

Aspergillus strain typing in the genomics era

C.H.W. Klaassen^{1*} and N. Osherov²

¹Department of Medical Microbiology and Infectious Diseases, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands; ²Department of Human Microbiology, Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

*Correspondence: Corné H.W. Klaassen, c.klaassen@cwz.nl

Abstract: Multiple reasons may justify a need for strain typing purposes, but the most common reason is to delineate the epidemiological relationships between isolates. The availability of whole genome sequences has greatly influenced our ability to develop highly targeted and efficient strain typing methods for these purposes. Some strain typing methods may serve dual goals: not only can they be used to discriminate between multiple isolates of a certain species, they can also aid in the recognition, identification, description and validation process of a fungal species.

Key words: AFLP, coding tandem repeats, high resolution typing, identification pathogenic aspergilli, MLST.

INTRODUCTION

Strain typing can fulfill many needs both in clinical settings and otherwise. Among the many potential applications for strain typing are outbreak analysis and environmental monitoring, patient monitoring and treatment follow-up, local and global epidemiology, database construction, strain identification (e.g. with production organisms) and many more. Apart from these applications at the subspecies level, molecular methods are also increasingly used at the genus level for the definition and recognition of fungal species.

Over the years, many different molecular methods have been developed for *Aspergillus* strain typing. Because of its clinical significance, these methods were primarily directed at *A. fumigatus*. The most promising techniques are either PCR based, such as analysis of microsatellite length polymorphisms (MLP)/ short tandem repeats (STR) (Bart-Delabesse *et al.* 1998; de Valk *et al.* 2005) and amplified fragment length polymorphism (AFLP) analysis (Warris *et al.* 2003; de Valk *et al.* 2007b), or based on non-coding repetitive sequences (such as the *Afut1* element) in combination with restriction fragment length polymorphisms (RFLP) (Girardin *et al.* 1993). Use of these and other methods has been reviewed by Varga (2006). Three recent additions to this diverse list are multilocus sequence typing (MLST) (Bain *et al.* 2007), coding tandem repeats (Balajee *et al.* 2007; Levdansky *et al.* 2007) and retrotransposon insertion-site context (RISC) typing (de Ruiter *et al.* 2007). Depending on the exact reason for strain typing and on the technical resources in a particular setting, the choice for either of these methods could be appropriate.

Classically, without the availability of genomic sequence information, the process of developing a new strain typing method often involved many laborious selection and optimisation experiments. At present, in the genomics era, the availability of whole genome sequences has had a great impact on our options to develop novel and state-of-the-art fingerprinting methods. We now can develop new fingerprinting methods using highly targeted approaches with much higher *à priori* chances of being successful than before. Naturally, as more genomic sequence information is becoming available, these chances will continue to increase.

Here, we will present a number of applications for several of these genotyping methods and discuss the impact of the availability of genomic sequence data on the applications of these methods.

High resolution exact strain typing using short tandem repeats

Microsatellites or STR"s are ubiquitously present in the genomes of many fungi including *Aspergillus spp.* Microsatellites, as tools for the identification of and discrimination between individual organisms, already have a relatively long history in human forensic applications where they currently comprise the global "gold standard" in the identification process of individuals. The use of STR"s offers a number of technical advantages over many other fingerprinting techniques including: ease of amplification, multiplex options, extremely high discriminatory power, an exact unambiguous (numerical) and highly portable and exchangeable typing result, ability to detect mixed samples, construction of databases, etc. Because of these advantages, there is a growing interest in the use of STR based methods for strain typing in the microbial field as well.

Bart-Delabesse *et al.* (1998) reported the first application of microsatellites for *A. fumigatus*. These markers were obtained by screening genomic DNA libraries of *A. fumigatus* for suitable, microsatellite containing sequences, a process that proved quite laborious in the pre-genomics era. A panel of 4 dinucleotide repeats was selected that performed well in comparative genotyping experiments (Lasker 2002). Recently, based on genomic sequence data that has become available, de Valk *et al.* (2005) reported a novel set of 9 tandem repeats for typing *A. fumigatus* isolates, the so-called STRAf assay (STR"s of *A. fumigatus*). In contrast to the previously developed typing scheme, this panel also contained tri- and tetranucleotide repeat markers and, in addition, all loci contained a single uninterrupted repeat element. By using multicolor multiplex approaches with these novel markers, large numbers of isolates can be analyzed in a short period of time. Because of the larger number of loci, the STRAf assay yielded a superior discriminatory power for typing *A. fumigatus* isolates. In Fig. 1, a graphical representation of

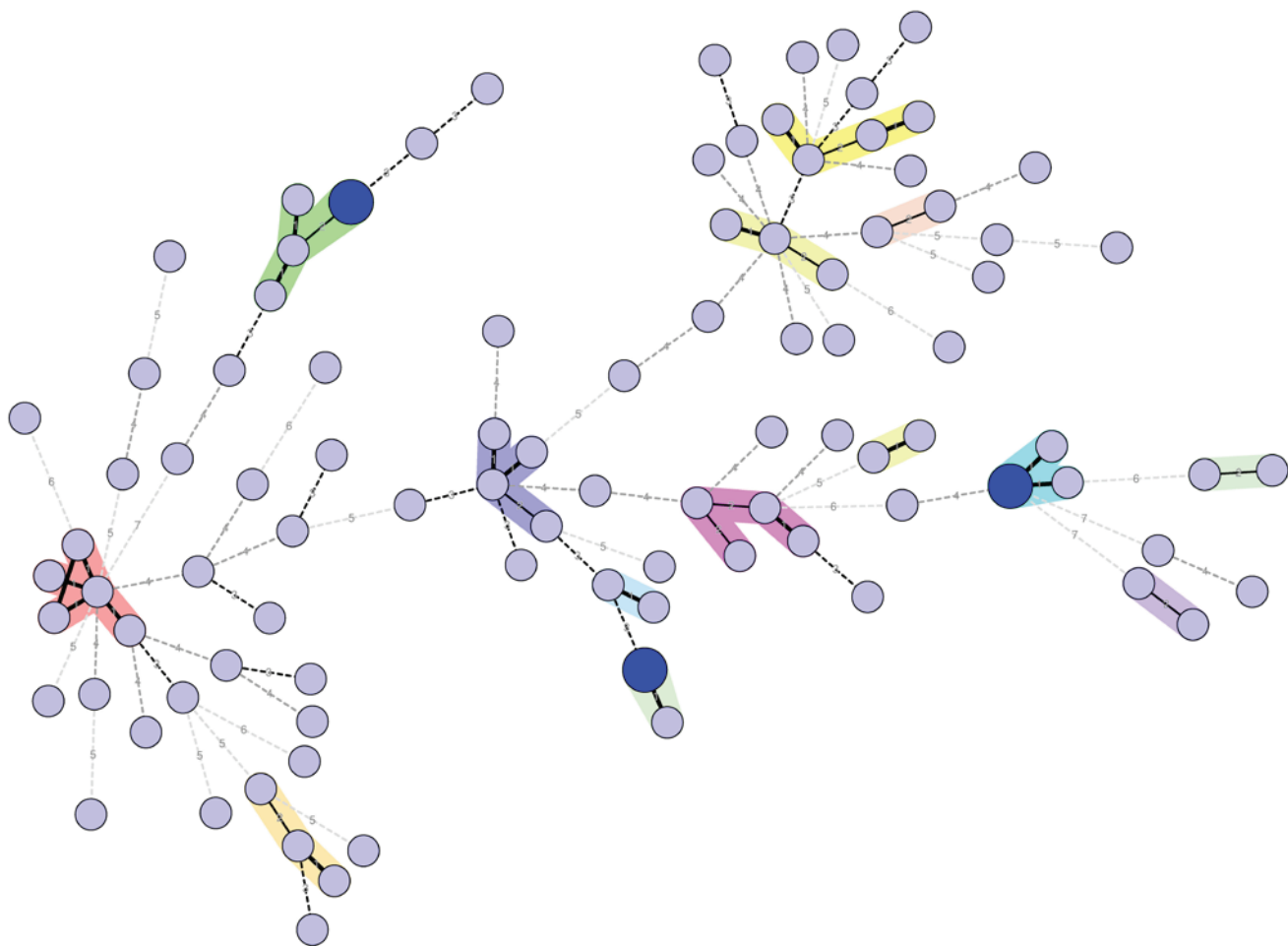


Fig. 1. Minimal spanning tree of 99 presumably unrelated *A. fumigatus* isolates based on microsatellite data. The tree was generated using the multi-state categorical similarity coefficient. Each circle represents a unique genotype. The size of the circle corresponds to the number of isolates with the same genotype. Genotypes connected through a shaded background differ by a maximum of 2 out of 9 markers. The numbers correspond to the number of different markers between the genotypes. No less than 96 different genotypes were discriminated. Data are taken from de Valk *et al.* (2005).

the diversity with the *A. fumigatus* population is shown. The minimal spanning tree represents 99 presumably unrelated *A. fumigatus* isolates. Almost all isolates could be discriminated from each other. The ones that could not be discriminated by the STRAf assay also proved to be indistinguishable using other molecular methods such as AFLP analysis. Furthermore, the STRAf assay proved to be an extremely robust typing assay. It has been shown that deliberate and significant changes to the experimental protocol did not lead to wrong typing results with this assay (de Valk *et al.* 2007a).

The key element in the use of microsatellites is to translate the electrophoretic mobility of the obtained fragment (reflected as the size of the fragment in bp and obtained on a high resolution electrophoretic platform) to the corresponding number of repeats. Unfortunately, this mobility is dependent on many factors such as the presence/absence of denaturing compounds, the sieving matrix that is used, the exact base composition and sequence of the fragment, run temperature, presence of different fluorescent labels and even something that may appear only trivial such as the sizing marker (de Valk *et al.* 2007a; Tu *et al.* 1998; Vainer *et al.* 1997). In order to transfer a microsatellite based assay to a different electrophoresis platform, a careful calibration of the new platform has to be established. Similar to the situation in human forensics, a series of allelic ladder was constructed that contain reference fragments with established repeat numbers. By running these allelic ladders, every platform can be calibrated to yield interchangeable typing data with any given set of isolates (de Valk

et al. submitted for publication). Thus, the STRAf assay has all the key ingredients to be successfully used for global standardisation of *A. fumigatus* typing.

Simultaneous identification and strain typing

In recent years, there has been a growing interest in the use of more accessible techniques such as MLST approaches for fungal identification purposes and strain typing. This approach that is exclusively based on sequencing data has the advantage of the development of accurate databases totally reliable for taxonomy. However, whereas MLST performs well at the genus and species level, in the case of *Aspergillus* (and in contrast to other species like *Candida*) the discriminatory power at the subspecies level turns out to be disappointing (Bain *et al.* 2007).

AFLP analysis is a highly discriminatory method at the intraspecies level. In AFLP analyses, fragments are amplified from random locations throughout an organisms' genome in a highly reproducible manner (Vos *et al.* 1995). The discriminatory power of AFLP analysis equals that of the STR panels (de Valk *et al.* 2007b) and *Afut1* RFLP analysis. However, like with any other fingerprinting method based on DNA banding patterns, its long-term stability and reproducibility may be quite challenging. Development of AFLP fingerprinting requires no prior sequence information. However, depending on the genome composition (GC-content and distribution, presence of multicopy elements), certain combinations

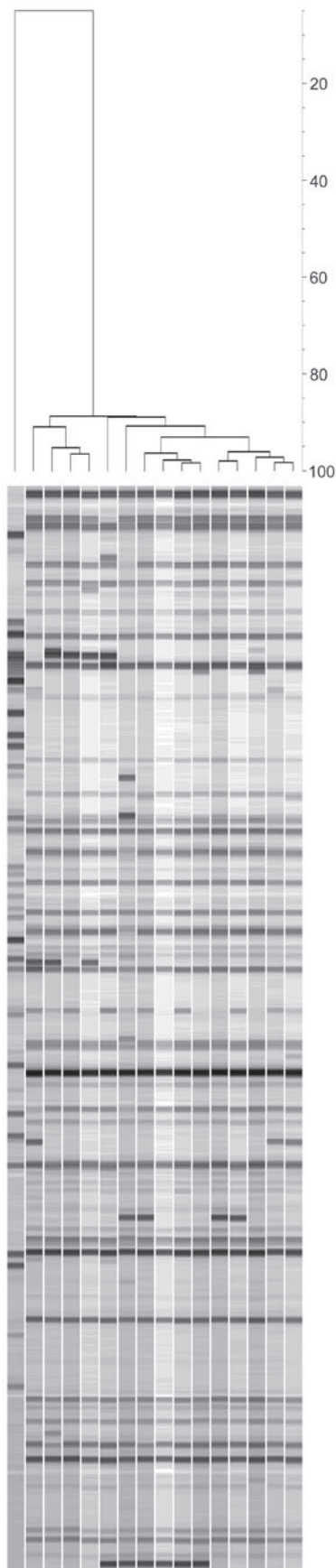


Fig. 2. Example of AFLP analysis showing the discriminatory power as well as the ability to discriminate between isolates belonging to different species. The figure shows 16 fingerprints from *A. fumigatus* cultured from 16 IA patients (1 isolate per patient). Based on the differential presence or absence of one or more bands, all isolates can be discriminated from each other. One isolate with a clearly different fingerprint turned out to be *N. fischeri*. The dendrogram was calculated by UPGMA clustering using the Pearson correlation coefficient. The scale bar indicates the percentage similarity.

of restriction enzymes and selective residues could prove to be more suitable than others. At present, based on available genomic sequence data, one can predict *in silico* which fragments will be obtained with any known genome (Bikandi *et al.* 2004).

Whereas AFLP has originally been presented as a tool for strain typing purposes, it is also very well suited to simultaneously resolve isolates belonging to different species from each other. This is relevant in the case where the identification of a microorganism may be uncertain such as the species from the morphologically very similar section *Fumigati*. If not properly identified, use of such typing data could easily lead to false conclusions. In a way, AFLP can be considered the perfect PCR alternative to DNA-DNA reassociation studies. Classical DNA-DNA reassociation studies rely on sequence similarities. If two species share a certain amount of sequence information, it is to be expected that they will also share a certain amount of similarity in banding patterns from an AFLP fingerprint. In fact, this has already been demonstrated for a variety of bacterial and fungal species. According to our own observations, AFLP fingerprints of isolates of the same species are usually > 60 % similar whereas fingerprints for isolates representing different species are usually < 40 % similar (Fig. 2). We also used AFLP analysis for confirmation of the identity of 67 isolates representing 26 species in *Aspergillus* section *Fumigati*. These isolates have previously been identified using a variety of other methods. Although the majority of isolates were correctly identified, this exercise clearly showed that: *i* some isolates were misidentified, *ii* that some recognised species are comprised of multiple clearly discernible subgroups and *iii* that several isolates that are currently recognised as different species should actually be grouped into a single species. Thus, AFLP can very well complement the use of MLST and other methods in the recognition and validation process of fungal species.

Strain typing based on coding tandem repeats

Coding tandem repeats are adjacent in-frame coding DNA sequences of 2 to 200 nucleotides in length that are directly repeated; these repeated units may be completely identical or partially degenerate (Li *et al.* 2004). The number of these coding-repeat copies often varies among different isolates leading to expansion or contraction of amino-acid blocks. Coding repeats have been observed in a number of prokaryotic and eukaryotic genomes where they play an important role in generating variability in cell-surface immunogenic antigens and adhesins, thereby evading the immune system or enhancing pathogenicity (Gravekamp *et al.* 1998; Jordan *et al.* 2003; Verstrepen *et al.* 2005). The inter-strain variability in the number of coding sequences can also serve as an extremely robust and rapid typing technique. Sequence analysis of a single, highly-variable gene, Protein A (*spa*) or clumping factor (*clfB*) has been successfully applied to strain differentiation amongst *Staphylococcus aureus* isolates which generally exhibit low variability and poorly discernible population genetic structure (Shopsin *et al.* 1999). Recently, an analysis of the genome of *A. fumigatus* identified as many as 292 genes with internal repeats. Fourteen of 30 selected genes showed size variation of their repeat-containing regions among 11 clinical *A. fumigatus* isolates. One of these, the cell wall protein *Afu3g08990* is involved in conidial germination and adhesion (Levdansky *et al.* 2007).

Importantly, the repeat containing region of *Afu3g08990* or CSP (cell-surface protein) was shown to vary significantly between *A. fumigatus* isolates from various origin (Levdansky *et al.* 2007; Balajee *et al.* 2007) (Fig. 3) By simply sequencing the *Afu3g08990*

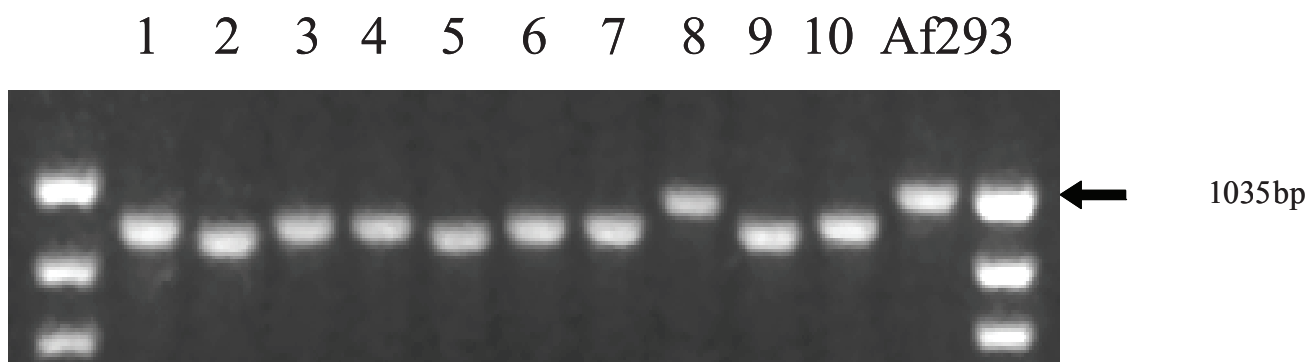


Fig. 3. The TR region in *A. fumigatus* gene *Afu3g08990* containing a leader sequence and GPI anchor shows size variability. The TR region of *Afu3g08990* was amplified by PCR in 11 clinical *A. fumigatus* isolates, including the genome-sequenced reference strain Af293. This figure shows size variation of the TR region between the strains.

repeat region in the various isolates and performing a phylogenetic analysis using the maximum parsimony method, it was possible to successfully "sub-type" fifty five epidemiologically linked *A. fumigatus* isolates from six nosocomial outbreaks of invasive aspergillosis. The results were concordant with another discriminatory genotyping technique, the *Afut1* RFLP typing method. However, while *Afut1* typing is labor and time intensive, needs specialised equipment and is not high throughput, *Afu3g08990* / CSP-typing requires only the ability to perform PCR and have access to an automated sequencer. Also, interpretation of the sequence information does not require sophisticated algorithms nor dedicated software and thus can be seamlessly integrated into any clinical microbiology laboratory.

It is worthy to note that in the *A. fumigatus* genome there is a substantial enrichment of putative cell-surface and/or secreted proteins that contain internal repeats. While 2.9 % of the ~9 900 genes in the *A. fumigatus* genome contain coding repeats, at least 12.5 % of all putative cell-wall encoding genes do so, a greater than 4-fold increase. This suggests that as found in a number of other fungal genomes, repeats in *A. fumigatus* may play an important role in generating variability in cell-surface immunogenic antigens and adhesins, thereby evading natural predators in its natural environment and the immune system in its inadvertent host.

CONCLUSIONS

Many different reasons may exist to explain the wish for being able to discriminate between different isolates of a given species. In the light of increasing reports of misidentified isolates, there is also clearly a need for more accurate and accessible methods to identify a fungal isolate to the species level. Ideally, these two parameters are combined in one typing/identification method. Based on the availability of genomic sequence data, highly targeted approaches allow strain typing methods to be directed at (a) highly specific part(s) of a specific fungal genome. Several of these typing methods can be used for typing purposes and may simultaneous be used for identification confirmation of fungal isolates: e.g. any clinical *A. fumigatus* isolate lacking the characteristic amplification products using a highly species specific typing method such as the STRAf assay or any isolate lacking the typical banding patterns obtained with the *Afut1* RFLP method is most likely not a true *A. fumigatus*. The true identity of such an isolate yet remains to be established using other methods. In contrast, more universally applicable typing methods that are not hampered by the species

barrier such as MLST and AFLP analysis are not only suitable for strain typing and identification purposes but they could additionally serve as parameter in the description and validation process of fungal species and in the delineation of the relationships between them.

REFERENCES

- Bain JM, Tavanti A, Davidson AD, Jacobsen MD, Shaw D, Gow NA, Odds FC (2007). Multilocus sequence typing of the pathogenic fungus *Aspergillus fumigatus*. *Journal of Clinical Microbiology* **45**: 1469–1477.
- Balajee SA, Tay ST, Lasker BA, Hurst SF, Rooney AP (2007). Characterization of a novel gene for strain typing reveals substructuring of *Aspergillus fumigatus* across North America. *Eukaryotic Cell* **6**: 1392–1399.
- Bart-Delabesse E, Humbert J-F, Delabesse E, Bretagne S (1998). Microsatellite markers for typing *Aspergillus fumigatus* isolates. *Journal of Clinical Microbiology* **36**: 2413–2418.
- Bikandi J, Millan RS, Rementeria A, Garaizar J (2004). In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* **20**: 798–799.
- Girardin H, Latge JP, Srikantha T, Morrow B, Soll DR (1993). Development of DNA probes for fingerprinting *Aspergillus fumigatus*. *Journal of Clinical Microbiology* **31**: 1547–1554.
- Gravekamp C, Rosner B, Madoff LC (1998). Deletion of repeats in the Alpha C protein enhances the pathogenicity of group B Streptococci in immune mice. *Infection & Immunity* **66**: 4347–4354.
- Jordan P, Snyder L, Saunders N (2003). Diversity in coding tandem repeats in related *Neisseria* spp. *BMC Microbiology* **3**: 23.
- Lasker BA (2002). Evaluation of performance of four genotypic methods for studying the genetic epidemiology of *Aspergillus fumigatus* isolates. *Journal of Clinical Microbiology* **40**: 2886–2892.
- Levdansky E, Romano J, Shadkchan Y, Sharon H, Verstrepen KJ, Fink GR, Oshero N (2007). Coding tandem repeats generate diversity in *Aspergillus fumigatus* genes. *Eukaryotic Cell* **6**: 1380–1391.
- Li Y-C, Korol AB, Fahima T, Nevo E (2004). Microsatellites Within Genes: Structure, Function, and Evolution. *Molecular Biology and Evolution* **21**: 991–1007.
- Ruiter MT de, de Valk HA, Meis JFGM, Klaassen CHW (2007). Retrotransposon insertion-site context (RISC) typing: a novel typing method for *Aspergillus fumigatus* and a convenient PCR alternative to restriction fragment length polymorphism analysis. *Journal of Microbiological Methods* **70**: 528–534.
- Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, Bost DA, Riehman M, Naidich S, Kreiswirth BN (1999). Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *Journal of Clinical Microbiology* **37**: 3556–3563.
- Tu O, Knott T, Marsh M, Bechtol K, Harris D, Barker D, Bashkin J (1998). The influence of fluorescent dye structure on the electrophoretic mobility of end-labeled DNA. *Nucleic Acids Research* **26**: 2797–2802.
- Vainer M, Enad S, Dolnik V, Xu D, Bashkin J, Marsh M, Tu O, Harris DW, Barker DL, Mansfield ES (1997). Short tandem repeat typing by capillary array electrophoresis: comparison of sizing accuracy and precision using different buffer systems. *Genomics* **41**: 1–9.
- Valk HA de, Meis JFGM, Curfs IM, Muehlethaler K, Mouton JW, Klaassen CHW (2005). Use of a novel panel of nine short tandem repeats for exact and high-

- resolution fingerprinting of *Aspergillus fumigatus* isolates. *Journal of Clinical Microbiology* **43**: 4112–4120.
- Valk HA de, Meis JFGM, Klaassen CHW (2007a). Microsatellite based typing of *Aspergillus fumigatus*: strengths, pitfalls and solutions. *Journal of Microbiological Methods* **69**: 268–272.
- Valk HA de, Meis JFGM, de Pauw BE, Donnelly PJ, Klaassen CHW (2007b). Comparison of two highly discriminatory molecular fingerprinting assays for analysis of multiple *Aspergillus fumigatus* isolates from patients with invasive aspergillosis. *Journal of Clinical Microbiology* **45**: 1415–1419.
- Varga J (2006). Molecular typing of aspergilli: Recent developments and outcomes. *Medical Mycology* **44** Suppl: 149–161.
- Verstrepen KJ, Jansen A, Lewitter F, Fink GR (2005). Intragenic tandem repeats generate functional variability. *Nature Genetics* **37**: 986–990.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407–4414.
- Warris A, Klaassen CHW, Meis JFGM, de Ruiter MT, de Valk HA, Abrahamsen TG, Gaustad P, Verweij PE (2003). Molecular epidemiology of *Aspergillus fumigatus* isolates recovered from water, air, and patients shows two clusters of genetically distinct strains. *Journal of Clinical Microbiology* **41**: 4101–4106.