Mol Gen Genomics (2007) 277:43–55
DOI 10.1007/s00438-006-0167-z

ORIGINAL PAPER

Structure and transcriptional impact of divergent repetitive elements inserted within *Phanerochaete chrysosporium* strain RP-78 genes

Luis F. Larrondo · Paulo Canessa · Rafael Vicuña · Philip Stewart · Amber Vanden Wymelenberg · Dan Cullen

Received: 26 July 2006 / Accepted: 6 September 2006 / Published online: 11 October 2006

Abstract We describe the structure, organization, and transcriptional impact of repetitive elements within the lignin-degrading basidiomycete, *Phanerochaete chrysosporium*. Searches of the *P. chrysosporium* genome revealed five copies of *pce1*, a ~1,750-nt non-autonomous, class II element. Alleles encoding a putative glucosyltransferase and a cytochrome P450 harbor *pce* insertions and produce incomplete transcripts. Class I elements included *pcret1*, an intact 8.14-kb *gypsy*-like retrotransposon inserted within a member of the multicopper oxidase gene family. Additionally, we describe a complex insertion of nested transposons within another putative cytochrome P450 gene. The disrupted allele lies within a cluster of >14 genes, all of which encode family 64 cytochrome P450s. Components of the insertion include a disjoint *copia*-like element, *pcret3*, the *pol* domain of a second retroelement, *pcret2*, and a duplication of an extended ORF of unknown function. As in the case of the *pce* elements, *pcret1* and *pcret2/3* insertions are confined to single alleles, transcripts of which are truncated. The corresponding wild-type alleles are apparently unaffected. In aggregate, *P. chrysosporium* harbors a complex array of repetitive elements, at least five of which directly influence expression of genes within families of structurally related sequences.

Keywords *Phanerochaete chrysosporium* · Retrotransposons · Multicopper oxidase · Cytochrome P450

Nucleotide sequence data reported are available in GenBank under accession numbers DQ400694, DQ444269, AY532142, DQ444268 and DQ677350.

L. F. Larrondo · P. Canessa · R. Vicuña
Departamento de Genética Molecular y Microbiología,
Facultad de Ciencias Biológicas,
Pontificia Universidad Católica de Chile and
Instituto Milenio de Biología Fundamental y Aplicada,
Santiago, Chile

P. Stewart · A. Vanden Wymelenberg
Department of Bacteriology,
University of Wisconsin,
Madison, WI 53706, USA

D. Cullen (✉)
USDA Forest Products Laboratory,
One Gifford Pinchot Drive,
Madison, WI 53726, USA
e-mail: dcullen@facstaff.wisc.edu

Present Address:
P. Stewart
Rocky Mountain Laboratories,
NIAID, NIH, Hamilton, MT, USA

Introduction

A diverse array of transposable elements (TEs) have been identified in fungi, most of which are categorized as class I (transposing via RNA intermediate) or class II (DNA intermediate) (reviewed by Kempken and Kuck 1998; Wostemeyer and Kreibich 2002; Daboussi and Capy 2003). Transposition is generally thought to influence genome organization and/or gene expression, although there is little experimental support for this view particularly within the major phylum, Basidiomycota.

In the widely studied lignin-degrading basidiomycete *Phanerochaete chrysosporium* (Cullen and Kersten

2004), a remnant of a class II element, *pce1*, is inserted within lignin peroxidase allele *lipI2*. The 1,747-nt insertion transcriptionally inactivates *lipI2* and is inherited in a strict Mendelian fashion (Gaskell et al. 1995). Southern blots of pulsed field gels localized four *pce*-like sequences to a single chromosome band, and segregation analysis detected distant linkage among three copies designated *pce1*, *pce2*, and *pce3* (Gaskell et al. 1995; Stewart et al. 2000). Class II elements of Homobasidiomycetes also include *Abr1* identified in *Agaricus bisporus* (Sonnenberg et al. 1999) and *Scooter*, which has been shown to disrupt two genes regulating signal transduction in *Schizophyllum commune* (Fowler and Mitton 2000).

A high quality draft assembly of the *P. chrysosporium* genome has been generated (http://www.genome.jgi-psf.org/whiterot1/whiterot1.home.html). Initial analysis revealed families of structurally related genes some of which are thought to be involved in lignocellulose degradation, e.g., cytochrome P450s, glycosyl hydrolases, copper radical oxidases, and multicopper oxidases (Martinez et al. 2004). The latter family occurs as a cluster of four genes designated *mco1* through *mco4* (Larrondo et al. 2003, 2004). Recently, Yadav and coworkers have examined the complex structure and organization of the cytochrome P450 families (Doddapaneni et al. 2005).

A pure whole genome shotgun sequencing strategy was used for the *P. chrysosporium* genome, and this approach tends to exclude highly repetitive sequences such as rDNA and TEs. Nevertheless, non-coding repetitive elements as well as class I and II transposons were identified (Martinez et al. 2004). The class I retroelements include *copia*-like, *gypsy*-like, and LINE sequences (reviewed by Kempken and Kuck 1998; Wostemeyer and Kreibich 2002; Daboussi and Capy 2003). Many of these elements appeared as truncated remnants and/or rearranged, and their long terminal repeats often lie apart as "solo LTRs" (Kim et al. 1998; Goodwin and Poulter 2000).

Most recently, a substantially improved assembly (v2.0) and revised gene models (v2.1) have become available (http://www.genome.jgi-psf.org/Phchr1/Phchr1.home.html). We have systematically examined the insertional context of repetitive elements within the latest assembly. *pce*-like insertional mutations were detected in a cytochrome P450 and a glycosyltransferase gene. Further, *gypsy*- and *copia*-like retroelements have insertionally mutated members of multicopper oxidase and cytochrome P450 gene families, respectively. Results are consistent with an important role for transposons in generating genetic variation in *P. chrysosporium*.

Materials and methods

Organism and culture conditions

Phanerochaete chrysosporium strains BKM-F-1767 and RP-78 were used throughout. Typical of basidiomycetes, the vegetative mycelium of BKM-F-1767 harbors two distinct haploid nuclei in constant association. This dikaryotic strain has been extensively used for decades in studies of *P. chrysosporium* genetics and physiology. Homokaryotic derivative RP-78 was isolated by regenerating protoplasts derived from BKM-F-1767 mycelium. The nuclear condition and physiological characteristics of RP-78 were established (Stewart et al. 2000) and the strain used for whole genome shotgun sequencing (Martinez et al. 2004). All alleles from the RP-78 haplotype have been arbitrarily assigned an "A" suffix. Both strains are available from several culture collections (e.g., Forest Mycology Center, Forest Products Laboratory, Madison, WI, USA).

For RNA, 200 ml defined medium (Eriksson and Hamp 1978) amended with 0.4% Avicel PH-101 (Fluka Chemie, Buchs, CH) was inoculated with 1×10^7 spores in 2-l flasks. Incubation was at 37°C, 250 rpm in a shaking incubator. The cultures were harvested after 6 days by filtration through Miracloth (Calbiochem, La Jolla, CA, USA). The mycelium was snap frozen in liquid N_2 and stored at −90°C.

Amplification of genomic and cDNA targets

PolyA RNA from frozen pellets was isolated by magnetic capture as described (Vanden Wymelenberg et al. 2002, 2005). First strand cDNA was synthesized by preparing a 500-µl reverse transcription master mix containing 1× PCR buffer (Promega Inc., Madison, WI, USA), 5 mM $MgCl_2$, 4 mM dNTPs, 500 units RNAsin (Promega), 105 pmol oligo dT_{15}, 1,250 units MMLV-RT (Invitrogen Corp., Carlsbad, CA, USA), and 100 µl mRNA. Fifty microliter aliquots were divided among 0.5-ml Eppendorf tubes and placed in a Perkin Elmer DNA Thermal Cycler 480 (Applied Biosystems, Foster City, CA, USA). Cycling conditions were 23°C 10 min, 42°C 45 min, 95°C 5 min. All tubes were combined and stored at −20°C.

Genomic and cDNA targets were amplified by PCR in a Perkin Elmer DNA Thermal Cycler 480. Each PCR reaction contained 1× PCR buffer, 1 mM $MgCl_2$, 1.25 units Taq DNA polymerase (Promega Inc., Madison, WI, USA), 10% DMSO (Sigma-Aldrich, St Louis, MO, USA), 11 pmol each primer and 10 ng DNA template. Cycling conditions were 94°C 6 min, 54°C 2 min, 72°C 40 min (1 cycle), followed by 94°C 1 min, 54°C 2 min,

72°C 5 min (35 cycles), 72°C 15 min (1 cycle). Using primer D (Table 1) and primer I (5′-CCGTTCCATCTGCACGGACA-3′), elongase (Invitrogen Corp., Carlsbad, CA, USA) was used for long-range PCR amplification of *pcret1*. Elongation time was 10 min and the annealing temperature was 58°C. Buffer A and B were used in a 2:3 proportion. Amplicons to be sequenced were first cloned using pGEM-T Easy (Promega). Primer sequences are listed in Tables 1 and 2.

Computational analysis

Overlapping approaches were used to detect transposon-interrupted coding regions in assemblies v1.0 and v2.0. Blastp and tblastn queries with well-known fungal TEs (Daboussi and Capy 2003), tyrosine recombinase retrotransposons (Goodwin and Poulter 2004), and helitrons (Poulter et al. 2003) identified numerous elements as did word or phrase searches (http://www.genome.jgi-psf.org:8080/annotator/servlet/jgi.annotation.-Annotation?pDb=Phchr1). Significant hits ($E<10^{-4}$) were rarely intact, and subsequent analysis of the adjacent regions for direct and inverted repeats were performed using GeneQuest (DNASTAR, Madison, WI, USA). In an iterative fashion, putative *P. chrysosporium* nucleotide repeats and TE-related proteins were used to re-scan the genome by blastn and blastp/tblastn, respectively. For each hit, the completeness of surrounding gene models was assessed. In the absence of adjacent gene models, the flanking regions were examined for extended ORFs and subject to blastx searches of NCBI. Finally, the 24-longest scaffolds (representing 95% of assembly v2.0) were examined for partial or rearranged TEs by manually scanning the browser's repeat track. In assembly v2.0, this track encompasses well-known fungal elements (e.g., *maggy*, *skippy*, etc.) and short

Table 1 PCR primers and predicted product lengths (nt)

Primers[a]	Sequence[b]	Homokaryon[c]		Dikaryon[d]	
		cDNA	Genomic	cDNA	Genomic
pce4					
D/C N-term	GCGTAGACAATGCCAGATAC CACGGATGTCGCGGAGT	244	244	244	244
D/A flanking	GCGTAGACAATGCCAGATAC TTCCGCGGCGAAGGGCCAG	1,386	3,605	1,386	1,858
B/A C-term	GCTGCCTATGACAAAGAGG TTCCGCGGCGAAGGGCCAG	421	636	421	636
pce5					
A/B N-term	GGAATGAGCCAGTGGAGG CGGCAAAGTGACCTCGAC	384	486	384	486
A/D flanking	GGAATGAGCCAGTGGAGG CTCTGGCTTGAACTCGTA	1,133	3,687	1,133	1,722
C/D C-term	CATACCCTGGCATTCACTT CTCTGGCTTGAACTCGTA	343	564	343	564
pcret1					
A/B N-term	GACGAACTACTACGACGGGA GAATCCTCCGTGAGCACAAA	1,008	1,537	1,008	1,537
A/E flanking	GACGAACTACTACGACGGGA CCGCGTCGGCAGGCTGTTCA	1,293	10,122	1,293	1,994
C/D C-term	GTACATTGGCCAGCCGCTGA CTAACGTCGTGCACCTGCTC	261	314	261	314
pcret2/3					
A/B N-term	TGGCGGAAATCTCTGCAAGT TTGAAGCTTGACGCTTGGATC	102	214	102	214
A/D flanking	TGGCGGAAATCTCTGCAAGT TGGTCTGACATTGGAAAGGCT	870	29,296	870	1,360
C/D C-term	ACGAACTCGACATGGTGCTAG TGGTCTGACATTGGAAAGGCT	504	711	504	711

Predictions based on manually curated genome sequence and experimentally determined cDNA sequence

[a] Relative position of primers illustrated in Figs. 2, 3, 4, and 5

[b] 5′ to 3′

[c] Homokaryotic strain RP-78. Calculations of cDNA amplicon lengths are based on precise excision of inserts

[d] Parental dikaryon BKM-F-1767. Both allelic targets may be present, but PCR conditions favor smaller amplicons and rarely amplify products over 3 kb. Only "B" allele predictions are listed

Table 2 Primers extending transcripts in RP-78 homokaryon

Primers[a]	Sequence[b]	cDNA[c]	Genomic
pce4 C-terminal extension (Fig. 2)			
A/B	TTCCGCGGCGAAGGGCCAG GCTGCCTATGACAAAGAGG	421	636
A/E	GCAGAGTATCAGAAGCTCCTGA	982	1,394
A/F	GTGGTGGATGAAATCGCAGCGA	1,112	1,576
A/G	CGTCTGAATGCTTGCGCGTCTG	1,217	1,681
A/H	GACGTCCAGGGCTGCATCCTCA	1,546	2,010
pce5 N-terminal extension (Fig. 3)			
A/B	GGAATGAGCCAGTGGAGG CGGCAAAGTGACCTCGAC	384	486
A/E	TTACTTGGAGGTGACATCCTAC	502	656
A/F	AGGTGCGCCATGTTGTCAATGC	900 (779)	1,054
A/G	GAGTAGATGCAGAGCATGAGTG	1,411 (1,290)	1,565
A/H	CTTCACACGCGCAGTTGCAACG	1,868 (1,747)	2,022
pce5 C-terminal extension (Fig. 3)			
C/D	CATACCCTGGCATTCACTT CTCTGGCTTGAACTCGTA	404	564
I/D	CTGTTTGTCATTGGTGTCGCCG	737	1,172
J/D	CGTCTTGATGCTTGTGCGTCTG	813	1,248
K/D	CGTTGCAACTGCGCGTGTGAAG	1,252	1,687
L/D	CACTCATGCTCTGCATCTACTC	1,709	2,144
pcret1 N-terminal extension (Fig. 4)			
A/B	GACGAACTACTACGACGGGA GAATCCTCCGTGAGCACAAA	1,008	1,537
A/F	TGTCCGTGCAGATGGAACGGAT	1,086	1,659
A/G	TTAGGCAGGCCAGCCGATGCGT	1,508	2,081
A/H	GGATTGTGTGGAGTCAGTGGAC	1,845	2,418
pcret2/3 N-terminal extension (Fig. 5)			
A/B	TGGCGGAAATCTCTGCAAGT TTGAAGCTTGACGCTTGGATC	102	214
A/F	TCCTTCATGTTGGCAGGATCGC	239	406
A/G	TGAAGGCCAAGCACTTCGCTTC	776	1,040
A/H	ACCTGCACGAAGTCTGGCAAAT	1,001	1,265
pcret2/3 C-terminal extension (Fig. 5)			
C/D	ACGAACTCGACATGGTGCTAG TGGTCTGACATTGGAAAGGCT	504	711
I/D	CACGTGGATGTTGGCTATGACC	561	764
J/D	CGAGAGACAGGGGCGAACGAAG	1,076	1,287
K/D	CGACAACGCCTATGGGCTGTGA	1,276	1,487

[a] Relative position of primers illustrated in Figs. 2, 3, 4, and 5. To eliminate redundancy, only underlined sequences are given

[b] 5′ to 3′

[c] Predictions based on manually curated genome sequence and experimentally determined cDNA sequence. Parenthetical sequence lengths based on processing of an intron within the *pce5* element

non-coding repeats. Because many transposon-related domains (integrases, reverse transcriptases) are not captured in final v2.1 "Best models" (Vanden Wymelenberg et al. 2006), manual examination included Fgenesh ab initio models, particularly those with unusually long intervening sequences.

Unless otherwise specified, our analysis emphasized assembly v2.0. This assembly was generated by a pure shotgun approach, and the homology-based gene predictions were recently described (Vanden Wymelenberg et al. 2006). The v2.0 and archived v1.0 assemblies, predicted genes, supporting evidence, annotations, and analysis are accessible through an interactive browser and tools at the Joint Genome Institute's web portal (www.jgi.doe.gov/whiterot). The protein information and related links for the most recent v2.1 genes can be rapidly accessed by appending model numbers to the end of the following URL: http://www.genome.jgi-psf.org/cgi-bin/dispGeneModel?db= Phchr1&id=. Throughout, similarity scores are based on the Smith–Waterman algorithm (Smith and Waterman 1981) using the BLOSUM62 matrix. ClustalW (Thompson et al. 1994) generated multiple alignments using DNASTAR software.

Results

Structure and organization of non-autonomous *pce1*-like repeats

Five sequences >40% identical to *pce1* (Gaskell et al. 1995) were identified by blastn searches of assembly

v2.0. Elements *pce2* (AF134289), *pce3* (AF134290), and *pce4* (AF134291) are nearly identical to *pce1*, as previously shown (Stewart et al. 2000). Designated *pce5*, a 1,967-nt paralog was identified on scaffold 9 (coordinates 438,871–440,837). The *pce5* nucleotide sequence is 44.5% identical to *pce1* and, like *pce1–pce4*, contains no extended ORFs. A truncated element on scaffold 33 (coordinates 5,333–6,260), *pce6*, extends only 928 nt but is otherwise identical to *pce1*. Blastn and blastx searches of NCBI or the v2.0 database failed to identify any additional *pce1*-like sequences.

Excluding the incomplete *pce6* element, the sequences show remarkable sequence conservation at their termini. All but *pce3* are flanked by dinucleotide repeats, and all feature imperfect inverted terminal repeats (ITRs, Fig. 1). But more surprisingly, the positions of these seven "imperfections" are consistently retained in *pce1*, *pce2*, *pce3*, and *pce4*. The more divergent element, *pce5*, retains five of these nucleotides and features two additional imperfect bases. In addition to the ITRs, a 16-bp direct repeat (GTTTGTGCATGTCTGT) was also conserved in *pce1*, *pce2*, *pce3*, and *pce4* (positions 1014 and 1602 in *pce1*).

Corroborating the long-range structure of assembly v2.0 and consistent with earlier segregation analysis (Stewart et al. 2000), *pce1*, *pce2*, and *pce3* were located on a single scaffold (number 19). Element *pce4* resides on a separate scaffold (number 8), an observation consistent with cosegregation analysis showing no detectable linkage (Stewart et al. 2000). However, the v2.0 assembly and segregation data are at variance with pulsed field gel blots which suggest *pce4* resides on the same 3.7-Mb chromosome as *pce1*, *pce2*, and *pce3* (Gaskell et al. 1995). Possibly, *pce4* is distantly linked, and scaffolds 8 and 19 may lie on the same chromosome. Elements *pce5* and *pce6* are located on scaffolds 9 and 33, respectively. Regions surrounding the elements were examined for the presence of incomplete gene models and/or extended ORFs. None were detected adjacent to *pce2* and *pce3*.

Integration context and transcriptional impact of *pce4*

Gene model 4905 (v2.1), encoding a putative UDP-glucosyltransferase (*ugt1A*), lies 184 nt to the left of *pce4* on scaffold 8 (coordinates 291,078–292,799). Blastp and blastx analysis, together with manual inspection of the region, suggested an incomplete N-terminus. Automated gene predictions failed to assign any models to the right of *pce4*, but an ORF potentially continuing *ugtA* was observed. PCR primers were designed within this ORF (Fig. 2a, primer D) and within the putative glucosyltransferase coding region (Fig. 2a, primer A). Using genomic DNA templates, the outermost primer pair (A/D) efficiently PCR amplified the 3,605-nt *ugt1A* allele from the homokaryon (Fig. 2b, lane 6) and the uninterrupted 1,858-nt *ugt1B* allele from the parental dikaryon (Fig. 2b, lane 12). Using dikaryon-derived RNA and the same three primer pairs, cDNAs corresponding to *ugt1B* were RT-PCR amplified (Fig. 2b, lanes 7–9) and the longest amplicon (1,386 nt) sequenced (GenBank DQ400694). (The 244-nt N-terminal *ugt1B* cDNA is only faintly visible in Fig. 2, lane 8.) Comparisons of *ugt1A* genomic sequence and the *ugt1B* cDNA pinpointed *pce4* insertion within *ugt1A* coding region. Using RP-78-derived RNA as template, RT-PCR amplifications of *ugt1A* (Fig. 2b, lanes 1–3) generated a 421-nt C-terminal cDNA. This transcript was subsequently shown to extend through the coding region (Fig. 2c, primer pairs A/B, A/E, A/F) but primers located >45 nt inside *pce4* (primer G, H) failed to amplify a *ugt1A* cDNA. Additional primers tested within this region yielded non-specific products (data not shown), and the precise transcriptional start point remains uncertain. Irrespective of the transcript start point, these results clearly indicate that *pce4* interferes with *ugt1A* transcription.

Blast analysis using the *ugt1B* cDNA and the corrected 4905 gene model showed greatest similarity to higher plant glucosyltransferases including *Oryza sativa*

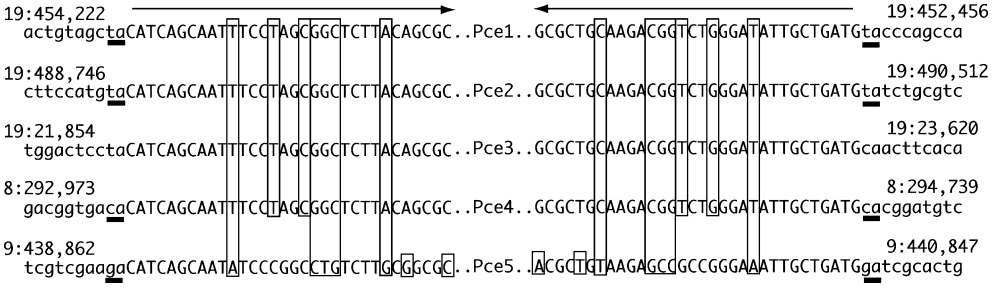

Fig. 1 Inverted terminal repeats (*capital letters*) of elements *pce1* through *pce5*. Imperfect bases are *boxed*. Ten base pairs of flanking region (*lower case letters*) are shown with dinucleotide repeats *underlined*. *Numbering* on *margins* indicate scaffold no.: position on scaffold

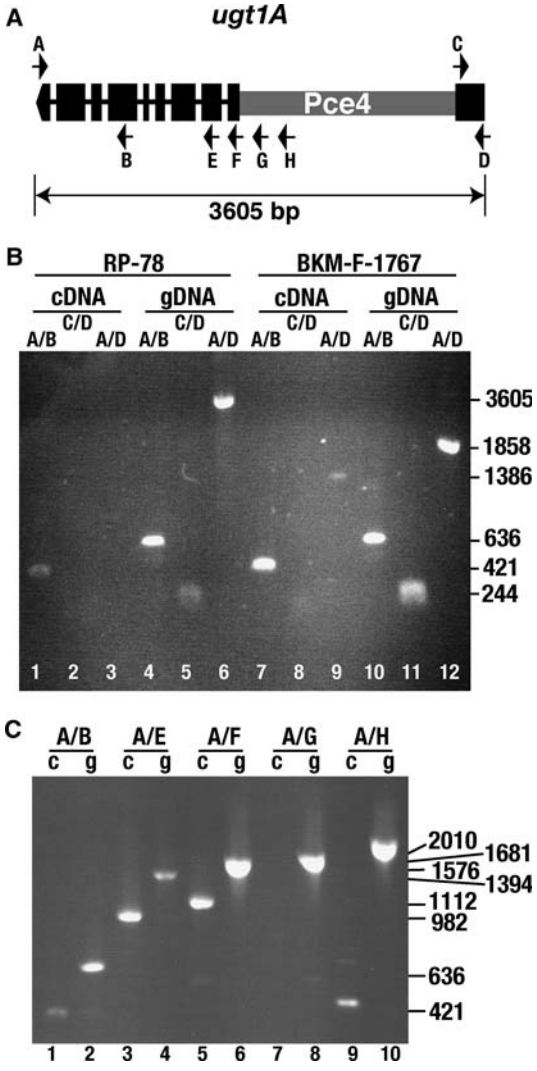

Fig. 2 Context and transcriptional consequence of *pce4* insertion within *ugt1A*. Position and orientation of all primers (*short arrows*) are shown in *panel a* relative to exons (*thick black boxes*), introns (*thin black lines*), and *pce4* (*gray line*). In *panel b*, genomic DNA (gDNA) and cDNA derived from homokaryotic (RP-78) and dikaryotic (BKM-F-1767) strains were amplified with primer pairs A/B, C/D, and A/D. Based on these results, the full-length of C-terminal transcript was further examined using RP-78 RNA (*panel c*, *c lanes*). RP-78 genomic DNA (*panel c*, *g lanes*) served as positive controls. Primer sequences and predicted product lengths are in Tables 1 and 2. Amplicon lengths in nt are shown to the *right* of the ethidium bromide-stained gels. Low molecular weight extraneous bands barely visible in *panel c*, *lanes 5* (~550 nt) and *9* (~700 nt, 440 nt) were sequenced and identified as non-specific amplicons

hydroquinone glucosyltransferase (gi 50252251, bit score 104) and a multifunctional triterpene/flavanoid glycosyltransferase from *Medicago truncatula* (gi 83753974, bit score 102). No significant fungal homologues were detected by blastn, blastx, or blastp searches of NCBI. In contrast, *P. chrysosporium* v1.0 gene models gx.22.38.1, pc.22.87.1, and pc.22.101.1 are 76, 66, and 53% identical to the deduced protein of *ugt1B*, respectively. (No corresponding v2.1 models were generated.) Sequences corresponding to these paralogs reside within a 16-kb region on v2.0 scaffold 6 (coordinates 869,664–885,611).

Integration context and transcriptional impact of *pce5*

pce5 lies on scaffold 9 between gene models 5526 and 5527 (v2.1), which are predicted to encode family CYP614/534 cytochrome P450 proteins. Close inspection of these automated model predictions suggested a single gene with the *pce5* insertion (Fig. 3a). Employing a strategy similar to that used for *pce4*, primers were designed to assess the transcriptional consequence of insertion (Fig. 3a). Using RNA derived from the parental dikaryon, an 1,133-nt cDNA corresponding to *cyp614/534B* was RT-PCR amplified, cloned, and sequenced (Fig. 3b, lane 9, GenBank DQ444269). Comparisons of cDNA and genomic sequences pinpointed the insertion to the 3′ end of an intron suggesting the possibility of *pce5* splicing. However, no flanking cDNA product was obtained from strain RP-78 (Fig. 3b, lane 3), indicating the insert is not spliced from the transcript, at least under the culture conditions examined.

Transcripts corresponding to the N- and C-termini of *cyp614/534A* were detected (Fig. 3b, lanes 1 and 2), and further examined by RT-PCR amplification of RP-78 RNA (Fig. 3c, d). The N-terminal (Fig. 3c, primer pairs A/E, A/F) and C-terminal transcripts (Fig. 3d, primer pair J/D) extended at least 512 and 68 nt, respectively, into *pce5*. Longer cDNAs were not detected (Fig. 3c, primers A/G, A/H; Fig. 3d, primer pairs K/D, L/D) even though these primer pairs efficiently amplified the corresponding genomic regions (Fig. 3c, lanes 8 and 10; Fig. 3d, lanes 8 and 10). RT-PCR amplification with primer pair A/F (Fig. 3c, lane 5) yielded a cDNA of 779 nt (GenBank DQ677350), 121 nt less than the predicted amplicon. Comparison with genomic DNA revealed the presence of an intron within this region of *pce5*. Although considerably larger than most *P. chrysosporium* introns (mode = 54 nt), canonical 5′ and 3′ splice sites (Martinez et al. 2004) were clearly present. Truncated chimeric transcripts have also been detected in *Fot1* mutated nitrate reductase (*niaD*) alleles of *Fusarium oxysporum* (Deschamps et al. 1999).

Blast analysis using the cDNA and corrected models showed the *cyp614/534A* sequence was most closely related to a hypothetical protein from *Gibberella zeae* (gi 42546901, bit score 164). *P. chrysosporium* models

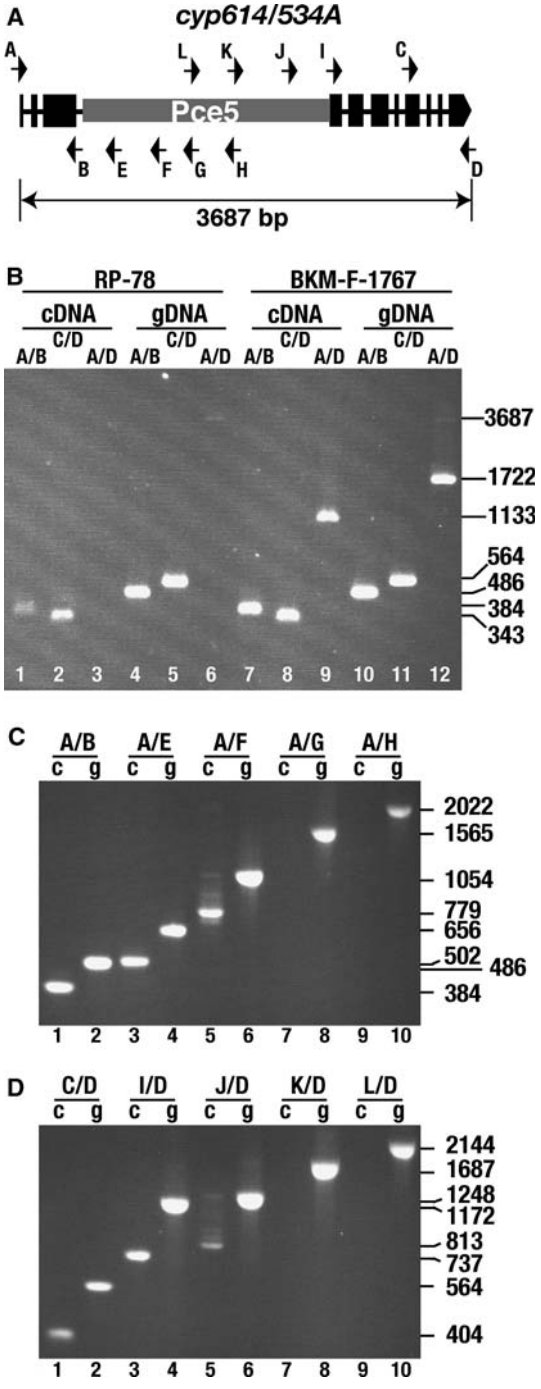

Fig. 3 Context and transcriptional consequence of *pce5* insertion within *cyp614/534A*. Using a strategy similar to *pce4* (Fig. 2a), primer position and orientation relative to exons (*thick black boxes*), introns (*thin black lines*), and *pce5* (*gray line*) are shown in *panel a*. Genomic DNA (gDNA) and cDNA derived from homokaryotic (RP-78) and dikaryotic (BKM-F-1767) strains were amplified with primer pairs A/B, C/D, and A/D (*panel b*). With increased brightness/contrast control, *faint bands* corresponding to genomic *cyp614/534A* (3,687 nt) were visible (*panel b*, *lanes 6* and *12*; data not shown). Amplification of cDNAs corresponding to the N- and C-termini of *cyp614/534A* (*panel b*, *lanes 1* and *2*), prompted further investigation of these transcripts using RP-78 cDNA (*panels c* and *d*). RP-78 genomic DNA template (*g lanes*) served as positive controls for cDNA amplifications (*c lanes*). Primer sequences and predicted product lengths are in Tables 1 and 2. *Faint bands* slightly larger than the target cDNAs (*panel c*, *lane 5*; *panel d*, *lane 5*) were sequenced and identified as splice variants with retained introns

4556, 3368, 4557, and 6351 were 47, 55, 58, and 60% identical, respectively, to the protein sequence deduced from cDNA. Of these, only models 4556 and 4557 are linked, occurring within a 5.3-kb region of scaffold 7. However, blastx analysis of a region 6.2 kb to the right of model 5527 (coordinates 448,750–452,500), reveals the presence of a more distantly related CYP614/514-like sequence. (No v2.1 models were generated in this region.)

Identification of *pcret1* in a *mco3* allele

On scaffold 9 of assembly v2.0, clustered multicopper oxidase genes *mco1A*, *mco2A*, and *mco4A* appear intact (Larrondo et al. 2004; Fig. 4a), but *mco3A*, located furthest toward the scaffold's left end, contains an extended intervening sequence within the 12th intron (http://www.genome.jgi-psf.org/cgi-bin/dispGeneModel?db=Phchr1&id=132237). Sequence analysis revealed the presence of an 8.14-kb insertion with features common to *gypsy*-class retroelements. Designated *pcret1*, the element contains identical LTRs of 336 bp flanking a 7.46-kb region with two extended ORFs (Fig. 4e). A 5-bp target site duplication (TSD) with the sequence AGTCT flanks the 5′ and 3′ LTRs (Fig. 4d). The first ORF present in *pcret1* extends 1,417 bp, and corresponds to a predicted *gag* protein of 475 residues. The second ORF of 5,222 bp encodes a putative *pol* protein, with domains corresponding to protease, reverse transcriptase, RNAseH, and integrase (Fig. 4e). Blastx searches of the NCBI database show the *pol* protein most closely related (bit score 786) to *Tricholoma matsutake marY1* (gi 5002510), and the full-length sequence is most closely related (bit score 709) to an unnamed *Cryptococcus neoformans* retrotransposon (gi 57227612).

As in the case of *pce*, *pcret1* insertion altered transcript patterns. cDNAs corresponding to the N-terminal, C-terminal, and flanking regions were detected in the parental dikaryon (Fig. 4f, lanes 7–9, respectively). In contrast, RT-PCR amplification of RNA derived from the homokaryon only yielded a 1,008-nt cDNA corresponding to the N-terminal region (Fig. 4f, lane 1). Additional RT-PCR amplifications of RP-78 RNA slightly extended this transcript (Fig. 4g, lane 3), but none was obtained using primers lying within the

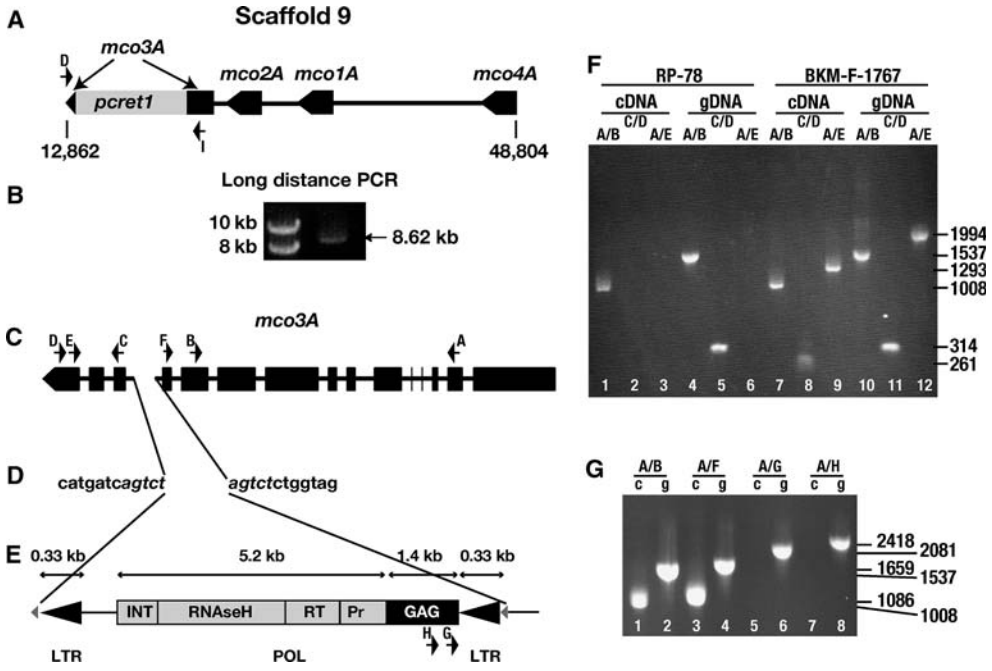

Fig. 4 The *gypsy*-like element *pcret1* is inserted within *mco3A*, a member of the multicopper oxidase family located on scaffold 9 (*panel a*). *Scaffold coordinates* are shown below *mco3A* and *mco4A*. Contiguity of the insertion was confirmed by long-range PCR using primers D and I (*panel b*). Positions of exons (*thick black lines*) and introns (*thin black lines*) were established by comparisons of cDNA and genomic sequences (*panel c*). Insertion occurred within the 12th intron of *mco3A* at which position a 5-nt duplication is observed (*italicized lettering*, *panel d*). *Panel e* shows the structural organization of *pcret1* with long terminal repeats (LTRs) and *pol* protein coding regions. Genomic DNA (gDNA) and cDNA derived from homokaryotic (RP-78) and dikaryotic (BKM-F-1767) strains were amplified with primer pairs A/B, C/D, and A/E (*panel f*). The length of the N-terminal *mco3A* transcript (*panel f*, *lane 1*) was further examined using RP-78 RNA (*panel g*, *c lanes*), and RP-78 genomic DNA (*g lanes*) served as positive controls. Primer sequences and predicted product lengths are in Tables 1 and 2

adjacent *gag* region (Fig. 4g, lanes 5 and 7). The absence of transcript flanking the insertion point and transcript 3′ to *pcret1* (Fig. 4f, lanes 2 and 3) strongly suggest that *pcret1* is not spliced, at least under the conditions tested.

pcret1-related elements

Blastn analysis showed a minimum of 36 sequences with >95% sequence identity to the *pcret1*-LTR. Often unpaired and without nearby retroelement domains (e.g., reverse transcriptase, integrase), the LTRs were distributed on 25 separate scaffolds. Examination of automated models in the surrounding regions and blastx submissions (NCBI) of flanking sequences failed to identify insertions within detectable genes. *pcret1*-like coding regions flanked by the highly conserved LTRs were detected on scaffold 2 (coordinates 2,656,822–2,673,069 and 2,464,083–2,471,738) and on scaffold 7 (coordinates 1,655,590–1,665,209 and 1,978,909–1,987,005). In these instances, the *pol* coding regions are interrupted by sequence ambiguities of varying lengths.

Identification of *pcret2/3* in a putative cytochrome P450 allele

Initial analysis of the draft genome assembly (v1.0) suggested the presence of several *copia*-like retroelements, including one located within a putative cytochrome P450 gene (Martinez et al. 2004). Analyses of the more recent assembly (v2.0) reveal an extensive cluster of >14 family 64 cytochrome P450s on scaffold 1, 5 of which lie within a 43.5-kb region (coordinates 2,490,057–2,533,590) (Fig. 5a). Models 132401, 133482, 133291, and 132914 appear intact while model 791 is incomplete due to a 27,869-nt insertion (Fig. 5a). The assemblies generally agree in this region, except the insertion is longer and more complex in v2.0 relative to v1.0.

Components of this complex insertion appear to include a disjoint *copia*-like element, *pcret3*, the *pol* domain of a second retroelement, *pcret2*, and a duplication of an extended ORF of unknown function (Fig. 5d). LTRs (356 bp) lie adjacent to ORFs corresponding to the carboxy (3′-*pcret3*) and N-terminal (5′-*pcret3*) regions of *copia*-like retrotransposons from

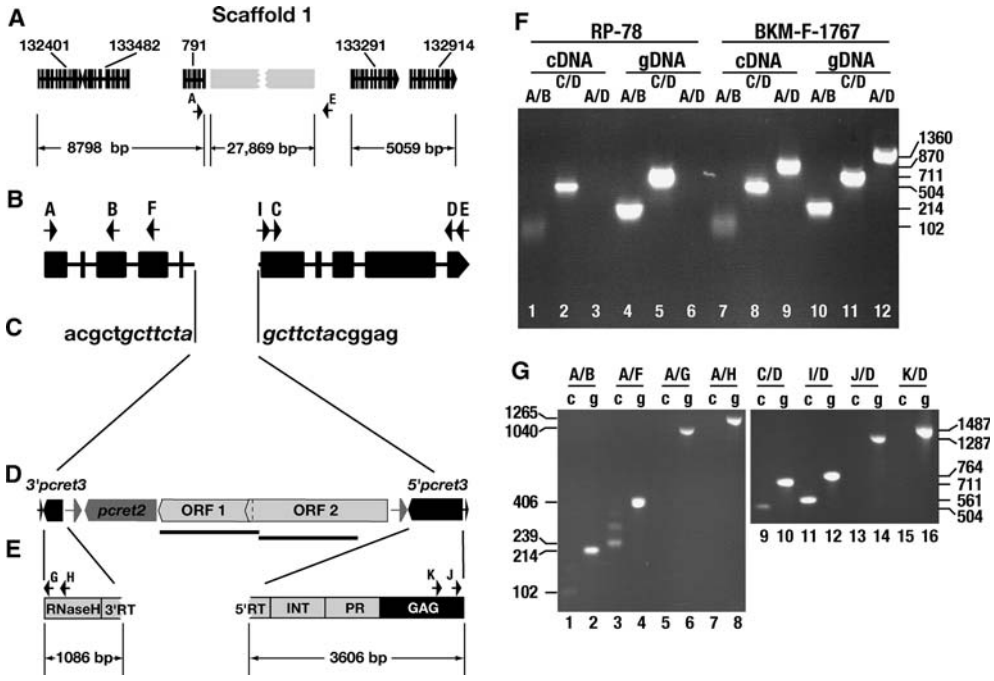

Fig. 5 Family 64 cytochrome P450 gene with complex insertion. Schematic of five CYP64 gene models (*black vertical lines* represent predicted exons) and complex insertion (*broken gray box*) is illustrated in *panel a*. Primers A (Table 1) and E (5′-CTCTGCCTCTGGAAAACGCA-3′) are designed to the predicted coding region of model 791 and to a potential ORF, respectively. Positions of exons (*thick black lines*) and introns (*thin black lines*) were established by comparison of dikaryon-derived cDNA encoding allele B with the genomic sequence of the RP-78 A allele (*panel b*). Intron sequence surrounding insertion point with potential duplication in *italics* is shown in *panel c*. Structural elements identified within the insertion (*panel d*) include five extended ORFs (*boxed arrows*) and two pairs of long terminal repeats (LTRs, *thin-lined arrows*). The outermost 356 bp LTRs lie adjacent to *copia*-like element *pcret3*. The *pcret3* ORFs of 1,086 and 3,606 bp are interrupted by another retroelement *pcret2* (4,566 bp), ORF1 (6,027 bp), ORF2 (9,366 bp), and 984 bp LTRs (*gray arrows*). An expanded view of *pcret3* showing domain organization typical of *copia* elements is shown in *panel e*. Genomic DNA (gDNA) and cDNA derived from homokaryotic (RP-78) and dikaryotic (BKM-F-1767) strains were amplified with primer pairs A/B, C/D, and A/D (*panel f*). cDNAs of 102 nt (*panel f*, *lane 1*) and 504 nt (*panel f*, *lane 2*) correspond to the N- and C-terminal P450 coding regions, respectively. Length of transcripts was further examined using RP-78 RNA (*panel g*, *c lanes*), and RP-78 genomic DNA (*g lanes*) served as positive controls. Primer sequences and predicted product lengths are in Tables 1 and 2. The non-target ~320-nt band (*panel g*, *lane 3*) was sequenced and identified as a splice variant with retained intron

Ipomoea batatas (gi 50838657, bit score 563), *O. sativa* (gi 77555860, bit score 516), and other higher plants. The approximate positions of *gag*, protease, integrase, reverse transcriptase/RNAseH domains were inferred from the presence of conserved motifs (Jordan and McDonald 1999) and alignment with the related *Candida albicans copia* element, *tca5* (gi 68492301, bit score 361; Plant et al. 2000; Fig. 5e).

The *pcret2* sequence (Fig. 5d) encodes a predicted protein of 1,511 residues with significant similarity to *pol* proteins from *Yarrowia lipolytica* (gi 50554773, bit score 344) and *Schistosoma mansoni* (gi 44829171, bit score 281, DeMarco et al. 2004). Conserved integrase core (pfam00665) and reverse transcriptase (pfam00078) domains align at amino acid positions 557–681 and 1164–1317, respectively. Immediately to the left of the *pcret2* ORF lies a 984-nt direct repeat that is duplicated approximately 20 kb to the right. It seems likely that the minimal active progenitor of *pcret2* is composed of these LTRs and the *pol* proteins contained within the 4,533-nt coding region. This view is supported by the structure of closely related "intact" elements within the *P. chrysosporium* genome (below).

In addition to the LTRs and retroelements within this large insertion, two extended ORFs are present (Fig. 5d). Overlapping by 165 nt, ORF1 and ORF2 encode nearly identical proteins of 2009 and 3122 amino acids, respectively. The duplicated nucleotide sequences (underlined in Fig. 5d) are 6,639 nt long and >99% identical. Blastp and blastx analysis of these ORFs shows no significant similarity to any proteins. However, they are distantly related (bit score 59) to a *P. chrysosporium* conceptual translation (gi 71148689) containing a putative retroviral aspartyl protease domain (RVP, pfam00077). Possibly, these ORFs represent duplicated fragments of divergent retroelements.

To ascertain the intron/exon context of the *pcret2/3* insertion, cDNA corresponding to the P450 "B" allele was RT-PCR amplified from the parental dikaryon (GenBank DQ444268). Sequence analysis extended and corrected errors in the truncated model (Fig. 5a, model 791). Consistent with a recent insertion, allele comparisons showed 96% nucleotide and 99% amino acid identity. The insertion was pinpointed to an intron with a putative 7-nt TSD (Fig. 5b, c).

As in the case of *pcret1*, insertion within an intron and the absence of deleterious mutation within the P450 coding regions suggested the possibility of an intact transcript through splicing. However, transcripts flanking the insertion point were RT-PCR amplified from the dikaryon (Fig. 5f, lane 9) but not the homokaryon (Fig. 5f, lane 3). As above, the data strongly suggest that the allele is functionally inactive due to the insertion.

Transcripts corresponding to the N- and C-termini of the mutated P450 gene were detected (Fig. 5f, lanes 1 and 2), and RT-PCR amplification of RP-78 RNA extended their lengths (Fig. 5g). However, in contrast to *pce5*, cDNAs did not extend into the element on either the N-terminus (Fig. 5g, primer pairs A/G and A/H) or the C-terminus (Fig. 5g, primer pairs J/D and K/D).

pcret2-related elements

Blast analysis of the *pcret2* LTRs and coding region yielded results similar to *pcret1*. Specifically, >10 highly conserved LTRs were identified some of which were adjacent to *pcret2*-like coding regions and others appeared to be "solo" LTRs. Tblastx searches of the *P. chrysosporium* genome revealed >15 highly similar (bit scores >200) *pol* proteins. On the left end of scaffold 15, *pcret2* and *pcret3* are again adjacent but the region is clearly truncated and the corresponding LTRs are absent.

pcret3-related elements

Using the reconstructed *pcret3* protein (Fig. 5e, 1563 amino acids), tblastn analysis of assembly v2.0 translations identified 49 sequences with significant similarity (bit scores >100). In most instances, further examination revealed significant similarity only to the 3′ *pol* proteins (integrase, reverse transcriptase, RNAseH). As in the case of *pcret1* paralogs, browser examination of nearby gene models, including ab initio FgenesH models, failed to identify additional mutations. These finding were confirmed by blastx queries (NCBI) using nucleotide sequences surrounding the retroelement paralogs.

Regions of extended *pcret3* homology, including conserved LTRs, were detected on assembly v2.0 scaffold 2 (coordinates 2,731,968–2,739,855) and scaffold 23 (coordinates 150,316–155,943). Although 250 nt shorter and containing sequence ambiguities, the scaffold 23 element was remarkably conserved, showing >94.1% nucleotide and 98.2% amino acid identity with *pcret3*. In contrast to blastn searches with *pcret1* LTRs, only four significant "hits" (bit scores >100) were obtained with the *pcret3* LTRs. Viewed together with the tblastn results, we conclude that *P. chrysosporium* strain RP-78 harbors many diverse *copia*-like elements, but *pcret3* copy numbers are relatively low. In this connection, it remains formally possible that additional copies may have been excluded from the shotgun assemblies.

Discussion

Transposable elements are widely distributed among the fungi and are presumed to play important roles in gene expression and genome organization. Most reports have focused on ascomycetes, although relatively few have emphasized TE location (Cambareri et al. 1998; Hua-Van et al. 2000; Thon et al. 2004) or identified spontaneous mutations in structural genes (Nishimura et al. 2000; Kang 2001; Kang et al. 2001; Fudal et al. 2005). Basidiomycetes have received comparatively less attention, but an increasingly diverse array of TEs have been characterized from *P. chrysosporium*, *A. bisporus*, *S. commune*, *T. matsutake*, *C. neoformans*, *Laccaria bicolor*, *Pisolithus microcarpus*, and *Microbotryum violaceum*. To our knowledge, *pce1* (Gaskell et al. 1995) and *scooter* (Fowler and Mitton 2000) represent the only examples of insertional mutagenesis in a basidiomycete. The *P. chrysosporium* draft genome provides an opportunity to extend our understanding of TE structure, organization and effects on gene expression.

Our analysis identified insertions in a minimum of five *P. chrysosporium* genes. These repeats included non-autonomous class II elements (*pce1*, *pce4*, *pce5*), a *gypsy*-like retroelement, *pcret1*, and a complex insertion containing rearranged retroelements *pcret2* and *pcret3*. Structurally, *pcret1* appears intact and similar to many other *gypsy*-class elements. In contrast, the *pcret2*/*pcret3* insertion features long, duplicated ORFs, and the LTRs are disjoint from their respective coding regions. Likely, this complex insertion represented nested and rearranged elements, a common feature of transposon organization in higher eukaryotes.

Independent *pcret2* and *pcret3* insertions are located elsewhere in the genome. Based on the relative position of reverse transcriptase and integrase domains, the overall structure of *pcret2* is similar to *gypsy*-class elements. However, a *gag* domain could not be clearly assigned in the absence of a zinc finger motif (Jordan and McDonald 1999). In contrast, *pcret3* organization is typical of numerous *copia*-class retroelements. Although relatively rare in fungi, an increasing number of *copia* elements have been identified in ascomycetes (reviewed by Daboussi and Capy 2003) and more recently in the basidiomycetes *L. bicolor*, *P. microcarpus* (Diez et al. 2003), and *M. violaceum* (Hood 2005).

In contrast to the retroelements, the *pce* sequences lack clear relationship to any known TE. Highly conserved, albeit imperfect, ITRs, together with putative TSDs are typical of non-autonomous class II elements. Such elements sometimes share ITR sequence identity with an autonomous transposase (Chomet et al. 1991; Rezsohazy et al. 1997; Hua-Van et al. 2000), but we were unable to identify any candidates in the RP-78 genome. Possibly, an active progenitor resides in other strains or in the parental dikaryon.

Undoubtedly, this analysis underestimates the frequency of TE-disrupted genes, in part because shotgun sequencing tends to exclude highly repetitive sequences. In addition, genome scanning relied on sequence similarity to known TEs and/or recognition of gene models with intervening sequences. Highly divergent genes, particularly those disrupted by long elements, might not be modeled by Fgenesh. Fortunately, in the case of *pce4* (Fig. 3a) and *pcret2/3* (Fig. 5a) insertions, at least one flanking model could be recognized. Still, our approach would likely overlook highly divergent genes including those that might have accumulated nonsense mutations following TE insertion. Tenuous examples of additional TE-disrupted genes include a putative family 3 glycoside hydrolase on scaffold 2 (coordinates 1,647,844–1,679,733). In other instances, TEs located adjacent to, but not within, gene models may influence gene expression. As a potential example, a solo *pcret1* LTR lies 1,012 bp upstream of the translational start of a putative cytochrome P450 gene (model 40563) on scaffold 5.

Interestingly, insertions seem biased toward gene families, members of which tend to cluster. Assuming genes within such families have redundant or overlapping functions, the impact of such insertions would be lessened. Further, in each case the corresponding allelic variant was uninterrupted and transcriptionally active. As a consequence, the negative affects of insertion would be further diminished by stable heterozygosity within the dikaryon. Such sheltering by heterozygosity might be expected to extend to single copy genes or even to essential genes. However, the sequenced strain, RP-78, is homokaryotic, and deleterious mutations would be negatively selected in the absence of some mitigating mechanism.

One such mechanism, well established in higher plants (Weil and Wessler 1990), *D. melanogaster* (Fridell et al. 1990) and *C. elegans* (Rushforth and Anderson 1996), involves splicing to remove TE sequences from mRNAs. Intron insertion points were identified for *pce5*, *pcret1*, and *pcret2/3*, prompting an examination of this possibility in *P. chrysosporium*. As in the case of certain *Fot1* insertion in *niaD* (Deschamps et al. 1999), we observed transcripts 5′ and/or 3′ to the insertion point (Figs. 3, 4, 5, lanes 1 and 2). In the case of *pce5*, the transcripts extended into the element (Fig. 3c, d). However, our RT-PCR analysis showed no splicing of the full-length elements (Figs. 3, 4, 5).

Conceivably, the truncated cDNAs might encode active protein, although this seems highly unlikely. In the case of *mco3A*, the cDNA lacks a Cys and related residues necessary for copper coordination (Canters and Gilardi 1993; Larrondo et al. 2003). Similarly, incomplete transcripts from *pce5*- and *pcret2/3*-interrupted sequences cannot give rise to active cytochrome P450s because the cDNAs lack either the active site Cys or the β1 domain conserved in all P450s (Graham and Peterson 1999). Finally, the previously described *pce1* insertion occurs at Ala135 within *lipI* (Gaskell et al. 1995), a position that divides residues (Arg43, His47, His176) essential for catalysis (Tien and Tu 1987). Thus, all five alleles are likely inactivated by the spontaneous insertions.

It remains unclear whether these loss-of-function mutations are generally targeted toward gene families. Future research may identify conditions favoring transposition. Ultimately, TEs might provide useful gene-tagging tools and thereby resolve fundamental questions about the role of gene multiplicity in *P. chrysosporium*.

Acknowledgments This work was supported by U.S. Department of Energy grant no. DE-FG02-87ER13712 and the Millenium Institute for Fundamental and Applied Biology and by grant 1030495 from FONDECYT, Chile. Nucleotide sequence data reported are available in GenBank under accession numbers DQ400694, DQ444269, AY532142, DQ444268, and DQ677350.

References

Cambareri EB, Aisner R, Carbon J (1998) Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. Mol Cell Biol 18:5465–5477

Canters GW, Gilardi G (1993) Engineering type 1 copper sites in proteins. FEBS Lett 325:39–48

Chomet P, Lisch D, Hardeman KJ, Chandler VL, Freeling M (1991) Identification of a regulatory transposon that controls the Mutator transposable element system in maize. Genetics 129:261–270

Cullen D, Kersten PJ (2004) Enzymology and molecular biology of lignin degradation. In: Brambl R, Marzulf GA (eds) The Mycota III biochemistry and molecular biology. Springer, Berlin Heidelberg New York, pp 249–273

Daboussi MJ, Capy P (2003) Transposable elements in filamentous fungi. Annu Rev Microbiol 57:275–299

DeMarco R et al (2004) Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. J Virol 78:2967–2978

Deschamps F, Langin T, Maurer P, Gerlinger C, Felenbok B, Daboussi MJ (1999) Specific expression of the *Fusarium* transposon *FotI* and effects on target gene transcription. Mol Microbiol 31:1373–1383

Diez J, Beguiristain T, Le Tacon F, Casacuberta JM, Tagu D (2003) Identification of Ty1-copia retrotransposons in three ectomycorrhizal basidiomycetes: evolutionary relationships and use as molecular markers. Curr Genet 43:34–44

Doddapaneni H, Chakraborty R, Yadav JS (2005) Genome-wide structural and evolutionary analysis of the P450 monooxygenase genes (P450ome) in the white rot fungus *Phanerochaete chrysosporium*: evidence for gene duplications and extensive gene clustering. BMC Genomics 6:92

Eriksson KE, Hamp SG (1978) Regulation of endo-1,4-β-glucanase production in *Sporotrichum pulverulentum*. Eur J Biochem 90:183–190

Fowler TJ, Mitton MF (2000) Scooter, a new active transposon in *Schizophyllum commune*, has disrupted two genes regulating signal transduction. Genetics 156:1585–1594

Fridell R, Pret A, Searles L (1990) A retrotransposon 412 insertion within an exon of the *Drosophila melanogaster* vermillion gene is spliced from the precursor RNA. Genes Dev 4:559–566

Fudal I, Bohnert HU, Tharreau D, Lebrun MH (2005) Transposition of MINE, a composite retrotransposon, in the avirulence gene ACE1 of the rice blast fungus *Magnaporthe grisea*. Fungal Genet Biol 42:761–772

Gaskell J, Vanden Wymelenberg A, Cullen D (1995) Structure, inheritance, and transcriptional effects of Pce1, an insertional element within *Phanerochaete chrysosporium* lignin peroxidase gene *lipI*. Proc Natl Acad Sci USA 92:7465–7469

Goodwin TJ, Poulter RT (2000) Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res 10:174–191

Goodwin TJ, Poulter RT (2004) A new group of tyrosine recombinase-encoding retrotransposons. Mol Biol Evol 21:746–759

Graham SE, Peterson JA (1999) How similar are P450s and what can their differences teach us? Arch Biochem Biophys 369:24–29

Hood ME (2005) Repetitive DNA in the automictic fungus *Microbotryum violaceum*. Genetica 124:1–10

Hua-Van A, Daviere JM, Kaper F, Langin T, Daboussi MJ (2000) Genome organization in *Fusarium oxysporum*: clusters of class II transposons. Curr Genet 37:339–347

Jordan IK, McDonald JF (1999) Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. Genetics 151:1341–1351

Kang S (2001) Organization and distribution pattern of MGLR-3, a novel retrotransposon in the rice blast fungus *Magnaporthe grisea*. Fungal Genet Biol 32:11–19

Kang S, Lebrun MH, Farrall L, Valent B (2001) Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. Mol Plant Microbe Interact 14:671–674

Kempken F, Kuck U (1998) Transposons in filamentous fungi-facts and perspectives. BioEssays 20:652–659

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 8:464–478

Larrondo L, Salas L, Melo F, Vicuna R, Cullen D (2003) A novel extracellular multicopper oxidase from *Phanerochaete chrysosporium* with ferroxidase activity. Appl Environ Microbiol 69:6257–6263

Larrondo L, Gonzalez B, Cullen D, Vicuna R (2004) Characterization of a multicopper oxidase gene cluster in *Phanerochaete chrysosporium* and evidence for altered splicing of the mco transcript. Microbiology 150:2775–2783

Martinez D et al (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP-78. Nat Biotechnol 22:695–700

Nishimura M, Hayashi N, Jwa NS, Lau GW, Hamer JE, Hasebe A (2000) Insertion of the LINE retrotransposon MGL causes a conidiophore pattern mutation in *Magnaporthe grisea*. Mol Plant Microbe Interact 13:892–894

Plant EP, Goodwin TJ, Poulter RT (2000) Tca5, a Ty5-like retrotransposon from *Candida albicans*. Yeast 16:1509–1518

Poulter RT, Goodwin TJ, Butler MI (2003) Vertebrate helentrons and other novel Helitrons. Gene 313:201–212

Rezsohazy R, van Luenen HG, Durbin RM, Plasterk RH (1997) Tc7, a Tc1-hitch hiking transposon in *Caenorhabditis elegans*. Nucleic Acids Res 25:4048–4054

Rushforth AM, Anderson P (1996) Splicing removes the *Caenorhabditis elegans* transposon Tc1 from most mutant pre-mRNAs. Mol Cell Biol 16:422–429

Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147:195–197

Sonnenberg AS, Baars JJ, Mikosch TS, Schaap PJ, Van Griensven LJ (1999) *Abr1*, a transposon-like element in the genome of the cultivated mushroom *Agaricus bisporus* (Lange) Imbach. Appl Environ Microbiol 65:3347–3353

Stewart P, Gaskell J, Cullen D (2000) A homokaryotic derivative of a *Phanerochaete chrysosporium* strain and its use in genomic analysis of repetitive elements. Appl Environ Microbiol 66:1629–1633

Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:2552–2556

Thon MR, Martin SL, Goff S, Wing RA, Dean RA (2004) BAC end sequences and a physical map reveal transposable element content and clustering patterns in the genome of *Magnaporthe grisea*. Fungal Genet Biol 41:657–666

Tien M, Tu C-PD (1987) Cloning and sequencing of a cDNA for a ligninase from *Phanerochaete chrysosporium*. Nature 326:520–523

Vanden Wymelenberg A et al (2002) Transcript analysis of genes encoding a family 61 endoglucanase and a putative membrane-anchored family 9 glycosyl hydrolase from *Phanerochaete chrysosporium*. Appl Environ Microbiol 68:5765–5768

Vanden Wymelenberg A et al (2005) The *Phanerochaete chrysosporium* secretome: database predictions and initial mass

spectrometry peptide identifications in cellulose-grown medium. J Biotechnol 118:17–34

Vanden Wymelenberg A et al (2006) Computational analysis of the *Phanerochaete chrysosporium* v2.0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveals complex mixtures of secreted proteins. Fungal Genet Biol 43:343–356

Weil C, Wessler S (1990) The effects of plant transposable element insertion on transcription initiation and RNA processing. Annu Rev Plant Physiol Plant Mol Biol 41:527–552

Wostemeyer J, Kreibich A (2002) Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. Curr Genet 41:189–198