

This article was downloaded by: [Advanced Institute of Sanita]
[Advanced Institute of Sanita]

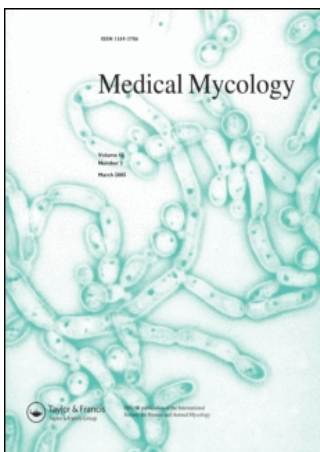
On: 4 March 2007

Access Details: [subscription number 769486267]

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



Medical Mycology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713694156>

Genomic analysis of allergen genes in *Aspergillus* spp.: the relevance of genomics to everyday research

To link to this article: DOI: 10.1080/13693780600972907

URL: <http://dx.doi.org/10.1080/13693780600972907>

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

Genomic analysis of allergen genes in *Aspergillus* spp.: the relevance of genomics to everyday research

PAUL BOWYER & DAVID W. DENNING

Faculty of Medicine and Human Sciences, University of Manchester; Wythenshawe Hospital, Manchester, UK

Full genomic sequencing of *Aspergillus fumigatus* and other genomes has allowed correction of *Aspergillus* allergen gene sequences and requires revision of Genbank and IUIS sequences of allergens. In addition allergens in other fungal species may be found in the aspergilli. We compared the published sequences of numerous allergens with recently available genome sequences. This analysis suggests that Asp f 56KD, Asp f 15 and Asp f 16 should be removed from the approved allergen list and that Asp f 17 is a larger protein than published. Additionally we propose likely gene candidates for Asp f GST (Afu6g09690) and Asp o lipase (AO090701000644). We suggest that the heat shock allergens should be re-classified: Asp f 12 should be classified as HSP90 (Asp f 12), HSP88 (Mala s 10) and HSP70 (Alt a 3, Cla h 4 and Pen c 19) according to human gene nomenclature. Comparison of fungal allergen databases with genome sequences suggests the presence of a core set of allergen – like proteins in all fungi. We also analysed allergens in the 3 sequenced aspergilli to look for internal homologies and this suggests that multi gene families may produce numerous cross-reactive allergens.

Keywords *Aspergillus*, allergen, genomics

Introduction

Allergy is an increasingly common and debilitating condition [1,2]. The causative agents of allergy are commonly eukaryotic proteins apparently mostly restricted to a number of functional classes [3,4]. Identification of the genes encoding allergens is an essential first step in characterization of the allergen, its epitopes and for studies involving diagnosis and cross-reaction where pure protein preparations are necessary. Much past and current effort has gone into discovering allergen gene sequences to define the nature of allergenic proteins [5–8] and allow manipulation of allergen gene sequences for definition of epitopes [9–11], production of component resolved diagnostics [12–15] or to direct design of a variety of immunotherapeutics [16–24]. Thus it is of critical importance that

the correct allergen sequences are known for existing allergens to prevent possible errors in epitope mapping or diagnostic studies. In diagnostic studies it is important to be confident that an allergen is confined to a particular species or genus. Similarly it is important to know whether a particular allergen has multiple homologues in a given species' genome in order to be sure that the allergen being studied is in fact the particular protein or cDNA that has been isolated and not merely a cross-reacting member of a large gene family present in a particular species.

Fungi are an important source of allergen proteins and have been shown to produce ~90 allergens from a total of ~750 published in databases (Bowyer, unpublished observations). The estimation of allergen numbers arises from uncertainties regarding identification of some allergens with some being multiply entered in databases under different names and others being possibly incorrect. Fungi are estimated to sensitise ~20% of the atopic population [25,26] although they may not be always the sole or causative agent. One important allergenic fungus is *Aspergillus fumigatus* with 20 described allergens [27–29]. This fungus is also

Received 7 July 2006; Accepted 23 August 2006

Correspondence: Paul Bowyer, Faculty of Medicine and Human Sciences, University of Manchester, Wythenshawe Hospital, Manchester, M23 9LT, UK. Tel: +44 (0)161 291 5847; Fax: +44 (0)161 291 5806; E-mail: paul.bowyer@manchester.ac.uk

important as a pathogen of immunocompromised individuals [30] and as a complicating allergenic factor in asthmatics [31,32]. Although ~20 allergens have been described in *A. fumigatus* only four are known in *A. oryzae*, 5 in *A. niger* and a further 2 in *A. flavus* [24 www.allergome.org]. A possible allergen Asp ni 2 (previously Asp nd1) has also been described in *A. nidulans* [33]. Other species of *Aspergillus* known to be allergenic such as *A. clavatus*, *A. versicolor*, *A. sydowii* and *A. ochraeus* have no allergens described. The genome of *A. fumigatus* has recently been sequenced [34].

Here we use the genome sequence to definitively catalogue or discard previously published cloned allergen genes, to show the presence of homologues to other fungal allergens in the genome of *A. fumigatus*, to assess the level of allergen homology in the genome in the form of multi-gene families and determine the gene and predicted protein sequences of allergens that had previously been partially characterized at the biochemical level. In order to search the genome sequences we have used BLAST (Basic Local Alignment Search Tool) algorithm which compares sequences and produces a report containing levels of identity and statistical measures of similarity. The analysis shows that three of the published Af allergens are duplications of existing allergen genes and provides gene sequences for a further three partially described *Aspergillus* allergens. Nine of 17 Af allergens belong to multi-gene families with six of these having one or more possible cross reacting homologues. Additionally the analysis suggests that fungi may possess a core set of at least 40 allergens or proteins that may cross-react.

Materials and methods

Source of sequences for comparison

Allergen gene sequences were compiled from databases at Allergome.org, allermatch, SDAP, and Swissprot [35–38]. Duplicated sequences were removed and peptide sequences checked to ensure that the longest sequences were represented. The final list contained 86 unique fungal allergen sequences. The database was compiled as a FASTA file (a type of simple sequence format). Genome sequences were obtained from The Institute for Genomic Resources (www.TIGR.org). Isolates represented are Af293 and CEA10 for Af, RIB40 for Ao and FGSC A4 for An. Allergens used in this study together with their organism of origin and predicted function are shown in Table 1.

Sequence analysis

Databases and individual sequences were used to search locally stored and formatted genome peptide database files using NCBI formatdb and NCBI standalone BLAST2.0 [39]. BLASTP, TBLASTN and BLASTX [40] were used to search nucleotide and protein sequence databases. Both unlimited searches and cut-offs of $1E^{-10}$ were used to derive data in Tables 2 and 3. E scores represent the probability of chance similarity to the query sequence being present in the database that is being searched so that lower scores indicate higher confidence in the result. We have used a highly conservative BLAST cut off score of $1E^{-10}$ to define a significant match. Although such a figure is necessarily rather arbitrary this level of homology corresponds to sequence homology >25% for proteins with >100 amino acids and it was reasoned that proteins with lower levels of homology are unlikely to be related to the query sequence. In general the cutoff score for data presented in Tables 2 and 3 is $1E^{-10}$ but for cases where no match was found the results were checked using a cutoff of $E=10$ to be sure that no match existed. Identity percentages were then checked manually to ensure that the whole predicted peptide sequence was included in the result. Matched regions were compared to predicted peptides from the automated gene calling analysis of the genome sequences and against other fungal homologues. Where the peptide was thought to be longer or shorter the surrounding sequence was analysed in all three reading frames using BLASTX and manually using DNASTAR Mapdraw to check that the peptide sequence had not been wrongly truncated for example by the presence of a stop codon containing intron or by sequencing error. Where partial sequences were used to find complete *A. fumigatus* genes the sequences were formatted as FASTA files then used to search locally stored databases.

Results

Characterization of partially sequenced or unsequenced allergens from *Aspergillus* species

Asp f I–Asp f II. These allergens are all present in the Af genome as described in the literature with no differences in the coding regions of the genes.

Asp f 12 [41]. This allergen is variously described as HSP70 or HSP 90 and grouped as a functional class with other heat shock protein allergens, Alt a 3, Mala s 10, Cla h 4, and Pen c 19. However when these genes are compared with *A. fumigatus* and human

Table 1 Allergens used in the comparative analysis. Allergens are listed according to their IUIS nomenclature alongside their organism of origin and predicted function.

Allergen name (IUIS)	Protein function	Organism name
Alt a 1		
Alt a 2		
Alt a 3	HSP70	
Alt a 4	Disulfide isomerase	
Alt a 6	Ribosomal P2	
Alt a 7		<i>Alternaria alternata</i>
Alt a 10	Aldehyde dehydrogenase	
Alt a 11	Enolase	
Alt a 12	Ribosomal P1	
Alt a GST	Glutathione S transferase	
Alt a NTF2	Nuclear transport factor 2	
Asp f 1	Mitogillin/ribonoclease toxin	
Asp f 2		
Asp f 3	Peroxiredoxin	
Asp f 4		
Asp f 5	Metalloprotease	
Asp f 6	Mn SOD	
Asp f 7		
Asp f 8	Ribosomal P2	
Asp f 9		<i>Aspergillus fumigatus</i>
Asp f 10	Aspergillopepsin	
Asp f 11	Cyclophilin	
Asp f 12	HSP90	
Asp f 13	Serine protease	
Asp f 17		
Asp f 18	Vacuolar serine protease	
Asp f 22	Enolase	
Asp f 23	Ribosomal L3	
Asp fl 1 (Asp fl 13)	Serine protease	<i>A. flavus</i>
Asp n 14	β -xylosidase	
Asp n 18	Vacuolar serine protease	<i>A. niger</i>
Asp n 25	Phytase B	
Asp n glucoamylase		
Asp o 13	Oryzin	<i>A. oryzae</i>
Asp o 21	α -amylase A	
Can a 1 (Cand a 1)	Alcohol dehydrogenase	
Can a peroxiredoxin		
(Can a 2)	Peroxiredoxin	<i>Candida albicans</i>
Cand a Enolase	Enolase	
Candida 37KD	Aldolase	
Cla h 3	Aldehyde dehydrogenase	
Cla h 5 (Cla h 7a)	Ribosomal P2	
Cla h 6	Enolase	
Cla h 12	Ribosomal P1	<i>Cladosporium herbarum</i>
Cla h CH	Hydrophobin	
Cla h HSP70 (Cla h 4)	HSP70	
Cla h NTF2	Nuclear transport factor 2	
Cop c 1		<i>Coprinus comatus</i>
Cop c 2	Thioredoxin	
Cop c 3		
Cop c 5		
Cop c 7		
Emb nz1	(Alt a 1-like)	<i>Alternaria</i> spp.
Emb t 1	(Alt a 1-like)	
Fus c 1	Ribosomal P2	<i>Fusarium culmorum</i>
Fus c 2	Thioredoxin	
Fus c 3	Helix-loop-helix protein	

Table 1 (Continued)

Allergen name (IUIS)	Protein function	Organism name
Mala f 1	Peroxiredoxin	<i>Malassezia furfur</i>
Mala f 2		
Mala f 3		
Mala f 4		
Mala s 1	Peroxiredoxin	<i>M. sympodialis</i>
Mala s 5		
Mala s 6		
Mala s 7		
Mala s 8	HSP88	<i>M. sympodialis</i>
Mala s 9		
Mala s 10		
Mala s 11		
Mala s 12	MnSOD	
Pen c 1	Alkaline serine protease	<i>Penicillium citrinum</i>
Pen c 18	Vacuolar serine protease	
Pen c 19	HSP70	
Pen c 2	Alkaline serine protease	
Pen c 3	Peroxiredoxin	
Pen c 22	Enolase	
Pen c 24		
Pen ch 68 KD		
Pen ch13	Alkaline serine protease	<i>P. chrysogenum (notatum)</i>
Pen n 13	Subtilisin	
Pen n 18	Serine protease	
Pen o 18	Vacuolar serine protease	<i>P. oxalicum</i>
Psi c 2	Cyclophilin	<i>Psilocybe cubensis</i>
Rho m 1	Enolase	<i>Rhodotorula mucilaginosa</i>
Rho m 2	Vacuolar serine protease	
Sac c Enolase	Enolase	<i>Saccharomyces cerevisiae</i>
Sac c MnSOD	Mn SOD	
Tri m 2	Serine protease	<i>Trichophyton meningotrichoides</i>
Tri m 4	Dipeptidyl-peptidaseV	
Tri r 2	Alkaline protease	<i>T. rubrum</i>
Tri r 4	Dipeptidyl-peptidase V	
Tri sc 4	Dipeptidyl-peptidaseV	<i>T. schoenlieni</i>
Tri t 4	Serine protease	<i>T. tonsurans</i>

genomes (whence the original gene notation is derived) it was found that they had homology to 3 distinct classes of heat shock protein. Asp f 12 should be classified as HSP90, Mala s10 as HSP88 and Alt a 3, Cla h 4 and Pen c 19 as HSP70 according to human gene nomenclature. Genes for HSP 70, 88 and 90 are present in all *Aspergillus* genomes.

Asp f 15 [42]. This gene is identical to *Asp f 13* (Afu2g12630). No other gene matching this sequence exists in the genome thus *Asp f 13* was taken to be the original and only allergen of this type.

Asp f 16 [43–47]. The predicted peptide sequences of the published *Asp f 16* peptide and the *Asp f 9* sequence derived from the Af293 genome sequence are aligned in Fig. 1 showing that regions of the two allergens are identical with others that do not match. The *Asp f 16*

protein does not appear to be encoded by the *A. fumigatus* Af293 genome. We also compared the sequence against the recently determined *A. fumigatus* CEA10 strain genomic sequence and again find only the *Asp f 9* sequence (shown as 1.m01486 in Fig. 1) with no gene matching *Asp f 16*. Comparison of the *Asp f 16* sequence with other *Aspergillus* genomes shows matches with only *Asp f 9* – like sequences with no evidence for the existence of an *Asp f 16* homologue. The closest match is to *Asp f 9* (Afu1g16190). In fact the sequence of the *Asp f 16* cDNA is almost identical (99.1%) to the *Asp f 9* DNA sequence and it appears that the published *Asp f 16* peptide sequence results from frameshift or other sequencing errors (shown by asterisks in Fig. 1) in the original cDNA clone. Assuming the existence of these sequencing errors the *Asp f 16* cDNA clone will produce a peptide sequence identical to *Asp f 9*. One explanation is that the *Asp f 9*

Table 2 Comparative analysis of fungal allergens. Allergens used as the basis for comparison are shown in bold type in shaded boxes. Proteins that have been compared to these allergen sequences are shown with % identity to the reference allergen in brackets. Matches to the predicted Af proteome are shown using CADRE nomenclature.

<i>Aspergillus fumigatus</i> Gene accession	Type allergens	Related fungal allergens
Asp f1 Afu5g02330 Asp f 2 Afu4g09580 Asp f 3 Afu6g02280		Cand b 2 (39); Pen c 3 (82); Can a 2 (40); Mala f 2 (44); Mala f 3 (46); Mala s 5 (?) (48); Mala s 1 (43)
Asp f 4 Afu2g03830 Asp f 6 Afu1g14550 Asp f 5 Afu8g07080 Asp f 7 Afu4g06670 Asp f 8 Afu2g10100 Asp f 9 Afu1g16190 Asp f 10 Afu5g13300 Asp f 11 Afu2g03720 Asp f 12 Afu5g04170 Asp f 13 Afu2g12630 Asp f 17 Afu4g03240 Asp f 18 Afu5g09210		Mala s 11 (52); Sac c MnSOD (53)
Asp f 22 Afu6g06770		Alt a 6 (82); Cla h 3 (80); Fus c 1 (81);
Asp f 23 Afu2g11850		Mala s 6 (62) (72 to Afu3g07430); Psi c 2 (61) Mal s 10 (36) Asp fl 13 (81) Pen c 13 fragment*; Tri t 4 fragment*
Afu1g12610 (52) Afu1g07440 (61) Afu1g06830 (80) Afu4g11800 (44) Afu2g09030 (59) Afu6g11430 (67) Afu5g06240 (59) Afu4g00610 (40) Afu3g10460 (58) Afu2g06150 (62) Afu1g02820 (69) Afu3g11690 (75) Afu1g16920 (63) Afu2g00690 (63) Afu4g10130 (78) Afu6g11330 (31) Afu5g09580 (28) Afu8g05020 (79) Afu1g11190 (65) Afu6g11030 (39) Afu8g01090 (53) Afu4g09090 (52) Afu1g09670 (42) Afu7g05740 (65) Afu6g09690 (94) Afu5g02040 (42)	Mala s 10 Alt a 3 Alt a 12 Asp o 13 Tri r 4 Alt a 10 Can a 1 Mala s 12 Alt a NTF2 Alt a 4 Alt a 7 Candida 37KDa allergen Asp n 14 Asp n glucoamylase Asp o 21 Asp n 25 Cla h CH Pen ch 20 Pen c 24 Cop c 3 Cop c 2 Fus c 2 Fus c 2 Fus c 3 Mala f 4 Alt a GST Asp o lipase	Asp n 18 (69); Pen n 18 (68); Pen n 13 (40); Pen o 18 (70); Pen c18 (72); Pen c2 (82); Pen c 1 (38); Tri r2 (40); Tri m 2 (46); Rho m 2 (52); Cla h 9 Alt a 11 (prev 5)(87); Cla h 6 (83); Pen c 22 (90); Cand a enolase (73); Rho m 1 (75); Sac c Enolase (72) Cla h 4 (85); Pen c19 (90) Cla h 12 (71) Asp fl 1 (78) Tri sc 4 (57); Tri m 4 (57) Cla h 3 (71) Cla h NTF2 (64) Cla h 5 (61) Fus c 2 (53) Cop c 2 (53)

*Peptide sequences <50AA are not scored.

sequence is erroneous. Given the 10 fold sequencing of the genome and lack of any known errors in this sequence for any of the *A. fumigatus* allergens together with the predicted presence of Asp f 9 – like rather than

Asp f 16 – like genes in the other aspergilli we suggest that the Asp f 9 sequence is correct and that Asp f 16 should henceforth be known as Asp f 9. It is also possible that the published Asp f 16 protein is an

Table 3 Occurrence of allergens and allergen-like proteins in multi gene families in the aspergilli. Numbers of matches to the reference allergen are shown in bold type. Matches less significant than a BLAST score of $1E^{-10}$ were excluded. Percent identity of the matches is shown bracketed after the number of matches. 'None' indicates no significant match with a BLAST cutoff of $1E^{-10}$ although no matches are found in these cases with a BLAST cutoff of $E = 1$.

Described Af allergen	Af homologues	An homologues	Ao homologues
Asp f1	1	None	None
Asp f2	1	6 (60, 52, 49, 46, 42)	7 (65, 44, 43, 42, 39, 35, 35)
Asp f3	2 (100, 62)	2 (90, 73)	3 (86, 68, 37)
Asp f4	3 (100, 61, 61)	3 (60, 58, 48)	3 (60, 55, 48)
Asp f5	1	None	2 (78, 64)
Asp f6	2 (100, 52)	2 (67, 51)	2 (61, 51)
Asp f7	3 (100, 82, 30)	3 (73, 69, 30)	3 (61, 43, 36)
Asp f8	1	2 (78, 64)	1 (81)
Asp f9	5 (100, 47, 44, 40, 34)	6 (68, 60, 43, 39, 35, 35)	7 (53, 52, 48, 43, 40, 39, 36)
Asp f10	3 (100, 40, 36)	7 (73, 42, 30, 29, 29, 28, 26)	7 (71, 44, 42, 30, 30, 29, 27)
Asp f11	9 (100, 66, 57, 51, 56, 62, 52, 47, 47)	7 (81, 61, 58, 55, 54, 54, 47)	8 (75, 69, 59, 56, 54, 54, 49, 46)
Asp f12	1	1 (95)	1 (95)
Asp f13	1	1 (52)	1 (58)
Asp f17	3 (100, 50, 37)	3 (45, 43, 29)	3 (57, 40, 29)
Asp f18	3 (100, 44, 43)	3 (78, 45, 24)	2 (86, 45)
Asp f22	1	1 (93)	1 (93)
Asp f23	1	1 (95)	1 (95)

isoform of Asp f 9 that occurs in the AF-102 isolate used in its isolation but not in other *Aspergillus* isolates or species. It would be relatively simple to resequence the gene from AF-102 to resolve this issue. It is interesting to note that where the Asp f 16 peptide sequence has been used to define epitope sequences (shown as shaded boxes in Fig. 1) or reactivity only those sequences matching Asp f 9 have produced a reaction suggesting this work has been largely valid although perhaps somewhat incomplete [44–47].

Asp f 17 [48]. Asp f 17 has previously been lodged in public data bases as a 191 AA fragment. The complete gene (Afu4g03240) encodes a protein of 284 amino acids, with a predicted MW of 27kD.

Asp f 18, Asp f 22, Asp f 23. Asp f 22 and Asp f 23 are present in the Af genome as described in literature with no differences in the coding regions of these genes.

Asp f GST. This allergen was predicted to occur through cross-reaction to Alt a 13 (also referred to as Alt a GST) an *Alternaria* glutathione s-transferase allergen [49]. Estimates of the allergen molecular weight and enzyme activity confirmed its identity as a ~26kD glutathione s-transferase. There are 15 predicted GST genes in the *A. fumigatus* genome (Afu1g01370, Afu1g17010, Afu2g00590, Afu2g08370, Afu2g17300, Afu3g07930, Afu3g10830, Afu4g01440, Afu4g14100, Afu4g14530, Afu6g00760, Afu6g03390, Afu6g09690, Afu7g05500, Afu8g02500). Several fit the

predicted size of 26kD for the Asp f GST allergen however only Afu6g09690 has significant homology (94.8%) to the cross-reacting Alt a 13 allergen. The other genes have <30% identity to Alt a 13. Thus we predict that Afu6g09690 encodes the Asp f GST allergen.

Asp f 56 kD [50]. The short peptide sequence reported for this allergen is not predicted to be encoded in any of the *Aspergillus* genomes studied. This may mean that the protein is not derived from *Aspergillus* or more likely that the peptide sequencing did not produce an accurate enough result for correct identification.

Asp o lipase. This allergen was described as an occupational lipase allergen in detergent workers [51,52] but was never characterized beyond determination of a short peptide sequence (XIPTTQLXXFKFWV-QYAAA). BLAST alignment of the partial sequence with the *A. oryzae* genome shows that the Asp o lipase allergen is encoded by gene AO090701000644 on chromosome 5. This gene is predicted to produce a 33.5 kD lipase protein.

Comparative analysis of allergen genes in fungi

In order to discover whether fungal allergens are unique to their particular species or occur in all fungi we compiled a database of 86 fungal allergens and used BLAST to compare the database with the *A. fumigatus* genome and against itself. Results of this analysis are shown in Table 2. We note that a small portion of this

Aspf16 1	MYFKYTAALA [*] AVLPLCSAQ [*] TWSKCNPLAETCPPNKGLAASTCTADFTSASALDQWEVTA 60
Asp f9 1	MYFKYTAALA [*] AVLPLCSAQ [*] TWSKCNPL +TCPPNKGLAAST TADFTSASALDQWEVTA 60
1.m01486	MYFKYTAALA [*] AVLPLCSAQ [*] TWSKCNPLEKTCPPNKGLAASTYTADFTSASALDQWEVTA
Aspf16 61	GKVPVGPQGA [*] EFTVAKQGD [*] APTIDTDFYFFFGKAEVVMKAAPGTGVVSSIVLESDDLDEV 120
Asp f9 61	GKVPVGPQGA [*] EFTVAKQGD [*] APTIDTDFYFFFGKAEVVMKAAPGTGVVSSIVLESDDLDEV 120
1.m01486	GKVPVGPQGA [*] EFTVAKQGD [*] APTIDTDFYFFFGKAEVVMKAAPGTGVVSSIVLESDDLDEV
Aspf16 121	DLVRLGGDTTQVQ [*] TNYFGK [*] GDTTTYDRGTYVPVATPQETFHTY [*] TI [*] DWTKDAVTWSIDGAV 180
Asp f9 121	D LGGDTTQVQ [*] TNYFGK [*] GDTTTYDRGTYVPVATPQETFHTY [*] TI [*] DWTKDAVTWSIDGAV 180
1.m01486	DWEVLGGDTTQVQ [*] TNYFGK [*] GDTTTYDRGTYVPVATPQETFHTY [*] TI [*] DWTKDAVTWSIDGAV
Aspf16 181	VRTLTYNDAKGG [*] TRFPQTPMRLRLAAGPAATPATPGH - -HRVGRWLDRLQRGTVHHVRQV 238
Asp f9 181	VRTLTYNDAKGG [*] TRFPQTPMRLRL + P+ P G D +V+ V 240
1.m01486	VRTLTYNDAKGG [*] TRFPQTPMRLRLGSwAGGDPSNPKGTIEWAGGLTDYSAGPYTMYVKSV
Aspf16 239	RPYRERQPRRVLHLLGQLWLLAEHQVRRLLRRYSSSSSVTSSTTSTASSASSTSSKTPSTS 298
Asp f9 241	R - IENANPAESY [*] TYSDNSGSWQSIKFDGSDV [*] DISSSSSVTSSTTSTASSASSTSSKTPSTS 299
1.m01486	R - IENANPAESY [*] TYSDNSGSWQSIKFDGSDV [*] DISSSSSVTSSTTSTASSASSTSSKTPSTS
Aspf16 299	TLATSTKATPTPSGTSSGSNSSSSAEPTTTGGSGSSNTGSwLRLLRLWLWYSSSTGSSTSA 358
Asp f9 300	TLATSTKATPTPSGTSSGSNSSSSAEPTTTGG+GSSNTGS SSGSSTSA 359
1.m01486	TLATSTKATPTPSGTSSGSNSSSSAEPTTTGGSGSSNTGSGSGSGSGSGSSSTGSSTSA
Aspf16 359	GASATPELSQGAAGSIKGSVT [*] PALWCSAPSLPCWHSKQNDDFGLMHDTHHEGDVRTIHFG
Asp f9 360	GASATPELSQGAAGSIKGSVTACALVFGAVA [*] AVLAF 395
1.m01486	GASATPELSQGAAGSIKGSVTACALVFGAVA [*] AVLAF 395
Aspf16 419	IGVSPSGFV 428

Fig. 1 Comparison of Asp f 9 and Asp f 16 genes. Peptide sequences of Asp f 9 and Asp f 16 were aligned. The Asp f 9 homologue, 1.m01486, from the CEA10 genome sequence is also presented for comparison. Asterisks mark the position of proposed errors in the Asp f 16 cDNA sequence.

data has been previously published [34] but is presented here for completeness. A number of Asp f allergens such as Asp f 18, Asp f 22 and Asp f 3 are clearly similar in function and protein identity to allergens from a wide range of fungi. Other major allergens such as Asp f 1, 7, 9 and 17 Alt a 1 and Alt a 2 as well as Cla h 1 and Cla h 2 have no recorded homologues in any other fungus. Interestingly 19 allergens from other fungal species have >50% identity with *A. fumigatus* genes, suggesting that the *A. fumigatus* proteins could possibly cross react with serum from patients sensitized to these fungi and that these 19 proteins are possible cryptic allergens in *A. fumigatus*. It is possible that the cross-homologous proteins and allergens shown in Table 2 represent a core set of allergens present in all fungi and that such comparative analyses can be used to predict the likely presence of allergen genes in all fungi.

Allergen-like proteins occur as families in allergenic and non-allergenic aspergilli

In order to determine whether allergen genes occur in genomes as single copies or as part of closely related multi-gene families sequences were matched against the Af, Ao and An genome DNA sequences and predicted peptide sequences. Results are presented in Table 3. As expected, all recorded Af allergen sequences with the exceptions noted below, were present in the Af genome sequence and in the predicted Af proteome. All allergen database sequences showed 100% identity to the genome sequence and predicted peptides. Af gene sequences are shown using CADRE gene nomenclature compatible with the curated Af genome sequence at <http://www.cadre.man.ac.uk>.

Af appears to possess many allergen homologues with sufficiently high identity to be classified as potential cross-reactive proteins. Each of the 17 Af

allergens tested was present as a single copy gene in Af. However 12 homologues of >50% identity and a further 8 homologues of between 40 and 50% identity were identified within the Af genome (Table 3). More significant potentially cross-reactive allergen homologues were detected in Ao than either An or Af despite Af to being the allergenic species. Ao has 25 predicted proteins with >50% identity to the Af allergens and 14 with between 40 and 50% identity whereas An has 25 predicted proteins with >50% identity to the Af allergens and 9 with between 40 and 50% identity. All primary allergen homologues from Ao and An are all >60% identical with f3, f5, f7, f8, f11, f12, f18, f22 and f23 all having >70% identity. This suggests that An and Ao should produce many proteins cross-reactive to those of Af. Given that the homologous proteins have predicted sizes close ($\pm 5\%$) to the characterized allergens and that individuals only become sensitized to a small number of the potential allergen repertoire it is possible to confuse allergen and allergen-like family members. Only two Af allergens, Asp f 1 and Asp f 5 did not have homologues in both An and Ao. The major allergen Asp f 1 did not have any homologues in either Ao or An. Asp f5 did not have any homologue in An although a close homologue exists in Ao. The significance of this observation awaits experimental elucidation however a naïve hypothesis could be that Af is the predominant allergenic organism due to the presence of Asp f 1. Asp f 1 has no paralogs in the *Aspergillus* genomes analysed here but closely related proteins are known to occur in *A. giganteus* and *A. clavatus*.

Discussion

The results presented above show the usefulness of genomic sequencing in gene discovery approaches. Knowledge of the genome sequence has enabled us to correct a variety of allergen sequences, discover the complete coding sequence of several partially described allergens and to make predictions as to the likely allergen complement of fungi. In respect of allergen occurrence in the fungi we note that many allergens such as the enolase class occur as highly homologous proteins, widely distributed in the fungal kingdom. In fact Table 2 suggests that a core set of 40 allergens may occur in many fungi with 79 of 86 allergens tested in this study being represented in this group. In fact high-throughput phage display analysis of allergens suggests the presence of at least 80 allergens in *A. fumigatus* [53]. Other fungal allergens such as Alt a 1 and Asp f 1 have no homology with any fungal genome tested and may be restricted to species or genus. Clearly not all of the

gene homologues implied in Table 2 will encode allergenic proteins however there is great potential for cross reactivity in many of the protein classes. Pragmatic studies of cross reactivity suggest that proteins with >50% identity to allergens are likely to cross-react with them [4]. In Table 2, 23/40 allergen types display >50% identity across more than one fungal species. The highly homologous proteins occur in yeasts, ascomycetes and basidiomycetes suggesting a wide distribution throughout the fungal kingdom. Thus even if many of these groups of allergen – like proteins do not cross-react there would still be many proteins capable of cross-reaction in myriad fungal species. This possibility should be urgently tested as the presence of such widespread cross-reactivity has important implications for both classical diagnostic reactivity testing using crude fungal extracts and more recent component resolved systems using recombinant proteins.

The genomic analysis has also revealed that several allergen classes exist as multi-gene families in the three *Aspergillus* genomes tested here. The levels of homology present within these gene families are also significant and imply the possibility of cross reactivity within proteins encoded by the same genome. This is significant for approaches aimed at cloning allergen genes by serum reactivity as cross-reactive gene homologues may be inadvertently cloned using these methods.

The potential cross-reactivity of the homologous protein classes can be simply tested by cloning and expressing gene candidates and testing the resulting protein against sera of known reactivity. The increasing number of fungal genomes available for analysis will allow more precise analysis of the presence of allergen homologues although better predictive methods for allergenicity are likely to be required before this information can be fully exploited.

References

- 1 The International Study of Asthma and Allergy in Childhood (ISAAC). Steering committee. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis and atopic eczema. *Lancet*; 1998; **351**: 1225–1232.
- 2 European Community Respiratory Health Survey. Variations in the prevalence of respiratory symptoms, self-reported asthma attacks and the use of asthma medications in the European community respiratory health survey (ECRHS) *Eur Respir J* 1996; **9**: 687–695.
- 3 Andersson K, Lidholm J. Characteristics and immunobiology of grass pollen allergens. *Int Arch Allergy Immunol* 2003; **130**: 87–107.
- 4 Aalberse RC. Structural biology of allergens. *J Allergy Clin Immunol* 2000; **106**: 228–238.

- 5 Edwards MR, Collins AM, Ward RL. The application of phage display in allergy research: characterization of IgE, identification of allergens and development of novel therapeutics. *Curr Pharm Biotechnol* 2001; **2**: 225–240.
- 6 Cramer R. High throughput screening: a rapid way to recombinant allergens. *Allergy* 2001; **56**(Suppl. 67): 30–34.
- 7 Cromwell O, Suck R, Kahlert H, *et al.* Transition of recombinant allergens from bench to clinical application. *Methods* 2004; **32**: 300–312.
- 8 Wallner M, Gruber P, Radauer C, *et al.* Lab scale and medium scale production of recombinant allergens in *Escherichia coli*. *Methods* 2004; **32**: 219–226.
- 9 Wu CH, Lee MF. Molecular characteristics of cockroach allergens. *Cell Mol Immunol* 2005; **2**: 177–180.
- 10 Valenta R. The future of antigen-specific immunotherapy of allergy. *Nat Rev Immunol* 2002; **2**: 446–453.
- 11 Dolen WK. Skin testing and immunoassays for allergen-specific IgE. *Clin Rev Allergy Immunol* 2001; **21**: 229–239.
- 12 Mothes N, Valenta R, Spitzauer S. Allergy testing: the role of recombinant allergens. *Clin Chem Lab Med* 2006; **44**: 125–132.
- 13 Kazemi-Shirazi L, Niederberger V, Linhart B, *et al.* Recombinant marker allergens: diagnostic gatekeepers for the treatment of allergy. *Int Arch Allergy Immunol* 2002; **127**: 259–268.
- 14 Valenta R, Lidholm J, Niederberger V, *et al.* The recombinant allergen-based concept of component-resolved diagnostics and immunotherapy (CRD and CRIT). *Clin Exp Allergy* 1999; **29**: 896–904.
- 15 Deinhofer K, Sevcik H, Balic N, *et al.* Microarrayed allergens for IgE profiling. *Methods* 2004; **32**: 249–254.
- 16 Jeong KY, Hongb CS, Yong TS. Recombinant allergens for diagnosis and immunotherapy of allergic disorders, with emphasis on cockroach allergy. *Curr Protein Pept Sci* 2006; **7**: 57–71.
- 17 Pons L, Palmer K, Burks W. Towards immunotherapy for peanut allergy. *Curr Opin Allergy Clin Immunol* 2005; **5**: 558–562.
- 18 Pons L, Burks W. Novel treatments for food allergy. *Expert Opin Investig Drugs* 2005; **14**: 829–834.
- 19 Sampson HA. Update on food allergy. *J Allergy Clin Immunol* 2004; **113**: 805–819.
- 20 Romagnani S. New therapeutic strategies in allergic diseases. *Drugs Today (Barc)* 2003; **39**: 849–865.
- 21 Nelson HS. Advances in upper airway diseases and allergen immunotherapy. *J Allergy Clin Immunol* 2006; **117**: 1047–1053.
- 22 Schmidt-Weber CB, Blaser K. Immunological mechanisms of specific allergen immunotherapy. *Inflamm Allergy Drug Targets* 2006; **5**: 15–21.
- 23 Nelson HS. Efficacy and safety of allergen immunotherapy in children. *Ann Allergy Asthma Immunol* 2006; **96**(Suppl. 1): S2–5.
- 24 Weber RW. Cross-reactivity of pollen allergens: recommendations for immunotherapy vaccines. *Curr Opin Allergy Clin Immunol* 2005; **5**: 563–569.
- 25 Mari A, Schneider P, Wally V, Breitenbach M, Simon-Nobbe B. Sensitization to fungi: epidemiology, comparative skin tests, and IgE reactivity of fungal extracts. *Clin Exp Allergy* 2003; **33**: 1429–1438.
- 26 Denning DW, O'Driscoll BR, Hogaboam CM, Bowyer P, Niven RM. The link between fungi and severe asthma: a summary of the evidence. *Eur Respir J* 2006; **27**: 615–626.
- 27 Ronning CM, Fedorova ND, Bowyer P, *et al.* Genomics of *Aspergillus fumigatus*. *Rev Iberoam Micol* 2005; **22**: 223–228.
- 28 Kurup VP. *Aspergillus* antigens: which are important? *Med Mycol* 2005; **43**(Suppl. 1): S189–196.
- 29 Rementeria A, Lopez-Molina N, Ludwig A, *et al.* Genes and molecules involved in *Aspergillus fumigatus* virulence. *Rev Iberoam Micol* 2005; **22**: 1–23.
- 30 Bille J, Marchetti O, Calandra T. Changing face of health-care associated fungal infections. *Curr Opin Infect Dis* 2005; **18**: 314–319.
- 31 Moss RB. Pathophysiology and immunology of allergic bronchopulmonary aspergillosis. *Med Mycol* 2005; **43**(Suppl. 1): S203–206.
- 32 Slavin RG, Hutcheson PS, Chauhan B, Bellone CJ. An overview of allergic bronchopulmonary aspergillosis with some new insights. *Allergy Asthma Proc* 2004; **25**: 395–399.
- 33 Calera JA, Ovejero MC, Lopez-Medrano R, *et al.* Characterization of the *Aspergillus nidulans* aspdn1 gene demonstrates that the ASPND1 antigen, which it encodes, and several *Aspergillus fumigatus* immunodominant antigens belong to the same family. *Infect Immun* 1997; **65**: 1335–1344.
- 34 Nierman WC, Pain A, Anderson MJ, *et al.* Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* 2005; **438**: 1151–1156.
- 35 Baker D, Sali A. Protein structure prediction and structural genomics. *Science* 2001; **294**: 93–96.
- 36 Mari A, Riccioli D. The Allergome Website – A database of allergenic molecules. aim, structure, and data of a web-based resource. *J Allergy Clin Immunol* 2004; **113**: S301.
- 37 Fiers MW, Kleter GA, Nijland H, *et al.* Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* 2004; **5**: 133.
- 38 Boeckmann B, Bairoch A, Apweiler R, *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003; **31**: 365–370.
- 39 Hileman RE, Silvanovich A, Goodman RE, *et al.* Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int Arch Allergy Immunol* 2002; **128**: 280–291.
- 40 Altschul SF, Madden TL, Schäffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
- 41 Kumar A, Reddy LV, Sochanik A, Kurup VP. Isolation and characterization of a recombinant heat shock protein of *Aspergillus fumigatus*. *J Allergy Clin Immunol* 1993; **91**: 1024–1030.
- 42 Kurup VP, Banerjee B. Fungal allergens and peptide epitopes. *Peptides* 2000; **21**: 589–599.
- 43 Denikus N, Orfaniotou F, Wulf G, *et al.* Fungal antigens expressed during invasive aspergillosis. *Infect Immun* 2005; **73**: 4704–4713.
- 44 Ramadan G, Davies B, Kurup VP, Keever-Taylor CA. Generation of cytotoxic T cell responses directed to human leucocyte antigen Class I restricted epitopes from the *Aspergillus* f16 allergen. *Clin Exp Immunol* 2005; **140**: 81–91.
- 45 Ramadan G, Davies B, Kurup VP, Keever-Taylor CA. Generation of Th1 T cell responses directed to a HLA Class II restricted epitope from the *Aspergillus* f16 allergen. *Clin Exp Immunol* 2005; **139**: 257–267.
- 46 Bozza S, Gaziano R, Lipford GB, *et al.* Vaccination of mice against invasive aspergillosis with recombinant *Aspergillus* proteins and CpG oligodeoxynucleotides as adjuvants. *Microbes Infect* 2002; **4**: 1281–1290.
- 47 Banerjee B, Kurup VP, Greenberger PA, Johnson BD, Fink JN. Cloning and expression of *Aspergillus fumigatus* allergen Asp f 16 mediating both humoral and cell-mediated immunity in allergic

- bronchopulmonary aspergillosis (ABPA). *Clin Exp Allergy* 2001; **31**: 761–770.
- 48 Hemmann S. Cloning, characterization and clinical evaluation of recombinant *Aspergillus fumigatus* allergens [PhD Thesis]. Zurich: University of Zurich; 1998
- 49 Shankar J, Gupta PD, Sridhara S, *et al.* Immunobiochemical analysis of cross-reactive glutathione-S-transferase allergen from different fungal sources. *Immunol Invest* 2005; **34**: 37–51.
- 50 Nigam S, Ghosh PC, Sarma PU. A new glycoprotein allergen/antigen with the protease activity from *Aspergillus fumigatus*. *Int Arch Allergy Immunol* 2003; **132**: 124–131.
- 51 Toida J, Kondoh K, Fukuzawa M, Ohnishi K, Sekiguchi J. Purification and characterization of a lipase from *Aspergillus oryzae*. *Biosci Biotechnol Biochem* 1995; **59**: 1199–1203.
- 52 Brant A, Hole A, Cannon J, Helm, *et al.* Occupational asthma caused by cellulase and lipase in the detergent industry. *Occup Environ Med* 2004; **61**: 793–795.
- 53 Rhyner C, Weichel M, Fluckiger S, *et al.* Cloning allergens via phage display. *Methods* 2004; **32**: 212–218.