Genetic dissection of the dwarfing effect of the apple rootstock 'M.9'

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Plant Biology at Massey University, Palmerston North, New Zealand.

Jean-Marc Celton

2007

ABSTRACT

The dwarfing phenomenon in apple is mainly characterized by the ability of a rootstock to reduce the vegetative growth and ultimate size of the scion grafted onto it. Many hypotheses have been proposed to explain the dwarfing effect of rootstocks, from the production and translocation of hormones to the altered anatomy of the graft union. However, despite the numerous studies undertaken, none of the studies support a single hypothesis. This study focussed on identifying genetic markers for genomic regions influencing the dwarfing effect of 'Malling 9' ('M.9').

Two rootstock populations derived from crosses between 'M.9' and the vigorous rootstock 'Robusta 5' ('R5') were propagated and phenotyped at the HortResearch Havelock North Research Centre (New Zealand). Bulked segregant analysis (BSA) and QTL analysis were performed using phenotypic data collected from scions grafted onto the rootstock populations. Several genomic regions in 'M.9' and 'R5' were found to have a significant influence on the dwarfing phenotype and other related traits. The percentage of variation explained by these QTLs ranged from 4.2% to 57.2%. A large number of traits had significant variation associated with the major dwarfing QTL *DW1* (identified previously by BSA), confirming the influence of this gene on tree architecture.

To identify the genes responsible for the dwarfing effect, a microarray analysis on RNA extracted from bark tissues was performed to detect genes differentially expressed among dwarfing and vigorous rootstocks derived from the 'M.9' × 'R5' family. Following the mapping of 16 markers developed from 12 candidate genes, their position was compared with those of the QTLs identified previously and co-localisations among genes and QTLs were identified. Results to date indicate that none of these particular CGs co-segregate with *DW1*.

In order to estimate the number of different genetic sources of dwarfing present in commercial rootstocks, two SSR markers mapping about 0.5 cM away from the dwarfing QTL *DW1* were screened over 58 rootstock accessions. The majority of the dwarf and semi-dwarf accessions screened carried the locus *DW1*, indicating that there may be only a single genetic source of dwarfing in apple rootstocks.

The identification of markers for dwarfing will have a major impact on apple rootstock breeding, which currently relies on laborious phenotyping of individuals in breeding populations that presently takes at least 5-7 years to adequately perform. The identification of the genetic function of *DW1* would provide an opportunity to develop

dwarfing rootstocks for other members of the Rosaceae family for which such rootstocks have not yet been developed.

ACKNOWLEDGEMENTS

I have always said that the first person in my acknowledgements would be Bev Breslin. So thank you Bev for the hundreds of sephadex plates and kilogrammes of gels you made for me in the last three years, I really appreciate! A big thanks also to everyone else in the lab (Heather, Emily, David, Mike, Claudia, Charmaine, Deepa, Dulantha, Chris, Wendy and Rosalie), without whom laboratory life would have been very monotonous. Heather and Emily, you have been great bench mates, and I will miss your company.

I would like to send my gratitude to Sue Gardiner who welcomed me in her lab when I was barely speaking English, and offered me this wonderful opportunity. Thanks also to Stuart Tustin and to all the people from Palmerston North, Havelock North and Auckland, who help me, directly or not, during these three years.

Thank you also David (Chagné), you taught me about mapping and was always here when I needed an advice. Thanks also Mike Cook, your robots saved me from going crazy. Thanks also for taking me out and make me discover that it is possible to have fun in Palmy!

Ludivine, thank you for your love, support and encouragement. Together we have been through this ordeal, and I am sure that the next chapter of our life will be just as exciting.

TABLE OF CONTENTS

List of figures

List of tables

List of abbreviations

ABA	abscisic acid
ANOVA	analysis of variance
AFLP	amplified fragment length polymorphism
BAC	bacterial artificial chromosome
bp	base pair
BSA	bulked segregant analysis
°C	degrees Celsius
cDNA	complementary DNA
CG	candidate genes
cm	centimetre
cM	centiMorgan
cv.	cultivar
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DW1	dwarfing locus 1
DW2	dwarfing locus 2
DW%	overall dwarfing phenotype
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
FDR	false discovery rate
GAs	gibberellins
gDNA	genomic deoxyribonucleic acid
GDR	Genome Database for Rosaceae
IAA	indole-3-acetic acid
indel	insertions and deletions
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	kilo base
KS	Kolmogorov-Smirnov
L	litre
LG	linkage group
LOD	logarithm of odds
m	metre
'M.9'	'Malling 9'
MAS	Marker assisted selection
mg	milligram

min	minute
ml	millilitre
mM	millimolar
ng	nanogram
MQM	multiple QTL model
mRNA	messenger RNA
ORFs	open reading frames
PCR	polymerase chain reaction
qPCR	quantitative PCR
QTL	quantitative trait loci
'R5'	'Robusta 5'
RAPD	randomly amplified polymorphic DNA
Rep	repetition
RFLPs	restriction fragment length polymorphism
RNA	ribonucleic acid
s	second
SCAR	sequence characterized amplified region
SIM	simple interval mapping
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
STSs	sequence tagged site
TAE	tris acetic acid EDTA
TAIR	The Arabidopsis Information Resource
TCA	trunk cross-sectional area
TIGR	The Institute for Genomic Research
Tris	tris(hydroxymethyl)aminomethane
Ty medium	tryptone yeast extract media
µg	microgram
µl	microlitre
µM	micromole
UTR	untranslated region
UV	ultra violet
Xgal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YAC	yeast artificial chromosome

1- INTRODUCTION

1.1 Apple and apple rootstocks

1.1.1 Overview of apple and apple rootstocks

1.1.1.1 Apple production

Apple is the most ubiquitous species among the temperate fruit crops and grows in high latitude regions of the world, where minimum temperatures may reach -40°C, up to a high elevation, as well as in the tropics where two crops may be grown in a single year (Janick, 1974). In terms of production, apples are the third biggest fruit crop in the world, with an estimated 62 million metric tonnes (MT)), behind banana (68 million MT) and grapes (66 million MT) (FAO 2005). Apple production relies on very few cultivars, mainly 'McIntosh', 'Jonathan', 'Cox's Orange Pippin', 'Granny Smith', 'Delicious', 'Golden Delicious', 'Braeburn', 'Royal Gala', 'Jonagold', 'Fuji' and 'Elstar'.

1.1.1.2 Taxonomy

Apple (*Malus*) belongs taxonomically to the Rosaceae family, along with other economically important temperate fruit crops. Together with other closely related fruit (*Pyrus* and *Cydonia*), and ornamental (*Amelanchier*, *Aronia*, *Chaenomeles*, *Crateagus*, *Pyracantha*, *Sorbus*, *Cotoneaster*) genera, they form the subfamily *Maloideae*. The genus *Malus* comprises some 55 species, although between eight and 79 species have been recognized (Harris *et al.* 2002). The scientific names that have been applied to the domesticated apple include *Malus pumila* Miller, *M. communis* Desf., *M. sylvestris* (L.) Miller, and *Malus domestica* (Borkh) (Mabberley *et al.* 2001, Harris *et al.* 2002). In this thesis, following Korban and Skirvin (1984), I will use *Malus domestica* (Borkh) when referring to apple, although most recent published opinion (Kartesz and Gandhi, 1992) argues for *M. sylvestris* (L.) Miller.
The origin of the *Maloideae* subfamily has also been the subject of some debate. The basic chromosome number x = 17 suggests a polyploid origin since other Rosaceae have x = 7, 8 or 9. Autopolyploidy was suggested by Darlington and Moffet (1930) but more recently the possibility of an allopolyploid origin with *Amygdaloideae* (x = 8) and *Spiraeoideae* (x = 9) has been advanced (Phipps *et al.* 1991; Lespinasse *et al.* 1999).

The common domesticated apple is usually designated *Malus* × *domestica* Borkh. (Korban and Skirvin, 1984) or *M. domestica* Borkh (Phipps *et al.* 1990). Most *Malus* species are diploids (2n=34), but triploid and tetraploid species have been observed (Way *et al.* 1989).

1.1.1.3 The apple tree

A whole apple tree used for commerce usually constitutes a rootstock onto which a scion is grafted. Occasionally, a third distinct component, the interstock, is grafted between the rootstock and the scion (Webster and Wertheim 2003) (Figure 1.1). Rootstocks have been developed by horticulturists originally as an aid in the propagation of selected scions but with apple today, also largely for control of scions growth.

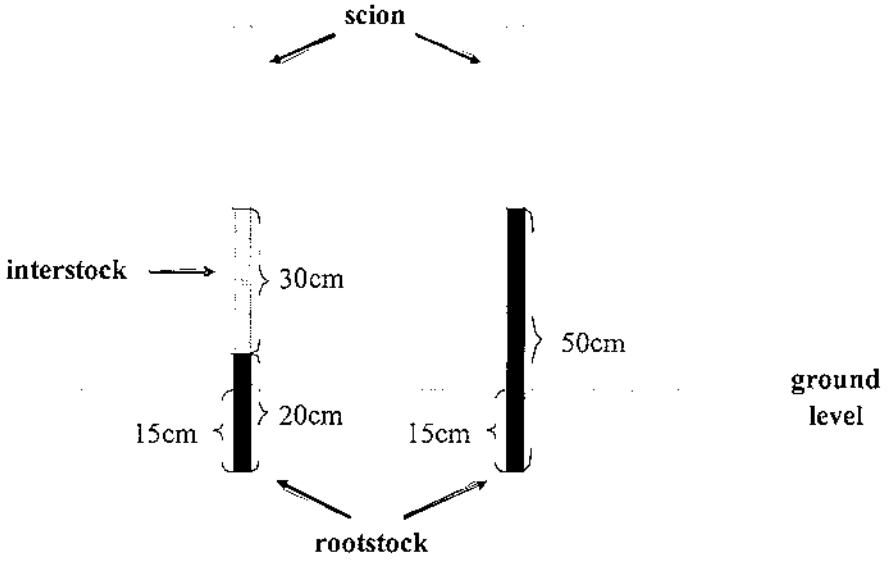

Figure 1.1. Standard size and position of rootstock/interstock segments of apple tree. Trees with both rootstock and interstock (left) and rootstock only (right) in the apple rootstock/interstock system are illustrated (Figure provided by Alla Seleznyova).

Most apple rootstock clones are derived from *Malus* species including *M. prunifolia* Willd., *M. baccata* (L), Borkh., *M. pumila* (Malling series), *M. sylvestris* Miller, *M. micromalus*, *M. niedzwetskyana* ('Red Standard'), *M. floribunda*, *M.* × *robusta* ('Robusta 5') and *M.* × *domestica* ('Northern Spy') (Ferree and Carlson 1987; Webster 2002).

1.1.1.4 Origins and History

Although the origin of apple is not entirely clear, it probably evolved from extensive forests of apples in central Asia, particularly in Kazakhstan, Kyrgystan, Uzbekistan, Turkmenistan and Tajikistan (Ferree and Warrington 2003, Harris *et al.* 2002).

It is established that rootstocks have been used in the propagation of apple trees for more than 2000 years (Webster and Wertheim 2003). Historians believe that dwarfing apple rootstocks were used in Greece during the Hellenistic period (after 323 BC). Alexander the Great was reported by the Greek botanist and philosopher Theophrastus to have brought dwarfing apple trees (or seeds) from Asia Minor (Turkey, Caucasus mountains, Iran).

The extension of apple cultivation throughout Europe was realised via the expansion of the Roman Empire. By the 1500s to 1800s, apples were grown on two general types of dwarfing rootstocks in France: Doucin and Paradise. In the 19th century, the apple rootstock 'Jaune de Metz' later renamed by Hatton (1916) as 'Malling 9' ('M.9') was discovered and has since become the industry standard for dwarfing rootstocks in apple. The parentage of this rootstock is unknown and the sequence of events associated with its selection is obscure (Carrière 1879).

1.1.1.5 Dwarfing rootstocks and interstocks

Dwarfing rootstocks are considered to be rootstocks which when grafted with scions, produce mature trees that are significantly smaller than trees grown on their own roots. Trees that grow slowly (by a reduced annual shoot growth) but at maturity are of large size are not considered as being part of the dwarfing category (Rusholme Pilcher *et al.* in press).

Interstocks were traditionally used when graft compatibility between the rootstock and the scion was a problem. Interstock use in apple is applicable when the desired dwarfing rootstock is difficult to propagate on its own roots or when the soil is not appropriate for the use of dwarfing rootstocks. The use of interstock can provide an alternative way of regulating the vigour of shoot growth on the scions (Wertheim and Callesen 2000).

1.1.1.6 Uses and benefits of dwarfing rootstocks and interstocks

Many of the important tree crops such as apple, peach, apricot, cherry or plum, use rootstocks as a way to improve production efficiency (Webster *et al.* 2000). Most fruit growers are seeking trees with limited height to facilitate more efficient production, harvesting, pruning and thinning operations. Smaller trees are also easier to spray, which reduces spray drift and increases efficiency of spray usage. Because they have reduced vegetative growth they can be planted close together to increase yield efficiency of orchards, especially precocious production; they require less pruning so labour cost is reduced and they produce more uniform sized and coloured fruits (Sarwar *et al.* 1998).

Clonal apple rootstocks offer a wide range of scion vigour control, ranging from super dwarfing to very vigorous. Currently, apple producers favour dwarfing selections, essentially 'M.9', resulting in a tree size of no more than three to four metres in height (Webster 2004).

In apple, a dwarf growth habit is the aim of many rootstock breeding programmes (Welander 1988). With apple, the degree of dwarfing of the scion can be influenced by the height of the graft. Increasing the length of rootstock stem between the roots and the graft union can increase the degree of dwarfing in the scion, especially with dwarfing and semi-dwarfing rootstocks (Parry 1986). Insertion of a dwarfing interstock between a combination of vigorous rootstock and scion reduces the scion vigour (Felius and Toorenaar 1959; Carlson and Robitaille 1970). The dwarfing effect increases with the length of the interstock, to an upper limit of 20 cm, beyond which no further dwarfing is attained (Kamboj *et al.* 1999a). In some cases, the vigour of trees using an interstock can be the same as if the same trees were grafted onto the corresponding rootstocks (Parry and Rogers 1972). All these observations imply that in apple, the dwarfing effect is partly due to the rootstock stem. It also suggests that these stems modify the exchange of hormones, water and nutrients between the root system and the scion crown and *vice versa.*

Dwarfing rootstocks and interstocks influence floral precocity (Tydeman 1937), either directly or indirectly by their effects on scion branching. Precocity is defined as the capacity to bear prolific quantities of blossom and comparatively heavy crops of fruit relatively early in the life of the tree (Tydeman 1937). The juvenile period, determined by the appearance of the first inflorescences, can be very long in apple and last up to 12 years (Fischer 1994). A short non-flowering phase is required to reach productivity in the three to four years after grafting. Using dwarfing rootstocks

has proven to be an efficient way to shorten the time until first flowering (Visser 1970; Aldwinckle 1975; Fischer 1994). All dwarfing rootstocks appear to have the effect of reducing the duration of the non-flowering phase (Tustin, personal communication). The precocity of flowering, and the ability of those flowers to set, retain, and size fruits, is crucial to the profitability of commercial tree fruit orchards. Substantial yields of high quality fruits early in the life of an orchard is essential if investments in fruits growing are to be worthwhile, particularly with high density plantings on dwarfing rootstocks (Webster 1995).

Rootstocks and interstocks are not the only means available to horticulturalists for the control of tree size. Each scion has an intrinsic vigour that influences its growth. Climatic factors, such as temperature, light and rainfall, as well as soil factors, like depth, mineral content and water availability must be taken into account. Management factors such as irrigation, nutrition, pruning (including root pruning) and weed control play an important part in the control of scion development. Tree factors like tree health and crop loading, which influences competition between fruit and shoot development for water, minerals, and assimilates also play a part in the architectural vigour of the developed scion (Webster 1995).

1.1.1.7 Overview of the major apple rootstocks and breeding programs

At the end of the 19th century, the number of different rootstocks used throughout Europe became relatively important, and because of the confusion arising due to their various effects on grafted scions, it became essential to classify them (Manhart 1995).

1.1.1.7.1 The Malling Apple rootstocks

In 1912, at the East Malling Research Station in Kent, England, all the different rootstocks known were gathered together and planted out for propagation. These rootstocks were standardized lines of well known vegetatively propagated rootstocks, some of them several centuries old.

From 1913 to 1935, R. Wellington followed by Lord Hatton and Dr. Tydeman identified 26 types of rootstocks, from 'M.1' to 'M.26' (Manhart 1995). Because of the various locations where the rootstocks were collected, the relationship among them is unknown. The dwarfing rootstock 'Malling 9' originates from this collection. 'Malling 9' ('M.9'), also called 'Jaune de Metz', 'Yellow Metz', 'Yellow Paradise of Metz', and 'Dieudonne', was selected as a chance seedling in France in 1879. The

leaves are large, shiny and flat (Figure 1.2). The shoots are straight, reddish yellow with silvery sheen and brittle. Scions grafted to this rootstock are dwarfed, about two metres high, and have a short juvenile phase (Webster 1995). 'M.9' is susceptible to woolly aphid (*Eriosoma lanigerum* Hausm.) and fire blight (*Erwinia amylovora* (Burrill) Winslow *et al.*), but resistant to collar rot.

Figure 1.2. Illustration of a shoot (left) and leaf (right) of a 'M.9' tree
(http://www.agric.nsw.gov.au/reader/7285)

1.1.1.7.2 The Malling-Merton apple rootstocks

In 1922, the John Innes Horticultural Institute then at Merton (England) and the East Malling Research Station at East Malling, began to raise a series of rootstocks by systematic plant breeding. This project was aimed at improving the range of stocks available, especially by introducing a resistance to woolly aphid which was then an important problem for fruit growers in the Southern Hemisphere former British colonies of South Africa, Australia and New Zealand. 'Northern Spy' was used as a parent for its resistance to woolly aphid, and various 'Malling' and other rootstocks were used as the other parent. The two stations released 'M25', 'M26', 'M27' and the 'Malling-Merton' ('MM') series, 'MM101' – 'MM115' (Cummins and Aldwinckle 1983). 'Merton 793', which is widely used, was also developed by the two institutes.

1.1.1.7.3 The Cornell Geneva ('CG') series

In 1968 the New York State Experimental Station in Geneva started a breeding programme principally aimed at resistance to fire blight and *Phytophthora*. Some rootstocks derived from this programme have recently been released. They cover a wide range of tree size control with improved winter hardiness, such as 'CG 202',

'CG 210' and 'CG 179'. Moreover, 'CG 179', 'CG 202' and 'CG 210' are resistant to fire blight, woolly apple aphid and *Phytophthora* (Wertheim 1998).

1.1.1.7.4 Important accessions from other breeding programmes

The other important rootstocks utilized in breeding programmes include 'Northern Spy' (originated at East Bloomfield in western New York in 1828), *Malus* 'Robusta 5', 'Ottawa 3' and 'Vineland 1' (from V. series) rootstocks developed in Canada, 'Bemali' selected in Balsgard in Sweden, 'Marubakaido' selected in Japan, and finally 'Aotea' (or 'Aotea 106') bred in New Zealand (Wertheim 1998).

1.1.1.8 Classification of rootstock according to their dwarfing effect

Rootstocks are traditionally classified into six categories: very dwarfing; dwarfing; semi-dwarfing; intermediate (also known as semi-vigorous); vigorous and very vigorous. Recently a new method of classification has been proposed. It consists in classifying rootstocks based on the percentage of growth of the scion compared to a tree grown on its own roots. However this method is suitable for a specific location only as the effects of the rootstocks vary depending on the environment in which the tree is grown (Fazio, personal communication).

The figure 1.3. represents a classification of the Malling series, the Malling-Merton series as well as some accessions originating from Cornell Geneva and other breeding programmes. These accessions have been classified into the six categories mentioned above, with the help of tree physiologist Stuart Tustin, according to growing conditions in the apple growing region of Hawkes Bay (North Island, New Zealand).

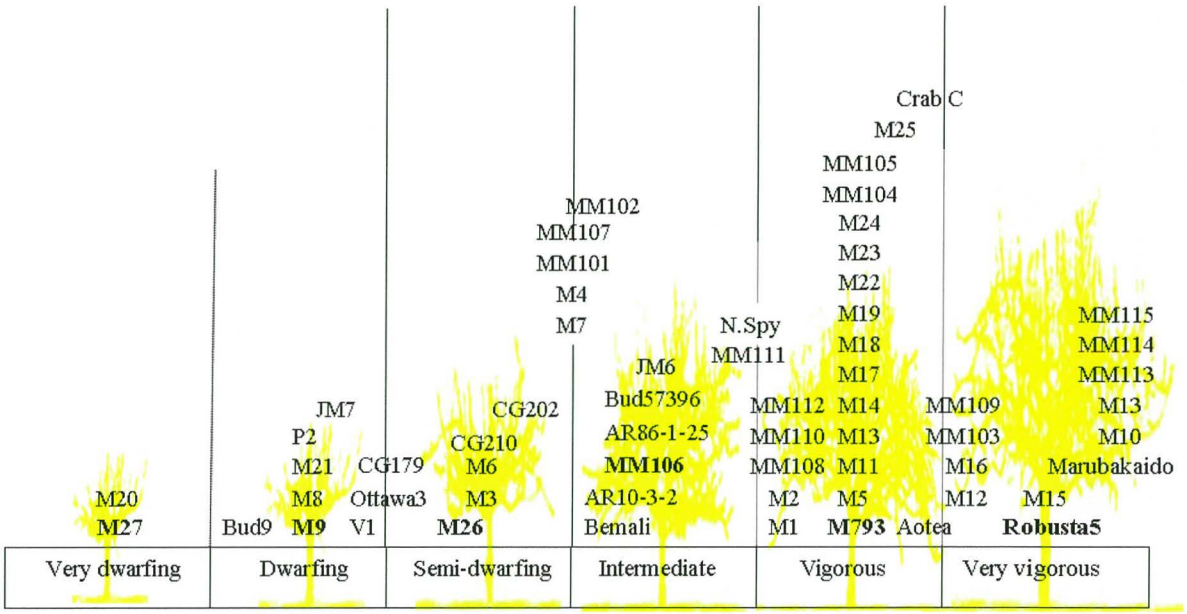

Figure 1.3. Classification of major apple rootstocks according to their dwarfing effect. Classification based on the growing conditions of Hawkes Bay (North Island, New Zealand). Rootstocks are organized in the order of their dwarfing effect on apple scions. The standard rootstock of each category is indicated in bold (ex. 'M.9' for the dwarfing category).

1.1.1.9 Dwarfing rootstocks in other members of the Rosaceae family

Dwarfing rootstocks for other deciduous (temperate) fruit species are of more recent origin than apple (Webster 2004). Rootstocks used for propagating stone fruits such as peaches, apricots, cherries and plums are very often of different species from the fruiting scion species, and some of these rootstocks may be hybrids (Webster and Wertheim 2003). It is only in recent years that dwarfing rootstocks have been reported for these crops and are commercially available.

Pear cultivars, particularly European pears, are mainly propagated by means of common pear (*Pyrus communis*) or quince (*Cydonia oblonga*) rootstocks. Some pear rootstocks such as 'Kosui' and 'La France' have a dwarfing growth habit when grafted to selected scions (Robbani *et al.* 2006).

1.1.2 Rootstock effect on scion development

1.1.2.1 Factors influencing scion development

The mechanisms by which rootstocks influence tree size in apple have not been fully understood. Over the years, several theories have been proposed in an attempt to explain, in anatomical, nutritional, hormonal, or other physiological terms, the dwarfing effect of rootstocks on scion performance (Webster 2004). The majority of

these theories suggest that the rootstock and interstock bring about their effect upon the scion by influencing the amount and/or ratios of growth promoting and inhibiting endogenous hormones and their distribution within the tree. The movement of assimilates (sugars and amino acids) or mineral elements between the scion and rootstock, and the amount of water taken up and moved through the rootstock/interstock to the scion have also been considered (Lockard and Schneider 1981).

1.1.2.2 Effects of dwarfing rootstocks on scion development

The dwarfed growth habit is characterized by a reduced number of vegetative shoots and a corresponding increase in the number of fruit spurs (Eaton and Lapin 1970). Similarly, Costes and Garcia-Villanueva (2007) found that the two main effects of dwarfing rootstock on the aerial development of trees were a reduction in the number of axes developed per tree and an enhancement of flowering. Seleznyova *et al.* (2003) determined that the dwarfing effect of 'M.9' is mainly due to a reduced number of nodes per extension growth unit of the scion. As the number of extension growth units produced in the next cycle is proportional to this, it results in the expression of the dwarfing phenotype over time.

This reduction in the number of nodes per extension growth unit is not the only effect brought about by the rootstock and many other outcomes have been documented (Webster 2004). Rootstocks can delay scion vegetative budburst and the onset of shoot extension in spring, resulting in a shorter shoot growth period. They can significantly influence both rates of shoot extension and the timing of termination of active extension growth (Webster 1995). Rootstocks influence tree development by having a significant effect on the degree of scion branching, the individual length of shoots (Warner 1991) and the angle of branching for scions with upright growth habit (as opposed to spreading growth habit) (Tworkoski and Miller, 2007). Seleznyova *et al.* (in press) also demonstrated that dwarfing rootstocks and interstocks accelerate the natural process of tree aging by reducing both the proportions of extension shoots and their cyclicity commencing in the second year of tree growth. Many studies have shown that branches grow less vigorously when they are inclined, either by tree training or naturally, towards the horizontal (Webster 2004). Rootstocks may also have an indirect consequence on tree habit by affecting the quantity of blossoms produced and the precocity of cropping: heavy or light fruit

1.1.2.4 Rootstock effect on resistance to diseases and gene expression in apple tree scion

The susceptibility of different rootstocks and scion cultivars to various diseases can vary substantially (Wilcox 1994). A given cultivar can have different levels of disease resistance depending on the rootstock to which it is grafted. For example, 'Delicious' cultivar scions grafted onto 'M.9' rootstocks have a higher susceptibility to fire blight compared to the same cultivar grafted onto 'M.7' rootstocks (Carlson *et al.* 1970; Cline *et al.* 2001). In general, it is recommended that scion cultivars susceptible to fire blight be grafted to resistant rootstocks (Jensen *et al.* 2003). This indicates that the rootstock may influence the expression of resistance genes in the grafted scion. Jensen *et al.* (2003) studied the level of expression of genes in 'Gala' scions grafted onto different rootstocks. Scions grafted to the dwarfing rootstock 'M.9 T337' (clone of 'M.9') showed higher expression of genes related to photosynthesis, transcription/translation and cell division, while scions grafted to the intermediate rootstock 'M.7 EMLA' showed increased stress-related gene expression. Some of these differentially expressed genes may be linked to the dwarfing effect of the 'M.9' rootstock.

1.1.2.5 Effect of rootstock organs on scions

When considering possible mechanisms of rootstock influence on scion development, it may be useful to evaluate which parts of the rootstock may be involved in the dwarfing effect.

The roots are the most important part of a rootstock and may contribute significantly to the dwarfing effect. Roots of dwarfing rootstocks have increased proportions of phloem and bark and reduced proportions of xylem (Beakbane 1949; 1953). Also many dwarfing rootstocks have smaller and shallower root systems than invigorating rootstocks.

Reports in the literature have often suggested the implication of the graft union in the dwarfing effect of some rootstocks. Its potential role in the dwarfing effect is presented in section *1.1.3.3.1.*

The stem or shank of the rootstock contributes significantly to the dwarfing effect, and its role is described below (section *1.1.3.3.2.*).

1.1.3 Hypotheses explaining the dwarfing effect of apple rootstocks

1.1.3.1 Changes in hormones production and translocation

The dwarfing effect of apple rootstocks is a complex subject that has been the object of a multitude of studies. The most investigated hypotheses concern the changes in hormone concentrations between the rootstock and the scion.

Plant hormones, also known as phytohormones, plant growth substances or plant growth regulators, are naturally occurring substances which, in low concentrations, regulate plant functions. Plant hormones encompass all compounds, natural and synthetic, which when applied to plants evoke a specific physiological response (Tomić, 1998). Five 'classical' endogenous plant hormones (groups) exist: auxins, gibberellins, cytokinins, abscisic acid and ethylene. All these hormones influence vegetative growth mainly by affecting cell division and elongation (Kende and Zeevaart, 1997).

Hartmann and Kester (1990) have hypothesised that dwarfing rootstocks would limit scion growth because of their reduced production of growth promoting hormones (mainly auxins, cytokinins and gibberellins), or by lowering the basipetal auxin transport (from the stem to the roots) in their tissues.

1.1.3.1.1 Auxins

1.1.3.1.1.1 Site of synthesis and role in plant development

Auxins, of which indole-3-acetic acid (IAA) (Thimann 1977) is the most abundant, are synthesized predominantly from tryptophan in growing regions of plants such as apical meristems and young roots (Kerr and Bennett 2007).

Auxins have a wide range of effects on plant development. They influence vascular tissue differentiation, flowering and fruit development, root and shoot development, phototropism, gravitropism and senescence, by up-regulating the transcription of genes whose promoter regions contain auxin-responsive elements (Kerr and Bennett 2007). This hormone may thus play a significant role in growth regulation of dwarfing apple rootstocks and scions.

1.1.3.1.1.2 Auxin translocation and effect on stem vascular differentiation

One of the most studied roles of auxins in dwarfing apple rootstocks is their effect on the rootstocks stem vascular tissues. Auxins have been shown to be important in stimulating cambial activity and xylem development in many species (Digby and

Wareing 1966; Zarkrewski 1983). In 'M.9' rootstocks, the xylem linking the bud to the rootstock contains fewer and smaller vessels (Figure 1.4) than in a rootstock with intermediate vigour such as 'MM106' (Soumelidou *et al.* 1994 a). This difference may indicate that the level of growth regulators, such as auxins, reaching these tissues in 'M.9' is lower than that in 'MM106'.

To summarize this hypothesis, it may be a failure of auxins from the bud to enter rootstock tissues, via the graft union, in sufficient quantities that is responsible for the low levels of xylem formation in the rootstock. Dwarfing rootstocks may also have an inactivation mechanism for metabolising auxin into an inactivate form (Soumelidou *et al.* 1994b), resulting in low levels of xylem formation. This in turn leads to a reduced supply of water and minerals to the scion with consequent reduction of scion growth (Soumelidou *et al.* 1994 a).

1.1.3.1.1.3 Auxins effect on root growth

The level of auxins in root tissues influences root growth by stopping or promoting organ growth at various stages of development (Celenza *et al.* 1995), which in turn influences the synthesis of other hormones such as cytokinins and gibberellins, which are then exported to the shoot via the xylem (Muday and Haworth 1994).

As plants maintain a constant shoot-root ratio, auxins produced in the shoot and translocated to the roots may be a mechanism for control of root growth by the shoot (Torrey 1976; Goodwin *et al.* 1978). A reduction in the level of shoot synthesized auxin reaching the root tissue, via a disorganised vascular system or through metabolic inactivation by the rootstock stem tissues (see section 1.1.3.1.1.2), may reduce root growth which would in turn reduce scion development.

1.1.3.1.1.4 Auxin effect on cell extensibility and metabolite availability

One important role of auxins is to regulate cell extensibility. Auxins may activate the enzyme dextranase that would break down a highly branched compound (dextran) located in the cell walls. The break down of this compound would result in an increase of the wall elasticity that is necessary for the cell elongation (Heyn 1970). A decrease in auxin concentration would affect cell extensibility and thus shorten internodes, which may be a factor contributing to dwarfed scions.

Auxins are also involved in carbohydrate metabolism (Colby 1935) and nutrient translocation (Davies and Wareing 1964). A reduction in auxin concentration would upset these mechanisms and potentially lead to a reduced plant growth (Figure 1.4).

1.1.3.1.2 Cytokinins

1.1.3.1.2.1 Site of synthesis and role in plant development

Cytokinins are synthesized primarily in roots and are translocated through the xylem sap to the shoot buds where they promote shoot growth and development (Kende 1965; Carr 1966; Itai and Vaadia 1965; Torrey 1976). Zeatin and zeatin riboside (ZR) constitute a significant proportion of xylem-mobile cytokinins in many plant species, including apple (Kamboj *et al.* 1999a).

Cytokinins mainly promote cell division and differentiation (organogenesis). These hormones enhance the "sink" effect of the tissues (Morris and Winfield 1972), which makes the shoot tip a more efficient competitor for carbohydrates and amino acids. Cytokinins may also affect cell membrane integrity (Shaw and Manocha 1965) and facilitate the movement of compounds in the tip region (Turvey and Patrick 1979). Despite numerous studies, little is known about the response on the whole plant to cytokinins as its actions are often dependant on interactions with other hormones (Hooijdonk *et al.* 2006).

1.1.3.1.2.2 Interaction of cytokinins with auxins

Cytokinins interact with other phytohormones, primarily with auxins. Production and translocation of cytokinins acropetally through the xylem is dependant on the amount of shoot-synthesized auxins reaching the roots (Lockard and Schneider 1981). Cytokinins and auxins levels are inversely correlated *in vivo* (Eklof *et al.* 2000) and auxin treatment can rapidly inhibit cytokinins by suppressing both the pool size and the synthesis of these hormones (Nordstrom *et al.* 2004).

1.1.3.1.2.3 Effects on scion growth

Several authors have observed a positive correlation between the growth potential of a rootstock and cytokinin levels in its xylem sap (Jones 1986; Kamboj *et al.* 1999a). The growth potential of apple rootstocks shows a positive correlation with the rate of cytokinin export from their roots (Kamboj *et al.* 1999a) (Figure 1.4). The differences in concentrations of cytokinins observed in the xylem sap among rootstocks could result from either selective transport or selective synthesis of cytokinins in the rootstocks (Kamboj *et al.* 1999a). In *Prunus*, the most vigorous rootstocks have a higher level of cytokinins compared to dwarfing rootstocks (Sorce *et al.* 2001).

In ungrafted rootstocks, shoot and root growth are well balanced, and the level of cytokinins is only slightly lower than auxins. In grafted plants, a dwarfing rootstock could upset such a balance, resulting in the long term in a reduced stature of the scion grafted onto it (Sorce *et al.* 2001).

1.1.3.1.3 Gibberellins

1.1.2.1.3.1 Site of synthesis and role in plant development

Gibberellins (GAs) are a complex family of tetracyclic diterpenoid plant hormones that mediate environmental and developmental signals (Bulley *et al.* 2005).
GAs are synthesized in shoot and root tips and have been detected in the phloem and xylem of a wide range of plants. GAs act by regulating cellular processes such as cell elongation and division and control seed germination, stem elongation, leaf expansion, trichome development as well as fruit and flower development. The level of bioactive GAs is controlled by several mechanisms such as transcriptional regulation of genes encoding enzymes from both catabolic and biosynthetic pathways (Olszewski *et al.* 2002).

1.1.2.1.3.2 Effect on scion growth

Many of the growth, flowering and fruiting characteristics of dwarfed apple trees, and their response to applied GAs, suggest that reduced levels of endogenous bioactive GAs are involved in the dwarfing mechanism (Robitaille and Carlson 1971; Richards *et al.* 1986) (Figure 1.4).
Richards *et al.* (1986) established that the whole of the 'M.9' dwarfing interstock is involved in the reduced transport of GA and its metabolites to the shoot. 'M.9' dwarfing interstocks have a tendency to overgrow rootstock and scion, indicating that 'M.9' tissues are either especially responsive to elevated GA levels or retain more GAs (Richards *et al.* 1986).
The down regulation of GA 20-oxidase, which catalyses the penultimate step in the formation of bioactive GAs, results in dwarfing of an apple scion independent of the rootstock on which it is grafted (Bulley *et al.* 2005). The dwarfing effect, in this transgenic plant, is a result of reduced level of bioactive GA in shoot tips. When transgenic scions were grafted onto vigorous rootstocks, the dwarfing effect was maintained, demonstrating that scion vigour can be controlled independently from the rootstock.

1.1.3.1.4 Abscisic acid

1.1.2.1.4.1 Site of synthesis and role in plant development

Abscisic acid (ABA) is a 15-carbon compound (a sesquiterpene), which is synthesized in a pathway that involves carotenoid intermediates. This hormone can be synthesized throughout the plant, where there are plastids (any of several pigmented cytoplasmic organelles found in plant cells) (Milborrow 2001), but is primarily made in roots. Following its synthesis, ABA can be transported through the xylem or the phloem (Hartung *et al.* 2002).

ABA has been reported to have multiple roles during the life cycle of plants, such as growth inhibition, gravitropism, stomatal closure and water relations, seed development and bud dormancy (Zeevaart and Creelman 1988).

1.1.2.1.4.2 ABA and relation to dwarfing

In apple, concentrations of ABA-like substances in some dwarfing rootstocks are reportedly higher than those in vigorous rootstocks (Yadava and Dayton 1972). Higher ABA levels have also been measured in stems of dwarfed apple scions compared to vigorous ones (Robitaille and Carlson 1976). These observations point to a possible implication of ABA in the dwarfing effect of apple rootstocks (Figure 1.4). Furthermore applying ABA by injection into the stem reduces shoot size in apple (Kim *et al.* 1984). High concentrations of ABA in dwarfing rootstocks could inhibit the transport or reduce the effects of other hormones such as auxins, cytokinins and gibberellins (Basler and McBride 1977; Jacqmar *et al.* 1995).

High ABA concentrations, which in turn increases the ABA:IAA ratio, also increase the synthesis of the rootstock bark (Kamboj *et al.* 1999b) caused by a greater differentiation of phloem and related tissues. The high bark to root ratio resulting from the increase in ABA concentrations has been used as a marker for the early selection of dwarfing rootstocks in rootstock breeding programmes (Rogers and Beakbane 1957).

1.1.3.2 Influence of phenolic compounds

Phenols are low molecular compounds ubiquitous in all tissues of higher plants which have a significant impact on plant development (Makoi and Ndakidemi 2007). Some phenolic compounds such as tyrosine and 4-hydroxybenzoic acid have been

found to promote bud formation in presence of kinetin and IAA, while ferulic and protocatechuic acids strongly inhibited it (Lee and Skoog 1965).

In apple, phenols are concentrated mainly in the bark of trees and have been regarded as possible growth controlling compounds in the control of tree size by dwarfing rootstocks. Some phenolic acids can enhance oxidative decarboxylation of auxins, resulting in the inhibition of tissue growth (Tomaszewski and Thimann 1966). The same hypothesis has been developed to explain the dwarfing effect of some citrus rootstocks (Mendel and Cohen 1962).

Yu and Carlson (1975) proposed that phenols may be involved in graft incompatibility. Noggle (1979) suggested that when freshly cut surfaces of stock and scion are brought together, the cells of the adjacent graft may be subjected to foreign phenolic compounds and may not have the appropriate enzymes to break down these compounds or convert them into non-toxic forms. These free phenols may then inhibit cell division, which would have the effect of a non-union or bad union of the graft (Figure 1.4). This would lead in the first case to the death of the plant, and in the second case to a disorganisation in the graft union that would disrupt the formation of xylem and phloem, and as a consequence disturb the translocation of hormones, nutrients and water.

1.1.3.3 Anatomical and physiological hypotheses

1.1.3.3.1 Anatomy of the graft union: role in mineral and water translocation

Abnormal features in the graft unions of apple trees have been reported to be associated with different degrees of dwarfing imposed by the rootstocks. Studies on the graft union showed that the vascular tissues that develop between the stock and the scion are arranged in a twisting pattern, and become necrotic during subsequent growth of the plant. It has been suggested that these atypical anatomical features are the result of incompatibility of the bud and rootstock tissues (Simon and Chu 1984).

Studies on the total solute and nutrient content of the xylem sap below and above the graft union have revealed that there is a depletion of sap nutrients, together with cytokinins, at the graft union with dwarfing rootstocks and interstocks (Jones 1984). The graft union clearly has a role to play in the dwarfing phenomenon.

1.1.3.3.2 Root anatomy of dwarfing rootstock

Many dwarfing rootstocks have naturally small root systems, which reduces the uptake of water and minerals and the production of hormones (Figure 1.4). Young apple trees grown within root restriction membranes are very similar in size irrespective of whether grown on 'M.9' or 'MM106' (Webster 1995). These results demonstrate the importance of the root system in the mechanism by which rootstocks dwarf scions.

The hydraulic conductivity of roots from apple dwarfing rootstocks is lower than those measured from vigorous rootstocks (Atkinson *et al.* 2003). By consequence, the amount of water and minerals reaching the shoot would be lower for trees grafted on dwarfing rootstock, and would result in a slower growth of the scion. The same result was found by Syvertsen and Graham (1985) on citrus dwarfing rootstocks.

1.1.3.4 Pathological hypotheses

Virus and mycoplasma-like organisms have been shown to affect the growth, yield and quality of apple cultivars (Posnette *et al.* 1963). Scions of the same cultivar are more vigorous when grafted onto virus-free rootstocks compared to their equivalent virus infected clones (Campbell 1980). However, smaller infected trees crop as well as the larger healthy ones in proportion to their size. Nevertheless, the effect of virus and micoplasma-like organisms is not consistent with the inheritance of the dwarfing character, and may only be a contributing factor to the dwarfing effect of some apple rootstocks.

1.1.3.5 Dwarfing hypotheses: summary

Despite all the studies and hypotheses considered over the past 100 years, the dwarfing phenomenon is still not fully understood and the cause and genetic basis of this effect have not yet been revealed. Dwarfing effects are probably associated with disturbances in the metabolism and translocation of auxins, cytokinins and gibberellins. Although their action is still unclear in the dwarfing response, growth inhibitors such as ABA and phenolic compounds possibly have a supplementary role to play. Rootstock anatomy, which includes the size of the root system, the stem and the graft union, may also partially explain the dwarfing effect of some rootstocks. Finally a form of incompatibility might explain some growth control associated with certain rootstocks but this cannot explain the effects induced by the majority of dwarfing rootstocks (Yu and Carlson 1975).

It is possible that the cause of dwarfing is not controlled by a single factor, and a combination of hypotheses may prove to be true. Any alteration in the rootstock/scion system, combined with the intrinsic effect of the rootstock can lead to a cascade of reactions that can dramatically transform the physiology of the plant and lead to the development of a dwarfed tree.

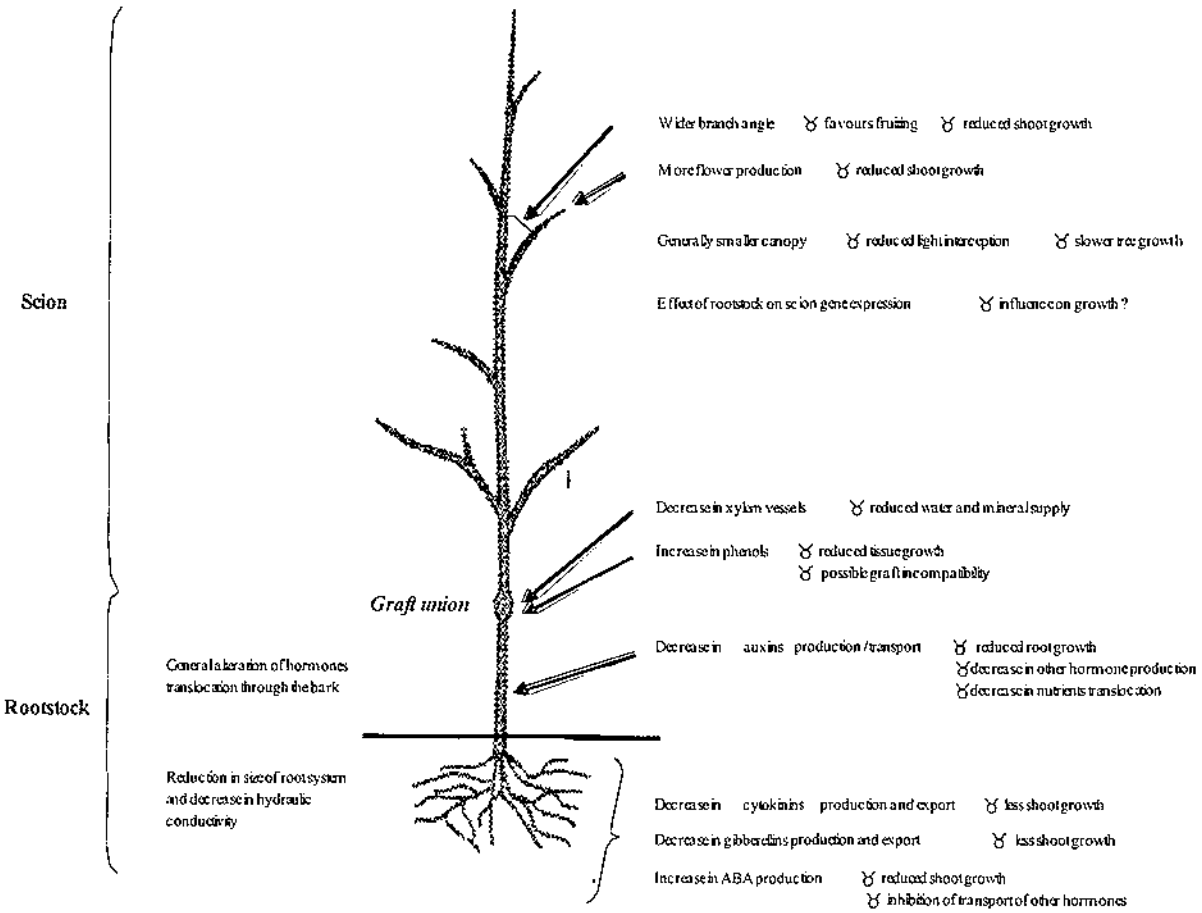

Figure 1.4. Summary of the major dwarfing hypotheses.

Until now, only physiological aspects of the dwarfing effect of some apple rootstocks have been considered, each author demonstrating the viability of their own hypotheses. Such approaches may have led to a restricted overview of the dwarfing phenomenon, and the variations observed by these authors between dwarfing and vigorous rootstocks may only be direct consequences of a more subtle alteration. A more general approach is certainly required to understand this complex phenomenon, which would require no underlying assumption concerning the cause of this effect. Such study could be conducted with the aim of describing the process

by which the scion is dwarfed. This type of study would have the advantage of not having a specific hypothesis of what the control mechanism may be, but would examine all the contributing factors of a plant structure and anatomy that may be implicated. Studies could also be conducted with the aim of understanding the genetic basis of dwarfing, which would eventually lead to an understanding of the physiology of the trait. Methods such as genetic mapping and other genomic approaches (closely linked with parallel physiological studies), which were unknown to researchers only a few decades ago, are now available. These methods, which include bulked segregant analysis, QTL analysis and candidate gene approach via microarray analysis will be described in the next paragraphs of this thesis.

1.1.4. Knowledge on the genetic control of the dwarfing effect by the apple rootstock 'M.9'

Bulked segregant analysis (BSA, see section 1.2.3.1) using randomly amplified polymorphic DNA (RAPD, see section 1.2.1.1) primers has already been performed in an F1 population derived from a cross between the dwarfing rootstock 'M.9' and the vigorous rootstock 'R5', and several markers linked to a major dwarfing locus (*DW1*) have been identified (Rusholme Pilcher *et al.* in press). Using simple sequence repeat (SSR, see section 1.2.1.4) markers developed by Liebhard *et al.* (2002), *DW1* was mapped to the top of the linkage group 5 (LG5) on the apple genome. However, the mapping of *DW1* was determined using only individuals with extreme phenotypes, and vigorous individuals amplifying a marker linked to *DW1* were not included in the analysis, which may have led to an incorrect estimation of the genetic distance between *DW1* and its closest markers (CH03a09).

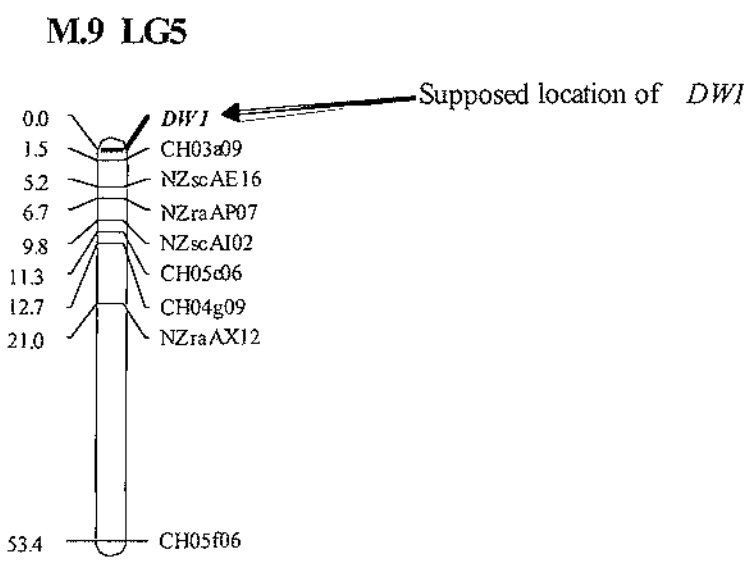

Figure 1.5. Location of *DW1* on the linkage group 5 of 'Malling 9' (from Rusholme *et al.* 2004). The names of the SSR markers (CH) from the literature were not changed. The sequence characterized amplified regions (SCAR) markers developed in HortResearch are prefixed with 'NZsc'. The RAPD markers are prefixed with 'NZra' followed by the Operon primer code. Numbers on the left side of the linkage group represent the distance in cM.

The absence of genetic markers on both sides of DW1, together with the fact that DW1 alone does not explain all the variation in the scion architecture, does not permit the implementation of a robust marker assisted breeding programme.

1.2 Genetic mapping and apple genetics

1.2.1 Molecular markers and mapping

Molecular markers are genetic markers that are based on DNA sequences, as opposed to morphological (expression of a trait), biochemical (chemical composition) and protein (isoenzymes) markers (Jones *et al.* 1997a). Markers occupy specific positions in the genome which are called loci (singular locus). Molecular markers can be coding or non-coding, they can be anonymous or of known sequence, they can be based on their length or on variations in their sequence and they can be dominant or codominant. In the following paragraphs, I will limit my description to the markers used in the context of this thesis, with a special emphasis on microsatellite markers.

1.2.1.1 Randomly amplified polymorphic DNA (RAPD)

RAPD markers are generated from PCR primers of 10 arbitrary nucleotides in length. Each product is derived from a region of the genome that contains two short segments in inverted orientation, on opposite strands, that are complementary to the primer and sufficiently close together for the amplification to work. RAPD primers generally amplify 5 to 12 DNA strands of between 200 and 2000 bp long, which lie between two inverted copies of the primer, one copy binding to each strand of the DNA. Priming occurs statistically once every million base pairs (Jones *et al.* 1997b). Polymorphisms arise because of sequence variations in the genome that alter the primer binding sites. RAPD markers are termed dominant markers because of their presence or absence at particular loci. Amplification products are electrophoretically resolved on agarose gels with staining using ethidium bromide.

RAPD markers are simple and less expensive than other markers such as Restriction Fragment Length Polymorphism (RFLPs) (based on the use of labelled DNA probes annealing genomic DNA cut by restriction enzymes) for the reason that no prior knowledge of sequence is required. There is no limit to the number of RAPD markers in a genome, making them very useful for mapping (Jones *et al.* 1997b). However, these markers have disadvantages: they cannot easily be transferred among species, because of the random nature of their generation and their short primer length, they have poor reliability and reproducibility, and they have a high sensitivity to experimental conditions (e.g.: concentration of salts, dNTPs, DNA) (Karp *et al.* 1996). Nevertheless they have proved useful in many studies in apple for the identification of molecular markers linked to resistance genes such as apple scab (*Venturia inaequalis*) (Yang *et al.* 1997; Gygax *et al.* 2004) and powdery mildew (*Podosphaera leucotricha* (Ell. *et* Ev.) Salm.) (Dunemann *et al.* 2004; James and Evans, 2004). RAPD markers have also been used for genetic map construction in apple (Hemmat *et al.* 1994).

1.2.1.2 Sequence characterized amplified regions (SCAR)

SCAR markers are derived from the sequencing of PCR products obtained from RAPD markers. Knowledge of the DNA sequence allows longer primers (from 20 to 25 bp) to be designed with the aim to make the amplification more specific. DNA sequence differences are manifested by the presence or absence of a single unique band. SCAR markers are more reproducible than RAPD markers and are usually

dominant markers (Staub *et al.* 1996). Like RAPD, SCAR markers are resolved using electrophoresis on agarose gels with staining using ethidium bromide.

1.2.1.3 Single nucleotide polymorphism (SNP)

Single DNA base differences between homologous DNA fragments, including small insertions and deletions (indels), are referred to as single nucleotide polymorphisms (SNP) (Zhu *et al.* 2002). In maize, the average is one SNP per 48 bp in non-coding regions, and one SNP per 131 bp in the coding regions (Rafalski 2002). Indel polymorphisms are present on average one in 126 bp, but almost exclusively in non-coding regions (Rafalski 2002). In apple, the average occurrence of SNP markers in coding regions is about one per 107 bp (HortResearch, unpublished data). SNP markers are usually bi-allelic and are very abundant throughout the genome. Genotyping of SNP generally involves the generation of allele-specific products of SNP of interest followed by their detection for genotype determination (Kim and Misra 2007). The majority of current genotyping technologies require a PCR amplification step. Several SNP genotyping technologies based on allele discrimination strategies have been described to characterise SNP polymorphism. These technologies are based on allele-specific PCR amplification or allele-specific biochemical reactions. Four popular methods have been developed: primer extension, hybridization, ligation and enzymatic cleavage (Kim and Misra 2007).
SNP markers have a wide range of applications such as construction of genetic maps, candidate gene mapping, genetic diagnostics and analysis of the genetic structure of populations (Batley and Edwards 2007).

1.2.1.4 Simple sequence repeats (SSR)

Simple Sequence Repeats (SSR) or microsatellites have been detected in the genome of every organism analysed so far, and are found at frequencies much higher than would be predicted purely on the grounds of base composition (Epplen *et al.* 1993). SSR markers are relatively rare in protein-coding regions but constitute a large fraction of non-coding DNA (Li *et al.* 2002a). They are an important class of DNA markers because of their abundance and length hypervariability. SSR markers have been used in cultivar identification, genetic analysis, genetic diversity analysis and genetic mapping (Guilford *et al.* 1997).

1.2.1.4.1 SSR definition and putative function

SSR markers are short tandem repeats of one to six base pairs with a maximum number of repetitions rarely greater than 60 (Taylor *et al.* 1999). SSR markers are generally classified in three categories following the nomenclature of (Weber 1990): 'perfect', when the motif is repeated without any point mutation (Figure 1.5 a); 'imperfect', where a point mutation is present in the repeat (Figure1.5 b); and composite (or compound), where the SSR is made up of two or more different types of motif (Figure 1.5 c).

(a)- Perfect dinucleotide repeat.

GTGTGTGTGTGTGTGTGTGT
CACACACACACACACACACA

(b)- Imperfect nucleotide repeat: contains a point mutation.

GTGTGTGTGTGGGTGTGTGT
CACACACACACCCACACACA

(c)- Composite SSR: transition from GT/CA to GA/CT.

GTGTGTGTGTGTGAGAGAGAGAGA
CACACACACACACTCTCTCTCTCTC

Figure 1.6. Diagram representing the different types of SSR. Panel (a) illustrates perfect repeats, where the SSR consists of a repeated motif; panel (b) illustrates imperfect repeats, where a point mutation is present (highlighted); and panel (c) illustrate a composite SSR repeat where there is a transition to another repeat motif (highlighted).

Even though SSR markers are generally considered as evolutionary neutral DNA markers, critical tests in various biological phenomena have proved the functional significance of a substantial part of them. Some SSR markers are probably involved in chromatin organisation, in the regulation of DNA metabolic processes (DNA

replication and recombination) and in the regulation of gene activity through transcription, protein binding and translation (Li *et al.* 2002b).

1.2.1.4.2 SSR Polymorphisms and mutation

The high polymorphism found in SSR markers is due to their high mutation rate. In plants this mutation rate is often quoted in the range of 10^{-2} to 10^{-6} per locus per generation. To make a comparison, eukaryotic DNA sequences mutate at a rate of approximately 10^{-9} per nucleotide per generation (Crow 1993). Mutation rate can vary among species as well as among loci. One major determinant of this variation is the length of the individual SSR. Long SSR repeats mutate more often than short ones. However other factors affect the rate of mutations at SSR loci including C/G content in flanking DNA, chromosome position, cell division (mitotic vs. meiotic), sex and genotype (Li *et al.* 2002b).

The majority of SSR mutations represent gains and losses of entire repeat units. Two mutational mechanisms can be invoked to explain these high rates of mutation. The first and the most important one involve DNA slippage during DNA replication. The second involves recombination between DNA strands.

Polymorphism of SSR markers depends on the size of the amplified fragments. Bands of different size for the same SSR primers can be considered alleles of that specific locus (Serrano *et al.* 2002).

1.2.1.4.3 Repeat type and repeat length

The majority of SSR markers found in many species are dinucleotides (48-67%) (Schug *et al.* 1998). The dinucleotide repeat SSR markers occurring with the highest frequency in plants are (AT)/(TA), with (AG)/(CT) and (AC)/(GT) as the second and third most frequent (Wang *et al.* 1994). Trinucleotide and tetranucleotide repeats are not as common but tend to generate fewer stutter bands, which makes them easier to score.

Di- and tri-nucleotide SSR markers occur in the apple genome at a frequency comparable with other plant species. The (GA) motif for example occurs about every 120 kb, compared to about every 225 kb in rice and 100 kb in tropical tree genomes (Condit and Hubbell 1991; Wu and Tanksley 1993). The GT repeats occur in the apple genome about every 190 kb (Wu and Tanksley 1993).

1.2.1.4.4 SSR amplification

SSR markers are flanked by specific regions that can be used to develop primers in PCR reactions (Serrano *et al.* 2002). In apple, the flanking regions of SSR markers tend to be conserved and the markers already developed work reliably on a wide range of cultivars (Liebhard *et al.* 2002).

SSR alleles can be distinguished from PCR artefacts by the presence of stutter bands. These stutter bands appear more often in markers that contain simple di-nucleotide repeats. These bands are amplified because of a slippage of the polymerase during the amplification, which result in the production of fragments that are reduced in length by a multiple of repeat units (Smeets *et al.* 1989).

1.2.1.4.5 Null alleles and heteroduplex formation

As mentioned above, SSR detection is PCR based. It is a requirement that both PCR primers match the flanking regions of all alleles. If a point mutation or a deletion is present in the primer binding site of a specific allele, that allele will not be amplified. Such alleles are called 'null alleles' (Schlotterer 1998).

Heteroduplex formations are generated by base pairing between complementary single strands derived from the different parental duplex molecules generated during genetic recombination (Ayliffe *et al.* 1994). The use of denaturating gels such as polyacrylamide gels eliminates heteroduplex formations, resulting in an easier analysis of the results. With non-denaturating gels, such as agarose gels and other types of gels such as the one used in the capillary electrophoresis system CePro 9600™ (Combisep, Ames, Iowa), heteroduplexes can be visualised as extra bands, usually of higher molecular mass. This apparent shift in the molecular mass of the heteroduplexes is not due to an increase in the length of the amplification products, but rather to the formation of nucleotide loops (because of the annealing of two alleles of different length), which slow down the migration of DNA strands through the gel during electrophoresis.

1.2.1.4.6 Methods for obtaining SSR flanking primers

1.2.1.4.6.1 Markers developed from SSR libraries

The traditional method used to obtain SSR flanking regions is to screen by probing genomic libraries with SSR sequences (Rassmann *et al.* 1991). Positive clones are then sequenced to identify the SSR and flanking regions for primer design. This

method has a low rate of SSR recovery and is very expensive and time consuming (Zane *et al.* 2002). Other methods involving repeated enrichment of a given library followed by cloning and sequencing have been widely used for SSR primer design (Zane *et al.* 2002). The majority of SSR developed in apple were developed using these methods (Liebhard *et al.* 2002; Silfverberg-Dilworth *et al.* 2006). SSR markers developed from anonymous genomic sequences often correspond to non-coding sequences.

1.2.1.4.6.2 Markers developed from Expressed Sequence Tags (EST) sequences

ESTs are sequenced portions of complementary DNA copies of mRNA. They represent part of the transcribed portion of the genome of an organism grown under various experimental conditions (Poncet *et al.* 2006). Polymorphisms associated with these sequences have been found, including SNP markers, introns and SSR markers. Sequence analyses revealed a range from 2.65 to 16.82% of ESTs containing SSR motifs in dicotyledonous species (Kumpatla and Mukhopadhyay 2005). In apple, over 250,000 EST sequences are now publicly available (Genome Database for Rosaceae (GDR)). SSR markers are particularly common in the 5'-untranslated region (UTR) and, to a lesser extent, in the 3'-UTR of transcribed plant sequences (Morgante *et al.* 2002). Out of the 160,719 apple EST sequences analysed by Han *et al.* (2006), more than 2,000 (12.5%) were found to contain an SSR. Of the 150,000 ESTs recently added to the database, Newcomb *et al.* (2006) found that 17% of the apple sequences contained one or more di-, tri- or tetranucleotide SSR markers. These ESTs represent an extensive source of ready to use SSR markers that is still mostly unexploited.

Together with the qualities common to all SSR markers, EST-SSR markers can be used to cross-reference genes among species for enhancing the resolution in comparative genomic studies and identifying conserved genomic regions among species and genera (Brown *et al.* 2001; Decroocq *et al.* 2003).

1.2.1.4.7 Transfer of SSR markers among species

For genomes with no or little DNA sequence available, the development of SSR markers can be expensive and time consuming. An alternative approach is to use SSR markers developed in other related species. This approach has been successful in many different plant species including *Eucalyptus* (Yasodha *et al.* 2005), *Pinus* (Chagné *et al.* 2004) as well as different species of bird (Eggert and Fleischer 2004)

and fish (Yue and Orban 2004). The major problems in transferability are the possible point mutations in the primer binding sequence and more drastically the complete absence of the locus.

Within the Rosaceae family, this approach has been successfully used. Several *Malus* SSR markers have been mapped in *Pyrus* (Yamamoto *et al.* 2004a) and *Cydonia* (Yamamoto *et al.* 2004b). *Prunus* SSR markers have been used in *Pyrus* (Yamamoto *et al.* 2004a) and *Fragaria* (Santiago *et al.* 2007) for the purpose of map alignment.

1.2.1.4.8 SSR application to apple

SSR markers have proved valuable in apple for framework map construction, marker assisted selection and cultivar fingerprinting. To date, over 300 SSR markers have been developed in apple, mainly from anonymous sequences (Hokanson *et al.* 1998; Liebhard *et al.* 2002; Silfverberg-Dilworth *et al.* 2006). As the majority of commercial cultivars of *Malus* × *domestica* are diploid, no more than two SSR alleles should be expected (Guilford *et al.* 1997). However, some SSR primer pairs can amplify more than two alleles for individual cultivars. These banding patterns may reveal ancestral chromosome duplication events (see section 1.2.2.4).

1.2.2 Gene mapping

1.2.2.1 Principle

Mapping implies measuring the relative genetic distances between markers and assigning them to their linkage groups on the basis of the recombination values from all their pairwise combinations. Mapping is based on the principle that genes and markers segregate via chromosome recombination (designated as crossing-over) during meiosis (i.e. sexual reproduction) (Collard *et al.* 2005). Recombination is the process by which new combinations of parental genes or characters arise. It occurs by independent segregation of unlinked loci or by crossover between loci that are linked (Jones *et al.* 1997a). The distance between two markers is proportional, although not linearly related (Hartl and Jones 2001), to the recombination frequency between them, and is measured in centiMorgan (cM). The number of different linkage groups found corresponds, given enough markers, to the basic chromosome number of the species.

The construction of a linkage map requires a segregating population. Ideally the parents selected for this mapping population differ for one or more traits of interest. The number of individuals composing the segregating population generally ranges from 50 to 250, but larger populations are required for high-resolution mapping (Mohan *et al.* 1997).

The next step in the construction of a linkage map is to identify molecular markers that reveal differences between parents and that segregate among the individuals composing the progeny. Depending on the size of the genome to be mapped (in cM), the number of markers required to saturate the genome will vary. A genetic map is considered saturated when each marker in the genome is linked to at least one other marker on the map and when the number of linkage groups identified is equivalent to the number of chromosome in the genome studied (Mohan *et al.* 1997).

Once polymorphic markers have been identified and screened over the segregating population, mapping software programmes such as Mapmaker/EXP (Lander *et al.* 1987; Lincoln *et al.* 1993), MapManager QTX (Manly *et al.* 2001) and JoinMap (Stam and Van Ooijen 1995) are used to construct the genetic map. Linkage between markers is usually calculated using odds ratios (i.e. the ratio of linkage versus non linkage). This ratio is more commonly expressed as the logarithm of the ratio and is called logarithm of odds (LOD) value or LOD score (Risch 1992). The LOD score is a ratio obtained by dividing the probability that two loci are linked by the probability that they are not. LOD values >3 are usually used to construct linkage maps. As an example, a LOD value of three between two markers indicates that linkage is 1000 times more likely than no linkage. LOD values can be lowered in order to detect a greater level of linkage or to place additional markers within maps constructed at higher LOD values.

High-density linkage maps are an essential tool for the determination of marker-trait associations using the genome scanning approach (Patocchi *et al.* 2005). Genetic maps permit localisation of genes of interest, and the identification of quantitative trait loci (QTL), by providing the framework to understand the biological basis of complex traits (Tanksley *et al.* 1989). When multilocus markers such as SSR, RFLP and SNP markers are employed, genetic maps are invaluable for identifying homoeologous chromosomal regions (Liebhard *et al.* 2002; Maliepaard *et al.* 1998a). Furthermore, the use of orthologous markers (e.g. transferable microsatellites or EST-based markers) can make it possible to align framework maps to other species

maps using comparative genome mapping. Finally, linkage maps enable marker assisted selection (MAS) of favourable alleles in parents and progenies. A major advantage of the use of markers is that they increase breeding efficiency by enabling early selection for adult traits. MAS also enables simultaneous selection for multiple traits, including resistance gene pyramiding, and selection for traits that are expensive to phenotype (Gardiner *et al.* 2007).

1.2.2.2 Mapping strategy used in apple

Unlike many other crop species, where segregating populations are commonly derived from backcross (BC_1) or F_2 crosses, in apple and other outbreeding species, full-sib families are used for genetic analysis. This substitute mapping method is called double pseudo-testcross mapping method (Weeden 1994). In this procedure, highly heterozygous cultivars are crossed and independent maps are constructed for loci segregating from each parent. Hence, different types of segregation for markers and QTL can occur simultaneously. If the parents are heterozygous for different alleles at marker loci, then new segregation types will occur in the progeny (Figure 1.6). For dominant RAPD, markers heterozygous in only a single parent segregate in a 1:1 present: absent ratio in the progeny (a), and markers heterozygous in both parents segregate in a 3:1 present: absent ratio (b). Doubly heterozygous markers are less informative than single heterozygous markers because the dominant allele progeny are composed of three indistinguishable genotypes: + +, + –, and – +. For co-dominant markers such as SSR markers, the two alleles of each parent can be detected and segregation types such as 1:2:1, 1:1:1:1 or 3:1 can be obtained in the progeny (c). When both parents are informative for several loci, the male and female maps can be aligned and a consensus map of the cross can be drawn (Figure 1.7).

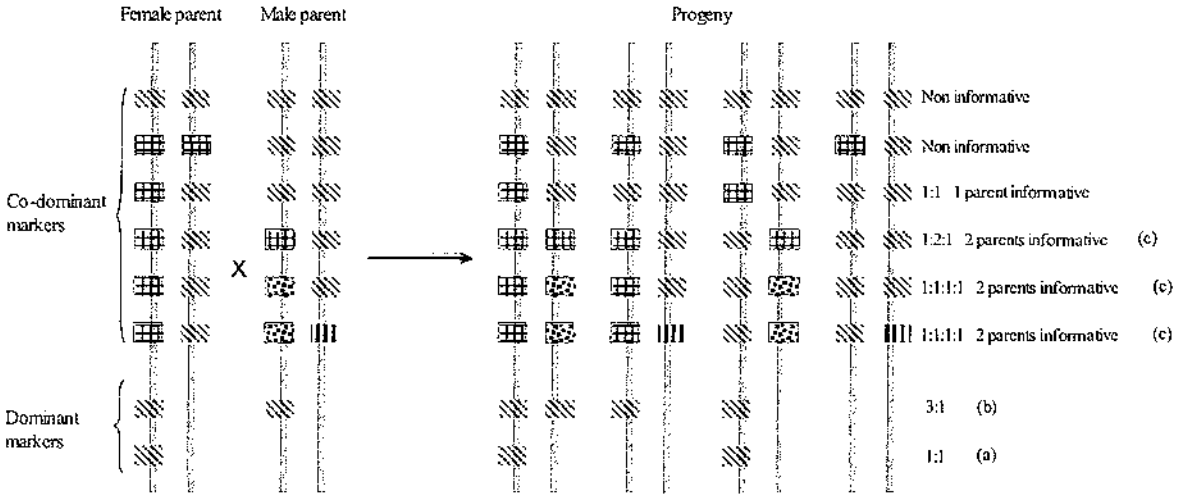

Figure 1.7. Double pseudo-testcross mapping strategy and different marker segregation types. ⊞ ⊠ ▯▯ ▨ represent different alleles of the loci.

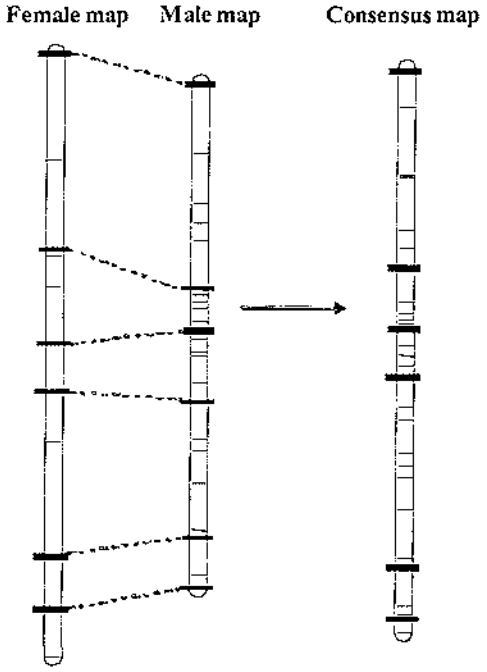

Figure 1.8. Construction of a consensus map. Male and female maps have been aligned using markers in common. Common markers among the maps are indicated by a bold line (━) and markers specific to each map are indicated by a simple line (—).

1.2.2.3 Genetic map construction in apple

The first genetic maps of apple, developed in the USA in the 1990s, included mainly RAPD markers and a small number of isoenzyme markers (Conner *et al.* 1997; Hemmat *et al.* 1994) (Table 1.1). Because of the poor transferability of RAPD markers, these maps were specific to the genetic background of the mapping parents. Later, genetic maps were constructed with co-dominant transportable markers, mostly RFLP markers and a few SSR markers (Maliepaard *et al.* 1998a). The first SSR markers mapped in apple included some of those identified by Guilford *et al.* (1997), Hemmat *et al.* (1997), and by Horticulture Research International (HRI). The alignment of 17 linkage groups, corresponding to the 17 chromosomes of the apple genome, and the calculation of the first integrated apple map was realised in a 'Prima' × 'Fiesta' population (Maliepaard *et al.* 1998a) (Table 1.1).

Currently, a genetic map constructed in a 'Fiesta' × 'Discovery' population of about 250 individuals contains the largest core of SSR markers. This map includes over 300 SSR markers identified by Liebhard *et al.* (2002), Gianfranceschi *et al.* (1998) and Silfverberg-Dilworth *et al.* (2006). The majority of these SSR markers are derived from enriched libraries, the rest is composed of apple EST-SSR and genomic SSR markers developed in *Pyrus* (Yamamoto *et al.* 2004a; Yamamoto *et al.* 2002a; Yamamoto *et al.* 2002b; Yamamoto *et al.* 2002c). The maps span a total of 1,145.3 cM ('Fiesta') and 1,417.1 cM ('Discovery') and the coverage is close to 100% (Silfverberg-Dilworth *et al.* 2006).

Table 1.1. Summary of published apple genetic maps. The population size (pop. size), number of markers for each parent, number of SSR markers mapped and length of the map (cM) for each parent is given.

Cross	Pop size	Number of markers		SSR markers mapped	Length of map (cM)	Reference
		Female	Male			
'Rome Beauty' × 'White Angel'	56	156	253	-	-, 950	Hemmat *et al.* 1994
'Wijcik McIntoch' × 'NY 75441-67'	114	238	110	-	1206 (integrated), 692	Conner *et al.* 1997
'Wijcik McIntosh' × 'NY 77441-58'	172	181	183	-	1206, 898	Conner *et al.* 1997
'Prima' × 'Fiesta'	152	194	163	10	842, 984	Maliepaard *et al.* 1998
'Fiesta' × 'Discovery'	112	202	227	118	914, 1015	Liebhard *et al.* 2002
'Fiesta' × 'Discovery'	267	439	499	129	1140, 1450	Liebhard *et al.* 2003b
'Fiesta' × 'Discovery'	44	-	-	No new SSR markers	1144, 1455 based on Liebhard *et al.* 2003b	Baldi *et al.* 2004
'Discovery' × 'TN10-8'	149	-	-	62	1,219 (integrated map)	Calenge *et al.* 2005
'Telamon' × 'Braeburn'	257	259	264	20	1039, 1245	Kenis and Keulemans 2005
'Fiesta' × 'Discovery'				168 new loci; total >300	1145, 1417	Silfverberg-Dilworth *et al.* 2006

1.2.2.4 Genome organisation and homoeology

There are several hypotheses about the origin of the domestic apple (see section 1.1.1.2), each involving a certain level of duplication within its genome. During the development of genetic maps, several authors have reported the mapping of loci, from the same marker, on at least two different linkage groups. Maliepaard *et al.* (1998) reported that linked sequences detected by RFLPs on one linkage group could also be found linked at another linkage group. The same observations were made by Liebhard *et al.* (2002) and Silfverberg-Dilworth *et al.* (2006) when some SSR markers amplified two loci mapping at different locations in the genome. These results suggest almost complete homoeology between some linkage groups (i.e. LG05 and LG10; LG13 and LG16) and partial homoeology among others (i.e. top of LG04 and top of LG06, bottom of LG02 and top of LG07). A partial map indicating homoeologous portions of the genome among 14 of the 17 linkage groups of apple was recently published by Van de Weg in Gardiner *et al.* (2007). These results also indicate that the ancestors of the original hybrid were closely related (Maliepaard *et al.* 1998a).

1.2.3 Methods used to map genes in apple

1.2.3.1 Bulked segregant analysis (BSA)

Developed by Michelmore *et al.* (1991), BSA is based on the comparison of marker amplification patterns between DNA bulks. The bulks are composed of approximately 10 individuals each, selected to have identical phenotype, and thus identical genotypes for a particular genomic region. The contrasting bulks are then screened to identify polymorphic markers differentiating them. The presence of polymorphisms between the amplification patterns of the two bulks (Figure 1.8) is expected only for those bands that are genetically linked to the gene of interest (Giovannoni *et al.* 1991) because markers differentiating the bulks are likely to be linked to the gene conferring the particular trait.

Polymorphisms between the DNA bulks result from either chromosomal changes in the amplified regions or base changes at the primer binding site (Michelmore *et al.* 1991).

Different types of markers can be employed to perform BSA, though RAPD primers provide the most efficient way of identifying new loci (Michelmore *et al.* 1991). Other more informative markers such as RFLPs, STSs (sequence tagged site) and SSR markers have successfully been used in BSA (Michelmore *et al.* 1991; James and Evans 2004; Rusholme Pilcher *et al.* in press). These markers offer the advantage of being transferable among crosses of different background and their location on the genome is known. When informative markers such as SSR markers are used in the context of BSA, the technique is often called whole genome scanning. James and Evans (2004) and Rusholme (unpublished) successfully used this technique to identify the location of some major resistance genes in apple.

The selection of individuals composing the bulks, based on their phenotype, must be meticulous because recombination between the target marker and the assayed polymorphic locus will result in diminishing distinction between the bulks (i.e. bulk of resistant individuals versus bulk of susceptible individuals). Thus, an accurate assessment of the desired trait is crucial to the success of BSA. In addition, the diminishing distinction between the two bulks may be the result of a decrease in the linkage of the marker to the gene.

Because of the relative DNA concentration of each of 10 individuals composing a bulk, segregating markers within a window of 10% recombination either side of the target locus will always be detectable, and many markers within 30% recombination window will also be detectable, at least as bands of unequal intensity (Michelmore *et*

al. 1991). In other terms, all markers closer than 15 cM are likely to be detected, whereas the limit of detection is located around 25 cM.

The probability of an unlinked locus being polymorphic between two bulks of 10 individuals was calculated to be 2×10^{-6} (Michelmore *et al.* 1991). Therefore, the frequency of unlinked loci will increase with the use of smaller bulks.

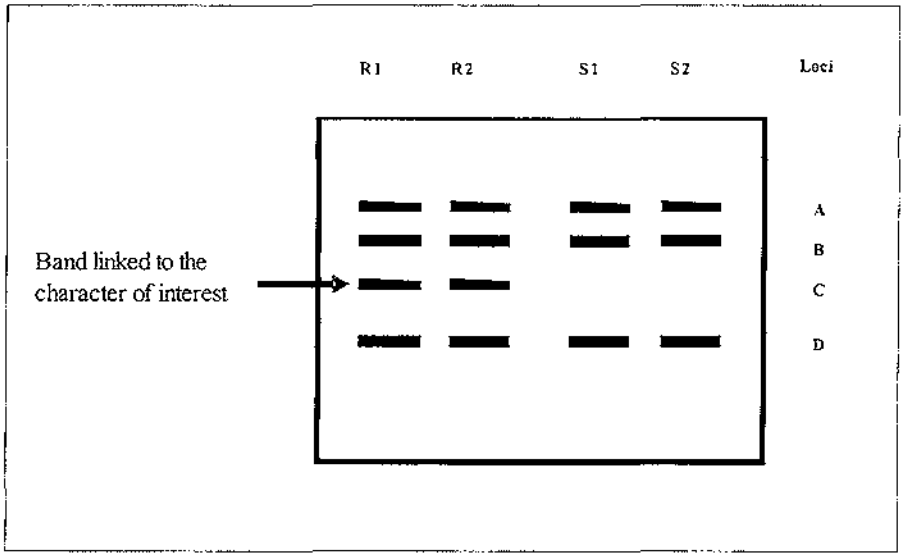

Figure 1.9. Illustration of bulked segregant analysis. Genotype of a RAPD marker screened on bulks derived from individuals homozygous for resistance (Bulks R1 and R2) or susceptibility (Bulks S1 and S2). The dominant allele at locus C is linked to the resistant allele (R) and therefore is polymorphic between the bulks. The other three loci (A, B and D) are not polymorphic between the 2 classes of bulks and therefore are unlinked to the R locus.

Markers showing polymorphism between the bulks are then screened over individual genotypes in the population and genetic distance between the trait and the marker is subsequently calculated (Michelmore *et al.* 1991).

BSA has been successfully used to find markers related to resistance genes in many species including lettuce (Michelmore *et al.* 1991.), tomato (Martin *et al.* 1991), bean (Haley *et al.* 1993) and apple (Markussen *et al.* 1995; Yang *et al.* 1997; Cheng *et al.* 1998; Gygax *et al.* 2004; James and Evans, 2004; Patocchi *et al.* 2004).

BSA has also been used to identify markers for major QTL (Quarrie *et al.* 1999). By grouping plants according to either high of low expression of a particular trait and extracting DNA from these two bulks, the process of genotyping the plants is reduced to only two DNA samples to be analysed instead of having to analyse DNA

independently from each of the plants composing the population (Quarrie *et al.* 1999).

1.2.3.2 Quantitative trait loci (QTL): principle

Kearsey and Pooni (1996) describe 'quantitative trait' as being a character for which the observed phenotypic variation is due to the segregation of several genes, and to the interaction between these genes and their environment. QTL analysis is based on the principle of detecting an association between phenotype and the genotype of markers (Collard *et al.* 2005).

Locating QTL provides a mean to answer fundamental questions about the genetic control of quantitative traits such as the number of genes involved in the control of the trait and the intensity of the effect attributable to each of these genes.

Knowledge of the location of QTL opens many opportunities to improve selection efficiency.

QTL analysis is a multi-step procedure. Each of the major components will be described in the following paragraphs.

1- Creation of a segregating population

Most commonly, a QTL mapping population is derived from the cross of two parental lines that show marked differences for the trait of interest. A typical QTL population consists of 100 to 300 individuals. The size of a population can greatly influence the outcome of a QTL analysis study. The precision of QTL location depends more on sample size than on the density of markers (Kearsey and Pooni 1996). The larger the population, the more accurate the mapping study and the more likely it is to allow detection of QTL with small effects (Haley and Andersson 1997; Tanksley 1993). An increase in population size provides gains in statistical power, estimates of gene effects and confidence of the locations of QTL (Beavis 1998; Darvasi *et al.* 1993).

2- Development of a genetic map (or linkage map)

The identification and localisation of QTL relies on the use of a linkage map covering the entire genome, with regularly spaced markers (15 to 20 cM between markers). Large gaps between markers on a linkage group can lead to inaccurate analyses. QTL located in regions that contain gaps between markers cannot be mapped precisely as their phenotypic effect will be underestimated because distant linkage cannot be differentiated from small phenotypic effect (Lander and Botstein

1989), and QTL on unmapped regions of the genome will be unobserved. According to Kearsey and Pooni (1996), no great increase in precision is obtained with more than five well-spaced markers per chromosome.

3- Phenotypic evaluation of the trait(s)

An essential part of QTL analysis is obtaining accurate estimates of the traits of interest for each individual in the mapping population. For most traits, this involves the use of randomized, replicated designs and measurement of multiple plants per line. Growing conditions and evaluation methods must be as uniform as possible across the whole population. Often a QTL population is evaluated in multiple locations or years, to determine whether the same or distinct QTL influence a trait under different environmental conditions (George *et al.* 2003; Hittalmani *et al.* 2002).

Following the collection of a set of phenotypic data, various analyses of this set can be performed prior to the actual QTL analysis. These include:

(i) Analysis of frequency distribution: graphing the number of individuals of a population that fall into different phenotypic classes is a useful place to start in examining phenotypic data. It is important to know if the frequency distribution of the population is normal or approximately so. If the trait is not normally distributed, transformation to obtain a more normal distribution might be considered; however, transformation may modify the data in such a way that genetic relationships may be obscured. Three main characteristics describe the frequency distribution of a trait: the degree of skewness which indicates whether phenotypic data tend to cluster on one end of the distribution curve or the other, the kurtosis which refers to whether the shape of a distribution is relatively short and flat, or tall and slender, or somewhere in-between those two extremes, and the modality which refers to the number of distinct peaks that appear within a distribution.

(ii) Estimation of heritability: The most basic question to be asked about a quantitative trait is whether or not the observed variation in the character is influenced by genes at all. Variations among phenotypes in a population arise from two sources. First, there are differences among the genotypes, and second, each genotype exhibits phenotypic variance because of environmental variation. The total phenotypic variation in a population (s^2_p) can be broken into two portions: the variance among genotypic means, also known as genetic variance (s^2_g), and the

environmental variance (s^2_e). The degree of heritability can be defined as the proportion of the total variance that is due to genetic variance:

$$H^2 = \frac{s^2_g}{s^2_p} = \frac{s^2_g}{s^2_g + s^2_e}$$

H^2 is called the broad sense heritability of the character, and ranges from 0 (variation entirely due to environment) to 1 (variation entirely due to genes). This measure of genetic influence indicates what proportion of the population's variation in phenotype can be assigned to variation in genotype. The higher the heritability estimate for a trait, the greater the proportion of total variability that is due to genetic variation, rather than environment. High heritability in a QTL study means that a great percentage of phenotypic variance can be accounted for by the QTL (Griffiths *et al.* 1993).

(iii) Correlation analysis among traits measured in the segregating population: Many related phenotypic traits can vary together in an imperfect or approximate way. The usual measure of the precision of a relationship between two variables is the correlation coefficient (R). A high correlation between two traits may indicate that the same QTL influence both traits (a condition known as pleiotropy). It may also indicate that linked QTL, rather than the same genes, are associated with the traits.

4- Identification of QTL

Many programmes have been developed for detecting QTL, including MAPMAKER/QTL (Lincoln *et al.* 1992), QTL Cartographer (Basten *et al.* 1994, 1997), Map Manager QT (Manly 1997; Manly and Elliott 1991), Multimapper (Sillanpää 1998) and MapQTL (Van Ooijen 2004).

In the next section, I will focus on the three methods that I used during my QTL analysis: single marker analysis, simple interval mapping and multiple-QTL model mapping (Liu 1998; Tanksley 1993).

Single-locus association (single-marker analysis):

This test considers each marker locus separately and does not require that the marker loci be mapped relative to one another. The statistical methods used for this analysis include t-test, analysis of variance (ANOVA) and linear regression. Linear regression is most commonly used because the coefficient of regresssion (R^2) from

the marker explains the phenotypic variation arising from the QTL linked to the marker. The disadvantage with this method is that the further a QTL is from a marker, the less likely it will be detected. This is because recombination may occur between the marker and the QTL. It usually causes the magnitude of the effect to be underestimated (Tanksley 1993). The use of a large number of markers covering the entire genome may minimize this problem. The QTL analysis software MapQTL® 5 uses the rank sum test of Kruskal-Wallis to detect association between marker and QTL. This test is nonparametric, meaning that no assumptions are made for the probability distribution of the quantitative trait (after fitting the QTL genotype). For the same reason, Kruskal-Wallis test can also be used when dealing with ordinal traits (traits classified in categories) (Van Ooijen, personal communication). The test is performed on each locus separately and no use is made of the linkage map other than for sorting the order of loci on the linkage groups. The test ranks all the individuals according to the quantitative trait, while it classifies them according to their marker genotype. A segregating QTL linked closely to the tested marker will result in large differences in average rank of the marker genotype classes (Van Ooijen 2004).

Interval mapping (Simple interval mapping and Multiple-QTL model mapping):
Simple interval mapping (SIM):
This method requires prior construction of a genetic map and is based on the analysis of intervals between adjacent pairs of linked markers along chromosomes (Lander and Botstein 1989). With this method, for each position on the genome (every centiMorgan) the likelihood for the presence of a segregating QTL is determined (the likelihood under the alternative hypothesis, H1). At the same time the genetic effects of the QTL and the residual variance are calculated. The likelihood under H1 is compared to the likelihood for the situation when a locus with zero genetic effect would segregate, i.e. there is no segregating QTL (the likelihood under the null-hypothesis, H0). This comparison is performed with the likelihood statistic ratio LOD (or the LOD score), which is the 10-base logarithm of the quotient of the two respective likelihoods. When the LOD score exceeds the predefined significance threshold (see permutation test) somewhere on a linkage group, a segregating QTL is detected.
The position with the largest LOD on the linkage group is the estimated position of the QTL on the map. The use of linked markers for this analysis compensates for the

recombination between the markers and the QTL, and is considered statistically more powerful compared to single marker analysis.

Multiple-QTL model mapping (MQM mapping):

MQM mapping was developed by Jansen (Jansen 1993; Jansen 1994) and Jansen and Stam (1994). Following QTL identification by interval mapping, markers close to detected QTL are selected as cofactors to take over the role of the nearby QTL in the approximate multiple-QTL models used in the subsequent MQM mapping. With this MQM mapping, a search over the genome is performed by testing for a single segregating QTL, while simultaneously fitting the selected cofactors, both under H0 and under H1. By doing this, the cofactors will reduce the residual variance of the QTL already detected. When a QTL explains a large proportion of the variation, the use of a linked marker as cofactor will enhance the power in the search for other segregating QTL. Several rounds of MQM mapping may be necessary to obtain the best possible final solution, and cofactor markers can be added or dropped according to the latest results (Van Ooijen 2004). As for interval mapping, the results of this test are presented using LOD score.

For both SIM and MQM mapping, the percentage of variance explained by a QTL (or percentage of explanation) was calculated as follow:

$$\% \text{ explanation} = \frac{100 \times \text{H0 var} - \text{var}}{\text{Population var}}$$

In which H0 var is the residual variance under the current null hypothesis.

QTL significance:

In order to establish whether a QTL has a significant effect on a trait, it is possible to calculate the significance threshold of the LOD score. The permutation test (Churchill and Doerge 1994) is a method widely used to determine this threshold. In this test, the trait values are randomly permuted among the progeny (1,000 to 10,000 times), destroying the relationship between the trait values and the genotypes of the marker loci. QTL parameters and a LOD value is then estimated for each permuted data set at regular intervals throughout the genome and the maximum LOD is recorded. To determine the significance threshold, a P-value is determined for each

linkage group (generally 0.05) and the LOD corresponding to this value is taken as the threshold (Van Ooijen 2004).

5- Reporting and describing QTL detected

The most common way of reporting QTL is by indicating the most closely linked markers in a table and/or as bars on linkage maps. The chromosomal regions represented by rectangles are usually the regions that exceed the significance threshold. QTL that are detected in common regions (based on different criteria or for related traits) are likely to be important QTL for controlling the trait (Collard *et al.* 2005). If the strength of the QTL is considered, three categories can be defined. Those which explain over 20% of the variance are strong QTL and can be considered almost Mendelian. Moderate QTL are those which explain between 1% and 20% of the variance. Weak QTL explain 1% or less of the trait variance and require at least a thousand progeny to detect them (Manly and Olson 1999).

As mentioned previously, the most likely position for a QTL is the map position at which the highest LOD score is detected. However, QTL are usually reported with confidence intervals. The simplest way to calculate a confidence interval is to find the region on both sides of a QTL peak that corresponds to a decrease of 2 LOD score (Van Ooijen 1992).

6- The last step in a QTL analysis study involves the confirmation of QTL locations and influence on the trait. Such confirmation studies may involve independent populations constructed from the same parental genotype or closely related genotypes used in the primary QTL mapping study. Making comparisons among maps and QTL locations can also be a way to verify the validity of the QTL detected. To perform such comparison, anchor markers, such as SSR markers, are needed in order to correlate information from one map to another. If common markers have been incorporated into different maps, they can be aligned together and QTL locations can be evaluated. Such comparisons can potentially reveal clusters of QTL on specific chromosomic regions.

1.2.4 QTL mapping in apple

Since the development of saturated genetic maps in apple, QTL influencing a wide range of traits have been studied. QTL have been identified for resistance to apple scab using reference genetic maps constructed in the populations 'Prima' × 'Fiesta'

(Durel *et al.* 2003) and 'Fiesta' × 'Discovery' (Liebhard *et al.* 2003a). QTL for resistance to powdery mildew (Calenge and Durel 2006; Stankiewicz-Kosyl *et al.* 2005) and fire blight (Calenge *et al.* 2005; Peil *et al.* 2006) have also been recently mapped. Various QTL controlling fruit quality such as number of fruit, fruit weight, fruit firmness, sugar content and acidity (King *et al.* 2000; Liebhard *et al.* 2003a) have also been identified.

QTL influencing tree growth and development have also been investigated and are described in the next section of this chapter as they are of particular relevance to this study.

1.2.5 Tree architecture and development QTL in apple

Many morphological and developmental traits in apple are believed to be under multigenic control. Using a map developed from a segregating population derived from a cross between 'Rome Beauty' and 'White Angel', Lawson *et al.* (1995) investigated the genetic control behind branching type, reproductive budbreak and root suckering. Several QTL were identified for these traits. Further work by Conner *et al.* (1998) used a population derived from a cross between Wijcik McIntosh and NY 75447-58 to position additional QTL influencing height increment, internode length, internode number, base diameter increment, base diameter, branch number and leaf break. The QTL influencing these developmental traits identified in both studies are important but the linkage maps used by the authors cannot be aligned with the consensus map (Liebhard *et al.* 2003b) because of a lack of transferable markers.

Using the 'Fiesta' × 'Discovery' linkage map, Liebhard *et al.* (2003a) undertook the analysis of several seedling and tree traits: stem diameter and leaf size (seedling); height increment, stem diameter, blooming time, number of branches, juvenile phase length and fruit harvest date (Figure 1.9). Using a population derived from a cross 'Telamon' × 'Braeburn', Kenis and Keulemans (2007) undertook the analysis of the following growth characteristics: main axis growth rate (GR), main axis height increment (HI), main axis internode number (IN), main axis mean internode length (IL), sylleptic branch number (SBN), sylleptic branch length (SBL), proleptic branch number (PBN), total branch number (TBN), total branch length (TBL) and growth increment (GI). These measurements were taken for plants on their own roots for the first and second year of growth (respectively OR1 and OR2), and for plants grafted on 'M.9' rootstock the first year after grafting (RS1). Thanks to the use of SSR

markers this genetic map has been aligned with the consensus map. Finally, Segura *et al.* (2007), using a population derived from a cross 'Starkrimson' × 'Granny Smith', studied a series of geometric, topological and phenological traits on one year old progeny. The approximate location of all the QTL for tree growth and development is represented in figure 1.9.

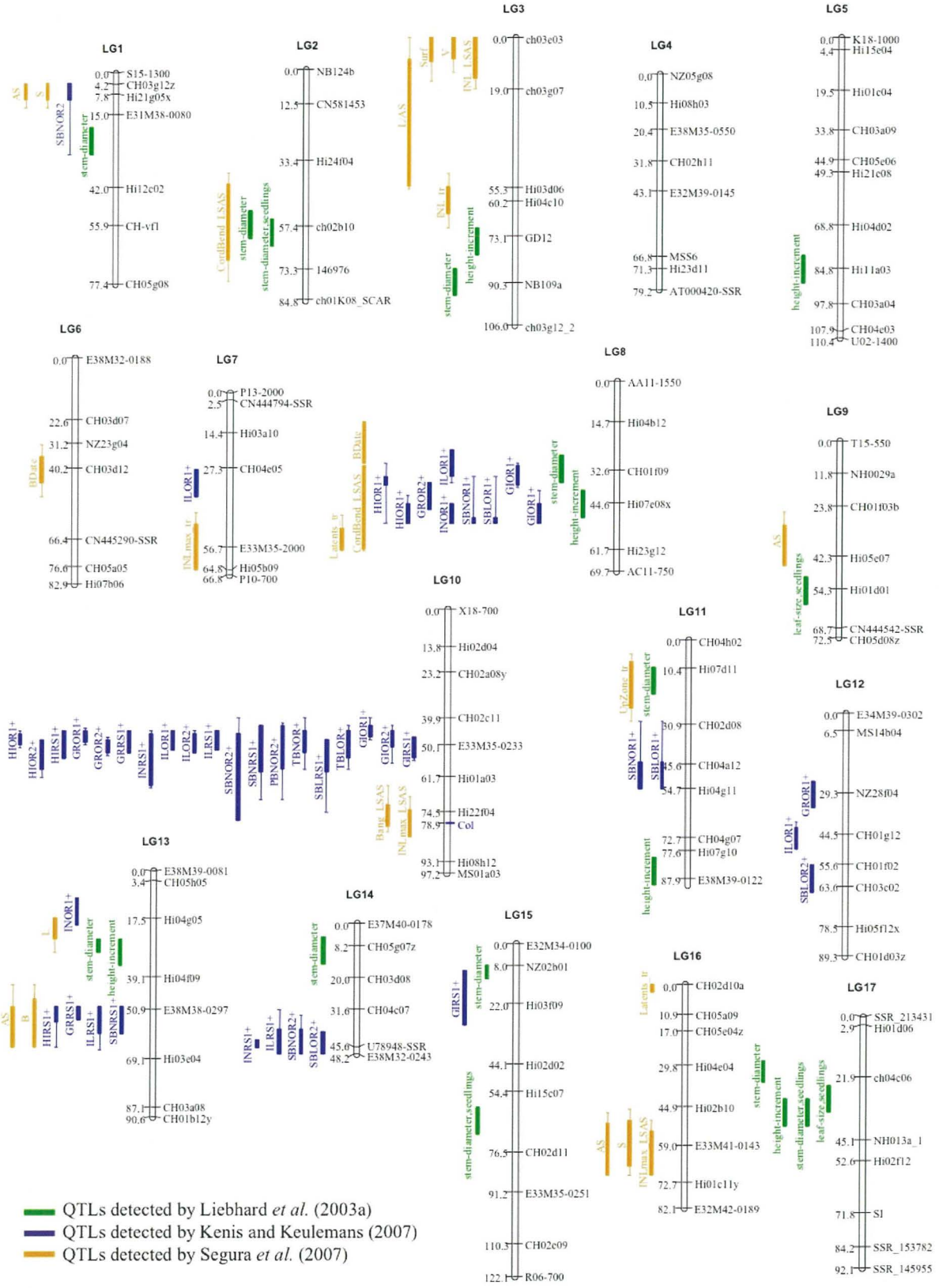

Figure 1.10. Representation the approximate location of all the QTL for growth traits on the linkage groups of a consensus map. LG represents the linkage groups following Liebhard *et al.* (2003b) numbering. The solid part of the bars of the QTL symbols indicate the most likely position of the QTL, the lines (when present) represent the confidence interval. QTL are located on the left side of LGs. Name of genetic markers is located on the right side of the LG.

QTL for tree architecture are located on all the LGs of the apple genome with the notable exception of LG4, LG6 and LG16. The results of these studies clearly show a clustering of the different QTL on LG8 and LG10. This is particularly relevant for the main axis trait QTL and the branching trait QTL. The main cluster is found on the LG10, in the region where a gene controlling the columnar habit had previously been mapped (Tian *et al.* 2005).

It is important to note that all the measurements in each study were made on genetically different scions grafted onto the same rootstock: 'M.27' in the case of the 'Fiesta' × 'Discovery' cross, and 'M.9' in the case of the 'Telamon' × 'Braeburn' cross. The influence of the rootstock on scion development has not been taken into account in any of these studies.

1.2.6 Bin Mapping: introduction and principles

A new method for mapping molecular markers and genes has recently emerged as a result of the development of saturated genetic maps. This method known as bin mapping or selective mapping was proposed by Vision *et al.* (2000) to improve the efficiency of mapping by significantly reducing the cost of genotyping new markers with a minimal loss of mapping precision. The method consists of a two-step process in which first a mapping population is used to construct a saturated framework map, and second, using a selected subset of highly informative plants, new markers are added to the map with lower precision. The number of plants in the subset is dependent on the size of the genome of the species studied, and the selection of the plants is based on the number and location of crossover sites. Ideally, for a given marker, the joint genotype of the selected subset of plants identifies a unique small bin in the genome (Figure 1.10).

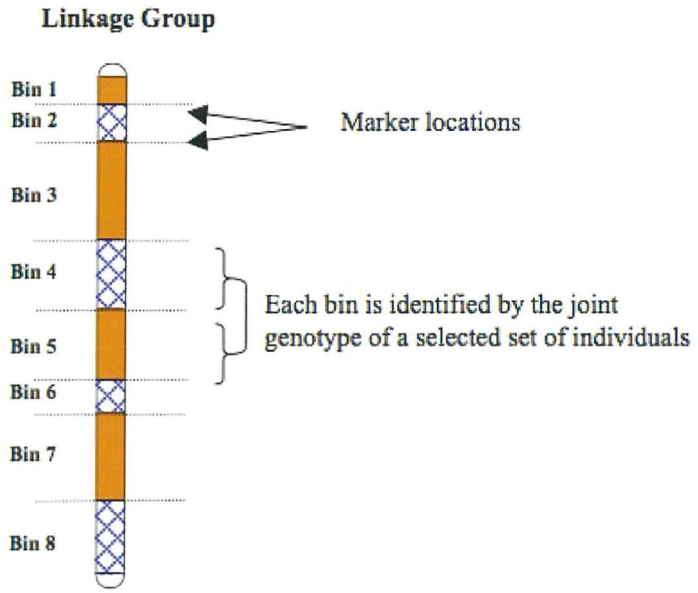

Figure 1.11. Example of a linkage group divided into bins. Each of the eight bin is unique and identified by the joint genotype of the individuals composing the bin set. The length of individual bins in a linkage group is dependent on the location of crossovers in individuals composing the bin set.

This method of mapping has already been successfully applied in *Prunus* (Howad *et al.* 2005), wheat (Johnson *et al.* 2007), and melon (Monforte *et al.* 2007).

1.2.7 The Candidate gene approach

The aim of the candidate gene approach is to identify markers within the gene directly controlling the trait of interest. This approach was made possible in apple by the creation and characterisation of 97 libraries of EST sequences (GDR, 2006) from 'Royal Gala', 'Pinkie', 'Pacific Rose', 'M.9', 'Aotea 1', 'Braeburn', 'Northern Spy', 'Goldrush', 'Jonagold', 'Granny Smith', 'Fuji', 'Suncrisp' and 'M.111'. The libraries represent 18 different tissue types, at 33 different stages of development, with a majority of them originating from different stages in fruit development (Newcomb *et al.* 2006; Arús and Gardiner 2007).

Candidate genes can be identified following two procedures. The first one consists of identifying potential candidates in the EST databases based on their homology to genes controlling traits in model plants. This procedure has already been successfully employed for various traits in many species, including members of the Rosaceae family, such as peach (Etienne *et al.* 2002), almond (Silva *et al.* 2005) and apple (Gardiner *et al.* 2003). The second is the microarray analysis approach (Pflieger *et*

al. 2001). Microarrays measure mRNA concentrations by labelling the sample with a dye and allowing it to hybridize to the spots on the array. Each spot contains either PCR-amplified cDNAs, or long oligonucleotides complementary to the transcribed parts of genes. The choice of spotting oligomers or cDNA sequences yields two different microarray technologies. Oligo arrays are generated by photolithography techniques to synthesize oligomers directly on the glass slide (Lipshutz *et al.* 1999). cDNA arrays are created by mechanical gridding, where prepared material is applied to each spot by ink-jet or physical deposition (Schena *et al.* 1995).

Generally, a one-to-one correspondence exists between spots and genes. However, various exceptions exist. Multiple genes may hybridize to the same spot if the DNA at that spot is not unique to a single gene; this problem is called cross-hybridization. A gene may also hybridize to more than one spot on the microarray if different spots cover different regions of the gene. Oligo microarrays can have non-specific binding, generally due to the presence of a single nucleotide polymorphism, which can alter their hybridization efficiency (Wu 2001).

The amount of hybridization to each spot is measured by the intensity of phosphorescent dye at each spot. mRNA samples are labelled with a dye before hybridization and the non-hybridized samples are washed off. The remaining hybridized and dye-labelled mRNA is then measured by a camera which records an intensity level. Problems due to artefacts arising from dust and other imperfections are handled by image processing software (Pflieger *et al.* 2001).

Data normalization:

During analysis of the microarrays, several factors have to be taken into account, such as the initial difference in mRNA concentration among samples; the concentration, brightness and relative binding affinity of the dye; the exposure time and the camera sensitivity. To correct these differences in intensity levels, a normalization, or bias correction needs to be performed (Wu 2001). In addition to these multiplicative effects, additive effects such as the background intensity level and the saturation effect in the hybridization process (amount of dye bound by an mRNA molecule) have to be estimated.

Evaluation of relative expression levels:

After data normalization, the relative expression of each gene can be evaluated according to its degree of differential expression. Because each gene is represented by a pair of expression values, the difference of expression can be evaluated by their difference or ratio. A threshold for selecting differentially expressed genes and *P* values can be calculated (Wu 2001).

Genes for which the *P* value is above the threshold can be considered as up or down regulated and consequently be evaluated as candidate genes.

Microarray analysis allows a significant initial reduction in the number of potential candidate genes that could be associated with a specific trait. Many studies in apple (Janssen *et al.* 2006; Schaffer *et al.* 2006; Schaffer *et al.* 2007) have demonstrated the power of this technology. The microarray analysis platform developed by HortResearch Ltd. (New Zealand) utilizes a 15,723 45-55 mer oligonucleotide array, representing 15,102 non redundant *Malus* sequences.

1.3 Aims and objectives

The overall aims of the project were to understand the genetic control of scion architectural modification induced the dwarfing apple rootstock 'M.9', and to develop genetic markers enabling an early detection of this desired phenotypic trait in apple rootstock. To achieve our goals, two populations segregating for the dwarfing characteristic were derived from crosses between 'M.9' and 'Robusta 5' (vigorous rootstock). Using these two populations, several strategies were used to achieve specific objectives.

I- In order to identify genetic markers closely linked to the major locus (*DW1*) influencing the dwarfing effect of 'M.9' (identified by Rusholme *et al.* 2004), and to uncover additional loci involved in the control of this trait, the bulked segregant analysis (BSA) strategy using RAPD primers was employed. To achieve these goals, various bulks were constructed from DNA of individuals based on their phenotype and genotype at the *DW1* locus.

II- To localize the dwarfing loci identified in objective I on the apple genome, and to gain a better understanding of the genetic control of this trait, the saturated genetic maps of 'M.9' and 'R5' were constructed using a variety of published molecular markers, as well as newly developed SSR markers from the apple EST database.

III- In order to ascertain the validity of the loci identified by BSA, and to identify additional loci with small genetic effects, a QTL analysis of the dwarfing phenotype and other related traits was performed using the genetic maps developed in objective II.

IV- With the aim of identifying the genes responsible for the dwarfing effect of 'M.9', candidate genes identified by microarray analysis were mapped and their location in the genome was compared to the position of the QTL previously identified. The identification of genes responsible for dwarfing would provide breeders with reliable markers located directly on the gene, enabling a very accurate marker assisted selection system to be implemented.

V- As it is often the case in many important food crops, apple scion cultivars are based on a very limited range of progenitor cultivars. This observation can also be applied to apple rootstocks since the majority of the commercially used rootstocks throughout the world are derived from 'M.9' or one of its seedlings. In order to determine the number of different dwarfing genetic sources, markers closely linked to the major locus *DW1* were screened over a set of rootstocks accessions, related or not to 'M.9'.