Theor Appl Genet (1992) 84:739–746

THEORETICAL AND APPLIED GENETICS

*Hpa*II library indicates 'methylation-free islands' in wheat and barley

W. Y. Cheung, G. Moore, T. A. Money, and M. D. Gale

Cambridge Laboratory, John Innes Centre, Colney Lane, Norwich, NR4 7UJ, UK

Received August 24, 1991; Accepted December 19, 1991
Communicated by J. W. Snape

Summary. A library of wheat genomic DNA *Hpa*II tiny fragments (HTF), sized below 500 bp, has been constructed. Of the clones in the library 80% belong to the single/low-copy category, while 12% of the clones are nuclear repetitive sequences and 8% originate from the chloroplast and mitochondrial DNA. This result shows a substantial enrichment in the single/low-copy sequences of the wheat genome, which contains at least 80% repetitive sequences. Twenty-nine random single/low-copy clones were analysed further for wheat chromosome location, cross-hybridisation to barley DNA and their association with rare-cutting, C-methylation-sensitive restriction sites. The results show that the HTF clones are associated more frequently than expected with *Not*I, *Mlu*I, *Nru*I and *Pst*I sites in wheat and barley genomic DNA. The 12% repetitive fraction of the clones contain both moderately and highly repetitive sequences, but no tandemly repeated sequences. The level of enrichment for single/low-copy sequences indicates that libraries of this type are a valuable source of probes for RFLP mapping. In addition, the close association of the HTF clones with rare-cutting restriction enzyme sites ensures that HTF clones will have a useful role in the construction of long-range physical maps in wheat.

Key words: Wheat – Barley – Methylation-free islands – *Hpa*II tiny fragments – RFLP

Introduction

Many higher plant genomes are large and are among the most highly methylated of eukaryotes. Plant nuclear DNA is heavily methylated at cytosine residues (Shapiro 1976), and the level of 5-methyl-cytosine (m^5C) can account for up to 30% of total cytosine residues, which are distributed between the sequences m^5CpG and m^5CpXpG, where X is any nucleotide (Gruenbaum et al. 1981). The level of methylation at methylatable sequences in plants is, as in vertebrates, less than complete. The relative location of methylated and non-methylated CpG (and also CpXpG in plants) has been investigated using methyl-sensitive restriction endonucleases to identify clusters of non-methylated sites in both vertebrates and higher plants (Bird 1986; Gardiner-Garden and Frommer 1987; Antequera and Bird 1988). In vertebrates, a small fraction of non-methylated DNA (1% of the genome) was detected by its cleavage to short fragments with the C-methyl-sensitive restriction enzyme *Hpa*II (CCGG). These DNA fragments are referred to as '*Hpa*II tiny fragments' (HTF). They differ from the bulk of genomic DNA by being unmethylated at CpG and by having relatively high G+C contents due to the lack of CpG suppression, but they show no extensive sequence consensus. HTF are found to occur in discrete domains, usually 1–2 kb long, and are dispersed in the genome (Bird et al. 1985; Bird 1986). These regions have been variously named HTF islands, CpG islands or methylation-free islands (MFI). HTF islands have been demonstrated in higher plants, including tobacco and three monocots, maize, rye and wheat (Antequera and Bird 1988). Surveys of genes in vertebrates (Bird et al. 1985; Bird 1987; Gardiner-Garden and Frommer 1987; Brown 1988; Fulton et al. 1988; Gardiner et al. 1990) and in higher plants (Antequera and Bird 1988) have shown that many HTF islands are associated with genes and regions where transcription begins, and they are non-methylated irrespective of gene expression (Bird 1987). Since plant genomes and vertebrate genomes have under-

Correspondence to: W. Y. Cheung

gone independent evolution, it has been suggested that HTF islands have arisen independently due to the selection pressure within large genomes in which most DNA is never transcribed. HTF islands thus provide, potentially, a means of identifying regions of chromosome carrying active genes. This suggestion has been supported in vertebrates by evidence from nuclease hyper-sensitivity (Wolf and Migeon 1985), the binding of a mammalian methyl-CpG-binding protein (MeCP) to non-island methylated DNA (Meehan et al. 1989), and the different chromatin structure of CpG islands as the result of a low level of H1 histones and higher level of acetylation in H3 and H4 histones (Tazi and Bird 1990).

CpG dinucleotides constitute 3.9% of the wheat genome, which has a G+C content of 48%. The theoretical CpG content expected is 5.8% (Gruenbaum et al. 1981). In mammals, CpG dinucleotides only constitute 1% of the whole genome, while the expected occurrence is 3.9% (Setlow 1976), indicating CpG suppression in the non-genic regions and the lack of such suppression in genic regions (Bird 1986). In comparison, the wheat genome does not seem to show CpG suppression in the non-genic regions, but there are certainly methylation-free islands (MFI), which have similar characteristics to the 'CpG islands' of mammals except for the aspect regarding CpG suppression. Given the enormous size and the large amount of repetitive sequences in the genomes of many economically important cereals (e.g. wheat, barley and rye) and the possible association of methylation-free islands with functional genes, as indicated by Cheung et al. (1991) in an analysis of α-amylase (*α-Amy-1*) multigene family in wheat, sequences from such islands (the HTF fraction of the genome) might be a source of probes for identification of RFLP in the region of active genes and for long-range physical mapping of chromosomes. We report here the cloning and characterisation of the HTF fraction of the DNA from an addition line containing a barley chromosome in a wheat background. The alien-wheat genetic stock was used in order that a defined HTF fraction of barley could be compared with that of wheat. The use of these sequences as potential landmarks in the physical mapping of wheat and barley chromosomes is also described.

Materials and methods

Library construction and screening

Total DNA from 'Chinese Spring'/*Hordeum vulgare* cv 'Betzes 5H' (CS/5H) was digested with the restriction endonuclease *Hpa*II. CS/5H is an addition line containing chromosome 5H (barley chromosome 7) in a wheat background (Islam et al. 1981). Digested DNA was separated on a 1.2% agarose gel. HTF fragments of sizes less than 0.5 kb were isolated onto a piece of DEAE membrane (Schleicher and Schuell) and recovered from the membrane as described by Dretzen et al. (1981) for ligation into pUC18 plasmid. Five micrograms of HTF DNA and 3 μg of pUC18 previously digested by *Acc*I and treated with calf intestine phosphatase (Boehringer Mannheim) were ligated in a volume of 20 μl with 2 U of T4 DNA ligase (BRL) at 15 °C overnight. *E. coli* strain DH5α cells were transformed with the ligation mixture using the $CaCl_2$-heat shock method (Mandel and Higa 1970). Recombinant colonies with wheat inserts were selected on ampicillin and X-gal/IPTG plates. Individual white colonies were picked and stored in glycerol on micro-titre plates.

A random selection of colonies were picked for plasmid mini-preparation, and an equal amount of DNA from each plasmid was digested with *Pst*I and *Xba*I to release the insert, which was separated on a 1.2% agarose gel. These gels, on which the insert DNA was well separated from the plasmid DNA, were blotted onto Hybond N^+ membrane (Amersham) and probed with labelled total wheat DNA to determine the proportion of repetitive sequences to single/low-copy sequences in the library, and with chloroplast and mitochondrial DNA to estimate the presence of these sequences in the repetitive fraction of the library.

Plant material, restriction digestion, electrophoresis and Southern analysis

Leaf genomic DNA from euploid wheat, *Triticum aestivum* cv 'Chinese Spring' (CS) and the full set of CS nullisomic-tetrasomic (NT) aneuploid lines (Sears 1954), barley *Hordeum vulgare* cvs 'Betzes', 'Golden Promise', 'Captain' and *H. spontaneum* ('IPSR #2370') were digested with *Hin*dIII and *Eco*RI (NBL). Electrophoresis of these DNA digests (10 μg DNA per track), alkaline blotting and hybridisation were as described by Sharp et al. (1988) except for the use of Hybond N^+ membrane. Washing and autoradiography of the filters were as described previously (Cheung and Gale 1990). These blots were used to determine the chromosome locations of the cloned sequences. Exposure times of around 10 days were required.

Wheat (CS) and barley ('Betzes') DNA were also digested with *Eco*RI and double digested with *Eco*RI and one of the four C-methylation sensitive restriction enzymes *Not*I, *Mlu*I, *Nru*I and *Pst*I (NBL) in order to investigate the association of clone sequences with rare-cutting enzymes sites.

Results

Characterization of the HTF library

Copy number. A total of 192 random genomic clones from a library of approximately 4,000 clones were surveyed for copy number. When blots of *Xba*I/*Pst*I digests of these clones were hybridised to ^{32}P-labelled sonicated total wheat DNA and washed under high stringency (0.2 × SSC, 1% SDS at 65 °C), 18 clones (9%) appeared, from hybridisation intensity, to be highly repetitive, 22 (11%) moderately repetitive and 152 (80%) single/low-copy sequences. Of the 40 moderately to highly repetitive clones, 16 were found to have originated from either the chloroplast and/or mitochondrial DNA. This classification was subsequently confirmed by hybridisation analysis to genomic DNA.

Analysis of clones with single- or low-copy sequences

More than two-thirds of the cloned fragments in this library were smaller than 300 bp in length. Twenty-nine

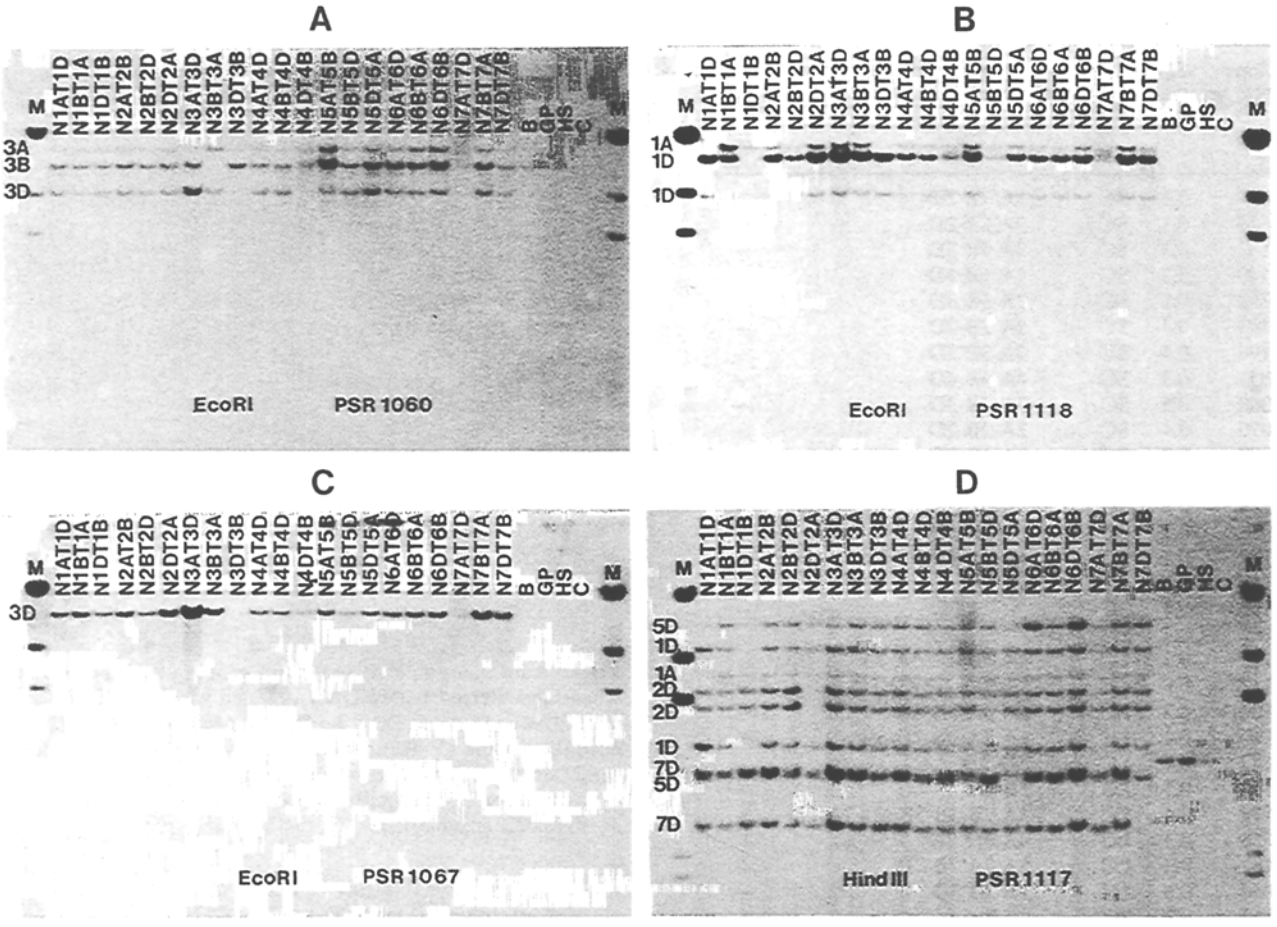

Fig. 1A–D. Autoradiographs derived from probing wheat nullisomic-tetrasomic and barley DNA with single-copy (**A–C**) and low-copy (**D**) genomic clones at high stringency (0.2 × SSC, 1% SDS, 65 °C). Four barley varieties are included in these filters: **B** 'Betzes', **GP** 'Golden Promise' **HS** *Hordeum spontaneum* ('IPSR #2370'), **C** 'Captain'. The chromosome locations in wheat of the hybridised fragments of each autoradiograph are shown on the *left* of the *λ-Hin*dIII size marker lane (*M*). (Note that the fragments missing in the N7AT7D digests are due to poor DNA digestions of DNA samples.) **A** *Eco*RI digests probed with PSR1060, a clone which detects single-copy sequences located on each of the group 3 chromosomes (note that N4DT4B probably carries a deletion for the region on 3D containing PSR1060); **B** *Eco*RI digests probed with PSR1118, which detects sequences located on only two of the group 1 chromosomes (note that the faint extra fragments observed only with N2BT2D, N6AT6D and N6BT6A digests are probably due to incomplete digestions of these samples); **C** *Eco*RI digests probed with PSR1067, which detects a single-copy sequence only on chromosome 3D; **D** *Hin*dIII digests probed with PSR1117, a low-copy clone located on more than one homoeologous group

randomly picked clones with single- or low-copy sequences 300 bp or more in length were analysed by Southern hybridisations using the set of nullisomic-tetrasomic lines to determine the chromosomal location(s) of the cloned sequences. A clone was classified as single copy if it hybridised to at most one fragment of each chromosome belonging to a single homoeologous group in at least one of the two restriction digests tested (e.g. *Eco*RI, *Hin*dIII) and if each hybridising fragment accounted for greater than 10% of the total signal in the track. Hybridisation to more than one fragment originating from chromosomes of different homoeologous groups resulted in classification as a low-copy clone. Eighteen out of the 29 clones tested belong to the single-copy category, 11 of which contain sequences that hybridised to a fragment from each of the three chromosomes within one homoeologous group (Table 1). An example of this class of clone (PSR1060) is shown in Fig. 1 A. Five clones hybridised to a fragment from two of the three homoeologous chromosomes within a group (e.g. PSR1118, Fig. 1 B), and three hybridised to a fragment from a single chromosome (e.g. PSR1067, Fig. 1 C) (Table 1). In the low-copy category, 4 clones were located on more than two homoeologous groups (e.g. PSR1117, Fig. 1 D), and 5 on more than one homoeologous group (Table 1).

Hybridisation to barley DNA was also tested with each of the 29 single/low-copy clones: 19 showed hybridisation to both wheat and barley ('Betzes'), 9 were wheat

Table 1. Characteristics of single/low-copy HTF clones

Clone PSR #	Size (kb)	Classification[a]	Wheat chromosome locations	Hybridisation to barley (Betzes)
1107	0.3	SC	7A 7B 7D	+
1114	0.3	SC	2A 2B 2D	–
1131	0.3	SC	7A 7B 7D	–
1140	0.3	SC	1A 1B 1D	+
1149	0.5	SC	3A 3B 3D	+
1193	0.4	SC	2A 2B 2D	+
1196	0.4	SC	3A 3B 3D	+
1051	0.3	SC	4A 4B 4D	+
1060	0.5	SC	3A 3B 3D	+
1070	0.4	SC	2A 2B 2D	–
1077	0.3	SC	3A 3B 3D	+
1118	0.3	SC	1A 1D	+
1194	0.3	SC	5B 5D	+
1056	0.3	SC	1B 1D	+
1179	0.5	SC	4A 4D	+
1157	0.3	SC	7A	–
1163	0.4	SC	3D	–
1067	0.3	SC	3D	–
1112	0.4	LC	4A 4B 4D 5A 7A 7B 7D	+
1117	0.6	LC	1A 1D 2D 4D 5D 7D	+
1101	0.3	LC	1A 5A 5B 5D	–
1160	0.4	LC	1A 2A 2B 2D 3B	+
1147	0.4	LC	3D 7D	–
1158	0.5	LC	3B 4B	–
1178	0.4	LC	1A 2B 3B 5B 6D 7A 7D	+
1058	0.3	LC	4A 4D 7A	+
1181	0.3	LC	1A 7A 7B 7D	+
1054	0.6	LC	2B 3B 3D	+
1153	0.3	SC	no hybridization	+

[a] SC, Single copy; LC, low copy
Total number of clones analysed = 29
Total number of clones hybridised to wheat and barley DNA = 19
Number of barley specific clones = 1

Table 2. Characteristics of highly/moderately repetitive nuclear HTF clones

Clone PSR #	Size (kb)	Classification based on hybridisation to wheat DNA	Classification based on hybridisation to barley DNA[a]
1104	0.5	M	M
1108	0.5	M	M
1132	0.2	M	–
1135	0.3	M	–
1137	0.5	M	M
1143	0.3	M	M
1180	0.3	M	M
1198	0.4	H	H
1004	0.3	M	M
1017	0.3	M	M
1050	0.3	M	–
1062	0.3	M	M
1068	1.0	H	H

[a] H, Highly repetitive; M, moderately repetitive
Total number of clones analysed = 13

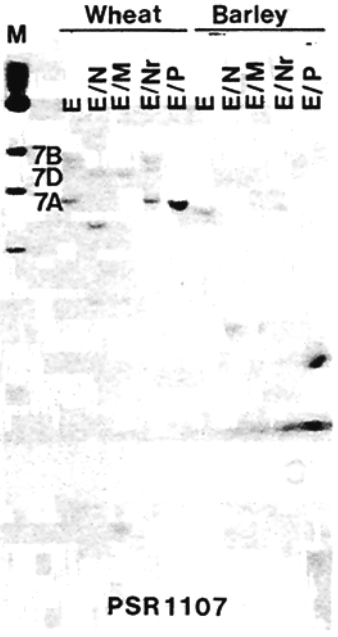

Fig. 2. Autoradiograph derived from probing 'Chinese Spring' wheat and 'Betzes' barley DNA digested with *Eco*RI (*E*), and with *Eco*RI and *Not*I (*E/N*), *Eco*RI and *Mlu*I (*E/M*), *Eco*RI and *Nru*I (*E/Nr*) and *Eco*RI and *Pst*I (*E/P*) with a single-copy clone (PSR1107) located on the group 7 chromosomes for analysis of the association of the *Hpa*II clone sequences with C-G enzymes. *M* λ-*Hin*dIII size markers

specific and 1 was barley specific. As the library was made with CS/5H DNA, the barley-specific sequence could only have come from the 'Betzes' 5H chromosome. This location was confirmed by hybridisation to 'Betzes' and the CS/5H single chromosome addition line (data not shown).

Analysis of clones with nuclear repetitive sequences

Thirteen clones, chosen at random from the 11% HTF repetitive fraction, were analysed by Southern hybridisation using blots with four different restriction digests (*Eco*RI, *Hin*dIII, *Eco*RV and *Dra*I) of both wheat and barley genomic DNA. A clone was classified as repetitive if more than 80% of the hybridisation signal was found in a continuous smear rather than discrete fragments. A further arbitrary sub-classification was based on exposure times. If the exposure time required was less than 1 day, the clone was classified as highly repetitive, and if more than a day was required the clone was classified as moderately repetitive (Table 2). There was a wide range in hybridisation patterns and copy number within a category. Most of the repetitive clones (10/13) hybridised to both wheat and barley DNA. Tandem repeats, which characteristically give rise to 'ladders' when hybridised to genomic DNA digested with restriction enzymes that recognise a variable site in the repeating unit, were absent in the repetitive nuclear HTF clones.

Table 3. Association of clone sequences with sites of other methylation-sensitive enzymes

Clone PSR #	*Eco*RI fragment(s) carrying sites of other methylation-sensitive enzymes							
	Wheat				Barley (Betzes)			
	*Not*I	*Mlu*I	*Nru*I	*Pst*I	*Not*I	*Mlu*I	*Nru*I	*Pst*I
1107	7A 7B 7D	7A 7B 7D	–	7A 7B 7D	+	+	–	–
1114	–	–	–	2B 2D	n/a	n/a	n/a	n/a
1131	7D	–	7B	7B	n/a	n/a	n/a	n/a
1140	–	–	–	1A 1B 1D	–	–	–	+
1149	–	3A 3B 3D	–	3A 3B 3D	–	–	–	–
1193	–	2A	2A	2A 2B	–	–	–	+
1196	–	3B 3D	–	3A 3B 3D	+	+	+	+
1051	–	–	–	4B 4D	+	+	+	–
1060	3A 3D	–	–	3A 3B 3D	+	–	–	+
1070	2B 2D	2B 2D	2B 2D	2A 2B 2D	n/a	n/a	n/a	n/a
1077	–	–	–	3A 3B	–	–	–	+
1118	–	–	1A 1D	1A 1D	–	–	–	+
1194	5B 5D	–	–	5B 5D	–	–	–	+
1056	1D	1B	–	1B 1D	+	–	–	+
1179	4B	4B	–	4B 4D	–	+	–	+
1157	–	–	–	7A	n/a	n/a	n/a	n/a
1163	3D	3D	–	3D	n/a	n/a	n/a	n/a
1067	–	–	–	3D	n/a	n/a	n/a	n/a
1112	4D	4B 4D 5A 7B 7D	4B 4D 7B	4B 5A 7A 7B	–	+	–	–
1117	1A 7D	5D 7D	–	1A 1D 2D 4D 5D 7D	–	–	–	–
1101	–	–	–	1A 5A 5B 5D	n/a	n/a	n/a	n/a
1160	–	1A	–	1A 2A 2B 2D	–	+	+	+
1147	–	–	–	3D 7D	n/a	n/a	n/a	n/a
1158	–	–	–	3B	n/a	n/a	n/a	n/a
1178	3B	–	–	2B 3B 5B 6D 7D	–	–	–	+
1058	–	4A 4D	–	4A 4D	–	–	+	–
1181	7B	1A 7A 7B 7D	7B	1A 7A 7B 7D	–	–	+	+
1054	2B 3B 3D	2B	2B	2B 3B 3D	+	–	+	+
1153	n/a	n/a	n/a	n/a	–	–	–	+

n/a, Not available due to absence of cross-hybridisation
Total number of wheat *Eco*RI fragments detected by the 29 clones analysed = 103
Total number of barley *Eco*RI fragments detected by the 19 clones cross-hybridised to barley DNA plus the barley specific clone AQ54 = 24
Total number of wheat *Eco*RI fragments detected by the same 19 cross-hybridised clones = 81

Association of single- or low-copy HTF clones with sites of other C-methylation sensitive enzymes

Restriction enzymes with recognition sites that are G+C rich and that contain one or more CpGs or CpXpGs (C-G enzymes) are expected to cut preferentially within HTF island DNA, rather than in inter-island DNA, due to the atypical sequence composition of the former (Brown and Bird 1986). The 29 single/low-copy clones listed in Table 1 were tested for their association (within 10 kb or less) with sites of four such enzymes, *Not*I (GCGGCCGC), *Mlu*I (ACGCGT), *Nru*I (TCGCGA) and *Pst*I (CTGCAG), by hybridisation of the cloned sequences to blots with *Eco*RI single digests of wheat and barley DNA, and double digests with *Eco*RI and one of the four C-G enzymes. The results are summarised in Table 3. For probe PSR1107, as shown in Fig. 2, it could be shown that all three wheat (7A, 7B and 7D) *Eco*RI fragments were cleaved by *Not*I, *Mlu*I and *Pst*I (tracks 3, 4 and 6), while none of the fragments contained an *Nru*I site (track 5). Two of the *Eco*RI/*Not*I fragments were the same size as the *Eco*RI/*Mlu*I fragments (compare tracks 3 and 4), thus making it probable that two of the *Not*I sites are close to the *Mlu*I sites. PSR1107 hybridised with a barley *Eco*RI fragment (track 7), which was cleaved by *Not*I and *Mlu*I, but not by *Nru*I or *Pst*I (tracks 8–11). Overall, about 22% (23/103) of the wheat *Eco*RI fragments detected by the 29 clones contained a *Not*I site, 35% (36/103) a *Mlu*I site, 12% (12/103) a *Nru*I site, and 85% (88/103) a *Pst*I site.

Nineteen out of the 29 clones hybridised with both wheat and barley DNA, and 1 clone (PSR1153) is barley specific. It was found that in the barley genome about 25% (6/23) of the cloned sequences are close to *Not*I

Table 4. Comparison of the occurrence of associated methylation-sensitive enzyme sites with the clone sequences in wheat and barley using 19 clones that hybridised to both barley and wheat DNA

Enzyme	Percentage of total fragments (detected by the 19 clones) associated with methylated-sensitive enzymes tested	
	Wheat (%)	Barley (%)
*Not*I	25	26
*Mlu*I	38	26
*Nru*I	11	26
*Pst*I	88	65

sites, 25% (6/23) close to *Mlu*I and *Nru*I sites, and 65% (15/23) close to *Pst*I sites. The results for wheat and barley for the 19 clones that co-hybridised to both species is summarised in Table 4. There appears to be little difference between barley and wheat in association of HTF with *Not*I and *Mlu*I sites, but there may be a higher incidence of *Nru*I sites near to the cloned sequences in barley and a lower incidence of *Pst*I sites close by compared to wheat.

Discussion

Wheat HTF library as a rich source of single- or low-copy sequences

Analysis of clones from the HTF library shows that this fraction of DNA in the wheat genome is extremely rich in single- or low-copy sequences. Considering that at least 80% of the wheat genome consists of repetitive sequences (Flavell 1986), HTFs are clearly not typical of the wheat genome as a whole. The HTF fraction from mouse has been cloned and analysed for single- or low-copy sequences (Bird et al. 1985). The results, which revealed that about 80% of the clones belonged to the single- or low-copy category, are comparable to the results reported here with wheat. Thus the HTF fraction of wheat DNA may be similar to that in vertebrate DNA. This lends strength to the suggestion of the independent evolution of 'methylation-free islands' in the proximity of genetically active regions in large genomes (Antequera and Bird 1988). Indeed, 13 of the 29 clones analysed detected positive signals in dot-blot analyses of leaf poly-A^+ RNA (W. Y. Cheung and T. A. Money, unpublished results), indicating that these clones are parts of transcripts of genes expressed in leaf.

HTF libraries may provide a means of generating low-copy clones of utility in RFLP studies, which will probe more and different HTF islands than those represented in *Pst*I genomic libraries. However, the sizes of inserts in the HTF clones are relatively small. This would not be expected to affect the level of polymorphism detected as shown by Miller and Tanksley (1990) in tomato, but has the disadvantage that small probes require a longer exposure time for autoradiography. Such drawbacks could be circumvented by making the library in a vector with SP6, T7 or T3 promoters around the cloning sites. This would allow probes to be labelled to high specificities with more than one type of radioactive nucleotide using transcription by SP6, T7 or T3 RNA polymerases (Melton et al. 1984).

Implications of the association of HTF fragments with sites of other C-G enzymes

Bird (1989) examined more than 30 kb of 'CpG island' DNA from human genes for the frequency of *Not*I, *Mlu*I and *Nru*I sites. He observed that *Not*I sites are present more frequently than expected at random, while *Mlu*I and *Nru*I sites occur less frequently than expected. The results obtained in this study for the occurrence of C-G enzyme sites in the proximity of HTFs indicate a different situation in cereals with an elevated incidence of each of these restriction sites. The average expected occurrence of *Not*I sites in the wheat genome is once every 600 kb, while *Mlu*I and *Nru*I sites are expected once every 300 kb (Moore et al. 1992). In wheat, *Mlu*I and *Not*I are found near HTFs with approximately 3 times and 2 times the frequency of *Nru*I sites, respectively, while in barley the three enzyme sites are found with similar frequencies.

*Pst*I (CTGCAG), also a C-methylation sensitive enzyme, is classified as a rare-cutter for plant DNA due to methylation at CpXpG residues. *Pst*I libraries have been a source of single- and low-copy probes used in RFLP mapping in tomato, maize and rice (Tanksley et al. 1987; Burr et al. 1988; McCouch et al. 1988), and in wheat, barley and rye (Devos et al. 1992; R. L. Harcourt unpublished results). The frequencies of cleavable sites of *Pst*I in both barley and wheat are higher than *Not*I, *Mlu*I and *Nru*I. There are *Pst*I sites in 85% of the wheat and 67% of the barley *Eco*RI fragments detected by the cloned HTF sequences. This suggests that the HTF DNA fraction in wheat or barley is closely associated with unmethylated *Pst*I sites. Indeed, the types of hybridisation patterns identified by HTFs when used to probe wheat DNA are similar to those observed by *Pst*I wheat genomic clones. The proportion of single/low-copy sequences (80%) is higher than that observed in *Pst*I libraries (50%) (Devos et al. 1992; R. L. Harcourt, personal communication), but this is likely to be simply a function of the smaller fragment size in the HTF library being less likely to include parts of the dispersed repetitive DNA that border the unique sequences in the regions of genes. The proportion of homoeologous clones, those present in sets across the three wheat genomes, to non-homoeologous

clones, those with copies on apparently unrelated chromosomes, is also similar in the HTF and *Pst*I genomic libraries. The significance of the latter clones is unknown, and is not accounted for by known evolutionary translocations (Naranjo et al. 1987; Liu et al. 1992). Although *Pst*I and HTF libraries are likely to sample similar regions of the genome, the latter are likely to include many more regions that are unclonable with *Pst*I or other rare methylation-sensitive C-G enzyme sites because of the greater number of *Hpa*II sites.

Conclusion

Wheat HTF DNA is a potentially rich alternative source of single- or low-copy sequences as markers to generate high density RFLP maps of plant genomes. The association of these sequences with rare-cutting C-methylation sensitive C-G enzymes also indicates that they will be useful markers for long-range physical mapping using C-G enzymes. The characteristics of the HTF library of wheat is similar to those of the analogous mouse library (Bird et al. 1985) in the proportion of single- or low-copy sequences to repetitive sequences, showing that although the two organisms have evolved quite independently and have very different genome sizes, the HTF fractions of the two genomes are remarkably similar. In the case of mammalian DNA, the HTF DNA is found to be derived from the 'methylation-free islands' that mark the genic regions of the genome. The plant HTF analogs probably play a similar role. Further surveys of wheat and barley gene sequences for the occurrence of under-methylated CpG-rich regions and the frequency of sites cleavable by C-G enzymes and *Hpa*II are likely to confirm their potential role as landmarks for transcribed genes during chromosome walking in wheat and barley.

Acknowledgements. We wish to thank Dr. R. M. D. Koebner for advice on the manuscript. This publication results from work undertaken in the Plant Gene Expression programme, which is supported by the DTI, AFRC and four companies as part of a LINK programme.

References

Antequera F, Bird AP (1988) Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J 7:2295–2299

Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209–213

Bird AP (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347

Bird AP (1989) Two classes of observed frequency for rare-cutter sites in CpG islands. Nucleic Acids Res 17:9485

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40:91–99

Brown WRA (1988) A physical map of the human pseudoautosomal region. EMBO J 7:2337–2385

Brown WRA, Bird AP (1986) Long-range restriction site mapping of mammalian genomic DNA. Nature 322:477–481

Burr B, Burr RA, Thompson KH, Albertsen MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Cheung WY, Gale MD (1990) The isolation of high molecular weight DNA from wheat, barley and rye for analysis by pulse-field gel electrophoresis. Plant Mol Biol 14:881–888

Cheung WY, Chao S, Gale MD (1991) Long-range physical mapping of the *α-amylase-1* (*α-Amy-1*) loci on homoeologous group 6 chromosomes of wheat. Mol Gen Genet (in press)

Devos KM, Atkinson MD, Chinoy CN, Gale MD (1992) RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor Appl Genet 83:931–939

Dretzen G, Bellard M, Sassone-Corsi P, Chambon P (1981) A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal Biochem 112:295–298

Flavell RB (1986) Repetitive DNA and chromosome evolution in plants. Philos Trans R Soc London B 312:227–242

Fulton TR, Bowcock AM, Smith AR, Deneshvar L, Green P, Cavalli-Sforza U, Doris-Keller H (1988) A 12-megabase restriction map at the cystic fibrosis locus. Nucleic Acids Res 17:271–284

Gardiner K, Horisberger M, Kraus J, Tantravahi U, Korenberg J, Rao V, Reddy S, Patterson D (1990) Analysis of human chromosome 21: correlation of physical and cytogenetic maps; gene and CpG island distribution. EMBO J 9:25–34

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196:261–282

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:860–862

Islam AKMR, Shepherd KW, Sparrow DHB (1981) Isolation and characterization of euplasmic wheat-barley addition lines. Heredity 46:161–174

Liu CJ, Atkinson MD, Chinoy CN, Devos KM, Gale MD (1992) Non-homoeologous translocation between group 4, 5 and 7 chromosomes within wheat and rye. Theor Appl Genet 83:305–312

Mandel H, Higa A (1970) Calcium dependent bacteriophage DNA infection. J Mol Biol 53:159–162

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated DNA. Cell 58:499–507

Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR (1984) Efficient in vitro synthesis of biological active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res 12:7035–7056

Miller JC, Tanksley SD (1990) Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. Theor Appl Genet 80:385–389

Moore G, Cheung W, Foote T, Gale MD, Koebner RMD, Leitch AR, Leitch IJ, Money T, Stancombe P, Yano M, Flavell RB (1992) Key features of cereal genome organization as revealed by the use of cytosine methylation-sensitive restriction endonucleases. Genomics (in press)

Naranjo T, Roca A, Goicoechea PG, Giraldez R (1987) Arm homoeology of wheat and rye chromosomes. Genome 2:873–882

Sears ER (1954) The aneuploids of common wheat. Univ Mo Agric Exp Stn Res Bull 572:1–58

Setlow P (1976) Nearest neighbor frequencies in deoxyribonucleic acids. In: Fasman GD (ed) Handbook of biochemistry and molecular biology, 3rd edn, vol II. CRC Press, Cleveland, Ohio, pp 312–318

Shapiro HS (1976) Distribution of purines and pyrimidines in deoxyribonucleic acids. In: Fasman GD (ed) Handbook of biochemistry and molecular biology, 3rd edn, vol II. CRC Press,Cleveland, Ohio, pp 258–262

Sharp PJ, Kries M, Shewry PR, Gale MD (1988) Location of β-amylase sequences in wheat and its relatives. Theor Appl Genet 75:286–290

Tanksley SD, Miller J, Paterson A, Bernatzky R (1987) Molecular mapping of plant chromosomes. In: Gustafson JP, Appels RA (eds) Chromosome structure and function. Plenum Press, New York, pp 157–173

Tazi J, Bird A (1990) Alternative chromatin structure at CpG islands. Cell 60:909–920

Wolf SF, Migeon BR (1985) Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as control elements of housekeeping genes. Nature 314:467–469