MOLECULAR AND CELLULAR BIOLOGY, June 1989, p. 2743–2747
0270-7306/89/062743-05$02.00/0

Vol. 9, No. 6

Molecular Cloning and Sequences of Lignin Peroxidase Genes of *Phanerochaete chrysosporium*

HEIDI SCHALCH,[1] JILL GASKELL,[2] TIMOTHY L. SMITH,[3] AND DANIEL CULLEN[2]*

Departments of Bacteriology[1] and Molecular Biology,[3] University of Wisconsin, Madison, Wisconsin 53706, and Institute for Microbial and Biochemical Technology, Forest Products Laboratory, Forest Service, U.S. Department of Agriculture, Madison, Wisconsin 53705-2398[2]

Received 28 November 1988/Accepted 10 February 1989

The genomic clones encoding lignin peroxidase isozyme H8 and two closely related genes were isolated from *Phanerochaete chrysosporium* BKM-1767, and their nucleotide sequences were determined. The positions and approximate lengths of introns were found to be highly conserved in all three clones. Analysis of homokaryotic derivatives indicated that the three clones are not alleles of the same gene(s).

Lignin depolymerization is catalyzed by extracellular peroxidases of white rot basidiomycetes such as *Phanerochaete chrysosporium* (9). In submerged culture, multiple lignin peroxidase (LiP) isozymes are present (10, 11), and production is derepressed under carbon, nitrogen, or sulfur limitation (8, 11). The roles of individual isozymes in lignin degradation and the underlying genetic regulation are poorly understood.

The nucleotide sequences of the cDNA encoding the predominant isozyme H8 (20), the genomic clone of H8 (17), and two additional cDNA clones (5) have been reported. Deduced amino acid sequences are 70 to 80% identical among all three clones. These clones were all derived from *P. chrysosporium* BKM-1767 (ATCC 24725), which is, quite probably, heterokaryotic (1, 19). Hence, the possibility of allelic relationships between these three clones cannot be formally excluded.

We report here remarkable structural homology among three LiP genomic clones. Analysis of homokaryotic derivatives demonstrated that the three are not alleles of the same gene.

Isolation and identification of LiP genomic clones. Standard Southern and colony hybridization techniques were used throughout (4, 12, 18). A 700-base-pair (bp) *Sph*I genomic fragment isolated from H8 (Fig. 1) was used to probe a 4,500-member *P. chrysosporium* library in cosmid pKBY2 (22). Given an approximate haploid genome size of 4.4×10^3 kilobases (14), there was a >0.97 probability of any gene being present in the cosmid library, although this figure would be reduced to some extent by heterokaryosis. Initially, low to moderate stringencies were used (30% formamide, 37°C), and 29 strongly hybridizing clones were identified. Subsequent Southern analyses of positive clones entailed digestions with *Eco*RV, *Xho*I, *Taq*I, and *Sau*3A, size fractionation on 1% agarose gels, blotting to Nytran (Schleicher & Schuell, Inc., Keene, N.H.), and probing with a 270-bp *Eco*RV fragment from genomic H8 (Fig. 1) under various stringencies.

Families of closely related clones were evident by several distinct patterns on Southern blot autoradiograms. Only one group, subsequently shown to contain the gene encoding H8 (17), hybridized at high stringencies (50% formamide, 60°C). Six clones fell into this group. Under reduced hybridization stringencies (50% formamide, 37°C), two additional patterns were recognized and designated 0282-like (nine cosmids) and V4-like (three cosmids). Seven cosmids gave no discernible signal, and four had unique patterns. A representative cosmid from each group was subcloned and partially sequenced. The four unique clones had no clear sequence homology with previously published cDNA sequences (5, 20). These cosmids probably represent complex rearrangements or separate genes. They were not characterized further. It is probable that additional related clones could be identified by using probes other than the 700-bp *Sph*I fragment from genomic H8. Furthermore, cosmid groups identified by Southern hybridization patterns, i.e., H8-, 0282-, and V4-like, may contain closely related genes or alleles. The exact number of LiP-specific alleles remains to be established.

Nucleotide sequence, predicted amino acid sequence, and comparative analysis of LiP clones. Cosmids H8, 0282, and V4 were subcloned and sequenced by the dideoxy method (15). The three clones showed regions of intense nucleotide homology (Fig. 1 and 2). Intron positions in 0282 and V4 were deduced by alignment with the genomic H8 sequence (17) and by the presence of highly conserved 5′ and 3′ splice sites (2). Intron position and length were conserved in all three genes, although sequence similarity within introns was

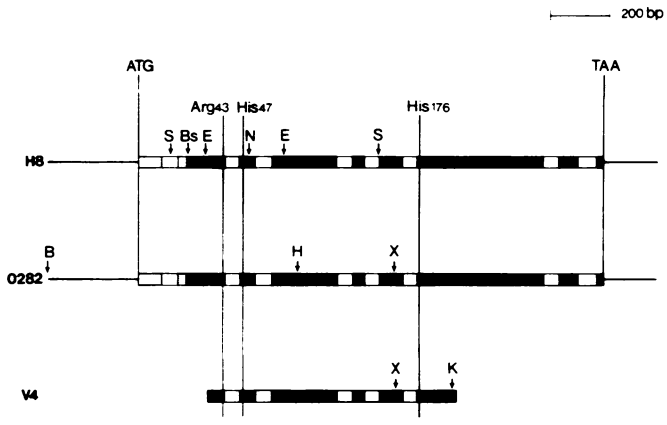

FIG. 1. Schematic representation of genomic clones showing alignments of signal sequences (□), introns (□), and coding regions for mature polypeptide (■). Positions of essential amino acid residues are shown by cDNA coordinates (20). For cosmid V4, only sequences surrounding essential amino acid residues were determined. Restriction site abbreviations: S, *Sph*I; Bs, *Bss*HII; E, *Eco*RV; N, *Nco*I; H, *Hin*dIII; X, *Xho*I; K, *Kpn*I.

* Corresponding author.

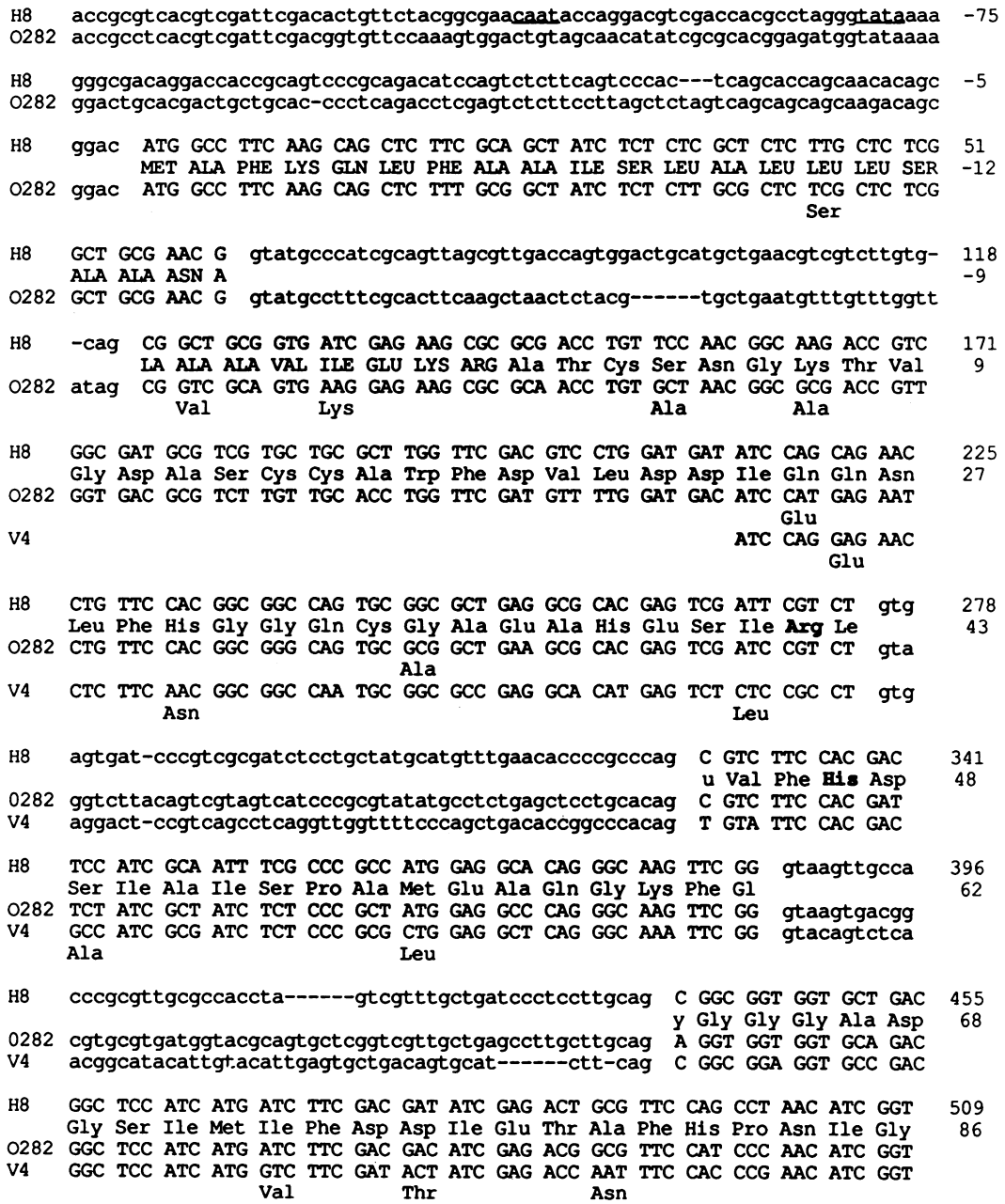

FIG. 2. Alignment of the genomic clone encoding H8 (EMBL accession number X06689) and two related clones, 0282 and V4 (partial sequence). Putative CAAT (−110) and TATA (−81) are underlined. Essential amino acids at positions 43, 47, and 176 are in boldface. Amino acids −1 to −28 (capitalized) represent the presumed signal and propeptide sequence. Gaps are introduced to maximize nucleotide alignments by the method of Wilbur and Lipman (21). All of H8 is translated, whereas 0282 and V4 translations are shown only where different from H8.

```
H8    CTC GAC GAG ATC GTC AAG CTC CAG AAG CCA TTC GTT CAG AAG CAC GGT GTC ACC   563
      Leu Asp Glu Ile Val Lys Leu Gln Lys Pro Phe Val Gln Lys His Gly Val Thr   104
O282  CTC GAC GAG ATT GTC AAG CTT CAG AAG CCG TTC GTC CAG AAG CAG AAC GTC ACC
                                                              Asn
V4    CTC GAC GAA ATC GTC CGC CTG CAG AAA CCG TTC GTT CAG AAG CAC GGT GTT ACT
                          Arg

H8    CCT GGT GAC TTC ATC GCC TTC GCT GGT GCT GTC GCG CTC AGC AAC TGC CCT GGT   617
      Pro Gly Asp Phe Ile Ala Phe Ala Gly Ala Val Ala Leu Ser Asn Cys Pro Gly   122
O282  CCT GGC GAT TTT ATT GCC TTC GCC GGG GCT GTC GCA CTC AGC AAC TGC CCT GGT
V4    CCT GGC GAC TTC ATC GCA TTC GCT GGT GCG GTA GGA CTG AGC AAC TGC CCG GGT
                                                  Gly

H8    GCC CCG CAG ATG AAC TTC TTC ACT GGT CGT GCA CCT G  gtatacctgcaaaactcgct   674
      Ala Pro Gln Met Asn Phe Phe Thr Gly Arg Ala Pro A                         134
O282  GCC CCT CAA ATG AAC TTC TTC ACC GGT CGC GCT CCT G  gtatgtgcttcattccattt
V4    GCG CCA CAA ATG AAC TTC TTC CTC GGT CGC --- CCA G  ctcgtgagtcgtttgtgatt
                                Leu

H8    ttgcggatggtacgaactaactgtctgctt--tag  CT ACC CAG CCC GCT CCT GAT GGC CTT   733
                                           la Thr Gln Pro Ala Pro Asp Gly Leu   143
O282  caccccaaggattcccaatgacattgggaccgtag  CT ACC CAG CCA GCT CCA GAT GGC CTT
V4    gaxatatgggcgcacctcatgtacc--gaatctag  CG ACC AAG GCC GCA CCG GAC GGT CTT
                                           Lys Ala

H8    GTC CCC GAG CCC TTC C  gtaagtggtcttatccaagcagttaggtgcggtctatactgactagca   797
      Val Pro Glu Pro Phe H                                                     148
O282  GTT CCC GAG CCG TTC C  gtaagtgcatacttctagacagacgtcaccgtactttcgctcacatct
V4    GTC ACA GAA CCC TTC C  gtactgaccgaaatctgcggatgcgaagttgctcaccagaagcataca

H8    tg---cag  AC ACT GTC GAC CAA ATC ATC AAC CGT GTC AAC GAC GCA GGC GAC TTC  849
                is Thr Val Asp Gln Ile Ile Asn Arg Val Asn Asp Ala Gly Glu Phe  164
0282  gcacccag  AC ACT GTC GAT CAA ATC ATC AGC CGT GTC AAT GAT GCC GGA CAG TTC
                                           Ser                          Gln
V4    ga---cag  AC TCC GTC GAT CAA ATC CTG GCT CGG GTG GCC GAC GCT GGC GAG TTT
                   Ser                 Leu Ala         Ala

H8    GAT GAG CTC GAC CTT GTC TGG ATG CTC TCC GC  gtaagtcactcactgttgacttcgacac   909
      Asp Glu Leu Glu Leu Val Trp Met Leu Ser Al                                174
0282  GAT GAG CTC GAG CTC GTA TGG ATG CTT TCG GC  gtaagtctcgagattgtgtgtcagttca
V4    GAC GAA CTC GAG ACT GTC TGG CTG CTC TCG GC  gtaagcttctccgcgttatgtgcgtatg
                     Thr         Leu

H8    tcccttcctgagacctcga------cag  G CAC TCC GTC GCA GCG GTG AAC GAC GTC GAC   962
                                    a His Ser Val Ala Ala Val Asn Asp Val Asn   185
0282  tcctatctgactgcctgg-------cag  G CAC TCC GTC GCG GCA GTC AAC GAC GTC GAT
V4    ttggtgtatttaccacccatgtgcgtag  G CAT TCT GTC GCT GCC GCC AAC GAC GTC GAC
                                                        Ala

H8    CCG ACC GTC CAG GGT CTG CCC TTT GAC TCG ACC CCC GGA ATC TTC GAC TCC CAG   1016
      Pro Thr Val Gln Gly Leu Pro Phe Asp Ser Thr Pro Gly Ile Phe Asp Ser Gln    203
0282  CCG ACT GTC CAA GGT CTG CCC TTC GAC TCG ACG CCC GGA ATC TTC GAC TCC CAG

H8    TTC TTC GTC GAG ACT CAG CTT CGT GGT ACC GCC TTC CCC GGC TCT GGT GGC AAC   1070
      Phe Phe Val Glu Thr Gln Leu Arg Gly Thr Ala Phe Pro Gly Ser Gly Gly Asn    221
0282  TTC TTC GTC GAG ACT CAG CTC CGT GGC ACC GCG TTT CCC GGG TCT GGC GGC AAC

H8    CAA GGC GAG GTC GAG TCG CCG CTC CCT GGC GAA ATT CGC ATC CAG TCC GAC CAC   1124
      Gln Gly Glu Val Glu Ser Pro Leu Pro Gly Glu Ile Arg Ile Gln Ser Asp His    239
0282  CAG GGT GAG GTC GAG TCG CCT CTT CCC GGC GAG TTC CGC ATC CAG TCG GAC CAC
                                                  Phe
```

FIG. 2—*Continued*

```
H8   ACT ATC GCC CGC GAC TCG CGC ACG GCG TGT GAA TGG CAG TCC TTC GTC AAC AAC   1178
     Thr Ile Ala Arg Asp Ser Arg Thr Ala Cys Glu Trp Gln Ser Phe Val Asn Asn    257
0282 ACC ATC GCC CGC GAC TCG GCC ACG GCG TGT GAA TGG CAG TCC TTT GTC AAC AAC
                           Ala

H8   CAG TCC AAG CTC GTC GAT GAC TTC CAG TTC ATC TTC CTC GCC CTC ACC CAG CTC   1232
     Gln Ser Lys Leu Val Asp Asp Phe Gln Phe Ile Phe Leu Ala Leu Thr Gln Leu    275
0282 CAG TCG AAG CTC GTC GAC GAC TTC CAG TTC ATC TTC CTC GCC CTC ACT CAG CTC

H8   GGC CAG GAC CCG AAC GCG ATG ACC GAC TGC TCG GAT GTT ATC CCG CAG TCC AAG   1286
     Gly Gln Asp Pro Asn Ala Met Thr Asp Cys Ser Asp Val Ile Pro Gln Ser Lys    293
0282 GGC CAG GAC CCG AAT GCG ATG ACC GAC TGC TCG GAT GTC ATC CCG CAA TCG AAG

H8   CCC ATC CCT GGC AAC CTC CCA TTC TCG TTC TTC CCC GCT GGC AAG ACC ATC AAG   1340
     Pro Ile Pro Gly Asn Leu Pro Phe Ser Phe Phe Pro Ala Gly Lys Thr Ile Lys    311
0282 CCC ATC CCC GGC AAC CTT CCG TTC TCG TTC TTC CCC GCA GGC AAG ACC ATA AAG

H8   GAC GTT GAG CAG GCG  gtgcgtatttcaccccaccatgcagtagagtggctgctgaacatcgcatg   1405
     Asp Val Glu Gln Ala                                                       316
0282 GAT GTT GAG CAG GCG  gtgcgtgatctgcatattcgtaggcgcgatgatcctgatctttgccttg-

H8   acag  TGT GCG GAG ACC CCC TTC CCG ACT CTC ACC ACT CTC CCG GGC CCC GAG     1457
           Cys Ala Glu Thr Pro Phe Pro Thr Leu Thr Thr Leu Pro Gly Pro Glu      332
0282 -cag  TGC GCG GAG ACC CCA TTC CCC ACA CTC ACG ACC CTC CCT GGA CCC GAG

H8   ACG TCC GTC CAG CGC AT  gtgagtacaatccatgagatctttcaggaaatgcaatctgggctgac   1521
     Thr Ser Val Gln Arg Il                                                    337
0282 ACC TCC GTC CAG CGC AT  gtgagtacacaatctaggttcagcccagaagcacgcactgacagcct

H8   atgctccttctccag  C CCT CCG CCT CCG GGT GCT TAA  atgatgccatacagaatactcct   1581
                      e Pro Pro Pro Pro Gly Ala END                            344
0282 ----------ttag  T CCG CCG CCC CCA GGT GCT TAA  acaaaaacaagtcgagaacgaca

H8   caaaccg-actgtaacggtggccggctaactc                                         1611
0282 gtatcttcactgtatcggtagctgatccagtc
```

FIG. 2—*Continued*

low relative to that in coding regions. Another genomic LiP recently cloned by Sims et al. (16) from *P. chrysosporium* ME446 shows the same basic structural features, i.e., eight introns and high sequence homology to H8 cDNA.

Amino acid and nucleotide similarities, expressed as percent identity within aligned regions, were calculated (Table 1). Predicted amino acid sequences were highly homologous, particularly between H8 and 0282, which were 96.5% identical. There was no apparent regional clustering of mismatched amino acids. The majority of mismatched bases occurred at the third position of codons and did not affect the amino acid sequence. As with previously studied LiP cDNA clones (5, 20), a high degree of codon bias was present. Arg-43, His-47, and His-176 (Fig. 1), believed essential for peroxidase activity (20), were present in all three clones. The amino acid sequences of H8 and 0282 were nearly identical in these regions, but V4 encoded single amino acid differences within two residues of Arg-43 (Ile-42 → Leu) and His-47 (Ser49 → Ala) (Fig. 2). In all three clones, the codons for the essential amino acids lay immediately adjacent to intron 2 or 6 (Fig. 1 and 2). The location of intron 8, adjacent to the last, proline-rich exon, and the position of intron 1, which splits a putative signal sequence-propeptide junction (13), are consistent with introns having played a role in shuffling functional domains (3). The putative propeptide (Ala-Ala-Val-Ile-Glu-Lys-Arg) follows the consensus splice site Ala-X-Ala (13) and is similar to the *Aspergillus niger* glucoamylase propeptide (Ser-Val-Ile-Ser-Lys-Arg) (7).

TABLE 1. Nucleotide and amino acid identities among *P. chrysosporium* lignin peroxidase genes

		Amino acids[a]				
		H8	0282	V4	CLG4[b]	CLG5[b]
Nucleotides[c]	H8		96.5	85.8	72.1	82.5
	0282	88.1		84.6	70.5	81.2
	V4	80.0	76.3		85.2	83.3
	CLG4	73.5	70.8	77.6		66.8
	CLG5	80.7	78.4	79.6	71.8	

[a] Percentage of matched nucleotides within exons as determined by Wilbur and Lipman (21), using K-tuple of 3, window size of 20, and gap penalty of 3.
[b] Data from reference 5.
[c] Percentage of matched amino acids within overlapping alignments (21).

Absence of allelic relationships among LiP clones. The potentially heterokaryotic nature of *P. chrysosporium* and the extremely high sequence homology, particularly between 0282 and H8, suggest the possibility of allelic relationships among these clones. This possibility was formally excluded by probing single-basidiospore derivatives with clone-specific fragments. Alleles of the same gene will segregate independently in homokaryotic progeny such that, if allelic, 0282, H8, and V4 would not be detected in the same single-basidiospore cultures.

Basidiospores were harvested from xylose-containing medium (6). For Southern analysis, DNA was purified from five single-basidiospore cultures, digested with restriction endonuclease, size fractionated on 0.6% agarose, blotted to Nytran, and probed with nick-translated fragments under stringent conditions (50% formamide, 60°C) (Fig. 3).

The pattern expected for individual genes, as opposed to segregating alleles, was observed in all cases. In one basidiospore culture, designated SB-9 (Fig. 3B, lane 9), the

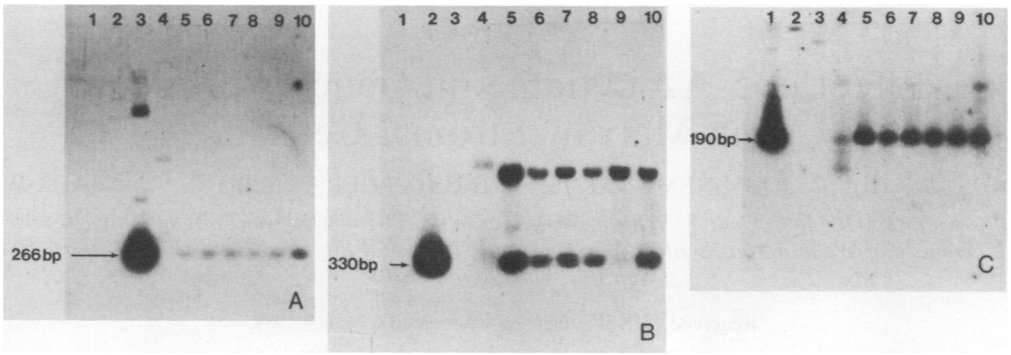

FIG. 3. Southern hybridizations of genomic DNA from single-basidiospore derivatives of *P. chrysosporium* BKM-1767. (A) *Eco*RV-digested DNA probed with a 266-bp *Eco*RV fragment of H8; (B) *Hin*dIII-*Xho*I-digested DNA probed with a 330-bp *Hin*dII-*Xho*I fragment of 0282; (C) *Xho*I-*Kpn*I-digested DNA probed with a 190-bp *Xho*I-*Kpn*I fragment of V4. Lanes: 1, 2, and 3, plasmid digests containing the V4, 0282, and H8 genes (to demonstrate specificity of probes); 4, DNA size marker (lambda *Hin*dIII plus pBR322 *Hin*fI); 5, BKM-1767 parental DNA (control); 6, single-basidiospore derivative 3 (SB-3); 7, SB-5; 8, SB-7; 9, SB-9; 10, SB-13.

expected 330-bp *Xho*I-*Hin*dIII band was lacking, although another, higher-molecular-weight band was visible in all single-basidiospore cultures (lanes 6 to 10) and in the parental culture (lane 5). After digestion with *Hin*dIII-*Bam*HI and probing with a 900-bp *Hin*dIII-*Bam*HI 0282 fragment (Fig. 1), a similar pattern was observed; i.e., the 0282-specific band was absent from SB-9 (data not shown). The deletion of 0282 in SB-9 may indicate aneuploidy. The 0282-like gene represented by the higher-molecular-weight band was not detected in the cosmid library.

These data, together with previously reported cDNA sequences, demonstrate that the LiPs of *P. chrysopsorium* are encoded by a large and complex gene family. The organization, transcriptional regulation, and isozyme and gene specificities of these genes need to be established.

We thank Kent Kirk and Sarah Covert for thoughtful comments and R. Berka for recognizing the putative propeptide-signal peptide cleavage site.

This work was supported by Department of Energy grant DE-FG02-87ER13712 and by the Biopulping Consortium of the University of Wisconsin Biotechnology Center, the Forest Products Laboratory, and member companies.

LITERATURE CITED

1. **Alic, M., C. Letzring, and M. Gold.** 1987. Mating system and basidospore formation in the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. Trans. Br. Mycol. Soc. **3**:285–294.
2. **Ballance, D. J.** 1986. Sequences important for gene expression in filamentous fungi. Yeast **2**:229–236.
3. **Blake, C. C. F.** 1978. Do genes-in-pieces imply protein-in-pieces? Nature (London) **273**:267.
4. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
5. **de Boer, H. A., Y. Z. Zhang, C. Collins, and C. A. Reddy.** 1987. Analysis and nucleotide sequence of two ligninase cDNAs from a white rot filamentous fungus, *Phanerochaete chrysosporium*. Gene **58**:93–102.
6. **Gold, M. H., and T. M. Cheng.** 1979. Conditions for fruit body formation in the white-rot basidiomycete *Phanerochaete chrysosporium*. Arch. Microbiol. **121**:37–41.
7. **Innis, M. A., M. J. Holland, P. C. McCabe, G. E. Cole, V. P. Wittman, R. Tal, K. W. K. Watt, D. H. Gelfand, J. P. Holland, and J. H. Meade.** 1985. Expression, glycosylation, and secretion of an *Aspergillus* glucoamylase by *Saccharomyces cerevisiae*. Science **228**:21–26.
8. **Jeffries, T. W., S. Choi, and T. K. Kirk.** 1981. Nutrient regulation of lignin degradation by *Phanerochaete chrysosporium*. Appl. Environ. Microbiol. **42**:290–296.
9. **Kirk, T. K., and R. L. Farrell.** 1987. Enzymatic combustion: the microbial degradation of lignin. Annu. Rev. Microbiol. **41**: 465–505.
10. **Kirk, T. K., R. L. Farrell, S. Croan, and M. Tien.** 1986. Production of multiple ligninases by *Phanerochaete chrysosporium*: effect of selected growth conditions and use of a mutant strain. Enzyme Microb. Technol. **8**:27–32.
11. **Leisola, M., D. Kozulic, D. Meussdierffer, and A. Fiechter.** 1987. Homology among multiple extracellular peroxidases from *Phanerochaete chrysosporium*. J. Biol. Chem. **262**:419–424.
12. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
14. **Raeder, U., W. Thompson, and P. Broda.** 1987. Establishing molecular genetics for *Phanerochaete chrysosporium*. Phil. Trans. R. Soc. London **321**:475–483.
15. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
16. **Sims, P. A. Brown, C. James, U. Raeder, A. Schrank, and P. Broda.** 1988. Molecular genetics of lignocellulose degradation in *Phanerochaete chrysosporium*, p. 365–373. *In* J. P. Aubert, P. Beguin, and J. Millet (ed.), The biochemistry and genetics of cellulose degradation. Academic Press, Inc., San Diego, Calif.
17. **Smith, T. L., H. Schalch, J. Gaskell, S. Covert, and D. Cullen.** 1988. Nucleotide sequence of a ligninase gene from *Phanerochaete chrysosporium*. Nucleic Acids Res. **16**:1219.
18. **Taub, F., and E. B. Thompson.** 1982. An improved method for preparing large arrays of bacterial colonies containing plasmids for hybridization: in situ purification and stable binding of DNA on paper filters. Anal. Biochem. **126**:222–230.
19. **Thompson, W., and P. Broda.** 1987. Mating behaviour in an isolate of *Phanerochaete chrysosporium*. Trans. Br. Mycol. Soc. **3**:285–294.
20. **Tien, M., and C. P. D. Tu.** 1987. Cloning and sequencing of a cDNA for ligninase from *Phanerochaete*. Nature (London) **326**:520–523.
21. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein databanks (global homology-optimal alignment). Proc. Natl. Acad. Sci. USA **80**:726–730.
22. **Yelton, M. M., W. E. Timberlake, and C. A. M. J. J. van den Hondel.** 1985. A cosmid for selecting genes by complementation in *Aspergillus nidulans*: selection of the developmentally regulated *yA* locus. Proc. Natl. Acad. Sci. USA **82**:834–838.