



Core collections building for ex-situ conservation

M21

Deliverable 2.1

Author(s): H  l  ne Muranty (INRAE)



**Funded by
the European Union**

Document Control Sheet

Project and deliverable/milestone Information

Project acronym	FRUITDIV
Project title	Exploiting the Untapped potential of Fruit tree Wild DIversity for Sustainable Agriculture
Project coordinator	INRAE
Project duration	60 months
Project start	1 January 2024
Funding instrument	Horizon Europe, Research and Innovation Action
Topic	HORIZON-CL6-2023-BIODIV-01
Deliverable no	D2.1
Deliverable title	Core collections building for ex-situ conservation
Deliverable type	Report ☒ Data ☐ Demonstration ☐ Other ☐
Dissemination level	Public (PU) X
Work package no and title	WP2: CWR genotypic characterisation
Task no and title	Task 2.3 Building CWR core-collections that will capture a maximum of genetic diversity at the species level
Task leader (name and short org. name)	Hélène Muranty (INRAE)
Lead beneficiary (name and short org. name)	Hélène Muranty INRAE-IRHS (INRAE)
Contributing beneficiary(ies) (name and short org. name)	See table 3
Status	Draft X Final ☐
Due date of deliverable	30 September 2025
Actual submission date	13 th of October 2025

Document history

Version	Date	Author/organisation
1	25 September 2025	Hélène Muranty (INRAE)
2	7 October 2025	Hélène Muranty (INRAE)
3	10 October 2025	Hélène Muranty (INRAE)

Document review

Reviewer	Date	Organisation	Reviewer/Organisation
1	XX Mon 202x	(short organisation name)	First and Last Name

Legal disclaimer

The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided “as is”. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The FRUITDIV Consortium Partners shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

Table of content

Summary	5
1. Background	6
2. Material and methods	6
2.1 Material.....	6
2.2 Methods.....	7
2.2.1 Genotyping	7
2.2.2 Structure analysis for species identification.....	8
2.2.3 Core-collection building.....	8
2.3 Contributing beneficiaries	9
3. Results.....	10
3.1 Amygdalus.....	10
3.2 Pyrus	10
3.3 Cerasus.....	11
3.4 Prunophora	12
3.5 Malus	14
Conclusions	15
Appendix 1: Names and sources of SSR markers.....	17

Summary

In order to enhance the genetic diversity and resilience of fruit tree crops, the FRUITDIV project studies and promotes the use of wild relatives. In this context, core-collections for five CWR taxa, *Malus*, *Pyrus*, *Prunus* section *Cerasus*, *Prunus* section *Prunophora* and *Prunus* section *Amygdalus*, will be designed and planted *ex-situ* in multiple sites, EU-wide, to conserve the widest genetic diversity and to establish a long-term experimental design allowing to disentangle genetic effects and GxE interactions in CWR. For all species, leaf samples from accessions in *ex-situ* collections and from *in-situ* CWR populations were received and analysed with 19 to 24 SSR markers or with an SNP array (for *Malus*) in a preliminary genotyping step aiming at setting aside crop/domestic individuals or crop x wild hybrids with a significant crop genetic component. After assigning samples to species, and possibly groups within species, including the crop compartment, individuals showing a low amount of admixture of the crop species were retained for core-collection building. Starting from sets of 206 *Prunus webbii*, 276 *Pyrus pyraeaster*, 289 *Prunus avium*, 643 individuals from the *Prunophora* taxon, representing a mix of *P. cerasifera*, *P. divaricata*, *P. cocomilia*, and *P. ramburii*, core-collections of 61, 86, 36 and 86 individuals, respectively, were built, that maximized diversity while minimizing redundancy. The 265 *Malus sylvestris* individuals were ordered according to their interest for building a core-collection that maximizes diversity while minimizing redundancy. As additional samples were collected in 2025 and their preliminary genotypic characterization data were not all available, the core-collections will likely evolve to consider these additional samples. Additionally, extending the core-collections by adding some redundancy in them, in order to repeat the genes in various backgrounds, will probably be necessary in order to obtain sufficient power for GWAS, which is a second goal of these core-collections.

Note: References should be placed in a footnote by using the British English style.

1. Background

The FRUITDIV project aims to enhance the genetic diversity and resilience of fruit tree crops through the study and use of wild relatives and diverse genetic resources. Work package 2 focuses on the genotypic characterization of crop wild relative (CWR) populations and collections in order to: 1) define core-collections for each genus and species to conserve the widest genetic diversity by planting *ex-situ* collections multi-sites, EU-wide, in the frame of WP3, 2) specify species status of in situ populations and identify key population(s) to preserve *in-situ*, together with WP6, 3) produce a robust genotypic dataset along with methodology and metadata for further genetics analyses in WP4 and WP5, and 4) provide a comparative view of species and population diversities and structures of fruit tree CWR. The present deliverable corresponds to the first objective.

2. Material and methods

The FruitDiv project deals with five CWR taxa: Malus, Pyrus, Prunus section Cerasus, Prunus section Prunophora and Prunus section Amygdalus. Initially, partners building the project thought they would build the core-collections on the basis of short-read sequencing data. However, it was feared that samples collected *in-situ* could be crop/domestic individuals or crop x wild hybrids with a significant crop genetic component, that would not correspond to FruitDiv goals regarding CWR. In order to avoid sequencing this kind of individuals, a preliminary screening step of the received samples was decided. As this delayed the launch of short-read sequencing, it was decided to build the core collections on the basis of these preliminary screening results.

2.1 Material

For all species, leaf samples from accessions in ex-situ collections and from in-situ CWR populations were received and DNA extracted. The numbers of samples received for each taxon, split by supposed species within taxon can be found in table 1.

Table 1: number of samples received per taxon

Taxon	Number of samples received	Number of samples kept for core collection building
Amygdalus	677 samples in 2024 <i>Prunus webbii</i> (N=415) <i>P. tenella</i> (N=177) <i>P. fenzliana</i> (N=40, 2 populations)	206 <i>Prunus webbii</i> <i>P. fenzliana</i> and <i>P. nairica</i> (75 individuals, 3 populations from FRUITDIV and one from previous project)

	<i>P. nairica</i> (N=20, 1 population) <i>P. scoparia</i> (N=19, 1 population) wild <i>P. dulcis</i> (N=6) 100 samples of Irakian <i>P. webbii</i> , <i>P. orientalis</i> , <i>P. scoparia</i> etc... are planned for study in 2025	Adding the other species in the core-collection will be discussed on a case-by-case basis
Pyrus	849 samples in 2024, ~300 samples expected in 2025 28 <i>Pyrus pyraeaster</i> populations CRA-W (partner 8) <i>Pyrus pyraeaster</i> collection (68 samples) 1 population from each of <i>Pyrus elaeagnifolia</i> , <i>Pyrus caucasica</i> , <i>Pyrus salicifolia</i>	276 <i>Pyrus pyraeaster</i>
Cerasus	516 wild <i>P. avium</i> samples in 2024, 133 additional samples expected in 2025	289 <i>P. avium</i>
Prunophora	665 samples in total <i>P. cerasifera</i> (N=350) <i>P. divaricata</i> (N=20) <i>P. cocomilia</i> (N=267) <i>P. ramburii</i> (N=28)	643 individuals, 404 individuals with seeds 433 individuals having 80% identity in Structure groups 268 individuals with seeds having 80% identity in Structure groups
Malus	727 samples in total	265 <i>Malus sylvestris</i> individuals

2.2 Methods

2.2.1 Genotyping

Genotyping in the preliminary screening step was performed using SSR markers for *Pyrus*, *Amygdalus*, *Cerasus* and *Prunophora* spp., and a 50K genotyping array for *Malus* spp. (Bianco et al 2016). For *Pyrus*, the chosen SSR markers were those routinely used at INRAE-IRHS for cultivar identification in *Pyrus communis*. For *Amygdalus*, *Prunophora* and *Cerasus*, the chosen SSR markers were identified from the literature. More details regarding SSR markers used are given in Appendix 1.

The numbers of markers used for genotyping in each taxon can be found in table 2, together with the numbers of markers that provided clear enough results for further use in core-collection building.

Table 2: number and types of markers used per taxon

Taxon	Number and type of markers used	Number of markers used for core-collection building
<i>Amygdalus</i>	24 nuclear SSR markers	15 SSR
<i>Pyrus</i>	22 nuclear SSR markers	20 SSR

Cerasus	10 nuclear SSR markers (nSSR) and 9 chloroplast SSR markers (cpSSR)	9 SSR
Prunophora	22 nuclear SSR markers	20 SSR
Malus	50K SNP markers	33,862 SNP

2.2.2 Structure analysis for species identification

For *Amygdalus*, *Pyrus*, *Cerasus* and *Prunophora*, sampled individuals were assigned to species, and possibly subdivided within species in genetic clusters, using Bayesian clustering analysis, performed with the Structure software. This analysis also enabled to identify individuals belonging to the crop related species of each taxon, and individuals resulting from crop x wild hybridization, or possibly from hybridization of the focus species with other wild species (ie *Cerasus* taxon). Individuals showing a low amount of admixture of the related crop species were retained for core-collection building. The tolerated amount of admixture was 20% for *Amygdalus*, *Pyrus* and *Cerasus*. For *Prunophora*, two options were considered: i) not selecting on admixture, because gene flow from the cultivated species (*P. domestica*, hexaploid, or *P. salicina* (Japanese plum), diploid but rare in Europe) was expected to be low, whereas gene flow from the unwanted species *P. spinosa* (tetraploid) was expected to result in sterile hybrids; gene flow between *P. cerasifera* and *P. cocomilia* was not foreseen to be a problem for the usefulness of the core-collection; ii) selecting on admixture, with a tolerated amount of 20%, which reduced the number of *P. cerasifera* individuals from 370 to 202, and the number of *P. cocomilia* individuals from 238 to 196. In addition, availability of seeds was considered or not in building the list of candidates.

For *Malus*, sampled individuals were assigned to species on the basis of the number of alleles specific of a species they had at some specific SNP. The list of SNPs, and alleles at these SNPs, that can be considered as specific of a species was shared by Nick Howard (FRESH FORWARD, Netherlands). A sample was considered pure *M. sylvestris* if at least 80% of the specific alleles belonged to that species.

2.2.3 Core-collection building

Using the genotypic data obtained in the preliminary screening step, and after selecting candidate individuals showing a low amount of admixture, CoreFinder v1.1 (Cipriani et al, 2010) was used, for *Amygdalus*, *Pyrus*, *Cerasus* and *Prunophora*, to rank the individuals in order of priority to be included into a core collection with the objective to maximize the number of alleles with the minimum of individuals. For *Malus*, CoreFinder was not able to cope with the large SNP dataset thus Darwin (Perrier and Jacquemoud-Collet, 2006) was used instead. In a first step, an unweighted neighbour joining (UNJ) tree was constructed with 500 bootstraps. In second step, a stepwise procedure, known as the *max length sub tree* method, was used to select a subset of individuals that reduces redundancy among individuals, while limiting loss in diversity. At each step an individual was removed to reduce the redundancy in the UNJ tree while aiming to maximize diversity. In other words, the individuals removed in the earlier steps are of less interest for a core collection and the set of individuals remaining in the later steps are of greater interest for a core collection.

2.3 Contributing beneficiaries

Table 3: Contributing beneficiaries

Contributing partner	People involved	Work performed
INRAE-BFP Amygdalus	Véronique Decroocq	Choice of candidates and populations, preparation of files for CoreFinder. Wrap up of results from Structure analysis
	Stéphane Decroocq	Bayesian and statistical analysis
	Wisam Khalid	Raw data (scans) analysis & Bayesian and statistical analysis
INRAE-IRHS Pyrus	Caroline Denancé	DNA extraction, PCR for SSR analysis, raw data (scans) analysis
	Jade Bruxaux	DNA extraction, Structure analysis for Pyrus and Malus, choice of candidate individuals for Pyrus
	Hélène Muranty	First and second draft of the deliverable report, WP2 leader
INRAE-AGAP	Marion Sinclair-Waters	CoreFinder and Darwin analysis
AUTH Cerasus	Phil Aravanopoulos	Selection of candidate populations, Cerasus genetic analysis report
	Nickos Tourvas	DNA extraction, raw data analysis , PCR for SSR analysis, Structure analysis
	Argyro Srampouli	DNA extraction, PCR for SSR analysis
CRAG Prunophora	Maria José Aranzana	Population structure and diversity analysis, select candidates and populations
	Iban Eduardo	Coordinate the SSR genotyping, select the SSR markers, raw data analysis
	Paloma Cabezas	DNA extraction, PCR and SSR genotyping, raw data analysis
	Pol Rey	Population structure and diversity analysis, select candidates and populations, preparation of files for CoreFinder
FEM Malus	Michela Troggio	Selection of samples for pre-screening
	Diego Micheletti	Analysis of the Apple Axiom 50K array and choice of candidate individuals for Malus
	Sylvia Lorenzi	DNA extraction and sample preparation
UU	Pascal Milesi	WP2 co-leader

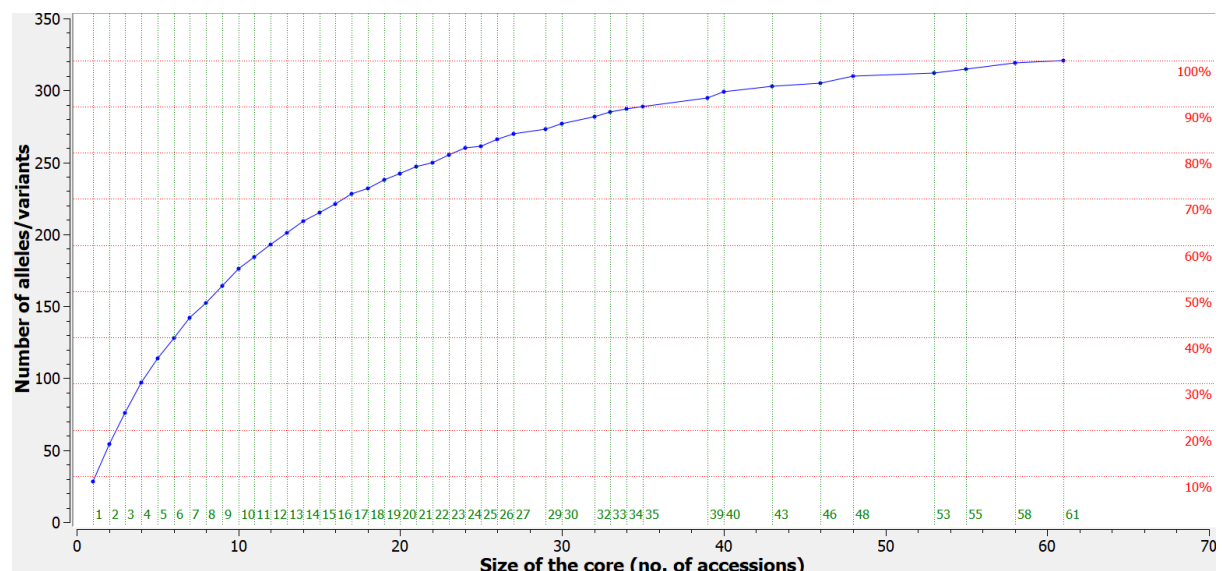
3. Results

3.1 Amygdalus

Starting from 206 *Prunus webbii* individuals genotyped for 15 SSR markers, a core-collection of 61 individuals, maximizing allelic diversity up to 100% was proposed. The increase in number of alleles represented with the number of individuals included in the core collection can be seen in figure 1. The list of individuals selected is depicted in the FRUITDIV sharepoint under the link:

https://sites.inrae.fr/site/project-fruitdiv/WP2%20Library/T2.3_Building%20core-collections/Amygdalus/CoreColl_Results/Pwebbii_CFresults.csv?d=wa1b75c499dbd4b7aa7d494f2851e664b

Figure 1: Number of alleles versus the number of individuals included in a hypothetical core collection in the case of *Prunus webbii*

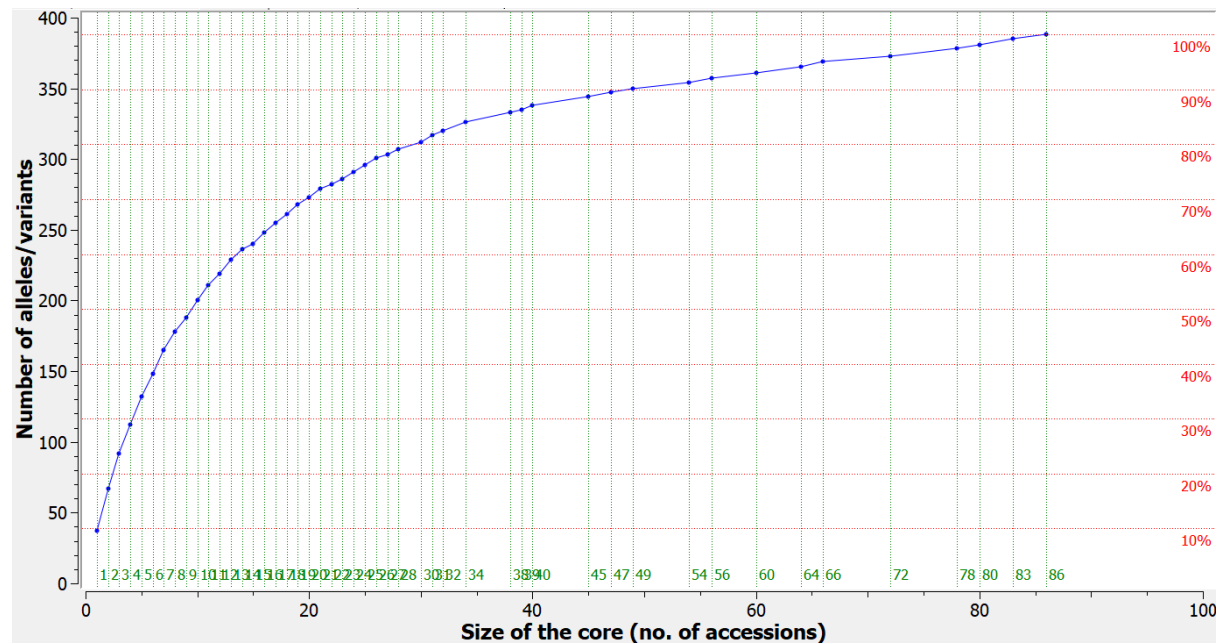


3.2 Pyrus

Starting from 276 *Pyrus pyraeaster* individuals genotyped for 20 SSR markers, a core-collection of 86 individuals, maximizing allelic diversity up to 100% was proposed. The increase in number of alleles represented with the number of individuals included in the core collection can be seen in figure 3. The list of individuals selected is depicted in the FRUITDIV sharepoint under the link:

https://sites.inrae.fr/site/project-fruitdiv/WP2%20Library/T2.3_Building%20core-collections/Pyrus/CoreColl_results/CoreCollection_Pyrus_CFresults.csv?d=w8ca2abcefd9345e581ca15846a1af86a

Figure 2: Number of alleles versus the number of individuals included in a hypothetical core collection in the case of *Pyrus pyrastrer*

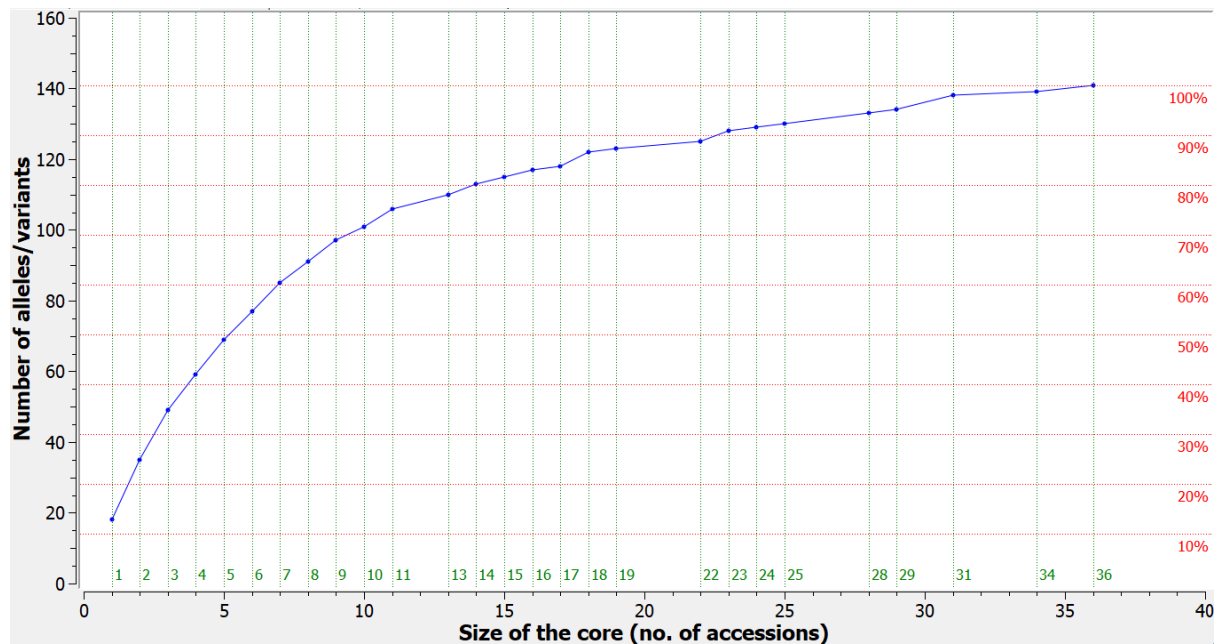


3.3 Cerasus

Starting from 289 *Prunus avium* individuals genotyped for 9 SSR markers, a core-collection of 36 individuals, maximizing allelic diversity up to 100% was proposed. The increase in number of alleles represented with the number of individuals included in the core collection can be seen in figure 3. The list of individuals selected is depicted in the FRUITDIV sharepoint under the link:

https://sites.inrae.fr/site/project-fruitdiv/WP2%20Library/T2.3_Building%20core-collections/Cerasus/CoreCollection_Cerasus_CFresults.csv?d=w4cc0ae465c4542a2853b79443e324fe8

Figure 3: Number of alleles versus the number of individuals included in a hypothetical core collection in the case of *Prunus avium*

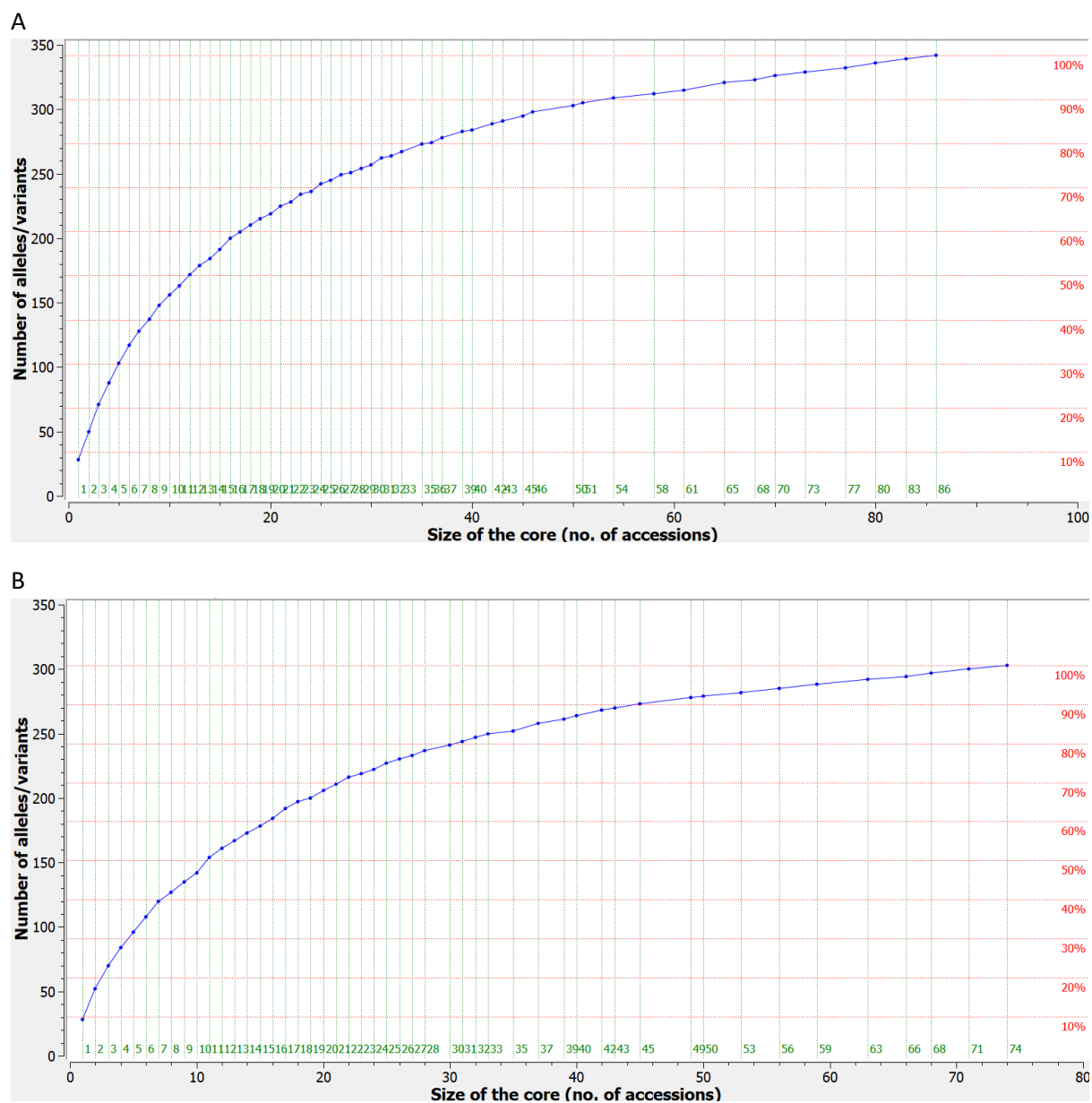


3.4 Prunophora

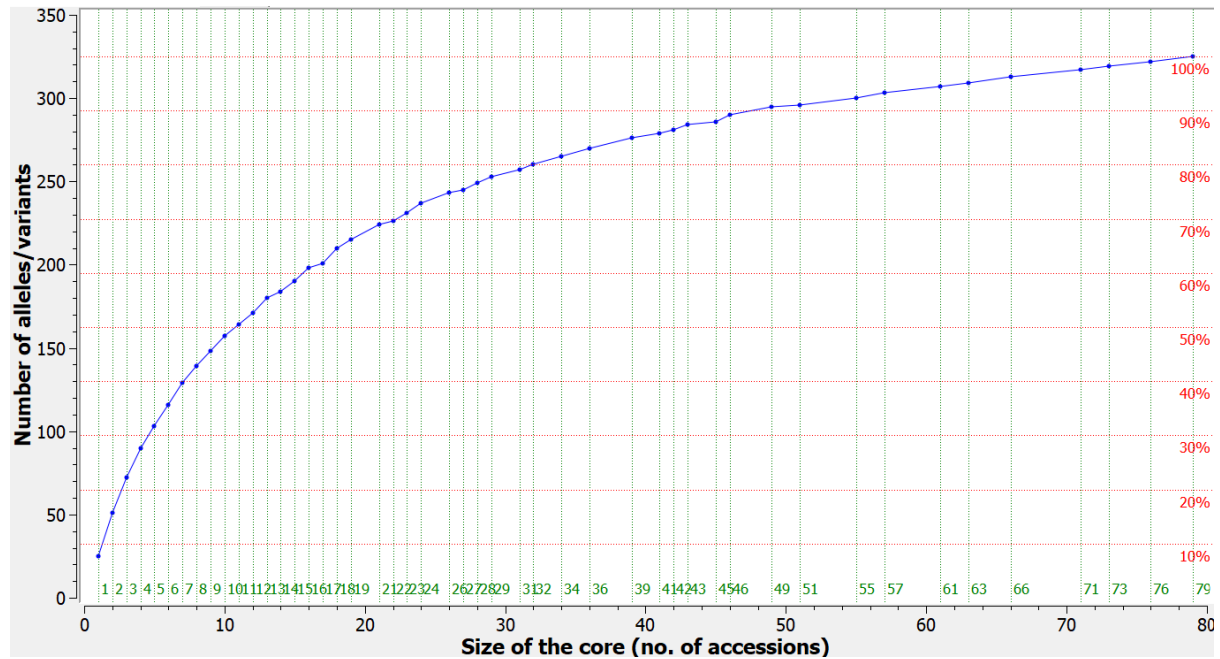
When neither selection on admixture nor on seed availability was performed, starting from 643 *Prunophora* individuals genotyped for 15 SSR markers, a core-collection of 86 individuals, maximizing allelic diversity, was proposed. When the individuals were restricted to those for which seeds were available and without selection on admixture, a core-collection of 74 individuals was proposed that represents 89% of the *prunophora* allelic diversity. When selection on admixture was performed, the starting number of individuals was reduced to 433, the core-collection proposed had a size of 79 individuals and represented 95% of the allelic diversity of the core-collection proposed when not selecting on admixture. When selection on admixture and availability of seeds was performed, the starting number of individuals was reduced to 268, the core-collection proposed had a size of 66 individuals and represented 85% of the allelic diversity of the core-collection proposed when not selecting on admixture nor on availability of seeds, and 96% of the allelic diversity of the core-collection proposed when only selecting on seed availability. The increase in number of alleles represented with the number of individuals included in the core collection can be seen in figure 4a for all individuals, 4b when considering only individuals with seeds, 4c when selecting on admixture only, and 4d when selecting on admixture and availability of seeds. The list of individuals selected is depicted in the FRUITDIV sharepoint under the link:

https://sites.inrae.fr/site/project-fruitdiv/WP2%20Library/T2.3_Building%20core-collections/Prunophora/CorrColl_Results, files : CoreCollection_Prunophora_CResults_{...}.csv

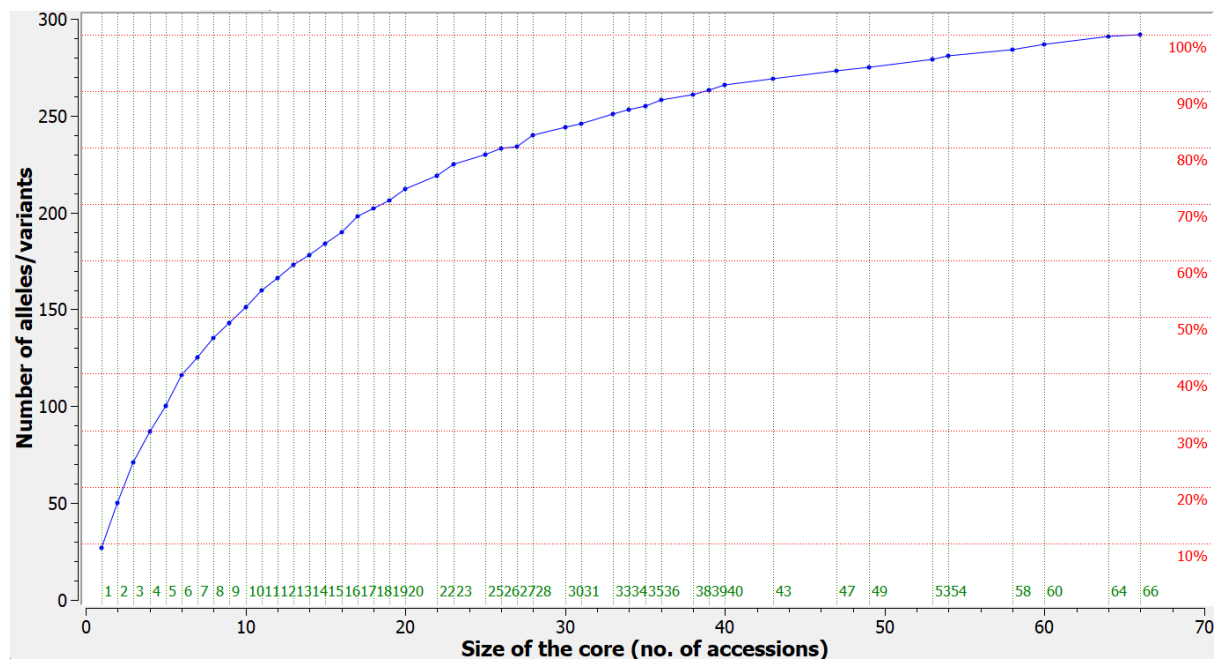
Figure 4: Number of alleles versus the number of individuals included in a hypothetical core collection in the case of individuals from the *Prunophora* taxon. A : all individuals, B : only individuals with seeds, C: individuals with more than 80% identity in Structure groups, D: individuals with more than 80% identity in Structure groups and with seeds



C



D



3.5 Malus

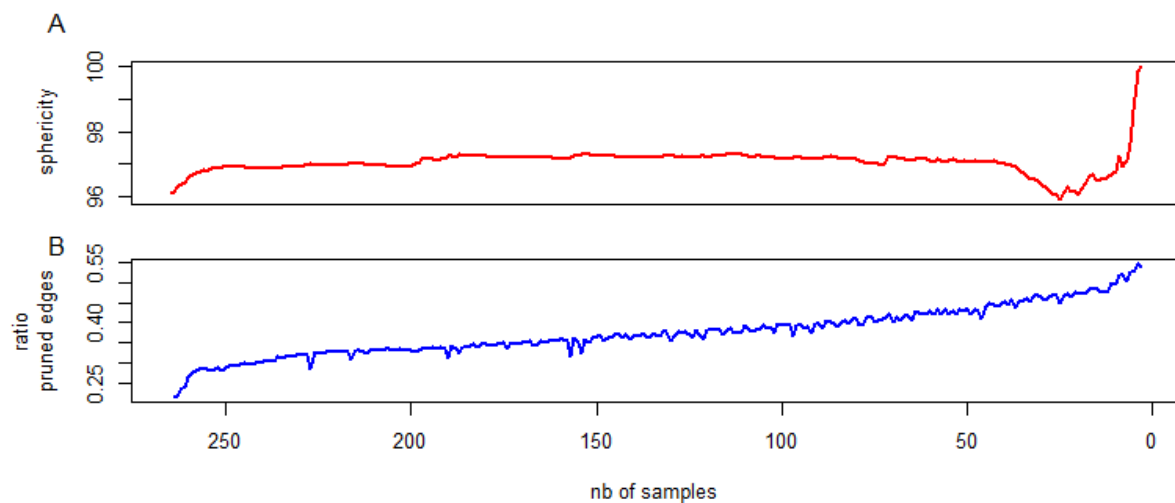
As a result of the Darwin analysis, the individuals were ordered according to their interest for building a core-collection, from the lowest interest, in the initial steps, to the highest interest, in the last steps. The procedure lets to the user the choice of the sample size to retain. The results can be found in the FRUITDIV sharepoint under the link:

https://sites.inrae.fr/site/project-fruitdiv/WP2%20Library/T2.3_Building%20core-collections/Malus/Msyt_noAdmixed_Darwin_results.txt

Figure 5 illustrates the decrease in redundancy and eventually diversity when removing individuals, with two indices (Darwin manual p 75):

- the 'sphericity index' that is the ratio of the lengths of external edges to the total length of the current sub UNJ tree. This index has low values for trees with numerous redundant units (=individuals) (a lot of small external edges) and tends towards 1 when all units are independent (a lot of long external edges, the UNJ tree tends towards a star-like tree)
- the length of the pruned edge expressed in percent of the initial UNJ tree length

Figure 5: A) Sphericity index and B) length of the pruned edge, in percent of initial UNJ tree length, versus the number of individuals in the UNJ tree. Please note that x-axis decreases from left to right.



Conclusions

Using genotypic characterization data obtained in a preliminary screening step, put in place in order to avoid sequencing crop/domestic individuals or crop x wild hybrids with a significant crop genetic component, we were able to propose core-collections maximizing diversity with a reduced number of individuals: in *Amygdalus*, the size of the proposed core-collection is 61 individuals, in *Pyrus*, 86 individuals, in *Cerasus*, 36 individuals, in *Prunophora*, 86 / 74 / 79 / 66 depending on the starting set of individuals. In *Malus*, the analysis approach did not indicate a number of individuals to include in the core collection but only the order of removal of the individuals to reduce redundancy while maximizing diversity.

It should be noted that preliminary genotypic characterization data were not available for a part of or all samples collected in 2025, thus the core-collections will likely evolve when these data will be available.

Additionally, the size of the core-collections has to be decided in view of the dual goal assign to them: on one hand, contribute to the ex-situ conservation of CWR genetic diversity, on the other hand, establish a long-term experimental design allowing to disentangle genetic effects and GxE interactions in CWR through an efficient phenotyping (see FruitDiv WP3 objectives), which mean enabling GWAS.

It should be recalled that maximizing diversity and having sufficient power for GWAS do not necessarily go hand in hand. Thus, extending the core-collections by adding some redundancy in them, in order to repeat the genes in various backgrounds, will probably be necessary.

Appendix 1: Names and sources of SSR markers

Taxon	SSR name	Reference	DOI	Developed for
Amygdalus	BPPCT001	Dirlewanger et al 2002	10.1007/s00122-002-0867-7	<i>Prunus persica</i>
	BPPCT010			
	BPPCT017			
	BPPCT025			
	BPPCT027			
	BPPCT036			
	CPDCT005	Mnejja et al 2005	10.1111/j.1471-8286.2005.00977.x	<i>Prunus dulcis</i>
	CPDCT025			
	CPDCT034			
	CPDCT045			
	CPDCT035*			
	CPPCT033	Aranzana et al. 2002	10.1046/j.1439-0523.2002.00656.x	<i>Prunus persica</i>
	CPSCT012	Mnejja et al 2004	10.1111/j.1471-8286.2004.00603.x	<i>Prunus salicina</i>
	CPSCT018			
	EPDCU5100	Mnejja et al 2010	10.1007/s11295-11010-10284-z	<i>Prunus dulcis</i>
	pchgms1	Sosinski et al 2000	10.1007/s001220051499	<i>Prunus persica</i>
	pchgms3			
	UDP96-001	Cipriani et al. 1999	10.1007/s001220051209	<i>Prunus persica</i>
	UDP96-003			
	UDP96-005			
	UDP96-013			
	UDP96-018			
	UDP97-401			
	UDP98-409			
	UDP98-408*			
Pyrus	CH01d08	Liebhard et al., 2002	10.1023/A:1020525906332	<i>Malus domestica</i>
	CH01d09			
	CH01f03b			
	CH01f07a			
	CH01h01			
	CH01h10			
	CH02b10			
	CH02c11			
	CH03d12			
	CH03g07			
	CH04c07			
	CH04e03			
	CH05c06			
	CH05f06			
	CH-Vf1	Vinatzer et al. 2004	10.1111/j.1439-0523.2004.00973.x	<i>Malus domestica</i>
	EMPc11			<i>Pyrus communis</i>

Taxon	SSR name	Reference	DOI	Developed for
	EMPc117	Fernández-Fernández et al., 2006	10.1111/j.1471-8286.2006.01422.x	
	GD142	Hokanson et al. 1998	10.1007/s001220050943	<i>Malus domestica</i>
	GD147			
	NAUpy40d	Fan et al 2013	10.1007/s11105-013-0586-z	<i>Pyrus bretschneideri</i>
	TsuENH080_354	Nishitani et al 2009	10.1270/jsbbs.59.391	<i>Pyrus pirifolia</i>
	TsuENH089			
Cerasus	EMPaS01	Vaughan & Russell 2004	10.1111/j.1471-8286.2004.00673.x	<i>Prunus avium</i>
	EMPaS02			
	EMPaS06			
	EMPaS10			
	EMPaS11			
	EMPaS12			
	EMPaS14			
	EMPA002	Clarke & Tobutt 2003	10.1046/j.1471-8286.2003.00517.x	<i>Prunus persica</i>
	EMPA004			
	EMPA005			
	EMPA015			
	EMPA017			
	EMPA018			
	UDP98-412	Testolin et al 2000	10.1139/g00-010	<i>Prunus persica</i>
	PceGA34	Downey & Iezzoni 2000	10.21273/JASHS.125.1.76	<i>Prunus cerasus</i>
Prunophora	CPPCT30	Aranzana et al. 2002	10.1046/j.1439-0523.2002.00656.x	<i>Prunus persica</i>
	CPPCT33			
	UDP96-018	Cipriani et al. 1999	10.1007/s001220051209	<i>Prunus persica</i>
	UDP96-013			
	UDP96-003			
	UDP97-401			
	UDP96-001			
	UDP98-408			
	UDP98-409			
	BPPCT001	Dirlewanger et al 2002	10.1007/s00122-002-0867-7	<i>Prunus persica</i>
	BPPCT010			
	BPPCT017			
	AMPA101	Hagen et al. 2004	10.1111/j.1471-8286.2004.00802.x	<i>Prunus armeniaca</i>
	AMPA100			
	EPPCU0532	Howad et al 2005	10.1534/genetics.105.043661	<i>Prunus persica</i>
	SSR5piso4Ga	Marandel et al 2009	10.1111/j.1365-3059.2008.02012.x	<i>Prunus persica</i>
	CPDCT025	Mnejja et al 2005	10.1111/j.1471-8286.2005.00977.x	<i>Prunus dulcis</i>
	CPDCT034			
	CPDCT018			
	pchgms03	Sosinski et al 2000	10.1007/s001220051499	<i>Prunus persica</i>
	pchgms1			
	UDA-021	Testolin et al 2004	10.1111/j.1471-8286.2004.00700.x	<i>Prunus dulcis</i>