Exposure to the oral host niche yields rapid phenotypic and genotypic diversification in *Candida albicans*

Forche, A.[1*], Cromie, G.[2], Gerstein, A.C.[3], Solis, N. V.[4], Pisithkul, Z.[1], Srifa, W.[1], Jeffery, E.[2], Filler, S.G.[4], Dudley, A.M.[2], Berman, J.[5]

[1]Department of Biology, Bowdoin College, Brunswick, ME USA

[2]Pacific Northwest Research Institute, Seattle, WA, USA

[3]Department of Microbiology & Immunology, University of Minnesota, Minneapolis MN USA

[4]Division of Infectious Diseases, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA USA; Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA USA

[5]School of Molecular Cell Biology & Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

ORCID for AMD 0000-0003-3644-0625

ORCID for AF 0000-0002-3004-5176

Keywords: *Candida albicans*, aneuploidy, loss of heterozygosity, oropharyngeal Candidiasis, hypervariability, colony phenotype

*To whom correspondence should be address:

Anja Forche PhD

Research Assistant Professor

Bowdoin College

Department of Biology

6500 College Station

Brunswick, ME 04011

27 USA

28 Email: aforche@bowdoin.edu

29 Phone: 1-207-725-3365

30

Abstract

In vitro studies suggest that stress may generate random standing variation, and that different cellular and ploidy states may evolve more rapidly under stress. Yet this idea has not been tested with pathogenic fungi growing within their host niche *in vivo*. Here, we analyzed the generation of both genotypic and phenotypic diversity during exposure of *Candida albicans* to the mouse oral cavity. Ploidy, aneuploidy, loss of heterozygosity (LOH) and recombination were determined using flow cytometry and ddRADseq. Colony phenotypic changes (CPs) in size and filamentous growth were evident without selection, and were enriched among colonies selected for LOH of the *GAL1* marker. Aneuploidy and LOH occurred on all chromosomes (Chrs), with aneuploidy more frequent for smaller Chrs and whole Chr LOH more frequent for larger Chrs. Large genome shifts in ploidy to haploidy often maintained one or more heterozygous disomic Chrs, consistent with random Chr missegregation events. Most isolates displayed several different types of genomic changes, suggesting that the oral environment rapidly generates diversity *de novo*. In sharp contrast, following *in vitro* propagation isolates were not enriched for multiple LOH events, except in those that underwent haploidization and/or had high levels of Chr loss. The frequency of events was overall 100 times higher for *C. albicans* populations following *in vivo* passage compared to *in vitro*. These hyperdiverse *in vivo* isolates likely provide *C. albicans* with the ability to adapt rapidly to the diversity of stress environments it encounters inside the host.

Author summary

Adaption is a continuous dynamic process that requires genotypic and phenotypic variation. Here we studied the effects of a single passage in a mouse oropharyngeal model of infection on the appearance of diversity in *C. albicans,* a common commensal of the human oral cavity and GI tract. We found that variation could be rapidly detected following oral colonization,

with the frequency of genome change being considerably higher with pre-selection for recombination and colony phenotypic changes. Importantly, one third of all isolates had multiple genome changes, significantly higher than expected by chance alone. We suggest that some cells in the population are naturally hypervariable and that they are a major source of diversity upon which selection can act in stressful conditions *in vivo* and *in vitro*.

Introduction

Candida albicans is a common commensal of the human GI tract and the oral cavity in healthy individuals, and also an opportunistic pathogen, especially in immunocompromised patients (CALDERONE 2012). In healthy people, the fungus is prevented from causing disease by the resident microbiota and the host immune system (LORENZ *et al.* 2004; RICHARDSON AND RAUTEMAA 2009). However, immune deficiencies or a minor imbalance of the microbiota (e.g., through administration of antibiotics) can be sufficient to cause superficial infections. During the course of infection, *C. albicans* encounters many different host environments to which it must adapt rapidly. Furthermore, it must cope with environmental fluctuations in established niches during long-term persistence in the host (STAIB *et al.* 2001). Determining the genetic and phenotypic changes that accompany the establishment of commensalism and the transition to pathogenicity (and hence how they can be prevented) is not known (NAGLIK *et al.* 2003; WILSON *et al.* 2009).

Several studies strongly suggest that *C. albicans* may have a very different arsenal of adaptation mechanisms when in direct contact with the host compared to laboratory conditions. For example, a novel cell phenotype (GUT) is unique to the commensal environment of the gastrointestinal tract (PANDE *et al.* 2013) and several genes (e.g., Cph2, Tec1) are specifically expressed under commensal conditions (ROSENBACH *et al.* 2010). Furthermore, cells in the commensal state express genes that suggest the presence of at least two sub-populations of exponentially growing cells alongside stationary-phase cells. In addition, the expression patterns

of several genes are clearly distinct during growth *in vivo* vs *in vitro* (NOBILE *et al.* 2006; VANDEPUTTE *et al.* 2011; FANNING *et al.* 2012; LOHBERGER *et al.* 2014).

Recent studies in fungi found that genome instability caused by large-scale chromosomal changes, including gross chromosomal rearrangements (GCRs), supernumerary Chrs (SNCs), aneuploidy and loss of heterozygosity (LOH), are more frequent under stress conditions *in vitro* and *in vivo* (RUSTCHENKO *et al.* 1997; SELMECKI *et al.* 2006; COYLE AND KROLL 2008; POLAKOVA *et al.* 2009; FORCHE *et al.* 2009a; FORCHE *et al.* 2011; HICKMAN *et al.* 2013). Aneuploidy in particular has been shown to be one of the mechanisms that can lead to antifungal drug resistance in pathogenic fungi including *Crypococcus neoformans* and *Candida glabrata* (SELMECKI *et al.* 2006; POLAKOVA *et al.* 2009; SIONOV *et al.* 2010). Interestingly, a recent study showed that *C. albicans* forms very large cells in response to acute micronutrient limitation, in particular to zinc. Cell size has been shown to be correlated with ploidy (HICKMAN *et al.* 2013) and indeed flow cytometry data support that these gigantic cells may be aneuploid (MALAVIA *et al.* 2017). In *C. neoformans,* Titan cells are large polyploid cells that can rapidly produce drug resistant aneuploid daughters upon exposure to the drug fluconazole (GERSTEIN *et al.* 2015), supporting the idea that aneuploidy is a common adaptation mechanism of pathogenic fungi.

Our previous study of ~80 *C. albicans* isolates recovered from the mouse model of hematogenously disseminated candidiasis (BSI) from mice kidneys (FORCHE *et al.* 2003; FORCHE *et al.* 2009a) provided a first glimpse into the types of genomic changes that *C. albicans* undergoes at the population level. We discovered higher rates of phenotypic and Chr-level genetic variation following passage of *C. albicans in vivo* relative to passage *in vitro*. In addition, missegregation events, including whole Chr aneuploidy and LOH, were positively associated with altered CPs.

The oral cavity is one of the few host niches that is both a commensal and pathogenic niche (VARGAS AND JOLY 2002; PATIL *et al.* 2015). *C. albicans* has been found as part of the commensal microflora in up to two thirds of the healthy population (VILLAR AND DONGARI-BAGTZOGLOU 2008; PANKHURST 2013). Oral and oropharyngeal candidiasis can develop as consequence of developed immunodeficiency (e.g. HIV/AIDS), underlying diseases such as diabetes, and treatment with broad-spectrum antibiotic, corticosteroids and chemotherapy (SOBUE *et al.* ; LYON *et al.* 2006; LU *et al.* 2017). In the oral niche, fungal-host interactions are highly dynamic due to a multitude of factors including the presence of antimicrobial salivary peptides and the microbiota of bacterial and fungal species that co-exist and compete for nutrients on epithelial cells (DEMUYSER *et al.* 2014; JAKUBOVICS 2015) and the highly fluctuating environmental conditions (e.g., temperature, pH) (PARK *et al.* 2009). Unexpectedly, we recently identified haploid, mating-competent *C. albicans* isolates for the first time, and most of these haploids were recovered after *in vivo* passage in an oral model of infection (HICKMAN *et al.* 2013). This extraordinary finding highlights the contribution of and the need for *in vivo* studies to the discovery of novel aspects of *Candida* biology in general and of host-pathogen interactions in particular.

To further our understanding on the acquisition of standing variation of *C. albicans* during infection, we performed experimental evolution of *Candida albicans* in mice to analyze the appearance of genotypic and phenotypic diversity during passage through the mouse oral cavity for 1, 2, 3 or 5 days using an oropharyngeal model of infection (KAMAI *et al.* 2001; SOLIS AND FILLER 2012). We found that diversity is rapidly generated after exposure to the oral host niche, and that many of these changes are identified in multiple mice. The overall high within-mouse diversity and multiple changes per isolate was high independent of the duration of infection. Surprisingly, the generation of multiple genetic changes in a single isolate appears to occur with higher frequency than would be expected by random chance alone. Taken together,

our results suggest that exposure to the host (and/or the transition from *in vitro* to *in vivo* growth conditions) generates highly variable isolates at a frequency 2 orders of magnitude higher than *in vitro*.

Methods

Isolate maintenance and DNA extraction

Strains used to generate parental strain YJB9318 are listed on Table S1. YJB9318 and recovered isolates were grown on YPD (2% glucose, 1% yeast extract, 1% bacto peptone, 20 mg/L uridine with 1.5% agar added for plate cultures). Gal phenotypes were assessed on MIN-Gal (0.67% yeast nitrogen base without amino acids, 2% galactose, 1.5% agar; only Gal^+ isolates grow) and 2-deoxygalactose medium (2DOG; 0.1% 2-deoxygalactose, 0.5% raffinose, glycerol, 1.5% agar; only Gal^- isolates grow). All isolates are stored long-term in 50% glycerol at -80°C. DNA extractions were performed as described previously (SELMECKI *et al.* 2005).

Construction of strain YJB9318

Plasmids and primers used in this study are listed in Table S1. YJB9318 is a derivative of strain RM1000 #2 (Table S1) in which one copy of *GAL1* was replaced with *URA3* (*GAL1/*Δ*gal1*::*URA3)*. First, the *URA3* marker was amplified from plasmid p1374 with primers 1672 and 1673 (Table S1), and transformed into isolate YJB7617 (RM1000#2) replacing one copy of *GAL1* (YJB8742) (LEGRAND *et al.* 2008). Correct disruption of *GAL1* was confirmed by diagnostic PCR using primers 1674 and 1675 (Table S1). To make YJB9318 prototrophic, *HIS1* was reintroduced into strain YJB8742 by transforming with plasmid p1375 (pGEM-*HIS1*) that was cut with restriction enzyme *Nru*I. Diagnostic PCR with primers 728 and 565 (Table S1) confirmed correct integration of *HIS1* at its native locus. To ensure that transformation did not cause any genomic changes to the parental strain, single nucleotide polymorphism (SNP) microarrays and SNP/Comparative genome hybridization arrays (SNP/CGH) were performed as

158 described previously (data not shown) (SELMECKI *et al.* 2005; FORCHE *et al.* 2009a; ABBEY *et al.*
159 2011).

160 *PCR conditions for transformation and diagnostic PCR*

161 PCRs for transformation were performed in a total volume of 50 µl with10 mM Tris–HCl (pH
162 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each dATP, dCTP, dGTP, and dTTP, 2.5 U rTaq
163 polymerase (TAKARA), 4 µl of 10 µM stock solution of each primer, and 1.0 µl of template
164 (p1374). PCRs were carried out for 34 cycles as followed: initial denaturation step for 5 min at
165 94°C, denaturation step for 1 min at 94°C, primer annealing step for 30 s at 55°C, extension
166 step for 1 min at 72°C, and a final extension step for 10 min at 72°C. Each PCR product was
167 checked by gel electrophoresis for the amplification of the desired PCR fragment. PCR products
168 were purified using ethanol precipitation.

169 Diagnostic PCR was performed in a final volume of 25 µl with 10 mM Tris–HCl (pH 8.0),
170 50 mM KCl, 1.5 mM $MgCl_2$, 100 µM each dATP,dCTP, dGTP, and dTTP, 2.5 U rTaq
171 polymerase, 2 µl of 10 µM stock solution of each primer, and 2.5 µl genomic DNA. PCRs were
172 carried out for 30 cycles as followed: initial denaturation for 3 min at 94°C, denaturation step or
173 1 min at 94°C, primer annealing step for 30 s at 55°C, extension step for 1 min at 72°C, and a
174 final extension step for 5 min at 72°C. Five microliters of PCR product was run on a 1% agarose
175 gel to verify that the fragment was of the appropriate size.

176

177 *Model of oropharyngeal Candidiasis (OPC)*

178 The OPC model was essentially performed as described previously (SOLIS AND FILLER 2012).
179 Briefly, male BALB/c mice (21-25 g; Taconic Farms) were immune-suppressed with cortisone
180 acetate (225 mg/kg, Sigma) on days -1, 1, and 3 of infection. For inoculum preparation, strain
181 YJB9318 was grown in MIN-Gal medium to ensure that no Gal^- cells arose prior infection. A
182 total of twenty mice were infected with 1×10^6 cells of strain YJB9318 (Table1). Of these, 17
183 mice survived to the scheduled dates of sacrifice. On days 1, 2, 3, and 5 post-infection (FIG.1,

FIG.S1), 4-5 mice were euthanized. The tongues were extracted, weighted, and homogenized. Next, appropriate dilutions were spread onto YPD agar plates for total CFU counts and onto 2DOG agar plates to determine the number of Gal^- cells. Recovered isolates were directly picked from the original YPD and 2DOG plates to 96well plates with 50% glycerol and stored at -80°C to avoid any changes to the isolates not acquired during *in vivo* passage.

To confirm the Gal status of recovered isolates, they were grown overnight in deep 96-well plates containing 300 µl YPD broth. Cultures were washed once with distilled water, 5 µl of each culture were spotted onto 150 x 100 mm YPD plates, MIN-Gal and 2DOG medium, and plates were incubated for 2 days at 30°C to assess growth. The frequency of LOH at the *GAL1* locus was determined using the ratio of total isolates recovered (CFUs on YPD) divided by the total number of $2DOG^R$ isolates. The *in vitro* frequency of *GAL1* loss in strain YJB9318 was measured as described previously (FORCHE *et al.* 2009).

Assessment of colony phenotypes (CPs) and selection of isolates for genotypic analysis

Previously, we showed that isolates with missegregation events (whole Chr aneuploidy and whole Chr LOH) exhibited CPs consistent with slow growth and abnormal filamentous growth (FORCHE *et al.* 2005; FORCHE *et al.* 2009). To increase the ability to identify genotypic changes, we plated all isolates for CPs on YPD at 30ºC and scored single colonies after 3 days. CPs were determined for colony diameter (smaller or larger than parental strain, first number) and filamentous growth (degree of wrinkling compared to parental strain, second number) resulting in a binary code for each of the 7 unique CPs (see FIG.2D for representative images). For further genotypic analysis, all isolates with altered CPs (6 Gal^+ and 158 Gal^-), and isolates with parental CP (148 Gal^+ and 116 Gal^-) from a total of 17 mice were chosen to yield a set of 429 isolates (FIG.S1, Table S2).

Determination of ploidy by flow cytometry

Ploidy of all recovered isolates was determined as described previously (ABBEY *et al.* 2011). Briefly, each isolate was streaked out to single colony onto YPD plates and incubated for 3 days at 30°C. Single colonies were transferred to deep 96-well plates containing 0.6 ml of YPD and cultures were grown overnight (16 hrs) at 300 rpm to stationary phase. Fifty microliters were transferred to new deep 96-well plates containing 250 µl YPD broth, and cultures were grown for 6 hrs at 30°C at 300 rpm. Two hundred fifty microliters of culture was transferred to round bottom 96-well plates, cells were spun down at 1,000 rpm, and resuspended in 20 µl of 50:50 buffer (50 mM Tris HCl, pH 8.0, 50 mM EDTA, pH 8.0). To fix cells, 180 µl of 95% ethanol was added to each well. Cells were treated with 0.1 µg/ml RNase (1 hr at 37°C) and 5 mg/ml Proteinase K (30 min at 37°C) followed by staining with SybrGreen for 1hr in the dark. After a final wash in 50:50 buffer, cells were resuspended in 50:50 buffer and run on a flow cytometer (FACSCALIBUR). A customized MATLAB script was used to calculate ploidy for each isolate using a diploid and a tetraploid isolate as controls (ABBEY *et al.* 2011).

Whole genome karyotyping using double digest restriction site associated DNA sequencing (ddRADseq)

ddRADseq was carried out using the restriction enzymes *Mfe*I and *Mbo*I (LUDLOW *et al.* 2013). For each lane of Illumina sequencing (up to 576 isolates/lane), raw read sequences were split into isolate-specific pools based on their associated 6 bp TruSeq multiplex and 4 bp inline barcode sequences, allowing 1 mismatch in the i7 barcodes and no mismatches in the inline barcodes. A minimum barcode quality of Phred = 20 was applied to all bases of the inline barcode. Reads were then aligned to the *C. albicans* reference (SC5314 v. A21-s02-m04-r01) using BWA (v.0.7.5) allowing 6 mismatches and quality trimming using the parameter -q 20. The SAMtools (v.0.1.17) (LI *et al.* 2009) mpileup command was then used to create a pileup file for each isolate, using the -q 20 and –C 50 parameters. From the pileup file, the count of all observed bases at each covered reference position was calculated.

Ploidy Estimation from ddRADseq

From the aligned-read SAM file, the position of the *Mfe*I end of each read was determined (5' end of forward reads, 3' end of reverse reads). Only reads with a Phred-scaled mapping alignment quality of at least 20 were considered. The occurrence of each end position was then counted, resulting in a set of marker positions for each isolate along with the number of reads aligning to each of those positions. For all isolates in a sequencing run, a matrix of observed read counts at all positions was generated. Because of the long tail of infrequently observed sites, counts of 1 were treated as counts of zero. Only positions counted in at least one isolate were retained. The matrix of counts was then edited to remove any isolates with ≥ 20% positions with 0 counts. Following this, marker positions were filtered to only include those occurring in >80% of remaining isolates. The edited count matrix was used to calculate relative ploidy at each marker position as follows. Each isolate was normalized for depth of sequencing by dividing all observed counts by the median value of all counts > 0. To control for marker-to-marker variation in coverage (largely due to the size of the associated DNA fragment), normalized coverage at each marker in each isolate was divided by the median coverage (ignoring zero values) of that marker across a set of control euploid isolates. Before plotting, globally noisy markers, with standard deviations >0.5 across all isolates (not including zero values), were removed and for each isolate a minimum raw count coverage of 10 or 20 was required at each position to remove low-confidence estimates for that isolate.

Ploidy was also assessed on a by-Chr basis. The unedited, original counts matrix was used to calculate the proportion of all reads in each isolate aligning to each Chr. The value for each Chr was then normalized by dividing the median value for that Chr across the set of euploid control isolates. To account for variation in genome size in aneuploids, values for each isolate were then further normalized by dividing each Chr value by the median Chr value for that isolate. For euploid isolates this should produce values of ~1 across all Chr. A diploid with one

trisomic Chr would have a value of ~1.5 for one Chr and 1 for the rest. On the assumption that most isolates are essentially diploid, values were converted into ploidy by multiplying by 2, 3 or 4 in the small number of isolates where this produced Chr copy number values closer to whole numbers.

Estimation of Allele Ratios

Heterozygosity was assessed at all heterozygous sites in parental strain SC5314 (MUZZEY *et al.* 2013) ignoring indels. For each Chr, SC5314 heterozygous sites were identified by aligning the two phased haploytpes ("A" and "B") from (MUZZEY *et al.* 2013) using the Mummer (v3.22) nucmer command with the parameters -c 100 -l 10 -b 200. Alignments were filtered using the delta-filter command with the –g parameter and snps were then called using the show-snps command with the parameters -r -C -H –T. the counts of the two expected alleles were extracted from the count of all observed alleles (as above) in each isolate for every expected het site covered by reads. The binomial probability of the observed counts was then calculated using models 1A:0B (homozygous A), 3A:1B, 2A:1B, 1A:1B, 1A:2B, 1A:3B and 0A:1B (homozygous B). For example for model 3A:1B the binomial probability of the observed data would be calculated based on P(A) = 0.75, P(B) = 0.25. For the homozygous models, observed counts of the "wrong" allele were assumed to be errors with a probability of 0.01, with the expected allele having a probability of 0.99. For each site, the set of possible models was then consolidated based on the copy number of the Chr, identified from the whole Chr read proportion analysis above. For disomic Chrs, models 1A:0B, 1A:1B and 0A:1B were compared and the best and second best models identified. For trisomic Chrs, the best and second best models were identified from models 1A:0B, 2A:1B, 1A:2B and 0A:1B. For tetrasomes, the models compared were 1A:0B, 3A:1B, 1A:1B, 1A:3B and 0A:1B. After identification of the best model, for each site in each isolate a LOD score (Log10 P (Best Model)/P(Second Best Model) was then calculated. For each dataset, markers were removed unless they were classified as

heterozygous (best model = 1A:1B with LOD > 1) in at least one isolate. For visualization, a median sliding window of size 7 was applied to the best model values, ordered by genome position.

Population frequency of changes

To obtain estimates of the population frequency of CPs and genotypic changes (i.e., not just for the analyzed isolates), we calculated the frequencies of phenotypic and genotypic changes for Gal+ and Gal- isolates based on the total number of *C. albicans* CFUs from the experiment (8.52×10^5), the number of Gal^+ isolates (8.51×10^5; 99.89% of the total), and the number of Gal^- cells (9.1×10^2; 0.11%). Extrapolation showed that the frequency of Gal^++ CP in the total population was 3.9×10^{-2} (~1/400 cells) and ~5.8×10^{-4} for Gal^-+CP (~6/10,000 cells) (FIG.S1). The frequency of genotypic changes (Chr1 changes excluded) for Gal^+ cells was 13×10^{-2} and 5.2×10^{-4} for Gal^- cells. (FIG.S1).

A customized MATLAB script was written to calculate the total likelihood for all the permutations for each number of genotypic changes. A chi square goodness-of-fit test was applied to ask whether the difference between the expected and observed numbers was due to sampling variation, or whether it was a real difference. A p-value of ≤ 0.05 was considered significant.

Diversity index

ddRADseq data was used to determine the number of unique karyotypes that were present in each mouse and the number of colonies that exhibited each karyotype. A karyotype was considered unique if there was either a unique whole or partial Chr aneuploidy or LOH event and/or there was variation in the LOH breakpoint, compared to other colonies isolated from the same mouse. Diversity was then calculated as Simpsons index ($1- \sum p_i^2$) (SIMPSON 1949), where

313 p_i is the proportional abundance of each colony type (note that if there is only one karyotype,
314 diversity is 0).
315
316 **Results**
317 To analyze diversification rates of *C. albicans* on a mucosal surface, we seeded the oral cavity
318 of 20 corticosteroid-treated mice with 10^6 cells originating from YJB9318, a single, colony-
319 purified *C. albicans* strain that was heterozygous for *GAL1* (GORMAN *et al.* 1992). Groups of 4-5
320 mice were sacrificed on days 1, 2, 3 and 5 and isolates were recovered from tongue
321 homogenates (FIG. 1 and S1). Initiating the experiment with a Gal+/- strain enabled us to
322 acquire evolved isolates from both YPD plates (Gal^+, 541 colony isolates, no selection, unknown
323 genomic changes) and 2-deoxygalactose (2DOG) plates (Gal^-, 360 colony isolates, minimum
324 genomic change of LOH at the *GAL1* locus) (FIG. 2A and S1). All recovered isolates were first
325 screened for changes in colony phenotype (CPs) and we identified 7 distinct CPs, 3 of them
326 were also detected among Gal^+ isolates and all 7 were detected among Gal^- isolates (FIG.S1,
327 Fig.2, and see below). As measured by flow cytometry, the majority (72%) of isolates retained a
328 diploid genome content (FIG. 2B). Strikingly, eleven isolates had haploid or near-haploid
329 genome (2%) content. An additional seven isolates were tetraploid or near-tetraploid while the
330 remainder (26%) had ploidy values consistent with aneuploid diploids (Table S2).
331
332 Twenty-four hours after infection, the oral fungal burden was approximately 10^2 CFUs
333 per g tissue, suggesting that only a small proportion of the starting inoculum initiated the oral
334 infections (FIG. 2A). The number of CFUs generally increased with time of infection and the
335 proportion with CPs increased slightly (Table 1). The proportion of $Gal1^-$ CFU increased
336 proportional with the total CFU (measured by comparing the frequency of $2DOG^R$ isolates and
337 YPD isolates). The overall frequency of LOH at *GAL1* was two orders of magnitude higher *in*

338 *vivo* compared to *in vitro* (FIG.2A, S2). Ploidy changes were much more prevalent in the
339 isolates selected for *GAL1* LOH (FIG. 2B) compared to isolates from YPD and non-diploid
340 isolates were frequently associated with reductions in both colony size and filamentous growth
341 (FIG. 2D).
342
343 **Genotypic diversity by RADSeq**
344 ddRADseq analysis was used to analyze Chr copy number and allele frequencies from
345 154 isolates off YPD and 275 isolates off 2DOG. The overall diversity was calculated using
346 Simpsons index of diversity (1-D) (see methods) for each mouse. The within-mouse diversity
347 was variable at day 1 and remained high thereafter (FIG. 2C). This suggests that diversity was
348 generated early after infection and was not highly deleterious to survival and growth within the
349 host.
350 We then clone-corrected the data set based upon the assumption that isolates with
351 identical genotype and CP from the same mouse were likely to be daughter isolates resulting
352 from a single mutational event. When the same event (genomic change) was found in different
353 mice, we expect that event was either frequent and/or subject to strong selective pressure in the
354 mice and was termed a 'recurrent' event.
355 Isolates that underwent ploidy shifts based on flow cytometry were re-analyzed by
356 ddRADseq. Interestingly, euploid shifts (the loss or gain of complete sets of Chrs) were
357 extremely rare; only three (of 10 confirmed) haploids and none of the 7 confirmed triploids or
358 tetraploids were euploid (FIG. 3A and B). By contrast, trisomy was detected for every Chr, with
359 higher trisomic frequencies for smaller Chrs and ChrR (FIG.4A and B) (Table 2, S3). There were
360 seven isolates in which the majority of Chrs (>4) were non-disomic or where Chrs were present
361 in multiple ploidy levels (e.g. monosomy, disomy and trisomy within the same isolate), providing
362 indirect evidence that euploid shifts likely preceded subsequent Chr missegregation events

(FIG.3C) (FORCHE *et al.* 2008; HARRISON *et al.* 2014; HICKMAN *et al.* 2015). Importantly, aneuploidy was detected in both Gal^+ and Gal^- isolates (FIG.S3).

Haploids were detected using flow cytometry optimized after detection of an initial haploid isolate from *in vitro* studies (HICKMAN *et al.* 2013). The detection of multiple haploid isolates (10/950 $2DOG^R$ isolates recovered initially from the mouse oral cavity) was unexpected and exciting. Of note, only three haploids were perfectly euploid, with 7 being near-haploid; all the haploids tested were relatively unstable and readily converted to the autodiploid state (HICKMAN *et al.* 2015), and data not shown), suggesting additional haploids may have been present *in vivo*. We identified nine distinct haploid or near-haploid genotypes that were recovered from 6 different mice: 3 single haploids from 3 different mice (Fig.3B), 2 unique haploids from one mouse (D3M2), 3 distinct haploids from one host (D5M5), and 2 identical haploids (different CPs but treated as likely clones, D3M1) from one mouse (Table S2). Interestingly, only 3 genotypes were identical between Hickman *et al.* (HICKMAN *et al.* 2013) and this study, which suggests that the original isolates were a mixed population (supported by mixed flow cytometry profiles, data not shown) although the instability of haploids may have contributed as well.

Whole Chr LOH was detected for all Chrs, with higher frequencies seen for the larger Chrs (ChrR, 1-3) (Fig.4C, Fig.S4) (Table 2, S3). The frequency of missegregation events is not entirely a function of Chr size, however, as the frequency of events on Chr3 (1.8 Mb), was higher than either Chrs 2 (2.1 Mb) or ChrR (2.3 Mb), which are 2.1 and 2.3 Mb, respectively. Of note, whole Chr LOH of Chrs 2, 4, 6, and 7, was biased towards allele A, consistent with the failure to detect homozygous allele B in haploids or other isolates (FORCHE *et al.* 2008; HICKMAN *et al.* 2013; FORD *et al.* 2015; HICKMAN *et al.* 2015; HIRAKAWA *et al.* 2015; HIRAKAWA *et al.* 2017). This suggests that the B alleles of these Chrs harbor lethal recessive alleles (FERI *et al.* 2016) and therefore cannot be entirely lost. We did identify isolates with segmental LOH towards allele B for Chrs2, 4, and 6 (FIG.5A).

389

390 **GAL1 LOH event characterization**

391 The isolate collection illuminates the diversity of molecular mechanisms that can yield a
392 Gal^- phenotype due to loss of the functional allele of *GAL1* from the B haplotype of Chr1
393 (Fig.5B). We classified 264 *GAL1* LOH events as either due to missegregation, (involving loss of
394 the entire B Chr) or recombination (involving LOH across the subsection of Chr1L
395 encompassing *GAL1*). Recombination was more frequent than missegregation (67% vs. 33% of
396 the total events respectively) consistent with aneuploidy of large Chrs as rare or deleterious
397 (Fig.5C) (Table S3). Among the Chr1 missegregation events, only the ten A-haplotype
398 monosomies can be explained by a single step process: non-disjunction of the B copy of Chr1
399 during mitosis, leading to progeny with a single copy of the A homolog. The remaining
400 aneuploids underwent at least two molecular events, either an increase in the copy number of
401 the A version of Chr1 by missegregation, followed by loss of the B homolog, or vice-versa.
402 Missegregation events were much more frequent than recombination on all other Chrs, with
403 whole Chr aneuploidy more frequent on smaller Chrs (Chr5-7) and whole Chr LOH more
404 prevalent on larger Chrs (Chr1-4, R) (Fig.5C).

405 Recombination events were categorized as: 1) LOH covering a Chr arm from the
406 recombination initiation site through the telomere (likely break induced replication, BIR); 2)
407 shorter-range LOH resulting from two crossover events that do not reach the telomere (double
408 crossovers or gene conversions, GC); and 3) segmental copy number variations (CNVs) which
409 includes truncations, amplifications and deletions. Recombination events on Chr1L were
410 dominated by BIR (94%), with GC events implicated in the remainder (6%, Fig.5B and 6). Rare
411 recombination events also were detected on all other Chrs in a few Gal^+ as well as in Gal^-
412 isolates and BIR was far more frequent than GC (Fig.5A and 6).

413 The sites of recombination across Chr1L appeared relatively randomly distributed
414 between *GAL1* and *CEN1* (Fig.5B). Four potential hot spot regions were identified by binning

Chr1L breakpoints every 50 kb along Chr1L (Fig.S5): two to the right of the *GAL1* (at ~450 kb) locus (451-500 kb and 501-550 kb), one between 701-750 kb and the last one between 851-900 kb (Fig.S5, Table S4). Most of these break regions were not near any genome elements known to promote double strand breaks such as transposable elements.

In addition to the missegregation and recombination patterns described above, a small number of complex rearrangements were also detected. These included Chr truncations and recombination events that involved segmental aneuploidies and could only arise through multiple sequential events on a single Chr (Fig.6). Complex events were only seen following 2DOG selection and were most frequent on Chr1 (Fig.6) (see Table S4 for break coordinates). In addition, several isolates had multiple crossover events on Chr1, some of them involving both Chr arms with LOH to AA and BB alleles at different positions (Fig.6). Taken together, these complex genotypes suggest that double-strand breaks (DSBs) were repaired via multiple, distinct mechanisms (see discussion).

Recurrent events

The finding that the same genomic changes appeared in multiple mice supports the idea that these are general responses of the genome to conditions encountered in the oral cavity during early stages of infection. Many missegregation events occurred in multiple mice (Fig.7, S6, Table S5). Homozygosis of Chr1 to the AA genotype was seen in every mouse following 2DOG selection, presumably because this is an efficient mechanism for *GAL1* LOH (Fig.4). Other recurrent whole Chr events, such as trisomy of Chr6, which appeared in the Gal^+ isolates as well (Fig.S6), was unexpected. Uniparental disomy of Chr3 and Chr5 trisomy were also prevalent. Whether these different missegregation events are advantageous during early infection or during the transition into and out of the host, remains to be determined.

Hypervariability in evolved isolates

Multiple combinatorial (i.e, recombination + missegregation) events were most frequent in Gal$^-$ isolates that also exhibit CPs (Fig.8A). Not surprisingly, multiple missegregation events were found together much more frequently than multiple recombination events—likely because aneuploidies and LOH arise during concerted Chr loss from tetraploid intermediates (FORCHE *et al.* 2008; HICKMAN *et al.* 2015). Recombination events were less prevalent in general and correspondingly the frequency of multiple recombination events was also much smaller.

We previously found that tetraploid isolates that underwent Chr loss yielded progeny with evidence of mitotic recombination that tended to involve multiple events on different Chrs (FORCHE *et al.* 2008). Combined with our observations of multiple and/or complex changes per isolate (Fig.5A and B, Fig.6), this suggests that once an isolate has undergone one mutational change it has an increased likelihood of additional changes. To quantitatively explore this idea, we calculated the frequencies of multiple *vs.* single events detected in the *in vivo* and *in vitro* samples (data not shown), as well as the frequency expected if each event arose randomly. The frequency of events was significantly greater than random for ≥ 5 changes per isolate for any genome change *in vivo* but not *in vitro* (only ≥ 7 changes was significant) (FIG.8B and C, Table S4), indicating that in the isolates studied, highly diverse isolates are overrepresented. This implies that rare individuals undergo high levels of recombination that involve multiple Chrs.

Discussion

To understand the evolutionary forces responsible for genomic rearrangements leading to fitter genotypes, one must first identify the types of changes that reshape the genome (CHADHA AND SHARMA 2014). Here, we provide the first population-level study of the standing variation that arises in *C. albicans* during oropharyngeal candidiasis by analyzing several hundred isolates recovered from 17 mice at different time points during the infection. Importantly, this study design provided the perspective of time within the host. Flow cytometry and ddRADseq of 429 isolates detected many types of events due to missegregation,

recombination and multiple events of both types. Our observations of missegregation and DSB-associated changes are consistent with two recent studies of genotypic and phenotypic intra-species variation and the evolution of drug resistance in single isolates of clinical *C. albicans* isolates (FORD *et al.* 2015; HIRAKAWA *et al.* 2015). While previous studies provide an important snapshot of ongoing changes in human infections, the lack of multiple isolates per time point makes it very difficult to recapitulate isolate genealogies throughout evolution. Of note, diversity was detectable even one day post infection, suggesting that either changes arise rapidly upon the shift in growth conditions from liquid medium to the mouse and back (JACOBSEN *et al.* 2008) or that exposure to the host environment for only 24 h of infection is sufficient to induce genotypic changes.

The detection of multiple independent haploid or near-haploid isolates with different genotypes was surprising, suggesting that haploidization repeatedly occurs in the oral cavity. We previously found Chr missegregation in isolates recovered after passage in a systemic model of infection and after *in vitro* exposure to physiologically relevant stressors (FORCHE *et al.* 2011). *In vitro,* the length of LOH tracts (short, long, whole Chr) was associated with the type and severity of stress applied. Here, all three types of LOH arose at appreciable frequencies along with high levels of aneuploidy, supporting the idea that *C. albicans* is exposed to significant combinatorial stress in the oral cavity even though it appears to flourish in the oral cavity during oropharyngeal candidiasis.

We detected a positive correlation between specific CPs and Chr missegregation. A large proportion of CPs were small in diameter and had completely smooth or less wrinkly colonies (Fig.2C), suggesting that they grow less well than the parental strain under the conditions tested and have defects in filamentous growth. This is reminiscent of the slow growth seen for aneuploidy *Saccharomyces cerevisiae* isolates grown in lab media (TORRES *et al.* 2007; THORBURN *et al.* 2013), which is thought to be the result of unbalanced protein stoichiometry, difficulty segregating aneuploidy Chrs or higher demands for DNA replication

493 (STORCHOVA *et al.* 2006; TORRES *et al.* 2008; PAVELKA *et al.* 2010; TORRES *et al.* 2010; BENNETT
494 *et al.* 2014; HIRAKAWA *et al.* 2015).

495 Mutants with filamentation defects cause less damage to epithelial and endothelial cells
496 *in vitro* (PHAN *et al.* 2000; TSUCHIMORI *et al.* 2000; BENSEN *et al.* 2002). This suggests that the
497 isolates with reduced filamentous growth may not express hyphal-specific genes (e.g., *ALS3*,
498 *SAP4* and *SAP6*) and/or may not be recognized as readily by the host immune cells. The
499 majority of isolates with small CM acquired whole Chr aneuploidy, supporting the idea that these
500 isolates may grow slowly under standard lab conditions, yet might have an advantage *in vivo*
501 (SEM *et al.* 2016). Interestingly, a subset of isolates recovered after a systemic infection in mice
502 also exhibited aneuploidy and LOH (FORCHE *et al.* 2009a).

503 Chr6 trisomy was much more frequent than other aneuploidies, and Chr6ABB was twice
504 as frequent as Chr6AAB. Chr6 harbors multiple members of important virulence gene families,
505 such as secreted aspartic proteases, lipases and adhesins (HUBE *et al.* 2000; NAGLIK *et al.*
506 2004; SCHALLER *et al.* 2005; HOYER *et al.* 2008; DJORDJEVIC 2010), the *NAG* gene cluster
507 important for alternative carbon utilization (KUMAR *et al.* 2000) and *RAD52*, a gene important for
508 DSB repair (CIUDAD *et al.* 2004; CIUDAD *et al.* 2005). Interestingly, overexpression of Rad52
509 increased genome instability (TAKAGI *et al.* 2008). Therefore, an extra copy of *RAD52* could
510 potentially lead to increased genome instability and amplification of specific advantageous
511 alleles (e.g. one extra copy of allele A) could promote adaption to specific environments such as
512 the oral cavity. Follow-up experiments will test the effect of Chr6 trisomy on survival,
513 persistence, and virulence of *C. albicans* in the oral cavity.

514 DSBs arise from endogenous sources including reactive oxygen species (e.g., produced
515 by immune cells), collapsed replication forks, and from exogenous sources including chemicals
516 that directly or indirectly damage DNA (SHRIVASTAV *et al.* 2007). The utilization of the *GAL1*
517 selection system not only allowed us to identify the major classes of genome changes and to
518 catalogue the types of LOH events that resulted in a Gal^- phenotype, but it also enabled us to

519 make hypotheses about the types of mechanisms that are involved in DSB repair. While the
520 majority of LOH was likely the result of BIR with or without crossover, more complex LOH
521 events also arose in a subset of isolates (see Fig.5C). The LOH signatures on Chr1 are
522 consistent with what would be observed after short and long patch mismatch repair using
523 different alleles as repair templates (COÏC *et al.* 2000; MARTINI *et al.* 2011; BOWEN *et al.* 2013).
524 Furthermore, more than one mismatch machinery may have been involved in repairing breaks.
525 Strikingly, similar LOH signatures were observed during mitotic DSB repair in *S. cerevisiae*
526 (GUO *et al.* 2017; HUM AND JINKS-ROBERTSON 2017), suggesting that these mechanisms may
527 have been conserved through evolution.

528 Whether genotypic variation arises through the parasexual cycle or via mitoic defects
529 followed by Chr missegregation accompanied by recombination events remains an outstanding
530 question. Our previous analysis of parasexual progeny showed that the majority of them were
531 aneuploid, that Chr missegregation predominated, that changes were observed for multiple
532 Chrs and that several isolates had short recombination tracts on multiple Chrs (FORCHE *et al.*
533 2008). A more recent study examined 32 parasexual progeny generated *in vitro* for a wide
534 range of virulence-associated traits and showed that parasexual mating can generate
535 phenotypic diversity *de novo*, and has important consequences for virulence and drug
536 resistance (HIRAKAWA *et al.* 2017) Direct evidence for the parasexual cycle *in vivo*, however,
537 remains elusive and the mechanism of mitotic failure followed by Chr loss events cannot be
538 ruled out (HARRISON *et al.* 2014).

539 Importantly, here we identified a substantial level of highly variable isolates, higher than
540 what one would expect by random chance alone. We hypothesize that hypervariable
541 subpopulations may be present in many natural populations, and that this diversity can enable
542 rapid adaptation in time of stress or environmental stochasticity. Whether the observed changes
543 are beneficial, detrimental or neutral remains to be determined, and is likely to be specific to the
544 particulars of the environment. The link between how specific genotypic changes affects

545 survival, persistence, and the virulence potential of *C. albicans,* and whether the host
546 recognizes and responds to this variation remains to be discovered.
547
548 **ACKNOWLEDGEMENTS**
549 We thank Carter Meyers for assistance with the animal studies and Darren Abbey for the
550 MAtlab script to determine likelihood for all permutations for each number of genotypic changes.
551 AMD, GC, EJ and AF were funded by NIH grant R15 AI090633 to AF. SGF was funded by NIH
552 grant R01DE022600 and JB was funded by an European Research Council Advanced Award
553 340087 (RAPLODAPT). ACG was supported by a CIHR Banting Postdoctoral Fellowship.
554
555
556 **Figure Legends**
557 FIG.1. Experimental overview of *in vivo* and *in vitro* experiments.
558
559 FIG.2. Genotypic and phenotypic diversity arises early during oral infection. **A.** CPs arise later in
560 the Gal^+ group (day 2, left graph) compared to the Gal^- group (day 1, right graph) but are not
561 observed *in vitro*. **B.** Ploidy changes (as measured by flow cytometry) do not arise in the Gal^+
562 group, but do so in the Gal^- group with ploidy shifts observed as early as day 1 post infection. **C.**
563 Simpson's D is high on average and depends strongly on composition of populations within
564 mice (Gal^+/Gal^- ration of isolates) left, overall Simpson's D; right, Simpson's D by days spent in
565 mice **D.** Bubble plot shows the number of haploid, diploid or aneuploid isolates (as measure by
566 flow cytometry) that exhibit indicated CPs determined by growth on YPD at 30°C for 3 days. CP
567 binary codes are shown in parenthesis (see also Table S2). Bubble size reflects the number of
568 isolates. Circles, total CFUs, green triangles, CPs.
569
570 FIG.3. Whole genome ploidy shifts are rare.

A. Shown is the ddRADseq whole genome karyotype for parental strain YJB9318. Allele status is indicated on the top. Bottom half of figure provides copy number for each Chr relative to diploid parent (1 = 2 copies) Chrs are colored in light grey and black to indicate start/end of each Chr. Color-coding is used throughout for indicated genotypes. Each dot on the lower part (copy number) is a copy number estimate for a restriction fragment based on the reads aligning to one end of the restriction fragment. The dots on the upper part (allele status) are maximum likelihood estimates of allele ratios at each (sequenced) known SNP site, constrained by the Chr or segment copy number and smoothed across x number of adjacent sites (see also methods). The colors for the allele status provide exact genotype for each Chr. Note: This strain background (RM1000 #2) has a preexisting Chr2L allele A homozygosis and a crossover on ChrR occurred during generation of the parental strain that was unmasked in isolates that became homozygous (see red arrow). In the case of whole Chr LOH, the genotype at the centromere was called (see black arrows). Gaps in allele coverage on Chrs3, 7, and R are due to lack of heterozygosity in the reference strain SC5314 used for analysis (FORCHE *et al.* 2004; VAN HET HOOG *et al.* 2007; BUTLER *et al.* 2009). **B.** Haploids and near haploids exhibit different genotypes, *(strain names in parenthesis from Hickman *et al.* 2013), y-axis, Chr copy number, x-axis, Chrs are ordered Chr1-7, and R. **C.** Isolates with > 2 ploidies/genome suggestive of ploidy shifts in progress.

FIG.4. Overview of missegregation events across Chrs. **A.** Whole Chr aneuploidies include trisomies and tetrasomies. The number of disomic Chrs from haploids and near haploids is shown in parentheses. Y-axis: normalized copy number relative to diploid parent. Aneuploid Chrs are boxed in. Note: Images show single whole Chr aneuploidies for clarity; most isolates carry more than one whole Chr aneuploidy. **B.** Single and double aneuploidies are detected both for Gal^+ and Gal^- isolates. Shown is number of isolates with 1 aneuploidy and 2 aneuploidies that were acquired *in vivo*. Chrs are shown from Chr1, Chr2-7, and R. **C.** Whole

597 Chr LOH more frequently occurs on larger Chrs 1-3, and R in Gal^- isolates. Combinations of
598 whole Chr LOH were not observed in Gal^+ isolates.
599
600 FIG.5. Recombination and missegregation events.
601 **A.** Crossover-associated events most often lead to *GAL1* loss *in vivo*. **B.** Location of LOH
602 breakpoints along Chr1. **C.** LOH breakpoints for Chr2-7, and R. Top horizontal black lines
603 represent the two homologs; black oval represents centromeres, arrows show location of the
604 major repeat sequence; Chr sizes are shown to the right of each Chr. The number of isolates for
605 each genotype are indicated at the left; for Chr1 the numbers are in shades of yellow/brown,
606 with higher numbers shaded darker; cyan, homozygous AA; magenta, homozygous BB; gray,
607 heterozygous AB. Breakpoints were mapped in 25 kb bins. Exact start/end coordinates of break
608 regions can be found in Table S4. Positions 1.6 - 2.8 Mb on Chr1 (indicated with 2 solid vertical
609 black lines) are not shown due to lack of any LOH events across this region. XO, crossover,
610 Maps are to scale.
611
612 FIG.6. Complex changes on individual Chrs include multiple recombination events on single
613 Chrs (mostly Chr1), segmental deletions, truncations, and amplifications. For legend, please
614 see Fig.5.
615
616 FIG.7. Recurrent missegregation events are frequent. Calculations were done for mice with *C.*
617 *albicans* populations size ≥ 12 (9 of 17 mice); bubble sizes reflect the percent mice where the
618 specific missegregation event (indicated on x-axis) was found. For example, whole Chr1 LOH
619 allele AA and whole Chr6 trisomy were found in all 9 mice (100%). Y-axis, Chr1-7, and R; x-
620 axis, missegregation genotypes.
621

622 FIG.8. Multiple changes (> 5) per isolate are significantly more frequent than what would be
623 expected by random chance alone *in vivo* but not *in vitro* **A.** Multiple combinatorial
624 (recombination (REC) + missegregation (MIS)) events are most frequent in Gal^- with CPs.
625 Percent of multiple event types for Gal^+ isolates (top left), Gal^+ plus CP (top right), Gal^- (bottom
626 left) and Gal^- plus CP (bottom right). Y-axis, number of recombination events/isolate; x-axis,
627 number of missegregation events per isolate. Bubble size represents the number of isolates
628 with indicated combinations, e.g. number of isolates that have 1 recombination and 1
629 missegregation event. Expected versus observed frequencies of changes *in vivo* (**B.**) and *in*
630 *vitro* (**C.**). Significance is indicated by ** (p = 0.01).
631
632 **Supplemental Figures**
633 FIG.S1. Detailed experimental overview. This figure is an expansion of Fig.1 and includes
634 detailed information about the total number of Gal^+ (YPD, no selection) and Gal^- (2DOG,
635 selection) isolates that were analyzed, that exhibit CPs, and genomic changes. Numbers are
636 parsed by mouse and the time that isolates spent in the mouse (in days). In addition, this figure
637 provides the actual total frequencies for CPs and genome changes determined by extrapolation
638 from the total number of CFUs.
639
640 FIG.S2. The frequencies of *GAL1* LOH is 2 orders of magnitude higher *in vivo*. Each open circle
641 represents one mouse (*in vivo*) or independent *in vitro* cultures (grown for 16 hrs).
642
643 FIG.S3. Single and double aneuploidies are detected for both Gal^+ (YPD, no selection) and Gal^-
644 (2DOG, selection) isolates. Shown are examples for isolates with 1 aneuploidy (1AN) and 2
645 aneuploidies (2AN) that were acquired *in vivo*. Chrs are shown from Chr1, Chr2-7, and R, gray
646 and black colors were used to show start and end of each Chr.
647

FIG.S4. Whole Chr LOH is more frequent for larger Chr1-3, and R. Shown are examples for whole Chr LOH either toward allele A or allele B and the number of isolates that acquired these specific genome changes. * For Chrs4 and 7 no isolate with a single whole Chr LOH was observed. Chrs exhibiting whole Chr LOH are boxed.

FIG.S5. Mapping of LOH breaks along Chr1L reveals 4 hotspot regions. Each of these regions contains between 15 and 22 breaks and is marked with a red arrow. Breaks were mapped in 25 kb bins due to low resolution of ddRADseq. *CEN1*, centromere 1

FIG.S6. Missegregation events are highly recurrent across mice.
Shown are summaries for 9 mice with N > 12. Chrs are indicated on the y-axis. Mouse IDs are indicated across the top. For each mouse there are two columns of pie charts. The first column shows the number of Chrs that are 1N, 2N, 3N, and 4N with shades of brown going from light (1N) to dark (4N). The right column shows allele status (heterozygous (gray), allele A (cyan), allele B (magenta)).

Table 1. Frequency of colony morphology phenotypes.

CM*	Total N = 429	Day1 N = 25	Day2 N = 54	Day3 N = 195	Day5 N = 155
00	0.62	0.8	0.74	0.66	0.49
01	0.03	0	0.04	0.04	0.02
02	0.03	0.04	0.04	0.02	0.05
03	0.01	0	0.00	0.02	0.01
10	0.06	0.04	0.07	0.04	0.10
11	0.07	0.04	0.02	0.05	0.12
12	0.14	0.08	0.06	0.12	0.21
13	0.01	0	0	0.02	0.01
20	0.02	0	0.04	0.04	0

*Colony morphology, see Fig. 2D for representative images.

667 Table 2. Summary of missegregation events by Chr.

Chromosome	Whole chromosome aneuploidy				Whole Chr LOH		
	Total	1N	3N	4N	Total	AA	BB
Chr1	18	10	7	1	81	81	0
Chr2	16	10	6	0	28	18	10
Chr3	11	10	1	0	41	11	30
Chr4	24	9	12	3	8	8	0
Chr5	34	10	24	0	6	5	1
Chr6	83	5	72	6	4	4	0
Chr7	19	5	12	2	5	5	0
ChrR	23	10	13	0	22	15	7

668 Number (events) for each genotype are shown separately for whole Chr aneuploidy (1N,
669 monosomic, 3N, trisomic, 4N, tetrasomic) and whole Chr LOH (AA, homozygous allele A, BB,
670 homozygous allele B)

671

672 **References**

673 Abbey, D., M. Hickman, D. Gresham and J. Berman, 2011 High-Resolution SNP/CGH
674 Microarrays Reveal the Accumulation of Loss of Heterozygosity in Commonly Used
675 Candida albicans Strains. G3 (Bethesda) 1**:** 523-530.

676 Bennett, R. J., A. Forche and J. Berman, 2014 Rapid mechanisms for generating genome
677 diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. Cold Spring Harb
678 Perspect Med 4.

679 Bensen, E. S., S. G. Filler and J. Berman, 2002 A forkhead transcription factor is important for
680 true hyphal as well as yeast morphogenesis in *Candida albicans*. Eukaryot Cell 1**:** 77-98.

681 Bowen, N., C. E. Smith, A. Srivatsan, S. Willcox, J. D. Griffith *et al.*, 2013 Reconstitution of long
682 and short patch mismatch repair reactions using Saccharomyces cerevisiae proteins.
683 Proceedings of the National Academy of Sciences of the United States of America 110**:**
684 18472-18477.

685 Butler, G., M. D. Rasmussen, M. F. Lin, M. A. S. Santos, S. Sakthikumar *et al.*, 2009 Evolution
686 of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459**:** 657-
687 662.
688 Calderone, R., 2012 *Candida and Candidiasis*. ASM, Washington, DC.
689 Chadha, S., and M. Sharma, 2014 Transposable elements as stress adaptive capacitors induce
690 genomic instability in fungal pathogen *Magnaporthe oryzae*. PLoS ONE 9**:** e94415.
691 Ciudad, T., E. Andaluz, O. Steinberg-Neifach, N. F. Lue, N. A. Gow *et al.*, 2005 Homologous
692 recombination in Candida albicans: role of CaRad52p in DNA repair, integration of linear
693 DNA fragments and telomere length. Mol Microbiol 56**:** 1396.
694 Ciudad, T., E. Andaluz, O. Steinberg-Neifach, N. F. Lue, N. A. R. Gow *et al.*, 2004 Homologous
695 recombination in Candida albicans: role of CaRad52p in DNA repair, integration of linear
696 DNA fragments and telomere length. Mol Microbiol 53**:** 1177-1194.
697 Coïc, E., L. Gluck and F. Fabre, 2000 Evidence for short-patch mismatch repair in
698 Saccharomyces cerevisiae. The EMBO Journal 19**:** 3408-3417.
699 Coyle, S., and E. Kroll, 2008 Starvation induces genomic rearrangements and starvation-
700 resilient phenotypes in yeast. Mol Biol Evol 25**:** 310-318.
701 Demuyser, L., M. A. Jabra-Rizk and P. Van Dijck, 2014 Microbial cell surface proteins and
702 secreted metabolites involved in multispecies biofilms. Pathogens and Disease 70**:** 219-
703 230.
704 Djordjevic, J., 2010 Role of phospholipases in fungal fitness, pathogenicity and drug
705 development- lessons from *Cryptococcus neoformans*. Frontiers Microbiol 1.
706 Fanning, S., W. Xu, N. Solis, C. A. Woolford, S. G. Filler *et al.*, 2012 Divergent Targets of
707 Candida albicans Biofilm Regulator Bcr1 In Vitro and In Vivo. Eukaryotic Cell 11**:** 896-
708 904.

709 Feri, A., R. Loll-Krippleber, P.-H. Commere, C. Maufrais, N. Sertour *et al.*, 2016 Analysis of
710 Repair Mechanisms following an Induced Double-Strand Break Uncovers Recessive
711 Deleterious Alleles in the Candida albicans Diploid Genome. mBio 7**:** e01109-01116.
712 Forche, A., D. Abbey, T. Pisithkul, M. A. Weinzierl, T. Ringstrom *et al.*, 2011 Stress alters rates
713 and types of loss of heterozygosity in *Candida albicans*. MBio 2.
714 Forche, A., K. Alby, D. Schaefer, A. D. Johnson, J. Berman *et al.*, 2008 The parasexual cycle in
715 *Candida albicans* provides an alternative pathway to meiosis for the formation of
716 recombinant strains. PLoS Biol 6**:** e110.
717 Forche, A., P. Magee, B. Magee and G. May, 2004 Genome-wide single-nucleotide
718 polymorphism map for Candida albicans. Eukaryotic cell 3**:** 705-714.
719 Forche, A., P. T. Magee, A. Selmecki, J. Berman and G. May, 2009 Evolution in Candida
720 albicans populations during a single passage through a mouse host. Genetics 182**:** 799-
721 811.
722 Forche, A., P. T. Magee, A. Selmecki, J. Berman and G. May, 2009a Evolution in *Candida*
723 *albicans* populations during single passage through a mouse host. Genetics 182**:** 799-
724 811.
725 Forche, A., G. May, J. Beckerman, S. Kauffman, J. Becker *et al.*, 2003 A system for studying
726 genetic changes in *Candida albicans* during infection. Fungal Genet Biol 39**:** 38-50.
727 Forche, A., G. May and P. T. Magee, 2005 Demonstration of loss of heterozygosity by single-
728 nucleotide polymorphism microarray analysis and alterations in strain morphology in
729 *Candida albicans* strains during infection. Eukaryot Cell 4**:** 156-165.
730 Ford, C. B., J. M. Funt, D. Abbey, L. Issi, C. Guiducci *et al.*, 2015 The evolution of drug
731 resistance in clinical isolates of Candida albicans. Elife 4**:** e00662.
732 Gerstein, A. C., M. S. Fu, L. Mukaremera, Z. Li, K. L. Ormerod *et al.*, 2015 Polyploid Titan Cells
733 Produce Haploid and Aneuploid Progeny To Promote Stress Adaptation. mBio 6**:**
734 e01340-01315.

735 Gorman, J. A., J. W. Gorman and Y. Koltin, 1992 Direct selection of galactokinase-negative
736 mutants of *Candida albicans* using 2-deoxygalactose. Curr Genet 21**:** 203-206.
737 Guo, X., Y. F. Hum, K. Lehner and S. Jinks-Robertson, 2017 Regulation of hetDNA Length
738 during Mitotic Double-Strand Break Repair in Yeast. Molecular Cell 67**:** 539-549.e534.
739 Harrison, B. D., J. Hashemi, M. Bibi, R. Pulver, D. Bavli *et al.*, 2014 A tetraploid intermediate
740 precedes aneuploid formation in yeasts exposed to fluconazole. PLoS Biol 12**:**
741 e1001815.
742 Hickman, M., G. Zeng, A. Forche, M. P. Hirakawa, D. Abbey *et al.*, 2013 The 'obligate diploid'
743 *Candida albicans* forms mating-competent haploids. Nature 494**:** 55-59.
744 Hickman, M. A., C. Paulson, A. Dudley and J. Berman, 2015 Parasexual Ploidy Reduction
745 Drives Population Heterogeneity Through Random and Transient Aneuploidy in Candida
746 albicans. Genetics 200**:** 781-794.
747 Hirakawa, M. P., D. E. Chyou, D. Huang, A. R. Slan and R. J. Bennett, 2017 Parasex generates
748 phenotypic diversity *de novo* and impacts drug resistance and virulence in *Candida*
749 *albicans*. Genetics 207**:** 1195-1211.
750 Hirakawa, M. P., D. A. Martinez, S. Sakthikumar, M. Z. Anderson, A. Berlin *et al.*, 2015 Genetic
751 and phenotypic intra-species variation in Candida albicans. Genome Res 25**:** 413-425.
752 Hoyer, L. L., C. B. Green, S. H. Oh and X. Zhao, 2008 Discovering the secrets of the Candida
753 albicans agglutinin-like sequence (ALS) gene family-a sticky pursuit. Med Mycol 46**:** 1 -
754 15.
755 Hube, B., F. Stehr, M. Bossenz, A. Mazur, M. Kretschmar *et al.*, 2000 Secreted lipases of
756 Candida albicans: cloning, characterization and expression analysis of a new gene
757 family with at least ten members. Archives of Microbiology 174**:** 362 - 374.
758 Hum, Y. F., and S. Jinks-Robertson, 2017 Mitotic gene conversion tracts associated with repair
759 of a defined double-strand break in *Saccharomyces cerevisiae*. Genetics 207**:** 115-128.

760 Jacobsen, M. D., A. D. Duncan, J. Bain, E. M. Johnson, J. R. Naglik *et al.*, 2008 Mixed *Candida*
761 *albicans* strain populations in colonized and infected mucosal tissues. FEMS Yeast
762 Research 8**:** 1334-1338.
763 Jakubovics, N. S., 2015 Intermicrobial Interactions as a Driver for Community Composition and
764 Stratification of Oral Biofilms. Journal of Molecular Biology 427**:** 3662-3675.
765 Kamai, Y., M. Kubota, T. Hosokawa, T. Fukuoka and S. G. Filler, 2001 New model of
766 oropharyngeal candidiasis in mice. Antimicrob Agents Chemother 45**:** 3195-3197.
767 Kumar, M. J., M. S. Jamaluddin, K. Natarajan, D. Kaur and A. Datta, 2000 The inducible N-
768 acetylglucosamine catabolic pathway gene cluster in Candida albicans: Discrete N-
769 acetylglucosamine-inducible factors interact at the promoter of NAG1. Proceedings of
770 the National Academy of Sciences of the United States of America 97**:** 14218-14223.
771 Legrand, M., A. Forche, A. Selmecki, C. Chan, D. T. Kirkpatrick *et al.*, 2008 Haplotype mapping
772 of a diploid non-meiotic organism using existing and induced aneuploidies. PLoS Genet
773 4**:** e1.
774 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence
775 Alignment/Map format and SAMtools. Bioinformatics 25**:** 2078-2079.
776 Lohberger, A., A. T. Coste and D. Sanglard, 2014 Distinct Roles of Candida albicans Drug
777 Resistance Transcription Factors TAC1, MRR1, and UPC2 in Virulence. Eukaryotic Cell
778 13**:** 127-142.
779 Lorenz, M. C., J. A. Bender and G. R. Fink, 2004 Transcriptional response of *Candida albicans*
780 upon internalization by macrophages. Eukaryot Cell 3**:** 1076-1087.
781 Lu, C.-W., Y. Tao, X.-H. Li, Y. Dong and D.-D. Zhou, 2017 Fungal chorioretinitis with systemic
782 candidiasis in an infant following treatment with broad spectrum antibiotics: A case
783 report. Experimental and Therapeutic Medicine 14**:** 286-288.
784 Ludlow, C. L., A. C. Scott, G. A. Cromie, E. W. Jeffery, A. Sirr *et al.*, 2013 High-throughput
785 tetrad analysis. Nat Meth PMID 23666411.

Lyon, J. P., S. C. da Costa, V. M. G. Totti, M. F. V. Munhoz and M. A. de Resende, 2006 Predisposing conditions for Candida spp. carriage in the oral cavity of denture wearers and individuals with natural teeth. Canadian Journal of Microbiology 52**:** 462-467.

Malavia, D., L. E. Lehtovirta-Morley, O. Alamir, E. Weiß, N. A. R. Gow *et al.*, 2017 Zinc Limitation Induces a Hyper-Adherent Goliath Phenotype in Candida albicans. Frontiers in Microbiology 8.

Martini, E., V. Borde, M. Legendre, S. Audic, B. Regnault *et al.*, 2011 Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair–Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways. PLoS Genetics 7**:** e1002305.

Muzzey, D., K. Schwartz, J. Weissman and G. Sherlock, 2013 Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. Genome Biol 14**:** R97.

Naglik, J., A. Albrecht, O. Bader and B. Hube, 2004 Candida albicans proteinases and host/pathogen interactions. Cell Microbiol 6**:** 915 - 926.

Naglik, J. R., S. J. Challacombe and B. Hube, 2003 *Candida albicans* secreted aspartyl proteinases in virulence and pathogenesis. Microbiol Mol Biol Rev 67**:** 400-428.

Nobile, C. J., D. R. Andes, J. E. Nett, F. J. Smith, F. Yue *et al.*, 2006 Critical role of Bcr1-dependent adhesins in *C. albicans* biofilm formation *in vitro* and *in vivo*. PLoS Pathog 2**:** e63.

Pande, K., C. Chen and S. M. Noble, 2013 Passage through the mammalian gut triggers a phenotypic switch that promotes *Candida albicans* commensalism. Nat Genet advance online publication.

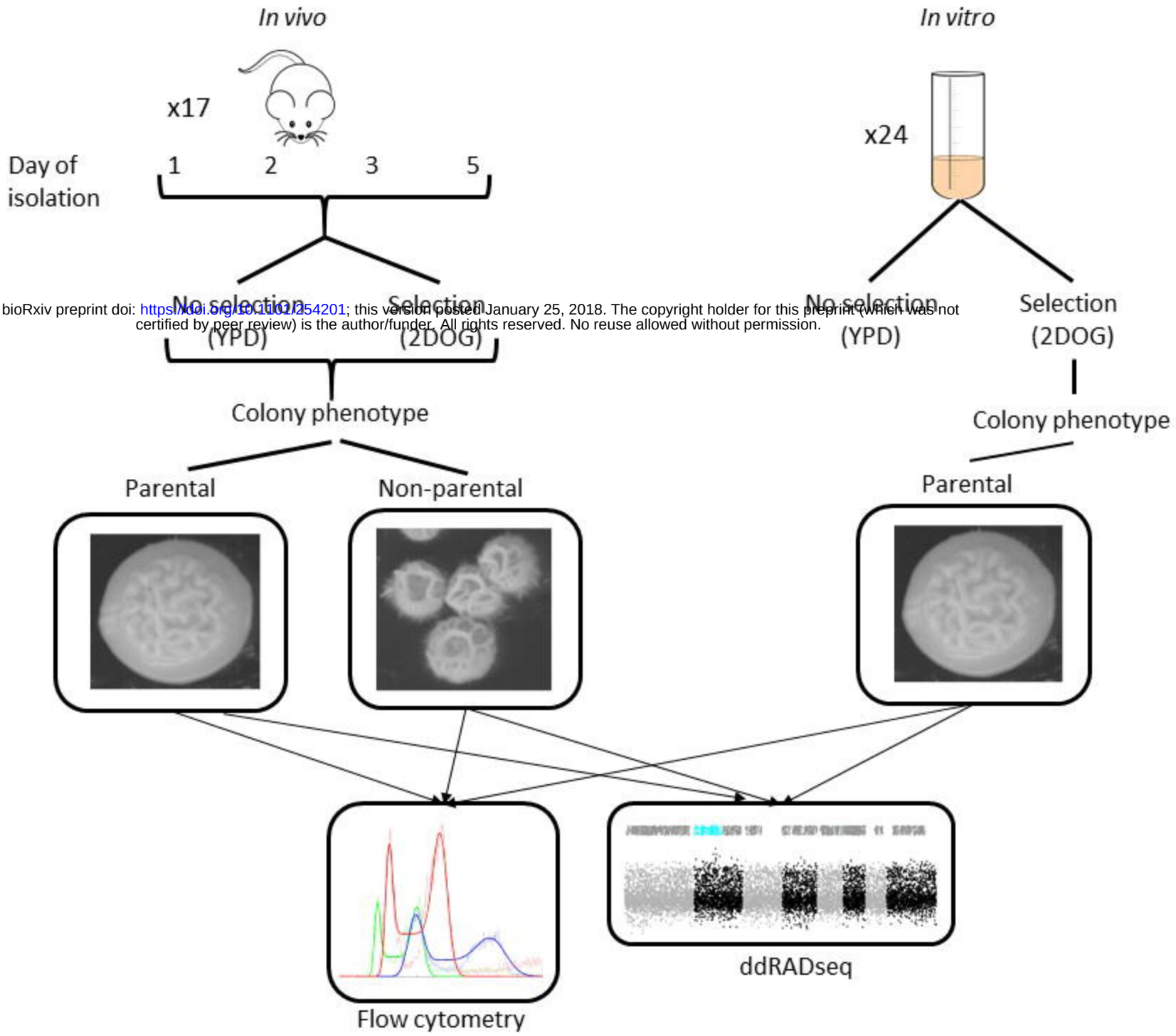
Pankhurst, C. L., 2013 Candidiasis (oropharyngeal). BMJ Clinical Evidence 2013**:** 1304.

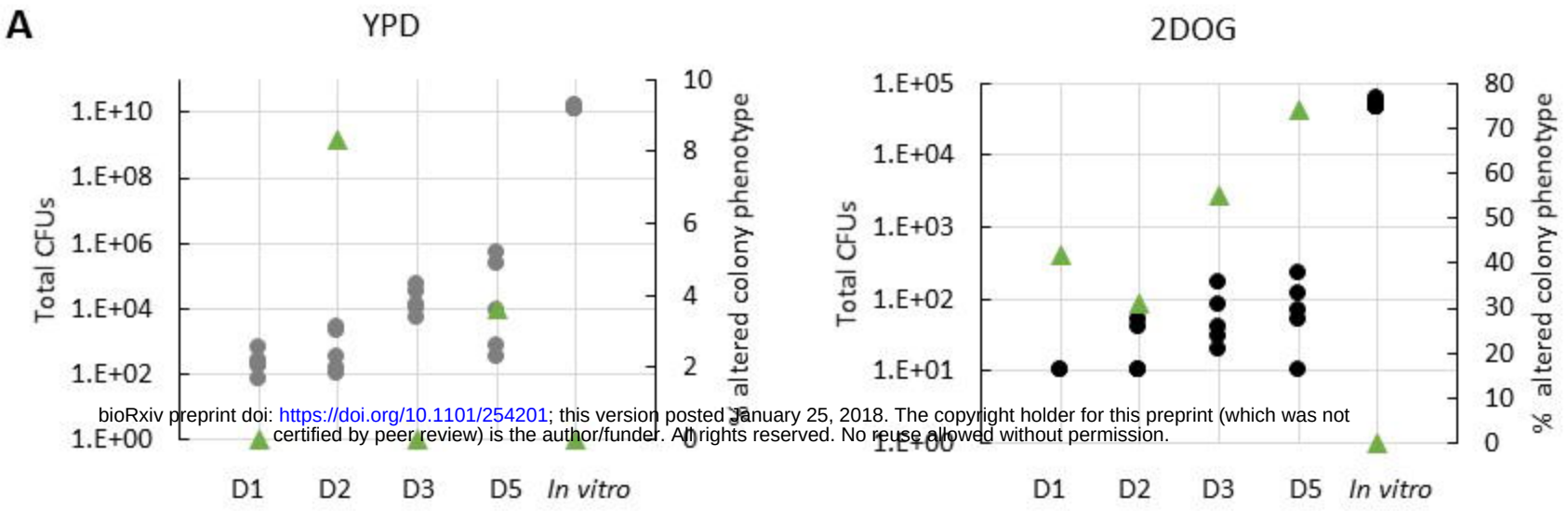
Park, H., Y. Liu, N. Solis, J. Spotkov, J. Hamaker *et al.*, 2009 Transcriptional Responses of Candida albicans to Epithelial and Endothelial Cells. Eukaryotic Cell 8**:** 1498-1510.

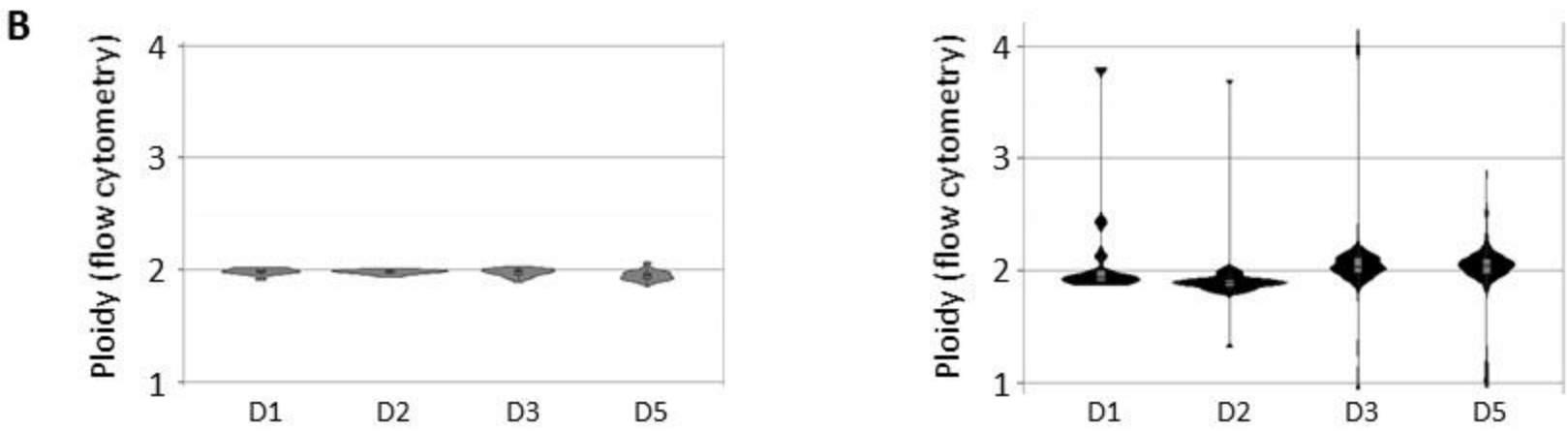
811 Patil, S., R. S. Rao, B. Majumdar and S. Anil, 2015 Clinical Appearance of Oral Candida
812 Infection and Therapeutic Strategies. Frontiers in Microbiology 6**:** 1391.
813 Pavelka, N., G. Rancati, J. Zhu, W. D. Bradford, A. Saraf *et al.*, 2010 Aneuploidy confers
814 quantitative proteome changes and phenotypic variation in budding yeast. Nature 468**:**
815 321-325.
816 Phan, Q. T., P. H. Belanger and S. G. Filler, 2000 Role of hyphal formation in interactions of
817 *Candida albicans* with endothelial cells. Infect. Immun. 68**:** 3485-3490.
818 Polakova, S., C. Blume, J. A. Zarate, M. Mentel, D. Jorck-Ramberg *et al.*, 2009 Formation of
819 new chromosomes as a virulence mechanism in yeast Candida glabrata. Proc Natl Acad
820 Sci U S A 106**:** 2688-2693.
821 Richardson, M., and R. Rautemaa, 2009 How the host fights against *Candida* infections. Front
822 Biosci 1**:** 246-257.
823 Rosenbach, A., D. Dignard, J. V. Pierce, M. Whiteway and C. A. Kumamoto, 2010 Adaptations
824 of Candida albicans for Growth in the Mammalian Intestinal Tract. Eukaryotic Cell 9**:**
825 1075-1086.
826 Rustchenko, E. P., D. H. Howard and F. Sherman, 1997 Variation in assimilating functions
827 occurs in spontaneous *Candida albicans* mutants having chromosomal alterations.
828 Microbiology 143**:** 1765-1778.
829 Schaller, M., C. Borelli, H. Korting and B. Hube, 2005 Hydrolytic enzymes as virulence factors of
830 Candida albicans. Mycoses 48**:** 365 - 377.
831 Selmecki, A., S. Bergmann and J. Berman, 2005 Comparative genome hybridization reveals
832 widespread aneuploidy in Candida albicans laboratory strains. Molecular Microbiology
833 55**:** 1553-1565.
834 Selmecki, A., A. Forche and J. Berman, 2006 Aneuploidy and isochromosome formation in
835 drug-resistant Candida albicans. Science 313**:** 367-370.

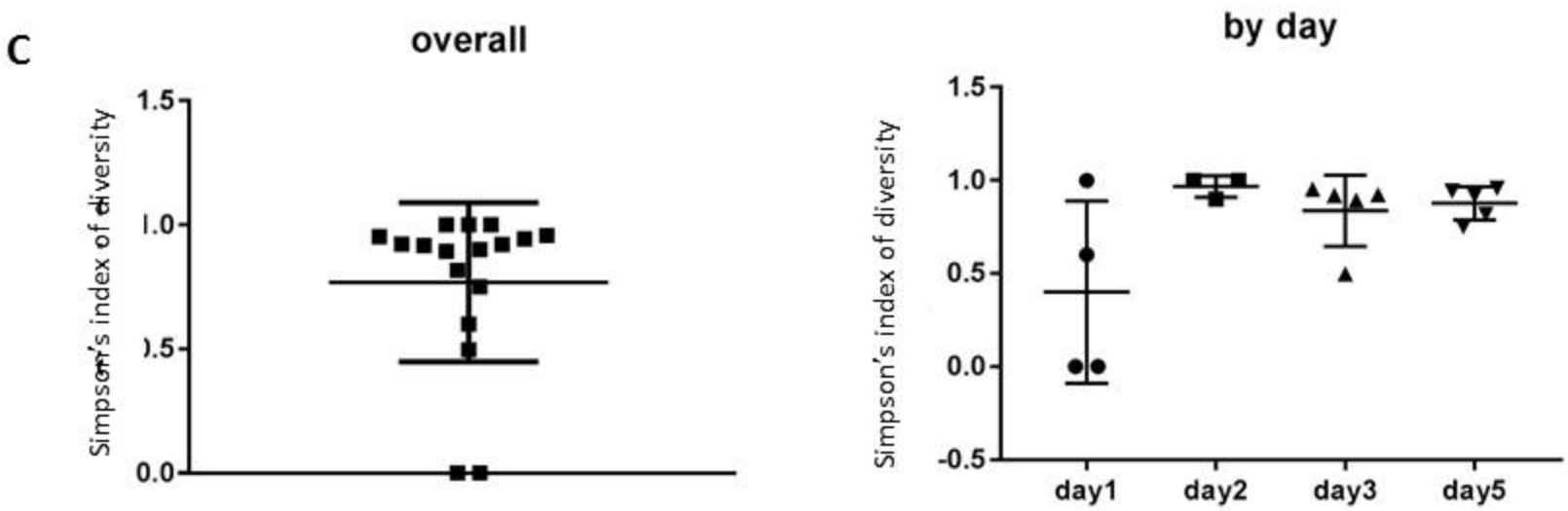
836 Sem, X., G. T. T. Le, A. S. M. Tan, G. Tso, M. Yurieva *et al.*, 2016 β-glucan exposure on the
837 fungal cell wall tightly correlates with competitive fitness of *Candida* species in the
838 mouse gastrointestinal tract. Frontiers in Cellular and Infection Microbiology 6**:** 186.
839 Shrivastav, M., L. P. De Haro and J. A. Nickoloff, 2007 Regulation of DNA double-strand break
840 repair pathway choice. Cell Research 18**:** 134.
841 Simpson, E. H., 1949 Measurement of diversity. Nature 163**:** 688.
842 Sionov, E., H. Lee, Y. C. Chang and K. J. Kwon-Chung, 2010 *Cryptococcus neoformans*
843 overcomes stress of azole drugs by formation of disomy in specific multiple
844 chromosomes. PLoS Pathog 6**:** e1000848.
845 Sobue, T., M. Bertolini, A. Thompson, D. E. Peterson, P. I. Diaz *et al.*, Chemotherapy-induced
846 oral mucositis and associated infections in a novel organotypic model. Molecular Oral
847 Microbiology**:** n/a-n/a.
848 Solis, N. V., and S. G. Filler, 2012 Mouse model of oropharyngeal candidiasis. Nat. Protocols 7**:**
849 637-642.
850 Staib, P., S. Wirsching, A. Strauss and J. Morschhauser, 2001 Gene regulation and host
851 adaptation mechanisms in *Candida albicans*. Int J Med Microbiol 291**:** 183-188.
852 Storchova, Z., A. Breneman, J. Cande, J. Dunn, K. Burbank *et al.*, 2006 Genome-wide genetic
853 analysis of polyploidy in yeast. Nature 443**:** 541-547.
854 Takagi, Y., R. Akada, H. Kumagai, K. Yamamoto and H. Tamaki, 2008 Loss of heterozygosity is
855 induced in *Candida albicans* by ultraviolet irradiation. Applied Microbiology and
856 Biotechnology 77**:** 1073.
857 Thorburn, R. R., C. Gonzalez, G. A. Brar, S. Christen, T. M. Carlile *et al.*, 2013 Aneuploid yeast
858 strains exhibit defects in cell growth and passage through START. Molecular Biology of
859 the Cell 24**:** 1274-1289.
860 Torres, E. M., N. Dephoure, A. Panneerselvam, C. M. Tucker, C. A. Whittaker *et al.*, 2010
861 Identification of aneuploidy-tolerating mutations. Cell 143**:** 71-83.

862 Torres, E. M., T. Sokolsky, C. M. Tucker, L. Y. Chan, M. Boselli *et al.*, 2007 Effects of
863 aneuploidy on cellular physiology and cell division in haploid yeast. Science 317**:** 916-
864 924.
865 Torres, E. M., B. R. Williams and A. Amon, 2008 Aneuploidy: Cells losing their balance.
866 Genetics 179**:** 737-746.
867 Tsuchimori, N., L. L. Sharkey, W. A. Fonzi, S. W. French, J. E. Edwards *et al.*, 2000 Reduced
868 Virulence of HWP1-Deficient Mutants of Candida albicans and Their Interactions with
869 Host Cells. Infection and Immunity 68**:** 1997-2002.
870 van Het Hoog, M., T. J. Rast, M. Martchenko, S. Grindle, D. Dignard *et al.*, 2007 Assembly of
871 the *Candida albicans* genome into sixteen supercontigs aligned on the eight
872 chromosomes. Genome Biol 8**:** R52.
873 Vandeputte, P., F. Ischer, D. Sanglard and A. Coste, 2011 *In vivo* systematic analysis of
874 *Candida albicans* Zn2-Cys6 transcription factors mutants for mice organ colonization.
875 PLOS One 6**:** e26962.
876 Vargas, K. G., and S. Joly, 2002 Carriage Frequency, Intensity of Carriage, and Strains of Oral
877 Yeast Species Vary in the Progression to Oral Candidiasis in Human Immunodeficiency
878 Virus-Positive Individuals. Journal of Clinical Microbiology 40**:** 341-350.
879 Villar, C. C., and A. Dongari-Bagtzoglou, 2008 Immune defence mechanisms and
880 immunoenhancement strategies in oropharyngeal candidiasis. Expert Reviews in
881 Molecular Medicine 10.
882 Wilson, D., S. Thewes, K. Zakikhany, C. Fradin, A. Albrecht *et al.*, 2009 Identifying infection-
883 associated genes of *Candida albicans* in the postgenomic era. FEMS Yeast Res 9**:** 688-
884 700.









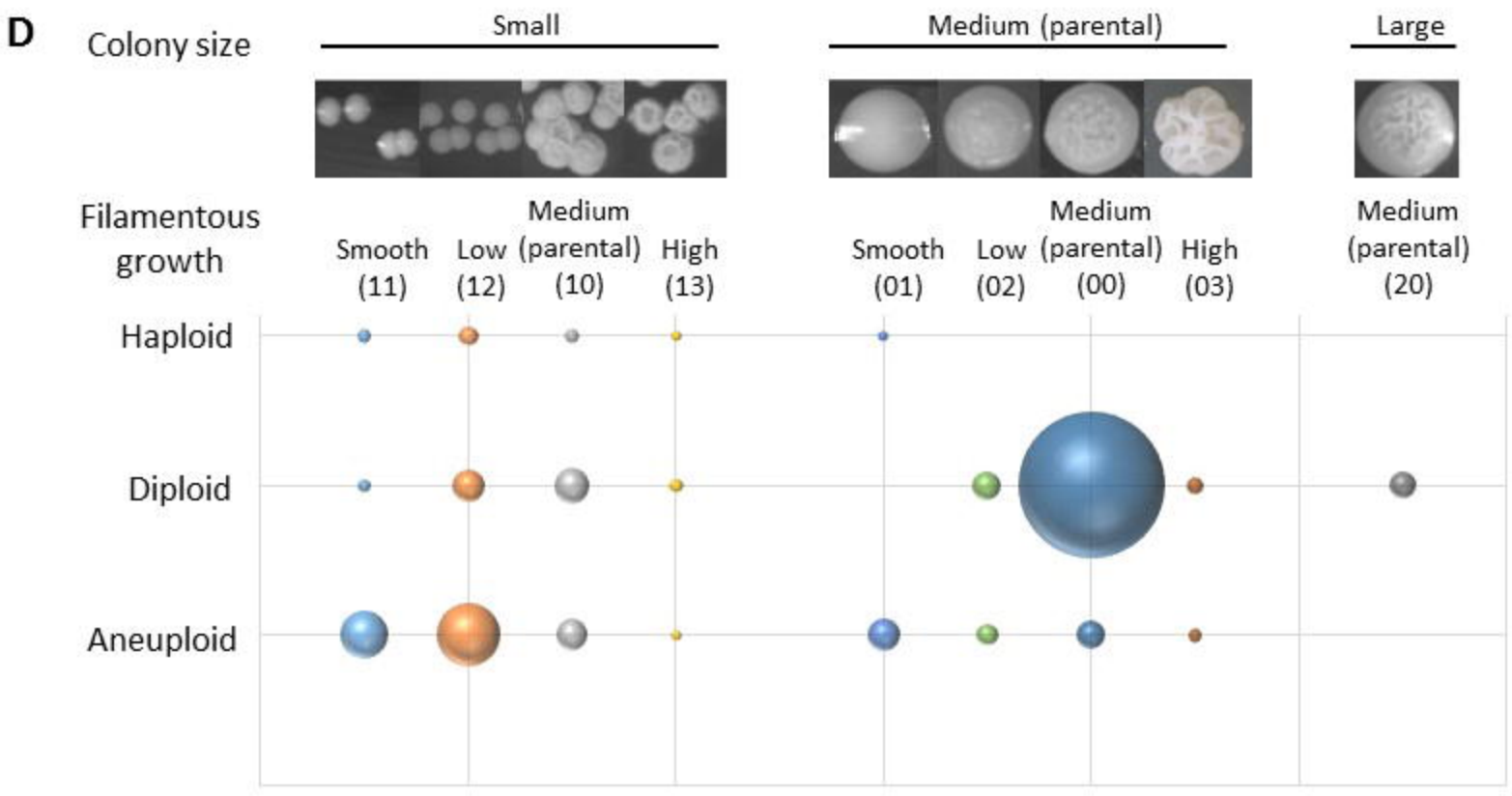
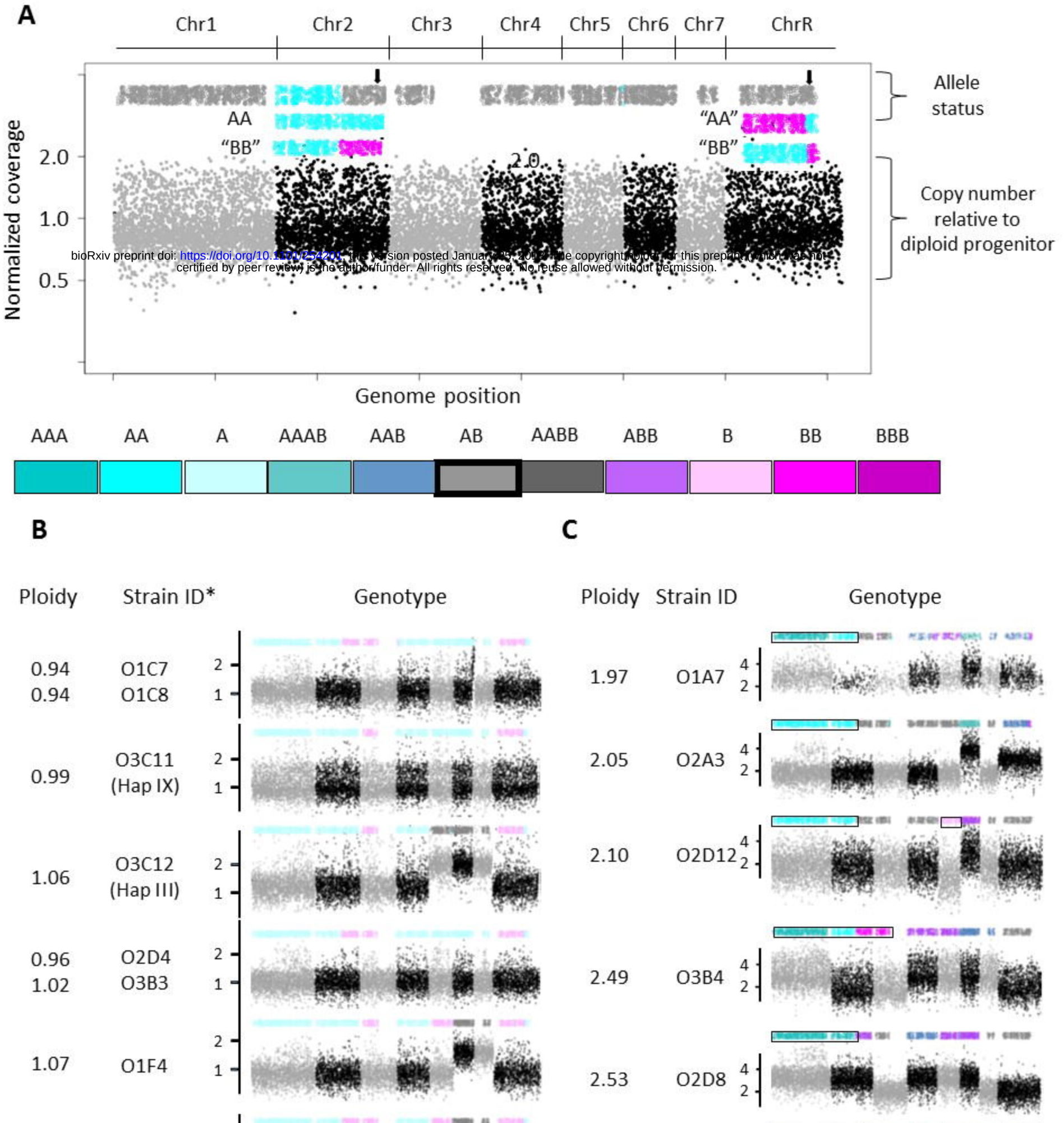


FIG2

A

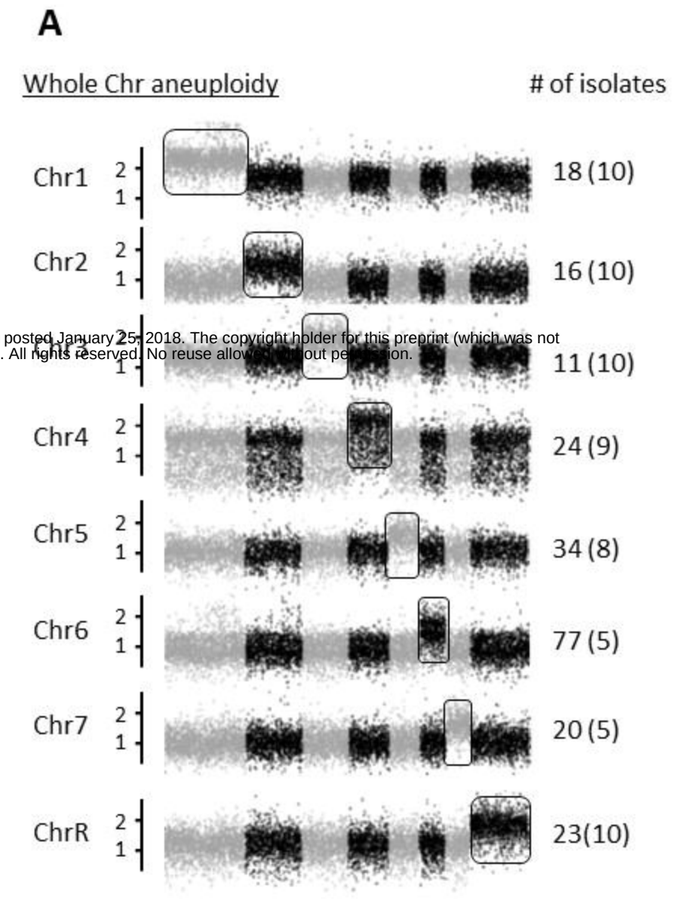


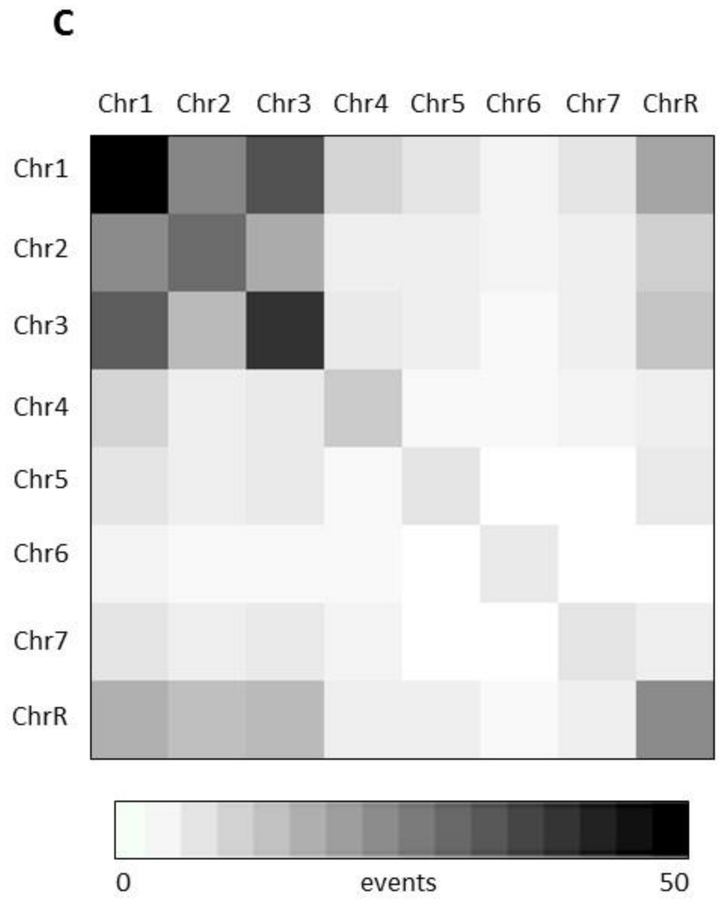
B

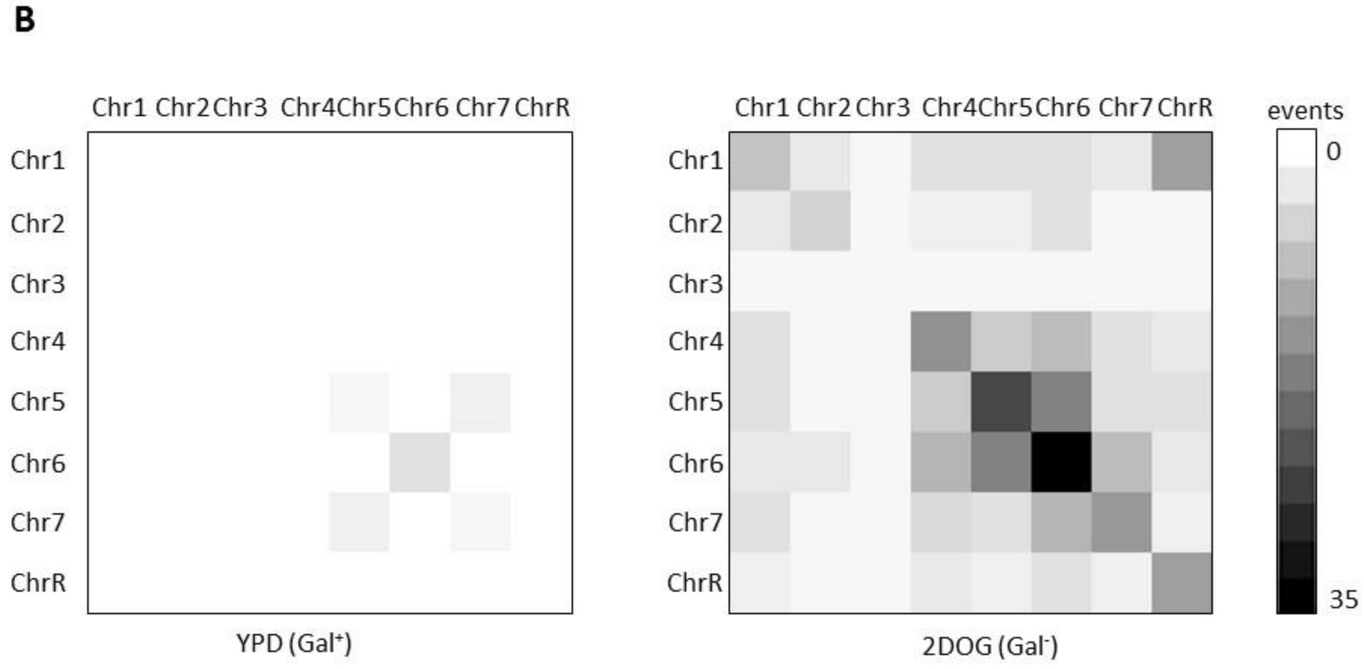
Ploidy	Strain ID*	Genotype
0.94	O1C7	
0.94	O1C8	
0.99	O3C11 (Hap IX)	
1.06	O3C12 (Hap III)	
0.96	O2D4	
1.02	O3B3	
1.07	O1F4	
1.13	O3C9 (Hap VIII)	
nd	O1F5	
1.31	O2C8	

C

Ploidy	Strain ID	Genotype
1.97	O1A7	
2.05	O2A3	
2.10	O2D12	
2.49	O3B4	
2.53	O2D8	
2.89	O2F4	
4.15	O1E9	







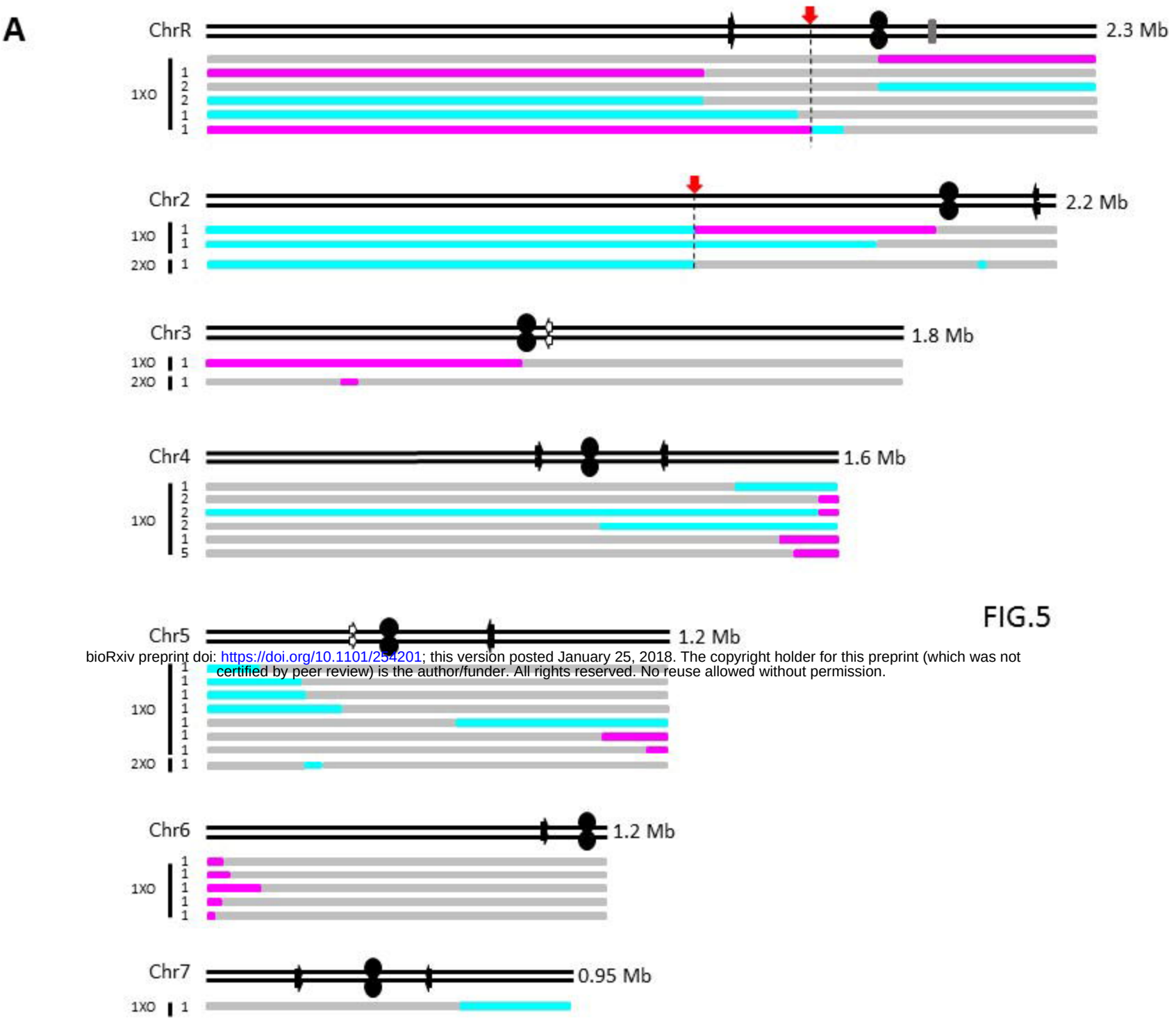
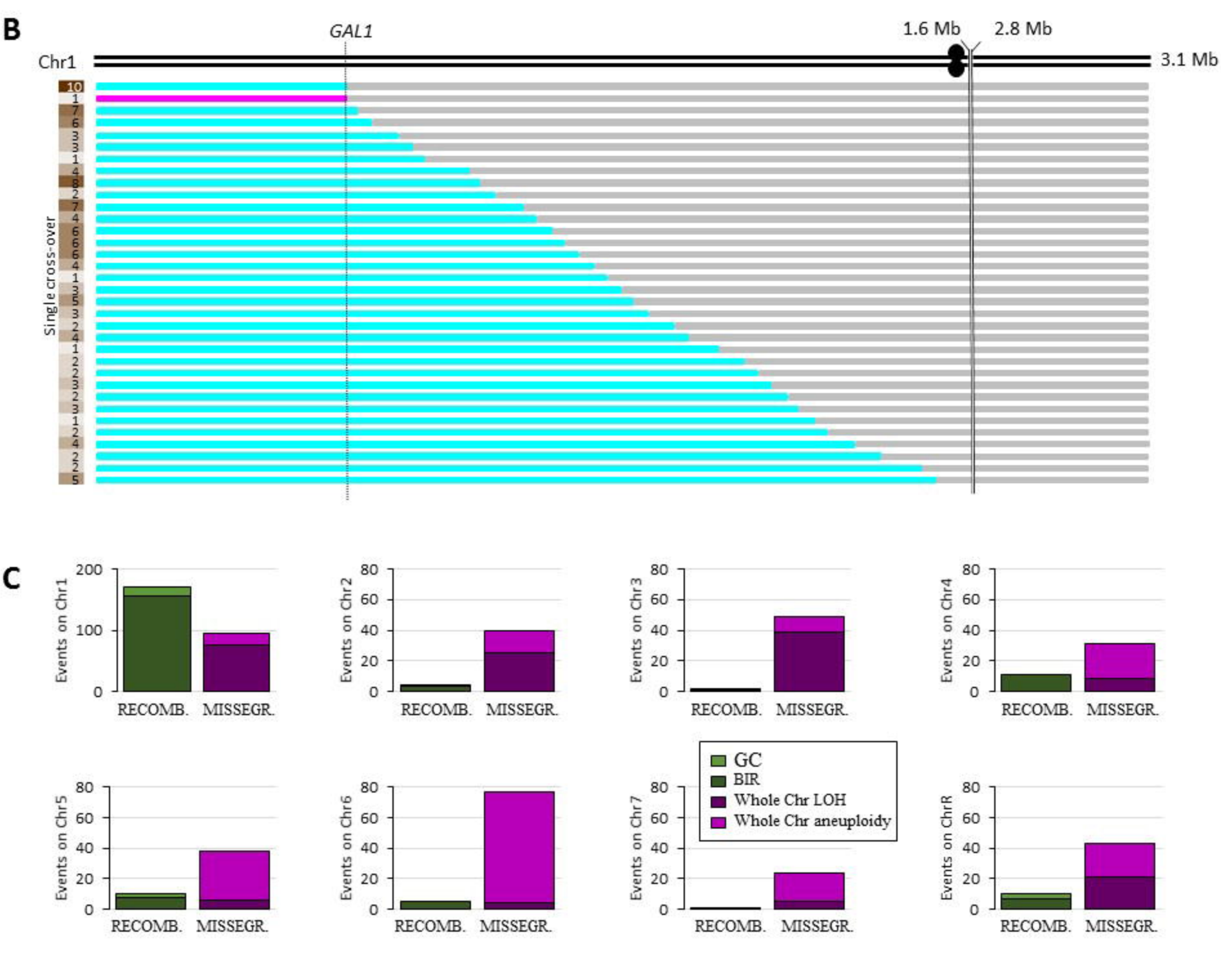
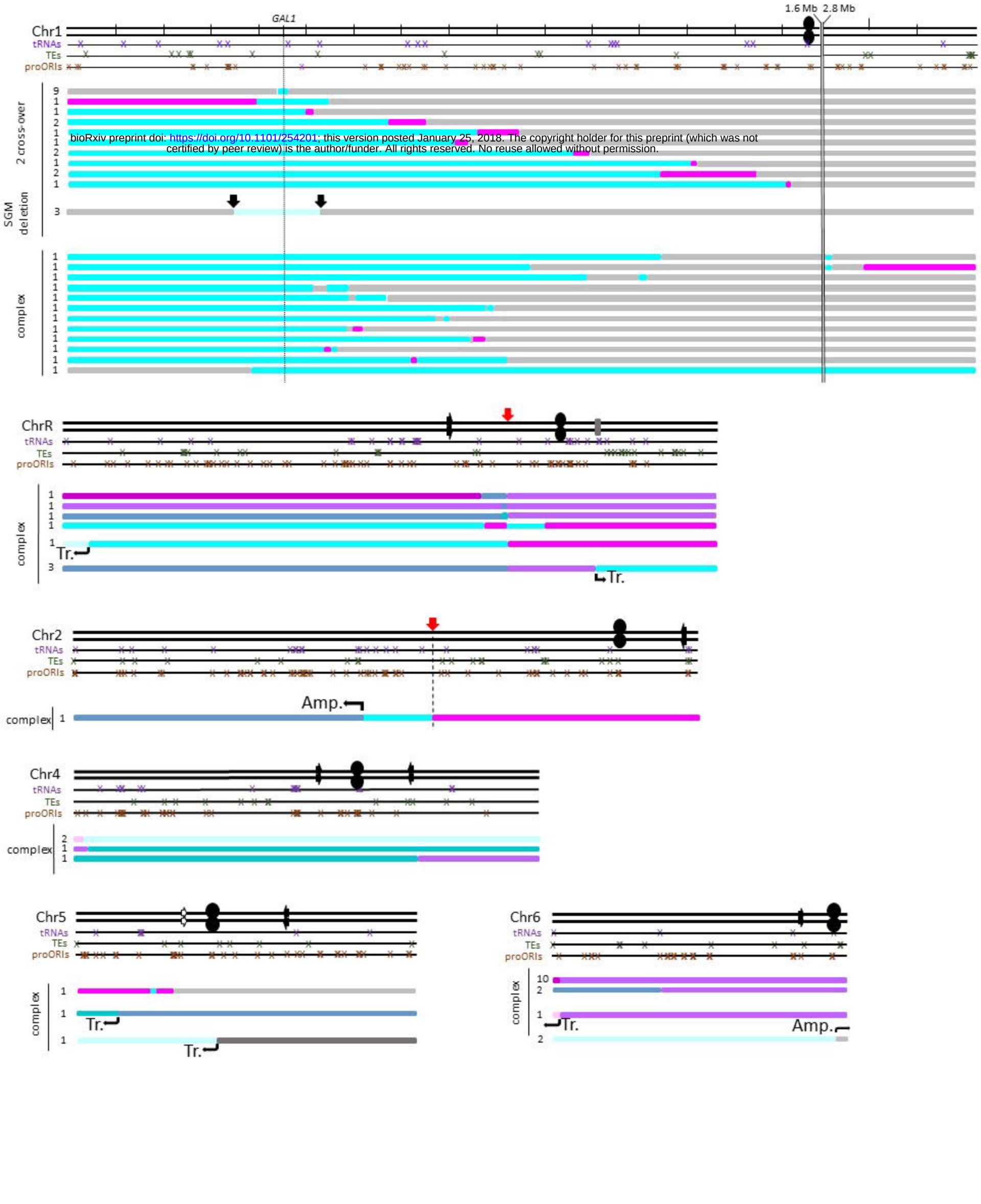
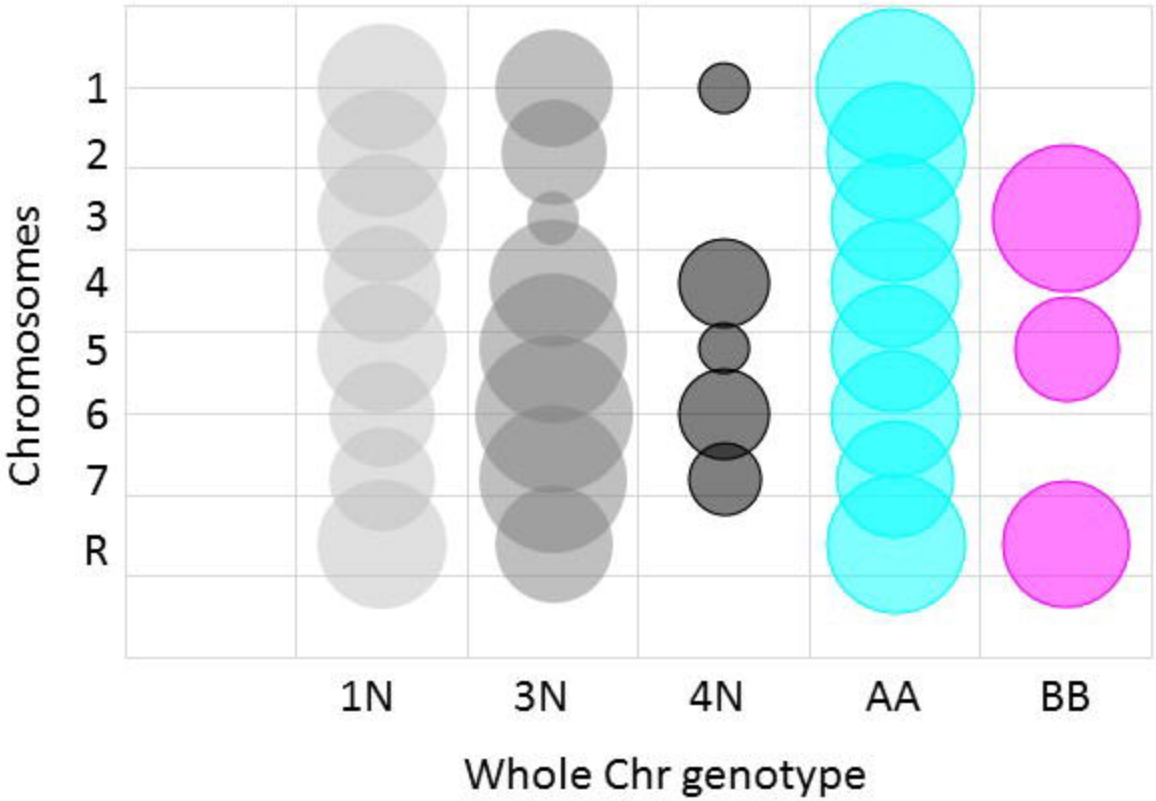
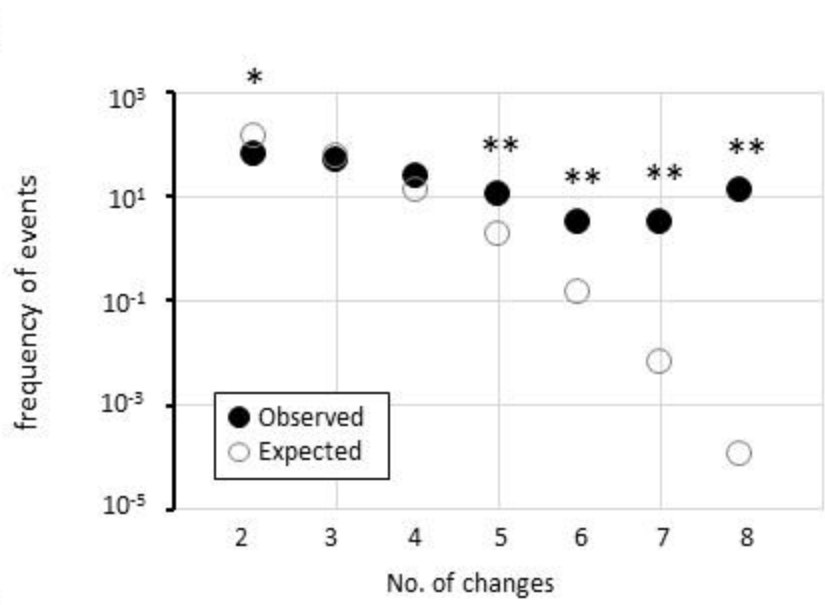


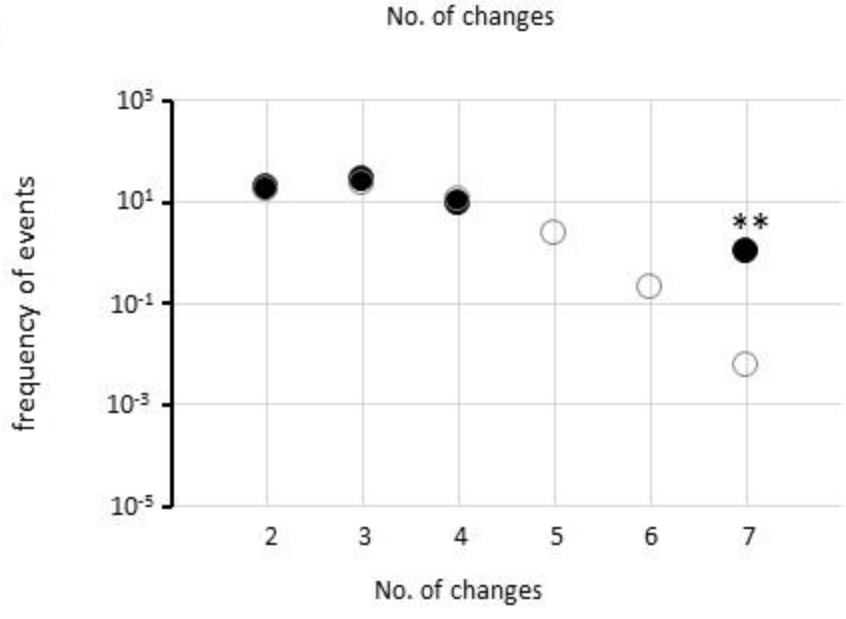
FIG.5






Chromosomes
1
2
3
4
5
6
7
R
1N
3N
4N
AA
BB
Whole Chr genotype


A
frequency of events
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
*
**
**
**
**
Observed
Expected
2
3
4
5
6
7
8
No. of changes


B
frequency of events
10^{3}
10^{1}
10^{-1}
10^{-3}
10^{-5}
**
2
3
4
5
6
7
No. of changes

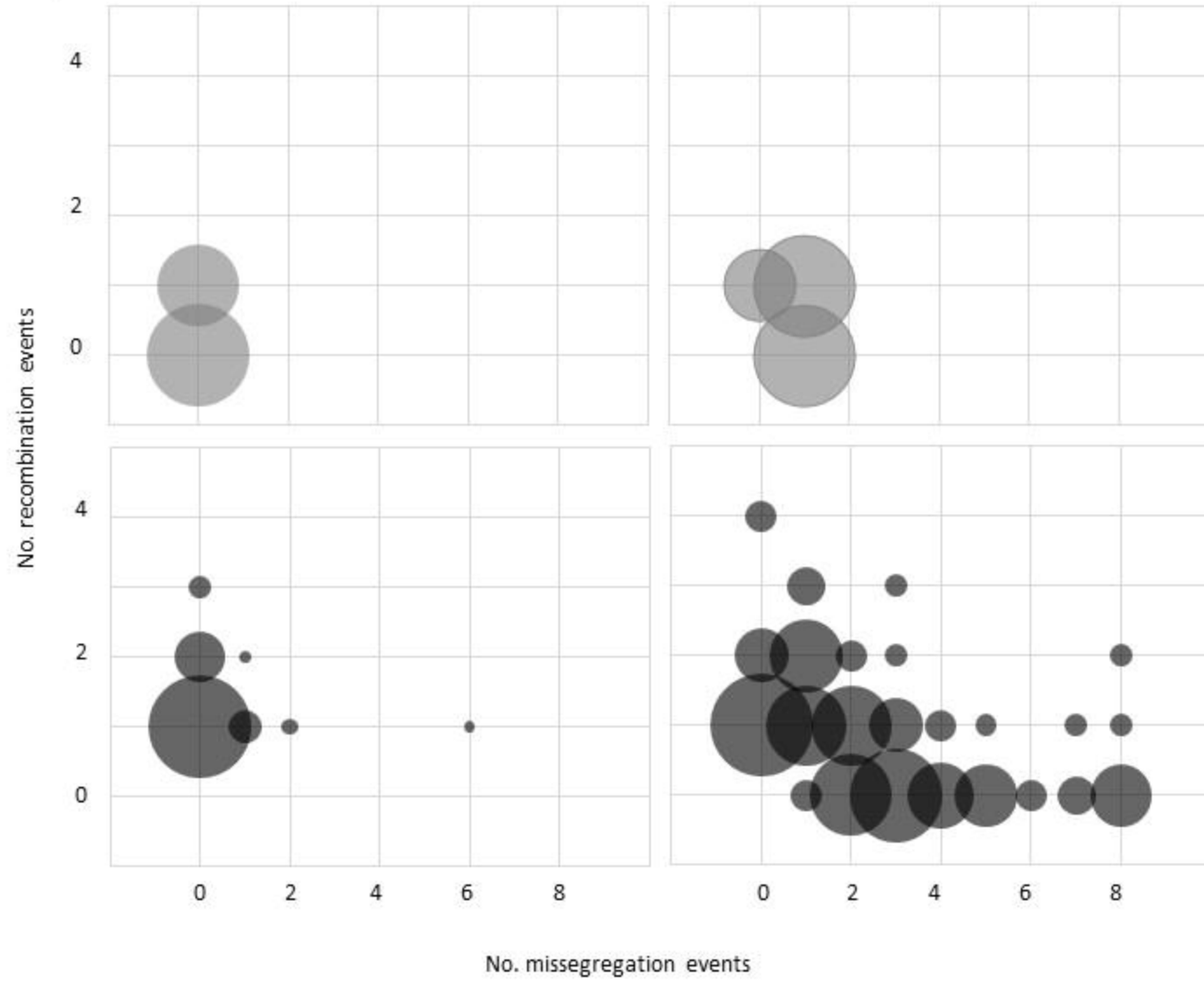

C
No. recombination events
4
2
0
4
2
0
0
2
4
6
8
0
2
4
6
8
No. missegregation events

FIG.8