Theor Appl Genet (2003) 106:1307–1315
DOI 10.1007/s00122-003-1200-9

V. Douhovnikoff · R. S. Dodd

Intra-clonal variation and a similarity threshold for identification of clones: application to Salix exigua using AFLP molecular markers

Received: 1 July 2002 / Accepted: 17 October 2002 / Published online: 20 February 2003

Abstract Although molecular methods are a major advance over phenological or root connectivity studies in the identification of clonal plants, there is still a level of ambiguity associated with two types of error: misidentification of genetically similar seedlings as clones and misidentification of dissimilar fingerprints from clones as genetically distinct individuals. We have addressed the second of these error types by determining the level of variation for AFLP fingerprints in *Salix exigua*, and then by developing a threshold value of Jaccard's similarity index for assigning individuals to clones or to siblings. Variation in AFLP banding patterns among clones was partitioned into three potential sources; clones, stems within-clones and foliage within-stems. Most of the variation was attributable to clones and then to stems within-clones. To provide an objective means of identifying clones, we developed a method for establishing a threshold similarity index to assign individuals to the same clone. Our method yielded a Jaccard similarity threshold of 0.983 that resulted in a potential pairwise error rate of 8.1% putative clone assigned to siblings and 1.5% sibling assigned to clones. The method was tested on independent clonal and sibling individuals resulting in the same threshold value and similar error rates. We applied our method to assign individuals to clones in a population of *S. exigua* along the Cosumnes River, California. A total of 11 clones were identified, with one clone including 43% of the individuals sampled. Our results show that this approach can be useful in the accurate identification of clones.

Keywords Clones · Siblings · Similarity threshold · *Salix exigua* · AFLP

Communicated by D.B. Neale

V. Douhovnikoff (✉) · R. S. Dodd
Department of Environmental Science, Policy and Management,
145 Mulford Hall, University of California at Berkeley,
California 94720-3114, USA,
e-mail: vlad@nature.berkeley.edu
Tel.: +1-510-6431635
Fax: +1-510-6423242

Introduction

It is tacitly assumed in many ecological theories that woody plants are non-clonal (Klimeš et al. 1997), with the result that each stem is considered a genetically and physically distinct unit. However, it has become increasingly apparent that clonal growth plays an important role in the ecology of many woody plant populations (Van Groenendael et al. 1996; Price and Marshall 1999). Klimeš et al. (1997) found that of 2,760 central European plant species studied 65.5% could be described as clonal, and Peterson and Jones (1997) compiled a partial list of representative clonal woody plants that included 86 species from 29 families of both gymnosperms and angiosperms. Therefore, in order to improve our understanding of woody plant ecology in areas such as demography, genetic diversity, gene flow, colonization, recruitment and evolutionary potential, the quantification of clonal structure in natural populations is becoming an important field of ecological investigation, necessitating an accurate means of identifying clones.

Until the recent development of molecular methods, ecologists were limited to identifying clones by root connectivity and by similarities in plant phenology and morphology. However, excavation, to expose roots, is not effective over large areas due to the high cost, root fragmentation and root grafting. Whereas root fragmentation removes the physical evidence of clonality, root grafting may connect non-clonal genets. In the dioecious genus *Salix*, it is not unusual to find genets of the opposite sex connected by root structures resulting from natural grafting. Phenological and morphological similarities are difficult to test for accuracy, are unreliable due to plasticity and require many hours of on-site observations over an extended period of time. Molecular genetic methods, on the other hand, allow for the identification of clones independent of environmental variation, can be easily tested for accuracy and reproducibility, and can give results after a single collection of vegetative material from the field.

Increasingly, studies are using Amplified Fragment Length Polymorphisms (AFLPs) for the identification of plant clones (Arens et al. 1998; Escaravage et al. 1998; Winfield et al. 1998; Pornon et al. 2000; Suyama et al. 2000; van der Hulst et al. 2000; Cabrita et al. 2001). AFLPs can reveal large numbers of markers with a high degree of reproducibility, overcoming problems inherent with isoenzyme and RAPD methods, and so are particularly suitable for surveys of fingerprint data.

However, as with all genetic fingerprinting methods, two potential sources of error exist. The first is the risk of classifying genetically similar siblings as clones. This problem has received some attention, particularly in the forensic literature, and probabilities of identity (P_{ID}) based on multilocus fingerprints have been developed for animal species, but these are rarely used with plants (Mills et al. 2000; Waits et al. 2001). Waits et al. (2001) warn that the use of such probabilities may be risky, because theoretical estimates of P_{ID} are consistently lower than observed values, and recommend computing P_{ID} between sibs to serve as a conservative upper bound for the probability of observing identical genotypes. If multilocus markers are not adequately polymorphic, siblings could have identical fingerprints. Therefore, a cluster of siblings might be misidentified as a clone.

The second source of error is the risk of assuming that individuals with non-identical fingerprints are genetically distinct individuals, when in fact they are clones. AFLP fingerprints are commonly not 100% identical for two samples from the same plant or genet (Tohme et al. 1996; Arens et al. 1998; Winfield et al. 1998; van der Hulst et al. 2000; Cabrita et al. 2001). This becomes particularly important when highly polymorphic AFLP fragments are used to maximize the sensitivity of the analysis for separation of sibs and clones. Differences between any two AFLP fingerprints can be considered to lie on a continuum where clones have few differences, and genetically distant individuals have many. The problem is to determine where along this continuum one can draw the threshold between non-identical clonal fingerprints and nearly identical sibling fingerprints. Thresholds have been used in animal systems to distinguish between half- and full-sib families (Lehmann et al. 1992; Gompper et al. 1997).

In the context of our broader work on the population ecology of riparian zones, we are interested in the dynamics and evolution of populations of narrow-leaved willow (*Salix exigua*) under varying levels of disturbance. This requires knowledge of the extent and spatial pattern of clonal groups. Narrow-leaved willow is a riparian plant of shrub to tree form, common to western North America from California to Alaska. It can grow in dense aggregations of hundreds of stems and readily forms clonal patches from root suckers. Single stems can exceed 8 m in height and 15 cm in diameter. From preliminary work, we found evidence that clones of *S. exigua* commonly have non-identical AFLP fingerprints, necessitating the development of an objective means of setting a threshold for pair-wise similarities of multilocus genotypes to distinguish putative clones from non-clonal individuals. To address this problem, we first determined the level of variation in AFLP fingerprints among and within clones of this species. We also set out to determine whether differences in fingerprints between members of the same clone were due to errors associated with the AFLP procedure, or whether part of the variation was inherent within the clones. We then developed a method to determine a similarity threshold for separating clones from non-clones, by equating the overlapping tails of the two distributions of pair-wise similarities for clones and for sibling plants raised from seed. The usefulness of this threshold was then tested with an independent set of clones and greenhouse-grown seedlings. Finally we applied the method to a field site on the Cosumnes River in California in which clonal structure was unknown.

Materials and methods

Plant material

Study sites and populations

Within-clone samples 1. We selected 11 well-separated, small putative clone clusters of stems along a 700 m stretch of the Cosumnes River in California. A portion of the roots of each cluster was excavated and inspected, to infer whether the stems that were selected were part of a single clone. Excavated samples were meticulously studied for any signs of root grafting. Even though root grafting cannot be excluded as a source of their physical connection, we believe the possibility of this was minimized. We were unable to verify the full extent of any single clone and as a result it is possible that some of our clone clusters were part of a larger clonal group. Within each cluster, three clonal stems were selected for sampling. Fresh foliage, with no evidence of infections, was collected from single branches of the upper stems of 33 individuals across 11 putative clones for DNA extraction and AFLP analysis. We sampled foliage from each stem three times to provide replication of our laboratory procedures.

Family population 1. For the sibling samples, we collected open-pollinated aments from female plants along the same riparian zone of the Cosumnes river as the within-clone samples 1, and a second set was collected from the adjacent Mokelumne river watershed. We sampled along two rivers to capture potential differences in population genetic structure. Seeds were extracted directly from these aments and allowed to germinate in a greenhouse. The resulting seedlings, identified by family (mother tree), were grown in the greenhouse until large enough to be sampled. Leaves were collected from 36 seedlings from five families for AFLP analysis. Although the family type is unknown from this sampling procedure, our collection methodology should capture proportionate samples of full sibs present in the collection areas.

Family population 2 and clonal group 2. In order to test the ability of the threshold to distinguish clones from siblings with limited error, we repeated the above analysis on an independent set of clones and sibs. Clones included 20 samples taken from five clonal clusters located directly upstream of the within-clone samples 1, and were again identified by root connections. The sibs included 20 seedlings from two families collected along the Cosumnes and Mokelumne rivers.

Application population. Finally, our method was used to evaluate clonal structure in a field site on a point bar along the Cosumnes River. We collected samples from 49 stems in 12 adjacent clusters.

Samples were collected when aments were available, allowing for the identification of each stem's sex. Due to the size of the site it was not possible to identify clones by digging up the root systems, and as a result it was impossible to establish "known" clones for testing purposes.

DNA isolation and AFLP analyses

DNA extraction

DNA was extracted according to the Cullings (1992) modification of Doyle and Doyle (1987). DNA concentrations were established by electrophoresis on agarose gels and comparisons with DNA lambda standards of known concentration.

AFLP analysis

The amplified fragment length polymorphism (AFLP) method developed by Vos et al. (1995) was performed with the following modifications: restriction digestion and ligation were performed simultaneously in a 50 μl solution containing 250 ng of genomic DNA, 5 U of *Eco*RI, 5 U of *Mse*I, 5 μl of 10 × restriction–ligation buffer (100 mM of Tris-Acetate, 100 mM of Mg-acetate, 500 mM of K-acetate, 50 mM of DTT), 1 U of T4 DNA ligase, 0.2 mM of ATP, 1.0 μM of *Mse*I adapter and 0.1 μM of *Eco*RI adapter. The restriction-ligation reaction was incubated for 4 h at 37 °C, then diluted to 200 μl with 1 × TE. Preamplification was performed in a 25-μl solution containing 2.5 μl of diluted restriction–ligation product, 0.2 mM of dNTPs, 0.3 μM of each primary amplification primer, 2.5 μl of 10 × PCR buffer (100 mM of Tris–HCl, 500 mM of KCl, 20 mM of $MgCl_2$, 13 mg/ml of BSA), and 0.5 U of *Taq* polymerase. For the primary amplification primers, the *Eco*RI primer was identical to the adapter sequence, whereas the *Mse*I primer had an extra "C" as a selective nucleotide. The PCR reaction was performed on a Techne Genius thermocycler for 28 cycles using the following cycling parameters: 30 s at 94 °C, 60 s at 60 °C and 60 s at 72 °C. The primary amplification product was then diluted to 250 μl with 1 × TE. Selective amplification was performed in a 25-μl solution containing 6.25 μl of diluted primary amplification product, 0.2 mM of dNTPs, 0.06 μM of *Eco*RI fluoresced selective primer, 0.3 μM of *Mse*I selective primer, 2.5 μl of 10 × buffer and 0.5 U of *Taq* polymerase. We pre-screened 32 selective primer pairs and chose three pairs that were reliable and highly polymorphic for this study (*Mse*I-CCAA/*Eco*RI-GTA, *Mse*I-CTC/*Eco*RI-TAC and *Mse*I-CGTG/*Eco*RI-GTA). The selective PCR reaction had two cycle sets: 13 cycles of 30 s at 94 °C, 30 s at 65 °C (annealing temperature was lowered 0.7 °C at each cycle) and 60 s at 72 °C, followed by 18 cycles of 30 s at 94 °C, 30 s at 56 °C and 60 s at 72 °C. Fingerprint data were obtained by running the amplified samples on an ABI Prism 377 DNA Sequencing System using PE Applied Biosystems protocols. Band scoring was completed with the Genescan and Genotyper software (PE Applied Biosystems, Foster City, Calif., USA).

Statistical analyses

Clonal population

The distribution of genetic variance among clones, among stems-within-clones and among foliage samples-within-stems was investigated by hierarchical AMOVA using WINAMOVA version 1.55 software (Excoffier et al. 1992). Input files of dominant AFLP markers were prepared using AMOVA-PREP (Miller 1998). As the Jaccard coefficient is not available as a metric in AMOVA-PREP, we used the simple matching coefficient, which is the proportion of matching phenotypes between two samples and includes presence and absence of bands, and was used as the similarity metric. Significances of the covariance components associated with each level in the AMOVA hierarchy were estimated by distributions generated from 1,000 random permutations. We also ran a principal components analysis on the full data set and tested the first ten vectors for significance attributable to the three sources of variation by analysis of variance. The procedure GLM in SAS (SAS Inc.) was used to estimate significance of *F*-ratios associated with the three levels (variance among clones, variance among stems-within-clones and variance among foliage samples-within-stems) in a nested analysis of variance. Variance components associated with these three levels were then estimated using the Type 1 method in the VARCOMP procedure in SAS. This method equates mean squares of the random effects with their expected values and is appropriate here, as all levels are treated as random effects.

Determining the similarity threshold from the clonal population and family population 1

To determine an appropriate threshold that can distinguish between non-identical clones and very similar siblings, the clonal population and siblings from family population 1 were compared. Jaccard similarity values were calculated between pairs of individuals within each clonal group and between pairs of siblings within families, but from the same river system (Sneath and Sokel 1973). We selected the Jaccard coefficient because it does not consider absence of a band in two samples as a match.

The frequency distributions of the pair-wise similarity values for clones and for seedlings were tested for normality using the Kolmogorov-Smirnoff test for goodness of fit of the UNIVARIATE procedure in SAS. Following this test, the two distributions did not depart from normality, so they were treated in further analyses as two normally distributed overlapping curves. The threshold for identifying clones from seedlings was therefore determined by equating the lower tail of the clonal distribution with the upper tail of the seedling distribution in the following way:

$$\frac{T-\mu_s}{\sigma_s}=\frac{\mu_c-T}{\sigma_c} \quad (1)$$

and

$$T=\frac{\sigma_c\mu_s+\partial\mu_c}{\sigma_s+\sigma_c} \quad (2)$$

where T is the threshold, μ_s, μ_c and σ_s, σ_c are the means and standard deviations of the sibling and clonal distributions of pair-wise similarities respectively.

Family population 2 and clonal group 2

In a replication of the method, the threshold for distinguishing between siblings and clones was determined for an independent group of siblings and clones in the same way as above.

Application population

Samples were assigned to the same clone if their pair-wise similarities were greater than the threshold determined from the clonal population and family population 1 above. Several tests were performed to evaluate the likely success of our method in identifying clones. First, we performed a UPGMA cluster analysis on a matrix of Jaccard similarities and compared these clusters with those obtained using the threshold approach. Then we compared the clonal structure based on our threshold approach with the spatial pattern of samples at the field site by means of Mantel tests. The Euclidean distance matrix computed from the point coordinates for all individuals was tested first, with a matrix of Jaccard similarities, and second, with a binary matrix in which pairs of individuals assigned to the same clone by our method were assigned the value 1 and all non-clone pairs were assigned the value 0. Significance of the correlations was tested by 1,000 random permutations. Genetic clusters were also checked for consistency with sex.

Table 1 Partition of variation of AFLP fingerprints among clones, stems-within-clones and foliage samples-within-stems in *S. exigua* from hierarchical AMOVA and from ANOVA of principal components vectors extracted from the full AFLP data set. The first seven principal components vectors are denoted as PC1–PC7. Variance components for ANOVA estimated using the Type 1 method in the VARCOMP procedure in SAS

Test		Variance components % of total			Probability		
		Clones	Stems (clones)	Fol (stems)	Clones	Stems (clones)	Fol (stems)
AMOVA		76.5	14.5	9.0	0.001	0.001	0.001
ANOVA							
Eigenvalue							
% of total							
PC1	0.25	89.4	9.2	1.4	0.0001	0.0001	0.83
PC2	0.09	0.7	57.4	41.9	0.0001	0.0001	0.28
PC3	0.08	94.2	5.8	0	0.0001	0.0003	0.23
PC4	0.06	81.7	18.3	0	0.0001	0.06	0.62
PC5	0.05	85.7	14.3	0	0.0001	0.07	0.37
PC6	0.04	78.7	15.8	5.5	0.0001	0.07	0.19
PC7	0.03	22.5	0	77.5	0.002	0.56	0.37

Results

We selected 177 polymorphic amplified fragments out of a total of 1,144. These were chosen in an initial screening across siblings, open-pollinated families and stems. Only well-defined fragments were selected, among which the most polymorphic fragments were chosen to maximize possible variation within clones. In this screening set fragments selected had a mean frequency of 0.560 and a variance of 0.083.

Variation within clones

No two samples were found to have identical fingerprints. Mean Jaccard similarity for foliage samples within stems was 0.990 with a minimum of 0.984. Pair-wise similarities for stems-within-clones were slightly lower with a mean of 0.980 and a minimum of 0.972. Hierarchical AMOVA showed most of the variance to be attributable to clones (76.5%) with the balance partitioned into stems-within-clone (14.5%) and foliage samples-within-stems (9.0%) (Table 1). Variance components for the three sources of variation were significant at the 0.001 probability level. The first ten principal components vectors explained 67% of the variance in the data set, but only the first six showed significance attributable to one of the three sources of variation (Table 1). Results from the GLM procedure of SAS revealed clonal variation to be significant for all six principal components vectors, stems-within-clones was significant for the first three vectors and foliage samples-within-stems was non-significant in all cases. The first principal components vector explained 25% of the variance, of which 89% was attributable to clones, 9% was attributable to stems-within-clones and 1.4% to foliage samples-within-stems following the VARCOMP procedure of SAS.

Similarity threshold

Within the sibling families a total of 331 pairwise similarity values were generated, ranging from 0.931 to 0.984 with a mean of 0.971 and a standard deviation of 0.009. Within the clonal groups a total of 99 pairwise similarity values were generated, ranging from 0.981 to 0.995 with a mean of 0.988 and a standard deviation of 0.004. These two distributions overlapped so that some pair-wise similarities among seedlings were greater than pair-wise similarities among some of the clones. We established our boundary between the two groups at the point at which the tails of the standardized distributions were equal. This yielded a similarity threshold of 0.983. Based on this threshold, eight (8.1%) clonal pairwise similarity values were lower than this threshold and five sibling pairwise similarity values (1.5%) were above the threshold (Fig. 1).

Application of the threshold to an independent set of siblings and clones

In the independent test set, within-family comparisons generated a total of 90 pairwise similarity values, ranging from 0.948 to 0.986 with a mean of 0.971 and a standard deviation of 0.008. Within the clonal groups 64 pairwise comparisons were generated, ranging from 0.981 to 0.994 with a mean of 0.989 and a standard deviation of 0.004. The calculation of T once again resulted in a similarity threshold of 0.983. Applying the threshold of 0.983, four clonal pairwise similarity values were below the threshold (6.2%), and three sibling pairwise similarity values were above the threshold (3.3%).

Application of the threshold on a field site

Out of the 49 samples tested, we detected 11 unique clones ranging in size from one to 17 stems, using our

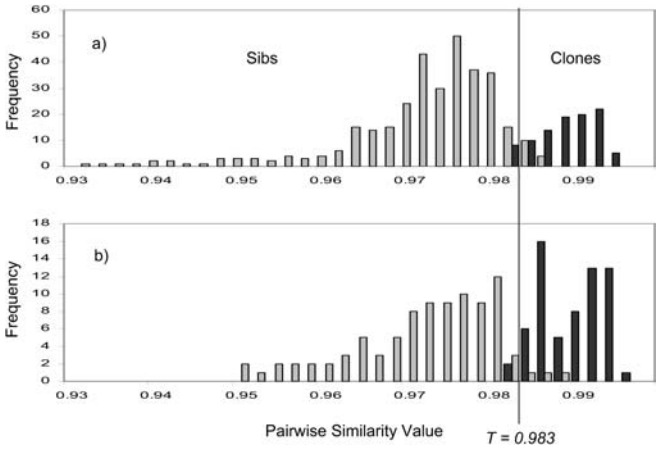

Fig. 1a, b Frequency histograms of pairwise Jaccard similarities for open-pollinated siblings (*grey bars*) and for putative clones (*solid bars*) of *S. exigua*. The *vertical bar* shows the estimated similarity threshold for distinguishing siblings and clones. **a**) Clone samples 1 from the Cosumnes River, California, and sibling families 1 from the Cosumnes and Mokelumne rivers (combined), California. **b**) Independent test clone samples 2 from the Cosumnes River and sibling families 2 from the Cosumnes and Mokelumne rivers (combined)

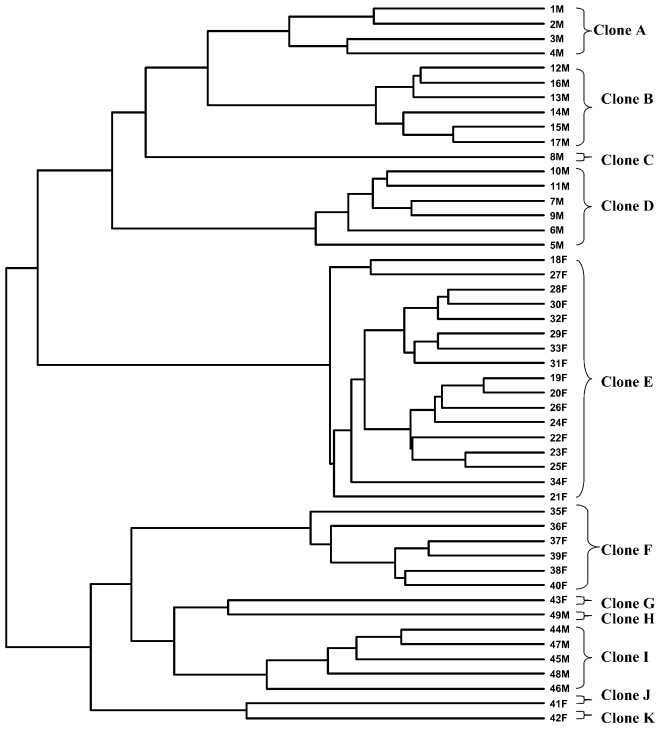

Fig. 2 UPGMA cluster analysis based on Jaccard similarities of 177 AFLP molecular markers among 49 individuals of *S. exigua* from a riparian zone along the Cosumnes River, California. Individuals were arbitrarily coded 1 to 49 and were identified as male (*M*), or female (*F*). Assignments to clones were based on a threshold level for the Jaccard similarity between pairs of individuals (see text for details)

similarity threshold criterion (Fig. 2). In almost every case similarity values were mutually supportive in the assignment of clones and siblings (e.g. each sample in clone B had pairwise similarity values above the threshold when compared to every other sample in clone B, but had no pairwise similarity values above the threshold with any samples outside the clone). When all similarity values are mutually supportive in this way, it affords great confidence in clonal assignments. Out of the 1,225 pairwise similarity values generated, there were only four cases in

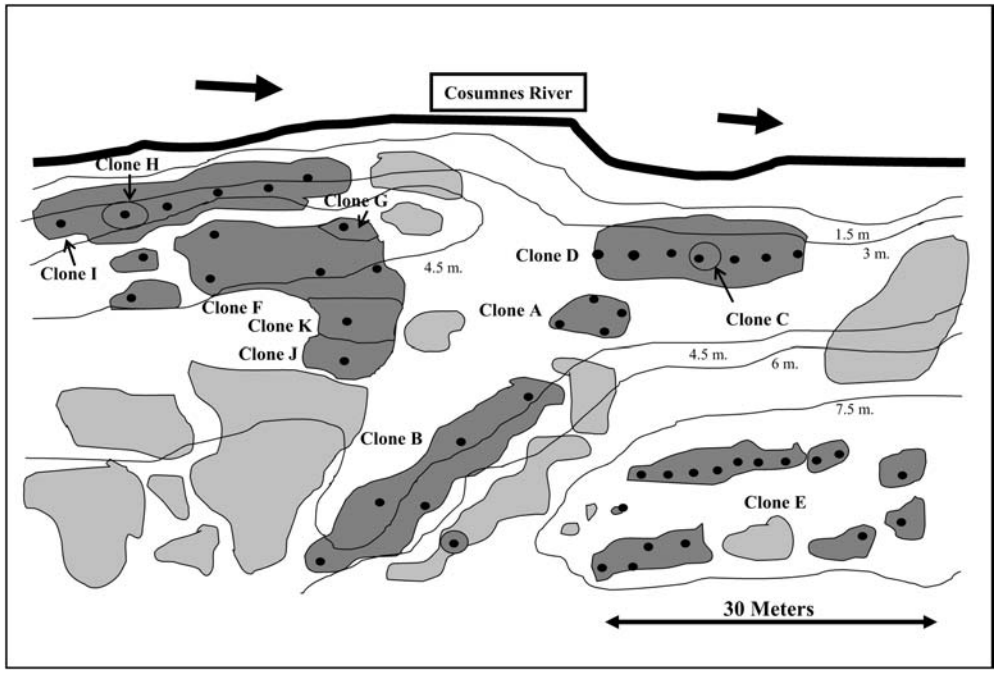

Fig. 3 Mapped distribution of *S. exigua* on a site along the Cosumnes River, California. Sampled areas are shown in dark grey representing 11 different putative clones identified using a threshold level for the Jaccard similarity between pairs of individuals based on 177 AFLP molecular markers (see text for details). Bounded areas in light grey are mapped individuals that were not sampled for molecular typing. Letters for clones as in Fig. 2

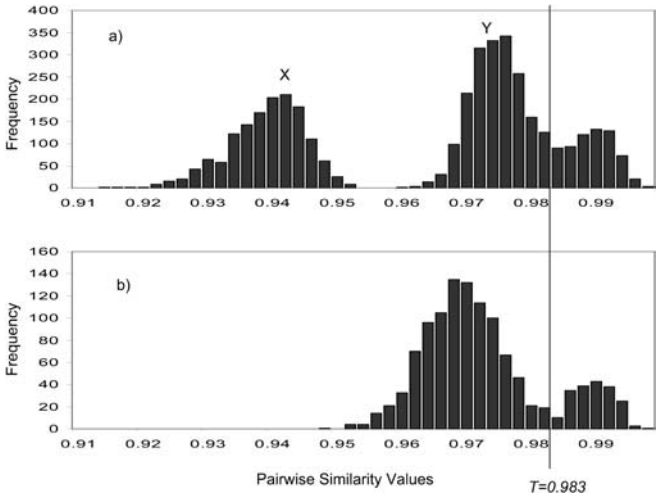

Fig. 4 Histograms of pairwise similarity values for the **a**) variation within clones data set and **b**) the field site data set. The peaks greater than the threshold represent similarities within clones. Peak X most likely represents similarity values between unrelated individuals while the intermediate peak Y represents values between related individuals. This suggests that all the genets on the field site **b**) are related

which intra-clonal similarities were not fully consistent. In three clones, all but one of the pairwise similarity values were mutually supportive above the threshold. In a fourth clone a single sample was found to have only one similarity value above the threshold. In these cases, we used the mean similarity values for each questionable sample in relation to all the other samples in the clone to determine if it should be included. This resulted in the former three pairs being included in the respective clones and with the latter sample being excluded. For example the similarity value for samples 18F and 22F was 0.982, slightly lower than our threshold for clonal assignment. But when these same samples were compared to all other samples in clone E, the similarity values were consistently above the threshold, averaging 0.986 and placing it within the clone.

The sex of each sample was compared within identified clones. In every clone, there was sexual consistency. In other words, no male samples were mismatched with a female sample in a clonal cluster. Clones A, B, C, H and I were identified as male, and clones D, E, F, G, J and K were identified as female.

The Mantel test revealed a significant negative correlation of 0.52 ($p < 0.002$) between the Jaccard similarities matrix (genetic data) and the Euclidean distances matrix (spatial data). When the Jaccard similarity matrix was replaced by a binary matrix corresponding to clonal pairs and non-clonal pairs, the Mantel correlation with geographic distance was 0.50 ($p < 0.001$). Assignment of individuals to clones from the genetic data was generally concordant with a spatial map (Fig. 3), but it is noteworthy that the clones may be irregular in form, fragmented, and several genetically distinct individuals may appear within a single cluster.

The frequency distribution of pairwise Jaccard similarities for individuals from the application site was compared to that of the within-clones sample 1 (Fig. 4). Individuals from the latter site were sampled from a greater geographic area and might capture more clonal variation. Whereas the frequency distribution of the application site was bimodal with modes at 0.988 and 0.970 similarity, the within-clones sample 1 site was trimodal, with two modes very close to those of the previous site (0.988 and 0.972) and a third mode at 0.938.

Discussion

When identifying clones, it is essential to verify an adequate amount of AFLP resolution for the species studied. If the number and polymorphism of AFLP markers is not high enough, genetically distinct but closely related individuals such as siblings could have similarity values as high as 1.000, and thus may be misidentified as clones. Therefore, selected primers must be able to consistently distinguish among siblings, and we recommend the use of known siblings in the screening of primers for clone identification. Once it was established that our primers were polymorphic enough to distinguish among siblings, concerns arose about excess variability and establishing a similarity threshold.

Variation within clones

Our results clearly show that in *S. exigua* clonal fingerprints are less than 100% identical, and that the source of this non-identity lies both in differences between ramets and in replicate runs of the same DNA template. By being highly selective in our choice of AFLP markers, we were able to minimize the experimental error, while at the same time maximizing the detection of variation. None of our AFLP fingerprints among all tested samples were identical. The Jaccard similarity coefficient among members of a clone averaged 0.988 (SE 0.0002). These values are well within the range of repeatability for AFLP studies (Becker et al. 1995; Huys et al. 1996; Tohme et al. 1996; Arens et al. 1998; Winfield et al. 1998; Cabrita et al. 2001). In fact, the Jaccard coefficient tends to give lower levels of similarity than the simple matching coefficient that was used in most of these studies, so the accuracy of our AFLP profiles compare very favorably with other reports.

From our replicated study of clonal variation, we estimated that more than 75% of the variance in AFLP fingerprints was attributable to differences among clones. Of the balance of this variance, stems-within-clones explained a greater proportion than did experimental error. From AMOVA 14.5% was attributable to stems-within-clones and 9.0% to error. Whereas analysis of variance revealed that stems-within-clones was significant for three of the principal components vectors

extracted from the AFLP data set, experimental error was not significant for any vectors.

In most studies in which lack of identity in AFLP profiles among clones is reported, replicate runs of the same sample have been used to quantify the acceptable degree of mismatch and to provide a lower limit of fingerprint similarities for the identification of clones. Arens et al. (1998) set their threshold of similarity for the identification of clones at 0.98 based on separate DNA extractions of the same clone. Winfield et al. (1998) found similarities between duplicate leaf samples to range from 0.96 to 1.00 and predicted that close neighbor trees having similarities of about 0.95 were within the range of scoring errors. We believe that basing the threshold of similarity only on duplicate runs of the same material, or on duplicate sets of leaves, will set a similarity limit that is too high and will exclude clonal members because it has failed to account for possible variation among ramets (Tuskan et al. 1996). In our study, genetic variation among stems of a clone of *S. exigua* was more important than variation due to the AFLP procedure. Silander (1978), Tuskan et al. (1996) and Winfield et al. (1998) allude to the possibility of somatic mutations occurring among clonal members. This deserves more detailed investigation, particularly with regard to its importance in long-lived woody plants. It has been acknowledged for some time in the horticultural literature that genetic variations exist among organisms produced from somatic embryogenesis (Larkin and Scowcroft 1981; Fourre et al. 1997; Hashmi et al. 1997; Vendrame et al. 1999; Hornero et al. 2001) and it has been suggested that this process may promote somatic mutations or somaclonal variation (Larkin and Scowcroft 1981). Another possible source of variation might be leaf sample contamination by fungi and other microorganisms in the field. AFLP analysis is non-specific and as a result has the potential to amplify any DNA present. Dyer and Leonard (2000) found significant variation in *Aphanomyces cochloides* AFLP products due to prokaryote contamination. Unfortunately, in a field study it is impossible to completely eliminate either type of variability, and as a result all possible variation should be accounted for in the calibration of a similarity threshold.

Similarity threshold

To establish the threshold, we used the logic that the maximum seedling pair-wise similarity should provide an upper bound for non-clonal individuals, and that the minimum clone pair-wise similarity should provide the lower bound for clonal individuals. By equating the tails of these two standardized distributions, we were able to derive a similarity with equal probabilities of being incorrectly assigned to the clonal or to the seedling group. The threshold similarity that we obtained for *S. exigua* under our experimental conditions was 0.983, which resulted in 8.1% of pair-wise similarities between true clones being misidentified and 1.5% of similarities between seedlings being misidentified. The inequality in proportions of seedlings assigned to clones and clones assigned to seedlings is due to deviations from normality of the two distributions of pair-wise similarities. This threshold is unique to this study of *S. exigua*, and would need to be derived for each experimental condition. Ideally, to obtain a maximum similarity among seedlings that can be used as a lower limit for clones, full-sib families should be used. However, to simplify the procedure for other researchers and natural resource managers we chose open-pollinated seedlings for our families. As a result the distribution of seedling pair-wise similarities is likely to have a smaller mean and larger variance than would be the case for full-sib families. Depending on the relative changes of these two parameters, our threshold could be higher with half sibs than with full sibs.

Our threshold held up well in a subsequent application to clonal and seedling samples taken from the same two watersheds in California. The independent data set and the initial threshold data set had very similar similarity value distributions and both resulted in threshold values (T) of 0.983. Once again, the proportion of clonal similarities that were incorrectly assigned was slightly greater than the number of seedling similarities incorrectly assigned.

In the field-site application, only four cases of questionable identity, due to inconsistent similarity values, arose. These were cases where a sample had similarity values among other members of the clone that were both below and above the threshold. This was easily resolved by using the average of the sample's within-clone similarity values. The error rate is reduced as the number of clonal members increases because of the mutual support provided by the increasing number of pair-wise similarities. In our field site, only six of the 49 samples were identified as not being part of a larger clone (e.g. more than two members) and as a result our error rate over the entire field data set was very low.

Our assignment of samples to clones was fully consistent with gender, there being no mismatches of males with females. The Mantel test suggested a tendency for spatial pattern in which shorter geographic distances were associated with greater genetic similarities, which is fully consistent with close proximity of clonal members. Nevertheless, our results show that a clone can be represented by a single aggregation of stems, or several aggregations of stems. In some cases a single aggregation can be made up of several smaller clones. This highlights the fact that aggregations alone are not a good substitute for genetic data in identifying clones.

In a comparison of 27 studies, Ellstrand and Roose (1987) calculated the proportion distinguishable (PD), which they defined as the number of genets detected divided by the sample size. They found a mean value of 0.17 for all 21 species. In other studies Pornon et al. (2000) found a PD of 0.16 to 0.18 in Rhododendron, while Suyama et al. (2000) found 0.43 for dwarf bamboo and Parks and Werth (1993) calculated a value of 0.51 for

bracken fern. In this study we calculated a PD value of 0.22. While this value is very similar to the average calculated by Ellstrand and Roose it is important to point out that comparisons between studies are almost impossible unless sampling schemes can be standardized for stem density and sample spacing.

The contrasting frequency distributions for Jaccard similarities in the application population and in the within-clone sample 1 population suggests that pairwise Jaccard similarities below the threshold for clone identification, may be identifiable into two groups; different clones that are closely related, perhaps of sibling origin and different clones that are unrelated, or only distantly related to one another. It is probable that all the non-clonal samples from the application site are closely related and that only at greater distances do we start to sample individuals that are unrelated.

In this preliminary study of a single field site it is clear that clonal growth is a very substantial component of willow site occupation. Clone E was found to represent as much as 43% of the stems sampled and spread over a distance of approximately 30 m. Only six stems sampled (12%) were not detected as part of a larger clone. In fact, assuming only sexual reproduction, and therefore that each stem represents a unique genet, the number of genets would be overestimated by approximately 500%. This extensive clonal habit may explain how willows can so readily colonize sites despite the fact that their seeds are so transient in the environment (McBride and Strahan 1984).

Conclusions

Several authors have found AFLPs to have less than 100% similarity values between members of the same clone. In this work, we were able to identify the largest part of that variation as between stem variation and not lab error. As a result we propose a quantitative approach to identify a pairwise similarity threshold for clone identification. Through the use of greenhouse grown siblings and excavated clones, we were able to verify an adequate level of AFLP resolution and calibrate the pairwise similarity threshold between genetically distinct individuals and members of the same clone. In a replication of the method, and several tests on the field site data, we found the quantitatively derived threshold to be very accurate in identifying clones.

Acknowledgements The authors thank Ignacio Chapela and Joe McBride for help in originating the questions addressed here, and Nasser Kashani, Matteo Garbelotto, Toby Bradshaw, Tom Bruns and George Roderick for help with the AFLP and ABI methods.

References

Arens P, Coops H, Jansen J, Vosman B (1998) Molecular genetic analysis of black poplar (*Populus nigra* L.) along Dutch rivers. Mol Ecol 7:11–18

Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249:65–73

Cabrita LF, Aksoy U, Hepaksoy S, Leitao JM (2001) Suitability of isozyme, RAPD and AFLP markers to assess genetic differences and relatedness among fig (*Ficus carica* L.) clones. Scien Hort 87:261–273

Cullings K (1992) Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. Mol Ecol 1:233–240

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Dyer AT, Leonard KJ (2000) Contamination, error, and nonspecific molecular tools. Phytopathology 90:565–567

Ellstrand NC, Roose ML (1987) Patterns of genotypic diversity in clonal plant species. Am J Bot 74:123–131

Escaravage N, Questiau S, Pornon A, Doche B, Taberlet P (1998) Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. Mol Ecol 7:975–982

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Fourre JL, Berger P, Niquet L, Andre P (1997) Somatic embryogenesis and somaclonal variation in Norway spruce: morphogenetic, cytogenetic and molecular approaches. Theor Appl Genet 94:159–169

Gompper ME, Gittleman JL, Wayne RK (1997) Genetic relatedness, coalitions and social behaviour of white-nosed coatis, *Nasua narica*. Anim Behavior 53:781–797

Hashmi G, Huettel R, Meyer R, Krusberg L, Hammerschlag F (1997) RAPD analysis of somaclonal variants derived from embryo callus cultures of peach. Plant Cell Rep 16:624–627

Hornero J, Martinez I, Celestino C, et al (2001) Early checking of genetic stability of cork oak somatic embryos by AFLP analysis. Int J Plant Sci 162:827–833

Hulst van der RGM, Mes THM, den Nijs JCM, Bachmann K (2000) Amplified fragment length polymorphism (AFLP) markers reveal that population structure of triploid dandelions (*Taraxacum officinale*) exhibits both clonality and recombination. Mol Ecol 9:1–8

Huys G, Coopman R, Janssen P, Kersters K (1996) High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. Int J Syst Bacteriol 46:572–580

Klimeš L, Klimesova J, Hendriks R, Van Groenendael J (1997) Clonal plant architecture: a comparative analysis of form and function. In: de Kroon H, Van Groenendael J (eds) The ecology and evolution of clonal plants. Backhuys Publishers, Leiden, pp 1–30

Larkin P, Scowcroft W (1981) Somaclonal variation – a novel source of variability from cell cultures for plant improvement. Theor Appl Genet 60:197–214

Lehmann N, Clarkson P, Mech LD, Meier TJ, Wayne RK (1992) A study of the genetic relationships within and among wolf packs using DNA fingerprinting and mitochondrial DNA. Behavioral Ecol Sociobiol 3:83–94

McBride JR, Strahan J (1984) Establishment and survival of woody riparian species on gravel bars of an intermittent stream. Am Midland Nat 112:235–245

Miller M (1998) AMOVA-PREP – a program for the preparation of AMOVA input files from dominant-marker raw data., Northern Arizona University, Flagstaff, USA

Mills LS, Citta JJ, Lair KP, Schwartz MK, Tallmon DA (2000) Estimating animal abundance using non-invasive DNA sampling: promise and pitfalls. Ecol Applic 10:283–294

Parks JC, Werth CR (1993) A study of spatial features of clones in a population of bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae). Am J Bot 80:537–544

Peterson CJ, Jones RH (1997) Clonality in woody plants: a review and comparison with clonal herbs. In: de Kroon H, Van Groenendael J (eds) The ecology and evolution of clonal plants. Backhuys Publishers, Leiden, pp 263–289

Pornon A, Escaravage N, Thomas P, Taberlet P (2000) Dynamics of genotypic structure in clonal *Rhododendron ferrugineum* (Ericaceae) populations. Mol Ecol 9:1099–1111

Price EAC, Marshall C (1999) Clonal plants and environmental heterogeneity: an introduction to the proceedings. Plant Ecol 141:3–7

Silander JA (1978) Microevolution in clonal plants. In: Jackson JBC, Buss LW, Cook RE (eds) Population biology and evolution of clonal organisms. Yale University Press, New Haven, pp 107–152

Sneath P, Sokal R (1973) Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman and Co., San Francisco, California

Suyama Y, Obayashi K, Hayashi I (2000) Clonal structure in a dwarf bamboo (*Sasa senanensis*) population inferred from amplified fragment length polymorphism (AFLP) fingerprints. Mol Ecol 9:901–906

Tohme J, Orlando Gonzalez D, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. Crop Sci 36:1375–1384

Tuskan GA, Francis KE, Russ SL, Romme WH, Turner MG (1996) RAPD markers reveal diversity within and among clonal and seedling stands of aspen in Yellowstone National Park, USA. Can J Forest Res 26:2088–2098

Van Groenendael JM, Klimes L, Klimesova J, Hendriks RJJ (1996) Comparative ecology of clonal plants. Phil Trans R Soc Lond B Biol Sci 351:1331–1339

Vendrame WA, Kochert G, Wetzstein HY (1999) AFLP analysis of variation in pecan somatic embryos. Plant Cell Rep 18:853–857

Vos P, Hogers R, Bleeker M, et al.(1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Winfield MO, Arnold GM, Cooper F, et al. (1998) A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the upper severn area of the UK using AFLP markers. Mol Ecol 7:3–10