Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fungal Genetics and Biology 42 (2005) 119–129
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Alt a 1 allergen homologs from *Alternaria* and related taxa: analysis of phylogenetic content and secondary structure

Soon Gyu Hong[a], Robert A. Cramer[b], Christopher B. Lawrence[b,1], Barry M. Pryor[a,*]

[a] *Division of Plant Pathology and Microbiology, Department of Plant Sciences, College of Agriculture, University of Arizona, Tucson, AZ 85721, USA*
[b] *Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523-1177, USA*

Received 31 March 2004; accepted 9 October 2004
Available online 25 November 2004

Abstract

A gene for the *Alternaria* major allergen, Alt a 1, was amplified from 52 species of *Alternaria* and related genera, and sequence information was used for phylogenetic study. Alt a 1 gene sequences evolved 3.8 times faster and contained 3.5 times more parsimony-informative sites than glyceraldehyde-3-phosphate dehydrogenase (*gpd*) sequences. Analyses of Alt a 1 gene and *gpd* exon sequences strongly supported grouping of *Alternaria* spp. and related taxa into several species-groups described in previous studies, especially the infectoria, alternata, porri, brassicicola, and radicina species-groups and the *Embellisia* group. The sonchi species-group was newly suggested in this study. Monophyly of the *Nimbya* group was moderately supported, and monophyly of the *Ulocladium* group was weakly supported. Relationships among species-groups and among closely related species of the same species-group were not fully resolved. However, higher resolution could be obtained using Alt a 1 sequences or a combined dataset than using *gpd* sequences alone. Despite high levels of variation in amino acid sequences, results of in silico prediction of protein secondary structure for Alt a 1 demonstrated a high degree of structural similarity for most of the species suggesting a conservation of function.

Keywords: *Alternaria*; Phylogenetics; Alt a 1 allergen; Secondary structute; *gpd*

1. Introduction

Alternaria alternata is considered one of the most prolific producers of fungal allergens. Many allergens produced are proteins such as enolase, ribosomal proteins, nuclear transport factor, and aldehyde dehydrogenase, which induce allergenic responses in humans (Achatz et al., 1995; Simon-Nobbe et al., 2000; Weichel et al., 2003). However, the biological functions of many other allergenic proteins are unknown (Achatz et al., 1995; De Vouge et al., 1996). For example, the major allergen produced by *A. alternata*, Alt a 1, is a protein with no known function in fungal metabolism or ecology (Barnes et al., 1996; De Vouge et al., 1996). Recently, a homolog of the gene was found to be highly up-regulated during the infection process of *Alternaria brassicicola* on *Arabidopsis thaliana*, suggesting the gene may be involved in plant pathogenicity (Cramer and Lawrence, 2003, 2004). A comparison between Alt a 1 homologs of *A. alternata* and *A. brassicicola* revealed greater sequence divergence than that found in similar comparisons of other ribosomal and protein-coding genes such as the internal transcribed spacer (ITS) and mitochondrial small subunit (mt SSU) rDNA, and glyceraldehyde-3-phosphate dehydrogenase (*gpd*) for the same two species (92% similarity vs. 96, 99, and 97%, respectively) (Cramer and Lawrence, 2003; Pryor and Bigelow, 2003). This level of sequence variation is worth noting, given the role the corresponding protein may play in human

* Corresponding author. Fax: +1 520 621 9290.
E-mail address: bmpryor@ag.arizona.edu (B.M. Pryor).
[1] Present address: Virginia Bioinformatics Institute, 1180 Pratt Drive, Research Building XV, Blacksburg, VA 24061, USA.

1087-1845/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2004.10.009

allergenic responses as well as potential roles in fungal pathogenicity. If Alt a 1 homologs are present in other related species, the gene may be useful in reconstructing phylogenetic relationships among *Alternaria* and related taxa. Furthermore, an analysis of nucleotide and amino acid conservation of Alt a 1 homologs might be valuable in understanding the evolution of this gene and its potential role in fungal biology.

Phylogenetic relationships among *Alternaria* and related species of *Ulocladium*, *Embellisia*, and *Nimbya* have been established based on analysis of ITS, mt SSU, and *gpd* sequences (Chou and Wu, 2002; de Hoog and Horre, 2002; Pryor and Bigelow, 2003; Pryor and Gilbertson, 2000). In these studies, it was revealed that the four genera composed a large monophyletic *Alternaria-Nimbya-Embellisia-Ulocladium* clade with *Stemphylium* as the sister taxon. It was also revealed that *Alternaria*, *Embellisia*, and *Ulocladium* are polyphyletic, and eight species-groups were proposed based on monophyletic relationships. Most of the species-groups were supported by high bootstrap values, especially the alternata, infectoria, porri, and radicina species-groups. However, relationships among species-group and among closely related species were not clearly resolved. To develop a more robust phylogeny of *Alternaria* and related taxa, additional analyses incorporating more variable genetic loci are required.

The objectives of this study were to determine if Alt a 1 gene homologs are present in other *Alternaria* spp. and related taxa, and to assess their potential phylogenetic content. In addition, the secondary structure of Alt a 1 homologs was predicted in silico to assess structural conservation, which may provide additional clues for assessing biological function and stimulation of allergenic response.

2. Materials and methods

2.1. Fungal strains, DNA extraction, PCR amplification, sequencing, and alignment

Thirty-seven species of *Alternaria*, four species of *Embellisia*, two species of *Nimbya*, five species of *Ulocladium*, three species of *Stemphylium*, and *Pleospora herbarum* (a teleomorph of *Stemphylium*) were used in this study (Table 1). All isolates were cultured on 0.05× potato dextrose agar (PDA, Pryor and Michailides, 2002) and diagnostic morphology was confirmed.

DNA extraction and purifications were conducted according to the previously described protocols (Pryor and Gilbertson, 2000). Amplification of *gpd* genes was conducted according to previously described protocols (Berbee et al., 1999). Amplification of Alt a 1 genes was conducted using primers Alt-for (5′-ATGCAGTTCAC CACCATCGC-3′) and Alt-rev (5′-ACGAGGGTGAY

Table 1
Isolates used in this study, their sources, and GenBank accession numbers for sequences used in phylogenetic analyses

Species name	Source	GenBank accession	
		Alt-a1	*gpd*
A. alternata	EGS 34-016	**AY563301**	AY278808
A. arborescens	EGS 39-128	**AY563303**	AY278810
A. argyranthemi	EGS 43-033	**AY563280**	**AY562400**
A. blumeae	EGS 40-149	**AY563291**	**AY562405**
A. brassicae	BMP 0322	**AY563309**	**AY562414**
A. brassicicola	EEB 2232	**AY563311**	AY278813
A. capsici	EGS 45-075	**AY563298**	**AY562408**
A. carontiincultae	EGS 26-010	**AY563287**	AY278798
A. cetera	EGS 41-072	**AY563278**	**AY562398**
A. cheiranthi	EGS 41-188	**AY563290**	AY278802
A. cinerariae	EGS 33-169	**AY563308**	**AY562413**
A. conjuncta	EGS 37-139	**AY563281**	**AY562401**
A. crassa	DDG Acr1	**AY563293**	AY278804
A. cucumerina	BMP 0188	**AY563300**	**AY562409**
A. dauci	ATCC 36613	**AY563292**	AY278803
A. dumosa	EGS 45-007	**AY563305**	AY562410
A. eryngii	EGS 41-005	**AY563313**	**AY562416**
A. ethzedia	EGS 37-143	**AY563284**	AY278795
A. euphorbiicola	EGS 42-049	**AY563314**	**AY562417**
A. japonica	ATCC 13618	**AY563312**	AY278814
A. limoniasperae	EGS 45-100	**AY563306**	**AY562411**
A. longipes	EGS 30-033	**AY563304**	AY278811
A. macrospora	DDG Am1	**AY563294**	AY278805
A. metachromatica	EGS 38-132	**AY563285**	**AY562404**
A. mimicula	EGS 01-056	**AY563310**	**AY562415**
A. mouchaccae	EGS 31-061	**AY563279**	**AY562399**
A. oregonensis	EGS 29-194	**AY563283**	**AY562403**
A. petroselini	EGS 09-159	**AY563288**	AY278799
A. photistica	EGS 35-172	**AY563282**	**AY562402**
A. porri	ATCC 58175	**AY563296**	AY278806
A. pseudorostrata	EGS 42-060	**AY563295**	**AY562406**
A. radicina	ATCC 96831	**AY563286**	AY278797
A. solani	ATCC 58177	**AY563299**	AY278807
A. sonchi	EGS 46-051	**AY563307**	**AY562412**
A. smyrniii	EGS 37-093	**AY563289**	AY278801
A. tagetica	EGS 44-044	**AY563297**	**Y562407**
A. tenuissima	EGS 34-015	**AY563302**	AY278809
E. allii	EGS 38-073	**AY563322**	AY278827
E. indefessa	EGS 30-195	**AY563323**	AY278828
E. novae-zelandiae	EGS 39-099	**AY563324**	AY278831
E. telluster	EGS 33-026	**AY563325**	AY562419
N. caricis	EGS 13-094	**AY563321**	AY278826
N. scirpicola	EGS 19-016	**AY563320**	AY278825
P. herbarum	ATCC 11681	**AY563277**	AY278823
S. botryosum	ATCC 42170	**AY563274**	AY278820
S. callistephi	EEB 1055	**AY563276**	AY278822
S. vesicarium	ATCC 18521	**AY563275**	AY278821
U. alternariae	BMP 0352	**AY563316**	AY278815
U. atrum	ATCC 18040	**AY563318**	AY278818
U. botrytis	ATCC 18043	**AY563317**	AY278817
U. chartarum	ATCC 18044	**AY563319**	AY278819
U. cucurbitae	EGS 31-021	**AY563315**	**AY562418**

Abbreviations for source: ATCC, American Type Culture Collection, Manassas, VA 20108; BMP, B. M. Pryor, Division of Plant Pathology and Microbiology, Department of Plant Sciences, University of Arizona, Tucson, AZ 85721; DGG, D. G. Gilchrist, Department of Plant Pathology, University of California, Davis, CA 95616; EEB, E. E. Butler, Department of Plant Pathology, University of California, Davis, CA 95616; EGS, E. G. Simmons, Mycological Services, Crawfordsville, IN 47933.

Fig. 1. Alt a 1 gene structure of *Alternaria alternata* (534 bp) and position of primers used in PCR amplification. Alt-for and Alt-rev primers were located at 1–20 and 493–512 nt positions, respectively.

GTAGGCGTC-3′), which were designed based on the conserved regions of *A. alternata* (De Vouge et al., 1996) and *A. brassicicola* (Cramer and Lawrence, 2003) Alt a 1 homologs (Fig. 1). Modified primers that were missing 1 or 2 bases at the 3′ end of each primer were used when PCR products were not obtained with Alt-for and Alt-rev primers. Reaction mixtures contained 2.5 μl of 10× ThermoPol reaction buffer [200 mM Tris–HCl, pH 8.3, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, and 1% Triton X-100], 2.5 μl of 10 mM $MgSO_4$, 10 ng template genomic DNA, 2 pmol of each primer, 2 μl of 2.5 mM dNTPs, 0.5 U of AmpliTaq polymerase (Applied Biosystems, Foster City, CA), and 0.0025 U of Vent DNA polymerase (New England Biolabs, Beverly, MA), and total volume was adjusted to 25 μl with deionized water. PCR amplification was conducted using the following conditions: an initial denaturation at 94 °C for 1 min, 35 amplification cycles of 94 °C for 30 s, 57 °C for 30 s and 72 °C for 1 min, and a final extension phase at 72 °C for 10 min. Sequences of the PCR products were determined with FS DyeTerminator reactions (Applied Biosystems, Foster City, CA) and an ABI automated DNA sequencer. PCR products were sequenced in both the forward and reverse directions for nucleotide confirmation. The resulting nucleotide sequences were proofread, edited, and aligned using the PHYDIT program version 3.1 (Chun, 1995; http://plaza.snu.ac.kr/~jchun/phydit/). The alignment was adjusted manually considering both nucleotide and deduced amino acid positions, and the presence of introns. For some species, *gpd* sequences that were included in the alignments were determined in previous studies (Pryor and Bigelow, 2003; Table 1).

2.2. Phylogenetic analysis of nucleotide sequences

The most parsimonious trees based upon DNA and amino acids sequences were obtained by heuristic search using TBR branch swapping on starting trees generated by random sequence addition (100 reps) using PAUP 4.0 10b (Swofford, 2002). Search settings were as follows: steepest descent option was not in effect, zero length branches were set to collapse to yield polytomies, and MULTREES option was in effect. Bootstrap values were evaluated by 1000 replications using a heuristic search with simple sequence addition, TBR branch swapping, and MULTREES on. Maximum likelihood (ML) trees of DNA sequences were obtained by heuristic search using the TrN + I + G model, which was deduced as the best fit for the data by the likelihood ratio test using MODELTEST ver3.06 (Posada and Crandall, 1998). Distance-based trees were reconstructed by the neighbor-joining method (Saitou and Nei, 1987) from data matrixes calculated using Kimura's 2-parameter model (Kimura, 1980). Sequence of *Stemphylium* spp. and *P. herbarum* were used as outgroups based on results from previous studies (Pryor and Bigelow, 2003; Pryor and Gilbertson, 2000). Data congruence between genes and domains were evaluated by the partition homogeneity test [incongruence length difference (ILD) test (Farris et al., 1995)] option of PAUP 4.0 10b (Swofford, 2002).

2.3. Tempo and mode of gene evolution

Uncorrected distance values were calculated with the pairwise distance values option of PAUP 4.0 10b (Swofford, 2002) to compare evolutionary rates among genes and domains. To assess the variation in levels of base substitution among sites, the number of steps per 5 consecutive bases was estimated for the most-parsimonious trees reconstructed from a combined dataset of Alt a 1 and *gpd* gene sequences using MacClade 4.0 (Maddison and Maddison, 2000). For these analyses, species lacking the alt a 1 intron or *gpd* intron 2 were excluded.

Amino acid identity among isolates was calculated. Of particular interest was the portion of Alt a 1 that corresponds to the antibody-binding epitope of *A. alternata* (Zhang et al., 1995). Probabilities for secondary structural units of Alt a 1 homologs were computed using the PSA Protein Structure Prediction Server (Stultz et al., 1993; http://bmerc-www.bu.edu/psa/index.html). To assess conservation of secondary structure, probabilities for helixes, strands, and turns were compared among species.

3. Results

3.1. PCR amplification, sequencing, and alignment

PCR amplification of the Alt a 1 gene from most of the isolates resulted in products that were approximately 510 bp in length. The PCR products from *P. herbarum*, *Stemphylium callistephi* and *Stemphylium vesicarium* were smaller than those of other species by approximately 60 bp (data not shown). By comparing the Alt a 1 cDNA sequence (Cramer and Lawrence, 2003) and PCR-amplified genomic DNA sequence of *A. brassicicola* (ATCC 34622), it was revealed that the Alt a 1 homolog of *A. brassicicola* has an intron at nucleotide position 334 from the start codon (data not shown). Comparisons between amplified DNA sequences and cDNA sequences of *A. brassicicola* (Cramer and Lawrence, 2003) and *A. alternata* (De Vouge et al., 1996)

revealed that all other taxa examined contained an intron at the same position except for *P. herbarum*, *S. callistephi* and *S. vesicarium*, which lacked the intron entirely. Additional confirmation for the presence of this intron in other species comes from the conserved sequence (GT...AG) of eukaryotic spliceosomal introns (Breathnach et al., 1978).

PCR amplification reactions of the *gpd* gene from most of the species produced PCR products approximately 620 bp in length. PCR products from *Alternaria cetera*, *Alternaria conjuncta*, *Alternaria ethzedia*, *Alternaria metachromata*, *Alternaria oregonensis*, and *Alternaria photistica* were smaller by about 60 bp (data not shown). By comparing sequences, it was revealed that these species were missing intron sequences defined as intron 2 in *Cochliobolus lunatus* (Osiewacz and Ridder, 1991) and *Cochliobolus heterostrophus* (Van Wert and Yoder, 1992). All sequences have been submitted to GenBank and corresponding accession numbers are provided in Table 1.

Alignment of Alt a 1 gene sequences resulted in a 486-character dataset composed of 331 exon 1 sites, 67 intron sites, and 88 exon 2 sites. Alignment of *gpd* sequences resulted in a 591-character dataset composed of 22 exon 2 sites, 60 intron 2 sites, 64 exon 3 sites, 122 intron 3 sites, and 323 exon 4 sites. Because of high variation in the intron sequences, it was difficult to achieve unambiguous alignment among distantly related taxa. In addition, *S. vesicarium*, *S. callistephi*, and *P. herbarum* were lacking the Alt a 1 intron and *A. cetera*, *A. conjuncta*, *A. ethzedia*, *A. metachromatica*, *A. oregonensis*, and *A. photistica* were lacking *gpd* intron 2. For these reasons, intron sequences were excluded in subsequent phylogenetic analyses. All alignments have been submitted to TreeBASE for review (SN2035).

3.2. Phylogenetic analysis

Alignment of Alt a 1 exon sequences resulted in a 419-character dataset of which 247 characters (58.9%) were variable and 190 characters (45.3%) were parsimony-informative. Parsimony analysis with the heuristic search option yielded 49 equally parsimonious trees (steps = 780, CI = 0.471, RI = 0.720), which differed primarily in the positions of the radicina and brassicicola species-groups relative to the other established groups, and in the monophyly of two clades in the *Ulocladium* group (Fig. 2). Most species-groups described in previous studies (Pryor and Bigelow, 2003) were revealed in these analyses as well. The infectoria, alternata, and radicina species-groups and the *Embellisia* group were well-defined with high bootstrap values (>89%). The porri species-group was defined with moderate bootstrap support (69%). *Alternaria sonchi* and *Alternaria cinerariae*

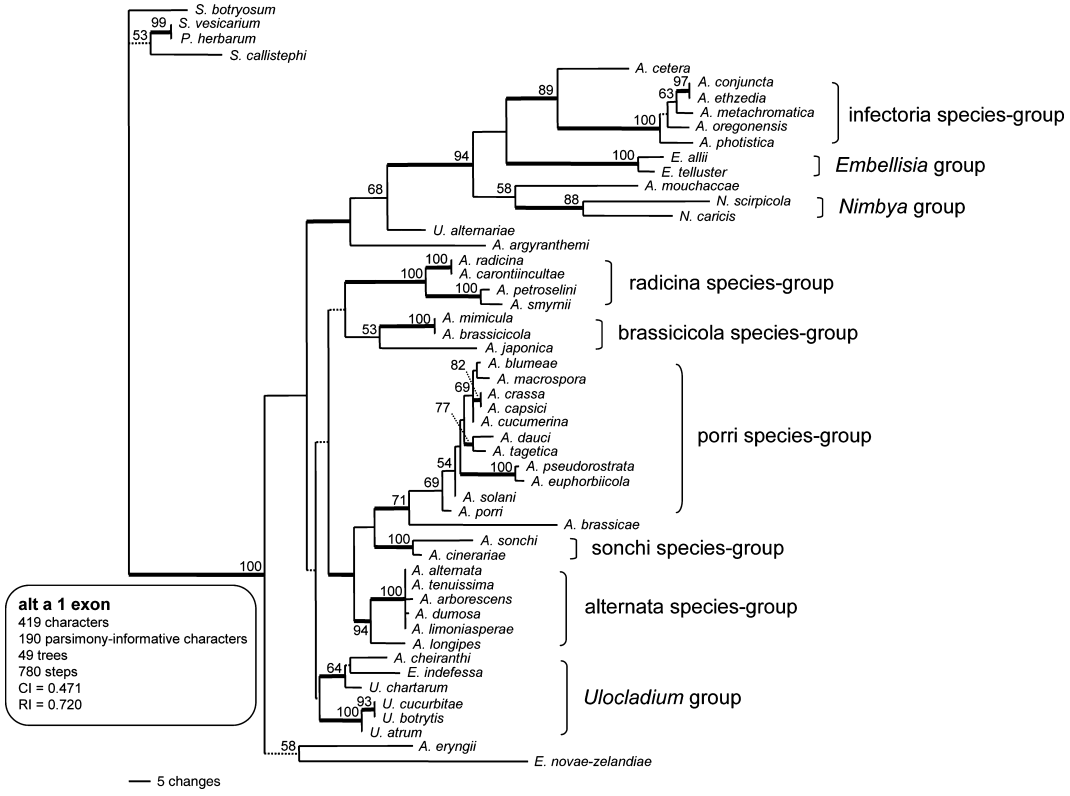

Fig. 2. One of 49 equally parsimonious trees based upon Alt a 1 exon sequences. Broken lines represent branches that were not conserved in the strict consensus tree of 49 equally parsimonious trees and thick lines represent branches conserved in analyses using distance, parsimony, and ML methods. Numbers represent parsimony bootstrap values from 1000 replicates. The scale bar indicates the number of nucleotide substitutions.

composed a well-supported monophyletic group (100%) and were circumscribed as the sonchi species-group. *Nimbya caricis*, *Nimbya scirpicola*, and *Embellisia allii*, which formed a monophyletic group and were circumscribed as the *Nimbya* group in a previous study (Pryor and Bigelow, 2003) did not form a monophyletic group, as *E. allii* clustered with *Embellisia telluster* in a strongly supported *Embellisia* clade (100%). *Embellisia novae-zelandiae* did not group within this clade but was poorly supported (58%) in a clade with *Alternaria eryngii*. Circumscription of *A. brassicicola*, *Alternaria mimicula*, and *Alternaria japonica* as the brassicicola species-group was not supported by a high bootstrap value (53%). Monophyly of the *Ulocladium* group comprising four *Ulocladium* species, *Embellisia indefessa*, and *Alternaria cheiranthi* was weakly supported (<50%) and was composed of two distinct sister clades: *Ulocladium atrum*, *Ulocladium botrytis*, and *Ulocladium cucurbitae* in one clade and *Ulocladium chartarum*, *E. indefessa*, and *A. cheiranthi* in the sister clade.

ML analysis of the Alt a 1 data revealed a tree with similar topology as that revealed in Fig. 2 with notable exception in the placement of *A. japonica*, which did not cluster in the brassicicola species-group but instead was placed as a sister taxon to the group comprising the infectoria species-group, *Embellisia* group, *Nimbya* group, *Ulocladium alternariae*, and *Alternaria argyranthemi* (data not shown). NJ analysis of the Alt a 1 data also revealed a tree with similar topology as that revealed in Fig. 2 with notable exception in the placement of *Alternaria mouchaccae*, which clustered with *E. allii* and *E. telluster*, and the position of the sonchi species-group, which was placed as sister group to the alternata species-group rather than the porri species-group (data not shown). All branches with more than 70% bootstrap support were consistent in all three phylogenetic analyses. These and other branches strictly conserved are shown as bold branches in Fig 2.

Several long-branched terminal taxa were evident in the Alt a 1 gene phylogeny including *A. cetera*, *N. scirpicola*, *N. caricis*, *A. mouchaccae*, *U. alternariae*, *A. argyranthemi*, *A. japonica*, *A. brassicae*, *A. eryngii*, and *E. novae-zelandiae*. Among them, *A. cetera* was related to the infectoria species-group with high bootstrap support (89%) and was maintained in all of the three tree reconstruction methods. Although phylogenetic relationships among *A. mouchaccae*, the group composed of *E. allii* and *E. telluster*, and the group composed of *N. caricis* and *N. scirpicola* were not clear, a monophyletic group comprising all of these species, *A. cetera*, and the infectoria species-group was supported by a high bootstrap value (94%) and maintained in all three of the different tree reconstruction methods. A close relationship between *A. brassicae* and the porri species-group was supported by moderate bootstrap value (71%). However, phylogenetic relationships of *U. alternariae*, *A. argyranthemi*, *A. japonica*, *A. eryngii*, and *E. novae-zelandiae* were not supported by high bootstrap values or maintained in the three tree reconstruction methods.

Parsimony analysis of the amino acid sequence dataset (139 total, 92 variable, and 72 parsimony-informative characters) produced 1993 equally parsimonious trees and the strict consensus of these trees revealed numerous polytomies (data not shown). Although species-groups revealed in amino acid sequence analysis were consistent with those revealed in DNA sequence analysis, resolution of relationships among groups and within groups was much lower. When step numbers for each codon position were examined on the 49 equally parsimonius trees based on DNA sequences, it was shown that the third codon position (427–431 steps depending on tree topologies) experienced 2.0 times and 3.0 times more substitutions than the first (209–212 steps) and second codon positions (142–145 steps), respectively.

Alignment of *gpd* exon sequences resulted in a 409-character dataset of which 81 characters (19.8%) were variable and 54 characters (13.2%) were parsimony-informative. Parsimony analysis with the heuristic search option yielded two equally parsimonius trees (steps = 172, CI = 0.552, RI = 0.789) with near identical topology (Fig. 3). Among groups suggested in a previous study (Pryor and Bigelow, 2003) and the Alt a 1 gene phylogeny (Fig. 2), the infectoria, alternata, and porri species-groups and the *Embellisia* group were maintained in the *gpd* gene phylogeny and supported by moderate to high bootstrap values (>76%). Among previously defined groups, sonchi species-group and the *Ulocladium* group were maintained, but were not supported by a high bootstrap value (67% and <50%, respectively). The brassicicola and radicina species-groups, and the *Nimbya* group were not resolved as monophyletic clades. Phylogenetic relationships of the species that were recognized as long-branched terminal taxa in the Alt a 1 tree were quite different in the *gpd* gene phylogeny. *Alternaria brassicae* was related to the *porri* species-group with moderate bootstrap support (71%) in the Alt a 1 gene phylogeny, but was related to the *N. caricis* in the *gpd* gene phylogeny. *A. japonica* formed a monophyletic group with *A. brassicicola* and *A. mimicula* in the Alt a 1 gene tree, but did not cluster with these species in the *gpd* tree. *A. mouchaccae*, which was related to *N. caricis* and *N. scirpicola* in the Alt a 1 gene tree, was related to *E. allii* and *E. telluster* in the *gpd* gene tree. The phylogenetic positions of *A. argyranthemi*, *A. eryngii*, *U. alternariae*, and *E. novae-zelandiae* changed in the *gpd* phylogeny, but their positions were poorly supported, which was also revealed in the Alt a 1 exon phylogeny.

ML analysis of the *gpd* data revealed a tree with very similar topology as that revealed in Fig. 3 with notable exception in the placement of *A. brassicae*, which was resolved as an independent branch within the

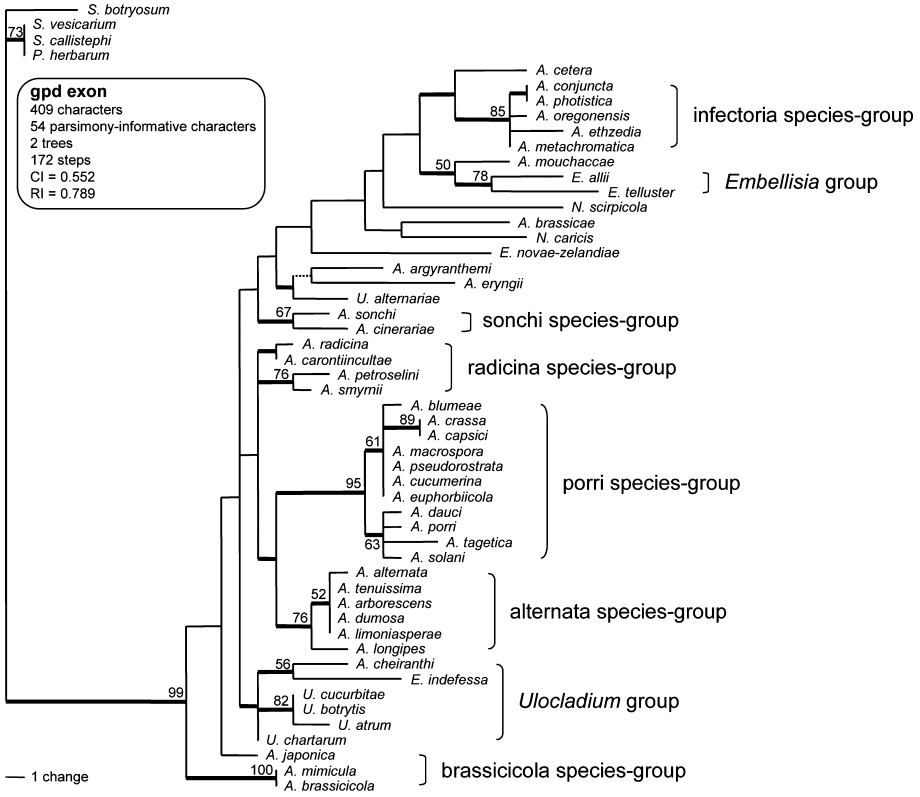

Fig. 3. One of two equally parsimonious trees based upon *gpd* exon sequences. Broken lines represent branches that were not conserved in the strict consensus tree of two equally parsimonius trees and thick lines represent branches conserved in analyses using distance, parsimony, and ML methods. Numbers represent parsimony bootstrap values from 1000 replicates. The scale bar indicates the number of nucleotide substitutions.

infectoria-*Nimbya*-*Embellisia* clade (data not shown). NJ analysis of the *gpd* data also revealed a tree with similar topology as that revealed in Fig. 3 regarding species-group composition but differed somewhat in the relationships among species-groups. Most notably, the sonchi species-group was placed as a sister taxon to the brassicicola species-group, both groups were placed as a sister taxon to the radicina species-group (data not shown), and the infectoria, *Nimbya*, and *Embellisia* groups were placed basal to all other groups. Branches strictly conserved in all three phylogenetic analyses are shown as bold branches in Fig. 3.

Amino acid sequences of the *gpd* gene were very well conserved and there were only five parsimony-informative characters. When step numbers for each codon position were examined on the two equally parsimonius trees based on DNA sequences, it was shown that the third codon position (165 steps) experienced 82.5 times and 33 times more substitutions than first (2 steps) and second codon positions (5 steps), respectively.

Parsimony analysis based upon combined data set of alt a 1 and *gpd* exon sequences (ILD test, $P = 0.09$) resulted in six most parsimonious trees (Fig. 4), which differ primarily in the position of *A. mouchaccae* and relationships among members of the infectoria species-group. Monophyletic grouping of the alternata, infectoria, porri, radicina, and sonchi species-groups and the *Embellisia* and *Nimbya* groups were supported by high bootstrap values (>89%). Bootstrap supports for brassicicola species-group and the *Ulocladium* group were low (62% and <50%, respectively). The *Ulocladium* group was divided into two monophyletic groups, one of which was composed of *A. cheiranthi*, *E. indefessa*, and *U. chartarum* and the other was composed of *U. cucurbitae*, *U. botrytis*, and *U. atrum*.

ML analysis of the combined data set of alt a 1 and *gpd* exon sequences revealed a tree with similar topology as that revealed in Fig. 4 with notable exception in the position of *Ulocladium* group, which was placed as a sister taxon to the group comprising the infectoria species-group, *Embellisia* group, *Nimbya* group, *U. alternariae*, and *A. argyranthemi* (data not shown). NJ analysis of the combined data set of alt a 1 and gpd exon sequences also revealed a tree with similar topology as that revealed in Fig. 4 with notable exception in the position of brassicicola species-group, which did not cluster with radicina species-group but instead was placed as a sister taxon to the group comprising alternata, porri, radicina, and sonchi species-groups, and the *Ulocladium* group (data not shown). Branches strictly conserved in all three phylogenetic analyses are shown as bold branches in Fig. 4.

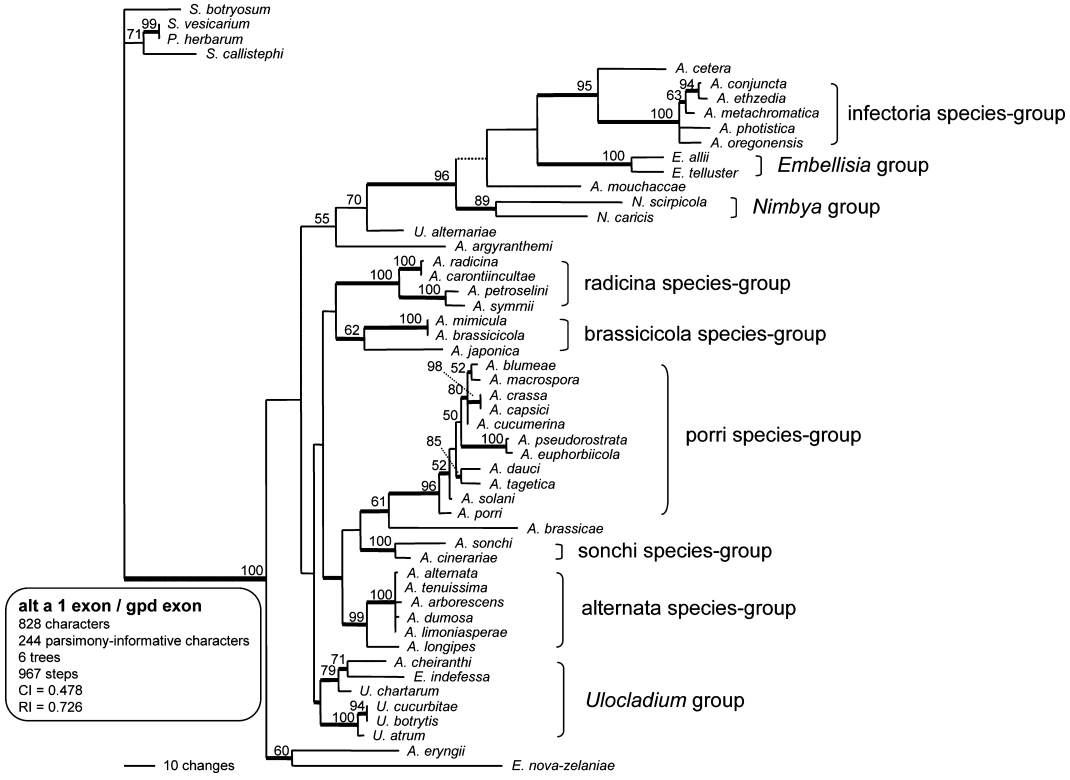

Fig. 4. One of six equally parsimonious trees of combined data set based upon Alt a 1 exon and *gpd* exon sequences. Broken lines represent branches that were not conserved in a strict consensus tree of six equally parsimonious trees and thick lines represent branches conserved in in analyses using distance, parsimony, and ML methods. Numbers represent parsimony bootstrap values from 1000 replicates. The scale bar indicates the number of nucleotide substitutions.

3.3. Tempo and mode of sequence evolution

To evaluate the evolutionary rates of the Alt a 1 and *gpd* genes, pairwise distance values of exon and intron regions for each gene were calculated and compared to each other (Fig. 5). By comparing distance values from exon regions, it was revealed that the evolutionary rate of the Alt a 1 gene was 3.8 times faster than the *gpd* gene (Fig. 5A). When distance values of the intron regions were compared to the exon regions, it was apparent that intron sequences evolved much faster. However, the difference in evolutionary rate between intron and exon regions was not the same for the Alt a 1 gene the *gpd* gene. The intron of the Alt a 1 gene evolved 2.7 times faster than the exon regions (Fig. 5B) and introns of the *gpd* gene evolved 6.1 times faster than the exon regions (Fig. 5C). When the intron of the Alt a 1 gene and introns of the *gpd* gene were compared, it was revealed that the Alt a 1 intron evolved 1.7 times faster than *gpd* introns (Fig. 5D). To examine the distribution of variable sites of each gene, parsimony step numbers over five base windows were depicted (Fig. 6). From the diagram, it is evident that sequence variation is distributed evenly throughout both genes except in intermittent conserved sites. The high evolutionary rate of intron domains is apparent from the diagram.

Amino acid identity between species ranged from 97 to 100% for the *gpd* gene. Amino acid identity between species ranged from 59 to 100% for the Alt a 1. For the antibody-binding epitope region of the Alt a 1 (Zhang et al., 1995) spanning aligned amino acid sites 22–41, amino acid identity between species ranged from 70 to 100%. Probabilities of protein secondary structural motifs were calculated for each aligned site from all of the sequences to examine if there is structural variation as a result of the primary sequence changes. By calculating probabilities for α-helix, β-strand, and turn structural motifs in silico, it was predicted that the Alt a 1 is composed of one α-helix at the N-terminal and from seven to nine β-strand motifs alternating with turn motifs. The variable number of β-strand motifs was due to variable structural probabilities among species between residues 21 and 29, which encompasses a portion of the antibody-binding epitope, and residues 44 and 58 (Fig. 7). Members of the infectoria and radicina species-groups had low probabilities for the first β-strand. The highest deviations from the general pattern were observed in *A. brassicae* and *E. novae-zelandiae*, which had eight probability peaks for the β-strand as did most other species, but probabilities for alternating turn motifs were low. To examine if there were changes in secondary structure prediction by sequence truncation,

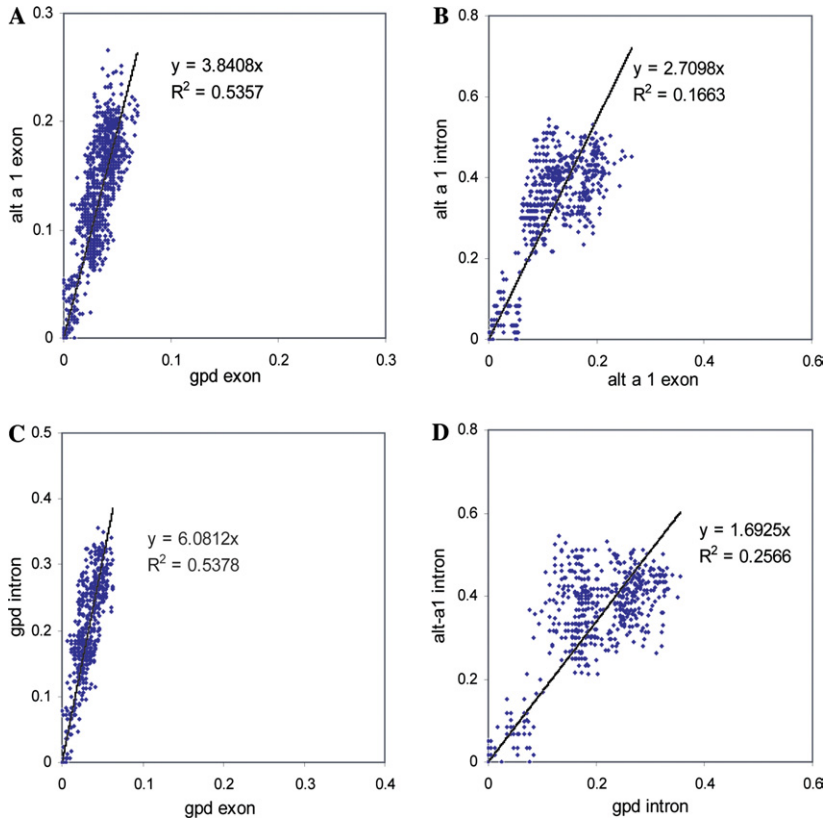

Fig. 5. Comparison of distance values between genes and domains. Uncorrected pairwise distance values of exon and intron domains were calculated for each pair of sequences and plotted. *X* and *Y* axes represent distance values for each domain.

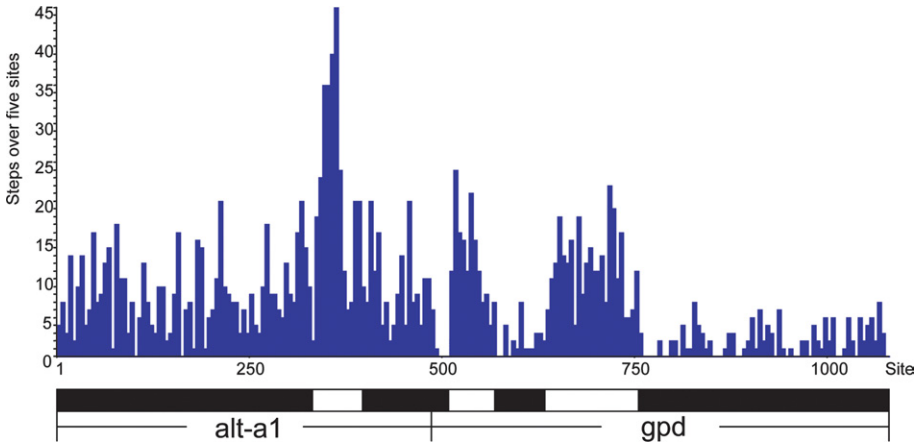

Fig. 6. Site variation of the Alt a 1 and *gpd* gene sequence of *Alternaria* and related species. The variation in base substitution numbers among sites across the length of the genes from 43 isolates of *Alternaria* and related species were assessed using a window size of five consecutive bases. The lower bar indicates the relative locations of exon and intron domains of Alt a 1 and *gpd* genes. Open boxes represent intron domains and filled boxes represent exon domains.

complete amino acid sequences retrieved from GenBank (AF288160 and AF499002) were used to compare secondary structure prediction. These results revealed that complete sequences were predicted to have two more β-sheet motifs between residues 116 and 152. Except for the β-strand spanning residues 116 and 124, prediction for all the other secondary structural motifs were not strongly affected by truncation of sequences.

4. Discussion

This study compares sequence evolution between two protein-coding genes, *gpd* and Alt a 1. As a molecular marker to study phylogenetic relationships, Alt a 1 exon regions contained considerably more parsimony-informative sites than *gpd* exon regions. Thus, Alt a 1 gene sequences produced phylogenetic trees with better

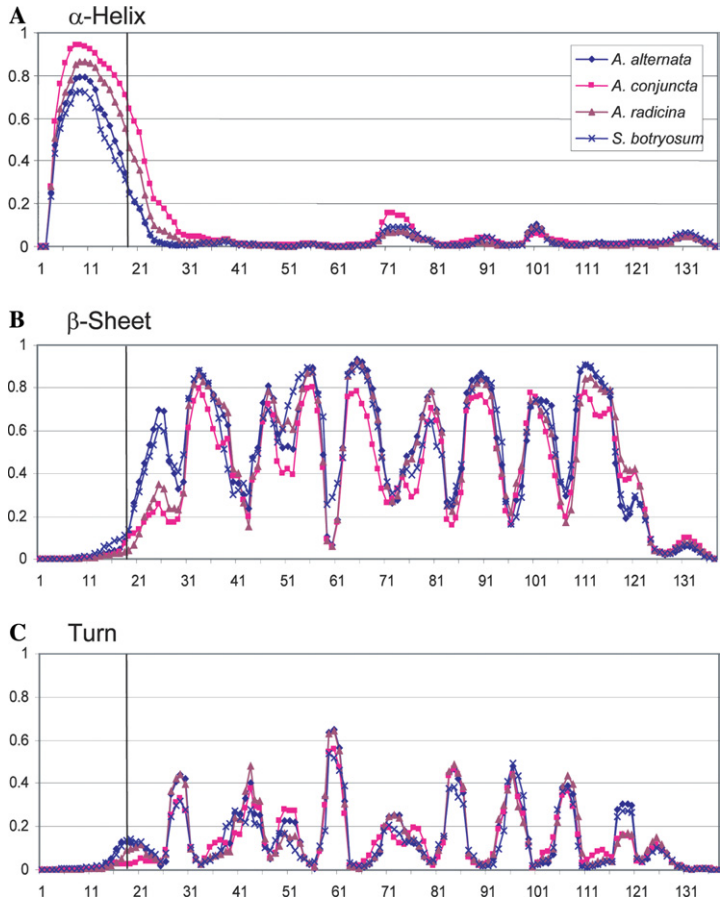

Fig. 7. Probabilities for secondary structural motifs of the Alt a 1 gene product for selected *Alternaria* and *Stemphylium* spp. predicted by PSA Protein Structure Prediction Server. *X* and *Y* axes represent amino acid position and probability for each structural motif, respectively. Vertical lines (aa position 19) indicate the N-terminus of the mature protein after dissociation of the hydrophobic signal sequence during protein processing.

resolution and higher bootstrap supports. Although there were some differences in relationships among certain species-groups and long-branched terminal taxa, many features of Alt a 1 and *gpd* gene phylogenies were similar.

In the phylogenetic analyses of Alt a 1 gene sequences, *gpd* sequences, and the combined data set, the previous proposal that the genus *Alternaria* is paraphyletic, and the genera *Embellisia* and *Ulocladium* are polyphyletic (Pryor and Bigelow, 2003; Pryor and Gilbertson, 2000) was supported. Grouping of *Alternaria* and related species into several species-groups was well supported for the infectoria, alternata, porri, and radicina species-groups. In this study, a new species-group, the sonchi species-group, was proposed for the monophyletic group consisting of *A. sonchi* and *A. cinerariae*. This group was strongly supported in the Alt a 1 gene tree and combined dataset tree. The phylogenetic position of *A. japonica*, resolved as a member of the brassicicola species-group in previous studies (Pryor and Bigelow, 2003), was not clearly resolved in this study.

The *Ulocladium* group described in previous works (Pryor and Bigelow, 2003; Pryor and Gilbertson, 2000) was maintained in this study in the *gpd* and combined data set trees, although bootstrap support was relatively low compared to most *Alternaria* species-groups. Regarding the *Embellisia* group, group membership established in this study was not the same as described by Pryor and Bigelow (2003). *E. allii* was included in the group in this study but was not in the previous study, whereas *E. novae-zelandiae* was not included in the group in this study but was in the previous study. In the previous work, *E. novae-zelandiae* was a member of a strongly supported *Embellisia* clade along with *E. hyacinthi*, *E. leptinellae*, and *E. proteae*. However, the latter three species were not included in this study due to difficulty in amplification of Alt a 1 homologs. As a result, *E. novae-zelandiae* did not resolve with any other taxa into a distinct clade. In the previous work, *E. allii* grouped with *Nimbya* spp. in a well supported *Nimbya* clade. However, with the inclusion of *E. telluster* in this study (not included in the previous work), a well supported *Embellisia* clade was resolved as a sister taxon to the *Nimbya* clade. This is more consistant with previously established morphotaxonomy as *E. allii* is the type species for the genus *Embellisia* and has features distinct from those of *Nimbya* species (Simmons, 1971).

Until more extensive phylogenetic studies are conducted with additional *Embellisia* taxa and/or loci, the *Embellisia* group established in this study may be considered a separate species-group distinct from the *Embellisia* group established previously. This would, in fact, support the hypothesized polyphyletic origin of *Embellisia* (Pryor and Bigelow, 2003).

The *Nimbya* group, which was supported in the previous study (Pryor and Bigelow, 2003), formed a monophyletic group in Alt a 1 and combined analyses, but not in *gpd* analysis alone as it did in the previous study. The difference between this study and the previous study, e.g., inclusion of additional taxa and exclusion of intron sequences, likely contributed to different findings. The fact that members of the *Nimbya* group have relatively long terminal branches, which are known to contribute to systematic error (Swofford et al., 1996), may also contribute to the instability of this group.

In contrast to the stable and well-defined grouping of taxa into several species-groups, relationships among species-groups and taxa with long terminal branches were not clear in this study, as in previous studies (Pryor and Bigelow, 2003; Pryor and Gilbertson, 2000). Regarding poor resolution in relationships among higher level lineages and among closely related species, it was shown that these controversial relationships could be better resolved by using more variable Alt a 1 gene sequences. However, these sequences alone could not resolve these relationships with a high level of statistical support. In the phylogenetic study of *Saccharomyces cerevisiae* and closely related species, a very stable and well supported phylogeny was obtained by combining sequence data from a genome scale study (Rokas et al., 2003). Likewise it is expected that phylogenetic relationships of rapidly diverging species of *Alternaria* and related genera could be better resolved by combining more sequence data with higher information content.

Data from DNA sequences revealed that the exon region of the Alt a 1 gene evolved more rapidly than the exon region of the *gpd* gene. This difference was even more pronounced in amino acid sequences. In the *gpd* sequences, most of the base substitutions were concentrated in the third codon position resulting in highly conserved amino acid sequences. In contrast, base substitution bias according to the codon position was not as high in the Alt a 1 gene resulting in highly variable amino acid sequences. These results suggested that the primary sequence of the Alt a 1 gene product is under relatively loose evolutionary constraints compared to glyceraldehyde-3-phosphate dehydrogenase. However, despite high sequence variation, overall secondary structure of Alt a 1 was predicted to be well preserved in most of the species. This result can be explained by the fact that a high proportion of amino acid changes in Alt a 1 are confined to the same physico-chemical groups (data not shown). Considering high primary sequence variation and conservation of secondary structure, it is proposed that structural architecture is under high selective constraint in the function of Alt a 1. The Alt a 1 gene was recovered by PCR amplification from most of the species examined in this study. However, a blastx search did not detect any homologs protein sequence from GenBank, which included completed or nearly completed genome sequence data for *S. cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Aspergillus nidulans*, and *Neurospora crassa*. This suggests that the Alt a 1 gene is unique to specific groups of fungi including *Alternaria* and related species.

The antibody-binding epitope of Alt a 1 was determined by human IgE and rabbit anti-Alt a 1 IgG recognition of the synthesized peptide corresponding to the amino acid sites 4–23 of the mature protein (Breitenbach and Simon-Nobbe, 2002; Zhang et al., 1995). It was also demonstrated that rabbit antibodies induced by the synthesized peptide recognized the native Alt a 1 allergen. The peptide sequence corresponded to the amino acid sites 22–41 in Fig. 7, which form the first and the second β-strand motifs. This region varied in probabilities for secondary structural motifs. Interestingly, other antibody-binding epitope regions (47–56 and 112–121 aa positions in Fig. 7) identified from IgE binding assay against synthetic peptides (Kurup et al., 2003), also had varied probabilities for secondary structural motifs. To better understand the epidemiology of human allergenic disorders caused by *Alternaria* species, it will be important to examine whether Alt a 1 homologs from diverse *Alternaria* species are cross-reactive to the IgE against specific Alt a 1 despite variation in primary and secondary structure.

Acknowledgments

This study was partially supported by the College of Agriculture and Life Sciences, University of Arizona; College of Agricultural Sciences, Colorado State University; the Arizona Disease Control Research Commission (ADCRC #7006); and the National Science Foundation (DBI #0227436). The authors would also like to acknowledge Drs. Yeisoo Yu and Rod A. Wing of the Arizona Genomics Institute for performing much of the DNA sequencing in this study.

References

Achatz, G., Oberkofler, H., Lechenauer, E., Simon, B., Unger, A., Kandler, D., Ebner, C., Prillinger, H., Kraft, D., Breitenbach, M., 1995. Molecular cloning of major and minor allergens of *Alternaria alternata* and *Cladosporium herbarum*. Mol. Immunol. 32, 213–227.

Barnes, C.S., Pacheco, F., Landuyt, J., Rosenthal, D., Hu, F., Portnoy, J., 1996. Production of a recombinant protein from *Alternaria* containing the reported N-terminal of the Alt a 1 allergen. In: Sehon,

A., Kraft, D., Hay Glass, K.T. (Eds.), Advances in Experimental Medicine and Biology, vol. 409. Plenum Press, New York, pp. 197–203.

Berbee, M.L., Pirseyedi, M., Hubbard, S., 1999. *Cochliobolus* phylogenetics and the origin of known, highly virulent pathogens, inferred from ITS and glyceraldehyde-3-phosphate dehydrogenase gene sequences. Mycologia 91, 964–977.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., Chambon, P., 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. Proc. Natl. Acad. Sci. USA 75, 4853–4857.

Breitenbach, M., Simon-Nobbe, B., 2002. The Allergens of *Cladosporium herbarum* and *Alternaria alternata*. In: Breitenbach, M., Crameri, R., Lehrer, S.B. (Eds.), Fungal Allergy and Pathogenicity. Chemical Immunology, vol 81. Karger, Basel, pp. 48–72.

Chou, W.Q., Wu, W.S., 2002. Phylogenetic analysis of internal transcribed spacer regions of the genus *Alternaria*, and the significance of filament-beaked conidia. Mycol. Res. 106, 164–169.

Chun, J., 1995. Computer-assisted classification and identification of actinomycetes (PhD thesis). University of Newcastle.

Cramer, R.A., Lawrence, C.B., 2003. Cloning of a gene encoding an Alt a 1 isoallergen differentially expressed by the necrotrophic fungus *Alternaria brassicicola* during *Arabidopsis* infection. Appl. Environ. Microbiol. 69, 2361–2364.

Cramer, R.A., Lawrence, C.B., 2004. Identification of *Alternaria brassicicola* genes expressed in planta during pathogenesis of *Arabidopsis thaliana*. Fungal Genet. Biol. 41, 115–128.

de Hoog, G.S., Horre, R., 2002. Molecular taxonomy of the *Alternaria* and *Ulocladium* species from humans and their identification in the routine laboratory. Mycoses 45, 259–276.

De Vouge, M.W., Thanker, A.M., Curran, I.H.A., Zhang, L., Muradia, G., Rode, H., Vijay, H.M., 1996. Isolation and expression of a cDNA encoding on *Alternaria alternata* Alt a 1 subunit. Int. Arch. Allergy Immunol. 111, 385–395.

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

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–112.

Kurup, V., Vijay, H.M., Kumar, V., Castillo, L., Elms, V., 2003. IgE binding synthetic peptides of Alt a 1, a major allergen of *Alternaria alternata*. Peptides 24, 179–185.

Maddison, W.P., Maddison, D.R., 2000. MacClade: Analysis of Phylogeny and Character Evolution. Version 4.0. Sinauer Associates, Sunderland, MA.

Osiewacz, H.D., Ridder, R., 1991. Genome analysis of imperfect fungi: electrophoretic karyotyping and characterization of the nuclear gene coding for glyceraldehyde-3-phosphate dehydrogenase (*gpd*) of *Curvularia lunata*. Curr. Genet. 20, 151–155.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Pryor, B.M., Bigelow, D.M., 2003. Molecular characterization of *Embellisia* and *Nimbya* species and their relationship to *Alternaria*, *Ulocladium*, and *Stemphylium*. Mycologia 95, 1139–1152.

Pryor, B.M., Gilbertson, R.L., 2000. Molecular phylogenetic relationships amongst *Alternaria* species and related fungi based upon analysis of nuclear ITS and mt SSU rDNA sequences. Mycol. Res. 104, 1312–1321.

Pryor, B.M., Michailides, T.J., 2002. Morphological, pathogenic, and molecular characterization of *Alternaria* isolates associated with *Alternaria* late blight of pistachio. Phytopathology 92, 406–416.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstruction phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Simmons, E.G., 1971. *Helminthosporium allii* as type of a new genus. Mycologia 63, 380–386.

Simon-Nobbe, B., Probst, G., Kajava, A.V., Oberkofler, H., Susani, M., Crameri, R., Ferreira, F., Ebner, C., Breitenbach, M., 2000. IgE-binding epitopes of enolases, a class of highly conserved fungal allergens. J. Allergy Clin. Immunol. 106, 887–895.

Stultz, C.M., White, J.V., Smith, T.F., 1993. Structural analysis based on state-space modeling. Protein Sci. 2, 305–314.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics, second ed. Sinauer Associates, Sunderland, MA, pp. 407–514.

Van Wert, S.L., Yoder, O.C., 1992. Structure of the *Cochliobolus heterostrophus* glyceraldehyde-3-phosphate dehydrogenase gene. Curr. Genet. 22, 29–35.

Weichel, M., Schmid-Grendelmeimer, P., Fluckiger, S., Breitenbach, M., Blaser, K., Crameri, R., 2003. Nuclear transport factor2 represents a novel cross-reactive fungal allergen. Allergy 58, 198–206.

Zhang, L., Curran, I.H.A., Murada, G., De Vouge, M.W., Rode, H., Vijay, J.M., 1995. N-terminus of a major allergen, Alt a 1, of *Alternaria alternata* defined to be an epitope. Int. Arch. Allergy Immunol. 108, 254–259.