/null	null [g]	0/94
	A. chinensis (4n)	C3 x male pool [e]	228/234 x [e]	not scorable	not scorable	-	-
	A. deliciosa (6n)	Koryoku x Matua	230 [h] x 230/234	230/234, 230	230/234, 230	null [i]	- [i]
	A. deliciosa (6n)	Koryoku x C10	119/231 x 119	119/231	119/231	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	230/234 x 230/234	230/234, 230, 234	230/234	null [i]	- [i]

[a] In crosses between polyploids (4n and 6n) multiple copies of alleles cannot be detected and therefore those occurrences are not marked in the profiles

[b] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 203/211/214 x 203/209

[c] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 192/217/223 x 223/226

[d] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 188/196/199/202 x 182/205/208

[e] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 228/234 x 228/234

[f] The presence of the allele 203 must be associated with the absence of the alternative allele 206 segregating from the male parent

[g] The presence of the null allele must be associated with the absence of the alternative allele 231 segregating from the male parent

[h] A further allele of 119 bp is present and has been considered belonging to a second locus

[i] Likely presence of a null not scorable allele

[l] The allele 223 likely linked to the sex in the male parent was present also in the female parent

2.6 Conclusions

The ddRADseq protocol for genotyping by sequencing, as described by Peterson et al. (2012), has shown to be a robust method for the identification and mapping of SNP markers in outcross species such as *A. chinensis*. The ability to obtain a controlled complexity reduction by means of the choice of two restriction enzymes coupled with size selection of fragments allowed to maintain adequate coverage across most RAD loci and thus yielding accurate heterozygous versus homozygous haplotype calling. Moreover, our results show that even in the presence of a reference genome, the genetic maps produced in this work can help to improve the genome assembly. Our maps can place about 120 Mbp of previously unanchored sequence of the genome assembly. Furthermore, its potential to correct some false joining that were originated in the first anchoring procedure of the genome assembly (Huang et al. 2013) was assessed.

The saturated map produced by merging ddRAD haplotypes and SSRs, with quite regularly spaced markers, easily helps to find genetic loci that control Mendelian traits like the gender and will help finding loci controlling quantitative traits (QTLs) in the extended population. The latter work is in progress.

2.7 References

Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics 1:171–182.

Chakravarti A, Lasher LK, Reefer JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. Genetics 128:175–182.

Cipriani G, Morgante M (1993) Evidence of chloroplast variation DNA variation in the genus *Actinidia* revealed by restriction analysis of PCR - amplified fragments. J Genet Breeding 47:319–326.

Crowhurst RN, Gleave AP, MacRae EA et al (2008) Analysis of expressed sequence tags from *Actinidia*: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. BMC Genomics 9:351.

Fraser LG, Tsang GK, Datson PM, De Silva HN et al (2009) A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X/Y sex-determining chromosomes. BMC Genomics 10:102. doi:10.1186/1471-2164-10-102

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity 90:33–38. doi:10.1038/sj.hdy.6800173

Huang WG, Cipriani G, Morgante M, Testolin R (1998) Microsatellite DNA in *Actinidia chinensis*: isolation, characterization, and homology in related species. Theor Appl Genet 97:1269–1278.

Huang S-X, Ding J, Deng D et al (2013) Draft genome of the kiwifruit *Actinidia chinensis*. Nat Communications 4:2640. doi:10.1038/ncomms3640

Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C (2012) Genomic selection for fruit quality traits in apple (*Malus x domestica* Borkh.). PLoS ONE 7(5):e36674. doi:10.1371/ journal.pone.0036674

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

McNeilage MA, Fraser LG, Tsang GK, Datson PM, De Silva HN, Crowhurst RN, Ferguson AR (2012) Molecular genetics and genomics and kiwifruit breeding. Acta Horticulturae 913:63–70.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Morgante M, Hanafey M, Powell W (2001) Microsatellites are preferentially associated with non repetitive DNA in plant genomes. Nat Genet 30:194–200.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7(5), e37135. doi:10.1371/journal.pone.003

Ronin Y, Mester D, Minkov D, Belotserkovski R, Jackson BN, Schnable PS, Aluru S, Korol A (2012) Two-phase analysis in consensus genetic mapping. G3 Genes Genomes Genetics 2(5):537–549. doi:10. 1534/g3.112.002428

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Methods Mol Biol, vol 132. Humana Press, Totowa, pp 365–386.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744.

Testolin R (2012) Breeding the future: what fruit breeders can learn from breeders of cows and chickens? Chronica Horticulturae 52(2):6–8.

Testolin R (2013) Kiwifruit breeding: from the phenotypic analysis of parents to the genomic estimation of their breeding value (GEBV). Acta Horticulturae 913:123–130.

Testolin R, Cipriani G, Costa G (1995) Sex segregation ratio and gender expression in the genus *Actinidia*. Sexual Plant Reprod 8:129–132.

Van Ooijen JW (2006) JoinMap® 4, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V, Wageningen

Van Os H, Stam P, Visser RGF, Eck HJ (2005) SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data. Theor Appl Genet 112:187–194.

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. Plos ONE. doi:10.1371/ journal.pgen.1000212

CHAPTER 3

EXPLORATORY QTL ANALYSIS FOR FRUIT-RELATED TRAITS IN KIWIFRUIT (*ACTINIDIA CHINENSIS* Pl.)

3.1 Abstract

Fruit quality is a key criterion used to select new cultivars in kiwifruit breeding programs. The aim of this chapter was to identify quantitative trait loci (QTLs) that control traits related to fruit quality in a F1 population derived from the cross of 'C8' (female) and 'A54.19' (male) diploid kiwifruit. A small female progeny was evaluated during 3 years for several traits: fruit weight, absorbance of chlorophyll (I_{AD}) values of the skin and flesh, flesh firmness, soluble solids content (SSC), titratable acidity (TA), pH, dry matter (DM) content and ripening time (RT). The existing genetic linkage maps of 'C8' and 'A54.19' discussed in the chapter 2 were used. Single or multiple QTL mapping models were applied separately for each year and all years combined. As a result, a total of 24 QTLs were detected on 12 different linkage groups (LGs). Some QTLs, such as flesh firmness, skin and flesh chlorophyll absorbance were reported for the first time in kiwifruit. However, QTLs were not consistent among years, and rarely identified in both parents in corresponding linkage groups. These results were expected due to the limited size of the progeny. This work was the first attempt to identify genomic regions that control fruit quality traits in the mapping cross population. The extension of the cross population to about 400 female individuals has been undertaken and it will guarantee a more reliable and robust identification of QTLs related to fruit quality in future analyses.

Keywords: Quantitative Trait Loci, pseudo-test cross, genetic linkage map, QTL mapping, female progeny, population size, massive screening

3.2 Introduction

Ripening represents a collection of physiological and biochemical processes that occur in fruit flesh, including but not limited to de-greening and accumulation of colored pigments different from chlorophylls, texture changes associated with cell wall metabolism and cell turgor variation leading to softening, and metabolic changes related to flavor, generally associated with accumulation of sugars, acids and volatiles, that culminate in a diverse array of tastes and smells (Giovannoni 2004). Attributes such as flesh firmness, sweetness, acidity, aroma and dry matter may affect the acceptance of kiwifruit. Thus, fruit quality is important for industry because it can affect the consumer preference (Crisosto et al. 2012).

Breeding for fruit quality traits is complex due to the expected polygenic nature of the genetic control of most of these traits. According to Murphy (2011), in several cases, complex traits are mostly regulated by one or a few major QTLs. Therefore, to improve the speed and efficiency of genotype selection, attention in recent years has focused on the identification of QTLs (Kenis et al. 2008). According to Broman (2001) and Mackay et al. (2009), the aim of QTL analysis is to identify genomic regions containing one or more candidate genes affecting the trait of interest.

Kiwifruit is one of the most recently domesticated fruit crops. The whole-genome sequence of *Actinidia chinensis* var. *chinensis* was published in 2013 by Huang et al. The publication of the kiwifruit genome has enabled the discovery of single nucleotide polymorphism (SNP) markers associated with major and minor genes (Cheng et al. 2016). Additionally, genotyping techniques such as genotyping-by-sequencing (GBS) or restriction-associated DNA (RAD) sequencing have allowed to identify allelic variants at relative low cost of a large number of coding sequences within the kiwifruit genome (Poland and Rife 2012; Hilario et al. 2015). Likewise, well-saturated maps can be easily obtained with the same new technologies (Zhang et al. 2015; Scaglione et al. 2015b). They provide a basis for fine mapping the positions of the major genes and QTLs, when the same mapping populations have been accurately phenotyped for traits of interest.

There are therefore several advantages to determine the inheritance of many qualitative and quantitative traits at the molecular level (Frett et al. 2014). Up to now few studies are available on QTL mapping in kiwifruit (McNeilage et al. 2011; Scaglione et al. 2015a). For this reason, the main objective of this chapter was to identify genetic regions associated with the most important quality fruit traits using the 'C8' x 'A54.19' mapping population discussed in the second chapter.

3.3 Material and methods

3.3.1 Plant material

The mapping population included ninety-four offsprings of the F1 progeny obtained from a pseudo-test cross between 'C8' (female parent) and 'A54.19' (male parent) of diploid kiwifruit (*Actinidia chinensis* Planch.). The cross was produced using parents genetically and geographically unrelated. Offsprings were grown in the open field at the University experimental farm in Udine (Italy). Once in reproductive maturity, gender of each offspring was recorded. The progeny resulted in: 41 females and 53 males, with a ratio slightly skewed toward an excess of males.

Parents and offspring were used to produce a linkage map based on SSR and SNP markers. However, only female individuals were observed for the phenotypic traits linked to fruit quality.

3.3.2 Fruit quality measurements

Fruit quality was evaluated for three seasons (2011, 2013 and 2014) using both nondestructive (fruit weight and the absorbance of chlorophyll (I_{AD}) of the skin) and destructive (I_{AD} of the flesh, flesh firmness, SSC) methods. Several variables such as titratable acidity, pH and dry matter were evaluated only in 2014.

Each season, at least ten randomly chosen fruit per genotype were collected weekly, starting about two weeks before the expected conventional time of harvest. Fruits were transferred to the lab, where the weight (g) per fruit was recorded and the I_{AD} was assessed in the equatorial zone of each cheek with a Kiwi-Meter (Sinteleia, Bologna, Italy). Yet, the I_{AD} of the flesh was also measured after removing the skin on each cheek. The SSC (°Brix) was determined by a thermo-compensated portable refractometer (Atago, Tokyo, Japan), and the flesh firmness (N) was measured with a hand-held penetrometer (Effegi, Verona, Italy) on each cheek after the skin removal, using a 7.9 diameter plunger with a penetration distance of 10 mm. Titratable acidity (TA, in g/100 mL), expressed as citric acid, was determined by titration of a mixture of the juice of three fruit with 0.1 N sodium hydroxide (NaOH) solution at pH 8.2. The pH of suspension was recorded before titration. Dry matter (average of three fruits) was measured by cutting two equatorial slices per fruit, 3-mm thick, and drying them at 82°C for 72 h. The fruit dry matter was calculated as a percentage of the initial wet

weight of the slices. Finally, harvest time was established when fruit sample reached 7-8 °Brix.

Table 5. List of abbreviations for fruit traits.

Abbreviation	Trait name	Unit
FW	Fruit weight	g
IS	I_{AD} of the fruit skin	I_{AD} unit
IF	I_{AD} of the flesh	I_{AD} unit
FF	Flesh firmness	N
SSC	Solid soluble content	°Brix
TA	Titratable acidity	%
PH	pH	-log[H^+]
DM	Dry matter	%
RT	Ripening time	d

It is important to precise that I_{AD} is an index based on fruit absorbance spectra measured by a vis/NIRs device, named Kiwi-meter. This index is related to the absorbance of chlorophyll (I_{AD}), calculated as the difference between A540–A750 in *Actinidia deliciosa* fruits and A640–A750 in *Actinidia chinensis* fruits, both near the peak of absorbance of chlorophyll-a (Costa et al. 2011). It has been proven that I_{AD} allows an indirect determination of the concentration of chlorophyll in the skin (Costa et al. 2015).

3.3.3 Data analysis

Statistical analysis of data was carried out to determine the variation, to test the distribution of data and to calculate correlation coefficients between the individual traits. All statistical analysis were carried out using the R software, version 3.1.2 (R Core Team, 2014). Histograms for each trait were constructed with all data set. Pearson's correlation coefficients among variables were calculated on the mean value for all years.

3.3.4 Genetic linkage maps

The genetic linkage maps of 'C8' and 'A54.19' discussed in the Chapter 2 (Scaglione et al. 2015b) were used for the QTL analysis. These genetic maps were produced using both genomic and EST-derived SSR markers from literature and SNPs produced with an approach of 'genotyping by sequencing' (GBS). The "two-ways pseudo-test cross" analysis resulted in a linkage map for each cross parent (Appendix A). Further details regarding genetic map construction can be found in Chapter 2. Linkage maps were drawn with the R/qtl package.

3.3.5 QTL analysis

QTL analysis was carried out using R/qtl package with the single QTL model (SQM) and restricted MQM procedures (Broman et al. 2003). Single Haley-Knott regression (Haley and Knott, 1992) was conducted for normalized traits and non-parametric method for non-normalized traits. The Haley–Knott (HK) regression method has been adopted because of the reduced computation need in comparison with the standard interval mapping (IM) described in their seminal article by Lander and Botstein (1989). The few drawbacks of the HK approach reported in the literature, such as the poorly performing in the presence of strong epistasis and/or association between QTLs, and non-normal distributions of data, were considered not determinant in the exploratory QTL analysis conducted in this work.

QTL analysis was performed separately for each year and with the overall mean (2011–2013-2014) using genetic linkage maps of each parent. The likelihood of the presence of a QTL was expressed as a log of odds (LOD) score. LOD significance thresholds were determined with the permutation test procedure; option settings included 1,000 permutations, and significance was set to p=0.05.

Multiple QTL regression was carried out with "stepwiseqtl" function. This approach uses forward/backward selection to identify a multiple-QTL model with inclusion of both main effect QTLs and pairwise interactions. Maximum QTL number was set to 5 for forward selection (max.qtl = 5). Model choice was made *via* a penalized LOD score (pLOD) which is the LOD score for the model (the log likelihood ratio comparing the full model to the null model without QTL) with penalties on the number of QTLs and pairwise QTL × QTL interactions. For each trait and year, specific penalties for main effect and digenic pairwise interaction terms were derived from 1,000 permutations of two-dimensional scan (the "scantwo" function, method = "hk", n.perm = 1,000) and

penalties at genome-wide error rate of 0.10 were used for multiple-QTL model fitting. The QTL model with the largest pLOD was identified as the most probable one. Once determined the multiple QTL model, we refined QTL position ("refineqtl" function) and estimate R^2 for the whole model and each term of the model, the individual LOD score of each term and the genotypic effect ("fitqtl" function). The "lodint" function was used to derive LOD-1 and LOD-2 QTL location confidence interval. Genome scan was performed with a 1 cM step.

3.4 Results and discussion

3.4.1 Phenotypic distribution and relation between traits

From the 94 F1 progeny used to construct the parental linkage maps, only 41 genotypes were female, which produced fruits on which measurements were taken. The distributions of all fruit quality traits evaluated are shown in Figure 4. Fruit weight (FW), flesh firmness (FF) and soluble solid content (SSC), presented a typical normal distribution. Ripening time (RT) and dry matter content (DM) showed a bimodal distribution. For skin chlorophyll absorbance (IS), flesh chlorophyll absorbance (IF) and titratable acidity (TA), a non-normal distribution was observed. In most cases of skewed distribution, the phenomenon could be attributed to the low sample sizes.

Notably, our small population exhibited considerable genotypic variation in traits such as FW, FF, SSC and RT.

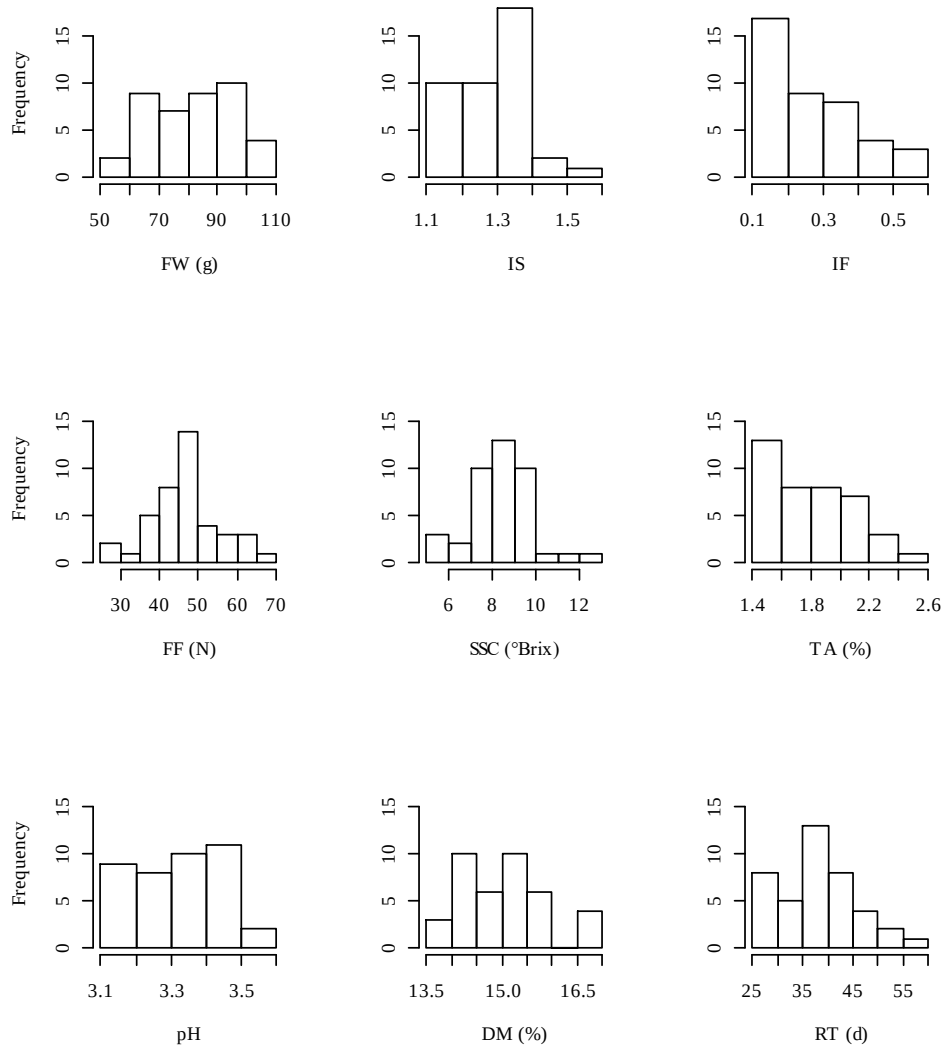


Figure 4. Frequency distributions of quality traits for the F1 female population. The distribution was calculated using one mean value per genotype averaged over three years, except for TA, PH and DM, which were evaluated only in 2014.

The correlation coefficients between the fruit traits are shown in Table 6. Correlation coefficients were calculated to determine the extent to which traits are correlated with each other because it has been noted that QTLs highly correlated often map to similar positions (Paterson et al. 1991). High positive correlations ($r > 0.60$) were found between IS, IF and FF. Conversely, SSC showed high negative correlations with traits such as IS, IF, FF, TA and RT. Likewise TA was negatively correlated with pH. Surprisingly, DM, an important quality index at harvest in kiwifruit (Crisosto et al. 2012), was the unique trait not correlated with any other variable.

Table 6. Correlation coefficients between fruit traits measured for 41 female genotypes in all years, except TA, pH and DM, which were evaluated only in 2014. The correlation coefficients were calculated with one mean value per genotype and year averaged over three to five replications.

Trait	FW	IS	IF	FF	SSC	TA	pH	DM	RT
FW									
IS	-0.24^{NS}								
IF	-0.37*	0.68***							
FF	-0.25^{NS}	0.61***	0.62***						
SSC	0.01^{NS}	-0.47**	-0.57***	-0.59***					
TA	-0.25^{NS}	0.13^{NS}	0.21^{NS}	0.15^{NS}	-0.42**				
pH	0.07^{NS}	0.14^{NS}	0.03^{NS}	0.02^{NS}	0.29^{NS}	-0.79***			
DM	-0.09^{NS}	0.16^{NS}	0.08^{NS}	0.29^{NS}	0.08^{NS}	-0.21^{NS}	0.30^{NS}		
RT	0.16^{NS}	0.27^{NS}	0.39*	0.38*	-0.88***	0.32*	-0.30^{NS}	-0.07^{NS}	

^{NS,*,**,***} Non significant or significant at $P \leq$ 0.05, 0.01, and 0.001, respectively

3.4.2 Genetic linkage maps

In this study, genetic maps used were produced through 'genotyping-by-sequencing' (GBS), where SNP markers were identified and mapped at the same time by sequencing all progeny of the cross and parent with next-generation sequencing (NGS) technologies (Peterson et al. 2012). These maps are highly saturated, as seen in Figure 5, with a total of 6,244 markers for the female parent and 6,371 markers for the male parent, and a mean distance between genetic bins of 2.78 and 2.40 cM for the female and male maps, respectively (see Chapter 2) (Scaglione et al. 2015b; Testolin and Cipriani 2016).

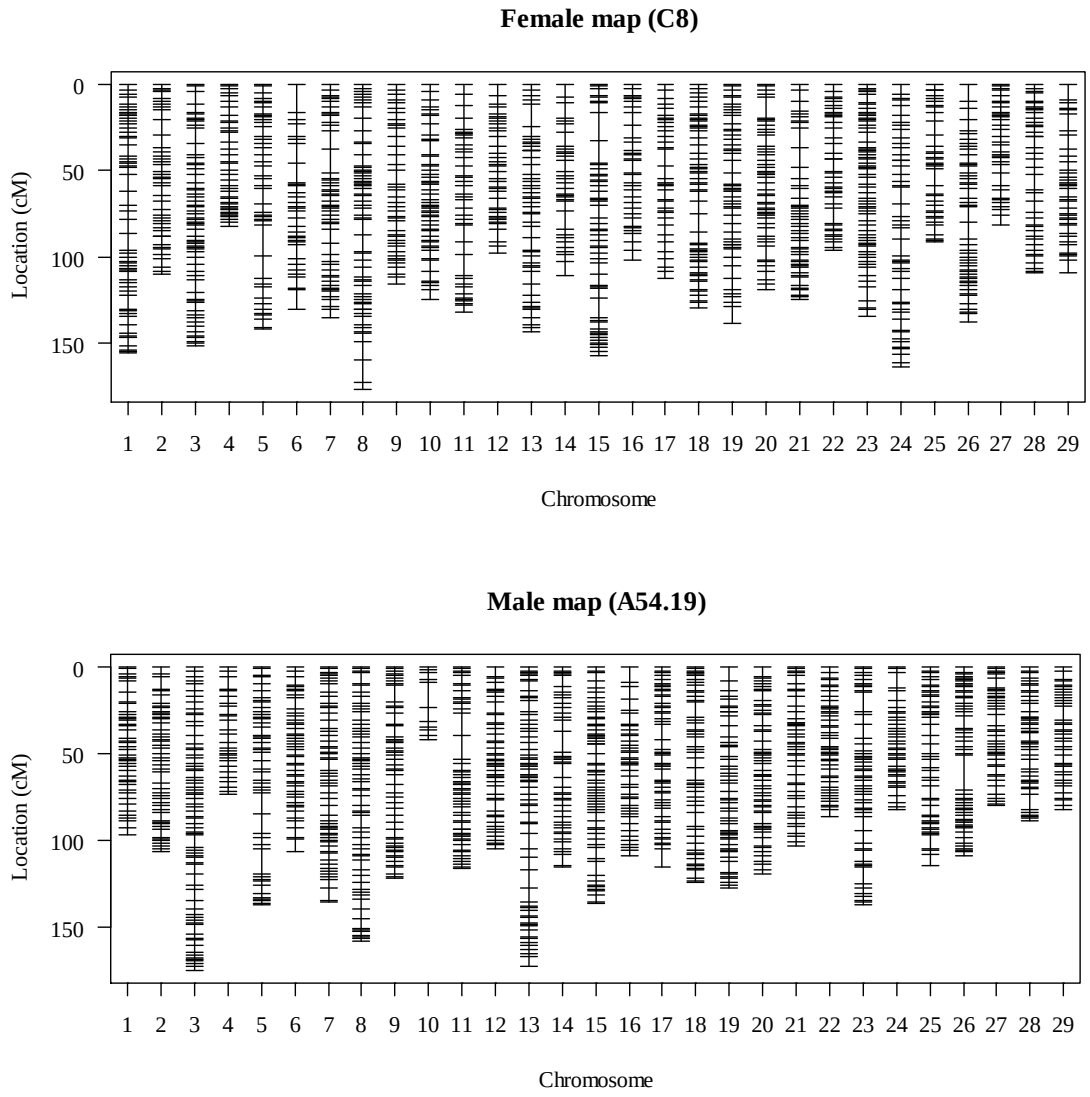


Figure 5. Genetic linkage maps of female ('C8') and male ('A54.19') parent in kiwifruit displaying 29 linkage groups.

3.4.3 Detection of QTLs

QTL analyses were carried out using fruit from the female offsprings of 'C8' x 'A54.19' population. Despite the small population size, considering the only fruit-bearing vines, we found 24 significant associations between markers and fruit traits (Table 7). Among them, 14 QTLs were mapped to the female map ('C8') and 10 to the male map ('A54.19'). None of the QTL detected were identified in both parents. The explained phenotypic variance ranged from 12.64 % for FF (flesh firmness) in 2011 (FF-2011) to 41.32 % for TA (titratable acidity) (TA-2014). The highest LOD score was 9.1 for IS considering the overall mean of years (IS-mean). DM (dry matter content) was the only

trait that did not show significant associations with markers of the maps. All significant QTLs were described below in detail based on the LGs where they were mapped (Figure 6).

For FW, a non-repeatable QTL was identified at the top of LG27 in male parent. McNeilage et al. (2011) and (Scaglione et al. 2015a) also identified a putative QTL for FW in male parent, but on a different LG.

Previously, a consistent association was indicated between IS (skin chlorophyll absorbance) and SSC (soluble solid content) in the same chromosomal region. The QTL analysis confirmed this result by showing an overlap of IS-2013, IS-mean and SSC-2013 at the bottom of LG6 in female parent (Figure 6). These co-located QTLs could correspond to distinct, closely linked genes or to a unique gene with a pleiotropic effect on several traits influenced by the same physiological process (Quilot et al. 2004). Interestingly, a further small QTL cluster was detected at the bottom of LG21 in male parent. In this case, the same location of QTLs for both IS (skin chlorophyll absorbance) and IF (flesh chlorophyll absorbance) was an expected result. These traits are indirect estimators of the concentration of chlorophyll in the fruit and both reflect the same process of chlorophyll degradation (Pilkington et al. 2012). Similarly, IS (skin chlorophyll absorbance) and RT (ripening time) overlapped on LG1 of the female parent, and this was largely expected on the basis of the measured correlation coefficients. Lastly for IS (skin chlorophyll absorbance), others two major QTLs on LG4 and one QTL on LG8 were detected in the female parent.

For FF (flesh firmness) we found a total of 7 QTLs considering both maps. Two major QTLs were detected at the bottom of LG11 of female parent, corresponding to 41.01 % and 37.23 % of phenotypic variance explained, respectively. Conversely, two QTLs on LG5, and a single QTL on LG7, LG9 and LG21 were identified in male parent. Others works have reported of the presence of several QTLs related to FF in apple and pear (Liebhard et al. 2003; Yamamoto et al. 2014).

In addition, we have reported QTLs associated with TA (titratable acidity) and pH in separate LGs. For TA, one major QTL was detected on LG6 in male parent. According to the literature a strong QTL for TA was located on LG3 in kiwifruit (Marsh and Harker 2016), but this could not be in contrast with our finding, considering than more than one genetic determinant could control such a quantitative trait. With regard to pH, two QTLs were identified in LGs 7 and 29, both mapped in the female parent, which explained 32.12 % and 27.86 % of the phenotypic variation, respectively.

Finally, for RT (ripening time) two QTLs were mapped on LG18 and LG1. The last, as mentioned above, co-located with IS (skin chlorophyll absorbance).

Table 7. Overview of QTSs controlling fruit quality traits detected in the segregating progeny of the C8 x A.54.19 cross, with results listed per linkage group (LG). For each QTL, the significance (LOD score) is presented first, followed by the value for the percentage of population variance explained by that QTL.

Code	Map	Year	LG1	LG4	LG5	LG6	LG7	LG8	LG9	LG11	LG18	LG21	LG27	LG29
FW	A54.19	2014											3.4/32.19	
IS	C8	2013		7.4/35.43		5.0/20.72						5.8/24.81		
	C8	Mean	7.0/26.89	9.1/39.64		4.1/13.34		4.5/14.63						
	A54.19	2014										4.1/37.54		
	A54.19	Mean										3.6/33.16		
IF	A54.19	2013										3.5/34.90		
FF	C8	2014								5.9/41.01				
										4.8/37.23				
	A54.19	2011			3.6/12.64		3.7/12.80		5.1/19.47			8.1/37.29		
	A54.19	Mean			3.0/28.31									
SSC	C8	2013				3.1/31.24								
TA	A54.19	2014				4.6/41.32								
PH	C8	2014					5.4/32.12							4.9/27.86
RT	C8	2014	4.5/34.19								4.0/29.31			

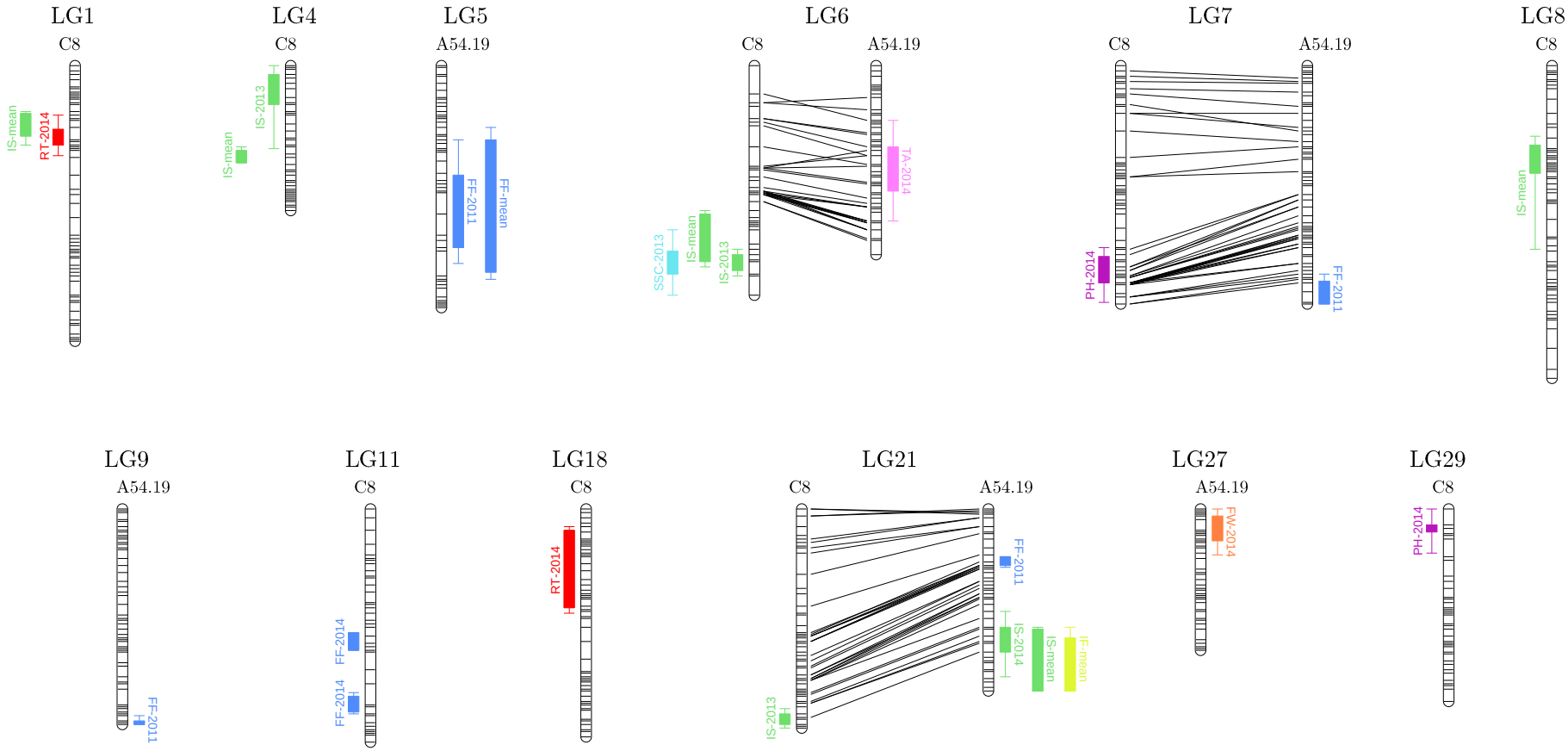


Figure 6. Location of putative QTLs identified for fruit quality traits in genetic linkage maps of *Actinidia chinensis* 'C8' and 'A54.19'. Significant QTLs are shown on the left side of LGs for 'C8' (female parent) and on the right side for 'A54.19' (male parent). Boxes and range lines indicate 1-LOD and 2-LOD support intervals, respectively. The QTLs detected with the mean were also represented. The trait codes are described in Table 5. Common markers on corresponding LGs are connected with a black line.

3.4.4 Limitations in QTL detection

In the literature there are few available paper on markers associated with traits of interest for breeders in kiwifruit (a review is available in Testolin and Cipriani 2016). This work reports a preliminary QTL analysis for fruit quality traits using the saturated maps produced by merging ddRAD SNPs and SSRs markers (Scaglione et al. 2015b) in *Actinidia chinensis*.

The results confirm the quantitative nature of most traits analyzed because generally more than one QTL was detected per trait in different linkage groups. Furthermore, many QTLs associated to fruit quality traits had large effects in terms of explained variance. However, most putative QTL loci showed poor stability from year to year and most of them were not identified in both parents.

Of course, we are aware that the weakness of this QTL analysis is due to the small size of the female population, and therefore, care should be taken when interpreting the significance of these results. According to Vales et al. (2005), the limited population sizes used in QTLs detection experiments can lead to underestimate the QTL number, to overestimate the effects of individual QTLs, and the combined result could be the failure to correct quantify a number of QTLs. Nevertheless, small population sizes in bi-parental mapping populations are commonly found in fruit crop literature, due to the cost of marker genotyping and the cost of trait phenotyping (Vales et al. 2005; Wang et al. 2012).

To localize stable QTLs we need more individuals in which recombination has occurred in the vicinity of the QTL so that only markers very close to the QTL on the chromosome remain linked to it (Mackay 2009). Therefore, the next step of this research must be the increase of the cross population size to obtain the necessary recombinations (Würschum 2012).

3.4.5 Extending population and future prospects

Consequently, one thousand seeds obtained from the same cross of the mapping population were produced and grown in greenhouse. Then, a total of 910 seedlings were screened using sex-linked SSR markers, specifically s_514T and s_776T that we identified as the suitable ones for the early gender screening (see Chapter 2). These markers were able to identify 457 male and 422 female individuals among the offspring. The rest of seedlings, 31 individuals, could not be identified due to lack of amplification

and/or the poor quality of DNA sample. As a result of this screening, all female plants were transplanted at the experimental farm of the University of Udine in May of the present year. In three more years, we expect to have the first phenotyping for fruit traits in a population of further 400 female individuals, which would represent a more adequate population size for QTL mapping at a higher resolution and reliable statistical confidence (Collard et al. 2005).

Future research prospects, beyond of this thesis, can be divided into three stages. In the first stage, we would map QTLs affecting the traits of interest with a more robust approach. In the second stage, we will focus on each QTL region to further narrow the genomic intervals containing the gene or genes affecting variation in the trait. The final stage is most challenging: to pinpoint the candidate genes (Mackay 2009). Independently from the last goal, the identification of robust markers closely linked to each QTLs will open the opportunity for a marker-assisted selection (MAS) for important fruit quality traits in kiwifruit breeding.

3.5 Conclusions

In this chapter, we report an exploratory search for QTLs dealing with fruit quality in kiwifruit (*Actinidia chinensis*) using a small female progeny produced from the controlled cross 'C8' x 'A54.19', genotyped at high saturation thorough an approach known as 'genotype-by-sequencing' that allowed the production of a linkage map based on RAD/SNP markers and integrated with SSR markers (see chapter 2).

Although the progeny size was limited, we detected 24 QTLs across 3 years of evaluation using the SQM and MQM mapping strategies. The co-localization and clustering of the some of the detected QTLs might indicate that these genes are tightly linked or pleiotropic effects may be involved. Undoubtedly, a more complete analysis is necessary with a larger population size. Our extended population will permit to validate these results, to detect new QTLs, and to have a more precise location of QTL in the genetic maps.

In conclusion, we hope that the future research will facilitate the development of new kiwifruit cultivars that bear high quality fruit, with a selection based on the modern and efficient marker-assisted selection (MAS).

3.6 References

Broman KW (2001) Review of statistical methods for QTL mapping in experimental crosses. Lab Anim (NY) 30:44–52. doi: 11469113

Broman KW, Wu H, Sen S, et al (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19(7):889–890. doi: 10.1093/ bioinformatics/btg112

Cheng C-H, Datson PM, Hanley Z (2016) Genome-based breeding. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 249–257

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169–196. doi: 10.1007/s10681-005-1681-5

Costa G, Bonora E, Fiori G, Noferini M (2011) Innovative non-destructive device for fruit quality assessment. Acta Hortic 913:575–582.

Costa G, Vidoni S, Rocchi L, et al (2015) Innovative non-destructive device for fruit quality assessment and early disease diagnosis. Acta Hortic 1096:69–78. doi: 10.17660/ActaHortic.2011.913.78

Crisosto GM, Hasey J, Zegbe JA, Crisosto CH (2012) New quality index based on dry matter and acidity proposed for Hayward kiwifruit. Calif Agric 66:70–75. doi: 10.3733/ca.v066n02p70

Frett TJ, Reighard GL, Okie WR, Gasic K (2014) Mapping quantitative trait loci associated with blush in peach [*Prunus persica* (L.) Batsch]. Tree Genet Genomes 10:367–381. doi: 10.1007/s11295-013-0692-y

Giovannoni J (2004) Genetic regulation of fruit development and ripening. Plant Cell 16:170–180.

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69(4): 315-24. doi:10.1038/hdy.1992.131

Hilario E, Barron L, Deng CH, et al (2015) Random tagging genotyping by sequencing (rtGBS), an unbiased approach to locate restriction enzyme sites across the target genome. PLoS One 10:1–15. doi: 10.1371/journal.pone.0143193

Huang S, Ding J, Deng D, et al (2013) Draft genome of the kiwifruit *Actinidia chinensis*. Nat Commun 4:1–9. doi: 10.1038/ncomms3640

Kenis K, Keulemans J, Davey MW (2008) Identification and stability of QTLs for fruit quality traits in apple. Tree Genet Genomes 4:647–661. doi: 10.1007/s11295-008-0140-6

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Liebhard R, Kellerhals M, Pfammatter W, et al (2003) Mapping quantitative physiological traits in apple (*Malus* × *domestica* Borkh.). Plant Mol Biol 52:511–526. doi: 10.1023/A:1024886500979

Mackay TFC (2009) Q&A: Genetic analysis of quantitative traits. J Biol 8:23. doi: 10.1186/jbiol133

Mackay TFC, Stone E a, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577. doi: 10.1038/nrg2612

Marsh KB, Harker FR (2016) Acid metabolism in kiwifruit. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 179–188

McNeilage M, Fraser LG, Tsang P, et al (2011) Molecular genetics and genomics and kiwifruit breeding. Acta Hortic 913:63–70.

Murphy DJ (2011) Plants, biotechnology and agriculture. CABI, pp 310.

Paterson AH, Damon S, Hewitt JD, et al (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197.

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7:1–11. doi: 10.1371/journal.pone.0037135

Pilkington SM, Montefiori M, Jameson PE, Allan AC (2012) The control of chlorophyll levels in maturing kiwifruit. Planta 236:1615–1628. doi: 10.1007/s00425-012-1723-x

Poland J a, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102. doi: DOI 10.3835/plantgenome2012.05.0005

Quilot B, Wu BH, Kervella J, et al (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana.* Theor Appl Genet 109:884–897. doi: 10.1007/s00122-004-1703-z

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Scaglione D, Fornasiero A, Cattonaro F, et al (2015a) A RAD-based linkage map of kiwifruit (*Actinidia chinensis* Pl.) and map position of genetic determinants of traits of interest to breeders. Acta Hortic 1096:191–198.

Scaglione D, Fornasiero A, Pinto C, et al (2015b) A RAD-based linkage map of kiwifruit (*Actinidia chinensis* Pl.) as a tool to improve the genome assembly and to scan the genomic region of the gender determinant for the marker-assisted breeding. Tree Genet Genomes. doi: 10.1007/s11295-015-0941-3

Testolin R, Cipriani G (2016) Markers, maps, and marker-assisted selection. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 85–99

Vales MI, Schön CC, Capettini F, et al (2005) Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor Appl Genet 111:1260–1270. doi: 10.1007/s00122-005-0043-y

Wang H, Smith KP, Combs E, et al (2012) Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. Theor Appl Genet 124:111–124. doi: 10.1007/s00122-011-1691-8

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210. doi: 10.1007/s00122-012-1887-6

Yamamoto T, Terakami S, Takada N, et al (2014) Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). Breed Sci 64:351. doi: 10.1270/jsbbs.64.351

Zhang Q, Liu C, Liu Y, et al (2015) High-density interspecific genetic maps of kiwifruit and the identification of sex-specific markers. DNA Res 22:367–375. doi: 10.1093/dnares/dsv019

CHAPTER 4

GENETIC VARIANCE AND BREEDING VALUES FOR PHENOTYPIC SELECTION IN KIWIFRUIT ESTIMATED BY EXPLORING MIXED MODELS AND PEDIGREE INFORMATION

4.1 Abstract

A mixed linear model approach was used to analyze kiwifruit breeding population and estimate variance components, heritabilities and breeding values (BV) for ten traits: time of bud break and flowering, number of flowers per cyme, flowering intensity, productivity, fruit weight, soluble solids content, flesh firmness and flesh color. The objectives of this study were to estimate genetic parameters and to identify superior parents, especially males for fruit-related traits. In order to improve accuracy of results, estimations were carried out using best linear unbiased prediction (BLUP) and residual maximum likelihood (REML). Pedigree information and temporal correlation were also accounted in the model. Results showed low heritabilities (≤ 0.20) for time of flowering, number of flowers per cyme, flowering intensity, productivity and yield, and moderate to high heritabilities (≥ 0.30) for time of bud break, fruit weight, soluble solid content, flesh firmness and flesh color. The contribution of parents to progeny for several traits was determined and selection of best male parents and cross families was successfully undertaken. Thus, these results will help to make better decisions in the selection among offsprings and in the selection of parents in the following breeding programs.

Keywords: Best linear unbiased predictions (BLUP), Residual maximum likelihood (REML), heritability, general combining ability (GCA), specific combining ability (SCA)

4.2 Introduction

The challenging objective of plant breeding programs is to improve crop performance through the development of new varieties. This requires the previous selection of best individuals to be used as parents in mating designs for traits of interest. To achieve the optimal selections it is important that the statistical methods used for analyzing data from these breeding trials, are as accurate and efficient as possible (Smith et al. 2005; De Faveri 2013).

The best linear unbiased prediction (BLUP) methodology has the property to account for selection of parents in a breeding population through the estimation of random effects of a mixed model (Robinson 1991). This method was originally developed in animal breeding for the estimation of breeding value of individuals (Henderson 1975). In contrast to the other procedures, BLUP makes use of all information on individuals for which records of phenotypic values and pedigree are available (Henderson 1976). Therefore, taking into account all the complex relationships among the relatives makes it possible to predict more accurately the genetic effects (Durel et al. 1998; Souza et al. 2000). Additionally, BLUP permits the use of flexible variance-covariance structures for genotype-by-environment interaction (Piepho et al. 2008). Thus, spatial variation typically present in the field and temporal correlation due to the repeated measures over the seasons can be included in the estimation of non-genetic effects (Smith et al. 2007).

Despite the advantages mentioned above, in plant breeding this procedure has not yet been used widely as it was in animal breeding. In particular, fruit tree crops present several experimental difficulties for the breeder like long generations, large size seedlings and methods for analyzing unbalanced data not yet well developed (Durel et al. 1998; De Souza et al. 2000).

In kiwifruit (*Actinidia* spp.), which is dioecious and crosses are generally done between female and male individuals, the selection of parents for crossing based on progeny testing is particularly valuable in the case of males, as they do not provide breeders with any phenotypic information on their genetic assets as far as fruit-controlling genes are concerned (Marsh et al. 1999; Testolin et al. 2011). Therefore, methods of estimating genetic parameters that efficiently make use of the information available, such as BLUP, can be extremely useful.

The objectives of the present chapter were first to estimate the genetic variance components useful in describing the heritability of a number of quantitative traits. Secondly, we wanted to rank parents, mainly male parents used to produce our experimental breeding populations for their capability of transferring traits of agronomic interest to their progenies (i.e. their breeding value).

4.3 Material and methods

4.3.1 Progeny population and pedigree

The present study was based on cross families of a kiwifruit breeding program carried out at the University of Udine, Italy. This program was designed to create new cultivars with high fruit quality. Three female (Lushan, Jintao and C9) and twelve male genotypes were selected as primary parents (Table 8). The mating design produced 35 cross families. The family sizes were unbalanced varying from 50 to 142 vines/cross. In total, 3023 female and male vines were established in an experimental field where the data were collected.

The thirty five families were analyzed taking account of the known relationships among parents and offsprings. The simple pedigree of each male parent involved was reconstructed. Female parents and some male accessions employed had unknown parentage (Table 8).

4.3.2 Phenotypic data measurement

The following attributes were measured during two consecutive seasons (2008 and 2009): 1) time of bud break, 2) time of flowering, 3) number of flowers per cyme, 4) flowering intensity, 5) productivity, 6) yield, 7) fruit weight, 8) soluble solids content (SSC), 9) flesh firmness, and 10) flesh color.

The time of bud break and the time of flowering were estimated by visual inspection once a week for each female and male vine, starting from 1st march and 1st may, respectively. At the appropriate time during the growing season, observations on the number of floral cymes, flowers, and fruits were made. The flowering intensity was rated for each vine at bloom time on six point scale: very high, high, medium, low, very low, and without flowers. In the same way, productivity per vine was evaluated in five categories, ranging from very high to very low. Yield (kg/vine) was estimated

starting from the total number of fruits per vine and the average fruit weight calculated over a sample of ten hand-pollinated fruit. The SSC (°Brix) was measured with a thermo-compensated refractometer (Atago, Tokyo, Japan). The flesh firmness (N) was determined with a hand-held penetrometer (Effegi, Verona, Italy) on each fruit cheek after the skin removal, using a 7.9 diameter plunger. Six categories were used for scoring flesh color: light green, green, yellow green, light yellow, yellow and intense yellow following a color chart based on flesh photos (Appendix C).

4.3.3 Statistical analyses

A simple mixed linear model can be written, in matrix notation, as

$$y = X\beta + Zu + Zv + e \tag{1}$$

where y is the vector of observations; β is a vector of fixed effect; u and v are vectors of random genetic and non-genetic effects, respectively; X and Z are the associated design matrices; and e is a random residual vector.

Because kiwifruit is dioecious, the following mixed linear model was employed in these analyses to incorporate effects of paternal and maternal parents and their interactions:

$$y = X\beta + Z_f u_f + Z_m u_m + Z_{fm} v_{fm} + Z_{fy} u_{fy} + Z_{my} u_{my} + Z_{fmy} v_{fmy} + e \tag{2}$$

where f denotes the random effect contributed by the maternal parent (the general combining ability of the genes in the crosses), and m the effect contributed by the paternal parent; fm is the interaction of paternal and maternal parents (specific combining ability of the genes in specific biparental combinations); fy is the interaction of maternal parent and year, my is the interaction of paternal parent and year; fmy is the interaction of both cross parents and year.

The random effects in the model were assumed to follow a multivariate distribution with means and variances defines by

$$\begin{bmatrix} u_f \\ u_m \\ v_{fm} \\ u_{fy} \\ u_{my} \\ v_{fmy} \\ e \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} A\sigma_f^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & A\sigma_m^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_c\sigma_{fm}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_y \otimes A\sigma_{fy}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_y \otimes A\sigma_{my}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_y \otimes \sigma_{fmy}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & R' \end{bmatrix} \right) \tag{3}$$

where 0 is a null matrix; A is the numerator relationship matrix which describes the additive genetic relationships among parents (Henderson 1976); I_c and I_y are identity matrices with order equal to c (the number of families), and y is the number of years, respectively; σ_f^2 and σ_m^2 are the general combining ability variance of female and male parent respectively, σ_{fm}^2 is the specific combining ability variance and R' is the matrix of errors.

Due to the unequal number of individuals and the missing parental combinations, there was a degree of unbalance within and between families. In order to estimate variance components and rank the parents, restricted maximum likelihood (REML) and the best linear unbiased predictors (BLUPs) were computed. The software used was ASReml (Butler 2009) in the R environment (R Core Team, 2014).

In this study common assumptions were applied to estimate genetic parameters. According to Falconer and Mackay (1996) the heritability in the narrow sense (h^2) determines the degree of average resemblance between relatives, and is defined as the ratio of additive genetic variance to phenotypic variance. Therefore, h^2 can be estimated as

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2} \tag{4}$$

The formula can be detailed in the variance components as follows

$$h^2 = \frac{2\left(\sigma_f^2+\sigma_m^2\right)}{\sigma_f^2+\sigma_m^2+\sigma_{fm}^2+\sigma_{fy}^2+\sigma_{my}^2+\sigma_{fmy}^2+\sigma_e^2} \tag{5}$$

In all what follows in this chapter, the term heritability is taken in the narrow-sense. The standard error method according to Kempthorne (1957) was used to indicate the precision of the estimates.

Table 8. Description of families, primary and secondary parents employed in the kiwifruit mating design.

Family	No. of offsprings	Primary parents[a]			Secondary parents[b]
		Female parent		Male parent	Paternal
Ac0358	132	Lushan	x	M3	-
Ac0359	121	Lushan	x	A0125.93	-
Ac0360	92	Lushan	x	Belén	-
Ac0361	114	Lushan	x	Ac0162.33	(A0136 x Zuva)[c]
Ac0362	73	Lushan	x	Ac0163.35	(A0136 x Belén)
Ac0363	112	Lushan	x	Ac0167.02	(A0136 x M3)
Ac0364	79	Lushan	x	Ac0168.32	(A0137 x A0125.46)
Ac0365	81	Lushan	x	Ac0171.01	(A0137 x A0134.41)
Ac0366	111	Lushan	x	Ac0174.46	(A0137 x M3)
Ac0367	146	Lushan	x	Ac0176.17	(A0145 x Zuva)
Ac0368	128	Lushan	x	Ac0178.15	(A0145 x A0134.41)
Ac0369	92	Lushan	x	Ac0181.21	(A0145 x M3)
Ac0370	68	Jintao	x	M3	-
Ac0371	50	Jintao	x	A0125.93	-
Ac0372	60	Jintao	x	Belén	-
Ac0373	67	Jintao	x	Ac0162.33	(A0136 x Zuva)
Ac0374	54	Jintao	x	Ac0163.35	(A0136 x Belén)
Ac0375	56	Jintao	x	Ac0167.02	(A0136 x M3)
Ac0376	0	Jintao	x	Ac0168.32	(A0137 x A0125.46)
Ac0377	73	Jintao	x	Ac0171.01	(A0137 x A0134.41)
Ac0378	78	Jintao	x	Ac0174.46	(A0137 x M3)
Ac0379	79	Jintao	x	Ac0176.17	(A0145 x Zuva)
Ac0380	62	Jintao	x	Ac0178.15	(A0145 x A0134.41)
Ac0381	72	Jintao	x	Ac0181.21	(A0145 x M3)
Ac0394	71	C9	x	M3	-
Ac0395	85	C9	x	A0125.93	-
Ac0396	90	C9	x	Belén	-
Ac0397	76	C9	x	Ac0162.33	(A0136 x Zuva)
Ac0398	87	C9	x	Ac0163.35	(A0136 x Belén)
Ac0399	123	C9	x	Ac0167.02	(A0136 x M3)
Ac0400	66	C9	x	Ac0168.32	(A0137 x A0125.46)
Ac0401	76	C9	x	Ac0171.01	(A0137 x A0134.41)
Ac0402	96	C9	x	Ac0174.46	(A0137 x M3)
Ac0403	80	C9	x	Ac0176.17	(A0145 x Zuva)
Ac0404	92	C9	x	Ac0178.15	(A0145 x A0134.41)
Ac0405	81	C9	x	Ac0181.21	(A0145 x M3)

[a] *Actinidia* derived from crosses and introduced as accessions are labelled with prefixes Ac and A, respectively

[b] Secondary parents are only known for male primary parents

[c] In each cross, the female parent is written first

4.4 Results and discussion

The present data set contained several drawbacks concerning the experimental design, the data and the scored traits. As described in the Material and methods, the controlled crosses were not designed for such an analysis of genetic parameters, because of the unbalanced cross family size and the failure of one of the programmed crosses (Ac0376, Jintao x Ac0168.32). The families were distributed in a single row in the same experimental field and not randomized, which is unfavorable for genetic analysis. Nevertheless the introduction of the pedigree information into the genetic model was a favorable compensation. In addition, the continuous observations on the same plant (experimental unit) during two consecutive seasons introduced correlations among data (Piepho et al. 2004), which were suitably accounted in the statistical analysis.

Table 9 shows the values of variance components estimated from the full data set for nine characters. Some traits have relatively low genetic component. In general, in all traits male general combining ability (GCA) effects were more important (lower p-values) than female GCA effects, except for soluble solid content (SSC) (data not shown).

Table 9. Variance components estimation for breeding population.

Trait	σ^2_m	σ^2_f	σ^2_{fm}	σ^2_{my}	σ^2_{fy}	σ^2_{fmy}	σ^2_e
Time of bud break	0.3650	1.0058	0.4525	4.30E^{-7}	0.5273	1.5982	3.2990
Time of flowering	2.16E^{-7}	0.1930	0.1181	2.16E^{-7}	2.16E^{-7}	0.2093	2.1353
Flowers per cyme	1.53E^{-3}	0.0021	0.0006	0.0004	0.0002	0.0002	0.0515
Flowering intensity	2.50E^{-7}	0.0023	0.0104	2.50E^{-7}	0.0102	0.0353	2.4703
Productivity	0.1000	1.72E^{-7}	0.0883	0.0483	1.72E^{-7}	1.72E^{-7}	1.7040
Yield	0.0215	6.88E^{-8}	0.0346	0.0350	6.88E^{-8}	0.0005	0.6795
Fruit weight[a]	0.0023	0.0009	0.0021	0.0002	1.43E^{-9}	0.0005	0.0142
Soluble Solid Content (SSC)	0.1414	1.3347	0.7266	0.0844	0.1906	0.0334	5.8481
Flesh firmness[b]	0.0815	1.4265	0.0289	1.54E^{-7}	0.1647	0.1075	1.5228
Flesh color	0.1589	0.1935	0.1415	-	-	-	1.4836

[a] Fruit weight (g) with a logarithmic transformation

[b] Flesh firmness (N) with a square root transformation

Heritability values were calculated and are listed in Table 10. The standard of Hansche et al. (1966) was used in this study to examine the heritability level - namely, very high (above 0.8), moderate to high (0.3-0.8), and low (below 0.3). Table 10 reveals that

for our set of data, time of bud break, fruit weight, SSC and color have moderate heritability, indicating that for these traits, the selection of superior seedlings as parent in successive crosses could result in rapid genetic improvement to subsequent generations (Daoyu et al. 2002). According to Durel et al. (1998), heritabilities values as low as 0.3-0.4 are still rather favorable and should guarantee the efficiency of mass selection in the field.

In contrast to other results found in the literature (Marsh et al. 2003; Testolin et al. 2011), flesh firmness had a very high heritability in this study. This high estimate of heritability is likely due to the harvesting time, that was selected when ripening time was enough advanced to display large differences either between and within cross families.

On the other hand, the time of flowering, number of flowers per cyme, productivity, yield and flowering intensity showed low heritabilities (Table 10). Results in time of flowering, number of flowers per cyme and yield are consistent with previous studies in kiwifruit (Beatson 1991; Testolin et al. 1995; Daoyu et al. 2002; Testolin et al. 2011). The others traits, that include productivity and flowering intensity, involved in this study were traits assessed on a discrete scale which is less precise than a continuous quantitative assessment, especially because of the subjectivity of observer (Durel et al. 1998). Casual variation in the trait assessment cannot be completely avoided, especially from year to year. Such background effects reduced the heritability estimation, as it can be observed in flowering intensity ($h^2 = 0.002$).

Table 10. Heritability estimates for several traits of interest for the breeder in kiwifruit.

Trait	Heritability (h^2)	s.e.
Time of bud break	0.378	0.076
Time of flowering	0.145	0.038
Flowers per cyme	0.128	0.037
Flowering intensity	0.002	0.007
Productivity	0.103	0.037
Yield	0.051	0.023
Fruit weight	0.318	0.072
Soluble solid content (SSC)	0.353	0.082
Flesh firmness	0.905	0.123
Flesh color	0.356	0.077

It is important to note that the heritability of a trait under selection, estimated from the plant material available, is of great interest to the breeder because it influences the decision on the selection strategy to be adopted (Testolin et al. 2011). In this case, moderate to high heritabilities allow phenotypic selection based on single individuals, whereas for traits with low heritability (below 0.20) it is advisable to proceed with a combined selection either between families and within families, to increase the chance of a significant gain in selection for the character under selection (Beatson 1991).

By including pedigree information, it was possible to predict the breeding values of genotypes where no data was available, in particular, primary and secondary parents of the breeding population. In order to show the most relevant results, all male parents are ranked in Table 11 by their predicted breeding values for several traits.

The best male parents for transmitting high yield were Ac0168.32, A0125.46 and Ac0171.01, while the poorest ones were M3, Ac0167.02 and Ac0174.46. In the same way the best parents for high fruit weight were Ac0171.01, Ac0181.21 and Ac0174.46, and the worst ones were Belén, Ac0163.35 and Zuva (Table 11).

However, the separate analysis of fruit weight and yield per vine does not helps the breeder in carrying out the selection of the best parents, considering that the two variables are usually negative correlated to each other. In kiwifruit, as in other fruit trees species, fruit weight is reduced by the increasing of number of fruits (Richardson and McAneney, 1990). An example of this negative relationship can be clearly observed in the results of Ac0174.46. Consequently, we consider to analyze both traits together for a better understanding of the results. Figure 7 shows the relationship between the predicted values of fruit weight and yield from mixed models of all male parents. As a result, the best male parents for transmitting high yield and fruit weight were Ac0168.32, Ac0171.01 and Ac0181.21 (evidenced in red on the right top quadrant of the Cartesian coordinates).

SSC and flesh firmness were characters evaluated at a fixed time, as mentioned previously, and then, there was high variability among individuals due to the differences in ripening time. For this reason, we consider the breeding values of these traits as indirect estimates of fruit ripening time. Generally speaking, in kiwifruit the two traits are in agreement with each other and low firmness corresponds to higher sugar content, indicating fruit closer to maturity, while high firmness corresponds to lower sugar content, would indicate fruit that are still immature (Testolin et al. 2011). Comparison among male parents of our breeding population showed that Ac0163.35, A0134.41, M3

and Belén induce an early maturation in the offsprings (Table 11). Conversely, Ac0168.32, A0125.46 and Ac0176.17 induce a later maturation.

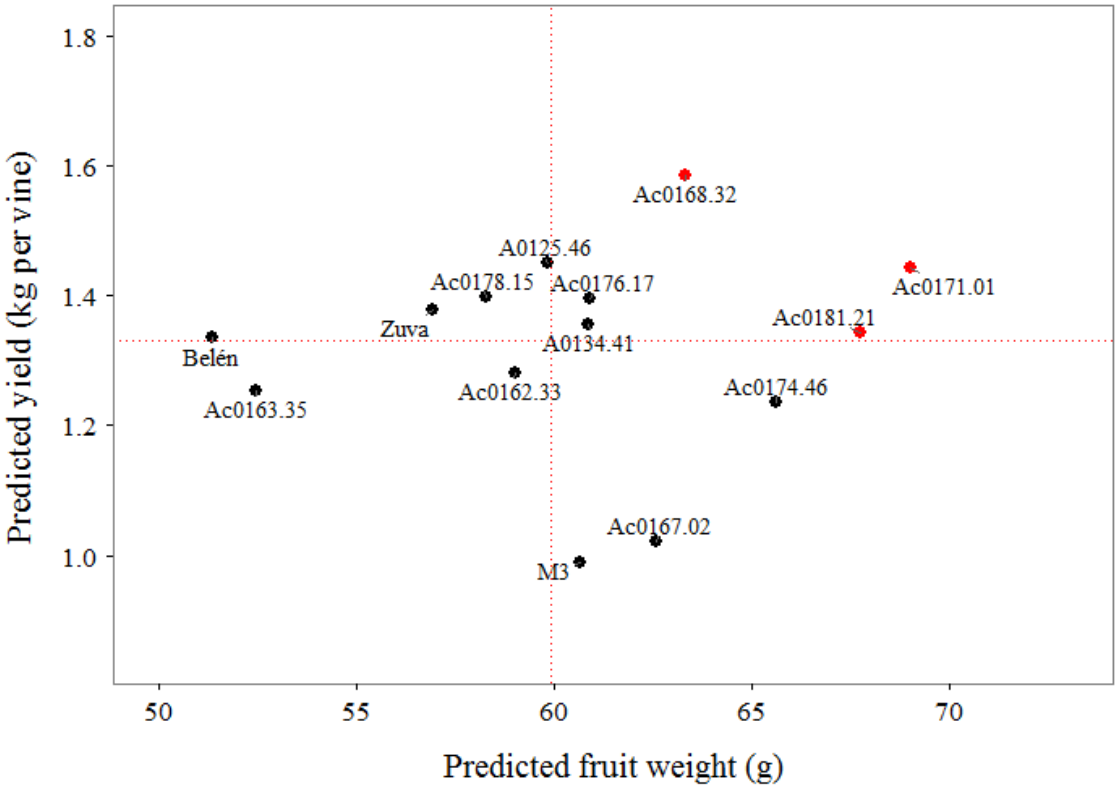


Figure 7. Relationship between predicted values of average fruit weight (g) and yield (kg per vine) of male parents. Red lines indicate mean value of each variable. Red points in the right top quadrant evidence the best genotypes.

Considering the female parents, no relevant differences were found in yield, but Lushan and C9 showed additive variance for high fruit weight. For ripening time, C9 and the accessions A0136, A0137 and A0145 evidenced an earlier maturation in comparison to Lushan and Jintao (data not shown). These data confirm in some way the phenotypic values directly observed in the parental genotype. Jintao produces fruit of small size compared with Lushan and C9 that have larger fruit sizes (Ferguson et al 2012). C9 is the most precocious in fruit ripening compared to all other female selections of the study.

Selecting the best families for cultivar development appears to be a reasonable strategy to maximize the probability of finding good genotypes (Davik and Honne 2005). Among

the 35 families evaluated in this program, the predicted performance varied between traits. The five poorest and the five best cross families for some characters are presented in Table 12, indicating a huge variation among families.

Once again, we present the relationship between fruit weight and yield, but in this case for families (Figure 8). For both traits among the top five ranked families (evidenced in red), three have parents that are ranked with high breeding values. C9 and Ac0163.35, both parents of the family Ac0398, however, presented low breeding values, providing a good example of the importance of the specific combing ability (SCA). Another example can be seen in the family Ac0396 which is also among the highest ranked, while its parents, C9 and Belén, had low breeding values for the same traits.

Regarding to ripening time, early and later families were also identified (Table 12). Although Iwanami et al. (2008) mention that ripening date is not usually a trait highly selectable in fruit breeding, early, middle, and later ripening cultivars are all useful and desirable for growers because cultivars with various ripening dates can help enlarging the harvest period. Moreover, this is a key information for appropriate management and efficient data collection in a large breeding population.

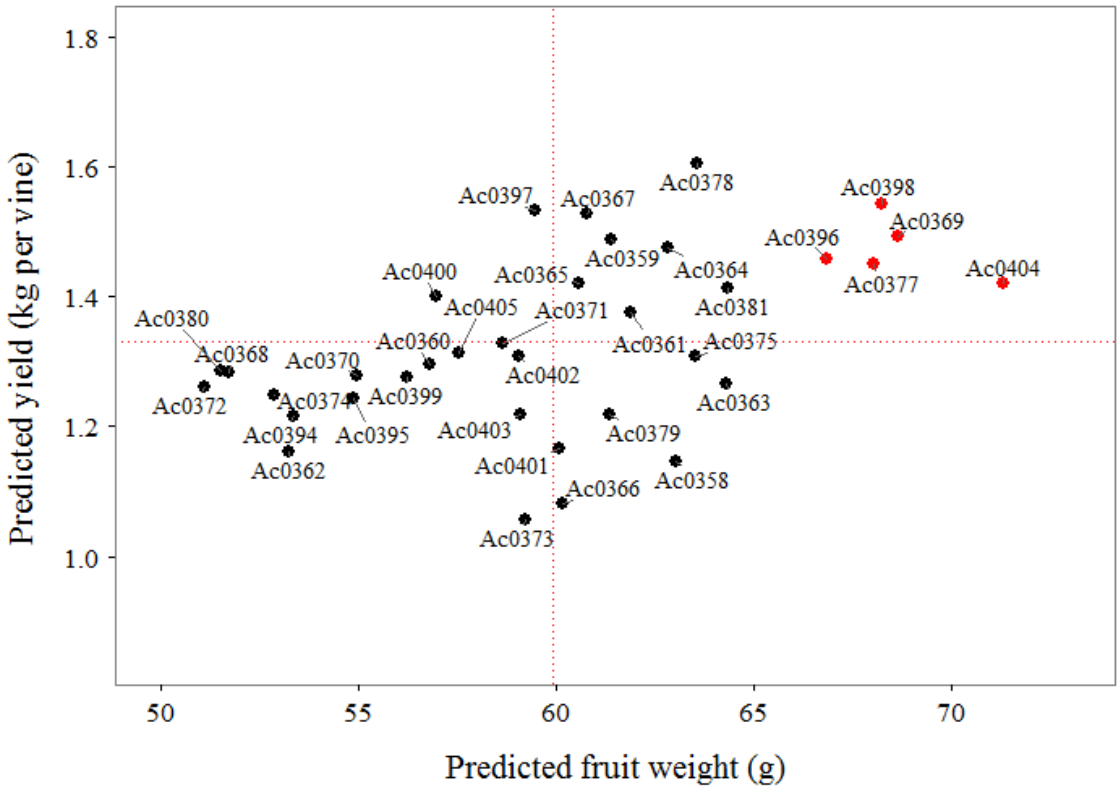


Figure 8. Relationship between predicted values of average fruit weight (g) and yield (kg per vine) of families. Red lines indicate mean value of each variable. Red points in the right top quadrant evidence the best families.

Table 11. Predicted breeding values (BV), standard error of the predictions (SEP) and rank of male parents for yield, fruit weight, SSC and flesh firmness.

Male parent	Yield (kg per vine)			Fruit weight (g)			SSC (°Brix)			Flesh firmness (N)		
	BV	SEP	Rank	Rank	BV	SEP	BV	SEP	Rank	Rank	BV	SEP
Ac0168.32	0.255	0.121	1	4	2.531	1.081	-0.522	0.789	14	14	5.072	0.537
A0125.46[a]	0.121	0.192	2	9	-0.968	1.123	-0.282	0.830	13	12	2.675	0.582
Ac0171.01	0.113	0.112	3	1	8.213	1.074	0.165	0.775	5	5	-1.258	0.532
Ac0178.15	0.068	0.109	4	11	-2.518	1.073	0.250	0.774	3	11	0.758	0.530
Ac0176.17	0.065	0.099	5	6	0.129	1.070	-0.090	0.765	11	13	3.039	0.525
Zuva	0.048	0.101	6	12	-3.854	1.070	-0.181	0.766	12	10	-0.162	0.526
A0134.41[a]	0.026	0.171	7	7	0.088	1.110	0.251	0.810	2	4	-1.260	0.566
Ac0181.21	0.014	0.101	8	2	6.945	1.071	0.006	0.765	8	6	-0.934	0.526
Belén	0.006	0.104	9	14	-9.451	1.072	0.234	0.772	4	3	-2.056	0.528
Ac0162.33	-0.049	0.111	10	10	-1.792	1.073	0.008	0.771	7	9	-0.284	0.531
Ac0163.35	-0.075	0.105	11	13	-8.351	1.071	0.260	0.769	1	2	-3.020	0.528
Ac0174.46	-0.093	0.103	12	3	4.832	1.071	-0.060	0.766	9	7	-0.863	0.527
Ac0167.02	-0.307	0.126	13	5	1.794	1.078	-0.081	0.777	10	8	-0.445	0.538
M3	-0.340	0.104	14	8	-0.151	1.070	0.148	0.764	6	1	-3.849	0.527

[a] Male secondary parents

Table 12. Predicted breeding values (BV) and standard error of the predictions (SEP) of the five top ranked and five bottom ranked families for yield, fruit weight, SSC and flesh firmness.

	Yield (kg per vine)			Fruit weight (g)			SSC (°Brix)			Flesh firmness (N)		
Rank	Family	BV	SEP	Family	BV	SEP	Family	BV	SEP	Family	BV	SEP
1	Ac0378	0.276	0.147	Ac0404	11.380	1.093	Ac0394	1.574	0.865	Ac0394	-5.939	0.560
2	Ac0398	0.212	0.141	Ac0369	8.731	1.093	Ac0362	1.383	0.873	Ac0365	-5.369	0.561
3	Ac0397	0.202	0.148	Ac0398	8.320	1.091	Ac0373	1.100	0.850	Ac0362	-4.504	0.562
4	Ac0367	0.196	0.137	Ac0377	8.101	1.094	Ac0372	1.045	0.833	Ac0359	-3.708	0.550
5	Ac0369	0.163	0.146	Ac0396	6.900	1.090	Ac0401	0.878	0.874	Ac0380	-2.086	0.558
31	Ac0401	-0.164	0.158	Ac0362	-6.680	1.096	Ac0400	-0.813	0.840	Ac0367	2.560	0.549
32	Ac0362	-0.168	0.156	Ac0374	-7.058	1.093	Ac0395	-0.944	0.828	Ac0377	2.794	0.557
33	Ac0358	-0.184	0.157	Ac0368	-8.230	1.093	Ac0375	-1.096	0.907	Ac0368	3.072	0.556
34	Ac0366	-0.249	0.153	Ac0380	-8.405	1.094	Ac0378	-1.205	0.843	Ac0395	3.803	0.550
35	Ac0373	-0.274	0.151	Ac0372	-8.846	1.090	Ac0364	-1.527	0.836	Ac0400	3.972	0.554

4.5 Conclusions

In the present chapter, genetic parameters were obtained with BLUP methodology in a large kiwifruit mating design where three female parents were crossed to 12 male parents to produce 35 cross families. One cross combination failed to set fruit. Although the estimation of genetic parameters had several limitations due to the unbalanced family sizes and the missing cross combination mentioned above, the information obtained appears all the same useful for a rough estimation of the genetic value of cross parents, especially the male ones, for fruit-related traits. Furthermore, the breeding values of recently selected families will be a valuable information for the efficient design of the future crosses

Currently, the cost of testing, managing and maintaining relatively large population over a number of years is a limiting factor in any woody plant breeding program, and this is particularly so for kiwifruit where expensive support structures are required. For these reasons, robust methods of genetic analysis of both parents and offsprings, such as BLUP, would be considerably beneficial for the breeder for the considerable amount of information that they carry and the reliability of information gathered, even in the presence of unbalanced mating designs and less appropriate experimental procedure in the management of plants in the field.

Finally, a simple mixed model allows the exploitation of the pedigree information and temporal correlation, but could be further improved by addition of spatial correlation or through a multivariate analysis not considered in the present chapter, where selection indexes can be applied to a set of variables simultaneously considered.

4.6 References

Beatson R (1991) Inheritance of fruit characters in *Actinidia deliciosa*. Acta Hortic 297:79–86.

Butler D (2009) asreml: asreml() fits the linear mixed model. R package version 3.0. www.vsni.co.uk

Daoyu Z, Lawes GS, Gordon IL (2002) Estimates of genetic variability and heritability for vegetative and reproductive characters of kiwifruit (*Actinidia deliciosa*). Euphytica 124:93–98.

Davik J, Honne BI (2005) Genetic variance and breeding values for resistance to a wind-borne disease [Sphaerotheca macularis (Wallr. ex Fr.)] in strawberry (*Fragaria* x *ananassa* Duch.) estimated by exploring mixed and spatial models and pedigree information. Theor Appl Genet 111:256–264. doi: 10.1007/s00122-005-2019-3

De Faveri J (2013) Spatial and temporal modelling for perennial crop variety selection trials. PhD Thesis, University of Adelaide, SA, Australia

De Souza VAB, Byrne DH, Taylor JF (2000) Predicted breeding values for nine plants and fruits characteristics of 28 peach genotypes. J Am Soc Hortic Sci 125:460–465.

Durel CE, Laurens F, Fouillet A, Lespinasse Y (1998) Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. Theor Appl Genet 96:1077–1085. doi: 10.1007/s001220050842

Falconer D.S and Mackay F.C (1996) Introduction to quantitative genetics. 4th ed. Longman, Essex, UK.

Ferguson AR, Huang HW, Testolin R (2012) Kiwifruit In: Finn CE and Clark JR (eds.) Register of new fruit and nut cultivars. List 46. HortScience 47:536-562.

Hansche PE, Beres V, Brooks RM (1966) Heritability and genetic correlation in the sweet cherry. Proc Amer Soc HortSci 88:173–183.

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447.

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83.

Iwanami H, Moriya S, Kotoda N, et al (2008) Estimations of heritability and breeding value for postharvest fruit softening in apple. J Amer Soc Hort Sci 133:92–99.

Kempthorne O (1957) An introduction to genetic statistics. John Wiley & Sons, New York 547 pp

Marsh H, McNeilage M, Gordon I (1999) Breeding value of parents in kiwifruit (*Actinidia deliciosa).* Acta Hortic 498:85–92.

Marsh HD, Paterson T, Seal AG, McNeilage MA (2003) Heritability estimates in kiwifruit. Acta Hortic 622:221–229.

Piepho HP, Buchse A, Richter C (2004) A mixed modelling approach for randomized

experiments with repeated measures. J Agron Crop Sci 190:230–247.

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 209–228. doi: 10.1007/s10681-007-9449-8

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Richardson AC, McAneney KJ (1990) Influence of fruit number on fruit weight and yield of kiwifruit. Sci Hortic (Amsterdam) 42:233–241. doi: 10.1016/0304-4238(90)90085-S

Robinson GK (1991) That BLUP is a good thing: the estimation of random effects. Stat Sci 6:15–32.

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143:449–462. doi: 10.1017/S0021859605005587

Smith AB, Stringer JK, Wei X, Cullis BR (2007) Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. Euphytica 157:253–266. doi: 10.1007/s10681-007-9418-2

Testolin R, Cipriani G, Costa G (1995) Estimate of variance components and heritability of characters in kiwifruit (*Actinidia deliciosa*). Acta Hortic 182–189.

Testolin R, Cipriani G, Fiori G, et al (2011) Evaluation of parental genotypes as a tool for planning the mating design in kiwifruit breeding programs. Acta Hortic 913:213–220.

CHAPTER 5

DE NOVO ASSEMBLY OF MALE GENOME IN KIWIFRUIT (*ACTINIDIA CHINENSIS* Pl.)

5.1 Abstract

The kiwifruit (*A. chinensis* (2n) 'Hongyang') genome sequence recently published (Huang et al. 2013) is a valuable source of genomic information for the genetic improvement of kiwifruit and for a better understanding of genome evolution. However, kiwifruit is a functionally dioecious plant, and gender determinant has been mapped on chromosome 25 of male. Therefore, sequencing a male genome in kiwifruit would represent a starting point for the exploration of sex-determining region. Thus, a F1 population was generated from a cross between 'C8' (female) and 'A54.19' (male) at experiment farm at University of Udine. The male parent, 'A54.19' was selected as a candidate genotype for sequencing and genome assembly. DNA libraries were prepared for Illumina sequencing platform. A total of 16,144 segregating restriction-site-associated DNA (RAD) loci were obtained through a genotyping-by-sequencing (GBS) approach based on haplotype calling of SNP markers identified by a modified double digest restriction-associated DNA sequencing (ddRADseq) protocol. This chapter describes the male genome assembly and the future steps towards the exploration of the sex-determining region.

Keywords: sex chromosomes, dioecy, genome assembly, single nucleotide polymorphism (SNP)

5.2 Introduction

Sex determination is a major switch in the evolutionary history of angiosperm, resulting 11 % monoecious and dioecious species (Zhang et al. 2014). In *Actinidia*, dioecy is one of the common traits of the genus. All *Actinidia* taxa appear to be functionally dioecious, although this has been unequivocally establish in only a few taxa such *A. chinensis* and *A. polygama* (Kawagoe and Suzuki 2004; Goodwin et al. 2013). The sex-determining system operates like an XX/XY system with male heterogametic (Testolin et al. 1995; Harvey et al. 1997; Gill et al. 1998), and apparently operate in all taxa, at diploid, tetraploid, hexaploid and octoploid levels, with disomic segregation into females and males in approximately 1:1 ratio (Testolin et al. 1995; Fraser and McNeilage 2016). A single Y chromosome appears to be sufficient for maleness at any ploidy level (Testolin et al. 1995; Testolin et al. 1999; Fraser et al. 2009). This mechanism probably involves at least two tightly linked genes: a dominant allele for pistil suppression closely linked to a dominant gene for pollen development on the Y chromosome, and on the equivalent part of the X chromosome, two recessive alleles that allows pistil development and pollen death, respectively (Fraser et al. 2009; Fraser and McNeilage 2016). For dioecy to be maintained, recombination in this part of the chromosome would need to be suppressed, even if not completely. In this model, as previously mentioned, the male is heterogametic, i.e. X_nX/X_nY (Testolin et al. 1995; Testolin et al. 1999). Therefore, in diploids, females would be XX and males XY, in tetraploids the females XXXX and the males XXXY, in hexaploids the females XXXXXX and the males XXXXXY, etc.

Furthermore, it is thought that the X and Y chromosomes are of relatively recent origin, being derived from an autosomal pair, and the sex-determining locus only occupies a portion of the chromosome. Several studies have mapped the gender determinant on a subtelomeric region of Y chromosome (Fraser et al. 2009; Scaglione et al. 2015; Zhang et al. 2015; Liu et al. 2016). In addition, Fraser and McNeilage (2016) reported that non-recombining region with genes segregating with gender is about 15 % of the Y chromosome in *A. chinensis* var. *chinensis*. Similarly, in papaya, another dioecious plant, Liu et al. (2004) mentioned that non-recombining region is about 10 % of the Y chromosome.

However, in spite of this clear delimitation of the region, the genes involved in initiating gender differences in dioecious plants have proven to be extremely difficult to identify

and have not yet been described in any genus, including *Actinidia* (Fraser and McNeilage 2016).

An example of the lack of alignment of markers of female and male parents in the kiwifruit sex-determining region is provided by a paper of McNeilage et al. (2011) where the most updated linkage maps were compared and tentatively merged (Figure 9).

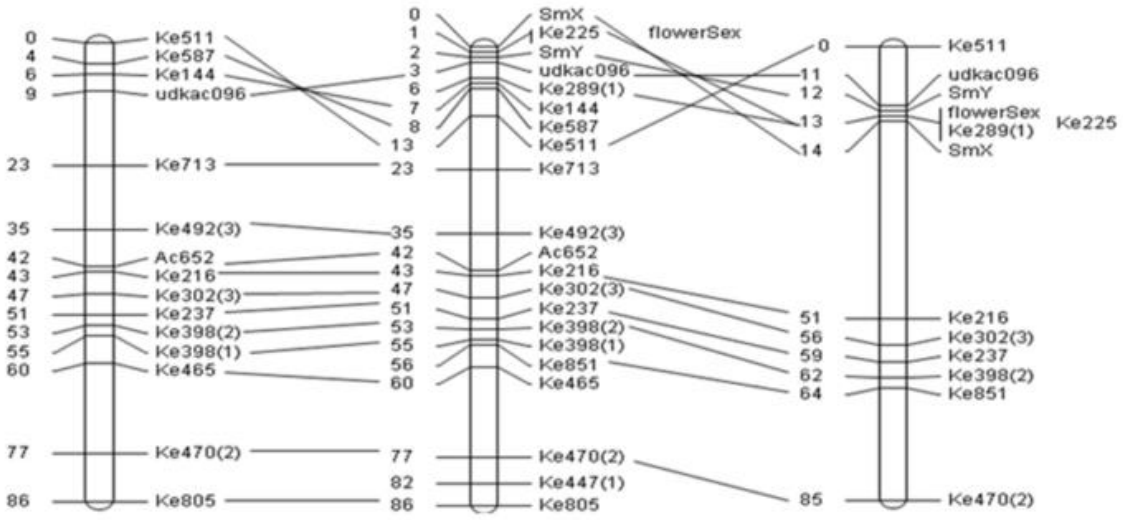


Figure 9. Inconsistency of the alignment of female genetic map (left) and male genetic map (right) in the consensus map (center) in the region of flower sex determination (top) due to haplotypic regions in female and male chromosome that do not recombine (after McNeilage et al. 2011).

On the other hand, the rapid advance of genomic technologies speeded up the investigation of sex determination genes and regulatory mechanisms of gene network (Zhang et al. 2014). Interestingly, the recently publication of the reference genome of a female *A. chinensis* 'Hongyang' (Huang et al. 2013) has been a fantastic starting points for genomics in kiwifruit.

However, the maleness determinant resides in the male genome in a species, like kiwifruit, where the heterogametic parent is the male one. Consequently, we started the exploration of sex control region in a male. The male parent, 'A54.19', of our mapping population was sequenced, and as the first step of this research is presented below.

5.3 Material and methods

5.3.1 Plant material and sequencing

As shown in Chapter 2, a pseudo-test cross population ('C8' x 'A54.19') was generated from a cross between 'C8' (female) and 'A54.19' (male). Parents are genetically and geographically unrelated and belong to the diploid (2n=58) *Actinidia chinensis*. In particular, 'A54.19' is a male originated from a seed introduced from the Beijing Botanical Garden. Young leaves of male were collected and a DNA library was prepared for Illumina sequencing. DNA was extracted following a CTAB-chloroform protocol, based on that of Doyle and Doyle (1990), modified by Lodhi et al (1994), Steenkamp et al (1994) and Manning (1991).

Briefly, 20 mg of young leaf tissue (two apical leaflets) was grinded in liquid nitrogen. Some 810 uL CTAB buffer with 1 % v/v β-mercaptoethanol was added, vortexed briefly and the tube left in the water bath at 60 °C for 20-30′ under the fume hood, mixing frequently. Suspension was cooled down to room temperature and 810 uL (24 : 1 v/v) chloroform : isoamyl alcohol was added. Vortexed for 10 min, opening the cap at intervals to allow gas leaking. The suspension was then transferred to 2 mL tubes and centrifuged at 13,000 rpm for 25-30′ at 15-20 °C. About 300 uL of the aqueous phase (the upper phase) was collected under the fume hood into another 1.5 ml centrifuge tube. The procedure continued with the protocol of Maxwell ® 16 LEV Plant DNA Kit (Blatter et al. 2006). 300 uL H_2O nuclease free was added together with 10 uL of RNAse A. The tube was briefly (10 sec) vortexed and spinned. The suspension was transferred to well #1 of Maxwell ® 16 LEV Plant DNA Kit reagent cartridge and the Plant protocol on Maxwell ® 16 was run on the Instrument (AS3050).

Once the DNA was extracted and controlled for the quality and concentration, sequencing and assembly was committed to the IGA Technology Services.

5.3.2 Preparation of paired-end and mate-pair libraries

The paired-end library was constructed following the ThruPlex DNA-Seq protocol and the mate-pair library was generated using Nextera Mate Pair protocol. Both libraries were sequenced on an Illumina HiSeq 2500 platform (Illumina Inc, San Diego, CA, USA). Raw reads were processed by removing PCR duplicates, low-quality reads, adaptor sequences and contaminated reads.

5.3.3 *De novo* assembly

A *de novo* assembly was generated with ALLPATHS (Butler et al. 2008) with default settings and the option HAPLOIDIFY set at TRUE. Short-insert paired-end reads and long- insert mate-pairs were used as an input to the assembler.

Briefly, raw Illumina reads were processed in order to de-multiplex samples on the basis of Illumina Truseq index and custom inline barcodes. Alignment on reference genome of *A. chinensis* 'Hongyang' (Huang et al. 2013), was carried out using Bowtie 2 aligner (Langmead et al. 2012) with default parameters and processed with Stacks software (Catchen et al. 2011).

5.4 Results and discussion

5.4.1 DNA sequencing

Genomic DNA was extracted from male individual coded as 'A54.19'. Two different DNA libraries were prepared with the aim of obtaining fragments with different insert size and reads with different length: short insert paired-end reads of 125 bp and 250 bp, and long insert mate-pair reads of 125 bp. The mode of insert size distribution was estimated following read alignment against the reference genome. The modal class was 435 bp for paired-end reads – forward/reverse (FR); 2,086 bp for mate-pair reads – reverse/forward (RF), and 93 bp with an additional peak at 624 bp for mate-pair reads –FR (Figure 10). In the mate-pair library, mayor peak in FR-reads is probably associated with the presence of adapters in not cleaned reads.

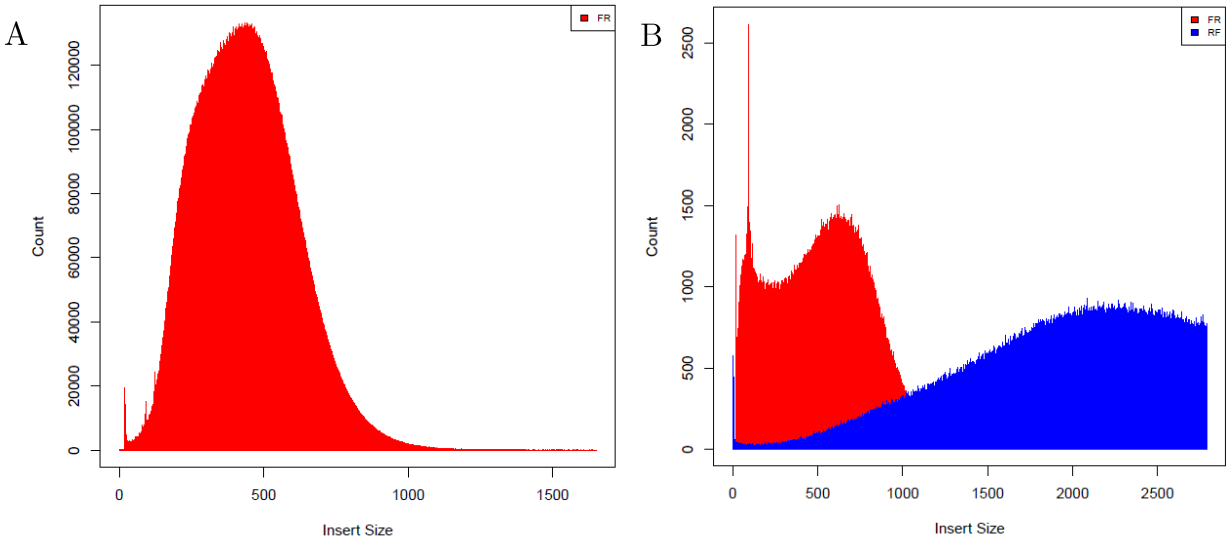


Figure 10. Plots of insert size (bp) for (A) paired-end and (B) mate-pair libraries. NGS libraries were prepared from genomic DNA of kiwifruit male 'A54.19'. Paired-end and mate-pairs reads were aligned against the reference genome. Data were extracted from BAM files using Picard tools.

Raw FASTQ data were processed for adapter removal, quality trimming e filtering for contaminants and duplicates (data not shown).

5.4.2 Male genome assembly

Male genome assembly is based on ALLPATHS assembly. ALLPATHS used approximately 164 million original fragment reads and 194 million jumping reads. Details of the number of reads per library and the fraction of assembled read are given in Table 13.

Table 13. Statistics of the libraries used by ALLPATHS.

Fragment type	Library	n_reads[a]	%_used[b]	n_pairs[c]
Fragment read	HiSeq 250-bp insert	130,588,182	64.2	58,739,291
Fragment read	HiSeq paired-end	33,880,768	70.5	12,781,467
Fragment read	Total	164,468,950	65.5	71,520,758
Jump read	HiSeq mate-pair	67,851,856	55.6	15,076,211
Jump read	Unclassified	127,065,318	48.6	27,752,081
Jump read	Total	194,917,174	51.1	42,828,292

[a] number of reads in input

[b] % of reads assembled

[c] number of valid pairs assembled

ALLPATHS produced scaffolds for a total genome length of approximately 697 Mbp, pretty close to the reference genome size, which was estimated in 758 Mbp according to the flow cytometry analysis (Hopping, 1994). The male genome assembly consists of 60,962 contigs and a total length of the contigs amounted to 565 Mbp. The total number of scaffolds is 7,955. Half of the genome is contained in scaffolds that are longer than 362 kb and in contigs longer than 20.9 kb. A summary of metrics comparing the final assemblies of male genome 'A54.19' and female genome 'Hongyang' is shown in Table 14.

Although the observed differences between genome assemblies may be due to the difference in bases of both genotypes, it is worth mentioning that male assembly covered 91 % of the estimated genome length, if we consider the total scaffold length with gaps, a value slightly better than the assembled 'Hongyang' genome, reported in 81 % (616 Mbp) (Huang et al. 2013).

Table 14. Comparison of metrics between the genome assemblies of the male 'A54.19' (this work) and the female 'Hongyang' (Huang et al. 2013).

Parameters	'A54.19'	'Hongyang'
Number of contigs [a]	60,962	26,721
Number of contigs per Mb	87.4	n.a[b]
Total contig length	565,353,733	604,217,145
Number of scaffolds	7,955	5,110
Number of scaffolds per Mb	11.40	n.a
Total scaffold length, with gaps	697,560,211	616,114,069
N50 contig size in kb	20.9	58.8
N50 scaffold size in kb	362	647
N50 scaffold size in kb, with gaps	425	n.a
Median size of gaps in scaffolds in kb	677	n.a

[a] > 1 kb for 'A54.19', > 2 kb for 'Hongyang'

[b] information not available

As the first male genome sequenced in kiwifruit, 'A54.19' can provide new evidence not only for the exploration of sex control region but also to improve the 'Hongyang' genome sequence. Huang et al. (2013) reported that their assembly contained 24 misassembled scaffolds. However, Scaglione et al. (2015) and Zhang et al. (2015) reported further scaffolds misassembly and misplacement. Clearly, no genome assembly of heterozygous individuals is likely to be complete or perfect. According to Crowhurst et al. (2016), it can only be a representation of the genome at that time and so should be viewed as a snapshot of the state of the art, where new research may lead to improve that representation.

5.4.3 Moving forward

With a male genome now available, the next step is to determine scaffold placement on kiwifruit pseudomolecules, with a high-density genetic map. A new map will be constructed using our F1 population derived from cross between 'A54.19' (male) and 'C8' (female) (see Chapter 2). Genotyping of each individual in the F1 population was done. The catalog of candidate segregating loci was populated by comparing male parental haplotype on the basis of their genomic coordinates. Reconstructed loci for

each progeny individual were then matched against this catalog in order to score segregating haplotypes. A total of 16,144 candidate segregating loci were already identified (Table 15) and the construction of a new male map is in progress.

Table 15. Summary of candidate segregating loci from a pseudo-test cross population.

Code	Description	n markers
lm x ll	locus heterozygous in the first parent	6,815
nn x np	locus heterozygous in the second parent	5,974
hk x hk	locus heterozygous in both parents, two alleles	1,526
ef x eg	locus heterozygous in both parents, three alleles	1,415
ab x cd	locus heterozygous in both parents, four alleles	414
	Total	16,144

Naturally, one of the expected results is the accurate ordering of scaffolds belonging to the linkage group 25, especially in the subtelomeric region, where the sex determinant was previously mapped (Appendix B). Subsequently, male pseudomolecule 25 can be improved by orienting contigs and scaffolds with the use of the dense genetic map produced and discussed previously in this thesis.

5.4.4 Future prospects

Our future prospects, beyond of this thesis, is firstly to refine the assembly of the subtelomeric region of chromosome 25; then to perform the gene annotation of that hot region. Gene annotation will be carried out using MAKER pipeline (Cantarel et al. 2008), which permits to identify repeats, aligns ESTs and proteins to a genome, produces ab-initio gene predictions and automatically synthesizes these data into gene annotations having evidence-based quality values. Specifically, Augustus and SNAP will be used for de novo gene prediction and, BLAST and Exonerate to align protein and EST sequences. In this way, we restricted of Y- specific candidate region according to clusters of genes related to flower development (floral organ identity). According to Huang et al. (2013), in a whole reference genome were predicted a total of 39,040 protein-encoding genes with an average coding sequence length of 1,073 bp and 4.6 exons per gene.

Subsequently, we hope to describe the complete sequence of the kiwifruit Y- specific region together with its X counterpart (reference genome), allowing a comprehensive

comparison of the gene content and repetitive content of a plant with a X/Y system of gender. Comparison between sequenced sex-determining regions has already been studied in papaya (*Carica papaya*), which is a trioecious tropical fruit tree with three sex types (female, male and hermaphrodite) controlled by an XY system (Wang et al. 2012). In fact, the genomic sequences of papaya sex chromosomes unveiled the molecular basis of recombination suppression in the sex determination region, and candidate genes for sex determination. Nevertheless, currently the lack of sex determination genes from dioecious species is still a major gap in our understanding of the origin of dioecy and sex chromosomes (Zhang et al. 2014). Hence, sequencing the male genome in kiwifruit provides a new resource for the future identification of sex determination genes and understanding their interacting gene network.

5.5 Conclusions

In the present chapter, we focused on a high-quality male genome in kiwifruit (*A. chinensis*), using Illumina paired-end and mate-pair reads. The assembly has been successful, covering approximately 91.3 % of the kiwifruit reference genome. Beyond the thesis, we are refining the assembly in the sex-determining region, by orienteering contigs and scaffolds. As soon as this work is completed, we will cope with the challenging task of gene prediction and annotation, which is harder than in other cases because genes involved in dioecy determination have not been yet discovered in any species. RNA-seq data will help in such a task.

Finally, in a long term view, we hope that dense maps available and their alignment with the genome assembly will allow mining also markers associated with the sex determinant that will be more stable, and a large transportability among genotypes and species. This will open the way to the early marker-assisted screening for sex in controlled crosses.

5.6 References

Blatter A, Bratz M, Gorshe R, Mandrekar P (2006) Isolation of DNA from plants: a comparison of Maxwell® 16, Manual Spin-Column and CTAB Methods. Available from: http://ita.promega.com/resources/pubhub/isolation-of-dna-from-plants-comparing-maxwellan d-ctabmethods/

Butler J, MacCallum I, Kleber M, Shlyakhter IA, et al (2008) ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. Genome Res 18: 810–820.

Cantarel BL, Korf I, Robb SMC, et al (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196. doi: 10.1101/gr.6743907

Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. G3: Genes, Genomes, Genetics 1:171–182.

Crowhurst R, Liu Y, Scaglione D (2016) The Kiwifruit Genome. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 101–114

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15.

Fraser LG, McNeilage MA (2016) Reproductive biology. In: Testolin R, Huang H-W, Ferguson AR (eds) The Kiwifruit Genome. Springer International Publishing, Cham, pp 65–84

Fraser LG, Tsang GK, Datson PM, et al (2009) A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X/Y sex-determining chromosomes. BMC Genomics 15:1–15. doi: 10.1186/1471-2164-10-102

Gill GP, Harvey CF, Gardner RC, Fraser LG (1998) Development of sex-linked PCR markers for gender identification in *Actinidia*. Theor Appl Genet 97:439–445. doi: 10.1007/s001220050914

Goodwin RM, McBrydie HM, Taylor M a (2013) Wind and honey bee pollination of kiwifruit (*Actinidia chinensis* "HORT16A"). New Zeal J Bot 51:229–240. doi: 10.1080/0028825x.2013.806934

Harvey CF, Gill GP, Fraser LG, McNeilage MA (1997) Sex determination in *Actinidia*. 1. Sex-linked markers and progeny sex ratio in diploid *A. chinensis*. Sex Plant Reprod 10:149–154. doi: 10.1007/s004970050082

Hopping ME (1994) Flow cytometric analysis of *Actinidia* species. N. Z. J. Bot. 32:85–93.

Huang S, Ding J, Deng D, et al (2013) Draft genome of the kiwifruit *Actinidia chinensis*. Nat Commun 4:1–9. doi: 10.1038/ncomms3640

Kawagoe T, Suzuki N (2004) Cryptic dioecy in *Actinidia* polygama: a test of the pollinator attraction hypothesis. Can J Bot 82:214–218. doi: 10.1139/b03-150

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. NatMethods 9:357–359.

Liu C, Zhang Q, Yao X, et al (2016) Characterization of genome-wide simple sequence repeats and application in interspecific genetic map integration in kiwifruit. Tree Genet Genomes 12:21. doi: 10.1007/s11295-016-0982-2

Liu Z, Moore P, Ma H et al. (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427:348–352.

Lodhi MA, Ye G-N, Weeden NF, Reisch BI (1994) A simple and effiecient method for DNA extraction from grapevine cultivars and Vitis species. Plant Mol Biol Rep 12: 6-13.

McNeilage M, Fraser LG, Tsang P, et al (2011) Molecular genetics and genomics and kiwifruit breeding. Acta Hortic 913:63–70.

Manning K (1991) Isolation of nucleic acids from plants by different solvent precipitation. Anal Biochem 195: 45-50.

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7:1–11. doi: 10.1371/journal.pone.0037135

Scaglione D, Fornasiero A, Pinto C, et al (2015) A RAD-based linkage map of kiwifruit (*Actinidia chinensis* Pl.) as a tool to improve the genome assembly and to scan the genomic region of the gender determinant for the marker-assisted breeding. Tree Genet Genomes. doi: 10.1007/s11295-015-0941-3

Steenkamp J, Wiid I, Lourens A, Van Helden P (1994) Improved method for DNA extraction from *Vitis vinifera*. Am J Enol Vitic 45: 102-106.

Testolin R, Cipriani G and Messina R (1999) Sex control in *Actinidia*. In: Ainsworth CC (ed) Sex determination in plants. BIOS Scientific Publisher Limited, Oxford, pp

137-146

Testolin R, Cipriani G, Costa G (1995) Sex segregation ratio and gender expression in the genus *Actinidia*. Sex Plant Reprod. doi: 10.1007/BF00242255

Wang J, Na J, Yu Q, et al (2012) Sequencing papaya X and Y h chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc Natl Acad Sci 109:13710–13715.doi:10.1073/pnas.1207833109//DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1207833109

Zhang J, Boualem A, Bendahmane A, Ming R (2014) Genomics of sex determination. Curr Opin Plant Biol 18:110–116. doi: 10.1016/j.pbi.2014.02.012

Zhang Q, Liu C, Liu Y, et al (2015) High-density interspecific genetic maps of kiwifruit and the identification of sex-specific markers. DNA Res 22:367–375. doi: 10.1093/dnares/dsv019

APPENDIX

A GENETIC MAPS

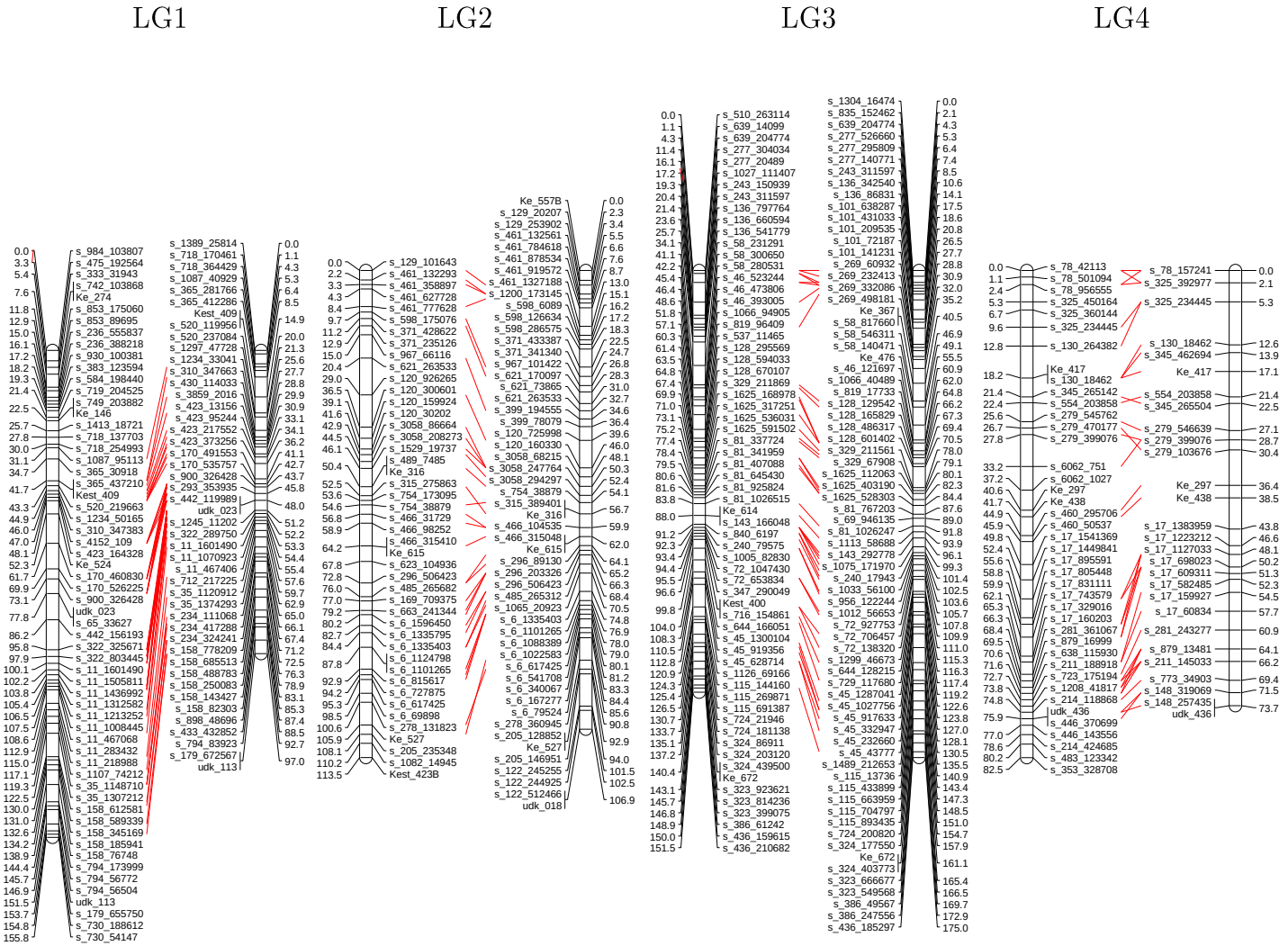
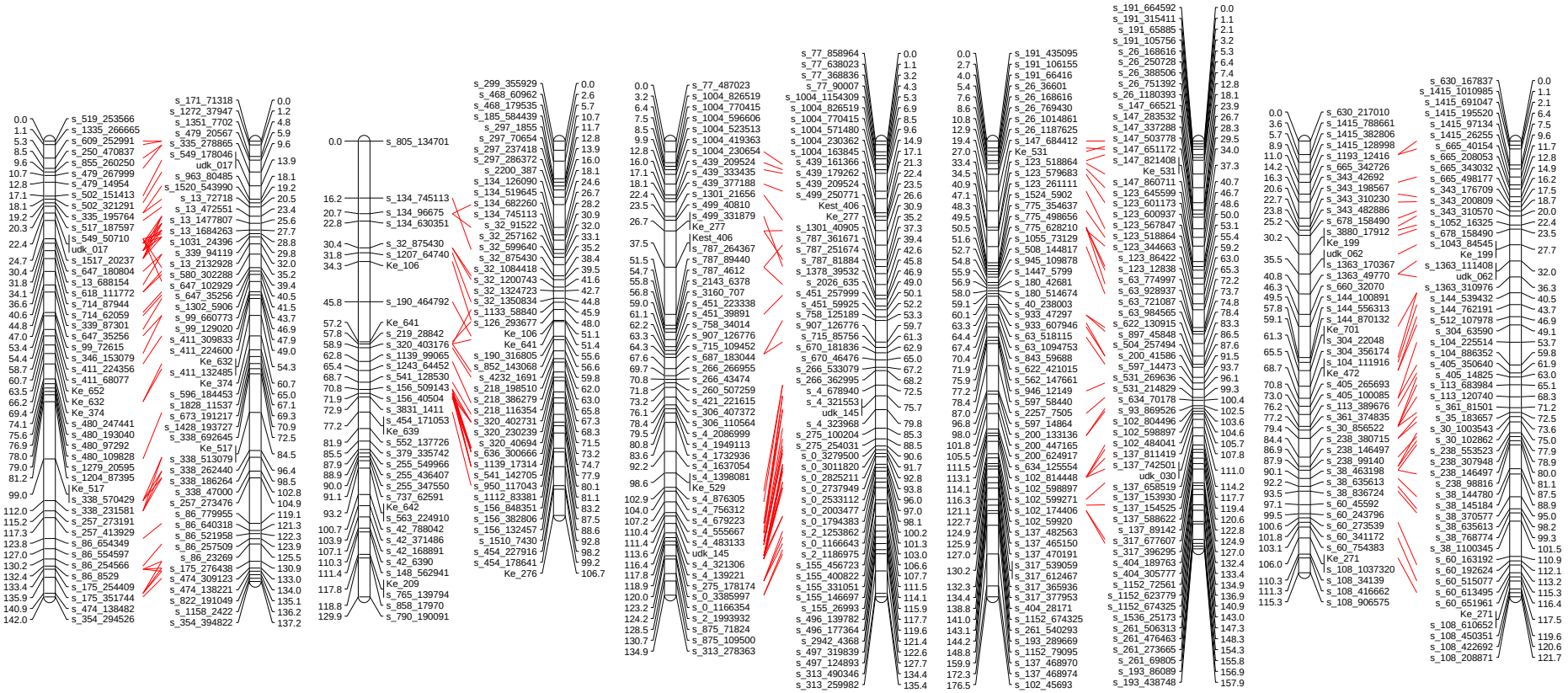
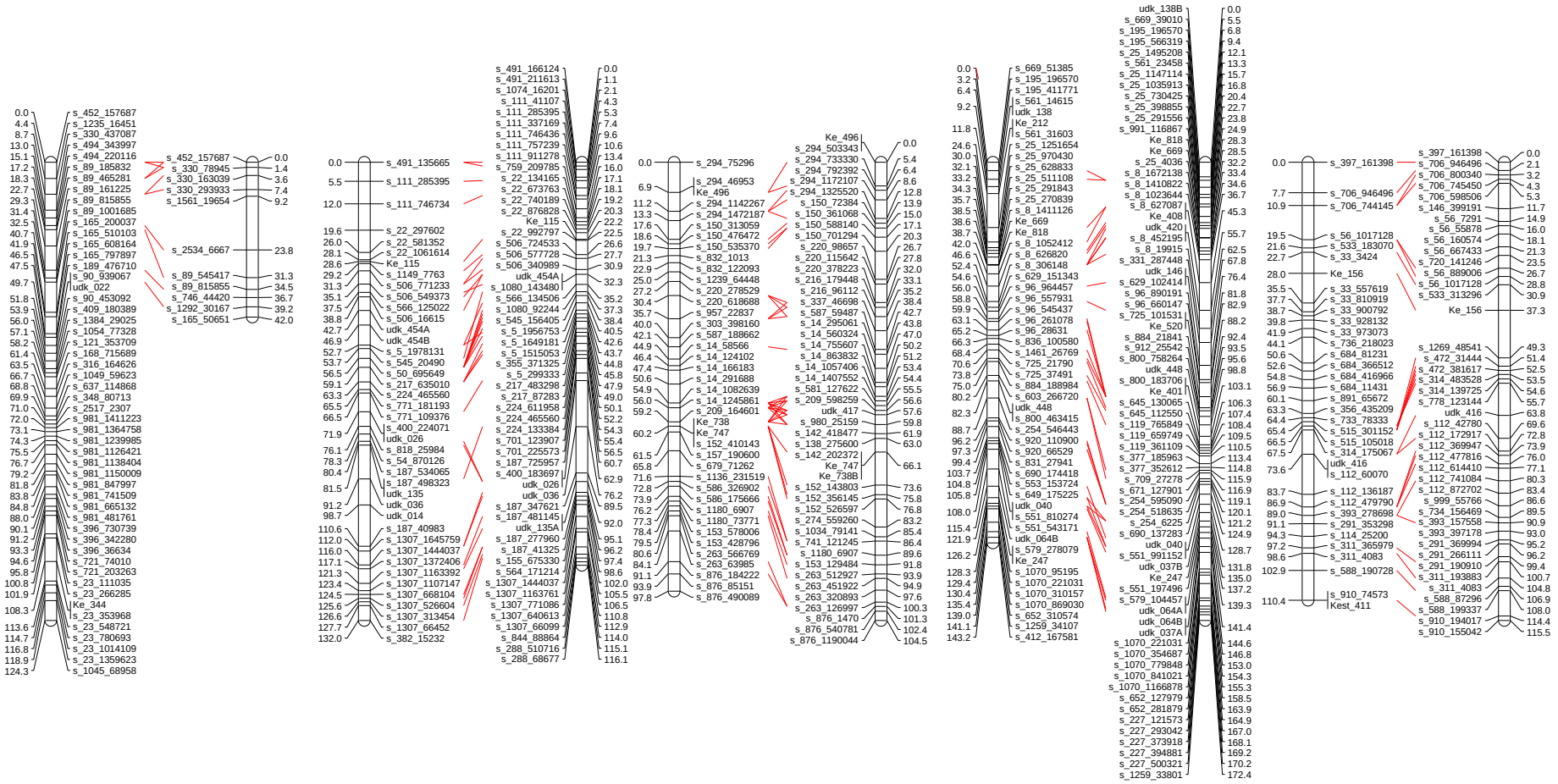


Figure 11. Alignment of parental linkage maps of a pseudo-testcross population of diploid kiwifruit (C8 x A54.19). In each linkage group: female parent first and male parent second; on the inner are the SSR and SNP markers and on the outer are the genetic distance in centiMorgan (cM). Red lines indicating homologous markers.

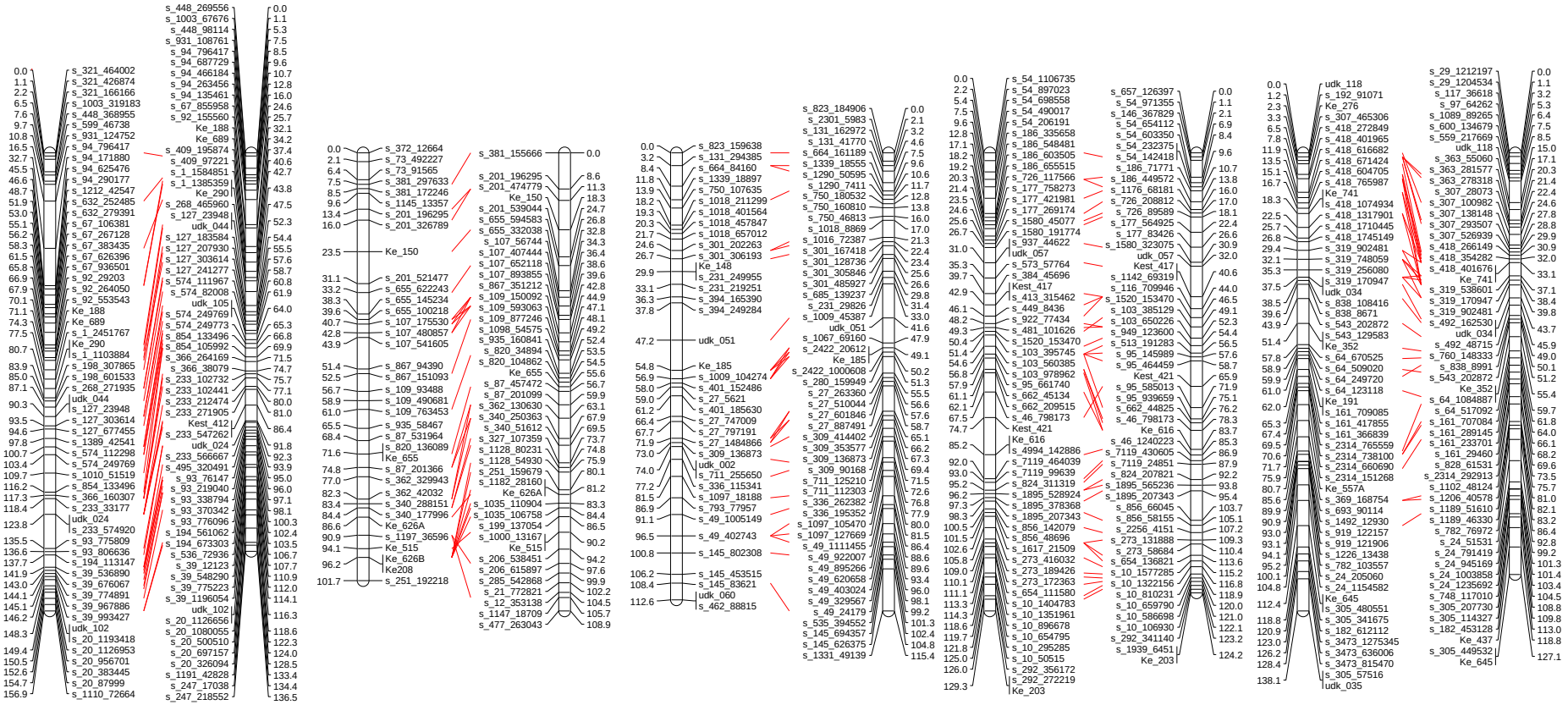
LG5 LG6 LG7 LG8 LG9



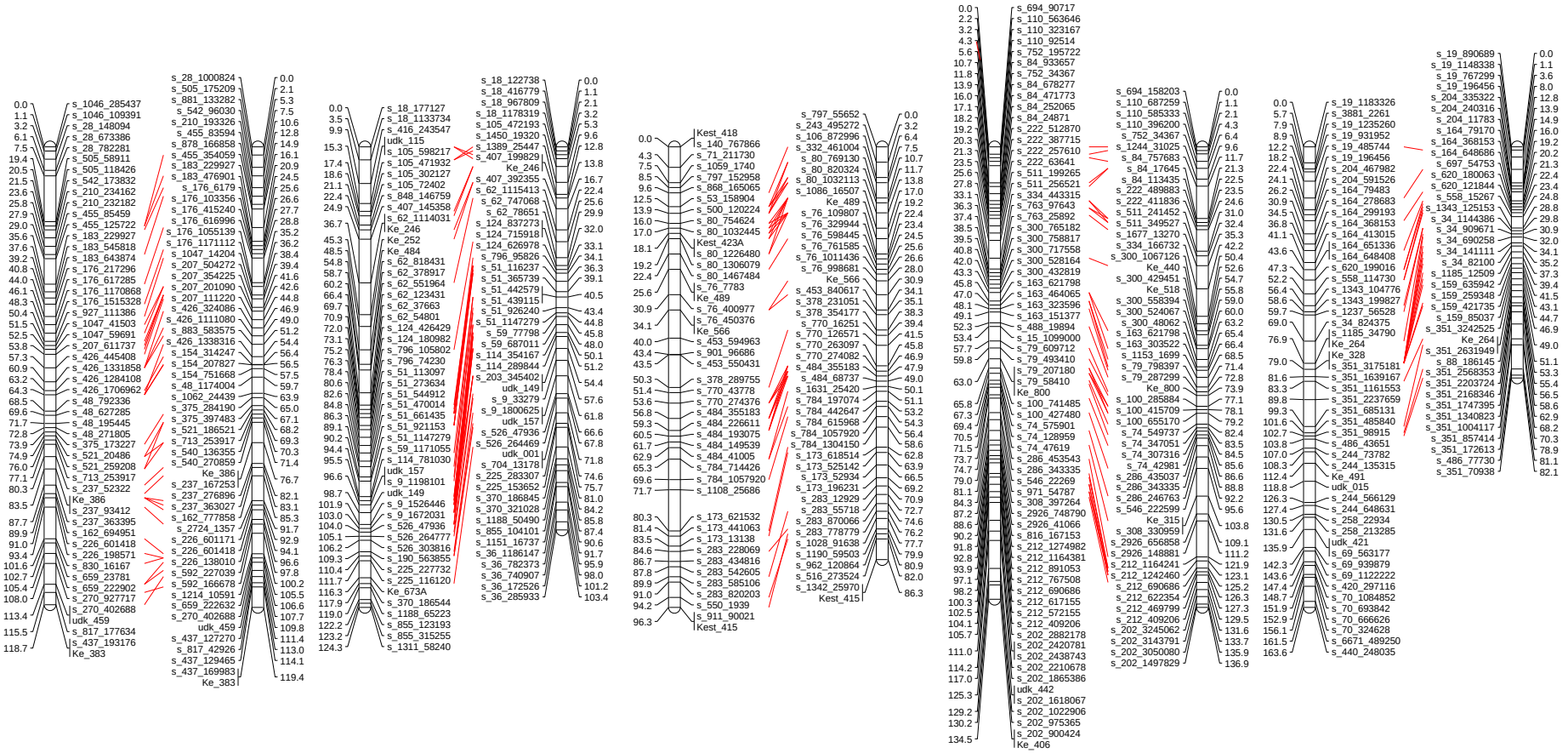
LG10 LG11 LG12 LG13 LG14



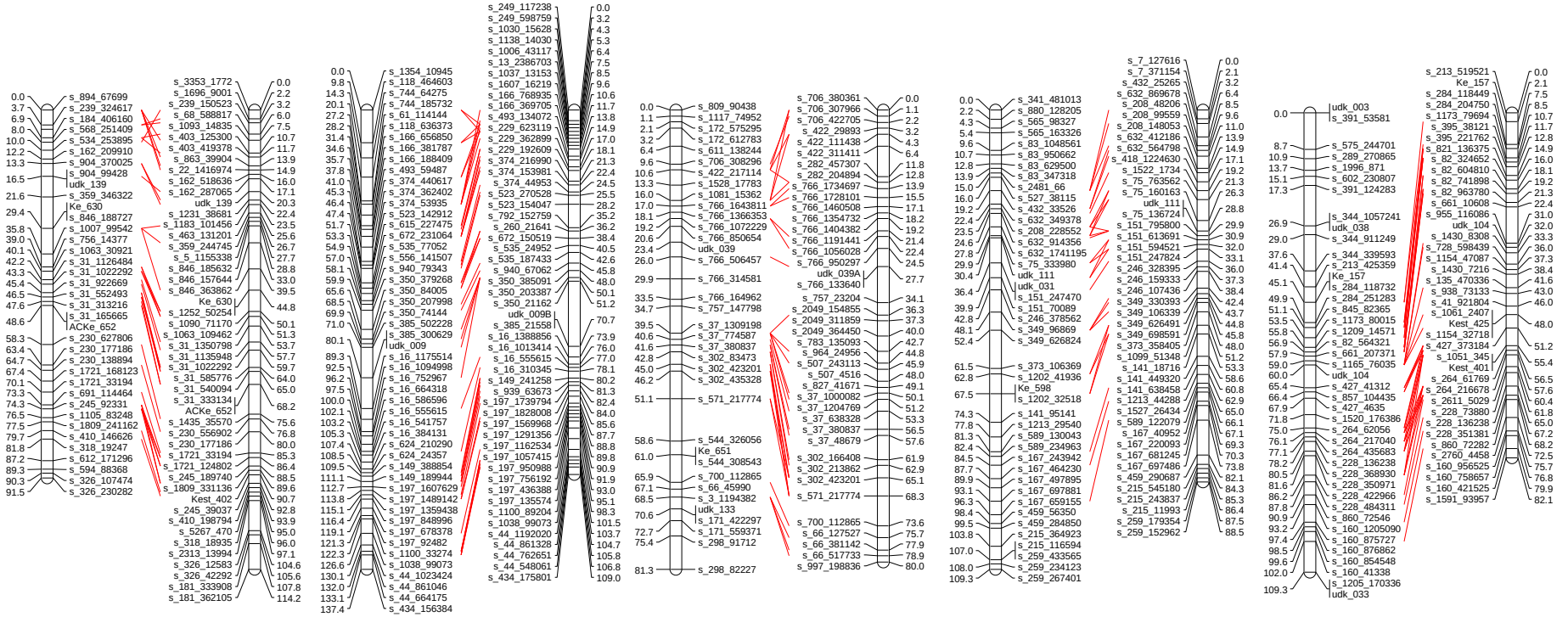
LG15 LG16 LG17 LG18 LG19



LG20 LG21 LG22 LG23 LG24



LG25 LG26 LG27 LG28 LG29



B CHROMOSOME 25

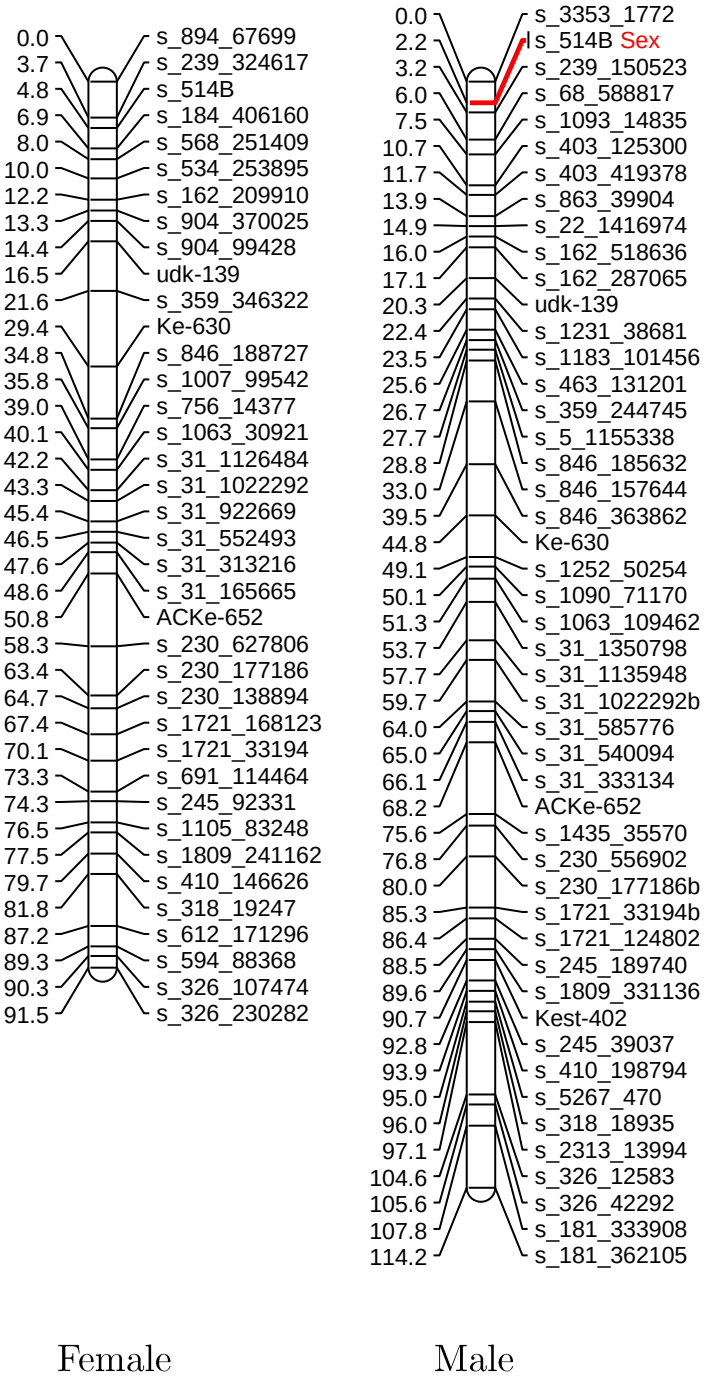
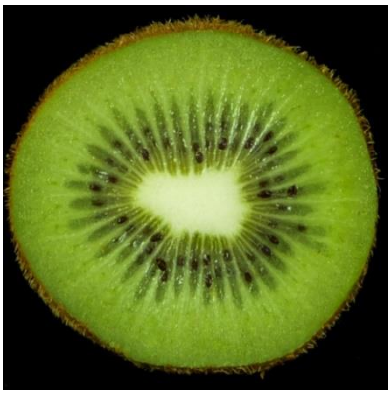
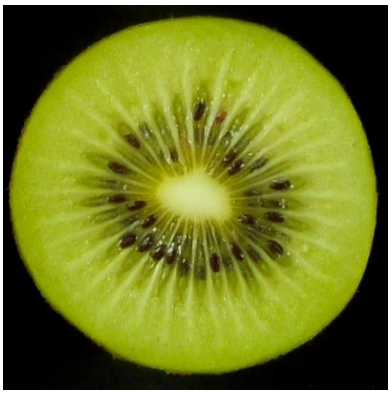


Figure 12. Chromosome 25 of female and male parent of a pseudo-test cross population of diploid kiwifruit (C8 x A54.19). On the right are the SSR and SNP markers and on the left are the genetic distance in centiMorgans (cM).

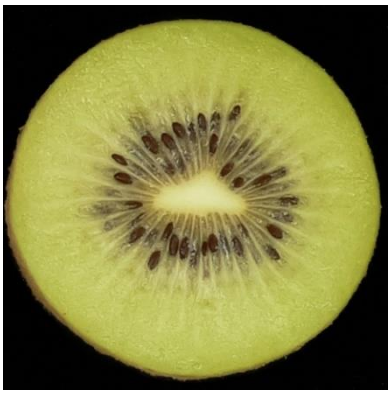
C FLESH COLOR CHART



Green

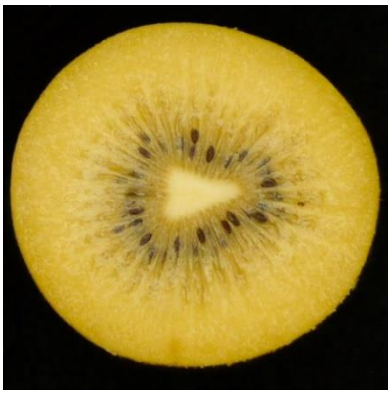


Light green



Yellow green

Light yellow



Yellow

Intense yellow