Regular Articles

Phylogenetic Signal in AFLP Data Sets

WIM J. M. KOOPMAN

Biosystematics Group, Nationaal Herbarium Nederland—-Wageningen University branch, Wageningen University, Generaal Foulkesweg 37, 6703BL Wageningen, The Netherlands; E-mail: wim.koopman@wur.nl
Present Address: Plant Research International B.V., P.O. Box 16, 6700 AA, Wageningen, The Netherlands

Abstract.—AFLP markers provide a potential source of phylogenetic information for molecular systematic studies. However, there are properties of restriction fragment data that limit phylogenetic interpretation of AFLPs. These are (a) possible nonindependence of fragments, (b) problems of homology assignment of fragments, (c) asymmetry in the probability of losing and gaining fragments, and (d) problems in distinguishing heterozygote from homozygote bands. In the present study, AFLP data sets of *Lactuca* s.l. were examined for the presence of phylogenetic signal. An indication of this signal was provided by carrying out tree length distribution skewness (g_1) tests, permutation tail probability (PTP) tests, and relative apparent synapomorphy analysis (RASA). A measure of the support for internal branches in the optimal parsimony tree (MPT) was made using bootstrap, jackknife, and decay analysis. Finally, the extent of congruence in MPTs for AFLP and internal transcribed spacer (ITS)-1 data sets for the same taxa was made using the partition homogeneity test (PHT) and the Templeton test. These analytical studies suggested the presence of phylogenetic signal in the AFLP data sets, although some incongruence was found between AFLP and ITS MPTs. An extensive literature survey undertaken indicated that authors report a general congruence of AFLP and ITS tree topologies across a wide range of taxonomic groups, suggesting that the present results and conclusions have a general bearing. In these earlier studies and those for *Lactuca* s.l., AFLP markers have been found to be informative at somewhat lower taxonomic levels than ITS sequences. Tentative estimates are suggested for the levels of ITS sequence divergence over which AFLP profiles are likely to be phylogenetically informative. [Amplified Fragment Length Polymorphism (AFLP) markers; congruence; Dollo parsimony; g_1 statistic; internal transcribed spacer (ITS) sequences; *Lactuca;* partition homogeneity test (PHT); phylogenetic signal.]

DNA sequences are the main source of information for molecular phylogenetic studies. For studies at the genus level and above, a wide range of sequences are available from the nuclear and plastid genomes. For studies among closely related genera or species, however, sequences often do not show enough variation. At these lower levels, restriction fragments may be suitable phylogenetic markers, providing they contain sufficient phylogenetic signal. The restriction fragment markers most commonly used are RFLPs (restriction fragment length polymorphisms; Grodzicker et al., 1974; Botstein et al., 1980), RAPDs (random amplified polymorphic DNAs; Williams et al., 1990; Welsh and McClelland, 1990), microsatellites or SSRs (simple sequence repeats; reviewed in Cregan, 1992), ISSRs (inter simple sequence repeats; Zietkiewicz et al., 1994), and AFLPs (Vos et al., 1995; often referred to as Amplified Fragment Length Polymorphisms, but officially a trademark rather than an abbreviation).

Elaborate evaluations and comparisons of these markers and their applications in systematic studies have been made by, e.g., Bachmann (1992), Powell et al. (1996), Karp et al. (1996), Jones et al. (1997), Milbourne et al. (1997), Russell et al. (1997), and McGregor et al. (2000). Focusing mainly on repeatability and information content, these studies show large differences among the various markers. Whereas the repeatability of RFLPs, SSRs, and AFLPs was found to be nearly 100%, that of RAPDs and ISSRs was much lower (75% to 85%, see Jones et al., 1997, and McGregor et al., 2000). Information content has typically been evaluated in terms of the number of polymorphic loci per gel lane (marker index; Powell et al., 1996) or assay unit (i.e., per primer or primer combination). In the above studies it was observed that the information content of RFLPs was lowest, whereas that of SSRs, RAPDs, and ISSRs was about twice as high. The information content of AFLPs was found to exceed that of RAPDs by a factor of 10 or more. This high information content is due to the high number of loci that are amplified simultaneously (called the multiplex ratio).

Due to the high number of polymorphisms per assay unit, the high reproducibility, and the fact that no prior sequence information is needed to perform the AFLP technique, its popularity in systematic studies has increased rapidly over the past few years. In these studies, a variety of techniques have been applied for data analysis. Clustering methods (usually UPGMA, e.g., Huys et al., 1996 (*Aeromonas*); Keim et al., 1997 (*Bacillus*); Kardolus et al., 1998 (*Solanum*); DeScenzo et al., 1999 (*Eutypa*); Kiers et al., 2000 (*Cichorium*); Werres et al., 2001 (*Phytophthora*); Pelser et al., 2003 (*Senecio*)) and neighbor joining (e.g., Lu et al., 1996 (*Pisum*); Angiolillo et al., 1999 (*Olea*); Tredway et al., 1999 (Clavicipitaceae); Giannasi et al., 2001 (*Trimeresurus*); Hodkinson et al., 2002

(*Miscanthus*)) are commonly used, as is heuristic parsimony (e.g., Keim et al., 1997; Kardolus et al., 1998; Tredway et al., 1999; Hodkinson et al., 2000 (*Phyllostachys*); Despres et al., 2003 (*Trollius*)).

To be suitable for phylogenetic analysis, restriction fragment data (such as AFLP markers) have to meet two basic requirements (Backeljau et al., 1995; Swofford et al., 1996): (1) the fragments must have evolved independently (Karp et al., 1996); and (2) fragments of equal length must be homologous (Black, 1993; Karp et al., 1996). As was pointed out by Karp et al. (1996) and Swofford et al. (1996), the problem of nonindependence of fragments and the problem of identifying homologous fragments potentially limit the phylogenetic interpretation of restriction fragment data. For AFLP markers used at the intraspecific level, homology assignment problems may vary, depending on the species. For example, in five relatively unrelated potato (*Solanum tuberosum*) genotypes, Rouppe van der Voort et al. (1997) identified 131 comigrating AFLP bands, 117 of which (89%) mapped to the same genomic regions. Twenty of these 117 putative homologous bands were sequenced, and the sequences of 19 bands (95%) were shown to be nearly identical. In contrast, Mechanda et al. (2004) detected intraspecific sequence identities in *Echinacea* as low as 75.82% for monomorphic AFLP bands, and 23.66% for polymorphic bands. Mechanda et al. (2004) and O'Hanlon and Peakall (2000) showed that the level of error in homology assignment for bands of the same size rapidly increased with taxonomic divergence. It reached 98.75% for polymorphic bands in different species of *Echinacea* (Mechanda et al., 2004), and 100% for species from different subtribes of Carduinae thistles (O'Hanlon and Peakall, 2000). Further, whereas nonhomologous fragments of equal length may be found in different genotypes (Rouppe van der Voort et al., 1997; O'Hanlon and Peakall, 2000; Mechanda et al., 2004), nonhomologous fragments of the same size are also found to occur within the same individual (Rouppe van der Voort et al., 1997; Hansen et al., 1999; Meksem et al., 2001; Mechanda et al., 2004). The extent of nonindependence of fragments in AFLP profiles has not yet been studied quantitatively.

A specific problem for phylogenetic analysis caused by error in homology assignment for increasingly diverged profiles is that this error results in a form of long branch attraction. The large proportion of nonhomologous fragments between the more distantly related genotypes results in high levels of homoplasy between these genotypes and the effect may mislead both parsimony and clustering methods of tree building. Inclusion of such distantly related AFLP profiles is thus undesirable, and recent research has involved developing tests to enable identification and removal of such genotypes (Koopman and Gort, 2004).

Other issues that are potentially significant for phylogenetic analyses of AFLP data include (1) asymmetry in the probability of losing and gaining fragments (loss of a fragment is much more probable than gain), and (2) the fact that AFLP markers are usually scored dominantly (i.e., without distinction between homozygotes and heterozygotes). Both features of the data may also limit the usefulness of AFLP profiles as a source for phylogenetic markers, as they will increase the amount of stochastic noise in the data (see Karp et al., 1996).

Although heuristic parsimony analyses of AFLP data have been published for many species groups, only a few studies (e.g., Giannasi et al., 2001; Despres et al., 2003) have made any attempt to address the question of the presence of phylogenetic signal in the data sets. Given the popularity of AFLP markers for systematic studies, the increasing number of phylogenetic analyses among these studies, and the limitations of AFLPs as phylogenetic markers, a more thorough examination of phylogenetic signal in AFLP data sets is warranted.

In this paper phylogenetic signal in AFLP data sets of *Lactuca* s.l. is examined. Techniques to examine phylogenetic signal in data sets are by their nature heuristic, and the subject of measuring phylogenetic signal is controversial (see, e.g., the recent paper by Grant and Kluge, 2003). Much discussion of this topic has appeared in cladistic literature, but it is beyond the scope of the present paper to engage in this debate. Rather, for the purpose of the present study the most widely used procedures have been employed with brief reference to the debate surrounding them. Notwithstanding the limitations of the various techniques, applied together they provide a measure of the extent of phylogenetic signal in *Lactuca* AFLP data sets. The findings from this study have been placed in the context of the findings of others as revealed by a literature survey. This comparison indicates that there is a general congruence of internal transcribed spacer signal (ITS), and AFLP-based tree topologies for a wide range of taxa in fungi, oomycetes, plants, and bacteria. Based on the analyses of the *Lactuca* s.l. data sets, and on the literature survey, tentative estimates are suggested for the levels of ITS sequence divergence over which AFLP profiles are likely to be phylogenetically informative.

Materials and Methods

The AFLP data sets were selected from larger sets studied in Koopman et al. (2001) and contained 84 accessions from 19 species of *Lactuca* and related genera. The accessions are identical to those reported in a previous ITS-1 sequence study (Koopman et al., 1998), but with the exclusion of *Lactuca sativa* CGN 5045, *Mycelis muralis* CGN 9367 and CGN 5005, *Prenanthes purpurea* W9534, and all accessions of *Sonchus*, *Taraxacum*, and *Chondrilla*. The first data set was generated with primer combination (pc) E35/M48 (EcoRI + ACA/MseI + CAC) and contained 530 polymorphic bands, 467 of which were parsimony informative. The second data set was generated with pc E35/M49 (EcoRI + ACA/MseI + CAG) and contained 500 polymorphic bands (434 parsimony informative).

The ITS-1 data set used as a reference in the present study was a subset from the data set of Koopman et al. (1998), containing the same 84 accessions as the AFLP data sets. The total aligned length was 261 nucleotides, including 131 polymorphic characters (116 parsimony

informative). The original multiple sequence alignment was corrected by hand to minimize the number of independent indel events. In the original ITS-1 data set of Koopman et al. (1998), each species was represented by at least two different accessions. For several species, the different accessions showed identical ITS-1 sequences. The presence of such duplicate sequences introduces structure in the data set. This structure will be detected by testing procedures and may be erroneously recognized as phylogenetic signal. To avoid this artifact, TLD and RASA (see below) for the ITS-1 data were calculated using a data set in which each sequence was present only once (leaving 46 unique sequences). The ITS most parsimonious trees (MPTs) presented in the results section were also calculated using this data set. Because tests of congruence require equal numbers of taxa in the data sets to be compared and the AFLP data set contained 84 taxa, congruence of the AFLP and ITS-1 data sets and MPTs was tested using the ITS-1 data set with 84 accessions.

As a first approach, the data sets were examined for the presence of phylogenetic signal. Three techniques for testing phylogenetic signal are commonly used: tree length distribution skewness (TLD; Hillis, 1991; Huelsenbeck, 1991; Hillis and Huelsenbeck, 1992), permutation tail probability testing (PTP; Archie, 1989; Faith and Cranston, 1991), and relative apparent synapomorphy analysis (RASA; Lyons-Weiler et al., 1996). Data decisiveness (Goloboff, 1991) is a related technique, but without the explicit claim that phylogenetic signal is measured (Carpenter, 1992).

TLD was used to measure phylogenetic signal because, although sometimes criticized (Källersjö et al., 1992; Lyons-Weiler et al., 1996), it is still widely employed. To determine phylogenetic signal based on TLD, a length distribution of randomly generated phylogenetic trees is assembled based on the observed data set. The skewness of this distribution is described by the g_1 statistic (Sokal and Rohlf, 1969). Negative g_1 values below a certain critical value (derived from tree length distributions based on random data sets) indicate significant phylogenetic signal (Hillis and Huelsenbeck, 1992). TLD and g_1 of 100,000 random trees were determined for the separate and combined AFLP data sets and for the ITS-1 data set, using PAUP* 4.0b8 (PPC/Altivec; hereafter PAUP*; Swofford, 1999).

The PTP tests of Archie (1989) and Faith and Cranston (1991) determine phylogenetic signal by comparing the length of a MPT based on an observed data set with the lengths of MPTs based on randomizations (usually 99 or 999) of that data set. The data set is randomized by randomly permuting the states of each of the characters across all taxa. The test statistic is the proportion of all data sets (i.e., the observed data set plus the randomizations) producing a tree as short or shorter than the MPT from the observed data set. Significant phylogenetic signal is concluded when this fraction is below a critical value. The PTP test has been extensively criticized (Carpenter, 1992; Källersjö et al., 1992; Steel et al., 1993; Lyons-Weiler et al., 1996; Carpenter et al., 1998), mainly

because the null hypothesis of random distribution of character states is considered invalid. For example, Carpenter (1992) argued that the use of a random null model is unfit to measure cladistic corroboration in the sense of Popper (and hence phylogenetic structure). In his reply, however, Faith (1992) demonstrated that the PTP test does measure Popperian corroboration, notwithstanding Carpenters claim for the opposite. Källersjö et al. (1992) criticized the fact that the PTP test may show significant phylogenetic structure for data sets yielding a set of MPTs whose strict consensus is entirely unresolved. Faith and Ballard (1994) responded by pointing out that a set of MPTs with an unresolved strict consensus tree does not exclude the presence of phylogenetic structure in the underlying data set. They demonstrated that the PTP test detected phylogenetic structure in the test data set of Källersjö et al. (1992), even though the data set yielded an unresolved strict consensus tree. In the present paper, the PTP test of Archie (1989) and Faith and Cranston (1991) is used as implemented in PAUP*. The test was performed on the separate (99 randomizations) and combined (999 randomizations) AFLP data sets, with randomization of either entire data sets or only ingroup taxa. Based on previous results (Koopman et al., 1998), *Prenanthes purpurea* L. was used as the outgroup. Heuristic searches were performed with random addition sequences (10 replicates) as well as with simple addition sequences using TBR branch swapping with "multrees" switched on. Parsimony settings were: acctran and collapse of zero-length branches.

RASA was presented as an alternative to TLD and PTP testing because it has been claimed that it is (1) an a priori measure of phylogenetic signal, (2) a tree-independent measure of phylogenetic signal, and (3) not at all based on the assumptions of maximum parsimony. To perform RASA, two measures are calculated for each taxon pair: a relative apparent synapomorphy score (RAS; representing the number of times that a taxon pair shares a character state to the exclusion of another taxon, summed over all characters), and the number of characters involved in the computation of RAS, called E. RAS, is plotted against E, and the observed slope of RAS on E is the measure of phylogenetic signal. A null slope is determined from a plot of RAS against E after reciprocal equiprobable redistribution of RAS and E. The test statistic for homogeneity of slopes (Myers, 1990) is used to compare the slopes, and an observed slope significantly steeper than the null slope indicates the presence of phylogenetic signal. Recently, Simmons et al. (2002), Faivovich (2002), and Farris (2002) claimed that RASA may detect strong phylogenetic structure in unstructured data sets (both hypothetical and observed), although it may fail to detect signal in data sets yielding strongly supported phylogenies. They suggested two main sources of error. First, the fact that phylogenetically uninformative characters are included in the calculation of RAS results in RAS being a measure of phenetic similarity rather than one of cladistic hierarchy. Second, the regression approach employed in the RASA procedure may be criticized because regression analyses are only valid using independent variables. In

the RASA procedure RAS and E are not independent, because both are calculated from the same three taxon statements. Despite these potential problems, RASA was applied here in its original form, using the RASA Web tool at http://bioinformatics.upmc.edu/RASA.html.

As a second approach, internal support was calculated for unweighted Wagner parsimony trees based on the combined AFLP data sets using PAUP*. MPTs were calculated in heuristic searches comprising 10,000 random addition sequences with TBR branch swapping and "multrees" switched off. Parsimony settings were: acctran and "collapse of zero-length branches." *Prenanthes purpurea* accessions W9505 and W9525 were used as the outgroup. Support for the unweighted MPT topologies was determined using the three most widely employed methods: nonparametric bootstrapping (Felsenstein, 1985a), jackknifing (Farris et al., 1996), and branch support (Bremer, 1988), the latter also known as decay index (Donoghue et al., 1992) or Bremer support (Källersjö et al., 1992). To avoid confusion, the term "internal support" will be used to refer to branch support in a general sense, whereas branch support sensu Bremer (1988) will be referred to as decay index (DI).

In decay analyses, the support for a clade in the MPT is determined from its presence in suboptimal trees. To establish the DI, trees are examined that are N steps longer than the MPT. The DI of a clade is determined as the number of extra steps N that correspond to the shortest tree in which that clade collapses. Because DIs are calculated using trees based on the same data set as the MPT, calculation of the DIs does not require any manipulation of the data set itself. Bootstrapping and jackknifing are fundamentally different in that the support for a clade is measured in terms of its frequency of occurrence in MPTs derived from resampled versions of the original data set.

DI values were calculated with AutoDecay version 3.03 (Eriksson and Wikström, 1996) using the heuristic search option with 10 random addition sequences, TBR branch swapping, and "multrees" switched on. Bootstrap values were calculated in 2500 replicates of a full heuristic search, with 10 random addition sequences in each replicate, and remaining settings as above. Jackknife values were calculated in a fast heuristic search with 25,000 replicates, nominal deletion of 37% of the characters in each replicate (according to Farris et al., 1996), and "Jac" resampling. Starting trees were obtained using random addition sequences without branch swapping.

DeBry (2001) has pointed out that "raw" DI values may give a misleading indication of the reliability of a branch, because the associated critical values increase with increasing branch lengths. The relationship between critical values and branch lengths is not straightforward, depending, e.g., on the total number of characters versus taxa in a tree, the number of splits in the lineages descending from a certain branch, and the number of contradictory characters on a branch. This complex relationship not only renders comparisons of DI values in the AFLP and ITS-1 MPTs potentially unreliable (even in the light of branch lengths), it also hampers significance

testing of these values. Nevertheless, following Felsenstein (1985b), DeBry (2001) concludes that DI values <4 should not be considered as strong support under any circumstances. Significance testing of bootstrap and jackknife values is equally problematic, although some authors suggest that a figure of 80% under some conditions indicates relatively strong support (DeBry, 2001). In the present study, decay analysis is applied as a method complementary to bootstrapping and jackknifing. Ideally, all three methods should (apart from scale factors) indicate similar internal support. However, because bootstrapping/jackknifing and decay analysis are fundamentally different approaches, possible peculiarities in the data sets (e.g., large amounts of contradictory characters on certain branches) may lead to discrepancies in internal support for the different methods. In order to trace such peculiarities, the bootstrap, jackknife, and DI values are discussed in relation to each other.

As a third approach, the congruence of the AFLP data and ITS-1 sequence data was examined, as well as the congruence of MPTs based on these data. Notwithstanding their drawbacks (reviewed in Álvarez and Wendel, 2003), ITS sequences are widely used phylogenetic markers (Baldwin, 1992), and congruence of ITS and AFLP data used to construct MPTs can be used to indicate whether or not AFLP data are suitable for phylogenetic analysis.

The congruence of AFLP and ITS-1 data sets was determined using the partition homogeneity test (PHT) of Farris et al. (1995), based on the incongruence length difference of Mickevich and Farris (1981). The test comprises the following steps: (1) determine the sum L_{x+y} of the lengths of the MPTs from both data sets; (2) randomly partition all characters into new data sets of the original sizes, and do this W times (e.g., 100); (3) determine the lengths of the MPTs from the partitioned data sets; (4) count the number S of MPTs that are longer than L_{x+y}; 5) the error rate on rejecting the null hypothesis of congruency is P = 1 − (S/(W + 1)). The PHT was performed with 500 replicates (= 499 repartitions) using the test implemented in PAUP*. Trees for each replicate were generated with heuristic searches starting with 1000 random addition sequences, TBR branch swapping, and "multrees" switched off. Parsimony settings were: acctran and "collapse of zero-length branches." Although the PHT is widely used, several authors have noticed serious limitations of the test (Barker and Lutzoni, 2002). Most importantly, there does not seem to be a straightforward relationship between congruence and phylogenetic accuracy (Cunningham, 1997; Yoder et al., 2001), the test is heavily influenced by the substitution model employed (Dowton and Austin, 2002; Barker and Lutzoni, 2002), and the test shows an excessive type I error rate (i.e., the probability of falsely rejecting the null hypothesis of congruence) when the matrices compared differ in their level of homoplasy (Dolphin et al., 2000; Yoder et al., 2001; Barker and Lutzoni, 2002). A detailed examination of the impact of these limitations on the use of the PHT test on AFLP data sets is beyond the scope of the present paper, and therefore

the PHT test is applied here, in spite of the potential problems.

To serve as a reference for the AFLP MPTs, ITS-1 MPTs were calculated in PAUP* using a heuristic search with simple taxon addition, TBR branch swapping, and "multrees" switched on. Parsimony settings were: acctran and "collapse of zero-length branches." *P. purpurea* accessions W9505 and W9525 were used as the outgroup. An additional analysis with 10,000 random addition sequences and "multrees" switched off was performed to identify possible islands of shorter trees. However, no such islands were found. Support for the ITS-1 MPT topologies was determined as described for AFLP MPTs.

Congruence of AFLP and ITS-1 MPTs was determined in two ways, using one arbitrarily selected MPT for each of the data sets.

First, to obtain a general estimate of congruence, the fit of both trees to either of the data sets was examined using the Templeton test (Templeton, 1983). A drawback of the Templeton test is that it may be biased when the trees compared are not derived independently of the data sets used for testing (Goldman et al., 2000). However, the applicability of alternative tests such as the Kishino-Hasegawa test (KH test; Hasegawa and Kishino, 1989; Kishino and Hasegawa, 1989) and the Shimodaira-Hasegawa test (SH test; Shimodaira and Hasegawa, 1999) is also limited because they all have their own specific drawbacks (discussed in, e.g., Shimodaira and Hasegawa, 1999, and Goldman et al., 2000). Therefore, testing of the fits of the trees to the data sets was limited to the application of the Templeton test. However, considering the fact that the trees examined are not derived independently of the data used for testing, the results should be regarded with some caution.

The original Templeton test is a one-tailed Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) that compares the number of changes required for each character on each of the trees (excluding ties). The difference for each character gets a signed rank number, and the negative rank sum is used as test statistic. The test statistic and the total number of ranks are converted into a probability statement based on the statistical tables from the Wilcoxon test, or, for a total number of ranks >25, using a normal approximation (Siegel, 1956). The two-tailed version of the Templeton test implemented in PAUP* was used.

Second, a comparison of the AFLP and ITS-1 MPT topologies was conducted by eye. The comparison involved a detailed examination of the presence of clades and splits common to both trees, in relation to their statistical support. The objective was to assess (1) the general congruence of the AFLP and ITS trees and (2) the taxonomic level at which the AFLP and ITS markers are informative (i.e., yield supported topologies).

In the comparison of AFLP and ITS-1 MPTs, unweighted Wagner parsimony was employed to calculate the AFLP MPT. The use of unweighted parsimony assumes equal probabilities of losing and gaining characters. However, restriction fragment data (such as obtained using AFLP markers) do not meet this assump-

tion, because the restriction sites are more easily lost than gained (e.g., any mutation in a restriction sites causes loss of the fragment, whereas only one specific backmutation restores the lost site). Focusing on this asymmetry, DeBry and Slade (1985) suggested the use of Dollo parsimony for animal mtDNA restriction site characters. To examine whether this approach could be valid for AFLP data, the *Lactuca* s.l. data set was reanalyzed in PAUP* using unrooted Dollo parsimony (all characters of type Dollo Up, no ancestor included in the analysis). MPTs were calculated in heuristic searches comprising 5000 random addition sequences with TBR branch swapping and "multrees" switched off. Parsimony settings were: acctran and "collapse of zero-length branches." The resulting trees were rooted using *P. purpurea* W9505 and W9525 as the outgroup. Bootstrap support for the trees was calculated in 1000 replicates of a full heuristic search, with 10 random addition sequences in each replicate and remaining settings as above.

Dollo parsimony, however, is rather restrictive because of the assumption that restriction sites can be gained only once. Therefore, Jansen et al. (1991) and Holsinger and Jansen (1993) suggested weighted parsimony (equivalent to a relaxed Dollo criterion, depending on the weights) as an alternative. Although this allows the loss/gain probabilities to more realistically fit the data, it introduces the problem of determining the proper weights. Jansen et al. (1991) and Holsinger and Jansen (1993) suggested using weights proportional to $-\ln(rt)$ for gains and to $-\ln(qt)$ for losses, with rt and qt being the expected numbers of gains and losses on a branch as determined from an initial Wagner tree. To determine rt and qt, separate frequency distributions are constructed for the site losses and site gains on the initial tree, describing the numbers of characters with 1, 2, 3, etc. transitions. Assuming that all sites change at an equal rate, the actual numbers of gains and losses, given that we know we have a character exhibiting at least one gain or loss, can be estimated by determining the Poisson parameters $\hat{\lambda}_g$ and $\hat{\lambda}_l$. These parameters are determined from the frequency distributions according to $\hat{\lambda}_g = \frac{1}{B \sum n_i} \sum i n_i$ and $\hat{\lambda}_l = \frac{1}{B \sum k_i} \sum i k_i$, where n_i is the number of characters that show i gains across the initial tree, k_i is the number of characters that show i losses, and B is the total number of branches. Next, the unconditional expectations for rt and qt are determined by solving the equations $\hat{\lambda}_g = \frac{rt}{1-e^{-rt}}$ and $\hat{\lambda}_l = \frac{qt}{1-e^{-qt}}$ (Jansen et al., 1991; K.E. Holsinger, personal communication). For the *Lactuca* s.l. data set, character state changes were determined on the first tree from the unweighted parsimony analysis, using MacClade 4 (Maddison and Maddison, 2000). Both the MIN and MAX options in the chart/changes platform were applied. The characters' state changes were determined only on the well-supported *sat/ser/dreg/alt* + *L. aculeata* clade (see Figure 1) because the presence of autapomorphies for more distantly related taxa on less supported branches would give rise to an overestimation of the number of gains. The resulting weight of 1.02 in favor of losses was subsequently used in a weighted parsimony analysis. All characters were weighted using
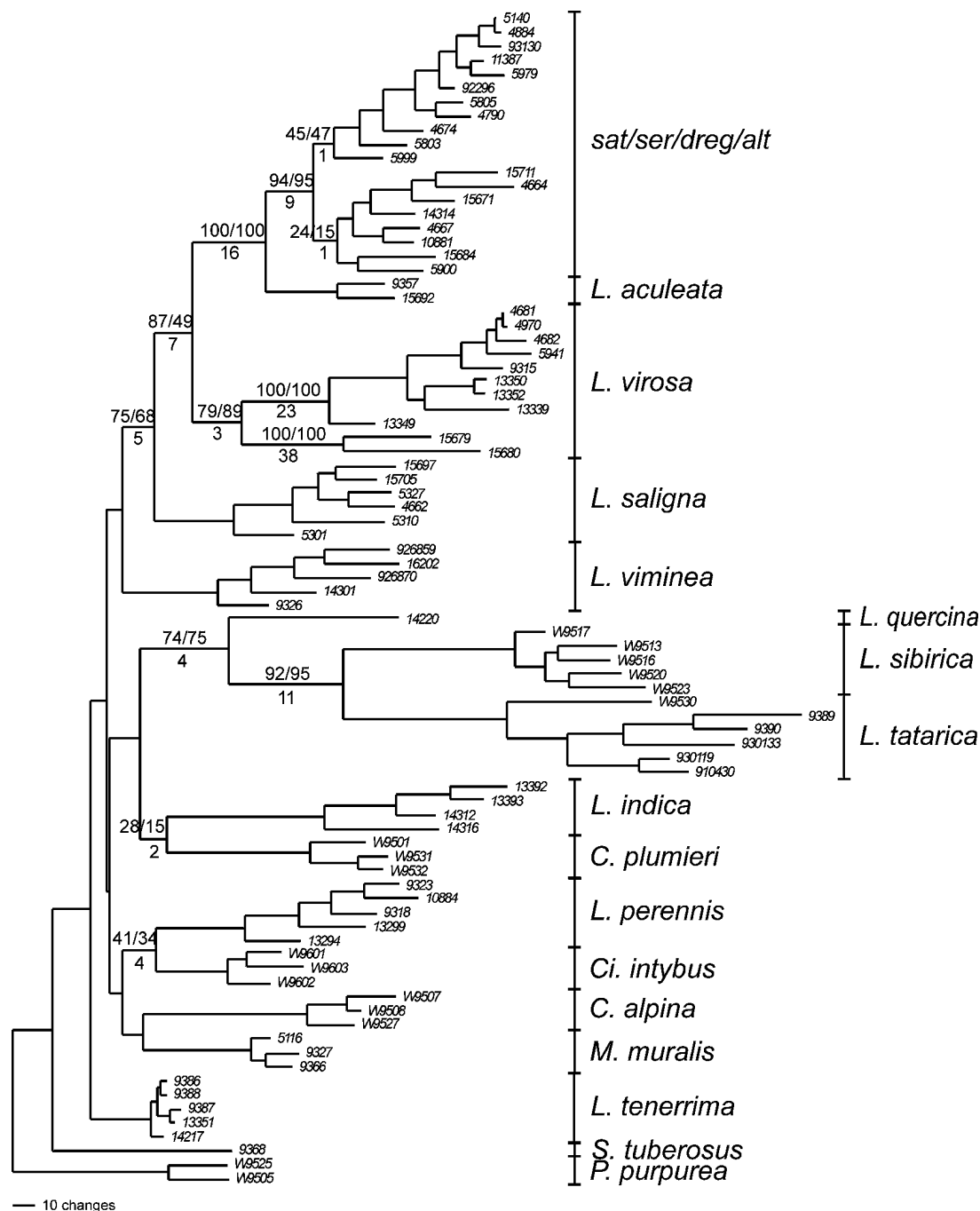
FIGURE 1.   One of the MPTs based the unweighted Wagner parsimony search on the combined AFLP primer combinations E35/M48 and E35/M49. Above branches: bootstrap values/jackknife values. Below branches: branch supports. *sat/ser/dreg/alt*: group of intermixed and closely related species *Lactuca sativa*, *L. serriola*, *L. dregeana*, and *L. altaica*. Genus abbreviations: *L = Lactuca*, *C = Cicerbita*, *Ci = Cichorium*, *M = Mycelis*, *S = Steptorhamphus*, *P = Prenanthes*.

the "user defined type" option and a 0/1 stepmatrix with a weight of 99 for losses and 101 for gains. MPTs were calculated in heuristic searches comprising 5000 random addition sequences with TBR branch swapping and "multrees" switched off. Parsimony settings were: acctran, "collapse of zero-length branches," and ancestor "standard" (=all missing) included in the analysis. The resulting trees were rerooted using *P. purpurea* W9505 and W9525 as the outgroup. Bootstrap support for the trees was calculated in 1000 replicates of a full heuristic search, with 10 random addition sequences in each replicate and remaining settings as above.

RESULTS
Phylogenetic Signal

The g_1 statistic from the TLD test was -0.52 for AFLP pc E35/M48, -0.46 for pc E35/M49, and -0.50 for the combined data sets. The lettuce data sets contain over 25 taxa and 500 or more variable characters, and therefore the critical value of -0.08 ($P = 0.01$) was used (Hillis and Huelsenbeck, 1992). All three g_1 values are considerably lower than this critical value, indicating the presence of significant phylogenetic signal in the AFLP data sets. The g_1 statistic for the ITS-1 data set was -0.59, which is considerably lower than the critical value of -0.12 (>25 taxa, 100 variable characters, $P = 0.01$), indicating significant phylogenetic signal.

The results from the PTP tests were highly significant in suggesting phylogenetic signal in both the separate (99 randomizations, all $\alpha = 0.01$) and combined (999 randomizations, all $\alpha = 0.001$) AFLP data sets. For pc E35/M48, the tree length for the observed data set was 1902 steps. The lengths of the shortest trees from the randomized data sets were 3959 (all taxa randomized, simple addition), 3951 (all taxa randomized, random addition), 3907 (only ingroup taxa randomized, simple addition), and 3905 steps (only ingroup taxa randomized, random addition). For pc E35/M49, the length for the observed data set was 1835 steps, and the lengths for the randomized data sets were 3649, 3637, 3617, and 3614 steps, respectively. For the combined pcs, the length for the observed data set was 3783 steps, and the lengths for the randomized data sets were 7854, 7850, 7769, and 7761 steps. The results from the PTP tests were also highly significant in suggesting phylogenetic signal in the ITS data set (99 randomizations, all $\alpha = 0.01$). The ITS MPT length was 279 steps, whereas the lengths of the shortest trees based on the randomized data sets were 698, 695, 695, and 697 steps, respectively.

The RASA test for pc E35/M48 showed an observed slope (β_{obs}) of 19.62, an expected slope (β_{null}) of 8.69, and a test statistic t_{RASA} of 31.62, with 3399 degrees of freedom (df). The test for E35/M49 showed a β_{obs} of 20.33, a β_{null} of 8.53, and a t_{RASA} of 34.04 ($df = 3399$). The combined AFLP data sets showed a β_{obs} of 20.33, a β_{null} of 8.53, and a t_{RASA} of 34.04 ($df = 3399$). The ITS data set showed a β_{obs} of 12.88, a β_{null} of 6.70, and a t_{RASA} of 19.85, with 986 degrees of freedom. In all cases, t_{RASA} indicates significant phylogenetic signal ($\alpha = 0.05$).

MPT Support

The unweighted search on AFLP data set E35/M48 yielded 240 MPTs of 1902 steps, a CI of 0.279, an RC of 0.191, and an RI of 0.685. The unweighted search with data set E35/M49 yielded 1238 MPTs of 1835 steps, a CI of 0.272, an RC of 0.181, and an RI of 0.666. The unweighted search using the combined data from E35/M48 and E35/M49 yielded 6 MPTs of 3783 steps, a CI of 0.272, an RC of 0.183, and an RI of 0.671. The six trees based on the combined primer combinations (pcs) differed only in a few terminal branches.

One of the MPTs based on the combined pcs is depicted in Figure 1. The MPT shows two moderately supported clades of species that are in accordance with the generally applied morphology-based classification of Feráková (1977). The first clade comprises all subsection *Lactuca* species in the present study: *L. sativa*, *L. serriola*, *L. dregeana*, *L. altaica*, *L. aculeata*, *L. virosa*, and *L. saligna*. The second clade comprises the section *Mulgedium* species *L. tatarica* and *L. sibirica*, and section *Lactucopsis* species *L. quercina*. Within the first clade, a subclade with *L. sativa*, *L. serriola*, *L. dregeana*, and *L. altaica* is well supported (94% bootstrap support, 95% jackknife support, and a DI of 9 steps, respectively), as is a larger clade including *L. sativa*, *L. serriola*, *L. dregeana*, *L. altaica*, and *L. aculeata* (100%, 100%, 16 steps). Two clades within *L. virosa*, possibly identifying intraspecific taxa, also have high supports (100%, 100%, 23 steps; and 100%, 100%, 38 steps). Within the second clade, the subclade with *L. tatarica* and *L. sibirica* is well supported (92%, 95%, 11 steps). For all but a few species, the bootstrap and jackknife supports were 100%, and the DI values exceeded 18 steps (not shown on the MPT). The only exceptions were *L. virosa* (79%, 89%, 3 steps), and *L. sativa*, *L. serriola*, *L. dregeana*, and *L. altaica*. The latter four species are probably conspecific (Koopman et al., 2001), and accessions of these species are intermixed in the MPT.

In general, the bootstrap/jackknife values and the DI values are well correlated (analyses not shown), without any anomalous DI values. The one exception concerns the bootstrap/jackknife values for the clade with *sat/ser/dreg/alt*, *L. aculeata* and *L. virosa*. The bootstrap value of 87% and the DI value of 7 indicate support for this clade, but it is not supported in the jackknife analysis (49%). The discrepancy may be explained by the presence of *L. virosa* in this clade. Previous work indicated that the position of *L. virosa* relative to *L. sativa* (*sat*), *L. serriola* (*ser*), and *L. saligna* differs for different marker systems. This behavior was explained by postulating a hybrid origin for *L. virosa* (Koopman et al., 1998). Considering the fact that AFLP markers are a more or less random sample of the genome, the hybrid nature of the *L. virosa* genome may give rise to contradictory information in the AFLP data set. In the present study, the jackknife analysis may be more sensitive to this contradictory information than the bootstrap and decay analyses (see Giribet, 2003, for more discussion on this).

Congruence

In the PHT test, 17 out of 500 trees were longer than the sum of tree lengths for the original data sets. The corresponding error rate on rejecting the hypothesis of congruence between the AFLP and ITS data sets is $1 - (17/500) = 0.966$, meaning that the data sets show significant congruence at $P = 0.034$.

The search with ITS-1 sequences yielded 558 trees of 279 steps, a CI of 0.667, an RC of 0.584, and an RI of 0.876. The trees differed only in the relationships within the species and in the relationships between *L. sibirica* and *L. viminea*. Because these differences were not expected to have a major impact on the AFLP/ITS-1 tree comparisons, only one arbitrarily chosen tree (Figure 2) was used for further analysis. Figures 3 and 4 show the 50%
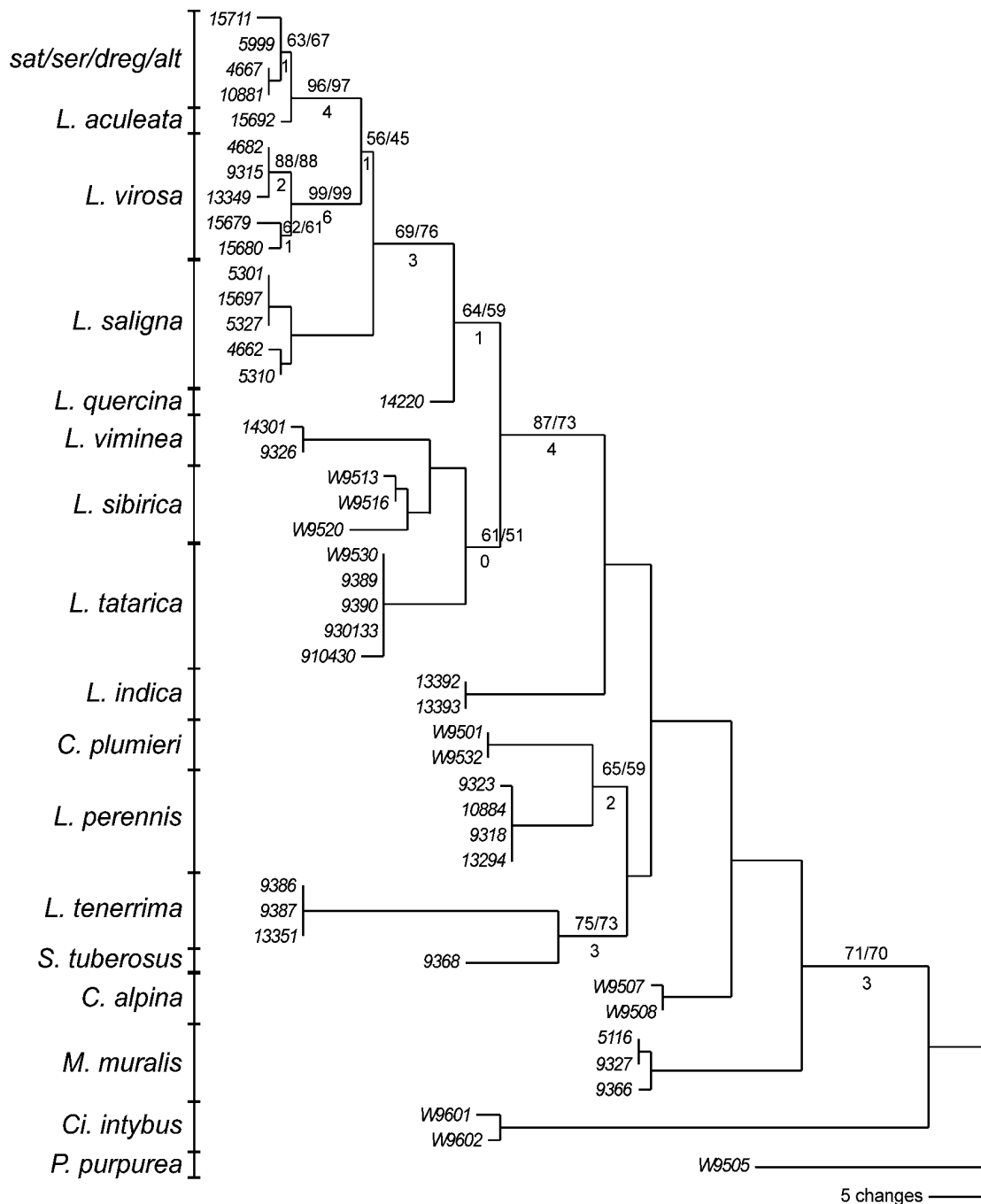
FIGURE 2.   One of the MPTs based on the unweighted Wagner parsimony search on the ITS-1 sequences. Above branches: bootstrap values/jackknife values. Below branches: branch supports. *sat/ser/dreg/alt*: group of intermixed and closely related species *L. sativa, L. serriola, L. dregeana*, and *L. altaica*. Genus abbreviations: *L = Lactuca, C = Cicerbita, Ci = Cichorium, M = Mycelis, S = Steptorhamphus, P = Prenanthes*.

majority rule consensus and the strict consensus ITS-1 and AFLP MPTs.

The Templeton test showed significant conflict in topologies between the AFLP and ITS-1 MPTs. Using the AFLP data set to compare the topologies, the AFLP MPT measured 3783 steps, the ITS MPT 4467. The AFLP MPT is significantly shorter (and thus incongruent) at $P < 0.0001$ (test statistic $T = 5185.5$, number of signed-ranks $N = 440$). Using the ITS-1 data set, the ITS-1 MPT measured 279 steps, the AFLP MPT 318. The AFLP MPT

is significantly longer (and thus the ITS tree (and thus incongruent) at $P < 0.0001$ ($T = 65$, $N = 34$).

Visual comparison of the ITS and AFLP MPT shows a general congruence for two moderately supported parts of the AFLP MPT. Similar to the AFLP MPT, the ITS MPT shows a clade with all subsection *Lactuca* species. This clade has a 69/76/3 support (bootstrap %, jackknife %, and DI, respectively), whereas the support for the subclades varies: the subclade with *L. sativa, L. serriola, L. dregeana*, and *L. altaica* has a 63/67/1 support, the larger
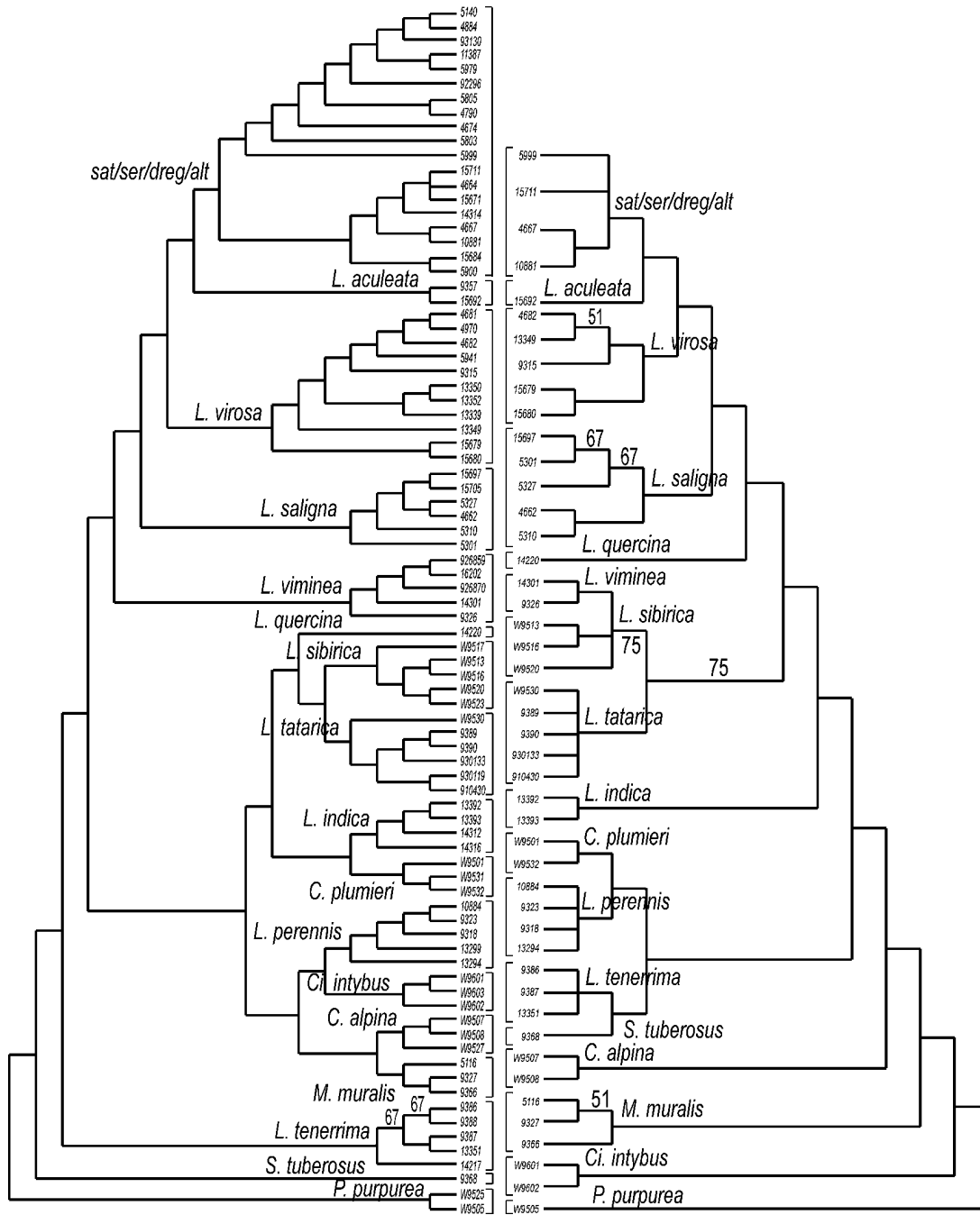
FIGURE 3.   The 50% majority rule consensus trees based on unweighted Wagner parsimony searches. Left: tree based on the combined AFLP primer combinations E35/M48 and E35/M49. Right: tree based on ITS-1 sequences. Above branches: percentage of trees showing the depicted topology. For branches on which no percentages are indicated, all trees showed the same topology.

clade including *L. sativa*, *L. serriola*, *L. dregeana*, *L. altaica*, and *L. aculeata* is supported with 96/97/4, and the two clades within *L. virosa* are supported with 88/88/2, and 62/61/1, respectively. The second of the moderately supported clades in the AFLP MPT is only partially reflected by the ITS MPT. In the AFLP MPT this clade comprises *L. tatarica*, *L. sibirica*, and *L. quercina*, whereas in the ITS MPT a clade is present comprising *L. tatarica*, *L. sibirica*, and *L. viminea*. The support values for the clade are low:

61/51/0, and there is no supported *L. tatarica/L. sibirica* subcluster.

Although their topologies are generally congruent, the exact taxonomic level at which groups are supported differs slightly for the AFLP and ITS MPTs. In both the AFLP and ITS MPTs, all species except the *sativa/serriola/dregeana/altaica* clade and the *virosa* clade have a 100% bootstrap and jackknife support. The DI for these groups is at least 19 steps in the AFLP MPT, but
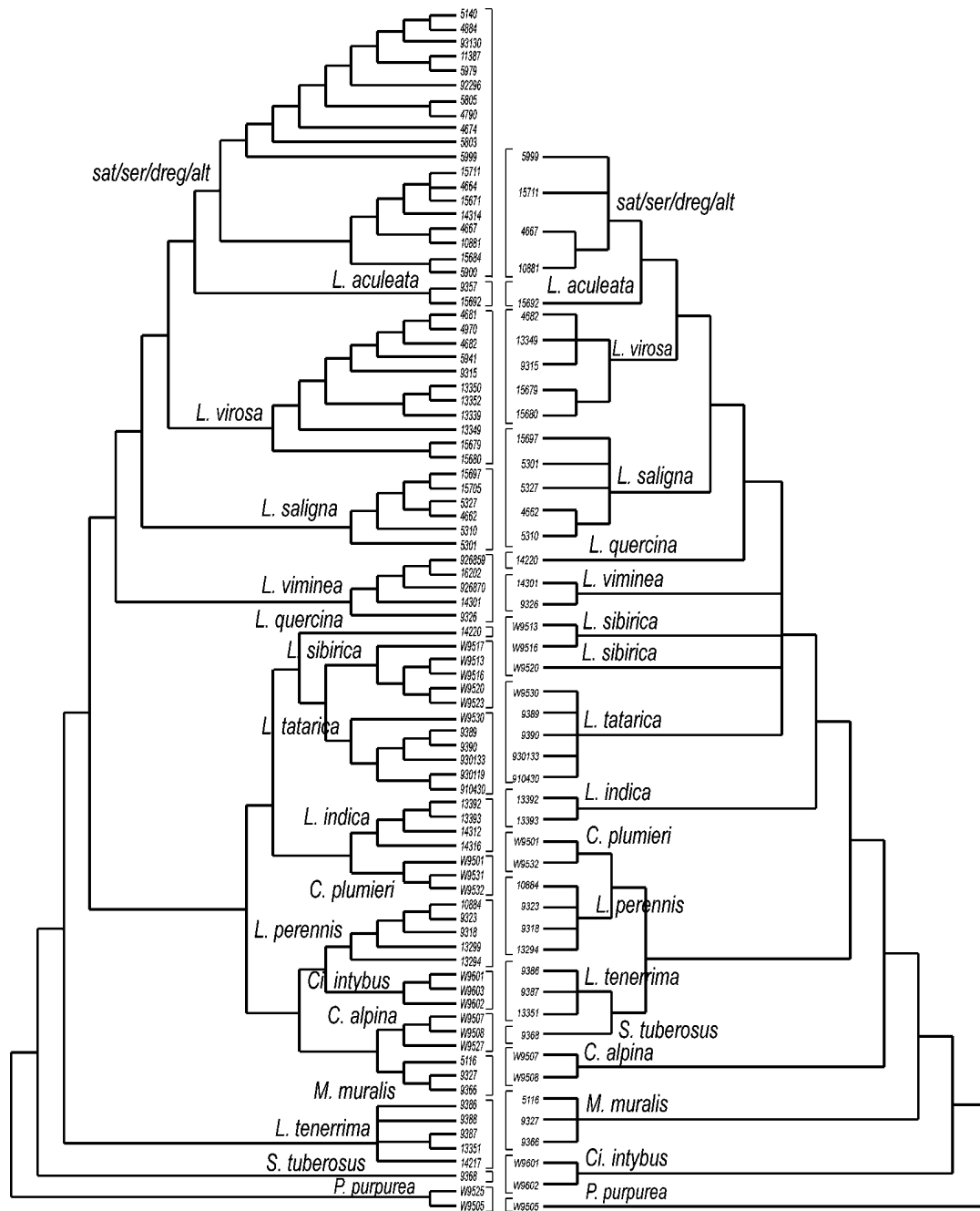
FIGURE 4. Strict consensus trees based on unweighted Wagner parsimony searches. Left: tree based on the combined AFLP primer combinations E35/M48 and E35/M49. Right: tree based on ITS-1 sequences.

much lower (down to 0 for *L. sibirica*) in the ITS MPT. The intraspecific differences in ITS-1 sequence range from 0 to 7 nucleotides, with many accessions sharing the same sequence. On the other hand, all accessions show different AFLP patterns. Apparently, AFLP markers provide more resolution at the intraspecific level than ITS-1 sequences do. The *sativa/serriola/dregeana/altaica* clade is well supported in the AFLP MPT (94/95/9), but much less in the ITS MPT (63/67/1). The ITS-1 divergence within the clade ranges from 0 to 3 nucleotides. For this clade, too, the moderate support in

the ITS MPT may indicate a lack of resolution for ITS sequences at lower taxonomic levels. A larger clade with *sativa/serriola/dregeana/altaica/aculeata* is well supported in both the ITS MPT (96/97/4) and in the AFLP MPT (100/100/16). The maximum ITS sequence divergence within this larger clade is 5 nucleotides. For the *sativa/serriola/dregeana/altaica/aculeata/virosa* clade and the *sativa/serriola/dregeana/altaica/aculeata/virosa/saligna* clade the support is only moderate, and decreases with increasing nucleotide diversity within the clades. The support is 87/49/1 for the clade including *L.*

virosa (with a maximum ITS-1 diversity of 18 nucleotides) and 75/68/5 for the clade including both *L. virosa* and *L. saligna* (with a maximum ITS-1 diversity of 19 nucleotides). The *L. virosa* clade consists of two subclades that are well supported in the AFLP MPT (100/100/23 and 100/100/38, respectively), but that are only moderately supported in the ITS MPT (88/88/2 and 62/61/1, respectively). The maximum ITS-1 sequence divergence within each of these subclades is 3 nucleotides. The clade including all *L. virosa* accessions shows the opposite pattern: it is well supported in the ITS MPT (99/99/6), but only moderately in the AFLP MPT (79/89/3). The maximum ITS-1 sequence divergence within this clade is 6 nucleotides. An explanation may be that within *L. virosa* the AFLP markers approach the maximum level of evolutionary differentiation at which they are informative, whereas the ITS markers approach their minimum level. In that case, the AFLP markers contain enough information to distinguish well-supported subclades within *L. virosa*, although their information content is too low for a high support of the relationship between the subclades. The ITS markers, on the other hand, contain not enough information for a high support of the subclades (probably due to a lack of variation), although they contain enough information to support the relationship between these clades. This interpretation of the support values in *L. virosa* is corroborated by the supports for the *L. sibirica/tatarica/quercina* clade, although the ITS-1 divergence within this clade is much higher. The closely related species *L. sibirica* and *L. tatarica* (both *Lactuca* section *Mulgedium*) are well supported by AFLP data (92/95/11), whereas the relationship of *L. sibirica/L. tatarica* with *L. quercina* (*Lactuca* sect. *Lactucopsis*) has only a moderate support (74/75/4). A comparison with the ITS markers is hampered by the fact that the ITS MPT indicates different (and poorly supported) relationships for *L. sibirica/L. tatarica/L. quercina*, but the maximum ITS-1 sequence divergence within the *L. sibirica/tatarica/quercina* clade is 19 nucleotides. None of the relationships among species outside the above clades are supported in the AFLP MPT, but some have a moderate support in the ITS tree. This difference in support may again indicate that ITS markers are informative at somewhat higher taxonomic levels (deeper in the tree) than are AFLP markers.

Dollo Parsimony

The Dollo parsimony analysis resulted in three MPTs of 7969 steps, a CI of 0.129, an RC of 0.116, and an RI of 0.894. The trees only differed in some intraspecific branches within the *L. perennis* and the *L. sativa/serriola/dregeana/altaica* clade. One of the MPTs is shown in Figure 5. Branches that collapse in the strict consensus tree are marked with dotted lines. A comparison of the Dollo and unweighted parsimony trees shows some striking differences. (1) In the unweighted analysis, *L. aculeata* is a well-supported sister group of the *L. sativa/serriola/dregeana/altaica* clade. In the Dollo analysis, a position of *L. aculeata* within the *L. sativa/serriola/dregeana/altaica* clade is well supported. (2) In the unweighted analysis, all *L. virosa* accessions

group together, with *L. virosa* accessions 15679 and 15680 as a sister group of a clade containing the other *L. virosa* accessions. In the Dollo analysis, *L. virosa* accessions 15679 and 15680 are the sister group of the *L. saligna* clade. (3) In the unweighted analysis, *L. sibirica* and *L. tatarica* group together with high bootstrap support (92%). In the Dollo analysis, *L. sibirica* groups with *L. quercina* (although the bootstrap support for this clade is low).

Weighted Parsimony

The weighted parsimony analysis resulted in two MPTs of 3800 steps. The trees only differed in the position of *L. tenerrima* accessions CGN 9386 and CGN 9388 relative to each other. In the one tree they group together, whereas in the other they branch off sequentially. Each of the MPTs is topologically identical to one of the MPTs from the unweighted parsimony search; only the branch lengths differ. The MPT that is topologically identical to Figure 1 is depicted in Figure 6. The MPTs from the weighted parsimony analysis are 17 steps longer than those from the unweighted search (3800 and 3783, respectively). Comparison of the branch lengths in the weighted and unweighted parsimony trees shows that the extra tree length is not concentrated on particular branches but is distributed throughout the entire tree. Moreover, although some branches are longer in the MPTs from the weighted parsimony search, other branches are longer in the MPTs from the unweighted search. For easy comparison, Figure 6 shows both the branch lengths from the weighted parsimony tree (derived from Figure 6 itself, given before brackets), and the lengths of the corresponding branches in the unweighted parsimony tree (derived from Figure 1, given between brackets). The identical topologies of the weighted and unweighted MPTs illustrate that the applied weight of 1.02 in favor of losses was too small to influence the tree topology.

DISCUSSION

Phylogenetic signal in *Lactuca* s.l. AFLP data sets was examined using three approaches. First, phylogenetic signal was tested directly in the data sets. Using TLD, PTP, and RASA, significant phylogenetic signal was detected in the data sets generated with both primer combinations (pcs). The signal for the combined data sets was also significant, indicating that the signal in the separate data sets was not conflicting. The similarity in results for RASA, TLD, and PTP testing is encouraging in that it suggests these different methods are all measuring similar properties of the data, despite various objections having been made against the methodology. This finding is significant because a literature search into heuristic parsimony studies using AFLP markers showed that testing AFLP data sets for phylogenetic signal prior to phylogenetic analysis is not common practice. Perhaps a reason for this is the assumed difficulty to measure phylogenetic signal. An exception to this general trend are the studies by Giannasi et al. (2001) and Despres et al. (2003) who used TLD to test for phylogenetic signal in their AFLP data sets. They reported significant signal. Giannasi et al.
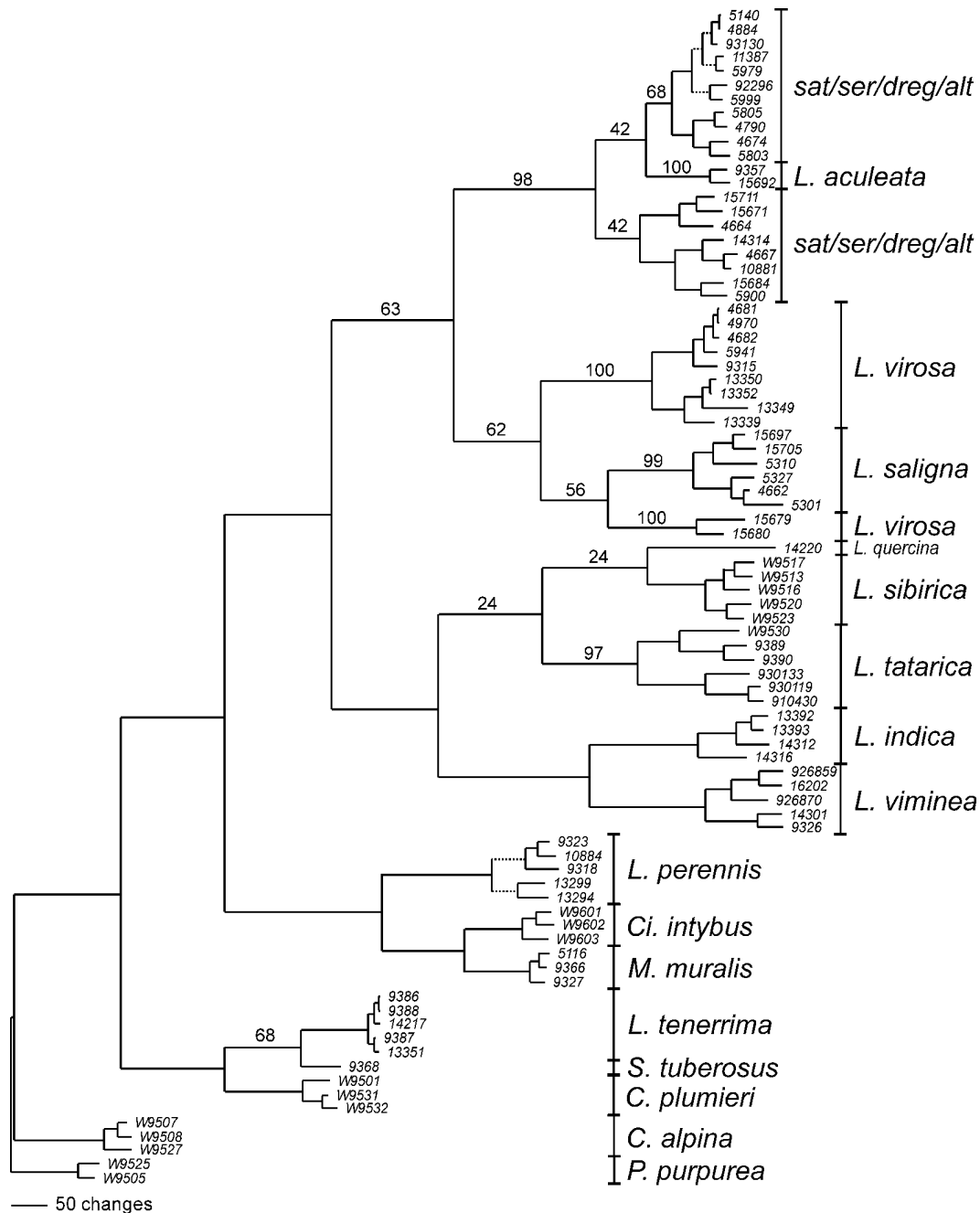
FIGURE 5.   One of the MPTs based on the Dollo parsimony search on the combined AFLP primer combinations E35/M48 and E35/M49. Above branches: bootstrap values (50% and higher; values below 50% are only depicted for the *sat/ser/dreg/alt* and *L. sibirica/L. tatarica/L. quercina* groups). *sat/ser/dreg/alt*: group of intermixed and closely related species *L. sativa, L. serriola, L. dregeana*, and *L. altaica*. Genus abbreviations: *L = Lactuca, C = Cicerbita, Ci = Cichorium, M = Mycelis, S = Steptorhamphus, P = Prenanthes*.

(2001) tested a data set comprising 27 *Trimeresurus* accessions (four species) and reported a g_1 value of -0.66. This value corresponds to a $P < 0.001$, indicating a highly significant result for the presence of phylogenetic signal in the data set. Despres et al. (2003) tested a data set of 34 individuals of 11 *Trollius* species and reported a significant g_1 of -0.35.

Second, branch supports for unweighted Wagner parsimony MPTs were compared. The MPT of the combined

Lactuca s.l. data sets showed two large clades with moderate support. Within these clades, various smaller groups showed high support, as did the clades for the individual species. These results seem to confirm the presence of phylogenetic signal in the data sets, but also indicate that the inferred signal is restricted to certain taxon relationships. The literature survey showed that in most phylogenetic AFLP analyses some kind of support is determined, usually bootstrap values. The general picture
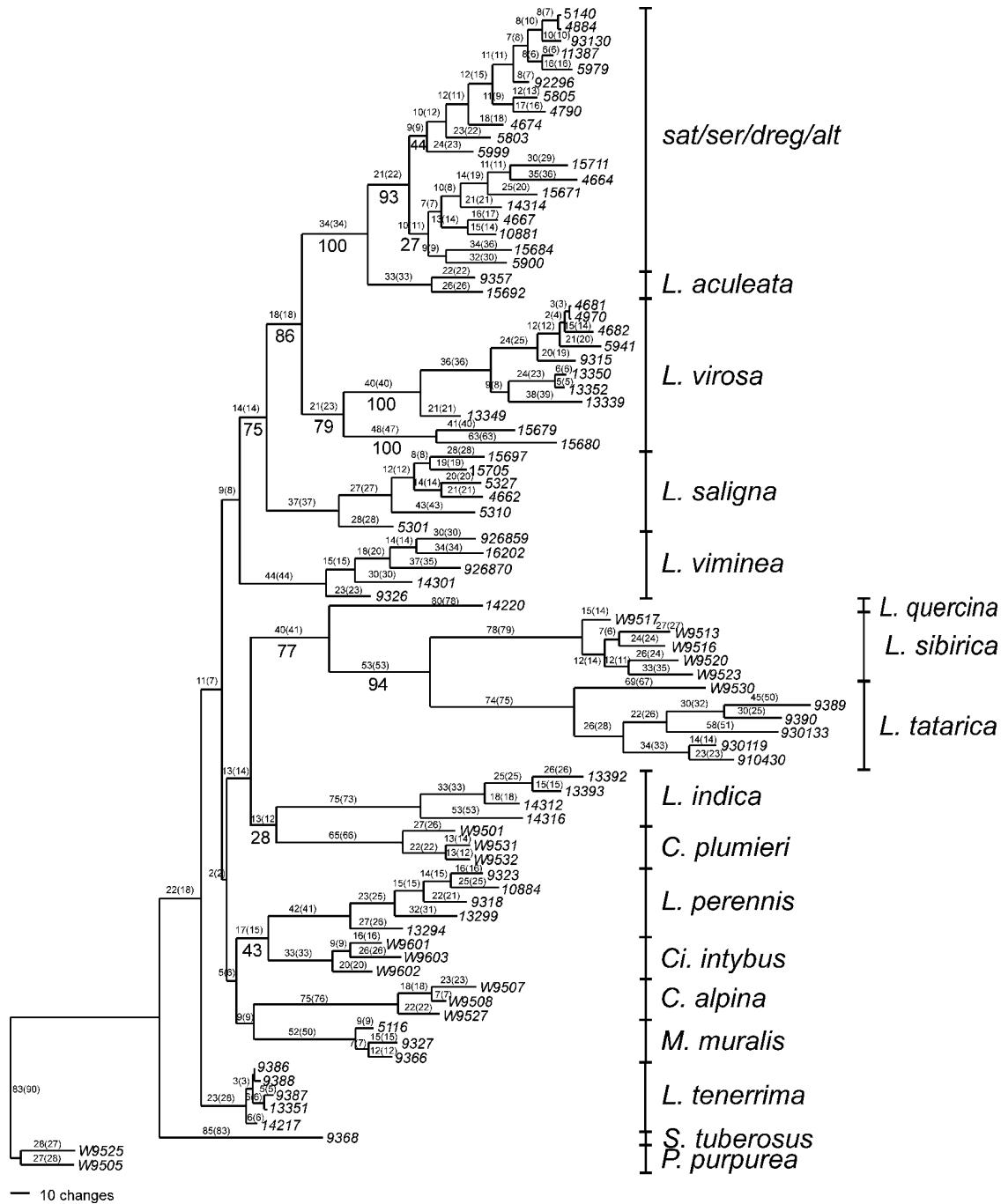
FIGURE 6.   One of the MPTs based on the weighted parsimony search on the combined AFLP primer combinations E35/M48 and E35/M49. Above branches, before brackets: branch lengths. Above branches, between brackets: length of the corresponding branch in the unweighted parsimony tree (Fig. 1). Below branches: bootstrap values (only 50% and higher). *sat/ser/dreg/alt*: group of intermixed and closely related species *L. sativa*, *L. serriola*, *L. dregeana*, and *L. altaica*. Genus abbreviations: *L = Lactuca, C = Cicerbita, Ci = Cichorium, M = Mycelis, S = Steptorhamphus, P = Prenanthes*.

from these studies is consistent with our finding that strength of phylogenetic signal is not evenly dispersed across reconstructed trees for *Lactuca*. These studies also demonstrate that the presence of well-supported topologies is a general phenomenon in MPTs based on AFLP data.

Third, congruence of AFLP and ITS data sets and MPT topologies was examined. The PHT showed a signifi-

cant congruence between the AFLP and ITS data sets. Comparison of the AFLP and ITS MPTs demonstrated that the moderately supported parts of the MPTs showed a general similarity, although some differences also existed. These differences were reflected in results from the Templeton test, which indicated significant topological incongruence. A closer examination of the similarities revealed that the well-supported clades in the AFLP

MPT correspond to ITS-1 sequence divergences of 0 to 7 nucleotides. This result suggests that robust phylogenetic hypotheses can be constructed from AFLP data for accessions that are 0 to 7 ITS-1 nucleotides apart. The moderately supported clades correspond to sequence divergences of 6 to 19 nucleotides, suggesting that AFLP data for accessions that are 6 to 19 ITS-1 nucleotides apart do contain phylogenetic information, but not enough to construct a robust phylogenetic hypothesis. Potential reasons for the low information content may be that the signal/noise ratio for these data is too low, or that the amount of data is not sufficient. In the latter case, addition of more data (primer combinations) could increase the support for the moderately supported clades. The absence of supported clades for groups of accessions with ITS-1 sequence divergences above 19 nucleotides may indicate a lack of phylogenetic information in AFLP patterns above this level of divergence. Scatterplots of ITS-1 distances versus AFLP distances (not shown) revealed that an ITS-1 distance of 19 nucleotides corresponds to a distance of approximately 300 AFLP bands. Given the fact that the combined AFLP data set contained 1030 variable bands, the maximum level of AFLP variation within supported groups is approximately 30%.

The phylogenies discussed above were constructed using unweighted parsimony, assuming equal probabilities of loss and gain of restriction sites. For restriction site data such as AFLPs, this assumption is violated, and therefore Dollo parsimony was used as an alternative (see DeBry and Slade, 1985). Using Dollo parsimony, the assumption is that a restriction site can be gained only once, but that it can be lost many times. The MPTs resulting from the Dollo parsimony analysis differed from the unweighted MPTs mainly in the position of three groups of accessions: *L. aculeata*, *L. virosa* accessions 15679 and 15680, and *L. sibirica*. In all three cases, the positioning in the unweighted MPTs is in accordance with the ITS MPTs and with independent taxonomic data (reviewed in Feráková, 1977, and Koopman et al., 1998), whereas the positioning in the Dollo MPTs is not. These results suggest that the unweighted parsimony criterion better fits the loss/gain probabilities in the *Lactuca* s.l. AFLP data set than the Dollo criterion does.

As an alternative to both unweighted and Dollo parsimony, the *Lactuca* AFLP data set was analyzed using weighted parsimony (or relaxed Dollo) according to Jansen et al. (1991), Holsinger and Jansen (1993), and Holsinger (personal communication), with a weight of 1.02 in favor of losses. The resulting tree topology was identical to that of the unweighted MPT, suggesting that either the loss/gain ratios in the AFLP data set are more or less equal, or that the applied weight is not appropriate for analyses of these data for some other reason. Further study is needed to understand the loss/gain ratio of fragments in AFLP data sets.

The present study showed a general congruence of the AFLP-based and ITS sequence-based MPTs, indicating the presence of phylogenetic signal in the AFLP data set. A detailed comparison of the ITS and AFLP MPTs showed that AFLPs are more variable markers than ITS

sequences, providing phylogenetic information where ITS sequences are too conserved. An extensive literature survey of biosystematic studies employing both ITS and AFLP markers showed that congruence of AFLP and ITS trees is the general rule in fungi, oomycetes, plants, and bacteria. However, in several cases the general congruence is also accompanied by local conflicts in topology for some of the species. Nevertheless, the congruence of ITS and AFLP trees can be used to establish the level of variation at which AFLP data can provide reliable relationship information.

Fungi

Tredway et al. (1999) examined seven species of *Epichloë*, *Neotyphodium* (the anamorph of *Epichloë*), and *Balansia* (Clavicipitaceae) using heuristic parsimony analyses of AFLP and ITS data. MPTs for these data were congruent as to the relationships among *Epichloë* and *Balansia* accessions, but the AFLP MPT was more resolved. The AFLP MPT showed poorly supported clades (54% or less) for species differing by zero to one nucleotide or by more than six nucleotides, and well-supported clades (87%) for species differing by three or four ITS nucleotides. The MPTs were entirely in conflict regarding the relationships among the *Neotyphodium* accessions. According to Tredway et al. (1999), this conflict may result from vegetative hybridization between *Neotyphodium* and *Epichloë*. Such hybridization events can result in an evolutionary history for ITS sequences that is not necessarily similar to that of the genome as a whole (see Tredway et al., 1999, for a more detailed discussion). Bakkeren et al. (2000) determined ITS sequences of 13 species of *Ustilago*, *Sporisorium*, and *Tilletia* (Ustilaginomycetes) and examined a subset of eight species using AFLP markers. The eight species form a single clade in the ITS tree, consisting of two subclades. The one subclade consists of two species and is basal to the other subclade that shows a polytomy of three smaller clades. The three smaller clades in the polytomy are a branch with one species, a resolved clade with two species, and a polytomy with three species, respectively. The AFLP tree shows the same clades as the ITS tree, but both polytomies present in the ITS tree are resolved in the AFLP tree. The AFLP tree shows poorly supported clades for species with identical ITS sequences (62%) or sequences differing seven nucleotides or more (70%), and well-supported clades (99% to 100%) for species differing zero to five nucleotides.

Montiel et al. (2003) used neighbor joining to analyze AFLP data on 24 isolates of four *Aspergillus* species. The AFLP-based relationships are in line with those apparent from ITS-2 sequencing, but AFLP markers provide a better resolution. The AFLP tree shows two large clades with a 100% bootstrap support. The first clade contains a mixture of *A. oryzae* and *A. flavus* accessions. The accessions within this clade show two types of ITS-2 sequences, differing by one gap: one type is shared by both species, the other type is confined to some of the *A. flavus* accessions. The second AFLP clade

comprises three subclades: one with only *A. sojae* accessions, one with only *A. parasiticus* accessions, and one with a single anomalous *A. parasiticus* accession. These subclades and their interrelationships all have a 100% bootstrap support in the AFLP tree, whereas they differ by zero to two ITS-2 nucleotides. The ITS-2 difference between the two larger clades ranges from five to seven nucleotides.

Cluster analyses of (dis)similarities derived from AFLP profiles also provide insight into the potential of phylogenetic analysis of AFLP data. Although interpretation of some of the studies is not straightforward and support values are not always included, a general picture emerges. First, accessions with identical or similar ITS sequences usually cluster together in the AFLP analyses. This was found by Gräser et al. (1999) for *Trichophyton* and by Gräser et al. (2000) for *Microsporum*, but detailed comparison of ITS and AFLP results is complicated by the fact that the AFLP results are analyzed and presented together with results obtained with PCR fingerprinting. In a study of Castella et al. (2002), 66 *Penicillium* accessions clustered in two large AFLP groups. Their ITS sequences showed only two variable positions, resulting in three different sequences. One sequence was common to both groups; the two remaining sequences were group specific. Wyand and Brown (2003) examined 63 isolates of four formae speciales of *Blumeria graminis* and detected three different ITS sequences. One sequence was shared by all f. sp. *tritici* and f. sp. *secalis* accessions, whereas the *tritici/secalis* sequence differed in 14 nucleotides from that of f. sp. *avenae*, and in 11 nucleotides from that of f. sp. *hordei*. The ITS sequences of f. sp. *avenae* and f. sp. *hordei* showed 17 nucleotide differences. The AFLP patterns of f. sp. *tritici* and f. sp. *secalis* were reported to be "similar," whereas the differences between the *tritici/secalis*, *avenae*, and *hordei* AFLP patterns were so large that the patterns could only be analyzed separately for each forma specialis. Second, AFLP patterns are usually more variable than ITS sequences. In a study of Bao et al. (2002), four groups of ITS sequences were found among 42 *Fusarium oxysporum* strains. For groups I, II, and IV, all accessions within a group shared the same sequence. The sequences in group III differed from each other and from those in group IV by single nucleotide polymorphisms only. In contrast, the AFLP analysis revealed considerable intragroup variation, with simple-matching similarities down to 0.7 (as estimated from the UPGMA dendrogram). Similar results are reported by Abeln et al. (2002) for *Phoma exigua*, Castella et al. (2002) for *Penicillium*, and Wyand and Brown (2003) for *Blumeria graminis*. Third, within the groups that show similar or identical ITS sequences, the relative positions of the accessions within clusters/clades are usually different in the AFLP and ITS trees (Bao et al., 2002; Abeln et al., 2002). Fourth, considering the relationships among the groups with similar or identical ITS sequences, there is a general congruence between ITS phylogenies and AFLP trees. This is apparent from the study of Gräser et al. (2000) for clades that are two to four nucleotides apart, but again the comparison is hampered by the fact that

AFLP and PCR fingerprinting results are presented together. In a study of Douhan and Rizzo (2003), the AFLP data allowed clustering of 20 *Hypomyces* isolates into two moderately supported groups (73% and 88% bootstrap). Each group was subdivided into two well-supported subgroups (99% to 100% bootstrap). Both the groups and subgroups were reflected in the ITS NJ tree, but without much support.

In addition to studies showing congruence at low levels of divergence, several studies indicate that above a certain level of ITS divergence, AFLP markers are sometimes too variable to properly detect relationships among groups. (Gräser et al., 1999, for clades that are 6 to 10 ITS nucleotides apart; Wyand and Brown, 2003, where groups of *Blumeria* accessions differing by 11 to 17 ITS nucleotides could not be matched).

Four studies are not in concordance with this general picture. In a study on *Eutypa*, DeScenzo et al. (1999) report ITS differences as large as 0 to 46 nucleotides between strains, whereas Bao et al. (2002) report 3 to 46 nucleotide differences within *Fusarium*. Nevertheless, the large groups in their AFLP trees are still in general accordance with those of the ITS MPT. Aquino de Muro et al. (2003) examined 48 isolates of *Beauveria bassiana* with AFLP markers and a subset of 26 using ITS sequences. They reported a maximum ITS sequence diversity of less than 2%, with many accessions showing identical sequences. The AFLP markers were much more variable, yielding different AFLP patterns for all accessions. Both the ITS and AFLP trees show a sequential branching-off of (groups of) accessions, but there seems to be little similarity in branching order between the trees. An anomalous study by Arroyo-Garcia et al. (2003) did not detect any supported structure in neither the ITS tree nor the AFLP tree in a study on *Fusarium*.

Oomycetes

In oomycetes, the most informative studies are those by Werres et al. (2001) and Mirabolfathy et al. (2001). Werres et al. (2001) examined 14 isolates of *Phytophthora ramorum* together with 8 related *Phytophthora* species. In both the ITS and AFLP trees the accessions clustered together according to their species designation. The distance between the most closely related species (*P. ramorum* and *P. lateralis*) in the ITS tree was 11 nucleotides (far more for less related species), and all clades had a moderate to good bootstrap support (73% to 100%). However, the AFLP tree showed totally different species relationships, even between *P. ramorum* and *P. lateralis*. The most closely related species in the AFLP UPGMA tree were *P. lateralis* and *P. cinnamomi*, with a Nei and Li distance of 0.65. Although not discussed by Werres et al. (2001), the discrepancy between the trees can be explained by assuming that the AFLP markers are too variable to reflect species relationships in *Phytophthora*. The results of Mirabolfathy et al. (2001) support this view. Their ITS neighbor-joining tree shows the well-supported relationships (>89% bootstrap) of eight *Phytophthora* species. The Pearson/UPGMA dendrogram supports the ITS-based

relationships for the four most closely related species (clustering at 0.3), but differs for the more distantly related species. The larger variability of AFLP markers relative to ITS sequences is also apparent from a study of Ivors et al. (2004) showing 31 distinct AFLP genotypes among 85 *Phytophthora ramorum* isolates, although all isolates had identical ITS sequences. Brasier et al. (1999) indicated a congruence of ITS and AFLP relationships for *Phytophthora* species differing in 12 positions (but not all varying among all species). Studies by Chowdappa et al. (2003a, 2003b) showed that ITS and AFLP markers identify similar groups in *Phytophthora*. The results of Rehmany et al. (2000) on *Peronospora* are in line with the observations on *Phytophthora*. In a study on 33 isolates of *Peronospora parasitica* and two related outgroup species, Rehmany et al. (2000) found five different ITS-1 sequences, each corresponding to a distinct group in the Jaccard/UPGMA AFLP tree. The groups with AFLP similarities of 0.3 or larger were well supported; groups with 0.12 or less similarity were not supported (<58% bootstrap support).

Plants

In plants, phylogenetic analysis of AFLP data seems to be more common than in Fungi and oomycetes. Most studies use heuristic parsimony, but neighbor joining is also frequently used and, more recently, Bayesian inference. Focusing on the parsimony studies, the general picture is comparable to that in Fungi. Most importantly: (1) Similar clades/groups are detected by both ITS and AFLP markers. (2) AFLP markers are usually more variable than ITS sequences. (3) There is a general congruence of ITS and AFLP MPTs, although exceptions may occur for individual accessions. Similar to the situation in Fungi, the congruence of ITS and AFLP MPTs depends on the evolutionary distance between the accessions involved. In *Phyllostachys* (Bambusoideae, Poaceae) (Hodkinson et al., 2000), the relationship between the sections *Phyllostachys* and *Heteroclada* (32 changes apart) is recovered in the AFLP MPT with 97% bootstrap support, as is the relationship between the two species in Sect. Heteroclada (12 changes apart, 81% support). On the other hand, the relationships within Sect. *Phyllostachys* (1 to 9 changes) are usually not recovered in the AFLP tree. In the study of Xu and Sun (2001) the accessions of *Amaranthus* fall in three clades that are 5 to 15 changes apart, and the relationships among these clades are recovered in the AFLP tree (100% bootstrap support). The accessions of *Trollius* studied by Despres et al. (2003) are 0 to 6 changes apart, but the resolution in their AFLP tree was too low to resolve among species relationships. The study by Pelser et al. (2003) on *Senecio* included 10 species that were present in both the ITS and AFLP analysis. Six species used as ingroup in the AFLP analysis differed by 1 to 20 ITS nucleotides, and the relationships among these species were totally unresolved in the strict consensus ITS tree. The AFLP strict consensus tree is partially resolved, showing well-supported relationships (97% and 71% bootstrap) between species that are 10 and 16 ITS

nucleotides apart. A polytomy in the AFLP tree includes four species that are 1 to 12 ITS nucleotides apart. The relationship of two outgroup species differing by 34 ITS nucleotides is well supported (100%). In a study on *Ipomoea* (Huang et al., 2002), the ITS consensus tree showed a grade comprising two species and a large polytomy. The polytomy consists of eight single accessions, a clade of two species, and a clade of three species. The grade is reflected in the AFLP MPT, but the polytomy is resolved there. Both groups that were nested within the polytomy in the ITS tree are present in the AFLP tree, with 78% and 81% bootstrap support.

Neighbor-joining analyses of AFLP data generally show a similar picture as maximum parsimony analyses. In a study by Hodkinson et al. (2002) on *Miscanthus*, ITS polymorphisms suggested that $M \times giganteus$ was a hybrid of *M. sinensis* and *M. sacchariflorus*. The AFLP NJ tree confirmed this view, showing $M \times giganteus$ in an intermediate position between *M. sinensis* and *M. sacchariflorus*. A study of Beardsley et al. (2003) in *Mimulus* showed topological congruence between the AFLP tree and the ITS MPT for ITS clades that are an estimated 1 to 14 changes apart. Beardsley et al. (2003) also conducted a heuristic parsimony analysis and reported that it recovered the same well-supported clades as did the neighbor-joining analysis. The work of Semerikov et al. (2003) on *Larix* showed well-supported topological congruence of the ITS MPT and the AFLP NJ tree for a group of six species differing by 1 to 49 nucleotides, but also conflicting sister group relationships in a group of three species differing by 4 to 20 nucleotides. Proposed causes of the conflicts were possible sampling error, differences in genome evolution between the species, and AFLP markers being too variable. A study of El-Rabey et al. (2002) on *Hordeum* highlights genome evolution events as a possible cause of discrepancies between AFLP and ITS phylogenies. The AFLP and ITS NJ trees show major topological incongruences, and only the AFLP topology is in accordance with independent data. Their explanation is that single loci (ITS sequences) can be misleading in the reconstruction of complex evolutionary histories such as those present in *Hordeum*. In contrast, these complex histories are better reconstructed using a multilocus sampling strategy as obtained with AFLP. In addition to congruence between ITS MPTs and AFLP phylogenies, congruence of ITS MPTs and AFLP phenograms was reported for *Datura* and *Brugmansia* (Mace et al., 1999a), *Solanum* (Mace et al., 1999b), *Cichorium* (Kiers et al., 1999), *Oxalis* (Tosto and Hopp, 2000; Emshwiller and Doyle, 1998), and *Soldanella* (Zhang et al., 2001). In such cases when representing complex evolutionary histories of plants, phylogenetic network methods such as Split Decomposition and NeighborNet appear to have promising potential (Perrie et al. 2003a, 2003b).

Bacteria

Complex evolutionary histories may be found in plants, but they are most obvious in bacteria, where exchange of genetic material between species is a common

phenomenon. In *Bradyrhizobium*, a general topological congruence of ITS NJ trees and AFLP phenograms was demonstrated, but incongruences occurred for many individual strains/species. Willems et al. (2001, 2003) show a *Bradyrhizobium* ITS tree consisting of two major clades and various subclades. The subclades largely correspond to previously determined AFLP groups (Willems et al., 2000). However, the level of variation within the AFLP groups corresponding to either of the ITS clades is strikingly different. One clade shows low AFLP similarity (50% to 55%, Willems et al., 2000) and moderate ITS sequence similarity (85% to 100%, Willems et al., 2001; >64.6%, Willems et al., 2003), whereas the other clade shows a higher AFLP similarity (55% to 90%, Willems et al., 2000) and a higher ITS similarity (94% to 100%, Willems et al., 2001; >92.5%, Willems et al., 2003). Both the incongruences between the ITS and AFLP trees and the high levels of variation in the one ITS clade are explained by lateral gene transfer, exchange, and recombination. In case of incongruences, the ITS results were better in line with DNA-DNA hybridization data than were the AFLP results. Therefore, ITS data are considered a more reliable indicator of taxonomic affinity. This is rather surprising, because one would expect the multilocus AFLP approach to best reflect the whole-genome DNA-DNA hybridization similarities (as seems to be the case in *Hordeum*). Clearly, the issue of species description based on independent multilocus (AFLP) data versus ITS data deserves attention in future studies in bacteria.

Summary

In summary, AFLP markers in Fungi seem to be most reliable at a level of variation corresponding to a difference of two to five ITS nucleotides. For genotypes differing by zero to two nucleotides, AFLP-based relationships are often not in accordance with ITS-based relationhips, or are poorly supported. The lack of congruence at this level of variation is probably caused by the ITS sequence variation being too small to yield reliable ITS-based relationships. The lack of support is more serious and may indicate a lack of sufficient signal in the AFLP data. However, the fact that AFLP markers usually show resolution even in the absence of ITS variation indicates that a limited amount of signal may be present. Increasing the number of AFLP markers (i.e., primer combinations) will therefore probably still result in well-supported topologies. At a level of six or more ITS nucleotide differences, AFLP markers are too variable to detect any phylogenetic signal. In oomycetes, detailed information on the levels of variation is scarce, but it may indicate that AFLP markers are useful at a higher level of ITS variation (up to 12 nucleotides?) than is the case in Fungi. However, additional information is needed for a better estimate in this group. In plants, AFLP-based relationships among genotypes that are 10 to 30 (10 to 35?) ITS nucleotides apart are usually recovered with good bootstrap support. For genotypes with ITS differences above 30 to 35 nucleotides, AFLP markers seem too variable to be useful as phylogenetic markers. For genotypes with differences below 10 ITS nucleotides, AFLP markers failed to detect congruent or well-supported relationships in several cases. However, in other cases well-supported AFLP relationships were recovered, indicating that phylogenetic signal can also be present when genotypes differ by less than 10 ITS nucleotides. Therefore, increasing the number of AFLP markers will probably yield better supported relationships at lower levels of ITS divergence, too. In bacteria, AFLP profiles should be interpreted with caution, because processes such as lateral gene transfer, exchange, and recombination complicate the issue of species identity.

Conclusions

All three approaches used in the present study indicated statistically significant phylogenetic signal in the *Lactuca* s.l. data sets, although significant conflict also existed in some parts of the AFLP and ITS MPTs. As stated in the introduction, restriction fragment markers have a number of limitations that theoretically could lead to a loss of phylogenetic signal in AFLP data sets. The presence of significant signal in the *Lactuca* s.l. test data sets indicates that, in practice, the influence of these limitations is limited. It should be noted, however, that the present conclusions only apply to data sets with closely related species, because AFLP markers are highly variable and the proportion of nonhomologous fragments increases with taxonomic divergence (O'Hanlon and Peakall, 2000). In data sets including more distantly related taxa, proportions of nonhomologous fragments among taxa may become so high that phylogenetic signal is lost. However, data sets can be tested on the presence of phylogenetic signal, and (parts of) data sets without signal can be discarded. The exact level of divergence that can be studied varies among taxa and should be determined for each group separately. A procedure to test for significant phenetic similarity among individual genotypes in AFLP data sets is described in Koopman and Gort (2004).

An extensive literature survey revealed topological congruence of (parts of) AFLP and ITS trees in a wide range of taxa, indicating the presence of phylogenetic signal in virtually all AFLP data sets. Gross topological incongruence of AFLP and ITS MPTs was reported in only a few cases, but these incongruences were usually limited to very specific parts of the trees. Thus, the results on *Lactuca* s.l. are generally corroborated by a large amount of data from the literature, indicating that the present study is representative for AFLP data sets in general.

The present study indicates that AFLP markers can be applied at similar taxonomic levels as ITS sequences, but that AFLP markers are generally somewhat more variable. Analyses of the *Lactuca* s.l. data sets, and inventory of the literature, enabled a rough estimate of the level of divergence at which ITS and AFLP markers are phylogenetically informative. In Fungi, AFLP markers are most valuable when the corresponding ITS variation is between 0 and 5 nucleotides. However, for ITS divergence levels between 0 and 2 nucleotides, the number

of AFLP markers should be increased in order to obtain sufficient phylogenetic information. In oomycetes, AFLP markers may be useful at divergence levels of up to 12 ITS nucleotides, but more information is needed for a better estimate. In plants, AFLP markers are most reliable at ITS divergence levels between 10 and 30 (35?) nucleotides, whereas above this level they are too variable. At divergence levels below 10 ITS nucleotides, an increased number of AFLP markers is needed to obtain reliable phylogenetic estimates. In bacteria, AFLP markers should be used with caution, because they may be sensitive to the genome rearrangement processes that are common in this group. Nevertheless, phylogenetic information seems to be present in bacterial AFLP data sets, too.

All in all, AFLP markers appear to be a valuable source of phylogenetic information among closely related taxa. Being somewhat more variable than ITS sequences, they are well suited for studying phylogenetic relationships when ITS sequences are too conserved. In general, however, for studies at that level to generate reliable phylogenetic hypotheses, the numbers of AFLP markers employed should be increased relative to what is presently customary.

RECOMMENDATIONS

As was stated in the introduction, two of the most important limitations of AFLP markers are their possible lack of independence and their possible lack of homology. Several studies have indicated that the problem of nonhomology is related to the phylogenetic distance among the taxa involved (see introductory section of this article), so it can be diminished by excluding taxa that are too distantly related. An attempt to estimate this distance is provided in Koopman and Gort (2004). No elaborate study into the nonindependence of AFLP fragments has been conducted yet. Given the importance of nonindependence of characters in parsimony analysis, detailed examination of the nonindependence of AFLP markers seems worthwhile.

Another serious drawback of AFLP markers as phylogenetic characters is their unequal loss-gain probability (see introductory section of this article). In the present study, two approaches were tested that theoretically could compensate for this unequal probability: Dollo parsimony and weighted parsimony. However, Dollo parsimony was too restrictive, whereas weighted parsimony yielded the same tree topology as did unweighted parsimony. Although it is merely speculation, these results could indicate that a proper weighing scheme should indicate weights somewhere between these two extremes. If this is true, the development of more refined weighing schemes could prove a fruitful way to more fully exploit the phylogenetic potential of AFLP markers.

ACKNOWLEDGMENTS

REFERENCES

Abeln, E. C. A., A. M. Stax, J. De Gruyter, and H. A. Van der Aa. 2002. Genetic differentiation of *Phoma exigua* varieties by means of AFLP fingerprints. Mycol. Res. 106:419–427.

Álvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417–434.

Angiolillo, A., M. Mencuccini, and L. Baldoni. 1999. Olive genetic diversity assessed using amplified fragment length polymorphisms. Theor. Appl. Genet. 98:411–421.

Aquino de Muro, M., S. Mehta, and D. Moore. 2003. The use of amplified fragment length polymorphism for molecular analysis of *Beauveria bassiana* isolates from Kenya and other countries, and their correlation with host and geographical origin. FEMS Microbiol. Lett. 229:249–257.

Archie, J. W. 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38:239–252.

Arroyo-Garcia, R., J. L. Cenis, J. Tello, J. M. Martinez-Zapater, and D. Cifuentes. 2003. Genetic relationships among seven specialized forms of *Fusarium oxysporum* determined by DNA sequencing of the ITS region and AFLPs. Spanish J. Agric. Res. 1:55–63.

Bachmann, K. 1992. Nuclear DNA markers in angiosperm taxonomy. Acta Bot. Neerl. 41:369–384.

Backeljau, T., L. De Bruyn, H. De Wolf, K. Jordaens, S. Van Dongen, R. Verhagen, and B. Winnepenninckx. 1995. Random amplified polymorphic DNA (RAPD) and parsimony methods. Cladistics 11:119–130.

Bakkeren, G., J. W. Kronstad, and C. A. Lévesque. 2000. Comparison of AFLP fingerprints and ITS sequences as phylogenetic markers in Ustilaginomycetes. Mycologia 92:510–521.

Baldwin, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Mol. Phylogenet. Evol. 1:3–16.

Bao, J. R., D. R. Fravel, N. R. O' Neill, G. Lazarovits, and P. Van Berkum. 2002. Genetic analysis of pathogenic and nonpathogenic *Fusarium oxysporum* from tomato plants. Can. J. Bot. 80:271–279.

Barker, F. K., and F. M. Lutzoni. 2002. The utility of the incongruence length difference test. Syst. Biol. 51:625–637.

Beardsley, P. M., A. Yen, and R. G. Olmstead. 2003. AFLP phylogeny of *Mimulus* section *Erythranthe* and the evolution of hummingbird pollination. Evolution 57:1397–1410.

Black, W. C. 1993. PCR with arbitrary primers: Approach with care. Insect Mol. Biol. 2:1–6.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

Brasier, C. M., D. E. L. Cooke, and J. M. Duncan. 1999. Origin of a new *Phytophthora* pathogen through interspecific hybridization. Proc. Natl. Acad. Sci. USA 96:5878–5883.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

Carpenter, J. M. 1992. Random cladistics. Cladistics 8:147–153.

Carpenter, J. M., P. A. Goloboff, and J. S. Farris. 1998. PTP is meaningless, T-PTP is contradictory: A reply to Trueman. Cladistics 14:105–116.

Castella, G., T. O. Larsen, J. Cabanes, H. Schmidt, A. Alboresi, L. Niessen, P. Farber, and R. Geisen. 2002. Molecular characterization of ochratoxin A producing strains of the genus *Penicillium*. Syst. Appl. Microbiol. 25:74–83.

Chowdappa, P., D. Brayford, J. Smith, and J. Flood. 2003a. Identity of *Phytophthora* associated with arecanut and its relationship with rubber and cardamom isolates based on RFLP of PCR-amplified ITS regions of rDNA and AFLP fingerprints. Curr. Sci. India 85:585–587.

Chowdappa, P., D. Brayford, J. Smith, and J. Flood. 2003b. Molecular discrimination of *Phytophthora* isolates on cocoa and their relationship with coconut, black pepper and bell pepper isolates based on rDNA repeat and AFLP fingerprints. Curr. Sci. India 84:1235–1238.

Cregan, P. B. 1992. Simple sequence repeat DNA length polymorphisms. Probe 2:18–22.

Cunningham, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. 46:464–478.

DeBry, R. W. 2001. Improving interpretation of the decay index for DNA sequence data. Syst. Biol. 50:742–752.

DeBry, R. W., and N. A. Slade. 1985. Cladistic analysis of restriction endonuclease cleavage maps within a maximum-likelihood framework. Syst. Zool. 34:21–34.

DeScenzo, R. A., S. R. Engel, G. Gomez, E. L. Jackson, G. P. Munkvold, J. Weller, and N. A. Irelan. 1999. Genetic analysis of *Eutypa* strains from California supports the presence of two pathogenic species. Phytopathology 89:884–893.

Despres, L., L. Gielly, B. Redoutet, and P. Taberlet. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. Mol. Phylogenet. Evol. 27:185–196.

Dolphin, K., R. Belshaw, C. D. L. Orme, and D. L. J. Quicke. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17:401–406.

Donoghue, M. J., R. G. Olmstead, J. F. Smith, and J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on rbcL sequences. Ann. Mo. Bot. Gard. 79:333–345.

Douhan, G. W., and D. M. Rizzo. 2003. Host-parasite relationships among bolete infecting *Hypomyces* species. Mycol. Res. 107:1342–1349.

Dowton, M., and D. Austin. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy—The behavior of the incongruence length difference test in mixed-model analyses. Syst. Biol. 51:19–31.

El-Rabey, H. A., A. Badr, R. Schafer-Pregl, W. Martin, and F. Salamini. 2002. Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. Plant Biol. 4:567–575.

Emshwiller, E., and J. J. Doyle. 1998. Origins of domestication and polyploidy in oca (*Oxalis tuberosa*: Oxalidaceae): nrDNA ITS data. Am. J. Bot. 85:975–985.

Eriksson, T., and N. Wikström. 1996. AutoDecay, version 3.03. Department of Botany, Stockholm University, Stockholm.

Faith, D. P. 1992. On corroboration: A reply to Carpenter. Cladistics 8:265–273.

Faith, D. P., and J. W. O. Ballard. 1994. Length differences and topology-dependent tests: A response to Källersjö et al. Cladistics 10:57–64.

Faith, D. P., and P. S. Cranston. 1991. Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. Cladistics 7:1–28.

Faivovich, J. 2002. On RASA. Cladistics 18:324–333.

Farris, J. S. 2002. RASA attributes highly significant structure to randomized data. Cladistics 18:334–353.

Farris, J. S., V. A. Albert, M. Källersjö, D. Lipscomb, and A. G. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99–124.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. Cladistics 10:315–319.

Felsenstein, J. 1985a. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Felsenstein, J. 1985b. Confidence limits on phylogenies with a molecular clock. Syst. Zool. 34:152–161.

Feráková, V. 1977. The genus *Lactuca* L. in Europe. Univerzita Komenského, Bratislava.

Giannasi, N., R. S. Thorpe, and A. Malhotra. 2001. The use of amplified fragment length polymorphism in determining species trees at fine taxonomic levels: Analysis of a medically important snake, *Trimeresurus albolabris*. Mol. Ecol. 10:419–426.

Giribet, G. 2003. Stability in phylogenetic formulations and its relationship to nodal support. Syst. Biol. 52:554–564.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652–670.

Goloboff, P. A. 1991. Homoplasy and the choice among cladograms. Cladistics 7:215–232.

Grant, T., and A. G. Kluge. 2003. Data exploration in phylogenetic inference: Scientific, heuristic, or neither. Cladistics 19:379–418.

Gräser, Y., A. F. A. Kuijpers, M. El Fari, W. Presber, and G. S. De Hoog. 2000. Molecular and conventional taxonomy of the *Microsporum canis* complex. Med. Mycol. 38:143–153.

Gräser, Y., A. F. A. Kuijpers, W. Presber, and G. S. De Hoog. 1999. Molecular taxonomy of *Trichophyton mentagrophytes* and *T. tonsurans*. Med. Mycol. 37:315–330.

Grodzicker, T., J. Williams, P. Sharp, and J. Sambrook. 1974. Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp. 39:439–446.

Hansen, M., T. Kraft, M. Christiansson, and N.-O. Nilsson. 1999. Evaluation of AFLP in *Beta*. Theor. Appl. Genet. 98:845–852.

Hasegawa, M., and H. Kishino. 1989. Confidence limits on the maximum-likelihood estimate of the hominoid tree from mitochondrial-DNA sequences. Evolution 43:672–677.

Hillis, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. Pages 278–294 *in* Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford University Press, New York.

Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83:189–195.

Hodkinson, T. R., M. W. Chase, C. Takahashi, I. J. Leitch, M. D. Bennett, and S. A. Renvoize. 2002. The use of DNA sequencing (ITS and trnL-F), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). Am. J. Bot. 89:279–286.

Hodkinson, T. R., S. A. Renvoize, G. N. Chonghaile, C. M. A. Stapleton, and M. W. Chase. 2000. A comparison of ITS nuclear rDNA sequence data and AFLP markers for phylogenetic studies in *Phyllostachys* (Bambusoideae, Poaceae). J. Plant Res. 113:259–269.

Holsinger, K. E., and R. K. Jansen. 1993. Phylogenetic analysis of restriction site data. Methods Enzymol. 224:439–455.

Huang, J., H. Corke, and M. Sun. 2002. Highly polymorphic AFLP markers as a complementary tool to ITS sequences in assessing genetic diversity and phylogenetic relationships of sweet potato (*Ipomoea batatas* (L.) Lam.) and its wild relatives. Genet. Resour. Crop Ev. 49:541–550.

Huelsenbeck, J. P. 1991. Tree-length distribution skewness: An indicator of phylogenetic information. Syst. Zool. 40:257–270.

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

Ivors, K. L., K. J. Hayden, P. J. M. Bonants, D. M. Rizzo, and M. Garbelotto. 2004. AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. Mycol. Res. 108:378–392.

Jansen, R. K., K. E. Holsinger, and R. G. Olmstead. 1991. Character weighting for phylogenetic analyses of restriction site data. Am. J. Bot. 78(Suppl.):193–194.

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. Van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breeding 3:381–390.

Källersjö, M., J. S. Farris, A. G. Kluge, and C. Bult. 1992. Skewness and permutation. Cladistics 8:275–287.

Kardolus, J. P., H. J. Van Eck, and R. G. Van den Berg. 1998. The potential of AFLPs in biosystematics: A first application in *Solanum* taxonomy (Solanaceae). Plant Syst. Evol. 210:87–103.

Karp, A., O. Seberg, and M. Buiatti. 1996. Molecular techniques in the assessment of botanical diversity. Ann. Bot.Lond. 78:143–149.

Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P. Jackson. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. 179:818–824.

Kiers, A. M., T. H. M. Mes, R. Van der Meijden, and K. Bachmann. 1999. Morphologically defined *Cichorium* (Asteraceae) species reflect lineages based on chloroplast and nuclear (ITS) DNA data. Syst. Bot. 24:645–659.

Kiers, A. M., T. H. M. Mes, R. Van der Meijden, and K. Bachmann. 2000. A search for diagnostic AFLP markers in *Cichorium* species with emphasis on endive and chicory cultivar groups. Genome 43:470–476.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

Koopman, W. J. M., and G. Gort. 2004. Significance tests and weighted values for AFLP similarities, based on *Arabidopsis* in silico AFLP fragment length distributions. Genetics 167:1915–1928.

Koopman, W. J. M., E. Guetta, C. C. M. Van de Wiel, B. Vosman, and R. G. Van den Berg. 1998. Phylogenetic relationships among *Lactuca* (Asteraceae) species and related genera based on ITS-1 DNA sequences. Am. J. Bot. 85:1517–1530.

Koopman, W. J. M., M. J. Zevenbergen, and R. G. Van den Berg. 2001. Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. Am. J. Bot. 88:1881–1887.

Lu, J., M. R. Knox, M. J. Ambrose, J. K. M. Brown, and T. H. N. Ellis. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. Theor. Appl. Genet. 93:1103–1111.

Lyons-Weiler, J., G. A. Hoelzer, and R. J. Tausch. 1996. Relative apparent synapomorphy (RASA) I: The statistical measurement of phylogenetic signal. Mol. Biol. Evol. 13:749–757.

Mace, E. S., C. G. Gebhardt, and R. N. Lester. 1999a. AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae). Theor. Appl. Genet. 99:634–641.

Mace, E. S., R. N. Lester, and C. G. Gebhardt. 1999b. AFLP analysis of genetic relationships among the cultivated eggplant, *Solanum melongena* L., and wild relatives (Solanaceae). Theor. Appl. Genet. 99:626–633.

Maddison, D. R., and W. P. Maddison. 2000. MacClade 4: Analysis of phylogeny and character evolution. Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

McGregor, C. E., C. A. Lambert, M. M. Greyling, J. H. Louw, and L. Warnich. 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. Euphytica 113:135–144.

Mechanda, S. M., B. R. Baum, D. A. Johnson, and J. T. Arnason. 2004. Sequence assessment of comigrating AFLP bands in *Echinacea*—implications for comparative biological studies. Genome 47:15–25.

Meksem, K., E. Ruben, D. Hyten, K. Triwitayakorn, and D. A. Lightfoot. 2001. Conversion of AFLP bands into high-throughput DNA markers. Mol. Gen. Genomics 265:207–214.

Mickevich, M. F., and J. S. Farris. 1981. The implications of congruence in *Menidia*. Syst. Zool. 30:351–370.

Milbourne, D., R. Meyer, J. E. Bradshaw, E. Baird, N. Bonar, J. Provan, W. Powell, and R. Waugh. 1997. Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Mol. Breeding 3:127–136.

Mirabolfathy, M., D. E. L. Cooke, J. M. Duncan, N. A. Williams, D. Ershad, and A. Alizadeh. 2001. *Phytophthora pistaciae* sp. nov. and *P. melonis*: The principal causes of pistachio gummosis in Iran. Mycol. Res. 105:1166–1175.

Montiel, D., M. J. Dickinson, H. A. Lee, P. S. Dyer, D. J. Jeenes, I. N. Roberts, S. James, L. J. Fuller, K. Matsuchima, and D. B. Archer. 2003. Genetic differentiation of the *Aspergillus* section *Flavi* complex using AFLP fingerprints. Mycol. Res. 107:1427–1434.

Myers, R. H. 1990. Classical and modern regression with applications. PWS-Kent, Boston.

O'Hanlon, P. C., and R. Peakall. 2000. A simple method for the detection of size homoplasy among amplified fragment length polomorphism fragments. Mol. Ecol. 9:815–816.

Pelser, P. B., B. Gravendeel, and R. Van der Meijden. 2003. Phylogeny reconstruction in the gap between too little and too much divergence: the closest relatives of *Senecio jacobaea* (Asteraceae) according to DNA sequences and AFLPs. Mol. Phylogenet. Evol. 29:613–628.

Perrie, L. R., P. J. Brownsey, P. J. Lockhart, E. A. Brown, and M. F. Large. 2003. Biogeography of temperate Australasian *Polystichum* ferns as inferred from chloroplast sequence and AFLP. J. Biogeogr. 30:1729–1736.

Perrie, L. R., P. J. Brownsey, P. J. Lockhart, and M. F. Large.. 2003. Allopolyploidy in New Zealand *Polystichum.* N. Z. J. Bot. 41:189–215.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breeding 2:225–238.

Rehmany, A. P., J. R. Lynn, M. Tor, E. B. Holub, and J. L. Beynon. 2000. A comparison of *Peronospora parasitica* (downy mildew) isolates from *Arabidopsis thaliana* and *Brassica oleracea* using amplified fragment length polymorphism and internal transcribed spacer 1 sequence analyses. Fungal Genet. Biol. 30:95–103.

Rouppe van der Voort, J. N. A. M., P. Van Zandvoort, H. J. Van Eck, R. T. Folkertsma, R. C. B. Hutten, J. Draaistra, F. J. Gommers, E. Jacobsen, J. Helder, and J. Bakker. 1997. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol. Gen. Genet. 255:438–447.

Russell, J. R., J. D. Fuller, M. Macaulay, B. G. Hatz, A. Jahoor, W. Powell, and R. Waugh. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs, and RAPDs. Theor. Appl. Genet. 95:714–722.

Semerikov, V. L., H. Q. Zhang, M. Sun, and M. Lascoux. 2003. Conflicting phylogenies of *Larix* (Pinaceae) based on cytoplasmic and nuclear DNA. Mol. Phylogenet. Evol. 27:173–184.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of loglikelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.

Siegel, S. 1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill Book Company, New York.

Simmons, M. P., C. P. Randle, J. V. Freudenstein, and J. W. Wenzel. 2002. Limitations of relative apparent synapomorphy analysis (RASA) for measuring phylogenetic signal. Mol. Biol. Evol. 19:14–23.

Sokal, R. R., and F. J. Rohlf. 1969. Biometry. The principles and practice of statistics in biological research. W. H. Freeman and Company, San Francisco.

Steel, M. A., P. J. Lockhart, and D. Penny. 1993. Confidence in evolutionary trees from biological sequence data. Nature 364:440–442.

Swofford, D. L. 1999. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic Inference. Pages 407–514 *in* Molecular systematics (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37:221–244.

Tosto, D. S., and H. E. Hopp. 2000. Suitability of AFLP markers for the study of genomic relationships within the *Oxalis tuberosa* alliance. Plant Syst. Evol. 223:201–209.

Tredway, L. P., J. F. J. White, G. B.S., P. V. Reddy, M. D. Richardson, and B. B. Clarke. 1999. Phylogenetic relationships within and between *Epichloe* and *Neotyphodium* endophytes as estimated by AFLP markers and rDNA sequences. Mycol. Res. 103:1593–1603.

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

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:7213–7218.

Werres, S., R. Marwitz, W. A. Man in 't Veld, A. W. A. M. De Cock, P. J. M. Bonants, M. De Weerdt, K. Themann, E. Ilieva, and R. P.

Baayen. 2001. *Phytophthora ramorum* sp. nov., a new pathogen on *Rhododendron* and *Viburnum*. Mycol. Res. 105:1155–1165.

Willems, A., R. Coopman, and M. Gillis. 2001. Comparison of sequence analysis of 16S–23S rDNA spacer regions, AFLP analysis and DNA–DNA hybridizations in *Bradyrhizobium*. Int. J. Syst. Evol. Micr. 51:623–632.

Willems, A., F. Doignon-Bourcier, R. Coopman, B. Hoste, P. De Lajudie, and M. Gillis. 2000. AFLP fingerprint analysis of *Bradyrhizobium* strains isolated from *Faidherbia albida* and *Aeschynomene* species. Syst. Appl. Microbiol. 23:137–147.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531–6535.

Willems, A., A. Munive, P. De Lajudie, and M. Gillis. 2003. In most *Bradyrhizobium* groups sequence comparison of 16S-23S rDNA internal transcribed spacer regions corroborates DNA-DNA hybridizations. Syst. Appl. Microbiol. 26:203–210.

Wyand, R. A., and J. K. M. Brown. 2003. Genetic and forma specialis diversity in *Blumeria graminis* of cereals and its implications for host-pathogen co-evolution. Mol. Plant Pathol. 4:187–198.

Xu, F., and M. Sun. 2001. Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; Amaranthaceae) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent inter-simple sequence repeat markers. Mol. Phylogenet. Evol. 21:372–387.

Yoder, A. D., J. A. Irwin, and B. A. Payseur. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. Syst. Biol. 50:408–424.

Zhang, L.-B., H. P. Comes, and J. W. Kadereit. 2001. Phylogeny and quarternary history of the European montane/alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. Am. J. Bot. 88:2331–2345.

Zietkiewicz, E., A. Rafalski, and D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183.

Cultivated lettuce and its wild relatives. Top row: *L. sativa* (cultivated lettuce), *L. serriola* (prickly lettuce), *L. aculeata*. Bottom row: *L. tatarica*, *L. perennis*, *L. tenerrima*.