Genet. Res., Camb. (1995), **65**, *pp.* 11–19 *With 5 text-figures*

Molecular analysis of *opaque-2* alleles from *Zea mays* L. reveals the nature of mutational events and the presence of a hypervariable region in the 5′ part of the gene

HANS HARTINGS[1], NADIA LAZZARONI[1], VINCENZO ROSSI[1], GIORGIA R. RIBOLDI[1], RICHARD D. THOMPSON[2], FRANCESCO SALAMINI[2] AND MARIO MOTTO[1]*

[1] *Istituto Sperimentale per la Cerealicoltura, Sezione di Bergamo, via Stezzano 24, 24126 Bergamo, Italy*
[2] *Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, 50879 Köln, Germany*

(*Received 3 October 1994 and in revised form 21 November 1994*)

Summary

Ten recessive *Opaque-2* (*O2*) alleles of independent origin were characterized at the molecular level. The results revealed a high level of polymorphism at the *O2* locus. In addition, our data suggest the possible cause for the recessive character of some of the alleles investigated, and allow us to infer some conclusions concerning the degree of relationship between the *o2* mutations. Comparison of genomic sequences spanning the first exon and obtained from a series of wild-type and recessive alleles revealed the presence of a hypervariable region, involving different dipeptides, in the N-terminal part of the O2 protein.

1. Introduction

Genetics has played a crucial role in elucidating the mechanisms behind cell specific gene expression in maize. Studies on genetic mutations that influence the accumulation of zein proteins have indicated the presence of several regulatory mechanisms controlling the expression of specific members of the zein multigene family (see Motto *et al.* 1989, for a review). The *Opaque-2* (*O2*) gene is one of the most widely studied loci involved in these regulatory mechanisms. The *o2* mutation causes, in addition to a modification of endosperm appearance, a severe reduction (50–70%) in zein synthesis, leading to a concomitant enhancement of lysine content in the seed (Mertz *et al.* 1964). In plants homozygous for *o2*, the synthesis of the α-class zeins, and in particular the 22 kDa zeins, is primarily reduced (Soave *et al.* 1976; Jones *et al.* 1977; Burr *et al.* 1982). This reduction is at least in part attributable to a decreased transcription rate (Kodrzycki *et al.* 1989). The *o2* mutant is furthermore unable to synthesize a number of other endosperm polypeptides, the most abundant of which is a 32 kDa albumin, termed b-32, whose expression is temporarily and quantitatively co-ordinated with the deposition of zeins (Soave *et al.* 1981).

The *O2* gene has been isolated by transposon tagging (Schmidt *et al.* 1987; Motto *et al.* 1988) and encodes a polypeptide containing a b-ZIP (basic/leucine zipper) DNA binding motif, characteristic of many transcriptional activators (Hartings *et al.* 1989; Schmidt *et al.* 1990). This eukaryotic DNA binding motif consists of a periodic repeat of leucines (the zipper), which is responsible for dimer formation, adjacent to a cluster of positively charged amino acids (the basic motif), which are responsible for sequence specific recognition of the target DNA (Landschultz *et al.* 1988; Vinson *et al.* 1989). It has been shown that the O2 protein binds the b-32, and 22 kDa α-zein gene promoters and regulates their transcription (Lohmer *et al.* 1991; Schmidt *et al.* 1992). Furthermore, studies in yeast have indicated that the O2 protein can substitute for the GCN4 protein, a yeast transcriptional activator of amino acid biosynthesis genes (Hinnebusch, 1990). Similarly to GCN4, the presence of three short upstream open reading frames (uORFs) in the leader of the O2 mRNA suggests a possible mechanism of control at the transcriptional level (Lohmer *et al.* 1993).

A number of spontaneous *o2* mutations have been described in past years. For example, Montanelli *et al.* (1984) have studied the behaviour of nine of these alleles in the presence of an autonomous *Bg* transposable element. Four of the investigated alleles, that is *o2R*, *o2-Charentes*, *o2-Crow* and *o2-G*, did not respond to the presence of *Bg*, while the remaining five alleles responded to *Bg*, producing variegated opaque seeds. The unstable *o2* alleles, carrying the *rbg* receptor

* Corresponding author.

element at the *O2* locus, have independent origins and were identified in maize strains from the United States [*o2-m(r)*], Italy (*o2-mh*; *o2-261*) and South America (*o2-Columbian*; *o2-Agroceres*).

In this paper, we describe the molecular analysis of this series of *o2* mutations. We demonstrate that all *Bg* responsive mutants are independent isolates of the *o2-m(r)* allele. In addition, our data suggest the possible cause for the recessive character of some of the alleles investigated, and allow us to infer some conclusions concerning the degree of relationship between the *o2* mutations.

2. Material and methods

(i) *Plant material*

The following *o2* alleles were investigated; *o2R*, *o2-Charentes*, *o2-Columbian*, *o2-Crow*, *o2-Italian*, *o2-Agroceres*, *o2-261*, *o2-mh*, *o2-33*, *o2-G* and *o2-m(r)*. The *o2R*(eference) allele was originally isolated from an Italian maize strain as *o2-Italian* (Salamini, 1980). This allele has been used extensively in our laboratory as an *o2* reference allele. All *o2* mutants analysed were part of the maize stock of the Istituto Sperimentale per la Cerealicoltura and were present in an A69Y background. For DNA extraction, two-week-old greenhouse grown seedlings were pooled and harvested in liquid nitrogen.

(ii) *Enzymes and chemicals*

T4 DNA ligase, Taq polymerase, M-MuLV reverse transcriptase and restriction enzymes were obtained from Bethesda Research Laboratories; random priming kit, [α-^{32}P]dCTP (3000 μCi) and Hybond-N membranes were purchased from Amersham International; sequence kits were from Applied Biosystems Int. All enzymes and kits were used as indicated by the manufacturer. The phagemid pGEM3Zf(+) was purchased from Promega Biolabs. Oligonucleotides were prepared on an ABI 391E oligonucleotide synthesizer. The following oligonucleotides were constructed:

O2/ATG – GTTGGACATGGAGCCTG;
O2/TGA – GCTCTTAGGGTCTCCATGTCTGC;
Bg3 – CCTGTTTATCGTCAACCTCCG;
#882 – CCGAATTCTTGCAGGTAGTTCGTC;
#1131 – CCCATGGTACCTAACTTGGTACG;
#1463 – CACTAGGAATTCAAGGAGATGCAG;
#1908 – ACGGATGGGATGGAGTACAATGCC;
#1913 – TCCTGCAGTATGGCATTGTACTCC;
#2818 – AGCATGCATCAGGAATAATCCAGTGC;
#2905 – TAGTAGCAAGCCTCACGCC;

(iii) *Southern analysis*

Genomic DNA for Southern analysis was prepared as previously described (Motto *et al.* 1988). Six μg of DNA were digested for 4–6 hours with 100 U of restriction enzyme. Electrophoresis was performed on a 0·8% agarose gel in standard TAE buffer as described by Sambrook *et al.* (1989). After treatment of the gels with denaturation and neutralization solutions, the DNA was blotted to Hybond-N membranes according to Southern (1975). DNA was fixed on the membranes by baking for 2 h at 80 °C.

Prehybridization was performed for 6 h at 68 °C in a solution containing 5 × SSPE, 0·5% SDS, 5 × Denhardt's and 50 μg/ml denaturated calf thymus DNA. Hybridization was performed at 68 °C for 16 h using 5×10^7 cpm of probe labelled by the random priming method of Feinberg & Vogelstein (1983). The hybridization solution was identical to the prehybridization solution. Filters were washed in 2 × SSPE, 0·1% SDS for 20 min at RT and in 0·1 × SSPE, 0·1% SDS for 40 min at 65 °C. Blots were exposed to Kodak XAR-5 films for 16 h using intensifying screens. Filters were used several times with different probes. Probes were removed by incubating the filters for 10 min in a 0·1% SDS solution at 95 °C.

Probes were recovered from a plasmid carrying the full-length O2 cDNA (Hartings *et al.* 1989), and consisted of a fragment spanning the 5′ part of O2 up to the *Eco*R I restriction site (probe 1 in Fig. 1), and a fragment containing the 3′ part of O2 starting from the same *Eco*R I site (probe 2 in Fig. 1).

(iv) *RNA preparation and Northern analysis*

RNA was prepared as described by Dean *et al.* (1985). Poly(A) RNA was subsequently purified by oligo(dT) cellulose chromatography as described by Sambrook *et al.* (1989). For Northern analysis, 2 μg of poly(A) RNA were separated on a 1·3% denaturing agarose gel. Blotting, hybridization and washing procedures were as described for Southern analysis.

(v) *PCR amplification and cDNA cloning*

PCR amplification of genomic DNA was performed in a reaction volume of 100 μl containing 50 mM Tris HCl pH 8·3; 50 mM KCl; 1·5 mM $MgCl_2$; 250 μM each of dATP, dTTP, dCTP, and dGTP; 0·5 mM of 5′ oligo; 0·5 mM of 3′ oligo; 1·0 unit of Taq polymerase; 100 ng of genomic DNA were used as a template. Mineral oil was used to prevent evaporation of the samples at high temperatures. Amplification was obtained through 30 cycles with the following temperature profile: 94 °C for 30 s; 55 °C for 30 s; 72 °C for 1 min. Amplification products were loaded on a 0·8% agarose gel containing EtBr, purified by gel electrophoresis followed by electro-elution, and subjected to direct sequencing.

Reverse transcription of mRNA was performed using 1 μg of poly(A) RNA in a total reaction volume of 20 μl containing 50 mM Tris HCl pH 8·3; 75 mM KCl; 15 mM $MgCl_2$; 1 mM each of dATP, dTTP, dCTP, and dGTP; 15 units of RNA-sin; 200 units of M-MuLV reverse transcriptase; 1 μg of d(T)17 primer. After incubation at 37 °C for 1 h, the reaction mixture was used directly to amplify cDNA fragments. For

https://doi.org/10.1017/S0016672300032961 Published online by Cambridge University Press

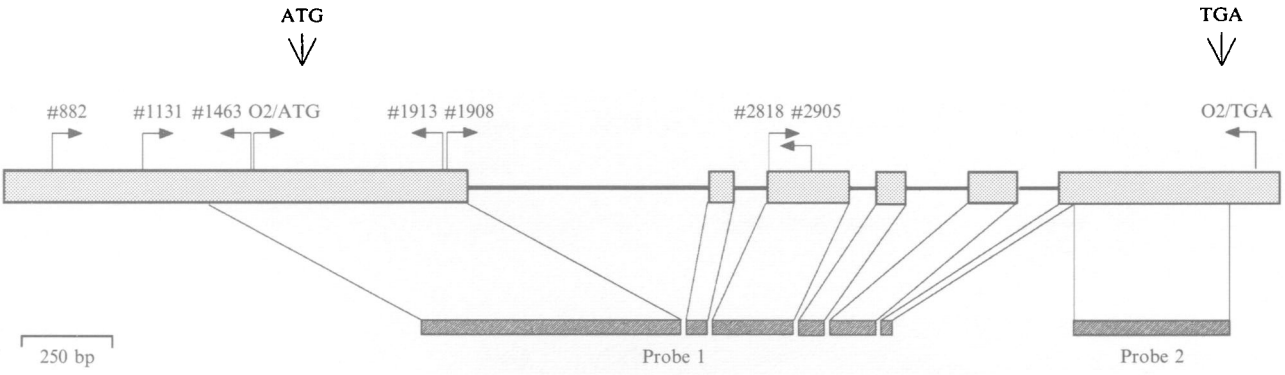

Fig. 1. Map of the *Opaque-2* allele. Intron sequences are indicated by a horizontal line; exon sequences are represented by filled boxes. The start and stop codons are shown. Position and orientation of oligonucleotides used for amplification are indicated by arrows. Fragments used as probes in Southern analyses are shown under the map.

this purpose, 1 μl of first-strand mixture was subjected to PCR amplification under the same conditions as described above, except for the extension phase at 72 °C which was increased to 2 min. Since several attempts to amplify the cDNA of *o2-Crow* with oligo #2818 failed, a new oligonucleotide (#1908), located 150 bp upstream of oligonucleotide #2818, was designed. The use of this oligonucleotide resulted in reliable PCR amplification of o2-Crow cDNA. Amplified fragments were gel purified, treated with T4 DNA polymerase and subcloned in *Sma* I cut pGEM3Zf(+) vector. In order to avoid artifacts introduced by the Taq polymerase, all fragments were isolated at least twice from independent PCR reactions.

(vi) *Sequence analysis*

Nucleotide sequences were determined on double-stranded templates. The Applied Biosystems 'Taq Dye Primer Cycle Sequence' kit was used to sequence cloned cDNA fragments. Multiple subclones of the same fragments were sequenced on both strands to avoid sequence ambiguities. Direct sequencing of amplified genomic fragments was performed with the use of the Applied Biosystems 'Taq Dyedeoxy Terminator Cycle Sequence' kit. Sequences were analysed on an Applied Biosystems model 373A automatic sequencer.

In all cases, sequences were verified by analysing several identical fragments, obtained from independent amplification reactions, in order to avoid any errors introduced by the Taq polymerase.

3. Results

(i) *Southern analysis*

Genomic DNA from 10 different A69Y inbred lines, each carrying independently isolated *o2* alleles, was cleaved with *Hin*d III endonuclease. *Hin*d III digests of genomic DNA from A69Y *o2R* and wild-type A69Y were used as controls. The digests were analysed with a molecular probe corresponding to a 1000 bp O2 cDNA fragment (probe 1 in Fig. 1). With the use of this probe, the mutable alleles were divided into four groups (Fig. 2A). As expected, *o2R* and *o2-Italian* carried a hybridizing fragment of identical size (12 kb). Moreover, *o2-m(r)*, *o2-Columbian*, *o2-Agroceres*, *o2-261* and *o2-mh* also possessed the same, 7 kb long, hybridizing fragment. The *o2-33*, *o2-G* and *o2-Crow* alleles carried a fragment identical in size to that found for the wild-type allele (16 kb). Finally, *o2-Charentes* contained a characteristic fragment of 17 kb, slightly larger than that of the wild-type allele.

Hybridization with probe 1 to *Eco*R I digests partly confirmed the data obtained from the *Hin*d III digest. With this probe, the *o2-33* mutant displayed an 8 kb hybridizing band, differing in size from the 9·5 kb fragment of the wild-type allele (Fig. 2B). Hybridization of the same digests with probe 2, corresponding to a 700 bp o2 cDNA fragment, revealed polymorphism between the *o2-G* and the wild-type allele, which displayed hybridizing bands of 10·5 and 5 kb respectively (Fig. 2C), fixing the number of polymorphic groups at a total of six. Results are summarized in Table 1.

(ii) *rbg-induced recessive alleles*

The *o2-m(r)*, *o2-Columbian*, *o2-Agroceres*, *o2-261* and *o2-mh* alleles were previously classified as responsive to the *Bg* transposable element (Montanelli *et al.* 1984), indicating in this way the presence of an *rbg* receptor element at the *O2* locus. Moreover, our results demonstrate that these alleles display identical patterns in Southern analysis. To investigate whether these mutants represent multiple isolates of the same allele, or actually constitute different genotypes, we performed PCR analyses on genomic DNA extracted from plants carrying the alleles of interest.

The *o2-m(r)* mutation is caused by the insertion of a non-autonomous *rbg* transposable element in the untranslated leader sequence of the *O2* gene (Hartings *et al.* 1991). Since the exact insertion point of the element has been determined, it is feasible to verify the presence of the *rbg* element by means of PCR analysis.

https://doi.org/10.1017/S0016672300032961 Published online by Cambridge University Press

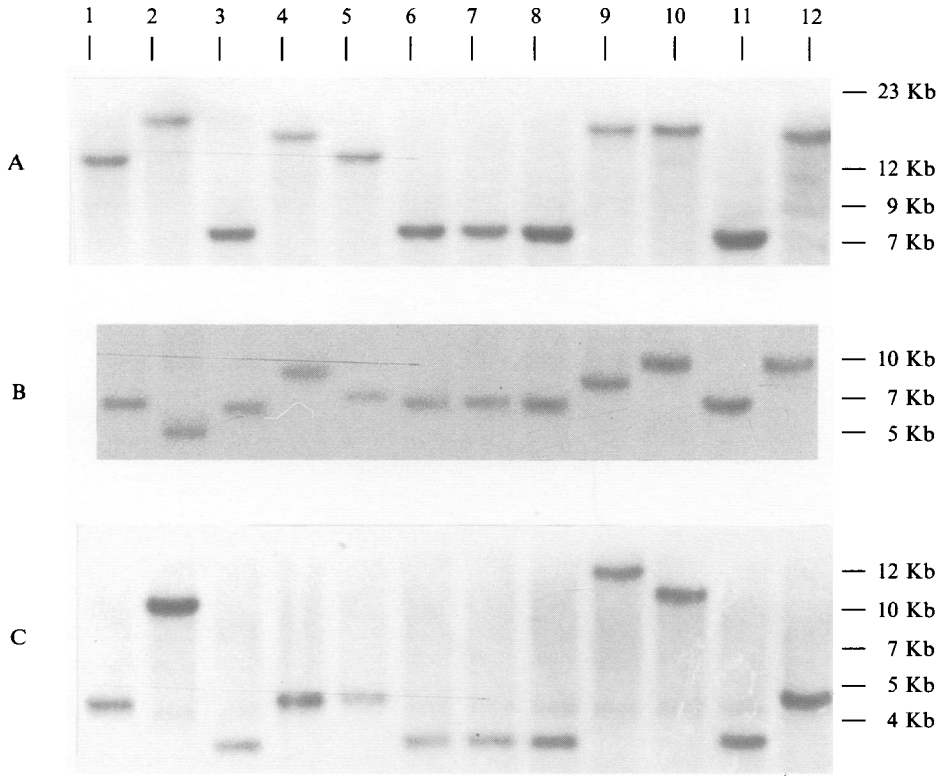

Fig. 2. Southern analyses. Molecular weights are indicated. (1) *o2R*, (2) *o2-Charentes*, (3) *o2-Columbian*, (4) *o2-Crow*, (5) *o2-Italian*, (6) *o2-Agroceres*, (7) *o2-261*, (8) *o2-mh*, (9) *o2-33*, (10) *o2-G*, (11) *o2-m(r)*, (12) wild type. (A) *Hind* III digests were hybridized with probe. 1. (B) *Eco*R I digests hybridized with probe 1. (C) *Eco*R I digests assayed with probe 2.

Table 1. *Summary of Southern analyses*

o2 mutant	*Hind* III-probe 1	*Eco*R I-probe 1	*Eco*R I-probe 2
wt[1]	16[2]	9·5	5·0
o2R	12	7·5	5·0
o2-Italian	12	7·5	5·0
o2-m(r)	7·0	7·0	4·0
o2-Columbian	7·0	7·0	4·0
o2-Agroceres	7·0	7·0	4·0
o2-261	7·0	7·0	4·0
o2-mh	7·0	7·0	4·0
o2-33	16	7·0	12·0
o2-G	16	9·5	10·5
o2-Crow	16	9·5	5·0
o2-Charentes	17	5·5	10·5

[1] Wild type.
[2] Fragment size in kb.

For this purpose, two oligonucleotides were constructed. The first (Bg3) is complementary to the subterminal 3′ end of the *Bg* element and directed towards the element's 3′ end, whilst the second (#1463) is complementary to the region of the *O2* gene flanking the *rbg* insertion point.

PCR amplification, performed on genomic DNA from A69Y *o2-m(r)*, with oligos Bg3 and #1463 gives rise to a characteristic 240 bp fragment (data not shown). PCR analysis carried out on genomic DNA from *o2-Columbian*, *o2-Agroceres*, *o2-261* and *o2-mh*, resulted in the amplification of the same sized fragment as that amplified from *o2-m(r)*. We therefore assume that the *o2-Columbian*, *o2-Agroceres*, *o2-261* and *o2-mh* alleles are independent isolates of the *o2-m(r)* allele.

(iii) *Northern analysis*

Poly(A) RNA was extracted from 15 DAP old endosperm of plants carrying the *o2R*, *o2-Crow*, *o2-m(r)*, *o2-Charentes* and *o2-G* alleles. Two μg of poly(A) RNA from each purification were separated on a denaturing agarose gel, together with 2 μg of poly(A) RNA extracted from 15 DAP old endosperms of A69Y wild-type plants. Hybridization of these poly(A)

https://doi.org/10.1017/S0016672300032961 Published online by Cambridge University Press

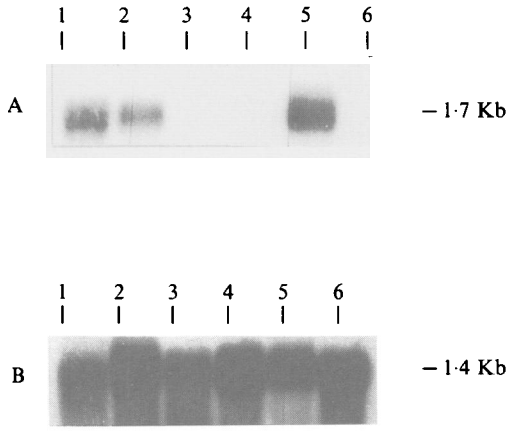

Fig. 3. Northern analysis. 15 DAP endosperm Poly(A) RNA from A69Y carrying the wild type [1], *o2R* [2], *o2-m(r)* [3], *o2-Charentes* [4], *o2-Crow* [5], or *o2-G* [6] allele was hybridized with a full-length O2 cDNA probe [A], or with a GAPDH probe in order to verify the presence of poly(A) RNA [B]. Molecular weights of hybridizing bands are indicated.

RNAs against a probe corresponding to a full-length O2 cDNA clone revealed the presence of hybridizing bands identical in size to the wild-type transcript in case of the *o2R* and *o2-Crow* alleles (Fig. 3).

(iv) *cDNA cloning and sequence analysis*

The o2 mRNA present in the *o2R* and *o2-Crow* genotypes was studied at the sequence level. For this purpose, four oligonucleotides (O2/ATG, O2/TGA, #2818, #2905) homologous to the O2 sequence were designed. The position of the oligonucleotides was chosen in such a way as to prime PCR amplification of two overlapping fragments (Fig. 1). The two overlapping fragments cover the complete O2 coding region, including approximately 90 bp upstream of the gene's start codon and 60 bp downstream of its stop codon.

Comparison of the cDNA sequence obtained for the *o2R* allele with the published O2 sequence

```
WT      MEVISMEEILGRFWELLPPPAPEPEPEPEPEREQPPVTGIVVGSVIDVA      50
Crow    ...........................--..................      48
o2R     ...........................----..................      46

WT      AAGHGDGDMMDQQHATEWT---FERLLEEEALTTSTPPPVVVVNSCCSG      97
Crow    ..................---.........................      95
o2R     .....H.GRRHDGS.ARHRVDL.........................      96

WT      ALNADRPPVMEEAVTMAPAAVSSAVVGDPMEYNAILRRKLEEDLEAFKMW     147
Crow    G.........G...........-.........................     144
o2R     ...V.........IEW................................     146

WT      RAASSVVTDDQRSQGSNNHTGGSSIRNNPVQNKLMNGEDPINNNHAQTAG    197
Crow    .---------------------------------------TWTEITEI.    154
o2R     ................................................     196

WT      LGVRLATSSSSRDPSPSDEDMDGEVEILGFKMPTEERVRKRKESNRESAR    247
Crow    ................................................     204
o2R     ..................GSRCLPRKE*                        225

WT      RSRYRKAAHLKELEDQVAQL                                267
Crow    ......PLT*                                          213
```

Fig. 4. Alignment of putative o2 protein sequences. All sequences were deduced from the respective nucleotide sequences. A partial sequence (first 267 amino acids) of the *O2* wild-type (wt) allele ending with the second leucine residue of the zipper domain is included for sequence comparison. Amino acid positions are shown on the right hand side of the sequences. Identical residues are marked by dots. Gaps, indicated by dashes, are introduced to maximize sequence alignment. Stop codons are indicated by asterisks. The O2 basic domain is distinguished by double underlining. Leucine residues of the zipper domain are in bold face.

https://doi.org/10.1017/S0016672300032961 Published online by Cambridge University Press

		hypervariable region	
A69Y	PP	PAPEPEPEPEPEREQ	PP
o2-Crow	PP	PAPEPEPEPE--REQ	PP
o2R	PP	PAPEPEPE----REQ	PP
W64A	PP	PAPEPE------REQ	PP
o2-G	PP	PAPEPEPE------Q	PP
B73	PP	PAPEPE--------Q	PP
Z.m.mexicana	PP	--PEPEPE------Q	PP
Z.m.huehuetenangensis	PP	PAPEPEPEPE--REP	PP

Fig. 5. Micro-satellite sequences present in the *O2* gene. Putative O2 protein sequences, deduced from the open reading frame found in the genomic sequences of different *O2* alleles, were aligned against the deduced O2 protein sequence of A69Y. Gaps, indicated by dashes, were introduced to optimize sequence alignment.

(Hartings *et al.* 1989), revealed 17 base substitutions. Moreover, the o2R sequence contained two insertions of respectively two and seven nucleotides, 41 bp apart. Since no stop codons are present in this sequence stretch, these insertions merely give rise to a short frame shifted stretch and consequent alteration of the deduced protein sequence in this region. Two deletions of 2 and 23 base respectively in the central region of the o2R sequence generate a second frame shift. This change in reading frame causes the premature termination of the coding sequence. As a consequence, the open reading frame of the *o2R* gene is only 255 codons long, with a deduced polypeptide devoid of one acidic region and the basic and zipper domains (Fig. 4).

Comparison of the o2-Crow and wild-type O2 cDNA sequences disclosed 27 base substitutions. In addition, the o2-Crow sequence contains five deletions which measure 3, 6, 120, 1 and 4 nucleotides with respect to wild type. Immediately following the 120 bp deletion, a 24 bp stretch devoid of homology with the O2 sequence is present in *o2-Crow*. Comparison of this nucleotide tract with the sequence of the first intron of *O2* reveals some homology. It is therefore possible that the 120 bp deletion was caused by a recombination event which has involved intron sequences. Whilst the first three deletions merely cause the omission of respectively 1, 2 and 40 residues from the deduced sequence, the fourth deletion generates a frame shift, which causes the premature termination of the o2-Crow polypeptide. As a consequence, the deduced o2-Crow protein is terminated after the basic domain (Fig. 4).

The nucleotide changes observed between the O2 and o2R sequences and between the O2 and o2-Crow sequences were used to estimate the average number of nucleotide substitutions per site according to Kimura's three substitution model (Kimura, 1981). The average distance calculated for the O2 and o2R sequence is $K = 0{\cdot}0106$, while *O2* and *o2-Crow* yield an average distance of $K = 0{\cdot}0186$. Taking these data, and considering a neutral nucleotide substitution rate of 5×10^{-9} substitutions/site/yr (Wolfe *et al.* 1987), the *o2R* and *O2* alleles should have diverged from a common sequence approximately 1 million yr ago. In a similar manner, *O2* and *o2-Crow* diverged approximately 1·86 million yr ago.

(v) *O2 contains a micro-satellite*

Comparison of the sequences from the first exon of *O2* revealed an interesting aspect of the O2 protein. The O2 protein sequence, as deduced from the A69Y cDNA sequence, carries a proline and glutamate rich stretch near the N-terminus. This region is 19 amino acids long and contains 10 proline residues (52·6%) and six glutamate residues (31·5%). Comparison of the deduced protein sequences from the cDNA O2 sequence (Hartings *et al.* 1989), the genomic O2 sequence (Maddaloni *et al.* 1989), and the cDNA sequences derived from the *o2R* and *o2-Crow* alleles, shows that this region is highly variable (Fig. 5). In fact, from 2 to 5 Pro-Glu repetitions could be identified in the deduced protein sequences. This observation prompted us to determine the extent of the variability in this region. For this purpose, DNA fragments corresponding to the first exon of the *O2* gene were amplified with the use of oligonucleotides O2/ATG and #1913, and subjected to sequence analysis. A total of 13 genes, comprising four recessive alleles (*o2-*

https://doi.org/10.1017/S0016672300032961 Published online by Cambridge University Press

Crow, *o2R*, *o2-G*, *o2-Charentes*) and nine wild-type alleles deriving from maize inbred lines (GM1417-MPI, A69Y, B37, W64A), a primitive Bolivian maize accession (Pisanckalla), and teosinte accessions (*Zea mays parviglumis*, *Zea mays huehuetenangensis* [two accessions], *Zea mays mexicana* [Nobogame]) were analysed. These 13 DNA sequences revealed a total of eight different amino acid sequences. Variability involved the number of Pro-Glu repetitions, a Pro-Ala block, and an Arg-Glu block (Fig. 5).

4. Discussion

The *Opaque-2* gene encodes a b-ZIP transcriptional factor that binds to the promoters and regulates the expression of both 22 kDa zein genes and the *b-32* gene (Lohmer *et al.* 1991; Schmidt *et al.* 1992). In order to acquire information on the mechanism of regulation exerted by *Opaque-2*, and to gain more insight into the regulation of synthesis of the O2 protein itself, ten different *o2* mutant alleles have been characterized at the molecular level.

Southern analysis performed on the *o2* mutant alleles has revealed a high level of polymorphism at the *O2* locus. Among the *o2* mutants analysed, six different restriction patterns could be distinguished.

The *o2-m(r)*, *o2-Columbian*, *o2-Agroceres*, *o2-261* and *o2-mh* alleles displayed identical patterns in Southern analysis. These alleles, which originated in maize strains from the United States, Italy and South America, have been previously analysed genetically. Since all displayed somatic mutability in the opaque phenotype in the presence of the *Bg* transposable element, the occurrence of a receptor element associated with the *O2* gene could be predicted.

Employing PCR analysis in combination of O2 and rbg-specific oligonucleotides, the occurrence of the *rbg* receptor element in each of these alleles could be confirmed. The presence of identical amplification products for all five alleles demonstrated that the *rbg* element occupies identical positions in these alleles. The *o2-Columbian*, *o2-Agroceres*, *o2-261* and *o2-mh* alleles could therefore be identified as independent isolates of the *o2-m(r)* recessive allele.

An o2 mRNA could be detected in the *o2R* and *o2-Crow* mutants, while *o2-m(r)*, *o2-Charentes* and *o2-G* did not reveal a transcript homologous to O2. In the *o2-m(r)* mutant, the absence of an O2 transcript results from the inactivation of *O2* by insertion of an *rbg* transposable element in the gene's untranslated leader sequence. When subjected to PCR analysis with different sets of oligonucleotides, spanning the entire O2 coding and promoter region, the *o2-G* and *o2-Charentes* alleles did not reveal deletions or insertions (data not shown). As neither the *o2-Charentes* nor the *o2-G* mutant give rise to an O2-homologous mRNA, the recessive character of these alleles might be due to nucleotide changes in the promoter region. In order to validate this hypothesis, the nucleotide sequence of the promoter region of these alleles should be determined.

Comparison of genomic sequences spanning the first intron of *O2* from a series of wild-type and recessive alleles, revealed the presence of a hypervariable region, involving different dipeptides, in the N-terminal part of the O2 protein, in which 8 different protein sequences were found in the 13 sequences analysed. The repetitive nature of this sequence stretch and the fact that the sequences differ in the number of repetitions are reminiscent of a micro-satellite. This finding is in good agreement with previous observations indicating that variation in the number of repetitions within a block of tandem repeats seems to be a universal feature of eukaroytic DNA, regardless of the length and type of repeats (Weber *et al.* 1989). From this point of view, Gerber *et al.* (1994) have proposed that variably sized polymeric amino acid stretches, which are often encoded by triplet repeats, can have a positive role in evolution. These stretches might be the main cause for modifications of transcription factor activity, resulting in subtle or overt genomic effects. Therefore, the variably sized repetitions found in the coding region of *O2* might alter the gene's transcription factor activity. Pro-Glu repeats have also been found in the E1A protein of adenovirus (Moran & Mathews, 1987), in the proteins of *Trypanosoma brucei* (Roditi *et al.* 1987), and in the transposable element *Activator* of maize (Junze *et al.* 1987). In all these cases it appears that the repeats provide an unstructured stretch as a hinge region to connect more strongly structured protein domains.

The cDNA sequences of the *o2R and o2-Crow* alleles were determined. Both loci have undergone numerous alterations in the coding region. In the case of *O2R*, a deletion in the coding region causes a frame shift in the open reading frame, resulting in premature termination of the protein. The predicted polypeptide is devoid of both the acidic and b-ZIP domain, as well as the C-terminal part of the protein.

The *o2-Crow* allele not only exhibits a large deletion in the coding region, but also contains a nucleotide stretch which does not show any homology with the O2 sequence. This nucleotide stretch displays homology with O2 intron sequences. Since this deletion is not only present at the cDNA level, but also in the corresponding genomic DNA of this allele (data not shown), aberrant or alternative splicing is not likely to be the cause for this sequence alteration. However, the presence of this DNA tract could be due to a recombination event involving intron sequences. Evaluation of the putative o2-Crow protein sequence suggested that this *o2* allele has the potential to encode only a truncated polypeptide, lacking the zipper region and C-terminal part of the O2 protein.

Using Kimura's 3ST evolutionary distance (Kimura, 1981), we calculated that the *O2* genes, now represented by the alleles *o2R* and *O2*, diverged from a common sequence approximately 1 million yr ago.

https://doi.org/10.1017/S0016672300032961 Published online by Cambridge University Press

Sequence data for the *o2-Crow* and *o2* alleles indicate that they diverged approximately 1·86 million yr ago. Considering the fact that part of the differences encountered between *O2*, *o2-Crow*, and *o2R* are due to the occurrence of the mutations leading to the recessive opaque-2 phenotypes, and that the rate of sequence change might have been higher once the alleles had lost their function, the given evolution distances should be lower but still quite significant. Available fossil and biosystematic evidence suggests that maize emerged within the past 10000 yr as a recent domesticated derivative of teosinte (Doebley *et al.* 1990; Iltis, 1987). Moreover, it has been argued that the evolution of maize from teosinte has required only a few thousand years (Iltis, 1987). If this is indeed the case, the genes which originated the *o2R* and *o2-Crow* alleles must have arisen in a teosinite background, that is, far before the domestication of maize from its presumed teosinte progenitor.

References

Burr, F. A. & Burr, B. (1982). Three mutations in *Zea mays* affecting zein accumulation. *Journal of Cell Biology* **94**, 201–206.

Dean, C., Van Den Elzen, P., Tamaki, S., Dunsmuir, P. & Bedbrook, J. (1985). Differential expression of the eight genes of the petunia ribulose biphosphate carboxylase small subunit multi-gene family. *EMBO Journal* **4**, 3055–3061.

Doebley, J. D., Stec, A., Wendel, J. & Edwards, M. (1990). Genetic and morphological analysis of a maize–teosinte F2 population: implications for the origin of maize. *Proceedings of the National Academy of Sciences, USA* **87**, 9888–9892.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6–13.

Gerber, H. P., Seipel, K., Georgiev, O., Höfferer, M., Hug, M., Rusconi, S. & Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**, 808–811.

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. & Thompson, R. D. (1989). The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO Journal* **8**, 2795–2801.

Hartings, H., Spilmont, C., Lazzaroni, N., Rossi, V., Salamini, F., Thompson, R. D. & Motto, M. (1991). Molecular analysis of the *Bg–rbg* transposable element system of *Zea mays* L. *Molecular and General Genetics* **227**, 91–96.

Hinnebusch, A. G. (1990). Involvement of an initiation factor and protein phosphorylation in translational control of GCN4 mRNA. *Trends in Biochemical Science* **15**, 148–152.

Iltis, H. H. (1987). Maize evolution and agricultural origins. In *Grass Systematics and Evolution* (ed. T. Soderstrom, K. Hilu, C. Campbell and M. Barkworth), pp. 195–213. Washington, D.C.: Smithsonian Institution Press.

Jones, R. A., Larkins, B. A. & Tsai, C. Y. (1977). Storage protein synthesis in maize. II. Reduced synthesis of a major component by the *opaque-2* mutant of maize. *Plant Physiology* **59**, 525–529.

Kimura, M. (1981). Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences, USA* **78**, 454–458.

Kodrzycki, R., Boston, R. S. & Larkins, B. A. (1989). The *opaque-2* mutation of maize differentially reduces zein gene transcription. *The Plant Cell* **1**, 105–114.

Kunze, R., Stochaj, U., Laufs, J. & Starlinger, P. (1987). Transcription of transposable element *Activator* (*Ac*) of *Zea mays* L. *EMBO Journal* **6**, 1555–1563.

Landschultz, W. H., Johnston, P. F. & McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759–1764.

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N., Hartings, H., Salamini, F. & Thompson, R. D. (1991). The maize regulatory locus *opaque-2* encodes for a DNA-binding protein which activates the transcription of the *b-32* gene. *EMBO Journal* **10**, 617–624.

Lohmer, S., Motto, M., Maddaloni, M., Salamini, F. & Thompson, R. D. (1993). Translation of the mRNA of the maize transcriptional activator *Opaque-2* is inhibited by upstream open reading frames present in the leader sequence. *The Plant Cell* **5**, 65–73.

Maddaloni, M., Di Fonzo, N., Hartings, H., Lazzaroni, N., Salamini, F., Thompson, R. D. & Motto, M. (1989). The sequence of the zein regulatory gene *Opaque-2* (*O2*) of Zea mays. *Nucleic Acids Research* **89**, 7532.

Mertz, E. T., Bates, L. S. & Nelson, O. E. (1964). Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science* **145**, 279–280.

Montanelli, C., Di Fonzo, N., Marotta, R., Motto, M., Soave, C. & Salamini, F. (1984). Occurrence and behaviour of the components of the *o2-m(r)-Bg* system of maize controlling elements. *Molecular and General Genetics* **197**, 209–218.

Moran, E. & Mathews, M. B. (1987). Multiple functional domains in the adenovirus E1A gene. *Cell* **48**, 177–178.

Motto, M., Di Fonzo, N., Hartings, H., Maddaloni, M., Salamini, F., Soave, C. & Thompson, R. D. (1989). Regulatory genes affecting maize storage protein synthesis. *Oxford Survey of Plant Molecular and Cell Biology* **6**, 87–114.

Motto, M., Maddaloni, M., Ponziani, G., Brembilla, M., Marotta, R., Di Fonzo, N., Soave, C., Thompson, R. D. & Salamini, F. (1988). Molecular cloning of *o2-m5* allele of *Zea mays* using transposon marking. *Molecular and General Genetics* **212**, 488–494.

Roditi, I., Carrington, M. & Turner, M. (1987). Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature* **325**, 272–274.

Salamini, F. (1980). Controlling elements at the *Opaque-2* locus of maize: their involvement in the origin of spontaneous mutations. *Cold Spring Harbor Symposia on Quantitative Biology* **45**, 467–476.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. New York: Cold Spring Harbor Laboratory Press.

Schmidt, R. J., Burr, F. A., Aukerman, M. J. & Burr, B. (1990). Maize regulatory gene *opaque-2* encodes a protein with a 'leucine-zipper' motif that binds to zein DNA. *Proceedings of the National Academy of Sciences, USA* **97**, 46–50.

Schmidt, R. J., Burr, F. A. & Burr, B. (1987). Transposon tagging and molecular analysis of the maize regulatory locus *opaque-2*. *Science* **238**, 960–963.

Schmidt, R. J., Ketudat, M., Aukerman, M. J. & Koschek, G. (1992). *Opaque-2* is a transcriptional activator that recognizes a specific target site in 22 kD zein genes. *The Plant Cell* **4** 689–700.

Soave, C., Righetti, P. G., Lorenzoni, C., Gentinetta, E. & Salamini, F. (1976). Expressivity of the *Opaque-2* gene at the level of zein molecular components. *Maydica* **21**, 61–75.

Soave, C., Tardani, L., Di Fonzo, N. & Salamini, F. (1981). Zein level in maize endosperm depends on a protein under control of the *Opaque-2* and *Opaque-6* loci. *Cell* **27**, 403–410.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments seperated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**, 911–916.

Weber, J. L. & May, P. E. (1989). An abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**, 388–396.

Wolfe, K. H., Li, W. H. & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* **84**, 9054–9058.