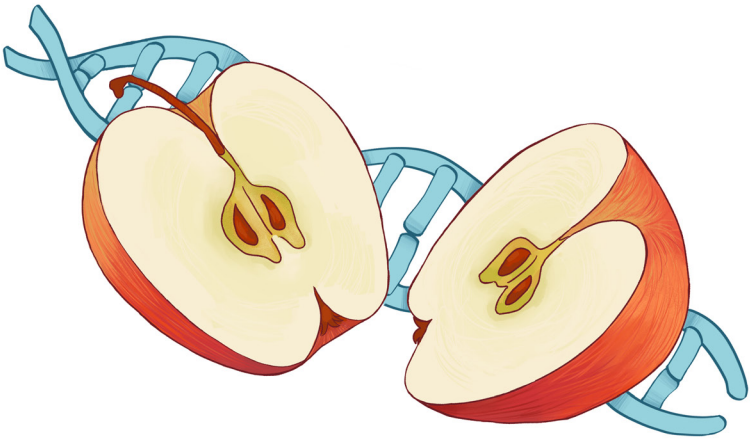
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2023:18
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Apple genomics for the Swedish breeding programme

Jonas Skytte af Sätra



Apple genomics for the Swedish breeding programme

Jonas Skytte af Sätra
Faculty of Landscape Architecture, Horticulture and Crop Production Sciences
Department of Plant Breeding
Alnarp

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Alnarp 2023

Acta Universitatis Agriculturae Sueciae
2023:18

Cover: A bisected apple and a DNA double helix. Illustration: Ulv Weström

ISSN 1652-6880
ISBN (print version) 978-91-8046-088-0
ISBN (electronic version) 978-91-8046-089-7
https://doi.org/10.54612/a.4usn6qgb7l

Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden
Print: SLU Service/Repro, Alnarp 2023

Apple genomics for the Swedish breeding programme

Abstract

Swedish apple (*Malus domestica* Borkh.) cultivation is a niche market setting specific requirements on new apple cultivars. Notably, the growing season is short, especially in central and northern Sweden, and there are relatively few pesticides available, making resistance to diseases important. The past decade has seen major developments in tools for genomics-led breeding in apple and targeted application of these tools could facilitate a major increase in efficiency. The aim of this thesis has been to lay the foundations for genomics-led breeding in the Swedish apple breeding programme. In a first step, the status of available genetic resources was investigated and curated (Paper I). A robust high-resolution virtual linkage map (Paper II) was then developed by analysis of two whole genome sequences and use of a high-density linkage map. Identification of pedigree relationships and use of a reliable map for marker ordering enabled production of highly curated and phase-resolved marker data. Apple germplasm was screened for resistance to European canker (*Neonectria ditissima*) and quantitative trait loci (QTL) in segregating offspring of 'Aroma' x 'Discovery' were mapped (Paper III). The metabolomic profiles of the parents during early stages of infection provided insight on the potential roles of two of the mapped QTL, and indicated that 'Santana' might provide a complementary source of resistance (Paper IV). The effect of previously published genomic regions and QTL intervals for date of flowering and harvest date under Nordic conditions were investigated in germplasm relevant for Nordic apple breeding (Paper V). A study on the genetic basis for adaptation to northern Sweden was initiated, with the timing of canopy senescence identified as a correlating trait (Paper VI). The applicability of the novel data obtained was illustrated by discussing future crosses that could be relevant for breeding to improve resistance to *N. ditissima* and adaptation to central and northern Sweden.

Keywords: plant breeding, *Malus*, QTL mapping, SNP array, European canker

Apple genomics for the Swedish breeding programme

Abstract

Den svenska marknaden för äppelodling är en nischmarknad med specifika krav på vissa odlingsegenskaper. Till exempel är odlingssäsongen kort, framförallt i mellersta och norra Sverige, och det finns relativt få kemiska bekämpningsmedel tillgängliga, vilket gör resistens mot olika sjukdomar extra viktigt. Under det senaste decenniet har det skett en kraftig utveckling av verktyg för genombaserad växtförädling av äpple, och att utnyttja dessa skulle kunna bidra till en kraftigt ökad effektivitet i växtförädlingsprogrammet på äpple. Därför har målet med denna avhandling varit att etablera grundläggande infrastruktur, i form av kunskap, färdighet och datainsamling, för att arbeta med genombaserad växtförädling i det svenska växtförädlingsprogrammet på äpple. Först undersöktes statusen på tillgängliga genetiska resurser (Paper I). Sedan utvecklades en robust och högupplöst virtuell genetisk karta efter att två helgenomsekvenser och en högupplöst kopplingskarta kurerats (Paper II). Tillgång till ett extensivt släktträd och en pålitlig genetisk karta gjorde det möjligt att generera kurerad och kopplingsfasad markördata. Motståndskraften mot fruktträdskräfta (*Neonectria ditissima*) undersöktes i en samling äppelsorter och kvantitativa loci (QTL) som styr egenskapen karterades i en segregerande avkomma av ’Aroma’ x ’Discovery’ (Paper III). Två av de identifierade QTL diskuterades i relation till metabolomprofilerna hos föräldrarna vid infektion (Paper IV). Effekten av tidigare publicerade QTL intervall och genomiska regioner för tid för blomning och fruktmognad under nordiska förhållanden undersöktes i en samling sorter relevanta för nordisk äppelförädling (Paper V). En undersökning av den genetiska basen för klimatanpassning påbörjades och visar att tid för höstsenescens är korrelerat med vilken odlingszon en sort är rekommenderad för (Paper VI). Till sist exemplifieras användningen av resultaten genom en diskussion om relevanta korsningar.

Keywords: växtförädling, *Malus*, QTL kartering, SNP array, fruktträdskräfta

Preface

"The Wheel of Time turns, and Ages come and pass, leaving memories that become legend. […] In one Age […] a wind rose in the Mountains of Mist. The wind was not the beginning. There are neither beginnings nor endings to the turning of the Wheel of Time. But it was *a* beginning."

-Robert Jordan, from Book One of The Wheel of Time.

Thus, while this thesis is not *the* ending, it is *an* ending to a journey. A journey that might have begun with *Ragnarök*: the end of the world and the birth of something new. Likewise, it is not just *an* ending but might also be the seed of *a* beginning.

Jonas Skytte af Sätra
Ängelholm, February 2023

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. Skytte af Sätra J., Troggio M., Odilbekov F., Sehic J., Mattisson H., Hjalmarsson I., Ingvarsson P.K., Garkava-Gustavsson L. (2020) Genetic status of the Swedish Central Collection of heirloom apple cultivars. *Scientia Horticulturae* 272. https://doi.org/10.1016/j.scienta.2020.109599

II. Larsen B.†, Skytte af Sätra J. †, Garkava-Gustavsson L., Howard N.P., van de Weg E. Integrating linkage and sequence-based maps in apple: a systematic evaluation resulting in polynomials to generate standardized virtual linkage maps. (manuscript) †Shared first authorship.

III. Skytte af Sätra J., Odilbekov F., Ingvarsson P.K., van de Weg E., Garkava-Gustavsson L. (2023) Parametric mapping of QTL for resistance to European canker in apple in 'Aroma' x 'Discovery'. *Tree Genetics and Genomes* 19(2). https://doi.org/10.1007/s11295-023-01587-w

IV. Garkava-Gustavsson L., Skytte af Sätra J., Odilbekov F., Abreu I., Johansson A.I., van de Weg E., Zhebentyayeva T. Resistance to *Neonectria ditissima* in apple: insights from metabolomics and lipidomics analyses. *Acta Horticulturae* (accepted)

V. Skytte af Sätra J., Røen D., Haikonen T., Rumpunen K., Hjeltnes S.H., Garkava-Gustavsson L. Validation of genomic regions associated with phenological traits in apple (*Malus domestica*) under Nordic conditions. (manuscript)

VI. Skytte af Sätra J., Hjalmarsson I., Ingvarsson P.K., Garkava-Gustavsson L. Characterization of autumn senescence in the Swedish Central Collection of heritage apple cultivars. *Acta Horticulturae* (submitted)

Papers I and III are reproduced with the permission of the publishers.

The contribution of Jonas Skytte af Sätra to the papers included in this thesis was as follows:

I. Conceptualised the study together with the co-authors, performed all data analysis and wrote the final version of the manuscript with input from the co-authors.

II. Contributed to the design of the study, performed the data analysis and wrote the final version of the manuscript with input from the co-authors.

III. Collected phenotypic data from the FS-family, conceived and performed all data analysis and wrote the final version of the manuscript with input from the co-authors.

IV. Performed all data analysis and contributed to writing the manuscript.

V. Conceived and performed all data analysis and wrote the final version of the manuscript with input from the co-authors.

VI. Designed the study, collected all data, performed all data analysis and wrote the final version of the manuscript with input from the co-authors.

Abbreviations

AUC	Area Under Curve
AUDPC	Area Under Disease Progression Curve
BLUE	Best Linear Unbiased Estimate
DPI	Days Post Inoculation
FS	Full-Sibling
HB	Haploblock
IBD	Identical By Descent
IBS	Identical By State
LD	Linkage Disequilibrium
LG	Linkage Group
MAPS	Marker-Assisted Parent Selection
MASS	Marker-Assisted Seedling Selection
NTTT	Not-True-to-Type
PBA	Pedigree-Based Analysis
PCA	Principal Component Analysis
QTL	Quantitative Trait Locus (singular) or Loci (plural)
SLU	Swedish University of Agricultural Sciences
SNP	Single Nucleotide Polymorphism
TTT	True-To-Type
WGS	Whole Genome Sequence

1. Background

1.1 The domesticated apple

The domesticated apple (*Malus domestica* Borkh.) is diploid, though triploid individuals occur. Triploid individuals make poor parents for breeding as they have low fertility, but are generally more vigorous and give larger fruits than diploids. Accordingly, they are relatively common among cultivars as these properties have been favoured by human selection. The domesticated apple most likely originated in the Tian Shan mountains, in today's Kazakhstan. From there, individuals of *Malus sieversii* appear to have spread to Europe along the Silk Road, arriving at around the time of the roman empire, and hybridised extensively with the European crabapple (*Malus sylvestris*). Individuals of *M. sieversii* also spread eastward along the Silk Road and hybridised with the Siberian crabapple (*Malus baccata*), leading to a slightly different kind of apple than in Europe. Interestingly, *M. sieversii* seems to have been largely self-domesticated, acquiring typical domestication traits such as large fruits that do not drop off the tree, probably as an adaptation to dispersal by large mammals. *M. domestica* shares many of the characteristics typical of domesticated perennial plant species, it is long lived and clonally propagated, is an self-incompatible outcrosser, and has an extended juvenile period (Cornille et al., 2014; Duan et al., 2017; Gaut et al., 2015; Spengler, 2019; Sun et al., 2020). More recently, other wild *Malus* species have been hybridised with *M. domestica* in order to introduce various traits, in particular disease resistance from e.g. *Malus floribunda* and *Malus robusta* (Bus et al., 2011, 2019; Janick, 2006). Taken together, this history has contributed to the very weak domestication bottleneck observed

in domesticated apple, its high degree of heterozygosity, and its large amount of accessible genetic variation.

1.2 The Swedish market

Sweden is part of the Scandinavian Peninsula, lying between 55°N and 69°N, and the growing season is short, especially in central and northern parts of the country. Approximately 20-25% of the dessert apples consumed come from domestic production and the vast majority of production is located in southern Sweden. Due to the small size of the domestic apple market, Swedish growers have limited access to pesticides, making resistances to diseases highly prioritised traits. Most of the pesticides applied in a typical orchard are aimed at controlling apple leaf scab (*Venturia inaequalis*) (Nybom, 2019).

Sweden is divided into nine growing zones (zones 1-8 and the alpine region) for perennial horticultural crops (Fernqvist, 1993). Some orchards have recently been established in zone 5 (around Umeå, 63°N), but most production is concentrated to zones 1 and 2.

1.2.1 Specific challenges

One of the main threats to Swedish commercial orchards is European canker, a fungal disease caused by *Neonectria ditissima*. Mild symptoms result in decreased fruit bearing capacity, while severe attacks can kill entire trees. In extreme cases, entire orchards have been destroyed due to this pathogen. In all cases, the profitability of apple production is affected by the disease. The complete lack of available fungicides to control *N. ditissima* further exacerbates the problem (Garkava-Gustavsson et al., 2016, 2013; Gustavsson et al., 2022). Thus, breeding for improved resistance to *N. ditissima* is of high importance.

The local climate also poses special challenges to domestic apple production. Notably, the growing season in Sweden is relatively short, even in the southernmost parts of the country. In addition, as spring becomes earlier due to ongoing global warming the risk of frost during flowering might increase in the future, though it is already perceived as an important problem (Hänninen, 2006; Legave et al., 2013; Monell, 2021; Nybom, 2019; Vitasse et al., 2018).

1.3 The Swedish breeding programme(s)

The first apple cultivars released from a domestic breeding programme in Sweden came from the experimental facility in Alnarp, initially known as *Permanenta Kommittén för Fruktodlingsförsök* and later as *Statens Trädgårdsförsök*. These cultivars included 'Snövit' (1955), 'Mio' (1955), 'Silva' (1970), and the tetraploid cultivar 'Alfa 68' (1953). In the 1940s, a dedicated breeding programme was established outside Kristianstad. It became known as the Balsgård programme, after the nearby mountain Balsberget, and operated as an independent entity until it became part of Swedish University of Agricultural Sciences (SLU) in 1970. A number of cultivars have been released from the Balsgård programme over the years, starting with 'Alice' (1963) followed by e.g. 'Katja' (1966), 'Aroma' (1973), 'Rödluvan' (1994), and 'Frida' (2007) (Nilsson, 1987; Svensson, 2005). At the time of writing, the Balsgård programme is in the process of being transferred to Alnarp. The breeding work is temporarily on hold, to free up the resources necessary for the transfer, and thus the cultivars described as originating from the Alnarp programme in this thesis refer to those coming from the experimental facility that operated during the 20th century.

Until work was paused, the Swedish apple breeding programme relied extensively on the 'breeder's eye' and phenotypic data for parent and seedling selection, although molecular markers were used to some extent. The programme was based upon clonal seedling selection with a recurrent breeding process implemented in a wide sense, with advanced selections or released cultivars being used as parents in future crosses.

1.3.1 Novel tools with potential

In the past 15 years, there has been rapid development in terms of tools available for genomic research and genome-informed breeding in apple. These include high-density single nucleotide polymorphism (SNP) marker arrays (Bianco et al., 2016, 2014; Chagné et al., 2012), calling and curation pipelines (Howard et al., 2021a; Vanderzande & Howard et al., 2019), several high-quality whole genome sequences (WGSs) (Daccord et al., 2017; Khan et al., 2022; Sun et al., 2020; Velasco et al., 2010; Zhang et al. 2019), a high-density linkage map (Di Pierro et al., 2016), pedigree-based analysis (PBA) implemented through FlexQTL™ (Bink et al., 2014, 2008), and a large number of published quantitative trait loci (QTL) intervals and genomic regions for relevant traits and markers, some of which have been

found to be reproducible (e.g. Allard et al., 2016; Bus et al., 2021, 2019; Karlström et al., 2022; Larsen et al., 2019; Migicovsky et al., 2021; Minamikawa et al., 2021; Urrestarazu et al., 2017). Exploiting these advances could facilitate transformation of the Swedish apple breeding programme into a more efficient and data-driven system. However, some inconsistencies have been reported between the iGLMap and the GDDH13v1.1 (Peace et al., 2019), and clarification of these would benefit the use of both maps.

One of the innovative approaches developed in recent years is the application of PBA for fruit tree breeding and genetic research. It is implemented in the software FlexQTL™ and facilitates QTL mapping from multiple pedigree-related families simultaneously. Availability of pedigree structures allows modelling of transmission of identical by descent (IBD) haplotypes through lineages. A Bayesian framework enables simultaneous estimation of number of QTL, position of QTL, and QTL effects. While FlexQTL has mostly been applied in PBA, it can also be used for QTL mapping in single biparental crosses, and predicted QTL genotypes can be utilized in genomic prediction (Bink et al., 2014, 2008; Sujeet and Whitaker, 2018). Additionally, FlexQTL has a number of useful functions, such as reporting of both Mendelian consistent and inconsistent errors, prediction of genotype probabilities, and calculation of IBD probabilities. It should be noted that FlexQTL requires a genetic or physical map, and that Mendelian consistent errors that could be caused by an erroneous marker order of the map will be flagged.

2. Aims and objectives of the thesis

There have been dramatic advances recently in the genomic tools available for apple genetic and breeding research, partly as a result of some very successful international projects, such as HiDRAS, RosBREED, and FruitBreedomics. The Swedish apple breeding programme has utilised these resources to a limited extent to date, even though they could improve the efficiency of the breeding programme and facilitate better integration with the core activities of the host organisation. In order to rectify this, SLU Grogrund funded a project entitled "Accelerated and cost-efficient cultivar development through genome-informed apple breeding" with the goal of establishing necessary infrastructure for utilisation of available novel tools in terms of both data collection and competence development. The PPP project "Nordfruit", running in parallel, extended the genotypic data-set and produced phenotypic data, which were of benefit to the SLU Grogrund-funded project. Against this background, the aims of this thesis were to:

- Evaluate the genetic status of the available gene pool of apple cultivars, starting with the Swedish Central Collection of heirloom cultivars.

- Improve the resolution of the available consensus genetic map for *M. domestica*, and resolve some of the reported inconsistencies between iGLMap and GDDH13v1.1 WGS data.

- Identify novel QTL for breeding for improved resistance to European canker that could be relevant for the Swedish breeding programme.

- Validate previously published QTL regions that could be relevant for Nordic conditions.

- Initiate preliminary studies on the genomic basis in apple for adaptation to conditions in central and northern Sweden.

Towards this end, genotyping in this thesis was carried out using the Illumina 20 K apple Infinium® SNP array (Bianco et al., 2014).

3. Results and discussion

3.1 Genetic status of available resources

Investigations on the genetic status of the Swedish Central Collection of heirloom apple cultivars identified a number of synonymous accessions and some accessions which were likely not-true-to-type (NTTT) (Paper I), prompting further investigations. For example, the tree representing the cultivar 'Björnegårdsäpple' seemed to have been lost from both the local clonal archives and the central collection, as the sample was synonymous to the rootstock 'A2'. In further investigations, budwood from the presumed mother tree of 'Björnegårdsäpple' was located and was found to be genetically unique in the collection of samples obtained, thus likely being true-to-type (TTT) (unpublished work). In another example, Paper I showed that the cultivar 'Hedemoraäpple' was synonymous with 'Grågylling', despite these generally being considered to be separate cultivars. Later analysis of a tree planted in the 1930's in a private garden close to Hedemora revealed that it was also synonymous with 'Grågylling' (unpublished data).
Five key founders among the Swedish heirloom cultivars were identified in Paper I. These were 'Gimmersta' (most likely a red sport of 'White Astrachan'), 'Grågylling', 'Vitgylling', 'Rosenhäger', and 'Klockhammarsäpple'. Since the publication of Paper I, additional material has been genotyped within this thesis project or in sister projects with material provided for genotyping by Lidija Bitz (Luke), Dag Røen (Graminor), Jakob Sandberg (Fredriksdal), and Gordana Đurić and Mirela Kajkut Zeljković (University of Banja Luka). In addition, Eric van de Weg and Bjarne Larsen (Wageningen University and Research) kindly shared 20K SNP array data for some genotypes, and Nicholas P. Howard has shared valuable pedigree information for some genotypes. Thus, by the end of the work in this thesis, 20K SNP array data had been obtained for over 700 unique diploid apple cultivars and advanced selections, as well as three large full-sibling (FS) families. Of these cultivars and selections, 59% were found to be involved in at least one parent-offspring relationship, most of which formed a large interconnected network (Figure 1a, unpublished work). In general, the results in Paper I indicated that most modern Swedish cultivars exploit only a fraction of the available genetic variation. In a principal component analysis (PCA) plot of the full data set (excluding some

genotypes with a very large contribution from exotic crab apples), the cultivars released from the Balsgård breeding program cluster separately from most of the Swedish heirloom cultivars (Figure 1b, unpublished work). The only exception being ‘Rödluvan’, the only cultivar released from the Balsgård programme that is recommended for cultivation in northern Sweden (zone 6) (Svensson, 2005). Indeed, some heirloom cultivars are considered to be very hardy and are recommended for cultivation in northern Sweden, e.g. ‘Antonovka Obyknovennaja’, ‘Grågylling’, ‘White Transparante’, ‘Sävstaholm’, and ‘Suislepp’. Thus, the modern Swedish cultivars appear to be genetically distinct from the Swedish heirloom cultivars. This might in part indicate that the need for hardy cultivars for use in central and northern Sweden has been given relatively little attention by the breeding programme.

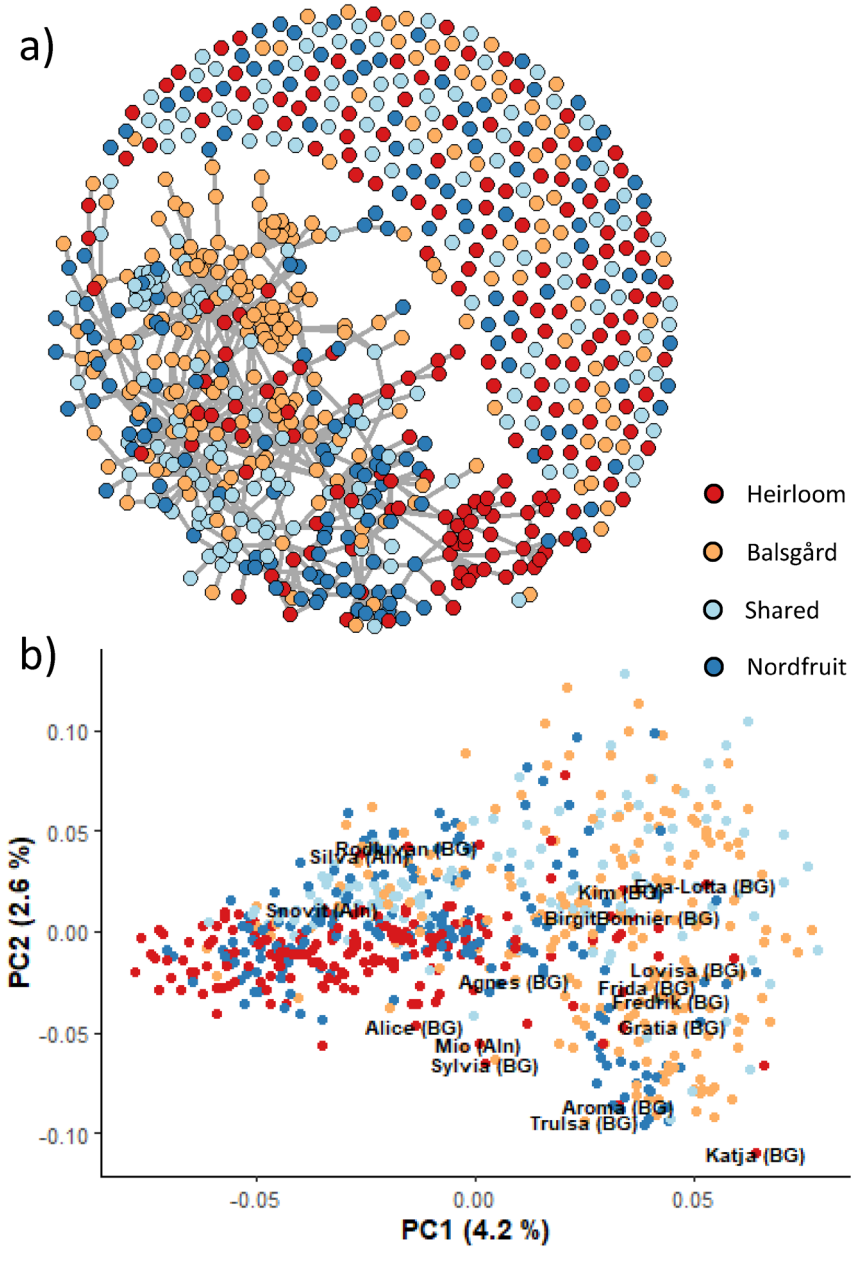


Figure 1. (a) Network plot and (b) principal component analysis (PCA) plot of apple germplasm. In the network plot each dot represents a unique diploid genotype and solid

(continued) lines indicate parent-offspring relationships. Parent-offspring relationships are based on tests for Mendelian consistency (Vanderzande & Howard et al., 2019. Cultivars released from the Swedish breeding programmes in (b) are labelled Alnarp (Aln) and Balsgård (BG). The PCA analysis was performed with PLINK1.9 (Chang et al., 2015), using a subset of SNPs from the 20K SNP array (Howard et al., 2021b) pruned for linkage disequilibrium (LD). The genotypes were coming from the Swedish Central Collection of heirloom cultivars ('Heirloom', red), the collection in Balsgård ('Balsgård', orange), bilateral exchange of data or samples ('Shared', light blue), and the PPP Nordfruit project ('Nordfruit', dark blue). Note that some genotypes might have been represented by synonymous samples from more than one source, but is only represented by a single dot in the figure. The three large full-sibling (FS) families analysed are not included in the figure.

3.2 iGLMap, GDDH13, HFTH1, and 15K-iGW-map

The work in Paper II began by investigating inconsistencies reported between the GDDH13v1.1 WGS and the iGLMap (Peace et al., 2019). However, shortly after the study began, the HFTH1 WGS was published (Zhang et al., 2019). Access to two independent WGSs allowed a more thorough analysis, and the HFTH1 WGS was found to be of sufficient quality to allow its use as base for a virtual genetic map, the iGW-map. Several contigs were found to be misassembled, misplaced, or misoriented on the GDDH13v1.1 WGS (Figure 2a), and accurate information about the composition, location, and order of these will be valuable for future work. Similarly, some of the haploblocks (HBs) of the iGLMap were found to be misplaced within their genetic bins or incorrectly composed. However, only one HBs was erroneously mapped, despite the large proportion of markers mapped, indicating the high quality of the iGLMap. This is particularly interesting since in construction of the iGLMap, special attention was paid to mapping loci that would normally be considered to exhibit segregation distortion, which are commonly discarded during linkage mapping. Segregation distortion can be caused by incorrect marker calls if there are e.g. secondary polymorphisms segregating for the marker or multiple loci affecting the marker calls. On the other hand, such loci can be highly informative if called correctly. During construction of the iGLMap, a forerunner of ASSIsT (Di Guardo et al., 2015) was used to call some types of null-allele segregation, and efforts were made to call at least some of the

markers segregating for more than one loci. Having confirmed the high quality of the iGLMap and the HFTH1 WGS, a polynomial regression of the genetic iGLMap positions on the physical HFTH1 positions for each linkage group (LG)/chromosome, using the 15K SNPs from the 20K SNP array that had been mapped to the iGLMap. This resulted in identification of a series of polynomial coefficients for each LG/chromosome that can be used to interpolate genetic positions from HFTH1 positions, combining the robustness of the iGLMap and accounting for variations in recombination rate across the genome with the high resolution of a WGS. The map generated by interpolating genetic position for the 15K SNPs mapped to the iGLMap was given the name '15K-iGW-map' (Figure 2b).

Analysis of the causes of the observed map errors provided new empirical knowledge on the reasons for these and highlighted some of the pitfalls in identifying genetic and WGS positions for a SNP probe. This is valuable information for all fields of research relying on genetic maps, sequence-based maps, or both.

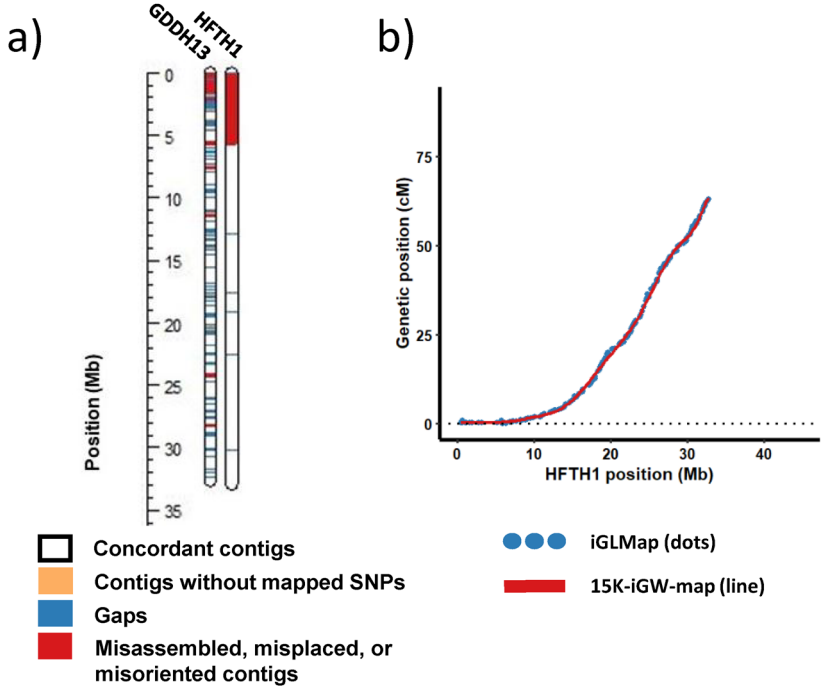


Figure 2. (a) Quality of chromosome 1 in the GDDH13v1.1 and HFTH1 whole genome sequences (WGSs) and (b) collinearity between linkage group (LG) 1 of the iGLMap and the 15K-iGW-map and chromosome 1 in the HFTH1 WGS. Adopted from parts of figures in Paper II.

3.2.1 Interlude

The reliability of genetic analyses using molecular marker data is highly dependent on the quality of the marker data used. A workflow for curation of SNP data based on pedigree information in outcrossing species, which accounts for Mendelian consistent and inconsistent errors, has been proposed by Vanderzande & Howard et al. (2019). Two of the main causes of such errors are problematic clustering and errors in the map positions used. Due to the extensive pedigree connections between the accessions genotyped in this thesis (Figure 1a), curation with the workflow described by Vanderzande & Howard et al. (2019) was possible for a large proportion of the genotyped germplasm. Subsequent work by Howard et al. (2021b) led to the identification of a subset of SNPs from the 20K SNP array with compatible calls between the Illumina Infinium® 20 K and Affymetrix Axiom® 480 K

SNP arrays available for apple (Bianco et al., 2016; Bianco et al., 2014). It was possible to limit the curation in Paper III and V to this set of robust SNPs, and to use curated cluster definitions kindly made available by Howard et al. (2021b). Similarly, the 15K-iGW-map provided a reliable marker order with high resolution. Taken together, these resources reduced the amount of work spent on marker curation during analysis for subsequent papers in this thesis, and are likely to do so also for some of the future work suggested in Chapter 5.

3.3 Resistance to *N. ditissima*

Quantitative trait loci (QTL) mapping of resistance to European canker was performed in Paper III, while Paper IV presents initial results from partial analysis of lipidomic and metabolomic data from a set of reference cultivars during infection by *N. ditissima*. Both studies used the same inoculation procedure (Figure 3a), as previously described in detail by Garkava-Gustavsson et al. (2016). For QTL mapping (Paper III) an FS family of 172 offspring of 'Aroma' x 'Discovery' (res. and suc., respectively, Figure 3b) was used, and the lesion length was assessed at seven time-points from approximately 42 to 126 days post inoculation (DPI). For the metabolomics study (Paper IV), samples were taken at 5, 15, 30 and 45 DPI. Thus, the first assessment in the QTL mapping study corresponded roughly to the last sampled time point in the metabolomic study. This allowed the results of the two studies to be integrated, although they derived from separate experiments. In the QTL mapping study, the assessment data were analysed in terms of a sigmoid growth model for lesion length and the area under the disease progression curve (AUDPC). Thus, Bayesian QTL mapping was performed using five parameters (Paper III). Five QTL with strong or decisive evidence were identified, two of which were associated with the LL_A1 parameter, i.e. the lesion length at around 42 DPI (Figure 4a and 4b). These two QTL were found on LG15 and LG16, the latter being designated as LG16b in Paper III. Data from previous cultivar screening experiments were also compiled, leading to identification of some cultivars with high levels of resistance with good potential for future breeding and research, some of which are known to be well-adapted to cultivation in northern Sweden.

The metabolomics analyses in Paper IV revealed clear patterns in the changes in metabolomic profiles during infection, particularly during the later stages of infection. Notably, 'Aroma' and 'Discovery' showed different patterns of accumulation of the suberin component suberic acid, with the susceptible 'Discovery' having a much lower concentration of this metabolite at 45 DPI (Figure 4c). The formation of a mature suberin layer enclosing the infection is known to be a key component of a successful resistance reaction in the *M. domestica – N. ditissima* pathosystem (Krähmer, 1980).

A predicted gene with similarity to a peroxidase (*TPX1* of *Arabidopsis thaliana*) is indicated in the GDDH13 genomic region corresponding to the LG15 QTL interval (Genome database for Rosaceae, GDR; Jung et al., 2018). Similarly, there is a predicted gene with similarity to a laccase (*LAC7* of *A. thaliana*) in the GDDH13 genomic region corresponding to the LG16b QTL interval. It has recently been suggested that peroxidases and laccases play an important role in suberin formation in hybrid poplar (*Populus tremula* x *Populus alba*) (Rains et al., 2018). Peroxidases have also been found to be involved in suberin synthesis in tomato (Quiroga et al., 2000) and laccases have been shown to be associated with response to biotic stresses in *e.g.* apple (Zhu et al., 2021) and cotton (Hu et al., 2018). While this provides further indication of the different roles of the identified resistance QTL during infection, key metabolites annotations should be verified and additional studies involving other omics-techniques are needed to shed further light on the resistance mechanisms in the *M. domestica – N. ditissima* pathosystem.

Interestingly, the resistant cultivar 'Santana' was not identical by state (IBS) for the favourable alleles at any of the QTL segregating in 'Aroma' x 'Discovery'. 'Santana' also showed a pattern for some metabolites that deviated from 'Aroma', for example for the Jasmonyl-Isoleucines (Figure 4d). The Jasmonyl-Isoleucines include the bioactive form of jasmonic acid, a phytohormone traditionally associated with response to necrotrophic fungi (Antico et al., 2012). Thus, 'Santana' is likely to provide a complementary source of resistance to *N. ditissima* that could be exploited in future breeding and genetic research.

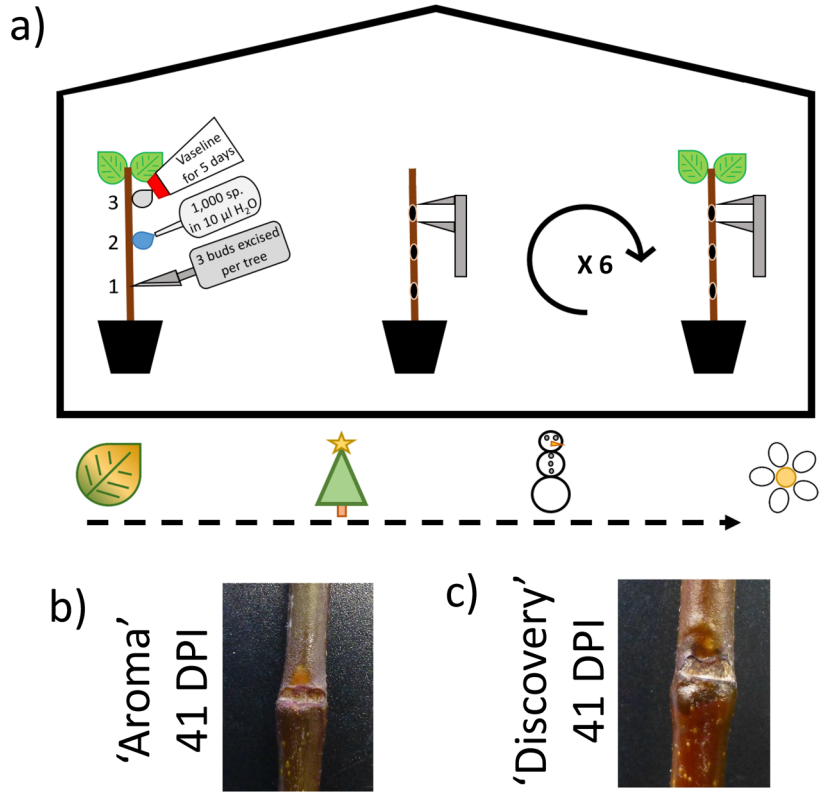


Figure 3. (a) Graphical summary of experimental protocol used in Paper III and IV whereby one year trees are inoculated in a glasshouse in late autumn. Briefly, three buds per tree are removed by a scalpel and 1,000 condia in 10 µl of water are then added to each wound. The wounds are then covered with petrolatum, which is removed after five days. The length of each wound is measured with a digital calliper, starting approximately 42 days post inoculation (DPI). Assessments are then repeated six more times, every second week, until the end of the experiment. (b, c) Examples of European canker symptoms on 'Aroma' and 'Discovery' 41 DPI, from a different experiment than those reported in Paper III and IV, but using the same protocol.

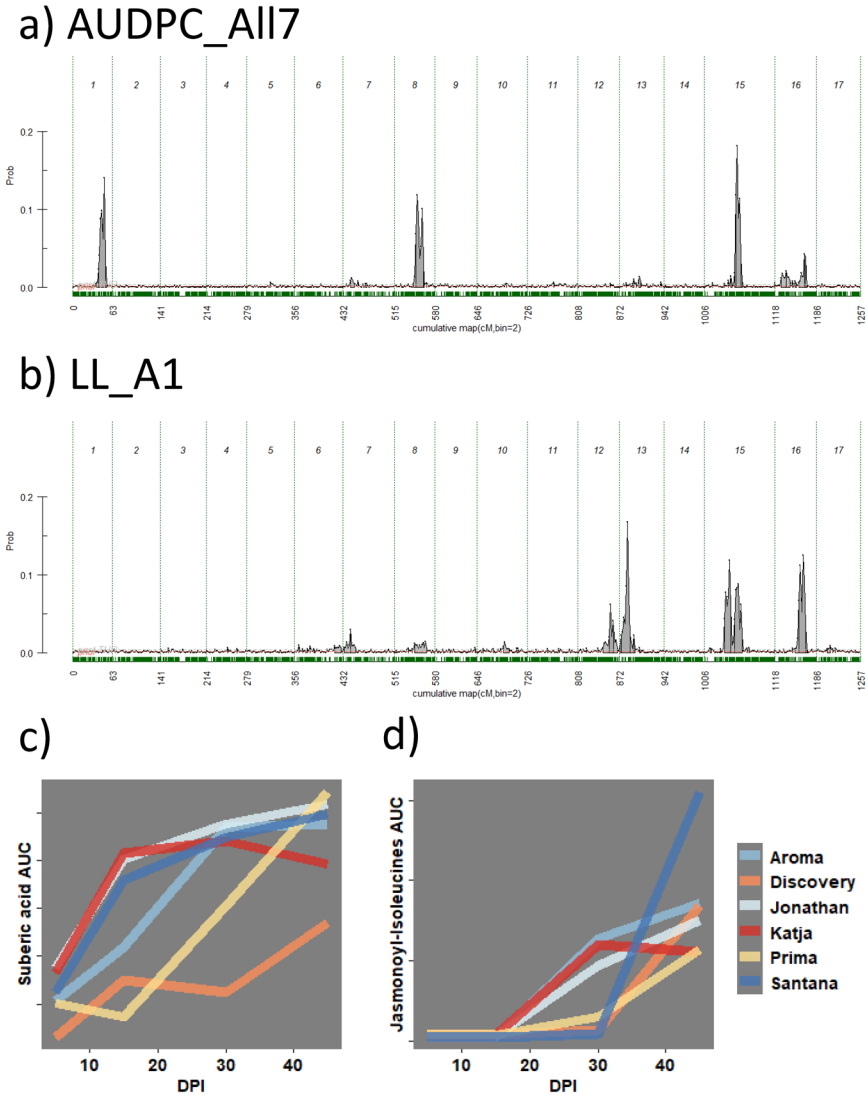


Figure 4. (a) Posterior QTL positions with the AUDPC_All7 parameter, (b) posterior QTL positions with the LL_A1 parameter, and average area under the curve (AUC) of (c) suberic acid and (d) Jasmonyl-Isoleucines metabolites over time in infected samples. Note that for plots (a) and (b), statistical evidence for a QTL is based on area under the peak. Metabolite data series are coloured by cultivar, from most susceptible ('Katja', dark red) to most resistant ('Santana', dark blue). (a) and (b) comes from the data analysis of Paper III, but from another Bayesian replicate than shown in the manuscript. (c) and (d) are modified from Paper IV.

3.4 Validation of genomic regions for phenological traits

Paper V investigated whether some previously published genomic regions and QTL intervals for flowering and harvest date could be validated in germplasm relevant for Nordic apple breeding under Nordic conditions. Phenotypic data were available from the PPP Nordfruit project, and associations with haplotypes in the germplasm were investigated. Post-hoc analyses were also performed to identify haplotypes associated with increasing and decreasing alleles at the validated loci.

Concerning flowering date, no association was found on the top of chromosome 9 above the threshold, despite this region showing the highest statistical power in previous studies (Allard et al., 2016; Urrestarazu et al., 2017). This is in line with Urrestarazu et al. (2017) who found no significant association for the part of the germplasm from SLU used in that study, although the work in Paper V was based on a different germplasm and compilation of phenotypic data from three locations across the Nordic countries. Cornelissen et al. (2020) mapped a QTL interval for chilling requirement for dormancy release to the top of LG9. If this was the same QTL as reported for flowering date, one would not expect it to have a large effect under Nordic conditions, as exposure to chilling is unlikely to be a limiting factor. The remaining genomic regions considered for validation had small effects, and several of the latest flowering cultivars were generally considered not to be suitable for cultivation under Nordic conditions (Paper V). Thus, a more promising approach to managing the risks of frost during flowering might be to breed for tolerance to frost during flowering, rather than avoidance. Previous work has identified locally adapted cultivars and advanced selections that could be used for further breeding and research on tolerance to frost during flowering (Nybom, 1992).

For harvest date, a significant association was found in the genomic region surrounding *NAC18.1*, which has previously been reported to control fruit ripening in apple (Migicovsky et al., 2021). Haplotypes associated with late and early harvest dates, respectively, were also identified, as were haplotype profiles from a reduced number of SNPs. Using the reduced SNP haplotypes, individuals that were heterozygous had on average a harvest date that was 26 days earlier than that of individuals homozygous for either of the late haplotypes (Figure 5a).

Regarding deployment in Nordic breeding programmes, care should be taken with parents that are homozygous for the late haplotypes. In particular,

special awareness is required for crosses where both parents are homozygous for the late haplotypes, as the offspring is expected to have a lower probability of being able to reach harvest maturity under Nordic conditions. This is particularly important when breeding for central and northern Sweden, which have a much shorter growing season than e.g. the Scania region in the south. Among the individuals that were found to be homozygous for the late alleles of the reduced SNP haplotypes, there were some cultivars of relevance for the Swedish breeding programme, including 'Dayton', 'Elise', 'Golden Delicious', 'Goldrush', 'John-Georg', 'Rubinola', and 'Scarlet O'Hara'. It should be noted that *NAC18.1* has also been reported to be associated fruit firmness, with the alleles for late harvest being associated with increased fruit firmness (Migicovsky et al., 2021, 2016). Thus, considering breeding for commercial apple production under short growing seasons the capacity of other firmness loci to counter the undesirable effect of the early allele at *NAC18.1* has to be considered.

3.5 Autumn senescence

Paper VI presents the first results from a study on autumn senescence in apple and its implication for adaptation to northern Sweden. The first steps in that study were to phenotype the Swedish Central Collection of mandate cultivars in Alnarp for canopy senescence over three years and calculate the average date of 50% senescent canopy. The results showed that the collection contains considerable variation for the trait, with the first and last genotypes to reach 50% senescent canopy being separated by 70 days on average (Figure 5b). The average date of 50% senescence was significantly negatively correlated with the highest recommended growing zone stated in pomological literature (Näslund, 2010; Nilsson, 1987; Svensson, 2005), when evaluated by both Pearson's correlation (p-value 0.01) and by Stuart's tau-c (99% confidence interval -0.34 to -0.01). The cultivars recommended for cultivation only in zone 1 reached 50% senescence on average two weeks later than those recommended for cultivation in zone 6 (Figure 5c, unpublished work). Previous studies on aspen have shown that genotypes collected from northern (Latitude 66°) and southern (Latitude 56.2°) parts of Sweden differ in date of onset of senescence by approximately one week when grown in Scania (Michelson et al., 2018). Thus, while the research on the genetic basis for adaptation to northern Sweden in apple is in its infancy,

canopy senescence appears to be an interesting candidate trait for further studies.

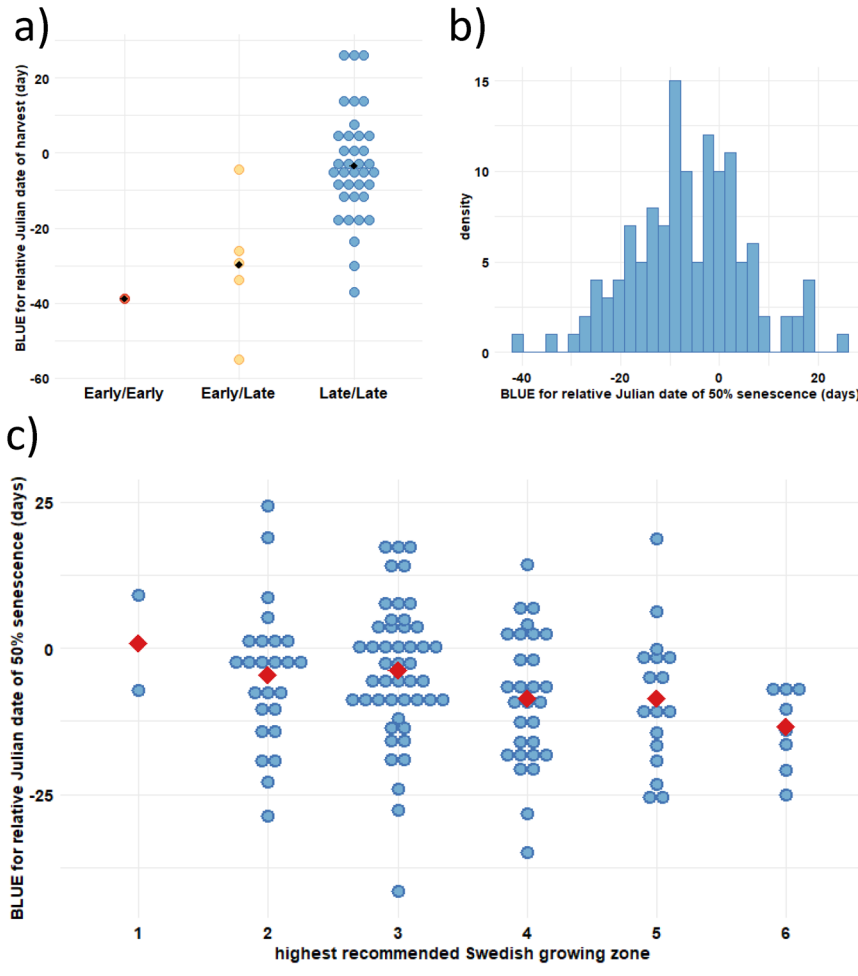


Figure 5. Distributions of Best Linear Unbiased Estimates (BLUEs) for (a) harvest date in relation to the reduced single nucleotide polymorphism (SNP) haplotypes, showing only individuals with effects assigned to both haplotypes, (b) distribution of BLUEs for date of 50% senescence in the Swedish Central Collection of heirloom cultivars, and (c) date of 50% senescence plotted against the highest recommended growing zone stated in pomological literature (Näslund, 2010; Nilsson, 1987; Svensson, 2005). Average values for each of the groups are indicated by black (a) or red (c) diamonds. (a) and (b) are modified from Paper V and IV, respectively.

3.6 Tracing haplotypes through pedigrees

Curated genotypic data in a pedigreed germplasm were used to trace the inheritance of desirable and undesirable alleles for identified genomic regions (Figure 6). Considering the QTL regions segregating for resistance to *N. ditissima* in 'Aroma' x 'Discovery', the only cultivar in the germplasm other than 'Aroma' appearing to have inherited the desirable allele from 'Filippa' at the LG15 locus was 'Gratia'. On the other hand, the desirable allele from 'Beauty of Bath' ("BeutBath" in Figure 6) seems to have been transmitted only to the susceptible cultivar 'Julia', which is likely not well suited for future breeding towards improved resistance to *N. ditissima*. Notably, 'Ingrid Marie' ("IngMarie" in Figure 6) appears to have three recombination events on the copy of LG15 coming from 'Cox's Orange Pippin' ("Cox" in Figure 6), potentially spanning a region sufficiently large to make it plausible. One of these recombination events might be within the consensus QTL region for the LG15 locus, although most of the consensus QTL region from 'Cox's Orange Pippin' in 'Ingrid Marie' seems to constitute the haplotype transmitted to 'James Grieve' ("JamesGr" in Figure 6) and 'Katja'. Note that 'Lord Lambourne' ("Lambourne" in Figure 6) seems to have inherited the undesirable allele from 'Worcester Pearmain' ("WorcPearm" in Figure 6) and one of the haplotypes from 'Cox's Orange Pippin' through 'James Grieve' (Figure 6a).

The LG16b locus was found to exhibit monohybrid segregation in 'Aroma' x 'Discovery', with the desirable allele coming from 'Worcester Pearmain'. While the susceptible cultivar 'Julia' seems to be the only individual in the current germplasm to have inherited the undesirable allele from 'Beauty of Bath', several individuals seem to have inherited the desirable allele from 'Worcester Pearmain' (Figure 6b). One of these being the advanced selection denoted 'Lillarosaröda' ("Lillarosarod" in Figure 6), which also seems to have inherited the desirable allele from 'Worcester Pearmain' at the LG8 locus (not shown). However, as the locus at LG16 exhibits monohybrid segregation in 'Aroma' x 'Discovery' dominance effects could not be assessed, and the potential presence of such effects should be taken into consideration when designing crosses based on this locus. Thus, considering the moderate estimated effect of this locus and potentially recessive resistance, the most robust interpretation would be to avoid using 'Beauty of Bath', 'Discovery', or 'Julia' as parents for further breeding.

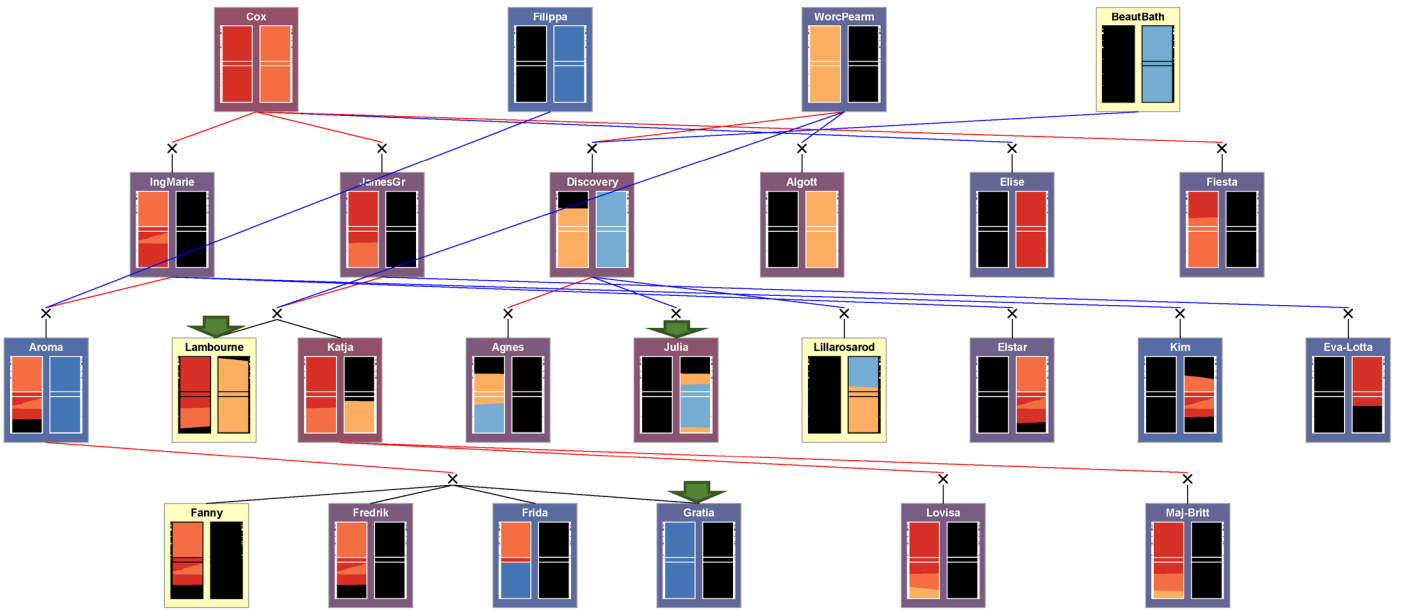
Considering the genomic region associated with harvest date of chromosome 3, the haplotypes of founder individuals associated with early and late maturity were similarly traced through the germplasm. The results showed that the cultivar 'Melba' appears to be heterozygous for haplotypes associated with early and late fruit maturity and has two half-siblings and several offspring. Of its offspring, 'Norland', 'Pirja', 'Silva', and the selection BM51880 from the Swedish breeding program seem to have inherited the haplotype associated with early maturity from 'Melba'. The cultivars 'Norland' and 'Parkland' are full siblings descending from 'Melba' and 'Rescue', and both have been reported to mature very early (Nybom, 2017). While 'Norland' seems to have inherited the early allele from 'Melba', 'Parkland' seems to have inherited the allele associated with late maturity. However, 'Rescue' is considered very winter hardy and is recommended for cultivation in the Swedish zone 6, where the growing season is very short. Thus, the rare haplotype of 'Rescue' might confer early maturity, and 'Rescue' might also carry genetic variation conferring early maturity at other loci. Among the other descendants of 'Melba' in the germplasm, 'Britemac', 'Maikki', 'Samo', and 'Vuokko' seem to have inherited the haplotype associated with late maturity coming from 'McIntosh'. Note that 'Britemac' appears to be homozygous IBD for a part of LG3 coming from 'McIntosh' as it is the offspring of 'Melba' x 'Cortland', which are half-siblings.

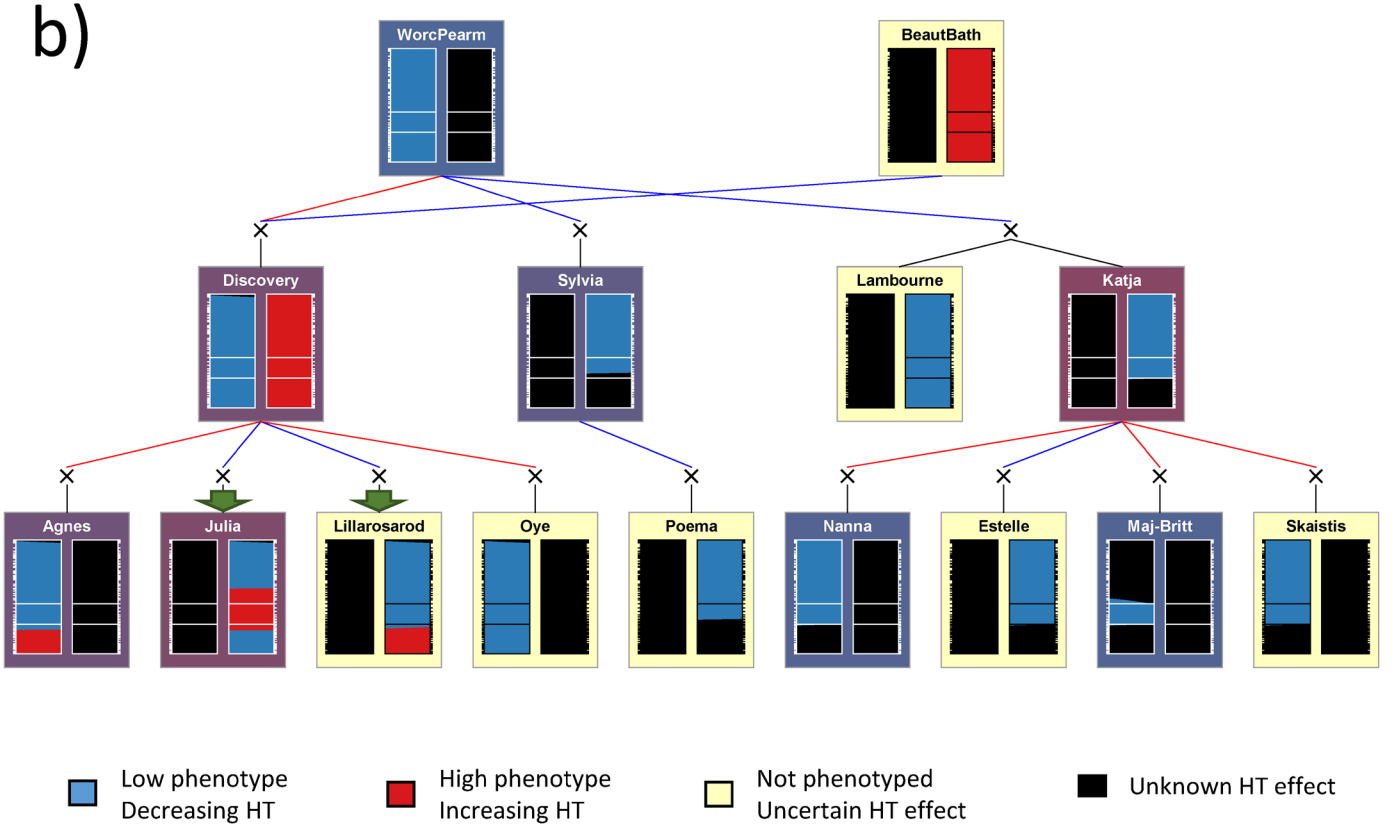
Thus, for future breeding one might consider e.g. crossing 'Lillarosaröda' with 'Gratia' to combine the desirable alleles for resistance to *N. ditissima* at LG8, LG15 and LG16. However, based on the performance of the ancestors, the offspring of such cross might not be suitable for cultivation in northern Sweden. 'Lillarosaröda' has one rare allele and one allele with ambiguous effect at LG3, while 'Gratia' has one rare allele and one allele associated with late maturity. Thus, while the offspring of such a cross might be hardy enough for cultivation in e.g. central Sweden, the fruit might not mature well at those latitudes.

In breeding for central or northern Sweden, one might instead consider crossing 'Santana' or 'Gratia' with one of the hardy cultivars identified as being resistant in the *N. ditissima* screening experiments, e.g. 'Rödluvan' ("Rodluvan" in Figure 6), 'Rescue', and 'Snövit' ("Snovit" in Figure 6). The cultivar 'Rödluvan' was released from the Swedish apple breeding programme at Balsgård in 1994 and is recommended for cultivation in

northern Sweden (zone 6) (Svensson, 2005), but not for commercial production. It is IBS to one of the desirable alleles segregating for resistance to *N. ditissima* in 'Aroma' x 'Discovery' and has one rare haplotype and one haplotype with ambiguous effect at LG3, although it reaches maturity at approximately the same time as 'Discovery' (not shown). Considering the risk of inbreeding, it should be noted that 'Lobo', one of the parents of 'Rödluvan', is also a grandparent of 'Santana' (Figure 6c). The cultivar 'Rescue' is considered to be very winter hardy and its offspring 'Norland' and 'Parkland' could also be considered for future crosses, although they have not been evaluated for resistance to *N. ditissima* at SLU. As both 'Norland' and 'Parkland' are very early maturing (Nybom, 2017), one might consider combining them with 'Santana', which is homozygous for haplotypes associated with late maturity at LG3. The cultivar 'Snövit' was released from the breeding programme in Alnarp in 1955 (Nilsson, 1987). It is recommended for cultivation in northern Sweden, has inherited the haplotype for early maturity from 'Stenbock' and reaches maturity approximately five days earlier than 'Discovery' (not shown). It should be noted that heirloom cultivars may express other traits that are undesirable under commercial production, e.g. the tendency to flower and bear fruit on the terminal buds of vegetative shoots rather than on spurs. Thus, additional work on these aspects might be needed to allow efficient exploitation of these resources.

a)


Cox
Filippa
WorcPearm
BeautBath
IngMarie
JamesGr
Discovery
Algott
Elise
Fiesta
Aroma
Lambourne
Katja
Agnes
Julia
Lillarosarod
Elstar
Kim
Eva-Lotta
Fanny
Fredrik
Frida
Gratia
Lovisa
Maj-Britt


b)
WorcPearm
BeautBath
Discovery
Sylvia
Lambourne
Katja
Agnes
Julia
Lillarosarod
Oye
Poema
Nanna
Estelle
Maj-Britt
Skaistis
Low phenotype
Decreasing HT
High phenotype
Increasing HT
Not phenotyped
Uncertain HT effect
Unknown HT effect

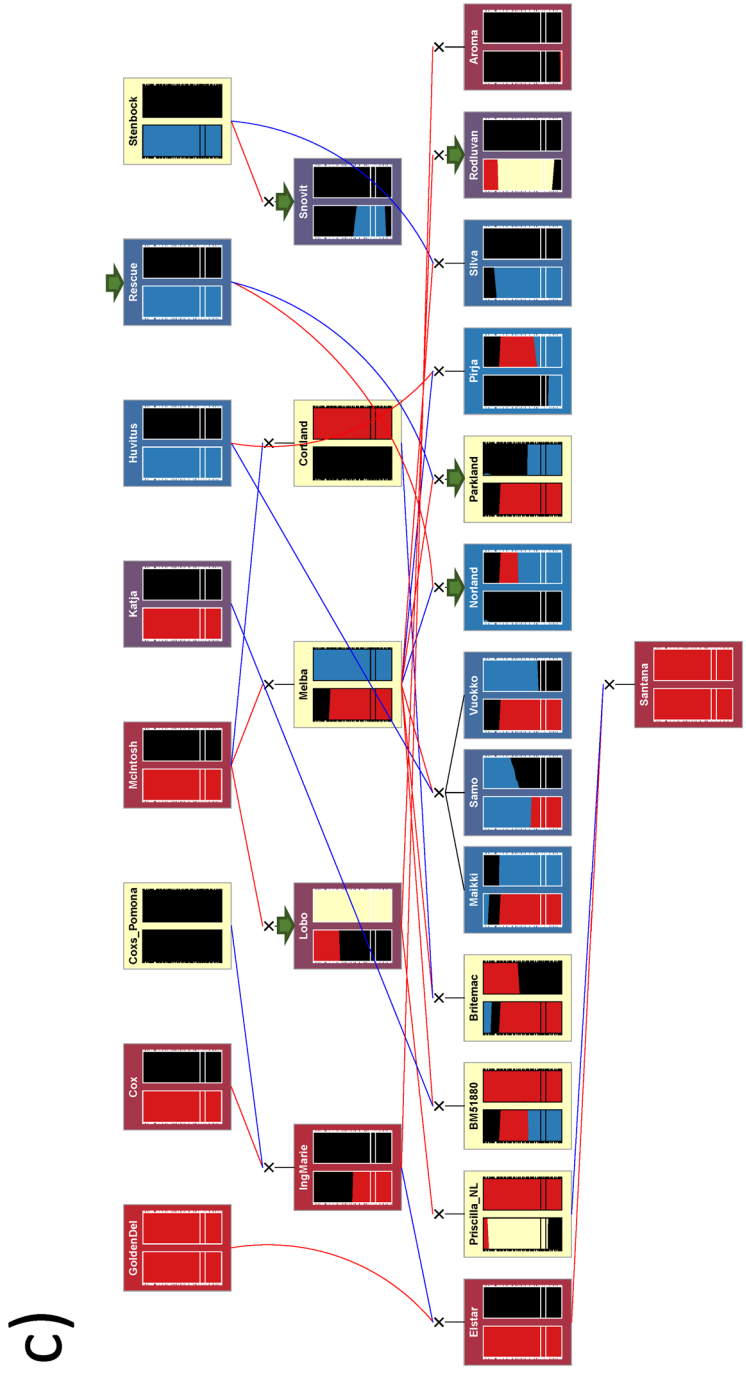

c)
Stenbock
Rescue
Huvitus
Katja
McIntosh
Coxs_Pomona
Cox
GoldenDel
Snovit
Cortland
Melba
Lobo
IngMarie
Aroma
Rodiuvan
Silva
Pirja
Parkland
Norland
Viokko
Samo
Maikki
Britemac
BM51880
Priscilla_NL
Elstar
Santana

Figure 6. Transmission of founder haplotypes for (a) resistance to *Neonectria ditissima* at linkage group (LG) 15, (b) resistance to *N. ditissima* at LG16 (the LG16b locus in Paper III), and (c) harvest date at LG3. Each box represents an individual with its two homologs for the respective LG. Horizontal lines in the homologs indicate boundaries of the consensus quantitative trait locus (QTL) intervals for loci for resistance to *N. ditissima*, and boundaries of the genomic region associated with harvest date. The boxes are coloured by scale according to their phenotypes for the two traits, from red (susceptible, late harvest) to blue (resistant, early harvest) or beige (no data). The homologs are similarly coloured by the effect of the allele within the boundaries, with haplotypes with unknown effect in black and haplotypes in (c) with ambiguous effect in beige. The lines connecting the individuals indicate maternal inheritance (red lines) and paternal inheritance (blue lines) when known from records; otherwise assignment is arbitrary as no cytoplasmic markers were used. Some individuals discussed in the text are indicated by green arrows. Note that the pedigree in (c) has been cut for space reasons, thereby assigning 'Katja' as a founder, though it is a descendant of 'Cox's Orange Pippin' through 'James Grieve' as can be seen in (a). The homolog of 'Katja' associated with late harvest that is transmitted intact to BM51880 seems not to be from 'Cox's Orange Pippin'. (a) and (b) are modified from Paper III.

4. Implications for the Swedish apple breeding programme

The long generation time and expensive phenotyping required in apple breeding, together with the limited resources of the Swedish breeding programme, pose great challenges for the Swedish apple breeding programme. Under such circumstances, efficient selection of parents for crosses is extremely important, as it takes several years until the offspring can be evaluated for most key traits. Thus, the deployment of marker-assisted parent selection (MAPS) allows for a more informed choice of parent combinations, improving the performance of the entire seedling population and increasing the possibility of finding a desirable individual. Greater genetic gain can be achieved from the same number of crosses, or the same genetic gain can be obtained from a reduced number of crosses. Given the weak economic driver for a higher rate of cultivars released to the Swedish market, the latter might be a preferable approach in the Swedish apple breeding programme. This could allow resources to be allocated to e.g. data collection and small-scale recurrent speed breeding programmes while maintaining a steady rate at which new cultivars are released to the market.

The work in this thesis served to establish a base for MAPS using phased SNP array data in a pedigreed germplasm, exemplified by the two key traits resistance to *N. ditissima* and fruit maturity date, as discussed in previous chapters. The results presented in this thesis could also improve the efficiency of the breeding programme further, as resources could be saved by reducing the size of the breeding germplasm maintained. Examples of accessions that could be culled from a slimmed breeding germplasm include genetic duplicates, accessions that are not TTT, and triploids.

5. Conclusions and future prospects

The work presented in this thesis lays the foundations for genome-informed breeding in the Swedish apple programme. Relating to the aims of the thesis there are some key conclusions:

- Some of the available germplasm was curated and a number of synonymous accessions and some accessions which were likely NTTT were identified. Five key founders among the Swedish heirloom cultivars were identified: 'Gimmersta', 'Grågylling', 'Vitgylling', 'Rosenhäger' and 'Klockhammarsäpple'. This also led to further work to locate and preserve other heirloom cultivars, which otherwise risked being lost.
- A virtual linkage map, the iGW-map, was developed based on the analysis of two independent WGSs and an existing high-density linkage map. A version of this map, the 15K-iGW-map, served as the base for Paper III and V in this thesis. The analysis also provided empirical information on causes for the reported inconsistencies between the GDDH13v1.1 WGS and the iGLMap, and pinpointed the occurrence of the different types of errors on the three maps.
- Five QTL for resistance to *Neonectria ditissima* were mapped in segregating offspring of 'Aroma' x 'Discovery', both parents being highly relevant to Swedish apple breeding. A preliminary analysis of metabolomic profiles of some apple cultivars during infection of *N. ditissima* indicated a potential role for two of the mapped QTL. Additionally, data from previous screening experiments were compiled, which led to the identification of some relatively resistant cultivars of relevance for further breeding.
- Haplotypes on LG3 associated with early and late harvest dates under Nordic conditions were identified in germplasm relevant for Nordic apple breeding. The locus for flowering date identified on LG9 in previous studies did not appear to have a significant effect under Nordic conditions.
- Preliminary studies exploring the genomic basis for adaptation to northern Sweden were performed. These showed that date of 50% canopy senescence was correlated with adaptation to northern

Sweden, and might be a relevant trait for further breeding and genetic research on cultivars adapted to such climate.

5.1.1 Future research

However, these novel findings are only the starting point. In order to transform the Swedish breeding programme so that it can tackle current and future challenges efficiently, further work is essential. Key targets for future work in relation to the Swedish breeding programme could be to:

- Identify complementary QTL for resistance to *N. ditissima* segregating in 'Santana'.
- Validate QTL for resistance to *N. ditissima*, using complementary experimental procedures and complementary genetic material, possibly allowing for estimation of potential interaction effects.
- Analyse Metabolomic, Proteomic, and Transcriptomic profiles of key founder cultivars towards an improved understanding of the molecular basis for differences in susceptibility to *N. ditissima*.
- Validate reported QTL regions with large or intermediate effects for important fruit quality traits, such as firmness, acidity, and sweetness, in particular in relation to alleles for harvest date at the LG3 locus.
- Establish phenotyping methods for tolerance to freezing during flowering, and identify tolerant material for further breeding and genetic research.
- Perform genetic mapping of QTL intervals for autumn senescence and undesirable traits in heirloom cultivars, e.g. by PBA.
- Identify candidate genomic regions underlying adaptation to northern Sweden using population genomic approaches, and establish germplasm collections for phenotyping in central and northern Sweden, which is crucial to advance this line of research.
- Validate markers for deployment in marker-assisted seedling selection (MASS) for key traits such as acidity, sweetness, and firmness, to allow efficient culling of the least promising seedlings during early stages of selection.

References

Allard, A., Bink, M.C., Martinez, S., Kelner, J.J., Legave, J.M., di Guardo, M., Di Pierro, E.A., Laurens, F., van de Weg, E.W., Costes, E. (2016) Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. *J Exp Bot* 67, 2875–2888. https://doi.org/10.1093/jxb/erw130

Antico, C.J., Colon, C., Banks, T., Ramonell, K.M. (2012) Insights into the role of jasmonic acid-mediated defenses against necrotrophic and biotrophic fungal pathogens. Front. Biol. 7, 48–56. https://doi.org/10.1007/s11515-011-1171-1

Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Théron, A., Poncet, C., Micheletti, D., Kerschbamer, E., Di Pierro, E.A., Larger, S., Pindo, M., Van De Weg, E., Davassi, A., Laurens, F., Velasco, R., Durel, C.E., Troggio, M. (2016) Development and validation of the Axiom® Apple 480K SNP genotyping array. *Plant J*. 86, 62–74. https://doi.org/10.1111/tpj.13145

Bianco, L., Cestaro, A., Sargent, D.J., Banchi, E., Derdak, S., Di Guardo, M., Salvi, S., Jansen, J., Viola, R., Gut, I., Laurens, F., Chagné, D., Velasco, R., van de Weg, E., Troggio, M. (2014) Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLoS One* 9, e110377. https://doi.org/10.1371/journal.pone.0110377

Bink, M.C., Jansen, J., Madduri, M., Voorrips, R.E., Durel, C.E., Kouassi, A.B., Laurens, F., Mathis, F., Gessler, C., Gobbin, D., Rezzonico, F., Patocchi, A., Kellerhals, M., Boudichevskaia, A., Dunemann, F., Peil, A., Nowicka, A., Lata, B., Stankiewicz-Kosyl, M., Jeziorek, K., Pitera, E., Soska, A., Tomala, K., Evans, K.M., Fernandez-Fernandez, F., Guerra, W., Korbin, M., Keller, S., Lewandowski, M., Plocharski, W., Rutkowski, K., Zurawicz, E., Costa, F., Sansavini, S., Tartarini, S., Komjanc, M., Mott, D., Antofie, A., Lateur, M., Rondia, A., Gianfranceschi, L., van de Weg, W.E. (2014) Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor Appl Genet* 127, 1073–1090. https://doi.org/10.1007/s00122-014-2281-3

Bink, M.C.A.M., Boer, M.P., Ter Braak, C.J.F., Jansen, J., Voorrips, R.E., van de Weg, W.E. (2008) Bayesian analysis of complex traits in pedigreed plant populations. *Euphytica* 161, 85–96. https://doi.org/10.1007/s10681-007-9516-1

Bus, V., Single, G., Horner, M., Jesson, L., Walter, M., Kitson, B., López-Girona, E., Kirk, C., Gardiner, S., Chagné, D., Volz, R. (2021) Preliminary genetic mapping of fire blight and European canker resistances in two apple breeding families. *Acta Hortic.* 1307, 199–203. https://doi.org/DOI

10.17660/ActaHortic.2021.1307.31
Bus, V.G., Rikkerink, E.H., Caffier, V., Durel, C.E., Plummer, K.M. (2011) Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*. *Annu Rev Phytopathol* 49, 391–413. https://doi.org/10.1146/annurev-phyto-072910-095339
Bus, V.G.M., Scheper, R.W.A., Walter, M., Campbell, R.E., Kitson, B., Turner, L., Fisher, B.M., Johnston, S.L., Wu, C., Deng, C.H., Singla, G., Bowatte, D., Jesson, L.K., Hedderley, D.I., Volz, R.K., Chagné, D., Gardiner, S.E. (2019) Genetic mapping of the European canker (*Neonectria ditissima*) resistance locus Rnd1 from *Malus* 'Robusta 5.' *Tree Genet. Genomes* 15, 1–13. https://doi.org/10.1007/s11295-019-1332-y
Chagné, D., Crowhurst, R.N., Troggio, M., Davey, M.W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R.P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J.D., Iezzoni, A., Mockler, T., Wilhelm, L., Weg, E. Van de, Gardiner, S.E., Bassil, N., Peace, C., Van de Weg, E., Gardiner, S.E., Bassil, N., Peace, C. (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS One* 7, e31745.
Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. https://doi.org/10.1186/s13742-015-0047-8
Cornelissen, S., Hefer, C.A., Rees, D.J.G., Burger, J.T. (2020) Defining the QTL associated with chill requirement during endodormancy in *Malus* × *domestica Borkh*. *Euphytica* 216, 1–12. https://doi.org/10.1007/s10681-020-02645-3
Cornille, A., Giraud, T., Smukders, M., Roldán-Ruiz, I., Gladieux, P. (2014) The domestication and evolutionary ecology of apples. *Trends Genet.* 30, 57–65. https://doi.org/http://dx.doi.org/10.1016/j.tig.2013.10.002
Daccord, N., Celton, J.M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E.A., Gouzy, J., Rees, D.J.G., Guerif, P., Muranty, H., Durel, C.E., Laurens, F., Lespinasse, Y., Gaillard, S., Aubourg, S., Quesneville, H., Weigel, D., van de Weg, E., Troggio, M., Bucher, E. (2017) High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet* 49, 1099–1106. https://doi.org/10.1038/ng.3886
Di Guardo, M., Micheletti, D., Bianco, L., Koehorst-van Putten, H.J., Longhi, S., Costa, F., Aranzana, M.J., Velasco, R., Arus, P., Troggio, M., van de Weg, E.W. (2015) ASSIsT: an automatic SNP scoring tool for in- and outbreeding species. *Bioinformatics* 31, 3873–3874. https://doi.org/10.1093/bioinformatics/btv446
Di Pierro, E.A., Gianfranceschi, L., Di Guardo, M., Koehorst-Van Putten, H.J., Kruisselbrink, J.W., Longhi, S., Troggio, M., Bianco, L., Muranty, H., Pagliarani, G., Tartarini, S., Letschka, T., Lozano Luis, L., Garkava-Gustavsson, L., Micheletti, D., Bink, M.C., Voorrips, R.E., Aziz, E., Velasco, R., Laurens, F., Van De Weg, W.E. (2016) A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing

species. *Hortic. Res.* 3, 16057. https://doi.org/10.1038/hortres.2016.57
Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X., Jiao, C., Legall, N., Mao, L., Wan, S., Wang, K., He, T., Feng, S., Zhang, Z., Mao, Z., Shen, X., Chen, Xiaoliu, Jiang, Y., Wu, S., Yin, C., Ge, S., Yang, L., Jiang, S., Xu, H., Liu, J., Wang, D., Qu, C., Wang, Y., Zuo, W., Xiang, L., Liu, C., Zhang, D., Gao, Y., Xu, Y., Xu, K., Chao, T., Fazio, G., Shu, H., Zhong, G.Y., Cheng, L., Fei, Z., Chen, Xuesen (2017) Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat. Commun.* 8, 1–11. https://doi.org/10.1038/s41467-017-00336-7
Fernqvist, I. (1993) RST växtatlas: med zonkartor, 3rd ed. Riksförbundet Svensk Trädgård, Täby.
Garkava-Gustavsson, L., Ghasemkhani, M., Zborowska, A., Englund, J.-E., Lateur, M., van de Weg, E. (2016) Approaches for evaluation of resistance to European canker (*Neonectria ditissima*) in apple. *Acta Hortic*. 75–82. https://doi.org/10.17660/ActaHortic.2016.1127.14
Garkava-Gustavsson, L., Zborowska, A., Sehic, J., Rur, M., Nybom, H., Englund, J.E., Lateur, M., van de Weg, E., Holefors, A. (2013) Screening of apple cultivars for resistance to european canker, Neonectria ditissima. *Acta Hortic*. 976, 529–536. https://doi.org/10.17660/ACTAHORTIC.2013.976.75
Gaut, B.S., Díez, C.M., Morrell, P.L. (2015) Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet*. https://doi.org/10.1016/j.tig.2015.10.002
Gustavsson, L., Östberg, J., Skytte af Sätra, J. (2022) Fruktträdskräfta – hotet mot äpplet. *Natur och Trädgård* nr 4.
Hänninen, H. (2006) Climate warming and the risk of frost damage to boreal forest trees: identification of critical ecophysiological traits. *Tree Physio*l. 26, 889–898. https://doi.org/https://doi.org/10.1093/treephys/26.7.889
Howard, N.P., Peace, C., Silverstein, K.A.T., Poets, A., Luby, J.J., Vanderzande, S., Durel, C.-E., Muranty, H., Denancé, C., van de Weg, E. (2021a) The use of shared haplotype length information for pedigree reconstruction in asexually propagated outbreeding crops, demonstrated for apple and sweet cherry. *Hortic. Res*. 2021 81 8, 1–13. https://doi.org/10.1038/s41438-021-00637-5
Howard, N.P., Troggio, M., Durel, C.-E., Muranty, H., Denancé, C., Bianco, L., Tillman, J., van de Weg, E. (2021b) Integration of Infinium and Axiom SNP array data in the outcrossing species *Malus* × *domestica* and causes for seemingly incompatible calls. *BMC Genomics* 22, 1–18. https://doi.org/10.1186/s12864-021-07565-7
Hu, Q., Min, L., Yang, X., Jin, S., Zhang, L., Li, Y., Ma, Y., Qi, X., Li, D., Liu, H., Lindsey, K., Zhu, L., Zhang, X. (2018) Laccase GhLac1 modulates broad-spectrum biotic stress tolerance via manipulating phenylpropanoid pathway and jasmonic acid synthesis. *Plant Physiol*. 176, 1808. https://doi.org/10.1104/PP.17.01628
Janick, J. (2006) The PRI apple breeding program. HortScience 41, 8–9.
Jung, S., Lee, T., Cheng, C.-H., Buble, K., Zheng, P., Yu, J., Humann, J., Ficklin,

S.P., Gasic, K., Scott, K., Frank, M., Ru, S., Hough, H., Evans, K., Peace, C., Olmstead, M., Devetter, L.W., Mcferson, J., Coe, M., Wegrzyn, J.L., Staton, M.E., Abbott, A.G., Main, D. (2018) 15 years of GDR: New data and functionality in the Genome Database for *Rosaceae*. *Nucleic Acids Res*. 47. https://doi.org/10.1093/nar/gky1000

Karlström, A., Gómez-Cortecero, A., Nellist, C.F., Ordidge, M., Dunwell, J.M., Harrison, R.J. (2022) Identification of novel genetic regions associated with resistance to European canker in apple. *BMC Plant Biol*. 2022 221 22, 1–16. https://doi.org/10.1186/S12870-022-03833-0

Khan, A., Carey, S.B., Serrano, A., Zhang, H., Hargarten, H., Hale, H., Harkess, A., Honaas, L., Carey, S.B., Serrano, A., Zhang, H., Hargarten, H., Hale, H., Harkess, A., Honaas, L. (2022) A phased, chromosome-scale genome of 'Honeycrisp' apple (*Malus domestica*). *Gigabyte* 2022, 1–15. https://doi.org/10.46471/GIGABYTE.69

Larsen, B., Migicovsky, Z., Jeppesen, A.A., Gardner, K.M., Toldam-Andersen, T.B., Myles, S., Ørgaard, M., Petersen, M.A., Pedersen, C. (2019) Genome-wide association studies in apple reveal loci for aroma volatiles, sugar composition, and harvest date. *Plant Genome* 12, 180104. https://doi.org/10.3835/plantgenome2018.12.0104

Legave, J.M., Blanke, M., Christen, D., Giovannini, D., Mathieu, V., Oger, R. (2013) A comprehensive overview of the spatial and temporal variability of apple bud dormancy release and blooming phenology in Western Europe. *Int J Biometeorol* 57, 317–331. https://doi.org/10.1007/s00484-012-0551-9

Michelson, I.H., Ingvarsson, P.K., Robinson, K.M., Edlund, E., Eriksson, M.E., Nilsson, O., Jansson, S. (2018) Autumn senescence in aspen is not triggered by day length. *Physiol Plant* 162, 123–134. https://doi.org/10.1111/ppl.12593

Migicovsky, Z., Gardner, K.M., Money, D., Sawler, J., Bloom, J.S., Moffett, P., Chao, C.T., Schwaninger, H., Fazio, G., Zhong, G.-Y., Myles, S. (2016) Genome to phenome mapping in apple using historical data. *Plant Genome* 9. https://doi.org/10.3835/PLANTGENOME2015.11.0113

Migicovsky, Z., Yeats, T.H., Watts, S., Song, J., Forney, C.F., Burgher-MacLellan, K., Somers, D.J., Gong, Y., Zhang, Z., Vrebalov, J., van Velzen, R., Giovannoni, J.G., Rose, J.K.C., Myles, S. (2021) Apple ripening is controlled by a NAC transcription factor. *Front. Genet.* 12, 908. https://doi.org/10.3389/fgene.2021.671300

Minamikawa, M.F., Kunihisa, M., Noshita, K., Moriya, S., Abe, K., Hayashi, T., Katayose, Y., Matsumoto, T., Nishitani, C., Terakami, S., Yamamoto, T. (2021) Tracing founder haplotypes of Japanese apple varieties : application in genomic prediction and genome-wide association study. *Hortic. Res.* 8, 49. https://doi.org/10.1038/s41438-021-00485-3

Monell, J. (2021) Växtförädlingsmål för framtida äppelförädling ur förädlares och odlares perspektiv. SLU.

Näslund, G.K. (2010) Svenska äpplen, 1st ed. Kärnhuset, Sigtuna.

Nilsson, A. (1987) Våra äpplesorter, 2:1. ed. Allmänna Förlaget, Stockholm.

Nybom, H., (2019) Apple production and breeding in Sweden. *Chron. Horticult.* 59, 21–25.

Nybom, H. (2017) Blomning och fruktmognad hos olika äpplesorter. LTV-fakultetens faktablad, SLU, Alnarp.

Nybom, H. (1992) Freeze damage to flower buds of some apple cultivars. *J. Hortic. Sci.* 67, 171–177. https://doi.org/10.1080/00221589.1992.11516234

Peace, C.P., Bianco, L., Troggio, M., van de Weg, E., Howard, N.P., Cornille, A., Durel, C.E., Myles, S., Migicovsky, Z., Schaffer, R.J., Costes, E., Fazio, G., Yamane, H., van Nocker, S., Gottschalk, C., Costa, F., Chagné, D., Zhang, X., Patocchi, A., Gardiner, S.E., Hardner, C., Kumar, S., Laurens, F., Bucher, E., Main, D., Jung, S., Vanderzande, S. (2019) Apple whole genome sequences: recent advances and new prospects. *Hortic Res.* https://doi.org/10.1038/s41438-019-0141-7

Quiroga, M., Guerrero, C., Botella, M.A., Barceló, A., Amaya, I., Medina, M.I., Alonso, F.J., De Forchetti, S.M., Tigier, H., Valpuesta, V. (2000) A tomato peroxidase involved in the synthesis of lignin and suberin. *Plant Physiol.* 122, 1119–1128. https://doi.org/10.1104/PP.122.4.1119

Rains, M.K., De Silva, N.D.G., Molina, I. (2018) Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues. *Tree Physiol.* 38, 340–361. https://doi.org/10.1093/TREEPHYS/TPX060

Rymenants, M., van de Weg, E., Auwerkerken, A., De Wit, I., Czech, A., Nijland, B., Heuven, H., De Storme, N., Keulemans, W. (2020) Detection of QTL for apple fruit acidity and sweetness using sensorial evaluation in multiple pedigreed full-sib families. *Tree Genet. Genomes* 16, 1–16. https://doi.org/10.1007/s11295-020-01466-8

Spengler, R.N. (2019) Origins of the apple: The role of megafaunal mutualism in the domestication of *Malus* and rosaceous trees. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2019.00617

Sujeet, V., Whitaker, V.M. (2018) Prediction of QTL genotypes and trait phenotypes using FlexQTL^{TM}: A pedigree-based analysis approach. *J Plant Biol Crop Res.* 2: 1006.

Sun, X., Jiao, C., Schwaninger, H., Chao, C.T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K., Cheng, L., Zhong, G.Y., Fei, Z. (2020) Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* 52, 1423–1432. https://doi.org/10.1038/s41588-020-00723-9

Svensson, H. (2005) Äpplen i Sverige, 2nd ed. Bokförlaget Prisma, Stockholm.

Urrestarazu, J., Muranty, H., Denancé, C., Leforestier, D., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Aubourg, S., Celton, J.M., Daccord, N., Dondini, L., Gregori, R., Lateur, M., Houben, P., Ordidge, M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Troggio, M., Bianco, L., Velasco, R., Poncet, C., Théron, A., Moriya, S., Bink, M.C.A.M., Laurens, F., Tartarini, S., Durel, C.E. (2017) Genome-wide association mapping of flowering and ripening periods in apple. *Front. Plant Sci.* 8, 1923.

https://doi.org/10.3389/fpls.2017.01923
Vanderzande, S., Howard, N.P., Cai, L., Da Silva Linge, C., Antanaviciute, L., Bink, M.C.A.M., Kruisselbrink, J.W., Bassil, N., Gasic, K., Iezzoni, A., Van de Weg, E., Peace, C. (2019) High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. *PLoS One* 14, e0210928. https://doi.org/10.1371/journal.pone.0210928
Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L.M., Gutin, N., Lanchbury, J., MacAlma, T., Mitchell, J.T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V.T., King, S.T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, A., Vecchietti, A., Kater, M.M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A.C., Bus, V., Chagné, D., Crowhurst, R.N., Gleave, A.P., Lavezzo, E., Fawcett, J.A., Proost, S., Rouzé, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R.P., Durel, C.-E.E., Gutin, A., Bumgarner, R.E., Gardiner, S.E., Skolnick, M., Egholm, M., Van de Peer, Y., Salamini, F., Viola, R. (2010) The genome of the domesticated apple (Malus × domestica Borkh.). *Nat Genet* 42. https://doi.org/10.1038/ng.654
Verma, S., Evans, K., Guan, Y., Luby, J.J., Rosyara, U.R., Howard, N.P., Bassil, N., Bink, M.C.A.M.A.M., van de Weg, W.E., Peace, C.P. (2019) Two large-effect QTLs, Ma and Ma3, determine genetic potential for acidity in apple fruit: breeding insights from a multi-family study. *Tree Genet. Genomes* 15, 1–17. https://doi.org/10.1007/s11295-019-1324-y
Vitasse, Y., Schneider, L., Rixen, C., Christen, D., Rebetez, M. (2018) Increase in the risk of exposure of forest and fruit trees to spring frosts at higher elevations in Switzerland over the last four decades. *Agric. For. Meteorol.* 248, 60–69. https://doi.org/10.1016/j.agrformet.2017.09.005
Zhang L., Hu J., Han X., Li J., Gao Y., Richards C.M., Zhang C., Tian Y., Liu G., Gul H., Wang D., Tian Y., Yang C., Meng M., Yuan G., Kang G., Wu Y., Wang K., Zhang H., Wang D., Cong P. (2019) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun*; 10: 1–13. http://dx.doi.org/10.1038/s41467-019-09518-x
Zhu, Y., Li, G., Singh, J., Khan, A., Fazio, G., Saltzgiver, M., Xia, R. (2021) Laccase directed lignification is one of the major processes associated with the defense response against *Pythium ultimum* infection in apple roots. *Front. Plant Sci.* 12, 1862. https://doi.org//doi.org/10.3389/fpls.2021.629776

Popular science summary

Most people probably have their own ideas about the most tasty types of apple. Some can even identify a few apple cultivars by name and have personal favourites among the different cultivars available. The question then is whether we need plant breeding to develop new cultivars, if we already have access to so many different tasty apples?

While there are indeed vast numbers of apple cultivars available, few can be grown commercially at a profit. Commercial apple producers struggle every day with the challenges of running an orchard and, as the global climate and cultivation practices change, new problems are emerging. The purpose of plant breeding is to develop new cultivars that are better adapted to current or future conditions, thereby easing the burden on growers and helping to improve the profitability of apple production. Advances achieved by breeding may include improved resistance to pests and diseases, to reduce the need for chemical treatments if such are available and better adaptation to the current and future climate (considering ongoing global warming).

Sweden's geographical location in the Scandinavian Peninsula in northern Europe presents special challenges to local apple growers, for example by a short growing season. In addition, the cool humid climate favours a serious disease known as European canker, which causes severe losses to growers by strangling branches or entire trees. Furthermore, with future global warming bud break and flowering is expected to occur earlier, which might lead to an increased risk of late spring frosts occurring during flowering, which is already an important problem that could ruin the yield of entire orchards.

Recognising the special challenges faced by Nordic apple growers, Sweden has had a dedicated domestic apple breeding programme running since the 1940s, with the goal of delivering new cultivars that can help farmers meet these special challenges. The aim of this thesis was to establish the infrastructure needed to make use of technological advances in plant breeding achieved in the past decade, so that

the Swedish apple programme can continue doing its important work, but with increased efficiency.

The work in the thesis started by investigating the genetic resources currently available (Paper I) and making a contribution to the available toolbox (Paper II). The information obtained was used to identify regions of the apple genome conferring resistance to European canker (Papers III & IV). The effect under Nordic conditions of some previously identified regions of the apple genome regulating date of flowering and fruit harvest was then assessed (Paper V). Date of 50% canopy autumn senescence was found to be associated with how far north a cultivar can be grown, which is an important first step in gaining better knowledge about the genetic basis of adaptation to central and northern Sweden (Paper VI). Lastly, use of the knowledge generated in the thesis was exemplified by a discussion on possible crosses for further breeding in relation to the key traits studied.

Populärvetenskaplig sammanfattning

De flesta människor har förmodligen sin egen ide om hur ett perfekt äpple ska vara. En del av oss kan till och med identifiera ett par äppelsorter vid namn och har personliga favoriter bland de sorter vi smakat. Man kan ställa sig frågan om vi egentligen behöver växtförädling för att ta fram nya äppelsorter, när det redan finns så många olika välsmakande äpplen?

Trots att det finns ett mycket stort antal äppelsorter att tillgå är det endast ett fåtal av dessa som det är ekonomiskt lönsamt att odla kommersiellt. Kommersiella äppelodlare kämpar dagligen med odlingstekniska utmaningar. Med den globala uppvärmningen och nya odlingsmetoder kommer också nya problem. Syftet med växtförädling är att utveckla nya sorter som är bättre anpassade till dagens och framtidens odlingsförhållanden, och på så sätt göra livet lite lättare för äppelodlarna och bidra till en förbättrad lönsamhet i deras verksamhet. Växtförädling kan bidra med t.ex. förbättrad motståndskraft mot olika sjukdomar, vilket kan minska behovet av kemisk bekämpning om sådan finns tillgänglig, och bättre anpassning till dagens eller framtidens klimat i ett visst område.

Sverige utgör geografiskt en del av den skandinaviska halvön i norra Europa, vilket innebär specifika utmaningar för landets äppelodlare som t.ex. en relativt kort odlingssäsong. Det svala och fuktiga klimatet är dessutom gynnsamt för en allvarlig växtsjukdom känd som fruktträdskräfta. Fruktträdskräfta orsakar stora förluster för odlare genom att strypa näringstransporten mellan rötterna och enskilda grenar eller hela trädet, beroende på var infektionen sitter. Med den pågående globala uppvärmningen förväntas knoppsprickning och blomning att ske tidigare på året, vilket kan leda till en ökad risk för att sen nattfrost ska infalla under blomning. Frost i blomning är redan i dag ett alvarligt problem som kan fördärva skörden för hela odlingar.

Mot bakgrund av de speciella utmaningar som landets äppelodlare möter har Sverige ett dedikerat inhemskt växtförädlingsprogram på äpple sedan 40-talet, som har som mål att förse svenska odlare med nya äpplesorter som bättre hanterar landets specifika utmaningar. Målet med den här avhandlingen har varit att etablera den infrastruktur som krävs för att utnyttja den tekniska utveckling som skett inom växtförädlingen det senast årtiondet, så att det svenska växtförädlingsprogrammet på äpple kan fortsätta sitt viktiga arbete med ökad effektivitet.

Arbetet i den här avhandlingen började med en undersökning av de tillgängliga genetiska resurserna (Paper I) och ett bidrag till den tekniska verktygslådan för växtförädling och forskning på äpple (Paper II). Informationen från dessa två studier användes sedan till att identifiera områden i äpplets arvsmassa som bidrar till ökat motståndskraft mot fruktträdskräfta (Paper III & IV). Tidigare studier har identifierat områden i arvsmassan som påverkar datum för blomning och skördemognad av frukt, och effekten av dessa regioner undersöktes under nordiska förhållanden (Paper V). Datumet för senescens av bladverket visade sig vara korrelerat med hur lång norrut en äpplesort kan odlas, vilket är ett första steg på vägen mot en ökad förståelse av hur anpassning till mellersta och norra Sverige styrs genetiskt (Paper VI). Till sist illustreras hur den samlade informationen kan användas inom växtförädlingen för att identifiera intressanta korsningar för vidare sortutveckling och forskning.

Acknowledgements

I would like to thank everyone who has contributed to this project and supported me along the way. First I would like to thank my principal supervisor *Larisa Gustavsson* for believing in me (most importantly during the recruitment process) and treating me with the respect of a peer (most of the time), which has allowed me to grow, both personally and scientifically. I would also like to thank my co-supervisors *Firuz Odilbekov*, *Kimmo Rumpunen*, and *Pär K. Ingvarsson* for generously sharing advice and knowledge in their different fields of expertise. A very special thanks goes to *Eric van de Weg* who has been something of an unofficial co-supervisor throughout this project, and I am deeply grateful for all the time he has invested in valuable discussions on various topics. Also, if it had not been for *Eric*'s prior efforts, I am convinced that this would have been a very different kind of project, and a much less interesting one.

I am very grateful for the nice collaborations with *Bjarne Larsen* and *Inger Hjalmarsson*, which have made this project much more interesting. Also, this project would have been much more challenging if it had not been for some friendly colleagues in the community: *Nick Howard* and *Stijn Vanderzande* have been very kind in answering my questions, and much of the work I have done is built upon their efforts to some degree. If it had not been for *Sujeet Verma* I might have abandoned the idea of using FlexQTL for QTL mapping at an early stage.

During my time as PhD student, *Elna Sjöström* and *Birger Olsson* have supplied around 3,300 apple trees of the highest quality (equivalent of around 1.3 ha of relatively intensive orchard), which have all been killed without mercy in the name of QTL mapping.

I would like to thank all the personnel at SLU whom I have encountered, who have contributed to the friendly culture and stimulating academic

environment. My time as a PhD student would have been a very different experience (and not for the better) if it had not been for the helpful and competent technical staff at the Department of Plant Breeding (especially *Helena Mattisson*, *Marina Kuzmenkova*, and *Oksana Korniienko*) and the equally helpful and competent field personnel at Balsgård: *Barbro Johnsson* and *Stefan Olsson*. The Sequencing and Genotyping platform at Fondazione Edmund Mach (FEM, Italy) have provided excellent genotyping service, despite my consistent delays in submitting samples.

This work benefitted from a Nordic collaboration through the PPP Nordfruit project, with a special thanks to *Tuuli Haikonen*, *Dag Røen*, *Vidmantas Bendokas*, *Hedi Kaldmäe*, and *Kristine Volens* for providing material and data through the project. Thanks also to *Lidija Bitz* (Luke), *Jakob Sandberg* (Fredriksdal), *Gordana Đurić* and *Mirela Kajkut Zeljković* (University of Banja Luka) for sharing material for genotyping.

This PhD project was funded by the SLU Grogrund project "Accelererad och kostnadseffektiv äpplesortutveckling genom genombaserad växtförädling".

Most importantly, none of this would have been possible if it had not been for the generous support of my family. Special thanks goes to *Ulv Weström* for providing the very nice cover illustration. Most importantly: *Erika*, *Hilda*, and *Stina*, thank you for reminding me of the important things in life.

I

Scientia Horticulturae 272 (2020) 109599

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Genetic Status of the Swedish Central collection of heirloom apple cultivars

Jonas Skytte af Sätra[a,*], Michela Troggio[b], Firuz Odilbekov[a], Jasna Sehic[c], Helena Mattisson[a], Inger Hjalmarsson[d], Pär K. Ingvarsson[e], Larisa Garkava-Gustavsson[a]

[a] Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden
[b] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Trento, Italy
[c] Department of Plant Breeding – Balsgård, Swedish University of Agricultural Sciences, Fjälkestadsvägen 459, 29194, Kristianstad, Sweden
[d] Department of Landscape Architecture, Planning and Management, Swedish University of Agricultural Sciences, Alnarp, Sweden
[e] Department of Plant Biology, Swedish University of Agricultural Sciences, 750 07, Uppsala, Sweden

ARTICLE INFO

Keywords:
Malus domestica
SNP array
Genotyping
Pedigree
Gene bank
Local cultivars
SSR markers

ABSTRACT

Cultivated apple is one of the most widely grown fruit crops worldwide. With the introduction of modern apple cultivars, from foreign and national breeding programs, the use of local cultivars decreased during the 20th century. In order to minimize genetic erosion and avoid loss of special genotypes, a number of local clonal archives were established across Sweden, with the goal of retaining old and local cultivars. About 220 apple cultivars, appointed for preservation, obtained the status of mandate cultivars. Initially, they were identified based on pomological traits, but prior to the establishment of the Swedish Central Collection they were genotyped with simple sequence repeat (SSR) markers. SSR markers helped to evaluate the status of the preserved material, as well as to find the best possible true-to-type source for propagation, thus guiding the establishment of the Central Collection. Recently, 215 accessions from this collection were genotyped using the 20 K apple Infinium® single nucleotide polymorphism (SNP) array, in order to gain insight into its genetic structure. The initial SSR analysis confirmed the identity of multiple samples with the same cultivar name grown in different locations and identified several mislabeled samples. In the subsequent SNP analysis we identified 30 clonal relationships and a number of parent-offspring relationships, including 18 trios. We also identified five cultivar samples with inconsistent ploidy levels between the SNP and SSR data, in some cases indicating problematic samples preserved in either the Central Collection or some of the local clonal archives. These cultivars need further investigation to ensure their true-to-typeness. Furthermore, the Swedish Central Collection has continued to grow since the onset of this work and now contains additional cultivars, which should be included in future studies. The results indicate that a number of the preserved mandate cultivars holds high potential value for modern breeding programs.

1. Introduction

The cultivated apple (*Malus domestica* Borkh.) is one of the world's most widely grown fruit crops. It belongs to the family Rosaceae, is diploid (2n = 34x), and has life-history traits characteristic of perennial fruit crops, i.e., it is outcrossing and to a large extent self-incompatible, has a long juvenile period and a long life span, and is often clonally propagated (Gaut et al., 2015). This generally leads to a high degree of heterozygosity and a large proportion of genetic diversity being retained following domestication (Miller and Gross, 2011). These life-history traits also facilitate close genetic relationships between very old cultivars and those emerging from modern breeding programs. In Sweden, a program studying pollination compatibility between cultivars was established in Alnarp in the 1920s. Seeds from interesting crosses were sown out and resulted in a number of cultivars being released (Nilsson, 1987). The current Swedish apple breeding program was established in Balsgård in the 1940s, and since then has released a number of cultivars adapted to the Scandinavian climate, both for commercial production and home gardening (Nybom, 2019). Some of the cultivars released from these two programs have been appointed for preservation, despite their modern source.

As modern cultivars increased in popularity among commercial growers, a national inventory and collection of local cultivars was initiated in 1979 by Nordiska Genbanken (now NordGen), to prevent loss of genetic resources. This resulted in the establishment of the first local clonal archives spread across the country and a compilation of

* Corresponding author.
E-mail address: Jonas.skytte.af.satra@slu.se (J. Skytte af Sätra).

https://doi.org/10.1016/j.scienta.2020.109599
Received 28 February 2020; Received in revised form 9 June 2020; Accepted 9 July 2020
Available online 17 July 2020

pomological descriptions (Nilsson, 1987). Today the responsibility for conservation of heirloom apple cultivars lies with the National Gene Bank of Vegetatively Propagated Crops, similarly to other Scandinavian countries, e.g. Norway (Gasi et al., 2016) and Finland (Heinonen and Bitz, 2019). Starting in 2013, the local clonal archives were complemented by the establishment of the Swedish Central Collection of mandate cultivars at SLU in Alnarp, with the majority of the trees being planted in 2014 and 2015. Mandate cultivars are heirloom cultivars assigned for preservation by the National Gene Bank and are defined as either i) being of local origin according to lore, or ii) originating from Swedish breeding programs, or iii) foreign cultivars with a long documented history of cultivation (Hjalmarsson, 2019, Hjalmarsson, 2020). Thus, the primary criteria for the selection of mandate cultivars was their specific cultural importance and not the inherent genetic diversity.

The Swedish Central Collection of mandate cultivars is expected to include genetic variation relevant for local adaptation, as it contains cultivars with a wide range of climate hardiness (Nilsson, 1987). This collection has the potential to become a rich source of favorable alleles for the Swedish apple breeding program. Current knowledge concerning relationships among the Swedish mandate cultivars is limited, and mainly based on historical pomological records (Nilsson, 1987), pedigree records in the breeding programs, and previous studies using SSR markers (Garkava-Gustavsson et al., 2008; Urrestarazu et al., 2016). Recently, the pedigrees of some cultivars in the collection have been elucidated using a large set of SNP markers (Muranty et al., 2020). Clarifying unknown pedigrees greatly enhances the prospects for germplasm utilization, by increasing the accuracy of pedigree-based quantitative trait loci analyses (Howard et al., 2017) and reducing the risk of unwanted inbreeding in future crosses.

Genotyping of gene bank material is crucial in order to ensure that genetic variation is maintained, but also to make gene bank material a useful resource for breeding. In the past, SSR markers have been widely used in germplasm studies of apple at the national, regional, and European level (Garkava-Gustavsson et al., 2013, 2008; Gasi et al., 2016; Heinonen and Bitz, 2019; Larsen et al., 2017; Lassois et al., 2016; Marconi et al., 2018; Patocchi et al., 2009; Routson et al., 2009; Urrestarazu et al., 2016; van Treuren et al., 2010). More recently, genotyping-by-sequencing (GBS), SNP arrays, and diversity array technology (DArT) markers have been used to unravel apple pedigrees (Larsen et al., 2018; Muranty et al., 2020; Ordidge et al., 2018; Vanderzande et al., 2017). The IRSC 8 K SNP array (Chagné et al., 2012) available for apple has been useful in elucidating pedigrees (Howard et al., 2017), while the 20 K Infinium® SNP array (Bianco et al., 2014) that followed is currently being used in an ongoing apple pedigree project (Howard et al., 2018). The latter has also been used in the construction of a high-density genetic map (Di Pierro et al., 2016) and QTL-discovery for several traits (Allard et al., 2016; Di Guardo et al., 2017; van de Weg et al., 2018). Thus, genotyping a collection with the 20 K SNP array not only provides robust assessments of pedigree inference and population structure, but also improves the usefulness of the collection by allowing integration with larger germplasms and facilitates its implementation in a breeding-by-design concept (Peleman and Van Der Voort, 2003) that utilizes published information on the genetic control of relevant traits.

In the present study, we first demonstrate how SSR markers were used to assist the establishment of the Swedish Central Collection of mandate cultivars by identifying the best possible true-to-type source of graft wood for the trees intended for preservation in the Central Collection. Recently, we used the 20 K apple SNP array to characterize 215 apple accessions from the Swedish Central Collection. The genotypic data was used to describe the status of the collection by identifying clonal relationships, putative 1st degree relationships, and possible parent-offspring relationships, and by assessing population structure.

2. Material and methods

2.1. Plant material

For SSR analysis, young leaves from a total of 204 apple cultivars were collected in May – early June (2007–2009) and stored at −80 °C until use. Samples were collected from different clonal archives, the stock collection of the Swedish Elite Plant Station (EPS, holding true-to-type and virus-free genotypes in stock orchards in Fjälkestad), the research- and breeding-oriented germplasm collection at the Swedish University of Agricultural Sciences (SLU-Balsgård) and the Institute of Horticulture in Kiev (Ukraine). If several trees of the same cultivar were present in different collections, all those were sampled and analyzed resulting in 340 samples (Supplementary File 1).

For SNP genotyping, leaf samples were collected from 215 accessions of *Malus domestica* Borkh. in the Swedish Central Collection of mandate cultivars (Alnarp, Sweden), including two rootstocks: 'A2' on which the collection is grafted and 'Bemali' (Supplementary File 1). Young leaves from growing shoots of a single tree were collected at the end of May 2018, freeze-dried, and stored at −80 °C. Extraction of total genomic DNA was performed using the DNeasy Plant Mini Kit (Qiagen) following the standard protocol. The quality and concentrations of the DNA samples were measured by spectrometry (Nanodrop, Thermo Scientific). The sample names used in this work are the accepted names according to the Swedish Utility and Cultivated Plants Database (SKUD) in January 2020, whenever such were available.

2.2. DNA-amplification and SSR analysis

Genomic DNA was isolated from approximately 100 mg of frozen leaves, using the Qiagen Dneasy™Plant Mini Kit (Qiagen) following the manufacturer´s protocol. Twelve primer pairs, developed by Liebhard et al. (2002) were used for SSR-analysis: CH01h02, CH02c06, CH02c09, CH02c11, CH02d08, CH04c06, CH04c06, CH04e05, COL, CH02b10, CH04e02 and CH02c02b. The forward primers were fluorescently labeled at the 5´-end using either FAM or HEX.

Amplifications were performed for each primer pair separately as described in (Garkava-Gustavsson et al., 2008). PCR-products were pseudo-multiplexed before being analyzed on a 3730 DNA analyzer (Applied Biosystems). Size of the amplified products was calculated based on an internal standard (500ROXTM Size Standard (Applied Biosystems)) and evaluated using GeneMapper ® software v. 3.0 (Applied Biosystems).

2.3. Flow cytometry

The flow cytometry analyses were conducted on the samples showing more than two alleles at the Plant Cytometry Services (JG Schijndel) in the Netherlands. The fresh leaves were chopped in ice-cold buffer with DAPI, and flow cytometry was performed on a CyFlow ML (Partec GmbH, Műnster, Germany) using *Lactuca sativa* as internal standard, and 'Cox Orange' as a ploidy level standard.

2.4. Analysis of SSR-data

The SSR fragments were scored in terms of loci and alleles, and thus the allelic composition of each sample was determined. Then Simple Matching coefficient (Sokal and Michener, 1958) was calculated for all the analyzed samples with NTSYS-pc statistical package v. 2.2 (Rohlf, 2005) to produce a similarity matrix and detect accessions with identical allelic profiles (a few mis-matches in one allele were allowed for the genotypes to be considered as identical). Allelic compositions were examined to verify the identity of cultivars with their known sports, to assure that accessions with the same name had identical profile, as well as to identify problematic cases (e.g., obvious identification mistakes and mislabeling).

Next, all triploid cultivars and a tetraploid cultivar were removed from the data set. The genetic diversity of the remaining 150 unique profiles of diploid cultivars was estimated using GenAlex 6.503 (Peakall and Smouse, 2006, 2012). Genetic diversity parameters analyzed included: number of different alleles (Na), number of effective alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe) and inbreeding coefficient (F).

2.5. SNP genotyping

The SNP genotyping was carried out using the 20 K Infinium® SNP array (Illumina Inc.) (Bianco et al., 2014). Samples comprising 200 ng of genomic DNA from each of the 215 accessions were analyzed, following the standard Illumina protocol detailed in Chagné et al. (2012). Intensity data were analyzed using the workflow described by Vanderzande et al. (2019), with one major deviation. In Vanderzande et al. (2019) the cluster definitions were obtained from ASSiST (Di Guardo et al., 2015), while in this project we used manually adjusted cluster definitions and a subset of SNPs obtained through an ongoing apple pedigree project (Howard et al., 2018) consisting of 10,368 SNPs. The cluster definitions were loaded to Genome Studio (GS), v 2.0 (Illumina Inc.) was implemented, and the SNP genotype calls obtained were used for further analyses. Sample quality was assessed using the overall call rate and the histogram distributions of the B-allele frequency and the R-parameter.

2.6. SNP-data analysis

Putative ploidy levels were determined using frequency plots according to Chagné et al. (2015), as described by Vanderzande et al. (2019). Duplicate individuals were identified using PLINK 1.9 (Chang et al., 2015), using the Pi_Hat parameter as an approximation of the global pairwise identity-by-decent (IBD) of individuals. PLINK utilizes a method-of-moments approach to estimate proportion of alleles shared IBD (PI_HAT, $\hat{\pi}$):

$$\hat{\pi} = P(Z = 1)2 + P(Z = 2),$$

where P(Z = 1) and P(Z = 2) is the global estimate of the proportion of loci with 1 and 2 alleles in common, respectively (Purcell et al., 2007).

Duplicates as well as known sports were used as quality controls, to verify the accuracy of the SNP genotyping procedure. In the case of unexpected duplicates, these were further examined for morphology of individual trees, records from sample preparations, historical descriptions and SSR data from local archives. All duplicates were removed from the dataset, which was then analyzed with PLINK to identify putative 1st degree relationships, i.e., pairs of individuals with coefficients of relationship equal to 0.5, such as full siblings and parent-offspring relationships.

The available pedigree records were verified using the R-script for counting Mendelian inconsistent errors, as described by Vanderzande et al. (2019). The pedigree records were adjusted for any errors detected and the script was re-run to identify possible parent-offspring combinations, using thresholds of 30 and 60 Mendelian inconsistent errors for parent-offspring (P-O) and parent-parent-offspring (P-P-O) combinations, respectively. To illustrate the results, network plots were generated using the iGraph package (Csardi and Nepusz, 2006) in RStudio version 1.1.463.

Assessment of population structure was performed following removal of synonymous samples, but without pruning for linkage disequilibrium (LD) or minor allele frequencies (MAF). Principal component analysis (PCA) was carried out in PLINK and plotted in Excel 2016 (Microsoft). Statistical analysis was carried out using a model-based clustering as implemented in the software STRUCTURE v2.3.4 (Falush et al., 2007, 2003; Pritchard et al., 2000). STRUCTURE was run with Parallel GNU (Tange, 2011) while assuming admixture, correlated allele frequencies between population, and implementing the LD model, using the genetic positions from Di Pierro et al. (2016). Initially, a single-replicate analysis was run assuming 1–10 subpopulations (K), in order to determine suitable settings. Five independent replicates were then performed, with a burn-in of 15,000 and a run length of 50,000 for each K from 1 to 5. Next, the most probable number of subpopulations was estimated using the method from Evanno et al. (2005), as implemented in Structure Harvester (Earl and von Holdt, 2012).

3. Results

3.1. Ploidy level and duplicates

All primer pairs produced clear and consistent DNA-profiles, which were used in the analysis. In all loci except two (CH04c06 L1 and CH02c02b) three alleles were found in some genotypes. Some of these were previously known triploid cultivars, e.g. 'Gravensteiner'. However, some cultivars were not previously described as being triploid and their ploidy level was confirmed by flow cytometry analyses, e.g. 'Sköldinge'. Analysis of the similarity matrix confirmed a number of samples with identical SSR profiles, as expected as multiple samples of the same cultivar had been collected from different locations. Others concerned samples with an SSR based indication of a need for further investigation, i.e. were 9 cases of samples with the same name having distinct SSR profiles when collected from different collections. In addition, there were 49 cases of multiple cultivars sharing the same SSR profile, indicating potentially synonymous cultivars or mislabeling (Supplemental File 1).

3.2. Genetic diversity

The number of alleles varied from 2 (CH04c06 L1) to 20 (CH02c06) (Table 1). Loci with a high number of alleles also produced high values for H_o and H_e, and vice versa. For each locus, H_o values (from 0.30 to 0.89) were quite similar to H_e values (from 0.34 to 0.88). The highest H_o was found at locus CH02b10, while the lowest at locus CH02c02b. Locus CH02b10 yielded the highest H_e value, while locus CH04c06 L1 had the lowest. All loci had F coefficients bellow 0.1, except CH02c02b which deviated with a value of 0.49.

3.3. SNP quality check and ploidy

Of the 10,368 SNPs considered here, 406 showed significant deviations from Hardy-Weinberg equilibrium ('ChiTest100′ column in the

Table 1
Genetic diversity in terms of number of different alleles (Na), number of effective alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho) expected heterozygosity (He), unbiased expected heterozygosity (uHe) and inbreeding coefficient (F) from the 150 unique diploid genotypes analyzed.

Locus	Na	Ne	I	Ho	He	UHe	F
CH01h02	12	3.06	1.44	0.67	0.67	0.68	0.00
CH02c06	20	7.77	2.34	0.86	0.87	0.87	0.01
CH02c09	11	7.58	2.13	0.85	0.87	0.87	0.02
CH02c11	14	6.97	2.15	0.83	0.86	0.86	0.03
CH02d08	15	4.41	1.85	0.79	0.77	0.78	−0.02
CH04c06 L1	2	1.51	0.52	0.37	0.34	0.34	−0.11
CH04c06 L2	12	6.45	2.07	0.86	0.84	0.85	−0.02
CH04e05	13	3.84	1.76	0.75	0.74	0.74	−0.01
COL	11	5.38	1.85	0.75	0.81	0.82	0.08
CH02b10	17	8.57	2.37	0.89	0.88	0.89	−0.01
CH04e02	11	4.29	1.68	0.76	0.77	0.77	0.01
CH02c02b	8	2.43	1.16	0.30	0.59	0.59	0.49
Mean	12	5.00	1.78	0.72	0.75	0.75	0.04
Total	146	62					

Table 2
Sample ploidy level inconsistencies. C_{SSR} indicates ploidy level expected from simple sequence repeat (SSR) data, validated by flow cytometry and C_{SNP} indicates ploidy level based on single nucleotide polymorphism (SNP) data.

Sample	C_{SSR}	C_{SNP}
'Bosebo Sötäpple'	3n	2n
'Cellini'	2n	3n
'Holländare'	3n	2n
'Skälbyäpple'	3n	2n
Pomme de Cannelle'	3n	2n

GS SNP table) in this collection. Six of these SNPs were not called, 35 displayed an excess of homozygosity and the remaining 365 SNPs displayed an excess of heterozygosity. Since the fraction of markers deviating from Hardy-Weinberg equilibrium was low (< 5%), these markers were included in the downstream analyses.

Using cluster definitions and a robust subset of SNPs from an ongoing apple pedigree project (Howard et al., 2018), all samples had an overall call rate above 0.88 (except for the tetraploid 'Alfa68′, which had a call rate of 0.80) and p50 GC values between 0.75 and 0.81. All samples had B-allele frequency (distinct heterozygous and homozygous peaks) and R-parameter histograms (low values on the x-axis) in GS, indicating good sample quality without contamination (distinct heterozygous and homozygous peaks), as described by Vanderzande et al. (2019). Next, we inspected the B-allele frequency plots generated using the method described in Chagné et al. (2015) which, together with the B-allele frequency histograms in GS, allowed us to identify sample ploidy. A few samples had ploidy levels differing from those revealed based on SSR markers (Table 2). Interestingly, 'Tegnérsäpple' showed a B-allele frequency plot indicating a translocation/aneuploidy at Linkage Group 15 (Supplementary File 2)

3.4. Duplicates in the Swedish Central Collection

Duplicate samples were defined as having global pairwise estimates of IBD > 0.999, as there were no sample pairs with IBD values between 0.99 and 0.7. Analysis for the presence of duplicates resulted in 185 unique samples out of the total 215 apple accessions tested. In the case of triploid synonymous samples, their B-allele frequencies were investigated to ensure that they had the same heterozygous state (i.e., either 'AAB' or 'ABB'). All synonymous triploid samples had B-allele frequencies with correlations above 0.99, indicating that they were identical. All expected duplicates were identified, both duplicate samples and color sports. However, we identified a number of unexpected duplicates and these were compared against the SSR data obtained from the local clonal archives. Two cases, 'Rödluvan'/'Borsdorfer' and 'A2′/'Björnegårdsäpple', were clear examples of mislabeling or propagation mistakes, and the true-to-type individuals were revealed based on examination of morphological characters. In the case of 'Cox's Orange Pippin'/'Skälbyäpple', the IBD coefficients to other cultivars were investigated and were found to be consistent with the pedigree of 'Cox's Orange Pippin'. Duplicate samples were excluded from further analysis (Table 3).

3.5. 1st degree relationships, P-O relationships, and possible crosses

Having removed duplicates, the remaining 185 samples were reanalyzed for pair-wise IBD. A threshold for putative 1st degree relationships was set at IBD = 0.484 (0.5–1/64), guided by the IBD values of samples with known relationships. A draft pedigree based on previous SSR studies (Urrestarazu et al., 2016) and historical records (Svensson, 2005) concerning the cultivars was then constructed and analyzed for pedigree errors using an R-script as described by Vanderzande et al. (2019). Erroneous parents (3 cases; Supplementary File 1) were removed and the pedigree was re-analyzed to search for possible parent-offspring combinations. Nineteen possible crosses were identified, all of which appeared plausible according to the historical record of cultivar ages (Nilsson, 1987; Svensson, 2005) (Table 4). Most of the offspring were local Swedish cultivars, with two exceptions, 'Transparante Blanche' and 'Gestreifter Wintercalvill'. These are thought to be of Baltic/Russian and German origin, respectively, and both have possible parents that are either very old or of unknown origin (Nilsson, 1987).

3.6. Possible parents

A number of possible parent-offspring relationships within the collection were also identified, heavily biased towards a few cultivars. The cultivars with the highest number of possible parent-offspring relationships were 'Gimmersta' (29), 'Grågylling' (19), 'Vitgylling' (16), 'Rosenhäger' (7), and 'Klockhammarsäpple' (6). For most cultivars, the number of possible parent-offspring relationships was in line with the number of 1st degree relationships estimated from IBD values (Supplementary File 1).

3.7. Network

Based on the possible parent-parent-offspring combinations identified, parent-offspring relations, and estimated 1st degree relationships, a single network was identified for the collection, comprising 115 cultivars interconnected through 1st degree relationships (Fig. 1). Forty-nine of the 185 unique cultivars were found to have no 1st degree relationships within the collection. A number of putative 1st degree relationships that are not supported by the test for Mendelian inconsistent errors were also highlighted. Notably, 'Sköldinge' and 'Alnarps Favorit' had a low number of possible parent-offspring relationships, while having a relatively high number of putative 1st degree relationships.

3.8. Genetic structure

In PCA, the first three principal components explained 9.0, 6.1, and 4.8 % of the genetic variation in the SNP set in the collection, with no evident population structure (Fig. 2). Accordingly, the STRUCTURE analysis yielded the highest $L(K)$ value for K = 1 and ΔK dropped to 5 already at K = 2, indicating that the Swedish Central Collection most likely consists of one sub-population (Supplementary File 1).

4. Discussion

4.1. SSR-analysis of local clonal archives

The SSR analysis of the local clonal archives elucidated several accession that were either duplicates or mislabelings. Some of the interesting cases involved the different accessions traditionally referred to as 'Antonovka'. 'Antonovka' from Bergianska trädgården has the same SSR profile as 'Antonovka' from Julita and 'Antonovka Pamtorutka' from Balsgård. The latter was found to be identical to 'Antonovka Polotora Funtovaja' by Urrestarazu et al. (2016), which has been the preferred name in subsequent studies (Muranty et al., 2020). In contrast, 'Antonovka' from Ekebyhov is identical to 'Antonovka Kamenichka' at Balsgård and a sample of 'Antonovka' imported from the Institute of Horticulture, Kiev (Ukraine). According to the same studies as above, the preferred name for this cultivar has been 'Antonovka Obyknovennaja' (Muranty et al., 2020; Urrestarazu et al., 2016). 'Antonovka Polutora Funtovaja' and 'Antonovka Obyknovennaja' was recently found to have a parent-offspring relationship, which was directed based on historical data with the latter suggested as being one of the possible parents (Muranty et al., 2020). For the establishment of the Central Collection 'Antonovka' from Bergianska trädgården was used as graft wood source, i.e. 'Antonovka Polotora Funtovaja'. Thus, the collection

Table 3
Synonymous sample names based on identity-by-decent (IBD) analysis. Samples 1-4: Synonymous sample names. SSR: Simple sequence repeat data from the samples collected from the local archives confirming that the samples are synonymous (Same) or indicating that the sampled tree was not true to type (Diff). TTT: Phenotypic characters of the sampled trees in situ indicated that one of the sampled trees was not true-to-type (Diff).

Sample 1 (used name)	Sample 2	Sample 3	Sample 4	SSR	TTT
'Grågylling'	'Grågylling, Skokloster'	'Rödgylling'	'Hedemoraäpple'	n/a	–
'Ullströmsäpple'	'Gislaved'	'Hagbyberga'		Same	–
'Åkerö'	'Åkerö från Gripsholm'	'Åkerö' (mother tree)		Same	–
'Cox's Orange Pippin'	'Skälbyäpple'			Diff	–
'A2′	'Björnegårdsäpple'			Same	Diff
'Aroma'	'Amorosa'			n/a	–
'Alexander'	'Annero'			Same	–
'Aspa'	'Eneroths Klaräpple'			n/a	–
'Brita Horn'	'Ringstad'			n/a	–
'Gravensteiner'	'Roter Gravenesteiner'			Same	–
'Gyllenkroks Astrakan'	'Gyllenkroks Astrakan' (mother tree)			n/a	–
'Suislepp'	'Gul Höstkalvill'			Same	–
'Kristinaäpple'	'Nådig Frun'			Same	–
'Melonenapfel'	'Rött Melonäpple'			Same	–
'Mio'	'Röd Mio'			Same	–
'Sävstaholm'	'Sävstaholm' (mother tree)	'P.J. Bergius'		n/a	–
'Vitgylling'	'Ramsta'			Same	–
'Borsdorffer'	'Rödluvan'			Diff	Diff
'Särsö'	'Spässerud'			Same	–
'Sjöholmsäpple'	'Äs Baron'	'Stor Klar Astrakan'		Same	–
'Calville Blanche d'Eté'	'Calville Rouge d'Èté'			n/a	–
'Stenkyrke'	'Stenkyrke' (mother tree)			n/a	–
'Transparante Blanche'	'Tistads Vaxgylling'			Same	–
'Kavlås'	'Wickstrands Favorit'			Same	–

Table 4
Possible crosses identified using the R-script described in Vanderzande et al. (2019). Offspring, Parent 1 & Parent 2: Possible parent-parent-offspring trios identified, maternal/paternal parent cannot be identified. Age Offspring, Age Parent 1 & Age Parent 2: Recorded approximate first appearance of offspring and parent cultivars, from Nilsson (1987). n/a = unknown origin, old = cultivars of unclear origin, but older than ~1850, *arrival in Sweden, ** from Svensson (2005).

Offspring	Parent 1	Parent 2	Age Offspring	Age Parent 1	Age Parent2
'Adelhill'	'Alexander'	'Oranie'	n/a	1800s*	n/a
'Åkerö'	'Grågylling'	'Vitgylling'	1700s	old	old
'Bläsingeäpple'	'Stenkyrke'	'Summer Pearmain'	1900s	1800s	old
'Edsele'	'Rosenhäger'	'Vitgylling'	1900s	n/a	old
'Fredriksdalsäpple'	'Ullströmsäpple'	'Klockhammarsäpple'	1900s	1800s	1800s
'Frösvidal'	'Hampus'	'Grågylling'	1800s	1700s	old
'Grönsö'	'Grågylling'	'Vitgylling'	1800s	n/a	old
'Hanaskogsäpple'	'Oranie'	'Vitgylling'	1800s	n/a	old
'Hornsberg'	'Hampus'	'Gimmersta'	1800s	1700s	1800s
'Mälardalens Vitgylling'	'Klockhammarsäpple'	'Vitgylling'	n/a	1800s	old
'Mölnbacka Skogsäpple'	'Holmserudsäpple'	'Rosenhäger'	1800s	n/a	n/a
'Råby Rubin'	'Sävstaholm'	Maglemer'	1900s**	1800s	1700s
'Sandbergs Röda'	'Grågylling'	'Vitgylling'	1900s	old	old
'Sjöholmsäpple'	'Gimmersta'	'Vitgylling'	1800s	1800s	old
'Sparreholm'	'Gimmersta'	'Rosenhäger'	1800s	1800s	n/a
'Stäringe Karin'	'Gimmersta'	'Grågylling'	1800s	1800s	old
'Gestreifter Wintercalvill'	'Holmserudsäpple'	'Rosenhäger'	n/a	n/a	n/a
'Transparante Blanche'	'Aspa'	'Vitgylling'	1800s *	1800s	old

might benefit from including 'Antonovka Obyknovennaja' as well as it is likely to have a long history of cultivation in Sweden, even though there has been some confusion in past as both of these cultivars have commonly been designated as simply 'Antonovka'.

There were also some clear cases of mislabelling, e.g. 'Eva-Lotta' in one of the local archives appears to be identical to 'Alice' and 'Hornsberg' in one of the local archives was identical to 'Hanaskogsäpple'. Revealing of such cases provided valuable support for the choice of graft wood source for the establishment of the central Swedish Central Collection. Noteably, some accessions from the breeding collection in Balsgård appears to be not true to type, including 'Domö Favorit', 'Menigasker' and 'Silva' which all have different SSR profiles than accessions with the same names at the local clonal archives. Though beyond the scope of this study, this might be a relevant note, as several international studies including old Swedish cultivars have been based on samples obtained from the Balsgård collection (Muranty et al., 2020; Urrestarazu et al., 2017, 2016).

Of the 12 SSR-markers utilized here, only 5 are common with those used by Urrestarazu et al. (2016) and their usefulness for comparison against a larger germplasm is therefore limited. Nevertheless, SSR profiles have been made available to the MUNQ-database and MUNQ genotype codes have been assigned as previously described, whenever possible (Muranty et al., 2020; Urrestarazu et al., 2016).

In the analysis for genetic diversity only diploid accessions were considered, as triploids are considerably less efficient as parents. Thus, diversity conserved in polyploids has limited impact as an actual resource. Considering the diversity parameters analyzed, the local archives seem to retain a high degree of genetic diversity. The mean observed heterozygosity (Table 1) is close to the expected heterozygosity, and quite high, which is expected considering the life history traits of the species and being in line with comparable studies (Gaut et al., 2015; Larsen et al., 2017; Lassois et al., 2016; Marconi et al.,

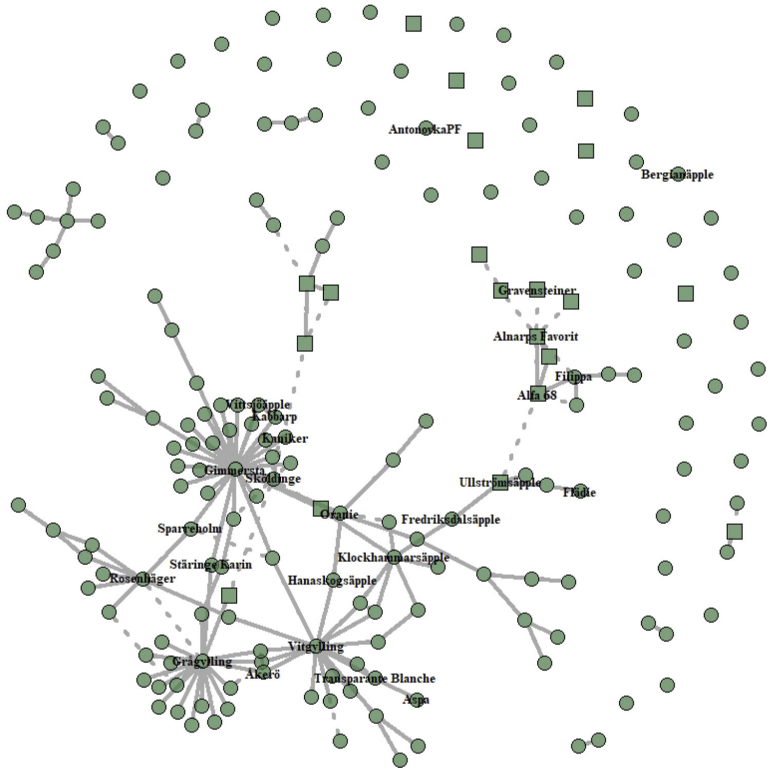


Fig. 1. Network plot of the Swedish Central Collection of heirloom apple cultivars. Diploid accessions are indicated as circles, triploid and tetraploid accessions as squares. Solid lines indicate possible parent-offspring and parent-parent-offspring relationships, including confirmed pedigree records; dotted lines indicate putative 1st degree relationship. Some of the cultivars mentioned in the discussion are designated with names.

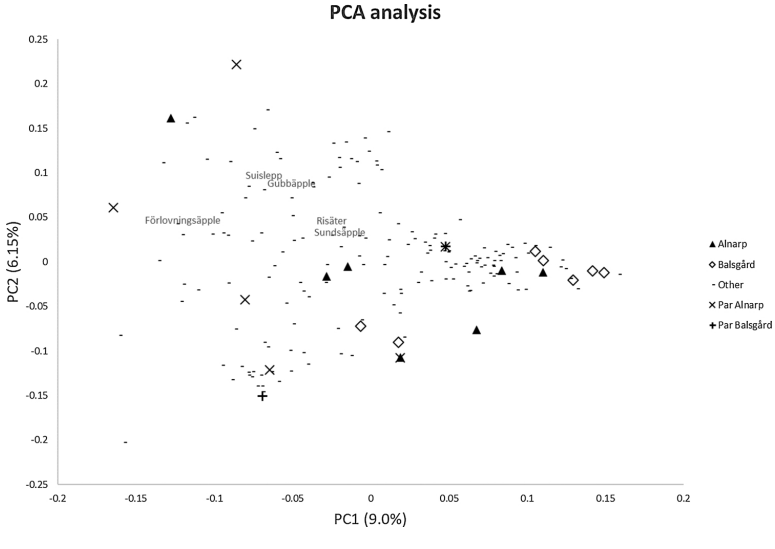


Fig. 2. A two-dimensional plot of principal component analysis (PCA) of the Swedish Central Collection of heirloom apple cultivars. Cultivars originating from the current Swedish breeding program in Balsgård (diamond), from the program in Alnarp (triangle) and their respective parents (star and cross) are indicated. Cultivars which can be considered as suitable parents for future breeding efforts, for northern Sweden, are labeled with names.

2018).

It should be noted, however, that the main criterion for the establishment of the Swedish Central Collection was not genetic diversity *per se*, but rather cultural diversity. Consequently, several duplicates with different names that were revealed by analysis with SSR markers were nonetheless considered as distinct mandate cultivars. While being genetically identical, duplicate accessions might still be sports, which can be difficult to distinguish without detailed investigations of morphology in common gardens. Another aspect is the richness in folk lore accompanying many of the old cultivars, which cases can be just as important as the genetic properties of an old cultivar.

4.2. Central collection - ploidy and synonyms

From the SNP markers analysis, we found the Swedish Central Collection of mandate cultivars to be comprised of 190 diploid, 24 triploid, and 1 tetraploid accessions, corresponding to 89, 10, and 1 % of all unique samples. Slightly higher frequencies of triploids have recently been reported for Belgian and Danish collections (15 % and 19 %, respectively). Similarly, we found the Swedish central collection to contain 86 % unique accessions, which is in line with the 79 % and 85 % reported for the Belgian and Danish collections, respectively (Larsen et al., 2018; Vanderzande et al., 2017).

When comparing ploidy levels with those obtained during SSR analysis of the same cultivars, but from the local clonal archives, we found five inconsistent cases. For one such case, 'Skälbyäpple', the sample analyzed from the local clonal archive was triploid and genetically identical with 'Gravensteiner', and accordingly a different source for budwood was used, which had not been analyzed prior to propagation. In the Central Collection, 'Skälbyäpple' was found to be morphologically similar to 'Cox's Orange Pippin'(Nilsson, 1987), while also having estimated global pairwise-IBD to other accessions that are in line with the pedigree of 'Cox's Orange Pippin'. According to pomological description 'Skälbyäpple' appears to be quite similar to 'Cox's Orange Pippin', thus it appears likely that 'Skälbyäpple' is synonymous with 'Cox's Orange Pippin'. The other accessions with inconsistent ploidy levels identified here were 'Bosebo Sötäpple', 'Cellini', 'Holländare' and 'lPomme de Cannelle' (Table 2). Similar to 'Skälbyäpple', 'lPomme de Cannelle' was propagated from source not analyzed by SSR markers, thus explaining the inconcistency. However, none of the other three samples appears to be synonymous with any other accession in the Central Collection, and this inconsistency thus requires further investigation.

The consistency between the use of SNP array data and flow cytometry to determine ploidy levels has previously been established (Chagné et al., 2015), making incorrect calling of ploidy levels unlikely. Consequently, the accessions from both the local clonal archives and the central collection should be analyzed using the same genetic markers, to clarify these inconsistencies. For example, 'Cellini' has previously been described as diploid (Urrestarazu et al., 2016), making it likely that the 'Cellini' accession present in the central collection is not true-to-type. Interestingly, although the 'Cellini' sample might not be true-to-type, it could still be a valuable part of the collection, as it does not appear to be synonymous with anything else in the collection. Similarly, 'Bergianäpple' has historically been sold as 'Antonovka' (Hjalmarsson, 2019), but it is now recognized as a separate cultivar with, according to our data, no close relationship to 'Antonovka' (IBD = 0.14). Regarding the synonymous samples, for nine out of the 24 pairs, SSR data were lacking for at least one of the accessions. This illustrates the usefulness of the current thorough examination of the collection, and complements the SSR-based information acquired during the establishment of the gene bank. Some of the duplicate pairs/trios reported here are either duplicates from different sources, e.g., 'Gyllenkroks Astrakan' (EPS) and 'Gyllenkroks Astrakan' (mother tree), or well-known sports such as 'Sävstaholm' and 'P.J. Bergius'. The remaining duplicate pairs would benefit from further genetic investigation of the material in the local clonal archives and from morphological studies, in order to establish whether they are synonymous accessions, mixed-up trees, or sports. For example, 'Hedemoraäpple' appears to be synonymous to 'Grågylling' in the Swedish Central Collection and these two cultivars are described as having rather similar morphology, but 'Hedemoraäpple' is not described as having the twisted stem that is characteristic of 'Grågylling' (Nilsson, 1987). The accessions of 'Rödgylling' from the local clonal archives also had the same SSR profiles as 'Hedemoraäpple' and some accessions labeled as 'Grågylling'. According to lore, 'Hedemoraäpple' is said to originate from a seedling, and thus it would be interesting to investigate whether it is a mix-up or a stem morphology sport, or whether the 'Hedemora' tree investigated previously was not old enough to express the twisted stem trait clearly. Another example is 'Björnegårdsäpple', which is described as "particularly valuable", with a taste and aroma reminiscent of both pineapple and wild strawberries (Nilsson, 1987). In both the Swedish Central Collection and the local clonal archive analyzed, it seems to have been lost and overtaken by shoots from the rootstock. Thus, a new source has to be found for this cultivar through inventory of sources closer to the supposed origin, preferably the mother tree.

4.3. 1st degree relationships

4.3.1. Parent-offspring test

Using a parent-offspring test for Mendelian inconsistent errors, we were able to reject some stipulated relationships. For example, 'Kaniker' was speculated by the early 1900s' pomologist Carl Dahl to be the parent of 'Oranie' (Dahl, 1943), but according to our current SNP marker analysis these cultivars clearly did not share a parent-offspring relationship. On the other hand, the data suggest that they might both be the offspring of 'Gimmersta', and their pairwise IBD estimates indicate either a half-sibling or a grandparent-offspring relationship (0.28). According to tradition 'Flädie' is supposed to originate from a seed of a 'Gravensteiner' apple (Dahl, 1943), which is not supported by our data. However, these cultivars appear to share a cryptic relationship (IBD = 0.37). Interestingly, Larsen et al. (2017) also reported a complete lack of parent-offspring relationships involving 'Gravensteiner', despite some suggested by pomologies. Yet another example that has been the cause of speculation is 'Sparreholm'. It was discovered as a pippin by the gardener L. G. Hedlund, who believed it to be the offspring of 'Rosenhäger' and an astrakhan-type cultivar (Dahl, 1943). According to our results, this speculation was correct, as it was identified as the possible offspring of 'Rosenhäger' and 'Gimmersta', the latter being of astrakhan type. More recently, Garkava-Gustavsson et al. (2008) speculated that 'Oranie' could be one of the parents of 'Hanaskogsäpple', based on SSR profiles, which our data seem to support. Similarly, 'Kabbarp' and 'Vittsjöäpple' were found to be genetically similar according to our SSR-analysis. Accordingly, our SNP data suggest that they are likely full siblings (IBD = 0.6), with 'Gimmersta' as one of the parents, as supported by the Mendelian inconsistent errors test.

4.3.2. Possible parent-parent-offspring combinations

From the analysis of the Swedish Central Collection we identified several possible parent-parent-offspring combinations, all of which seem plausible based on historical records of the cultivar ages. All accessions involved were diploid, except for the triploid 'Ullströmsäpple' which is a possible parent of 'Fredriksdalsäpple' (Table 4, Supplementary File 1). Triploid apples perform much better as pollen donors than as mothers (Sato et al., 2007), so it is reasonable to assume that 'Ullströmsäpple' is the father. However, the procedure used here calls triploids only as heterozygous, regardless of their heterozygous state, and some proportion of true heterozygous SNPs can be expected not to be called, as they will have signals outside the heterozygous clusters. Thus, further clarification of putative parent-offspring relationships would clearly benefit from automated SNP calling of triploid samples, allowing more robust integration of triploids with diploid datasets.

4.3.3. Possible parent-offspring combinations

The most common possible parents were all old cultivars of unknown origin, e.g., 'Gimmersta' (29 P-Os), 'Grågylling' (19 P-Os), 'Vitgylling' (16 P-Os), 'Rosenhäger' (7 P-Os) and 'Klockhammarsäpple' (6 P-Os). Thus 'Gimmersta', a sparsely spread local cultivar considered to be related to an astrakhan-type cultivar (Nilsson, 1987), surprisingly emerged as having the highest number of possible parent-offspring relationships in the collection. At the same time, a very old cultivar which was previously among the most common grown in central Sweden, 'White Astrachan' (Nilsson, 1987), is currently missing from the Central Collection. In our SSR analysis of local archives, 'Gimmersta' was found to be identical to samples of 'White Astrachan', as well as the not true-to-type accession of 'Arvidsäpple' preserved in the breeding collection in Balsgård. The accession of 'Arvidsäpple' preserved at Balsgård has previously been found to be synonymous to 'Astrakan Bilyi' (Urrestarazu et al., 2016). In our analysis of SNP data from the Swedish central collection, 'Stäringe Karin' was found to be the offspring of 'Gimmersta' x 'Grågylling', whereas Muranty et al. (2020) found the same cultivar to be the offspring of 'White Astrachan' x 'Grågylling'. While 'Gimmersta' is described as having red fruit over color (Nilsson, 1987) a plausible explanation is that 'Gimmersta' is in fact a sport of 'White Astrachan', which would then explain the rather large number of possible parent-offspring relationships.

4.3.4. Residual putative 1st degree relationships

Two cultivars, 'Sköldinge' and 'Alnarps Favorit', had far more putative 1st degree relationships than possible parent-offspring relationships. Many of the 1st degree relationships suggested by IBD values, but not passing the test threshold for Mendelian inconsistent errors, involved triploid cultivars or occurred within the network of 1st degree relationships. Due to their higher heterozygosities, triploids are expected to have higher IBD using the method-of-moments technique. Despite this, a few recorded heteroploid relationships were confirmed by the test for Mendelian inconsistent errors ('Alfa 68′ - 'Filippa' and 'Alnarp Favorit' - 'Alfa 68′).

Remaining couples (those not involving 'Sköldinge' and 'Alnarps Favorit') with putative 1st degree relationships, but failing the test threshold for Mendelian inconsistent errors, generally had quite high IBD values (up to 0.6). However, they also had Z0 parameters above 0.1 (Supplementary File 1), illustrating that the method-of-moments technique implemented in PLINK generally inflates IBD values (Morrison, 2013). The structure of the Swedish Central Collection is similar to the Danish collection, in that an equal proportion of the collection belongs to a single network of parent-offspring relationships and has a similar distribution of 1st degree relationships (Larsen et al., 2017).

4.4. Genetic structure

The three first principal components explained a greater proportion of the genetic variation (9.0 and 6.15 % for the first two dimensions) than in two recent studies with similar numbers of cultivars that used GBS (2.87 and 2.47 %) or the 8 K apple SNP array (5.9 and 4.5 %) for genotyping (Larsen et al., 2018; Vanderzande et al., 2017). No population structure was evident from the PCA analysis, even though the dataset was not pruned for markers in LD. As reviewed by Miller and Gross (2011), perennial fruit crops generally exhibit limited population structure due to their life history traits such as clonal propagation, outcrossing, and long juvenile periods. Thus, as the Swedish Central Collection was established with the aim of preserving mainly local cultivars, a lack of population structure is to be expected. Comparable studies have found other European collections to consist of different subpopulations. However, they also contain modern cultivars (Lassois et al., 2016; Marconi et al., 2018; Pereira-Lorenzo et al., 2017; Vanderzande et al., 2017) and wild material (Larsen et al., 2018), in addition to local cultivars. An investigation of Norwegian *ex situ* collections found separate clustering between cultivars from southern countries and traditional Scandinavian or very winter-hardy cultivars (Gasi et al., 2016). Similar clustering patterns have been proposed elsewhere (Garkava-Gustavsson et al., 2013; Urrestarazu et al., 2016), but as the Swedish Central Collection is well curated, containing only mandate cultivars, little clustering is expected. From a genetic resource point of view, a lack of population structure indicates that allelic variants underlying e.g., adaptation to a northern climate are present in an otherwise diverse genetic background, making it a promising germplasm to identify those underlying regions. Based on our PCA analysis, the cultivars released from the Swedish National breeding program at Balsgård seem to encompass only a small fraction of the genetic variation available in the mandate cultivar collection along the two first principal components. Furthermore, a majority of the parents of those cultivars are not represented in the collection (Fig. 2). The Swedish central collection can therefore be expected to contain considerable variation that can invigorate future breeding efforts. Based on the clustering of cultivars released from the Balsgård breeding program, diploid cultivars with coordinates below zero along principal component one and above zero along principal component two might serve to broaden the genepool (upper-left part of Fig. 2). Considering cultivation in northern Sweden as a potential breeding goal for the Swedish breeding program, the cultivars 'Gubbäpple', 'Risäter' and 'Suislepp' as well as 'Förlovningsäpple' and 'Sundsäpple' might be interesting for crosses. 'Gubbäpple' is described as having big fruits similar to 'Gravensteiner', with a balanced sweetness and acidity as well as crispy flesh and an aroma reminiscent of almonds. 'Risäter' is described primarily suited as a cooking apple but is hardy, resistant to diseases and having broad branch angles. 'Suislepp' is a cultivar of Estonian origin, which is considered to combine excellent hardiness and a good, aromatic, taste. 'Förlovningsäpple' and 'Sundsäpple' are very briefly described in the literature, but originate from the northern part of Sweden and can thus be expected to be adapted to the climate (Hjalmarsson, 2019; Nilsson, 1987; Svensson, 2005). To our knowledge, these are not the recorded parents of any of the cultivars released from the Swedish breeding program in Balsgård, which has mainly been directed at commercial apple production in Sweden's most southern province Scania, but this does not exclude the possibility that these cultivars have been used in unsuccessful crosses.

5. Conclusions

SSR markers were useful for first screening of the material preserved in local clonal archives and the collection at SLU-Balsgård and the Swedish Elite Plant Station in order to pinpoint putative identification mistakes as well as providing guidance in choosing the best possible source of graft wood for the Swedish Central Collection. The later has subsequently been analyzed using a subset of 10,368 SNPs from the 20 K SNP array. Thus, we obtained robust information from characterization of the Swedish central collection of mandate cultivars. A few, previously unknown, clonal relationships were identified and these might merit revision of cultivar names. It was also possible to identify putative ploidy levels and possible parent-offspring relationships. Notably, a few cases were identified where the SNP data were in conflict with data obtained from the presumed graft-wood source using SSR markers. In order to guarantee the true-to-typeness and genetic integrity of the Swedish Central Collection and its congruence with local clonal archives, these conflicting cases should be investigated further to identify the sources of the errors. Furthermore, the collection has continued to grow since the onset of this study, and thus molecular-based efforts should continue. Despite considerable phenotypic variation in traits related to hardiness and diverse geographic origins, the collection exhibits limited genetic structure. This information is expected to help improving the future development of the Swedish Central Collection and to facilitate use of the material in breeding programs.

CRediT authorship contribution statement

Jonas Skytte af Sätra: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Michela Troggio:** Methodology, Validation, Investigation, Resources, Writing - review & editing. **Firuz Odilbekov:** Methodology, Validation, Investigation, Resources, Data curation, Writing - review & editing. **Jasna Sehic:** Methodology, Validation, Investigation. **Helena Mattisson:** Methodology, Validation, Investigation. **Inger Hjalmarsson:** Conceptualization, Validation, Resources, Writing - original draft, Writing - review & editing. **Pär K. Ingvarsson:** Methodology, Validation, Writing - review & editing, Supervision. **Larisa Garkava-Gustavsson:** Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are very grateful to the curators of the Swedish local clonal archives and Head of the Swedish Elite Plant Station for providing plant material as well as MSc Anna Zborowska for assistance with compiling the SSR data and accompanying cultivar information. We would also like to thank Caroline Denancé for providing MUNQ genotype codes. Financial support was obtained from the Swedish Board of Agriculture and the GroGrund "Accelerated and Cost efficient cultivar development through genome informed apple breeding". The data analysis benefitted from the Udda server, Department of Plant Breeding, at the LTV Faculty, SLU.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.scienta.2020.109599.

References

Allard, A., Bink, M.C., Martinez, S., Kelner, J.J., Legave, J.M., di Guardo, M., Di Pierro, E.A., Laurens, F., van de Weg, E.W., Costes, E., 2016. Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. J. Exp. Bot. 67, 2875–2888. https://doi.org/10.1093/jxb/erw130.

Bianco, L., Cestaro, A., Sargent, D.J., Banchi, E., Derdak, S., Di Guardo, M., Salvi, S., Jansen, J., Viola, R., Gut, I., Laurens, F., Chagné, D., Velasco, R., van de Weg, E., Troggio, M., Chagne, D., Velasco, R., van de Weg, E., Troggio, M., 2014. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (Malus × domestica Borkh). PLoS One 9, e110377. https://doi.org/10.1371/journal.pone.0110377.

Chagné, D., Crowhurst, R.N., Troggio, M., Davey, M.W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R.P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J.D., Iezzoni, A., Mockler, T., Wilhelm, L., Van de Weg, E., Gardiner, S.E., Bassil, N., Peace, C., 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One 7, e31745. https://doi.org/10.1371/journal.pone.0031745.

Chagné, D., Kirk, C., Whitworth, C., Erasmuson, S., Bicknell, R., Sargent, D.J., Kumar, S., Troggio, M., 2015. Polyploid and aneuploid detection in apple using a single nucleotide polymorphism array. Tree Genet. Genomes 11. https://doi.org/10.1007/s11295-015-0920-8.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7. https://doi.org/10.1186/s13742-015-0047-8.

Csardi, G., Nepusz, T., 2006. The Igraph Software Package for Complex Network Research.

Dahl, C.G., 1943. Pomologi : Beskrivningar Över De Viktigaste I Sverige Odlade Fruktsorterna. D. 1 Äpplen, 2nd ed. Bonnier, Stockholm.

Di Guardo, M., Micheletti, D., Bianco, L., Koehorst-van Putten, H.J., Longhi, S., Costa, F., Aranzana, M.J., Velasco, R., Arus, P., Troggio, M., van de Weg, E.W., 2015. ASSIsT: an automatic SNP scoring tool for in- and outbreeding species. Bioinformatics 31, 3873–3874. https://doi.org/10.1093/bioinformatics/btv446.

Di Guardo, M., Bink, M.C.A.M., Guerra, W., Letschka, T., Lozano, L., Busatto, N., Poles, L., Tadiello, A., Bianco, L., Visser, R.G.F., van de Weg, E., Costa, F., 2017. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. J. Exp. Bot. 68, 1451–1466. https://doi.org/10.1093/jxb/erx017.

Di Pierro, E.A., Gianfranceschi, L., Di Guardo, M., Koehorst-van Putten, H.J., Kruisselbrink, J.W., Longhi, S., Troggio, M., Bianco, L., Muranty, H., Pagliarani, G., Tartarini, S., Letschka, T., Lozano Luis, L., Garkava-Gustavsson, L., Micheletti, D., Bink, M.C., Voorrips, R.E., Aziz, E., Velasco, R., Laurens, F., van de Weg, W.E., 2016. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic. Res. 3, 16057. https://doi.org/10.1038/hortres.2016.57.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361. https://doi.org/10.1007/s12686-011-9548-7.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study. Mol. Ecol. 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

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. https://doi.org/10.1080/13235818.1996.10673675.

Falush, D., Stephens, M., Pritchard, J.K., 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7, 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x.

Garkava-Gustavsson, L., Kolodinska Brantestam, A., Sehic, J., Nybom, H., 2008. Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis. Hereditas 145, 99–112. https://doi.org/10.1111/j.0018-0661.2008.02042.x.

Garkava-Gustavsson, L., Mujaju, C., Sehic, J., Zborowska, A., Backes, G.M., Hietaranta, T., Antonius, K., 2013. Genetic diversity in Swedish and Finnish heirloom apple cultivars revealed with SSR markers. Sci. Hortic. (Amsterdam). 162, 43–48. https://doi.org/10.1016/j.scienta.2013.07.040.

Gasi, F., Kanlić, K., Stroil, B.K., Pojskić, N., Asdal, Å., Rasmussen, M., Kaiser, C., Meland, M., 2016. Redundancies and genetic structure among ex situ apple collections in Norway examined with microsatellite markers. HortScience 51, 1458–1462. https://doi.org/10.21273/HORTSCI11212-16.

Gaut, B.S., Díez, C.M., Morrell, P.L., 2015. Genomics and the contrasting dynamics of annual and perennial domestication. Trends Genet. https://doi.org/10.1016/j.tig.2015.10.002.

Heinonen, M., Bitz, L., 2019. How to discover traditional varieties and shape in a national germplasm collection: the case of finnish seed born apples (Malus × domestica borkh.). Sustainability 11, 7000. https://doi.org/10.3390/su11247000.

Hjalmarsson, I. (Ed.), 2019. Äldre äppelsorter för dagens trädgårdar, 1st ed. SLU, Alnarp.

Hjalmarsson, I., n.d. Conservation of fruit genetic resources in Sweden. Acta Hortic EUCARPIA symposium, in press.

Howard, N.P., van de Weg, E., Bedford, D.S., Peace, C.P., Vanderzande, S., Clark, M.D., Teh, S.L., Cai, L., Luby, J.J., 2017. Elucidation of the "Honeycrisp" pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (Malusxdomestica) pedigree-connected SNP data set. Hortic. Res. 4, 17003. https://doi.org/10.1038/hortres.2017.3.

Howard, N.P., Albach, D.C., Luby, J.J., 2018. The identification of apple pedigree information on a large diverse set of apple germplasm and its application in apple breeding using new genetic tools. Ecofruit. 88–91.

Larsen, B., Toldam-Andersen, T.B., Pedersen, C., Ørgaard, M., 2017. Unravelling genetic diversity and cultivar parentage in the Danish apple gene bank collection. Tree Genet. Genomes 13. https://doi.org/10.1007/s11295-016-1087-7.

Larsen, B., Gardner, K., Pedersen, C., Ørgaard, M., Migicovsky, Z., Myles, S., Toldam-Andersen, T.B., 2018. Population structure, relatedness and ploidy levels in an apple gene bank revealed through genotyping-by-sequencing. PLoS One 13. https://doi.org/10.1371/journal.pone.0201889.

Lassois, L., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Hibrand-Saint-Oyant, L., Poncet, C., Lasserre-Zuber, P., Feugey, L., Durel, C.E., 2016. Genetic diversity, population structure, parentage analysis, and construction of core collections in the french apple germplasm based on SSR markers. Plant Mol. Biol. Report. 34, 827–844. https://doi.org/10.1007/s11105-015-0966-7.

Marconi, G., Ferradini, N., Russi, L., Concezzi, L., Veronesi, F., Albertini, E., 2018. Genetic characterization of the apple germplasm collection in central Italy: the value of local varieties. Front. Plant Sci. 9, 1460. https://doi.org/10.3389/fpls.2018.01460.

Miller, A.J., Gross, B.L., 2011. From forest to field: perennial fruit crop domestication. Am. J. Bot. 98, 1389–1414. https://doi.org/10.3732/ajb.1000522.

Morrison, J., 2013. Characterization and correction of error in genome-wide ibd estimation for samples with population structure. Genet. Epidemiol. 37, 635–641. https://doi.org/10.1002/gepi.21737.

Muranty, H., Denancé, C., Feugey, L., Crépin, J.L., Barbier, Y., Tartarini, S., Ordidge, M., Troggio, M., Lateur, M., Nybom, H., Paprstein, F., Laurens, F., Durel, C.E., 2020. Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. BMC Plant Biol. 20, 2. https://doi.org/10.1186/s12870-019-2171-6.

Nilsson, A., 1987. Våra äpplesorter, 2:1. ed. Allmänna Förlaget, Stockholm.

Nybom, H., 2019. Apple production and breeding in Sweden. Chron. Horticult. 59, 21–25.

Ordidge, M., Kirdwichai, P., Fazil Baksh, M., Venison, E.P., George Gibbings, J., Dunwell, J.M., 2018. Genetic analysis of a major international collection of cultivated apple varieties reveals previously unknown historic heteroploid and inbred relationships.

PLoS One 13. https://doi.org/10.1371/journal.pone.0202405.
Patocchi, A., Fernández-Fernández, F., Evans, K., Gobbin, D., Rezzonico, F., Boudichevskaia, A., Dunemann, F., Stankiewicz-Kosyl, M., Mathis-Jeanneteau, F., Durel, C.E., Gianfranceschi, L., Costa, F., Toller, C., Cova, V., Mott, D., Komjanc, M., Barbaro, E., Kodde, L., Rikkerink, E., Gessler, C., Van De Weg, W.E., 2009. Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. Tree Genet. Genomes 5, 211–223. https://doi.org/10.1007/s11295-008-0176-7.
Peakall, R., Smouse, P.E., 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x.
Peakall, R., Smouse, R.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460.
Peleman, J.D., Van Der Voort, J.R., 2003. Breeding by design. Trends Plant Sci. 8, 330–334. https://doi.org/10.1016/S1360-1385(03)00134-1.
Pereira-Lorenzo, S., Urrestarazu, J., Ramos-Cabrer, A.M., Miranda, C., Pina, A., Dapena, E., Moreno, M.A., Errea, P., Llamero, N., Díaz-Hernández, M.B., Santesteban, L.G., Laquidain, M.J., Gogorcena, Y., Urbina, V., Dalmases, J., Ascasíbar-Errasti, J., Royo, J.B., 2017. Analysis of the genetic diversity and structure of the Spanish apple genetic resources suggests the existence of an Iberian genepool. Ann. Appl. Biol. 171, 424–440. https://doi.org/10.1111/aab.12385.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.
Rohlf, 2005. NTSYSpc Numerical Taxonomy and Multivariate Analysis System.
Routson, K.J., Reilley, A.A., Henk, A.D., Volk, G.M., 2009. Identification of historic apple trees in the Southwestern United States and implications for conservation. HortScience 44, 589–594. https://doi.org/10.21273/HORTSCI.44.3.589.
Sato, M., Nyui, T., Takahashi, H., Kanda, H., 2007. Comparison of flowering and fruiting of seedlings from reciprocal crosses between diploid and triploid apple cultivars. J. Japan. Soc. Hort. Sci.
Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 38, 1409–1438.
Svensson, H., 2005. Äpplen I Sverige, 2nd ed. Bokförlaget Prisma, Stockholm.
Tange, O., 2011. GNU Parallel - the Command-line Power Tool. https://doi.org/10.5281/zenodo.16303.
Urrestarazu, J., Denance, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Poncet, C., Lateur, M., Houben, P., Ordidge, M., Fernandez-Fernandez, F., Evans, K.M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Miranda, C., Gassmann, J., Kellerhals, M., Suprun, I., Pikunova, A.V., Krasova, N.G., Torutaeva, E., Dondini, L., Tartarini, S., Laurens, F., Durel, C.E., 2016. Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. BMC Plant Biol. 16, 130. https://doi.org/10.1186/s12870-016-0818-0.
Urrestarazu, J., Muranty, H., Denancé, C., Leforestier, D., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Aubourg, S., Celton, J.M., Daccord, N., Dondini, L., Gregori, R., Lateur, M., Houben, P., Ordidge, M., Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Troggio, M., Bianco, L., Velasco, R., Poncet, C., Théron, A., Moriya, S., Bink, M.C.A.M., Laurens, F., Tartarini, S., Durel, C.E., 2017. Genome-wide association mapping of flowering and ripening periods in apple. Front. Plant Sci. 8, 1923. https://doi.org/10.3389/fpls.2017.01923.
van de Weg, E., Di Guardo, M., Jänsch, M., Socquet-Juglard, D., Costa, F., Baumgartner, I., Broggini, G.A.L., Kellerhals, M., Troggio, M., Laurens, F., Durel, C.-E., Patocchi, A., 2018. Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. Mol. Breed. 38. https://doi.org/10.1007/s11032-017-0755-0.
van Treuren, R., Kemp, H., Ernsting, G., Jongejans, B., Houtman, H., Visser, L., 2010. Microsatellite genotyping of apple (Malus × domestica Borkh.) genetic resources in the Netherlands: application in collection management and variety identification. Genet. Resour. Crop Evol. 57, 853–865. https://doi.org/10.1007/s10722-009-9525-0.
Vanderzande, S., Micheletti, D., Troggio, M., Davey, M.W., Keulemans, J., 2017. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. Tree Genet. Genomes 13. https://doi.org/10.1007/s11295-017-1206-0.
Vanderzande, S., Howard, N.P., Cai, L., Da Silva Linge, C., Antanaviciute, L., Bink, M.C.A.M., Kruisselbrink, J.W., Bassil, N., Gasic, K., Iezzoni, A., Van de Weg, E., Peace, C., 2019. High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. PLoS One 14, e0210928. https://doi.org/10.1371/journal.pone.0210928.

III

Tree Genetics & Genomes (2023) 19:12
https://doi.org/10.1007/s11295-023-01587-w

ORIGINAL ARTICLE

Parametric mapping of QTL for resistance to European canker in apple in 'Aroma'×'Discovery'

Jonas Skytte af Sätra[1] · Firuz Odilbekov[1,2] · Pär K. Ingvarsson[3] · Eric van de Weg[4] · Larisa Garkava-Gustavsson[1]

Received: 28 October 2022 / Revised: 21 December 2022 / Accepted: 25 January 2023

Abstract
Resistance to European canker (*Neonectria ditissima*) in apple is currently one of the most important breeding targets for commercial production in Sweden. Previous research has identified significant genetic variation in susceptibility to the disease, with the local Swedish cultivar 'Aroma' considered as one of the most resistant cultivars. Identification of genetic regions underlying the resistance of this cultivar would be a valuable tool for future breeding. Thus, we performed Bayesian quantitative trait loci (QTL) mapping for resistance to European canker in a full-sib family of 'Aroma' × 'Discovery'. Mapping was performed with the area under the disease progression curves (AUDPCs) from all seven (AUDPC_All7) and the first four assessments (AUDPC_First4), and three parameters of a sigmoid growth model for lesion length. As a scale for the effect of the different parameters, historic phenotypic data from screenings of a genetically diverse germplasm was compiled and re-analyzed. The parametrization of the data on lesion growth increased the number of QTL that could be identified with high statistical power, and provided some insight into their roles during different stages of disease development in the current experimental setup. Five QTL regions with strong or decisive evidence were identified on linkage groups 1, 8, 15, and 16. The QTL regions could be assigned to either of the parameters lesion length at the first assessment ('LL_A1'), the maximal lesion growth rate (lesion length doubling time, 't_gen'), and the lesion length at girdling ('LL_G'). Three of these QTL were traced along the pedigrees of some known relatives of the FS family, and discussed in relation to future crosses for breeding and genetic research.

Keywords *Neonectria ditissima* · *Malus domestica* · FlexQTL · Sigmoid growth · IBD

Introduction

European canker, caused by the fungus *Neonectria ditissima*, is currently one of the most serious threats to Swedish commercial apple (*Malus domestica* Borkh.) production. The fungus can infect trees through natural or artificial wounds throughout the year and is favored by cool and wet climates, such as North-West Europe, New Zealand, and Chile (Weber 2014). Infected wounds develop into necrotic lesions that may girdle branches and the stem, thereby killing the distal parts of the shoot or the entire tree. Apart from yield losses caused by reduced bearing capacity, the fungus can also infect fruits and thereby cause storage losses (Brown et al. 1994; Swinburne 1975).

N. ditissima commonly colonizes the phloem and xylem of the apple stem and then continues to spread until it reaches a defensive mature suberized cell layer (Krähmer 1980). However, beyond the importance of suberization, little is known about the defense mechanisms against *N. ditissima* in apple. The disease is managed by the removal of infected tissues, cultivation of resistant cultivars, and application of fungicides (Weber 2014). As applications of fungicides are limited in Sweden and pruning of cankered wounds is time consuming, there is a strong need

Communicated by D. Chagné.

✉ Jonas Skytte af Sätra
jonas.skytte.af.satra@slu.se

[1] Department of Plant Breeding, Swedish University of Agricultural Sciences, Lomma, Sweden

[2] Lantmännen Lantbruk, Svalöv, Sweden

[3] Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[4] Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands

for locally adapted cultivars with improved resistance to *N. ditissima*.

Despite the importance of the disease, there is currently no consensus on how to screen for resistance. Reported approaches differed in type of plant material (trees, detached shoots), cultivation (planted, potted, outdoors, or in a glasshouse), sources of inoculum (e.g., wild collected vs single-spore isolate), type of isolates (wild collected and in vitro propagated single-spore isolates), and type of spores (ascospores, conidia) as well as wounding techniques (e.g., naturally occurring leaf scars or manually inflicted by rasp or scalpel) (Bus et al. 2019; Delgado et al. 2022; Garkava-Gustavsson et al. 2016; Gómez-Cortecero et al. 2016; Karlström et al. 2022; van de Weg 1989; Wenneker et al. 2017). Also, various parameters to quantify resistance have been used, including area under the disease progression curve (AUDPC), infection rate, the slope of the lesion growth regression line, lesion length at a given timepoint, and indexes quantifying the establishment of secondary infections (Bus et al. 2019; Garkava-Gustavsson et al. 2016; Gómez-Cortecero et al. 2016; Karlström et al. 2022; Wenneker et al. 2017). Different parameters might capture the effect of different resistance mechanisms depending on the timing of assessment in relation to the disease development of the experiment, due to variations in, e.g., isolate virulence, the conductivity of the experimental procedure to infection, physiological stage of the trees, and environmental conditions affecting disease progression (e.g., temperature). This diversity might account for variation in cultivar rankings in screening experiments as well as different outcomes in quantitative trait loci (QTL) mapping and validation studies. While likely to be of relevance for field conditions, assessment of colonization and infection rates might not be feasible for QTL mapping studies with large populations as it requires a high number of inoculation points (trees). In cases when repeated measurements have been made, the AUDPC can be used to average out the data on disease progression over time. This can help reduce the noise in the phenotypic data, as compared to using data from individual assessments. However, AUDPC does not provide information on the temporal dynamics of lesion growth. In the *M. domestica – N. ditissima* pathosystem, the lesion growth rate based on a linear regression over a defined time period has been proposed as a reproducible parameter of disease resistance (Wenneker et al. 2017). Parameters of non-linear models such as exponential-, logistic-, and Gompertz curves have been used successfully for functional QTL identification of growth in, e.g., *Populus* (Wu et al. 2003) and *Arabidopsis* (Bac-Molenaar et al. 2015), and provide information that has direct implications for the understanding of a trait's genetic architecture. Employing such models to disease progression in the *N. ditissima – M. domestica* pathosystem might improve our understanding of the functional characteristics of different mapped QTL, thereby enabling a better understanding of underlying candidate genes, as well as facilitating better integration of results from different experiments.

Previous studies have shown that cultivars such as 'Aroma', 'Santana', and 'Golden Delicious' are partially resistant, while cultivars such as 'Discovery', 'Elise', and 'Katja' are highly susceptible to *N. ditissima* (Garkava-Gustavsson et al. 2016; van de Weg 1989; Wenneker et al. 2017). QTL for resistance to *N. ditissima* in apple have been mapped in a number of studies, e.g., 'Scired' x A045R14T055, 'M9' × 'Robusta 5', 'Jonathan' × 'Discovery', and 'Golden Delicious' × 'Discovery' (Bus et al. 2021; Bus et al. 2019; van de Weg et al. 2020) and recently in a set of five families including 'Aroma' as a parent of one full-sib (FS) family (Karlström et al. 2022). The cultivars 'Aroma' and 'Discovery' are both important parents and founders in the Nordic apple breeding programs and mapping of QTL for resistance to *N. ditissima* in these varieties is of high priority for the Swedish breeding program. Previous studies have indicated a complex inheritance of resistance to *N. ditissima* in *M. domestica* (Gómez-Cortecero et al. 2016; Karlström et al. 2022). Therefore, a Bayesian approach is likely better suited to QTL analysis than conventional interval-mapping, as it allows for simultaneous estimation not only for the position of the largest QTL, but for the number of QTL, the additive effect sizes, and credible interval of all QTL (Bink et al. 2014, 2008; van de Weg et al. 2018). Once QTL regions have been mapped, tracing the inheritance of the mapped loci through pedigrees can provide valuable information for further breeding and genetic research. This can be facilitated through calculation of Identity-by-Decent (IBD) probabilities of the founder haplotypes in a pedigreed germplasm. This can provide information on the presence of QTL regions in specific individuals and recombination events in the founder haplotypes, which can be valuable for increasing the applicability of the results.

The purpose of this study was to gain insight into the genetic control of partial resistance to *N. ditissima* in apple. To this end, an FS family from 'Aroma' × 'Discovery' was phenotyped, genotyped, and analyzed through Bayesian QTL mapping using two AUDPC parameters as well as parameters of a sigmoid growth model of lesion length. Finally, the use of the most relevant QTL and related parameters were evaluated through Identity-by-State (IBS) and IBD approaches, using historic phenotypic data compiled from previous cultivar screening experiments.

Materials and methods

Plant material

For QTL discovery, an FS family consisting of 172 offspring from an 'Aroma' × 'Discovery' cross, was phenotyped in two independent experiments over two consecutive years (2019 and 2020), together with their parents and six reference cultivars 'Cox's Orange Pippin', 'Prima', 'Jonathan', 'Katja', 'Golden Delicious', and 'Santana'. In the second year of phenotyping, two grandparents ('Filippa' and 'Ingrid Marie') and three half-siblings ('Fredrik', 'Frida', and 'Julia') of the FS family were included. To trace QTL intervals detected in the FS family and to provide a scale for the disease resistance parameters used, we also compiled and reanalyzed phenotypic data from 98 genotypes that had been screened previously. These cultivar screening experiments were performed over 5 years (2011–2016), with seven reference cultivars being included in all experiments: 'Aroma', 'Cox's Orange Pippin', 'Discovery', 'Elise', 'Golden Delicious', 'Jonathan', and 'Santana' (File S1). Results for a subset of 15 cultivars from the screening germplasm were published previously (Garkava-Gustavsson et al. 2016) using various parameters for estimation of resistance.

Phenotyping

Phenotyping of the FS family and a set of reference cultivars as controls was performed essentially as described previously (Garkava-Gustavsson et al. 2016). Each autumn, new trees grown in 3L pots were obtained, which had been grafted on rootstock 'M9' in the spring of the same year. A single-spore isolate of *N. ditissima* ('SLU-E1') was propagated, from a spore solution maintained at − 80 °C in 50% glycerol, on 20% Potato Glucose Agar (70,139, Sigma-Aldrich) supplemented with bactoagar (B1000, Saveen o Werner) for 4–6 weeks on a laboratory bench under ambient light and temperature. Three buds (typically the 11:th, 14:th, and 17:th axillary buds from the apex) were removed with a scalpel (File S1) and the wounds were inoculated with a mixture of micro- and macroconidia (mostly macroconidia with two septae or less) to a total of 1000 conidia in 10 µl water. Once the suspension was absorbed by the plant tissue, the wound was covered with white petrolatum (Special Care Vitt vaselin, ACO Sweden). The petrolatum was removed 5 days after inoculation, using tissue paper. The inoculations were performed block-wise in late autumn (November–December), with all trees of a block being inoculated on the same date. The inoculation dates of the blocks were separated by approximately 1 week, thus leaves were typically fully present at the first inoculation occasion and mostly shed at the fourth and final inoculation occasion. Throughout the trial period, the trees were kept in a glasshouse with minimal heating and no artificial cooling or light, with a temperature fluctuating between 9 and 11 °C until the end of February. From then on, changes in outdoor climate led to increased fluctuations in temperature, ranging between 5 °C (night) and occasionally above 20 °C (day). Lesions were measured using a digital caliper at seven time points, every second week, starting approximately 6 weeks after inoculation when the first visual symptoms had appeared in a meaningful proportion of the block (typically more than half of the wounds having developed lesions at least 5 mm long). Lesions shorter than 5 mm were recorded as '0'. In cases where two or three lesions merged during the trial period, their total length was measured and divided between the lesions according to the relative proportions of their last individual measurement. In case lesions could not be measured because the stems were girdled, thereby strangling the distal parts and preventing further upward growth of the lesion, the lesions were assigned the last true measurement for the remainder of the trial. For the experiments with the FS family, approximately half of the developed lesions girdled. The previous screening experiments had been performed similarly, but using a wild-collected spore solution (mostly with three septae) prepared as described previously (Garkava-Gustavsson et al. 2016) and one year using rootstock 'B9' (File S1). All current and previous experiments were performed using a randomized complete block design with three trees per individual in each block, except for the FS family where each individual was represented by a single tree in each of four blocks for practical reasons (File S1).

Parameters of disease resistance and heritability

For the AUDPC parameter, we initially considered all seven assessments ('AUDPC_All7') as the phenotypic parameter for QTL mapping. Next, to investigate how robust the detected QTL regions were to artifacts introduced by girdling of wounds, we considered the AUDPC from the first to the fourth ('AUDPC_First4') assessment, as most wounds that girdled typically did so after the fourth assessment. For the sigmoid growth curve model we used the parameters initial lesion length ('LL_A1'), the shortest doubling time of the length of the lesion ('t_gen'), and the lesion length at girdling ('LL_G').

For all parameters, wounds that had not developed any lesions at the 6th assessment were treated as missing values. The AUDPC parameters were calculated as:

$$AUDPC = \sum_{i=1}^{n} [\frac{Y_{i+1} + Y_i}{2}][X_{i+1} - X_i] \quad (1)$$

where Y_i is the lesion length at the ith assessment, X_i is the number of days at the ith observation and n is the total number of observations. To estimate the lesion growth rate, a logistic regression was performed over each pseudo-replicate (wound) using the R (R Core Team 2020) package GrowthCurveR (Sprouffske and Wagner 2016), fitting the model:

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}} \quad (2)$$

where N_t is the lesion length at time t, K is the carrying capacity (maximum lesion length reached at girdling of the wound), N_0 is the lesion length at the beginning of the growth curve, and r is the intrinsic growth rate of the lesion, which would occur if there were no restrictions. The generation time (t_gen) is the shortest time it could take for the lesion to double in length under the model:

$$t_{gen} = \frac{ln(2)}{r} \quad (3)$$

where r is the intrinsic growth rate obtained from (2). To improve the ability of the model to fit the growth rate parameter, an artificial time point was added to the assessment data for the logistic regression, corresponding to the time of inoculation, which was given the phenotypic value '0' for all wounds. The regression curves of all wounds with outlier residual standard errors were inspected manually, and wounds with poorly fitting regression curves were excluded from further analysis. The lesion length at the last assessment for all wounds that had girdled was used as the parameter for the maximal lesion length (LL_G), i.e., the plateau of the sigmoid growth curve.

Best linear unbiased estimates (BLUEs) combined over the experiments were used as phenotypes and BLUEs were calculated using the R package 'lme4' (Bates et al. 2015) with Genotype treated as a fixed effect and with Tree nested within Block, and Block in turn nested within Year, as random effects. Wounds with outlier residuals, identified visually from QQ-plots, were investigated manually and excluded from the analysis if the assessment data were problematic.

Broad sense heritability estimates ($\hat{H}^2$) were calculated using the regression method previously described (Schmidt et al. 2019). In short, the model described above was fitted to calculate BLUEs from each data set, and a similar model but with the Genotype as a random effect was used to calculate Best Linear Unbiased Predictors (BLUPs). By regressing BLUPs on BLUEs, the regression coefficient can be used as an approximation of $\hat{H}^2$. A Genotype x Year ('G x Y') interaction term was not included in the models, as this was not considered biologically significant for the current experimental protocol.

The residuals of all parameters (AUDPC_All7, AUDPC_First4, LL_A1, t_gen, and LL_G) showed slight deviations from a normal distribution in both the FS family and the screening population (File S2). Initially, various transformation methods were considered, where Box-Cox transformation gave distributions that were closest to normal. However, preliminary QTL-mapping analysis indicated that Box-Cox-transformed and untransformed data for the AUDPC parameters gave very similar results. In addition, the application of Box-Cox transformation to mixed models is problematic and makes biologically relevant interpretations of BLUEs and, e.g., QTL effects complicated. Thus, untransformed data were used for all QTL analyses.

Genotyping and marker curation

DNA from the FS family and part of the screening population was extracted from lyophilized leaf tissue using the DNeasy 96 Plant Kit (Qiagen) and was used for genotyping with the 20 K apple Infinium® single nucleotide polymorphism (SNP) array (Bianco et al. 2014). Marker data were initially curated at the level of individual SNPs, as described by Vanderzande and Howard et al. (2019), with the modification that a subset of 10 K SNPs retained in a previous study on marker integration (Howard et al. 2021) were called in Genome Studio (GS) v 2.0 (Illumina Inc.), using cluster definition files kindly made available by the authors. For computational efficiency of the QTL analysis in the FS family, the marker set was pruned for SNPs with pairwise r^2 above 0.2 in 2 Mb sliding windows with a 200-kb shift using PLINK 1.9 (Chang et al. 2015) and physical positions on the HFTH1 whole genome sequence (WGS) (Eric van de Weg, unpubl. data). This was done separately for SNPs exhibiting maternal monohybrid segregation, paternal monohybrid segregation, and dihybrid segregation, resulting in a set of 4506 SNPs in approximate linkage equilibrium. Marker data for linkage groups (LGs) with QTL loci in the FS family were further curated as haploblocks (HBs) as described by Vanderzande and Howard et al. (2019) in the screening germplasm, resulting in 370 HBs representing 2348 SNPs on linkage groups (LGs) 1, 8, 15, and 16. Genotypic 20 K SNP array data on some of the accessions in the screening population, and close relatives, was made available from other previous and ongoing projects (Hjeltnes et al. 2019; Skytte af Sätra et al. 2020). For curation as well as subsequent QTL analysis, genetic positions were taken from an advanced draft of a virtual linkage map (the 15K-iGW-map, Eric van de Weg, unpubl. data) resulting from integration of the iGLMap (Di Pierro et al. 2016) and the HFTH1 WGS (Zhang et al. 2019).

QTL analysis

QTL analysis was conducted using a Bayesian approach as embedded in FlexQTL™ (www.flexqtl.nl), which uses Markov chain Monte Carlo (MCMC) simulations and bi-allelic QTL models to evaluate possible combinations of the number of QTL, QTL positions, QTL effects, and their transmission through pedigrees (Bink et al. 2014, 2008). In this study, each parameter was analyzed four times using different seeds and priors for the number of QTL (1 or 3). The maximum number of QTL was set to 15 for all runs, and all runs consisted of at least 100,000 iterations with a thinning of 100 using additive genetic models with normal prior distributions and random (Co) variance matrix diagonals. While the phenotypes of parents are usually not included in QTL analyses, initial runs consistently identified 'Aroma' as the susceptible parent and 'Discovery' as the resistant parent. By adding the phenotypes of the parents to the model, they were correctly identified as resistant/susceptible, and as the same QTL regions were identified with or without the phenotypes of the parents in preliminary FlexQTL runs, these were included for further analysis.

All analyses were run until the effective chain samples exceeded 100, i.e., until convergence. The level of evidence provided for the presence of a QTL was indicated by two times the natural logarithm of the Bayes factors ('2lnBF') for an incremental number of QTL per LG through a pair-wise comparison. A 2lnBF value above 0, 2, 5, and 10 is considered to indicate hardly any, positive, strong, and decisive evidence for the presence of a QTL, respectively. Further analysis was limited to the five QTL with a distinct peaks in posterior position for the AUDPC_All7 parameter. QTL regions were defined as successive 2-cM bins with 2lnBF above 2. For individual QTL regions, the Bayesian probability (B-Prob.) of a QTL is based on the posterior QTL intensity, i.e., the frequency at which any of the 2-cM bins within the QTL region were included in the model. Given the probability of more than one QTL per LG, the Bayesian probability of a QTL region might exceed 1.0 in case multiple genetic bins within the QTL region are frequently included in the model. The proportion of phenotypic variance explained (PVE) by a QTL was calculated using the formula:

$$PVE = \left(\frac{AVt}{V_P}\right) \times 100 \quad (4)$$

where V_P is the total phenotypic variance and AVt is the additive variance explained by the QTL region (from PostQTL analysis). For all QTL regions (Table 1), trace plots of QTL positions for all runs were investigated to confirm convergence for the QTL position, i.e., that the QTL regions were not included in the model alternatingly. In FlexQTL outputs Q and q denotes alleles with increasing and decreasing effects on the phenotype, respectively, and this notation is used throughout this paper. In addition to the model with additive effects only, FlexQTL also supports models with dominance and additive effects. Tests of these models with the AUPDC parameters indicated similar QTL regions as in the additive effects only model, but with very low statistical power and very small estimated dominance effects. Thus, dominance was not considered further for QTL analysis. While there is little documentation accompanying FlexQTL, some previous publications include informative material and methods sections and supplementary material that has been very valuable to the current QTL analysis (Mangandi et al. 2017; van de Weg et al. 2018).

Compound QTL genotypes

QTL intervals were identified as two or more consecutive 2-cM genetic bins with a 2lnBF above five, in all four runs performed for the parameter analyzed. Individuals with recombination events within these bins were excluded from the analyses of compound genotype effects. Segregation and effects of the QTL regions were investigated by fitting a series of linear models using the non-recombinant genotypes in base R with the 'car' an 'lmtest' packages (Fox and Weisberg 2019; R Core Team 2020; Zeileis and Hothorn 2002; File S2). Nested models were evaluated by likelihood ratio tests and non-nested models were compared based on their Akaike information criterion (AIC). The models with the lowest AIC were also evaluated by their Bayesian information criterion (BIC).

Search for epistatic loci

As the Q alleles at the LG1 and LG8 as well as the q alleles at the LG15 and LG16b loci did not have significantly different additive effects, they were grouped into subsets based on their total number of q alleles for AUDPC_First4 across both respective loci. These subset of individuals with 0, 1, 2, or 3 q alleles for LG1/LG8 and LG15/LG16b, respectively, were used to identify potentially epistatic loci segregating in the population. Only individuals without recombination events in consensus QTL regions were used, resulting in subsets consisting of 17, 49, 35, and 10 individuals for the LG1/LG8 loci, and 12, 42, 49, and 12 individuals for the LG15/LG16b loci. A single FlexQTL run was performed for each subset and parameter, using AUDPC_All7, AUDPC_First4, and t_gen for the LG1/LG8 loci, and AUDPC_All7, AUDPC_First4, and LL_A1 for the LG15/LG16b loci. Additionally, subsets of 89 and 71 individuals carrying the q and Q alleles from 'Discovery' at the LG16a locus, respectively, were also used to search for additional epistatic loci for the LL_G parameter.

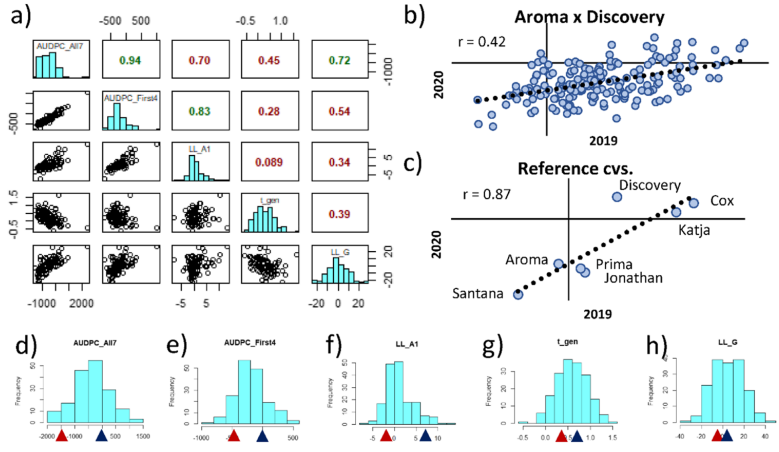


Fig. 1 Distributions of, and correlations between, parameters and years. **a** Correlogram from the screening population with histograms for each parameter in the diagonal, Pearson correlation coefficient (*r*) in the upper off-diagonal, and pairwise scatter plots in the lower left off-diagonal. **b** Correlations between the two years of phenotyping for the QTL mapping in the FS family and **c** the reference cultivars that were common between the 2 years. **d**–**h** Distributions in the FS family for the different parameters where phenotypic values of the parents are indicated by red ('Aroma') and blue ('Discovery') arrows

Prevalence of the segregating haplotypes in the screening germplasm

Genotypes that were IBS to the haplotypes segregating in the FS family were identified from HB genotypes, and individuals having identical haplotype calls along the entire consensus QTL intervals were considered IBS. Next, IBD genotypes for estimation of allele effects were identified as IBS genotypes where the haplotype could be traced through their known pedigree in cultivars with phased genotypic data to the segregating haplotype from three of the FS-family's grandparents 'Filippa', 'Worcester Pearmain', or 'Beauty of Bath' and two of its grand-grandparents, 'Cox's Orange Pippin' and 'Cox's Pomona'. For visualization of QTL haplotype inheritance through the pedigree, the probability of IBD was calculated in FlexQTL and used as input together with AUDPC_First4 BLUEs for visualization in PediMap (Voorrips et al. 2012).

Results

Phenotypic data

The infection rate was above 95% of inoculated wounds in all experiments, except in 2012 (89.4%). All parameters (AUDPC_All7, AUDPC_First4, LL_A1, t_gen, and LL_G) were heteroscedastic and had residual distributions that deviated from normality (File S2). $\hat{H}^2$ was moderate to high for the FS family across both years (0.78, 0.78, 0.67, 0.73, and 0.52 for AUDPC_All7, AUDPC_First4, LL_A1, t_gen, and LL_G, respectively). Similarly, moderate to high $\hat{H}^2$ point estimates were observed for the screening germplasm across all years (0.87, 0.87, 0.62, 0.74, and 0.75 for AUDPC_All7, AUDPC_First4, LL_A1, t_gen, and LL_G, respectively). The two AUDPC parameters, AUDPC_All7 and AUDPC_First4, were highly correlated, while the three parameters relating to the sigmoid growth model were very weakly correlated with each other (Fig. 1a). The year-year correlation for the FS family (represented by 1 tree/block) was low for all parameters (Pearson corr. = 0.43, 0.42, 0.34, 0.54, 0.12, and 0.35 for AUDPC_All7, AUDPC_First4, LL_A1, t_gen, and LL_G, respectively, Fig. 1b), but the correlation between years was moderate to high for the control cultivars included both those years (represented by 3 trees/block) (Pearson Corr. = 0.83, 0.87, 0.90, 0.74, and 0.75 for AUDPC_All7, AUDPC_First4, LL_A1, t_gen, and LL_G, respectively, Fig. 1c). The low correlation between years for the FS-family is likely caused by extensive experimental noise in single-year data when using only one tree per block. As a consequence, differences between years cannot be reliably estimated in the FS family, and the BLUEs over both years were used as phenotypes for QTL mapping.

QTL mapping

First, we considered the AUDPC over all seven assessments (AUDPC_All7), and subsequently we also assessed

Table 1 QTL regions are summarized from four replicate FlexQTL runs for each parameter. For each parameter (Param.) and QTL, the highest and lowest LG-wide 2lnBF* value for one QTL over no QTL, range of genetic bins with strong local evidence in all four runs (Range), the width of the QTL range, the range of the two genetic bins with the highest joint probabilities averaged over all four runs (Peak), average Bayesian probability of the QTL across the four runs (B-Prob.), average PVE over the four replicates, and the parental genotypes with strong evidence. Note that PVE is calculated based on the entire QTL interval, defined as a series of consecutive genetic bins with 2lnBF values above 2, which is wider than the indicated range. In addition, the consensus QTL regions across all parameters are noted (i.e., the outer boundaries of the range and peak across parameters)

Param	QTL	2lnBF*	Range (cM)	Width (cM)	Peak (cM)	B-prob	PVE (%)	'Aroma'	'Discovery'
AUDPC_All7	LG1	5.6–6.1	44–54	10	50–54	0.6	13.4	–	*Q*/*q*
	LG8	6.1–7.3	38–50	12	40–44	0.7	14.0	–	*Q*/*q*
	LG15	5.4–6.2	51–61	10	51–55	0.7	10.9	*Q*/*q*	*Q*/*q*
	LG16a	4.0–4.4	20–22**	–**	20–22**	0.2	8.7	–	*Q*/*q*
	LG16b		48–50**	–**	48–50**	0.2	9.8	–	*Q*/*q*
AUDPC_First4	LG1	5.8–6.5	40–52	12	44–48	0.7	11.8	–	*Q*/*q*
	LG8	10.6–14.0	38–50	12	46–50	1.0	14.0	–	*Q*/*q*
	LG15	10.9–12.4	51–57	6	51–55	1.0	13.6	*Q*/*q*	*Q*/*q*
	LG16a	–	–	–	–	–	–	–	–
	LG16b	8.7–9.8	44–50	6	46–50	0.9	16.1	–	*Q*/*q*
LL_A1	LG1	–	–	–	–	–	–	–	–
	LG8	2.1–2.4	50–52**	–**	50–52**	0.2	8.7	–	–
	LG15	8.0–9.7	49–57**	6**	53–57**	1.0	14.0	*Q*/*q*	–
	LG16a	–	–	–	–	–	–	–	–
	LG16b	6.1–7.8	38–50	12	46–50	0.7	12.1	–	*Q*/*q*
t_gen	LG1	3.4–3.7	50–54**	4	50–54	0.4	9.8	–	*Q*/*q*
	LG8	9.2–9.7	42–52	10	46–50	1.0	13.0	–	*Q*/*q*
	LG15	–	–	–	–	–	–	–	–
	LG16a	–	–	–	–	–	–	–	–
	LG16b	–	–	–	–	–	–	–	–
LL_G	LG1	–	–	–	–	–	–	–	–
	LG8	–	–	–	–	–	–	–	–
	LG15	–	–	–	–	–	–	–	–
	LG16a	5.3–6.2	20–24	4	20–24	0.6	13.9	–	*Q*/*q*
	LG16b	–	–	–	–	–	–	–	–
Consensus	LG1	–	40–54	14	44–54	–	–	–	–
	LG8	–	38–52	14	40–50	–	–	–	–
	LG15	–	51–61	10	51–55	–	–	–	–
	LG16a	–	20–24	4	20–24	–	–	–	–
	LG16b	–	38–50	12	46–50	–	–	–	–

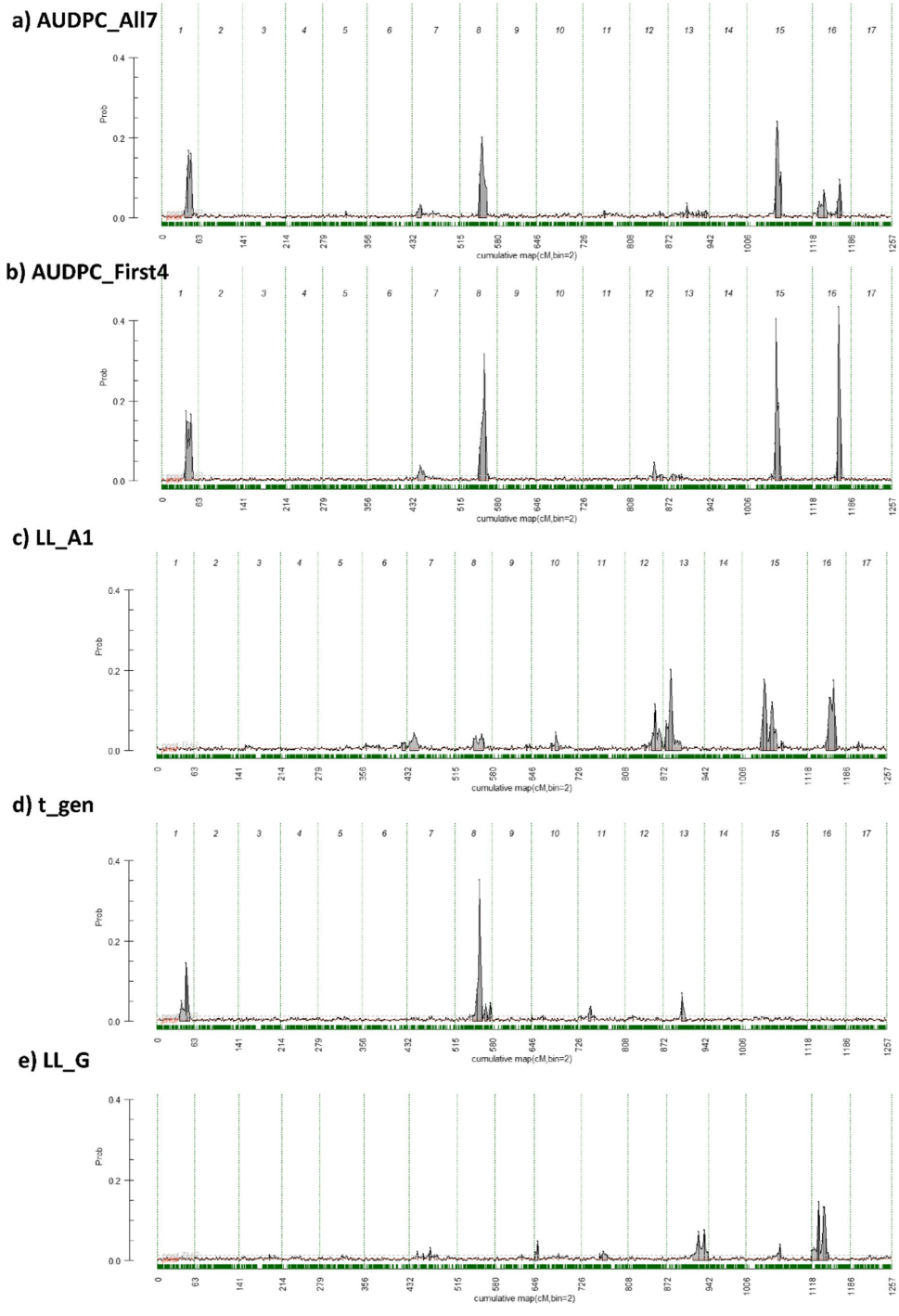
*A 2lnBF value above 0, 2, 5, and 10 is considered to indicate no, positive, strong, and decisive statistical evidence for the presence of a QTL, respectively.

**Not represented by two or more consecutive bins with 2lnBF above 5 in all four replicate runs, but shown for consistency with QTL regions for other parameters.

the AUDPC over only the first four assessments (AUDPC_First4), when most wounds had not started to girdle. Next, QTL mapping was performed on three parameters of a sigmoid growth curve: initial lesion length (LL_A1), growth rate (t_gen), and maximum lesion length before the wounds girdle (LL_G).

For AUDPC_All7 there was strong statistical evidence for three QTL, one each on LG1, LG8, and LG15. In addition, there was positive evidence for one QTL on LG16, although this evidence was the joined result of two peaks in the posterior probability far apart, while there was hardly any evidence for 2 QTL (2lnBF: 0.6 – 1.3) (Table 1, Fig. 2, File S2).

AUDPC_First4 provided decisive evidence for one QTL each on LG8 and LG15, and strong evidence for one QTL each on LG1 and LG16. In contrast to AUDPC_All7, there was a single consistent peak in posterior intensity for LG16, coinciding with one of the peaks for AUDPC_All7 on that LG, denoted LG16b (Table 1, Fig. 1, File S2).


a) AUDPC_All7
b) AUDPC_First4
c) LL_A1
d) t_gen
e) LL_G
Prob
cumulative map(cM,bin=2)

◂**Fig. 2** Posterior QTL positions from FlexQTL for each parameter, from a run with 3 prior QTL, **a** AUDPC_All7, **b** AUDPC_First4, **c** LL_A1, **d** t_gen, and **e** LL_G. Note that statistical evidence is based on the area under a peak rather than the height of a peak

The LL_A1 parameter provided strong evidence for one QTL, and positive evidence for a second QTL on LG15, one of which overlapped with the QTL detected for AUDPC_All7 and AUDPC_First4 on this LG. There was also strong evidence for a QTL for LL_A1 on LG16, which overlapped with the LG16b locus detected for AUDPC_First4, and positive evidence for a QTL on LG8, overlapping with the region identified for other parameters. Additionally, there was strong evidence for a QTL for LL_A1 on LG13, which was not detected for any of the other parameters analyzed and was not considered further (Table 1, Fig. 2, File S2).

The parameter t_gen provided strong evidence for one QTL on LG8, overlapping with the QTL on LG8 detected for AUDPC_All7 and AUDPC_First4. In addition, one QTL with only positive evidence was detected on LG1, overlapping with the QTL for AUDPC_All7 and AUDPC_First4 on this LG (Table 1, Fig. 2, File S2).

LL_G provided strong evidence for one QTL on LG16, which coincided with one of the QTL for AUDPC_All7 on LG16, but not with the LG16b locus. Thus, we denote this locus LG16a (Table 1, Fig. 2, File S2).

All identified QTL regions explained a low to moderate proportion of the variance of the respective parameter (Table 1). None of the parameters showed strong evidence ($2\ln BF > 5$) for more than one QTL on any LG, in any replicate FlexQTL run. All FlexQTL runs converged (ECS > 100 for all parameters) and QTL intervals were reproducible between runs (File S2). The LG15 locus was the only one with strong evidence for segregation in 'Aroma', while there was strong evidence for segregation in 'Discovery' for all five loci (Table 1). The two peaks for the LL_A1 parameter on LG15 (Fig. 2c) seemed to be fitted alternatingly according to the trace plot and were thus interpreted as representing a single QTL with an uncertain position. All other identified QTL regions showed independence in their discovery according to the trace plots (File S2). The consensus regions of all 5 QTL intervals were investigated (i.e., the outer boundaries of the range across parameters), and the presence of segregating markers and recombination events could be confirmed in all five cases.

Compound QTL genotypes

Thus, five QTL regions related to different parameters of a sigmoid growth curve for lesion length were identified and considered further. Next, we analyzed the segregation patterns (monohybrid vs dihybrid) and potential interaction effects within and between loci, considering the pairs of loci affecting the same parameter. To evaluate the behavior of the compound QTL genotypes, individuals without recombination in the consensus QTL regions for the relevant loci were selected. For LG1 and LG8, the t_gen parameter was used as phenotype as there were no other QTL regions with LG-wide positive evidence for that parameter. For the LG15 and LG16b loci, the LL_A1 parameter was used and the LL_G parameter was used for the LG16a locus, despite LG-wide strong and positive evidence for a locus segregating at LG13, respectively. The genotypes for the QTL regions were then analyzed by a series of nested linear models to investigate segregation patterns and interaction effects (File S2). None of the loci exhibited skewed segregation.

For the t_gen parameter QTL at LG1 and LG8, 111 individuals were identified without recombination events in the QTL regions. Considering the possible compound genotypes across both loci, the model with paternal segregation at LG1 and dihybrid segregation at LG8 was significantly better than the alternative models. The desirable *Q* alleles for t_gen came from 'Beauty of Bath' at LG1, and from 'Ingrid Marie' and 'Worcester Pearmain' at LG8. Adding terms for dominance or allele-specific epistasis between the two loci did not cause any significant improvements over the additive effects only model. Indeed, a simple linear model accounting for the total number of *Q* alleles across both loci had lower AIC (62.0) and BIC (70.0) than a model accounting for the segregation at each locus separately (65.0 and 78.5, respectively). Adding an extra *Q* allele increased the average t_gen by 0.07, 0.21, and 0.25 days when going from zero to one, from one to two, and from two to three *Q* alleles, respectively, corresponding to an average decrease in AUDPC_All7 of 149, 335, and 339 mm*days (Fig. 3a). As the average increase in t_gen is about three times larger when going from one to two, or from two to three *Q* alleles than when going from zero to one *Q* allele, there might be some synergistic epistasis for the number of *Q* alleles at the LG1 and LG8 loci. Indeed, alternative models accounting for epistasis but otherwise similar to the simple model (accounting for the total number of *Q* alleles across both loci) all had AICs in the same range (within 2 units of the minimum). However, there were only two alternative full models with a BIC that was less than 3 units above the simple model, indicating that they might be equally competitive. These models were ones that had an additional $\log_{10}$-transformed term for the number of *Q* alleles across both loci, and one which instead of the total number of *Q* alleles, had one term for going from zero to one *Q* allele, and one term for additional *Q* alleles beyond one.

For the LL_A1 parameter QTL at LG15 and LG16b, 115 individuals were identified without recombination events in the LG15 or LG16b QTL regions. Considering the additive effects of the four possible alleles at each locus,

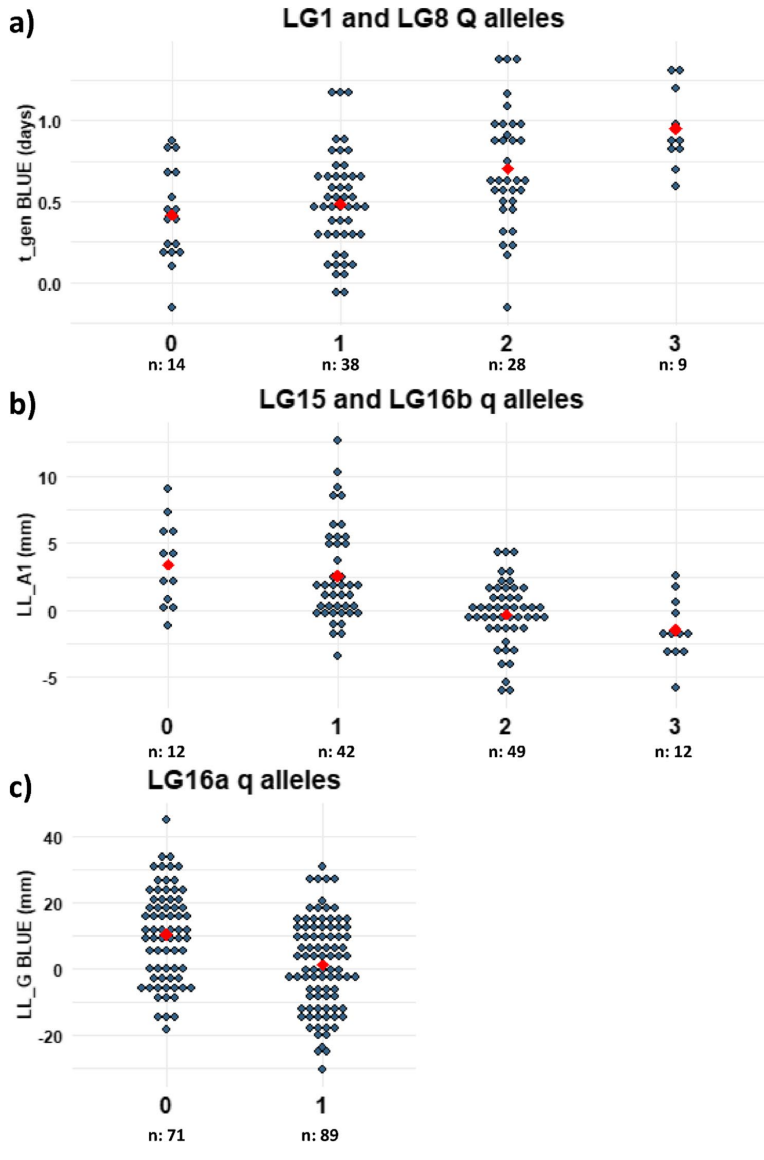


Fig. 3 Phenotype distributions of compound genotypes within the ArDi FS family, grouped by **a** number of *Q* alleles for t_gen at the LG1 and LG8 loci, **b** number of *q* alleles for LL_A1 at the LG15 and LG16b loci, and **c** number of *q* alleles for LL_G at the LG16a locus. Individuals with recombination events within the QTL regions of the respective QTL loci are excluded. Note that for LG1, LG16a, and LG16b the homozygous state of 'Aroma' cannot be determined, thus those alleles are not counted

the best model was for dihybrid segregation at LG15 and paternal monohybrid segregation at LG16b. The desirable *q* alleles for LL_A1 came from 'Filippa' and 'Beauty of Bath' at LG15, and from 'Worcester Pearmain' at LG16b. There were, however, no significant effects of dominance or allele-specific epistasis. As the three *q* alleles appeared to have similar effects, different linear models with the total number of *q* alleles regardless of locus were fitted and compared to the alternative models with individual alleles. Adding an extra *q* allele decreased LL_A1 by 0.85, 2.84, and 1.18 mm when going from zero to one, from one to two, and from two to three alleles, respectively (Fig. 3), corresponding to an average decrease in AUDPC_All7 of 2, 423, and 523 mm*days. The model with the lowest AIC accounted only for linear additive effects from the total number of *q* alleles across both loci, although a number of models which

had additional terms for non-linear or allele-specific effects, had AICs that were less than 2 units larger than the simple model. The latter had a BIC value 3.7 units lower than the alternative full model with the second lowest BIC. Similar to the case for the LG1 and LG8 loci, the second-best model was one that instead of the total number of *Q* alleles, had one term for going from zero to one *Q* allele and one term for additional *Q* alleles beyond one. While this might indicate the presence of some degree of synergistic epistasis for the LG1 and LG8 loci, and the LG15 and LG16b loci with regard to the number of *q* alleles, it could also be an artefact from the data as the distribution is skewed, particularly LL_A1 for which there are 'zero' observations (Fig. 1 and File S2).

There were 159 individuals without recombination events within the LG16a QTL region. Considering either maternal or paternal alleles separately, or compound genotypes, provided statistical evidence only for maternal segregation, with the decreasing *q* allele coming from 'Worcester Pearmain'. Individuals carrying the *q* allele had lesions that were on average 9.18 mm shorter when they girdled, and a 325 mm*days lower AUDPC_All7.

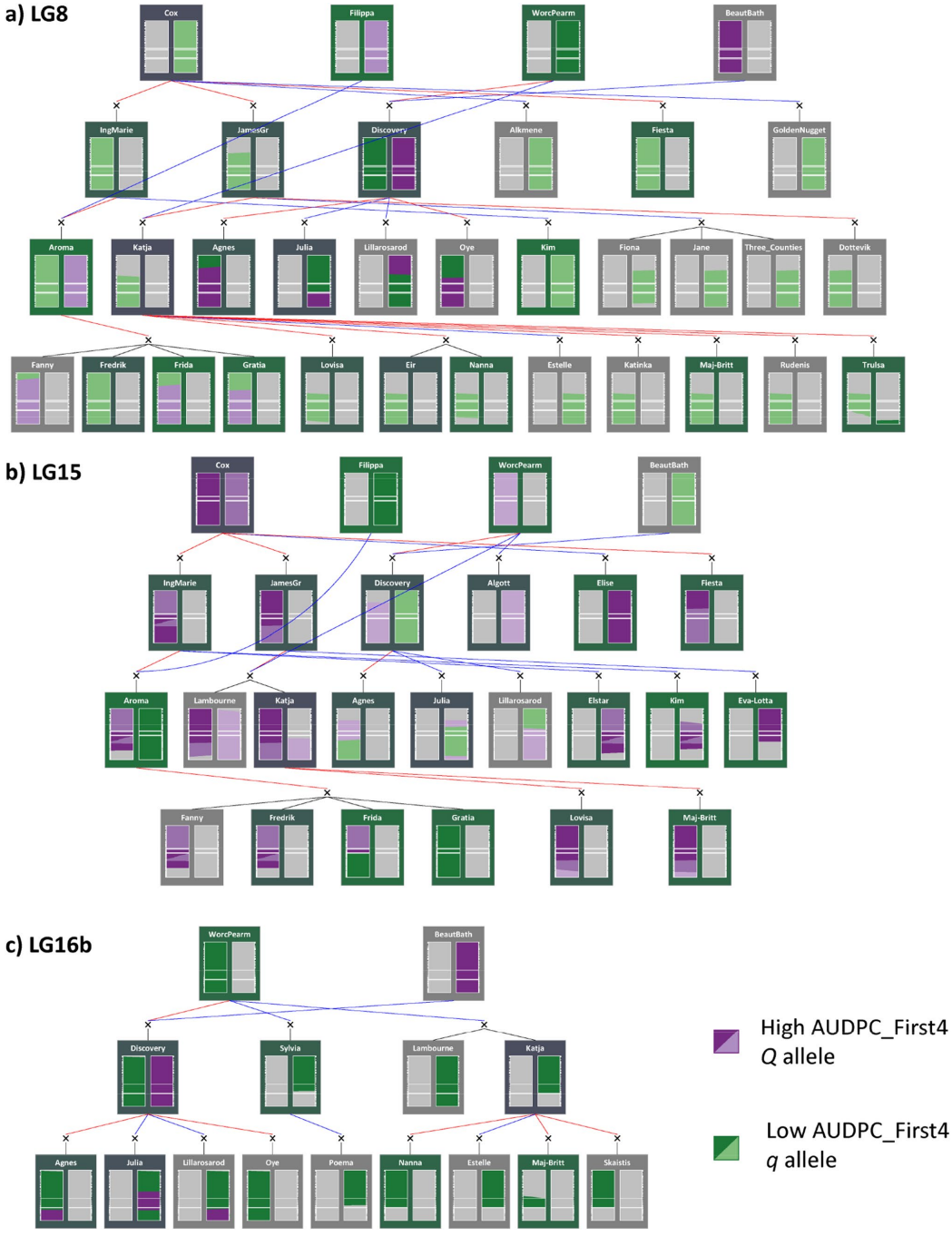
Search for epistatic loci

As the phenotypic distributions of some of the subgroups of the FS family seemed to indicate some residual segregation (e.g., the two *Q* alleles subgroup in Fig. 3a and the one *q* allele subgroup in Fig. 3b), we considered the possibility of epistatic interactions specific to each subgroup. To search for potential epistatic loci, we divided the FS population into four subsets with individuals having zero to three desirable alleles at the LG1 and LG8 loci, four subsets with individuals having zero to three desirable alleles at the LG15 and LG16b loci, and two groups carrying the *q* or *Q* allele from 'Discovery' at the LG16a locus. However, as this resulted in very small subsets, this must be considered very preliminary results. For individuals with one *Q* allele for t_gen at LG1 or LG8, there was strong LG-wide evidence for a QTL on LG3 for the t_gen parameter (2lnBF = 5.6), which was not found in the full FS family for any of the parameters. In the LG16a subgroups there was LG-wide positive evidence for one QTL on LG3 (2lnBF = 3.4) and on LG10 (2lnBF = 2.6), in the subgroups with one and zero *q* allele for LL_G, respectively.

In the original mapping using the full FS-family, strong LG-wide evidence for a QTL on LG13 was found for the LL_A1 parameter (2lnBF = 5.5–6.1). Investigation of the genotype probabilities assigned by FlexQTL for one of the replicate runs with the full mapping population and LL_A1 as parameter indicated that individuals with a probability above 0.5 for having a *Qq* genotype at the LG13 locus occurred at a frequency close to 0.5 for both the 1*q* and 2*q* subgroups (0.5 and 0.43, respectively). Thus, the LG13 locus is expected to segregate in both subgroups, but positive evidence was found for one QTL on LG13 for the AUDPC_All7 and AUDPC_First4 (2lnBF = 4.2 and 3.1, respectively) parameters only in the subgroup carrying a single *q* allele for AUDPC or LL_A1 at either the LG15 or the LG16b locus. That the locus on LG13 is only detected in the subgroup with one *q* allele might indicate that there is a locus on LG13 that is partially recessive to the other loci mapped in this study, although the absence of detectable segregation in the 2*q* subgroup could be due to scale effects. However, this is expected to be less of an issue with the AUDPC_All7 parameter than with LL_A1 (File S2). Furthermore, positive LG-wide evidence was found for the LL_G parameter (2lnBF = 2.8–3.6) on the same LG, although there was barely any evidence using the AUDPC or t_gen parameters (Fig. 2, File S2). The peak for the LL_A1 parameter appears on the lower end of LG13 together with the LG16b locus on the upper end of LG16, and the peak for the LL_G parameter appear of the upper end of LG13 together with the LG16a locus on the lower end of LG16 (Fig. 2). Chromosome 13 is known to be homoeologous to chromosome 16 (Daccord et al. 2017), although the QTL mapping was done based on genetic distances, which do not follow a linear relationship to the physical position across the entire LG. Due to the low power for the LG13 loci and the large QTL intervals, the potential homoeology between the two LG13 peaks and the LG16a and LG16b loci was not investigated further, although it might indicate a complex interaction between these regions.

Prevalence of the segregating haplotypes in the screening germplasm

Several of the phenotyped individuals in the screening germplasm were related to either of the parents of the FS family, with phased marker data available. Thus, transmission of IBD haplotypes can be traced and haplotypes being IBS to those segregating in the FS family can be identified. While validation of QTL with moderate effects on a polygenic trait can be difficult in a genetically diverse germplasm, it can still provide information on the effect of a QTL and guide future crosses for breeding and genetic research. Among the identified QTL regions, we focused on LG8, LG15, and LG16b for further analysis due to the low prevalence of the desirable allele from 'Beauty of Bath' at LG1 and the uncertain biological effect of the *q* allele at LG16a. The LG1 locus was not considered further, as the donor of the *q* allele, 'Beauty of Bath', had very few descendants in the germplasm studied. Also, the LG1 locus had lower Bayesian probability across parameters (Table 1). The LG16a locus was omitted due to its small effect, uncertainties in its segregation (low Bayesian probability, relatively few segregating SNP markers in the QTL region, and *q* allele of LG16a being in coupling phase with the *q* allele at the LG16b locus) and ambiguity in which allele should be considered desirable


a) LG8
Cox
Filippa
WorcPearm
BeautBath
IngMarie
JamesGr
Discovery
Alkmene
Fiesta
GoldenNugget
Aroma
Katja
Agnes
Julia
Lillarosarod
Oye
Kim
Fiona
Jane
Three_Counties
Dottevik
Fanny
Fredrik
Frida
Gratia
Lovisa
Eir
Nanna
Estelle
Katinka
Maj-Britt
Rudenis
Trulsa
b) LG15
Cox
Filippa
WorcPearm
BeautBath
IngMarie
JamesGr
Discovery
Algott
Elise
Fiesta
Aroma
Lambourne
Katja
Agnes
Julia
Lillarosarod
Elstar
Kim
Eva-Lotta
Fanny
Fredrik
Frida
Gratia
Lovisa
Maj-Britt
c) LG16b
WorcPearm
BeautBath
Discovery
Sylvia
Lambourne
Katja
Agnes
Julia
Lillarosarod
Oye
Poema
Nanna
Estelle
Maj-Britt
Skaistis
High AUDPC_First4
Q allele
Low AUDPC_First4
q allele

◂**Fig. 4** Transmission of QTL haplotypes of the QTL **a** LG8, **b** LG15, and LG16b along the pedigrees of some cultivars, which have inherited parts of the haplotypes of the consensus QTL regions segregating in 'Aroma' and 'Discovery'. The color of the frames indicates the cultivars BLUE for AUDPC_First4, ranging from purple (high value, susceptible) to green (low value, resistant). The boundaries of the consensus QTL regions are indicated by white lines. The founder haplotypes segregating in 'Aroma' × 'Discovery' are indicated as dark/light green (*q*) and dark/light purple (*Q*), all other haplotypes are grey. Red lines indicate maternal inheritance and blue lines indicate paternal inheritance. Note that both haplotypes of 'Cox's Orange Pippin' are traced for LG15, as 'Ingrid Marie' seems to have a recombination event within the consensus QTL region, such that the first 3 cM comes from one haplotype (light purple) and the upper 7 cM from the other haplotype (dark purple) of 'Cox's Orange Pippin'. Additionally, 'Ingrid Marie' seems to have two subsequent double recombination events on LG15 from 'Cox's Orange Pippin', including a 10-cM region upstream of the QTL regions where 'Cox's Orange Pippin' is homozygous for the markers included in this study

(see the "Discussion" section). Thus, haplotypes of the LG8, LG15, and LG16b loci that segregated in the FS family were traced along the known pedigrees of some cultivars related to 'Aroma' and 'Discovery' (Fig. 4), and individuals with IBS haplotypes were identified in the germplasm (File S1).

Concerning LG8, 'Lillarosaröda' (an unreleased selection from the Swedish breeding program, 'Lillarosarod' in Fig. 4) and the highly susceptible 'Julia' are the only individuals, apart from 'Discovery', to have inherited the *q* allele for AUDPC from 'Worcester Pearmain'. However, the *q* allele for AUDPC coming from 'Cox's Orange Pippin' has been transmitted to several descendants, and there are 15 haplotypes that are IBS to either of the *q* alleles (Fig. 4a, File S1).

Regarding the LG15 locus, 'Gratia' seems to be the only descendant of 'Aroma' to share the *q* allele by IBD, and the only individual that shares the *q* allele of 'Beauty of Bath' by IBD is the susceptible cultivar 'Julia'. Notably, 'Ingrid Marie' seems to have a recombination event within the consensus QTL region, so it is unclear which of the descendants of the 'Cox's Orange Pippin' inherited the *Q* allele segregating in the 'Aroma' × 'Discovery' family. However, 10 haplotypes in the screening germplasm are IBS to either of the *Q* alleles, while only 5 are IBS to either of the *q* alleles, all of which are IBD (Fig. 4b, File S1).

The *q* allele from 'Worcester Pearmain' at the LG16b locus has been transmitted to six individuals in the screening germplasm, and several other cultivars which were not phenotyped. Offspring of 'Worcester Pearmain' that have inherited the *q* allele include 'Discovery', 'Katja', and 'Lord Lambourne' (Fig. 4c, File S1). These have in turn transmitted the *q* allele to cultivars such as 'Agnes', 'Lillarosaröda', 'Øye', 'Nanna', 'Estelle', 'Maj-Britt', and 'Skaistis' ('Estelle' and 'Maj-Britt' are unreleased selections from the Swedish breeding program).

While not considered further, it should be noted that the LG16a *q* allele for AUDPC_All7 is in coupling phase with the *q* allele of the LG16b locus in 'Worcester Pearmain'. Also, a relatively large number of haplotypes in the screening germplasm are IBS to either the *Q* or the *q* alleles (17 haplotypes of each allele, File S1).

Discussion

QTL related to resistance to European canker

We identified several QTL for resistance against *N. ditissima* segregating in 'Aroma' × 'Discovery', five of which were considered for further analysis. Two QTL on LG15 and LG16 (LG16b) were identified for the LL_A1 parameter as well as AUDPC_First4. Two QTL on LG1 and LG8 were identified for the t_gen parameter and both AUDPCs. Lastly, a second QTL on LG16 (LG16a) was identified for the LL_G parameter. The AUDPC_All7 parameter gave strong evidence for only three QTL and no evidence for presence of two QTL on LG16. However, further analysis by parametrization provided strong evidence for two QTL on LG16 with effects during different stages of lesion growth. Thus, parametrization of the phenotypic data allowed us to identify more QTL with better statistical power, and provided a first insight into their roles during different stages of disease development under these current experimental conditions. For example, 'Discovery' segregated for a *q* allele for LL_G at the LG16a locus. The *q* allele results in a shorter lesion length when the wound girdles, which implies a smaller lesion with less potential as a source for secondary infections. On the other hand, a shorter lesion length when the wound girdled implies that distal parts of the stem dies quicker. From an epidemiological point the *q* allele might be desirable while the effect of the *Q* allele might be preferred by the farmer as the distal parts of the branch can remain productive for a longer time before it dies off. Thus, the *Q* allele could be considered a tolerance allele rather than a susceptibility allele. In the current study, we found no statistically significant dominance effects for the five loci, although this could be due to a lack of statistical power in the available data and monohybrid segregation as for, e.g., the LG16b locus. Thus, further validation studies would benefit from being designed to enable the assessment of dominance effects at the QTL of interest. On the other hand, we identified potentially epistatic loci on LGs 3 and 13, with effects specific to individuals in the FS family carrying one *q* allele for AUDPC at the LG1 and LG8 or the LG15 and LG16b loci, respectively. We also identified a potential epistatic interaction between the LG3 and LG10 loci for subsets of the FS family that have one and zero *q* allele at the LG16a locus, respectively. However, these potentially epistatic loci were identified based on small subsets of the FS family, with the parameters of each subgroup suffering from scale effects to different degrees. Thus, their status as epistatic loci must be

considered very preliminary, and interpretations should be made with great caution until their effects have been further verified. Putative epistatic QTL interactions have previously been identified in other pathosystems of *M. domestica*, e.g., resistance to fire blight (van de Weg et al. 2018).

In the current study, the LG8, LG15, and LG16b loci had the highest probabilities across parameters (Table 1). Bus et al. (2021) also identified QTL for resistance against European canker on LG8 and LG16 segregating in a bi-parental family. The SNP being most strongly linked to the LG8 locus in that study was within the consensus QTL peak of the LG8 locus in the current study and might thus represent the same locus. Similarly to the current study, the LG8 locus was segregating in the cultivar considered to be susceptible ('Scired'). On the other hand, the SNP being reported by Bus et al. (2021) as most tightly linked to the LG16 locus is located in between the LG16a and LG16b loci identified in the current study, outside the consensus QTL regions of both loci. Additionally, Karlström et al. (2022) recently performed QTL mapping through Pedigree Based Analysis (PBA) using several disease parameters and found positive or strong evidence for QTL regions on several linkage groups, including LGs 8, 15 and 16. While the QTL regions on LG15 of that study barely coincided with the QTL at LG15 in the current study, the statistical power was much higher in the current study, with decisive evidence for the AUDPC_First4 parameter. This might in part be due to the much smaller families used by Karlström et al. (2022) (60–69 individuals). Considering a trait that is controlled by several QTL with moderate effect, large families will be required for both breeding and QTL mapping. For further efforts in pre-breeding or validation of QTL effects, one might for example consider the cross 'Gratia' × 'Lillarosaröda' where 'Gratia' is carrying the *q* allele from 'Filippa' at LG15 and 'Lillarosaröda' is carrying the *q* alleles for AUDPC from 'Worcester Pearmain' at the LG8 and LG16b loci (Fig. 4).

Experimental procedures

The experimental procedures of the current study provided phenotypic data of high quality, which facilitated high statistical power in QTL detection. By dissecting the phenotypic data further, in light of a sigmoid growth curve for lesion growth, these QTL regions were found to play key roles during specific stages of the infection. This indicates that the time points at which assessments are being made can have a crucial effect on which QTL regions are detected in a study. Wenneker et al. (2017) suggested LGR as the most robust metric for quantification of resistance to European canker. In the current study, we used the parameter t_gen, which is conceptually similar to LGR as both are measures of growth rate. In line with Wenneker et al. (2017), we found the t_gen parameter to provide high heritability and correlations between years in both the screening germplasm and the FS family. While the t_gen parameter provided strong evidence for the LG8 QTL, it did not provide any evidence for the LG15, LG16a, or LG16b QTL regions.

While it has been questioned whether screening tests in controlled conditions are relevant to field conditions (Delgado et al. 2022), there are large differences between experimental set-ups that might appear similar at a first glance, e.g., potted trees experiment. Karlström et al. (2022) also made use of a controlled conditions experiment with potted trees, which in some respects was similar to the experimental set-up of the current study. There were, however, several differences in the procedure, including physiological growth stage of the trees, climatic conditions, number of replicates and pseudo-replicates, number of conidia applied to each wound, and assessment intervals. They reported a broad sense heritability of 0.46 for their potted trees experiments, which is much lower than what is reported here (0.78 for AUDPC_All7). Consequently, they reported a much weaker statistical support for the QTL regions identified in their study.

Future prospects

As the identified QTL all had low to moderate effects, and the estimates of broad-sense heritability were high, each of the identified QTL will likely have a low predictiveness by themselves. Thus, maximal genetic gain will likely be obtained by first identifying the most suitable parent-parent combination, followed by a two-step selection process, first by culling the worst genotypes using marker-assisted selection and subsequently by phenotypic screening of the remaining individuals (Vanderzande et al. 2018; Ru et al. 2016). However, this requires that the identified QTL haplotypes can be converted to a small number of markers that can be routinely analyzed on a relatively cheap platform. As the SNPs for the 20 K SNP array were selected from a narrow genetic base of only 14 individuals, mostly with high MAF, single SNPs that are diagnostic of a given QTL region are unlikely to be identified. Rather, subsets of SNPs that can distinguish a given haplotype in a given breeding germplasm could be searched for (Chagné et al. 2019). Additionally, QTL effects from desirable alleles that are rare in a germplasm, such as the *q* allele at LG15 from 'Filippa', are difficult to verify, and even more so in the presence of several additional QTL of moderate effect segregating in the germplasm. Thus, future efforts to validate the QTL regions identified in the current study, and potential subsets of SNPs for marker-assisted seedling selection, would require FS-families that are expected to segregate

for the specific QTL alleles to be validated. Furthermore, a promising source of resistance for the Swedish breeding program, the cultivar 'Santana', was not IBS for the favorable alleles at any of the QTL loci. Thus, 'Santana' might pose a supplementary source of resistance, and further breeding efforts would benefit from the mapping of resistance QTL segregating in that cultivar.

Data archiving statement

BLUEs for phenotypic parameters for the screening germplasm and phased SNP calls of the parents of the FS family for each consensus QTL region are given in File S1. SNP calls and BLUEs for the different parameters of the FS-family can be made available upon request.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11295-023-01587-w.

Acknowledgements We gratefully acknowledge access to germplasm for genotyping granted by Inger Hjalmarsson and the Programme for Diversity of Cultivated Plants (Swedish Central Collection, Alnarp). We are appreciative of Tuuli Haikonen (Luke, Finland), Dag Røen (Graminor, Norway), and Vidmantas Bendokas (Lithuanian Research Centre of Agriculture and Forestry, Lithuania) for providing material through the PPP Nordfruit project. Excellent genotyping service was provided by the Sequencing and Genotyping platform at Fondazione Edmund Mach (FEM, Italy). Thanks to Jasna Sehic, Marjan Ghasemkhani, Anna Zborowska, and Malin Dörre for technical assistance as well as Adam Flöhr and Jan-Eric Englund for statistical support. Thanks also to Stijn Vanderzande, Nicholas P. Howard, and Sujeet Verma for valuable comments on various aspects of data curation and the use of FlexQTL. Data analysis benefited from the Udda server at the Department of Plant Breeding, LTV Faculty, SLU.

Funding Open access funding provided by Swedish University of Agricultural Sciences. This work was funded by the SLU GroGrund project "Accelerated and cost efficient cultivar development through genome informed apple breeding", Formas (grant nr 2014–00924 and grant nr 2017–00496), Nordic Co-operation ("PPP Nordfruit"), and SLF (grant V1036010). SNP data were shared by the Fruitbreedomics Project No. 265582, which was funded by the EU seventh Framework Programme, and the PPP Nordfruit, which was funded by the Nordic Co-operation.

Declarations

Conflict of interest The authors declare no competing interests.

References

Bac-Molenaar JA, Vreugdenhil D, Granier C, Keurentjes JJB (2015) Genome-wide association mapping of growth dynamics detects time-specific and general quantitative trait loci. J Exp Bot 66:5567. https://doi.org/10.1093/JXB/ERV176

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/JSS.V067.I01

Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagné D, Velasco R, van de Weg E, Troggio M (2014) Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). Plos One 9:e110377. https://doi.org/10.1371/journal.pone.0110377

Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104:751–762. https://doi.org/10.1007/s00122-001-0796-x

Bink MCAM, Boer MP, Ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE (2008) Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161:85–96. https://doi.org/10.1007/s10681-007-9516-1

Bink MC, Jansen J, Madduri M, Voorrips RE, Durel CE, Kouassi AB, Laurens F, Mathis F, Gessler C, Gobbin D, Rezzonico F, Patocchi A, Kellerhals M, Boudichevskaia A, Dunemann F, Peil A, Nowicka A, Lata B, Stankiewicz-Kosyl M, Jeziorek K, Pitera E, Soska A, Tomala K, Evans KM, Fernandez-Fernandez F, Guerra W, Korbin M, Keller S, Lewandowski M, Plocharski W, Rutkowski K, Zurawicz E, Costa F, Sansavini S, Tartarini S, Komjanc M, Mott D, Antofie A, Lateur M, Rondia A, Gianfranceschi L, van de Weg WE (2014) Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. Theor Appl Genet 127:1073–1090. https://doi.org/10.1007/s00122-014-2281-3

Brown AE, Muthumeenakshi S, Swinburne TR, Li R (1994) Detection of the source of infection of apple trees by *Cylindrocarpon heteronema* using DNA polymorphisms. Plant Pathol 43:338–343. https://doi.org/10.1111/j.1365-3059.1994.tb02693.x

Bus VGM, Scheper RWA, Walter M, Campbell RE, Kitson B, Turner L, Fisher BM, Johnston SL, Wu C, Deng CH, Singla G, Bowatte D, Jesson LK, Hedderley DI, Volz RK, Chagné D, Gardiner SE (2019) Genetic mapping of the European canker (*Neonectria ditissima*) resistance locus Rnd1 from *Malus* 'Robusta 5.' Tree Genet Genomes 15:1–13

Bus V, Single G, Horner M, Jesson L, Walter M, Kitson B, López-Girona E, Kirk C, Gardiner S, Chagné D, Volz R (2021) Preliminary genetic mapping of fire blight and European canker resistances in two apple breeding families. Acta Hortic 1307:199–203

Chagné D, Vanderzande S, Kirk C, Profitt N, Weskett R, Gardiner SE, Peace CP, Volz RK, Bassil NV (2019) Validation of SNP markers for fruit quality and disease resistance loci in apple (*Malus* × *domestica* Borkh.) using the OpenArray® platform. Hortic Res 6:30. https://doi.org/10.1038/s41438-018-0114-2

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, Di Pierro EA, Gouzy J, Rees DJG, Guerif P, Muranty H, Durel CE, Laurens F, Lespinasse Y, Gaillard S, Aubourg S, Quesneville H, Weigel D, van de Weg E, Troggio M, Bucher E (2017) High-quality de novo assembly of the apple genome and methylome

dynamics of early fruit development. Nat Genet 49:1099–1106. https://doi.org/10.1038/ng.3886

Delgado Á, García-Fernández B, Gómez-Cortecero A, Dapena E (2022) Susceptibility of Cider Apple Accessions to European Canker - Comparison between Evaluations in Field Planted Trees and Rapid Screening Tests. Plants (Basel). https://doi.org/10.3390/plants11091145

Di Pierro E, Gianfranceschi L, Di Guardo MM, Koehorst-van H, Kruisselbrink J, Longhi S, Troggio M, Bianco L, Muranty H, Pagliarani G, Tartarini S, Letschka T, Luis LL, Garkava-Gustavsson L, Micheletti D, Bink M, Voorrips R, Aziz E, Velasco R, Laurens F, van de Weg E (2016) A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic Res 3:16057

Fox J, Weisberg S (2019) An R companion to applied regression, 3rd edn. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Garkava-Gustavsson L, Ghasemkhani M, Zborowska A, Englund J-E, Lateur M, van de Weg E (2016) Approaches for evaluation of resistance to European canker (*Neonectria ditissima*) in apple. Acta Hortic 1127:75–82. https://doi.org/10.17660/ActaHortic.2016.1127.14

Gómez-Cortecero A, Saville RJ, Scheper RWA, Bowen JK, Agripino De Medeiros H, Kingsnorth J, Xu X, Harrison RJ (2016) Variation in host and pathogen in the *Neonectria/Malus* interaction; toward an understanding of the genetic basis of resistance to European canker. Front Plant Sci 7:1365. https://doi.org/10.3389/fpls.2016.01365

Hjeltnes S-H, Garkava-Gustavsson L, Alsheikh M, Borve J, Davik J, Haikonen T, Ikase L, Keldmaë H, Karhu S, Kjos B, Odilbekov F, Parikka P, Pedersen C, Rentanen M, Rumpunen K, Røen D, Sasnauskas A, Volens K, van de Weg E (2019) Public Private Partnership project NORDFRUIT: meeting the future challenges in Nordic apple and strawberry production [Poster presentations]. XV EUCARPIA Fruit Breeding and Genetics Symposium, Prague

Howard NP, Troggio M, Durel C-E, Muranty H, Denancé C, Bianco L, Tillman J, van de Weg E (2021) Integration of Infinium and Axiom SNP array data in the outcrossing species *Malus* × *domestica* and causes for seemingly incompatible calls. BMC Genomics 22:1–18. https://doi.org/10.1186/s12864-021-07565-7

Karlström A, Gómez-Cortecero A, Nellist CF, Ordidge M (2022) Dunwell JM, Harrison RJ (2022) Identification of novel genetic regions associated with resistance to European canker in apple. BMC Plant Biol 221(22):1–16. https://doi.org/10.1186/S12870-022-03833-0

Krähmer H (1980) Wound reactions of apple trees and their influence on infections with *Nectria galligena*. J Plant Dis Prot 87(2):97–112

Mangandi J, Verma S, Osorio L, Peres NA, van de Weg E, Whitaker VM (2017) Pedigree-based analysis in a multiparental population of octoploid strawberry reveals QTL alleles conferring resistance to *Phytophthoracactorum*. G3 Genes. Genomes Genet 7:1707–1719

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ru S, Hardner C, Carter PA, Main Evans K, D, Peace C, (2016) Modeling of genetic gain for single traits from marker-assisted seedling selection in clonally propagated crops. Hortic Res 3:16015. https://doi.org/10.1038/hortres.2016.15

Schmidt P, Hartung J, Rath J, Piepho HP (2019) Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci 59:525–536. https://doi.org/10.2135/CROPSCI2018.06.0376

Sprouffske K, Wagner A (2016) Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics 17:172. https://doi.org/10.1186/s12859-016-1016-7

Skytte af Sätra J, Troggio M, Odilbekov F, Sehic J, Mattisson H, Hjalmarsson I, Ingvarsson PK, Garkava-Gustavsson L, (2020) Genetic status of the Swedish central collection of heirloom apple cultivars. Sci Hortic (Amsterdam) 272:109599. https://doi.org/10.1016/j.scienta.2020.109599

Swinburne TR (1975) European canker of apple (*Nectria galligena*). Rev Plant Pathol 54:787–799

van de Weg E, Di Guardo M, Jänsch M, Socquet-Juglard D, Costa F, Baumgartner I, Broggini GAL, Kellerhals M, Troggio M, Laurens F, Durel C-EE, Patocchi A (2018) Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. Mol Breed 38:1–18. https://doi.org/10.1007/s11032-017-0755-0

Vanderzande S, Howard NP, Cai L, Da Silva Linge C, Antanaviciute L, Bink MCAM, Kruisselbrink JW, Bassil N, Gasic K, Iezzoni A, van de Weg E, Peace C (2019) High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. Plos One 14:e0210928. https://doi.org/10.1371/journal.pone.0210928

Voorrips RE, Bink MCAM, van ee Weg WE (2012) Pedimap: Software for the Visualization of Genetic and Phenotypic Data in Pedigrees. J. Hered. 103:903–907. https://doi.org/10.1093/JHERED/ESS060

Weber RWS (2014) Biology and control of the apple canker fungus *Neonectria ditissima* (syn. *N. galligena*) from a Northwestern European perspective. Erwerbs-Obstbau 56:95–107. https://doi.org/10.1007/s10341-014-0210-x

Wenneker M, Goedhart PW, van der Steeg P, van de Weg WE, Schouten HJ (2017) Methods for the quantification of resistance of apple genotypes to European fruit tree canker caused by *Neonectria ditissima*. Plant Dis 101:2012–2019. https://doi.org/10.1094/PDIS-04-17-0576-RE

Wu R, Ma CX, Zhao W, Casella G (2003) Functional mapping for quantitative trait loci governing growth rates: A parametric model. Physiol Genomics 14:241–249. https://doi.org/10.1152/physiolgenomics.00013.2003

van de Weg E, van der Steeg P, Wenneker M, Koerst-van Putten H, Smulders MJM, Schouten H (2020) Resistance to *Neonectria ditissima* in apple: what QTL discovery teaches us about plant resistance tests and resistance parameters [Conference presentation]. 4th International workshop on European fruit tree canker and resilient orchards, Online via Zoom

van de Weg EW (1989) Breeding for resistance to *Nectria galligena* in apple: differences in resistance between seedling populations. In Gessler C, Butt DJ, Koller B (eds) Integrated control of pome fruit diseases, vol II. Proceedings of a workshop of the Subgroup on integrated control of pome fruit diseases of the Working Group: "Integrated Plant Protection in Orchards" pp 137–145

Vanderzande S, Piaskowski JL, Luo F, Edge Garza DA, Klipfel J, Schaller A, Martin S, Peace C (2018) Crossing the finish line: How to develop diagnostic DNA tests as breeding tools after QTL discovery. J Hortic 5:1. https://doi.org/10.4172/2376-0354.1000228

Zeileis A, Hothorn T (2002) Diagnostic checking in regression relationships. R News 2(3):7–10. https://CRAN.R-project.org/doc/Rnews/

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, Wang D, Tian Y, Yang C, Meng M, Yuan G, Kang G, Wu Y, Wang K, Zhang H, Wang D (2019) Cong P (2019) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun 10:1–13

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2023:18

In recognition of the special challenges faced by domestic growers, the Swedish apple breeding programme has been operating since the 1940s. The purpose of this thesis has been to establish the infrastructure necessary for genomic-led breeding in the Swedish apple programme. Towards this end, genetic resources have been curated, a novel genetic tool has been developed, and key traits such as resistance to European canker and phenology traits have been studied.

Jonas Skytte af Sätra received his doctoral education at the Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp. He received his MSc in Horticulture from the Swedish University of Agricultural Sciences, Alnarp.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-088-0
ISBN (electronic version) 978-91-8046-089-7