

This article was downloaded by:[Althoff, David M.]
[Althoff, David M.]

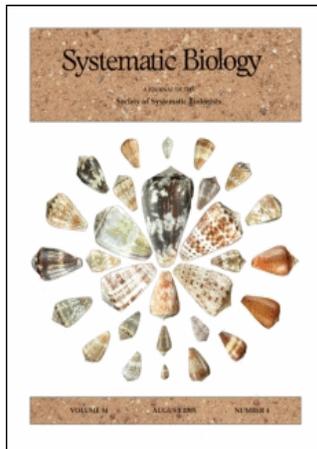
On: 18 June 2007

Access Details: [subscription number 779485701]

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Systematic Biology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713658732>

The Utility of Amplified Fragment Length Polymorphisms in Phylogenetics: A Comparison of Homology within and between Genomes

To cite this Article: Althoff, David M., Gitzendanner, Matthew A. and Segraves, Kari A. , 'The Utility of Amplified Fragment Length Polymorphisms in Phylogenetics: A Comparison of Homology within and between Genomes', *Systematic Biology*, 56:3, 477 - 484

To link to this article: DOI: 10.1080/10635150701427077

URL: <http://dx.doi.org/10.1080/10635150701427077>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

The Utility of Amplified Fragment Length Polymorphisms in Phylogenetics: A Comparison of Homology within and between Genomes

DAVID M. ALTHOFF,¹ MATTHEW A. GITZENDANNER,² AND KARI A. SEGRAVES¹

¹Department of Biology, Syracuse University, 130 College Place, Syracuse, NY 13244, USA; E-mail: dmalthof@syr.edu (D.M.A.)

²Department of Botany, University of Florida, P.O. Box 118526, Gainesville, FL 32611, USA

Abstract.—The amplified fragment length polymorphism (AFLP) technique is being increasingly used in phylogenetic studies, especially in groups of rapidly radiating taxa. One of the key issues in the phylogenetic suitability of this technique is whether the DNA fragments generated via the AFLP method are homologous within and among the taxa being studied. We used a bioinformatics approach to assess homology based on both chromosomal location and sequence similarity of AFLP fragments. The AFLP technique was electronically simulated on genomes from eight organisms that represented a range of genome sizes. The results demonstrated that within a genome, the number of fragments is positively associated with genome size, and the degree of homology decreases with increasing numbers of fragments generated. The average homology of fragments was 89% for small genomes (<400 Mb) but decreased to 59% for large genomes (>2 Gb). Fragment homology for large genomes can be increased by excluding smaller fragments, although there is no clear upper limit for the size of fragments to exclude. A second approach is to increase the number of selective nucleotides in the final selective amplification step. For strains of the same organism, homology based on chromosome location and sequence similarity of fragments was 100%. Fragment homology for more distantly related taxa, however, decreased with greater time since divergence. We conclude that AFLP data are best suited for examining phylogeographic patterns within species and among very recently diverged species. [AFLP; bioinformatics; DNA sequence; genome; homology; phylogenetics; species radiations.]

Since its introduction by Vos et al. (1995), the amplified fragment length polymorphism (AFLP) technique has been used extensively in studies ranging from population genetics to phylogeography to phylogenetics (Mueller and Wolfenbarger, 1999; Bensch and Akesson, 2005). Genomic DNA is first digested using two restriction endonucleases, typically one with a 6-bp recognition sequence (usually *EcoRI*) and one with a 4-bp recognition sequence (usually *MseI*). Adapters of known sequence are then ligated to each end of the fragments and two successive rounds of selective PCR amplification are performed. The first round of PCR (preselective or +1 amplification) uses primers that match the adapters on the *EcoRI* end and *MseI* end of the fragments plus one extra nucleotide. The second round (selective or +3 amplification) has an additional two nucleotides added to the +1 primers sequences. These rounds of selective amplification reduce the resulting pool of DNA fragments to a size more manageable for analysis. Although the DNA fragments are anonymous, the method is remarkably reliable and consistent (Vos et al., 1995). This technique is readily adapted to new taxa because no taxon-specific information is needed and AFLPs are suitable for use in both prokaryotic and eukaryotic organisms. Moreover, the technique surveys the entire genome, is relatively inexpensive, and generates many potential phylogenetic markers.

Although the AFLP technique offers many advantages for generating phylogenetic markers, the phylogenetic utility of the markers is limited by the assumption of size homology—identifying the homology of the fragments is not simple or commonly done (Bussell et al., 2005). Homology, from a phylogenetic perspective, entails information about both the location of a fragment in the genome and the degree of sequence similarity

of fragments of the same size. Because fragments are anonymous and sorted by size via electrophoresis, there is the potential for homoplasy of fragments at several levels. Within an individual, multiple fragments from different regions of the genome may comigrate. These fragments may or may not contain the same sequence. Comigrating fragments that have similar sequences may be orthologous copies of genes, paralogous copies, pseudogenes, or repeated sequences of unknown function. Furthermore, when making comparisons among taxa, fragments that are the same size may not necessarily come from the same locus in the respective genome or have the same sequence. This uncertainty in homology is a particularly important issue in phylogenetic studies—especially because use of the technique is rapidly expanding to examine species radiations that have been difficult to resolve based on other markers such as organellar and nuclear DNA sequences (e.g., Albertson et al., 1999; Giannasi et al., 2001; Després et al., 2003; Brouat et al., 2004; Sullivan et al., 2004; Mendelson and Shaw, 2005). Although studies have demonstrated that AFLPs can resolve phylogenies when other markers have failed, and that AFLP phylogenies corroborate morphology, many phylogeneticists, with good rationale, remain concerned about the homology of markers and their utility as phylogenetic characters.

A number of studies have begun to assess fragment homology via empirical and theoretical approaches (reviewed in Koopman and Gort, 2004; Bussell et al., 2005). Empirical approaches have consisted of comparing fragment profiles from known crosses (Waugh et al., 1997) to extracting fragments from gels and sequencing them to assessing sequence similarity across individuals (Parsons and Shaw, 2001). O’Hanlon and Peakall (2000) outlined a simple method for detecting fragment

sequence similarity by comparing the fragment profile from the final round of selective amplification (usually 3 selective nucleotides) to the profile from another round with additional selective nucleotides (such as 4 selective nucleotides). The results from these approaches have suggested that AFLPs are suitable for phylogenetic studies because sequence similarity has ranged from 95% to 100% in very closely related taxa. Empirical approaches, however, are labor intensive and only a small number of fragments has been examined. Furthermore, studies have shown that co-migration of fragments does occur (see Koopman and Gort, 2004; Mechanda et al., 2004).

Theoretical approaches have surveyed larger sets of markers by modeling the distribution of fragment lengths and generating probabilities of comigration and homoplasy. For example, Vekemans et al. (2002) used analytical theory, based on the work by Innan et al. (1999), and numerical simulations to determine the distribution of fragments generated by AFLPs and the relationship between fragment size and homology. Based on Monte Carlo simulations and empirical results from two plant species, they found that homoplasy may be quite high and was likely to be even higher for shorter fragment sizes. Koopman and Gort (2004) compared theoretical predictions of fragment length distributions to empirical distributions generated from *in silico* AFLP for the *Arabidopsis thaliana* genome. Based on the degree of homoplasy detected from this comparison, they developed significance tests and weighting values as a means to correct for size homoplasy when calculating similarities based on band sharing.

Studies that have examined the homology of fragments have either used sequence similarity as a proxy for homology or have examined the number of bands that may be comigrating for a given size. In most cases, only one or a few organisms have been examined. Consequently, we still do not have a good understanding of the homology of fragments within single individuals, and the issue of homology of AFLP fragments across species remains unresolved. Furthermore, there has not been a comprehensive survey across taxa that differ in genome size to assess the influence of genome size on the degree of fragment homology.

In this study, we take a bioinformatics approach to assess the homology of fragments generated by the AFLP technique. Using a computer program that mimics the AFLP procedure, we examine fragment homology in whole genomes, including organellar, from eight species across the tree of life. We address the following questions: (1) What is the degree of positional and sequence homology of fragments generated from a single individual? (2) How does genome size influence the number of fragments produced and, subsequently, how does genome size influence fragment homology? (3) What is the degree of fragment homology between closely related strains of the same species and between species that have diverged within the last 7 million years? (4) Does the degree of homology change as a function of selective primer sequence?

METHODS

Species Surveyed

We limited the species chosen for analysis according to several criteria: completeness of genome sequencing, assignment of all sequences to chromosomal location, taxonomic breadth, and genome size. Although there are currently many genomes being sequenced, most of these are in the initial stages of assembly and remain incomplete, especially in terms of assigning all sequence data to chromosomal location. We assessed within-organism fragment homology by surveying species along a continuum from smaller to larger genomes. The genomes used were *Bacillus anthracis* strain A2012 from the NCBI Entrez Genome database, GenBank accession number AAAC01000001 (ca. 5.23 Mb); *Saccharomyces cerevisiae* RM11-1a from the NCBI Entrez Genome database via the Broad Institute (ca. 12.07 Mb); *Caenorhabditis elegans* from the NCBI Entrez Genome database (ca. 97 Mb); *Arabidopsis thaliana* from the NCBI Entrez Genome database (119.2 Mb); *Drosophila melanogaster* from the NCBI Entrez Genome database (ca. 180 Mb); *Oryza sativa* variety *japonica* from the International Rice Genome Sequence Project Build 4.0 (ca. 389 Mb); *Mus musculus* Build 36 from the Mouse Genome Sequencing Consortium via the University of California genome browser (ca. 2.5 Gb); *Homo sapiens* Genome Build 36.1 from the International Human Genome Project sequencing centers via the University of California genome browser (ca. 2.9 Gb). The mouse and human genome are not 100% complete in terms of assigning all sequence data to chromosomal locations, but we included these genomes because they were much larger than the other genomes, and we were interested in whether the pattern of homology changed for very large genomes.

We also assessed the homology of fragments between organisms very closely related and more distantly related. We compared the homology of fragments between two strains of yeast, *S. cerevisiae* RM11-1a and *S. cerevisiae* YJM789 (NCBI Entrez Genome database). This comparison would be analogous to a phylogeographic study among populations of the same species. We also compared fragment homology between three fruit fly species *D. melanogaster*, *D. simulans* (*D. simulans* White 501, Genome sequencing center at Washington University School of Medicine), and *D. yakuba* (NCBI Entrez Genome through Genome sequencing center at Washington University School of Medicine Build 2.1). These species have diverged 2 to 7 million of years ago (Russo et al., 1995). We caution that the results may not completely reflect the pattern of homology because the *D. simulans* and *D. yakuba* genomes still contain unsigned sequence data. These three species, however, were the best possible candidates for comparisons among species.

Program Overview

We used a program written in Perl to first survey genomes for sequence fragments containing specific

nucleotides on either end, and then to BLAST those sequences against one another to assess sequence similarity. Specifically, the program conducts two nested searches for strings of nucleotides specified by the researcher. The first search identifies all sequence fragments that contain nucleotide sequences for an *EcoRI* restriction site (GAATTC) on the 5' end and *MseI* restriction site (TTAA) on the 3' end, and vice versa. This search is analogous to the restriction digest in the AFLP technique. In the second search, the subset of fragments is reexamined for an additional three nucleotides adjacent to the 3' end of each enzyme recognition sequence. This search is analogous to the +3 selective amplification step in which the polymerase chain reaction technique is used to selectively amplify fragments that have an additional three nucleotides flanking the restriction site recognition sequences. We searched for the same three selective nucleotides (AAT) adjacent to the *EcoRI* recognition sequence and varied the selective nucleotides adjacent to the *MseI* recognition sequence. We used four sets of three selective nucleotides for the *MseI* recognition sequence—CAA, CAC, CAG, and CAT. By using different *MseI*-associated selective nucleotides, different groups of fragments are detected. For the comparison among the *Drosophila* species we used an additional set of *MseI*-associated selective nucleotides (TTA, TTC, TTG, TTT) to increase the sample size and power of our test of the number of homologous bands shared among the species.

The program sorts the sequence fragments based on length (bp). We only used fragments between 50 and 500 bp, and binned fragments by their length with a 1-bp resolution. This is the standard size range and resolution of AFLP fragments used in population genetic, phylogeographic, and phylogenetic studies. The program creates a set of folders based on the bin sizes found among the fragments and creates text files for each fragment detected. The text files contains a unique identifier and the DNA sequence. For example, there may be a set of folders titled 50, 150, and 250, and within these folders are text files with the DNA sequences of the fragments found. In some cases, there may be multiple fragments within a bin. The program then performs a BLAST search using BLAST 2.2.12 from NCBI to compare the sequence similarity at the 100% and 95% similarity levels for the fragments within each bin. In this way, we assessed the sequence similarity of fragments of the same length.

Analyses

For the within-organism analyses in which species had the genome packaged into chromosomes, we searched for fragments by chromosome. The resulting fragment profiles for each chromosome were then combined to give the organism-wide profile. The organism-wide profile represents the fragment profile that would be visualized via electrophoresis in the laboratory. In cases in which the same fragment size was found multiple times within or among more than one chromosome,

we assessed the sequence similarity of the fragments. The between-organism analyses were also conducted on a chromosome-by-chromosome basis. For the comparison between yeast genomes, only a very small number of fragments were generated using three selective nucleotides. Therefore, we used only two selective nucleotides after each restriction enzyme rather than three in order to increase the number of fragments for the analyses.

Genomes were downloaded from the sources mentioned above and were converted to standard GenBank format. For the three *Drosophila* species, the genome data from NCBI are packaged into chromosome arms. We compared each arm separately across species and then checked the fragment pattern from the entire (both arms) chromosome to determine if fragments were homoplasious or homologous. We also ran an additional set of four +3 selective nucleotide primer pairs to increase the sample size and power of our test of the number of homologous bands shared among the *Drosophila* species. We used least squares regression to examine the relationship between genome size and the overall number of fragments produced and between the number of fragments produced and the number of homoplasious fragments both within and between chromosomes. We also conducted Wilcoxon tests to examine whether there were differences in the number of fragments produced by each of the selective nucleotide combinations. Statistical analyses were conducted using JMP 5.0.1.2 (SAS Institute, Cary, NC).

RESULTS

The organellar genomes for all of the species did not yield any fragments. The number of fragments produced was dependent on genome size (Table 1; Fig. 1a). On average, larger genomes produced more fragments (Fig. 1a, $y = 0.06x + 8.19$; $F_{1,26} = 545.18$, $P < 0.0001$). There was no difference in the average number of fragments produced for each of the +3 selective nucleotides sets (Wilcoxon $\chi^2 = 1.85$, $df = 3$, $P < 0.60$). The number of homoplasious fragments both within and among chromosomes was a function of the number of fragments produced (Fig. 1b, c). Fragment homology decreased (homoplasia increased) with increasing fragment production both within chromosomes ($y = 0.08x - 0.61$; $F_{1,26} = 71.82$, $P < 0.0001$) and among chromosomes ($y = 0.37x - 4.80$; $F_{1,26} = 456.64$, $P < 0.0001$). The nucleotide sequence of the *MseI* +3 selective nucleotides did not influence fragment homology within a chromosome (Wilcoxon $\chi^2 = 0.72$, $df = 3$, $P = 0.87$) or within an entire genome (Wilcoxon $\chi^2 = 0.80$, $df = 3$, $P = 0.85$). Fragments of smaller sizes were more likely to be homoplasious than larger fragments (Fig. 1d) across all of the genomes sampled. This pattern was driven by the large number of fragments generated from the mouse and human genomes (431 out of the 454 homoplasious fragments from all genomes).

The comparison of homology between genomes yielded contrasting results (Table 2). The four *EcoRI*

TABLE 1. Summary of the number of fragments detected by each +3 selective nucleotide pair (AAT was always used on the *EcoRI* end), and the number of fragments homoplasious within chromosomes (i.e., multiple same-sized fragments but different sequences) and homoplasious among chromosomes (i.e., same size fragment from more than one chromosome) for each species.

Species	Genome size (Mb)	No. of chromosomes	<i>MseI</i> +3 nucleotides	No. fragments visualized	No. fragments homoplasious	
					Within chromosomes	Among chromosomes
<i>Bacillus anthracis</i>	5.23	1	caa	0	0	0
			cac	0	0	0
			cag	0	0	0
			cat	1	0	0
			Total	1	0	0
<i>Saccharomyces cerevisiae</i>	12.07	16	caa	2	0	0
			cac	0	0	0
			cag	4	0	0
			cat	3	0	0
			Total	9	0	0
<i>Caenorhabditis elegans</i>	97	6	caa	34	1	1
			cac	14	1	0
			cag	17	0	0
			cat	27	0	0
			Total	92	2	1
<i>Arabidopsis thaliana</i>	119.2	5	caa	34	3	4
			cac	16	0	2
			cag	10	0	1
			cat	19	1	0
			Total	79	4	7
<i>Drosophila melanogaster</i>	180	4	caa	27	0	1
			cac	16	1	2
			cag	7	0	0
			cat	28	2	1
			Total	78	3	4
<i>Oryza sativa japonica</i>	389	12	caa	43	1	3
			cac	20	1	2
			cag	24	2	2
			cat	43	3	5
			Total	130	7	12
<i>Mus musculus</i>	2500	20	caa	166	6	60
			cac	134	23	27
			cag	157	7	49
			cat	182	18	68
			Total	639	54	204
<i>Homo sapiens</i>	2900	23	caa	189	16	71
			cac	134	16	41
			cag	148	6	44
			cat	174	15	71
			Total	645	53	227

+2/*MseI* +2 selective nucleotide combinations yielded a total of 121 fragments for *S. cerevisiae*. Fragment homology both in terms of chromosome position and sequence similarity was 100% between the yeast strains, except for one sequence that shared greater than 95% similarity. In contrast, the homology of fragments for the comparison between *D. melanogaster*, *D. simulans*, and *D. yakuba* was lower. The eight *EcoRI* +3/*MseI* +3 selective nucleotide combinations produced 211 fragments for *D. melanogaster*, 219 fragments for *D. simulans*, and 209 fragments for *D. yakuba*. Of these fragments, 26 of the 44 shared fragments between *D. melanogaster* and *D. simulans* were homologous, 4 of the 17 shared fragments between *D. melanogaster* and *D. yakuba* were homologous, 2 of the 14 shared fragments between *D. simulans* and *D. yakuba* were homologous, and 3 of the 11 fragments shared by all three species were homologous for all three species. Six of the eight remaining fragments

shared by all three species were only homologous between *D. melanogaster* and *D. simulans*.

DISCUSSION

Because of the anonymous nature of the fragments generated by the AFLP technique, there is concern over whether fragments shared among individuals are indeed homologous, especially as divergence time increases. This concern is particularly relevant for phylogenetic studies that rely on homologous characters to determine the evolutionary history of a group of taxa. The AFLP technique has increasingly been used to elucidate the evolutionary relationships among rapidly radiating taxa, resolving relationships that were equivocal based on one or several gene sequences and producing species phylogenies that are consistent with morphological and behavioral data (Albertson et al., 1999; Parsons and Shaw,

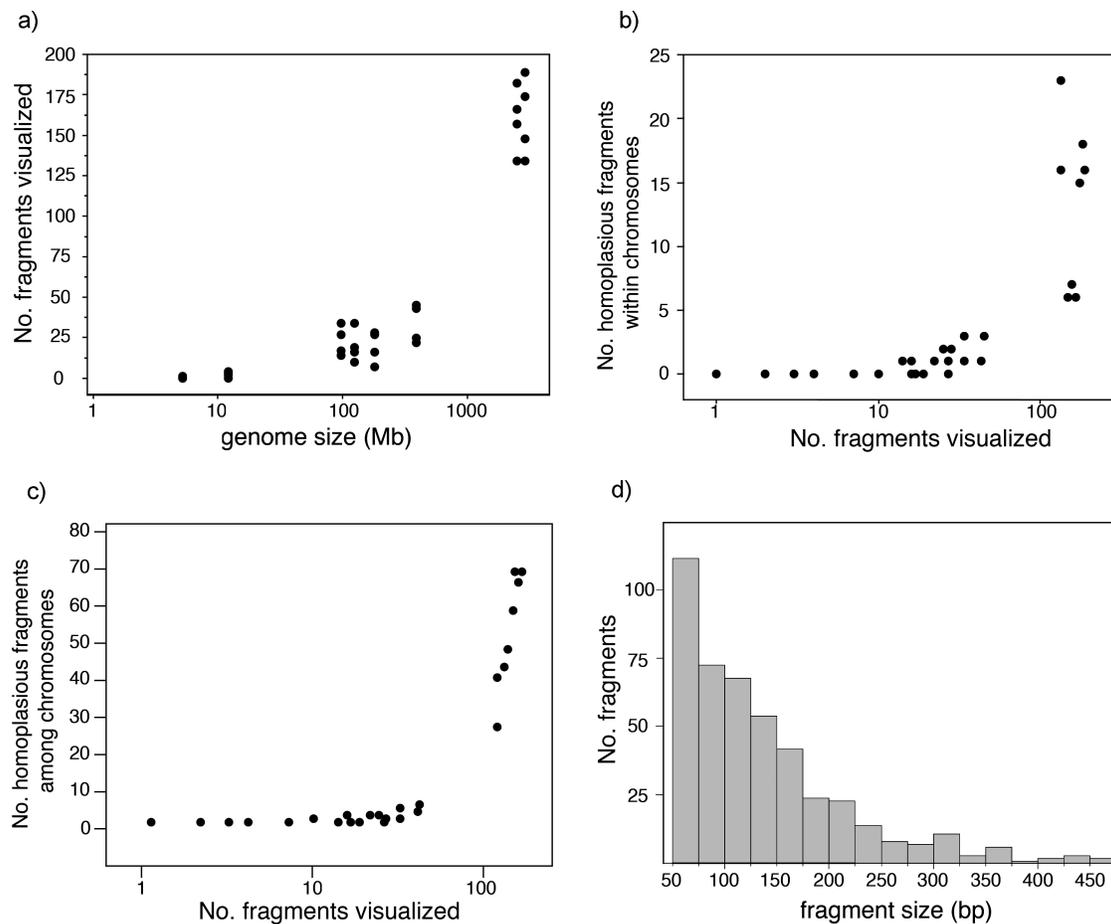


FIGURE 1. Relationship between genome size and homoplasmy of AFLP fragments. Only genomes that produced bands are displayed. (a) The number of fragments produced as a function of genome size. (b) The number of homoplasious fragments (i.e., comigrating) within a chromosome as a function of the number of fragments detected. (c) The number of homoplasious fragments (i.e., comigrating) from across the entire genome as a function of number of fragments detected. (d) Frequency distribution of the fragment sizes that were homoplasious from the eight genomes sampled (431 of 454 fragments were from the mouse and human genomes).

2001; Ogden and Thorpe, 2002; Després et al., 2003; Sullivan et al., 2004; Althoff et al., 2006). Many researchers, including the authors, have operated under the assumption that AFLP fragments are homologous or that only a very small percentage of the fragments are homoplasious. Moreover, another assumption is that the resulting phylogenetic signal from homologous fragments would overcome any weak signal from homoplasious markers. Koopman (2005) surveyed the phylogenetic signal present in AFLP datasets by comparing AFLP phylogenetic analyses to analyses of ITS sequence data in a wide range of taxa (plants, fungi, and bacteria). He found that ample phylogenetic signal is present within AFLP datasets and the phylogenetic results were largely consistent with those derived from sequence data.

Our results from using bioinformatics tools to perform *in silico* AFLP procedures on entire genomes demonstrated that fragment homology within individuals can be quite high, but is dependent on genome size. For the genomes under 400 Mb, on average 11% (range 0% to 18.65%) of the fragments comigrating were homopla-

sious. This result suggests that 89% of the fragments generated by the AFLP technique are homologous. For the mouse and human genomes that are over 2 Gb in size, however, fragment homology was considerably lower. On average, 41.5% (range 33% to 49%) of the fragments were homoplasious. This pattern is driven by the number of fragments produced per genome. The proportion of homologous fragments was negatively correlated with the number of AFLP fragments, and larger genomes produced more fragments. There are two potential ways to reduce homoplasmy in these larger genomes. One is to exclude smaller sized fragments from analysis. Vekemans et al. (2002) demonstrated that smaller sized fragments are more likely to be homoplasious. Our results confirm this observation, although there was no clear cut off point for fragment size and a decrease in homoplasmy (Fig. 1d). For example, eliminating fragments less than 125 bp from the genomes analyzed would help reduce the levels of homoplasmy by approximately one third on average, but there are still markers greater than 125 bp that are homoplasious. A second way to

TABLE 2. (a) Summary of the homology of fragments detected by each *MseI*+2 selective nucleotide set (AA was always used on the *EcoRI* end), for two strains of *Saccharomyces cerevisiae*. Each shared fragment between *S. cerevisiae* strains originated from the same chromosome. (b) Fragment homology for each *MseI*+3 selective nucleotide set (AAT was always used on the *EcoRI* end) for *Drosophila melanogaster*, *D. simulans*, and *D. yakuba*. In contrast with *S. cerevisiae*, many shared fragments originated from different chromosomes and were not homologous.

(a) <i>Saccharomyces cerevisiae</i> strain RM11-1a compared to strain YMJ789							
<i>MseI</i> +2 nucleotides	No. fragments visualized		No. fragments shared	Fragments with sequence similarity of			
	RM11-1a	YMJ789		100%	>95%		
ca	37	37	37	37	—		
cc	20	20	20	20	—		
cg	25	25	25	25	—		
ct	39	39	39	38	1		
Total	121	121	121	120	1		
(b) Homology of fragments detected in <i>Drosophila melanogaster</i> , <i>D. simulans</i> , and <i>D. yakuba</i>							
<i>MseI</i> +3 nucleotides	No. fragments visualized			No. shared fragments between (no. homologous)			
	<i>melanogaster</i>	<i>simulans</i>	<i>yakuba</i>	<i>melanogaster</i> and <i>simulans</i>	<i>melanogaster</i> and <i>yakuba</i>	<i>simulans</i> and <i>yakuba</i>	All species
caa	27	27	25	7 (6)	2 (1)	1 (1)	1 (1 <i>mel.</i> & <i>sim.</i> , 0 for all)
cac	17	14	16	2 (2)	0	0	0
cag	7	7	15	0	0	0	0
cat	28	21	22	3 (2)	3 (2)	0	2 (2 for all)
tta	23	31	20	7 (4)	0	4 (1)	0
ttc	17	16	18	2 (2)	1 (0)	0	1 (1 for all)
ttg	43	42	37	11 (6)	5 (1)	2 (0)	4 (2 <i>mel.</i> & <i>sim.</i> , 0 for all)
ttt	49	61	56	12 (4)	6 (0)	7 (0)	3 (3 <i>mel.</i> & <i>sim.</i> , 0 for all)
Total	211	219	209	44 (26)	17 (4)	14 (2)	11 (6 <i>mel.</i> & <i>sim.</i> , 3 for all)

reduce homoplasmy is to use more stringent primers (such as four selective nucleotides). Indeed, a simulation using a four selective nucleotide primer pair (aatt and catt) on the mouse genome yielded 33 fragments, only one of which was homoplasious. There is a tradeoff in that the number of phylogenetic markers decreases with more selective conditions, but for large genomes this may be the only means of increasing marker homology.

The comparisons of fragment homology between closely related taxa suggested that homology and the phylogenetic usefulness of AFLPs will vary, depending on time since divergence. We used time since divergence as a proxy for whole-genome divergence, under the assumption that increasing time since divergence equates with an increase in whole genome divergence. Comparisons among baker's and brewer's yeast strains demonstrated complete homology of the AFLP fragments generated. If we assume that the evolutionary distance among the yeast strains approximates evolutionary distance among populations within a species, AFLP markers should be quite useful in determining phylogeographic patterns. We must highlight the caveat that the high homology of fragments for our comparison of yeast may be biased by the fact that yeast has a small genome. The comparison between *D. melanogaster*, *D. simulans*, and *D. yakuba*, however, indicates that there is an upper limit to fragment homology among diverged taxa. *Drosophila melanogaster* and *D. yakuba* are estimated to have diverged approximately 6 million years ago, and *D. melanogaster* and *D. simulans* about 2 million years ago (Russo et al., 1995). The *D. melanogaster*–*D. simulans* genomes shared more bands (44) than either shared with *D. yakuba* (17 for *melanogaster*–*yakuba* and 14

for *simulans*–*yakuba*), and the proportion of homologous bands was also higher. Although more than 200 bands were generated for each species, only a small number of these were shared among all of them. This finding has two implications. First, many of the fragments will be unique to a taxon rather than shared among taxa. Second, only a very small proportion of the fragments will determine the relationships among taxa and a portion of these will be homoplasious. The results from this study fit well with the caution given by others that AFLPs should only be used to assess relationships of very closely related taxa (O'Hanlon and Peakall, 2000; Bussell et al., 2005; Koopman, 2005). We also suggest that AFLPs should be used only in instances where DNA sequence data are insufficiently variable to resolve the phylogeny. For larger genomes (greater than 400Mb), fragment homology can be increased by discarding smaller fragments and using more stringent selective primers in the final selective amplification step.

The Future of AFLPs in Phylogenetics

The appeal of AFLPs for phylogenetics is that the technique generates large numbers of markers and surveys the entire genome rather than just a single gene or a small number of genes. The issue of fragment homology, however, is a major concern in using AFLPs in phylogenetic studies. The current study, combined with previous work, suggests that this issue is one that will not be resolved easily without considerable effort to compare homology for organisms with small and large genomes and at all hierarchical levels—from within individuals

to between species within genera. The completion of ongoing genome sequencing projects of closely related species, especially *Drosophila*, will be extremely valuable in testing the homologous nature of AFLP markers as a function of time since divergence among taxa. The current results demonstrate that not all AFLP markers of a given length are homologous, and we suggest that researchers should approach the interpretation of phylogenetic trees with caution, especially without the added benefit of congruent relationships from other data such as morphology, behavior, or other molecular markers.

Another issue confronting the use of AFLPs in phylogenetics is the limited avenues of phylogenetic analysis. Currently, AFLPs are analyzed in predominately two ways—by using parsimony on the presence and absence of fragments or converting the AFLP fragment patterns into a distance measure and using minimum evolution or neighbor joining search algorithms to determine the best phylogenetic tree. Both approaches assume that fragments of the same size are homologous across all the taxa analyzed. Our results clearly demonstrate that this assumption is going to be violated for some taxonomic comparisons. An additional, and much more tenuous assumption that we have not tested here is that the absence of a fragment is homologous as well. Alternatively, converting the fragment patterns into a distance matrix assumes that genetic distance is a good measure of evolutionary divergence and evolutionary divergence translates into evolutionary history. The use of both parsimony and distance approaches have been a source of contention in phylogenetics and have led to the use of alternative analytical approaches. We are somewhat constrained in using likelihood and Bayesian phylogenetic approaches because there are currently very few models of evolution describing the pattern of marker gain and loss, and these models are based on restriction fragment patterns rather than the AFLP technique per se. For example, the F81 model for binary data is available for use in maximum likelihood analyses implemented in TreePuzzle (Schmidt et al., 2002), and the RESTRICTION model that allows different rates of fragment gain and loss, and rate differences across markers is available in Bayesian methods implemented in MrBayes v3.1 (Ronquist and Huelsenbeck, 2003). Models based on restriction site patterns represent a simplified picture of AFLP fragment evolution (Felsenstein, 2004) and to date there has been no empirical evaluation of their performance for AFLPs. Moreover, as the results demonstrated, genome size may be a critical factor in determining the homology of fragments, and their gain and loss. In addition, other issues such as marker nonindependence and clustering of markers based on G+C content need to be evaluated and incorporated into evolutionary models of AFLP markers. The continuing development of such models will advance the utility of AFLP markers in phylogenetic studies, although the current work indicates that homology assessment will continue to be an essential component of any study.

ACKNOWLEDGMENTS

We thank Jack Sullivan for spurring us to think critically about the homology of AFLP fragments, Jason Machacek for generating a preliminary version of the software used in the manuscript, and Celeste Brown for facilitating the interaction between biologists and computer programmers. We also thank the IBEST program at the University of Idaho for computer time and support. Frank (Andy) Anderson and two anonymous reviewers provided very valuable criticism that greatly improved the manuscript. This work was funded by NSF grant DEB 0321293 to David Althoff and Olle Pellmyr and NSF grant DBI 0306164 to Kari Segraves.

REFERENCES

- Albertson, R. C., J. A. Markert, P. D. Danley, and T. D. Kocher. 1999. Phylogeny of a rapid evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proc. Natl. Acad. Sci.* 96:5107–5110.
- Althoff, D. M., K. A. Segraves, J. Leebens-Mack, and O. Pellmyr. 2006. Patterns of speciation in the yucca moths: Parallel radiations within the *Tegeticula yuccasella* species complex. *Syst. Biol.* 55:398–410.
- Bensch, S., and M. Akesson. 2005. Ten years of AFLP in ecology and evolution: Why so few animals? *Mol. Ecol.* 14:2899–2914.
- Brouat, C., D. McKey, and E. J. P. Douzery. 2004. Differentiation in a geographical mosaic of plants coevolving with ants: Phylogeny of the *Leonardoxa africana* complex (Fabaceae:Caesalpinioideae) using amplified fragment length polymorphism markers. *Mol. Ecol.* 13:1157–1171.
- Bussell, J. D., M. Waycott, and J. A. Chappill. 2005. Arbitrarily amplified DNA markers as characters for phylogenetic inference. *Perspect. Plant Ecol. Evol. Syst.* 7:3–26.
- Després, L., L. Gielly, W. 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.
- Giannasi, N., R. 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.
- Innan, H., R. Terauchi, G. Kahl, and F. Tajima. 1999. A method for estimating nucleotide diversity from AFLP data. *Genetics* 151:1157–1164.
- Koopman, W. J. M. 2005. Phylogenetic signal in AFLP data sets. *Syst. Biol.* 54:197–217.
- 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.
- Mechanda, S. M., B. R. Baum, D. A. Johnson, and J. T. Arnason. 2004. Sequence assessment of comigrating AFLP (TM) bands in *Echinacea*—implications for comparative biological studies. *Genome* 47:15–25.
- Mendelson, T. C., and K. L. Shaw. 2005. Rapid speciation in an arthropod. *Nature* 433:375.
- Mueller, U. G., and L. L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* 14:389–394.
- Ogden, R., and R. S. Thorpe. 2002. The usefulness of amplified fragment length polymorphism markers for taxon discrimination across graduated fine evolutionary levels in Caribbean *Anolis* lizards. *Mol. Ecol.* 14:437–445.
- O'Hanlon, P. C., and R. Peakall. 2000. A simple method for the detection of size homoplasy among amplified fragment length polymorphism fragments. *Mol. Ecol.* 9:815–816.
- Parsons, Y. M., and K. L. Shaw. 2001. Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Mol. Ecol.* 10:1765–1772.
- Russo, C. A. M., N. Takezaki, and M. Nei. 1995. Molecular phylogeny and divergence times of *Drosophilid* species. *Mol. Biol. Evol.* 12:391–404.
- Sullivan, J. P., S. Lavoue, M. E. Arnegard, and C. D. Hopkins. 2004. AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* 58:825–841.

- Vekemans, X., T. Beauwens, M. Lemaire, and I. Roldan-Ruiz. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol.* 11:139–151.
- 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.
- Waugh, R., N. Bonar, E. Baird, B. Thomas, A. Graner, P. Hayes, and W. Powell. 1997. Homology of AFLP products in three mapping populations of barley. *Mol. Gen. Genet.* 255:311–321.

*First submitted 27 September 2006; reviews returned 24 November 2006;
final acceptance 8 March 2007
Associate Editor: Frank Anderson*