Alma Mater Studiorum – Università di Bologna

Dipartimento di Colture Arboree

DOTTORATO DI RICERCA IN

BIOTECNOLOGIE CELLULARI E MOLECOLARI

Ciclo XXIII

Settore scientifico-disciplinare: AGR/03, Arboricoltura Generale e Coltivazioni Arboree

METABOLIC AND MOLECULAR ASPECTS OF CYANOGENESIS IN APRICOT SEED

Dissertazione presentata dalla Dott.ssa Claudia Cervellati

Coordinatore Dottorato

Prof. Santi Spampinato

Relatore

Prof. Andrea Masia

Co-relatori

Prof. Bernd Schneider
Dr. Luca Dondini

Esame finale anno 2011

Index

Abstract

An holistic investigation on the physiological and genetic factors influencing the apricot fruit quality must consider the characteristic of the whole fruit, seed included. In fact, usually only the edible part of the fruit is taken in consideration when quality is defined. However, especially in stone fruits, the seed represents an important portion of the fruit itself and the major by-product in fruit processing. Nowadays, huge amounts of seeds are discarded yearly when processing apricot fruits. This not only wastes a potentially valuable resource but also aggravates an already serious disposal problem (Femenia et al., 1995; Schieber et al., 2001). Besides, apricot seeds could be a very interesting source of food, thanks to their significant content in dietary protein, oil and fibre (Gòmez et al., 1998). Unfortunately, seeds from several apricot varieties contain also a remarkable amount of the cyanogenic glycoside amygdalin. This represents a big constraint to their use for human or animal nutrition, both for its bitter taste and, moreover, for the toxicity of the hydrogen cyanide (HCN) produced by its dissociation. So, development and cultivation of high quality cultivars with sweet kernel may represent a quantum change for the apricot production system, as far as it would decrease health hazards and in parallel increase the marketability of this by-product.

Cyanogenic glycosides (CNGs) are β-glucosides of α-hydroxynitriles derived from amino acids and have been found in more than 2,650 plant species distributed among 130 families in pteridophytes, gymnosperms and angiosperms. Such a widespread occurrence among many different taxonomic groups implies that in plants the ability to synthesize CNGs is at least 300 million years old (Bak et al., 2006). The most known role of CNGs is linked with plant defences. In fact, secondary metabolites play an important role in these responses and CNGs represent an important group of phytoanticipins: constitutive defence compounds produced independently from the pathogen attack (Zagrobelny et al., 2004). These natural products are instantly activated upon tissue damage or pathogen attack. Briefly, CNGs are stored in the vacuoles (Vetter, 2000) and, when the cell integrity is lost, they are brought into contact with the specific β-glucosidases and α-hydroxynitrile lyases that hydrolyze them, thereby causing a release

of toxic hydrogen cyanide. This binary system provides plants with an immediate defence against intruding herbivores and pathogens that cause tissue damage (Møller and Seigler, 1999; Møller, 2010). Thus, cyanogenesis requires the presence of two biochemical pathways: the former controlling the synthesis of the cyanogenic glycoside and the latter controlling the production of the specific degradative enzymes. The balance of these routes determines the cyanogenic potential of the plant.

In addition, accumulation of cyanogenic glucosides in certain angiosperm seeds may provide a storage deposit of sugars and nitrogen for the developing seedlings (Swain et al., 1992b; Swain and Poulton, 1994). Other evidences demonstrate that cyanogenic compounds also have a function in the metabolism and transport of nitrogen (Gleadow and Woodrow, 2000; Busk and Møller, 2002), in the legume-rhizobium interaction and in reducing the damages caused by excessive light (Møller, 2010). These roles do not, however, need to be seen as an alternative to the herbivore defence: just as some plants store nitrogen as inactive Rubisco, cyanogenic plants may store nitrogen in a toxic form for the dual purpose of defence (Gleadow and Woodrow, 2002).

Apricot and other species belonging to the Rosaceae family are among the plants producing the highest levels of CNGs (Swain and Poulton, 1994). CNGs inheritance has been studied in various *Prunus* species. In peach, the sweet kernel behaves as a recessive trait, controlled by a single gene (sk), linked to the fuzzless skin (nectarine) trait and therefore assigned to the linkage group 5 of a genetic map based on an interspecific almond×peach cross (Bliss et al., 2002). In almond, sweet has been reported as a dominant trait inherited as a simple Mendelian factor (Dicenta et al., 2007; Sánchez-Pérez et al., 2008). In this species, as well as in apricot, all the seeds produced by a single tree show the same phenotype (bitter, slightly-bitter or sweet); thus, the presence of CNGs is a seed-parent trait (Negri et al., 2008; Sánchez-Pérez et al., 2010). Regarding the apricot, different hypotheses on the inheritance of the bitter/sweet trait have been proposed (Kostina, 1977; Bassi and Negri, 1991) but no definitive model has been demonstrated. Negri et al. (2008), given that both the bitter and the sweet phenotypes were represented in populations from bitter×bitter and sweet×sweet crosses as well as from self-pollination of either bitter- or sweet-seeded trees, elaborated a model in which five non-linked genes (three for the anabolic and two for the catabolic way) were involved in the determination of this quantitative character.

A very efficient approach to study quantitative trait loci (QTLs) implies the use of Candidate Genes (CGs, Pflieger et al., 2001; Etienne et al., 2002). This approach is based

on the *a priori* choice of one or more genes that seem to be functionally connected to the examined character: a correlation between the phenotypic data and an allelic polymorphism validates the hypothesis that the chosen gene controls, at least partly, the studied trait (Causse et al., 2004). On these CGs it would then be possible the development of functional markers, based either on direct sequence polymorphisms (SCAR: Sequenced Characterized Amplified Region) or on differences in restriction sites (CAPS: Cleaved Amplified Polymorphic Sequence) or on a different number of microsatellite repetitions (SSR: Short Sequence Repeat). These particular kind of molecular markers would eventually provide the breeders with a valuable method to carry on an early selection of the plants with the desired traits. In fact, being a direct analysis of the plant genotype, the MAS (Marker Assisted Selection) makes possible the recognition of the individuals carrying certain characteristics way before these would be showed in the field. To support these selection programmes, the availability of a densely populated linkage map is a prerequisite. Four apricot maps were realized for the apricot accessions Lito, BO81604311 (San Castrese×Reale di Imola; Dondini et al., 2007), Harcot and Reale (Dondini et al., 2010).

Beyond functional markers and genetic linkage maps, the identification of the DNA regions and of the candidate genes that underline the multifactorial trait bitterness presupposes a fine phenotypization. Only a very precise analysis, in fact, could provide the characterization of the various QTLs accountable for the variation of the phenotypic trait, so telling us about the genetic architecture of this character. The nuclear magnetic resonance (NMR) is considered to be one of the most robust and precise analytical methods in research, and it's slowly becoming more appreciated for the phenotyping purpose (Moing et al., 2004; Terskikh et al., 2005; Pereira et al., 2006). Under precise "quantitative conditions" (Pauli et al., 2005 and 2007) NMR enables a unique and quantitative determination of the relative amount of molecular groups, so being a real powerful tool to quantify entire molecular structures even in mixtures. Moreover, quantitative NMR (qNMR) analyses are time-saving, thanks to the ability of quantifying a single compound in complex mixtures without requiring fractionation or isolation procedures.

Thus, to investigate the metabolic and molecular aspects of the cyanogenesis in apricot, a candidate gene approach coupled with a precision phenotyping *via* qNMR was used on two F1 apricot population with the aims of: (i) identify and sequence some

apricot genes involved in the cyanogenesis; (ii) develop functional markers linked to those genes and implement them in four existent maps; (iii) achieve a fine phenotyping of the amygdalin content of the seeds collected from the seedlings of both progenies; (iv) identify the genetic loci involved in the determination of the bitterness; (v) localize the amygdalin directly *in situ* with Raman imaging.

Screening the Lito's BAC library with the heterologous primers designed on sequences available in public databases resulted in the achievement of the apricot sequences for: UDPG-glucosyl transferase, amygdalin and prunasin hydrolases, mandelonitrile lyase. The primer walking on those sequences is not yet completed, but the length achieved by now is sufficient enough to give high homology results with the other available *Prunus* sequences.

Seven new functional markers developed on three candidate genes involved in the catabolism of CGs were added to the maps of the two apricot accessions Lito and BO81604311. Four of them were added in Lito: an SSR for an MDL gene in L1, an SSR for an isoform of AH in L7, two loci of a CAPS marker for a PH gene in L2 and L7. The remaining three were positioned in the map of BO81604311: a CAPS marker for an MDL isoform in B1, another linked to a PH gene in B6 and an SSR for an AH in B7. Other eight markers, seven developed on two catabolic enzyme sequences and one on an anabolic one, were instead mapped on the cultivars Harcot and Reale di Imola. Five were mapped in Harcot: an SSR for UGT in H3; an SSR and two CAPS for a PH gene in H1 and H6, respectively; and an SSR for an isoform of AH in H7. The other three were located in the map of Reale di Imola: a CAPS marker for an AHL isoform in R6A, another linked to a PH gene in R7 and an SSR for an AH always in the LG7. Analogously to almond (Sánchez-Pérez et al., 2010), none of the CGs mapped in this work was found to be in a QTL region. However, other isoforms of these genes could still be involved in the determination of the bitter phenotype.

The major QTLs for bitterness, already identified in almond and peach (Joobeur et al., 1998, Dirlewanger et al., 2004, Sánchez-Pérez et al., 2007), was found in the linkage group 5 mainly in Harcot (30.63 of LOD), but also in Reale di Imola, even if with a relative lower effect. Another QTL has been identified in the bottom part of R4. As for Lito and BO81604311, putative QTLs were found in L1, L4 and L6, as well as in B1, B4 and B6, seeming to be located in the same position in both parents. In Lito another significant locus is found in L3. The Lito×BO81604311 cross produces a progeny without

any sweet seedling, consequently it was very important the high precision of the NMR quantification in order to identify significant QTLs. In fact the genomic regions identified in this study for the first time are likely to be the ones responsible for the variability of the bitter taste among the individuals and not for the bitter/sweet phenotype.

The use of qNMR to determine the amygdalin content of the seeds allowed also the quantification of the small amygdalin amounts present in some sweet H×R kernels and showed that amygdalin is the only cyanogenic glycoside of mature apricot seeds. Plotting the classes-clustered amygdalin content measured for the L×B population among the three harvesting years resulted in a double-peak distribution, a trend that is not uncommon in this population, which has often shown differences between the early ripening individuals and the late ones. The peaks' position is slightly shifted depending on the years and the shape of the distribution is also changing, with the 2007 and 2009 populations seeming to be more spread, showing very low and very high amygdalin content (with the 2009 harvest a little bit shifted toward the less-bitter values), whereas the 2008 one looks more close to the average value. This high effect of the environmental conditions is a clear confirmation of the quantitative nature of the bitterness trait in apricot. A statistical survey of a few environmental variables (maximum, minimum and average temperatures, mm of rain) clustered the rainfall in the same node as the amygdalin content and put the maximum temperatures as the most distant parameter.

A particular result of the qNMR measurements was the sometimes huge differences found between the amygdalin values determined in different quarters of the same seed. Several imaging experiments done in Raman microspectroscopy made clear that amygdalin is unevenly distributed in the seed and, moreover, its localization seems not to follow a unique pattern. This findings are consistent with the hypotheses that consider the storage as the principal function of the CNGs in the seeds (Swain et al., 1992b; Swain and Poulton, 1994; Gleadow and Woodrow, 2000 and 2002; Busk and Møller, 2002; Jørgensen et al., 2005), since it seems more important the presence itself of the compound than its precise localization.

In conclusion, this thesis has revealed some new loci, involved in the shaping of the bitterness degree; has proven the complexity of the bitter trait in apricot, reporting an high variance of the amygdalin content found among the years and a correlation with the environmental parameter rainfall; has showed the critical importance of the phenotyping

step, whose precision and accuracy is a pre-requisite when studying such a multifactorial character; has given new insights for the *in situ* localization of the amygdalin.

1. Introduction

1.1 Cyanogenic glycosides

The knowledge about hydrogen cyanide (HCN) formation in plants is antique. In ancient Egypt, traitorous priests in Memphis and Thebes were poisoned to death with pits of peaches (Davis, 1991). The first known detection of HCN liberated from damaged plant tissue was made in 1802 by the pharmacist Bohm in Berlin, upon distillation of bitter almonds and, later on, in 1830, Robiquet and Boutron-Chalard discovered the structure of the HCN-liberating compound in bitter almonds (Lechtenberg and Nahrstedt, 1999). Because the compound was isolated from *Prunus amygdalus* (synonym *Prunus dulcis*), it was named amygdalin. Amygdalin has subsequently been found widespread in seeds of other members of the Rosaceae like in apples (*Malus* spp.), peaches (*Prunus persica*), apricots (*Prunus armeniaca*), black cherries (*Prunus serotina*), and plums (*Prunus* spp.; Sánchez-Pérez et al., 2008).

The most known role of HCN is linked with plant defences. In fact, plants are sessile organisms and, accordingly, they respond to environmental challenges and attacks from herbivorous and microbial pathogens by defending themselves rather than escaping. Secondary metabolites play an important role in these responses (Bak et al., 2000). Phytoalexins are defence compounds elicited by pathogen attack. Their formation from remote precursors is dependent on *de novo* enzyme synthesis and follows a lag phase of several hours after infection. In contrast, phytoanticipins are constitutive defence compounds that are produced in specific plant tissues and at specific developmental stages, independently of pathogen attack. These natural products are instantly activated upon tissue damage or pathogen attack. The activation is mediated by compartmentalized plant enzymes released as a result of loss of cell integrity. Phytoanticipins are therefore among the first chemical barriers to potential predators, herbivores and pathogens (VanEtten et al., 1994; Osbourn, 1996).

Cyanogenic glycosides (CNGs) constitute an important group of phytoanticipins.

CNGs have been found in more than 2,650 higher plant species distributed among 130 families in pteridophytes (ferns), gymnosperms and angiosperms (flowering plants; Conn 1981; Siegler and Brinker 1993). Such a widespread occurrence implies that they are ancient biomolecules in terrestrial plants (Bak et al., 2006; Zagrobelny et al., 2004). CNGs are β-glucosides of α-hydroxynitriles derived from aminoacids (Fig. 1.1). Cyanogenic ferns and gymnosperm species contain aromatic cyanogenic glycosides derived from either tyrosine or phenylalanine, whereas angiosperms contain aliphatic as well as aromatic cyanogenic glycosides derived from isoleucine/valine/leucine or tyrosine/phenylalanine, respectively. In addition, a few plant species, such as *Passiflora*, contain cyanogenic glycosides derived from the non-protein amino acid cyclopentenyl glycine (Jaroszewski et al., 2002).

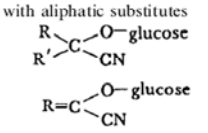
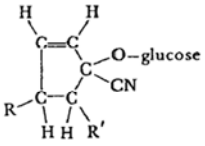
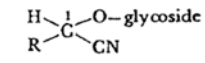
	Substituent	Glycoside	Sugar	Configuration at C1	Occurrence
(A) Glycosides with aliphatic substitutes 	R=R′=CH_3-	linamarin	D-glucose		*Linum* spp. *Trifolium* spp.
	R=CH_3-, R′=CH_3CH_2-	lotaustralin	D-glucose		*Lotus* spp. *Maniholt* spp.
	R=$(CH_3)_2CH$	aciapetalin	D-Glucose		*Acacia* spp.
	R=HCO_2CH=$CH(CO_2HCH_2)C$	Triglochinin	D-glucose		*Triglochin* spp.
	R=H, R′=H	deidaclin	D-glucose		*Deidamia* spp.
		tetraphyllin A	D-glucose		*Tetrapathaea* spp.
	R=OH, R′=H	tetraphyllin B	D-Glucose		*Tetrapathaea* spp.
	R=OH, R′=H	Gynocardin	D-glucose		*Gynocardia* spp. *Pangium* spp.
(B) Glycosides with aromatic substituents 	phenyl	prunasin	D-glucose	D	*Prunus* spp.
	phenyl	amygdalin	gentiobiose	D	*Prunus* spp.
	phenyl	lucumin	primeverose	D	*Lucuma* spp.
	phenyl	vicianin	vicianose	D	*Vicia* spp.
	phenyl	sambunigrin	D-glucose	L	*Sambucus* spp.
	p-hydroxyphenyl	dhurrin	D-glucose	L	*Sorghum* spp.
	p-hydroxyphenyl	taxiphyllin	D-glucose	D	*Taxus* spp.
	p-Hydroxyphenyl	Zierin	D-Glucose	–	*Zieria* spp.
	p-Glucosyloxyphenyl	Proteacin	D-Glucose	L	*Macadamia* spp.

Fig. 1.1. Chemical structures of some cyanogenic glycosides (Vetter, 2000).

In plants, CNGs are stored in the vacuoles (Vetter, 2000). When plant tissue is disrupted, e.g. by herbivore attack, CNGs are brought into contact with β-glucosidases and α-hydroxynitrile lyases that hydrolyze them, thereby causing a release of toxic hydrogen

cyanide (HCN). This binary system —two sets of components chemically inert when separated— provides plants with an immediate defence against intruding herbivores and pathogens that cause tissue damage. (Møller and Seigler, 1999). The efficacy of these defences can be deducted by considering the high toxicity of hydrogen cyanide, mainly due to its affinity for the terminal cytochrome oxidase in the mitochondrial respiratory pathway (Brattsten et al., 1983). The acute lethal dose of cyanide for vertebrates lies in the range of 35–150 mmol/kg, if applied in a single dose. Much higher amounts of HCN can be tolerated if consumed or administered over a longer period (Davis and Nahrstedt, 1985). There are many well-documented examples of deaths and serious illness from consumption of cyanogenic plants in humans (e.g., Cock, 1982; Cardoso et al., 1998; Banea-Mayambu et al., 2000), cattle (e.g., Robinson, 1930; Finnemore et al., 1935; Boyd et al., 1938; Hopkins, 1995; Zentek, 1997), sheep (e.g., Robinson, 1930; Cooper-Driver et al., 1977; Crush and Caradus, 1995), koalas (e.g., Pratt, 1937), goats (e.g., Webber et al., 1985), and other grazing mammals (e.g., Harborne, 1982; Saucy et al., 1999).

1.1.2 Metabolism

Biosynthesis and degradation of CNGs are well documented in many plants (Jones et al., 2000; Lechtenberg and Nahrstedt, 1999) The main metabolic processes resulting in synthesis, degradation and detoxification of CNGs in plants are shown in Fig. 1.2.

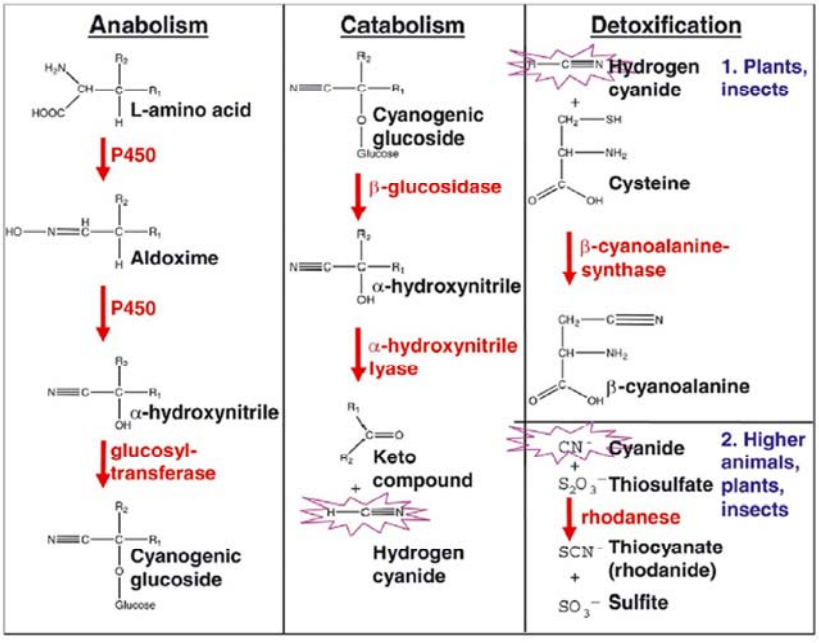


Fig. 1.2. Metabolism of CNGs in plants, insects and higher animals. Enzymes involved are shown in red. HCN is highlighted in purple (Zagrobelny et al., 2004).

The first two committed steps in CNG biosynthesis are catalyzed by cytochromes P450. The first P450 catalyzed step proceeds *via* two successive N-hydroxylations of the amino group of the parent amino acid, followed by decarboxylation and dehydration (Sibbesen et al., 1994). The aldoxime formed is subsequently converted to an α-hydroxynitrile through the action of a second cytochrome P450 (Bak et al., 1998; Kahn et al., 1997). This reaction involves an initial dehydration that forms a nitrile and is followed by hydroxylation of the alpha carbon to generate a cyanohydrin. The final step in CNG synthesis, the glycosylation of the cyanohydrin moiety, is catalyzed by a Family1 UDPG-glycosyltransferase (Jones et al., 1999). These biosynthetic enzymes form a metabolon to facilitate the channelling of the otherwise toxic and reactive intermediates to the end product in the pathway (Bak et al., 2006). A metabolon is a multienzyme complex whose proteic constituents are held together by noncovalent interactions and often stabilized by membrane anchoring. Allowing the direct passage of a product from one enzymatic reaction to the consecutive enzyme, the intermediate diffusion is limited, their different pools are kept separated, the turnover of lable or toxic intermediates is facilitated and it is also prevented an undesired crosstalk between different metabolic pathways (Møller, 2010).

Catabolism of CNGs is initiated with an enzymatic hydrolysis catalyzed by a β-glucosidase to afford the corresponding α-hydroxynitrile which, at pH values above 6, spontaneously dissociates into a sugar, a keto compound, and HCN. At lower pH values, the dissociation reaction is driven by an α-hydroxynitrile lyase.

HCN is detoxified by two main reactions (Møller and Poulton, 1993). The first route involves the formation of β-cyanoalanine from cysteine and is catalyzed by β-cyanoalanine synthase. β-cyanoalanine is subsequently converted into asparagine (Miller and Conn, 1980). For example, in sorghum seedlings HCN is converted into the amide group of asparagine, and without any separation of the C and N atoms of the cyanide group (Blumenthal et al., 1968). The second route proceeds by conversion of HCN into thiocyanate and is catalyzed by rhodanese (Bordo and Bork, 2002). The detoxification route involving β-cyanoalanine is common in plants and possibly also in insects, while the thiocyanate pathway occurs mainly in vertebrates but also in some plants and insects.

It may be advantageous for plants which contain cyanogenic glycosides to be capable of metabolizing cyanide, and for those plants which contain high levels of cyanogenic glycosides to have high levels of cyanide metabolizing activity. The route of β-cyanoalanine synthase may also be advantageous since many plants can further metabolize the β-cyanoalanine to asparagine which can then be incorporated into the general metabolism of the plant (Blumenthal et al., 1968).

So, cyanogenesis requires the presence of two biochemical pathways: one controlling the synthesis of the cyanogenic glycoside and the other controlling the production of the specific degradative β-glucosidase. The balance of these routes determines the cyanogenic potential of the plant.

In sorghum and *L. japonicus*, the cyanogenic potential decreases as a function of developmental age. This indicates that synthesis primarily occurs in young and developing tissues and that the levels found in older plant parts decrease because *de novo* synthesis proceeds at a lower rate than catabolic turnover (Bak et al., 2006). The presence of high amounts of cyanogenic glycosides in young and developing tissues supports the function as defence compounds. In the works of Swain et al., (1992a) and Zheng and Poulton (1995), biochemical changes related to cyanogenesis were monitored during the maturation of black cherry fruits. It was shown that, concomitant with cotyledon development during phase II, the seeds begin to accumulate both the amygdalin and the catabolic enzymes amygdalin hydrolase (AH), prunasin hydrolase (PH), and mandelonitrile lyase (MDL). Thus, from that time onward, they are therefore highly cyanogenic when disrupted (showing a maximum at approximately 49 days after flowering, DAF). In contrast, the pericarp remains acyanogenic throughout the entire ripening process because it lacks the catabolic enzymes.

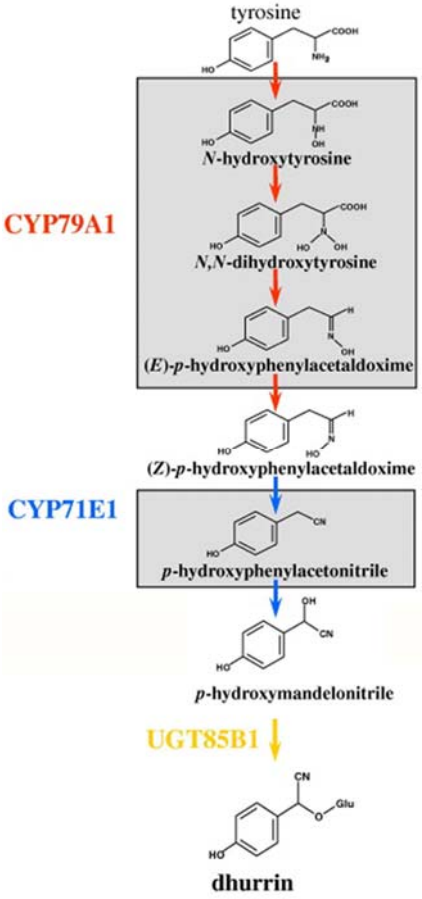
The inheritance, the genetic control of cyanogenesis, has not an unique mechanism. The results regarding different cyanogenic plant taxa are different. Variations occur within individual plants or genotypes, depending on different physiological or ecological factors (Vetter, 2000). In white clover (*Trifolium repens*) the

discrete form of variation is controlled by two genes according to the classic result of Corkill (1942). The presence or absence of both the glycosides (linamarin and lotaustralin) is regulated by alleles of a single gene, designated Ac, whereas the presence or absence of the enzyme linamarase is governed by alleles of another, independently inherited, gene (Li). Only plants which posses dominant functional alleles of both genes liberate HCN when damaged. In peach (*Prunus persica*) the sweet kernel behaves as a recessive trait, controlled by a single gene (sk), linked to the fuzzles skin (nectarine, Bliss et al., 2002), while in almond (*Prunus amygdalus*) sweet has been reported to be a dominant trait inherited as a simple mendelian factor (Dicenta et al., 2007; Sánchez-Pérez et al., 2008) In the *Prunus* spp., for a given tree, the phenotype of the kernels with respect to bitterness is generally uniform and determined by the genotype of the tree (Dicenta et al., 2000). Candidate sites for the control of amygdalin accumulation, therefore, include all vegetative parts of the tree as well as the kernel seed coat – which is derived from ovule integument (Haughn and Chaudhury 2005) – but excludes other kernel tissues that have pollination-derived genotypes.

In addition to the defence function, the accumulation of cyanogenic glucosides in certain angiosperm seeds may provide a storage deposit of sugar and nitrogen for the developing seedlings (Swain et al., 1992b; Swain and Poulton, 1994). Other evidences demonstrate that cyanogenic compounds also have a function in the metabolism and transport of nitrogen (Gleadow and Woodrow, 2000; Busk and Møller, 2002), in the legume-rhizobium interaction and in reducing the damages caused by excessive light (Møller, 2010). These roles do not, however, need to be seen as an alternative to their role in herbivore defence. Just as some plants store nitrogen as inactive Rubisco, cyanogenic plants may store nitrogen in a toxic form for the dual purpose of defence (Gleadow and Woodrow 2002).

1.1.3 Key enzymes for the anabolism.

Concerning the biosynthesis of CNGs, *Sorghum bicolor* (L.) Moench is the model plant (Bak et al., 1998). Sorghum contains the tyrosine-derived cyanogenic glucoside dhurrin. Dhurrin biosynthesis is catalyzed by two multifunctional membrane-bound



cytochrome P450 (Cyt P450) enzymes: CYP79A1 and CYP71E1 (Sibbesen et al., 1994 and 1995; Bak et al., 1998 and 2000), and it is shown in Fig. 1.3.

CYP79A1 catalyzes the conversion of Tyr into Z-phydroxyphenylacetaldoxime (Sibbesen et al., 1995) and CYP71E1 catalyzes the conversion of the Z-phydroxyphenylacetaldoxime into p-hydroxymandelonitrile (Kahn et al., 1997; Bak et al., 1998). Conversion of this labile cyanohydrin into dhurrin is catalyzed by a soluble UDP-Glc (UDPG)-glucosyltransferase UGT85B1 (Jones et al., 1999; Hansen et al., 2003).

Fig. 1.3. The biosynthetic pathway of the cyanogenic glucoside dhurrin from sorghum. The reaction catalyzed by the two cytochromes P450 are boxed (Nielsen et al., 2008).

Cytochromes P450 are a family of heme-thiolate enzymes which catalyse a vast array of different reactions on endogenous and exogenous substrates. In plants, cytochromes P450 are involved in the biosynthesis of many different secondary metabolites, such as phenylpropanoids, terpenoids, alkaloids and cyanogenic glucosides, as well as in the detoxification of xenobiotics. Plant cytochromes P450 have been divided into an A and a non-A group. Members of the A group are believed to be derived from a common plant cytochrome P450 ancestor. The non-A group is more heterogeneous than the A group and does not form a group in a phylogenetic sense (Bak et al., 1998). Cytochrome P-450 dependent monooxygenases are small electron transport chains consisting of a flavin-containing reductase transferring the reducing equivalents from NADPH to a terminal CYP450 oxygenase, which is then responsible for substrate binding and O_2-activation. Generally, the CYP450 involved in biogenesis exhibit high substrate specificity whereas the CYP450 responsible for catabolic processes have a broad substrate specificity. In particular it has been demonstrated that CYP79A1 shows very high substrate specificity, whereas CYP71E1 has a less stringent substrate requirement (Kahn et al., 1999)

The final step in biosynthesis of CNGs is catalyzed by a Family 1 glycosyltransferase (Jones et al., 1999; Vogt and Jones, 2000). Family 1 glycosyltransferases are soluble proteins with a molecular mass of 45–60 kDa, which utilize UDP-activated sugar moieties as the donor molecules to glycosylate the acceptor molecules. As for cytochromes P450, Family 1 glycosyltransferases are encoded by a multigene family and are ubiquitously found in plants, animals, fungi, bacteria and viruses (Paquette et al., 2003). Almond it is thought to have two kind of UDPG-glucosyltransferase, one able to glucosylate mandelonitrile into prunasin (GT1) (Franks et al., 2008) and the other which glucosylate prunasin into amygdalin (GT2) (Sánchez-Pérez et al., 2008). The same research team found that almond leaf lamina shows low UDPG mandelonitrile glucosyltransferase activity over the entire growth phase independently from the genotype (bitter or sweet). In fruit tissues, the activity of this glucosyltransferase was more dominant in the bitter compared to the sweet variety, but an activity was indeed observed. In contrast to these results, UDPG prunasin glucosyltransferase activity was essentially restricted to the cotyledon with similar activities in the sweet and bitter variety. Although the tegument of the genotype S3067 (sksk, bitter) showed a high content of prunasin whereas prunasin was barely detectable in Ramillete (SkSk, sweet), the radiolabeling experiments did not indicate major differences in biosynthetic capacity between the bitter and sweet genotypes (Sánchez-Pérez et al., 2008).

Franks et al., (2008) characterized the GT1 of almond: a mandelonitrile GT that they named UGT85A19. This enzyme, detected in developing kernels of almond, is evidently synthesised there *de novo* and is not derived from accumulated levels in maternal floral tissues (i.e. pistil) because the protein was not detected in the very early kernel developmental phase. Whether or not the kernels were bitter, kernel protein extracts exhibited mandelonitrile GT activity, and the activity in the bitter types was approximately three-fold greater than that in non-bitter types.

1.1.4 Key enzymes for the catabolism

In plants, degradation of CNGs is catalyzed by β-glucosidases and α-hydroxynitrile lyases (Conn, 1980; Hösel and Conn, 1982; Poulton, 1990).

β-glucosidases catalyze the hydrolysis of glycosidic linkage in aryl and alkyl β-glucosides and in cellubiose. They are present in bacteria, fungi, plants and animals. β-glucosidases generally have a subunit molecular mass of 55–65 kDa, acidic pH optima

(pH 5-6) and an absolute specificity towards β-glucosides (Esen, 1993). Plant β-glucosidases involved in cleavage of CNGs exhibit a high specificity towards the aglycone moiety of CNGs present in the same plant species (Hösel et al., 1987; Hösel and Conn, 1982; Nahrstedt, 1985). Sequence analysis (Zheng and Poulton, 1995) indicates that β-glucosidases fall into two distinct families, designated by Béguin (1990) as families BGA and BGB. The BGA family includes β-glucosidases, phosphor-β-glycosidases, thio-β-glucosidases, and β-galactosidases from organisms as diverse as archaebacteria, bacteria, plants, and mammals. The BGB family includes fungal enzymes and the β-glucosidases of rumen bacteria. The common structural features shown by BGA enzymes suggest a shared mechanism for the enzymatic hydrolysis of β-glycosidic bonds. The BGA enzymes catalyze glycoside hydrolysis by a double-displacement mechanism in which an enzymatic nucleophile attacks the substrate, so forming a glucosyl-enzyme intermediate (Sinnott, 1990). This latter is then hydrolyzed, releasing the sugar with overall retention of anomeric configuration. Aglycone group departure may be aided by a protonated amino acid side chain with acid catalytic function. Amygdalin hydrolase (AH) and prunasin hydrolase (PH) are the β-glucosidases expecially involved in cyanogenesis in *Prunus* spp. (Poulton, 1993). AH cleaves the β-glycosidic bond of amygdalin, yelding the monoglucoside (R)-prunasin, which is subsequently hydrolyzed to (R)-mandelonitrile by PH (Zheng and Poulton, 1995).

Hydroxynitrile lyases (HNLs) catalyzes the final step in the biodegradation pathway of cyanogenic glycosides releasing HCN and the corresponding carbonyl components (Nanda et al., 2005). α-hydroxynitrile lyases have been characterized in plants, where they appear to be located in the same tissues as the CNG degrading β-glucosidases, though their activity is observed in protein bodies (Swain et al., 1992b), instead of in chloroplasts or apoplastic space as typically reported for β-glucosidases (Hickel et al., 1996). α-Hydroxynitrile lyases constitute two broad phylogenetically distinct groups that have convergently evolved to the same function (Møller and Poulton, 1993). One homogeneous group comprises monomeric FAD-containing glycosylated enzymes. These enzymes have only been found in two subfamilies within the Rosaceae (isolated from members of the Prunoideae and Maloideae subfamilies by Poulton, 1988; Møller and Poulton, 1993) and utilize the aromatic cyanohydrin mandelonitrile as a substrate. The other group is more heterogeneous and comprises dimeric or oligomeric non-FAD containing enzymes that typically are not glycosylated. These enzymes have been found in di- and monocotyledonous plant families (e.g.: *Sorghum bicolor*, *Manihot*

esculenta, *Linum usitatissimum*, *Hevea brasiliensis*, *Ximenia americana*, and *Phlebodium aureum*) and their natural substrates may be p-hydroxymandelonitrile as well as acetone cyanohydrin, this latter being the commonest substrate (Zagrobelny et al., 2004). The mandelonitrile lyase (MDL) is one of the most well studied HNLs. It is mainly found in Rosaceae species (Genus: *Prunus*) and catalyzes the formation of benzaldehyde and HCN from (R)-mandelonitrile (Nanda et al., 2005). One puzzling aspect of the biochemistry of rosaceous stone-fruit MDLs is their microheterogeneity (Hu and Poulton, 1999). With only one known exception (Xu et al., 1986), all MDLs isolated from seeds of members of the Prunoideae and Maloideae subfamilies exist as multiple forms (Hickel et al., 1996). Within a given species, these isoforms display slight differences in molecular mass, isoelectric point, electrophoretic mobility and specific activity, but they have identical antigenic properties. Likely sources of MDL microheterogeneity include allelic differences at the structural locus for the polypeptide (i.e. allozymes), alteration of the translated polypeptide (e.g. N- and C-terminal processing, glycosylation, and phosphorylation), and gene duplication leading to multigene families (Weeden, 1983). Among the five black cherry MDLs isolated till now, the amino acid identity ranged from 75.3% to 87.6%. This degree of similarity suggests that they represent members of a multigene family rather than multiple alleles (Hu and Poulton, 1999). The intron locations is conserved, while their length varies considerably among the five genes, presumably as a result of insertions and/or deletions. The mandelonitrile lyase gene mdl3 can be considered as a typical example of the characteristics of this multigene family. Mdl3 and its corresponding cDNA (MDL3) were isolated from black cherry (*Prunus serotina*) by Hu and Poulton, 1997. The gene coding region is interrupted by three short introns, one of them possessing the unusual CC-AC splice junction dinucleotides. This gene encodes a polypeptide of 573 amino acids that includes a putative signal sequence, 13 potential N-glycosylation sites, and a presumptive flavin adenine dinucleotide-binding site. Studies with transgenic tobacco, in which the putative mdl3 promoter was fused to the GUS reporter gene, allowed to localize the expression of mdl3. Developing seeds constituted a major site of expression, whereas the youngest immature embryos showed no detectable activity. As the embryos grew, the expression of the gene increased. These results suggest a developmental regulation of mdl3 expression. These temporal and spatial patterns of activity agree well with earlier work on MDL expression in developing black cherry seeds (Zheng and Poulton, 1995), in which it was showed that MDL transcripts were first detectable in cotyledonary parenchyma cells at approximately 40 DAF, when embryos

became visible to the naked eye. Transcript levels peaked at approximately 49 DAF before declining to negligible levels at fruit maturity (82 DAF). Confirming previously published data (Swain et al., 1992a), MDL enzyme activity was undetectable until 44 DAF; it then increased rapidly, essentially reaching a plateau at fruit maturity (Zheng and Poulton, 1995). In addition to its major location in seeds, black cherry MDL is also found in cotyledons, epicotyls, and hypocotyls of germinating seedlings (Swain and Poulton, 1994). Thus, MDL, which constitutes approximately 10% of the soluble protein of black cherry seeds (Zheng and Poulton, 1995), is believed to be multifunctional, serving both as a storage protein and in cyanogenesis (Swain et al., 1992b). *In situ* hybridization analysis carried on by Zheng and Poulton (1995) showed that the spatial expression of MDL1 mRNA differs greatly from that of AH1 mRNA, although both transcripts show similar temporal expression patterns in developing seeds. Moreover, in contrast to the procambial location of AH1 transcripts, MDL1 mRNA exhibited a spatial expression pattern more characteristic of storage proteins, being restricted to the cotyledonary parenchyma cells.

1.1.5 Cyanogenesis in the Prunus *genus*

Among the most highly cyanogenic species known are the rosaceous stone fruits (e.g. plums, apricots, peaches, and cherries), whose kernels are a rich source of (R)-amygdalin (the β-gentiobioside of (R)-mandelonitrile) and its catabolic enzymes, whereas their flesh is considered acyanogenic (Poulton, 1983). Containing both the cyanogenic diglucoside and its catabolic enzymes, these seeds constitute miniature cyanide bombs capable of a rapid HCN release upon tissue disruption. Fortunately, such large-scale, suicidal cyanogenesis is apparently avoided in undamaged tissue by compartmentation (Poulton and Li, 1994). So, non-bitter and bitter kernel phenotypes can be dramatically discriminated by measurement of amygdalin (Dicenta et al., 2002). The cyanogenicity of rosaceous stone fruits has long been known (Liebig and Wohler, 1837). Whereas leaves and immature fruits of these species contain the cyanogenic monoglucoside (R)-prunasin, the predominant cyanogen in mature fruits is the corresponding diglucoside (R)-amygdalin. Upon seed disruption, amygdalin is rapidly degraded to HCN, benzaldehyde, and glucose by the sequential action of amygdalin hydrolase, prunasin hydrolase, and MDL.

P. turneriana. All its tissue are highly cyanogenic, containing combination of the cyanogenic glycosides (R)-prunasin, (S)-sambunigrin and amygdalin (Miller et al., 2004). The progeny of a single parent tree varied markedly and continuously in their cyanogenic glycoside content, indicating that this variation in genetically based. There was an increased allocation of cyanogenic glycosides to old, expanded and photosynthetically productive leaves, a pattern which appears to be inconsistent with predictions of optimal defence theories, and the results of other studies. Miller and her team suggested that such a strategy may be advantageous for seedlings of tree species that can only reach a reproductive stage following the creation of a canopy gap. Amygdalin was found restricted to seeds, while prunasin was found in all plant parts. This distribution pattern is typical of other Rosaceae. The amount of amygdalin generally increases at the expense of the prunasin during seed maturation, and there is evidence that this cyanogenic di-glycosides in seeds are metabolized upon germination, and used for synthesis of non-cyanogenic compounds during seedling development (Miller et al., 2004).

P. amygdalus* (almond*). While the kernels borne by cultivated almond genotypes are pleasantly flavoured, wild almond kernels can be distasteful and poisonous (Ladizinsky 1999). Bitterness in almond is determined by the content of the cyanogenic diglucoside amygdalin. Sweet kernel in almond is one of the principal breeding target for almond breeders and growers (Sánchez-Pérez et al., 2008). Sweet kernel in almond has been shown to be a monogenic trait and the bitter kernel trait to be recessive (Heppner, 1923, 1926; Dicenta and Garcìa, 1993; Dicenta et al., 2007). The gene conferring sweetness (Sweet kernel [Sk] gene) belongs to linkage group five (Joobeur et al., 1998; Bliss et al., 2002; Sánchez-Pérez et al., 2007) but its precise localization and function remain mostly unknown. During the entire season, prunasin can be detected in the vegetative part of all, bitter and sweet, genotypes, but the content was always several fold higher in the bitter ones (Dicenta et al., 2002). Previous studies in almonds have shown that prunasin is transformed into amygdalin during fruit ripening (Frehner et al., 1990).

If the biosynthetic steps are less-well known, the enzymes and genes involved in the amygdalin degradation have been extensively studied and characterized (Fig. 1.4). The first step in amygdalin degradation is mediated by the β-glucosidase amygdalin hydrolase, and results in the formation of prunasin and concomitant release of glucose. Prunasin is subsequently hydrolyzed by another β-glucosidase, the prunasin hydrolase,

to form mandelonitrile and glucose. Mandelonitrile is finally converted into benzaldehyde and HCN by the action of mandelonitrile lyase.

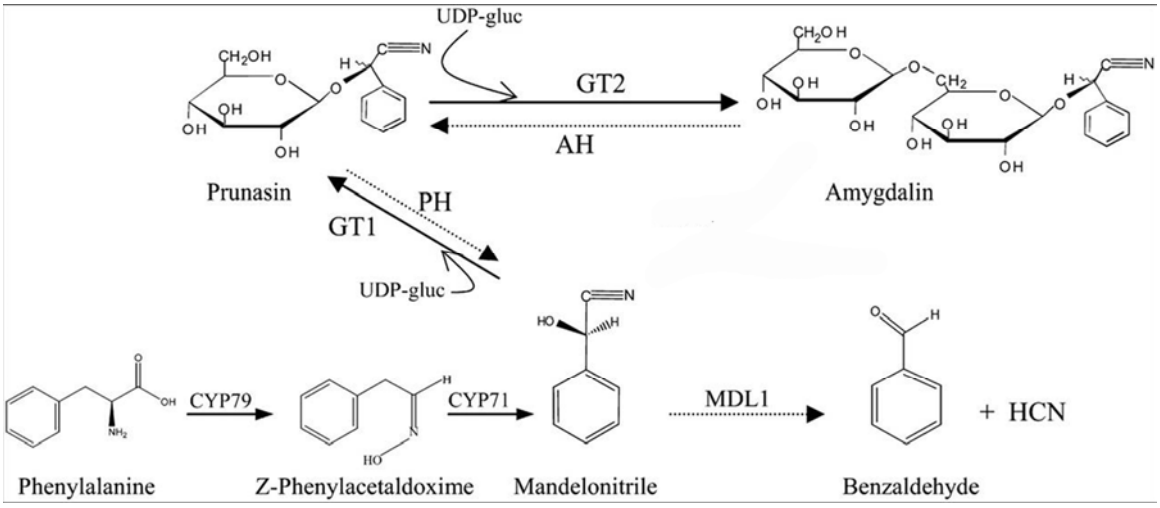


Fig. 1.4. The metabolic pathways for synthesis and catabolism of the cyanogenic glucosides prunasin and amygdalin in almonds. Biosynthetic enzymes are highlighted with black lines, catabolic enzymes with dashed lines. (Sánchez-Pérez et al., 2008).

At the beginning of the growth season, the content of prunasin increases in leaf lamina and stems. This may be related to mobilization of reserves stored in lignified tissues throughout the winter. The prunasin content in stems, petioles, and leaf laminas decreases in the subsequent period (April). This coincides with the initiation of secondary growth and the formation of terminal shoot sprouts, that later on results in a significant expansion of the tree crown and in the formation of the different fruit tissues (exocarp, mesocarp, endocarp, and kernel; Girona and Marsal, 1995). Between March and April the fruits reach maximum size, and prunasin levels increase in all genotypes. In contrast to the results found for the vegetative tissues, prunasin was only detected in the fruits of the bitter genotype (Sánchez-Pérez et al., 2008). In the sweet variety Ramillete, the staining experiments with Fast Blue BB salt showed the existence of a continuous β-glucosidase-rich cell layer in the inner epidermis of the tegument, facing the nucellus. If transport of prunasin from tegument to cotyledon proceeds *via* the symplast, this would indicate that the prunasin synthesized in the tegument of the sweet variety (SkSk) is degraded when it has to pass this cell layer. But why would prunasin first be synthesized in tegument tissue of the developing almond fruit in the sweet variety and subsequently be subjected to degradation? It is possible that in almond an active cytochrome P450 system (similar to that working in sorghum) could convert the excess of free phenylalanine transported to or synthesized within the tegument into prunasin. Degradation of the prunasin by the action of the cyanogenic β-glucosidase would then result in the release of HCN, which,

by the action of β-cyano-Ala synthase, would first be converted into β-cyano-Ala and subsequently converted into asparagine and aspartic acid. Recently a different pathway for cyanogenic glycoside catabolism involving nitrilase heterodimers has been suggested to operate in sorghum (Jenrich et al., 2007; Kriechbaumer et al., 2007). In this pathway, the nitrogen atom of the nitrile group is recovered into ammonia without the release of toxic HCN. The two pathways offer the opportunity to convert Phe into ammonia, Asp, and Asn, which could be used as general precursors for amino acid and protein synthesis in the developing cotyledons. The sweet almond varieties may thus profit by having a more balanced and alternative direct supply of free amino acids for protein synthesis in the developing cotyledons whereas the bitter variety accumulating amygdalin in the cotyledons may profit from the protection offered by this secondary metabolite toward herbivores and pests (Sánchez-Pérez et al., 2008). While prunasin is found only in the vegetative part (roots and leaves), amygdalin was detected only in the kernels, and mainly in bitter genotypes. Its content was detected in kernels by HPLC, but with variable values (between 0 and 411 mg CN-/100 g) depending on the genotype (Dicenta et al., 2002). Prunasin thus appears to be the form of cyanogenic glycoside transported in the plant while amygdalin is utilized for storage, as previously suggested by Frehner et al., (1990).

Prunus armeniaca **(apricot):** Apricot seeds contain the toxic cyanogenic glycoside amygdalin, accompanied by minor amounts of prunasin. These CNGs are synthesized by the maternal tissues and transported to the developing seeds. Indeed, in apricot as well as in almond (Kester and Asay 1975) all the seeds produced by a single tree show the same phenotype, i.e., either bitter, sweet (non-bitter) or slightly bitter; thus, the presence of cyanoglucosides in seeds has to be considered as a seed-parent trait. How these compounds can move in plant tissues without being cleaved by catabolic enzymes (which would result in HCN production) remains an unanswered question. Mostly unknown, too, are the underlying mechanisms of differential accumulation of cyanoglucosides in plant organs: this may be attributed either to a tissue specific synthesis or, in the acyanogenic organs, to an enzymatic cleavage without HCN production, with CNGs immediately converted into non-cyanogenic compounds by detoxifying enzymes such as β-cyanoalanine synthase (Miller and Conn 1980; Selmar et al., 1988; Swain and Poulton 1994).

Different hypotheses on the inheritance of the bitter/sweet trait have been proposed (Kostina 1977; Bassi and Negri 1991) but no definitive model has been

demonstrated. In 2008 Negri et al., given that both the bitter and the sweet phenotypes were represented in populations from bitter×bitter and sweet×sweet crosses as well as from self-pollination of either bitter- or sweet-seeded trees, postulated that more than one gene is involved in this trait. So, taking into account the evidence against both the quantitative and the monofactorial inheritance, that research team assumed that several genes may be involved in seed bitterness. Therefore they considered that seed bitterness could result from two distinct biochemical pathways: an "additive" pathway (either the biosynthesis of cyanoglucosides or their transport or both) leading to the storage of these metabolites in seeds, and a "catabolic" pathway of cyanoglucosides, possibly associated with their simultaneous conversion into noncyanogenic compounds. Moreover, each of these pathways could be made up of one or possibly more steps, controlled by genes that can also be unlinked. The functioning of both the anabolic and the catabolic pathways would be ensured by the dominant alleles at each one of the involved loci. The bitter phenotype would then result from a complete anabolic pathway, due to the presence of at least one functional allele at each one of the involved loci (that there named B) in the absence of any inhibitor gene product, associated with the breakdown of the catabolic function, attributable either to recessive homozygosis at one or more of the corresponding loci (called S) or to inhibitor gene dominant allele(s). Conversely, the sweet phenotype could be due either to the breakdown of the anabolic pathway, or to the triggering of the catabolic function, or to both (Fig. 1.5).

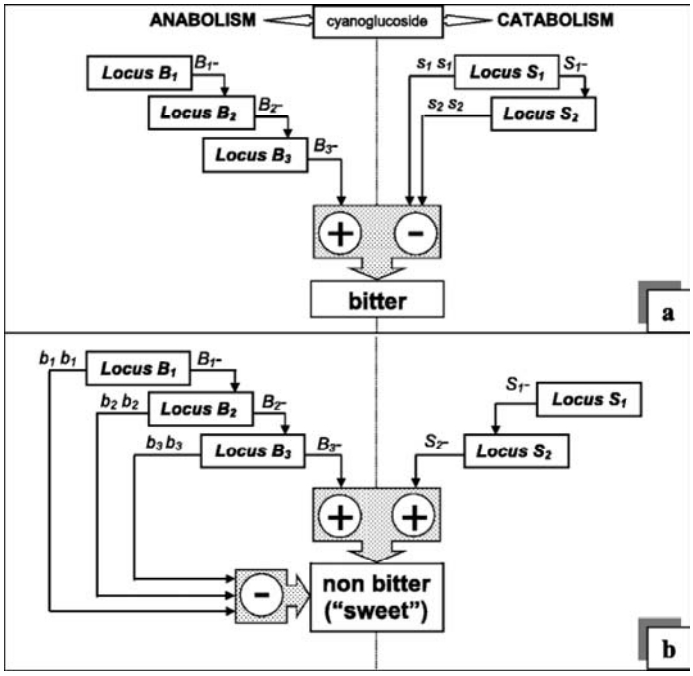


Fig. 1.5. Flow chart of the inheritance hypothesis of Negri et al. 2008. The bitter (a) or non-bitter phenotype (b) results from success (+) or fail (-) of anabolism (3 genes) and catabolism (2 genes)

In this model, the sweet seedlings issued from self-pollinated bitter genotypes (supposedly lacking the "subtractive" function) can be explained by the segregation of recessive alleles (either at any of the loci responsible for the anabolic pathway, or at the catabolism inhibitor gene). At the same time, this hypothesis accounts for the bitter seedlings present in a self-pollination population of a sweet tree. Given that the sweet phenotype can be due to either the catabolism or the inhibition of anabolic enzymes, the bitter seedlings would then result from the segregation of recessive alleles at one or more of the corresponding loci. According to this model, the bitter phenotype would correspond to 40 (B1- B2- B3- s1s1--, or B1- B2- B3- S1- s2s2) out of 243 possible genotypes. All the segregation ratios observed by Negri et al., (2008) matched with that inheritance mechanism of five, non-linked genes (Fig. 1.6).

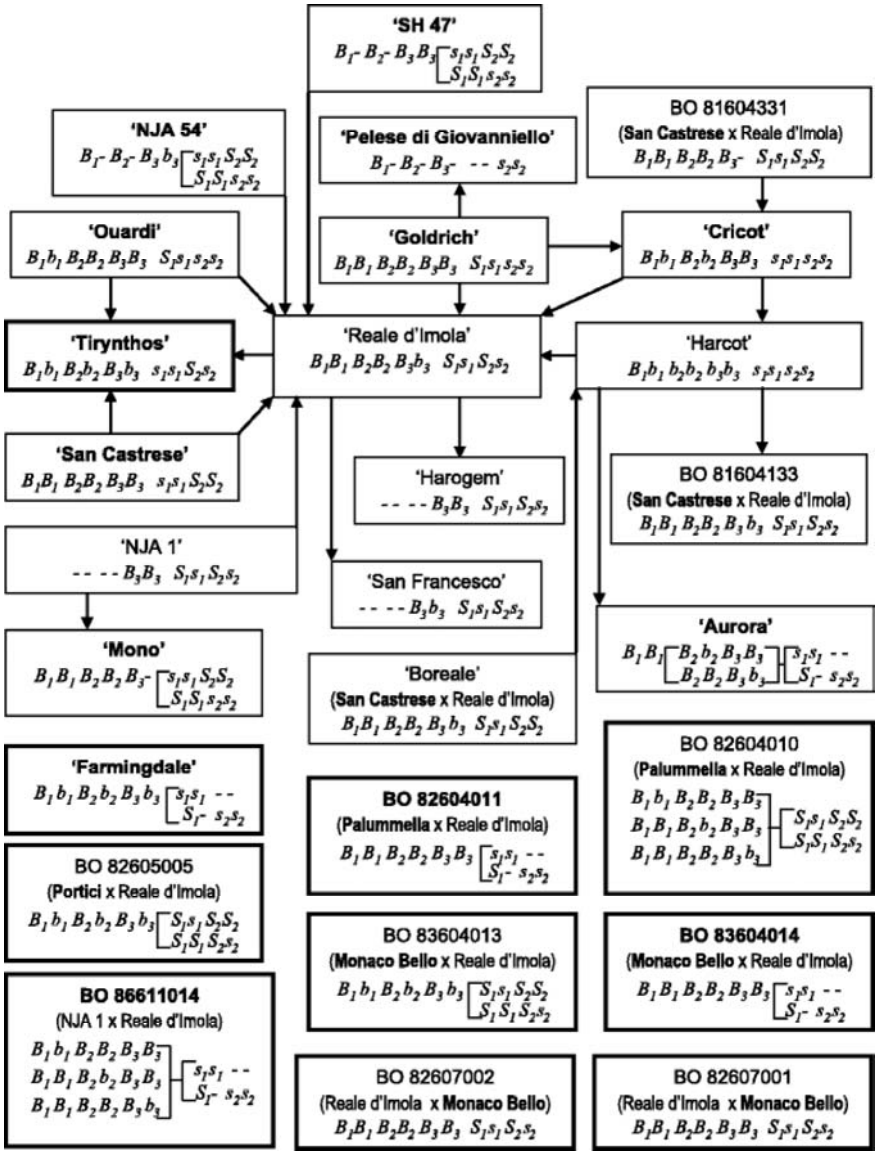


Fig. 1.6. Tentative genotypes (in italics) assigned consistently with all the observed segregation (Negri et al., 2008). Typed in bold: bitter parents. Framed in bold: self-pollinations. Arrow (pointing to the pollen donor): evaluated cross.

1.2 The apricot system

1.2.1 Systematic placement.

The apricot was known in Armenia during ancient times, and has been cultivated there for so long that it is often thought to be native there. Its scientific name *Prunus armeniaca* (Armenian plum) derives from that assumption. But the most probable geographical origin is the north-east area of China, with second derivation centres in the Xinjiang region and in central Asia. From the border between China and Russia the apricot cultivation would have been extended toward the west, until reaching Armenia. Its later introduction to Greece is attributed to Alexander the Great. But its diffusion in the Mediterranean area was consolidated only afterwards from the Arabians: in fact the name 'apricot' comes from the Arabic word 'Al-barquq'. Today the cultivars have spread to all parts of the globe with climates that support it. The apricot is a small tree with a dense, spreading canopy. The leaves are ovate, with a rounded base, a pointed tip and a finely serrated margin. The flowers have five white to pinkish petals andare produced before of the leaves, singly or in pairs, in early spring. The fruit is a drupe (from the greek drypepa: that ripe [verb peptein] on the tree [drys]), coloured from yellow to orange and often red on the side most exposed to the sun, with an external thin membrane (called exocarp or skin, velvety to the touch and often pubescent), a central fleshy part (mesocarp or pulp), an inner woody part (endocarp or pit), and a single seed enclosed in the hard stony shell (Fig. 1.7).

Fig. 1.7. Examples of apricots from the parental lines BO81604311 and Lito and of one of their hybrids (L×B 64)

1.2.2 Genetic improvement

The apricot genome is relatively small ($5.9*10^8$ bp/2n, only the double of the *Arabidopsis'* one), the base chromosome number is 8 and all the known species are diploids. The Russian and Asiatic cultivars show a huge diversification regarding the majority of the agronomic traits of interest, whereas the European group, deriving from a small number of lines, is characterized from the less genetic variability. The big goals for the apricot breeders, apart from the organoleptic and commercial quality, focus on the biotic and abiotic resistances and the adaptability to different pedoclimatic conditions, as summarized in the following table (Tab. 1.1). Special attention is given to the extension of the ripening period, objective that has become very important in the latest years, most of all for what concern the 'late ripening' trait (Bassi and Rizzo, 2004).

Tab. 1.1. Major goals of the apricot genetic improvement.

Organoleptic and commercial fruit quality traits	ϒ High size. ϒ Orange skin with brilliant red overcolour. ϒ Tasty and aromatic pulp with equilibrated rate between acids and sugars. ϒ High consistency. ϒ Tolerance to manipulation and refrigeration.
Favourable fruiting traits	ϒ Small vegetative habits. ϒ Flower fertility. ϒ Fruiting efficiency. ϒ Uniform ripening. ϒ Autocompatibilty
Biotic stress resistance	ϒ Fungi (*Monilia laxa* and *fructigena*) ϒ Bacteria (*Pseudomonas*, *Xanthomonas*) ϒ Virus (Sharka) ϒ Phytoplasma
Environmental adaptability	ϒ Resistance to the cold winters. ϒ Tolerance to the spring colds. ϒ Late flowering ϒ High and constant fruiting in every environment.
Extension of the ripening calendar	ϒ Earliness. ϒ Lateliness.

A genetic improvement program is based on a series of crossing, in order to completely tap the natural existing variability inside the same species (intra-specific cross) or in related species (inter-specific cross). This last method is particularly used for the species with a small genetic background, for which the new useful traits have to be found in wild related species (Foulongne et al., 2003). After crossing, the next step is the selection of the best genotypes. This is a very long process, because only when the plants have overcome the juvenile phase is really possible to evaluate the productiveness related characteristic and the fruit quality. This represents the biggest disadvantage of the genetic improvement of the woody plants.

1.2.3 Early selection and molecular markers

All that said, is clear the importance of having available methods allowing the earliest possible selection of the cultivars with the desired characteristics, and, moreover, without any contemporary correlation of these traits with some negative ones (Quilot et al., 2004). The only way, hence, not only to speed up but also to lower costs and spaces usually required for the traditional breeding, is the use of molecular markers for the early assisted selection. The MAS (Marker Assisted Selection) is done by direct analysis of the plant genotype with a marker tightly linked to the desired trait. That means the possibility of recognizing the individuals with certain characteristics way before these would be showed in the field.

There are basically four kind of markers: morphological markers, for which the analysis is done at the phenotypic level (e.g.: pigmentation); cytogenetic markers, with analysis conducted via microscope techniques (e.g.: heterochromatic areas of chromosomes); biochemical markers, that rely on electrophoretical analysis of proteinic preparates (e.g.: isoenzymes); molecular markers, whose analysis can be directly done at the DNA level. The use of molecular markers in the breeding programmes has allowed to solve quite a lot of its common problems, just like the time-consuming feature, the difficulties in the determination of the character, the interaction between the genotype and the environment, the selection's complications for the quantitative traits). A molecular marker can be defined as a genomic locus (of variable dimension, generally between 50 and 3000 base pairs) detectable with specific probes or primers, whose simple

presence is sufficient to mark in a characteristic and unequivocal way a chromosomal stretch. The fundamental attributes of molecular markers are: being able to distinguish the parents and segregating in the progeny. In this way they allow to follow the chromosomal stretch passing from one generation to the other. These differences in the DNA level can be then visualized by gel electrophoresis on agarose or polyacrylamide. Compared to other kind of markers, the molecular ones offer quite a lot of advantages: they are widespread and easily available in the genome, they are present in good agronomic genotypes, they are not affected neither by the environment, nor by the developmental stage of the plant, they are phenotipically neutral (that is there aren't any kind of epistatic interaction or pleiotropic effect, because the presence of one allele or another is not influencing the external phenotype) and moreover, having a simple mendelian heredity, they are ideal for genetic studies. Beyond these, other idyllic features sought after in a potential marker are: high polymorphic level at the single locus, codominance, possibility of identify more loci, transferability intra and inter species, low analysis costs.

The most useful molecular markers are the functional ones. These markers are polymorphic genomic traits amplified by primers designed particularly on the codifying sequence of the genes involved in the character determination. Their high specificity allow to work with elevate annealing temperature in PCR, so assuring an almost total absence of false positive and a very good reproducibility. As far as now their principal limit was the absolute need of knowing the candidate gene's sequence, but now, thanks to the enrichment of the databases, the situation has highly improved. It's still open, instead, the problem related to the high conservation of the most important genes inside the same species and also among species; that makes difficult sometimes to find polymorphisms. The SCAR markers (Sequence Characterized Amplified Regions) were born as 'transfer markers': they convert random markers in specific and transferable ones. They are obtainable directly from the amplification of a DNA fragment by using two primers specifically drawn on the nucleotidic sequence. So, to identify this kind of marker is required to clone and sequencing a PCR product previously amplified with heterologous primers or with RAPD or AFLP techniques. On the obtained sequence, is then possible to draw the new specific primers. Beyond the high specificity, the SCAR markers have the advantage of being often codominant. They have been extensively used

also because they require small and not so purified initial amounts of DNA, so allowing to screen high numbers of individuals in small periods of time. A disadvantage was, until now, the high economic cost for their identification. But the availability on the internet of lots of genetic sequences has highly reduced the negativity of this aspect. When direct polymorphisms are not present, is possible to discriminate the PCR products by an enzymatic digestion, obtaining in this way another kind of marker, so called CAPS (Cleaved Amplified Polymorphic Sequence). In this way is possible to pick out differences that were invisible before, due to the identity in the molecular weights. This kind of molecular markers have many advantages, just like specificity, reproducibility, codominance, all that at the expense of the costs, quite high also in this case. It has to be noticed that in apricot species the necessity of knowing the sequence of the studied genes is still an open problem because, no matter its economic relevance, the identification of genes controlling the fruit quality or the agronomic traits is extremely limited (Grimplet et al., 2005). In fact, when searching the databases containing all the EST (Expressed Sequence Tags) annotated to this day, is remarkable to notice that, whereas for the peach 79.580 sequence are published, for the apricot only less than a quarter are available: 15.105 (and this is the same number of the 2007!) This is a very low score, barely equivalent to the half of the known EST of plants with a low economic value, like the wild willowleaf lettuce (*Lactuca saligna*) or the black maidenhair fern (*Adiantum capillus-veneris*); to one tenth of the Chinese cabbage (*Brassica rapa* sub. *pechinensis*) ones, one sixteenth of the lotus (*Lotus japonicus*) ones and to only one twenty second of the loblolly pine (*Pinus taeda*) ones.[1]

1.2.4 Genetic maps

One of the main application of molecular markers, beyond their use in the assisted selection, is the building of genetic maps. Only polymorphic loci in the parents and whose polymorphism segregate in a cross population can be mapped. Considering that bigger is the distance between linked markers, higher are the odds that a recombining event could occur in the region included amid this markers, the recombinant percentage among markers become the unit of measurement of the molecular genetic maps. So, a genetic map results to be a linear representation of the chromosomes, reporting the relative distances between the various loci in recombinant

[1] **http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html**

frequency units (map units, m.u. or centimorgans, cM). The centimorgan is equal to a 1% chance that a marker at one genetic locus on a chromosome will be separated from a marker at a second locus due to crossing over in a single generation. The number of base-pairs to which a cM corresponds varies widely across the genome (different regions of a chromosome have different propensities towards crossover), on average, 1 centimorgan corresponds to about 1 million base pairs in humans. To give an example: *Plasmodium falciparum*, a protozoan parasite, has an average recombination distance of ~15 kb per centimorgan, that means that markers separated by 15 kb of DNA (15,000 nucleotides) have a 1% chance of being separated by crossing over in a single generation. The building of a genetic map is done in three steps

1) Partition of the loci in linkage groups.
2) Calculus of the distances between all the loci's (or gene's or marker's) couples inside a linkage group.
3) Ordering of the loci inside every linkage group.

Statistical software are available for this kind of association analysis (e.g. Joinmap, Mapmaker) allowing to assemble the markers in linkage groups. Two markers are considered associated or not on the basis of the χ^2 test, which is used to compare experimental versus expected results. To build a genetic map is preferable to use for first codominant markers [like RFLP (Restriction Fragment Length Polymorphism) and SSR (Simple Sequence Repeat)], namely markers able to discriminate the heterozygous, to which add, subsequently, AFLP (Amplified Fragment Length Polymorphism) to saturate (that is to cover the whole genome with really small gaps between each marker) the basic structure. To gain a very detailed map the identification a high number of polymorphic markers well distributed among the genome of the studied species is needed: they have to be able, in fact, to represent in the map all the chromosomal regions. The genetic maps applications are many and fundamental. In fact a genetic map:

- is the starting point for the isolation of genes of interest by positional cloning;
- allows to optimize the choice of the markers for the assisted selection;
- permit to evaluate the positional effect of a gene on a chromosome on the same gene's expression;

- facilitate the transferring of information between species by exploiting the syntenic relations among the various maps (Pierantoni et al., 2004, Dondini et al., 2007);
- is the basis for the identification of those genes controlling the quantitative traits (Dondini et al., 2004; Illa et al., 2010).

Concerning the P*runus* genus, the genomic tools are still developing (Etienne et al., 2002). But from the experiments carried out till now it was already possible to determine how the internal evolution of the chromosome set was leading only to a small number of inversions and translocations. The *Prunus* genus, at least at the genomic level, is behaving as a single genetic entity (Dilewanger et al., 2004). These results are concord to the various crossing patterns presents inside this genus, where the inter-specific hybridizations, also across bridge species, are usually possible and the hybrids often fertile. The presence of this high level of synteny (co-linearity, especially between the genomes of almond, apricot, cherry, peach, *P. cerasifera*, *P. ferganensis* e *P. davidiana*) make easier every kind of genetic investigation. This is true most of all for what regard the saturation of the genetic maps and the searching of candidate genes, thanks to the possibility of using linkage data already known and DNA stretches already sequenced from related species as a starting point for the new analysis.

For the researchers working on the *Prunus* a very useful instrument is the Prunus Reference Map, a saturated genetic map elaborated on a F_2 population derived from crossing almond (cv. Texas) with peach (cv. Earlygold). Initially the eight linkage groups (corresponding to the eight chromosome characteristic of this genus) were identified using only 246 markers: 11 isoenzymes and 235 RFLP (Joobeur et al., 1998). A remarkable improvement was done by the research group of Aranzana with the addition of 96 SSR, which allowed to close some gaps and to increase the map density, beyond the big advantage given by the possibility to use some of these SSRs to discriminate different cultivars and to anchor the TxE map to other *Prunus* maps (Aranzana et al. 2003). Today the Prunus Reference Map contains 562 markers with a total covered distance of 519 cM (map density: 0,92 cM per marker). Among the numerous advantages of having this kind of map there are: possibility of using a highly polymorphic population for all the association studies, establishment of a common nomenclature for the linkage groups, development of a set of transferable markers with

known position, opportunity of compare the maps of different *Prunus* species and positioning on a unique map of many major genes and QTL.

For what concern more especially the apricot, the first map was elaborated by Hurtado et al., (2002) and was designed on a F_1 progeny (Goldrich×Valenciano) segregating for the Plum Pox Virus (Sharka) resistance. A total of 33 RAPD, 82 AFLP, 4 RFLP and 17 SSR were positioned on the parent Goldrich and 19 RAPD, 48 AFLP, 16 RFLP and 7 SSR on Valenciano, so giving rise to a map with eight linkage groups in 511 cM for the first parental and seven LG in 467.2 cM for the second. It was clear at this point the need of having available a higher number of microsatellite markers, to gain eight LG also in Valenciano and to be able of really compare the genomic organization between the apricot and the other *Prunus* species, with a precise hint to the Prunus Reference Map (Hurtado et al. 2002).

A subsequent map was made by Vilanova et al. (2003) on a F2 population coming from a self-fertilization of Lito and segregating for the resistance to PPV and the self-incompatibility. A total of 211 markers (180 AFLP, only 29 SSR and 2 agronomic traits) individuated 11 association groups for a length of 602 cM. The presence of several and quite wide gaps (between 10 and 20 cM), the high number of markers that remained un-linked with the others and the identification of three LG more than the basic number of the *Prunus* chromosomes, imposed again the necessity of improving the numbers of the usable markers.

So, a third map was designed on the cross Polonais×Stark EarlyOrange from Lambert et al., (2004). For its building numerous markers from the T×E map were used, with a particular attention on the multiallelic and codominant SSR, ideal for being used as anchor loci. These markers, resulting polymorphic also on the experimental population, allowed to obtain a perfect alignment with the Prunus Reference Map. Eight LG were achieved for both the parental, and the homology between these and the ones belonging to the T×E turned out to be really high, also for what was concerning the order of the markers on the chromosomes, almost identical.

At last two other maps were realized in 2007 from Dondini et al., (see Fig. 1.8a and b) for the apricot accessions Lito and BO81604311 (a selection derived from San Castrese×Reale di Imola), segregating for the Sharka resistance (Dondini et al., 2010), the fruit size and the organic acid and soluble solid content (Ruiz et al., 2010). The two maps are long respectively 532 and 601 cM, with 161 and 168 markers respectively, 96 of which are anchor markers showing a complete colinearity between the two genomes

and an optimum alignment with all the other maps based on SSR and available in *Prunus*, both for what concerns the association groups division and the maintenance of the marker's order inside them.

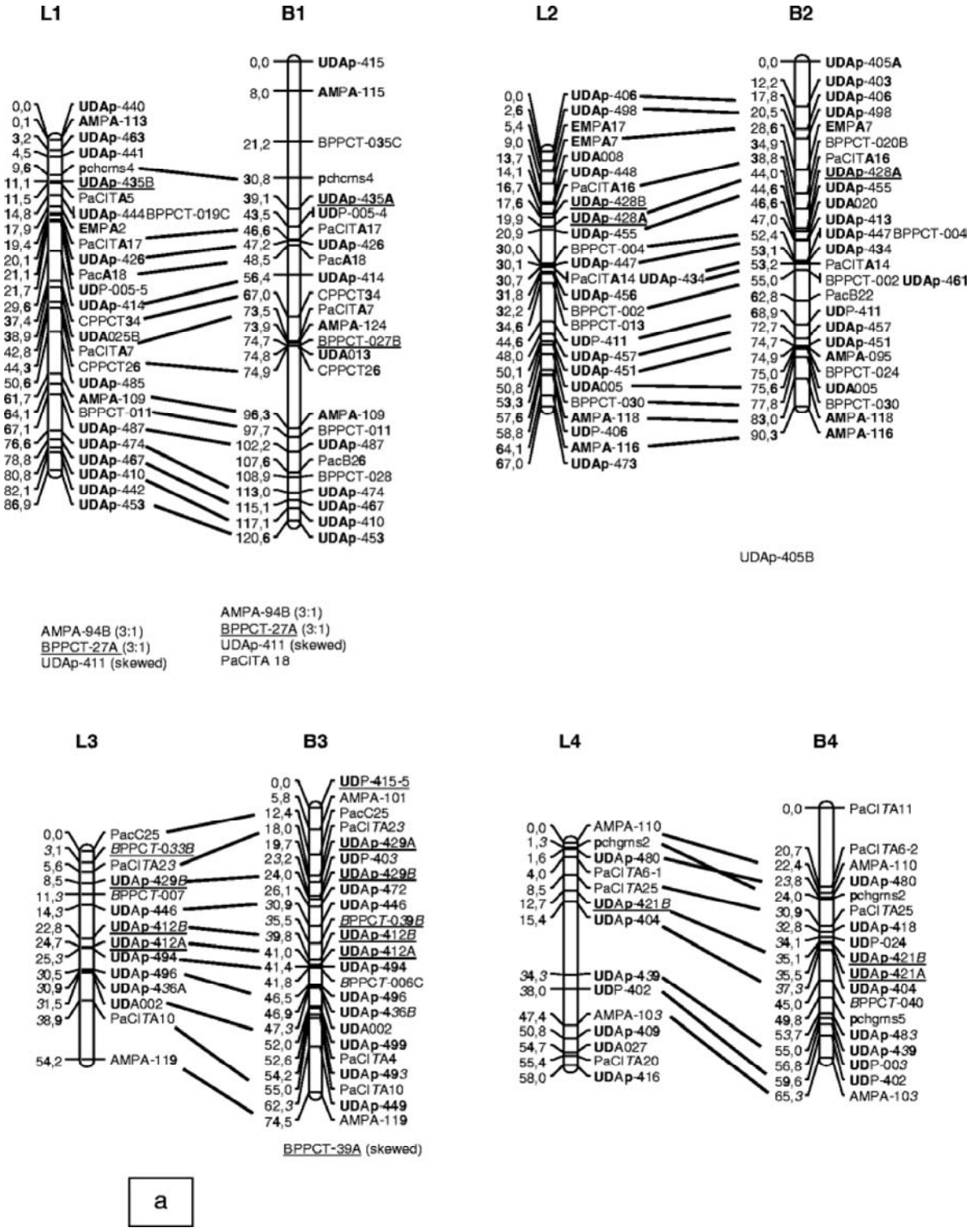


Fig. 1.8a, b. Map of Lito(L) and BO81604311(B). Markers style: assigned but discarded, underneath relative LG; multilocus, underlined with individual loci indicated by letters; not previously mapped, bold; identifying new loci: italics. (Dondini et al., 2007)

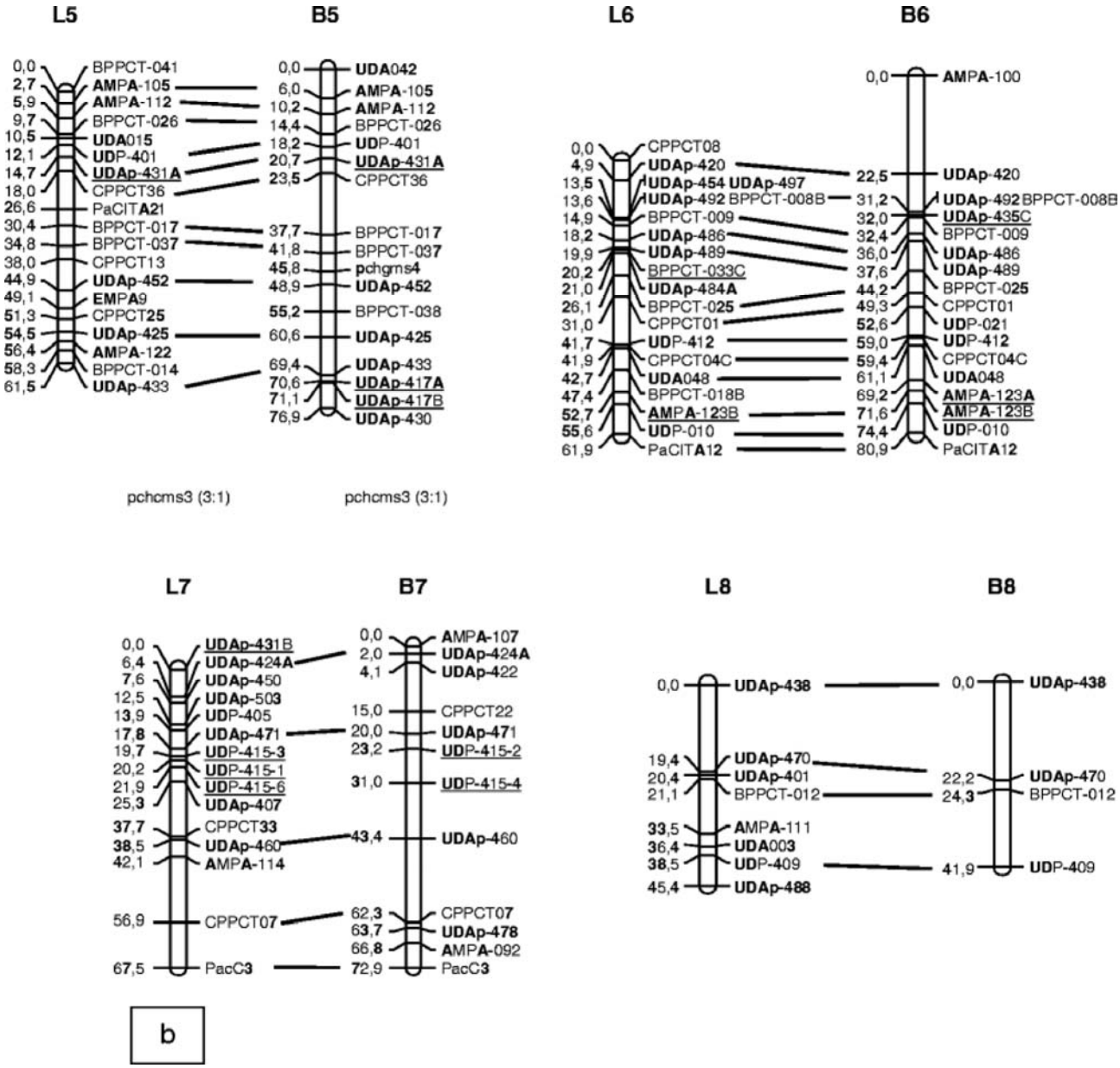


Fig. 1.8a, b: map of Lito(L) and BO81604311(B). Markers style: assigned but discarded, underneath relative LG; multilocus, underlined with individual loci indicated by letters; not previously mapped, bold; identifying new loci: italics. (Dondini et al., 2007)

1.2.5 Quantitative trait loci analysis

Most of the phenotypic traits of the woody plants are not consequent of the a single gene expression. Quite the opposite, lots of the characters whose improvement is wanted (plant growing rate and bearing, productivity, ripening and harvesting time, fruit quality, etc…) are actually the result of the expression of a multitude of genes and of the interaction between them and with the environment. These kind of characters, named quantitative ones, are indicated with the acronym QTL: Quantitative Trait Loci. Their

distribution is continuous and their analysis is complex and almost impossible to carry on with the mere help of the formal analysis that is usually suitable for the qualitative characters. This is true even if, using adequate crossing plans, it is technically possible to resolve the total variance associated in the single components. The genetic variability that lies under the quantitative phenotypes comes out, in fact, from the segregation of numerous QTLs, everyone of them responsible for only a part of the variance. A very efficient approach to study these QTLs is the one called of the Candidate Genes (Pflieger et al., 2001; Etienne et al., 2002). It is based on the *a priori* choice of one or more genes that seem to be functionally connected to the examined character: a correlation between the phenotypic data and an allelic polymorphism validate the hypothesis that the chosen gene controls, at least partly, the studied trait (Causse et al., 2004; Illa et al., 2010; Le Dantec et al., 2010). On the contrary its involvement in the variation of the character is excluded, the chosen gene discarded, and the study will go on with the choice of a new candidate gene.

The associations between the agronomic traits under quantitative control and the molecular markers have been extensively studied in the last decade starting from the herbaceous cultivated species. This because the perennial fruit plants are typified by long generation times and big size, features that represent a huge obstacle to the genetic studies, leading, logically, to a slower breeding progress (Joobeur et al., 1998; Hurtado et al., 2002). Therefore is not that difficult to understand how the QTLs identification was initially focused (and successful) on the herbaceous species in which the self-fertilization was also a possibility. In these cases the studies could have been based on near isogenic lines (that is, lines that differ only in the QTL zone) and on the analysis on thousands of individuals (Morgante and Salamini, 2003). Chronologically the first experiments gave a result on maize in 1987, and in the following years it was possible to clone numerous QTLs always on this species (apical dominancy, flowering time), but also in *Arabidopsis* (flowering date and hypocotyl elongation), rice (numerous loci, at least 14, for the flowering time) and tomato (sugar content, pH, weigh and shape of the fruit), only to quote the most well documented example (Paran and Zamir, 2003).

The quantitative characters identification in the woody fruit plants started a full decade after, and initially only with the identification of QTLs for the growing and the developing of the apple and for the Plum Pox Virus resistance in peach (Wang et al.,

2000). Afterwards were mapped: in *Pyrus* the bacterial fire blight resistance, in *Malus* the scab one, in *Citrus* the number of fruit per plant, in *Vitis* the berry weight and the seed number. As far as the *Prunus*, QTLs have been mapped in peach for: powdery mildew resistance (Foulongne et al., 2003), productivity, fresh weight, pH, titratable acidity, soluble solid content, ripening time, sugar metabolism (Dirlewanger et al., 1999 and 2006), kernel size, red colour of the skin, diameter of the suture (Quilot et al., 2004); in cherry for: flowering and ripening date, death percentage of pistils, pollen germinability, fruit weight and soluble solid content (Wang et al., 2000); and in almond for: leaves developing and flowering time, self compatibility, productivity, ripening date and fruit texture (Sanchez-Pérez et al., 2006). The number of QTLs related to fruit organoleptic or nutritional quality in stone fruit is likely to increase when the results of ongoing international projects, such as ISAFRUIT (Callesen, 2009), will be released (Audergon et al., 2009; Illa et al., 2010).

The apricot is the species less genetically characterized among all the ones comprised in the *Prunus* genus (Ruiz et al., 2010). Until now only the hereditability of a few traits has been determined: Sharka resistance (Lambert et al., 2007), self compatibility and male sterility (Hurtado et al., 2002). An improvement in the saturation and precision of the genetic maps and in the number of polymorphic markers available for the association analysis is so essential to be able to identify and locate the QTLs in the various linkage groups of that plant (Foulongnge et al., 2003, Audergon et al., 2009).

1.2.6 Fruit quality

The introduction in cultivation of new apricot varieties cannot disregard from the study of the agronomic traits of the plant and the fruit as well as of the satisfaction of the customers (Ruiz et al., 2010). In fact, as previously reported, numerous are the studies conducted and planned to reach these objectives: on the adaptability of the cultivar to the environment, on the biotic and abiotic stress resistance, on the commercial qualities of the fruits (size, storability, tolerability to the refrigeration and the manipulation during the transports). An aspect that until now was pushed into the background and only recently have been under the eyes of everyone, is the one regarding the organoleptic qualities of the fruit (mostly defined by the sugar and acid composition), the taste and the aroma.

These are traits no longer ignorable and whose improvement is time to effect, insofar as, although the appearance of the apricot (big size, bright colour and overcolour) is important for the customer judgment, the intrinsic organoleptic characteristic are the most important for his/hers final verdict (Mellano et al., 2006). Tendentially the fruits perceived as qualitatively superior are the one with low pulp consistency, high juiciness, and good taste. This last parameter is linked with a high sugar contents (>12.0° Brix), an acid concentration between 190 and 240 meq/l and a sugar/acid ratio (this very factor used ad a common quality evaluating index) included between 0.5 and 0.7 (Bartolozzi et al., 1997).

Is within the ripening process that take place the important changes for the fruit chemical composition that will lead to modifications in colour, taste, texture, aroma and so to the sensation it will rise on the consumption moment (Almela, 2006). In the following table are listed the relations between the sensorial quality criteria, the chemical constituents involved and the changes that they go through during the ripening process.

Tab. 1.2. Relation among fruit quality parameters and its physical-chemical ripening changes.

Quality factors	Chemical constituents	Parameter variations
Colour	chlorophyll carotines xantophyllis flavonoids anthocyans	Modification of skin and pulp colour. Yellow-red coloration.
Taste	starch organic acids tannins carbohydrates proteins	Sweetness increase. Acidity lowering. Nutritional quality improving. Astringency decrease.
Aroma	aromatic compouns	Aroma and fragrance developing.
Texture	protopectines soluble pectines	Consistency decrease. Fruit softening.

1.2.7 Seed quality

If usually only the fleshy part of the fruits is taken in consideration when quality is defined, especially in stone fruits, is the seed that represents a huge, important portion of the fruit itself and the major by-product in fruit processing. In fact, currently large amounts of seeds are discarded yearly at processing apricot fruits. This not only wastes a potentially valuable resource but also aggravates an already serious disposal problem (Femenia et al., 1995). Moreover, apricot seeds could be a very interesting food, because they contain important levels of dietary protein (Nout et al., 1995) and also significant amounts of oil and fibre (Gòmez et al., 1998). But the presence of the cyanoglycoside amygdalin is a big constraint to their use for human or animal nutrition. So, development and cultivation of high quality cultivars with sweet kernel is a big object for the apricot production system, as far as it would decrease health hazards and in parallel increase the marketability of this by-product.

To satisfy this 'need for breed' two things are needed: knowing the heredity mechanism of the bitter/sweet character and developing some molecular markers linked with the sweet genotype. A crucial step to reach these goals is the exact biochemical phenotyping of the amygdalin content of the seeds. Only a very precise analysis, in fact, will provide the characterization of the variability among bitter and sweet phenotypes, thus allowing to broaden the knowledge of the putative genetic loci and the kind of heredity of the bitter/sweet phenotype.

1.3 Fine phenotyping

The most common way to measure the amygdalin presence in an extract is by HPLC. Frehner et al., (1990) used to extract the seed overnight in MeOH, than take the extract to dryness, resuspend it in water, centrifuge and filter it before injection in the system. Ressler et al., (2001) used a 10% solution of trichloroacetic acid (to prevent hydrolysation of CNGs) as extractant, defatted the sample 4 times with ether, stored everything o/n before centrifugation and concentration steps. Hwang et al., (2002) extracted the seeds in MeOH for 5-6 hours, took the extract to dryness, resuspended it in water and defatted with hexane 3 times prior to make HPLC. Lv et al., (2005) used MeOH or water and citric acid as extractant, carried on the extraction for 5 hours, then concentrated the sample by solid-phase-extraction. Joo et al., (2006) used to carry on the

extraction under reflux in water for 3 hours, filter the mixture and defat it with hexane and filter it again. All these HPLC protocols have some contraindications: they require a huge handling of the sample, that means a lot of time, risk of sample losing during the steps, higher risk to get false positives and to make mistakes. Furthermore HPLC requires an internal reference for each analyte.

These problems can be overcome by using the nuclear magnetic resonance (NMR). This analytical method relies directly on the physical intrinsic properties of the matter, being based on its magnetic features. The nuclei of the atoms with an odd protons' number, an odd neutrons' number or both things, rotate on themselves, therefore having a magnetic moment. This magnetic moment is characterized by a quantum spin number. Since the atomic nuclei possess a charge, a rotating nucleus is giving rise to a little electric current and has a little magnetic field associated. When a nucleus with a spin is put in a magnetic field it become, like the compass needle, subjected to a couple of forces that make it turn to be aligned with the external field. If these oriented nuclei are then irradiated with an electromagnetic radiation of appropriate frequency, the ones with low energy (aligned with the field direction) absorb an energy quantum, rotate their spin to gain the high energy state (opposite to the field direction) and then radiate this energy back out, so giving an observable signal. When these transitions of spin happens, it is said that the nuclei are in resonance with the radiation applied, from this the name NMR. The frequency of the electromagnetic radiation needed for the resonance depends both on the strength of the external magnetic field and on the examined nucleus features. Moreover, the NMR signal of a certain nucleus results shifted on the spectrum to higher or lower frequencies depending on its chemical environment, principally because of the shielding effect that the electronic density plays on the external field: the bigger is this shielding, the lower will be the active magnetic field that could act on the nuclei. The resulting shift in the NMR signal of a given nucleus (chemical shift, measured in ppm) is the measuring principle of the NMR spectroscopy: by understanding different chemical environments, the chemical shift can be used to obtain structural information about the molecules in a sample. For example, for the ^{1}H-NMR spectrum for ethanol (CH_3CH_2OH), one would expect three specific signals at three specific chemical shifts: one for the CH_3 group, one for the CH_2 group and one for the OH group. The chemical shift is reported as a relative measure from some reference resonance frequency, and for the nuclei ^{1}H, ^{13}C, and ^{29}Si, the tetramethylsilane is commonly used as a reference, because of the high shield of its carbons and protons. On

this scale the simple hydrocarbon protons tends to absorb in the 0.5-1.5 ppm region, the protons of the carbon near the carbonyl are shifted toward 2 ppm and the more electronegative atoms (oxygen and halogens) move the protons to 3-5 ppm, only to quote some values. In this way is possible to determine the region to which the complex compounds belong, just like the sugar region (3-4 ppm) or the aromatic one (6-9 ppm). The most commonly studied nuclei in biomolecules are ^{1}H (the most NMR-sensitive isotope after the radioactive ^{3}H) and ^{13}C. Nuclei from isotopes of many other elements (e.g. ^{2}H, ^{10}B, ^{11}B, ^{14}N, ^{15}N, ^{17}O, ^{19}F, ^{23}Na, ^{29}Si, ^{31}P, ^{35}Cl, ^{113}Cd, ^{129}Xe, ^{195}Pt) are studied by high-field NMR spectroscopy as well. One of the most important feature of the NMR absorbance is that its intensity is directly proportional to the number of nuclei that give rise to the signal, that is, the area below a certain peak (the integral) is dependent on the concentration of the molecular groups to which this peak belong. So an absolute concentration is computable by comparing the desired metabolite peak to a standard reference compound one.

In the lab practice the NMR is used to determine the content of various compounds in complex cellular extracts without having a previous knowledge of the samples (Moing et al., 2004). In fact this spectroscopic method allow to measure the soluble sugars, organic acids, aminoacids and other secondary metabolites in extracts from fruits, roots and leaves (Pereira et al., 2006). To quote some results: the group of A. Moing has used the ^{1}H-NMR to phenotype the fruit of a progeny of strawberry (so being able to identify some QTLs for the quality) and to compare the metabolic composition of wild type and transgenic tomato roots (revealing that the environmental factors can modify in a significant way the metabolic state of the plant, thus hiding or emphasizing the genetic background expression). Instead Pereira and co-workers (2006) have exploited this technique to distinguish wine samples on the basis of their geographical origin (because of the different amino acid content) or on the different vintage years (mostly thanks to the sugar and phenolic composition).

All that said explains clearly why the NMR is considered to be one of the most robust and precise analytical methods in research: it enables a unique and quantitative determination of the relative amount of molecular groups, so being a real powerful tool to quantify entire molecular structures even in mixtures. And it has also some crucial practical sample-handling advantages: in comparison to traditional sample preparation it requires a less initial amount (mg vs. g); it's time-saving, thanks to its ability to quantify a single compound in complex mixtures without requiring fractionation or isolation

procedures. This latter is a key advantage for the analysis of seed extracts, insofar as their fat content usually represents a problem for an HPLC assay, because of the risk of sticking the column. By using the NMR to analyze the amygdalin content of apricot seeds it is possible to simply not care about oil, because the signal of the amygdalin is in another region of the spectrum. And these particular benefits have to be added to the normal advantages of using such a technique for the phenotyping: the possibility of a simultaneous determination of more than one analyte in a mixture, its non-destructive character, the relatively short-measurement time.

1.3.1 Quantitative NMR

To successfully use NMR as a quantitative method, anyway, some particular experimental details have to be established. According to the reviews of Pauli et al. (2005 and 2007) these recommended "quantitative conditions" are:

- Acquiring data on a static sample. So it is avoided the risk to have fastidious sidebands that may overlap to the main signal and make the integration really difficult.
- Removing the carbon satellites. The ^{13}C isotope is present to the extent of 1.1% in the sample and it will couple to the proton resonances, producing satellites peaks that flank the principal proton resonance. These satellite resonances can overlap with other signals of the primary analyte or with signals arising from impurities present in the sample. Performing a heteronuclear decoupling (like the Globally-optimized Alternating-phase Rectangular Pulses (GARP) scheme) at the radio frequency of the C nuclei will make the satellite collapse.
- Use a relaxation delay adequate for both analytes and internal standard. This delay is inserted to allow the excited nuclei to re-establish their equilibrium *z*-magnetization after the acquisition of the FID information and prior to the application of the next pulse, in order to avoid distortion of integrated signal intensity due to relaxation effects. So the internal standard has to be chosen not only regarding its being available in a highly pure form, inexpensive, stabile and chemically inert, non-volatile and non-hygroscopic, soluble in all (or most) of the NMR solvents that are used routinely but also regarding its resonance window, that has to be close to the analyte's one

1.4 Raman spectroscopy

If some is clear about the sub-cellular localization of the cyanogenic glycosides in almond, plum and black cherry (these metabolites were detected in plum and black cherry seeds by tissue printing from Poulton and Li in 1994, and show to occur solely in the cotyledonary parenchyma while being absent from the procambium and endosperm, an opposite pattern respect of their catabolic enzymes), few is known about the global distribution of the amygdalin in the mature apricot seeds. A very suitable method for this kind of *in planta* measurements is the Raman spectroscopic imaging (Thygesen et al. 2003).

The Raman effect occurs when light interacts with the electron cloud and the bonds of a molecule. When a monochromatic radiation of frequency ν_0 is incident on a sample, some of the radiation is scattered. In the scattered radiation, in addition to a radiation with the same frequency as the incident radiation (elastically scattered radiation—Rayleigh radiation), a radiation of different frequencies (inelastically scattered radiation—Raman radiation) is also observed. The basic idea of the inelastic scattering may be described as follows: the interaction of the photon of energy $h\nu_0$ with the molecule may lead to the annihilation (virtual absorption) of the initial photon and to a rapid subsequent creation (life time of virtual state in the range of femto to pico seconds) of a new photon of energy $h(\nu_0 - \nu_M)$, accompanied by the transition of the molecule on which scattering occurs to a state with energy higher by $h\nu_M$ (usually an excited vibrational state). Instead, if the molecule is initially in the excited vibrational state, it will be observed a scattering that leads to the annihilation of the initial photon of energy $h\nu_0$ and creation of a new photon of energy $h(\nu_0 + \nu_M)$ accompanied by the transition of the molecule to a state with energy lower by $h\nu_M$ (Kudelski 2008). In Raman spectroscopy, the sample is radiated with monochromatic visible or near infrared light from a laser and, normally, the scattered photon has a lower frequency than the laser light (Stokes Raman scattering), and the difference in frequency (given in reciprocal centimetres) between the frequency of the laser and that of the scattered photon is called the Raman shift.

Analogously to spectra obtained with other vibrational techniques (e.g., infrared absorption, vibrationally resonant sum frequency generation, high resolution electron energy loss spectroscopy), Raman spectrum can be treated as a compound's fingerprint. This makes Raman spectroscopy a useful technique for the identification of many compounds. And since measured Raman signals are often characteristic for analyzed

compounds, Raman spectroscopy allows for analysis of very complex samples, including, for example, in vivo analysis of live tissues. Moreover, when the Raman spectrometer is combined with the microscope, the analysis is spatially refined, and by carrying out Raman measurements at various places of the sample (mapping the sample) it is possible to obtain detailed information regarding the distribution of specific compounds.

For the Raman microscopy of biological and medical specimens near-infrared (NIR) lasers are the light sources usually used, in order to reduce the risk of damaging the specimen by applying higher energy wavelengths. On the contrary, Raman microscopy of inorganic specimens, such as rocks, ceramics and polymers, can use a broader range of excitation wavelengths.

Of great practical importance for every kind of sample, is the non-destructive feature of the Raman methodology. For example, Reitzenstein et al., (2007) showed that on the basis of the measured Raman spectra it is possible to distinguish rapeseed (*Brassica napus*) seedlings from the classic line Drakkar from the new genetically modified line T-mix. And the seedlings were not at all affected from the Raman spectroscopic measurements, so that it was possible to cultivate each of them afterwards. It is also worth to note that water, which content in the biological samples is usually above 50%, is a weak Raman scatterer and usually does not interfere with Raman measurements (in contrary to the IR absorption measurements).

So, samples of microscopic size can be analysed directly without fixatives, markers or stains, in air, at ambient temperature and pressure, wet or dry, and in many cases without destroying the sample (Krafft et al., 2003). Moreover, it is possible to obtain both qualitative and quantitative information using the Raman microspectroscopy. A number of organic compounds and functional groups can be identified by their unique pattern of absorption, and the intensity of the absorption may be used for the calculation of the relative concentration in the sampled entity (Wetzel and LeVine, 1999).

Knowing that from the Raman spectrum is possible to identify the various molecules present in the sample, if thousands of Raman spectra are acquired from all over the field of view the data can then be used to generate images showing *in situ* the location and amount of these different components. Raman microspectroscopy may be combined with at least three different mapping techniques: point, line and area. With point acquisition several spectra are measured from different places in a sample selected after visual inspection through the microscope (i.e. the spectra are not systematically related to each other spatially). Line mapping defines a series of spectra to be obtained

along one dimension and can be used to investigate changes in a chemical component along a certain direction (i.e. a profile). An area map uses two dimensions, providing a spectroscopic image that can be directly compared to the corresponding visual image, but with an entire spectrum in each pixel instead of a simple colour. By using a series of partly overlapping acquisition areas it is possible to obtain a smoothing of a profile or a map. Using a standard microscope equipped with an XY stage it may take up to several hours to map a few square millimetres, even for an FT-based instrument. However, new instruments that record an entire row (i.e. a line scan) simultaneously by use of a CCD array will make area mapping much less time consuming (Thygesen et al. 2004).

The nitrile group(–C≡N) present in the amygdalin as well as in the other cyanogenic glycosides has special vibrational characteristics that are due to the relative rigidity of the triple bond. What is more, the nitrile vibration band can be considered a really highly specific distinctive sign for these metabolites, as this group is rarely found in natural compounds (Micklander et al., 2002). Its signal arises near 2240 cm^{-1} in the Raman spectrum (see Fig. 1.9), in an area which is almost free from interference from other chemical components. The aromatic ring found in the amygdalin molecule is also strongly Raman active, but the specificity is low, because aromatic compounds are ubiquitous in nature.

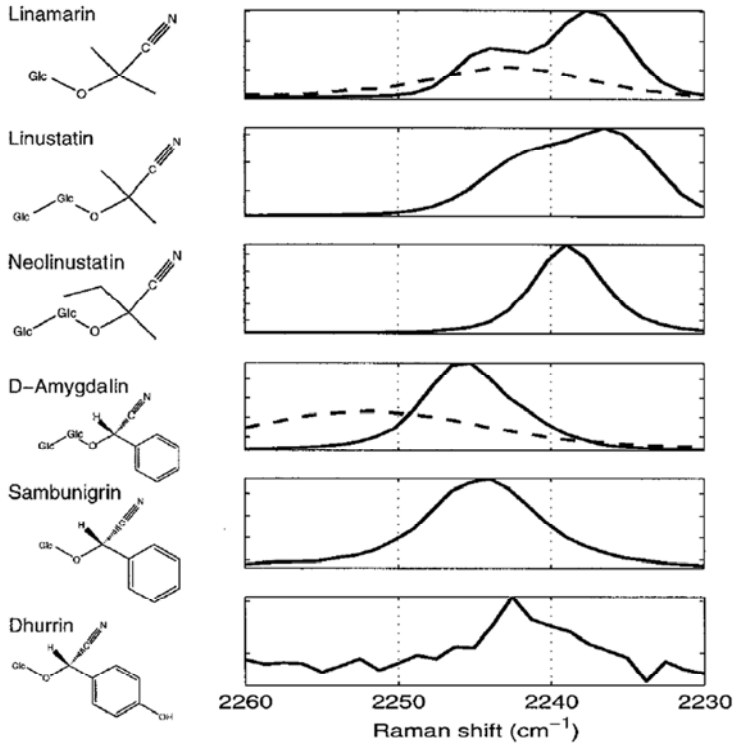


Fig. 1.9. Chemical structures of different cyanogenic glucosides and corresponding Raman signal from the nitrile group using crystalline samples. Spectra recorded from linamarin and amygdalin dissolved in water are shown with dashed lines (Thygesen et al., 2004).

So, thanks to the strong Raman activity of the amygdalin and to the possibility to use Raman spectroscopy as an imaging technique, it became possible, in theory, to investigate directly *in situ* the localization of this compound in the apricot seeds.

2. Aims of the work

Deepen the apricot genetic understanding is more than fundamental for the development of new breeding programmes aimed at the constitution of improved or new cultivars (Ruiz et al., 2010). In fact, these programmes have necessarily to be based on the knowledge of the map position of the single genes and/or of the QTLs and on the marker assisted selection. With a coupled biochemical and molecular approach, involving a candidate gene analysis and a fine phenotyiping *via* qNMR, this study aims to obtain information on the cyanogenetic mechanism in apricot, a species of great importance in the Italian agricultural system, but whose genetic data are still limited.

Thus in this work, using two apricot F1 populations, Lito×BO81604311 and Harcot×Reale di Imola, the following objectives were pursued:

- Identification and sequencing of some candidate genes involved in the cyanogenesis.

The availability of data about the metabolic pathways for the synthesis and the catabolism of the cyanogenic glucosides in almond (Sànchez-Pérez et al., 2008) was the starting point for this work. Databases of annotated sequences will be screened in order to align different *Prunus* genes to design heterologous primers. These will be used for fishing the apricot sequence of the selected candidate genes in a BAC library of the cultivar Lito. A primer walking strategy will be adopted with the aim to enlarge the identified sequences.

- Implementation of four existent maps with functional markers.

The candidate genes sequences will be screened to develop functional markers: SSR tandem repeats, indel as well as restriction sites polymorphisms will be identified and the relative SSR, SCAR and CAPS markers will then be tested on the two F1 populations available, looking for their usefulness in the parental maps saturation.

- Phenotyping of the amygdalin content of the seeds.

The seeds of the L×B population (collected over three years, from 2007 to 2009) and of the H×R one will be analyzed for their content in amygdalin. A first overview will be done using a colorimetric method on some sample groups, while a fine phenotyping of the entire

collection of seeds will be done *via* quantitative NMR. Distribution, environmental effect and ecological significance will be evaluated.

- Quantitative trait loci analysis.

The phenotypic data obtained by the qNMR experiments will be used to make an extensive QTL analysis on the two F1 populations. The presence or not of a correlation between the identified regions and the candidate genes position will be commented, as well as the co-localization with other regions involved in the bitter phenotype determination already identified in peach (Bliss et al., 2002) and almond (Dicenta et al., 2007).

- *In situ* localization of amygdalin.

Distribution of the amygdalin content will be determined by Raman spectroscopic analysis in seed slices, with the aim to identify where this cyanogenic glucoside is accumulated.

From the practical point of view, the work will begin from the molecular analysis linked to the candidate gene determination in the laboratory of the Fruit Tree and Woody Plant Sciences Department (University of Bologna, Italy). The biochemical phenotyping of the amygdalin, crucial pre-requisite for the QTL analysis, will be done in Jena (Germany) at the Max Planck Institute for Chemical Ecology, under the supervision of Prof. Bernd Schneider, leader of the Biosynthesis/NMR Group. The Raman imaging experiments will be possible thanks to a collaboration established with Dr. Krafft, Institute for Photonic Technology (Jena, Germany).

3. Materials and methods

3.1 Starting material

3.1.1 Apricot populations

The apricot plants used in this work belong to two F1 populations. The first one is made by 118 individuals coming from the cross Lito×BO81604311. Lito is a commercial variety derived from Stark Early-Orange×Tyrinthos and it's resistant to the D-strain of the plum pox virus (PPV, causal agent of the Sharka disease). BO81604311, the male parent, is a breeding line derived from the very fine varieties San Castrese and Reale d'Imola, it possess excellent qualitative characteristics but is susceptible to the PPV.

The second population is made by 99 plants, coming from the cross between Harcot and Reale d'Imola. The female parent is a precocious Canadian variety characterized by medium-big fruits with extended red blush and good fruit quality, even if very sensible to manipulations. The male line is a local variety of Emilia-Romagna, typified by big-sized fruits with light orange colour and 30% of red blush. Perfumed and aromatic, its sweet pulp is very suitable for the production of jams, cakes and juices.

As far as the seed features, the Lito×BO81604311 cross is a bitter×bitter one with all bitter offspring, while the Harcot×Reale one is a sweet×sweet cross with some bitter offspring.

3.1.2 Lito's BAC library

A genomic BAC library for the cultivar Lito has been made from the Lucigen Corporation for the University of Bologna together with the Universities of Udine and Milano. This library is organized in 30336 bacterial clones, everyone carrying a 130 Kbp insert of genomic DNA (see Fig. 3.1) such that the collection of cloned DNA molecules represents the entire genome of the source organism.

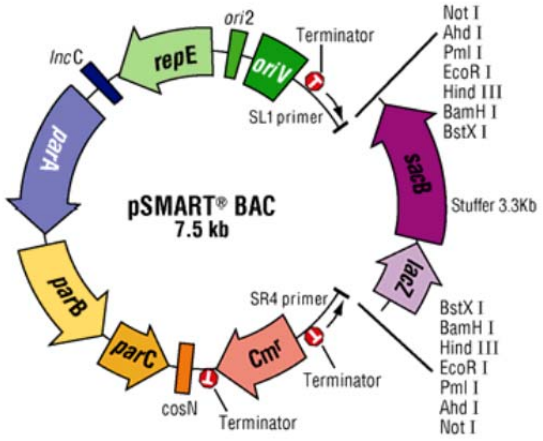


Fig. 3.1. Lucigen pSMART BAC vector: ori2, repE, IncC, origin of replication (single copy); oriV, inducible origin of replication; parA-B-C, partition genes; Cm^r, chloramphenicol resistance; cosN, lambda packaging signal; T, CloneSmart transcription terminators; sacB, sucrase gene; lacZ, alpha peptide portion of the beta galactosidase gene.

The 30336 unique clones are divided in 79 plates with 384 wells each. In order to make the screening procedures easier and cheaper, a series of pools has been made (see Fig. 3.2). In particular a 77 wells Plate Pool (PP), containing in each well all the clones of one original plate; a Column-Row Pool (CRP), whose row 1 is the pool of all the rows 1 of the 79 original plates, the row 2 the pool of all the rows 2 of the 79 plates and so on (the same being true for each column, so giving a plate with 24 columns and 16 rows); two Pin Point Pools (PPP1 and 2, each one made by 4 96-wells plates), with every A1 clones from the plates 1-40 in PPP1's A1, every A2 of the plates 1-40 in PPP1's A2 and so on, and the same for the plates 41-79 in PPP2.

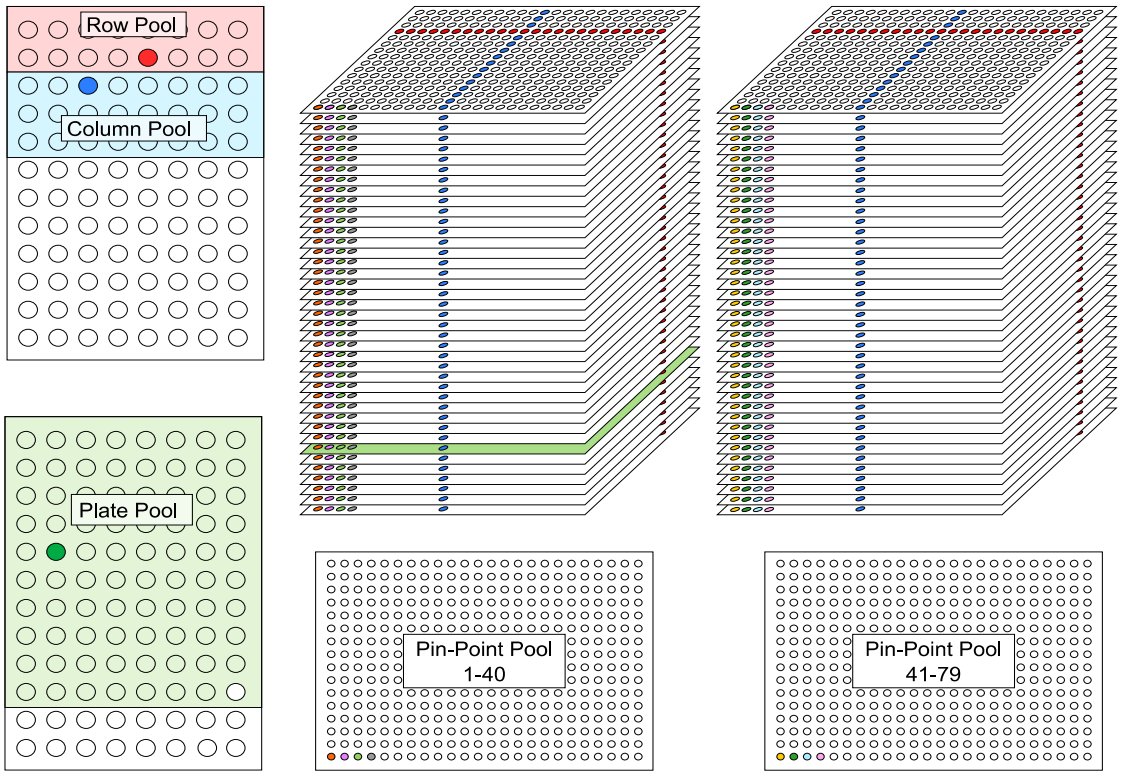


Fig. 3.2. Lito's BAC library pools organization.

3.2 DNA extraction and quantification

The DNA was extracted from young freeze-dried leaves, because from this kind of leaves the extraction is generally more efficient, thanks to the higher cellular density. Extracting DNA from apricot is not an easy task, because of the high amount of polysaccharides and polyphenols that diminish the efficiency of the enzymatic manipulations. For the extraction it was used a specific method to minimize the sugar/phenolic contamination elaborated by Mercado et al. (1999). From every freeze-dried leaf 0.05 g are weighed in a 2.0 ml Eppendorf tube and grounded in the "Mixer Mill" with some silicon carbide grains (carborundum) for 6 min at a frequency of 29 beats per second. The fine powder so obtained is resuspended in 1 ml of washing buffer (see Tab. 3.1) and subsequently centrifuged at 3000 rpm for 10 min. The supernatant is then discarded and the washing procedure repeated in order to get a pellet with a very low polysaccharidic content. The pellet is then resuspended again in 0.65 ml of washing buffer to which are added, as described in the table, 0.15 ml of NaCl 5 M, 0.1 ml of N-laurylsarkosine 10% (to improve the extraction efficiency) and 0.1 ml of CTAB 10% (the contemporary presence of two detergents is needed to maximize the yeld).

Tab. 3.1. Solutions for the DNA extraction according to Mercado et al., 1999	
Washing Buffer *(quantities given for a total of 100 ml of solution)*	
Sodium acetate 100 mM pH 5	0.82 g
EDTA 20 mM	4 ml
Sorbitol 200 mM	3.6444 g
PVP 40000 WT 2%	2 g
β-Mercaptoethanol 1%	1 ml
Extraction buffer - CTAB *(quantities given for each sample)*	
Washing buffer	0.65 ml
NaCl 5 M	0.15 ml
N-Laurylsarkosyne	0.1 ml
CTAB 10%	0.1 ml

After a short and gentle shaking the tubes are left in incubation at 60°C for 30 min. Then, to get rid of the carborundum, an equal volume of dichloromethane:isoamyl alcohol (24:1) is added. In this way a separation is obtained between an upper aqueous part and a lower organic one. This has to be shaken vigorously until an uniform emulsion is gained.

The tubes are then centrifuged for 5 min at 10000 rpm, forming again two phases: a lower one of a very emerald green colour that contains also the silica grains, and an upper one in which is the DNA, divided by a thin layer made by protein and cellular debris. The

supernatant is then transferred in a clean tube and 5 µl of RNase are added. The tube is then put in the oven for 30 min at 37°C. A second extraction step is then carried out. After a new recollection of the supernatant in a clean tube, 0.8 volumes of freeze cold isopropanol are added to precipitate the DNA. This step is let acting for about 20 min at -20°C. A subsequent centrifugation at 10000 rpm for 10 min and a quick washing with an ethanol 80% solution (to remove the buffer's salts) allow to obtain a whitish pellet that is then air dried and dissolved in 100-150 µl of sterile water. After an accurate resuspension the DNA is ready to be quantified. The concentration of DNA of every sample has been measured with a bench spectrophotometer (Nano-Drop Tech. ND-1000). Unlike the conventional spectrophotometers, in this instrument the cell is made by the sample drop, of a known volume (1 to 1.5 µl), that has to be measured. This allows to obtain in a very rapid way and using a very low amount of sample, a highly precise quantification of the DNA concentration (ng/µl).

3.3. Amplifications, electrophoresis and screening for markers

3.3.1 Primer design and PCR optimization

Looking to the metabolic pathways for the synthesis and the catabolism of the cyanogenic glucosides in almond (Sànchez-Pérez et al., 2008), the candidate genes chosen were the ones coding for: amygdalin hydrolase (AH), prunasin hydrolase (PH), mandelonitrile lyase (MDL), UDPG-glucosyl transferase (UGT). To amplify the desired apricot DNA traits, heterologous primer were designed on sequences already available in the databases and coming from both other *Prunus* species and other genus, like *Malus*. The annotated sequences found on the different databases were aligned with the ClustalW software and the primers were designed on the traits showing the highest homology in order to maximize the transferability to the apricot. Another software, Primer3, was used to correct the primers sequences according to various parameters: melting temperature (Tm), GC content, 3'stability and self-complementarity. Except for UDPG-glucosyl transferase primers, that were directly taken from an article by Franks et al. (2008), a total of ten primers were designed, based on annotated sequences from *Prunus serotina*, *Prunus amygdalus*, *Prunus dulcis* and *Malus×domestica* (Tab. 3.2). Two forward primers were done for the PH gene: PHC, designed on the sequence of the isoform PH I (PH-S1), and PHL, designed on the isoform PH

A (PH-L2). Also for the AH gene two forward primers were designed: AHC and AHL, this last conceived to amplify a longer fragment containing two introns.

Tab. 3.2. Sequence and main features of the designed and used primers.

Candidate gene		Primer sequence	Length	Tm°C
UDPG-glucosyl transferase	UGT	F: GAGAAAATGAGTCCAGTTGC	20	56
		R: AGAACAAGATGGAATTGAATCA	22	55
Prunasin hydrolase	PHC	F: CCATCTCATGGTCTAGATTGTTACC	25	60
		R: CTTTACTCGATCACCAAATTCCT	23	59
	PHL	F: AGGTTCTCTATCTCATGGTCCAG	23	59
		R: CTTTACTCGATCACCAAATTCCT	23	59
Mandelonitrile lyase	MDL	F: CATTGTGTTAGCGGCATGAA	20	61
		R: AGATAGAAGCCCTGAGGATGG	21	60
Amygdalin hydrolase	AHL	F: GGAACGATTACCAAATTTCACG	22	60.5
		R: TGAGGATCTGTTATGTAGCTTGC	23	58.5
	AHC	F: TGTGACAATCTATCATTGGGATCT	24	60
		R: TGAGGATCTGTTATGTAGCTTGC	23	58.5

Initially the polymerase chain reaction was carried on only on the parental genotypes Lito and BO81604311, in order to find the optimal condition for the primer functioning in apricot and starting from the standard reaction mix here described:

- Bi-distillated sterile water: 13.525 µl
- Reaction buffer 10X: 1.75 µl
- $MgCl_2$: 0.6 µl
- Nucleotides mix (dNTP 10mM): 0.175 µl
- Forward & reverse primer (10mM each): 0.175 µl each
- Taq polymerase (10 U/µl): 0.1 µl
- DNA template (50 ng/µl): 1 µl

(total volume of final amplicon so obtained = 17.5 µl)

The chosen buffer is without magnesium (an indispensable cofactor for the polymerase activity), to leave the possibility of adapting the master mix to the need of every couple of primers. Other variables that can be adjusted are: the primer concentration, the amount of Taq enzyme and the annealing temperature of the PCR machine.

The standard thermocycler steps are the following:

- 94°C for 3' (denaturation step to separate the double helix);
- 35 repetition of:
    - Annealing/gradient temperature (depending from the primers' features, see Tab 3.2) for 1';
    - 72°C for 1' (extension step);

- 94°C for 20'' (again to denature the double strands);

- Annealing / gradient temperature for 1' again;
- 72°C for 10' (final elongation);
- 10°C forever (for a short storage of the reaction).

Then the samples can be used immediately, kept for one day at 4°C or stored for longer periods at -20°C.

The optimal conditions for every pair of primers have been determined using a gradient thermocycler (Gradient Cycler PTC-200, MJ-research) and two different $MgCl_2$ concentrations for each temperature tested (see the figure below for an example).

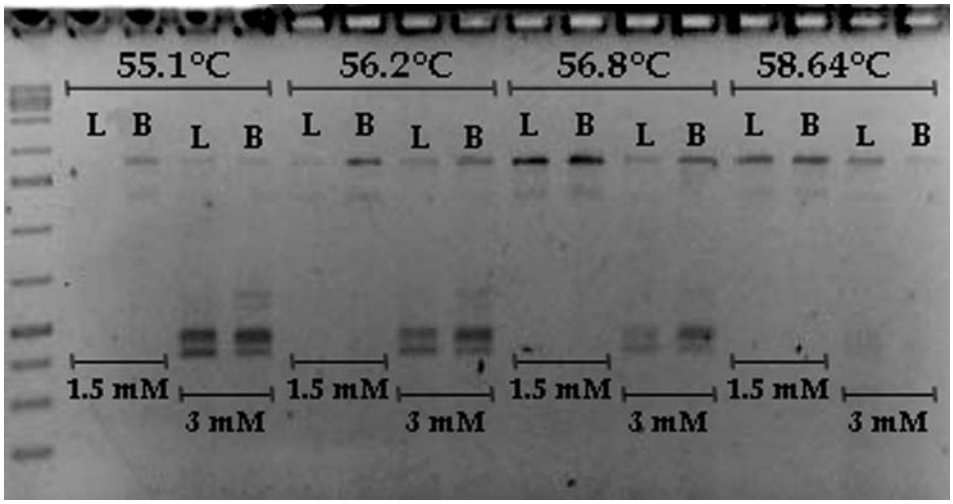


Fig. 3.3. Example of gradient test with variation of temperature and $MgCl_2$ concentration.

The parameters choice was done trying to find the best compromise between a high signal and the more stringent conditions possible (to avoid dirty amplicons or unspecifics). In the end of these preliminary tests the optimal setting for every candidate gene amplification were the ones listed in the following table:

Tab. 3.3. Optimal amplification conditions for every couple of primers

Candidate gene	Annealing temperature (°C)	[$MgCl_2$] (mM)
UGT	56	1.5
PHC	63	1.5
PHL	63	1.5
MDL	63	1.25
AHC	61	1.5
AHL	58	1.5

Afterwards, the parental DNA and of only six offspring was amplified with the optimized conditions. This preventive screening is used to test just on a limited number of individuals the potential presence of SCAR polymorphisms and, in the negative case, to test the possibility to develop CAPS markers by restriction enzyme trials. When polymorphisms are identified, the analysis can then be extended to the whole population.

3.3.2 Agarose gel electrophoresis

For all the agarose elecrophoresis the gel was done at a concentration of 1.5% in TAE buffer (see Tab. 3.4). For every sample were loaded 10 µl of PCR product mixed with 2 µl of loading dye (Blu juice) or, anyway, a quantity of loading dye around 1/6 of the amplicon's available quantity was used. The ladders employed were the Range ruler 20 bp, the Gene ruler 100 bp or the Gene ruler 1 kb Plus. Gels were run in TAE buffer 1X at a voltage not exceeding the 150 V (chamber model 80-6061-57, Amersham Biosciences (SF) Coop., San Francisco, California).

TAB 3.4. Solutions for the agarose gel electrophoresis	
GEL DI AGAROSIO 2%	
Water TAE 50X Agarose	196 ml 4 ml 3 g
TAE buffer 50X (1 liter)	
Tris Glacial acetic acid 1M EDTA 0.5 M pH8	242 g 57.1 ml 100 ml
LOADING DYE	
Water TAE 50X Ficoll 400 Bromophenol blue	9 ml 1 ml 1.25 g 12.5 mg
SOLUZIONE DI COLORAZIONE	
Ethidium bromide solution (1g/100ml) Water	35 µl 700 ml

Bands were visualized by ethidium bromide, an intercalating agent fluorescent when illuminated with UV light. To acquire the fluorescent image the "Image Station 440 CF" (Kodak Digital Science) was used. This instrument is connected with a software that allow to set the number and the frequency of the acquisitions, so giving a really sharp image, even if sometimes the use of a graphic program was needed to further improve the picture quality.

3.3.3 Polyacrylamide gel electrophoresis

To better visualize the restriction polymorphisms the sample can be loaded in a polyacrylamide gel. This gel is characterized by a higher separative capacity, thus allowing a more accurate analysis of the digestion results.

The glass sheets used have a dimension of 35×45 cm. The treatments done on those that will became the inner surfaces of the gel chamber were:

- Cleaning: with an ethanol solution, first 100% and then 40%.
- "Bind" sheet treatment: 1 ml of a solution containing ethanol 95% and acetic acid 5% it's mixed with 3 µl of Bind Silane[2]. This mixture has to be uniformly spread on the sheet's surface.
- "Repel" sheet treatment: on its surface is spread 1 ml of Sigma Cote[3]. This solution readily forms a covalent and microscopically thin film on the glass, making it really water repellent, and thus allowing a perfect detachment of the gel after the run.

(To avoid the risk of contaminations that will lead to breaks in the gel, these special treatments are never done contemporarily).

- Second cleaning: with a 40% ethanol solution.

Spacers of 0.3 mm are then placed between the glass sheet, so forming the gel chamber.

For the polyacrylamide gel preparation see Tab 3.5. It's very important to add the catalyst (TEMED) and the polymerizing agent (ammonium persulfate, APS) only at the last moment and in this order, to avoid a premature polymerization of the matrix. When the gap is completely filled with the gel, is time to insert the comb upside down, to build a uniform, straight front. The polymerization reaction is generally ended after a couple of hours.

The two chambers of the electrophoresis apparatus are filled with a pre-heated TBE buffer 1×. The comb is extracted and re-inserted with the teeth touching the gel, so creating the wells. Before loading the samples a 30 min pre-run is done with a 1:1 denaturant:buffer solution with the aim of uniforming the migration front and bringing the gel at a temperature of approximately 40°C.

[2] Methacryloxypropyltrimethoxysilane (Sigma Aldrich)
[3] a special silicone solution in heptane (Sigma Aldrich)

Then the denaturing agent (see Tab. 3.5) is added to the samples and the reaction is carried on in the thermocycler at 95°C for 5 min. After this the samples are put immediately on ice to avoid the renaturation. The wells have to be cleaned from the urea with a syringe, then 4-5 µl of sample can be loaded. The run is done at a constant power of 65 W (chambers models: S3S, Owl Separation Systems Inc., Portsmouth, NH; DASG-400-50, CBS Scientific Co., Dal Mar, California) and the timing is dependent both from the absolute length of the DNA fragments and from their relative dimension, in order to obtain a good spatial resolution without losing the smallest bands.

TAB 3.5. Solution for the polyacrylamide gel electrophoresis.	
5% polyacrylamide gel	
TBE- urea	54 ml
Liquabis (bis-acrylamide)	9 ml
Liquacryl (acrylamide)	9 ml
TEMED	45 µl
APS solution 10%	300 µl
TBE-UREA	
Urea	210 g
TBE 5X	100 ml
Bi-distilled water	170 ml
TBE 5X (1 liter)	
Tris	54 g
Boric acid	27.5 g
EDTA 0.5 M pH 8	20 ml
DENATURING LOADING AGENT	
Bromophenol blue	25 mg
Xylene cyanol	25 mg
Formamide 98%	24.5 ml
EDTA 0.5 M	0.5 ml

After the run the gel, that has to remain attached to the "Bind" sheet, is subjected to three treatments: fixing, coloration and developing (see Caetano-Anollés and Gresshof (1994) for a more detailed description).

For fixing, the gel is put in a basin with 2 litres of a 10% acetic acid solution (see Tab. 3.6) in slow agitation for 30 min. This has the aim of preventing the DNA diffusion and to eliminate the loading dye.

After this treatment the gel is moved to another basin and subjected to two washing steps in bi-distilled water of 5 min each to remove the fixing solution. This one is kept in a beaker, because it will be needed after.

Later than is the colouring treatment: the gel is put in a new basin with a silver nitrate solution (Tab. 3.6) and let in slow agitation for 30 min. The principle of this staining is that the silver ions are able, thanks to their positive charge, to bind the negative DNA molecules. Formaldehyde is added in the silver solution to improve sensitivity and contrast.

To develop the staining, an anhydrous sodium carbonate solution (always Tab. 3.6) is prepared to make the environment slightly basic and allow the thyosulfate to reduce the silver, so making the DNA bands visible. The developing solution has to have a temperature of 10°C or less, to slow down the reaction in order to keep the colouring under control. Just before the use the sodium thiosulfate and the formaldehyde are added.

The gel is quickly washed (10 s) to remove the excess of silver, and then immersed in the first basin with the developing solution. As soon as the first bands start to be visible, the gel is moved in the second staining basin, where the development continues until the coloration is adequate.

The staining reaction is stopped when is reached a compromise between the highest sharpness of the bands and the lowest background possible. To block the development is employed half a litre of the fixing solution previously used.

When the bubbling caused by the reaction is finished, the gel can be moved to a basin filled with water, where it will rest for around 15 min.

Then the gel is put, vertical, to dry overnight. The day after the images will be acquired with an Epson A3 scanner and the software "Photoshop 7.0".

TAB 3.6: Solutions for the polyacrylamide gel coloration (silver staining)

FIXING SOLUTION(2 litres)	
Acetic acid	200 ml
COLOURING SOLUTION (1 liter)	
Silver nitrate	1 g
Formaldehyde 37%	3 ml
DEVELOPING SOLUTION (2 litres)	
Sodium carbonate	60 g
Formaldehyde 37%	3 ml
Sodium thyosulfate	400 μl

3.4 Enzymatic digestions

Where no polymorphisms are detected, the amplicons are treated with a series of endonucleases, trying to find some differences in the cutting sites of these enzymes (CAPS markers). The following restriction enzymes were employed:

- *Rsa*I (Optimal cutting temperature: 37°C)
- *Taq*I (Optimal cutting temperature: 65°C)
- *Mse*I (Isoschizomer of Tru9I; optimal cutting temperature: 65°C)
- *Pst*I (Optimal cutting temperature: 37°C)
- *Xho*I (Optimal cutting temperature: 37°C)
- *Alu*I (Optimal cutting temperature: 37°C)
- *Bsp143*I (Optimal cutting temperature: 65°C)
- *Bsu*RI (Isoschizomer of Hae III; optimal cutting temperature: 65°C)
- *Sau3*AI (Optimal cutting temperature: 37°C)

The digestion mix for every sample is:

- Restriction enzyme 0.15 µl
- Buffer (specific for every endonuclease) 1.5 µl
- Water 3.85 µl
- Amplicon 10 µl

The reaction is left in incubation at the particular temperature for every enzyme for a minimum of one hour. Afterwards the fragments are visualized by agarose or polyacrylamide gel electrophoresis.

3.5 BAC screening

3.5.1 Positive clones picking

The specific primers for the candidate genes (Tab. 3.2) have been used to screen *via* PCR the Lito's BAC library. An example of the BAC screening method is explained in the following figure. The first step is the PCR of the column-row pools, the cross of which is then used to individuate the positive clones. The plate pool is then analyzed to see if it is possible to amplify only one Pin Point Pool (namely, if positives are found only in the first – plates 1-40 – or in the second – plates 41-79 – half). Afterward are amplified one or both of the PPPs, and the complete "address" (colony and plate) of the positive clones is so determined.

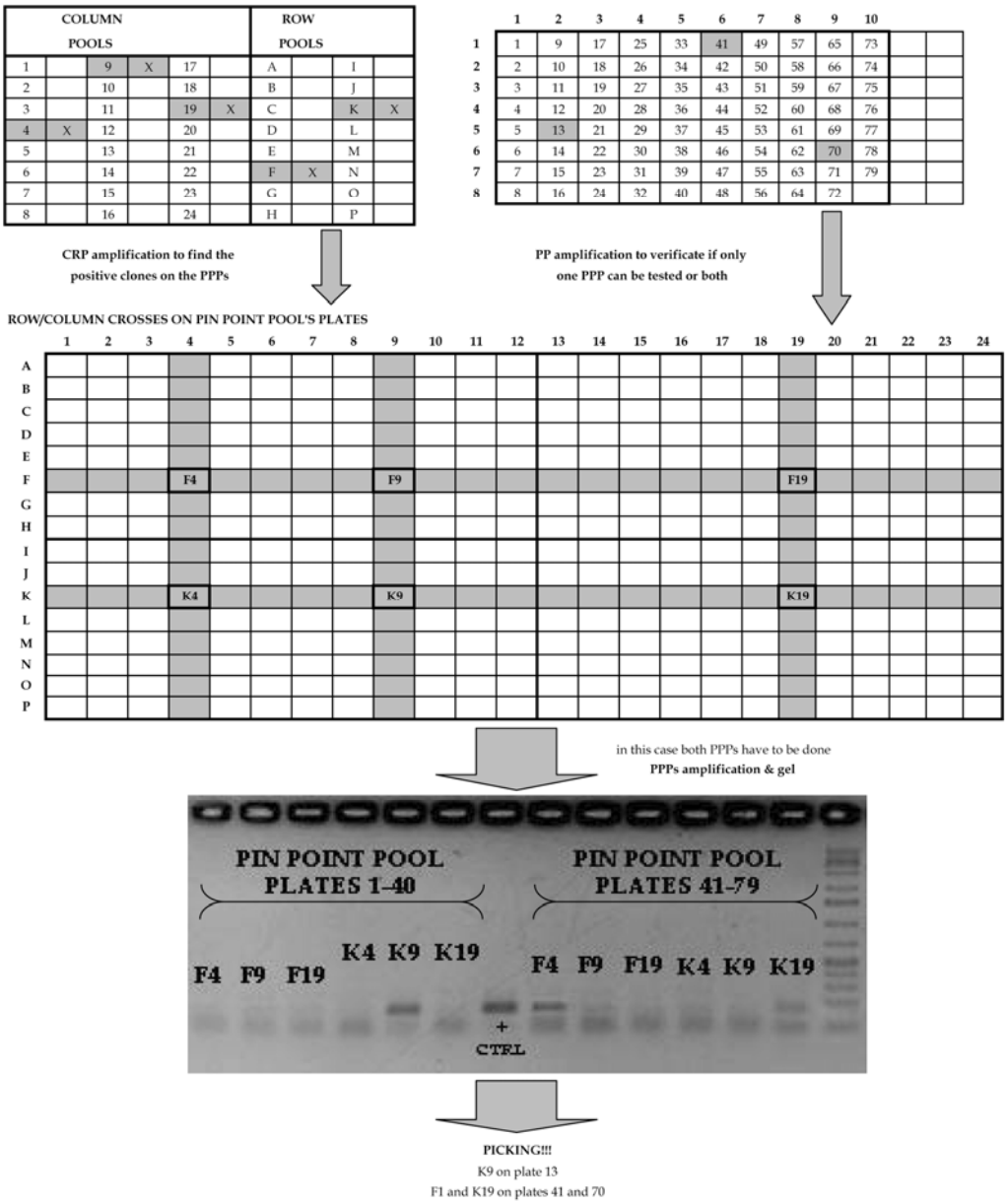


Fig 3.4. Scheme of the BAC screening through different pools.

3.5.2 Plasmid extraction

Once that the positive colonies are individuated, is possible to pick them from their plate of origin. The bacteria are put to grow in a Petri dish with LB-agar medium added with 12.5 µg/ml chloramphenicol. After a bacterial PCR control the plasmid is extracted from the positive clones, according to the following protocol (adapted from Bimboim and Doly, 1979):

- Put the bacteria to grow in 5 ml of LB medium (see Tab. 3.7) added with chloramphenicol (1 µl/ml); let them grow under vigorous shaking until confluence. Centrifuge for 5 min at 7000 rpm to pellet the cells and discard the supernatant.

Resuspend the pellet in 200 µl of LYR, transfer everything in a 2 ml Eppendorf tube and let incubate at room temperature for 5-10 min.

- Add 400 µl of Alkaline SDS, mix gently by inverting the tubes and incubate again 5 min at room temperature. Add 300 µl of KoAc 5 M pH 4.8 and mix well until complete flocculation. Incubate 30 min on ice.
- Centrifuge for 5 min at maximum speed and transfer the supernatant in a new tube. Repeat this at least once. Add 1 ml of icy-cold isopropanol and let the DNA precipitate for 5 min at room temperature.
- Centrifuge 10 min at maximum speed, discard the supernatant and wash the pellet with 1 ml of 80% ice-cold ethanol. Centrifuge again at max speed for 5 min, discard the supernatant and make the pellet dry in speed-vacuum. Resuspend well in 100 µl of water and quantify the amount and quality of the extracted plasmid with the Nano-Drop.

Tab. 3.7: Solutions for the plasmid extraction	
LB MEDIUM (1 litre)	
Bacto-Tryptone Bacto-Yeast extract NaCl	10 g 0.5 g 5 g
LYR	
glucose EDTA pH 8 Tris-HCl pH8 Lysozyme RNase A	10% w/v 10 mM 25 mM 2 mg/ml 2 mg/ml
ALKALINE SDS *(prepare immediately before the use)*	
NaOH SDS	0.1 M 1%
KOAc 5M pH 4.8	
KOAc Acetic acid Water	147.3 g 115 ml bring to 500 ml volume.

3.5.3 Primer walking and sequencing

The first sequencing was done on the cloned fragment, using the original primers designed for the BAC screening, from the BioFab Research Company. The chromatograms were viewed with the free software FinchTV (www.geospiza.com/**finchtv**/) and the sequence manipulated with ApE (http://biologylabs.utah.edu/jorgensen/wayned/ape/). A good sequencing reaction appears with high and well resolved peaks:

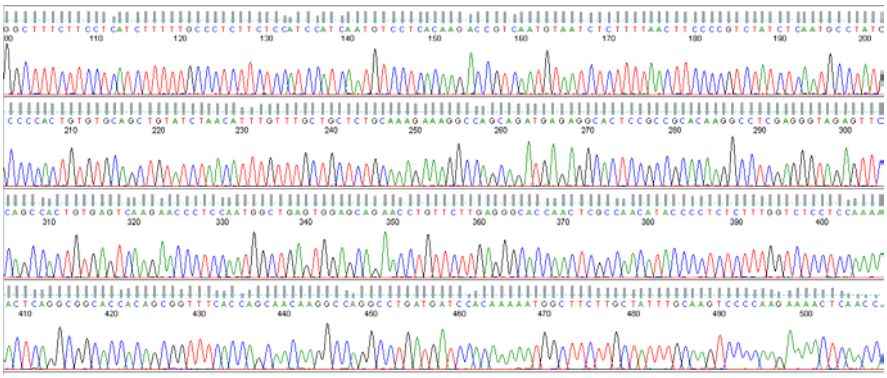


Fig. 3.5. Example of a well done sequencing reaction.

A sequence with overlapped or even absent peaks means that the reaction has not properly worked.

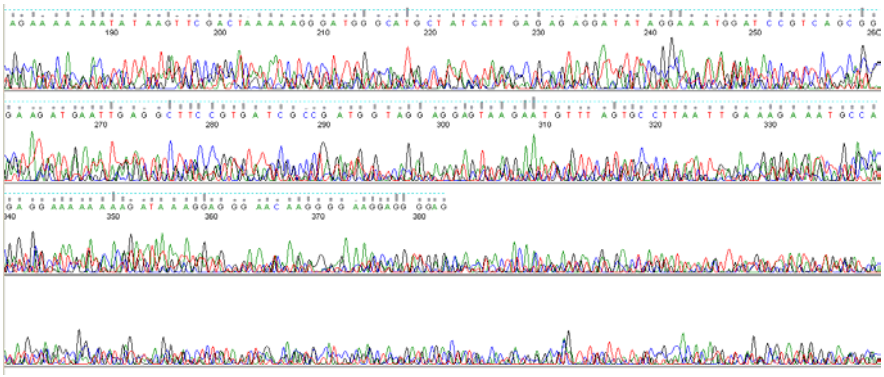


Fig. 3.6. Example of sequencing problems.

The factors that can lead to an inaccurate sequencing in our cases can be:

- Bacterial colonies not well singularized. In this case, as far as every colony could have different inserts, for every position more than one base can be found.
- Amplification reaction not correctly done, because of too low or too high amounts of template or because of a scarcely purified plasmid.

On the good sequences were then designed the primers to elongate them on both sides, called for every gene studied BAC1forward and BAC1reverse. On this second sequence another couple of elongation primers were prepared: BAC2for and BAC2rev. And so on until the extremity of the genes were reached (this was known thanks to the high synteny that the apricot genome has with the peach one).

All the obtained sequences were assembled manually and double-checked with the Lasergene SeqMan program. All the missing bases, the redoubling and the consensus conflicts were manually corrected. The work-in-progress sequences were step by step submitted to the BLAST (Basic Local Alignment Search Tool) software to control the

correspondence with the candidate gene sequence and the progresses of the primer walking. Since it has been published the whole peach genome (1st April 2010) it was possible to align the obtained sequence with it and to use it for a welcomed help for the primers' design.

3.6 Quantification and localization of CNGs

3.6.1 Colorimetric method

To quantify the total cyanide content of the seeds the colorimetric method developed by Masia and Cabrini (1994) was used, upon some steps of optimization, according to the specific starting material. The final protocol is the following:

- Weigh a piece of 10 mg of seed and grind it with liquid nitrogen.
- Add 270 µl of NaPi 500 mM buffer (pH 8.0, see Tab. 3.8) and mix very well.
- Add 30 µl of β-glucosidase (0.4 mg/ml) and leave overnight at room temperature for incubation (this will ensure a complete degradation of the CNGs).
- Prepare the blank solution (for the spectrophotometric measure): 270 µl of buffer and 30 µl of β-glucosidase.
- Prepare the 100 ppm standard by adding 900 µl of NaPi buffer 50 mM (pH 8.0) to 100 µl of CN-mother solution (Tab. 3.8)
- Add 300 µl of this standard to 700 µl of buffer to obtain the 30 ppm standard. From this stock are prepared all the other standards: 15 ppm, 7.5 ppm, 3.75 ppm and 0 ppm (only the buffer). 100 µl of this stocks solutions are there transferred to the "Standards" glass tubes and 2.5 ml of NaPi 50 mM buffer are added.
- Put in the "Samples" glass tubes 2.59 ml of NaPi 50 mM buffer and in the "Blanks" glass tubes only 2.5 ml.
- Centrifuge the Eppendorf at 5000 g for 5 min.
- Add to the glass tubes 10 µl of sample, being very careful in taking only the really limpid supernatant, without sucking the oil layer. Add to the Blanks 100 µl of the blank solution.

- Add to all 100 µl of Chloramine-T (0.12 mg/ml, this will lead to cyanogen chloride formation) and vortex.
- Only to the Samples and the Standards add 200 µl of 1,3-Dimethylbarbituric acid (0.24 mg/ml, glutaconic aldehyde is developed). Instead, add 200 µl of NaPi 50 mM buffer to the Blanks.
- Under fume hood: add to all 100 µl of pyridine and vortex (the chromogen compound is so finally producted).
- Let rest for 15 min. During this time the intensity of the colour will increase until a plateau is reached.
- Measure the intensity of the colours with a spectrophotometer, using the Blanks to make the zero.

Thanks to the measurement of the Standards is possible to obtain a calibration curve and to use its equation to quantify the total cyanide that was evolved from the β-glucosidase's degradation of every sample.

TAB 3.8: Solutions for the colorimetric method to quantify the total cyanide content (Masia and Cabrini, 1994)	
NaPi 500mM BUFFER (500 ml) *(dilute 1:10 to obtain the 50 mM buffer)*	
$Na_2HPO_4 \cdot 2H_2O$ H_3PO_4	44.49 g to bring the pH at 8.0
β-GLUCOSIDASE SOLUTION *(keep at 4°C for maximum two weeks)*	
Almond β-glucosidase NaPi buffer 50mM	2 mg 5 ml
CHLORAMINE-T *(keep at room temperature for maximum two weeks)*	
Chloramine-T NaPi buffer 50 mM	600 mg 5 ml
1,3-DIMETHYLBARBITURIC ACID SOLUTION *(keep at room temperature for maximum two weeks*	
1,3-Dimethylbarbituric acid NaPi buffer 50 mM	600 mg 10 ml
KOH 10 mM, pH 12	
KOH distilled H_2O	0.5611 g 1 l
KCN 1000 ppm – MOTHER SOLUTION *(validity: 2 months)*	
KCN KOH 10 mM, pH 12	125.13 mg 45 ml

3.6.2 Quantitative NMR

The colorimetric method was used as a preliminary analysis, because it is not so precise and it cannot discriminate between CNGs. To obtain a more reliable and accurate analysis a qNMR approach was used.

For the extraction, seeds were cut in fourth, according to Fig. 3.7, and every quarter was analyzed independently. The grinding apparatus (always Fig. 3.7) was cooled down with liquid nitrogen, the seeds were also frozen in nitrogen before being pulverized and then the powder was kept in ice-cold Eppendorf tubes until every quarter was grinded. The tubes were then carefully weighed before adding 1.5 ml of solvent (deuterated dimethyl sulfoxide, DMSO-d_6) and putting them to shake at 1400 rpm, room temperature for 5 hours.

The samples were then centrifuged for 15 min before being filtered with cartridges. The filtrate (600 µl) was placed in a 5 mm NMR tube and 10 µl of a phloroglucinol solution at 6% m/v were added as internal standard.

Fig. 3. 7. Seed cutting scheme and grinding apparatus

The NMR spectra were acquired on a Bruker Avance DRX-500 with BACS autosampler and direct-observe-probe (5 mm). Acquisition of spectra was carried out with TOPSPIN software (version 1.3). The ^{1}H NMR spectra were recorded with zgig30 pulse program. The spectral window was 12 ppm, and data were collected into 64k data points after 128 scans plus 2 dummy scans. The relaxation delay (d1) was set to 2 s. All experiments were carried out with an automatically determined receiver gain (around 100), automatic tuning and matching (ATM) with GRADSHIM tools and GARP decoupling scheme to collapse the ^{13}C satellites.

All the acquired spectra were calibrated on the residual DMSO solvent signal at 2.49 ppm. The phase of the spectra was manually corrected by selecting the submenu "Phase Correction", and the baseline was adjusted by creating a macro performing an automatic baseline correction by subtracting a polynomial function only in the spectral region

containing the amygdalin and the phloroglucinol peaks. The integration of signals was manually carried out and done independently twice.

The quantification of the amygdalin content of every sample was done according to the following formula (Malz and Jancke, 2005):

Quantification formula:

$$m_{Amy} = \frac{N°nuclei_{Phl} * n_{Phl} * Intval_{Amy} * M_{Amy}}{N°nuclei_{Amy} * Intval_{Phl}}$$

m = mass
n = number of moles
Intval = area of the peak
M = molecular mass

3.6.3 Raman imaging

The tissue specimens for the Raman analysis were collected using a cryo-microtome Leica CM1850. The apricot seeds were embedded in Jung tissue freezing medium (Leica Microsystems Nussloch GmbH, Nussloch, Germany) and freezed using the freezing shelf (it allows an extremely fast freezing down to –60 °C with the Peltier unit). the chamber temperature was set on –17°C and the orientations of the blade and the glass anti-roll guide were adjusted to obtain very flat slices of 15 µm each.

Raman images were collected in backscattering mode using a commercial microscopic Raman system (RXN1 microprobe, Kaiser Optical Systems, USA). The system consists of a multi-mode diode laser at 785 nm emission (Invictus), an f/1.8 spectrograph with a holographic transmissive grating (Kaiser) and a Peltier-cooled, back-illuminated, deep-depletion CCD detector (Andor). The microscope is coupled to the Raman system by fiber optics. Raman line measurements were performed approximately every 70 µm along vertical and horizontal lines across the seed slice, whereas the spectral maps were obtained by collecting spectra with a step size of 50 µm nearby the centre of the slice.

Data sets were imported into Matlab (The Mathworks, USA) for pre-processing. In-house written scripts were applied for background subtraction and removal of low intensity spectra. A vertex component analysis (VCA) was used to retain all relevant information by representing the image raw data in a space of smaller dimensionality. The aim of this method is that the endmembers (reference substances) represent spectra of most dissimilar chemical constituents. Then, scores denote the concentration of the endmember spectra. According to Krafft et al. 2011, the vertex component analysis procedure can be described as follows:

1. the VCA algorithm iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined;
2. the new endmember signature corresponds to the extreme of the projection;
3. the algorithm iterates until all endmembers are exhausted;
4. the dimensionality of the procedure is reduced to a few endmembers;
5. all other spectra are expressed as linear combinations of these endmembers.

3.7 QTL analysis

MapQTL 4.0 is a program developed for the calculus of the quantitative traits positions and of their significance on genetic linkage maps. This software, beyond the finding of the putative QTLs, permits to evaluate possible correlation with the functional markers present in the map.

The phenotypic data have to be expressed in a numerical way, so allowing to calculate the average effect due to each allele of each marker. To properly work the software need three files, all in .txt format. The first one contains the whole map dataset with the segregations of every marker, the second gives the genetic map information (marker order for each LG and distances in cM) for each marker of the previous file, the third contains the phenotypic data.

When these files are loaded, it's time to choose which test to use for the statistical analysis. For a preliminary overview the Kruskal-Wallis test is the best. The results of this function is the probability that every single marker is linked to the studied trait and is indicated by a number of asterisks, between one and seven, that means a significance (P) values of 0.10, 0.05, 0.01, 0.005, 0.001, 0.0005, and 0.0001 respectively. This test is separately made for every locus, one marker per time, and its main feature is that it works without any assumption, only on the real data, so it can be employed as a control and integration of every other approach.

To refine the analysis the "Interval Mapping" test can be used. It calculates the logarithm of odds of the presence of a QTL against its absence for every interval between two markers on the map. This value is then indicated by a LOD data, that represent the Log10 of the ratio of the two hypotheses (presence vs. absence). The final result is a probability map (QTL likelihood map): when there aren't enough real data (i.e. when the chosen interval to elaborate the map is 5 cM while in the real map the distances between some pairs of markers are wider) the program use an estimated value. In the output there are

also the percentage of variability explained by every founded locus and the graph of the QTL presence probability along the chromosome.

In the end, a permutation test has to be done, in order to estimate the LOD threshold over that a QTL is considered valid.

4. Results and discussion

4.1 Identification of apricot candidate genes involved in the CNGs metabolism

4.1.1 PCR optimization and first sequencing step

The first amplification of all the selected genes was done only on the accessions Lito and BO81604311. Every PCR protocol required an optimization step to find the best amplification conditions (annealing temperature and magnesium concentration), in order to avoid unspecific signals and to obtain, when possible, a single DNA band. Different Taq polymerases have also been tested, to be sure of the reproducibility of the results and of the quality of the sequences. The size of the PCR products obtained are listed in the following table:

Tab. 4.1. Length of the amplicons obtained for each candidate gene.

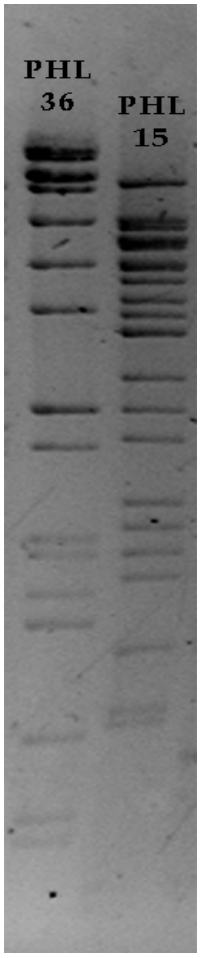
Candidate gene	Approximate amplified length (bp)
UGT	1500
PHC	600
PHL	Two bands: 500 and 700
MDL	300
AHC	150
AHL	1000

The first sequencing attempt was done on cut bands from a polyacrylamide gel. The sequence results were positive for all the samples. A BLAST analysis against the whole peach genome sequence (that has been made public the 1st of April 2010, http://www.phytozome.net/peach.php) confirmed the identity only of the longest sequences. In the meantime, to make easier the sequencing approach and avoid interferences due to presence of both alleles of each gene in the bands extracted from the polyacrylamide gels, a BAC screening was started and positive clones picked.

4.1.2 BAC screening

The first screening step of the Lito's BAC library for the amygdalin and the prunasin hydrolases (AHC/AHL and PHC/PHL respectively), the mandelonitrile lyase (MDL) and the UDPG-glucosyl transferase (UGT) genes resulted in a quite high number of putative positive clones. This was an expected result, considering both that the BAC used is a 10× and that the candidate genes chosen are all part of different multigenic families (MDL to the aldehyde lyases, AH and PH to the glycoside hydrolases and UGT to the glycosyltransferases).

The bacterial PCR of the identified colonies helped to reduce the numbers of the positive clones and in the end, by making the amplification condition as stringent as possible, only one clone per candidate gene was chosen for sequencing.



Concerning PHL, two clones kept giving a strong signal, so they were cut with *Eco*RI and, as far as the pattern resulted to be very different (Fig 4.1), they were both taken in consideration and named PHL15 and PHL36 from the identification number of their plate of origin. Such a high divergence in the digestion pattern of the two clones allow to suppose that they are not organised in a contig but belonging to two distinct chromosomes (or regions of the same chromosome).

Fig 4.1. Different cutting patterns of the two positive clones of PHL.The *Eco*RI digestion visualized on 1% agarose gel run at 70W for 5h.

4.1.3 Primer walking

The plasmids from the final number of positive clones so identified were sent for sequencing. On every obtained sequence a Forward and a Reverse primer were designed, to extend the sequencing. Every subsequent step was named progressively BAC1, BAC2 and so

on, according to a primer-walking approach. The sequencing results were not always of good quality, so some steps were repeated designing new BAC walking primers on different regions of the sequence.

UGT. The primer walking on the UGT gene gave very positive results on the "Rev" side, but no readable chromatograms were obtained when trying to extend the sequence to the other side. Modifying the assembling parameters to decrease the astringency allow to insert two BAC-For sequences, but takes out from the contig the UGT-Rev (see Fig. 4.2). This results in a slightly better result when blasting this second contig versus the nucleotide collection of the NCBI (organism selected: *Prunus* taxid:3754. Identity = 98% with 0% gaps, against an identity = 90% with 1% gaps) than when blasting it against the whole peach genome (Scaffold 1, identity: 88.8% against identity: 91.6%)

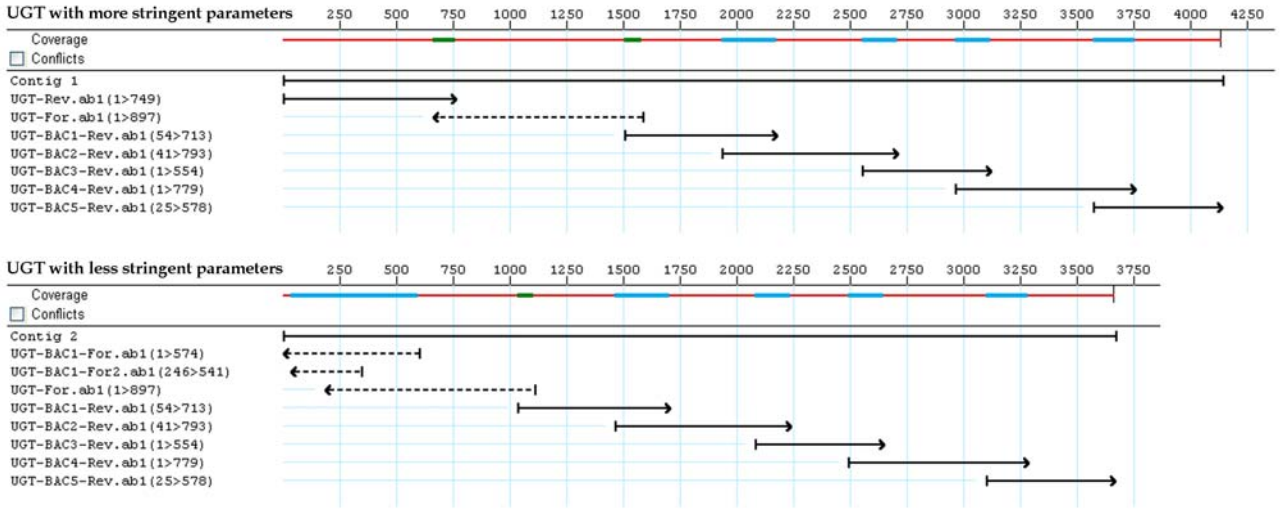


Fig 4.2. Contigs built for the UGT primer walking with the Lasergene software SeqMan. Differences in the assembling parameters bring differences in the final consensus sequence.

PHC. No BAC extension steps were done on PHC, because the first sequencings were almost not legible, and in any case no trustworthy enough to design primers on their basis. Lowering the parameters of SeqMan was not making any better, so a manual alignment was tried. Blasting this putative consensus in the peach genome resulted in a significant similarity in the scaffold 7, in a region where is annotated a 'β-glucosidase, lactase phlorizinhydrolase, and related' protein. The NCBI blast of this sequence gave an 82% identity value with a *Prunus serotina* prunasin hydrolase isoform PH B precursor (PH-L3; AF411928.1) gene (E-value: 0.0, coverage of 42%). Immediately after came: another 82% with the PH isoform PH C precursor (PH-L4;AF413213.1. E-value: 0.0, 41% coverage) and a 83% identity with the PH I precursor (PH-S1; AF414608.1, the original sequence on which the first

primers were designed. E-value always 0.0 but covering a little bit lesser: 34%). This demonstrates that a breach of trust in the technology is not good.

PHL15 and PHL36. The primer walking of the two PHL clones was going much more smoothly, with three forward and two reverse steps for PHL15 and exactly the reverse for PHL36 (Fig. 4.3). Both sequences gave positive homology results in the NCBI blast search (PHL15: 93% identity on 23% coverage with AF413214.1, isoform PH A precursor (PH-L2) gene; PHL36: 99% identity on 78% coverage with AF413213.1, isoform PH C precursor (PH-L4) gene) and when blasted against the whole peach genome (PHL15: scaffold 6 starting at 11001682 bp, identity: 85.8% with a base pair coverage of 2270/2646; PHL36: scaffold 6 starting at 3651691 bp, identity: 93.1% with a base pair coverage of 1946/2090).

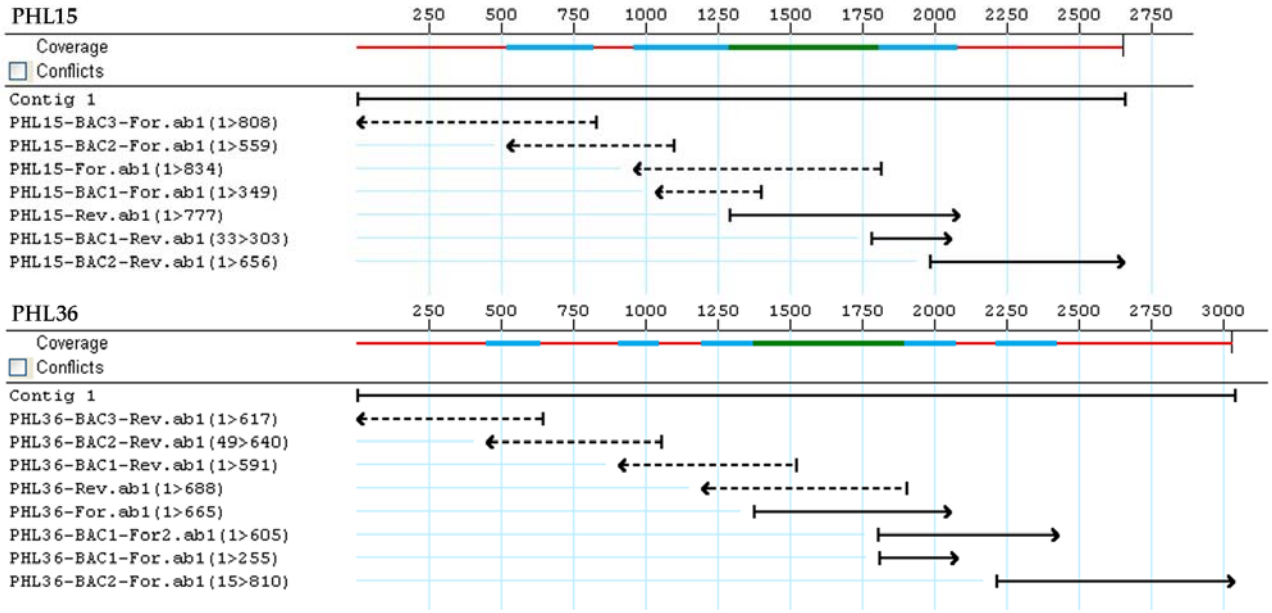


Fig 4.3. Consensus built for PHL15 and PHL36 with SeqMan (Lasergene). The two sets of sequences are too different to be assembled only in one contig.

Another proof of the difference between the two PHL clones is that it is impossible to build a single contig with SeqMen, but the sequences, no one excluded, are assembled always in two different contigs.

MDL. The only sequencing reactions working on MDL were the ones done on the parental genotypes Lito and BO81604311, sequences that were quite well in agreement one with the other (see Fig. 4.4).

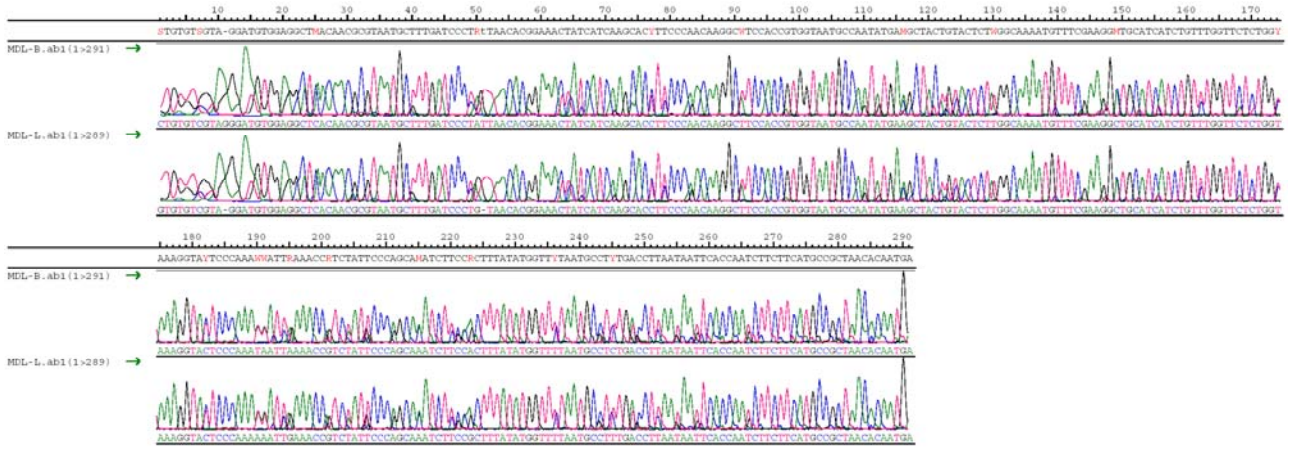


Fig 4.4. Alignment of the parental MDL sequences with SeqMan. Conflicts in the consensus were corrected looking the chromatograms quality or the expanded letter code (R, Y M, K, S) was used, to mark the flexible positions.

The consensus so built, trying to fix the conflicts arising from a bad chromatogram reading, gave optimum results in the NCBI blast firstly (91% of identity on a 94% query coverage with Y08211.1, *P. amygdalus* mRNA for mandelonitrile lyase) and then also in the peach genome search (two main homologies both in scaffold 1: identity: 86.6% (253/292 bp) starting at 6561336 bp and identity: 83.2% (233/280 bp) starting at 639718 bp).

AHC and AHL. Primer walking on both the AHs was very difficult, with numerous chromatograms showing poor or overlapping signals or even no peaks at all. Almost only the "forward" sequences were finally giving a nicely readable signal and were so inserted in a contig by the Lasergene SeqMan software (Fig 4.5).

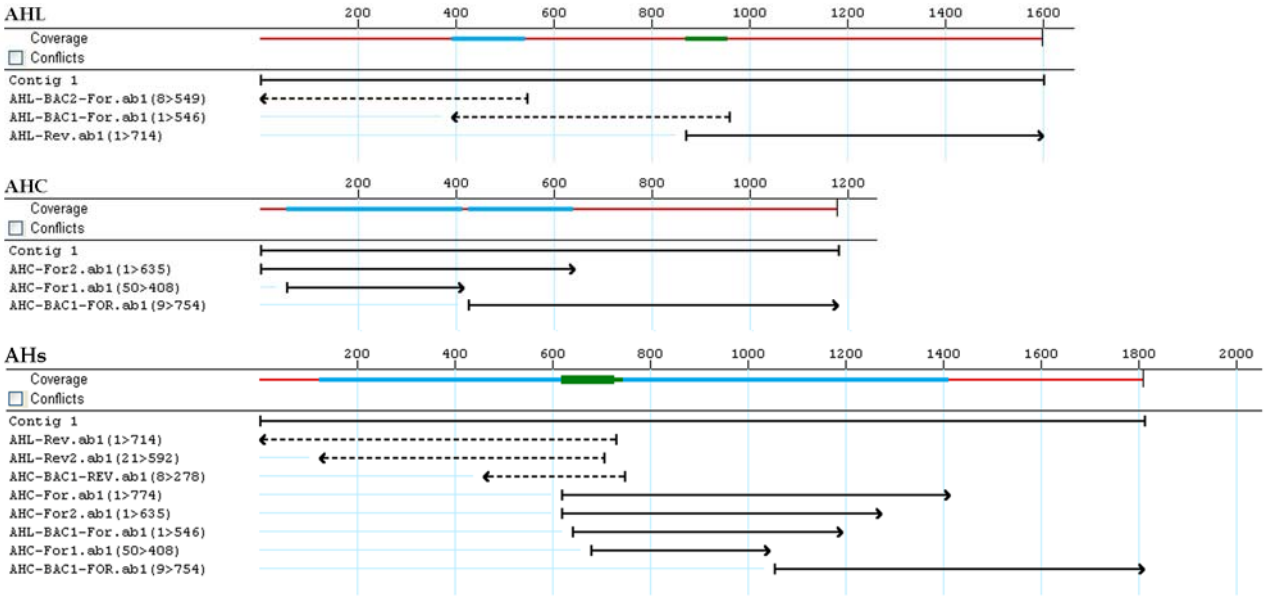


Fig 4.5. Assembling of the sequence sets for AHL, AHC and of all the AH sequences together (SeqMan software, Lasergene).

The final consensus get a bit longer when putting together all the AH sequences. Considering that AHC and AHL are only a shorter and a longer version of the same AH isoform, and were designed with this difference in length to increase the probability of finding a SCAR or a CAPS marker, the sequence deriving from the merging of the two contigs was taken as good for the BLAST analysis and the possible eventual annotation of a complete apricot AH sequence.

AHL gave a positive feedback when blasted in the NCBI nucleotide collection, showing an 80% identity with AF414606.1 (*Prunus serotina* amygdalin hydrolase isoform AH I precursor gene) on a coverage of 80%.

Also the alignment with the peach genome is quite good: 90% of identity on 1389 bp over 1533 in scaffold 7 (from 13891722 bp to 13893249). To confirm the being of AHC a shorter version of AHL, its blast result against the peach genome is found to be between 13892186 bp and 13892736, exactly inside the region covered by AHL.

This sequence extension work on the BAC plasmids has revealed itself to be more difficult than thought, very likely because of the high numbers of copies of these genes present in the genome.

This is demonstrated also from a brief survey of the identities found with the obtained CGs sequences and the peach genome. For example 15 MDL homologous are present in scaffold 1 only between 6535495 and 7206633 bp. That is why, owing also to some practical sequencing problems in obtaining good chromatograms (not quite well understood even after discussion with the BioFab technicians), the primer walking is still a work in progress.

For some candidate genes a complete sequence is almost reached, as summarized in Tab. 4.2, together with the best matches found on the peach genome and the relative annotated proteins.

Tab. 4.2. Summarizing tables of the primer walking results. The best matches with the peach genome are listed, comprising the homologous proteins and the bp still missing to get the whole apricot sequence for the protein.

Candidate Gene	scaffold	Blast values (identity)	Homologous protein	Homologous protein blast	Actual size and missing bp
UGT	1	91.6% (1059/1156 bp)	ppa004968m (UDP-glucuronyl transferase, UDP-glucosyl transferase)	*Prunus dulcis* mandelonitrile glucosyltransferase UGT85A19 mRNA.	3458 bp, missing 266 bp reverse.
PHC	7	81.6% (169/207 bp)	ppa018933m (β-glucosidase, lactase phlorizinhydrolase, and related proteins)	*Prunus serotina* prunasin hydrolase isoform PH B precursor.	376 bp, missing 1865 bp forward and 949 reverse.
PHL15	6	85.8% (2270/2646 bp)	ppa020368m (β-glucosidase, lactase phlorizinhydrolase, and related proteins)	*Prunus serotina* prunasin hydrolase isoform PH I precursor	3082 bp, missing 248 bp forward.
PHL36	6	93.1% (1946/2090 bp)	ppa021137m (β-glucosidase, lactase phlorizinhydrolase, and related proteins)	*Prunus serotina* prunasin hydrolase isoform PH C precursor	3028 bp, missing 755 bp forward.
MDL	1	88.5% (246/278 bp)	ppa003414m (Glucose dehydrogenase/ choline dehydrogenase/ mandelonitrile lyase: GMC oxidoreductase family)	*Prunus serotina* (R)-(+)-mandelonitrile lyase isoform MDL2 precursor	289 bp, missing 296 bp forward and 1526 reverse.
AHC	7	89.2% (495/555 bp)	ppa022513m (β-glucosidase, lactase phlorizinhydrolase, and related proteins)	*Prunus serotina* amygdalin hydrolase isoform AH I precursor	1179 bp, missing 153 bp forward and 341 reverse.
AHL	7	90.6% (1389/1533 bp)	ppa022513m (β-glucosidase, lactase phlorizinhydrolase, and related proteins)	*Prunus serotina* amygdalin hydrolase isoform AH I precursor	2517 bp, missing 1121 bp reverse

4.2 Implementation of the L×B and H×R maps

4.2.1 Functional markers development

All the primer developed for BAC screening were tested on the L×B and H×R progenies for their mapping. Being these markers not polymorphic if used as SCARs, they were tested as CAPSs: MDL cut with *Tru1*I and PHL cut with *Tru1*I or *Rsa*I were polymorphic in the L×B cross, whereas in H×R were scored AHL cut with *Rsa*I and PHC cut with *Rsa*I or *Tru1*I. In particular, the digestion of PHL and PHC with *Tru1*I allowed us to identify two different loci, named PHL/Tru1 and PHL/Tru2 in L×B and PHC/Tru1 and PHC/Tru2 in H×R. In this latter population also the cut of PHC with *Rsa*I was marking two loci, so named PHC/Rsa1 and PHC/Rsa2.

When SSR domains were identified on the available sequences, new primers were designed (Tab 4.3). Two of them resulted polymorphic in the L×B maps (AHssr and MDLssr) and three in the H×R (UGTssr, PHLssr and AHssr).

Tab. 4.3. Primer pairs developed on microsatellite repetitions found in the apricot CG sequences.

microsatellite	repetition	Primer sequence	T_m°C
UGTssr	$(CAC)_5$	F: CCCACTCATTCGAATTTTGC	60.4
		R: GTGACGACGGTGACTTTCAG	59.3
PHLssr	$(GA)_{28}$	F: ATGCTCGAACCCCAATAACA	60.3
		R: AATTTGGGATTGGGCTGATT	60.5
MDLssr	$(TA)_9$	F: GGTTCACTTCGATAAAATGAGCA	60.4
		R: GGGAGAGAAGCCACTCTGATT	59.8
AHssr	$(TA)_{17}$	F: GGCGTGAATATGGATGGAAT	59.6
		R: GCTTGAGGAAGATCCCAATG	59.6

4.2.2 Markers mapping in L×B

The genetic map of Lito, upgraded in a recent work (Dondini *et al.*, 2010), is long 532 cM and it has 161 markers at a mean separation of 3.3 cM. The map of the male parent BO81604311 (Dondini et al., 2007) covers 601 cM with 168 markers (average distance: 3.6 cM).

Two functional markers were developed for the MDL gene, a microsatellite and a CAPS one, being polymorphic, respectively, in Lito or in BO81604311 only. Both markers map in the very beginning of the LG1, so confirming the high sinteny between the two cultivars. The positioning of a gene encoding for a mandelonitrile lyase at the top of LG1 is coherent with the study of Joobeur et al. (1998) and is confirmed also by the BLAST analysis again the whole peach genome sequence (http://www.phytozome.net/search.php?show=blast).

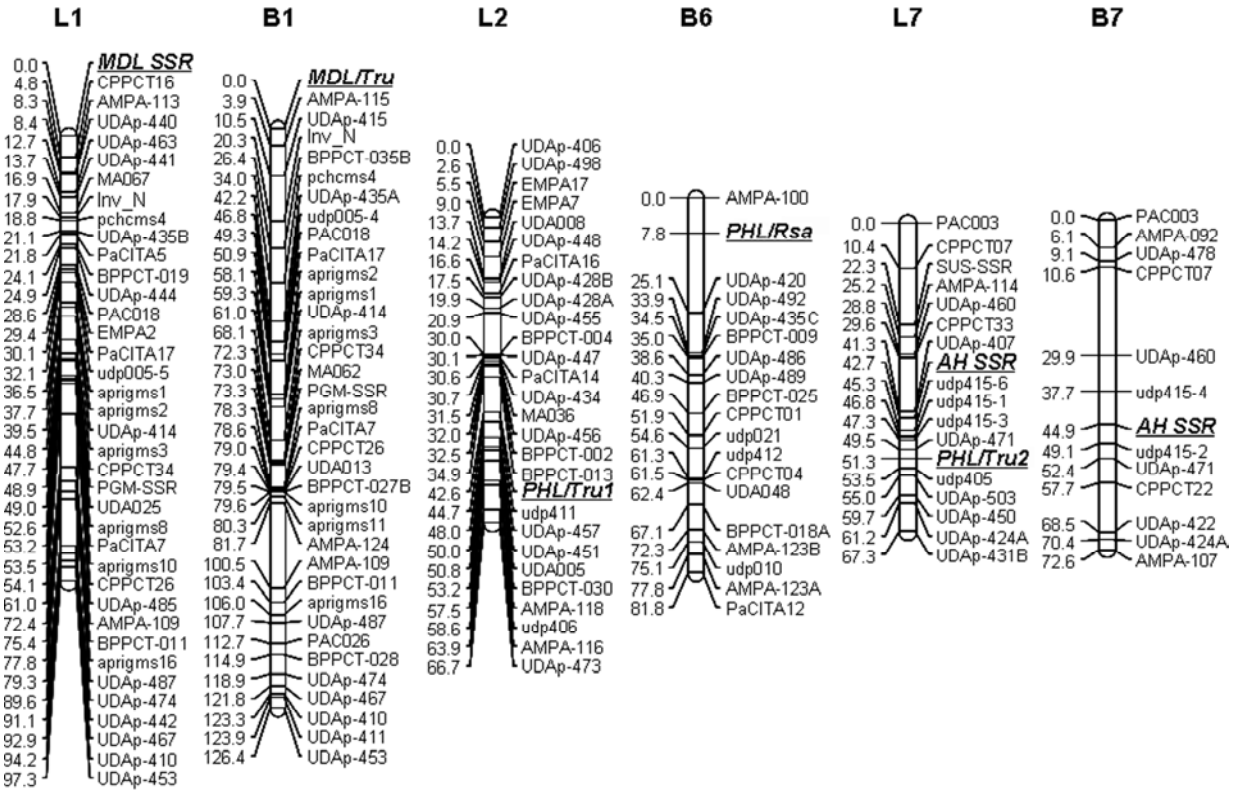


Fig 4.6 Candidate genes mapped on the parental lines Lito and BO81604311. The SSR and the CAPS markers related to the genes involved in amygdalin metabolism are highlighted in bold.

The AHssr marker showed a co-dominant segregation, and it was mapped on both parents in the LG 7 (Fig. 4.6). An amygdalin hydrolase (isoform I) results in fact localized in the scaffold 7 of the peach genome, at around 13,8 millions bp; as expected it is close to the flanking Lito's markers UDAp-407 and UDAp-471, whereas in BO81604311 is located between the markers UDAp-460 and UDAp-471. In a recent work Sánchez-Pèrez et al. (2010) localized an amygdalin hydrolase isoform I precursor in the LG1 of almond. This data was confirmed also by our BLAST analysis, which found an isoform of this gene in the scaffold 1, but with a lesser homology degree with the gene identified in the present study (70,1% over 970 bp against 90,6% over 1533 bp). Amygdalin hydrolase is a member of the β-glucosidase family, which in turn belongs to the family 1 of the glycosyl hydrolases families. Glycosyl hydrolases are a widespread group of enzymes which cleave the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (Xu et al., 2004). Zhou et al. (2002) have described four isoforms of amygdalin hydrolase in black cherry, with a 37-65% identity with the other members of the Family 1. Thus it is easy to hypothesise that the size of this enzyme class and the number of the AH sequences could be similar in the apricot genome.

A total of three CAPS markers were developed for the PHL isoform, each one showing a dominant segregation and, by consequence, mappable in Lito or BO81604311 only. Two loci were positioned in the LG2 and the LG7 of Lito. Their positions are confirmed by BLAST analysis on the peach genome both in scaffold 2 with the flanking marker BPPCT-013 and UDAp-457 and in scaffold 7 between the SSRs UDAp-471 and UDAp-503. The third locus (PHL/Rsa) was located near the top of the LG 6 of BO81604311, between the SSRs AMPA-100 and UDAp-420, a region in which there was a gap. Also in this case the colinearity with the peach genome was confirmed. Sánchez-Pèrez et al. (2010) mapped a PH candidate gene in the bottom of LG 1 of almond. Five isoform of prunasin hydrolases have been reported in black cherry so far (Zhou et al., 2002), being combined in a multigenic family characterized by an inner degree of homology of 71-93%, a sequence identity of 67-71% with the amygdalin hydrolases family and a degree of similarity of 49-92% with the other members of the glycoside hydrolase family 1. Therefore, is very likely the presence of more than one isoform of the PH gene also in almond and apricot, and the founding of them spread on the whole genome, as well as it is found to be in Arabidopsis (Xu et al., 2004).

4.2.3 Markers mapping in H×R

Harcot's genetic map has recently been upgraded within a work on the Sharka disease resistance (Dondini et al., 2010). Especially its linkage group 1 has been saturated, oriented and aligned to that of Lito. The whole map now is 525 cM with 98 markers. The map of Reale di Imola is less saturated, due to the very low level of polymorphisms found in this cultivar. Thus, some linkage groups are still divided in an upper and a lower part, being not polymorphic the markers that in other cultivars allow to unify the two frames, and still it has not been found a sufficient number of new markers to fill the gaps. All that is likely due to the low genetic diversity between the two parental lines, as reported in Geuna et al., 2003. To date the map of Reale di Imola is long 525 cM and has 62 markers.

As in L×B, also in H×R the microsatellite marker developed for the AH gene shows a co-dominant segregation and map on both parents in the LG 7 (Fig. 4.7), when in the peach scaffold 7 is an amygdalin hydrolase at 13,8 millions bp. The flanking markers are: for Harcot CPPCT33, 16.7 millions bp, and udp415, 12,9 millions bp; for Reale di Imola CPPCT7, 21 millions bp, and again udp415, as it's true in Lito and BO81604311. Also a CAPS marker was developed for this gene by cutting AHL with *Rsa*I, and was found to be in R6A, the upper part of the linkage group 6 of Reale. The blast results confirm this data, being the scaffold 6

the second region identified in the homology ranking (90.6% identity of AH with scaffold 6 (1389/1533 bp) and 70.1% identity with scaffold 6 (1007/1436 bp).

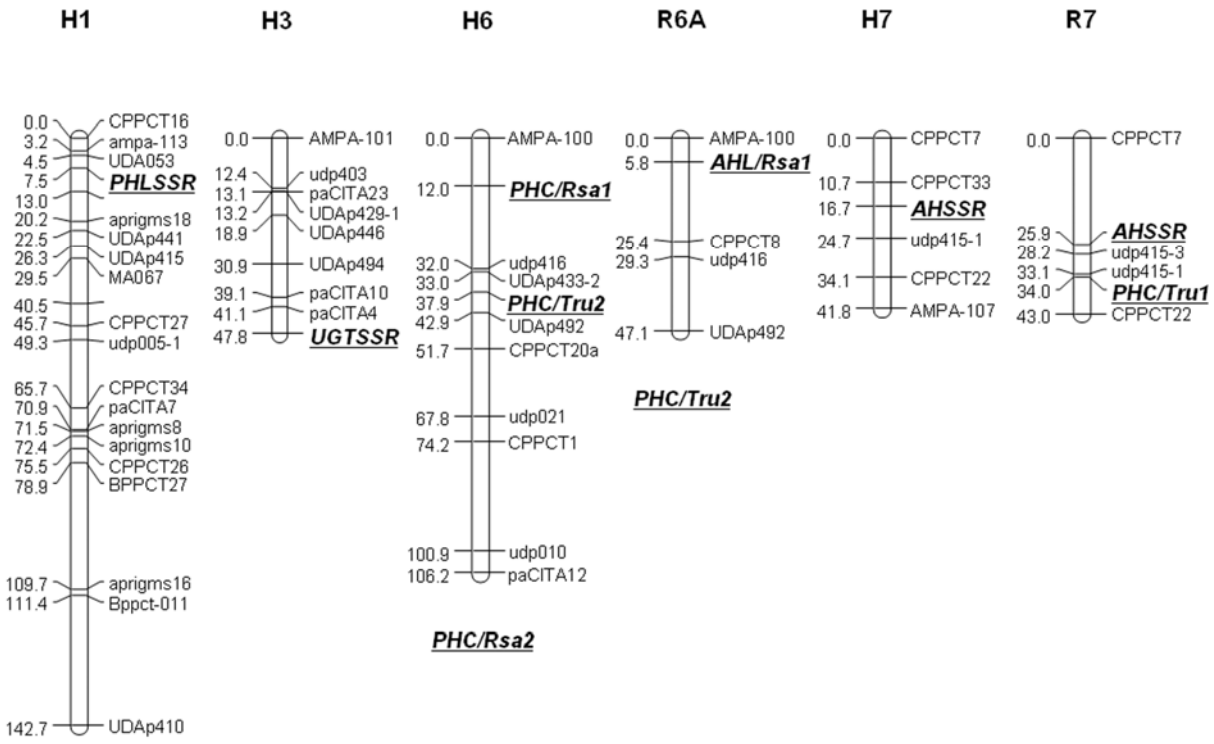


Fig 4.7. Candidate genes mapped on the parental lines Harcot and Reale. The SSR and the CAPS markers related to the genes involved in amygdalin metabolism are highlighted in bold. CAPS assigned to a LG but discarded during the map assembly are reported underneath the relative LG.

Four functional markers were developed for the prunasin hydrolase: an SSR and three CAPS resulting from the cut of PHC with *Rsa*I or *Tru1*I, this latter identifying two loci. PHssr was mapped in the upper part of Harcot's linkage group 1. Blasting the SSR and the near markers sequences against the whole peach genome this positioning is confirmed, being AMPA-113 at 0.7 millions bp, PHssr at 1 million bp and aprigms18 at 4.4 millions bp. Sánchez-Pérez et al. (2010) have also mapped a PH candidate gene (for the isoform PH C precursos) in the of LG 1 of almond, but in its bottom region. This difference in the map positions can be explained when considering that in the region of the peach genome where PHssr is aligned is annotated a protein (ppa003891) whose BLAST reported a high degree of homology with another isoform: PH B precursor. As far as the CAPS, PHC/Rsa1 and PHC/Tru2 map in H6, in the upper part near AMPA-100 the first, and around 30 cM further down the second. A CAPS marker derived from a cut with *Rsa*I was found in the same position also in BO81604311. The other locus identified by the cut with *Tru1*I, PHC/Tru1, was also assigned to H6 (Fig. 4.7), but then discarded when building the map because some

inconsistencies in the segregation made impossible the assignment of a phase by the program (the same problem was found also with PHC/Tru2 and R6A). In any case, the first locus of the CAPS found with *Tru1*I shows a co-dominant segregation, making so possible to map it on the other parental, in R7. This result is consistent with the peach sequence, where a PH is found in the scaffold 7, at 11.2 millions bp. The flanking markers, udp415 and CPPCT22, are located at 12.9 and 10 millions bp, respectively.

No functional markers were polymorphic in this population for the MDL gene, but it was possible to map the SSR designed for the UDPG-glucosyl transferase, UGTssr, in the bottom of H3. In the lower-middle part of the linkage group 3 of almond Sánchez-Pérez et al. (2010) have mapped a SNP marker for the glucosyl transferase. The positioning of UGTssr could, in fact, be not really accurate, because on the peach sequence UGTssr seems to be at 13 millions bp, whereas its near markers paCITA4, paCITA10 and, a little bit distant, UDAp 494 are located, respectively, at 14.8, 14.1 and 7.8 millions bp. This convey the idea of an inversion, where UGTssr may be situated between paCITA10 and UDAp494.

4.3 Amygdalin phenotyping

4.3.1 Colorimetric method

The first quantification attempt of the amygdalin in the seeds was done using a colorimetric method elaborated from Masia and Cabrini (1994). The protocol implies a complete degradation of the CNGs present in the sample and their development in a chromogenic compound by making the hydrogen cyanide reacting with the Chloramine-T, the so formed cyanogen chloride reacting with 1,3-barbituric acid to form glutaconic aldehyde that, by the action of the pyridine, finally results a purple-coloured solution (Fig.. 4.8).

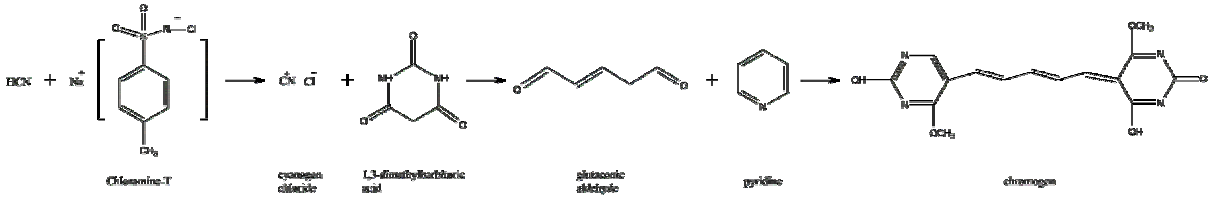


Fig 4. 8. Steps of the colorimethric method, starting from the hydrogen cyanide developed by the β-glucosidase degradation of the CNGs to the chromogen compound.

The intensity of the colour, directly proportional to the initial amount of CNGs, is measured by spectrophotometer using a calibration curve previously done with standard samples (see Fig. 4.9).

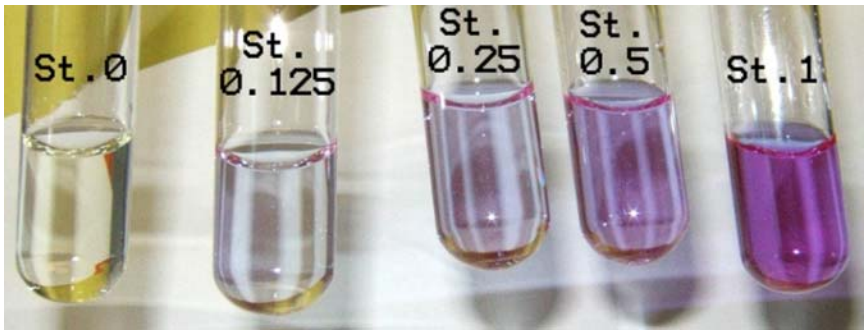


Fig 4.9. Example of the standard solutions prepared for a calibration curve (concentrations are given in ppm).

The protocol was adjusted using the L×B seeds already available, in order to determine the method sensitivity, the standard curve calibration, the initial seed quantity necessary and the variability between the seeds of the same individual. Then the developed method was tried on the seed of the H×R individuals that were been more productive (for some plants we got only one or two seeds), in order to get a first glance without using all the material. In the Fig 4.10. are reported some results obtained from this preliminary analyses.

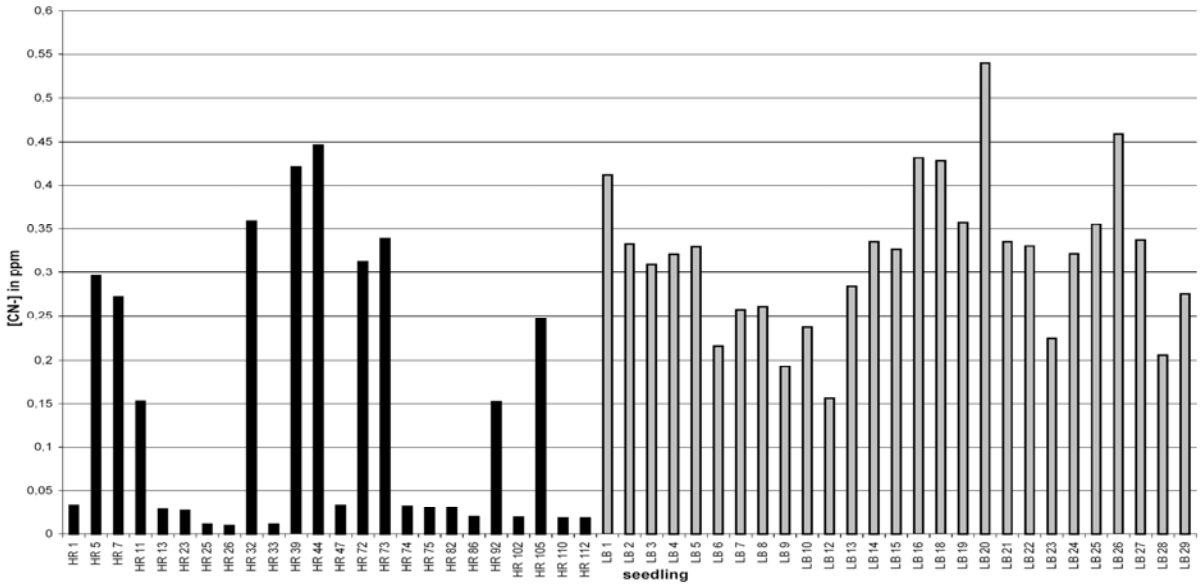


Fig 4.10. Total cyanide content of some HxR (black columns) and of some LXB (grey columns) seeds determinated by the colorimetric method.

The outcome was not satisfying enough, mainly because of a low reproducibility of the measures. In fact, it was indispensable to repeat a new calibration curve before of every analysis, because their results were quite different among the various trials (Fig. 4.11).

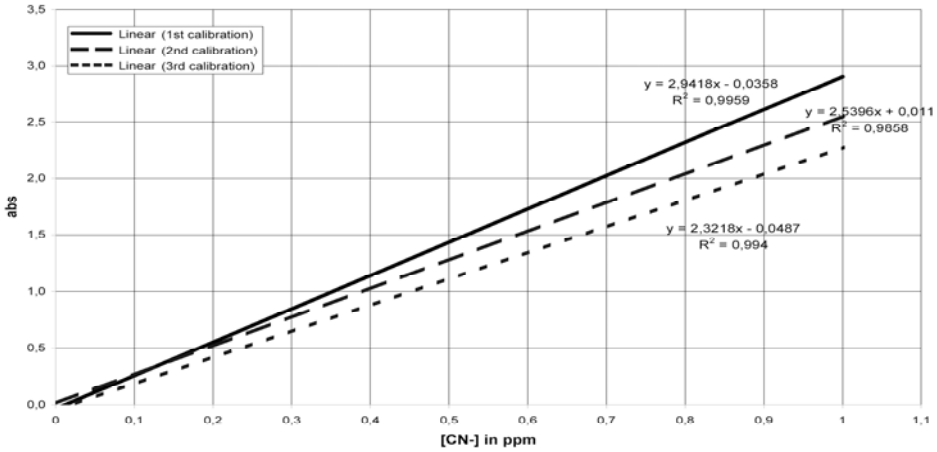


Fig 4.11. Example of the differences occuring among different calibration curves.

Moreover, the fact that the method cannot discriminate between different CNGs, being only able to show the total cyanide developed by the degradation of all of them, was an obstacle to the precise and accurate phenotyping that is needed to make a reliable QTL analysis. Thus, in order to examine which CNGs are present in the apricot seed and to quantify them in a precise and robust way, a quantitative NMR analysis was done.

4.3.2 Quantitative NMR

qNMR revealed that apricot seeds contains only amygdalin. No prunasin was detected, not even in traces. Fig. 4.12 shows an expansion of the region of interest in the seed extract spectrum, namely the region containing the amygdalin and the internal standard peaks.

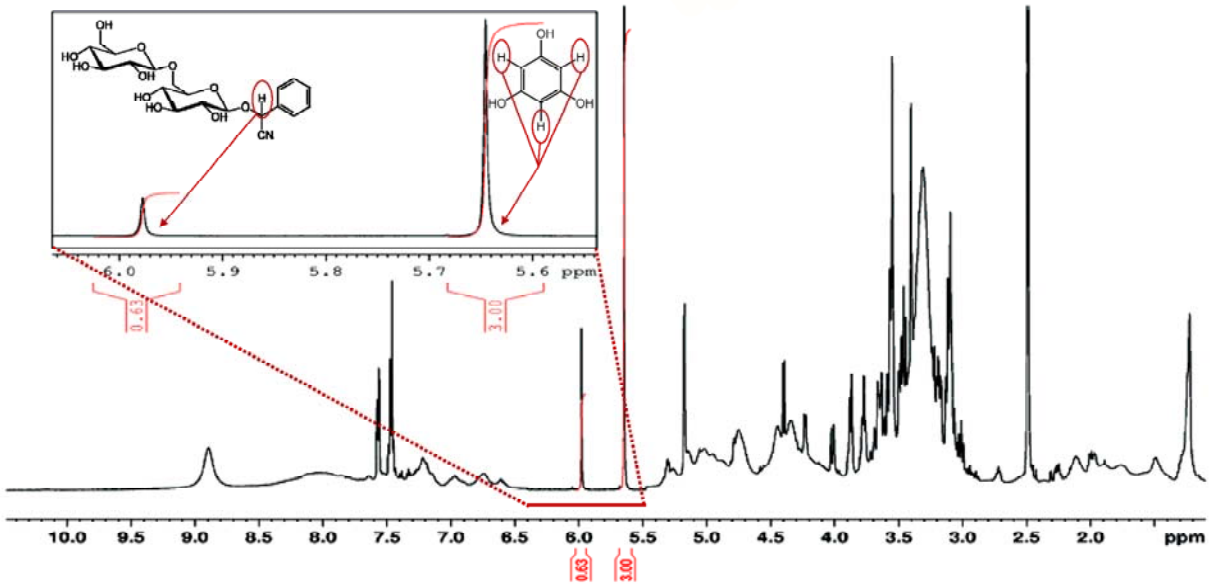


Fig 4.12. Enlargement of the spectral windows containing the amygdalin and the phloroglucinol (internal standard) peaks. In the frame are indicated the protons responsible for the signals (amygdalin on the right, phlg on the left).

The phenyl ring NMR signal of amygdalin corresponds to a chemical shift of δ 5.98 for the pure compound in DMSO-d6 and of δ 5.97 for the seed extract. According to the literature, the most popular method for determination of absolutes values by qNMR is the main component analysis using the internal standard (Malz and Jancke, 2005; Pauli et al., 2005; del Campo et al., 2005). In choosing the internal standard for quantification, several factors have to be taken into account: the standard compound has to give a clear single signal, it has to be soluble in the chosen solvent and, without overlapping, it has to possess a resonance window close to the analyte one. Phloroglucinol (from here on abbreviated as phlg) gives a clear and strong singlet signal, it is soluble in DMSO, it is not overlapping with any other signal of the seed extract and its peak, with the chemical shift of δ 5.64, is located really near to the amygdalin one (Fig. 4.12).

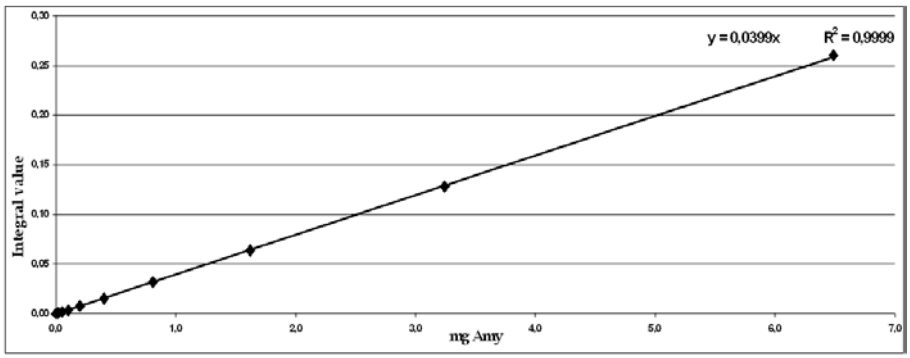


Fig 4.13. Linearity test for the qNMR analysis.

To check the linearity of the method, 20 serial dilution of pure amygdalin solved in DMSO-d6 were prepared. Fig. 4.13 shows the integral values of the phenyl ring signal of amygdalin experimentally determined versus the amygdalin mass present in the NMR tube. Linear regression yielded a correlation coefficient of 0.9999.

Amygdalin content in H×R population. Both parental lines Harcot and Reale d´Imola exhibit a sweet phenotype, whereas the population resulting from their cross shows a percentage of bitter individuals. As shown in Fig. 4.14, bitter and sweet seed can easily be distinguished from their spectrum.

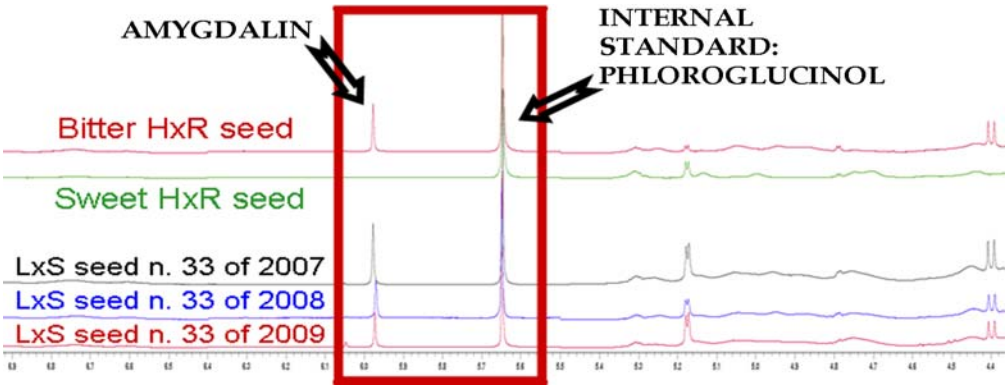


Fig 4.14. Upper spectra: bitter and sweet seeds from HxR. Lower spectra: the same individual from LxB among the three harvesting years.

The qNMR analysis of this cross population has revealed that also some seeds that are sweet to the taste contain traces of amygdalin, in a range that goes between 0,0006 to 0,03% (Fig 4.15).

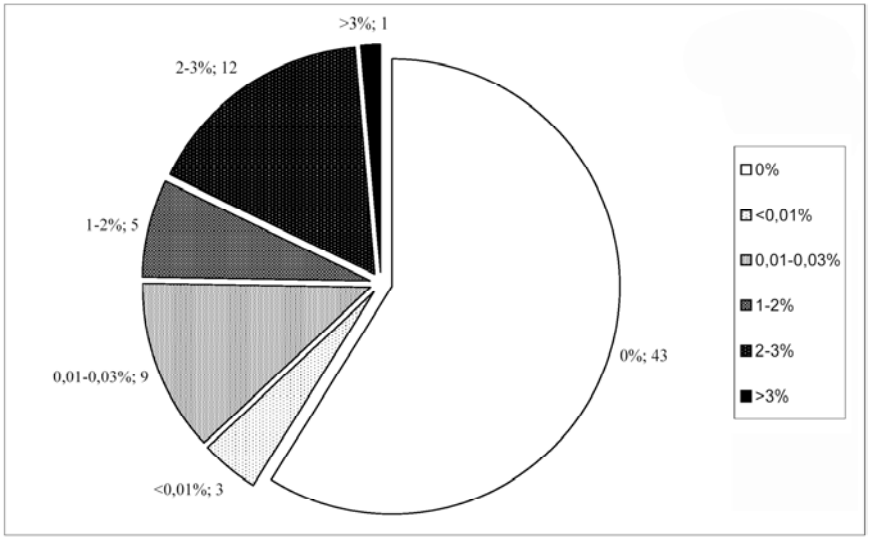


Fig 4.15. Amygdalin distribution in the HxR population. Individual percentage values were grouped in class, as shown in the legend.

As summarized in table 4.4., 59 individuals are completely sweet, without any trace of amygdalin, whereas 16 individuals shows very little amount of this cyanogenic glucoside, even if it cannot be perceived by tasting. This means that among the H×R offspring the 22.2% are bitter-tasting (18 individuals over 73) and the remaining are sweet. It would be interesting to try some experiments with standard dilution of amygdalin to find at which percentage is found the bitter-sensitivity threshold of the tongue.

Tab. 4.4. Percentage amygdalin content distribution in HxR population (pop). Double line: boundary between sweet (upper part of the table) and bitter (lower part) taste of the seed. N: number of seedlings. %: percent of the population. POP: total number of individuals in the population

Content	N	%
0%	43	59%
<0,01%	3	4%
0.01-0.03%	9	12%
0.03-1.5%	0	0%
1.5-2%	5	7%
2-2.5%	8	11%
2.5-3%	4	5%
3-3.5%	1	1%
3.5-4%	0	0%
4-4.5%	0	0%
>4.5%	0	0%
POP	**73**	

The use of the NMR spectroscopy was crucial for allowing the quantification of these little amounts and this new set of data will be implemented in a QTL mapping program to confirm the association between the bitter character and the linkage groups 4 and 5 previously found only by tasting the seeds (data not shown).

Amygdalin content in L×B population. The second population analyzed is derived from the bitter×bitter cross Lito×BO81604311, being this latter a selection derived from the breed of the bitter San Castrese with the sweet Reale d´Imola. All the L×B offspring are bitter, but the distribution of this bitterness changes among the three years of harvesting.

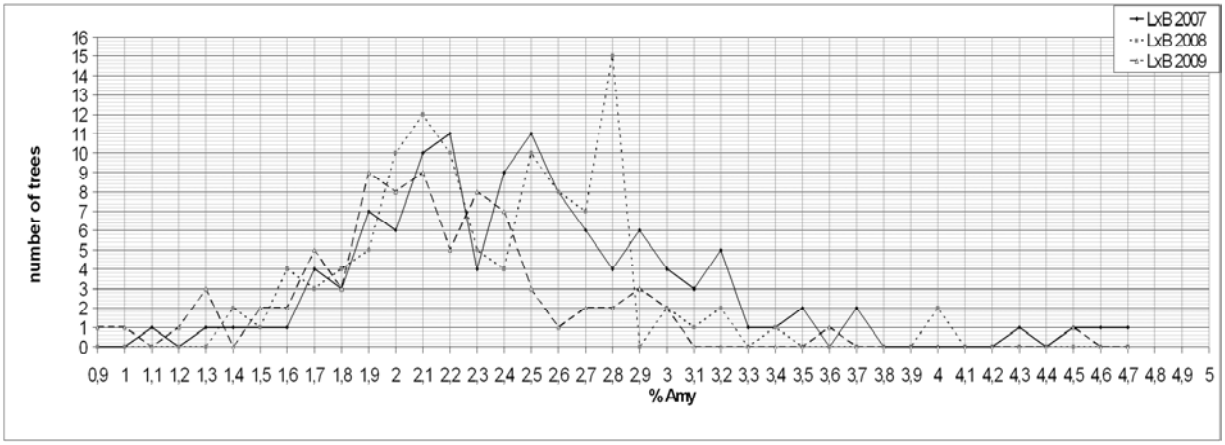


Fig 4.16. Amygdalin distribution in LxB population trend. Black continuous line: harvesting year 2007; dotted line: year 2008; broken line: 2009.

In Fig. 4.16 are put in a graph the average percentage amygdalin content (X axis) and the number of trees having it (Y axis). It become so visible a tri-modal trend, whose peaks result slightly shifted depending from the year. When the data are clustered in classes, the distribution begins to normalize, showing just two peaks (Fig. 4.17). This bi-modal trend is not uncommon in this population, which has often showed differences between the early ripening individuals and the late ones.

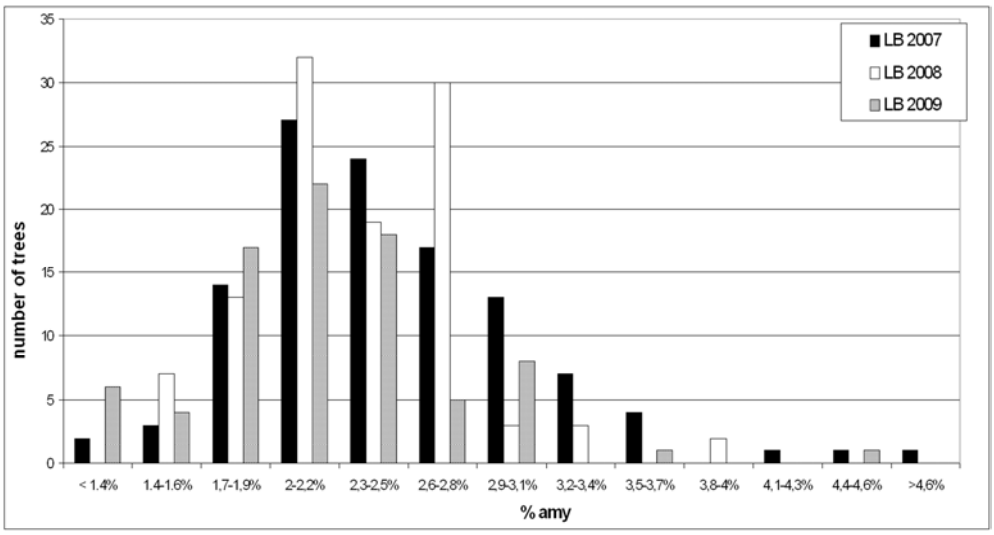


Fig 4.17. Average percentage classes distribution among the three harvesting years fo the L×B progeny. Black columns: 2007; white columns: 2008; grey columns: 2009.

It has to be noticed that the 2007 and 2009 populations seem to be more spread, showing very low and very high amygdalin content (with the 2009 harvest a little bit shifted toward the less-bitter values) while the 2008 one is more close to the average value. This trend is summarized in the Table 4.5.

Tab. 4.5. Percentage amygdalin content distribution in LxB population among the different harvesting years. N: number of seedlings; %: percent of the population; POP: total number of individuals in the population.

population	LB 2007		LB 2008		LB 2009	
Content	N	%	N	%	N	%
<1.5%	3	3%	2	2%	6	7%
1.5-2%	16	14%	18	17%	21	26%
2-2.5%	40	35%	41	38%	37	45%
2.5-3%	34	30%	40	37%	14	17%
3-3.5%	14	12%	6	6%	2	2%
3.5-4%	4	4%	0	0%	1	1%
4-4.5%	1	1%	2	2%	0	0%
>4.5%	2	2%	0	0%	1	1%
POP	**114**		**109**		**82**	

The average amygdalin content over the three years is 2.3%. Taking the flanking classes to the ones containing the average value, it is in fact found that: in 2007 the 79% of the individuals fell in the average range, in 2008 the 91% and in 2009 the 88%. The seedlings exceeding the 3.5% of amygdalin content are 6% in 2007 and 2% in both 2008 and 2009. In this latter years it has increased the number of individuals under the 1.5% of amygdalin: 7% versus 2% in 2008 and 3% in 2007.

The minimum and the maximum percentage content of amygdalin is reported, also for the H×R population, in the table 4.6., with again the individuals coming from the L×B 2007 holding the bitterness record, the L×B 2009 holding the less-bitter record and L×B 2008 being assessed in the middle showing the higher minimum values. Interestingly, the lower maximum values are the ones determined in H×R.

Tab. 4.6. Minimum and maximum amygdalin content (and their average) in HxR and LxB populations, subdivided into the four quarters value. Last column: average amygdalin content over all the population. All the amygdalin values are given as percentages. For HxR only the bitter tasting were considered. *Italic bold*: higher maximum values; bold: lower maximum values; dotted underlined: lower minimum values; doubled underlined: higher minimum values.

population	amygdalin content	1st quarter	2nd quarter	3rd quarter	4th quarter	average	total average amygdalin content
HR	min	1,2	1,4	0,8	1,2	1,5	2.2
	max	**3,5**	**3,4**	**4,1**	**3,1**	**3,1**	
LB 2007	min	0,9	1,0	1,2	1,1	1,1	2,5
	max	***5,3***	***5,4***	***5,0***	***4,9***	***4,7***	
LB 2008	min	1,2	1,3	1,4	1,1	1,4	2,3
	max	4,3	4,2	4,2	4,0	4,0	
LB 2009	min	0,8	0,8	1,0	0,9	0,9	2,2
	max	4,5	4,5	4,4	4,9	4,5	

From all this variations in the amygdalin accumulation among different years in L×B (summarized at a glance in the box-plot of Fig. 4.18), it is clear that the environmental conditions are playing a quite important role in the determination of this quantitative trait.

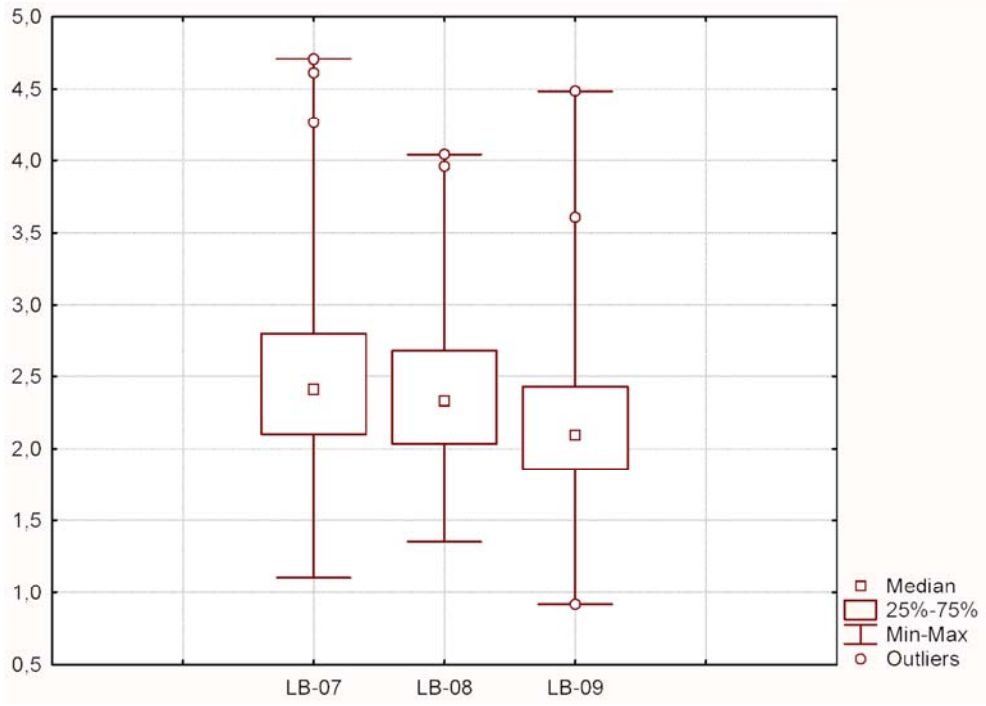


Fig 4.18. Box-plot of the amygdalin content found in LxB in the three analyzed years. Y axis: amygdalin percentage content.

Thus the environmental data regarding the temperatures and the rain of the months going from March to June (namely the more important month for fruit development) of every of the three years of L×B harvesting were retrieved from the archives of a meteorological station near the fields and analyzed with the software Statistica 7.0 (StatSoft). Looking to the mm of rain felt, the maximum, the minimum and the average temperatures related to every single month, it seems that: in 2007 the concentration of amygdalin was more correlated to the rain felt in April, May and June, whereas in 2008 it was more dependent from the rain felt in March (and then in April), and in 2009 the rains of May and April were equally important and it was arising also the significance of the minimum temperature of March. For every year the less correlated variables were the maximum temperatures of June, and of May in 2007 and 2009. When the average data of all the parameters were analyzed, the clusters resulted as the one reported in Fig. 4.19.

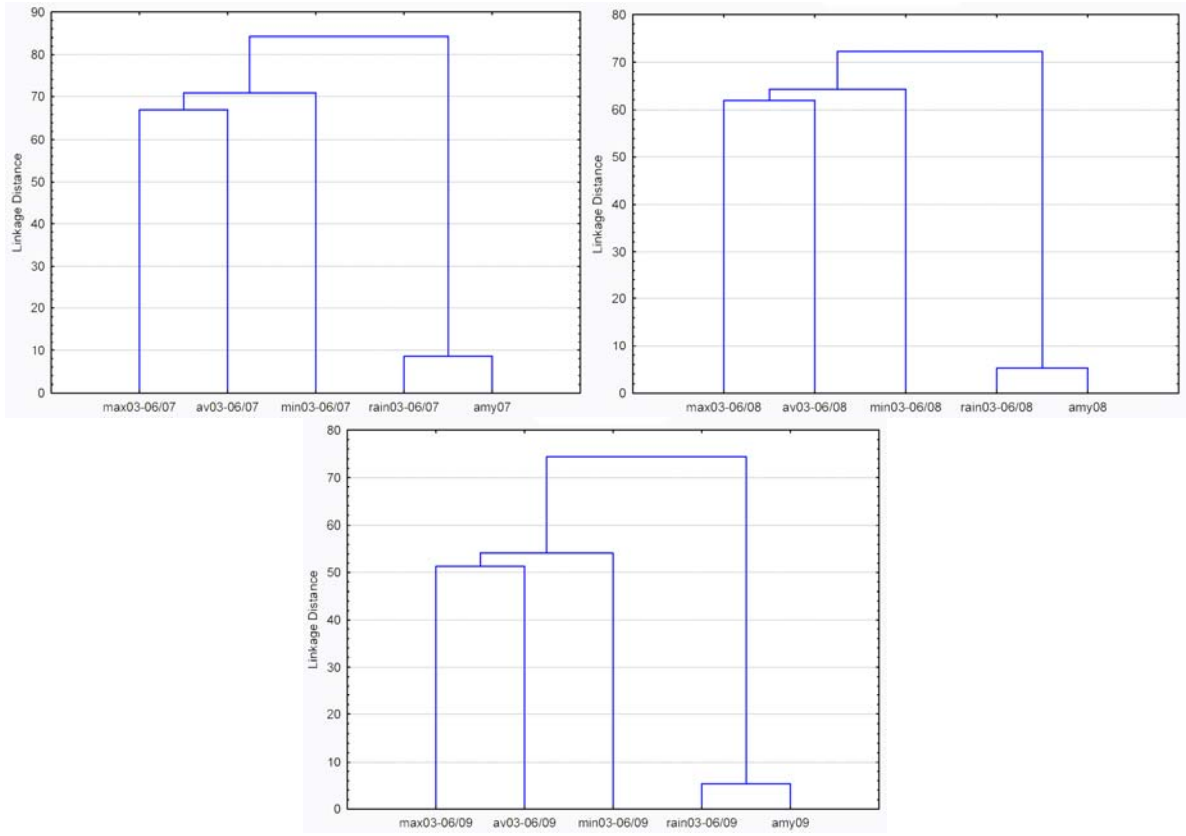


Fig 4.19. Linkage analysis among the amygdalin content and the average environmental parameters: max (maximum temperature), min (minimum temperature), av (average temperature) and rain (mm of rain fallen). Euclidean distances are shown.

These average values confirms that the amygdalin content is much more strictly correlated with the rainfall than with every other environmental variable examined. The clusters shows the very same trend for every analyzed year, with marginal differences in the Euclidean distances. The maximum and also the average temperatures are quite less involved, where only the minimum seems to may play a role. The projection of the variable on the factor plane allow to understand the direction of this correlation: an increase of the amygdalin content is correlated with a decrease of the rainfall. This correlation resulted then to be more stringent in the 2008, a little less in 2007 and still lesser in 2009 (Fig. 4.20).

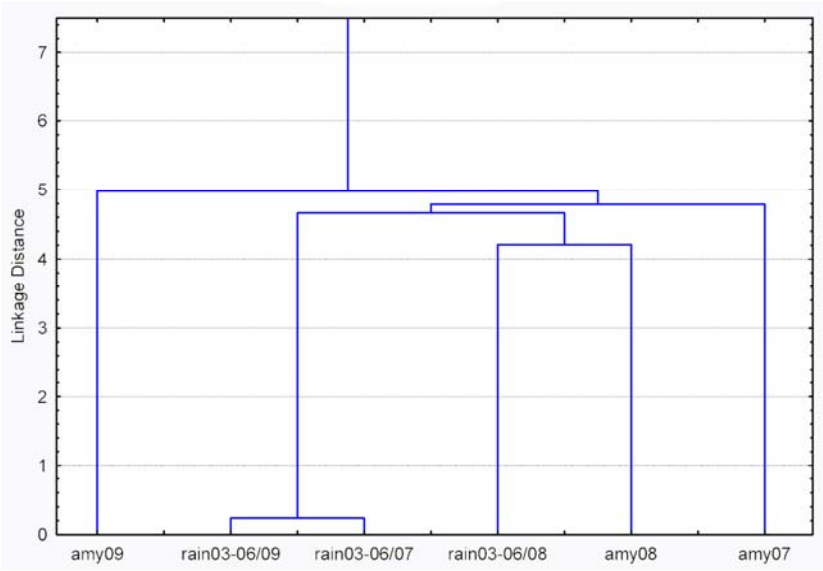


Fig 4.20. Vertical cluster visualization of the Euclidean distances between the amygdalin content measured in the different years and the respective average rainfall.

Summarizing, the distribution of the amygdalin content was found to be more shifted toward "extreme" values in 2007 and 2009. The year 2008, whose growing season was characterized by a higher rainfall, shows a higher percentage of individuals close to the average value. Moreover, the accumulation of amygdalin seems negatively correlated with the number of rainy days. To give an ecological reason of this trend counting only upon these data is entering in the speculation field. But it is still interesting to make some hypotheses. For first, the negative correlation amygdalin–rain can be explained by a change in the carbon:nitrogen ratio. When it rains the plants can keep the stomata opened for more time, thus being able to photosynthesize more. Then, to produce CNGs, the nitrogen could become a limiting factor, being its adequate presence a fundamental prerequisite (as demonstrated by the increase in cyanide potential upon nitrate administration reported by Busk and Møller, 2002, and Jørgensen et al., 2005). Another explanation could be that more rainy days means less sunny days, and this can influence the CNGs production because the first two anabolic enzymes are two cytochrome P450-dependent monooxygenases and their activity is light-driven (Schröder et al., 1999; Jensen et al., 2011). Instead, to explain why in less-rainy 2007 and 2009 the amygdalin distribution is more distant from the average value, we can postulate a field effect. When it rains no plants have problems for the water supply, but when it doesn't different positions in the field (i.e. near the trenches) could mean different water intake, here hence different CNGs production. These hypotheses could also match the concept that consider the CNGs as storage compounds, because no water stress would imply no necessity of an increased production of stock metabolites. In any case, three

harvesting years done on only one population provide not enough data to build a more detailed and reliable set of hypothesis.

But from the survey of all the qNMR results something else was found to be impressive: if in between the different quarters of a seed there was sometimes a little or very little difference in the percentage amygdalin content, between some others the difference resulted to be more than 1 percentage point, with peaks of 1.4, 1.7 and 2.5 (see following figure).

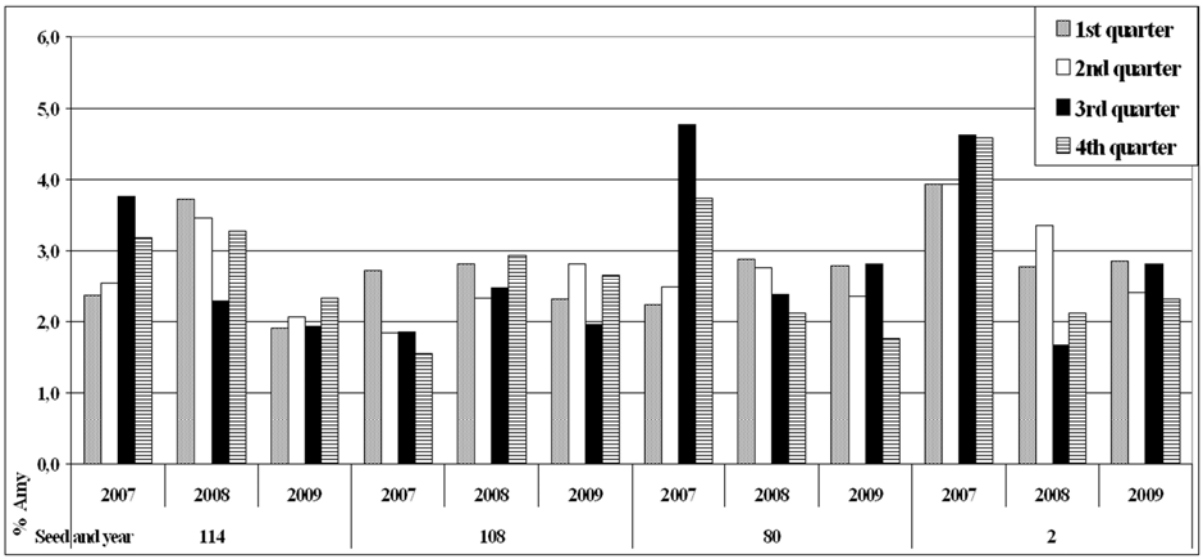


Fig 4.21. Example of some seeds showing great differences in the amygdalin content among their quarters.

These dissimilarities in the amygdalin content of the seed quarters are likely due to an uneven distribution of this secondary metabolite in the seed. So far there are few studies and few agreement on the localization of the CGs in the seeds, mostly because some of their features (i.e.: being unpigmented, water-soluble and with low molecular weight) make them notoriously difficult to localize at the tissue and subcellular levels (Poulton and Li, 1994). So, if the tissue prints made in plum and black cherry seeds by Poulton and Li showed a sort of uniform distribution of the CGs (excluding the procambial strands and the longitudinal edges of the seeds), an experiment of Raman imaging carried on by Micklander et al. (2002) in almond resulted in a not measurable amount of amygdalin in the centre of the seed. Therefore it was decided to try a Raman imaging approach on some sample, in order to get a direct look on the amygdalin localization in the apricot seeds.

4.4 *Amygdalin* in situ *localization*

Raman spectroscopy offers several advantages for microscopic analysis. Since it is a scattering technique, specimens do not need to be fixed or prepared in particular ways. From the Raman spectrum, then, is possible to identify the various molecules present in the sample. In Fig. 4.22 is shown a set of Raman spectra collected from three different seed slices. The CN group of the amygdalin gives rise to a clear signal at 2246 cm^{-1}.

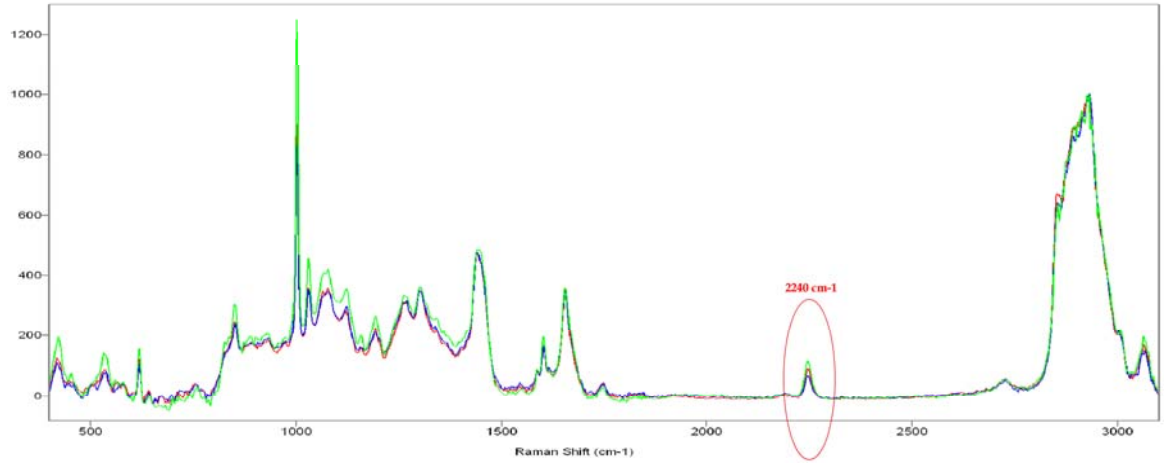


Fig 4.22. Raman spectra of three seed slices (15 µm thick). The CN peak is red-circled.

Two kind of approaches were done. In the first, the Raman spectra were collected along a line horizontally or vertically crossing the seeds. One result of this experiment is shown in the following figure.

The concentration of amygdalin is different in the two directions, being higher at the horizontal line margins, but not in both the vertical edges.

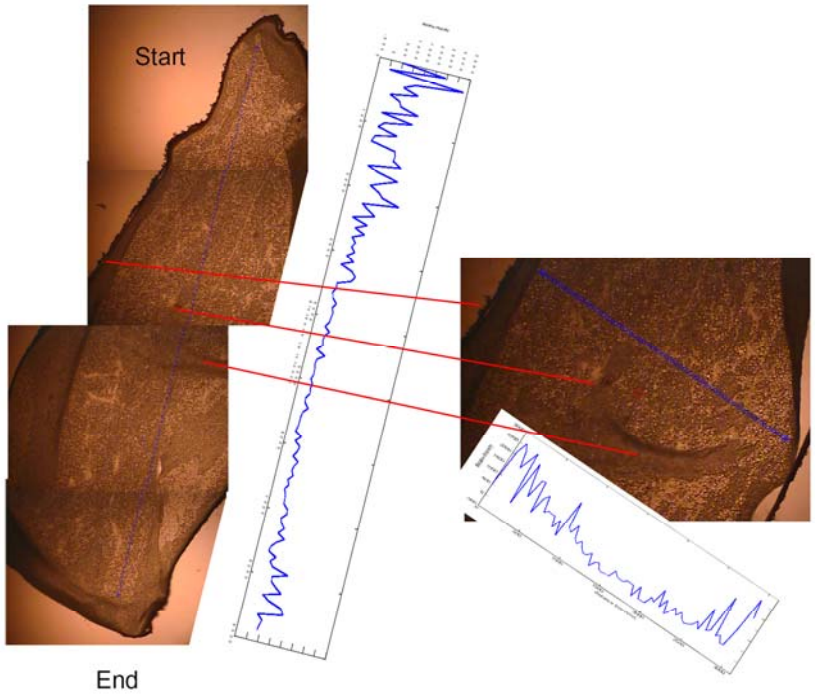


Fig 4.23. Raman lines spectra. Vertical line: 120 points at a step size of 74.11 µm; horizontal line: 80 points, step size 37.98 µm; red lines indicates the points of correspondence between the two images.

The second approach was oriented at building a Raman map, acquiring thousands of spectra from all over the field of view to generate an image showing the localization and the amount of amygdalin. In the seed used as sample for this experiment and shown in Fig. 4.24, the amygdalin resulted to be present almost only on the right side (map localization) and more concentrated in the centre than in the upper and bottom part (line analysis).

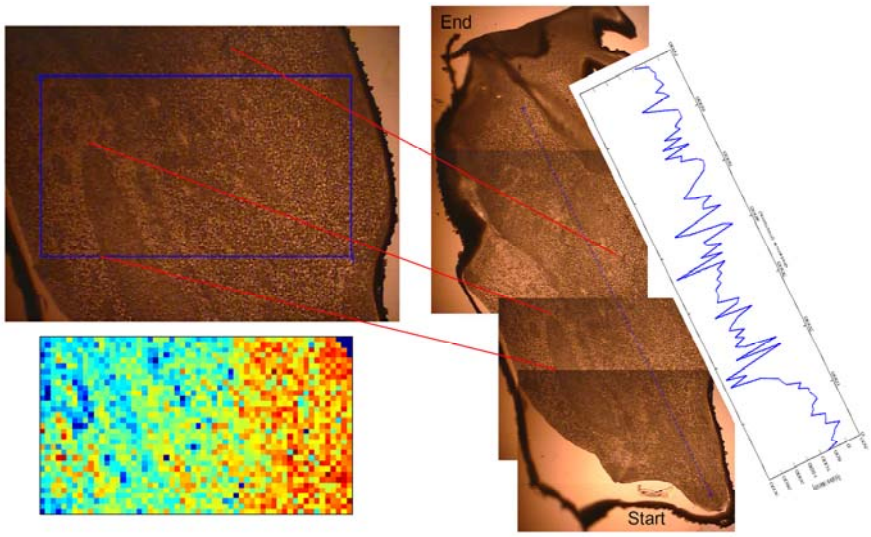


Fig 4.24. Raman line and map analysis. Map: 58x33=1914 points, step size 50 µm; line: 100 points, step size 70.45 µm; red lines indicates the points of correspondence between the two images.

The spatial resolution of Raman microspectroscopy in the low micrometer scale allows to map even single cells (Krafft et al., 2003). Thus, using a smaller step size it was possible to visualize the cell walls, and to obtain an image mapping the amygdalin and the lipid content (Fig. 4.25).

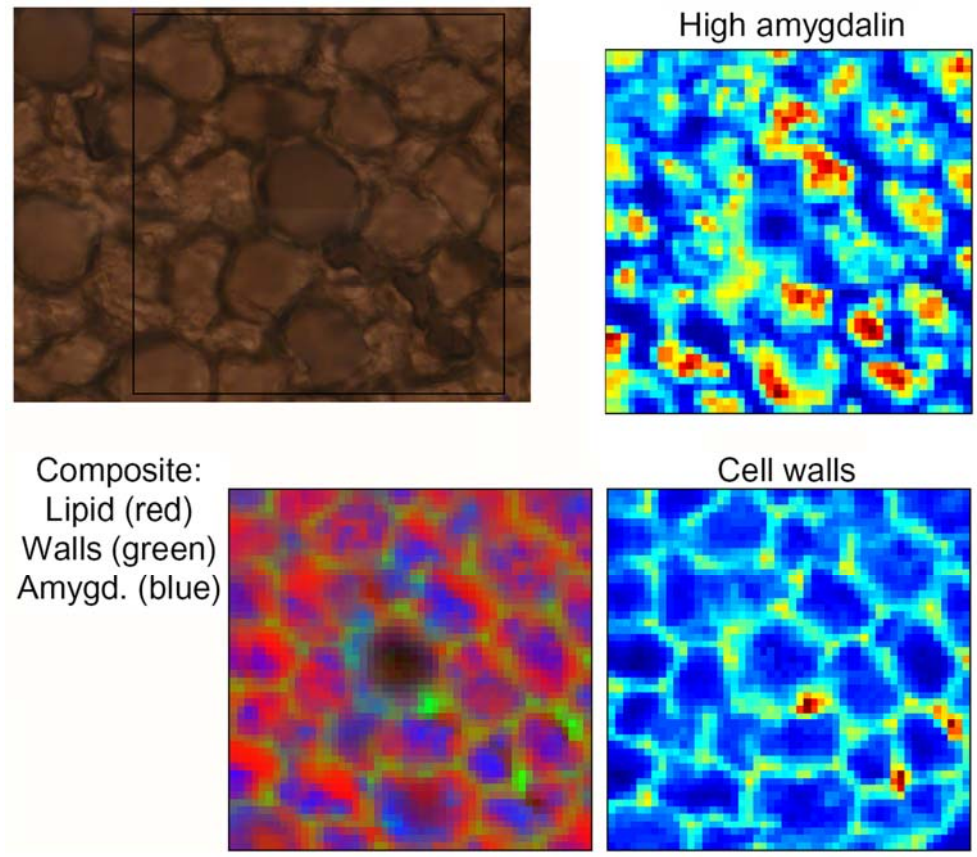


Fig 4. 25. Raman local map analysis. 50x50=2500 points at a step size of 2.5 µm.

From these analyses it is clear that the amygdalin is not uniformly distributed in the seed and, moreover, it seems that its localization does not follow a unique rule for all the seeds. In addition, in almost all the analyzed samples this compound was located in a different way with respect of what has been found in almond's Raman images (Micklander et al.; 2002; Thygesen et al., 2003), namely not with its maximum concentration near the seed's margins (see Fig. 4.26). This findings are consistent with the hypotheses that see in the storage one of the principal function of the CNGs in the seeds (Swain et al., 1992b; Swain and Poulton, 1994; Gleadow and Woodrow, 2000 and 2002; Busk and Møller, 2002; Jørgensen et al., 2005). That would in fact mean that the main focus is not on where the amygdalin is located but on its presence itself.

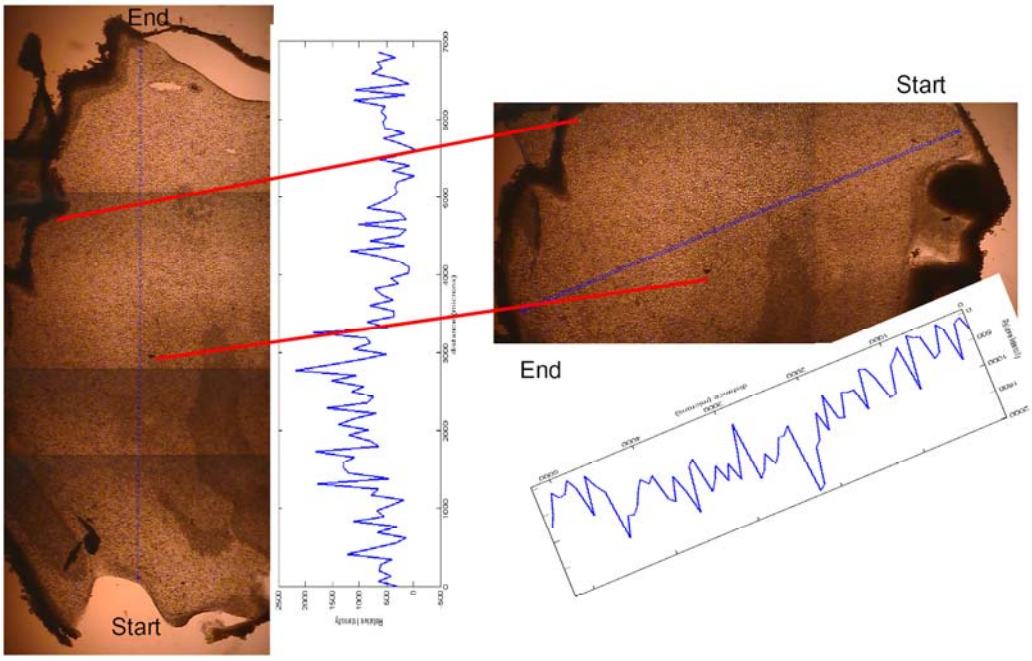


Fig 4.26. Raman line analysis. Vertical line: 100 points, step size 69.34 µm; horizontal line: 75 points, step size 70.05 µm; red lines indicates the points of correspondence between the two images

4.5. Quantitative trait loci analysis

4.5.1 Genetic regions identified in the L×B population

All the phenotypic data coming from the quantification of the amygdalin *via* NMR were used to make a comprehensive QTL analysis, thus allowing to follow the loci related both to the single quarters values and to the average amygdalin content of every seed for each single year and on the mean of the three years. The results were similar using either the Kruskal-Wallis test (KW) or the Interval Mapping (IM) algorithm. The LOD threshold was estimated at LOD 3 by permutation test. Several putative QTLs were identified in LG1 and LG6 in both parental lines, and in LG4 and LG3 in Lito only (Tab. 4.7).

Tab. 4.7. Marker association to the bitter phenotype as determined by Interval Mapping on the parental lines Lito and BO81604311. For every LG are listed the molecular markers underlying the peak, their position on the genetic map (in cM), the relative LOD scores and percentage of explained variance (% exp). Highest LOD scores for every peak are highlighted in bold; ns means not significant.

Marker	LG	cM	2007		2008		2009		average	
			LOD	%exp	LOD	%exp	LOD	%exp	LOD	%exp
UDAp-441	L1	13.7	ns	-	ns	-	3.42	18.7	ns	-
CPPCT34a		47.7	ns	-	ns	-	**4.89**	25.1	ns	-
PGM-SSR		48.9	ns	-	ns	-	4.75	24.4	ns	-
AMPA-109		72.4	ns	-	ns	-	3.49	18.3	ns	-
BPPCT-011		75.4	ns	-	ns	-	ns	-	**3.22**	12
PGM-SSR	B1	73.3	ns	-	ns	-	**4.81**	24.3	ns	-
BPPCT-028		114.9	ns	-	ns	-	ns	-	**3.4**	13.7
UDAp-446	L3	14.4	**3.02**	12.5	ns	-	ns	-	ns	-
UDAp-404	L4	15.4	**3.25**	13.3	ns	-	ns	-	**3.44**	13.9
BPPCT-040	B4	44.6	ns	-	2.92	14.3	ns	-	ns	-
UDAp-454	L6	13.5	ns	-	ns	-	ns	-	**3.51**	14.4
UDAp-492		13.6	ns	-	ns	-	ns	-	**3.51**	14.4
BPPCT-009		14.9	ns	-	ns	-	ns	-	3.39	15.7
CPPCT01		31	ns	-	ns	-	2.97	15.8	ns	-
BPPCT-009	B6	35	ns	-	**3.29**	14.8	ns	-	**4.21**	16.9
CPPCT01		51.9	ns	-	ns	-	2.96	15.7	ns	-

A significant QTL for the female parent is clearly visible in the 2009 data and it is located in Lito's linkage group 1 (L1), with a peak of LOD 4.89 (accounting for the 25,1% of the variance) on the marker CPPCT34 (see Fig. 4.27). Two minor peaks are visible in correspondence of UDAp-441 on the top of the LG and of AMPA-109 at 72.4 cM. Knowing that we have mapped a candidate gene for an MDL isoform on the top of this LG, the best BLAST match for our UGT sequence was on the scaffold 1 of the peach genome, Joobeur et al. (1998) have found an Mdl1 in the same region, and Sánchez-Pèrez et al. (2010) have located in the same position an Ah1, an involvement of this linkage group in the determination of a bitter phenotype seems clear. If in 2007 and 2008 the LOD threshold is not reached, in the average of the years the significance is re-established, even if the QTL is slightly shifted toward the end of the LG (peak at 75.4 cM). The main QTL determined in L1 and related to the same year 2009, found a correspondence also in the linkage group 1 of BO81604311 (B1, see Fig. 4.27), with comparable LOD and percentage of explained variance. Unlike Lito, BO81604311 shows a single wide peak, centred at 73.3 cM; but, similarly to its

female partner, no significant QTL emerged in 2007 or 2008, but the threshold is reached again in the analysis done on the means of the three years, also this time more shifted toward the bottom of the LG (peak at 114.9 cM).

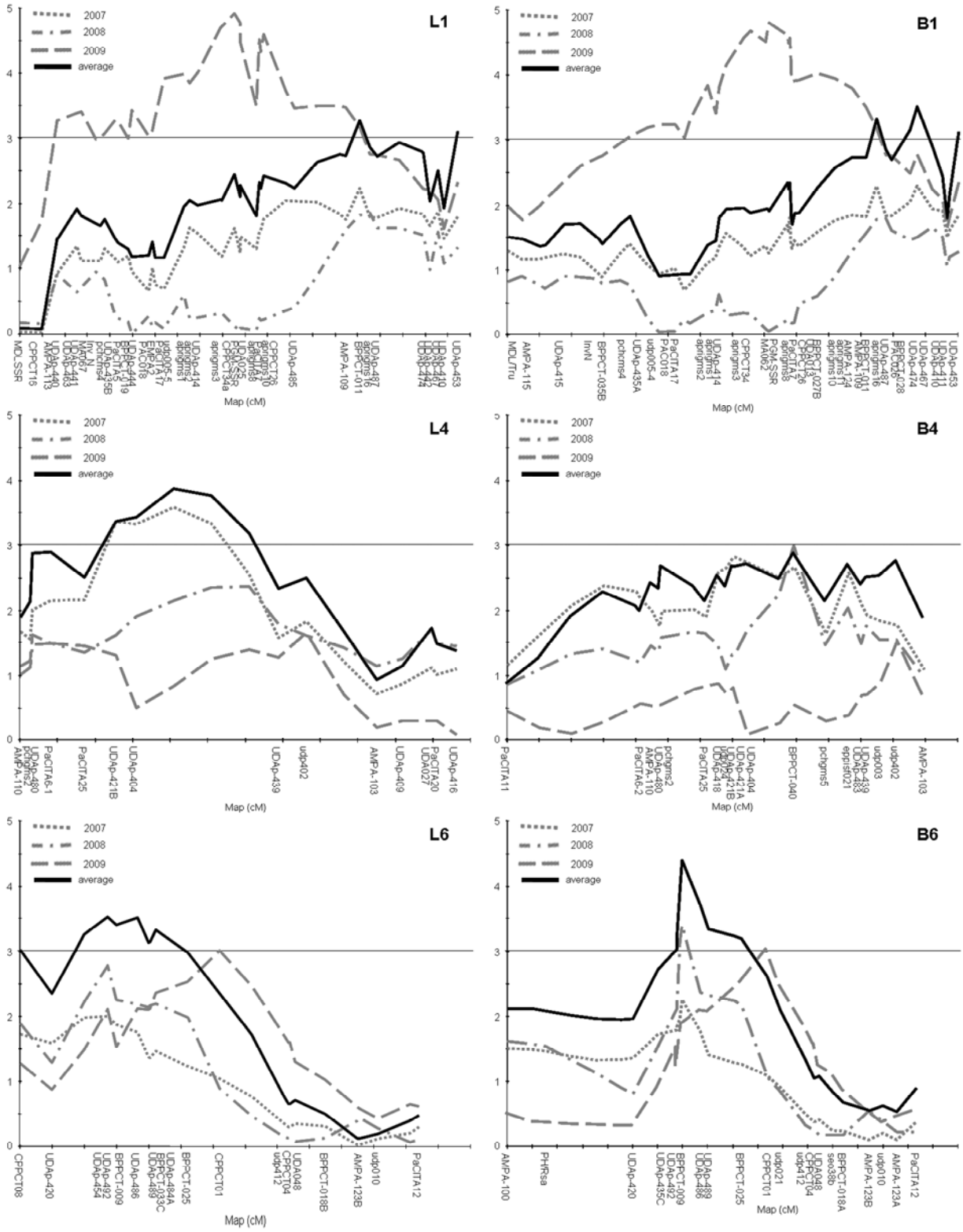


Fig 4.27. QTLs identified in Lito and BO81604311. X axis: molecular markers position; Y axis: LOD value (threshold highlighted by a continuous line). Grey curves correspond to single years values; black curve corresponds to the average of the data collected on the three years.

A second significant QTL was found in the middle of B6, centred on the marker BPPCT-009 with a LOD of 4.21 in the average of the years and of 3.29 in 2008 (Fig. 4.27). Again, no QTL is found in 2007. In 2009 the threshold of 3 is almost reached, with a LOD of 2.96 and a peak on the marker CPPCT01. The very identical situation is found in L6, where the highest peak of 2009 is on CPPCT01 and shows a LOD of 2.97. Another region of L6, near the top, resulted very significant when the mean of the three years is analyzed: a peak covering 1.4 cM (from 13.5 to 14.9) with a mean LOD of 3.49, in total accounting for the 73.3% of the variance. The marker on the lower side of the region is again BPPCT-009.

Another linkage group that seems involved in the shaping of the bitter trait is L4 (Fig. 4.27). Here, a minor QTL is found in 2007 on the marker UDAp-404 and arise again in the same position, in the middle-high portion of the linkage group, and with a slightly higher LOD when looking at the average of the years. In BO81604311 the LOD values did not reach the threshold, being the LOD slightly under 3 (see Tab. 4.7).

An additional putative QTL was located in L3, with a LOD score just over the threshold in 2007 (LOD 3.02, only 12.5% of explained variance), but with the significance of the marker accountable for the peak (UDAp-446) supported by the K* index.

The relative low LOD of these QTLs is probably due to the fact that Lito×BO81604311 is a bitter×bitter cross with all bitter offspring, entailing that the big QTLs, the loci responsible for the sweet or bitter determination, are clearly not visible analyzing this population. But that particular situation also means that the "accessory" loci, the secondary QTLs, the ones responsible for the shaping of the 'bitter' trait and thus for the determination of the level of bitterness, can finally stand out. This can also explain why other QTLs for bitterness, such as those described in the linkage group 5 by Dirlewanger et al. (2004) and Sánchez-Pérez et al. (2007), were not detected in the L×B population.

Another peculiarity of this study was that the phenotyping done via qNMR was performed on the seed cut in quarters. This fine analysis allowed us to appreciate dissimilarities in the LOD scores, dissimilarities that sometimes were big enough to mean a boundary between significance and not. For example, the highest LOD of the linkage group S2 (peak on marker UDA20, see Fig. 4.28) was 1.2 when the first series of quarters' data were analyzed, it was then 3.4 for the second series, 3.7 for the third and 1.5 for the fourth and the total LOD value for UDA20 was 2.9.

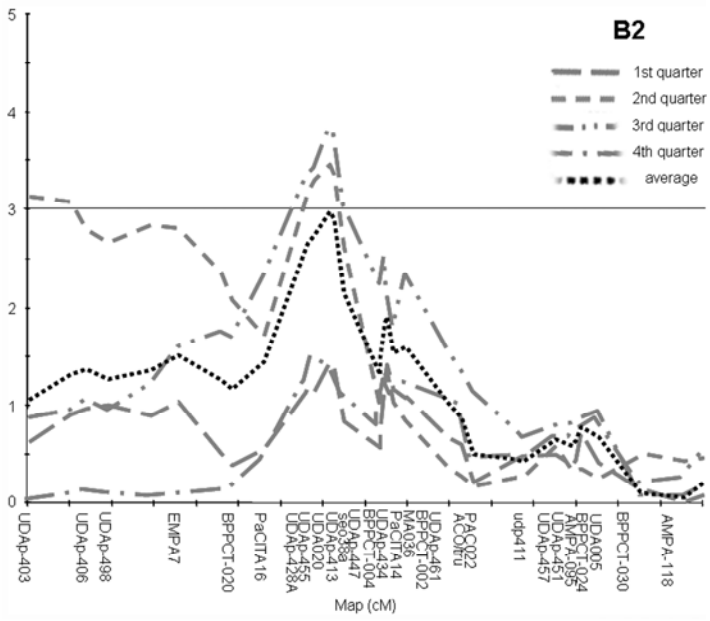


Fig. 4.28. Example of QTL analysis by using the phenotypic data collected on the single quarters of the seeds: grey lines for every quarter, black line for their average.

Also some major QTLs (even if the LOD score was always over the threshold) have showed differences of more than 1 point. QTLs whose significance was not confirmed by the average LOD value, but only arising in the particular data of some quarters were not considered acceptable.

This means that an accurate phenotyping can be more than a crucial step for a right QTL analysis. Even more so when the character studied is dependent from a complex mechanism of differential accumulation that is not completely known (Negri et al., 2008).

4.5.2 Genetic regions identified in the H×R population

As far as the H×R population, only one year of seed harvesting was available. Also in this case the QTL analysis was done either with KW and IM, with comparable results. The LOD threshold calculated by the permutation test was different for the two parents, because of their differences in map saturation and length. The estimated LOD threshold values are 13 for Harcot and 2.3 for Reale di Imola. According to this parameter three QTLs were found, in LG5 in both parents and in LG5 in Reale di Imola only (Tab. 4.8).

Tab. 4.8. Marker association to the bitter phenotype as determined by Interval Mapping on the parental lines Harcot and Reale. For every LG are listed the molecular markers underlying the peak, their position on the genetic map (in cM), the relative LOD scores and percentage of explained variance (% exp). Highest LOD scores for every peak are highlighted in bold.

marker	LG	cM	LOD	%exp
CPPCT14	R4B	0.00	4.23	23.7
paCITA20		3.07	**4.49**	25.3
CPPCT20b		11.02	3.56	30.1
CPPCT13	R5	26.4	**25.79**	93.2
UDAp439	H5	24.3	27.22	93.2
paCITA21		46.1	**30.63**	93.2
udp401		52.8	16.7	83.8
CPPCT36		58.6	14.42	83.9
AMPA-112		62.8	16.88	83.9

A major QTL, exceeding the LOD score of 30, is located in the linkage group 5 of Harcot (Fig. 4.29). The significant region covers 28.5 cM, between the markers UDAp439 and udp401, with the peak on the SSR paCITA21 (accounting for the 93.2% of the variance). A minor QTL is present toward the end of the LG. As far as its peak falls in a gap it could be evaluated as not trustworthy, but the zone nearby is clearly involved, because three markers (CPPCT36, AMPA-112 and UDA042) have a LOD score over the threshold. Looking the meanrank means of the segregation classes, is visible that this QTL is almost all dependent from the Harcot alleles. This main QTL determined in H5, found a correspondence also in the linkage group 5 Reale d'Imola, with 25.79 of LOD score and 93.2 of explained variance (see Tab. 4.8). Unlike Harcot, where the peak is in the middle of the LG, in Reale the acme is located in the bottom part. The segregation values of Reale's markers confirm Harcot alleles as the main component in determining this QTL.

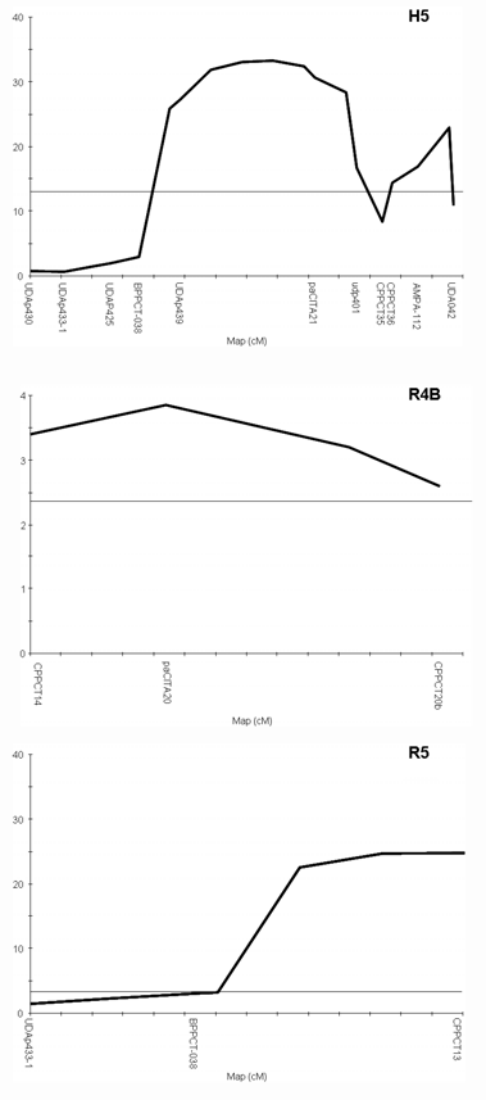


Fig 4.29. QTLs identified in Harcot and Reale. X axis: molecular markers position; Y axis: LOD value. The threshold is highlighted by a continuous line and corresponds to 13 for H5 and to 2.3 for R4B and R5.

As expected, the major QTL for bitterness, already identified in LG5 by various authors (Joobeur et al., 1998; Bliss et al., 2002; Dirlewanger et al., 2004; Sánchez-Pérez et al., 2007), is clearly visible in this population, constituted by both bitter and sweet individuals. In addition to the NMR data, the H×R QTL analysis was done also on the result of a panel-test where the seeds were tasted from a group of six people. The IM of these data found a 'perfect fit!' on paCITA21 and UDAp439 in H5, so confirming once again the involvement of this genomic region in the determination of the sweet/bitter phenotype. Unfortunately no CG have mapped here until now, but, as previously said, the number of their isoforms is quite high, so it's possible that one of them could be located here and thus a functional marker for it could be found later on.

A second significant QTL was found in the bottom region of the LG4 in Reale (Fig. 4.29), centred on the SSR paCITA20 with a LOD score of 4.49. The meanrank means of the segregation classes shows that the determinants for this QTL are the alleles of Reale. So, it seems that there are two regions involved in the determination of the sweet/bitter phenotype: one located in H5 and the second in R4B.

This confirms at least in part the inheritance model elaborated by Negri et al. (2008) and especially the fact that the functioning of both the anabolic and the catabolic pathways would be ensured by the dominant alleles at each one of the involved loci. In fact, when one considers the two QTLs found, the one on LG4 coming from Reale and the one on LG5 dependent from Harcot, and that only recessive mutation on different genes, just like the ones Negri et al. (2008) have hypothesized are in action to determine the sweet phenotype, complements, the following simplified hypothesis can be made:

- H5 is responsible for the turning on of the catabolic reaction (S- genotype, see Fig.1.5). Its synthetic system can be either on or off, as far as in any case the CNGs would be degradated (i.e. by a β-glucosidase rich cell layer, as it has been found in almond by Sánchez-Pérez et al., 2008). Considering that no QTLs have been found in H4, it can be assumed that the synthesis is on (B- genotype).
- R4B is responsible for the breakdown of the anabolic pathway (bb genotype). Also in this case it does not matter if the degradative system is on or off, because the CNGs are not produced. Considering that a QTL has been found in R5, but the allele responsible for it is Harcot's, the assumption is degradation off (ss genotype).

Tab. 4.9. Punnet square for the putative simplified genotypes hypothesized for Harcot and Reale di Imola.

gametes		Harcot			
		SB	S-	-B	--
Reale di Imola	sb	*SsBb*	*SsBb* Ssbb	*SsBb* **ssBb**	Ssbb *SsBb* ssbb **ssBb**

The cross results for these putative and very simplified genotypes are reported in table 4.9 and can be explained as follows:

- ssBb: CNGs degradation off and synthesis on → bitter individual;
- Ssbb and ssbb: degradation either on or off, but no CNGs synthesis → sweet individual;
- SsBb: both CNGs degradation and synthesis on → the individual can be bitter or sweet depending from which S allele is functional, S_1 or S_2 (Negri et al., 2008). The present hypothesis consider S_1 functional, so resulting in a sweet individual.

According to this model, the final result of the H×R cross would give 2 bitter and 7 sweet (see Tab. 4.9) individuals. This would mean a 22% of bitter offspring and it is consistent with the data obtained for this cross when analyzed by qNMR (see paragraph 4.3.2, subsection "Amygdalin content in H×R population").

Coming to an exact knowledge of the genetic basis of the bitterness trait is the essential prerequisite to the breeding and the selection of new high quality cultivars with sweet kernel; being this a really desirable quantum change for the apricot production system, as it would decrease health hazards and increase the marketability of the by-product "seed".

5. Conclusions

In apricot, the bitter flavour of seeds is determined by their content of amygdalin, a cyanogenic glycoside whose cleavage by endogenous enzymes, upon crushing, results in a release of toxic hydrogen cyanide. The presence of such a poisonous compound is an obstacle to the use and commercialization of apricot seeds for human or animal nutrition.

To investigate the determinants for the bitter phenotype a combined molecular and biochemical approach was used on two F1 apricot progenies, involving a candidate gene study, a BAC library screening, a fine phenotyping *via* quantitative nuclear magnetic resonance and a QTL analysis. Moreover, a series of Raman imaging experiments was carried out to localize directly *in situ* the amygdalin.

The primer walking on the positive clones identified by screening Lito's BAC library with the primers developed on the five candidate gene chosen (amygdalin and prunasin hydrolases, mandelonitrile lyase, UDPG-glucosyl transferase) resulted in the first, though not completed yet, apricot sequences identified for these genes. All of them gave optimum BLAST results with the annotated sequences in the NCBI and with the whole peach genome.

On these apricot sequences several functional markers were developed, either SSR or CAPS. Seven of them were positioned on the genetic maps of the parental lines Lito and BO81604311 (in particular four in L and three in B) and eight on the maps of Harcot and Reale di Imola (five in H and three in R). The markers that were mappable on more than one cultivar show a perfect correlation: MDL was found in the top of both L1 and B1, the CAPS developed by cutting PH with *Rsa*I were located in the upper part of B6 and H6 (and the cut of AH with the same endonuclease was found in the same position in R6A); PH cut with *Tru1*I was mapped in the bottom region of L7 and R7; the microsatellite developed on AH was located in the middle of the linkage group 7 of every cultivar. All the map positions determined have found correspondence with the peach genome.

The use of qNMR to measure the amygdalin content of the seeds allow to quantify even the traces detectable in some sweet H×R kernels and showed that amygdalin is the only cyanogenic glycoside present in mature apricot seeds. Plotting the amygdalin content measured for the L×B population among the three harvesting years showed a bi-modal trend with slightly shifted peaks depending on the year. The distribution shape was also found to be different within the years, especially the 2008 one resulted more closed around the mean value, whereas in 2007 and 2009 more extreme values (toward a higher content and toward a lower one, respectively) were present. This environmental effect, clear confirmation of the quantitative character of the bitterness trait in apricot, was statistically found to be more dependent from the rainfall than from the temperatures. This inverse correlation with the precipitations may be a corroboration of the CNGs' role of storage compounds and, at the same time, an effect of the light-sensitivity of the first two enzymes of their anabolic pathway: CYP79 and CYP71, cytochrome dependent monooxigenases.

The qNMR experiments were done on the seeds divided longitudinally in quarters and the amygdalin content was found to be variable among the quarters of the same seeds. Thus, a set of Raman imaging experiments have been set up, in order to get some insights on the localization of this compound in the seed. The results of this *in situ* analysis seem to demonstrate that the presence itself of the amygdalin is more important than its precise localization, being this not always the same in every analyzed seed. Moreover, the fact that the amygdalin distribution proceeds toward lower levels as moving to the centre of the seed was not found to be the majority of the cases. This is a further support to the hypothesis of C and N storage compounds as major role for CNGs in seeds.

The QTL analysis, done with the data coming from the qNMR phenotyping, allowed to identify several putative QTLs in the four parental lines analyzed. Four were spotted in Lito (in L1, L3, L4 and L6) and two in BO81604311 (in B1 and B6) for what concerns the L×B population, whereas, regarding the H×R, one was found in Harcot (H5) and two in Reale di Imola (R4B and R5). The Harcot contribution for the QTL on linkage group 5 is higher than the one of Reale di Imola. The loci identified in the bitter×bitter population have never been described before, because, being minor QTLs, namely the ones responsible for the bitterness level and not for the bitter/sweet phenotype distinction, they are likely to be masked in population which segregate also for the sweetness. Concerning the QTLs identified in HxR, also the QTL in R4 has never been depicted before this work, while the major QTL found in

the LG5 of Harcot and Reale di Imola, has already been reported as the region mainly controlling the sweet/bitter trait in almond and peach.

In conclusion, this thesis has revealed some loci involved in the shaping of the bitterness degree; has proven the complexity of the bitter trait in apricot, reporting an high variance of the amygdalin content found among the years and a correlation with the environmental parameter rainfall; has showed the critical importance of the phenotyping step, whose precision and accuracy is a pre-requisite when studying such a multifactorial character; has given new insights for the *in situ* localization of the amygdalin.

Future perspectives would be the completion of the sequencing of the apricot candidate genes chosen, the saturation of the Harcot×Reale map and the analysis of its seeds among more years, the development of new functional markers linked to other isoforms of the studied GCs and the analysis of the two cytochrome dependent monooxigenases responsible for the first two steps of the anabolic pathway.

References

Almela, F.R., 2006. Fisiologia della maturazione e qualità delle albicocche. Rivista di frutticoltura e di ortofrutticoltura. LXVIII(6), 22–24.

Aranzana, M.J., Pineda, A., Cosson, P., Dirlewanger, E., Ascasibar, J., Cipriani, G., Ryder, C.D., Testolin, R., Abbott, A., King, G.J., Iezzoni, A.F., Arús, P. 2003. A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. Theor. Appl. Genet. 106, 819–825.

Audergon, J.M., Ruiz, D., Bachellez, A., Ferréol, A.M., Lambert, P., Pascal, T., Poëssel, J.L., Signoret, V., Quilot, B., Boudehri, K., Renaud C., Dirlewanger, E., Dondini, L., Deborde, C., Maucourt, M., Moing A., Monllor, S., Gouble, B., Grotte, M., Bogé, M., Reiling, P., Reich M., Bureau, S., Arús, P., 2009. ISAFRUIT – study of the genetic basis of *Prunus* fruit quality in two peach and two apricot populations. Proceedings of the 12th Eucarpia Symposium on Fruit Breeding and Genetics, Zaragoza, Spain, 16-20 September 2007., International Society for Horticultural Science (ISHS) 523–528.

Bak, S., Kahn, R.A., Nielsen, H.L., Møller , B.L., Halkier, B.A., 1998. Cloning of three A type cytochromes P450, CYP71E1, CYP98, and CYP99 from Sorghum bicolour (L.) Moench by a PCR approach and identification by expression in Escherichia coli of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant Mol. Biol. 36, 393–405.

Bak, S., Olsen, C.E., Halkier, B.A., Møller, B.L., 2000. Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. Plant Physiol., 123, 1437–48.

Bak, S., Paquette, S.M., Morant, M., Rasmussen, A.V., Saito, S., Bjarnholt, N., Zagrobelny, M., Jørgensen, K., Hamann, T., Osmani, S., 2006. Cyanogenic glycosides: a case study for evolution and application of cytochromes p450. Phytochem. Rev. 5, 309–329.

Banea-Mayambu, J. P., Tylleskar, T., Tylleskar, K.,Gebre-Medhin, M., Rosling, H., 2000. Dietary cyanide from insufficiently processed cassava and growth retardation in children in the Democratic Republic of Congo (formerly Zaire). Ann. Trop. Paediatr. 20, 34–40.

Bartolozzi, F., Bertazza, G., Bassi, D., Cristoferi, G., 1997. Simultaneous determination of soluble sugars and organic acids as their trimethylsilyl derivatives in apricot fruits by gas-liquid chromatography. J. Chromatogr. A, 758, 99–107.

Bassi, D., Negri, P., 1991. Ripening date and fruit traits in apricot progenies. Acta Hort. 293, 133–140.

Bassi, D., Rizzo, M., 2004. Albicocco: si seleziona per il tardivo. Rivista di frutticoltura e di ortofrutticoltura. LXVI(6), 30–35.

Béguin, P., 1990. Molecular biology of cellulose degradation. Annu. Rev. Microbiol. 44, 219-248.

Bimboim, H.C., Doly, J., 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7(6), 1513–1523

Bliss, F.A., Arulsekar, S., Foolad, M.R., Becerra, A.M., Gillen, A., Warburton, M.L., Dandekar, A.M., Kocsisne, G.M., Mydin, K.K., 2002. An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. Genome 45, 520–529.

Blumenthal, S.G., Hendrickson, H.R., Abrol, Y.P., Conn, E.E., 1968. Cyanide metabolism in higher plants III. The biosynthesis of β-cyanoalanine. J. Biol. Chem. 243, 5302–5307.

Bordo, D., Bork, P., 2002. The rhodanese/Cdc25 phosphatase superfamily – sequence-structure-function relations. EMBO Reports 3, 741–746.

Boyd, F. T., Aamodt, O. S., Bohstedt, G., Truog, F., 1938. Sudan grass management for control of cyanide poisoning. J. Am. Soc. Agron. 30, 569–582.

Brattsten, L.B., Samuelian, J.H., Long, K.Y., Kincaid, S.A., Evans, C.K., 1983. Cyanide as a feeding stimulant for the southern armyworm, Spodoptera eridania. Ecol. Entomol. 8, 125–132.

Busk, P.K., Møller, B.L., 2002. Dhurrin synthesis in sorghum is regulated at the transcriptional level and induced by nitrogen fertilization in older plants. Plant Physiol. 129, 1222–1231.

Caetano-Anollés, G., Gresshof, P.M., 1994. Staining nucleic acids with silver: an alternative to radioisotopic and fluorescent labelling. Promega Notes Magazine 45, 13–20.

Callesen, O., 2009. ISAFRUIT: the total chain approach. J. Hort. Sci. Biotech. ISAFRUIT Supplement 1.

Cardoso, A. P., Ernesto, M., Cliff, J., Egan, S.V., Bradbury, J.H., 1998. Cyanogenic potential of cassava flour: Field trial in Mozambique of a simple kit. Int. J. Food Sci. Nutr. 49, 93–99.

Causse, M., Duffe, P., Gomez, M.C., Buret, M., Damidaux, R., Zamir, D., Gur, A., Chevalier, C., Lemaire-Chamley, M., Rothan, C., 2004. A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J. Exp. Bot. 55, 1671–85.

Cock, J. 1982. Cassava: A basic energy source in the tropics. Science 218, 755–762.

Conn, E.E., 1980. Cyanogenic compounds. Annu. Rev. Plant. Physiol. 31, 433–451.

Conn, E.E., 1981. Cyanogenic glycosides. In: E. E. Conn (Ed.), The Biochemistry of Plants. A Comprehensive Treatise, Vol. 7, Secondary Plant Products. Academic Press, New York, 479–500.

Cooper-Driver, G.., Finch, S., Swain, T., 1977. Seasonal variation in secondary plant compounds in relation to the palatability of *Pteridium aquilinum*. Biochem. Syst. Ecol. 5, 177–183.

Corkill, L., 1942. Cyanogenesis in white clover. 5. The inheritance of cyanogenesis. N. Z. J. Sci. Technol. (Ser. B) 23, 178–193.

Crush, J. R., Caradus, J. R., 1995. Cyanogenesis potential and iodine concentration in white clover (Trifolium repens L.) cultivars. N. Z. J. Agric. Res. 38, 309–316.

Davis, R.H., Nahrstedt, A., 1985. Cyanogenesis in insects. In: Kerkut, G.A., Gilbert, L.I. (Eds.), Comprehensive Insect Physiology, Biochemistry and Pharmacology. Pergamon Press, Oxford, 635–654.

Davis, R.H., 1991. Cyanogens. In: J.P.F. D'Mello, C.M. Duffus, J.H. Duffus (Eds.), Toxic Substances in Crop Plants. Royal Society of Chemistry, Cambridge, UK, 202–225.

del Campo, G., Berregi, I., Caracena, R., Santos, J., 2005. Quantitative analysis of malic and citric acids in fruit juices using proton nuclear magnetic resonance spectroscopy. Anal. Chim. Acta 556, 462–468.

Dicenta, F., Garcìa, J.E., 1993. Inheritance of the kernel flavour in almond. Heredity 70, 308–312.

Dicenta, F., Martìnez-Gòmez, P., Ortega, E., Duval, H., 2000. Cultivar pollinizer does not affect almond flavour. Hort. Sci. 35(6), 1153–1154.

Dicenta, F., Martìnez-Gòmez, P., Granè, N., Martìn, M.L., Leòn, A., Cànovas, J.A., Berenguer, V., 2002. Relationship between cyanogenic compounds in kernels, leaves, and roots of sweet and bitter kernelled almonds. J. Agric. Food Chem. 50, 2149–2152.

Dicenta, F., Ortega, E., Martinez-Gomez, P., 2007. Use of recessive homozygous genotypes to assess the genetic control of kernel bitterness in almond. Euphytica 153, 221–225.

Dirlewanger, E., Moing, A., Rothan, C., Svanella, L., Pronier, V., Guye, A., Plomion, C., Monet, R., 1999. Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch). Theor. Appl. Genet. 98, 18–31.

Dirlewanger, E., Cosson, P., Howad, W., Capdeville, G., Bosselut, N., Claverie, M., Voisin, R., Poizat, C., Lafargue, B., Baron, O., 2004. Microsatellite genetic linkage maps of myrobalan plum and an almond-peach hybrid – location of root-knot nematode resistance genes. Theor. Appl. Genet. 109(4), 827–838.

Dirlewanger, E., Cosson, P., Renaud, C., Monet, R., Poëssel, J.L., Moing, A., 2006. New detection of QTLs controlling major fruit quality components in peach. Acta Hort. (ISHS) 713, 65–72.

Dondini, L., Pierantoni, L., Gaiotti, F., Chiodini, R., Tartarini, S., Bazzi, C., Sansavini, S., 2004. Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. Mol. Breed. 14, 407–418.

Dondini, L., Lain, O., Geuna, F., Banfi, R., Gaiotti, F., Tartarini, S., Bassi, D., Testolin, R., 2007. Development of a new SSR based linkage map in apricot and analysis of synteny with existing Prunus maps. Tree Genet. Genomes 3, 239–249.

Dondini, L., Lain, O., Vendramin, V., Rizzo, M., Vivoli, D., Adami, M., Guidarelli, M., Gaiotti, F., Palmisano, F., Buzzoni, A., Boscia, D., Geuna, F., Tartarini, S., Negri, P., Castellano, M., Savino, V., Bassi, D., Testolin, R., 2011. Identification of QTL for resistance to plum pox virus strains M and D in Lito and Harcot apricot cultivars. Mol. Breeding, 27(3) 289–299.

Esen, A., 1993. β-glucosidases - overview. ACS. Symp. Series 533, 1–14.

Etienne, C., Rothan, C., Moing, A., Plomion, C., Bodénès, C., Svanella-Dumas, L., Cosson, P., Pronier, V., Monet, R., Dirlewanger, E., 2002. Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. Theor. Appl. Genet. 105, 145–159.

Femenia, A., Rossello, C., Mulet, A., Canellas, J., 1995. Chemical composition of bitter and sweet apricot kernels. J. Agric. Food Chem. 43(2), 356–361.

Finnemore, H., Reichard, S. K., Large, D. K., 1935. Cyanogenetic glucosides in Australian plants. Part 3. *Eucalyptus cladocalyx*. J. Proc. R. Soc. N. S. W. 69, 209–214.

Foulongne, M., Pascal, T., Arùs, P., Kervella, J., 2003. The potential of *Prunus* davidiana for introgressione into peach [*Prunus persica* (L.) Batsch] assessed by comparative mapping. Theor. Appl. Genet. 107, 227–238.

Franks, T.K., Abbas, Y., Wirthensohn, M.G., Guerin, J.R., Kaiser, B.N., Sedgley, M., Ford, C.M., 2008. A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. Funct. Plant Biol. 35, 236–246.

Frehner, M., Scalet, M., Conn, E.E., 1990. Pattern of the cyanide-potential in developing fruits. Plant Physiol. 94, 28–34.

Geuna, F., Toschi, M., Bassi, D., 2003. The use of AFLP markers for cultivar identification in apricot. Plant Breed. 122, 526–531.

Girona, J., Marsal, J., 1995. Estrategias de RDC en almendro. In: M. Zapata, P. Segura (Eds.), Riego Deficitario Controlado. Fundamentos y Aplicaciones. Mundi Prensa, Madrid, 99–118.

Gleadow, R.M., Woodrow, I.E., 2000. Temporal and spatial variation in cyanogenic glycosides in *Eucalyptus cladocalyx*. Tree Physiol. 20, 591–598.

Gleadow, R.M., Woodrow, I.E., 2002. Mini-Review: constraints on effectiveness of cyanogenic glycosides in herbivore defense. J. Chem. Ecol. 28(7), 1301–1313.

Gòmez, E., Burgos, L., Soriano, C., Marin, J., 1998. Amygdalin content in the seeds of several apricot cultivars. J. Sci. Food Agric. 77, 184–186.

Grimplet, J., Romieu, C., Audergon, J.-M., Marty, I., Albagnac, G., Lambert, P., Bouchet, J.-P., Terrier, N., 2005. Transcriptomic study of apricot fruit (*Prunus armeniaca*) ripening among 13.006 expressed sequence tags. Physiol. Plantarum, 125, 281–292.

Hansen K.S., Kristensen C., Tattersall D.B., Jones P.R., Olsen C.E., Bak S., Møller B.L., 2003. The in vitro substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. Phytochemistry 64,143–151.

Harborne, J. B., 1982. Introduction to Ecological Biochemistry. Academic Press, New York.

Haughn, G., Chaudhury, A., 2005. Genetic analysis of seed coat development in *Arabidopsis*. Trends Plant Sci. 10, 472–477.

Heppner, J., 1923. The factor for bitterness in the sweet almond. Genetics 8, 390–392.

Heppner, J., 1926. Further evidence on the factor for bitterness in the sweet almond. Genetics 11, 605–606.

Hickel, A., Hasslacher, M., Griengl, H., 1996. Hydroxynitrile lyases: functions and properties. Physiol. Plantarum 98, 891–898.

Hopkins, A., 1995. Factors influencing cattle bracken-poisoning in Great Britain, In: R. Thornton Smith and J. A. Taylor (Eds.), Bracken: An Environmental Issue. Bracken 94 Conference, Aberystwyth, Wales, July 1994. International Bracken Group Special Publication No. 2, 120–123.

Hösel, W., Conn, E.E., 1982. The aglycone specificity of plant ß-glycosidases. Trends Biochem. Sci. 7(6), 219–221.

Hösel, W., Tober, I., Eklund, S.H., Conn, E.E., 1987. Characterization of β-glucosidases with high specificity for the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench seedlings. Arch. Biochem. Biophys. 252, 152–162.

Hu, Z., Poulton, J.E., 1997. Sequencing, genomic organization, and preliminary promoter analyses of a black cherry (R)-(+)-mandelonitrile lyase gene. Plant Physiol. 115, 1359–1369.

Hu, Z., Poulton, J.E., 1999. Molecular analysis of (R)-(+)-mandelonitrile lyase microheterogeneity in black cherry. Plant Physiol. 11, 1535–1546.

Hurtado, M.A., Romero, C., Vilanova, S., Abbot, A.G., Llacer, G., Badenes, M.L. 2002. Genetic linkage map of two apricot cultivars (*Prunus armeniaca* L.) and mapping of PPV (sharka) resistance. Theor. Appl. Genet. 106, 819–825.

Hwang, E.-Y., Lee, S.-S., Lee, J.-H., Hong, S.-P., 2002. Development of quantitative extraction method of amygdalin without enzymatic hydrolysis from tonin (Persicae Semen) by high performance liquid chromatography. Arch. Pharm. Res. 25, 453–456.

Illa, E., Eduardo, I., Audergon, J.M., Barale, F., Dirlewanger, E., Li, X., Moing, A., Lambert, P., Dantec, L., Gao, Z., Poëssel, J.-L., Pozzi, C., Rossini, L., Vecchietti, A., Arús, P., Howad, W., 2010. Saturating the *Prunus* (stone fruits) genome with candidate genes for fruit quality. Mol. Breed. Doi: 10.1007/s11032-010-9518-x.

Jaroszewski, J.W., Olafsdottir, E.S., Wellendorph, P., Christensen J., Franzyk, H., Somanadhan, B., Budnik, B.A., Jørgensen, L.B., Clausen, V., 2002. Cyanohydrin glycosides of Passiflora: distribution pattern, a saturated cyclopentane derivative from *P. guatemalensis*, and formation of pseudocyanogenic alpha-hydroxyamidesas isolation artefacts. Phytochemistry 59, 501–511.

Jenrich, R., Trompetter, I., Bak, S., Olsen, C.E., Møller, B.L., Piotrowski, M., 2007. Evolution of heteromeric nitrilase complexes in Poaceae with new functions in nitrile metabolism. PNAS 104, 18848–18853.

Jensen, K., Jensen, P.E., Møller, B.L., 2011. Light-Driven Cytochrome P450 Hydroxylations. ACS chemical biology, in press.

Jones, P.R., Møller, B.L., Höj, P.B., 1999. The UDP-glucose: phydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* - isolation, cloning, heterologous expression, and substrate specificity. J. Biol. Chem. 274, 35483–35491.

Jones, P.R., Andersen, M.D., Nielsen, J.S., Höj, P.B., Møller, B.L., 2000. The biosynthesis, degradation, transport and possible function of cyanogenic glucosides. In: Romero, J.T., Ibrahim, R., Varin, L., De Luca, V. (Eds.), Evolution of Metabolic Pathways. Elsevier Science, New York, 191–247.

Joo, W.-S., Jeong, J.-S., Kim, H., Lee, Y.-M., Lee, J.-H., Hong, S.-P., 2006. Prevention of epimerization and quantitative determination of amygdalin in Armeniacae Semen with Schizandrae Fructus solution. Arch. Pharm. Res. 29, 1096–101.

Joobeur, T., Viruel, M.A., De Vicente, M.C., Jàuregui, B., Ballester, J., Dettori, M.T., Verde, I., Truco, M.J., Messeguer, R., Battle, I., 1998. Construction of a saturated linkage map for *Prunus* using an almond × peach F2 progeny. Theor. Appl. Genet. 97, 1034–1041.

Jørgensen, K., Bak, S., Busk, P.K., Sørensen, C., Olsen, C.E., Puonti-Kaerlas, J., Møller, B.L., 2005. Cassava Plants with a Depleted Cyanogenic Glucoside Content in Leaves and Tubers. Distribution of Cyanogenic Glucosides, Their Site of Synthesis and Transport, and Blockage of the Biosynthesis by RNA Interference Technology. Plant Physiol. 139, 363–374.

Kahn, R.A., Bak, S., Svendsen, I., Halkier, B.A., Møller, B.L., 1997. Isolation and reconstitution of cytochrome P450ox and in vitro reconstitution of the entire biosynthetic pathway of the cyanogenic glucoside dhurrin from sorghum. Plant Physiol. 115, 1661–1670.

Kahn, R.A., Fahrendorf, T., Halkier, B.A., Møller, B.L., 1999. Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. Arch. Biochem. Biophys. 363(1), 9–18.

Kester, D.E., Asay, R.N., 1975. Almonds. In: Janick J., Moore J.N. (Eds.), Advances in fruit breeding. Purdue University Press, West Lafayette, IN 387–419.

Kostina, K.F., 1977. Breeding apricot in the southern zone of the USSR. Sadovodstvo 7, 24–25.

Krafft, C., Knetschke, T., Siegner, A., Funk, R.H.W., Salzer, R., 2003. Mapping of single cells by near infrared Raman microspectroscopy. Vib. Spectrosc. 32, 75–83.

Krafft, C., Diderhoshan, M.A., Recknagel, P., Miljkovic, M., Bauer, M., Popp, J., 2011. Crisp and soft multivariate methods visualize individual cell nuclei in Raman images of liver tissue sections. Vib. Spectrosc. 55, 90–100.

Kriechbaumer, V., Park, W.J., Piotrowski, M., Meeley, R.B., Gierl, A., Glawischnig, E., 2007. Maize nitrilases have a dual role in auxin homeostasis and β-cyanoalanine hydrolysis. J. Exp. Bot. 58, 4225–4233.

Kudelski, A., 2008. Analytical applications of Raman spectroscopy. Talanta 76, 1–8.

Ladizinsky, G., 1999. On the origin of almond. Genet. Resour. Crop Ev. 46, 143–147.

Lambert, P., Hagen, L.S., Arùs, P., Audergon, J.M., 2004. Genetic linkage maps of two apricot cultivars (*Prunus armeniaca* L.) compared with the almond Texas × peach Earlygold reference map for *Prunus*. Theor. Appl. Genet. 108, 1120–1130.

Lambert, P., Dicenta, F., Rubio, M., Audergon, J.M., 2007. QTL analysis of resistance to sharka disease in the apricot (*Prunus armeniaca* L.) Polonais×Stark Early Orange F1 progeny. Tree Genet. Genomes 4(4), 299–309.

Le Dantec, L., Cardinet, G., Bonet, J., Fouché, M., Boudehri, K., Monfort, A., Poëssel, J.-L., Moing, A., Dirlewanger, E., 2010. Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. Tree Genet. Genomes 6, 995–1012.

Lechtenberg, M., Nahrstedt, A., 1999. Cyanogenic glycosides. In: R. Ikan, (Ed.), Naturally Occurring Glycosides. John Wiley and Sons, Chichester, UK, 147–191.

Lv, W.-F., Ding, M.-Y., Zheng, R., 2005. Isolation and quantitation of amygdalin in Apricot kernel and *Prunus tomentosa* Thunb. by HPLC with solid-phase extraction. J. Chromatogr. Sci. 43, 383–387.

Malz, F., Jancke, H., 2005. Validation of quantitative NMR. J. Pharmaceut. Biomed. 38, 813–823.

Masia, A., Cabrini L., 1994. Determinazione dei glucosidi cianogenici in semi di drupacee. Atti II Giornate Scientifiche S.O.I., 1069, 215–216.

Mellano, M. G., Valentini, N., Rolle, L., Zeppa, G., Botta, R., 2006. Impiego della "consumer science" per la valutazione qualitativa di albicocche prodotte in Piemonte. Rivista di frutticoltura e di ortofrutticoltura. LXVIII(6), 54–57.

Mercado, J.A., El Mansouri, I., Jimenez-Bermudez, S., Plieco-Alfaro, F., Quesada, M., 1999. A convenient protocol for extraction and purification of DNA from *Fragaria*. In Vitro Cell. Dev-Pl. 35, 152–153.

Micklander, E., Brimer, L., Engelsen, S.B., 2002. Noninvasive Assay for Cyanogenic Constituents in Plants by Raman Spectroscopy: Content and Distribution of Amygdalin in Bitter Almond (*Prunus amygdalus*). Appl. Spectrosc. 56, 1139–1146.

Miller, J.M., Conn, E.E., 1980. Metabolism of hydrogen cyanide by higher plants. Plant Physiol. 65, 1199–1202.

Miller, R.E., Gleadow, R.M., Woodrow, I.E., 2004. Cyanogenesis in tropical *Prunus turneriana*: characterisation, variation and response to low light. Funct. Plant Biol. 31, 491–503.

Moing, A., Maucourt, A., Renaud, C., Gaudillere, M., Brouquisse, R., Lebouteiller, B., Gousset-Dupont, A., Vidal, J., Granot, D., Denoyes-Rothan, B., Lerceteau-Köhler, E., Rolin, D., 2004. Quantitative metabolic profiling by 1-dimensional 1H-NMR analyses: application to plant genetics and functional genomics. Funct. Plant Biol. 31(9), 889–902.

Møller, B.L., Poulton, J.E., 1993. Cyanogenic glucosides. In: Lea, P.J. (Ed.), Methods in Plant Biochemistry, vol. 9. Academic Press, San Diego, 183–207.

Møller, B.L., Seigler, D.S., 1999. Biosynthesis of cyanogenic glucosides and related compounds, in: Singh B.K. (Ed.), Plant amino acids. Marcel Dekker, New York, 563–609.

Møller, B.L., 2010. Functional diversifications of cyanogenic glucosides. Current Opin. Plant. Biol. 13, 338–347.

Morgante, M., Salamini F., 2003. From plant genomics to breedeing practice. Curr. Opin. Biotechnol. 14, 214–219.

Nahrstedt, A., 1985. Cyanogenic compounds as protecting agents for organisms. Plant Syst. Evol. 150, 35–47.

Nanda, S., Kato, Y., Asano, Y., 2005. A new (R)-hydroxynitrile lyase from *Prunus mume*: asymmetric synthesis of cyanohydrins. Tetrahedron 61, 10908–10916.

Negri, P., Bassi, D., Magnanini, E., Rizzo, M., Bartolozzi, F., 2008. Bitterness inheritance in apricot (*P. armeniaca* L.) seeds. Tree Genet. Genomes 4, 767–776.

Nout, M.J.R., Tunçel, G., Brimer, L., 1995. Microbial degradation of amygdalin of bitter apricot seeds (*Prunus armeniaca*). Int. J. Food Microbiol. 24, 407–412.

Osbourn, A.E., 1996. Preformed antimicrobial compounds and plant defence against fungal attack. Plant Cell 8,1821–1831.

Paquette, S.M., Møller, B.L., Bak, S., 2003. On the origin of family 1 plant glycosyltransferases. Phytochemistry 62, 399–413.

Paran, I., Zamir, D., 2003. Quantitative traits in plants: beyond the QTL. Trends Genet. 19(6), 303–6.

Pauli, G.F., Jaki, B.U., Lankin, D.C., 2005. Quantitative 1H NMR: development and potential of a method for natural products analysis. J. Nat. Prod. 68, 133–149.

Pauli, G.F., Jaki, B.U., Lankin, D.C., 2007. A routine experimental protocol for qHNMR illustrated with taxol. J. Nat. Prod. 70, 589–595.

Pereira, G.E., Gaudillere, J-P, Van Leeuwen, C., Hilbert, G., Maucourt, M., Deborde, C., ,Moing, A., Rolin, D., 2006. ^{1}H NRM metabolite fingerprints of grape berry: comparison of vintage and soil effects in bordeaux grapevine growing areas. Anal. Chim. Acta 563 (1-2), 346–352.

Pflieger, S., Lefebvre, V., Causse, M., 2001. The candidate gene approach in plant genetics: a review. Mol. Breeding 7, 275–291.

Pierantoni, L., Cho, K-H., Shin, I-S., Chiodini, R., Tartarini, S., Dondini, L., Kang, S-J., Sansavini, S. 2004. Characterization and transferability of apple SSR to two European pear F1 populations. Theor. Appl. Genet. 109, 1519–1524.

Poulton, J.E., 1983. Cyanogenic compounds in plants and their toxic effects. In: R.F. Keeler, A.T. Tu (Eds.), Handbook of Natural Toxins, Vol. 1: Plant and Fungal Toxins. Marcel Dekker, New York, 117–157.

Poulton, J.E., 1988. Localization and catabolism of cyanogenic glycosides. In: D. Evered and S. Harnett (Eds.), Cyanide Compounds in Biology. John Wiley & Sons, Chichester, United Kingdom. 67–91.

Poulton, J.E., 1990. Cyanogenesis in plants. Plant Physiol. 94, 401–405.

Poulton, J.E., 1993. Enzymology of cyanogenesis in rosaceous stone fruits. In: A. Esen (Ed.), Symposium Series 533. American Chemical Society, Washington, DC, 170–190.

Poulton, J.E., Li C.P., 1994. Tissue level compartmentation of (R)-amygdalin and amygdalin hydrolase prevents large-scale cyanogenesis in undamaged *Prunus* seeds. Plant Physiol. 104, 29–35.

Pratt, A., 1937. The Call of the Koala. Robertson & Mullens, Melbourne.

Quilot, B., Wu, B. H., Cervella, J., Génard, M., Foulongne, M., Moreau, K., 2004. QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. Theor. Appl. Genet. 109, 884–897.

Reitzenstein, S., Rösch, P., Strehle, M. A., Berg, D., Baranska, M., Schulz, H., Rudloff, E., Popp, J., 2007. Nondestructive analysis of single rapeseeds by means of Raman spectroscopy. J. Raman Spectrosc. 38, 301–308.

Ressler C., Tatake, J.G., 2001. Vicianin, prunasin, and beta-cyanoalanine in common vetch seed as sources of urinary thiocyanate in the rat. J. Agr. Food Chem. 49, 5075–5080.

Robinson, M. E., 1930. Cyanogenesis in plants. Biol. Rev. 5, 126–142.

Ruiz, D., Lambert, P., Audergon, J.M., Dondini, L., Tartarini, S., Adami, M., Gennari, F., Cervellati, C., De Franceschi, P., Sansavini, S., Bureau, S., Gouble, B., Reich, M., Renard, C.M.G.C., Bassi, D., Testolin, R., 2010. Identification of QTLs for fruit quality traits in apricot. Acta Hort. 862, 587–592.

Sánchez-Pérez, R., Howad, W., Dicenta, F., Arùs, P., Martìnez-Gòmez, P., 2007. Mapping major genes and quantitative trait loci controlling agronomic traits in almond. Plant Breeding 126, 310–318.

Sánchez-Pérez, R., Jørgensen, K., Olsen C.E., Dicenta F., Møller B.L., 2008. Bitternes in almond. Plant Physiol. 146, 1040–1052.

Sánchez-Pérez, R., Howad, W., Garcia-Mas, J., Arús, P., Martínez-Gómez, P., Dicenta, F., 2010. Molecular markers for kernel bitterness in almond. Tree Genet. Genomes 6(2), 237–245.

Saucy, F., Studer, J., Aerni, V., Schneiter, B., 1999. Preference for acyanogenic white clover (*Trifolium repens*) in the vole *Arvicola terrestris*: I. Experiments with two varieties. J. Chem. Ecol. 25,1441–1454.

Schieber, A., Stintzing, F.C., Carle, R., 2001. By-products of plant food processing as a source of functional compounds – recent developments. Trends Food Sci. Tech. 12(11), 401–413.

Schröder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., De Luca, V., Schröder, J., 1999. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. FEBS letters 458, 97–102.

Selmar, D., Lieberei, R., Biehl, B., 1988. Mobilization and utilization of cyanogenic glycosides, the linustatin pathway. Plant Physiol. 86, 711–716.

Sibbesen, O., Koch, B., Halkier, B.A., Møller, B.L., 1994. Isolation of the heme-thiolate enzyme cytochrome P-450TYR, which catalyzes the committed step in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. PNAS 91, 9740–9744.

Sibbesen, O., Koch, B., Halkier, B.A., Møller, B.L., 1995. Cytochrome P-450TYR is a multifunctional hemethiolate enzyme catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. J. Biol. Chem. 270, 3506–3511.

Siegler, D.S., Brinker, A.M., 1993. Characterisation of cyanogenic glycosides, cyanolipids, nitroglycosides, organic nitro compounds and nitrile glycosides from plants. In: Dey P.M., Harborne J.B. (Eds.), Methods of plant biochemistry, alkaloids and sulfur compounds. Academic, New York, 51–93.

Sinnott, M.L., 1990. Catalytic mechanisms of glycosyl transfer. Chem. Rev. 90, 1171–1202.

Swain, E., Li, C.P., Poulton, J.E., 1992a. Development of the potential for cyanogenesis in maturing black cherry (*Prunus serotina* Ehrh.) fruits. Plant Physiol 98, 1423–1428.

Swain, E., Li, C.P., Poulton, J.E., 1992b. Tissue and subcellular localization of enzymes catabolizing (R)-amygdalin in mature *Prunus serotina* seeds. Plant Physiol 100, 291–300.

Swain, E., Poulton, J.E., 1994. Utilization of amygdalin during seedling development of *Prunus serotina*. Plant Physiol. 106, 437–445.

Terskikh, V.V., Allan Feurtado, J., Borchardt, S., Giblin, M., Abrams, S.R., Kermode, A.R., 2005. In vivo ^{13}C NMR metabolite profiling: potential for understanding and assessing conifer seed quality. J. Exp. Bot. 56(418), 2253–2265.

Thygesen, L.G., Løkkey, M.M., Micklander, E., Engelsen, S.B., 2003. Vibrational microspectroscopy of food. Raman vs. FT-IR. Trends Food Sci. Tech. 14, 50–57.

Thygesen, L.G., Jørgensen, K., Møller, B.L., Engelsen, S.B., 2004. Raman spectroscopic analysis of cyanogenic glucosides in plants: development of a flow injection surface-enhanced Raman scatter (FI-SERS) method for determination of cyanide. Appl. Spectrosc. 58, 212–217.

VanEtten, H.D., Mansfield, J.W., Bailey, J.A., Farmer, E.E., 1994. Two classes of plant antibiotics: phytoalexins versus "phytoanticipins". Plant Cell 6, 1191–1192.

Vetter, J., 2000. Plant cyanogenic glycosides. Toxicon 38, 11–36.

Vilanova, S., Romero, C., Abbot, A.G., Llacer, G., Badenes, M.L., 2003. An apricot (*Prunus armeniaca* L.) F2 progeny linkage map based on SSR and AFLP markers, mapping plum pox virus resistance and self-incompatibility traits. Theor. Appl. Genet. 107, 239–247.

Vogt, T., Jones, P., 2000. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci. 5, 380–386.

Wang, D., Karle, R., Iezzoni, A.F., 2000. QTL analysis of flower and fruit traits in sour cherry. Theor. Appl. Genet. 100, 535–544.

Webber, J. J., Roycroft, C. R., Callinan, J. D., 1985. Cyanide poisoning of goats from sugar gums (*Eucalyptus cladocalyx*). Aust. Vet. J. 62, 28.

Weeden, N.F., 1983. Evolution of plant isozymes. In: S.D. Tanksley, T.J. Orton (Eds.), Developments in Plant Genetics and Breeding: Isozymes in Plant Genetics and Breeding, Part A. Elsevier, New York 175–205.

Wetzel, D.L., LeVine, S.M., 1999. Imaging Molecular Chemistry with Infrared Microscopy Science 285, 1224–1225.

Wohler, F., Liebig, J., 1837. Über die Bildung des Bittermandelöls. Annu. Chem. 22, 1–24.

Xu, L.L., Singh, B.K., Conn, E.E., 1986. Purification and characterization of mandelonitrile lyase from *Prunus lyonii*. Arch. Biochem. Biophys. 250, 322–328.

Xu, Z., Escamilla-Treviño, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.-L., Shih, M.-C., Poulton, J.E., 2004. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. Plant Mol. Biol. 55(39), 343–367.

Zagrobelny, M., Bak, S., Rasmussen, A.V., Jørgensen, B., Naumann, C.M., Møller, B.L., 2004. Cyanogenic glucosides and plant-insect interactions. Phytochemistry 65, 293–306.

Zentek, J., 1997. Case report – bloat and diarrhoea in calves. [German] Deutsch. Tier. Wochenschr. 104,153–155.

Zheng L., Poulton J. E., 1995. Temporal and spatial expression of amygdalin hydrolase and (R)-(+)-mandelonitrile lyase in black cherry seeds. Plant Physiol 109,31–39.

Zhou, J., Hartmann, S., Shepherd, B.K., Poulton, J.E., 2002. Investigation of the microheterogeneity and aglycone specificity-conferring residues of black cherry prunasin hydrolases. Plant Physiol., 129, 1252–1264.