Theor Appl Genet (1992) 84: 535–543

Moderately repeated, dispersed, and highly variable (MRDHV) genomic sequences of common wheat usable for cultivar identification

Y.-G. Liu *, T. M. Ikeda, and K. Tsunewaki

Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Sakyo-ku, Kyoto 606-01, Japan

Received December 30, 1991; Accepted January 9, 1992
Communicated by G. Wenzel

Summary. A 4.1-kb DNA clone (pTag546), which when used as a probe produces hypervariable DNA fingerprints in common wheat, was found among the genomic clones of *Triticum aestivum* cv 'Chinese Spring'. Nulli-tetrasomic analyses revealed that the sequences hybridizing to this clone were located at 12 loci on ten chromosomes of the A, B, and D genomes of common wheat. The complete nucleotide sequence of pTag546 was shown to have a transposable element-like structure within it, though no open reading frame was detected. The sequences located in the A and D genomes were assumed to have been derived from the B genome by transposition. Using this clone as a probe, we were able to identify 56 common wheat cultivars, some of which are closely related, by their DNA fingerprints. This suggests that pTag 546 will be useful for cultivar identification as well as for germ plasm evaluation in wheat.

Key words: Wheat – *MRDHV* sequences – *DNA* fingerprint – Transposable element – Cultivar identification

Introduction

Since the identification of cultivars or lines of different crops is important for the documentation of genetic resources and the protection of the plant breeder's rights, it is necessary to improve upon the older and sometimes inaccurate usage of morphological and/or physiological characteristics. Because morphology and physiology are more or less influenced by environmental factors, they are not always reliable criteria. To complement these criteria in wheat, attempts have been made to utilize isozyme markers, but because of the small number of the isozyme loci available and the low level of allelic variation at most loci, this method has been of no great value. However, restriction fragment length polymorphisms (RFLPs) can be used for this purpose because of an extremely large number of potential RFLP loci (Soller and Backmann 1983).

Theoretically, it is expected that even closely related cultivars can be distinguished from one another if large numbers of probes and restriction endonucleases are used in combination. If probes that hybridize to hypervariable sequences are available, one or a few of them may be sufficient to distinguish among individual cultivars. Indeed, such hypervariable probes have been isolated from the human genome (Jeffreys et al. 1985a) and M13 phage (Vassart et al. 1987), both of which belong to the minisatellite DNA family and contain many arrays of short tandem repeats. The Southern hybridization patterns produced by these hypervariable probes, called DNA fingerprints, are individual specific in humans (Jeffreys et al. 1985b) and animals (Wetton et al. 1987) and cultivar specific in plants (Dallas 1988; Rogstad et al. 1988). Thus, DNA fingerprinting provides a new means for identifying any one plant cultivar out of many. So far, however, hypervariable probes have not been isolated from plant genomes, the exception being the spacer sequences between rRNA genes (Appels and Dvorak 1982).

In this paper we report the discovery of a hypervariable clone, pTag546, from a genomic DNA library of common wheat and a description of its molecular characteristics. Further, we illustrate how the fingerprints produced with this clone can be used to identify common wheat cultivars.

* *Present address:* Mitsui Plant Biotechnology Institute, Tsukuba Center Inc, Sengen 2-1-6, Tsukuba, Ibaraki 305, Japan
Correspondence to: K. Tsunewaki

Materials and methods

Plant materials

Eight common wheats and a durum wheat (see the legend of Fig. 1) and 56 common wheat cultivars (Table 2) were tested for DNA fingerprints. A set of the nulli-tetrasomic lines of *Triticum aestivum* cv 'Chinese Spring' (CS) produced by Sears (1966) was used to determine which chromosomes carry the sequences that hybridize with pTag546. Five diploid species of *Triticum* and *Aegilops* (see the legend of Fig. 5) were also tested for sequences homologous to pTag546.

DNA preparation, Southern hybridization and DNA fingerprint recording

The procedures for preparing genomic DNA from wheat plants and for digesting DNA with restriction endonucleases, and the agarose gel electrophoresis and Southern hybridization procedures have been previously described (Liu et al. 1990). The hybridization was conducted overnight at 66 °C in a mixture of 5 × SSC, 0.1% sodium *N*-lauroyl sarcosinate, 0.02% sodium dodecyl sulfate (SDS), 0.5% Blocking Reagent™ (Boehringer Mannheim), and 5% dextran sodium sulfate, containing ^{32}P-labelled probe at about 5×10^6 dpm/ml. The final wash was carried out under the stringent condition of 0.2 × SSC and 0.1% SDS at 65 °C for 40–50 min.

Hybrid band patterns of all cultivars were scored as figure patterns, e.g., 101100101101011, with "1" for the presence and "0" for the absence of each band, and the collected data were analyzed using a computer program developed by Liu (unpublished results).

Estimation of copy number per genome of sequences hybridized with pTag546

A known quantity of CS DNA and different quantities of pTag546 DNA were spotted onto Hybond-N membrane, and each was hybridized to a known quantity of ^{32}P-labelled pTag546 DNA. The copy number (n) of sequences homologous to pTag546 DNA per haploid genome was estimated according to the following formula;

$$n = \frac{i}{j} \times \frac{1.45 \times 10^7}{4.1},$$

where 4.1 and 1.45×10^7 are the size of pTag546 and the haploid genome size of common wheat in kb, respectively, and i and j are the known quantities of loaded DNA of pTag546 and CS, respectively.

DNA sequencing

The complete nucleotide sequence of pTag546 was determined by the dideoxy chain-termination method with Sequenase (Version 2.0, 7-deaza-dGTP edition, United States Biochemical) or AmpliTaq™ (Takara Shuzo), using the universal and sequence-specific primers, respectively. Deletions of the insert generated by exonuclease III/mung bean nuclease were also used for sequencing. The sequence data were analyzed using DNASIS (Hitachi Software) and IDEAS (Integrated Database and Extended Analysis System for Nucleic Acids and Proteins, Kyoto University).

Coefficient of relationship and similarity of DNA fingerprints between cultivars

As a parameter of relatedness between pairs of cultivars, a coefficient of relationship that is given by the following formula was used:

$$x = \sum_{i=1}^{m} (1/2)^n,$$

where m is the number of pathways for transmission of an allele, and n is the number of cultivar generations between cultivars in the i^{th} pathway.

The similarity in DNA fingerprints (y) was given by the proportion of common fragments between the two cultivars probed with pTag546.

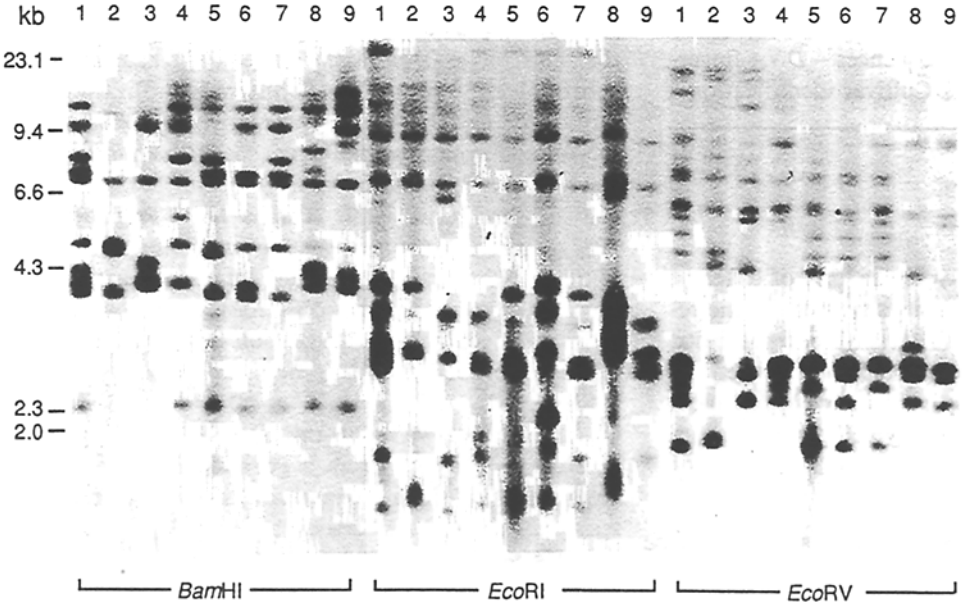

Fig. 1. Southern hybridization patterns of genomic DNA of eight common wheats and an emmer wheat, probed with pTag546. *Lanes 1–9 T. aestivum* cv "Chinese Spring", *T. compactum* cv "No. 44", *T. spelta* var "*duhamelianum*", *T. aestivum* var "*erythrospermum*", *T. sphaerococcum* var "*rotundatum*", *T. aestivum* ssp. *tibetanum*, *T. aestivum* cv "S-615", *T. macha* var "*subletschchumicum*", *T. durum* var "*reichenbachii*".

Results

DNA fingerprinting of nine polyploid wheats using pTag546 as a probe

In order to develop RFLP linkage maps of common wheat chromosomes, a *Pst*I genomic library of CS was constructed with plasmid pUC119 as a vector. The clones obtained were used as probes to detect RFLPs among eight common wheats, including CS, and an emmer wheat (Liu et al. 1990).

Among these wheats the pTag546 clone produced extremely variable Southern hybridization patterns when it was hybridized to their total DNAs, which were digested with five restriction endonucleases, *Bam*HI, *Dra*I, *Eco*RI, *Eco*RV, and *Hin*dIII. The DNA fingerprints of the nine wheat accessions clearly differed from one another after digestion by all of the five endonucleases (Fig. 1).

On the basis of the fingerprint data, the probabilities that two accessions share a single hybrid fragment or all of the hybrid fragments in common were calculated for each restriction endonuclease (Table 1). The results indicate that the probability is on the order of 10^{-2} for most endonucleases, which means that with one or two enzymes it is possible to distinguish among more than 100 cultivars. With pTag546 as probe, *Eco*RV and *Bam*HI are more effective than the other three enzymes as a means for identifying cultivar specificity.

Molecular characteristics of pTag546

The nucleotide sequence of pTag546 was determined (Fig. 2). A total of 4093 bp make up the primary structure of pTag546; the GC content is 45.6%; a 2159-bp region, from position 1783 to 3942, was found to have a pair of short (6 bp), inverted repeats with a 1 base mismatch at the distal ends. This region was flanked by a pair of 5-bp direct repeats, AGGGC. The structure of this particular region can be shown as follows: 5'-AGGG*CTGTTGG*...(2147 bp)...*CCAAGA*AGGGC-3' (the inverted repeats are indicated in italic).

In this 2159-bp region five tRNA-like sequences were also detected. Two sequences of 132 bp and 140 bp were found to have 80% homology and to be tandemly arrayed near the middle of the region. The IDEAS search indicated that about a 60-bp TC/AG stretch in each of these tandem repeats shows high homology (80–88%) to as many as 33 different genes reported in human, mouse, and other mammals. In addition, hexaplicated tandem repeats of the sequence TAAGA were found.

Chromosome mapping of the sequences probed with pTag546

The carrier chromosomes of the sequences homologous to pTag546 were determined by nulli-tetrasomic analysis.

Table 1. Probability of 1 or all fragments being found in common between two common wheat cultivars

Restriction enzyme	Average number of fragments per cultivar	Probability of having one fragment in common	Probability of having all fragments in common
*Bam*HI	8.3	0.58	1.0×10^{-2}
*Dra*I	9.2	0.71	4.3×10^{-2}
*Eco*RI	10.5	0.70	2.4×10^{-2}
*Eco*RV	14.3	0.67	3.3×10^{-3}
*Hin*dIII	9.5	0.72	4.4×10^{-2}
Average	10.4	0.67	1.5×10^{-2}

These homologous sequences are located on ten chromosomes, 5A, 6A, 7A, 2B, 3B, 6B, 7B, 2D, 4D, and 5D, distributed among the three genomes of CS (Fig. 3). To determine whether the different fragments found on the same chromosome are within identical or different gene loci, their linkage relationships were analyzed using an F_2 progeny of the cross, CS × *T. spelta* var '*duhamelianum*' (Spelta in abbreviation) (Liu and Tsunewaki 1991). Two of the 3 fragments on chromosome 5A revealed complete linkage, but the third fragment showed recombination with both. One of the 2 fragments located on chromosome 6B was present in both parents, whereas the second fragment was nullimorph in Spelta, indicating a different gene locus. In total, 12 loci homologous to pTag546 were identified on ten chromosomes. It is worthwhile to mention that 6 of 7 RFLP loci were nullimorphic in either CS or Spelta.

The copy number of sequences homologous to pTag546 per haploid genome of CS was estimated by a dot-blot analysis. The intensity of the hybrid signal of 400 μg CS DNA was between 1.2 and 1.8 pg, the intensity of pTag546 DNA (Fig. 4). Inserting these values to the formula given in Materials and methods, we estimated the copy numbers to be between 10 and 15. This indicates that most of the 12 loci identified above contain only a single copy of the sequence.

Distribution of sequences homologous to pTag546 within the putative ancestors of common wheat

Southern hybridization patterns probed with the pTag546 of one emmer wheat (4*x*, genome constitution AABB), six einkorn wheats (2*x*, AA), and one each of *Aegilops speltoides* (2*x*, SS) and *Ae. squarrosa* (2*x*, DD) accessions were compared with that of common wheat (CS, 6*x*, AABBDD) (Fig. 5).

All of the einkorn wheat accessions and the *Ae. squarrosa* accession produced only two to three hybrid bands, mostly with weak signals, when probed with pTag546. On the contrary, *Ae. speltoides* produced a large number of hybrid bands with strong signals, almost comparable to those observed in polyploid wheats.

```
CTGCAGCCTCGTCTATTGACGGACTTGTAATGATTGTGGGCACCATCCTTGTGTCATCTACACTTGATGCATGCACCATC      80
CTTAATATCGGTGTAAACGAACTTTTACTATCCACTGTATTGAGTAACCACTGCTTTCCATGACTACTTTGACAATCCCA     160
TGTTTATCTCAAATATATGTGAGTATATAGTTAAAAATACCTCTTATTCTCACCAGAACGGCGAAGTACAAACTGAGCCC     240
ATGAACTAAATTCTAGATTATGCAAAACAATTTAAGATCACATATGTTTCAGATTTAAGAAGAATATGCACCTATGTTAT     320
CAATATTAAAAAGCAGAAATACCAAGCAACCCTTAGGGAAAGAAGGCTCGTATCTTATAAAACATTCTTCCCCTGTATAC     400
TCTTCATTGAATTTATCAATTAATCCACATTCTCATCCATGAACGCATGCATCATATGAGAATTTCAGCAAGCACAGATT     480
AGCTAAATAAAAATCCAACTACGTCCACATAACAAGTTAGGTCGAAAGAGAAGACACGACACATGCGGCCATCGGATTGA     560
GTTTTCTCGAACATGTCCGGTCCTGTGGCAGGTCGACCGACCTGTAATGATTGTTGGCCCACCCCTATCTCATCTACACC     640
AACCCCACATTGTGTGTAATCTTTTTTTCATAGGGTGCAATTGTTCAGGTCTAGAGTCATGCAATTGTTGCTAGATAAAT     720
TTGAATGAAATTACTAAAGGTCTTACATGATGAAGCTCGCCGGCGAGGGATCGAATTCATTGAGAATGTCACAGACGATT     800
CCTACGCATCTTCCATCCTATCTGAATCATCCGAGCCTAAATCTGATGGGTAGTTCACACGACAAACAGTTGATACAATA     880
AACAGAAAACAGTAACAACAAAGCATCATATAGTAGCAGCAAAAATCACATTCCTTCTCTCTTATTTTTTCTCATAATAG     960
CAGCAAATCCGAAACCCGTCCTAACTCAGGAGAGTATCTTATAGTGGCCGCCGCCTTGTTGGCACATAGCCTAAACCAAC    1040
CGTAGCACACCAGCCCCAAGCACTGATCCAGTTTCTCTAATAAATTAAAAATTAAACTGTAATAGGGATGTGGCAAAGTC    1120
AGTAGCGGTGGCATCAACCTGGGGATTGGCCTTGAGGCGCGGCGTTATAAAGCTTGATTTTTTTTTGCTTCACCTGTAAT    1200
GCATGAAAAATATCAGTTCATTCACCAAGACTTGCTCACGTTTCATCATGGGATATTTTCTTTTGATGCTCATATGTGCT    1280
CATGAATTATAGAAATCAACAGGAGCTTAAGGGCGTCGATTCAGGGATACAGACCAAAAAAGCAGACCAAACGGTACATT    1360
AAGCAACATTACAAATTGATTGTAAAAACAACATACATCCACATATATCGGGATACTGGCCAAAAAGCAAGTTAAACAAT    1440
TTACACTACTTGAACACACCATATTGATTGTGAAAATGCAGTCCAAGCCGTACAATTCCATTCTAATATATAACTGAAAA    1520
AGAAGATTTTCAAACCATGTATGTACAGTTGTACACTACTGCAACAGTACAGACTCCATTCATACAAATACTGCAACACT    1600
CGTCTTGTGGAAAAAATTACCATGGGCATATCCTTAGAGGAATAGGGATATCAACAGCTCGGGAATTGATTCTGGTGGAG    1680
AAGCACATGAGTTTGCTGCTGAGGAGCTGTATGGTGATGTCCGAGAAGAACAGCCGACCGACAGAGACTCGGATAGACGG    1760
AGCAGTTGTTCAATCAGAGGGCTGTTGGCAGCCAACCCAGGTGTAGCCGAGTGGACGTGAGATACAGTTCGAGTCCTTCA    1840
CCACGTAAATCTTCGGCAGAGGTTCCTTGAGGTCTGTCTGCACAAACACAAATCAGCCTGAACAATCCCTTTGTTGTCTT    1920
GAGCAGGAAAAGCAAAAACAATAGCCATGTCAATTTTTCAAAGTTGATTTCACACCCGGTGCAAAAGCATGGAGATTCCT    2000
TGTGGTTAACTCGTGCTTCGAAGAAAAAATTATCCCTGTTGCGTTCTCTTTTCGAAAGGTAACAACTTGGAACAATACTA    2080
AACACATCTTTTCACATGCTAAATGCATTTAGGGCCTCTCACATCCAATGCTGCATGAATCTCCTGCCTGATCATGTCCT    2160
TCTAACTGAAGTTTGATTTTACCCTGAAATATATGTGCACGGTGAGGACTGCACTAACTTTGTCGCCCCTGCTGCAATCA    2240
AATTTGTCAATATTCATAGACCATATTCCACAAATACTCTGACTTTCTAATGAGAATTAGAATTTAGGGGCCAAATACTT    2320
TCAGATTTGGGGAGAAAGTTATAGATAAATGTGTTGCCGTGAAGCCTGATGCCCTTTGTCTTGAAGCTCCTGTACTACAA    2400
AAAAAATATCAGTTCAGTGTTGTTCTAATCTAGACCGTGAATCATGTTGCTCCCTGCTGTTTTCTCGAGCAGGGACTTGT    2480
CCTTGTCGTCAGCCTATGGCTCTGGAAGGGAGAGAGGGAAGGAGGTCTCACCATTATATGAGTAGGAGGGTGCTGGAATC    2560
CATCTTGCGCAGTTCTCCTCCTCCAAGTGCAGGTCGCCCTGCCGCCGGCCGATCACCGGAGCGCCACCGCCAGCGACCGT    2640
CTCCGTCACCTTGCGCCCATTCTCCTTCTCCTTTCTTCCTCTCTTTCCTCTCTCCCTCATGCCTGTCTCTCTCTTTCTCT    2720
GGACGAGACGGAGCGCCGCTGCGAGCAAGCCCCCGCACCAGCCGCGAGATCCCTCCCTCATCTTGCGACACTTCTCCTTC    2800
TATGCCTTTCTCTCTTTCCTCTCCCCCTCACACCCTGTCTCTCTCTTTAACTGGAATGAGCAAGACATCGCCGCCGCCAC    2880
CGGAGCAAGCCGCCGCACCGGCCGGCAGATCCGCCGCGCCCGGATCCGCGCGTCTTCCTCGCGCTGTCGAACGCTTTATC    2960
ACTTTGAAAGGGCCGCGGGTTCAATACAGGAAACTACAGGGTTTCTTGCAAAAACAATAGTCAGAACGAAGGGTTTTCTC    3040
ACCTAATAAGGGACCGCGGGTTGAATATTAGAAAGTGCATGGCTTTTTTGCAAAAATGCGCGACGGACGACCAGAAGCAA    3120
TGCGTGGTTTATTAGTAGGTAAGATAAGATAAGATAAGATAAGATAAGATATACGAAGGCTTGAACACGTACGTAATTGC    3200
TCTGTTGATCTAGTTCGTATGTACGTACGGTGTGGTGTGACGAATTCGAGACAGCAGACACGTGTAAGTTCGCATAAGTT    3280
CCCTTTCTAGCCGTTTTTCAAATTTTGTTTGACGATATGATACGAAGATATTTCGTTGATTATTACATGGATATTTCGTT    3360
GATTTCATATTTGCCATTTTGCCTCATAATCGGAACAACAAACTGAAGATCACGCAAGGTTACAGGACCGTCTTGGTCAG    3440
CTAAGCATGCACGGGCACTGAACTCCATTCCAGCCGAAAGAAAAGGGCAGGTTAGACCGTCAGAGACGAGATGACATGCA    3520
TGCAGAGATTGTCCCTCCGATGGTGGCGCCGGGCCGCATACGCATCAGCGGACAAGCCCGGCCACAGCCAGCCAGCGAGA    3600
TCGCCGGAGCCCGCTGGTCACACTATCTTCCGCCATAAATTACCAGGCACTTCATCCATTAGAATCACACAGCTCGCACC    3680
ACAAGGCACATACACATAACACCAATAAAATCTCTTGGATCAATCGAGAAGATGGCCTCCTCCTCCAGGATGCTCACGGT    3760
GGTGGTGCTGGGGCGCTGGTCGCCGGTGCGATGTGCGCCGTGAGGGTGAAGTTGACGGTGGAGAAGGGGTCCGACAAAAA    3840
GAAGCTGGCGCTCAAGATCGACTACACAAGGCCAGGCGACAGCCTGTCAGAGGTGGAGCTCCGGCAGCACGGCTCAGAGG    3920
AGTGGCAGCCGTTGACCAAGAAGGGCGACGTGTGGGAGGTCTCGTGCTCCAAGCCACTGGTTGGCCCCTTCAACTTCCGC    4000
TTCTTGTCCAAGAATGGCATGAAGAACGTCTTCGACGAGGTCTTCTCCACCGATTTCAATTATTTATCAAGGATGATTGG    4080
TCTCCCCCTGCAG                                                                        4093
```

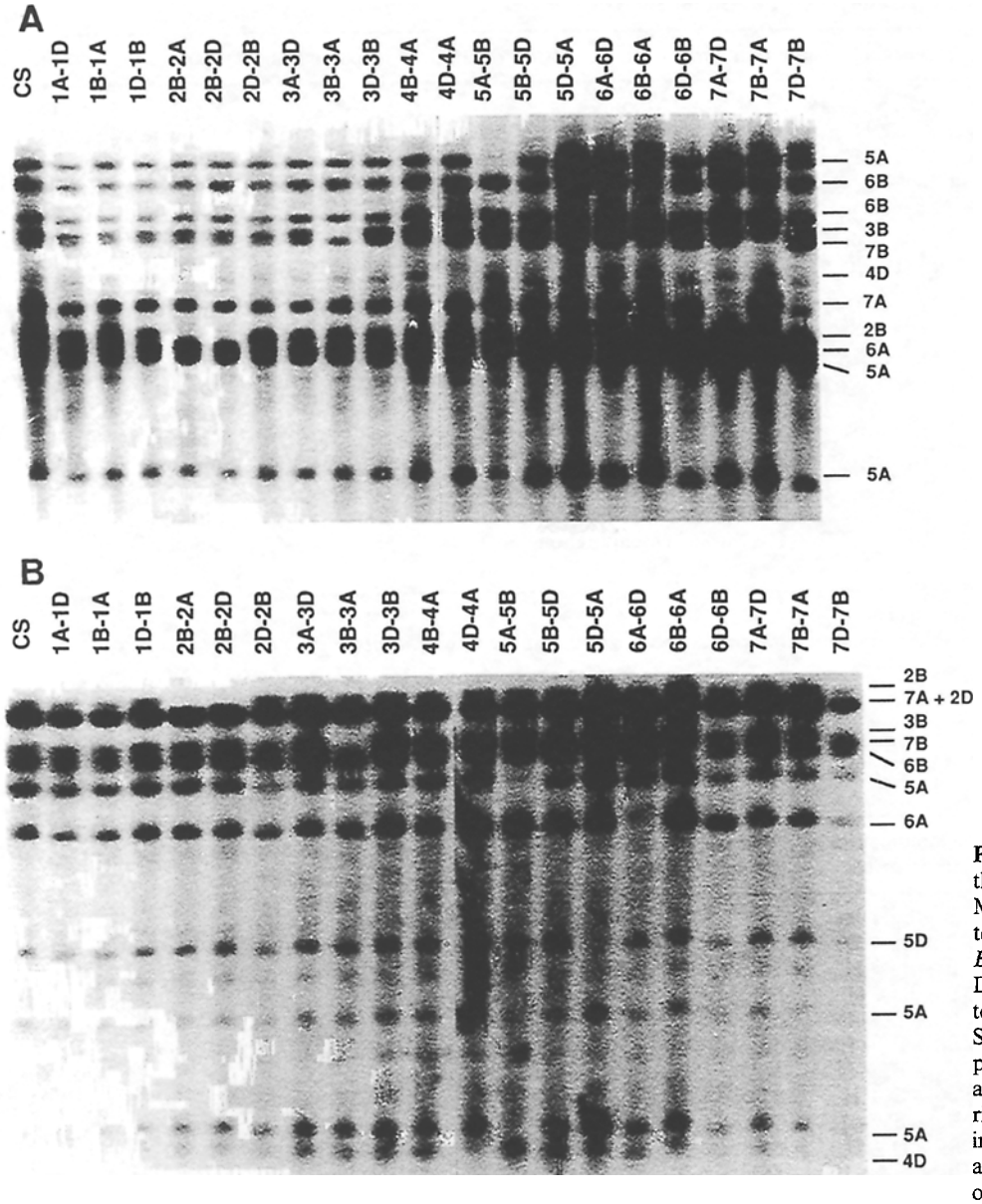

Fig. 3A, B. Determination of the chromosomes carrying the MRDHV sequences by nullitetrasomic analysis. **A** and **B** *Bam*HI and *Dra*I digests of DNA of normal and 20 nullitetrasomic lines of "Chinese Spring" were probed with pTag546 after electrophoresis and Southern transfer. The carrier chromosomes corresponding to the individual fragments are indicated in the *right margin* of each figure

Identification of common wheat cultivars using pTag546 as a probe

The usefulness of pTag546 for cultivar identification in common wheat was examined using 56 cultivars bred in different countries (Table 2). The pedigrees of 29 Japanese cultivars have been reported by Fukunaga and Inagaki (1985); these pedigrees indicate that many of them have very close relationships to each other (the coefficients of relationship are equal to or larger than 0.5).

The Southern hybridization patterns were obtained after digestion of genomic DNA with six restriction endonucleases, *Bam*HI, *Bgl*II, *Bst*I, *Dra*I, *Eco*RV, and *Hin*dIII, all 6-bp cutters, and probed with pTag546. Some examples of DNA fingerprints are shown in Fig. 6. The numbers of different patterns produced by these enzyme combinations are given in Table 3. The use of any single enzyme was not sufficient to distinguish among all of the cultivars, but as the number of enzymes increased to three all of the cultivars could be identified.

Fig. 2. Complete nucleotide sequence of pTag546. *Double underlines* 5-bp direct repeats (AGGGC); *short horizontal arrows with a large head* 6-bp inverted sequences flanked to the 5-bp direct repeats; *short horizontal arrows with a small head* six tandem direct repeats of 5 bp (TAAGA); *long horizontal arrows* long tandem repeats; *boxes* tRNA-like sequences; *broken overline* a TC/AG stretch showing high homology to a large number of genes reported in mammals, including humans

Table 2. Common wheat cultivars used for DNA fingerprint analysis

Code number	Cultivar[a]	Origin[b]	Code number	Cultivar	Origin	Code number	Cultivar	Origin
CS	Chinese Spring	C	19	Mikuni-k.	J	38	Jinghong 5	C
1	Akakomugi	J	20	Muka-k.	J	39	Nongda 139	C
2	Akasabishirazu	J	21	Nambu-k.	J	40	Bison	A
3	Aoba-k.	J	22	Nichirin-k.	J	41	Bledsoe	A
4	Asakaze-k.	J	23	Sakigake-k.	J	42	Contender	A
5	Chihoku-k.	J	24	Norin 29	J	43	Enigma	A
6	Chikushi-k.	J	25	Shirasagi-k.	J	44	Gaines	A
7	Ebisu-k.	J	26	Takune-k.	J	45	Kancom	A
8	Fujimi-k.	J	27	Ushio-k.	J	46	Knox 62	A
9	Hachiman-k.	J	28	Yutaka-k.	J	47	N. I. Triumph	A
10	Haruhikari	J	29	Norin 4	J	48	Penjamo 62	A
11	Haruminori	J	30	Norin 10	J	49	Rosetta	A
12	Hiyoku-k.	J	31	Norin 26	J	50	S-615	A
13	Horoshiri-k.	J	32	Norin 27	J	51	Wakeland	A
14	Hosokan	J	33	Shinchunaga	J	52	Warrior	A
15	Junrei-k.	J	34	Norin 52	J	53	Benno	E
16	Kitakami-k.	J	35	Norin 61	J	54	Perseus	E
17	Kobushi-k.	J	36	Dongfanghong 3	C	55	Salmon	J
18	Kokeshi-k.	J	37	Jing 1-38	C			

[a] k stands for "Komugi", meaning wheat
[b] A, C, E and J refer to the North American, Chinese, European and Japanese origin, respectively

Table 3. Number of groups into which 56 common wheat cultivars were classified based on DNA fingerprints with pTag546 as a probe

Combination of fingerprints[a]	Number of groups	Combination of fingerprints	Number of groups
*Bam*HI (B)	37	B + Bg + Bs	55
*Bgl*II (Bg)	22	B + Bg + D	53
*Bst*I (Bs)	31	B + Bg + H	55
*Dra*I (D)	31	B + Bs + H	55
*Eco*RV (EV)	35	Bg + Bs + D	55
*Hin*dIII (H)	29	Bg + Bs + EV	55
B + Bg	51	Bg + Bs + H	56[b]
Bg + Bs	53	Bg + D + H	54
Bg + H	51	Bg + EV + H	54
Bs + D	49	Bs + D + H	55
		Bs + EV + H	56[b]

[a] Fingerprints are indicated by restriction enzymes used for DNA digestion. Example: B + Bg refers to two DNA fingerprints produced by *Bam*HI and *Bgl*II digests
[b] All 56 cultivars can be distinguished by DNA fingerprints

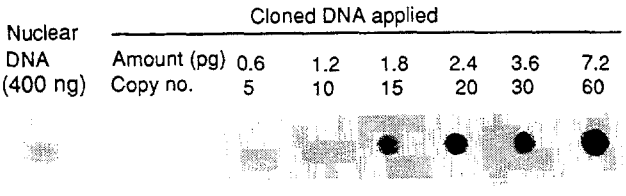

Fig. 4. A reconstruction experiment for estimating the copy number per genome of sequences homologous to pTag546. Hybrid signal intensity of the 400 ng nuclear DNA(2n genome size is 2.9×10^7 kb) was between those of 1.2 and 1.8 pg cloned DNA (size is 4.1 kb), suggesting 10–15 copies per genome

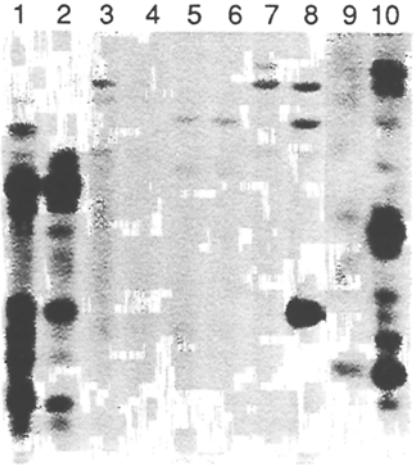

Fig. 5. Distribution of sequences homologous to pTag546 among the putative ancestors of common wheat. The genomic DNAs were digested with *Hin*dIII and probed with pTag546 after electrophoresis and Southern transfer. *Lanes 1-10 T. aesticum* cv "Chinese Spring" (nuclear genomes AABBDD), *T. durum* var "*reichenbachii*" (AABB), *T. boeoticum* ssp. *boeoticum* (AA), *T. boeoticum* ssp. *thaoudar* (AA), *T. urartu* var "*spontaneoalbum*" (AA), *T. urartu* var "*albonigrum*" (AA), *T. monococcum* var "*flavescens*" (AA), *T. monococcum* var "*vulgare*" (AA), *Ae. squarrosa* (DD), *Ae. speltoides* (SS ≒ BB)

The correlation coefficient (r_{xy}) between the coefficient of relationship (x) and similarity in DNA fingerprint (y) for 29 Japanese cultivars was 0.696, which is significant at the 0.1% level (Fig. 7). The fingerprint of each cultivar is undoubtedly determined mostly by its pedigree.

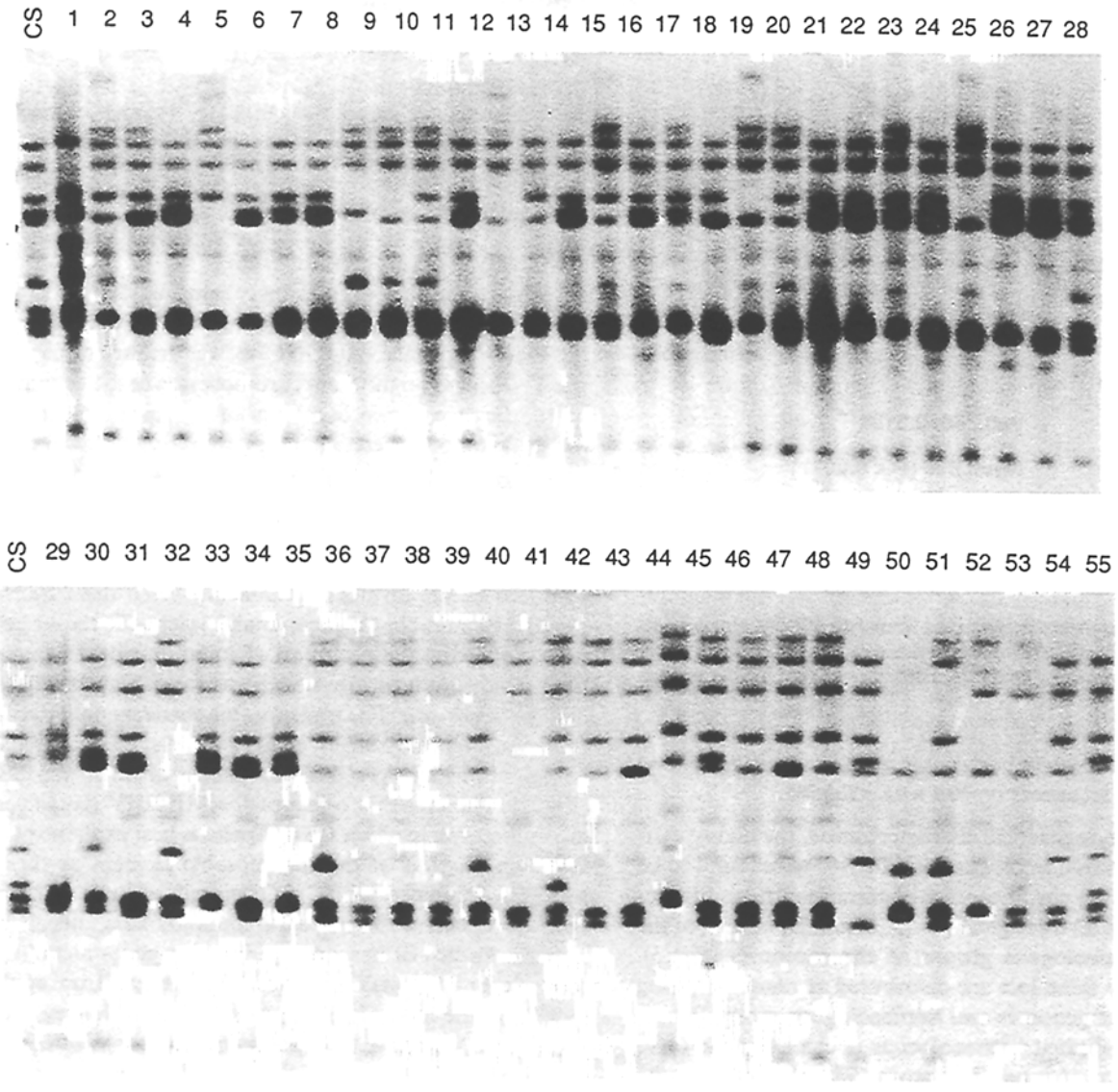

Fig. 6. DNA fingerprints of 56 common wheat cultivars. Genomic DNA was digested with *Bam*HI and probed with pTag546 after electrophoresis and Southern transfer. For code numbers of the cultivars, refer to Table 1

Discussion

Transposable element-like characteristics of pTag546

The complete nucleotide sequence of 4093 bp of pTag546 (Fig. 2) was revealed to have 6-bp inverted repeat sequences, TGTTGG and CCAAGA (1 base mismatch), located at the two ends of a 2159-bp region; this region is flanked by 5-bp (AGGGC) direct repeats at both ends. These are common structural features found among many transposable elements (Finnegan 1985); the 5-bp sequence is considered to be the target of a transposable element, and its direct repeats are assumed to have been created by the insertion of an element (Farabaugh and Fink 1980). A large number of transposable elements known in other organisms, including two cases, one in maize (Shepherd et al. 1984; Johns et al. 1985) and the other in wheat (Harberd et al. 1987; Flavell et al. 1988), have the inverted repeats of a consensus sequence, TGTTGG. . . . CCAACA, at their terminal ends, as does pTag546.

The sequences corresponding to the long terminal repeats (LTRs), present in many transposable elements, were not found in pTag546. The *Cin1* transposon family of maize has no such LTRs, but instead the entire structure of the element is similar to an LTR (Shepherd et al. 1984). These facts suggest that the 2159-bp region found in pTag546 originated from a transposable element, although the contemporary element is assumed to exist in a truncated form since it lacks open reading frames and one LTR.

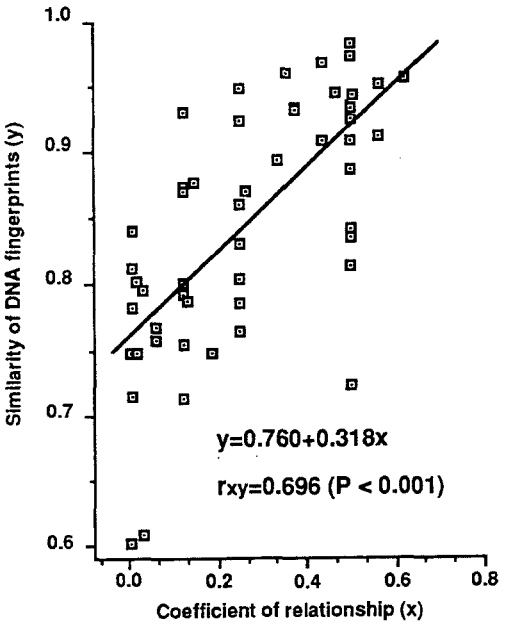

Fig. 7. Relationship between the similarity of DNA fingerprints (y) and the coefficient of relationship (x) observed in 29 Japanese common wheat cultivars. Estimates of the two parameters are given in the text

The probable origin and causes of high variability among the sequences probed with pTag546

By way of cytogenetical characterization, the highly variable sequences homologous to pTag546 were found to be located on ten chromosomes of common wheat, distributed among all three genomes (A, B and D) and six of the seven homoeologous groups of chromosomes (Fig. 3). Apparently these loci are distributed at random among the different genomes and homoeologous groups rather than as triplicated homoeoloci as is characteristic of so many structural genes (McIntosh 1988). In addition, the copy number per locus was estimated to be about one, that is, these sequences are moderately repeated, dispersed, and highly variable ('MRDHV' sequences in abbreviation). These sequences differ from the hypervariable sequences that are present within the minisatellite DNA family in the human genome, which consist of variable numbers of tandem repeats ('VNTR' in abbreviation) (Jeffreys et al. 1985a, b; Nakamura et al. 1987).

The three genomes (A, B and D) of common wheat are assumed to have been derived from einkorn wheat (most likely from *T. urartu*) (Kihara 1924; Dvorak 1988; Tsunewaki et al. 1991), *Ae. speltoides* (Sarkar and Stebbins 1956; Riley et al. 1958; Tsunewaki 1989), and *Ae. squarrosa* (Kihara 1944; McFadden and Sears 1944), respectively. The sequences homologous to pTag546 are distributed unevenly among the genomes of these putative ancestors, the greater number showing strong homology with sequences in the S genome (Fig. 5). On the contrary, 12 loci of MRDHV sequences were distributed almost equally among the three genomes of common wheat, i.e., 4, 5, and 3 loci in the A, B and D genome, respectively (Fig. 3).

In the case of the minisatellite DNA family, variability is caused by frequent changes in the number of tandemly arrayed repeating units. However, the cause of high variability among MRDHV sequences is expected to differ from that of the minisatellite DNA sequences since their structures are different.

We assume the following mechanisms for the origin and cause of high variability among MRDHV sequences: the target sequence (AGGGC) of the transposable element is located on different chromosomes of the genomes of the three putative ancestors of common wheat (all diploid). Among these ancestors only *Ae. speltoides* carried active transposable elements. Following the origin of the tetraploid progenitor (genome constitution AABB) by amphidiploidization of a hybrid between *Ae. speltoides* and *T. urartu*, the transposable elements were introduced by the former's genome and from there transposed first into the A genome and later into the D genome following the evolution of common wheat by amphidiploidization of a hybrid between an emmer wheat and *Ae. squarrosa*. The transposable elements underwent frequent internal changes, mainly by intramolecular recombinations between different classes of direct repeats (Fig. 2). The fact that 6 loci of the MRDHV sequences are nullimorph in either CS or Spelta is best explained by insertion/excision of the transposable element at these loci. The presence of TC/AG stretch regions in long tandem repeats was first found in plants here, although they are known of in many mammalian genes including the human U2 small nuclear RNA gene (Htun et al. 1985) and the mouse endogeneous provirus (Jeffrey and Robins 1988). These regions may have contributed to creating variability among the MRDHV sequences. At the same time, the insertion/excision of transposable elements may have resulted in structural changes in the flanking regions, as is known in maize (Schwarz-Sommer et al. 1985). The validity of the present hypothesis must be tested by analyzing the nucleotide sequences of MRDHV sequences in different loci.

Cultivar identification based on DNA fingerprints obtained using pTag546 as a probe

In order to distinguish large numbers of cultivars by individual DNA fingerprints, many probes will be required if they possess ordinary numbers of RFLPs. Soller and Backmann (1983) estimated that 10–20 polymorphic probes should be used for this purpose. In fact, Wang and Tanksley (1989) reported that only 58 of 70 rice cultivars tested could be identified using 10 RFLP probes combined with five restriction endonucleases.

The task has been simpler in our studies in which the pTag546 clone can be used to identify wheat cultivars by

detection of RFLPs of genomic MRDHV sequences. Based on the data shown in Table 1, we judged that two enzymes would be sufficient for discriminating among 100 cultivars, but in fact three enzymes were required to differentiate among 56 wheat cultivars. The discrepancy may be due to the fact that the Japanese cultivars tested are extremely closely related. At any rate, 56 bread wheat cultivars were distinguished using pTag546 as a probe in combination with three different endonucleases. The probe was also effective in differentiating among 50 accessions of two wild tetraploid wheats, *T. dicoccoides* and *T. araraticum*, using digests of but a single endonuclease (Mori et al. unpublished results). These results show that pTag546 is an extremely useful probe for cultivar identification among polyploid wheats; this DNA clone will be useful in the registration of new wheat cultivars, the evaluation of wheat germ plasm, and in the diagnosis of pedigree relationships among wheat cultivars.

Acknowledgments. Contribution No. 522 from the Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Japan. We are greatly indebted to Dr. T. Terachi, Kyoto Sangyo University, for his generous help in primer DNA synthesis, and Mr. Y. Nagano, Faculty of Agriculture, Kyoto University, for his help in the use of IDEAS. We also wish to thank Dr. H. Hirochika, National Agricultural Bioresources Research Institute, for his valuable advice on the present work and Dr. Val W. Wooward, University of Minnesota, for his endeavors to improve the manuscript. The work was supported in part by a Grant-in-Aid (No. 02404001) from the Ministry of Education, Science and Culture, Japan.

References

Appels R, Dvorak J (1982) The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. Theor Appl Genet 63:337–348

Dallas JF (1988) Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc Natl Acad Sci USA 85:6831–6835

Dvorak J (1988) Cytogenetical and molecular inferences about the evolution of wheat. In: Miller TE, Koebner RMD (eds) Proc 7th Int Wheat Genet Symp Cambridge, pp 187–192

Farabaugh PJ, Fink GR (1980) Insertion of eukaryotic transposable element Tyi creates a 5-base pair duplication. Nature 286:352–356

Finnegan D (1985) Transposable elements in eukaryotes. Int Rev Cytol 93:281–326

Flavell RB, Harris N, O'Dell M, Sardana RK, Jackson S (1988) Transposable elements and the control of ribosomal RNA gene expression in wheat. In: Miller TE, Koebner RMD (eds) Proc 7th Int Wheat Genet Symp Cambridge, pp 33–37

Fukunaga K, Inagaki M (1985) Genealogical pedigrees of Japanese wheat cultivars. Jpn J Breed 35:89–92 (in Japanese)

Harberd NP, Flavell RB, Thompson RD (1987) Identification of a transposon-like insertion in a *Glu-1* allele of wheat. Mol Gen Genet 209:326–332

Htun H, Lund E, Westin G, Pattersson U, Dahlberg JE (1985) Nuclease S1-sensitive sites in multigene families: human U2 small nuclear RNA genes. EMBO J 4:1839–1845

Jeffrey BS, Robins DM (1988) An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell 55:247–254

Jeffreys AJ, Wilson V, Thein SL (1985a) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73

Jeffreys AJ, Wilson V, Thein SL (1985b) Individual-specific 'fingerprints' of human DNA. Nature 316:76–79

Johns MA, Mottingger J, Freeling M (1985) A low copy number, *Copia*-like transposon in maize. EMBO J 4:1093–1102

Kihara H (1924) Cytologische und genetische Studien bei wichtigen Getreidearten mit besonderer Rücksicht auf das Verhalten der Chromosomen und die Sterilität in den Bastarden. Mem Coll Sci Kyoto Imp Univ Series B:1–200

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric Hortic Tokyo 19:889–890 (in Japanese)

Liu YG, Tsunewaki K (1991) Restriction fragment length polymorphism (RFLP) analysis in wheat. II. Linkage maps of the RFLP sites in common wheat. Jpn J Genet 66:617–633

Liu YG, Mori N, Tsunewaki K (1990) Restriction fragment length polymorphism (RFLP) analysis in wheat. I. Genomic DNA library construction and RFLP analysis in common wheat. Jpn J Genet 65:367–380

McFadden ES, Sears ER (1944) The artificial synthesis of *Triticum spelta*. Rec Genet Soc Am 13:26–27

McIntosh RA (1988) A catalogue of gene symbols for wheat. In: Miller TE, Koebner RMD (eds) Proc 7th Int Wheat Genet Symp Cambridge, pp 1225–1323

Nakamura Y, O'Connell P, Leppert M, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number of tandem repeat (VNTR) marker for human gene mapping. Science 235:1616–1622

Riley R, Unrau J, Chapman V (1958) Evidence on the origin of the B genome of wheat. J Hered 49:91–98

Rogstad SH, Patton JC II, Schaal BA (1988) M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. Proc Natl Acad Sci USA 85:9176–9178

Sarkar P, Stebbins GL (1956) Morphological evidence concerning the origin of the B genome in wheat. Am J Bot 43:297–304

Schwarz-Sommer Z, Gierl A, Cuypers H, Pererson PA, Saedler H (1985) Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J 4:591–597

Sears ER (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis KR (eds) Chromosome manipulations and plant genetics. Oliver and Boyd, Edinburgh, pp 29–45

Shepherd NS, Schwarz-Sommer Z, vel Spalve JB, Gupta M, Wienand U, Saedler H (1984) Similarity of the Cin 1 repetitive family of *Zea mays* to eukaryotic transposable elements. Nature 307:185–187

Soller M, Backmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Tsunewaki K (1989) Plasmon diversity in *Triticum* and *Aegilops* and its implication in wheat evolution. Genome 31:143–154

Tsunewaki K, Takumi S, Mori N, Achiwa T, Liu YG (1991) Origin of polyploid wheats revealed by RFLP analyses. In: Sasakuma T (ed) Proc Int Symp Wheat Cytoplasmic Engineer. Sapporo, pp 31–39

Vassart G, Georges M, Monsieur R, Brocas H, Lequarre AS, Christophe D (1987) A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. Science 235:683–684

Wang ZY, Tanksley SD (1989) Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113–1118

Wetton JH, Carter RE, Parkin DT, Walters D (1987) Demographic study of a wild house sparrow population by DNA fingerprinting. Nature 327:147–152