Correlating Genotype and Phenotype in the Asexual Yeast Candida orthopsilosis Implicates ZCF29 in Sensitivity to Caffeine

Kontxi Martinez de San Vicente,1 Markus S. Schröder,2 Lisa Lombardi, Elise Iracane, and Geraldine Butler3
School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland

ABSTRACT Candida orthopsilosis is diploid asexual yeast that causes human disease. Most C. orthopsilosis isolates arose from at least four separate hybridizations between related, but not identical, parents. Here, we used population genomics data to correlate genotypic and phenotypic variation in 28 C. orthopsilosis isolates. We used cosine similarity scores to identify 65 variants with potential high-impact (deleterious effects) that correlated with specific phenotypes. Of these, 19 were Single Nucleotide Polymorphisms (SNPs) that changed stop or start codons, or splice sites. One variant resulted in a premature stop codon in both alleles of the gene ZCF29 in C. orthopsilosis isolate 185, which correlated with sensitivity to nystatin and caffeine. We used CRISPR-Cas9 editing to introduce this polymorphism into two resistant C. orthopsilosis isolates. Introducing the stop codon resulted in sensitivity to caffeine and to ketoconazole, but not to nystatin. Our analysis shows that it is possible to associate genomic variants with phenotype in asexual Candida species, but that only a small amount of genomic variation can be easily explored.

KEYWORDS Candida genotypetype pathogenic phenotype yeast

Population genomics studies of fungi are increasing, ranging from small-scale studies (approximately 20 isolates) to extremely large-scale studies (>1000 isolates) (Talas et al. 2016; Taylor et al. 2017; Peter et al. 2018; Ropars et al. 2018). These studies have revealed the high level of genetic variation among isolates, and sometimes uncovered evidence of distinct populations (Liti et al. 2009; Ellison et al. 2011; Schröder et al. 2016; Cissé et al. 2018; Ropars et al. 2018).

The association of phenotype with genetic variation can be studied in a number of ways, including linkage analysis and characterization of Quantitative Trait Loci (QTL), population genomics, and Genome Wide Association Studies (GWAS). All these approaches have been successfully applied in fungi, for example to identify loci associated with virulence, stress response and metabolism (Taylor et al. 2017). In Saccharomyces species QTL have been identified that correlate with response to stress (Kvítek et al. 2008), sugar transport (Bergström et al. 2014), xylose metabolism (Wenger et al. 2010) and adaptation to low temperature (García-Ríos et al. 2017), among others (Liti et al. 2017). QTL analysis has been used successfully in other fungi also, including Lachancea kluyveri (Sigwalt et al. 2016), Neurospora (Kim et al. 2007) and plant fungal pathogens (Lendenmann et al. 2016). GWA studies identified polymorphisms associated with virulence in fungal pathogens of plants and animals (Dalman et al. 2013; Talas et al. 2016; Zhong et al. 2017; Desjardins et al. 2017). A very large population genomics study of 1,011 S. cerevisiae isolates identified 35 variants associated with 14 conditions (Peter et al. 2018).

Standard linkage analysis cannot be applied in many fungal pathogens of humans, such as some in the Candida clade, where the species are asexual or have very rare, cryptic, sexual cycles (Wolfe and Butler 2017). Some studies however have identified causative variants in populations. For example, a population genomics study of 21 non-sexual C. albicans isolates identified a genetic variant causing a premature stop codon in the transcription factor EFG1 that is correlated with decreased
virulence (Hirakawa et al. 2015). Ropars et al. (Ropars et al. 2018) speculated that the presence of stop codons in transcription factor genes in one clade of C. albicans is associated with reduced virulence. However, there is a lack of systematic analyses of the contribution of individual genes to phenotype in pathogenic Candida species.

C. albicans belongs to the Lodderomyces group of the CTG-Ser clade, in which the CUG codon is translated as serine rather than leucine (Santos and Tuite 1995). Most (and perhaps all) members of the Lodderomyces clade have diploid genomes and do not undergo meiosis. The Lodderomyces clade includes the Candida parapsilosis sensu lato species group - C. parapsilosis, Candida orthopsilosis and Candida metapsilosis (Tavanti et al. 2005). Mating or meiosis have never been observed in C. parapsilosis sensu lato (Sai et al. 2011). C. parapsilosis is the most commonly isolated pathogen from this group, and it is of particular concern in very low weight neonates (Chow et al. 2012). C. orthopsilosis is second or third depending on the geographical region (Cantón et al. 2011; Bonfietti et al. 2012; Chen et al. 2015).

C. parapsilosis isolates have highly homozygous genomes, with approximately one heterozygous Single Nucleotide Polymorphism (SNP) per 15 kb (Butler et al. 2009) and some Copy Number Variations (CNVs), which may have resulted from recombination (Pryszcz et al. 2013). In contrast, C. metapsilosis isolates have very heterozygous genomes, and most likely arose from hybridization between two related species (Pryszcz et al. 2015). The first C. orthopsilosis genome sequence, with a size of ~12.66 Mb, was from a highly homozygous isolate (Riccombeni et al. 2012). Subsequent analysis showed that the genomes of most C. orthopsilosis isolates have high levels of heterozygosity. For example, the average number of SNPs is approximately 1 in 32 bp in C. orthopsilosis (Schröder et al. 2016) compared to 1 in 330 in C. albicans and 1 in 576 bp in Lodderomyces elongisporus (Butler et al. 2009). C. orthopsilosis isolates probably originated from at least four separate hybridization events between the same pair of parental species (Pryszcz et al. 2014; Schröder et al. 2016).

Here, we investigate the correlation between genotype and phenotype in 28 isolates of C. orthopsilosis, using cosine analysis. We identified >60 high-impact variants that are potentially correlated with phenotype. We implemented CRISPR/Cas9-based gene editing to confirm the effect of one high-impact variant on caffeine sensitivity.

**MATERIALS AND METHODS**

**Strains and growth conditions**

C. orthopsilosis strains (Table S1) were stored at -80°C in media containing 15% glycerol before being subcultured onto YPD agar medium (1% yeast extract, 2% peptone, 2% glucose and 2% agar) and incubated at 30°C. For phenotype analysis, isolates were inoculated as 2x2 arrays (two independent cultures with one technical replicate of each) into 96-well plates and incubated at 30°C for 24 h. Stocks were diluted in 96-well plates containing 200 µl of YPD broth (1% yeast extract, 2% peptone, 2% glucose) in 96-well plates and incubated at 30°C for 72 h. For phenotyping measurements, plates were photographed and growth was compared to YPD using SGAtools (Wagih et al. 2013). All phenotype images and growth scores are available at FigShare (File S1). SGAtools was designed to analyze synthetic genetic interactions and assumes that average growth on a plate does not vary. This was not true for several media, where many strains grew poorly. We therefore compared the growth of each strain on the test media to the growth of the same strain on YPD, using the raw data extracted from SGAtools.
For each strain in each analyzed growth condition, the SGAtools scores were converted to a binary score where 0 represents everything with a growth ratio above 0.45 (no, or minor, growth defect) and 1 represents scores below or equal to 0.45 (major growth defect). The SGAtools scores are normally distributed with a mean of 0.99, a median of 1.04, a 1st quartile of 0.76 and a 3rd quartile of 1.25. The cut-off of 0.45 was manually chosen after inspection of the phenotype plates. This score corresponds to a major growth defect.

Correlation between genotype and phenotype

Variants in each genome relative to the reference *C. orthopsilosis* 90-125 were identified as previously described (Schröder *et al.* 2016). Insertions, deletions and single nucleotide variants were identified using the GATK HaplotypeCaller (McKenna *et al.* 2010) with -nct 42 threads. Variants with mapping quality scores <40 and read depth <15 were removed. Variants from all strains were merged using GATK CombineVariants and are available at FigShare (File S2). Variants were functionally annotated with SnpEff (Cingolani *et al.* 2012), using the reference genome sequence from *C. orthopsilosis* 90-125 and gene annotation from the Candida Gene Order Browser (Fitzpatrick *et al.* 2010; Maguire *et al.* 2013). Variants with a high putative impact on a gene, as identified by SnpEff, were retained (available at FigShare, File S2). All high impact variants in each strain were converted to a binary score, with 1 representing a heterozygous or homozygous variant, and 0 representing absence of the variant. The correlation of the phenotype and genotype binary scores was then assessed using a cosine similarity measure in R (Gentleman *et al.* 2004). 1415 high impact variants with cosine similarity measures above or equal to 0.85 were retained at this stage. Variants that had an adjacent high impact variant within 10 bases or that were present in more than 20 alleles were subsequently excluded from further analysis, because they most likely result from poor alignment. Step 1 reduced the number of variants to 423, and step 2 reduced the variant number to 65 (Table S3). All remaining 65 variants were manually inspected in a local genome browser.

**Editing ZCF29 with CRISPR-Cas9**

A 20 bp protospacer sequence (guide RNA) targeting *C. orthopsilosis* ZCF29 (CORT0G04310) was designed using the Eukaryotic Pathogen CRISPR guide RNA Design Tool (Peng and Tarleton 2015). The guide RNA was generated by annealing of two short sequences.
oligos (gCoZCF29_TOP/BOT; Table S2) and the insert was cloned into the SapI-digested pCP-tRNA plasmid to generate plasmid pCP-tRNA-ZCF29, as previously described (Lombardi et al. 2019). The repair template (RT-CoZCF29) carrying the desired stop codon and 1 synonymous SNP to disrupt the PAM sequence was generated by primer extension (rCoZCF29_TOP/BOT, Table S2) using the ExTaq DNA polymerase (TaKaRa Bio, USA). C. orthopsilosis strains 1825 and 151 were transformed with 5 μg pCP-tRNA-ZCF29 and 25 μl of unpurified RT-CoZCF29 using the method previously described for C. parapsilosis (Holland et al. 2014). Transformants were selected on YPD agar plates containing 200 μg/ml nourseothricin (NTC), incubated at 30° for 48 h. Two NTC resistant transformants for each strain were sequenced using sCoZCF29_FWD/pCoZCF29_REV. The pCP-tRNA-ZCF29 plasmid was cured by growing the cells in the absence of selection on YPD until they could not grow anymore in the presence of NTC.

Galleria mellonella infection
Inoculation
Larvae were kept in individual glass vials containing 50 μL of stock suspensions. After inoculation, larvae were kept at 30° in the dark. The viability of larvae was monitored every 24 h, for four days. Virulence was analyzed by comparing the survival curves over time (by Kaplan-Meier estimate with a log-rank test). The statistical analyses were performed using the IBM SPSS Statistics 24 software package.

Data availability
All strains and plasmid are available upon request. Images and data analysis files (FileS1), VCF files (FileS2) and supplementary tables are available on Figshare. Table S1 contains a list of strains used, Table S2 contains the sequences of oligonucleotides used, Table S3 contains the raw and filtered high impact variants that correlate with phenotype data. Supplemental material available at Figshare: https://doi.org/10.25387/g3.8864960.

RESULTS
Phenotype analysis of C. orthopsilosis
Most C. orthopsilosis isolates fall into four or five clades that arose from independent hybridization events between related, but not identical parents termed A and B (Schröder et al. 2016). Two isolates (strains 90-125 and 428) have highly homozygous genomes and represent parent A. The remaining isolates demonstrate varying levels of heterozygosity (Schröder et al. 2016). We examined the growth characteristics of the 28 sequenced C. orthopsilosis isolates, representing all clades, in 60 different conditions, including alternative carbon sources, heavy metals, and exposure to antifungal drugs. Cultures were pinned to solid media including at least four replicates per strain, and growth was scored after 72 h growth using SGAtools (Wagih et al. 2013). SGAtools calculates colony size and averages all replicates. We compared the growth of C. orthopsilosis isolates on various media to growth on YPD, a rich media. Colonies with scores ≤0.45 (i.e., colony size was less than or equal to 45% on the test condition relative to growth on YPD) were assumed to have a moderate to severe growth defect in the tested conditions. Figure 1 shows the growth defects of all strains in 22 different conditions. All images and scores are available in File S1.

Some phenotypes are specific to one isolate or a small group of isolates. For example, strain 185 (clade 3) grows less well than all other isolates on 14 μg/ml nystatin, whereas strains 428 and 1799 grow poorly in the presence of SDS (Figure 1).

Correlating genotype and phenotype
We previously reported the genome sequences of the 28 C. orthopsilosis isolates (Riccombeni et al. 2012; Schröder et al. 2016). Homozygous

| Isolate | Media | Gene ID | Gene common | Impact | Copy No | Mutation
|---|---|---|---|---|---|---|
| B03950 | Caffeine 30 mM, Nystatin 14 μg/ml | E01960 | YND1 | Stop gained | Het | Arg182
| E01960 | YND1 | F01150 | SFL2 | Stop gained | Het | Gin456
| H02450 | CDC39 | B03390 | SLF1 | Start lost | Het | Met1
| C04260 | TSA1 | G04310 | ZCF29 | Stop gained | Hom | Leu135
| D04060 | RGD2 | G04310 | ZCF29 | Stop gained | Hom | Gin814
| D04150 | RVS162 | C01460 | RAD50 | Stop gained | Het | Tyr599
| E03200 | Stop gained | H01350 | Stop gained | Het | Leu32
| H01350 | Stop gained | G04310 | Stop gained | Het | Tyr56
| A02170 | Stop gained | E04360 | Stop gained | Het | Arg143
| E04360 | Stop gained | H00190 | Stop gained | Het | Ser540
| B06120 | Stop gained | H00190 | Stop gained | Het | Tyr20
| B06120 | Stop gained | H00190 | Stop gained | Het | Ser540
| B11130 | Start lost | B11240 | Start lost | Hom | Met1
| E01950 | Stop gained | E04360 | Stop gained | Het (428 Hom) | Lys215

1 = Non-synonymous SNP leading to stop codon at indicated amino acid position, Met1? = Methionine (start codon) at amino acid position 1 is lost.

Table 1 List of SNPs correlated with phenotype. See Table S3 for full list of variants.
and heterozygous single nucleotide polymorphisms and insertions and deletions relative to the reference genome *C. orthopsilosis* 90-125 (Riccombeni et al. 2012) were identified (Schröder et al. 2016). Here, we attempt to correlate these variants with specific phenotypes.

Putative high impact variants were identified using SnpEff, which predicts coding effects of variants (Cingolani et al. 2012). High impact variants are defined as those that cause frameshifts, introduce stop codons, lose start codons, or lie within splice acceptor or splice donor sites. A total of 6679 high impact variants were identified in 28 *C. orthopsilosis* isolates relative to the reference annotation. The number of unique high-impact variants varies between strains (Figure 2A), with isolate 599 having the least and isolate 428 the most. Some variants are shared by isolates within a clade (Figure 2B).

Population structure can interfere with the identification of variants associated with phenotypes, especially in asexual species where the structure can result from clonal reproduction (Taylor et al. 2017). Some of these issues can be addressed using Linear Mixed Models (Peter et al. 2018). Because we are characterizing a small number of very heterozygous isolates, we decided to use simple cosine measurements to investigate relationships between genotype and phenotype. The presence of each variant in each strain was represented using a binary score, where 1 represents the presence of a heterozygous or homozygous variant, and 0 represents absence of the variant. Growth phenotype vectors were also scored as either 1 (representing a growth defect, i.e., a score ≤0.45), or 0 (representing no growth defect). Correlation between growth scores and high impact variants was then determined using cosine similarity. The cosine similarity indicates how alike two vectors are to each other. In our case, one vector contains information about the presence or absence of a variant across all strains, and the second vector records the presence or absence of a growth defect across all strains. 65 variants with a cosine similarity value >0.85 were identified following filtering, as described in Methods (Table S3).

The majority of the identified variants that potentially correlate with phenotype result in predicted frameshifts. Some of these may be artifacts resulting from repetitive regions in the sequence, or genome assembly errors in the reference genome. We therefore restricted our analysis to highly supported SNPs that result in gain or loss of stop codons, loss of start codons, or variants that affect splicing. There were 19 variants of these types (Table 1). Because many isolates have more than one variant that correlates with the same phenotype, it is difficult to identify the causative allele using this method alone.

Strain 185 (CAS08-0185, (Schröder et al. 2016)) is sensitive to growth on caffeine, nystatin, ketoconazole, caspofungin, cobalt chloride and cadmium chloride (Figure 1). Cosine analysis suggested that sensitivity to caffeine and/or nystatin may be associated with variants that are specific to this isolate. These include four heterozygous SNPs resulting in premature stop codons, one heterozygous SNP resulting in a premature start, and two homozygous SNPs resulting in premature stops in both alleles at two loci (*TSA1* and *ZCF29*) (Table 1). *TSA1* encodes a predicted alkyl hydroperoxide peroxidase C, that is likely to be associated with oxidative stress signaling (Urban et al. 2005), whereas deleting *ZCF29* in *C. albicans* results in increased sensitivity to caffeine (Homann et al. 2009). The C to T SNP in *ZCF29* in *C. orthopsilosis* isolate 185 replaces glutamine 814 with a stop codon (Q814*), resulting in a protein that is 222 amino acids shorter than the wild type (Figure 3). The *ZCF29* gene was therefore chosen for additional analysis.

Because most *C. orthopsilosis* isolates are hybrids of two parents (called A and B) (Schröder et al. 2016), the *ZCF29* alleles in sample 185 could be derived from either parent. Phylogenetic analysis of *ZCF29* from all isolates (Figure 3) shows that in sample 185 and 20 other isolates, both alleles are derived from parent B. In another 6 isolates both alleles are derived from parent A. One isolate (sample 498) is unusual - the 5’ end of the gene suggests that one allele is derived from parent A and one from parent B, but the remaining of the gene has undergone LOH so that both alleles resemble parent B.

**Figure 3** Overview of Zcf29. (A) The presence of a premature stop codon in strain 185 is represented by the lollipop. The protein domains are shown as blue and green squares. AA = Amino acid. (B) Phylogenetic relationship of all ZCF29 sequences from *C. orthopsilosis* isolates. Most *C. orthopsilosis* strains are derived from hybridization between two parents (A and B). The phylogenetic tree shows that most strains retain both alleles of ZCF29 from the B parent after hybridization. Strain 185 (premature stop codon) is highlighted with a rhombus. Strains are color-coded based on their clade of origin (Clade 1 in green, Clade 2 in red, Clade 3 in orange, Clade 4 in blue, homozygous strains in purple).

**Editing ZCF29 in C. orthopsilosis**

To test if the homozygous stop codon variant in *ZCF29* in sample 185 would confer sensitivity to caffeine, nystatin or other drugs in other genetic backgrounds, we used the CRISPR-Cas9 technology to
introduce the variant into C. orthopsilosis isolates 151 (CAS08-0151) and 1825 (CAS10-1825), which belong to the same clade as sample 185 (Schröder et al. 2016) but are not sensitive to caffeine or nystatin (Figure 1).

A guide RNA was expressed on a plasmid between a tRNA and a ribozyme (Lombardi et al. 2019) generating plasmid pCP-tRNA-CoZCF29, and was transformed together with a repair template RT-CoZCF29, designed to introduce a stop codon at amino acid 814 (Q814*), reproducing the polymorphism seen in strain 185. A synonymous SNP was also introduced to destroy the PAM site targeted by the guide RNA (Figure 4A). The edited mutants (151eZCF29 and 1825 eZCF29) were confirmed by sequencing.

Editing ZCF29 confers sensitivity to caffeine

Cosine analysis suggested that variants in strain 185 correlated with sensitivity to caffeine and nystatin (Table 1). However, strain 185 is also sensitive to ketoconazole, caspofungin, cobalt chloride, and cadmium chloride (Figure 1). We therefore tested the growth of the edited strains in all conditions. Strains with an introduced premature stop codon in ZCF29 showed reduced growth on caffeine and ketoconazole compared to parental strains. (C) Kaplan-Meier curves illustrating the survival over time of G. mellonella larvae injected with C. orthopsilosis strains. Twenty larvae were used per strain. Log-rank test results comparing wild type strains (in green) and ZCF29 edited strains (in pink) show no significant differences between them (P = 0.109 in isolate 151 and P = 0.446 in isolate 1825).

![Figure 4](image-url) Introducing stop codons in ZCF29 in C. orthopsilosis confers sensitivity to caffeine and ketoconazole, and does not affect virulence. (A) CRISPR-editing was used to introduce the homoygous variant in ZCF29 in sample 185 into strains 151 and 1825. After the Cas9-induced double-strand break (scissors), a repair template was used to introduce one synonymous SNP to disrupt the PAM (in yellow) and the C > T SNP resulting in the premature stop codon. The guide sequence is boxed in black. The SNPs introduced in the mutated sequence are highlighted in red. (B) Wild type and ZCF29 edited versions of strains 151 and 1825, and the original strain 185, were grown in YPD overnight and diluted in PBS before plating on YPD agar plates containing nystatin, caffeine, ketoconazole (KCZ), caspofungin (CAS), cobalt chloride (CoCl₂), or cadmium chloride (CdCl₂) at the indicated concentrations. Plates were photographed after 72h at 30°C. Strains with an introduced premature stop codon in ZCF29 showed reduced growth on caffeine and ketoconazole compared to parental strains. (C) Kaplan-Meier curves illustrating the survival over time of G. mellonella larvae injected with C. orthopsilosis strains. Twenty larvae were used per strain. Log-rank test results comparing wild type strains (in green) and ZCF29 edited strains (in pink) show no significant differences between them (P = 0.109 in isolate 151 and P = 0.446 in isolate 1825).
ZCF29 showed reduced growth in caffeine and ketoconazole (0.5 μg/ml), compared to parental strains (Figure 4B). No differences in growth between the edited strains and their parental wild type strains were observed in the presence of nystatin, caspofungin, cobalt chloride, or cadmium chloride (Figure 4B).

**Editing ZCF29 does not alter virulence of C. orthopsilosis**

In *C. albicans*, mutating ZCF29 results in reduced filamentation and decreased colonization of mouse organs (Vandeputte et al. 2011). To test if disrupting ZCF29 reduces the pathogenicity of *C. orthopsilosis*, we compared the virulence of the edited strains to their wild-type strains in the model host *G. mellonella* (Figure 4C). The virulence of the edited strains is not significantly different from their respective wild type parental strains (log-rank test, Figure 4C).

**DISCUSSION**

Identification of specific polymorphisms associated with phenotypes is challenging, and usually requires linkage analysis, or combining linkage studies with genome sequencing. However, it is difficult to carry out association studies in asexual species. In this study, we identified 65 high impact variants in the asexual yeast *C. orthopsilosis* that are linked to specific phenotypes. 19 are SNPs that likely represent high confidence variants.

We validated our approach by testing the effect of one variant which introduces a premature stop codon in both alleles of the transcription factor ZCF29 and is present only in *C. orthopsilosis* strain 185. This variant correlated with decreased growth relative to other strains when caffeine or nystatin is present in the media. We found that we could recapitulate the caffeine sensitivity phenotype by introducing the variant into ZCF29 in two different *C. orthopsilosis* isolates. Nystatin sensitivity may be associated with other variants in this isolate, such as the homozygous premature stop in TSA1, or heterozygous stops in 5 other alleles (Table 1). However, we found that editing ZCF29 also reduced ketoconazole sensitivity, although we did not identify a significant correlation between ketoconazole sensitivity and the variant (Table 1). It is therefore likely that ketoconazole sensitivity is caused by many different variants in different *C. orthopsilosis* isolates.

Zcf29 belongs to the Zn(II)2Cys6 zinc cluster family of transcription factors in fungi. In *C. albicans*, mutations in ZCF29 result in sensitivity to the TOR (target of rapamycin) inhibitor beaucovirin, to caffeine and to the oxidative stress inducer menadione (Homann et al. 2009; Vandeputte et al. 2011; Shekhar-Guturja et al. 2016). The role of ZCF29 in caffeine sensitivity is conserved between *C. albicans* and *C. orthopsilosis*. *C. albicans* mutants are also unable to filament at 42° or in serum containing media or to colonize mouse kidneys (Vandeputte et al. 2011). At least some of these phenotypes result from regulation of expression of multidrug transporter genes (Shekhar-Guturja et al. 2016).

The presence of stop codons in ZCF29 in isolates belonging to one *C. albicans* clade has been associated with reduced virulence in this species (Ropars et al. 2018). Editing ZCF29 in *C. orthopsilosis* did not have any effect on virulence. However, the stop codon was introduced after the predicted DNA binding and chromatin structure-remodelling domains (Figure 3). It is therefore possible that the truncated protein retains some function. In addition, we monitored virulence using the mini-host *G. mellonella*. There is some evidence that the response of *Galleria* to infection can mimic the response of the mammalian host. For example, in a recent study, 10 of 18 *C. albicans* deletion strains tested had similar effects on fungal load in mice and in *G. mellonella* larvae (Amorim-Vaz et al. 2015). Deleting *C. albicans* SPT20 also reduces virulence in both mice and larvae (Tan et al. 2014). However, other studies found inconsistencies. For example, Frenkel et al. (2016) found that a *C. albicans* strain isolated from the vaginal tract (M39) was highly virulent in mice, but avirulent in the insect model. Our interpretation that *C. orthopsilosis* ZCF29 does not impact virulence should therefore be treated with some caution.

We have shown that it is possible to use cosine similarity analysis to link high impact variants with specific phenotypes in *C. orthopsilosis*. However, we have explored only a small amount of the genetic variation in the species. The high number of homozygous and heterozygous polymorphisms in hybrid species such as *C. orthopsilosis* (one SNP per ~32 bp), and in particular the high number of apparent indels, CNVs and frameshifts, makes identifying phenotypic-genotypic associations particularly difficult. The approach may therefore be more successful in more homoyzgous species such as *C. parapsilosis* (one SNP per 15 kb).

**ACKNOWLEDGMENTS**

This work was supported by Science Foundation Ireland (12/IA/1343), the Wellcome Trust (097419/Z/11/Z), and the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. H2020-MSCA-ITN-2014-642095.

**LITERATURE CITED**

Amorim-Vaz, S., E. Delarze, F. Ischer, D. Sanglard, and A. T. Coste, 2015 Examining the virulence of *Candida albicans* transcription factor mutants using *Galleria mellonella* and mouse infection models. Front. Microbiol. 6: 367. https://doi.org/10.3389/fmicb.2015.00367

Bergström, A., J. T. Simpson, F. Salinas, B. Barré, L. Parts et al., 2014 A high-definition view of functional genetic variation from natural yeast genomes. Mol. Biol. Evol. 31: 872–888. https://doi.org/10.1093/molbev/msu037

Bonfietti, L. X., M. W. Szeszis, M. R. Chang, M. A. Martins, S. R. B. S. Pukinskas et al., 2012 Ten-year study of species distribution and antifungal susceptibilities of *Candida* bloodstream isolates at a Brazilian tertiary hospital. Mycopathologia 174: 389–396. https://doi.org/10.1007/s11046-012-9566-3

Butler, G., M. D. Rasmussen, M. F. Lin, M. A. S. Santos, S. Sakthikumar et al., 2009 Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459: 657–662. https://doi.org/10.1038/nature08064

Cantón, E., J. Pemán, G. Quindós, E. Eraso, I. Miranda-Zapico et al., 2011 Prospective multicenter study of the epidemiology, molecular identification, and antifungal susceptibility of *Candida parapsilosis*, *Candida orthopsilosis*, and *Candida metapsilosis* isolated from patients with candidemia. Antimicrob. Agents Chemother. 55: 5590–5596. https://doi.org/10.1128/AAC.00466-11

Chen, C.-Y., W.-H. Sheng, S.-Y. Huang, W.-C. Chou, M. Yao et al., 2015 Clinical characteristics and treatment outcomes of patients with candidaemia due to *Candida parapsilosis* sensu lato species at a medical centre in Taiwan, 2000–12. J. Antimicrob. Chemother. 70: 1531–1538. https://doi.org/10.1093/jac/dku540

Chow, B. D. W., J. R. Linden, and J. M. Bliss, 2012 *Candida parapsilosis* and the neonate: epidemiology, virulence and host defense in a unique patient setting. Expert Rev. Anti Infect. Ther. 10: 935–946. https://doi.org/10.1586/eri.12.74

Cingolani, P., A. Platts, L. W. Wang, M. Coon, T. Nguyen et al., 2014 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPeFF: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80–92. https://doi.org/10.4161/fly.19695

Cissé, O. H., L. Ma, D. Wei Huang, P. P. Khil, J. P. Dekker et al., 2018 Comparative population genomics analysis of the mammalian...
analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303. https://doi.org/10.1101/gr.107524.110

Peng, D., and R. Tarleton, 2015 EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. Microb. Genom. 1: e000033. https://doi.org/10.1099/mgen.0.000033

Peter, J., M. De Chiara, A. Friedrich, J.-X. Yue, D. Pfieger et al., 2018 Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature 556: 339–344. https://doi.org/10.1038/s41586-018-0030-5

Przybywa, L. P., T. Németh, A. Gácsér, and T. Gabaldón, 2014 Genome comparison of Candida orthopsilosis clinical strains reveals the existence of hybrids between two distinct subspecies. Genome Biol. Evol. 6: 1069–1078. https://doi.org/10.1093/gbe/evu082

Przybywa, L. P., T. Németh, E. Saus, E. Kisezopolska, E. Hegedűsöv et al., 2015 ‘The genomic aftermath of hybridization in the opportunistic pathogen Candida metapsilosis. PLoS Genet. 11: e1005626 (erratum: PLoS Genet. doi.org/10.1371/journal.pgen.1006202). https://doi.org/10.1371/journal.pgen.1006526

Riccombeni, A., G. Vidanes, E. Proux-Wéra, K. H. Wolfe, and G. Butler, 2012 Sequence and analysis of the genome of the pathogenic yeast Candida orthopsilosis. PLoS One 7: e35750. https://doi.org/10.1371/journal.
pone.0035750

Ropers, J., C. Maurais, D. Diogo, M. Marceyt-Houben, A. Perin et al., 2018 Gene flow contributes to diversification of the major fungal pathogen Candida albicans. Nat. Commun. 9: 2253. https://doi.org/10.1038/s41467-018-04787-4

Sai, S., L. M. Holland, C. F. McGee, D. B. Lynch, and G. Butler, 2011 Evolution of mating within the Candida parapsilosis species group. Eukaryot. Cell 10: 578–587. https://doi.org/10.1083/EC.00276-10

Santos, M. A., and M. F. Tuite, 1995 The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. Nucleic Acids Res. 23: 1481–1486. https://doi.org/10.1093/nar/23.9.1481

Schröder, M. K., S. K. de San Vicente, T. H. R. Prandini, S. Himmel, D. G. Higgins et al., 2016 Multiple origins of the pathogenic yeast Candida orthopsilosis by separate hybridizations between two parental species. PLoS Genet. 12: e1006404. https://doi.org/10.1371/journal.pgen.1006404

Shekhar-Guturja, T., W. A. Tebung, H. Mount, N. Liu, J. R. Köhler et al., 2016 Beaurevericin potentiates azole activity via inhibition of multidrug efflux, blocks C. albicans morphogenesis, and is effluxed via Yor1 and circuitry controlled by Zdr2. Antimicrob. Agents Chemother. AABC0959–16.

Sigwalt, A., C. Caradee, C. Brion, J. Hou, J. de Montigny et al., 2016 Dissection of quantitative traits by bulk segregant mapping in a protoploid yeast species. FEMS Yeast Res. 16: fow056. https://doi.org/10.1093/femsyr/fow056

Talas, F., R. Khalil, T. Miedaner, and B. A. McDonald, 2016 Genome-Wide Association Study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of Fusarium graminearum. Mol. Plant Microbe Interact. 29: 417–430. https://doi.org/10.1094/MPMI-09-15-0218-R

Tan, X., B. B. Fuchs, Y. Wang, W. Chen, G. J. Yuen et al., 2014 The role of Candida albicans SP270 in filamentation, biofilm formation and pathogenesis. PLoS One 9: e94468. https://doi.org/10.1371/journal.
pone.0094468

Tavanti, A., A. D. Davidson, N. A. R. Gow, M. C. J. Maiden, and F. C. Odds, 2005 Candida orthopsilosis and Candida metapsilosis spp. nov. to replace Candida parapsilosis groups II and III. J. Clin. Microbiol. 43: 284–292. https://doi.org/10.1128/ICM.43.1.284-292.2005

Taylor, J. W., S. Branco, C. Gao, C. Hann-Soden, L. Montoya et al., 2017 Sources of fungal genetic variation and associating it with phenotypic diversity. Microbiol. Spectr. 5. FUNK-0057-2016. https://doi.org/10.1128/microbiolspec

Urban, C., X. Xiong, K. Sohn, K. Schröppel, H. Brunner et al., 2005 The moonlighting protein Tsa1p is implicated in oxidative stress response and
in cell wall biogenesis in *Candida albicans*. Mol. Microbiol. 57: 1318–1341. https://doi.org/10.1111/j.1365-2958.2005.04771.x

Vandeputte, P., F. Ischer, D. Sanglard, and A. T. Coste, 2011 In vivo systematic analysis of *Candida albicans* Zn2-Cys6 transcription factors mutants for mice organ colonization. PLoS One 6: e26962. https://doi.org/10.1371/journal.pone.0026962

Wagih, O., M. Usaj, A. Baryshnikova, B. VanderSluis, E. Kuzmin et al., 2013 SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. Nucleic Acids Res. 41: W591–W596. https://doi.org/10.1093/nar/gkt400

Warnes G. R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber et al., 2016 gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.

Wenger, J. W., K. Schwartz, and G. Sherlock, 2010 Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. PLoS Genet. 6: e1000942. https://doi.org/10.1371/journal.pgen.1000942

Wolfé, K. H., and G. Butler, 2017 Evolution of mating in the Saccharomycotina. Annu. Rev. Microbiol. 71: 197–214. https://doi.org/10.1146/annurev-micro-090816-093403

Zhong, Z., T. C. Marcel, F. E. Hartmann, X. Ma, C. Plissonneau et al., 2017 A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. New Phytol. 214: 619–631. https://doi.org/10.1111/nph.14434

Communicating editor: B. Andrews