Theor Appl Genet (2004) 109: 543–551
DOI 10.1007/s00122-004-1675-z

ORIGINAL PAPER

P. C. Bundock · R. J. Henry

Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley

Received: 8 August 2003 / Accepted: 19 March 2004 / Published online: 14 May 2004

Abstract The *Isa* gene from barley—an intronless gene expressed in maternal tissues of the seed—has a likely role in defence against pathogens. The protein product—bi-functional α-amylase/subtilisin inhibitor—inhibits the seed's own amylase in addition to the bacterial protease subtilisin and fungal xylanase. Sixteen barley genotypes were targeted to amplify and sequence the *Isa* gene region to detect sequence polymorphisms, since little is known about genetic diversity at this locus. A total of 80 single nucleotide polymorphisms (SNPs) and 23 indels were detected in 2,164 bp of sequence containing the *Isa* transcript, promoter and 3′ non-transcribed region (overall one SNP per 27 bp and one indel per 94 bp), with eight sequence-based haplotypes distinguishable amongst the 16 varieties. Sequencing a polymorphic region in the promoter in an additional 27 barley genotypes increased the number of sequence-based haplotypes discovered to 11. However there is low haplotype diversity amongst the cultivated barley varieties sampled, with most varieties represented by a single haplotype. There was minor amino acid diversity in the protein, with five out of ten SNP sites in the coding region predicted to produce amino acid substitutions. SNP analysis indicated a history of recombination events—a minimum of seven based on the initial eight haplotypes from the whole sequenced region. Most of the recombination events occurred in the highly polymorphic regions, the 3′ non-transcribed region and sequences flanking a microsatellite in the *Isa* promoter.

Communicated by P. Langridge

P. C. Bundock (✉) · R. J. Henry
Molecular Plant Breeding CRC, Centre for Plant Conservation Genetics, Southern Cross University,
Lismore, NSW, 2480, Australia
e-mail: pbundock@scu.edu.au
Tel.: +61-2-66203581
Fax: +61-2-66222080

Introduction

The *Isa* gene of barley (Hejgaard et al. 1984) is a single-copy intronless gene that codes for the single-polypeptide protein bi-functional α-amylase/subtilisin inhibitor (BASI, Leah and Mundy 1989). BASI has been reported to selectively inhibit the high isoelectric point group of barley α-amylases (AMY-2) coded by the *amy-1* locus in addition to the bacterial serine protease subtilisin (Mundy et al. 1983) and most recently, fungal xylanase (Sancho et al. 2003). The role of this apparently multifunctional protein in the seed has not yet been established. The inhibition of the plant's α-amylase is a function that appears to have evolved specifically in the Triticeae tribe (Henry et al. 1992). In developing grain it has recently been shown that the *Isa* promoter directs expression of green fluorescent protein in the pericarp, the vascular tissue, the nucellar projection cells and the endosperm transfer cells and also, in mature germinating grain in the embryo (Furtado et al. 2003).

Although the *Isa* gene has not as yet been associated with any phenotypic effect in barley, based on its expression pattern and the in vitro inhibitory effects of the protein product, *Isa* is likely to be involved in the seed's defence against pathogens. Genetic variation for the quantity of BASI protein in barley grain has been found (Jarrett et al. 1997). However studies of the mechanisms of BASI inhibition (Abe et al. 1993; Rodenburg et al. 2000) and the crystal structure of the AMY-2–BASI complex (Vallee et al. 1998) have been based on only one or two genotypes. In addition the expression pattern has been studied using the promoter from one genotype only (Furtado et al. 2003), and the extent of sequence variation within cultivated barleys and between cultivated and wild barleys is not known.

We have generated sequence information for a number of domesticated barleys and compared these with a small group of wild barleys to determine the extent of nucleotide variation in the *Isa* coding and regulatory sequences in domesticated barley.

Materials and methods

Plant material

Sixteen barley varieties/accessions were used for PCR amplification for re-sequencing of 2,164 bp of the *Isa* gene (Table 1, region I). Barley varieties were selected from those used in the Australian Barley Molecular Marker Program, which included a single wild barley—*Hordeum spontaneum* 35. Three additional wild barleys were used for amplification and sequencing of the entire *Isa* region. A further 30 varieties, landraces or wild barley accessions were selected for the second part of the study to assess the distribution of polymorphisms around the microsatellite site in the promoter region (Table 1, region II).

PCR amplification, SSR sizing and re-sequencing of the *Isa* gene

For amplification for re-sequencing from the 16 barley varieties, the *Isa* gene sequence was divided into four regions (Fig. 1; Table 2, products 1–4). One wild barley (*H. spontaneum* 39) would not PCR at the 3′ end (product 4), and two alternative products (products 5 and 6, Table 2) were amplified and sequenced for this accession. Design of PCR primers was based on the sequence obtained by Furtado et al. (2003). Re-sequencing a small region (~290 bp) flanking the microsatellite was also carried out for the additional 30 varieties (Table 1, region II). Sizing of a PCR product harbouring the dinucleotide microsatellite (SSR) located in the promoter region of the *Isa* gene was carried out for all 46 varieties. The M2F primer was fluorescently labelled (6FAM), and sizing of the resulting PCR product was carried out using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, Calif., USA).

The following concentrations of reaction components were used in PCR amplifications: 0.2 μM each primer, 0.2 mM each dNTP, 0.05 U/μl *Taq* polymerase (Roche), with 1/10 volume of reaction buffer and 0.8 ng/μl of template DNA.

Glycerol at a concentration of 10% (v/v) was also added to the reactions for the amplification of PCR product 3, since this region has a high GC content. The primer sequences and cycling conditions are presented in Table 2. All singleton single nucleotide polymorphism (SNP) and indel sites were checked by carrying out sequencing in both forward and reverse orientations of the PCR products with the exception of the region flanking the microsatellite. Polymorphisms found in the microsatellite flanks were checked by carrying out at least two forward or two reverse sequencing reactions, depending on the side of the flank in which the polymorphism was found. Amplification was carried out using a 96-well GeneAmp PCR System 9700 thermocycler (Applied Biosystems). PCR products for re-sequencing were purified using Amicon columns (Millipore, Bedford, Mass., USA). Sequencing reactions were performed using ABI PRISM Dye Terminator Version 3.0 (Applied Biosystems) at 1–4 μl per 16-μl reaction, with separation of products being carried out by the Australian Genome Research Facility using an automated sequencer (ABI3700).

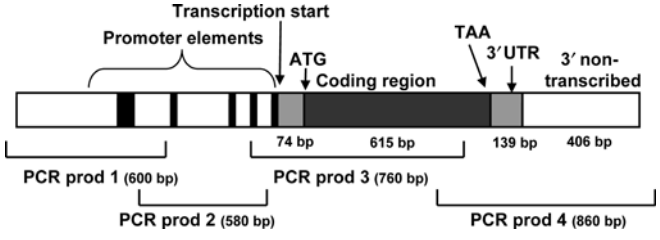

Fig. 1 Representation of the DNA region containing the *Isa* gene that was amplified and sequenced from 16 barley varieties. The relative positions of the four PCR products amplified for sequencing are indicated

Table 1 List of barley varieties used for amplification and sequencing of the *Isa* gene

Barley variety/accession	Country of origin	Region sequenced
Arapiles	Australia	I
Barque	Australia	I
Chebec	Australia	I
Clipper	Australia	I
Franklin	Australia	I
Galleon	Australia	I
Sloop	Australia	I
Alexis	Germany	I
Halcyon	UK	I
Haruna Nijo	Japan	I
Harrington	Canada	I
Sahara 3771	North African	I
Hordeum spontaneum 35 CPI 7128448–mountain	Israel	I
H. spontaneum 37 CPI 771292–coast	Israel	I
H. spontaneum 38 CPI 7713711–temperate	Israel	I
H. spontaneum 39–desert	Israel	I
Algerian	Algerian	II
Amaji Nijo	Japan	II
Andre	USA	II
Bearpaw	USA	II
Brindabella	Australia	II
Caminant	Denmark	II
Carmague	Germany	II
Chariot	UK	II
Ethiopia CI 3576	Ethiopia	II
CIMMYT 42002		II
Ellice	Canada	II
Ethiopia 183	Ethiopia	II
Igri	Germany	II
Kino Nijo	Japan	II
Lara	Australia	II
Malebo	Australia	II
Morex	USA	II
Namoi	Australia	II
Norbet	Canada	II
O'Connor	Australia	II
Osiris	France	II
Prior	Australia	II
Proctor	UK	II
Richard	Canada	II
Shapporo Shine 5	Japan	II
Sultan	Netherlands	II
Triumph	Germany	II
Ulandra	Australia	II
Yerong	Australia	II
H. spontaneum 36 CPI 7128527 – desert	Israel	II

Table 2 Primers used for PCR amplification and sequencing of the *Isa* gene

Product number, region	Product size (bp)	Primer name[a]: sequence	Cycling conditions
1, Upstream from the promoter	600	1F: ACTGGGCTCGAAACTAAAATAAGAACATG 1R: GGGCTTCACCTTTGTGAGACTCTCAC	94°C for 2 min, (94°C for 30 min, 52°C for 30 min, 72°C for 1 min)×30, 72°C for 5 min, 4°C hold
2, Promoter region to transcription start site	580	2F: ACTGCTTCCCCGTCATTCTGTC 2R: TGAAACCTCTGCTGGAGTGTCCT 2SF: CAAACGCTTGATTCGCA 2SR: GCATCACGAGTGCTTCAT	94°C for 5 min, (94°C for 30 min, 52°C for 30 min, 72°C for 30 min)×40, 72°C for 7 min, 4°C hold
3, Coding and 5′ non-translated sequence	760	3F: CACATCACGCAATCCACCAGAAG 3R: GAACACGACGACATGGTATGGCTC	94°C for 5 min, (94°C for 30 min, 50°C for 30 min, 72°C for 1 min)×35, 72°C for 5 min, 4°C hold
4, 3′ UTR[b], terminator and downstream	860	4F: CTACACGACGTGTCTGCAGTCCA 4R: GTGCTTGCCCATTCGATTATGTCGAG	94°C for 3 min, (94°C for 1 min, 60°C for 1 min, 72°C for 1 min −1°C per cycle)×10, (94°C for 30 min, 50°C for 30 min, 72°C for 1 min)×30, 72°C for 3 min, 4°C hold
5, 3′ UTR (used for wild barley 39 only)	434	5F: AACGCCTTCCGCATCGAGAA 5R: GGGGTTAGACGCACAGGTTAGTTT	94°C for 5 min, (94°C for 30 min, 58°C for 30 min, 72°C for 1 min)×35, 72°C for 5 min, 4°C hold
6, NTR[c] (used for wild barley 39 only)	286	6F: GCAAAGTAGCAGCAAGAAGAAATCA 6R: GCAACAAATGTGGTAAGTATGGGTA	94°C for 5 min, (94°C for 30 min, 52°C for 30 min, 72°C for 1 min)×35, 72°C for 5 min, 4°C hold
M1, Microsatellite region for sequencing	~290	M1F: CCAACTACTCCCTTCGTTTTTA M1R: GTGCTTGCCCATTCGATTATGTCGAG	94°C for 5 min, (94°C for 30 min, 50°C for 30 min, 72°C for 30 min)×35, 72°C for 7 min, 4°C hold
M2, Microsatellite region for allele sizing	~290	M2F: CAACTACTCCCTTCGTTTTT M2R: (6FAM) TTCTTCTGGTGGATTGCGTGAT	94°C for 5 min, (94°C for 20 min, 48°C for 20 min, 72°C for 20 min)×35, 72°C for 5 min, 4°C hold

[a]*F* Forward primer, *R*reverse primer, *S* primer used for sequencing
[b]*UTR* Untranslated region
[c]*NTR* Non-transcribed region

SNP identification and recombination history

Sequences were aligned, and SNPs and indels were identified from sequencing chromatograms using the program Sequencher (Gene Codes, Ann Arbor, Mich., USA). The distribution of polymorphisms across the *Isa* gene was tested for random expectations using a chi-square test. For this purpose the 2,164-bp region was divided into three sections covering 930 bp of upstream sequence (harbouring promoter elements), an 828-bp region covering the transcript [5′ untranslated region (UTR), coding region and 3′ UTR] and 406 bp of the 3′ downstream non-transcribed region (NTR) sequence. The program RecMin (Unix version, Myers and Griffiths 2003) was used to calculate the minimum number of recombination events evident in the *Isa* haplotypes and identify where (between which polymorphic sites) these recombination events occurred. Singleton sites were excluded as part of the analysis. Polymorphic sites (SNPs and indels) excluding microsatellite repeat length from across the entire sequence were considered for analysis, which excluded singleton sites. For the three tri-allelic sites, the two most common alleles were assigned a 1 or a 0, with the rare third allele designated as

unknown data (using a question mark). The program calculated both theK_m value of Hudson and Kaplan (1985) and the K_h of Myers and Griffiths (2003). Sequence comparisons were carried out using BLASTN and TBLASTX at the National Center for Biological Information Web site (http://www.ncbi.nlm.nih.gov/BLAST/) on non-redundant and EST databases.

Results

Location and frequency of SNP and indel polymorphisms

Sequence polymorphisms were detected among 16 barley varieties/accessions across 2,164 bp of sequence, which included 928 bp of upstream (5′ NTR; Tables 3,4), 828 bp of *Isa*gene transcript (5′ UTR, coding and 3′ UTR) and 406 bp of downstream (3′ NTR, Table 5). From the 16 barley genotypes 80 SNPs were detected (Table 6), giving an average SNP frequency of one SNP every 27 bases. Excluding differences in the number of repeats found within a microsatellite occurring in the promoter region (−297 to −286), there were 19 indels (insertions/deletions), consisting of nine single-base indels and ten indels ranging in size from 4 to 306 bases (Table 6). The average frequency of indels was one every 114 bases. These polymorphisms defined a total of eight haplotypes among the 16 barley varieties for which complete sequence information was obtained.

The frequency of SNPs was greatest in the 3′ NTR region (one SNP per 16 bases), and this was due entirely to SNPs in the wild barleys and the Sahara landrace accession (Table 6). A lower-than-average SNP frequency of one SNP per 75 bases was found in the transcribed region (Table 6), with ten SNPs occurring in the 615-bp coding region, none in the 74-bp 5′ UTR and a single SNP in the 139-bp 3′ UTR. The region flanking a microsatellite repeat in the promoter (100 bp either side) was particularly abundant in polymorphisms with 12 SNPs, and nine indels compared with an expected seven SNPs and two indels based on the frequency found overall. There were no indels found in the transcribed region, although based on the overall frequency, approximately seven indels would be expected in a region of this size. It is to be noted that there are three SNP sites (from the 80) that are tri-allelic, with two occurring in the promoter region (−315 and −273), and one in the downstream region (177; Tables 6, 7).

The observed distribution of SNP sites and indel sites was found to be significantly different (for SNPs χ^2, 2 *df*, $P<0.0001$; for indels $P<0.01$) from an expected even distribution across the three defined regions (upstream, transcript, downstream; Table 6). The uneven distribution of polymorphisms was due particularly to the low frequency found in the transcribed region (Table 6).

Five out of ten SNP sites in the coding region were predicted to produce amino acid substitutions. This is well below the expected frequency of amino acid changes resulting from random single-base substitutions in codon triplets, which on average across all codon triplets is 76%. Two non-synonymous substitutions were unique to the North African landrace accession Sahara 3771, two were shared by two *H. spontaneum* accessions (35 and 38) and one shared by three *H. spontaneum* accessions (35, 37 and 38). For the ‘rare’ allele for each of the five synonymous base substitutions, one was shared by two *H. spontaneum* accessions (35 and 38), one was unique to a single *H.*

Table 3 Single nucleotide polymorphism (SNP) and indel positions in the *Isa* gene—upstream region, 5′ of the $(AT)_n$ SSR site. SNPs relative to the most common sequence (haplotype 1) are indicated in*boldface*. The numbering of SNP positions is relative to the sequence published for var. Grimmett (Furtado et al. 2003). A *horizontal dash* indicates the absence of the indicated base, insertion or deletion from the sequence for that haplotype

Haplotype[a]	−921	−906	−899	−898	−874	−871	−860	−854	−846	−837	−828	−820	−818	−817	−812	−806	−805
1, 2, 3	T	T	A	C	G	–	G	C	C	T	T	G	C	T	C	T	T
4	T	T	A	C	G	–	G	C	C	T	T	G	C	T	C	T	T
5	T	T	A	**T**	**A**	–	**A**	C	C	**C**	T	G	C	T	C	T	T
6	T	T	A	**T**	**A**	–	**A**	C	C	**C**	T	G	C	T	C	T	T
7	**C**	T	A	C	G	–	G	**T**	**T**	**C**	**C**	**A**	**T**	**C**	**T**	**C**	**C**
8	T	**A**	**G**	C	G	306 bp del	–	–	–	–	–	–	–	–	–	–	–

Haplotype[a]	−781	−761	−740	−722	−706	−696	−666	−654	−652	−627	−613	−602	−579	−559	−481	−456	−430	−430	−350
1, 2, 3	C	G	A	A	C	A	G	T	A	C	C	C	C	A	A	T	T	–	T
4	C	G	A	A	C	A	G	T	A	C	C	C	C	A	A	T	T	–	C
5	C	G	**G**	A	–	A	G	**C**	A	C	C	**T**	C	**T**	A	**C**	–	–	**C**
6	C	G	**G**	A	–	A	**A**	**C**	A	C	C	**T**	C	**T**	A	T	–	–	**C**
7	**T**	**A**	**G**	**T**	C	**G**	G	**C**	**C**	**T**	**T**	C	**G**	**T**	–	**C**	T	–	**C**
8	–	–	–	–	–	–	–	–	–	–	–	–	–	**T**	–	**C**	T	**T**	**C**

[a]Varieties corresponding to each haplotype are: *1*Alexis, Arapiles, Barque, Chebec, Franklin, Galleon, Halcyon, Harrington and Sloop; *2* Haruna Nijo; *3* Sahara; *4* Clipper; *5 H. spontaneum 35*; *6 H. spontaneum 38*; *7 H. spontaneum 37*; *8 H. spontaneum 39*

Table 4 SNP and indel positions in the region flanking a $(TA)_n$ SSR in the promoter region of the *Isa* gene. SNPs relative to the most common sequence (haplotype 1) are indicated in *boldface*. The numbering of SNP positions is relative to the sequence published for Grimmett (Furtado et al. 2003). *A horizontal dash* indicates the absence of the indicated base, insertion or deletion from the sequence for that haplotype

Haplotype[a]	−330	−315	−301–298	−297–286	−283	−278	−277	−273	−271	−254–244	−239	−202	−134
1, 3	A	T	CAAATAAATAAA	$(TA)_n$	T	T	T	T	G	GTTATGT -TGTACG	T	A	C
2	A	T	CAAATAAATAAA	$(TA)_n$	T	T	T	T	G	GTTATGT -TGTACG	T	**T**	C
4	**T**	**G**	CAAA - - - - - - - -	$(TA)_n$	**G**	-	-	**G**	G	GC**G**ATGT**A**TG- - –	**C**	A	C
5, 6	A	T	- - - - - - - - - - - - - -	$(TA)_n$ $(\mathbf{G}A)_5$ $(TA)_2(\mathbf{G}A)_2$	**G**	T	T	**G**	**A**	GTTATGT**A**TGTACG	**C**	A	**T**
7	**T**	**C**	CAAA - - - - - - - - -	$(TA)_n$	**G**	T	T	**G**	G	- - - - - - - - - - - - - - - -	**C**	A	C
8	**T**	**C**	CAAA - - - - - - - - -	$(TA)_n$	T	T	T	**A**	G	GC**T**ATGT**A**TG - - - -	**C**	A	C
9	**T**	**C**	CAAA - - - - - - - - -	$(TA)_n$ $(\mathbf{TG})_2(TA)_8$	**G**	T	T	**G**	G	TCTTTTTC- - - - - - -	**C**	A	C
10	A	T	CAAATAAA - - - -	$(TA)_n$	T	T	T	T	G	GTTATGT-TGTACG	T	**T**	C
11	?	**C**	- - - - - - - - TAAA	$(TA)_n$	**G**	T	T	**G**	G	GC**T**ATGT**A**TG - - - -	**C**	A	C

[a]Varieties for each haplotype are: *1* Alexis, Andre, Arapiles, Barque, Bearpaw, Caminant, Carmague, Chariot, Chebec, CI3576, CIMMYT42002, Ellice, Franklin, Galleon, Halcyon, Harrington, Igri, Lara, Norbet, Proctor, Richard, Sloop and Triumph; *2* Haruna Nijo, Amaji Nijo, Kino Nijo and Shapporo Shine 5; *3* Sahara; *4*Clipper, Malebo, Morex, O'Connor, Osiris, Prior and Yerong; *5H. spontaneum* 35; *6H. spontaneum* 38; *7H. spontaneum* 37; *8H. spontaneum* 39; *9H. spontaneum* 36; *10* Sultan and Namoi; *11* Ethiopia 183

spontaneum accession (39), one was found in Clipper, Sahara and all four of the *H. spontaneum* accessions, and two are found in Sahara and three *H. spontaneum* accessions (35, 37 and 38). A 306-bp indel in the 5′ NTR of *Isa* appears to be the result of an insertion event. One fully sequenced wild barley (*H. spontaneum* 39) was found to have a 306-bp segment missing from the 5′ NTR with a second wild barley (*H. spontaneum* 36—PCRed but not sequenced) sharing this deletion. The 306-bp segment was found to match extremely closely to regions of sequence found in a DCINA-induced protein of unknown function from barley, a section of BAC clone containing the *Mla* locus and to a number of EST sequences.

Haplotype diversity and evidence for recombination

Based on the whole length of the 2,164 bases sequenced, eight haplotypes were detected with 9 of the 11 cultivated varieties sharing the same haplotype (haplotype 1; Tables 3, 4, 5). The variety Haruna Nijo differed from the 'cultivated' haplotype at a single SNP site (Table 3). As might be expected, most of the rare SNP and indel alleles were found in the wild barley accessions, with the landrace Sahara 3771 contributing a few unique SNP sites (Tables 5, 6). Clipper was the only cultivated variety to have sequence that diverged strongly from the cultivated haplotype, with SNP and indel sites mostly restricted to the region flanking the microsatellite repeat (Table 4). Clipper had several unique SNP and indel alleles, but was much less divergent from the cultivated haplotype than any of the wild barleys (Table 6).

After the removal of singleton sites, 48 polymorphic sites remained in the analysis of recombination. The Haruna Nijo haplotype—differentiated on the basis of one singleton site—was thus redundant, leaving seven haplotypes. Using the program RecMin (Myers and Griffiths 2003), it was found that a minimum of seven recombination events explain the pattern of polymorphisms in the seven haplotypes identified from the sequences across the entire *Isa* region. Seven recombination events were found using the algorithm of Hudson and Kaplan (1985), with no increase in this when Myers' and Griffiths' (2003) method was used. Recombinations are calculated to have occurred between the following informative sites within each of the three regions: a minimum of three recombinations in the 5′ NTR (−481 to −456, −298 to −263, −273 to −254), one in the transcribed region (within coding sequence, 360–425) and three in the 3′ NTR (361–369, 378–383, 422–460; Fig. 2). The consequences of recombination events are evident in the pattern of polymorphisms in the Clipper haplotype (haplotype 4), when compared to the other haplotypes. The Clipper sequence is the same as the commonest haplotype (haplotype 1) in the first 470 bases of sequence in the 5′ NTR (−921 to −350, Table 3). However across the remainder of the promoter region there are 12 polymorphisms, most of which Clipper shares with one or more wild barleys (Table 4). Throughout the coding and downstream regions Clipper is again virtually

Table 5 SNP and indel positions in the transcribed and downstream regions of the *Isa* gene. SNPs relative to the most common sequence (haplotype 1) are indicated in *boldface*. The numbering of SNP positions is relative to the sequence published for Grimmett (Furtado et al. 2003). A*horizontal dash* indicates the absence of the indicated base, insertion or deletion from the sequence for that haplotype. Insertions are as indicated

Haplotype[a]/	91	131	138	268	323	351	360	413	425	623	737	106	139–140	150	166	171	177	186	197	198
1, 2	C	C	C	A	G	T	A	T	T	G	G	A	–	G	A	A	T	C	C	T
3	**T**	C	**T**	A	G	T	A	**C**	**C**	**C**	G	A	AAAGTA	**A**	**C**	**G**	**A**	–	–	T
4	C	C	C	A	G	T	A	**C**	T	G	G	A	–	G	A	A	T	C	C	T
5	C	**G**	C	**T**	G	**C**	**G**	**C**	**C**	**C**	**A**	A	AAAGTA	**A**	**C**	**G**	**C**	–	–	T
6	C	**G**	C	**T**	G	**C**	**G**	**C**	**C**	**C**	**A**	A	AAAGTA	**A**	**C**	**G**	**C**	–	–	T
7	C	C	C	**T**	G	T	A	**C**	**C**	**C**	G	A	AAAGTA	**A**	**C**	**G**	**C**	–	–	T
8	C	C	C	A	**T**	T	A	**C**	T	G	G	**G**	–	G	A	A	T	C	**T**	**C**

Haplotype[a]/	221–222	236	245	270	270–271	272	285	293–301	306	314	315	361	368	369	378	383	389	405	422	450	453	460
1, 2	–	C	C	T	–	C	C	–	T	A	A	C	G	G	G	C	C	C	G	C	T	G
3	GTG**T**	**T**	C	T	–	C	**T**	–	T	A	A	**T**	G	**A**	**A**	**T**	C	**T**	**A**	**T**	T	G
4	–	C	C	T	–	C	C	–	T	A	A	C	G	G	G	C	C	C	G	C	T	G
5	GTGC	**T**	C	T	–	C	**T**	–	T	A	A	**T**	G	G	**A**	**T**	**T**	**T**	**A**	C	T	G
6	GTGC	**T**	C	T	–	C	**T**	–	T	A	A	**T**	G	**A**	**A**	**T**	**T**	**T**	**A**	C	T	**A**
7	GTGC	**T**	C	T	–	C	**T**	–	T	A	A	**T**	G	**A**	**A**	C	C	C	**A**	C	**A**	**A**
8	–	C	**T**	**A**	**A**	**T**	C	9 b del	**C**	**G**	**G**	**T**	**C**	G	G	C	C	C	G	C	T	G

[a]Varieties corresponding to each haplotype are: *1*Alexis, Arapiles, Barque, Chebec, Franklin, Galleon, Halcyon, Harrington and Sloop; *2* Haruna Nijo; *3* Sahara; *4* Clipper; *5H. spontaneum* mountain 35; *6H. spontaneum 38*; *7H. spontaneum 37*; *8H. spontaneum 39*

Table 6 Distribution of SNP and indel sites across the barley *Isa* gene based on sequences from sixteen barley varieties. Indel numbers and frequencies are in *parentheses*. The numbers of SNPs and indels are relative to the most common haplotype (haplotype 1), with differences in the number of repeats in the microsatellite excluded from the tally. Tally only includes SNPs and indels for the 16 varieties (eight haplotypes) for which sequence was obtained across the whole region (i.e. excluding haplotypes 9, 10 and 11)

Barley variety	5′ region (930 bp)	Transcript (828 bp)	3′ region (406 bp)	Total (2,164 bp)
Haruna Nijo	1 (0)	0 (0)	0 (0)	1 (0)
Clipper	9 (4)	1 (0)	0 (0)	10 (4)
Sahara	0 (0)	5 (0)	13 (4)	18 (4)
H. spontaneum 35	15 (4)	8 (0)	12 (4)	36 (8)
H. spontaneum 37	29 (3)	4 (0)	12 (4)	45 (7)
H. spontaneum 38	16 (4)	8 (0)	14 (4)	39 (8)
H. spontaneum 39	10 (5)	2 (0)	11 (1)	23 (6)
Observed total	44 (13)	11 (0)	25 (6)	80 (19)
Expected (even distribution)	34.4 (8.2)	30.6 (7.3)	15 (3.5)	NA[a]
Frequency of bases/SNP (indel)	21 (72)	75 (−)	16 (68)	27 (114)

[a]*NA* Not applicable

identical to haplotype 1 (Table 5). This suggests that the *Isa* sequence found in Clipper has resulted from at least two recombination events in the past relative to the commonest haplotype (or vice versa).

The sequence from the variety Sahara also provides evidence of recombination. The (930-bp) 5′ upstream region of Sahara is identical to the commonest haplotype, but in the transcribed region there are five differences between the two haplotypes, whilst the 3′ downstream region (where there are 13 SNPs and four indel differences) is almost identical to the sequence from three wild barley accessions (35, 38 and 37; Tables 3, 4, 5). In a comparison of Sahara with the commonest haplotype and wild haplotypes, there is thus evidence for at least one recombination event, evident in the transcribed region (Fig. 3).

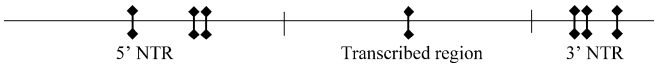

Fig. 2 Location of recombination events in the *Isa* gene region. The approximate location of each of seven recombination events is indicated on the *horizontal bar* representing the length of the region sequenced. *Fine vertical lines* represent the borders between the transcribed region and the 5′ non-transcribed region (NTR) and 3′ NTR

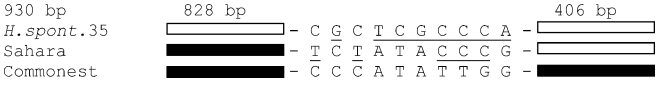

Fig. 3 Evidence of recombination in the *Isa* gene. *Letters* indicate single nucleotide polymorphism sites present in the transcribed region (see Table 6) with differences from haplotype 1 (cultivated barleys) *underlined*. The sequence of the upstream and downstream regions are indicated by a *clear box* for the *Hordeum spontaneum* 35 (haplotype 5) sequence and a *black box* for the haplotype 1 sequence

Table 7 SSR repeat length polymorphism in the promoter of the barley *Isa* gene

Flanking sequence haplotype	Variety	PCR product size (bp)	Number of TA (GA) repeats
1	Barque, Caminant, Chebec, CI3576, Sloop, Triumph	310	11
1	Alexis, Andre, Arapiles, Bearpaw, Carmague, Chariot, CIMMYT 42002, Ellice, Franklin, Galleon, Halcyon, Harrington, Igri, Lara, Norbet, Proctor, Richard	312	12
2	Haruna Nijo, Amaji Nijo, Kino Nijo, Shapporo Shine 5	312	12
3	Sahara	328	20–21
4	Clipper, Malebo, Morex, O'Connor, Osiris, Prior, Yerong	294	8 (1)
5, 6	*H. spontaneum* 35–mountain, *H. spontaneum* 38–temperate	301	8 (7)
7	*H. spontaneum* 37–coast	340	ND
8	*H. spontaneum* 39–desert	301	11 (1)
9	*H. spontaneum* 36–desert	307	13 (1)
10	Sultan, Namoi	310	12
11	Ethiopia 183	297	8 (1)
ND[a]	Brindabella	294, 312	ND
ND	Ulandra	294, 310	ND
ND	Algerian	335	ND

[a]*ND* Not determined

Evaluation of polymorphisms in a wider sample of germplasm

In the 100 bases flanking either side of the TA repeat in the *Isa* promoter, the frequency of SNPs was around twice the overall average, and the frequency of indels more than four times the overall frequency (based on sequence from the 16 barley varieties). Sequence diversity was explored further in this region by sequencing products amplified from an additional 30 varieties. However from these 30 varieties, two varieties, Ulandra and Brindabella, were found to have at least two haplotypes of the *Isa* gene, being either heterozygous or composed of a mixture of genotypes and could not be sequenced with clarity. In addition the landrace Algerian was found to have an extremely long amplification product, probably due to extensive repeats in the microsatellite region and could also not be sequenced. Sequence from the microsatellite-flanking region differentiates nine haplotypes from the 43 varieties for which complete sequence information was obtained (Table 4). However two of these nine haplotype groupings each consists of two haplotypes that can be differentiated based on sequence information obtained from outside this region (i.e. 11 haplotypes can be distinguished based on all sequence information).

The size of PCR amplification products containing the repeat region was determined against a standard for all 46 varieties using capillary electrophoresis. Generally sequence-based haplotypes were monomorphic for microsatellite repeat length; however, haplotype 1 was found to be composed of two repeat length alleles (310 bases and 312 bases, Table 7). The varieties Ulandra and Brindabella were found to be heterozygous for microsatellite allele size (Table 7). Two cases of homoplasy are evident where the amplified product was 301 bases (haplotypes 5, 6 and 8) and 310 bases (haplotypes 1 and 10, Table 7). The three additional haplotypes (haplotypes 9, 10 and 11) found based on sequence from all 43 barley varieties in this region increased the estimated minimum number of recombinations (calculated using RecMin) in the region flanking the microsatellite from two to four.

Discussion

This study has shown that in a sample of modern barley cultivars there is limited sequence diversity at the *Isa* locus, with most cultivars possessing an identical or nearly identical haplotype. Across the whole length (2,164 bp) of the *Isa* gene region, 9 out of 11 cultivated varieties were identical in sequence. A similar lack of diversity in modern cultivars was found when a larger number (36 modern cultivars, five wild barleys, two land races) were compared across a smaller region (280 bp) of *Isa* gene sequence. Greater haplotype diversity was however found in a small sample of wild barleys from Israel. These results indicate that for this region of the genome, a single ancestor has been the basis of selection for most of the germplasm utilised for barley production. Such declines in genetic diversity for domesticated varieties versus wild relatives are well documented for many crop species, e.g. soyabean (Zhu et al. 2003).

Reflecting a probable history of selection against deleterious mutations, the frequency of SNPs and indels within the three regions of the *Isa* gene (upstream, transcript and downstream) was unequal. There were fewer SNPs and indels than expected in the transcribed region, but more SNPs than expected (based on the overall frequency) in the 3′ downstream region. There were, however, no indels and only one SNP in the 139-bp 3′ UTR, although there are anecdotal observations that in general, 3′ UTRs have a higher frequency of polymorphism than coding regions (Bhattramakki et al. 2002; Kota et al. 2001). A high frequency of SNPs and indels in the region flanking a microsatellite repeat (as observed in the *Isa* promoter) has also been observed for microsatellite sites in maize (Mogg et al. 2002) and soyabean (de Barros et al. 2000).

The patterns of polymorphism in the genotypes surveyed in this study indicate a history of recombination at the *Isa* locus, which has contributed to haplotype diversity. Effective recombination, which is dependent on heterozygosity, must be much rarer in predominantly selfing species such as the wild progenitors of cultivated barley. One study of wild barley indicated an outcrossing rate of 1.6%, with consequently low heterozygosity (Brown et al. 1978). Lin et al. (2002) used this observation to calculate that the effective recombination rate in *H. spontaneum* would be expected to be 50-fold lower compared with an obligate outcrossing species. Thus effective recombination would be expected to be rarely observed in barley haplotypes. In line with this expectation, Cummings and Clegg (1998) found no evidence for recombination in a study of the *adh1* locus (1,362 bp) sequenced in 45 wild barley accessions. However when *adh2* (1,980 bp) was studied in 25 of the same barley accessions, a minimum of two recombination events were identified from the haplotypes (Lin et al. 2002). The putative *adh3* locus (1,873 bp) was also studied in this set of germplasm (Lin et al. 2001) and found to consist of two clusters of haplotypes with evidence of interallelic recombination between these clusters. Also in a study of 22 loci from chromosome 1H of barley sequenced in up to four varieties, Collins et al. (2001) found sequence structures that suggest a history of recombination events in at least three loci. In the present study from the eight haplotypes identified by sequencing the entire *Isa* region, there is evidence based on sequence structures of a minimum of seven recombination events having occurred in these haplotypes with a further two events found in haplotypes identified in the microsatellite flank (nine recombinations in all). Based on all the studies, there would appear to be considerable variation in the number of detectable recombination events between loci, even within the same sample of germplasm and in the case of the *adh1* and *adh2*genes, between closely linked loci.

Results from studying human genome sequences indicate that haplotypes tend to possess a block structure due to the non-uniform distribution of recombination events, i.e. the occurrence of recombination hot spots (e.g. Daly et al. 2001; Patil et al. 2001). Recombination hotspots have previously been identified in a number of species that have been the subject of thorough genetic study, such as yeast (*Sacharomyces cereviseae*), fungi (e.g.

Neurospora), maize, nematode (*Caenorhabditis elegans*) and mice (Lichten and Goldman 1995). There may be selection for recombination hotspots in the vicinity of defence genes in particular, since this would help maintain genetic variability in the face of selective sweeps centred on defence loci. Since the *Isa* gene has a likely role in defence, it may be associated with recombinationally rich sequences which it has acquired as a result of selection. Differences in the observed frequency of recombination events between barley loci may thus be a product of the sequence or regional dependence of recombination, with the *Isa* gene possibly representing a recombination hotspot.

To explain observed recombination events (and low LD) in *adh2* in wild barley, Lin et al. (2002) suggest that wild barley (and hence modern cultivated barleys) may be derived from predominantly outcrossing ancestors relatively recently in its evolutionary history. In *Arabidopsis thaliana* (which has an estimated outcrossing rate of 0.3%), Kuittinen and Aguade (2000) have suggested that heterosis in rare progeny from outcrossing with subsequent selection for recombinant individuals might explain the preservation and extent of recombinant haplotypes that they observed in nucleotide sequences from seven gene regions. This scenario suggested for *Arabidopsis* may also be relevant to the ancestors of wild and cultivated barley.

In conclusion, this study indicates that there is low diversity at the *Isa* locus amongst cultivated barley compared with wild barley. Such reductions in genetic diversity are commonly found in other domesticated crop species. A history of recombination at the *Isa* locus has increased the haplotype diversity and may indicate that this locus is a hotspot for recombination. This may be due to selection for recombinationally rich sequences at the *Isa* locus possibly due to its likely role in the seeds defence against pathogens.

Acknowledgements We would like to acknowledge the contribution of Linh Nguyen for carrying out sequencing and allele sizing of the microsatellite region and Giovani Cordeiro for running RecMin.

References

Abe J, Sidenius U, Svensson B (1993) Arginine is essential for the α-amylase inhibitory activity of the α-amylase/subtilisin inhibitor (BASI) from barley seeds. Biochem J 293:151−155

de Barros G, Tingey S, Rafalski JA (2000) Sequence characterisation of hypervariable regions in the soyabean genome: leucine-rich repeats and simple sequence repeats. Genet Mol Biol 23:411–415

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC, Tingey SV, Rafalski A (2002) Insertion–deletion polymorphisms in 3′ regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 48:539–547

Brown AHD, Zohary D, Nevo E (1978) Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. Heredity 41:49–62

Collins NC, Lahaye T, Peterhansel C, Freialdenhoven A, Corbitt M, Schulze-Lefert P (2001) Sequence haplotypes revealed by sequence-tagged site fine mapping of the *Ror1* gene in the centromeric region of barley chromosome 1H. Plant Physiol 125:1236–1247

Cummings MP, Clegg MT (1998) Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc Natl Acad Sci USA 95:5637–5642

Daly MJ, Rioux JD, Schaffner SE, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229–232

Furtado A, Henry RJ, Scott KJ, Meech S (2003) The promoter of the *Isa* gene directs expression in the maternal tissues of the seed in transgenic barley. Plant Mol Biol 52:787–799

Hejgaard J, Bjorn SE, Nielsen G (1984) Localization to chromosomes of structural genes for the major protease inhibitors of barley grains. Theor Appl Genet 68:127–130

Henry RJ, Battershell VG, Brennan PS, Oono K (1992) Control of wheat α-amylase using inhibitors from cereals. J Sci Food Agric 58:281–284

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164

Jarrett SJ, Marschke RJ, Symons MH, Gibson CE, Henry RJ, Fox GP (1997) α-amylase/subtilisin inhibitor levels in Australian barleys. J Cereal Sci 25:261–266

Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas 135:145–151

Kuittinen H, Aguade M (2000) Nucleotide variation at the chalcone isomerase locus in *Arabidopsis thaliana*. Genetics 155:863–872

Leah R, Mundy J (1989) The bifunctional α-amylase/subtilisin inhibitor of barley: nucleotide sequence and patterns of seed-specific expression. Plant Mol Biol 12:673–682

Lichten M, Goldman ASH (1995) Meiotic recombination hotspots. Annu Rev Genet 29:423–444

Lin JZ, Brown AHD, Clegg MT (2001) Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). Proc Natl Acad Sci USA 98:531–536

Lin JZ, Morrell PL, Clegg MT (2002) The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley (*Hordeum vulgare* ssp*spontaneum*). Genetics 162:2007–2015

Mogg R, Batley J, Hanley S, Edwards D, O'Sullivan H, Edwards KJ (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. Theor Appl Genet 105:532–543

Mundy J, Svendsen I, Hejgaard J (1983) Barley α-amylase/subtilisin inhibitor. I. Isolation and characterisation. Carlsberg Res Commun 48:81–90

Myers SR, Griffiths RC (2003) Bounds on the minimum number of recombination events in a sample history. Genetics 163:375–394

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, et al (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723

Rodenburg KW, Vallee F, Juge N, Aghajari N, Guo XJ, Haser R, Svensson B (2000) Specific inhibition of barley α-amylase 2 by barley α-amylase/subtilisin inhibitor depends on charge interactions and can be conferred to isozyme 1 by mutation. Eur J Biochem 267:1019–1029

Sancho AI, Faulds CB, Svensson B, Bartolme B, Williamson G, Juge N (2003) Cross-inhibitory activity of cereal protein inhibitors against α-amylases and xylanases. Biochim Biophys Acta 1650:136–144

Vallee F, Kadziola A, Bourne Y, Juy M, Rodenburg KW, Svensson B, Haser R (1998) Barley α-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 angstrom resolution. Struct Fold Des 6:649–659

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134