Genomic Context of Azole Resistance Mutations in *Aspergillus fumigatus* Determined Using Whole-Genome Sequencing

Alireza Abdolrasouli,a,b Johanna Rhodes,c 🏚️ Mathew A. Beale,c,d Ferry Hagen,e Thomas R. Rogers,f,g Anuradha Chowdhary,h Jacques F. Meis,a,c,d Darius Armstrong-James,a Matthew C. Fisher

National Heart & Lung Institute,a Department of Medical Microbiology, Charing Cross Hospital,b and Department of Infectious Disease Epidemiology; Imperial College London, London, United Kingdom; Institute of Infection and Immunity, St. George’s University of London, London, United Kingdom; Department of Medical Microbiology and Infectious Diseases, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands; Department of Clinical Microbiology, Trinity College Dublin, Dublin, Ireland; St. James’ Hospital, Dublin, Ireland; Department of Medical Mycology, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi, India; Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, The Netherlands

A.A. and J.R. contributed equally to this article.

ABSTRACT A rapid and global emergence of azole resistance has been observed in the pathogenic fungus *Aspergillus fumigatus* over the past decade. The dominant resistance mechanism appears to be of environmental origin and involves mutations in the *cyp51A* gene, which encodes a protein targeted by triazole antifungal drugs. Whole-genome sequencing (WGS) was performed for high-resolution single-nucleotide polymorphism (SNP) analysis of 24 *A. fumigatus* isolates, including azole-resistant and susceptible clinical and environmental strains obtained from India, the Netherlands, and the United Kingdom, in order to assess the utility of WGS for characterizing the alleles causing resistance. WGS analysis confirmed that TR34/L98H (a mutation comprising a tandem repeat [TR] of 34 bases in the promoter of the *cyp51A* gene and a leucine-to-histidine change at codon 98) is the sole mechanism of azole resistance among the isolates tested in this panel of isolates. We used population genomic analysis and showed that *A. fumigatus* was panmictic, with as much genetic diversity found within a country as is found between continents. A striking exception to this was shown in India, where isolates are highly related despite being isolated from both clinical and environmental sources across >1,000 km; this broad occurrence suggests a recent selective sweep of a highly fit genotype that is associated with the TR34/L98H allele. We found that these sequenced isolates are all recombining, showing that azole-resistant alleles are segregating into diverse genetic backgrounds. Our analysis delineates the fundamental population genetic parameters that are needed to enable the use of genome-wide association studies to identify the contribution of SNP diversity to the generation and spread of azole resistance in this medically important fungus.

IMPORTANCE Resistance to azoles in the ubiquitous ascomycete fungus *A. fumigatus* was first reported from clinical isolates collected in the United States during the late 1980s. Over the last decade, an increasing number of *A. fumigatus* isolates from the clinic and from nature have been found to show resistance to azoles, suggesting that resistance is emerging through selection by the widespread usage of agricultural azole antifungal compounds. Aspergillosis is an emerging clinical problem, with high rates of treatment failures necessitating the development of new techniques for surveillance and for determining the genome-wide basis of azole resistance in *A. fumigatus*. 

Fungal infections are increasingly recognized as a threat to human (1, 2), animal, and plant (3) populations. *Aspergillus* species are ubiquitous, saprophytic fungi with airborne conidia that grow on organic matter. *Aspergillus fumigatus* is the principal causative agent of human aspergillosis, which can range from allergic syndromes to invasive aspergillosis (IA), a life-threatening infection in immunocompromised hosts. Profoundly neutropenic patients receiving chemotherapy or hematopoietic stem cell or solid organ transplantation are at high risk of IA (4). Oral triazole antifungal drugs (itraconazole, voriconazole, and posaconazole) are effective against *A. fumigatus* and remain the front-line therapy in the management and prophylaxis of IA (5). However, the emergence of azole resistance in *A. fumigatus* isolates from Europe, Asia, the Middle East, and recently, Africa (6–8) is a global and evolving public health problem (9), although surprisingly, the United States seems to be exempt from this increase in azole resistance (10). Treatment failure due to azole-resistant strains is a major clinical concern in the management of patients with IA, with high rates of patient mortality being observed (11–13). The true prevalence of azole resistance in *Aspergillus* species is largely unknown; however, frequencies of resistance varying between 0 and 6% have been determined for different clinical centers.
Azole-resistant *A. fumigatus* strains have been isolated from both azole-naïve and -exposed patients, as well as broadly from the natural environment in several countries. This has led to the hypothesis that the widespread use of structurally similar azoles in agriculture has led to increases in resistance in the clinical context (9, 15, 16).

The molecular basis of resistance to triazoles in *A. fumigatus* mainly involves point mutations at several codons in the *cyp51A* gene, which encodes lanosterol 14-α-sterol demethylase, combined with tandem repeats in the promoter region (17, 18). This enzyme is required for the biosynthesis of ergosterol, an essential component of the fungal cell membrane. Single-nucleotide polymorphisms (SNPs) in *cyp51A* cause structural alterations due to amino acid substitutions. Numerous SNPs that confer increased resistance to triazoles *in vitro* have been reported in this gene in *A. fumigatus* (19–23). However, a predominant resistance mutation, TR34/L98H, consists of a tandem repeat (TR) of 34 bases in the promoter of the *cyp51A* gene and a leucine-to-histidine change at codon 98 (24); this mutation is globally widespread in environmental isolates of the fungus. Recently, a novel *cyp51A*-mediated resistance mutation that leads to high-level voriconazole resistance, TR65/H1102F, has been described as occurring in environmental and clinical isolates of *A. fumigatus* found in the Netherlands (25), Belgium (26), Denmark (27), Germany (28), Tanzania (7), and India (29).

Typing methods, such as multilocus sequence typing (MLST), have fallen out of favor due to their low discriminatory power and considerable cost (30); the *Aspergillus fumigatus* MLST database is no longer curated, and higher-resolution methods are becoming more widely accessible for the analysis of small eukaryotic genomes due to falling costs (31). Driven by rapid technological advances, the application of whole-genome sequencing (WGS) to type microbial genomes is increasingly used to improve patient care (32–36). WGS is now positioned to become an essential primary platform for the management of antimicrobial resistance, including detection and surveillance of emerging drug-resistant microorganisms (33). While WGS is still a novel tool for the growing challenge of infectious diseases, broad applications of this technology beyond bacterial and viral infections have been swiftly embraced, and WGS has recently been used to study separate outbreaks of human fungal infections using whole-genome SNP phylogenetic analysis (37–39). Therefore, WGS is now poised to overtake other typing methods due to the high resolution of the data generated and the relatively low labor and monetary costs.

In this study, we determined the genetic background upon which *cyp51A* mutations occur by sequencing a diverse panel of clinical and environmental isolates of this fungus, and by doing so, we have demonstrated the utility of WGS to describe the evolutionary processes that underpin the emergence of azole resistance in *A. fumigatus*.

### RESULTS

#### Species identification

Twenty-two isolates were evaluated in our study (Table 1). The isolates were identified as *A. fumigatus* sensu stricto based on either sequencing of the internal transcribed spacer (ITS) region and amplification of parts of the β-tubulin and calmodulin genes (for the eight isolates from India and eight isolates from the Netherlands) or matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (for the six isolates obtained from Leeds, United Kingdom). All

<table>
<thead>
<tr>
<th>Country, city</th>
<th>Source</th>
<th>Yr of isolation</th>
<th>Collector</th>
<th>Culture reference</th>
<th>Triazole resistance</th>
<th>No. of reads aligned (millions)</th>
<th>Depth of coverage (%)</th>
<th>% of reference genome covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK, Leeds</td>
<td>Clinical</td>
<td>2012</td>
<td>R.C.B.</td>
<td>12-7505446</td>
<td>Resistant</td>
<td>31.9</td>
<td>174.1</td>
<td>94.4</td>
</tr>
<tr>
<td>UK, Leeds</td>
<td>Clinical</td>
<td>2012</td>
<td>R.C.B.</td>
<td>12-7505220</td>
<td>Susceptible</td>
<td>48.3</td>
<td>161.9</td>
<td>95.0</td>
</tr>
<tr>
<td>UK, Leeds</td>
<td>Clinical</td>
<td>2009</td>
<td>R.C.B.</td>
<td>09-7500806</td>
<td>Susceptible</td>
<td>43.6</td>
<td>146.7</td>
<td>95.2</td>
</tr>
<tr>
<td>UK, Leeds</td>
<td>Clinical</td>
<td>2012</td>
<td>R.C.B.</td>
<td>12-7504652</td>
<td>Susceptible</td>
<td>34.7</td>
<td>145.6</td>
<td>93.1</td>
</tr>
<tr>
<td>UK, Leeds</td>
<td>Clinical</td>
<td>2012</td>
<td>R.C.B.</td>
<td>12-7504462</td>
<td>Susceptible</td>
<td>47.4</td>
<td>158.8</td>
<td>95.4</td>
</tr>
<tr>
<td>UK, Leeds</td>
<td>Clinical</td>
<td>2012</td>
<td>R.C.B.</td>
<td>12-7505054</td>
<td>Susceptible</td>
<td>50.1</td>
<td>168.7</td>
<td>93.2</td>
</tr>
<tr>
<td>Netherlands, Nijmegen</td>
<td>Clinical</td>
<td>2003</td>
<td>J.F.M.</td>
<td>08-12-12-13</td>
<td>Resistant</td>
<td>34.9</td>
<td>117.5</td>
<td>93.2</td>
</tr>
<tr>
<td>Netherlands, Nijmegen</td>
<td>Clinical</td>
<td>2005</td>
<td>J.F.M.</td>
<td>08-36-03-25</td>
<td>Resistant</td>
<td>45.2</td>
<td>152.5</td>
<td>94.8</td>
</tr>
<tr>
<td>Netherlands, Nijmegen</td>
<td>Clinical</td>
<td>2004</td>
<td>J.F.M.</td>
<td>08-31-08-91</td>
<td>Resistant</td>
<td>45.8</td>
<td>153.6</td>
<td>94.0</td>
</tr>
<tr>
<td>Netherlands, Berghem</td>
<td>Environmental</td>
<td>2008</td>
<td>J.F.M.</td>
<td>08-19-02-61</td>
<td>Resistant</td>
<td>31.7</td>
<td>173.2</td>
<td>93.7</td>
</tr>
<tr>
<td>Netherlands, Berghem</td>
<td>Environmental</td>
<td>2008</td>
<td>J.F.M.</td>
<td>08-19-02-30</td>
<td>Susceptible</td>
<td>44.5</td>
<td>150.1</td>
<td>94.8</td>
</tr>
<tr>
<td>Netherlands, Nijmegen</td>
<td>Clinical</td>
<td>2010</td>
<td>J.F.M.</td>
<td>10-01-02-27</td>
<td>Resistant</td>
<td>46.5</td>
<td>155.9</td>
<td>94.5</td>
</tr>
<tr>
<td>Netherlands, Nijmegen</td>
<td>Environmental</td>
<td>2008</td>
<td>J.F.M.</td>
<td>08-19-02-46</td>
<td>Resistant</td>
<td>48.6</td>
<td>163.1</td>
<td>93.8</td>
</tr>
<tr>
<td>Netherlands, Nijmegen</td>
<td>Environmental</td>
<td>2008</td>
<td>J.F.M.</td>
<td>08-19-02-10</td>
<td>Resistant</td>
<td>51.8</td>
<td>173.6</td>
<td>93.5</td>
</tr>
<tr>
<td>India, Delhi</td>
<td>Clinical</td>
<td>2009</td>
<td>A.C.</td>
<td>AuF 942/09</td>
<td>Resistant</td>
<td>44.2</td>
<td>148.1</td>
<td>93.6</td>
</tr>
<tr>
<td>India, Delhi</td>
<td>Clinical</td>
<td>2009</td>
<td>A.C.</td>
<td>AuF 1042/09</td>
<td>Resistant</td>
<td>41.8</td>
<td>139.8</td>
<td>93.5</td>
</tr>
<tr>
<td>India, Delhi</td>
<td>Clinical</td>
<td>2011</td>
<td>A.C.</td>
<td>AuF 343/P/11</td>
<td>Resistant</td>
<td>38.4</td>
<td>128.9</td>
<td>94.8</td>
</tr>
<tr>
<td>India, Delhi</td>
<td>Clinical</td>
<td>2012</td>
<td>A.C.</td>
<td>AuF 591/12</td>
<td>Resistant</td>
<td>34.4</td>
<td>115.4</td>
<td>93.5</td>
</tr>
<tr>
<td>India, Delhi</td>
<td>Environmental</td>
<td>2011</td>
<td>A.C.</td>
<td>AuF 124/El</td>
<td>Resistant</td>
<td>50</td>
<td>167.1</td>
<td>93.6</td>
</tr>
<tr>
<td>India, Bihar</td>
<td>Environmental</td>
<td>2011</td>
<td>A.C.</td>
<td>AuF 166/E1</td>
<td>Resistant</td>
<td>42.8</td>
<td>143.6</td>
<td>93.6</td>
</tr>
<tr>
<td>India, Bihar</td>
<td>Environmental</td>
<td>2011</td>
<td>A.C.</td>
<td>AuF 257/E1</td>
<td>Resistant</td>
<td>33.8</td>
<td>113.6</td>
<td>93.3</td>
</tr>
<tr>
<td>India, Delhi</td>
<td>Environmental</td>
<td>2011</td>
<td>A.C.</td>
<td>AuF 218/E1</td>
<td>Resistant</td>
<td>46.5</td>
<td>155.8</td>
<td>93.5</td>
</tr>
<tr>
<td>UK</td>
<td>Clinical</td>
<td>1997</td>
<td>NCPF</td>
<td>AF65 (NCPF 7097)</td>
<td>Susceptible</td>
<td>57.5</td>
<td>192.8</td>
<td>95.1</td>
</tr>
<tr>
<td>UK</td>
<td>Clinical</td>
<td>1993</td>
<td>NCPF</td>
<td>AF293 (NCPF 7367)</td>
<td>Susceptible</td>
<td>49.3</td>
<td>164.5</td>
<td>98.1</td>
</tr>
</tbody>
</table>

* a) Resistance to triazole antifungal drugs was defined as an itraconazole MIC of >2 µg/liter using CLSI broth microdilution methods.

* b) The AF293 genome was the reference genome for the number of reads aligned, the corresponding depth of coverage, and the percentage of the reference genome covered by reads.
二十个鉴定和两个对照菌株（NC PF 7097/AF65）和NC PF 7367/AF293）被进一步分析使用全基因组测序（WGS）。

### Antifungal Susceptibility Testing

抗真菌敏感性测试（Table 2）显示18株菌株的MICs高于根据建立的表型切割值的伊曲康唑（≥2 mg/liter）。所有的野型型菌株显示低MICs对所有真菌。

### Sequence Analysis

所有24个基因组被映射到超过92%的de novo AF293参考基因组（覆盖100X）的全部菌株（平均152.7X）（Table 1）。原始测序读数覆盖至少93.1%的AF293参考基因组的全部菌株（Table 1）；98.1%的AF293参考基因组被测序由测序的AF293隔离。我们发现73 SNPs也在被测序的AF293隔离（数据未显示），反映由于重复亚属（40）或测序错误引起的突变。

被测序的全部基因组深度的覆盖率被观察到可能没有多变的事件（Fig. 1）。除了24个A. fumigatus菌株被序列在显示拷贝数变异（CNV）或染色体CNVs的大小的他们在多个染色体中被观察。令人惊讶的是，超过300,000 bp的序列被在所有菌株除了AF293和AF65在两个染色体1和6中，没有为丢失的129和124个基因，分别。没有为一个特定的生物学过程或功能没有被发现的基因在每一种中。然而，三个真菌内的这些在被删除的基因已经知道与azole抗真菌：Af u 1 g14330，一个ABC运输器涉及在氟康唑运输；Af u 1 g14050，一个F-box蛋白的转录是被诱导的voriconazole；和Af u 6 g10130，一个未表征的开放阅读框架（ORF）的转录是得到调节在响应voriconazole。一个较小的删除60,000 bp被观察在染色体8在所有菌株除了09-7500806, 12-7504652, 12-7504462, 和AF65; 这个删除覆盖了一个基因组的编码16个基因。除了其他的删除，没有为生物过程或功能在被删除的基因集中。

### SNP Analysis

在全部24个基因组，一个总为1,895,038 SNPs被识别，217,498个为共有的或者一个或更多菌株。所有SNPs在cy p 51A中认为一个非同义氨基酸的改变被总结在Table 2的。16个伊曲康唑（ITC）-抗性菌株显示了呈现的L98H突变。基因组的六个azole-敏感的-和-抗性A. fumigatus菌株。如在表中所示，全部24个基因组的中的17个被影响的Itraconazole MICs为≥4 mg/liter得到的L98H氨基酸的改变在增加到34-成对的核苷酸的野生型的重复在他们的促进器区域中。如已经被目前的与Itraconazole耐药性。所有的一组Itraconazole MICs为≤1 mg/liter有一个野生型的（例如，观察在参考菌株AF293）cy p 51A氨基酸的改变或者有Y 46F, V172M, T248N, E163K, D248N, S297T, K427E，好F495I。
of Afu 343/P/11, which is a clinical isolate and may have been acquired somewhere other than India (for instance through migrant working or tourism). Except for the tightly clustered isolates from India, there was no discernible relationship between the geographical location and the genotype of the cultured isolate. The phylogeny showed that clinical and environmental isolates were genotypically indistinguishable. Furthermore, phylogenetic analysis was unable to discriminate between azole-resistant and azole-susceptible strains, showing that the TR34/L98H allele occurs in diverse genetic backgrounds for both clinical and environmental isolates.

Genetic diversity. When compared against the reference isolate AF293, isolates from the Netherlands and the United Kingdom had 155,728 and 148,979 high-confidence SNPs, respectively, whereas there were 83,447 high-confidence SNPs within the Indian isolates. The average pairwise SNP diversity was higher in isolates collected from the United Kingdom (27,914) and the Netherlands (23,623) than in those from India (5,100), showing that the TR34/L98H allele occurs in diverse genetic backgrounds for both clinical and environmental isolates.

To evaluate the genetic divergence among populations, Wright’s fixation indexes ($F_{ST}$) were calculated for pairs of populations. The pairwise $F_{ST}$ value for the United Kingdom and the Netherlands was zero, implying complete panmixis and interbreeding between these two populations. This finding of intermixed diversity among these two populations is supported by phylogenetic analysis (Fig. 2).

Pairwise $F_{ST}$ values between Indian, United Kingdom, and Netherlands populations showed that there was a pronounced phylogeographic structure at the continental level, with an $F_{ST}$ value of 0.314 for India versus the Netherlands and an $F_{ST}$ value of 0.381 for India versus the United Kingdom.

Recombination analysis. Population-wide recombination rates were estimated for each population (defined as a single country) using the program LDhat (42), which identifies patterns of linkage disequilibrium using Hudson’s composite likelihood method. The interval method used in this study specifically estimates a variable recombination rate using a Bayesian reversible-jump Markov chain Monte Carlo method under a crossing-over model. Evidence for the frequency of recombination in each population is presented in Fig. 3; the greater numbers of recombination hot spots in the United Kingdom and the Netherlands data...
reflect the high genetic diversity (Fig. 3). Given the lower genetic diversity observed in India (in terms of SNP numbers and also as seen in the phylogenetic analysis in Fig. 2), we anticipated this data set to be highly clonal, but on the contrary, the LDhat analysis (Fig. 3) detected similar numbers of recombination hot spots for the United Kingdom and the Netherlands, implying that the Indian data set is derived from isolates that are undergoing sexual or parasexual recombination. Overall, the rate of recombination in the Indian population ($\rho = \text{0.0726}/\text{bp}^{-1}$) was significantly higher (calculated using a t test) than the recombination rates observed in United Kingdom and Netherlands populations ($\rho = \text{0.0070}/\text{bp}^{-1}$ and $\rho = \text{0.0104}/\text{bp}^{-1}$, respectively).

**FIG 2** Phylogenetic analysis of *A. fumigatus* isolates representing azole-resistant and -susceptible genotypes from India, the Netherlands, and the United Kingdom. Bootstrap analysis was performed on WGST SNP data from 24 *A. fumigatus* genomes to generate an unrooted maximum-likelihood phylogeny, with all branches supported to 87% or higher. Isolates are color coded according to country of origin (red, India; green, the Netherlands; blue, United Kingdom), and environmental isolates are indicated by a tree symbol. Isolates marked with a black circle contain the TR34/L98H mutation in the cyp51A gene. Branch lengths represent the numbers of SNPs between taxa.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Afu 942/09</th>
<th>Afu 1042/09</th>
<th>Afu 343/P/11</th>
<th>Afu 591/12</th>
<th>Afu 124/E11</th>
<th>Afu 166/E11</th>
<th>Afu 257/E11</th>
<th>Afu 218/E11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afu 942/09</td>
<td>2207</td>
<td>21,150</td>
<td>3067</td>
<td>1954</td>
<td>2406</td>
<td>2928</td>
<td>2353</td>
<td></td>
</tr>
<tr>
<td>Afu 1042/09</td>
<td>2233</td>
<td>21,261</td>
<td>3232</td>
<td>2148</td>
<td>2625</td>
<td>3100</td>
<td>2617</td>
<td></td>
</tr>
<tr>
<td>Afu 343/P/11</td>
<td>5136</td>
<td>5221</td>
<td>5479</td>
<td>4861</td>
<td>5017</td>
<td>5491</td>
<td>4941</td>
<td></td>
</tr>
<tr>
<td>Afu 591/12</td>
<td>1943</td>
<td>2082</td>
<td>20,369</td>
<td>1484</td>
<td>1704</td>
<td>2401</td>
<td>1615</td>
<td></td>
</tr>
<tr>
<td>Afu 124/E11</td>
<td>2564</td>
<td>2732</td>
<td>21,485</td>
<td>3218</td>
<td>2383</td>
<td>3359</td>
<td>2331</td>
<td></td>
</tr>
<tr>
<td>Afu 166/E11</td>
<td>2371</td>
<td>2564</td>
<td>20,996</td>
<td>1738</td>
<td>2950</td>
<td>2950</td>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>Afu 257/E11</td>
<td>1645</td>
<td>1791</td>
<td>20,222</td>
<td>2242</td>
<td>1466</td>
<td>1702</td>
<td>1623</td>
<td></td>
</tr>
<tr>
<td>Afu 218/E11</td>
<td>2539</td>
<td>2777</td>
<td>21,141</td>
<td>2925</td>
<td>1907</td>
<td>2212</td>
<td>3092</td>
<td></td>
</tr>
</tbody>
</table>
In order to refute the existence of clonality in the Indian data set, we estimated the index of association (IA) and the modified statistic rBarD (73) to confirm the presence of recombination within each population of this data set. We assumed that the null hypothesis of no linkage disequilibrium and, therefore, no recombination would not be rejected if the resulting values of both statistics were not significantly different from the distribution of values obtained from 999 resamplings. For all three populations (India, United Kingdom, and the Netherlands), the null hypothesis could not be rejected, implying no significant linkage among loci, and we concluded, therefore, that all populations are recombining.

PCR and mining of de novo assembly of sequence reads has shown that both MAT1-1 and MAT1-2 loci are present in all populations investigated here (Table 4). While a chi-square test showed no significant difference in the frequency of either locus in the United Kingdom and Netherlands populations, a bias was observed in the Indian population, which contained significantly fewer MAT1-2 idiomorphs (P < 0.0304).

<table>
<thead>
<tr>
<th>Country</th>
<th>% (no. of isolates) with mating type</th>
<th>Chi-square value</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>66.6 (4) 33.4 (2)</td>
<td>0.41</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>62.5 (5) 37.5 (3)</td>
<td>0.47</td>
</tr>
<tr>
<td>India</td>
<td>87.5 (7) 12.5 (1)</td>
<td>0.034</td>
</tr>
</tbody>
</table>

**DISCUSSION**

To investigate the role whole-genome sequencing can play in the detection of azole resistance in *A. fumigatus*, we sequenced the genomes of 24 clinical and environmental isolates obtained from India, the Netherlands, and the United Kingdom. In the present study, we aimed to verify the applicability of using WGS to characterize known mutations in the cyp51A locus and to determine the genome-wide patterns of genetic diversity upon which these mutations reside.

We were able to characterize the genotype of cyp51A for all isolates in the study with high confidence. Of these, sixteen isolates with itraconazole MICs of ≥4 mg/liter had the L98H amino acid substitution accompanied by a characteristic 34-base-pair nucleotide tandem repeat in their promoter region. We did not detect any amino acid substitutions associated with resistance in cyp51A genes in the strains with itraconazole MICs of ≤1 mg/liter. Several mutations specific to the cyp51A gene, such as substitutions at codons G54, G138, P216, F219, M220, and G448, have also been described as responsible for resistance to azoles in clinical isolates of *A. fumigatus* (14–18), though none of these mutations were present in our analyzed sample collection. Interestingly, and in contrast with previous observations where all TR34/L98H mutations are accompanied by a pair of amino acid substitutions, S297T and F495I (44), only two of our 17 isolates with TR34/L98H showed S297T and F495I substitutions. This suggests that intragenic recombination in cyp51A may have occurred to unlink these mutations in this set of isolates. Whether or not this means that there have been multiple origins of the TR34/L98H allele will re-

![FIG 3 Recombination analysis, using LDhat interval (42), of *A. fumigatus* isolates from the United Kingdom, the Netherlands, and India. The black peaks represent the recombination rate across the whole *A. fumigatus* genome, and the vertical red lines mark the chromosome boundaries.](image-url)
quire broader WGS surveys; however, the observation that this pair of mutations has never been found unlinked to date suggests that the probability of *de novo* origination for this allele is low to nonexistent. In support of this interpretation, it is worth noting that a novel environmental *cyp51A*-mediated resistance mechanism (TR<sub>46</sub>/Y121F/T289A) was recently reported in the Netherlands (25), Belgium (26), Denmark (27), Germany (28), Tanzania (7), and India (29), superimposed on a more homogenous genetic background than TR<sub>34</sub>/L98H. This recently detected resistance mechanism has been observed to rapidly increase in frequency in Dutch hospitals and homes, where it is associated with voriconazole therapy failure (6, 25). Therefore, it appears that the contemporary rise in the frequency of TR<sub>46</sub>/L98H is not a unique event and that the emergence of *A. fumigatus* isolates in nature showing *cyp51A*-associated resistance alleles is an ongoing phenomenon.

Our demonstration of high levels of SNP diversity and a panmictic population structure within the United Kingdom and Netherlands is testament to the ability of this highly aerosolized (airborne) fungus to passively disperse. Our findings suggest that the TR<sub>34</sub>/L98H allele has existed long enough for the twin forces of (airborne) fungus to passively disperse. Our findings suggest that mictic population structure within the United Kingdom and Netherlands is nonexistent. In support of this interpretation, it is worth noting that we sequenced were collected from locations separated by over 1,000 km. This suggests that there has been a recent expansion of this genotype within India and it is likely not a coincidence that these genotypes all contain the TR<sub>34</sub>/L98H allele. To further investigate the possibility that this allele has been introduced into the Indian genetic background recently, further sampling and sequencing of both azole-resistant and azole-sensitive isolates from this region are needed to clarify our finding.

There have been significant advances recently in our understanding of azole resistance mechanisms in other pathogenic fungi, such as *Candida albicans* (45) and *Cryptococcus neoformans* (46). In both species, aneuploidy for the chromosome bearing the *ERG11* gene has been reported and found to be integral to antifungal resistance. We did not detect any chromosomal aneuploidy or copy number variation for *cyp51A* or *cyp51B*, the ortholog of *ERG11*, in any of the *A. fumigatus* isolates investigated here. However, we did detect large deletions in chromosomes 1, 6, and 8 in many of the isolates (Fig. 1). Analyzing the gene content of the deleted regions revealed two genes in chromosome 1 (Afu1g14330, an ABC transporter, and Afu1g14050, encoding an F-box protein) and one gene in chromosome 6 (Afu6g10130, uncharacterized but downregulated in response to voriconazole) that had a known function associated with azole antifungals. However, given the large-scale nature of the deletions (both over 300,000 bp in size), we do not believe that these deletions are part of an evolutionary process that is enabling *A. fumigatus* to evolve resistance to azole antifungals. Indeed, we see no deletions specific to either azole-sensitive or azole-resistant isolates, suggesting that these deletions occurred in response to a different process. Deletions on this scale have not previously been reported in *A. fumigatus*, representing an opportunity for further investigation.

While most mutations in azole-resistant *A. fumigatus* isolates were single nucleotide substitutions in the target gene (*cyp51A*), mutations at other genes have been reported as conferring resistance to azoles in this fungus. Furthermore, other resistance mechanisms could be responsible for azole resistance in *A. fumigatus*. The non-*cyp51A*-mediated resistance can potentially be explained by alternative resistance mechanisms that have been described previously, including (i) higher basal expression of the *cdr1B* efflux transporter (47), (ii) a mutation in the CCAAT-binding transcription factor complex subunit *HapE* (48), which was not seen in our data, and (iii) high *cyp51B* expression (49). A recent publication reported drug resistance due to RNA interference (RNAi)-dependent epimutations in *Mucor circinelloides* (50); we found five orthologs of genes involved in the RNAi pathway in *Aspergillus nidulans*, *Cryptococcus neoformans*, and *Schizosaccharomyces pombe* present in *A. fumigatus* isolate 09-750806. These five genes (Afu3g11010, Afu8g05280, Afu5g11790, Afu4g02930, and Afu5g11440) all had protein domains associated with RNA silencing: PIWI and PAZ domains were present in both Afu3g11010 and Afu8g05280, Dicer domains were present in both Afu5g11790 and Afu4g02930, and a domain for Argonaute small interfering RNA (siRNA) chaperone complex subunit Arb1 was present in Afu5g11440. Transcriptome profiling has revealed that all five genes are expressed in *A. fumigatus* A1163 (51). While these five genes do not represent the full RNAi pathway, we believe that epigenetic regulation could be involved in mediating antifungal drug resistance within *A. fumigatus*.

The limited set of mutations in *cyp51A* conferring resistance makes them relatively simple targets for molecular detection because their roles in resistance are known *a priori*. However, the evolution of novel azole resistance mechanisms, such as has been seen for TR<sub>46</sub>/Y121F/T289A and *cdr1B*, makes the detection of these “known unknowns” much more challenging. As our method captures the entirety of SNP diversity for each isolate of *A. fumigatus*, it is theoretically possible to determine the contribution of each nucleotide in the genome to the generation of the resistance phenotype within the framework of a genome-wide association study (GWAS). This “hypothesis-free” approach will ultimately extend the WGS-enabled description of known mutations, such as those reported here, to describe the full suite of drug resistance alleles and their epistatic interactions that occur in *A. fumigatus*. However, before this ambition is realized, significant hurdles need to be addressed. Most fundamentally, a description of the genome-wide global population structure of *A. fumigatus* needs to be undertaken in order to delineate the essential population parameters that underpin a GWAS; namely, the extent of population genetic structure and genome-wide rates of linkage disequilibrium. From the limited set of isolates that we sequenced, it is already evident that genome-wide recombination occurs and that SNPs are shared across the United Kingdom, the Netherlands, and India, as well as within these countries. This is an encouraging finding, especially as mixed-model GWAS frameworks are now being used and shown to perform well when handling structured populations by estimating the phenotypic covariance due to genetic relatedness (43). Although GWAS has not yet been broadly applied to microbial populations, a recent study by Sheppard et al. (53) illustrated the requirement for novel techniques in order to take these population genetic variables into account, and similar
approaches are likely amenable to analyzing fungal genome-wide SNP datasets such as we have described here.

We have shown the population of *A. fumigatus* to be broadly panmictic and recombinogenic, a phenomenon that has been previously observed and confirmed by independent groups (54–57). The process of recombination is expected to accelerate adaptation to novel environmental conditions, as has previously been described in the related nonpathogenic fungus *Aspergillus nidulans* (52). Furthermore, while there were fewer recombination hot spots observed within the Indian population, the genome-wide population recombination rate was higher than that seen in the United Kingdom and the Netherlands. Seven of eight Indian isolates were typed as being MAT1-1, with the chi-square test showing that this bias is statistically significant. Together, these findings provide tentative evidence that an outcrossing event may have resulted in the origin of the Indian genotypic cluster, and it is worth speculating that the resulting MAT1-1 TR34/L98 progeny are undergoing a rapid selective sweep by virtue of being highly fit. Clustered sexually recombining neighborhoods have previously been described in a related pathogenic fungus, *Talaromyces (Penticillium) marneffei* (58), suggesting that this type of fungal population structure may be more common than previously recognized. Whether our findings reflect the process of natural selection in *A. fumigatus* populations in response to the widespread use of agricultural azoles in Europe and India remains to be seen; however, the genomic approaches that we have detailed here will clearly underpin future research to decipher the population structure of *A. fumigatus* and map the future evolutionary trajectory of clinically relevant phenotypes.

### MATERIALS AND METHODS

**Isolates.** Twenty-four *A. fumigatus* isolates representing three geographical locations were included in the analysis (Table 1). Six clinical isolates, 12-7505446, 12-7505220, 09-7500806, 12-7504652, 12-7504462, and 12-7505054, were obtained from the Mycology Reference Centre, Leeds Teaching Hospitals National Health Service Trust, Leeds, United Kingdom. The eight isolates from the Netherlands included four clinical isolates (08-12-12-13, 08-36-03-25, 08-31-08-91, and 10-01-02-27) and four environmental soil isolates (08-19-02-61, 08-19-02-30, 08-19-02-46, and 08-19-02-10). Of the eight Indian isolates included in this study, four were clinical (Afu 942/09, Afu 1042/09, Afu 343/P/11, and Afu 591/12) and the remaining four (Afu 124/E11, Afu 166/E11, Afu 257/E11, and Afu 218/E11) originated from environmental soil sources. Additionally, the following two control isolates were included for comparison. (i) AF293 (NCBI 7367; National Collection of Pathogenic Fungi at Mycology Reference Laboratory, Public Health England, Bristol, United Kingdom) is a clinical isolate initially cultured in 1993 from lung tissue of a neutropenic patient with invasive aspergillosis in the United Kingdom; the whole-genome sequence of this strain was published in 2005 by Nierman et al. (59). (ii) AF65 is a clinical isolate originally cultured from a lung biopsy specimen from a patient with acute leukemia; this isolate was deposited at the European Nucleotide Archive under project accession number NCBI 7097 (60).

**Aspergillus fumigatus identification** All 24 isolates were identified by macroscopic and microscopic morphological characteristics as *A. fumigatus* species complex; growth at 50°C differentiated *A. fumigatus* from *Aspergillus lentulus*. Six strains from Leeds, United Kingdom, were further identified by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS), using a Bruker Daltonics Biosystem (Bruker, Bremen, Germany). All of the *A. fumigatus* isolates obtained from the Netherlands (*n* = 8) and India (*n* = 8) were previously described as *A. fumigatus* by sequencing of the internal transcribed spacer (ITS) region. In addition, the presence of any cryptic species within *Aspergillus section Fumigati* among these isolates was previously excluded by amplification of parts of the β-tubulin and calmodulin genes (61).

**Antifungal susceptibility testing.** The *in vitro* antifungal susceptibility profiles of triazole antifungal drugs, determined according to the CLSI M38-A2 broth microdilution method (62), were previously described and published for isolates from India and the Netherlands (61). MICs were interpreted based on proposed breakpoints (63). For six isolates, the MICs of itraconazole were primarily determined at Mycology Reference Centre, Leeds, United Kingdom, using E test strips (bioMérieux, Marcy l’Etoile, France) on RPMI 1640 agar supplemented with 2% glucose, according to the manufacturer’s instructions, and the results were interpreted as described by Pfaffer et al. (64). To ensure consistency of results, MICs for all 24 isolates were further confirmed using the CLSI M38-A2 broth microdilution method.

**Conidial harvest.** All fungal isolates were subculured on potato dextrose agar (PDA) plates and incubated at 35°C for 5 days until sporulation. Stock conidial suspensions were prepared by washing the surface of the PDA plates with 10 ml of sterile water containing 0.05% Tween 20. The conidial suspensions were filtered using Miracloth (EMD Chemicals, San Diego, CA, United States) to remove fungal hyphae, transferred to 50-ml sterile conical tubes, and centrifuged at maximum speed (10,000 × g) for 10 min. The supernatants were discarded, and the pellets were resuspended in 5 ml of sterile distilled water. The concentrations of the suspended conidial stocks were determined by counting the conidia using a hemocytometer chamber at ×400 magnification. Harvested conidia at concentrations of 2 × 10⁷/ml were subjected to DNA extraction.

**DNA extraction and quality assessment.** High-molecular-weight DNA was extracted with an optimized MasterPure yeast DNA purification kit (Epicentre Biotechnologies, Cambridge, United Kingdom) with an additional bead beating step included. Harvested conidia were homogenized using 1.0-mm-diameter zirconia/silica beads (BioSpoc Products, Bartlesville, OK) in a FastPrep-24 system (MP Biomedicals, Solon, OH) at 4.5 m/s for 45 s. Genomic DNA (gDNA) was quantified using a Qubit 2.0 fluorometer and dsDNA BR (double-stranded DNA, broad-range) assay kit (Life Technologies, Carlsbad, CA). Quality control of extracted gDNA samples prior to library preparation was performed using the TapeStation 2200 system (Agilent, Santa Clara, CA) and gDNA ScreenTape assays. Purified gDNAs were stored at −20°C until further use.

**Mutation screening.** Isolates were screened for the presence of a tandem repeat insertion in the promoter region of the *cyp51A* gene, as well as for the presence of the common mutations (e.g., L98H, G54, and M220), by using a mixed-format real-time PCR as previously described (10).

**Library preparation.** Genomic DNA libraries were constructed according to protocols provided by Illumina (TruSeq Nano DNA Sample Preparation Guide) with the TruSeq Nano kit (Illumina, San Diego, CA). Briefly, gDNA was sheared into 350-base-pair fragments using an S220 ultrasonicator (Covaris, Woburn, MA) and AFA fiber Snap-Cap microTubes, followed by end repair and solid-phase reversible immobilization (SPRI) bead-based size selection. The final libraries were quality controlled on the TapeStation 2200 system (Agilent) with D1K ScreenTape assays (Agilent) and quantified with quantitative PCR (qPCR) on an Applied Biosystems 7300 instrument (Life Technologies) using the Kapa library quantification kit (Kapa Biosciences, Boston, MA). DNA libraries were normalized to 10 nM, and randomized indexed pools were pooled into three libraries of 8 samples.

**Illumina whole-genome sequencing.** Prepared whole-genome libraries (*n* = 24) were sequenced on three lanes of the same flow cell using a HiSeq 2500 sequencer (Illumina) at Medical Research Council Clinical Genomics Centre, Imperial College London, Hammersmith, United Kingdom, generating 100-bp paired-end reads in high-output mode. All raw reads and relevant information in this study have been submitted to the European Nucleotide Archive under project accession number PRJEB8623.

**Alignment.** Raw Illumina WGS reads were quality checked using FastQC (version 0.10.1; Babraham Institute) and aligned against the

---

Abdolrasouli et al.
AF293 reference genome (59) using the short-read alignment component (aln) of the Burrows-Wheeler Aligner (BWA) alignment tool (65) with a quality threshold of 15. Duplicate reads were masked using Picard (version 1.112).

Identification of SNPs. Single-nucleotide polymorphism (SNP) detection was conducted using UnifiedGenotyper from the Genome Analysis Toolkit (GATK) package (66, 67). To ensure high confidence in the SNPs called, variants were filtered according to whether they were present in 80% of reads, the mapping quality, and the depth of coverage at each base to provide a list of high-confidence variants, as described in Rhodes et al. (31). SNPs in cyp51A were mapped using VCF-annotator (Broad Institute, Cambridge, MA).

Identification of tandem repeat sequences. Whole-genome sequence reads for each isolate were used as the input for de novo assembly using SPAdes 3.1.1 (68). The resulting fasta file was mined for the tandem repeat sequences.

Phylogenetic analysis. Whole-genome SNP data were converted into relaxed interleaved Phytip format. Maximum-likelihood trees were constructed using RAxML (41) and visualized in FigTree version 1.4.0. The rapid bootstrap algorithm was used to bootstrap all phylogenies using 100 replicates.

Analysis of gene diversity. Three groups, comprising 6 isolates from the United Kingdom, 8 isolates from the Netherlands, and 8 isolates from India, were formed for the purpose of analyzing genetic diversity. The fixation statistic FST was calculated between each population pair using the package BEDASSLE (69) for R (70).

Recombination analysis. Two statistics commonly used for describing linkage disequilibrium, the index of association (IA) and rBarD, were recommended in the user manual (42). The resulting fasta file was mined for the tandem repeat sequences.

Analysis of gene diversity. The program was executed for 2 million iterations with samples from 8 chromosomes of A. fumigatus, partitioned by their country of original isolation. The program was executed for 2 million iterations with sampling every 200 iterations after a 20,000-iteration burn-in period (as recommended in the user manual [http://ldhat.sourceforge.net/manual.pdf]). The output was summarized using LDhat stat.

ACKNOWLEDGMENTS

J.R. and M.A.B. were supported by a Medical Research Council grant (MRC MR/K000373/1) awarded to M.C.F. This study was supported by a grant from Gilead (WMNF P47129) to D.A.-J., The Royal Brompton and Harefield Respiratory Biomedical Research Unit, and the Imperial College London Healthcare Biomedical Research Centre Infectious Diseases Stream.

We thank Elizabeth M. Johnson and Adrien Szekely (National Collection of Pathogenic Fungi and Mycology Reference Laboratory, Public Health England, Bristol, United Kingdom) and Richard C. Barton (Mycology Reference Centre, Leeds Teaching Hospitals National Health Service Trust, Leeds, United Kingdom), who submitted isolates for this study.

J.F.M. has received grants from Astellas, Basilea and Merck. He has been a consultant to Astellas, Basilea, and Merck and has received speaker’s fees from Merck and Gilead. D.A.-J. has received grants from Gilead and Pfizer and is a consultant to Pulmocide.

REFERENCES


