Theor Appl Genet (2005) 112: 187–194
DOI 10.1007/s00122-005-0124-y

ORIGINAL PAPER

Hans van Os · Piet Stam · Richard G. F. Visser
Herman J. van Eck

SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data

Received: 12 August 2004 / Accepted: 28 October 2004 / Published online: 29 October 2005

Abstract High-density genetic linkage maps can be used for purposes such as fine-scale targeted gene cloning and anchoring of physical maps. However, their construction is significantly complicated by even relatively small amounts of scoring errors. Currently available software is not able to solve the ordering ambiguities in marker clusters, which inhibits the application of high-density maps. A statistical method named SMOOTH was developed to remove genotyping errors from genetic linkage data during the mapping process. The program SMOOTH calculates the difference between the observed and predicted values of data points based on data points of neighbouring loci in a given marker order. Highly improbable data points are removed by the program in an iterative process with a mapping algorithm that recalculates the map after cleaning. SMOOTH has been tested with simulated data and experimental mapping data from potato. The simulations prove that this method is able to detect a high amount of scoring errors and demonstrates that the program enables mapping software to successfully construct a very accurate high-density map. In potato the application of the program resulted in a reliable placement of nearly 1,000 markers in one linkage group.

Keywords High-density genetic linkage maps · Software · Simulations · Scoring errors

Introduction

Linkage maps based on molecular markers are important tools in genetic analysis. They are useful for the localization of genes underlying quantitative traits, marker assisted breeding and map based gene cloning. Molecular marker systems like AFLP® (Vos et al. 1995) allow that many markers can be generated in short time. This leads to the construction of highly saturated maps to enable fine-scale genetic mapping and the anchoring of physical maps (Klein et al. 2000).

In principle, these highly saturated or high-density maps can be constructed with the same software as genetic linkage maps of normal density. Commonly used programs like Joinmap (Stam 1993; Stam and Van Ooijen 1995) and MapMaker (Lander et al. 1987) are very suitable for low-density genetic linkage map construction. However, these methods have difficulty in solving the increasing ordering ambiguities in denser maps (Lincoln and Lander 1992; Van Os et al. 2005). Denser maps have more loci than normal maps, but the offspring genotypes contain the same amount of recombinations. With these fixed amounts of recombinations, the increased number of markers in denser maps are separated on an average by less recombination events. Moreover, mapping algorithms based on pairwise distances will try to determine the order within clusters of markers; even for co-segregating markers or markers that only differ in a few scoring errors, but in fact share the same genetic position. In high-density maps, errors do not only give problems within marker clusters, but also across recombination events and thus severely complicate the establishment of the true marker order.

An accurate marker order is indispensible for further application of the map, like for instance, map based cloning. We state that marker order is more important than estimated map distances. Map distance estimates are trivial as they may vary across mapping studies by several centimorgans. For clusters of markers that co-segregate, the order is indeterminate. Therefore it is not correct to suggest non-existing distance between markers caused by scoring errors or missing values.

The troublesome data points in the data are most likely to be caused by inaccurate scoring, but some data points that cause ambiguities in the marker order can

Communicated by G. Wenzel

H. van Os · P. Stam · R. G. F. Visser · H. J. van Eck (✉)
Laboratory of Plant Breeding, Wageningen University,
P.O. Box 386, 6700 AJ Wageningen, The Netherlands
E-mail: herman.vaneck@wur.nl
Tel.: +31-317-482837
Fax: +31-317-483457

also be caused by double recombination events, gene conversions, mutations and other biological phenomena. These various causes of ambiguous data are collectively called singletons. The term 'singleton' in the context of mapping data has first been used to indicate the misclassification of a marker phenotype (Nilsson et al. 1993). A singleton is, in fact, a single locus in one plant that appears to have recombined with both its directly neighbouring loci (see Fig. 1). During map calculation, every singleton has to be treated as the unlikely event of a double recombination. We propose to identify and temporarily remove the singletons from the data. By eliminating these singletons, most ordering ambiguities are solved, including those in marker dense clusters.

As a consequence, mapping algorithms will not be hampered by these ordering ambiguities and can calculate the best possible map for the data set. This ideal framework map will be used to refit the raw data, supplying the verification of this map and a quality label for each marker that specifies the number of singletons these markers contain.

In this paper a statistical method is presented that can identify and remove the most obvious singletons in genetic linkage data. The method uses the marker order calculated by mapping software before eliminating the singletons. It is implemented in a computer program for the case of a first generation backcross population. This paper illustrates the advantages and potential pitfalls of eliminating singletons and provides concise instruction on how this method should be applied.

The method is tested on simulated data for different aspects like error percentage, population size and marker density. Besides the results of the simulation studies, experiments with real mapping data from potato are discussed as well.

Materials and methods

Software

The idea behind the identification of a singleton at a particular marker locus, i, is to compare the observed marker score at locus i, y_i, with a local prediction of the marker score, $\hat{y}_i$. The observed marker score, y_i, takes the value 1 when the allele is identified as coming from one of the sister chromatids, and 0 when coming from the non-sister chromatids. The local prediction of the marker score $\hat{y}_i$ is calculated as the weighted w_j average of the observed scores y_j within a defined number of loci L flanking locus i on either side:

$$\hat{y}_i = \frac{\sum_{j \in L} w_j y_j}{\sum_{j \in L} w_j},$$

with, $L = \{j : j \leq \delta, j \neq 0\}$, where δ is the maximum number of flanking loci around locus i that contributes to the local prediction for the marker score at i. Various weighing regimes were tested in combination with different choices for δ, but these parameter settings were rather immaterial to the performance of the procedure. For that reason we only present the results for $\delta = 15$ loci and with weights declining in a roughly quadratic fashion ($w_1 = 0.998$; $w_2 = 0.981$; $w_3 = 0.934$; $w_4 = 0.857$; $w_5 = 0.758$; $w_6 = 0.647$; $w_7 = 0.537$; $w_8 = 0.433$; $w_9 = 0.342$; $w_{10} = 0.265$; $w_{11} = 0.202$; $w_{12} = 0.151$; $w_{13} = 0.112$; $w_{14} = 0.082$; $w_{15} = 0.059$).

The absolute difference between observed marker score and predicted marker score, $d = |y_i - \hat{y}_i|$, is proportional to the probability that the marker score at i represents a singleton. Threshold values for d, above which singletons are identified, were adaptively chosen.

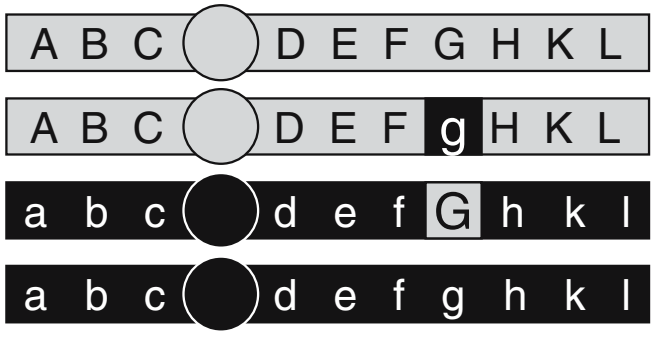

Fig. 1 Four meiotic products after a double recombination event involving two non-sister chromatids causing a singleton at locus 'G/g'. The *bar* indicates a chromatid, whereas the centromere is represented by a *circle*

Illustration of application

Before SMOOTH can be applied to the data, a preliminary marker order has to be established. This map is the starting point for singleton detection and is still far from ideal. Although a singleton is context dependent, the most obvious singletons are clearly perceptible even in less ideal maps.

In Fig. 2, the difference d for one female gamete in the ultra-dense AFLP map of one chromosome of potato (Isidore et al. 2003) is shown. This gamete was the result of two recombination events: one recombination event occured between locus 76 and locus 77 and one recombination event occured between locus 944 and locus 954. Around the recombination events, the value of d approaches 0.5. The most likely singletons ($d = 1$) can be observed at loci 2, 21, 105, 474, 508, 536, 615, 735, 793, 898 and 918.

When the value d of each data point is calculated, a threshold for singleton removal can be set. The singletons are removed in an iterative process, alternately using a mapping algorithm and SMOOTH. In other words, a mapping algorithm like RECORD (Van Os et al. 2005) is used to calculate the marker order, subsequently SMOOTH is used to remove singletons, after which the marker order is recalculated with RECORD, etc. In principal, all mapping algorithms can be used, but data sets containing 500 markers demand only

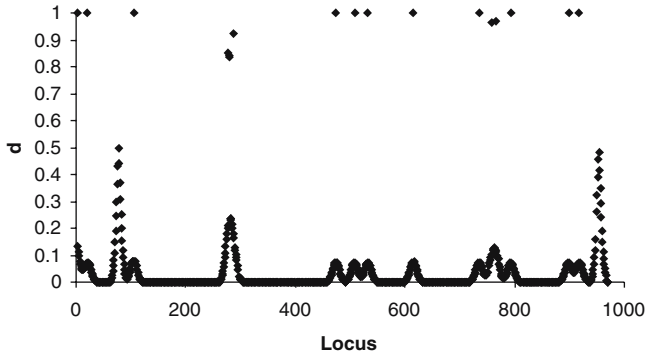

Fig. 2 The difference *d* for all 971 loci in a particular gamete in the ultra-dense potato map. Data from chromosome I (Isidore et al. 2003)

20 min for analysis with RECORD on a 350 MHz processor. By comparison, JoinMap will take 9 days to calculate a map from a linkage group containing 500 markers on the same type of machine. In the first cycle of the iteration, a high threshold (0.99) is set and the most likely singletons are replaced by missing values. During the following cycles, the threshold is slightly decreased. Using more cycles in the iteration and smaller decreasing steps in the threshold, singleton removal is more accurate. In this experiment, the iteration is continued for 15 cycles while decreasing the threshold with 0.02, until the final threshold of 0.70 is reached. At this point, most singletons are removed from the data. Empirical evidence will be provided below, that at threshold 0.70, the amount of singletons that remain in the data set is in balance with the number of correct data points that are unjustly removed.

After removing all singletons, an unambiguous framework map can be constructed. Subsequently, the original marker data can be fit into the framework map by maximum likelihood, providing a verification of the framework map and also a quality label for each marker.

Simulations are used to demonstrate that the program detects the singletons and that eventually the correct marker order is obtained. The practical applicability is established by the analysis of an experimental data set of potato comprising 971 markers in 130 individuals.

Simulated data

The power of SMOOTH to detect singletons was tested on simulated data. For this purpose, several first generation backcross (BC1) populations were generated varying in the number of loci, population size and error percentage as shown in Table 1.

The simulated data were produced as follows: a given number of loci were randomly positioned (according to a Poisson process) along a single chromosome of 50 cM; genotypes were generated for a BC1 progeny following standard Mendelian segregation (assuming no crossover interference). Errors were randomly introduced in the data set and the positions of these errors were stored in a logfile. The range of error percentages increased from 1 to 25% thus creating 25 data sets for each population. In Experiment I, emphasis is put on both error percentage and marker density. In Experiment II, the effect of error percentage and population size is evaluated.

Corrected data sets were obtained from each simulated data set with introduced errors by calculating marker orders with RECORD (Van Os et al. 2005) and removing singletons with SMOOTH. In the mean time, SMOOTH kept track of all the data points that were removed during the mapping and cleaning process. After completion of the process, this list of removed data points was compared with the list of introduced errors. From this comparison, the number of errors were counted that were found and missed by SMOOTH. Also the number of correct data points that should not have been removed were counted. The marker order before and after cleaning with SMOOTH was compared with the original simulated map, using Spearman's rank-order correlation coefficient r_s, between the expected marker position on the simulated map and the observed marker position on the map calculated by ORD.

Experimental data

Besides the simulations, actual mapping data were analyzed from the ultra-dense genetic map of potato (Isidore et al. 2003). From the data set of this outbreeding population, the AFLP markers segregating from only one parent were considered. Both parental maps were analyzed separately. In the maternal map, 4,187 marker were segregating and 3,413 markers segregated in the paternal map. Grouping was done with JoinMap 2.0 and divided the data in 12 groups. A preliminary marker order was used to assign the linkage phase to all markers based on their flanking markers. After linkage phase assignment, the data could be treated as if it were a first generation backcross. This approach, also called two-way pseudo-testcross (Grattapaglia and Sederoff 1994), is commonly applied for map construction in populations descending from non-inbred parents.

Marker ordering was done by RECORD, while SMOOTH cleaned the data from singletons applying the same approach as was used for the simulations.

Table 1 Values for simulation variables used in the two different simulation experiments

Variables	Experiment I	Experiment II
Map length (cM)	50 (fixed)	50 (fixed)
Number of loci	10, 25, 50, 100, 250, 500	100 (fixed)
Population size	100 (fixed)	50, 100, 150
Percentage scoring errors	1, 2, 3,...,25%	1, 2, 3,...,25%
Percentage missing observations	0% (fixed)	0% (fixed)

Results

Simulated data

The utility of SMOOTH in obtaining an accurate marker order was evaluated by simulation experiments. In Experiment I, the effect of error percentage and marker density were assessed. The accuracy of the marker order with and without the application of SMOOTH was examined using the rank correlation coefficient between the calculated marker order and the simulated marker order. The quality of dense genetic maps can be improved considerably by the application of SMOOTH. The value of SMOOTH was most obvious in the data set with the highest marker density in Experiment I. The rank correlation coefficients for this data set consisting of 500 loci and 100 individuals are shown in Fig. 3. Results are generated for error percentages ranging from 1 to 25%. Rank correlation coefficients are shown for both approaches, i.e. before and after cleaning with SMOOTH. Without SMOOTH, marker orders with intolerable inaccuracy are produced, when more than 5% error is present. However, SMOOTH enables mapping software to calculate accurate maps from data sets with error levels up to 20%. Obviously SMOOTH is able to recognize most of the singletons in the data and enables the mapping software to accurately position the markers.

To understand the process of singleton removal in detail, the detected singletons were compared with the introduced errors in the data sets. In this comparison we monitored the unjustly removal of correct data points and the errors that were not detected by SMOOTH. Figure 4 shows the percentage of errors that were not detected by SMOOTH and the percentage of data points that were unjustly removed from the total amount of data points in the same data set as mentioned in Fig. 3. SMOOTH recognizes the vast majority of errors, e.g. at 10% error level, 95% of the errors were detected, reducing the amount of errors to 0.5%. The number of errors that were not detected and the number of data points that were unjustly removed are more or less similar for lower error levels. This indicates that the choice to stop SMOOTH at a final threshold of $d = 0.7$ is justified. By decreasing this threshold even further, the number of data points that are unjustly removed would increase and surpass the number of undetected errors.

Close inspection of the position of errors that were not detected or data points that were unjustly removed, revealed that they occurred near recombinations and at the ends of the map. Close to recombination events, the flanking markers at either side of the recombination offer contradicting information. Therefore error detection in the vicinity of recombination events is more complicated. At the ends of the map, the difference between the last recombination or the last singleton cannot be determined. Therefore the last recombination event should be confirmed by at least two markers distal to that recombination.

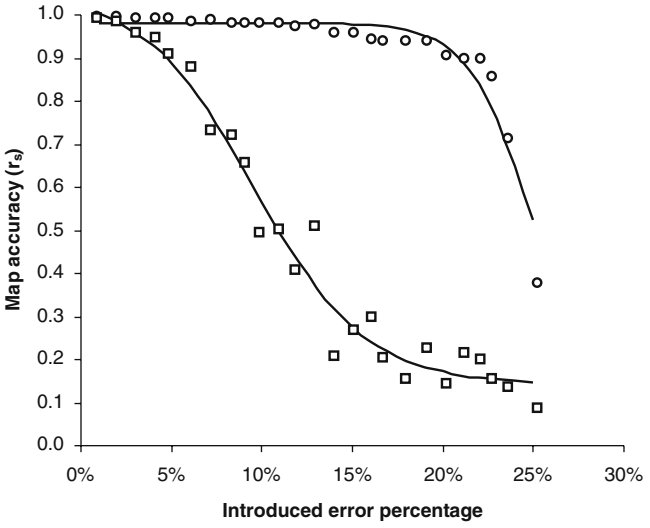

Fig. 3 The rank correlation coefficient between the calculated map and the original simulated map before ('*open square*') and after ('*open circle*') using SMOOTH for different levels of scoring errors based on simulated data sets with 500 loci on 50 cM and 100 individuals

Marker density is an important factor to enable error detection, as can be observed from the results of Experiment I shown in Fig. 5. The percentage of undetected errors is lower in data sets with a higher marker density. This is not surprising because the concept of smoothing genetic linkage data is based on the redundancy in genetic information. In high density data sets, the required amount of 30 neighbouring data points at

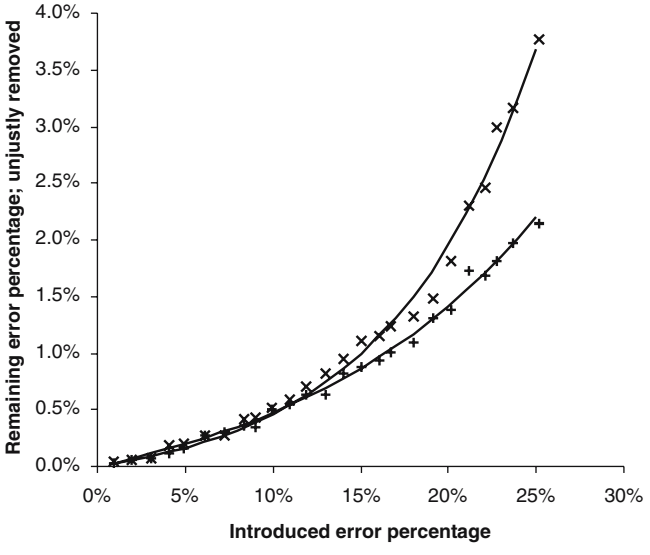

Fig. 4 The percentage of remaining errors from the total data set ('×') and the percentage of unjustly removed data points ('+') for different levels of scoring errors based on simulated data sets with 500 loci on 50 cM and 100 individuals

close genetic distance is available, but data sets with ten markers per linkage group only contain up to nine neighbouring data points over a large distance to predict the marker score.

The effect of population size was analyzed in Experiment II. Figure 6 shows that marker ordering is more accurate in larger populations. In fact, this is not the result of applying SMOOTH, but due to the increased performance of the mapping algorithm. As population size increases, more recombination events between a pair of markers can be observed, which adds to the resolution between the markers. The ordering of the markers will be more accurate and the relative impact of missing observations and singletons will decrease. Furthermore, the marker score predictions by SMOOTH will be more precise due to the more accurate order of the markers.

Experimental data

To compare the results from the simulations with real data, the software was tested on a data set from the high-density map of potato (Isidore et al. 2003). After cleaning the data with SMOOTH, the data were visually inspected for any undetected errors. This revealed a systematic error caused by a group of markers based on AFLP primer combinations from a batch of newly isolated DNA. The confusion of genotypes was solved by removing these individuals from the new set of markers.

When all data ambiguities were removed, a vast amount of redundancy was observed. For instance in chromosome I and IV, a large cluster of cosegregating markers, presumably the centromeric region, contained more than half of the total amount of markers in both the maternal and paternal map. Finally, by deleting the

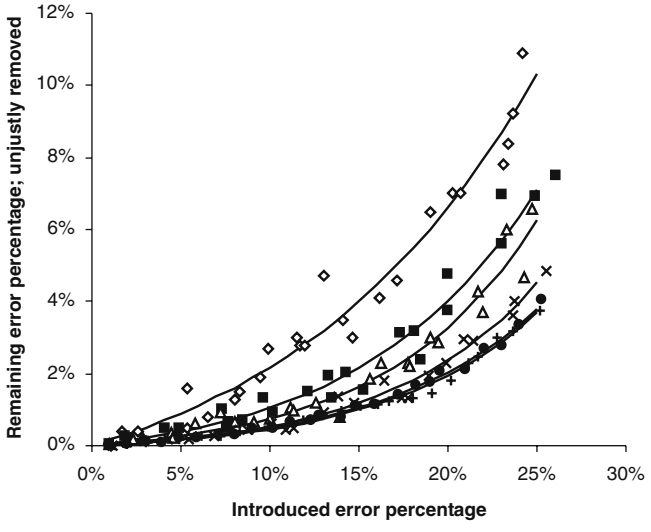

Fig. 5 The percentage of remaining errors from the total data set for different levels of scoring errors based on simulated data sets with 100 individuals. The number of loci is indicated by: '*open diamond*' for 10; '*filled square*' for 25; '*open triangle*' for 50; '×' for 100; '+' for 250 and '*filled circle*' for 500 loci

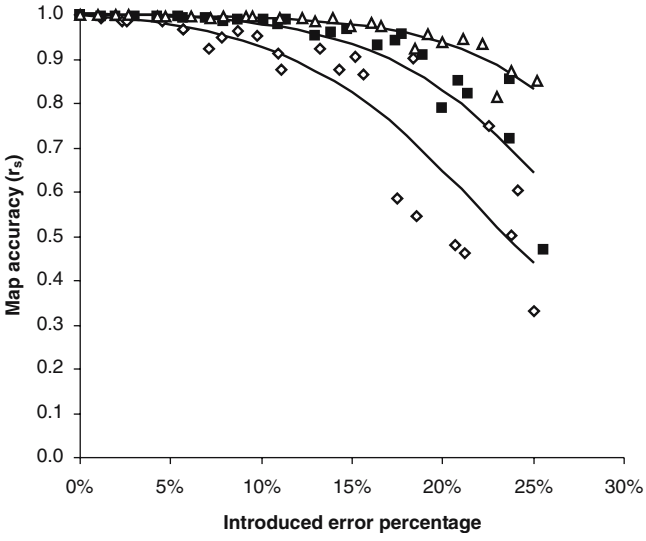

Fig. 6 The rank correlation coefficient between the calculated map and the original simulated map after using SMOOTH for different levels of scoring errors based on simulated data sets with 100 loci. The population size is indicated by: '*open diamond*' for 50; '*filled square*' for 100 and '*open triangle*' for 150 individuals

redundant markers from all linkage groups, framework maps were obtained that only consist of unique corrected markers. These markers were converted into bin signatures by restoring all missing values that were not flanked by recombination events. A bin is a unique and most accurate representation of a marker at a certain genetic position. A bin contains at least 1 marker and cannot be divided within the given population. Bins are numbered consecutively, based on the recombination events. As a consequence, the bin numbers can be directly translated into map units. Both parental framework maps were free from ambiguities and all the markers were fitted into the most likely bin by maximum likelihood. The map was inspected for possible inconsistencies with the refitted markers and some minor corrections were made to the bins. Redundant and empty bins were removed; bins that appeared to contain a recombination event were split up and missing values in the bin signature were restored if possible.

To illustrate the difference between a framework map as described above and a conventional map obtained with JoinMap or RECORD, two linkage groups are shown in Fig. 7. These linkage groups were derived from the high-density map of potato and represent the paternal map of chromosome III and IX respectively. Linkage group III comprised 124 AFLP markers and linkage group IX comprised 190 AFLP markers. No clustering of markers was observed for linkage group III, but linkage group IX contained a centromeric cluster of 27 cosegregating markers. The marker order from RECORD is basically the same as the order in the framework map. However, four markers with an exceptional high number of scoring errors are positioned at the end of the linkage group; a commonly observed artifact of mapping software. Major ordering ambigui-

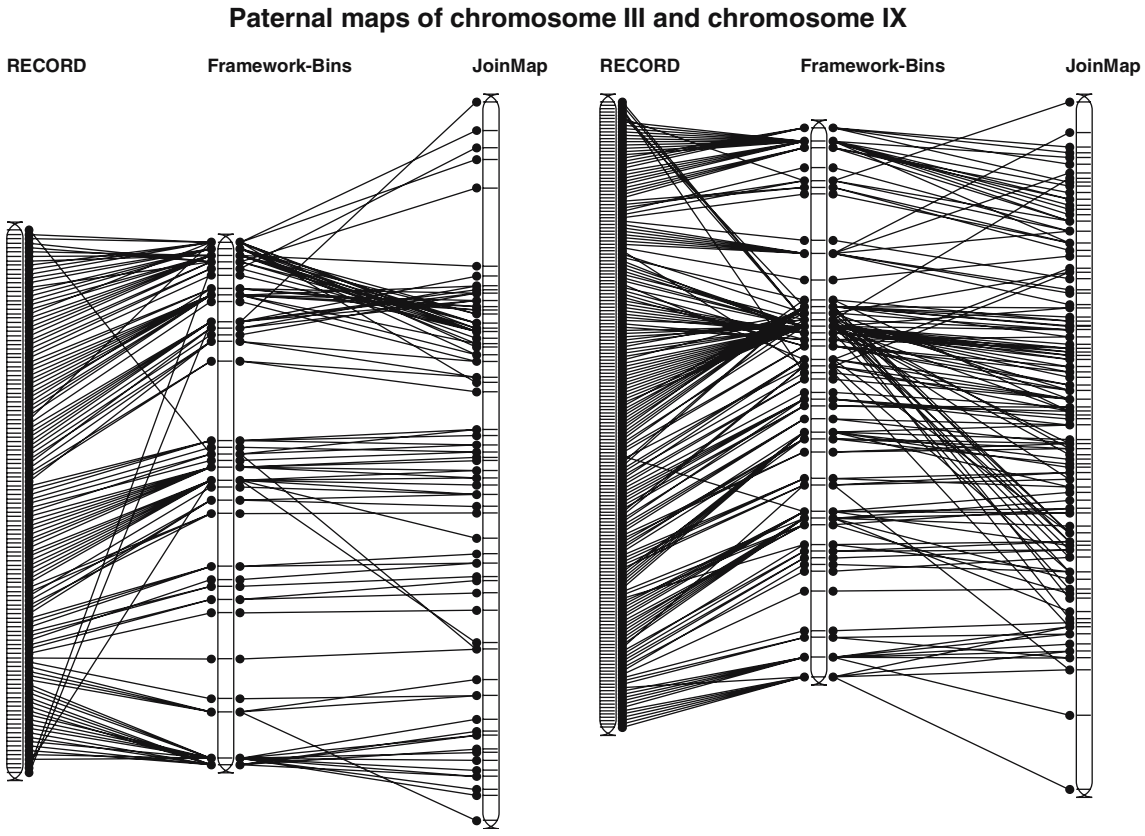

Fig. 7 Comparison between three methods of linkage map construction on two different linkage groups. The framework map in the middle is obtained by SMOOTH and the uncleaned markers have been refitted in the bins. Flanking maps have been constructed from the uncleaned data set by RECORD and JoinMap. The paternal map of potato linkage group III from the the high-density map of potato (Isidore et al. 2003) is shown on the *left*, the paternal map of potato linkage group IX is shown on the *right*. Relative marker positions are displayed by aligning the results of the three methods of linkage map construction. RECORD produces a marker order; distances are proportional to the number of markers. The distances between the bins in the framework map depend on the number of recombination events, which are transformed into centiMorgans. The markers on the map produced by JoinMap are displayed at their corresponding centiMorgan position

ties can be observed around the centromeric cluster in linkage group IX. JoinMap produces a map which is in length roughly similar to the framework map. However, some map inflation can be observed at both ends of the linkage groups. Ordering ambiguities are more abundant in marker dense areas: markers from the centromeric cluster with elevated levels of singletons are pushed away from the centromeric region and dispersed towards the distal ends of the map.

In the data set of the high density genetic linkage map of potato, the number of singletons for each marker was calculated by comparing the original data of each marker with the signature of its most likely bin (see Fig. 8). The average number of singletons per marker was 3.9 in 130 individuals (3.0%). In contrast with the simulations, the distribution of singletons in the experimental data was not random. In the maternal map, one third of the markers did not contain any singletons, which provides a verification for the framework map. However, some markers contained up to 38 singletons. In fact, 10% of the markers were responsible for more than half of the scoring errors. Despite the fact that in reality singletons are not randomly distributed, SMOOTH was able to detect them to enable the construction of a solid framework map.

Discussion

Singletons, whether or not caused by biological phenomena or human error, seriously hamper high-density genetic linkage map construction. For calculating a reliable linkage map or marker order, these singletons have to be removed. We devised a statistical method to detect and remove singletons from high-density genetic linkage data. The approach is based on predicting marker scores on the basis of the available neighbouring data points, which are more abundant in denser maps. Although in denser maps the rising amount of errors becomes increasingly difficult to handle with current mapping software, this new method takes advantage of the redundancy in high density data sets. The excess of markers within a close genetical range, are the basis of a reliable estimate of the marker score. By removing

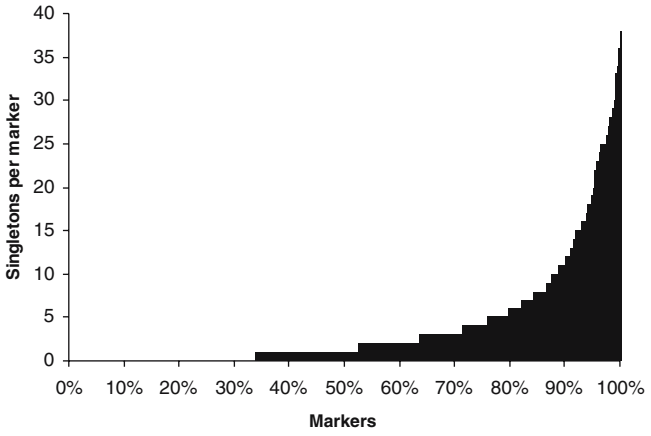

Fig. 8 Quality of the data of the ultra dense map of potato (Isidore et al. 2003). The number of singletons per marker is calculated after determining the most likely position of the markers on the framework map. The markers are sorted from left to right according to their increasing amount of singletons

highly unlikely marker scores from the data, the true recombination events will remain in the data and facilitate marker ordering.

SMOOTH has been extensively tested on simulated data. The results have provided convincing evidence that more than 95% of the singletons can be detected. With a large amounts of errors present in the data, a reliable and accurate map can only be constructed when applying SMOOTH. Therefore we conclude that the program has great utility in high density mapping, which is proven by the successful application to the experimental data set of potato for the construction of an accurate framework map.

It is advisable to use SMOOTH for data sets with at least 100 loci per linkage group, because the error detection is dependent on the amount of neighbouring markers. Although the program is therefore not intended for globally smoothing low-density maps, it can be useful in cleaning up marker dense clusters in low density maps. These marker clusters are regularly observed in genetic maps (Strommer et al. 2002) and are often situated around the centromere of the chromosome where recombination is suppressed.

The error detection works less on the two distal ends of the chromosome and close to recombination events. Here, the predicted value of the data points is based on two sets of data points with contradicting information. In these situations, there is a risk of removing data points that are correct. However, the consequences of removing too many data points are not severe. In fact, the removal of correct data points in the vicinity of recombinations causes a local decrease of the effective population size and has therefore the same effect as the removal of an individual offspring genotype from the mapping population. The consequences of these unjustified removals can be solved by correcting the framework map using the original data. This verification of the framework map is done by maximum likelihood comparison of the original markers with the framework bins. Moreover, the risk of cleaning data points that were not erroneous is sufficiently reduced by employing the method in an iterative process with the mapping algorithm.

The verification of the framework map by refitting the original data does not provide indisputable evidence for the true marker order. Nevertheless, it provides a detailed overview of the ambiguities in the data. The accuracy of the ultra-dense marker order can only be assessed in simulation studies where the true marker order is known. For potato, the consistency of the genetic map with a physical map is expected to provide the evidence for the current marker order.

The program has been applied for the construction of the ultra-dense genetic linkage map of potato. All linkage groups of this map contain more than 100 markers. Accurate mapping of these large linkage groups was not possible, despite the even small amounts of scoring errors. Most of these errors could be erased by manual re-evaluation of the AFLP gels, but in spite of these time-consuming efforts, accurate marker ordering was still severely complicated. With SMOOTH, the ambiguities of the data were removed to construct a framework map that provided accurate marker placement.

To a certain extent, error detection is available in the current version of MapMaker (Lincoln and Lander 1992). Instead of removing possible errors, MapMaker takes the possibility for a data point to be erroneous into account and avoids potential map inflation. The errors remain in the data set and still cause ordering problems, therefore MapMaker is not adequate to calculate high density maps.

Besides backcross populations, the concept of SMOOTH can also be suitable for analyzing other populations like F2. Dominance will nevertheless decrease the detection power of singletons. In this case, the marker density should be higher than in backcross populations to ensure a reliable singleton detection.

In conclusion, with the advent of ultra-dense genetic linkage maps, a completely new approach of data analysis is required. In combination with RECORD (Van Os et al. 2005), this method provides a fast and accurate way of positioning genetic markers along an unambiguous framework map.

Acknowledgements We are very grateful to Dr. Fred van Eeuwijk for providing valuable comments. This work was carried out under the EU FAIR programme grant: FAIR5-PL97-3565.

References

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-test-cross: mapping strategy and RAPD markers. Genetics 137:1121–1137

Isidore E, Van Os H, Andrzejewski S, Bakker J, Barrena I, Bryan GJ, Caromel B, Van Eck H, Ghareeb B, De Jong W, Van Koert P, Lefebvre V, Milbourne D, Ritter E, Rouppe van der Voort J,

Rousselle-Bourgeois F, Van Vliet J, Waugh R (2003) Toward a marker-dense meiotic map of the potato genome: Lessons from linkage group I. Genetics 165:2107–2116

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput aflp-based method for constructing integrated genetic and physical maps: progress towards a sorghum genome map. Genome Res 10:789–807

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lincoln SE, Lander ES (1992) Systematic detection of errors in genetic linkage data. Genomics 14:604–610

Nilsson NO, Säll T, Bengtsson BO (1993) Chiasma and recombination data in plants: are they compatible?. Trends Genet 9:344–348

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Stam P, Van Ooijen JW (1995) JoinMap™ version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen

Strommer J, Peters J, Zethof J, De Keukeleire P, Gerats T (2002) AFLP maps of Petunia hybrida: building maps when markers cluster. Theor Appl Genet 105:1000–1009

Van Os H, Stam P, Visser RGF, Van Eck HJ (2005) RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet (in press and aiming for back-to-back publication)

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