UNIVERSIDADE FEDERAL DE PELOTAS
Faculdade de Agronomia Eliseu Maciel
Programa de Pós-Graduação em Agronomia

Tese

Associação genômica ampla para resistência a bacteriose em germoplasma de pessegueiro com base em SNPs

Liane Bahr Thurow

Pelotas, 2018

Liane Bahr Thurow

Associação genômica ampla para resistência a bacteriose em germoplasma de pessegueiro com base em SNPs

Tese apresentada ao Programa de Pós-Graduação em Agronomia da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutora em Ciências (área do conhecimento: Fitomelhoramento).

Orientadora: Caroline Marques Castro, Dra. - Embrapa Clima Temperado

Co-Orientadores:

Maria do Carmo Bassols Raseira, PhD. - Embrapa Clima Temperado

Sandro Bonow, Dr. - Embrapa Clima Temperado

Pelotas, 2018

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

T542a Thurow, Liane Bahr

Associação genômica ampla para resistência a bacteriose em germoplasma de pessegueiro com base em SNPs / Liane Bahr Thurow ; Caroline Marques Castro, orientadora ; Maria do Carmo Bassols Raseira, Sandro Bonow, coorientadores. — Pelotas, 2018.

102 f. : il.

Tese (Doutorado) — Programa de Pós-Graduação em Agronomia, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, 2018.

1. Prunus persica. 2. Genotipagem por sequenciamento. 3. Fenotipagem. 4. REML/BLUP. 5. GWAS. I. Castro, Caroline Marques, orient. II. Raseira, Maria do Carmo Bassols, coorient. III. Bonow, Sandro, coorient. IV. Título.

CDD : 634.25

Elaborada por Gabriela Machado Lopes CRB: 10/1842

Liane Bahr Thurow

Associação genômica ampla para resistência a bacteriose em germoplasma de pessegueiro com base em SNPs

Tese aprovada, como requisito parcial, para obtenção do grau de Doutora em Ciências, Programa de Pós-Graduação em Agronomia, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas.

Data da Defesa: 14 de março de 2018

Banca Examinadora:

.....
Prof. Dra. Caroline Marques Castro (Orientadora)
Doutora em Genética pela Universidade Estadual Paulista Júlio de Mesquita Filho

.....
Prof. *PhD.* Antonio Costa de Oliveira
Ph.D. em *Genetics* pela *Purdue University*

.....
Prof. Dr. Leandro José Dallagnol
Doutor em Fitopatologia pela Escola Superior de Agricultura “Luiz de Queiroz”

.....
Prof. Dr. Valmor João Bianchi
Doutor em Agronomia pela Universidade Federal de Pelotas

Aos meus pais Neldo Buss Thurow e Ilma Bahr Thurow,
dedico esta conquista,
meu carinho e toda minha vida,
pela vida toda.

Agradecimentos

A Deus, por sempre guiar meu caminho.

Aos meus pais, Neldo e Ilma, meus maiores exemplos de amor incondicional e dedicação. Grata pela vida, pelo apoio, cuidado e carinho.

Aos melhores irmãos que eu poderia ter, Liamara e Márcio, juntos formamos um trio que fez esta jornada mais fácil, mais segura e mais próxima daquilo que é efetivamente importante.

Ao Daniel Hornke, meu amado noivo, parceiro de vida e de sonhos, pelo incentivo, paciência, apoio e amor dedicados ao longo destes anos.

À minha orientadora Dra. Caroline Marques Castro, grata pela confiança, por todo o aprendizado, pelas oportunidades de crescimento profissional, pelo incentivo e amizade. Talvez não consiga expressar em palavras o carinho e gratidão que tenho.

À Dra. Maria do Carmo Bassols Raseira e a Dra. Ksenija Gasic, que honra contar com duas melhoristas tão renomadas, exemplos de dedicação e amor pelo trabalho que desenvolvem.

Ao Programa de Pós-graduação em Agronomia da UFPel, pela oportunidade de formação profissional.

À CAPES pela bolsa de estudos no Brasil e ao CNPq pela concessão da bolsa de estudos no exterior.

À Embrapa Clima Temperado pela oportunidade, infraestrutura e apoio financeiro disponibilizado para o desenvolvimento deste trabalho.

À todos dos laboratórios de Biologia Molecular, Fitopatologia e Melhoramento Genético, pela agradável convivência, experiências compartilhadas e amizade. Em

especial para Angela, Raquel, Natércia, Naci, Joyce, Franciele, Tuane, Daiana, Fernanda, Luiz Felipe, Juliana, Jaqueline, Cristiano e Dani.

Ao Dr. Bernardo Ueno, por todos os ensinamentos na área de fitopatologia.

Aos Drs. Willian Barros e Maicon Nardino pelo auxílio nas análises estatísticas com modelos mistos.

À Dra Ksenija Gasic, por todo o aprendizado durante os 12 meses de doutorado sanduiche junto a Universidade de Clemson (EUA) e pelas incontáveis trocas de e-mails sanando dúvidas e construindo conhecimento. Agradeço também aos Drs. Guido Schnabel, Gregory Reighard e Douglas Bielenberg pela ajuda sempre que foi necessário. Aos meus colegas e amigos do *Gasic lab* Asma, May, Brad, Trey, Collenn, Brianna, Amanda e Verónica.

À todos os professores que tive em minha vida, especialmente aqueles que amam o que fazem, sejam doutores ou de ensino fundamental incompleto, vossos ensinamentos nortearão sempre a minha vida e futura carreira.

À todos aqueles que ajudaram a construir esta tese, àqueles que sempre acreditaram em mim e me motivaram a seguir em frente, aos familiares e aos amigos não nominalmente citados, mas nunca esquecidos, pelo significado de suas presenças em minha vida.

“If I have seen further it is by standing on the shoulders of Giants.”

Isaac Newton

Resumo

THUROW, Liane Bahr. **Associação genômica ampla para resistência a bacteriose em germoplasma de pessegueiro com base em SNPs.** 2018. 102f. Tese (Doutorado em Agronomia - Fitomelhoramento) - Programa de Pós-Graduação em Agronomia. Universidade Federal de Pelotas, Pelotas, 2018.

O pessegueiro (*Prunus persica*) é uma das espécies decíduas geneticamente melhor caracterizadas e a terceira frutífera mais importante de clima temperado em todo o mundo. A cultura é vulnerável à bacteriose causada pelo patógeno *Xanthomonas arboricola* pv. *pruni* (*Xap*) e o melhoramento genético visando resistência têm sido a melhor forma de controle da doença. Com o objetivo de contribuir para desenvolvimento de cultivares com maior resistência e implementar análises de associação genômica ampla (GWAS), foi explorado o uso de genotipagem por sequenciamento (GBS) para descoberta e genotipagem simultânea de SNPs em larga escala e realizada caracterização fenotípica para resposta à *Xap*, em um painel de 220 genótipos de pessegueiro, representativos do germoplasma disponível para melhoramento no Brasil. Um total de 93.353 marcadores SNPs foram descobertos e após filtragem de alta qualidade 18.373 SNPs foram utilizados. Destes, 34% estavam localizados em regiões genômicas, sendo 70% destes em regiões codificantes. Foi detectada forte estrutura genética de população e a distribuição dos genótipos dentro de subpopulações baseou-se principalmente em características relacionadas à fruta: polpa fundente e polpa não-fundente. Padrões de desequilíbrio de ligação (LD) sugeriram uma queda média de LD em relação à distância e a extensão do LD altamente dependente da subpopulação e das regiões do genoma. Avaliações de campo e através do bioensaio de folhas destacadas mostraram resultados confiáveis e complementares para identificar fontes resistentes à *Xap*. Os genótipos 'Norman', 'Cristal Taquari', 'La Feliciana' e 'Precocinho' foram considerados fontes altamente resistentes e podem ser alternativas efetivas para melhorar a resistência à *Xap* em pessegueiro. Em geral, o germoplasma avaliado mostrou grande variabilidade para resposta à *Xap*, permitindo a identificação de genótipos contrastantes para a característica de interesse. Os genótipos com resistência podem ser preferencialmente utilizados em áreas de produção com maior ocorrência da doença, ou utilizados como genitores no programa de melhoramento visando aumentar o nível de resistência. Análises de GWAS validaram e definiram com mais precisão as regiões genômicas identificadas com resistência ao patógeno em estudos anteriores, bem como possibilitaram a identificação de novos genes candidatos que necessitam ser melhor

estudados. Vários SNPs informativos foram funcionalmente anotados em genes envolvidos em mecanismos de defesa à infecção por patógenos, com destaque para duas regiões genômicas, localizadas no cromossomo 1 (2,59 Mpb) e 2 (2,85 Mpb), respectivamente, ambas identificadas com vários genes R. Os resultados encontrados abrem novos caminhos para o melhoramento genético visando resistência a *Xap*, com grande potencial para subsequente aplicação de seleção assistida por marcadores.

Palavras-Chave: *Prunus persica*; genotipagem por sequenciamento; fenotipagem; REML/BLUP; GWAS

Abstract

THUROW, Liane Bahr. **Genome-wide association mapping for bacterial spot resistance in peach germplasm based on SNPs.** 2018. 102f. Thesis (Doctoral degree in Agronomy - Plant Breeding) - Graduate Program of Agronomy, Federal University of Pelotas, Pelotas, 2018.

Peach (*Prunus persica*) is one of the best genetically characterized deciduous trees and the third most important temperate fruit crop worldwide. This crop is vulnerable to bacterial spot caused by *Xanthomonas arboricola* pv. *pruni* (*Xap*) and breeding for resistance has been the main choice to control the disease. With the aim to contribute with breeding for more resistant cultivars, and perform genome-wide association analysis (GWAS), we explored the use of genotyping by sequencing (GBS) for large-scale SNP discovery and simultaneous genotyping and performed high quality phenotyping for *Xap* response, among a panel with 220 peach genotypes, representative of the Brazilian breeding germplasm. A total of 93,353 SNP markers were discovered, and after filtering 18,373 high quality SNPs were used in analyses. Thirty-four percent of selected SNPs were located in genic regions and 70% of these in the coding sequence. Strong population genetic structure was detected, with the distribution of genotypes within subpopulations based mainly on fruit-related traits: melting and non-melting flesh. Linkage disequilibrium (LD) patterns suggested a medium LD level, with the extent of LD highly dependent on the subpopulations and genome regions. Field evaluation and detached leaf assessments were reliable and complementary methods to identify *Xap* resistant sources. The genotypes 'Norman', 'Cristal Taquari', 'La Feliciana' and 'Precocinho' were considered highly resistant sources and may be effective alternatives to improve *Xap* resistance in peach. Overall, the germplasm evaluated showed great variability for response to *Xap*, allowing the identification of contrasting genotypes for the trait of interest. Genotypes with resistance could be preferred for peach production areas more subjected to disease occurrence, or used as parents by the breeding program to improve resistance. GWAS analysis validated and defined more accurately the known genomic regions underlying *Xap* resistance, as well as identified novel candidate genes that provide useful targets for further investigation. Several informative SNPs were functionally annotated in genes involved in defense mechanisms against pathogen infection, highlighting two genomic regions, located on chromosome 1 (2.59 Mbp) and 2 (2.85 Mbp), respectively, both housing several R genes. Our results provide new insights into breeding for *Xap* resistance in peach, with great potential for subsequent application of marker-assisted selection.

Keywords: *Prunus persica*; genotyping by sequencing; phenotyping; REML/BLUP; GWAS

Lista de Figuras

CAPÍTULO 1: Genome-wide SNP discovery through genotyping by sequencing, population structure and linkage disequilibrium in Brazilian peach breeding germplasm.

CAPÍTULO 2: Selection of peach germplasm with resistance to *Xanthomonas arboricola* pv. *pruni* based on field and detached leaf assessments using the REML/BLUP approach.

CAPÍTULO 3: Genome-wide association mapping for bacterial spot resistance in *Prunus persica*.

Lista de Tabelas

CAPÍTULO 1: Genome-wide SNP discovery through genotyping by sequencing, population structure and linkage disequilibrium in Brazilian peach breeding germplasm.

CAPÍTULO 2: Selection of peach germplasm with resistance to *Xanthomonas arboricola* pv. *pruni* based on field and detached leaf assessments using the REML/BLUP approach.

CAPÍTULO 3: Genome-wide association mapping for bacterial spot resistance in *Prunus persica*.

Sumário

INTRODUÇÃO GERAL

O pessegueiro [*Prunus persica* (L.) Batsch] é a terceira espécie frutífera de clima temperado mais predominante em todo o mundo, com produção estimada em 22,8 milhões de toneladas por ano, em uma área de 1,5 milhões de hectares (FAOSTAT, 2014).

Assim como na maioria das espécies, programas de melhoramento de pessegueiro são continuamente confrontados com a necessidade de desenvolver plantas geneticamente superiores para solucionar os mais diversos problemas, relacionados a doenças e pragas, assim como associados à constantes alterações climáticas (VERDE et al., 2013).

A expansão da cultura para regiões subtropicais úmidas, por exemplo, só foi possível graças aos esforços do melhoramento na seleção e desenvolvimento de cultivares adaptadas a baixa necessidade de frio (BYRNE et al., 2012). No entanto, áreas quentes e úmidas, com alta frequência de ventos fortes e solos arenosos favorecem a disseminação e infecção de patógenos como *Xanthomonas arboricola* pv. *pruni* (*Xap*).

Xap pode atacar todas as espécies cultivadas de *Prunus* e seus híbridos (STEFANI, 2010). O patógeno encontra-se amplamente distribuído nas principais regiões produtoras de frutos ao redor do mundo (EPPO, 2017). Sintomas ocorrem em folhas, frutos e ramos e infecções severas causam desfolha precoce, enfraquecimento progressivo da planta, além de reduzir a produtividade e a qualidade do fruto, a ponto de inviabilizar a comercialização (RITCHIE, 1995; OEPP/EPPO, 2006; PALACIO-BIELSA et al., 2015).

O controle químico é caro e muitas vezes pouco eficiente, aumentando o interesse no desenvolvimento de cultivares resistentes ao patógeno (SACHET et al., 2013; GASIC et al., 2015).

Estudos indicam que o controle para resistência a *Xap* é de natureza genética quantitativa (SOCQUET-JUGLARD et al., 2013; YANG et al., 2013; FRETT, 2016). Desta forma, combinar alelos favoráveis apenas com seleção fenotípica é uma tarefa difícil, principalmente na seleção de múltiplos alelos de resistência e na seleção de caracteres que sofrem grande influência ambiental (BLISS, 2010; RU et al., 2015).

Nas últimas décadas, o melhoramento genético deixou de ser um processo baseado apenas em seleção fenotípica e passou a incorporar o uso de informações baseadas no DNA (VARSHNEY et al., 2014). Os consideráveis avanços da genômica, a exemplo do sequenciamento completo do genoma do pessegueiro (227.4 Mbp) (VERDE et al., 2013; VERDE et al., 2017) e o rápido desenvolvimento das tecnologias de sequenciamento de nova geração (NGS), possibilitaram o sequenciamento de múltiplos indivíduos em larga escala e a identificação de polimorfismos de único nucleotídeo (SNPs) com alta cobertura do genoma, revolucionando o melhoramento de plantas (BARABASCHI et al., 2016).

Dentre estas metodologias NGS, destaca-se a genotipagem por sequenciamento (GBS) que permite a descoberta e genotipagem simultânea de SNPs por todo o genoma, de forma rápida, simples, multiplex, com alto rendimento e baixo custo por amostra (ELSHIRE et al., 2011).

GBS têm sido utilizada com sucesso em programas de melhoramento de diversas espécies, com destaque para espécies frutíferas perenes (GARDNER et al., 2014; GUAJARDO et al., 2015; BIELENBERG et al., 2015; GÜRCAN et al., 2016; SALAZAR et al., 2017; KUMAR et al., 2017). Apresenta ampla gama de aplicações que vão desde a identificação de SNPs, caracterização de germoplasma, até mapeamento de QTLs (*quantitative trait locus*), seleção genômica (GS) e estudos de associação genômica ampla (GWAS) (HE et al., 2014).

Em estudos de GWAS, visando identificar associações significativas entre polimorfismos genéticos e a variação fenotípica, o número de SNPs requeridos e a resolução do mapeamento é determinado pela extensão do desequilíbrio de ligação (LD)

entre locos. Alta densidade de SNPs cobrindo o genoma é necessária para observar a frequência de recombinação na população (FLINT-GARCIA et al., 2003; BRACHI et al., 2011). Além disso, a estrutura genética da população, bem como os efeitos devidos à relação de parentesco devem ser considerados para evitar associações espúrias, dificultando a distinção de locos que realmente afetam o caráter de interesse (BRADBURY et al., 2007; KHAN; KORBAN, 2012).

O germoplasma de pessegueiro disponível para uso no programa de melhoramento genético apresenta grande variabilidade genética à nível de DNA (THUROW et al., 2017) e estudos anteriores relataram ampla variação fenotípica para resistência à *Xap* (RASEIRA et al., 2008; SACHET et al., 2013; PALACIO-BIELSA et al., 2015). A utilização de ferramentas genômicas (a exemplo de GBS) aliada a seleção fenotípica tradicional pode facilitar a introgressão de resistência à *Xap* no desenvolvimento de novas cultivares, assim como possibilitar a identificação de regiões genômicas associadas com resistência ao patógeno visando implementar o uso de seleção assistida por marcadores.

Frente ao exposto, o presente trabalho teve como objetivo explorar o uso de genotipagem por sequenciamento (GBS) no painel associativo de pessegueiro representativo do germoplasma disponível no Brasil, avaliar a resposta do germoplasma à infecção de *Xap* através da fenotipagem no campo e utilizando bioensaio de folhas destacadas, além de identificar regiões genômicas associadas com resistência ao patógeno utilizando análises de associação genômica ampla (GWAS).

A tese foi dividida em três capítulos, o primeiro abrange o uso de GBS para identificação de alta densidade de SNPs com cobertura total do genoma, explora a variabilidade genética, a estrutura de população e a extensão do desequilíbrio de ligação (LD). O segundo capítulo aborda a resposta do germoplasma em relação à infecção de *Xap* através de escala de classificação de severidade no campo e em bioensaio de folhas destacadas em ambiente controlado, assim como estimação dos parâmetros genéticos associados com a resistência e predição de valores genotípicos através do uso de modelos mistos (REML/BLUP). O terceiro capítulo integra os dados obtidos através de GBS e os dados de fenotipagem para resistência para implementar análises de GWAS e identificar regiões genômicas e genes candidatos associados com resistência à *Xap*.

CAPÍTULO 1

Genome-wide SNP discovery through genotyping by sequencing, population structure and linkage disequilibrium in Brazilian peach breeding germplasm

1.1 Introduction

Peach [*Prunus persica* (L.) Batsch] belongs to the Rosaceae family and is one of the best genetically characterized fruit tree species (BYRNE et al., 2012). It was domesticated in China more than 4,000 years ago, from where it was dispersed throughout the world, becoming adapted to a wide range of climates (FAUST; TIMON, 1995).

From the late 1950s, peach breeding started in southern Brazil (Pelotas, RS at lat. 31° 42’ S, long. 52° 24’ W) with the aim to develop peaches adapted to the mild winter and high relative humidity conditions (RASEIRA; NAKASU, 2006). The raw germplasm used as the basic genetic material for the breeding program, originated from seedling selections of locally adapted varieties, enriched by thousands of open-pollinated and hybrid seeds from the North American breeding programs (RASEIRA et al. 2008). At that time, only two cultivars were grown in southern Brazil, with a harvest season spanning only 15 days (RASEIRA et al., 1992). Since then, considerable improvements have been made, extending this period to more than 100 days, improved disease resistance, productivity, fruit quality and adaptation (RASEIRA; NAKASU, 2012).

The program has evaluated and crossed a wide range of local and foreign accessions. Since its inception, several cultivars were released, which are being grown extensively in the south and southeast regions of Brazil and also being used for breeding purposes. Hence, exploring this low-chill genetic variability is crucial to breed new peach cultivars adapted to current challenges. Therefore, this germplasm needs to be evaluated and used to a greater extent, for development of DNA based information to support breeding decisions.

The advances in development of genomics resources, such as the availability of the whole genome sequence (VERDE et al., 2013; VERDE et al., 2017) and the rapid development of next generation sequencing (NGS) technologies have greatly improved the understanding of the genetic base of important agronomic traits in peach. DNA based genetic markers have been extensively used in peach to characterize different germplasm panels, to evaluate diversity within breeding programs, to analyze population structure and to perform QTL mapping. Valuable genetic tools, such as 9K SNP peach array (VERDE et al., 2012) were developed and used in whole-genome diversity and QTL mapping studies (YANG et al., 2013; FRETT et al., 2014; MICHELETTI et al., 2015; FRESNEDO-RAMÍREZ et al., 2015; FRESNEDO-RAMÍREZ et al., 2016), providing resources that will facilitate the development of new varieties.

Advances in NGS technologies continuously improved the throughput and cost-efficiency, offering the possibility of shifting from pre-defined SNP panels to direct sequencing of populations of interest, producing unbiased markers across the entire genome (BARABASCHI et al., 2016). In this regard, genotyping by sequencing (GBS) has become a promising approach for comprehensive genotyping on a genome-wide scale (ELSHIRE et al., 2011; PETERSON et al., 2014). GBS uses enzyme-based complexity reduction and allows simultaneous marker discovery and genotyping across a whole germplasm set of interest, with and without reference genomes (ELSHIRE et al., 2011; POLAND; RIFE, 2012). It has been used in a wide number of crop species, providing new opportunities for breeders (KIM et al., 2016) and successfully applied in perennial tree fruit species such as apple (GARDNER et al., 2014), sweet cherry (GUAJARDO et al., 2015), peach (BIELENBERG et al., 2015), apricot (GÜRCAN et al., 2016), japanese plum (SALAZAR et al., 2017) and pear (KUMAR et al., 2017).

In the current study, we explored suitability of GBS as a genome scan to discover and genotype SNPs in a panel composed of 220 diverse peach cultivars and advanced selections from the Brazilian Agricultural Research Corporation (Embrapa) peach breeding program. We provide a comprehensive overview of population structure, variability parameters and detailed estimation of LD decay patterns. SNP markers reported here will enable better understanding and utilization of variability and population stratification in the available germplasm, support breeding decisions and provide basis for genome-wide association studies (GWAS) in peach.

1.2 Material and methods

1.2.1 Plant Material

A panel of 220 diverse peach cultivars and advanced selections [*Prunus persica* (L.) Batsch] conserved at Embrapa (Pelotas-RS, located at lat. 31°42'S, long. 52°24'W and an altitude of 57m above sea level), was used in this study (Supplementary Table S1). All peach trees were grafted on seedling rootstocks of *P. persica*, spaced 3m between trees and 5m between the rows, trained as open center system and standard horticultural practices were applied.

This panel was selected to represent the whole germplasm available for the breeding program, based on prior knowledge of contrasting phenotypes to bacterial spot and brown rot resistance and tolerance to abiotic stresses such as heat tolerance at the flowering stage and chilling requirement.

Accessions adapted to the low and medium chill zones mainly composed this selected panel. In addition, cultivars from different peach breeding programs and some foreign cultivars (United States, Bolivia, Spain, Italy, Canary Islands, Mexico, Japan, Uruguay and Taiwan) were also included in this study.

1.2.2 DNA isolation, library preparation and sequencing

Genomic DNA was isolated from powdered, freeze-dried young leaves of all the 220 *P. persica* accessions, following the miniprep method based on Dellaporta et al. (1983). DNA samples were first quality tested with a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies; Wilmington, DE, USA) and then quantified using Hoechst 33258 (Sigma-Aldrich; St. Louis, MO, USA), against a λ standard DNA dilution series, with a Synergy H-T fluorimeter (BioTek; Winooski, VT, USA). Finally, DNA concentrations were normalized to 10ng μl^{-1} and subsequently used for library preparation.

Ninety-six-plex libraries comprising each 95 peach DNA samples and a negative control (no DNA) were prepared according to the 'genotyping by sequencing' (GBS) method, as described by Elshire et al. (2011). High quality genomic DNA (100ng) from individual samples was digested with the *Ape*KI methylation sensitive restriction enzyme and barcode adapters were ligated to the ends of genomic DNA fragments. The adapters comprised a set of 96 forward adaptors each with a unique barcode and a single common reverse adaptor. Oligonucleotide sequences of the *Ape*KI barcode adapters, used for multiplex sequencing, were those provided in Elshire et al. (2011).

After adapter ligation, DNA samples were pooled by plate into a single library and clean-up with QIAquick PCR Purification Kit (Qiagen; Valencia, CA, USA), according to the manufacturer's instructions. Each library in duplicate was amplified by PCR (polymerase chain reaction), to selectively enrich for those fragments in the library with adapters at both ends, PCR products were purified again, as above, and combined to get more concentrated library. Finally, pooled, amplified libraries were submitted to David H. Murdock Research Institute (DHMRI; Kannapolis, NC, USA) for sequencing. Aliquots of all libraries were run on the Agilent BioAnalyzer/High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA) to check for library fragment sizes and presence/absence of unwanted peaks of adapter and primer dimers. Libraries were then sequenced with single-end reads on the Illumina HiSeq2000 platform.

1.2.3 Processing of raw sequence data and SNP calling

The multiple steps of GBS analysis pipeline were carried out using default parameters implemented in TASSEL v4.0 GBS pipeline (BRADBURY et al., 2007; GLAUBITZ et al., 2014).

Filtered sequence reads were aligned to the peach genome (Peach v2.0.a1, Verde et al. 2017) (https://www.rosaceae.org/species/prunus_persica/genome_v2.0.a1) as a reference using Bowtie v2.1 (LANGMEAD; SALZBERG, 2012), and SNP calls were exported as vcf files. Further, raw SNPs were imported into TASSEL v5.0 GUI (www.maizegenetics.net) for SNP and taxa filtering based upon coverage and SNP filtering based on minor allele frequency (MAF). The pipeline parameters used to filter the raw SNPs were as follows, MAF > 0.05 and a 25% threshold for maximum missing data.

1.2.4 Summary statistics and genetic variability

Distribution of SNP markers across the peach genome was calculated using Geno Summary analysis tool in TASSEL v5.0 software (BRADBURY et al., 2007) as well as filtering for genotype quality and missing data using taxa and site filter tools. Deviations from Hardy-Weinberg equilibrium (HWE) and MAF were assessed for each SNP using PLINK v.1.9 (PURCELL et al., 2007).

Estimations of observed heterozygosity (Ho), expected heterozygosity (He) and inbreeding coefficient (F) were performed for each genotype using the “--het” option in PLINK, as described in Anderson et al. (2010).

1.2.5 Structural annotation of SNPs

The SNP physical position in base pairs (bp) identified by GBS approach was correlated with the GFF3 file containing the peach genome annotation (Peach v2.0. a1, VERDE et al., 2017). Analysis based on the genomic distribution of each SNP within genic (exons, introns and untranslated regions - UTRs) and intergenic regions (‘noncoding DNA’

surrounding genes containing promoters and other regulatory elements) were performed using custom scripts written in Perl language.

1.2.6 Population genetic structure

To investigate patterns of population structure, both Principal Coordinate Analysis (PCoA) and a Bayesian clustering approach were used. Multidimensional scaling (MDS) through PCoA was performed on the distance matrix (1-IBS) generated by PLINK v.1.9 (PURCELL et al., 2007) using the “cmdscale” function from R (v 3.3.1).

In a second approach, population stratification was inferred using fastSTRUCTURE (RAJ et al., 2014). Due to the program assumptions, SNPs with more than 25% missing data and with a MAF < 0.05 were removed, before the analysis. To avoid the strong influence of tightly linked SNPs, LD-based SNP pruning in PLINK 1.9 (PURCELL et al., 2007) was also used to generate a final subset of SNP markers that are in approximate linkage equilibrium with each other. Pruning was performed using the “indep” function by defining a window of 50 SNPs, with a step size of five SNPs to shift the window and a variance inflation factor (VIF) threshold of 2 (which recursively removes SNPs within the sliding window if $R^2 > 0.5$). LD-Pruned SNPs were converted to “plink bed” format using “make-bed” option in PLINK v.1.9 (PURCELL et al., 2007) and output files (.bed, .bim, .fam) used as input in the fastSTRUCTURE. This software implements a Bayesian framework similar to STRUCTURE with variational algorithms that allows inferring population structure in large SNP data sets. FastSTRUCTURE was run with a “simple prior” option and remaining default parameters. K (number of populations) values, ranging from 1 to 10 were tested and the most probable number of populations was chosen running the build-in script for multiple choices of K (RAJ et al., 2014). The admixture proportions of each genotype, estimated by fastSTRUCTURE, were visualized using DISTRUCT plots (ROSENBERG, 2004). Accessions were assigned to a specific subpopulation when the estimated membership coefficients (Q) were above 0.75.

Pairwise fixation index (F_{st}) estimates among the subpopulations identified by fastSTRUCTURE were calculated with R package adegenet 2.0.1 (JOMBART; AHMED, 2011), using Nei’s (1972) distance.

1.2.7 Linkage disequilibrium (LD)

Genome-wide LD analysis were performed in the whole panel, as well as separately in the subpopulations inferred by fastSTRUCTURE and in the admixed genotypes. To assess LD, SNPs with more than 25% missing data and with a MAF < 0.05 were discarded for each of the genotype datasets. Additionally, the extent of LD was calculated across each of the eight peach chromosomes using whole panel genotype data.

LD was estimated using the squared correlation (r^2) based on genotype allele counts between pairs of SNP markers as implemented in PLINK v.1.9 (PURCELL et al., 2007). The plink command was set to measure pairwise LD between 50 subsequent SNPs among a distance of 2,000 kbp. Intrachromosomal r^2 values were then plotted against the physical distance (Kbp) and LD decay fitted using a locally weighed polynomial regression (LOESS) curve obtained with R software (v 3.3.1).

1.3 Results

1.3.1 Sequencing and SNP identification

Within our 220 *Prunus persica* accessions, Illumina sequencing generated total of 314.8 Mbp of sequence, with an average 1.45 million single-end 100 bp reads per genotype. Three genotypes (‘Linda’, ‘Ingo’ and c.2009.173.33) showed low initial read numbers and were removed for further analysis, resulting in a final germplasm set composed of 217 peach genotypes with all accessions having between 524,092 and 3,663,145 reads.

Out of the 1,324,595 high-quality filtered GBS sequence tags, the alignment analysis indicated that 388,803 (29.35%) had no match within the peach reference genome (Peach v2.0). Of the 70.65% GBS fragments that aligned perfectly to the peach genome, 602,068 (45.45%) aligned to single genomic locations while 333,724 (25.20%) aligned to multiple positions due to the presence of repetitive DNA.

The TASSEL GBS pipeline initially generated a total of 93,353 SNP markers, evenly distributed through the eight major scaffolds of the peach genome, across all 217 peach genotypes analyzed. The number of SNPs on each scaffold ranged from 18,323 on scaffold 1 to 8,447 on scaffold 5. The proportion of missing SNP markers ranged from 42.88 % on scaffold 5 to 45.69 % on scaffold 8, with a mean value of 44.16 %, whereas the proportion of heterozygous SNP markers ranged from 13.23 % on scaffold 2 to 15.63 % on scaffold 7, with a mean value of 14.42 % (Table 1).

Of the 93,353 SNPs, after depth and quality filtering, 18,373 SNP markers were selected, with MAF > 0.05 and a call rate of 75%. These 18,373 SNPs were used for subsequent analysis, and correspond to an average of 80.8 SNPs/Mbp (considering peach whole-genome size ~227.4 Mbp) which is equivalent to 35.4 SNPs/cM of the *Prunus* reference map (519 cM; according to Dirlewanger et al. (2004)). The greatest number of SNPs was observed on scaffold 1 (3,513 SNPs), while the lowest was detected on scaffold 5 (1,780 SNPs). Chromosome 2 showed the lowest heterozygosity rate (Ho = 0.267) while the highest heterozygosity was detected on chromosome 7 (Ho = 0.313). Detailed information of the 18,373 SNP markers is provided in table 1. After applying stringent filtering quality scores, the proportion of missing SNP markers decreased from 44.16 % to 7.54 % and the proportion of heterozygous SNPs increased from 14.42 to 29.36 %.

Table 1 - Distribution of single-nucleotide polymorphisms (SNPs) across the 8 major scaffolds of the peach genome, missing and heterozygosity rate including SNP panels with 93,353 (93K) and 18,373 (18K) SNPs, respectively.

Scaffold	Panel of 93,353 SNPs (93K)			Panel of 18,373 SNPs (18K)		
	Number of SNPs	Missing rate (%)	Heterozygosity rate (%)	Number of SNPs	Missing rate (%)	Heterozygosity rate (%)
1	18,323	43.08	14.19	3,513	7.19	29.86
2	13,525	44.98	13.23	2,568	7.32	26.70
3	10,963	44.83	14.67	2,264	8.18	28.83
4	11,432	45.14	14.87	2,261	7.32	30.03
5	8,447	42.88	14.81	1,780	7.29	30.35
6	12,037	43.31	13.89	2,290	7.80	29.66
7	8,950	43.38	15.63	1,837	7.38	31.27
8	9,676	45.69	14.05	1,860	7.85	28.14
Total	93,353	-	-	18,373	-	-
Average	-	44.16	14.42	-	7.54	29.36

Deviations from Hardy-Weinberg equilibrium (HWE) were tested for all 93,353 SNP markers and 60.13 % failed HWE test at $p < 0.001$ cutoff. For the 18K SNP panel 48.54 % of the markers showed significant deviation from HWE.

The site frequency analyses for all 217 peach genotypes revealed 40 % of SNPs in the 93K panel, with MAF < 0.1, and 25 % of those SNPs being rare (MAF < 0.05) (Figure 1a). The 18K SNP panel, accounted for 27 % of the total SNPs, exhibited similar distribution of SNPs per MAF category after removal of SNPs with MAF < 0.05 (Figure 1b).

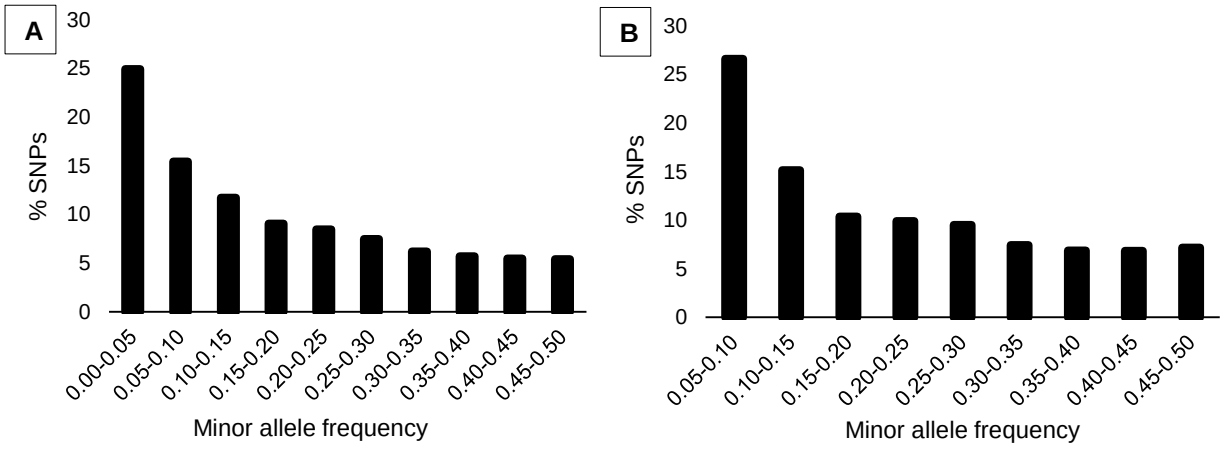


Figure 1 - Frequency distribution of the minor alleles (MAF) in 217 peach genotypes based on: A) 93,353 SNPs and B) 18,373 SNPs genotype datasets, respectively.

Considering all 217 peach genotypes analyzed with 18,373 SNP markers, the observed mean heterozygosity (Ho) rate per genotype was 0.291, ranging from 0.133 in 'Cristal-Taquari' to 0.462 in 'Turmalina'. The average expected heterozygosity (He) was 0.305, with an average inbreeding coefficient per individual (F=1-(Ho/He)) of 0.045, ranging from -0.519 to 0.562.

1.3.2 Structural Annotation of SNPs

The structural annotation of 18,373 genome-based SNPs was determined by comparing the physical position of each SNP against the annotated peach genome (Peach v2.0. a1). Results showed 6,268 SNPs (34%) covering 2,127 genes, with an

average frequency of 2.9 SNPs/gene, and 12,105 SNPs (66%) in the intergenic regions (Figure 2a).

The annotation of SNPs based on their structural components of genes revealed a higher percentage of SNPs in exonic regions (70%), followed by intronic regions (15%). The remaining 15% SNPs were situated in untranslated regions, 5'UTR (9%) and 3'UTR (6%) (Figure 2b). Most of the allelic variation was structurally annotated in non-coding sequence components with only 24% discovered in coding gene regions.

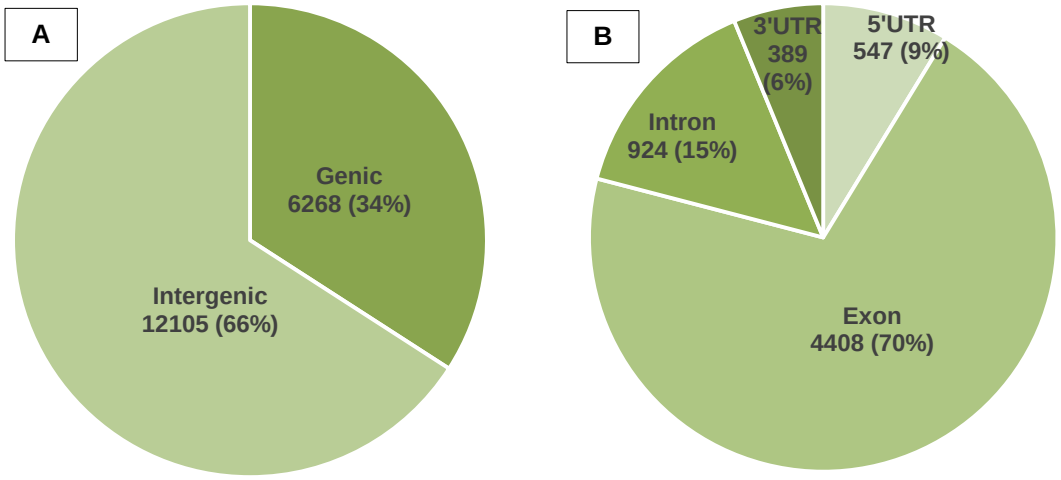


Figure 2 - Distribution of SNPs in different genomic regions using the physical location of each SNP on Peach v2.0. a1. A) Structural occurrence in genic and intergenic regions. B) Distribution of SNPs in genes (exonic, intronic and untranslated regions).

1.3.3 Population Structure

Population structure and correlation among the 217 peach genotypes was addressed using two different approaches (MDS and fastSTRUCTURE). First approximation of population stratification was obtained using multidimensional scaling through PCoA for all 93,353 SNPs, which provided evidence of genetic variation among genotypes in accordance with morphological traits related to the fruit flesh type: melting and non-melting genotypes (Figure 3).

In a second approach, to obtain a more detailed stratification in the germplasm panel, we used the software fastSTRUCTURE. Patterns of population structure were evaluated using 5,378 genome-wide and unlinked SNPs which, based on model

complexity that maximizes marginal likelihood, estimated the most likely number of populations at K = 3 (Figure 4).

Considering the grouping threshold of Q > 0.75, the three inferred subpopulations were: POP I (78 genotypes), POP II (1 genotype) and POP III (62 genotypes). The remaining 76 genotypes were classified as admixed (ADM), since they had less than 75 % shared ancestry with one of the three main distinct subpopulations. Individual's assignment to these subpopulations is provided in Supplementary Table S1.

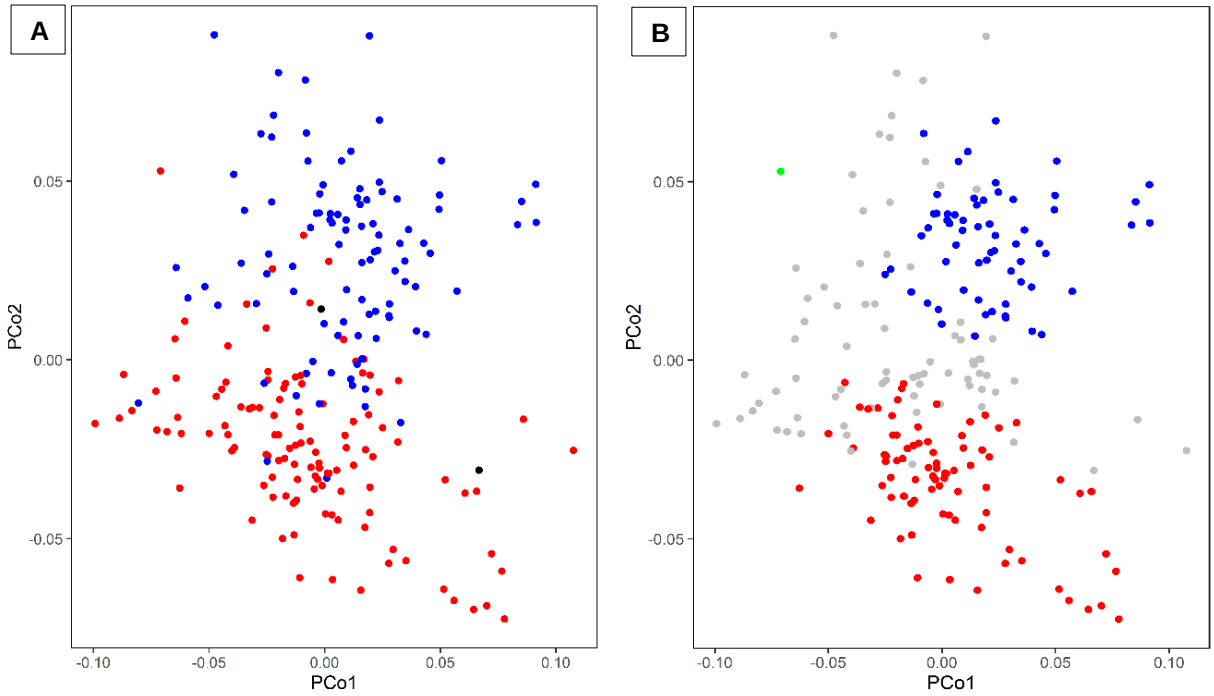


Figure 3 - Principal Coordinate Analysis (PCoA) based on the 93,353 genome-wide SNPs among 217 *Prunus persica* genotypes. A) Principal coordinate plot overlaid with fruit morphology traits: *red* represents melting genotypes and *blue* represents non-melting genotypes, based on phenotypic evaluations by the peach breeding program, in *black* genotypes with unknown fruit flesh texture. B) The genotypes are colored with respect to the three subpopulations inferred by fastSTRUCTURE analysis (membership coefficient > 0.75): POP I (*red*), POP II (*green*) and POP III (*blue*). *Light grey* indicates unstructured genotypes (ADM).

Both PCoA and fastSTRUCTURE analyses revealed population grouping based mainly on the fruit traits of accessions: melting and non-melting flesh (Figure 3b). POP I accounts for majority of the melting genotypes. Such cultivars are used mainly for fresh market and include: advanced selections and cultivars released by EMBRAPA breeding program (51.3%), cultivars from Agronomic Institute of Campinas (IAC) (3.8%) and majority of the introductions from North American peach breeding programs (30.8%), as

well as additional few cultivars from Japan, Mexico, Spain, Italy and Taiwan (10.2% in total) and genotypes with unknown origin (3.8%). POP II included only 'Mollares Hierro', the melting peach from Canary Islands, while POP III comprised majority of the non-melting cultivars and advanced selections bred for processing purpose by Embrapa's program (90.3%), few accessions from IAC (4.8%) and few introduced from Bolivia (4.8%). At K = 3, it was clear that the distribution of the genotypes in populations reflected the fruit-type characteristics. Increasing the number of populations (K) from three to five had almost no difference between K-values, and maintained the membership in the initial three populations almost invariable with additional populations empty under a membership coefficient above 0.75.

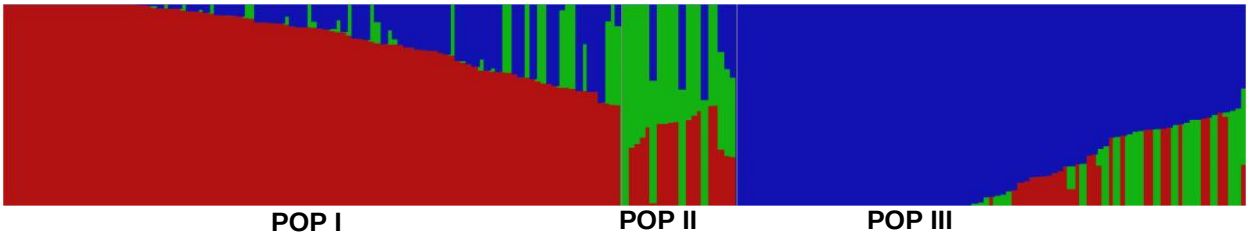


Figure 4 - Genome-wide SNP-based population genetic structure among 217 *Prunus persica* genotypes, inferred by fastSTRUCTURE at K = 3. Each genotype is shown as a vertical bar.

Genetic differentiation between the populations inferred by fastSTRUCTURE was tested using Fst statistics estimated from pairwise analysis. Pairwise Fst values ranged from 0.011 (between POP I and ADM genotypes) to 0.107 (between POP I and POP II) (Table 2).

Table 2 - Pairwise Fst among populations identified by fastSTRUCTURE.

	ADM	POP I	POP II
POP I	0.011		
POP II	0.078	0.107	
POP III	0.013	0.014	0.099

Summarized statistics of cultivar genetic variability, including observed and expected heterozygosity rates and inbreeding coefficients were also provided for each population and for the entire germplasm panel, with all 217 genotypes (Table 3). POP II was removed before analysis, since it had just one genotype assigned. POP I exhibited higher level of heterozygosity (Ho = 0.286) than POP III (Ho = 0.259). The average

inbreeding coefficient (F) was 0.061, 0.15 and -0.06 for POP I, POP III and ADM, respectively.

Table 3 - Summary statistics of genetic variability within each subpopulation inferred by fastSTRUCTURE and within the whole panel including all 217 genotypes.

	Observed Heterozygosity (Ho)	Observed Heterozygosity (Ho) range	Expected Heterozygosity (He)	F coefficient	F coefficient range
POP I	0.286	0.176 to 0.385	0.305	0.061	-0.259 to 0.419
POPIII	0.259	0.133 to 0.355	0.305	0.150	-0.168 to 0.562
ADM	0.323	0.164 to 0.462	0.305	-0.060	-0.519 to 0.461
Whole panel	0.291	0.133 to 0.462	0.305	0.045	-0.519 to 0.562

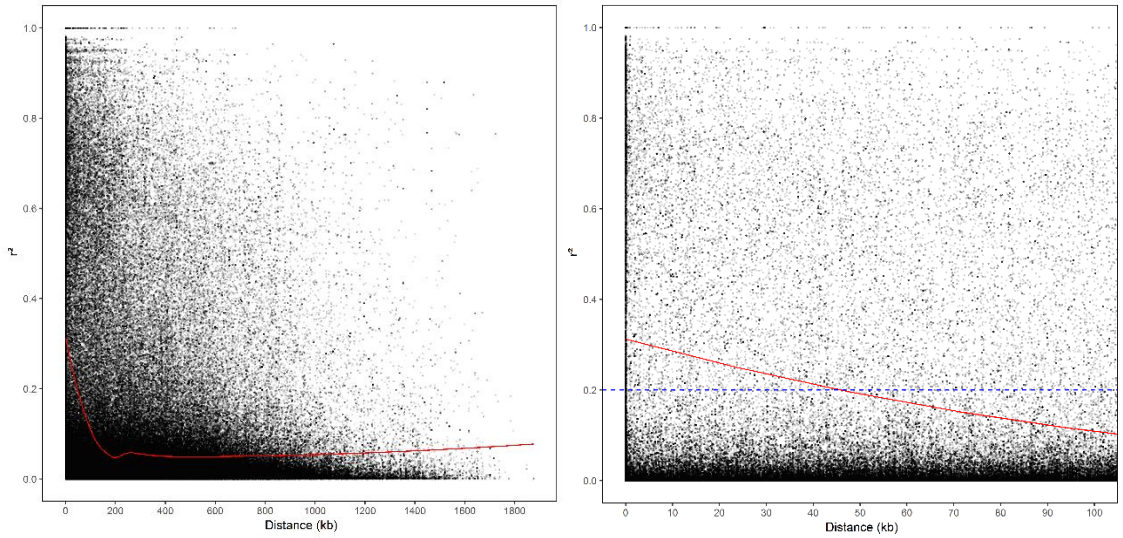
1.3.4 Linkage disequilibrium (LD)

The LD estimates (measured as r^2) and extent of LD decay were calculated using SNP markers with MAF > 0.05 and less than 25% missing data. Pairwise r^2 was measured using 17,505 SNP markers for POP I, 14,098 for POP III and 24,276 for ADM genotypes. LD was also calculated for the entire panel using 18,373 SNP markers and within chromosomes of the entire panel. LD was not estimated for POP II due to only one genotype assigned to this population.

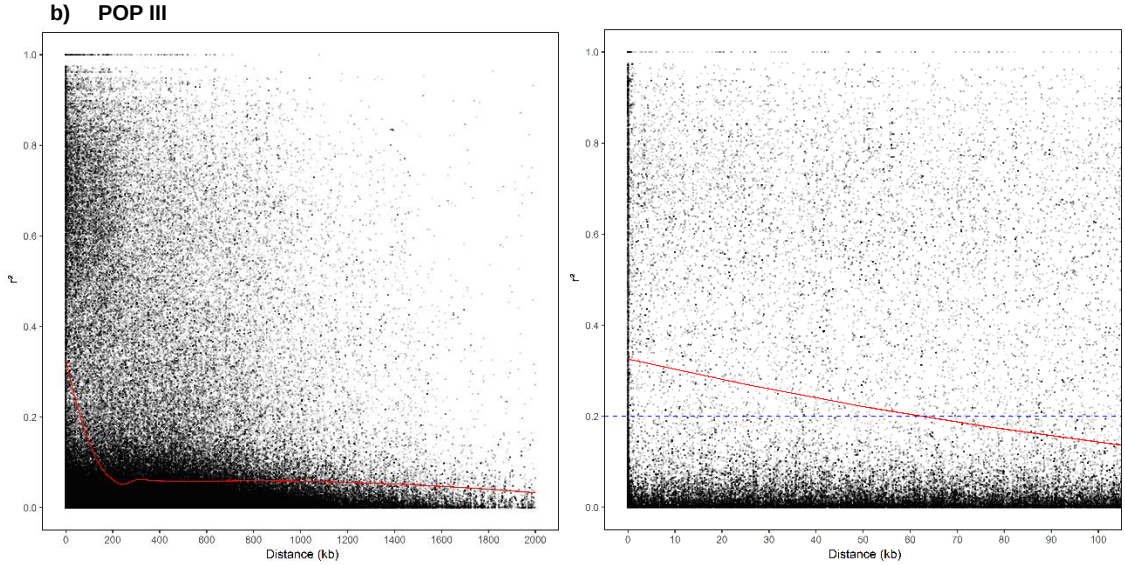
The average value estimated of intra-chromosomal r^2 was 0.093 in POP I, 0.101 in POP III and 0.099 in ADM genotypes. On the other hand, the average value for inter-chromosomal r^2 was smaller, showing values of 0.020 and 0.027 for POP I and POP III, respectively.

On the average, intra-chromosomal LD displayed differential patterns among different populations declining below 0.2 at around 46.5 Kbp in POP I, at 63.2 Kbp in POP III and at 34.3 Kbp in ADM genotypes (Figure 5). For the entire germplasm panel this critical r^2 value was observed within a distance of about 38 Kbp (Figure 5d).

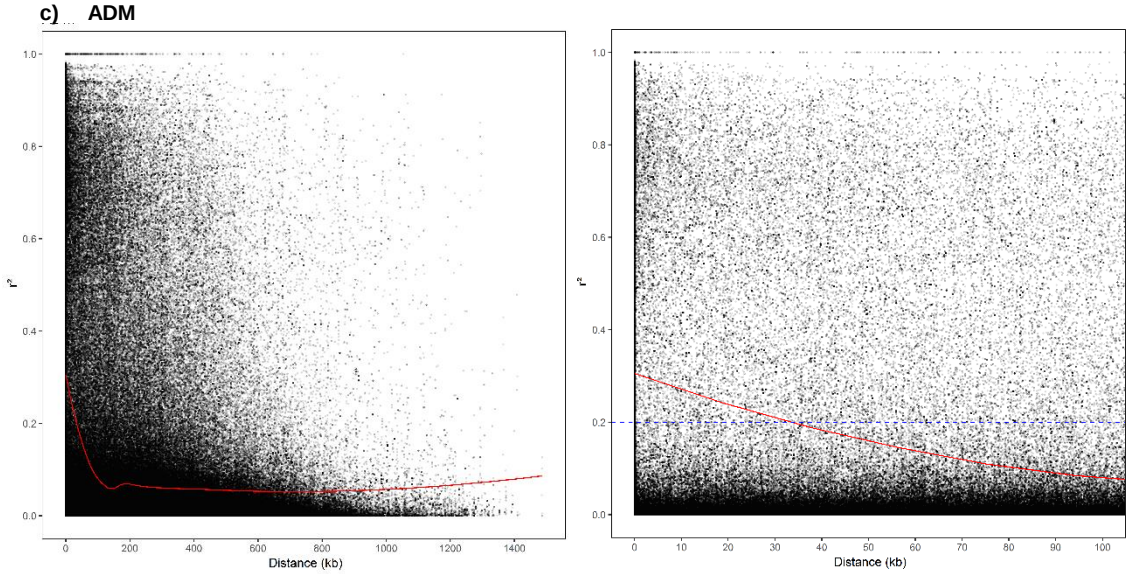
a) POP I



b) POP III



c) ADM



d) All 217 genotypes

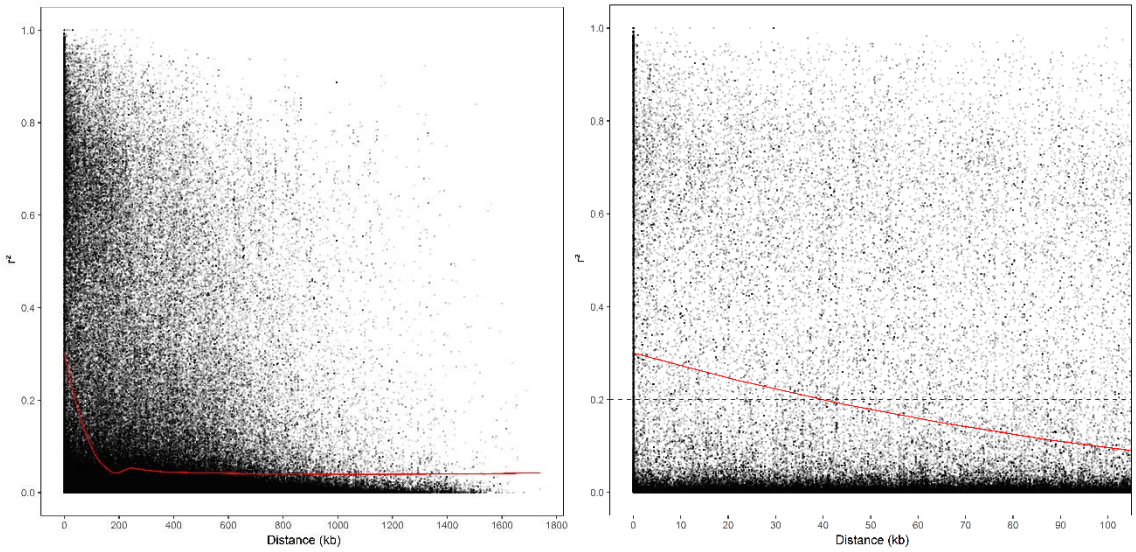
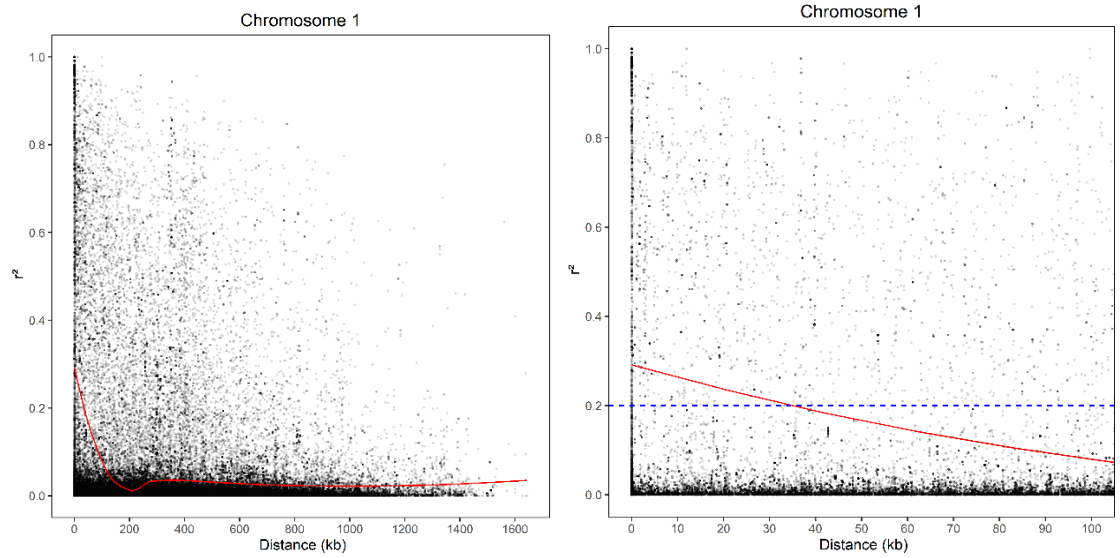
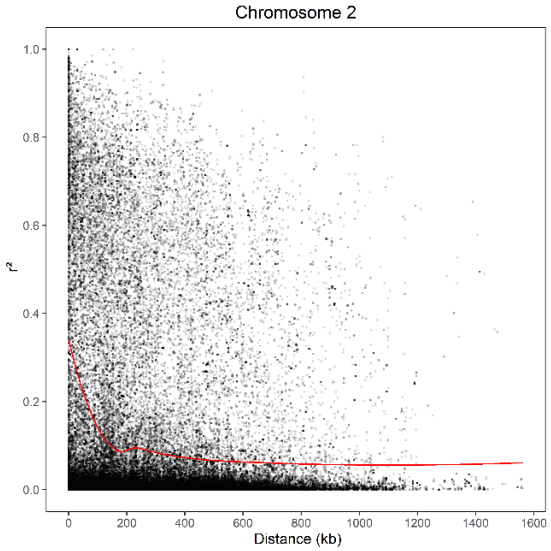
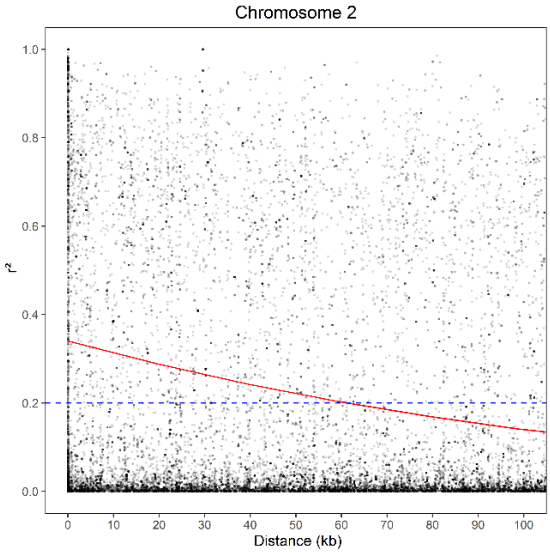


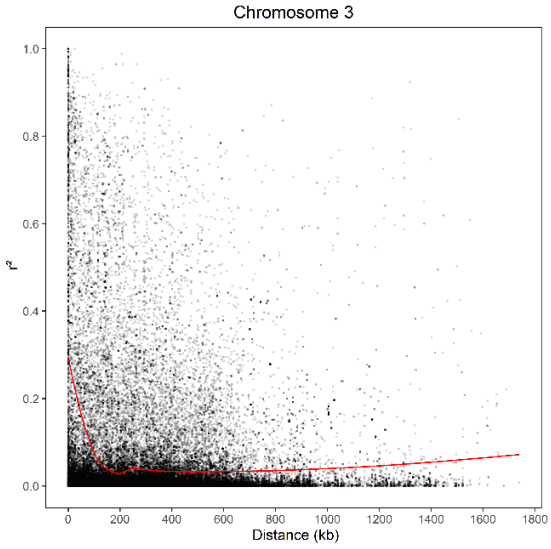
Figure 5 - Linkage disequilibrium measures (r^2) against physical distance between pairs of SNP markers for POP I, POP III, ADM genotypes and all 217 genotypes (left-hand side). Zoom-in figure of LD decay until a 100 Kbp physical distance (right-hand side). The red line represents the LOESS fitting curve of LD decay. The horizontal dashed line indicate a fixed r^2 value of 0.2.

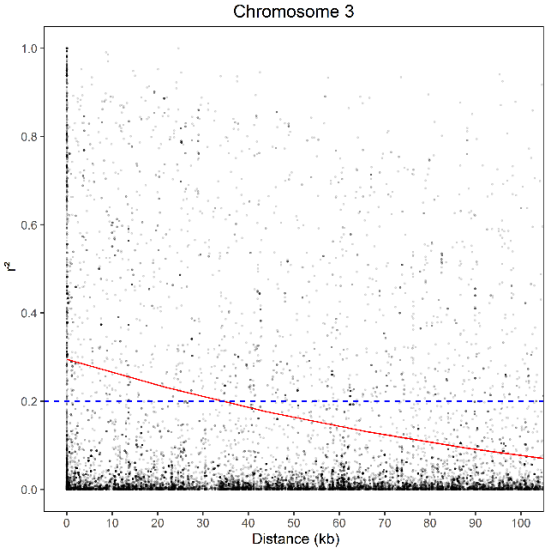
Different LD patterns have also been observed between the eight peach chromosomes (Figure 6). LD declined below 0.2 over the shortest distance on chromosome 5 (around 23 Kbp) while the greatest distance for LD decay was observed on chromosome 4 (76 Kbp) which decayed to its half value (r^2 = 0.1) at around 240 Kbp.

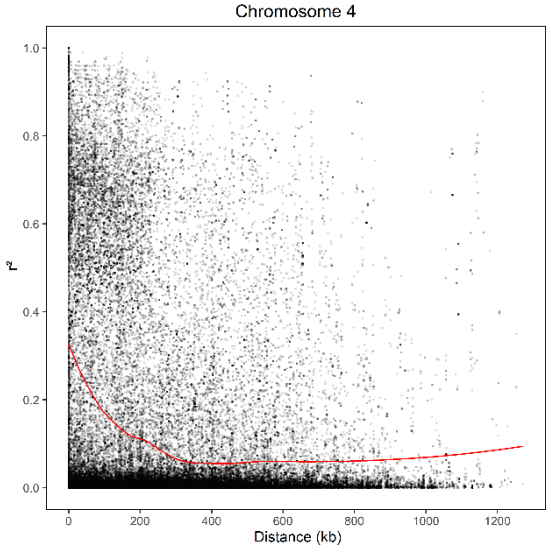


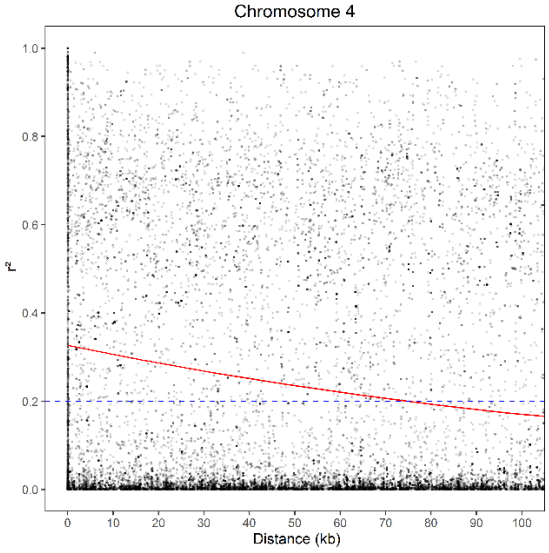

Chromosome 2
r²
Distance (kb)

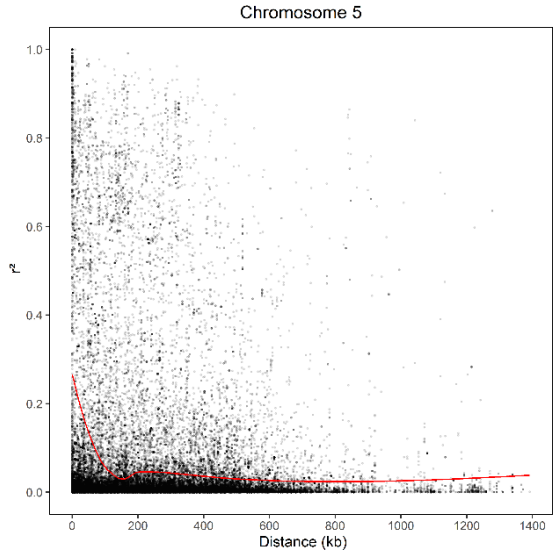

Chromosome 2
r²
Distance (kb)

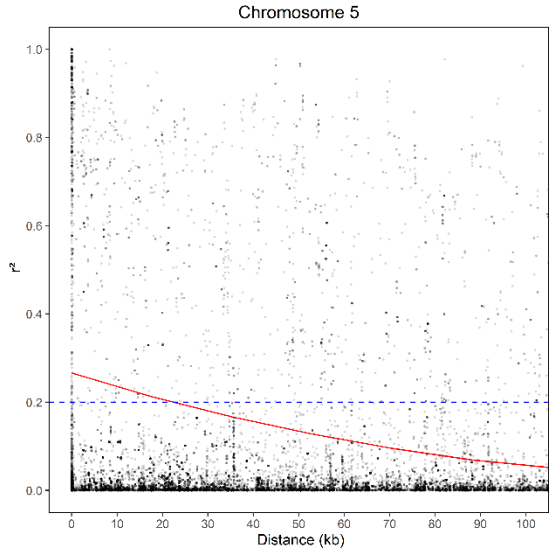

Chromosome 3
r²
Distance (kb)

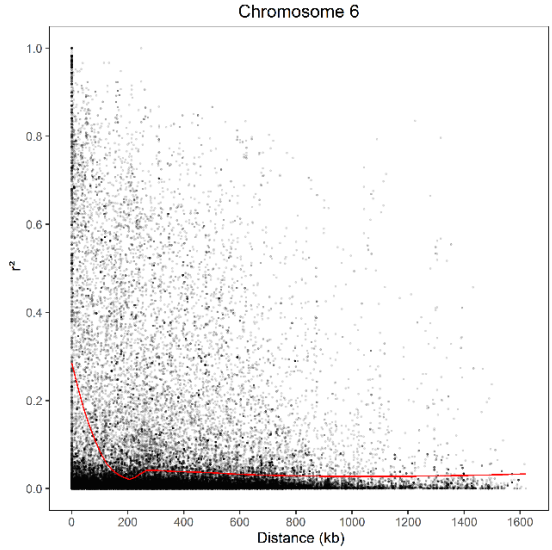

Chromosome 3
r²
Distance (kb)

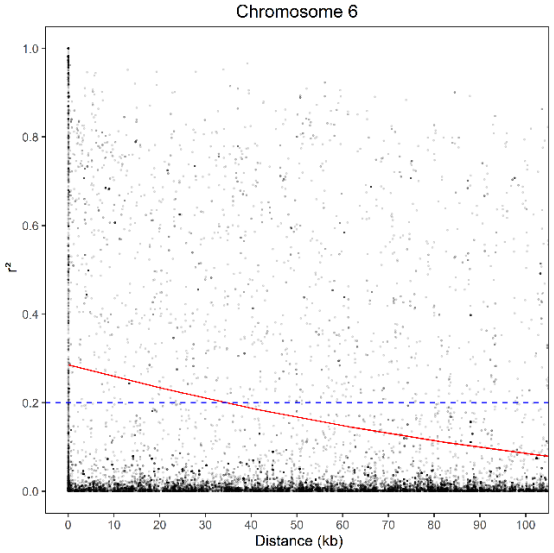

Chromosome 4
r²
Distance (kb)

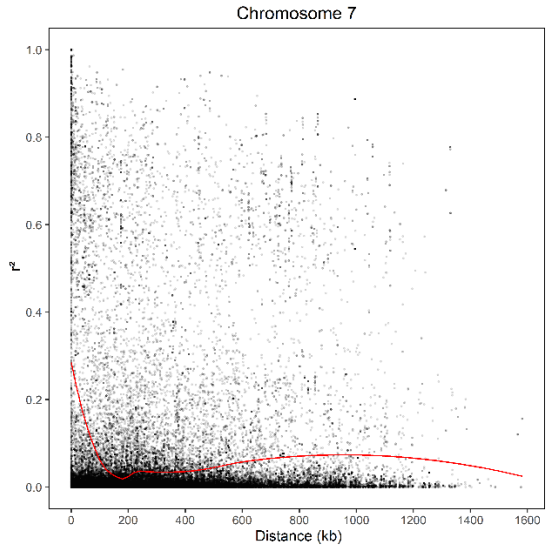

Chromosome 4
r²
Distance (kb)

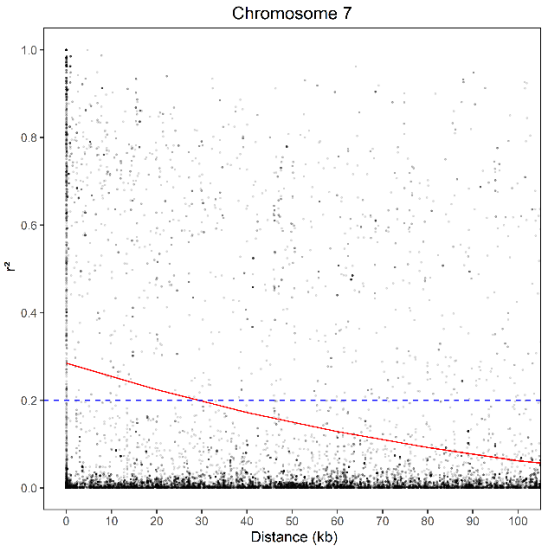

Chromosome 5
r²
Distance (kb)


Chromosome 5
r²
Distance (kb)


Chromosome 6
r²
Distance (kb)


Chromosome 6
r²
Distance (kb)


Chromosome 7
r²
Distance (kb)


Chromosome 7
r²
Distance (kb)

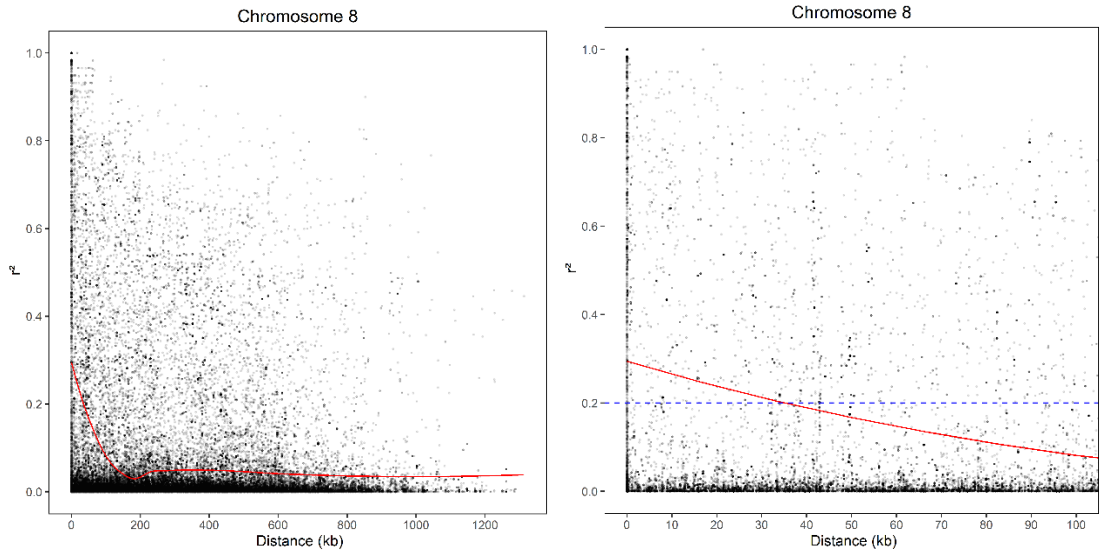


Figure 6 - Linkage disequilibrium measures (r^2) against physical distance between pairs of SNP markers along each peach chromosome (left-hand side). Zoom-in figure of LD decay until a 100 Kbp physical distance (right-hand side). The red line represents the LOESS fitting curve of LD decay. The horizontal dashed line indicates a fixed r^2 value of 0.2.

1.4 Discussion

This study provides the first attempt to evaluate genetic variation among Brazilian peach germplasm on a genome-wide scale. Here we report the use of GBS, a rapid, high throughput and cost-effective tool to increase the breeding efficiency in perennial tree fruit species (BADENES et al., 2016). Furthermore, GBS offers several advantages, since no preliminary sequence information is required and all newly discovered markers originate from the germplasm being genotyped, thus removing ascertainment bias (DESCHAMPS et al., 2012). Despite its benefits, GBS is a low coverage sequencing technology, which results in a high missing rate and large number of SNPs with very low frequency. These issues can be solved by increasing sequencing depth or by using data filtration tools.

We successfully applied the GBS approach to generate a high-density SNP coverage across the entire peach genome. The 93,353 SNPs obtained among the 217 peach genotypes constitute an important genomic resource and will facilitate peach breeding efforts. Additionally, 18,373 high-quality SNPs, obtained after filtering, will enable the use of genome-wide association studies, marker-assisted selection and genetic diversity analysis.

Structural annotation of the 18K SNPs showed approximately three times less SNPs in genic regions than in intergenic regions, probably because genic regions are evolutionarily more conserved compared to intergenic regions, which evolve faster and accumulate higher level of polymorphism. However, SNPs located in intergenic regions can also be functional harboring promoters and regulatory elements.

Complementary information obtained from PCoA and fastSTRUCTURE, indicated that three distinct populations define the genetic variation of the *P. persica* germplasm analyzed and available at EMBRAPA. Population stratification clearly supported the separation between melting and non-melting flesh cultivars/selections. Previous studies with SSR markers (ARANZANA et al., 2010; LI et al., 2013; CHAVEZ et al., 2014; THUROW et al., 2017) and, more recently, using SNPs (MICHELETTI et al., 2015), also reported strong population stratification in *Prunus persica* with similar grouping.

Our findings are consistent with the peach breeding history in Southern Brazil, reinforcing the use of different genetic resources for the development of peach cultivars for fresh consumption and canning purposes in early breeding (BYRNE et al., 2000; BYRNE, 2003; RASEIRA et al., 2003). The use of a limited number of principal founders in the breeding program has driven the concomitant formation of distinct subpopulations seen in our germplasm panel.

POP I represents the melting peach genotypes derived from locally adapted cultivars such as the founder ‘Delicioso’ and most of the germplasm introduced from North American breeding programs including also the nectarine founder ‘Panamint’. While POP II includes only ‘Mollares Hierro’, a melting peach introduced from Canary Islands and not often used by the breeding program, due to its long cycle from blooming to fruit harvest and high susceptibility to fungal diseases. On the other hand, POP III groups essentially the germplasm of non-melting peaches derived from locally adapted cultivars including the founder varieties ‘Aldrighi’ and ‘Abóbora’, which also grouped in this population.

The fact that the melting-flesh advanced selections and cultivars released by Embrapa program grouped with majority of the cultivars introduced from North American breeding programs suggests a common gene pool for the development of fresh consumption cultivars. These results are in agreement with the active exchange of germplasm between the breeding programs of these two countries and the use of North

American cultivars as basis in Embrapa breeding efforts (RASEIRA; NAKASU, 2006). In addition, majority of the nectarines (20 from 32 nectarines in total) were also included in the same population as melting-flesh peaches (POP I), indicating no separate breeding efforts were maintained for these two fruit types. One of the reasons might be the small fruit size of nectarine under subtropical conditions, suggesting that the design of crosses accounted for useful variability from both fruit types.

The low mean observed heterozygosity rate per genotype for the whole germplasm panel analyzed of Ho = 0.29 was similar to that reported for a larger collection of 1,580 Occidental and Oriental peach accessions, Ho = 0.28 (MICHELETTI et al., 2015). Furthermore, genetic variability calculated for each of the populations showed relevant differences for observed heterozygosity rates between POP I (Ho = 0.286) and POP III (Ho = 0.259). This could be due to their genetic background. At the establishment of the Embrapa program, breeding of melting peach cultivars used as basis several introductions from different countries and breeding programs, whereas the non-melting peach breeding efforts utilized mainly local varieties. Thus, the gene pool to select genotypes for processing was probably smaller than the gene pool available to select peaches for fresh consumption. In addition, worldwide, the number of peach breeding programs that breed non-melting flesh types adapted to low chill areas are only a few compared to melting fresh market peach breeding programs, limiting germplasm exchange.

Pairwise Fst indicated that there were expressive genetic differences between the populations, especially pairwise comparisons with POP II. Our results confirm that this introduction from Canary Islands represents a different germplasm source from those commonly used for breeding in Brazil, since this genotype did not cluster with any other genotype.

The degree of population structure influences LD patterns within the genome. A rapid LD decay was observed in all populations. As expected, LD decayed faster with distance in the ADM population, once those genotypes retain variability from both melting and non-melting founders with different allele frequencies and recombination rates that weakens LD (FLINT-GARCIA et al., 2003). POP III displayed a much slower LD decay than POP I, probably due to the lower genetic variability archived in the first population. Inbreeding contributes to LD maintenance over distance, limiting the effective

recombination rates. Previous studies have measured LD decay over distance in different peach germplasm collections with different low-to-medium density markers. Several studies estimated a high level of LD conservation in peach (ARANZANA et al., 2010; CAO et al., 2012; FONT I FORCADA et al., 2013; LI et al., 2013; CHAVEZ et al., 2014). However, those authors used SSR markers, therefore their results are not comparable to those reported in this study. Micheletti et al. (2015) reported average LD distance values varying from 0.8 to 1.8 Mbp in different populations using SNP markers, with an average decay of 1.4 Mbp in Occidental accessions used in modern breeding programs.

LD estimates in this study based on 18K SNPs indicated a more rapidly decay than previously reported in peach, declining below 0.2 within a distance of about 38 Kbp in the entire germplasm panel. One reason for the divergence in LD patterns could be the much larger number of gene regions covered in our study. Differences can also be partly explained by different germplasm sets analyzed and the methods used for estimating LD decay distances.

Our results of LD decay were similar of those achieved in a large-scale sequencing study of 84 Chinese *Prunus* accessions, representing the majority of the ecotypes in the world (CAO et al., 2014). The LD decay was slower in both groups belonging to *P. persica*: ornamental peach (56 Kbp) and edible peach (14 Kbp), when compared to the wild species group (5 Kbp).

The number of SNPs required for GWAS is justified by the LD decay over distance. Considering the LD decay of our germplasm panel (38 Kbp), about 5,984 SNPs covering the total peach genome (227.4 Mbp) should be sufficient to carry out association analysis. However, domestication regions containing key genes exhibit faster LD decay than across total genome and more SNPs might need to be identified (CAO et al., 2016).

Furthermore, observed variable level of LD in different chromosomes might be due to artificial selection, which led to the fixation of a higher number of LD blocks especially around genes carrying important agronomic traits (SOTO-CERDA; CLOUTIER, 2012).

1.5 Conclusions

In the present study, we successfully applied the GBS approach for the identification of high quality genome-wide SNPs in Brazilian peach germplasm. Our results detected strong population genetic structure, with the distribution of genotypes within populations based mainly on fruit-related traits: melting and non-melting flesh. LD patterns suggested a medium LD level in peach germplasm with the extent of LD being highly dependent on the populations and genome regions. The valuable SNP resources generated in this study will facilitate peach breeding efforts and support subsequent genome-wide association studies.

CAPÍTULO 2

Selection of peach germplasm with resistance to *Xanthomonas arboricola* pv. *pruni* based on field and detached leaf assessments using the REML/BLUP approach

2.1 Introduction

Bacterial spot, caused by *Xanthomonas arboricola* pv. *pruni* (*Xap*), is a serious disease that threats stone fruits worldwide. This pathogen has been reported from all five continents, expanding its range to almost all countries where *Prunus* species are grown (EPPO, 2017).

Currently considered the most important bacterial disease of peach and nectarine, *Xap* is especially devastating when highly susceptible cultivars are grown in warm, humid areas, accompanied by wind and sandy soils (RITCHIE, 1995; STEFANI, 2010).

Symptoms include varied size spots on leaves, fruit lesions, and twig cankers. Severe infections leads to premature defoliation, progressive tree weakening, as well as reduced fruit quality to the point of unmarketable, and overall decline in production (RITCHIE, 1995; OEPP/EPPO, 2006; PALACIO-BIELSA et al., 2015).

Antibacterial chemicals offer only limited control, therefore, the development of peach cultivars with resistance to *Xap* has been a priority in several breeding programs. The basis of screening for reduced susceptibility has generally been field observations, especially in humid production areas with the existing pathogen pressure, like the

southern Brazil and south and eastern North America (BYRNE et al., 2012; SACHET et al., 2013; FRETT, 2016; THUROW et al., 2016).

Field disease phenotyping, based on the expression of symptoms, experienced great progress and the different levels of resistance observed, indicate a quantitative nature of resistance to *Xap* in peach (YANG et al., 2013; FRETT, 2016). However, the accuracy of phenotyping in the field can be compromised by environmental factors that affect disease expression, such as orchard location, temperature, humidity, distribution of the pathogen inoculum and the presence of other diseases.

To overcome these main drawbacks and enhance the resistance screening for *Xap*, detached leaf assays are often performed. This is rapid, efficient and allows one to differentiate levels of resistance to *Xap* (RANDHAWA; CIVEROLO, 1985; KRETZSCHMAR et al., 1998; MEDEIROS et al., 2011; FRETT, 2016).

Previous studies report considerable variability in the resistance to *Xap* across the germplasm available at the Brazilian Agricultural Research Corporation (EMBRAPA) peach breeding program (SACHET et al., 2013; PALACIO-BIELSA et al., 2015; THUROW et al., 2016). Nevertheless, successful selection of genotypes with resistance to *Xap* depends not only on the variability of the peach germplasm but also on the accuracy of the employed selection methods.

Mixed model methodologies are increasingly used as an optimum selection method in breeding of diverse perennial plant species. It involves the estimation of variance components by means of the restricted maximum likelihood model (REML) developed by Patterson and Thompson (1971), and the prediction of genotypic values by the best linear unbiased predictor (BLUP) (HENDERSON, 1975), resulting in a more accurate selection process than selection indices or phenotypic selection (RESENDE, 2002).

REML can be applied to unbalanced data and produces the same results as ANOVA for balanced data sets. Using REML, the components of variance are estimated without being affected by the fixed effects of the model and the degrees of freedom related to the estimation of the fixed effects are considered, producing unbiased estimates (RESENDE et al., 2014).

Specifically in the cases of mixed models with the assumption of treatments as random effects, BLUP properties allows maximize selective accuracy, minimize prediction

error, predict unbiased genetic values, maximize genetic gain per selection cycle and also the probability of selecting the best genotypes (RESENDE, 2004).

Thus, this study aimed to identify sources of resistance to *Xap* across a diverse peach germplasm panel based on field ratings and detached leaf assessments, as well as estimate the genetic parameters associated with resistance to the disease and predict the individual genotypic values, using the mixed model methodology REML/BLUP.

2.2 Material and methods

2.2.1 Plant material

A panel consisted of 186 diverse peach cultivars and advanced selections (*Prunus persica*), maintained at Embrapa (Pelotas-RS, located at lat. 31°42'S, long. 52°24'W and an altitude of 57m above sea level) was evaluated for response to *Xap*.

This germplasm was selected based on prior knowledge of contrasting phenotypes for bacterial spot, brown rot and tolerance to abiotic stresses such as heat tolerance at the flowering stage and chilling requirement and represents the genetic variability available for breeding.

All peach trees were grafted on seedling rootstocks of *P. persica*, trained as open center system, spaced 3m between trees and 5m between the rows and standard horticultural practices were applied. Detailed information about the member genotypes was previously reported in chapter 1 (Supplementary Table S1).

2.2.2 Field condition

Two experiments, aiming to evaluate *Xap* resistance in peach, were performed from 2014 to 2016. Data of Embrapa environmental conditions (temperature, humidity and rainfall) collected from the local weather station, from August to December of the studied years are presented in table 1.

According to the Köppen classification, the region has a predominant Cfa type of climate, defined as a humid subtropical climate, with hot summers and annual rainfall between 1,000 and 2,000 mm, well distributed throughout the seasons (Table 1). Warm

temperatures combined with very high relative humidity (up to 80%), frequent periods of rainfall and extended heavy dews enable the occurrence of severe *Xap* infection and disease development.

Table 1 - Temperature, rainfall and relative humidity weather condition data for each month (August - December) collected by Embrapa local weather station in 2014, 2015, and 2016.

		Aug.	**Sept.**	**Oct.**	**Nov.**	**Dec.**	**Average**
2014	Average temperatures (°C)	14.5	16.5	19.4	21.4	22.5	18.86
	Average minimum temperatures (°C)	10.1	12.8	15.9	16.7	18.6	14.82
	Average maximum temperatures (°C)	20.1	20.9	24.2	27	27.4	23.92
	Rainfall (mm)	82	179.8	213.8	91.9	148.7	143.24
	Rainy days	16	23	15	12	22	17.6
	Relative humidity (%)	82.3	86.8	85.1	78.2	80.9	82.66
		Aug.	**Sept.**	**Oct.**	**Nov.**	**Dec.**	**Average**
2015	Average temperatures (°C)	18.1	15.1	16.5	18.9	22.1	18.14
	Average minimum temperatures (°C)	14.1	11.7	13.1	15.3	18.4	14.52
	Average maximum temperatures (°C)	22.7	19.4	20.4	23.1	26.6	22.44
	Rainfall (mm)	116.3	277.8	321.4	192.2	261.8	233.9
	Rainy days	16	21	17	16	17	17.4
	Relative humidity (%)	83.6	84.1	87.9	84.4	83.7	84.74
		Aug.	**Sept.**	**Oct.**	**Nov.**	**Dec.**	**Average**
2016	Average temperatures (°C)	14.3	13.9	13.9	19.4	22.5	16.8
	Average minimum temperatures (°C)	10.1	10.3	10.3	14.1	17.7	12.5
	Average maximum temperatures (°C)	19.7	18.6	18.6	25.3	28.4	22.12
	Rainfall (mm)	267.6	141.6	141.6	200.4	137.6	177.76
	Rainy days	19	19	19	11	15	16.6
	Relative humidity (%)	84.2	86.6	86.6	78.2	78.7	82.86

2.2.3 Experiment I - Field disease assessment

All 186 peach genotypes evaluated in this study experienced routine exposure to *Xap* for infection, and no artificial bacterial inoculum was applied in the field. This approach was used because the local environmental conditions are considered as a hotspot for the pathogen and the disease is observed every year. Although some degree of severity variation among years may occur. In addition, multiple years of disease assessments provide adequate reliability of the phenotypic expression response.

Leaf symptoms were evaluated during a one-week period in December, from 2014 to 2016. All germplasm panel was phenotyped following the categorical scale developed by Yang (2012) (Figure 1) ranking from 0 (no symptoms on leaves) to 5 (> 50% diseased leaves or observed defoliation) (Table 2), with intermediate steps of 0.5. The disease score was visually estimated according to the overall intensity and distribution of symptoms on each individual tree. At least three plants per genotype were observed, and the average recorded.

Figure 1 - Bacterial spot symptom severity on peach leaves showed in six different severity categories applied for leaf field evaluation. Source: YANG, 2012.

Table 2 - Phenotypic scale used to assess bacterial spot (*Xap*) infection on peach leaves, described in Yang (2012).

Class	Leaf symptoms
0	No leaves with symptoms
1	1-5% diseased leaves or observed defoliation
2	6-10% diseased leaves or observed defoliation
3	11-25% diseased leaves or observed defoliation
4	25-50% diseased leaves or observed defoliation
5	> 50% diseased leaves or observed defoliation

2.2.4 Experiment II - Detached leaf bioassay

To overcome the potential environmental differences in the field, a detached leaf bioassay was developed, using a modified protocol based on the one described by Randhawa and Civerolo (1985).

To perform the assay, 109 genotypes from the whole peach panel were evaluated between October to December in the years of 2015 and 2016. The bioassay experiment was arranged in an augmented incomplete block design with two replications (2015 and 2016). Blocks consisted of 9 to 14 genotype entries randomly chosen, plus two check cultivars (Eldorado and Gaúcho). Under field conditions, the cultivar Eldorado is considered moderately susceptible (RASEIRA et al., 2014) and the cultivar Gaúcho is resistant to *Xap* infection (Raseira, personal communication). The experimental unit was represented by one leaf with eight inoculated sites and 10 leaf replicates were assayed by genotype.

The *Xap* inoculum was isolated from peach symptomatic leaves in 2014, identified by microscopic observation and finally stored in 25% glycerol at -80°C. Diseased leaves were collected in two distinct locations of the field experimental area, resulting in two different *Xap* isolates used to perform the assay.

Third to fifth young fully expanded leaves (from 20-25 cm long shoots) from actively growing trees in the field were collected in the morning, placed in plastic bags on ice and brought to the lab. The leaves were surface-sterilized by soaking in a 0.5% sodium hypochlorite solution for 5 min, rinsed with sterile distilled water and water excess removed with sterile filter paper. Subsequently leaves were immediately used for inoculations.

Inoculum suspensions were prepared from cultures incubated for 36 h at 28 °C on Kado 523 culture media. A loopful of bacterial growth was suspended in sterile phosphate buffered saline (1X PBS) containing 0.05 % Tween 20. Bacterial concentration was then adjusted using UV spectrophotometer to 0.2 optical density at 600 nm, equivalent to 10^8 CFU ml^{-1}. Thereafter, an inoculum mixture of the two different *Xap* isolates, in equal proportion, was used to perform the assay.

Leaves were placed abaxial side up on sterile filter paper and infiltrated with a sterile 3 ml syringe (without needle) loaded with the inoculum. Gentle pressure was applied for leaf infiltration, until a clear water-soaked spot appeared around the inoculation point. If inoculum spilled around, excess was lightly wiped from the surface with sterile filter paper. Inoculated leaves were then placed in petri dishes containing agar (1.5%) + fungicide 2.0 ppm in water solution, with the abaxial side up, sealed to conserve moisture

and incubated at 25°C ± 2°C, under a 12h photoperiod. After 10 days, each lesion length was measured and data recorded. The lesion length (mm) was determined by measuring the longest length (A) and width (B) of each lesion and calculated using the formula: (A + B)/2. The lesion area (mm^2) was estimated based on the lesion length measures, using the formula: $\pi \cdot r^2$.

2.2.5 Data analysis

Statistical analyses were performed using a mixed linear model methodology. Variance components were estimated by restricted maximum likelihood (REML) method, and phenotypic values of the peach genotypes were predicted by best linear unbiased prediction (BLUP) (RESENDE, 2007b).

The following linear mixed model was applied for genotype response to *Xap* at field level:

$$y = Xm + Wp + e \ ,$$

where *y* is the data vector, *m* is the vector of the assessment years effects (assumed as fixed) added to the overall average, *p* is the vector of permanent effects of the peach genotypes (genotypic effects + permanent environmental effects) (assumed as random), and *e* is the vector of errors or residuals (random effects). *X* and *W* are matrices of incidence of the *m* and *p* effects, respectively.

The mean components (individual BLUPs) based on the permanent phenotypic effect of 186 evaluated peach genotypes were obtained aiming to rank and select genotypes conferring resistance to *Xap*.

For the detached leaf bioassay, conducted based on an augmented incomplete block design with two replications, the analysis followed the statistical model:

$$y = Xf + Zg + Wb + Ti + e \ ,$$

where *y* is the data vector, *f* is the vector of fixed effects (check cultivars average and genotype overall average in each year), *g* is the vector of genotypic effects (assumed as random), *b* is the block effect (assumed as random), *i* is the vector of genotype x

environment (G x E) interaction effects (random), and *e* is the vector of errors or residuals (random). *X*, *Z*, *W* and *T* represent the matrix of incidence for the respective effects.

REML estimates of the variance components were obtained using the expectation-maximization (EM) algorithm according to Resende et al. (2014).

The following parameters were estimated: genotypic variation among genotypes ($\hat{\sigma}_g^2$); environmental variance between blocks ($\hat{\sigma}_b^2$); variance of the genotype x environment interaction ($\hat{\sigma}_i^2$); residual variance ($\hat{\sigma}_e^2$); individual broad-sense heritability $\hat{h}_g^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_b^2 + \hat{\sigma}_i^2 + \hat{\sigma}_e^2}$; coefficient of determination of the block effect $c_b^2 = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_g^2 + \hat{\sigma}_b^2 + \hat{\sigma}_i^2 + \hat{\sigma}_e^2}$; coefficient of determination of the G x E interaction effects $c_i^2 = \frac{\hat{\sigma}_i^2}{\hat{\sigma}_g^2 + \hat{\sigma}_b^2 + \hat{\sigma}_i^2 + \hat{\sigma}_e^2}$; and genotypic correlation between the behavior of genotypes in different environments $\hat{r}_{gA} = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_i^2}$.

The empirical BLUP predictors of genotypic values free of the interaction were given by $\hat{\mu} + \hat{g}_i$, where $\hat{\mu}$ is the average of all environments, and $\hat{g}_i$ is the genotypic effect free of the G x E interaction.

All analyses were performed using the Selegen-REML/BLUP software (RESENDE, 2007b; RESENDE, 2016).

2.2.6 Germplasm classification for resistance to *Xap*

The genotypes were grouped into different classes, according to their responses to *Xap*, based on the normal distribution of the predicted genotypic values (individual BLUPs).

The standard deviation (SD) of predicted genotypic values was calculated for the response to *Xap* in the field and in the detached leaf bioassay (length and area of the lesion). Assuming a normal distribution of genotypes in the curve, genotypes with positive deviations were classified as susceptible, while genotypes with negative deviations were classified as resistant.

This SD value was used to discriminate the peach genotypes in six classes: HS (highly susceptible): > +2 SD; S (susceptible): > +1 SD and < +2 SD ; MS (moderately susceptible): > $\bar{X}$ and < +1 SD; MR (moderately resistant): > -1 SD and < $\bar{X}$; R (resistant):

> -2 SD and < -1 SD; HR (highly resistant): < -2 SD; where $\bar{X}$ is the overall mean of the variable in study.

2.3 Results and discussion

2.3.1 Estimates of genetic parameters

Estimates of the phenotypic and genetic parameters based on leaf response to *Xap* are presented in Table 3. The mean value of *Xap* severity among the 186 peach genotypes evaluated under field conditions was 2.50. Estimated value of permanent phenotypic variance among plants (σ^2_{pp}), composed by the genotypic variance and the permanent environment variance over crop seasons, was higher than the estimate of temporary environment variance (σ^2_{te}), indicating the possibility for successful selection of *Xap* resistant sources. Consequently, the broad sense heritability (h^2) was considered of high magnitude (0.56 ± 0.09). According to Stansfield (1974), heritability values higher than 50% are considered high.

The correlation between the predicted genotypic values and the true genotypic values can be evaluated based on the selective accuracy (A_{cm}). The A_{cm} value obtained with three repeated measures reached a high magnitude (0.89) (Table 3), indicating high precision according to the limiting values proposed by Resende and Duarte (2007).

Table 3 - Estimates of variance components (individual REML) and genetic parameters related to the severity of bacterial leaf spot caused by *Xap* in peach.

Parameter	Leaf response to *Xap*
σ^2_{pp}	0.57
σ^2_{te}	0.44
σ^2_{p}	1.01
$r=h^2$	0.564 ± 0.09
rm	0.80
A_{cm}	0.89
Overall average	2.50

σ^2_{pp} permanent phenotypic variance among plants, σ^2_{te} temporary environment variance, σ^2_{p} individual phenotypic variance, $r = h^2$ individual repeatability and its confidence interval, *rm* repeatability of the average of crop seasons or repeated measures, A_{cm} selection accuracy based on the average of *m* crop seasons or repeated measures.

For the detached leaf bioassay, the estimates of phenotypic variance (σ^2_p) decomposed into genotypic variance (σ^2_g), environmental variance between blocks (σ^2_b), variance of the genotype x environment interaction (σ^2_i), and residual variance (σ^2_e) are presented in Table 4. For genotype reaction most of the variance for both the length (mm) and area (mm^2) of the lesion is due to random effects (residue), followed by the environmental variance between blocks for the first trait and variance of the G x E interaction for the second (Table 4). Genotypic variance accounted for 16.7% and 14.3% of the phenotypic variance for lesion length and lesion area, respectively.

Table 4 - Estimates of variance components (individual REML) and genetic parameters related to the severity of bacterial leaf spot in peach detached leaves.

Parameter	Lesion length	Lesion area
σ^2_g	0.047	3.427
σ^2_b	0.068	5.348
σ^2_i	0.064	7.460
σ^2_e	0.103	7.679
σ^2_p	0.282	23.914
h^2_g	0.167	0.143
c^2_b	0.240	0.224
c^2_i	0.228	0.312
r_{gA}	0.422	0.315
Overall average	5.33	23.08

σ^2_g genotypic variance, σ^2_b variance between blocks, σ^2_i variance of the genotype x environment interaction, σ^2_e residual variance, σ^2_p individual phenotypic variance, h^2_g individual broad-sense heritability of total genotypic effects, c^2_b coefficient of determination of the block effect, c^2_i coefficient of determination of the environmental x genotype interaction effects, r_{gA} genotypic correlation between the behavior of genotypes in different environments.

Due to the significant environmental influence on disease expression, individual heritability in the broad sense (h^2_g) was of low magnitude (0.167 and 0.143 for length and area of the lesion, respectively). However, it is important to take into consideration that h^2_g is free of G x E interaction and, therefore, may not be compared with individual phenotypic heritability.

The coefficients of determination of the block effect (c^2_b = 0.24 and 0.22) for both variables were slightly above the recommended by Sturion and Resende (2005), who considered values around 0.10 of low magnitude in perennial plants.

A significant part of the variation observed was attributed to the environmental x genotype interaction as showed by the coefficients of determination of the G x E interaction effects within the block (c^2_i) (Table 4), which justifies the capture of these effects to select peach genotypes with resistance to *Xap*. This could be achieved by selecting genotypes based on average genotypic values in the environments or by selecting based on genotypic values for specific environments.

The genotypic correlation between the behavior of genotypes in different environments (r_{gA}), in this study, two different years of assessments, was of moderate magnitude, indicating possible complex interactions between genotypes. Only 42% and 31.5% matched when selecting specific genotypes for each year-assessment, considering length and area of the lesion, respectively. According to Resende (2007a), r_{gA} values ≥ 0.70, or 70% of correlation, indicates a simple type of interaction and < 0.70 indicates a complex interaction.

2.3.2 Genotypic values of resistance to *Xap*

The variation of leaf response to *Xap* among the 186 peach genotypes evaluated under field conditions is shown in Figure 2. The genotypic values of field ratings for *Xap* severity ranged from 0.91 to 4.09 with an overall average of 2.50 and a standard deviation of 0.67. In terms of resistant sources to *Xap*, 87 genotypes showed lower predicted genotypic values than the general mean and therefore are considered more resistant to *Xap* infection. On the other hand, 99 genotypes exhibited higher *Xap* severity predicted values than the general mean, which means more susceptible genotypes. Predicted genotypic values and genotype rank are shown in Supplementary Table S2.

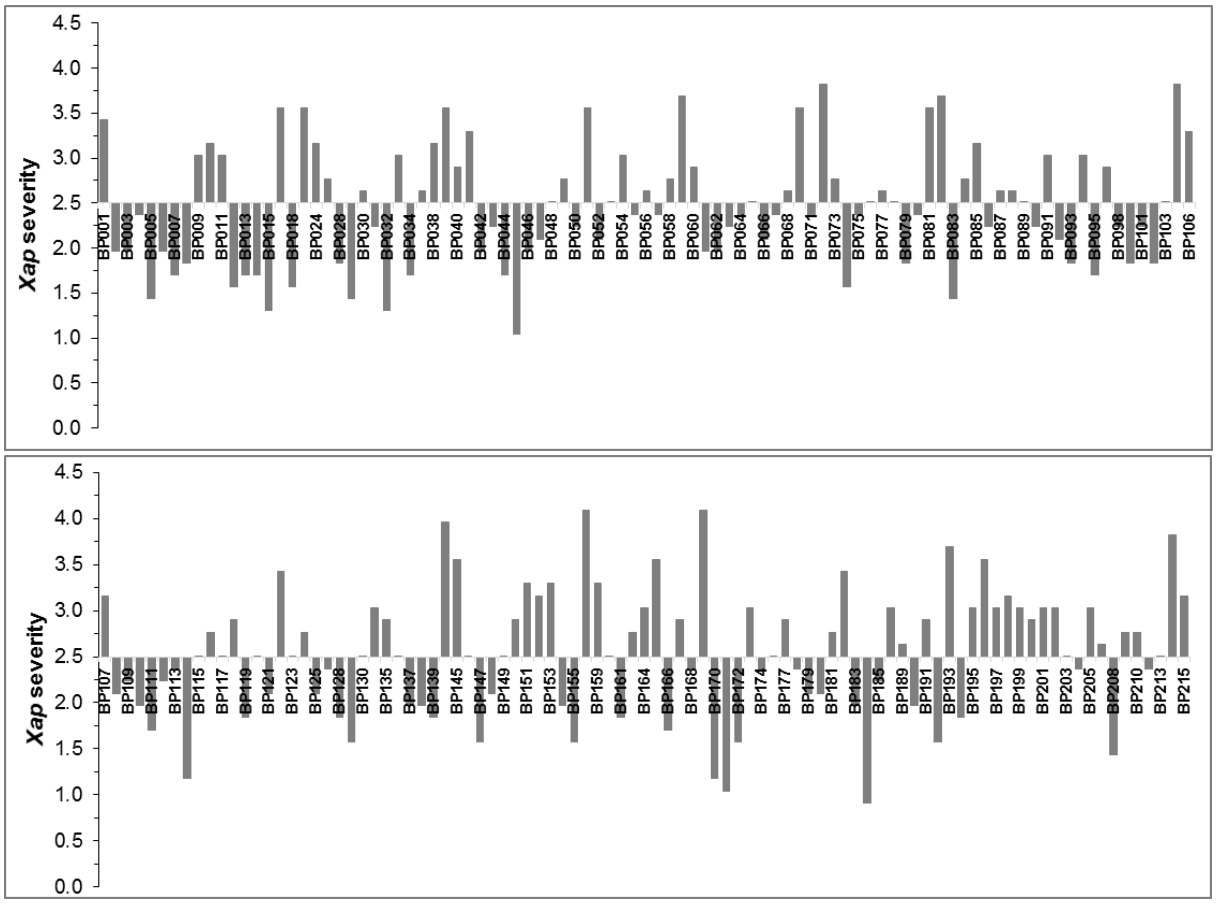


Figure 2 **-** Predicted genotypic values for Xap severity among 186 peach genotypes evaluated under field conditions. Genotype predicted values are plotted against the general mean (2.50).

The detached leaf assay assessed the variables lesion length and lesion area as a result for *Xap* infection. As expected, the correlation between both variables was high (0.99), representing a perfect agreement between length and estimated lesion area.

Genotypic values for lesion length exhibited an average of 5.33 mm among all 109 peach genotypes assessed. The values ranged from 5.05 to 5.63 mm, and showed a standard deviation of 0.13 mm. The germplasm distribution for lesion length in response to *Xap* infection is shown in Figure 3a. In terms of resistant sources to *Xap*, 54 genotypes exhibited negative deviations with respect to the overall mean and therefore, considered with some degree of resistance and 55 genotypes showed positive deviations, which indicates genotypes with higher susceptibility for the disease development.

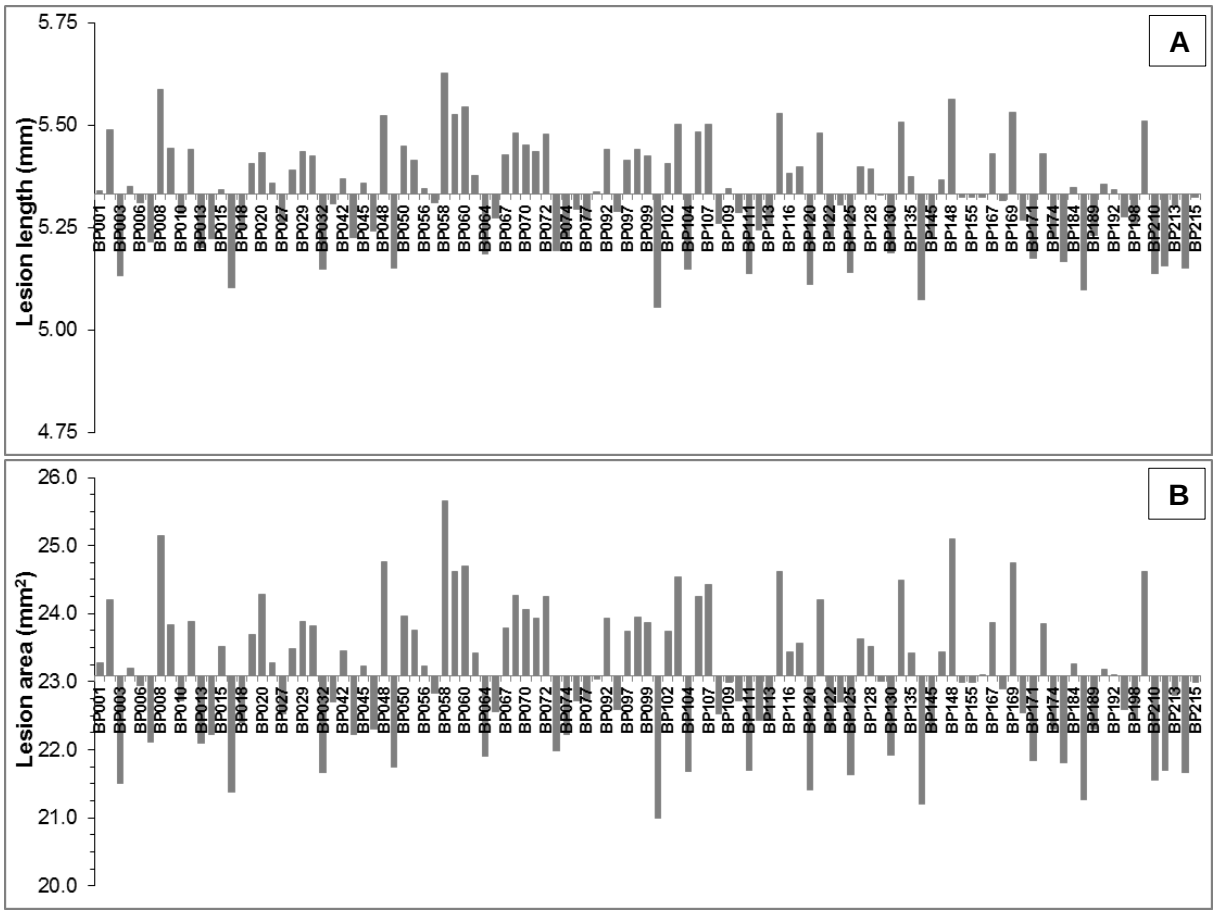


Figure 3 - Predicted genotypic values for lesion length (mm) (A) and lesion area (mm^2) (B) caused by *Xap* among 109 peach genotypes evaluated using a detached leaf bioassay. Genotype predicted values are plotted against the general mean: 5.33 mm (A) and 23.08 mm^2 (B).

For the variable lesion area, predicted genotypic values exhibited an average of 23.08 mm^2 among all 109 peach genotypes. The values ranged from 21.00 to 25.67 mm^2, and showed a standard deviation of 1.00 mm^2. The genotype distribution for lesion area affected by *Xap* infection is shown in Figure 3b. A total of 55 genotypes exhibited negative deviations with respect to the overall mean and therefore, can be considered with some degree of resistance and 54 genotypes showed positive deviations, which refers to genotypes with higher susceptibility for the disease development.

Genotype BLUP scores for both variables assessed by detached leaf bioassay and genotype rank are given in Supplementary Table S2. According to Resende (2007b) these predicted genotypic values can be extrapolated to locations outside this experimental network, once the performance of the genotypes is free from G x E interaction. It is expected that the same genotypic means for resistance to *Xap* would be

observed under different environments, due to the conservative nature of the $\hat{\mu} + \hat{g}_i$ estimates (MAIA et al., 2009). On the other hand, selection of *Xap* resistance sources under field conditions, based on permanent phenotypic values, will be effective only for the region where they were evaluated, once these values capitalize the genetic variation and the permanent environment effects.

2.3.3 Comparison between field rating and detached leaf bioassay

The general rating of susceptible (genotypes with positive deviations with respect to the mean) and resistant genotypes (genotypes with negative deviations with respect to the mean) on response to *Xap* under field conditions and on the detached leaf assay coincides in 57.4% of the genotypes.

However, about 23.2% were considered resistant under field conditions and susceptible on the detached leaf assay. This behavior is expected, once there is an increased amount of disease pressure, most due to the partial loss of physiological resistance in detached leaf tissue compared with the whole plant. On the detached leaf assay the *Xap* inoculum is directly infiltrated into leaf tissue, thus the pathogen already overcomes the pre-formed structural defenses (cuticle, cell wall, stomata, trichomes) that help in limiting pathogen attachment, invasion and infection. Disease escape also plays an important role in reducing infection rates, since genotypes will not be screened under sufficient pathogen pressure (PANGULURI; KUMAR, 2013).

The presence of a susceptibility factor not detected by the detached leaf assay may be the reason why 19.4% of the genotypes were rated as susceptible under field conditions and resistant on the detached leaf assay. In addition, majority of those discrepant genotypes are part of the moderate susceptible (MS) group in the field rating and classified as moderate resistant (MR) or resistant (R) in the detached leaf assessment. This variation may also reflect in part, the difficulty in determining small differences on the measurements, as well as possible factors, which could have limited pathogen growth.

Pathogen variability also need to be considered for developing reliable detection methods, since resistance may differ dramatically due to unique pathogenic races of the

bacteria as reported before by Martins (1996) who identified wide differences in virulence between *Xap* strains obtained from peach, nectarine and plum orchards of different regions in Brazil. In the present study, the *Xap* strains used were isolated from diseased leaves of two susceptible cultivars in the experimental field area.

2.3.4 Germplasm classification for resistance to *Xap*

Peach genotypes were grouped into six different classes of response to *Xap* based on the standard deviation of the predicted genotypic values (Table 5). Among the 186 genotypes evaluated for severity of *Xap* symptoms, under field conditions, the cultivars Norman, Cristal Taquari and La Feliciana were considered highly resistant (HR). These results agree with those reported in the literature, considering these genotypes as sources of resistance to *Xap* (OKIE, 1998; RASEIRA; NAKASU, 2006). Regarding the detached leaf assay, the cultivar La Feliciana was considered MR and the cultivars Norman and Cristal Taquari were classified as MS.

Approximately 12.9 % of the genotypes were classified as resistant (R), 32.3 % as moderately resistant (MR), 39.2 % as moderately susceptible (MS), 12.4 % as susceptible (S) and 1.6 % as highly susceptible (HS) (Table 5). The genotypes Cascata 1423, Conserva 1127 and ‘Ingo’ were those most severely affected by the disease with high rates of premature defoliation and, therefore classified as HS, under field conditions. Conserva 1127 was also previously identified as one of the most susceptible genotypes for reaction to *Xap* by Sachet et al. (2013). Ingo was the only cultivar assessed by detached leaf, and classified as one of the most susceptible genotypes.

One hundred nine genotypes were also tested by detached leaf assay to assess genotype response to *Xap* infection. The distributions of them in different classes based on lesion length and lesion area showed high similarity (Table 5). According to the classification (based on the standard deviation of the predicted genotypic values), only the cultivar Precocinho was considered highly resistant for both assessed variables. This result corroborate with the field rating and with previous report considering this cultivar as a resistant source (PALACIO-BIELSA et al., 2015). On the other hand, two genotypes (‘Flordagrande’ and ‘Arlequim’) were considered the most highly susceptible (HS) for the

variable lesion length and three HS genotypes (‘Flordagrande’, ‘Arlequim’ and Cascata 1513) for lesion area assessments.

Table 5 - Distribution of peach genotypes with respect to their response to *Xap* evaluated at the field level (*Xap* severity) and by a detached leaf bioassay (lesion length and lesion area).

Class	***Xap* severity**		**Lesion length (mm)**		**Lesion area (mm^2)**	
	N° of genotypes	Range of predicted values	N° of genotypes	Range of predicted values	N° of genotypes	Range of predicted values
HS	3	3.95 - 4.09	2	5.58 - 5.63	3	25.10 - 25.66
S	23	3.29 - 3.83	15	5.47 - 5.56	15	24.20 - 24.76
MS	73	2.49 - 3.17	38	5.33 - 5.45	36	23.08 - 24.06
MR	60	1.83 - 2.37	34	5.21 - 5.33	36	22.09 - 23.03
R	24	1.17 - 1.71	19	5.07 - 5.20	18	21.20 - 21.99
HR	3	0.90 - 1.05	1	5.05 - 5.05	1	21.00 - 21.00

HS highly susceptible, *S* susceptible, *MS* moderately susceptible, *MR* moderately resistant, *R* resistant, *HR* highly resistant.

The overall screening results indicate the genotypes, ‘Norman’, ‘Cristal Taquari’, ‘La Feliciana’ and ‘Precocinho’ as good sources of resistance. Therefore, they may be effective alternatives to improve *Xap* resistance in peach. The genotypes Cascata 1423, Conserva 1127 and ‘Ingo’ were the most susceptible evaluated, under field conditions, and ‘Flordagrande’, ‘Arlequim’ and Cascata 1513 the worst assessed by detached leaf.

It should be noted that, in this study, the cultivar Flordagrande was considered MS in field conditions, in agreement with Byrne et al. (2000), who also described this cultivar as MS. This result confirms that in detached leaf assessments, the pathogen may be more aggressive, once already overcame the pre-formed structural defenses that help limit infection and disease development.

Although some discrepancy occurred in the classification of peach genotypes for resistance to *Xap*, based on symptoms expressed under field conditions and in the detached leaf assessments, it was possible to identify several genotypes with resistance. These genotypes could be preferred for peach production areas more subjected to disease occurrence, or used as parents by the breeding program to improve resistance (Supplementary Table S2).

2.4 Conclusions

Field evaluation and detached leaf assessments are reliable and complementary methods to identify *Xap* resistant sources in peach.

The germplasm evaluated showed great variability for response to *Xap*, identifying contrasting genotypes for the trait of interest.

CAPÍTULO 3

Genome-wide association mapping for bacterial spot resistance in *Prunus persica*

3.1 Introduction

Bacterial spot, caused by *Xanthomonas arboricola* pv. *pruni* (*Xap*) is one of the most serious diseases threatening peach production worldwide (EPPO, 2017).

Symptoms occur on leaves, twigs and fruits. Severe infection results in premature leaf defoliation, progressive tree weakening, as well as reduced fruit quality and yield (OEPP/EPPO, 2006; STEFANI, 2010; PALACIO-BIELSA et al., 2015).

Infection and development of *Xap* is highly dependent on environmental conditions. The disease is especially devastating when highly susceptible cultivars are grown in warm, humid areas, accompanied by wind and sandy soils (RITCHIE, 1995; YANG et al., 2013).

Due to limited anti-bacterial chemical control and an environmentally conscious public, developing peach cultivars with resistance to *Xap* has been a priority in many breeding programs (GASIC et al., 2015; FRETT, 2016).

Differential levels of resistance were previously reported (SACHET et al., 2013; THUROW et al., 2016), indicating that there is no complete resistance, because numerous genes or quantitative trait loci (QTLs) control the disease.

Previous studies identified several QTLs related with the genetic control of *Xap* resistance in *Prunus* (SOCQUET-JUGLARD et al., 2013; YANG et al., 2013; FRETT, 2016). However, all previous QTLs co-segregating with *Xap* resistance were identified

using single bi-parental segregating populations or more recently, using pedigree-based approach (PBA).

An apricot population consisting of 101 F1 individuals allowed to identify a major QTL for *Xap* leaf resistance on linkage group (LG) 5 (SOCQUET-JUGLARD et al., 2013). Fourteen QTLs, including five major QTLs on LG 1, 4(2), 5, and 6, were identified with additive effects on *Xap* resistance in a peach F2 population derived from 'O'Henry' (susceptible) and 'Clayton' (resistant) (YANG et al., 2013).

Recently, Frett (2016) used a PBA QTL analysis approach and identified 19 reliable QTLs, out of all, six consensus QTLs associated with both resistance in fruit and leaf on LG 1(2), 2(2), 5 and 6 and four associated with only fruit resistance on LG 1, 2, 3 and 8. PBA approach uses a larger genetic background to enhance the ability to detect more QTL loci for *Xap* resistance compared to single bi-parental mapping.

Fortunately, the rapid development of new sequencing technologies has created the opportunity to enhance our understanding of the genetic basis of *Xap* resistance. Next-generation sequencing technologies have decreased the cost of SNP genotyping and expanded the availability of numerous markers (DAVEY et al., 2011). High throughput genotyping methods, such as genotyping by sequencing (GBS) (ELSHIRE et al., 2011), can produce high-density SNP coverage spanning the entire genome and supporting genome-wide association studies (GWAS).

GWAS presents a powerful tool to reconnect a phenotype defined across many individuals, back to its underlying genetics, becoming an important and effective tool for plant breeding (BRACHI et al., 2011; KORTE; FARLOW, 2013).

Association mapping is most commonly performed in collections of unrelated diverse germplasm, in order to maximize the diversity of alleles and haplotypes. Thus, phenotypic data accumulated every year in conventional breeding programs could also be useful for association tests. GWAS takes advantage of the linkage disequilibrium (LD) present between SNPs, as well as historical recombinations within the gene pool available, to identify significant associations between DNA polymorphisms and trait variation (KHAN; KORBAN, 2012; VARSHNEY et al., 2014).

The objective of this study was to conduct a GWAS approach for *Xap* resistance using a diverse peach germplasm panel and GBS-derived SNPs. GWAS was performed

in order to identify genomic regions and candidate genes associated with *Xap* resistance aiming subsequent application of marker-assisted selection (MAS) in the develop of new varieties with improved resistance for this pathogen.

3.2 Material and methods

3.2.1 Phenotypic and genotypic data

Phenotypic data for *Xap* resistance was obtained through a three consecutive-year field assessment of a panel consisted by 186 diverse peach (*Prunus persica*) cultivars and advanced selections. Disease severity was visually assessed using the categorical scale developed by Yang (2012), ranking from 0 (no symptoms on leaves) to 5 (> 50% diseased leaves or observed defoliation), with intermediate steps of 0.5. A mixed linear model approach was used to obtain the best linear unbiased prediction (BLUP) values for all genotypes in the panel. These BLUP values were taken as the trait phenotypes in association analysis (available in chapter 2).

SNP marker data was obtained through genotyping by sequencing, as reported earlier in chapter 1. The parameters used to filter the raw SNPs were a 25% threshold for maximum missing data and a MAF of 0.05 as cut-off, which resulted in a total of 18,373 SNP markers (hereafter referred to as 18K).

3.2.2 Population structure and kinship estimation

Population structure was determined using the model-based Bayesian statistics implemented in the software fastSTRUCTURE (RAJ et al., 2014). To avoid bias due to tightly linked markers, the 18K SNPs were pruned based on linkage disequilibrium (LD), to give a final subset of 5,378 genome-wide and unlinked SNPs, as reported earlier in chapter 1. Inferred subpopulation memberships (Q matrix) of each genotype for the identified population structure (K = 3) were used as covariate in association analysis.

The genetic relatedness of the genotypes in the panel was computed using the default "Centered_IBS" (identical by state) kinship method (ENDELMAN; JANNINK,

2012), implemented in the software TASSEL v5.2 (BRADBURY et al., 2007). This method provides a better estimate of additive genetic variance. It codes genotypes as 2, 1, or 0, corresponding to the count of one of the alleles at that locus, followed by imputation of missing genotype values with the average genotypic score at that locus. The relationship matrix (K matrix) was estimated using the 18K SNP markers.

3.2.3 Association mapping

A genome-wide marker-trait association analysis was conducted to test for association between SNPs and *Xap* resistance using the software TASSEL v5.2 (BRADBURY et al., 2007).

Two types of methods were employed, a general linear model (GLM) and a mixed linear model (MLM). In GLM method, the Q matrix comprising the subpopulation membership estimates was used as covariate in the model to avoid spurious associations. Multiple testing corrections were implemented, in order to control the experiment-wise error rate, by running 1000 permutations. The critical *p*-value for assessing the significance of associations was calculated based on the Bonferroni multiple test correction (significance level divided by total number of SNPs tested) at the 5% significance threshold for 18K markers, corresponding to a *p*-value ≤ 2.7e-06.

To improve statistical power, in MLM method, along with the phenotypic data, kinship matrix (K matrix) was included as a random effect within the model, in addition to the genotypic and Q matrix, which were considered as fixed effects. Q + K matrices correct the bias for both population structure and relatedness. Highly stringent and conservative *p*-values may exclude real associations, therefore the significant threshold was set at the probability value $p < 0.001$ for MLM method.

The comparison of GLM and MLM analysis was made using quantile-quantile (Q-Q) plots by showing the deviation of the observed *p*-values against expected values (null hypothesis). To provide a complimentary summary of putative QTLs, Manhattan plots were generated using R software (R Core Team, 2016).

3.2.4 Candidate gene mapping

Flanking sequences of the candidate SNPs associated with *Xap* resistance were extracted and aligned to the *Prunus persica* whole genome assembly v2.0 annotation v2.1 (VERDE et al., 2017) using basic local alignment search tool (BLASTn).

SNPs were named according to scaffold and base pair position (physical location) within the peach genome assembly v2.0 (VERDE et al., 2017). SNP names contain “S” meaning scaffold followed by the scaffold number, an underscore, and characters denoting the base position. A SNP at physical location 46,645,443 on scaffold 1 would therefore have been named ‘S1_46645443’.

3.3 Results and discussion

3.3.1 *Xap* resistance and SNP markers

Out of the 186 peach genotypes evaluated in the field, large phenotypic variation was observed, with best linear unbiased prediction (BLUP) values for disease severity ranging from 0.91 to 4.09 and an overall average of 2.50. The phenotypic distribution of disease severity based on BLUP estimates can be visualized in figure 1.

The whole peach panel was genotyped using GBS. A set of 18,373 SNPs was selected across the whole peach panel, with MAF > 0.05 and a call rate of 75%, which corresponds to an average density of one marker per 12.4 kbp (considering peach whole-genome size ~227.4 Mbp).

Two genotypes (‘Linda’ and ‘Ingo’) evaluated for *Xap* resistance were removed before the association analysis was performed due to low initial read numbers, resulting in a final association panel composed by 184 peach genotypes.

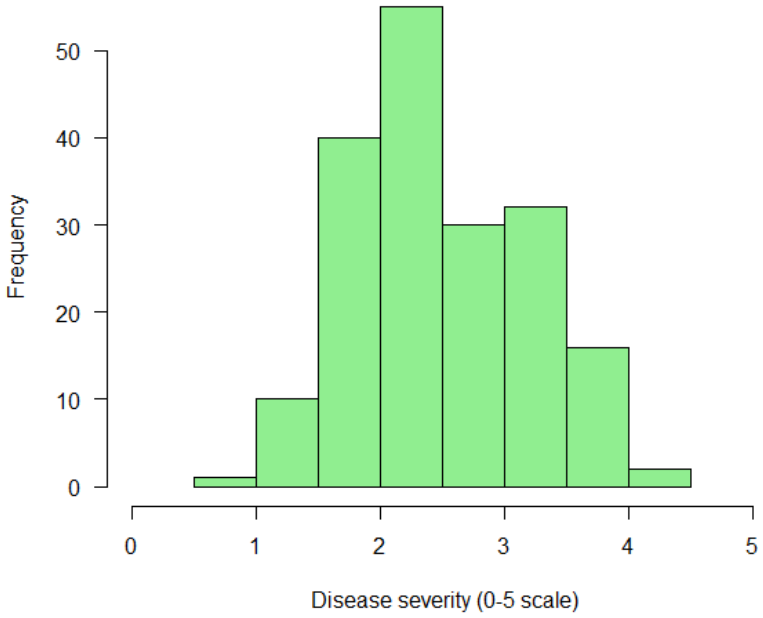


Figure 1 - Phenotypic distribution of disease severity based on best linear unbiased prediction (BLUP) estimates for 186 peach genotypes evaluated under field conditions.

3.3.2 Association mapping and defense mechanisms in response to *Xap*

Genome-wide association was tested between 18,373 SNP markers and disease severity BLUP values for 184 peach genotypes evaluated in the field. The Q-Q plots indicated that both models controlled false positive associations. However, including both, the genetic relatedness and the genetic structure information (K + Q) in the MLM model, resulted in a better fit between observed and expected p-values than only the correction for population structure (Q) in the GLM model (Figure 2).

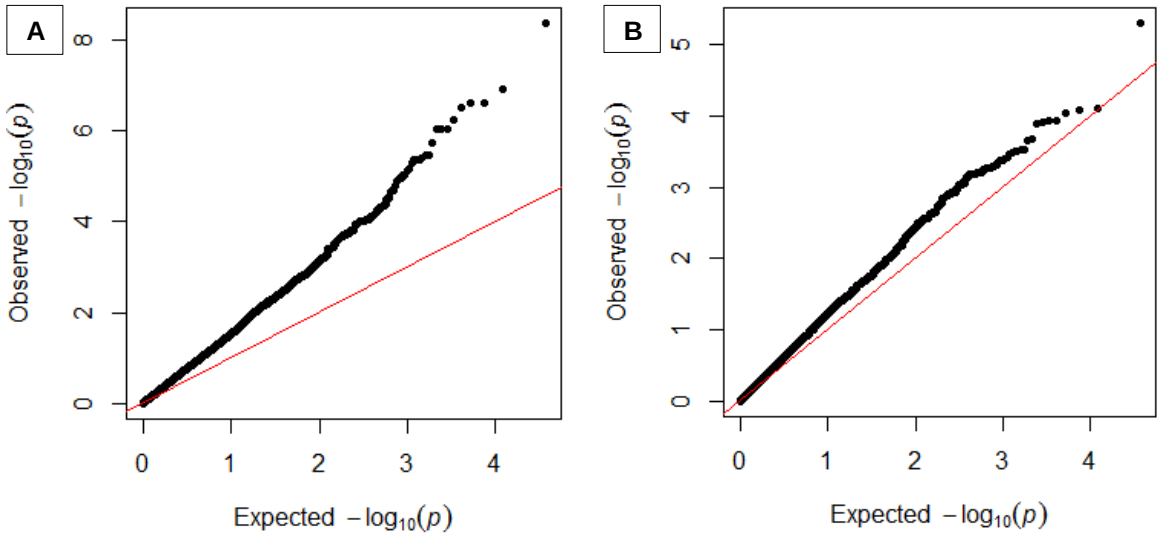


Figure 2 - Quantile-quantile (Q-Q) plots of p-values for *Xap* resistance association analysis. A) Obtained by GLM analysis including only population structure correction. B) Obtained by MLM analysis including both kinship matrix and population structure correction.

A total of 10 SNPs, spanning a 5.15 Mbp region on chromosome 1, remained statistically significant after Bonferroni multiple adjustments (*p*-value ≤ 2.7e-06) in GLM analysis (Figure 3). At a less stringent threshold of 0.10 significance level (*p*-value ≤ 5.4e-06), seven additional SNP markers were identified, two within the same genomic region of chromosome 1, and 5 SNPs located on chromosomes 2, 3 and 4. Interestingly, within the 5.15 Mbp region (41.97 Mbp to 47.12 Mbp) on chrom 1, besides the 10 SNP markers that met stringent criterion, additional 17 SNPs exceeded *p*-values < 0.0001, providing strong evidence of a key region involved with *Xap* resistance.

Using Pedigree-Based Analysis (PBA) approach and the same phenotypic scale to assess *Xap* infection in the field as used in this study, Frett (2016) identified a consensus QTL for *Xap* resistance in fruit and leaf. A region of 12.39 Mbp on chrom 1, located between 34.46 Mbp and 46.85 Mbp confidence interval, which is co-localized with the chrom 1 region identified by GLM analysis in this study.

By MLM method, using a cut-off *p*-value < 0.0001, four SNP markers on chrom 1, located between 44.06 Mbp and 46.65 Mbp, met this stringent criterion (Figure 4). These four SNPs were also identified in the GLM analysis showing *p*-values < 1e-6, shrinking the 5.15 Mbp genomic region identified by GLM approach to 2.59 Mbp in MLM analysis.

These four SNP markers with strongest association (*p*-value < 0.0001) defined a 2.59 Mbp interval, which contains 18 TIR-NB-LRR genes (TNL genes) (Figure 5). TNL genes constitute a major family of effector-triggered immunity (ETI) R genes in dicots (VAN GHELDER; ESMENJAUD, 2016).

It is well known that the first line of defense to pathogen attacks involves the early detection of pathogen-associated molecular patterns (PAMPs) through PAMP-triggered immunity (PTI). Pathogens then secrete effectors that manipulate plant immunity and suppress PTI. These effectors are recognized, directly or indirectly, by specific disease resistance (R) genes, through a second line of defense known as effector-triggered immunity (ETI). Most R genes encode nucleotide binding-leucine rich repeat (NB-LRR) proteins, and if one effector (Avr protein), is recognized by a corresponding NB-LRR protein, ETI occurs (JONES; DANGL, 2006).

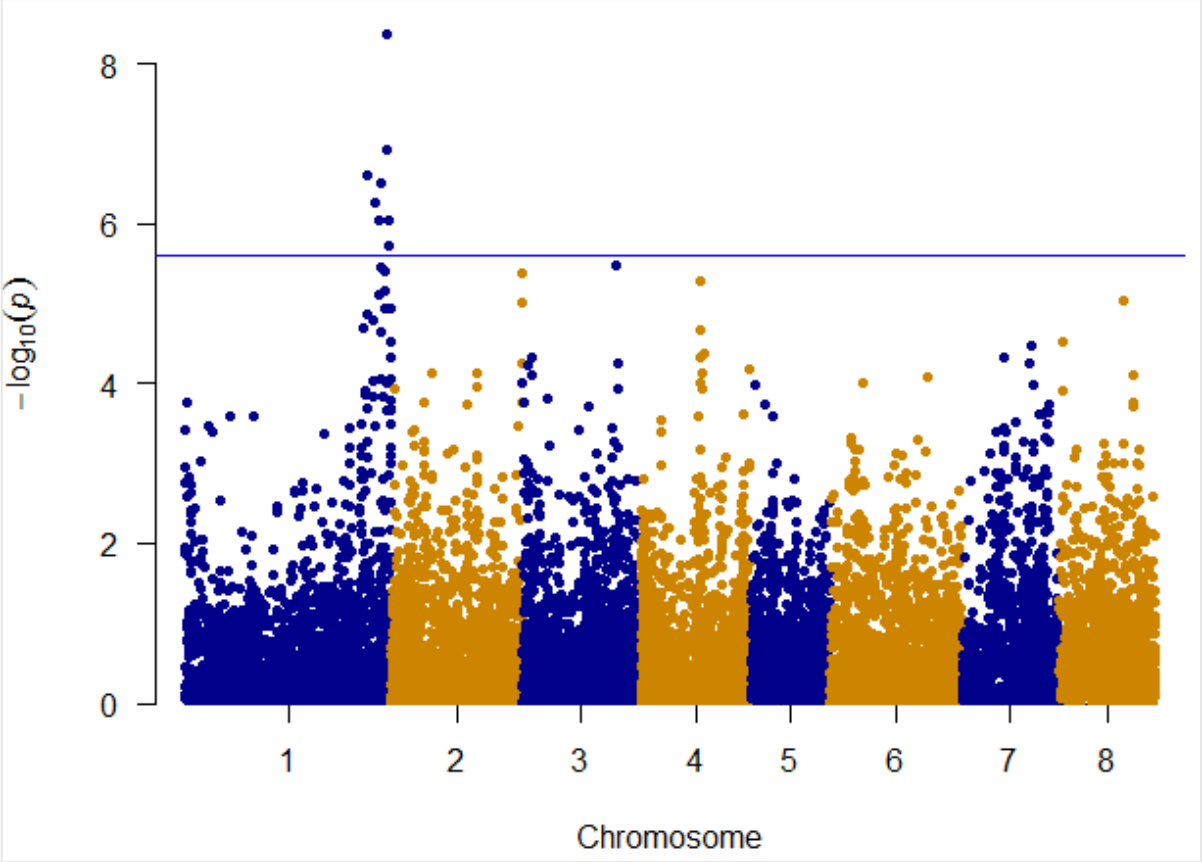


Figure 3 - Genome-wide association study (GWAS) for *Xap* resistance in peach. The 184 peach genotypes were scanned with 18K SNPs using a GLM approach taking into account population structure (Q). The vertical axis plots the $-\log_{10}(p)$ values of the association between the SNP markers and disease severity. The horizontal line denotes Bonferroni-corrected 0.05 significance level (*p*-value ≤ 2.7e-06).

For the most significant SNP marker (S1_46645443) associated with *Xap* resistance in both, GLM and MLM analysis, we found three genotypes CC, CT and TT. The allele C probably confers resistance to *Xap*. For the peach panel evaluated, 45 individuals carries the favorable homozygous CC alleles with an average disease severity rating of 2.03, 66 are heterozygous (CT) with an average of 2.48 and 69 are homozygous transferring TT alleles, probably the susceptible allele, with an average disease severity of 2.78.

Based on a *p*-value < 0.001 for MLM method, 60 trait-SNP associations were considered as candidates for *Xap* resistance (Figure 4; Supplementary Table S3). BLASTn search was used for all SNPs considered candidates for *Xap* resistance and the functionally annotated SNPs likely involved with disease response are presented. Some SNPs were detected in *P. persica* uncharacterized genes and, for this reason, not mentioned.

Out of the 60 trait-SNP associations identified by MLM analysis, 30 SNPs were concordant to GLM method with *p*-value < 0.0001 (Supplementary Table S3). MLM model includes an additional covariate of kinship so that false discovery of association is better regulated than in GLM method, because there is an effective control for population structure and relatedness.

Majority of the 60 SNP markers were localized on chrom 1 (36 SNPs). From the 2.59 Mbp region physically located between 44.06 Mbp and 46.65 Mbp, eight additional SNPs were identified within the region. Spanning up to 47.66 Mbp and down to 41.50 Mbp, additional nine SNPs in both sides of the region were detected, resulting in the co-localized region identified by the GLM approach.

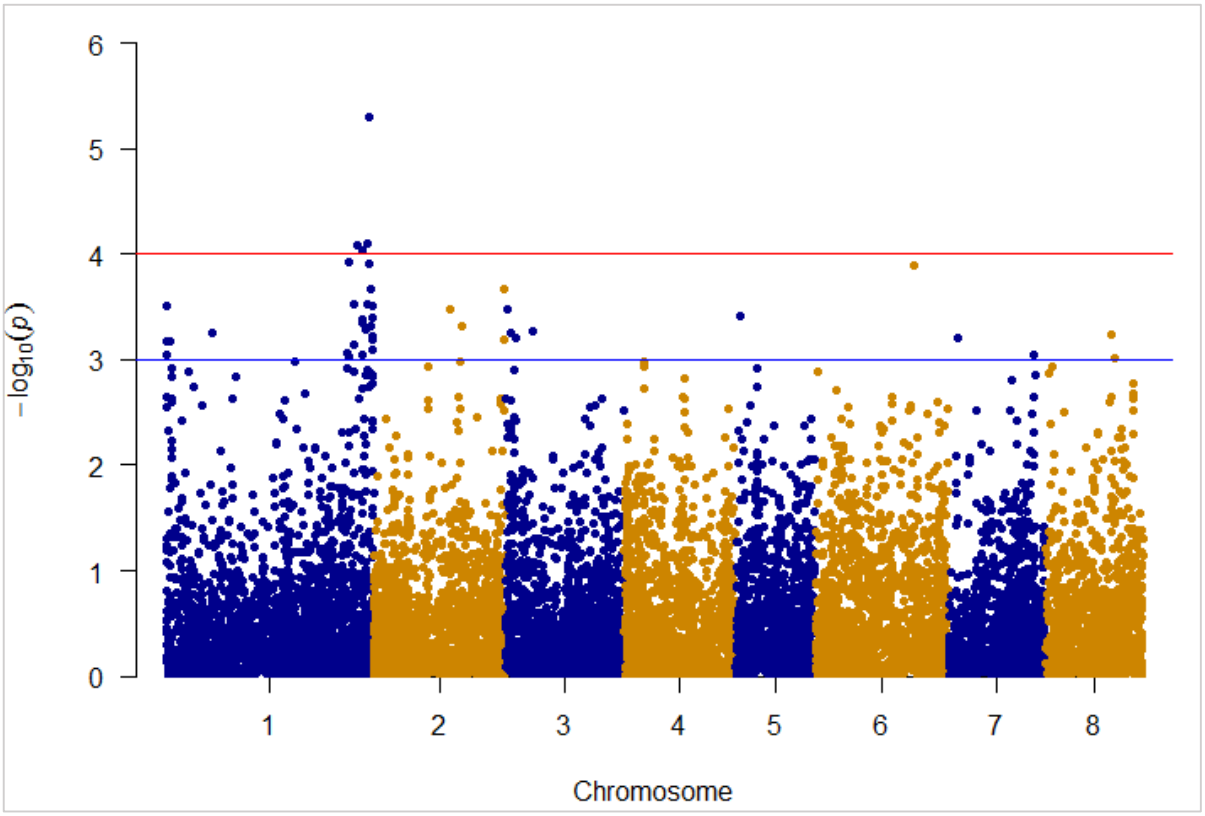


Figure 4 - Genome-wide association study (GWAS) for *Xap* resistance in peach. The 184 peach genotypes were scanned with 18K SNPs using a MLM approach taking into account both population structure and genetic relatedness (Q + K). The vertical axis plots the $-\log_{10}(p)$ values of the association between the SNP markers and disease severity. The horizontal orange line denotes a significance threshold of $p < 0.0001$ and the blue line a p-value < 0.001.

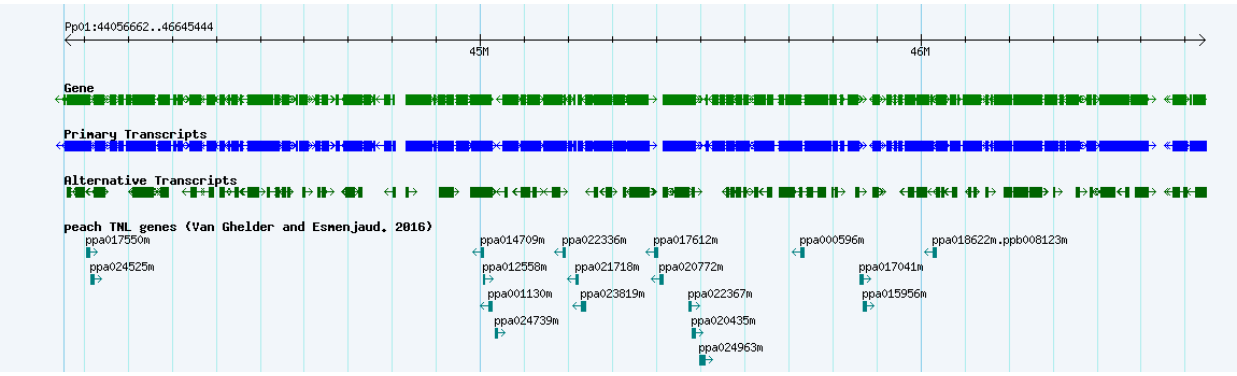


Figure 5 - Screen shot image of peach chromosome 1, showing a 2.59 Mbp genomic region, housing 18 TIR-NB-LRR genes (TNL genes) and the four SNP markers with the strongest association to *Xap* resistance (p-value < 0.0001). Image was captured from rosaceae.org.

Four additional SNPs, in different physical locations on chrom 1, were also associated with *Xap* response (Table 1). The SNP S1_97027 was functionally annotated

in a gene that encodes a Zinc finger CCCH domain-containing protein 16. S1-179632 was housed in a Glucan endo-1,3-beta-glucosidase 6, related with physiological defense response. S1_199125 located in HT1 gene, encoding a serine/threonine-protein kinase, involved in signaling pathways and plant defense (AFZAL et al., 2008). Finally, the SNP S1_10545345 is housed in a gene that encodes a bidirectional sugar transporter (SWEET17).

Table 1 - Candidate genes identified in association with *Xap* resistance using MLM approach.

Chrom	**SNP**	**Functional annotation***
1	S1_97027	Zinc finger CCCH domain-containing protein 16
1	S1_179632	Glucan endo-1,3-beta-glucosidase 6
1	S1_199125	HT1 gene, encoding a serine/threonine-protein kinase
1	S1_10545345	bidirectional sugar transporter (SWEET17)
2	S2_30116222; S2_30148134; S2_30148135; S2_30148164; S2_30170988	cellulose synthase-like protein D3
3	S3_6284771; S3_6284813; S3_6284820	26.5 kDa heat shock protein (HSP)
6	S6_22698094	plasma membrane calcium-transporting ATPase 2
7	S7_19439914; S7_19439917; S7_19439918; S7_19439921	probable phospholipase A2 homolog 1
8	S8_15134675	RPP8 gene (type of R gene for the CC-NBS-LRR class)

* *Prunus persica* whole genome assembly v2.0 annotation v2.1 (VERDE et al., 2017) used for annotation.

The chrom 2 showed seven trait-SNP associations, two of these SNPs (S2_17483624 and S2_20331739) are flanking a region of 2.85 Mbp, which contains 16 TIR-NB-LRR genes (TNL genes) (Figure 6). This identified genomic region is also co-located within a consensus QTL on chrom 2 identified by Frett (2016), between 12.44 Mbp and 26.64 Mbp. Also in chrom 2, another genomic region of 54.77 Kbp with five SNPs, was found to be associated with *Xap* resistance. The sequence of this region was blasted and showed the presence of cellulose synthase-like protein D3 (Table 1). Cellulose synthases are required for secondary cell wall formation, which is one of the barriers that pathogens need overcome to successfully infect and colonize plant tissues (MIEDES et al., 2014).

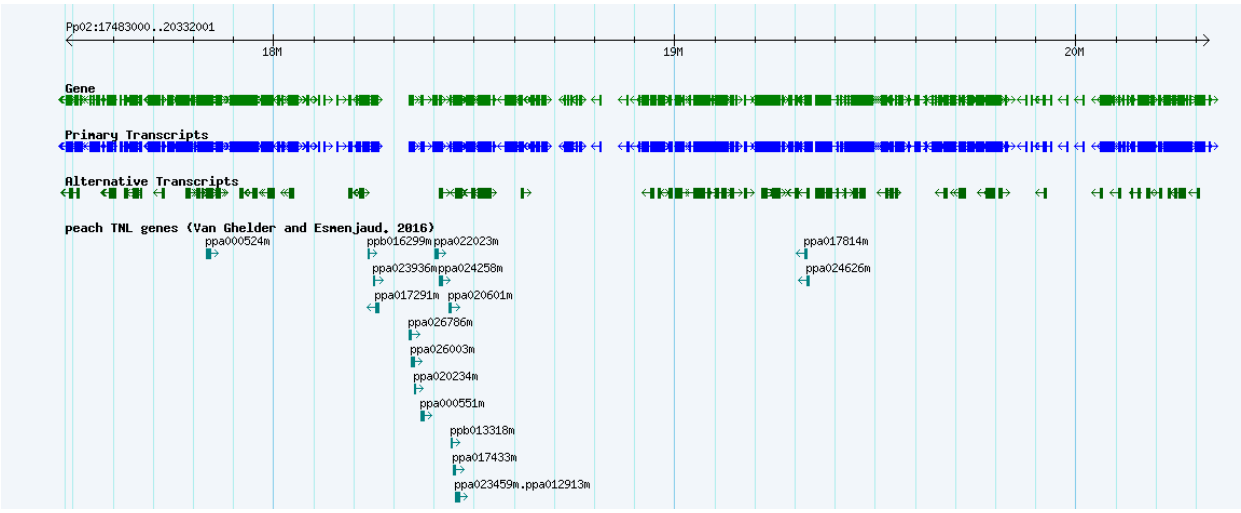


Figure 6 - Screen shot image of peach chromosome 2, showing a 2.85 Mbp genomic region, flanked by the markers S2_17483624 and S2_20331739, both significantly associated with *Xap* resistance (*p*-value < 0.001). This genomic region houses 16 TIR-NB-LRR genes (TNL genes). Image was captured from rosaceae.org.

The chrom 3 showed eight SNP associations. Here, emphasis is given to three SNP markers (S3_6284771, S3_6284813, S3_6284820) housed within a gene that encodes a 26.5 kDa heat shock protein (HSP). HSPs play an indispensable role as molecular chaperones and are key components for quality control of PRRs and R proteins in the innate immune response (PARK; SEO, 2015).

Only one candidate SNP (S6_22698094) was associated with *Xap* resistance in chrom 6. This SNP is housed in a plasma membrane calcium-transporting ATPase 2. Plasma membrane Ca^{2+} ATPases are general regulatory components of receptor kinases that mediate signaling pathways important for plant immunity and development, likely through the regulation of Ca^{2+} fluxes in the cytosol (FREI DIT FREY et al., 2012).

Four close SNPs in chrom 7 were housed in a probable phospholipase A2 homolog 1 (Table 1). Phospholipases were previously reported in plant defense signaling, with roles in oxylipin and jasmonate biosynthesis and also been correlated with reactive oxygen species (ROS) production (CANONNE et al., 2011).

Finally, the SNP S8_15134675 on chrom 8 was found to be also associated with *Xap* resistance. The RPP8 gene, a type of R gene for the CC-NBS-LRR class, is housing this SNP, encoding a disease resistant protein RPP8.

The use of SNPs derived from GBS approach were a powerful tool to elucidate *Xap* disease resistance. Some genomic regions previously identified by bi-parental crosses were not supported in our findings, probably due to different germplasm backgrounds.

Genotypes carrying favorable alleles for specific genomic regions could be used in breeding crosses aiming improve *Xap* resistance in peach. Moreover, the results from this study have revealed candidate SNP markers that can be exploited for use in marker-assisted selection and stacking of genes for *Xap* resistance in peach breeding. To fully understand the value of these loci in breeding for *Xap* resistance, the potentially novel loci identified in this study should be further validated by determining functional haplotypes, by comparing haplotypes with the phenotypic trait of interest. Validated markers underlying resistance will be, subsequently incorporated into MAS and used routinely in peach breeding for the prediction of *Xap* resistance.

3.4 Conclusions

GWAS approach validated and defined more accurately the known genomic regions underlying *Xap* resistance, as well as identified novel candidate genes that provide useful targets for further investigation. Several informative SNPs were functionally annotated in genes involved in defense mechanisms against pathogen infection, highlighting two genomic regions, located on chrm 1 (2.59 Mbp) and chrm 2 (2.85 Mbp), respectively, both housing several R genes. Our results provide new insights into breeding for *Xap* resistance in peach, with great potential for subsequent application of MAS.

CONSIDERAÇÕES FINAIS

A técnica de genotipagem por sequenciamento (GBS) foi altamente eficiente para a identificação direta de SNPs no germoplasma em estudo, possibilitando ampla cobertura do genoma.

Os resultados encontrados reforçam a necessidade de considerar a informação da estrutura genética em análises de GWAS para o controle de associações falso-positivas, uma vez que forte estrutura de população foi detectada, baseada principalmente em características relacionadas ao fruto, separando genótipos de polpa fundente e polpa não-fundente.

A extensão do desequilíbrio de ligação (LD) no germoplasma em estudo determina o número de marcadores necessários para obter sucesso em análises de associação. Forte LD foi observado em uma distância de até 38 Kbp no genoma, sendo a variação dos padrões de LD altamente dependente da subpopulação e das regiões genômicas em análise. O número de SNPs utilizados neste estudo (18K) possibilitou uma cobertura três vezes maior que a mínima necessária.

Avaliações de campo e através do bioensaio de folhas destacadas mostraram resultados confiáveis e complementares para identificar fontes resistentes à *Xap*. Grande variabilidade foi observada, permitindo a identificação de genótipos contrastantes para a característica de interesse. Os genótipos com maior resistência podem ser preferencialmente utilizados em áreas de produção com maior ocorrência da doença, ou como genitores no programa de melhoramento visando aumentar o nível de resistência.

Análises de GWAS validaram e definiram com maior precisão as regiões genômicas identificadas contendo genes associados à resistência ao patógeno em

estudos anteriores, bem como possibilitaram a identificação de novos genes candidatos que necessitam ser melhor estudados. Vários SNPs informativos foram funcionalmente anotados em genes envolvidos em mecanismos de defesa à infecção por patógenos, com destaque para duas regiões genômicas, localizadas no cromossomo 1 (2,59 Mpb) e 2 (2,85 Mpb), respectivamente, ambas identificadas com vários genes R.

Os resultados encontrados abrem novas perspectivas para o melhoramento genético visando resistência à *Xap*, com grande potencial para subsequente aplicação de seleção assistida por marcadores. Além disso, os SNPs de alta qualidade identificados por GBS podem ser utilizados para identificar regiões genômicas associadas à diversas características de interesse no melhoramento genético do pessegueiro, dependendo apenas de uma fenotipagem eficiente para a caráter e assim incorporando a seleção assistida por marcadores moleculares aos programas.

REFERÊNCIAS

AFZAL, A. J.; WOOD, A. J.; LIGHTFOOT, D. A. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. **Molecular Plant-Microbe Interactions**, v. 21, n. 5, p. 507-517, 2008.

ANDERSON, C. A.; PETTERSSON, F. H.; CLARKE, G. M.; CARDON, L. R.; MORRIS, A. P.; ZONDERVAN, K. T. Data quality control in genetic case-control association studies. **Nature Protocols**, v. 5, n. 9, p. 1564-1573, 2010.

ARANZANA, M. J.; ABBASSI, E.; HOWAD, W.; ARÚS, P. Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. **BMC Genetics**, v. 11, p. 69, 2010.

BADENES, M. L.; FERNÁNDEZ I MARTÍ, A.; RÍOS, G.; RUBIO-CABETAS, M. J. Application of genomic technologies to the breeding of trees. **Frontiers in Genetics**, v. 7, n. 198, p. 1-13, 2016.

BARABASCHI, D.; TONDELLI, A.; DESIDERIO, F.; VOLANTE, A.; VACCINO, P.; VALÈ, G.; CATTIVELLI, L. Next generation breeding. **Plant Science**, v. 242, p. 3-13, 2016.

BIELENBERG, D. G.; RAUH, B.; FAN, S.; GASIC, K.; ABBOTT, A. G.; REIGHARD, G. L.; OKIE, W. R.; WELLS, C. E. Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. **PLOS ONE**, v. 10, n. 10, p. e0139406, 2015.

BLISS, F. A. Marker-Assisted Breeding in Horticultural Crops. **Acta Horticulturae**, v. 859, p. 339-350, 2010.

BRACHI, B.; MORRIS, G. P.; BOREVITZ, J. O. Genome-wide association studies in plants: the missing heritability is in the field. **Genome Biology**, v. 12, p. 232, 2011.

BRADBURY, P. J.; ZHANG, Z.; KROON, D. E.; CASSTEVENS, T. M.; RAMDOSS, Y.; BUCKLER, E. S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633–2635, 2007.

BYRNE, D. H. Founding clones of low chilling fresh market peach germplasm developed in the USA and Brazil. **Acta Horticulture**, v. 606, p. 17-21, 2003.

BYRNE, D. H.; RASEIRA, M. B.; BASSI, D.; PIAGNANI, M. C.; GASIC, K.; REIGHARD, G. L.; MORENO, M. A.; PÉREZ, S. Peach. In: BADENES, M. L.; BYRNE, D. H. (Eds.). **Fruit Breeding**: Handbook of Plant Breeding. New York: Springer Science+Business Media, 2012. p. 505-569.

BYRNE, D. H.; SHERMAN, W. B.; BACON, T. A. Stone fruit genetic pool and its exploitation for growing under warm winter conditions. In: EREZ, A. (Ed.). **Temperate fruit crops in warm climates**. Dordrecht: Kluwer Academic Publishers, 2000. p. 157-230.

CANONNE, J.; FROIDURE-NICOLAS, S.; RIVAS, S. Phospholipases in action during plant defense signaling. **Plant Signaling & Behavior**, v. 6, n. 1, p. 13-18, 2011.

CAO, K.; WANG, L.; ZHU, G.; FANG, W.; CHEN, C.; LUO, J. Genetic diversity, linkage disequilibrium, and association mapping analyses of peach (*Prunus persica*) landraces in China. **Tree Genetics and Genomes**, v. 8, p. 975-990, 2012.

CAO, K.; ZHENG, Z.; WANG, L.; LIU, X.; ZHU, G.; FANG, W. et al. Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. **Genome Biology**, v. 15, n. 7, p. 415, 2014.

CAO, K.; ZHOU, Z.; WANG, Q.; GUO, J.; ZHAO, P.; ZHU, G.; FANG, W.; CHEN, C.; WANG, X.; WANG, X.; TIAN, Z.; WANG, L. Genome-wide association study of 12 agronomic traits in peach. **Nature communications**, v. 7, p. 13246, 2016.

CHAVEZ, D. J.; BECKMAN, T. G.; WERNER, D. J.; CHAPARRO, J. X. Genetic diversity in peach [*Prunus persica* (L.) Batsch] at the University of Florida: past, present and future. **Tree Genetics and Genomes**, v. 10, p. 1399-1417, 2014.

DAVEY, J. W.; HOHENLOHE, P. A.; ETTER, P. D.; BOONE, J. Q.; CATCHEN, J. M.; BLAXTER, M. L. Genome-wide genetic marker discovery and genotyping using next generation sequencing. **Nature Reviews Genetics**, v. 12, p. 499-510, 2011.

DESCHAMPS, S.; LLACA, V.; MAY, G. D. Genotyping-by-sequencing in plants. **Biology**, v. 1, p. 460-483, 2012.

DELLAPORTA, S. L.; WOOD, J.; HICKS, J. B. A plant DNA minipreparation: version II. **Plant Molecular Biology Reporter**, v. 1, n. 14, p. 19-21, 1983.

DIRLEWANGER, E.; GRAZIANO, E.; JOOBEUR, T.; GARRIGA-CALDERÉ, F.; COSSON, P.; HOWAD, W.; ARÚS, P. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 26, p. 9891–9896, 2004.

ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S.; MITCHELL, S. E. A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. **PLOS ONE**, v. 6, n. 5, p. e19379, 2011.

ENDEKMAN, J. B.; JANNINK, J. L. Shrinkage Estimation of the Realized Relationship Matrix. **G3: Genes Genomes Genetics**, v. 2, p. 1405-1413, 2012.

EPPO. EPPO Global Database, 2017. Available online: <https://gd.eppo.int/taxon/XANTPR/distribution>. Accessed 22 January 2018.

FAOSTAT data (2014) http://www.fao.org/faostat/en/#data/QC. Acessado em 20 de novembro de 2017.

FAUST, M.; TIMON, B. Origin and dissemination of peach. **Horticultural Reviews**, v. 17, p. 331-379, 1995.

FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. **Annual Review of Plant Biology**, v. 54, p. 357-374, 2003.

FONT I FORCADA, C.; ORAGUZIE, N.; IGARTUA, E.; MORENO, M.A.; GOGORCENA, Y. Population structure and marker-trait associations for pomological traits in peach and nectarine cultivars. **Tree Genetics and Genomes**, v.9, p.331-349, 2013.

FREI DIT FREY, N.; MBENGUE, M.; KWAAITAAL, M.; NITSCH, L.; ALTENBACH, D.; HÄWEKER, H.; LOZANO-DURAN, R.; NJO, M. F.; BEECKMAN, T.; HUETTEL, B.; BORST, J. W.; PANSTRUGA, R.; ROBATZEK, S. Plasma membrane calcium ATPases are important components of receptor-mediated signaling in plant immune responses and development. **Plant Physiology**, v.159, p. 798-809, 2012.

FRETT, Terrence J. **Genetic determinism of *Xanthomonas arboricola* pv. *pruni* (*Xap*) resistance, fruit quality, and phenological traits in peach and incorporation of marker-assisted selection (MAS) in the University of Arkansas peach and nectarine breeding program.** 2016. 734f. Ph.D. Dissertation in Plant Science, University of Arkansas, Fayetteville, 2016.

FRESNEDO-RAMÍREZ, J.; BINK, M. C. A. M.; VAN DE WEG, E.; FAMULA, T. R.; CRISOSTO, C. H.; FRETT, T. J.; GASIC, K.; PEACE, C. P.; GRADZIEL, T. M. QTL mapping of pomological traits in peach and related species breeding germplasm. **Molecular Breeding**, v. 35, p. 166, 2015.

FRESNEDO-RAMÍREZ, J.; FRETT, T. J.; SANDEFUR, P. J.; SALGADO-ROJAS, A.; CLARK, J. R.; GASIC, K.; PEACE, C. P.; ANDERSON, N.; HARTMANN, T. P.; BYRNE, D. H.; BINK, M. C. A. M.; VAN DE WEG, E.; CRISOSTO, C. H.; GRADZIEL, T. M. QTL mapping and breeding value estimation through pedigree-based analysis of fruit size and weight in four diverse peach breeding programs. **Tree genetics and genomes**, v. 12, p. 25, 2016.

FRETT, T. J.; REIGHARD, G. L.; OKIE, W. R.; GASIC, K. Mapping quantitative trait loci associated with blush in peach [*Prunus persica* (L.) Batsch]. **Tree Genetics and Genomes**, v. 10, p. 377-381, 2014.

GARDNER, K. M.; BROWN, P.; COOKE, T. F.; CANN, S; COSTA, F.; BUSTAMANTE, C.; VELASCO, R.; TROGGIO, M.; MYLES, S. Fast and cost-effective genetic mapping in apple using next-generation sequencing. **G3: Genes Genomes Genetics**, v. 4, n. 9, p. 1681-1687, 2014.

GASIC, K.; REIGHARD, G.; OKIE, W.; CLARK, J.; GRADZIEL, T.; BYRNE, D.; PEACE, C.; STEGMEIR, T.; ROSYARA, U.; IEZZONI, A. Bacterial Spot Resistance in Peach: Functional Allele Distribution in Breeding Germplasm. **Acta Horticulture**, v. 1084, p. 69-74, 2015.

GLAUBITZ, J. C.; CASSTEVENS, T. M.; LU, F.; HARRIMAN, J.; ELSHIRE, R. J.; SUN, Q.; BUCKLER, E. S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. **PLOS ONE**, v. 9, n. 2, p. e90346, 2014.

GUAJARDO, V.; SOLÍS, S.; SAGREDO, B.; GAINZA, F.; MUÑOZ, C.; GASIC, K.; HINRICHSEN, P. Construction of high density sweet cherry (*Prunus avium* L.) linkage maps using microsatellite markers and SNPs detected by genotyping-by-sequencing (GBS). **PLOS ONE**, v. 10, n. 5, p. e0127750, 2015.

GÜRCAN, K.; TEBER, S.; ERCISLI, S.; YILMAZ, K. U. Genotyping by Sequencing (GBS) in Apricots and Genetic Diversity Assessment with GBS-Derived Single-Nucleotide Polymorphisms (SNPs). **Biochemical Genetics**, v. 54, n. 6, p. 854-885, 2016.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z.; LIU, H.; LI, Z. Genotyping-by sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Frontiers in Plant Science**, v. 5, p. 484, 2014.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v. 31, n. 2, p. 423-447, 1975.

JOMBART, T.; AHMED, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. **Bioinformatics**, v. 27, n. 21, p. 3070-3071, 2011.

JONES, J. D. G.; DANGL, J. L. The plant immune system. **Nature**, v. 444, n. 16, p. 323-329, 2006.

KHAN, M. K.; KORBAN, S. S. Association mapping in forest trees and fruit crops. **Journal of Experimental Botany**, v. 63, n. 11, p. 4045-4060, 2012.

KIM, C.; GUO, H.; KONG, W.; CHANDNANI, R.; SHUANG, L.; PATERSON, A. H. Application of genotyping by sequencing technology to a variety of crop breeding programs. **Plant Science**, v. 242, p. 14-22, 2016.

KORTE, A.; FARLOW, A. The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods**, v. 9, p. 29, 2013.

KRETZSCHMAR, A. A.; ROSSETTO, E. A.; MARTINS, O. M. Resistência de algumas cultivares de pessegueiro a *Xanthomonas campestris* pv. *pruni* inoculadas por infiltração em folhas destacadas. **Revista Brasileira de Fruticultura**, v. 20, n. 2, p. 213-219, 1998.

KUMAR, S.; KIRK, C.; DENG, C.; WIEDOW, C.; KNAEBEL, M.; BREWER, L. Genotyping-by-sequencing of pear (*Pyrus* spp.) accessions unravels novel patterns of genetic diversity and selection footprints. **Horticulture Research**, v. 4, p. 17015, 2017.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357-359, 2012.

LI, X.; MENG, X.; JIA, H.; YU, M.; MA, R.; WANG, L.; CAO, K.; SHEN, Z.; NIU, L.; TIAN, J.; CHEN, M.; XIE, M.; ARÚS, P.; GAO, Z.; ARANZANA, M. J. Peach genetic resources: diversity, population structure and linkage disequilibrium. **BMC Genetics**, v. 14, p. 84, 2013.

MAIA, M. C.C.; RESENDE, M. D. V.; PAIVA, J. R.; CAVALCANTI, J. J. V.; BARROS, L. M. Seleção simultânea para produção, adaptabilidade e estabilidade genotípicas em clones de cajueiro, via modelos mistos. **Pesquisa Agropecuária Tropical**, v. 39, n. 1, p. 43-50, 2009.

MARTINS, O. M. Evaluation of virulence of strains of *Xanthomonas campestris* pv*. pruni* on peach and plum cultivars. **Fruit Varieties Journal**, v. 50, n. 4, p. 221-225, 1996.

MEDEIROS, J. G. S.; CITADIN, I.; SANTOS, I.; ASSMANN, A. P. Reaction of peach tree genotypes to bacterial leaf spot caused by *Xanthomonas arboricola* pv. *pruni*. **Scientia Agricola**, v. 68, p. 57-61, 2011.

MICHELETTI, D.; DETTORI, M. T.; MICALI, S.; ARAMINI, V.; PACHECO, I.; LINGE, C. S. et al. Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. **PLOS ONE**, v. 10, n. 5, p. e0136803, 2015.

MIEDES, E.; VANHOLME, R.; BOERJAN, W.; MOLINA, A. The role of the secondary cell wall in plant resistance to pathogens. **Frontiers in Plant Science**, v. 5, n. 358, p. 1-13, 2014.

NEI, M. Genetic distances between populations. **The American naturalist**, v. 106, n. 949, p. 283-292, 1972.

OEPP/EPPO. *Xanthomonas arboricola* pv. *pruni*. **Bulletin OEPP/EPPO**, Bulletin 36, p. 129-133, 2006.

OKIE, W. R. **Handbook of peach and nectarine varieties**: performance in the southeastern United States and index of names. Washington: USDA/ARS Agriculture, 1998. 808 p.

PALACIO-BIELSA, A.; BERRUETE, I. M.; LÓPEZ, M. M.; PENALVER, J.; MORENTE, C.; CUBERO, J.; PENALVER, J.; MORENTE, C.; CUBERO, J.; GARITA-CAMBRONERO, J.; SABUQUILLO, P.; REDONDO, C.; MITIDIERI, M.; GOMES, C. B.; UENO, B.; CASTRO, L. A. S. de; LEONI, C.; SILVEIRA, E. La mancha bacteriana de los frutales de hueso y del almendro (*Xanthomonas arboricola* pv. *pruni)* en España y Sudamérica. **Phytoma España**, v. 271, p. 21-28, 2015.

PANGULURI, S. K.; KUMAR, A. A. **Phenotyping for Plant Breeding**: applications of phenotyping methods for crop improvement. New York: Springer Science+Business Media, 2013. 211 p.

PARK, C.; SEO, Y. Heat Shock Proteins: A Review of the Molecular Chaperones for Plant Immunity. **Plant Pathology Journal**, v. 31, n. 4, p. 323-333, 2015.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, n. 3, p. 545-554, 1971.

PETERSON, G. W.; DONG, Y.; HORBACH, C.; FU, Y. Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. **Diversity**, v. 6, p. 665-680, 2014.

POLAND, J. A.; RIFE, T. W. Genotyping-by-sequencing for plant breeding and genetics. **The Plant Genome**, v. 5, n. 3, p. 92-102, 2012.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; BAKKER, P. I. W.; DALY, M. J.; SHAM, C. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American journal of human genetics**, v. 81, n. 3, p. 559-575, 2007.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. Disponível em: https://www.R-project.org/.

RAJ, A.; STEPHENS, M.; PRITCHARD, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. **Genetics**, v. 197, p. 573-589, 2014.

RANDHAWA, P. S.; CIVEROLO, E. L. A detached-leaf bioassay for *Xanthomonas campestris* pv. *pruni*. **Phytopathology**, v. 75, n. 9, p.1060-1063, 1985.

RASEIRA, M. C. B.; BYRNE, D. H.; FRANZON, R. C. Pessegueiro: tradição e poesia. In: BARBIERI, R. L.; STUMPF S. T. (Eds.). **Origem e evolução de plantas cultivadas**. Brasília: Embrapa Informação Tecnológica, 2008. p. 679-705.

RASEIRA, M. C. B.; CASTRO, C. M.; UENO, B. Intercâmbio e utilização de recursos genéticos de Prunóideas. In: MARIANTE, A. S; SAMPAIO, M .J. A.; INGLIS, M. C. V. (Org.). **Informe nacional sobre a situação dos recursos ftogenéticos para a alimentação e a agricultura do Brasil**. Brasília (DF): Embrapa, 2008b. p. 54-56.

RASEIRA, M. C. B.; HERTER, F.; POSSER, C. A. S. The EMBRAPA/ Clima Temperado peach breeding program and adaptation to subtropical regions. **Acta Horticulturae**, v. 606, p. 45-50, 2003.

RASEIRA, M. C. B.; NAKASU, B. H. Peach breeding program in Southern Brazil. **Acta Horticulturae**, v. 713, p. 93-97, 2006.

RASEIRA, M. C. B.; NAKASU, B. H. Breeding peaches for mild winters: Recent results of the non-melting peach breeding program of Embrapa, in Southern Brazil. **Acta Horticulturae**, v. 962, p. 29-34, 2012.

RASEIRA, M. C. B.; NAKASU, B. H.; BARBOSA, W. Cultivares: descrição e recomendação. In: RASEIRA, M. C. B.; PEREIRA, J. F. M.; CARVALHO, F. L. C. (Eds.). **Pessegueiro**. Brasília: Embrapa, 2014. p. 73-141.

RASEIRA, M. C. B.; NAKASU, B. H.; SANTOS, A. M.; FORTES, J. F.; MARTINS, O. M.; RASEIRA, A.; BERNARDI, J. The CNPFT/EMBRAPA fruit breeding program in Brazil. **HortScience**, v. 27, n. 11, p. 1154-1157, 1992.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007a. 561 p.

RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: Embrapa Florestas, 2004. 57 p.

RESENDE, M. D. V. **Selegen-REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Colombo: Embrapa Florestas, 2007b. 359 p.

RESENDE, M. D. V. Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breeding and Applied Biotechnology**, v. 16, p. 330-339, 2016.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v. 37, n. 3, p. 182-194, 2007.

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional**: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão, aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência. Viçosa: Suprema, 2014. 881 p.

RITCHIE, David F. Bacterial spot. In: OGAWA, J. M.; ZEHR, E. I.; BIRD, G. W.; URIU, K.; UYEMOTO, J. K. (Eds.) **Compendium of stone fruit diseases**. St. Paul: APS Press, 1995. p. 50-52.

ROSENBERG, N. A. Distruct: a program for the graphical display of population structure. **Molecular Ecology Notes**, v. 4, p.137-138, 2004.

RU, S.; MAIN, D.; EVANS, K.; PEACE, C. Current applications, challenges, and perspectives of marker-assisted seedling selection in Rosaceae tree fruit breeding. **Tree Genetics and Genomes**, v. 11, 2015.

SACHET, M. R.; CITADIN, I.; SCARIOTTO, S.; SANTOS, I.; ZYDEK, H.; RASEIRA, M. B. Reaction of peach genotypes to bacterial leaf spot: correlations with environmental conditions, leaf phenology, and morphology. **HortSciense**, v. 48, n. 1, p. 28-33, 2013.

SALAZAR, J. A.; PACHECO, I.; SHINYA, P.; ZAPATA, P.; SILVA, C.; ARADHYA, M.; VELASCO, D.; RUIZ, D.; MARTÍNEZ-GÓMEZ, P.; INFANTE, R. Genotyping by sequencing for SNP-based linkage analysis and identification of QTLs linked to fruit quality traits in japanese plum (*Prunus salicina* Lindl.). **Frontiers in Plant Science**, v. 8, p. 476, 2017.

SOCQUET-JUGLARD, D.; DUFFY, B.; POTHIER, J.F.; CHRISTEN, D.; GESSLER, C.; PATOCCHI, A. Identification of a major QTL for *Xanthomonas arboricola* pv. *pruni* resistance in apricot. **Tree Genetics and Genomes**, v. 9, p. 409-421, 2013.

SOTO-CERDA, B. J.; CLOUTIER, S. Association mapping in plant genomes. In: CALISKAN, M (Ed.). **Genetic Diversity in Plants**, 2012. p. 29-54.

STANSFIELD, W. D. **Genética**. São Paulo: McGraw-Hill do Brasil, 1974. 958 p

STEFANI, E. Economic significance and control of bacterial spot/canker of stone fruits caused by *Xanthomonas arboricola* pv. *pruni*. **Journal of Plant Pathology**, v. 92, p. 99-103, 2010.

STURION, J. A.; RESENDE, M. D. V. Eficiência do delineamento experimental e capacidade de teste no melhoramento genético da erva-mate (*Ilex paraguariensis* St. Hil). **Boletim de Pesquisa Florestal**, n. 50, p. 3-10, 2005.

THUROW, L. B.; RASEIRA, M. B.; UENO, B.; BONOW, S.; CASTRO, M. C. Caracterização fenotípica para resistência à bacteriose (*Xanthomonas arborícola* pv. *pruni*) em germoplasma de pessegueiro. In: Encontro de Iniciação Científica e Pós-

graduação da Embrapa Clima Temperado, 6., 2016, Pelotas. **Anais**... Pelotas: Embrapa Clima Temperado, 2016. p. 282-284.

THUROW, L. B.; RASEIRA, M. C. B.; BONOW, S.; ARGE, L. W. P.; CASTRO, C. M. Population genetic analysis of Brazilian peach breeding germplasm. **Revista Brasileira de Fruticultura**, v. 39, n. 5, p. e-166, 2017.

VAN GHELDER, C.; ESMENJAUD, D. TNL genes in peach: insights into the postLRR domain. **BMC Genomics**, v. 17, p. 317, 2016.

VARSHNEY, R. K.; TERAUCHI, R.; McCOUCH, S. R. Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. **PLOS Biology**, v. 12, n. 6, p. e1001883, 2014.

VERDE, I.; ABBOTT, A. G.; SCALABRIN, S.; JUNG, S.; SHU, S. Q.; MARRONI, F. et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. **Nature genetics**, v. 45, p. 487-494, 2013.

VERDE, I.; BASSIL, N.; SCALABRIN, S.; GILMORE, B.; LAWLEY, C. T.; GASIC, K. et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. **PLOS ONE**, v.7, p. e35668, 2012.

VERDE, I.; JENKINS, J.; DONDINI, L.; MICALI, S.; PAGLIARANI, G.; VENDRAMIN, E. et al. The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. **BMC Genomics**, v. 18, p. 225, 2017.

YANG, Nannan**. Mapping quantitative trait loci associated with resistance to bacterial spot (*Xanthomonas arboricola* pv. *pruni*) in peach.** 2012. 158f. Ph.D. Dissertation in Plant and Environmental Sciences, Clemson University, Clemson, 2012.

YANG, N.; REIGHARD, G.; RITCHIE, D.; OKIE, W.; GASIC, K. Mapping quantitative trait loci associated with resistance to bacterial spot (*Xanthomonas arboricola* pv. *pruni*) in peach. **Tree Genetics and Genomes**, v. 9, p. 573-586, 2013.

VITAE

LIANE BAHR THUROW, filha de Neldo Buss Thurow e Ilma Bahr Thurow, nasceu na cidade de Canguçu, Rio Grande do Sul, em 18 de julho de 1988. Em 2007, ingressou no Curso de Agronomia da Universidade Federal de Pelotas (UFPel), graduando-se em março de 2012. Durante a graduação realizou estágio extra-curricular no Laboratório de Pós-Colheita, Industrialização e Qualidade de Grãos sob orientação do prof. Dr. Moacir Cardoso Elias por dois semestres e foi bolsista do Programa Institucional de Bolsas de Iniciação Científica (PIBIC-CNPq) durante três anos, desenvolvendo atividades no laboratório de Cultura de Tecidos e Biologia Molecular sob orientação do prof. Dr. Valmor João Bianchi. Realizou estágio de conclusão de curso no Laboratório de Biologia Molecular da Embrapa Clima Temperado, por seis meses, sob orientação do Dr. Sandro Bonow. Em março de 2012, ingressou no Mestrado pelo Programa de Pós-Graduação em Agronomia na área de concentração em Fitomelhoramento (UFPel), sob a orientação da Dra. Caroline Marques Castro. Após um ano e meio de curso fez Progressão de Nível, tornando-se aluna de Doutorado desta mesma Instituição. Foi contemplada com uma bolsa de estudos do CNPq na modalidade de doutorado sanduiche no exterior (SWE), permanecendo por 12 meses (11/2013 a 10/2014) na *Clemson University* situada na Carolina do Sul nos Estados Unidos, sob a orientação da Dra. Ksenija Gasic, desenvolvendo atividades no programa de melhoramento de pessegueiro relacionadas a nova geração de sequenciamento, desenho e validação de marcadores para qualidade de fruto e resistência a doenças e desenvolvimento de protocolo visando fenotipagem para resistência a bacteriose em pessegueiro.

APÊNDICES

Table S1 - Cultivars and advanced selections of Brazilian peach breeding germplasm genotyped by GBS, reported parentage, origin, fruit characteristics and population stratification.

ID	Genotype	Reported parentage [a]	Origin [b]	Fruit traits[c]	Stratification[d]
BP001	Abóbora	unknown	Embrapa	PYNRC	POP III
BP002	Ágata	(68201041 x C2 R19 T182) F2	Embrapa	PYNRC	POP III
BP003	Aldrighi	unknown	Embrapa	PYNRC	POP III
BP004	Alpes	Aldrighi - seleção 439 x Tapes	Embrapa	PYNRC	POP III
BP005	Amarillo	unknown	Bolivia	N??RS	POP III
BP006	Anita	unknown	Embrapa	NWMRC	POP I
BP007	Apote	unknown	Bolivia	P?NR?	POP III
BP008	Arlequim	Lake City x Toschina	IAC	PWMRF	ADM
BP009	Atenas	Jade op.	Embrapa	PYNRC	POP III
BP010	Aurora 1	Ouromel-3 op.	IAC	PYMRC	POP I
BP011	Aurora 2	unknown	IAC	PYMRC	ADM
BP012	Y-babcock	unknown	Embrapa	PYNRC	POP III
BP013	Barbosa	unknown	Embrapa	PWMRF	ADM
BP014	Bolinha	unknown	Embrapa	PYNRC	POP III
BP015	BR1	Delicioso x Panamint	Embrapa	PWMRC	POP I
BP016	BR3	Pala op. = (Coral x Panamint) op.	Embrapa	PWMRS	POP I
BP017	Cai	Delicioso x Lake City	Embrapa	PWMRC	ADM
BP018	Capdeboscq	(Lake city x Intermediário) op.	Embrapa	PYNRC	ADM
BP019	Cardeal	(338-90 FV) op. = (24401 x 17825) op.	Embrapa	PYMRC	ADM
BP020	Regalo	Chula x Chimarrita	Embrapa	PWMRC	POP I
BP021	Cascata 805	Chimarrita x Della Nona	Embrapa	PWMRS	POP I
BP022	Cascata 828	Taquari 19 x FLA 6-12	Embrapa	PWMF?	POP I
BP023	Cascata 838	Taquari 19 x FLA 6-12	Embrapa	PYMFS	ADM
BP024	Fascínio	(Chimarrita x Linda) op.	Embrapa	PWMRS	POP I
BP025	Cascata 1303	Cascata 951 x Maciel	Embrapa	PYMRC	ADM
BP026	Cascata 1373	Cascata 828 op.	Embrapa	PYMFC	ADM
BP027	Cerrito	(Lake city x Intermediário) op.	Embrapa	PYNRC	POP III
BP028	Chimarrita	Babcock x Flordabella	Embrapa	PWMRS	POP I
BP029	Chiripá	Delicioso x Nectared 5	Embrapa	PWMRF	ADM
BP030	Conserva 334	unknown	Embrapa	PYNRC	ADM
BP_31	Conserva 594	(Capdeboscq x Madrugador) op.	Embrapa	PYNRC	POP III
BP032	Conserva 672	Topázio x Conserva 334	Embrapa	PYNRC	POP III
BP033	Conserva 930	Conserva 531 (=Convênio x RR 53272) x Eldorado	Embrapa	PYNRC	ADM
BP034	Conserva 947	Bolinha x P60 -22	Embrapa	PYNRC	POP III
BP035	Cascata 1281	(BR3 x A333) op.	Embrapa	PWMRC	POP I
BP036	Conserva 1578	Ametista x Conserva 594	Embrapa	PYNRC	POP III
BP037	Conserva 1596	Conserva 672 x Leonense	Embrapa	PYNRC	ADM
BP038	Conserva 1600	Conserva 672 x A334	Embrapa	PYNRC	POP III

BP039	Conserva 1612	Atenas x Riograndense	Embrapa	PYNRC	POP III
BP040	Conserva 1666	Conserva 1182 x Precocinho	Embrapa	PYNRC	POP III
BP041	Conserva 1798	Jubileu x Conserva 1248	Embrapa	PYNRC	POP III
BP042	Convênio	(Amsden x Abóbora) op.	Embrapa	PYNRC	POP III
BP043	Coral	(Delicioso x Interlúdio) op.	Embrapa	PWMRS	ADM
BP044	Coral 2	mutation of cultivar Coral	Embrapa	PWMRC	POP I
BP045	Cristal-taquari	unknown	Embrapa	PWNRC	POP III
BP046	Delicioso	unknown	Embrapa	PWMRF	POP I
BP047	Della Nona	(Delicioso x Nectared 5) op.	Embrapa	PWMRF	POP I
BP048	Diamante	Convênio x seleção Pelotas 77 (Cardeal x Aldrighi, op.)	Embrapa	PYNRC	POP III
BP049	Dulce	NJN67 x Pala (Coral x Panamint)	Embrapa	NWMRS	POP I
BP050	Edmundo Perret	unknown	Embrapa	PWMRC	ADM
BP051	Eldorado	Gaudério x Serrano	Embrapa	PYNRC	ADM
BP052	Eragil	unknown	USA	PYMRF	POP I
BP053	Esmeralda	Alpes x RR 37-201	Embrapa	PYNRC	POP III
BP054	Ewtrin	unknown	Trinidad	NWMRC	ADM
BP055	Farrapos	(Edmundo Perret x Aldrighi) op.	Embrapa	PYNRC	ADM
BP056	Flordabella	Fla 16-6 (Southland x Hawaiian, F2) x Flordawon	Florida	PYMRF	ADM
BP057	Flordaglo	Sundowner x Maravilha	Florida	PWMRC	POP I
BP058	FlordaGrande	FLA5-58 x (Flordasun x Springtime)	Texas	PYMRS	ADM
BP059	Flordaking	FLA9-67 x Early Amber	Florida	PYMRC	POP I
BP060	Flordaprince	FLA2-7 x Maravilha	Florida	PYMRC	POP I
BP061	Galaxy	P34-106 x D33-1	California	PWMFC	ADM
BP062	Gaúcho	unknown	Embrapa	PWMRS	ADM
BP063	Gaúcho de Porto Alegre	Delicioso op.	Embrapa	PWMRS	ADM
BP064	Gaudério	(Delicioso x Interlúdio) op.	Embrapa	PYMRC	POP I
BP065	Granada	Granito op. = (Alpes x Conserva 102) op.	Embrapa	PYNRC	ADM
BP066	Interlúdio	(Southland x Jewel) op.	Embrapa	PYMRS	ADM
BP067	Jade	(Alpes x RR 53-272) op.	Embrapa	PYNRC	POP III
BP068	Josefina	op. (Ouromel x Rubrosol) F2	IAC	NWMRF	ADM
BP069	Jubileu	Bolinha x Conserva 662	Embrapa	PYNRC	POP III
BP070	Kampai	Chimarrita x Flordaprince	Embrapa	PWMRS	ADM
BP071	Leonense	(Brilhante x NJC 97) F2	Embrapa	PYNRC	ADM
BP072	Libra	Conserva 594 (Capdeboscq x Madrugador) x Pepita	Embrapa	PYNRC	POP III
BP073	Linda	(NJ 238 x Sunred) op.	Embrapa	NYMRC	*
BP074	Lord	(Taquari 96) op. = (Abóbora x Taquari precoce) op.	Embrapa	PYNRC	ADM
BP075	Maciel	Conserva 171 x Conserva 334	Embrapa	PYNRC	POP III
BP076	Madrugador	(Aldrighi x Taquari precoce) op.	Embrapa	PYNRC	POP III
BP077	Magno	Ambrósio Perret x Tapes	Embrapa	PYNRC	POP III
BP078	Mara	Nectared 9 x Sunred (Rubrosol)	Embrapa	NYMRS	ADM

BP079	Marli	Delicioso x Prelúdio	Embrapa	PWMRS	ADM
BP080	Minuano	Carapuça op.	Taquari	PYMRS	ADM
BP081	Mollares Hierro	unknown	Canary Islands	PWMRF	POP III
BP082	Morro Redondo	(Lake city x Intermediário) op.	Embrapa	PYNRC	POP III
BP083	Natal	Suber x Tos China	IAC	PWMRC	POP III
BP084	Necta 468	BR3 x Necta IAPAR	Embrapa	NWMRC	POP I
BP085	Necta 480	Conserva 1134 x Sungen	Embrapa	NYNRC	ADM
BP086	Necta 496	Necta 343 (Caí x N. Remanso) x Sabrina	Embrapa	NWMRC	POP I
BP087	Necta 511	Conserva 1166 op. = (Cerrito x A 334 CN) op.	Embrapa	NYNRC	POP III
BP088	Necta 512	Conserva 1166 op. = (Cerrito x A 334 CN) op.	Embrapa	NYNRC	POP III
BP089	Necta 528	Dulce x Sunsnow	Embrapa	NWMRS	ADM
BP090	Necta 532	Necta 432 x Sunblaze	Embrapa	NYMRF	ADM
BP091	Okinawa	unknown	Japan	PWMRF	POP I
BP092	Olímpia	Bolinha x 7-28	Embrapa	PYNRC	POP III
BP093	Ônix	Farrapos op. x unknown	Embrapa	PYNRC	POP III
BP094	Pampeano	unknown	Embrapa	PWMRS	ADM
BP095	Panamint	(Babcock x Boston) x (Goldmine x Rio Oso Gem)	California	NYMRF	POP I
BP096	Pepita	Precocinho op.	Embrapa	PYNRC	POP III
BP097	Pérola de Itaquera	unknown	IAC	PWNRC	POP III
BP098	Pilcha	Precoce Rosado op.	Embrapa	PYMRC	POP I
BP099	Piratini	Ambrósio Perret x Tapes	Embrapa	PYNRC	POP III
BP100	Planalto	Coral x Babcock	Embrapa	PWNRC	POP I
BP101	Precocinho	Diamante op.	Embrapa	PYNRC	POP III
BP102	Premier	(Cardeal x 15 de novembro) F2	Embrapa	PWMRS	ADM
BP103	Princesa	Hawai x Southland	Embrapa	PYMRS	POP I
BP104	Real	unknown	IAC	PYNRC	POP III
BP105	Riograndense	(Brilhante x seleção NJC 97) op.	Embrapa	PYNRC	ADM
BP106	Rubimel	Chimarrita x Flordaprince	Embrapa	PYMRS	POP I
BP107	Safira	Ambrósio Perret x Cerrito	Embrapa	PYNRS	POP III
BP108	São Pedro	Flordasun x Springtime	Florida	PYMRS	POP I
BP109	Santa Aurea	Cerrito x NJC 88	Embrapa	PYNRC	POP III
BP110	Sentinela	Premier op.	Embrapa	PWMRC	POP I
BP111	Sinuelo	Prelúdio x Amarelinho	Embrapa	PYMRS	POP I
BP112	Sulina	Princesa x Premier	Embrapa	PWMRC	POP I
BP113	Sunblaze	FLA3-4N x FLA5-9	Florida	NYMRS	ADM
BP114	Sunhigh	J. H. Hale x (Carman x Slappey)	New Jersey	PYMRS	ADM
BP115	Sunmist	Flordaglo x Mayfire	Florida	NWMRC	POP I
BP116	Taquari 80	Delicioso x Interlúdio	Taquari	PWMRS	POP I
BP117	Tarumã	Aldrighi x (Amsdem x Abóbora) op.	Embrapa	PYNRC	POP III
BP118	Topázio	[Convênio x Pelotas 76 (=Aldrighi x Taquari precoce)] op.	Embrapa	PYNRC	POP III

BP119	Tropic Beauty	FLA3-2 x Flordaprince	Florida	PYMRC	POP I
BP120	Tropic Blush	{[(Southland x Jewel) op.] x Kaygold} op.	Florida/Texas	PYMRC	POP I
BP121	Tropic snow	FLA7-11 x Maravilha	Texas	PWMRS	POP I
BP122	Turquesa	(Convênio x Cerrito) op.	Embrapa	PYNRC	ADM
BP123	Vanguarda	(Alpes x RR 55-272) op.	Embrapa	PYNRC	POP III
BP124	Alvorada	Cardeal op.	Embrapa	PYMRC	ADM
BP125	Âmbar	Esmeralda x Conserva 555	Embrapa	PYNRC	ADM
BP126	Ametista	(Alpes x RR-37-201) op.	Embrapa	PYNRC	ADM
BP127	Aztec Gold	[(Mexican Cling x Sunred) F2] op.	Mexico	PYNRC	ADM
BP128	Babygold 7	(Lemon Free x PI35201) x [(J. H. Hale x Goldfinch) op.]	New Jersey	PYNRC	ADM
BP129	Babygold 9	PI35201 x PI43137	New Jersey	PYNRC	ADM
BP130	Baronesa	[(Hawai x Southland) op.] op.	Taquari	PYMRS	POP I
BP131	Blancona	unknown	Bolivia	PWMRF	POP III
BP132	Bonão	[Conserva 594 (=Capdeboscq x Madrugador)] x Pepita	Embrapa	PYNRC	POP III
BP133	Conserva 985	Eldorado x Conserva 611[=Conserva 253 ((=G.Elb. Cling x Aldrighi) op.) x NJC.88]	Embrapa	PYNRC	ADM
BP134	Conserva 1844	Leonense x Bolinha	Embrapa	PYNRC	POP III
BP135	Conserva 1153	Conserva 677 [(=Brilhante x NJC 97) op.] x Granada	Embrapa	PYNRC	POP III
BP136	Carapuça	(Southland x Jewel) op.	Embrapa	PYMRS	POP I
BP137	Cascata 349	(NJ 230 x FLA2631) op.	Embrapa	PYMRS	POP I
BP138	Cascata 700	Cascata 564 x Escarlate	Embrapa	PYMRC	POP I
BP139	Cascata 727	Conserva 327 x Taquari 19	Embrapa	PWNRC	ADM
BP140	Cascata 1005	C92-16 (Chimarrita x Cristal Taq.) op.	Embrapa	PYNRC	ADM
BP141	Cascata 1015	Ametista x A170	Embrapa	PYMRC	ADM
BP142	Cascata 1055	Chinoca x Granada	Embrapa	PWMR?	ADM
BP143	Cascata 1067	BR3 x A333	Embrapa	PWMR?	POP I
BP144	Cascata 1423	Fascínio op.	Embrapa	PWMRS	POP I
BP145	Cascata 1429	Fascínio op.	Embrapa	PWMRS	POP I
BP146	Cascata 1493	(Cascata 253 x A425) op.	Embrapa	PWNRC	ADM
BP147	Cascata 1511	Cascata 972 x Ruipan 2	Embrapa	PWMFC	POP I
BP148	Cascata 1513	Cascata 845 x Chimarrita	Embrapa	PWMRS	POP I
BP149	Cascata 1577	Tropic Snow x Marfim	Embrapa	PWMRS	POP I
BP150	Cascata 1669	Cascata 805 x Aurora 1	Embrapa	PWMRS	POP I
BP151	Chato 10	(Peento MF) op.	Embrapa	PWMFF	ADM
BP152	Chato 11	Capdeboscq x Cascata 69	Embrapa	PYMFC	POP III
BP153	Chato 13	(Aldrighi x Cascata 69) op.	Embrapa	PWMFF	POP III
BP154	Chula	Delicioso x Panamint	Embrapa	PWMRF	POP I
BP155	Conserva 657	(Brilhante x NJC 97) op.	Embrapa	PYNRC	ADM
BP156	Conserva 1824	Conserva 672 x A334	Embrapa	PYNRC	POP III
BP157	Conserva 1127	Maciel x A320	Embrapa	PYNRC	ADM
BP158	Conserva 1218	Conserva 672 x Maciel	Embrapa	PYNRC	POP III

BP159	Conserva 1278	Conserva 1125 op.	Embrapa	PYNRC	ADM
BP160	Conserva 1526	Conserva 672 x A334	Embrapa	PYNRC	ADM
BP161	Conserva 1556	Conserva 672 x A334	Embrapa	PYNRC	ADM
BP162	Conserva 1806	Conserva 1062 x Maciel	Embrapa	PYNRC	POP III
BP163	Conserva 1203	Cerrito x A-249	Embrapa	PYNRC	POP III
BP164	Douradão	Dourado-1 (Tutu x Maravilha) op.	IAC	PYMRF	POP I
BP165	Dourado 2	[Tutu (IAC 1353-1)] x [Maravilha (FLA 13-72)]	IAC	PYMRF	ADM
BP166	Early Diamond	unknown	USA	NYMRC	POP I
BP167	Flordastar	Flordagold x EarliGrande	Florida	PYMRC	POP I
BP168	Granito	Alpes x Conserva 102	Embrapa	PYNRC	POP III
BP169	Ingo	unknown	Taiwan	PWMRF	*
BP170	July Elberta	unknown	Louisiana	PYMRF	POP I
BP171	La Feliciana	Dixigem op.	Louisiana	PYMRF	POP I
BP172	Lotus	(Sunhigh x Redcrest) op.	USA	NYMRF	POP I
BP173	Maravilha	Sunred x 28-48 op (Okinawa x Highland)	Florida	PWMRC	POP I
BP174	Marfim	Coral x pollen from China (Gang Shan Suo Shang)	Embrapa	PWNRC	ADM
BP175	Necta 422	IACN30-74-49 x A334CN	Embrapa	NYNRC	ADM
BP176	Necta 466	(Eldorado x A403CN) op.	Embrapa	NYNRC	ADM
BP177	Necta 508	(Sunred x Rayon) op.	Embrapa	NYMRS	POP I
BP178	Necta 529	Tropic Snow x Marfim	Embrapa	NYMRS	POP I
BP179	Necta 531	Tropic Snow x Marfim	Embrapa	NWMRS	POP I
BP180	Necta 543	Sun Snow x [Necta 420 (Branca x Linda)]	Embrapa	NWMRF	POP I
BP181	Necta 3973	unknown	Mexico	NWMRC	POP I
BP182	Necta Morena	unknown	Spain	NYMRC	POP I
BP183	Nectared 5	NJ53939(Candoka x Flaming Gold) x NJN14 (Nectalate op.)	New Jersey	NYMRC	POP I
BP184	Norman	Sunhigh x Redskin	North Carolina	PYMRF	POP I
BP185	Piazito	(C 79.53.22) op.	Embrapa	PWMRS	ADM
BP186	Rayon	[(Okinawa x Panamint) op.] x Kaygold	Florida	PYMRC	ADM
BP187	Rei Del Monte	unknown	Uruguai	PYMR?	ADM
BP188	c.2009.77.15	Conserva 1510 x Libra	Embrapa	PYN??	ADM
BP189	Cascata 967	Sinuelo x Fla 3-2	Embrapa	P?M??	POP I
BP190	Cascata 1065	Cascata 864 op. [(=Escarlate x A.379)op.]	Embrapa	PYM??	ADM
BP191	Sel. Bolinha 26	Bolinha op.	Embrapa	PYNRC	POP III
BP192	Sprincrest	[(Firglowx Hiley)x Fireglow] x Springtime	Georgia	PYMRC	POP I
BP193	Suncoast	FLA9-12N x FLA7-3N (Sungold x ArmKing)	Florida	NYMRF	POP I
BP194	Sunlite	[Fla 8B-27 (Okinawa x Panamint)] x NJN21	IAC	NYMRS	POP I
BP195	Sunred	Panamint x Fla R9T10	Florida	NYMRC	POP I
BP196	Super morena	unknown	Spain	NYMRS	POP I
BP197	Talismã	Rei da Conserva x Jewel	IAC	PWMRC	ADM
BP198	Taq 98	(B. XVI-16 x Delicioso) op.	Taquari	PWMRC	ADM
BP199	Tsukuba 1	Okinawa x Akame	Japan	P?M??	POP I

BP200	Tx 1A 95	TX1193-1 op.	Texas	PYMRC	ADM
BP201	Tx 1A 100	TXW1192-2 x Earligrande	Texas	PYMR?	POP I
BP202	Tx 1A 125	Tropic Beauty x FLA 84-4	Texas	PYMRC	POP I
BP203	Tx 1A 150	Victor op. (Tropic Beauty x Goldprince)op.	Texas	PYMRC	POP I
BP204	Tx 2A 232 LWN	Sunmist x Arctic Star	Texas	NWMRC	POP I
BP205	Vila Nova	Cristal x Princesa	Embrapa	PYMRF	POP I
BP206	Cascata 1020	Eldorado x Cascata 727 (=Conserva 327 x Taq 19)	Embrapa	PYMRC	ADM
BP207	Conserva 685	Conserva 471 op. (=Alpes x Conserva 102 op.)	Embrapa	PYNRC	POP III
BP208	Maria Bianca	Dew Hale x Michelini	Italy	PWMRF	POP I
BP209	Sensação	Granito op. = (Alpes x Conserva 102) op.	Embrapa	PYNRC	ADM
BP210	Turmalina	Conserva 334 x Conserva 594 (=Capdeboscq x Madrugador)	Embrapa	PYNRC	ADM
BP211	Sel. Bolinha 9	Bolinha op.	Embrapa	PYNRC	POP III
BP212	Sel. Bolinha 17	Bolinha op.	Embrapa	PYNRC	POP III
BP213	Sel. Bolinha 25	Bolinha op.	Embrapa	PYNRC	POP III
BP214	Hu Shou da Tao	unknown	Taiwan	PWMR?	POP I
BP215	Hu Sao	unknown	Taiwan	PWMR?	POP I
BP216	Conserva 1215	Conserva 657 [(=Brilhante x NJC 97) op.] x Conserva 655 (=Conserva 497 x Diamante)	Embrapa	PYNRC	POP III
BP217	c.2009.173.74	(Olimpia x Seleção Bolinha 26) op.	Embrapa	PYNRC	ADM
BP218	c.2009.173.33	(Cascata 805 x Aurora 1) op.	Embrapa	P????	*
BP219	c.2006.198.7	(Jubileu x Hu-so Tao) op.	Embrapa	P????	ADM
BP220	c.2006.45.4	Olimpia x Seleção Bolinha 26	Embrapa	PYNRC	POP III

op = open-pollinated; *IAC* = Agronomic Institute of Campinas-São Paulo/Brazil; Embrapa = Brazilian Agricultural Research Corporation. Taquari = Peach program that started in 1953 at Taquari Experiment Station, in 1958 was incorporated to now EMBRAPA program.

[a] Reported parentage = Pedigree data were obtained from Raseira and Nakasu (1998); Raseira et al (2014); Raseira, personal communication

[b] Origin = Breeding program/Country

[c] Fruit traits are as follows = First letter: *P* peach, *N* nectarine. Second letter: *W* white, *Y* yellow. Third letter: *N* non-melting flesh, *M* melting flesh. Fourth letter: *R* round shape, *F* flat shape. Fifth letter: *C* clingstone, *S* semi-clingstone, *F* freestone

[d] Population stratification defined by fastSTRUCTURE, considering a membership coefficient above 0.75.

? = No information available

* = Peach genotype removed due to GBS low initial read numbers

Table S2 - Rank of individual BLUP scores for field rating, lesion length and lesion area in detached leaf assessments, and classification of genotypes in response to *Xap* infection.

		Field			Lesion length (mm)			Lesion area (mm²)		
ID	**Genotype**	$\hat{\mu}+\hat{p}_p$	**Rank**	**Class**	$\hat{\mu}+\hat{g}_i$	**Rank**	**Class**	$\hat{\mu}+\hat{g}_i$	**Rank**	**Class**
BP001	Abóbora	3.43	19	S	5.34	54	MS	23.27	47	MS
BP002	Ágata	1.97	136	MR	5.49	13	S	24.21	17	S
BP003	Aldrighi	1.97	137	MR	5.13	106	R	21.51	106	R
BP004	Alpes	2.37	100	MR	5.35	48	MS	23.20	51	MS
BP005	Amarillo	1.44	176	R						
BP006	Anita	1.97	138	MR	5.31	65	MR	22.94	63	MR
BP007	Apote	1.70	160	R	5.22	91	MR	22.11	91	MR
BP008	Arlequim	1.84	149	MR	5.59	2	HS	25.15	2	HS
BP009	Atenas	3.03	35	MS	5.44	20	MS	23.84	29	MS
BP010	Aurora 1	3.16	28	MS	5.28	72	MR	22.72	67	MR
BP011	Aurora 2	3.03	36	MS	5.44	22	MS	23.88	25	MS
BP012	Y-babcock	1.57	168	R						
BP013	Barbosa	1.70	161	R	5.20	92	R	22.10	92	MR
BP014	Bolinha	1.70	162	R	5.22	90	MR	22.22	88	MR
BP015	BR1	1.31	180	R	5.34	52	MS	23.51	39	MS
BP016	BR3	3.56	10	S	5.10	108	R	21.37	108	R
BP018	Capdeboscq	1.57	169	R	5.25	81	MR	22.40	82	MR
BP019	Cardeal	3.56	11	S	5.41	34	MS	23.70	35	MS
BP020	Regalo				5.43	26	MS	24.29	13	S
BP024	Fascínio	3.16	29	MS	5.36	45	MS	23.27	46	MS
BP027	Cerrito	2.76	61	MS	5.26	78	MR	22.52	77	MR
BP028	Chimarrita	1.84	150	MR	5.39	39	MS	23.48	40	MS
BP029	Chiripá	1.44	177	R	5.44	25	MS	23.88	24	MS
BP030	Conserva 334	2.63	72	MS						
BP031	Conserva 594	2.23	115	MR	5.43	30	MS	23.82	30	MS
BP032	Conserva 672	1.31	181	R	5.15	102	R	21.66	103	R
BP033	Conserva 930	3.03	37	MS						
BP034	Conserva 947	1.70	163	R	5.31	66	MR	22.71	71	MR
BP037	Conserva 1596	2.63	73	MS						
BP038	Conserva 1600	3.16	30	MS						
BP039	Conserva 1612	3.56	12	S						
BP040	Conserva 1666	2.90	51	MS						
BP041	Conserva 1798	3.29	22	S						
BP042	Convênio	1.97	139	MR	5.37	43	MS	23.46	41	MS
BP043	Coral	2.23	116	MR						
BP044	Coral 2	1.70	164	R	5.23	88	MR	22.22	89	MR
BP045	Cristal-taquari	1.04	184	HR	5.36	46	MS	23.22	50	MS

BP046	Delicioso	1.97	140	MR						
BP047	Della Nona	2.10	126	MR	5.24	83	MR	22.31	84	MR
BP048	Diamante	2.50	81	MS	5.52	8	S	24.76	4	S
BP049	Dulce	2.76	62	MS	5.15	99	R	21.75	98	R
BP050	Edmundo Perret	2.23	117	MR	5.45	19	MS	23.96	20	MS
BP051	Eldorado	3.56	13	S	5.33*	56		23.08*	55	
BP052	Eragil	2.10	127	MR						
BP053	Esmeralda	2.50	82	MS	5.41	33	MS	23.75	32	MS
BP054	Ewtrin	3.03	38	MS						
BP055	Farrapos	2.37	101	MR						
BP056	Flordabella	2.63	74	MS	5.35	50	MS	23.23	49	MS
BP057	Flordaglo	2.37	102	MR	5.31	64	MR	22.82	65	MR
BP058	FlordaGrande	2.76	63	MS	5.63	1	HS	25.66	1	HS
BP059	Flordaking	3.69	7	S	5.53	7	S	24.62	9	S
BP060	Flordaprince	2.90	52	MS	5.55	4	S	24.69	6	S
BP061	Galaxy	1.97	141	MR						
BP062	Gaúcho	1.97	142	MR	5.33*	57		23.08*	56	
BP063	Gaúcho POA	2.23	118	MR	5.38	41	MS	23.42	45	MS
BP064	Gaudério	2.37	103	MR	5.18	95	R	21.91	95	R
BP065	Granada	2.50	83	MS						
BP066	Interlúdio	2.10	128	MR	5.27	74	MR	22.56	75	MR
BP067	Jade	2.37	104	MR	5.43	29	MS	23.78	31	MS
BP068	Josefina	2.63	75	MS	5.48	16	S	24.27	14	S
BP070	Kampai	3.56	14	S	5.45	18	MS	24.06	19	MS
BP071	Leonense	2.37	105	MR	5.44	24	MS	23.93	23	MS
BP072	Libra	3.82	4	S	5.48	17	S	24.26	16	S
BP073	Linda	2.76	64	MS	5.19	93	R	21.99	93	R
BP074	Lord	1.57	170	R	5.23	86	MR	22.22	90	MR
BP075	Maciel	2.37	106	MR	5.30	69	MR	22.71	70	MR
BP076	Madrugador	2.50	84	MS						
BP077	Magno	2.63	76	MS	5.27	75	MR	22.72	69	MR
BP078	Mara	2.50	85	MS						
BP079	Marli	1.84	151	MR						
BP080	Minuano	2.37	107	MR						
BP081	Mollares Hierro	3.56	15	S						
BP082	Morro Redondo	3.69	8	S						
BP083	Natal	1.44	178	R	5.34	55	MS	23.03	57	MR
BP084	Necta 468	2.76	65	MS						
BP085	Necta 480	3.16	31	MS						
BP086	Necta 496	2.23	119	MR						
BP087	Necta 511	2.63	77	MS						

BP088	Necta 512	2.63	78	MS						
BP089	Necta 528	2.50	86	MS						
BP090	Necta 532	2.23	120	MR						
BP091	Okinawa	3.03	39	MS						
BP092	Olímpia	2.10	129	MR	5.44	23	MS	23.93	22	MS
BP093	Ônix	1.84	152	MR	5.29	70	MR	22.60	73	MR
BP094	Pampeano	3.03	40	MS						
BP095	Panamint	1.70	165	R						
BP097	Pérola de Itaquera	2.90	53	MS	5.41	32	MS	23.73	34	MS
BP098	Pilcha	2.23	121	MR	5.44	21	MS	23.95	21	MS
BP099	Piratini	1.84	153	MR	5.42	31	MS	23.87	26	MS
BP101	Precocinho	2.23	122	MR	5.05	111	HR	21.00	111	HR
BP102	Premier	1.84	154	MR	5.41	35	MS	23.75	33	MS
BP103	Princesa	2.50	87	MS	5.50	12	S	24.54	10	S
BP104	Real	3.82	5	S	5.15	101	R	21.68	101	R
BP106	Rubimel	3.29	23	S	5.48	14	S	24.26	15	S
BP107	Safira	3.16	32	MS	5.50	11	S	24.43	12	S
BP108	São Pedro	2.10	130	MR	5.26	79	MR	22.52	78	MR
BP109	Santa Aurea	2.23	123	MR	5.34	51	MS	23.00	59	MR
BP110	Sentinela	1.97	143	MR	5.29	71	MR	22.72	68	MR
BP111	Sinuelo	1.70	166	R	5.14	104	R	21.69	99	R
BP112	Sulina	2.23	124	MR	5.24	82	MR	22.43	81	MR
BP113	Sunblaze	2.37	108	MR	5.26	77	MR	22.43	80	MR
BP114	Sunhigh	1.17	182	R						
BP115	Sunmist	2.50	88	MS	5.53	6	S	24.62	8	S
BP116	Taquari 80	2.76	66	MS	5.38	40	MS	23.43	42	MS
BP117	Tarumã	2.50	89	MS						
BP118	Topázio	2.90	54	MS						
BP119	Tropic Beauty	1.84	155	MR	5.40	36	MS	23.56	37	MS
BP120	Tropic Blush	2.50	90	MS	5.11	107	R	21.41	107	R
BP121	Tropic snow	2.10	131	MR	5.48	15	S	24.21	18	S
BP122	Turquesa	3.43	20	S	5.23	87	MR	22.29	86	MR
BP123	Vanguarda	2.50	91	MS	5.30	67	MR	22.70	72	MR
BP124	Alvorada	2.76	67	MS						
BP125	Âmbar	2.10	132	MR	5.14	103	R	21.63	104	R
BP127	Aztec Gold	2.37	109	MR	5.40	37	MS	23.63	36	MS
BP128	Babygold 7	1.84	156	MR	5.39	38	MS	23.52	38	MS
BP129	Babygold 9	1.57	171	R	5.33	58	MR	23.00	58	MR
BP130	Baronesa	2.50	92	MS	5.19	94	R	21.91	94	R
BP132	Bonão	3.03	41	MS	5.51	10	S	24.49	11	S
BP135	Conserva 1153	2.90	55	MS	5.38	42	MS	23.43	44	MS

BP136	Carapuça	2.50	93	MS	5.07	110	R	21.20	110	R
BP137	Cascata 349	1.97	144	MR						
BP138	Cascata 700	1.97	145	MR						
BP139	Cascata 727	1.84	157	MR						
BP144	Cascata 1423	3.96	3	HS						
BP145	Cascata 1429	3.56	16	S	5.23	89	MR	22.28	87	MR
BP146	Cascata 1493	2.50	94	MS						
BP147	Cascata 1511	1.57	172	R	5.37	44	MS	23.43	43	MS
BP148	Cascata 1513	2.10	133	MR	5.56	3	S	25.10	3	HS
BP149	Cascata 1577	2.50	95	MS						
BP150	Cascata 1669	2.90	56	MS						
BP151	Chato 10	3.29	24	S						
BP152	Chato 11	3.16	33	MS						
BP153	Chato 13	3.29	25	S						
BP154	Chula	1.97	146	MR	5.32	60	MR	22.98	62	MR
BP155	Conserva 657	1.57	173	R	5.32	59	MR	22.99	60	MR
BP157	Conserva 1127	4.09	1	HS						
BP159	Conserva 1278	3.29	26	S	5.32	61	MR	23.08	54	MS
BP160	Conserva 1526	2.50	96	MS						
BP161	Conserva 1556	1.84	158	MR						
BP162	Conserva 1806	2.76	68	MS						
BP164	Douradão	3.03	42	MS						
BP165	Dourado 2	3.56	17	S						
BP166	Early Diamond	1.70	167	R						
BP167	Flordastar	2.90	57	MS	5.43	28	MS	23.86	27	MS
BP168	Granito	2.37	110	MR	5.32	63	MR	22.89	64	MR
BP169	Ingo	4.09	2	HS	5.53	5	S	24.74	5	S
BP170	July Elberta	1.17	183	R	5.27	76	MR	22.55	76	MR
BP171	La Feliciana	1.04	185	HR	5.17	96	R	21.85	96	R
BP172	Lotus	1.57	174	R						
BP173	Maravilha	3.03	43	MS	5.43	27	MS	23.86	28	MS
BP174	Marfim	2.37	111	MR	5.23	85	MR	22.30	85	MR
BP176	Necta 466	2.50	97	MS						
BP177	Necta 508	2.90	58	MS						
BP178	Necta 529	2.37	112	MR						
BP179	Necta 531	2.10	134	MR						
BP180	Necta 543	2.10	135	MR						
BP181	Necta 3973	2.76	69	MS						
BP182	Necta Morena	3.43	21	S						
BP183	Nectared 5	1.97	147	MR	5.17	97	R	21.81	97	R
BP184	Norman	0.91	186	HR	5.35	49	MS	23.26	48	MS

BP185	Piazito	2.23	125	MR						
BP186	Rayon	3.03	44	MS	5.10	109	R	21.27	109	R
BP189	Cascata 967	2.63	79	MS	5.23	84	MR	22.31	83	MR
BP190	Cascata 1065	1.97	148	MR						
BP191	Sel. Bolinha 26	2.90	59	MS	5.36	47	MS	23.19	52	MS
BP192	Sprincrest	1.57	175	R	5.34	53	MS	23.08	53	MS
BP193	Suncoast	3.69	9	S						
BP194	Sunlite	1.84	159	MR						
BP195	Sunred	3.03	45	MS						
BP196	Super morena	3.56	18	S						
BP197	Talismã	3.03	46	MS	5.28	73	MR	22.59	74	MR
BP198	Taq 98	3.16	34	MS	5.26	80	MR	22.45	79	MR
BP199	Tsukuba	3.03	47	MS						
BP200	Tx 1A 95	2.90	60	MS						
BP201	Tx 1A 100	3.03	48	MS						
BP202	Tx 1A 125	3.03	49	MS						
BP203	Tx 1A 150	2.50	98	MS						
BP204	Tx 2A 232 LWN	2.37	113	MR						
BP205	Vila Nova	3.03	50	MS	5.51	9	S	24.63	7	S
BP207	Conserva 685	2.63	80	MS						
BP208	Maria Bianca	1.44	179	R						
BP209	Sensação	2.76	70	MS						
BP210	Turmalina	2.76	71	MS	5.14	105	R	21.56	105	R
BP211	Sel. Bolinha 9	2.37	114	MR	5.16	98	R	21.69	100	R
BP213	Sel. Bolinha 25	2.50	99	MS	5.30	68	MR	22.81	66	MR
BP214	Hu Chou da Thu	3.82	6	S	5.15	100	R	21.66	102	R
BP215	Hu Sao	3.16	27	MS	5.32	62	MR	22.99	61	MR

* Peach cultivars used as checks in the detached leaf assessments.
$\hat{\mu} + \hat{p}_p$: permanent phenotypic values
$\hat{\mu} + \hat{g}_i$: predicted genotypic values free of the G x E interaction
HS highly susceptible, *S* susceptible, *MS* moderately susceptible, *MR* moderately resistant, *R* resistant, *HR* highly resistant.

Table S3 - SNP markers associated with *Xap* resistance in peach germplasm evaluated in the field, based on MLM and GLM association analysis.

Marker*	Chrom.	SNP position (bp)	MLM model		GLM model	
			p-value	Marker R^2 (%)#	*p*-value	Marker R^2 (%)#
S1_46645443	1	46645443	4.97E-06	15.47	4.27E-09	19.80
S1_46187295	1	46187295	7.80E-05	11.05	6.88E-06	12.46
S1_44056662	1	44056662	8.33E-05	10.97	5.55E-07	14.90
S1_45237186	1	45237186	9.16E-05	10.86	3.05E-07	15.46
S1_41971822	1	41971822	1.17E-04	10.94	2.47E-07	16.18
S1_41971811	1	41971811	1.17E-04	10.94	2.47E-07	16.18
S1_46773369	1	46773369	1.25E-04	10.47	1.17E-07	16.36
S6_22698094	6	22698094	1.30E-04	10.43	8.34E-05	9.98
S1_47121392	1	47121392	2.13E-04	9.82	8.85E-07	14.45
S2_30116222	2	30116222	2.16E-04	11.40	5.68E-05	12.12
S1_43394291	1	43394291	2.94E-04	9.43	1.59E-05	11.64
S1_46271580	1	46271580	2.95E-04	10.68	3.86E-06	14.62
S1_199125	1	199125	3.04E-04	9.39	-	-
S1_47642117	1	47642117	3.12E-04	9.86	4.62E-05	10.77
S3_556835	3	556835	3.35E-04	10.83	-	-
S2_17483624	2	17483624	3.37E-04	7.41	-	-
S5_1099578	5	1099578	3.80E-04	12.20	-	-
S1_47440542	1	47440542	3.96E-04	9.50	1.12E-05	12.18
S1_45053077	1	45053077	4.18E-04	9.01	7.68E-06	12.36
S1_45282063	1	45282063	4.23E-04	8.99	3.47E-06	13.13
S1_45129341	1	45129341	4.45E-04	11.13	8.82E-05	12.04
S1_47103863	1	47103863	4.76E-04	9.47	1.87E-06	14.37
S2_20331739	2	20331739	4.79E-04	7.02	-	-
S1_45865835	1	45865835	5.10E-04	6.94	-	-
S1_45865822	1	45865822	5.10E-04	6.94	-	-
S3_6284820	3	6284820	5.39E-04	10.46	-	-
S3_6284813	3	6284813	5.39E-04	10.46	-	-
S3_6284771	3	6284771	5.39E-04	10.46	-	-
S3_1129486	3	1129486	5.47E-04	9.93	-	-
S1_10545345	1	10545345	5.53E-04	6.85	-	-
S8_15134675	8	15134675	5.75E-04	9.85	-	-
S1_47642151	1	47642151	6.04E-04	9.02	-	-
S1_47662090	1	47662090	6.14E-04	9.15	-	-
S3_2286578	3	2286578	6.19E-04	8.53	-	-
S3_2286576	3	2286576	6.19E-04	8.53	-	-
S3_2286574	3	2286574	6.19E-04	8.53	-	-

S7_2081503	7	2081503	6.30E-04	8.70	-	-
S1_47642144	1	47642144	6.40E-04	8.95	-	-
S2_30148164	2	30148164	6.49E-04	8.82	4.22E-06	13.40
S2_30148135	2	30148135	6.49E-04	8.82	4.22E-06	13.40
S2_30148134	2	30148134	6.49E-04	8.82	4.22E-06	13.40
S2_30170988	2	30170988	6.51E-04	9.01	9.51E-06	12.86
S1_179632	1	179632	6.65E-04	8.45	-	-
S1_966785	1	966785	6.69E-04	9.94	-	-
S1_966782	1	966782	6.69E-04	9.94	-	-
S1_43405192	1	43405192	7.09E-04	8.37	9.26E-05	9.88
S1_43405191	1	43405191	7.09E-04	8.37	9.26E-05	9.88
S1_43405190	1	43405190	7.09E-04	8.37	9.26E-05	9.88
S1_43405189	1	43405189	7.09E-04	8.37	9.26E-05	9.88
S1_47386666	1	47386666	8.06E-04	8.22	3.06E-05	10.99
S1_41499593	1	41499593	8.81E-04	8.18	-	-
S1_97027	1	97027	8.87E-04	9.96	-	-
S1_45054778	1	45054778	8.94E-04	10.21	9.10E-07	16.98
S1_45054777	1	45054777	8.94E-04	10.21	9.10E-07	16.98
S7_19439921	7	19439921	9.10E-04	11.05	-	-
S7_19439918	7	19439918	9.10E-04	11.05	-	-
S7_19439917	7	19439917	9.10E-04	11.05	-	-
S7_19439914	7	19439914	9.10E-04	11.05	-	-
S1_41969063	1	41969063	9.46E-04	8.02	1.33E-05	11.82
S8_15915684	8	15915684	9.64E-04	8.46	-	-
S3_21853664	3	21853664	-	-	3.40E-06	13.30
S4_13924048	4	13924048	-	-	5.21E-06	14.24
S8_14818040	8	14818040	-	-	9.07E-06	12.74

* SNP markers were ranked based on significance threshold (smallest *p*-value) detected in MLM analysis and concordant SNPs identified with a *p*-value < 0.0001 in GLM analysis were also presented.

[#] Marker R^2 (%): reports the proportion of phenotypic variation explained by corresponding marker after fitting other model terms (population structure, relatedness).