Role of Ectopic Gene Conversion in the Evolution of a *Candida krusei* Pleiotropic Drug Resistance Transporter Family

Erwin Lamping,*,† Jing-yi Zhu,*† Masakazu Niimi,*‡ and Richard David Cannon*

*Sir John Walsh Research Institute and †Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand, and ‡Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand

ORCID IDs: 0000-0002-4942-4782 (E.L.); 0000-0002-1333-1939 (M.N.); 0000-0002-5398-2066 (R.D.C.)

**ABSTRACT** Gene duplications enable the evolution of novel gene function, but strong positive selection is required to preserve advantageous mutations in a population. This is because frequent ectopic gene conversions (EGCs) between highly similar, tandem-duplicated, sequences, can rapidly remove fate-determining mutations by replacing them with the neighboring parent gene sequences. Unfortunately, the high sequence similarities between tandem-duplicated genes severely hamper empirical studies of this important evolutionary process, because deciphering their correct sequences is challenging. In this study, we employed the eukaryotic model organism *Saccharomyces cerevisiae* to clone and functionally characterize all 30 alleles of an important pair of tandem-duplicated multidrug efflux pump genes, *ABC1* and *ABC11*, from seven strains of the diploid pathogenic yeast *Candida krusei*. Discovery and functional characterization of their closest ancestor, *C. krusei ABC12*, helped elucidate the evolutionary history of the entire gene family. Our data support the proposal that the pleiotropic drug resistance (PDR) transporters Abc1p and Abc11p have evolved by concerted evolution for ~134 MY. While >90% of their sequences remained identical, very strong purifying selection protected six short DNA patches encoding just 18 core amino acid (aa) differences in particular trans membrane span (TMS) regions causing two distinct efflux pump functions. A proline-kink change at the bottom of Abc11p TMS3 was possibly fate determining. Our data also enabled the first empirical estimates for key parameters of eukaryotic gene evolution, they provided rare examples of intron loss, and PDR transporter phylogeny confirmed that *C. krusei* belongs to a novel, yet unnamed, third major Saccharomycotina lineage.

**KEYWORDS** gene duplication and gene conversion; copy number variation; evolution of multi-gene families; PDR transporters; *Candida krusei*

---

**Gene duplication** is one of the major driving forces for the evolution of novel gene function. It enables organisms to rapidly adapt to changing environments (Ohno 1970; Li 1997), and it is also important for speciation (Lynch and Conery 2003). Large proportions (20–60%) of genes in all kingdoms of life are generated by gene duplication (Rubin *et al.* 2000; Li *et al.* 2001; Zhang 2002). The birth rate of duplicated genes in eukaryotes has been estimated to be ~1 gene⁻¹ 100 MY⁻¹ (Lynch and Conery 2000), which is similar to their silent-site nucleotide substitution rate of ~0.25–1.6 nt⁻¹ 100 MY⁻¹ (Li 1997; Lynch and Conery 2000). However, most duplicated genes are inactivated relatively quickly (Walsh 1995). *Saccharomyces cerevisiae* contains duplicated genes that arose from an ancient whole genome duplication event (Wolfe and Shields 1997; Kellis *et al.* 2004), which was recently identified as an ancient interspecies hybridization event (Marcet-Houben and Gabaldon 2015). However, almost 90% of all duplicated genes were lost within a few million years at an estimated half-life of ~1 MY (Lynch and Conery 2000, 2003). The fate of duplicated genes depends on a number of factors. Davis and Petrov (2004) reported preferential duplication of conserved genes, while genes encoding proteins of multi-protein complexes or dosage-sensitive genes were less likely to generate stable duplicates (Papp *et al.* 2003a). A comprehensive study by Wapinski *et al.* (2007) found that duplicated genes typically diverge with respect to regulatory control, but rarely in their biochemical function (Wapinski *et al.* 2007). The authors concluded that gene...
duplication may eventually simplify, rather than increase, the complexity of a system by partial “division of labor.” That is, “subfunctionalization” (SF) enables the optimization of overlapping functions of an ancestral protein (Wapinski et al. 2007). However, in smaller populations, mutation and genetic drift can increase functional complexity without any necessary adaptive benefit to the organism (Lynch 2007). The SF of a component of the fungal V-ATPase proton pump is a case in point (Finnigan et al. 2012).

Gene conversion is a “copy and paste” function between highly similar (≏90%) sequences. It can occur between alleles, typically during meiosis (allelic gene conversion), but also occurs between paralogous, duplicated, genes during mitosis [ectopic gene conversion, EGC (Fawcett and Innan 2011)]. EGC was first recognized as important for maintaining sequence identity between repeat copies of genes within the large rRNA gene cluster, prompting the generation of a model for the evolution of multi-gene families termed “concerted evolution” (Brown et al. 1972). Many examples of multi-gene families (e.g., histones, GPCRs, ubiquitins, immunoglobulins, and MHC genes) supposedly shaped by concerted evolution followed. However, this model does not explain how members of multi-gene families evolve novel functions, and why there are so many inactive pseudogenes (~2–5% of yeast genes (Lafontaine and Dujon 2010) and >10,000 in humans (Zheng et al. 2007)). This prompted the formulation of the “birth-and-death” model, which suggested that new genes are created by gene duplication and some stay in the genome for a very long time, whereas most are inactivated or deleted from the genome (Nei and Rooney 2005). The controversy over which of the two models best describes the evolution of multi-gene families continues (Arguello and Connaillon 2011), partly because neither model explains the biological evidence satisfactorily.

Many factors affect the gene conversion rate such as genome location, the distance and similarity between similar tracts, and the minimum tract length or minimum efficient processing segment (MEPS)—the minimum length of highly similar repeat sequences required for homologous recombination. Most studies estimate the gene conversion rate to be ~10–100 times higher than the synonymous nucleotide substitution rate (Mansai et al. 2011). The outcome of gene conversion can be either beneficial (e.g., rRNA gene family) or harmful, with an increasing number of genetic disorders being ascribed to EGC (Chen et al. 2010).

But how can duplicated genes evolve a novel function (i.e., neo-functionalize) and fix an advantageous mutation(s) in a population when EGC is active? Population genetic theory indicates that strong selection against EGC is required for the stable maintenance of a mutation (i.e., a fate determining mutation, as it provides an evolutionary advantage to the organism) ultimately resulting in the evolution of a novel function (i.e., neo-functionalization (NF)) of a duplicated gene copy (Innan and Kondrashov 2010). Gao and Innan noticed that the battle between these two opposing evolutionary forces (i.e., maintenance of an advantageous mutation vs. frequent EGCs between tandem-duplicated genes) continues for a relatively long time (Gao and Innan 2004). This, ultimately, leads to a local peak of sequence divergence between these tandem-duplicates with low divergence in regions away from the mutation, and the outcome of NF in the face of EGC is a recognizable pattern of highly similar sequences interspersed by highly divergent sequences (Osada and Innan 2008; Innan and Kondrashov 2010). A prominent example is the human red and green opsin genes that have arisen by gene duplication after the split of New and Old World monkeys ~30–40 MYA with only two aa variations of exon 5 accounting for >80% of the functional differences between the red and green eye pigments (Nathans et al. 1986; Neitz et al. 1991; Winderickx et al. 1993; Shyue et al. 1994).

Although EGC, natural mutation, and positive Darwinian selection are important factors shaping the evolution of gene duplicates, and complex mathematical models can predict possible outcomes, more experimental data are needed to improve existing models (Innan and Kondrashov 2010). Such data are hard to obtain because sequencing duplicated genes from a representative sample of individuals from higher eukaryotes is extremely difficult due to their large size and, often, complex gene arrangements, which makes it almost impossible to correctly resolve such repeat sequences (Blow 2015). The discovery of an ancient pair of tandem-duplicated genes (ABC1 and ABC11), encoding drug efflux pumps, undergoing prolonged concerted evolution in Candida krusei provided a rare opportunity to extract actual values for important evolutionary factors shaping the evolution of duplicated genes. Extensive polymorphism data from geographically diverse isolates gave novel insights into the population structure of C. krusei. They enabled direct estimates for the gene conversion rate and the size of the MEPS, and they confirmed “GC-bias” for gene conversion (Galtier 2003; Marais 2003; Noonan et al. 2004) in C. krusei. Intramolecular gene deletions and intermolecular nonhomologous mitotic recombination between tandem-duplicates also caused copy number variations (CNV), and the formation of ABC11-1 or ABC1-11 chimeras with novel pump functions. The discovery of a third close homolog, C. krusei multidrug efflux pump ABC12, and discovery of ABC11 and ABC1 orthologs in the closest sequenced relative, Pichia membranifaciens, were crucial to tentatively infer the evolutionary history of all three C. krusei family members, and provided clues about the amino acid changes in Abc1p and Abc11p that were most likely fate-determining, and caused the evolution of two functionally distinct efflux pumps.

Materials and Methods

Strains and culture conditions

C. krusei isolates and yeast strains used in this study are listed in Supplemental Material, Table S1 in File S1. All S. cerevisiae strains created in this study are based on ADΔ (Lamping et al. 2007). Fungal strains were grown in yeast extract, peptone
and glucose (YPD) medium: 1% (w/v) Bacto-yeast extract (Difco Laboratories, Detroit, MI), 2% (w/v) Bacto-peptone (Difco) and 2% (w/v) glucose. Yeast transformants were selected on plates containing 0.077% (w/v) complete supplement mixture without uracil (CSM-ura; Bio 101, Vista, CA), 0.67% (w/v) yeast nitrogen base without amino acids (Difco), 2% (w/v) glucose, and 2% (w/v) agar (Difco). Plasmids were maintained in Escherichia coli strain DH5α. E. coli cells were grown in Luria-Bertani (LB) medium, to which ampicillin was added (100 μg/ml) as required.

**Materials**

Molecular biology reagents, restriction and modifying enzymes were from New England Biolabs (Beverly, MA) or from Roche Diagnostics N.Z. (Auckland, New Zealand). Lyophilized desalted DNA oligonucleotides (Table S2 in File S1) were purchased from Sigma-Aldrich Pty. (Sydney, Australia).

PCR-amplified DNA fragments were purified using kits from Qiagen Pty. (Clifton Hill, Victoria, Australia). Genomic DNA (gDNA) was isolated from yeast using the Y-DER Yeast DNA Extraction Reagent Kit from Pierce (Rockford, IL), or the DNeasy Tissue Kit from Qiagen. Yeast were transformed using transformation cassette without uracil (CSM-ura; Bio 101, Vista, CA), and 2% (w/v) glucose. Yeast transformants were selected on plates containing 0.077% (w/v) complete supplement mixture without uracil (CSM-ura; Bio 101, Vista, CA), 0.67% (w/v) yeast nitrogen base without amino acids (Difco), 2% (w/v) glucose, and 2% (w/v) agar (Difco). Plasmids were maintained in Escherichia coli strain DH5α. E. coli cells were grown in Luria-Bertani (LB) medium, to which ampicillin was added (100 μg/ml) as required.

**Compounds**

The chemicals and antifungal agents were purchased from the following sources: fluconazole (FLC; Difucan; Pfizer Laboratories, Auckland, New Zealand), itraconazole (ITC; Janssen Research Foundation, Beerse, Belgium), aureobasidin A (AUR; Takara Bio, Shiga, Japan), clotrimazole (CLT; Bayer, Osaka, Japan), miconazole (MCZ), voriconazole (VRZ), rhodamine 6G (R6G), rhodamine 123 (R123), cycloheximide (CHX), nigericin (NIG), trifluoperazine (TFP), doxorubicin (DOX), daunorubicin (DAU), enniatin (ENI), and oligomycin (OLI) (Sigma-Aldrich New Zealand, Auckland, New Zealand). FK506 was a gift from Astellas Pharma (Tokyo, Japan), and the milbemycins α11, α20, β9, and β11 were a gift from Sankyo (Tokyo, Japan).

**Multilocus sequence typing (MLST)**

To determine the diploid sequence type (DST) for each isolate, we used the publicly available C. krusei MLST database (http://pubmlst.org/ckrusei/; Jacobsen et al. 2007). Using the recommended DNA primers, we PCR-amplified parts of the ORFs of six housekeeping genes (HIS3, LEU2, NMT1, TRP1, ADE2, and LYS2) from gDNA extracted from individual C. krusei strains, and sequenced the fragments in both directions with the same primers. Genotypes that were included in the database were numbered accordingly, and new genotypes were given new numbers by extending the count of identified genotypes.

**Genetic manipulations**

Plasmid pBluescriptISK(+) (Stratagene, La Jolla, CA) was used to generate ABC1 and ABC11 subclones. Inverse PCR fragments were amplified from gDNA extracted from different C. krusei isolates as previously described (Lamping et al. 2009). In short, ~1 μg of gDNA was digested to completion with restriction enzymes, purified by phenol extraction followed by ethanol precipitation, and redissolved in 100 μl of sterile milliQ water. The fully digested, purified, gDNA (~50 ng) was then ligated (20 μl reaction) to form circular DNA molecules. Portions of the ligated gDNA mix (1 μl) were used as templates for inverse PCR reactions (50 μl). Inverse forward and reverse PCR primer pairs specific for ABC11 (11-P1, 11-P2, and 11-P3), or ABC1 (1-P3; Figure S1B) were designed to be close to, but pointing away from, each other. The following PCR cycling protocol was used: 94°C for 5 min, (94°C for 20 sec, 55°C for 10 sec and 68°C for 1 min/kb) repeated 34 times, and with a final extension time of 10 min at 68°C. The inverse PCR fragments were gel-purified and further characterized. Some were used to create restriction maps to ascertain which inverse PCR products covered the largest portions of the uncharacterized genome sequence, while others were sequenced directly.

**Verification of an 88 bp intron in ABC11**

The ABC11 intron was verified experimentally by amplifying the ABC11 cDNA ORF from total RNA isolated from B2399 cells grown in YPD medium and harvested at late logarithmic growth. First strand ABC11 cDNA was amplified from total RNA with SuperscriptIII Reverse Transcriptase (Invitrogen New Zealand, Auckland, NZ), and the ABC11-specific primer ABC11-Stop. A portion of this reaction was then used as DNA-template to PCR-amplify the entire ABC11 cDNA ORF, and DNA sequencing confirmed the 88 bp intron.

**Yeast strains overexpressing individual ABC1 and ABC11 alleles with or without an intron**

In order to compare the function of ABC1 and ABC11 alleles with or without an intron, the AB₆-αΔ (Abc1p and Abc12p), the proteins were overexpressed in S. cerevisiae ADΔ. The construction of ABC1 overexpressing yeast strains ADΔ/CkABC1g and ADΔ/CkABC1c was described previously (Lamping et al. 2007). Because ABC1 was toxic to E. coli, and it was impossible to clone and propagate plasmid pABC3-CkABC1 in E. coli, an alternative cloning strategy was developed (Lamping et al. 2007). However, this strategy required multiple PCR steps (>60 PCR cycles) to amplify the entire transformation cassette (see top of Figure S2), which caused a high mutation rate. We therefore developed a new cloning strategy that is fast, highly reliable, and accurate, and could also be useful for cloning and sequencing individual alleles of other large genes. Rather than using multiple PCR steps to create the entire transformation cassette in one piece, we directly transformed ADΔ with separate
PCR fragments that overlapped by 25 bp (center and bottom of Figure S2), and let the individual pieces fuse by homologous recombination. This required four or five separate homologous recombination events, which are exceptionally accurate and efficient in S. cerevisiae. Instead of using the ABC11-cDNA from the above experiment to create ADΔ/CkABC11c, ABC11c was PCR-amplified from gDNA as two separate DNA fragments that overlapped by 25 bp, with the intron eliminated by primer design (see Figure S2). Equimolar amounts (total amounts were ~0.6–1.2 µg) of these three or four (see center and bottom of Figure S2), gel-purified, PCR fragments were used to transform ADΔ. All Ura+ transformants (~10–30) were confirmed by colony PCR, and >80% had the correct DNA sequence. The transformation rate using either three or four separate overlapping PCR fragments was only ~5–10 times lower than the rates achieved for transformation with a single cassette.

Identification of individual alleles of the entire ABC11-ABC1 locus of C. krusei 89021, V2b, 90147, and 89221

To ascertain which allele of ABC11 of strain 89021 was adjacent to which allele of ABC1, the 1.3 kb XhoI/BglII fragment (see Figure S1B) was subcloned, and six individual clones were sequenced. gDNA sequences for the ABC11-ABC1 locus and sequences of individual ABC1 alleles for V2b, 90147 and 89221 were also determined.

Functional analysis of wild-type and recombinant yeast strains

The susceptibility of yeast to antifungal agents was measured by a microdilution assay as described previously (Niimi et al. 2004; Holmes et al. 2006). Agarose diffusion assays were performed to test the susceptibility of S. cerevisiae strains overexpressing C. krusei Abc1p or Abc1p to different xenobiotics. In brief, a 10 ml YPD overnight culture of each test strain was diluted 1:20 into 3 ml CSM medium, and grown at 30°C for another 4 h to mid-logarithmic growth phase (OD600 ~1; ~10^7 cells/ml). The cells were diluted to OD600 = 0.008 in 5 ml of molten CSM medium containing 0.6% agarose (45°C), and overlaid on 20 ml CSM agar base medium. Whatman 3MM paper disks (5 mm diameter) containing xenobiotics were placed onto the solidified top agarose medium, and the plates were incubated at 30°C for 48 h, or until growth inhibitory zones became clearly visible. The amounts of xenobiotics used are determining the toxicity of each drug) or FLC at 0.13× the MIC of FLC (MICFLC). The cell suspension was poured into a rectangular Omnitray plate (126 × 86 × 19 mm; Nunc, Roskilde, Denmark) that contained 20 ml of CSM solidified with 0.6% agarose either without (control), or with FLC. Whatman 3MM paper disks containing potential drug pump inhibitors (1 µg of milbemycins α11, α20, β9, and β11, 0.2 µg ENI, 1 µg beauvericin, 5 µg FK506, and 25 nmol OLI) were placed on the overlay, and the plates were incubated at 30°C for 48 h.

Analysis of phylogenetic relationships and identification of potential yeast transcription factor binding sites (TFBS)

Phylemon 2.0 (Sanchez et al. 2011), a publicly available web-tool, was used for sequence alignments, tree reconstruction, and the estimation of synonymous and nonsynonymous distances. YeTFaSCo (de Boer and Hughes 2012), another web-tool, was used to search for potential yeast TFBSs upstream of ABC1, ABC11, and ABC12. Sequence alignments were performed with ClustalW v2.0.10 (Thompson et al. 1994) and viewed, and, if necessary, edited, with Jalview14.0 (Waterhouse et al. 2009). FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) was used as a graphical viewer of phylogenetic trees and to produce publication-ready figures. Phylogenetic relationships were calculated by maximum likelihood analysis using PhyML v3.0 (Guindon and Gascuel 2003; Anisimova and Gascuel 2006). Evolutionary tests for the synonymous and nonsynonymous substitution rates between pairs of genes were performed with CodeML and yn00 (Yang and Nielsen 2000) from PAML v4.4c (Yang 2007).

Helical wheel projections and determination of lipid facing surfaces (LIPS) of individual TMS

The surfaces of individual TMs that most likely face the lipid bilayer were predicted with the LIPS server (http://tanto.bioengr.uic.edu/lips/; Adamian and Liang 2006) using the sequences of the TMSs of 42 cluster A PDR transporters as input [a list of these transporters and their alignment, with the predicted TMS boundaries highlighted, can be found in Lamping et al. (2010)]. The helical wheel projections of individual TMSs were created online (http://rzlab.ucr.edu/scripts/wheel/wheel.cgi).

GenBank accession numbers

ABC11 and ABC1 sequences: 16.4 kb ABC11-ABC1 gDNA of 89021 (JX679710); 13.9 kb of the ABC11-ABC1 gDNA of 89221 (JX679711) and 90147 (JX679712); 7.7 kb ABC11, including 3.1 kb upstream, gDNA of V2b (JX679713) and 4.9 kb ABC1, including ~100 bp upstream and downstream, gDNA of V2b (JX679714); 10 kb ABC11-ABC1 locus (from ABC11 start to ABC1 stop) of the A allele (FJ445757) and the B allele (FJ445767) of 89021; 10 kb ABC11-ABC1 locus of IFO0011 (KT716056) and B2399 (KT716053); 15.3 kb ABC11-ABC1-11-chimera1-ABC1 locus of 89102 (KT716059) and 15.3 kb ABC11-ABC1-11-chimera2-ABC1 locus of 89102 (KT716059); 5.3 or 4.9 kb ABC1 A and B alleles (including the promoter and terminator) of 89221, 90147,
and V2b (FJ445760-5); 6.3 kb HindIII/BamHI fragment and ABC1 cDNA of B2399 [DQ903906-7 (Lamping et al. 2009)]; 4654 bp ABC11-1 chimera of IFO0011 (KT716057), ABC1-1 chimera1 (KT716054) and ABC11-1 chimera2 (KT716055) of B2399, and ABC11-1 chimera of 89102 (KT716060). ERG11 sequences: ERG11 A (FJ445752) and ERG11 B (FJ445753) alleles of 89102.

Data availability

Strains are available upon request. Sequence data are available at GenBank and the accession numbers are listed above in the Materials and Methods section. File S1 contains Tables S1–S4, and detailed descriptions of all Supplemental Material (i.e., legends for Figure S1, Figure S2, Figure S3, Figure S4, Figure S5, Figure S6, Figure S7, Figure S8, Figure S9, Figure S10, Figure S11, and Figure S12).

Results

Characterization of a representative set of C. krusei isolates

Although C. krusei, also known as Issatchenka orientalis or Pichia kudriavzevii, is a relatively poorly studied ascomycetous yeast, it is important for natural food fermentation (Chan et al. 2012), especially of cocoa beans for chocolate production. It is also an opportunistic fungal pathogen of humans that has reduced susceptibility to azole antifungals due, in part, to pump-mediated drug efflux (Lamping et al. 2009). Odds and colleagues developed a MLST database for clinical C. krusei isolates from all parts of the world (Jacobsen et al. 2007). Their data indicated that C. krusei is diploid (there was no mention of aneuploid or triploid strains), and there was no geographical association for particular subtypes. The high proportion (77%) of singletons (i.e., unique diploid sequence types; DSTs), despite almost saturated levels of haplotypes, was taken as a sign that sexual recombination may play an important part in the C. krusei life cycle (Jacobsen et al. 2007).

The seven isolates used in the present study comprise a diverse range of isolates from humans, animals, and the environment; isolates from three different continents, different body sites, and different years of isolation (see Table S1 in File S1). MLST confirmed that strain 89021 was closely related to 90147 and V2b, and, to a lesser extent, 89221, whereas the genotypes of 89102, IFO0011, and B2399 were somewhat different (see Table S3 in File S1). As expected from Odds and colleagues’ study (Jacobsen et al. 2007), only six (out of 42; ~14%) new haplotypes were identified, yet six of the seven isolates (~86%) had unique DSTs (see Table S3 in File S1). Interestingly, New Zealand isolate 90147 had a DST (63) that was identical to an isolate from Italy (see Table S3 in File S1). The peak height ratios of single nucleotide polymorphisms (SNPs) from sequencing chromatograms easily distinguished a diallelic (1:1 ratio) from a triallelic (1:2 ratio) sequence. All SNPs of 89021, 89221, 90147, V2b and IFO0011 had 1:1 peak ratios. However, all B2399 SNPs had 1:2 peak ratios, whereas 89102 exhibited 1:2 peak ratios for ERG11 and ABC1, but 1:1 peak ratios for all housekeeping genes (see Table S4 in File S1). This implied that C. krusei B2399 was triploid, 90102 was aneuploid with a trisomy in the ERG11- and ABC1-containing chromosome, and 89021, 89221, 90147, V2b, and IFO0011 were diploid, which was confirmed by measuring the fluorescence of cells stained with the DNA intercalating agent SYBR green (see Figure S3).

Discovery of the tandem-duplicated multidrug efflux transporter pair ABC11-ABC1 of C. krusei 89021

We previously characterized 6.3 kb (between the dashed vertical arrows; Figure 1A) of C. krusei B2399 gDNA containing the major multidrug efflux pump ABC1 (Lamping et al. 2009). Overexpression of Abc1p in S. cerevisiae AΔ caused multidrug resistance that could be reversed with known efflux pump inhibitors (Lamping et al. 2007, 2009). But, although C. krusei B2399 was triploid, averaging one SNP/400 nt for the ~10 kb ERG11 locus (Lamping et al. 2009), ABC1 of C. krusei B2399 had no detectable SNPs (Lamping et al. 2009). Southern blot results for C. krusei B2399 gDNA, however, gave two unexpected EcoRI bands, indicating allelic variation or a recent tandem-duplication of ABC1 (Lamping et al. 2009). This led ultimately to the discovery of the tandem-duplicated ABC11-ABC1 locus of C. krusei 89021 (Figure 1A). Both ends of the 16.4 kb sequence contained partial ORFs (996 and 973 bp, respectively) of unknown function, and an additional 1512 bp ORF with high homology to the MATE (Multidrug And Toxic compound Extrusion) family of drug/sodium antiporters ~2 kb upstream of ABC11 (Figure 1A). ABC11 and ABC1 of 89021 had 19 and 16 SNPs, respectively (i.e., an average SNP frequency of one SNP/532 nt; Figure 1B). The SNP frequency for the noncoding regions was ~2.3 times higher (one SNP/244 nt), similar to the previously reported ~threefold difference between coding and noncoding regions for the 10 kb ERG11 locus of B2399 (Lamping et al. 2009). The average SNP frequency of one SNP/437 nt for the entire 16.4 kb was comparable with diploid Candida CTG clade pathogens such as C. albicans (one SNP/330–390 nt) or C. tropicalis (one SNP/576 nt; (Butler et al. 2009)). Interestingly, ABC1 and ABC11 also shared one identical intron (gray box; Figure 1A).

Sequence identity interspersed by four short stretches of sequence divergence distinguishes ABC1 from ABC11

We determined the sequences of 22 ABC1 and ABC11 alleles for the seven C. krusei strains (Figure 2) in order to identify the minimum (core) differences between ABC1 and ABC11. Strains B2399, IFO0011 and strain 89102 had only one ABC11 and one ABC1 allele each (Figure 2B). However, the remaining four diploid C. krusei strains 89021, V2b, 90147, and 89221 each had two ABC1 and two ABC11 alleles (Figure 2A). In fact, all their ABC11-A, ABC1-A, and ABC1-B alleles were identical, apart from 89221 ABC1-B, which had one additional synonymous SNP (sSNP) in patch 1 (Figure 2A), and strain 89221, which had the two SNPs immediately
upstream of patch II of ABC11-A on Chr A homogenized with ABC1 sequence by a very recent EGC event (black arrow; Figure 2A). In contrast, the four ABC11-B alleles differed slightly from each other (Figure 2A). Overall, there were 92 core nucleotide differences (plus one extra codon near the stop codon of ABC11; see Figure S4) between ABC11 and ABC1 limited to six short (415 nt or 8.9% of the ORF) DNA stretches I–VI colored blue and green, respectively, in Figure 2. The five “fixed” SNP differences (blue and green SNPs; Figure 2) in both ABC11 and ABC1 near patches II, IV, and VI found in most strains were excluded from the core differences because they were homogenized by four EGCs in strains 89221, IFOO011 and B2399, respectively (all in the direction from ABC1 to ABC11; black arrows in Figure 2). The six DNA patches I–VI that encoded the core differences between ABC11 and ABC1 were conserved in all but one of the 22 ABC11 and ABC1 alleles, with ABC11 of strain 89102 having adopted ABC1-specific patches III and V by two separate EGCs from the neighboring ABC1 allele (Figure 2B). It would appear that ABC11 patches III and V are possibly less important for the biologically preserved Abc11p function.

**ABC1 and ABC11 are under very strong purifying selection and have evolved by concerted evolution for ~134 MY**

The vast majority of SNPs between individual alleles of ABC1 and of ABC11 were synonymous. Strain 89021, for example, had 16 ABC1 and 19 ABC11 SNPs, 15 and 16 of which were sSNPs, respectively (short vertical lines; Figure 1B). The non-synonymous to synonymous nucleotide substitution ratio dN/dS (ω) of 0.052 for ABC1 and 0.053 for ABC11 of 89021 indicated that both genes were under very strong purifying selection. Even the ω value for the concatenated 252 nt of patches II–IV was not much higher (0.101; Table 1), suggesting that Abc1p and Abc11p subfunctionalized rather rapidly after duplication, and have since been under very strong purifying selection. The ω value for the concatenated 159 nt of patches I and VI, encoding the N- and C-termini, respectively, was much higher (0.347), suggesting that these 53 aa were of little functional consequence and, thus, under reduced purifying selection. ABC11 was 97.8% identical to ABC1, with a calculated dS of 0.05 for the entire ORF, suggesting a rather recent duplication event ~3.2 MYA (Table 1). However, dS within “protected” patches II–IV was much higher (1.18; Table 1), indicating that the duplicates were possibly much older (~71 MY; Table 1).

This rather high estimate of ~71 MY for a 97.8% identical tandem-duplicated pair of paralogs had to be further revised with the discovery of the *P. membranifaciens* ABC11-ABC1 tandem-duplicate. The synteny of the ABC11-ABC1 locus between *C. krusei* and *P. membranifaciens* was well conserved, which strongly suggested common ancestry (Figure 3A). The *P. membranifaciens* ORFs were also closely linked (separated by just 548 nt), and PmfABC1p and PmfABC11p each were 79% identical, and 87 and 89% homologous to CkAbc1p and CkAbc11p, respectively. Sequence comparison of PmfABC1 (391367–395992 of scaffold2-contig42) with PmfABC11 (386144–390818 of scaffold2-contig42; Figure 3B) clearly showed how parts of these two paralogs are, like *C. krusei* ABC1 and ABC11, still evolving by concerted evolution, but in a different fashion (cf. Figure 3, B and C). PmfABC1 and PmfABC11 had a few shorter identical sequence patches in regions encoding the less conserved nucleotide-binding domain 1 (NBD1), but alignments of the 3’ halves of PmfABC1 with PmfABC11 encoding the highly conserved NBD2 and the TMD2 regions showed a somewhat similar pattern to CkABC1/CkABC11 (Figure 3, B and C). As with *C. krusei*, the estimated age for this pair of paralogs varied dramatically depending on which dS values were used for the calculations. Using dS for the entire ORF (0.41) gave an estimated age of

---

**Figure 1 Discovery of the tandem-duplicated multi-drug efflux pump pair ABC11-ABC1, and polymorphism data for the ABC11 and ABC1 alleles of *C. krusei* 89021.** ABC11 and ABC1 gene sequences are blue and green, respectively, and short and tall vertical lines are synonymous (sSNPs) and non-synonymous SNPs (nSNPs) between the alleles of each gene, respectively. The two vertical dashed arrows in (A) indicate previously characterized sequence (Lamping et al. 2009). (A) The 16.4 kb locus surrounding the *C. krusei* 89021 tandem-duplicated gene-pair ABC11-ABC1. ORFs and their orientation are indicated; gray boxes within ABC1 and ABC11 represent the common 88 bp intron. (B) A SNP map of ABC11 and ABC1 for the two chromosomes (Chr) of the diploid *C. krusei* 89021. Allelic differences between the ORF sequences are shown as magenta (Chr A) and purple (Chr B) SNPs, respectively. The map underneath shows the polymorphism pattern for the diploid sequence with red SNPs indicating polymorphisms that are the same in both ABC11 and ABC1 and black SNPs representing unique polymorphisms.
26 MY, the dS for the C-terminal half (0.15) gave an even lower estimate of ~10 MY, but dS for the N-terminal half (0.75), which has almost completely escaped concerted evolution, gave an age of 46 MY (Table 1). The most accurate age-estimate for the duplication event of this pair of paralogs was obtained, however, when we compared the two pairs of orthologs with each other (Figure 3, D and E), because they had evolved independently from each other after speciation. There was surprisingly low (~76%) sequence homology across the entire ORF (Figure 3, D and E) reflected in surprisingly high dS values for the ABC1 (2.13) and ABC11 (2.01) orthologs, providing an average age estimate of 134 MY since speciation (Table 1).

Thus, we conclude that the majority of C. krusei ABC1 and ABC11 evolved by concerted evolution for possibly >134 MY; however, patches II-V of C. krusei ABC11–ABC1 were strongly protected against frequent EGCs for ~70 MY in order to preserve important functional differences between the two gene products. This hypothesis was supported by the fact that the majority of SNPs between patches I-VI of ABC1 and ABC11 were shared polymorphisms (10 red SNPs; Figure 1B)—a characteristic feature for the concerted evolution of duplicated genes (Innan 2003; Osada and Innan 2008; Arguello and Connallon 2011).

Intramolecular and intermolecular (non)homologous recombination events caused gene deletions and CNVs, and generated a number of ABC11-1 chimeras

The chromosomal arrangements for ABC1 and ABC11 of B2399, IFO0011 and 89102 were difficult to unravel because of apparent CNVs. After careful consideration, we were able to successfully amplify, clone, sequence, and functionally express the “hidden” gene copies in S. cerevisiae ADΔ. These copies comprised two ABC1–11 and four ABC11–1 chimeras (Figure 4A), which is why these “hidden” genes could not be PCR amplified with ABC1- or ABC11-specific primer pairs. ABC11 and ABC1 of IFO0011 and B2399 were very similar to the A alleles on Chr A of all other diploid strains (Figure 2).
A different, intermolecular, nonhomologous mitotic recombinant, with all three CkABC1-1626 E. Lamping of three different DNA stretches of sequence identities between of three different DNA stretches of sequence identities be-
tween of three different DNA stretches of sequence identities be-
tween of three different DNA stretches of sequence identities be-
venal allele types (see Figure S6C).

Table 1 Silent-site nucleotide substitution rates dS of C. krusei 89021 and P. membranifaciens ABC1, ABC11, and ABC12, and estimated times of duplication

| Sequences Compared         | NBDs-TMDs | Patches II–IV |
|----------------------------|-----------|---------------|
| dS | dN | dN/dS | dS | dN | dN/dS | T (MY) |
|---------------------------------|------------|---------------|
| CkABC1 - CkABC11              | 0.05        | 0.009        | 0.191 | —    | — | — | — | 3.2 |
| PmfABC1 - PmfABC11             | 0.41        | 0.061        | 0.150 | —    | — | — | — | 71 |
| CkABC11 - PmfABC11             | 0.16        | 0.040        | 0.261 | —    | — | — | — | 66 |
| CkABC1 - CkABC12               | 0.23        | 0.038        | 0.160 | —    | — | — | — | 130 |
| CkABC11 - CkABC12              | 0.24        | 0.038        | 0.160 | —    | — | — | — | 130 |
| CkABC1 - CkABC1                | 0.23        | 0.038        | 0.160 | —    | — | — | — | 130 |
| CkABC11 - CkABC1               | 0.24        | 0.038        | 0.160 | —    | — | — | — | 130 |
| ScPDR5 - ScPDRDR15             | 1.55        | 0.152        | 0.098 | —    | — | — | — | 100 |
|                                 | —           | —            | —     | —    | — | — | — | 100 |

a Estimated time since duplication or speciation (when comparing ORFs of C. krusei with P. membranifaciens). ScPDR5 and ScPDRDR15 are ohnologs of a whole genome duplication event ~100 MYA (Wolfe and Shields 1997; Kellis et al. 2004). The times of retrotransposition of CkABC12 (~149 MYA; i.e., from between 131 to 156 MYA), and duplication of C. krusei and P. membranifaciens ABC1/11 (~134 MYA; i.e., from between 130 to 137 MYA) from their respective ancestors was estimated by comparing the silent-site substitution rates dS for the concatenated sequences of the conserved regions encoding NBD1-TMD1, NBD2-TMD2, or both NBDs and TMDs (NBDs-TMDs; the concatenated sequences included regions encoding amino acids Q99-E823 and N847-S1437 of sequences of the conserved regions encoding NBD1-TMD1, NBD2-TMD2, or both i.e. ScPDR5/C24). The polymorphism patterns in Figure 1B suggested that ten recent EGCs (Figure 5A) between individual patches of sequence divergence between physically linked ABC11 and ABC1 gene copies of C. krusei 89021 homogenized an entire tract in one event rather than in multiple smaller EGC events. This conclusion is based on the observation that tracts exposed to frequent EGCs between strongly protected patches of sequence divergence contained either almost exclusively shared polymorphisms or allele-specific SNPs randomly distributed across entire tracts (Figure 1B). If, for example, the most recent EGC homogenized only half a tract between ABC11 and ABC1, then one half would be expected to contain shared polymorphisms only (red SNPs), and the other half a number of allele-specific SNPs (black SNPs), but that clearly was not the case. In addition, as mentioned above, there was also no noticeable interference with the observed polymorphism patterns by allelic gene conversions or meiotic crossover events.

However, both strains had lost one gene copy each, probably via intramolecular homologous recombination between one of three different DNA stretches of sequence identities between ABC11 and ABC1 on IFO0011 ChrB and B2399 ChrA’ and ChrB (Figure 4, A and B). This generated three different ABC11-1 chimeras, either possessing the majority of the 10 shared A or the 10 shared B allele-specific SNPs of strain 89021 (see Figure S5), with all three ABC11-1 chimeras being put under the control of the ABC11 promoter (Figure 4A). A different, intermolecular, nonhomologous mitotic recombination event between patches I and II of ABC1 on ChrA and ABC11 on ChrB in the ancestor of 89102 (Figure 4C) is likely to have generated a single ABC11-1 chimera on ChrA, and an additional ABC1-11 chimera between ABC11 and ABC1 on ChrB (Figure 4A). A more detailed description of how 89102 possibly obtained its complex chromosomal arrangement since that initial intermolecular, nonhomologous, recombination event can be found in Figure S6B.

The fact that all physically linked ABC11 and ABC1 ORFs grouped into either A or B allele types (see Figure S6C) clearly demonstrated that intermolecular meiotic crossovers and allelic gene conversions between neighboring copies of the ~10 kb ABC11-ABC1 locus were much rarer than EGCs; otherwise, we would not have found such clear demarcations of almost exclusively shared (red) or allele-specific (black; Figure 1B) SNPs between individual tracts of sequence divergence, and, most importantly, we would have expected a significant reduction in the SNP frequency in parts that had experienced a recent allelic gene conversion, which clearly was not the case—random SNP distribution between individual alleles of ABC11 and ABC1 is evident in Figure 1B.

First direct estimates for the gene conversion rate and the size of the MEPS

Polymorphism-based methods can overcome the methodological limitations that preclude precise estimates for the rate of EGC (Arguello and Connallon 2011). Therefore, we exploited the sequence information we had collected from a sample of strains representative for the C. krusei population to obtain the first “direct,” empirical, estimates for the gene conversion rate without relying on computer modeling.

The polymorphism patterns in Figure 1B suggested that ten recent EGCs (Figure 5A) between individual patches of sequence divergence between physically linked ABC11 and ABC1 gene copies of C. krusei 89021 homogenized an entire tract in one event rather than in multiple smaller EGC events. This conclusion is based on the observation that tracts exposed to frequent EGCs between strongly protected patches of sequence divergence contained either almost exclusively shared polymorphisms or allele-specific SNPs randomly distributed across entire tracts (Figure 1B). If, for example, the most recent EGC homogenized only half a tract between ABC11 and ABC1, then one half would be expected to contain shared polymorphisms only (red SNPs), and the other half a number of allele-specific SNPs (black SNPs), but that clearly was not the case. In addition, as mentioned above, there was also no noticeable interference with the observed polymorphism patterns by allelic gene conversions or meiotic crossover events.

Given the clarity of these results, we applied a molecular clock (assuming an estimated nucleotide substitution rate of 0.8 per silent-site per 100 MY for fungi (Lynch and Conery 2000), and one, rather than multiple, gene conversions per tract) to estimate the age of these 10 most recent EGC events (Table 2). The age of the 10 most recent EGC events was calculated by counting the number of sSNPs that had accumulated between individual patches since the last EGC and calculating dS. However, these age estimates could be up to ~8 times higher or ~2 times lower, because the nucleotide substitution rates can range anywhere from ~0.1 to ~1.6 per silent-site per 100 MY in different organisms (Lynch and Conery 2000). The paucity of new SNPs in EGC tracts reflects the higher rate of EGC than natural mutation. With these limitations in mind our best estimates for the average age of the four recent EGCs between the two largest tracts (1.2 and 1.7 kb; i.e., EGCs 3 and 8 and EGCs 1 and 6; Figure 5A) were 0.43 and 0.13 MY, respectively, while the calculated average age for the three smaller tracts (339–399 bp; i.e., EGCs 2 and 7, EGCs 4 and 9, and EGCs 5 and 10; Figure 5A)
ranged from 0.53 to 1.68 MY (Table 2). Assuming a constant

gene conversion rate, these equated to 30–385 gene conver-
sions per gene per 100 MY (Table 2). Because all five tracts

between the “protected” patches of sequence divergence of

the two paralogs, ABC1 and ABC11, were separated by the

same distance (~5.3 kb), there appeared to be little

Table 2 Estimation of the gene conversion rate between patches of individual alleles of the ABC11-ABC1 locus of C. krusei 89021

| DNA Between | Tract Length (bp) | sSNPs | sSNPs | Y (10^-6) Since Last Gene Conversion | Average Time Since Last Gene Conversion (MY) | Gene Conversion Rate/Gene/100 MY |
|-------------|-------------------|-------|-------|-------------------------------------|----------------------------------------------|---------------------------------|
| I–II        | 1650              | 492   | 0     | 0.13 < 0.13                         | 0.13                                         | 385                             |
| III–IV      | 1209              | 360   | 4     | 0.17 < 0.17                         | 0.69                                         | 116                             |
| V–VI        | 399               | 117   | 1     | 0.53 < 0.53                         | 0.53                                         | 94                              |
| IV–V        | 375               | 112   | 5     | 0.56 2.79                           | 1.68                                         | 30                              |
| II–III      | 339               | 93    | 1     | 0.67 < 0.67                         | 0.67                                         | 75                              |

The 88 bp intron was removed for this analysis.

* Ns is the calculated number of possible synonymous sites.

* These are the numbers of sSNPs in the A or B alleles between individual patches of ABC11 and ABC1.

* The time since the last gene conversion event was calculated by using the synonymous nucleotide substitution rate (dS) as a molecular clock, assuming an estimated dS of 0.8 per synonymous site per 100 MY for fungi (Lynch and Conery 2000) (Ns is twice the above listed value because one gene conversion event “homogenizes” the sequences of both paralogs, ABC1 and ABC11).
interference with EGC by allelic gene conversion, and, because the effect of genome location could be ignored (as all events occurred in the same location), we could tentatively investigate the influence of tract length on the EGC rate in \textit{C. krusei}. Despite the limited number of EGC events there was an inverse correlation between the average age of the most recent EGC and its tract length (Figure 5C). Extrapolating the graph in Figure 5C to \( y = 0 \) gave a direct estimate for the MEPS of 106 nt for duplicated loci separated by 5.3 kb in \textit{C. krusei}.

**Gene conversion in \textit{C. krusei} shows GC-bias**

There is substantial evidence that frequent EGCs lead to increased GC-content, especially at the third positions in codons (GC3) that are under little selective constraint, due to GC-bias in mismatch repair (Galtier 2003; Noonan et al. 2004). Others, however, have found little evidence for GC-bias due to mismatch repair, and argued that GC-content may favor EGC rather than GC being GC-biased (Casola et al. 2010; Arguello and Connallon 2011). The average GC3-content of those parts of \textit{ABC1} and \textit{ABC11} that were exposed to frequent EGCs was significantly higher (~50%) than the GC3-content (~35%) for patches I–VI that were strongly protected against frequent EGCs (Figure 5D). The GC3-content of patches I–VI was similar to the GC-content (38.3%) for the entire \textit{C. krusei} M12 genome (Chan et al. 2012), and, most importantly, the GC3-content of all neighboring genes (Figure 3A), and all \textit{C. krusei} PDR transporters including \textit{ABC12}, which shares a common ancestor with \textit{ABC1} and \textit{ABC11} (see below), were equally low (~38%) (data not shown). Our data therefore favor a model of GC-bias of mismatch repair for EGC in \textit{C. krusei}.

**Elucidation of the ABC11-ABC1 locus**

\textit{ABC1} and \textit{ABC11} belong to cluster A of the nine clusters (A–H2) of fungal PDR transporters (Lamping et al. 2010). We searched recently released draft genome sequences of \textit{C. krusei} M12 (Chan et al. 2012), SD108 and NBRC 1279T, environmental isolates from Malaysia, the United States and Russia, respectively (see Table S1 in File S1), for all possible PDR transporter homologs, and identified three cluster A (\textit{ABC1}, \textit{ABC11} and \textit{ABC12}), six cluster D (two \textit{S. cerevisiae} Pdr12p and four \textit{S. cerevisiae} Snq2p homologs), and one cluster F member (data not shown). No \textit{ABC1} or \textit{ABC11} ORF sequences were found in NBRC 1279T, although contig 614 contained ~9 kb DNA upstream of \textit{ABC11}, and included \textit{ABC11} patch I, and contig 294 contained ~6 kb DNA downstream of \textit{ABC11}, and included \textit{ABC11} patch VI (see Figure S7). However, five M12 contigs and five SD108 scaffolds contained ~50% of \textit{ABC1} and \textit{ABC11} including ~14 kb (contig 327) and ~113 kb (scaffold 21) of \textit{ABC11} upstream DNA and ~6 kb (contig 351) and ~41 kb (scaffold 21) of \textit{ABC11} downstream DNA (see Figure S7). SD108 scaffold 21, the longest contig (161 kb), contained a predicted \textit{ABC11}-I chimera with an ~600 bp N-terminal insertion following \textit{ABC11} patch I (see Figure S7). Despite some unresolved regions of M12, the available \textit{ABC1} and \textit{ABC11} ORF sequences matched the A alleles of 89021 (100% identity), whereas the available SD108 sequences matched the B alleles, with only 8 nt differences to the B alleles of 89021, all of which were also detected in some B alleles of the other six \textit{C. krusei} strains (data not shown). Scaffold 21 also contained the \textit{C. krusei} azole drug target \textit{ERG11} just 59.9 kb upstream of the \textit{ABC11}-I chimera, confirming our observation that \textit{ERG11}, \textit{ABC1}, and \textit{ABC11} are closely linked on the same chromosome.

**Discovery of \textit{ABC12}**

We named the third cluster A PDR transporter discovered in another part of the \textit{C. krusei} genome Abc12p, because it showed highest homology (84%) to Abc1p and Abc11p. All three recent \textit{C. krusei} genome databases agreed on the chromosomal assembly for \textit{ABC12} encoded on M12 contig 30, NBRC 1279T contig 223, and SD108 scaffold 33. \textit{ABC12} of \textit{C. krusei} M12 had only four or eight SNPs compared to \textit{ABC12} of SD108 or NBRC 1279T, respectively, with just 1 aa difference in Abc12p between M12 and SD108 (H1424R) or NBRC 1279T (K781E). Interestingly, \textit{P. membranifaciens} had only two cluster A PDR transporters,
PmfABC1 and PmfABC11, indicating it had lost its ABC12 ortholog. Comparing a map for the C. krusei ABC12 locus with a map of the same chromosomal region of P. membranifaciens (Figure 6A) indicated how P. membranifaciens may have lost ABC12. There was an inversion of ~20 kb (gray; Figure 6A) immediately upstream of C. krusei ABC12 (pink; Figure 6A), and ABC12 may have been lost during this inversion event (PmfABC12 is the only missing ORF at this locus; Figure 6A). Interestingly, three of the seven C. krusei strains (89021, 89221, and 90147) had only one intact copy of ABC12, with the second allele being an inactive pseudogene [i.e., a C2275T mutation caused Q759 (CAA codon) to become a TAA stop codon].

**C. krusei belongs to a recently discovered third major Saccharomycotina lineage**

Although yeast comprise a broad phylogenetic range of species (Kurtzman 2011), research within the large Saccharomyces family has primarily focused on members of the Saccharomycotina family and the Candida CIG lineages (Tsui et al. 2008; Dujon 2010). This is why the third major Saccharomyces lineage has only recently been discovered.
support values of 100 replicates are shown above individual branches, and branches with 100% bootstrap support are marked with an asterisk. (C) Predicted topology of the Abc1p and Abc11p TMDs with the location of the 18 core aa differences within patches II, III, IV, and V highlighted. The gray background represents the membrane bilayer. Open boxes represent TMS1–6. (A) Physical maps of the ABC12 loci of C. krusei and P. membranifaciens, and a topology map of aa differences between multidrug efflux pumps C. krusei Abc1p and Abc11p. (A) Physical maps of the ABC12 loci of C. krusei (top) and P. membranifaciens (bottom). ORFs are named according to orthology with the S. cerevisiae genome database (SGD) where possible. The C. krusei ABC12 assembly is a consensus between M12 contig 30 (ABC12 = 56553–61015; Abc12p = 1486 aa), SD108 scaffold 33 (ABC12 = 62859–67319) and NBR1 1279T contigs 223 (ABC12 = 39451–43911) and 259 (3′ end from DST1). The P. membranifaciens locus was within the ~153 kb scaffold2-contig136 contig (from 56000 to 87000). There was no ABC12 homolog in P. membranifaciens. It appears that P. membranifaciens lost ABC12 (pink ORF in C. krusei) during an inversion of the sequence (~18 kb) highlighted in gray just upstream of C. krusei ABC12. (B) Maximum likelihood tree for C. krusei and P. membranifaciens Abc1p and Abc11p and C. krusei Abc1p and Abc11p and their phylogenetic relationship to the cluster A PDR transporters of representative species of the three major Saccharomycotina lineages: Saccharomycetaceae (green) species S. cerevisiae (Sc) and C. glabrata (Cg), Candida CTG (magenta) species C. albicans (Ca) and Pichia stipitis (Ps), and the third major Saccharomycotina lineage (blue) species H. parapolymorpha (Hp), P. membranifaciens and C. krusei. Y. lipolytica cluster B PDR transporter YliQ6cET9 was used as the outgroup. The bootstrap confidence values of 100 replicates are shown above individual branches, and branches with 100% bootstrap support are marked with an asterisk. (C) Predicted topology of the Abc1p and Abc11p TMDs with the location of the 18 core aa differences within patches II, III, IV, and V highlighted. The gray background represents the membrane bilayer. Open boxes represent TMS1–12 (from left to right), with ELs and ILs connecting individual TMSs in yellow and blue, respectively. The regions corresponding to patches II–V are in gray boxes, and amino acids that differed between Abc1p and Abc11p of C. krusei B2399 are shown as circles. Sequence comparison with their common ancestor Abc12p revealed which differences between Abc1p and Abc11p were due to changes in Abc1p (two green circles), Abc11p (13 blue circles), or in both Abc1p and Abc11p (three red circles). Amino acids with thick circles are unique changes when compared with all other fungal cluster A PDR transporters.

(Morales et al. 2013; Hittinger et al. 2015; Riley et al. 2016). It includes the methylotrophic Pichia pastoris (De Schutter et al. 2009) and Hansenula polymorpha (Ravin et al. 2013), the nitrate-assimilating yeasts H. polymorpha (Ravin et al. 2013), Dekkera bruxellensis (Curtin et al. 2012), and Kuraishia capsulata (Morales et al. 2013), and also C. krusei and P. membranifaciens (Riley et al. 2016). It is now clear that the Saccharomycetaceae lineage is the oldest branch, and the Candida CTG clade and the third major Saccharomycotina lineage separated later (Morales et al. 2013; Riley et al. 2016). A ClustalW-alignment of all cluster A PDR transporters of five Candida CTG, seven Saccharomycetaceae and two species (H. polymorpha and C. krusei) of the third Saccharomycotina lineage is shown in Figure S8. Abc1p, Abc11p, and Abc12p, as well as H. polymorpha HpCdr2p and HpCdr4p, had two characteristic 3-aa deletions at the center of extracellular loop EL3 and the beginning of EL6 (see Figure S8 and Figure S9). A maximum likelihood tree of all cluster A PDR transporters from two or three species of each of the three Saccharomycotina lineages matched their species tree (Morales et al. 2013; Riley et al. 2016) with good bootstrap support (Figure 6B). These findings confirmed that C. krusei and P. membranifaciens belong to a third, only recently discovered, major Saccharomycotina lineage.

**Intron loss in ABC12 and deintronization in ABC1 and ABC11**

The ClustalW alignment of ABC sequences (see Figure S8) highlighted a 30-aa N-terminal insertion in C. krusei Abc1p and Abc11p, absent in all other homologs (Figure 7A shows an alignment of this region for C. krusei and P. membranifaciens Abc1p, Abc11p, and C. krusei Abc12p). We hypothesized that the common ancestor of ABC1, ABC11, and ABC12 had at least two introns, which were lost in ABC12 by retrotransposition into another part of the genome, while ABC1 and ABC11 converted (i.e., deintronized) one intron into coding sequence, but kept the second intron intact. To test this hypothesis, we searched the DNA sequences encoding the 15–30 aa insertion in ABC1 and ABC11 for remnants of the conserved yeast splice site (ss) signals [i.e., 5′ ss, 3′ ss, and branch site (BS) signal; (Kupfer et al. 2004)]. There was reasonable homology between ABC1/ABC11 and ABC12 before and after the predicted intron boundaries, particularly...
Figure 7 Evidence for a possible deintronization event in the ancestor of *C. krusei* and *P. membranifaciens* ABC1 and ABC11, and evidence for intron loss by retrotransposition in *C. krusei* ABC12. Identical sequences are highlighted in dark blue and residues that are >80% conserved in (A) are in light blue. (A) Alignment of the N-termini of Abc1p, Abc11p, and Abc12p revealed a 30-aa insertion (the area highlighted in gray) in CkAbc1p and CkAbc11p, and slightly shorter 15 and 24 aa insertions in PmfAbc1p and PmfAbc11p, respectively. (B) Alignment of the DNA sequences encoding the amino acids boxed in (A). Remnants of the 5' and 3' splice site (ss; dark gray) and branch site (BS) signals (underlined) were clearly visible in CkABC1 and CkABC11; PmfABC1 and PmfABC11 may have experienced further deletions (i.e., 45 and 18 nt, respectively) in this region to accommodate the possibly detrimental functional effects of the initial deintronization event. Changes to the conserved ss signals in CkABC1 and CkABC11 are highlighted with bold-type letters underneath the alignment. Deintronization of the "ancient" 92 bp intron may have caused the formation of a pseudogene, the function of which was possibly recovered by the 10 bp insertion (marked in gray) just 21 nt downstream. (C) Sequence alignment of *C. krusei* and *P. membranifaciens* ABC1 and ABC11 with *C. krusei* ABC12 surrounding the ABC1/ABC11 intron [the position in the protein where the intron is inserted is marked with an arrow in (A)] confirmed the loss of this "ancient" intron in ABC12. Conserved 5' ss, 3' ss and BS signals are marked in dark gray.
for *C. krusei ABC1* and *ABC11*, and splice site remnants were still visible (Figure 7B). The predicted 5’ ss GTAG was changed to ACACG, the BS TACTAAC was changed to TTCAAAC, and the potential 3’ ss TTNTTAG was changed to ATNN in *C. krusei ABC1* and *ABC11* (Figure 7B). The two most significant changes, each of which would abolish splicing, are the GT to AC changes of the 5’ ss, and the T to A change of the BS. Deintronization of the inferred 92 bp intron may have caused a frameshift and a temporary pseudogene, which was possibly recovered by the inferred 10 bp insertion just 21 nt downstream (Figure 7B). Because all four Abc1p/Abc11p paralogs share this unusual insertion (Figure 7A), it would seem that, after the proposed initial deintronization event (possibly not long before speciation; ~134 MYA), *P. membranifaciens ABC1* and *ABC11* responded quite differently to the functional disturbance caused by the ~30-aa insertion; *PmfABC1* had a 45- and *PmfABC11* had an 18-nt deletion (Figure 7B). However, without a cluster A PDR transporter sequence that still contains this “ancient” intron, the proposed deintronization could have also been a random insertion event instead.

An alignment of *C. krusei* and *P. membranifaciens ABC1* and *ABC11* with *C. krusei ABC12*, inferred from protein alignments surrounding the confirmed 88 bp intron (Figure 7A), revealed conserved splice site signals and intron boundaries (Figure 7C), strongly supporting *ABC12* intron loss by retrotransposition. Interestingly, *PmfABC1* and *PmfABC11* introns had conserved splice site signals, but, unlike the 88 bp *CkABC1/111* intron, they were completely different from each other, both in sequence and length (78 bp vs. 109 bp; Figure 7C).

**Functional characterization of Abc1p, Abc11p, and Abc12p**

To investigate functional differences between Abc1p, Abc11p, and Abc12p, we overexpressed each in the host *S. cerevisiae* ADα, a strain that is deleted in seven ABC pumps, and thus exquisitely sensitive to many xenobiotics, and which achieves exceptionally high expression of ABC proteins (Lamping *et al.* 2007). Both gDNA and cDNA ORFs encoding Abc1p and Abc11p of B2399 were cloned separately to see if the presence of introns affected the yeast phenotype. The presence of the intron did not alter drug resistance conferred by Abc1p or Abc11p significantly [see Figure S10 and Lamping *et al.* (2009)]. Strains overexpressing Abc1p or Abc11p were found to be resistant to a number of xenobiotics: i.e., resistant to azole antifungals (FLC, CLT, ITc, MCZ, and VRZ), fluorescent substrates (R6G and R123), large ionophores (NIG), translation inhibitors (CHX and ANI), and anticancer drugs (DOX and DAU; see Figure 8A and Figure S10). The antifungal AUR and the antipsychotic drug TFP, however, were not substrates of either pump (see Figure S10). Abc12p had the same substrate specificities as Abc1p and Abc11p (the seven Abc1/11p substrates tested in Figure 8A were also Abc12p substrates). Although all three multidrug efflux-pumps transported the same range of substrates, their transport activities for individual compounds differed significantly (Figure 8A). Abc1p appeared to pump the selected test compounds best overall. However, Abc12p was the best transporter of the smaller test compounds [i.e., low molecular weight (MW); ANI, CHX, FLC, and CLT], and Abc11p was the best transporter of the larger (i.e., high MW) test compounds (R6G and NIG). Perhaps surprisingly, Abc11p of 89102 (pump strain A in Figure 8A), a chimera with two Abc1p-specific amino acids in TMS6 (patch III) and one in TMS11 (patch V), was an even better multidrug efflux pump (i.e., it was more resistant to most test compounds) than Abc1p. Additional variations in substrate specificities were observed for the remaining *ABC1-1* and *ABC1-11* chimeras (Figure 8A, pump strains B–E and F–G, respectively). Although all chimeras were clearly functional, some effluxed certain compounds even more efficiently than wild-type Abc1p or Abc1p (e.g., FLC efflux by Abc1-11p chimeras; Figure 8A, pump strains F and G) whereas some Abc11-1p chimeras showed severely reduced drug efflux, especially of ANI, KTC, or NIG (Figure 8A, pump strains B, C, and E).

Differences were also observed in relative sensitivities conferred by Abc1p and Abc11p to known efflux pump inhibitors. Milbemycins (see Figure S11) are acaricides, insecticides, and anthelmintics widely used in agriculture and veterinary medicine. FK506 is a commonly used immunosuppressant; ENI and beauvericin are ionophoric depsipeptides with antibiotic and insecticidal activities, and OLI is an ATPase inhibitor. Abc1p was more sensitive than Abc1p to all eight efflux pump inhibitors tested, especially milbemycin β11 (Figure 8B), and, as expected, Abc11p of 89102 conferred inhibitor sensitivities that were between those for Abc1p and Abc11p of B2399 (Figure 8B) because it contains elements of Abc1p (pump strain A in Figure 8A).

**SF demonstrated by Abc1p and Abc11p**

The 18 core aa differences between Abc1p and Abc11p in TMS3, -4, -6, -7 and -11 (Figure 6C) clearly provided *C. krusei* with two distinct biological functions. V582I (IL1) and T795A (TMS6; Figure 6C) were allelic variations. Because ABC12 shared a common ancestor with *ABC1* and *ABC11* (Figure 6B), we compared the 18 core aa differences between *C. krusei* Abc1p and Abc11p with Abc12p to infer which had likely changed from their common ancestor. In addition, to ascertain which of these changes were possibly fate determining, we assessed the severity of changes by comparing them with 53 fungal homologs [see Figure S8 and Lamping *et al.* (2010)]. It appeared that Abc1p experienced only 5 aa (green and red circles on Figure 6C) and Abc11p 16 aa (blue

(5’ and 3’ ss) or are underlined (BS). Concerted evolution kept identical 88 bp introns in CkABC1 and CkABC11, but the 77 bp PmfABC1 and the 109 bp PmfABC11 introns escaped such concerted evolution. Not only are these two introns completely different in size and sequence, but each has also gained an additional 11 bp (PmfABC1) and 16 bp (PmfABC11) polypyrimidine tract (dark gray).
and red circles on Figure 6C) changes compared to their common ancestor, three of which were changes to the same position in both transporters (red circles; Figure 6C). The most significant changes, not found in any of the other 53 fungal homologs, have thick circles on Figure 6C. Abc11p experienced the majority of changes, which is also reflected in the maximum likelihood tree for the concatenated sequences of patches II–V (see Figure S9B). Compared to 53 fungal homologs, Abc11p had a unique V616 at the bottom of TMS3, and Abc1p had two unique amino acids, M628 and G1209, near the top of TMS3 and the center of TMS7 (thick circles; Figure 6C). Most changes were in, or near, the predicted lipid-facing surfaces (LIPS) of individual TMSs (see Figure S12). Four of the five changes predicted to face the substrate-binding pocket in Abc11p were either too high (TMS3, -4 and -6) or too low (TMS7) to contribute directly to the substrate-binding pocket, leaving S1355, at the center of TMS11, the only candidate likely to directly interact with substrates.

Sub-NF of the regulatory control of ABC1, ABC11, and ABC12

A genome-wide study of the evolution of cis-regulatory elements in duplicated genes in yeast found that the number of shared motifs rapidly decreased with time, whereas the total number of cis-regulatory motifs remained constant (Papp et al. 2003b). This suggested that NF played an important role in the evolution of duplicated genes in yeast (Papp et al. 2003b). Similar studies in yeast and humans later showed that most duplicated genes evolved by rapid SF followed by prolonged NF, a model termed sub-neo-functionalization (SNF) (He and Zhang 2005).

To test whether the regulatory control of ABC1 and ABC11 also evolved by SNF, and to confirm possible NF of ABC12, we searched their promoters with the YeTFaSCo database (de
of transcription factor binding sites (TFBSs; Figure 9). As expected, we found similar numbers of regulatory motifs in each promoter (13 in ABC1, 14 in ABC11, 16 in ABC12; Table 3 and Figure 9). The ABC1 promoter shared only four TFBSs (HCM1, GTS1, and two GCR1) with the ABC11 Table 3 List of S. cerevisiae TFBSs identified in the promoters of C. krusei ABC1, ABC11, and ABC12

| Groups of TFBSs | PDR Network | General Stress Response | Weak Acid or Alkaline Stress, Arsenic Resistance, and Quiescence | Diauxic Shift and Hypoxia | Glycolysis and Nitrogen Metabolism | Mitosis, Meiosis, Mating, and Sporulation |
|----------------|-------------|-------------------------|---------------------------------------------------------------|--------------------------|----------------------------------|----------------------------------------|
| TFBSs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| ABC1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ABC11 | 2 | 4 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ABC12 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

a 1=PDR1, 2=PDR3, 3=PDR8, 4=YRR1, 5=STB5, 6=MSN2/4, 7=CRZ1, 8=YAP3, 9=HAH1, 10=RM101, 11=ARR1/YAP8, 12=AFT2, 13=STB3, 14=DAT1, 15=AFT2, 16=HCM1, 17=GCR1, 18=DAT1, 19=ARO80, 20=UME6, 21=SUC1, 22=STE12, 23=GTS1.

b 662 bp ABC1 and 1000 bp ABC11 and ABC12 upstream sequences were searched for potentially meaningful TFBSs (numbers indicate how many times that site was detected; also see Figure 9).
promoter, and three TFBSs (PDR3 and two STB5) with the ABC12 promoters, and the ABC11 promoter shared five TFBSs (YAP3, AFT2, GAT1, and two MSN2/4) with the ABC12 promoter, but no TFBS were shared between all three promoters. The ABC1 promoter had five (PDR1, AZF1, UME6, and two SUM1), the ABC11 promoter two (HAA1 and AR080), and the ABC12 promoter eight (PDR8, YRR1, CRZ1, RIM101, ARR1/YAP8, STB3, DAT1, and STE12) unique TFBSs (Table 3). Regulation of expression of PDR transporters involves a complex network of transcription factors (TFs) belonging to the PDR network (i.e., PDR1, PDR3, PDR8, YRR1, and STB5), and often involves TFs of the general stress response (i.e., MSN2, MSN4, YAP3, and CRZ1; (Sipos and Kuchler 2006)). Interestingly, ABC12 had evolved five TFBSs of the PDR network and four TFBSs of the general stress response network, whereas regulation by these two networks was clearly divided between ABC1 (five for PDR) and ABC11 (six for general stress response; Table 3 and Figure 9).

Discussion

The C. krusei population is surprisingly homogenous

This study provides important insights into the population structure of C. krusei. There was surprisingly little sequence difference between isolates from vastly different locations, and, importantly, our data suggest that clinical (i.e., 89021, 89221, 89102, 90147, and V2b) C. krusei isolates may be of environmental origin. Although generally considered diploid, one C. krusei strain (89102) was aneuploid and B2399 was triploid, which is not uncommon in yeast and possibly a response to environmental stress. B2399 may have gained an additional set of chromosomes in the same way as some triploid D. bruxellensis hybrid strains (Borneman et al. 2014). D. bruxellensis is a close relative of C. krusei, and a major contaminant of industrial fermentations in biofuel production and the Australian wine industry (Borneman et al. 2014). It is also possible that B2399 is triploid as a result of a failure in cell division, but that would have had to happen in the distant past given its three distinct sets of chromosomes with typical SNP frequencies.

We have previously shown that the constitutive expression of Abc1p may tip the balance in favor of innate azole resistance for C. krusei (Lamping et al. 2009). In this report, we show that there may yet be other factors contributing to the innate azole resistance of C. krusei, namely the expression of additional multidrug efflux transporters Abc11p and Abc12p.

First direct estimates for the EGC rate and the MEPS and evidence for GC-bias of EGC in C. krusei

ABC1 and ABC11 are an exemplary pair of tandem-duplicated genes that enabled a detailed analysis of the major forces shaping their evolution. The striking pattern of identical sequences interspersed by four short (158, 19, 73, and 1 nt), highly divergent (22.2% dissimilarity), and strongly protected, sequences is a predicted outcome for the concerted evolution of tandem-duplicated paralogs of distinct function (Innan and Kondrashov 2010). Interestingly, ABC1 and ABC11 initially appeared to have experienced concerted evolution for >70 MY, which is significantly longer than the previous estimate of 25 MY for the average period of concerted evolution in yeast (Gao and Innan 2004). However, the discovery of the “same” pair of tandem-duplicated paralogs in P. membranifaciens extended that estimate to >134 MY.

Because the five DNA tracts that were homogenized by frequent EGCs varied from 0.3 to 1.7 kb in size (Table 2), and they were separated by an equal distance (5.3 kb; Figure 5A), and because we obtained polymorphism data for a set of C. krusei alleles representative of the entire C. krusei population (e.g., no additional ABC1- or ABC11-specific SNPs were detected in any of the three C. krusei whole genome shotgun (WGS) sequenced strains) it was possible to obtain, to our knowledge for the first time, a direct empirical estimate for the C. krusei gene conversion rate and the MEPS. The gene conversion rate increased proportionately with the tract length, and was ~38–480× higher than the estimated silent-site nucleotide substitution rate for yeast. However, depending on the true nucleotide substitution rate of C. krusei, these estimates could be up to ~8× lower (i.e., ~5–60) or ~2× higher (i.e., ~76–960). The calculated MEPS for C. krusei was 106 bp. These results were in reasonable agreement with previous “indirect” estimates of 10–100× higher gene conversion than nucleotide substitution rates, and a MEPS of ~200 bp for yeast (reviewed in Mansai et al. 2011), although our data suggest that previous estimates may have been slightly conservative. The GC-bias of gene conversion was also confirmed.

Gene deletions and CNVs combined with frequent EGCs between tandem-duplicated genes provide additional means for the evolution of novel gene function

Three out of seven strains possessed gene deletions and CNVs that, together with frequent EGCs between individual gene copies, created additional ABC11-1 and ABC1-11 chimeras with significantly altered functions. We previously characterized 28 different TMD chimeras between C. albicans Cdr1p and Cdr2p, two closely related paralogs (Figure 6B) (Holmes et al. 2006), and, in that study as well, many chimeras were severely altered in their substrate specificity and inhibitor sensitivities (Tanabe et al. 2011). It would appear that creating chimeras, and/or exchanging regulatory elements between duplicated paralogs, by intramolecular or intermolecular (non)homologous recombination and/or EGC are important evolutionary forces creating novel functionality.

Model for the evolution of C. krusei ABC11, ABC1, and ABC12

The discovery of ABC12 and P. membranifaciens ABC11-ABC1 was critical for unraveling the evolutionary history of C. krusei ABC1 and ABC11 (Figure 10). ABC12 appeared to
be the product of gene-duplication by retrotransposition (step 1) of their common ancestor ~149 MYA (Table 1), causing the loss of at least two inferred introns. The ancestor of \( ABC1 \) or \( ABC11 \) deintronized (step 4) one of these introns possibly soon after duplication (step 3), and before speciation of the \( C. krusei \) and \( P. membranifaciens \) lineages ~134 MYA, which may have caused a temporary pseudogene. A functional ORF was recovered (step 5) not long after by a 10 bp insertion, and the deintronized version was then copied by EGC (step 6) onto its neighboring duplicate. The \( ABC1 \) and \( ABC11 \) ancestors optimized their efflux pump functions by SF of their common ancestor \( ABC1/11’ \). While the majority of their ORFs have evolved by concerted evolution for ~134 MY, four small stretches of \( C. krusei \) \( ABC1 \) and \( ABC11 \) have been under very strong purifying selection (step 7), which led to an extremely clear pattern of sequence identity interspersed by very short peaks of sequence divergence. \( ABC12 \), on the other hand, is likely to have evolved by NF of both its regulatory control and pump function (step 2) since its retrotransposition into a completely different part of the genome. The experimental evidence indicates that, although all three genes are under very strong purifying selection, \( ABC1 \), \( ABC11 \), and \( ABC12 \) are not dosage sensitive genes, and one gene copy appears to be enough for the survival of \( C. krusei \) (triploid B2399 and diploid IFO0011 each had only one intact \( ABC1 \) and \( ABC11 \) gene copy, and diploid strains 89021, 89221, and 90147 each had only one intact \( ABC12 \) gene copy).

\( P. membranifaciens \) \( ABC1 \) and \( ABC11 \) followed a different evolutionary trajectory after speciation from \( C. krusei \), maybe to compensate for the loss of the third multidrug efflux pump \( ABC12 \). Like \( C. krusei \) \( ABC1 \) and \( ABC11 \), however, they too experienced concerted evolution for ~134 MY, and exhibit a unique pattern of sequence identities interspersed with stretches of sequence divergence. Interestingly, the nitrate assimilating Saccharomyces yeast \( C. utilis \) also has similar PDR multidrug efflux pump tandem-duplicates that were difficult to assemble because of the complex DNA patterns caused by concerted evolution (Watanasrisin et al. 2016). It would seem that prolonged concerted evolution of tandem-duplicates is perhaps more common than generally assumed, at least in the case of PDR transporters.

**Different types of intron loss in \( C. krusei \) PDR transporters**

Because all Saccharomycotina cluster A PDR transporters share a common ancestor with \( ABC1 \), \( ABC11 \), and \( ABC12 \) (Figure 6B), it is likely that these too contained at least two introns, which they have lost by a number of independent retrotransposition events. We also identified one N-terminal intron each in both \( C. krusei \) homologs of \( S. cerevisiae \) \( PDR12 \) encoding a weak organic acid transporter (Piper et al. 1998).

Both predicted proteins were highly homologous to \( S. cerevisiae \) Pdr12p (75 and 77% identity, respectively), and 86% homologous to each other. We therefore named them \( PDR12 \) (SD108 scaffold 20: 24,998–29,583) and \( PDR121 \) (i.e., \( PDR12 \) one; SD108 scaffold 7: 101,770–106,331). Although the ORF sequences were almost identical, their introns varied in size from 103–104 (\( PDR12 \)) to 66–80 (\( PDR121 \)) nucleotides between the three WGS sequences. This was in contrast to the \( C. krusei \) \( ABC1 \) and \( ABC11 \) intron, which was identical for all A and B alleles, demonstrating the overpowering force of EGC between closely linked and highly similar stretches of DNA that are under no or little positive selection. Thus, similar to the \( ABC1/11/12’ \) ancestor, the common ancestor of the \( PDR12 \) homologs had at least two introns, because the \( PDR12 \) and \( PDR121 \) introns clearly had distinct origins (i.e., they were inserted 49 aa apart, just after K209 and K258, respectively—a highly conserved region of NBD1 between the Q-loop and the ABC signature motif; see Figure S8). Because \( PDR121 \) lacked the \( PDR12 \) intron just after K209, \( PDR121 \) likely lost this ancestral intron by homologous recombination between a cDNA copy of its mRNA and the region surrounding that intron. However, unlike \( PDR12 \) and \( PDR121 \), that each had two polypyrimidine tracts, one before and one after the BS signal, and \( P. membranifaciens \) \( ABC11 \) and \( ABC1 \), that each had one polypyrimidine tract after the BS signal (Figure 7C), the conserved \( C. krusei \) \( ABC1-ABC11 \) intron had
no polypyrimidine tract. However, the 31 nt between its BS and 3’ ss signal was well within the maximum 45 nt required for efficient mRNA splicing in yeast (Meyer et al. 2011). Although yeast genes contain few introns [S. cerevisiae contains, on average, 0.05 per gene (Hirschman et al. 2006) compared to 8.1 introns per gene for H. sapiens (International Human Genome Sequencing Consortium 2004)], it is now commonly accepted that the fungus-animal ancestor had an intron density comparable to the highest known modern intron densities, underscoring the importance of intron loss during eukaryotic evolution (Stajich et al. 2007). It seems that ABC1 and ABC11, as well as PDR12 and PDR121, are “ancient” transporters that have not yet lost all their introns.

**Gene duplication, a possibly fate determining proline change, and (SN)F of the regulatory control enabled the evolution of distinct biological functions**

Since 2006, a number of structures for bacterial (Dawson and Locher 2006), lower eukaryotic (Jin et al. 2012; Kodan et al. 2014), and mouse (Aller et al. 2009) ABCB-type multidrug efflux transporters, functional homologs of PDR transporters, have been solved. The structures revealed a centrally located substrate-binding cavity (Dawson and Locher 2006; Aller et al. 2009; Jin et al. 2012) into which substrates enter from the inner leaflet of the bilayer via a gate formed by a “flexible” TMS4 caused by strategically positioned G and P residues (Kodan et al. 2014). A high frequency of helix kinks, often caused by P residues, is one of the hallmarks of membrane protein structure (Yohannan et al. 2004). Unique G and P residues were also involved in the evolution of Abc1p and Abc11p. It is possible that the unique P change at the bottom of TMS3 of Abc11p (P616V) may have been fate-determining, and induced the SF of Abc1p and Abc11p. Changing the conserved proline-kink at the bottom of TMS3 of Abc11p by an angle of ~100° possibly caused significant TMS packing disorder, which was compensated for by changes to TMS3, TMS4, IL2, TMS6, TMS7, and TMS11 (Figure 6C), the majority of which were along one face of TMS3 (four changes) and TMS7 (four changes; see Figure S12). The involvement of the TMS7 surface, which includes G1209 as well as V1201, V1205, V1209, and I1213 of Abc11p, in critical helical rearrangements (i.e., helix–helix contacts) during the transport cycle of PDR transporters was previously noted by Kolaczkowski et al. (2013), who found that an A1208V mutation of C. albicans Cdr1p (equivalent to S1212 of Abc11p, which is also part of that surface; see Figure S12) dramatically altered its substrate transport activities (Kolaczkowski et al. 2013).

As a consequence of the changes to Abc11p, Abc1p function may have been optimized by modulation of amino acids in TMS3, TMS6, and TMS7, with M628 and G1209 possibly critically important for that process (Figure 6C). There is also strong evidence that S1355N, the patch V Abc11p/Abc1p amino acid at the center of TMS11 and possibly facing the substrate-binding pocket (see Figure S12), is important for substrate transport and inhibitor sensitivity. Mutation of the equivalent amino acid in S. cerevisiae Pdr5p (S1360A/F) and C. albicans Cdr1p (T1351A/F) severely affected substrate transport, and also eliminated the pumps’ sensitivities to FK506 (Egner et al. 2000; Saini et al. 2005). In the absence of better alternatives, the YeTFaSCo database provides a valuable tool for the search of TFs that possibly regulate gene expression in members of the Saccharomyces family. Our discovery of functionally related TFBSs was consistent with the SNF of the regulatory controls of ABC1 and ABC11, and the NF of the regulatory control of ABC12.

**Summary**

C. krusei is an innately azole resistant diploid fungal pathogen with a tendency to create aneuploid strains or triploid hybrids when exposed to environmental stress. This study confirms that it belongs to a third major Saccharomyces lineage that has yet to be classified. A detailed study of the tandem-duplicated multidrug efflux pump genes ABC11 and ABC1 (a) provided the first empirical estimates for the gene conversion rate and the size of the MEPS for any eukaryotic organism; (b) confirmed GC-bias for EGC in C. krusei; (c) revealed concerted evolution for the majority of the two ORFs lasting for >134 MY; and (d) highlighted the importance of gene deletions, nonhomologous mitotic recombination, and CNVs combined with frequent EGCs between tandem-duplicated genes, for the evolution of novel functionality. We also provided first examples of intron loss by retrotransposition, deintronization, and, possibly, also by homologous recombination for any fungal PDR transporter. Discovery of a third C. krusei multidrug efflux transporter, ABC12, and the discovery of the tandem-duplicated orthologs P. membranifaciens ABC1 and ABC11, helped unravel the evolutionary history of an entire ABC transporter multi-gene family. It revealed a unique proline-kink change at the bottom of TMS3 of C. krusei ABC11 that possibly sealed the fate of Abc11p and induced the SF of ABC1 and ABC11. The regulatory control of ABC1 and ABC11, however, “escaped” concerted evolution and appeared to have evolved by SNF. We propose that, given the experimental difficulties in recognizing and correctly assembling highly similar, tandem-duplicated, gene sequences, their abundance, and, in turn, the impact of EGC on gene duplication and evolution, may be significantly underestimated by the scientific community.

**Acknowledgments**

The authors thank Judith Berman, Pete and Bebe Magee, and Mark McClelland, Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, USA, for their hospitality and expert supervision with fluorescent cell sorting experiments, as well as Sankyo Pharmaceuticals Ltd., Tokyo, Japan, for providing the milbemycins, and Astellas Pharma Inc., Tokyo, Japan, for providing FK506. This work was supported by the Japan Health Sciences Foundation, the National Institutes of Health, USA (R21DE015075-RDC; R01DE016885-01-RDC).
the Foundation for Research Science and Technology of New Zealand (IIOF grant UOOX0607-RDC), and the Marsden Fund of the Royal Society of New Zealand (UOO1305-RDC-EL). The authors declare that they have no competing interests.

**Literature Cited**

Adamian, L., and J. Liang, 2006  Prediction of transmembrane helix orientation in polytopic membrane proteins. BMC Struct. Biol. 6: 13.

Aller, S. G., J. Yu, A. Ward, Y. Weng, S. Chittaboina et al., 2009 Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–1722.

Anisimova, M., and O. Gascuel, 2006 Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55: 539–552.

Arguello, J. R., and T. Connallon, 2011 Gene duplication and ectopic gene conversion in Drosophila. Genes (Basel) 2: 131–151.

Blow, N., 2015 Decoding the unsequenceable. Biotechniques 58: 52–58.

Borneman, A. R., R. Zeppel, P. J. Chambers, and C. D. Curtin, 2014 Insights into the Dekkera bruxellensis genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. PLoS Genet. 10: e1004161.

Brown, D. D., P. C. Wensink, and E. Jordan, 1972 A comparison of the ribosomal RNAs of Xenopus laevis and Xenopus mulleri: the evolution of tandem genes. J. Mol. Biol. 63: 57–73.

Butler, G., M. D. Rasmussen, M. F. Lin, M. A. Santos, S. Sakhthiham et al., 2009 Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459: 657–662.

Casola, C., C. L. Ganote, and M. W. Hahn, 2010 Nonallelic gene conversion in the genus Drosophila. Genetics 185: 95–103.

Chan, G. F., H. M. Gan, H. L. Ling, and N. A. Rashid, 2012 Genome sequence of Pichia kudriavzevii M12, a potential producer of bioethanol and phytase. Eukaryot. Cell 11: 1300–1301.

Chen, J. M., C. Ferec, and D. N. Cooper, 2010 Gene conversion in human genetic disease. Genes (Basel) 1: 550–563.

Curtin, C. D., A. R. Borneman, P. J. Chambers, and I. S. Pretorius, 2012 De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast Dekkera bruxellensis AWR11499. PLoS One 7: e33840.

Davis, J. C., and D. A. Petrov, 2004 Preferential duplication of conserved proteins in eukaryotic genomes. PLoS Biol. 2: E55.

Dawson, R. J., and K. P. Locker, 2006 Structure of a bacterial multidrug ABC transporter. Nature 443: 180–185.

de Boer, C. G., and T. R. Hughes, 2012 YeTFaScO: a database of evaluated yeast transcription factor sequence specificities. Nucleic Acids Res. 40: D169–D179.

De Schutter, K., Y. C. Lin, P. Tiels, A. Van Hecke, S. Glinka et al., 2009 Genome sequence of the recombinant protein production host Pichia pastoris. Nat. Biotechnol. 27: 561–566.

Dujon, B., 2010 Yeast evolutionaries genomics. Nat. Rev. Genet. 11: 512–524.

Egner, R., B. E. Bauer, and K. Kuchler, 2000 The transmembrane domain 10 of the yeast Pdr5p ABC antifungal efflux pump determines both substrate specificity and inhibitor susceptibility. Mol. Microbiol. 35: 1255–1263.

Fawcett, J. A., and H. Innan, 2011 Neutral and non-neutral evolution of duplicated genes with gene conversion. Genes (Basel) 2: 191–209.

Finnigan, G. C., V. Hanson-Smith, T. H. Stevens, and J. W. Thornton, 2012 Evolution of increased complexity in a molecular machine. Nature 481: 360–364.

Gallicer, N., 2003 Gene conversion drives GC content evolution in mammalian histones. Trends Genet. 19: 65–68.

Gao, L. Z., and H. Innan, 2004 Very low gene duplication rate in the yeast genome. Science 306: 1367–1370.

Guindon, S., and O. Gascuel, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52: 696–704.

He, X., and J. Zhang, 2005 Rapid subfunctionalization accompanied by prolonged and substantial nonfunctionalization in duplicate gene evolution. Genetics 169: 1157–1164.

Hirschman, J. E., R. Balakrishnan, K. R. Christie, M. C. Costanzo, S. S. Dwight et al., 2006 Genome snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. Nucleic Acids Res. 34: D442–D445.

Hittinger, C. T., A. Rokas, F. Y. Bai, T. Boekhout, P. Goncalves et al., 2015 Genomics and the making of yeast biodiversity. Curr. Opin. Genet. Dev. 35: 100–109.

Holmes, A. R., S. Tsao, S. W. Ong, E. Lamping, K. Niimi et al., 2006 Heterozygosity and functional allelic variation in the Candida albicans efflux pump genes CDR1 and CDR2. Mol. Microbiol. 62: 170–186.

Hughes, T. R., and C. G. de Boer, 2013 Mapping yeast transcriptional networks. Genetics 195: 9–36.

Innan, H., 2003 A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. Proc. Natl. Acad. Sci. USA 100: 8793–8798.

Innan, H., and F. Kondrashov, 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11: 97–108.

International Human Genome Sequencing Consortium, 2004 Finishing the euchromatic sequence of the human genome. Nature 431: 931–945.

Jacobsen, M. D., N. A. Gow, M. C. Maiden, D. J. Shaw, and F. C. Odds, 2007 Strain typing and determination of population structure of Candida krusei by multilocus sequence typing. J. Clin. Microbiol. 45: 317–323.

Jin, M. S., M. L. Oldham, Q. Zhang, and J. Chen, 2012 Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. Nature 490: 566–569.

Kellis, M., B. W. Birren, and E. S. Lander, 2004 Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 428: 617–624.

Kodan, A., T. Yamaguchi, T. Nakatsu, K. Sakiyama, C. J. Hipolito et al., 2014 Structural basis for gating mechanisms of a eukaryotic P-glycoprotein homolog. Proc. Natl. Acad. Sci. USA 111: 4049–4054.

Kolaczkowski, M., K. Sroda-Pomianek, A. Kolaczkowska, and K. Michalak, 2013 A conserved interdomain communication pathway of pseudosymmetrically distributed residues affects substrate specificity of the fungal multidrug transporter Cdr1p. Biochim. Biophys. Acta 1828: 479–490.

Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu et al., 2004 Introns and splicing elements of five diverse fungi. Eukaryot. Cell 3: 1088–1100.

Kurtzman, C. P., 2011 Phylogeny of the ascomycetous yeasts and the renaming of Pichia anomala to Wickerhamomyces anomalous. Antonie van Leeuwenhoek 99: 13–23.

Lafontaine, I., and B. Dujon, 2010 Origin and fate of pseudogenes in Hemiascomycetes: a comparative analysis. BMC Genomics 11: 260.

Lamping, E., B. C. Monk, K. Niimi, A. R. Holmes, S. Tsao et al., 2007 Characterization of three classes of membrane proteins involved in fungal azole resistance by functional hypereexpression in Saccharomyces cerevisiae. Eukaryot. Cell 6: 1150–1165.

Michalak, 2013 A conserved interdomain communication pathway of pseudosymmetrically distributed residues affects substrate specificity of the fungal multidrug transporter Cdr1p. Biochim. Biophys. Acta 1828: 479–490.

Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu et al., 2004 Introns and splicing elements of five diverse fungi. Eukaryot. Cell 3: 1088–1100.

Kurtzman, C. P., 2011 Phylogeny of the ascomycetous yeasts and the renaming of Pichia anomala to Wickerhamomyces anomalous. Antonie van Leeuwenhoek 99: 13–23.

Lafontaine, I., and B. Dujon, 2010 Origin and fate of pseudogenes in Hemiascomycetes: a comparative analysis. BMC Genomics 11: 260.
the balance in favor of innate azole resistance in *Candida krusei*. Antimicrob. Agents Chemother. 53: 354–369.

Lamping, E., P. V. Baret, A. R. Holmes, B. C. Monk, A. Goffeau et al., 2010 Fungal PDR transporters: phylogeny, topology, motifs and function. Fungal Genet. Biol. 47: 127–142.

Li, W. H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Li, W. H., Z. Gu, H. Wang, and A. Nekrutenko, 2001 Evolutionary analyses of the human genome. Nature 409: 847–849.

Lynch, M., 2007 The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA 104(Suppl. 1): 8597–8604.

Lynch, M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.

Lynch, M., and J. S. Conery, 2003 The evolutionary demography of duplicate genes. J. Struct. Funct. Genomics 3: 35–44.

Mansai, S. P., T. Kado, and H. Innan, 2011 The rate and tract length of gene conversion between duplicated genes. Genes (Basel) 2: 313–331.

Marais, G., 2003 Biased gene conversion: implications for genome and sex evolution. Trends Genet. 19: 330–338.

Marcet-Houben, M., and T. Gabaldon, 2015 Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker’s yeast lineage. PLoS Biol. 13: e1002220.

Meyer, M., M. Plass, J. Perez-Valle, E. Eyras, and J. Vilardebell, 2011 Deciphering 3’s selection in the yeast genome reveals an RNA thernosensor that mediates alternative splicing. Mol. Cell 43: 1033–1039.

Morales, L., B. Noel, B. Porcel, M. Marcet-Houben, M. F. Hullo et al., 2013 Complete DNA sequence of *Kuraishiia capsulata* illustrates novel genomic features among budding yeasts (*Saccharomyces*). Genome Biol. Evol. 5: 2524–2539.

Nathans, J., D. Thomas, and D. S. Hogness, 1986 Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science 232: 193–202.

Nei, M., and A. P. Rooney, 2005 Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39: 121–152.

Neitz, M., J. Neitz, and G. H. Jacobs, 1991 Spectral tuning of pigments underlying red-green color vision. Science 252: 971–974.

Niimi, K., D. R. Harding, R. Parshot, A. King, D. J. Lun et al., 2004 Chemosensitization of *the* [*Saccharomyces*] genome. PLoS Biol. 2: e1000305.

Papp, B., C. Pal, and L. D. Hurst, 2003a Dosage sensitivity and the evolution of gene families in yeast. Nature 424: 194–197.

Papp, B., C. Pal, and L. D. Hurst, 2003b Evolution of cis-regulatory elements in duplicated genes of yeast. Trends Genet. 19: 417–422.

Piper, P., Y. Mahe, S. Thompson, R. Pandjaitan, C. Holyoak et al., 1998 The pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. EMBO J. 17: 4257–4265.

Ravin, N. V., M. A. Eldarov, V. V. Kadinikov, A. V. Beletsky, J. Schneider et al., 2013 Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DLI. BMC Genomics 14: 837.

Riley, R., S. Haridas, K. H. Wolfe, M. R. Lopes, C. T. Hittinger et al., 2016 Comparative genomics of biotechnologically important yeasts. Proc. Natl. Acad. Sci. USA 113: 9882–9887.

Rubin, G. M., M. D. Yandell, J. R. Wortman, G. L. Gabor Miklos, C. R. Nelson et al., 2000 Comparative genomics of the eukaryotes. Science 287: 2204–2215.

Saini, P., T. Prasad, N. A. Gaur, S. Shukla, S. Jha et al., 2005 Alanine scanning of transmembrane helix 11 of Cdr1p ABC antifungal efflux pump of *Candida albicans*: identification of amino acid residues critical for drug efflux. J. Antimicrob. Chemotherapy. 56: 77–86.

Sanchez, R., F. Serra, J. Tarraga, I. Medina, J. Carbonell et al., 2011 *Phylemon* 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res. 39: W470–W474.

Shyue, S. K., L. Li, B. H. Chang, and W. H. Li, 1994 Intrinsic gene conversion in the evolution of human X-linked color vision genes. Mol. Biol. Evol. 11: 548–551.

Sipos, G., and K. Kuchler, 2006 Funnel ABC-binding cassette (ABC) transporters in drug resistance & detoxification. Curr. Drug Targets 7: 471–481.

Stajich, J. E., F. S. Dietrich, and S. W. Roy, 2007 Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. Genome Biol. 8: R223.

Tanabe, K., E. Lamping, M. Nagi, A. Okawada, A. R. Holmes et al., 2011 Chimeras of *Candida albicans* Cdr1p and Cdr2p reveal features of pleiotropic drug resistance transporter structure and function. Mol. Microbiol. 82: 416–433.

Thompson, J. D., D. G. Higgins, and T. J. Gibson, 1994 CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.

Walsh, J. B., 1995 How often do duplicated genes evolve new functions? Genetics 139: 421–428.

Wapinski, I., A. Pfeffer, N. Friedman, and A. Regev, 2007 Natural history and evolutionary principles of gene duplication in fungi. Nature 449: 54–61.

Watanasrisin, W., S. Iwatanli, T. Oura, Y. Tomita, S. Ikushima et al., 2016 Identification and characterization of *Candida utilis* multidrug efflux transporter CuCdr1p. FEBS J. 283: 2215–2221.

Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.

Yohannan, S., S. Faham, D. Yang, J. P. Whitelegge, and J. U. Bowie, 2004 The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc. Natl. Acad. Sci. USA 101: 959–963.

Zheng, D., A. Frankish, R. Baertsch, P. Karpanov, A. Reymond et al., 2007 Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res. 17: 839–851.
Role of Ectopic Gene Conversion in the Evolution of a Candida krusei Pleiotropic Drug Resistance Transporter Family

Erwin Lamping, Jing-yi Zhu, Masakazu Niimi and Richard David Cannon