**ABSTRACT** Candida glabrata is an important fungal pathogen which develops rapid antifungal resistance in treated patients. It is known that azole treatments lead to antifungal resistance in this fungal species and that multidrug efflux transporters are involved in this process. Specific mutations in the transcriptional regulator PDR1 result in upregulation of the transporters. In addition, we showed that the PDR1 mutations can contribute to enhance virulence in animal models. In this study, we were interested to compare genomes of two specific C. glabrata-related isolates, one of which was azole susceptible (DSY562) while the other was azole resistant (DSY565). DSY565 contained a PDR1 mutation (L280F) and was isolated after a time-lapse of 50 d of azole therapy. We expected that genome comparisons between both isolates could reveal additional mutations reflecting host adaptation or even additional resistance mechanisms. The PacBio technology used here yielded 14 major contigs (sizes 0.18–1.6 Mb) and mitochondrial genomes from both DSY562 and DSY565 isolates that were highly similar to each other. Comparisons of the clinical genomes with the published CBS138 genome indicated important genome rearrangements, but not between the clinical strains. Among the unique features, several retrotransposons were identified in the genomes of the investigated clinical isolates. DSY562 and DSY565 each contained a large set of adhesin-like genes (101 and 107, respectively), which exceed by far the number of reported adhesins (63) in the CBS138 genome. Comparison between DSY562 and DSY565 yielded 17 nonsynonymous SNPs (among which the was the expected PDR1 mutation) as well as small size indels in coding regions (11) but mainly in adhesin-like genes. The genomes contained a DNA mismatch repair allele of MSH2 known to be involved in the so-called hyper-mutator phenotype of this yeast species and the number of accumulated mutations between both clinical isolates is consistent with the presence of a MSH2 defect. In conclusion, this study is the first to compare genomes of C. glabrata sequential clinical isolates using the PacBio technology as an approach. The genomes of these isolates taken in the same patient at two different time points exhibited limited variations, even if submitted to the host pressure.

Infectious diseases caused by fungal pathogens represent a major threat to human health. It is estimated that ~2 billion people suffer from several forms of fungal diseases worldwide, from which ~1.5 million deaths will occur every year (Brown et al. 2012). Invasive fungal infections cause on average 50% mortality, with total number of deaths comparable to tuberculosis and malaria (Brown et al. 2012). The most problematic fungal pathogens include Candida, Aspergillus, and Cryptococcus species (Richardson and Lass-Flörl 2008). Within Candida species, Candida albicans and C. glabrata are known as the most-recovered species from infected patients (Marchetti et al. 2004). The current reports on the epidemiology of fungal diseases and their associated impact on human health reflect that the available therapeutic options, which include antifungal drugs, have limited efficacy. Only four major classes of antifungal drug are used to treat patients including azoles, polyenes, echinocandins, and pyrimidine analogs. Each of these agents have specific cellular targets and are used in several formulations...
with different derivatives (Ruhnke 2014). In general, most *Candida* species are susceptible to these drugs *in vitro*.

However, the use of antifungal agents has resulted in the inevitable occurrence of antifungal resistance. Antifungal resistance in a specific fungal isolate can be defined as a lack of activity of a given drug at concentrations that are higher than in a wild-type isolate. Antifungal resistance has been reported for virtually all existing antifungal agents and major fungal pathogens (Perlin et al. 2015). Frequencies at which antifungal resistance occur in hospitalized patients are variable based on data collected by population-based surveillance programs and available for major fungal pathogens including *C. albicans* and *C. glabrata*. In general, antifungal resistance rates for *C. albicans* are low. In a study performed between 2008 and 2011, resistance to fluconazole (cut-off: $\geq 64$ μg/ml) or echinocandins (cut-off: $\geq 4$ μg/ml) ranged from between 1 and 2% in bloodstream isolates (Cleveland et al. 2012). Resistance rates in *C. glabrata* are much higher: *C. glabrata* infections as a cause of invasive candidiasis increased from 18% of all BSI isolates in 1992–2001 to 25% in 2001–2007. Fluconazole resistance rates in *C. glabrata* increased from 9 to 14% in the same time period (cut-off: $\geq 64$ μg/ml) (Pfaller et al. 2003, 2009). Resistance of *C. glabrata* to the class of echinocandins is also increasing: within a 10-yr period (2001–2010) in a US hospital, the echinocandin resistance increased from 4.9 to 12.3% (Alexander et al. 2013). Similar trends are documented in Europe, however resistance rates are reported to be between 1 and 4% (Arendrup and Perlin 2014).

Recent studies have also documented the occurrence of multidrug resistance not only in *C. albicans* but also in *C. glabrata* (Sanglard 2016). *C. glabrata* cross-resistance between azoles and candins has captured more attention recently (Perlin 2015). For example, among echinocandin-resistant isolates sampled between 2008 and 2013 in two US surveillance hospital sites, 36% were also resistant to azoles (Pham et al. 2014). Between 2005 and 2013 in another US site, 10.3% *C. glabrata* isolates from cancer patients were resistant to caspofungin, from which ~60% were cross-resistant to azoles (Farmakiotis et al. 2014). Intriguingly, the latter study reported that caspofungin exposure alone could induce multidrug resistance without azole exposure. While *C. glabrata* azole resistance is mediated principally by efflux-dependent mechanisms, candin resistance is the consequence of mutations in the genes (*FKS1* and *FKS2*) encoding the candin target enzyme (B-glucan synthase) (Perlin 2015; Vale-Silva and Sanglard 2015). It has been shown recently that *C. glabrata* can exhibit high mutation rates in the presence of antifungal agents. This property is based on the decreased activity of the DNA repair machinery that exists in specific clinical isolates (Healey et al. 2016).

We have shown in the past years that *C. glabrata* azole resistance is coupled with selective fitness advantages in animal models of infection (Ferrari et al. 2009; Vale-Silva et al. 2013). Specific isolates exhibit upregulation of the adhesin *EPA1* when the transcriptional activator responsible for azole resistance CgPDR1 is mutated. When *EPA1* is upregulated, the consequence is that *C. glabrata* exhibits increased adherence to host cells, which is one of the factors that contributes to enhanced virulence of this yeast species which develops azole resistance (Vale-Silva et al. 2016). Our recent data also highlighted that *C. glabrata* can exhibit different degrees of adhesion to host cells. Given that adhesins are implicated in this process, it becomes relevant to address the repertoire of adhesins in specific *C. glabrata* lineages. Genome data are critical to achieve this goal, however the only available *C. glabrata* genome is up to now from the isolate CBS138 (Dujon et al. 2004). It is estimated that *C. glabrata* contains 63 ORFs with adhesin properties (de Groot et al. 2013). In a recent study, we compared CBS138 to other clinical isolates and found that a specific isolate (DSY562) exhibited the much higher adherence capacities to epithelial cells as compared with CBS138. This result was not dependent on *EPA1* expression since DSY562 expressed *EPA1* at much lower levels than CBS138.

These results stimulated our interest into the acquisition of the DSY562 genome. At the same time, it became also pertinent to compare the genome of this isolate with one (DSY565) that acquired azole resistance within a time-lapse of 50 d from DSY562 in a patient with oropharyngeal candidiasis (Sanglard et al. 1999). In this way, evolutionary pressure on the derived isolate could be estimated at the level of genome differences.

Here we report genome sequences from both related isolates. We used several approaches to resolve the genome with high accuracy. We analyzed the occurrence of polymorphisms between the two strains and also addressed the repertoire of genes encoding adhesins.

**MATERIALS AND METHODS**

**Strains and media**

Strains DSY562 and DSY565 were reported earlier (Sanglard et al. 1999) and were grown in YEPD at 30°C under constant agitation.

**DNA isolation**

Genomic DNA from DSY562 and DSY565 was obtained by the spheroplasting method as previously described (Calabrese et al. 2000).

**Multilocus sequence typing**

Multilocus sequence typing (MLST) was performed according to the study of Dodgson et al. (2003). Sequencing data of individual loci were analyzed using an online web tool (http://cgglabrata.mlst.net) to determine the sequence type (ST).

**Genome sequencing**

To produce high-quality genomic DNA from *C. glabrata* isolates, overnight cultures (5 ml) were first grown in YNB minimal medium [0.67% yeast nitrogen base (Difco) with 2% glucose] at 30°C under constant shaking to obtain $2 \times 10^8$ cells for each strain. The DNA isolation protocol followed instructions of the Gentra Puregene Yeast/Bact kit (Qiagen) with slight modifications. First, yeast cell lysis was performed with Zymolyase 100T (3 μg/μl end concentration) for 30 min at 37°C. In addition, after RNAse A treatment of precipitated nucleic acids, phenol/chloroform extractions were carried out according to recommendations issued by Pacific Bioscience (PacBio SampleNet—Shared Protocol) for the use of phenol/chloroform/isoamyl alcohol. After final precipitation with NaOHAc and several washes with 80% alcohol, the genomic DNA was dissolved carefully in a small volume of elution buffer (10 mM Tris-HCl, pH 8.5). Aliquots of 5 μg of extracted, high-quality genomic DNA was diluted to 150 μl using elution buffer at 30 μg/μl. Long insert SMRTbell template libraries were prepared (20-kbp insert size) according to PacBio protocols. In total, nine SMRT cells per strain were sequenced using P4 and P5 polymerase binding and C2 and C3 sequencing kits with 120 min (five first SMRT cells) and 180 min acquisition on PacBio RSII. *De novo* genome assemblies were produced using PacBio’s SMRT Portal (v2.0.0) and the hierarchical genome assembly process (HGAP version 3.0), with default settings and a seed read cut-off length of 6000 bp.

Illumina-based genome sequencing was performed by FASTERISSA (Plan-les-Ouates, Switzerland). Briefly, 4 μg of each DSY562 and DSY565 genomic DNA was fragmented by nebulization and the genomic libraries were prepared following the procedure described in the Genomic DNA Sample Prep Kit (Illumina, FC-102-1003). To obtain
inserts of 300 bp, a size selection was performed on agarose gel before a 10-cycle PCR amplification. The resulting libraries were sequenced in a 2 × 38-bp run on a Genome Analyzer IIx using the Sequencing Kit v4 (Illumina).

### Genome comparisons

The final PacBio assemblies of DSY562 and DSY565 were analyzed using GeneroUS (9.1.8). Genome comparisons were performed on the basis of each chromosome’s homolog using the Mauve aligner plug in (progressive Mauve algorithm) (Darling et al. 2004). The alignments were next inspected with a SNP discovery tool integrated in the software.

For comparisons between CBS138 and the Illumina-based genome sequencing of DSY562 and DSY565, reads were imported in the CLC genomic workbench software (v.9.5.2) (Qiagen). Variant detection was performed from the mapped reads using both probabilistic and quality-based variant detection tools available in CLC. A pipeline of the analysis of genome data are included in Supplemental Material, Figure S4.

### RNAseq

RNA extractions from the *C. glabrata* strains DSY562 and DSY565 were carried out as previously published (Ferrari et al. 2011). The strains were grown to logarithmic phase in 5 ml YEPD medium [1% Bacto peptone (Difco), 0.5% yeast extract (Difco), 2% glucose (Fluka)] at 30°C under constant shaking. RNA libraries for RNAseq were prepared with a TruSeq Stranded Total RNA Library Prep Kit (Illumina). The resulting libraries were processed using the Illumina TruSeq PE cluster kit v3. The resulting libraries were sequenced using the Illumina HiSeq2000 system using Illumina SBS Kit v3 reagents. Sequencing data were processed using Illumina Pipeline software version 1.82 to obtain fastq files.

### Sanger sequencing

Sequencing reactions were carried with a BigDye Terminator DNA Sequencing Kit (Thermo Fisher Scientific, Waltham, MA) and analyzed with a 3130xl genetic analyzer (Applied Biosystems). Primers used are listed in Table S1. Forward and reverse tag primer tails (5’-CGACGCCGCTGATA-3’ and 5’-GTCCGGGAGCCATG-3’) were added to each primer pair (labeled by suffixes "F" and "R"). PCR were performed on DSY562 and DSY565 DNA with the forward and reverse primers which was followed by sequencing with the tag primers.

### Adhesins predictions

DSY562 ORFs were analyzed with the help of the software FungalRV (http://fungalrv.igib.res.in/index.html), which helps to predict the occurrence of adhesin-like proteins in fungal genomes. To select for potential adhesins, a value of 0.5 was set as a threshold, as recommended (Chaudhuri et al. 2011). The list of potential adhesins was further subjected to several additional tests including a search for PFAM domains (http://pfam.xfam.org/search#tabview=tab1) and the presence of specific signatures relevant to the detection of adhesin-like proteins. These signature included (i) the presence of signal peptides and absence of internal transmembrane domains using the online programs SignalP and TMHMM (http://www.cbs.dtu.dk/services/); (ii) the presence of consensus sequence for GPI anchor ([NSGDAC]–[GAS-VIETKDLF]–[GASV]–X(4,19)–[FILMVAGPSTCYWNQ](10)–), as recommended by Weig et al. (2004); (iii) the presence of AWP motifs (VSHTT) (de Groot et al. 2013); (iv) the presence of EPA motifs (EPA CC: CX(27)CX(40)CX(44)DDX(13)CC) and EPA: GCSX(8,9)GL) following Diderrich et al. (2015) and Gabaldón et al. (2013); and (v) the presence of PIR motifs (Q[IV]XDGQ[IVP]Q) following (de Groot et al. 2008).

### Protein alignments

Protein alignments were performed with a Web tool available at http://mafft.cbrc.jp/alignment/server (MAFFT version 7) and default settings. A phylogenetic tree was constructed from the alignment produced from the tool. The produced alignments were next visualized after average linkage (UPGMA) and a bootstrapping with 1000 resamplings.

### Detection of splicing junctions

RNAseq data obtained from DSY562 and DSY565 were imported as fastq files into GeneroUS (9.1.8). Splice junctions were detected using the TopHat RNAseq aligner (Trapnell et al. 2009).

### Counterclamped homogeneous electric field (CHEF) analysis

Two or three colonies of each yeast strain were inoculated in YEPD liquid medium and incubated overnight at 30°C with agitation. A culture volume corresponding to 10⁶ cells was pelleted at 3000 × g for 5 min, and cells were washed in 5 ml 50 mM EDTA (pH 9). The pellet was resuspended in 330 μl 50 mM EDTA. Next, 110 μl of SCE [1 M sorbitol,

| Table 1 Characteristics of DSY562 and DSY565 |
|-----------------------------------------------|
| C. glabrata Clinical Strains | ST | MIC (μg/ml) | PDR1 Mutation | Time Elapsed (d) |
|-----------------------------------------------|
| | | Fluconazole | Voriconazole | Amphotericin B | |
| DSY562 | ST8 | 4 | 0.25 | 0.5 | — | — |
| DSY565 | ST8 | 64 | 1 | 0.5 | L280F | 50 |

ST, sequence type.

| Table 2 PacBio sequencing analysis of DSY562 and DSY565 |
|--------------------------------------------------------|
| Strain/assembly type | PacBio | |
| | DSY562 | DSY565 | CBS138 |
| SMRT cells | 9 (8 P4C2 + 2 P5C3) | 9 (8 P4C2 + 2 P5C3) | — |
| Number of bases (Gb) | 3.4 (coverage: 261x) | 2.8 (coverage: 215x) | — |
| Mean read length (kb) | 6.5 | 5.7 | — |
| Major contigs (including mitochondrial DNA) | 14 | 15 | 14 |
| Minor contigs (4-20 kb) | 5 | 9 | — |
| Total of assembled bases | 12,729,188 | 12,728,164 | 12,338,305 |
| Mitochondrial genome | 20,086 | 20,086 | 20,063 |
| N50 | 1,178,607 | 1,064,616 | — |
10 mM EDTA, 100 mM sodium citrate (pH 5.8), adjusted with citric acid] was added and then completed with 5% β-mercaptoethanol and Zymolyase 100T (100 units/ml) (solution I). The solution was gently mixed with 560 ml of GTG agarose and rapidly distributed into molds of 10 by 5 by 2 mm. The solidified plugs were immersed in solution II [450 mM EDTA, 10 mM Tris-HCl (pH 8), 7.5% β-mercaptoethanol] and incubated overnight at 37°C without agitation. Solution II was replaced by solution III [450 mM EDTA, 10 mM Tris-HCl (pH 8), 1% N-lauryl- Sarcosyl, and 150 mg/ml proteinase K]. Plugs were incubated for 6 hr at 65°C without agitation and kept for 10 min on ice. Finally, plugs were transferred to a 0.5 M EDTA (pH 9) solution and stored at 4°C for several months. One-third of the plug was loaded into wells of 0.5% GTG agarose gel in 0.5% Tris-borate-EDTA. The gel was then placed into the electrophoresis chamber of a CHEF DR II (Bio-Rad, Zürich, Switzerland) apparatus. Migration was performed in 0.5% Tris-borate-EDTA at 14°C with the following steps: block 1, 60- to 120-min switch, 6 V/cm, and 120°C over 24 hr; block 2, 120- to 300-min switch, 4.5 V/cm, and 120°C over 22 hr. After gel electrophoresis, gels were stained with ethidium bromide to reveal DNA under ultraviolet light.

**Table 3** Details of PacBio assembly of DSY562 and DSY565

|       | Length (bp) |       | Length (bp) |
|-------|-------------|-------|-------------|
| Chr A | 482,027     | Chr A | 482,027     |
| Chr B | 511,306     | Chr B | 511,306     |
| Chr C | 646,952     | Chr C | 628,526     |
| Chr D-L| 695,320     | Chr D-L| 691,913     |
| Chr E | 711,760     | Chr E | 712,009     |
| Chr F-L-I| 1,594,519  | Chr F-L-I| 1,594,552  |
| Chr G | 1,026,668   | Chr G | 1,009,061   |
| Chr H | 1,064,558   | Chr H | 1,064,559   |
| Chr I-L-D| 752,897     | Chr I-L-D| 744,977     |
| Chr J | 1,234,082   | Chr J1| 1,048,338   |
|       |             | Chr J2| 181,774     |
| Chr K | 1,305,194   | Chr K | 1,305,194   |
| Chr L | 1,178,610   | Chr L | 1,176,450   |
| Chr M-J| 1,424,660   | Chr M-J| 1,419,287   |
| Mitochondrial DNA | 20,086 | Mitochondrial DNA | 20,086 |
| Total Chr + mitochondrial DNA | 12,648,639 | Total Chr + mitochondrial DNA | 12,594,059 |
| unitig13_20kb | 20,067 | unitig5_7kb | 7847 |
| unitig5_14kb | 14,784 | unitig6_11kb | 11,645 |
| unitig2_20Kb | 23,113 | unitig12_22kb | 22,969 |
| unitig20_10kb | 10,815 | unitig16_17kb | 17,373 |
| unitig16_11kb | 11,770 | unitig17_9kb | 9902 |
|       |             | unitig18_13kb | 13,680 |
|       |             | unitig20_14kb | 14,791 |
|       |             | unitig22_16kb | 16,821 |
|       |             | unitig32_19kb | 19,077 |
|       |             | unitig18_13kb | 13,680 |
|       |             | unitig20_14kb | 14,791 |
|       |             | unitig22_16kb | 16,821 |
|       |             | unitig32_19kb | 19,077 |
| Total small contigs | 80,549 | Total small contigs | 134,105 |
| Total all contigs | 12,729,188 | Total all contigs | 12,728,164 |

**Figure 1** Chromosome structures of DSY562 as compared to CBS138. A circular map representing the major translocations between the CBS138 and DSY562 genomes was produced using J-Circos v2 (An et al. 2015).
Data availability

Strains described here are available upon request. PacBio assemblies are available under Bioproject PRJNA374542 (SRR5298594–SRR5298602 for DSY565, SRR5298595 and SRR5298512–SRR5298519 for DSY562). RNAseq data are available under Bioproject PRJNA374632 (SRR5296197–SRR5296202). Paired-end reads from Illumina-based genome sequencing are available under Bioproject PRJNA374542 (SRR5459192–SRR5459193). The final assembled genomes for DSY562 and DSY565 and their annotations are available under accession numbers MVOE00000000 and MVOF00000000, respectively.

RESULTS

Genome data acquisition

*C. glabrata* DSY562 and DSY565 are paired clinical isolates that were isolated from an HIV-positive patient presenting oropharyngeal candidiasis. DSY565 was isolated after 50 d of fluconazole therapy and exhibited azole resistance (Sanglard *et al.* 1999) (Table 1). The relationship between the two isolates was assessed at this time by RFLP which concluded a high relationship between the isolates. Before undertaking genome sequencing, both isolates were subjected to MLST typing according to Dodgson *et al.* (2005) (Table 1), and thus the high relationship between both isolates was confirmed.

Genome sequencing was first approached by Illumina HiSequation 38-bp, paired-end sequencing. Briefly, trimmed reads of both DSY562 and DSY565 were aligned to the reference genome of CBS138 using the software CLC Genomics Workbench. In both cases, ~95% of reads aligned to the CBS genomes and coverage was 72- to 64-fold for DSY562 and DSY565 data, respectively. Details on comparisons can be found in File S1 and File S2. Both comparisons yielded a high number of changes (85,711–85,986 including SNPs and indels) and resulted in ~10,390 nonsynonymous changes as compared to CBS138. Given that the genomes were aligned and compared directly to CBS138, they did not reflect actual genome rearrangements of strains DSY562 and DSY565. We therefore undertook an alternative genome-sequencing approach using PacBio technologies enabling de novo assembly of large reads. Since this technology is prone to random intrinsic high error rates, we aimed to reach genome sequencing coverage between 200- and 300-fold, which therefore required the use of several sequencing SMRT cells. The PacBio data are summarized in Table 2. As summarized in this table, the number of major assembled contigs (size >20 kb) were 14 and 15 for DSY562 and DSY565, respectively. N50s were 1.06 and 1.17 Mb for both isolates, thus highlighting the high quality of the assemblies. Table 3 summarizes the sums of nucleotides for each assembled contig, which were assigned to chromosomes according to the current CBS138 genome data (see below). DSY562 and DSY565 accounted for totals of 12,729,188 and 12,728,164 nucleotides. This represents a difference of 1024 nucleotides between the two genomes. The difference in nucleotide counts for major contigs (including mitochondrial DNA) is larger (58,400 nucleotides), however this is compensated for by nucleotide counts of the small contigs which could not be consistently assembled to larger contigs.

Genome comparisons between clinical strains and CBS138

Assembled contigs almost reconstituted the entire set of chromosomes that is known from the CBS138 genome. Figure 1 shows the chromosome maps of DSY562 as compared to CBS138 as well as a circular map to highlight major translocations between the reference and the DSY562 genomes. For practical reasons, the naming of chromosomes (A to M) was kept as proposed initially for CBS138. The data highlight

---

**Table 4 DSY562 and DSY565 annotation features**

| Chr | CDS | RNA | tRNA | rRNA | Introns | Conformed Introns | Chr | CDS | RNA | tRNA | rRNA | Introns | Conformed Introns | Chr | CDS | RNA | tRNA | rRNA | Introns | Conformed Introns |
|-----|-----|-----|------|------|--------|---------------|-----|-----|-----|------|------|--------|---------------|-----|-----|-----|------|------|--------|---------------|
| A   | 205 | 9   | 4    | 10   | 7      | 9             | B   | 220 | 9   | 4    | 10   | 6      | 9             | C   | 235 | 9   | 4    | 10   | 7      | 9             |
| D   | 288 | 9   | 4    | 10   | 7      | 9             | E   | 300 | 9   | 4    | 10   | 6      | 9             | F   | 387 | 9   | 4    | 10   | 7      | 9             |
| G   | 382 | 9   | 4    | 10   | 7      | 9             | H   | 387 | 9   | 4    | 10   | 6      | 9             | I   | 440 | 9   | 4    | 10   | 7      | 9             |
| J   | 523 | 9   | 4    | 10   | 7      | 9             | K   | 523 | 9   | 4    | 10   | 6      | 9             | L   | 523 | 9   | 4    | 10   | 7      | 9             |
| M   | 523 | 9   | 4    | 10   | 7      | 9             | N   | 523 | 9   | 4    | 10   | 6      | 9             | O   | 523 | 9   | 4    | 10   | 7      | 9             |
| P   | 523 | 9   | 4    | 10   | 7      | 9             | Q   | 523 | 9   | 4    | 10   | 6      | 9             | R   | 523 | 9   | 4    | 10   | 7      | 9             |
| S   | 523 | 9   | 4    | 10   | 7      | 9             | T   | 523 | 9   | 4    | 10   | 6      | 9             | U   | 523 | 9   | 4    | 10   | 7      | 9             |
| V   | 523 | 9   | 4    | 10   | 7      | 9             | W   | 523 | 9   | 4    | 10   | 6      | 9             | X   | 523 | 9   | 4    | 10   | 7      | 9             |
| Y   | 523 | 9   | 4    | 10   | 7      | 9             | Z   | 523 | 9   | 4    | 10   | 6      | 9             | a   | 523 | 9   | 4    | 10   | 7      | 9             |

---

**Volume 7 August 2017 | Candida glabrata Genomes | 2417**
several chromosomal rearrangements between the two isolates. While Chr A, B, E, G, H, and K kept similar structures between both strains, other chromosomes underwent several modifications. Chr C, D, J, and M in DSY562 underwent exchanges from CBS138 chromosomes essentially at chromosomal extremities. Parts of Chr L and I were added to Chr F (which was renamed as Chr F-L-I) in DSY562, thus yielding the largest chromosome in this species (1.59 Mb). Chr D and L were joined to the shortened Chr I (which was renamed as Chr I-L-D). The left arms of Chr L and C were inverted as compared to CBS138. Remarkably, the separate DSY565 assembly yielded the same chromosome arrangement as DSY562, with the exception that Chr J was not completely assembled in DSY565. This chromosome was still split in two parts (Chr J1 and J2; see Table 4 and Figure S1) in DSY565. A CHEF was performed with strains CBS138 and DSY562/DSY565 to compare these assemblies with physical chromosome separations. Remarkably, size predictions from the assemblies corresponded well to physical separations (see Figure 2). Red arrows shown in Figure 2 correspond to size conservation of Chr A, B, G, H, and J between the two strains, which is consistent with predictions. Chr C, D, E, and F-L-I were shifted to higher sizes (yellow color) by CHEF in DSY562, which is also consistent with assembly predictions. Chr M and I-L-D exhibited decreased sizes (blue color) as compared to CBS138, which is also predicted from assemblies. Chr K from DSY562 showed a discrepancy from the expected patterns from CBS138, however this result may be due to intrinsic chromosome size variations observed by others for this type of strain (Rader et al. 2012).

With these data, we first compared the DSY562 PacBio-assembled data (DSY562-HGAP3) with the CBS138 published genome (CGD, version s02-m04-r02). The DSY562-HGAP3 genome was first annotated with the tools available in the software Geneious (version 9.1.8, “live-annotation” tool) with a threshold level of 95% between comparisons. DSY562-HGAP3 was also further annotated on the basis of a recently published annotation using CBS138 and RNAseq data (Linde et al. 2015). As observed from Table 4, the number of deduced CDS (locus tag prefix B1J91) in DSY562 was 5294, which was near to the 5293 predicted CDS from CBS138. The DSY565 genome was next annotated on the basis of DSY562 annotations with the same Geneious tool and yielded 5298 CDS (locus tag prefix B1J92). Table 4 indicates that the number of CDS in mitochondrial DNA and numbers of tRNA were identical between strains. rDNAs were identified in DSY562 and DSY565, and were located at the telomeric ends of Chr L and M-J as in CBS138.

The similarity between the total number of CDS in the three strains (5293–5298 CDS) obscures, however, the following divergence: when comparing the CDS of each isolate, only a total of 5225 CDS were shared between them. 65 CDS from CBS138 were not detected in DSY562 and DSY565 (Figure 3 and File S3). 78 CDS from DSY562 and DSY565 were not found in CBS138. From these 78 CDS, 61 were common to DSY562 and DSY565; thus highlighting unique CDS for each of the strains. These above-mentioned results suggest that several selection events resulted in the loss and gain of several genes between these strains. When inspecting the identity of these 61 genes, we found that most of the genes gained in DSY562 and DSY565 as compared to CBS138 belonged to adhesin-like genes (34). Nine genes originating from DSY562 and DSY565 (named here Tkp5-1 to -9) were most similar to the Tkp5 protein from Vanderwaltozyma polyspora. This type of protein encodes elements of Ty-like retrotransposons (see Discussion). Other common genes between DSY strains underwent duplication events, including B1J91/B1J92_D02794g (endoplasmic reticulum membrane protein), B1J91/B1J92_M05115g (a putative retrotransposon protein), B1J91/B1J92_H04257g, and B1J91/B1J92_H04279g (encoding copper metallothionein MT-II and MT-IIB in DSY562 and DSY565 respectively). Interestingly, B1J91/B1J92_H04257g and B1J91/B1J92_H04279g were found in four

![Figure 2 CHEF chromosome separations of isolates CBS138, DSY562, and DSY565. Chromosome sizes are indicated for CBS138 and were extracted from public databases. Colored arrows show the size changes between the CBS and DSY562/DSY565 genomes. Red, yellow, and blue colors signify size conservation, increased sizes, and decreased sizes between CBS138 and the two different genomes, respectively. The attached table lists the chromosome sizes deduced from genome data of CBS138 and DSY562/DSY565. Plotting the relative migration distance (Rf values) of known CBS138 chromosome sizes with the obtained sizes of DSY562 and DSY565 chromosomes yielded an R² value of 0.981.](image-url)
tandem repeats on Chr H as opposed to CBS138, in which only one repeat is found. The remaining 10 genes out of the 61 common to DSY562 and DSY565 were ranked as pseudogenes in the current CGD (some of which are also described as adhesin-like genes), thus explaining their absence from the CBS genome used in our analysis. Genes that were CBS138 specific (65) included, among others, adhesin-like genes (6), CAGL0B01243g (MATα1), and CAGL0G07175g (RNA binding protein with RNA-directed DNA polymerase activity). Notably, MATα1 is still present in DSY562 and DSY565 but as another CDS (B1J91/B1J92_B00242g). Most of the remaining CBS138-specific genes were of unknown function (48) (see details on File S3).

At the level of single nucleotides, large differences were observed between DSY562 and CBS138 (Table 5). Nuclear and mitochondrial genomes diverged by 0.7% between the two strains. These data indicate extensive divergence between the two genomes, however they are in general agreement with the analysis deduced from other C. glabrata genomes (Singh-Babak et al. 2012). The high number of nonsynonymous SNPs (11,014) underscores a large number of amino acid substitutions. A high number of SNPs occurred in genes encoding adhesins, and the following section will discuss the resolution of this important gene category. A few genes underwent frame shifts resulting in protein truncations. For example, B1J91/B1J92_D00154g (AQY1, role in iron transport) was truncated in DSY562 and DSY565 (CDS length of 459 bp) as compared to CBS138 (CDS length of 873 bp). Interestingly, another homolog is present in DSY562 and DSY565 (B1J91/B1J92_A01221g), thus likely compensating for the probable loss of function of B1J91/B1J92_D00154g. B1J91/B1J92_C02343g (ARBI1: a protein with ATPase activity and a role in cellular response to drugs, a ribosomal small subunit exported from the nucleus and cytosol), which contains an intron in CBS138, lacks a reading frame in DSY562 and DSY565. Removal of the intron in this gene restores a truncated CDS as compared to CBS138. B1J91/B1J92_E01309g (EKII: choline kinase activity, a role in the phosphatidylcholine biosynthetic process) exhibited an N-terminal deletion due to the use of other intron splicing as compared to the CBS138 homolog. B1J91/B1J92_M05687g (NPY1: has NAD+ diphosphatase activity and a role in the NADH metabolic process), B1J91/B1J92_G03487g (TMN3: has a role in cellular copper ion homeostasis), B1J91/B1J92_I10725g (YGR122W: has a role in negative regulation of transcription from the RNA polymerase II promoter), B1J91/B1J92_E01155g (RPA14: RNA polymerase I activity), and B1J91/B1J92_H06017g (FLR1: a multidrug transporter of the major facilitator superfamily) exhibit N-terminal truncated proteins as compared to CBS138. Whether the identified protein truncations may result in deficient functioning remains to be established. We believe that these changes do not reflect PacBio errors since (i) the DSY562 and DSY565 assemblies were produced independently to produce the same changes as compared to CBS138, and (ii) Sanger sequencing verifications confirmed the absence of PacBio errors [see example of B1J91/B1J92_H06017g (FLR1) sequencing in Figure S2]. We also noticed that some previously annotated tandem CDS in DSY562 were now forming a single CDS in DSY562. For example, B1J91/B1J92_I02816g and B1J91/B1J92_I02838g (AZF1: asparagine-rich zinc finger protein) were now generating a single CDS (B1J91/B1J92_I02838g) in DSY562. B1J91/B1J92_E03421g and B1J91/B1J92_E03432g (close homologs of Saccharomyces cerevisiae CDA1 and CDA2 genes: chitin deacetylases) were now forming a single CDS (B1J91/B1J92_E03421g). It is interesting to observe that there are ~370000 nt differences between the CBS138 and DSY562 genomes (the latter having the largest genome), thus highlighting that additional/novel genes may be present in the DSY strains (see below). As a last step in these comparisons, we analyzed the presence of introns in strains DSY562 and DSY565 as compared to CBS138. This was performed first by the transfer of annotations from the recently updated CBS138 genome. We next used RNAseq data from both DSY562 and DSY565 transcriptomes to detect spliced RNAs. Table 5 shows that DSY562 and DSY565 genomes possessed at least 144 intron-containing genes, which is close to the number identified in CBS130 (150). RNAseq data from DSY562 and DSY565 showed that 114 and 115 splice sites, respectively, coincided with the predicted splice sites (File S4).

The cluster of adhesins in DSY562
C. glabrata is known for its extensive repertoire of adhesins. The genes encoding adhesins often contain repeated motifs and are mostly localized at chromosome extremities. These two characteristics make it challenging to establish a detailed and complete listing of adhesins from C. glabrata genomes. Here we have used PacBio genome assembly approaches to address these challenges and we used a systematic analysis (described in Materials and Methods) to identify adhesin-like genes in DSY562 and DSY565. Briefly, all DSY562 CDS were screened with the help of the software FungalRV, which helps to predict the occurrence of adhesin-like proteins in fungal genomes (Chaudhuri et al. 2011). We next analyzed the list of potential adhesins for the occurrence of PFAM

---

Table 5 Whole genome differences between DSY562 and CBS138

|                | CBS138 vs. DSY562 |
|----------------|-------------------|
|                | Nonsynonymous SNP | Synonymous SNP | In/Outside SNPs Indels | Indels | Total Percentage SNP-Indel vs. CBS |
| Total          | 11,014            | 25,845         | 86,915                 | 5353   |
| Percentage vs. CBS | 0.09             | 0.21           | 0.7                    | 0.04   |
| Mitochondrial  | 34                | 11             | 140                    | 57     |
| Percentage vs. CBS | 0.17             | 0.055          | 0.7                    | 0.285  |

Figure 3 Venn diagram representation of genes shared between DSY562, DSY565, and CBS138. The diagram was generated with a Web tool available at http://bioinformatics.psb.ugent.be/webtools/Venn/.
domains and the presence of specific signatures, including (i) the presence of signal peptides and absence of internal transmembrane domains, (ii) the presence of AWP motifs (de Groot et al. 2013), (iv) the presence of motifs typical for the class of EPA proteins following Diderrich et al. (2015) and Gabaldón et al. (2013), (v) the presence of PIR motifs, and lastly (vi) the occurrence of hyphally regulated cell wall proteins (Hyphal_reg_CWP, IPR021031). For practical reasons and for comparison purposes with CBS138, we used the same CAGL0 gene naming that is currently accepted for CBS138. When several highly similar proteins were identified in DSY562, we add a numerical suffix to them (i.e., 1, 2, etc.). According to our analysis, we proposed a list of 101 adhesin-like proteins in DSY562 (File S5 and File S6), which well exceeds the number of deduced adhesins in CBS138 (66) according to de Groot et al. (2013). Figure 4 shows the genomic localization of adhesins detected in DSY562 as compared to CBS138. Notably, the repertoire of adhesins in DSY565 is very similar to DSY562 (Figure S1, see below). ~50% of the adhesin genes were located in DSY562, near to the telomeric ends (region of 50 kb down to chromosome ends) of the chromosomes. One example of the expansion of adhesins in DSY562 is given by the end of Chr C (Figure 4). The size difference between Chr C between DSY562 and CBS138 may be explained by the tandem arrangement of a group of four adhesins. The DSY562 potential adhesins were aligned and a phylogenetic tree was assembled (Figure 5). This tree was merged with conserved domain features and with CBS138 protein homologs to, on the one hand, help distinguish between these
proteins and, on the other hand, to better delineate extensions of protein families in DSY562. Several protein clusters (highlighted by colors in Figure 5) were observed. Members of the EPA family (19) were well discerned in this tree (blue color of Figure 5) and harbor typical signature sof EPA adhesins, including PA14 domains as well as EPA-specific domains (see details in Material and Methods), with the exception of B1J91_K00170g. Duplication of specific EPAs occurred in DSY562 for EPA2, EPA6, EPA8, EPA12, and EPA22. The largest cluster of proteins contained VSHITT-motif repeats and PA14 motifs, however, they lacked EPA signatures (green color of Figure 5). CBS138 PWP homologs (PWP1,-2, -3, and -5) were found in this cluster. VSHITT-motif repeats were also found in another cluster (red color of Figure 5) that essentially lacked PA14 motifs. The largest extension occurred for adhesin-like proteins with VSHITT repeats that include members of the PWP and AWP family (Figure 5). Another well-distinguished cluster lacked both VSHITT and PA14 motifs and contained CBS138 AWP homologs (AWP6 and -7). Only one gene (B1J91_M11726g) contained a PIR motif and was extensively unrelated to other genes of the phylogenetic tree. The remaining proteins could not be assembled in clusters with common features. Taken together, the data suggest that 61 novel adhesin-like proteins were assigned to DSY562, while 40 remained in common with CBS138 (Figure 5, only 50 adhesin-like proteins without frameshifts were taken into account in these calculations). The majority of the novel proteins exhibited VSHITT domains.

**Genome comparisons between DSY562 and DSY565**

As we mentioned earlier, DSY562 and DSY565 are two highly related isolates that were obtained from a patient treated with fluconazole for oropharyngeal candidiasis. DSY565 was isolated ∼50 d after initiation of fluconazole treatment and was azole resistant. We reported that azole resistance was due to a mutation in the transcription factor CgPDR1. This mutation was contributing to an increase of virulence as compared to a parental strain and was measured in different animal models (Ferrari et al. 2009; Vale-Silva et al. 2016). Given that virulence traits may be multifactorial, we reasoned that other mutations may contribute to this phenotype. Direct genome comparisons with PacBio-generated data may provide some cues into this hypothesis. We first annotated the DSY562 genome to a high resolution as described above and used this basis as a tool to annotate the DSY565 genome. The use of CBS138 as an annotation basis provided similar results (data not shown). Table 2, Table 3, and Table 4 compare several features between DSY562 and DSY565. At the genomic level, a few changes occurred. While 101 adhesin-like genes were identified in DSY562, 107 adhesin-like genes were found in DSY565 (Figure S1). 95 adhesins were shared between both isolates (Figure 6). 95 adhesins were shared between both isolates (Figure 6) and the discrepancies between the two genomes were principally due to the presence of adhesin-like genes in small contigs, which could not be mapped to the chromosomes (Figure S1). 12 and three adhesin-like and unique genes were identified in the small contigs of DSY565 and DSY562, respectively. In addition, DSY562 possesses one gene (B1J91_E00110g) located at the extremity of Chr J that is not present in DSY565 (Chr J1). This may be due to an incomplete assembly of Chr J1 in this isolate. One other specific change between DSY562 and DSY565 occurred at the left arm of Chr C. A size difference of ∼29 kb exists at the left arm of Chr C between DSY562 and DSY565 (Figure S1). While a tandem repeat of B1J91_C01067g is present in DSY562 at the Chr C extremity, only one copy remained in DSY565. However, this gene is now situated between the adhesins B1J92_C00968g1 and B1J91_C01133g1 of DSY565 (Figure 6). One remaining copy of this specific adhesin (B1J92_C01067g1) is situated in the small contig unitig16_17kb, which suggests that this small contig
may be assigned to the missing nucleotides of the DSY565 Chr C left arm.

Another major difference between DSY562 and DSY565 is that DSY565 lacks the retrotransposon Tkp5-7 that is situated at the right arm of Chr M in DSY562. Interestingly, DSY565 has retained a single long terminal repeat (LTR) at this locus as a trace of the presence of the retrotransposon. A PCR-based approach confirmed the absence of Tkp5-7 in DSY565 while leaving the LTR signature (see Figure S3).

At the single nucleotide level, differences between DSY562 and DSY565 include synonymous and nonsynonymous SNP variations as well as indels. Table 6 gives an overview of the observed differences. First we noticed that most of the indels in these comparisons were in homopolymeric sequences, which are long stretches (10–30) of identical nucleotides (108 in total). Since it is known that PacBio can yield these differences, we undertook systematic verifications of these sequences by separate Sanger sequencing of small PCR fragments.

### Table 6 Nucleotide variations between DSY562 and DSY565

| Chr name | DSY562 vs. DSY565 |  |
|----------|------------------|---|
|          | Nonsynonymous SNP | Indels in CDS | Synonymous SNP | In/Outside SNP | Indels | Corrected Indels in Homopolymers |
| Chr A    | 2                 | 0             | 0              | 3             | 0      | 4                               |
| Chr B    | 0                 | 0             | 0              | 1             | 0      | 1                               |
| Chr C    | 1                 | 1             | 0              | 3             | 1      | 4                               |
| Chr D-L  | 3                 | 1             | 1              | 4             | 1      | 7                               |
| Chr E    | 2                 | 2             | 2              | 4             | 0      | 8                               |
| Chr F-L-I| 1                 | 1             | 3              | 7             | 3 (2)\* | 17                              |
| Chr G    | 0                 | 0             | 1              | 1             | 2 (1)\* | 9                               |
| Chr H    | 1                 | 0             | 0              | 2             | 3      | 4                               |
| Chr i-L-D| 3                 | 3             | 1              | 6             | 0      | 4                               |
| Chr J1   | 0                 | 1             | 2              | 5             | 3      | 3                               |
| Chr J2   | 0                 | 1             | 1              | 0             | 0      | 1                               |
| Chr K    | 1                 | 0             | 0              | 1             | 0      | 10                              |
| Chr L    | 0                 | 1             | 0              | 9             | 0      | 4                               |
| Chr M-J  | 3                 | 0             | 0              | 6             | 1      | 16                              |
| Mitochondrial | 0              | 0             | 0              | 0             | 0      | 3                               |
| DNA      | 0                 | 0             | 0              | 0             | 0      | 3                               |
| Total    | 17                | 11            | 11             | 52            | 14     | 95                              |

\*Number in parentheses indicates the number of single nucleotide indels.
### Table 7 Nonsynonymous SNPs between DSY562 and DSY565

| Chr | Locus_tag | Gene Name | Nucleotide Position | Nucleotide Change | Amino Acid Change | Function |
|-----|------------|-----------|---------------------|------------------|------------------|----------|
| Chr A | B1J91_A01804g | HXT3/HXT1 | 185747 | C → G | V560 → L | Hexose transporter |
| Chr A | B1J91_A00451g | PDR1 | 53729 | G → C | L280 → F | Transcription factor |
| Chr C | B1J91_C00968g3 | CAGL0C00968g | 41110 | GCA → TTG | A993 → N | Putative adhesin-like cell wall protein |
| Chr D-L | B1J91_D02442g | LEM3 | 408506 | T → C | V81 → A | Translocation of phospholipids and alkylphosphocholine transporters |
| Chr D-L | B1J91_D00429g | GSD2 | 648612 | C → T | L633 → F | Glycine dehydrogenase |
| Chr D-L | B1J91_D03762g | GIC2 | 274967 | G → T | T280 → K | CDC42 GTase-binding protein |
| Chr E | B1J91_E01414g | SAN1 | 160147 | C → T | P575 → S | Ubiquitin-protein ligase |
| Chr E | B1J91_C00209g2 | AWSP7 | 20343 | A → G | I71 → T | Putative adhesin-like cell wall protein |
| Chr F-L-I | B1J91_F05709g | ATC1 | 584210 | A → C | 268E → D | Nuclear protein; possibly involved in regulation of cation stress responses |
| Chr H | B1J91_H02057g | GARC1 | 873919 | C → T | 92T → I | Nuclearol RNA processing protein |
| Chr I-L-D | B1J91_I01001g | NOT3 | 85356 | T → A | E118 → D | Subunit of CCR4-NOT global transcriptional regulator |
| Chr I-L-D | B1J91_I07601g | COQ3 | 333430 | G → A | P171 → L | Mitochondrial ubiquinone biosynthesis |
| Chr I-L-D | B1J91_I08659g | YSCB4 | 471495 | C → T | G75 → R | Actin-binding protein |
| Chr K | B1J91_K04565g | YGR065c | 434920 | A → G | I218 → M | Vitamin H transporter |
| Chr M-J | B1J91_M08954g | YOR019w | 522442 | C → G | Q546 → E | Protein of unknown function; may interact with ribosomes |
| Chr M-J | B1J91_M00946g | YLR435w | 1307238 | C → A | D101 → Y | Protein with a potential role in pre-rRNA processing |
| Chr M-J | B1J91_M13233g | MNR2 | 105774 | T → R | 243R → H | Vacular membrane protein; putative magnesium transporter |

DSY562 locus_tag is given here. 
Gene name is inferred from the CBS138 annotations. 
Nucleotide position is from the DSY562 genome.
wild-type allele (Vale-Silva et al. 2016). Interestingly, a separate study demonstrated that DSY562 was a *C. glabrata* isolate with the highest capacity among those tested to bind *C. albicans* hyphae (Tati et al. 2016). Binding of *C. glabrata* involved the *C. albicans* Als3 adhesin. All these data suggested the potential presence of other adhesins in the genetic background of DSY562 and its related strain DSY565. Given these reasons, we aimed here to obtain the two *C. glabrata* genomes with a high resolution to identify all possible differences between the two genomes and to determine all possible adhesin and adhesin-like genes in the two genomes.

At the genomic level, DSY562 and DSY565 underwent several major chromosomal rearrangements. No data are currently available on PacBio-assembled *C. glabrata* genomes of sequential isolates, and it is therefore difficult to compare our results with other studies. One recent short report described the genomes of three pairs of *C. glabrata* isolates from patients with candidemia (Håvelsrud and Gaustad 2017). The genomes were assembled de novo using Illumina reads, thus resulting in different genome outputs (119–186 scaffolds). Comparison between the DSY genomes and the three published pairs using the same chromosomal fragments highlighted divergences as high as those between DSY isolates and CBS138. This underlines the high genetic diversity existing between *C. glabrata* isolates of different origins which has been observed by others using different approaches. One other recent study also applied Illumina-based de novo assembly of 33 genomes (Carreté et al. 2017) which identified 17 different chromosome patterns. Some patterns are similar to those observed in DSY562 and DSY565. While Chr D-L and Chr L inversions, as well as Chr I-L-D are among the common patterns, Chr F-L-I and Chr C inversions seem unique to DSY562 and DSY565. Another separate study established the genome of a *C. glabrata* isolate (CCTCC M202019) which is an industrial yeast strain widely used to produce α-oxocarboxylic acid (Xu et al. 2016). Genome data were established on the basis of a 150-bp, pair-end library sequenced by Illumina approaches and 111 contigs and 74 scaffolds were generated. Interestingly, this *C. glabrata* isolate was remarkably closely related to CBS138 (205 SNP differences) (Xu et al. 2016).

Taken together, these studies underline the high genetic diversity existing between *C. glabrata* isolates of different origins. It is likely that future studies will reveal additional patterns. In proportion to the size of their genomes, the number of changes between the DSY strains is rather small. We noticed only 17 nonsynonymous changes between these isolates (Table 7). Besides the known L280F change in *PDR1*, several other proteins with diverse functions were affected. The role played by these alterations is not clear at this time. One might speculate that they could contribute to the fitness of DSY565 or that they might be accidental. Other larger changes were detected between both strains (Table 8) and they occurred mainly in adhesin-like genes. This could reflect some adaptation due to host selection pressure.

It has been reported earlier that the mismatch-repair gene MSH2 can modulate drug resistance when its activity is compromised by specific SNPs, giving rise to a hyper-mutator phenotype (Healey et al. 2016b). Interestingly, DSY562 and DSY565 carry a V239L nonsynonymous mutation in MSH2, which is known to favor this hyper-mutator phenotype (Healey et al. 2016b). One might expect that this phenotype could generate genetic diversity within the context of host conditions. DSY565 was isolated with a time-lapse of 50 d in the oral cavity of a patient as compared to DSY562. Within this elapsed time, the MSH2 defect may lead to mutation accumulations. In a study probing the accumulation of mutations due to MSH2 defects in *S. cerevisiae* over 170 generations in vitro, mutation rates (mutations per bp per generation) were in the range of 5–9 × 10⁻⁸, which corresponded to 8–16 single base pair substitutions and 110–180 deletions/insertions in the
haploid *S. cerevisiae* genome (Lang et al. 2013). The number of generations that occurred within the time-lapse of DSY562 and DSY565 isolation may correspond to at least 250–500 generations, when taking into account 5–10 generations/d according to published data (Forche et al. 2009; Roetzer et al. 2011). Taking into account the data of Table 7 and Table 8, the number of SNPs and indels between DSY562 and DSY565 are in agreement with numbers reported in *S. cerevisiae*. One may expect that in *C. glabrata* the MSH2 defect may generate random mutations with a negative effect on *in vivo* fitness. The host conditions may serve as a selection pressure to eliminate progeny with fitness costs. We only sampled successful progenitors from the clinic, which of course strongly reduces the diversity of isolates for further analysis. This underpins that genetic diversity by the MSH2 defect may be much higher than observed from the collected clinical strains. The current literature on the effect of DNA mismatch-repair defects also highlighted that the occurrence of mutations is not random: DNA regions with a greater density of repeats are more mutable in mismatch-repair defective cells (Ma et al. 2012; Lang et al. 2013). In the context of adhesins, which are enriched in repetitive elements, MSH2 defects may perhaps contribute to increase the diversity of this gene family over generations. It has been observed that 44–50% of isolated isolates (*n = 625*) were carrying *MSH2* defects which thus indicates that the *MSH2* defect may confer some selective advantage in *C. glabrata* (Dellière et al. 2016; Healey et al. 2016a,b). In agreement with our hypothesis, it is interesting to mention that CBS138 and CCTCC M202019 do not exhibit *MSH2* defects and they contain a limited repertoire of adhesins as compared to DSY562/DSY565.

Besides the above-mentioned SNPs, the DSY562 and DSY565 genomes harbored transposon-like genes not reported previously in *C. glabrata*. These Ty-like retrotransposons genes (named here Tkp5-1 to -9) were most similar to the Tkp5 protein from *V. polyspora*, a yeast representing the postwhole-genome duplication lineage most divergent from *S. cerevisiae* (Scannell et al. 2007). We observed that, while Tkp5-2 and Tkp5-5 to -9 encoded the same proteins, Tkp5-3 and -4 were different (66% identity) but still similar to the Tkp5 protein from *V. polyspora*. Tkp5-1 shared the C-terminal ends of Tkp5-3 and -4. As expected from retrotransposons, Tkp5-2 and Tkp5-5 to -9 were flanked by the same LTRs, while Tkp5-3 and -4 were flanked by other distinct LTRs. Curiously, Tpk5-1 was located next to B1J91/B1J92_M05115g on Chr G, which is described as similar to Ty5-6, a *S. paradoxus* retrotransposon. Both genes are flanked by the same LTR. Taken together, these features suggest that DSY562 and DSY565 have acquired retrotransposons from different sources and that these strains have undergone several transposition events. Notably, DSY565 did not retain Tkp5-7, which is otherwise situated at the right arm of Chr M in DSY562. Given that a single LTR remained at this location in DSY565, this suggests an active transposition system. This retrotransposon was not identified in another location in DSY565.

We determined the presence of >100 adhesin-like genes in both DSY strains, which was not yet anticipated from other genome-wide studies. This number exceeds by far the numbers published for CBS138 (50–66, depending on the criteria used for selection) and therefore suggests that an expansion of this gene family occurred in our isolates. A genome report on a separate *C. glabrata* isolate closely related to CBS138 identified 49 adhesin-like genes (Xu et al. 2016). Since no equivalent *C. glabrata* genome assembly has been yet published using a PacBio approach, we still cannot confirm whether or not the investigated isolates constitute a unique case or that a high repertoire of adhesin genes (>100 genes) is shared in several *C. glabrata* clades. The number of identified adhesions between both isolates differs substantially (101 in DSY562 vs. 107 in DSY565). It is unlikely that this discrepancy could reflect an expansion of adhesins in DSY565 as compared to DSY562 in such a short time-lapse between sampling of isolates. We rather attribute this difference to technical issues during the assembly of PacBio reads. These reads produced small contigs, as illustrated in Figure S1, which are enriched in adhesin genes. These genes probably form part of chromosomal ends, but remain presently unassembled to major contigs.

To help distinguish the adhesins in DSY562 and DSY565, we used a combination of hierarchical clustering and mapping of PFAM and functional domains. With this approach, we could distinguish four major groups including the EPA-like genes, two VSHIT-containing groups, and another group with genes lacking EPA and VSHIT signatures. This grouping is less complex than proposed earlier using CBS138 data (seven clusters) (de Groot et al. 2008). The overlap between both grouping showed consistently that cluster I contained the EPA-like genes and that clusters V and VII and clusters II and III were grouped in two categories that include VSHT-containing adhesins. The other clusters (IV and VI) are located in distinct branches, the latter being grouped in genes lacking EPA and VSHIT signatures. It is presently difficult to validate one or the other categorization, since no systematic functional analysis of all adhesins has been yet carried out with the exception of EPA-like genes (Zupancic et al. 2008). Now that a more extensive repertoire of adhesins is available from our studies, such analysis may be undertaken in the future.

**ACKNOWLEDGMENTS**

The authors are thankful to F. Ischer for technical assistance and the LGIT (University of Lausanne) for technical support. This work was supported by the Swiss National Research Foundation grant no. 31003A_146936. The purchase of the Pacific Biosciences RSII instrument at the University of Lausanne was financed in part by the “Loterie Romande” through the “Fondation pour la Recherche en Médecine Génétique.”

**LITERATURE CITED**

Alexander, B. D., M. D. Johnson, C. D. Pfeiffer, J. Jimenez-Ortigosa, J. Catania et al., 2013 Increasing Echinocandin resistance in Candida glabrata: clinical failure correlates with presence of FKS mutations and elevated minimum inhibitory concentrations. Clin. Infect. Dis. 56: 1724–1732.

An, J., L. Lai, A. Sajjanhar, J. Batra, C. Wang et al., 2015 J-Circos: an interactive Circos plotter. Bioinformatics 31: 1463–1465.

Arendrup, M. C., and D. S. Perlin, 2014 Echinocandin resistance: an emerging clinical problem? Curr. Opin. Infect. Dis. 27: 484–492.

Bader, O., A. Schwarz, E. A. Kraneveld, M. Tangwattanchuleeporn, P. Schmidt et al., 2012 Gross karyotypic and phenotypic alterations among different progenies of the Candida glabrata CBS138/ATCC2001 reference strain. PLoS ONE 7: e52218.

Brown, G. D., D. W. Denning, N. A. B. Gower, S. M. Levitz, M. G. Netea et al., 2012 Hidden killers: human fungal infections. Sci. Transl. Med. 4: 165rv13.

Calabrese, D., J. Bille, and D. Sanglard, 2000 A novel multidrug efflux transporter gene of the major facilitator superfamily from Candida albicans (FLU1) conferring resistance to fluconazole. Microbiology 146: 2743–2754.

Carreté, L., E. Ksiezopolska, C. Pequeroles, E. Gómez-Molero, E. Saus et al., 2017 Patterns of genomic variation in the opportunistic pathogen Candida glabrata suggest the existence of mating and a secondary association to the human host. bioRxiv 105791 DOI: https://doi.org/10.1101/105791.

Chaudhuri, R., F. A. Ansari, M. V. Raghuwandanan, and S. Ramachandran, 2011 FungalRV: adhesin prediction and immunoinformatics portal for human fungal pathogens. BMC Genomics 12: 192.
