SPSD II

STUDYING APPLE BIODIVERSITY: OPPORTUNITIES FOR CONSERVATION AND SUSTAINABLE USE OF GENETIC RESOURCES (APPLE)

W. KEULEMANS, I. ROLDAN-RUIZ , M. LATEUR

ATMOSPHERE AND CLIMATE

MARINE ECOSYSTEMS AND BIODIVERSITY

TERRESTRIAL ECOSYSTEMS AND BIODIVERSITY

NORTH SEA

ANTARCTICA

BIODIVERSITY

PART 2

GLOBAL CHANGE, ECOSYSTEMS AND BIODIVERSITY

GENETIC RESOURCES (APPLE)

SCIENTIFIC SUPPORT PLAN FOR A SUSTAINABLE DEVELOPMENT POLICY
(SPSD II)

Part 2:
Global change, Ecosystems and Biodiversity

FINAL REPORT

Studying apple biodiversity: opportunities for conservation and sustainable use of genetic resources (APPLE)

EV/28

Prof. Dr. Ir. Wannes Keulemans
Laboratory for Fruit Breeding and Biotechnology (LFBB)
and Fruitteeltcentrum (FTC)
Katholieke Unviersiteit Leuven (KULeuven)

Prof. Dr. Isabel Roldán-Ruiz
Institute for Agricultural and Fisheries Research Plant Unit (ILVO-PLANT)
Melle

Dr. Ir. Marc Lateur
Centre Wallon de Recherches Agronomiques (CRA-W)
Gembloux

Subcontractors
Dr. Ir. Kristine Vander Mijnsbrugge,
Research Institute for Nature and Forest (INBO)
Geraardsbergen

Dr. Ir. Dominique Jacques
Centre de Recherche de la Nature, des Forêts et du Bois (CRNFB)
Gembloux

Mr. Ludo Royen
Nationale Boomgaarden Stichting (NBS)
Hasselt

June 2006

BELGIAN SCIENCE POLICY

D/2007/1191/36
Published in 2007 by the Belgian Science Policy
Rue de la Science 8
Wetenschapsstraat 8
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11 – Fax: +32 (0)2 230 59 12
http://www.belspo.be

Contact person:
Mrs Aline Vander Werf
Secretariat: +32 (0)2 238 36 13

TABLE OF CONTENTS

GENERAL INTRODUCTION

As in other parts of Europe, *Malus sylvestris* has become one of the most endangered tree species in Belgium (Kleinschmitt *et al.* 1998, Maes and Rövekamp, 1999). Forest decline and forest fragmentation have caused a drastic reduction of suitable habitats for this species and today only isolated individuals and small populations are found in nature (Hokanson *et al.* 1998, Kleinschmit *et al.* 1998). Therefore, little time is left to save the last wild apple relicts in Belgium and the set up of a sound conservation programme for the wild apple is urgently needed. As for all biodiversity conservation programs, the aim of this conservation program for wild apple should be to safeguard the evolutionary potential of the species. The extent and distribution of genetic variation within the species are of fundamental importance to its evolutionary potential, as they determine the species' chance for long-term survival. Assessment of the diversity present in Belgian populations of wild apple is therefore of key relevance to develop effective in-situ and/or ex-situ conservation strategies.

However, other human activities impose a more specific threat to the biodiversity and genetic identity of wild apples. Spontaneous hybridisation with cultivated apple varieties (*M. domestica*) is known to occur. Moreover, during inventories in Belgium and Germany, apple trees with intermediate phenotype were found at forest sites, suggesting a hybrid origin between wild and cultivated gene pools (Wagner 1998; Maes and Rovenkamp 1999 and 2000; Vorman and Gebhardt 2000). These factors raised doubts about the wild nature of some of the apple trees present in the Belgian forests. If we take into account that the last relicts of wild apple trees in Belgian forests are often old exemplars (average age is estimated at 150-200 years), apple varieties that were grown in the region by the time that these trees germinated should be taken as reference in a study of wild apple, as these are the varieties that might have been introgressed in wild populations. Studying these old varieties is thus necessary to estimate the extent of past and present hybridisation between wild and cultivated gene pools. At present old apple varieties, which have no longer importance in apple production but which have indisputable cultural-historical value, are maintained in collections. Mainly two organisations focus on the conservation of old regional varieties in Belgium: CRA-W and NBS, both involved in this project.

The conservation of wild apples and old regional varieties is not only important from a nature conservation or cultural-historical point of view but they can also serve as source of novel genes in commercial apple breeding. For example, it can be expected that natural populations, subjected to selection pressures during several

generations in nature, will be enriched for genes involved in stress and disease resistance (Schlosser *et al.* 1991). Therefore, natural genotypes of *M. sylvestris* represent a potential source of as yet unexploited genetic variation that can be used in breeding programs of cultivated apple. Also old regional varieties, present in collections of CRA-W and NBS, might carry useful traits to enrich gene pools of contemporary breeding programs. This factor becomes more relevant if we take into account that apple breeders might be operating within a gene pool of greatly reduced genetic diversity, what can lead to serious levels of inbreeding in the near future (Noiton and Alspach 1996). Therefore, a broadening of the genetic base is needed in sustainable breeding strategies (Robinson *et al.* 2001). However, the rational use of wild resources in breeding programs has to be based on knowledge of their genetic potential, a knowledge that is also of paramount importance for the design of optimal conservation strategies. The interests of conservationists of apple genetic resources and those of apple breeders can be easily combined in a synergic action that is advantageous for both. This synergic action is technically feasible because the species boundaries within the genus *Malus* are not sharp, and the molecular tools developed for one species are in most cases directly useful in other species.

Specific objectives of the project were to:

1. Study the genetic diversity present in wild apple populations and old regional varieties at different levels of organisation: neutral and functional nuclear diversity and cytoplasmic diversity
2. Analyse past and recent hybridisation processes between wild and cultivated apple gene pools and determine the present degree of distinctness between wild apples and cultivated varieties
3. Asses the viability of wild apple populations through the analysis of demographic parameters and fertility-related traits
4. Implement all the above results into conservation guidelines for wild apple in Belgium (in-situ where possible, ex-situ where necessary).
5. Develop an efficient management plan to conserve the biodiversity present in existing collections of old regional varieties in Belgium. In this way, the project helped to coordinate the efforts for the conservation of Belgian regional apple varieties at CRA-W and NBS.
6. Disseminate these conservation strategies to a wide audience, including forestry and nature conservation agencies, and non-governmental organisations involved in conservation of wild species.
7. Determine the value of wild apple trees and regional varieties to expand the genetic base of current apple breeding programmes (i.e. introgression of new sources of stress and disease resistance)

PARTNERS AND THEIR CONTRIBUTION TO THE PROJECT

The basis for this corporation was the shared interest in the conservation and sustainable use of apple genetic resources. The partnership brings together foresters, breeders and molecular biologists from Flanders and Wallonia, that are experts in complementary fields, all essential to develop sound conservation guidelines for the endangered wild apple and collections of old cultivars in Belgium.
The consortium consisted of 5 partners:

Laboratory for Fruit Breeding and –Biotechnology (LFBB), KUL-FTC

LFBB and Fruitteeltcentrum KULeuven are reference groups in fruit breeding research, mostly focused on apples. They concentrate on the genetic control of plant processes such as disease resistance, fruit set and fruit ripening, vegetative and reproductive development of fruit trees and also on the diversity of major genes involved in these processes. Lobke Vanwynsberghe is a PhD student of Wannes Keulemans, working on genetic diversity in *M. domestica*. In this project she was involved in (i) microsatellite fingerprinting of *Malus* genotypes in collaboration with ILVO-PLANT and data analysis, (ii) observations on NBS accessions, (iii) scab infections on subsets of the populations used in the project. Rozemarijn Dreesen is a post doc senior researcher, specialised in gene expression and in molecular genetic studies of SI genes.

Instituut voor Landbouw en Visserijonderzoek - Eenheid Plant, ILVO-PLANT

Els Coart is a post-doc working in the research group led by Isabel Roldán-Ruiz, who is senior scientist. Their main research interests are (agro)biodiversity, the relationship between crops and their wild relatives and the genome organisation of disease resistance and reproduction-related genes in different plant species.
They coordinated all scientific reports (including this final report) and were involved in several work packages. Their main tasks within the project were (i) microsatellite fingerprinting of *Malus* genotypes in collaboration with KUL-FTC and data analysis for the wild species and (ii) the study of hybridisation through analyses of microsatellites and chloroplast diversity.

Centre Wallon de Recherches Agronomiques, CRA-W

Marc Lateur is a senior scientist working in the department of Biological control and plant genetic resources. His main interests are (1) the collection and the management of fruit tree biodiversity; (2) the evaluation and valorization of these genetic resources for various uses and (3) the breeding of new apple varieties with a high level of durable disease resistance intended for commercial fruit production. Bernard Watillon was in charge of the genetic analysis using S-SAP and REMAP markers and Adriana Antofi developed the database.

CRA-W was mainly involved in work packages concerning (i) the collection and the description of the material of *M. domestica* and *M. sylvestris*, (ii) the development of a conservation strategy for *M. domestica* in Belgium, (iii) database design and (iv) S-SAP and REMAP fingerprinting.

CHAPTER 1	Coordinator	**CRA-W**
	Other partners involved	**KUL-FTC, CRNFB, NBS**

COLLECTION AND PHENOTYPIC DESCRIPTION OF MATERIAL

PUBLICATIONS

Jacques D., Lateur M., Watillon B., Lemaire S., Coart E., Roldan Ruiz I.,Vander Mijnsbrugge K., Vanwijnsberghe L.& Keulemans W. [2003]. Développement d'un programme de gestion de la diversité génétique du pommier sauvage (*Malus sylvestris* Mill.) en Belgique : application en Région Wallonne. *Les Naturalistes belges*, 84, 2-3-4 : 149-161.

INTRODUCTION

In this Chapter the inventory, collection and the methodology used to describe the plant materials studied in the project are outlined. In addition, for *M. sylvestris* estimates of heritability for some of the characters analysed are presented and the distribution and abundance of this species across Belgium are discussed. Methodologies for the description of *M. domestica* cultivars in collections are discussed. The main structure of the database designed for the storage and management of *Malus* data is also described.

1.1. INVENTORY AND COLLECTION OF PLANT MATERIAL

1.1.1. WILD BELGIAN *MALUS SYLVESTRIS* GENOTYPES

The inventory and collection work took much more time and effort than initially planned, especially in the Walloon Region. In Wallonia, the *M. sylvestris* prospecting was based on a partial inventory of the Regional Forest Service (DNF) initiated in 1999. This list was completed with information coming from several other sources like the Regional Forest Inventory, data available on sites of high biological interest and specific investigations. The final objective was to cover as much as possible all Forest Districts of Wallonia. This allowed us to evaluate the total variability of the species in the region. In the Flemish Region, the same principles were applied.

As shown in Table 1.1, a total of 977 trees of putative *M. sylvestris* were located in the wild and positioned on 1/25000 maps and with GPS coordinates. All trees were physically labelled with durable labels and information has been circulated in order to protect these trees *in situ*. From this total amount, 764 trees are located in Wallonia and 213 in Flanders. For many accessions, pictures (Figure 1.1) were taken.

Table 1.1: Final results of the Belgian survey of Malus sylvestris trees; W: Wallonia; Fl: Flanders

N°	Forest district or location	Region	# trees discovered and labelled	# genotypes grafted (2003-2005)	% grafted genotypes
1	Meerdaal forest	Fl	182	182	100.0
2	Heverlee forest	Fl	13	13	100.0
3	Kortenberg	Fl	1	1	100.0
4	Aarschot	Fl	1	1	100.0
5	St Katelijne Waver	Fl	3	3	100.0
6	Arendonk	Fl	1	1	100.0
7	Vorselaar	Fl	1	1	100.0
8	Lier	Fl	2	2	100.0
9	Bilzen	Fl	1	1	100.0
10	Heuvelland	Fl	1	1	100.0
11	Torhout	Fl	7	7	100.0
	Sub-total F	Fl	213	213	100.0

Table 1.1 (continued)

N°	Forest district or location	Region	# trees discovered and labelled	# genotypes grafted (2003-2005)	% grafted genotypes
1	***Arlon***	W	6	6	100.0
2	***Aywaille***	W	28	16	57.1
3	***Beauraing***	W	15	15	100.0
4	***Bertrix***	W	2	2	100.0
5	***Bièvre***	W	15	12	80.0
6	***Bouillon***	W	17	13	76.5
7	***Bullange***	W	19	10	52.6
8	***Chimay***	W	23	16	69.6
9	***Couvin***	W	26	12	46.2
10	***Dinant***	W	26	14	53.8
11	***Elsenborn***	W	22	8	36.4
12	***Eupen 1***	W	3	2	66.7
13	***Eupen 2***	W	13	10	76.9
14	***Florenville***	W	43	12	27.9
15	***Habay-la-Neuve***	W	10	10	100.0
16	***La Roche***	W	40	16	40.0
17	***Libin***	W	11	10	90.9
18	***Liège***	W	9	6	66.7
19	***Malmédy***	W	2	0	0.0
20	***Marche***	W	37	15	40.5
21	***Mons***	W	4	2	50.0
22	***Namur***	W	19	13	68.4
23	***Nassogne***	W	19	11	57.9
24	***Neufchateau***	W	66	16	24.2
25	***Paliseul***	W	18	12	66.7
26	***Philippeville***	W	17	14	82.4
27	***Rochefort***	W	14	13	92.9
28	***Saint-Hubert***	W	2	2	100.0
29	***Saint-Vith***	W	33	10	30.3
30	***Spa***	W	16	14	87.5
31	***Thuin***	W	34	17	50.0
32	***Verviers***	W	21	14	66.7
33	***Vielsalm***	W	41	15	36.6
34	***Viroinval***	W	45	14	31.1
35	***Virton***	W	28	14	50.0
36	***Wellin***	W	20	14	70.0
	Sub-total W	***W***	***764***	***400***	***52.4***
	TOTAL		***977***	***613***	***62.7***

At the same time, all characterisation and evaluation data were collected. From the 977 trees identified, bud wood was been collected on a selection of 400 and 213 in Walloon Region and Flemish Region respectively in order to multiply them in our nurseries. Due to the large amount of *M. sylvestris* trees found in the Walloon region, a selection had to be made in view of collecting the maximum diversity of the species and to collect a representative sample of material originating from all forest districts. Table 1.1 shows that in total 62% of the

genotypes have been selected and grafted for *ex situ* conservation. A total of 3,616 and 1,000 graft operations have been realised in Walloon Region and Flemish Region respectively, in order to obtain five trees per genotype. During the prospection, leaves of selected trees were collected for the study of genetic diversity.

1.1.2. BELGIAN *M. DOMESTICA* CULTIVARS

From the NBS and the CRA-W collections, a selection of 256 and 248 old cultivars respectively, were chosen for phenotypic and genetic diversity studies. Cultivars were chosen based on their Belgian origin or former common presence in Belgium. Furthermore, 104 "modern" cultivars were included from the breeding collection of KUL-FTC. The phenotypic evaluations of these trees were carried out from 2003 to 2005.

1.1.3. COLLECTION OF OTHER WILD GENOTYPES AND *M. DOMESTICA* CULTIVARS

To further enlarge the sample of cultivated apple genotypes analysed and create a broader reference collection, we also included in the genotypic analysis 31 old Danish cultivars and a representative sample of 30 cider apple cultivars originating from Great Britain and France. Also for *M. sylvestris*, additional materials from Denmark, France, Germany and Central Asia were analysed. An additional 77 samples of different *Malus* spp. genotypes were obtained from the Collection of USDA, Geneva, USA.

Table 1.2 gives an overview of the materials included in the project. We achieved largely our initial goals, as a total of 1,498 accessions of diverse *Malus* spp. was characterised. In some instances it was not possible to collect as many trees as planned e.g. for wild material from the Flemish region (no more wild genotypes were found) and we could not get *Malus* spp. materials from Dresden Pillnitz, due to an epidemic of fire blight and of Apple Proliferation phytoplasm.

Table 1.2: Overview of the samples included in the study

Species	Aim total project	Done in 2003	Done in 2004	Done in 2005	Total achieved
M. domestica					
Old cvs (CRA-W)	250	256	0	0	256
Cider cvs	30	0	30	0	30
Old cvs (NBS)	250	240	8	0	248
Modern cvs (KUL-FTC)	100	104	0	0	104
Old cvs (DK)	0	0	31	0	31
Sub total M. domestica	***630***	***600***	***69***	***0***	***669***
M. sylvestris					
Wallony	400	335	65	0	400
Flanders	300	147	90	0	237
Europe (G, FR, DK)	40	43	91	0	134
Central Asia (Genebank Geneva)	0	12	0	0	12
Sub total M. sylvestris	***740***	***1247***	***507***	***193***	***783***
Wild species					
M. baccata (KUL-FTC)	10	6	0	0	6
M. mandshurica (KUL-FTC)	10	1	0	0	1
M. orientalis (KUL-FTC)	10	2	0	0	2
M. prunifolia (KUL-FTC)	10	7	0	0	7
M. sieversii (KUL-FTC)	10	4	1	0	5
M. floribunda (CRA-W)	10	0	0	0	0
M. praecox (CRA-W)	10	0	0	0	0
M. sieboldii (CRA-W)	10	0	0	0	0
M. dasyphyllus (CRA-W)	10	0	0	0	0
M. florentina (KUL-FTC)	2	1	0	0	1
M. fusca (KUL-FTC)	2	0	0	0	0
M. hupehensis (KUL-FTC)	2	1	0	0	1
M. kansiensis (KUL-FTC)	2	0	0	0	0
M. kirghisorum (KUL-FTC)	2	0	0	0	0
M. komarovii (KUL-FTC)	2	0	0	0	0
M. transitoria (KUL-FTC)	2	0	0	0	0
M. x. adstringens (KUL-FTC)	2	1	0	0	1
M. zhaojiaoensis (KUL-FTC)	2	2	0	0	2
M. toringoides	0	3	0	0	3
M. pumila	0	1	0	0	1
Sub total other Malus species	***108***	***29***	***1***	***0***	***30***
TOTAL	***1478***	***1876***	***577***	***193***	***1482***

1.2. DEFINITION OF STANDARD DESCRIPTORS FOR CHARACTERISATION AND EVALUATION OF THE MATERIAL

1.2.1. *MALUS SYLVESTRIS*: DESCRIPTORS FOR PASSPORT DATA, ECOLOGICAL DATA AND PHENOTYPIC DATA

A total of 64 descriptors were defined in coordination between CRNFB and CRA-W and used to characterise the *M. sylvestris* trees chosen throughout Belgium. The geographical coordinates of each studied tree were also recorded and the locations were described using rough ecological descriptors. All data were stored in a central database. These data were used to evaluate the intraspecific variability and to prepare a conservation programme to be applied in Belgium. Furthermore, standards for taking pictures of fruits and pips were defined and followed. Data analysis concentrated on eleven descriptors (Table 1.3) for which sufficient data were available either for *in situ* or *ex situ* evaluations. In addition, bud burst phenology was recorded in nursery at one location in order to test the hypothesis of the existence of adapted "ecotypes". Other characteristics mainly concern morphology, physiology and susceptibility to diseases (Table 1.3).

Table 1.3: Most important descriptors used for data analysis

Descriptors	***Unit***
Height	M
Circumference	Cm
Leaf hairiness	scale (1 to 9)
Twig hairiness	scale (1 to 9)
Fruit diameter	Mm
Skin colour	scale (1 to 6)
Over colour	scale (1 to 6)
Canker	scale (1 to 9)
Leaf scab	scale (1 to 9)
Fruit scab	scale (1 to 9)
Powdery mildew	scale (1 to 9)

M. sylvestris trees were grafted and the ones which succeeded were evaluated during a two-year period (2004 and 2005) under nursery conditions (see 1.2.2). The measurements performed on the grafts concern leaf and twig hairiness, susceptibility to diseases (scab and powdery mildew) and flushing. In addition to a general descriptive analysis, these characteristics were analysed through a one-way analysis of variance to derive genetic parameters. The factor "Clone" used in this model is considered as random. The model is defined as: $Y_{ij} = \mu + C_i + \varepsilon_{ij}$ where:Y_{ij} = j^{th} observation on the i^{th} treatment , (where i = 1,…,k; j=1,…,n_i); n_i = number of replicates in treatment i; μ = population mean, C_i = Clonal effect i, ε_{ij} = residual deviation associated with the ij^{th} observation (ramet in clone and other residual effects).

Genotypic heritability on means per clone (h^2_G) was defined as the part of the global variability observed between genetic elements if each of these genetic elements is kept identical to itself after reproduction (NANSON 1970). Genotypic heritability was therefore derived from the ANOVA analysis using the following formula (NANSON, 1970):
$\mathbf{h^2_G = \sigma^2_C / \sigma^2_P}$, and its estimation is derived from the present model by the following ratio: $\mathbf{h^2_G = (CM_C - CM_r) / CM_C}$ where: h^2_G= broad-sense heritability, CM_C = mean square for the "Clone" factor, CM_r = residual mean square.

1.2.2. *MALUS DOMESTICA*: DESCRIPTORS FOR PHENOTYPIC OBSERVATIONS AND SHORT DESCRIPTION OF THE MATERIAL AND METHODS USED

A common descriptors list was designed and priority descriptors were identified for *M. domestica*, in coordination between CRA-W, KUL-FTC and NBS. The aim was to use descriptors that are in accordance with the 'ECP/GR' (European Cooperative Programme for Plant Genetic Resources Networks) and 'EURISCO' (European Plant Genetic Resources Search Catalogue) frameworks. We therefore used the *Malus* passport descriptor list adopted by the ECP/GR *Malus*/*Pyrus* Working Group which is based on the FAO Multicrop Passport Descriptors. This is a list of 20 descriptors. For evaluation and characterisation, the 'Minimum descriptor list' defined by the ECP/GR *Malus*/Pyrus Working Group was followed. For fruit description and disease assessment, 14 descriptors were selected, two for flower description (intensity and period of blossoming) and one for tree architecture. Some descriptors were adapted or more precisely defined, e.g. figures and numeral references were added to some of them to increase the accuracy of the evaluation. For disease resistance evaluations, we used the descriptors defined for *M. sylvestris* (see above).

1.3. PHENOTYPIC DIFFERENTIATION OF *MALUS SYLVESTRIS* AND *MALUS DOMESTICA*

1.3.1. DIFFERENTIATION BETWEEN WILD AND HYBRID FORMS OF *MALUS SYLVESTRIS*

To evaluate the phenotypic variability and the geographical distribution of *M. sylvestris* in Belgium, the dataset containing information on *in situ* evaluations was cleared of putative hybrid forms using phenotypic information. Fruit size diameter and leaf hairiness, according to Fellenberg (2001) and Wagner (1995a, 1995b, 1996) were used as indication of 'wildness'. Limits of these two characteristics for wild apple were arbitrary fixed to a diameter of 35 mm for the mean fruit size and to a scale of 4 for leaf hairiness. Most accessions did not carry fruits in the field, but when fruits were available, fruit mean diameter was measured and a photo database was developed that currently contains 328 pictures of fruits and 181 pictures of pips (illustrated in Figure 1.2).

Figure 1.2: Photograph of apple and pips

One-hundred out of the 285 trees for which these two characteristics were available appeared to be putative hybrid forms (Figure 1.3). As postulated by Kleinschmit *et al.* (1998) and Stephen *et al.* (2003) for Germany, a strong introgression seems to exist between the wild and cultivated forms in Belgium. On the other hand, the number of hybrids defined on leaf hairiness and fruit diameter is approximately one third of the inventory, which is not in accordance with the results of Coart *et al*. (2003) based on microsatellite markers. The difference in size of the two samples and the covered geographical areas could explain these differences. However, hairiness might not be the best discriminatory characteristic (Coart *et al*. (2003); Larsen *et al.* (2006)).

For fruit diameters corresponding to putative hybrids, a broad range of hairiness was observed (upper part of Figure 1.3) including many hybrids (according to fruit size) with a low degree of hairiness. Fruit diameter appears more efficient as discriminative criterion without being totally efficient. Unfortunately, fructification is not regularly observable in the field and can strongly vary from one year to another which limits the efficiency of such a criterion applied in forests.

According to the classification shown in Figure 1.3, putative hybrids are very common in farm lands (86.7%) and along roads (66.7%), and to a lesser extent also along forest edges (46.2%) or in hedges (41.3%). On the other side, in forests a proportion below 20 percent hybrids (17.9%) was observed.

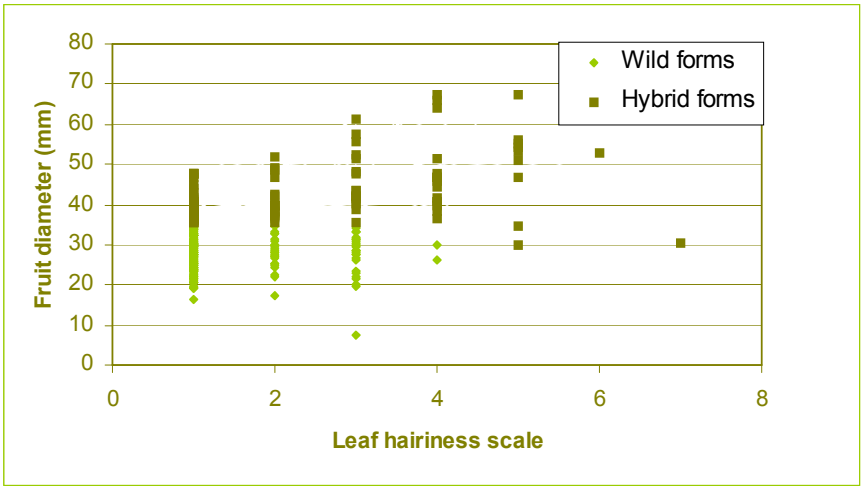


Figure 1.3: Identification of 185 wild apple out of a sample of 285 trees according to a fixed limit for leaf hairiness scale (1: hairless to 9: intense hairiness) and fruit diameter (mm)

1.3.2. GEOGRAPHICAL DISTRIBUTION OF *MALUS SYLVESTRIS* IN BELGIUM

Wild apple is still common in the southern part of Belgium but mixed with hybrids spread everywhere in this region. On the contrary, it appears to be very rare north of the Sambre and Meuse axis. Except in a very limited number of sites as Meerdaal Forest, where a large population still survives (more than 150 individuals), wild apple confirms its status of scattered species as only 30 % of trees are neighboured by other wild apple trees.

High densities of wild apples are observed mainly in two regions: Ardenne and Low Plateau of Meuse (Figure 1.4). The high cover of forests in these two regions, respectively over 50 % and between 15 and 50 % comparatively with less than 15 % the Northern part of the Sambre and Meuse axis (LECOMTE *et al.* 2002), could partly explain this situation. The majority of the wild apples was found in forest (59.5 %) and to a lesser extent in hedges (25.4 %). The rest of the populations were mainly found in forest borders or along roads (Figure 1.5).

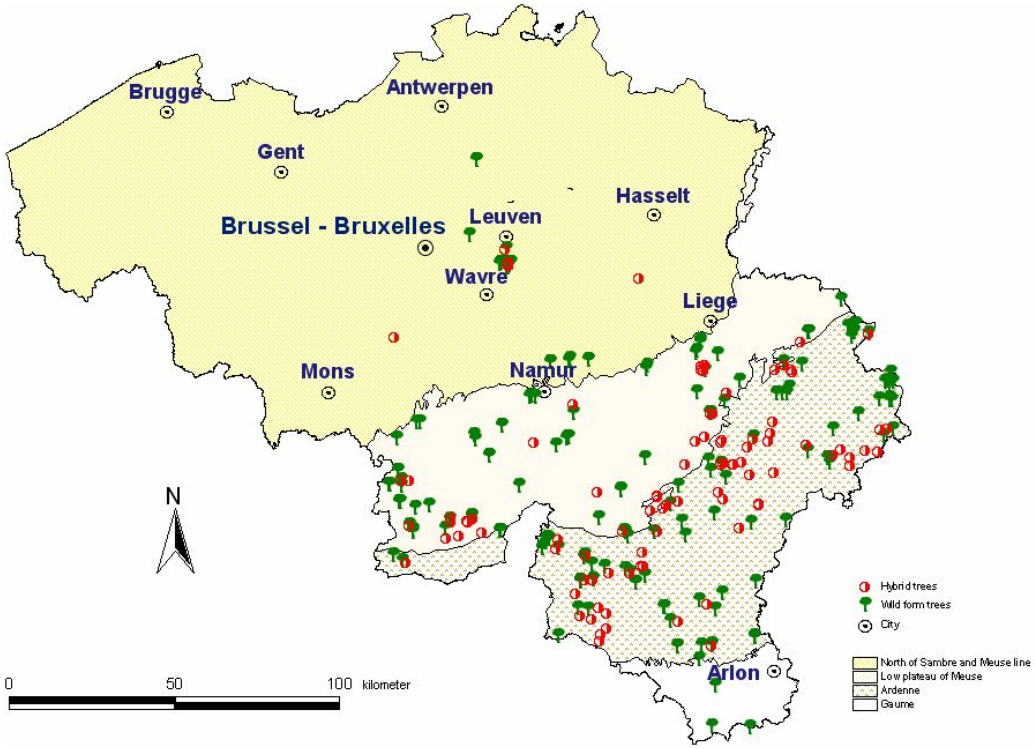


Figure 1.4: Distribution map of the *Malus sylvestris* samples split in wild and hybrid forms for different Belgian regions of provenance (n=185).

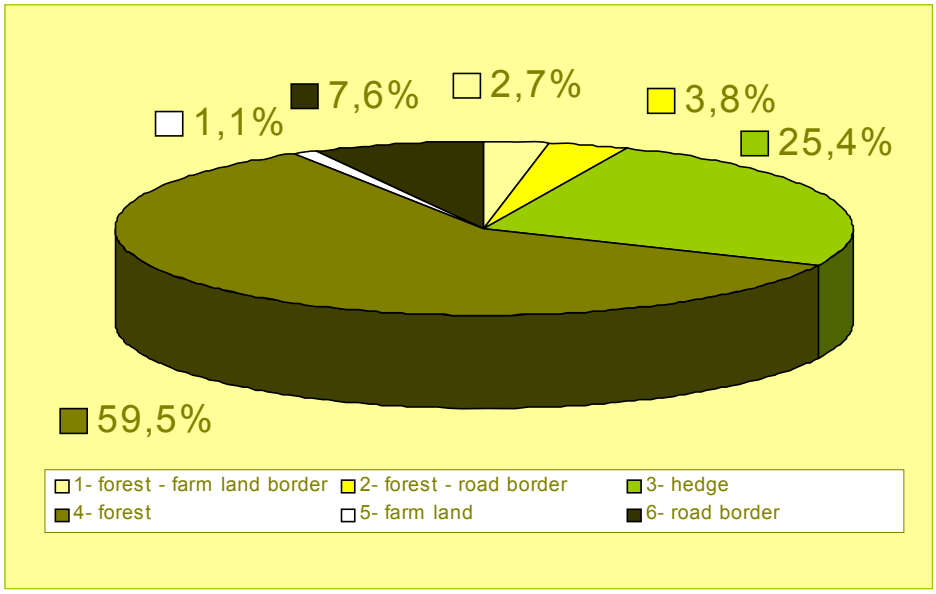


Figure 1.5: Percentage of observed wild apples according to habitat (n = 185).

1.3.3. PHENOTYPIC CHARACTERISATION OF *MALUS SYLVESTRIS IN SITU* AND *EX SITU*

Results of the eleven characteristics that have been evaluated are presented in Table 1.4. These results provide precious information to appreciate the phenotypic variability of the species.

Table 1.4: Descriptive data for 11 characteristics observed on wild apple in the field and in nursery (n=185)

	Unit	Genotype	Mean	Standard deviation	Min	Max
Height	m	180	10.3	3.2	1.5	20.3
Circumference	cm	185	74.6	37.1	13.0	234.0
Leaf hairiness	Scale (1 to 9)	185	1.4	0.8	1	4
Twig hairiness	Scale (1 to 9)	118	3.2	1.6	1	8
Fruit diameter	mm	185	27.7	4.5	7.7	35.3
Skin colour	Scale (1 to 6)	160	2.1	1.1	1	5
Over colour	Scale (1 to 6)	160	1.7	1.2	1	5
Canker	Scale (1 to 9)	165	1.4	1.0	1	5
Leaf scab	Scale (1 to 9)	185	3.6	1.5	1	7
Fruit scab	Scale (1 to 9)	184	3.1	1.8	1	7
Powdery mildew	Scale (1 to 9)	159	1.2	0.8	1	7

Although the species is generally considered as a shrub in reference flora (De Langhe *et al*. 1978; Rameau *et al. 1989*), the observed wild trees reach a mean height of 10.3 meters. Moreover, it is not rare to observe individuals with a total height over 15 meters and a circumference at breast height over 120 cm. Twig and leaf hairiness show a large variability, which is not in accordance with the description of reference floras. One third of the observed trees show a slight pubescence on the lower surface of the leaf blade. A majority of trees also show hairy twigs. On the contrary, fruit diameter and colour fit well with the general description of the species. The observed fruit mean diameter is 28 mm. The most frequent skin colour is pale green with orange over colour.

Wild apple trees do not appear very susceptible to diseases. Cancer and powdery mildew symptoms were very limited but leaf scab infections were slightly more important. Fruit scab (*Venturia inaequalis* Cooke), the most important fungal disease on cultivated apples, was also observed on wild genotypes (Table 1.5). The infection level in the field was intermediate but the variability was very large. As average, scab observed on leaf in nursery was slightly more aggressive on wild forms than on hybrids. This significant difference could be explained by a higher pressure of selection applied for the horticultural needs. Wild forms seem also to be a little less susceptible to powdery mildew but differences were not significant.

Concerning the evaluation of *in situ* fruit scab susceptibility, 80% of the evaluated wild and hybrid trees show a good tolerance and about 30% are free of any scab symptom (Figure 1.6). Nevertheless, compared to the expected normal Gauss curve, the frequency of class 1 was overestimated, probably because a significant proportion of trees escaped a sufficient high scab pressure level for expressing any symptom.

Table1.5: Mean susceptibility of wild and hybrid forms of *M. sylvestris* for different diseases and level of significance of the differences between these two forms

Disease	*Site*	*Wild type*	*Hybrid type*	*p-value*
Cancer	In situ	1.2	1.6	0.013 *
Leaf scab	Nursery	1.9	1.6	0.026 *
Fruit scab	In situ	3.0	3.1	0.767 ns
Powdery mildew	Nursery	1.6	1.9	0.166 ns

*p<0.05; ** p<0.01; ***p<0.001; ns: not significant

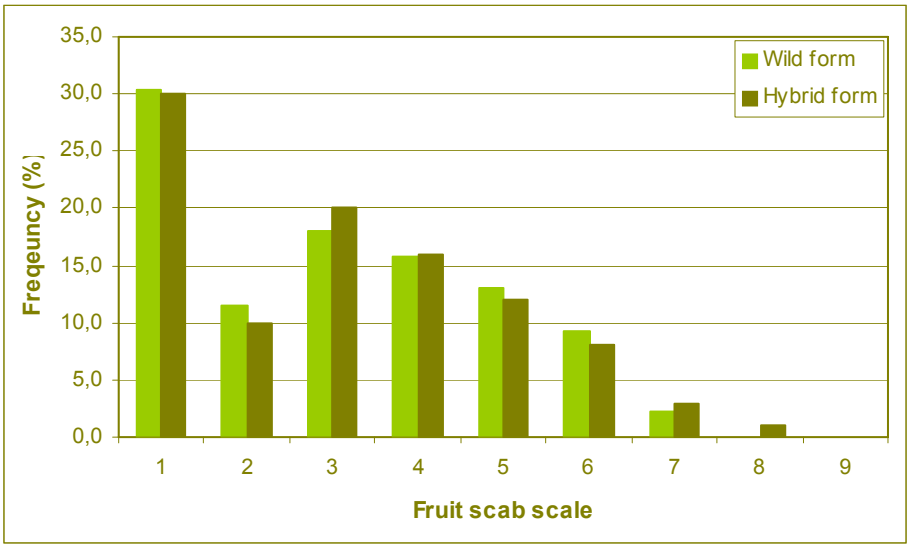


Figure 1.6: Frequency of fruit scab scores on 185 and 100 genotypes of respectively wild and hybrid forms of *M. sylvestris* evaluated in situ. 1: no symptoms; 9: high susceptibility

1.3.4. GENOTYPIC HERITABILITY OF CHARACTERISTICS OBSERVED IN NURSERY FOR *MALUS SYLVESTRIS*

Hairiness, susceptibility to scab and powdery mildew, and flushing were observed and evaluated on grafted material in the nursery during two years and genotypic heritability has been calculated from these observations.

Twig and leaf hairiness appear to be heritable characteristics when observed in nursery (average h= 0.85 and 0.78 respectively over two years). However, when these results were compared with the *in situ* data, large differences were apparent (Figure 1.7). A higher hairiness was observed in the nursery and only a low percentage of trees are free of hairs, compared to the *in situ* observations. This suggests an important influence of the environment on this characteristic like age of tree, position in the tree, age of leaves, light intensity… It confirms once more the risk of using this characteristic to differentiate wild and hybrid forms in the field even if hairiness is a heritable characteristic.

Leaf scab and powdery mildew are also heritable characteristics even though the first results of 2004 were less clear (h= 0.04 and 0.41 for 2004 and 0.53 and 0.73 for 2005 respectively).

The low intensity of scab and powdery mildew symptoms observed in 2004 undoubtedly did not allow a good expression of the real susceptibility on a significant part of the tested clones. This result shows that a large genetic diversity still exist in this wild species and confirms the possibility of selection for horticultural purposes, especially for cancer resistance, as it is the case for the old varieties created before the development of fungicides (Populer *et al*. 1998).

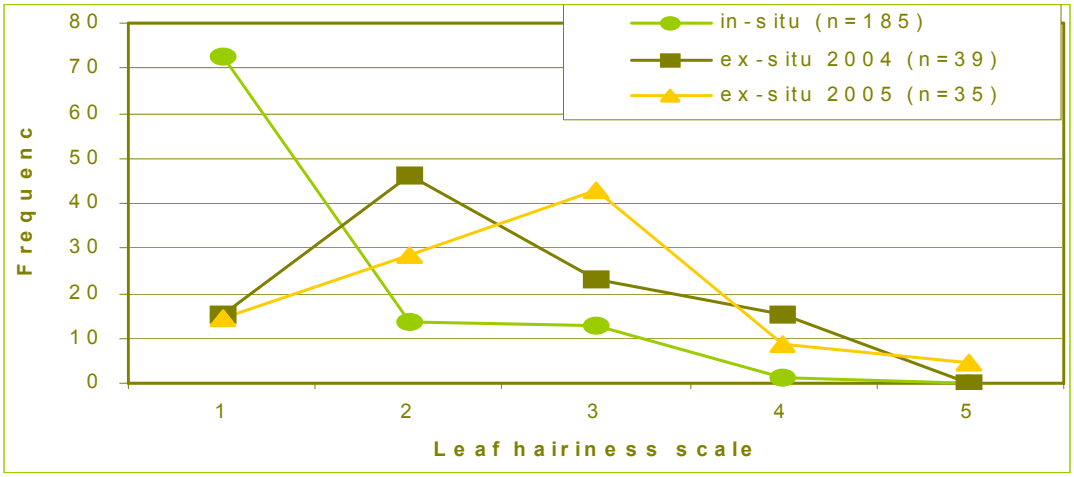


Figure 1.7: Observed frequency of hairiness leaf susceptibility evaluated on clones of *Malus sylvestris* observed in situ and ex situ (2004 and 2005)

Flushing was observed during two years (2004-2005) on respectively 10 and 38 clones. A very high heritability was found for this characteristic over years (0.93 and 0.86 for 2004 and 2005 respectively). This is in accordance with numerous results published for forest trees (Nanson 1971 and 1973, Wuelhlisch *et al*. 1993, Vernier & Teissier Du Cros 1996). Flushing is a particularly interesting characteristic for studying the adaptation to the environment in Western Europe, where late frost is regularly observed in spring and could have an important impact on the survival of young plants. Belgium can be divided into different regions of provenance to organise and to optimise the transfer of seeds and plants of forest species at the country level. These areas are delineated on the basis of climatic conditions (temperature, precipitation, altitude…) which assumes a potential differentiation within the species. Wild apples are well represented in two regions of provenance (Ardenne and the low plateau of Meuse). We compared the characteristics for which a high heritability was found in these two regions using variance analysis (Table 1.6). However, no significant differences were found between regions.

Table 1.6: Comparison of Malus sylvestris populations split into regions of provenance for flushing, susceptibility to powdery mildew and to leaf scab evaluated in 2005

Characteristic	*Region of provenance*		*p-value*
	Low plateau of Meuse	*Ardenne*	
Flushing	4.97	4.62	0.312
Susceptibility to powdery mildew	1.46	1.75	0.316
Susceptibility to leaf scab	2.08	1.79	0.267

1.3.5 CHARACTERIZATION OF APPLE CULTIVARS IN BELGIAN COLLECTIONS

Characterisation and evaluation data of cultivars are useful for (i) identification of the material (ii) studying the phenotypic diversity and (iii) the evaluation of future use of the material. Characterisation data are normally not dependent on the environmental conditions and are mostly "qualitative data". Evaluation data are, on the contrary "quantitative data" that are strongly under the influence of the environment and that vary from year to year. Using the 'common descriptor list' (see 1.1.1), we characterised the cultivated accessions with a total of 16.000 data points collected during the project.

Not all results can be presented in this report due to space limitations but some are illustrated below as case-studies on the development of a good common methodology. A general remark should be made about the difficult comparisons of observations made on cultivars of different origins, in different environments, on different ages of trees and orchards managed under different tree and orchard management methods.

The cultivated accessions were evaluated for the main apple diseases in Belgium: scab (*Venturia inaequalis*), Powdery Mildew (*Podosphaera leucotricha*) and *Nectria* cancer (*Nectria galligena*). Results obtained in three different collections (old cultivars of CRA-W and NBS collections and modern cultivars of the KUL-FTC collection) are shown in Figure 1.8 Results from CRA-W were very different from those coming from KUL-FTC and NBS. The results from NBS are not in accordance with literature reports, in which immunity to scab is not present among old cultivars due to the fact that this resistance is commonly of polygenic origin. The high percentage of immune cultivars at NBS is perhaps due to the difficulty of scab evaluations on leaves on high stem trees. The results obtained by KUL-FTC for modern cultivars are probably due to the fact that the KUL-FTC collection has been evaluated on two-year-old material, in a young orchard where the inoculum is at a low level. In the KUL-FTC material there are also a number of cultivars that possess the *Vf* major gene for scab resistance. Furthermore, on average, above class 3, fruit scab intensity on fruits is lower than on leaves (results not shown).

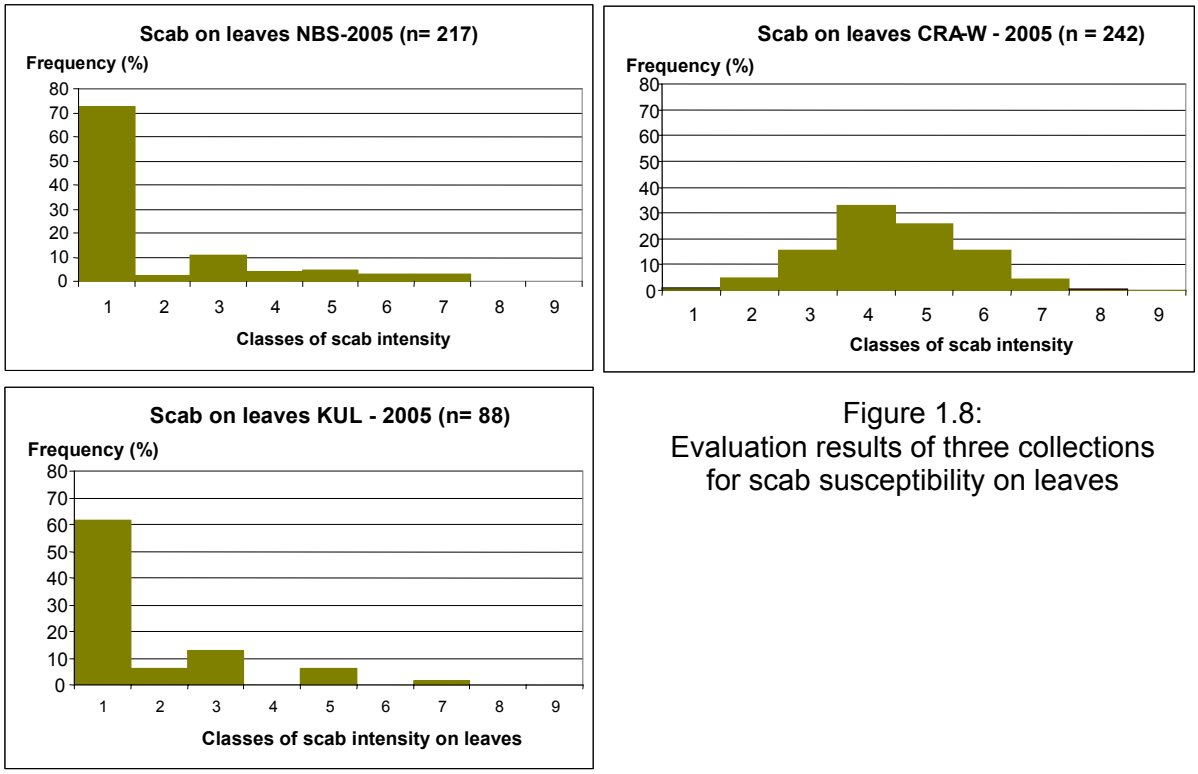


Figure 1.8:
Evaluation results of three collections for scab susceptibility on leaves

Concerning Powdery Mildew, results of KUL-FTC follow the same patterns as scab and correspond to a young orchard with very low infection pressure. NBS and CRA-W results show that there is no source of immunity to Powdery Mildew among old cultivars. For cancer, tendencies between CRA-W and NBS are more or less similar but cancer pressure seems to be at a much higher level in the CRA-W evaluation orchards. On the other hand, high stem trees, as in the NBS orchards, might be more tolerant to cancer (results not shown).

Figure 1.9 shows results obtained for other characters studied. Due to the fact that the size of the three samples was heterogeneous and that for some characters the period of observation was too short, interpretation of the different histograms is not straightforward. For flowering period (first row of panels), the CRA-W results are expressed as an average class of flowering period and the distribution curve is quite normal. The KUL-FTC collection of modern cultivars seems to present a lower diversity for this character. Concerning fruit size the three collections follow the same tendency with the particularity of the KUL-FTC collection to have a larger proportion of fruit ranked in class 6. This illustrates that this collection contains more commercial cultivars. Also the lower phenotypic diversity within this collection can be noted for "crowning at apex fruit end". Furthermore, these preliminary results show that more extended observations over a longer period are absolutely needed and that, even when the same descriptors are used, more training time is needed to harmonize their practical use.

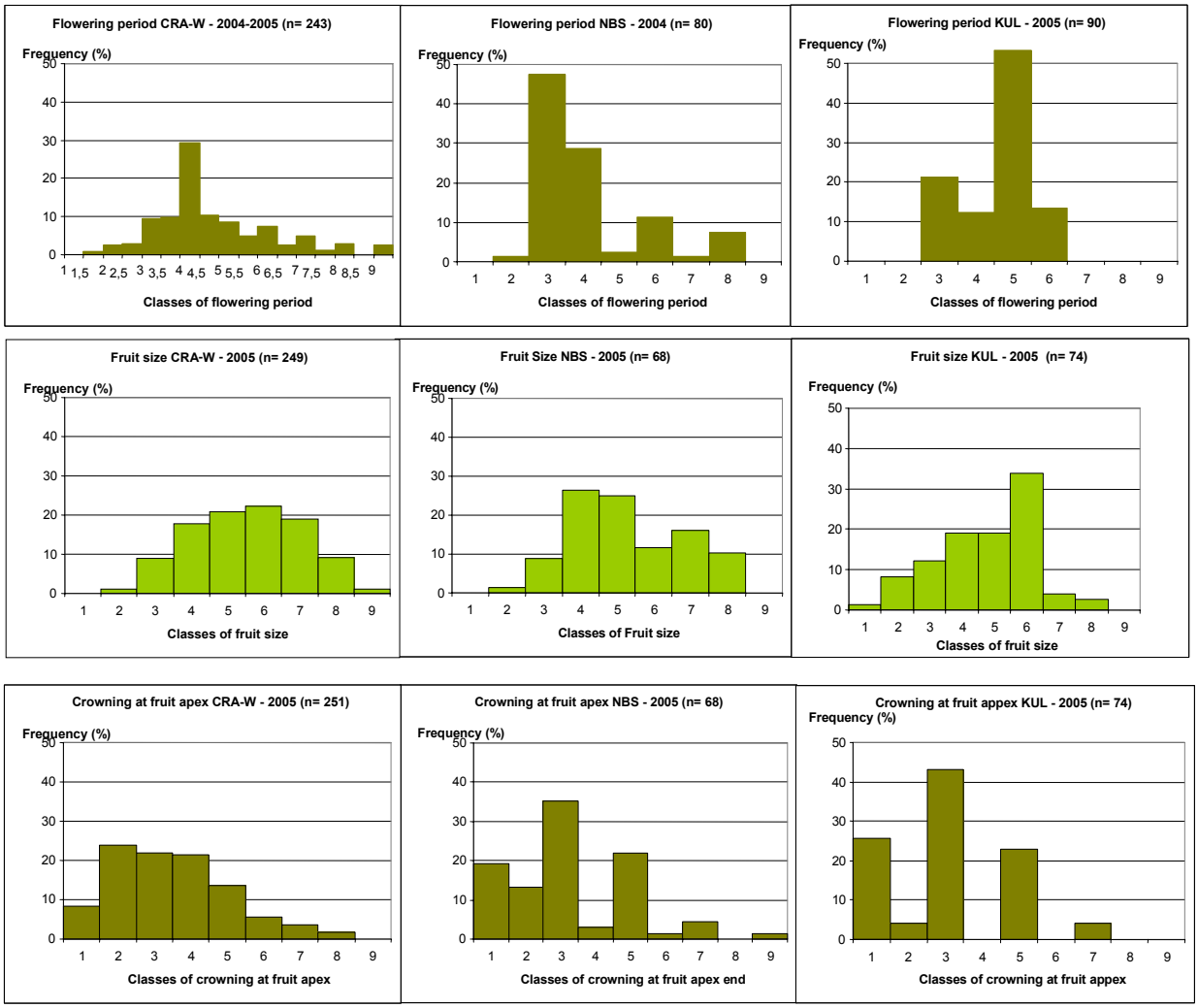


Figure 1.9: Results of traits that have been characterised and evaluated on three samples of *M. domestica*

Based on data collected at CRA-W, the phenotypic diversity was analysed using cluster analysis. Here the "Fastclus" method was used (SAS 9.1, 2002). The FASTCLUS procedure performs a disjoint cluster analysis on the basis of distances computed from one or more quantitative variables. Observations are divided into clusters such that every observation belongs to one and only one cluster; the clusters do not form a tree structure. By default, the FASTCLUS procedure uses Euclidean distances, so the cluster canters are based on least-squares estimations. This method is often called a *k-means model*, since the cluster centres are the means of the observations assigned to each cluster when the algorithm is run to complete convergence. Iteration reduces the least-squares criterion until convergence is achieved. The DISTANCE procedure computes various measures of distance, dissimilarity, or similarity between the observations. These proximity measures are stored as a lower triangular matrix that can then be used as input to the MDS (multidimensional scaling) procedure. MDS was used to create a plot that provides an overview of the different clusters.

The study was concentrated on characters that are polymorphic and assumed to be stable under different environments or otherwise, that could be compared with standard controls

with the same relative ranking. Table 1.7 shows, for eight characters, the mean value observed for each identified cluster and the relative weight of them in the clustering determination. Some characters have a higher discriminative weight compared to others. Amount of russet on fruit, fruit stalk length, crowning at fruit end apex and flowering period are characters with the highest discriminative power, whereas tree architecture seems not to be discriminative.

Table 1.7: Mean values of eight characters for each identified cluster (a) and Table of the discriminate weight of the characters in the clustering identification (b)

(a)

Cluster	Flowering period	Harvest maturity	Eating maturity	Keeping ability	Crowning at apex	Amount russet	Length stalk	Tree architecture
1	7,12	6,38	4,31	3,92	2,31	6,54	2,69	5,46
2	4,13	5,63	3,25	3,29	2,63	1,46	2,29	5,27
3	4,25	6,00	3,44	3,06	2,19	3,25	5,88	4,81
4	2,63	1,63	1,75	1,13	3,38	1,13	3,13	5,50
5	3,73	6,07	4,14	4,07	2,82	7,11	2,93	5,25
6	4,11	5,95	3,63	3,68	4,37	4,74	1,95	5,26
7	4,80	7,47	4,87	4,53	2,07	2,00	2,20	5,53
8	4,05	6,28	4,08	4,26	5,18	1,77	3,44	5,56
9	7,16	6,14	3,73	3,77	4,91	1,55	2,09	6,05

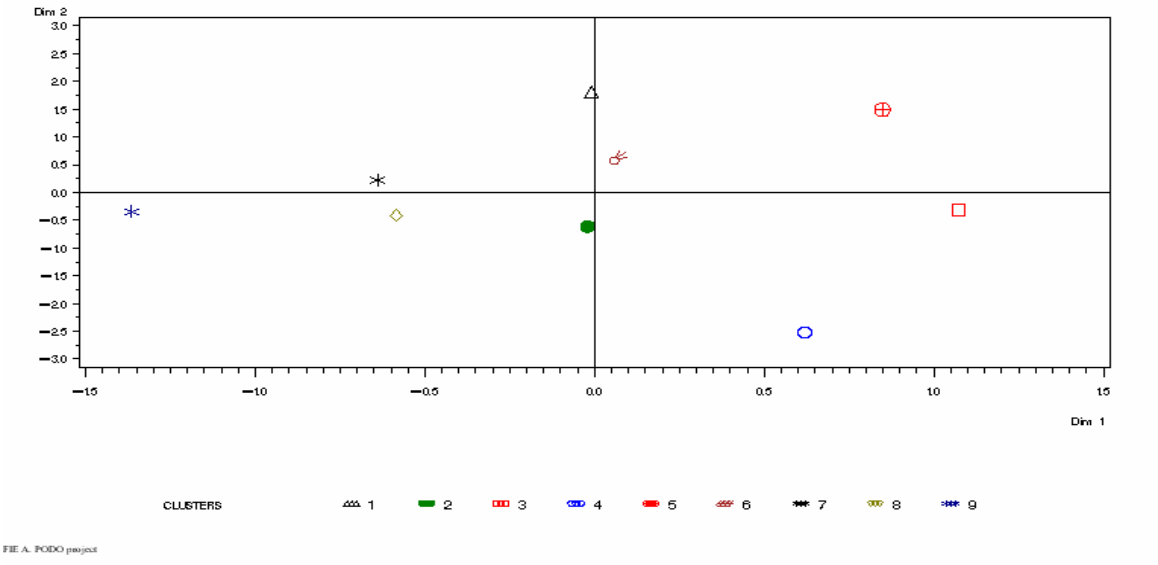
(b)

Cluster	Flo_per	Har_mat	Eat_mat	Keep_ab	Crow_ap	Amo_rus	Leng_sta	Tree_arc
1	++				--		--	
2					--	--	--	
3					--			
4	--	--	--	--		--		
5					--	++	--	
6							--	
7		++			--	--	--	
8						--		
9	++					--	--	

Figure 1.10 shows the inferred clusters. The first figure shows the global repartition of the nine clusters or families of accessions that are identified in this sample of old apple cultivars. Mean distance between the clusters constitute an interesting tool for the expression of the phenotypic diversity (see Chapter 5). Table 1.8 shows the distance between pairs of clusters.

Table 1.8: Distance matrix between inferred clusters

Clusters	1	2	3	4	5	6	7	8	9
1	0,00								
2	6,52	0,00							
3	5,96	4,05	0,00						
4	9,39	4,81	6,17	0,00					
5	3,54	5,87	5,33	8,31	0,00				
6	4,49	3,77	4,53	6,51	3,29	0,00			
7	5,29	2,68	4,37	7,24	5,40	4,10	0,00		
8	6,63	3,07	4,04	6,32	5,84	3,21	3,56	0,00	
9	6,03	4,13	5,84	7,24	6,96	4,19	4,11	3,57	0,00



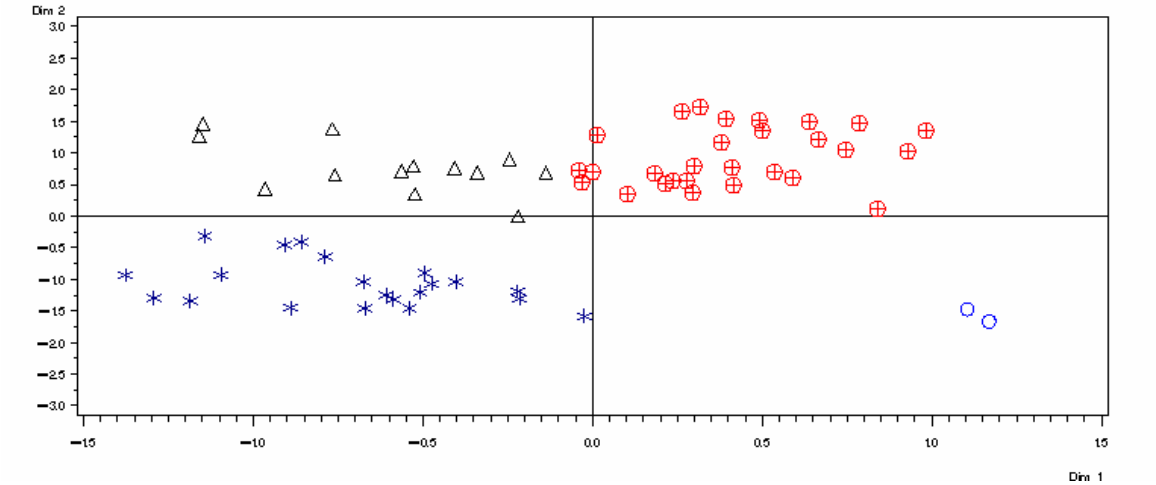


Figure 1.10: Results of the cluster analysis based on eight characters for old cultivars of the CRA-W collection.

1.4. DATABASE

The general aim of the database was to store the data collected by the different partners and to facilitate the sharing of information. The design of a database is made at three levels: the conceptual data model (CDM), the logical data model (LDM) and the physical data model (PhDM). The conceptual data model describes the structure. The logical data model is a detail of the CDM and describes the link existing between all information of the database. The physical data model is the translation of the logical data model into a real structure of the database. On this level each entity defined in the LDM is converted into a table in the PhDM. With the aim to include information from different work packages and after a first analysis concerning the needs of the partners, the database was divided in several groups as illustrated below (Figure 1.11).

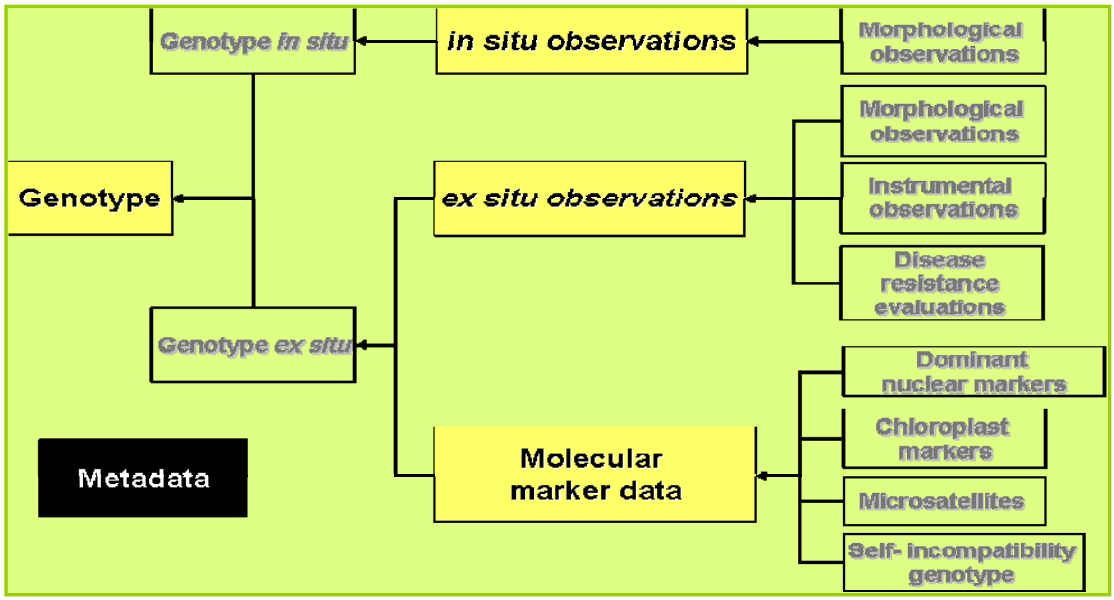


Figure 1.11: General structure and conceptual data model of the "Applebiodiv database"

The general structure is divided into four classes as "genotype", "*in situ* observations", "*ex situ* observations" and "molecular marker data". In the diagram, the classes are designed with a yellow background. A class is a generic entity for the structuring of data which are more or less similar. The "**Genotype**" class includes two subclasses: "*in situ* Genotype" and "*ex situ* Genotype". In the inferior level of the virtual architecture are the subclasses. The subclass is a specialization of the class. The "***In situ* Genotype**" subclass included the genotypes identified and described *in situ*. This subclass makes the link between "**Genotype**" class and "***in situ* observation**" class. The "***In situ* observation**" class includes the subclass "*Morphological observation*" and lists the morphological observations (e.g. size of fruit, leaves hairiness, etc.) made on the genotypes (trees) *in situ.* "**Genotype *ex situ***" subclass includes the genotypes (trees) conserved *ex situ*. This subclass makes the link between "**Genotype**" class and two others classes: "***ex situ* observation**" and "**molecular marker data**". The three subclasses are included into "***ex situ* observation**" class: "*Morphological observations*", "*Disease resistance evaluations*" and "*Instrumental observations*". This class list the observations make on the genotype conserved in the

collections. The class "**Molecular marker data**" includes "*Dominant nuclear marker*", "*Chloroplast marker*", "*Microsatellites*" and "*Self-incompatibility genotype*". The last data group are the **metadata** that consist of the structure of each table and the format of each field in the tables. The metadata are managed by the database system itself. In the logical data model level, each subclass defined in the conceptual data model, is converted in one or more entity. Each entity corresponds to a table in the physical model. The genotype is the hard core of the database that makes the link between genotypic and phenotypic information. To collect data from the partners, a template in Excel format was created for each table. Partners sent the completed excel files to the database manager, who transferred the tables to the database. Extraction of data (only for partners) and some general queries can be done using the website.

1.5 CONCLUSIONS AND PERSPECTIVES

In summary, *M. sylvestris* is a disseminated tree in Belgium, still well represented in Wallonia (especially South of the river Meuse) but very rare in Flanders except for the large population (circa 200 trees) in Meerdaal forest. The species is most frequently encountered in high forest. In open landscapes, *M. sylvestris* is commonly found in hedges. A high phenotypic variability was measured *in situ* and *ex situ*: this is the case for tree shape, forking, pubescence of leaves and twigs, disease susceptibility, fructification, fruit form and flushing. This opens the way for a more detailed analysis including comparisons with molecular marker results.

The results obtained for *M. domestica* highlight the major importance of a validation procedure that need to be adopted before analysing the data. Validation includes the reproducibility of the data over years, locations and users. Some characters are more stable than others (characterisation data) but the method of data assessment need to be validated. The evaluation data are under the influence of the environment and therefore data collection procedures need to be repeated during a couple of years, in order to obtain a level at which the polymorphism of the trait is expressed. The results presented here allow the definition of a methodology to study phenotypic diversity. These methods will be further discussed in Chapter 5 (conservation strategies).

CHAPTER 2	Coordinator	**ILVO-PLANT**
	Other partners involved	**KUL-FTC, CRA-W**

CHARACTERISATION OF GENETIC DIVERSITY BY USE OF NEUTRAL DNA MARKERS

PUBLICATIONS

A.S. Larsen , C.B. Asmussen, E. Coart, D.C. Olrik, E.D. Kjær (2006) Hybridization and genetic variation in Danish populations of European crab apple (*Malus sylvestris*). Tree Genetics & Genomes, in press.

INTRODUCTION

In this Chapter levels of genetic diversity in wild apple populations and cultivar collections are presented and compared. Microsatellite (SSR) markers represent a well-established tool in conservation genetics studies and have been successfully applied to characterise apple diversity by the partnership of this project (Coart *et al*. 2003, Kenis *et al*. 2005). Numerous primers for amplification of polymorphic SSR loci in *Malus* have been developed and published to date (e.g. Hokanson *et al*. 1998, Gianfranceschi *et al*. 1998, Maliepaard *et al*. 1998, Guilford *et al*. 1997, Liebhard 2002). The map position of the loci amplified by these SSR primers has also been determined in *M. domestica*. The availability of this large set of mapped SSRs allowed us to select a subset of 10 SSR loci spread over the apple genome, to fingerprint all the genotypes described in Chapter 1. We assumed that these primer sets amplify homologous loci in *M. sylvestris*. A total of 1,452 *M. sylvestris* and *M. domestica* samples were included in the analysis.

2.1 PLANT MATERIALS AND DNA EXTRACTIONS

Leaf samples of all 1,452 *M. domestica* and *M. sylvestris* genotypes described in Chapter 1 (Table 1.2) were used for genetic analysis. Total Plant DNA was isolated from 40 mg of ground dried leaf material, using the CTAB extraction procedure described in Dumolin *et al.* (1995). DNA concentrations were estimated and standardised against known concentrations of λ DNA on 1.5% agarose gels. All 1,452 genotypes were analysed with SSR markers. A subset of 250 genotypes was chosen for S-SAP and REMAP analysis.

2.2 ANALYSIS OF THE LEVEL OF DIVERSITY AT 10 SSR LOCI

Preliminary remark: More detailed technical and methodological information can be found in Coart *et al*. (2003).

2.2.1 METHODS

2.2.1.1. Description of SSR loci

All accessions described in WP1 were typed for 10 SSR loci, spread over different chromosomes: NZ02b01, NZ04h11, NZ05g08, NZ23g04, NZ28f04, CH01h10,

CH01f02, CH01h01, CH02b12 and CH02c06. Normalised nomenclature of the loci follows Liebhard *et al*. 2002, with loci with prefix 'NZ' from Guilford *et al*. (1997) and loci with prefix 'CH' from Gianfranceschi *et al*. (1998). Information on loci NZ02b01, NZ04h11, NZ05g08 and NZ23g04 was produced at KUL-FTC. The other six loci were typed at ILVO-PLANT. The choice of these 10 SSR loci was based on degree of polymorphism and applicability over different *Malus* species (Coart *et al*. 2003).

2.2.1.2 Data analysis

Identification of hybrids

The software *STRUCTURE* (Pritchard *et al*. 2000, Falush *et al.* 2003) was used to analyze the genotypic data of all *M. sylvestris* and *M. domestica* accessions from different regions and countries (Table 1.3). The software implements a model-based Bayesian clustering approach for probabilistic assignment of individuals to a number of clusters with simultaneous estimation of the unknown allele frequencies within them. Assignment of individuals and inference of allele frequencies is performed in such a way that departures from Hardy-Weinberg and gametic-phase disequilibrium within the clusters are minimized. Cultivated genotypes were used as learning samples ('population information model') and the number of clusters was set at 2 (as in Beaumont *et al*. 2001). This allowed the identification of hybrids between both gene pools and the detection of genotypes that were outliers in their sample of origin and that in fact belonged to another gene pool.

Diversity estimates of *M. sylvestris* populations

Prior to data analysis, the putative hybrids detected in the *STRUCTURE* analysis described above were removed from the dataset. F-statistics (Weir and Cockerham, 1984) were estimated for each microsatellite locus independently using the software GENE-SURVEY (Vekemans & Lefèbvre 1997). The following statistics of within-population genetic variation were computed as average over loci: mean number of alleles per locus A; average observed heterozygosity H_O; average gene diversity H_E, computed according to Nei (1978); and Wright's inbreeding coefficient F_{IS} corrected for small sample sizes (Kirby 1975). Except for the locations Treignes, Rance and Houyet, the Walloon accessions did not consist of real populations. In practice, in each 'cantonnement' several forests were sampled but only one tree per forest. The Walloon 'populations' were named according to the 'cantonnement' of origin. These artificial groups are considered for the calculation of population statistics. Therefore, the nature of these 'populations' should be kept in mind for the interpretation of the results. .Nei genetic distances (Nei 1978) between populations, between collections and regions of origin were also calculated.

Diversity estimates of *M. domestica* collections

Prior to data analysis, duplicates (see 2.2.2.3) were removed from the dataset. The analysis of genetic diversity in these *M. domestica* collections was complicated by the presence of triploid apple genotypes. Because most software programmes for population genetics and genetic diversity studies only deal with diploid marker data, the number of programmes applicable to our data set was limited.

SPAGeDi 0.0 (Hardy and Vekemans, 2002) was used to calculate for each collection the (average) number of alleles per locus, the allele frequencies and the expected heterozygosity H_e (Nei, 1978). Observed heterozygosity H_o was calculated per locus in Excel based on the raw SSR-data. Following, Wrights inbreeding coefficient F_{IS} was calculated as $(H_e-H_o)/H_e$. The effective number of alleles A_e was calculated per locus as $1/(1-H_e)$. For *M. domestica*, collections were considered as populations and population genetics results should be interpreted with this information kept in mind.

2.2.2 RESULTS AND DISCUSSION

2.2.2.1. Descriptive statistics

Table 2.1 reports on the number of alleles and F-statistics for each of the ten SSR loci typed. All loci were highly polymorphic with the number of alleles per locus ranging from 10 to 36 in *M. sylvestris* and from 11 to 38 in *M. domestica*.

Loci NZ04h11 and NZ05g08 (marked in bold) showed inbreeding values which are one order or magnitude higher than the other 8 loci (Table 2.1). These *Fis* values are in fact too high for a typically outcrossing species. It was therefore decided to remove the information derived from these two loci from the dataset. These two loci were included in the *STRUCTURE* analysis to identify hybrids, but this had little influence on the final conclusions (results not shown).

Table 2.1: Number of alleles, ranges and F-statistics of Gene diversities, calculated with GENE-SURVEY.

	M. domestica		*M. sylvestris*				
Locus	***# alleles***	***Range (bp)***	***# alleles***	***range***	***Fit***	***Fis***	***Fst***
CH01f02	28	162-230	25	162-230	0.061	0.041	0.021
CH01h01	22	91-148	21	106-148	0.050	0.030	0.020
CH01h10	28	87-145	26	87-145	0.028	0.002	0.026
CH02b12	24	97-149	23	97-149	0.035	0.018	0.018
CH02c06	38	201-286	36	201-286	0.086	0.064	0.024
NZ02b01	23	213-246	10	213-246	0.052	0.020	0.033
NZ04h11	11	201-225	22	201-225	**0.416**	**0.321**	**0.140**
NZ05g08	25	105-221	16	105-221	**0.627**	**0.621**	**0.015**
NZ23g04	17	84-120	16	84-120	0.099	0.071	0.031
NZ28f04	12	93-114	11	95-114	0.086	0.050	0.039

2.2.2.2. Identification of hybrids

Figure 2.1 shows the assignment of accessions to both gene pools, ranked according to their assignment to the wild gene pool.

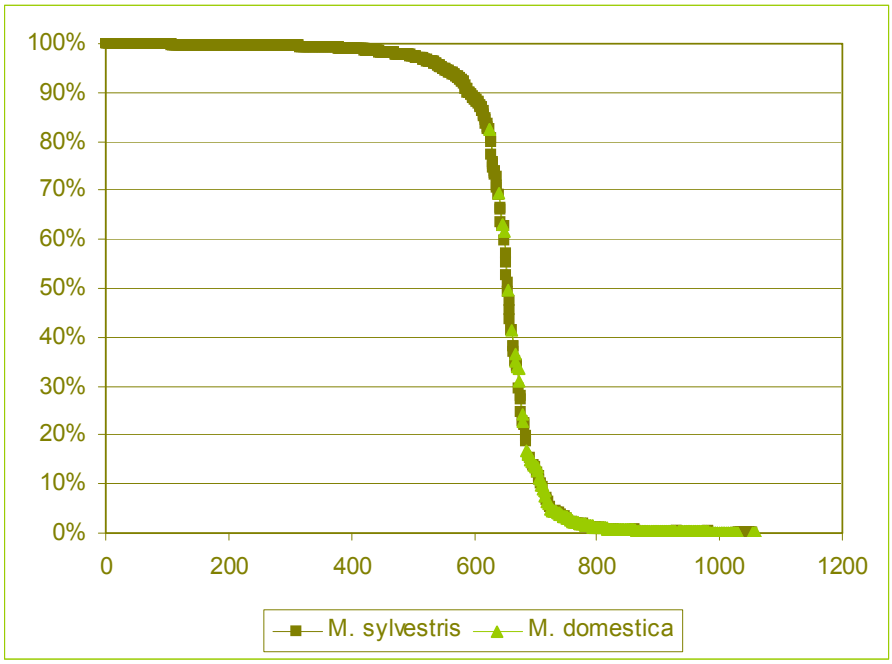


Figure 2.1: Results of the *structure* analysis with 'population information model'. For each genotype, the proportion of its genetic information assigned to the wild gene pool is shown. *M. sylvestris* refers to trees sampled in the field. Note that some of them fit better in the cultivated gene pool

The STRUCTURE analysis resulted in the overall assignment of 87.31% of the genetic information of the accessions collected in the field to the wild gene pool and 12.69% to the cultivated gene pool. For *M. domestica* accessions, 4% of their genetic information was attributed to the *M. sylvestris* gene pool and 96% to the cultivated gene pool.

Using the arbitrary cut-off values of 15 and 85%, accessions were defined as 'genuine' wild genotypes (<15% assigned to cultivated gene pool), hybrids (between 15 and 85% assigned to cultivated gene pool) or cultivars (>85% assigned to the cultivated gene pool). Cultivars found in the wild are further referred to as feral cultivars. Table 2.2 and Figure 2.2 give an overview of the number of hybrids and (feral) cultivars detected in each 'population'. For *M. sylvestris*, Denmark was the sample with the smallest percentage of hybrids and feral cultivars, while among the genotypes from Central Asia, 20% were identified as feral cultivar and 30% as hybrids. Between the two Belgian regions, Wallonia contained a higher percentage of feral cultivars and hybrids (11 and 12% respectively) than Flanders (7 and 2% respectively). This information was used to select the genotypes to be included in the *M. sylvestris* core collection described in Chapter 5.

Table 2.2: Number (on left hand side) and percentage (on right hand side) of cultivars, hybrids and wild genotypes for each population and species.

Population	N	#(Feral) Cultivar	#Hybrid	#Wild	%(Feral) Cultivar	%Hybrid	%Wild
Flanders	225	15	5	205	0.07	0.02	0.91
Wallonia	367	39	45	283	0.11	0.12	0.77
Denmark	94		1	93	0.00	0.01	0.99
Germany	13	2		11	0.15	0.00	0.85
France	21	1		20	0.05	0.00	0.95
Central Asia	10	2	3	5	0.20	0.30	0.50
Total M. sylvestris	***730***	***59***	***54***	***617***	***0.08***	***0.07***	***0.85***
NBS gene bank	82	78	4	0	0.95	0.05	0.00
CRA gene bank	148	138	10	0	0.93	0.07	0.00
Danish gene bank	25	21	4	0	0.84	0.16	0.00
Modern cultivar	50	49	1	0	0.98	0.02	0.00
Total M. domestica	***328***	***308***	***20***		***0.94***	***0.06***	***0.00***

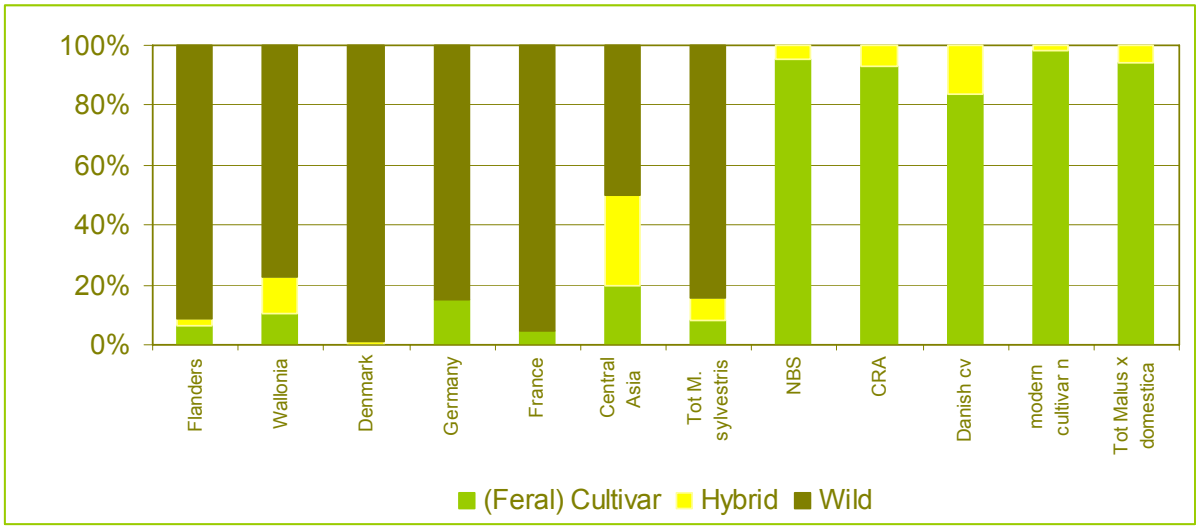


Figure 2.2: Percentage of (feral) cultivars, hybrids and wild genotypes for each population and species.

2.2.2.3. Occurrence of duplicates among *M. domestica* accessions

Among the 669 *M. domestica* accessions, several genotypes were deliberately replicated to confirm their identity. The accurate determination of the identity of a cultivar is of utmost importance to rationalise the management of collections. Information on the duplicated samples was provided once the analyses were completed, to avoid bias during the scoring process. Based on the information of ten microsatellite loci, 104 genotypes occurred at least twice within the SSR dataset. In total, these 104 duplicated genotypes correspond to 295 different accession numbers in the project database. One genotype even occurred 13 times. It should be stated however, that the identity control as done based on 10 microsatellite loci does not allow to discriminate sports or mutants of the same genotype (sports derived from the same original cultivar will share the same SSR multi-locus fingerprint).

The results revealed sets of identical genotypes that were expected to be identical, but also other interesting results were found. Trees that were found to be closely related based on previous phenotypic evaluations, proved to be genetically identical using 10 SSR loci. On the other hand, some phenotypically very distinct trees displayed the same genotype and some trees that were believed to be identical carried different alleles. It is clear that these data have to be contrasted with the phenotypic datasets for further evaluation (see Chapter 5).

2.2.2.4. SSR diversity of *M. sylvestris* populations

Within population diversity

All estimates of within population diversity were high, as expected in *M. sylvestris*. The percentage of polymorphic loci (PLP) was 100% in all populations (Table 2.3). The allelic richness of all populations was also rather high, oscillating between 3.6 for the smallest population (HabaylaNeuve, 3 individuals) and 16.9 for the largest population (Meerdal, 175 individuals). Estimates of observed and expected heterozygosity were comparable, resulting in low levels of heterozygosity deficiencies or excesses as expected for an obligate cross-pollinator. However, as most of the populations do not represent biological populations, F_{is} values cannot be interpreted in terms of population genetics.

Table 2.3: Within-population diversity estimates for 43 M. sylvestris populations

Population	n	PLP	A	H_O	H_E	F_{is}
Aywaille	5	100.0	5.8	0.9000	0.8633	-0.0526
Beauraing	3	100.0	4.0	0.7917	0.8167	-0.0210
Bièvre	8	100.0	7.6	0.8906	0.8604	-0.0392
Bouillon	6	100.0	6.5	0.8125	0.8693	0.0588
Bullange	5	100.0	5.6	0.7333	0.8444	0.1198
Chimay	7	100.0	6.6	0.6964	0.8516	0.1781
Couvin	8	100.0	7.5	0.7813	0.8448	0.0716
Dinant	4	100.0	5.0	0.7604	0.8372	0.0619
Elsenborn	7	100.0	6.5	0.8214	0.8462	0.0242
Eupen	10	100.0	9.0	0.9000	0.8618	-0.0469
Florenville	10	100.0	8.4	0.7944	0.8309	0.0381
HabaylaNeuve	3	100.0	3.6	0.8125	0.8292	-0.0171
Houyet	13	100.0	8.3	0.9122	0.8460	-0.0776
LaRoche	8	100.0	6.9	0.7902	0.8488	0.0608
Liège	6	100.0	5.5	0.7917	0.8220	0.0516
Marche	12	100.0	8.6	0.8333	0.8279	-0.0083
Namur	12	100.0	8.4	0.8513	0.8368	-0.0169
Nassogne	9	100.0	8.4	0.8715	0.8770	0.0033
Neufchâteau	8	100.0	8.1	0.7969	0.8708	0.0813
Paliseul	6	100.0	6.3	0.8750	0.8295	-0.0622
Philippeville	10	100.0	8.3	0.8375	0.8329	-0.0080
Rance	18	100.0	9.5	0.7627	0.8196	0.0687
Rochefort	10	100.0	7.1	0.8250	0.8349	0.0093
Spa	8	100.0	7.4	0.8906	0.8760	-0.0206
StVith	5	100.0	6.5	0.7937	0.8788	0.0885
Thuin	13	100.0	8.3	0.8330	0.8256	-0.0096
Treignes	16	100.0	9.6	0.7885	0.8452	0.0658
Verviers	8	100.0	6.9	0.7656	0.8365	0.0811
Vielsalm	9	100.0	8.1	0.8333	0.8513	0.0181
Viroinval	7	100.0	7.3	0.7619	0.8533	0.1025
Virton	8	100.0	7.3	0.7813	0.8292	0.0541
Wellin	12	100.0	9.8	0.9063	0.8800	-0.0316
DenmarkHE	23	100.0	10.1	0.7753	0.8243	0.0588
DenmarkKA	20	100.0	7.6	0.7831	0.7751	-0.0099
DenmarkKO	23	100.0	10.0	0.8099	0.8186	0.0101
DenmarkNO	27	100.0	8.9	0.8142	0.8044	-0.0106
France	20	100.0	11.5	0.7813	0.8361	0.0650
Germany	11	100.0	7.0	0.7341	0.8033	0.0830
Heverleebos	11	100.0	5.9	0.6867	0.7510	0.0898
Meerdaal	175	100.0	16.9	0.8156	0.8231	0.0091
Voeren	8	100.0	6.5	0.8125	0.7500	-0.0876
Weynendaelbos	4	100.0	4.6	0.8438	0.7964	-0.0705
Central Asia	5	100.0	7.0	0.7250	0.9250	0.2144
	mean		7.64	0.8088	0.8369	0.0274
	std		0.0433	0.0325		

Among populations diversity

Estimates of gene diversity among the 43 populations indicate a low level of genetic differentiation with an overall low but significant *Fst* value of 0.026. The values were of the same order of magnitude obtained in a similar study of *M. sylvestris* at a smaller scale (Coart *et al*. 2003). Based on Nei's genetic distance between pairs of populations, a neighbour joining tree was constructed to summarise the relationships among all *M. sylvestris* populations (Figure 2.3). The low bootstrap values at forks indicate that the clustering of samples has no strong support. The Belgian populations do not cluster according to their geographic location. At the European level, the clustering of all Danish populations shows their distinctness from other *M. sylvestris* populations.

When we group the *M. sylvestris* populations in six regions (Flanders, Wallonia, Germany, France, Denmark and Central Asia), the genetic differentiation between regions is also low with *Fst* = 0.036. This small differentiation was also significant. Table 2.4 shows the pairwise Nei's distance between regions. The Flemish and Walloon regions were the least differentiated (pairwise distance of 0.008), while the Central Asian region was the most differentiated from all other regions what is in accordance with the larger geographic distance to other regions (Nei's distance ranging from 0.052 to 0.086).

Table 2.4: Corrected Nei's distance (Nei 1978) among 6 regions

	C. Asia	***Denmark***	***Flanders***	***France***	***Wallonia***	***Germany***
Central Asia	0.000					
Denmark	0.086	0.000				
Flanders	0.071	0.030	0.000			
France	0.063	0.048	0.039	0.000		
Wallonia	0.052	0.031	0.008	0.024	0.000	
Germany	0.077	0.028	0.033	0.051	0.033	0.000

Relationship between genetic and geographic distances at the Belgian scale

Using the software SPAGEDI, an autocorrelation analysis was performed between the pairwise genetic distance among individuals (expressed as *Fst*/(1-*Fst*)) and their geographic distance, expressed in meters. As shown in Figure 2.4, no geographical structuring of genetic diversity among Belgian populations is present. Up to 10 m, genotypic and geographical distances are clearly correlated. However, beyond 10 m this relationship drops dramatically to become non-significant at a distance of 100 m.

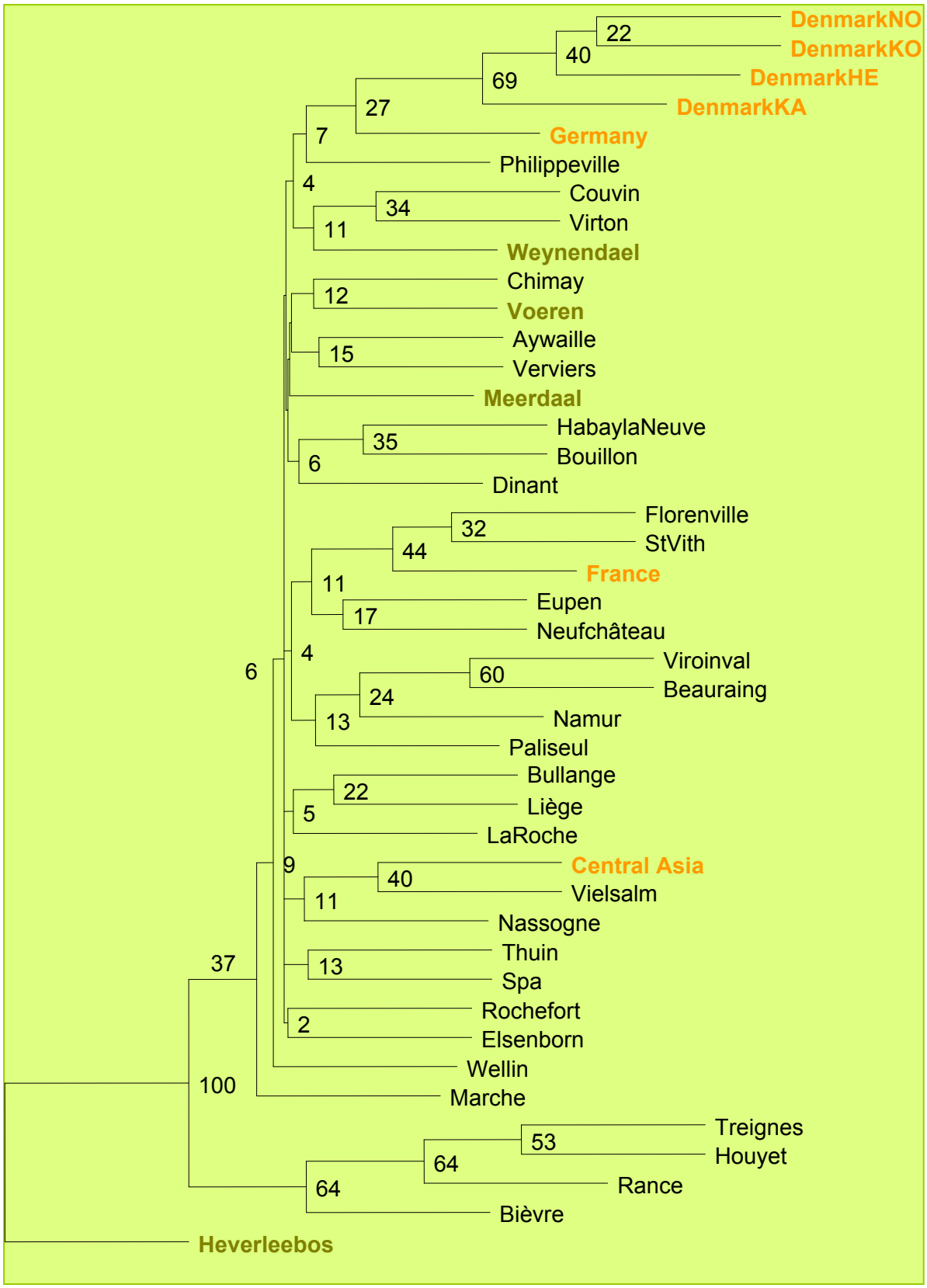


Figure 2.3: Neighbour joining tree based on Nei's genetic distance between *M. sylvestris* populations. Foreign populations are marked in orange; Flemish populations are marked in green; other populations originate from the Walloon region.

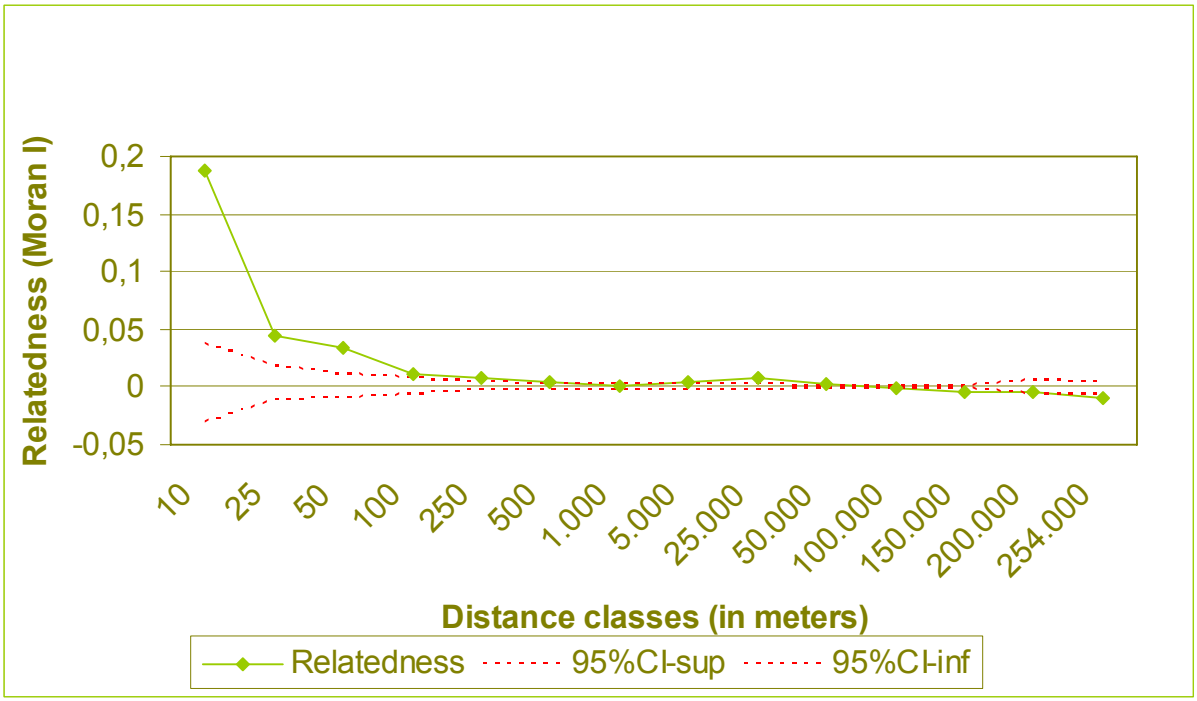


Figure 2.4: Pairwise relatedness of genetic and geographic distance between individuals

2.2.2.5. SSR diversity of *M. domestica* collections

Within collections diversity

As for *M. sylvestris*, all estimates of within population diversity were high. The percentage of polymorphic loci (PLP) was 100% in all collections (Table 2.5). The allelic richness of all populations was also rather high, oscillating between 7.5 for the smallest collection (cider cultivars, 23 individuals) and 12.4 for the largest collection (CRA-W cultivars, 193 individuals). Estimates of observed and expected heterozygosity were comparable, resulting in low levels of heterozygote deficiencies or excesses as expected for this obligate cross-pollinated crop. However, as the collections do not constitute populations in genetics terms, , Fis values cannot be interpreted in terms of population genetics. Expected heterozygosity values are comparable over collections.

Table 2.5: Within-collection diversity estimates in *M. domestica*

Population	n	PLP	A	H_O	H_E	F_{is}
Modern	80	100.0	11.8	0.7625	0.7643	0.0082
NBS	133	100.0	12.0	0.8180	0.7796	-0.0543
CRA-W	193	100.0	12.4	0.7969	0.7795	-0.0232
Denmark	25	100.0	9.7	0.7956	0.8318	0.0445
Cider	23	100.0	7.5	0.7739	0.8003	0.0423
	mean		10.68	0.7894	0.7911	0.0035
	std		0.92	0.0097	0.0117	0.01906

Among collections diversity

In total, 11 alleles were detected which were frequent (f > 5%) in one collection, but rare (f < 5%) or absent in the other collections. The cider apple collection holds five of these alleles, and the Danish collection and the modern apple collection each three. When only the modern and Belgian old collections were considered, eight alleles were found which were frequently present in the old apple collections but not in the modern apple collection and five were frequent alleles in the modern apple collection but not in the old collections. Also, 44 collection-specific alleles were found over the five collections, but for only 4 of them, the frequency was above 1%. When again only the modern and Belgian old apple collections were considered, 37 collection-specific alleles were detected, but only four had a frequency higher than 1%. All were detected in the modern apple collection.

Estimates of gene diversity among the five populations indicate a low level of genetic differentiation with an overall low but significant *Fst* value of 0.036. Table 2.6 shows the pairwise distances between collections. Strikingly, the differentiation between the Belgian collections of old cultivars (CRA-W and NBS) was zero, illustrating the similarity of both collections.

Table 2.6: Corrected Nei's distance (=DNei78) among five *M. sylvestris* collections

	Cider cv	***CRA-W cv***	***Dan cv***	***Modern cv***	***NBS cv***
Cider cv	0.000				
CRA-W cv	0.030	0.000			
Danish cv	0.031	0.014	0.000		
Modern cv	0.057	0.035	0.028	0.000	
NBS cv	0.027	0.000	0.014	0.027	0.000

2.2.2.6. Differentiation between *M. sylvestris* and *M. domestica*

Allelic differences were calculated between both species after exclusion of hybrids and feral cultivars as identified with *STRUCTURE*. In total, 79 alleles private to *M. sylvestris* were identified. However, only 3 of these alleles occurred with a frequency higher than 5%. Five alleles were private to the cultivated gene pool, but all had frequencies lower than 5%.

Most alleles were shared by both species, and only differed in frequency. In a few cases, the range of allele sizes was larger for *M. sylvestris* than for *M. domestica* as illustrated for locus CH01h01 in Figure 2.5. Two of the private alleles for *M. sylvestris* with f>0.05 are from this locus (sizes 142 and 144).

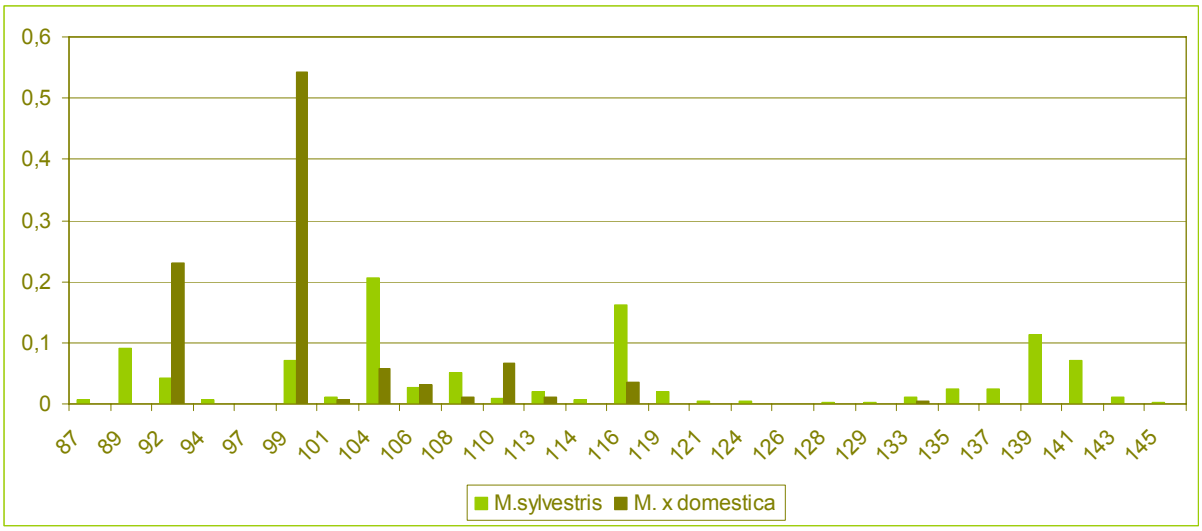


Figure 2.5: Allele frequencies for *M. sylvestris* and *M. sylvestris* for SSR locus CH01h01

Although *STRUCTURE* analyses showed that both species consist of separated gene pools, the genetic differentiation measured as *Fst* was only 0.072, a value often encountered in literature between conspecific populations. Table 2.7 shows the pairwise genetic distances among *M. sylvestris* regions and *M. domestica* collections. Based on these genetic distances, a Neighbour joining tree was constructed (Figure 2.6). This tree clearly shows the differentiation at the species level, although the overall differentiation value was low. As could already be deduced from the *STRUCTURE* analyses, the collections of old cultivars are slightly closer related to *M. sylvestris* than the modern cultivars.

Table 2.7: Pairwise Nei's distance between regions of *M. sylvestris* and *M. sylvestris* collections

	Asia	***Den***	***Fland***	***Fran***	***Wall***	***Germ***	***Cider***	***CRA***	***Dan***	***mode***	***NBS***
Asia	0.000										
Denm	0.086	0.000									
Fland	0.071	0.030	0.000								
France	0.063	0.048	0.039	0.000							
Wall	0.052	0.031	0.008	0.024	0.000						
Germ	0.077	0.028	0.033	0.051	0.033	0.000					
Cider cv	0.090	0.123	0.106	0.090	0.079	0.105	0.000				
CRA cv	0.111	0.145	0.127	0.112	0.102	0.135	0.030	0.000			
Dan cv	0.078	0.137	0.113	0.107	0.090	0.125	0.031	0.014	0.000		
modern	0.125	0.148	0.133	0.126	0.108	0.152	0.057	0.035	0.028	0.000	
NBS cv	0.115	0.149	0.127	0.114	0.102	0.132	0.027	0.000	0.014	0.027	0.000

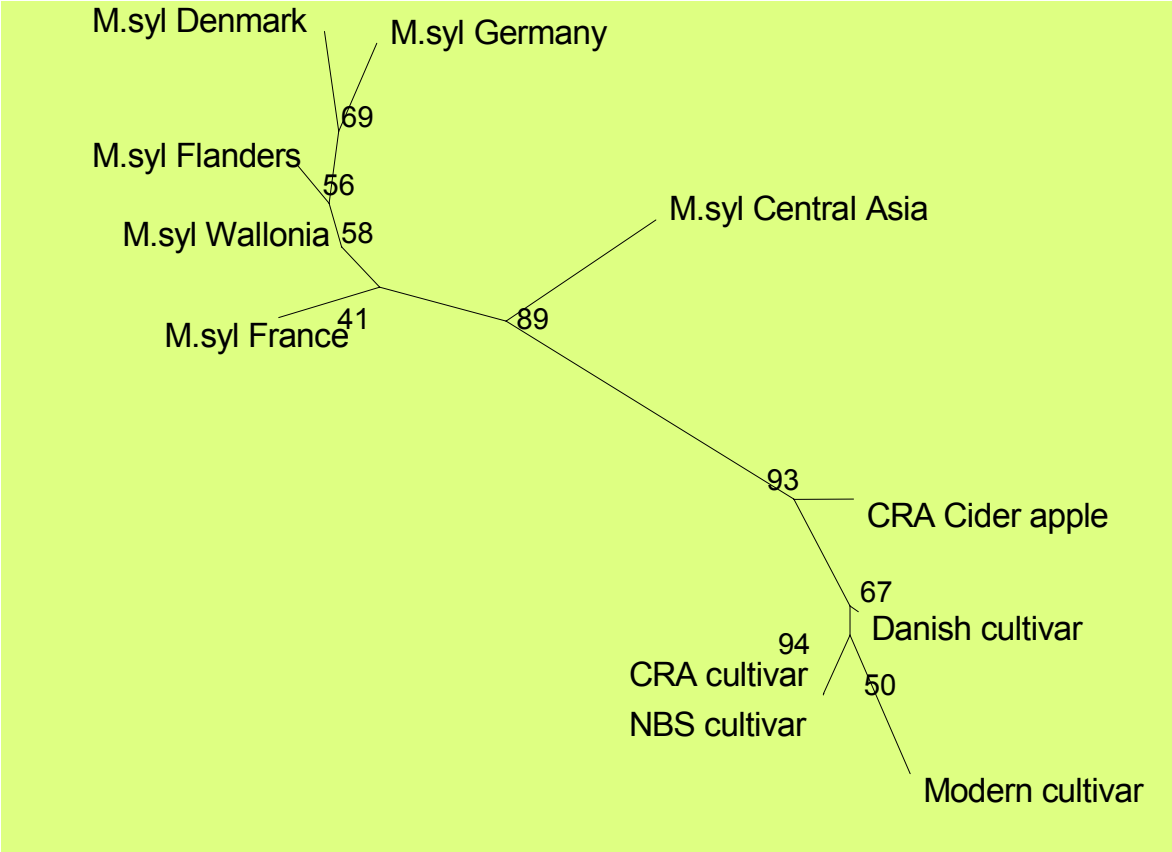


Figure 2.6: Neighbour joining tree based on Nei's genetic distance

2.2.2.7 Integration of SSR data and phenotypic data for *M. sylvestris*

For 249 *M. sylvestris* accessions, SSR data and phenotypic evaluations were available. The number of genotypes among the different genotypic categories wild, hybrid and feral cultivars are not balanced as they contain 193, 52 and only 4 genotypes respectively. For these genotypes, the correlation between both datasets was calculated.

As stated in Chapter 1, fruit characteristics are considered the most discriminative between wild and cultivated apples. Unfortunately, fructification was not observed on enough genotypes to calculate correlations between the phenotype and genotype. Also hairiness of leaves and twigs is often used as discriminatory variable between species and we therefore focused on these characters. Figures 2.7 and 2.8 show the frequency of different hairiness classes (1 = no hairs to 9 = very high pubescence) over different genotypic categories for leaf and twig pubescence respectively. For twig hairiness, variability is very high for both wild and hybrid genotypic categories with all nine hairiness classes detected in both groups. For leaf hairiness, variability is still high in all three genotypic categories but the highest hairiness classes (6 to 9) where never scored among the wild trees or the hybrids.

The pubescence of leaves and twigs is significantly correlated with the three genotypic categories (Table 2.8). As expected, wild trees appear to have less hairy twigs and leaves. A Kruskal Wallis test shows that both hairiness of twigs and leaves are significantly different

(with $p \leq 0.001$) between the different categories of genotypes. A similar test done only on the wild and hybrid categories, showed that these are not significantly different ($p = 0.149$).

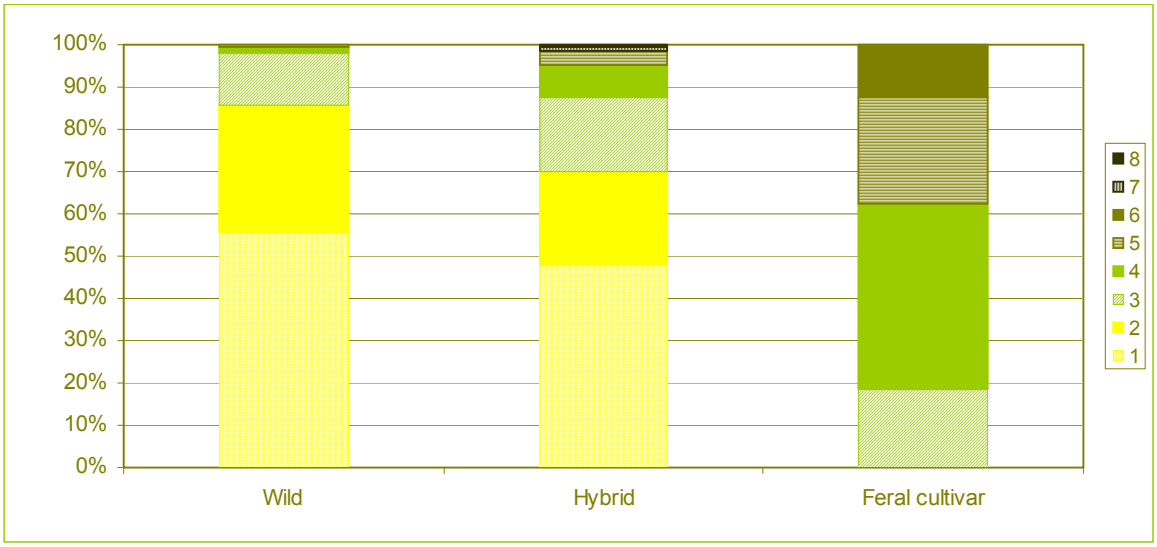


Figure 2.7: Leaf hairiness over different genotypic categories as defined by the *Structure* analysis

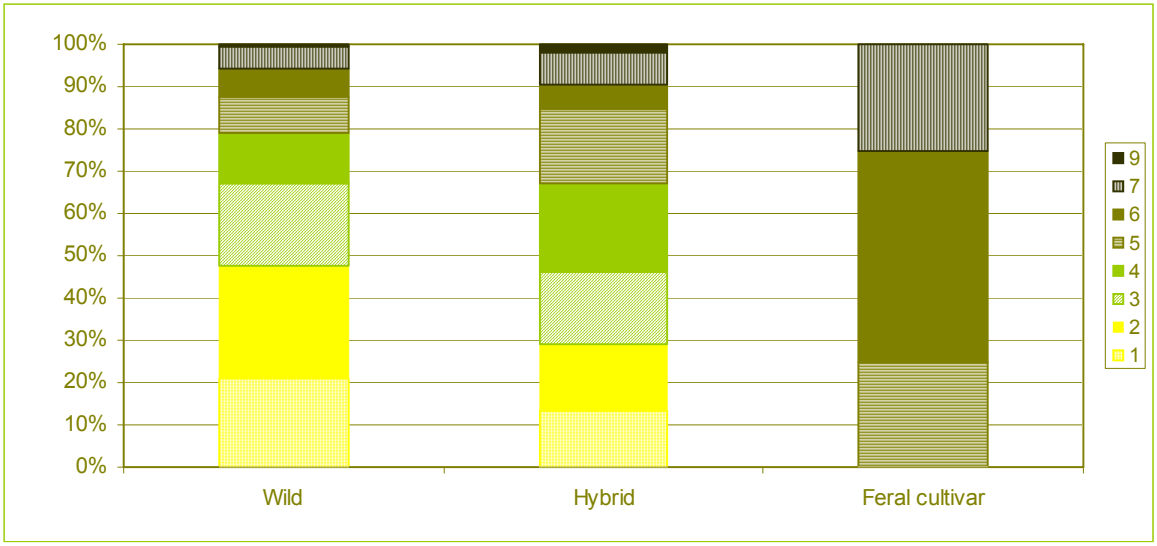


Figure 2.8: Twig hairiness over different genotypic categories *as defined by the Structure analysis*

Table 2.8: Correlation (Spearman's Rho) between hairiness of leaves and twigs and genotypic categories for apple trees sampled in nature.

		Hairiness of twigs	***Hairiness of leaves***	***Genotypic category***
Hairiness of twigs	***Correlation coef.***	1.000	.291**	-.203**
Hairiness of leaves	***Correlation coef.***	.291**	1.000	-.226**

In conclusion, it is clear that hairiness of leaves is a better indicator than hairiness of twigs and that hairiness can only be used to identify feral cultivars in the field. The variation for these traits among the wild trees is too large to discriminate 'genotypic' hybrids from 'genotypic' wild trees.

Correlations between all other phenotypic parameters (including disease susceptibility) and site descriptors on the one hand and the genotypic status of the trees on the other hand, were also calculated. The only parameters significantly correlated with the genotype were vigour, forking and proximity of a road (all $p < 0.01$). Wild trees tend to be less vigorous, less forked and less often found in roadsides than hybrids and feral cultivars.

2.3 ANALYSIS OF THE LEVEL OF DIVERSITY USING S-SAP AND REMAP

As complement to the SSR analysis, a multi-locus DNA-marker system based on the recently identified copia-like retrotransposable elements in cultivated apple was applied (Tignon *et al.* 2001). Like SSR sequences, retrotransposable elements are abundant and dispersed throughout the eucromatic regions of chromosomes. The unique retrotransposition mechanism results in insertions of retrotransposon into new sites without losing the parental copies. As these retrotransposon insertions are irreversible, they are useful not only for the estimation of genetic differentiation parameters, but also for the analysis of phylogenetic (parental lineage) relationship. Here the Sequence Specific Amplification Polymorphism (SSAP) technique was applied, which is a multiplex method able to detect a large number of highly informative individual retrotransposon insertions as bands on a sequencing gel (Waugh *et al.* 1997). This technique was developed in apple by CRA-W based on the information derived from the characterisation of a "copia-like" retrotransposon in this species (Research Project n° S-6067, Ministry of Traders and Agriculture). We planned to obtain multiplex insertional polymorphism patterns using the SSAP method on a subset of samples (at least 300 genotypes, including 150 wild Belgian apples, 100 cultivated varieties and 50 wild genotypes from Kazakhstan).

By using a combination of one LTR-primer and one adapter primer, the presence of a retrotransposon at a given distance from a restriction site is analysed (S-SAP). An LTR-primer combined with a microsatellite primer may be used for the detection of the association of a retrotransposon copy with microsatellite loci (REMAP). A full-length retrotransposon copy was previously isolated from an apple cultivar and characterized (Tignon *et al.* 2001). Based on the sequence of the retrotransposon LTR3' region, primers were designed and used for REMAP and S-SAP fingerprinting reactions (Table 2.9).

Table 2.9: Primers used for S-SAP and REMAP reactions.

LTR primers	LTR 3.1: 5' TAG GCT AGT CAA CTT TCT T 3', LTR 3.2 : 5' GCA TGA AAT AAA CAT GTT G 3', LTR 3.3 : 5' GTG TTA AAA TGC ATG AAA T 3'.	S-SAP, REMAP
Microsatellite repeat primers	<u>Degenerated :</u> µsat01 : 5' GAG AGA GAG AGA GAG A**H** 3' µsat02 : 5' GTG TGT GTG TGT GTG T**H** 3' µsat03 : 5' TCT CTC TCT CTC TCT C**H** 3' <u>Selective :</u> µsat11 : 5' GAG AGA GAG AGA GAG AGA **T** 3' µsat12 : 5' GAG AGA GAG AGA GAG AGA **C** 3' µsat13 : 5' GAG AGA GAG AGA GAG AGA **A** 3' µsat21 : 5' GTG TGT GTG TGT GTG TGT **T** 3' µsat22 : 5' GTG TGT GTG TGT GTG TGT **C** 3' µsat23 : 5' GTG TGT GTG TGT GTG TGT **A** 3'	REMAP

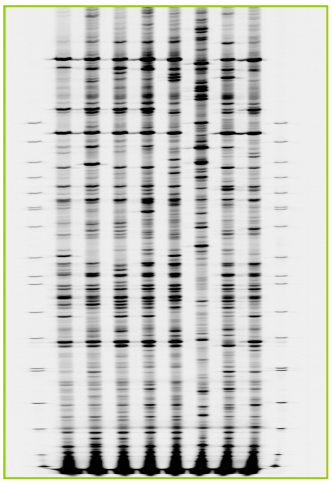


Figure 2.9: Representative REMAP profiles obtained for *M. sylvestris* samples.

The S-SAP and REMAP protocols described in literature were tested and adapted to the fingerprinting of apple genotypes. In particular, different primer combinations were tested on several apple genotypes and resulting profiles were evaluated and compared for their reproducibility and information content. For S-SAP, the best results were obtained using the *Mse*I – CAA and *Mse*I – CTG AFLP primers in combination with LTR3.2 primer. For REMAP, reproducible and highly polymorphic fingerprint profiles were obtained using several primer combinations, especially with primer combinations LTR 3.2 - µsat01 and LTR 3.3 - µsat11 (Figure 2.9). Therefore, this method was selected for further analysis of the subset of 250 *Malus* samples. However, when techniques were applied on the larger set of samples, results did not seem reliable and no data analysis was performed. As a consequence, no comparison with SSR fingerprinting could be made.

2.4 CONCLUSIONS

This study confirmed that *M. sylvestris* and *M. domestica* represent clearly differentiated gene pools that can be separated based on allele frequencies at 10 SSR loci. However, 7% of the sampled *M. sylvestris* trees was identified as hybrid, indicating ongoing gene-flow between both species. Strikingly, hybrids were almost entirely restricted to Belgium. Next to environmental factors that might enhance contact between the species in Belgium (e.g. smaller sizes and fragmentation of forests), this might also be the reflection of the apple cultivars analysed. Many ancient apple cultivars of Belgian origin or often planted in Belgium were analysed, but only a small selection of old Danish varieties were included and no German or French cultivars. Eight % of the accessions, sampled as *M. sylvestris* in nature,

were identified as feral cultivar. Feral cultivars were detected among *M. sylvestris* samples of most countries.

M. sylvestris still contains high genetic diversity within most Belgian and European populations. It is clear that no geographical structuring of this diversity is present at the regional level. Only the Danish populations group together and can be considered a separate gene pool. When pooling the populations into five regions, it becomes apparent that Flanders and Wallonia are the most related regions. Within populations, genotypic and geographical distances are positively correlated for individuals up to 10 m apart. Beyond 10 m this relationship drops dramatically to become non-significant at a distance of 100 m.

Within *M. domestica* collections, high allelic richness and expected heterozygosities were found. The modern and the cider apple collections were genetically the most distant collections. This reflects very likely the influence of breeding on the genetic make-up of a cultivated species. Breeding for different purposes has resulted in the creation of two gene pools, characterised by other alleles and allele frequencies. Strikingly, the collections of old Belgian cultivars of CRA-W and NBS were extremely similar with Nei's genetic distance between both < 0.001. Also within collection diversities were very similar and all alleles were shared between both collections.

The results of this genetic diversity analysis in the modern apple collection compared to the collections of old apple varieties showed that modern apple breeding has probably not led to a decrease of genetic diversity in apple. Based on pedigree analysis, Noiton (1996) suggested that the apple gene pool used by modern breeders is probably too narrow to guarantee breeding progress in future. However, in this study, no strong evidence was found for the occurrence of genetic erosion in apple since modern apple breeding started, at least at neutral loci.

CHAPTER 3	Coordinator	**ILVO-PLANT**
	Other partners involved	**none**

STUDY OF ADMIXTURE BETWEEN *M. DOMESTICA* AND *M. SYLVESTRIS* GENE POOLS

PUBLICATIONS

Coart E, Van Glabeke S, De Loose M, Larsen AS and Roldán-Ruiz I (2006) Chloroplast diversity in the genus *Malus*: New Insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*M. domestica* Borkh.). Molecular Ecology, 15, 2171-2182.

Koopman W*, Li Y*, Coart E*, van de Weg W, Vosman B, Smulders M.J.M., Roldán-Ruiz I. Reconstruction of gene flow among wild apple (*Malus sylvestris* (L.) Mill.) populations and hybridization with cultivated apple (*M. domestica* Borkh.) using multilocus microsatellite haplotype sharing. *: the three authors contributed equally. Molecular Ecology, in press.

INTRODUCTION

In addition to the purely academic interest in the origin of the domesticated apple, studies on the relationship between the domesticated apple and its wild relatives are relevant as they might point towards taxa that can serve as sources of novel genes for breeding purposes (Robinson *et al.* 2001). By the start of this project, the most widely accepted theory, based mainly on morphological and molecular evidence, pointed to series *Malus* and specifically towards *M. sieversii*, a wild species of Central Asia, as the most likely maternal ancestor of *M. domestica*. However, those studies were based on small numbers of accessions and more domesticated apples should be analysed to ensure that rare hybridization events with other *Malus* species that might have contributed to the early domestication of the apple had not been overlooked. Furthermore, the postulated close relationship between *M. sieversii* and *M. domestica* should be confirmed using additional molecular evidence. We analysed levels of chloroplast diversity in a large sample of *Malus* species. The results were contrasted with diversity estimates derived from nuclear SSR markers. The main research questions addressed here were: (i) are *M. sylvestris* and *M. domestica* as clearly differentiated at the chloroplast-DNA level as for their nuclear genomes?; (ii) could *M. sylvestris* have been a maternal progenitor of *M. domestica*?; (iii) is the relationship between *M. sieversii* and *M. domestica* as close as is currently assumed?

In the second part of this Chapter we report on the development and testing of an innovative methodology for the analysis of hybridisation between populations and species. This methodology makes use of haplotypes at linked SSR loci to reconstruct past and recent hybridisation events. This can potentially provide further insights into the amount and regional distribution of recent hybridisation and introgression events

since one would expect significant associations between linked markers under a scenario of (ongoing) introgression but not from common descent caused by ancient use of the wild species in the domestication process (Riesenberg *et al.* 2000; Falush *et al.* 2003). Furthermore, as the expected life span of a haplotype depends on the amount of linkage, the length (in cM) of the haplotype shared contains information on the number of generations passed.

3.1 ANALYSIS OF CHLOROPLAST DIVERSITY

Preliminary remark: In the project description this part is referred to as 'analysis of the matK-gene', but the study was extended to other chloroplast-regions. All technical information can be found in Coart *et al.* 2006.

3.1.1 PLANT MATERIALS AND DNA-EXTRACTIONS

A subsample of *Malus* genotypes was used in the study of chloroplast diversity (Table 3.1). This set of plants constitutes a representative sample of *Malus* species with different geographical origin. Total Plant DNA was isolated from 40 mg of ground dried leaf material, using the CTAB extraction procedure described in Dumolin *et al.* (1995).

Table 3.1: Plant materials included in the study of chloroplast diversity

Population of origin	*N*
M. sylvestris Belgium (WL)	225
M. sylvestris Belgium (FL)	207
M. sylvestris Denmark	73
M. sylvestris France	19
M. sylvestris Germany	9
M. sylvestris Central Asia	9
M. domestica Belgium (NBS)	139
M. domestica Belgium (CRA-W)	149
M. domestica Denmark	23
M. domestica Modern	86
M. domestica Cider apples	25
M. orientalis	10
M. prunifolia	9
M. sieversii	20
M. baccata	11

3.1.2 METHODS

A combination of nuclear and chloroplast data may provide complementary views of the hybridisation process between *Malus* species. In particular, Robinson *et al.*

(2001) hypothesized that a specific region of the chloroplast matK gene could be used to distinguish between edible apple cultivars and *M. sylvestris* individuals. In this gene, two duplications, located 39bp upstream from the 3'-end of the matK-coding region are present. Duplication I is an imperfect 8bp duplication, duplication II is a perfect 18bp duplication. Duplication I seems to be present in most apples of the section *Malus* and is therefore not useful to discriminate between *M. sylvestris* and *M. domestica* individuals. However, all but one of the nine *M. domestica* trees included in the study by Robinson *et al.* (2001) were shown to have duplication II and this duplication was absent in the two *M. sylvestris* individuals included in the study. To establish the validity of this trait to differentiate between M. sylvestris and *M. sylvestris*, we sequenced this region of the matK gene for 18 additional individuals. Subsequently, a marker able to detect the presence or absence of duplication II as a length polymorphism was developed and all accessions of the project were subsequently typed.

We developed a DNA-marker to detect a second polymorphism in the matK region in *Malus*. Sequencing of the matK fragment had shown that at position 338, a point mutation was present (T→G). When a G is present at position 338, a recognition site for the restriction enzyme Hpy99I is formed. The genotypes were scored after restriction as mutation absent (position 338 is a T, one DNA fragment of 643 bp after restriction of the PCR product) or mutation present (position 338 is a G, one restriction fragment of 338 bp and one of 305 bp). This allowed us to detect the presence/absence of this point mutation in each individual genotype.

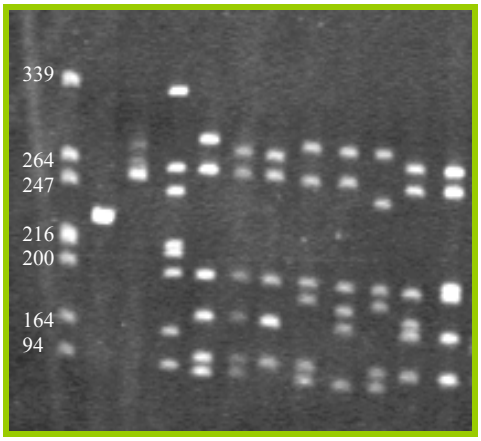


Figure 3.1 PCR-RFLP restriction patterns obtained with HK1/*Eco*RI and DT/*Mse*I. Lanes 1 and 4 are length markers λ/*Pst*I, lengths of this fragments are given in bp. Other lanes are restriction patterns of HK1/*Eco*RI (lanes 2 and 3) and DT/*Mse*I (lanes 5 to 12). Pattern 1 and 2 obtained with HK1/*Eco*RI also share a fragment of circa 1700 bp that is not visible on this gel.

Additional chloroplast haplotyping was performed to identify other chloroplast polymorphisms. For haplotyping, cpDNA fragments were amplified with conserved primers and subsequently cut with restriction-enzymes in order to detect polymorphisms among genotypes. In a first step, ten pairs of universal chloroplast primers (trnH/trnK1, trnC/trnD, trnD/trnT, psbC/trnS, trnS/trnfM, psaA/trnS, trnS/trnT, trnM/rbcL, trnT/trnF and trnK1/trnK2) were tested for amplification on 10 randomly

chosen *Malus* genotypes belonging to different species. Primer pairs are as described in Dumolin-Lapègue *et al.* (1997) and Lofty *et al.* (2003). The three primer pairs which were able to amplify clear DNA fragments in the 10 genotypes analysed were selected (trnH/trnK1, trnD/trnT and trnS/trnT). In a second step, the PCR products amplified by these three primer pairs in 48 genotypes from different species were digested with 16 different restriction endonucleases (AluI, MboI, DpnI, MunI, MluI, MseI, EcoRI, HinfI, HindIII, PvuII, XmiI, XbaI, SalI, RsaI, TaqI, PstI). Following restriction, polymorphisms were only revealed in the region trnD/trnT using MseI and HinfI and in the region trnH/trnK1 with MseI, HindIII, EcoRI, MunI and PvuI. Restriction patterns derived for each of these cpDNA regions using different restriction enzymes provided identical information on the 48 genotypes analysed (results not shown), therefore the enzymes resulting in the clearest restriction patterns were selected to type the complete set of *Malus* accessions: MseI for region trnT/trnD and EcoRI for region trnH/trnK1 (Figure 3.1). All accessions were typed for both cp-regions. When all marker data were combined, 16 different chloroplast haplotypes (Cpht) could be defined (Table 3.2).

Table 3.2: Definition of 16 chloroplast haplotypes based on all mutational combinations. *: 0 = duplicationII absent, 1 = duplicationII present.

cp haplotype	*matKdupII**	*matK 338*	*HK1-EcoRI*	*DT-MseI*
1	0	T	1	1
2	0	T	1	4
3	0	T	1	6
4	0	T	2	1
5	0	T	2	2
6	0	T	2	3
7	0	T	2	4
8	0	G	1	1
9	0	G	2	1
10	0	G	2	2
11	0	G	2	3
12	0	G	2	7
13	0	G	2	8
14	1	T	1	1
15	1	T	1	2
16	0	G	1	2

3.1.3 RESULTS

3.1.2.1. Distribution of chloroplast haplotypes in different Malus species

Table 3.3 summarises the frequency of chloroplast haplotypes across species and populations. Eleven haplotypes were only detected in series *Malus*, two haplotypes (cpht2 and cpht3) only in series Baccata and one haplotype (chpt6) only in series Kansuenses. Although this indicates that some haplotypes can be useful to differentiate between the different series, these results should be taken with care given the limited number of accessions from series Baccata, Florentina and Kansuenses that were included in the study.

M. sylvestris and *M. domestica* shared eight haplotypes (cpht1, cpht5, cpht9, cpht10, cpht11, cpht14, cpht15, cpht16), although with different frequencies. However, the same three chloroplast haplotypes (cpht1, cpht10 and cpht14) accounted for more than 95% of accessions of each species. These haplotypes also occurred in other *Malus* species but only in one of the analysed *M. sieversii* accessions (cpht14). Two haplotypes (cpht4 and cpht8) were differentially found in *M. sylvestris* and *M. domestica*. Cpht4 was not present in *M. sylvestris*, but was shared by *M. domestica*, *M. prunifolia*, *M. baccata*, *M. mandschurica*, *M. toringoides* and *M. transitoria*. Cpht8, which is not present in *M. domestica*, is shared by *M. sylvestris* and *M. fusca*. Cpht11, which is the one found most frequently in *M. sieversii*, was also present in three *M. domestica* and four *M. sylvestris* genotypes.

The distribution of haplotypes 5, 9, 15 and 16, which were only detected in *M. domestica* and *M. sylvestris*, is remarkable. Cpht5 was detected three times: in one Danish *M. sylvestris* individual and in two Danish cultivars. Cpht16 was detected only in one Belgian *M. sylvestris* tree and one Belgian cultivar. Cpht9 was detected in five Belgian cultivars and four Belgian *M. sylvestris* individuals. Finally cpht15 was detected only in seven old Belgian cultivars and in one Belgian *M. sylvestris* accession. These results suggest a close relationship between the two species, even at the local level

Table 3.3: Frequency of 16 chloroplast haplotypes across *Malus* species and origins. Last two lines: frequency of chloroplast haplotypes in hybrid genotypes and in feral cultivars sampled in the nature and originally classified as *M. sylvestris*.. FL: Flanders (North of Belgium); WL: Wallonia (South of Belgium). Sec/Ser: section and series according to Phipps *et al.* (1990). M: *Malus*; B: Baccata; S: Sorbo*Malus*; F: Florentinae; K: Kansuenses

Population of origin	Sec/	Nb	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
M. sylvestris Belgium	M/M	225	0.60								0.01	0.15	0.02			0.21		
M. sylvestris Belgium	M/M	207	0.56									0.08				0.35	0.01	0.01
M. sylvestris Denmark	M/M	73	0.91				0.01			0.01		0.01				0.04		
M. sylvestris France	M/M	19	0.84													0.16		
M. sylvestris Germany	M/M	9	1.00															
M. sylvestris Central	M/M	9	0.77													0.22		
Total M. sylestris		542	0.64				0.00			0.00	0.00	0.09	0.01			0.23	0.00	0.00
M. domestica Belgium	M/M	139	0.10									0.34	0.01			0.55		
M. domestica Belgium	M/M	149	0.08								0.03	0.26	0.01			0.57	0.05	0.01
M. domestica Denmark	M/M	23	0.13			0.04	0.08					0.34				0.39		
M. domestica modern	M/M	86	0.19			0.02						0.08	0.01			0.69		
M. domestica cider	M/M	25	0.40									0.04				0.56		
Total M. domestica		422	0.13			0.00	0.00				0.01	0.24	0.01			0.58	0.02	0.00
M. kirghisorum	M/M	3											0.67	0.33				
M. orientalis	M/M	10											0.60	0.30	0.10			
M. prunifolia	M/M	9				0.22			0.33				0.11			0.33		
M. sieversii	M/M	20											0.85	0.10		0.05		
M. baccata	M/B	11	0.18	0.36	0.09	0.18			0.18									
M. hupehensis	M/B	3	0.66									0.33						
M. mandschurica	M/B	3		0.33		0.33			0.33									
M. florentina	S/F	1	1.00															
M. fusca	S/K	2								1.00								
M. kansuenses	S/K	1													1.00			
M. toringoides	S/K	3				0.66		0.33										
M. transitoria	S/K	2				1.00												
Hybrids M.sylvestris-M.	-	71	0.46			0.00					0.01	0.17	0.01			0.34		
Feral M. domestica	-	21	0.19			0.04						0.19				0.57		

3.1.2.2 Comparison of results at chloroplast and nuclear level

M. sylvestris accessions belonging to the three major chloroplast haplotypes (cpht 1, 10 and 14) were significantly different for introgression at nDNA level (F = 11.80, P < 0.001). Trees carrying cpht1 had significantly lower percentage of their genotypic information attributed to the cultivated gene pool than trees carrying cpht10 or cpht14 (Table 3.4). This value (0.087) was also lower than the average percentage of assignment to the cultivated gene pool found in *M. sylvestris* (0.13). The same analysis was done for *M. domestica* but for this species the groups carrying different chloroplast haplotypes were not significantly different (F = 0.001, p = 0.999) with regards to the percentage of assignment to the cultivated sample.

Table 3.4: Comparison of introgression at the nuclear level for *M. sylvestris* and *M. sylvestris* accessions belonging to the three major chloroplast haplotypes.

Cpht	Mean % assigned to M. domestica gene pool*	Standard error
M. sylvestris		
1	0.087 a	0.008
10	0.190 b	0.032
14	0.206 b	0.022
M. domestica		
1	0.950 c	0.010
10	0.949 c	0.113
14	0.949 c	0.005

*: based on nSSR and STRUCTURE results; a,b,c: different letters indicate significantly different values (P<0.05).

3.1.2.3 Within-species level of chloroplast diversity

Table 3.5 summarises diversity statistics for all samples containing more than 3 accessions. Among the *M. sylvestris* samples, the two Belgian origins contain the highest levels of gene diversity (h) and allelic richness (r). This could be indicative of greater admixture with other *Malus* gene pools than for the rest of the origins of *M. sylvestris* studied. Within *M. sylvestris*, the sample of modern cultivars exhibits the lowest diversity value but the sample of cider cultivars has the lowest allelic richness.

The average gene diversity and allelic richness are higher for *M. sylvestris* (h = 0.59 and r = 1.86) than for *M. sylvestris* (h = 0.33 and r = 0.97). The gene diversity found in *M. sieversii* (h = 0.28) is lower than the mean value for either *M. sylvestris* or *M. sylvestris*. The allelic richness in the M. sieversii sample (r = 1.05) falls in between the mean value for both species. Of all samples, M. prunifolia and M. baccata exhibit the highest gene diversity (h = 0.81 and 0.84 respectively) and allelic richness (r =

2.89 and 3.56 respectively). Finally, M. orientalis shows moderately high diversity with h = 0.600 and r = 1.80.

Among the cultivar groups, old cultivars comprise more cpDNA diversity than modern cultivars (h = 0.58, 0.60, 0.73 for different samples of old cultivars, and h = 0.49 for modern cultivars).

Table 3.5: Chloroplast diversity in *Malus* species and samples. FL: Flanders (North of Belgium); WL: Wallony (South of Belgium).

Population of origin	N	h	se(h)	r(8)
M. sylvestris Belgium (WL)	225	0.5701	0.029	2.86
M. sylvestris Belgium (FL)	207	0.5561	0.022	2.53
M. sylvestris Denmark	73	0.1582	0.057	1.63
M. sylvestris France	19	0.2812	0.116	1.83
M. sylvestris Germany	9	0.000	0.000	0.00
M. sylvestris Central Asia	9	0.3892	0.164	2.00
Overall M. sylvestris		0.515a		2.61
M. domestica Belgium (NBS)	139	0.579	0.026	2.63
M. domestica Belgium (CRA-W)	149	0.604	0.034	3.09
M. domestica Denmark	23	0.731	0.057	3.66
M. domestica Modern	86	0.489	0.055	2.62
M. domestica Cider apples	25	0.547	0.054	2.31
Overall M. domestica		0.594a		2.97
M. orientalis	10	0.600a	0.131	2.80
M. prunifolia	9	0.806b	0.089	3.89
M. sieversii	20	0.279c	0.123	2.05
M. baccata	11	0.836b	0.080	4.56

All parameters were calculated with CONTRIB (Petit *et al*. 1998). h: gene diversity, se(h): standard error of h; r(8): allelic richness after rarefaction to a common sample size of 8 genes. a,b,c and 1,2: different letters and numbers indicate significantly different levels of diversity. Letters refer to differences between species.

3.1.4 DISCUSSION

3.1.4.1 Overall Levels of Diversity

All samples considered in this study, with the exception of the *M. sylvestris* sample from Germany, displayed some degree of chloroplast diversity. *M. prunifolia* and *M. baccata*, displayed the highest allelic richness, while *M. sylvestris* and *M. sieversii* displayed the lowest values (0.974 and 1.05 respectively). These figures illustrate the high levels of polymorphism present in the genus *Malus* in the chloroplast regions analysed. However, the actual figures might be much higher, as we did not perform sequence analysis. Our choice was to develop simple PCR and restriction based tests which enabled us to analyse specific polymorphisms in numerous samples without the need to generate sequence information. By doing this we have unravelled

polymorphisms which might have remained undetected if sequence information was generated for only a limited number of individuals of each species, as is usually done. On the other hand, sequence information is still required to determine the levels of homoplasy present, especially in the restriction patterns of the trnH/trnK1 and trnD/trnT regions.

3.1.4.2 Relationship between M. sylvestris, M. sieversii and M. domestica

Our data suggest a much closer relationship between *M. sylvestris* and *M. domestica* than up to now appreciated, with the detection of three frequent chloroplast haplotypes (cpht1, cpht14 and, to a lesser extent, cpht10), which are shared by the two species. Moreover, the sharing of rare haplotypes by old cultivars of a given country and *M. sylvestris* collected in the same geographic area can be interpreted as an indication of gene exchange between the two species on a local scale. Surprisingly, the three main chloroplast haplotypes were almost absent in the analysed *M. sieversii* accessions, while this species has formerly been proposed as the most important maternal progenitor of the domesticated apple mainly based on DNA polymorphisms in the matK region (Robinson *et al*. 2001, Harris *et al.* 2002, Forte *et al.* 2002). This hypothesis should be reconsidered, as it is not supported by the results of the comprehensive survey of chloroplast DNA variation in *Malus* species presented here. Furthermore, our analyses of 10 nuclear microsatellite loci for the same genotypes (see Chapter 2), demonstrate that the nuclear genome of *M. sieversii* is almost as divergent from that of *M. domestica* as the genome of *M. sylvestris* (*Fst* = 0.09 between *M. sieversii* and *M. domestica* and 0.11 between *M. sylvestris* and *M. domestica*). The analysis of a larger sample of *M. sieversii* individuals might help to clarify this point.

In a similar study of chloroplast DNA variation in domesticated and wild *Prunus avium* genotypes, three haplotypes were detected among cherry cultivars (Panda *et al*. 2003). These haplotypes were also the haplotypes most frequently found in wild cherries, from which cultivars have been derived. However, as expected, the wild species showed higher chloroplast DNA diversity, with mutational combinations resulting in sixteen different haplotypes (Mohanty *et al*. 2001; Panda *et al*. 2003). Based on patterns of haplotype sharing between wild and domesticated *P. avium* trees, ancient and recent domestication events of this species were traced. Lower cpDNA diversity in cultivars compared with their wild relatives has been observed in many other plant species (e.g. in olives (Amane *et al*. 1999) and soybean (Xu *et al.* 2001)).

In the case of *M. sylvestris* and *M. domestica*, both species largely share eight haplotypes. A possible explanation is that in the past, local *M. sylvestris* genotypes

were mainly used as female parent in apple selection and breeding. This is further supported by the fact that some rare haplotypes are shared by cultivars and *M. sylvestris* populations with the same geographical origin. *Malus* trees with larger and tasteful fruits might have been introduced in these regions and used as pollinators. If these pollinators were *M. sieversii* genotypes (possibly originating from the Tien Shan region described by Juniper *et al*. (1999) and Harris *et al.* (2002) where they still grow in the wild), this would explain the intermediate position of *M. domestica* in-between *M. sylvestris* and *M. sieversii* as can be deduced from nuclear genetic information. The fact that old cultivars comprise more cpDNA diversity than modern cultivars (h = 0.58, 0.60, 0.73 for different samples of old cultivars, and h = 0.49 for modern cultivars) could also be the result of the former use of local trees in apple selection and breeding, whereas current breeding programmes focus on a limited already domesticated gene pool. The high cpDNA diversity detected in *Malus* cultivars could be further explained by the often proposed hypothesis of the complex hybrid origin of *M. domestica* (briefly reviewed in Robinson *et al*. 2001 and Hokanson *et al.* 1998) and thus admixture of genetic information.

However, the current pattern of chloroplast variation in *M. sylvestris* can also be partially caused by the cytoplasmic introgression of chloroplast haplotypes from the omnipresent domesticated apple, especially in Belgium. This could explain the relatively high levels of cp diversity of Belgian *M. sylvestris* samples (h equals 0.57 and 0.56 in the Belgian samples compared to the mean for *M. sylvestris* of 0.33), comparable to the levels of diversity found in domesticated samples (mean h = 0.59).

In conclusion, the present study indicates the occurrence of hybridisation between *M. sylvestris* and *M. sylvestris* but the nature of this relationship is difficult to establish. Probably the detected gene flow is bi-directional and brought about by the use of (local) wild *Malus* genotypes for the (local) cultivation process of apple and the later cytoplasmic introgression of chloroplast haplotypes into *M. sylvestris* populations from the domesticated apple. The present results also form a warning against drawing phylogenetic conclusions based on limited sequence variation of a limited number of genotypes and it reopens the exciting discussion on the origin of apple cultivars.

To further unravel the phylogenetic relationship between these apple species, one could look for the ultimate chloroplast mutation, unique to *M. sylvestris*, *M. sylvestris* or other apple species, sections or series. However, this might result impossible, if hybridisation is as common as our results indicate. More promising is the analysis of allelic associations at linked nuclear markers, as shown in 3.2.

3.2 ANALYSIS OF LINKAGE DISEQUILIBRIUM AT EIGHT SSR-LOCI

Preliminary remark: Detailed methodological information can be found in Koopman *et al*. 2006.

3.2.1 PLANT MATERIALS AND DNA-EXTRACTIONS

We analyzed a total of 204 pure *M. sylvestris* trees (including 124, 32, 32 and 16 trees from Flanders, Wallonia, France and Germany respectively), as well as 36 hybrids between *M. sylvestris* and *M. domestica* and 6 feral cultivars (as inferred from 10 SSR loci and *structure* analysis). Furthermore, 130 *M. domestica* genotypes were analysed, including 40 modern cultivars, 12 cider cultivars and 72 old varieties from the CRA-W collection (40) and NBS collection (32). For the present study, the *M. domestica* samples were treated as one population.

3.2.2 METHODS

We created a dataset containing eight microsatellites located at variable distance from each other on Linkage Group 10 of the *M. domestica* map (Silverberg-Dilworth et al. 2006). These markers are linked along one chromosome. Their genetic linkage, as derived from the frequency of recombination between the markers found in a segregating population, is expressed as centiMorgans (cM). The linkage group 10 markers are spaced at variable distances, up to 47.1 cM apart. As no linkage map of *M. sylvestris* has been published to date, we assumed that the map location on the microsatellite loci in this species was similar to their location in the *M. domestica* map. Table 3.6 provides general information on the SSR loci analysed. The loci were amplified in multiplexed PCRs as described in (Silverberg-Dilworth *et al*. 2006). PCR products were separated on an ABI Prism 377 DNA Analyzer or on a capillary AB 3130 DNA Analyzer (Applied Biosystems). Allele sizes were estimated with Genotyper or Genemapper software (Applied Biosystems) and manually inspected.

The presence of LD was analyzed using POPGENE 1.32 (Yeh & Wang 1999), as Burrow's composite measure of linkage disequilibrium between pairs of loci, and associated chi-square significance levels. Based on the results of the LD analyses (see Results section for details), data set LG10 was split into a data set with a group of only loosely linked loci (Ch02b07, Ch02a10, Ch02c11, Ch03d11, spanning 32.4 cM; further referred to as data set LG10U), and one with a set of tightly linked loci (U50187, Ch01f07a, Ms02a01, Ch02b03b, spanning only 1.5 cM; further referred to as data set LG10L).

We started by inferring the haplotypes from the sampled genotypes. Unlike *Arabidopsis*, which is nearly completely homozygous so that genotypes can be

derived directly from genotyping information (e.g., Toomajian *et al*. 2006), apple is an obligatory outbreeder. Haplotype-based studies in outbreeders typically follow a two-step procedure: first, haplotypes are inferred from a sample of phase-unknown genotypes using a computational algorithm, and second, inferred haplotypes are fed into the multilocus analysis, where they are treated as having been directly observed (Schouten *et al*. 2005). The latter step introduces some uncertainty (Schouten *et al*. 2005) and hence possible error (see Discussion). We inferred haplotypes using PHASE (Stephens et al. 2001) in its revised version (Stephens & Donnelly 2003).

We tested five sets of loci spanning an increasingly larger part of chromosome 10: 1) the four loci from LG10L (covering 1.5 cM); 2) LG10L + Ch03d11 (covering 14.7 cM); 3) LG10L + Ch03d11 + Ch02c11 (26.1 cM); 4) LG10L + Ch03d11 + Ch02c11 + Ch02a10 (36.2 cM); 5) all eight loci from LG10 (47.1 cM). For each set 20 runs with different starting points were performed. Each run consisted of 500 final iterations, 100 burnin iterations, and a thinning interval of 10. We employed the Parent-Independent Mutation model for all loci (-d1 option). The run with the highest overall likelihood was saved as the final result. Based on the results, we calculated the amounts of shared haplotypes between species and between populations.

Within each haplotype, each centiMorgan (cM) distance between two markers corresponds to a 1% chance that a marker allele at one locus was separated from a marker allele at another locus due to crossing-over (recombination) during meiosis in the cross between the commercial cultivars 'Fiesta' x 'Discovery' (Silverberg-Dilworth et al. 2006), i.e. in a single generation. As no linkage map of *M. sylvestris* has been published to date, we have assumed that the genetic linkage of the microsatellite loci in this species is similar to that of the *M. domestica* cross used to produce a segregating population. The haplotype results in this study are consistent with this order of markers. Haplotypes spanning smaller genetic distances will survive longer than haplotypes spanning larger genetic distances in populations. Unfortunately, the estimation of average survival time (in number of generations) of a haplotype across generations is not a straightforward extrapolation of recombination frequency within a generation, as the decay of LD due to actual recombinations is influenced by historical population size, population structure, and the occurrence of selection (Toomajian *et al*. 2006). If we assume the same recombination rate in each generation, the probability P that a given haplotype did not change from its ancestor G generations ago is $P = (1-r)\exp(G)$, with r = recombination and mutation rate (eq 1 in Stephens *et al*. 1998). When we assume r is comparable to the map recombination distance (cM/100), then 50% of the original haplotypes will theoretically be lost after 46 generations for the set of 4 loci (spanning 1.5 cM), 5 generations for the set of 5 loci (14.7 cM), 3 generations for the set of 6 loci (26.1 cM), and 2 generations for the

set of 7 (36.2 cM) or 8 loci (47.1 cM). We assumed 10 or 15 years as effective generation time (then 46 generations equals 460 or 675 years, in both cases well back into medieval times).

This gradual decrease in survival time for the different 'haplotype lengths' was used to reconstruct patterns of gene flow between *M. domestica* and *M. sylvestris*, by analysing the occurrence and proportion of shared haplotypes between the cultivar group and the wild populations. Almost all shared haplotypes had a (much) higher frequency in *M. domestica*, and they were therefore assumed to indicate introgression into *M. sylvestris*.

The average percentage of assignment to the cultivated gene pool (as defined by unlinked microsatellite data and structure analysis) was compared for two groups of *M. sylvestris* genotypes based on PHASE: trees carrying 'cultivated haplotypes' and trees without 'cultivated haplotypes'. 'Cultivated haplotypes' were defined as haplotypes that were detected at least once among the *M. domestica* accessions.

3.2.3 RESULTS AND DISCUSSION

3.2.3.1. General characteristics of the SSR loci analyzed

Levels of observed and expected heterozygosity estimated for the different loci were similar in *M. sylvestris* and *M. domestica*. One exception was locus Mald4.03a, for which the observed and expected heterozygosity were much lower in wild than in cultivated apple (Table 3.6).

Table 3.6: Summary information on the SSR loci typed

		Original publication		***This study M. sylvestris***		***This study M. domestica***	
Locus	LG (cM)	# alleles	Size range	# alleles	size range	# alleles	size range
CH02b07	10 (17.1)	7	180-202	12	92 - 127	14	92 - 142
CH02a10	10 (28.0)	6	143-177	10	140-170	13	138-176
CH02c11	10 (38.1)	7	219-239	13	208-240	15	20 -264
CH03d11	10 (49.5)	6	115-181	12	93-132	12	100-196
U50187	10 (62.7)	5	18 -204	14	174-212	13	174-210
CH01f07a	10 (63.6)	8	174-206	19	172-214	16	172-204
MS02a01	10 (64.2)	6	170-194	16	138-219	20	138-219
CH02b03b	10 (64.2)	8	77-109	20	71-116	16	71-106

3.2.3.2 Extent of Linkage Disequilibrium across the *M. domestica* and the *M. sylvestris* genomes

Figures 3.2a-b show, for all pairwise comparisons of loci, the number of allele combinations that were in significant linkage disequilibrium. LD on linkage group 10 was due, for the major part, to the four loci that are located within 1.5 cM of each other: Ch02b03b, Ms02a01, Ch01f07a, and U50187. Among these loci, LD of the pair Ch02B03b/Ch01F07a stands out for both *M. domestica* and *M. sylvestris*. Among the remaining LG10 loci, the pairs Ch02A10/Ch02B07 and Ch02A10/Ch02C11 show relatively high LD in *M. domestica*, but not in *M. sylvestris*. When the pairs of alleles in LD were expressed as permillage of the total number of allele combinations and plotted against the genetic distance along the linkage group (Figures 2c-d), a strong decrease in LD was observed with increasing recombination distance, as can be expected, but with some differences between *M. domestica* and *M. sylvestris*: 1) the amount of LD among the four closely linked loci is generally higher for *M. domestica* than for *M. sylvestris*; 2) the relatively high LD found in *M. domestica* for locus pairs Ch02A10/Ch02B07 and Ch02A10/Ch02C11 at approx. 10 cM distance is not present in *M. sylvestris*; 3) the level of LD among pairs of less closely linked loci is relatively even for both species, but slightly higher for *M. domestica* (0-30‰) than for *M. sylvestris* (0-20‰).

Based on these results, we subdivided data set LG10 into a data set with a group of loosely linked loci (Ch02b07, Ch02a10, Ch02c11, Ch03d11; data set LG10U), and one with a set of tightly linked loci (U50187, Ch01f07a, Ms02a01, Ch02b03b; data set LG10L).

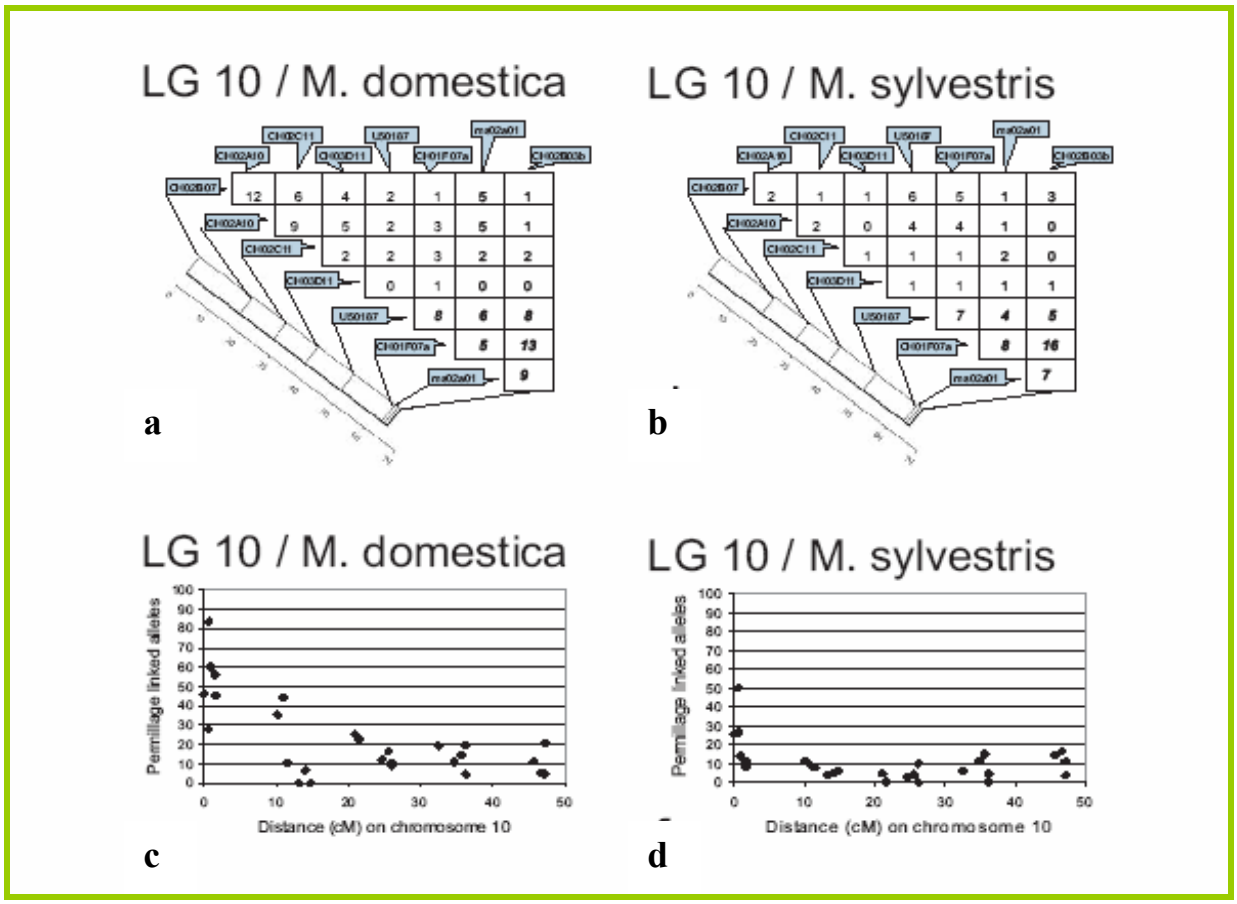


Figure 3.2: Linkage disequilibrium in apple.
a-b: pairs of alleles at two loci in significant LD for M. domestica and M. sylvestris.
c-d: relationship between genetic distance (cM) on linkage group 10 and the permillage of allele pairs in significant LD, across all loci of LG10.

3.2.3.3. Distinction of wild and cultivated apple using linked loci and haplotype frequencies and comparison of analyses using unlinked markers

Table 3.7 gives an overview of the total number inferred of haplotypes for haplotype lengths ranging from 8 tot 4 loci. We then defined cultivated haplotypes as haplotypes that occurred at least once among *M domestica* accessions. All other haplotypes are further referred to as *M. sylvestris* haplotypes.

Table 3.7: Number of haplotypes for haplotype lengths of 8 to 4 loci as inferred by PHASE.

Length of haplotypes	***Tot # Ht***	***# Ht M. sylvestris (´% shared Ht)***	***# Ht M. domestica***	***# Ht shared***
8 loci)	328	186 (1,0%)	110	2
7 loci	308	177 (1,1%)	99	2
6 loci	262	161 (4,3%)	85	7
5 loci	189	127 (7,1%)	54	9
4 loci	120	91 (16,5%)	40	15

Figure 3.3 shows the amount of cultivated haplotypes detected among *M. sylvestris*, *M. sylvestris* hybrids and feral cultivars for haplotypes consisting of 4, 5, 6, 7, and 8 loci, expressed as percentage of shared haplotypes, summed over all plants in a given group. Cultivated haplotypes that also occur in *M. sylvestris* are relatively rare, and nearly completely limited to haplotypes of 4, 5 and 6 loci. These haplotypes may have originated from *M. domestica* and have been introgressed into *M. sylvestris* populations, and this is supported by the fact that they are more frequent in *M. domestica* than in *M. sylvestris*. Overall, the proportion of shared haplotypes decreases rapidly with haplotype length, consistent with a longer survival times of "shorter" haplotypes (spanning shorter genetic distance). For all lengths of haplotypes, the proportion of cultivated haplotypes is lower for the 'pure' *M. sylvestris* category than for the identified hybrids and feral cultivars. Furthermore, structure analyses are designed to detect recent hybridization. The 'pure' *M. sylvestris* genotypes contain only two genotypes that carry a 8 loci-long (i.e. recently received) cultivated haplotype. These plant could therefore be direct hybrids, but also offspring of hybrids in which LG10 has not yet been destroyed by recombination. This suggests a certain agreement between the use of unlinked SSR spread over the different chromosomes and the linked SSRs in the identification of *M. sylvestris* genotypes with cultivated influence. However, the 'pure' *M. sylvestris* genotypes do contain a significant amount of old cultivated haplotypes (16,5 % carries a short cultivated haplotype of 4 loci), suggesting that historic introgression has taken place. This is in agreement with the chloroplast results where extensive sharing of chloroplast haplotypes has been shown.

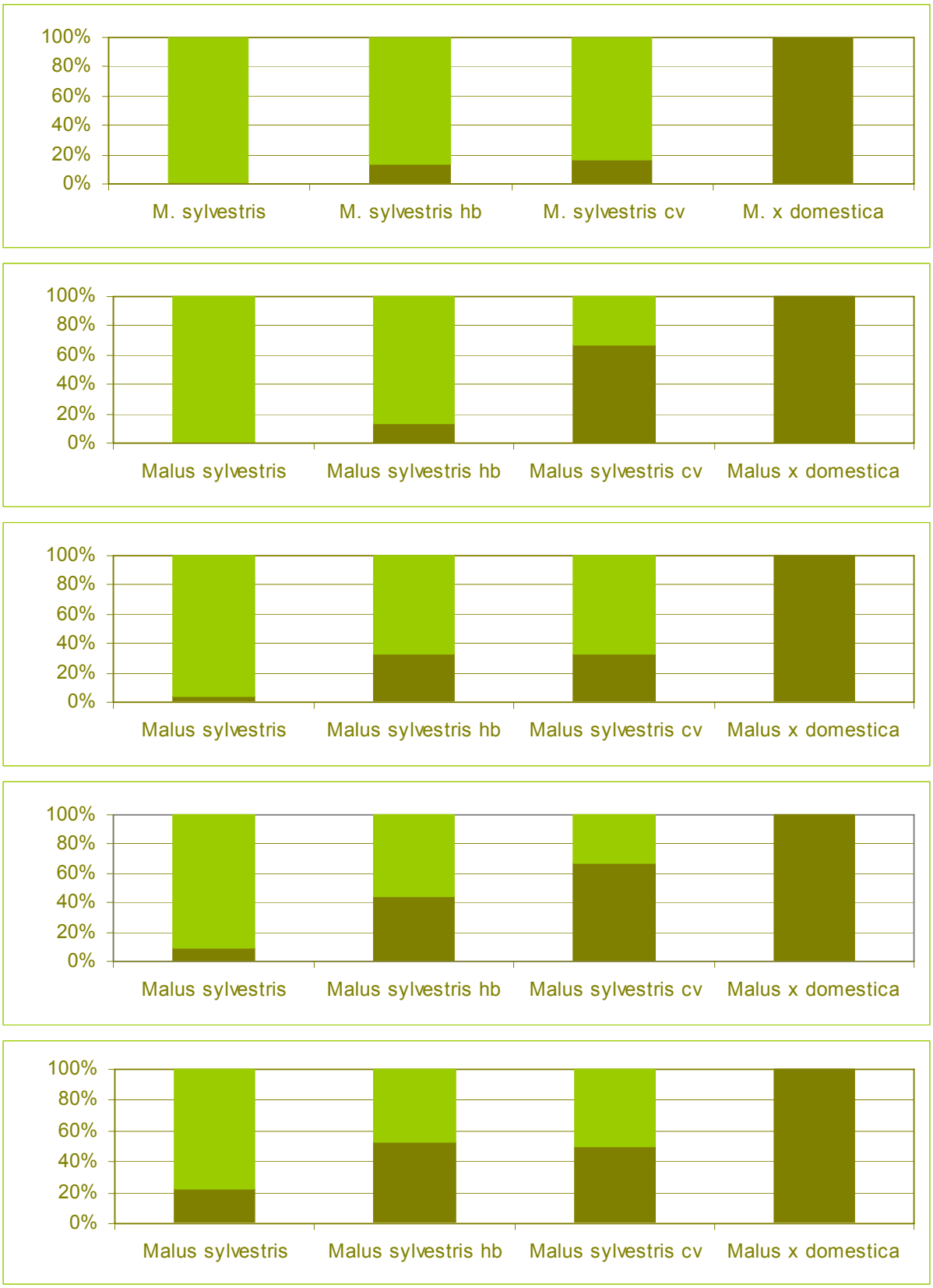


Figure 3.3: Distribution of haplotypes among species for haplotypes consisting of 8 (first panel) to 4 loci (last panel). *M. sylvestris* hb and cv: hybrids and feral cultivars respectively as inferred from unlinked SSRs and structure analysis.
■ = cultivated Ht; ■ = *M. sylvestris* Ht

3.2.4. Conclusions

Our analysis of sharing of multilocus microsatellite haplotypes from linked loci yields general information on population structure and population differentiation that is largely comparable with the results obtained with unlinked loci, as based on Bayesian analysis implemented in *STRUCTURE*. The strength of the method may be in generating information on the temporal scale of processes as multilocus haplotypes made it possible to distinguish past and recent hybridisation events. In addition, identifying a small number, or even one long haplotype shared between species indicates gene flow in a qualitative way, even when F_{ST} analysis, based on quantitative differences in allele frequencies, is not conclusive for instance because of low sample sizes. These examples demonstrate the usefulness and added value of sharing of multilocus haplotypes from linked microsatellite markers as a source of population genetic information.

The study of linked loci identified more accessions that might result from hybridisation with *M. domestica* than the analyses of unlinked SSRs. However, in most cases, these newly detected hybrids show signs of historic hybridisation, even dated in the medieval period. This approach thus provides further insight into the evolution of both species rather than identifying genotypes that have to be considered as true hybrids with *M. domestica*. Therefore, we propose not to take into account information on these historic introgression events when composing a core collection of pure *M. sylvestris* genotypes for future conservation (see chapter 5).

CHAPTER 4	Coordinator	**KUL-FTC**
	Other partners involved	**CRA-W**

STUDY OF FUNCTIONAL DIVERSITY: RESISTANCE AND FERTILITY TRAITS

INTRODUCTION

The use of evolutionary neutral genetic markers such as microsatellites is in rapid expansion in conservation genetics studies, as they allow the estimation of population parameters such as gene diversity, gene flow or population differentiation (Hendrick 2001, Haig 1998). However, the differences revealed by neutral molecular markers among wild populations, can be indicative of adaptive differentiation but can also result from random genetic processes (e.g. genetic drift, founder events) or limitations of the techniques applied. Therefore, the value of markers associated with potentially useful adaptive traits (functional markers) to study functional differences among populations and gene pools is becoming widely recognised (Hendrick 2001). However, the inclusion of this type of data in conservation genetics studies is still lagging behind. One of the reasons is the lack of knowledge about gene sequences in the species studied. The application of this kind of approach in the genus *Malus* was however possible, as sequence information for important gene families involved in e.g. disease resistance and fertility is available in international databases. In this project we aimed at developing functional markers for these two characteristics and use them to characterise *M. sylvestris* populations.

4.1 DEVELOPMENT OF GENE PROFILING TECHNIQUES

4.1.1 DEVELOPMENT OF GENE PROFILING TECHNIQUES: FERTILITY

4.1.1.1 Methods

Self-incompatibility is the genetically controlled mechanism which guarantees cross pollination and is the most effective reproductive barrier in angiosperms. The fact that apple cannot be fertilized by its own (or genetically related) pollen is determined by the action of genes at a single locus, which is highly polymorphic (the S-locus). The gene which is expressed in the pistil has been characterized in apple and other *Rosaceae* and encodes an S-RNase. This knowledge and the availability of apple S-RNase gene sequences in public databases made the development of a suitable fingerprinting technique a rather straightforward task.

Universal primers were designed to amplify all S-alleles present in a sample by means of a single PCR-reaction. This is not novel: for pear such an approach was already published in 1999 (Ishimizu *et al*.) and adapted for apple by Matsumoto and Kitahara (2000). Priming occurs within two highly conserved regions which are separated from each other by an allele-specific region and one intron. Specific bands for each allele, for which length polymorphisms are very likely, can be visualized as such by gel electrophoresis. These published primer sequences were our starting point. By comparing all available sequences (status January 2004) in the public databases, some adaptations were made to the original primer sequences in order to ensure the amplification of all S-alleles known at that moment.

In an additional approach, S-allele-specific PCR reactions were carried out for a sub set of known S-alleles (Broothaerts, 2003; Broothaerts *et al*. 2004; Van Nerum *et al*. 2001). For the visualization of S8, S10 and S16, additional primer sets were designed (Table 4.1) and amplification occurred under standard conditions, using 58ºC for annealing. PCR products were separated by either 4% agarose electrophoresis or capillary electrophoresis in case of the general S-genotyping. For the S-allele-specific assays, this was always done by 0.8% agarose electrophoresis.

Sequencing of fragments was done by cloning PCR-products in the pGEM-TEasy™ vector (Promega). Three positive clones were sequenced using the Thermosequenase cycle sequencing kit (USB) and a LiCor™ automated sequencer. A part of the fragments was sequenced by a direct sequencing approach on the cleaned PCR product. Primers used were standard (T7/SP6) or the FTQQYQ primer. Sequence determination was done by the e-Seq software (LiCor). Further analysis was performed with common bio-informatical tools involving the Blast algorithm and ClustalW analysis.

Table 4.1: primers used for S-allele genotyping and sequencing

Purpose	*Primer Set*	*Sequence*	*Fragment Size*
general S-genotyping	FTQQYQ	5'-TTTACGCAGCAATATCAG-3'	depending on S-genotype
	anti-I/MIWPNV	5'-ACGTTCGGCAAATA/CATT-3'	
S8-genotyping	S8(F)	5'-CGATTATTTTCAATTTACGCTTCA-3'	200bp
	S8(R)	5'-AGGTTGTTTCTTTGCAATACTCTG-3'	
S10-genotyping	S10-sense	5'-AACAAATCTTAAAGCCCAGC-3'	320bp
	S10-antisense	5'-GGTTTCTTATAGTCGATACTTTG-3'	
S16-genotyping	S16-5P	5'-GCGCACGGGTGTATACTTTT-3'	350bp
	S16-3P	5'-GCGTATGGCAATTTCAAGGT-3'	

4.1.1.2 Results

A pilot study was performed on a selection of commercial cultivars from KUL-FTC collection. This selection contained the majority of the presently known S-alleles and contained also the reference cultivars suggested by Broothaerts (2003). The approach gave good results: The PCR-reaction yields reproducible bands which – proven by sequencing- all are the predicted S-alleles. Non-specific gene products are rare and if so this can always be clearly discriminated from true ones. The results published by Matsumoto and Kitahara (2000) were confirmed. In addition results for other S-alleles were obtained (summarized in Table 4.2). The fragment sizes range from 318bp ("S_7") to 1400bp ("S_3").

For the S-RNase alleles "S_8", "S_{10}" and "S_{16}", problems were encountered with the general primer approach, resulting in no amplification at all or some very irreproducible large, faint bands (>2500bp). We suspect that the origin of the problem lies in the fact that the intron is too large, impeding proper amplification of the fragment. PCR conditions and type of *Taq* polymerase was checked but to no satisfying results. This problem can be dealt with by using specific primer pairs, designed to amplify a small region within the first exon.

Using the general approach, the size of the marker is very similar for a part of the S-alleles: in the region 340-350bp four S-alleles are located, in the region 360-370bp five S-alleles, and in the region 510-530 three S-alleles (Table 4.2). This makes it difficult to determine the exact genotype in these cases. We have optimised a PCR-method using specific primer pairs which made it feasible to detect 18 different S-alleles in apple. (Broothaerts, 2003; Broothaerts *et al*., 2004; Van Nerum *et al.,* 2001). We decided to combine this approach with the general marker system to confirm unclear data.

Table 4.2: Overview of results obtained by PCR-based S-allele genotyping used in this study. For the nomenclature of the S-alleles, we followed Broothaerts (2003). For others, we followed the name assigned in the Genbank accessions. *1: This work 2: Broothaerts, 2003, x: no specific PCR primers available.

S-allele	*PCR fragment size general approach*	**S-allele-specific assay*	*S-allele*	*PCR fragment size general approach*	**S-allele-specific assay*
S1	+/- 530bp	2	***S11=S14***	371bp	x
S2=S33	347bp	2	***S16***	?	1
S3	+/- 1400bp	2	***S17***	368bp	2
S4	338bp	2	***S19***	368bp	2
S5	+/- 1250bp	2	***S20***	508bp	2
S6=S12	365bp	x	***S21***	+/- 370bp	2
S7	318bp	2	***S22***	338bp	2
S8	?	1	***S23***	344bp	2
S9	344bp	2	***S24***	525bp	2
S10	?	1	***S26***	361bp	2

4.1.2 DEVELOPMENT OF GENE PROFILING TECHNIQUES: DISEASE RESISTANCE

4.1.2.1 Methods

The R-gene class includes a large number of genes involved in molecular processes associated with resistance against diverse pathogens in plants. Most functionally defined R-genes belong to the "nucleotide binding site – leucine rich repeat" (NBS-LRR) supergene family. In many plant species it was demonstrated that they are present in plant genomes as large multigenic families. Products of these genes are involved in signal transduction and protein-protein interactions. Due to the presence of well-conserved protein domains, putative members of this gene family may be conveniently isolated using primers corresponding to these domains (e.g. the NBS domain).

In order to test the possibility to develop a gene profiling technique for members of this gene family in apple, genomic DNA fragments were amplified using primer pairs where one primer corresponds to the retro-transposon LTR sequence and the other to well-conserved NBS-LRR domains. PCR-products were cloned and sequenced.

4.1.2.2 Results and discussion

Sequence comparison of two of the clones with nucleic acid databases revealed that their sequences are very similar to a *M. domestica* NBS-LRR sequence (AccNº AY369218, 97,5 % sequence identity), differing only by SNPs. Moreover, the regions corresponding to nucleotide positions 1 to 450 within these clones share 90% sequence identity with an apple R-gene homolog previously identified in cv.

"Freedom", as well as with other members of the NBS-LRR gene family from other species. Logically the sequences present at the other ends of these clones correspond to the *copia*-like LTR3' sequence. The sequence of one of these fragments is presented in Figure 4.1. The highlighted primer sequences (D1-3 and R 1-3) were selected according to the presence of conserved domains. These results demonstrate that a copy of an apple NBS-LRR gene, interrupted by a retro-transposon insertion, has been identified. Primers corresponding to well-conserved protein domain positions within this sequence (D1, D2, D3 and R1, R2, R3), were designed and used, in combination with retro-transposon LTR primers to develop a gene-profiling technique.

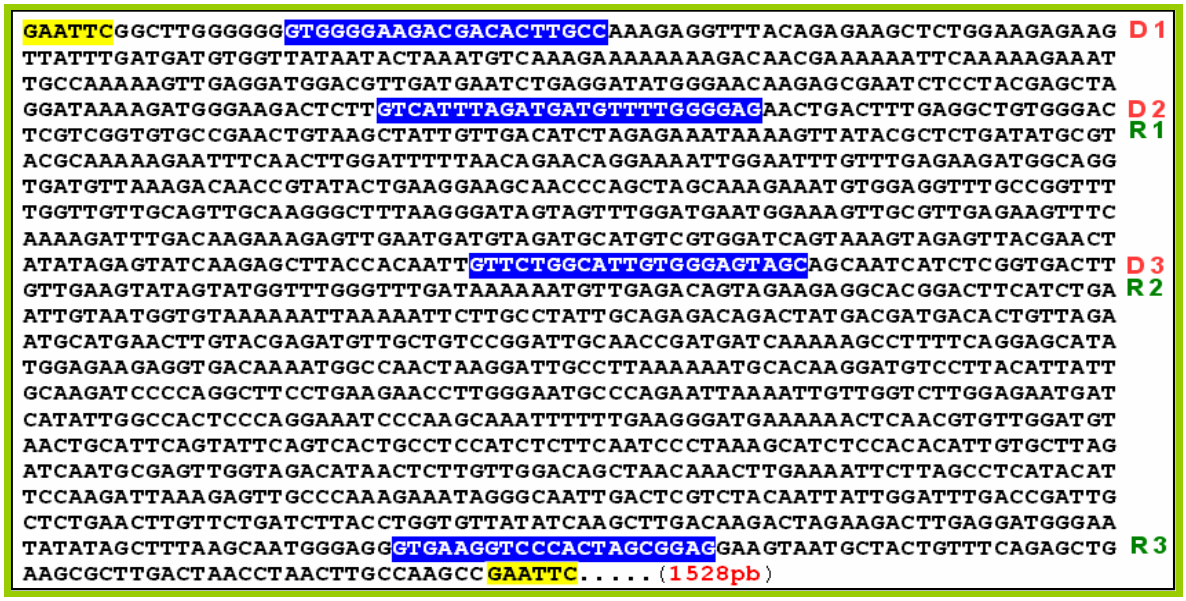


Figure 4.1: DNA sequence of one of the amplified fragments (clone 4) and localisation of the primers used for subsequent fingerprinting experiments

Several primer combinations were tested on genomic DNA samples corresponding to either apple cultivars or *M. sylvestris* individuals. Amplified fragments were separated by agarose gel electrophoresis and DNA fragments were visualized using ethidium bromide staining. The information content and reproducibility of the obtained profiles was evaluated. Representative profiles for 2 primer combinations giving polymorphic profiles are presented in Figure 4.2.

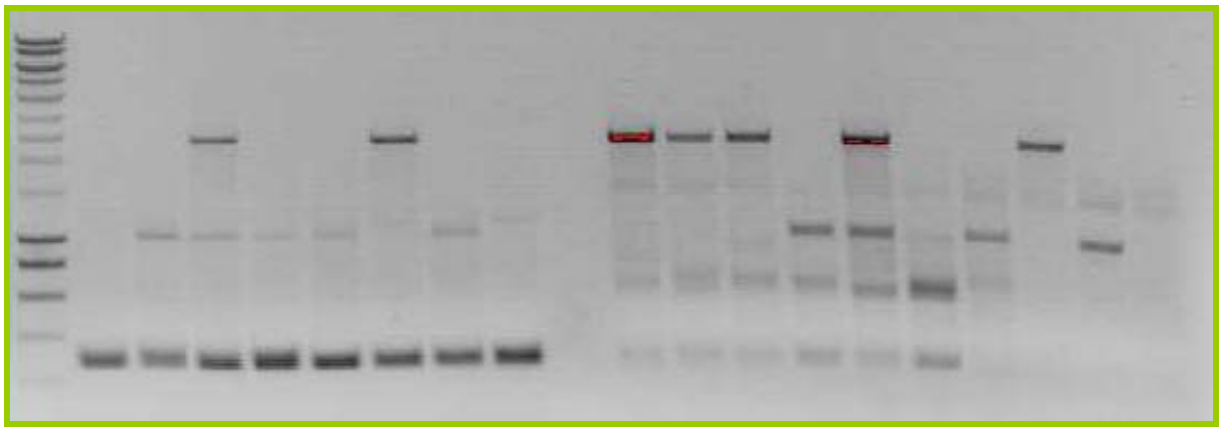


Figure 4.2: DNA fragments obtained by amplification of apple genotypes using D2 and LTR 3.2 primers (lanes 2-9) or D2 and LTR 3.1 primers (lanes 11-20).
Lane 1: molecular weight marker

It is therefore possible to obtain polymorphic DNA profiles using primer pairs which target well-conserved domains in NBS-LRR-type resistance genes and LTR3' *copia*-like retro-transposon sequences. Profiles obtained with the D3 – LTR3.1 primer pair turned out to be complex and polymorphic and this primer pair was selected for a first survey of the functional diversity of a subset of the apple genotypes. As a positive control, DNA was amplified from plasmid DNA containing the cloned insert.

4.2. VALIDATION OF MARKERS IN A SEGREGATING POPULATION

Since self-incompatibility is codified by a single locus which has already been mapped in apple by Maliepaard and co-workers (1998), this part of the work package was not applied for the self-incompatibility trait.

An attempt was made to produce a population by hand-pollination of an apple cultivar with *M. sylvestris* pollen in the spring of 2003. This yielded only a small number of fruits with seeds which were sown in the next winter period. Unfortunately no seedlings were obtained. In spring 2004, new pollinations were carried out by pollinating respectively 278 and 209 flowers from two apple cultivars with pollen from *M. sylvestris*. The cultivars were chosen by their sensitivity towards canker and scab disease. The resulting seeds were sown in the winter of 2004-2005 but unfortunately, no seedlings were obtained due to problems in the greenhouse. Alternatively, a segregating population of KUL-FTC was available. This consists of approximately 250 progeny plants derived from the cross between the cultivars Telamon and Braeburn. For both parents a genetic map is available based on AFLP and SSR markers. DNA of 100 individuals of this population was used for mapping.

The primers developed for amplification of NBS-LRR sequences as described in 4.1.2. were applied on these 100 genotypes. However, scoring of the marker patterns in this apple family was too complex and irreproducible. Attempts were made to deal

with these problems and to create better fingerprints. Unfortunately this was without any success. With the obtained results it was not possible to study functional diversity of disease resistance genes, nor was it possible to map disease resistance genes in a segregating population.

4.3 FUNCTIONAL DIVERSITY OF THE *MALUS* GENE POOL

4.3.1 PLANT MATERIALS AND METHODS

For the study of the functional diversity in the *Malus* gene pool a set of 325 individuals was used, which is representative for the whole collection used in this project. The selection consists of 225 *M. sylvestris* genotypes belonging to 12 populations in Flanders, Wallonia, France and Germany and of 100 *M. domestica* genotypes (34 modern and 66 old cultivars). Table 4.4 provides further details on the material used. As no reliable results could be obtained for the analysis of disease resistance, functional diversity was only tested for the self-incompatibility trait.

For the assessment of diversity of the S-alleles, in addition to visualization on agarose gels, sequencing was performed for those bands that might represent putative novel S-alleles.

4.3.2 RESULTS AND DISCUSSION

All 325 *Malus* accessions were subjected to the general S-allele genotyping method. After agarose gel electrophoresis we were able to asses the genotype for those S-alleles with a unique band size. For PCR-product sizes ranging from 350 to 380 and slightly larger than 500bp, additional S-allele specific assays were necessary. As a control, the majority of bands was sequenced (54%). This mostly confirmed our predictions. Some S-alleles, uncommon in commercial cultivars, but for which sequences have been deposited in the database, were detected, as well as the S_t- and $S_{g'}$ allele which were isolated previously from *Malus transitoria* (Matsumoto *et al.* 2000 and 2001). Moreover, also 20 novel sequences were brought to light (Table 4.3). For some sequences this was predicted by a distinctive length polymorphism, for others the novel S-alleles were found among bands in regions where known alleles were situated. Mostly their sequence is related to such S-alleles, but distinctively different at nucleotide level in the amplified exon sequence. In very few cases, we found S-alleles where the intron sequence was different but the exon sequence was identical. As they might correspond to almost identical proteins, they are considered to be an identical S-allele in further calculations.

Table 4.3: Overview of additional s-alleles and potential novel S-alleles and their PCR-product size in the general S-allele genotyping assay. Note: the nomenclature of the potential novel S-alleles is not definitive and based on annotation in sequence analyses (in most cases the apple S-allele with most homology)

S-allele	PCR fragment size general approach	Remarks	S-allele	PCR fragment size general approach
S29	450bp	cultivar unknown	***S19"***	+/- 360-370bp
S31	535bp	cultivar York Imperial	***S17"***	+/- 360-370bp
St	370bp	*Malus* transitoria	***S17'''***	+/- 360-370bp
Sg'	510bp	functionally identical to S20	***S17'***	+/- 360-370bp
S1'	+/- 520-530bp		***S24-400***	+/-400bp
Sg"	+/- 520-530bp		***Sg-400***	+/-400bp
S24'	+/-520-530bp		***S17-1250***	+/- 1250bp
S33'	+/- 350bp		***Sj-spadona***	+/- 950bp
S23'	+/- 350bp		***S29-700***	+/- 720bp
S26'	+/- 350bp		***S23'***	+/- 350bp
S10'	+/- 350bp		***S17''''***	+/- 360-370bp
S2'	+/- 350bp			

As the ploidy level is known for all individuals in the data set, it is known how many S-alleles have to be identified. For 20% of the accessions, not all S-alleles were amplified. This can be caused by (i) the occurrence of double bands on agarose gels; (ii) too long intron sizes for PCR amplification; (iii) homozygosity for the S-locus. Based on careful observations of the electrophoresis results, we assume that double bands might be present in 26% of the genotypes and too long PCR products in 33%. However, for 41% the reason is not identified. We assumed that *M. sylvestris* carries a fully functional self-incompatibility system and high degrees of heterozygosity were expected at the S-locus in this species. However, this assumption is mainly based on studies in commercial cultivars and the degree of homo- or heterozygosity at the S-locus in *M. sylvestris* populations is difficult to investigate, as the amplification of one single allele in a PCR reaction can have different explanations. Comparisons with levels of heterozygosity at neutral SSR loci (as the ones used in this study, Chapter 3) can help to clarify this point in the future.

Among the 225 *M. sylvestris* accessions, 13 genotypes were identified as hybrids with *M. domestica* based on SSR typing and STRUCTURE analysis (Chapter 2). These trees were omitted from the dataset for further analysis. Figure 4.3 gives an overview of the frequency of S-alleles in *Malus domestica* and *Malus sylvestris*.

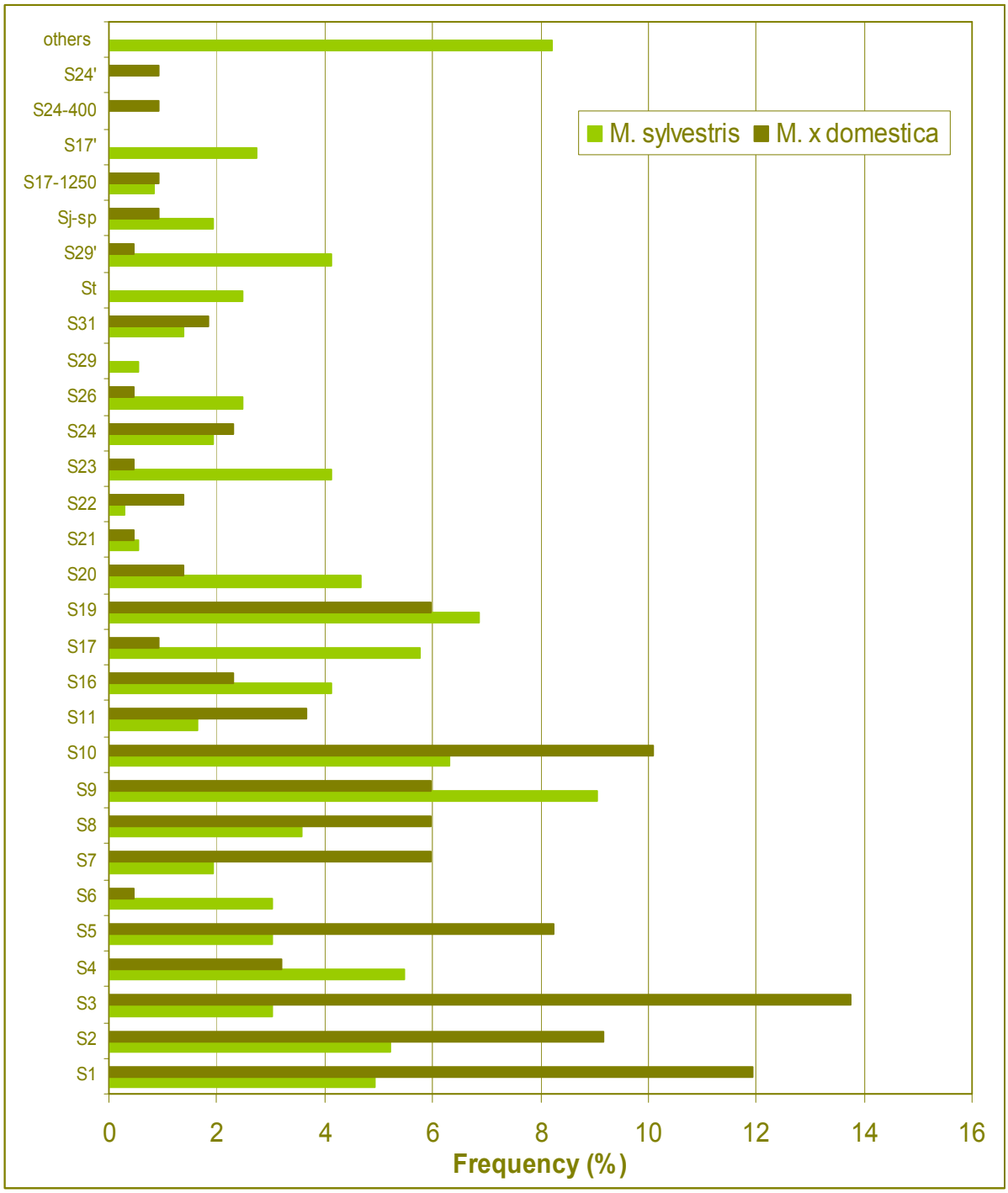


Figure 4.3: Frequency of S-alleles in *Malus sylvestris* (panel A) and *Malus* domestica (panel B). Note: $S_{g'}$ is considered as S_{20}

For *M. domestica* dominance of alleles S_1, S_2, S_3 and S_9 was observed. This was also found by Broothaerts *et al*. (2004) who observed a much larger dataset (150 cultivars). The dominance of these four S-alleles is probably caused by the fact that the majority of the commercial cultivars are derived from a small number of parents carrying these alleles, such as Golden Delicious (S_2S_3). The alleles S_6, S_{17}, S_{21}, S_{23} and S_{26} occur with a very low frequency. The presence of S_{11} and S_{19} can be attributed to the old cultivars. S_t and S_{29} were not detected at all, whereas they are

present in some of the *Malus sylvestris* individuals. The majority of the novel S-alleles was found in the wild species. Also in *Malus sylvestris*, S_1, S_2 and S_9 are important S-alleles. In addition, S_4, S_{10}, S_{17} and S_{19} are quite frequent.

Table 4.4 shows the amount of S-alleles found in the different species and populations. It is clear that diversity of S-alleles is present in all populations. However, in spite of the allelic diversity detected, in some populations a significant part of the individuals shares one S-allele.

Table 4.4: Number of S-alleles in different subpopulations

Malus domestica		***# individuals***	***# S-alleles***
Modern cultivars		66	24
Old cultivars		34	14
Malus sylvestris	***Population***		
Flanders	Meerdaalwoud	65	33
	Heverleebos	14	10
	Voeren	8	9
Wallonia	Houyet	20	13
	Rance	20	11
	Treignes	19	13
	Verviers	4	7
	Vielsalm	14	18
	Spa	10	12
	La Roche	6	11
Germany		12	12
France		19	22

In the future, promising alternatives can be developed for apple. Recently, the male S-gene has been cloned for the first time in a *Rosaceae* species and encodes an F-box type of gene which is also highly divergent (*Prunus*). Vaughan and co-workers (2006) used this sequence information and combined it with the knowledge of the S-RNases to develop a PCR-based technique in *Prunus* which is similar as the one used here but is based on both genes. The combinations of the respective sizes for the two resulting fragments (S-RNase and F-box gene) enables a more accurate identification of the S-genotype involved. As such they evaluated a large set of *Prunus* individuals by one single PCR. However, in *Prunus* the size differences of the fragments are limited between 200 and 400 bp, which is much smaller than in apple.

4.4 CONCLUSIONS

By a straightforward, PCR-based technique we were able to visualize S-genotypes as polymorphic bands. However, additional specific PCR-assays were necessary to retrieve the precise S-genotype for a part of the S-alleles of similar length, impeding proper determination. The fact that sometimes the fragments are too large is another reason to combine the two PCR-based methods. Currently, the use of capillary electrophoresis for more accurate size determination is studied. But, given the length of some S-alleles (up to almost 1.5kb) also this technique will have limitations. The combination of the different approaches remains necessary.

For a dataset of 325 individuals the S-genotype was determined. Clear differences were observed between *M. domestica* and *M. sylvestris*, although they shared 24 S-alleles. Approximately 20 novel S-allelic fragments were isolated and sequenced, mostly originating from *M. sylvestris* genotypes. For 20% of accessions, not all S-alleles were detected. This could be caused by the presence of unknown S-alleles with a too large intron for proper amplification or double bands on agarose gels. However, the possibility exists that the S-genotype is homozygous in some genotypes. For the different local populations of *M. sylvestris*, a rather large diversity of S-alleles in comparison to the amount of individuals was observed, which is positive for the generation of offspring. However, for example in "Meerdaalwoud" a significant part of the individuals shares one S-alleles. This means that half of the male gametes will not be able to fertilize flowers and give raise to fruit. Under conditions of high pollen availability, this will give comparable fruit set compared to full compatibility. However, when the pollen availability is not optimal (e.g. bad weather conditions, no (semi) compatible trees in the near vicinity) this could cause problems towards the viability of the population.

The developed gene profiling techniques demonstrate the possibility to obtain polymorphic DNA profiles using primer pairs which target well-conserved domains in NBS-LRR-type resistance genes and LTR3' *copia*-like retro-transposon sequences. Profiles obtained with the D3 – LTR3.1 primer pair turned out to be complex and polymorphic and this primer pair was selected for a first survey of the functional diversity of a subset of the apple genotypes. However, scoring of the marker patterns in the 'Braeburn' x 'Telamon' family was too complex and irreproducible. Attempts were made to deal with these problems and to create better fingerprints. Unfortunately this was without any success. It was therefore not possible to study functional diversity of disease resistance genes, nor was it possible to map disease resistance genes in a segregating population.

CHAPTER 5	Coordinator	**INBO**
	Other partners involved	**CRA-W, KUL-FTC, NBS, ILVO-PLANT**

DEVELOPMENT OF CONSERVATION STRATEGIES AND DISSEMINATION OF RESULTS

PUBLICATIONS

Lateur, M. [2003]. The integration of different sectors is a key factor for the conservation, the evaluation and the utilisation of our Belgian fruit tree biodiversity. *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique – Biologie,* 73-SUPPL.: 85-95.

Villette I., Lateur M. & Delpierre L. [2003]. Création d'un réseau wallon de conservation *in situ* de ressources génétiques fruitières. *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique – Biologie* 73-SUPPL.: 97-101.

Jacques D., Lateur M., Watillon B., Lemaire S., Coart E., Roldan Ruiz I.,Vander Mijnsbrugge K., Vanwijnsberghe L.& Keulemans W. [2003]. Développement d'un programme de gestion de la diversité génétique du pommier sauvage (*Malus sylvestris* Mill.) en Belgique : application en Région Wallonne. *Les Naturalistes belges*, 84, 2-3-4: 149-161.

Lateur M., Blazek J. [2004]. Evaluation descriptors for *Malus*. *In*: Maggioni L., M. Fischer, M. Lateur, E.-J. Lamont and E. Lipman, compilers. Report of Working Group on *Malus/Pyrus*. Second Meeting, 2-4 May 2002, Dresden-Pillnitz, Germany. *International Plant Genetic Resources Institute*, Rome, Italy: 76-82

Lateur M. [2004]. Research in Belgium on wild *Malus*. *In*: Maggioni L., M. Fischer, M. Lateur, E.-J. Lamont and E. Lipman, compilers. Report of a Working Group on *Malus/Pyrus*. Second Meeting, 2-4 May 2002, Dresden-Pillnitz, Germany. *International Plant Genetic Resources Institute,* Rome, Italy: 67.

INTRODUCTION

As demonstrated in previous Chapters, this study has generated an extensive description of the biodiversity of the Belgian *Malus* gene pool, revealed information on the present distinctness between wild and cultivated apples and given insights into past and present hybridisation events. Based on these findings, we have defined (and implemented) in-situ and ex-situ conservation guidelines for *M. sylvestris* in Belgium. Furthermore, the usefulness of wild apple trees and regional varieties to expand the genetic basis of current apple breeding programs has been determined.

5.1. CONSERVATION STRATEGIES FOR *MALUS SYLVESTRIS*

5.1.1 GEOGRAPHIC DISTRIBUTION AND PHENOTYPIC AND GENETIC DIVERSITY

Especially in Flanders, local populations are severely reduced in size. Total extinction is thus a serious threat for all these populations, except for one large population in Meerdaalwoud. In the Walloon region *M. sylvestris* is more common and populations mostly consist of larger numbers of individuals in comparison to the situation in Flanders. No immediate menace of extinction exists.

Three main phenotypic characteristics were observed: leaf and twig hairiness and fruit size. It became clear that twig hairiness is less heritable in comparison to leaf hairiness and fruit size. None of the morphologic descriptors (or a combination of descriptors) could be used for reliable identification of hybrids. Differences in bud burst were found and although results were not significant (probably due to the small population sizes and relatively large variances), a geographic-based variation was suspected in the Walloon region for this trait. Because of this result, different regions of provenance were defined for *M. sylvestris* in Belgium.

From the results of the cpDNA and self-incompatibility analyses it was concluded that the gene pool of *M. sylvestris* and cultivated apples are closely related. However, these data are not directly informative for the discrimination between wild and hybrid genotypes. The cpDNA analyses of a diverse sample of *Malus* species provided, however, new insights on the origin of the cultivated apple. The analysis of S-alleles has resulted in the identification of novel alleles in *Malus*.

Based on the SSR results, it was possible to distinguish between wild types, hybrids and cultivated apple trees (feral cultivars). In Wallonia more hybrids and cultivated apple trees were encountered in comparison to Flanders, but this may reflect the less strict selection criteria followed to sample trees in the field. No geographic structure was found in relation to genetic relationships among Belgian populations. For the construction of a core collection of *M. sylvestris* it is of utmost importance to discriminate between wild, hybrid and cultivated apple trees. We therefore combined phenotypic (based on leaf hairiness and fruit size) and SSR classification where possible and otherwise used only SSR classification to identify genuine *M. sylvestris* trees and to construct a core collection for future conservation of the species.

5.1.2 CONSTRUCTING A CORE COLLECTION FOR *MALUS SYLVESTRIS*

A core collection is a sub-sample of a larger germplasm collection that contains, with a minimum of repetitiveness, the maximum possible genetic diversity of the species

in question. In this case, the target population corresponds to the wild type apple trees growing in Belgium.

The selection of individuals for the core collection of *M. sylvestris* has been set-up in several steps. As described in the previous Chapters, around 600 trees spread over the Belgian territory were characterised. Among these 600 trees, 115 genotypes were identified as genuine wild type by phenotypic (fruit diameter <35 mm, leaf hairiness <5 on a scale of 1 to 9) and genetic data. Two-hundred twenty-seven genotypes for which no phenotypic data were available, were identified as genuine wild type by SSR analysis alone.

Of these 342 (115+ 227) trees, 201 trees were selected for the "Belgian *Malus sylvestris* core collection", based on various selection criteria. We suspect some degree of structuring in the *M. sylvestris* gene pool, based on phenological observations (bud burst), corresponding to regions of provenance. These are characterised by specific ecological conditions and geographical characteristics. Therefore, the core collection is stratified into 4 groups: 'North of Sambre and Meuse line', 'Low plateau of Meuse', 'Ardenne' and 'Gaume'. Each group contains an optimal number of genotypes (between 60 and 70) that are geographically distributed all over the region of provenance. Ideally, each group contains the whole range of variability found in total collection of trees in the corresponding region or provenance.

5.1.3. COMPOSITION OF THE CORE COLLECTION

Tables 5.1 and 5.2 show the total number of wild trees found in the four regions of provenance and the number of trees included in the core collection. Table 5.3 illustrates the large phenotypic diversity present in the core collection.

Table 5.1: Numbers of wild type genotypes in the total Belgian collection

Region of provenance	*Wild type 1*	*Wild type 2*	*Total*	*In core collection*
North of S & M	26	175	201	61
Low plateau of Meuse	40	37	77	61
Ardenne	76	45	121	61
Gaume	2	15	17	18
Total	144	272	416	201

Wild type 1: according to phenotypic and genetic data; Wild type 2: according to SSR data only

Table 5.2: Detailed selection of trees in the core collection from four provenance region

District	Wild type 1	Wild type 2	Total
North of Samber & Meuse			
Antwerpen	2	2	4
Brugge	0	5	5
Hasselt	0	1	1
Leuven	14	32	46
Namur	5	0	5
Subtotal N Sambre & Meuse	21	40	61
Condroz & Famenne			
Aywaille	2	0	2
Chimay	2	4	6
Couvin	1	3	4
Dinant	1	4	5
Eupen	0	1	1
Liège	6	0	6
Marche	0	3	3
Namur	4	1	5
Philippeville	5	1	6
Rochefort	3	2	5
Thuin	3	3	6
Verviers	1	5	6
Viroinval	0	1	1
Wellin	2	3	5
Subtotal Condroz & Famenne	30	31	61
Ardenne			
Aywaille	1	0	1
Beauraing	2	0	2
Bertrix	2	0	2
Bièvre	3	0	3
Bouillon	1	2	3
Bullange	2	1	3
Chimay	1	0	2
Elsenborn	4	0	4
Eupen II	4	0	4
Florenville	3	0	3
Habay-la-Neuve	2	0	2
La Roche	4	0	4
Libin	0	1	1
Marche	2	1	3
Nassogne	2	0	2
Neufchâteau	4	0	4
Paliseul	3	0	3
Spa	3	0	3
St Hubert	1	0	1
St-Vith	1	1	2
Verviers	0	2	2

Vielsalm	1	2	3
District	***Wild type 1***	***Wild type 2***	***Total***
Viroinval	1	1	1
Wellin	3	0	3
Subtotal Ardenne	50	11	61
Gaume			
Arlon	1	5	6
Florenville	0	2	2
Virton	1	9	10
Subtotal Gaume	2	16	18
Total Belgium	103	98	201

Wild type 1: according to phenotypic and SSR data; Wild type 2: according to SSR data only

Table 5.3: Variability present in the core collection of 201 *M. sylvestris* trees

	North of S & M			***Condroz & Famenne***			***Ardennes***			***Gaume***		
	Mean	***Min***	***Max***	***Mean***	***Min***	***Max***	***Mean***	***Min***	***Max***	***Mean***	***Min***	***Max***
Height (m)	10.1	1.5	20	11.0	4.5	20.3	10.1	3.8	17.5	15.0	7.4	20.5
Circumfer. (cm)	53	3	154	69.6	29	177	89	14	234	97	68	152
2005 Twig hairiness	-	-	-	2.9	1	6	3.1	1	5	-	-	-
2005 Leaf hairiness	-	-	-	2.4	1	4.5	2.4	1	3.3	-	-	-
Leaf scab	2.7	2	5	3.1	1	6	-	-	-	-	-	-
Fruit diameter	29	22	35	28.8	19	34	27	17	35	-	-	-
Fruit scab	3.2	1	6	2.8	1	6	3.6	1	7	-	-	-
Cancer	-	-	-	1.1	1	4	1.4	1	4	1.2	1	4

5.1.4. *EX-SITU* CONSERVATION

A seed orchard (1 ha) containing all wild types from the provenance region North of Sambre and Meuse, has been planted in Dentergem (Flanders). This place is owned by the Flemish community. The plants are grafts on M9 root stock. This plantation includes not only the genotypes selected for the core collection but all collected wild type individuals for this provenance region. In total 194 genotypes are present, all in two copies. Planting distance is 5 m by 5 m.

The purpose of the plantation is twofold: conservation of the genotypes and production of seed for the production of planting stock. The seed orchard will be officially approved as a 'seed source' under the category 'source identified'. The latter means that no information is available on stem and wood characteristics. In the case of *Malus sylvestris*, this is not relevant. The approval as a seed source gives private nurseries the possibility to certify the harvested seeds. The ultimate purpose is to

stimulate private nurseries to collect these seeds and grow planting stock, so that it will become commercially available.

At FTC in Rillaar an orchard is planted with trees from the core collection of *M. sylvestris* and a representative subset of old varieties from the NBS collection. Trees were planted during winter 2004-2005 at 3.5 by 0.75 m. These trees will be used in 2006 for observations on disease susceptibility (scab, mildew) under field conditions. After 2006 the NBS trees will be replanted at a new location under the supervision of NBS and remain available for further characterisation and genetic studies. Trees of the core collection of *M. sylvestris* will be used later on for a screen of functional genetic diversity (ripening genes, self-incompatibility genes, resistance genes,….) and for breeding.

The plantation of a National Core orchard of Belgian *M. sylvestris* has been initiated in spring 2006 in Philippeville (Wallonia). A total of 279 trees representing 112 genotypes are already planted (Figure 5.1). The core orchard of 1 ha is planned to be finalised in spring 2008 and will include the complete core collection (see Table 5.2). In total, five copies of all 201 genotypes of the core collection will be planted with planting distances of 4 m by 2.5 m. This plantation will be used as a seed orchard to produce pure *M. sylvestris* seeds for forest and landscape purposes. The orchard is owned by the Ministry of Walloon Region, Division of Nature and Forests.

Figure 5.1: Planting of the Belgian *M. sylvestris* "Core orchard" during spring 2006 at Philippeville.

Finally a subset of 15 genotypes of *M. sylvestris* are entering in the CRA-W collection and will be further evaluated for their use in breeding programs.

5.1.5 IN SITU CONSERVATION

5.1.5.1. Flanders

At several locations in Flanders only one or a few wild type trees are still present. These populations are too small in size to survive over a long time period. The only way to preserve these populations is restocking with plant material originating from the proper provenance region. In addition, it is absolutely necessary to cut hybrids and cultivated apple trees in the immediate surroundings.
In Wijnendale forest, in the province of West-Flanders, a restocking project has already started. Only 5 wild type individuals are still present in the forest. Around 50 apple trees that were planted as wild types some 20 years ago, turned out to be cultivated apple trees and will be cut in 2006-2007. Also, a restocking will take place in 2006 with 25 wild genotypes that were grown at the nursery of INBO. These plants originate from Meerdaal forest. They will be planted at the southern border of the forest in a stand that will be managed as coppice with standards. In this way the trees will have the possibility for fructification.

Apart from Wijnendale forest, four other locations in Flanders concern relict populations of wild types that grow in or near forests of public ownership. Restocking projects will be elaborated through information of the stake holders and the growth of cuttings or seedlings of wild types.

In Meerdaalwoud, a forest owned by the Flemish community, the rangers are now aware of the importance of this exceptional large population of *M. sylvestris*. Fructification of the trees will be promoted through cutting of surrounding trees and improving light conditions. Here, trees are up to 20 m in height.

5.1.5.2. Wallonia

The *in situ* conservation strategy in Wallonia includes two main objectives: the follow-up and the conservation of the *M. sylvestris* trees located in the framework of this project and the selection of three stands with a large population compatible with the seed collection for forestry purposes. To reach the first objective, the forest rangers will be invited to promote the different *M. sylvestris* trees identified in their district by adapted thinning to favour their growth and their fructification. Moreover, special attention will be paid to all other wild apple trees at the time of the thinning to support their conservation and development in the forests where it is still present. Furthermore, three specific stands have been identified and selected. They are described in Table 5.4.

Table 5.4: Description of the three stands selected as seed source.

General description	Stand Sivry-Rance	Stand Houyet	Stand Treignes
District	Thuin	Dinant	Viroinval
Owner	Walloon region	Community	Community
Acreage	4.5 ha	10 ha	4.5 ha
Dominant circumference (cm)	96.8	91.2	102.5
Dominant height (m)	17.4	12.9	15.6
Mature tree number	95	69	70

We will encourage the forest rangers to manage these three stands by thinning of the trees in competition with the *M. sylvestris* to promote their development and seed production. The final purpose is to give the opportunity to the Walloon Seed Centre and the private nurseries to collect these seeds, so that it will become commercially available. To give the possibility for official control of the forest reproductive material produced in these stands, the possibilities of including this species in the list of species submitted to control will be evaluated.

5.2. CONSERVATION STRATEGIES FOR *MALUS DOMESTICA*

5.2.1 THE CONSERVATION OF AGRO-BIODIVERSITY

Future plant selection and breeding efforts - not only for disease or pest resistance but also for fruit taste, specific fruit uses, tree habit, adaptation to low fertilizing, etc... - are dependent on the genetic variation which is stored in old cultivars and some of their wild relatives. Agro-biodiversity is therefore a genetic heritage of the past that must be included in a dynamic and continuous process for a better adaptation to the present constrains and requirements of commercial cultivars. It represents an immeasurable potential value for the present and future generations.

Since the Belgian ratification, in November 1996, of the legally binding 'Convention on Biological Diversity' (CBD) and moreover since its adoption in June 1996 of 'The Global Plan of Action for the Conservation and sustainable Utilization of Plant Genetic Resources for Food and Agriculture' (GPA), under the auspices of the 'Food and Agriculture Organisation' (FAO), many topics must be put in this new context. Belgium has more than ever a moral obligation to set-up a National Programme and to provide '*ad hoc*' financial support for planning and coordinating the conservation and the utilization of its 'Plant Genetic Resources' –also called 'agricultural biological diversity'- which are a part of the global biodiversity. Specifically on the agricultural matters, the CBD ask to each country member of the Working Party to implement the

Convention for the conservation and sustainable use of agricultural biological diversity by setting up and/or making progress in the following main actions:

- Implementing a national strategy, programmes and plans which ensure the development and successful implementation of policies and actions that lead to sustainable use of agrobiodiversity components;
- Establishing or enhancing mechanisms for increasing public awareness and understanding of the importance of the sustainable use of agrobiodiversity components;
- Helping to implement the Global Plan of Action for the Conservation and Sustainable Utilization of Plant Genetic Resources;
- Promoting regional and thematic co-operation within this framework of the programme of work on agricultural biological diversity;
- Providing financial support for implementation of the programme of work on agricultural biological diversity;
- Co-ordinating its position in both the Convention on Biological Diversity and the International Undertaking on Plant Genetic Resources.

The 'Global Plan of Action' goes into more detail and strengthens the implementation of "National programmes for the conservation and sustainable utilization of plant genetic resources for food and agriculture" (Cooper *et al.* 1998). Many European countries have already developed regional and national structures which have the responsibility to manage the genetic resources (Lefort *et al*.1998; Schierscher, Kleijer 1999). In Belgium there is very limited initiative in this area in the "National strategy" designed by the CBD National Focal point.

5.2.2 CURRENT SITUATION OF *MALUS DOMESTICA* IN BELGIUM

M. domestica is one of the most important cultivated fruit species in the world, and between 10,000 and 15,000 cultivars are described in the literature. One of the most important threats that raised the lasts decades, is the drastic erosion of the genetic base of the commercial cultivated apple. Less than five commercial cultivars represent more than half of the world production (~70 millions tons). This narrow base is also caused by of the fact that the cultivar 'Golden Delicious' is the most important progenitor of most of the commercial cultivars. The origin of the species is complex and is probably the result of several interspecific crosses between different wild species that have been selected by men throughout time and space. Apple is clonally propagated by using grafting techniques which combine scion buds and different rootstocks that controls tree vigour and soil conditions adaptation.

Belgium is proportionally to its surface, very active in the collection and the management of apple genetic resources. Many specific cultivars have been raised in de past in Belgium by tremendously active peasant population. This category of

material is ranged in the terminology "landraces". In the public sector, the CRA-W started a research programme on Belgian fruit genetic resources from 1975 onwards and collected more than 1,650 apple accessions that are systematically evaluated for disease resistance and agronomic characters (Lateur, 2003). The NBS association also collected a lot of accessions (~1,500) that are mostly planted as high stem trees in view of the development of orchards with landscape value. The KUL-FTC developed a work collection for their breeding programme that contains mostly commercial cultivars (cvs) but also old cvs (~90). Our large fruit biodiversity constitutes an historic, ethnobotanic and genetic heritage with a high potential value for breeding work.
In Belgium, two main teams are currently active in the field of apple breeding. The first one is "Beter3fruit", a "Spin off" of the KUL-FTC with private partners (Boomkwekerij Johan Nicolaï, Veiling Haspengouw,…) and public partners like KUL-FTC (Fruitteelt Centrum Rillaar) and the Province. IWT is funding their research work focused on breeding commercial apple cultivars (e.g. "club" cvs like 'Greenstar" and 'Kanzi' have been created by this team). The second team is working at CRA-W Gembloux with as principal objective to broaden the genetic basis of cultivated apple by using as parents original Belgian genetic resources that are crossed with commercial cultivars. The aim is to create commercial cultivars with a high level of polygenic disease resistance and with original and high fruit quality. An additional team is working at ILVO-PLANT (Melle) with the main objective of selecting new ornamental apple cultivars. Genetic resources also have a potential economic value by direct valorisation of some selected old cultivars for e.g. the nursery trade, diversification of commercial apple growing sector and for processing applications (juice, cider, baby food, bakery industry,…).

To enhance the practical use of the collections we firstly need to develop and apply a proper and extensive characterisation and evaluation method. Furthermore, a good data organisation (database) is necessary and finally we need the development of a dynamic network of potential users. The management of such important collections is quite a large work and rationalisation procedures are needed before developing safe conservation strategies. Therefore the definition of safe conservation strategies is the last process of the steps defined above. The identification of the unique and original material held in the collections is necessary basic knowledge in order to avoid, as far as possible, gaps and redundancies.

Since 1980, at the European level, the European Cooperative Programme for Crop Genetic Resources Network (ECP/GR – IPGRI) plays a central role as an umbrella for the coordination of collection, characterisation, evaluation, utilisation and safe conservation of plant genetic resources. Thirty-nine countries belong to ECP/GR and

each of them contributes financially to its subsistence. ECP/GR is organized in 18 different Crop Working Groups (WG's) and one of them is dealing with *Malus* and *Pyrus* genetic resources – Belgium is currently chairing this WG. Inside this network, 45 European Central Crop Databases (ECCDB) are managed - in kind - each by one country. United Kingdom is responsible for the European *Malus* CCDB and Belgium for the *Pyrus* one. A common strategy for long term conservation of fruit tree genetic resources progressively developed at the level of ECP/GR. The principle is laid on the definition of National collections and the virtual European collection philosophy is to compare national collection lists, sorting out gaps and redundancies and sharing the responsibilities between European countries for the safe conservation of the most original material that enlarge the diversity of traits.

The objective of this project was, based on a limited sample of material, to start to work on the definition of regional and national collections and to develop common characterisation and evaluation methods. Future steps could be the sharing of responsibilities for the long term conservation strategies of the "Belgian apple collection" concept.

At this level of study the definition of strategies for a safe conservation of apple genetic resources is at the preliminary stage because the different partners have very different levels of experience with the management of their collections. Another reason is the lack of phenotypic data that need to be collected and validated (at least 4 to 5 years are needed). Furthermore, there are multiple objectives pursued by the curators of collections. For example, the KUL-FTC collection was a "work collection" for breeding purposes. It is therefore a dynamic collection focused on their breeding aims and not *per se* to be conserved. The NBS collection is managed by amateur growers and it is a mix between collected material from old orchards and collection of cultivars from different nurseries and from other collections, even from abroad. Numerous redundancies are probably present in this collection. The CRA-W collection is focused on the collection, the evaluation, the characterisation and the valorisation of our Belgian apple genetic resources and its main interest is to conserve the original material which significantly enlarges the existing diversity. The CRA-W database contains more than 20 years evaluation and characterisation data and quality data management procedures are developed.

5.2.3 TOWARD THE DEFINITION OF A SAFE BELGIAN CONSERVATION STRATEGY FOR GENETIC RESOURCES OF *MALUS DOMESTICA*.

5.2.3.1. Theoretical concept

Due to the large diversity of fruit cultivars and the scarce financial support at the different regional, federal and European levels we need to develop a coordinated

strategy for a safe conservation of our apple genetic resources. One of the proposals that take more and more adhesion inside the ECP/GR Working Groups is the concept of sharing the responsibility between regional, national and European collections by the establishment of decentralized collections of fruit accessions, to ensure long-term conservation and easy access to the important biodiversity for European horticulture, cultural heritage or science.

The principle is to define a "National collection" of apple genetic resources. This "National collection" is composed by the different regional collections. The "National collection" concept is a dynamic process and has quite a broader objective than the concept of "core collection". The objective is to coordinate efforts of individual Belgian collections to conserve and make available *Malus* accessions originating in Belgium or otherwise important to Belgian horticulture, cultural heritage or science. The availability will be subject to applicable international and national legislation and for the purpose of utilization and conservation for research, breeding and training for food and agriculture.

Furthermore, the Belgian *Malus* Collections should be decentralized and cost-effective *ex situ* collections comprising accessions held by participating gene banks and available for distribution. Accessions to be regarded initially as Belgian Collection accessions are those held in various regional collections and accepted as Belgian Collection accessions by an *ad hoc* expert Working Group. Material that belongs to those collections needs to be summarised and therefore the first step is to define the common criteria that will be used for selection of the accessions. One of the important results of this project is to define Belgian criteria as presented in Table 5.5. It is essential to dispose of a minimum set of passport data on the material that belongs to the different collections and it is of utmost importance to have validated characterisation and evaluation data that lead to the procedure of guaranteed identification of the material. This also means that there must be a "reference collection" that develops the proper identification of the reference material to which the other collections can compare their material in order to validate their identification. It is logic to develop safe conservation strategies only for unique, original and non redundant material. After this stage, material can be identified in order to preserve the socio historical heritage and to fulfil the present and future needs of the Belgian scientific and agronomic activities. The next step is the establishment of protocols for the network structure and sharing the responsibility for the save conservation strategy of this priority collection list.

Table 5.5: Belgian criteria that will be used for the definition of regional and national collections of *M. sylvestris* genetic resources

Type of criteria	***Criteria***
1. Minimum data	a) Material need to have minimum passport data b) Material need to have minimum validated data from characterization and evaluation work.
2. Guaranteed identity of the material and screening unique and original material.	a) Material need to be identified by morphologic and pomological data. Molecular markers are very useful as validation method. b) Material from different collections needs to be compared and it is important to verify if it is already present in other European collections. c) Material need to be defined as original and unique by the above procedures.
3. Criteria for inclusion in the National collection.	a) Old cultivars raised in Belgium either as landraces, as chance seedling or selected by a breeder. b) Old cultivars formerly grown and adapted to Belgium and that had a socio economic importance and therefore a historic link with the country. c) Cultivars that extend the diversity of characters already known. d) Cultivars of interest for scientific, breeding or economic purpose. e) Cultivars not present and not conserved in other European collections.

Such procedures have to be followed both at the regional level, at the national level and finally at the European level where the 'European Collection' consists of a coordinated network of "National collections" with the view of sharing the responsibilities and avoiding duplication of efforts. The global management of the network will be coordinated by the ECP/GR, in collaboration with the respective European Central Crop Database managers. A pilot project called "A European Genebank Integration System" (AEGIS) started in 2004 and has the aim to study the feasibility of these principles.

5.2.3.2 Integration of phenotypic and genotypic data as a source of information for the conservation of Belgian *Malus domestica* collections

The first task was to develop a case study on the utilization of both phenotypic and genomic data with the objective to sort out synonyms and gaps inside and between the different collections. One of the most recurrent problems of old fruit cultivars which hampers the efficient management of collections, is the number of synonyms

that they have received during their travel through time and countries. There are many sources of synonyms, one is the translation or transliteration of original names into local languages but there are also orthographic errors due to the transcription of these names in data collection registers. Through the years, nurserymen have also often changed names either for fun or for commercial reasons. A last source is the "appropriation" by people, over time, of introduced foreign cultivars in their own local tradition by giving them a new local name adapted to their language or dialect.

In this project we have used 10 microsatellites as main source of information for the classification of a sample of 513 accessions into "genotype families", where a genotype family consist of individuals that have identical SSR-fingerprints. Subsequently, phenotypic data of reference cultivars were used as validation procedure and hereby three categories of results could be obtained. Reference cultivars are identified by using phenotypic data that are compared either with reference pomological descriptions (breeder named cultivars) or by comparing different provenances of cultivars that are putatively the same ('landraces' that have never been described in any pomological book). Very interesting information was derived from this exercise:

- Firstly, 13.1% of the accessions that had a pomological name in the collections turned out to be mislabelled and could be identified by comparing them with reference cultivars. In some cases they could only be pointed out as mislabelled but without providing the correct name (e.g. cv ID 135 labelled as 'Speeckaert' that does not belong to the 'Speeckaert' family).
- Secondly, material of the collections that was sampled in the countryside from old orchards as standard trees, often has no name because it is not known to the collector/owner. This material receives a 'curator name' that is provisory. It is possible to identify these accessions based on phenotypic and/or molecular data when they are ranked in a specific genotype family that unravels their pomological identity. In this way, the comparison of the current data resulted in the identification of 15.8 % of the accessions.
- Finally, as validation procedure, accessions that had been identified by pomological work or that were labelled with a pomological name needed to be compared with reference cultivars from the same or between different collections. This work could validate the identity of 16.6 % of the studied material.

Table 5.6. shows some cases where the data validation was based on available phenotypic data. Figure 5.2 shows one of the quick validation procedure using fruit pictures. Furthermore, in some cases the "ploidy level" could help to sort out potential conflicts or to confirm the cultivar identity.

Table 5.6: Examples of results for the accession identification obtained by using 10 SSR's and validation of these results by phenotypic data where available.

TYPES of EXAMPLES	Tree_ID	CRA_NBS Accessio Nb	CRA Genotype Nb	Curator or provisory Name	Accession name	Synonyme	SSRngenotype	Country of origin	Belgian Region	Ploidy	Flow cyto	Collector	Data first validation	PHOTO Validation
1. ERRORS														
	230			Clivia	M9 ?		79					KUL	?	
	804			Discovery	M9 ?		79					KUL	?	
	808			Telamon	M9 ?		79					KUL	?	
	368	149		Keuleman	Pomme d'Or?		19	?				NBS	?	
	440	243		Linneous Pippin (HR-1-11-)	Pomme d'Or?		19	?				NBS	?	
	829	100		Present v Engeland (p. d'Or ?)	Pomme d'Or?		19	?				NBS	?	
	41	41		Speeckaert	Speeckaert	Speeckaert Appel	47	B	Fl.			NBS	OK	
	539	29	29	Speeckaert Appel	Speeckaert	Speeckaert Appel	47	B	Fl.			CRA	OK	
	135	135		Speeckaert	Unkown	Unkown						NBS	?	
2. IDENTIFICATION														
	696	600	600	Camousse	Holaart Doux	Camousse	7	NI				CRA	OK	OK
	715	650	600	Camousse	Holaart Doux	Camousse	7	NI				CRA	OK	OK
	532	14	14	Holaart Doux	Holaart Doux	Camousse	7	NI				CRA	OK	OK
	690	589	589	Comtesse	Reinette Baumann	Comtesse = Double Agathe	11	B	Wal			CRA	OK	OK
	647	346	346	Double Agathe = S 7 ?	Reinette Baumann	Comtesse = Double Agathe	11	B	Wal			CRA	OK	OK
	388	169		Frieslandse Rtte	Reinette Baumann	Comtesse = Double Agathe	11	B	Wal			NBS	OK	OK
	89	89		Onbekende appel H-V	Reinette Baumann	Comtesse = Double Agathe	11	B	Wal			NBS	OK	OK
	701	618	618	Reinette Baumann	Reinette Baumann	Comtesse = Double Agathe	11	B	Wal			CRA	OK	OK
3. VALIDATION	566	71	428	Calville des Vergers	Cwastresse Double	Cwastresse Double, Triomphe du Luxembourg, Calville des Vergers	2	B	Wal	3n	3n	CRA	OK	OK
	653	396	428	Triomphe du Luxembourg	Cwastresse Double	Cwastresse Double, Triomphe du Luxembourg, Calville des Vergers	2	B	Wal	3n	3n	CRA	OK	OK
	592	152	152	Grand Alexandre	Grand Alexandre	Kaizer Alexandre	4	Rus				CRA	OK	OK
	728	694	152	Pomme Beaurieux	Grand Alexandre	Kaizer Alexandre	4	Rus				CRA	OK	OK
	721	661	661	Pépin d'Or	Pépin d'Or		8	B	Wal			CRA	OK	OK
	776	1055	661	Pépin d'Or	Pépin d'Or		8	B	Wal			CRA	OK	OK
	609	195	195	Reinette de Geer	Reinette de Geer		16	B	Wal ?			CRA	OK	
	127	127		Reinette de Geer (HR-1-34-10)	Reinette de Geer		16	B	Wal ?			NBS	OK	
	633	1072	1072	Ribston Pippin	Ribston Pippin		27	UK			3	CRA	OK	
	270			Ribston Pippin	Ribston Pippin		27	UK				KUL	OK	

A high variability exists among the collections for all three categories of results (error, identification and validation) as shown in Figure 5.3.

Figure 5.2: Comparison of fruit pictures from different accessions of the same genotype-family from different origin that confirms their membership to the 'Reinette Baumann' family.

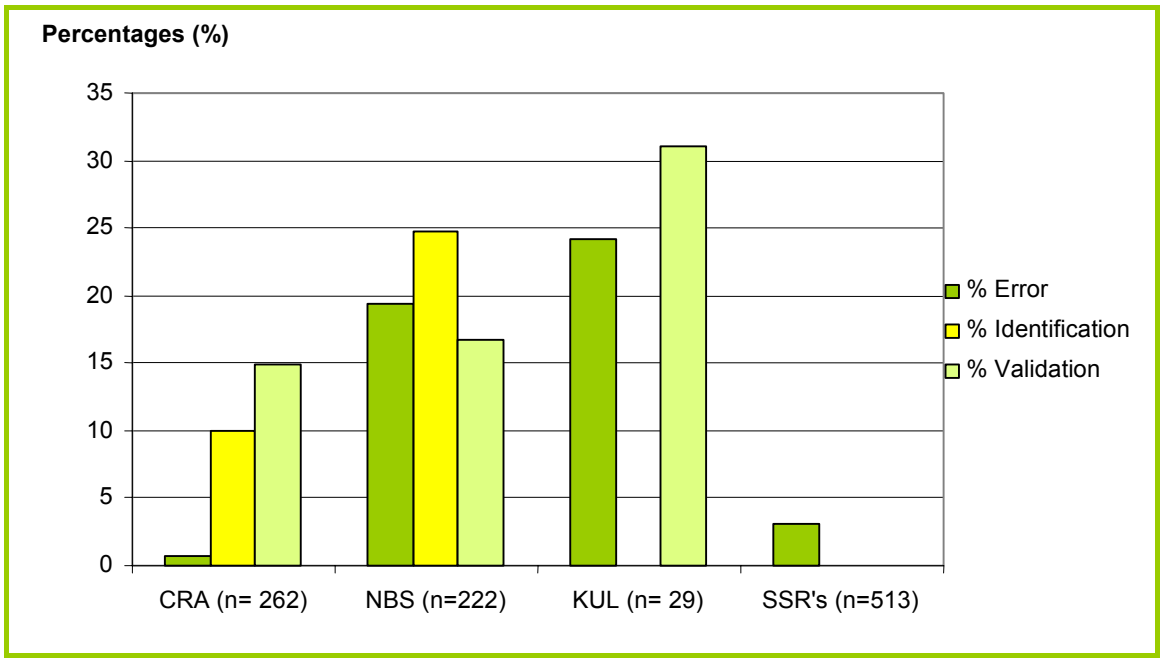


Figure 5.3: Situation of different collections concerning the identification status of a sample of their material

Due to the long experience of the CRA-W in the systematic characterisation and evaluation works on the collected material, percentage of error of CRA-W was very low (less than 1 %). It is also the objective of this Institute to concentrate its efforts on reference cultivars. But even in this situation, the use of molecular data helped to recognise 10 % of unnamed cultivars by comparing them, in 85 % of the cases, with internal reference cultivars. The methodology could also validate the identity of 15 % of accessions that were already grouped in pomological names on the basis of phenotypic data. Using in most of the cases the CRA-W material as reference, the molecular markers results permitted to achieve high score of identification of errors in both other collections and high percentage of identification of unnamed accessions that are stored in the NBS collection. Finally many accessions identities belonging to these two collections have also been validated.

Nevertheless, in a limited set of cases (3.1%), inexplicable discrepancies have been observed between phenotypic and molecular data. In some cases, reference accessions (or their mutants) did not belong to the correct genotype-family, for other accessions belonging to a closely related genetic pool were not grouped in their core family but showed discrepancies in one or a few alleles. Another case concerns cultivars that are put in the same genotype-family and where this is not validated by the phenotypic data. The cluster analysis developed in Chapter 1 is therefore a useful complementary validation tool in such cases. The origin of these discrepancies could be either caused by errors during sampling, mislabelling of DNA samples or scoring mistakes. For this limited number of cases new samples could be analysed (if

necessary with a few of extra markers) in order to validate the data and also to extend the genetic material explored. Examples of these discrepancies are shown in Table 5.7 and Figure 5.4 illustrates the case of 'Joseph Musch'.

Table 5.7: Examples of discrepancies and useful pedigree information

TYPES of EXAMPLES	Tree_ID	CRA_NBS Accessio Nb	CRA Genotype Nb	Curator or provisory Name	Accession name	Synonyme	SSRnrgenotype	Country of origin	Belgian Region	Ploidy	Flow cyto	Collector	Data first validation	PHOTO Validation
1. Joseph Musch	645	342	342	Double Copette = Musch ?	Joseph Musch	Double Copette	12	B	Wal		3	CRA	OK	OK
	819	20		J. Musch	Joseph Musch	Double Copette	12	B	Wal			NBS	OK	
	47	47		Joseph Musch	Joseph Musch	Double Copette	12	B	Wal			NBS	OK	
	242			Joseph Mush	Joseph Musch	Double Copette	12	B	Wal			KUL	OK	
	733	712	712	Pomme Sokay	Joseph Musch Rouge			B	Wal	3n	3	CRA	Problem = 12	
	528	7	7	Joseph Musch	Joseph Musch			B	Wal	3n	3	CRA	Problem = 12	
2. Belle-Fleur Large Mouche	198			Hunt 7	Belle-Fleur Large Mouche Rouge		45	B				NBS	OK	
	657	416	416	Verdia Rouge	Belle-Fleur Large Mouche Rouge Rouge		45	B		3n	3	CRA	OK	OK
	851	216		B.F. type Machiels (HR-21-einde)	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?			NBS	= 45?	
	576	102	101	Belle-Fleur Large Mouche	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?	3n	3	CRA	= 45?	
	402	196		Hunt 5	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?			NBS	= 45?	
	107	107		Large Mouche	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?			NBS	= 45?	
	1147	189		Pommische (Darcis)	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?			NBS	= 45?	
	84	84		Rambeau	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?			NBS	= 45?	
	122	122		Sletsing 1	Belle-Fleur Large Mouche	Pommische, Sletsing, Dubbele Bellefleur	46	B	?			NBS	= 45?	
3. Melkappel	562	66	66	Belle de Furnes	Belle de Furnes		82	B	Fl.			CRA	><	
	603	189	189	Melkappel	Melkappel		82	B	Fl.			CRA	OK	
	85	85		Melkappel	Melkappel		82	B	Fl.			NBS	OK	
	379	160		Wasappel (HR-1-3-26)	Melkappel		82	B	Fl.			NBS	OK	
4. Boskoop and Reinette Coulon	202			Boskoop	Schone Van Boskoop	Boskoop	74	NL				KUL		
	367	148		Boutersem nr 1	Schone Van Boskoop	Boskoop	74	NL				NBS		
	77	77		Boutersem nr 6	Schone Van Boskoop	Boskoop	74	NL				NBS		
	366	147		Boutersem nr 6	Schone Van Boskoop	Boskoop	74	NL				NBS		
	607	193	193	Reinette Coulon	Reinette Coulon	Reinette de Kerkvoorde	75	B	Wal	3n	3	CRA		
	421	222		Reinette Coulon (onder)	Reinette Coulon	Reinette de Kerkvoorde	75	B	Wal			NBS		
	662	435	435	Reinette de Kerkvoorde	Reinette Coulon	Reinette de Kerkvoorde	75	B	Wal		3	CRA		

Figure 5.4: Pictures of the reference tree of 'Joseph Musch' (B 16) and one of his red mutant 'R 39' that does not belong to the 'Joseph Musch' SSR family.

The quick validation based on fruit pictures is corroborated with the cluster analysis presented in Chapter 1 where 'B16' and 'R39' clearly belong to the same cluster 8 and on the opposite, the cv 'Melkappel' in classified in another cluster (cluster 9) than the cv 'Belle de Furnes' (cluster 2). In the same way, 'Verdia Rouge' and 'Belle-Fleur

Large Mouche belong to the same cluster. The classification obtained by the clustering analysis of phenotypic data can be used as a validation indication of relationship between cultivars but it will never provide certainty on the identity of a cultivar. On the other hand, when cultivars belong to distinct clusters there is a very low probability that they are synonyms.

It is interesting to point out that for instance, 'Reinette Coulon' that is described in the literature as a mutant or a sub-type of the cv 'Schone van Boskoop' was confirmed by both the molecular marker data (same genotypic family) and the cluster analysis.

All the work devoted to the identification of this sample of *M. domestica* collections, constitutes only the first step of the definition of a national collection. From the present results it becomes clear that many accessions have different names, although they share the same molecular fingerprint. Some of them are sports from each other or mutants from the same initial genotype (e.g. the group of 'Jonagold' or 'Elstar' mutants or the different sports of the 'Belle Fleur' group). Sports and mutants can not be distinguished by the SSR analysis but they do have different phenotypes. We can assume that among the 513 accessions studied, 324 accessions constitute different genotypes. This means that 37 % are duplicates, synonyms or mutants of well known cultivars. Among this sample of cultivars, 130 are of Belgian origin (40 %). The next step towards the definition of a 'National collection' is the setting up of a concrete network of partners that share the responsibility of a safe conservation of Belgian apple genetic resources but this is still open to further work.

5.3 DISSEMINATION OF RESULTS

5.3.1 COMMUNICATION OF RESULTS TO THE GENERAL PUBLIC

On November 23rd 2004, a workshop was organised in Brussels. The aim of the project was to disseminate the current results to the members of the users' committee and other stakeholders and also obtain feed- back from this group to proceed with the project. The invitation to the workshop, including the program, is shown in Figure 5.5. In total, 57 persons attended the workshop. Many participants were enthusiastic about the project and were interested in attending a similar workshop where final results will be presented.

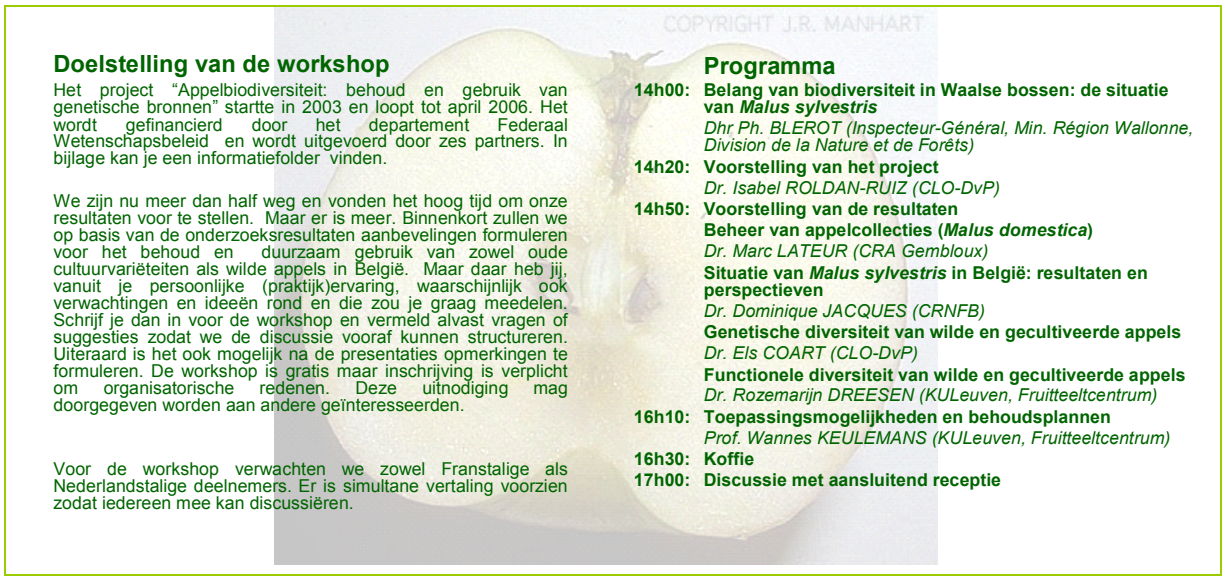


Figure 5.5: Invitation to the workshop on apple biodiversity.

5.3.2 WEBSITE

A web site was designed (see Figure 5.6) and is available at http://applebiodiv.cra.wallonie.be. General information on the project and queries on the data are available to the public. Row data are only accessible for the partners of the project. This website is hold and managed by the CRA-W partner, with a MySQL database in the Linux environment. The webpage are developed with the Dreamweaver software in PHP language.

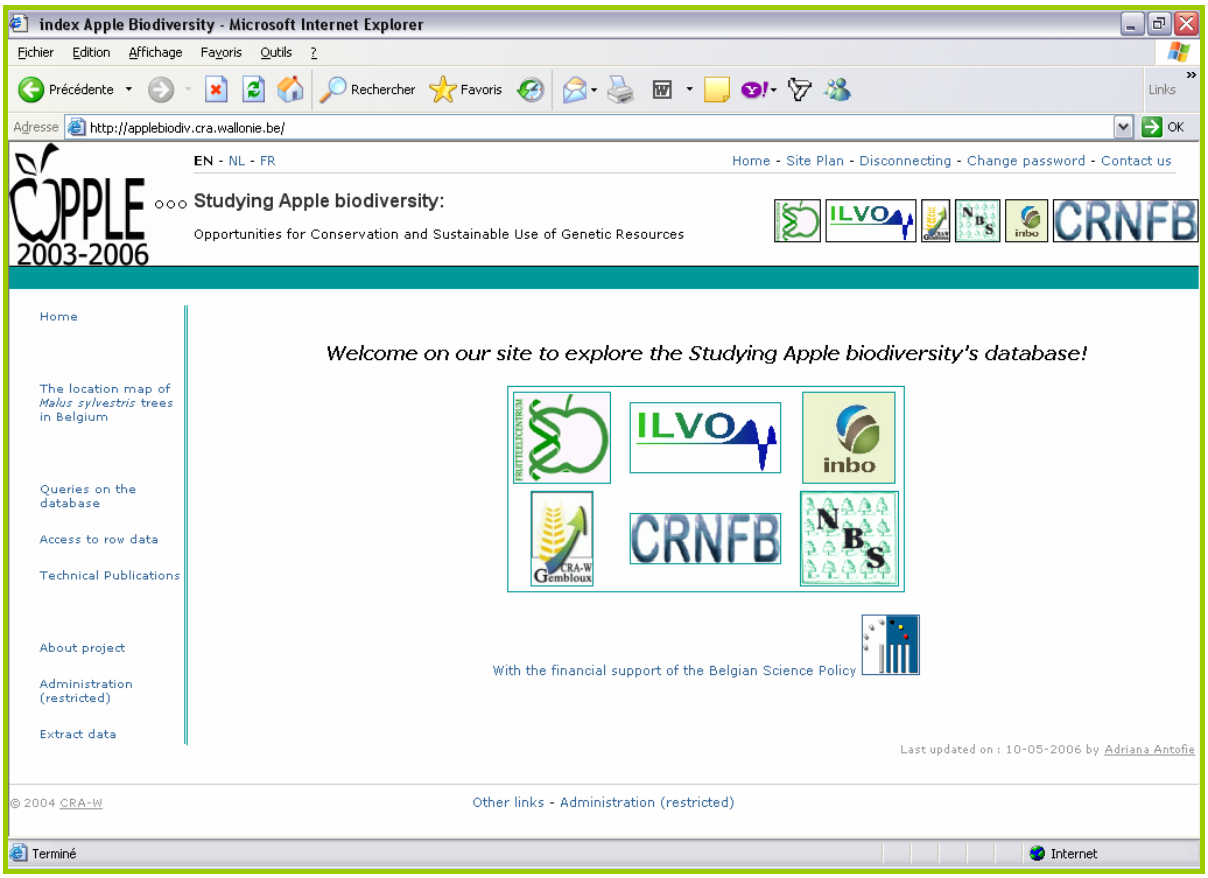


Figure 5.6: Homepage of the apple website.

5.4 CONCLUSIONS

As the first European country, Belgium has a complete inventory of the occurrence of the rare forest tree species *M. sylvestris*. The species is very rare in Flanders but rather common in Wallonia. Based on thorough phenotypic and genotypic analyses, we were able to discriminate between pure *M. sylvestris* genotypes, hybrids with *M. domestica* or feral cultivars. Throughout Belgium, hybrids were identified with a total of 7% of apple genotypes sampled in nature. Furthermore, large phenotypic and genotypic diversity was still present within populations.

A core collection consisting of 200 pure *M. sylvestris* genotypes was constructed, using the results of this project. This collection will be completely planted in Phillipeville for conservation and seed production purposes. In Dentergem, all Flemish wild genotypes are planted together with the same aims in mind. Also in-situ conservation measures are currently taken, including restocking of the small Flemish populations. Responsibilities for in-situ conservation are shared among scientists and field managers by providing the field managers with detailed distribution data of the species and suggestions for the forest management.

For *M. domestica*, one of the important results is to have defined the main lines of a methodology for the definition of a Belgian strategy that has been based on this case study. When sufficient reference material is available, the use of molecular markers for the identification of duplicates within and between collections is very promising but the complementary phenotypic evaluation is of utmost importance. This study further highlights the importance of a clear definition of each step in the development of the 'National collection' concept, and more specific for validation procedures. This is of particular significance for the characterisation and the evaluation of data and to a lower extent, to the molecular data.

GENERAL CONCLUSIONS

As demonstrated below, most objectives defined by the start of the project have been achieved. We can state that the results obtained constitute an important contribution for future conservation efforts of *M. sylvestris* and *M. domestica,* not only in Belgium, but also in other European regions. Through the combination of different types of information we have been able to get new insights on the diversity present in the Belgian *Malus* gene pool. This project has considered fundamental questions such as the origin of the cultivated apple, methodological questions on the combination of different types and data from different collections but has also gone far beyond the 'pure' academic aspects and has resulted in the implementation of concrete conservation efforts for the species studied. The combination of the complementary expertises of the different partners has been a clear added-value.

Furthermore, the methodology developed and the experience gained constitutes a valuable basis for the design of conservation approaches in similar species. Here all results are summarised and discussed in the context of the objectives formulated by the start of the project.

1. Study the genetic diversity present in wild apple populations and old regional varieties at different levels of organisation: neutral and functional nuclear diversity and cytoplasmic diversity

For *M. sylvestris*, high phenotypic diversity has been detected, especially for tree shape, forking, pubescence of leaves and twigs, flushing, fruit form but also for disease resistance traits. Furthermore, *M. sylvestris* populations still harbour high genetic diversity, and no clear signs of genetic erosion were detected in this scattered forest species in Belgium. Estimates of observed and expected heterozygosity were comparable, resulting in low levels of inbreeding or heterozygote excesses as expected for an obligate cross-pollinator. No geographical structuring of genetic diversity was detected at the regional level. Only the Danish populations group together and can be considered a separate gene pool. When pooling the populations into five regions, it becomes apparent that Flanders and Wallonia are the most related areas. Within populations, genotypic and geographical distances are positively correlated for individuals up to 10 m apart. Beyond 10 m this relationship drops dramatically to become non-significant at a distance of 100 m.

The collections of old *Malus* varieties studied harbour a significant amount of diversity. Phenotypic results highlight the major relevance of a validation procedure

that needs to be adopted prior to data collection and analysis. A striking result was the similarity between the collections of CRA-W and NBS for genetic and phenotypic diversity. The collection of modern cultivars was shown to contain a similar amount of genetic diversity as the collections of old varieties, indicating that the threat of inbreeding in the modern apple gene pool is probably not as relevant as often suggested.

2. Analyse past and recent hybridisation processes between wild and cultivated apple gene pools and determine the present degree of distinctness between wild apples and cultivated varieties

This study has confirmed that *M. sylvestris* and *M. domestica* represent clearly differentiated gene pools that can be separated on the basis of allele frequencies at 10 SSR loci. However, 7% of the sampled *M. sylvestris* trees was identified as hybrid with *M. domestica*, indicating ongoing gene-flow between both species. Strikingly, hybrids were almost entirely restricted to Belgium. Next to environmental factors that might enhance contact between the species (e.g. smaller forest sizes, fragmentation of forests), this might also be the reflection of the apple cultivars analysed in the present study. Many ancient apple cultivars of Belgian origin or often planted in Belgium were part of the project, whereas only a small selection of other European cultivars were included. Eight % of the accessions, sampled as *M. sylvestris* in nature, were identified as feral cultivars.

Clear differences exist between the species at the self-incompatibility locus although they shared 24 S-alleles. Most novel S-alleles identified were found in *M. sylvestris* genotypes.

On the other hand, the study at chloroplast level indicates the occurrence of hybridisation between *M. sylvestris* and *M. domestica* by the extensive sharing of chloroplast haplotypes. The nature of this relationship is difficult to establish but probably the detected gene flow is bi-directional and brought about by the use of (local) wild *Malus* genotypes for the (local) cultivation process of apple and the later cytoplasmic introgression of chloroplast haplotypes into *M. sylvestris* populations from the domesticated apple.

From a methodological point of view, an innovative strategy based on the use of linked SSR markers was developed and tested to detect hybridisation. The study of linked SSR loci identified more accessions that might be the result of hybridisation with *M. domestica* than the analyses of unlinked SSRs. However, in most cases, these newly detected hybrids show signs of historic hybridisation, even dated back in the medieval period. This technique thus provides further insight into the evolution of

both species rather than identifying genotypes that have to be considered as hybrids with *M. domestica*. Therefore, these historic introgression events were not be taken into account when composing a core collection of pure *M. sylvestris* genotypes for future conservation (see objective 4).

3. Asses the viability of wild apple populations through the analysis of demographic parameters and fertility-related traits

According to the data collected, fructification of *M. sylvestris* in Belgian forests is very low, and most trees do not carry fruits. This can hamper *in situ* conservation efforts. We have taken initiatives to promote fruit production of natural populations (see objective 4).

For the different local populations of *M. sylvestris*, a rather large diversity of S-alleles in comparison to the amount of individuals was observed, which is positive for the generation of offspring. However, for example in "Meerdaal forest" a significant part of individuals shares one S-allele. This means that half of the male gametes will not be able to fertilize flowers and give raise to fruit. Under conditions of high pollen availability, this will give comparable fruit set compared to full compatibility. However, when the pollen availability is limited (e.g. bad weather conditions, no (semi) compatible trees in the near vicinity) this could cause problems towards the viability of the population.

4. Implement all the above results into conservation guidelines for wild apple in Belgium (in-situ where possible, ex-situ where necessary). We will assess the possibilities for in-situ conservation for wild apple. The collected grafts of wild apple trees can be used to construct permanent gene banks

As the first European country, Belgium has now a complete inventory of the occurrence of the rare forest tree species *M. sylvestris*. Especially in Flanders, local populations are severely reduced in size. Total extinction is a serious threat for these populations, except for one large population in Meerdaal forest. The only way to preserve the small populations is restocking with plant material originating from the proper provenance region. In addition, it is absolutely necessary to cut hybrids and cultivated apple trees in the immediate surroundings. In Meerdaal forest, owned by the Flemish community, the rangers are aware of the importance of the exceptional large population of *M. sylvestris*. Fructification of the trees will be promoted through cutting of surrounding trees and improving light conditions.

In the Walloon region *M. sylvestris* is more densely distributed and populations consist of larger numbers of individuals in comparison to the situation in Flanders. No

immediate menace of extinction is at hand. The *in situ* conservation strategy in Wallonia includes two main objectives: the follow up and the conservation of the *M. sylvestris* trees and the selection of three stands with a large population compatible with the seed collection for forestry purposes. To reach the first objective, the forest rangers are invited to promote the different *M. sylvestris* trees identified in their district by adapted thinning to favour their growth and their fructification. Furthermore, three specific stands have been identified and selected. Forest rangers are encouraged to manage these three stands by thinning of the trees in competition with the *M. sylvestris* to promote their development and seed production. The final purpose is to give the opportunity to the Walloon Seed Centre and private nurseries to collect these seeds, so that the material will become commercially available. To give the possibility for official control of the forest reproductive material produced in these stands, the possibility of including *M. sylvestris* in the list of species submitted to control will be evaluated.

For the construction of a core collection of *M. sylvestris*, it is of utmost importance to discriminate between wild, hybrid and cultivated apple trees. We therefore combined phenotypic (based on leaf hairiness and fruit size) and SSR classification where possible and otherwise used only SSR classification. A core collection consisting of 201 pure *M. sylvestris* genotypes has been constructed. This collection will be completely planted in Phillipeville for conservation and seed production purposes. In Dentergem, all Flemish wild genotypes are planted together for the same purposes.

5. *Develop an efficient management plan to conserve the biodiversity present in existing collections of old regional varieties.*

When sufficient reference material is available the use of molecular markers for the identification of duplicates within and between collections is very useful but the complementary phenotypic evaluation is still of utmost relevance. Our study further highlighted the importance of a clear definition of each step in the development of the 'National collection' concept, and more specific for validation procedures. This is of particular significance for the characterization and the evaluation data and to a lower extent, to the molecular data.

The validation of genotypic analyses with phenotypic data on reference cultivars, led to three categories of results. In total, 13% of the accessions turned out to be mislabelled, whereas 16% of formerly unidentified *Malus* accessions were identified. Furthermore, the identity of 17% of accessions as formerly defined based on pomological basis was confirmed. We can assume that among the 513 accessions studied, 324 constitute different genotypes. This means that 37 % are duplicates,

synonyms or mutants of well-known cultivars. Among this sample of cultivars, 130 are of Belgian origin (40 %).

For *M. domestica*, one of the first important results of this project is the definition of the main lines of a methodology for the definition of a Belgian strategy. The next step towards the definition of a 'National collection' is the setting-up of a concrete network of partners that share the responsibility of a safe conservation of Belgian apple genetic resources but this is still open to further work.

6. Determine the value of wild apple trees and regional varieties to expand the genetic base of current apple breeding programmes (i.e. introgression of new sources of stress and disease resistance)

Wild apple trees do not appear very susceptible to diseases. Cancer and powdery mildew symptoms are very limited but leaf scab infections are slightly more important. Fruit scab, the most important fungal disease on cultivated apples, is also observed on wild genotypes. The infection level on the field is medium but the variability is very large. More than 70% of wild trees was ranked in class 1 (no symptoms) for cancer susceptibility. Furthermore, leaf scab and powdery mildew also appear to be heritable characteristics. These results show that a large genetic diversity still exist in this wild species and confirms the possibility of selection for horticultural purposes especially for cancer resistance. In total 15 *M. sylvestris* genotypes are entering in the CRA-W collection to be further evaluated for their use in breeding programs.

Also cultivated accessions belonging to the three collections under study, show a high variability for disease resistance and fruit characteristics. Although no immunity was detected among these cultivated trees for scab, powdery mildew or cancer, it is clear that they contain valuable genetic information for future breeding programs.

7. Disseminate these conservation strategies to a wide audience, including forestry and nature conservation agencies, and non-governmental organisations involved in conservation of wild species

At the onset of the project, a website with general information on the project was developed. Also a handout with our major goals and approaches was created in three languages to disseminate among all contacts held by the different partners.

On November 23rd 2004, a workshop was organised in Brussels to disseminate the current results to the members of the users committee and other stakeholders and also obtain feedback from this group to proceed with the project. Many participants were enthusiastic about our work and were interested in attending a similar workshop where final results will be presented.

REFERENCES

Amane M, Lumaret R, Hany N, Debain C, Vivier G, Deguilloux MF (1999) Chloroplast-DNA variation in cultivated and wild olive (*Olea europeae* L.). *Theoretical and Applied Genetics*, 99, 133-139.

Beaumont, M. A., Barratt, E. M., Gotelli, D., Kitchener, A. C., Daniels, M. J., Pritchard, J. K., & Bruford, M. W. (2001) Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology,* 10, 319-336.

Broothaerts W. (2003) New findings in apple S-genotype analysis resolve previous confusion and request the re-numbering of some S-alleles. *Theoretical and Applied Genetics* 106:703-714

Broothaerts W,, Van Nerum I, and Keulemans J. (2004) Update on and review of the incompatibility (S-) genotypes of appel cultivars. *Hortscience* 39(5): 943-947

Coart E, Van Glabeke S, De Loose M, Larsen AS and Roldán-Ruiz I (2006) Chloroplast diversity in the genus *Malus*: New Insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*M. domestica* Borkh.). *Molecular Ecology*, 15, 2171-2182.

Coart E, Vekemans X, Smulders MJM, Wagner I, Van Huylenbroeck J, Van Bockstaele E, Rolddán-Ruiz I (2003) Genetic variation in the endangered wild apple (*Malus sylvestris* (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. *Molecular Ecology*, 12, 845-857.

De Langhe J.-E., Delvosalle J., Lambinon J., Vanden berghen C. (1978) Nouvelle flore de Belgique, du Grand-Duché de Luxembourg, du Nord de la France et des Régions voisines (Ptéridophytes et Spermatophytes) Deuxième édition: 899 p. Edition du Patrimoine du Jardin botanique national de Belgique, B-1860 Meise.

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, 4, 129-131.

Dumolin S, Demesure B, Petit J (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics*, 91, 1253-1256.

Dumolin-Lapègue S, Pemonge MH, Petit RJ (1997) An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology*, 6, 393-397.

Falush, D., Stephens, M., & Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics,* 164, 1567-1587.

Fellenberg U. (2001) Beurteilung von Wildobst – Voraussetzung für geeignetes Vermehrungsgut zur Erhaltung von Waldgenressourcen. *Forst und Holz* 56(2): 50-54.

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. *Theoretical and Applied Genetics*, 96, 1069-1076.

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatelittes in *Malus x domestica*: abundance, polymorphism and cultivar identification. *Theoretical and Applied Genetics*, 94, 249-254.

Hardy OJ & Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial structure at the individual or population levels. *Molecular Ecology Notes*, 2, 618-620.

Hokanson SC, Swezc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* borkh. core subset collection. *Theoretical and Applied Genetics*, 97, 671-683.

Ishimizu, T., Inoue K., Shimonaka M., Saito T., Terai O., and Norioka S. (1999) PCR-based method for identifying the S-genotypes of Japanese pear cultivars. *Theoretical and Applied Genetics*, 98, 961

Juniper, B. E., Watkins, R., Harris, S. A., Tobutt, K. R., & Alston, F. H. (1999) The origin of the apple. I: Tobutt, K.R. & Alston, F.H.(eds.). Proceedings of the Eucarpia Symposium on Fruit Breeding and Genetics. 16.Sep.1999. Acta Horticulturae. 484, 27-33.

Kirby GC (1975) Heterozygote frequencies in small subpopulations. *Theoretical Population Biology*, 8, 31-48.

Kleinschmit J, Stephan R, Wagner I (1998) Wild fruit trees (Prunus avium, Malus sylvestris and Pyrus pyraster). In: Noble hardwoods network; IPGRI/EUFORGEN Report of the second meeting, 22-25 March 1997, Lourizan, Spain. Compilers Turok J, Collin E, Demesure B, Erikson G, Kleinschmit J, Rusanen M, Stephan R. p51-60.

Koopman W*, Li Y*, Coart E*, van de Weg W, Vosman B, Smulders M.J.M., Roldán-Ruiz I. (2006) Reconstruction of gene flow among wild apple (*Malus sylvestris* (L.) Mill.) populations and hybridization with cultivated apple (*M. domestica* Borkh.) using multilocus microsatellite haplotype sharing. *: the three authors contributed equally. *Molecular Ecology*, in press.
Lecomte H., Florkin P., Morimont JP, Thirion M. (2002) La forêt wallonne, état de la resoource à la fin du 20ème siècle: 69 p. MRW, DGRNE, Division de la Nature et des Forêts.
Liebhard R, Koller B, Gianfranceshi L, Gessler C (2003) Creating a saturated reference map for the apple (*Malus x domestica* Borkh.) genome. *Theoretical and Applied Genetics*, 106, 1497-1508.
Lofty S, Luro F, Carreel F, froelicher Y, Rist D, ollitrault P (2003) Application of cleaved amplified polymorphic sequence method for analysis of cytoplasmic genomes among Aurantioideae intergeneric somatic hybrids. *Journal of the American Society of Horticultural Science*, 128, 225-230.
Maes N en Rovekamp C (1999) Onderzoek naar autochtone genenbronnen in de ecologische impulsgebieden, vertrouwelijke rapport, AMINAL afdeling Bos en Groen, Brussel.
Maes N en Rovekamp C (2000) Onderzoek naar autochtone genenbronnen in de regionale landschappen Vlaamse Ardennen en Westvlaamse heuvels en houtvesterijen Hechtel en Bree, vertrouwelijke rapporten, AMINAL afdeling Bos en Groen, Brussel.
Maliepaard C, Alston FH, Van Arkel G, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill) using multi-allelic markers. *Theoretical and Applied Genetics*, 97, 60-73.
Matsumoto S. and Kitahara K. (2000) Discovery of a new self-incompatibility allele in apple. *Hortscience* 35(7):1329-1332
Matsumoto S., Suzuki M., Kitahara K. and Soejima J. (2000) Possible Involvement of a New S-gene 'St-RNase' (Accession No. AB035928) in the Wild Apple Possessing High Similarity to the 'S3-' and 'S5-RNase' in the Japanese Pear. (PGR00-025*). Plant Physiology*. 122: 620
Matsumoto S., Kitahara K., Komatsu H. and Soejima J. (2001) A functional S-allele, 'Sg' in the wild apple possessing a single amino acid, S-RNase 'Sg'-RNase', different from 'Sg-RNase' in *Malus* x *domestica* cultivars. *J. Hortic. Sci. Biotechnol*. 76: 163-166
Mohanty A, Martin JP, Aguinagalde I (2001) A population genetic analyses of chloroplast DNA in wild populations of *Prunus avium* L. in Europe. *Heredity*, 87, 421-427.
Nanson A.(1970) L'héritabilité et le gain d'origine génétique dans quelques types d'expériences. *Silvae genetica.* 19(4**)**: 113-121.
Nanson A. (1971) Tests de descendances d'épicéa. Trav. Sta. Rech. Eaux et Forêts Groenendaal, Sér. E, n° 4, 64 p.
Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89, 583-590.
Noiton DAM, Alspach PA (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. *J. Am. Soc. Hortic. Sci*. 121, 773-782.
Panda S., Martin J.P., Aguinagalde I., Mohanty A. (2003) Chloroplast DNA variation in cultivated and wild *Prunus avium* L.: a comparative study. *Plant Breeding*, 122, 92-94.
Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, 12, 844-855.
Populer C., Lateur M., Wagemans, C. (1998) Ressources génétiques et résistance aux maladies des arbres fruitiers *Biotechnol. Agron. Soc. Environ* **2(1)**, 46-58.
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics,* 155, 945-959.
Rameau J.C., Mansion, D. & Dume G. (1989) Flore Forestière Française. Guide écologique illustré. 1 Plaines et collines: 1785 p. IDF et Ministère de l'agriculture et de la forêt, France.
Rieseberg LH, J Whitton, K Gardner (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152, 713-727
Rieseberg LH, Baird SJE, Gardner KA (2000) Hybridization, introgression and linkage evolution. *Plant Molecular Biology*., 42, 205-224.
Robinson JP, Harris SA, Juniper BE (2001) Taxonomy of the genus *Malus* Mill. (*Rosaceae*) with emphasis on the cultivated apple, *Malus x domestica* Borkh. *Plant Systematics and. Evolution*, 226, 35-58.
Schlosser S, Reichhoff L, Hanelt P (1991) Wildpflanzen Mitteleuropas, Nutzung und Schutz. DLV Berlin.
Schneider D., Stern R.A. and Goldway M. (2005) A comparison between semi and fully compatible apple pollinators grown under sub-optimal pollination conditions. *Hortscience* 40: 1280-1282
Silverberg-Dilworth E, Matasci CL, Van de Weg WE, *et al*. (2006) Microsatellite markers spanning the apple (*Malus* x domestica Borkh.) genome. *Tree Genetics & Genomes*. In press

Tignon M, Watillon B and. Kettmann R. (2001) Identification of copia-like retrotransposable element by apple. *Acta Hort*. 546 : 515-520.
Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang C, Zheng H, Zhao K, Calabrese P, Dean C, Nordborg M (2006) A nonparametric test reveals selection for rapid flowering in the Arabidopsis genome. *PLoS Biology* **4**, e137.
Vaughan S.P., Russell K., Sargent D.J. and Tobutt K.R. (2006) Isolation of S-locus F-box alleles in Prunus avium and their application in a novel method to determine self-incompatibility genotype. *Theor Appl Genet.* 112: 856-866
Van Nerum I, Geerts M, Van Haute A, Keulemans J, Broothaerts W (2001) Re-examination of the self-incompatibility genotype of apple cultivars containing putative 'new' S-alleles. *Theoretical and Applied Genetics* 104: 584-591.
Vekemans X & Lefèbvre C (1997) On the evolution of the heavy-metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. *Journal of Evolutionary Biology*, 10, 175-191.
Vernier E. & Teissier Du Cros E. (1996) Variabilité génétique du hêtre. Importance pour le reboisement en Picardie et en Normandie. *Rev. For. Fr.* XLVIII (1): 7-20.
Vorman B and Gebhard K (2000) PCR-based markers reveal genetic identity and diversity in subset collections of wild and cultivated apple. *Acta Hort*, 530, 463-467.
Yeh FC, Wang R-C (1999) POPGENE version 3.31. Department of Renewable Resources, University of Alberta, Canada.
Wagner I. (1995a) Identifikation von Wildapfel (Malus sylvestris (L.) Mill.) und Wildbirne (Pyrus pyraster (L.) Burgsd.): Voraussetzung zur Generhaltung des einheimischen Wildobstes. *Forstarchiv* 66(2): 39-47.
Wagner I. (1995b) Artenschutz bei Wildapfel. Die Blattbeharung von 116 Apfelklonen auf zwei Samenplantagen. *Forst und Holz* 53(1): 40-43.
Wagner I. (1996) Zusammenstellung morphologischer Merkmale und ihrer Ausprägungen zur Unterscheidg vonun Wild- und Kultur-formen des Apfel- (Malus) und Birnbaumes (Pyrus), *Mitteilungen der Deutschen Dendrologischen Gesellschaft* **82**:87-108
Waugh, R., McLean, K. and Flavell, A.J. (1997) Genetic distribution of Bare1-like retrotransposable element in the barley genome revealed by sequence-specific amplification polymorphism (S-SAP). *Mol. Gen. Genet*. 253:687-694.
Wuehlisch G. von, Jacques D.and Muhs H.-J. (1993) Phenological differences between beech provenances. In : Muhs, H.-J. and Wuehlisch, G. von (Eds): The scientific basis for the evaluation of forest genetic resources of beech. EC-workshop proceedings, Ahrensburg, working document for the EC, DG VI, Brussels, 229-232.

ACKNOWLEDGEMENTS

We thank Nancy Mergan, Sabine van Glabeke (ILVO-PLANT) and Katrien Luyten (LFBB) for excellent laboratory assistance and Veerle Van Den Eynde (INBO) for collection of *M. sylvestris* material. Also thanks to Sébastien Lemaire (CRNFB and CRA-W) for selection and description of the trees in the forest and the coordination of the nursery activities. The collaboration with René Smulders, Wim Koopman (Plant Research International, Wageningen, The Netherlands) and Anders Soendergaard Larsen (Danish Centre for Forest, Landscape and Planning, Hørsholm, Denmark) is greatly appreciated.

We thank the Federal Science Policy office for funding this project.