*Plant Pathology* (2016) **65**, 968–977 Doi: 10.1111/ppa.12479

# Cytogenomic characterization of *Colletotrichum kahawae*, the causal agent of coffee berry disease, reveals diversity in minichromosome profiles and genome size expansion

A. S. Pires[ab], H. G. Azinheira[ac]*, A. Cabral[ac], S. Tavares[a], D. Tavares[d], M. Castro[d], V. Várzea[ac], M. C. Silva[ac], R. Abranches[b], J. Loureiro[d] and P. Talhinhas[abc]

[a]*CIFC – Centro de Investigação das Ferrugens do Cafeeiro, Instituto Superior de Agronomia, Universidade de Lisboa, Quinta do Marquês, 2784-505 Oeiras;* [b]*ITQB/UNL – Plant Cell Biology Laboratory, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. República, 2780-157 Oeiras;* [c]*LEAF – Linking Lanscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa; and* [d]*CFE – Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, 3000-456 Coimbra, Portugal*

*Colletotrichum kahawae* is an emerging fungal pathogen, which has recently undergone a speciation process from a generalistic '*C. gloeosporioides* species complex' background by acquiring the unique capacity to infect green coffee berries, thus causing coffee berry disease. This is a severe and widespread disease in Africa and an imminent threat to Arabica coffee cultivation in Asia and America, if the pathogen enters those continents. Genetic diversity within *C. kahawae* is low but notorious differences in pathogen aggressiveness have been described. This work characterized two cytogenomic traits (genome size and minichromosome profiles) of a collection of *C. kahawae* isolates, representing the breadth of its genetic diversity and distinct aggressiveness classes, along with closely related taxa. The results obtained constitute the first flow cytometry-based genome size estimation in the genus *Colletotrichum* and show a *c.* 8 Mb genome size expansion between *C. kahawae* (79·5 Mb on average) and its closest relatives (71·3 Mb), corroborating evidence indicating that *C. kahawae* (i.e. the coffee berry disease pathogens) should remain as a distinct species. Results have also shown the presence of two to five minichromosomes in *C. kahawae*, suggesting a positive relationship between the number of minichromosomes and the level of aggressiveness of the different isolates analysed, while no correlation could be established between aggressiveness and whole genome size. Overall, these results may be the basis for the identification of pathogenicity/aggressiveness-related factors in such minichromosomes, and may provide clues to the characterization of specific markers for aggressiveness classes.

*Keywords:* coffee berry disease, *Colletotrichum gloeosporioides* species complex, *Colletotrichum kahawae*, cytogenomics, genome size, minichromosomes

## Introduction

*Colletotrichum kahawae*, the causal agent of coffee berry disease (CBD), is an emerging fungal pathogen and the major limiting factor to the production of Arabica coffee (*Coffea arabica*) in Africa. Crop losses occur upon infection of green berries, with the formation of dark sunken lesions leading to anthracnose-like symptoms that cause their premature dropping and mummification, and up to 70–80% of yield losses if no control measures are adopted (Silva *et al.*, 2006; Van der Vossen & Walyaro, 2009). Occurrence of CBD depends mostly on climate conditions and altitude ranges; regions of higher altitude are more prone to the disease than lower altitude regions, and mild temperatures favour the disease (Gichuru *et al.*, 2008; Manuel *et al.*, 2010). This disease was first detected in 1922 in Kenya, and is so far restricted to Africa, but there are major concerns about the risk of its introduction to America and Asia, mimicking another major coffee disease, coffee leaf rust, which has spread from Africa to Asia and America over the last 150 years (Silva *et al.*, 2006). Understanding the biology and pathology of *C. kahawae* is therefore of utmost importance for a deeper comprehension of pathogen–host interactions and subsequently for an adequate deployment of resistance breeding strategies, not only in Africa but also in Asia and America.

*Colletotrichum kahawae* is very closely related to other members of the *C. gloeosporioides* species complex, but only *C. kahawae* is capable of infecting green coffee berries. Other *Colletotrichum* species (including other members of the *C. gloeosporioides* species complex along with *C. acutatum*) are also found in coffee berries, but they act as saprobes by invading mature coffee berries and are unable to infect green berries (Chen *et al.*, 2003). In fact, a phylogenetic analysis has shown

*E-mail: hgazinheira@gmail.com

*Published online 14 December 2015*





Plant Pathology

evidence of a recent host jump underlying the speciation of *C. kahawae* from a population of *C. gloeosporioides sensu lato* isolates from diverse hosts (Silva *et al.*, 2012). Such a close relationship between *C. kahawae* and non-CBD-causing *Colletotrichum* fungi led several researchers to consider that, based on molecular data, the CBD pathogens should not be considered as a separate species (Sreenivasaprasad *et al.*, 1993; Weir *et al.*, 2012), although pathogenicity and biological data suggest otherwise (Waller *et al.*, 1993; Silva *et al.*, 2012). Currently, the CBD pathogens are either recognized as a separate species, *C. kahawae* (Silva *et al.*, 2012), or as a subspecies, *C. kahawae* subsp. *kahawae* (Weir *et al.*, 2012). The phylogenetically closest, non-CBD-causing fungi have multiple geographic origins and hosts and, as a group, are either regarded as an undescribed group (Silva *et al.*, 2012) or as members of *C. kahawae*, clustering under *C. kahawae* subsp. *ciggaro* (Weir *et al.*, 2012).

Several DNA and isoenzymatic studies coincided in recognizing little variability among CBD isolates (Derso & Waller, 2003; Bridge *et al.*, 2008; Manuel *et al.*, 2010; Loureiro *et al.*, 2011; Silva *et al.*, 2012), despite its phylogeographical structure in three distinct populations (Angola, Cameroon and East Africa) according to a multigene nucleotide diversity analysis (Silva *et al.*, 2012). However, variations in aggressiveness among *C. kahawae* isolates are notorious (Manga *et al.*, 1997; Várzea *et al.*, 1999; Silva *et al.*, 2006) and not explained by those DNA and isoenzymatic studies.

In several pathosystems, cytogenetic and cytogenomic studies have provided important clues to demonstrate links between pathogenicity traits and genome structure or chromosomal composition, including several examples from ascomycetes such as *Alternaria alternata* (Hatta *et al.*, 2002), *Fusarium* (Ma *et al.*, 2010), *Leptosphaeria maculans* (Balesdent *et al.*, 2013) and *Mycosphaerella graminicola* (Stukenbrock *et al.*, 2010). Supernumerary, often small-sized, chromosomes are abundant in fungi and of relevance to plant pathology (Covert, 1998). Moreover, events leading to whole genome size variation, namely related to polyploidy, aneuploidy and mobile elements activity, also play important roles in genome evolution, population dynamics and pathogenic specialization in fungi (Stukenbrock & Croll, 2014). No such relationships between pathogenicity traits and cytogenetic or cytogenomic characteristics have been established so far for *Colletotrichum* fungi, although genome size variation and supernumerary chromosomes have been reported in *C. lindemuthianum* and *C. gloeosporioides*, respectively (He *et al.*, 1998; O'Sullivan *et al.*, 1998).

Considering all this, the aim of this work was to evaluate whether cytogenomic traits (namely genome size and supernumerary chromosomes) could contribute to explain host specialization of the CBD pathogens, i.e. the differentiation of *C. kahawae*, and its diversity. A collection of 15 *C. kahawae* isolates representing different levels of aggressiveness and the breadth of its genetic diversity was analysed along with a selection of eight phylogenetically closest non-CBD-causing isolates. Genome size was determined using flow cytometry, currently the gold standard for genome size determination in fungi (D'Hondt *et al.*, 2011), and the small supernumerary chromosomes (minichromosomes) profile was obtained by pulsed-field gel electrophoresis (PFGE), a technique enabling separation of chromosomes up to 10 Mb (Wieloch, 2006). Additionally, the karyotype was observed for a selection of isolates by microscopic examination of metaphase-arrested fungal samples.

## Materials and methods

### Fungal isolates

This study included isolates of *Colletotrichum* spp. associated with coffee and other hosts. CBD-causing isolates were previously obtained from infected green *C. arabica* berries in five African countries representing the three genetic groups that were previously differentiated (Silva *et al.*, 2012): Angola (isolates Ang6, Ang21, Ang29 and Ang81); Cameroon (isolates Cam1 and Cam2); and East Africa (isolates Que2, Zim1, Zim12, Uga2, Uga3, Uga4, Uga5, Uga6 and Uga9, from Kenya (Que), Zimbabwe (Zim) and Uganda (Uga), respectively). Non-CBD-causing isolates were selected in order to represent two populations closely related to *C. kahawae* (Silva *et al.*, 2012), *Colletotrichum* sp. 1 (= *C. kahawae* subsp. *ciggaro* (Weir *et al.*, 2012); isolates C1262.12, C1206.3 and PR432) and *Colletotrichum* sp. 2 (= *C. aotearoa* (Weir *et al.*, 2012); isolate C1282.4), along with *C. gloeosporioides sensu strictu* (isolates Uga8, PR220, PT800 and PT808; Talhinhas *et al.*, 2009). Isolates are stored at the CIFC collection.

### Aggressiveness assays

Although CBD-causing isolates employed in this study were selected according to contrasting aggressiveness profiles (Várzea *et al.*, 1999; authors' unpublished data), their aggressiveness was re-evaluated for direct comparison. Hypocotyls of *Coffea arabica* 'Caturra' (CIFC 19/1) were inoculated as previously described (Figueiredo *et al.*, 2013) using *C. kahawae* conidia suspensions ($2 \times 10^6$ conidia $mL^{-1}$). Symptoms were scored in hypocotyls over 1 month, daily over the first week, and every other day thereafter, according to the following disease severity scale: 0, no symptoms; 1, discrete lesions (necroses) less than 2 mm in length; 2, discrete lesions (necroses) less than 6 mm in length; 3, lesions surrounding at least half of the hypocotyl; 4, lesions totally surrounding the hypocotyl; 5, dead seedling. Four independent experiments were conducted for each isolate, with 16 seedlings in each. Average disease severity scores were used to calculate disease progression curves, the area under the disease progression curve (AUDPC) and the number of days required to reach severity score 4 (a point from which death of the seedling is unavoidable) for each isolate and experiment.

Non-CBD-causing isolates were previously confirmed as non-pathogenic to green coffee berries (Silva *et al.*, 2012).

### Genome size

The nuclear DNA content of each isolate was estimated by flow cytometry using 7- to 10-day-old mycelium grown in Petri dishes containing potato dextrose agar (Difco) medium at 21°C. Nuclei were released simultaneously from *Colletotrichum* hyphae and from leaves of the internal reference standard *Arabidopsis thaliana* 'Col-0' (2C = 0·32 pg or 313 Mb; Tavares

*et al.*, 2014) following the procedure of Bourne *et al.* (2014). In brief, nuclear suspensions were prepared by chopping 5–10 mg of fungal hyphae and *c.* 50 mg of *A. thaliana* leaf tissue with a razor blade in a glass Petri dish containing 1 mL woody plant buffer (WPB; 0·2 M Tris-HCl, 4 mM $MgCl_2$, 1% Triton X-100, 2 mM $Na_2EDTA$, 86 mM NaCl, 20 mM sodium metabisulphite, 1% PVP-10, pH 7·5; Loureiro *et al.*, 2007).

The suspensions were then filtered through a 30 $\mu$m nylon filter to remove large fragments, and nuclei were stained with propidium iodide (PI; Fluka) at 50 $\mu$g $mL^{-1}$. To prevent staining of double-stranded RNA, 50 $\mu$g $mL^{-1}$ RNase (Fluka) was also added. Samples were kept at room temperature and analysed within a 5 min period using a CyFlow Space flow cytometer (Partec GmbH) equipped with a 30 mW green solid-state laser emitting at 532 nm for optimal PI excitation.

Each day, prior to analysis, the overall instrument quality was evaluated using calibration beads green concentrate (Partec GmbH). After the analysis of the first sample of each isolate, the amplifier system was set to a constant gain, with the $G_1$ peak of the internal reference standard being positioned at a specific channel (between channel positions 400 and 600 on a 0–1024 scale). For each isolate, at least three independent replicate measurements were performed in order to assess the reproducibility of the results.

Data from each measurement were acquired using FloMax v. 2.4d software (Partec GmbH) in the form of three graphs: fluorescence pulse integral in linear scale (FL); fluorescence pulse integral in linear scale versus time (to monitor fluorescence stability over time); and fluorescence pulse integral in linear scale versus side light scatter in logarithmic scale (SSC). A region of interest comprising mostly the isolated nuclei was defined in the FL versus SSC cytogram and subsequently used to gate all the other graphs (see example in Fig. 1). Whenever possible, at least 3000 nuclei were analysed per sample.

The holoploid genome size in pg (1C for fungi; *sensu* Greilhuber *et al.*, 2005) was estimated using the following formula:

$$\frac{\text{Mean } G_1 \text{ fluorescence of sample nuclei}}{\text{Mean } G_1 \text{ fluorescence of reference standard}} \times \text{genome size of the reference standard}$$

The genome size in Mb was also calculated for all isolates using the factor 1 pg = 978 Mb (Doležel & Bartoš, 2005).

The reliability of the results was assessed through the careful evaluation of the symmetry and coefficient of variation (CV, %) of $G_1$ peaks in the FL histograms, and through the CV of the genome size estimation of each isolate based on the independent replicate measurements. Following the criteria established by Bourne *et al.* (2014), nuclear DNA content estimates were only considered when the CV values of DNA peaks were below 10%.

Genome sizes were compared among isolates using Tukey's Honestly Significant Difference mean comparison test at 95% confidence (STATISTICA v. 8.0; StatSoft Inc.). Nested design approaches were employed to compare *C. kahawae* and its groups (Angola, Cameroon and East Africa) against non-CBD-causing isolates.

## Minichromosome profile

Intact chromosomal DNA was prepared for each *Colletotrichum* following cultivation in 40 mL MP medium (3% malt extract, 0·5% peptone) at 22°C for 72 h. Mycelium was broken by vortexing using glass beads, and used as inoculum for further incubation in three 1 L Roux flasks containing 150 mL MP medium at 22°C for 72–96 h.

Mycelium was harvested by filtration through a nylon cloth, squeezed to remove all liquid and roughly chopped with a scalpel. Cells were suspended in OM buffer (1·2 M $MgSO_4$/10 mM sodium phosphate, pH 5·8) according to a ratio of 2 g mycelium:10 mL OM buffer with 12 mg $mL^{-1}$ lysozyme (Sigma-Aldrich), followed by incubation for 4·5 h at 28°C with gentle swirling. The suspension was filtered through nylon cloth and 30 mL filtrate containing the protoplasts were transferred to centrifuge tubes and overlaid carefully with 4 mL ST buffer (0·8 M sorbitol/10 mM Tris-HCl, pH 7·0).

Tubes were centrifuged at 1500 ***g*** in a SX4750 swinging bucket rotor (Beckman Coulter) for 20 min at 4°C. Protoplasts were recovered at the interface by pipetting, and subsequently centrifuged at 1000 ***g*** for 10 min. The pellet was washed three times with five volumes of ST buffer and the protoplasts were finally recovered in an equal volume of STC buffer (0·8 M sorbitol, 50 mM Tris-HCl pH 7·5, 10 mM $CaCl_2$).

An equal volume of 1·6% low melting point agarose (Bio-Rad) in STC buffer precooled to 50°C was added, and the agarose plugs were allowed to solidify at 4°C for at least 30 min.

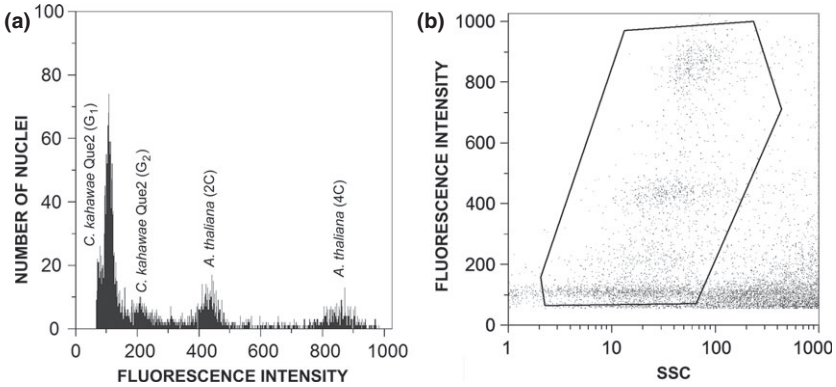


Figure 1 Example of genome size measurement results for *Colletotrichum kahawae* using flow cytometry. Flow cytometric histogram (a) of relative fluorescence intensities of propidium iodide-stained nuclei simultaneously isolated from *C. kahawae* isolate Que2 mycelium and the plant DNA reference standard, *Arabidopsis thaliana* (2C = 0·32 pg DNA), and dot-plot of side light scatter (SSC) versus fluorescence pulse integral in linear scale (b) applying one gating region to exclude partial nuclei and other types of debris as much as possible in order to improve the quality of the histogram.

The agarose plugs were incubated in 0·4 M EDTA, pH 8·0, 0·1% sodium *N*-lauroylsarcosinate and proteinase K (1 mg $mL^{-1}$) at 50°C for 96 h. Finally, the plugs were washed three times in water and stored at 4°C in 10 mM Tris-HCl pH 8·0, 100 mM NaCl.

Chromosomes were separated in 0·8 or 1% pulsed field certified agarose gel in 0·5× TBE (6·5 mM Tris base, 2·25 mM boric acid, 0·125 mM EDTA, pH 8·0) at 12°C in a CHEF-DR III system (Bio-Rad) device. Run times and conditions for each electrophoresis are described in the legends of Figures 3 and 4. The gels were stained with ethidium bromide (1 mg $mL^{-1}$) for 20 min and then destained in water for 2 × 30 min. For each run condition, each isolate was analysed by PFGE at least twice.

*Saccharomyces cerevisiae* chromosomal DNA (size marker, 0·225–2·2 Mb; Bio-Rad) was used as marker for calculation of chromosome size. Isolates were compared for the size of their minichromosomes, and a binary matrix was constructed to score chromosomes of similar sizes. The software NTSYSpc v. 2.02 h (Applied Biostatistics Inc.) was used to perform a cluster analysis by the unweighted pair group mean averages (UPGMA) method upon the calculation of a similarity matrix using the Dice coefficient.

### Chromosome count

The karyotypes of *C. kahawae* isolates Que2 and Ang29 were obtained using conidial suspensions released from potato dextrose agar cultures incubated on slides at 25°C until germination. Upon optimization of experimental conditions for each isolate, metaphases were arrested by incubation in 100 μL of 1 mg $mL^{-1}$ thiabendazole (Sigma-Aldrich; a metaphase-arresting agent that inhibits the formation of microtubules; Mehrabi *et al.*, 2007) for 1 h. Material was then fixed in methanol:acetic acid (9:1), flame dried and stained with 4′,6-diamidino-2-phenylindole (DAPI). Slides were then mounted in antifade solution (Vectashield mounting medium; Vector Laboratories Inc.). Observations were made by UV excitation using an epifluorescence microscope (DM6000B; Leica Microsystems) equipped with a iXon+ CCD camera (Andor Technologies).

## Results

### Genome size

In general, fluorescence histograms of good quality were obtained (Fig. 1a) when employing flow cytometry to estimate the genome size of *Colletotrichum* isolates under study after gating FL histograms with a polygonal region drawn around nuclei in the FL versus SSC cytogram (Fig. 1b) to improve the quality of the histograms. This way, partial nuclei and other types of debris were excluded from the analysis, resulting in $G_1$ fluorescence peaks with CV values below 10% (usually between 5 and 8%), which is within the range of accepted values for fungal species (Bourne *et al.*, 2014). The analysis was not affected by the endopolyploid nature of *A. thaliana* (Kudo & Kimura, 2001) because the two peaks of plant DNA reference standard visible in the scale set were readily assignable to 2C and 4C nuclei, respectively.

The average genome size of *C. kahawae* isolates ranged from 69·4 to 87·6 Mb (isolates Uga6 and Cam1, respectively), with an overall average of 79·5 Mb. The average genome sizes of isolates from Angola, Cameroon and East Africa groups was 78·1, 81·9 and 79·3 Mb respectively, with no significant differences between these groups (Table 1). Only a few *C. kahawae* isolates differed significantly for their genome sizes. For instance, in the East African group, the genome of isolate Uga6 was significantly smaller than those of isolates Zim1 and Zim12, while in the Cameroon samples, the genome of isolate Cam2 was significantly smaller than that of isolate Cam1.

The genome size of isolates clustering in *C. kahawae* subsp. *ciggaro* ranged from 68·4 to 73·9 Mb, with an average of 71·3 Mb, while the *C. aotearoa* isolate had a genome of 76·3 Mb. The genomes of *C. gloeosporioides* isolates ranged from 61·4 to 69·3 Mb, with an average of 65·6 Mb. Statistical analysis showed that *C. gloeosporioides* genomes are significantly smaller than any other species in this study, while *C. aotearoa* did not differ from *C. kahawae* (Table 1). In contrast, *C. kahawae* subsp. *ciggaro* differed significantly from *C. kahawae*, but not from *C. aotearoa*.

### Chromosome count

Treatment with thiabendazole was successful for obtaining fully condensed chromosomes that enabled reliable chromosome counting. The number of chromosomes was estimated to be between 11 and 14 for isolate Ang29 (Fig. 2a) and between 11 and 12 for isolate Que2 (Fig. 2b). Within each metaphase spread, the relative size of the chromosomes varied, including the presence of small dot-shaped chromosomes (Fig. 2). These could potentially correspond to the minichromosomes detected in the electrophoretic profiles.

### Minichromosomes

*Colletotrichum* minichromosomes were resolved by PFGE. Running conditions were optimized to separate bands below *c.* 2 Mb, because preliminary analyses showed no chromosomes in the 2–6 Mb range. Chromosome separations showed a variable number and size of minichromosomes across isolates, ranging from two to five minichromosomes, and from *c.* 270 kb to *c.* 1600 kb (Fig. 3). Isolates with similar patterns were Ang6 and Que2, which showed a pattern of two minichromosomes of *c.* 740 and 1070 kb, and isolates Uga2, Uga6 and Uga9, all sharing the same five minichromosomes. Cluster analysis (Fig. 3) further showed two main groups, one containing isolates from the Cameroon (Cam1 and Cam2) and the other containing isolates from East Africa (Que2, Uga2, Uga6, Uga9, Zim1 and Zim12). Remarkably the three Angolan isolates did not cluster together, with isolates Ang29 and Ang67 being remotely related to the East Africa and Cameroon groups, while Ang6 clustered together with Que2.

Among non-CBD-causing isolates, one minichromosome was identified in *C. gloeosporioides* isolate Uga8

Table 1 Genome size of *Colletotrichum kahawae* isolates, including other isolates from the *C. gloeosporioides* species complex, estimated by flow cytometry

| Species/group[a] | | Isolate | Genome size[b] | | | | | | Homogeneous groups[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean (fg) | Mean (Mb) | *n* | SD (Mb) | Range (Mb) | CV (%) | By isolate | By group | By species |
| *C. gloeosporioides* | | PR220 | 70·9 | 69·3 | 3 | 1·4 | 68·0–70·8 | 5·8 | abcd | a | a |
| | | PT800 | 62·7 | 61·4 | 3 | 6·0 | 58·1–69·2 | 7·2 | a | | |
| | | PT808 | 68·0 | 66·5 | 3 | 2·1 | 66·2–70·2 | 6·9 | abc | | |
| | | Uga8 | 66·6 | 65·1 | 4 | 2·1 | 63·6–68·2 | 6·9 | ab | | |
| *C. aotearoa* | | C1282.4 | 78·0 | 76·3 | 6 | 4·4 | 71·1–85·9 | 8·1 | cdef | bc | bc |
| *C. kahawae* subsp. *ciggaro* | | PR432 | 75·5 | 73·9 | 7 | 4·5 | 63·2–75·7 | 5·8 | abcde | b | b |
| | | C1206.3 | 69·9 | 68·4 | 5 | 6·5 | 63·2–78·6 | 6·7 | abc | | |
| | | C1262.12 | 73·3 | 71·7 | 4 | 6·0 | 65·8–78·4 | 9·9 | abcde | | |
| *C. kahawae* | East Africa | Que2 | 80·9 | 79·1 | 6 | 4·4 | 72·3–86·3 | 6·6 | defg | c | c |
| | | Uga2 | 81·8 | 80·0 | 3 | 3·8 | 77·2–84·3 | 7·7 | defg | | |
| | | Uga5 | 75·8 | 74·2 | 3 | 5·7 | 69·5–80·5 | 5·7 | abcdef | | |
| | | Uga6 | 71·0 | 69·4 | 3 | 6·5 | 64·8–84·3 | 7·1 | abcde | | |
| | | Uga9 | 82·6 | 80·7 | 3 | 6·4 | 75·7–87·9 | 6·8 | defg | | |
| | | Zim1 | 89·0 | 87·0 | 5 | 6·4 | 78·7–96·7 | 8·7 | g | | |
| | | Zim12 | 86·8 | 84·9 | 4 | 2·7 | 82·1–87·6 | 9·5 | fg | | |
| | Cameroon | Cam1 | 89·6 | 87·6 | 6 | 3·2 | 83·3–91·5 | 8·2 | g | c | |
| | | Cam2 | 77·9 | 76·2 | 6 | 3·7 | 69·6–80·7 | 6·3 | cdef | | |
| | Angola | Ang6 | 77·3 | 75·6 | 6 | 3·0 | 71·1–78·8 | 7·7 | bcdef | c | |
| | | Ang29 | 79·3 | 77·6 | 3 | 4·4 | 73·7–82·4 | 4·6 | cdefg | | |
| | | Ang67 | 82·9 | 81·1 | 6 | 6·4 | 75·9–91·4 | 8·8 | defg | | |

[a]Species within the *C. gloeosporioides* complex and groups among *C. kahawae* according to Weir *et al.* (2012) and Silva *et al.* (2012), respectively.
[b]*n*, number of samples measured; SD, standard deviation of genome sizes measured for each isolate; range, minimum and maximum genome sizes measured for each isolate; CV, average of coefficients of variation obtained in each measurement.
[c]Homogeneous groups obtained by Tukey's HSD test ($\alpha$ = 0·05) comparison among isolates, and among groups or species (isolates nested respectively in groups or species). Isolates, groups or species affected by the same letter do not differ significantly.

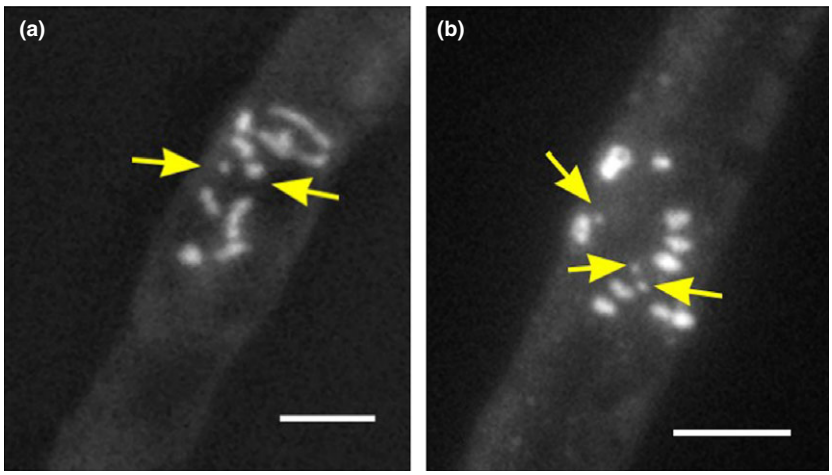


Figure 2 Karyotype of *Colletotrichum kahawae*. Well-separated metaphase chromosomes in intact germ tube cells for isolates Que2 (a) and Ang29 (b); arrows indicate potential minichromosomes. Bar = 2 $\mu$m.

and in *C. aotearoa* isolate C1282.4, while one to three minichromosomes were found in isolates C1262.12 and C1206.3 of *C. kahawae* subsp. *ciggaro* (Fig. 4).

## Aggressiveness

Results from aggressiveness assays allowed the differentiation of three groups of isolates according to high, medium or low aggressiveness levels (Fig. 5; Table 2). Ang29, Uga6 and Uga9 were classified as highly aggressive as they all achieved level 4 disease severity around 8 days after inoculation (d.a.i.), as well as AUDPC scores above 110. Ang6, Ang67, Que2 and Zim1 were the least aggressive isolates, as AUDPC scores were below 70 and they only achieved level 4 disease severity after 26 and 28 d.a.i. for Que2 and Zim1, respectively, while Ang6

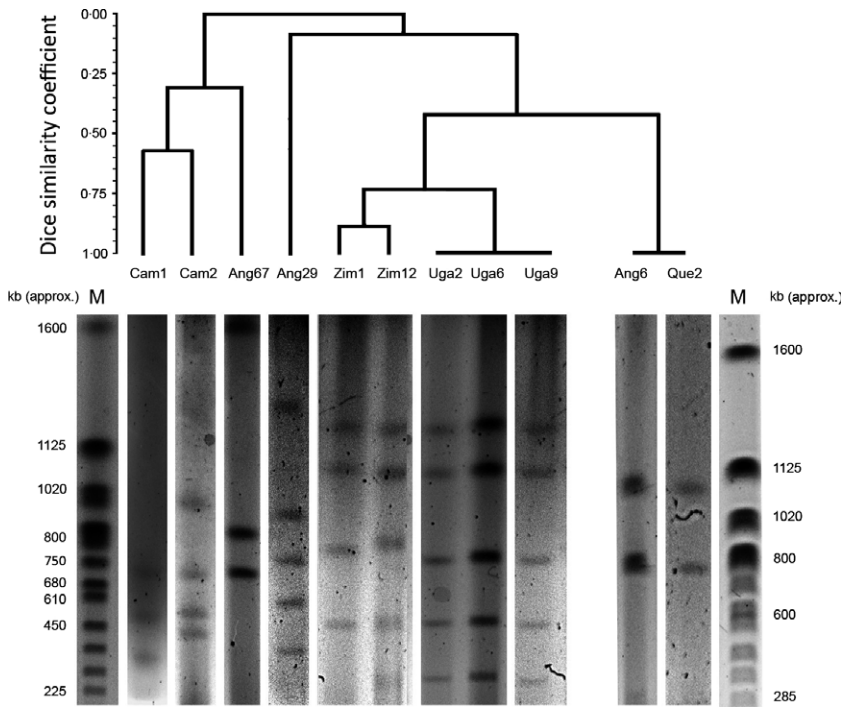


Figure 3 Cluster analysis (UPGMA) upon Dice similarity calculated among *Colletotrichum kahawae* isolates concerning minichromosome profiles were obtained by pulse field gel electrophoresis. Chromosomes were separated in a 0·8 or 1% pulsed field certified agarose gel in 0·5 × TBE at 12°C. Running conditions were as follows: Ang6 and Que2, 0·8% agarose gel, 32 h at 3 V $cm^{-1}$ with a 350–550 s switch time ramp at an included angle of 120°, followed by 13 h at 4·5 V $cm^{-1}$ with a 80–120 s switch time ramp at an included angle of 120°; others, 1% agarose gel, 24 h at 6 V $cm^{-1}$ with a 60–120 s switch time ramp at an included angle of 120°; M, size marker, 0·225–2·2 Mb *Saccharomyces cerevisiae* chromosomal DNA.

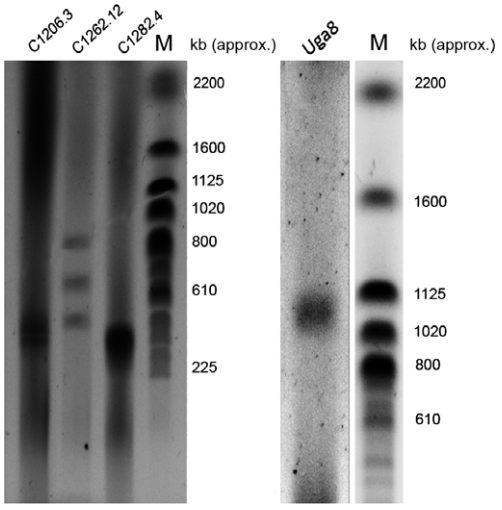


Figure 4 Minichromosomes of *Colletotrichum kahawae* allied taxa (non-CBD-causing isolates). Running conditions: 0·8% agarose gel, 50 h at 1·5 V $cm^{-1}$ with a 1000–2500 s switch time ramp at an included angle of 106°, followed by 16 h at 3·5 V $cm^{-1}$ with a 100–300 s switch time ramp at an included angle of 106°; M, size marker, 0·225–2·2 Mb *Saccharomyces cerevisiae* chromosomal DNA.

and Ang67 did not even reach this severity level (maximum 3·4 and 3·6, respectively). A medium level of aggressiveness was observed in isolates Cam1, Cam2, Uga2 and Zim12, with AUDPC values ranging between 93 and 104 and a severity level of 4 achieved 10–12 d.a.i.

## Discussion

This work has provided the first flow cytometry-based genome size data for *Colletotrichum* species, showing that the average genome size of *C. kahawae* is 79·5 Mb (±5·4 Mb), while that of *C. gloeosporioides* s.s. is 65·6 Mb (±3·3 Mb). The latter value is slightly higher than that obtained from genome sequencing (55·6 Mb; Gan *et al.*, 2012), a difference that could be due either to intraspecific variability in *C. gloeosporioides*, as different isolates were analysed in both studies, or to differences inherent to the techniques employed. Considering that genome size estimations obtained from flow cytometry can vary by a considerable fraction from those based on genome sequencing (Bennett *et al.*, 2003), it is still worth noting that the genome size obtained in this study for *C. kahawae* is within the range of genome sizes obtained (based on genome sequencing) for other species, such as *C. acutatum* (49·0 Mb; Baroncelli *et al.*, 2014); *C. higginsianum* (49·3 Mb; O'Connell *et al.*, 2012); *C. graminicola* (50·9 Mb; O'Connell *et al.*, 2012); and *C. orbiculare* (88·3 Mb; Gan *et al.*, 2012). Together, and considering also the genome size estimation for

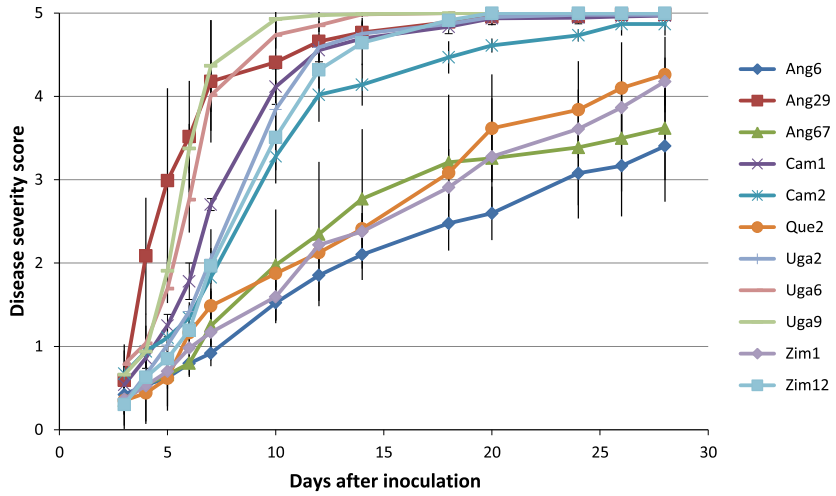


Figure 5 Disease severity of coffee berry disease symptoms scored on hypocotyls of *Coffea arabica* 'Caturra' upon inoculation with *Colletotrichum kahawae* isolates. For each data point, vertical bars represent standard deviation.

Table 2 Assignment of *Colletotrichum kahawae* isolates into aggressiveness groups according to the number of days after inoculation required to reach severity score 4 and to the area under the disease progression curve (AUDPC)

| Isolate | Aggressiveness group | No. of days to reach severity score 4 | AUDPC | No. of minichromosomes[a] |
|---|---|---|---|---|
| Uga9 | High | 7·8 (±1·5) | 113·5 (±0·7) | 5 |
| Ang29 | | 8·0 (±2·7) | 112·2 (±6·0) | 5 |
| Uga6 | | 8·5 (±1·7) | 111·6 (±1·4) | 5 |
| Cam1 | Medium | 10·0 (±0·0) | 103·3 (±1·7) | 4 |
| Uga2 | | 11·5 (±1·0) | 100·2 (±1·6) | 5 |
| Zim12 | | 11·5 (±1·0) | 98·3 (±2·4) | 5 |
| Cam2 | | 11·5 (±1·0) | 93·1 (±2·6) | 3 |
| Que2 | Low | 25·5 (±3·9) | 66·1 (±7·2) | 2 |
| Ang67 | | Never reached | 63·3 (±15·5) | 3 |
| Zim1 | | 26·7 (±2·3) | 62·1 (±4·1) | 4 |
| Ang6 | | Never reached | 52·6 (±6·2) | 2 |

Values in brackets refer to standard deviation.
[a]The number of minichromosomes (according to Figs 3 & 4) correlates with the AUDPC at $r^2 = 0{\cdot}673$.

*C. aotearoa* (78·0 Mb) and *C. kahawae* subsp. *ciggaro* (71·3 Mb) obtained in this study, these results set the average genome size in *Colletotrichum* at 66·5 Mb, *c.* 20 Mb above the overall average for fungi (45·9 Mb; Ramos *et al.*, 2015). Moreover, the results presented here show a genome size expansion of *c.* 8 Mb in *C. kahawae* as compared to the closely related fungus, *C. kahawae* subsp. *ciggaro*, underlying, at the cytogenomic level, the host jump speciation process that led to current *C. kahawae* populations (Silva *et al.*, 2012). In this context, it is also relevant to stress that no differences in genome size could be traced between the three genetically distinct *C. kahawae* populations identified in a multigene nucleotide diversity study (Silva *et al.*, 2012).

Besides determining the genome size, this work also reports the karyotype for *C. kahawae* based on the analysis of two genetically contrasting isolates. Karyological analysis by direct visualization of metaphase spreads allowed chromosome counting and the assessment of relative sizes. Between 11 and 12 chromosomes were estimated for isolate Que2 and 11–14 for isolate Ang29. Recently, the cytological analysis of the karyotype of the genus *Colletotrichum* (Taga *et al.*, 2014, 2015) revealed that the basic chromosome number (bCN) determined for several *Colletotrichum* species was $n = 10$ with or without an additional one or two minichromosomes, suggesting that bCN is rather conserved despite extensive speciation in this genus. In fact in the present work, the number of chromosomes estimated varied concomitantly with the number of minichromosomes observed by PFGE analysis (between two and five minichromosomes), suggesting a bCN for *C. kahawae* of $n = 9$. This value is within the range of that determined by Taga *et al.* (2015). Previously, the number of chromosomes in two distinct *C. gloeospori-*

*oides* lineages was estimated as $n = 6–8$ and $n = 13–15$, including 3–10 minichromosomes (Masel *et al.*, 1990). The comparison of the number of minichromosomes identified in *C. kahawae* in the current study (between two and five) and in related taxa, namely *C. kahawae* subsp. *ciggaro*, *C. aotearoa* and *C. gloeosporioides* s.s. (one to three) suggests no clear differences in this matter. However, the variability in number and size of minichromosomes among *C. kahawae* isolates revealed two main groups, one clustering all isolates from the East Africa genetic type and the other containing all isolates of the Cameroon genetic type. Surprisingly, a wide diversity of minichromosome profiles was identified among isolates of the Angola genetic type, with isolate Ang6 showing an identical profile to that of the Kenyan isolate Que2, isolate Ang67 being remotely related to the isolates from the Cameroon and isolate Ang29 placed apart from the two main groups. Although limited by the small number of isolates in the study, this larger variability in minichromosome profile of the Angola isolates is in agreement with the older phylogenetic origin estimated for this population (Silva *et al.*, 2012).

Variability in *C. kahawae* aggressiveness has long been recognized, with no clear relationship with genetic diversity. Although information on aggressiveness was available for several isolates used in this study, such information was scattered in various sources. The aggressiveness of the isolates involved in this study was therefore re-assessed for comparative purposes, revealing three groups (high, medium and low aggressiveness), each of them comprising isolates from various genetic groups, and thus confirming the absence of a clear relationship between genetic differentiation and aggressiveness. Interestingly, however, a relationship could be drawn between the number of minichromosomes and the level of aggressiveness. In particular, while most low aggressive isolates exhibited two to three minichromosomes, the medium and high aggressive isolates presented four to five minichromosomes (Table 2). A couple of exceptions include isolates Zim1 (low aggressiveness but four minichromosomes) and Cam2 (medium aggressiveness but three minichromosomes). Altogether, a moderate ($r^2 = 0{\cdot}673$) positive correlation between number of minichromosomes and aggressiveness was noted, while no correlation ($r^2 = 0{\cdot}0203$) could be established between aggressiveness and the whole genome size (or between genome size and number of minichromosomes; $r^2 = 0{\cdot}00799$), suggesting that polymorphism in minichromosomes is more relatable to aggressiveness than that in larger chromosomes. Besides the higher number of minichromosomes pointed out for the highly aggressive isolates, it is worth noting that the three isolates (Ang29, Uga6 and Uga9) rated as the most aggressive have in common only one of their five minichromosomes, at *c.* 700 kb, which in turn is absent from the remaining isolates (i.e. those classified as medium or low aggressiveness). Also, isolates Zim1 and Zim12 cluster together based on their minichromosome profiles, differing only in the absence/presence (for Zim1 and Zim12, respectively) of the *c.* 260 kb minichromosome, although they differ substantially in their aggressiveness, with Zim12 being more aggressive than Zim1. In both examples, the presence of these minichromosomes (i.e. the *c.* 700 kb minichromosome of isolates Ang29, Uga6 and Uga9, and the *c.* 260 kb minichromosome of isolate Zim12) in more aggressive isolates, and their absence in less aggressive isolates, could be an indication that such minichromosomes may harbour factors conditioning aggressiveness. In *L. maculans*, the causal agent of oilseed rape stem canker, the single minichromosome contains an effector gene, and its loss generates a fitness deficit (Balesdent *et al.*, 2013).

The surprising diversity of minichromosome patterns within *C. kahawae*, coupled with other reports on plant pathogenic fungi where polymorphisms in the number of small chromosomes among isolates can be involved in pathogenicity traits (O'Sullivan *et al.*, 1998; Ma *et al.*, 2010), prompts the need for further studies aimed at improving understanding of the molecular basis of the *C. kahawae* infection process, leading to better informed resistance breeding strategies. The identification of genetic or cytogenomic markers that may be linked to differences in aggressiveness could also be very important to epidemiology and aetiology, thus conditioning disease management strategies in each particular geographical region. Finally, the identification of a significant genome size expansion between *C. kahawae* and its closest relatives corroborates biological and pathological evidence (Silva *et al.*, 2012) indicating that *C. kahawae* should remain regarded as a distinct species, comprising solely the CBD pathogens. Significant genome size variability among *C. kahawae* isolates, not relatable to genetic diversity or aggressiveness, further suggests the occurrence and accumulation of presumably neutral variation in these populations.

## Acknowledgements

This work was supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia (Research unit GREEN-it 'Bioresources for Sustainability' UID/Multi/04551/2013, project PTDC/AGR-GPL/114949/2009 and grants SFRH/BPD/65686/2009, SFRH/BPD/84508/2012, SFRH/BD/89617/2012 and SFRH/BPD/88994/2012, attributed to A.S.P., A.C., M.C. and P.T., respectively). Authors are grateful to Dr Fátima Lopes (Instituto de Biologia Experimental e Tecnológica, Portugal) and Professor Luisa Brito (LEAF) for technical support and advice on PFGE, to Dr Ana Paula Pereira and Ana Vieira (CIFC) for the preparation of coffee seedlings and to Professor Ana Paula Ramos (LEAF), and Drs Peter Johnston and Bevan Weir (Landcare Research, New Zealand) for the isolates provided. The authors declare no conflicts of interest.

## References

Balesdent MH, Fudal I, Ollivier B *et al.*, 2013. The dispensable chromosome of *Leptosphaeria maculans* shelters an effector gene

conferring avirulence towards *Brassica rapa*. *New Phytologist* **198**, 887–98.
Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E, 2014. Draft genome sequence of *Colletotrichum acutatum* sensu lato (*Colletotrichum fioriniae*). *Genome Announcements* **10**, e00112–4.
Bennett MD, Leitch IJ, Price HJ, Johnston JS, 2003. Comparisons with *Caenorhabditis* (approximately 100 Mb) and *Drosophila* (approximately 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be approximately 157 Mb and thus approximately 25% larger than the *Arabidopsis* genome initiative estimate of approximately 125 Mb. *Annals of Botany* **91**, 547–57.
Bourne EC, Mina D, Gonçalves SC, Loureiro J, Freitas H, Muller LAH, 2014. Large and variable genome size unrelated to serpentine adaptation but supportive of cryptic sexuality in *Cenococcum geophilum*. *Mycorrhiza* **24**, 13–20.
Bridge PD, Waller JM, Davies D, Buddie AG, 2008. Variability of *Colletotrichum kahawae* in relation to other *Colletotrichum* species from tropical perennial crops and the development of diagnostic techniques. *Journal of Phytopathology* **156**, 274–80.
Chen ZJ, Ribeiro A, Silva MC *et al.*, 2003. Heat shock-induced susceptibility of green coffee leaves and berries to *Colletotrichum gloeosporioides* and its association to PR- and *hsp70* gene expression. *Physiological and Molecular Plant Pathology* **63**, 181–90.
Covert SF, 1998. Supernumerary chromosomes in filamentous fungi. *Current Genetics* **33**, 311–9.
Derso E, Waller JM, 2003. Variation among *Colletotrichum* isolates from diseased coffee berries in Ethiopia. *Crop Protection* **22**, 561–5.
D'Hondt L, Hofte M, Van Bockstaele E, Leus L, 2011. Applications of flow cytometry in plant pathology for genome size determination, detection and physiological status. *Molecular Plant Pathology* **12**, 815–28.
Doležel J, Bartoš JAN, 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* **95**, 99–110.
Figueiredo A, Loureiro A, Batista D *et al.*, 2013. Validation of reference genes for normalization of qPCR gene expression data from *Coffea* spp. hypocotyls inoculated with *Colletotrichum kahawae*. *BMC Research Notes* **6**, 388.
Gan P, Ikeda K, Irieda H *et al.*, 2012. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytologist* **197**, 1236–49.
Gichuru EK, Agwanda CO, Combes MC *et al.*, 2008. Identification of molecular markers linked to a gene conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica*. *Plant Pathology* **57**, 1117–24.
Greilhuber J, Doležel J, Lysák MA, Bennett MD, 2005. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Annals of Botany* **95**, 255–60.
Hatta R, Ito K, Hosaki Y *et al.*, 2002. A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics* **161**, 59–70.
He C, Rusu AG, Poplawski AM, Irwin JA, Manners JM, 1998. Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. *Genetics* **150**, 1459–66.
Kudo N, Kimura Y, 2001. Flow cytometric evidence for endopolyploidy in seedlings of some *Brassica* species. *Theoretical and Applied Genetics* **102**, 104–10.
Loureiro J, Rodriguez E, Doležel J, Santos C, 2007. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany* **100**, 875–88.
Loureiro A, Guerra-Guimarães L, Lidon FC, Bertrand B, Silva MC, Várzea V, 2011. Isoenzymatic characterization of *Colletotrichum kahawae* isolates with different levels of aggressiveness. *Tropical Plant Pathology* **36**, 287–93.
Ma LJ, van der Does HC, Borkovich KA *et al.*, 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **18**, 367–73.
Manga MB, Bieysse D, Mouen-Bedimo JA, Akalay I, Bompard E, Berry D, 1997. Observation sur la diversité de la population de *Colletotrichum kahawae* agent de l'anthracnose des baies du caféier Arabica. Implications pour l'amélioration génétique. In: *Proceedings of the 17th International Conference on Coffee Science, 1997, Nairobi, Kenya*. Paris, France: Association Scientifique Internationale du Café, 604–12.
Manuel L, Talhinhas P, Várzea V, Neves-Martins J, 2010. Characterisation of *Colletotrichum kahawae* isolates causing Coffee Berry Disease in Angola. *Journal of Phytopathology* **158**, 310–3.
Masel A, Braithwaite K, Irwin J, Manners J, 1990. Highly variable molecular karyotypes in the plant pathogen *Colletotrichum gloeosporioides*. *Current Genetics* **18**, 81–6.
Mehrabi R, Taga M, Kema GHJ, 2007. Electrophoretic and cytological karyotyping of the foliar wheat pathogen *Mycosphaerella graminicola* reveals many chromosomes with a large size range. *Mycologia* **99**, 868–76.
O'Connell RJ, Thon MR, Hacquard S *et al.*, 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genetics* **44**, 1060–5.
O'Sullivan D, Tosi P, Creusot F *et al.*, 1998. Variation in genome organization of the plant pathogenic fungus *Colletotrichum lindemuthianum*. *Current Genetics* **33**, 291–8.
Ramos AP, Tavares S, Tavares D, Silva MC, Loureiro J, Talhinhas P, 2015. Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the largest fungal genome reported – 2489 Mbp. *Molecular Plant Pathology* **16**, 1006–10.
Silva MC, Várzea V, Guerra-Guimarães L *et al.*, 2006. Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Brazilian Journal of Plant Physiology* **18**, 119–47.
Silva DN, Talhinhas P, Cai L *et al.*, 2012. Host-jump drives rapid and recent ecological speciation of the emergent fungal pathogen *Colletotrichum kahawae*. *Molecular Ecology* **21**, 2655–70.
Sreenivasaprasad S, Brown AE, Mills PR, 1993. Coffee Berry Disease pathogen in Africa: genetic structure and relationship to the group species *Colletotrichum gloeosporioides*. *Mycological Research* **97**, 995–1000.
Stukenbrock E, Croll D, 2014. The evolving fungal genome. *Fungal Biology Reviews* **28**, 1–12.
Stukenbrock EH, Jorgensen FG, Zala M, Hansen TT, McDonald BA, Schierup MH, 2010. Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. *PLoS Genetics* **6**, e1001189.
Taga M, Kaminakaya H, Tanaka K, Ogawa M, Kubo Y, 2014. Cytological analysis of karyotype evolution in the genus *Colletotrichum*. In: *Proceedings of the XVI International Congress on Molecular Plant–Microbe Interactions, 2014. Rhodes, Greece*. St Paul, MN, USA: International Society of Molecular Plant–Microbe Interactions, 642.
Taga M, Tanaka K, Kato S, Kubo Y, 2015. Cytological analyses of the karyotypes and chromosomes of three *Colletotrichum* species, *C. orbiculare*, *C. graminicola* and *C. higginsianum*. *Fungal Genetics and Biology* **82**, 238–50.
Talhinhas P, Neves-Martins J, Oliveira H, Sreenivasaprasad S, 2009. The distinctive population structure of *Colletotrichum* species associated with olive anthracnose in the Algarve region of Portugal reflects a host–pathogen diversity hot spot. *FEMS Microbiology Letters* **296**, 31–8.
Tavares S, Ramos AP, Pires AS *et al.*, 2014. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant Science* **5**, 422.
Van der Vossen HAM, Walyaro DJ, 2009. Additional evidence for oligogenic inheritance of durable resistance to coffee berry disease (*Colletotrichum kahawae*) in Arabica coffee (*Coffea arabica* L.). *Euphytica* **165**, 105–11.
Várzea VMP, RodriguesCJ Jr, Silva MC, Pedro JP, Marques DM, 1999. High virulence of a *Colletotrichum kahawae* isolate from Cameroon as

compared with other isolates from other regions. In: *Proceedings of the 18th International Conference on Coffee Science, 1999, Helsinki, Finland*. Paris, France: Association Scientifique Internationale du Café, 131.

Waller JM, Bridge PD, Black R, Hakiza G, 1993. Characterization of the coffee berry disease pathogen, *Colletotrichum kahawae* sp. nov.. *Mycological Research* **97**, 989–94.

Weir BS, Johnston PR, Damm U, 2012. The *Colletotrichum gloeosporioides* species complex. *Studies in Mycology* **73**, 115–80.

Wieloch W, 2006. Chromosome visualisation in filamentous fungi. *Journal of Microbiological Methods* **67**, 1–8.