Theor Appl Genet (2003) 106:676–682
DOI 10.1007/s00122-002-1116-9

**P. C. Bundock · J. T. Christopher · P. Eggler**
**G. Ablett · R. J. Henry · T. A. Holton**

# Single nucleotide polymorphisms in cytochrome P450 genes from barley

Received: 3 December 2001 / Accepted: 11 July 2002 / Published online: 22 October 2002


**Abstract** Plant cytochrome P450s are known to be essential in a number of economically important pathways of plant metabolism but there are also many P450s of unknown function accumulating in expressed sequence tag (EST) and genomic databases. To detect trait associations that could assist in the assignment of gene function and provide markers for breeders selecting for commercially important traits, detection of polymorphisms in identified P450 genes is desirable. Polymorphisms in EST sequences provide so-called perfect markers for the associated genes. The International Triticeae EST Cooperative data base of 24,344 ESTs was searched for sequences exhibiting homology to P450 genes representing the nine known clans of plant P450s. Seventy five P450 ESTs were identified of which 24 had best matches in Genbank to P450 genes of known function and 51 to P450s of unknown function. Sequence information from PCR products amplified from the genomic template DNA of 11 barley varieties was obtained using primers designed from six barley P450 ESTs and one durum wheat P450 EST. Single nucleotide polymorphisms (SNPs) between barley varieties were identified using five of the seven PCR products. A maximum of five SNPs and three haplotypes among the 11 barley lines were detected in products from any one primer pair. SNPs in three PCR products led to changes between barley varieties in at least one restriction site enabling genotyping and mapping without the expense of a specialist SNP detection system. The overall frequency of SNPs across the 11 barley varieties was 1 every 131 bases.

**Keywords** Barley cytochrome · P450 · SNP marker

Communicated by P. Langridge

P.C. Bundock · J.T. Christopher · P. Eggler · G. Ablett
R.J. Henry (✉) · T.A. Holton
Cooperative Research Centre for Molecular Plant Breeding,
Centre for Plant Conservation Genetics,
Southern Cross University, PO Box 157,
Lismore, NSW 2480, Australia
e-mail: rhenry@scu.edu.au
Tel.: +61-2-6620-3010, Fax: +61-2-6622-2080

## Introduction

Cytochrome P450-dependent mono-oxygenases (P450s) are haem-containing, membrane-bound enzymes that are pivotal in many metabolic processes in bacteria, fungi, animals and plants (Bolwell et al. 1994; Schuler 1996; Archakov et al. 2001). A large number of different P450 genes occur in the genomes of individual plant and animal species commensurate with the wide range of metabolic functions they perform. Many of the best-characterised animal P450s are inducible enzymes, involved in detoxification, and have relatively broad substrate specificity (Bolwell et al. 1994; Feynereisen 1999; Paquette et al. 2000). In contrast, many of the best characterised plant P450s are involved in synthetic pathways, are highly substrate specific, not inducible and expressed at low levels (Bolwell et al. 1994; Schuler et al. 1996). It has been estimated that the number of P450s is lower in animal genomes (80 to 100 genes; Nelson 1999; Paquette et al. 2000) than in plant genomes (300 to 350 genes; Nelson 1999; Paquette et al. 2000).

Plant P450s are involved in a wide range of metabolic functions that may be of considerable commercial importance, especially in crop species. These include synthetic pathways leading to phenylpropanoids, alkaloids, terpenes, lipids, cyanogenic glycosides and glucosinolates (Bak et al. 2001, 2000; also reviewed in Bolwell et al. 1994; Schuler 1996; Chapple 1998). Many of these products influence important properties of a plant such as flavour and colour. They are also important for herbivore and pathogen resistance. Plant P450s are also essential in the synthetic pathways of the plant growth regulators, the giberellins, brassinosteroids and jasmonic acid.

Specific examples of P450 genes influencing traits of potential interest to plant breeders include defective mutants of the the genes encoding CYP88 causing a lesion in gibberellin biosynthesis in the *dwarf3* maize mutant and CYP85 in the tomato *dwarf* mutant thought to be involved in brassinosteriod synthesis (Winkler and Helentjaris 1995; Bishop et al. 1996). Molecular biologists with an interest in transforming plants are also in-

terested in a number of P450 genes, controlling such characters as herbicide resistance, insecticide resistance and flower colour (Holton et al.1993; Brugliera et al. 1999; Ohkawa et al. 1999; Werck-Reinhart et al. 2000; Wang et al. 2001).

Despite all this interest, advances in identifying the function of plant P450s has been slow (Winkler et al. 1998), but with recent advances in genomics the discovery of plant P450 genes has been rapid. Currently 273 *Arabidopsis* P450 genes have been identified as a result of EST and genome-sequencing efforts (http://www.biobase.dk/p450/chrmaps/chrmap n.jpg); however, the function of most of these is unknown. Given the diverse range of important functions ascribed to the small proportion of P450s of known function it appears likely that a significant number of the other P450s also perform important functions. This suggests that markers for plant P450 genes may be correlated with economically important traits. Mapped P450 gene markers could be used for trait association to ascribe function to more plant P450s. It would, therefore, be desirable to have markers for as many P450 genes as possible.

Plant P450 genes are often duplicated and found in large clusters of up to 14 genes (Nelson 1999). Genes encoding P450s involved in the synthesis of important defence compounds are clustered on chromosome 4 in maize (Frey et al. 1995, 1997). In order to distinguish the individual effects of P450s in such clusters perfect markers will be of particular importance.

Markers for plant P450 genes may also provide an additional resource for studying genetic diversity between barley varieties and for varietal identification.

We set out to determine whether it is possible to identify useful PCR-based molecular markers to allow genetic mapping of P450 genes in barley. The potential to use EST data to develop perfectly linked simple sequence repeat (microsatellite; SSR) or single nucleotide polymorphism (SNP) markers for P450 genes in barley was investigated. Here we describe an approach for identifying SNPs in barley P450 genes that can provide polymorphic markers between barley varieties.

## Materials and methods

### Identification of ESTs with sequence homology to P450 genes

A group of plant P450 genes was chosen so as to include a full-length sequence representing at least one P450 from each of the nine clans of cytochrome P450 genes found in plants using cereal or grass genes where possible (http://drnelson.utmem.edu/biobloD.html; Table 1). The International Triticeae EST Cooperative database (ITEC: http://wheat.pw.usda.gov/genome/) of 24,344 EST sequences was searched using representatives from each of the nine clans using the program TFASTXY (Pearson and Lipman 1988) to find the best matches. Subsequently the best matching sequences in the Genbank database for each candidate P450 EST were identified using NCBI BLAST-X + BEAUTY (Worley et al. 1995, 1998). P450 status was confirmed where candidate genes had best matches to P450 sequences in Genbank. Overlapping EST sequences were identified using Assembly LIGN 1.0.9c (Oxford Molecular, Campbell, Calif.).

### Search for SSRs

The ITEC EST sequence database was screened for all possible dinucleotide ($n \geq 6$) and trinucleotide ($n \geq 5$) SSRs, using the program FASTA (Pearson and Lipman 1988) as previously described (Holton et al. 2000). The ESTs containing SSRs were then checked against the P450 ESTs identified as above.

### Identification of SNPs

Oligonucleotide primers were designed to amplify segments from 27 of the P450 ESTs using the program MacVector (Oxford Molecular, Campbell, Calif.). Primers were designed to target the 3′ untranslated region of the genes where possible. Primers were tested for amplification using genomic barley template DNA. Those that amplified clear single bands were used to sequence genomic template DNA from 11 barley varieties used in the Australian National Barley Molecular Marker program; Alexis, Arapiles, Chebec, Clipper, Franklin, Galleon, Halcyon, Harrington, Haruna Nijo, Sahara and Sloop. Since EST-derived PCR primers from wheat and barley are often transferable (Holton et al. 2000), the primers were also tested for amplification using wheat genomic template DNA. Eleven wheat varieties were targeted including ten varieties used in the Australian National Wheat Molecular Marker program; Cranbrook, Halberd, Egret, Sunstar, Sunco, Tasman, Katepwa, CD87, Cascades and Cadoux plus the international reference variety Chinese Spring. Primers that successfully amplified a single PCR product of the expected size in barley were then used as forward and reverse sequencing primers to sequence PCR products amplified from the 11 barley varieties. Single nucleotide polymorphisms (SNPs) were detected from sequencing chromatograms using Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, Mich.). An SNP in PCR products from primers designed using barley EST ISC004.A02 between barley varieties Halcyon and

**Table 1** Plant P450 genes used to identify P450 ESTs in the International Triticeae EST Consortium (ITEC) database. P450 Clan, P450 gene, species of origin, accession and function where known

| Clan | P450 | Species | Accession | Function |
|---|---|---|---|---|
| CYP51 | CYP51 | *Triticum aestivum* | Y09291 | Obtusifoliol 14α-demethylase |
| CYP71 | CYP98A1 | *Sorghum bicolor* | AF029856 | Dhurrin synthesis |
| | CYP75A1 | *Petunia hybrida* | Z22545 | Flavonoid-3′-5′-hydroxylase |
| CYP72 | CYP72A1 | *Catharanthus roseus* | L19074 | Unknown |
| CYP74 | CYP74A3 | *Hordeum vulgare* | AJ250864 | Allene oxide synthase |
| CYP85 | CYP85A1 | *Lycopersicon esculentum* | U54770 | Unknown (dwarf protein) |
| CYP86 | CYP86A1 | *Arabidopsis thaliana* | X90458 | Unknown |
| CYP97 | CYP97A1 | *A. thaliana* | CAB64216 | Unknown |
| CYP710 | CYP710A1 | *A. thaliana* | AAC26691.1 | Unknown |
| CYP711 | CYP711A1 | *A. thaliana* | AC004484 | Unknown |

**Table 2** ESTs from the ITEC (International Triticeae EST Coop.) database identified as belonging to the cytochrome P450 gene family. The ESTs are grouped according to the function of the best matching sequence from Genbank

| ITEC EST | | Best match in Genbank | | | | |
|---|---|---|---|---|---|---|
| EST name | EST origin* | Species | Function | Prob. | aa Identity | CYP no. |
| 0253_B08 | W | *Hordeum vulgare* | Allene oxide synthase | 5.70E-29 | 60/66 (90%) | 74A |
| 0302_d03[1] | W | *H. vulgare* | Allene oxide synthase | 1.30E-24 | 58/59 (98%) | |
| 0302_d05[1] | W | *H. vulgare* | Allene oxide synthase | 1.30E-24 | 58/59 (98%) | |
| HWM002.E11[2] | B | *H. vulgare* | Allene oxide synthase | 3.10E-38 | 83/85 (97%) | 74A |
| HWM015cB.03r[2] | B | *H. vulgare* | Allene oxide synthase | 9.00E-39 | 83/83 (100%) | 74A |
| AWB009.E11 | W | *H. vulgare* | Allene oxide synthase | 1.20E-41 | 86/93 (92%) | 74A |
| MCG001.A05 | B | *Catharanthus roseus* | Cinnamate 4-hydroxylase | 8.60E-50 | 101/134 (75%) | 73 |
| MCG016.B10 | B | *Populus* | Cinnamate 4-hydroxylase | 1.70E-93 | 175/243 (72%) | 73 |
| JJL003.C01[a] | B | *Glycine max* | Cinnamate 4-hydroxylase | 5.60E-58 | 77/100 (77%) | 73 |
| MWL020.E10[a] | W | *Populus kitakamiensis* | Cinnamate 4-hydroxylase | 6.30E-32 | 59/98 (60%) | 73 |
| MWL036.F02[a] | W | *Citrus sinensis* | Cinnamate 4-hydroxylase | 1.70E-51 | 101/113 (89%) | 73 |
| WWS01.E6[b] | W | *Mesembryanth. crystallinum* | Cinnamate 4-hydroxylase | 1.40E-24 | 57/80 (71%) | 73 |
| WWS017.C5[b] | W | *C. sinensis* | Cinnamate 4-hydroxylase | 2.60E-78 | 134/174 (77%) | 73 |
| TAS005.B11 | W | *Sorghum bicolor* | Dhurrin biosynthesis | 3.20E-30 | 74/118 (62%) | 71E1 |
| PSR7243 | W | *Lycopersicon esculentum* | Dwarf protein | 2.20E-35 | 74/124 (59%) | 85 |
| SCU011.B10[3] | W | *Zea mays* | Gibberellin biosynthesis | 8.20E-77 | 144/206 (69%) | 88A1 |
| SCU011.C10[3] | W | *Oryza sativa* | Gibberellin biosynthesis | 4.80E-45 | 81/116 (69%) | 88A1 |
| MTD006.A07 | D | *Cucumis sativus* | Fatty acid hydroperoxide lyase | 3.20E-28 | 60/101 (59%) | |
| SCL072.F07 | W | *Helianthus tuberosus* | Fatty acid hydroxylase | 1.40E-30 | 66/121 (54%) | 81B1 |
| MWL029.C02 | W | *A. thaliana* | Ferulate 5-hydroxylase | 1.40E-15 | 42/59 (71%) | 84A1 |
| MTD015.B08 | D | *L. esculentum* | Ferulate 5-hydroxylase | 2.20E-42 | 88/120 (73%) | 84 |
| MTD015.G03 | D | *S. bicolor* | Obtusifoliol 14 α-demethylase | 9.30E-37 | 78/127 (61%) | 51 |
| MTD005.F10[4] | D | *Triticum aestivum* | Obtusifoliol 14 α-demethylase | 1.70E-68 | 120/120 (100%) | 51 |
| MTD007.A12[4] | D | *T. aestivum* | Obtusifoliol 14 α-demethylase | 1.10E-63 | 129/129 (100%) | 51 |
| 0252_b03 | W | *Nicotiana plumbaginifolia* | Unknown | 1.00E-29 | 64/171 (37%) | |
| 0260_C10 | W | *H. tuberosus* | Unknown | 4.80E-16 | 45/101 (44%) | 81B1c |
| 0260_G04 | W | *S. bicolor* | Unknown | 1.00E-44 | 89/98 (90%) | 98A1 |
| 0304_C08 | W | *A. thaliana* | Unknown | 1.30E-27 | 58/137 (42%) | |
| 0304_H01 | W | *S. bicolor* | Unknown | 3.80E-23 | 53/75 (70%) | 98A1 |
| 0305_C11 | W | *A. thaliana* | Unknown | 8.90E-17 | 46/119 (38%) | |
| AWB002.G04 | W | *A. thaliana* | Unknown | 2.30E-26 | 56/100 (56%) | |
| AWB003.F10 | W | *A. thaliana* | Unknown | 3.90E-22 | 56/159 (35%) | |
| AWB006.C09 | W | *A. thaliana* | Unknown | 2.20E-51 | 99/159 (62%) | |
| AWB010.E08 | W | *A. thaliana* | Unknown | 1.30E-46 | 98/142 (69%) | 78A9 |
| CNW01PL0207 | W | *A. thaliana* | Unknown | 2.30E-25 | 58/105 (55%) | 89A2 |
| CSB006.G04 | W | *Capsicum annuum* | Unknown | 2.10E-29 | 79/191 (41%) | |
| CSB007.H02 | W | *C. annuum* | Unknown | 7.60E-22 | 49/102 (48%) | |
| HWM004.E02 | B | *A. thaliana* | Unknown | 1.10E-07 | 36/57 (63%) | |
| HWM004.F10 | B | *Z. mays* | Unknown | 2.10E-58 | 110/214 (51%) | 71C4 |
| HWM009.g12 | B | *Glycine max* | Unknown | 8.20E-22 | 52/95 (54%) | 71D10 |
| HWM013cE.07r | B | *N. plumbaginifolia* | Unknown | 5.80E-44 | 93/200 (46%) | |
| ISC004.A02[c] | B | *Persea americana* | Unknown | 4.60E-22 | 51/98 (52%) | 71A1 |
| ISC009.H03[c] | B | *Solanum verbascifolium* | Unknown | 4.40E-06 | 28/89 (31%) | 71A2 |
| MCG005.H01 | B | *A. thaliana* | Unknown | 9.70E-42 | 87/206 (42%) | |
| MCG006.C04 | B | *A. thaliana* | Unknown | 1.30E-27 | 80/217 (36%) | |
| MCG007.D10 | B | *A. thaliana* | Unknown | 4.00E-32 | 54/102 (52%) | |
| MCG010.C12 | B | *A. thaliana* | Unknown | 1.40E-38 | 57/112 (50%) | |
| MCG014.D01 | B | *S. bicolor* | Unknown | <5.1E-68 | 132/151 (87%) | 98A1 |
| MTD007.B10 | D | *N. plumbaginifolia* | Unknown | 3.00E-29 | 67/127 (52%) | |
| MTD008.B02 | D | *N. tabaccum* | Unknown | 9.90E-43 | 79/117 (67%) | |
| MTD010.G09 | D | *A. thaliana* | Unknown | 9.10E-13 | 48/124 (38%) | |
| MTD012.G04 | D | *A. thaliana* | Unknown | 3.20E-34 | 69/122 (56%) | |
| MWL001.A04 | W | *A. thaliana* | Unknown | 1.70E-22 | 61/125 (48%) | 81E1 |
| MWL001.G10 | W | *N. tabaccum* | Unknown | 9.80E-32 | 59/101 (58%) | 71A10 |
| MWL006.D06 | W | *A. thaliana* | Unknown | 3.70E-26 | 35/72 (48%) | 88A3 |
| MWL024.A02 | W | *L. esculentum* | Unknown | 5.80E-17 | 30/46 (65%) | |
| MWL024.B03 | W | | Unknown | 2.40E-12 | 28/79 (35%) | |
| MWL032.C01 | W | | Unknown | 4.90E-07 | 29/68 (42%) | |
| PSR6476 | W | Chaco potato | Unknown | 1.40E-40 | 87/207 (42%) | 71D6 |
| SCC001.F07 | S | *Z. mays* | Unknown | 2.00E-19 | 49/91 (53%) | 71C3 |
| SCE001.f07 | | *Z. mays* | Unknown | 4.50E-22 | 58/120 (48%) | 71C3 |
| SCL081.A02 | W | *A. thaliana* | Unknown | 6.00E-29 | 55/89 (61%) | 75A2 |
| SCU001.E06 | W | *P. americana* | Unknown | 1.90E-27 | 74/171 (43%) | 71A1 |
| SFR006.E10 | B | *Streptomyces coelicolor* | Unknown | 1.00E-06 | 18/39 (46%) | 94A1 |

**Table 2** (continued)

| ITEC EST | | Best match in Genbank | | | | |
|---|---|---|---|---|---|---|
| EST name | EST origin* | Species | Function | Prob. | aa Identity | CYP no. |
| SUN000.A12 | W | *A. thaliana* | Unknown | 5.50E-21 | 34/57 (59%) | 86A1 |
| SUN003.F06 | W | *A. thaliana* | Unknown | 5.00E-21 | 37/52 (71%) | 86A1 |
| SUN004.B07 | W | *G. max* | Unknown | 3.50E-37 | 76/135 (56%) | 71D10 |
| SUN010.F11 | W | *A. thaliana* | Unknown | 3.80E-22 | 52/101 (51%) | |
| TAS003.H11 | W | *A. thaliana* | Unknown | 1.70E-17 | 30/65 (46%) | |
| TJC001d_b03 | W | *A. thaliana* | Unknown | 2.70E-17 | 46/109 (42%) | |
| TJC040.g05 | W | *A. thaliana* | Unknown | 4.60E-14 | 46/111 (41%) | 89A2 |
| WHE0025.F11 | W | *G. max* | Unknown | 5.00E-23 | 51/80 (63%) | 71D10 |
| WHE0026.F05 | W | *G. max* | Unknown | 1.10E-33 | 66/104 (63%) | 71D10 |
| WWR020.E7 | W | *A. thaliana* | Unknown | 8.20E-29 | 61/92 (66%) | 81B1c |
| WWS016.G8 | W | *S. bicolor* | Unknown | 8.20E-45 | 93/109 (85%) | 98A1 |

* W = bread wheat (*Triticum aestivum*), B = barley (*Hordeum vulgare*), D = durum wheat (*Triticum turgidum*) and S = rye (*Secale cereale*)

[1, 2, 3, 4] Sequence alignment indicates that ESTs with the same superscript number are likely to be from the same gene

[a, b, c] Sequence alignment suggests that sequences with the same superscript letter are likely to be from homologous or homoeologous loci

**Table 3** Primers used to amplify the P450 sequences from barley genomic DNA. ITEC ESTs used to design PCR primer pairs, source species of the cDNA library from which the EST was derived, forward primer sequence, reverse primer sequence, $MgCl_2$ concentration used in PCR reactions and product size predicted from the EST sequence

| EST | Source | Forward sequence | Reverse sequence | ($MgCl_2$) mM | Product size (bp) |
|---|---|---|---|---|---|
| HWM002.E11 | *H. vulgare* | TTCTCGGCACCAAGGTTGAGTT | CGGCTTATTACACTGGGCAAGATT | 2.0 | 302 |
| HWM004.E02 | *H. vulgare* | TGCGTGCTGGAAGAGTTTGAA | TGACGATGGACCAACATACCAAGTA | 3.0 | 258 |
| HWM009.G12 | *H. vulgare* | GCTCGGGTTGACGTGGACTTTA | GCTCGTTCCTCTTACTCTGTTATGTCTT | 3.0 | 356 |
| ISC004.A02 | *H. vulgare* | GATGAAGGAACCAATTAGCACTT | GGAGAACGCCGAGGAGTT | 2.0 | 335 |
| MCG001.A05 | *H. vulgare* | CTTATACAAACACATGCCCGCAT | AGGATGCCAAGCTCACTGGTTA | 1.5 | 442 |
| MCG007.D10 | *H. vulgare* | GACGGCAAGGAGGTGGACAT | GCTACTCGTTCTTACAAAACCCTGATA | 1.5 | 237 |
| MTD005.F10 | *T. turgidum* | GATCATGCTGCTTCGCCAAT | CCCCCAAAGGAGATGTACGAGAA | 3.0 | 222 |

Sloop was mapped using a doubled-haploid population from the Australian National Barley Molecular Marker Program.

### PCR amplification conditions

PCR reactions contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.25 mM dNTPs, 0.25 µM of each primer, 0.8 ng/µL of template DNA and 0.05 U/µL of *Taq*. Magnesium chloride concentrations were optimised for each primer pair (see Table 3). During PCR amplification one cycle of 94 °C for 3 min was followed by a "touchdown" consisting of 10 cycles of denaturing for 1 min at 94 °C followed by annealing for 1 min then extension for 1 min at 72 °C. Annealing temperatures were progressively decreased by 1 °C per cycle from 60 °C to 50 °C. The PCR continued for 25 additional cycles at 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s. The reaction ended with an extension at 72 °C for 3 min. PCR products were purified using Microcon PCR clean up columns (Millipore, Bedford, Mass. 01730, USA) prior to sequencing. Sequencing reactions were carried out using Big Dye Terminator Ready Reaction mix (Applied Biosystems, Foster City, Calif.).

## Results

### Identification of ESTs with sequence homology to P450 genes

A total of 75 ESTs or 0.3% of the ESTs in the ITEC database were identified as belonging to the P450 gene family (see Tables 2, 3). On the basis of sequence alignments it was estimated that a total of 61 P450 genes were represented in the ESTs giving a redundancy rate of 20%. This estimate most likely represents a lower limit for redundancy, as ESTs from the same gene but without overlap would not have been detected.

High levels of sequence homology to genes of known function in the Genbank database allowed the function of genes represented by the ESTs to be tentatively assigned in 24 cases (Table 2). These included: allene oxide synthase (represented by both barley and wheat ESTs), obtusifoliol 14α-demethylase (from durum wheat), ferulate 5-hydroxylase (from durum wheat) and also several wheat and several barley ESTs that are likely to be from cinnamate 4-hydroxylase genes (Table 2). Most of the ESTs (51 ESTs or 68%) had best matches to gene sequences classified as P450s but

**Table 4** Summary of SNP data from barley P450 sequences

| EST | Function of best-matching Genbank sequence | Amino-acid identities | Number of SNPs | Number of haplotypes | RS change |
|---|---|---|---|---|---|
| HWM002.E11 | Allene oxide synthase | 83/83 (100%) | 0 | 1 (11) | |
| HWM004.E02 | Unknown | | 4 | 2 (9,2) | N |
| HWM009.G12 | Unknown | | 1 | 2 (9,2) | N |
| ISC004.A02 | Unknown | | 3 | 3 (8,2,1) | Y (2) |
| MCG001.A05 | Cinnamate 4-hydroxylase | 101/134 (75%) | 5 | 3 (8,1,1) | Y (3) |
| MCG007.D10 | Unknown | | 1 | 2 (7,4) | Y (1) |
| MTD005.F10 | Obtusifoliol 14 α-demethylase | 129/129 (100%) | 0 | 1 (11) | |

**Table 5** Summary of data from wheat P450 ESTs identified in ITEC. ITEC EST designation, species of origin of cDNA library containing ESTs, function of best matching gene in Genbank to the EST and amino-acid identity

| EST | Source | Function of best-matching Genbank sequence | Amino-acid identities |
|---|---|---|---|
| 0302_b03 | *T. aestivum* | Allene oxide synthase | 58/59 (100%) |
| CSB006.G04 | *T. aestivum* | Unknown | |
| MTD005.F10 | *T. turgidum* | Obtusifoliol 14 α-demethylase | 129/129 (100%) |
| MTD015.G03 | *T. turgidum* | Unknown | |
| SCL081.A02 | *T. aestivum* | Unknown | |
| SCU011.B10 | *T. aestivum* | Unknown | |
| SUN004.Bo7 | *T. aestivum* | Unknown | |

with unknown function, many of these originating from *Arabidopsis* (Table 2).

## Search for SSRs

Although 1,366 ESTs containing dinucleotide ($n > 6$) and trinucleotide ($n > 6$) repeats have been identified in the ITEC database (Holton et al. 2000) none of these corresponded to any of the identified P450 ESTs.

## Identification of SNPs

Primers were designed to amplify the sequence for 13 P450 ESTs from hexaploid wheat, nine from barley, four from durum wheat and one from rye. Using genomic barley DNA as a template it was possible to sequence PCR products amplified using seven primer pairs; six pairs derived from barley ESTs and one from a durum EST (Table 3). SNPs between the 11 barley varieties were detected in five of the seven PCR products with up to five SNPs between varieties detected in products from any one primer pair (Table 4; SNP alleles, position in EST and distribution in varieties is given in the Appendix: Table A). In the seven PCR products only a single indel of one base was detected (Appendix, HWM004.E02 base 324). The overall frequency of SNPs in the seven PCR products sequenced for the 11 barley varieties was 1 every 131 bases (priming sites excluded). For three PCR products SNPs caused a change of restriction site between varieties (Table 4). In these cases genotyping can be carried out by simple restriction digestion of PCR products and agarose-gel electrophoresis.

As an example, the SNPs detected using primers derived from barley EST ISC004.A02 are illustrated in Fig. 1. Three SNPs between the 11 lines were detected

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chebec | CCTGGC | CTTAAG | TCTGCCTTGATCCTTGCCG | **T** | AAAACCTTTCTAG | | | | | | | | | | | | | |
| Galleon | CCTGGC | CTTAAG | TCTGCCTTGATCCTTGCCG | **T** | AAAACCTTTCTAG | | | | | | | | | | | | | |
| Halcyon | CCTGG | **G** | CTT | **C** | AGTCTGCCTTGATCCTTGCCG | **T** | AAAACCTTTCTAG | | | | | | | | | | | |
| Others[a] | CCTGGC | CTTAAG | TCTGCCTTGATCCTTGCCG | **C** | AAAACCTTTCTAG | | | | | | | | | | | | | |
| AA[b] | | P | G | L | K/**Q** | S | A | L | I | L | A | A/**V** | K | P | F | * | | |
| | | | | CTT | AAG | | | | | | | | | | | | | |
| | | | | *Afl*II[c] | | | | | | | | | | | | | | |

**Fig. 1** Alignment of DNA sequences amplified from 11 barley varieties using primers designed to target barley EST ISC004.A02. Three haplotypes are evident, one for varieties Chebec and Galleon, one for Halcyon and one for the other 8 varieties. SNPs are indicated in bold text and the restriction site for *Afl*II is underlined. For 2 of the 3 SNP sites alternative amino acids are encoded

and three different haplotypes identified. An SNP between Halcyon and the other ten varieties leads to the loss of an *Afl*II restriction site in Halcyon, allowing mapping of the P450 EST in Halcyon crosses. Using a population of 166 doubled-haploids made from a cross between barley varieties Halcyon and Sloop this EST was mapped to the same position on chromosome 1H as the Scottish Crop Research Institute SSR marker Bmag0382.

PCR primers designed to target barley EST SSRs can be transferable to wheat (Holton et al. 2000) while nine wheat ESTs and four durum ESTs were used to design P450 EST primers (Table 3). It was, therefore, of interest to determine whether the P450 EST primers could amplify products using genomic wheat DNA templates and if so whether SNPs could be detected among wheat varieties. Primer pairs derived from seven ESTs gave products of the anticipated size range including five pairs derived from hexaploid wheat ESTs and two from durum wheat ESTs (Table 5). One durum EST-derived primer pair was amplified in both wheat and barley (MTD005.F10; Table 5); however, no barley EST-derived primer pairs amplified in wheat. Attempts to assign wheat PCR products to chromosomes using nullitetrasomic lines failed in all instances due to amplification of multiple products of

similar size. For several of the EST-derived primer pairs high-resolution separation of PCR products amplified from wheat revealed the presence of more than one band. Cloning and sequencing of the PCR fragments confirmed the presence of multiple sequences. Commonly small insertions/deletions were present in the different sequences from a cloned amplification. Products amplified from barley genomic DNA and their sequences showed no evidence of multiple locus amplification. Efforts to design locus-specific primers for wheat are continuing.

## Discussion

We have demonstrated that SNP markers can be used to differentiate between P450 genes from different barley varieties and that these SNPs can be discovered using an EST database. In the ITEC database of 24,344 ESTs, 75 P450 ESTs were detected representing 0.3% of the total.

Using seven pairs of P450 EST-derived primers to amplify and sequence barley P450s we were able to detect SNPs between 11 barley varieties in the products from five primer pairs (Table 3). Interestingly, a number of the SNPs that we identified could be differentiated on the basis of changes in restriction sites between lines. Thus, mapping of these SNPs may be achieved without the expense of specialist SNP detection equipment. If this high success rate for detecting SNPs between barley P450s were maintained when larger numbers of P450s are tested, we would predict that P450 SNPs might prove highly informative.

One of the barley ESTs used to design primers that amplify products with a SNP in barley exhibited high levels of sequence homology to a plant P450 of known function (Table 4). The barley EST MCG001.A05 exhibited 75% amino-acid sequence homology to cinnamate 4-hydroxylase, which is involved in the phenylpropanoid pathway that gives rise to important metabolites including lignin, flavonoids, hydroxycinnamic acid esters, lignans, stilbenes and many other secondary metabolites. Traits associated with this marker have yet to be demonstrated; however, the important processes affected by this enzyme suggest that changes at this locus could be of importance.

The other four ESTs used to design primers yielding SNPs in barley exhibited sequence homology to P450s of unknown function. Therefore, traits that map to the same location as these markers will be of potential use in assigning a function to these genes.

SSRs occur less frequently than SNPs in the plant genome and we were unable to demonstrate the presence of any useful SSRs in the 75 P450 ESTs identified. Clearly this robust PCR-based marker system is unlikely to prove as useful for mapping plant P450s as SNPs.

The P450 EST-derived primers were not suitable for identifying SNPs between wheat varieties due to problems with amplification of multiple PCR products. Since there was no evidence of multiple products in amplifications from barley, it is likely that the amplification of multiple products from wheat is due to the presence of the A, B and D genomes. The highly similar sequences present in clones from wheat amplification products are reminiscent of the same gene from different organisms, and the multiple wheat amplicons probably originate from homoeologous loci rather than amplification from other members of the P450 gene family. Presumably the diversity in the P450 gene family means that primers designed to amplify one member are generally gene specific. Identification of SNPs in P450s from bread wheat may have to rely on EST databases with sufficient redundancy to allow for the confident construction of contigs for each homoeolog of each particular P450 gene.

P450 SNPs between barley varieties can also give information on genetic relationships. Cluster analysis of genetic distance data calculated for a set of 80 barley accessions using RFLP information from a total of 1,324 polymorphic fragments has demonstrated that Sahara and Halcyon are clearly separated from the other nine barley lines used in the present study (Chalmers et al. 2001). Sahara had an uncommon haplotype for four of the seven PCR products whilst Halcyon had the least common haplotype for three (Table 4, Appendix Table A). Thus, the P450 SNP data presented here support the RFLP data.

The frequency of SNPs found among the 11 barley varieties (1 per 131 bases or 0.0076) is lower, though of similar magnitude, to a large study conducted in maize where the frequency of SNPs was found to be 1 every 83 bases for eight maize inbreds (Bhattramakki and Rafalski 2001). It was suggested that a high frequency of SNPs would be expected for maize as it is open-pollinated (Bhattramakki and Rafalski 2001). Be that as it may, the abundance of SNPs found is dependent on the diversity of the varieties sampled and we would predict that the inclusion of wild barley varieties in the present study would increase the frequency of SNPs detected.

We anticipate that the use of EST databases to find SNPs between barley P450s will provide useful markers for plant breeders and will provide additional markers for studying genetic relationships in barley. Barley P450 ESTs could also provide a useful resource for determining the function of some of the majority of plant P450s for which function is unknown. With the rapidly increasing numbers of plant ESTs and genes being added to data bases it is anticipated that the potential to detect SNPs in plant P450 genes between the individual parents of mapping populations should increase rapidly.

**Acknowledgements** This work was funded by the Cooperative Research Centre for Molecular Plant Breeding and the Grains Research and Development Corporation. We would also like to thank Gai McMichael for her assistance with mapping.

## Appendix

**Table A** The positions of single nucleotide polymorphisms (SNPs) found in five cytochrome P450 genes amplified from barley

| EST name and source variety | Genbank accession no. | Base position of SNP[a] | Rare SNP allele | Common SNP allele | Varieties with rare allele |
|---|---|---|---|---|---|
| HWM004.E02 | BE420973 | 155 | A | T | Arapiles/Sahara/Barke(EST) |
| Barke | | 158 | C | G | Arapiles/Sahara/Barke(EST) |
| | | 167 | A | G | Arapiles/Sahara/Barke(EST) |
| | | 262 | A | G | Arapiles/Sahara/Barke(EST) |
| | | 279 | A | G | EST only |
| | | 324 | T insert | – | Arapiles/EST |
| HWM009.G12 Barke | BE421487 | 336 | T | C | Harrington/Sahara |
| ISC004.A02 | BE411284 | 125 | A | G | Galleon/Chebec/Halcyon |
| Nure | | 147 | G | T | Halcyon |
| | | 151 | C | G | Halcyon |
| MCG001.A05 | BE412402 | 180 | A | G | Sahara/Halcyon |
| unknown | | 208 | A | G | Sahara/Halcyon |
| | | 253 | C | G | Sahara/Halcyon |
| | | 283 | C | T | Sahara/Halcyon |
| | | 319 | A | C | Sahara/Halcyon |
| MCG007.D10 unknown | BE412662 | 257 | C | G | Haruna nijo/Harrington/ Halcyon/ Sahara |

[a] SNP positions are given relative to parent EST sequences

## References


Archakov A, Lisitsa A, Gusev S, Koyamans L, Janssen P (2001) Inventory of the cytochrome P450 superfamily. J Mol Modeling 7:140–142

Bak S, Olsen CE, Halkier BA, Moller BL (2000) Transgenic tobacco and Arabidopsis plants expressing the two multifuntional sorghum cytochrome P450 enzymes, CYP79A1 and CYO71E1, are cyanogenic and accumulate metabolites derived from intermediates of dhurin biosynthesis. Plant Physiol 123:1437–1448

Bak S, Tax FE, Feldmann KA, Galbraith DW, Feyereisen R (2001) CYP83B1, a cytochrome P450 at the metabolic branch point in the auxin and indole glucosinolate biosynthesis in *Arabidopsis*. Plant Cell 13:101–111

Bhattramakki D, Rafalski A (2001) Discovery and application of single nucleotide polymorphism markers in plants. In: Henry RJ (ed) Plant genotyping: the DNA fingerprinting of plants. CABI Publishing, Oxon, UK, pp 179–191

Bishop GJ, Harrison K, Jones JDG (1996) The tomato *dwarf* gene isolated by heterologous transposon tagging encodes the first member of a new cytochrome P450 family. Plant Cell 8:959–969

Bolwell GP, Bozak K, Zimmerlin A (1994) Plant cytochrome P450. Phytochemistry 37:1491–1506

Brugliera F, Barri-Rewell G, Holton TA, Mason JG (1999) Isolation and characterization of a flavonoid 3′-hydroxylase cDNA clone corresponding to the Ht1 locus of *Petunia hybrida*. Plant J 19:441–451

Chalmers KJ, Jeffries SP, Langridge P (2001) Comparison of RFLP and AFLP marker systems for assessing genetic diversity in Australian barley varieties. In: Henry RJ (ed) Plant genotyping: the DNA fingerprinting of plants. CABI Publishing, Oxon, UK, pp 161–178

Chapple C (1998) Molecular-genetic analysis of plant cytochrome P450-dependant monoxygenases. Annu Rev in Plant Physiol Mol Biol 49:311–343

Feynereisen R (1999) Insect P450 enzymes. Annu Rev Entomol 44:507–533

Frey M, Kliem R, Saedler H, Gierl A (1995) Expression of a cytochrome P450 gene family in maize. Mol Gen Genet 246:100–109

Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP, Simcox K, Gierl A (1997) Analysis of a chemical defence mechanism in grasses. Science 277:696–699

Holton TA, Brugliera F, Lester DR, Tanaka Y, Hyland CD, Menting JGT, Lu CY, Farcy E, Stevenson TW, Cornish EC (1993) Cloning and expression of the cytochrome-P450 genes controlling flower colour. Nature 366:276–279

Holton TA, McClure L, Christopher JT, Henry RJ (2000) Mapping barley with SSRs and SNPs. International Barley Genetics Symposium, Adelaide, Australlia

Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys 369:1–10

Ohkawa H, Tsumii H, Shimoji M, Imajaku Y, Imaishi H (1999) Cytochrome P450 biodiversity and plant protection. J Pesticide Sci 24:197–203

Paquette SM, Bak S, Feyereisen R (2000) Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. DNA Cell Biol 19:307–317

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Nat Acad Sci USA 85:2444–2448

Schuler MA (1996) Plant cytochrome P450 monoxygenases. Crit Rev Plant Sci 15:235–284

Wang FM, Wang R, DeParasis J, Loughrin JH, Gan SS, Wagner GJ (2001) Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. Nature Biotechnol 19:371–374

Werck-Reichart D, Hehn A, Didierjean L (2000) Cytochrome P450 for engineering herbicide tolerance. Trends Plant Sci 5:116–123

Winkler RG, Frank MR, Galbraith DW, Feyereisen R, Feldmann KA (1998) Systematic reverse genetics of transfer-DNA-tagged lines of *Arabidopsis*: isolation of mutations in the cytochrome P450 gene superfamily. Plant Physiol 118:743–749

Winkler RG, Helentjaris T (1995) The maize *dwarf3* gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis. Plant Cell 7:1307–1317

Worley KC, Wiese BA, Smith RF (1995) BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. Genome Res 5:173–184

Worley KC, Culpepper P, Wiese BA, Smith RF (1998) BEAUTY-X: enhanced BLAST searches for DNA queries. Bioinformatics 14:890–891