Euphytica **96:** 353–361, 1997.

Determining paternity in polyploids: Hexaploid simulation studies

Mario I. Buteler[1,*], Don. R. LaBonte[1] & Raúl E. Macchiavelli[2]
[1] *Department of Horticulture, Louisiana Agricultural Experiment Station, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, U.S.A.;* [2] *Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, U.S.A.; (* author for correspondence)*

Received 23 April 1996; accepted 9 April 1997

Key words: codominant markers, genealogy reconstruction, most likely male parent, polyploid, polysomic segregation

Summary

The breeding of new sweetpotato varieties is a highly inefficient process, confounded by incompatibility, poor fertility, open-pollination and hexaploidy. Upwards of 12–20 lines are combined in open pollinated nurseries based on horticulturally important characteristics. After several years of selection most progeny can be traced back to just 3 or 4 maternal lines. A method that would identify the paternal parent of superior progeny would enable breeders to combine parents that exhibit superior combining ability in more efficient, smaller nurseries. The objective of this work is to explore by means of computer simulation the application of genealogy reconstruction techniques on hexaploid individuals based on co-dominant marker data. The progeny obtained from each female parent is categorically assigned to each male with non-zero exclusion probability based on its paternity likelihood. Computer simulations show that even with polysomic segregation types, it is possible to discriminate between putative parents with few errors or mis-assignments. The number of loci scored for a 10 parent population should not be less than 20 in the case of 3 alleles per locus, and no more than 10 loci for a five allele model. An increment in the number of alleles or loci increases the discriminatory power with the number of alleles yielding a far more important effect than the number of loci. This study also demonstrates the feasibility of using simulations to determine the minimum requirements, i.e. number of loci to be genotyped, for unambiguous parentage allocation in polyploids.

Introduction

The sweetpotato is a natural hexaploid (2n = 6x = 90) (Ozias-Akins & Jarret, 1994; Magoon et al., 1970) with a disputed ancestry; both, allopolyploid and autopolyploid origins have been proposed (Magoon et al., 1970; Nishiyama et al., 1975; Shiotani, 1987). This species has complex self-incompatibility system and high cross-incompatibility (Jones et al., 1986; Martin, 1982) that make it difficult to obtain seeds from controlled crosses. Thus, open pollination by insects and mass selection are preferred in breeding programs. The breeding efficiency is low; for instance, usually less than 0.01% of the progeny from a 13–15 line crossing block are selected in the first year of evaluation in the Louisiana Agricultural Experiment Station breeding program. The majority of the extant lines are traced to just 2 or 3 of the original maternal parents. It is further likely that paternity is limited to just a few lines.

Parental selection based on horticulturally important characteristics, i.e. disease and insect resistance, yield and quality, and combining ability of the parental lines could significantly enhance the efficiency of sweetpotato breeding programs. Few sweetpotato breeding programs, with the exception of the Japanese and Chinese programs (Yamakawa, 1989), evaluate lines for combining ability by hand-crossing; in fact, most programs are financially unable to do this. Male paternity determination in open-pollinated nurseries will enable breeders to test the combining ability of breeding lines, and use this information to favor recombination among lines with high combining ability.

In plants, genealogy reconstruction from genetic data is used in natural diploid populations studies (Ellstrand, 1984; Adams et al., 1992; Devlin et al., 1992; Meagher, 1986; Meagher & Thompson, 1986, 1987); in cultivated species it has been intended for varietal protection (Wang et al., 1994). Usually, the determination of each individual's maternal parent is straightforward, mother-offspring pairs can be observed directly while collecting the seeds; only the male parent assignment is unknown. Two different approaches exist for assigning paternity. In one case the intent is a categorical answer to the paternity likelihood for each specific combinational triplet of offspring, female parent and putative male progenitor, i.e. human paternity analysis or paternity assignment in a breeding population. In contrast, no categorical response is needed in natural population studies where the objective is to discern the mating behavior (Devlin et al., 1988; Smouse & Meagher, 1994).

Polyploids, like sweetpotato, complicate parental analysis by possessing numerous copies of the same gene in a given genotype and complex meiotic behavior. Three segregation types are recognized for polyploids showing multivalent pairing at meiosis (Burnham, 1962): chromosome segregation, random chromatid segregation, and maximum equational segregation. The expected segregation ratios for genetic characters or markers are markedly different for each of them.

The basic marker requirements for paternity determination are: i) the markers should be unambiguously inherited, ii) segregate independently in the population, and iii) lead to lower levels of ambiguity than the parentage uncertainty to be solved (Smouse & Meagher, 1994). Biochemical genetic markers have been successfully used for paternity analysis in natural plant populations (Ellstrand, 1984; Meagher, 1986; Meagher & Thompson, 1986; Devlin et al., 1992) but they present some disadvantages, i.e. limited number of polymorphic markers and relatively low allele number. Molecular genetic techniques, such as Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD) or more recently Simple Sequence Repeat (SSR), provide an a priori source of numerous polymorphic DNA markers for parentage analysis and genealogy reconstruction (Lewis & Snow, 1992; Milligan & McMurry, 1993; Lynch, 1988; Hughes & Queller, 1993; Pellissier Scott & Williams, 1993). However, some difficulties are still present, with RFLP, the cost and technical complexity, and in the case of RAPD its dominant inheritance. In both cases the number of possible alleles for each loci is limited too, particularly with RAPD.

SSR markers present several advantages over RFLP or RAPD. In contrast to RAPD, SSRs are co-dominant and segregate in a Mendelian fashion (Cregan et al., 1994; Jarret & Bowen, 1994); also, uncertainty about the homology of similar sized bands is eliminated. Some of the complexities and expenses of RFLP techniques are avoided. SSR loci usually have a greater number of alleles than RFLP loci (Rafalski & Tingey, 1993; Cregan et al., 1994; Rongwen et al., 1995), are highly polymorphic and somatically stable (Cregan et al., 1994; Jarret & Bowen, 1994; Saghai Maroof et al., 1994; Rongwen et al., 1995). Several reports indicate that the high degree of polymorphism observed in SSR loci is present between individuals within a species (Cregan et al., 1994; Jarret & Bowen, 1994; Saghai Maroof et al., 1994; Rongwen et al., 1995).

SSR markers have been successfully applied on animal and plant diploid populations to determine genealogy structure and kin relationships (Morin et al., 1994; Saghai Maroof et al., 1994; Adato et al., 1995). They have been widely used in the human genome linkage map, in plant genome linkage mapping (Cregan et al., 1994), characterization of plant genetic resources (Jarret & Bowen, 1994), and in animal linkage map projects (Crawford et al., 1994). There are no known data relative to the use of SSR for parentage determination in plants; however, the advantages over other DNA markers could make them especially suitable for paternity analysis in polyploids. Indeed, preliminary results obtained using SSR on sweetpotato (data not shown) show it is possible to resolve multiple alleles following a polysomic segregation pattern.

The objectives of this study were: i) to adapt the current genetical/statistical techniques for parentage determination to be used with polyploids; ii) to develop a computer program suitable for parentage analysis of hexploids; and iii) to use computer simulation studies to determine number of loci and number of co-dominant alleles for unambiguous paternity assignment in hexaploids.

Methods

Paternity analysis

Two basic approaches are used to obtain a categorical answer for establishing paternity; both were developed for human population studies and forensic applica-

tions, and still not universally accepted in their entirety (Chakraborty et al., 1974; Thompson, 1975; Valentin, 1980; Thompson, 1986). One approach, paternity exclusion, is based on the nearly conclusive proof of non-paternity based on parent-offspring marker genotype data incompatibility (Chakraborty et al., 1988). Basically, simple exclusion compares the progeny genotype with the female parent genotype, subtracts the maternal contribution, and compares the remaining paternal gametic contribution with all putative male parent genotypes. The individuals who can not produce the paternal gametic contribution are excluded, and paternity is assigned to the remaining group; Ellstrand (1984) successfully used this approach for natural population studies. Chakraborty et al. (1988) clearly demonstrated that the exclusionary criteria alone can not solve the parentage assignment problem in natural populations. Further, they theoretically showed that a high frequency of ambiguous cases will remain, even if the number of informative markers were extended to enhance the average probability of exclusion to values close to one. As the number of loci are increased, the probability of linkage increases and undermines the basic assumption of independently segregating loci. Although the exclusionary power also depend on the allele number and frequency, more alleles do not necessarily increase the exclusionary power (Chackraborty et al., 1988).

The second approach, sometimes called the most likely parent method, calculates the paternity likelihood based on segregation (Mendelian) probabilities. Parentage is assigned to the putative parent with the maximum likelihood value, introducing a relaxation in the certainty of paternity. In the case of ties, no parent is selected. This approach allows paternity assignment to a higher number of progeny (Devlin et al., 1988; Smouse & Meagher, 1994); these authors also pointed out two limitations. First, categorical assignments are not possible for all the progeny because of ambiguous progeny genotypic profiles and redundancies among male genotypes. Second, a statistical bias in favor of homozygotes for a homozygous putative parent will always give them a higher likelihood score for a given genetic locus than a heterozygous individual. However, the bias can be reduced by increasing the number of genetic markers.

The most likely parent approach, developed and applied by Meagher (1986) and Meagher & Thompson (1986), involves the consideration of the log-likelihood scores over a restricted range because a mathematical indetermination can appear if genetic exclusions are permitted, i.e. log of a zero probability value. In actuallity, the most likely parent is chosen after applying simple exclusion, complementing this approach. Meagher & Thompson (1987) stated the likelihood-based categorical allocation is also robust under small fluctuations of allele frequencies.

For these reasons we chose the most likely parent method for the present study. The computational framework can be deduced from the following considerations developed by Meagher (1986) and Meagher & Thompson (1986): Consider an ordered triplet of genotypes (g_o, g_f, g_m) for the individuals O, F, M. The main intent is to identify the male parent given a known female progenitor; only two genealogical situations need to be considered:

(1) Relationship *A*: F is the female parent of O and M is unrelated.
(2) Relationship *B*: Both F and M are parents of O.

The triplets conditional probabilities given the true relationship $R = A$ or B are denoted as $P(g_o, g_f, g_m|R)$. Therefore,

$$P(g_o, g_f, g_m \mid A) = P(g_o \mid g_f, -)P(g_f)P(g_m)$$
$$P(g_o, g_f, g_m \mid B) = P(g_o \mid g_f, g_m)P(g_f)P(g_m)$$

where $P(g_i)$ is the probability of the genotype g_i in a random mating population and $P(g_o|g_f, -)$, $P(g_o|g_f, g_m)$ specify Mendelian probabilities. Let x_f and x_m denote for female and male gametes, respectively and $P(x_f|g_f)$, $P(x_m|g_m)$ the gamete segregations from parental genotypes with hexaploid segregation ratios. Therefore,

$$P(g_o \mid g_f, -) = \sum_{x_f} P(g_o \mid x_f)P(x_f \mid g_f)P(x_m)$$
$$P(go \mid gf, gm) = \sum_{x_f}\sum_{x_m} P(g_o \mid x_f, x_m)$$
$$P(x_f \mid g_f)P(x_m \mid g_m)$$

where $P(g_o|x_f)$ and $P(g_o|x_f, x_m)$ will take values 0 or 1 describing the offspring genotype, only segregation ratios differ based on ploidy.

The joint genotype probability given the relationship R is the likelihood of R given the set of genotypes. To compare the likelihoods of different relationships the difference in the log-likelihoods is considered (LOD scores). In the present study the hypothesized relationship is *B* since *A* is known, i.e. the female parent, so it can be used as the base-point alternative. Given the offspring genotypic information, loci are independent conditional on the parental genotypes; so the LOD scores and their means and variances are additive over independently inherited loci. The necessity of independently segregating loci requires the use

of a limited number of markers, as we noted before. Thus, the LOD score for the parent is:

$$L(B \mid g_o, g_f, g_m) = \sum_{loci} log_e \left[\frac{P(g_o \mid g_f, g_m)}{P(g_o \mid g_f, -)} \right]$$

As previously noted, the distribution of these LOD scores can not be considered over their entire range. If a genetic exclusion exists ($P(g_o|g_f, g_m) = 0$) then $L(B) = -\infty$. Since only genetically possible parents should be compared via the likelihood ratio, the probability of each triplet must be conditioned on non-exclusion of the relationship *B*. The non-exclusion probability is, like the likelihoods, multiplicative over loci; thus, the conditioning on non-exclusion can be done for each locus separately (Thompson & Meagher, 1987). Thompson & Meagher (1987) also stated that the statistic employed,

$$\Lambda(B) = log_e \left[\frac{P(B)}{P(A)} \right]$$

is sufficient for comparisons of the alternative hypotheses *B* and *A*.

Theoretical hexaploid segregation ratios

Of the three possible polysomic segregation types, chromosome segregation and maximum equational segregation provide the most extreme segregation ratios (Burnham, 1962). The segregation ratios for the chromosome type are calculated considering each chromatid derived from a different chromosome; therefore, the total number of possible gametes for each independent loci is:

$$\binom{6}{3} = 20$$

The gamete frequency distribution follows the multivariate hypergeometric distribution, thus each gamete theoretical probability can be calculated as follows:

$$\frac{\binom{M}{m}\binom{N}{n}\binom{O}{o}\binom{P}{p}\binom{Q}{q}\binom{R}{r}}{\binom{6}{3}};$$

$$m \leq M; n \leq N; o \leq O; p \leq P; q \leq Q; r \leq R;$$
$$M + N + O + P + Q + R = 6;$$
$$m + n + o + p + q + r = 3;$$
$$M, N, O, P, Q, R = 0, 1, 2, ..., 6;$$
$$m, n, o, p, q, r = 0, 1, 2, 3;$$

where *M, N, O, P, Q, R* are the total number of three possible alleles of the genotype of interest, and *m, n, o, p, q, r* the number of copies of such alleles present in the gamete.

In cases of maximum equational segregation and random chromatid segregation, gametes with alleles derived from sister chromatids can be expected for some loci. That is, if a cross over takes place between the centromere and a specific loci, multivalents are formed, and these sister chromatids go to the same pole during the first division of meiosis. Gametes with both sister chromatids are then produced. This phenomenon is called double reduction (Burnham, 1962). The frequency of double reduction is maximum when multivalents are always formed and there is always one effective cross over between the loci and the centromere, i.e. maximum equational segregation (Burnham, 1962). Under these circumstances, the maximum expected frequency of double reduced gametes for a hexaploid is $\frac{3}{10}$, while in the case of random chromatid segregation this expected frequency is $\frac{3}{11}$. Since the second value is smaller than the first, the theoretical maximum is only achieved under the maximum equational segregation. The ease of calculating random chromatid segregation frequencies, the small difference between the two values, and the specific conditions required for maximum equational segregation favored the use of the former segregation type in the present study.

Segregation ratios for the random chromatid segregation are derived by randomly sampling three chromatids from a set of twelve without replacement. As in the first segregation type, the frequencies follow the multivariate hypergeometric distribution; thus, the total number of possible gametes is:

$$\binom{12}{3} = 220$$

and the respective gamete frequencies are calculated as before by replacing the respective number of alleles with twice that value, i.e. the number of chromatids bearing each allele.

Computer algorithms

Two algorithms were developed to fulfill the objectives of this study. The first algorithm was developed to simulate a population of parents and offspring, allowing the study of real scenarios which define the number of marker loci needed as a function of the number of

alleles per loci for unambiguous parentage assignment. Since there is no information regarding the allele frequency distributions of genetic markers in polyploids, we used the uniform distribution as the simplest possible distribution scenario (Lewis & Snow, 1992). Two other basic exigencies are: i) the simulation of variable numbers of double reduced gametes, and ii) a randomization process leading to the generation of genotypes and populations without sequential correlations.

A second algorithm performed the actual paternity analysis based on the computational framework developed above, i.e., LOD scores. The algorithm has two principal routines; the first one excludes all the putative parents that show genetic incompatibility with each offspring. To perform this, all possible gametes from the female (known) and male progenitors are extracted from the respective genotypes; the female contribution is subtracted from the offspring genotype and the remainder is compared with the putative male parent contribution. The second routine calculates the likelihood of paternity for each non-excluded putative male parent.

The two algorithms were programmed in FORTRAN (Microsoft FORTRAN\Power Station 1993) to be run in a PC. Simulated populations of 10 parents and 2 offspring per parental pair were used to determine the minimum loci number needed for unambiguous parentage assignment. Each simulated population was independently generated 4 times under each condition resulting from a factorial combination of the following alternatives:

(1) Chromosome segregation with 3 and 5 alleles per loci. In simulations with 3 alleles, 12, 15, 17, 19, and 21 loci alternatives were tested; while 3, 5, 7, and 9 loci were assayed with 5 alleles. Preliminary simulations showed that 2 alleles per loci provide a resolving power incompatible with a set of loci segregating independently. The upper number of alleles per loci is compatible with the reported resolving power of microsatellites (Cregan et al., 1994).

(2) Random chromatid segregation in 2% of the loci selected at random for each run. The same number of alleles and loci alternatives as before were considered. Since there are no citations about double reduction frequencies in hexaploids, the frequency chosen was close to the highest detected using biochemical markers in tetraploid potato (Haynes & Douches, 1993); this value seems to be a safe assumption because of the cultivated potato autopolyploid nature.

(3) Equal allele frequency distribution. For the 3 allele condition each allele has a frequency = 1/3, and in the 5 allele condition each allele was present with frequency = 1/5. This type off allele frequency distribution provides the maximum discriminatory power (Meagher & Thompson, 1987).

(4) Uneven allele frequency distribution. In the case of 3 alleles/loci the frequencies selected were: 0.45 for 2 alleles and 0.1 for one allele. In the 5 alleles situation the frequencies were: 0.3 for 3 allele and 0.05 for each of the remaining two. These alternatives were set to study the discriminatory power stability in relationship with allele frequency fluctuations. Allele frequency was set at 0.05 since Ott (1992) considered this the lowest informative frequency in diploids.

Populations of 15 and 20 parents with 2 offspring per parental pair were also generated, but only 21 loci and 9 loci with 3 and 5 alleles respectively were considered. These larger parental populations were tested to determine the effect of increased parent number on paternity assignment. Parental populations of this size approximate a normal sweetpotato breeding population. The mean number of mis-assigned male parents/offspring and the mean number of non-excluded male parents/offspring, and their respective standard errors over the 4 repetitions were calculated for each different alternative. Preliminary trials suggested that 4 repetitions were sufficient to determine the main trend and simultaneously were conservative enough for a realistic minimum loci number determination.

Results and discussion

Variation in allele number

The three and five allele models were selected in our study as tenable given current SSR marker technology (Jarret & Bowen, 1994; Jarret et al., 1995). However, there are reports of polymorphisms exceeding 10 alleles/loci in several cultivated species (Saghai Maroof et al., 1994; Rongwen et al., 1995). It is unlikely to find in a breeding population a uniform number of alleles across loci. In fact, the simulation results illustrate the possible discriminatory power that could be achieved with the most likely parent method. Even with 3 alleles per loci, a low number for a hexaploid, and 21 loci, it was possible to identify the male-female-offspring triplet with less than 0.003 mis-assigned male parent/offspring without double reduction (Figure 1). The

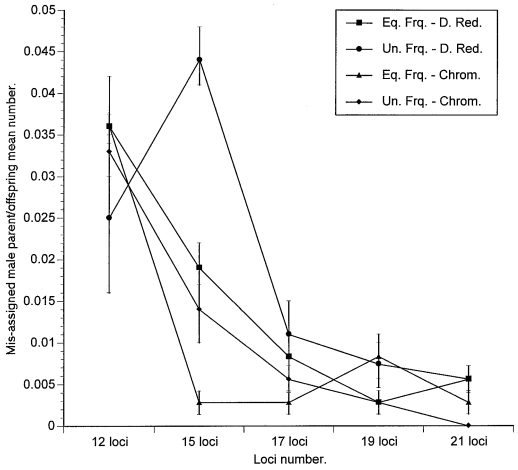

Figure 1. Relationship between loci number genotyped with 3 allele and mis-assigned male parent/offspring mean number (± S.E.M.) over 4 replications, in a 10 parent and 90 offspring population. (Eq. Frq = even allele frequency distribution: 1/3, 1/3, 1/3; Un. Frq. = skewed allele frequency distribution: 0.45, 0.45, 0.10; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

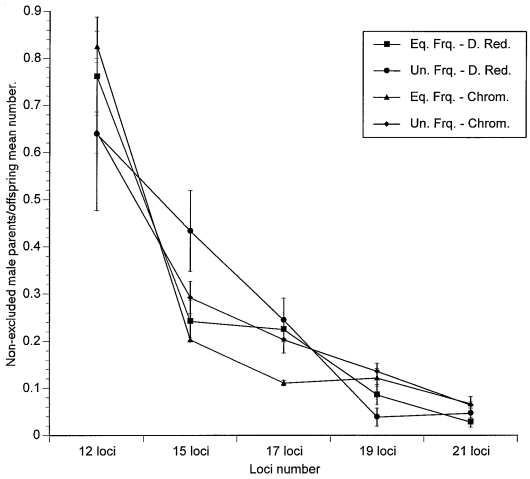

Figure 2. Relationship between loci number genotyped with 3 allele and non-excluded male parent/offspring mean number (± S.E.M.) over 4 replications, in a 10 parent and 90 offspring population. (Eq. Frq = even allele frequency distribution: 1/3, 1/3, 1/3; Un. Frq. = skewed allele frequency distribution: 0.45, 0.45, 0.10; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

mis-assignments were less than 0.006 mis-assigned male parent/offspring, and the variability of the estimated mean increased (standard error of the mean (S.E.M.) = 0.0016 vs 0.0014) when double reduction was introduced (Figure 1). The uneven allele frequency distribution combined with random chromatid segregation and 21 loci produced the same result as random chromatid segregation with even frequency distribution (Figure 1). Moreover, the skewed allele frequency distribution combined with the chromosome segregation alternative produced no mis-assignments with 21 loci (Figure 1).

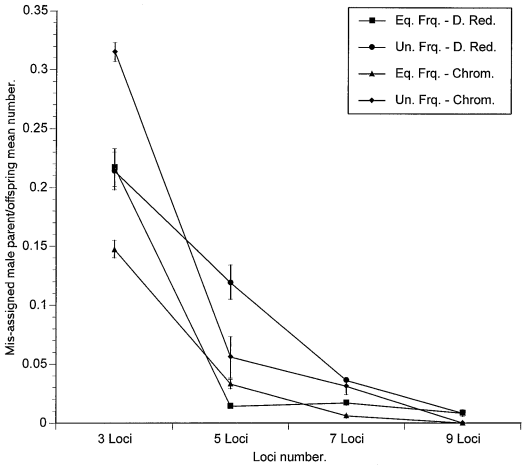

Figure 3. Relationship between loci number genotyped with 5 allele and mis-assigned male parent/offspring mean number (± S.E.M.) over 4 replications in a 10 parent and 90 offspring population. (Eq. Frq = even allele frequency distribution: 0.2, 0.2, 0.2, 0.2, 0.2; Un. Frq. = skewed allele frequency distribution: 0.3, 0.3, 0.3, 0.05, 0.05; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

It is important to stress the greater discriminatory power of the most likely parent compared with the simple exclusion, despite the bias pointed out by Chakraborty et al. (1988) and Devlin et al. (1992) for the former methodology. Our results show with 21 loci there were consistently more non-excluded parents (Figure 2) than mis-assigned parents (Figure 1); but, concomitantly the dispersion of the estimates tended to be narrower than with the most likely parent method. Briefly, in a 10 parent population the number of loci needed to achieve a reasonable discriminatory power must exceed 20, considering a relatively low allele number, skewed alleles frequencies, and double reduction. The number of offspring in the population had no effect since the most likely parent technique does not use the genotypic information jointly.

The simulations using 5 alleles per loci produced results similar to those obtained with 3 alleles per loci (Figures 3, 4). The only difference was a complete triplet reconstruction with the 5 allele model and 9 loci, i.e., the number of mis- assigned parents = 0, if double reduced gametes were not produced (Figure 3).

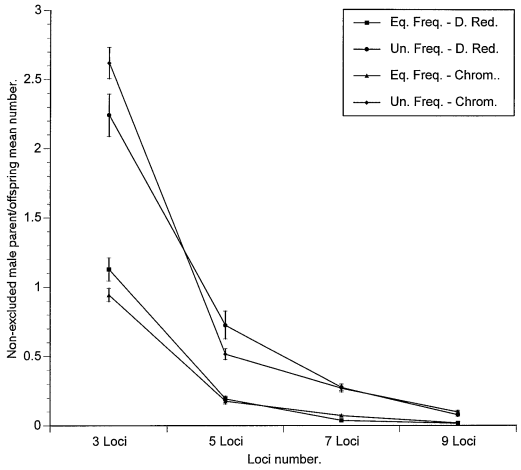

Figure 4. Relationship between loci number genotyped with 5 allele and non-excluded male parents/offspring mean number (± S.E.M.) over 4 replications, in a 10 parent and 90 offspring population. (Eq. Freq. = even allele frequency distribution: 0.2, 0.2, 0.2, 0.2, 0.2; Un. Frq. = skewed allele frequency distribution: 0.3, 0.3, 0.3, 0.05, 0.05; D. Re = random chromatid segregation type; Chrom = chromosome segregation type).

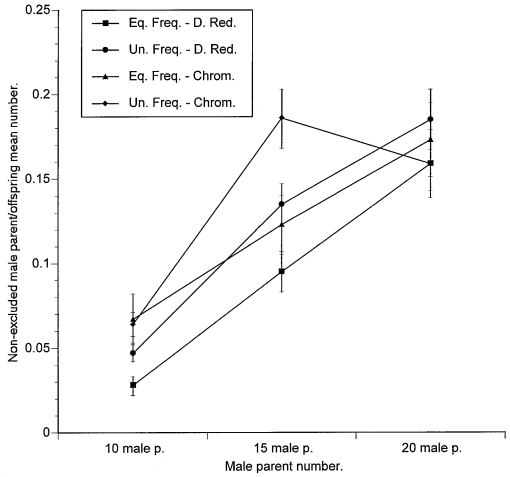

Figure 6. Relationship between parent number and non-excluded male parents/offspring mean number (± S.E.M.) over 4 replications, for 21 loci genotyped with 3 allele. (Eq. Frq = even allele frequency distribution: 1/3. 1/3, 1/3; Un. Frq. = skewed allele frequency distribution: 0.45, 0.45, 0.10; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

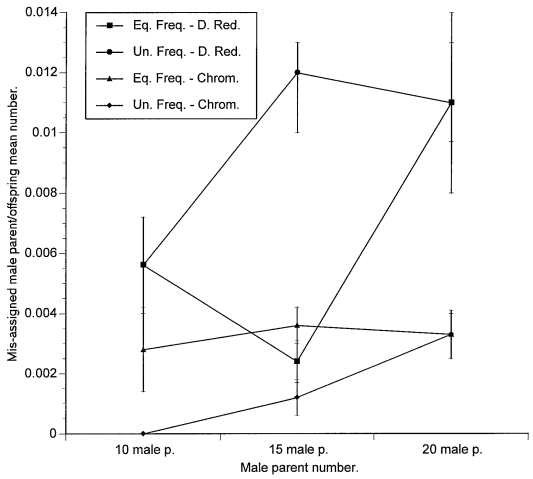

Figure 5. Relationship between parent number and mis-assigned male parent/offspring mean number (± S.E.M.) over 4 replications, for 21 loci genotyped with 3 allele. (Eq. Frq = even allele frequency distribution: 1/2, 1/3, 1/3; Un. Frq. = skewed allele frequency distribution: 0.45, 0.45, 0.10; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

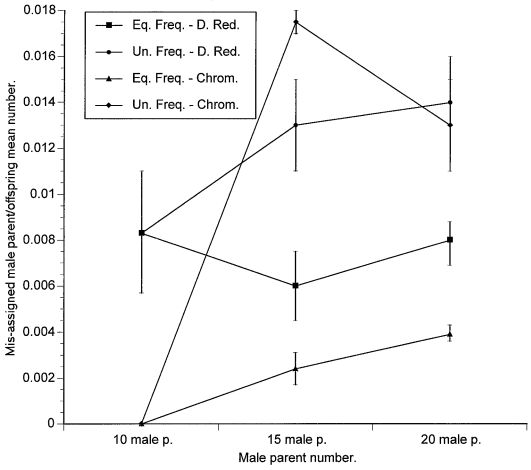

Figure 7. Relationship between parent number and mis-assigned male parent/offspring mean number (± S.E.M.) over 4 replications, for 9 loci genotyped with 5 allele. (Eq. Frq = even allele frequency distribution: 0.2, 0.2, 0.2. 0.2, 0.2; Un. Frq. = skewed allele frequency distribution: 0.3, 0.3, 0.3, 0.05, 0.05; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

Another important consequence of the allele number increment was a consistent reduction in the variance of the estimated means in all scenarios (Figures 1–4).

On the other hand, the simple exclusion method with 5 alleles did not allow the reconstruction of all the triplets in any scenario (Figure 4). Compared with the 3 allele alternative, the reduction in the optimum loci number with 5 alleles was not proportional to the allele number increment (9 vs 21). As before, double reduced gametes did not allow paternity assignment to the progeny without some error, e.g., 0.0083 (S.E.M. = 0.0026) mis-assigned male parent per offspring, 9 loci, equal and skewed allele frequencies, and random chromatid segregation type vs. 0.0 mis-assigned male par-

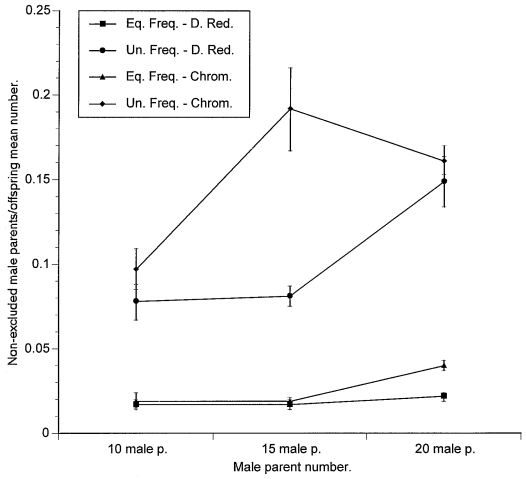

Figure 8. Relationship between parent number and non-excluded male parents/offspring mean number (± S.E.M.) over 4 replications for 9 loci genotyped with 5 allele. (Eq. Frq = even allele frequency distribution: 0.2, 0.2, 0.2, 0.2, 0.2; Un. Frq. = skewed allele frequency distribution: 0.3, 0.3, 0.3, 0.05, 0.05; D. Re. = random chromatid segregation type; Chrom = chromosome segregation type).

ent per offspring, equal and skewed allele frequency, random chromosome segregation.

In summary, both allele number examples showed that it is possible to determine *a priori* the minimum number of loci needed to discriminate among paternal parents.

Variation in parent number

Figure 5, mis-assigned male parent/offspring, and Figure 6, non-excluded male parent/offspring, show the results of the increment in the parental population size for the 3 allele and 21 loci alternatives. The trends were similar for a reduction in the loci number, an increment in the variability of the estimates, and in the number of mis-assignments. The skewed allele frequency and the presence of double reduced gametes exacerbated these results. The 5 allele, 9 loci situation produces similar effects but less pronounced (Figure 7, mis-assigned parents; Figure 8, non-excluded parents). The results also show a difference in slopes between the 3 and the 5 allele scenarios for non-excluded parents per offspring (Figures 6, 8). In the first scenario the variable steadily increase with the parent number, while in the second the increment is less pronounced. This result is a consequence of the less discriminatory power of a reduced allele number per loci.

Conclusions

The present study demonstrated the feasibility of applying parental analysis techiques to polyploids using codominant markers. Also, it shows that, in spite of the drawbacks the likelihood based categorical parental allocation has, it is powerful enough to allocate without error all the parental set in a breeding population provided there are enough loci genotyped bearing more than three alleles (Figure 3). Although the presence of double reduction produces male parent mis-assignment, this does not strongly bias the results. It is important to note that the selection of highly polymorphic loci, i.e., large allelic families, is a prerequisite with polyploids to obtain high discriminatory power, especially in the case of large breeding populations.

We also demonstrated that it is possible to use a genotype simulation program to project results for specific conditions. We are now applying this model to SSR marker data for calibration.

References

Adams, W.T., A.R. Griffin & G.F. Moran, 1992. Using paternity analysis to measure effective pollen dispersal in plant populations. Am Nat 140: 762–780.

Adato, A., D. Sharon, U. Lavi, J. Hillel & S. Gazit, 1995. Application of DNA fingerprints for identification and genetic analysis of mango (*Mangifera indica*) genotypes. J Am Soc Hort 120: 259–264.

Burnham, C.R., 1962. Discussions in Cytogenetics. Burguess Publishing Co. Minnesota, USA. 375 p.

Chakraborty, R., T.R. Meagher & P.E. Smouse, 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118: 527–536.

Chakraborty, R., M. Shaw & W.J. Schull, 1974. Exclusion of paternity: the current state of the art. Am J Hum Genet 26: 477–488.

Crawford, A.M., G.W. Montgomery, C.A. Pierson, T. Brown, K.G. Dodds, S.L.F. Sunden, H.M. Henry, A.J. Ede, P.A. Swarbrick, T. Berryman, J.M. Penty & D.F. Hill, 1994. Sheep linkage mapping: nineteen linkage groups derived from the analysis of paternal half-sib families. Genetics 137: 573–579.

Cregan, P.B., A.A. Bhagwat, M.S. Akkaya & J. Rongwen, 1994. Microsatellite fingerprinting and mapping in soybean. Methods Mol Cell Biol 5: 49–61.

Devlin, B., K. Roeder & N.C. Ellstrand, 1988. Fractional paternity assignment: theoretical development and comparison to other methods. Theor Appl Genet 76: 369–380.

Devlin, B., J. Clegg & N.C. Ellstrand, 1992. The effect of flower production on male reproductive success in wild radish populations. Evolution 46: 1030–1042.

Ellstrand, N.C., 1984. Multiple paternity within the fruits of the wild radish, *Raphanus sativus.* Am Nat 123: 819–828.

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

Hughes, C.R. & D.C. Queller, 1993. Detection of highly polymorphic microsatellite loci in species with little allozyme polymorphisms. Mol Ecol 2: 131–138.

Jarret, R.L. & N. Bowen, 1994. Simple Sequence Repeats (SSRs) for sweetpotato germplasm characterization. Plant Genet Res Lett 100: 9–11.

Jarret, R.L., N. Bowen, S. Kresovich & Z. Liu, 1995. Simple sequence repeats (SSRs) in sweetpotato (*Ipomoea batatas* (L.) Lam.). HortScience 30: 841.

Jones, A., P.D. Dukes, J.M. Schalk, 1986. Sweet potato breeding. In: M.J. Basset (Ed.), Breeding Vegetable Crops, pp. 1–35. The AVI Publishing Company, Inc., USA.

Lewis, P.O. & A.A. Snow, 1992. Deterministic paternity exclusion using RAPD markers. Mol Ecol 1: 155–160.

Lynch, M., 1988. Estimation of relatedness by DNA fingerprinting. Mol Biol Evol 5: 584–599.

Magoon, M.L., R. Krishnan & K. Vijaya Bai, 1970. Cytological evidence on the origin of sweet potato. Theor Appl Genet 40: 360–366.

Martin, F.W., 1982. Analysis of the incompatibility and sterility of sweet potato. In: R.L. Villareal & T.D. Griggs (Eds.), Sweet Potato, pp. 275–283. Proceedings of the First International Symposium. AVRDC, 1982, Taiwan.

Meagher, T.R., 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of most-likely parents. Am Nat 128: 199–215.

Meagher, T.R. & E. Thompson, 1987. Analysis of parentage for naturally established seedlings of *Chamaelirium luteum* (Liliaceae). Ecology 68: 803–812.

Meagher, T.R. & E. Thompson, 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor Popul Biol 29: 87–106.

Milligan, B.G. & C.K. McMurry, 1993. Dominant vs. codominant genetic markers in the estimation of male mating success. Mol Ecol 2: 275–283.

Morin, P.A., J.J. Moore, R. Chakraborty, L. Jin, J. Goodall & D.S. Woodruff, 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265: 1193–1201.

Nishiyama, I., T. Miyazaki & S. Sakamoto, 1975. Evolutionary autoploidy in the sweet potato (*Ipomoea batatas* (L.) Lam.) and its progenitors. Euphytica 24: 197–208.

Ott, J., 1992. Analysis of human genetic linkage. Rev Ed John Hopkins University Press, USA.

Ozias-Akins, P. & R.L. Jarret, 1994. Nuclear DNA content and ploidy levels in the genus *Ipomoea*. J Am Soc Hort Sci 119: 110–115.

Pellissier Scott, M. & S.M. Williams, 1993. Comparative reproductive success of communally breeding burying beetles as assessed by PCR with randomly amplified polymorphic DNA. Proc Natl Acad Sci USA 90: 2242–2245.

Rafalski, J.A. & S.V. Tingey, 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. TIG 9: 275–279.

Rongwen, J., M.S. Akkaya, A.A. Bhagwat, U. Lavi & P.B. Cregan, 1995. The use of microsatellites DNA markers for soybean genotype identification. Theor Appl Genet 90: 43–48.

Saghai Maroof, M.A., R.M. Biyashev, G.P. Yang, Q. Zhang & R.W. Allard, 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91: 5466–5470.

Shiotani, I., 1987. Genomic structure and the gene flow in sweet potato and related species. In: Exploration, Maintenance and Utilization of Sweet Potato Genetic Resources, pp. 61–73. Report of the First Sweet Potato Planning Conference. CIP, 1987, Lima, Peru.

Smouse, P.E. & T.R. Meagher, 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). Genet 136: 313–322.

Thompson, E.A., 1975. The estimation of pairwise relationship. Ann Hum Genet 39: 173–188.

Thompson, E.A., 1986. Pedigree analysis in human genetics. John Hopkins University Press, Baltimore, MD, USA. 213 p.

Thompson, E.A. & T.R. Meagher, 1987. Parental and sib likelihoods in genealogy reconstruction. Biometrics 43: 585–600.

Valentin, J., 1980. Exclusions and attributions of paternity: practical experiences of forensic genetics and statistics. Am J Hum Genet 32: 420–431.

Wang, G., S. Castiglione, J. Zhang, R. Fu, J. Ma, W. Li, Y. Sun & F. Sala, 1994. Hybrid rice (*Oryza sativa* L.): identification and parentage determination by RAPD fingerprinting. Plant Cell Rep 14: 112–115.

Yamakawa, O., 1989. Annual report of sweet potato breeding. Sweet Potato Breeding Laboratory, Kyushu National Agricultural Experiment Station, Ministry of Agriculture, Forestry and Fisheries, Japan.