Association mapping in tetraploid potato

Björn B. D'hoop

Promotoren:

Prof. Dr. Richard G.F. Visser
Hoogleraar in de Plantenveredeling, Wageningen Universiteit
Prof. Dr. Fred A. van Eeuwijk
Hoogleraar in de Toegepaste Statistiek, Wageningen Universiteit

Co-promotor:

Dr. Ir. Herman J. van Eck
Universitair docent, Laboratorium voor Plantenveredeling,
Wageningen Universiteit

Promotiecommissie:

Dr. Ir. Marnik Vuylsteke, Universiteit Gent, België
Prof. Dr. Ir. Johan A.M. van Arendonk, Wageningen Universiteit
Prof. Dr. Rolf F. Hoekstra, Wageningen Universiteit
Dr. Christiane Gebhardt, Max Planck Instituut voor Plantenveredeling, Keulen, Duitsland

Dit onderzoek werd uitgevoerd binnen de onderzoeksschool “Experimental Plant Sciences”.

Association mapping in tetraploid potato

Björn B. D'hoop

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
Prof. Dr. M. J. Kropff,
In het openbaar te verdedigen
Op maandag 2 maart 2009
Des namiddags om vier uur in de Aula.

Association mapping in tetraploid potato

Björn B. D'hoop

PhD Thesis Wageningen University, The Netherlands, 2009
With references – with summaries in English, Dutch and French

ISBN 978-90-8585-333-6

Contents

Chapter 1 ..... **1**

General introduction

Chapter 2 ..... **13**

Association Mapping of Quality Traits in Potato (*Solanum tuberosum* L.)

Chapter 3 ..... **29**

Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections; comparing a designed set of trials with a historical set of trials

Chapter 4 ..... **49**

Population structure and LD in tetraploid potato

Chapter 5 ..... **69**

Marker-trait associations in potato

Chapter 6 ..... **89**

General discussion

Literature cited **99**

Summary **113**

Samenvatting **115**

Résumé **117**

Dankwoord **119**

Curriculum Vitae **123**

List of Publications **124**

Education Certificate **125**

Supplementary Figures and Tables Chapter 5 ..... **126**

Supplementary Figure 1 126
Supplementary Figure 2 130
Supplementary Figure 3 134
Supplementary Figure 4 138
Supplementary Figure 5 142
Supplementary Figure 6 143
Supplementary Table 1 144
Supplementary Table 2 148

Appendix **159**

Chapter 1

General introduction

Björn B. D'hoop
Herman J. van Eck
Fred A. van Eeuwijk
Richard G.F. Visser

Potato breeding

Potato, the third most important food crop of the world on a consumption per capita per year base (FAOSTAT, 04 Oct 2008), originates from a single domestication process in the Peruvian Andes (Spooner et al. 2005). Its expansion as a global food crop started with the Spanish conquistadores who introduced potato as a "botanical" novelty from the "New World" into Europe. According to Salaman (1949) and Glendinning (1983) there is evidence that there have been two separate introductions of potato into Europe in the late 16th century and that most probably their origin lies in the north Andean region (Colombia/Venezuela). These introductions were described as of an irregular shape with disfiguring outgrowths and fairly deep-eyed (Salaman 1926), which contrasts with modern cultivars that are shallow-eyed and uniform of shape. Initially and until the end of the 18th century, potato was considered as a garden peculiarity by botanists, unaware of the increasing cultivation of potato as a food crop in monasteries (Oliemans 1988). By the early 19th century potato established itself as a major European food crop. This development together with emerging plant diseases, such as viruses, nematodes, and of course the devastating late blight disease, responsible for the great Irish Famine of 1847, triggered potato breeding efforts (Glendinning 1983).

Before 1850 mainly selection among seedlings was practiced. Most likely the seedlings were raised from seeds from spontaneous berries, which are the result from unintended crosses. The focal traits at that time were adaptation to day-length, yield, size and shape of tubers. In the course of the 19th century there have been some well documented introductions of germplasm from Latin America to broaden the gene pool which was decimated due to the late blight epidemics. Rough Purple Chili is the most well known example, as it appears as ancestor in the pedigrees of an impressive list of varieties in Europe as well as in the USA (Glendinning 1983; Plaisted and Hoopes 1989; van Berloo et al. 2007). Ortiz (1998) nicely illustrated the contribution of Rough Purple Chili to the USA gene pool. From open-pollinated seedlings the cultivar Garnet Chili was selected. Subsequently Garnet Chili was the parent of Early Rose, which is among the ancestors of Russet Burbank, one of the most widely grown cultivars in the USA. A scan through the potato pedigree database (http://www.plantbreeding.wur.nl/potatopedigree, update of 16 February 2008), going 25 generations deep, learns that Early Rose is present in about 60% of the pedigrees of 7074 accessions. Therefore Rough Purple Chili can be considered as among the few but influential founders of today's potatoes. From 1850 onwards breeders started with deliberate crossing between cultivars. In the 20th century introductions of wild species germplasm through repeated backcrossing took place to improve pathogen resistance. Although modern potato cultivars usually have wild relatives and primitive cultivars in their ancestries, due to selection about 80% of the genes they contain are derived from a small amount of cultivars grown early in the 20th century (Glendinning 1983).

This has consequences for potato research. The contemporary commercial potato cultivars are highly interrelated, both because of their narrow genetic base (Love 1999), but also because of the low number of sexual generations since the start of modern breeding. Furthermore, the level of recombination is limited to little more than one cross-over per chromosome, leaving almost 50 percent of the chromatids without recombination (van Os et al. 2006). Therefore, linkage disequilibrium is expected to be more extended in the potato genome than in other outbreeding crops such as maize. This immediately implies that fine mapping will probably remain an utopia within potato even if association mapping is used as methodology where mapping resolution

can be increased through the use of a set of genetically more varying cultivars as compared to full sib mapping populations that are commonly used in QTL mapping studies.

Suggestions to improve on potato breeding efficiency have been numerous (Bradshaw and Mackay 1994), but none of them have dramatically changed commercial breeding practices up to date, mostly because of the nature and complexity of tetrasomic inheritance. The modest reproduction rate of vegetative propagation via seed tubers and the many years of clonal selection are in sharp contrast to other crops. In greenhouse crops, such as cucumber or tomato for example, up to four selection cycles a year can be reached when both hemispheres are used. Breeding at the diploid level by use of parthenogenetic development of unfertilised eggs in combination with unreduced gametes has been proposed as a practical solution to widen the genetic base of potato and to introgress novel genetic diversity from wild relatives (Hougas and Peloquin 1958). Even distantly related wild relatives, not cross compatible with tetraploid potato, have become accessible by techniques such as protoplast fusion and embryo rescue (Bradshaw and Mackay 1994; Ortiz 1998). Attempts to create true homozygous potato genotypes by repetitive selfing have always failed, because of severe inbreeding depression, causing the loss of fertility and vigour (Douches et al. 1996). Again due to potato's tetrasomic inheritance and highly heterozygous nature, mutation breeding – a practice successfully applied in asexually propagated ornamentals – is equally unlikely to become a straightforward alternative to regular potato breeding practices since mutations will be masked by the other alleles, although some success has been reported for yield and late blight resistance (Kowalski and Cassells 1999), as well as for starch properties (Jacobsen et al. 1989; Muth et al. 2008).

With the advent of molecular markers and linkage maps in potato, marker assisted selection emerged as a useful tool in both plant and animal breeding programmes. Specifically for potato, marker assisted selection can be very useful – whether applied for negative or positive selection – for example to speed up recovery of *tuberosum* related background when introgressing traits from wild relatives, to follow a certain amount of traits simultaneously during early generation selection or to screen the complete breeding programme for relevant traits to aid on parental choice. An overview of achieved results with marker assisted selection within potato has been presented by Barone (2004).

In retrospect, potato breeding is still a recurring process of crossing and selection and all of the improvements mentioned above to enhance the breeding efficiency had very little impact. At most these innovations have contributed to progenitor clone development. For the near future potato breeding will remain an extensive process of clonal phenotyping across environments before candidate varieties can be identified with some level of confidence.

Phenotypic and genetic analysis of multi-environment breeding data

To discover potato clones with wide adaptation, i.e. a good performance across a wide range of agro-ecological conditions, the majority of the potato breeding companies have been testing their candidate cultivars and breeder's clones for a variety of traits in a huge number of trials and thus environments before releasing them on the market. This implies that commercial breeders have very valuable multi-environment datasets for genetic studies.

The statistical framework to analyse multi-environment trials aims at providing estimates for the average performance of specific genotypes across a range of environments on the one hand and performance estimates for specific environments on the other hand. A review on different approaches to statistically analyse multi-environment datasets is presented by van Eeuwijk and colleagues (2005).

The more traditional approach has been to use basic additive models to describe the phenotypic behaviour of different genotypes across environments. Although this approach has the advantage of being statistically well defined and straightforward, the results of the analysis are cumbersome for interpretation. They only provide estimates for genotypic effects and environmental effects that are valid within the range of the underlying trials. Performance predictions are only roughly possible for new environments that were within the quality range of the environments included in the multi-environment trials, and physiological aspects cannot be assessed. The second generation of statistical approaches are of the AMMI type: Additive Main effects and Multiplicative Interaction effects models. In this statistical framework the genotype by environment interaction (GEI), i.e. phenotypic differences between genotypes depend on the environment, enters the previous additive model in the form of a variable number of bi-linear terms that describe the GEI (Gauch 1988). Each bi-linear term consists of genotypic sensitivities to an environmental characterisation. Underlying physiological processes of GEI become accessible with this approach when environmental parameters are regressed on explicit environmental measures (Vargas et al. 1999). It is however only through the application of factorial regression models that GEI can be properly characterised by differential sensitivity to explicit environmental variables. Physiological knowledge can be applied to determine which sets of environmental (co)-variables have to be implemented in the model for GEI (van Eeuwijk et al. 1996; Voltas et al. 2002).

One step further would be to include also genetic covariables that describe differences in genotypic means across environments. In this way quantitative trait loci (QTL) can be mapped (Haley and Knott 1992). With the same philosophy the factorial regression model can be expanded to model QTL by environment interaction. Following this approach, prediction of differential genotypic responses to environmental changes using marker information and environmental covariables is amenable (Boer et al. 2007; Vargas et al. 2006). Through introduction of non-linearity in the genotype dependent response functions for the phenotype across the environments, an essential improvement towards real eco-physiological crop models is realised, because most physiological processes do behave non-linearly. Malosetti and colleagues (2006) have illustrated this approach in a mixed-model framework to senescence in potato.

These kind of advanced statistical models have only to some extent been implemented in practice for diploid potato populations (Malosetti et al. 2006) and for tetraploid potato populations first results of QTL mapping have been reported by Bradshaw et al. (2004, 2008) and Bryan et al. (2004). Association mapping methodology to infer QTL on a variety of traits on a collection of tetraploid potato cultivars using multi-environment datasets is presented in this thesis. However, modelling of the genetic effects is restricted to the genotypic main effects, i.e. the average response across environments.

In addition to the use of sophisticated statistical models to analyse phenotypic data, we need to inspect the data itself. To avoid the 'garbage in – garbage out' principle both a researcher or breeder should define an explicit and standardised procedure to ensure reproducibility of the phenotypic data. For many quality traits the breeders assign values for "general impression" with many implicit assumptions on the nature of the trait.

Before embarking on a description of genotype by environment interactions for multiple environments, breeders and researchers better first look at a detailed analysis of individual trials and some simple summary measures for the responses across multiple environments. Thus, one can estimate variance components for genotypic main effects. Furthermore, one can estimate heritabilities, or repeatabilities, of the traits under study. For our type of data, collections of cultivars, it is more appropriate to talk of repeatability than of heritability, although we will use the terms interchangeably. Traits that can be scored 'objectively' tend to have higher heritabilities than more 'subjective' traits, because of less GEI and smaller error, and maybe because of higher genetic variance. Heritabilities are by no means constants and always should be interpreted with care for any trait. They depend on the range of genotypes and environments that are included, as well as on the amount of replication at various levels of variation. Nevertheless, heritabilities can loosely be used as predictors for the probability to identify QTL. They indicate to what extent a trait is likely to improve through breeding by selection on the phenotype.

Marker technology

Molecular marker techniques are used with the purpose to visualise DNA sequence variation at a specific genetic locus. The methods have been developed in the past two to three decades when DNA sequencing was prohibitively expensive for large scale data production. The first markers made use of DNA sequence variation in the recognition sites of restriction enzymes. The variation in length between the restriction fragments resulted in polymorphic patterns which were interpreted to represent a RFLP marker locus. After the immediate adoption of the PCR machine in molecular genetics another range of techniques were developed using length variation between the primer annealing sites or restriction site polymorphism in the PCR product. Single locus marker techniques were then replaced by multi locus marker techniques of which AFLP has gained wide adoption in molecular genetic studies. In this thesis two contrasting marker techniques have been chosen for specific reasons. AFLP is cost effective and SSRs are rich in their number of different alleles.

AFLP™

Since its first description by Vos and colleagues (1995), the AFLP method has become a well-established and routinely used marker technology in mapping and breeding applications and genetic diversity studies in a wide range of organisms, whether plant or animal, irrespective of their ploidy (Meudt and Clarke 2007). The number of citations of this milestone publication, by now over 4400, is still increasing, witnessing the importance of this marker technique. The AFLP technique is based on selective PCR amplification of restriction fragments resulting from a digestion of genomic DNA. Typically about 100 fragments are amplified per primer combination per reaction. These fragments are absence/presence polymorphisms mainly caused by point mutations in the restriction recognition site or the selective nucleotides added to the AFLP core-primers. Only occasionally insertions or deletions in the sequence between the primers result in a length polymorphism (Vos et al. 1995). AFLP is known for its high

reproducibility, genome-wide coverage, cost-efficiency due to a high multiplex ratio and does not require prior DNA sequence information (Breyne et al. 1997; Karp et al. 1996; Meudt and Clarke 2007; Powell et al. 1996). The technique is easily optimised for organisms with uncommon genome sizes (Han et al. 1999). Due to centromeric suppression of recombination the markers tend to cluster at the centromeric region of chromosomes, but methylation sensitive restriction enzymes such as *Pst*I can counter this effect.

However when polyploids such as potato are considered, AFLP has the disadvantage that dominantly scored marker data (i.e. presence/absence) does not contain as much information to the researcher as compared to co-dominant marker techniques. Notwithstanding this aspect of AFLP, some researchers have investigated the possibility of scoring AFLP in a semi co-dominant way, using band intensities. One of the first attempts has been performed by Castiglioni et al. (1999) on diploid maize using image analysis software to quantify band intensity and convert these values into band zygosity. Through application of mixture models on optical band densities in a gel-based system or fluorescence band intensities in capillary systems, it is possible to obtain co-dominant AFLP marker data, as illustrated by Piepho and Koch (2000) in diploid sugar beet as model and for tomato by Gort et al. (personal communication). However, overlapping regions persist between the distributions of the heterozygous and homozygous genotypes for diploids, hence full reliable genotyping remains an issue in these cases (Jansen et al. 2001), and generally only a certain percentage of fragments are suitable to co-dominant scoring (Meudt and Clarke 2007). Nonetheless, there is no apparent loss in accuracy at stake when trying to exploit quantitative information from AFLP band intensities. In the worst case, the quantitative method will perform equally well as the qualitative, i.e. presence/absence method (Piepho 2001). If co-dominant scoring of AFLP patterns would become robust enough to be performed on a routinely basis, so that cost-efficiency increases to even higher levels and polyploid genomes become fully accessible, this surely would be a development of substantial impact for plant science (Meudt and Clarke 2007).

SSRs

Microsatellites or Simple Sequence Repeats (SSRs), also known as Simple Sequence Length Polymorphisms or Short Tandem Repeats, were reported as a marker system in combination with the PCR reaction for the first time in 1989 by three independent groups (Weber and May 1989; Litt and Luty 1989; Tautz 1989). SSRs are polymorphic loci, present in nuclear and organellar DNA, that consist of repeating units of one to six base pairs in length, surrounded by unique DNA sequences on which specific primers can be designed (Cregan and Quigley 1997). The causal mechanism causing repeat number variation is mainly polymerase slippage during DNA replication or DNA repair (Tautz 1989).

Microsatellite markers have been thoroughly described for the first time and predicted to be useful for genotype identification, mapping and population studies within plants (soybean) by Akkaya et al. (1992) and applied for the first time by Bell and Ecker (1994) for linkage mapping within *Arabidopsis thaliana*. Since then numerous applications, such as functional genomics, association and QTL mapping, diversity studies, genome mapping, comparative mapping and gene tagging, have been reported, as reviewed by Varshney et al. (2005). The advantages of SSRs lie in their reproducibility, multi-allelism, co-dominance, genome coverage, possibility to multiplex (either at the PCR step or at the electrophoresis step) and wide applicability (Gupta and Varshney 2000; Sharopova 2008; Varshney et al. 2005). SSRs also have disadvantages:

for example stutter bands and null alleles (Varshney et al. 2005). Another challenge rises when more specifically one wants to infer phylogenetic relationships based on SSR marker patterns. The multi-allelic character of microsatellites can cause over- or underestimation of true relatedness (Provan et al. 1996a). Albeit, these interpretation hurdles do not outweigh the advantages of working with microsatellites (McGregor et al. 2000). Recent findings suggest an even more important role than previously assumed for SSRs than merely mapping in a breeding context: there is evidence that SSRs are involved in transcriptional and posttranscriptional regulation of gene expression and may even be involved in regulation of recombination (Kashi et al. 1997; Kashi and King 2006).

For potato the first microsatellite markers have been developed by Provan et al. (1996a, 1996b) and Kawchuk et al. (1996). Following this pioneering work, a much wider set of microsatellite markers has been developed. Milbourne and colleagues (1997) developed 112 SSR primer pairs of which 98 appeared polymorphic within a small set of six potato clones. Of these, 65 could be mapped in two independent mapping populations. Ghislain and co-researchers at CIP in Peru tested 70 previously developed SSR loci on a sample of potato cultivars belonging to nine different taxonomic groups for transferability of SSRs between cultivated potato species and suggested a subset of 18 high quality markers as a potato genetic identification kit (Ghislain et al. 1999). In 2005, another set of 61 EST-derived microsatellites was added to the already existing set of about 100 SSRs developed for potato (Feingold et al. 2005). The mapping of microsatellite markers is easily achieved in diploid mapping populations, but their use in tetraploid populations is however not that straightforward (Esselink et al. 2004; Landergott et al. 2006).

QTL mapping

Methodology

Association of segregating molecular markers with phenotypic trait data in full sib populations allows the identification of loci involved in quantitative traits, also called QTL (Quantitative Trait Loci) mapping (Paterson et al. 1988). Basically, QTL analysis identifies the linkage between marker alleles and a quantitative trait. QTL mapping can be performed with any kind of mapping population, whether F2, BC (Back Cross), DH (Doubled Haploid) or RIL (Recombinant Inbred Line) populations (Collard et al. 2005; Kearsey 1998; Kearsey and Farquhar 1998). Reviews on QTL mapping strategies have been presented by Tanksley (1993), Kearsey and Farquhar (1998), Mackay (2001) and Borevitz and Chory (2004).

QTL analysis in Potato

Most of the QTL mapping studies in potato have been performed on statistically better accessible diploid mapping populations. Loci involved in quantitative resistance for late blight (Bradshaw et al. 2006; Collins et al. 1999; Costanzo et al. 2005; Leonards-Schippers et al. 1994; Park et al. 2005; Trognitz et al. 2002; Visker et al. 2003), potato cyst nematodes *Globodera rostochiensis* and *pallida* (Bryan et al. 2002; Kreike et al. 1993; Rouppe van der Voort et al. 1998, 2000), *Verticillium albo-atrum* (Simko et al. 2004b) and insects: trichome mediated resistance (Bonierbale et al. 1994) and thrips resistance (Galvez et al. 2005) have been mapped within diploid mapping populations. Likewise, QTL for agronomic and quality traits like specific gravity and dormancy (Ewing et al. 2004; Freyre and Douches 1994; Freyre et al. 1994), tuberisation (Fernandez-Del-Carmen et al. 2007), leaf senescence (Malosetti et al.

2006), chlorosis (Simko et al. 2008), tuber shape, regularity of tuber shape, eye depth and flesh colour (Sliwka et al. 2008; van Eck et al. 1994a,b), chip colour (Douches and Freyre 1994), cold sweetening (Menendez et al. 2002), enzymatic discolouration (Werij et al. 2007), glycoalkaloid content (Sørensen et al. 2008; Yencho et al. 1998), yield and starch content (Schäfer-Pregl et al. 1998) have been detected with success within a diploid, usually interspecific, genetic background. Some of these QTL have been confirmed within a tetraploid potato genetic background as well, namely resistance QTL to *Verticillium albo-atrum* (Simko et al. 2004b) and resistance QTL to the potato cyst nematode *Globodera pallida* (Bryan et al. 2002).

Only a limited number of QTL studies used tetraploid mapping populations, because of the complexity to construct a molecular marker map. Invariably those studies were based on the tetraploid mapping population 12601ab1 x Stirling, developed at the Scottish Crop Research Institute. QTL for yield, late blight and potato cyst nematode resistance, agronomic and quality traits have been localised (Bradshaw et al. 1998; 2004; 2008). Any further report using a tetraploid population for mapping studies reported only partial linkage groups with traits based on a single gene or major QTL. For example, a major QTL has been located for corky ring spot disease resistance (Khu et al. 2008), the *Rx* gene for potato virus X resistance (Bendahmane et al. 1997), and *R2* involved in late blight resistance (Li et al. 1998).

Association mapping

Association mapping is usually defined as the detection of non random associations between molecular markers on the one hand and phenotypic traits on the other hand in populations of genotypes without a simple genetic structure, which contrasts association mapping with traditional QTL mapping approaches (Gupta et al. 2005; Lander and Schork 1994). Association mapping, like traditional QTL mapping, makes use of linkage disequilibrium (LD), also known as gametic (phase) disequilibrium. The level of linkage disequilibrium will decay with increasing genetic distance, the number of generations from the coalescent and population structure. Various assumptions can be made about the genetic relationships between genotypes included in an association study. The genotypes can be considered to be independent, but also all kinds of relationship, both close and distant are possible (Flint-Garcia et al. 2003; Gupta et al. 2005; Nordborg and Tavare 2002; Tanksley 1993; Zhu et al. 2008). The applicability of the method to arbitrarily structured populations is one of the major characteristics distinguishing association mapping from traditional QTL mapping. An additional advantage is the sampling of more meiotic recombination events that occurred across generations since departure of the coalescent. Therefore a higher mapping resolution can be achieved but a higher marker density is a prerequisite (Flint-Garcia et al. 2003; Jannink and Walsh 2002; Nordborg and Tavare 2002). Furthermore, the results are applicable to a much wider germplasm (Zhu et al. 2008). As such, association mapping can be regarded as an alternative, and complementary approach to traditional linkage mapping because linkage mapping is a useful first step to perform a genome-wide scan for QTL, while association mapping can render a more precise location of individual QTL (Glazier et al. 2002; Gupta et al. 2005; Mackay 2001; Zhu et al. 2008). Alternatively, an association study can deliver a number of interesting QTL that may be confirmed in more traditional mapping studies (van Berloo et al. 2008).

Association mapping was initially developed and applied in human genetics to identify causal mutations for common complex human genetic disorders, such as cystic fibrosis (Kerem et al. 1989) or Alzheimer (Corder et al. 1994). The last few years the association mapping approach

received the attention of plant scientists because it has the potential of identifying the causal polymorphisms underlying complex traits (Flint-Garcia et al. 2003; Gupta et al. 2005). This led to an increase of association studies in plants with *Arabidopsis* (Nordborg et al. 2002), maize (Remington et al. 2001) and rice (Lu et al. 2005) as the pioneering model plant species. Applications of association mapping are numerous. They range from the analysis of genome-wide LD and estimation of LD decay, for example in sugarcane (Raboin et al. 2008) and maize (Stich et al. 2005), to marker-trait associations with a perspective for marker assisted selection for discrete traits such as endosperm colour in maize (Palaisa et al. 2003), as well as quantitative traits, like kernel size and milling quality in wheat (Breseghello and Sorrells 2006a), yield and yield stability in barley (Kraakman et al. 2004) or sugar content and yield in sugar beet (Stich et al. 2008c).

Reports about association mapping studies in other plant species created high expectations about this technique and inspired researchers to exploit its possibilities within tetraploid potato. LD decay has been assessed using single nucleotide polymorphisms within a stretch of DNA of more than 25kb revealing an expected decline of LD below 0.1 at a distance of about 10cM (Simko et al. 2006), while Gebhardt et al. (2004) identified linkage equilibrium at 0.9cM. Malosetti and colleagues (2007) have shown that association mapping is a valid alternative approach to traditional linkage based QTL mapping with the successful mapping of late blight resistance within two sets of tetraploid potato cultivars using resistance loci targeted markers and an advanced mixed modelling approach. Li and co-researchers (2005; 2008) proved that the candidate gene association mapping approach can be equally successful within potato as it was in maize (Wilson et al. 2004). They succeeded to associate natural DNA variation at candidate loci, i.e. QTL that co-localise with a candidate gene whose function is known to be involved with a certain trait, with cold sweetening using a straightforward non-parametric t-test approach (Li et al. 2005) and with chip colour, yield, starch content and starch yield using various regression and mixed models (Li et al. 2008). Previously, late blight and maturity had already been mapped deploying similar t-test based candidate gene association mapping methodology within a set of 600 tetraploid potato cultivars (Gebhardt et al. 2004) and *Verticillium dahliae* resistance QTL within a set of 137 tetraploid cultivars (Simko et al. 2004a).

Objectives and outline of this thesis

This thesis deals with association mapping in potato. Following a methodological proof of principle, a number of steps are described, resulting in a comprehensive association analysis in potato.

Chapter 2: In this pilot experiment, the methods and preliminary results of a regression based association mapping approach on agro-morphological and quality traits are presented. Phenotypic data originated from recent trials performed by five breeding companies and genotypic data resulted from five AFLP primer combinations. Decay of linkage disequilibrium was estimated by calculating the squared correlation between pairs of markers using log-transformed marker band intensities. Marker-trait associations were investigated by fitting single marker regression models for phenotypic traits on marker band intensities with and without correction for population structure. The potential of association mapping in tetraploid potato is hereby illustrated, because existing phenotypic data, a modest number of AFLP markers, and a relatively simple statistical analysis allowed identification of interesting associations.

Chapter 3: Analysis of two differing phenotypic datasets was performed to deliver insight in variance components for the genotypic main effects and the GEI. Furthermore, adjusted genotypic means were calculated. This is a first step towards association mapping in Chapter 5. One dataset was obtained through a field trial at two locations with two replicates in 2006 containing 205 diverse cultivars. The other dataset was contributed by five breeding companies and contained 299 genotypes. The 2006 field trial data were nicely balanced, because the trial was designed in that way. The historical breeding dataset was highly unbalanced, which is typical for data obtained during generations of clonal selection. Both phenotypic datasets were analysed separately within a mixed model framework including terms for genotype by environment interactions. We describe both phenotypic datasets by comparing variance components, heritabilities (=repeatabilities), intra-dataset relationships and inter-dataset relationships. Broader aspects related to phenotypic datasets and their analysis, such as data management issues, methodological issues, pros and cons of different experimental set-ups are discussed as well.

Chapter 4: The second step prior to an association mapping study is to gather information regarding population structure and genetic relatedness, and to estimate the extent of linkage disequilibrium in potato germplasm. Using a collection of 430 diverse tetraploid potato cultivars, genotyped with 41 AFLP primer combinations and 53 microsatellites, we aimed to detect hidden population structure. Two methods were used: a Bayesian approach and a distance-based clustering approach. The study showed a weak level of structure in our material. Groups were detected which complied with criteria such as their intended market niche, as well as groups differing in their year of first registration on a national list. Linkage disequilibrium, using r2 as measure, appeared to decay across linkage groups at about 5cM on average. The results obtained so far are promising for association mapping research in potato. The odds are high that truly linked associations can be detected and the potential mapping resolution will suffice for detection of QTL in an association mapping context.

Chapter 5: A comprehensive genome-wide association mapping study is presented. The adjusted genotypic means from two association mapping panels calculated previously were combined with marker information in two association mapping models. One baseline model without correction for population structure and one more advanced model with correction for population structure and genetic relatedness. Genetic relatedness was estimated from marker information. Interesting QTL could be identified for 19 agro-morphological and quality traits. The observed QTL partly confirm previous studies e.g. for tuber shape and frying colour, but also new QTL have been detected e.g. for after baking darkening and enzymatic browning.

In the final chapter, the general discussion, results of preceding chapters are evaluated and their implications for research as well as breeding are discussed.

Chapter 2

Association Mapping of Quality Traits in Potato (*Solanum tuberosum* L.)

Björn B. D'hoop
M. João Paulo
Rolf A. Mank
Herman J. van Eck
Fred A. van Eeuwijk

Published in Euphytica (2008) 161: 47-60

Abstract

In this paper, we describe the assessment of linkage disequilibrium and its decay in a collection of potato cultivars. In addition, we report on a simple regression based association mapping approach and its results to quality traits in potato. We selected 221 tetraploid potato cultivars and progenitor lines, representing the global diversity in potato, with emphasis on genetic variation for agro-morphological and quality traits. Phenotypic data for these agro-morphological and quality traits were obtained from recent trials performed by five breeding companies. The collection was genotyped with 250 AFLP® markers from five primer combinations. The genetic position of a subset of the markers could be inferred from an ultra dense potato map. Decay of linkage disequilibrium was estimated by calculating the squared correlation between pairs of markers using marker band intensities. Marker-trait associations were investigated by fitting single marker regression models for phenotypic traits on marker band intensities with and without correction for population structure. The paper illustrates the potential of association mapping in tetraploid potato, because existing phenotypic data, a modest number of AFLP markers, and a relatively simple statistical analysis, allowed identifying interesting associations.

Keywords: AFLP, Association Mapping, Linkage Disequilibrium, Potato, Quality

Introduction

Since quality demands of consumers and potato processing industry have become increasingly stringent, breeding efforts need to focus on quality traits in potato. Association mapping, also known as linkage disequilibrium mapping or gametic phase disequilibrium mapping — originally developed to study genetic disorders in humans — is a recently emerging tool in plant breeding research. Association mapping differs from classical Quantitative Trait Locus (QTL) mapping. An important distinction is that no segregating offspring have to be grown and phenotyped, since association mapping can deal with collections of existing cultivars. Another distinction is that parent choice is less of a dilemma, nor a limitation on the genetic diversity. A further difference is that phenotyping efforts may be reduced as existing phenotypic data from national lists, gene banks and breeding companies can be used in addition to or as a replacement of present day trials. Still, designed trials for the evaluation of selected germplasm will provide more reliable results. Lastly, some forms of association mapping, like the one presented in this paper and the one in Malosetti et al. (2007), provide a relatively simple approach to the identification of QTL in tetraploids, comparable to the use of marker regression in segregating populations, but applied to a wider genetic background.

Association mapping has been and remains a popular research tool in human and animal genetics. Human disease genetics was the first area for which association mapping methodology was developed and where successes were achieved (Carlson et al. 2004; Jorde 2000; Lander and Schork 1994). In animal genetics, most concern was about the LD patterns in breeding populations to determine to what extent LD holds and the marker density required to fine map genes, e.g. in cattle (Farnir et al. 2000), pig (Nsengimana et al. 2004) and sheep (McRae et al. 2002). In plant genetics, however, fewer papers have been produced so far on association mapping, except for model plant systems like *Arabidopsis thaliana* (Hagenblad and Nordborg 2002; Nordborg et al. 2002), maize (Ching et al. 2002; Palaisa et al. 2004; Palaisa et al. 2003; Parisseaux and Bernardo 2004; Rafalski and Morgante 2004; Remington et al. 2001; Tenaillon et al. 2001) and rice (Garris et al. 2003; Lu et al. 2005; Semon et al. 2005), where association mapping is gaining importance due to the development of high-throughput marker systems and the availability of genome sequences. Other plant systems have been investigated for LD patterns and associations as well. In barley Kraakman et al. (2004, 2006) performed association analysis for agronomical, resistance related and morphological traits. Conifers were examined using the candidate gene approach to dissect complex traits (Neale and Savolainen 2004). Barnaud et al. (2006) reported on the LD pattern of grapevine. In *Lolium perenne*, Skøt (2005) found associations for heading date. LD patterns were examined deploying SNPs in soybean (Zhu et al. 2003), AFLPs in sugar beet (Kraft et al. 2000) and RFLPs in sugarcane (Jannoo et al. 1999). Kernel size and milling quality were fine mapped applying association mapping in wheat (Breseghello and Sorrells 2006a). A complete overview on association mapping and the status of this approach in plants has been published recently (Gupta et al. 2005). In conclusion, although the amount of publications on association mapping in plant genetics is not as high as it is for animal or human genetics, there is a clear trend visible towards an increasing number of plant species on which association mapping research has been performed with success.

Genetic and QTL studies in potato have predominantly been performed on diploid populations, e.g. leaf blight resistance QTL have been detected in a di-haploid clone (Bradshaw et al. 2006),

in a diploid *S. phureja* x *S. stenotomum* hybrid population (Costanzo et al. 2005; Simko et al. 2006), in diploid *S. vernei* (Sørensen et al. 2006), in diploid *S. microdontum* (Bisognin et al. 2005) and in diploid *S. tuberosum* (Visker et al. 2005). Cold sweetening QTL have been mapped by Menendez et al. (2002) in diploid potato. Recently, tetraploid potato populations have been used as well. Simko and co-researchers mapped *Verticillium* resistance genes in tetraploid populations using various association mapping approaches, such as haplotype association — a whole genome approach — and candidate genes — a local approach (Simko et al. 2004a,b). With a limited number of targeted markers, Bormann et al. (2004) tagged QTL for maturity corrected late blight resistance in tetraploid potato. However, the QTL studies in the tetraploid mapping population 12601ab1 x Stirling are unique in demonstrating the complexity of map construction and QTL analysis (Bradshaw et al. 2004;Bryan et al. 2004).

Association mapping methodology in potato till now is mainly restricted to standard two sample t-tests or Mann-Whitney U tests. Gebhardt et al. (2004) for example, deployed the non-parametric Mann-Whitney U test for an association assay on 415 cultivars, genotyped with five PCR markers closely linked to previously mapped QTL for late blight resistance and plant maturity, with evaluation data for late blight resistance and maturity. Simko et al. (2004a,b) detected significant association between *Verticillium dahliae* resistance and a SSR marker closely linked to a candidate gene locus for *Verticillium* resistance (*StVe1*) in tetraploid potato using a standard two sample t-test. *Verticillium albo-atrum* resistance appeared associated with a specific haplotype of the same candidate gene locus in tetraploid potato, detected using haplotype association with the Wilcoxon two sample test. Recently, Malosetti et al. (2007) presented a general mixed model approach to LD mapping, using potato data for illustration, giving attention to the incorporation of relationships between genotypes, whether induced by population substructure, pedigree or other, and the combination of data arising from multiple environments. Their approach is broadly applicable, without the need for specific software. The approach explained by Malosetti et al. (2007) is very similar to the mixed model approach of Yu et al. (2006). In the latter paper, the mixed methodology was illustrated by an example in maize.

In this paper, we will present preliminary results on the utility of a simple association mapping strategy for tetraploid potato using existing data from five potato breeding companies. With a limited number of markers and relatively simple regression models we detected marker-trait associations for a number of quality traits such as cold sweetening (Li et al. 2005;Menendez et al. 2002), after cooking darkening (ACD) (Friedman 1997;Wang-Pruski and Nowak 2004), enzymatic browning and/or blackspot bruising as catalysed by Polyphenol oxidase (PPO) (Lærke et al. 2002). These first results are very promising considering this was a preliminary experiment performed to sort out association mapping methodology before embarking on a more detailed and thorough follow-up study that is presently undertaken.

Materials and methods

Plant material

With the aim to collect a representative subset of worldwide available potato germplasm, we selected 221 tetraploid potato cultivars and progenitor clones. The selection was based on (1) the acreage for seed potato production in the Netherlands, (2) the country of origin, (3) the year of commercial introduction, (4) the market niche (table, starch and processing), (5) phenotypic diversity for quality traits and (6) availability of the cultivars, especially for older genotypes. The

material was kindly provided by Dutch breeding companies and several gene banks (see acknowledgements). Leaf material was harvested from greenhouse-grown and in-vitro-grown genotypes, was frozen with liquid nitrogen and stored at -80˚C until DNA extraction.

Molecular marker analysis

DNA extraction was according to van der Beek et al. (1992). DNA quality and concentration were visually examined using ethidiumbromide stained 1% agarose gels.

AFLP markers were generated according to Vos et al. (1995) using five well known *Eco*/*Mse* primer combinations: E+AAC/M+CAC, E+AAC/M+CAG, E+AAC/M+CCA, E+ACA/M+CAC, E+AGA/M+CAT. Fragments were separated using a capillary sequencer (MegaBACE 1000, Molecular Dynamics & Amersham, serial number 13757) according to van Eijk et al. (2004), each primer combination being labelled with either FAM, NED or JOE. The ROX channel was used for the MegaBACE^{TM} ET900-Rox size standard from GE Healthcare (Amersham Biosciences).

Pseudo gel images were scored at Keygene N.V. Marker nomenclature was based on primer combination and fragment mobility.

Normalisation of signal intensity variation between capillaries due to DNA loading effects was performed on log-transformed band intensity values using the ANOVA procedure in GenStat, release 8.11 (VSN International Ltd., Oxford, UK).

Position information of AFLP markers was retrieved from the ultra dense potato map, using the parental diploid genotypes SH83-92-488 and RH89-039-16 as internal reference (van Os et al. 2006; http://potatodbase.dpw.wau.nl/UHDdata.html).

Phenotypic data collection

An overview of the agro-morphological and quality traits of this study is shown in **Table 1**.

The agro-morphological traits tuber shape, flesh colour and maturity are scored by breeders on an ordinal scale. Trait values do not imply breeder or consumer preference. The flesh colour shades go from white to orange and can be interpreted as a ranking. Underwater weight was assessed on a continuous scale and was based on a 5 kg sample size.

Quality traits were all scored on an ordinal scale (**Table 1**). Cold sweetening was not measured as such, but through the highly correlated characteristics frying and chipping colour. All tubers were stored at 8˚C before frying analysis. Frying colour was determined after three minutes of frying at 180˚C, chipping colour after frying at 180˚C until no more air bubbles appeared. After baking darkening was determined following two minutes of pre-frying at 160˚C, merely scoring the amount of greyness which has a different causal background than frying colour. After cooking darkening was phenotyped after steam cooking at 100˚C, as the severity of greyness. To determine the bruising score, a sample of 50 tubers representative for the field plot was taken — when available — and shaken on a mechanical shaker. One week later, the tubers were peeled and subdivided into classes of severity of bruising (mild, medium, severe) according to size and frequency of blue spots. To determine the bruising score, tuber amounts per severity class were counted and weighed applying the following formula: (50/number of

tubers used) * ((number of mild + 2*number of medium + 3*number of severe)/3). For enzymatic browning, tubers were grated and exposed to air and at a fixed time point enzymatic browning severity was scored. Cooking type was scored following steam cooking.

Table 1: Overview of phenotypic traits, their data types, scales and the numbers of genotypes for which data were provided by five breeding companies for these 11 phenotypic traits. Total number of genotypes was 221. Absence of information is indicated by -- .

Trait	Type	Scale	Agrico	Averis	CMK	HZPC	Van Rijn
Tuber shape	Ordinal	1 = long, 5 = oval, 9 = round with intermediates	124	6	41	145	20
Flesh Colour	Ordinal	4 = white, 5 = cream, 6 = light yellow, 7 = yellow, 8 = dark yellow, 9 = orange	123	6	41	145	20
Underwater Weight	Continuous	g per 5kg tubers	78	6	41	142	20
Maturity	Ordinal	minimum 1 (extremely early) to 9 (extremely late) with increments of 1	124	6	41	142	20
Cooking Type	Ordinal	1 (A, firm) to 7 (D, very mealy) with increments of 1	110	--	36	141	20
After Cooking Darkening	Ordinal	minimum 3 (very dark) to 8 (very light) with increments of 1	113	--	36	141	20
After Baking Darkening	Ordinal	minimum 3 (very dark) to 8 (very light) with increments of 1	--	--	--	126	--
Frying Colour	Ordinal	minimum 3 (very dark) to 8 (very light) with increments of 1	111	--	16	124	18
Chipping Colour	Ordinal	minimum 3 (very dark) to 8 (very light) with increments of 1	110	--	31	109	16
Blackspot Bruising	Ordinal	minimum 0 (no tubers bruised) to maximum 50 (all tubers heavily bruised) with intermediates	--	6	41	130	19
Enzymatic Browning	Ordinal	minimum 3 (very dark) to 8 (very light) with increments of 1	--	--	--	52	--

Phenotypic data came from field trials performed by five Dutch breeding companies (see acknowledgements). For a number of traits, the trait values for individual genotypes represented summary statistics across years and locations, calculated following company specific procedures that were not accessible for us. The phenotypic dataset was unbalanced because the breeding companies analysed different sets of cultivars and traits (**Table 1**).

Statistical procedures

Linkage disequilibrium (LD) between loci was quantified by the squared correlation coefficient, r^2, between the (logarithm of) band intensities (Flint-Garcia et al. 2003; Remington et al. 2001; Zhao et al. 2005). LD decay was investigated by plotting r^2 versus map distance.

We chose band intensities for the estimation of LD and marker-trait associations and not band presence/absence, because in a specially designed pilot experiment we found band intensity to

be related to allele dosage (unpublished). The main conclusion of the pilot was a clear intensity difference between no allele and at least one allele, but not a very strong trend between intensity and allele dosage in the range of one to four alleles. We have experimented with different functions representing the relation between band intensity and allele dosage, but in the end decided that at the moment of writing the most robust approach was to use (the logarithm of) band intensity as such. The use of band intensity produced results that were very closely the same to results using band presence/absence. In genetic terms this means that we worked with a model in which band presence is fully dominant, i.e. it does not matter whether a band corresponds to an allele dosage of 1, 2, 3 or 4, the effect on the phenotype will be the same for any dosage.

The problem of possible population structure in the potato cultivar collection was approached in two ways. Firstly, as an a posteriori approach, using MEGA 3.1 (Kumar et al. 2004), a neighbour joining dendrogram was calculated based on Euclidian distances of the AFLP band intensities. This dendrogram was subsequently inspected for possible clusters. These clusters would then indicate subpopulations. A contrasting, a priori approach consisted in assuming that the breeding companies that developed the cultivars represented subpopulations. In the association analysis, correction for population structure can then be achieved by inclusion of a factor whose levels are defined by the breeding companies.

A two step approach to association mapping was used. First, Best Linear Unbiased Estimators (BLUEs) were calculated per trait for the genotypes, using GenStat, release 8.11 (VSN International Ltd., Oxford, UK). For each trait an additive analysis of variance model was fitted, response = breeding company + genotype + error, where the breeding company factor referred to the company that provided the data for the particular trait (see **Table 1**). As the breeding companies differed in the set of genotypes they evaluated, the data were unbalanced and therefore the BLUEs for the genotypes were adjusted means.

The traits were all analysed as quantitative traits, assuming the error variation to be normally distributed with constant variance, although the majority of the traits was measured on an ordinal scale. Inspection of diagnostic plots for residuals indicated no strong violations of these assumptions.

After calculation of genotypic BLUEs, markers were screened for association with traits. For each trait and for each of the 149 markers, two models were fitted;

1) response = marker + error
2) response = breeder + marker + error.

The marker information that entered the models consisted of the logarithm of the band intensity, so that the markers were introduced as continuous variables for which a slope parameter was estimated by standard least squares. The breeder term refers to the breeding company that developed the particular cultivar. This term is supposed to correct for possible population structure.

Results

LD decay and population structure

Markers which could be scored in an unambiguous manner were retained. In total 551 AFLP fragments were distinguished on 227 genotypes, including a three-fold repetition of the diploid mapping parents SH83-92-488 and RH89-039-16 (van Os et al. 2006), which were included as internal reference genotypes. From these 551 AFLP markers 149 markers could be assigned to mapped AFLP bands in the ultra dense potato map. We assumed that the position of these 149 markers in the diploid UHD map is essentially not different from the position of these bands in the collection of tetraploid genotypes. Besides position information, the AFLP bands represent alleles on two homologous chromosomes from the female and male mapping parents. Although male and female centiMorgan (cM) positions may differ due to differences in recombination in male and female meiosis, we did not attempt to assign sex-averaged recombination distances between the markers. Markers from the same homology group were tested for LD while ignoring from which homologous chromosome they were derived.

Linkage disequilibrium (LD) decay across the potato genome was calculated deploying r^2, the squared correlation between the (log-) band intensities of AFLP marker loci. A graphical representation, plotting the resulting r^2-values in relation to the UHD map distance in cM between the AFLP markers, was constructed. Linkage disequilibrium between markers mapped on different non-homologous chromosomes is embedded in the same graph by using ∞ cM as symbolic distance between these unlinked markers (**Figure 1**). In this AFLP marker dataset r^2 seems to drop below 0.1 at about 3cM, which is promising for LD mapping in potato and its applications in marker assisted breeding. For the current sample size of 221 and a significance level of 0.01, the critical value for r^2 is 0.03. Applying this value as reference, LD seems to disappear after about 8cM.

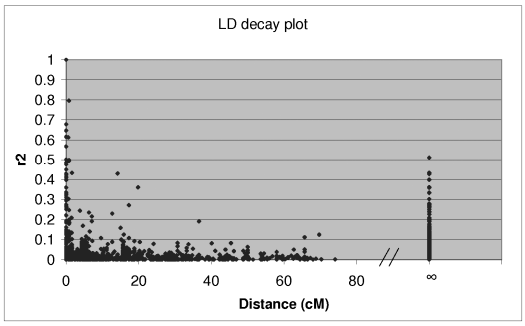


Figure 1: Linkage disequilibrium (LD) decay plot. The squared correlation between paired marker intensities on the y-axis is plotted against the distance between pairs of markers in centiMorgan (cM) on the x-axis. LD between pairs of markers on different non-homologous chromosomes is represented by the symbol ∞. The commonly accepted reference value for r^2 of 0.1, indicates LD to decay at about 3cM. The critical test value for r^2 is 0.03 for a test level of 0.01. Using this critical value as threshold, LD decays at about 8cM.

Population structure of our cultivar and progenitor clone collection was examined, using the neighbour joining facilities of the package MEGA 3.1 for clustering (Kumar et al. 2004) and defining distance between genotypes by the Euclidian distance measure on the basis of the AFLP (log-transformed) band intensities. Although there is a suggestion of some clusters of cultivars sharing country of origin or market niche, the neighbour joining analysis revealed no clear structural relationships, as demonstrated by the very short branches separating clusters from the central point of the astral projection of the dendrogram (**Figure 2**).

Marker-trait associations

Numbers of marker loci associated with traits at a test level of 0.01 are given for the models with and without correction for breeding origin of the material in **Table 2**. The breeder term was significant for all traits. As expected, the number of marker-trait associations typically decreased after inclusion of the correction for breeder. Since all 149 AFLP markers in the genotypic dataset had a known map location, the loci associated with particular traits could be allocated to a potato chromosome. **Table 2** offers an overview per trait of the detected loci together with their chromosome allocations.

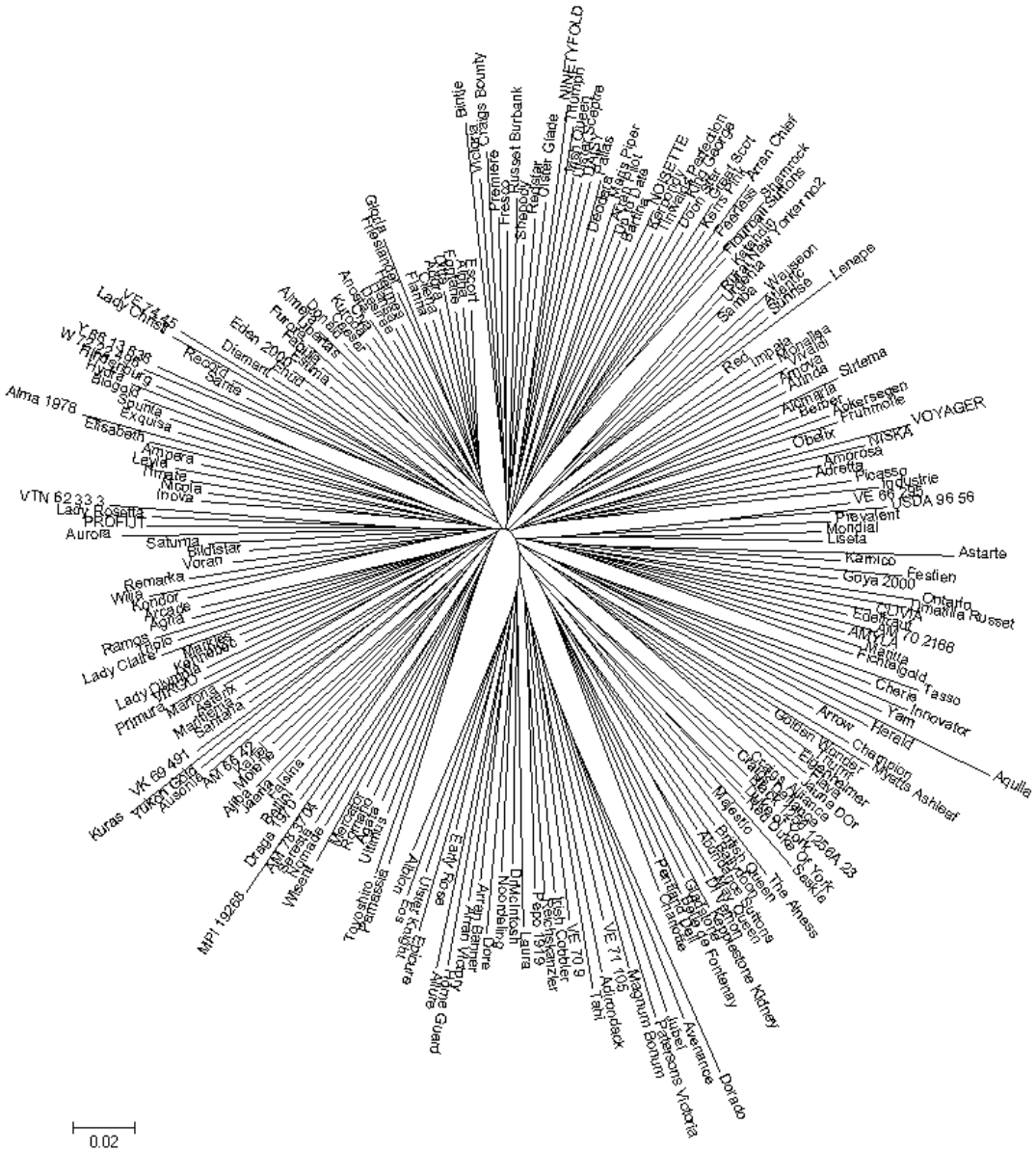


Figure 2: Neighbour joining dendrogram visualised with MEGA3.1 (Kumar et al. 2004). Euclidian distances between band intensities for 149 AFLP markers defined distance between genotypes. The population of genotypes comprised 221 potato cultivars and progenitor clones.

Table 2 also presents a comparison between results described in literature and our association mapping results. In most instances, detected QTL coincide with chromosomes for which associations with the respective trait were previously reported. For cooking type and tuber shape, our QTL chromosomes differed from literature reports. For maturity, after cooking darkening, chipping colour and enzymatic browning more loci were detected than described in

literature. For the other traits like flesh colour, underwater weight and frying colour less loci were detected than reported elsewhere. This paper presents the first QTL reported for after baking darkening and black spot bruising.

Table 2: This table presents on the left side the number of detected marker-trait associations deploying the model without (Simple model) and deploying the model with correction for population structure (Structure model) together with their chromosomal location, based on the UHD map, for the 11 traits studied. On the right side a comparison is presented with associations described in literature. Precise chromosomal locations are presented whenever there is a chromosome match between literature and our association mapping results.

Trait	Number of associated loci with p-value ≤ 0.01		Chromosomal location of loci detected with "Structure model"	Literature		
	Simple model	"Structure model"		Number of loci	Chromosomal location	Reference
Tuber shape	**6**	**5**	**2, 4, 11**	**1**	**10**	van Eck et al. (1994b)
Flesh Colour	**7**	**6**	**1, 2, 3 (31.5cM), 4, 6, 7**	**5**	**3 (Y-locus, 20cM), 10**	van Eck et al. (1994a); Bonierbale et al. (1988); Jacobs et al. (1995)
Under Water Weight	**16**	**6**	**1 (8.4, 9.2 & 24.9cM), 4 (21.3cM), 5 (16.2cM), 9(46cM)**	**--**	**1 (ts(s), 5cM), 4 (ts(i), 7cM), 5 (ts(a), 14cM), 9 (ts(o), 24cM),…**	Gebhardt et al. (2005)
Maturity	**13**	**11**	**1, 4, 5 (16.2cM), 6, 9, 10**	**2**	**3, 5 (GP21, 15cM)**	Gebhardt et al. (2004); Visker et al. (2005)
Cooking Type	**6**	**6**	**1, 2, 6, 10**	**StTLRP candidate gene**	**9**	Kloosterman 2006
After Cooking Darkening	**17**	**7**	**1 (0.8 & 24.9cM), 2, 4 (22.8cM), 5, 7 (54.5cM), 10**	**ACO, CIS, HQT candidate genes,…**	**1 (CIS, 42cM), 7 (ACO, 42cM), …**	Chen et al. (2001); Wang-Pruski and Nowak (2004)
After Baking Darkening	**10**	**10**	**1 (9.2 & 24.9cM), 3, 4 (27.5, 30 & 32.9cM), 6, 7 (55.3cM), 11**	--	--	
Frying Colour	**4**	**3**	**1 (4.6cM), 2 (3.7cM), 7 (63cM)**	**24**	**1(Sug1b, 10cM), 2(Sug2a, 9cM), 7 (Sug7a,c,e, SPS, 60cM),…**	Menendez et al. (2002)
Chipping Colour	**4**	**6**	**1, 2 (3.7cM), 4 (27.5cM), 7, 10 (49.3cM)**	**6**	**2 (I10.1), 4 (H14.1, 35cM), 5, 10 (G13.3, 50cM)**	Douches and Freyre (1994)
Blackspot Bruising	**5**	**2**	**1, 3**	**--**	**--**	
Enzymatic Browning	**6**	**6**	**1, 4, 7, 8 (10.6cM), 11**	**POT candidate genes**	**8 (PPO-III, 25cM)**	Bachem et al. (1994)

Illustration for two marker-trait associations

Choosing a significance level of 0.01, which involves a very modest multiple testing correction, 11 AFLP marker loci were detected for maturity, scattered over six chromosomes of the potato genome, namely *1*, *4*, *5*, *6*, *9* and *10* (**Table 2**). The observed distribution of maturity is depicted in **Figure 3**. The histogram in **Figure 3** illustrates the spectrum of earliness as determined throughout the collection of 221 potato cultivars and progenitor clones. Unfortunately, there was

no maturity data available for all 221 genotypes, only 155 genotypes were represented in the phenotypic dataset.

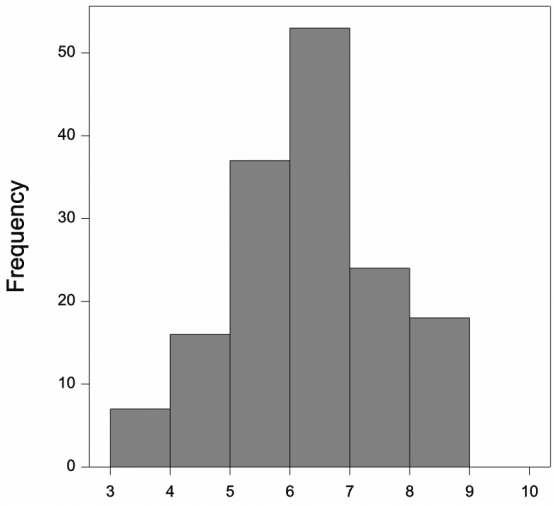


Figure 3: Histogram for maturity. Number of observations was 155.

Because literature has reported several times on a major QTL influencing maturity residing on chromosome *5*, we were interested in the effect of a chromosome *5* AFLP marker associated with maturity in our collection of genotypes. **Figure 4** illustrates the effect of this chromosome *5* marker in the context of the observed data when portrayed by boxplots; marker presence will result in an earlier ripening phenotype. Observed difference between the medians of genotypes with and without band was 0.83 units. The estimate for the effect of the band on maturity in the association model was 0.97 units with an estimated standard error of 0.20. The F-test for association of this marker with maturity had a P-value of 10e-6.

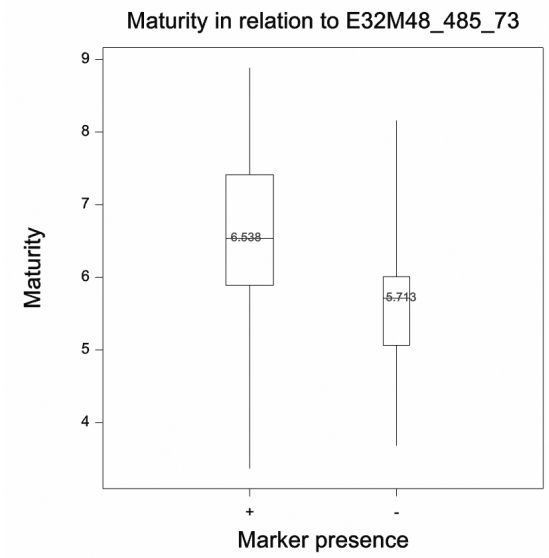


Figure 4: Boxplots for maturity using presence and absence of an associated AFLP marker on chromosome *5* to define the groups. The presence and absence of the specific marker was determined using peak detection on the pseudo gel image. The median values are written in the boxes, the difference in median between genotypes with and without band was 0.83 points on the maturity scale, meaning that a genotype having this marker band will in general be earlier ripening in the field.

Frying colour, an indicator for cold sweetening, appeared associated with three AFLP marker loci at the significance level of 0.01, spread over three chromosomes of the potato genome, namely *1*, *2* and *7* (**Table 2**). The distribution of this trait is depicted in **Figure 5**, where the histogram reflects the spectrum of frying colour as determined throughout the collection of 221 potato cultivars and progenitor clones. Unluckily, there was again no frying colour data available on all 221 genotypes, only 145 genotypes were represented in the phenotypic dataset.

Literature has reported on the location of cold sweetening influencing loci throughout the whole potato genome (Menendez et al. 2002). Therefore we decided to look at the effect of the molecular marker showing the strongest significant association with frying colour in our collection of genotypes. **Figure 6** illustrates the effect of this chromosome *7* marker. As can be deduced from the boxplot, marker presence will result in a slightly better frying quality, more specific a less dark frying colour, meaning a less severe cold sweetening phenotype. Observed difference in median frying colour between genotypes with and without the band at this marker

was 0.38 units. The effect of band presence for this marker was estimated to be 0.37 units with a standard error of 0.13. The F-test for association for this marker had a P-value of 0.004.

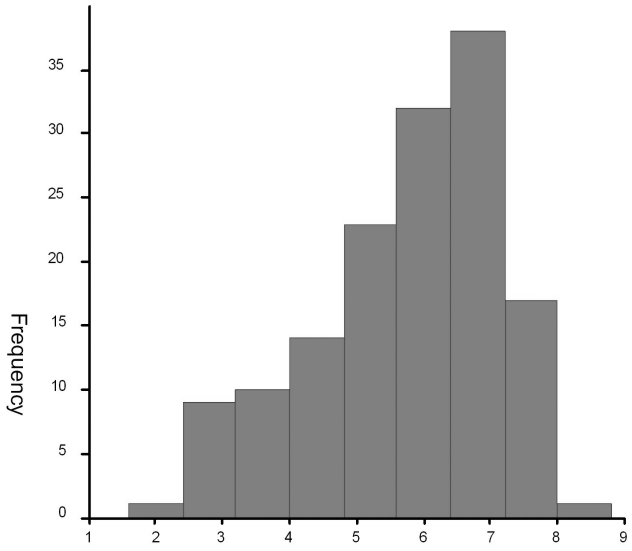


Figure 5: Histogram for frying colour. Number of observations was 145.

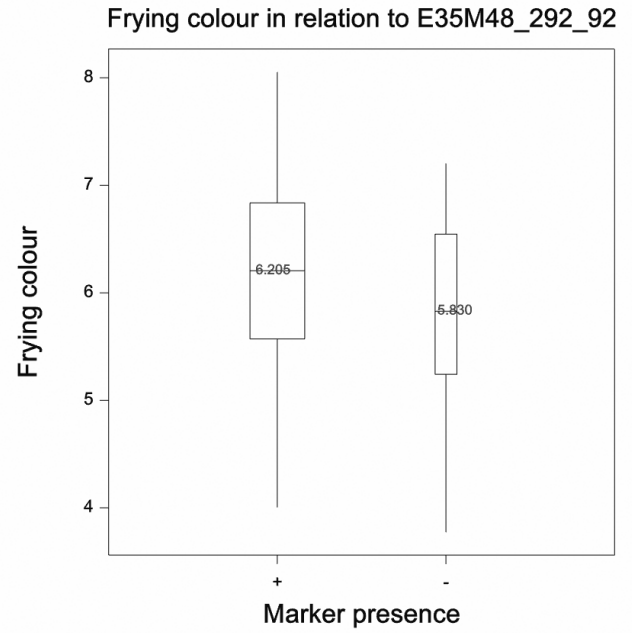


Figure 6: Boxplots for cold sweetening using presence and absence of an associated AFLP marker on chromosome *7* to define the groups. The presence and absence of the specific marker was determined using peak detection on the pseudo gel image. The median values are written in the boxes. The difference in median between genotypes with and without the band was 0.38 points on the frying colour scale, meaning that a genotype having this marker band will in general have a less dark frying colour, i.e. a higher frying quality, than a genotype not expressing this marker band.

Discussion

Association mapping approach

When comparing LD mapping in tetraploid potato with QTL analysis on segregating populations, LD mapping is less cumbersome than standard QTL mapping: it is less time-consuming as no segregating offspring need to be grown and phenotyped and no parent choice needs to be made. For LD mapping, the relationships between marker and trait loci are not merely explained by genetic distance, but also by the number of meiotic generations separating genotypes from each other. On the one hand, this renders the genetics more complex, because population admixture, selection or genetic drift, can bias detected associations. On the other hand, due to recombination events occurring each meiotic generation, stretches of linkage disequilibrium will tend to be somewhat shorter, which can be advantageous to pinpoint genetic markers closer to a trait locus of interest. In LD studies in potato, the role of recombination may be restricted. Gebhardt et al. (2004) showed for a subset of German potato germplasm that there are only

four to five meiotic generations separating heirloom potatoes from their recent descendants. In contrast, our collection contains exceptional cultivars, like for example Biogold being separated by ten meiotic generations from his ancestors (Hutten and van Berloo 2001).

LD decay

We estimated the linkage disequilibrium pattern of the potato genome deploying r^2 between (log-) band intensities, thereby following an approach similar to Remington et al. (2001) and Flint-Garcia et al. (2003). Applying the commonly used reference value of 0.1 for r^2, we found that LD seems to drop at about 3cM. LD decay in potato has already been addressed by Gebhardt et al. (2004), where they examined four DNA markers within 1cM on chromosome *5*. For two markers residing within 0.3 cM LD was maintained, whereas for markers being separated by 0.6 cM LD had decreased and for markers being 0.9 cM apart linkage equilibrium had been reached. Simko et al. (2006) found indications of a fast decaying linkage disequilibrium on short range (r^2= 0.208 at 1kb) but then afterwards a slower decreasing LD (r^2 = 0.137 at 70kb) was found. They concluded from their data that LD decayed below 0.10 at about 10cM genetic distance.

Population structure

We examined population structure with the neighbour joining algorithm using Euclidian distance on the AFLP marker band intensities. This revealed no clear clusters, although some geographically isolated cultivars (USA and UK) and some market niche cultivars (starch processing) tended to group together (**Figure 2**). Cultivars known to be mutants of one another, for example Duke of York (Eersteling) and Red Duke of York (Rode Eersteling) or cultivars known to be closely related to each other, such as Craig's Defiance and Craig's Alliance or Arran Banner and Arran Victory, clustered nicely together. Certainly, the clustering of cultivars could change if more marker data had been included in the cluster analysis, but still, this dataset, although small, sets a trend and more or less confirms what could be expected, namely that other continents and different market niches use to some extent a different gene pool.

Our correction for origin of germplasm in the form of inclusion of breeding company in the association analyses was successful in that this term was judged to be significant in the models for all phenotypic traits. Thus, it made sense to include this correction for population structure in the models for association.

Relations between traits

As illustrated in **Table 2**, some different types of discolouration have loci on the same chromosomes, namely chromosomes *2* and *7*. For the traits chipping colour and frying colour this confirmed our expectations, since both traits are used as indirect indicators of cold sweetening. Moreover, both traits were correlated in our dataset (r = 0.49) and the detected loci on chromosomes *2* and *7*, at 3.7cM and 63.0cM respectively, seemed pleiotropic because they were associated with the same AFLP markers. Contrarily, we did not expect after cooking darkening to share loci with cold sweetening nor with enzymatic browning, since these traits are caused by different biochemical processes. Indeed, no associated markers were found for those traits that showed association with more than one trait, thereby providing more evidence for close linkage of different QTLs than pleiotropic effects of a single QTL. Nevertheless, with the

current precision for marker positions, it is not possible to draw firm conclusions on pleiotropy versus linkage as long as the markers are within 5 to 10 cM from each other.

The traits after cooking darkening and after baking darkening shared loci on chromosomes *1*, *4* and *7* (see **Table 2**). Except for one marker tagging chromosome *1*, the detected loci were not associated with the same AFLP markers. Nevertheless, the markers of these associated loci map closely together on the UHD map which could indicate a common QTL. This confirms our expectations, since both traits are assumed to be caused by the same process (see M&M), the only difference between both traits lies in the temperature to which the samples have been exposed before darkening took place.

Prospects

We used AFLP markers for an association study in potato using a compilation of existing data belonging to different breeding companies. We found in total 69 associated AFLP marker loci for 11 phenotypic traits at the significance level of 0.01, spread throughout the entire potato genome (**Table 2**), showing that even this simply constructed phenotypic historic data set in combination with a modest number of markers of the widely available and easy to use AFLP marker system and a relatively straightforward statistical analysis was able to produce interesting results for applied breeding. Our association approach identified markers that could be useful in the development of marker assisted breeding strategies. These initial results obtained with association mapping should be compared with the power to detect loci involved in quality traits with conventional QTL mapping approaches (both at the diploid and at the tetraploid level). At this moment we regret that in spite of the number of QTL studies on pathogen resistance, the number of QTL studies on quality traits is limited. We anticipate that association mapping will appear a valid method to analyse complex traits in tetraploid potato and will be used along with diploid QTL mapping studies. We can, however, compare the input required by either approach. Our association approach and QTL mapping in diploid mapping populations are both relatively simple in methodology and modestly demanding with respect to genotyping. This is in contrast to the genotyping and methodological demands for QTL mapping using tetraploids (Luo et al. 2006). A similar conclusion was obtained by Malosetti et al. (2007). Although these promising results were based on existing phenotypic data, we will invest in an improved phenotypic dataset. This will be achieved through a field trial with replicates including all cultivars and progenitor clones of the present study. Furthermore, we aim to expand the current marker dataset with additional AFLP markers, but also candidate genes and microsatellites. To further improve on the methodological part of the association approach, research is going on with more advanced mixed models that are able to correct for pedigree relationships and can introduce different marker types into the model. Finally, we aim at obtaining allele dosage observations and estimates for inclusion in our genetic models.

Acknowledgements

We appreciate the involved breeding companies for sharing phenotypic data, namely: Agrico Research (Bant, Netherlands), Averis seeds (Valthermond, Netherlands), C Meijer (Rilland, Netherlands), HZPC Research (Metslawier, Netherlands) and Van Rijn (Emmeloord, Netherlands). We thank Keygene NV and especially, Mariëlle Sengers, for performing the AFLP genotyping work. Furthermore we express our gratitude to the gene banks that were willing to provide us with tubers of old and sparse potato cultivars. It concerns Agriculture and Agri-Food

(Canada), Arche Noah (Austria), IPK Gatersleben (Institut für Pflanzengenetik und Kulturpflanzen-forschung, Germany), INRA (Institut National de la Recherche Agronomique, France), SASA (Scottish Agricultural Science Agency, Scotland), Teagasc (Irish Agriculture and Food Development Authority, Ireland) and USDA (United States Department of Agriculture, USA). Finally, we would like to acknowledge other people who added through different ways to the present study, namely Mariëlle Sengers, Theo Borm, Dirkjan Huigen, Tae-Ho Park, Jeroen Werij, Ronald Hutten, Ralph van Berloo and PPW-Unifarm. Project P8 was (co)financed by the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research.

Chapter 3

Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections; comparing a designed set of trials with a historical set of trials

Björn B. D'hoop
M. João Paulo
Richard G.F. Visser
Herman J. van Eck
Fred A. van Eeuwijk

Abstract

We present analyses on two diverse association panels of tetraploid potato cultivars. One panel of 205 diverse cultivars was evaluated in two trials in 2006, one on sandy soil, the other on clay soil. Data for the second panel, comprising 299 genotypes including 190 recent breeds and a series of standard cultivars, were compiled from contributions from five breeding companies and included 66 locations and 18 years. We calculated variance components for genotypic main effects and genotype-by-environment interactions. Furthermore, we produced estimates for the random genotypic main effects. The genotypic main effects for different traits served to investigate genetic correlations between traits within the two data sets. In addition, for the overlapping genotypes between the two association panels we investigated the genetic correlations between the designed and historical data. The comparison of the general pattern of genetic correlations between the designed and historical data sets and the inspection of the genetic correlations for specific traits for the overlapping genotypes in both association panels provides the basis for an evaluation of the feasibility of the use of company breeding data for association mapping purposes. We will further discuss issues related to data management and definition of traits.

Keywords: genotype-by-environment interaction, mixed models, heritabilities, potato, agro-morphological and quality traits

Introduction

Performance of cultivars across environments can be captured in the form of cultivar main effects. Good estimates of these effects are of great value for compiling national lists of crop species, to produce reliable recommendations. Also for potato such estimates are valuable. However, the production of such estimates requires caution. Potato cultivars are maintained as vegetatively propagated clones. The origin and physiological status of the planting material (seed tubers) may already have a profound impact on the plant (Asiedu et al. 2003). Therefore it is required to plant trial fields with seed tubers which have been propagated, stored and sprouted under identical conditions for at least one propagation cycle. Potato cultivars may serve different purposes and market niches. Since the seventies and eighties of the previous century potato breeding companies have specialised and consequently their gene pools have increasingly diversified. Currently, cultivars are developed for specific markets such as processing industry (pre-baked frozen fries, crisps and flakes), fresh consumption (domestic and export markets), or starch industry, each with their own specific requirements (Kirkman 2007).

Studies on the inheritance of complex traits are currently no longer limited by the number of genetic data points. High multiplex SNP assays or DNA marker techniques have enabled the construction of marker-dense linkage maps (van Os et al. 2006), as well as marker-dense genome-wide association scans (Zhu et al. 2008). Accurate phenotypic information of biological samples has become the limiting step in both linkage and association mapping studies. In plants large F_2, backcross or recombinant inbred mapping populations can be generated easily, which in animal or human genetics is not so straightforward. In human genetics association genetics is habitually used by necessity. In plants association studies have gained popularity, mostly because association mapping allows historical phenotypic data to be used for QTL mapping purposes (Flint-Garcia et al. 2003).

It is generally assumed that association studies using well-known cultivars (both contemporary and historical) have two major advantages as compared to QTL studies. First, in a QTL mapping population the number of segregating alleles is limited to specific alleles of the crossing parents, which is a far from comprehensive reflection of the genetic and phenotypic trait diversity of the entire gene pool. Second, phenotypic data may already be available for these cultivars. In potato, for example, the ECPGR Potato Working Group (a continuation of EU project RESGEN-CT95-34/45) aims to collect phenotypic data (www.europotato.org). Unfortunately, it appeared that these data emphasise on disease resistance, are highly incomplete with respect to quality traits, and highly incomplete with respect to cultivars released in the recent years.

In the framework of a public-private research effort (www.cbsg.nl) two association panels have been composed. One panel comprises 205 contemporary and historical cultivars, the other panel contains 190 recent potato breeds from five participating breeding companies. The contemporary and historical cultivars (hereafter referred to as historical cultivars, in distinction to the recent breeds) were chosen to represent the commercial potato gene pool (excluding native potato land races from Latin America) of the previous century with emphasis on the oldest “founding fathers”, the most widely used in cultivation or widely used as crossing parent (major contributing ancestors, see Love 1999), as well as progenitor clones, representing widely used

donors of disease resistance genes. The group of recent breeds covers mostly advanced clones just before granting Plant Breeders' Rights. Obviously, breeders are hesitant to share this material for centralised phenotypic analysis, but in this project breeders gladly shared all existing multi-year-multi-location data obtained during the selection process so far. For the historical cultivars, it appeared problematic to use historic data for many reasons. The data were highly incomplete, were obtained in periods with outmoded agronomic practices (e.g. obsolete cultivars), evaluated with different methods, or recorded on scales that do not relate to recent practices in varietal assessment. Therefore, it initially appeared necessary to collect new phenotypic values on all contemporary quality traits for all historical potato cultivars included in our study. This data set is hereafter referred to as the "2006 field trial".

Table 1: Characteristic differences between two phenotypic data sets of potato cultivars obtained from a 2006 trial field and multi-year-multi-location trial fields.

Aspects\ Data set	2006	MYML
Material	Historical cultivars (N=205 clones)	Recent breeds (N=190 and a variable number of standards)
Availability	Breeders and gene banks	Not yet available to the market, except for the breeding company
Standards	Belong within the set	Derived from the set of historical cultivars Standards may differ per company per year
Trait evaluations	Comprehensive on those traits that were deemed relevant by the market niche of the breeding company	All kinds of traits, but highly incomplete
Locations	2 (Clay + sandy soil)	Variable (ranging from 1 to 66 locations)
Years	2006 only	Variable (ranging from 2 to 18 years)
Replications	2 replications	Variable (ranging from 3 to 255 records)
Performed by	Wageningen University, in collaboration with breeders	Data from 5 different breeding companies
Protocol	Generalised protocol	Company specific protocols

In this paper we describe, analyse and compare two phenotypic data sets ("MYML" and "2006", see **Table 1**). The MYML data set is based on existing multi-year-multi-location observations of recent breeds at five different breeding companies. The 2006 data set is based on a single year field experiment with historical potato cultivars grown on a river-clay and sandy soil trial field in two replicates. Both data sets are intended to be used for a genome-wide association study, and this paper will describe the approach to generate adjusted mean values per genotype. Furthermore, the strengths and weaknesses of both the MYML data set and the 2006 data set will be analysed. Finally, broader aspects related to phenotypic data sets and their analysis are discussed as well.

Materials & methods

Plant material

Tubers of 205 contemporary and historical potato cultivars (listed in **Appendix 1** of this thesis) have been collected from the participating breeding companies or were kindly provided by gene banks (Agriculture and Agri-Food (Canada), Arche Noah (Austria), IPK Gatersleben (Germany), INRA (France), SASA (Scotland), Teagasc (Ireland) and USDA (USA)). Uniform planting material was generated (Asiedu et al. 2003) by propagation of seed tubers in 2005 to perform the 2006 field trials, with a few exceptions for in-vitro-grown cultivars donated by overseas gene banks.

Field trial

In 2006, the 205 historical potato cultivars were grown on trial fields at two locations near Wageningen, one with sandy soil and one with clay soil. Both locations harboured two replicates. The cultivars were arranged into maturity blocks according to their maturity type (early maturing – main crop – late maturing) to avoid unequal competition as much as possible. The four plant plots were randomised within the maturity blocks, and the three maturity blocks were randomised within each replicate.

Phenotypic observations

Agro-morphological traits were evaluated in a similar way for both data sets. For the 2006 field trial these traits were recorded during the growing season in the field and using the harvested material of these fields. Phenotypic values of quality traits for both the designed 2006 field trial data set and the multi-year-multi-location data set were assessed according to standard protocols of the breeding companies. In this study we focus on the non-enzymatic discolouration due to oxidation of an iron-chlorogenic acid complex (McKenzie et al. 2005) which is observed after baking of French fries (After Baking Darkening or ABD) or after boiling potatoes (After Cooking Darkening ACD), as well as on the Maillard discolouration of French fries due to reducing sugars (Frying Colour FryingCol). The traits ABD and FryingCol were assessed after varying months of storage of harvested material. Another form of discolouration, observed on peeled or grated tubers, is caused by the enzymatic oxidation of phenolic compounds by polyphenol-oxidase (Enzymatic Browning), see Laerke et al. (2002). Tuber flesh colour of potato is usually recorded as an ordinal trait, where values below 5.5 are regarded as white fleshed, whereas yellow fleshed potato are ranked six and higher depending on the intensity of the yellow colour. The white – yellow contrast, controlled by the well-known *Y*-locus on chromosome *3* encoding the gene beta-carotene hydroxylase (Brown et al. 2006), was not regarded as a relevant quality trait. In this study only the intensity of the yellow pigmentation is recorded, which means that a truncated scale was used starting from light yellow six to orange nine, in an attempt to study the genetic control of tuber flesh yellowness, both in raw tuber flesh, as well as cooked tuber tissue (FleshColY and CookColY). Furthermore, we recorded cooking type, maturity type, tuber shape, tuber size and starch content (underwater weight). An overview with descriptions and scales of all the traits measured is presented in **Table 2**.

Data analysis

As both data sets (2006 and MYML) will be used for genome-wide association analysis, it is important to study the variance components for the cultivar main effect and genotype-by-environment interactions, as well as the heritabilities, or better, repeatabilities, to get a first indication for the feasibility of an association mapping approach on this material. Furthermore, we will look at the correlations between traits in both the designed and historical data set. Finally, we will look at correlations between data sets for the genotypes that are common to both association panels. The correlations will be calculated from the estimates for the random genotypic main effects as produced from appropriate mixed models for the individual traits in both data sets. For the analysis, we used the mixed model facilities of GenStat, 10th edition (VSN International Ltd., Oxford, UK).

Table 2: Abbreviated trait names of phenotypic traits along with a short description of the protocol, the data type and the scale.

Trait	Trait description	Description protocol	Type	Scale
ABD_Nov8c	After baking darkening	Shades of gray scored 1h after frying in November after storage at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
ABD_Feb8c	After baking darkening	As above after storage until February at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
ABD_Apr8c	After baking darkening	As above after storage until April at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
ABD_Apr4c	After baking darkening	As above after storage until April at 4˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
ABD_24hNov8c	After baking darkening	Shades of gray scored 24h after frying in November after storage at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
ACD_1h	After cooking darkening	shades of gray scored 1h after cooking	Ordinal	3 (very dark) to 8 (very light) with intermediates
ACD_24h	After cooking darkening	Shades of gray scored 24h after cooking	Ordinal	3 (very dark) to 8 (very light) with intermediates
CookColY	Intensity of yellow flesh in cooked tissue	Scored 1h after 30 minutes of steam cooking	Ordinal	6 light yellow, 7 yellow, 8 dark yellow, 9 orange
Cooking Type	Cooking type	Scored 1h after 30 minutes of steam cooking	Ordinal	1 (A, firm) to 7 (D, very mealy) with intermediates
EnzBrow_30min	Enzymatic browning	Scored 30 minutes after exposure of grated raw tuber tissue to air	Ordinal	3 (very dark) to 8 (very light) with intermediates
FleshColY	Intensity of yellow flesh in fresh tissue	Scored immediately after cutting tubers	Ordinal	6 light yellow, 7 yellow, 8 dark yellow, 9 orange
FryingCol_Nov8c	Maillard discolouration of French fries	Scored 1h after 3 minutes frying at 180 ℃ using IBVL chart in November after storage at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
FryingCol_Feb8c	Maillard discolouration of French fries	As above after storage until February at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
FryingCol_Apr8c	Maillard discolouration of French fries	As above after storage until April at 8˚C	Ordinal	3 (very dark) to 8 (very light) with intermediates
FryingCol_Apr4c	Maillard discolouration of French fries	As above after storage at 4˚C until April	Ordinal	3 (very dark) to 8 (very light) with intermediates
Maturity	Maturity	Visual field observations end of July, half September	Ordinal	3 (extremely early) to 9 (extremely late) with intermediates
Shape	Tuber shape	Visual observation on tubers from 4 plants in ordinal classes	Ordinal	1 long, 3 oval, 5 round with intermediates for 2006 data. The MYML were recorded in 9 classes
Size	Visual impression of the proportion of the tubers with a tuber size exceeding 55mm	Visual observation on tubers from 4 plants in ordinal classes	Ordinal	3 (predominantly undersized tubers) to 9 (predominantly large tubers) with intermediates
Underwater Weight	Dry matter content; in particular starch composition	Measurements on tubers from 4 plants	Continuous	g per 5kg tubers

Designed trials

For the analysis of the 2006 field trials, the following model was used:

Response = Location + Maturity Block + (Location-by-Maturity Block interaction) + (Replicate within Location) + (SecGr within Replicate) + (Green within Replicate) + Genotype + (Genotype-by-Location interaction) + Error

Most terms in the model will be self explanatory. Random terms are underlined. SecGr and Green are two covariates representing secondary growth and tuber greening, respectively. Both disturbances are caused by delayed and excessive rain in the 2006 growing season.

Heritabilities or repeatabilities for cultivar means across locations and replicates were calculated per maturity block by:
$h^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_{gL}/n_L + \sigma^2_e/n_e n_r)$, with σ^2_g the variance component for genotypic main effects, σ^2_{gL} the variance component for genotype-by-location interaction, σ^2_e the error variance, $n_e = 2$ for the number of locations and $n_r = 2$ for the number of replicates. Variance components σ^2_g and σ^2_{gL} were estimated for each maturity class separately. Coefficients of variance (%CV) were calculated according to: $\%CV = 100*\sqrt{(\sigma^2_e)} /mean$.

Furthermore, overall heritabilities were calculated by a weighted average across the three maturity classes with weights proportional to the number of genotypes in the classes. Likewise, overall coefficients of variance were calculated.

Most of the traits were measured on an ordinal scale. Nevertheless, we analyzed these traits without transformation as we found that the diagnostics plots did not show serious violations of the standard assumptions for the application of mixed models. This was true for both the designed trials and the historical trials.

Historical trials

The MYML data set, contributed by five Dutch breeding companies, is composed of phenotypic values of 190 advanced clones (38 per breeding company), supplemented with additional observations on a variable series of standard potato cultivars, leading to a total of 299 genotypes in this data set. The standards are well-known and widely grown cultivars used by each of these breeding companies, and are also included in the 2006 field trial. Obviously, the companies use different sets of standard cultivars, in view of their difference in market niches. Data of the MYML data set were assembled throughout a variable number of successive years of clonal selection and commercial testing of these recent breeds. **Table 3** presents an overview of the structure of the MYML data set. The number of tested genotypes is highly variable, depending on the trait, simply because cultivars for starch industry are habitually not evaluated for cooking and processing traits. Conversely, observations on processing and starch properties (not part of this study) are poorly known for cultivars for fresh consumption. **Table 3** gives also the number of locations, the number of years, and the number of trials in which a particular trait was measured. As illustrated by **Table 3**, the phenotypic observations on agro-morphological traits are abundant, whereas data on processing traits are scarce. Plot sizes were in general 16 plants per genotype.

The model for the trait values in the historical data set was:

Response = Trial + (Repeat within Trial) + Genotype + (Genotype-by-Year interaction) + (Genotype-by-Trial-within-Year interaction) + Error

Trial represents the factorial combination of location and year. (Repeat within Trial) is a kind of replicate within trial effect. We use the term repeat instead of replicate, because replicate, as in the 2006 field trial means fully duplicated for all traits, whereas repeat in case of the MYML data set indicates that a trait has been assessed in a per trait differing number of replicates, depending on the breeding company. Again, random terms are underlined, from which it becomes evident that for these data we chose to use a fully random model. The large number of trials made it statistically attractive to use a purely random model.

Table 3: Structure of the unbalanced MYML data set. For each phenotypic trait, the number of genotypes indicates on how many genotypes the trait has been evaluated, followed by a percentage that illustrates how this number is related to the total number of genotypes in the data set. The number of locations and the respective percentage indicate on how many locations the trait has been evaluated and how this relates to the entire data set. The same reasoning holds for the number of years and the accompanying percentage. The number of trials should be interpreted as in how many combinations of year and location a specific trait has been tested, the percentage shows on how many trials a trait has been assessed as compared to the total number of trials.

Traits	No. genotypes	% genotypes	No. locations	% locations	No. years	% years	No. trials	% trials
ABD_Nov8c	151	50.5	9	10.6	8	44.4	39	11.4
ABD_Feb8c	130	43.5	5	5.9	8	44.4	21	6.1
ABD_Apr8c	113	37.8	3	3.5	3	16.7	9	2.6
ABD_Apr4c	77	25.8	3	3.5	7	38.9	15	4.4
ABD_24hNov8c	29	9.7	14	16.5	2	11.1	16	4.7
ACD_1h	256	85.6	33	38.8	8	44.4	103	30.1
ACD_24h	116	38.8	26	30.6	5	27.8	54	15.8
CookColY	151	50.5	5	5.9	8	44.4	37	10.8
Cooking Type	258	86.3	33	38.8	8	44.4	103	30.1
EnzBrow_30min	98	32.8	1	1.2	3	16.7	3	0.9
FleshColY	267	89.3	50	58.8	18	100.0	219	64.0
FryingCol_Nov8c	180	60.2	25	29.4	8	44.4	66	19.3
FryingCol_Feb8c	161	53.8	8	9.4	8	44.4	34	9.9
FryingCol_Apr8c	133	44.5	5	5.9	5	27.8	19	5.6
FryingCol_Apr4c	108	36.1	6	7.1	7	38.9	29	8.5
Maturity	275	92.0	21	24.7	15	83.3	117	34.2
Shape	295	98.7	66	77.6	18	100.0	249	72.8
Size	252	84.3	64	75.3	8	44.4	173	50.6
Underwater Weight	292	97.7	58	68.2	18	100.0	255	74.6
Total data set	299	100.0	85	100.0	18	100.0	342	100.0

Heritabilities or repeatabilities were calculated as:
$h^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_{gy} + \sigma^2_{gty} + \sigma^2_e)$, with σ^2_g the variance component for the genotypic main effects, σ^2_{gy} the variance component for the genotype-by-year interaction, σ^2_{gty} the variance component for the genotype-by-trial-within-year interaction, σ^2_e the error variance. As it was not possible to retrieve the exact numbers of locations, years and trials we did not divide the variance components by these numbers. Thus, we calculated heritabilities that are underestimating the real values.

However, to be able to compare heritabilities between both the 2006 field trial and the MYML data set, we calculated another set of heritabilities according to:
$h^2_{cp} = \sigma^2_g / (\sigma^2_g + \sigma^2_{gty}/2 + \sigma^2_e/4)$, where the variance component for genotype-by-year interaction was left out of the formula as this component could not be tested with the designed data set. Analogous to the heritabilities calculation for the 2006 data set we divided the genotype-by-trial-within-year variance component by two and the error component by four. Also for this data set we calculated coefficients of variance (%CV) according to: $\%CV = 100*\sqrt{(\sigma^2_e)}$ /mean.

Comparative analysis of the designed and historical data

The genotypic main effects from the above mixed models were used to inspect correlation structures within the designed and historical data set. In addition, for individual traits we inspected the correlation between the designed and historical data set. As we calculated the

correlations based on the estimates for the random genotypic main effects, we are effectively investigating the structure of the genetic correlations within and between the two data sets. To visualize the within-set genetic correlations we produced biplots based on a principal component analysis of standardized genetic effects. For the biplots we used the software SC-Biplot (Cosinus Computing BV, Waalwijk, NL).

Results

More than fifteen thousand data points were collected during the 2006 field trial of 205 historical potato genotypes on two locations with two replicates for 19 phenotypic traits (205*2*2*19=15,580). The MYML data set is even larger (299*85*18=457,470 data points) but has a highly unbalanced data structure due to many incomplete trait observations.

Estimates for variance components and their standard errors for the designed trials are shown in **Table 4**. A similar overview for the MYML data set is shown in **Table 5**. Comparable heritabilities are plotted in **Figure 1**.

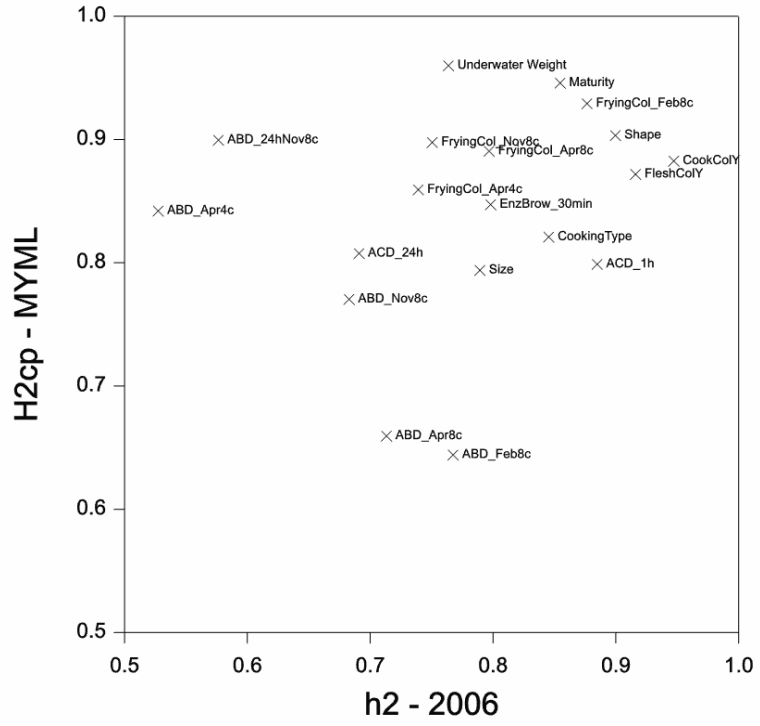


Figure 1: Comparison of 2006 and MYML heritability estimates for 19 phenotypic traits. For the MYML data set the comparably scaled h^2_{cp} estimates were used.

In general high heritabilities have been obtained, which is not abnormal for highly variable traits of clonally propagated crops. The high heritabilities (both for 2006 and MYML) for tuber shape and maturity confirm that these traits are very stable across locations and years. The high heritability for underwater weight can be interpreted as the absence of strong genotype-by-environment (GxE) interactions. The same interpretation holds for frying colour, as apparently for these traits, there was no severe GxE interaction observed, as illustrated by their low GxE variance component in comparison with their genetic variance component (**Tables 4+5**).

Notwithstanding the truncated scales of flesh colour and cooking colour, now reflecting yellow colour intensity only, the heritability of these traits is still high. Tuber size in the MYML data set has a low genetic variance component which is also reflected in the lower heritability for this trait in that data set and is probably the result of variation introduced by the size and physiological age of the seed tuber, as these factors are known to have a profound effect on size grading variability within a genotype. The genetic variance components for discolouration phenotypes and cooking type are low when compared to the respective error variance components for the MYML data set. The same observations hold for these traits for the 2006 data set and for ABD the GxE variance components are high as well. This resulted in lower heritabilities for discolouration traits and cooking type and is partly because the trait values describe composite

effects. ABD observations are influenced by the dark, cold sweetening induced frying colours. Cooking type is also a composite trait, much influenced by starch content.

Table 4: Results of the phenotypic analysis of the 2006 field trial data set. Vg represents the genetic variance component per maturity class, likewise, VgL contains the estimates for the variance component for genotype-by-location interaction. %CV tabulates the coefficient of variation and % present is indicative of the percentage of genotypes for which data could be obtained.

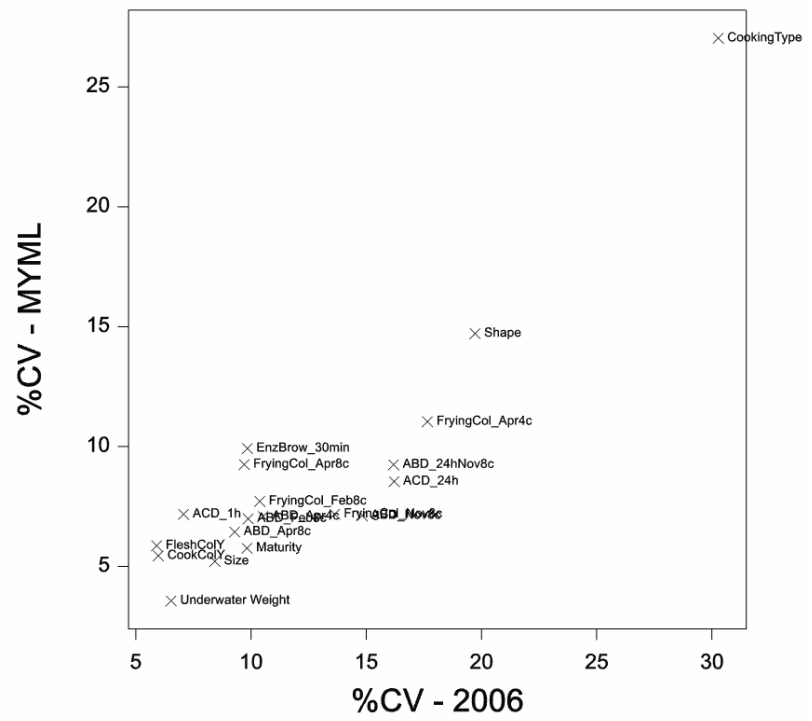
Traits	Maturity	Vg (± s.e.)	VgL (± s.e.)	Error	h^2	Predicted means	%CV	% present
Shape	Early	1.75 (± 0.48)	0.06 (± 0.14)	0.32 (± 0.02)	0.90	2.61	19.73	100
	Mid	2.32 (± 0.40)	0.38 (± 0.14)			2.88		
	Late	1.43 (± 0.44)	0.25 (± 0.20)			3.27		
Size	Early	0.61 (± 0.23)	0.08 (± 0.15)	0.31 (± 0.02)	0.79	6.43	8.43	100
	Mid	1.04 (± 0.25)	0.67 (± 0.19)			6.69		
	Late	1.70 (± 0.48)	0.04 (± 0.15)			6.48		
Maturity	Early	0.26 (± 0.13)	0	0.39 (± 0.03)	0.85	8.04	9.83	100
	Mid	1.83 (± 0.33)	0.30 (± 0.13)			6.76		
	Late	4.17 (± 1.09)	0.57 (± 0.28)			4.69		
Underwater Weight	Early	479.17 (± 179.30)	231.72 (± 123.20)	504.50 (± 38.40)	0.76	324	6.53	100
	Mid	1206.76 (± 223.49)	431.30 (± 106.95)			341		
	Late	2716.23 (± 786.01)	948.96 (± 319.85)			377		
Cooking Type	Early	0.51 (± 0.20)	0.02 (± 0.13)	0.42 (± 0.03)	0.85	1.90	30.27	100
	Mid	0.68 (± 0.14)	0.05 (± 0.09)			2.11		
	Late	1.60 (± 0.49)	0.23 (± 0.20)			2.48		
ACD_1h	Early	0.26 (± 0.13)	0	0.22 (± 0.02)	0.88	6.90	7.08	100
	Mid	0.75 (± 0.15)	0.06 (± 0.09)			6.63		
	Late	1.19 (± 0.37)	0.13 (± 0.17)			6.22		
ACD_24h	Early	0.17 (± 0.11)	0	0.85 (± 0.06)	0.69	6.39	16.21	100
	Mid	0.90 (± 0.19)	0.16 (± 0.10)			5.72		
	Late	0.67 (± 0.23)	0			5.09		
ABD_Nov8c	Early	0.98 (± 0.33)	0.23 (± 0.18)	0.86 (± 0.06)	0.68	6.18	14.82	100
	Mid	1.02 (± 0.19)	0			6.20		
	Late	0.10 (± 0.20)	0.34 (± 0.24)			6.43		
ABD_24hNov8c	Early	0.89 (± 0.36)	0.55 (± 0.26)	0.88 (± 0.07)	0.58	5.93	16.19	100
	Mid	1.11 (± 0.22)	0.27 (± 0.12)			5.77		
	Late	0	0.84 (± 0.25)			5.70		
ABD_Feb8c	Early	0.90 (± 0.40)	0.26 (± 0.29)	0.41 (± 0.04)	0.77	6.25	9.88	80
	Mid	0.76 (± 0.20)	0.35 (± 0.15)			6.51		
	Late	0.69 (± 0.26)	0.08 (± 0.17)			6.56		
ABD_Apr8c	Early	1.04 (± 0.42)	0	0.38 (± 0.04)	0.71	6.45	9.29	69
	Mid	0.71 (± 0.21)	0.37 (± 0.18)			6.62		
	Late	0.35 (± 0.29)	0.54 (± 0.32)			6.65		
ABD_Apr4c	Early	0.36 (± 0.28)	0	0.49 (± 0.06)	0.53	6.51	10.53	51
	Mid	0.29 (± 0.14)	0.09 (± 0.15)			6.64		
	Late	0	0.57 (± 0.26)			6.70		
CookColY	Early	1.32 (± 0.41)	0.11 (± 0.16)	0.16 (± 0.01)	0.95	6.88	5.98	86
	Mid	1.47 (± 0.29)	0.09 (± 0.10)			6.77		
	Late	1.23 (± 0.41)	0			6.53		
FleshColY	Early	1.03 (± 0.33)	0.07 (± 0.15)	0.15 (± 0.01)	0.92	6.51	5.91	77
	Mid	0.99 (± 0.22)	0.16 (± 0.12)			6.61		
	Late	0.83 (± 0.35)	0			6.49		
FryingCol_Nov8c	Early	1.07 (± 0.39)	0.48 (± 0.25)	0.67 (± 0.05)	0.75	5.97	13.63	100
	Mid	1.08 (± 0.21)	0.12 (± 0.10)			5.98		
	Late	0.35 (± 0.22)	0.22 (± 0.20)			6.09		
FryingCol_Feb8c	Early	2.84 (± 0.74)	0.20 (± 0.18)	0.38 (± 0.03)	0.88	5.23	10.38	99
	Mid	2.79 (± 0.46)	0.24 (± 0.12)			5.99		
	Late	1.04 (± 0.46)	0.82 (± 0.35)			6.58		
FryingCol_Apr8c	Early	2.22 (± 0.89)	1.00 (± 0.55)	0.34 (± 0.03)	0.80	5.63	9.71	79
	Mid	1.36 (± 0.30)	0.40 (± 0.17)			6.03		
	Late	0.98 (± 0.45)	0.61 (± 0.33)			6.39		
FryingCol_Apr4c	Early	0.91 (± 0.34)	0	0.89 (± 0.07)	0.74	4.77	17.66	79
	Mid	0.82 (± 0.18)	0			5.42		
	Late	0.59 (± 0.33)	0.56 (± 0.29)			5.79		
EnzBrow_30min	Early	0.80 (± 0.33)	0.40 (± 0.23)	0.39 (± 0.03)	0.80	6.57	9.84	100
	Mid	0.77 (± 0.15)	0			6.30		
	Late	0.32 (± 0.19)	0.21 (± 0.19)			6.10		

Table 5: Overview of the estimates of variance components, accompanied by their respective standard errors for the MYML data set: Vg represents the genetic variance component, Vgy the variance component for genotype-by-year interaction and Vgty the variance component for genotype-by-trial-within-year interaction. H^2_{cp} is the heritability estimate calculated for comparison reasons. %CV presents the coefficient of variation, % present indicates on how many genotypes data could be obtained.

Traits	Vg (± s.e.)	Vgy (± s.e.)	Vgty (± s.e.)	Error (± s.e.)	h^2	h^2_{cp}	Predicted Mean (± s.e.)	%CV	No. of years	No. of locations	% present
Shape	2.55 (± 0.22)	0.07 (± 0.01)	0.26 (± 0.02)	0.57 (± 0.01)	0.74	0.90	5.15 (± 0.02)	14.72	18	66	73.2
Size	0.23 (± 0.02)	0.01 (± 0.00)	0.04 (± 0.00)	0.16 (± 0.00)	0.52	0.79	7.76 (± 0.01)	5.22	8	64	58.6
Maturity	1.41 (± 0.13)	0.02 (± 0.00)	0.09 (± 0.01)	0.14 (± 0.00)	0.85	0.95	6.38 (± 0.02)	5.76	15	21	43
Underwater Weight	2837.00 (± 244.60)	77.50 (± 7.60)	132.20 (± 8.20)	209.40 (± 6.00)	0.87	0.96	405.49 (± 0.63)	3.57	18	58	61.8
Cooking Type	0.99 (± 0.10)	0.03 (± 0.01)	0.11 (± 0.01)	0.65 (± 0.02)	0.56	0.82	2.99 (± 0.02)	27.03	8	33	50.4
ACD_1h	0.35 (± 0.04)	0.03 (± 0.01)	0.06 (± 0.01)	0.23 (± 0.01)	0.52	0.80	6.67 (± 0.01)	7.18	8	33	49.4
ACD_24h	0.43 (± 0.08)	0.05 (± 0.02)	0.05 (± 0.02)	0.31 (± 0.04)	0.51	0.81	6.55 (± 0.02)	8.54	5	26	8.7
ABD_Nov8c	0.29 (± 0.04)	0.03 (± 0.01)	0.06 (± 0.01)	0.24 (± 0.01)	0.48	0.77	6.88 (± 0.02)	7.12	8	9	18.5
ABD_Feb8c	0.13 (± 0.02)	0.01 (± 0.01)	0.03 (± 0.00)	0.23 (± 0.01)	0.33	0.64	6.84 (± 0.02)	6.99	8	5	12.4
ABD_Apr8c	0.16 (± 0.03)	0	0.07 (± 0.02)	0.19 (± 0.02)	0.38	0.66	6.82 (± 0.03)	6.44	3	3	5.1
ABD_Apr4c	0.31 (± 0.07)	0.16 (± 0.04)	0.01 (± 0.03)	0.22 (± 0.03)	0.45	0.84	6.54 (± 0.04)	7.09	7	3	3.4
ABD_24hNov8c	0.96 (± 0.232)	0.04 (± 0.06)	0.01 (± 0.06)	0.42 (± 0.08)	0.68	0.90	6.99 (± 0.08)	9.24	2	14	1.4
CookColY	0.26 (± 0.03)	0.01 (± 0.00)	0.01 (± 0.00)	0.13 (± 0.00)	0.65	0.88	6.57 (± 0.01)	5.45	8	5	29.6
FleshColY	0.29 (± 0.03)	0.01 (± 0.00)	0.01 (± 0.00)	0.15 (± 0.00)	0.63	0.87	6.52 (± 0.01)	5.86	18	50	50.1
FryingCol_Nov8c	0.92 (± 0.11)	0.03 (± 0.01)	0.10 (± 0.01)	0.22 (± 0.01)	0.72	0.90	6.59 (± 0.02)	7.18	8	25	23.4
FryingCol_Feb8c	1.23 (± 0.15)	0.04 (± 0.01)	0.08 (± 0.01)	0.22 (± 0.01)	0.78	0.93	6.13 (± 0.03)	7.72	8	8	14.7
FryingCol_Apr8c	1.07 (± 0.15)	0.05 (± 0.02)	0.11 (± 0.03)	0.31 (± 0.02)	0.7	0.89	5.97 (± 0.04)	9.25	5	5	7.8
FryingCol_Apr4c	1.13 (± 0.18)	0.00 (± 0.02)	0.21 (± 0.04)	0.32 (± 0.03)	0.68	0.86	5.09 (± 0.04)	11.04	7	6	6.5
EnzBrow_30min[a]	0.59 (± 0.11)			0.43 (± 0.05)	0.58	0.85	6.59 (± 0.07)	9.92	3	1	1.9

[a] For enzymatic browning a more simple model could be used for analysis due to the trait being analyzed on one location only.

For the 2006 field trial the heritabilities range from as low as 0.53 for ABD_Apr4c to as high as 0.95 for CookColY. For the MYML data set the underestimated heritabilities range from 0.33 for ABD_Feb8c to 0.87 for underwater weight. However, when we compare the 2006 heritabilities with the comparably scaled MYML heritabilities there is for the majority of traits a clear trend for the MYML heritabilities to be higher (see also **Figure 1**).



The coefficients of variation of both data sets are plotted in **Figure 2**. It can be deduced from **Figure 2** that there is consistency for similar traits between both data sets. Traits that have a higher coefficient of variance in one data set also show a higher coefficient of variance in the other data set, likewise for the traits with a lower %CV.

Figure 2: Comparison of the coefficients of variation between the 2006 and MYML data set as calculated for 19 agro-morphological and quality traits.

Figure 3 gives a biplot showing the traits only following a principal components analysis on standardized genotypic effects (Gabriel 1971, 1978). **Figure 3** thus gives an impression of the genetic correlations within both data sets, with the designed 2006 field trial experiment in **Figure 3a** (46% of the original variation) and the MYML data set in **3b** (40% of the original variation). Acute angles between trait directions represent positively correlated traits, obtuse angles represent negative correlations and orthogonal directions represent independent traits. The quality of representation of individual traits can be read off from the distance from the origin. For traits close to the origin, one should not try to reconstruct the original correlation from the biplot.

In **Figure 3b** FryingCol_Apr8c, FryingCol_Nov8c, FryingCol_Feb8c and FryingCol_Apr4c are highly correlated. The same holds for ABD_Nov8c, ABD_Feb8c, ABD_Apr8c and ABD_Apr4c. These are both expected clusters of related traits as they only differ in their time point of evaluation or in their storage temperature before evaluation. The protocols to assess the traits are exactly the same. Also the relationships between cooking colour and flesh colour, cooking type and underwater weight were predictable. The first two measure the same phenotype with just one processing step in between. The correlation between the latter two can be explained since cooking type is influenced by dry matter content: tubers with higher dry matter content produce a looser cooking type. The strong negative correlation between maturity and underwater weight (and indirectly also cooking type) can be explained by the development of a higher dry matter content with later maturing genotypes. However, there also seems to be a relationship between ACD_24h, cooking colour and flesh colour, which is unexpected. A possible cause for this artefact could lie in the difficulty for breeders to score after cooking darkening objectively after a 24h period in naturally darker coloured potato genotypes.

In **Figure 3a** the same cluster of correlated frying colour traits persisted, although the correlations are less pronounced: the angles are less sharp. The cluster of after baking darkening traits became less obvious as well, only ABD_Apr8c and ABD_Feb8c appear highly correlated, whereas ABD_Nov8c and ABD_Apr4c shifted away from the cluster. The

relationship between ACD_1h and ACD_24h became more apparent in this single-year data set whereas the relationships between cooking colour and flesh colour, and between cooking type and underwater weight became less obvious. The negative correlation between maturity and underwater weight persisted also in this data set.

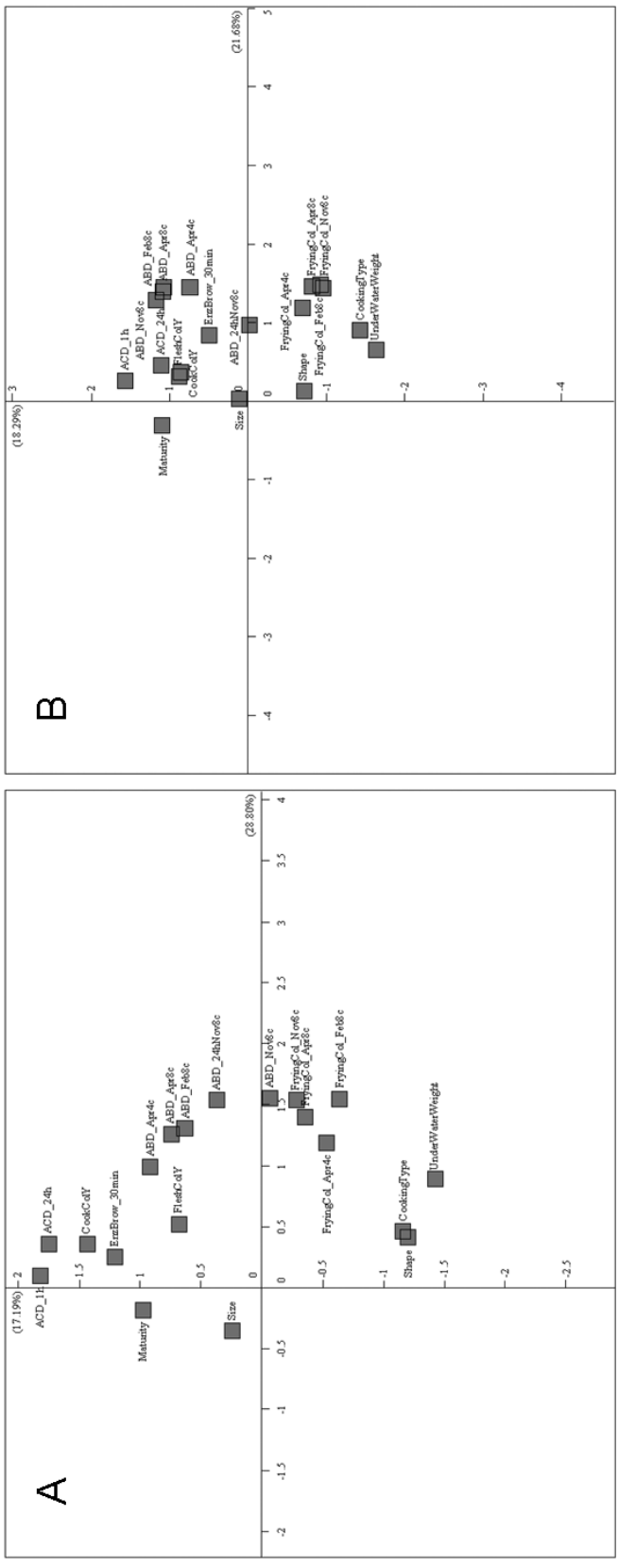


Figure 3: Biplot showing genetic correlations. Clustered squares, sharing approximately the same direction (a sharp angle) represent highly correlated traits. Squares at opposite sides of the origin (obtuse angles) represent negatively correlated traits. Squares with orthogonal directions represent independent traits.

A. Trait correlations as deduced from the 2006 data
B. Trait correlations as deduced from the MYML data

Figure 4 illustrates individual trait relationships between the two data sets and also contains information about the magnitude of variances for specific traits present in both data sets. When the points are concentrated along the diagonal, this is an indication of similar behaviour of traits in both data sets and high correlation. By comparing the range of the genotypic main effects, it can be deduced whether both data sets withhold the same amount of information, i.e. variance, for a particular trait. For example tuber shape, tuber size, underwater weight, maturity, cooking type, ACD_1h, flesh colour, cooking colour, enzymatic browning and the frying colour traits all have their points more or less along the diagonal, an indication that these traits correlate well. Tuber size, underwater weight, cooking colour and the frying colour traits all have comparable ranges for genotypic main effects in both data sets, indicating that equal variance is present in both data sets for these traits. Tuber shape, maturity, cooking type, ACD_1h, flesh colour and EnzBrow_30min have a slightly higher range for their genotypic main effects in the MYML data set, which suggests that the MYML data set has a slightly higher information content for these traits. This illustrates what

already could be interpreted from the correlations presented in **Table 6**, namely that, at least for the traits with a high correlation, both data sets are equally valuable. As an important consequence, historical data may serve equally well for association studies in particular traits as a designed trial.

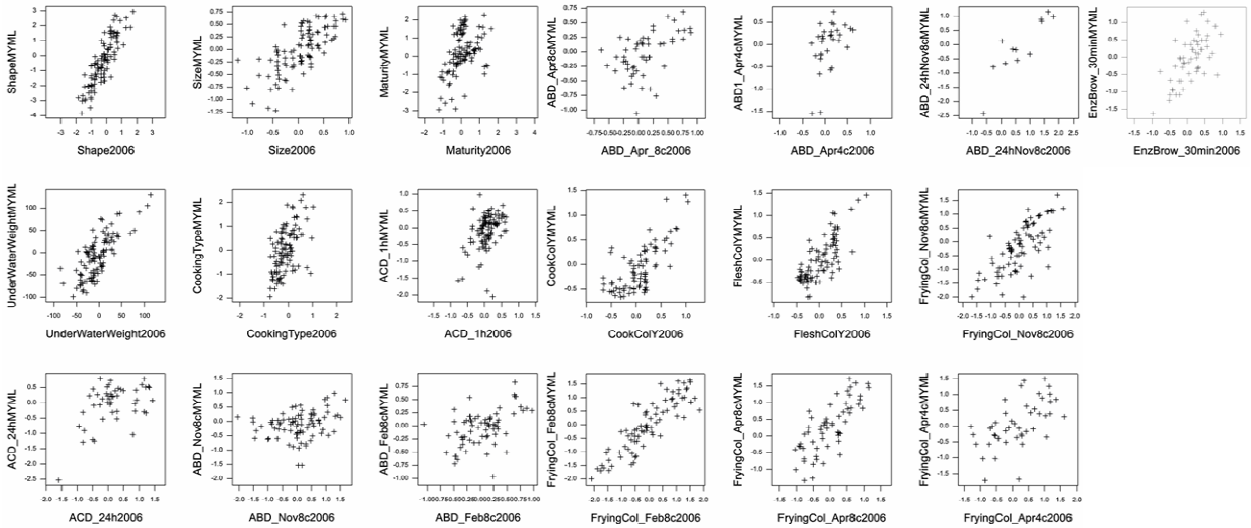


Figure 4: Comparison of 2006 versus MYML genotypic effects for 19 traits. The scatter plots display the variability of the traits across genotypes as observed in 2006 and MYML data, as well as the pair-wise correlation of the traits between both data sets.

Discussion

We encountered various obstacles during the compilation of the historical data set. One of the first things one has to deal with while assembling data from different companies is to match traits with each other. This problem can pose itself in various ways, the trait names can be different while the same phenotype is measured, but the other way around is also possible. The trait names can be the same whereas a different phenotype is measured or the scales that are used can be (slightly) different whether in order or in increments or in range. Not to mention that the protocols to measure the same phenotype can differ. A clear trait ontology, analogous to the plant structure ontology intended for anatomy and morphology of flowering plants (Ilic et al. 2007), with an accurate trait definition and protocol description to evaluate a certain trait could be very helpful in the case of potato. Some nice examples of such an attempt already exist in cereal genomics (Jaiswal et al. 2002) and in tomato (Brewer et al. 2006). Moreover, in view of association mapping and to increase robustness and power to detect QTL, it would be advisable to develop or apply objective measurement tools to evaluate traits that could replace the more subjective scores by breeder's eye still applied in today's potato breeding practice.

When in the end reliable data sets have been established as a result of trait ontological considerations applied to the different sublevel data structures, it becomes feasible to compare the patterns of variation and correlations in these data sets. Here we compared one data set that resulted from a designed single-year-two-locations field trial with a historical data set that resulted from multi-year-multi-location (MYML) trials.

Table 6: Correlations between genotypic effects in designed and historical data. On the diagonal the correlations between the two assessments for the same trait. Values higher than 0.3 are highlighted.

	ABD_Nov8c_2006	ABD_Apr8c_2006	ABD_Apr4c_2006	ABD_Feb8c_2006	ABD_24hNov8c_2006	ACD_1h_2006	ACD_24h_2006	CookColY_2006	CookingType_2006	EnzBrow_30min_2006	FleshColY_2006	FryingCol_Nov8c_2006	FryingCol_Feb8c_2006	FryingCol_Apr8c_2006	FryingCol_Apr4c_2006	Maturity_2006	Shape_2006	Size_2006	UnderwaterWeight_2006
ABD_Nov8c_MYML	0.20	0.32	0.34	0.37	0.29	0.40	0.37		0.12	0.36		0.16	0.15	0.18				0.21	
ABD_Apr8c_MYML	0.31	0.31	0.20	0.31	0.41	0.32	0.41	0.12		0.20		0.32	0.17	0.24	0.16				0.11
ABD_Apr4c_MYML	0.23	0.22	0.29	0.23	0.31	0.18	0.21		0.15	0.22		0.22	0.11		0.14		0.14		0.13
ABD_Feb8c_MYML	0.17	0.28	0.31	0.35	0.27	0.27	0.39		0.13	0.30		0.16	0.11	0.14					0.11
ABD_24hNov8c_MYML	0.24	0.16	0.20	0.21	0.25	0.13	0.20	0.14		0.13	0.10	0.24	0.21	0.29	0.17			0.15	
ACD_1h_MYML		0.21	0.18	0.23		0.39	0.41	0.20		0.34						0.13			
ACD_24h_MYML		0.18	0.23	0.22	0.16	0.23	0.35	0.17		0.14									
CookColY_MYML	0.19	0.12	0.26	0.17	0.26		0.11	0.57			0.71	0.22			0.11				
CookingType_MYML	0.39	0.23		0.29	0.36				0.48			0.38	0.47	0.31	0.31		0.28		0.50
EnzBrow_30min_MYML	0.20		0.22	0.22	0.20		0.10			0.42		0.14	0.11						
FleshColY_MYML	0.19	0.28	0.20	0.29	0.27		0.17	0.69		0.17	0.75	0.18		0.13	0.12				
FryingCol_Nov8c_MYML	0.59	0.42	0.34	0.41	0.55		0.11		0.18		0.13	0.59	0.70	0.63	0.51				0.31
FryingCol_Feb8c_MYML	0.65	0.40	0.28	0.43	0.61		0.12		0.19		0.13	0.65	0.73	0.57	0.49		0.10		0.30
FryingCol_Apr8c_MYML	0.59	0.35	0.33	0.36	0.54		0.11		0.14		0.13	0.57	0.62	0.47	0.38				0.19
FryingCol_Apr4c_MYML	0.34	0.27	0.33	0.31	0.30		0.15					0.36	0.40	0.38	0.38				0.12
Maturity_MYML																0.44			
Shape_MYML	0.15				0.10				0.25			0.15	0.25	0.12	0.24		0.86		0.32
Size_MYML										0.14								0.70	
UnderwaterWeight_MYML	0.39	0.18	0.12	0.24	0.31				0.51			0.40	0.49	0.42	0.46		0.33		0.76

Heritabilities, coefficients of variation, trait relationships within and between data sets, and phenotypic variance attributable to genotype have been estimated and compared. In **Tables 4** and **5** an overview of the analyses of both data sets are presented. Especially the columns with the magnitude of the genotypic variance and the heritability are interesting because they give an impression of the feasibility of genetic improvement and possibilities for QTL detection. **Figure 1** shows the average for the MYML heritabilities to be higher than the ones from the 2006 field trial. **Figure 2** demonstrates that traits in both data sets behaved consistently when their coefficients of variance are considered, indicating that both data sets contain data of comparable quality. Trait relationships within each data set are depicted through biplots in **Figure 3** and correlations between traits in both data sets are tabulated in **Table 6**. Basically, the same trait relationships persisted within each data set, but the strength of the relationships varied between data sets. Agro-morphological traits have higher pair-wise correlations between the two data sets than quality traits, except for frying colour traits. This trend is visually expressed in **Figure 4**, where the individual pair-wise trait correlations are depicted in scatter plots.

The fact that the designed trial heritabilities were on average clearly higher than the historical ones can be interpreted in two ways. The MYML data set included more environments. The more environments for which trait data are available the more conservative heritability estimates will be, due to a balancing effect between different environments. Extreme weather conditions leading to deviating trait observations will be buffered by trait observations obtained from environments exposed to normal climatic circumstances. A second factor that could inflate the heritabilities observed for the 2006 data set could result from the choice for diversity among the cultivars in this set. About one third of the cultivars had a first registration on a national list in Europe dating before 1930, and in those days for most traits more deviating values were accepted that are no longer observed in contemporary cultivars.

Phenotypic values of traits with high heritabilities are expected to correlate well between the 2006 and MYML data sets. Indeed, this is confirmed by the correlations between genotypic main effects, as shown in **Table 6**. These correlations range from 0.2 for ABD_Nov8c to 0.86 for tuber shape. Tuber shape, underwater weight, flesh colour and cooking colour do not only have higher heritability estimates, but are also higher correlated between the two data sets than quality traits like after baking darkening and after cooking darkening. Unexpectedly, the high heritability for maturity is not equalled by a strong correlation between the data sets. This could have been caused by the exceptional summer of 2006 initially causing drought induced senescence for certain genotypes and renewal of foliage development for other genotypes later in the season once abundant rain was available. Although the data structure of both data sets is quite different most of the traits correlate reasonably well, which is promising for future genome-wide association analyses. When inspecting off-diagonal correlations, some striking observations appear. For example, the expected relationship between flesh colour and cooking colour appears even when correlating their trait values from two different data sets. The same holds for the traits describing frying colour. Surprisingly, there seems to be a rather high correlation between ABD traits and frying colour traits, which partly confirms the composite effect resulting in discolouration phenotypes.

Figure 3 allows us to look at genetic correlations between traits within data sets. The clusters of frying colour traits and after baking darkening traits that were detected in both data sets are

according to expectation as these traits measure exactly the same phenotypes. The only differences between the traits are the temperature at which the tubers have been stored and the duration of storage prior to evaluation. Theoretically spoken, there are no genes involved in the discolouration process itself: for frying colour the discolouration is a result of the Maillard reaction, whereas for after baking darkening it is the result of oxidation of an iron-chlorogenic acid complex (McKenzie et al. 2005). However genetics does influence the amount of reaction products available for discolouration processes. Olsson and colleagues (2004) provided evidence that storage time and temperature do not influence the ranking of genotypes with respect to frying colour that much, which could explain the strong relationships between this set of traits. Wang-Pruski and Nowak (2004) say that for after baking darkening storage time and temperature only have an influence on the severity of the discolouration, nothing is mentioned about change of ranking of genotypes. The positive correlation between cooking type and underwater weight that persisted in both data sets is equally expectable. Underwater weight is a measure for dry matter content of potato tubers and indirectly also for starch content: the higher the underwater weight, the higher the starch content of a cultivar. Starch granules create swelling pressure while being cooked and therefore can contribute to disintegration of the cooked tuber structure which creates a looser cooking type (van Marle et al. 1997). The relationship between flesh colour and cooking colour is even more predictable: both traits measure exactly the same phenotype, the mere difference is that the first is evaluated on raw tubers and the latter on cooked tubers. The negative correlation between maturity and underwater weight can be explained as follows, taking the characteristics of the traits into account: later maturing cultivars reside longer in the field and thus have the opportunity to assemble more sucrose as their photosynthesis remains active for a longer time period, and so, can stack larger amounts of starch in their tubers which increases their underwater weight.

To get an impression of the genetic variation in our data sets with respect to other studies, it is interesting to compare obtained heritabilities with previously reported heritabilities, where the remark has to be made that in our study we calculated a ratio of variance components that strictly spoken is not really a heritability, but rather a repeatability. Maris (1969) reports heritabilities pooled over 12 progenies of potato. For maturity and underwater weight he obtained values of 0.85 and 0.76 respectively, similar to our estimates (see also **Tables 4** and **5**). Tai and Young (1984) reported a somewhat lower heritability of 0.67 for maturity and 0.65 for specific gravity, a trait directly related to underwater weight. Yildirim and Çalişkan (1985) obtained 0.77 as heritability estimate on a clone mean basis for tuber shape evaluated in a field trial performed over two years and three locations, likewise comparable to what we computed. Love et al. (1997) reported heritabilities on a clone mean basis of 27 traits obtained from a study involving three clonal generations with 200 clones. For shape they calculated a heritability of 0.79, we obtained 0.74 for the MYML data set and 0.84 for the 2006 field trial data set. Thompson et al. (1983) estimated tuber size heritability to be 0.79 for a *Solanum andigena* population propagated by true potato seed, we obtained lower values: 0.52 within the MYML data set and 0.71 within the 2006 data set. Neele et al. (1991) estimated heritabilities for maturity, shape and underwater weight to be 0.77, 0.61 and 0.68 respectively. Bradshaw et al. (2008) reported heritabilities on a clone mean basis for maturity of 0.92, dry matter 0.81, tuber size 0.87, tuber shape 0.85, after cooking darkening 0.60, frying colour at 4˚C 0.79 and frying colour at 10˚C 0.78. These estimates are slightly higher than or comparable to the values we obtained (see **Tables 4** and **5**). The paper of Bradshaw (2008) also reveals a worrying aspect. Albeit that these high heritabilities were obtained with a field trial of very high quality, they could

only detect very few QTL with a high explained variance (>10%). There appears to be a discrepancy between two kinds of traits: oligogenic traits, like maturity and tuber shape, where few QTL were detected with high explained variance, in contrast to true polygenic traits like after cooking darkening and frying colour, where more QTL with lower explained variance were detected. However, it is plausible that the amount of markers used in that study, namely 514, together with the presence of maximum eight haplotypes in the mapping population analysed, was perhaps not powerful enough.

Multi-year-multi-location (MYML) data sets would always be preferred over single-year-single-location data sets when scientists had the opportunity to choose. Just for the sake of being able to perform sound statistics and obtain more reliable results. Often dilemmas have to be solved before the start of projects in the genre of: do we opt for one year of field trials with several locations and replicates, or for more years of trials with several locations and less replicates. Looking back at the outcome of our phenotypic analyses it is obvious that the heritabilities obtained with the two types of data sets are different. But, when trait correlations within or between data sets are concerned, the same conclusions could be drawn. So, what is the best choice? Both approaches have their pros and cons. When choosing for just one year of field trials, one runs the risk of having a bad climate and/or obtaining overestimated heritabilities due to a smaller number of environments, lack of seasonal balancing effects or the effect of some extreme genotypes. Nevertheless, a bad climate season does not necessarily pose such a big risk, since extreme weather conditions can be corrected for and genotype rankings, at least in our case, were not affected. When opting for MYML, one chooses for certainty of obtaining reliable data, heritabilities will be more conservative and probably more realistic and more variable (environment-sensitive) traits will become better accessible. However with MYML data some aspects can change in the course of time, for instance reference cultivars can change as new standards are being set by breeding companies or scoring methodology of traits can be altered due to new techniques or insights. Usually the GxE analysis will become increasingly complex as more trials are to be included in the data analysis and the unbalancedness will only increase. Not to mention that it simply takes longer before analyses can be performed. In all cases, it is advisable to investigate into repeatability of detection of QTL over years or locations, as this offers possibilities in detection of stable QTL and to unmask incidental environmental QTL.

Plot size is another factor to consider when designing field trials. Polemics about this issue are the order of the day, some breeders claim that small plot size, e.g. four plants per genotype, suffice to evaluate most of the traits, others say that you need at least a plot size of 16 plants per genotype to get a reliable idea of the performance of a genotype. Considering the phenotypic analysis of our 2006 field trial experiment, given acceptable environmental conditions and good experimental design, for most simple traits, such as maturity, tuber shape or flesh colour, small plot size indeed suffices to obtain good quality phenotypic data. But when more variable traits, like discolouration traits, have to be assessed, a larger plot size can improve the results of the trial by rendering more sample material and it offers a buffer to extreme environmental conditions.

In order for marketing brochures or national lists to be really informative about a cultivar's performance for a certain phenotypic trait, it is necessary to present only true unbiased trait values. The philosophy behind these kinds of lists is to offer trait values with which cultivars can

be ranked with respect to each other, so that conclusions can be drawn. Therefore it is perhaps advisable to include adjusted trait mean values, corrected for GxE interaction, as obtained in this study, instead of simply averaging trait values over performed trials. At least in our case, these adjusted means offered robust data that could be reliably compared throughout data sets resulting from different field trials.

The phenotypic analysis presented in this article is the first step of an encompassing two-step association mapping study. We intend to integrate a high amount of genotypic marker information to detect associations for the traits mentioned. The outcome of the current analyses produced adjusted genotypic means that subsequently will be used as entry values for further association analysis, a similar approach as described in D'hoop et al. (2008) and Li et al. (2008).

Acknowledgements

We appreciate Agrico Research (Bant, Netherlands), Averis seeds (Valthermond, Netherlands), C Meijer (Rilland, Netherlands), HZPC Research (Metslawier, Netherlands) and Van Rijn (Emmeloord, Netherlands) for sharing their phenotypic data and for their support and help in the 2006 field trial. We would also like to thank the "2006 field trial team" for their help.
Project P8 was financed by the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

Chapter 4

Population structure and LD in tetraploid potato

Björn B. D'hoop
M. João Paulo
Krissana Kowitwanich
Mariëlle Sengers
Richard G.F. Visser
Herman J. van Eck
Fred A. van Eeuwijk

Abstract

Association mapping is regarded as an important alternative strategy for the identification of QTL as compared with traditional QTL mapping. An important prerequisite for association analysis however, is detailed information regarding hidden population structure and the extent of linkage disequilibrium (LD). A collection of 430 tetraploid potato cultivars, comprising two association panels, has been analysed with 41 AFLP™ primer combinations yielding 3364 AFLP fragments, as well as with 53 SSR loci. LD, using r^2 as a criterion, was on average observed up to approximately 5cM. Using a Bayesian approach and a distance-based clustering approach population structure was investigated, and only very little structure could be found. We detected groups differing in year of market release and market segment (starch, processing and fresh consumption). These results, LD up to 5cM and mild population structure, are promising for association mapping research in tetraploid potato, because there seems to be a good chance that linkage-based marker-trait associations can be identified at moderate marker densities.

Keywords: linkage disequilibrium, population structure, potato, tetraploid

Introduction

Association mapping has become a valuable tool for plant breeders in identifying quantitative trait loci (QTL) responsible for variation in complex phenotypic traits, complementary to traditional QTL mapping. Advances in molecular marker technology and statistical methods have made association mapping accessible and affordable (Zhu et al. 2008). There are two major advantages inherent to association mapping: (1) assessing an ample amount of genetic diversity through the use of diverse collections of cultivars and breeding material enables a wide applicability of obtained results, and (2) higher mapping resolution may be reached as more meiotic recombination events are sampled through a set of diverse genotypic material (Flint-Garcia et al. 2003; Gaut and Long 2003; Jannink and Walsh 2002).

Specifically the aspect of wide genetic diversity can also hamper the interpretation of obtained results as population structure and relationships among genotypes can negatively affect the outcome of association mapping by causing false positives (Flint-Garcia et al. 2003; Zhu et al. 2008). It is therefore a common strategy to first inspect the collection of genotypes to be used in an association mapping study for possible population structure followed by incorporation of correction factors for group effects. The idea here is that only true associations, i.e. caused by physical linkage, will remain (Yu et al. 2006).

There are several ways to detect population structure in a collection of cultivars and subsequently incorporate that information into association analysis. One of the more popular ways is to apply the approach developed by Pritchard et al. (2000), implemented and freely available through the software package STRUCTURE. Within a Bayesian framework, group membership probabilities are assigned to genotypes using molecular marker information. The estimated group membership probabilities can subsequently be used to assign genotypes to different groups, after which marker-trait association analysis takes place within the identified groups (Remington et al. 2001; Simko et al. 2004b). Alternatively, the group memberships can be translated into an extra set of covariables or a factor in a statistical model relating phenotypes to genotypes (Thornsberry et al. 2001). Another way to classify genotypes is to apply standard multivariate analysis methodology like clustering, where the input matrix of genotypic distances can be based on molecular marker data. Similarly, one may derive a distance matrix from pedigree information. Identified groups from cluster analysis can subsequently be used as a correction factor in association analysis (Kraakman et al. 2004; Simko et al. 2004a). One can also opt for a direct approach where a genotypic relationship matrix is constructed based on neutral marker info or pedigree data to impose structure on the variance-covariance matrix of the genetic effects (Malosetti et al. 2007; Parisseaux and Bernardo 2004). Yu et al. (2006) include both a factor representing population structure and a marker-based relationship matrix in a mixed model framework for association analysis.

The feasibility of association mapping within a given crop species, i.e. the power to assess marker-trait associations, depends on the rate of decay of linkage disequilibrium (LD) between loci. This decay rate is a function of the number of generations that passed since domestication. The lower the decay rate the easier it will be to detect marker-trait associations. However, the faster LD decays the better the opportunities for fine mapping. For slow LD decay rates, whole-genome LD scans become practical (Breseghello and Sorrells 2006a; Mackay and Powell 2007).

Potato, an autotetraploid crop species (2n=4x=48), is our crop model. Population structure and LD have previously been assessed in potato within an association mapping context by Simko et al. (2004a,b), Gebhardt et al. (2004), Malosetti et al. (2007) and D'hoop et al. (2008). None of them have detected statistically significant population structure, whereas LD decay estimates varied from rapidly decaying (<1cM: Gebhardt et al. 2004) over a more slow decay (~3cM: D'hoop et al. 2008) to a long range decay of about 10cM (Simko et al. 2004a). Unfortunately all these estimates were based on, or a limited amount of marker information, or just a few DNA sequences. Successful QTL mapping research has already been performed at the diploid level (e.g. Costanzo et al. 2005; Malosetti et al. 2006; Werij et al. 2007) and at the tetraploid level (Bradshaw et al. 2004, 2008; Khu et al. 2008). Nonetheless, we anticipate that association mapping in tetraploid potato has an added value in that it results in QTL in a relevant genetic background and at the relevant ploidy level.

In this paper, we present the inspection of a large amount of collected AFLP and SSR marker information for evidence of population structure in a big collection of tetraploid potato cultivars using both a Bayesian approach and a neighbour joining cluster analysis approach. Results obtained with both approaches are compared and discussed. The same marker information is subsequently used to inquire into the LD pattern along the potato genome. The resulting information on structure and LD decay will be deployed in an encompassing association mapping study currently undertaken.

Materials & methods

Plant material

We collected a representative subset of worldwide available commercial potato germplasm containing 221 tetraploid potato cultivars and progenitor clones. For details about criteria used for the composition of this set we refer to D'hoop et al. (2008). This initial core set was expanded in a later stage with the parents of the SHxRH diploid mapping population (van Os et al. 2006) to enable marker positioning, 17 extra tetraploid potato cultivars selected to enlarge diversity coverage and 190 advanced breeder's clones selected to represent the current breeding germplasm pool. The material of in total 430 potato genotypes was kindly provided by five Dutch breeding companies and several gene banks (see acknowledgements). Leaf material was harvested from greenhouse-grown and in-vitro-grown genotypes, was frozen with liquid nitrogen and stored at -80˚C until DNA extraction.

Molecular marker analysis

DNA extraction was according to van der Beek et al. (1992). DNA quality and concentration were visually examined using ethidiumbromide stained 1% agarose gels.

AFLP™

AFLP markers were generated according to Vos et al. (1995) using 26 well-known *Eco*/*Mse* and 15 well-known *Pst*/*Mse* primer combinations: E+AAC/M+AGG, E+AAC/M+CAC, E+AAC/M+CAG, E+AAC/M+CCA, E+AAC/M+CCT, E+AAC/M+CTC, E+AAC/M+CTG, E+AAG/M+ACC, E+AAG/M+AGC, E+AAG/M+CAC, E+AAG/M+CGA, E+ACA/M+CAC, E+ACA/M+CAG, E+ACA/M+CCT, E+ACA/M+CTG, E+ACC/M+AGT, E+ACC/M+CAT, E+ACC/M+CTT, E+ACT/M+CTC, E+AGA/M+ACC, E+AGA/M+CAG, E+AGA/M+CAT,

E+AGA/M+CTC, E+AGT/M+CAG, E+ATG/M+CTA, E+ATG/M+CTC, P+AC/M+AAC, P+AC/M+ACT, P+AC/M+AGC, P+AC/M+AGG, P+AC/M+AGT, P+AC/M+ATG, P+AG/M+AAG, P+AG/M+ACC, P+AG/M+AGC, P+AG/M+AGT, P+AT/M+AAC, P+AT/M+AGG, P+CA/M+ACT, P+CA/M+AGG, P+CT/M+AGG. Fragments were separated using a capillary sequencer (MegaBACE 1000, Molecular Dynamics & Amersham, serial number 13757) according to van Eijk et al. (2004), each primer combination being labelled with either FAM, NED or JOE. The ROX channel was used for the MegaBACE™ ET900-Rox size standard from GE Healthcare (Amersham Biosciences). Pseudo gel images were scored with proprietary software at Keygene N.V. (Wageningen, the Netherlands). Marker nomenclature was based on the restriction enzyme combination, selective nucleotides and fragment mobility relative to a ROX labelled size ladder.

Normalisation of signal intensity variation between capillaries due to DNA loading effects was performed by deducting genotype means (lane/column means) from the log-transformed band intensity values. Effectively, the residual log intensities were retained after fitting a genotype main effect to the genotype-by-marker log intensities, using the ANOVA procedure in GenStat, 11th edition (VSN International Ltd., Oxford, UK).

Information on the position of the AFLP markers was retrieved from the ultra dense potato map, using the parental diploid genotypes SH83-92-488 and RH89-039-16 as internal reference (van Os et al. 2006; http://www.plantbreeding.wur.nl/potatomap/).

SSR

53 microsatellite primer pairs (**Table 1**), previously designed based on expressed sequence tag database information (Bradshaw et al. 2008; Feingold et al. 2005; Ghislain et al. 2004; Kawchuk et al. 1996; Milbourne et al. 1998; Rios et al. 2007), were selected using the following criteria: (1) amplification products should map to a single locus, (2) their quality score when available, (3) their linkage group and (4) their map location within a linkage group pursuing at least one SSR marker for each chromosome arm. Linkage group *8* is overrepresented in this set with 18 markers to be able to test our scoring methodology and linkage disequilibrium measurement by deduction of genomic marker order from LD. Primer sequences, labelled with HEX, NED or 6-FAM, were modified by adding pigtail nucleotides according to Brownstein et al. (1996) to avoid as much as possible the appearance of stutter bands in the electropherograms. Microsatellites were amplified by separate PCR in a 20µL reaction volume, containing 10ng genomic DNA, 75mM Tris-HCl pH 9.0, 20mM $(NH_4)_2SO_4$ 0.1% (w/v) Tween 20, 2.5mM $MgCl_2$, 100µM of each dNTP (Fermentas), 4pmol of each primer and 0.3 U Goldstar *Taq* DNA polymerase (Eurogentec). The optimised PCR conditions were one cycle of 94˚C for 3min, followed by 40 cycles of 94˚C for 30s, 50˚C for 30s, 72˚C for 45s and a final extension at 72˚C for 10min.

PCR amplification products were visually examined using ethidiumbromide stained 2% agarose gels along a 1kb ladder (Invitrogen). Differently labelled PCR products (6-FAM, HEX and NED) were combined in appropriate amounts to obtain optimal peak patterns for detection. The fluorescently labelled products were separated by capillary electrophoresis using an ABI Prism 3700 DNA Analyzer (Applied Biosystems). Electropherograms were created automatically using GeneScan Analysis Software version 3.7 (Applied Biosystems). Peak mobilities and areas were determined using ABI PRISM Genotyper® 3.6 NT software.

Allele calling was supported by peak mobility distribution plots because peak mobilities can vary within electrophoretic runs and between runs. Quantitative peak area information as provided by the software was used to calculate pair-wise peak area ratios in order to determine allele copy numbers in individual samples according to the microsatellite allele counting-peak ratios (MAC-PR) methodology described in Esselink et al. (2004). Whenever four alleles were detected, single dosage was assumed for each allele as our samples were obtained from tetraploid individuals. When a single allele was detected, a dosage of four was suggested under the assumption of absence of null alleles. We do acknowledge the possibility of presence of non-detected null alleles as potato is highly heterogeneous, however when no obvious proof of presence was available absence of null alleles was maintained as valid genotypic model assumption. When unambiguous evidence of null alleles was available (e.g. a peak area ratio of ~2 or ~0.5 when only two alleles were detected or pair-wise peak area ratios of ~1 when only three alleles were detected), null alleles were defined and called as a separate allele "0" in the genotypic model as they represent an extra haplotype and so contain extra information. Also in this case we acknowledge that it is likely that obvious null alleles are not all of the same size and should therefore not be classified within one single allelic class. For convenience and to keep the statistical analysis simple we chose to lump all null alleles for a certain SSR locus into one single allelic class "0".

Position information regarding the microsatellites was obtained from literature and when polymorphic within the SHxRH population from the ultra dense potato map, using the parental diploid genotypes SH83-92-488 and RH89-039-16 as internal reference (van Os et al. 2006; http://www.plantbreeding.wur.nl/potatomap/).

As an approximate descriptive statistic for tetraploid potato, the Polymorphism Information Content (PIC) value was calculated according to Nei's statistic (Nei 1973), namely $PIC = 1-\sum p_i^2$, where p_i is the allele frequency, using co-dominant scores, of the ith allele of a certain microsatellite locus detected in the germplasm.

Population structure

The possibility of hidden population structure in our diverse collection of potato cultivars was addressed with two different approaches. The first approach is based on the Bayesian modelling environment implemented in the programme STRUCTURE, version 2.1 (Pritchard et al. 2000). This program identifies putative subgroups based on the assumption of Hardy-Weinberg and absence of LD inside subgroups. LD originates then principally from differences in allele frequencies between subpopulations. Based on position information and linkage configurations of the markers, as obtained from the ultra dense genetic map of potato (van Os et al. 2006), we created four different sets of polymorphic AFLP markers for the analysis with STRUCTURE. These four subsets are (1) a redundant set of 315 AFLP markers, (2) a subset of 103 approximately independent AFLP markers that were about equidistantly spaced every 5cM, (3) a non redundant subset of 37 markers available from the 48 positions at the 12 centromeric marker clusters from both parental maps, and linked in trans configuration, and therefore haplotype specific, and (4) a subset of 48 AFLPs from the 48 telomeric positions from 12 chromosomes in two parents. STRUCTURE was run under the assumption of admixture with independent allele frequencies and no *a priori* population information was used. Simulations were performed for the number of subgroups, K, ranging from two to 20 with two independent repeats for each K and with a total of 150,000 iterations, of which the first 50,000 were

considered as burn-in. Apart from this, STRUCTURE was also applied to the set of 53 microsatellite loci with the same model assumptions and settings, although approximate mapping distance information was provided to STRUCTURE to account for positional clustering of some microsatellite loci. In all cases with AFLP data the ploidy was set to haploid, while with the SSRs ploidy was set to four with phase unknown.

The second approach is a cluster analysis based on the program DARWIN (Dissimilarity Analysis and Representation for Windows, version 5.0.155, Perrier and Jacquemoud-Collet 2006). DARWIN was run with four sets of AFLP data: (1) a full set of 3364 AFLP markers with normalised log-transformed band intensity information, (2) a subset of 1700 polymorphic AFLP markers using the dominantly scored presence/absence information, (3) a subset of 315 absence/presence polymorphic AFLPs with known map location and (4) an equidistant subset of 103 AFLPs with presence/absence information and spaced approximately every 5cM on the genetic map of potato. Apart from this, a set of 53 SSR loci was used as well for cluster analysis with DARWIN, however due to the presence of null alleles, the genetic dissimilarities had to be calculated with GenStat, 11th edition, prior to data import in DARWIN. For the calculation of the dissimilarity matrix, we opted for each analysis with presence/absence or allele dosage data for the Jaccard dissimilarity index, whereas for the continuous band intensity data we opted for an Euclidian type of dissimilarity index. For each dissimilarity calculation prior to cluster analysis with DARWIN, 100 bootstraps and a minimal proportion of 50% of valid data for pair-wise dissimilarity calculations were selected as options. For tree construction we opted for hierarchical clustering using the Ward minimal variance methodology. Information concerning market niche, year of registration and country of origin was used as identifier set.

LD assessment

Linkage disequilibrium (LD) between loci was quantified with the squared correlation coefficient, r^2 (Flint-Garcia et al. 2003; Zhao et al. 2005), between log-transformed normalised AFLP band intensities (D'hoop et al. 2008) using a set of 720 AFLP markers with known map location on the ultra dense genetic map of potato (van Os et al. 2006). LD decay was visualised per chromosome by plotting r^2 versus map distance in centiMorgans.

The pattern of LD along the potato genome was also investigated for a set of 53 microsatellites, where we adapted the method of Flajoulot and colleagues (2005) for the calculation of r^2 based on co-dominant microsatellite data to include full zygosity information on SSR alleles. As a measure for LD between two loci we first assessed which allele was the most frequent one for each of both loci and then simply calculated the squared ordinary Pearson product moment correlation between the copy numbers, with possible values 0, 1, 2, 3 or 4, of these two most frequent alleles. A graphical representation of the LD pattern along the potato genome was accomplished with GenStat, 11th edition (VSN International Ltd., Oxford, UK).

Results

Molecular marker analysis

The 41 AFLP fingerprints were analysed with proprietary software at Keygene, and those markers which could be scored unambiguously were retained. In total 3364 AFLP fragments were distinguished on 430 genotypes, including a threefold repetition of the diploid mapping parents SH83-92-488 and RH89-039-16 which were included as internal reference genotypes.

On average a primer combination produced 82 fragments. Of the 3364 AFLP fragments, 628 did not display absence/presence segregation, but these “constant” bands may segregate as allele dosage polymorphisms (i.e. reflect a very high AFLP allele frequency rendering predominantly quadruplex, triplex and perhaps some duplex genotypes). In addition, 1144 markers showed an absence/presence polymorphism and could be scored in a dominant qualitative fashion. To allow us to use all 3364 fragments for population structure and LD analyses, we worked with the quantitative band intensities. See D’hoop et al. (2008) for a justification of working with band intensities.

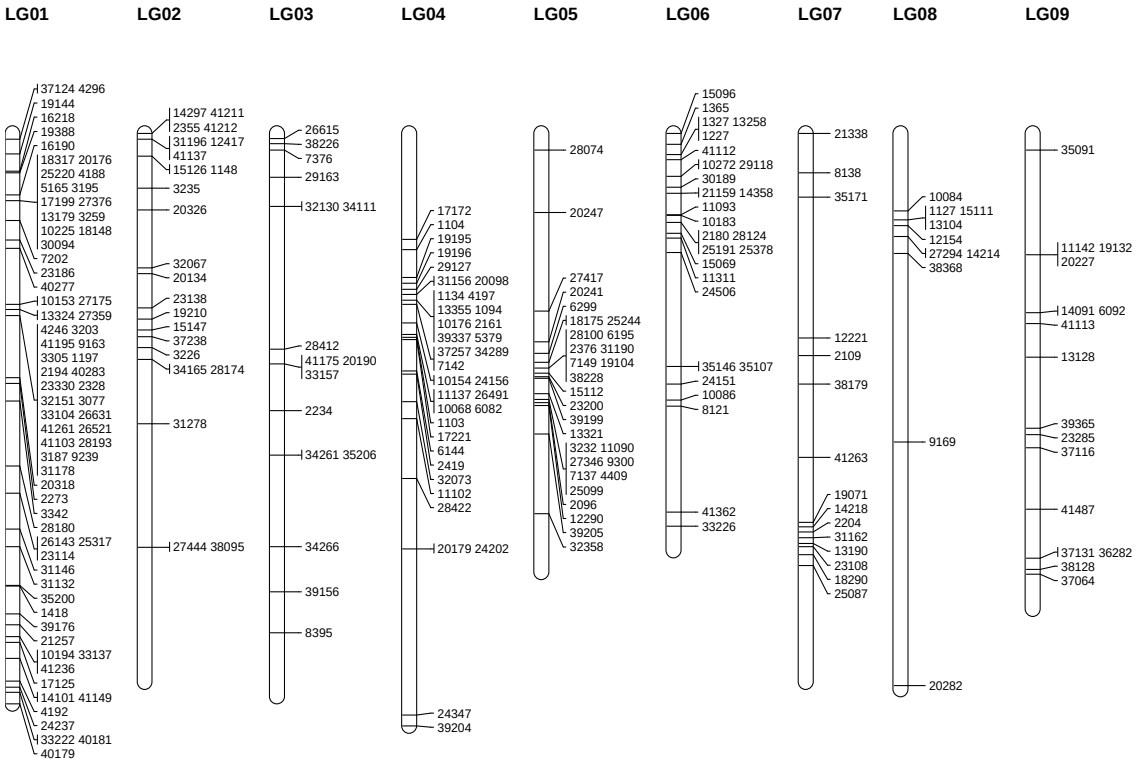


Figure 1: AFLP marker distribution along the potato genome. Map positions are according to van Os et al. (2006).

In total 720 markers, of which 315 were scored qualitatively, could be assigned to a genetic position on the ultra dense potato map (van Os et al. 2006). We assumed that the position of these 720 markers in the diploid UHD map is essentially not different from the position of these bands in the collection of tetraploid genotypes. If position information is available, then it is also known if these markers are from the different parents, as well as their linkage in cis or trans configuration. Markers from different parents or linked in trans configuration tag different haplotypes. Although male and female cM positions may differ due to differences in recombination in male and female meiosis, we did not attempt to assign sex-averaged recombination distances between the markers. In **Figure 1** the distribution of mapped markers across the 12 potato chromosomes is illustrated.

Table 1: Overview of used microsatellite loci including the number of alleles, average number of alleles per genotype, number of genotypes and number of unique genotypes detected, the frequency of the most abundant genotype, the PIC-value and linkage group according to literature.

SSR Locus	Number of alleles	Average number of alleles per genotype	Number of unique genotypes	Number of genotypes	Frequency most abundant genotype	PIC	LG	Source of SSR
STI009	15	2.23	56	125	24	0.778	1	Feingold et al 2005
STM1029	11	1.52	34	59	121	0.709	1	Milbourne et al 1998
STM1049	10	1.45	9	24	187	0.576	1	Milbourne et al 1998
STM5127	8	1.64	29	63	113	0.67	1	Rios et al. 2007
STI029	16	2.67	80	162	15	0.83	2	Feingold et al 2005
STI052	10	2.19	27	71	47	0.734	2	Feingold et al 2005
STM0038	12	2.08	23	71	45	0.766	2	Milbourne et al 1998
STM1030	9	1.65	19	42	85	0.591	2	Milbourne et al 1998
STM2022	7	1.42	10	30	192	0.514	2	Milbourne et al 1998
STM1054	8	1.02	7	11	380	0.055	3	Milbourne et al 1998
STI001	12	2.07	38	78	96	0.667	4	Feingold et al 2005
STI012	13	2.74	47	123	17	0.814	4	Feingold et al 2005
STI055	9	2.24	26	57	53	0.656	4	Feingold et al 2005
STM3016	10	1.71	30	64	60	0.776	4	Milbourne et al 1998
STM3020	2	0.84	1	2	358	0.003	4	Milbourne et al 1998
STM3023	9	1.55	10	35	72	0.716	4	Milbourne et al 1998
STM0013	25	2.37	55	103	76	0.737	5	Milbourne et al 1998
STM1041	6	1.79	7	23	93	0.585	5	Milbourne et al 1998
STM5148	17	2.77	152	232	15	0.865	5	Bradshaw et al. 2008
STI016	14	1.78	31	62	51	0.723	6	Feingold et al 2005
STI045	8	1.84	8	24	147	0.495	6	Feingold et al 2005
STM0001	16	1.67	58	101	132	0.722	6	Milbourne et al 1998
STM1100	26	1.64	51	83	59	0.718	6	Milbourne et al 1998
STI040	10	1.02	10	23	231	0.561	7	Feingold et al 2005
STM0028	12	2.35	28	70	42	0.719	7	Milbourne et al 1998
STM0052	22	1.47	54	94	53	0.802	7	Milbourne et al 1998
SSR1	14	2.70	98	168	16	0.82	8	Kawchuk et al. 1996
STGBSS	11	2.32	27	75	35	0.742	8	Ghislain et al 2004
STI003	19	2.21	50	88	53	0.692	8	Feingold et al 2005
STI022	8	1.93	13	45	49	0.765	8	Feingold et al 2005
STM0024	13	1.46	36	65	196	0.551	8	Milbourne et al 1998
STM1001	9	1.71	35	73	90	0.731	8	Milbourne et al 1998
STM1005	8	1.08	11	25	219	0.45	8	Milbourne et al 1998
STM1016	7	2.28	19	65	41	0.736	8	Milbourne et al 1998
STM1024	10	1.93	18	49	93	0.56	8	Milbourne et al 1998
STM1055	10	1.73	19	39	120	0.683	8	Milbourne et al 1998
STM1057	7	1.80	9	45	118	0.619	8	Milbourne et al 1998
STM1104	8	1.73	19	47	117	0.597	8	Milbourne et al 1998
STM1105	14	2.54	86	151	25	0.8	8	Milbourne et al 1998
STM3010	7	1.60	9	30	71	0.658	8	Milbourne et al 1998
STM3015	20	1.89	81	132	88	0.761	8	Milbourne et al 1998
STSS1	14	2.64	73	132	29	0.777	8	Kawchuk et al. 1996
STWAX1	10	1.82	22	58	116	0.606	8	Kawchuk et al. 1996
STWAX2	18	2.45	79	131	30	0.783	8	Ghislain et al 2004
STM1051	20	1.86	73	121	109	0.662	9	Milbourne et al 1998
STM0051	6	1.69	8	21	132	0.498	10	Milbourne et al 1998
STM1106	19	1.23	29	57	211	0.582	10	Milbourne et al 1998
STM2012	19	1.72	46	92	42	0.79	10	Milbourne et al 1998
STI018	8	2.25	34	66	90	0.663	11	Feingold et al 2005
STI028	13	1.88	34	67	110	0.535	11	Feingold et al 2005
STM2005	13	2.41	30	82	29	0.823	11	Milbourne et al 1998
STM0003	16	2.00	59	116	24	0.817	12	Milbourne et al 1998
STM2028	15	1.66	38	74	78	0.667	12	Milbourne et al 1998
Average	**12.32**	**1.89**	**36.89**	**74.45**		**0.663**		

Only microsatellite markers which could be scored reliably in a co-dominant way were retained. A total of 653 alleles were detected with these 53 microsatellites when scored on 430 genotypes, with an average of somewhat more than 12 alleles per locus ranging from two for STM3020 to 26 for STM1100 and an average of somewhat less than two alleles per genotype. The number of unique genotypes per locus was 37 and the total number of genotypes 75 on average. PIC-values based on allele frequencies ranged from 0.003 for STM3020 to 0.865 for STM5148 and were on average 0.663 (**Table 1**). Presence of null alleles was obvious for all but one locus, namely STM3020, although in view of the low number of detected alleles for this locus and knowing the highly heterogeneous nature of potato, presence of null alleles in that locus is very likely.

Population structure

The results obtained with STRUCTURE 2.1 analysis using the admixture model and assuming independent allele frequencies showed in all cases a continuous increase of the goodness of fit statistic, Ln[P(D)], versus the number of groups, K, increasing from two to 20 (see **Figure 2**) indicating that no optimal solution could be found. From the use of the subset of 37 centromeric AFLP markers from different haplotypes, no obvious groups could be inferred since genotypes never had a group membership probability exceeding 0.75. The subset of 48 telomeric AFLP markers, chosen to represent distal independent markers, revealed no clear groups either. Again only very few genotypes had membership probabilities higher than 75% and there were no apparent characteristics linking them.

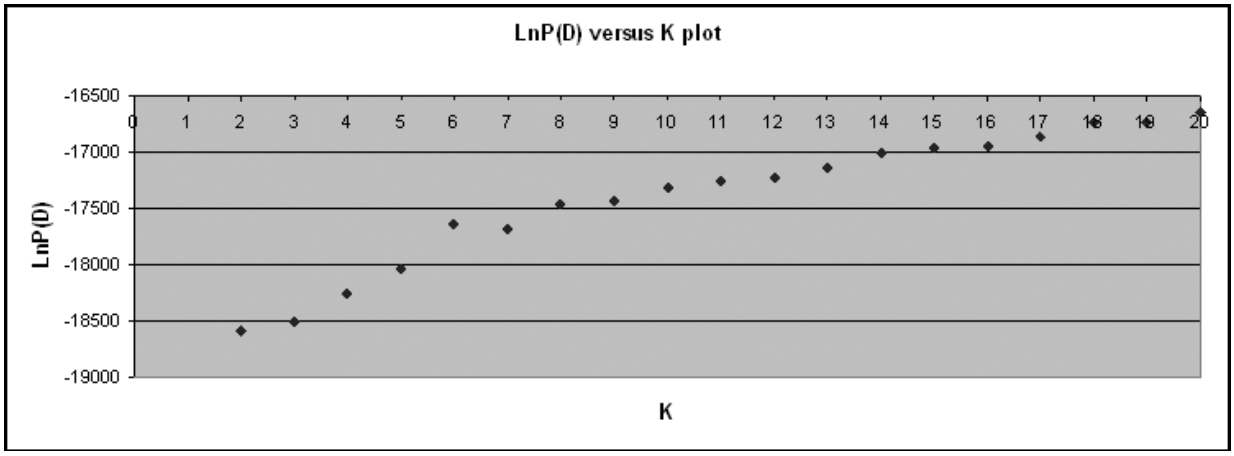


Figure 2: Example of a typical goodness of fit, Ln[P(D)], versus number of groups, K, plot for our data. The graph represents the resulting plot obtained with the AFLP data set including 103 5cM spaced markers. K is the number of assumed clusters for a run with STRUCTURE, starting with two and Ln[P(D)] is the logarithm of the likelihood averaged over all iterations since the end of the burn-in, i.e. over 100,000 iterations in our case.

However, using a larger set of 315 AFLP markers, more obvious and plausible groups appeared. Based on group characteristics and group membership probabilities, the six group solution seemed most adequate. Therefore we decided to perform another analysis with STRUCTURE using 103 equidistantly spaced, at approximately every 5cM, AFLP markers to prevent bias due to linkage disequilibrium between markers without diluting the marker set too much, to see whether the same groups reappeared. In **Figure 3**, the result of this analysis is depicted. Again six groups seemed the most plausible solution, as too few groups did not result in consistent groups: some consistent groups like Starch, Processing and Ancient already

appeared at a lower number of hypothetical groups (K), but members were swapped between repeats of the STRUCTURE runs. Larger number of groups resulted in too high amounts of admixture affecting group membership probabilities. We named the eventual groups in the six groups solution SH, Ancient, Processing, Starch, Fresh and Rest group.

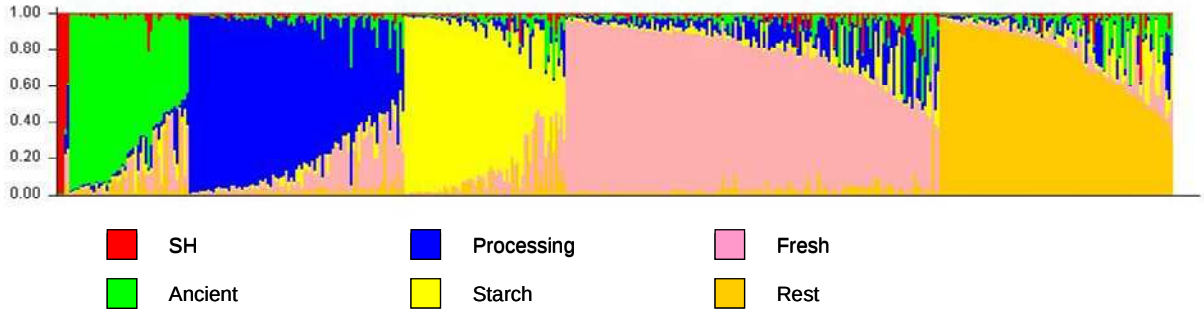


Figure 3: Bar plot of individual potato cultivars generated by STRUCTURE 2.1 using the admixture model with independent allele frequencies, based on 103 AFLP markers, spaced every 5cM on the ultra dense potato map (van Os et al. 2006). Groups are represented by colours, as indicated at the bottom. Each column (430 in total) represents a cultivar's genotype and is partitioned into segments indicating its genetic composition, the longer a segment the more a genotype resembles one of the inferred six groups.

The SH group included only the three repeats of diploid mapping parent SH83-92-488. The Ancient group comprised cultivars released before 1950 and these were mostly from the UK. The Processing group included cultivars related to Agria, a frying cultivar widely used as crossing parent, as well as having an above average underwater weight: cultivars with a Processing group membership exceeding 60% had an average underwater weight exceeding 400 grams, which is typical for frying cultivars. The Starch group included mostly those cultivars that have been specifically bred for the starch processing market, moreover their average underwater weight is 471 grams. The Fresh group is a rather large group of mainly European cultivars registered later than 1950, used for fresh consumption and having an average underwater weight of 371 grams. The Rest group is comprised of a few progenitor clones, the diploid mapping parent RH89-039-16 and miscellaneous European cultivars.

The population structure analysis with 53 microsatellite loci did only partly support the results of the AFLP structure analysis. Although the groups Starch and Ancient persisted until the assumption of five underlying clusters, the other clusters found previously with the AFLPs were not consistent enough to be considered as a group. A possible explanation for this inconsistency is that the majority of microsatellites were developed based on expressed sequences which implies that they only reflect a small part of the genome.

DARWIN analysis resulted in all five cases in the recognition of clades comprising the majority of the cultivars previously assigned to the groups Ancient and Starch by STRUCTURE. Three DARWIN analyses, using the largest number of AFLP markers, showed a clustering of recently bred Dutch material for fresh consumption market, while in the analyses with a more diluted marker set this cluster split up into several parts. In contrast to the clear visually observable clusters, there was only a weak statistical support for these results. **Figure 4** shows the Ward tree obtained with the entire AFLP data set comprising 3364 markers.

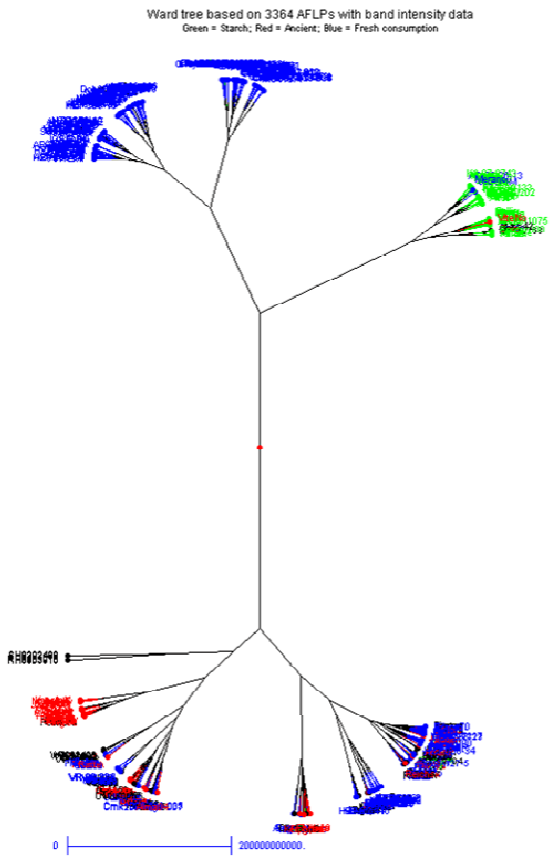


Figure 4: Ward tree created with DARWIN 5 based on 3364 AFLP fragments using normalised log transformed band intensities. Labels are abbreviated cultivar names. The exact name however is of no importance here, it is their colour according to their characteristics that deserves attention. Green coloured genotypes are bred for starch processing industry, red coloured genotypes are registered before 1950, blue coloured genotypes are bred for table consumption. Some groups that were also detected with STRUCTURE (Figure 3) reappear in this tree. The blue group on the top left contains recently bred consumption clones, the green group on the top right contains mainly starch cultivars and in the bottom part there is a group containing mainly old cultivars. The length of the branches is not representative for genetic distances between cultivars, the tree merely shows the branching pattern.

Assessment of Linkage Disequilibrium

AFLP markers were tested for linkage disequilibrium by using the LD statistic r^2. Markers belonging to the same linkage group were tested for LD, which resulted for each of the 12 linkage groups in an overview of LD decay, as shown in **Figure 5**. The LD decay is based on log-transformed band intensities of 720 mapped AFLP markers. The general trend observed across the 12 linkage groups suggests that LD decays below the 0.1 threshold when the genetic distance exceeds 5cM. On several chromosomes secondary peaks of significant LD are visible beyond 5cM. This is an artefact mainly caused by the difference in the genetic position of the centromere on the parental linkage groups of the ultra dense map. The difference between centromeric positions is caused by parent specific differences in recombination frequencies. A disproportionally large number of AFLP markers map to centromeric marker clusters. In spite of their localisation on independent haplotypes from either a maternal or paternal centromeric cluster strong LD values have been observed between such centromeric markers. The position of the secondary peak coincides exactly with the difference between the parental centromeric positions; i.e. 15, 2, 5, 2, 2, 9, 2, 7, 8, 12, 11 and 5cM for chromosome *1-12*, respectively. For chromosome *1* our explanation of the secondary peak was confirmed (**Figure 6**). The number of markers in this linkage group was sufficiently large to generate separate LD decay plots. These separate plots show only LD between markers that retrieved their genetic position from our reference map from either the paternal or the maternal linkage group. The secondary peak at 15cM indeed disappeared, but on the other hand some minor secondary peaks remained (e.g. at 20cM). Those minor secondary peaks may as well be due to marker clustering (cold spots for recombination) along haplotypes at a few cM distance.

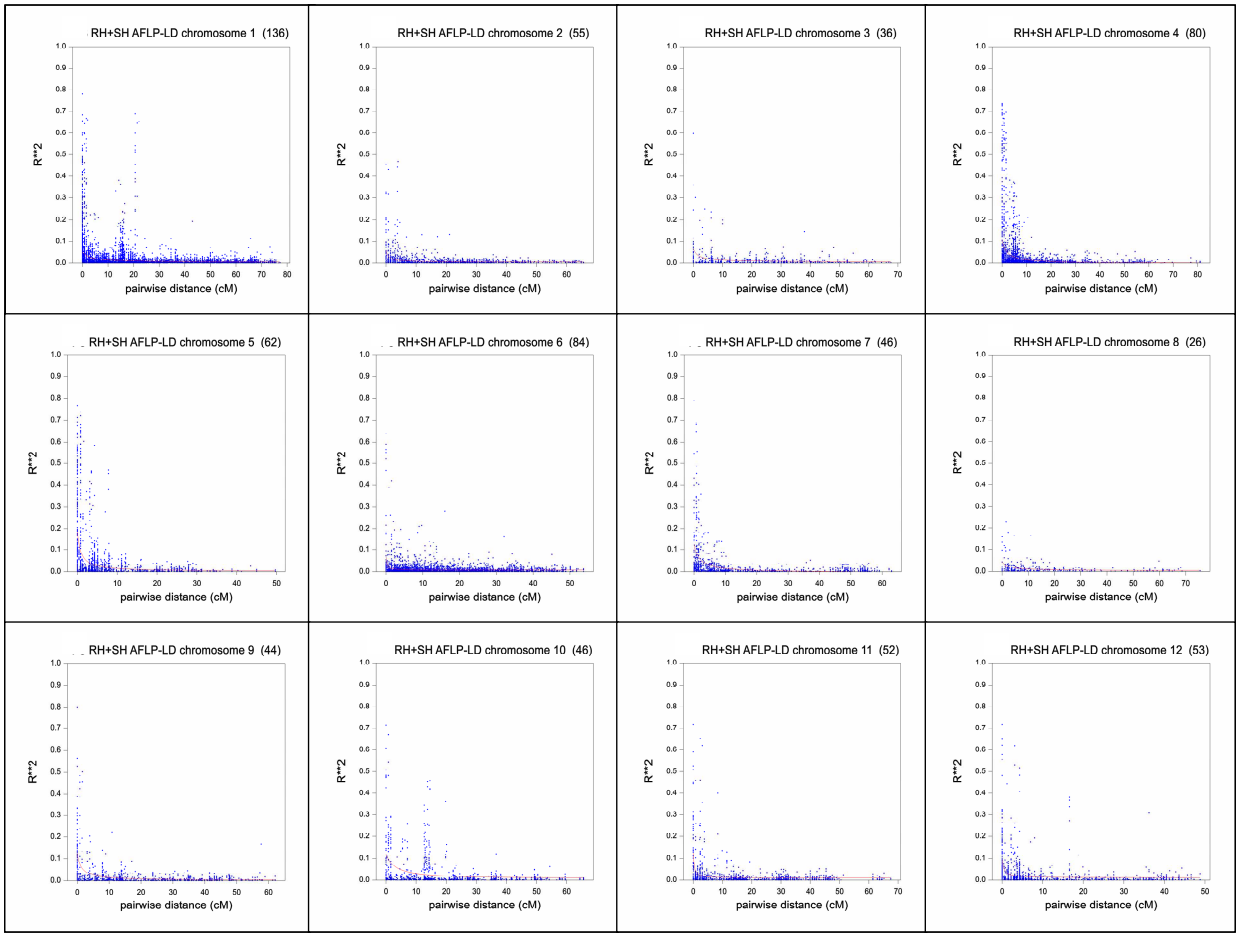


Figure 5: Overview of linkage disequilibrium (LD) decay across all 12 potato chromosomes based on 720 AFLP markers using normalised log-transformed band intensities. As LD measure r^2 has been used. Marker map positions in cM were deduced from the ultra dense potato map (van Os et al. 2006). Each plot represents the LD pattern of one chromosome, between brackets the amount of markers of a particular chromosome is shown.

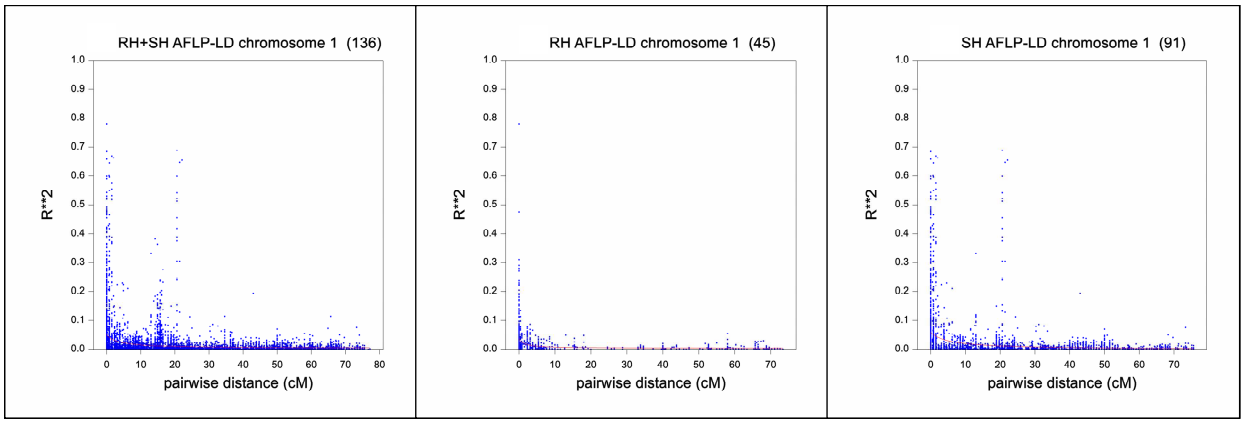


Figure 6: Illustration of the effect of differential parental recombination on the LD decay plot of chromosome *1*. On the left, the decay plot is shown for all markers combined, in the middle the decay plot is shown for markers only residing on the paternal map (RH) and on the right the decay plot is shown for markers only residing on the maternal map (SH).

LD between the 53 microsatellite markers as calculated with our modified statistic for r^2 (see M&M) resulted in a heat map of the LD pattern along the potato genome, and is shown in **Figure 7**. The microsatellites are sorted according to their order on the chromosomes. No

common cM positions can be given as the microsatellites were mapped in different diploid mapping populations. It can be deduced from the pattern that there is a block of persisting LD present on chromosome *8*. This was expected as these microsatellites reside in or in the vicinity of the Granule Bound Starch Synthase (GBSS) locus. On chromosome *8* we deliberately elevated the number of SSR loci, even including multiple GBSS SSRs within a few kb physical distance, as a positive control to demonstrate that our LD statistic and our scoring method of microsatellites works as expected. This block of high LD values does not violate our previous estimation of LD decay based on AFLP in potato as these SSRs are all localised within a 5cM genetic interval. There are no other blocks of LD distinguishable in this marker set, which is according to our expectations. The distant localisation of the microsatellite loci was chosen to uniformly cover the genome, mapping at genetically independent loci.

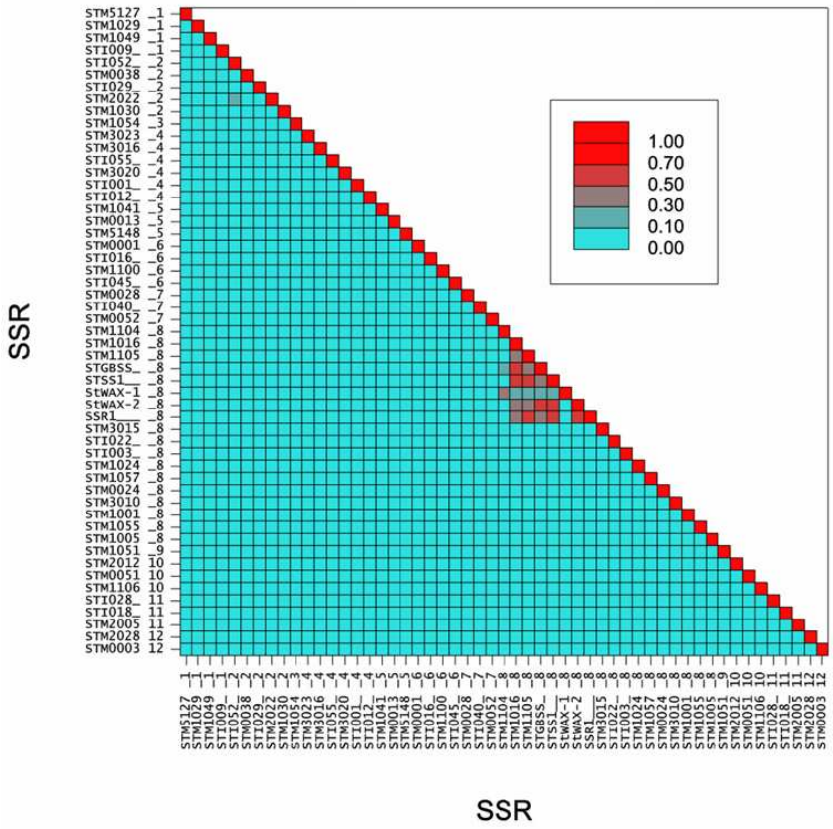


Figure 7: Heat map of the LD pattern along the potato genome based on genotypic data of 53 microsatellites. The LD measure r^2 was used to construct the map. Labels consist of the name of the microsatellite followed by its chromosome. Markers are ordered according to their relative position on a chromosome. A continuous blue-to-red colour spectrum indicates the intensity of LD between loci.

Discussion

AFLP

Although AFLP markers have several disadvantages when applied in diversity and association mapping studies in polyploids, such as their dominant inheritance and the risk of homoplasy (Woodhead et al. 2005; Zhu et al. 2008), the major advantage of AFLP fingerprinting of potato cultivars is the high multiplex ratio. This multiplex ratio of ~80 markers per assay, results per data point in the most cost effective method, when compared to any other currently available system to detect genetic variation in potato (McGregor et al. 2000; Meudt and Clarke 2007). We aim to investigate marker-trait associations within cultivated potato (*Solanum tuberosum* spp. *tuberosum*), which is a tetraploid crop species with a high level of heterozygosity. To study intraspecific genetic diversity, AFLP might even be preferred, as there are only four to ten times more dominant AFLP loci required to gain the same efficiency as with co-dominant markers and

given the high heterogeneity of potato more frequent low-informative marker loci are likely more efficient than less frequent but highly informative marker loci (Mariette et al. 2002).

In this study, we worked with qualitative absence/presence AFLP polymorphisms, as well as with quantitative log-transformed band intensities, which allowed to compare both data types when inference of population structure is at stake. Where qualitatively scored AFLP data will have a certain error rate due to misclassification by the scoring software, the quantitative intensity values are not biased by these classification errors. In D'hoop et al. (2008), the authors explain why they opted for band intensities for the estimation of LD and marker-trait associations. Referring to the comparable results obtained when analysing population structure within potato germplasm applying DARWIN to these two types of AFLP data, there is no reason why not to opt for band intensities. Moreover, there is no apparent loss in accuracy at stake when trying to exploit quantitative information from AFLP band intensities. In the worst case, the quantitative method will perform equally well as the qualitative (presence/absence) method (Piepho 2001).

SSR

Within potato a lot of work has been done to develop a wide range of applicable microsatellite markers. Pioneering work in potato was performed in 1996 through the development of the first 14 polymorphic SSR primer pairs within potato (Provan et al. 1996) and through the investigation on the abundance, length, composition and stability of 40 detected SSRs in potato and the application of three site-specific microsatellites with success for tetraploid potato cultivar identification (Kawchuk et al. 1996). Following this, a much larger set of microsatellite markers has been developed and subsequently mapped to diploid mapping populations (Ghislain et al. 1999; Feingold et al. 2005; Milbourne et al. 1998). In our study we applied 53 of these previously developed microsatellite markers to a diverse set of 430 tetraploid potato cultivars, where we intended, considering our plan to perform association mapping with these markers in follow-up research, to score these SSRs in a co-dominant way. We detected in total 653 alleles with these 53 microsatellites, with an average of 12 alleles per locus. PIC values ranged from 0.450 to 0.865, ignoring the two low extremes—STM3020 and STM1054 (**Table 1**). For comparison, Ghislain and colleagues (1999) obtained with their microsatellite diversity study performed with 156 SSRs on ~930 cultivated potato accessions belonging to eight taxonomic groups an allele number ranging from three to 27 and PIC values ranging from 0.280 to 0.892. Even though we only intended to use PIC values in tetraploid potato as a means for rough comparison, this does indicate that the gene pool of tetraploid potato has not been narrowed due to the efforts of commercial breeding, as sometimes suggested (Pavek and Corsini 2001). Furthermore, these results suggest that potato breeders are not likely to suffer from lack of genetic diversity in their future breeding efforts.

To obtain co-dominant microsatellite data we adopted the MAC-PR methodology described by Esselink et al. (2004). This scoring method could be applied to all microsatellite loci tested, albeit several times quite cumbersome and time-consuming. Problems encountered during analysis of electropherograms were e.g. differential PCR amplification efficiency of certain alleles and decreasing fluorescence signal intensity, i.e. fade-out, with larger sized alleles (Kimpton et al. 1993; Suenaga and Nakamura 2004). These observations were taken into account while converting the peak area into allele zygosity. We did not attempt to multiply ratios or peaks with correction factors. We also did not investigate any made assumption of null allele

presence by developing a new PCR primer based on sequence information of cloned microsatellite loci as has been done by Esselink et al. (2004) for one microsatellite. Instead we declared null alleles in cases with obvious evidence for their presence and classified them into a new allelic class “0”. Considering the heterogeneous nature of potato, we do acknowledge that the fitted genotypic model when only one allele is detected (quadruplex presence) or when there is a 1:1 peak area ratio between two alleles (duplex-duplex) is not unambiguous since the possibility exists of non-detected null alleles. However when no obvious proof of presence was available absence of null alleles was maintained as genotypic model assumption. In the cases where we did detect and score null alleles, we acknowledge that it is likely that obvious null alleles are not all of the same size and should therefore not be classified within one single allelic class. However, for convenience in subsequent statistical analyses we decided to lump null alleles for a certain SSR locus into one single allele “0”.

Population structure

In this study we describe a population structure analysis performed with two different approaches, Bayesian and Neighbour Joining, within the commercial potato gene pool using AFLP and SSR marker loci. With the Bayesian approach, using the program STRUCTURE (Pritchard et al. 2000), we observed continuously increasing goodness of fit, Ln[P(D)], for increasing number of groups, K, when applied to all sets of markers, indicating that no optimal solution could be identified (**Figure 2**). This is in agreement with what previously has been published concerning population structure in potato germplasm. Simko et al. (2004b) inspected without success a set of 150 tetraploid North American potato cultivars for population structure using segregating bands from 27 ISSR and RAPD markers. Simko et al. (2006) investigated 66 DNA fragments from 47 accessions for evidence of population structure using the Bayesian approach and reported no significant overall substructure, but did report that individual fragments in proximity of an introgression locus for resistance did reveal subgroup presence. Gebhardt et al. (2004) reported about the lack of meiotic generations separating cultivars from one another in a set of 600 potato cultivars and the high amount of close relationships between them pointing at a lack of population structure.

We investigated the possible presence of population structure using several (sub)sets of AFLP markers differing in their amount of markers and/or their spacing, and a set of microsatellites to ensure that we assessed a sufficient number of analyses to support any inference on structure before embarking into association mapping analysis. We first analysed a set of 315 AFLP markers, the largest set we could obtain with our data, with STRUCTURE, which pointed at six possible subgroups in our germplasm of which five allowed consistent interpretation. However the number of genotypes with membership probabilities exceeding 0.7 was rather low. Therefore, to ascertain our findings, we decided to run STRUCTURE using a subset of the original 315 markers, containing 103 markers, spaced approximately every 5cM to avoid disturbance of LD between markers (see **Figure 5**), without diluting the marker set too much. Again six subgroups seemed the most plausible solution, this time with a higher amount of genotypes with high cluster membership probabilities. In the series of STRUCTURE runs performed with this marker set, some of these subgroups already appeared at stages as early as K=3, and all the repeats performed at step K=6 revealed exactly the same subgroup composition. Going beyond K=6 diffused the subgroups to such an extent that no logical group inference could be made anymore.

Continuing on the same hypothesis that lowering the number of markers would induce clearer group assignments because of less loci being dependent led us to two even smaller marker subsets, one with 37 centromeric AFLP markers and one with 48 telomeric AFLP markers. Both of these were chosen to represent a set with truly independent markers. We assume the centromeric markers to be haplotype independent as they are close to the centromere regions where recombination is suppressed and because they belong to different linkage phases according to the SHxRH ultra dense genetic map of potato (van Os et al. 2006). We assume the telomeric markers to be independent because they reside on the distal ends of each of the chromosomes, physically and genetically far enough from each other to assume linkage equilibrium. Unfortunately, no groups could be inferred using any of these marker subsets. Group membership probabilities were almost never higher than 0.75. These marker sets were thus not informative for consistent group identification and assignment.

The structure analysis with the 53 microsatellites supported the results of the AFLP analysis to some extent. The Starch and Ancient subgroups reappeared; the other subgroups found previously with the AFLPs were not consistent. A possible explanation for this lack of support is that the majority of microsatellites were developed based on expressed sequence tags which implies that they only reflect a small part of the genome, whereas the AFLPs are more randomly spread throughout the genome (**Figure 1**).

DARWIN supported in all five analyses the findings from STRUCTURE, revealing clusters containing the majority of the Starch group as well as the majority of the Ancient group. In three analyses, including the analysis based on the highest number of markers, recently bred Dutch material for the fresh consumption market was clustering together (see **Figure 4**), while in the analyses with a less dense marker set this cluster split up into several parts. In contrast to these visually distinguishable trends, there was only weak statistical support for these clustering results.

The Bayesian as well as the Neighbour Joining approach to population structure discovery indicate that breeding germplasm intended for starch processing industry differs from other breeding germplasm, that recently bred material diverges to some extent from older, more alike, potato germplasm and that the majority of the companies breeding for the fresh consumption and processing market extract their cultivars from the same potato gene pool. These findings suggest that to be able to perform reliable association mapping, i.e. with appropriate control for false positives, it will be necessary to account for the relationships between the genotypes, where the marker information can be used to estimate those relationships.

LD

We unravelled the LD pattern within the potato genome with both an ample AFLP marker set and a set of SSR loci. We deployed the squared correlation r^2 as LD measure, thereby following a similar approach as Flint-Garcia et al. (2003) and Zhao et al. (2005). When applying the generally used cut-off value of 0.1 as detection limit for linkage disequilibrium, we found LD to decay on average at about 5cM throughout the potato genome (**Figure 5**). The pattern revealed by the microsatellite information did not contradict these findings as no LD was detected between any two loci except when residing in close proximity, i.e. less than 5cM apart from each other (**Figure 7**). Our findings hold the middle between previous findings: Gebhardt et al. (2004) examined four markers within 1cM on chromosome *5*. For two markers residing within

0.3cM LD was maintained, whereas for markers being separated by 0.6cM LD had decreased and for markers being 0.9cM apart linkage equilibrium had been reached. Simko et al. (2006) concluded from their data that LD decayed below 0.1 at about 10cM. D'hoop et al. (2008) estimated using a set of ~150 AFLP markers scattered throughout the genome an LD decaying below 0.1 at a distance of 3cM.

An interesting LD pattern can be distinguished on several chromosomes: there seems to be significant LD beyond a distance of 5cM (**Figure 5**). This is an artefact caused by mutual action of differential recombination between the parental genotypes of the diploid SHxRH mapping population and the typical feature of AFLP markers to cluster into centromere regions. Due to differential recombination, centromere regions of the two parental maps have been assigned different BIN positions during the construction of the genetic map (**Fig.2** in van Os et al. 2006). It is likely that quite some AFLP markers in our data set do reside in these centromere regions which results in such kind of LD artefacts, as illustrated for chromosome *1*, where the initial marker set has been split into a separate set for each of the parental maps (**Figure 6**). The remaining unexplained pair-wise marker combinations exhibiting higher LD are likely to be due to long distance covering haplotype blocks in potato which could have been detected in our data set as there are several cold spots of recombination where markers cluster on chromosomes (van Os et al. 2006).

As a measure for LD between two microsatellite loci, we adapted the method of Flajoulot et al. (2005) to calculate r^2 based on the most frequent allele, to enable the use of zygosity information. Flajoulot et al. (2005) tested their LD measure using one SSR on each chromosome thereby illustrating that their LD measure was not likely to detect LD between unlinked loci. Our adapted method also did not discover LD between loci on different chromosomes and it did detect LD between closely mapped SSR loci, as expected.

Prospects

The findings described within this paper indicate that there is, although rather limited, evidence of population structure within the potato gene pool suggesting that correction for population substructure within association mapping research in potato will be necessary in order to obtain linkage-based associations between phenotypic trait loci and genotypic marker loci. With the inclusion of structure information and/or marker-inferred kinship relationships to correct for spurious associations, we hope that we can detect valuable associated markers, that may be useful for marker assisted breeding in potato. Referring to the obtained patterns for LD decay throughout the potato genome, a marker density of at least one marker every 5 to 10cM seems advisable.

Acknowledgements

We express our gratitude to the potato breeding companies and gene banks that were willing to provide us with tubers of potato cultivars. It concerns Agrico Research (Netherlands), Averis seeds (Netherlands), C Meijer (Netherlands), HZPC Research (Netherlands) and Van Rijn (Netherlands), Agriculture and Agri-Food (Canada), Arche Noah (Austria), IPK (Germany), INRA (France), SASA (Scotland), Teagasc (Ireland) and USDA (USA). Our special thanks go to Paul Keizer for his artistic creations within GenStat. The help of Brigitte Uwimana and Hanneke van der Schoot in SSR data generation was highly appreciated.

This study was financed by the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

Chapter 5

Marker-trait associations in potato

Björn B. D'hoop
Paul L.C. Keizer
M. João Paulo
Richard G.F. Visser
Fred A. van Eeuwijk
Herman J. van Eck

Abstract

A genome-wide association mapping study was conducted using two association mapping panels. The first panel contained 205 historical and contemporary potato cultivars that were phenotyped in a field trial at two locations with two replicates each in 2006. The second panel consisted of 299 potato cultivars and combined five times 38 recent breeds obtained from five Dutch potato breeding companies and some reference cultivars. Phenotypic data for the second panel was generated by breeders during clonal selection programs at the individual breeding companies. Interesting QTL could be identified for 19 agro-morphological and quality traits. Two association mapping models were used: one baseline model without correction for population structure or genetic relatedness, and a second model correcting for population structure and genetic relatedness. Genetic relatedness was estimated from marker information. The identified QTL partly confirmed previous studies e.g. for tuber shape and frying colour, but also new QTL were found like for after baking darkening or enzymatic browning.

Keywords: association mapping, linkage disequilibrium, potato, quality

Introduction

Potato is according to FAOSTAT (www.faostat.org) the third most important food crop in the world on a per capita per year base. Due to increasingly stringent environmental and consumer related demands, potato breeders are stimulated to continuously produce new cultivars (Bradshaw 2007). Potato breeding methodology has not changed dramatically over the years though, as it mainly consists of making crosses between two complementary parental clones, thereby creating sufficiently diverse offspring. Several years of clonal selection aim to identify candidate cultivars. In recent years, developments in genetics and DNA technology have led to numerous cost effective methods to create large data sets of molecular markers for the construction of dense genetic linkage maps (van Os et al. 2006) or genome-wide marker scans using unstructured populations. Consequently, breeders look forward to adopt marker assisted breeding strategies in progenitor development and introgression breeding. Furthermore parent choice for crossing programs can be optimized, not only on the basis of phenotypically complementary characteristics but also based on genetic information (Bradshaw 2007; Simko 2004a). To assure the relevance of marker information, newly emerged methods like association mapping should be adopted, because this approach can cope with the diversity of the tetraploid gene pool.

Association mapping is generally defined as the detection of statistically significant associations between phenotypic trait loci on the one hand and molecular marker loci on the other hand in populations of complex genetic architecture (Gupta et al. 2005). The technique was initially developed to overcome limitations inherent to traditional family-based disease linkage studies in humans (Carlson et al. 2004). Association mapping proved more powerful to detect common disease loci through the recruitment of unrelated individuals which simultaneously led to narrower disease-associated regions, which is an extra advantage. Since its initial development, numerous successes have been achieved in a wide variety of genetic disorders, e.g. Alzheimer's disease (Corder et al. 1994) and diabetes (Scott et al. 2007), a recent overview of successes has been presented by Slatkin (2008).

In plant breeding research, association mapping has equally demonstrated its promises. Pioneering association studies have been performed in model plant species such as *Arabidopsis* (Aranzana et al. 2005; Nordborg et al. 2002; Zhao et al. 2007), maize (Remington et al. 2001; Thornsberry et al. 2001; Yu et al. 2006) and rice (Lu et al. 2005; Mather et al. 2007). More recently, the focus moved from model organisms to crop species as soon as the molecular and statistical requirements had been fulfilled (Zhu et al. 2008). Association mapping has resulted in the identification of quantitative trait loci in several crop species, e.g. barley (Caldwell et al. 2006; Cockram et al. 2008; Kraakman et al. 2004), *Brassica rapa* (Zhao et al. 2007), potato (D'hoop et al. 2008; Malosetti et al. 2007; Simko et al. 2004a, 2006), soybean (Wang et al. 2008), sugar beet (Kraft et al. 2000; Stich et al. 2008b, 2008c), sugarcane (Raboin et al. 2008), sunflower (Fusari et al. 2008; Kolkman et al. 2007) and wheat (Breseghello and Sorrells 2006a; Maccaferri et al. 2005; Stich et al. 2008a). The number of successes achieved clearly indicates that association mapping will become equally important as conventional QTL analysis for the development of marker assisted selection or breeding (Ersoz et al. 2007).

Compared to traditional linkage-based QTL mapping, association mapping has several promising advantages. First, a higher number of meiotic recombination events are sampled by

adopting distantly related individuals, which can lead to an increased mapping resolution (Flint-Garcia et al. 2003; Jannink and Walsh 2002; Nordborg and Tavare 2002). Second, the results obtained are representative and applicable to a much wider germplasm (Zhu et al. 2008). Association mapping also has disadvantages, such as false positive associations which are not necessarily caused by true genetic linkage. Linkage disequilibrium (LD), the phenomenon underlying both traditional QTL mapping and association mapping, is affected by several factors like recombination and allele frequencies. Consequently, association mapping results should be interpreted with care as processes such as genetic drift, admixture and selection could influence the extent of LD between two loci. Therefore these processes should be taken into account during association mapping analysis, because they can increase the number of false positives detected (Flint-Garcia et al. 2003; Gupta et al. 2005). For that reason, a good estimation of hidden population (sub)structure and genetic relatedness is essential for association mapping analyses to be reliable (Breseghello and Sorrells 2006b; Jannink and Walsh 2002; Mackay and Powell 2007; Zhu et al. 2008).

Association mapping research can be realised in a number of ways, depending on the interests of the researchers and the available data. One approach to association mapping research can be described as the candidate gene approach where a limited number of predefined alleles are tested for association with a phenotype, another approach can be defined as whole-genome association mapping or as genome scan, where the entire genome is surveyed for genetic variation associated with a phenotype (Zhu et al. 2008). Likewise, the possible problem of hidden population structure can be tackled by different strategies. Most popular in association mapping research is to inspect the genotype collection with the method described by Pritchard et al. (2000), where a Bayesian approach is used to assign group membership probabilities to genotypes using molecular marker information. An alternative is to use multivariate statistics approaches to calculate genetic distances from molecular marker information or pedigree information, which can subsequently serve as input for cluster analysis (Kraakman et al. 2004; Simko et al. 2004a) or in a direct approach to impose structure on the variance-covariance matrix between genotypes in a mixed model environment (Malosetti et al. 2007; Parisseaux and Bernardo 2004). Both strategies can be combined as well, as described by Yu et al. (2006). Similarly, phenotypic analysis can precede or can be incorporated in the association mapping analysis. In the first case phenotypic observations are analysed separately, succeeded by the estimation of adjusted genotypic means which subsequently serve as entry values for the actual association mapping analysis. This strategy is usually referred to as the two-step association mapping approach. When phenotypic analysis and association analysis are integrated into one single stage, this is termed a one-step association mapping approach (Stich et al. 2008a). The way we performed association mapping can be classified as a genome-wide two-step association mapping approach because we had access to a high number of random generated marker information scattered across the genome and because we computed adjusted genotypic means prior to association mapping analysis. Population structure was tackled in a straightforward way by imposing structure on the variance-covariance matrix using marker correlations.

Potato (*Solanum tuberosum* L.) is an autotetraploid crop species (2n=4x=48) demonstrating a tetrasomic inheritance (Bradshaw et al. 2008). Specifically this complication explains the modest progress in potato breeding as compared to less complex crop species. Potato genetics has for obvious reasons mainly been performed on a reduced ploidy level. Diploid mapping

populations have been the preferred study object for many traditional QTL mapping studies in potato so far, whether the objectives were to improve resistance (e.g. Collins et al. 1999; Visker et al. 2003) or quality traits (e.g. Douches and Freyre 1994; Menendez et al. 2002; Werij et al. 2007). In recent years, QTL mapping studies within a tetraploid mapping population have been reported as well (Bradshaw et al. 2004, 2008; Khu et al. 2008). Even though classical QTL mapping efforts, whether at the diploid or tetraploid level, have proven successful, it is likely that the use of tetraploid potato cultivars in association mapping offers interesting opportunities both through testing of detected QTL at the relevant ploidy level as within a suitable genetic background for the crop. The first examples exploiting tetraploid potato cultivar collections for association mapping have been published recently (Gebhardt et al. 2004; Li et al. 2005, 2008; Malosetti et al. 2007; Simko et al. 2004a,b). The majority of these association mapping studies followed a candidate gene based approach. In D'hoop et al. (2008) a first successful example of a genome-wide, albeit with a modest number of markers, association mapping approach has been presented. Hence, we anticipate association mapping to become an attractive and widely-used alternative to traditional QTL mapping for potato, especially when in the near future the Potato Genome Sequencing Consortium (PGSC) will release the genome sequence to the potato research community (http://www.potatogenome.net).

We present the results of a genome-wide two-stage association mapping approach in potato, using two differently structured phenotypic data sets and a large amount of molecular marker information. With an advanced mixed model strategy we were able to detect new promising associations as well as to confirm previously known associations for a variety of agro-morphological and quality traits.

Materials and methods

Plant material

To obtain a representative subset of worldwide available commercial potato germplasm, we collected in total 430 tetraploid potato cultivars, progenitor clones and advanced breeds. For further details about the composition and criteria used for selection, we refer to D'hoop et al. (2008) and Chapter 4. Tuber material was kindly provided by five Dutch breeding companies: Agrico Research (Bant, Netherlands), Averis seeds (Valthermond, Netherlands), C Meijer (Rilland, Netherlands), HZPC Research (Metslawier, Netherlands) and Van Rijn (Emmeloord, Netherlands), as well as several gene banks: Agriculture and Agri-Food (Canada), Arche Noah (Austria), IPK Gatersleben (Germany), INRA (France), SASA (Scotland), Teagasc (Ireland) and USDA (USA). Leaf material was harvested from greenhouse-grown and in-vitro-grown genotypes, was frozen with liquid nitrogen and stored at -80˚C until DNA extraction.

Phenotypic observations

Two partly overlapping association panels were compiled based on the above mentioned 430 genotypes. The first mapping panel is based on freely available historical and contemporary potato cultivars and their phenotypic values were obtained from two trial fields near Wageningen in 2006. We will refer to this dataset hereafter as the "2006 field trial". Secondly, a multi-year-multi-location (MYML) data set was contributed by five Dutch breeding companies. This MYML data set contains the data collected during clonal selection programs of in total 299 genotypes: 190 recent breeds (38 per breeding company), some of them have not entered the market yet, supplemented with additional observations from each breeding company on a series

of standard potato cultivars. There was a substantial overlap of 101 cultivars between the two association panels, illustrating that the cultivars from the first panel have frequently served as standards in the trials at breeding companies. For further details about the characteristics of and the specific differences between both the 2006 field trial and the MYML data set we refer to Chapter 3.

Agro-morphological and quality traits were assessed according to standard protocols of the breeding companies. We focused on the non-enzymatic discolouration due to oxidation of an iron-chlorogenic acid complex (McKenzie et al. 2005), observed after baking of French fries (After Baking Darkening ABD) or after cooking potatoes (After Cooking Darkening ACD), as well as on the Maillard discolouration of French fries due to reducing sugars (Frying Colour FryingCol). The traits ABD and FryingCol were tested after varying months and temperatures of storage of harvested material. Another discolouration type, enzymatic discolouration, caused by the enzymatic oxidation of phenols by the enzyme polyphenol-oxidase (Laerke et al. 2002) was recorded 30 minutes after grating of tubers (EnzBrow_30min). For tuber flesh colour we used a truncated scale starting from light yellow six to orange nine, thereby excluding the white–yellow contrast, controlled by the well-known *Y*-locus on chromosome *3* encoding the gene beta-carotene hydroxylase (Brown et al. 2006), which was not regarded as a relevant quality trait. As such, only the intensity of the yellow pigmentation was considered in an attempt to unravel the genetic control of tuber flesh yellowness, both in raw tuber flesh, as well as in cooked tuber tissue (FleshColY and CookColY). Furthermore, we evaluated Cooking Type, Maturity, tuber Shape and Size and starch content (Underwater Weight). An overview with description and scale of all the phenotyped traits was previously presented in Chapter 3.

Molecular marker analysis

DNA extraction was according to van der Beek et al. (1992). DNA quality and concentration were visually examined using ethidiumbromide stained 1% agarose gels.

AFLP™

AFLP markers were generated according to Vos et al. (1995) using 26 well-known *Eco*/*Mse* and 15 well-known *Pst*/*Mse* primer combinations. Specifications on which combinations were used and how AFLP marker information was generated and processed can be found in Chapter 4. Position information of AFLP markers was retrieved from the ultra dense potato map, using the parental diploid genotypes SH83-92-488 and RH89-039-16 as internal reference (van Os et al. 2006).

SSR

Microsatellite peak patterns were generated with 53 SSR primer pairs, previously reported in literature. For further reading about which specific SSR loci we selected, the protocol that was used to obtain peak patterns and the methodology deployed to convert peak patterns into reliable zygosity information, we suggest Chapter 4. Map positions were obtained from literature and, when microsatellites were polymorphic within the SHxRH population, from the ultra dense potato map, using the parental diploid genotypes SH83-92-488 and RH89-039-16 as internal reference (van Os et al. 2006).

Association analysis

Presence of population structure in our collection of potato cultivars was investigated prior to association mapping. Both a Bayesian modelling approach (STRUCTURE, Pritchard et al. 2000) and a hierarchical clustering approach (DARWIN, Perrier and Jacquemoud-Collet 2006) were used to make enquiries about hidden population structure. With the Bayesian approach a six group solution seemed the most adequate as consistent groups were obtained repeatedly. The clustering approach supported these findings to some extent: visually distinguishable clades were detected which allowed consistent interpretation. However in both cases statistical support was weak. For a more elaborate overview of our population structure analyses we refer to Chapter 4. Our findings indicated that certainly important genetic relationships do exist within our association panels and that we will need to take these relationships into account, one way or another, to avoid a large number of false positives, i.e. marker-trait associations not following from linkage but from population structure or other population genetic processes inducing genetic relatedness. We followed Malosetti et al. (2007), but also see Kang et al. (2008) and Zhao et al. (2007), in that a marker-based genetic relatedness matrix can effectively structure the variance-covariance matrix of the random genotypic effects, allowing a simple and powerful protection against false positive associations.

GenStat, 11th edition (VSN International Ltd., Oxford, UK) was used as environment for our association analyses. We opted for a two-step association mapping approach, where first adjusted genotypic means, i.e. Best Linear Unbiased Predictors (BLUP), were computed. For the 2006 field trial we calculated two sets of adjusted means, one set representing the average of the adjusted genotypic means across the two locations sand and clay, and another set representative for the difference between the adjusted genotypic means for sand and clay, intended to detect loci influenced by genotype-by-environment interaction. See Chapter 3 for further details about specific models used to generate these sets of adjusted means. In a second step, the actual association analysis, the previously calculated adjusted means are used as entry values. Three different marker sets were designed for association mapping: (1) a set of 720 AFLP markers with known map location, (2) the full set of 3364 AFLP markers and (3) a set of 53 microsatellite loci. For the AFLP marker sets normalised log-transformed band intensities were used, whereas for the SSRs allele dosages were exploited. Two association mapping models were constructed.

(1) A basic association mapping model, the baseline model, without correction for population structure or genetic relatedness:

$$\underline{\text{Response}} = \text{Marker} + \underline{\text{error}}$$

where Response refers to the set of adjusted means and Marker represents either AFLP or SSR marker information. Random terms are underlined.

(2) A more advanced association mapping model where we correct for population structure and genetic relatedness:

$$\underline{\text{Response}} = \text{Marker} + \underline{\text{Genotype}} + \underline{\text{error}}$$

Again, Response refers to the set of adjusted means and Marker represents or AFLP or SSR information. A random term for Genotype has been added to the model. The variance-covariance matrix for the genotypic effects was structured by a relatedness matrix calculated from the marker information. Random terms are underlined.

For both association mapping models we only consider the associations that were obtained with a P-value smaller than 0.001 or similarly with a -10logP bigger than three. We used the standard test for fixed effects in a mixed model context to assess significance for the marker-trait associations; a Wald test (Verbeke and Molenberghs 2000).

As mentioned previously, we used two panels differing in their genotypic composition for association mapping. Consequently, genetic relatedness correction was restricted to those genotypes included in each of the two association panels. In addition, two differing marker types were exploited for association mapping analysis, namely AFLPs and SSRs. One could thus advocate that genetic relatedness should be corrected in a different way for both marker types, e.g. by using the microsatellite marker information when SSRs are used and the AFLP information when AFLPs are used. We did try correcting for genetic relatedness with SSR information in case where SSRs were deployed for association analysis, however no reduction in number of detected associations was observed when compared to the baseline model (data not shown). Therefore we decided that for both marker types, the available genome-wide AFLP marker information would be used to account for genetic relatedness. Thus, the variance-covariance matrix of the genotypic effects in the more advanced model has been structured according to correlations as obtained between normalised log-transformed AFLP marker band intensities. The full set of 3364 AFLP markers was used for genetic relatedness correction, but it should be noted that 628 markers were effectively "constant" bands (see Chapter 4).

Results

Phenotypic analysis

During the 2006 field trial in total 15,580 data points were recorded for 19 phenotypic traits on 205 contemporary and historical potato cultivars. The MYML data set is even larger with more than 450,000 data points but had a highly unbalanced data structure due to many incomplete trait observations. A quantification of this highly unbalanced data structure is provided by the relative amount of about 77 % empty cells. High heritabilities, or better said, repeatabilities, have been obtained for both data sets for almost all traits. For more elaborate background information about the specific characteristics of the phenotypic data used in this study see Chapter 3.

Molecular marker analysis

In total 3364 AFLP fragments were distinguished on 430 genotypes, including a threefold repetition of the diploid mapping parents SH83-92-488 and RH89-039-16 which were included as internal reference genotypes. An average of 82 fragments was generated per primer combination. In total 1144 AFLP markers could be scored in a qualitative fashion (presence/absence). To permit use of the full number of 3364 fragments for association mapping quantitative band intensities were retained. For a justification of working with band intensities we refer to D'hoop et al. (2008). A subset of 720 polymorphic fragments could be assigned to a map location in the ultra dense potato map (van Os et al. 2006).

The application of 53 microsatellite loci to a set of 430 diverse potato cultivars led to the detection of 653 alleles, with on average somewhat more than 12 alleles per locus. To allow extraction of a maximal amount of information in an association mapping context, discrete allele

dosage data were pursued (0,1,2,3,4). See Chapter 4 for further background and a detailed description of obtained results concerning AFLP and SSR markers.

Population structure and linkage disequilibrium

As mentioned before STRUCTURE and DARWIN analyses were implemented using available AFLP and SSR marker information. Some characteristic groups of cultivars were repeatedly detected with both approaches. Cultivars bred for the starch industry as well as older cultivars and cultivars intended for the fresh consumption market or the processing industry were found, albeit with weak statistical support. Linkage disequilibrium in our cultivar collection was assessed by using the r^2 statistic as LD measure. We found indications for LD to decay below the 0.1 threshold at ~5cM on average in potato. For an elaborate clarification of population structure and LD in our potato collection, we make reference to Chapter 4.

Association analysis

With and without correction for genetic relatedness

With our basic association mapping model applied to the set of 720 AFLP markers with known map location, 158 and 428 associated loci were detected for 19 agro-morphological and quality traits of potato with the 2006 field trial and the MYML data set respectively. When a more advanced association mapping model correcting for genetic relatedness was used to analyse the same data sets, 43 and 24 associated loci persisted respectively (**Table 1**). The full marker set with 3364 AFLPs delivered 701 and 2875 marker-trait associations for the 2006 and the MYML data set correspondingly with the basic model. The model accounting for genotypic relatedness reduced these high numbers of obtained marker-trait associations down to 157 and 150 for the 2006 and MYML data respectively (**Table 2**). Similarly, the numbers of associated SSR loci were lowered from 15 and 99 for the designed and historical data set respectively to 11 and 13 when genetic relatedness correction was applied (**Table 3**). An identical reducing effect could be perceived when loci associated with genotype-by-environment interaction were concerned. This is illustrated in **Table 4** for the designed field trial data set and the set of 720 AFLP markers, comparable effects were observed for the full set of 3364 AFLP markers and the set of 53 SSR loci (data not shown). **Figure 1** displays the associated loci detected with the relatedness corrected association mapping model for both the 2006 field trial and the MYML data set along the potato genome. **Supplementary Figures 1** and **2** provide a per chromosome overview of associated loci between the 2006 field trial and the set of 720 AFLP markers, respectively without and with genotypic relatedness correction. An analogous survey is presented with **Supplementary Figures 3** and **4** where associated loci for the MYML data set are displayed. **Supplementary Figures 5** and **6** illustrate associations for the SSR markers and the 2006 field trial (5) or the MYML data set (6), without (5a and 6a) and with relatedness correction (5b and 6b).

Table 1: Results obtained with association mapping for the 2006 field trial and the MYML data set with the set of 720 AFLP markers with known map location. On the right an overview of literature reports is presented.

	Number of associated loci with p ≤ 0.001 for the 2006 trial: main effect QTL only			Number of associated loci with p ≤ 0.001 for the MYML data set			Diploid and tetraploid QTL mapping studies		Association mapping studies	
Trait	Basic model	Relatedness corrected	Chromosome nr (position cM)	Basic model	Relatedness corrected	Chromosome nr (position cM)	Chromosomes	Reference	Chromosomes (position cM)	Reference
ABD_Nov8c	6	1	4 (27.5cM)	26	1	10 (0cM)	--	--	1 (9.2, 24.9cM), 3, 4 (27.5, 30, 32.9cM), 6, 7 (55.3cM), 11	Chapter 2
ABD_Feb8c	5	1	3 (31.5cM)	7	0	--	--	--	1 (9.2, 24.9cM), 3, 4 (27.5, 30, 32.9cM), 6, 7 (55.3cM), 11	Chapter 2
ABD_Apr8c	10	5	3 (31.5cM), 4 (0cM), 9 (15.8cM), 11 (19.4, 44.6cM)	2	0	--	--	--	1 (9.2, 24.9cM), 3, 4 (27.5, 30, 32.9cM), 6, 7 (55.3cM), 11	Chapter 2
ABD_Apr4c	5	3	1 (67.2cM), 9 (59.6cM), 10 (65.6cM)	5	2	1 (24.9cM), 8 (6.9cM)	--	--	--	--
ABD_24hNov8c	5	1	4 (27.5cM)	10	2	1 (49.2cM), 6(37cM)	--	--	1 (9.2, 24.9cM), 3, 4 (27.5, 30, 32.9cM), 6, 7 (55.3cM), 11	Chapter 2
ACD_1h	7	2	3 (56.5cM), 11 (35.5cM)	22	0	--	1 (CIS), 2, 4, 6, 7 (ACO), 11	Bradshaw et al. (2008); Chen et al. (2001); Wang-Pruski and Nowak (2004)	1 (0.8, 24.9cM), 2, 4 (22.8cM), 5, 7 (54.5cM), 10	Chapter 2
ACD_24h	0	0	--	0	0	--	1 (CIS), 2, 4, 6, 7 (ACO), 11	Bradshaw et al. (2008); Chen et al. (2001); Wang-Pruski and Nowak (2004)	1 (0.8, 24.9cM), 2, 4 (22.8cM), 5, 7 (54.5cM), 10	Chapter 2
CookColY	20	3	1 (9.2cM), 3 (31.5cM), 8 (11.8cM)	9	0	--	--	--	--	--
Cooking Type	5	2	2 (2.3cM), 11 (35.5cM)	20	2	1 (9.2cM), 6 (22.9cM)	9 (StTLRP)	Kloosterman (2006)	1, 2, 6, 10	Chapter 2
EnzBrow_30min	4	3	5 (36.4, 37.2cM), 6 (10.7cM)	1	1	1 (56.5cM)	1 (ED-1), 3 (ED-3), 8 (ED-8a, ED-8b, PPO-III)	Bachem et al. (1994); Werij et al. (2007)	1, 4, 7, 8 (10.6cM), 11	Chapter 2

FleshColY	12	5	1 (8.8cM), 5 (8.5cM), 6 (38.5cM), 9 (16.6cM), 12 (8.8cM)	41	3	3 (31.5cM), 6 (13.7, 38.5cM)	3 (Y/y-locus), 4, 10, 12	Bonierbale et al. (1988); Jacobs et al. (1995); Sliwka et al. (2008)	1, 2, 3 (31.5cM), 4, 6, 7	Chapter 2
FryingCol_Nov8c	7	0	--	31	1	10 (57.6cM)	1(Sug1b), 2(Sug2a), 6, 7 (Sug7a,c,e, SPS), 11	Bradshaw et al. (2008); Menendez et al. (2002)	1 (4.6cM), 2 (3.7cM), 7 (63cM)	Chapter 2
FryingCol_Feb8c	9	2	2 (3.7, 8.4cM)	28	1	4 (0.0cM)	1(Sug1b), 2(Sug2a), 6, 7 (Sug7a,c,e, SPS), 11	Bradshaw et al. (2008); Menendez et al. (2002)	1 (4.6cM), 2 (3.7cM), 7 (63cM)	Chapter 2
FryingCol_Apr8c	9	7	1 (43.8cM), 2 (3.7cM), 4 (19.7cM), 5 (31.3cM), 6 (51.8cM), 9 (15.8, 16.6cM)	16	2	6 (35cM), 9 (15.8cM)	1(Sug1b), 2(Sug2a), 6, 7 (Sug7a,c,e, SPS), 11	Bradshaw et al. (2008); Menendez et al. (2002)	1 (4.6cM), 2 (3.7cM), 7 (63cM)	Chapter 2
FryingCol_Apr4c	5	1	9 (16.6cM)	13	5	1 (24.9cM), 6 (6.6cM), 7 (53.9cM), 9 (15.8cM), 12 (43.6cM)	1(Sug1b), 2(Sug2a), 6, 7 (Sug7a,c,e, SPS), 11	Bradshaw et al. (2008); Menendez et al. (2002)	--	--
Maturity	10	2	1 (74.9cM), 3 (44cM)	69	1	2 (7.5cM)	3, 5, 6, 12	Bradshaw et al. (2004; 2008); Malosetti et al. (2006); Visker et al. (2005)	1, 4, 5 (16.2cM), 6, 9, 10	Chapter 2; Gebhardt et al. (2004)
Shape	10	4	2 (3.7cM), 10 (0cM), 11 (3.8cM), 12 (39.8cM)	28	1	6 (51.8cM)	2, 5, 7, 10 (Ro/ro), 11, 12	Bradshaw et al. (2008); Sliwka et al. (2008); Sorensen et al. (2006); van Eck et al. (1994)	2, 4, 11	Chapter 2
Size	4	1	8 (75.5cM)	7	1	4 (15.9cM)	5	Bradshaw et al. (2008)	--	--
Underwater Weight	25	0	--	93	1	2 (62.8cM)	1 (ts(s)), 2, 4 (ts(i)), 5 (ts(a)), 9 (ts(o)),…	Gebhardt et al. (2005)	1 (8.4, 9.2, 24.9cM), 4 (21.3cM), 5 (16.2cM), 9 (46cM)	Chapter 2

Between phenotypic data sets

Comparison of obtained results for both the 2006 field trial and the MYML data set revealed a clear difference in total number of detected associations with a trend for the MYML data set to deliver more associations than the 2006 field trial data when considering the basic association mapping model. An opposite trend became apparent when the more complex model is concerned, except for the SSR loci (**Tables 1-3**). Most likely this effect is due to a concerted action of (1) the MYML data set harbouring a higher number of higher related cultivars relative to the 2006 field trial, see Chapter 4 and (2) a higher amount of phenotypic trait variation—at least for some traits—present in the MYML data set in comparison to the 2006 field trial (Chapter 3).

Table 2: Results observed when association mapping was performed with the full set of 3364 AFLPs.

	Number of associated markers with p ≤ 0.001 for the 2006 trial: main effect QTL only		Number of associated markers with p ≤ 0.001 for the MYML data set	
Trait	Basic model	Relatedness corrected	Basic model	Relatedness corrected
ABD_Nov8c	24	3	245	8
ABD_Feb8c	29	5	60	10
ABD_Apr8c	26	11	20	8
ABD_Apr4c	7	6	23	8
ABD_24hNov8c	38	10	45	13
ACD_1h	49	13	128	10
ACD_24h	9	8	3	2
CookColY	66	14	38	3
Cooking Type	38	12	103	7
EnzBrow_30min	14	7	2	3
FleshColY	28	12	248	10
FryingCol_Nov8c	27	5	206	7
FryingCol_Feb8c	48	5	185	9
FryingCol_Apr8c	33	14	137	6
FryingCol_Apr4c	24	5	75	15
Maturity	30	7	504	9
Shape	48	12	198	8
Size	11	4	21	3
Underwater Weight	152	4	634	11

When inspecting **Table 1** for matching associated loci between traits in both phenotypic data sets, it is striking that only very few identical loci can be discerned. Only for the traits FleshColY (locus on chromosome *6* 38.5cM), FryingCol_Apr8c (chromosome *9*: 15.8-16.6cM) and FryingCol_Apr4c (chromosome *9*: 15.8-16.6cM) there are some identical or closely residing loci detected. However this is due to the severe correction for relatedness: in **Table 1** only the loci that persisted when relatedness correction was applied are mentioned, and when inspecting the uncorrected models of both phenotypic data sets for the traits with the highest heritabilities in both data sets (Chapter 3), the majority of the associated loci detected for the 2006 data set also reappeared with the MYML data set albeit with varying P-values (**Supplementary Tables 1** and **2** respectively). For the SSRs such a comparison could not be drawn as in general very little associated loci were detected, while most of the AFLP were of unknown map position.

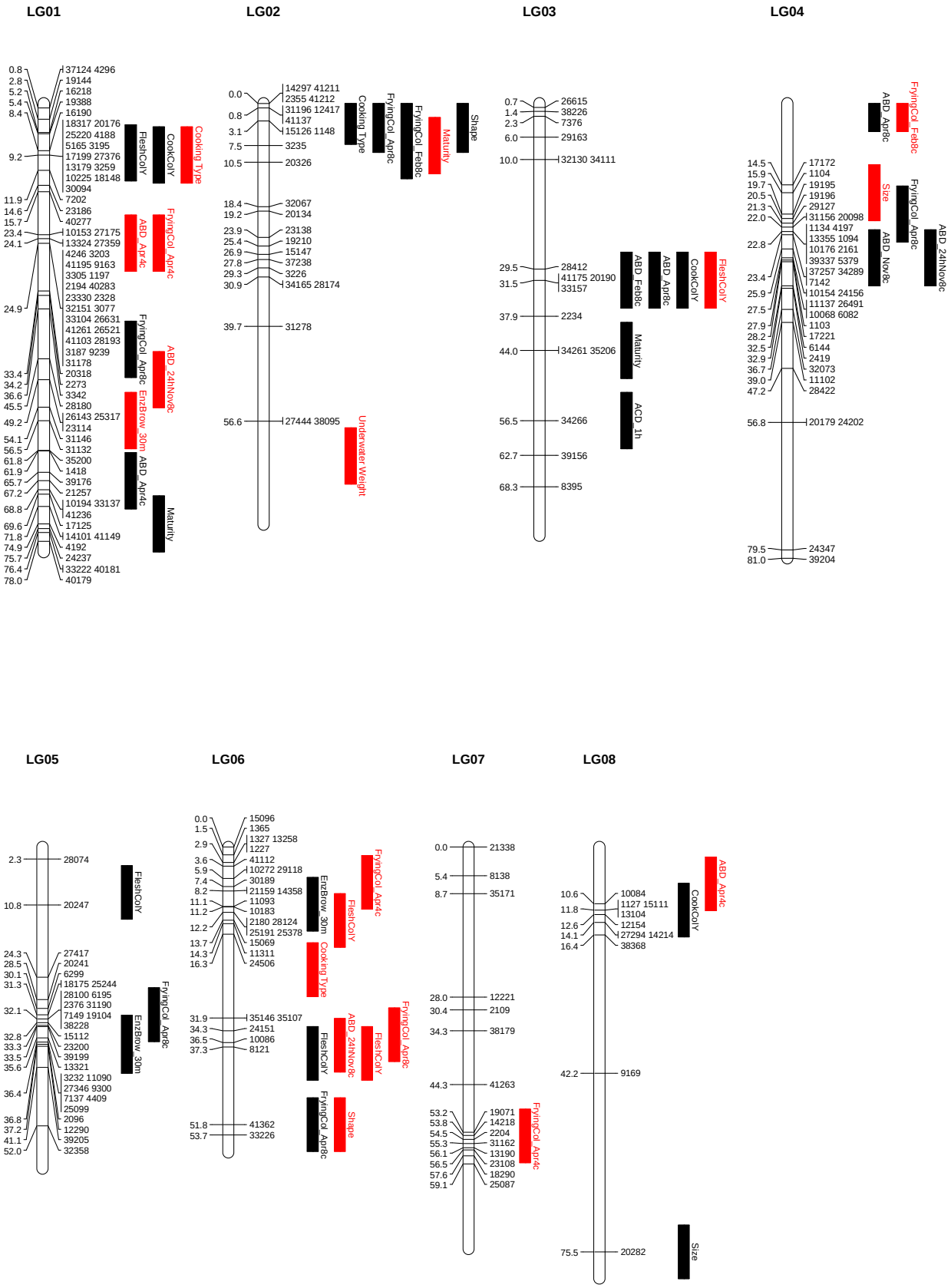
Table 3: Associations detected when association mapping was applied to the set of 53 microsatellites.

	Number of associated SSR loci with p ≤ 0.001 for the 2006 trial: main effect QTL only			Number of associated SSR loci with p ≤ 0.001 for the MYML data set		
Trait	Simple model	Relatedness corrected	SSR locus – Chromosome number	Simple model	Relatedness corrected	SSR locus – Chromosome number
ABD_Nov8c	0	1	STI012 - 4	5	0	--
ABD_Feb8c	1	0	--	3	1	STM0024 - 8
ABD_Apr8c	1	0	--	2	1	STM0024 - 8
ABD_Apr4c	0	0	--	4	3	STI009 - 1, STM1030 - 2, STM1106 – 10
ABD_24hNov8c	0	1	STI012 - 4	1	1	STM5148 – 5
ACD_1h	0	0	--	3	2	STGBSS - 8, StWAX-1 - 8
ACD_24h	0	0	--	0	0	--
CookColY	0	0	--	0	0	--
Cooking Type	0	0	--	6	0	--
EnzBrow_30min	0	0	--	1	1	STM1024 - 8
FleshColY	0	0	--	9	1	STM0003 - 12
FryingCol_Nov8c	1	1	STI003 - 8	5	0	--
FryingCol_Feb8c	0	0	--	6	2	STM1055 - 8, STM1005 - 8
FryingCol_Apr8c	1	2	STM1054 - 3, STM0051 - 10	5	0	--
FryingCol_Apr4c	1	0	--	0	1	STM0003 - 12
Maturity	3	3	StWAX-2 - 8, STM1055 - 8, STM1005 - 8	17	0	--
Shape	3	1	STI055 - 4	8	0	--
Size	3	2	STI012 - 4, STM0013 - 5	1	0	--
Underwater Weight	1	0	--	23	0	--

Trait relationships within data sets

Table 1 reveals common associations between some traits for each phenotypic data set, emphasizing the presence of relationships between traits. These trait relationships are more abundant within the 2006 field trial data set than within the MYML data set. This is partly attributable to the more severe effect of genetic relatedness correction within the germplasm collection of the MYML data set which effectively led to the disappearance of common trait loci when compared to the outcome of the uncorrected association mapping model. Therefore we will mainly focus on the relationships apparent within the 2006 data set.

Although relationships between agro-morphological traits like underwater weight, maturity type and cooking type were anticipated based on the relationships observed between their genotypic main effects (Chapter 3), no evidence of common QTL could be discerned when the association mapping model accounting for genetic relatedness was applied. However, when no genetic relatedness correction would have been performed, almost all loci detected for maturity and cooking type would have been found shared with underwater weight within both the 2006 and MYML data set (**Supplementary Tables 1** and **2** respectively). The two traits reflecting tuber tissue yellowness, FleshColY and CookColY, shared one single QTL on chromosome *1* (8.8-9.2cM) within the 2006 data set. When inspecting these two traits across both phenotypic data sets another common locus is revealed, namely on chromosome *3* at 31.5cM. Again, when the uncorrected data are inspected a higher number of common loci appears.


LG01
LG02
LG03
LG04
LG05
LG06
LG07
LG08
FleshColY
CookColY
Cooking Type
FryingCol_Apr4c
ABD_Apr4c
FryingCol_Apr8c
ABD_24hNov8c
EnzBrow_30m
Maturity
Shape
FryingCol_Feb8c
Underwater Weight
ABD_Feb8c
ACD_1h
Size
ABD_Nov8c

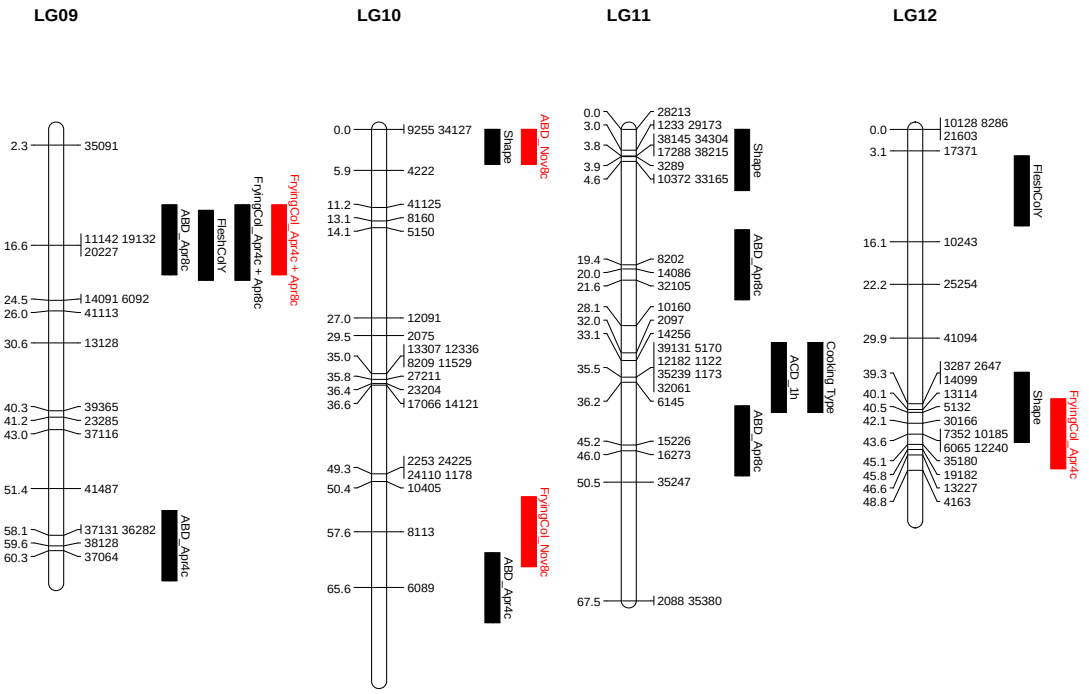


Figure 1: Distribution of QTL detected with the relatedness corrected model along the potato genome. Borders of QTL loci were estimated by adding or subtracting 5cM to the associated marker positions. QTL in black were obtained with the 2006 field trial data set, QTL in red with the MYML data set.

Unlike for the agro-morphological traits, common QTL between quality traits were maintained even after relatedness correction. Two out of four frying colour traits (FryingCol_Apr8c and FryingCol_Apr4c) share a QTL on chromosome *9* (15.8-16.6cM), also across the two phenotypic data sets. The frying colour traits that were measured after storage at 8˚C share a locus on chromosome *2* (3.7cM), at least within the 2006 data set. For after baking darkening (ABD) common QTL could be detected as well, although not one QTL was shared by all assessed ABD treatments. It concerns a QTL on chromosome *3* (31.5cM) and one on chromosome *4* (27.5cM). Also in **Table 3**, a common SSR locus was found between two ABD traits on chromosome *4* (STI012). After cooking darkening (ACD) is probably related to some extent with after baking darkening since both detected QTL for ACD, on chromosome *3* at 56.5cM and on chromosome *11* at 35.5cM, could be linked with QTL for ABD (chromosome *3* at 31.5cM, chromosome *11* at 44.6cM) as they reside more or less close-by on the same chromosome, although not within the predicted LD decay range (Chapter 4). Also here these trait commonalities were expected as they could be partly deduced at the stage of phenotypic analysis (Chapter 3). Similar as for the agro-morphological traits, the trait relationships between sets of quality traits, like ABD or frying colour become more obvious when the results of the baseline model are examined (**Supplementary Tables 1** and **2**).

Genotype-by-location QTL

Table 4 demonstrates that far less QTL involved in genotype-by-location interaction have been detected when compared to main effect QTL and that the ones that have been detected sometimes match with main effect QTL. When a main effect QTL coincides with a QTL for genotype-by-environment interaction, we should not try to interpret the main effect, but better concentrate on the difference between the sand and clay location. **Table 4** shows a number of main effect QTL that are better interpreted purely in terms of genotype-by-environment

interaction. For ABD, we mention the QTL on chromosome *3* with genotype-by-environment interaction at 29.5cM (-10logP = 3.4) and main effect at 31.5cM (-10logP = 7.1). As the main effect is a lot stronger, the difference between sand and clay seems relatively small in relation to the main effect. On chromosome *11* at 44.6cM we find another example (-10logP for main effect is -5.1 versus the -10logP for the genotype-by-location QTL of 3.8). For FleshColY we find a QTL at chromosome *9* (16.6cM: -10logP main effect 3.1 versus 3.2 for the genotype-by-location QTL), indicating that this QTL should be interpreted only in terms of a genotype-by-location interaction. Similarly, CookColY has a main effect QTL on chromosome *8* (11.8cM) which is in the proximity, albeit separated by 13cM, of a genotype-by-location QTL at 25cM on chromosome *8*. The appearance of genotype-by-location QTL for the ABD traits was predictable as during the phenotypic analyses of these traits, the estimated variance component for genotype-by-location was high in comparison to the genetic variance component (Chapter 3). No genotype-by-location QTL could be detected for frying colour traits which is not so surprising as they had low genotype-by-location variance components relative to their respective genetic variance components.

Table 4: Association mapping results as obtained with the 2006 field trial and the set of 720 AFLPs with known map location. On the left side the main effect QTL are mentioned, on the right side the genotype-by-location interaction QTL.

	Number of associated loci with p ≤ 0.001 for the 2006 trial: main effect QTL			Number of associated loci with p ≤ 0.001 for the 2006 trial: QTL depicting GxE		
Trait	Simple model	Relatedness corrected	Chromosome nr (position cM)	Simple model	Relatedness corrected	Chromosome nr (position cM)
ABD_Nov8c	6	1	4 (27.5cM)	0	0	--
ABD_Feb8c	5	1	3 (31.5cM)	1	1	3 (29.5cM)
ABD_Apr8c	10	5	3 (31.5cM), 4 (0cM), 9 (15.8cM), 11 (19.4, 44.6cM)	2	3	6 (15.2cM), 11 (44.6cM), 12 (43.6cM)
ABD_Apr4c	5	3	1 (67.2cM), 9 (59.6cM), 10 (65.6cM)	0	0	--
ABD_24hNov8c	5	1	4 (27.5cM)	1	1	10 (0cM)
ACD_1h	7	2	3 (56.5cM), 11 (35.5cM)	1	1	12 (43.6cM)
ACD_24h	0	0	--	0	0	--
CookColY	20	3	1 (9.2cM), 3 (31.5cM), 8 (11.8cM)	1	1	8 (25cM)
Cooking Type	5	2	2 (2.3cM), 11 (35.5cM)	1	1	10 (36.4cM)
EnzBrow_30min	4	3	5 (36.4, 37.2cM), 6 (10.7cM)	1	1	1 (24.9cM)
FleshColY	12	5	1 (8.8cM), 5 (8.5cM), 6 (38.5cM), 9 (16.6cM), 12 (8.8cM)	5	2	9 (16.6cM), 10 (14.1cM)
FryingCol_Nov8c	7	0	--	1	1	6 (15.2cM)
FryingCol_Feb8c	9	2	2 (3.7, 8.4cM)	0	0	--
FryingCol_Apr8c	9	7	1 (43.8cM), 2 (3.7cM), 4 (19.7cM), 5 (31.3cM), 6 (51.8cM), 9 (15.8, 16.6cM)	0	0	--
FryingCol_Apr4c	5	1	9 (16.6cM)	0	0	--
Maturity	10	2	1 (74.9cM), 3 (44cM)	3	1	4 (22.8cM)
Shape	10	4	2 (3.7cM), 10 (0cM), 11 (3.8cM), 12 (39.8cM)	2	2	1 (24.9cM), 7 (55.3cM)
Size	4	1	8 (75.5cM)	4	4	2 (0.8cM), 10 (35, 59.5cM), 12 (3.1cM)
Underwater Weight	25	0	--	3	2	7 (56.5cM), 12 (46.6cM)

Discussion

Genetic relatedness correction

It is a commonly accepted strategy that for reliable association mapping research it is necessary to investigate first whether population structure is likely to cause false positives and whether genetic relatedness can be of influence to the outcome of the association mapping analysis (Breseghello and Sorrells 2006b; Jannink and Walsh 2002; Mackay and Powell 2007; Zhu et al. 2008). Relying on this reasoning we scanned our association panels for presence of population structure and genetic relationships between individual genotypes prior to association mapping analysis (Chapter 4). The findings of that exercise resulted in the creation of a more advanced association mapping model where adequate control for genetic relatedness could take place by imposing structure on the variance-covariance matrix through marker-based correlations. In retrospect, examining the obtained results in **Tables 1** to **3** and **Supplementary Tables 1** and **2**, it is clear that genetic relatedness correction had a severe deflating effect on the number of associated QTL in general as well as per trait. For almost all cases, as expected, the highest significant associations detected with the basic association mapping model were retained with the genetic relatedness corrected model as well, albeit generally with a less significant P-value (**Supplementary tables 1** and **2**). Still, it should be mentioned that the associated loci that were discarded or that dropped below a predefined significance threshold, e.g. $P \leq 0.001$ in our case, when the advanced model was applied, are not necessarily meaningless. Consequently, a careful inspection and comparison of detected QTL whether obtained with an association mapping model not correcting for relatedness as with one that does correct for relationships is advisable. In exceptional cases also new associations can appear, and in these instances it is equally sensible to treat these “orphan” loci with caution.

Trait relationships

The apparent trait relationships that were observed within our association mapping results all have a plausible explanation. During the phenotypic analysis, relationships between genotypic main effects of quality and agro-morphological traits were observed (Chapter 3). But there is also a biological explanation. When quality traits are concerned, common loci were predictable for the ABD traits and the FryingCol traits as these two trait sets measure exactly the same discolouration type following exactly the same protocol, the mere difference lies in the storage specifications. Likewise, the commonalities between ABD and ACD traits are reasonable since both observe a similar discolouration process (oxidation of a colourless iron-chlorogenic acid compound to a coloured iron-chlorogenic acid complex). The shared QTL found for the agro-morphological traits FleshColY and CookColY were evenly expected, both measure the colour of tuber tissue with just one processing step in between.

Genotype-by-environment QTL

As can be deduced from **Table 4** the number of detected genotype-by-location QTL is not that high. Still, there are a number of genotype-by-location interaction QTL present, albeit not coinciding with main effect QTL, when referring to the genetic relatedness corrected model, for agro-morphological traits like underwater weight, tuber shape and size, maturity and cooking type. Location or in this case soil effects seem reasonable for agro-morphological traits as soil can indeed influence the performance of a cultivar for these traits. Moreover when inspecting the variance components estimated for genotype-by-location interaction during phenotypic analyses (Chapter 3) for some of the agro-morphological traits like tuber size and underwater

weight, their height when compared to the genetic variance components is not that low that no effects should be expected. Clay soil has a better water retaining capacity than sandy soil which could lead to an environment favouring a later maturity type and equally well a higher underwater weight as well as a larger tuber size. Furthermore it is known that soil can influence tuber shape as tubers have in general a rounder shape in clay soil than in sandy soil. For cooking type the environmental influence can be seen as a result of its close relationship with underwater weight. The effect of soil type on quality traits like after cooking darkening, after baking darkening and enzymatic browning is in our situation probably related to the exceptional summer during the growing season of 2006 where abundant rain towards the end of the growing season was following a period of drought, leading to secondary growth. A bit surprising were the genotype-by-location QTL for yellowness of tuber flesh and cooked tuber tissue. One could speculate in this case that on sandy soil due to abundant rainfall hills were eroded, allowing sunlight to penetrate deeper into the soil which may have caused tuber greening. This tuber greening effect may have led to a bias for higher yellowness scores for tubers harvested on the sandy soil.

Comparison with QTL mapping studies

A comparison with results previously obtained for potato, whether in traditional QTL mapping studies both at the diploid and tetraploid level or in association mapping, is presented in **Table 1**. We will refer repeatedly to our previous association mapping study to compare results, where it should be stated that a different set of phenotypic trait data had been used, consisting mainly of summary statistics across years and locations, calculated according to company specific procedures. After baking darkening traits have been reported in literature as a result of an association mapping study by D'hoop et al. (2008). When combining all the ABD trait loci together, an overlap is clear: we can see that chromosomes *3*, *6* and *11* were previously found to be involved in ABD and for the loci on chromosome *4* there is even a perfect match at 27.5cM. On chromosome *1* several QTL for ABD were detected however they do not all reside in each other's vicinity, except for the locus at 24.9cM obtained with the MYML data set. It is however not unlikely that a common haplotype has been identified by the other QTL on chromosome *1*. For after cooking darkening a locus on chromosome *11* (between 5 and 19cM) has been described by Bradshaw et al. (2008), perhaps the same haplotype was detected in our study with a QTL at 35.5cM. More interesting is the link with a QTL on chromosome *11* (19.4cM) that was detected for ABD in our study. Since both ABD and ACD measure similar discolouration processes we may have identified the same locus. The observed loci for cooking type on chromosomes *2* and *6* were in agreement with our previous results (D'hoop et al. 2008). Recently, a study combining candidate gene analysis with QTL and expression analysis discerned several enzymatic discolouration loci (Werij et al. 2007), one QTL was mapped to chromosome *1* between 32.7 and 46.6cM, we found a QTL on the same chromosome at 56.5cM and in a previous association mapping study a locus on chromosome *1* was presented as well (D'hoop et al. 2008). Sliwka et al. (2008) identified a flesh colour locus on the top of chromosome *12*, and the *Y*-locus involved with beta-carotene hydroxylase resides on chromosome *3* (Bonierbale et al. 1988). Another set of flesh colour influencing loci were detected on chromosomes *1* and *6* (D'hoop et al. 2008). For comparison, we obtained QTL on chromosome *3* at 31.5cM, on chromosome *12* at 8.8cM as well as on chromosomes *1* (8.8cM) and *6* (13.7, 38.5cM). When we compile the loci indicative of the Maillard discolouration (frying colour) and put them side by side with other mapping studies, we can see that loci on chromosomes *1*, *2*, *6* and *7* have been reported before (Bradshaw et al. 2008; D'hoop et al.

2008; Menendez et al. 2002), and an exact match with our previous association mapping study (D'hoop et al. 2008) on chromosome *2* at 3.7cM can be found. Similar as with our results, a QTL for maturity has been observed on chromosome *3* (Visker et al. 2005) as well as on chromosome *1* (D'hoop et al. 2008). For tuber shape we discerned QTL on chromosomes *2* (3.7cM), *10* (0cM), *11*(3.8cM) and *12* (39.8cM) similarly as has been reported by Bradshaw et al. (2008), Sliwka et al. (2008) and D'hoop et al. (2008) for chromosome *2* and *11*, and by Sørensen et al. (2006) for chromosome *12*. A locus for tuber shape on chromosome *10* has previously been defined by van Eck et al. (1994b). Finally, for underwater weight, the obtained QTL on chromosome *2* has likely been observed by Gebhardt et al. (2005) as well.

Prospects

As has been illustrated in this paper, genome-wide association mapping can be a valuable tool to identify candidate loci for agro-morphological as well as quality traits in potato. Quantitative trait loci that have repeatedly been observed when analysing different phenotypic data sets are surely promising candidate QTL for further elaborate studies whether it concerns fine mapping by more traditional mapping studies or even functional studies where one inquires if a link between obtained QTL and for instance expression or metabolite data can be detected. If promising candidate loci could be translated into easily accessible PCR markers and if subsequently these PCR markers can be shown to confirm the link with a phenotypic trait, they could be useful for breeding purposes.

Acknowledgements

We would like to express our gratitude to the gene banks that were willing to provide us with tubers of scarce potato cultivars and we appreciate the involved potato breeding companies for kindly sharing their phenotypic data and for their support and help in the 2006 field trial. We would also like to thank the "2006 field trial team" for their help. We are indebted to the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research for financing this study.

Chapter 6

General discussion

Björn B. D'hoop
Herman J. van Eck

This thesis presents the results obtained within project P8 “Association mapping and family genotyping in potato” of the Centre for BioSystems Genomics (www.cbsg.nl). The anticipated results of this project should be the selection of a representative set of potato cultivars, genotyped with a sufficient number of AFLP and SSR markers to enable the estimation of the level of linkage disequilibrium (LD) across the potato genome. Furthermore it should give insight into the population structure and genetic relatedness present in the cultivated potato germplasm, next to the identification of marker-trait associations. In the original project proposal a number of open questions were raised, which were of great importance for an effective design of the experimental work.

1. How many existing potato varieties should be included?
2. What phenotypic data for which traits are available and can be used?
3. What structure is to be encountered in potato germplasm?
4. What kind of marker system should be used?
5. What marker density should be pursued for a genome-wide marker scan to ensure sufficient genome coverage?

The discussion will reflect on the experimental chapters of this thesis, thereby putting the original project goals and aforementioned questions in perspective. This will provide an overview of the progress made as well as a number of recommendations for further research.

The plant material

The results of this thesis rely on the initial choice made when selecting cultivars for inclusion in the association mapping panels. The goal was to represent worldwide (cultivated) potato germplasm without losing too much focus on the (Dutch) commercial potato gene pool. To meet with these final attainments, selection criteria accounting for geography, time point of cultivar introduction, intended market segment, acreage, and importance for the seed potato industry were adopted. Certainly, our collection comprises wide phenotypic and genetic diversity, in spite of the fact that commercial potato has only very few founding fathers. Based on calculations using pedigree information available from the potato pedigree database (van Berloo et al. 2007), we estimated the number of founders of commercial potato somewhere between five and ten cultivars (unpublished data). This inevitably leads to a narrow genetic base and high degree of interrelatedness for potato germplasm (Love 1999), which is off course also reflected in our association panels (Chapter 4). Remarkably this relatedness is not obviously deduced from DNA marker data. DNA marker data rather display high levels of polymorphism, suggesting a large gene pool. At this moment there is no obvious explanation for this contradiction, but speculatively one could imagine that (1) alleles in the potato gene pool are highly divergent DNA sequences, contributing to the level of DNA polymorphism, (2) that inbreeding depression is effectively preventing allele loss, and (3) that those few founding genotypes from the 19th century were probably highly heterozygous and unrelated. Further knowledge on the breeding history of contemporary potato cultivars may clarify certain marker-trait associations, as well as identify false positives. To avoid a worst case scenario where the majority of obtained marker-trait associations merely reflects hidden population structure or relatedness, resulting in false positives, the aspect of population structure requires an adequate statistical correction (Jannink and Walsh 2002; Mackay and Powell 2007). Knowledge on the breeding history of potato has hardly been used so far. Future analysis of our data will certainly allow the identification of ancient alleles which are no longer available in contemporary germplasm, and of alleles

appearing in time in resistant cultivars derived from wild species introgression segments, as well as the estimation of the effect of selection on allele frequency.

Phenotypic data

One major assumption prior to the start of this project was that no additional phenotypic data would have to be collected, because our germplasm was chosen from widely known cultivars. In addition, large phenotypic data sets were already available to us. One particularly relevant phenotypic data set intended for use in P8 was collected during a collaborative European project, RESGEN-CT95-34/45. The aim for that project was to gather phenotypic data from official variety testing institutes, gene banks and participating potato breeding companies. Unfortunately, initial inspection of this database revealed a severe bias to disease resistance test data, shortage of data concerning recently bred potato cultivars, trait values that were often obtained using outdated phenotyping methods or in periods marked by outmoded agronomic practices, and last but not least suffered from incompleteness (The RESGEN consortium is currently continued through the ECPGR Potato Working Group, and maintains the online database www.europotato.org).

This compelled us to investigate other ways to obtain the necessary phenotypic data to serve our purpose of association mapping. One of the options was to collect multi-year-multi-location (MYML) data as obtained through in-house performed field trials at participating breeding companies. This would certainly solve the problem of insufficient information on recent cultivar data as well as the bias of the RESGEN database towards resistance data, since this project aims to target quality and processing traits. Phenotypic data released by breeding companies are certainly representative of current breeding practices and contemporary agronomical standards as well as consumer preferences. However, one of the major pitfalls here is that these kind of data are usually highly unbalanced: cultivars emerge and disappear from National List field trials throughout the years. Cultivars intended for the processing industry are usually not well documented when starch characteristics are at stake. Likewise, cultivars bred for starch industry are usually not well phenotyped for traits relevant to the fresh consumption or processing market. Furthermore, breeding companies differ in field trial practices and trait evaluation.

In an attempt to obtain a balanced phenotypic data set with relevant trait values, we performed a field trial ourselves with the assistance of the involved potato breeding companies to assure similarity in trait data between both the MYML data set and the 2006 designed field trial. The evaluation of the entire set of cultivars under identical conditions resulted in valuable data, but validity of the data across years (and locations) will remain unknown. In our case, the 2006 growing season was characterised by exceptional sunny and dry weather conditions in the first two months and abundant rain in the third month. This surely resulted in inadequately estimated genotype-by-environment interaction and consequently in overestimated heritabilities. Furthermore, since we used four plant plots only, the available harvested tuber material for phenotyping was limited. Therefore the data of the designed 2006 field trial may give rise to a few spurious marker-trait associations.

Both our 2006 and MYML data set have been used to generate adjusted genotypic mean values for phenotypic traits. Clear marker-trait associations should give rise to some confidence in our results, similar to confidence that is based on marker-trait associations that consistently

appear across years and environments. Menendez et al. (2002) performed a QTL analysis across a series of years by concatenating QTL analyses for individual years and then comparing the detected QTL for consistency across years. Because these QTL analyses are done per year, the power to detect marker-trait associations will not benefit from similar experiments in other years. This lower power may create the impression that QTL are less consistent than they actually are as weak QTL will be missed in particular years due to chance factors drowning the QTL expression. In our association mapping approach, we first estimate the average genotypic effect across years, using all available information on the included genotypes. Subsequently, we relate the adjusted genotypic means to marker information. This approach targets the constant differences between genotypes across years and locations. To get an impression of the relative importance of genotype-by-environment interaction and QTL-by-environment interaction, one can compare the variance components for the genotypic main effect with those for the genotype-by-environment interaction. In **Table 4** (Chapter 3) we see that for most traits the genotype-by-environment interaction was small and thus we feel justified in adopting an association analysis approach that specifically targets the genotypic main effects, i.e. the consistent QTL expression across environments.

Although all the phenotypic data for this thesis have been collected with the help of clear protocols used by breeding companies, it has been recognised that in peer reviewed literature there is little consistency on trait names or procedures for data collection. For example, specific gravity, underwater weight, starch content, dry matter content are highly similar but not synonymous traits. Likewise, breeder's preference, visual appearance, size grading, tuber number and average tuber size are non-synonymous but correlated traits. It is perceived that the current efforts to develop controlled vocabularies (ontologies) that describe plant structures, growth and developmental stages (www.plantontology.org, Avraham et al. 2008; Ilic et al. 2007) should be taken as an example towards an ontological description of potato quality traits.

Population structure

Chapter 4 deals with population structure in cultivated potato germplasm. Two different approaches to discover population structure and genetic relatedness did, in essence, not reveal strong statistically supported results. Cluster analysis did not result in bootstrap values (as performed with MEGA in Chapter 2) and likewise weak group membership probabilities were obtained with STRUCTURE. The modest statistical support for population subdivision described in Chapter 4 is in agreement with literature data about population structure in commercial potato germplasm (Gebhardt et al. 2004; Simko et al. 2004b, 2006). So far the dendrograms in this thesis have remained unrooted, although the oldest cultivars like cv. Early Rose and cv. Paterson's Victoria, could have been assigned as outgroup. This is surely something we are going to perform in the future.

Based on the results of our analysis on population structure, we decided that a marker-based genetic relatedness matrix should be sufficiently effective and powerful to avoid too many false positives, thereby following a similar strategy as for example Malosetti and colleagues (2007) used. In our opinion it was not necessary to include an extra factor to correct for relatedness, in the form of a "Q"-matrix based on group membership probabilities obtained from STRUCTURE (Pritchard et al. 2000), as suggested by Yu et al. (2006), because intrinsically by structuring the variance-covariance matrix of the random genotypic effects based on marker correlations one

already corrects for population structure as well as for more subtle relatedness effects. Also, we did not find strong group structure.

Choice of marker system

In this project two different marker types have been used. First the AFLP™ marker system, which results in polymorphic patterns due to SNP variation in the restriction enzyme recognition sites and/or the selective nucleotides (Brugmans et al. 2002). Hence the AFLP system is representative for any other method to visualise SNP data. Second, SSR markers have been used. The molecular mechanism leading to SSR length polymorphisms is most likely replication slippage of DNA polymerases during DNA replication (Ellegren et al. 2004). In view of the distinction between the molecular mechanisms that cause AFLP and SSR polymorphisms, there is no prior reason to assume that SSR or AFLP data would provide a similar glimpse of the potato genome.

Besides this fundamental difference between both polymorphisms there are also practical considerations typical for each method. AFLPs have a high multiplex ratio which renders them cost effective, they are reliable and reproducible and no prior information is needed (Meudt and Clarke 2007). The availability of an ultra dense potato map with 10,000 AFLP markers (van Os et al. 2006) made AFLP also a convenient choice. However, AFLPs have their shortcomings as well. They are dominant bi-allelic markers (sometimes even called mono-allelic, referring to the absence of the alternate allele), although quantitative information on band intensity can be obtained. AFLP data capture can thus be performed in three different ways. First, absence/presence data can be recorded and was possible for 1144 bands. Absence/presence data will not be free from data error due to intensity misclassifications by image analysis software threshold settings. Second, intensity values can be obtained for all fragments. The intensity values will not only reflect absence/presence information, but to a certain degree also band zygosity. Furthermore, for the monomorphic (i.e. not absent/present) AFLP bands it is very hard to decide if they indicate genetic fixation at that locus (all quadruplex), or reflect a very high AFLP allele frequency rendering predominantly quadruplex, triplex and perhaps some duplex genotypes. Intensity values represent the data in the purest form without interpretations and thus not affected by classification errors. Most of the analyses in this thesis were performed with intensity data. Third, for all markers we are in the process of classifying band intensity into genotypic classes. This would allow to replace a quantitative analysis into a real genetic analysis of the potato cultivars. The classification however is not a trivial issue and so far we cannot comment on the validity of the latter approach. Size homoplasy, i.e. the occurrence of alleles identical in size but not by descent, can complicate the interpretation of banding patterns (Koopman 2005; Koopman and Gort 2004). It is perceived that homoplasy should not be a major problem for those markers for which position information is available. These markers proved to behave properly as single copy, single locus markers. The number of exceptions is roughly known and below 1%. One chapter in the recent PhD thesis of Theo Borm (http://library.wur.nl/wda) shows a method to identify marker homoplasy and to disentangle the signal, allowing the identification of the two (independent) map positions for each of the superimposed fragments.

Microsatellites are multi-allelic which enables the detection of rare alleles, they contain more information per locus, they provide a good genome coverage and they can be scored in a qualitative (presence/absence of alleles) or a quantitative (allele dosage) fashion (Sharopova

2008; Varshney et al. 2005). However, they suffer from PCR failure, resulting in stutter bands or null alleles and specifically their multi-allelism poses challenges for analysis and interpretation (Provan et al. 1996a; Varshney et al. 2005). And also for microsatellite loci size homoplasy can play a role (Lia et al. 2007).

Reliable marker band intensity quantification is a prerequisite for genotyping of polyploids

For both SSRs and AFLPs it has become obvious that the tetraploid level of cultivated potato is not a trivial issue. The SSR pattern as observed from gel or electropherogram may clearly reveal a specific number of alleles, but as long as less than four different alleles can be discerned, the interpretation of the pattern into a genotype will require reliable quantification of allele intensities to allow an estimation of the exact allele dosage. In the case of AFLP markers the zygosity of the AFLP fragment is even more difficult to assess. The AFLP data from this study have been generated by a MegaBACE capillary sequencer, which is expected to be of the highest accuracy in both fragment mobility as well as dynamic range to capture fragment intensity. However, raw data from a MegaBACE cannot be understood without image analysis software to capture the AFLP marker data. This creates dependence on settings and thresholds for band calling when using image analysis software. Indeed, visual inspection of printouts of pseudo gel images does not always fully support the outcome of band calling software. However, visual inspection, although recommended, is prohibitively time consuming. As an alternative to band calling and zygosity determination we used band fluorescence intensities as a quantitative interpretation of the AFLP fingerprints. In the worst case these band intensities will offer an equal amount of information as conventional presence/absence data (Piepho 2001).

Should other marker techniques have been explored? A large number of marker techniques have been proposed to visualise SNP variation, but the aspects discussed before, affecting the AFLP data quality will equally result in ambiguities for other marker systems. For example, the GoldenGate assay developed by Illumina has been tested in our lab recently (data not shown, manuscript in preparation). It was shown that diploid mapping populations can be analysed with extremely high accuracy, but unstructured populations of tetraploids again resulted in ambiguous interpretations.

In this discussion we aim to propose an explanation for the problematic conversion of SNP zygosity based signal intensity values into allele copy numbers in the tetraploid. The foremost reason should be sought in the extremely high number of SNPs observed between potato alleles. Most SNP detection platforms scan between 15 and 50 nucleotides of DNA sequence in the proximity of a target SNP polymorphism. In the case of AFLP the 15 nucleotides are composed of the six and four cutter restriction sites plus the +3/+3 selective nucleotides minus the target SNP. This is only the lower value, primers for PCR amplification or ligation methods generally infer much larger nucleotide stretches. As soon as one additional SNP is within the proximity of the target SNP, the assay will fail, and will cause the presence of a null allele. Therefore the expected number of intensity classes is no longer based on the five possible classes for a bi-allelic marker polymorphism (AAAA, AAAB, AABB, ABBB, BBBB). The null allele will add another nine genotypic classes (AAAO, AAOO, AOOO, AABO, ABOO, ABBO, BBBO, BBOO and BOOO) with a specific signal intensity differing from the expected five classes. The issue of null alleles as discussed here for AFLPs equally applies to SSR data.

In conclusion this research project has shown that excessive SNP variation is an intrinsic problem for the genetic analysis of unstructured tetraploid populations. However the effect of extreme SNP variation will never be prohibitive for any marker system. Null alleles will certainly cause some noise to the data, but this seems unavoidable and can be taken into consideration while analysing polymorphic marker patterns.

Marker number

The current project has resulted in a genome-wide marker scan of 3364 AFLPs and 53 SSR markers. Is this a sufficient coverage to allow the detection of QTL via marker-trait associations? For an individual potato cultivar this requires the coverage of 48 chromosomes, and suggests that the genetic length of the potato genome (~1000cM) requiring marker coverage has to be multiplied at least four times.

Calculations or a simulation study that aim to answer the question of marker coverage have to take a number of uncertainties into account. First, it is required to consider the non-random placement of the markers on the potato genome. Non-random marker positioning is caused by localised and parent-specific variation in recombination affecting the estimated marker distances (in cM) on the genetic map. Typically the centromeric suppression of recombination results in marker-dense clusters at the centromeric position of a linkage group (van Os et al. 2006). With the current marker number it is therefore very probable that many chromosome arms were not tagged by an AFLP marker. Adding more and more AFLP markers would not considerably improve map coverage though, because of “the law of diminishing returns”. After 41 AFLP primer combinations, the relation describing added markers per extra primer combination will have reached a nearly horizontal slope. Further marker saturation should be achieved with and based on targeted and gene based SNPs only. Here the large collection of EST sequences can be exploited to predict SNPs (Tang et al. 2006, 2008) which can subsequently be tested with e.g. the GoldenGate assay. Addition of more SSR markers does not seem to have any value, as the number of single locus mapped SSR markers is already largely exploited.

A second aspect is the non-random distribution of markers across haplotypes. This is in fact the reflection of the non-random distribution of SNPs (see **Figure 1**). The more deviating alleles (for example wild species introgressions) may give rise to much more AFLP markers, resulting in marker redundancy, whereas there may be a severe underrepresentation of AFLP markers tagging the more common alleles.

In conclusion we still wait to answer the question if our genome-wide marker scan had sufficient coverage to expect the detection of important marker-trait associations. The non-random placement of AFLP markers across a linkage group and the non-random placement of AFLPs across haplotypes cannot be quantified at this moment. For SSR markers the distribution of marker alleles across haplotypes is more obvious, because the definition of SSR allele and haplotype are converging for those SSRs with many alleles. These non-random aspects effectively prevent us from reaching an answer to the question on genome coverage with markers.

Prospects

With the research conducted during this four year project a number of promising marker-trait associations have been found for agro-morphological traits like tuber size, cooking type, tuber shape and underwater weight as well as for quality traits like cold sweetening (indirectly through loci detected for frying colour), other non-enzymatic discolouration processes like after cooking darkening and after baking darkening, and enzymatic discolouration. These QTL, when translated into diagnostic markers with validated discriminatory power for each of the corresponding phenotypic traits, may become useful in marker assisted selection.

However, the validated discriminatory power of an AFLP marker is not easily retained during marker conversion. Excision and sequencing of the AFLP fragment may allow the development of a nice single locus marker, but this is not the relevant point. It must have been a specific SNP that conveyed the significant marker-trait association. Hence the AFLP causing SNP has to be identified. Brugmans et al. (2003) have described an incontestable procedure to capture the AFLP causing SNP. Only this SNP assay will have predictive power, even if the distance and orientation relative to the trait locus remains unknown. In view of the resolution of LD studies, the trait locus should be expected within less than 5-10cM, which is an acceptable distance for marker assisted selection. Caution remains important, because specific haplotypes may extend over much larger cM intervals. Furthermore when translating an AFLP marker into a simple single locus marker it is also highly recommended to achieve full resolution of all haplotypes as illustrated in **Figure 1**. Using full resolution haplotyping at the marker locus, it will be possible to test all individual alleles as well as allele interactions in their effect on trait variation.

With the advent of the potato genome sequence ahead (www.potatogenome.net), much more of the yet unmapped markers will receive a map position, including the QTL associated to some of these markers. One of the current complications is the perceived redundancy among markers associated with specific traits. Prior to (expensive) marker translations the putative marker redundancy has to be identified. For example there could be a situation with several unmapped markers which are associated to the same trait locus. Are these markers tagging different haplotypes, or are there marker pairs that bracket the most relevant trait allele?

There are two ways to analyse redundancy in marker-trait associations: (1) using a genetics approach which will study marker position in terms of map position and haplotype, and (2) using a statistical approach by applying multiple regression that aims to identify for each trait a minimal number of markers that can explain a maximal proportion of the trait variance. It is not excluded that both approaches can be integrated.

So far the use of pedigree information has received little attention. One of the first applications of the known familial relations could be the identification and removal of some false scores in the marker data e.g. due to presence of undetected null alleles. In a later stage the analysis of marker-trait associations should also benefit from this important information on familial relations.

Would high heritabilities warrant the identification of significant marker-trait associations? A recent review about human length, serving as a model for truly polygenic traits (Weedon and Frayling 2008), but also a recent QTL mapping study in potato genetics (Bradshaw et al. 2008) showed that high heritability traits may result in the identification of very many, very small sized

QTL. For our traits these two papers suggest that two scenarios can be discerned, concerning the detection of QTL: (1) oligogenic traits where surprisingly few major-effect QTL can be distinguished with the obvious opportunities for marker assisted selection, contrasting to (2) truly polygenic traits where the explained variance invariably remains below e.g. 5%. Most likely potato breeding companies will hesitate to invest in and adopt diagnostic markers that only offer a potential improvement of a trait by less than 5%.

This project has resulted in a number of interesting marker-trait associations, but LD studies in tetraploids are still in their infancy. Within a CBSG bottleneck project the genetic composition of potato alleles of candidate genes was studied using a re-sequencing approach. This resulted in more insight in allelic variation underlying trait variation and the consequences for marker development. In **Figure 1** the SNP positions are illustrated for ten alleles of a functional gene. We propose that the multiple SNPs that characterise each individual allele are sufficiently powerful to allow to conclude that repeated observations of the same haplotype are indicative for identity by descent.

Individual unphased SNPs have much less power as compared to haplotypes. The SNPs with a grey background in **Figure 1** (indicated with large arrows) are not unique to one haplotype thus representing SNPs that lump a variable number of haplotypes. The other SNPs indicated with a shorter arrow are indicative for a unique haplotype. Only a comprehensive set of these latter tagging SNPs would allow to fully model the genetic effects of individual alleles and the allele interactions on trait values.

Imagine that the financial budget spent during this project on marker loci could be spent today on next generation sequencing approaches! To our opinion re-sequencing of allelic variation of expressed genes would immediately reveal SNPs tagging unique haplotypes. This new level of information is expected to enable a complete reconstruction of the haplotype structure of tetraploid potato cultivars.

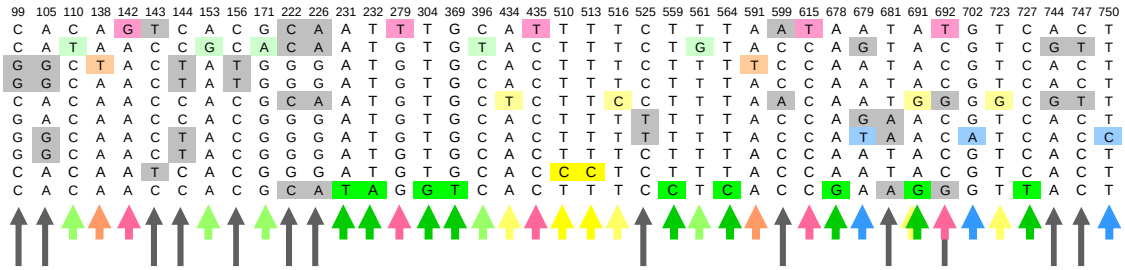


Figure 1: Results obtained by re-sequencing alleles of a candidate gene locus. Ten different haplotypes are shown. The numbers above indicate the position in base pairs within the re-sequenced stretch of DNA. Arrows indicate where lumping (large grey arrows) SNPs are detected as opposed to tagging (arrow colours specific for the haplotype) SNPs.

In retrospect, this project has delivered and validated a valuable core set of potato cultivars representing worldwide commercial potato germplasm with which it has been shown that, provided a decent statistical modelling environment, relevant QTL can be detected using a genome-wide association mapping approach irrespective of how phenotypic data has been obtained or what marker system has been used.

Literature cited

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length Polymorphisms of Simple Sequence Repeat DNA in Soybean. Genetics 132:1131-1139

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes. PLoS Genetics 1:e60

Asiedu SK, Astatkie T, Yiridoe EK (2003) The Effect of Seed-Tuber Physiological Age and Cultivar on Early Potato Production. Journal of Agronomy and Crop Science 189:176-184

Avraham S, Tung CW, Ilic K, Jaiswal P, Kellogg EA, McCouch S, Pujar A, Reiser L, Rhee SY, Sachs MM, Schaeffer M, Stein L, Stevens P, Vincent L, Zapata F, Ware D (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. Nucleic Acids Research 36:D449-D454

Bachem CWB, Speckmann G-J, van der Linde PCG, Verheggen FTM, Hunt MD, Steffens JC, Zabeau M (1994) Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. Biotechnology 12:1101-1105

Barnaud A, Lacombe T, Doligez A (2006) Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. TAG Theoretical and Applied Genetics 112:708-716

Barone A (2004) Molecular marker-assisted selection for potato breeding. American Journal of Potato Research 81:111-117

Bell CJ, Ecker JR (1994) Assignment of 30 Microsatellite Loci to the Linkage Map of *Arabidopsis*. Genomics 19:137-144

Bendahmane A, Kanyuka K, Baulcombe DC (1997) High-resolution genetical and physical mapping of the *Rx* gene for extreme resistance to potato virus X in tetraploid potato. TAG Theoretical and Applied Genetics 95:153-162

Bisognin DA, Douches DS, Buszka L, Bryan G, Wang D (2005) Mapping late blight resistance in *Solanum microdontum* Bitter. Crop Science 45:340-345

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A Mixed-Model Quantitative Trait Loci (QTL) Analysis for Multiple-Environment Trial Data Using Environmental Covariables for QTL-by-Environment Interactions, With an Example in Maize. Genetics 177:1801-1813

Bonierbale MW, Plaisted RL, Pineda O, Tanksley SD (1994) QTL analysis of trichome mediated insect resistance in potato. TAG Theoretical and Applied Genetics 87:973-987

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Current Opinion In Plant Biology 7:132-136

Bormann CA, Rickert AM, Ruiz RAC, Paal J, Lubeck J, Strahwald J, Buhr K, Gebhardt C (2004) Tagging quantitative trait loci for maturity-corrected late blight resistance in tetraploid potato with PCR-based candidate gene markers. Molecular Plant Microbe Interactions 17:1126-1138

Bradshaw J, Hackett C, Pande B, Waugh R, Bryan G (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). TAG Theoretical and Applied Genetics 116:193-211

Bradshaw JE (2007) Potato-Breeding Strategy. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, MacKerron DKL, Taylor MA, Ross HA (eds) Potato biology and biotechnology - Advances and perspectives. Elsevier Ltd., Amsterdam, pp 157-177

Bradshaw JE, Hackett CA, Lowe R, McLean K, Stewart HE, Tierney I, Vilaro MDR, Bryan GJ (2006) Detection of a quantitative trait locus for both foliage and tuber resistance to late blight [*Phytophthora infestans* (Mont.) de Bary] on chromosome *4* of a dihaploid potato clone (*Solanum tuberosum* subsp. *tuberosum*). TAG Theoretical and Applied Genetics 113:943-951

Bradshaw JE, Hackett CA, Pande B, Waugh R, Bryan GJ (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp *tuberosum*). TAG Theoretical and Applied Genetics 116:193-211

Bradshaw JE, Mackay GR (1994) Potato genetics. CAB International, Wallingford

Bradshaw JE, Meyer RC, Milbourne D, McNicol JW, Phillips MS, Waugh R (1998) Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker-assisted selection. TAG Theoretical and Applied Genetics 97:202-210

Bradshaw JE, Pande B, Bryan GJ, Hackett CA, McLean K, Stewart HE, Waugh R (2004) Interval mapping of quantitative trait loci for resistance to late blight [*Phytophthora infestans* (Mont.) de Bary], height and maturity in a tetraploid population of potato (*Solanum tuberosum* subsp *tuberosum*). Genetics 168:983-995

Breseghello F, Sorrells ME (2006a) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165-1177

Breseghello F, Sorrells ME (2006b) Association Analysis as a Strategy for Improvement of Quantitative Traits in Plants. Crop Science 46:1323-1330

Brewer MT, Lang L, Fujimura K, Dujmovic N, Gray S, van der Knaap E (2006) Development of a Controlled Vocabulary and Software Application to Analyze Fruit Shape Variation in Tomato and Other Plant Species. Plant Physiology 141:15-25

Breyne P, Boerjan W, Gerats T, VanMontagu M, VanGysel A (1997) Applications of AFLP(TM) in plant breeding, molecular biology and genetics. Belgian Journal of Botany 129:107-117

Brown C, Kim T, Ganga Z, Haynes K, De Jong D, Jahn M, Paran I, De Jong W (2006) Segregation of total carotenoid in high level potato germplasm and its relationship to beta-carotene hydroxylase polymorphism. American Journal of Potato Research 83:365-372

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20:1004-1010

Brugmans B, del Carmen AF, Bachem CWB, van Os H, van Eck HJ, Visser RGF (2002) A novel method for the construction of genome wide transcriptome maps. Plant Journal 31:211-222

Brugmans B, van der Hulst RGM, Visser RGF, Lindhout P, van Eck HJ (2003) A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. Nucleic Acids Research 31:e55

Bryan GJ, McLean K, Bradshaw JE, De Jong WS, Phillips M, Castelli L, Waugh R (2002) Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei*. TAG Theoretical and Applied Genetics 105:68-77

Bryan GJ, McLean K, Pande B, Purvis A, Hackett CA, Bradshaw JE, Waugh R (2004) Genetical dissection of *H3*-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. Molecular Breeding 14:105-116

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557-567

Carlson CS, Eberle MA, Kruglyak L, Nickerson DA (2004) Mapping complex disease loci in whole-genome association studies. Nature 429:446-452

Castiglioni P, Ajmone-Marsan P, van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: co-dominant scoring and linkage group distribution. TAG Theoretical and Applied Genetics 99:425-431

Chen X, Salamini F, Gebhardt C (2001) A potato molecular-function map for carbohydrate metabolism and transport. TAG Theoretical and Applied Genetics V102:284-295

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3

Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O'Sullivan DM (2008) Association mapping of partitioning loci in barley. BMC Genetics 9

Collard B, Jahufer M, Brouwer J, Pang E (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196

Collins A, Milbourne D, Ramsay L, Meyer R, Chatot-Balandras C, Oberhagemann P, De Jong W, Gebhardt C, Bonnel E, Waugh R (1999) QTL for field resistance to late blight in potato are strongly correlated with maturity and vigour. Molecular Breeding 5:387-398

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM, Schmader KE, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180-184

Costanzo S, Simko I, Christ BJ, Haynes KG (2005) QTL analysis of late blight resistance in a diploid potato family of *Solanum phureja* x *S. stenotomum*. TAG Theoretical and Applied Genetics 111:609-617

Cregan PB, Quigley CV (1997) Simple sequence repeat DNA marker analysis. In: Caetano-Anolles G, Gresshoff PM (eds) DNA markers, protocols, applications and overviews. Wiley, New York

D'hoop BB, Paulo MJ, Mank RA, van Eck HJ, van Eeuwijk FA (2008) Association mapping of quality traits in potato (*Solanum tuberosum* L.). Euphytica 161:47-60

Douches DS, Freyre R (1994) Identification of Genetic-Factors Influencing Chip Colour in Diploid Potato (*Solanum* spp.). American Potato Journal 71:581-590

Douches DS, Maas D, Jastrzebski K, Chase RW (1996) Assessment of potato breeding progress in the USA over the last century. Crop Science 36:1544-1552

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics 5:435-445

Ersoz ES, Yu J, Buckler ES (2007) Applications of Linkage Disequilibrium and Association Mapping in Crop Plants. In: Varshney RK, Tuberosa R (eds) Genomics-Assisted Crop Improvement. Springer Verlag, Germany, pp 97-119

Esselink GD, Nybom H, Vosman B (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. TAG Theoretical and Applied Genetics 109:402-408

Ewing EE, Simko I, Omer EA, Davies PJ (2004) Polygene mapping as a tool to study the physiology of potato tuberization and dormancy. American Journal of Potato Research 81:281-289

FAOSTAT (2008) www.faostat.fao.org.

Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Research 10:220-227

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). TAG Theoretical and Applied Genetics 111:456-466

Fernandez-Del-Carmen A, Celis-Gamboa C, Visser RGF, Bachem CWB (2007) Targeted transcript mapping for agronomic traits in potato. Journal of Experimental Botany 58:2761-2774

Flajoulot S, Ronfort J, Baudouin P, Barre P, Huguet T, Huyghe C, Julier B (2005) Genetic diversity among alfalfa (*Medicago sativa*) cultivars coming from a breeding program, using SSR markers. TAG Theoretical and Applied Genetics 111:1420-1429

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54:357-374

Freyre R, Douches DS (1994) Isoenzymatic identification of quantitative traits in crosses between heterozygous parents - mapping tuber traits in diploid potato (*Solanum* spp.). TAG Theoretical and Applied Genetics 87:764-772

Freyre R, Warnke S, Sosinski B, Douches DS (1994) Quantitative trait locus analysis of tuber dormancy in diploid potato (*Solanum* spp.). TAG Theoretical and Applied Genetics 89:474-480

Friedman M (1997) Chemistry, biochemistry, and dietary role of potato polyphenols. A review. Journal of Agricultural and Food Chemistry 45:1523-1540

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of Single Nucleotide Polymorphisms and analysis of Linkage Disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biology 8

Gabriel KR (1971) Biplot graphic display of matrices with application to principal component analysis. Biometrika 58:453

Gabriel KR (1978) Biplot as a tool for diagnosing models. Biometrics 34:156-157

Galvez HF, Fernandez EC, Hautea DM (2005) Molecular mapping of resistance to thrips in potato. Philippine Agricultural Scientist 88:268-280

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 165:759-769

Gauch HG, Jr. (1988) Model Selection and Validation for Yield Trials with Interaction. Biometrics 44:705-715

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502-1506

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schuler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Molecular Breeding 13:93-102

Gebhardt C, Menendez CM, Chen X, Li L, Schäfer-Pregl R, Salamini F (2005) Genomic approaches for the improvement of tuber quality traits in potato. Acta Horticultura (ISHS) 684:85-92

Ghislain M, Rodriguez F, Villamon F, Nunez J, Waugh R, Bonierbale M (1999) Establishment of microsatellite assays for potato genetic identification. CIP program report 1999-2000. CIP, Lima, pp 167-174

Ghislain M, Spooner DM, Rodriguez F, Villamon F, Nunez J, Vasquez C, Waugh R, Bonierbale M (2004) Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. TAG Theoretical and Applied Genetics 108:881-890

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding Genes That Underlie Complex Traits. Science 298:2345-2349

Glendinning DR (1983) Potato introductions and breeding up to the early 20th century. New Phytologist 94:479-505

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology 57:461-485

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163-185

Hagenblad J, Nordborg M (2002) Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. Genetics 161:289-298

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324

Han TH, van Eck HJ, De Jeu MJ, Jacobsen E (1999) Optimization of AFLP fingerprinting of organisms with a large-sized genome: a study on *Alstroemeria* spp. TAG Theoretical and Applied Genetics 98:465-471

Hougas R, Peloquin S (1958) The potential of potato haploids in breeding and genetic research. American Journal of Potato Research 35:701-707

Hutten RCB, van Berloo R (2001) An online potato pedigree database. http://www.dpw.wau.nl/pv/query.asp

Ilic K, Kellogg EA, Jaiswal P, Zapata F, Stevens PF, Vincent LP, Avraham S, Reiser L, Pujar A, Sachs MM, Whitman NT, McCouch SR, Schaeffer ML, Ware DH, Stein LD, Rhee SY (2007) The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant. Plant Physiology 143:587-599

Jacobs JME, Eck HJ, Arens P, Verkerk-Bakker B, Hekkert BL, Bastiaanssen HJM, El-Kharbotly A, Pereira A, Jacobsen E, Stiekema WJ (1995) A genetic map of potato (*Solanum tuberosum*) integrating molecular markers, including transposons, and classical markers. TAG Theoretical and Applied Genetics 91:289-300

Jacobsen E, Hovenkamp-hermelink JHM, Krijgsheld HT, Nijdam H, Pijnacker LP, Witholt B, Feenstra WJ (1989) Phenotypic and genotypic characterisation of an amylose-free starch mutant of the potato. Euphytica 44:43-48

Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, Pan X, Clark K, Teytelman L, Cartinhour S, Stein L, McCouch S (2002) Gramene: development and integration of trait and gene ontologies for rice. Comparative and Functional Genomics 3:132-136

Jannink J-L, Walsh B (2002) Association mapping in plant populations. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB International Publishing, pp 59-68

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. TAG Theoretical and Applied Genetics 99:1053-1060

Jansen RC, Geerlings H, Van Oeveren AJ, Van Schaik RC (2001) A comment on co-dominant scoring of AFLP markers. Genetics 158:925-926

Jorde LB (2000) Linkage disequilibrium and the search for complex disease genes. Genome Research 10:1435-1444

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient Control of Population Structure in Model Organism Association Mapping. Genetics 178:1709-1723

Karp A, Seberg O, Buiatti M (1996) Molecular techniques in the assessment of botanical diversity. Annals of Botany 78:143-149

Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends in Genetics 13:74-78

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends in Genetics 22:253-259

Kawchuk LM, Lynch DR, Thomas J, Penner B, Sillito D, Kulcsar F (1996) Characterization of *Solanum tuberosum* simple sequence repeats and application to potato cultivar identification. American Potato Journal 73:325-335

Kearsey M (1998) The principles of QTL analysis (a minimal mathematics approach). Journal of Experimental Botany 49:1619-1623

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? Heredity 80:137-142

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073-1080

Khu DM, Lorenzen J, Hackett CA, Love SL (2008) Interval mapping of quantitative trait loci for corky ringspot disease resistance in a tetraploid population of potato (*Solanum tuberosum* subsp *tuberosum*). American Journal of Potato Research 85:129-139

Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. Genome Research 3:13-22

Kirkman MA (2007) Global Markets for Processed Potato Products. In: Vreugdenhil D, Bradshaw J, Gebhardt C, Govers F, MacKerron DKL, Taylor MA, Ross HA (eds) Potato Biology and Biotechnology: Advances and Perspectives. Elsevier Ltd., Amsterdam, pp 27-44

Kloosterman B (2006) Transcriptomic analysis of potato tuber development and tuber quality traits using microarray technology. PhD Dissertation, Wageningen University

Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics 177:457-468

Koopman WJM (2005) Phylogenetic signal in AFLP data sets. Systematic Biology 54:197-217

Koopman WJM, Gort G (2004) Significance Tests and Weighted Values for AFLP Similarities, Based on *Arabidopsis* in silico AFLP Fragment Length Distributions. Genetics 167:1915-1928

Kowalski B, Cassells A (1999) Mutation breeding for yield and *Phytophthora infestans* (Mont.) de Bary foliar resistance in potato (*Solanum tuberosum* L. cv. Golden Wonder) using computerized image analysis in selection. Potato Research 42:121-130

Kraakman ATW, Martinez F, Mussiraliev B, Van Eeuwijk FA, Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Molecular Breeding 17:41-58

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

Kraft T, Hansen M, Nilsson N-O (2000) Linkage disequilibrium and fingerprinting in sugar beet. TAG Theoretical and Applied Genetics 101:323-326

Kreike CM, Dekoning JRA, Vinke JH, Vanooijen JW, Gebhardt C, Stiekema WJ (1993) Mapping of loci involved in quantitatively inherited resistance to the potato cyst nematode *Globodera rostochiensis* pathotype *Ro1*. TAG Theoretical and Applied Genetics 87:464-470

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5:150-163

Lærke PE, Christiansen J, Veierskov B (2002) Colour of blackspot bruises in potato tubers during growth and storage compared to their discolouration potential. Postharvest biology and technology 26:99-111

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037-2048

Landergott U, Naciri Y, Schneller J, Holderegger R (2006) Allelic configuration and polysomic inheritance of highly variable microsatellites in tetraploid gynodioecious *Thymus praecox* agg. TAG Theoretical and Applied Genetics 113:453-465

Leonards-Schippers C, Gieffers W, Schafer-Pregl R, Ritter E, Knapp SJ, Salamini F, Gebhardt C (1994) Quantitative resistance to *Phytophthora infestans* in potato - a case study for QTL mapping in an allogamous plant species. Genetics 137:67-77

Li L, Paulo M-J, Strahwald J, Lübeck J, Hofferbert H-R, Tacke E, Junghans H, Wunder J, Draffehn A, van Eeuwijk F, Gebhardt C (2008) Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. TAG Theoretical and Applied Genetics 116:1167-1181

Li L, Strahwald J, Hofferbert HR, Lubeck J, Tacke E, Junghans H, Wunder J, Gebhardt C (2005) DNA variation at the invertase locus *invGE/GF* is associated with tuber quality traits in populations of potato breeding clones. Genetics 170:813-821

Li X, van Eck HJ, Rouppe van der Voort JNAM, Huigen DJ, Stam P, Jacobsen E (1998) Autotetraploids and genetic mapping using common AFLP markers: the *R2* allele conferring resistance to *Phytophthora infestans* mapped on potato chromosome *4*. TAG Theoretical and Applied Genetics 96:1121-1128

Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology 11:2453-2465

Lia V, Bracco M, Gottlieb A, Poggio L, Confalonieri V (2007) Complex mutational patterns and size homoplasy at maize microsatellite loci. TAG Theoretical and Applied Genetics 115:981-991

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics 44:397-401

Love S, Werner B, Pavek J (1997) Selection for individual traits in the early generations of a potato breeding program dedicated to producing cultivars with tubers having long shape and russet skin. American Journal of Potato Research 74:199-213

Love SL (1999) Founding clones, major contributing ancestors, and exotic progenitors of prominent North American potato cultivars. American Journal of Potato Research 76:263-272

Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR, Tai TH (2005) Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. Crop Science 45:66-76

Luo ZW, Zhang Z, Leach L, Zhang RM, Bradshaw JE, Kearsey MJ (2006) Constructing genetic linkage maps under a tetrasomic model. Genetics 172:2635-2645

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Molecular Breeding 15:271-289

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends in Plant Science 12:57-63

Mackay TFC (2001) The genetic architeccture of quantitative traits. Annual Review of Genetics 35:303-339

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879-889

Malosetti M, Visser RGF, Celis-Gamboa C, van Eeuwijk FA (2006) QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. TAG Theoretical and Applied Genetics 113:288-300

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139-145

Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. Molecular Ecology 11:1145-1156

Maris B (1969) Studies on maturity, yield, under water weight and some other characters of potato progenies. Euphytica 18:297-319

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The Extent of Linkage Disequilibrium in Rice (*Oryza sativa* L.). Genetics 177:2223-2232

McGregor CE, Lambert CA, Greyling MM, Louw JH, Warnich L (2000) A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. Euphytica 113:135-144

McKenzie MJ, Sowokinos JR, Shea IM, Gupta SK, Lindlauf RR, Anderson JAD (2005) Investigations on the role of acid invertase and UDP-glucose pyrophosphorylase in potato clones with varying resistance to cold-induced sweetening. American Journal of Potato Research 82:231-239

McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. Genetics 160:1113-1122

Menendez CM, Ritter E, Schäfer-Pregl R, Walkemeier B, Kalde A, Salamini F, Gebhardt C (2002) Cold sweetening in diploid potato: Mapping quantitative trait loci and candidate genes. Genetics 162:1423-1434

Meudt HM, Clarke AC (2007) Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. Trends in Plant Science 12:106-117

Milbourne D, Meyer R, Bradshaw JE, Baird E, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Molecular Breeding 3:127-136

Muth J, Hartje S, Twyman RM, Hofferbert HR, Tacke E, Prufer D (2008) Precision breeding for novel starch variants in potato. Plant Biotechnology Journal 6:576-584

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends in Plant Science 9:325-330

Neele A, Nab H, Louwes K (1991) Components of Visual Selection in Early Clonal Generations of a Potato Breeding Programme. Plant Breeding 106:89-98

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. Proceedings of the National Academy of Sciences of the United States of America 70:3321-3323

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30:190-193

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends in Genetics 18:83-90

Nsengimana J, Baret B, Haley CS, Visscher PM (2004) Linkage disequilibrium in the domesticated pig. Genetics 166:1395-1404

Oliemans WH (1988) Het brood van de armen: de geschiedenis van de aardappel temidden van ketters, kloosterlingen en kerkvorsten. SDU, 's-Gravenhage. 348 pp.

Olsson K, Svensson R, Roslund C-A (2004) Tuber components affecting acrylamide formation and colour in fried potato: variation by variety, year, storage temperature and storage time. Journal of the Science of Food and Agriculture 84:447-458

Ortiz R (1998) Potato breeding via ploidy manipulations. In: Janick J (ed) Plant Breeding Reviews. John Wiley & Sons, Inc.

Palaisa KA, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proceedings of the National Academy of Sciences of the United States of America 101:9885-9890

Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795-1806

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. TAG Theoretical and Applied Genetics 109:508-514

Park TH, Vleeshouwers V, Kim JB, Hutten RCB, Visser RGF (2005) Dissection of foliage and tuber late blight resistance in mapping populations of potato. Euphytica 143:75-83

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726

Pavek J, Corsini D (2001) Utilization of potato genetic resources in variety development. American Journal of Potato Research 78:433-441

Perrier X, Jacquemoud-Collet JP (2006) DARWIN software. http://darwinciradfr/darwin

Piepho H-P (2001) Exploiting quantitative information in the analysis of dominant markers. TAG Theoretical and Applied Genetics 103:462-468

Piepho HP (2005) Statistical tests for QTL and QTL-by-environment effects in segregating populations derived from line crosses. TAG Theoretical and Applied Genetics 110:561-566

Piepho HP, Koch G (2000) Co-dominant analysis of banding data from a dominant marker system by normal mixtures. Genetics 155:1459-1468

Plaisted RL, Hoopes RW (1989) The past record and future prospects for the use of exotic potato germplasm. American Potato Journal 66:603-627

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2:225-238

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945-959

Provan J, Kumar A, Shepherd L, Powell W, Waugh R (1996a) Analysis of intra-specific somatic hybrids of potato (*Solanum tuberosum*) using simple sequence repeats. Plant Cell Reports 16:196-199

Provan J, Powell W, Waugh R (1996b) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). TAG Theoretical and Applied Genetics 92:1078-1084

Raboin L-M, Pauquet J, Butterfield M, D'Hont A, Glaszmann J-C (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. TAG Theoretical and Applied Genetics 116:701-714

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends in Genetics 20:103-111

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences of the United States of America 98:11479-11484

Rios D, Ghislain M, Rodriguez F, Spooner DM (2007) What Is the Origin of the European Potato? Evidence from Canary Island Landraces. Crop Science 47:1271-1280

Rouppe van der Voort J, Lindeman W, Folkertsma R, Hutten R, Overmars H, van der Vossen E, Jacobsen E, Bakker J (1998) A QTL for broad-spectrum resistance to cyst nematode species (*Globodera* spp.) maps to a resistance gene cluster in potato. TAG Theoretical and Applied Genetics 96:654-661

Rouppe van der Voort J, van der Vossen E, Bakker E, Overmars H, van Zandvoort P, Hutten R, Lankhorst RK, Bakker J (2000) Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. TAG Theoretical and Applied Genetics 101:1122-1130

Salaman RN (1926) Potato varieties. Cambridge University Press, Cambridge

Salaman RN (1949) The history and social influence of the potato. Cambridge University Press, Cambridge

Schäfer-Pregl R, Ritter E, Concilio L, Hesselbach J, Lovatti L, Walkemeier B, Thelen H, Salamini F, Gebhardt C (1998) Analysis of quantitative trait loci (QTLs) and quantitative trait alleles (QTAs) for potato tuber yield and starch content. TAG Theoretical and Applied Genetics 97:834-846

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341-1345

Semon M, Nielsen R, Jones MP, McCouch SR (2005) The population structure of African cultivated rice *Oryza glaberrima* (Steud.): evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. Genetics 169:1639-1647

Sharopova N (2008) Plant simple sequence repeats: distribution, variation, and effects on gene expression. Genome 51:79-90

Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004a) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. TAG Theoretical and Applied Genetics 108:217-224

Simko I, Costanzo S, Ramanjulu V, Christ BJ, Haynes KG (2006b) Mapping polygenes for tuber resistance to late blight in a diploid *Solanum phureja* x *S. stenotomum* hybrid population. Plant Breeding 125:385-389

Simko I, Haynes KG, Ewing EE, Costanzo S, Christ BJ, Jones RW (2004b) Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. Molecular Genetics and Genomics 271:522-531

Simko I, Haynes KG, Jones RW (2006a) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics 173:2237-2245

Simko I, van den Berg JH, Vreugdenhil D, Ewing EE (2008) Mapping loci for chlorosis associated with chlorophyll b deficiency in potato. Euphytica 162:99-107

Skøt LL (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* L. Molecular Breeding 15:233-245

Slatkin M (2008) Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9:477-485

Sliwka J, Wasilewicz-Flis I, Jakuczun H, Gebhardt C (2008) Tagging quantitative trait loci for dormancy, tuber shape, regularity of tuber shape, eye depth and flesh colour in diploid potato originated from six *Solanum* species. Plant Breeding 127:49-55

Sørensen K, Kirk H, Olsson K, Labouriau R, Christiansen J (2008) A major QTL and an SSR marker associated with glycoalkaloid content in potato tubers from *Solanum tuberosum* × *S. sparsipilum* located on chromosome *I*. TAG Theoretical and Applied Genetics 117:1-9

Sørensen KK, Madsen MH, Kirk HG, Madsen DK, Torp AM (2006) Linkage and quantitative trait locus mapping of foliage late blight resistance in the wild species *Solanum vernei*. Plant Breeding 125:268-276

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ (2005) A single domestication for potato based on multi-locus amplified fragment length polymorphism genotyping. Proceedings of the National Academy of Sciences of the United States of America 102:14694-14699

Stich B, Melchinger AE, Heckenberger M, Möhring J, Schechert A, Piepho H-P (2008c) Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). TAG Theoretical and Applied Genetics 117:1167-1179

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. TAG Theoretical and Applied Genetics 111:723-730

Stich B, Mohring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008a) Comparison of Mixed-Model Approaches for Association Mapping. Genetics 178:1745-1754

Stich B, Piepho H-P, Schulz B, Melchinger A (2008b) Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). TAG Theoretical and Applied Genetics 117:947-954

Suenaga E, Nakamura H (2004) Rapid Individual Identification by Minisatellite Variant Repeat (MVR)-PCR at *D1S8* Locus Using Exponential Law. Analytical Sciences 20:1629-1633

Tai G, Young D (1984) Early generation selection for important agronomic characteristics in a potato breeding population. American Journal of Potato Research 61:419-434

Tang J, Leunissen J, Voorrips R, van der Linden CG, Vosman B (2008) HaploSNPer: a web-based allele and SNP detection tool. BMC Genetics 9:23

Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JAM (2006) QualitySNP: A pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. BMC Bioinformatics 7

Tanksley SD (1993) Mapping Polygenes. Annual Review of Genetics 27:205-233

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17:6463-6471

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome *1* of maize (*Zea mays* ssp. *mays* L.). Proceedings of the National Academy of Sciences of the United States of America 98:9161-9166

Thompson P, Mendoza H, Plaisted R (1983) Estimation of genetic parameters for characters related to potato propagation by true seed (TPS) in an *andigena* population. American Journal of Potato Research 60:393-401

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28:286-289

Trognitz F, Manosalva P, Gysin R, Nino-Liu D, Simon R, Herrera MD, Trognitz B, Ghislain M, Nelson R (2002) Plant defense genes associated with quantitative resistance to potato late blight in *Solanum phureja* x dihaploid *S. tuberosum* hybrids. Molecular Plant-Microbe Interactions 15:587-597

van Berloo R, Hutten R, van Eck H, Visser R (2007) An Online Potato Pedigree Database Resource. Potato Research 50:45-57

van Berloo R, Zhu A, Ursem R, Verbakel H, Gort G, van Eeuwijk F (2008) Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. TAG Theoretical and Applied Genetics 117:89-101

van der Beek JG, Verkerk R, Zabel P, Lindhout P (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome *1*. TAG Theoretical and Applied Genetics 84:106-112

van Eck HJ, Jacobs JME, Stam P, Ton J, Stiekema WJ, Jacobsen E (1994b) Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics 137:303-309

van Eck HJ, Jacobs JME, van den Berg PMMM, Stiekema WJ, Jacobsen E (1994a) The inheritance of anthocyanin pigmentation in potato (*Solanum tuberosum* L.) and mapping of tuber skin colour loci using RFLPs. Heredity 73:410 -421

van Eeuwijk FA, Denis JB, Kang MS (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: Kang MS, Gauch HGJ (eds) Genotype-by-environment interaction: new perspectives. CRC Press, Boca Raton, FL, pp 15-49

van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Australian Journal of Agricultural Research 56:883-894

van Eijk MJ, Broekhof JL, van der Poel HJ, Hogers RC, Schneiders H, Kamerbeek J, Verstege E, van Aart JW, Geerlings H, Buntjer JB, van Oeveren AJ, Vos P (2004) SNPWave: a flexible multiplexed SNP genotyping technology. Nucleic Acids Research 32: e47

van Marle J, de Vries R, Wilkinson E, Yuksel D (1997) Sensory evaluation of the texture of steam-cooked table potatoes. Potato Research 40:79-90

van Os H, Andrzejewski S, Bakker E, Barrena I, Bryan GJ, Caromel B, Ghareeb B, Isidore E, de Jong W, van Koert P, Lefebvre V, Milbourne D, Ritter E, Rouppe van der Voort JNAM, Rousselle-Bourgeois F, van Vliet J, Waugh R, Visser RGF, Bakker J, van Eck HJ (2006) Construction of a 10,000-marker ultra dense genetic recombination map of potato: Providing a framework for accelerated gene isolation and a genome-wide physical map. Genetics 173:1075-1087

Vargas M, Crossa J, van Eeuwijk FA, Ramirez ME, Sayre K (1999) Using Partial Least Squares Regression, Factorial Regression, and AMMI Models for Interpreting Genotype x Environment Interaction. Crop Science 39:955-967

Vargas M, van Eeuwijk F, Crossa J, Ribaut J-M (2006) Mapping QTLs and QTL × environment interaction for CIMMYT maize drought stress program using factorial regression and partial least squares methods. TAG Theoretical and Applied Genetics 112:1009-1023

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23:48-55

Verbeke G, Molenberghs G (2000) Mixed models for longitudinal data. Springer

Visker M, Keizer LCP, Van Eck HJ, Jacobsen E, Colon LT, Struik PC (2003) Can the QTL for late blight resistance on potato chromosome *5* be attributed to foliage maturity type? TAG Theoretical and Applied Genetics 106:317-325

Visker MHPW, Heilersig HJB, Kodde LP, van de Weg WE, Voorrips RE, Struik PC, Colon LT (2005) Genetic linkage of QTLs for late blight resistance and foliage maturity type in six related potato progenies. Euphytica 143:189-199

Voltas J, van Eeuwijk FA, Igartua E, García del Moral LF, Molina-Cano JL, Romagosa I (2002) Genotype by environment interaction and adaptation in barley breeding: Basic concepts and methods of analysis. In: Slafer GA, Molina-Cano JL, Savin R, J.L. A, Romagosa I (eds) Barley Science: Recent Advances from Molecular Biology to Agronomy of Yield and Quality The Harworth Press Inc., New York, pp 205-241.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414

Wang J, McClean PE, Lee R, Goos RJ, Helms T (2008) Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. TAG Theoretical and Applied Genetics 116:777-787

Wang-Pruski G, Nowak J (2004) Potato after-cooking darkening. American Journal of Potato Research 81:7-16

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American Journal of Human Genetics 44:388-396

Weedon MN, Frayling TM (2008) Reaching new heights: insights into the genetics of human stature. Trends in Genetics 24:595-603

Werij JS, Kloosterman B, Celis-Gamboa C, de Vos CHR, America T, Visser RGF, Bachem CWB (2007) Unravelling enzymatic discoloration in potato through a combined approach of candidate genes, QTL, and expression analysis. TAG Theoretical and Applied Genetics 115:245-252

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ESIV (2004) Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. Plant Cell 16:2719-2733

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M, Powell W (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (*Pteridophyta*) using AFLPs and SSRs from anonymous and transcribed gene regions. Molecular Ecology 14:1681-1695

Yencho GC, Kowalski SP, Kobayashi RS, Sinden SL, Bonierbale MW, Deahl KL (1998) QTL mapping of foliar glycoalkaloid aglycones in *Solanum tuberosum* x *S. berthaultii* potato progenies: quantitative variation and plant secondary metabolism. TAG Theoretical and Applied Genetics 97:563-574

Yildirim M, Çalişkan C (1985) Genotype X environment interactions in potato (*Solanum tuberosum* L.). American Journal of Potato Research 62:371-375

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203-208

Zhao H, Nettleton D, Soller M, Dekkers JCM (2005) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genetical Research 86:77-87

Zhao J, Paulo MJ, Jamar D, Lou P, van Eeuwijk FA, Bonnema G, Vreugdenhil D, Koornneef M (2007) Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. Genome 50:963

Zhao K, Aranzana M, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An *Arabidopsis* Example of Association Mapping in Structured Samples. PLoS Genetics 3:e4

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. The Plant Genome 1:5-20

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134

Summary

The results of a four year project within the Centre for BioSystems Genomics (www.cbsg.nl), entitled “Association mapping and family genotyping in potato” are described in this thesis. This project was intended to investigate whether a recently emerged methodology, association mapping, could provide the means to improve potato breeding efficiency.

In an attempt to answer this research question a set of potato cultivars representative for the commercial potato germplasm was selected. In total 240 cultivars and progenitor clones were chosen. In a later stage this set was expanded with 190 recent breeds contributed by five participating breeding companies which resulted in a total of 430 genotypes.

In a pilot experiment, the results of which are reported in Chapter 2, a subset of 220 of the abovementioned 240 cultivars and progenitor clones was used. Phenotypic data was retrieved through contributions of the participating breeding companies and represented summary statistics of recent observations for a number of traits across years and locations, calculated following company specific procedures. With AFLP marker data, in the form of normalised log-transformed band intensities, obtained from five well-known primer combinations, the extent of linkage disequilibrium (LD), using the r^2 statistic, was estimated. Population structure within the set of 220 cultivars was analysed by deploying a clustering approach. No statistically supported population structure was revealed and the LD seemed to decay at about 3cM with this set of marker data. Furthermore, marker-trait associations were investigated by fitting single marker regression models for phenotypic traits on marker band intensities with and without correction for population structure. Population structure correction was performed in a straightforward way by incorporating a design matrix into the model assuming that each breeding company represented a different breeding germplasm pool. The potential of association mapping in tetraploid potato has been demonstrated in this pilot experiment, because existing phenotypic data, a modest number of AFLP markers, and a relatively straightforward statistical analysis allowed identification of interesting associations for a number of agro-morphological and quality traits.

These preliminary results encouraged us to engage into an encompassing genome-wide association mapping study in potato. Two association mapping panels were compiled. One panel comprising 205 genotypes, all of which were also present in the set used for the pilot experiment, and another panel containing in total 299 genotypes including the entire set of 190 recent breeds together with a series of standard cultivars, about 100 of which are in common with the first panel. Phenotypic data for the association panel with 205 genotypes were obtained in a field trial performed in 2006 in Wageningen at two locations with two replicates (“2006 field trial”). Phenotypic data for the other panel with 299 genotypes was contributed by the five breeding companies and consisted of multi-year-multi-location data obtained during generations of clonal selection. The 2006 field trial data were nicely balanced, because the trial was designed in that way. The historical breeding data set was highly unbalanced. Analysis of these two differing phenotypic data sets was performed to deliver insight in variance components for the genotypic main effects and the genotype-by-environment interaction (GEI), besides the estimation of genotype main effects across environments. Both phenotypic data sets were analysed separately within a mixed model framework including terms for GEI. In Chapter 3 we describe both phenotypic data sets by comparing variance components, heritabilities

(=repeatabilities), intra data set relationships and inter data set relationships. Broader aspects related to phenotypic data sets and their analysis are discussed as well.

To retrieve information about hidden population structure and genetic relatedness, and to estimate the extent of LD in potato germplasm, we used marker information generated with 41 AFLP primer combinations and 53 microsatellite loci on a collection of 430 genotypes. These 430 genotypes contained all genotypes present in the two association mapping panels introduced before plus a few extra genotypes to increase potato germplasm coverage. Two methods were used: a Bayesian approach and a distance-based clustering approach. Chapter 4 describes the results of this exercise. Both strategies revealed a weak level of structure in our material. Groups were detected which complied with criteria such as their intended market niche, as well as groups differing in their year of first registration on a national list. LD, using the r^2 statistic, appeared to decay below the threshold of 0.1 across linkage groups at about 5cM on average. The results described in Chapter 4 are promising for association mapping research in potato. The odds are reasonable that useful marker-trait associations can be detected and that the potential mapping resolution will suffice for detection of QTL in an association mapping context.

In Chapter 5 a comprehensive genome-wide association mapping study is presented. The adjusted genotypic means obtained from two association mapping panels as a result of phenotypic analysis performed in Chapter 3 were combined with marker information in two association mapping models. Marker information consisted of normalised log-transformed band intensities of 41 AFLP primer combinations and allele dosage information from 53 microsatellite loci. A baseline model without correction for population structure and genetic relatedness and a more advanced model with correction for population structure and genetic relatedness were applied. Population structure and genetic relatedness were estimated using available marker information. Interesting QTL could be identified for 19 agro-morphological and quality traits. The observed QTL partly confirm previous studies e.g. for tuber shape and frying colour, but also new QTL have been detected e.g. for after baking darkening and enzymatic browning.

In the final chapter, the general discussion, results of preceding chapters are evaluated and their implications for research as well as breeding are discussed.

Samenvatting

De resultaten van een vierjarig project binnen het Centre for BioSystems Genomics (www.cbsg.nl), getiteld “Association mapping and family genotyping in potato“, worden beschreven in deze doctoraatsthesis. Dit project had als doel te onderzoeken of een recent ontwikkelde methodiek, associatieanalyse, een alternatief kan zijn voor traditionele QTL analyse. De laatste is gebaseerd op de analyse van nakomelingschappen van kruisingen tussen twee ouders, terwijl associatieanalyse zich richt op de analyse van collecties van genotypen zonder duidelijke nakomelingstructuur.

Om deze onderzoeksvraag te beantwoorden werd een collectie aardappelcultivars, representatief voor de commerciële aardappelgenenpool samengesteld. In eerste instantie werden 240 cultivars en introgressieklonen geselecteerd. In een later stadium werd deze set uitgebreid met 190 recente kwekersklonen aangedragen door vijf deelnemende aardappelveredelingsbedrijven tot een totaal van 430 genotypen.

In een proefexperiment waarvan de resultaten in Hoofdstuk 2 zijn gerapporteerd, werden 220 van de bovengenoemde 240 cultivars en introgressieklonen gebruikt. Fenotypische gegevens werden verkregen via de veredelingsbedrijven in de vorm van langjarige gemiddelden van observaties voor een aantal kenmerken, verzameld over jaren en locaties heen, en berekend volgens bedrijfsspecifieke procedures. Met AFLP merker gegevens, in de vorm van genormaliseerde en logaritmisch omgezette bandintensiteiten, gegenereerd met vijf bekende primercombinaties werd de mate waarin merker loci al dan niet in koppelingsevenwicht zijn binnen het aardappelgenoom geschat. Hierbij werd gebruik gemaakt van de meeteenheid r^2, de gekwadrateerde correlatie tussen merkerparen. Populatiestructuur binnen de 220 cultivars werd onderzocht met een clusteringmethode. Geen statistisch ondersteunde populatiestructuur werd gevonden en koppelingsevenwicht tussen merker loci leek op te treden vanaf een genetische afstand van ca. 3cM binnen deze merkerset. Verder werd gezocht naar associaties tussen merkers en kenmerken door het merkersgewijs toepassen van regressiemodellen voor fenotypische kenmerken op de merker bandintensiteiten met en zonder correctie voor populatiestructuur. Populatiestructuur correctie werd uitgevoerd door het opnemen van een ontwerpmatrix in het associatieanalyse model onder de veronderstelling dat elk veredelingsbedrijf een verschillende genenpool vertegenwoordigde. De mogelijkheden voor associatieanalyse binnen tetraploïde aardappel werden geïllustreerd met dit proefexperiment omdat met bestaande fenotypische gegevens, een bescheiden aantal AFLP merkers en een relatief eenvoudige statistische analyse het identificeren van interessante associaties voor een aantal agromorfologische en kwaliteitskenmerken mogelijk bleek.

Deze initiële resultaten zetten ons aan om een genoombrede associatiestudie in aardappel uit te voeren. Daartoe werden twee rassensets voor associatieanalyse samengesteld. Een eerste set met 205 genotypen, die allen ook aanwezig waren in de set die voor het proefexperiment werd gebruikt. De tweede set telde 299 genotypen met daarin alle 190 recente kwekersklonen en een reeks standaardrassen en had ongeveer 100 genotypen gemeenschappelijk met de eerste set. Fenotypische gegevens voor de associatieset met 205 genotypen werden verzameld tijdens een praktijkexperiment in 2006 in Wageningen op twee locaties met twee herhalingen (“praktijkexperiment 2006“). De fenotypische gegevens voor de tweede associatieset met 299 genotypen werden aangedragen door de vijf aardappelveredelingsbedrijven in de vorm van

multi-jaar-multi-locatie gegevens die tijdens generaties van klonale selectie werden verzameld. De praktijkexperiment 2006 gegevens waren keurig gebalanceerd omdat de proef zodanig was ontworpen. De multi-jaar-multi-locatie data set was daarentegen sterk ongebalanceerd. De analyse van deze twee verschillende fenotypische data sets werd uitgevoerd om inzicht te verwerven in de variantiecomponenten voor de genotypische hoofdeffecten en de genotype-milieu interactie effecten (GMI). Daarnaast werden ook milieuafhankelijke genotypeneffecten geschat. Beide fenotypische data sets werden afzonderlijk geanalyseerd met behulp van gemengde statistische modellen inclusief termen voor GMI. In Hoofdstuk 3 beschrijven we beide fenotypische data sets door zowel hun variantiecomponenten te vergelijken, als de mate waarin eigenschappen erfelijk overdraagbaar of experimenteel herhaalbaar blijken. Intra data set relaties en inter data set relaties worden evengoed bestudeerd. Tevens worden bredere aspecten aangaande fenotypische data sets en hun analyse besproken.

Om informatie over verborgen populatiestructuur en genetische verwantschap te verkrijgen, en om de mate waarin merkers zich al dan niet in koppelingsevenwicht in de aardappelgenenpool bevinden in te schatten, gebruikten we merkerinformatie gegenereerd met 41 AFLP primercombinaties en 53 microsatellieten over 430 genotypen. Deze 430 genotypen bevatten alle genotypen aanwezig in de twee eerder vermelde associatiesets plus een aantal extra genotypen om een hogere dekkingsgraad over de aardappelgenenpool te verwerven. Twee methodes werden gebruikt: een Bayesiaanse benadering en een afstandsgebaseerde clusteringbenadering. Hoofdstuk 4 vermeldt de resultaten van deze exercitie. Beide strategieën toonden een zwak niveau van populatiestructuur aan in ons materiaal. Groepen werden ontdekt die beantwoordden aan criteria zoals hun voorgenomen marktsegment, evenals groepen die in hun jaar van eerste registratie op een nationale rassenlijst verschilden. De mate van koppelingsonevenwicht tussen merkers berekend met r^2 daalde, gemiddeld over alle koppelingsgroepen heen, bij een genetische afstand van ongeveer 5cM onder de drempelwaarde van 0.1. De in Hoofdstuk 4 beschreven resultaten zijn veelbelovend voor associatiestudies binnen aardappel. Wellicht zullen nuttige merker-kenmerk associaties worden gedetecteerd en de potentiële karteringsresolutie zal volstaan voor het opsporen van loci betrokken bij kwantitatieve kenmerken (QTL) binnen een associatieanalyse context.

In Hoofdstuk 5 presenteren we de resultaten van een genoombrede associatiestudie. De tijdens de in Hoofdstuk 3 beschreven fenotypische analyses geschatte genotypische hoofdeffecten werden gecombineerd met merkerinformatie in twee associatieanalyse modellen. Merker informatie was aanwezig als genormaliseerde logaritmisch omgezette bandintensiteiten afkomstig van 41 AFLP primercombinaties en als alleldosis informatie van 53 microsatellieten. Een basismodel zonder correctie voor populatiestructuur of genetische verwantschap en een geavanceerder statistisch model met correctie voor populatiestructuur en genetische verwantschap werden opgesteld. Populatiestructuur en genetische verwantschap werden geschat met behulp van de beschikbare merkerinformatie. Interessante QTL werden geïdentificeerd voor 19 agromorfologische en kwaliteitskenmerken. De gedetecteerde QTL bevestigen gedeeltelijk voorgaande studies b.v. voor knolvorm en bakkleur, maar ook nieuwe QTL werden ontdekt b.v. voor verkleuring na bakken en enzymatische verkleuring.

In het laatste hoofdstuk, de algemene discussie, worden de resultaten van voorafgaande hoofdstukken geëvalueerd en hun implicaties voor onderzoek en veredeling besproken.

Résumé

Les résultats d'un projet de quatre ans au Centre for BioSystems Genomics (www.cbsg.nl), intitulé "Association mapping and family genotyping in potato" sont décrits dans cette thèse de doctorat. Ce projet était prévu pour rechercher si une méthodologie récemment développée, l'analyse d'associations, pourrait fournir les moyens d'améliorer la pomme de terre avec une efficacité plus élevée.

Afin de répondre à cette question scientifique, un assortiment de cultivars de pomme de terre représentatif des ressources génétiques de la pomme de terre commerciale a été compilé. D'abord 240 cultivars et clones ancestraux ont été choisis. Plus tard cet assortiment a été accru de 190 clones récemment améliorés, provenant de cinq entreprises collaboratrices ce qui résulta en un total de 430 génotypes.

Dans une expérience préliminaire, dont les résultats sont rapportés en Chapitre 2, un sous-ensemble de 220 des 240 cultivars et clones ancestraux a été employé. Les données phénotypiques ont été apportées par les entreprises participantes et représentaient une synthèse statistique des observations récentes sur un certain nombre de caractères, calculée selon des procédures spécifique aux entreprises. Les données de marqueurs d'AFLP (intensité des bandes) normalisées après log transformation et obtenues à partir de cinq combinaisons d'amorces bien connues ont permis l'estimation du déséquilibre de liaison (LD) par r^2 (la corrélation au carré). La structure de population au sein de cet ensemble de 220 cultivars a été analysée en déployant une approche en cluster. L'analyse n'indiquait aucune structure statistiquement significative dans la population et le LD semblait décroître, au-dessous du seuil de 0.1, à une distance génétique d'environ 3cM avec cet ensemble de marqueurs. En outre, les associations entre caractères et marqueurs ont été étudiées par modèles de régression pour chaque marqueur séparément, pour des caractères phénotypiques sur les intensités de bandes, avec et sans prise en compte de la structure de la population. La prise en compte de la structure de la population a été accomplie en incorporant directement une matrice de conception dans le modèle supposant que chaque entreprise représentait un pool de ressources génétiques différent. Le potentiel de la cartographie par association pour la pomme de terre tétraploïde a été démontré dans cette expérience préliminaire, puisqu'un nombre modeste de marqueurs d'AFLP, des données phénotypiques existantes, et une analyse statistique relativement simple ont permis l'identification des associations intéressantes pour un certain nombre de caractères agro-morphologiques et de qualité.

Ces résultats initiaux nous ont encouragés à engager une étude de cartographie par association sur la globalité du génome de la pomme de terre. Deux ensembles de cultivars prévus pour la cartographie par association ont été assemblés. Un panel comportait 205 génotypes, également présents dans l'ensemble employé pour l'expérience préliminaire, tandis que l'autre panel contenait au total 299 génotypes dont 190 clones récemment améliorés et une série de cultivars de référence. Environ 100 génotypes étaient en commun dans les deux panels. Des données phénotypiques pour le panel d'association des 205 génotypes ont été obtenues lors d'une étude de terrain exécutée en 2006 à Wageningen. Lors de l'étude, chaque génotype était présent en double à deux endroits ("l'étude de terrain 2006"). Des données phénotypiques pour l'autre panel des 299 génotypes ont été collectées auprès des cinq entreprises et se composaient de données obtenues pendant plusieurs années de sélection

clonale à différents sites. Ces données sont fortement déséquilibrées alors que les données de l'étude de terrain 2006 étaient bien équilibrées. Ces deux ensembles de données phénotypiques ont été analysés pour estimer la variance des effets génotypiques et leur interaction avec leur environnement spécifique (GEI). L'effet génotypique fut aussi estimé à travers tous les environnements. Les deux ensembles de données phénotypiques ont été analysés séparément par des modèles mixtes comprenant des termes de GEI. Le Chapitre 3 détaille les deux ensembles de données phénotypiques en comparant les composants des variances, les héritabilités (=répétabilités) et les relations inter et intra ensemble de données. Les aspects plus complexes liés aux ensembles de données phénotypiques et leur analyse sont aussi débattus.

Afin de rechercher des informations sur la structure de la population, la parenté génétique et l'ampleur du LD dans les ressources génétiques de pomme de terre, nous avons utilisé des informations de marqueurs générées avec 41 combinaisons d'amorces d'AFLP et 53 microsatellites sur une collection de 430 génotypes. Ces 430 génotypes contenaient tous les génotypes actuels des deux panels d'association présentés ci-dessus et quelques génotypes supplémentaires pour augmenter la couverture de la diversité génétique. Deux méthodes ont été employées: une approche Bayésienne et une approche en cluster. Le Chapitre 4 rapporte les résultats de cet exercice. Les deux stratégies ont révélé une faible structure au sein de la population. Plusieurs génotypes ont été groupés par rapport à leur niche commerciale, ou leur année d'enregistrement sur les listes nationales. Utilisant le r^2, le LD semblait décroître au-dessous du seuil de 0.1 à travers les groupes de liaison à une distance génétique d'environ 5cM en moyenne. Les résultats présentés en Chapitre 4 montrent que chez la pomme de terre, la cartographie par association est envisageable. Il est très probable que des associations avantageuses entre des caractères phénotypiques et des marqueurs moléculaires puissent être détectées. En plus il est probable que la résolution de cartographie potentielle sera suffisante à la détection de QTL lors de cartographie par association.

Dans le Chapitre 5 une étude compréhensive est présentée de cartographie par association sur la globalité du génome. Les valeurs statistiques obtenues par les deux panels d'association suite à l'analyse phénotypique présentée en Chapitre 3, ont été combinées avec des informations de marqueurs dans deux modèles de cartographie par association. Ces informations sont constituées d'une part des intensités de bande normalisées après log transformation provenant de 41 combinaisons d'amorces d'AFLP et d'autre part des informations de dosage d'allèle de 53 microsatellites. Deux modèles ont été utilisés, un modèle de base sans correction pour la structure de la population et la parenté génétique, et un modèle plus avancé avec correction pour ces deux critères. La structure de la population et la parenté génétique ont été estimées en utilisant les informations de marqueurs disponibles. Des QTL intéressants ont pu être identifiés pour 19 caractères agro-morphologiques et de qualité. Les QTL observés confirment en partie les études précédentes par exemple pour la forme des tubercules et la couleur des frites, mais également de nouveaux QTL ont été détectés comme ceux pour la décoloration après friture et pour la décoloration enzymatique.

Dans le dernier chapitre, discussion générale, les résultats des chapitres précédents sont évalués et leurs implications pour la recherche scientifique aussi bien que pour l'amélioration de la pomme de terre sont discutées.

Dankwoord

De lay-out perikelen zijn achter de rug nu. De tijd is dus rijp om terug te blikken op de afgelopen vier jaar en iedereen die van direct of indirect belang was voor het tot stand komen van dit boekje te bedanken. Maar ik begin met een anekdote.

Mijn eerste kennismaking met de Wageningse atmosfeer vond plaats in oktober 2004. Ik was uitgenodigd voor een job interview bij het laboratorium voor plantenveredeling. Na de lange reis van West-Vlaanderen tot Wageningen, bleek het laboratorium voor plantenveredeling (toen nog op de Binnenhaven) onvindbaar. Ik was verdwaald op het Agro-Business Park. In zo accentloos mogelijk Nederlands vroeg ik de weg. De mensen die ik lastig viel bleken het echter niet te kennen of misschien begrepen ze me gewoon niet. Ik zal het nooit weten. Net op tijd vond ik dan toch de Binnenhaven. Naar Belgische traditie had ik mij in een net pak gehesen naar aanleiding van het interview. Een ontnuchtering volgde echter al snel: niemand van mijn interviewers, noch Herman noch Fred, droeg een pak. Typisch Wageningen zou ik later al snel ontdekken. Diezelfde avond nog op de terugweg ter hoogte van de grens kreeg ik een telefoontje van een Nederlandse student die een kamer in onderhuur had per november 2004. De verrassing was compleet, wist ik toen veel dat Herman al contact had opgenomen via het bulletin board met de mededeling dat ik dringend een kamer zocht in Wageningen…

Eerst en vooral wil ik mijn begeleiders bedanken. Zonder hun inzet, tot in de nachtelijke uren en het weekend toe op het laatst, en hun kritische geest was dit boekje nooit op tijd klaar geraakt. Herman, heel erg bedankt voor je inbreng! Bij elke discussie bleek je weer vol interessante en prikkelende vragen te zitten die het onderzoek ten goede kwamen. Je hebt me veel bijgebracht over aardappelgenetica. Je was ook altijd te vinden voor een al dan niet filosoferend gesprek. Ik moet wel bekennen dat ik al snel leerde om na vijven bij je langs te komen met vragen waar ik snel een antwoord op wou. Fred, bedankt voor alles! Onder jouw vleugels heb ik mij kunnen ontwikkelen tot een meer multifunctionele wetenschapper. Ik spreek nu niet meer alleen de taal van een bio-ingenieur met een genetische en biotechnologische achtergrond, maar heb ook geleerd te communiceren met statistici. Een enorme verrijking voor mijn wetenschappelijke bagage. Je bent ook bijzonder efficiënt. Nooit eerder had ik meetings waar binnen een beperkt tijdsbestek zoveel zaken op een professionele wijze konden behandeld worden. Toen je professor werd en je eigen groep mensen ging leiden bij Biometris vreesde ik even dat onze samenwerking en het project op een dood spoor zouden belanden. Gelukkig bleek die vrees ongegrond. Ik herinner me ook de borrels in de Vlaamsche Reus waar een grappige of cynische anekdote van jouw kant nooit ver weg was. Richard, hoewel je het vaak (te) druk had, was je toch altijd bij de pinken om problemen snel het hoofd te bieden, waarvoor mijn oprechte dank. Vooral tijdens het laatste jaar ben je nadrukkelijker op de voorgrond getreden wat er mee toe heeft geleid dat heel snel na het verstrijken van het vierde jaar de leesversie werd afgerond. Je verdient bovendien een pluim voor je snelle revisies.

Time to switch to English now. The enormous amount of marker data within this project could never have been produced by myself alone. That is obvious. Brigitte, my first (and only) MSc student, I enjoyed guiding your first steps into the lab and performing SSR experiments with you. I appreciated the often revealing conversations we had on history and culture while leaf picking in the greenhouse and I will never forget your first “experience” with chloroform. I wish you all the best with your PhD! The majority of SSRs were run by Krissana. Today I am still

impressed with the amount of data you assembled. It took me a year to analyse it all! I am happy that you started your own PhD now and I am confident that you will succeed. The last SSR repeats were generated by Hanneke. Thanks for being so precise and meticulous. I know it must have been hell to get all the samples organised. Thanks to Marleen from Greenomics for keeping the DNA Analyzers up and running. AFLP data were generated by Keygene where I especially want to thank Mariëlle and Rolf for a nice cooperation.

No association mapping without phenotypical data. Dirkjan, thanks for arranging greenhouse space and taking care of my plants. Ronald, thanks for your tips in organising field trials, the arrangements you took to order tubers and to arrange field trial sites and definitely thanks for maintaining the pedigree database. My appreciation to PPW for their help in planting and harvesting the 2006 field trial. Henk and Brigitte, thanks for your help with the determination of underwater weight of all 960 plots. Thanks to the wonderful “2006 field trial team”! What a tremendous job we did: collecting more than 5,000 samples in less than two days! I did not need St Nicholas to remind me to acknowledge your help. And of course I want to express my utmost gratitude to the breeding companies involved in project P8 of CBSG. Bedankt Guus, Nick, Jeroen, Jacqueline, Kees, Robert, Mariëlle, Nico, Sjefke en stagiaires voor het vrijgeven van jullie veldboek gegevens, voor de assistentie in de veldproef en voor alle waardevolle suggesties, kritische vragen en tips geopperd tijdens de vele meetings of gewoon over de telefoon. Ik heb veel bijgeleerd over aardappels, ook vanuit een veredelingsperspectief, en dit is grotendeels jullie verdienste! Een aardappel zal nooit meer zomaar een aardappel zijn voor mij. Verder gaat mijn dank ook uit naar de VAVI (Leo en Martin) voor het mee scoren van de frietsamples bij de veldproef van 2006.

No association mapping without statistics either. João, I enjoyed working with you! I learned a lot from discussions about mixed models and other statistical topics. But even more, I appreciated being able to talk to you whenever things were worrying me or when things did not work out the way I wanted them. Paul, you are amazing! You impersonated my GenStat helpdesk throughout the past four years and were of an incredible value during the last weeks of this thesis. I learned a lot about programming in GenStat thanks to you. But most importantly: you were my coach. How many hours did we spend talking on the phone? I have no clue, but each minute was well spent.

Dear officemates (Thierry, Arend, Ralph, Remco, Animesh, Mina, Alvaro, Jan and Dunia), I admire you. It must not have been easy, sharing an office with me. Thanks for maintaining a good atmosphere and for keeping up with my sometimes noisy behaviour (e.g. whistling, humming or finger tapping when listening to a nice song, or cursing when my computer was not behaving properly). Dear secretaries (Annie, Letty, Mariame), thanks for taking care of the administration! People often don’t realise how important that is! Dear lab staff, thanks for making the labs function properly and thanks for helping me out whenever I needed something.

I was part of several discussion groups: the statistical genetics group, the biodiversity group, the potato interest group. A warm thank you to all the colleagues that were part of these groups and were eager to comment on my research and its progress. Your ideas and suggestions have certainly been useful.

Starting a PhD in Wageningen was a hard nut to crack four years ago. Not knowing anybody here and leaving my friends behind was tough. But I must admit that I didn't regret at any moment the choice I made! What a bunch of wonderful friends would I not have met! For example my lunch and/or coffee break buddies: Bjorn (aka "Klossie"), Jeroen, Japie, Marieke, Benoit, Marcos, João, Remco, Antoine, Anoma, Estelle, Nico, Krissana, Mathieu, Pierre-Yves, Mariame, Adillah, Zheng, Luigi, Stefano, Alireza, Ningwen, Anitha, Hulya, Paula, Animesh, Arwa, Nasim and so many more. Thanks to the same people for sharing so many fun activities, e.g. het WK veldrijden, the Batavierenrace, We-days, fitness work-outs, weekend trips, birthday parties, drinks, beer tasting evenings, congresses, dinners,… Thanks to the Thai group for the many cosy Friday evenings at the Vlaamsche Reus and the very tasty dinners afterwards. I hope you all enjoyed the trip to Belgium, I surely had a great time showing you around in Ghent. Aan mijn "Murphy's Quiz Night" maatjes van het worst kaas scenario (Martijn, Jeroen, Ries, Olga en Frank): bedankt voor de vele leuke dinsdagavonden die vaak beëindigd werden met een slaapmutsje in de XL. Ooit gaan we die verdomde quiz eens winnen! Bedankt aan de festivalgroep voor de altijd weer gezellige sfeer. Bedankt aan mijn ex-huisgenoten van Psychotel, jullie hebben er in mijn eerste maanden voor gezorgd dat ik mij hier kon aarden en maakten me duidelijk dat appelmoes in Nederland geen volwaardig maaltijdonderdeel is. Heel erg bedankt aan mijn geweldige afdelingsgenoten van Asserpark 3B! Wat hebben jullie geleden de afgelopen maanden telkens als ik 's avonds laat chagrijnig van vermoeidheid, stress en honger thuis kwam. Gelukkig ben ik wat meer relaxt nu. Ik had nooit gedacht toen ik voor het eerst op 3B kwam dat ik er zolang zou blijven. Het feit dat we er met zijn vieren (Emmy, Anna, Myrthe en mezelf) tegelijk kwamen wonen en dat Martijn er nog lang is gebleven hebben daar zeker toe bijgedragen. Als ik binnenkort verhuis hoop ik dan ook dat we elkaar zullen blijven opzoeken en dat we de vriendschap die er nu is kunnen onderhouden.

Nico and Jeroen, my dearest paranymphs, it is great that you will assist me at my defence. Honestly, I could not have imagined defending my thesis without two of my best friends by my side. Nico (together with Pierre-Yves and Antoine) thanks for your assistance in the improvement of the French summary. I really appreciated the trip to Thailand, it was certainly one of my best travelling experiences so far. Good luck with your PhD and your forthcoming "little one"! Jeroen, hoe goed kan je met iemand opschieten? Menig gewicht hebben we opgedrukt en menig biertje gedronken. Het was steeds plezant om met jou als reisgezel op congres te gaan. Ik herinner me nog levendig de afsluitende avond in Carlow en de trip naar Zuid-Korea en China (in het gezelschap van Klossie) kon tellen als geslaagde reis. Ik vond het ook een hele eer toen ik als daggast aanwezig mocht zijn bij je huwelijk. Heel veel succes nog met je carrière bij Enza en je toekomstige promotie en veel geluk in Hoorn! Let's keep in touch!

Bert en Stephen, samen met mij de West-Vlaamse connectie of de drie musketiers toen we nog studenten waren op het boerenkot in Gent. Bedankt voor de bezoekjes, de reizen en citytrips die we samen gemaakt hebben en de quizavonden. Ook al zien we elkaar veel minder dan vroeger, als we elkaar zien is het net weer als vroeger. En ja Stephen, ik beloof dat ik je kom opzoeken in Zwitserland! Heel veel succes met jullie verdere carrières en tot binnenkort!

Aan mijn KLJ makkers uit Ingelmunster en omstreken. Jullie hebben er steeds voor gezorgd dat de weekenden dat ik terug in België was een feest waren. Om het met de gevleugelde woorden van Flip Kowlier te zeggen: "Min moaten, min moaten, ik zie ulder geern, wuk zoe kik doen zonder gidder?"

“Jij gaat nog in Nederland belanden!” Dit waren mijn moeders woorden kort na mijn afstuderen als bio-ingenieur. En zie: het was geen leugen. Ma, pa, het is zover! Ik ben klaar met studeren! Een oprechte “dank jullie wel” voor de onvoorwaardelijke steun die ik kreeg bij gelijk welke keuze die ik ooit maakte. Eén van de belangrijkste levenslessen die ik onthouden heb uit mijn opvoeding is dat je nergens komt zonder discipline en doorzettingsvermogen. Eens te meer bleek dat een waarheid als een koe. Ik kan me geen betere ouders voorstellen! Tot slot nog dit: “Meme, ik ben heel erg blij dat je er nog bij kunt zijn en ik hoop dat pepe Albert dit op een of andere manier toch ook nog meekrijgt. Dit is eigenlijk ook een beetje voor jullie!”

Björn

Curriculum Vitae

Björn Benny D'hoop was born on the 9th of November 1981 in Izegem (Belgium). After obtaining his high school diploma (VWO) in 1999 at the St.-Josephs College in Tielt (Belgium), he started his undergraduate studies in bio-engineering sciences with specialization cell and gene biotechnology at the Faculty of Agricultural and Applied Biological Sciences of the University of Ghent (Belgium). During this five year undergraduate period, he participated in a study project involving the composition and analysis of metabolic network models of bacterial fermentations at the Department of Applied Mathematics, Biometrics and Process Control. For his MSc thesis he constructed a genetic map of maize with AFLP markers at the Department of Plant Production. It is there that he became fascinated by applied plant research. In July 2004 he obtained his Master degree in Bioengineering Sciences with distinction. In November 2004 he started his PhD at the Laboratory of Plant Breeding of Wageningen University (the Netherlands) under the supervision of Dr. Ir. Herman J. van Eck, Prof. Dr. Fred A. van Eeuwijk and Prof. Dr. Richard G.F. Visser. The research focused on association mapping of agro-morphological and quality traits in potato. The results of this research project are presented in this PhD thesis. After defending his PhD thesis, he will continue working on association mapping in potato as a postdoc at the Laboratory of Plant Breeding of Wageningen University in the Netherlands.

List of publications

D'hoop B.B. (2008) Association mapping in tetraploid potato using SSRs and AFLPs. In: 17th Triennial conference of the EAPR, Abstracts of papers and posters, Brasov, Romania, 06-10 July 2008: p.376

D'hoop B.B., Paulo M.J., Mank R.A., van Eck H.J., van Eeuwijk F.A. (2008) Association mapping of quality traits in potato (*Solanum tuberosum* L.). Euphytica 161:47-60

D'hoop B.B., Paulo M.J., Kowitwanich K., Visser R.G.F., van Eck H.J., van Eeuwijk F.A. (2008) Association mapping in tetraploid potato using SSRs and AFLPs. In: Molecular mapping and marker assisted selection in plants, Programme and abstracts, Vienna, Austria, 03-06 February 2008: p.88

D'hoop B.B., Uwimana B., Paulo M.J., Visser R.G.F., van Eck H.J., van Eeuwijk F.A. (2007) Association mapping in tetraploid potato (*Solanum tuberosum* L.) using SSRs. In: The 4th Solanaceae Genome Workshop 2007, Jeju Island, Korea, 09-13 September 2007: p.75

D'hoop B.B., Paulo M.J., Vosman B., van Eck H.J., Visser R.G.F., van Eeuwijk F.A. (2006) Association mapping in tetraploid potato (*Solanum tuberosum* L.) – a pilot study. In: EAPR-EUCARPIA The science of selection: potato breeding methodology for the 21st century, Abstracts of papers and posters, Carlow, Ireland, 20-22 November 2006: p.9

D'hoop B.B., Paulo M.J., Vosman B., van Eck H.J., Visser R.G.F., Mank R.A., van Eeuwijk F.A. (2006) Association mapping in potato (*Solanum tuberosum* L.). Agriculturae Conspectus Scientificus 71Supp1: p.6

Submitted:
D'hoop B.B., Paulo M.J., Visser R.G.F., van Eck H.J., van Eeuwijk F.A. Phenotypic analyses of multi-environment data for two diverse tetraploid potato collections; comparing a designed set of trials with a historical set of trials.

D'hoop B.B., Paulo M.J., Kowitwanich K., Sengers M., Visser R.G.F., van Eck H.J., van Eeuwijk F.A. Population structure and LD in tetraploid potato.

D'hoop B.B., Keizer P.L.C., Paulo M.J., Visser R.G.F., van Eeuwijk F.A., van Eck H.J. Marker-trait associations in potato.

Education Statement of the Graduate School
Experimental Plant Sciences

Issued to: **Björn B. D'hoop**
Date: **March 02, 2009**
Group: **Laboratory of Plant Breeding, Wageningen University**

1) Start-up phase	*date*
► **First presentation of your project**	
Association mapping and family genotyping in potato	May 02, 2005
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
► **MSc courses**	
Modern statistics for the life sciences	Jan-Feb 2005
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*7.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD students day 2005, Nijmegen	Jun 02, 2005
EPS PhD students day 2006, Wageningen	Sep 19, 2006
PhD Joint Retreat 2008 (EPS, SDV and IMPRS), Wageningen	Oct 02-03, 2008
► **EPS theme symposia**	
EPS theme symposium 'Genome Plasticity', WU, Wageningen	Dec 09, 2004
EPS theme symposium 'Genome Plasticity', WU, Wageningen	Dec 09, 2005
EPS theme symposium 'Genome Plasticity', RU, Nijmegen	Dec 08, 2006
EPS theme symposium 'Genome Plasticity', LU, Leiden	Dec 07, 2007
► **NWO Lunteren days and other National Platforms**	
ALW meeting Lunteren 2005	Apr 04-05, 2005
ALW meeting Lunteren 2006	Apr 03-04, 2006
ALW meeting Lunteren 2007	Apr 02-03, 2007
ALW meeting Lunteren 2008	Apr 07-08, 2008
► **Seminars (series), workshops and symposia**	
CBSG summit 2005	Feb 21-22, 2005
► CBSG summit 2006	Mar 06-07, 2006
CBSG summit 2007	Feb 06-07, 2007
CBSG summit 2008 - delivering the deliverables	Mar 17-18, 2008
CBSG clustermeeting potato 2005	Sep 08, 2005
CBSG clustermeeting potato 2006	Oct 05, 2006
► CBSG clustermeeting potato 2007	Aug 31, 2007
Flying seminar: Joseph Ecker	Sep 26, 2005
Seminar on breeding and genetics (Dr. W.(Bill) G. Hill and Dr. Jack C.M. Dekkers	Apr 27, 2007
Workshop on patenting - CBSG	Jun 23, 2005
The truth of science - conference	Dec 08, 2005
Modeling for the Life Sciences - Lustrum symposium	Jun 21, 2007
Genomics Researchers Event 2007	Nov 29, 2007
Plant breeding research day	Jun 17, 2008
Genstat user day 2008	Jun 18, 2008
Statistical methods for linkage disequilibrium analysis	Jun 19, 2008
► **Seminar plus**	
Joseph Ecker	Sep 26, 2005
► **International symposia and congresses**	
XIII EUCARPIA Biometrics in Plant Breeding Section Congres Zagreb (Croatia)	Aug 30-Sep 01, 2006
EAPR 2006 The science of selection Carlow (Ireland)	Nov 20-22, 2006
Solanaceae 2007 Jeju Island (South Corea)	Sep 09-13, 2007
Molmap and MAS plants 2008 Vienna (Austria)	Feb 03-06, 2008
EAPR 2008 Triennual general meeting Brasov (Roumania)	Jul 06-10, 2008
► **Presentations**	
CBSG summit 2005 - oral	Feb 21-22, 2005
CBSG clustermeeting potato 2005 - oral	Sep 09, 2005
CBSG summit 2006 - oral	Mar 06-07, 2006
CBSG clustermeeting potato 2006 - oral	Oct 05, 2006
XIII EUCARPIA Biometrics in Plant Breeding Section Congres Zagreb (Croatia) - oral	Aug 30-Sep 01, 2006
EAPR 2006 The science of selection Carlow (Ireland) - oral	Nov 20-11, 2006
CBSG summit 2007 - oral	Feb 06-07, 2007
CBSG clustermeeting potato 2007 - oral	Aug 31, 2007
Solanaceae 2007 Jeju Island (South Corea) - poster + oral	Sep 09-13, 2007
EPS theme 4 Genome Plasticity symposium 2007	Dec 07, 2007
Molmap and MAS plants 2008 Vienna (Austria)	Feb 03-06, 2008
EPS PhD Joint Retreat 2008	Oct 02-03, 2008
► **IAB interview**	
► **Excursions**	
Organization of visit to McCain company for Plant Breeding on September, 15th 2008	Sep 15, 2008
Subtotal Scientific Exposure	*25.7 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Multivariate analysis	Apr 19-27, 2006
Bioinformatics - a user's approach	Mar 13-16, 2007
► **Journal club**	
Member of literature discussion group	2004-2007
► **Individual research training**	
Subtotal In-Depth Studies	*5.7 credits**

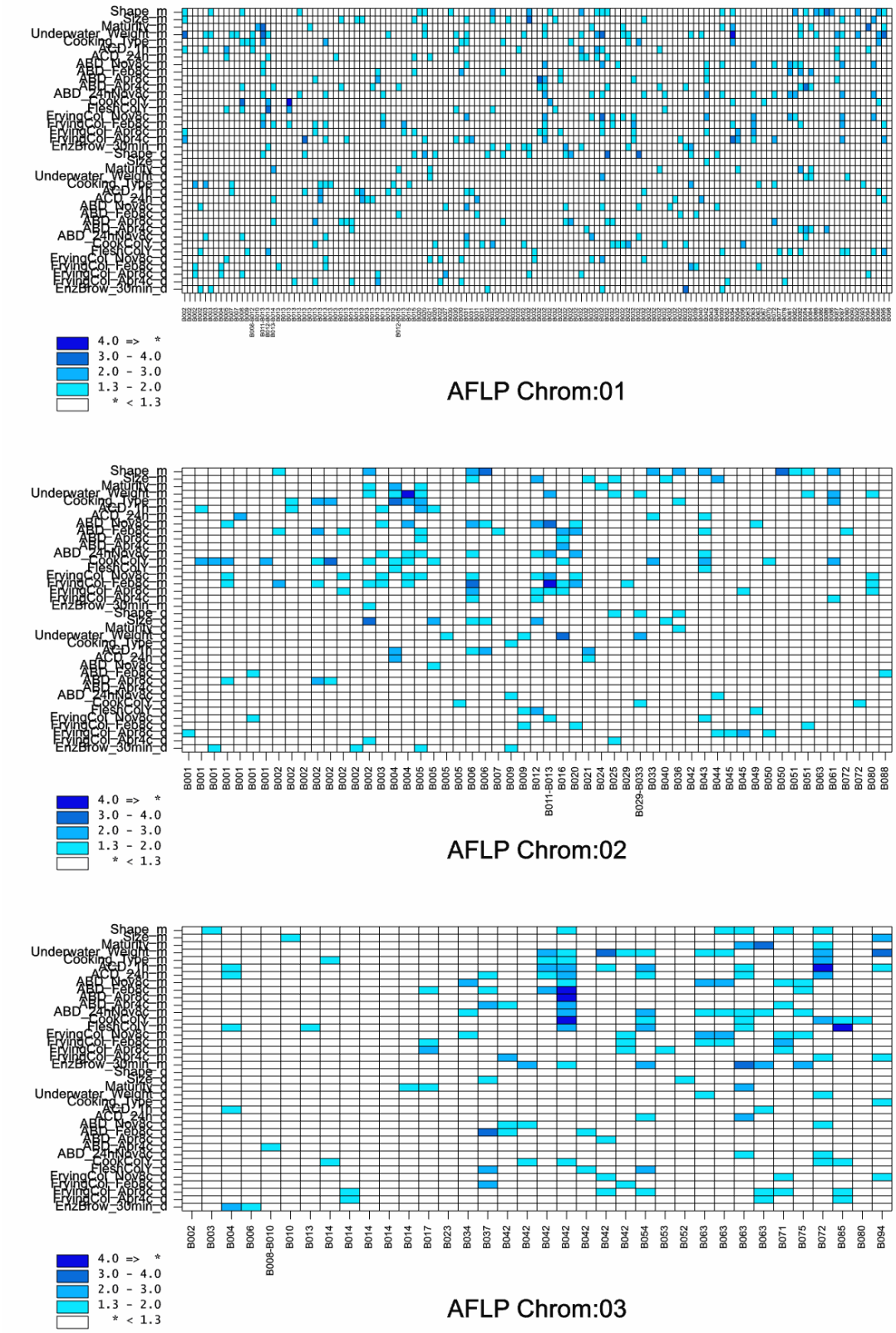
4) Personal development	*date*
► **Skill training courses**	
NWO talent days	Oct 03, 2006
Afstudeervak organiseren en begeleiden	Nov 13-14, 2006
PhD Competence assessment	Mar 16, 2005
PhD Presentation Skills	Apr 24-May 15, 2006
PhD The art of writing	Jun 07-28, 2005
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
PhD Council membership	2006 - 2007
Subtotal Personal Development	*5.0 credits**

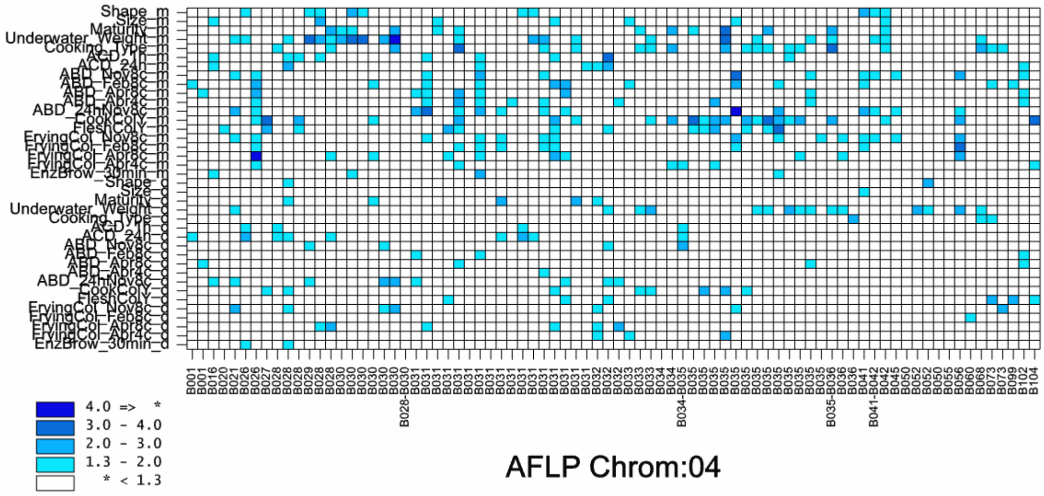
TOTAL NUMBER OF CREDIT POINTS*	**43.9**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 credits

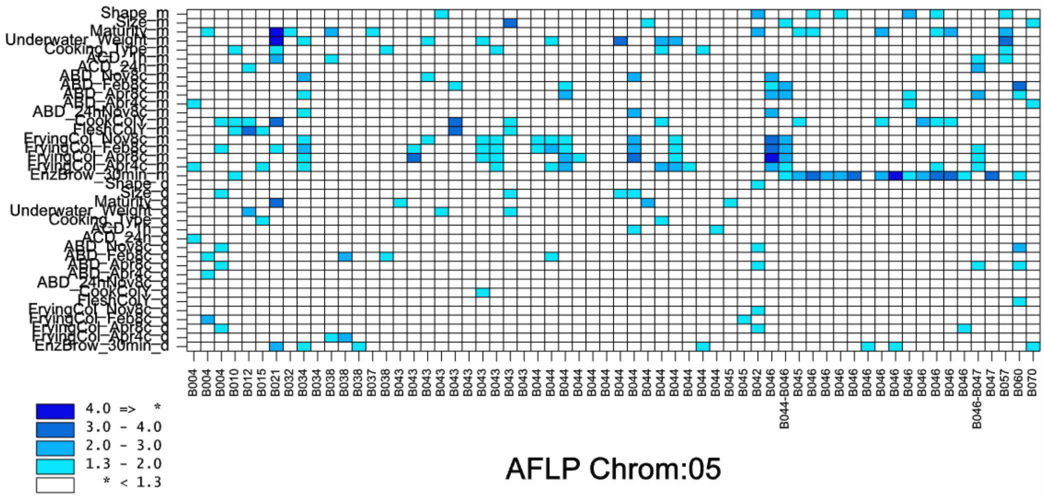
** A credit represents a normative study load of 28 hours of study*

Supplementary Figures and Tables Chapter 5

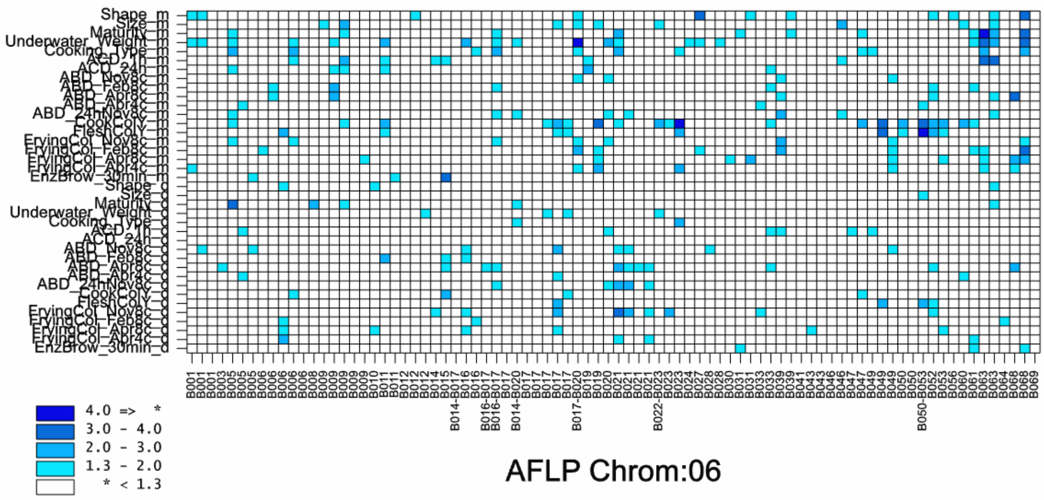




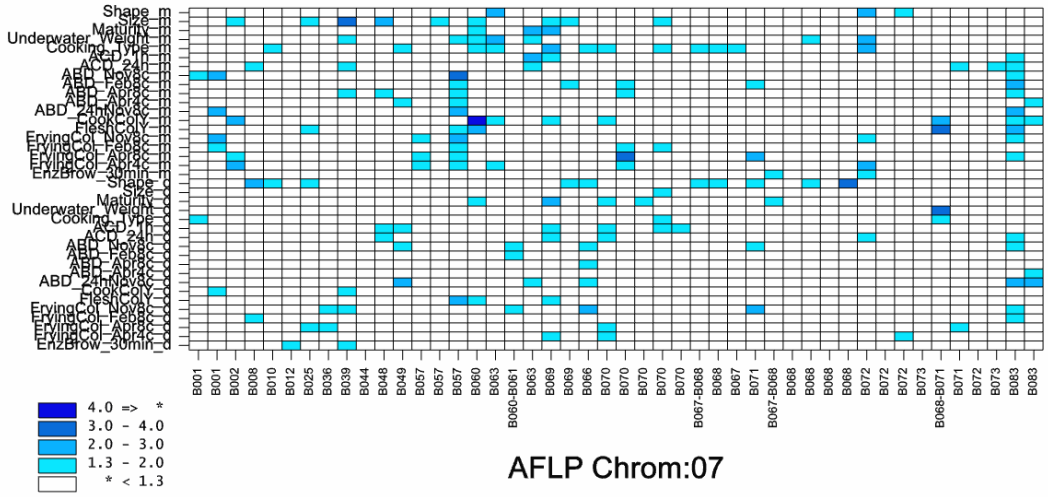
AFLP Chrom:04

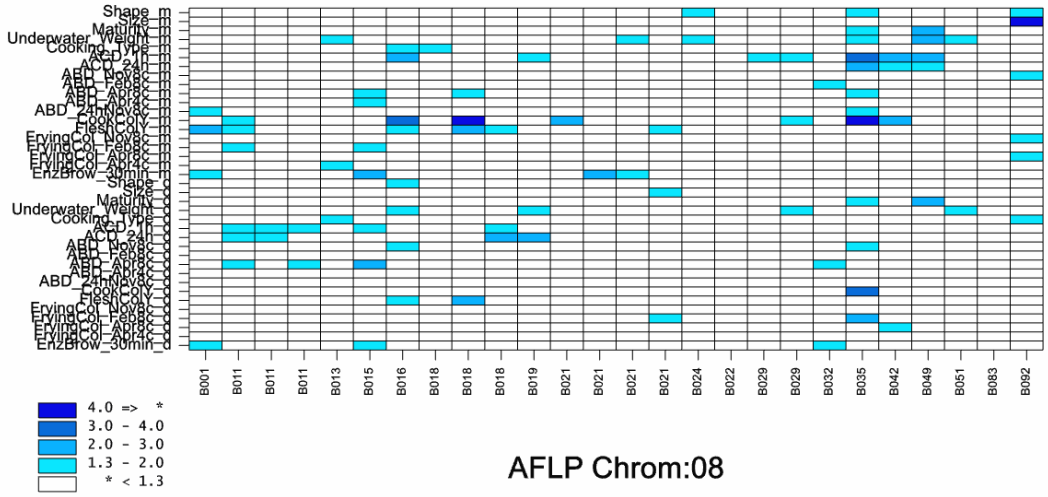


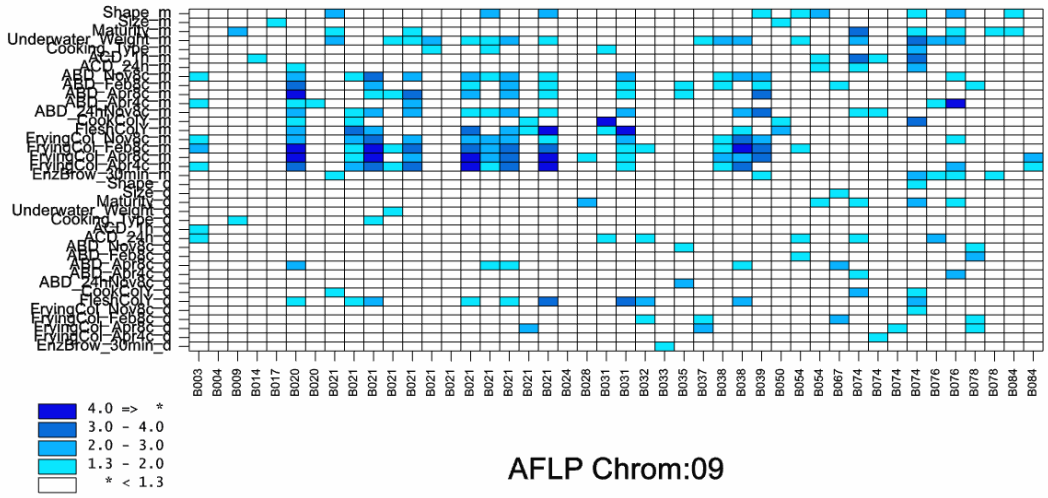
AFLP Chrom:05

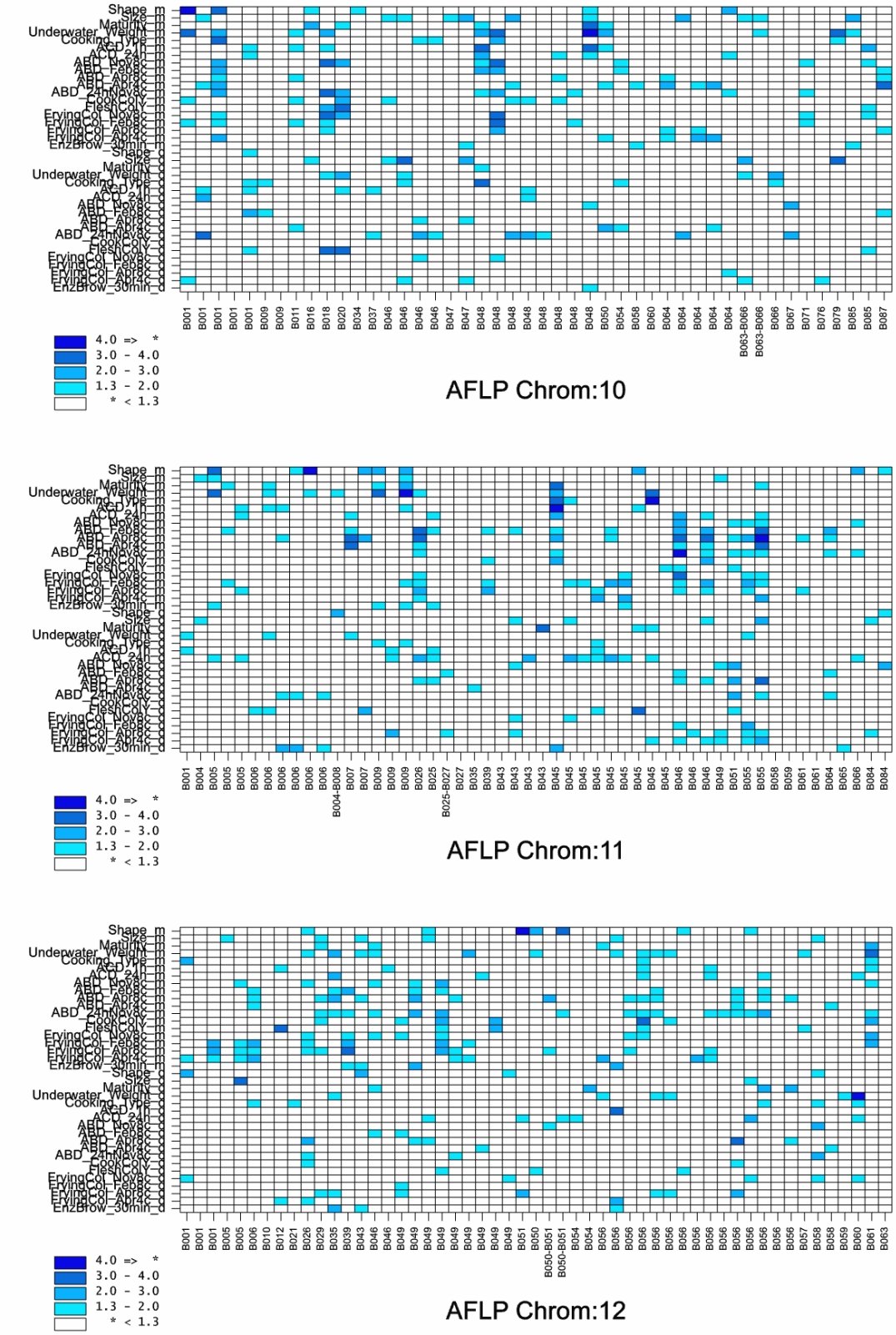


AFLP Chrom:06

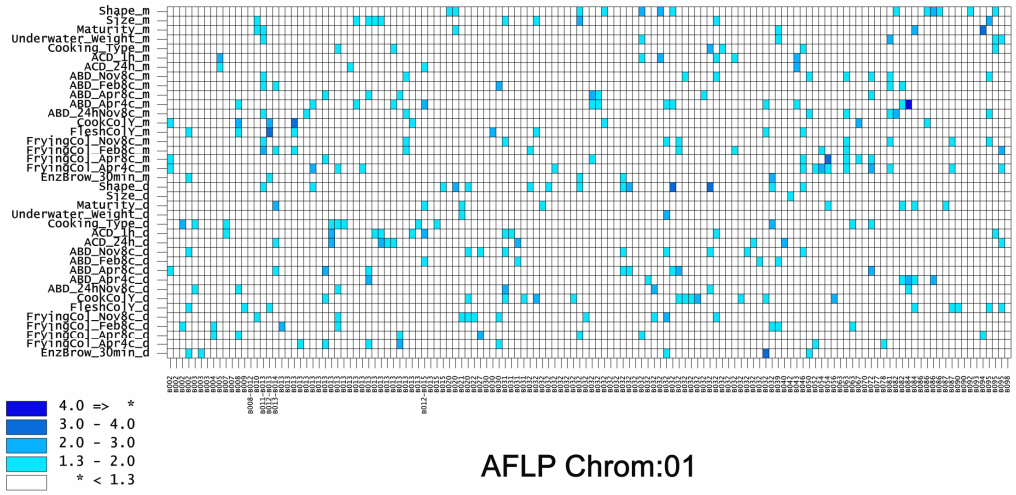

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
AcD_7h_m
ACD_24h_m
ABD_Novac_m
ABD_Feb4c_m
ABD_Apr4c_m
ABD_Apr4x_m
ABD_24hNovac_m
CookColy_m
FleshColy_m
ErvingCol_Novac_m
ErvingCol_Feb4c_m
ErvingCol_Apr4c_m
ErvingCol_Apr4x_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
AcD_7h_d
ACD_24h_d
ABD_Novac_d
ABD_Feb4c_d
ABD_Apr4c_d
ABD_Apr4x_d
ABD_24hNovac_d
CookColy_d
FleshColy_d
ErvingCol_Novac_d
ErvingCol_Feb4c_d
ErvingCol_Apr4c_d
ErvingCol_Apr4x_d
EnzBrow_30min_d
B001 B001 B002 B008 B010 B012 B025 B036 B039 B044 B048 B049 B057 B057 B057 B060 B063 B060-B061 B063 B069 B069 B066 B070 B070 B070 B070 B070 B067-B068 B068 B067 B071 B067-B068 B068 B068 B068 B068 B072 B072 B072 B073 B068-B071 B071 B072 B073 B083 B083
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:07

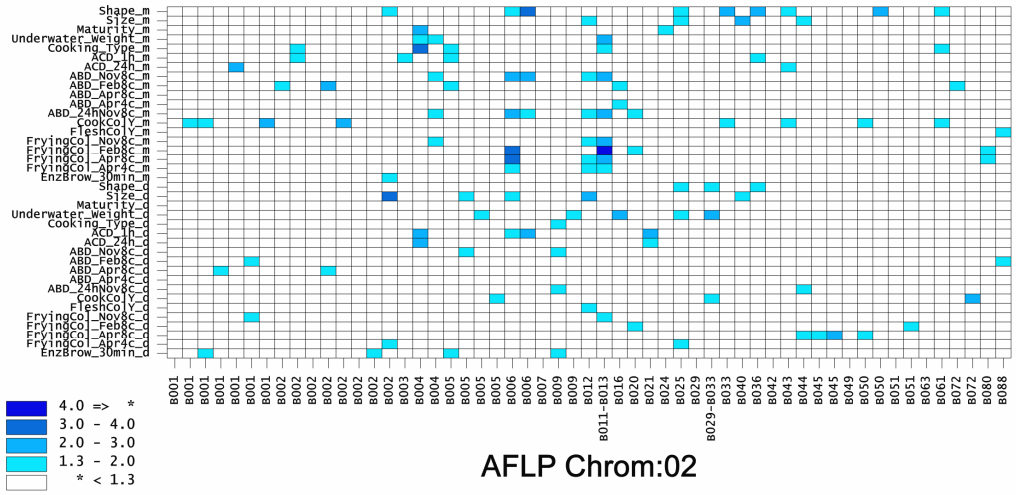

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
AcD_7h_m
ACD_24h_m
ABD_Novac_m
ABD_Feb4c_m
ABD_Apr4c_m
ABD_Apr4x_m
ABD_24hNovac_m
CookColy_m
FleshColy_m
ErvingCol_Novac_m
ErvingCol_Feb4c_m
ErvingCol_Apr4c_m
ErvingCol_Apr4x_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
AcD_7h_d
ACD_24h_d
ABD_Novac_d
ABD_Feb4c_d
ABD_Apr4c_d
ABD_Apr4x_d
ABD_24hNovac_d
CookColy_d
FleshColy_d
ErvingCol_Novac_d
ErvingCol_Feb4c_d
ErvingCol_Apr4c_d
ErvingCol_Apr4x_d
EnzBrow_30min_d
B001 B011 B011 B011 B013 B015 B016 B018 B018 B018 B019 B021 B021 B021 B021 B024 B022 B029 B029 B032 B035 B042 B049 B051 B083 B092
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:08

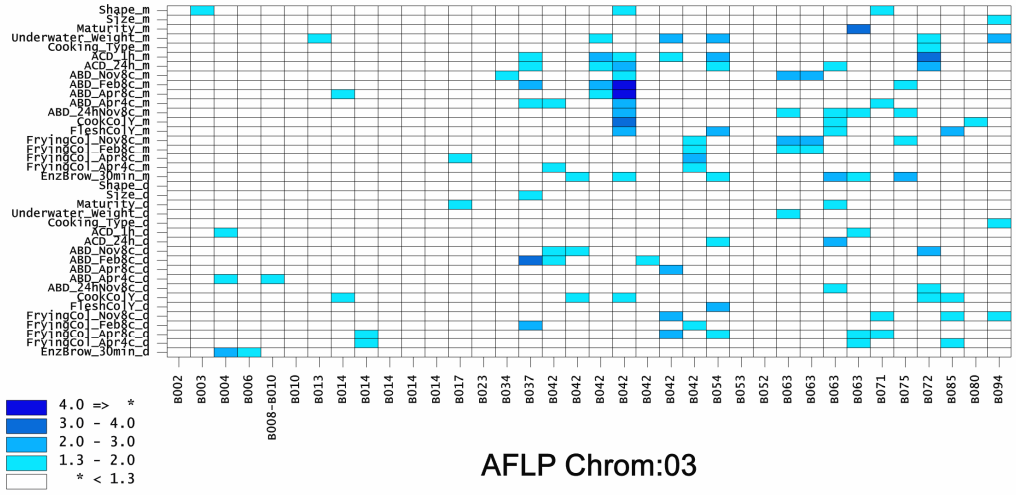

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
AcD_7h_m
ACD_24h_m
ABD_Novac_m
ABD_Feb4c_m
ABD_Apr4c_m
ABD_Apr4x_m
ABD_24hNovac_m
CookColy_m
FleshColy_m
ErvingCol_Novac_m
ErvingCol_Feb4c_m
ErvingCol_Apr4c_m
ErvingCol_Apr4x_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
AcD_7h_d
ACD_24h_d
ABD_Novac_d
ABD_Feb4c_d
ABD_Apr4c_d
ABD_Apr4x_d
ABD_24hNovac_d
CookColy_d
FleshColy_d
ErvingCol_Novac_d
ErvingCol_Feb4c_d
ErvingCol_Apr4c_d
ErvingCol_Apr4x_d
EnzBrow_30min_d
B003 B004 B009 B014 B017 B020 B020 B021 B021 B021 B021 B021 B021 B021 B021 B021 B021 B021 B021 B024 B028 B031 B031 B032 B033 B035 B037 B038 B038 B039 B050 B054 B054 B067 B074 B074 B074 B074 B076 B076 B078 B078 B084 B084
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:09

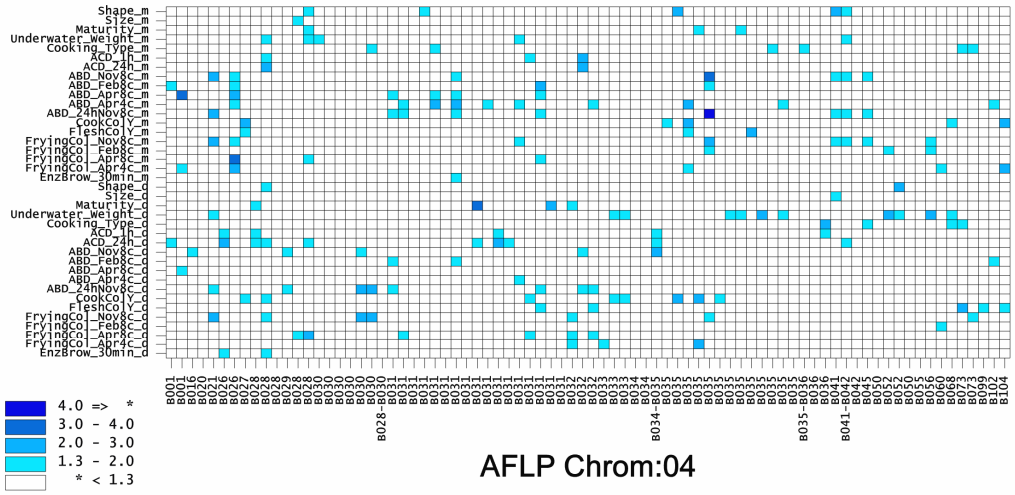


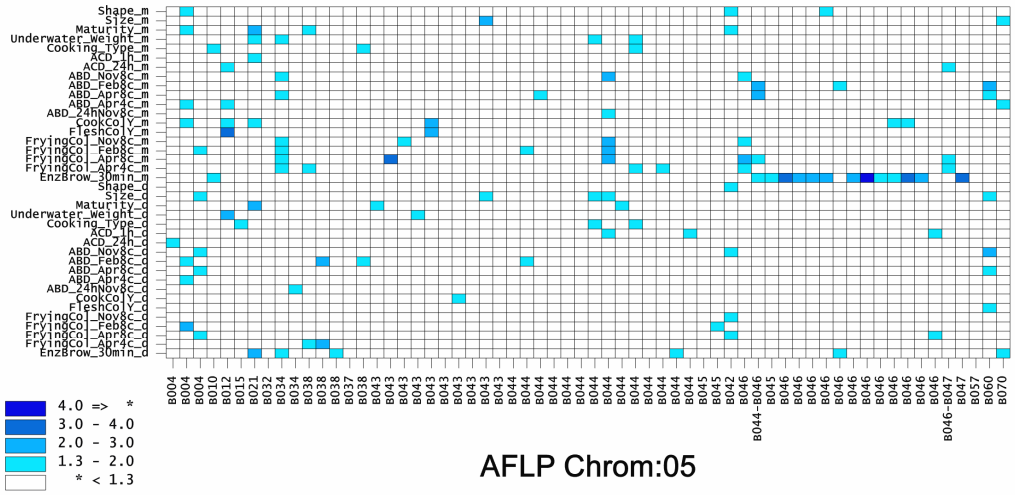
Supplementary Figure 1: A per chromosome overview of obtained associations with the basic association mapping model applied to the set of 720 AFLPs with known map location and the 2006 field trial. On the left side of each plot the trait names are indicated followed by a "m" for mean effect and a "d" for genotype-by-location interaction effect. Beneath each plot, the BIN positions are denoted, ordered from left to right in an ascending fashion. Identified associations are coloured according to increasing "-10logP"-value.

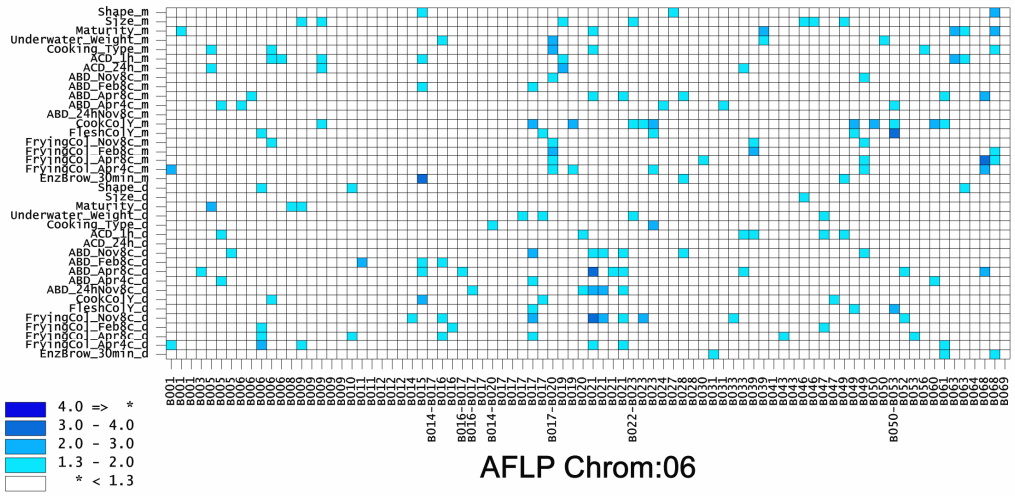

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8C_m
ABD_Feb8C_m
ABD_Apr8C_m
ABD_Apr4C_m
ABD_24hNov8C_m
CookColY_m
FleshColY_m
FryingCol_Nov8C_m
FryingCol_Feb8C_m
FryingCol_Apr8C_m
FryingCol_Apr4C_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8C_d
ABD_Feb8C_d
ABD_Apr8C_d
ABD_Apr4C_d
ABD_24hNov8C_d
CookColY_d
FleshColY_d
FryingCol_Nov8C_d
FryingCol_Feb8C_d
FryingCol_Apr8C_d
FryingCol_Apr4C_d
EnzBrow_30min_d
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:01

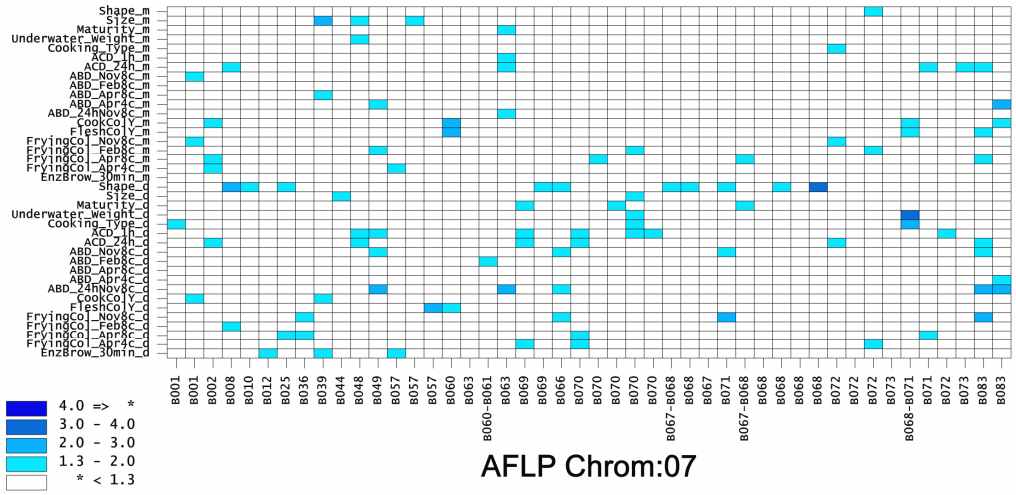

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8C_m
ABD_Feb8C_m
ABD_Apr8C_m
ABD_Apr4C_m
ABD_24hNov8C_m
CookColY_m
FleshColY_m
FryingCol_Nov8C_m
FryingCol_Feb8C_m
FryingCol_Apr8C_m
FryingCol_Apr4C_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8C_d
ABD_Feb8C_d
ABD_Apr8C_d
ABD_Apr4C_d
ABD_24hNov8C_d
CookColY_d
FleshColY_d
FryingCol_Nov8C_d
FryingCol_Feb8C_d
FryingCol_Apr8C_d
FryingCol_Apr4C_d
EnzBrow_30min_d
B001 B001 B001 B001 B001 B001 B002 B002 B002 B002 B002 B002 B002 B002 B003 B004 B004 B005 B005 B005 B006 B006 B007 B009 B009 B012 B011-B013 B016 B020 B021 B024 B025 B029 B029-B033 B033 B040 B036 B042 B043 B044 B045 B045 B049 B050 B050 B051 B051 B063 B061 B072 B072 B080 B088
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:02

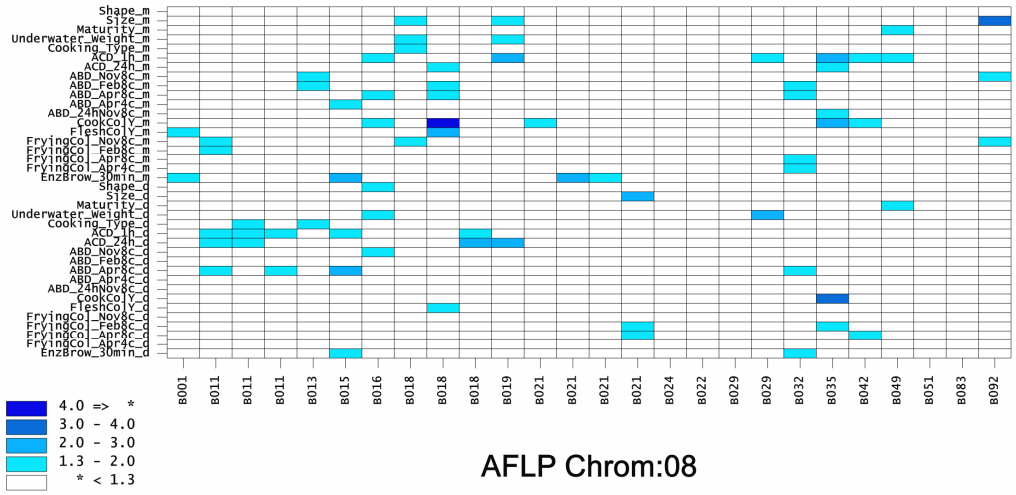

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8C_m
ABD_Feb8C_m
ABD_Apr8C_m
ABD_Apr4C_m
ABD_24hNov8C_m
CookColY_m
FleshColY_m
FryingCol_Nov8C_m
FryingCol_Feb8C_m
FryingCol_Apr8C_m
FryingCol_Apr4C_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8C_d
ABD_Feb8C_d
ABD_Apr8C_d
ABD_Apr4C_d
ABD_24hNov8C_d
CookColY_d
FleshColY_d
FryingCol_Nov8C_d
FryingCol_Feb8C_d
FryingCol_Apr8C_d
FryingCol_Apr4C_d
EnzBrow_30min_d
B002 B003 B004 B006 B008-B010 B010 B013 B014 B014 B014 B014 B014 B017 B023 B034 B037 B042 B042 B042 B042 B042 B042 B042 B054 B053 B052 B063 B063 B063 B063 B071 B075 B072 B085 B080 B094
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:03

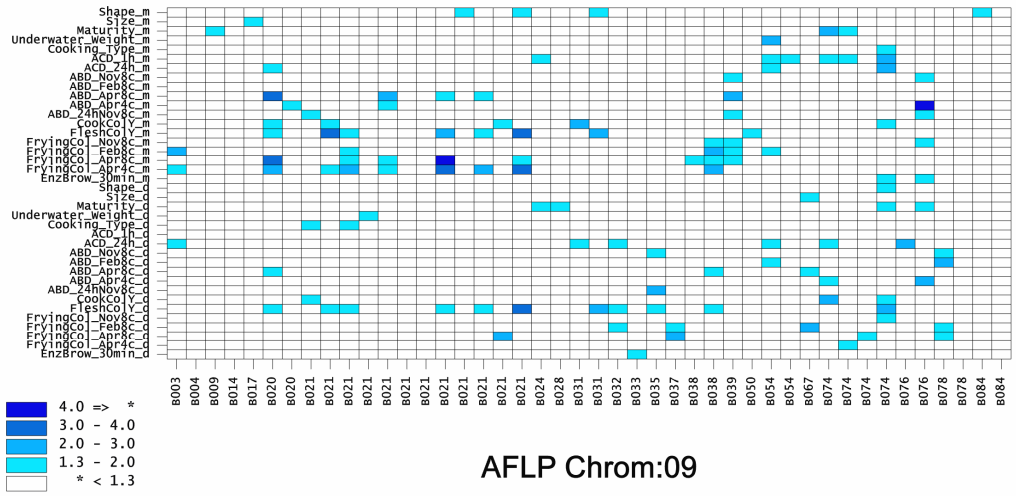

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8c_m
ABD_Feb8c_m
ABD_Apr8c_m
ABD_Apr4c_m
ABD_24hNov8c_m
CookColY_m
FleshColY_m
FryingCol_Nov8c_m
FryingCol_Feb8c_m
FryingCol_Apr8c_m
FryingCol_Apr4c_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8c_d
ABD_Feb8c_d
ABD_Apr8c_d
ABD_Apr4c_d
ABD_24hNov8c_d
CookColY_d
FleshColY_d
FryingCol_Nov8c_d
FryingCol_Feb8c_d
FryingCol_Apr8c_d
FryingCol_Apr4c_d
EnzBrow_30min_d
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:04

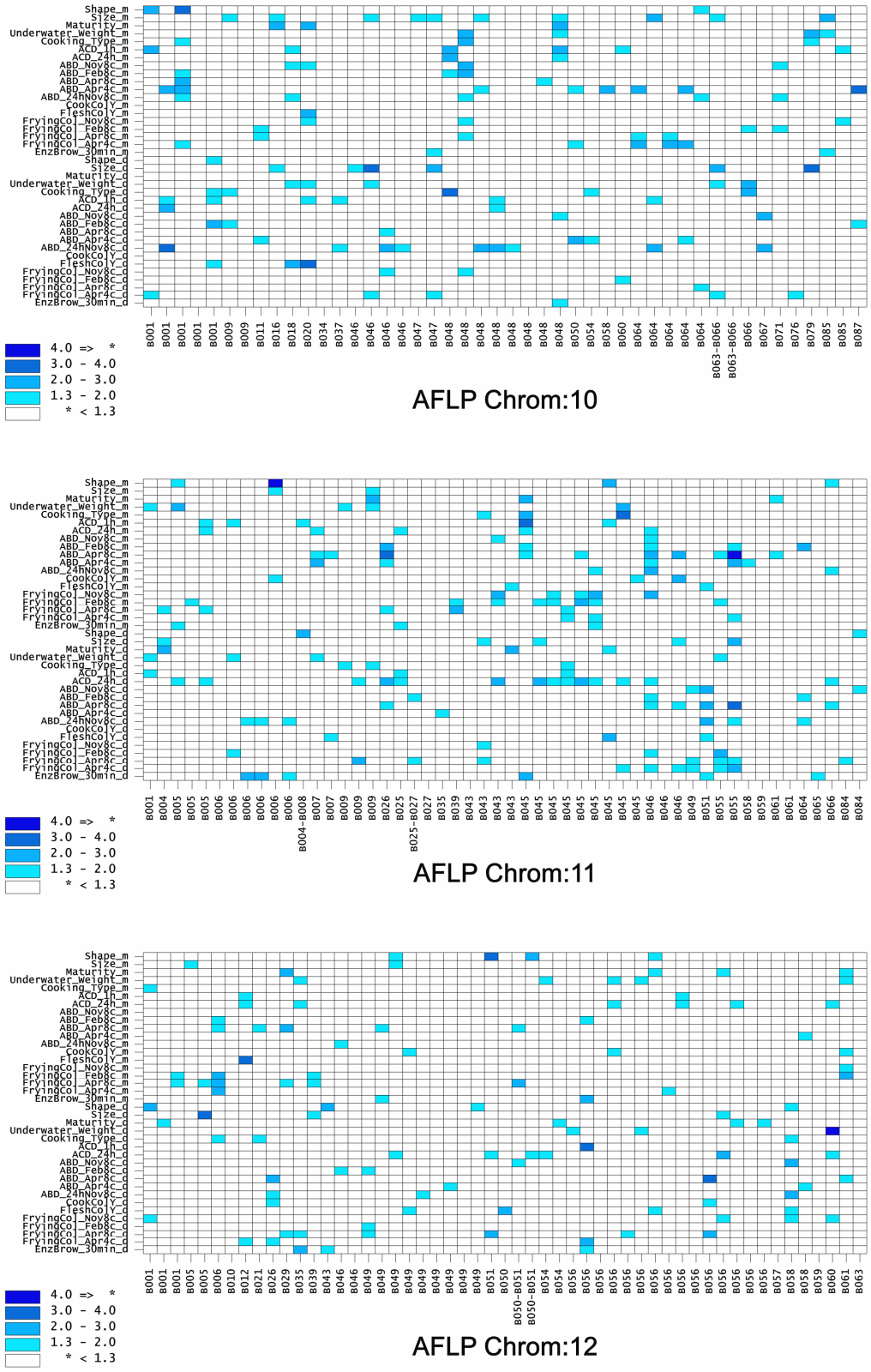

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8c_m
ABD_Feb8c_m
ABD_Apr8c_m
ABD_Apr4c_m
ABD_24hNov8c_m
CookColY_m
FleshColY_m
FryingCol_Nov8c_m
FryingCol_Feb8c_m
FryingCol_Apr8c_m
FryingCol_Apr4c_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8c_d
ABD_Feb8c_d
ABD_Apr8c_d
ABD_Apr4c_d
ABD_24hNov8c_d
CookColY_d
FleshColY_d
FryingCol_Nov8c_d
FryingCol_Feb8c_d
FryingCol_Apr8c_d
FryingCol_Apr4c_d
EnzBrow_30min_d
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:05


Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8c_m
ABD_Feb8c_m
ABD_Apr8c_m
ABD_Apr4c_m
ABD_24hNov8c_m
CookColY_m
FleshColY_m
FryingCol_Nov8c_m
FryingCol_Feb8c_m
FryingCol_Apr8c_m
FryingCol_Apr4c_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8c_d
ABD_Feb8c_d
ABD_Apr8c_d
ABD_Apr4c_d
ABD_24hNov8c_d
CookColY_d
FleshColY_d
FryingCol_Nov8c_d
FryingCol_Feb8c_d
FryingCol_Apr8c_d
FryingCol_Apr4c_d
EnzBrow_30min_d
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:06

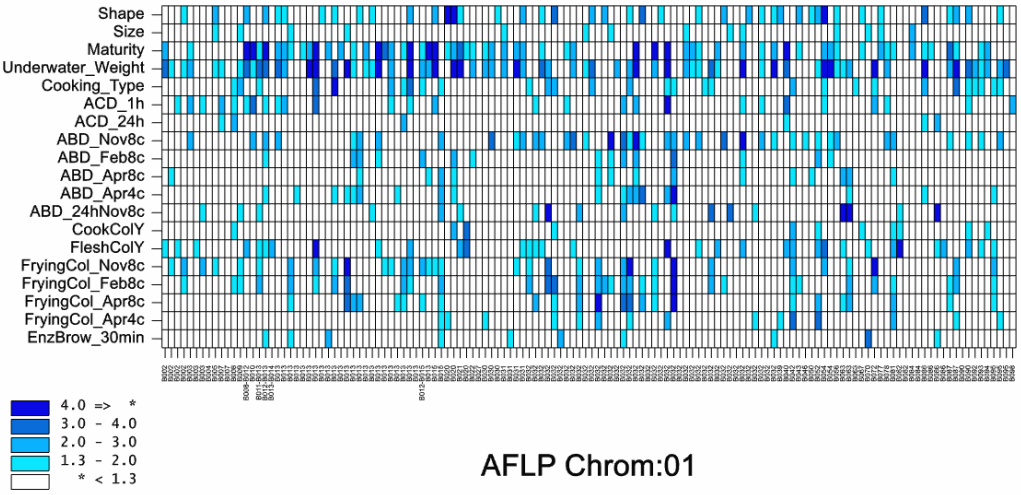

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8C_m
ABD_Feb8C_m
ABD_Apr8C_m
ABD_Apr4C_m
ABD_24hNov8C_m
CookColY_m
FleshColY_m
FryingCol_Nov8C_m
FryingCol_Feb8C_m
FryingCol_Apr8C_m
FryingCol_Apr4C_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8C_d
ABD_Feb8C_d
ABD_Apr8C_d
ABD_Apr4C_d
ABD_24hNov8C_d
CookColY_d
FleshColY_d
FryingCol_Nov8C_d
FryingCol_Feb8C_d
FryingCol_Apr8C_d
FryingCol_Apr4C_d
EnzBrow_30min_d
B001 B001 B002 B008 B010 B012 B025 B036 B039 B044 B048 B049 B057 B057 B057 B060 B063 B060-B061 B063 B069 B069 B066 B070 B070 B070 B070 B067-B068 B068 B067 B071 B067-B068 B068 B068 B068 B072 B072 B073 B068-B071 B071 B072 B073 B083 B083
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:07


Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8C_m
ABD_Feb8C_m
ABD_Apr8C_m
ABD_Apr4C_m
ABD_24hNov8C_m
CookColY_m
FleshColY_m
FryingCol_Nov8C_m
FryingCol_Feb8C_m
FryingCol_Apr8C_m
FryingCol_Apr4C_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8C_d
ABD_Feb8C_d
ABD_Apr8C_d
ABD_Apr4C_d
ABD_24hNov8C_d
CookColY_d
FleshColY_d
FryingCol_Nov8C_d
FryingCol_Feb8C_d
FryingCol_Apr8C_d
FryingCol_Apr4C_d
EnzBrow_30min_d
B001 B011 B011 B011 B013 B015 B016 B018 B018 B018 B019 B021 B021 B021 B021 B024 B022 B029 B029 B032 B035 B042 B049 B051 B083 B092
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:08

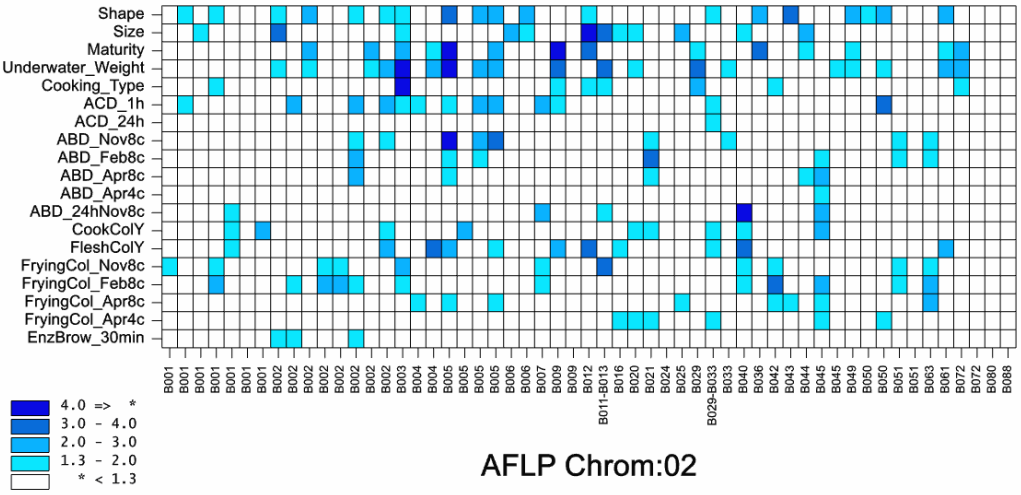

Shape_m
Size_m
Maturity_m
Underwater_Weight_m
Cooking_Type_m
ACD_1h_m
ACD_24h_m
ABD_Nov8C_m
ABD_Feb8C_m
ABD_Apr8C_m
ABD_Apr4C_m
ABD_24hNov8C_m
CookColY_m
FleshColY_m
FryingCol_Nov8C_m
FryingCol_Feb8C_m
FryingCol_Apr8C_m
FryingCol_Apr4C_m
EnzBrow_30min_m
Shape_d
Size_d
Maturity_d
Underwater_Weight_d
Cooking_Type_d
ACD_1h_d
ACD_24h_d
ABD_Nov8C_d
ABD_Feb8C_d
ABD_Apr8C_d
ABD_Apr4C_d
ABD_24hNov8C_d
CookColY_d
FleshColY_d
FryingCol_Nov8C_d
FryingCol_Feb8C_d
FryingCol_Apr8C_d
FryingCol_Apr4C_d
EnzBrow_30min_d
B003 B004 B009 B014 B017 B020 B020 B021 B021 B021 B021 B021 B021 B021 B021 B021 B021 B021 B024 B028 B031 B031 B032 B033 B035 B037 B038 B038 B039 B050 B054 B054 B067 B074 B074 B074 B074 B076 B076 B078 B084 B084
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:09



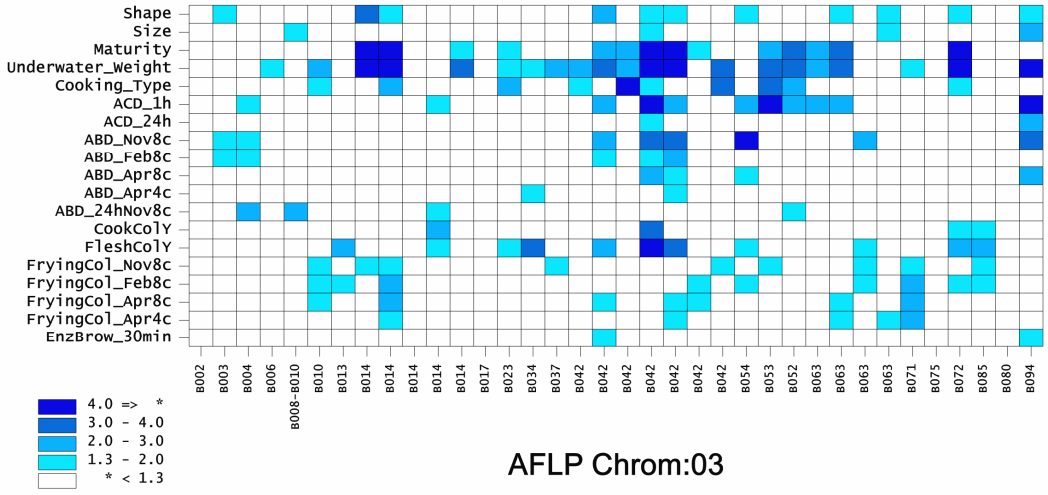
Supplementary Figure 2: A per chromosome overview of obtained associations with the genetic relatedness correcting association mapping model applied to the set of 720 AFLPs with known map location and the 2006 field trial. On the left side of each plot the trait names are indicated followed by a "m" for mean effect and a "d" for genotype-by-location interaction effect. Beneath each plot, the BIN positions are denoted, ordered from left to right in an ascending fashion. Identified associations are coloured according to increasing "-10logP"-value.



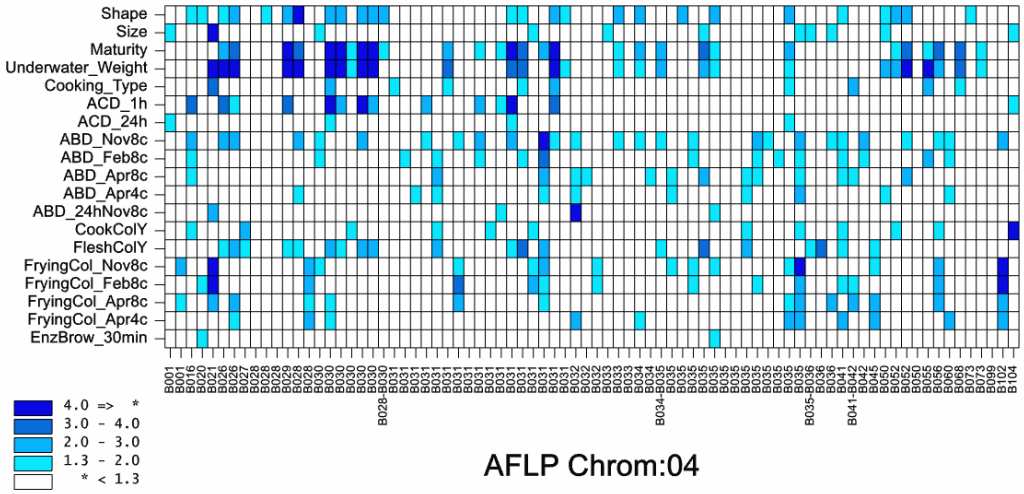
AFLP Chrom:01



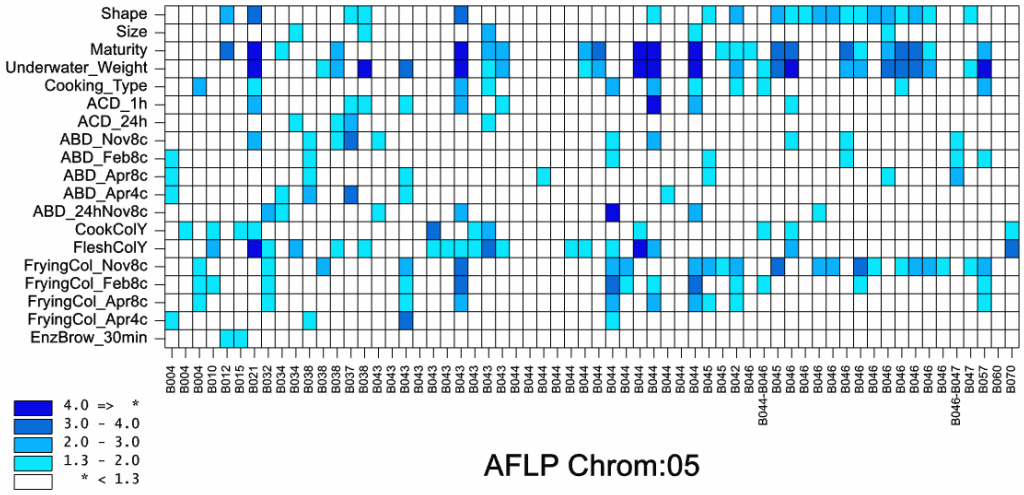
AFLP Chrom:02



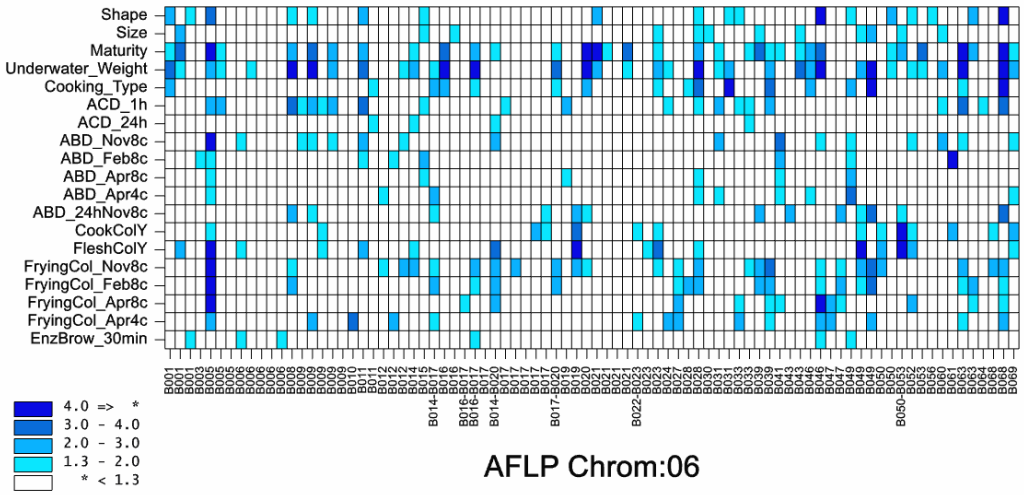
AFLP Chrom:03



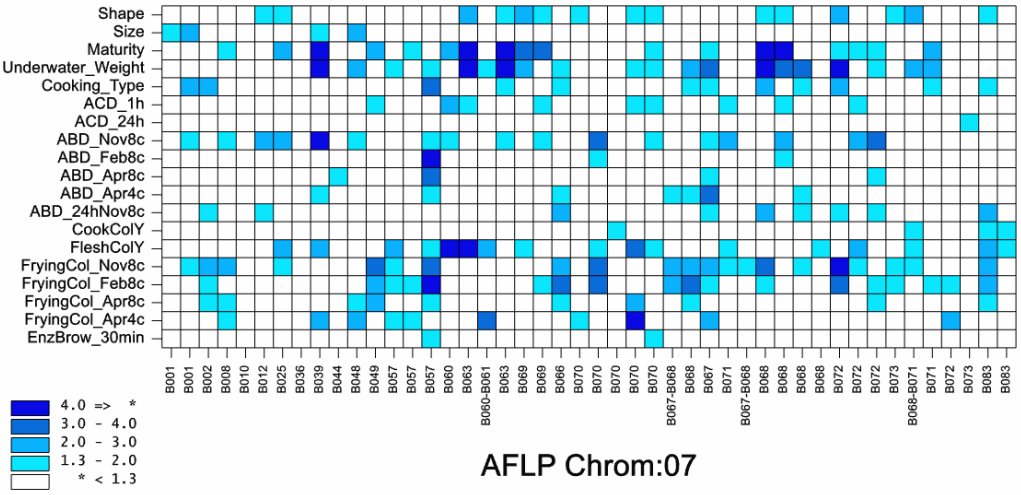
AFLP Chrom:04



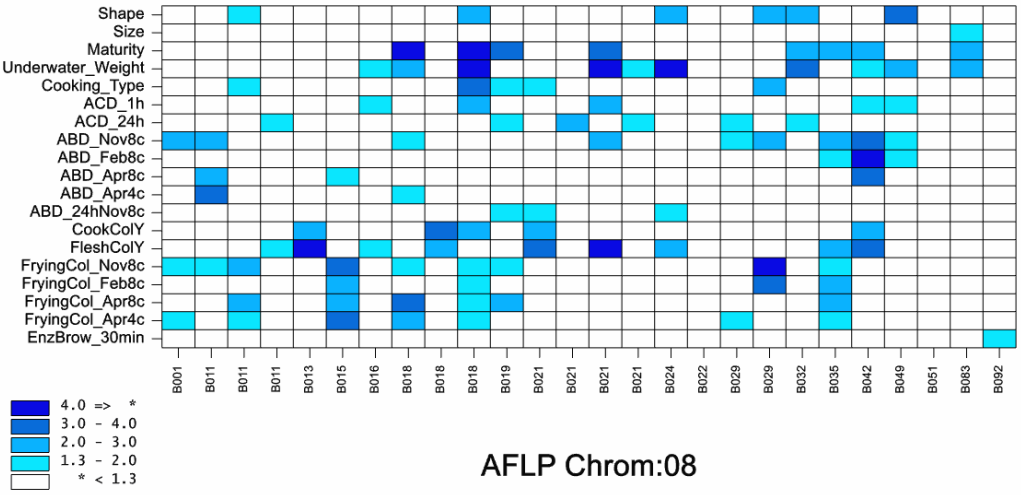
AFLP Chrom:05



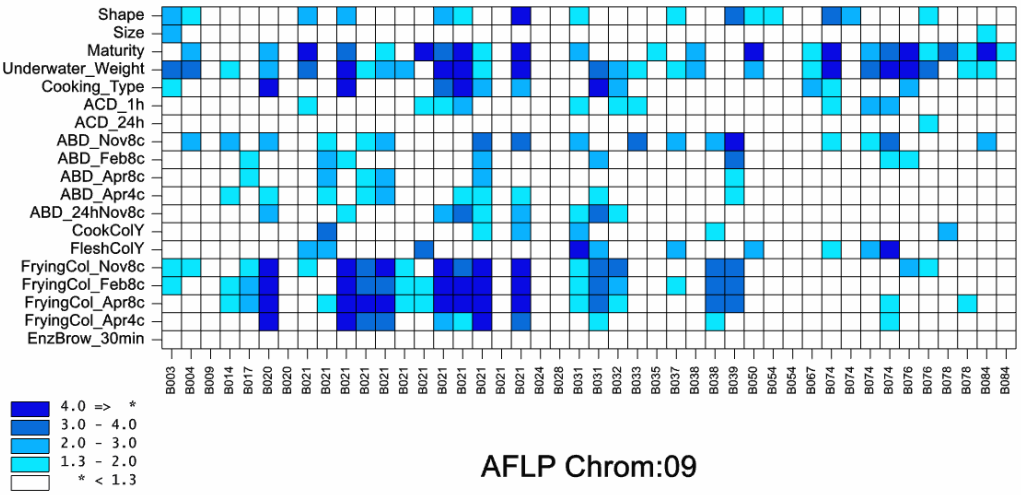
AFLP Chrom:06



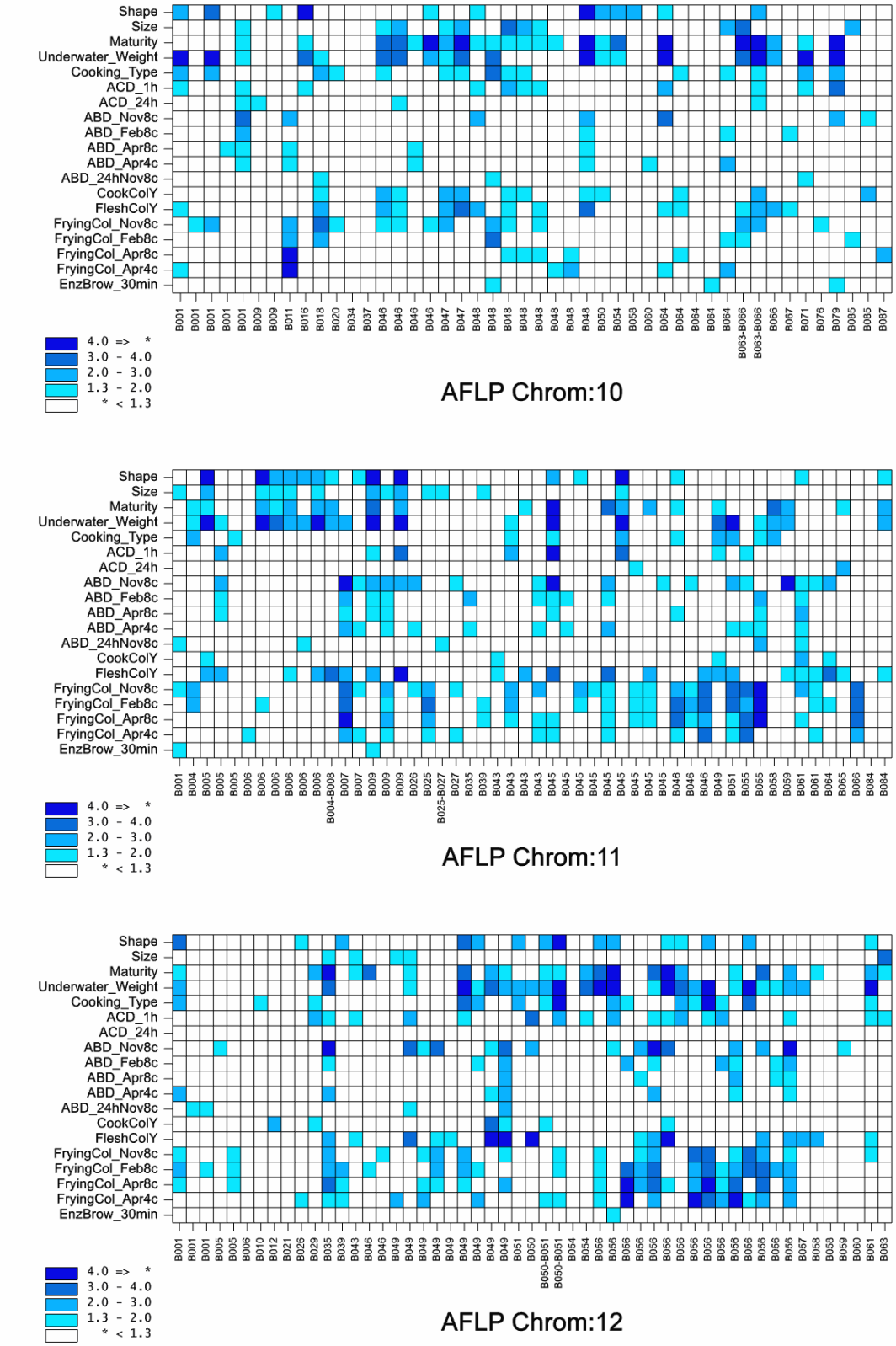
AFLP Chrom:07



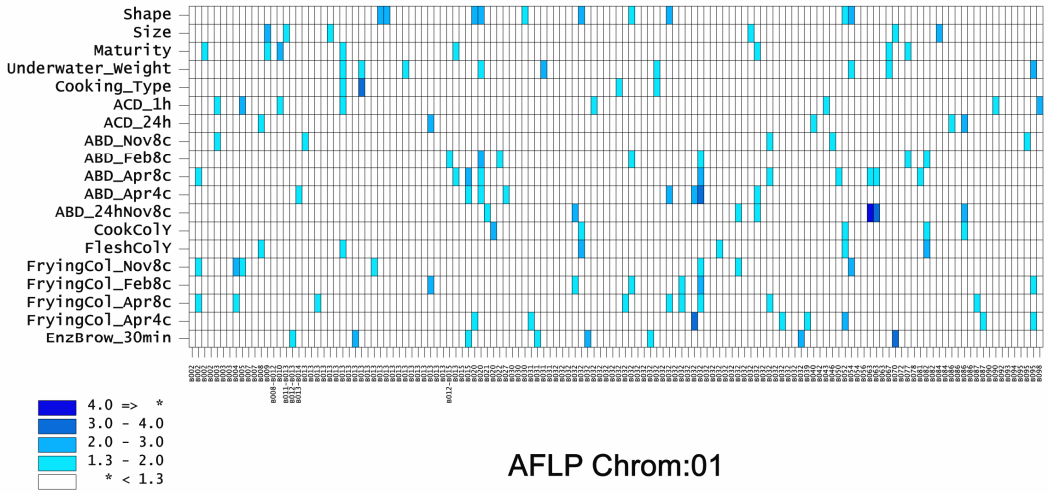
AFLP Chrom:08



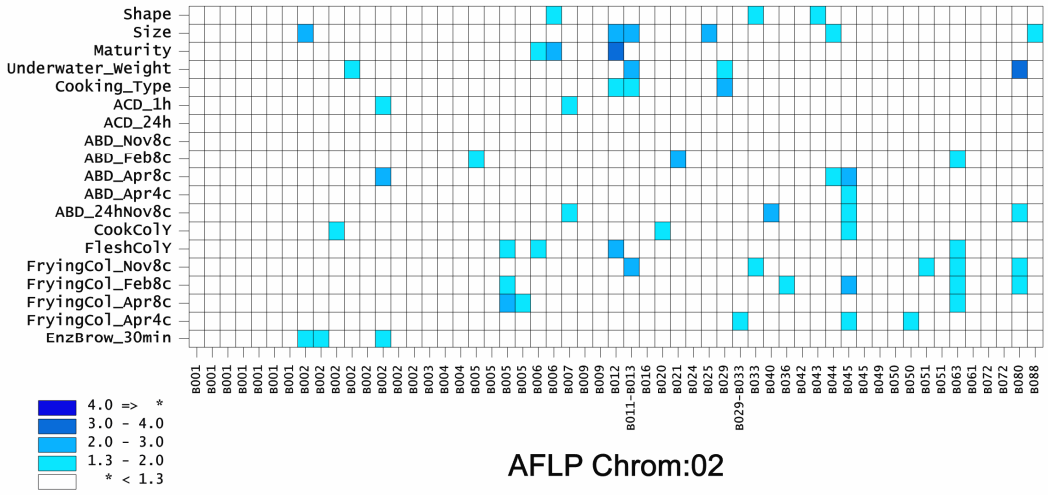
AFLP Chrom:09



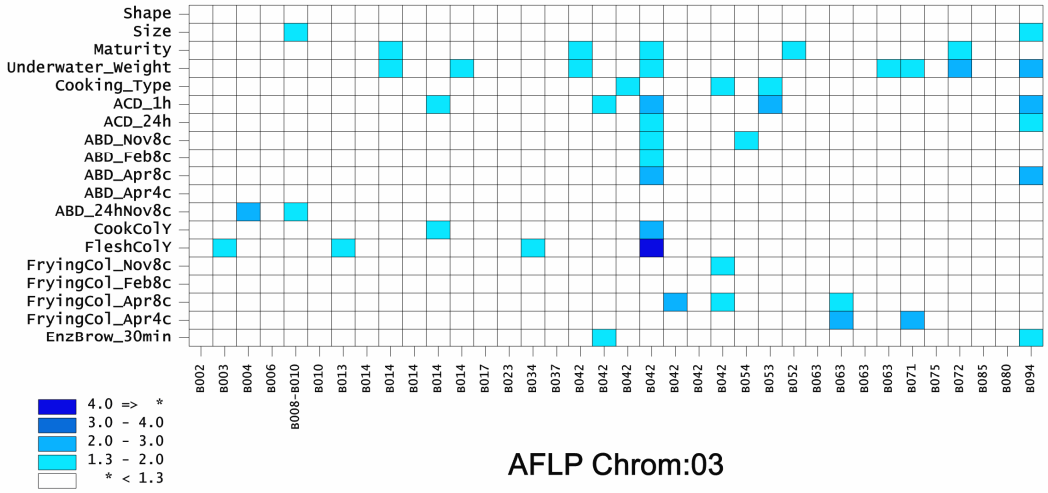
Supplementary Figure 3: A per chromosome overview of obtained associations with the basic association mapping model applied to the set of 720 AFLPs with known map location and the MYML data set. On the left side of each plot the trait names are indicated. Beneath each plot, the BIN positions are denoted, ordered from left to right in an ascending fashion. Identified associations are coloured according to increasing "-10logP"-value.



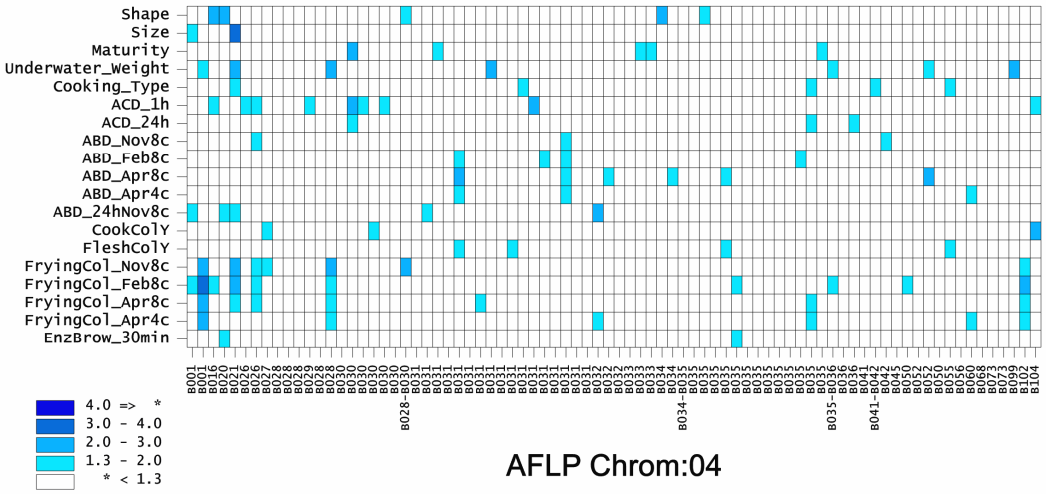
AFLP Chrom:01

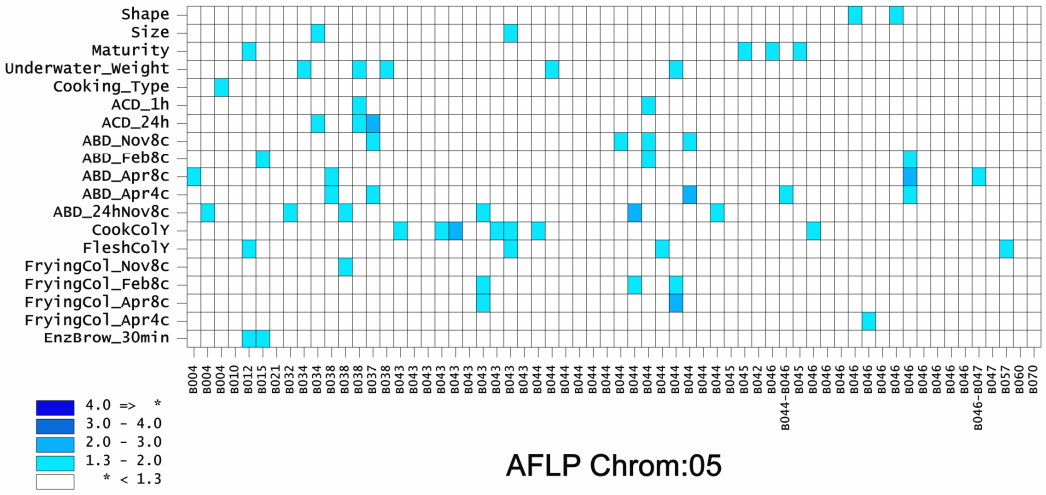


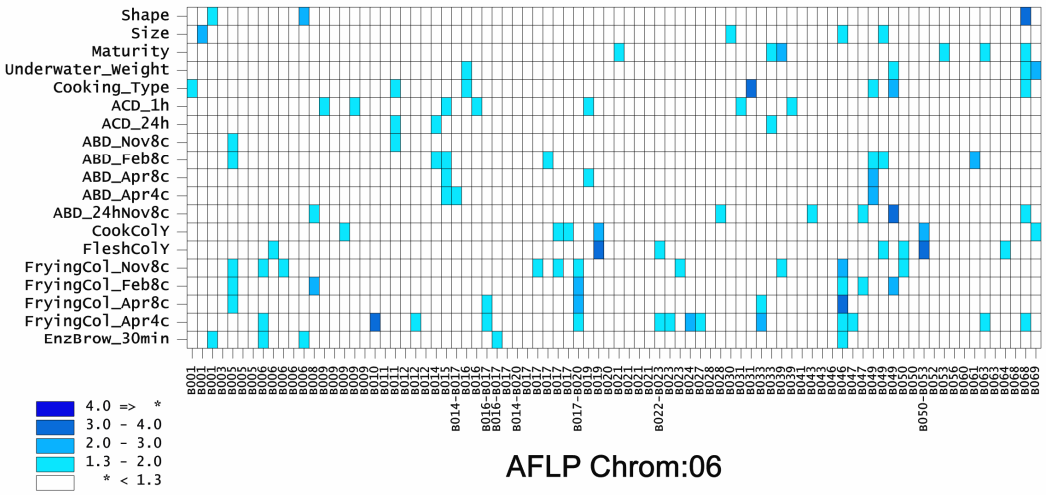
AFLP Chrom:02

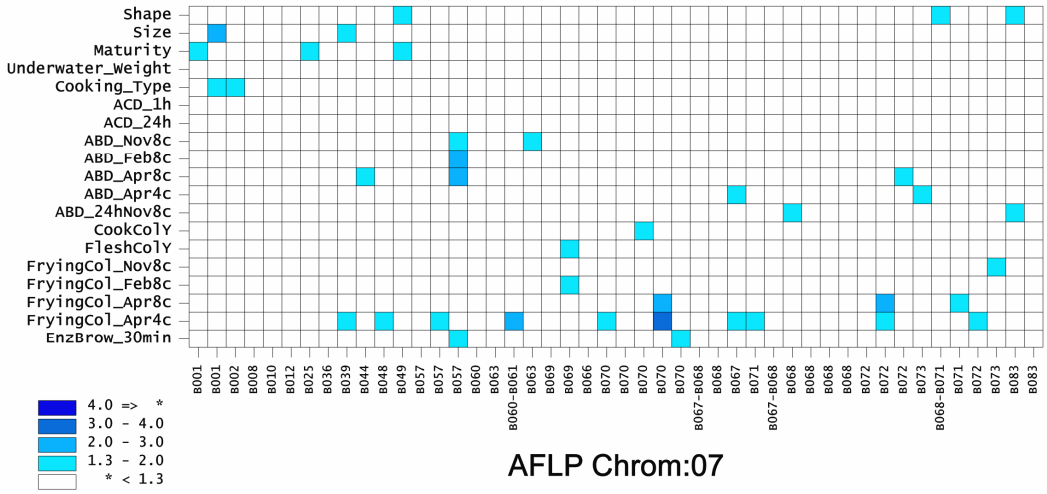


AFLP Chrom:03

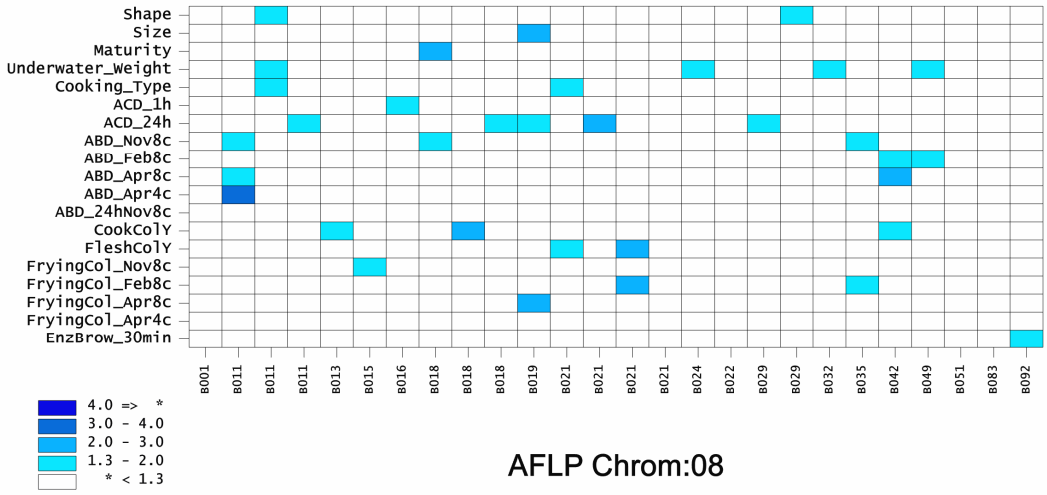

Shape
Size
Maturity
Underwater_Weight
Cooking_Type
ACD_1h
ACD_24h
ABD_Nov8c
ABD_Feb8c
ABD_Apr8c
ABD_Apr4c
ABD_24hNov8c
CookColY
FleshColY
FryingCol_Nov8c
FryingCol_Feb8c
FryingCol_Apr8c
FryingCol_Apr4c
EnzBrow_30min
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:04


Shape
Size
Maturity
Underwater_Weight
Cooking_Type
ACD_1h
ACD_24h
ABD_Nov8c
ABD_Feb8c
ABD_Apr8c
ABD_Apr4c
ABD_24hNov8c
CookColY
FleshColY
FryingCol_Nov8c
FryingCol_Feb8c
FryingCol_Apr8c
FryingCol_Apr4c
EnzBrow_30min
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:05

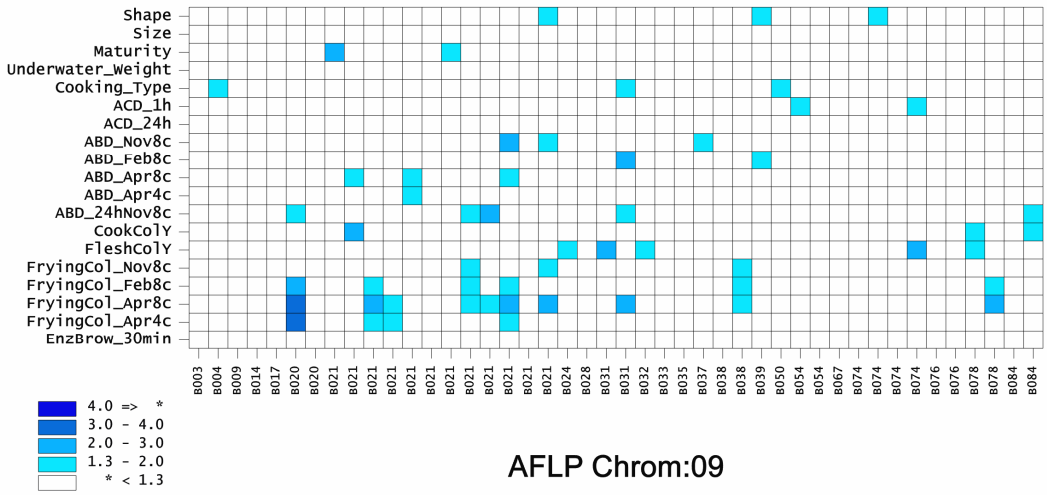

Shape
Size
Maturity
Underwater_Weight
Cooking_Type
ACD_1h
ACD_24h
ABD_Nov8c
ABD_Feb8c
ABD_Apr8c
ABD_Apr4c
ABD_24hNov8c
CookColY
FleshColY
FryingCol_Nov8c
FryingCol_Feb8c
FryingCol_Apr8c
FryingCol_Apr4c
EnzBrow_30min
4.0 => *
3.0 - 4.0
2.0 - 3.0
1.3 - 2.0
* < 1.3
AFLP Chrom:06



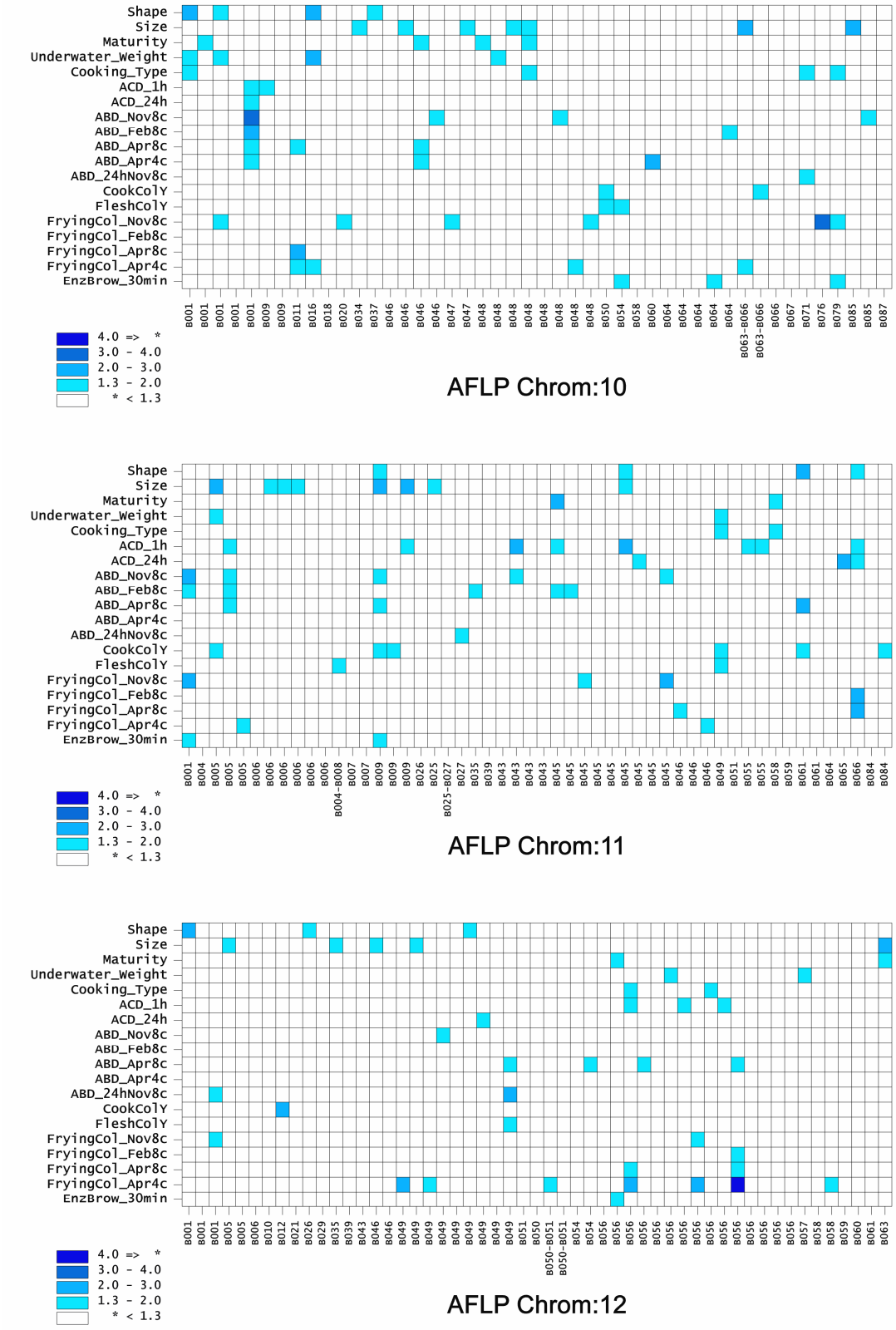
AFLP Chrom:07



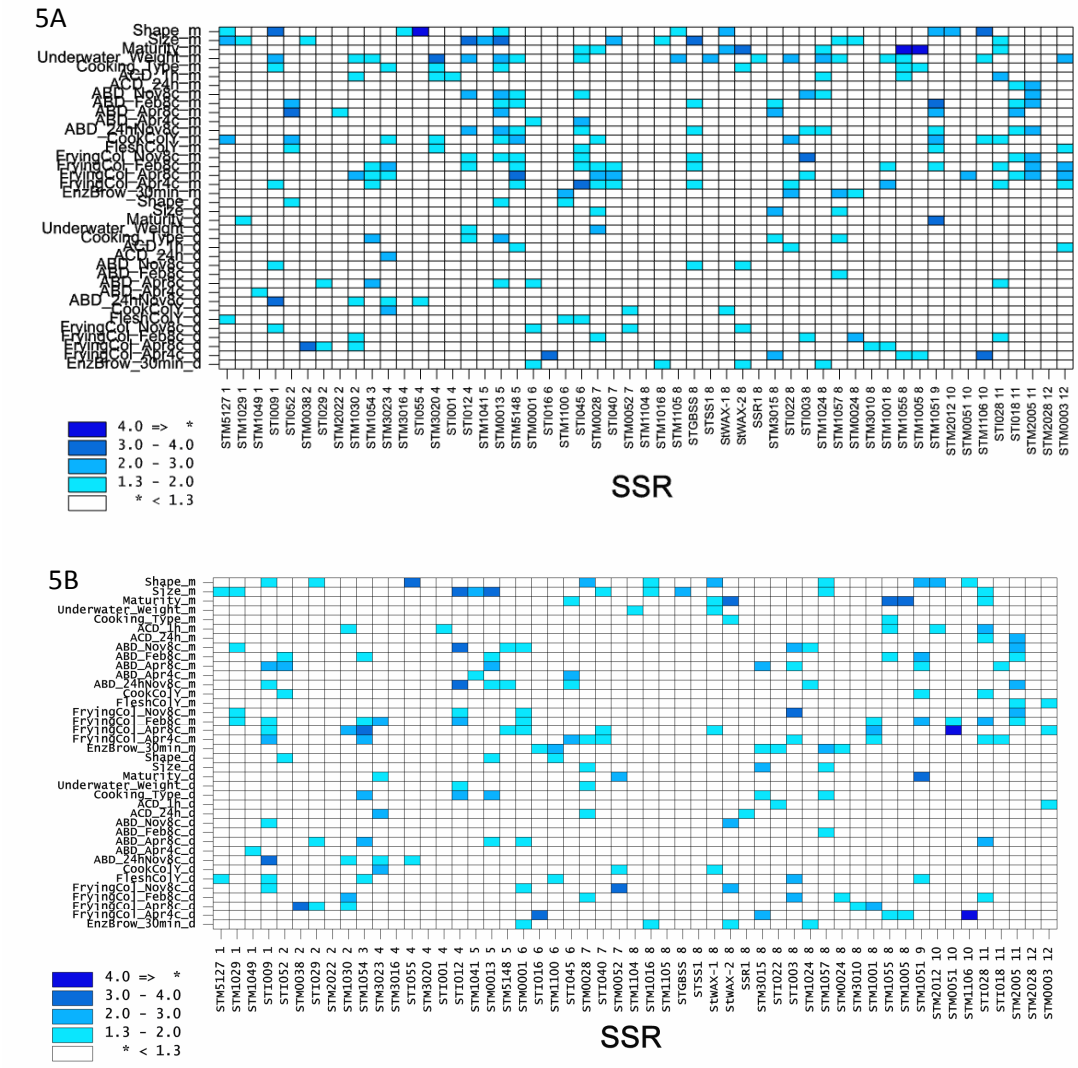
AFLP Chrom:08



AFLP Chrom:09



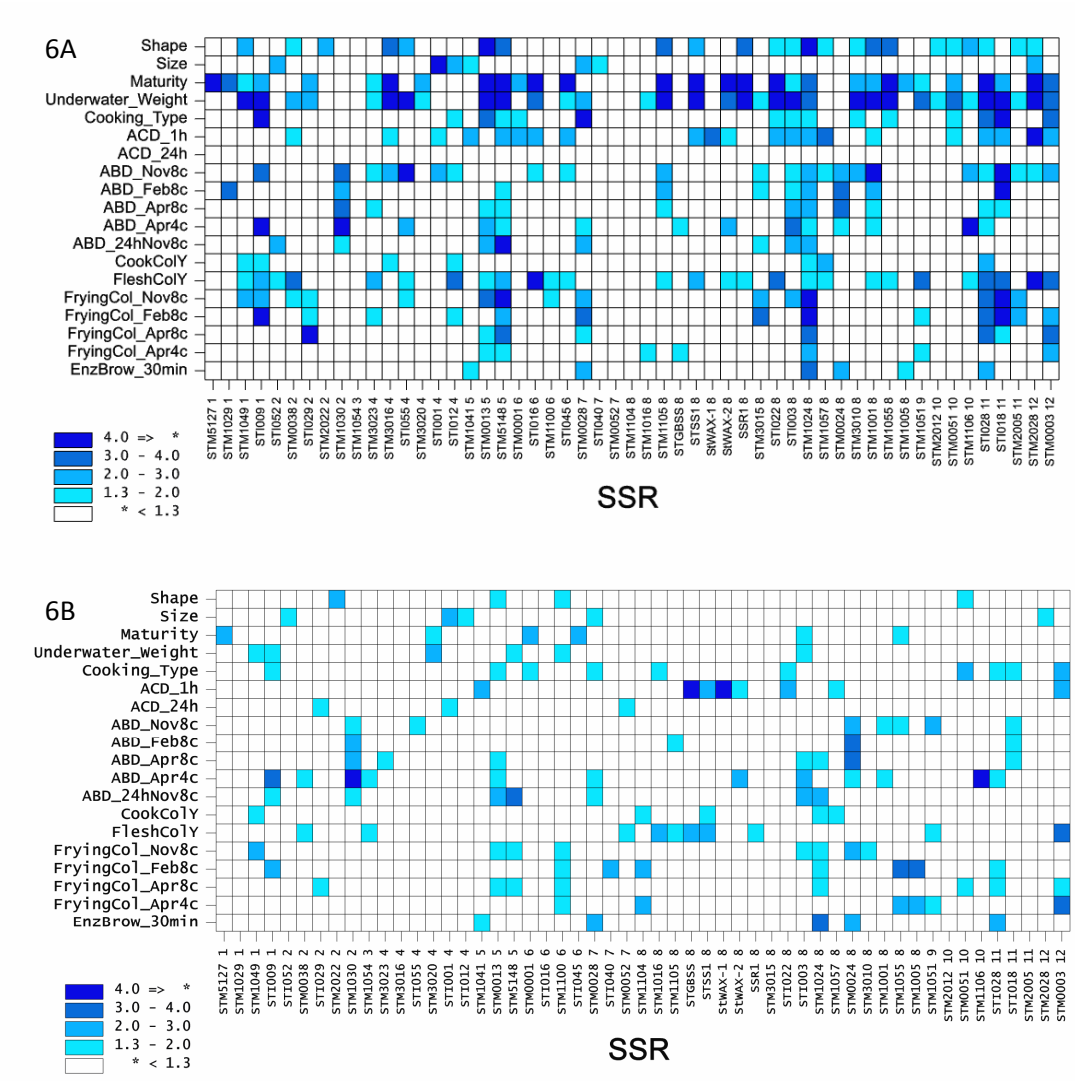
Supplementary Figure 4: A per chromosome overview of obtained associations with the genetic relatedness correcting association mapping model applied to the set of 720 AFLPs with known map location and the MYML data set. On the left side of each plot the trait names are indicated. Beneath each plot, the BIN positions are denoted, ordered from left to right in an ascending fashion. Identified associations are coloured according to increasing "-10logP"-value.



Supplementary Figure 5: A genome-wide overview of identified associations with the set of 53 microsatellite loci and the 2006 field trial. The labels represent the names of the SSRs followed by the chromosome to which they are mapped. The SSRs are ordered per chromosome from left to right.

A. Basic model

B. Genetic relatedness corrected model



Supplementary Figure 6: A genome-wide overview of identified associations with the set of 53 microsatellite loci and the MYML data set. The labels represent the names of the SSRs followed by the chromosome to which they are mapped. The SSRs are ordered per chromosome from left to right.

A. Basic model

B. Genetic relatedness corrected model

Supplementary Table 1: Detected marker-trait associations between the 2006 field trial and the set of 720 mapped AFLP markers. To the left, the associations identified with the basic model are tabulated, to the right the ones with the relatedness corrected model. Each time the observed significance "-10logP", the chromosome, the map position in cM and the marker name are listed.

	Basic model				Genetic relatedness corrected model					Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Shape	3.6	1	68.8	33137	3.1	2	3.7	17392	ACD_1h	4.1	3	56.5	34266	3.3	3	56.5	34266
	3.9	2	3.7	17392	3.1	10	0	762		3.4	4	23.4	34289	3.6	11	35.5	1122
	3.6	2	35.6	31277	4.2	11	3.8	38145		3.2	6	48	22101				
	3.3	6	19.6	14558	3	12	39.8	28308		3.1	6	48	22216				
	3.4	6	51.8	41362						3.3	8	25	22240				
	4.4	10	0	2139						3.2	9	58.1	22293				
	3.8	10	0	762						3.4	9	58.1	37131				
	3.2	11	3	1233						3.9	10	36.4	6140				
	5.4	11	3.8	38145						3.4	10	36.6	28221				
	4.3	12	39.8	28308						4.6	11	35.5	1122				
	3.7	12	40.5	23181					ABD_Nov8c	3.4	2	8.4	8142	3.5	4	27.5	1068
Size	3.1	1	24.9	4102	3	8	75.5	20282		3.2	4	27.5	1068				
	3.5	5	31.3	30298						3.1	7	44.3	41263				
	3.2	7	30.4	2109						3.3	9	16.6	8230				
	4	8	75.5	20282						4	10	13.1	8160				
Maturity	3	1	8.4	16190	3.1	1	74.9	33190		3.9	10	36.4	23204				
	4	1	74.9	33190	3.8	3	44	35206	ABD_Feb8c	8.8	3	31.5	20190	7.1	3	31.5	20190
	3.6	3	44	35206						3.1	5	47.3	5203				
	3.1	4	27.5	8166						3.3	9	15.8	14163				
	4.6	5	16.2	1165						3.2	11	19.4	8202				
	4.3	6	48	22101						3.2	11	44.6	9385				
	3.3	6	51.8	41362					ABD_Apr8c	3.2	1	24.9	5198	4.1	3	31.5	20190
	3.8	9	58.1	22293						5.3	3	31.5	20190	3.3	4	0	1791
	3.8	10	36.6	28221						3	6	51.8	29315	3.5	9	15.8	14163
	3.2	11	35.5	1122						4.8	9	15.8	14163	3.8	11	19.4	8202
Underwater Weight	3.1	1	0.8	4296						3.2	9	16.6	9265	5.1	11	44.6	9385

		Basic model				Genetic relatedness corrected model				Genetic relatedness corrected model				Basic model				Genetic relatedness corrected model			
	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Underwater Weight		3.5	1	8.4	16190					ABD_Apr8c	3.8	9	30.6	13128							
		3.5	1	24.9	16268						3.3	11	4.6	10372							
		4.6	1	43.8	981						3.7	11	19.4	8202							
		3.2	1	76.4	40181						3.7	11	36.2	6145							
		4.6	2	2.3	9181						3.1	11	36.2	24279							
		3.5	3	31.5	33157						6.6	11	44.6	9385							
		3.8	3	68.3	8395					ABD_Apr4c	3.6	1	67.2	10171	4.5	1	67.2	10171			
		3.3	4	21.2	4190						5.2	9	59.6	38128	5.2	9	59.6	38128			
		3.7	4	22	15176						3.6	10	65.6	689	3.2	10	65.6	689			
		3.6	4	22	16429						3	11	4.6	10372							
		4.2	4	22	31156						3	11	44.6	9385							
		3.1	4	27.5	8166					ABD_24hNov8c	3.2	4	22.8	1134	4.4	4	27.5	1068			
		4.2	5	16.2	1165						4.4	4	27.5	1068							
		3.7	5	32.1	8132						3.2	9	30.6	13128							
		3.6	5	41.1	39205						3.1	10	13.1	8160							
		4.2	6	13.3	6415						4	11	36.2	6145							
		3.6	6	48	22101					CookColY	3.6	1	5.4	19388	3.1	1	9.2	4188			
		3.2	6	51.8	41362						5.1	1	9.2	4188	3.5	3	31.5	20190			
		3.4	9	58.1	37131						3.6	2	0.8	27163	4.3	8	11.8	13104			
		3.6	10	0	2139						5.4	3	31.5	20190							
		3.4	10	36.4	23204						3.7	4	20.5	18309							
		4.2	10	36.6	28221						3.3	4	27.5	682							
		3	10	59.5	10517						3.1	4	27.5	12263							
		3.9	11	3	1233						3.1	4	81	39204							
		3.7	11	6.1	12320						3.2	5	16.2	1165							
		4.3	11	6.1	27355						3.7	5	31.3	18175							
		3.3	11	35.5	35239						3	6	13.7	1569							
		3.7	12	47.3	17135						5.8	6	16.7	10389							
	Cooking_Type	3.2	2	2.3	3612	3.3	2	2.3	3612		3.7	6	36.5	12229							

	Basic model				Genetic relatedness corrected model					Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Cooking_Type	3.3	4	22.8	5379	3.3	11	35.5	35239	CookColY	3.5	6	38.5	23142				
	3.2	4	27.9	1103						4.1	7	46.6	8411				
	3	10	0	762						3.3	8	10.6	1084				
	3	11	35.5	1122						4.3	8	11.8	13104				
	4.1	11	35.5	35239						4.3	8	25	22240				
										4.6	9	24.5	692				
										3.1	9	58.1	37131				
										3.7	12	43.6	7110				
FleshColY	3.4	1	8.8	6361	3	1	8.8	6361	FryingCol_Apr8c	4	4	19.7	19195	3.5	1	43.8	19230
	4	3	61.6	20432	3.4	5	8.5	26159		3.9	5	31.3	5156	3.2	2	3.7	2166
	3.4	4	27.5	16258	3.7	6	38.5	23142		3.2	5	32.1	1374	3.5	4	19.7	19195
	3.2	5	8.5	26159	3.2	9	16.6	5135		4.7	5	33.5	39199	3.3	5	31.3	5156
	3.4	5	31.3	18175	3.1	9	16.6	23430		3.3	7	53.9	14165	3.1	6	51.8	29315
	3	6	36.5	12229	3.1	12	8.8	37224		6.1	9	15.8	14163	3.3	9	15.8	14163
	5.2	6	38.5	23142						4.9	9	16.6	8230	4	9	16.6	12230
	3	7	56.5	23108						7.1	9	16.6	12230				
	4	9	16.6	5135						3.8	9	16.6	19132				
	3.1	9	16.6	12230						4.4	9	16.6	23430				
	4.2	9	16.6	23430						3.1	9	30.6	13128				
	4.1	9	24.5	1491						3.2	12	29.9	4194				
	3.6	10	14.1	5150					FryingCol_Apr4c	3.3	1	9.2	10225	3.6	9	16.6	12230
	3.2	12	8.8	37224						3.5	1	43.8	981	3.4	9	16.6	23430
FryingCol_Nov8c	3.1	1	24.9	16268						3.9	9	15.8	14163				
	3.5	5	33.5	39199						3.8	9	16.6	8230				
	3.6	9	16.6	8230						3.4	9	16.6	9265				
	3.4	9	30.4	15233						5.6	9	16.6	12230				
	3	10	13.1	8160						3.8	9	16.6	19132				
	3.7	10	36.4	23204						5.6	9	16.6	23430				
	3.3	11	36.2	6145						3	9	30.4	15233				

	Basic model				Genetic relatedness corrected model				Basic model				Genetic relatedness corrected model			
	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
	FryingCol_Feb8c	3.5	2	3.7	2166	3.5	2	3.7	2166	EnzBrow_30min	3.4	3	44	34264	3.3	5	36.4	3133
		4.1	2	8.4	8142	4.4	2	8.4	8142		3.7	5	36.4	3133	4.1	5	36.4	1190
		3	4	44.1	7191						3.2	5	36.4	7137	3.3	5	36.4	27346
		3.6	5	33.5	39199						4.5	5	36.4	1190	3.5	5	37.2	12290
		3.4	6	51.8	41362						3.8	5	36.4	27346	3.5	6	10.7	23376
		4.2	9	15.8	14163						3.3	5	36.4	27420				
		5.3	9	16.6	8230						4	5	37.2	12290				
		3.1	9	16.6	9265						3.5	6	10.7	23376				
		3.7	9	16.6	12230													
		3.1	9	16.6	19132													
		3.4	9	16.6	23430													
		4.1	9	30.4	15233													
		3.1	9	30.6	13128													
		3.3	10	36.4	23204													

Supplementary Table 2: Detected marker-trait associations between the MYML data set and the set of 720 mapped AFLP markers. To the left, the associations identified with the basic model are tabulated, to the right the ones with the relatedness corrected model. Each time the observed significance "-10logP", the chromosome, the map position in cM and the marker name are listed.

	Basic model				Genetic relatedness corrected model					Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Shape	5.3	1	14.6	11151	3.5	6	51.8	41362	Underwater Weight	3	1	0.8	4296	4	2	62.8	10366
	6.5	1	14.6	23186						3	1	6.9	15263				
	3.7	1	24.9	377						4	1	8.4	16190				
	3.9	1	24.9	12514						4	1	8.8	6361				
	6.2	1	43.8	981						3	1	9.2	3195				
	3.1	1	68.8	10194						5	1	9.2	10225				
	3.1	2	3.1	1148						11	1	9.2	11148				
	3.1	2	29.3	3226						4	1	9.2	17169				
	3.7	3	10	3145						6	1	9.2	18144				
	4.3	4	21.3	2181						9	1	9.2	25220				
	3.8	5	16.2	1165						10	1	9.2	38309				
	3.2	5	31.3	25244						6	1	11.9	7202				
	3.4	6	2.9	1227						5	1	14.6	23186				
	4.9	6	35	161						4	1	15.4	14200				
	7	6	51.8	41362						6	1	24.1	13324				
	3.3	8	37.6	15190						3	1	24.9	2194				
	5	9	16.6	23430						3	1	24.9	5198				
	3.3	9	30.6	13128						4	1	24.9	12158				
	3.9	9	58.1	22293						7	1	24.9	16268				
	3.5	10	0	762						3	1	24.9	28193				
	4.5	10	11.2	41125						5	1	24.9	38250				
	4.6	10	36.6	28221						4	1	24.9	41261				
	4.4	11	3	1233						3	1	31.1	9194				
	4.2	11	3.8	17288						9	1	43.8	981				
	6.4	11	6.1	12320						5	1	43.8	19230				
	4.4	11	6.1	27355						4	1	58	7394				

Trait	Basic model				Genetic relatedness corrected model			
	-10log*p*	Chromosome	Map position	Marker name	-10log*p*	Chromosome	Map position	Marker name
Shape	4.2	11	35.5	17303				
	3.2	12	0	8286				
	3.5	12	39.3	15116				
	5	12	40.5	23181				
Size	3.4	2	0.7	31267	3.2	4	15.9	1104
	4.1	2	7.5	3235				
	3.5	2	8.4	8142				
	4.1	4	15.9	1104				
	3.3	10	36.6	3121				
	3.3	10	49.6	25326				
	3.1	12	48.8	4163				
Maturity	4.5	1	6.9	15263	3.8	2	7.5	3235
	4.4	1	6.9	21278				
	5.4	1	8.8	6361				
	8.2	1	9.2	11148				
	4.8	1	9.2	25220				
	3.5	1	9.2	26111				
	4.9	1	9.2	38309				
	4.5	1	10.5	41138				
	6.9	1	11.9	7202				
	3.4	1	15.4	14200				
	4.6	1	24.9	12158				
	4.1	1	24.9	15118				
	8.1	1	24.9	16268				
	4.2	1	31.1	9194				
	4	1	69.6	17125				
	5.6	2	3.1	1148				
	4.2	2	6	4309				
	3.6	2	7.5	3235				

Trait	Basic model				Genetic relatedness corrected model			
	-10log*p*	Chromosome	Map position	Marker name	-10log*p*	Chromosome	Map position	Marker name
Underwater Weight	5	1	68.8	10194				
	8	1	69.6	39341				
	4	1	71.8	41149				
	3	1	76.4	40181				
	6	2	1.5	17445				
	7	2	3.1	1148				
	3	2	6	4309				
	4	2	8.4	8142				
	4	2	22.3	23139				
	7	4	15.9	1104				
	4	4	18.9	14246				
	5	4	19.7	19195				
	5	4	21.2	4190				
	7	4	21.3	2181				
	9	4	22	15176				
	6	4	22	16429				
	6	4	22	2098				
	5	4	22	31156				
	3	4	22.8	8314				
	4	4	22.8	14466				
	3	4	22.8	16346				
	8	4	22.8	30217				
	5	4	38.1	37430				
	4	4	43.3	5118				
	4	4	50.7	1188				
	9	5	16.2	1165				
	5	5	30.1	6299				
	3	5	31.3	9136				
	9	5	31.3	25244				

	Basic model				Genetic relatedness corrected model					Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name	Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Maturity	3.9	2	27.8	37238					Underwater Weight	4	5	32.1	28100				
	3.7	4	19.7	19195						14	5	32.1	3099				
	4.5	4	21.2	4190						5	5	32.1	38228				
	3.5	4	21.3	2181						3	5	35.6	13321				
	9.7	4	22	15176						5	5	36.4	3133				
	5.5	4	22	16429						3	5	36.4	18264				
	6.9	4	22	2098						3	5	36.4	2599				
	6.9	4	22	31156						4	5	36.4	27346				
	5.4	4	22.8	14466						6	5	41.1	39205				
	3.2	4	22.8	16346						4	6	0	1596				
	4.8	4	22.8	30217						5	6	5.1	3993				
	3.1	4	27.5	8166						6	6	5.9	14201				
	3.1	4	38.1	37430						3	6	7	18196				
	3.1	4	44.1	7191						4	6	11.2	10183				
	3.2	4	50.7	1188						5	6	11.9	30182				
	3.6	5	8.5	26159						3	6	13.3	6415				
	5.9	5	16.2	1165						6	6	14.3	11311				
	7.1	5	31.3	25244						5	6	20.3	26195				
	3.1	5	32.1	8132						3	6	31.9	35146				
	5.1	5	32.1	28100						5	6	35	161				
	9	5	32.1	3099						5	6	37	15283				
	4.3	5	32.1	38228						12	6	48	22101				
	3	5	35.6	13321						11	6	51.8	41362				
	3.6	5	36.4	3133						5	7	30.4	2109				
	3.5	5	36.4	9125						5	7	48.9	3132				
	3.5	5	36.4	2599						6	7	48.9	22346				
	3.5	5	36.4	27346						3	7	54.5	24480				
	3.7	6	0	16110						5	7	55.3	12145				
	4	6	2.9	1227						4	7	55.3	12166				

Trait	Basic model				Genetic relatedness corrected model			
	-10log$p$	Chromosome	Map position	Marker name	-10log$p$	Chromosome	Map position	Marker name
Maturity	3.2	6	5.9	14201				
	3.6	6	11.2	10183				
	5	6	14.3	11311				
	5	6	15.2	17465				
	3.2	6	15.2	17545				
	3.5	6	29.2	3140				
	3.7	6	35	161				
	3.7	6	39.6	3596				
	8.4	6	48	22101				
	5.9	6	51.8	41362				
	4.2	7	30.4	2109				
	4.5	7	48.9	3132				
	5.7	7	48.9	22346				
	3.2	7	53.2	1971				
	3.1	7	53.2	41100				
	4.9	7	55.3	12145				
	4.5	7	55.3	12166				
	5.1	8	11.8	1127				
	6.9	8	11.8	15111				
	3.1	8	12.6	12154				
	3.4	8	14.1	29137				
	9.4	9	16.6	4161				
	3.2	9	16.6	8230				
	5.3	9	16.6	11254				
	3.7	9	16.6	12230				
	4	9	16.6	16224				
	4.2	9	16.6	23430				
	5	9	40.3	39365				
	4.4	9	58.1	22293				
Underwater Weight	4	7	55.3	27336				
	6	7	55.4	3386				
	8	8	11.8	15111				
	6	8	14.1	29137				
	5	8	16.4	38368				
	4	8	22.6	8105				
	4	9	2.3	3591				
	4	9	3.1	1175				
	4	9	16.6	4161				
	5	9	16.6	8230				
	6	9	16.6	12230				
	5	9	16.6	16224				
	4	9	16.6	23430				
	4	9	24.5	1491				
	5	9	58.1	22293				
	3	9	58.1	36282				
	5	9	58.1	37131				
	6	9	59.6	18433				
	4	9	59.6	38128				
	5	10	0	2139				
	5	10	0	762				
	3	10	11.2	41125				
	3	10	35	8209				
	4	10	35	11529				
	4	10	35.8	27211				
	4	10	36.4	23204				
	9	10	36.6	28221				
	5	10	49.3	1178				
	4	10	49.6	25326				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Maturity	4	9	58.1	37131				
	5	9	59.6	18433				
	3.3	9	60.3	3764				
	4	9	64.4	12126				
	3.9	10	35	8209				
	3.6	10	35	11529				
	5	10	35	13307				
	4.7	10	35.8	27211				
	7.5	10	36.6	28221				
	3.1	10	41.8	19190				
	5.6	10	49.3	1178				
	4.1	10	49.6	25326				
	5.4	10	49.6	30140				
	4	10	59.5	10517				
	3	11	6.1	12320				
	9.4	11	35.5	1122				
	3.8	11	35.5	12182				
	3.3	11	45.2	15226				
	4.9	12	27	17181				
	3.5	12	37.1	8370				
	3.7	12	39.3	15116				
	4	12	43.6	1119				
	8	12	43.6	1131				
	3.6	12	43.6	7352				
	6.3	12	43.6	10185				
	3.2	12	43.6	26584				
Cooking_Type	3.8	1	9.2	11148	3.1	1	9.2	17169
	4.5	1	9.2	17169	4	6	22.9	19164
	3.8	1	9.2	38309				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Underwater Weight	7	10	49.6	30140				
	4	10	54.2	885				
	4	10	59.5	10517				
	6	11	3	1233				
	4	11	3.8	17288				
	4	11	3.8	34304				
	5	11	3.8	38215				
	7	11	6.1	12320				
	8	11	6.1	27355				
	8	11	35.5	1122				
	6	11	35.5	17303				
	4	11	38.5	2793				
	4	11	39.9	6327				
	4	12	27	17181				
	8	12	39.3	15116				
	3	12	39.3	23229				
	7	12	40.5	23181				
	3	12	42.1	30166				
	5	12	43.6	1119				
	7	12	43.6	1131				
	5	12	43.6	10185				
	4	12	43.6	10275				
	4	12	43.6	14230				
	6	12	43.6	24446				
	5	12	47.3	17135				
ABD_Feb8c	3	1	24.9	17400				
	3	2	13.5	19208				
	4	4	22.8	19244				
	4	6	46.4	5248				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
Cooking_Type	3.5	1	69.6	39341				
	4.4	2	1.5	17445				
	3.4	4	15.9	1104				
	3.8	6	20.3	26195				
	4.9	6	22.9	19164				
	3	6	29.2	6294				
	4.7	6	37	15283				
	4.4	6	51.8	41362				
	3	7	44.3	41263				
	3.4	8	11.8	15111				
	4.3	9	15.8	14163				
	4.7	9	16.6	8230				
	3.7	9	16.6	12230				
	5.1	9	16.6	16224				
	4.7	9	24.5	1491				
	3.1	10	36.4	23204				
	3.4	12	39.3	15116				
	4.8	12	40.5	23181				
	4.7	12	43.6	14230				
	3.8	12	43.6	24446				
ACD_1h	3.6	1	6.9	21278				
	3.4	1	9.2	11148				
	4.4	1	24.9	16268				
	3.3	1	31.1	9194				
	3.3	2	35.6	31277				
	3.5	4	11.4	3080				
	3	4	18.9	14246				
	3.2	4	21.2	4190				
	5.7	4	22	15176				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
ABD_Feb8c	5	7	44.3	41263				
	4	8	34	8117				
	4	9	30.6	13128				
ABD_Apr8c	3	7	44.3	41263				
	3	8	34	8117				
ABD_Apr4c	3	1	24.9	12514	3	1	24.9	17400
	4	1	24.9	17400	3	8	6.9	8133
	3	5	28.5	20241				
	3	6	36.5	1086				
	3	7	54.5	24480				
	3	8	6.9	8133				
ABD_24hNov8c	4	1	24.9	2328	4	1	49.2	23114
	4	1	24.9	28193	4	1	49.2	25317
	4	1	24.9	32151	3	6	37	15283
	6	1	49.2	23114				
	5	1	49.2	25317				
	4	1	68.8	33137				
	5	2	26.9	15147				
	4	4	23.4	7142				
	4	5	32.1	1374				
	4	6	37	15283				
	3	6	51.8	41362				
	3	9	16.6	16224				
	4	9	24.5	1491				
CookColY	3	1	15.7	40277				
	4	4	81	39204				
	3	5	31.3	18175				
	3	6	13.7	1569				
	5	6	38.5	23142				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
ACD_1h	4.3	4	22	2098				
	4.3	4	22.8	14466				
	3	4	22.8	30217				
	4.9	5	32.1	3099				
	3.4	6	5.1	3993				
	3.2	6	7	18196				
	3.1	6	48	22101				
	3.9	6	51.8	41362				
	3.8	10	59.5	10517				
	3.1	11	6.1	27355				
	5	11	35.5	1122				
	3.8	11	35.5	17303				
	3.8	12	40.1	13114				
ABD_Nov8c	3.5	1	23.4	27175	3.8	10	0	34127
	3.9	1	24.9	3305				
	5.1	1	24.9	9163				
	3.1	1	24.9	9483				
	4.7	1	24.9	12158				
	3.6	1	24.9	31287				
	5.8	1	24.9	38250				
	4.2	2	3.1	1148				
	3.4	2	3.1	30183				
	4.9	4	22.8	19244				
	3.8	5	28.5	20241				
	5.5	6	2.9	1227				
	3.2	6	30.7	21183				
	4	7	30.4	2109				
	3.2	7	53.9	14165				
	3.4	7	55.4	41325				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
CookColY	3	8	11.8	13104				
	4	9	16.6	5135				
	3	12	39.3	23229				
FleshColY	5	1	9.2	11148	5	3	31.5	20190
	3	1	15.7	40277	4	6	13.7	1569
	5	1	24.9	16268	3	6	38.5	23142
	3	1	43.8	981				
	4	1	65.7	32269				
	3	2	2.3	9181				
	4	2	7.5	3235				
	4	2	26.9	15147				
	4	4	22.8	16346				
	4	4	27.5	8166				
	3	4	28.2	1273				
	4	5	16.2	1165				
	3	5	31.3	30298				
	6	5	32.1	28100				
	3	5	52	32358				
	4	6	2.9	1227				
	3	6	12.2	8126				
	7	6	13.7	1569				
	4	6	16.7	10389				
	4	6	36.5	12229				
	9	6	38.5	23142				
	6	7	46.6	8411				
	4	7	48.9	3132				
	3	7	53.9	31297				
	4	8	8.3	38334				
	3	8	14.1	14214				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
ABD_Nov8c	3.5	8	34	8117				
	3.1	9	16.6	19132				
	3.7	9	16.6	23430				
	3.5	9	26	41113				
	5.3	9	30.6	13128				
	4	9	58.1	37131				
	3.7	10	0	34127				
	3.2	10	49.3	1178				
	5.3	11	4.6	10372				
	5.5	11	35.5	1122				
	4.7	11	46	16273				
	4.9	12	27	17181				
	3.1	12	39.3	3287				
	3.6	12	39.3	12293				
	3.5	12	39.3	29243				
	4.1	12	43.6	7352				
	3.9	12	43.6	10185				
	5.4	12	43.6	30234				
FryingCol_Nov8c	5.4	1	9.2	18144	3.4	10	57.6	8113
	3.7	1	24.9	2328				
	4.3	1	24.9	11309				
	4.5	1	24.9	17400				
	4.1	1	58	7394				
	3.2	2	8.4	8142				
	5.6	4	15.9	1104				
	4.4	4	27.5	28216				
	4.9	4	79.5	24347				
	3	5	31.3	25244				
	3	5	35.6	13321				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
FleshColY	4	8	14.1	29137				
	3	8	34	8117				
	3	9	16.6	11254				
	10	9	24.5	692				
	7	9	58.1	37131				
	3	10	35.8	27211				
	3	10	36.6	28221				
	4	11	3.9	3289				
	4	11	6.1	27355				
	3	11	35.5	1122				
	3	11	35.5	12182				
	4	11	49.8	15231				
	3	12	39.3	3287				
	4	12	39.3	23229				
	5	12	39.3	29243				
	5	12	40.1	13114				
	5	12	43.6	10185				
FryingCol_Apr8c	3	1	9.2	18144	4	6	35	161
	5	1	24.9	9131	4	9	15.8	14163
	3	1	24.9	9483				
	4	1	24.9	11309				
	5	1	24.9	17400				
	6	6	2.9	1227				
	4	6	35	161				
	3	8	11.8	1127				
	10	9	15.8	14163				
	10	9	16.6	8230				
	4	9	16.6	9231				
	4	9	16.6	9265				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
FryingCol_Nov8c	3.2	5	36.4	9300				
	5.4	6	2.9	1227				
	3.4	6	29.2	6294				
	3.1	6	37	15283				
	3.5	7	38.2	38185				
	3.2	7	44.3	41263				
	3.8	7	53.9	14165				
	3.2	7	55.3	12145				
	4.3	7	55.4	3386				
	3.5	8	9.8	2431				
	5.5	8	20.3	21283				
	9.2	9	15.8	14163				
	6.6	9	16.6	8230				
	3.5	9	16.6	9231				
	5.3	9	16.6	9265				
	8.8	9	16.6	12230				
	3.2	9	16.6	16224				
	4.1	9	16.6	19132				
	8.6	9	16.6	23430				
	3.8	9	24.5	1491				
	3.5	9	25.6	30107				
	3.3	9	30.4	15233				
	3.4	9	30.6	13128				
	3.4	10	13.1	8160				
	3.3	11	4.6	10372				
	3.7	11	36.2	24279				
	3.7	11	39.9	6327				
	3.3	11	44.6	9171				
	5.3	11	44.6	9385				

	Basic model				Genetic relatedness corrected model			
Trait	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
FryingCol_Apr8c	5	9	16.6	12230				
	5	9	16.6	16224				
	7	9	16.6	19132				
	6	9	16.6	23430				
	4	9	24.5	1491				
	3	9	30.4	15233				
	3	9	30.6	13128				
	5	10	7.7	33269				
	5	11	4.6	10372				
	4	11	36.2	6145				
	4	11	44.6	9171				
	4	11	44.6	9385				
	4	11	51.3	36341				
	3	12	27	17181				
	7	12	43.6	665				
	3	12	43.6	7352				
	4	12	43.6	14230				
	3	12	43.6	23146				
	3	12	43.6	26584				
FryingCol_Apr4c	3	1	33.4	20318	3	1	24.9	16268
	3	1	41.4	11269	3	6	6.6	11324
	3	5	31.3	9136	4	7	53.9	31297
	3	6	6.6	11324	3	9	15.8	14163
	3	7	48.9	5129	5	12	43.6	23146
	4	7	53.9	31297				
	4	8	9.8	2431				
	6	9	15.8	14163				
	5	9	16.6	8230				
	4	9	16.6	9231				

Trait	Basic model				Genetic relatedness corrected model			
	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
FryingCol_Nov8c	3.4	11	51.3	36341				
	3.9	12	43.6	12240				
	3.6	12	43.6	14230				
	3.3	12	43.6	24446				
FryingCol_Feb8c	3.9	1	9.2	18144	3	4	0	1791
	3.5	1	24.9	2328				
	3.9	1	24.9	377				
	5.2	1	24.9	17400				
	3.3	2	28.5	34169				
	5.9	4	15.9	1104				
	3.2	4	22.8	10112				
	4.4	4	79.5	24347				
	3.4	5	31.3	25244				
	3.3	5	32.1	1374				
	3.5	5	32.1	38228				
	4.7	6	2.9	1227				
	3.4	6	37	15283				
	4.4	7	44.3	41263				
	3.9	7	53.8	14218				
	3.3	7	53.9	14165				
	3.3	7	54.5	2388				
	4	7	55.4	3386				
FryingCol_Feb8c	3.1	8	20.3	21283				
	10.8	9	15.8	14163				
	7.7	9	16.6	8230				
	3.3	9	16.6	9231				
	4	9	16.6	9265				
	6	9	16.6	12230				
	4.2	9	16.6	16224				

Trait	Basic model				Genetic relatedness corrected model			
	-10logp	Chromosome	Map position	Marker name	-10logp	Chromosome	Map position	Marker name
FryingCol_Apr4c	3	9	16.6	9265				
	5	9	16.6	19132				
	3	9	16.6	23430				
	4	10	7.7	33269				
	4	11	36.2	24279				
	3	11	44.6	9171				
	5	12	43.6	665				
	5	12	43.6	12240				
	4	12	43.6	14230				
	6	12	43.6	23146				
EnzBrow_30min	4	1	56.5	31132	4	1	56.5	31132

Trait	Basic model				Genetic relatedness corrected model				Genetic relatedness corrected model			
	-10log*p*	Chromosome	Map position	Marker name	-10log*p*	Chromosome	Map position	Marker name	-10log*p*	Chromosome	Map position	Marker name
FryingCol_Feb8c	5	9	16.6	19132								
	4.5	9	16.6	23430								
	3.5	9	24.5	1491								
	3.7	9	30.4	15233								
	3.6	9	30.6	13128								
	3.2	10	36.4	23204								
	3.2	11	4.6	10372								
	3.4	11	20	1486								
	3.2	11	36.2	6145								
	3.6	11	36.2	24279								
	3.9	11	39.9	6327								
	5.1	11	44.6	9385								
	4	11	51.3	36341								
	3.7	12	43.6	665								
	3.6	12	43.6	7352								
	3.6	12	43.6	14230								
	3.2	12	43.6	24446								
	3.4	12	43.6	26584								

Appendix 1: Cultivars accompanied by their code used in project P8.

P8code	Cultivar	P8code	Cultivar	P8code	Cultivar	P8code	Cultivar	P8code	Cultivar
P80001	1256A(23) = Black 1256	**P80093**	Gladstone	**P80185**	SPUNTA	**P8B037**	HZPC06-37	**P8B129**	Averis06-15
P80002	Abundance (Sutton's)	**P80094**	GLORIA	**P80186**	Sunrise	**P8B038**	HZPC06-38	**P8B130**	Averis06-16
P80003	Ackersegen	**P80095**	Golden Wonder	**P80187**	Tahi	**P8B039**	VR06-01	**P8B131**	Averis06-17
P80004	Adirondack	**P80096**	Goya (2000)	**P80188**	Tasso	**P8B040**	VR06-02	**P8B132**	Averis06-18
P80005	ADORA	**P80097**	Great Scot	**P80189**	The Alness	**P8B041**	VR06-03	**P8B133**	Averis06-19
P80006	ADRETTA	**P80098**	HANSA	**P80190**	TIMATE	**P8B042**	VR06-04	**P8B134**	Averis06-20
P80007	AGATA	**P80099**	Herald	**P80191**	Tinwald's Perfection	**P8B043**	VR06-05	**P8B135**	Averis06-21
P80008	AGRIA	**P80100**	HERMES	**P80192**	Toyoshiro	**P8B044**	VR06-06	**P8B136**	Averis06-22
P80009	Ajiba	**P80101**	Hindenburg	**P80193**	Triplo	**P8B045**	VR06-07	**P8B137**	Averis06-23
P80010	Albion	**P80102**	Home Guard	**P80194**	Triumf	**P8B046**	VR06-08	**P8B138**	Averis06-24
P80011	ALCMARIA	**P80103**	Hydra	**P80195**	Triumph	**P8B047**	VR06-09	**P8B139**	Averis06-25
P80012	Allure	**P80104**	IMPALA	**P80196**	Ulster Glade	**P8B048**	VR06-10	**P8B140**	Averis06-26
P80013	Alma (1978)	**P80105**	Industrie	**P80197**	Ulster Knight	**P8B049**	VR06-11	**P8B141**	Averis06-27
P80014	ALMERA	**P80106**	INNOVATOR	**P80198**	Ulster Sceptre	**P8B050**	VR06-12	**P8B142**	Averis06-28
P80015	ALPHA	**P80107**	INOVA	**P80199**	ULTIMUS	**P8B051**	VR06-13	**P8B143**	Averis06-29
P80016	AM 66-42	**P80108**	Irish Cobbler	**P80200**	Umatilla Russet	**P8B052**	VR06-14	**P8B144**	Averis06-30
P80017	AM 70-2166	**P80109**	Irish Queen	**P80201**	Up to Date	**P8B053**	VR06-15	**P8B145**	Averis06-31
P80018	AM 78-3704	**P80110**	JAERLA	**P80202**	URGENTA	**P8B054**	VR06-16	**P8B146**	Averis06-32
P80019	AMOROSA	**P80111**	JAUNE D'OR = Fransen	**P80203**	USDA 96-56	**P8B055**	VR06-17	**P8B147**	Averis06-33
P80020	AMPERA	**P80112**	Jubel	**P80204**	VE 66-295	**P8B056**	VR06-18	**P8B148**	Averis06-34
P80021	AMYLA	**P80113**	KARNICO	**P80205**	VE 70-9	**P8B057**	VR06-19	**P8B149**	Averis06-35
P80022	ANOSTA	**P80114**	KARTEL	**P80206**	VE 71-105	**P8B058**	VR06-20	**P8B150**	Averis06-36
P80023	Aquila	**P80115**	Katahdin	**P80207**	VE 74-45	**P8B059**	VR06-21	**P8B151**	Averis06-37
P80024	ARCADE	**P80116**	KENNEBEC	**P80208**	VICTORIA	**P8B060**	VR06-22	**P8B152**	Averis06-38
P80025	ARINDA	**P80117**	Kepplestone Kidney	**P80209**	VIRGO	**P8B061**	VR06-23	**P8B153**	Meijer06-01
P80026	ARNOVA	**P80118**	Kerpondy	**P80210**	VIVALDI	**P8B062**	VR06-24	**P8B154**	Meijer06-02
P80027	Arran Banner	**P80119**	Kerr's Pink	**P80211**	VK 69-491	**P8B063**	VR06-25	**P8B155**	Meijer06-03
P80028	Arran Chief	**P80120**	King George	**P80212**	VORAN	**P8B064**	VR06-26	**P8B156**	Meijer06-04
P80029	Arran Pilot	**P80121**	KONDOR	**P80213**	VOYAGER	**P8B065**	VR06-27	**P8B157**	Meijer06-05
P80030	Arran Victory	**P80122**	KURAS	**P80214**	VTN 62-33-3	**P8B066**	VR06-28	**P8B158**	Meijer06-06
P80031	Arrow	**P80123**	KURODA	**P80215**	W 72-22-496	**P8B067**	VR06-29	**P8B159**	Meijer06-07
P80032	ASTARTE	**P80124**	LADY CHRISTL	**P80216**	Wauseon	**P8B068**	VR06-30	**P8B160**	Meijer06-08
P80033	ASTERIX	**P80125**	LADY CLAIRE	**P80217**	Wilja	**P8B069**	VR06-31	**P8B161**	Meijer06-09
P80034	Atlantic	**P80126**	LADY OLYMPIA	**P80218**	Winston	**P8B070**	VR06-32	**P8B162**	Meijer06-10

P8code	Cultivar	P8code	Cultivar	P8code	Cultivar	P8code	Cultivar	P8code	Cultivar
P80035	Aurora	**P80127**	LADY ROSETTA	**P80219**	Wisent	**P8B071**	VR06-33	**P8B163**	Meijer06-11
P80036	AUSONIA	**P80128**	Laura	**P80220**	Y 66-13-636	**P8B072**	VR06-34	**P8B164**	Meijer06-12
P80037	Avenance	**P80129**	Lenape	**P80221**	Yam	**P8B073**	VR06-35	**P8B165**	Meijer06-13
P80038	Ballydoon	**P80130**	LEYLA	**P80222**	Yukon Gold	**P8B074**	VR06-36	**P8B166**	Meijer06-14
P80039	BARTINA	**P80131**	LIBERTAS	**P80223**	SH83-92-488	**P8B075**	VR06-37	**P8B167**	Meijer06-15
P80040	Belle de Fontenay	**P80132**	LISETA	**P80224**	RH89-039-16	**P8B076**	VR06-38	**P8B168**	Meijer06-16
P80041	Bellini	**P80133**	Magnum Bonum	**P80225**	Merano	**P8B077**	AR06-01	**P8B169**	Meijer06-17
P80042	BERBER	**P80134**	Majestic	**P80226**	Pink Firr Apple	**P8B078**	AR06-02	**P8B170**	Meijer06-18
P80043	BILDTSTAR	**P80135**	MARFONA	**P80227**	Folva	**P8B079**	AR06-03	**P8B171**	Meijer06-19
P80044	BINTJE	**P80136**	MARIS PIPER	**P80228**	Nadine	**P8B080**	AR06-04	**P8B172**	Meijer06-20
P80045	Biogold	**P80137**	MARITIEMA	**P80229**	Panda	**P8B081**	AR06-05	**P8B173**	Meijer06-21
P80046	British Queen	**P80138**	Maritta	**P80230**	Arielle	**P8B082**	AR06-06	**P8B174**	Meijer06-22
P80047	CAESAR	**P80139**	MARKIES	**P80231**	Marabel	**P8B083**	AR06-07	**P8B175**	Meijer06-23
P80048	Champion	**P80140**	May Queen	**P80232**	Vitelotte Noir	**P8B084**	AR06-08	**P8B176**	Meijer06-24
P80049	Charlotte	**P80141**	MERCATOR	**P80233**	Quarta	**P8B085**	AR06-09	**P8B177**	Meijer06-25
P80050	Cherie	**P80142**	MONALISA	**P80234**	Princess	**P8B086**	AR06-10	**P8B178**	Meijer06-26
P80051	CILENA	**P80143**	MONDIAL	**P80235**	Ratte	**P8B087**	AR06-11	**P8B179**	Meijer06-27
P80052	CIVA	**P80144**	MORENE	**P80211bis**	VK 69-491 (HZPC)	**P8B088**	AR06-12	**P8B180**	Meijer06-28
P80053	CLIVIA	**P80145**	MPI 19268	**P80236**	Alma (1904)	**P8B089**	AR06-13	**P8B181**	Meijer06-29
P80054	Craigs Alliance	**P80146**	Myatt's Ashleaf	**P80237**	Stirling	**P8B090**	AR06-14	**P8B182**	Meijer06-30
P80055	Craigs Bounty	**P80147**	NICOLA	**P80238**	Iwa	**P8B091**	AR06-15	**P8B183**	Meijer06-31
P80056	Craigs Defiance	**P80148**	NINETYFOLD	**P80239**	Urineka	**P8B092**	AR06-16	**P8B184**	Meijer06-32
P80057	DAISY	**P80149**	NISKA	**P8B001**	HZPC06-01	**P8B093**	AR06-17	**P8B185**	Meijer06-33
P80058	Deodara	**P80150**	NOISETTE	**P8B002**	HZPC06-02	**P8B094**	AR06-18	**P8B186**	Meijer06-34
P80059	DESIREE	**P80151**	NOMADE	**P8B003**	HZPC06-03	**P8B095**	AR06-19	**P8B187**	Meijer06-35
P80060	Di Vernon	**P80152**	Noordeling	**P8B004**	HZPC06-04	**P8B096**	AR06-20	**P8B188**	Meijer06-36
P80061	DIAMANT	**P80153**	OBELIX	**P8B005**	HZPC06-05	**P8B097**	AR06-21	**P8B189**	Meijer06-37
P80062	DITTA	**P80154**	Ontario	**P8B006**	HZPC06-06	**P8B098**	AR06-22	**P8B190**	Meijer06-38
P80063	DONALD	**P80155**	PALLAS	**P8B007**	HZPC06-07	**P8B099**	AR06-23		
P80064	Doon Star	**P80156**	Parnassia	**P8B008**	HZPC06-08	**P8B100**	AR06-24		
P80065	Dorado	**P80157**	Paterson's Victoria	**P8B009**	HZPC06-09	**P8B101**	AR06-25		
P80066	Dore	**P80158**	Peerless	**P8B010**	HZPC06-10	**P8B102**	AR06-26		
P80067	Dr McIntosh	**P80159**	Pentland Dell	**P8B011**	HZPC06-11	**P8B103**	AR06-27		
P80068	Draga (1970)	**P80160**	Pepo (1919)	**P8B012**	HZPC06-12	**P8B104**	AR06-28		
P80069	Duke of York = Eersteling	**P80161**	PICASSO	**P8B013**	HZPC06-13	**P8B105**	AR06-29		

P8code	Cultivar	P8code	Cultivar	P8code	Cultivar	P8code	Cultivar	P8code	Cultivar
P80070	Early Rose	**P80162**	PREMIERE	**P8B014**	HZPC06-14	**P8B106**	AR06-30		
P80071	Edeltraut	**P80163**	PREVALENT	**P8B015**	HZPC06-15	**P8B107**	AR06-31		
P80072	EDEN (2000)	**P80164**	PRIMURA	**P8B016**	HZPC06-16	**P8B108**	AR06-32		
P80073	EHUD	**P80165**	PROFIJT	**P8B017**	HZPC06-17	**P8B109**	AR06-33		
P80074	Eigenheimer	**P80166**	Ramos	**P8B018**	HZPC06-18	**P8B110**	AR06-34		
P80075	Elisabeth	**P80167**	RECORD	**P8B019**	HZPC06-19	**P8B111**	AR06-35		
P80076	Eos	**P80168**	Red Duke of York = Rode Eersteling	**P8B020**	HZPC06-20	**P8B112**	AR06-36		
P80077	Epicure	**P80169**	RED SCARLETT	**P8B021**	HZPC06-21	**P8B113**	AR06-37		
P80078	ESCORT	**P80170**	REDSTAR	**P8B022**	HZPC06-22	**P8B114**	AR06-38		
P80079	ESTIMA	**P80171**	Reichskanzler	**P8B023**	HZPC06-23	**P8B115**	Averis06-01		
P80080	EXQUISA	**P80172**	REMARKA	**P8B024**	HZPC06-24	**P8B116**	Averis06-02		
P80081	FABULA	**P80173**	ROMANO	**P8B025**	HZPC06-25	**P8B117**	Averis06-03		
P80082	FELSINA	**P80174**	Rural New Yorker no.2	**P8B026**	HZPC06-26	**P8B118**	Averis06-04		
P80083	Festien	**P80175**	RUSSET BURBANK	**P8B027**	HZPC06-27	**P8B119**	Averis06-05		
P80084	FIANNA	**P80176**	SAMBA	**P8B028**	HZPC06-28	**P8B120**	Averis06-06		
P80085	Fichtelgold	**P80177**	SANTANA	**P8B029**	HZPC06-29	**P8B121**	Averis06-07		
P80086	Flava	**P80178**	SANTE	**P8B030**	HZPC06-30	**P8B122**	Averis06-08		
P80087	Flourball (Sutton's)	**P80179**	SASKIA	**P8B031**	HZPC06-31	**P8B123**	Averis06-09		
P80088	FONTANE	**P80180**	SATURNA	**P8B032**	HZPC06-32	**P8B124**	Averis06-10		
P80089	FRESCO	**P80181**	SERESTA	**P8B033**	HZPC06-33	**P8B125**	Averis06-11		
P80090	FRIESLANDER	**P80182**	Shamrock	**P8B034**	HZPC06-34	**P8B126**	Averis06-12		
P80091	Fruhmolle	**P80183**	SHEPODY	**P8B035**	HZPC06-35	**P8B127**	Averis06-13		
P80092	FURORE	**P80184**	SIRTEMA	**P8B036**	HZPC06-36	**P8B128**	Averis06-14		

This project was financed by the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research.

Cover design by Björn D’hoop and Martijn Hendriksen.
Cover pictures by Björn D’hoop.