Plant Breeding **132**, 718–724 (2013) doi:10.1111/pbr.12102

Plant Breeding

Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome-wide high-density genotyping using the SolCAP SNP array

Benjamin Stich[1], Claude Urbany[1], Per Hoffmann[2] and Christiane Gebhardt[1,3]

[1]Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829, Köln, Germany; [2]Institute for Human Genetics, Biomedical Center, Sigmund-Freud-Straße 25, 53127, Bonn, Germany; [3]Corresponding author, E-mail: gebhardt@mpipz.mpg.de

With 5 figures and 1 table

Received October 15, 2012 / Accepted July 2, 2013
Communicated by J. Léon

Abstract

Genome-wide association (GWA) mapping in potato requires high-density genotyping. With the Illumina SolCAP potato single-nucleotide polymorphism (SNP) array, a first tool for GWA mapping in potato became available. Thirty-six tetraploid varieties and eight diploid breeding clones were genotyped for 8303 SNPs using this array. The objectives of our study were to examine in this set of germplasm: (i) the degree of polymorphism of the SolCAP SNPs in European germplasm, (ii) the population structure, (iii) temporal trends of genetic diversity and (iv) the genome-wide extent of linkage disequilibrium (LD). Three-quarters of the SNPs were polymorphic. In the principal coordinate analysis, a clear separation of tetraploid from diploid genotypes was observed, whereas no distinct subgroups among the tetraploid varieties were detected. The nonlinear trendline of the LD measure r^2 vs. the physical map distance decayed within 275 bp to an r^2 value of 0.10, indicating that theoretically, about 3 million equally distributed SNPs are required for GWA mapping in this diverse set of germplasm. As the LD decay changes with the population selected for GWA mapping, the number of required markers might be different in other germplasm.

Key words: potato — SNP array — diploid and tetraploid

The cultivated potato (*Solanum* Group *tuberosum*) (Spooner et al. 2007) is the world's fourth most important crop after rice, wheat and maize. Recent results suggested a single domestication event for cultivated potato from *S. brevicaule* (Spooner et al. 2005) which occurred in the Andes Mountains of South America about 10 000 years ago. Since this time, different plant breeding techniques (for review see Simmonds 1993) have been used to adapt potato to the human requirements. Due to the tetraploidy of potato and the high number of selection criteria (Barocka and Ross 1985), annual gain from selection in potato breeding was in the last 50 years lower than for other crop species (Douches et al. 1996). The gain from selection, however, can be potentially increased by using marker-assisted selection approaches.

A prerequisite for marker-assisted selection in potato is the identification of markers, which are diagnostic for superior or inferior alleles at the loci contributing to the variation of a trait of interest (quantitative trait loci, QTL). Until recently, such markers have been identified in potato using linkage mapping approaches (e.g. Leonards-Schippers et al. 1994, Zimnoch-Guzowska et al. 2000, Bradshaw et al. 2004, , 2008, Khu et al. 2008). The problem of such approaches, which are based on individual bi-parental segregating populations, is the low mapping resolution due to the low number of meioses that have occurred. Furthermore, the limited genetic background represented by bi-parental segregating population decreases the probability that the identified QTL can be broadly applied in marker-assisted selection programmes (for review see Stich and Melchinger 2010).

In contrast, association mapping approaches are based on germplasm collections rather than individual bi-parental segregating populations. This allows to make use of historical recombination events and results in a high mapping resolution (Ersoz et al. 2007). Furthermore, when association mapping is performed in appropriate breeding materials, the results are expected to be broadly applicable (Stich and Melchinger 2010). The most important constraint to the use of association mapping is the increased rate of false-positive associations, when the germplasm studied is genetically structured, that is, some genotypes are more closely related than the average (cf. Thornsberry et al. 2001). Therefore, this aspect requires special attention.

A limited number of association mapping experiments have been performed so far in potato. Most of those were based on genotyping of candidate genes (e.g. Gebhardt et al. 2004, Li et al. 2005, Malosetti et al. 2007, Achenbach et al. 2009, Pajerowska-Mukhtar et al. 2009, Urbany et al. 2011). Although the candidate gene approach was successful in identifying marker–trait associations, it is biased, because the candidate genes are selected based on information available from genetic, biochemical or physiological studies in both model and non-model plant species. The candidate gene approach is only straightforward when restricted to traits, of which the biochemical and molecular basis is well understood (Hall et al. 2010). Furthermore, association mapping based on candidate genes always runs the risk of missing important, but unknown genes. In addition, candidate genes are often initially discovered from loss-of-function mutations in laboratory strains, and it is not clear how well such mutations describe the variation that actually underlies quantitative trait variation in natural populations (Hall et al. 2010).

One association study used genome-wide distributed amplified fragment length polymorphism (AFLP) markers for genotyping (D'hoop et al. 2008). However, AFLP markers are almost exclusively dominant which introduces a number of problems when using them for estimating population structure (Ritland 2005) or for direct use in mapping (Hall et al. 2010). Moreover, AFLPs are anonymous and cannot readily be linked to the potato genome sequence (Xu et al. 2011). Single-nucleotide polymor-

phisms (SNPs) are the markers of choice for genome-wide association (GWA) mapping, as SNP alleles are co-dominant and the allele dosage in polyploid species can be estimated (Rickert et al. 2003). In 2010, the Illumina SolCAP SNP genotyping array became available for genome-wide high-density SNP genotyping in potato. The SNPs were derived from the North American varieties 'Atlantic' (1976), 'Premier Russet' (2006), 'Snowden' (1990), 'Kennebec' (1948) and 'Shepody' (1980), and the old variety 'Bintje' (1910) from the Netherlands (Hamilton et al. 2011).

With this study, we want to set the stage for GWA mapping in potato using high-density SNP genotyping. Thus, our objectives were to examine: (i) how informative the SolCAP potato array would be in a set of tetraploid varieties bred over the past hundred years in Europe, (ii) whether population structure was detectable in this material, (iii) whether trends of genetic diversity can be observed over time and (iv) the genome-wide extent of linkage disequilibrium (LD). For comparison, we included in the analysis some diploid potato clones, which were parents of experimental populations used previously for linkage mapping of QTL.

Materials and Methods

Plant materials: Our study was based on 36 tetraploid and eight diploid potato clones (Table 1). The tetraploid varieties were selected based on the following criteria: (i) varieties that were important progenitors in the 20th century potato breeding history in Europe (http://www.plantbreeding.wur.nl/potatopedigree/index.html) and (ii) varieties with contrasting phenotypes for susceptibility to tuber bruising (Urbany et al. 2011). The diploid clones included the parents of F_1 families, which have been used to construct molecular maps and to map qualitative and quantitative traits in potato (references in Table 1). Three diploid clones carried, based on pedigree information, genetic material from wild potato species. The clone P40 is an interspecific hybrid between *S. tuberosum* and *S. speggazzinii* (Barone et al. 1990). DG83 has in its pedigree *S. tuberosum*, *S. chacoense* and *S. yungasense* (Zimnoch-Guzowska et al. 2000), and G87 is a mosaic genome with contributions of *S. tuberosum*, *S. kurtzianum*, *S. tarijense*, and *S. vernei*. Genomic DNA was extracted from freeze-dried leaves as described by Li et al. (2005)

Table 1: Potato varieties and clones used for single-nucleotide polymorphisms analyses

Variety or clone	Year of release[2]	Breeder (country)[2]	Ploidy	Origin of plant material
Adretta	1975	NORIKA (D)	4n	SaKa Pflanzenzucht GbR, 2010
Agila	2006	NORIKA (D)	4n	Urbany et al. (2011)
Agria	1985	Kartoffelzucht Böhm (D)	4n	SaKa Pflanzenzucht GbR, 2010
Alcmaria	1969	PGV (NL)	4n	MPI Plant Breeding Research, 1994
Aquila	1942	Biologische Reichsanstalt für Land und Forstwirtschaft (D)	4n	KKS 10234, 2010 [1]
Aspirant	2004	EUROPLANT (D)	4n	Urbany et al. (2011)
Assia	1980	Uniplanta (D)	4n	Gebhardt et al. (1989)
Berber	1983	Friese Maatschappij (NL)	4n	Goerg et al. (1992)
Calla	1990	EUROPLANT (D)	4n	Urbany et al. (2011)
Cara	1973	Dr. H.W. Kehoe, Oak Park Research, Carlow (IRL)	4n	KKS 12020, 2010
Carmona	2000	SaKa Pflanzenzucht (D)	4n	Urbany et al. (2011)
Désireé	1962	HZPC (NL)	4n	MPI Plant Breeding Research, 1994
Fitis	2005	NORIKA (D)	4n	Urbany et al. (2011)
Flava	1931	Pommersche Saatzucht Gesellschaft (D)	4n	KKS 10083, 2010
Gala	2002	NORIKA (D)	4n	Urbany et al. (2011)
Granola	1975	Saka Pflanzenzucht (D)	4n	MPI Plant Breeding Research, 1994
Hindenburg	1916	Von Kameke (D)	4n	KKS 10332, 2010
Impala	1989	Kartoffelzucht Böhm (D)	4n	SaKa Pflanzenzucht GbR, 2010
Industrie	1900	H. Modrow (D)	4n	KKS 10015, 2010
Innovator	1999	HZPC (NL)	4n	SaKa Pflanzenzucht GbR, 2010
Kolibri	1998	NORIKA (D)	4n	Urbany et al. (2011)
Krone	2002	Bavaria Saat (D)	4n	Urbany et al. (2011)
Kuba	2005	Niehoff (D)	4n	Urbany et al. (2011)
Lady Rosetta	2004	Meijer (NL)	4n	Urbany et al. (2011)
Logo	2003	Saatzucht Firlbeck (D)	4n	Urbany et al. (2011)
Lolita	2003	Saatzucht Firlbeck (D)	4n	Urbany et al. (2011)
Maris Piper	1963	Plant Breeding Institute, Trumpington, Cambridge (UK)	4n	KKS 10024, 2010
Maxilla	1981	NORIKA (D)	4n	Urbany et al. (2011)
Nicola	1973	Saatzucht Soltau-Bergen (D)	4n	Goerg et al. (1992)
Olga	2003	EUROPLANT (D)	4n	Urbany et al. (2011)
Panda	1986	Saka Pflanzenzucht (D)	4n	Urbany et al. (2011)
Quarta	1979	EUROPLANT (D)	4n	Urbany et al. (2011)
Rafaela	2004	Saka Pflanzenzucht (D)	4n	Urbany et al. (2011)
Remarka	1993	HZPC (NL)	4n	Urbany et al. (2011)
Saturna	1964	E. Scholten (NL)	4n	Gebhardt et al. (1989)
Schwalbe	1956	Inst. für Pflanzenzüchtung, Groß-Lüsewitz (D)	4n	KKS 10574, 2010
P3	–	MPI Plant Breeding Research	2n	Schaefer-Pregl et al. (1998)
P18	–	MPI Plant Breeding Research	2n	Gebhardt et al. (2003)
P38	–	MPI Plant Breeding Research	2n	Schaefer-Pregl et al. (1998)
P40	–	MPI Plant Breeding Research	2n	Gebhardt et al. (2003)
P49	–	MPI Plant Breeding Research	2n	Leonards-Schippers et al. (1994)
P54	–	MPI Plant Breeding Research	2n	Menendez et al. (2002)
DG83	–	Mlochow Research Center (PL)	2n	Zimnoch-Guzowska et al. (2000)
G87	–	GERMICOPA, Quimper (F)	2n	Oberhagemann et al. (1999)

[1] Accession No. of the IPK gene bank, Groß-Lüsewitz, Germany.
[2] Sources: van Berloo et al. (2007); Bundessortenamt (2005).

and purified using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) according to the supplier's instructions.

SNP genotyping: The forty-six potato individuals were genotyped for 8303 SNPs using the SolCAP Potato genotyping array (Hamilton et al. 2011). The array was processed by the Department of Genomics, Life & Brain Center, Bonn on an Illumina iScan system using the Infinium assay. For each SNP locus, one of five possible genotypes (AAAA, AAAB, AABB, ABBB or BBBB) was manually assigned to the individuals using the GenomeStudio Software version 2011.1 (Illumina). In the possible case that there is a third allele present in the population, this would not be detected by the assay. Wherever possible, the cluster positions on the SNP Graph of the diploid individuals with the genotype AA, AB or BB were used as reference for the tetraploid individuals.

Statistical analyses: From the entire set of 8303 SNPs, those with an Illumina GenTrain score < 0.6 were discarded. Furthermore, all SNPs with a minor allele frequency < 0.05 as well as a rate of missing data > 20 % were excluded. The SNPs which passed these filters were used for the following analyses.

The Rogers' distance (RD) was calculated according to Rogers (1972). Associations among the groups were revealed with principal coordinate analysis (Gower 1966) based on RD estimates between all pairs of accessions as well as between all pairs of tetraploid clones. The gene diversity was calculated according to Nei (1987). In order to adjust for the different sample sizes, a resampling simulation was performed as described by Stich et al. (2005). The polymorphic information content (PIC) value was calculated for each SNP according to Botstein et al. (1980).

Bi-locus LD was characterized by the LD measure r^2 (Hill and Robertson 1968, Achenbach et al. 2009), which was calculated for all pairs of SNPs located on the same superscaffold (Xu et al. 2011). Curves of decay of LD with physical distance were fitted by nonlinear regression, where the expectation of r^2 between adjacent sites (Hill and Weir 1988) was considered (cf. Heuertz et al. 2006). Due to the small sample size of diploid clones, LD was only examined in the set of tetraploid clones. For all analyses described above, the dosage of the alleles at the individual SNPs was considered.

All analyses have been performed using the statistical software R (R Development Core Team 2011).

Results

From the total number of 8303 SNPs available on the SolCAP SNP array, 6119 SNPs fulfilled the criteria applied during SNP filtering. The PIC values of these SNPs ranged for the entire set of 44 potato clones from 0.04 to 0.50 with an average of 0.35 (Fig. 1).

Rogers' distance estimates between all pairs of clones ranged from 0.14 to 0.37 (Fig. 2). In principal coordinate analysis based on RD estimates of all clones, the first two principal coordinates explained 14.2% and 9.2% of the molecular variance (Fig. 3). The tetraploid potato clones were clearly separated from the diploid clones with respect to the first principal coordinate. The diploid clones DG83, G87 and P40 were clearly separated from the other five diploid clones by the second principal coordinate.

In principal coordinate analysis based on RD estimates of the 38 tetraploid potato clones, the first two principal coordinates explained 9.1% and 6.8% of the molecular variance (Fig. 4). This analysis revealed no distinct subgroups among the tetraploid potato clones. We observed no significant ($\alpha = 0.05$) correlation between the RD estimates among all pairs of tetraploid potato clones and the difference in the year of release. Furthermore, the gene diversity calculated for four groups of tetraploid potato clones based on their release dates (1900–1960, 1961–1980, 1981–2000, and 2001–2010) revealed no significant ($\alpha = 0.05$) difference.

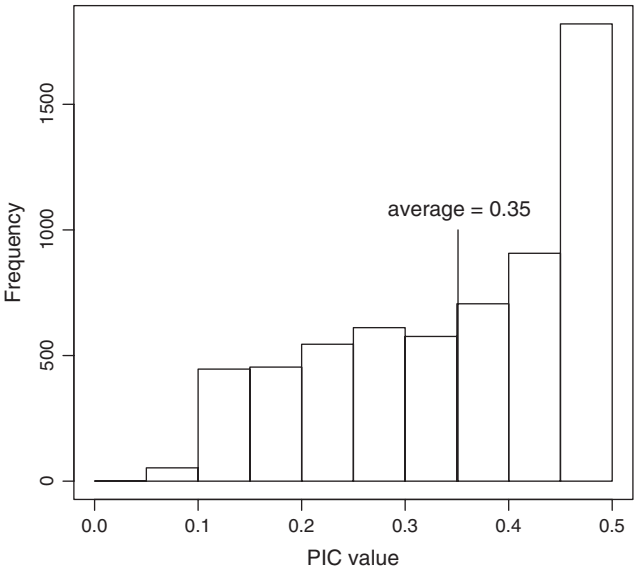

Fig. 1: Distribution of polymorphic information content (PIC) values calculated for all 44 potato clones across the 6119 single-nucleotide polymorphism markers

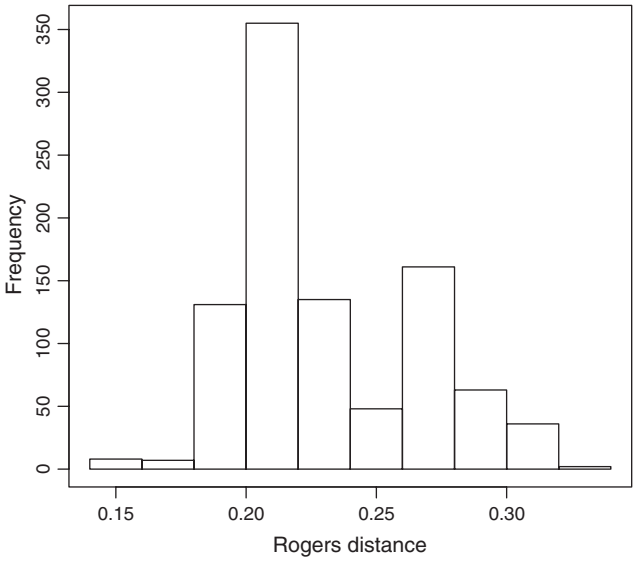

Fig. 2: Distribution of Rogers' distance estimates among all pairs of potato clones

The proportion of loci pairs with r^2 values > 0.1 ($P_{r^2 > 0.1}$) was 19.8% in the set of tetraploid potato clones, where $P_{r^2 > 0.8}$ was 0.7%. There was a rapid reduction in the magnitude of r^2 as physical map distance between the SNP markers increased (Fig. 5). The trendline from the nonlinear regression reached an r^2 value of 0.1 at 265 bp.

Discussion

In a pilot experiment, we evaluated how informative the SolCAP potato genotyping array would be for European potato varieties released over the last hundred years. Thirty-six tetraploid varieties and eight diploid clones of various origin were genotyped for 8303 SNPs using the Illumina IScan platform. The genotypes for each SNP were manually called, as the available GenomeStudio software was not able to perform this task automatically for tetraploid individuals. The eight diploid individuals included in

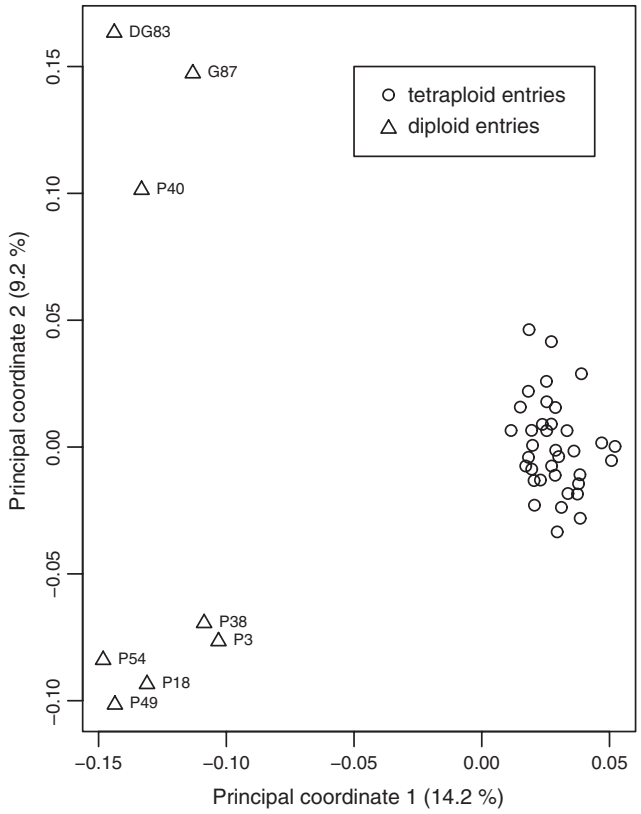

Fig. 3: Principal coordinate analysis of 8 diploid and 36 tetraploid potato clones based on Rogers' distance estimates. Numbers in parentheses refer to the proportion of variance explained by the principal coordinate

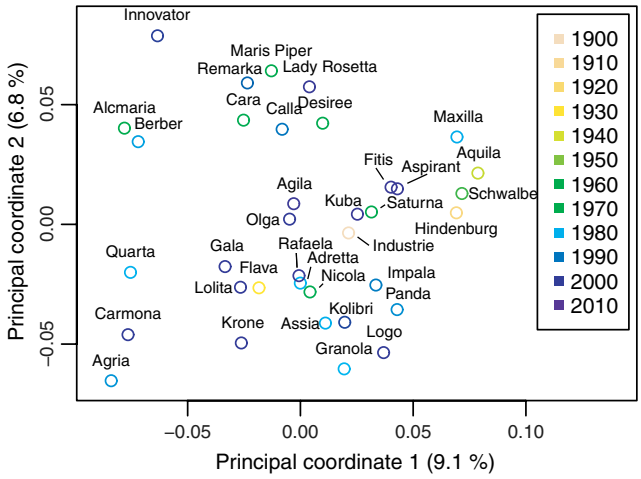

Fig. 4: Principal coordinate analysis of 36 tetraploid potato clones based on Rogers' distance estimates. Numbers in parentheses refer to the proportion of variance explained by the principal coordinate. The different colours indicate the year of release of the corresponding clone

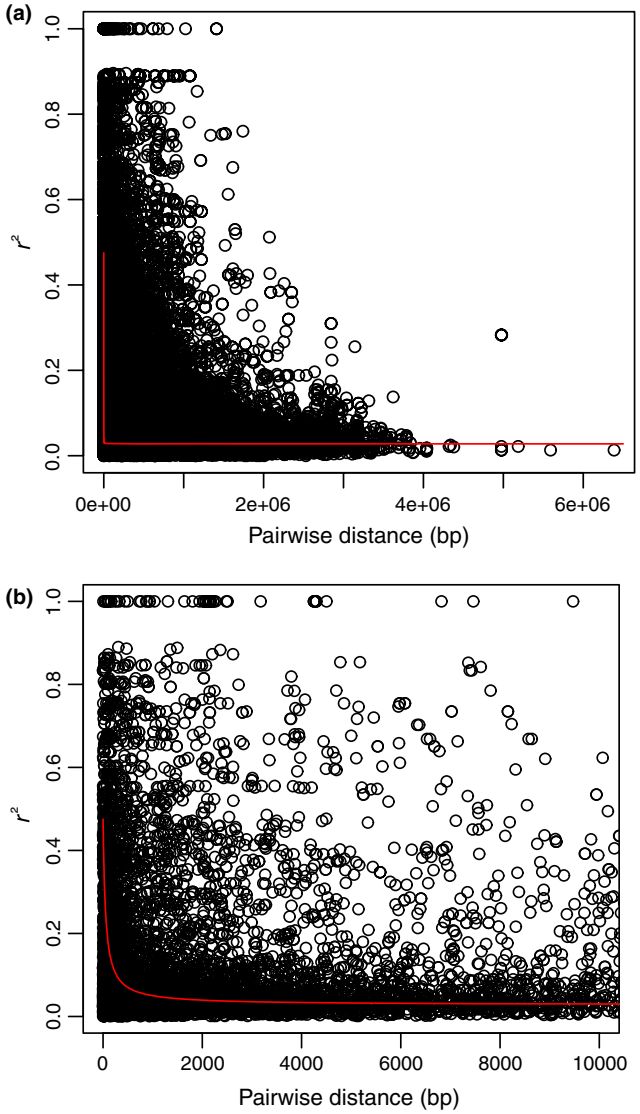

Fig. 5: Linkage disequilibrium measure r^2 plotted vs. the physical map distance between all pairs of single-nucleotide polymorphism (SNP) markers from the same superscaffold (top) between all pairs of SNP markers from the same superscaffold located within 10 000 bp (bottom) calculated for all 36 tetraploid potato clones. The trend line of the non-linear regression of r^2 vs. the physical map distance between the SNP markers is indicated in red

the set of tetraploid varieties were highly useful for this purpose, because the fluorescence signals for the genotypes AA, AB and BB marked the positions of tetraploid AAAA, AABB and BBBB genotypes and thereby facilitated the identification of AAAB and ABBB genotypes. After genotype calling, we excluded from the total number of 8303 SNPs present on the SolCAP potato genotyping array 959 SNPs (11.5%) with an Illumina GenTrain score < 0.6 (e.g. Namjou et al. 2009). The GenTrain score evaluates the confidence of the genotyping for each SNP on all samples and depends on the distance between the clusters and the fluorescence intensity. This proportion of SNPs with a low GenTrain score is in good accordance with the value reported for pea (Deulvot et al. 2010). Our finding suggested that with manually edited genotype classes, the SolCAP potato genotyping array can provide also for tetraploid species SNP genotyping results with similar reliability as for diploid species. Recently, a software package for genotype calling in tetraploid species was published (Voorrips et al. 2011), which might solve this problem in the future.

From the remaining SNPs, only 23 had to be removed due to a high rate of missing data and 1202 SNPs based on the minor allele frequency threshold. Furthermore, 99.1 % of the SNPs showed a PIC value ≥ 0.1 and 44.6 % a PIC value ≥ 0.4 (Fig. 1). These results demonstrate that high polymorphism rates are detectable in historical and modern European germplasm with the SolCAP potato genotyping array, which has been developed based on five potato varieties from North

America (Atlantic, Kennebec, Premier Russet, Shepody and Snowden) and one European variety (Bintje) (Hamilton et al. 2011). The large number of SNP alleles shared between North American and European varieties can be explained by the history, according to which the potato was introduced into North America from Europe (Love 1999). The SolCAP potato genotyping array can therefore be broadly applied for SNP genotyping in the *S. tuberosum* gene pool. It might be less suitable though in detecting more recent introgressions originating from national breeding programmes, for example, of resistance traits (Ross 1986).

Population structure and genetic diversity of *S. tuberosum*

The principal coordinate analysis revealed a clear separation between the tetraploid and the diploid potato clones with respect to the first principal coordinate (Fig. 3). Our finding is in disagreement with the results of Simko et al. (2006), who analysed 1 monoploid, 17 diploid and 29 tetraploid potato clones with 1100 SNPs present in approximately 25 kbp potato genomic sequence. These authors observed no subgroups, despite the fact that seven of the eight diploid genotypes of our study were also included in the study of Simko et al. (2006). The reason for this discrepancy might be the lower number and low genome coverage of the SNPs analysed by Simko et al. (2006) or differences between the methods for detecting subgroups. The clear separation between the tetraploid and diploid genotypes in our study can explain why molecular markers that are diagnostic for a given trait in diploid linkage mapping populations might be of low diagnostic value in tetraploid varieties and breeding populations (Niewöhner et al. 1995). In addition, our results indicate that association analysis should be performed separately for diploid and tetraploid potato genotypes.

The principal coordinate analysis revealed furthermore a clear separation of the diploid genotypes DG83, G87 and P40 from the other five diploid genotypes by the second principal coordinate (Fig. 4). The reason for this observation might be that, based on pedigree information, large proportions of the genomes of DG83, G87 and P40 originated from wild Solanum species (Barone et al. 1990, Oberhagemann et al. 1999, Zimnoch-Guzowska et al. 2000). The distinct group of five diploid genotypes originated from the diploid breeding programme conducted at the MPI for Plant Breeding Research in the seventies and eighties of the last century (Gebhardt et al. 1989). As there is no pedigree information available for these clones, it is not possible to assess whether and which wild species contributed to their genomes.

In order to examine in detail the population structure of the tetraploid potato clones, a principal coordinate analysis was performed separately for these clones. This analysis revealed no distinct subgroups for the tetraploid potato clones (Fig. 4). Our finding is in accordance with earlier results (Gebhardt et al. 2004, Simko et al. 2004a,b, Malosetti et al. 2007, D'hoop et al. 2008) and can be explained by the fact that potato breeders disregard population structure when choosing the parents of a cross. This practice is expected to lead to the absence of distinct subpopulations.

The principal coordinate analysis revealed no correlation between population structure and the release date, despite a wide range of release dates with more than hundred years were covered in our study. Furthermore, the correlation between the RD estimates and the difference between the release dates were not significantly ($\alpha = 0.05$) different from 0. Our observation is in contrast to the results of D'hoop et al. (2010), who observed in a set of 430 tetraploid potato cultivars groups differing in year of market release. This difference might be due to the considerably higher number of cultivars evaluated by D'hoop et al. (2010) compared to our study, which increases the power to detect such correlations.

The analysis of the temporal trends of gene diversity of the tetraploid potato clones grouped by their release dates did not reveal significant differences ($\alpha = 0.05$). This finding also contrasts studies in other crop species (e.g. Malysheva-Otto et al. 2007, Bus et al. 2011) and might be due to our limited sample size. On the other hand, this finding might be because introgression and incorporation (Simmonds 1993) were more frequently applied in potato breeding than in other crop species. Therewith, our results suggest that potato breeding over the last hundred years neither significantly reduced nor dramatically increased genetic diversity. Thus, no genetic diversity enhancement programmes are urgently needed for European potato breeding programmes.

Genome-wide LD in tetraploid potato

In contrast to linkage mapping, association mapping uses the LD which is present in a germplasm set. We observed that the nonlinear trendline of the LD measure r^2 vs. the physical map distance decayed within 275 bp to an r^2 value of 0.10 (Fig. 5). This finding is in contrast to earlier results of tetraploid potato. Simko et al. (2006) observed that the r^2 value reached a value of 0.21 within about 1 kb distance and a value of 0.10 at a distance of about 10 cM, which suggests a considerably slower decay of LD compared to our study. The reason for this might be the different ways the SNPs were sampled in the two studies. Whereas Simko et al. (2006) used SNPs located within 100 bp, which were derived from the ends of BAC (bacterial artificial chromosome) insertions of approximately 70 kbp, the SNPs on the SolCAP array originated from genome-wide-distributed loci with, on average, much larger physical distances between them (Hamilton et al. 2011). The former study's estimate of LD was based therefore on shorter physical distances than the estimate of our study, with the result of larger LD estimates. Furthermore, the BACs, of which the end sequences were analysed by Simko et al. (2006), have been selected for containing resistance (R) gene homologs (Rickert et al. 2003). Genomic regions containing R genes frequently show large sequence differences between alleles due to introgressions from wild species, which reduces recombination rates and thereby increases LD in such regions.

D'hoop et al. (2010) reported that LD decayed on average at about 5 cM to r^2 values of 0.1. This figure was derived from 3300 AFLP and 650 microsatellite alleles. Based on a size of the potato genome of 844 Mbp (Xu et al. 2011) and a length of the genetic map in the order of 1000 cM (Gebhardt et al. 1991), 5 cM corresponds to approximately 4 Mbp. This estimate for LD decay is orders of magnitude larger than the value found in our study and indicates the presence of large haplotype blocks in the population analysed by D'hoop et al. (2010).

Considering a complex trait, a large QTL may explain approximately 10 % of the phenotypic variation. If a marker is in LD with the trait-coding locus with an r^2 value of 0.1, this marker explains 10 % of the total QTL variation and consequently 1 % of the phenotypic variation (Ersoz et al. 2007). Results from computer simulations suggest that with currently used sizes of association mapping populations of about 300 genotypes such marker–trait associations could be detected with a reasonable statistical power (cf. Yu et al. 2006, Stich and Melchinger 2009).

The observed very rapid decay of LD suggests that theoretically, about 3 million genome-wide equally distributed SNPs are required to reach this goal in potato. According to this figure, the chance of detecting marker–trait associations by genotyping 8000 SNPs would be very low. In contrast, the association mapping experiments performed so far in populations of tetraploid potato genotypes did identify marker–trait associations by probing only minute physical portions of the potato genome, mostly at candidate loci for the traits of interest (Gebhardt et al. 2004, Simko et al. 2004a, Li et al. 2005, , 2008, Malosetti et al. 2007, Pajerowska-Mukhtar et al. 2009) but also when using a few hundred random AFLP markers (D'hoop et al. 2008). This discrepancy can have several reasons. First, the very rapid decay of LD observed in our study might be due to the use of a set of highly diverse genotypes when compared to the populations of tetraploid varieties and advanced breeding clones used for the abovementioned association mapping experiments. Second, the candidate loci are indeed the trait-coding loci or within short physical distance from the trait-coding loci.

Acknowledgements

This study was funded by the Max-Planck society and the German Federal Ministry for Economy and Technology (project ZIM-KF2408401). The authors thank Dr Josef Strahwald (Saka Pflanzenzucht GbR, Windeby, Germany) for suggestions for the selection of clones and Sandra Mäurer for technical assistance. The authors declare no conflict of interest.

References

Achenbach, U., J. Paulo, E. Ilarionova, J. Lubeck, J. Strahwald, E. Tacke, H. R. Hofferbert, and C. Gebhardt, 2009: Using SNP markers to dissect linkage disequilibrium at a major quantitative trait locus for resistance to the potato cyst nematode *Globodera pallida* on potato chromosome V. Theor. Appl. Genet. **118**, 619–629.

Barocka, K.-H., and H. Ross, 1985: *Lehrbuch der Züchtung landwirtschaftlicher Kulturpflanzen*, chapter Knollen-und Wurzelarten, pages 212–287. Paul Parey.

Barone, A., E. Ritter, U. Schachtschabel, T. Debener, F. Salamini, and C. Gebhardt, 1990: Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. Mol. Gen. Genet. **224**, 177–182.

van Berloo, R., R. Hutten, H. vanEck, and R. Visser, 2007: An online potato pedigree database resource. Potato Res. **50**, 45–57.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis, 1980: Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32**, 314–331.

Bradshaw, J. E., B. Pande, G. J. Bryan, C. A. Hackett, K. McLean, H. E. Stewart, and R. Waugh, 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.

Bradshaw, J. E., C. A. Hackett, B. Pande, R. Waugh, and G. J. Bryan, 2008: QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum subsp tuberosum*). Theor. Appl. Genet. **116**, 193–211.

Bundessortenamt (editor) 2005: Beschreibende Sortenliste Kartoffeln. Deutscher Landwirtschaftsverlag GmbH, Hannover.

Bus, A., N. Körber, R. Snowdon, and B. Stich, 2011: Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*. Theor. Appl. Genet. **123**, 1413–1423.

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

D'hoop, B. B., M. J. Paulo, K. Kowitwanich, M. Sengers, R. G. F. Visser, H. J. van Eck, and F. A. van Eeuwijk, 2010: Population structure and linkage disequilibrium unravelled in tetraploid potato. Theor. Appl. Genet. **121**, 1151–1170.

Deulvot, C., H. Charrel, A. Marty, F. Jacquin, C. Donnadieu, I. Lejeune-Henaut, J. Burstin, and G. Aubert, 2010: Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics **11**, 468.

Douches, D. S., D. Maas, K. Jastrzebski, and R. W. Chase, 1996: Assessment of potato breeding progress in the USA over the last century. Crop Sci. **36**, 1544–1552.

Ersoz, E. S., J. Yu, and E. Buckler, 2007: *Genomics-assisted crop improvement*, chapter Application of linkage disequilibrium and association mapping in crop plants, pages 97–119. Springer.

Gebhardt, C., C. Blomendahl, U. Schachtschabel, T. Debener, F. Salamini, and E. Ritter, 1989: Identification of 2n breeding lines and 4n varieties of potato *Solanum tuberosum* ssp *tuberosum* with RFLP fingerprints. Theor. Appl. Genet. **78**, 16–22.

Gebhardt, C., E. Ritter, A. Barone, T. Debener, B. Walkemeier, U. Schachtschabel, H. Kaufmann, R. D. Thompson, and M. W. Bonierbale, 1991: RFLP maps of potato and their alignment with the homoeologous tomato genome. Theor. Appl. Genet. **83**, 49–57.

Gebhardt, C., B. Walkemeier, H. Henselewski, A. Barakat, M. Delseny, and K. Stuber, 2003: Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. Plant J. **34**, 529–541.

Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schuler, 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. Mol. Breed. **13**, 93–102.

Goerg, R., U. Schachtschabel, E. Ritter, F. Salamini, and C. Gebhardt, 1992: Discrimination among 136 tetraploid potato varieties by fingerprints using highly polymorphic DNA markers. Crop Sci. **32**, 815–819.

Gower, J. C., 1966: Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika **53**, 325–338.

Hall, D., C. Tegström, and P. K. Ingvarsson, 2010: Using association mapping to dissect the genetic basis of complex traits in plants. Brief. Funct. Genomics **9**, 157–165.

Hamilton, J. P., C. N. Hansey, B. R. Whitty, K. Stoffel, A. N. Massa, A. Van Deynze, W. S. De Jong, D. S. Douches, and C. R. Buell, 2011: Single nucleotide polymorphism discovery in elite north american potato germplasm. BMC Genomics **12**, 302.

Heuertz, M., E. De Paoli, T. Kallman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux, and N. Gyllenstrand, 2006: Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. Genetics **174**, 2095–2105.

Hill, W. G. and A. Robertson, 1968: Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38**, 226–231.

Hill, W. G. and B. S. Weir, 1988: Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. **33**, 54–78.

Khu, D. M., J. Lorenzen, C. A. Hackett, and S. L. Love, 2008: Interval mapping of quantitative trait loci for corky ringspot disease resistance in a tetraploid population of potato (*Solanum tuberosum subsp tuberosum*). Am. J. Potato Res. **85**, 129–139.

Leonards-Schippers, C., W. Gieffers, R. Schafer-Pregl, E. Ritter, S. J. Knapp, F. Salamini, and C. Gebhardt, 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., J. Strahwald, H. R. Hofferbert, J. Lubeck, E. Tacke, H. Junghans, J. Wunder, and C. Gebhardt, 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, L., M.-J. Paulo, J. Strahwald, J. Luebeck, H.-R. Hofferbert, E. Tacke, H. Junghans, J. Wunder, A. Draffehn, F. van Eeuwijk, and C. Gebhardt, 2008: Natural DNA variation at candidate loci is associated with

potato chip color, tuber starch content, yield and starch yield. Theor. Appl. Genet. **116**, 1167–1181.

Love, S. L., 1999: Founding clones, major contributing ancestors, and exotic progenitors of prominent North American potato cultivars. Am. J. Potato Res. **76**, 263–272.

Malosetti, M., C. G. van der Linden, B. Vosman, and F. A. van Eeuwijk, 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.

Malysheva-Otto, L., M.W. Ganal, J. R. Law, J. C. Reeves, and M. S. Röder, 2007: Temporal trends of genetic diversity in European barley cultivars (*Hordeum vulgare* L.). Mol. Breed. **20**, 309–322.

Menendez, C. M., E. Ritter, R. Schaefer-Pregl, B. Walkemeier, A. Kalde, F. Salamini, and C. Gebhardt, 2002: Cold sweetening in diploid potato: mapping quantitative trait loci and candidate genes. Genetics **162**, 1423–1434.

Namjou, B., C. Gray-McGuire, A. L. Sestak, G. S. Gilkeson, C. O. Jacob, J. T. Merrill, J. A. James, E. K. Wakeland, Q. Z. Li, C. D. Langefeld, J. Divers, J. Ziegler, K. L. Moser, J. A. Kelly, K. M. Kaufman, and J. B. Harley, 2009: Evaluation of *C1q* genomic region in minority racial groups of lupus. Genes Immun. **10**, 517–524.

Nei, M. 1987: Molecular Evolutionary Genetics. Colombia University Press, New York.

Niewöhner, J., F. Salamini, and C. Gebhardt, 1995: Development of PCR assays diagnostic for RFLP marker alleles closely linked to alleles Gro1 and H1, conferring resistance to the root cyst nematode *Globodera rostochiensis* in potato. Mol. Breed. **1**, 672–678.

Oberhagemann, P., C. Chatot-Balandras, R. Schafer-Pregl, D. Wegener, C. Palomino, F. Salamini, E. Bonnel, and C. Gebhardt, 1999: A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Mol. Breed. **5**, 399–415.

Pajerowska-Mukhtar, K., B. Stich, U. Achenbach, A. Ballvora, J. Lübeck, J. Strahwald, E. Tacke, H. R. Hofferbert, E. Ilarionova, D. Bellin, B. Walkemeier, R. Basekow, B. Kersten, and C. Gebhardt, 2009: Single nucleotide polymorphisms in the *allene oxide synthase 2* gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. Genetics **181**, 1115–1127.

R Development Core Team 2011: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3- 900051-07-0.

Rickert, A. M., J. H. Kim, S. Meyer, A. Nagel, A. Ballvora, P. J. Oefner, and C. Gebhardt, 2003: First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. Plant Biotechnol. J. **1**, 399–410.

Ritland, K., 2005: Multilocus estimation of pairwise relatedness with dominant markers. Mol. Ecol. **14**, 3157–3165.

Rogers, J. S., 1972: Measures of genetic similarity and genetic distance. In: Wheeler, M.R., (ed.) Studies in Genetics VII, 145–173. University of Texas Publication 7213.

Ross, H., 1986: Advances in Plant Breeding 13. Potato Breeding Problems and Perspectives. Paul Parey, Berlin, Germany.

Schaefer-Pregl, R., E. Ritter, L. Concilio, J. Hesselbach, L. Lovatti, B. Walkemeier, H. Thelen, F. Salmini, and C. Gebhardt, 1998: Analysis of quantitative trait loci (QTLs) and quantitative trait alleles (QTAs) for potato tuber yield and starch content. Theor. Appl. Genet. **97**, 834–846.

Simko, I., S. Costanzo, K. G. Haynes, B. J. Christ, and R. W. Jones, 2004a: Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor. Appl. Genet. **108**, 217–224.

Simko, I., K. G. Haynes, E. E. Ewing, S. Costanzo, B. J. Christ, and R. W. Jones, 2004b: Mapping genes for resistance to *Verticilliumalboatrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. Mol. Genet. Genom. **271**, 522–531.

Simko, I., K. G. Haynes, and R. W. Jones, 2006: Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics **173**, 2237–2245.

Simmonds, N. W., 1993: Introgression and Incorporation - Strategies for the use of crop genetic-resources. Biol. Rev. Camb. Philos. Soc. **68**, 539–562.

Spooner, D. M., K. McLean, G. Ramsay, R. Waugh, and G. J. Bryan, 2005: A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proc. Natl Acad. Sci. U.S.A. **102**, 14694–14699.

Spooner, D. M., J. Nunez, G. Trujillo, M. d. R. Herrera, F. Guzman, and M. Ghislain, 2007: Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. Proc. Natl Acad. Sci. U.S.A. **104**, 19398–19403.

Stich, B. and A. E. Melchinger, 2009: Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics **10**, 94–94.

Stich, B. and A. Melchinger, 2010: Association mapping in plants. CAB Rev. **5**, 39.

Stich, B., A. E. Melchinger, M. Frisch, H. P. Maurer, M. Heckenberger, and J. C. Reif, 2005: Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor. Appl. Genet. **111**, 723–730.

Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E. S. Buckler, 2001: *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28**, 286–289.

Urbany, C., B. Stich, L. Schmidt, L. Simon, H. Berding, H. Junghans, K. H. Niehoff, A. Braun, E. Tacke, H. R. Hofferbert, J. Lubeck, J. Strahwald, and C. Gebhardt, 2011: Association genetics in *Solanum tuberosum* provides new insights into potato tuber bruising and enzymatic tissue discoloration. BMC Genomics **12**, 7.

Voorrips, R. E., G. Gort, and B. Vosman, 2011: Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics **12**, 172.

Xu, X., S. Pan, S. Cheng, B. Zhang, D. Mu, P. Ni, G. Zhang, S. Yang, R. Li, J. Wang, G. Orjeda, F. Guzman, M. Torres, R. Lozano, O. Ponce, D. Martinez, G. De la Cruz, S. K. Chakrabarti, V. U. Patil, K. G. Skryabin, B. B. Kuznetsov, N. V. Ravin, T. V. Kolganova, A. V. Beletsky, A. V. Mardanov, A. Di Genova, D. M. Bolser, D. M. A. Martin, G. Li, Y. Yang, H. Kuang, Q. Hu, X. Xiong, G. J. Bishop, B. Sagredo, N. Mejia, W. Zagorski, R. Gromadka, J. Gawor, P. Szczesny, S. Huang, Z. Zhang, C. Liang, J. He, Y. Li, Y. He, J. Xu, Y. Zhang, B. Xie, Y. Du, D. Qu, M. Bonierbale, M. Ghislain, M. del Rosario Herrera, G. Giuliano, M. Pietrella, G. Perrotta, P. Facella, K. O'Brien, S. E. Feingold, L. E. Barreiro, G. A. Massa, L. Diambra, B. R. Whitty, B. Vaillancourt, H. Lin, A. Massa, M. Geoffroy, S. Lundback, D. DellaPenna, C. R. Buell, S. K. Sharma, D. F. Marshall, R. Waugh, G. J. Bryan, M. Destefanis, I. Nagy, D. Milbourne, S. J. Thomson, M. Fiers, J. M. E. Jacobs, K. L. Nielsen, M. Sonderkaer, M. Iovene, G. A. Torres, J. Jiang, R. E. Veilleux, C. W. B. Bachem, J. de Boer, T. Borm, B. Kloosterman, H. van Eck, E. Datema, B. T. L. Hekkert, A. Goverse, R. C. H. J. van Ham, and R. G. F. Visser, 2011: Genome sequence and analysis of the tuber crop potato. Nature **475**, 189.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, J. F. Doebley, M. D. Mc-Mullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler, 2006: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**, 203–208.

Zimnoch-Guzowska, E., W. Marczewski, R. Lebecka, B. Flis, R. Schaefer-Pregl, F. Salamini, and C. Gebhardt, 2000: QTL analysis of new sources of resistance to *Erwinia carotovora* ssp. *atroseptica* in potato done by AFLP, RFLP, and resistance-gene-like markers. Crop Sci. **40**, 1156–1167.