

Evolutionary Capture of Viral and Plasmid DNA by Yeast Nuclear Chromosomes^{▽†}

A. Carolin Frank^{1,2*} and Kenneth H. Wolfe¹

Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland,¹ and School of Natural Sciences, University of California, Merced, California²

Received 14 April 2009/Accepted 28 July 2009

A 10-kb region of the nuclear genome of the yeast *Vanderwaltozyma polyspora* contains an unusual cluster of five pseudogenes homologous to five different genes from yeast killer viruses, killer plasmids, the 2 μ m plasmid, and a *Penicillium* virus. By further database searches, we show that this phenomenon is not unique to *V. polyspora* but that about 40% of the sequenced genomes of Saccharomycotina species contain integrated copies of genes from DNA plasmids or RNA viruses. We propose the name NUPAVs (nuclear sequences of plasmid and viral origin) for these objects, by analogy to NUMTs (nuclear copies of mitochondrial DNA) and NUPTs (nuclear copies of plastid DNA, in plants) of organellar origin. Although most of the NUPAVs are pseudogenes, one intact and active gene that was formed in this way is the *KHS1* chromosomal killer locus of *Saccharomyces cerevisiae*. We show that *KHS1* is a NUPAV related to M2 killer virus double-stranded RNA. Many NUPAVs are located beside tRNA genes, and some contain sequences from a mixture of different extrachromosomal sources. We propose that NUPAVs are sequences that were captured by the nuclear genome during the repair of double-strand breaks that occurred during evolution and that some of their properties may be explained by repeated breakage at fragile chromosomal sites.

It is well known that the nuclear genomes of most eukaryotes contain integrated fragments of organellar DNA called NUMTs (nuclear copies of mitochondrial DNA) and NUPTs (nuclear copies of plastid DNA, in plants) (26, 29, 44, 45, 57). These fragments are usually pseudogenes, although some NUMTs and NUPTs have become incorporated into functional nuclear genes (38). The NUMTs present in the nuclear genomes of Saccharomycotina yeast species were recently analyzed by Sacerdot et al. (48).

In addition to their mitochondrial genomes, yeast species contain a variety of other extranuclear DNA and RNA elements, including viruses and plasmids. These extrachromosomal elements are usually considered to be autonomous entities that do not interact with nuclear DNA. When our laboratory sequenced the genome of the yeast *Vanderwaltozyma polyspora* (synonym: *Kluyveromyces polysporus*) (49), we were therefore surprised to find the genomic region we describe here, which contains integrated fragments of several plasmid- and virus-like sequences. We propose that this region was formed by the capture of plasmid and viral sequences by the same mechanism that captures mitochondrial DNA to form NUMTs (43, 65). In a literature search, we could find only one previous report of a similar finding: Utatsu et al. (59) reported the sequences of two regions of nuclear DNA from *Zygosaccharomyces rouxii* that were highly similar to parts of the 2 μ m-like plasmid pSR1 from that species, but rearranged.

Before describing the *V. polyspora* region, and similar re-

gions found in other species, we will first briefly introduce the extrachromosomal RNA and DNA entities that are known to exist in yeasts. Extrachromosomal nucleic acids are relatively uncommon in yeasts: a broad survey of 1,800 strains from 600 species by Fukuhara (14) found that 196 strains (11%) contained some sort of extrachromosomal entity. Among these, 105 strains had a double-stranded RNA (dsRNA), 28 had a linear dsDNA plasmid, and 53 had a circular DNA plasmid of the 2 μ m family. These elements typically also have a patchy distribution within a species, being found in some individuals or strains but not in others. For instance, Nakayashiki et al. (37) surveyed 70 “wild” strains of *Saccharomyces* (mostly *S. cerevisiae*) for the presence of five extrachromosomal elements (2 μ m DNA plasmid, L-A and L-BC helper RNA viruses, and W and T RNA entities) and found each element to be present in between 1 and 38 of the strains, with 1 strain even containing all five elements simultaneously.

Yeast killer systems. Killer systems are genetic elements present in some cells that cause the death of other cells that do not contain the element. These systems vary hugely in their genetic organizations and mechanisms of action, and the killer phenotype appears to have multiple evolutionary origins (31). Most killer systems involve either cytoplasmic dsRNA (“killer viruses,” such as the M1 virus of *S. cerevisiae*) or cytoplasmic dsDNA (“killer plasmids,” such as the pGKL1 plasmid of *Kluyveromyces lactis*). More rarely, polymorphic nuclear genes have been implicated in killer phenotypes (“chromosomal killers”); the best known of these are the *KHS1* and *KHR1* genes of *S. cerevisiae* (17, 18) and the *SMK1* gene of *Pichia farinosa* (56). Killer yeasts produce and secrete extracellular toxin proteins to which they are immune or resistant but which can kill sensitive strains of the same or different species.

Different toxin molecules attack different molecular targets

* Corresponding author. Mailing address: 4225 Hospital Rd., Atwater, CA 95301-5142. Phone: (209) 201-4530. Fax: (209) 228-4675. E-mail: cfrank3@ucmerced.edu.

† Supplemental material for this article may be found at <http://ec.asm.org/>.

▽ Published ahead of print on 7 August 2009.

in sensitive cells; known mechanisms include inhibition of DNA synthesis (*S. cerevisiae* K28 toxin), disruption of ionic potential across membranes (*S. cerevisiae* K1 toxin), and destruction of tRNA molecules by an endonuclease (*K. lactis* zymocin) (33, 51). The killer phenotype (that is, the production of a toxin and immunity to the effect of one's own toxin) is widespread in yeasts living in natural habitats (6), particularly in the diverse insect-colonized yeast communities living in fruit, where the killer phenotype provides a competitive advantage (55). The fact that not all yeast species are killers and that polymorphism for the phenotype exists in some species suggests that there is a cost associated with toxin production (42, 54).

dsRNA killer viruses. The dsRNA virus-based killer system of *S. cerevisiae* has been studied extensively (reviewed in references 51 and 52). In this system, killer cells contain two types of cytoplasmic dsRNA molecules: a satellite virus that encodes the toxin-coding gene and a helper virus without which the satellite virus cannot survive. Three distinct toxin proteins (K1, K2, and K28) and their corresponding satellite viruses (M1, M2, and M28) have been identified. Even though these three satellites do not have any apparent sequence similarity to one another, they all depend on the same helper virus, L-A, for their replication and encapsidation, and any killer yeast cell will carry only one of these satellite viruses because they exclude one another (52). The M and L-A dsRNAs are separately encapsidated into virus-like particles in the cytoplasm of the infected cell. The L-A helper virus genome is 5 kb in size and codes for two proteins: a coat protein and an RNA polymerase that is expressed as a fusion with the coat protein by ribosomal frameshifting (10, 13, 21). The M1, M2, and M28 satellites are only 1 to 2 kb in size, and each contains only one gene, coding for the corresponding toxin (5, 9, 34, 53). There is no separate gene coding for an immunity function in these dsRNA viruses; the unprocessed precursor of the mature secreted killer toxin confers immunity. A second helper virus, L-BC, with a genomic organization similar to that of L-A, has also been identified in some *S. cerevisiae* strains (40), but no killer virus satellites of L-BC are known. Many strains of *S. cerevisiae* contain L-A and/or L-BC without a killer satellite virus being present (1, 37).

In addition to the L-A/L-BC viruses and the M satellites, two other small (2- to 3-kb) cytoplasmic linear dsRNA molecules have been identified in *S. cerevisiae* and have been named T and W (reviewed in references 12 and 46). These molecules are not well understood. T and W each contain just one gene, coding for an RNA-dependent RNA polymerase (RDRP), and the T and W RDRPs are highly divergent from each other. T and W do not become encapsidated into virus-like particles, but their RDRPs have low but significant similarity to viral RDRPs, so they are apparently of viral origin. Single-stranded RNA molecules corresponding to one strand of T and W, called the 23S and 20S RNAs, respectively, accumulate in large amounts during sporulation and may be replication intermediates (12).

Linear dsDNA killer plasmids. *K. lactis* provides the best-characterized example of a DNA plasmid-based killer system (19, 20, 50). The toxin, called zymocin, consists of three subunits (α , β , and γ) that are encoded on the 9-kb cyto-

plasmic linear dsDNA plasmid pGKL1. The toxicity of the zymocin complex resides solely within the smallest subunit (γ), encoded by open reading frame 4 (ORF4). The α and β subunits are both encoded by ORF2. The α subunit is a chitinase that, together with the hydrophobic β subunit, is putatively involved in the uptake of the γ subunit into the target cell. As well as coding for the zymocin precursor subunits, pGKL1 contains a gene (ORF3) that confers immunity on the host cell. pGKL1 is a satellite of a helper plasmid, pGKL2, on which it depends for essential functions, such as replication; pGKL2 is 13 kb in size and contains 10 genes (50, 58). A similar system of two linear dsDNA plasmids exists in *Pichia acaciae*, with a toxin/immunity satellite plasmid (pPac1-2) depending on a larger helper plasmid (pPac1-1) for its maintenance (39). When characterized, all of the linear dsDNA plasmids from other yeast species that have been studied in detail have turned out to be similar to these killer plasmids, although not all of the strains that contain them show a killer phenotype (14).

Circular 2 μ m dsDNA plasmids. The *S. cerevisiae* 2 μ m plasmid and its relatives are the only circular plasmids that have been found to occur naturally in yeasts (14). In a survey of 2,500 strains of 500 species, the 2 μ m family was found only in the genera *Saccharomyces*, *Zygosaccharomyces*, *Kluyveromyces*, and *Torulaspora* (4). This is a much narrower host range than is seen in the linear dsDNA plasmids. The *S. cerevisiae* 2 μ m plasmid is a 6-kb circular dsDNA plasmid present in the nuclei of many natural isolates of *S. cerevisiae* at a copy number of about 60 (15, 37). All plasmids in the 2 μ m family contain two identical copies of an ~600-bp repeat sequence in opposite orientations that divide the rest of the molecule into two single-copy regions of approximately equal sizes (60). The plasmids undergo flip-flop recombination across this repeat so that two isomeric forms of the plasmid exist in equimolar quantities. This recombination is carried out by the FLP recombinase, a member of the DNA-breaking-rejoining family that also includes bacterial tyrosine recombinases. *FLP* is one of the four genes on the 2 μ m plasmid, all of which have functions related to replication or maintenance (60).

MATERIALS AND METHODS

The *V. polyspora* VPEI (for *V. polyspora* extrachromosomal island) genomic region discussed here is located in contig 2001 (GenBank accession number DS480388) of the genome sequence. Although most of the genome of this species was sequenced by a whole-genome shotgun (WGS) approach, the sequence of the VPEI region was finalized by completely sequencing a fosmid clone (fos53_d05) of 31,743 bp that spans VPEI and that closed a gap between two contigs in the initial assembly of the genome (49). Agreement between the fosmid sequence and the data from plasmids sequenced in the shotgun phase shows that the observed frameshifts and internal stop codons are not artifacts of cloning or sequence assembly. The five gene remnants in the region were given the names Kpol_2001.23a ψ through Kpol_2001.26a ψ (Fig. 1). These remnants were detected using BLASTX searches against the NCBI nonredundant (nr) database. A summary of the hits is shown in Fig. 2. Two of the hits were very weak and were annotated using an iterative approach (41), where the first hit was used as a query for a new search of the nr database.

To search for similar regions in other fungal genomes, we first constructed a data set of fungal plasmid- and virus-encoded proteins (hereafter called the FPV data set) (see Table S1 in the supplemental material) from NCBI. We began by including all genes from the 17 plasmid or viral genomes that have at least one homolog among the gene relics in the VPEI. For example, the *V. polyspora* pseudogene Kpol_2001.24 ψ is homologous to immunity genes on

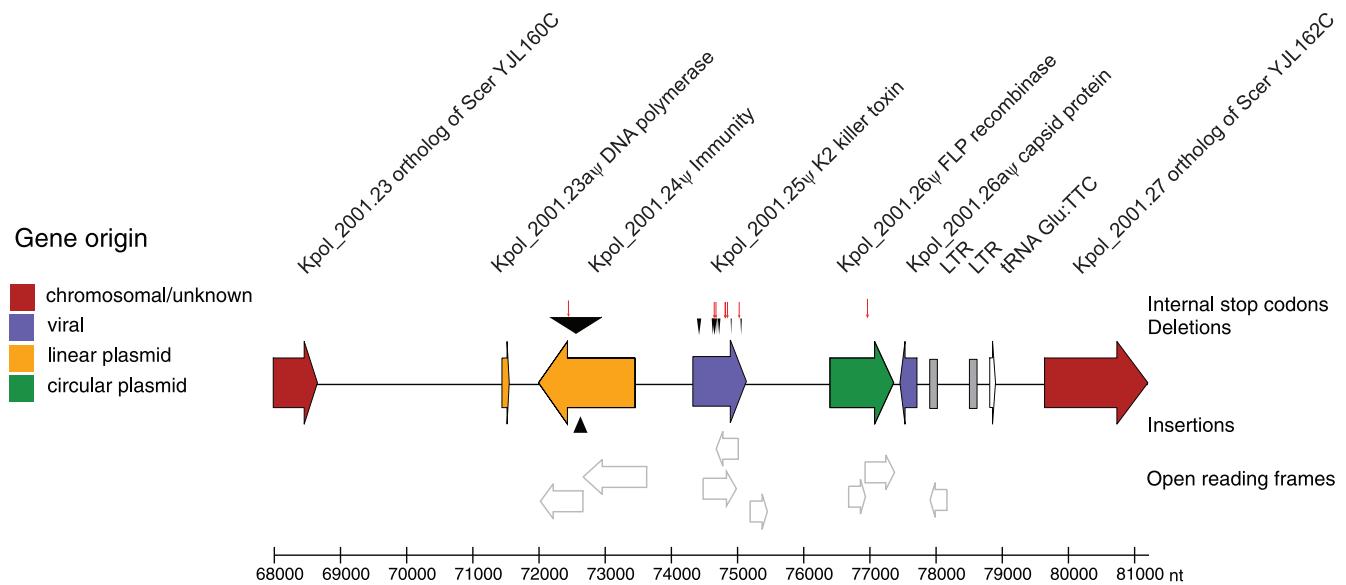


FIG. 1. Structure of the VPEI region, an example of a NUPAV. The positions of internal stop codons and deletions and insertions in the pseudogenes relative to intact homologs are indicated, as are the locations of ORFs in the VPEI region. The scale shows positions in the chromosomal contig (GenBank accession number DS480388; NCBI gi 156117034).

three killer plasmids, so all genes from these three killer plasmids were included in the FPV data set. It turned out that the resulting data set included representatives of all the yeast DNA plasmid families sequenced to date. However, the data set included only two viruses, so to increase the representation of other viral genomes beyond those with homologs in *V. polyspora*, we included the K1 and K28 killer toxin proteins of *S. cerevisiae* and the KP1, KP4, and KP6 killer toxins of *Ustilago maydis*, the only other killer virus toxins for which there are protein sequences. We also included the two proteins of the *S. cerevisiae* L-A and L-BC helper viruses. Thus, the FPV data set includes all sequenced yeast plasmids and viruses, though plasmids from other fungi (such as the *Neurospora* mitochondrial plasmids) were excluded (see Table S1 in the supplemental material). We subsequently filtered the FPV data set to remove RNA polymerase proteins because they are parts of large families and cannot readily be recognized as having a plasmid or viral origin through a BLAST search alone. The filtered FPV data set contained 71 proteins.

The FPV data set was annotated to simplify the classification of their homologs in fungal genomes. First, we clustered the 71 proteins into families using an all-against-all BLASTP search (E value $\leq 1e-5$). Then, each family was annotated for gene function (e.g., DNA polymerase or immunity for clusters containing at least one gene of known function or unknown1, etc., for clusters with unknown functions) and for genome type origin (linear dsDNA, circular dsDNA, or dsRNA). By our definition of gene function, some genes not previously annotated as immunity genes were annotated as such. *K. lactis* pGKL1 ORF1 and *Lachancea kluyveri* pSKL ORF1 shared some sequence similarity with the known immunity gene *K. lactis* pGKL1 ORF3 and were therefore considered immunity homologs.

We then used the FPV data set to search for homologs, or remnants thereof, in complete fungal genomes and in fungal WGS data. We downloaded 16 complete fungal genomes and 59 fungal WGS genomes from NCBI on 12 September 2007 (the species are listed in Table S2 in the supplemental material). The following BLAST searches were conducted: BLASTP of all proteins from complete fungal genomes against the FPV data set (BLASTP_complete), BLASTX of complete genome sequences divided into 1-kb pieces against the FPV data set (BLASTX_complete), and BLASTX of WGS sequences divided into 1-kb pieces against the FPV data set (BLASTX_WGS). Only the top hit in the FPV data set for each query in the fungal genomes was saved, and only hits with E values of $\leq 1e-10$ were considered. Four hits to plasmid sequences in the WGS sequences (from *Kluyveromyces waltii* and *Saccharomyces kudriavzevii*) were excluded from the results. Query regions with BLASTX hits that overlapped with BLASTP protein hits were ignored. The region beside each genomic FPV homolog was scanned for adjacent tRNA genes (annotated using tRNAScan-SE [30]) and the presence of other FPV homologs.

Genomic regions with BLAST hits to the FPV data set were classified as either

a “putatively intact gene” (if the BLAST hit overlapped with an intact ORF covering at least 70% of the query protein) or a “putative pseudogene” (otherwise). BLAST hits in which a query protein hit a genomic DNA region in more than one reading frame were classified as “putative pseudogenes.” Where a genomic sequence contains multiple frameshifts or stop codons and is clearly too damaged to be functional, we use the word “pseudogene” without qualification.

Because the plasmid killer toxin α subunit (PKT α) is related to a large family of chitinases, the BLAST searches with PKT α queries resulted in hits to a multitude of genomic chitinases. To look for genuine genomic homologs of PKT α proteins, we saved a list of all genomic regions that had PKT α hits and then searched those genomic sequences against each other and against the PKT α sequences. Only those genomic regions that were reciprocal best hits with a PKT α and not with a genomic chitinase were considered PKT α homologs and added to the set of genomic FPV homologs.

The resulting 88 BLAST high-scoring pairs from the BLASTP_complete, BLASTX_complete, and BLASTX_WGS searches were compiled (see Table S3 in the supplemental material) and manually inspected for hits to the same FPV homologs and for adjacent FPV homologs (clusters). For each genomic region in which these searches detected one or more homologs of the FPV data set, we also manually inspected the results of a BLASTX search against the FPV database to locate additional FPV homologs whose BLAST E values were weaker than the cutoff used in our initial search.

RESULTS

The VPEI, an extrachromosomal island in the *V. polyspora* genome. During the annotation of the *V. polyspora* genome (49), we identified a 10-kb region with unusual properties (Fig. 1). Because of its similarities to bacterial genomic islands and its apparent extrachromosomal origin, we refer to it as the VPEI. It is located between the *V. polyspora* genes Kpol_2001.23 and Kpol_2001.27, which are orthologs of *S. cerevisiae* YJL160C (a member of the PIR cell wall protein family) and YJL162C (JJ2), respectively. VPEI is located beside a tRNA^{Glu} gene that has an ortholog at the corresponding location in *S. cerevisiae* and two long terminal repeats from Tkp4, a common retrotransposon in the *V. polyspora* genome.

VPEI contains five pseudogenes homologous to virus and plasmid genes from fungi in various stages of degradation. In

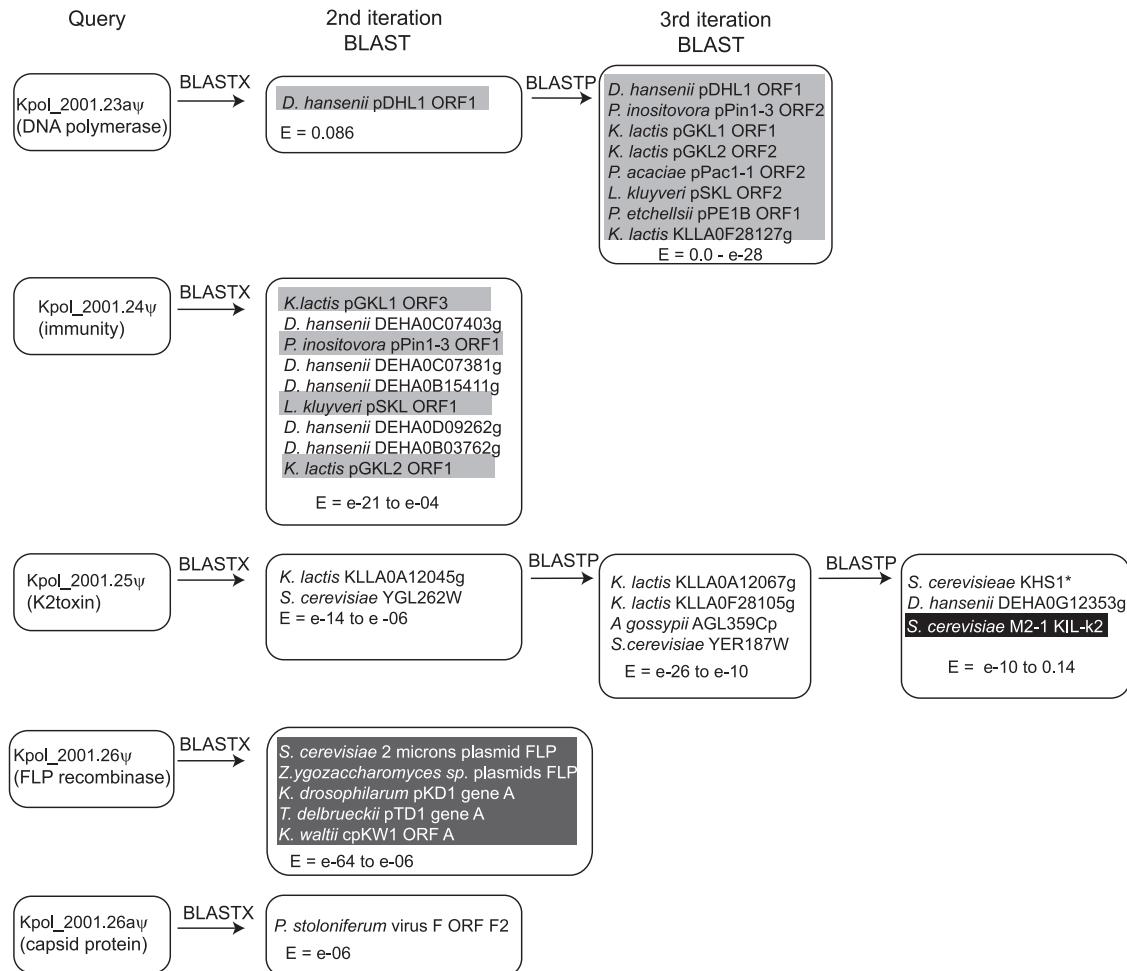


FIG. 2. Summary of BLAST search results using pseudogenes in the VPEI region as queries. The left column names the VPEI pseudogenes that were used as queries in BLASTX searches. The second column summarizes the hits seen in these searches and the range of BLAST E values. The third and fourth columns indicate the results of subsequent searches using as queries proteins that were hit in the previous searches, as shown by the arrows (41). Black background shading indicates one protein encoded by a dsRNA virus, dark-gray backgrounds show proteins from circular plasmids, light-gray backgrounds show proteins from linear plasmids, and the absence of background shading indicates chromosomal genes.

order (from left to right in Fig. 1), they are Kpol_2001.23aψ, a short, heavily degraded homolog of a killer plasmid DNA polymerase; Kpol_2001.24ψ, a nearly intact homolog of a killer plasmid immunity gene; Kpol_2001.25ψ, a degraded homolog of the viral K2 killer toxin; Kpol_2001.26ψ, a nearly intact homolog of the FLP recombinase from a 2μm plasmid; and Kpol_2001.26aψ, a degraded remnant homologous to a mycovirus capsid protein.

For convenience, we omit the prefix Kpol_2001 from the locus names in the following discussion. The 23aψ and 26aψ loci are so truncated and damaged that they are more accurately described as gene relics (27) than as pseudogenes. Figure 2 summarizes the results of BLASTX searches of the five *V. polypora* loci against the NCBI nr database, and sequence alignments are included in Fig. S1 in the supplemental material. Some of the BLASTX hits were to genes on plasmids and viruses, whereas others were to chromosomal genes in other yeast species.

The 23aψ (DNA polymerase) pseudogene and 24ψ (im-

munity) putative pseudogene have similarity to genes on linear dsDNA plasmids, including the pGKL1 killer plasmid. The similarity of 23aψ to plasmid DNA polymerases extends over only a short segment of 111 bp and is of borderline significance when considered alone (E value, 0.086 for the top BLASTX hit, to ORF1 of the *Debaryomyces hansenii* plasmid pDHL1). Further BLASTP searches with pDHL1 ORF1 showed that it is related to the DNA polymerases of linear DNA plasmids in several other yeast species (Fig. 2, second-iteration BLAST). However, the adjacent 24ψ putative pseudogene had a very significant hit to immunity genes from the same family of linear plasmids (for example, pGKL1, Pin1-3, and pSKL) and is only one stop codon away from being an intact ORF. The 24ψ sequence is most similar to ORF3 of *K. lactis* pGKL1, but interestingly, it also has homologs in the *D. hansenii* nuclear genome (Fig. 2). The presence of the putative immunity pseudogene 24ψ greatly increases the likelihood that the match between the adjacent 23aψ region and the DNA polymerases from the

same family of plasmids is also genuine; otherwise, it would be an extraordinary coincidence.

The 25ψ locus is a relatively long (974 bp) but degraded pseudogene with similarity to a family of chromosomal genes that includes the *S. cerevisiae* chromosomal killer gene *KHS1* (17). The KHS (killer, heat-sensitive) phenotype involves a chromosomally encoded thermolabile toxin that kills sensitive strains. *KHS1* is located near the telomere of *S. cerevisiae* chromosome V and is not syntenic with the VPEI locus in *V. polyspora*. We analyzed the sequences of *KHS1* available in GenBank from various *S. cerevisiae* strains and found that the species is polymorphic for active and null alleles. *KHS1* is intact and codes for a 350-amino-acid protein in *S. cerevisiae* strains YJM789 (protein name, SCY_1690), M22, and most of the strains sequenced by the *Saccharomyces* Genome Resequencing Project, but it contains an internal stop codon in the reference strain S288c and in strains RM11-1a and YPS163 (data from references 8, 11, 28, 47, and 62). The ORF YER187W in S288c corresponds to a 3' part of SCY_1690 from YJM789. We believe that there are errors in the original report of the sequence of *KHS1*, from *S. cerevisiae* strain no. 115 (17), because it differs from all other sequences by many frameshifts and by a 1.4-kb inversion that ends at a restriction site used in cloning. More interestingly, the coding region of *KHS1* from *S. cerevisiae* strains YJM789 and M22 is much more similar to its ortholog in *Saccharomyces paradoxus* (99% DNA sequence identity) than to the null alleles in other strains of *S. cerevisiae* (89% identity). This result suggests horizontal exchange of *KHS1* between *S. paradoxus* and some strains of *S. cerevisiae*, but we cannot infer the direction of transfer. *KHS1* is intact (an ORF encoding 407 amino acids) in *S. paradoxus* but contains disabling mutations in *Saccharomyces mikatae* and *Saccharomyces bayanus* (data from references 7 and 22).

KHS1 also has chromosomal homologs in *K. lactis* (KLLA0A12045g, which has the highest similarity to the *V. polyspora* 25ψ pseudogene, and KLLA0C19327g), *Eremothecium gossypii* (AGL359C), and *D. hansenii* (DEHA2C 11286g), as well as being similar to *S. cerevisiae* YGL262W. Starting with KLLA0C012045g, two further iterations of BLASTP against the NCBI nr database, using the first hit as a query for a new search, identified the K2 toxin of the *S. cerevisiae* M2 satellite killer virus as a homolog of this family (Fig. 2). Similarly, a significant relationship between the *KHS1* locus and the K2 toxin was recovered from two iterations of PSI-BLAST, using SCY_1690 from *S. cerevisiae* strain YJM789 as the initial query. These results suggest that a cDNA copy of an M2 killer virus became integrated into yeast nuclear chromosomes on two occasions: once to form the *KHS1* locus, which is an intact and active gene in some *S. cerevisiae* strains, and once to form the 25ψ pseudogene in the single strain of *V. polyspora* that has been examined.

The 26ψ locus of the VPEI is a putative pseudogene homologous to the FLP recombinase from the 2μm circular DNA plasmid family. It is nearly intact, with only one frameshift and a missing start codon. Its hypothetical translation product has higher sequence identity (38%) to the *S. cerevisiae* 2μm FLP

recombinase than to any of the other sequenced recombinases from the 2μm plasmid family.

Finally, the 26aψ pseudogene has weak but significant similarity (BLASTX E value, 2e-6) to a capsid protein (P46 protein) from segment 2 of *Penicillium stoloniferum* virus F. This segment is one of three small dsRNA molecules that comprise the viral genome, and the capsid gene is the only gene present on segment 2 (23). The host species, *P. stoloniferum*, is a filamentous ascomycete (Pezizomycotina), in contrast to the hosts of the extrachromosomal homologs of the other four pseudogenes in the VPEI, which are all Saccharomycotina. The P46 protein has no homologs in the NCBI databases, except for high similarity to the putative translation of some expressed sequence tags that were isolated from rice plants infected with another filamentous ascomycete, *Magnaporthe grisea*, and which probably derive from an unidentified *M. grisea* virus.

Homologs of plasmid and viral genes in other genomes. We used BLAST searches to investigate whether the integration of plasmid and viral genes is a general phenomenon in fungi. We made a database (FPV) containing all the proteins encoded by plasmid or viral genomes that have homologs in the VPEI, plus some additional proteins of potential interest (see Materials and Methods). We then searched for homologous genomic sequences in all 75 available fungal genome sequences (64 ascomycetes, 8 basidiomycetes, 1 microsporidian, 1 chytrid, and 1 from a basal fungal lineage; the species are listed in Table S2 in the supplemental material). Because the VPEI is located beside a tRNA gene, we also checked whether any genomic regions identified in these searches also contained tRNA genes.

These searches revealed 46 homologs to fungal plasmid and viral proteins (FPV homologs). The results are given in detail in Table S3 in the supplemental material and summarized in Tables 1 and 2. During subsequent inspection of the genomic regions with FPV hits, we identified a further eight FPV homologs whose BLAST scores were weaker than the cutoff used in our initial search (see Table S4 in the supplemental material; these hits are included in Table 1 and Table 2). Some of the hits to FPV sequences are physically clustered in the nuclear genomes in a manner similar to that in the VPEI; the 54 hits are located in 33 separate genomic regions (Table 1; see Table S3 in the supplemental material). There are 13 clusters with 2 to 5 FPV homologs (see Table S5 in the supplemental material) and 20 single FPV homologs. For most of the clusters, the FPV homologs they contain all come from the same type of extrachromosomal element, except for the VPEI (discussed above) and cluster 12 from *D. hansenii*, which contains a linear plasmid immunity gene homolog and a dsRNA virus toxin homolog (see Table S5 in the supplemental material).

Examples of some of the genomic regions detected are shown in Fig. 3 (for *D. hansenii*) and Fig. 4 (for other species). In six of the 33 regions, the FPV homologs are located directly adjacent to tRNA genes. In *S. bayanus*, integrated copies of two linear plasmid genes are found at one end of the tandem array of ribosomal DNA (rDNA) genes, between the 5S rDNA and the neighboring gene, *RNH203* (Fig. 4C).

All 11 species with FPV homologs in their nuclear genomes are in the Saccharomycotina clade (Table 1). Thus, the pro-

TABLE 1. Distribution of NUPAVs among Saccharomycotina genomes

Species	No. of hits ^a	No. of regions	No. of hits to:			No. of regions adjacent to tRNA	Described extrachromosomal element in species
			Linear plasmid	Circular plasmid	Virus		
<i>C. albicans</i>	2	2	2			1	
<i>C. tropicalis</i>	1	1			1	1	
<i>Clavispora lusitaniae</i>	1	1	1				
<i>D. hansenii</i>	18	12	12		6	2	Linear plasmids pDHL1/2/3
<i>K. lactis</i>	7	4	3	1	3		Linear plasmids pGKL1/2
<i>L. waltii</i>	2	2	1	1			Circular plasmid pKW1
<i>P. guilliermondii</i>	4	3	4				
<i>P. stipitis</i>	5	1			5		
<i>S. bayanus</i>	3	2	2	1		1	
<i>S. mikatae</i>	2	1		2			
<i>V. polyspora</i>	9	4	6	1	2	1	
Total	54	33	31	6	17	6	

^a A hit is defined as the presence of a gene, pseudogene, or gene relic with similarity to a protein in the FPV database.

portion of species in which we detected integrated extrachromosomal elements is 42% among the Saccharomycotina (11 of 26 species) but only 15% among all fungi (11 of 75 species). The apparent phylogenetic bias toward Saccharomycotina probably reflects a bias of ascertainment, because the majority of fungal plasmids and viruses that have been studied (that is, sequenced) also come from the Saccharomycotina. The distri-

bution of integrated elements is also quite uneven within the Saccharomycotina, with *D. hansenii* having by far the most (18 elements in 12 regions; some are shown in Fig. 3), followed by *K. lactis* and *V. polyspora* with six FPV homologs each. The plasmid and viral integrations reported here are likely to reflect only a subset of all plasmid and viral integrations in fungal genomes because our FPV data set is limited by sequence

TABLE 2. Functional categories of origin of the 54 NUPAV genes identified in Saccharomycotina genomes

Protein functional category ^a	Host element	No. of NUPAV genes in:										Total
		<i>C. albicans</i>	<i>C. tropicalis</i>	<i>Clavispora lusitaniae</i>	<i>D. hansenii</i>	<i>K. lactis</i>	<i>L. waltii</i>	<i>P. guilliermondii</i>	<i>P. stipitis</i>	<i>S. bayanus</i>	<i>S. mikatae</i>	
FLP recombinase	Circular plasmid									1	1	2
REP2 (<i>Kluyveromyces drosophilicola</i> pKD1)	Circular plasmid									1		1
2μm REP1	Circular plasmid									1		1
2μm REP2	Circular plasmid									1		1
REP (ORF D of <i>L. waltii</i> pKW1)	Circular plasmid								1			1
Immunity, <i>K. lactis</i> -like	Linear plasmid					3	1	1		1	3	9
Immunity, <i>P. acaciae</i> -like	Linear plasmid					1						1
Unknown1 (ORF1 of <i>K. lactis</i> pGKL2)	Linear plasmid					3					1	4
Unknown5 (ORF8 of <i>K. lactis</i> pGKL2)	Linear plasmid								1			1
Unknown6 (ORF9 of <i>K. lactis</i> pGKL2)	Linear plasmid								1			1
Putative terminal recognition factor	Linear plasmid	1										1
DNA polymerase	Linear plasmid					1	1		1	1	1	5
Unknown15 (ORF3 of <i>Pichia etchellsii</i> pPE1A)	Linear plasmid					1						1
Plasmid killer toxin, large (αβ) subunit	Linear plasmid	1		1	2	1		1			1	7
Viral K2 killer toxin	dsRNA virus		1		4	3					1	9
Pol (RNA polymerase)	L-A virus				1					1		2
Gag (capsid protein)	L-A virus				2				4			6
Capsid protein	<i>P. stoloniferum</i> virus F										1	1
Total		2	1	1	18	7	2	4	5	3	2	54

^a From our annotation of the FPV data set.

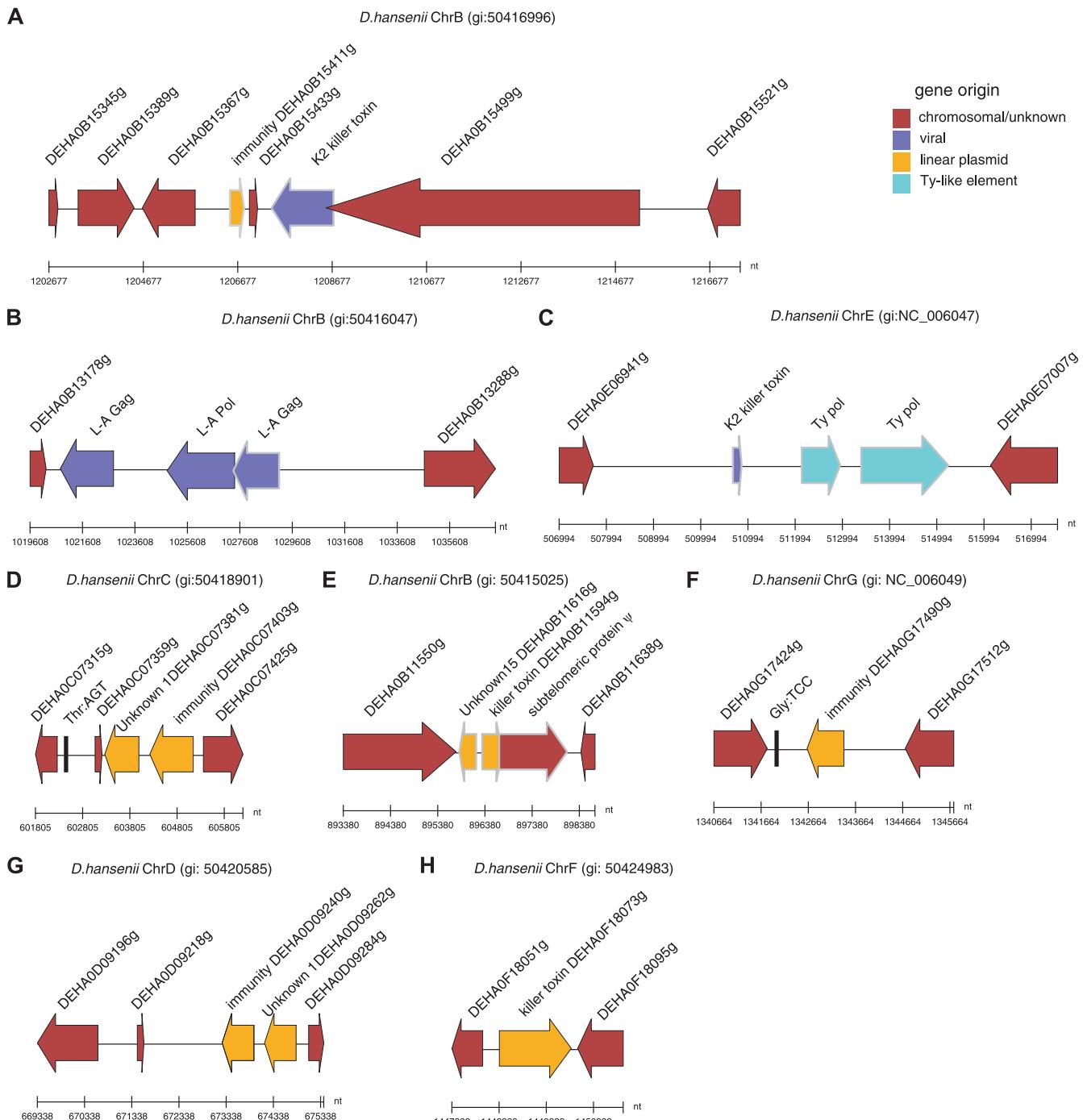


FIG. 3. Organizations of some NUPAVs in the *D. hansenii* genome. Gray outlines indicate pseudogenes, and black outlines indicate intact ORFs.

availability. We did not detect any examples of two different species sharing the same integration event, which suggests that we may be able to detect only relatively recent and hence species-specific events.

More than half of the integrated extrachromosomal sequences we detected are derived from linear DNA killer plasmids (Table 1). We found 10 nuclear fragments of immunity genes from these plasmids (Table 2). We also found

seven integrated nuclear fragments of genes coding for the large ($\alpha\beta$) subunit precursor of PKT (Table 2). In view of this, it is surprising that we did not find any copies of the toxin γ -subunit gene. However, toxin γ subunits vary extensively in sequences and modes of action, even between otherwise related killer plasmids (24, 25), so it is not unlikely that even if such genes existed in yeast nuclear genomes we could fail to detect them. One integrated copy of a toxin

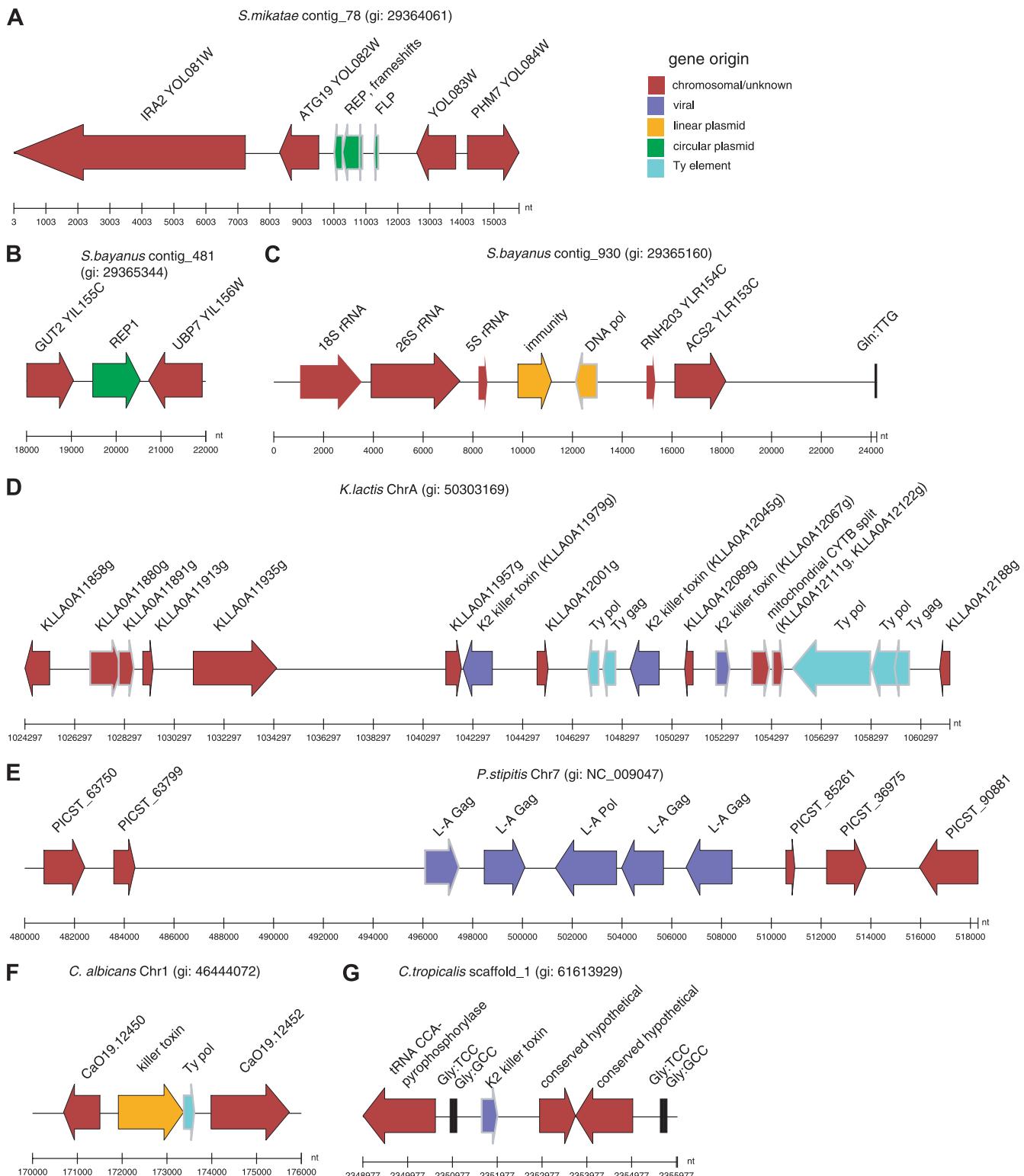


FIG. 4. Organizations of NUPAVs in other genomes in *Saccharomycotina* species. Gray outlines indicate pseudogenes, and black outlines indicate intact ORFs. *C. tropicalis*, *Candida tropicalis*.

$\alpha\beta$ -subunit gene in *D. hansenii* appears to be completely intact (gene DEHA0F18073g on chromosome F) (Fig. 3H). Full-length, potentially intact PKTs are also present in the genomes of *Candida albicans* (Fig. 4F) and *Pichia guillier-*

mondii (not shown), both of which do not carry any immunity gene detectable by our search criteria.

Integrations derived from dsRNA virus sequences are rarer, though we found nine nuclear fragments of K2 toxin genes

from M2-like dsRNA viruses (Table 2). The genomes of *D. hansenii* and *Pichia stipitis* both contain integrated sequences similar to the complete genome of the *S. cerevisiae* L-A helper virus, repeated in tandem arrays (Fig. 3B and 4E). Integrations derived from the 2 μ m family of circular DNA plasmids are the least common (Table 1).

DISCUSSION

We have detected sequences in fungal nuclear genomes with similarity to fungal plasmid and viral proteins, and we propose that they are the result of the evolutionary capture of sequences of viral or plasmid origin by the nuclear genome. In principle, the direction of evolution could be the reverse; genes located on plasmids and viruses could have a nuclear origin. However, because the phylogenetic distribution of FPV homologs in nuclear genomes is patchy (we did not detect any examples of two different species sharing the same FPV homologs at conserved genomic locations) and because most of the nuclear sequences are pseudogenes, we infer that the direction of transfer is more likely to have been from the extrachromosomal elements to the nuclear genome than vice versa.

How were the islands of integrated extrachromosomal DNA formed? They have several properties that are similar to structures seen in experiments on the repair of double-strand DNA breaks, as described below. We suggest that the islands were formed, over evolutionary time, by occasional events of aberrant repair of double-strand breaks in yeast nuclear DNA. This is the same mechanism that Ricchetti et al. and Sacerdot et al. proposed for the evolutionary origin of NUMTs in yeasts (43, 48). The lengths of integrated extrachromosomal sequences that we detected, their occurrence at clustered locations, and their association with tRNA genes are all very similar to the patterns seen in yeast NUMTs. We therefore propose the analogous name NUPAVs (nuclear sequences of plasmid and viral origin) for the loci we report here. It is also notable that among the studied species, *D. hansenii* has the highest incidence of both NUMTs (145 sites) (48) and NUPAVs (12 sites) (Table 1), although none of the NUPAV sites in the *D. hansenii* genome coincides with a NUMT site (data not shown).

Double-strand breaks are lethal to haploid cells if they are not repaired. Repair of a dsDNA break in haploid cells usually proceeds via the NHEJ pathway, but laboratory experiments with *S. cerevisiae* have shown that a small percentage of breaks are repaired aberrantly by a mechanism(s) that patches the chromosome back together by capturing some other fragment of DNA at the site of the joint. The sequences captured in these experiments included partial cDNAs of Ty elements and pieces of mitochondrial DNA (35, 43, 65). Similarly, partial cDNAs from several other genes have been found integrated into arrays of Ty elements at sites of chromosomal rearrangement in senescent yeast cells (32). It seems plausible that DNA from killer plasmids or 2 μ m-like plasmids could be used similarly to repair double-strand breaks, particularly since these extrachromosomal elements are present at high copy numbers. We suggest that the NUPAVs derived from dsRNAs may originate by a two-step process in which a single-stranded RNA (either a viral mRNA or a replication intermediate) is first reverse transcribed by Ty-encoded reverse transcriptase, and then the cDNA is used for dsDNA break repair. However, we

note that the use of virus or 2 μ m plasmid sequences to repair breaks was not detected in the *S. cerevisiae* laboratory experiments cited above, even though most laboratory strains do contain the 2 μ m plasmid (28) and the L-A, L-BC, and 20S RNAs (63, 64).

An aberrant DNA repair model can explain why fragments of integrated extrachromosomal DNA derived from different sources are sometimes clustered at adjacent sites in the nuclear genome, as occurs at two of the nine NUPAV clusters we identified (the VPEI and cluster 12) (see Table S5 in the supplemental material). There are two possible explanations for clustering. One is the simultaneous capture of more than one foreign molecule in a double-strand break, as seen in some *S. cerevisiae* experiments (43, 65). Alternatively, repeated cycles of breakage and repair, over millions of years, may have occurred at sites that are particularly fragile. The fact that some clusters, such as the VPEI, are located beside tRNA genes supports the latter idea. Ty elements tend to integrate beside tRNA genes, and their mechanism of integration involves the formation of a staggered dsDNA break in the chromosome (61). If each Ty integration event has a small probability of going wrong, then over evolutionary time the genomic regions close to tRNA genes will tend to have been broken and repaired repeatedly. We found previously that tRNA loci are hot spots for chromosomal rearrangements during evolution and that genes that have been added to the *S. cerevisiae* genome during recent evolution tend to be located beside tRNA genes, consistent with repeated cycles of breakage and repair (16). We hypothesize that a single mechanism—the repair of double-strand breaks by using whatever nucleic acid is available at the time—is responsible for the capture of extrachromosomal DNA to form NUMTs (as first proposed in references 43 and 65) and NUPAVs and for the evolutionary addition of some genes to genomes (16).

The presence of plasmid and viral genes in yeast genomes raises questions about their potential for functioning and their impact on genome evolution. Although most of the captured sequences are clearly incapable of being functional, the process of DNA capture during double-strand break repair has the potential to add functional new genes to the genome. One gene that was formed by this process in *S. cerevisiae* (*KHS1*) is known to be functional, and a family of genes homologous to it exists in many Saccharomycotina. The repertoire of genes that is carried by eukaryotic extrachromosomal elements is probably quite limited, so the main evolutionary consequence of the break-repair process is likely the formation of chromosomal killer and immunity loci from their plasmid or viral sources. However, it is also possible that the process could introduce more novel genes if viruses or plasmids have broad host ranges (as suggested by the presence of a *Penicillium* virus homolog in *V. polyspora*) and are capable of occasionally carrying sequences derived from their hosts.

We speculate that, like NUMTs and NUPTs, the occurrence of NUPAVs is a eukaryote-wide phenomenon. In yeasts, we observed capture of both plasmid and viral DNA. In animal and plant cells, the most common types of extrachromosomal nucleic acids are virus genomes, so we would expect these to occasionally be captured during the repair of dsDNA breaks, in addition to the previously reported capture of organelle DNA. Indeed, integration of hepatitis B virus DNA at dsDNA

breaks has been shown experimentally in chicken cell lines (3), and integration of geminiviral DNA during the recent evolution of the tobacco plant genus has been inferred bioinformatically (2, 36). Accidental double-strand breaks are likely to occur spontaneously in the germ line cells of all cellular organisms. DNAs from plasmids, viruses, and organelles occur in higher copy numbers than the nuclear DNA, so their frequent incorporation may simply reflect their ubiquity. Our results suggest that double-strand breaks are occasionally repaired by incorporation of any nucleic acid molecules available in the cell at the time.

ACKNOWLEDGMENTS

This work was supported by a Swedish Research Council (Vetenskapsrådet) postdoctoral fellowship, UC Merced startup funds, and Science Foundation Ireland.

We are grateful to two referees for helpful comments.

REFERENCES

1. Baeza, M. E., M. A. Sanhueza, and V. H. Cifuentes. 2008. Occurrence of killer yeast strains in industrial and clinical yeast isolates. *Biol. Res.* **41**:173–182.
2. Bejarano, E. R., A. Khashoggi, M. Witty, and C. Lichtenstein. 1996. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci. USA* **93**:759–764.
3. Bill, C. A., and J. Summers. 2004. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc. Natl. Acad. Sci. USA* **101**:11135–11140.
4. Blaisonneau, J., F. Sor, G. Cheret, D. Yarrow, and H. Fukuhara. 1997. A circular plasmid from the yeast *Torulaspora delbrueckii*. *Plasmid* **38**:202–209.
5. Bostian, K. A., Q. Elliott, H. Bussey, V. Burn, A. Smith, and D. J. Tipper. 1984. Sequence of the preprotoxin dsRNA gene of type I killer yeast: multiple processing events produce a two-component toxin. *Cell* **36**:741–751.
6. Buzzini, P., and A. Martini. 2000. Biodiversity of killer activity in yeasts isolated from the Brazilian rain forest. *Can. J. Microbiol.* **46**:607–611.
7. Cliften, P., P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B. A. Cohen, and M. Johnston. 2003. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**:71–76.
8. Dietrich, F. S., J. Mulligan, K. Hennessy, M. A. Yelton, E. Allen, R. Araujo, E. Aviles, A. Berno, T. Brennan, J. Carpenter, E. Chen, J. M. Cherry, E. Chung, M. Duncan, E. Guzman, G. Hartzell, S. Hunicke-Smith, R. W. Hyman, A. Kayser, C. Komp, D. Lashkari, H. Lew, D. Lin, D. Mosedale, K. Nakahara, A. Namath, R. Norgren, P. Oefner, C. Oh, F. X. Petel, D. Roberts, P. Sehl, S. Schramm, T. Shogren, V. Smith, P. Taylor, Y. Wei, D. Botstein, and R. W. Davis. 1997. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome V. *Nature* **387**(Suppl.):78–81.
9. Dignard, D., M. Whiteway, D. Germain, D. Tessier, and D. Y. Thomas. 1991. Expression in yeast of a cDNA copy of the K2 killer toxin gene. *Mol. Gen. Genet.* **227**:127–136.
10. Dinman, J. D., T. Icho, and R. B. Wickner. 1991. A –1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc. Natl. Acad. Sci. USA* **88**:174–178.
11. Doniger, S. W., H. S. Kim, D. Swain, D. Corcueria, M. Williams, S. P. Yang, and J. C. Fay. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* **4**:e1000183.
12. Esteban, R., N. Rodriguez-Cousino, and L. M. Esteban. 1993. Genomic organization of T and W, a new family of double-stranded RNAs from *Saccharomyces cerevisiae*. *Prog. Nucleic Acid Res. Mol. Biol.* **46**:155–182.
13. Fujimura, T., and R. B. Wickner. 1988. Gene overlap results in a viral protein having an RNA binding domain and a major coat protein domain. *Cell* **55**:663–671.
14. Fukuhara, H. 1995. Linear DNA plasmids of yeasts. *FEMS Microbiol. Lett.* **131**:1–9.
15. Futcher, A. B. 1988. The 2 micron circle plasmid of *Saccharomyces cerevisiae*. *Yeast* **4**:27–40.
16. Gordon, J. L., K. P. Byrne, and K. H. Wolfe. 2009. Additions, losses and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet.* **5**:e1000485.
17. Goto, K., H. Fukuda, K. Kichise, K. Kitano, and S. Hara. 1991. Cloning and nucleotide sequence of the KHS killer gene of *Saccharomyces cerevisiae*. *Agric. Biol. Chem.* **55**:1953–1958.
18. Goto, K., Y. Iwatsuki, K. Kitano, T. Obata, and S. Hara. 1990. Cloning and nucleotide sequence of the KHR killer gene of *Saccharomyces cerevisiae*. *Agric. Biol. Chem.* **54**:979–984.
19. Gunge, N. 1983. Yeast DNA plasmids. *Annu. Rev. Microbiol.* **37**:253–276.
20. Hishinuma, F., K. Nakamura, K. Hirai, R. Nishizawa, N. Gunge, and T. Maeda. 1984. Cloning and nucleotide sequences of the linear DNA killer plasmids from yeast. *Nucleic Acids Res.* **12**:7581–7597.
21. Icho, T., and R. B. Wickner. 1989. The double-stranded RNA genome of yeast virus L-A encodes its own putative RNA polymerase by fusing two open reading frames. *J. Biol. Chem.* **264**:6716–6723.
22. Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**:241–254.
23. Kim, J. W., E. Y. Choi, and J. I. Lee. 2005. Genome organization and expression of the *Penicillium stoloniferum* virus F. *Virus Genes* **31**:175–183.
24. Klassen, R., and F. Meinhardt. 2003. Structural and functional analysis of the killer element pPin1-3 from *Pichia inositovora*. *Mol. Genet. Genomics* **270**:190–199.
25. Klassen, R., S. Teichert, and F. Meinhardt. 2004. Novel yeast killer toxins provoke S-phase arrest and DNA damage checkpoint activation. *Mol. Microbiol.* **53**:263–273.
26. Kleine, T., U. G. Maier, and D. Leister. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu. Rev. Plant Biol.* **60**:115–138.
27. Lafontaine, I., G. Fischer, E. Talla, and B. Dujon. 2004. Gene relics in the genome of the yeast *Saccharomyces cerevisiae*. *Gene* **335**:1–17.
28. Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, I. N. Roberts, A. Burt, V. Konopanec, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. O'Kelly, A. van Oudenaarden, D. B. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. A. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E. J. Louis. 2009. Population genomics of domestic and wild yeasts. *Nature* **458**:337–341.
29. Lopez, J. V., N. Yuhki, R. Masuda, W. Modi, and S. J. O'Brien. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **39**:174–190.
30. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
31. Magliani, W., S. Conti, M. Gerloni, D. Bertolotti, and L. Polonelli. 1997. Yeast killer systems. *Clin. Microbiol. Rev.* **10**:369–400.
32. Maxwell, P. H., and M. J. Curcio. 2007. Retrosequence formation restructures the yeast genome. *Genes Dev.* **21**:3308–3318.
33. Mehlgarten, C., and R. Schaffrath. 2004. After chitin docking, toxicity of *Kluyveromyces lactis* zymocin requires *Saccharomyces cerevisiae* plasma membrane H⁺-ATPase. *Cell. Microbiol.* **6**:569–580.
34. Meskauskas, A. 1990. Nucleotide sequence of cDNA to yeast M2-1 dsRNA segment. *Nucleic Acids Res.* **18**:6720.
35. Moore, J. K., and J. E. Haber. 1996. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* **383**:644–646.
36. Murad, L., J. P. Bielawski, R. Matysek, A. Kovarik, R. A. Nichols, A. R. Leitch, and C. P. Lichtenstein. 2004. The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. *Heredity* **92**:352–358.
37. Nakayashiki, T., C. P. Kurtzman, H. K. Edskes, and R. B. Wickner. 2005. Yeast prions [URE3] and [PSI⁺] are diseases. *Proc. Natl. Acad. Sci. USA* **102**:10575–10580.
38. Noutsos, C., T. Kleine, U. Armbruster, G. Dalcorso, and D. Leister. 2007. Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. *Trends Genet.* **23**:597–601.
39. Paluszynski, J. P., R. Klassen, and F. Meinhardt. 2007. *Pichia acaciae* killer system: genetic analysis of toxin immunity. *Appl. Environ. Microbiol.* **73**:4373–4378.
40. Park, C. M., J. D. Lopinski, J. Masuda, T. H. Tzeng, and J. A. Bruenn. 1996. A second double-stranded RNA virus from yeast. *Virology* **216**:451–454.
41. Park, J., S. A. Teichmann, T. Hubbard, and C. Chothia. 1997. Intermediate sequences increase the detection of homology between sequences. *J. Mol. Biol.* **273**:349–354.
42. Pintar, J., and W. T. Starmer. 2003. The costs and benefits of killer toxin production by the yeast *Pichia kluveri*. *Antonie van Leeuwenhoek* **83**:89–97.
43. Ricchetti, M., C. Fairhead, and B. Dujon. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**:96–100.
44. Richly, E., and D. Leister. 2004. NUPTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* **21**:1081–1084.
45. Richly, E., and D. Leister. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol. Biol. Evol.* **21**:1972–1980.
46. Rodriguez-Cousino, N., A. Solorzano, T. Fujimura, and R. Esteban. 1998. Yeast positive-stranded virus-like RNA replicons. 20 S and 23 S RNA terminal nucleotide sequences and 3' end secondary structures resemble those of RNA coliphages. *J. Biol. Chem.* **273**:20363–20371.
47. Ruderfer, D. M., S. C. Pratt, H. S. Seidel, and L. Kruglyak. 2006. Population genomic analysis of outcrossing and recombination in yeast. *Nat. Genet.* **38**:1077–1081.
48. Sacerdot, C., S. Casaregola, I. Lafontaine, F. Tekaia, B. Dujon, and O. Ozier-Kalogeropoulos. 2008. Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. *FEMS Yeast Res.* **8**:846–857.
49. Scannell, D. R., A. C. Frank, G. C. Conant, K. P. Byrne, M. Woolfit, and K. H. Wolfe. 2007. Independent sorting-out of thousands of duplicated gene

pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl. Acad. Sci. USA* **104**:8397–8402.

50. Schaffrath, R., and K. D. Breunig. 2000. Genetics and molecular physiology of the yeast *Kluyveromyces lactis*. *Fungal Genet. Biol.* **30**:173–190.

51. Schmitt, M. J., and F. Breinig. 2002. The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol. Rev.* **26**:257–276.

52. Schmitt, M. J., and F. Breinig. 2006. Yeast viral killer toxins: lethality and self-protection. *Nat. Rev. Microbiol.* **4**:212–221.

53. Schmitt, M. J., and D. J. Tipper. 1995. Sequence of the M28 dsRNA: preprotoxin is processed to an alpha/beta heterodimeric protein toxin. *Virology* **213**:341–351.

54. Starmer, W. T., P. F. Ganter, and V. Aberdeen. 1992. Geographic distribution and genetics of killer phenotypes for the yeast *Pichia kluyveri* across the United States. *Appl. Environ. Microbiol.* **58**:990–997.

55. Starmer, W. T., P. F. Ganter, V. Aberdeen, M. A. Lachance, and H. J. Phaff. 1987. The ecological role of killer yeasts in natural communities of yeasts. *Can. J. Microbiol.* **33**:783–796.

56. Suzuki, C., and S. Nikkuni. 1994. The primary and subunit structure of a novel type killer toxin produced by a halotolerant yeast, *Pichia farinosa*. *J. Biol. Chem.* **269**:3041–3046.

57. Timmis, J. N., and N. S. Scott. 1983. Sequence homology between spinach nuclear and chloroplast genomes. *Nature* **305**:65–67.

58. Tommasino, M., S. Ricci, and C. L. Galeotti. 1988. Genome organization of the killer plasmid pGK12 from *Kluyveromyces lactis*. *Nucleic Acids Res.* **16**:5863–5878.

59. Utatsu, I., T. Imura, and A. Toh-e. 1988. Possible gene interchange between plasmid and chromosome in yeast. *Yeast* **4**:179–190.

60. Volkert, F. C., D. W. Wilson, and J. R. Broach. 1989. Deoxyribonucleic acid plasmids in yeasts. *Microbiol. Rev.* **53**:299–317.

61. Voytas, D. F., and J. D. Boeke. 2002. Ty1 and Ty5 of *Saccharomyces cerevisiae*, p. 631–662. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, DC.

62. Wei, W., J. H. McCusker, R. W. Hyman, T. Jones, Y. Ning, Z. Cao, Z. Gu, D. Bruno, M. Miranda, M. Nguyen, J. Wilhelmy, C. Komp, R. Tamse, X. Wang, P. Jia, P. Luedi, P. J. Oefner, L. David, F. S. Dietrich, Y. Li, R. W. Davis, and L. M. Steinmetz. 2007. Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc. Natl. Acad. Sci. USA* **104**:12825–12830.

63. Wesolowski, M., and R. B. Wickner. 1984. Two new double-stranded RNA molecules showing non-Mendelian inheritance and heat inducibility in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**:181–187.

64. Wickner, R. B. 1996. Double-stranded RNA viruses of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **60**:250–265.

65. Yu, X., and A. Gabriel. 1999. Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* **4**:873–881.