The double reduction landscape in tetraploid potato as revealed by a high-density linkage map


*Wageningen UR Plant Breeding, Wageningen University and Research Centre, P.O. Box 386, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands.
Running title
Double reduction landscape in tetraploids

Key words
Linkage mapping, tetraploid, double reduction, potato, multivalents.

1Corresponding author
Chris Maliepaard. Mail: Wageningen UR Plant Breeding, Wageningen University and Research Centre, P.O. Box 386, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands. Phone: +31 (0) 317 480855. E-mail: chris.maliepaard@wur.nl
ABSTRACT

The creation of genetic linkage maps in polyploid species has been a long-standing problem for which various approaches have been proposed. In the case of autopolyploids, a commonly-used simplification is that random bivalents form during meiosis. This leads to relatively straightforward estimation of recombination frequencies using maximum likelihood from which a genetic map can be derived. However, autopolyploids such as tetraploid potato (*Solanum tuberosum* L.) may exhibit additional features such as double reduction, not normally encountered in diploid or allopolyploid species. In this study we produced a high-density linkage map of tetraploid potato and used it to identify regions of double reduction in a bi-parental mapping population. The frequency of multivalents required to produce this degree of double reduction was determined through simulation. We also determined the effect that multivalents or preferential pairing between homologous chromosomes have on linkage mapping. Low levels of multivalents or preferential pairing do not adversely affect map construction when highly-informative marker types and phases are used. We reveal the double reduction landscape in tetraploid potato, clearly showing that this phenomenon increases with distance from the centromeres.
INTRODUCTION

Polyploid species constitute a very important group among cultivated crops. Polyploids themselves can be further divided into auto- and allo-polyploids, with autopolyploids showing random association between homologous chromosomes and allopolyploids showing non-random or preferential pairing during meiosis. Linkage mapping in autopolyploid species remains a challenging exercise despite recent advances in genotyping technology and mapping methodology. Breeding work in many autopolyploid crops has yet to benefit from the use of markers in breeding programs. This is partly due to the lack of software to perform linkage mapping and QTL analysis in polyploids, but is also due to the complicated nature of autopolyploid genomes and genetics. The software program TetraploidMap (HACKETT AND LUO 2003) is a notable exception to this, but is constrained by the relatively low numbers of markers it can handle (currently 800 is the maximum) and the need to manually assign marker phase which may become infeasible with large datasets.

One autopolyploid species where large advances in genetic analysis have been made is tetraploid potato (*Solanum tuberosum* L.), in terms of the availability of a high-quality reference sequence (PGSC 2011), many published linkage maps (MEYER *et al.* 1998; VAN OS *et al.* 2006; FELCHER *et al.* 2012; HACKETT *et al.* 2013) as well as methods for performing linkage mapping at the polyploid level (LUO *et al.* 2001; BRADSHAW *et al.* 2004; HACKETT *et al.* 2013). In comparison to other economically-important autotetraploid species such as alfalfa, rose or leek, the pairing behavior of potato is thought to be relatively well-understood, with random bivalent pairing during prophase I of meiosis being generally assumed (SWAMINATHAN AND HOWARD 1953; MILBOURNE *et al.* 2008). Although a certain proportion of multivalents is known to occur, these are not deemed to occur at a sufficient frequency to merit their inclusion in a pairing model (BRADSHAW 2007).

The simplest marker segregation type to map in a tetraploid cross are simplex x nulliplex marker types which are expected to segregate in a 1:1 fashion. In a tetraploid, we employ the term simplex x nulliplex to collectively refer to 1x0, 3x0, 3x4 and 1x4 markers (with 0x1, 0x3, 4x3 and 4x1 markers being nulliplex x simplex). A relabeling of allele dosages is sufficient to convert all these markers to their simpler form. These have traditionally been the markers most favored in tetraploid mapping because of their simple segregation, reliability in genotype-calling and high information content in coupling-phase. One important practical advantage is that these markers can be mapped using advanced mapping software developed for diploids such as JoinMap (VAN OIJEN 2006) which can efficiently map large numbers of
markers as well as providing many checks on map and data quality. Simplex x nulliplex markers also provide the clearest linkage information to cluster markers into separate homologous chromosomes, forming the basis of homologue maps. In our population, simplex x nulliplex markers were also the most abundant marker segregation type. We therefore restricted our analysis to simplex x nulliplex markers, which nevertheless allowed us to map a total of 3273 markers across both parents.

Simplex x nulliplex markers are also the most useful markers to provide direct evidence of one of the observable consequences of multivalent formation, namely double reduction (DR). In autopolyploid species, pairing may occur between all homologous chromosomes which can lead to complicated pairing structures during the first meiotic division (Milbourne et al. 2008). In cases where a cross-over occurs between two sets of sister chromatids which subsequently migrate to the same pole, it is possible for a chromatid and its recombinant copy (segment) to end up in the same gamete, a situation which can never occur in diploids. For a simplex x nulliplex marker with the segregating allele on the recombinant segment in question, this can lead to a duplex score in that offspring.

By simulating comparable mapping populations genotyped with the same mapped markers we were able to estimate the rate of multivalent formation that would account for the observed levels of DR. We also performed a simulation study using populations with different rates of multivalent formation and preferential pairing to investigate the effect that the assumption of random bivalent formation has on the estimation of recombination frequency and marker phase.

MATERIALS AND METHODS

Plant material
An F1 mapping population of 237 individuals was created from the cross between two tetraploid potato varieties, cultivars ‘Altus’ (hereafter referred to as parent one, P1) and ‘Colomba’ (P2).

DNA extraction and genotyping
DNA was extracted from leaf material using KingFisher Flex according to the manufacturer’s instructions (Thermo Scientific). The concentration of DNA was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) and the DNA concentration was adjusted to ~50 ng µl⁻¹ (Vos et al. 2015). For DNA concentrations in the range of 25 – 50 ng µl⁻¹ the
sample was also used; samples having concentrations lower than 25 ng µl⁻¹ were discarded and DNA isolation was performed again. The samples were genotyped on the SolSTW Infinium SNP array which assayed 17,987 SNPs as described by (Vos et al. 2015). Of these SNPs, 4179 also form part of the SolCap SNP array (Felcher et al. 2012). The arrays were processed according to the manufacturer’s protocol at ServiceXS, Leiden, the Netherlands. Each parent was genotyped in duplicate using two biological replicates. 1662 other tetraploid accessions were sampled in a similar fashion, as well as 516 diploid accessions (for use in another study as well as helping marker dosage fitting).

**Assignment of dosages**

The X and Y allele signal intensities were imported from the Illumina data output into the R programming environment (R Core Team 2015). SNPs were initially filtered so that the average of their total signal intensity (the sum of the X and Y allele signal intensities) over all samples was greater than 0.2. The marker intensities were converted into allele dosages using the fitTetra package for R (Voorrips et al. 2011). Changes to the default settings of the `saveMarkerModels` function of fitTetra were as follows: `p.threshold` was decreased from 0.99 to 0.95, `peak.threshold` was increased from 0.85 to 0.99 and `sd.target` was set to 0.04, where `p.threshold` is the “minimum P-value required to assign a genotype to a sample”, `peak.threshold` is the “maximum allowed fraction of the scored samples that are in one peak” and `sd.target` is used to specify the maximum non-penalised standard deviation of the fit on a transformed scale (Voorrips et al. 2011). All diploid and tetraploid samples were included in the fitting because this generally results in a better fit of the dosage classes. Following fitting with fitTetra, the marker dosage scores were screened to ensure consistency between parental and offspring genotypes. Markers with up to 3% invalid scores (scores that were not expected based on the parental genotypes and bivalent chromosome pairing) were allowed. A high frequency of many invalid scores suggests that either the marker performed poorly, there was some consistent error in dosage assignment, or one or both of the parents had been incorrectly genotyped. Highly-skewed markers (p < 0.001) were also removed at this stage.

**Marker conversion**

Markers that segregated in a 1:1 fashion were re-labelled as simplex x nulliplex (or nulliplex x simplex) for mapping and double reduction analysis. Considering markers whose segregating
allele is inherited from P1, these consisted of triplex x nulliplex, triplex x quadruplex and simplex x quadruplex markers. For example, a triplex x nulliplex marker is expected to produce 50% dosage ‘1’ and 50% dosage ‘2’ among the offspring, with observable double reduction scores appearing as dosage ‘0’ (a double-copy of the ‘0’ allele from P1). Relabelling ‘2’ as ‘0’ and ‘0’ as ‘2’ (with the parents re-labelled as simplex and nulliplex) achieves the desired result of marker conversion.

**Linkage map construction**

Simplex x nulliplex marker data were recoded to JoinMap 4.1 cross-pollinator format (lm x ll). ‘Impossible’ genotypes (invalid scores) were made missing before importation into JoinMap. One pair of identical individuals was identified in the dataset (similarity of 0.9922), therefore we removed individual #202. Markers were assigned to linkage groups with a minimum LOD of 4 (a higher LOD was used if clusters broke into large sub-clusters at a higher LOD). Marker clusters were assigned to physical chromosomes based on the position of markers on the physical sequence (PGSC 2011). Mapping was first performed using the groupings from the Groupings Tree using maximum likelihood (ML). Homologues were then identified by large gaps in the estimated map distances ($\geq 60$ cM), which was also often accompanied by a transition in estimated marker phase. Marker data for separate homologues was exported from JoinMap in .loc files and re-imported for creation of the homologue maps. After an initial mapping of the homologues, individual #067 was found to contribute unrealistic numbers of recombinations in many linkage groups across both parents and was therefore removed, resulting in a final mapping population of 235 individuals. Mapping was performed using ML with three rounds of map optimization using the default settings for spatial sampling thresholds. Haldane’s mapping function was used to convert recombination frequency estimates to map distances as has previously been used for linkage map construction in tetraploid potato (MEYER et al. 1998; HACKETT et al. 2013). In a number of cases, we used linkage information from the duplex x nulliplex and simplex x simplex markers to connect sub-homologue linkage groups that had poor internal linkage among simplex x nulliplex markers. Map data was exported from JoinMap as text files and imported into MapChart 2.3 (VOORRIPS 2002) for further plotting.

**Comparison of genetic and physical maps**

The genetic positions of markers were compared with their physical positions as defined in (VOS et al. 2015). It was found that some markers did not map to the same chromosome as expected from the physical map; a list of such markers is included in Supplementary table 1.
The physical position of the centromere boundaries was initially adopted from previously-published values (Sharma et al. 2013). These were not found to coincide precisely with the points of inflection on the genetic-physical map, following which the approximate centromere bounds were re-defined by examination of the aligned genetic-physical plots (also referred to as Marey Maps (Chakravarti 1991)) and calculating an approximate physical position between marker pairs flanking the points of inflection on these plots. The order of the genetic map was reversed in cases where the genetic maps were found to be inversely ordered with respect to the physical map.

**Conversion rate of physical to genetic distance**

The conversion rate between genetic and physical distance was determined by regressing the genetic positions on their physical positions per homologue arm. The slopes of the regression lines for each homologue arm were tested for equality in an analysis of covariance by introducing, where necessary, up to three dummy variables (to code for the presence or absence of a homologue) per chromosome arm per parent (Andrade and Estévez-Pérez 2014). An average genome-wide estimation of the genetic to physical conversion rate was calculated after excluding a single outlying value from the northern arm of homologue 2 of chromosome 1 in parent 1. This genome-wide recombination rate was used to convert the physical map to a pseudo-integrated genetic map for use in the simulation studies.

**Rates of double reduction**

After recoding the 1:1 segregating marker data, duplex marker scores in the offspring were taken as possible evidence for double reduction. Duplex scores can also arise as a result of genotyping errors. Therefore, we used a relatively strict criterion to decide if such scores were evidence of double reduction: a string of three consecutive duplex scored markers on a homologue map was required in order to be considered strong enough evidence for double reduction. This could theoretically lead to some under-estimation of the rates of double reduction, but the simplex marker density was sufficient that in most cases a double reduction region would contain at least three (segregating) simplex x nulliplex markers.

A routine was written in R to identify strings of three or more duplex scores. The rate of double reduction was determined for each marker by counting the number of times it formed part of a double reduction segment and dividing this by the number of non-missing values scored for that marker across the population. We then derived the average rate of double reduction for 1Mb windows north and south of the centromeric bounds by calculating the
mean rate of double reduction over all markers within that window. These means were aggregated to give a single average rate of double reduction for each 1Mb window distance from the centromeres across all chromosomes and both parents.

**Simulation of double reduction and prediction of quadrivalent formation**

An approximate “integrated” genetic linkage map was produced using the average cM to Mb conversion rate and physical positions of the simplex markers. Only markers for which the assigned linkage group and physical chromosome corresponded were considered. Marker phase was determined according to the homologue assignment of all markers. Phased marker genotypes and a consensus genetic map position are the basic input for the simulation software PedigreeSim (VOORRIPS AND MALIEPAARD 2012), which simulates (diploid or) polyploid populations with specified levels of multivalents and / or preferential pairing. One thousand separate populations of 235 individuals were generated using the same simplex marker data and approximated map under a range of different fractions of quadrivalents. The algorithm for estimating double reduction was applied to the simulated datasets, allowing us to deduce the relationship between the rate of double reduction and the frequency of multivalents underlying meiosis.

**Estimation of the rate of preferential pairing**

Repulsion-phase simplex marker data can be used to investigate whether preferential pairing occurs, as the estimates for recombination frequency in repulsion are expected to differ under disomic and tetrasomic inheritance (QU AND HANCOCK 2001). We have adapted the approach of (QU AND HANCOCK 2001) to correct for multiple-testing using the false-discovery rate (FDR) (BENJAMINI AND HOCHBERG 1995), confining our analysis to within chromosomes to reduce the overall number of tests (coupling or repulsion linkage have no meaning when marker pairs from separate linkage groups are considered).

For two markers A and B we define \( n_{00} \) as the number of individuals with dosage 0 at both markers, \( n_{01} \) as the number of individuals with dosage 0 at marker A and dosage 1 at marker B and so on. The explicit ML estimator for the recombination frequency \( r \) in coupling phase under both disomic and tetrasomic inheritance is invariant \( \left( \frac{n_{01}+n_{10}}{n_{00}+n_{01}+n_{10}+n_{11}} \right) \), whereas in repulsion phase the ML estimator under disomic inheritance is \( r_{diso} = \frac{n_{00}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}} \) and under tetrasomic inheritance, \( r_{tetra} = \frac{2(n_{00}+n_{11})-(n_{01}+n_{10})}{n_{00}+n_{01}+n_{10}+n_{11}} \). If the mode of inheritance is tetrasomic, \( r_{diso} \) should never fall below the value of 1/3, whereas in the case of disomic
inheritance, \( r_{\text{disom}} \in [0,0.5) \). This forms the basis of an exact Binomial test, with \( H_0: r_{\text{disom}} = 1/3 \) and \( H_1: r_{\text{disom}} < 1/3 \). Correction for multiple testing was performed using the FDR procedure with \( \alpha = 0.05 \), as described in (Benjamini and Hochberg 1995).

**Simulation of mapping under different rates of quadrivalent formation and preferential pairing**

One of the hypotheses we wanted to test was whether bivalent formation predominates in tetraploid potato as is commonly assumed. We also wanted to see the effect that deviations from this assumption could have on recombination frequency estimates that are based on a bivalent model. In this study we limited our focus to 1:1 segregating markers. We used PedigreeSim to simulate new mapping populations of 250 individuals with the fraction quadrivalents varying from zero to one in increments of 0.1. For each setting, one thousand simulated populations were generated. The simulated genome had a single chromosome of 100 cM with 51 simplex x nulliplex markers randomly distributed at positions no closer than 0.1 cM apart and the centromere at 25 cM. The true and estimated recombination frequencies between the first marker and the other 50 markers on the chromosome were recorded, as well as the LOD and assigned phase (“coupling” or “repulsion”). Recombination frequencies between marker pairs were estimated using ML, for which explicit estimators can be derived in the case of simplex marker pairs (c.f. previous section). Phase was determined by choosing the lowest estimate for the recombination frequency in the range \([0,0.5)\) which we term phasing by the minimum recombination frequency (MINR). This differs from previous studies, where the maximum of the log-likelihood (MLL) was used to assign the most likely phase (Luo et al. 2001; Hackett et al. 2013). Negative estimates for \( r \) can occur due to Mendelian sampling variation under weak repulsion linkage. For strongly-negative values \( (r < -0.05) \), a recombination frequency of 0.499, LOD of zero and phase “Unknown” were assigned and in the case \(-0.05 \leq r < 0\), the recombination frequency was set to zero and the LOD and phase were left unchanged. The recombination frequency estimates were regressed on their true values for both coupling and repulsion phase in order to evaluate how close to the true value the estimates fell for each pairing scenario. The proportion of correctly-assigned phases for coupling and repulsion phase markers was also recorded.
RESULTS

Genotyping and dosage assignment
Of the 17,987 SNPs assayed, only 40% were found to be acceptable and segregating in this population (Table 1). Acceptable markers were those for which dosages could be assigned by fitTetra, for which parental dosages were scored consistently between replicates and for which parental dosages and offspring segregation patterns were consistent. Approximately 85% of the markers could be assigned dosages by fitTetra, after which a further 5% were rejected for having inconsistent parental – offspring dosages, or for being too highly-skewed ($X^2$ test with $p < 0.001$). 1:1 segregating markers formed the largest group among the 7214 segregating markers in our population (Table 2), accounting for over 45% of useable markers.

Mapping of the 1:1 segregating markers
Almost no simplex x nulliplex markers dropped out during the mapping stage. Of the 1549 simplex x nulliplex markers in P1, 1544 were mapped (Table 3), and 1729 out of the 1733 P2 markers were mapped. The unmapped markers were lost due to poor linkage (either no chromosome assignment or extremely weak linkage within a linkage group) or large numbers of missing values.

Marker coverage over all chromosomes was well spaced with on average over 270 markers per chromosome. Only chromosomes 10 and 12 had fewer than 200 mapped markers (126 and 168 markers respectively) with chromosomes 2 and 5 having the highest marker coverage (390 and 388 markers mapped respectively). A number of homologues were split up over more than one linkage group due to insufficient linkage information. In these cases, duplex x nulliplex and simplex x simplex markers were used to provide linkage information between homologue fragments. An example of the four homologue maps of chromosome 1 in parent 2 is shown in Figure 1. In total, 30 mapped markers were found to have a discrepancy between their assignment to a linkage group in this population and their assigned chromosome on the physical sequence (FELCHER et al. 2012; VOS et al. 2015). Of these, two solcap markers (solcap_snp_c2_42265 and solcap_snp_c2_32337) were found to have positions at two physical locations but mapped to a single genetic position. A further 25 mapped markers were found to have an unknown physical position from the published datasets of marker positions (FELCHER et al. 2012; VOS et al. 2015). We provide a list of these markers with their mapped positions in Supplementary table 1. None of the 30 markers which showed linkage-group
discrepancies were included in the analysis of cM / Mb conversion rates or double reduction, but they were included on the final genetic maps due to their unambiguous genetic position.

**Position of the centromeres**

A graphical comparison of the aligned genetic and physical maps allowed an estimation of the centromeric bounds (Figure 2). When compared to previously-published centromere boundaries (SHARMA et al. 2013) the results do not correspond precisely for chromosomes 4, 5, 7, 9, 10, 11 and 12. It is possible that the discrepancies are due to the fact that our estimates are based on a tetraploid population rather than a diploid one (FELCHER et al. 2012; SHARMA et al. 2013) since the method used to determine the boundaries was essentially the same. Supplementary table 2 provides our estimates for the centromere bounds, used in the calculation of relative distance from the centromere for the double reduction analysis.

**Conversion rate between genetic and physical distance**

The cM / Mb conversion rate was determined per homologue arm across all chromosomes in both parents by linear regression of genetic distance on the physical distance (Figure 3). Apart from one clearly outlying value (due to insufficient marker coverage) the recombination rate was found to be relatively constant across all chromosomes, with an average value of 3.07 ± 0.09 (standard error of the mean).

**Double reduction**

Double reduction events were identified on all twelve chromosomes, suggesting that multivalent pairing structures can form among all potato chromosomes. Of the 235 individuals in the mapping population, 112 (47.7%) showed evidence of double reduction coming from P1 meioses and 89 (37.9%) showed double reduction segments from P2. Forty-six individuals showed evidence of having inherited a double reduction segment from both parents (but not necessarily from the same chromosome), which corresponds well with the 42.5 individuals expected under independence of parental meioses. The distribution of duplex string lengths shows that singleton duplex scores predominate in this dataset (Supplementary Figure 1). Here we have chosen to consider singleton duplex scores as unsupported evidence for double reduction which cannot be distinguished from errors in dosage estimation. We also use an algorithm which allows for possible missing scores within a string of duplex values. Using this approach, we were able to reveal the relationship between double reduction and the average distance from the centromere (Figure 4) by pooling the estimates from all 96 homologue maps, giving the average rate of double reduction as a function of distance from
the centromere. The rate of double reduction close to the centromeres approaches zero while towards the telomeres it increases substantially. Within the centromeres themselves there were twenty-two P1 markers and five P2 markers with duplex scores in the offspring. Of these, 18 were single occurrences which were probable errors (for example, the centromeric marker “PotVar0014900” which mapped to chromosome 1, homologue 4 in P1 gave five separate duplex scores. This marker was also found to have 16.2% missing values, suggesting a lower reliability. Other isolated cases would require a double recombination at both sides of the markers, which is highly unlikely to have occurred). There remained five cases of longer strings of duplex scores which partially entered the boundaries of the centromeric regions (Supplementary table 3), suggesting that in a very limited number of cases, recombination may occur within what is considered to be a non-recombining region.

PedigreeSim has previously been used to determine the rate of double reduction in simulated populations and to visualize the relationship between (genetic) distance from the centromere and double reduction (Voorrips and Maliepaard 2012). In this study we simulated phased marker data and a mapping population size of 235 in order to empirically fit a pairing model to the observed data. The observed rates of double reduction and those predicted by simulation overlap well when the fraction of quadrivalents was simulated in the range 0.2 – 0.3. Towards the telomeres the average rate of double reduction exceeded the expected rates (within a 95% confidence interval), although the confidence intervals were found to widen greatly in these regions. This may be due to the limited number of markers at these distances from the centromeres, causing greater uncertainty in the estimates.

Evidence for preferential pairing

Using the repulsion-phase marker data, we investigated whether there was any evidence for preferential pairing in this population. We found almost no evidence for preferential pairing (correcting for multiple testing using the FDR correction). On chromosomes 5 and 8 in P1 there were four marker pairs (out of 18336 and 17205 pairs, respectively) which did show possible evidence of disomic pairing, but this was not considered strong enough evidence to support a hypothesis of preferential pairing. In P2, no markers displayed disomic-like behavior. It was therefore concluded that potato follows tetrasomic inheritance as is generally assumed.
Effect of quadrivalents on mapping of simplex markers

Our analysis of double reduction suggests that quadrivalents may account for between 20% and 30% of all meiotic pairing configurations in this population. Given that previous mapping studies in potato have assumed that the rate of quadrivalent formation is negligible, we wanted to examine what effect quadrivalents have on recombination frequency estimates (and hence on linkage mapping). We compared pairwise ML estimators for \( r \) to their true underlying value (Figure 5) for different rates of quadrivalents. Overall, the effect of quadrivalents on coupling-phase estimates for simplex marker pairs was relatively minor, as shown by the gradual decrease in the slope of the regression between the true and estimated values (Figure 6.b). Correct phasing in the coupling phase was also unaffected by quadrivalents (Figure 6.a). For a quadrivalent rate between 0.2 and 0.3, the effect on coupling-phase estimates can likely be ignored. For repulsion-phase marker pairs, a greater effect was found although remarkably, the assignment of marker phasing actually improves slightly with higher numbers of quadrivalents (Figure 6.a). Of the 2374 incorrect repulsion phase assignments in the purely bivalent situation, only 14 had an associated LOD score greater than one. This suggests that as a precaution against incorrect phase assignment within a linkage group, an “unknown” phase be assigned in cases where the LOD falls below a certain threshold (e.g. LOD of one).

Effect of preferential pairing on mapping of simplex markers

Our study on the effect of preferential pairing on estimates of \( r \) revealed that preferential pairing has no effect on these estimates in coupling phase but has a dramatic impact in repulsion phase (Figure 7.b). This fact has already been reported (Qu and Hancock 2001; Koning-Boucoiran et al. 2012) and forms the basis for a test of preferential pairing which we also exploit in this study. It is evident that preferential pairing can have a severe impact on the correct assignment of repulsion phase (Figure 7.a), regardless of whether MINR or MLL is used for phase-assignment (data not shown). Since we found no evidence to suggest that any systematic preferential pairing occurred we can be fairly confident that the estimates for recombination frequency and phase were accurately performed, as confirmed by the simulation study.
DISCUSSION

Linkage maps
A recent publication describes the methods used to produce a high-density SNP linkage map of a well-studied tetraploid mapping population (HACKETT et al. 2013) using the Infinium 8300 SolCap array (FELCHER et al. 2012). Although we have not attempted to include all marker types in the current linkage maps, we have mapped a large number of markers (3273) in a tetraploid population which to the best of our knowledge is the highest-yet reported marker density of a tetraploid potato map. This has given us adequate coverage to recover all homologous chromosomes and develop an accurate picture of the double reduction landscape in this tetraploid species. We have presented separate homologue maps rather than a single consensus integrated map per chromosome as achieved by (HACKETT et al. 2013). Separate homologue maps give one the ability to infer the phasing of markers directly from the map (long-range haplotyping) without recourse to hidden Markov models (HACKETT et al. 2013), although ultimately integrated maps and genotype probabilities estimated using the integrated map will lead to greater power in subsequent QTL studies. Our finding that the large-scale conversion rate between genetic and physical distance is essentially constant outside the centromeric regions (genome-wide recombination rate) has shown that the prospects for integrating maps across homologues and between parents are good and should not impose undue stress on the underlying homologue maps. We also found little evidence of recombination “hot-spots” or “cold-spots” outside the centromeres, as evidenced by the high R²-values associated with our genetic-physical distance regressions (Supplementary table 4).

Potato cytology
Information on the pairing behavior of polyploids has traditionally been generated from cytological studies. One of the more influential publications on potato cytology has been the 1953 review of Swaminathan and Howard who summarized the findings of previous researchers such as Cadman, Lamm and Bains for the mean number of multivalents per cell at diakinesis and first metaphase in tetraploid *S. tuberosum*, ranging from 1.70 to 5.24 (SWAMINATHAN AND HOWARD 1953). This cytological evidence has been used to support the use of a simplified pairing model in potato mapping and QTL analysis since then (HACKETT et al. 2001; LUO et al. 2001; HACKETT et al. 2003; BRADSHAW et al. 2008; HACKETT et al. 2013). In our study we have used marker data to estimate the rate of double reduction and from this to extrapolate the likely frequency of multivalents (we only consider quadrivalents)
involved. A fraction of 20-30% quadrivalents translates to between 2.4 and 3.6 quadrivalents per cell, consistent with the original cytological findings of Lamm performed on the cultivar ‘Deodara’ and the line ‘36/209’ from the cross Greta x Fürstenkrone in 1945 (Lamm 1945).

**General polyploid model**

Attempts have previously been made to develop a general theory of linkage mapping in tetraploids which simultaneously considers the possibility of preferential pairing and multivalent formation (Wu et al. 2004). According to the authors, if the preferential pairing factor is set to 0 (for the case of random pairing) their model implies that the fraction of quadrivalents will equal 2/3 and that of bivalents 1/3. This is consistent with the random-end pairing model that assumes pairing initiation occurs at one set of telomeres, with probability of 1/3 that the pairing at the other telomeres will result in a separation into bivalents (John and Henderson 1962). Our data shows that preferential pairing does not occur in potato yet we have not found a fraction of quadrivalents as high as 2/3. Our findings on quadrivalent pairing are in line with a previous review of autopolyploid meiosis which found a mean multivalent frequency (trivalents and quadrivalents) of 28.8% over 93 different studies (Ramsey and Schemske 2002). It has also been shown that low numbers of multivalents does not necessarily suggest that preferential pairing behavior occurs (Sybenga 1992; Sybenga 1994).

**Identification of double reduction**

We decided to take a more stringent approach than studies which consider two or even a single locus as sufficient evidence for double reduction (Luo et al. 2006; Hackett et al. 2013). This is likely to have led to an under-estimation of DR on our part. However, all quantification of DR using marker data are likely to under-estimate the true rate of double reduction to some extent. For instance, DR segments can be hidden (no segregating allele carried on the segment), or due to limited numbers of markers one might only recover part of a double reduction segment. Higher-density linkage maps (where all homologue parts are covered by segregating markers) will lead to more accurate estimates of the rate of double reduction, unless a strong bias exists in how markers are distributed or where DR occurs. In this study, with over 3000 well-distributed simplex x nulliplex markers, we feel we have sufficient marker coverage for a detailed understanding of the double-reduction landscape. Simplex x nulliplex markers give the most unambiguous information about the presence of double reduction when compared with other marker segregation types. Other marker classes
could have been used as well (for example simplex x simplex markers, which are expected to show triplex scores in 50% of the cases of double reduction involving one of the simplex alleles). However, no marker class other than simplex x nulliplex allow DR scores to be distinguished directly as a double-reduction product. Maximum likelihood approaches that estimate the rate of double reduction such as that described in (Luo et al. 2006) may be useful for the identification of double reduction in cases where it is not clear, although we feel that flanking simplex x nulliplex marker information which supports the duplex score should be used as we have done here.

**Double reduction increases towards the telomeres**

It has been widely reported that the rate of double reduction is expected to increase towards the telomeres (Mathers 1936; Fisher 1947; Butruille and Boiteux 2000; Stift et al. 2008; Nemorin et al. 2012; Zielinski and Scheid 2012), given that the probability of a cross-over occurring between the centromere and a locus should increase as that locus is situated further from the centromere. Nevertheless, it has only rarely been experimentally verified. The clearest evidence we found in the literature came from an analysis of tetraploid potato using isozyme markers, although the numbers of markers used were rather few, with less than 50 loci considered (Haynes and Douches 1993). In our study we have clearly shown, using high-density marker data of over 3000 markers, that the rate of double reduction steadily increases with distance from the centromere. We have furthermore been able to visualize this phenomenon which has not previously been reported.

The fact that the frequency of double reduction increases towards the telomeres is perhaps cause for some concern as this could be considered a systematic source of error in the marker data. Nevertheless, with dense marker data it is now possible to accurately estimate the rate of double reduction in a mapping population. In cases where the rate of double reduction is low and marker number high, it is questionable whether highly complicated models with many parameters to be estimated are actually useful, particularly if they do not distinguish between singleton double reduction scores and genotyping errors. Our simulations have shown that even with fully quadrivalent pairing, pairwise estimators for recombination frequency between coupling-phase simplex x nulliplex markers under a bivalent pairing model are close to being exact (and as these are the most informative pairing scenario, they are the most important estimates for linkage map construction). We look forward to comparing our estimates for double reduction in tetraploid potato with other polyploid species and in gaining
a deeper understanding of why these rates differ in what are otherwise classified collectively as autopolyploids.

**Double reduction in mapping**

Some authors claim that double reduction should be included in map estimation and QTL analysis to increase the power and accuracy of the analysis (Lí et al. 2011). Our findings show that quadrivalents have little effect on the mapping of simplex markers in the highly-informative coupling-phase. In potato at least, our data shows that the level of quadrivalent formation (and preferential pairing) is very low and is therefore not likely to be of serious worry for linkage mapping. However, confirmation of this finding for other marker types is still needed.

It is also worth pointing out that quadrivalent formation not only leads to double reduction but can also result in the formation of homologue combinations of more than two parental homologues (Sved 1964) which can result from pairing-partner switches (Jones and Vincent 1994) along the chromosome. The fact that this is already part of the simulation process in PedigreeSim (Voorrips and Maliepaard 2012) increases the accuracy of our approach, not only in terms of modelling double reduction but also in our study of the effect of quadrivalents on map estimation.

**Double reduction in breeding**

Double reduction has many implications for polyploid breeding. One consequence that has been described is its potential to lead to a higher inbreeding coefficient in dihaploids derived from tetraploid lines (Haynes and Douches 1993). Given the efforts currently underway towards hybrid potato breeding (Lindhout et al. 2011), DR may have unwanted impacts on genetic diversity at the diploid level if future diploid founder material is derived from tetraploid lines. On the other hand, hybrid breeding is dependent on the production of highly-homozygous inbred lines. Tetraploid potato breeding might welcome greater levels of homozygosity in a crop that is often complicated by high heterozygosity (Uitdewilligen et al. 2013), as well as the potential purging effect that DR can have by exposing deleterious alleles to selection (Butruille and Boiteux 2000). DR could also speed up the accumulation of rare but favorable alleles through marker-assisted selection. Here we have developed the tools for the identification of DR in a segregating population which could be applied by breeders in the selection of founder parents for subsequent crossings or for confirmation studies of QTL positions.
CONCLUSIONS

In this study we constructed 96 separate homologue linkage maps of tetraploid potato using 1:1 segregating simplex markers. We estimated the approximate rate of double reduction (1.5% or more at the distal regions) and predicted by simulation that a fraction quadrivalents of 20 – 30% is required to account for this level of double reduction. We found no evidence of preferential pairing in our data, consistent with previous reports on the mode of inheritance in potato. Simulation studies using simplex x nulliplex markers revealed that marker phasing and recombination frequency estimation under a simplifying bivalent-pairing model are relatively robust, even when some level of multivalent pairing occurs.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS’ CONTRIBUTIONS

PB finalised the linkage maps, performed the analysis of double reduction and simulation studies and drafted the manuscript. CM conceived of the study, performed the linkage mapping and assignment of homologous chromosome groups, and helped draft the manuscript. REV conceived of the study, performed marker data analysis in fitTetra and helped draft the manuscript. RGFV participated in coordination and helped draft the manuscript. All authors read and approved the final manuscript.

AUTHORS’ INFORMATION

*Wageningen UR Plant Breeding, Wageningen University and Research Centre, P.O. Box 386, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

ACKNOWLEDGEMENTS

The authors would like to acknowledge Peter Vos for assistance with marker genotype calling and HZPC and Averis for providing potato varieties. The authors would also like to thank the anonymous reviewers for their comments which helped improve the manuscript. Funding for this research was provided through the TKI polyploids project “A genetic analysis pipeline for polyploid crops”, project number BO-26.03-002-001.
REFERENCES


Bradshaw, J. E., B. Pande, G. J. Bryan, C. A. Hackett, K. McLean et al., 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.


Swaminathan, M. S., and H. Howard, 1953 Cytology and genetics of the potato (Solanum tuberosum) and related species. Bibliographia Genetica 16: 1–192.


Voorrips, R., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93: 77-78.


ILLUSTRATIONS AND FIGURES

Figure 1. Homologue linkage maps of potato chromosome 1 for parent 2

Chromosome 1

<table>
<thead>
<tr>
<th>h1</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Homologue linkage maps of potato chromosome 1 for parent 2" /></td>
<td><img src="image.png" alt="Homologue linkage maps of potato chromosome 1 for parent 2" /></td>
<td><img src="image.png" alt="Homologue linkage maps of potato chromosome 1 for parent 2" /></td>
<td><img src="image.png" alt="Homologue linkage maps of potato chromosome 1 for parent 2" /></td>
</tr>
</tbody>
</table>
Figure 2. Comparison of genetic to physical distance with homologue maps of potato chromosome 1

Approximate centromere bounds are shown as dashed lines, corresponding to the inflection points in the curve (averaged over P1 and P2). Homologue maps were aligned prior to graphing by re-defining the 0 cM positions if necessary.
Figure 3. Average recombination rate across homologous chromosome arms

Rates calculated per homologue arm (north or south of the centromere) by linear regression of marker positions on the physical versus genetic distance plots. Points are colored by chromosome, with upward-pointing triangles denoting north (p) arms and downward denoting south (q) arms. P1 data is shown by filled triangles, P2 data by empty triangles.
Figure 4. Average rate of double reduction versus distance from the centromeres
Shaded areas represent 95% confidence regions around the simulated mean rate of DR arising from a fraction quadrivalents of 0.2 and 0.3. The standard deviation of the simulated mean rate of DR increases towards the telomeres, coinciding with greater fluctuations in the true rate of DR in these regions.
Figure 5. True versus estimated $r$ (using maximum likelihood) for coupling-phase simplex markers with fraction quadrivalents 0.2
The green line ($y = x$) shows the line of perfect correspondence between true and estimated values.
Figure 6 a. Proportion of incorrectly-phased markers pairs under different levels of quadrivalent formation; b. Effect of quadrivalents on accuracy of \( r \) ML estimates for coupling and repulsion-phase simplex marker pairs
Figure 7  a. Proportion of incorrectly-phased marker pairs with different levels of preferential pairing; b. Effect of preferential pairing on accuracy of $r$ ML estimates for coupling and repulsion-phase simplex marker pairs.
TABLES AND CAPTIONS

Table 1. Breakdown of SNP marker numbers after quality filtering

<table>
<thead>
<tr>
<th>Steps in SNP filtering</th>
<th>#SNPs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SolSTW Infinium array total # SNPs</td>
<td>17987</td>
<td>100.0</td>
</tr>
<tr>
<td>Dosages assigned by fitTetra a</td>
<td>15266</td>
<td>84.9</td>
</tr>
<tr>
<td>Both parents ok</td>
<td>15137</td>
<td>84.2</td>
</tr>
<tr>
<td>F1 pattern acceptable b</td>
<td>13767</td>
<td>76.4</td>
</tr>
<tr>
<td>F1 monomorphic</td>
<td>6553</td>
<td>36.4</td>
</tr>
<tr>
<td>F1 polymorphic</td>
<td>7214</td>
<td>40.0</td>
</tr>
</tbody>
</table>

a Markers not scored were either monomorphic or not clearly resolved
b Criteria for lack of F1 fit: presence of null alleles, > 3% invalid scores, highly-skewed segregation ($p > 0.001$)

Table 2. Tetraploid marker segregation types by number

<table>
<thead>
<tr>
<th>Parental dosage</th>
<th>Segregation</th>
<th>#SNP a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplex x nulliplex</td>
<td>1:1</td>
<td>1549</td>
</tr>
<tr>
<td>Nulliplex x simplex</td>
<td>1:1</td>
<td>1733</td>
</tr>
<tr>
<td>Duplex x nulliplex</td>
<td>1:4:1</td>
<td>466</td>
</tr>
<tr>
<td>Nulliplex x duplex</td>
<td>1:4:1</td>
<td>421</td>
</tr>
<tr>
<td>Simplex x simplex</td>
<td>1:2:1</td>
<td>949</td>
</tr>
<tr>
<td>Simplex x triplex</td>
<td>1:2:1</td>
<td>441</td>
</tr>
<tr>
<td>Duplex x simplex</td>
<td>1:5:5:1</td>
<td>714</td>
</tr>
<tr>
<td>Simplex x duplex</td>
<td>1:5:5:1</td>
<td>640</td>
</tr>
<tr>
<td>Duplex x duplex</td>
<td>1:8:18:8:1</td>
<td>303</td>
</tr>
</tbody>
</table>

7214

a Number of SNP markers after simplifying marker conversions have been performed.
Table 3. Composition of parental homologue maps

<table>
<thead>
<tr>
<th>Chr.</th>
<th>Parent 1</th>
<th>Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>h1 (^a)</td>
<td>h2</td>
</tr>
<tr>
<td>1</td>
<td>98.4 (44)</td>
<td>60.8 (34)</td>
</tr>
<tr>
<td>2</td>
<td>71.5 (44)</td>
<td>76.7 (34)</td>
</tr>
<tr>
<td>3</td>
<td>91.5 (17)</td>
<td>59.0 (23)</td>
</tr>
<tr>
<td>4</td>
<td>20.3 (5)</td>
<td>95.1 (98)</td>
</tr>
<tr>
<td>5</td>
<td>75.1 (32)</td>
<td>74.3 (101)</td>
</tr>
<tr>
<td>6</td>
<td>67.7 (12)</td>
<td>76.6 (35)</td>
</tr>
<tr>
<td>7</td>
<td>72.4 (21)</td>
<td>60.0 (36)</td>
</tr>
<tr>
<td>8</td>
<td>61.2 (40)</td>
<td>58.1 (99)</td>
</tr>
<tr>
<td>9</td>
<td>97.0 (8)</td>
<td>78.8 (62)</td>
</tr>
<tr>
<td>10</td>
<td>66.4 (22)</td>
<td>64.8 (18)</td>
</tr>
<tr>
<td>11</td>
<td>59.7 (34)</td>
<td>50.3 (37)</td>
</tr>
<tr>
<td>12</td>
<td>33.9 (15)</td>
<td>77.6 (27)</td>
</tr>
</tbody>
</table>

**PARENT 1**

<table>
<thead>
<tr>
<th>Chr.</th>
<th>h1 (^a)</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
<th>Total (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72.1 (25)</td>
<td>48.7 (61)</td>
<td>87.5 (60)</td>
<td>94.5 (37)</td>
<td>183</td>
</tr>
<tr>
<td>2</td>
<td>76.6 (74)</td>
<td>71.7 (17)</td>
<td>76.7 (55)</td>
<td>62.9 (89)</td>
<td>235</td>
</tr>
<tr>
<td>3</td>
<td>53.7 (110)</td>
<td>26.6 (26)</td>
<td>60.2 (47)</td>
<td>62.0 (42)</td>
<td>225</td>
</tr>
<tr>
<td>4</td>
<td>52.9 (32)</td>
<td>70.6 (37)</td>
<td>113.2 (19)</td>
<td>66.3 (46)</td>
<td>134</td>
</tr>
<tr>
<td>5</td>
<td>60.4 (69)</td>
<td>68.6 (34)</td>
<td>74.0 (78)</td>
<td>83.7 (15)</td>
<td>196</td>
</tr>
<tr>
<td>6</td>
<td>61.3 (24)</td>
<td>59.8 (51)</td>
<td>53.2 (28)</td>
<td>66.1 (50)</td>
<td>153</td>
</tr>
<tr>
<td>7</td>
<td>51.2 (12)</td>
<td>55.8 (48)</td>
<td>66.8 (28)</td>
<td>50.5 (46)</td>
<td>134</td>
</tr>
<tr>
<td>8</td>
<td>60.0 (12)</td>
<td>69.9 (37)</td>
<td>48.8 (31)</td>
<td>66.3 (24)</td>
<td>104</td>
</tr>
<tr>
<td>9</td>
<td>72.6 (15)</td>
<td>71.8 (39)</td>
<td>70.9 (12)</td>
<td>68.1 (41)</td>
<td>107</td>
</tr>
<tr>
<td>10</td>
<td>45.7 (5)</td>
<td>45.3 (4)</td>
<td>75.1 (11)</td>
<td>55.9 (23)</td>
<td>43</td>
</tr>
<tr>
<td>11</td>
<td>44.5 (23)</td>
<td>59.0 (34)</td>
<td>77.5 (21)</td>
<td>53.4 (51)</td>
<td>129</td>
</tr>
<tr>
<td>12</td>
<td>54.1 (8)</td>
<td>61.1 (39)</td>
<td>11.9 (19)</td>
<td>23.6 (20)</td>
<td>86</td>
</tr>
</tbody>
</table>

**PARENT 2**

\(^a\) Homologue map lengths in cM using Haldane’s mapping function, with number of mapped markers in brackets. h1 refers to homologue 1 etc.

\(^b\) Total number of mapped markers.