The Double-Reduction Landscape in Tetraploid Potato as Revealed by a High-Density Linkage Map

Peter M. Bourke, Roeland E. Voorrips, Richard G. F. Visser, and Chris Maliepaard[1]

Wageningen UR Plant Breeding, Wageningen University and Research Centre, Wageningen 6708 PB, The Netherlands

ORCID IDs: 0000-0002-0665-6508 (P.M.B.); 0000-0003-4388-9244 (R.E.V.)

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 biparental 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 has 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.

KEYWORDS linkage mapping; tetraploid; double reduction; potato; multivalents

POLYPLOID species constitute a very important group among cultivated crops. Polyploids themselves can be further divided into auto- and allopolyploids, with autopolyploids showing random association between homologous chromosomes and allopolyploids showing nonrandom 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 data sets.

One autopolyploid species in which 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 (Potato Genome Sequencing Consortium 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 with other economically important autotetraploid species such as alfalfa, rose, and 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.* 2009). 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 is the simplex × nulliplex marker type, which is expected to segregate in a 1:1 fashion. In a tetraploid, we employ the term *simplex × nulliplex* to collectively refer to 1×0, 3×0, 3×4, and 1×4 markers (with 0×1, 0×3, 4×3, and 4×1 markers being nulliplex × 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 Ooijen 2006), which can efficiently map large numbers of markers as well as providing many checks on map and data quality. Simplex × nulliplex markers also provide the clearest linkage information to cluster markers into separate homologous chromosomes, forming the basis of homolog maps. In our population, simplex × nulliplex markers were also the most abundant marker segregation type. We therefore restricted our analysis to simplex × nulliplex markers, which nevertheless allowed us to map a total of 3273 markers across both parents.

Simplex × 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. 2009). In cases where a crossover occurs between two sets of sister chromatids that 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 that can never occur in diploids. For a simplex × 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 F_1 mapping population of 237 individuals was created from the cross between two tetraploid potato varieties, cultivars *Altus* (hereafter referred to as parent one, P_1) and *Colomba* (P_2).

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 (Vos et al. 2015). For DNA concentrations in the range of 25–50 ng/μl, the sample also was used; samples having concentrations lower than 25 ng/μl 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. A total of 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 nonpenalized SD 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 relabeled as simplex × nulliplex (or nulliplex × simplex) for mapping and DR analysis. Considering markers whose segregating allele is inherited from P_1, these consisted of triplex × nulliplex, triplex × quadruplex, and simplex × quadruplex markers. For example, a triplex × nulliplex marker is expected to produce 50% dosage 1 and 50% dosage 2 among the offspring, with observable DR scores appearing as dosage 0 (a double copy of the 0 allele from P_1). Relabeling 2 as 0 and 0 as 2 (with the parents relabeled as simplex and nulliplex) achieves the desired result of marker conversion.

Linkage-map construction

Simplex × nulliplex marker data were recoded to JoinMap 4.1 cross-pollinator format ($lm \times ll$). "Impossible" genotypes (invalid scores) were removed before importation into

Table 1 Breakdown of SNP marker numbers after quality filtering

Steps in SNP filtering	No. of SNPs	Percent
SolSTW Infinium array total number of SNPs	17,987	100.0
Dosages assigned by fitTetra[a]	15,266	84.9
Both parents assigned	15,137	84.2
F_1 pattern acceptable[b]	13,767	76.4
F_1 monomorphic	6,553	36.4
F_1 polymorphic	7,214	40.0

[a] Markers not scored were either monomorphic or not clearly resolved.
[b] Criteria for lack of F_1 fit: presence of null alleles, >3% invalid scores, highly skewed segregation ($P < 0.001$).

Table 2 Tetraploid marker segregation types by number

Parental dosage	Segregation	No. of SNP[a]
Simplex × nulliplex	1:1	1549
Nulliplex × simplex	1:1	1733
Duplex × nulliplex	1:4:1	466
Nulliplex × duplex	1:4:1	421
Simplex × simplex	1:2:1	949
Simplex × triplex	1:2:1	441
Duplex × simplex	1:5:5:1	714
Simplex × duplex	1:5:5:1	640
Duplex × duplex	1:8:18:8:1	303
Total	—	7214

[a] Number of SNP markers after simplifying marker conversions have been performed.

JoinMap. One pair of identical individuals was identified in the data set (similarity of 0.9922); therefore, we removed individual 202. Markers were assigned to linkage groups with a minimum LOD score of 4 (a higher LOD score was used if clusters broke into large subclusters at a higher LOD score). Marker clusters were assigned to physical chromosomes based on the position of markers on the physical sequence (Potato Genome Sequencing Consortium 2011). Mapping was first performed using the groupings from the groupings tree using maximum likelihood (ML). Homologs then were identified by large gaps in the estimated map distances (≥60 cM), which also were often accompanied by a transition in estimated marker phase. Marker data for separate homologs were exported from JoinMap in .loc files and reimported for creation of the homolog maps. After an initial mapping of the homologs, individual 067 was found to contribute unrealistic numbers of recombinations in many linkage groups across both parents and therefore was 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 been used previously 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 × nulliplex and simplex × simplex markers to connect subhomolog linkage groups that had poor internal linkage among simplex × nulliplex markers. Map data were 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 Supporting Information, Table S1. 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 redefined by examining 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 (Figure S2). 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 the physical positions per homolog arm. The slopes of the regression lines for each homolog 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 homolog) 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 homolog 2 of chromosome 1 in parent 1. This genome-wide recombination rate was used to convert the physical map to a pseudointegrated genetic map for use in the simulation studies.

Rates of DR

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

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

Table 3 Composition of parental homolog maps

Chromosome	h1[a]	h2	h3	h4	Total[b]
Parent 1					
1	98.4 (44)	60.8 (34)	67.3 (26)	89.9 (54)	158
2	71.5 (44)	76.7 (34)	56.0 (31)	46.0 (46)	155
3	91.5 (17)	59.0 (23)	56.4 (12)	86.0 (57)	109
4	20.3 (5)	95.1 (98)	91.9 (21)	69.4 (20)	144
5	75.1 (32)	74.3 (101)	114.7 (9)	66.3 (50)	192
6	67.7 (12)	76.6 (35)	75.4 (15)	69.9 (14)	76
7	72.4 (21)	60.0 (36)	57.3 (13)	55.4 (35)	105
8	61.2 (40)	58.1 (99)	58.6 (20)	56.1 (27)	186
9	97.0 (8)	78.8 (62)	86.9 (24)	101.8 (23)	117
10	66.4 (22)	64.8 (18)	58.2 (13)	64.0 (30)	83
11	59.7 (34)	50.3 (37)	61.1 (22)	56.4 (44)	137
12	33.9 (15)	77.6 (27)	73.1 (20)	52.2 (20)	82
Total	—	—	—	—	1544
Parent 2					
1	72.1 (25)	48.7 (61)	87.5 (60)	94.5 (37)	183
2	76.6 (74)	71.7 (17)	76.7 (55)	62.9 (89)	235
3	53.7 (110)	26.6 (26)	60.2 (47)	62.0 (42)	225
4	52.9 (32)	70.6 (37)	113.2 (19)	66.3 (46)	134
5	60.4 (69)	68.6 (34)	74.0 (78)	83.7 (15)	196
6	61.3 (24)	59.8 (51)	53.2 (28)	66.1 (50)	153
7	51.2 (12)	55.8 (48)	66.8 (28)	50.5 (46)	134
8	60.0 (12)	69.9 (37)	48.8 (31)	66.3 (24)	104
9	72.6 (15)	71.8 (39)	70.9 (12)	68.1 (41)	107
10	45.7 (5)	45.3 (4)	75.1 (11)	55.9 (23)	43
11	44.5 (23)	59.0 (34)	77.5 (21)	53.4 (51)	129
12	54.1 (8)	61.1 (39)	11.9 (19)	23.6 (20)	86
Total	—	—	—	—	1729

h1, homolog 1; h2, homolog 2; etc.
[a] Homolog map lengths in centimorgans using Haldane's mapping function, with number of mapped markers in brackets.
[b] Total number of mapped markers.

per homologue for each 1-Mb window distance from the centromeres across all chromosomes and both parents. The average rate per chromosome was estimated by multiplying the homologue rates by a factor of four.

Simulation of DR and prediction of quadrivalent formation

An approximate "integrated" genetic linkage map was produced using the average cM/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 homolog assignment of all markers. Phased marker genotypes and a consensus genetic map position are the basic inputs 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 DR was applied to the simulated data sets, allowing us to deduce the relationship between the rate of DR 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 because 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 has 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, *i.e.*,

$$\frac{n_{01} + n_{10}}{n_{00} + n_{01} + n_{10} + n_{11}}$$

whereas in repulsion phase the ML estimator under disomic inheritance is

$$r_{\text{disom}} = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}}$$

and under tetrasomic inheritance is

$$r_{\text{tetra}} = \frac{2(n_{00} + n_{11}) - (n_{01} + n_{10})}{n_{00} + n_{01} + n_{10} + n_{11}}$$

If the mode of inheritance is tetrasomic, r_{disom} 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 0 to 1 in increments of 0.1. For each setting, 1000 simulated populations were generated. The simulated genome had a single chromosome of 100 cM with 51 simplex × 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
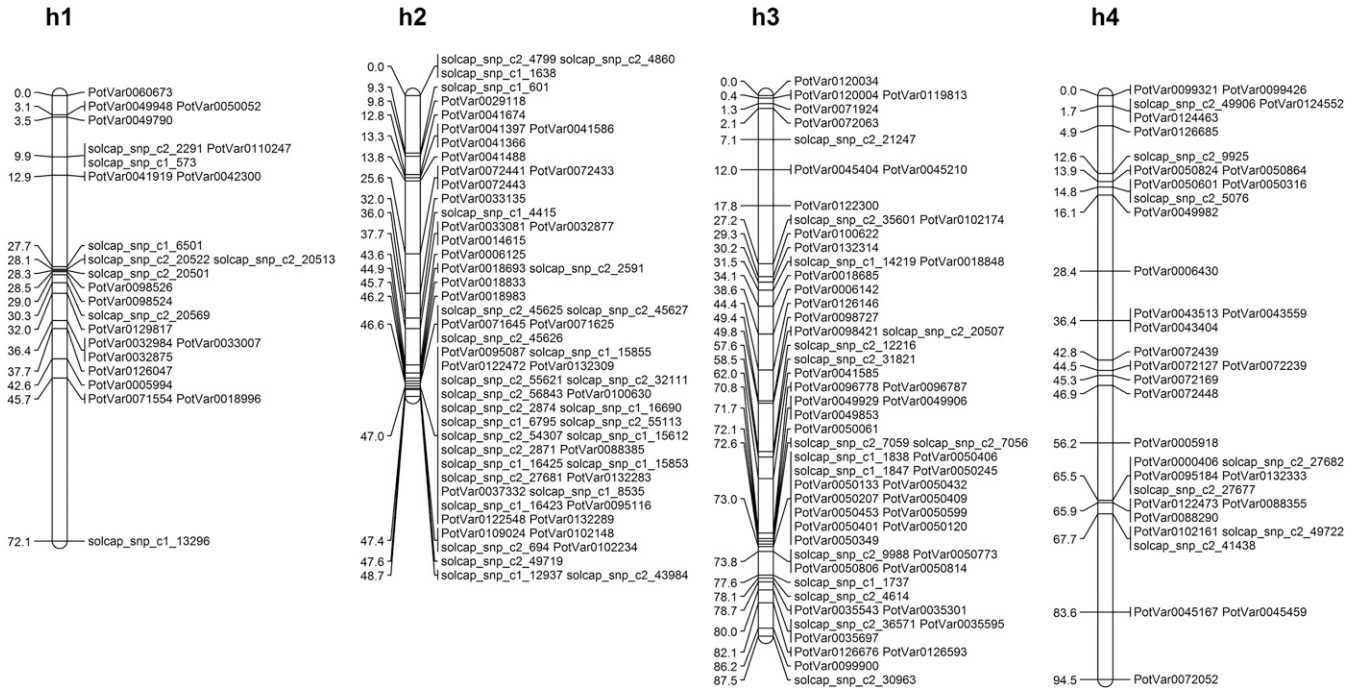
Chromosome 1

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

50 markers on the chromosome were recorded, as well as the LOD score 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 (see the preceding 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 owing to Mendelian sampling variation under weak repulsion linkage. For strongly negative values ($r < -0.05$), a recombination frequency of 0.499, a LOD score of 0, and phase "unknown" were assigned, and in the case $-0.05 \leq r < 0$, the recombination frequency was set to 0, and the LOD score and phase were left unchanged. The recombination frequency estimates were regressed on their true values for both coupling and repulsion phase 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.

Data availability

Genotype data of the 1:1 segregating markers used in this study are provided as supplementary data (Table S5). The genetic map positions of these markers are also provided (Table S6).

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 (χ^2 test with $P < 0.001$). Markers that segregated 1:1 formed the largest group among the 7214 segregating markers in our population (Table 2), accounting for over 45% of usable markers.

Mapping of the 1:1 segregating markers

Almost no simplex × nulliplex markers dropped out during the mapping stage. Of the 1549 simplex × nulliplex markers in P_1, 1544 were mapped (Table 3), and 1729 of the 1733 P_2 markers were mapped. The unmapped markers were lost owing 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
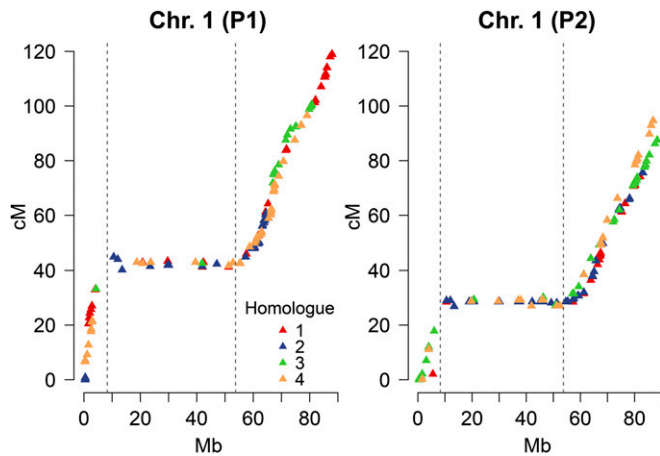
Figure 2 Comparison of genetic-to-physical distance with homolog maps of potato chromosome 1. Approximate centromere bounds are shown as dashed lines, corresponding to the inflection points in the curve (averaged over P_1 and P_2). Homolog maps were aligned prior to graphing by redefining the 0-cM positions, if necessary.

and 5 having the highest marker coverage (390 and 388 markers mapped, respectively). A number of homologs were split up over more than one linkage group as a result of insufficient linkage information. In these cases, duplex × nulliplex and simplex × simplex markers were used to provide linkage information between homolog fragments. An example of the four homolog 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 data sets of marker positions (Felcher *et al.* 2012; Vos *et al.* 2015). We provide a list of these markers with their mapped positions in Table S1. None of the 30 markers that showed linkage group discrepancies were included in the analysis of cM/Mb conversion rates or DR, but they were included on the final genetic maps because of 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) because the method used to determine the boundaries was essentially the same. Table S2 provides our estimates for the centromere bounds used in the calculation of relative distance from the centromere for the DR analysis.

Conversion rate between genetic and physical distance

The cM/Mb conversion rate was determined per homolog 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 (owing 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 (SEM).

Double reduction

DR events were identified on all 12 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 DR coming from P_1 meioses, and 89 (37.9%) showed DR segments from P_2. Forty-six individuals showed evidence of having inherited a DR 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 data set (Figure S1). Here we have chosen to consider singleton duplex scores as unsupported evidence for DR that cannot be distinguished from errors in dosage estimation. We also use an algorithm that allows for possible missing scores within a string of duplex values.

Using this approach, we were able to reveal the relationship between DR and the average distance from the centromere (Figure 4) by pooling the estimates from all 96 homolog maps, giving the average rate of DR as a function of distance from the centromere. The rate of DR close to the centromeres approaches zero, while toward the telomeres it increases substantially. Within the centromeres themselves, there were 22 P_1 markers and 5 P_2 markers with duplex scores in the offspring. Of these, 18 were single occurrences that were probable errors. (For example, the centromeric marker PotVar0014900, which mapped to chromosome 1, homolog 4 in P_1, gave five separate duplex scores. This marker also was 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 that partially entered the boundaries of the centromeric regions (Table S3), suggesting that recombination may occur within what is considered to be a nonrecombining region in a very limited number of cases.

PedigreeSim has been used previously to determine the rate of DR in simulated populations and to visualize the relationship between (genetic) distance from the centromere and DR (Voorrips and Maliepaard 2012). In this study, we simulated phased marker data and a mapping population size of 235 to empirically fit a pairing model to the observed data. The observed rates of DR and those predicted by simulation overlap well when the fraction of quadrivalents was simulated in the range 0.2–0.3. Toward the telomeres, the average rate of DR exceeded the expected rates (within a 95%
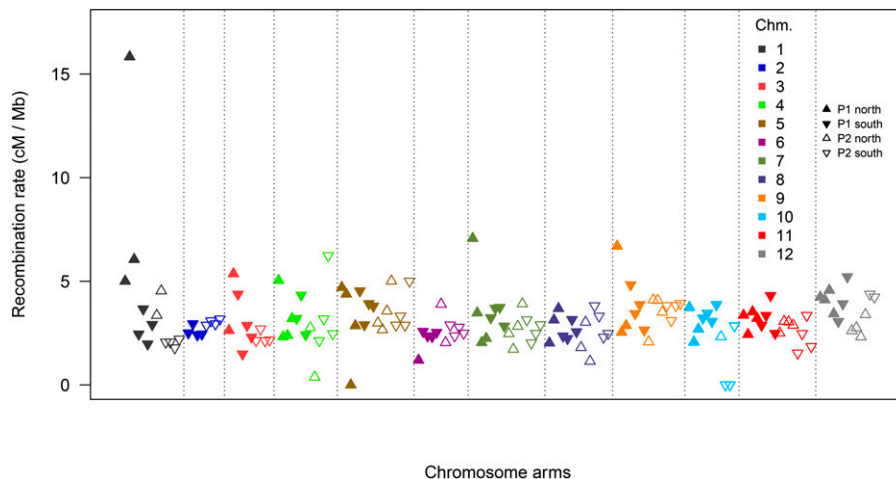
Figure 3 Average recombination rate across homologous chromosome arms. Rates calculated per homolog arm (north or south of the centromere) by linear regression of marker positions on the physical *vs.* genetic distance plots. Points are colored by chromosome, with upward-pointing triangles denoting north (*p*) arms and downward denoting south (*q*) arms. P_1 data are shown by filled triangles, P_2 data by empty triangles.

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 P_1, there were four marker pairs (of 18,336 and 17,205 pairs, respectively) that did show possible evidence of disomic pairing, but this was not considered strong enough evidence to support a hypothesis of preferential pairing. In P_2, 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 DR 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 values (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 6B). Correct phasing in the coupling phase also was unaffected by quadrivalents (Figure 6A). For a quadrivalent rate between 0.2 and 0.3, the effect on coupling-phase estimates likely can 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 6A). Of the 2374 incorrect repulsion-phase assignments in the purely bivalent situation, only 14 had an associated LOD score greater than 1. This suggests that as a precaution against incorrect

phase assignment within a linkage group, an "unknown" phase should be assigned in cases where the LOD score falls below a certain threshold (*e.g.*, a LOD score of 1).

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 7B). 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 that 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 7A) regardless of whether MINR or MLL is used for phase assignment (data not shown). Because 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 performed accurately, 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 that 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 DR landscape in this tetraploid species. We have presented separate homolog maps rather than a single consensus integrated map per chromosome, as achieved by Hackett *et al.* (2013). Separate homolog maps give one the ability to infer the phasing of markers directly
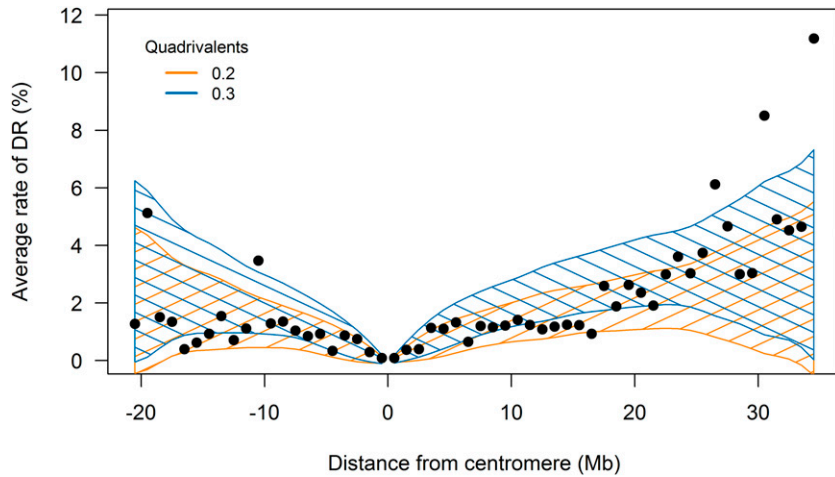
Figure 4 Average rate of DR *vs.* distance from the centromeres. Shaded areas represent 95% confidence regions around the simulated mean rate of DR arising from fraction quadrivalents 0.2 and 0.3. The SD of the simulated mean rate of DR increases toward the telomeres, coinciding with greater fluctuations in the true rate of DR in these regions.

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 homologs and between parents are good and should not impose undue stress on the underlying homolog maps. We also found little evidence of recombination hot or cold spots outside the centromeres, as evidenced by the high R^2 values associated with our genetic-physical distance regressions (Table S4).

Potato cytology

Information on the pairing behavior of polyploids traditionally has been generated from cytological studies. One of the more influential publications on potato cytology has been the early review of Swaminathan and Howard (1953), who summarized the findings of previous researchers such as Cadman (1943), Lamm (1945), and Bains (1951) for the mean number of multivalents per cell at diakinesis and first metaphase in tetraploid *S. tuberosum* as ranging from 1.70 to 5.24. 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, 2003, 2013; Luo *et al.* 2001; Bradshaw *et al.* 2008). In our study, we have used marker data to estimate the rate of DR and from this to extrapolate the likely frequency of multivalents involved (we only consider quadrivalents). A fraction of 20–30% quadrivalents translates to between 2.4 and 3.6 quadrivalents per cell, consistent with the original cytologic findings of Lamm (1945) performed on the cultivar *Deodara* and the line 36/209 from the cross Greta × Fürstenkrone.

General polyploid model

Attempts have been made previously to develop a general theory of linkage mapping in tetraploids that simultaneously considers the possibility of preferential pairing and multivalent formation (Wu *et al.* 2004). According to these 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, which assumes that 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 show 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 that found a mean multivalent frequency (trivalents and quadrivalents) of 28.8% over 93 different studies (Ramsey and Schemske 2002). It also has been shown that low numbers of multivalents do not necessarily suggest that preferential pairing behavior occurs (Sybenga 1992, 1994).

Identification of DR

We decided to take a more stringent approach than studies that consider two or even a single locus as sufficient evidence for DR (Luo *et al.* 2006; Hackett *et al.* 2013). This is likely to have led to an underestimation of DR on our part. However, all quantification of DR using marker data are likely to underestimate the true rate of DR to some extent. For instance, DR segments can be hidden (no segregating allele carried on
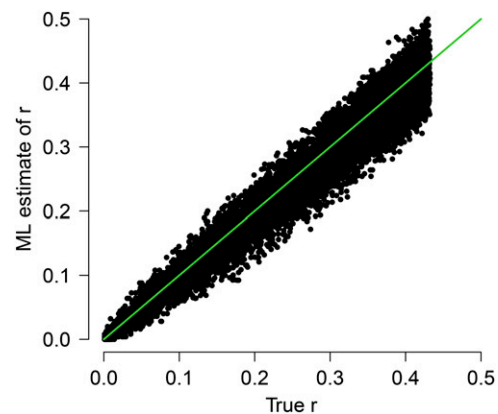

Figure 5 True *vs.* estimated *r* (using ML) for coupling-phase simplex markers with fraction quadrivalents 0.2. The green line ($y = x$) shows the line of perfect correspondence between true and estimated values.
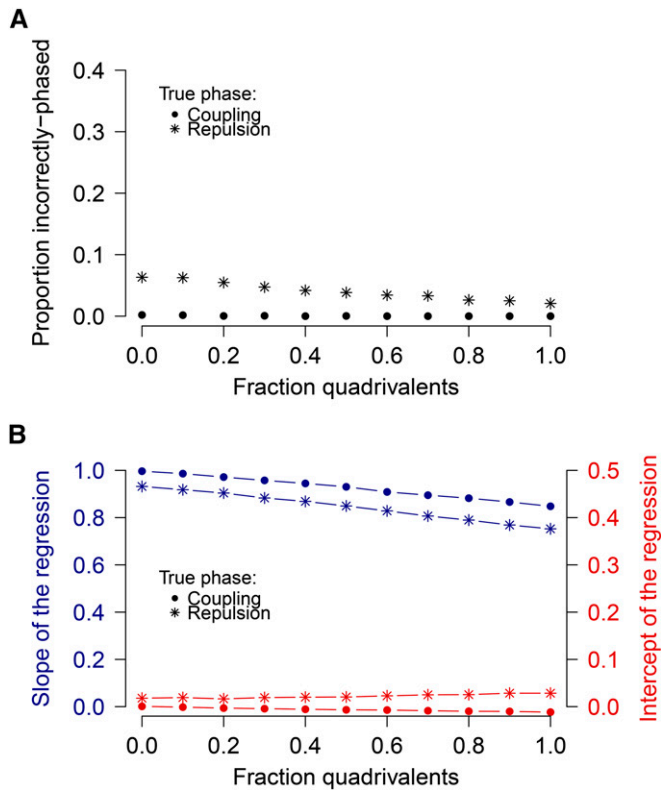
Figure 6 (A) Proportion of incorrectly phased markers pairs under different levels of quadrivalent formation. (B) Effect of quadrivalents on accuracy of *r* ML estimates for coupling- and repulsion-phase simplex marker pairs

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

the segment), or owing to limited numbers of markers, one might recover only part of a DR segment. Higher-density linkage maps (where all homolog parts are covered by segregating markers) will lead to more accurate estimates of the rate of DR unless a strong bias exists in how markers are distributed or where DR occurs. In this study, with over 3000 well-distributed simplex × nulliplex markers, we feel that we have sufficient marker coverage for a detailed understanding of the DR landscape.

Simplex × nulliplex markers give the most unambiguous information about the presence of DR when compared with other marker segregation types. Other marker classes could have been used as well (*e.g.*, simplex × simplex markers, which are expected to show triplex scores in 50% of the cases of DR involving one of the simplex alleles). However, no marker class other than simplex × nulliplex allow DR scores to be distinguished directly as a DR product. ML approaches that estimate the rate of DR, such as that described in Luo *et al.* (2006), may be useful for the identification of DR in cases where it is not clear, although we feel that flanking simplex × nulliplex marker information that supports the duplex score should be used, as we have done here.

DR increases toward the telomeres

It has been widely reported that the rate of DR is expected to increase toward the telomeres (Mather 1936; Fisher 1947; Butruille and Boiteux 2000; Stift *et al.* 2008; Nemorin *et al.*
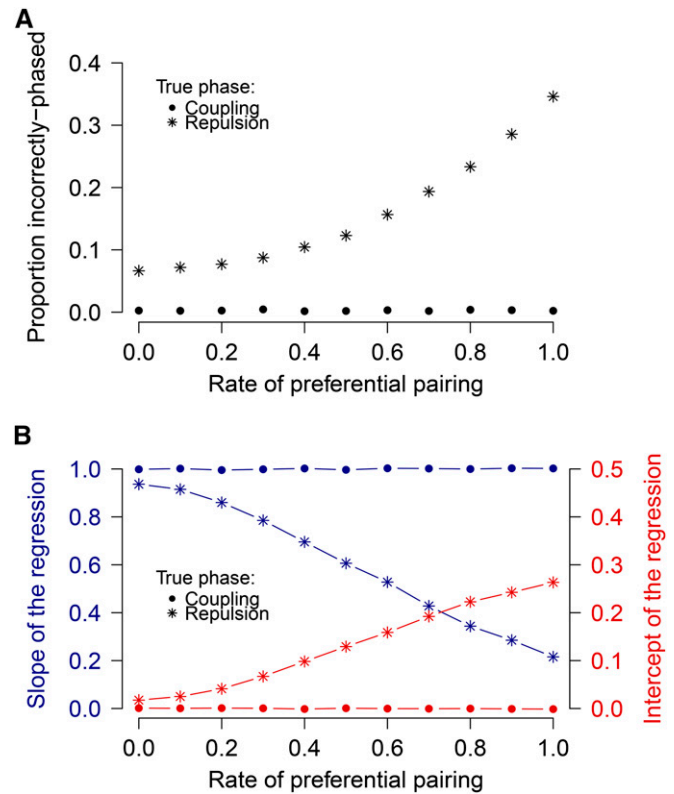
2012; Zielinski and Scheid 2012), given that the probability of a crossover occurring between the centromere and a locus should increase as that locus is situated further from the centromere. Nevertheless, this has been experimentally verified only rarely. The clearest evidence we found in the literature came from an analysis of tetraploid potato using isozyme markers, although the number of markers used was rather limited, with fewer 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 DR steadily increases with distance from the centromere. We have furthermore been able to visualize this phenomenon, which has not been reported previously.

The fact that the frequency of DR increases toward the telomeres is perhaps cause for some concern because 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 DR in a mapping population. In cases where the rate of DR 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 DR scores and genotyping errors. Our simulations have shown that even with fully quadrivalent pairing, pairwise estimators for recombination frequency between coupling-phase simplex × nulliplex markers under a bivalent pairing model are close to being exact

(and because these constitute the most informative pairing scenario, they are the most important estimates for linkage-map construction). We look forward to comparing our estimates for DR 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.

DR in mapping

Some authors claim that DR should be included in map estimation and QTL analysis to increase the power and accuracy of the analysis (Li *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 show that the level of quadrivalent formation (and preferential pairing) is very low and therefore is 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 DR but also can result in the formation of homolog combinations of more than two parental homologs (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 modeling DR but also in our study of the effect of quadrivalents on map estimation.

DR in breeding

DR 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 toward 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. However, hybrid breeding depends 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 also could 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 that 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 homolog linkage maps of tetraploid potato using 1:1 segregating simplex markers. We estimated the approximate rate of DR (6% or more at the distal regions) and predicted by simulation that a fraction of quadrivalents of 20–30% is required to account for this level of DR. 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 $\times$ 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.

Literature Cited

Andrade, J., and M. Estévez-Pérez, 2014   Statistical comparison of the slopes of two regression lines: a tutorial. Anal. Chim. Acta 838: 1–12.

Bains, G. S., 1951   Cytological studies in the genus *Solanum*, sect. *Tuberarium*. MSc. Dissertation, Univ. Cambridge.

Benjamini, Y., and Y. Hochberg, 1995   Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B 57: 289–300.

Bradshaw, J., 2007   The canon of potato science. 4. Tetrasomic inheritance. Potato Res. 50: 219–222.

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.

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.

Butruille, D., and L. Boiteux, 2000   Selection–mutation balance in polysomic tetraploids: impact of double reduction and gametophytic selection on the frequency and subchromosomal localization of deleterious mutations. Proc. Natl. Acad. Sci. USA 97: 6608–6613.

Cadman, C. H., 1943   Nature of tetraploidy in cultivated European potatoes. Nature, Lond. 152: 103–04.

Chakravarti, A., 1991 A graphical representation of genetic and physical maps: the Marey map. Genomics 11: 219–222.

Felcher, K. J., J. J. Coombs, A. N. Massa, C. N. Hansey, J. P. Hamilton *et al.*, 2012 Integration of two diploid potato linkage maps with the potato genome sequence. PLoS One 7: e36347.

Fisher, R. A., 1947 The theory of linkage in polysomic inheritance. Philos. Trans. R. Soc. Lond. B Biol. Sci. 233: 55–87.

Hackett, C., and Z. Luo, 2003 TetraploidMap: construction of a linkage map in autotetraploid species. J. Hered. 94: 358–359.

Hackett, C., J. Bradshaw, and J. McNicol, 2001 Interval mapping of quantitative trait loci in autotetraploid species. Genetics 159: 1819–1832.

Hackett, C., B. Pande, and G. Bryan, 2003 Constructing linkage maps in autotetraploid species using simulated annealing. Theor. Appl. Genet. 106: 1107–1115.

Hackett, C. A., K. McLean, and G. J. Bryan, 2013 Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PLoS One 8: e63939.

Haynes, K., and D. Douches, 1993 Estimation of the coefficient of double reduction in the cultivated tetraploid potato. Theor. Appl. Genet. 85: 857–862.

John, B., and S. Henderson, 1962 Asynapsis and polyploidy in *Schistocerca paranensis*. Chromosoma 13: 111–147.

Jones, G., and J. Vincent, 1994 Meiosis in autopolyploid *Crepis capillaris*. II. Autotetraploids. Genome 37: 497–505.

Koning-Boucoiran, C., V. Gitonga, Z. Yan, O. Dolstra, C. Van der Linden *et al.*, 2012 The mode of inheritance in tetraploid cut roses. Theor. Appl. Genet. 125: 591–607.

Lamm, R., 1945 Cytogenetic studies in *Solanum*, sect. Tuberarium. Hereditas 31: 1–129.

Li, J., K. Das, G. Fu, C. Tong, Y. Li *et al.*, 2011 EM algorithm for mapping quantitative trait loci in multivalent tetraploids. Int. J. Plant Genomics 2010: 216547.

Lindhout, P., D. Meijer, T. Schotte, R. C. Hutten, R. G. Visser *et al.*, 2011 Towards F_1 hybrid seed potato breeding. Potato Res. 54: 301–312.

Luo, Z., C. Hackett, J. Bradshaw, J. McNicol, and D. Milbourne, 2001 Construction of a genetic linkage map in tetraploid species using molecular markers. Genetics 157: 1369–1385.

Luo, Z. W., Z. Zhang, L. Leach, R. M. Zhang, J. E. Bradshaw *et al.*, 2006 Constructing genetic linkage maps under a tetrasomic model. Genetics 172: 2635–2645.

Mather, K., 1936 Segregation and linkage in autotetraploids. J. Genet. 32: 287–314.

Meyer, R., D. Milbourne, C. Hackett, J. Bradshaw, J. McNichol *et al.*, 1998 Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (Phytophthora infestans). Mol. Genet. Genomics 259: 150–160.

Milbourne, D., J. E. Bradshaw, and C. A. Hackett, 2009 Molecular mapping and breeding in polyploid crop plants, pp. 355–394 in Principles and Practices of Plant Genomics, Vol. 2: Molecular Breeding, edited by C. Kole and A. G. Abbott. CRC Press, Boca Raton, FL.

Nemorin, A., K. Abraham, J. David, and G. Arnau, 2012 Inheritance pattern of tetraploid Dioscorea alata and evidence of double reduction using microsatellite marker segregation analysis. Mol. Breed. 30: 1657–1667.

Potato Genome Sequencing Consortium (PGSC), 2011 Genome sequence and analysis of the tuber crop potato. Nature 475: 189–195.

Qu, L., and J. Hancock, 2001 Detecting and mapping repulsion-phase linkage in polyploids with polysomic inheritance. Theor. Appl. Genet. 103: 136–143.

Ramsey, J., and D. W. Schemske, 2002 Neopolyploidy in flowering plants. Annu. Rev. Ecol. Syst. 33: 589–639.

R Core Team, 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sharma, S. K., D. Bolser, J. de Boer, M. Sønderkær, W. Amoros *et al.*, 2013 Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 3: 2031–2047.

Stift, M., C. Berenos, P. Kuperus, and P. H. van Tienderen, 2008 Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. Genetics 179: 2113–2123.

Sved, J. A., 1964 The relationship between diploid and tetraploid recombination frequencies. Heredity 19: 585–596.

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

Sybenga, J., 1992 *Cytogenetics in Plant Breeding* (Monographs in Theoretical and Applied Genetics, Vol. 17). Springer, Berlin.

Sybenga, J., 1994 Preferential pairing estimates from multivalent frequencies in tetraploids. Genome 37: 1045–1055.

Uitdewilligen, J. G., A.-M. A. Wolters, B. B. D'hoop, T. J. Borm, R. G. Visser *et al.*, 2013 A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8: e62355.

Van Ooijen, J., 2006 *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Kyazma BV, Wageningen, The Netherlands.

van Os, H., S. Andrzejewski, E. Bakker, I. Barrena, G. J. Bryan *et al.*, 2006 Construction of a 10,000-marker ultradense genetic recombination map of potato: Providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173: 1075–1087.

Voorrips, R., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93: 77–78.

Voorrips, R. E., and C. A. Maliepaard, 2012 The simulation of meiosis in diploid and tetraploid organisms using various genetic models. BMC Bioinformatics 13: 248.

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.

Vos, P. G., J. G. A. M. L. Uitdewilligen, R. E. Voorrips, R. G. F. Visser, and H. J. van Eck, 2015 Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. Theor. Appl. Genet. .DOI:10. 1007/ s00122-015-2593-y

Wu, R., C.-X. Ma, and G. Casella, 2004 A mixed polyploid model for linkage analysis in outcrossing tetraploids using a pseudo-test backcross design. J. Comput. Biol. 11: 562–580.

Zielinski, M.-L., and O. M. Scheid, 2012 Meiosis in polyploid plants, pp. 33–55 in Polyploidy and *Genome Evolution*, edited by P. S. Soltis and D. E. Soltis. Springer, Berlin.

Communicating editor: A. H. Paterson

GENETICS

The Double-Reduction Landscape in Tetraploid Potato as Revealed by a High-Density Linkage Map

Peter M. Bourke, Roeland E. Voorrips, Richard G. F. Visser, and Chris Maliepaard

Table S1. Markers for which a discrepancy was found between the mapped position and physical position (or physical position unknown).

Marker	Chr.	cM	Phys. chr.[a]	Phys. bp
solcap_snp_c2_54303	1	43.02	-	-
solcap_snp_c2_54306	1	43.02	-	-
PotVar0014240	1	43.42	10	38714113
PotVar0118576	1	41.52	10	52034130
PotVar0071270	1	43.02	11	34315041
solcap_snp_c2_54307	1	28.56	-	-
PotVar0014615	1	37.86	10	28576951
PotVar0125072	2	37.44	8	45790082
solcap_snp_c2_42265 [b]	2	44.08	2	37393961
			12	45024176
solcap_snp_c2_47201	2	83.17	-	-
PotVar0085038	3	4.35	6	49818714
R2_E_LG04	4	14.43	-	-
solcap_snp_c2_36957	4	5.2	-	-
solcap_snp_c1_11008	4	4.77	-	-
solcap_snp_c2_53782	4	22.02	-	-
R2_D_LG04	4	9.91	-	-
solcap_snp_c2_50286	5	40.72	-	-
PotVar0134951	5	62.47	-	-
solcap_snp_c2_48821	5	62.49	-	-
PotVar0007814	5	76.75	12	17623063
PotVar0134955	5	39.16	-	-
solcap_snp_c2_48820	5	39.59	-	-
PotVar0024151	5	40.67	-	-
PotVar0024137	5	40.67	-	-
PotVar0047235	5	22.35	11	39417958
PotVar0084430	5	40.67	4	58263426
PotVar0084432	5	40.67	4	58263375
PotVar0032700	7	26.51	4	13034726
PotVar0032779	7	26.51	4	13034026
PotVar0037615	7	26.51	12	50380397
PotVar0032614	7	26.51	4	13565406
PotVar0032617	7	24.5	4	13035443
solcap_snp_c2_57589	8	20.19	-	-
solcap_snp_c2_37600	8	19.34	-	-
solcap_snp_c2_57588	8	19.34	-	-
solcap_snp_c2_57591	8	20.19	-	-
solcap_snp_c2_37599	8	19.34	-	-
solcap_snp_c2_53177	8	19.63	-	-
PotVar0118202	8	0.86	11	45421226
PotVar0118200	8	0.86	11	45421108
PotVar0081045	8	19.34	1	25032740

P. M. Bourke *et al.*

solcap_snp_c2_54581	8	19.34	1	24708276
PotVar0069362	8	42.84	6	30358249
solcap_snp_c2_21331	9	35.92	-	-
solcap_snp_c2_21314	9	35.92	-	-
PotVar0131182	9	35.92	-	-
solcap_snp_c2_4567	9	38.07	7	12696253
PotVar0118577	9	59.81	10	52034143
solcap_snp_c2_32337 [b]	11	37.78	1	57541294
			11	10391627
PotVar0101550	11	44.31	4	4925336
PotVar0101542	11	44.31	4	4925258
PotVar0124931	12	63.5	8	45789083
PotVar0124993	12	63.5	8	45789554
PotVar0125939	12	16.43	5	8135275
PotVar0022107	12	4.19	7	1503568

[a] Chromosome assignments based on ST4.03 version of the potato reference sequence (SHARMA *et al.* 2013).

[b] Both solcap_snp_c2_42265 and solcap_snp_c2_32337 appear at two positions on the potato physical sequence version ST4.03 (http://potato.plantbiology.msu.edu/cgi-bin/gbrowse/potato/) but mapped in this study to a single genetic position.

Table S2. Estimates for the centromere bounds as derived from comparison of genetic and physical marker positions.

Chr.	Start	End
1	8276035	53724444
2	4840214	14625580
3	11792597	36000000
4	11000000	48500000
5	18700000	41855486
6	7047440	32627834
7	10000000	26941412
8	6000000	33622322
9	12714380	42082705
10	7716726	41361076
11	17230017	31908912
12	20805053	49044694

P. M. Bourke *et al.*

Table S3. Markers for which a duplex score was found within the centromeres.

Marker	Chr.	Homologue	Phys. bp	Offspring ID	Parent
PotVar0114353	1	1	42372618	166	1
solcap_snp_c2_43984	1	2	13369757	119	2
PotVar0102234	1	2	51293649	119	2
PotVar0014900	1	4	52868606	5	1
PotVar0014900	1	4	52868606	6	1
PotVar0014900	1	4	52868606	15	1
PotVar0014900	1	4	52868606	16	1
PotVar0014900	1	4	52868606	114	1
PotVar0109627	2	1	11293249	7	1
PotVar0109627	2	1	11293249	109	1
solcap_snp_c2_730	2	1	12188989	109	1
solcap_snp_c1_3347	3	1	16008414	76	2
PotVar0085581	3	1	35357169	108	1
solcap_snp_c2_14930	4	2	19581827	179	1
PotVar0073568	4	2	28597148	105	1
PotVar0086359	7	3	17995566	201	1
solcap_snp_c1_8212	9	2	21192684	4	1
solcap_snp_c1_8212	9	2	21192684	102	1
solcap_snp_c1_8212	9	2	21192684	190	1
solcap_snp_c2_681	9	2	36646088	190	1
solcap_snp_c2_53559	9	3	20712307	74	1
solcap_snp_c2_53559	9	3	20712307	138	1
solcap_snp_c2_16276	9	3	31520789	138	1
solcap_snp_c2_22758	9	3	32236211	138	1
solcap_snp_c2_33008	10	3	39738068	164	2
PotVar0113312	11	2	28752104	26	1
solcap_snp_c2_49334	12	4	26423442	188	2

Table S4. Summary of the results of the regression of genetic distances with physical distances outside the centromeres.

Chr.	Hom.[a]	Arm	Slope	R^2_{adj}	Parent	Chr.	Hom.	Arm	Slope	R^2_{adj}	Parent
1	1	N	5.00E-06	0.981	1	7	2	S	2.00E-06	0.98	2
1	1	S	2.40E-06	0.991	1	7	3	N	2.00E-06	-	1
1	1	S	2.10E-06	0.995	2	7	3	N	2.80E-06	0.951	2
1	2	N	1.58E-05	0.45	1	7	3	S	3.70E-06	1	1
1	2	S	3.70E-06	0.942	1	7	3	S	2.50E-06	0.98	2
1	2	S	2.10E-06	0.978	2	7	4	N	2.20E-06	0.926	1
1	3	N	3.40E-06	0.974	2	7	4	N	3.90E-06	0.978	2
1	3	S	2.00E-06	0.951	1	7	4	S	2.80E-06	0.998	1
1	3	S	1.80E-06	0.998	2	7	4	S	2.90E-06	0.954	2
1	4	N	6.00E-06	0.98	1	8	1	N	2.00E-06	0.974	1
1	4	N	4.50E-06	0.998	2	8	1	N	1.80E-06	0.986	2
1	4	S	2.90E-06	0.97	1	8	1	S	2.40E-06	0.985	1
1	4	S	2.20E-06	0.996	2	8	1	S	3.80E-06	0.982	2
2	1	S	2.50E-06	0.992	1	8	2	N	3.10E-06	0.992	1
2	1	S	2.90E-06	0.99	2	8	2	N	3.00E-06	0.847	2
2	2	S	3.00E-06	0.992	1	8	2	S	2.20E-06	0.952	1
2	2	S	3.10E-06	1	2	8	2	S	3.30E-06	0.988	2
2	3	S	2.40E-06	0.995	1	8	3	N	3.70E-06	0.903	1
2	3	S	2.90E-06	0.999	2	8	3	S	3.20E-06	0.998	1
2	4	S	2.40E-06	0.958	1	8	3	S	2.30E-06	0.936	2
2	4	S	3.20E-06	0.984	2	8	4	N	1.10E-06	0.849	2
3	1	N	2.60E-06	0.842	1	8	4	S	2.60E-06	0.955	1
3	1	S	4.40E-06	0.919	1	8	4	S	2.50E-06	0.943	2
3	1	S	2.10E-06	0.968	2	9	1	N	2.10E-06	0.953	2
3	2	S	1.50E-06	0.943	1	9	1	S	4.80E-06	0.928	1
3	2	S	2.70E-06	0.997	2	9	1	S	3.80E-06	0.964	2
3	3	S	2.90E-06	0.861	1	9	2	N	6.70E-06	-	1
3	3	S	2.10E-06	0.994	2	9	2	N	4.10E-06	0.997	2
3	4	N	5.40E-06	0.961	1	9	2	S	3.40E-06	0.996	1
3	4	S	2.30E-06	0.951	1	9	2	S	3.10E-06	0.973	2
3	4	S	2.20E-06	0.895	2	9	3	N	2.60E-06	0.93	1
4	1	N	5.00E-06	0.957	1	9	3	N	4.10E-06	0.995	2
4	1	S	2.10E-06	0.95	2	9	3	S	3.90E-06	1	1
4	2	N	2.30E-06	0.991	1	9	3	S	3.80E-06	0.999	2
4	2	S	3.20E-06	0.947	1	9	4	N	2.90E-06	0.973	1
4	2	S	3.20E-06	0.977	2	9	4	N	3.50E-06	0.973	2
4	3	N	2.40E-06	0.988	1	9	4	S	2.60E-06	0.92	1
4	3	N	2.80E-06	0.77	2	9	4	S	3.90E-06	0.965	2
4	3	S	4.30E-06	0.881	1	10	1	N	3.70E-06	0.999	1
4	3	S	6.20E-06	0.983	2	10	1	N	2.30E-06	1	2
4	4	N	3.20E-06	0.984	1	10	1	S	3.20E-06	0.933	1
4	4	N	4.00E-07	0.824	2	10	1	S	0	-	2

4	4	S	2.40E-06	0.901	1	10	2	N	2.10E-06	0.967	1
4	4	S	2.50E-06	0.961	2	10	2	S	3.40E-06	0.965	1
5	1	N	4.70E-06	0.994	1	10	3	S	3.10E-06	0.989	1
5	1	N	3.00E-06	0.927	2	10	3	S	0	-	2
5	1	S	4.50E-06	0.939	1	10	4	N	2.70E-06	0.918	1
5	1	S	2.90E-06	0.966	2	10	4	S	3.90E-06	-	1
5	2	N	4.40E-06	0.905	1	10	4	S	2.80E-06	0.988	2
5	2	N	2.70E-06	0.997	2	11	1	N	3.40E-06	0.992	1
5	2	S	2.90E-06	0.878	1	11	1	N	2.50E-06	0.994	2
5	2	S	3.30E-06	0.997	2	11	1	S	2.90E-06	0.98	1
5	3	N	0	-	1	11	1	S	1.50E-06	0.997	2
5	3	N	3.60E-06	0.952	2	11	2	N	2.40E-06	0.994	1
5	3	S	3.90E-06	0.979	1	11	2	N	3.10E-06	0.991	2
5	3	S	2.90E-06	0.99	2	11	2	S	3.40E-06	0.951	1
5	4	N	2.90E-06	0.982	1	11	2	S	2.50E-06	0.966	2
5	4	N	5.00E-06	0.994	2	11	3	N	3.50E-06	0.993	1
5	4	S	3.80E-06	0.973	1	11	3	N	3.00E-06	0.626	2
5	4	S	5.00E-06	0.969	2	11	3	S	4.30E-06	0.92	1
6	1	S	2.60E-06	0.999	1	11	3	S	3.40E-06	0.998	2
6	1	S	2.90E-06	0.99	2	11	4	N	3.20E-06	0.993	1
6	2	N	3.90E-06	-	2	11	4	N	2.90E-06	1	2
6	2	S	2.40E-06	0.967	1	11	4	S	2.50E-06	0.998	1
6	2	S	2.40E-06	0.872	2	11	4	S	1.80E-06	0.961	2
6	3	N	1.20E-06	0.997	1	12	1	N	4.20E-06	0.993	1
6	3	S	2.30E-06	0.99	1	12	1	N	2.60E-06	-	2
6	3	S	2.80E-06	0.916	2	12	1	S	3.10E-06	0.985	1
6	4	N	2.00E-06	0.966	2	12	1	S	4.40E-06	0.999	2
6	4	S	2.50E-06	0.999	1	12	2	N	4.10E-06	0.98	1
6	4	S	2.50E-06	0.934	2	12	2	N	2.80E-06	0.915	2
7	1	N	7.10E-06	0.91	1	12	2	S	3.90E-06	0.979	1
7	1	N	2.50E-06	0.925	2	12	2	S	4.20E-06	0.988	2
7	1	S	3.20E-06	0.985	1	12	3	N	4.60E-06	0.975	1
7	1	S	3.10E-06	0.926	2	12	3	N	2.30E-06	0.954	2
7	2	N	3.50E-06	0.992	1	12	3	S	5.20E-06	0.996	1
7	2	N	1.70E-06	0.943	2	12	4	N	3.40E-06	0.991	1
7	2	S	3.70E-06	0.949	1	12	4	N	3.40E-06	0.995	2

[a] homologue

Tables S5-S6

Available for download as .xlsx files at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.181008/-/DC1

Table S5. Genotype data of the 1:1 segregating markers for 2 parents and 235 F1 offspring of the AxC population

Table S6. Homologue linkage maps for all 12 potato chromosomes, with 8 separate homologue maps per chromosome (4 homologues x 2 parents)
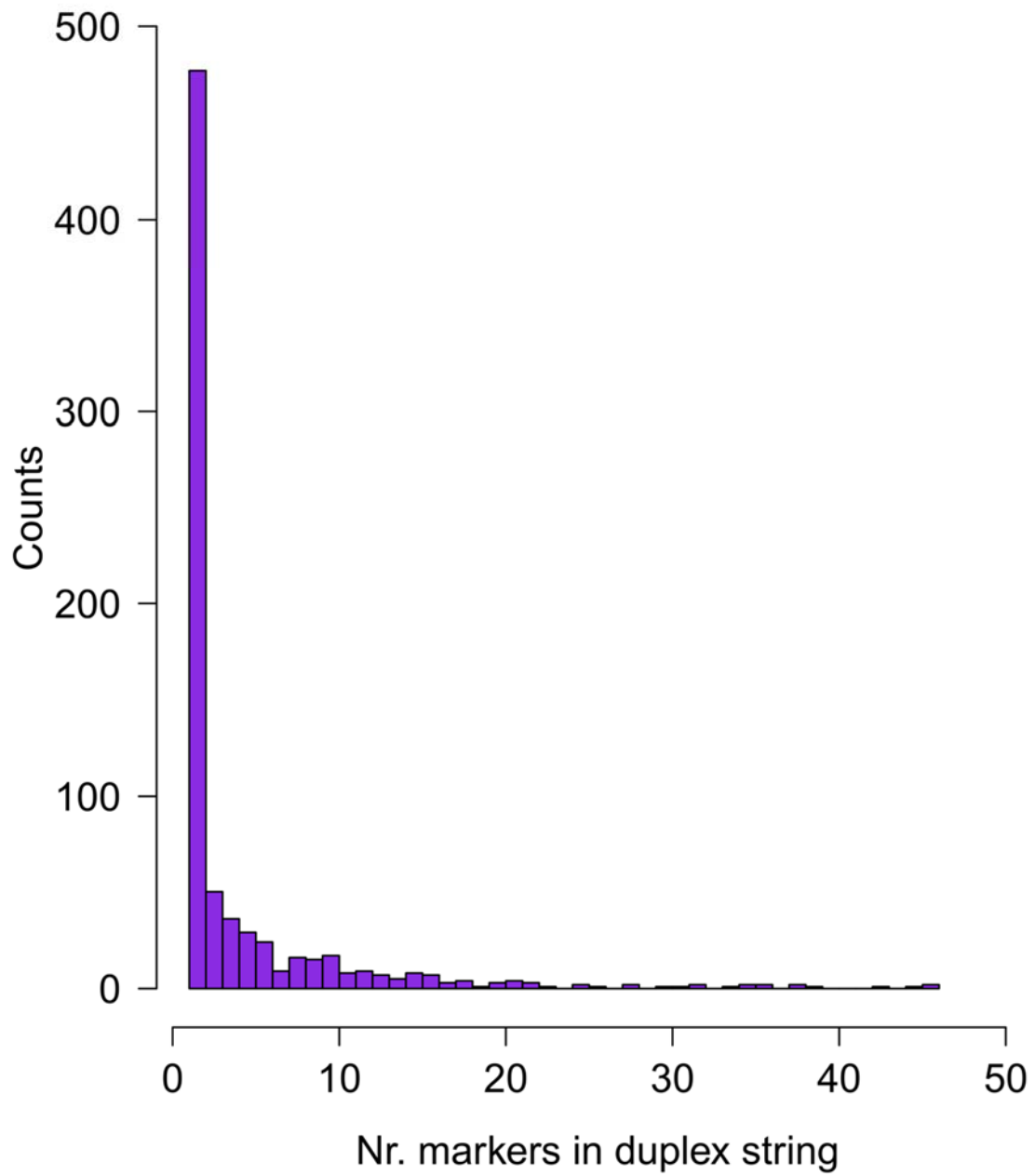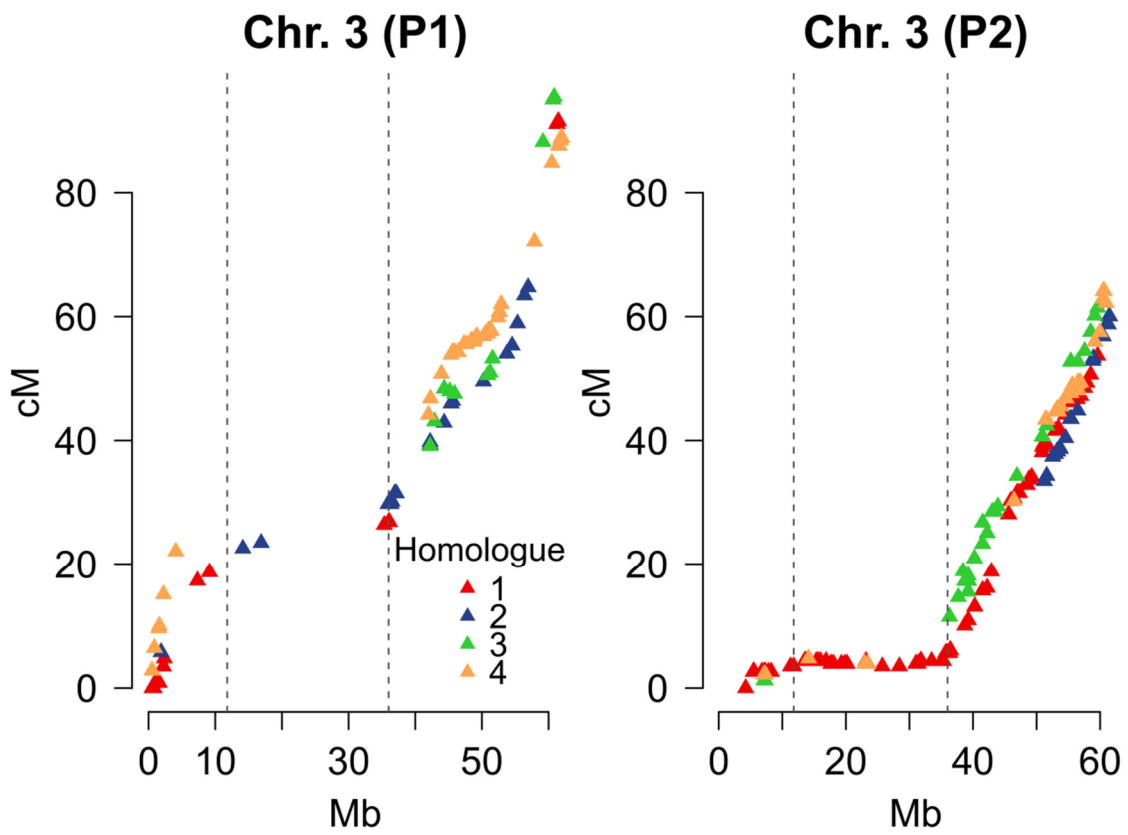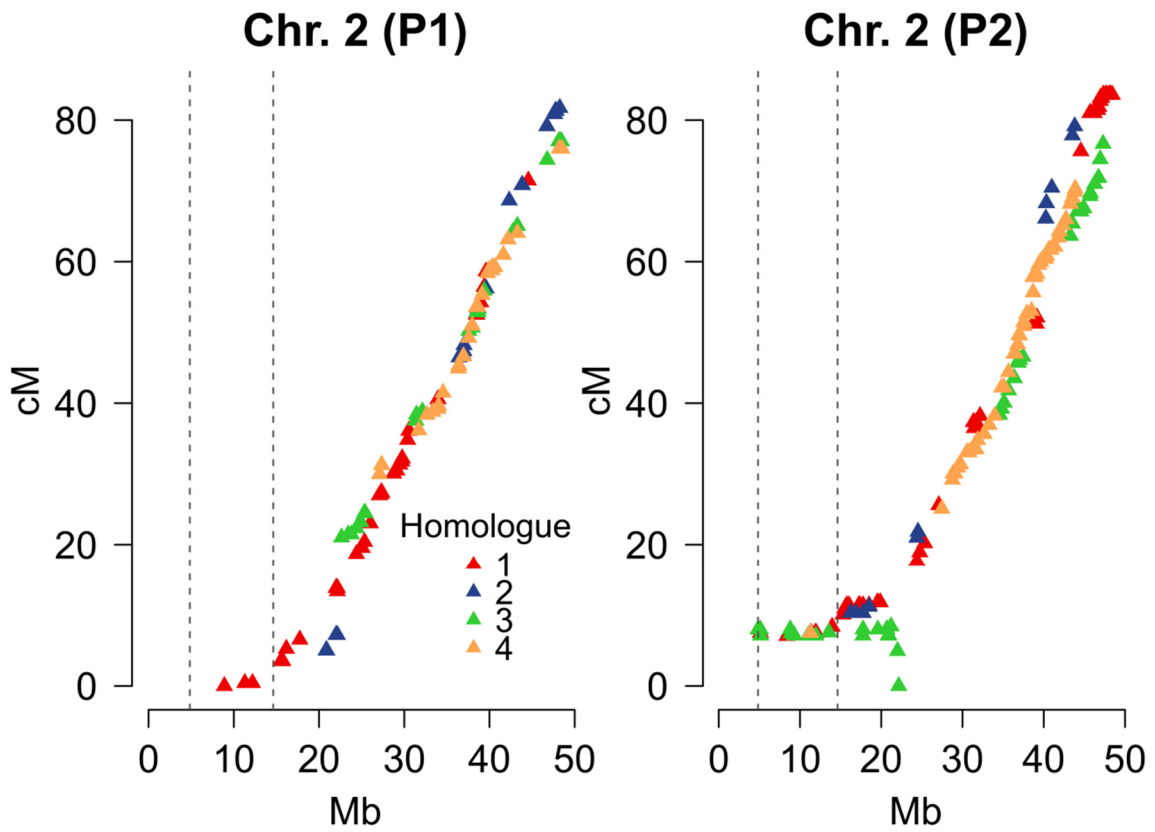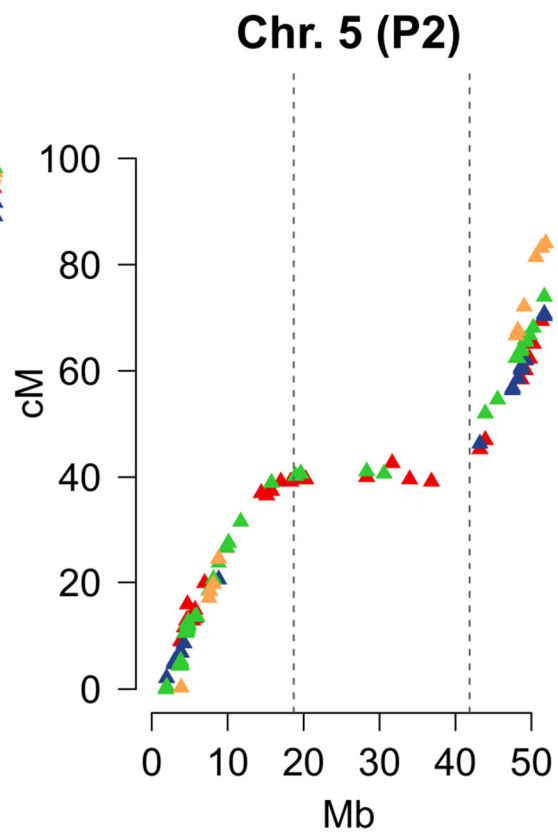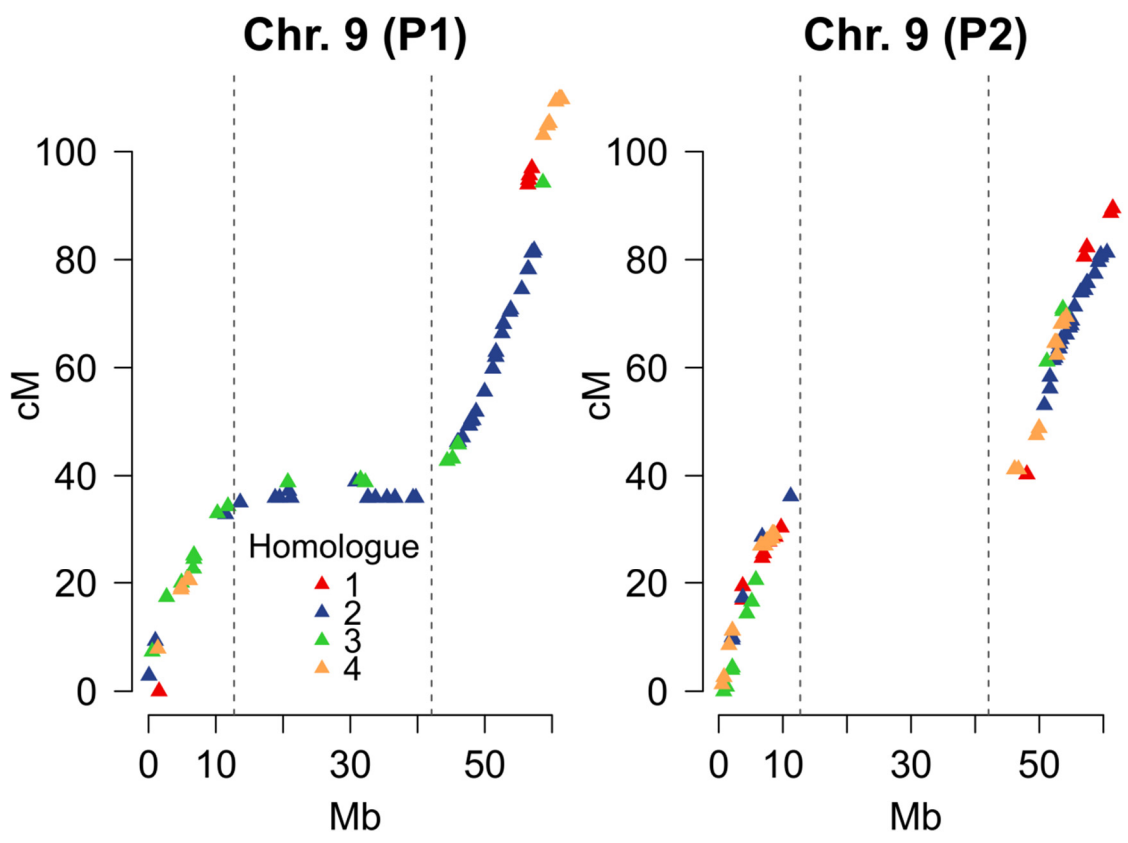
P. M. Bourke *et al.*

Figure S1. Histogram showing distribution of lengths of strings of duplex scores in the dataset
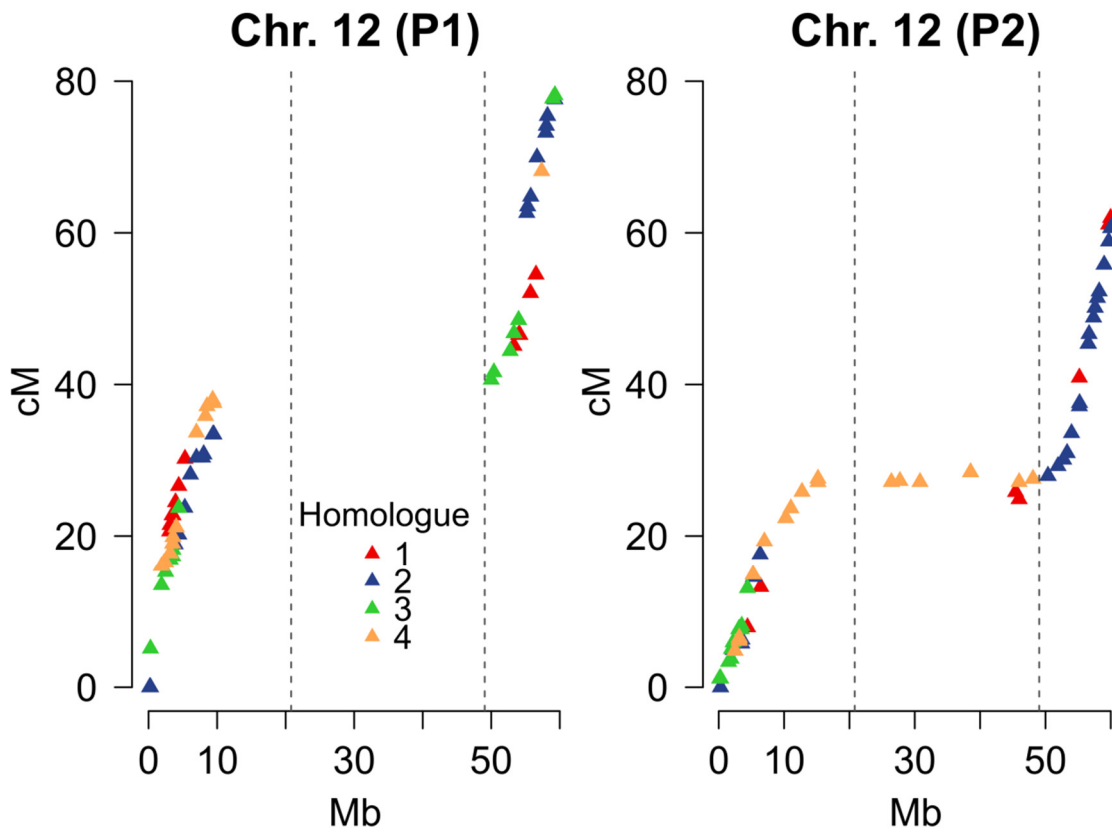
P. M. Bourke *et al.*
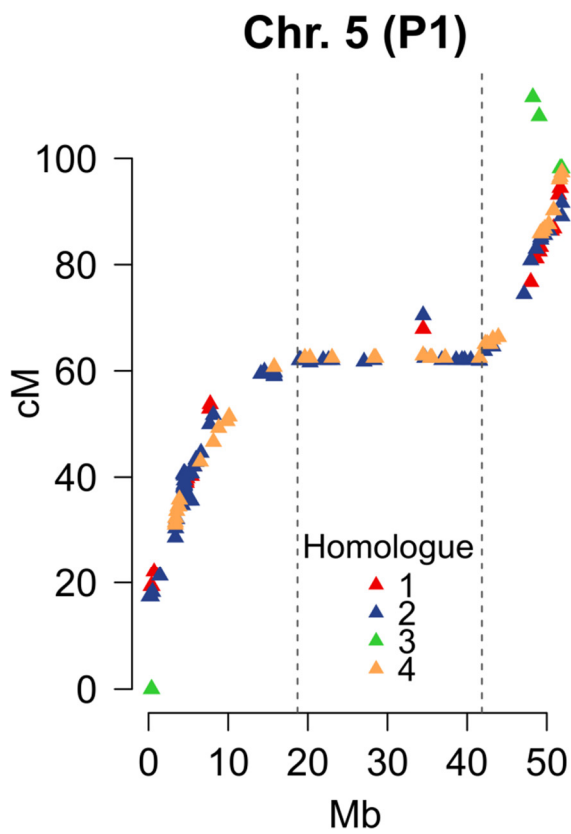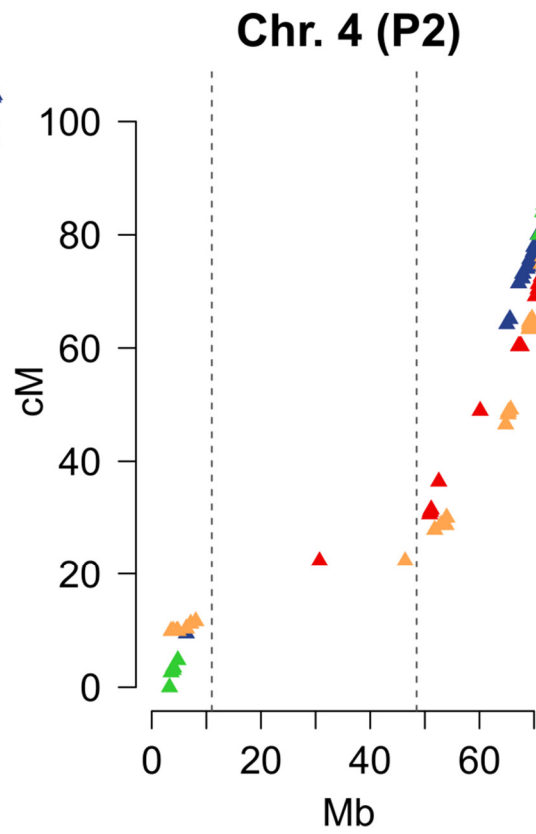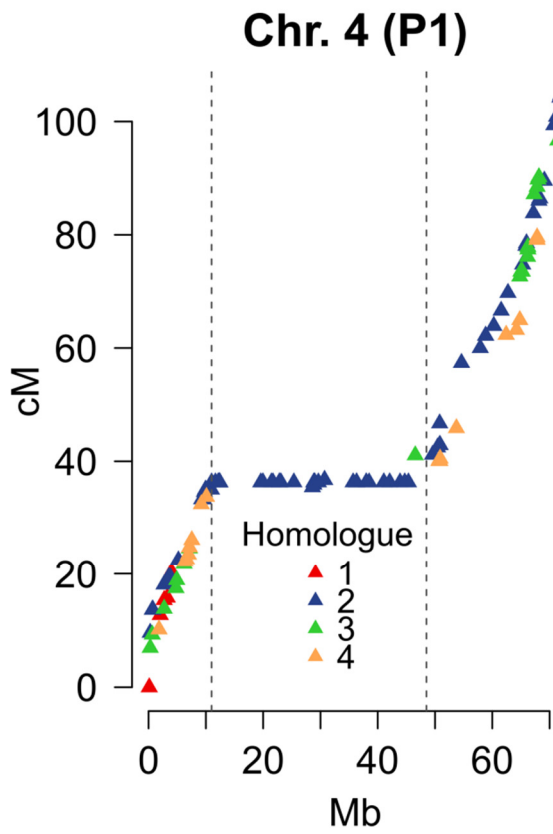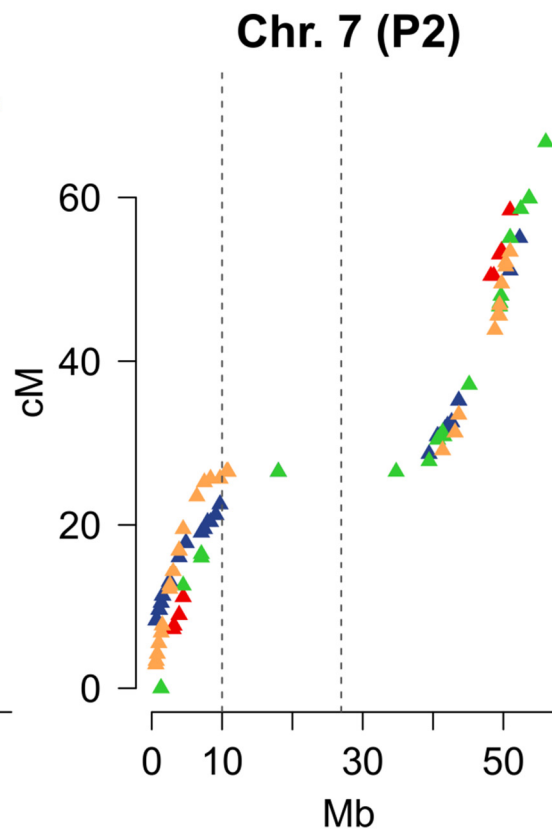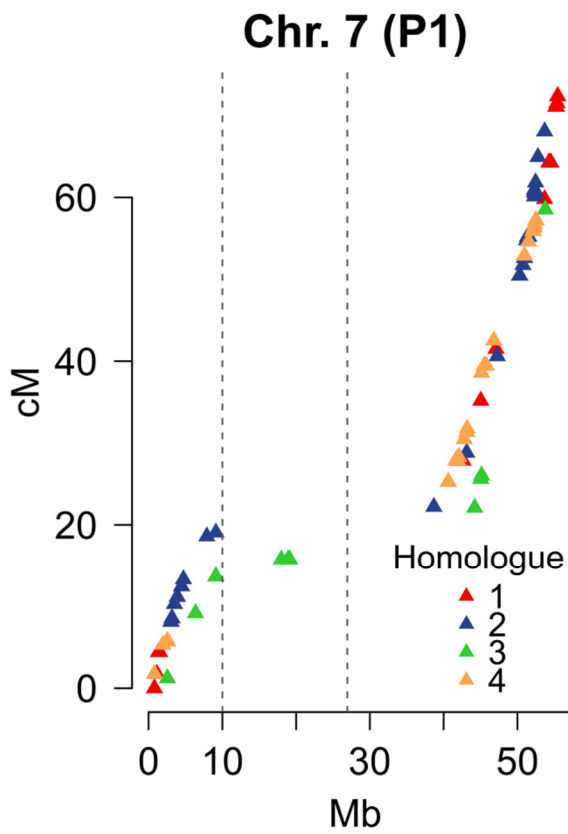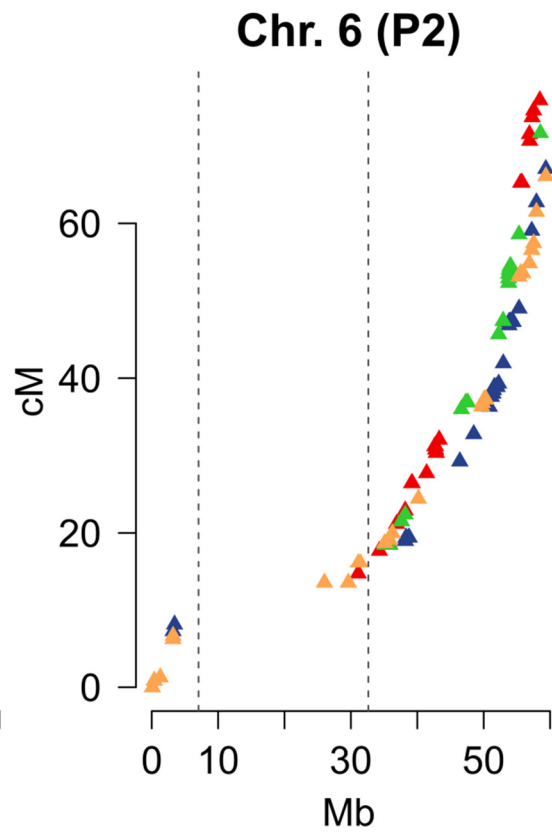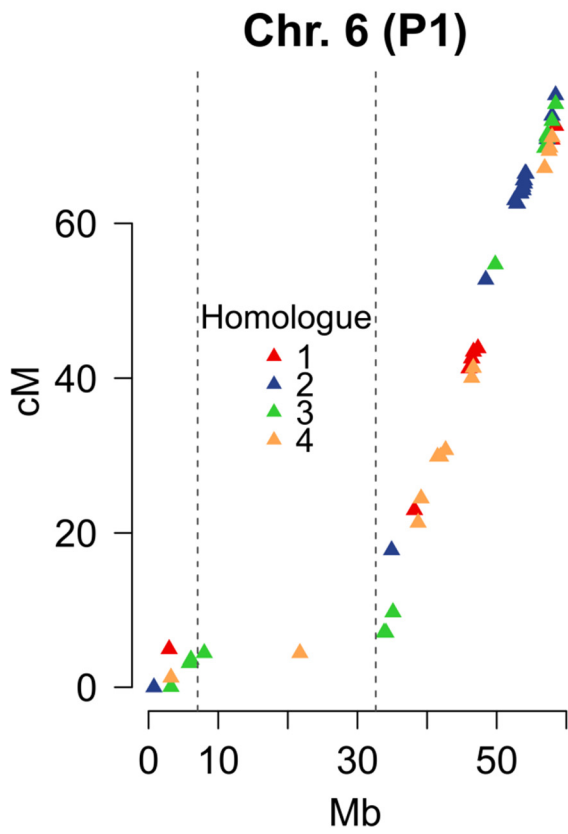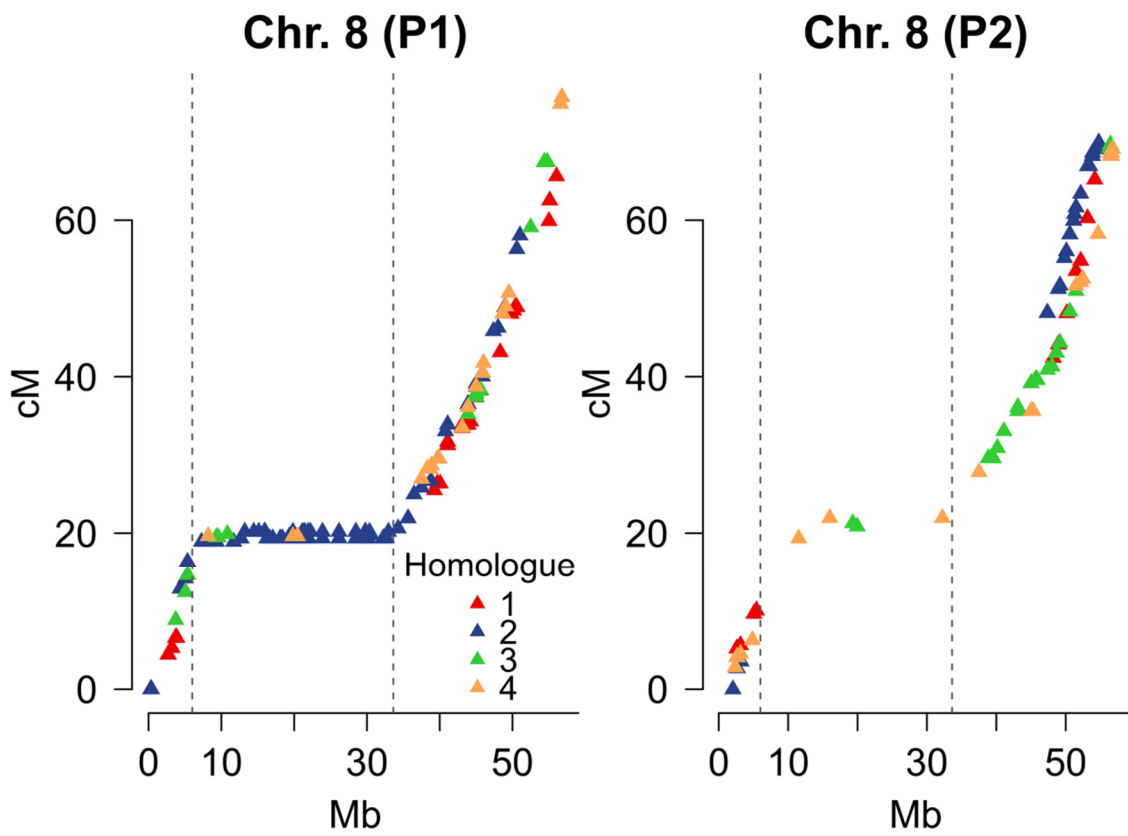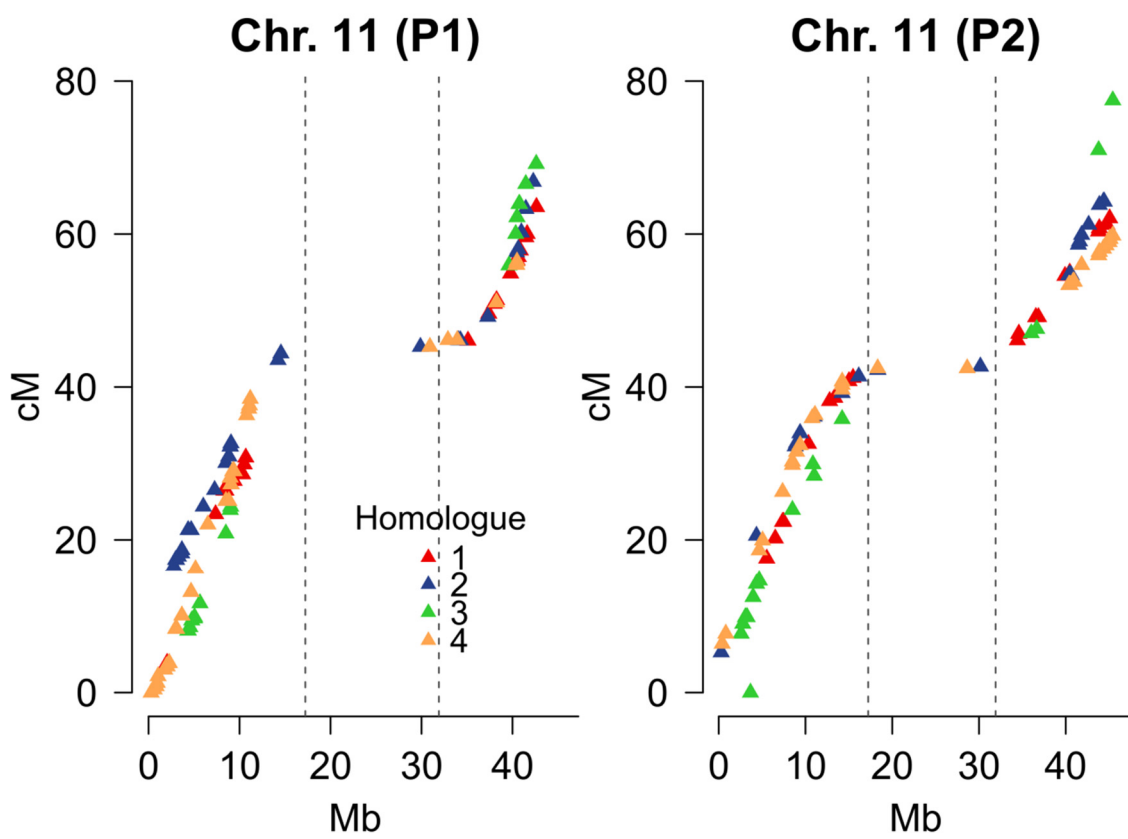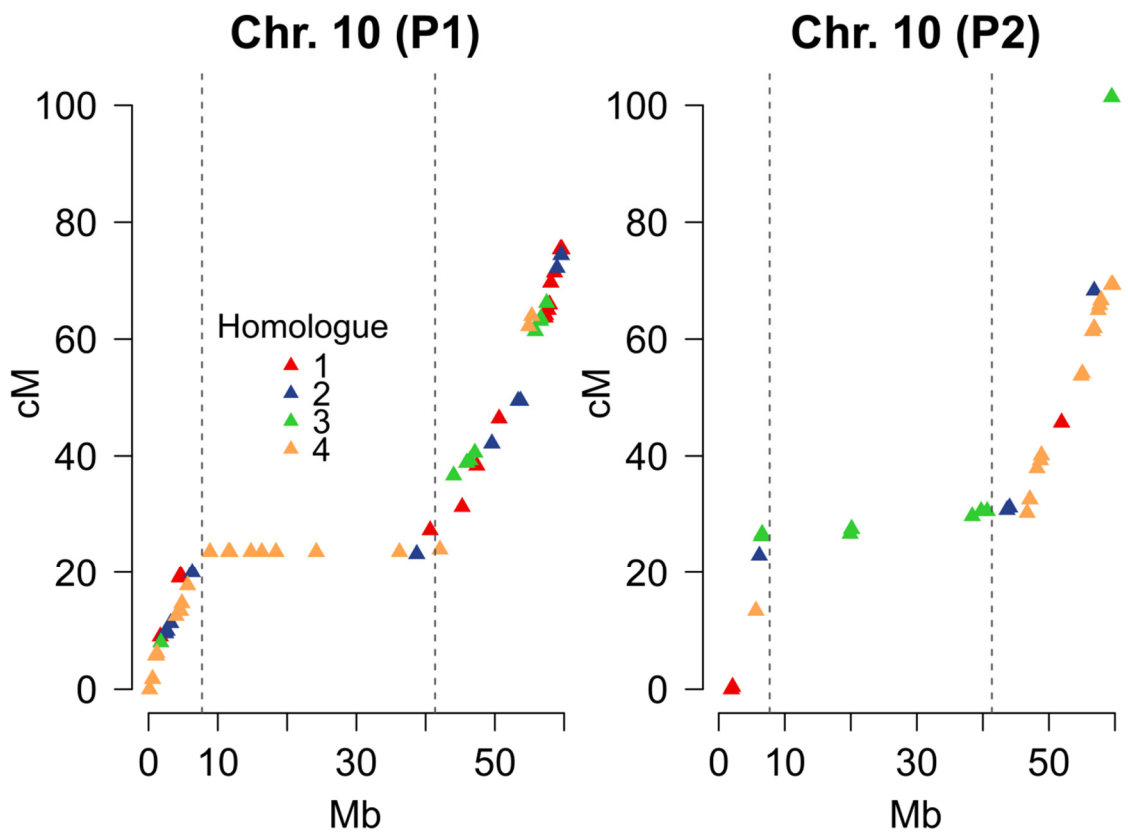
Figure S2. Comparison of genetic to physical distance with homologue maps of potato chromosomes 2-12.