Chromosomal rearrangements and loss of subtelomeric adhesins linked to clade-specific phenotypes in *Candida auris*

José F. Muñoz[1], Rory M. Welsh[2], Terrance Shea[1], Dhwani Batra[3], Lalitha Gade[2], Anastasia P. Litvintseva[2], and Christina A. Cuomo[1]

[1] Broad Institute of MIT and Harvard, Cambridge, MA, USA; [2] Mycotic Diseases Branch, and [3] Division of Scientific Resources, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, GA, USA.

ABSTRACT

Candida auris is an emerging fungal pathogen of rising concern due to its increasing incidence, its ability to cause healthcare-associated outbreaks and antifungal resistance. Genomic analysis revealed that early cases of *C. auris* that were detected contemporaneously were geographically stratified into four major clades. Clade II, also termed East Asian clade, consists of the initial isolates described from cases of ear infection, is less frequently resistant to antifungal drugs and to date, the isolates from this group have not been associated with outbreaks. Here, we generate nearly complete genomes ("telomere-to-telomere") of an isolate of this clade and of the more widespread Clade IV. By comparing these to genome assemblies of the other two clades, we find that the Clade II genome appears highly rearranged, with 2 inversions and 9 translocations resulting in a substantially different karyotype. In addition, large subtelomeric regions have been lost from 10 of 14 chromosome ends in the Clade II genomes. We find that shorter telomeres and genome instability might be a consequence of a naturally occurring loss-of-function mutation in *DCC1* exclusively found in Clade II isolates, resulting in a hypermutator phenotype. We also determine that deleted subtelomeric regions might be linked to clade-specific adaptation as these regions are enriched in Hyr/Iff-like cell surface proteins, novel candidate cell surface proteins, and an ALS-like adhesin. The presence of these cell surface proteins in the clades responsible for global outbreaks causing invasive infections suggests an explanation for the different phenotypes observed between clades.

IMPORTANCE

Candida auris was unknown prior to 2009 and since then it has quickly spread around the world, causing outbreaks in healthcare facilities and representing a high fraction of candidemia cases in some regions. The emergence of *C. auris* is a major concern, since it is often multidrug-resistant, easily spread between patients, and causes invasive infections. While isolates from

three global clades cause invasive infections, isolates from Clade II primarily cause ear infections and have not been implicated in outbreaks, though cases of Clade II infections have been reported on different continents. Here, we describe genetic differences between Clade II and Clades I, III and IV, including a loss-of-function mutation in a gene associated with telomere length maintenance and genome stability, and the loss of cell wall proteins involved in adhesion and biofilm formation, that may suggest an explanation for the lower virulence and potential for transmission of Clade II isolates.

OBSERVATION

The emerging fungal pathogen *Candida auris* is increasingly reported as the cause of infections and has become a leading cause of invasive candidiasis in some hospitals, often in severely ill patients (1). *C. auris* isolates are commonly resistant to one or more antifungal drugs and can survive for long periods both in the clinical environment and as a commensal on skin (2). Initially identified in cases of ear infection in Japan and South Korea (3, 4), cases of systemic infection were soon after reported in India, South Africa, and Venezuela (5–7).

Initial genomic analysis of the outbreak identified four major genetic groups corresponding to these geographic regions or Clades I, II, III, and IV (8). Clades I, III, and IV are responsible for the ongoing and difficult to control outbreaks in healthcare facilities worldwide (9). Clade II, also termed the East Asia clade, is predominantly associated with cases of ear infection and appears to be less resistant to antifungals than other clades (10). While a reference genome assembly of a Clade I isolate is commonly used for SNP analyses, the karyotype is known to vary based on whole genome alignment with an assembly of a Clade III isolate (11) and wider analysis of chromosomal sizes (12). To better understand the emergence of this species and phenotypic differences between clades, here we leverage complete reference genomes for isolates from Clades II and IV. We find that the genome of Clade II is highly rearranged and is missing large subtelomeric regions that include candidate cell wall proteins found in all of the other three clades, which may help explain the major difference in clinical presentation between isolates from this clade and those from the global expanding clades causing outbreaks.

Large chromosomal rearrangements and deletions in *C. auris* Clade II

To investigate genomic differences between clades, we generated complete chromosome scale assemblies for isolates from Clades II and IV. Genome assemblies of B11245 (Clade IV) and B11220 (Clade II) consisted of 7 nuclear contigs corresponding to complete chromosomes with

telomeres at both ends, excluding one end that corresponds to rDNA in each assembly and one additional end in B11220 (**Supplementary Table 1**). While the number of chromosomes and total genome size is similar (average 12.3 Mb), chromosome lengths can differ by up to 1.1 Mb in Clade II relative to Clades I, III and IV (**Figure 1a**; **Supplementary Table 1**). By aligning these genomes with those previously published (11), we found evidence of large chromosomal rearrangements (> 10 kb) between the *C. auris* clades (**Figure 1b** to **d**; **Supplementary Table 1**). Notably, the genome of Clade II (B11220) is the most highly rearranged compared to the other three, with two inversions and 9 translocations resulting in large changes in chromosome size (**Figure 1b**). Fewer chromosomal rearrangements were detected in the most distantly related isolate Clade IV (B11245), including two inversions located in chromosome 1 (**Figure 1c**). The Clade III genome (B11221) contains one inversion in chromosome 1 and two translocations in chromosome 1 and 3 ((11); **Figure 1d**). These alterations in karyotype, most dramatically of Clade II, likely serve as a barrier to the production of viable progeny following mating and recombination.

In addition to rearrangements, we identified large genomic regions that were missing in one or more clades, predominantly subtelomeric regions that are absent in Clade II relative to Clades I, III and IV. Comparing the Clade I (B8441) and Clade II (B11220) genomes, we identified 11 large regions (>5 kb) absent in Clade II that encompassed 226 kb and 74 genes; 10 of these 11 regions are subtelomeric in B8441 (**Figure 1b**; **Supplementary Table 2**). Comparing with assemblies of Clade III (B11221) and Clade IV (B11245), we confirmed that these regions were also subtelomeric and only absent in Clade II. These subtelomeric deletions are a common feature of isolates from Clade II, as these regions are also absent in five other Clade II isolates from United States and South Korea, based on aligning whole genome sequence to B8441 (**Supplementary Table 3**). To search for genetic variation that could explain these dramatic changes in genome integrity in Clade II, we examined loss-of-function mutations found exclusively in Clade II isolates (**Supplementary Note**). We found that Clade II isolates have a loss-of-function mutation in *DCC1* (B9J08_000232), specifically a point mutation near the beginning of the protein (amino acid change=Y10*; codon change=taC/taG; **Supplementary Table 3**). As mutations in *DCC1* in *Saccharomyces cerevisiae* result in shorter telomeres (13) and genome instability (14), this suggests that this naturally occurring loss of function mutation in Clade II might contribute to shorter telomeres and genome rearrangements observed in this clade.

Depleted *Hyr/Iff* and species-specific cell-wall protein families in *C. auris* Clade II

The subtelomeric regions deleted in Clade II likely contribute to the phenotypic differences of this clade, most notably by the loss of fourteen candidate adhesins present in Clades I, III and IV. These include two sets of genes that contain predicted GPI anchors and secretion signals, one set sharing sequence similarity to *C. albicans* adhesins from the *Hyr/Iff* family and a second set of clustered genes only found in *C. auris* and the closely related species *C. haemulonii and C. duobushaemulonii* (**Figure 2a**; **Supplementary Table 2**). The *Hyr/Iff* gene family was previously noted to be the most highly enriched family in pathogenic *Candida* species and has been associated with pathogenicity and virulence (15). Six of eight *Hyr/Iff* proteins found in *C. auris* contain intergenic tandem repeats, which modulate adhesion and virulence (16); five *Hyr/Iff* genes are deleted in Clade II (**Figure 2a**; **Supplementary Table 2**). The second set, *C. auris*-specific candidate adhesins, are small proteins with serine/threonine-rich regions (9.0% to 17.3% Serine; 17.5% to 22.8% Threonine), and are tandemly located in subtelomeric regions conserved in Clades I, III and IV, but absent in Clade II (**Figure 2a** and **b**; **Supplementary Table 2**). The subtelomeric location and serine/threonine-rich region are properties shared with *C. glabrata EPA* adhesins (17), and the expansion of *EPA* adhesins is linked to the emergence of the ability to infect humans in the *C. glabrata* lineage (18). Several of the genes in deleted regions of Clade II (three of the *C. auris*-clade specific adhesins, one *HYR-like* gene, and other cell-wall associated proteins (*ALS4*, *CSA1* and *RBR3*)) were induced during developing *C. auris* biofilms ((19); **Supplementary Table 2**), suggesting they play a role in biofilm formation. The loss of adhesins-like genes in Clade II isolates could explain differences in infection and environmental phenotypes between this clade and the others that are more commonly observed.

Phenotypic differences between *C. auris* clades correlate with adherence ability

Candida auris can be easily transmitted within health care facilities, accelerated by the ability to persist on plastic surfaces common in health care settings (2), which may be enhanced by its ability to form biofilms. In addition, the ability of *C. auris* to form biofilms is associated with increased resistance to all classes of antifungals (19) and might also enhance *C. auris* capacity to colonize patients' skin further increasing patient to patient transmission and potentiating outbreaks. These are features of isolates from Clades I, III and IV that are the primary cause of invasive infections and are rarely reported from ear infections (10). The identification of these candidate cell wall proteins that are absent in the highly rearranged genome of Clade II, genomic changes likely due to a natural loss-of-function variant in *DCC1* in Clade II, highlights

the major differences that can occur between otherwise closely related isolates of a species. These candidate cell wall proteins are strong candidates to explain the different phenotypic properties of these clades, and further understand the emergence of this species.

Methods

DNA Purification

High molecular weight DNA for long read sequencing was obtained using the Epicentre MasterPure yeast DNA purification kit (MPY80200). DNA for Illumina sequencing was extracted using the ZYMO Research ZR Fungal/Bacterial DNA MiniPrep kit.

Genome assembly

Chromosome-level assemblies for Clade IV (South America) strain B11245 (CDC AR 0386) and Clade II (East Asia) strain B11220 (CDC AR 0381) were generated using an Oxford Nanopore Technology Ligation Sequencing Kit 1D (SQK-LSK108), sequenced on a MinIon Flow Cell R9.4 (FLO-MIN106) and basecalled with Albacore v2.0.2. The total read depth was 88X (21.2 kb N50) for B11245 and 51X (20.6 kb N50) for B11220. Reads were assembled using Canu v1.5 (genomeSize=12000000; stopOnReadQuality=false; correctedErrorRate=0.075) and Flye v 2.4.2 (genome-size=12000000) (20, 21). The most contiguous assembly was obtained with Canu for B11245 and Flye for B11220. A tandem motif (AGACACCACCTA{1,2}GAAA{1,2}CC{1,2}) was identified at contig ends; contig ends missing this motif were aligned to the unassembled contigs and manually extended. Each assembly had five iterations of Illumina read error correction using Pilon v1.12 (22). Assemblies were aligned to each other and to B8441 and B11221 (11) using NUCmer (MUMmer v3.22) (23), and rearrangement sites were manually inspected for read support. Gene annotation was performed using RNA-Seq to improve gene structure predictions as previously described ((11); **Supplementary Note**). The predicted gene number was highly similar across all *C. auris* genomes, totaling 5,328 for B11220 and 5,506 for B11245. GPI anchored proteins were predicted with PredGPI using the general model and selecting proteins with high probability (> 99.90% specificity) (24).

Genome alignments

Shared synteny regions of at least 10 kb were identified using NUCmer v3.22 (23). Chromosomal rearrangements (translocations, inversions and deletions) were identified from the alignments blocks based on alignment length, chromosome mapping, and orientation.

Illumina read alignments were manually inspected in Integrative Genomics Viewer (IGV) v2.3.72 (25) to confirm that the rearrangement junctions are well supported in each assembly. Illumina sequence of 5 additional clade II isolates (B11808, B11809, B12043, B12081, and B14308) were aligned to the B8441 genome using BWA mem v0.7.12 (26) and deleted regions identified using CNVnator v0.3 (1 kb windows; *p*-value < 0.01) (27). Variants were identified between these isolates using GATK v3.7 (28) (**Supplementary Note**).

Data and resource availability

The whole genome sequence and assemblies of B11220 and B11245 were deposited in NCBI under BioProject PRJNA328792. Illumina sequence of Clade II isolates is available in BioProject PRJNA328792. Isolates are available from the CDC and FDA Antimicrobial Resistance (AR) Isolate Bank, https://www.cdc.gov/drugresistance/resistance-bank/index.html.

Disclaimer

The use of product names in this manuscript does not imply their endorsement by the US Department of Health and Human Services. The finding and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Acknowledgements

This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under award U19AI110818 to the Broad Institute. CAC is a CIFAR fellow in the Fungal Kingdom Program.

197 References

198 1. van Schalkwyk E, Mpembe RS, Thomas J, Shuping L, Ismail H, Lowman W, Karstaedt AS,
199 Chibabhai V, Wadula J, Avenant T, Messina A, Govind CN, Moodley K, Dawood H,
200 Ramjathan P, Govender NP. 2019. Epidemiologic Shift in Candidemia Driven by *Candida*
201 *auris*, South Africa, 2016–2017. Emerging Infectious Diseases 25:early online.

202 2. Welsh RM, Bentz ML, Shams A, Houston H, Lyons A, Rose LJ, Litvintseva AP. 2017.
203 Survival, Persistence, and Isolation of the Emerging Multidrug-Resistant Pathogenic Yeast
204 *Candida auris* on a Plastic Health Care Surface. J Clin Microbiol 55:2996–3005.

205 3. Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. 2009. *Candida*
206 *auris* sp. nov., a novel ascomycetous yeast isolated from the external ear canal of an
207 inpatient in a Japanese hospital. Microbiol Immunol 53:41–44.

208 4. Kim M-N, Shin JH, Sung H, Lee K, Kim E-C, Ryoo N, Lee J-S, Jung S-I, Park KH, Kee SJ,
209 Kim SH, Shin MG, Suh SP, Ryang DW. 2009. *Candida haemulonii* and closely related
210 species at 5 university hospitals in Korea: identification, antifungal susceptibility, and clinical
211 features. Clin Infect Dis 48:e57-61.

212 5. Chowdhary A, Sharma C, Duggal S, Agarwal K, Prakash A, Singh PK, Jain S, Kathuria S,
213 Randhawa HS, Hagen F, Meis JF. 2013. New clonal strain of *Candida auris*, Delhi, India.
214 Emerging Infect Dis 19:1670–1673.

215 6. Magobo RE, Corcoran C, Seetharam S, Govender NP. 2014. *Candida auris*-associated
216 candidemia, South Africa. Emerging Infect Dis 20:1250–1251.

217 7. Calvo B, Melo ASA, Perozo-Mena A, Hernandez M, Francisco EC, Hagen F, Meis JF,
218 Colombo AL. 2016. First report of *Candida auris* in America: Clinical and microbiological
219 aspects of 18 episodes of candidemia. J Infect 73:369–374.

8. Lockhart SR, Etienne KA, Vallabhaneni S, Farooqi J, Chowdhary A, Govender NP, Colombo AL, Calvo B, Cuomo CA, Desjardins CA, Berkow EL, Castanheira M, Magobo RE, Jabeen K, Asghar RJ, Meis JF, Jackson B, Chiller T, Litvintseva AP. 2017. Simultaneous Emergence of Multidrug-Resistant *Candida auris* on 3 Continents Confirmed by Whole-Genome Sequencing and Epidemiological Analyses. Clin Infect Dis 64:134–140.

9. Chow NA, Gade L, Tsay SV, Forsberg K, Greenko JA, Southwick KL, Barrett PM, Kerins JL, Lockhart SR, Chiller TM, Litvintseva AP, US Candida auris Investigation Team. 2018. Multiple introductions and subsequent transmission of multidrug-resistant *Candida auris* in the USA: a molecular epidemiological survey. Lancet Infect Dis.

10. Welsh RM, Sexton DJ, Forsberg K, Vallabhaneni S, Litvintseva A. 2019. Insights into the Unique Nature of the East Asian Clade of the Emerging Pathogenic Yeast *Candida auris*. Journal of Clinical Microbiology 57:e00007-19.

11. Muñoz JF, Gade L, Chow NA, Loparev VN, Juieng P, Berkow EL, Farrer RA, Litvintseva AP, Cuomo CA. 2018. Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species. Nat Commun 9:5346.

12. Bravo Ruiz G, Ross ZK, Holmes E, Schelenz S, Gow NAR, Lorenz A. 2019. Rapid and extensive karyotype diversification in haploid clinical *Candida auris* isolates. Curr Genet.

13. Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, Coker C, Krauskopf A, Kupiec M, McEachern MJ. 2004. A genome-wide screen for *Saccharomyces cerevisiae* deletion mutants that affect telomere length. Proc Natl Acad Sci USA 101:8658–8663.

14. Yuen KWY, Warren CD, Chen O, Kwok T, Hieter P, Spencer FA. 2007. Systematic genome instability screens in yeast and their potential relevance to cancer. Proc Natl Acad Sci USA 104:3925–3930.

15. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJ, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PW, Harris D, Hoyer LL, Hube B, Klis FM, Kodira C, Lennard N, Logue ME, Martin R, Neiman AM, Nikolaou E, Quail MA, Quinn J, Santos MC, Schmitzberger FF, Sherlock G, Shah P, Silverstein KA, Skrzypek MS, Soll D, Staggs R, Stansfield I, Stumpf MP, Sudbery PE, Srikantha T, Zeng Q, Berman J, Berriman M, Heitman J, Gow NA, Lorenz MC, Birren BW, Kellis M, Cuomo CA. 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459:657–62.

16. Bates S, de la Rosa JM, MacCallum DM, Brown AJ, Gow NA, Odds FC. 2007. *Candida albicans* Iff11, a secreted protein required for cell wall structure and virulence. Infect Immun 75:2922–8.

17. De Las Peñas A, Pan S-J, Castaño I, Alder J, Cregg R, Cormack BP. 2003. Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to RAP1- and SIR-dependent transcriptional silencing. Genes Dev 17:2245–2258.

18. Gabaldón T, Martin T, Marcet-Houben M, Durrens P, Bolotin-Fukuhara M, Lespinet O, Arnaise S, Boisnard S, Aguileta G, Atanasova R, Bouchier C, Couloux A, Creno S, Almeida Cruz J, Devillers H, Enache-Angoulvant A, Guitard J, Jaouen L, Ma L, Marck C, Neuvéglise C, Pelletier E, Pinard A, Poulain J, Recoquillay J, Westhof E, Wincker P, Dujon B, Hennequin C, Fairhead C. 2013. Comparative genomics of emerging pathogens in the Candida glabrata clade. BMC Genomics 14:623.

19. Kean R, Delaney C, Sherry L, Borman A, Johnson EM, Richardson MD, Rautemaa-Richardson R, Williams C, Ramage G. 2018. Transcriptome Assembly and Profiling of

Candida auris Reveals Novel Insights into Biofilm-Mediated Resistance. mSphere 3:e00334-18.

20. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736.

21. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546.

22. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one 9:e112963.

23. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12.

24. Pierleoni A, Martelli PL, Casadio R. 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics 9:392.

25. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nature Biotechnology 29:24–26.

26. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio].

27. Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 21:974–984.

28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–303.

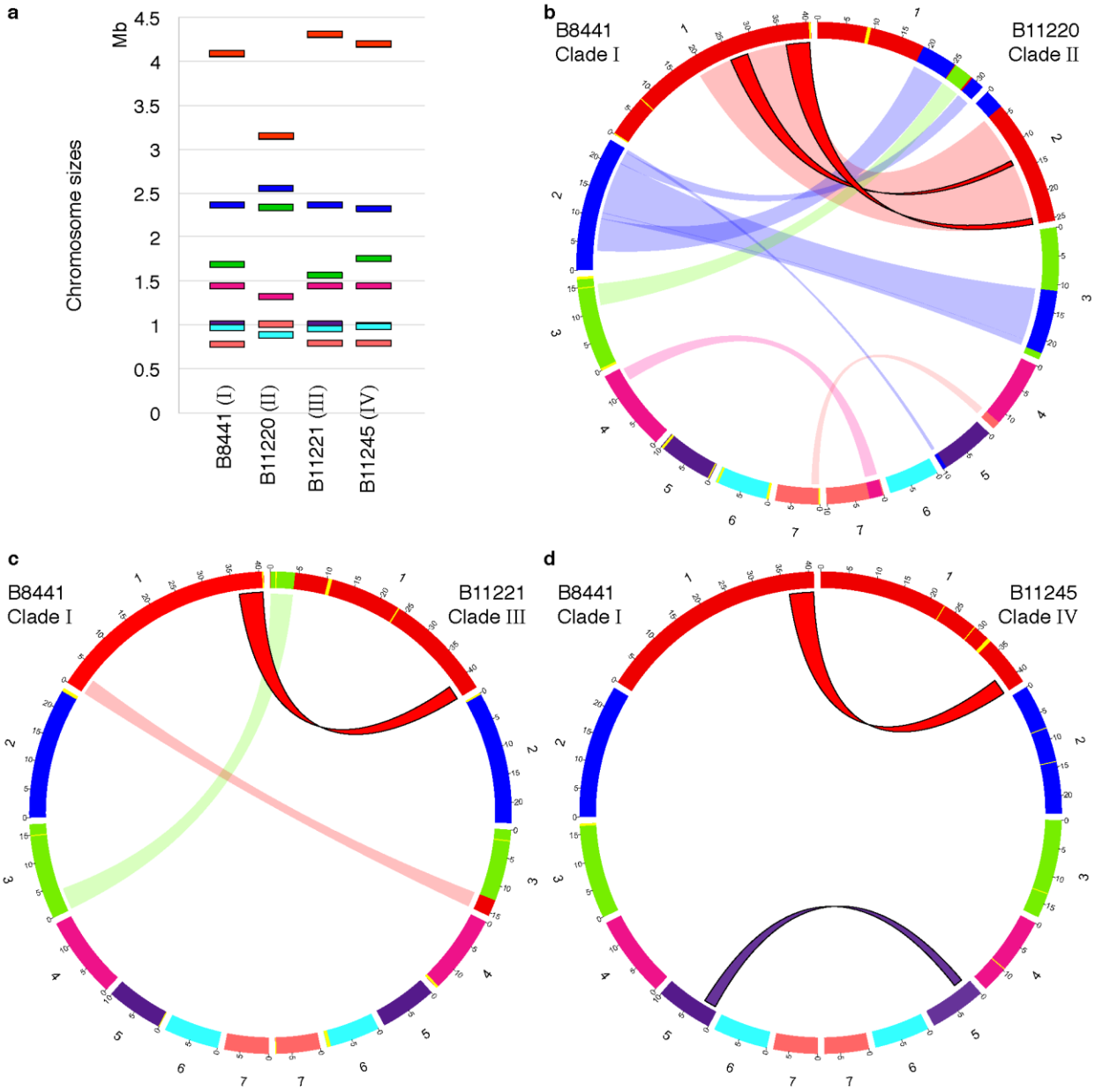

Figure 1. Karyotypic variation and chromosomal rearrangements in *Candida auris*. (a) Chromosome sizes are based on complete chromosome scale contigs. **(b-d)** Circos plots showing syntenic chromosomes by color and links for inversions (twisted filled links) and translocations (flat transparent links) using B8441 (Clade I) as reference compared to B11220 (Clade II; **b**), B11221 (Clade III; **c**), and B11245 (Clade IV; **d**). Scaffolds/contigs to chromosome mapping for these genome assemblies is included in **Supplementary Table 4**.

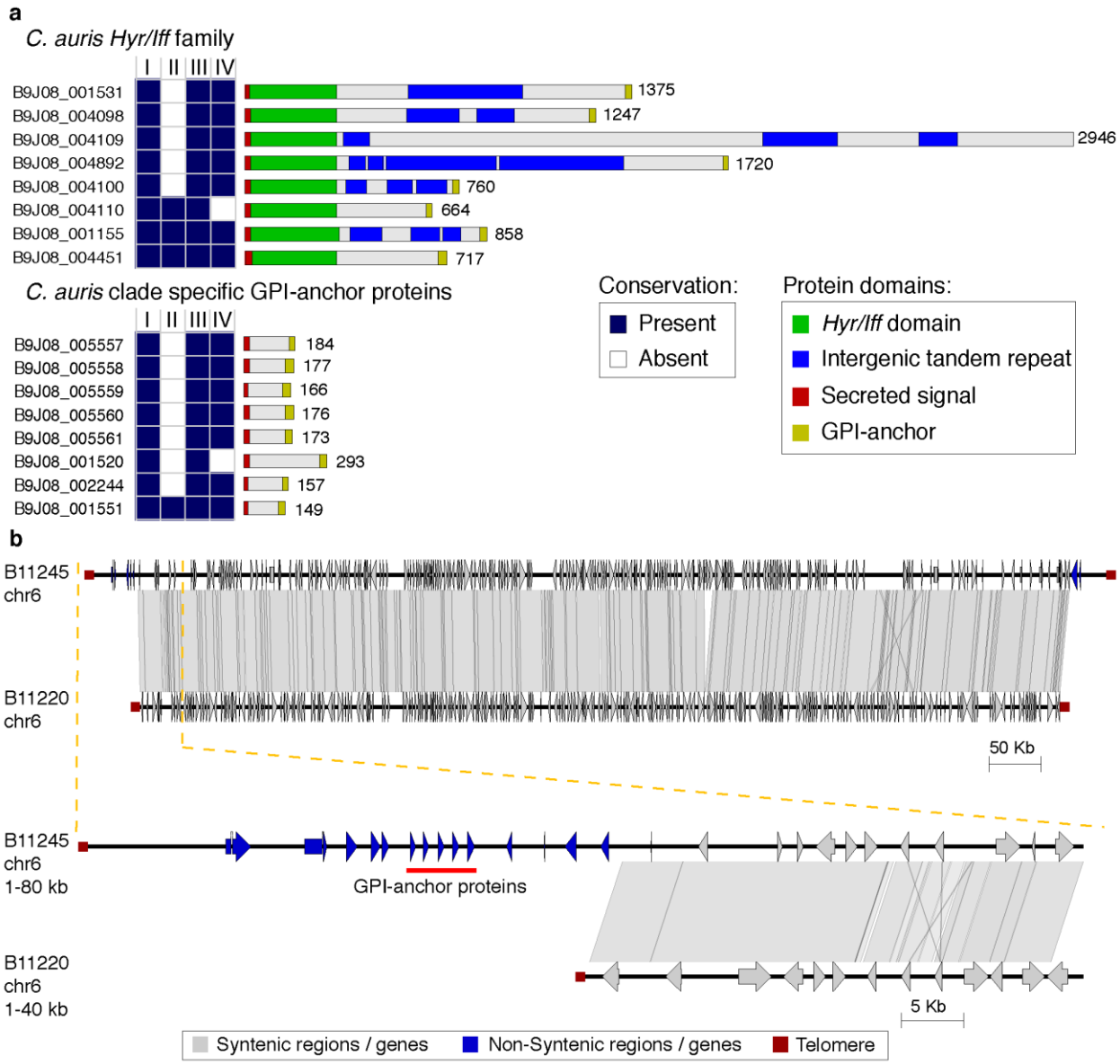

Figure 2. Differences in the repertoire GPI-anchor proteins in *Candida auris*. (a) Conserved domains in two clusters of GPI-anchor found in *C. auris* B8441. (top) *Hyr/Iff* GPI-anchor family and (bottom) *C. auris* clade-specific GPI-anchor protein. Conservation across *C. auris* strains representing Clades I, II, III and IV is color-coded indicating whether the gene is present (dark blue) or absent (white). **(b)** Chromosome wide synteny between B11245 (Clade IV) and B11220 (Clade II). Chromosome 6 includes telomeres at both ends in both strains (dark red square). Shared synteny regions based on genome alignment (blastn) are depicted in gray vertical blocks connecting the chromosome regions. Depicted genes in blue and light gray arrows showing the direction of transcription are color-coded according to the location in conserved (gray) or non-conserved (blue) regions. The top comparison corresponds to the entire chromosome 6, and the bottom comparison corresponds to a zoom in of the subtelomeric region depleted in Clade II (B11220), which encompasses *auris*-clade specific adhesins in tandem (red line).