# Molecular Characterisation of the *et1* gene of *Zea mays* L.

**Dissertation**

A thesis submitted to the
Fachbereich Biologie, Universität Hamburg
for the degree of
*doctor rerum naturalium*

by

Preeti Garg
New-Delhi, India

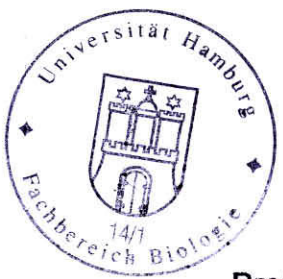
Hamburg
2002

Genehmigt vom
Fachbereich Biologie der
Universität Hamburg
auf Antrag von Herrn Professor Dr. U. Wienand

Weitere Gutachter der Dissertation:
Herr Professor Dr. H. Lörz

Tag der Disputation: 01. Februar 2002

Hamburg, den 16. Januar 2002




Professor Dr. A. Frühwald
Prodekan

# Table of Contents

# Abbreviations

## Chemicals and Solutions

| | |
|---|---|
| CSPD | 3-(4-methoxyspiro{1,2-dioxetan-3,2´-(5´chloro)tricyclo[3.3.1.1$^{3,7}$]decan}-4-yl) Phenylphosphate, Disodium salt |
| DEPC | **die**thyl **p**yro**c**arbonate |
| DEPC-$H_2O$ | DEPC treated $H_2O$, for RNA preparation and analysis |
| Dig / DIG | Digoxigenin |
| DTT | Dithiothreitol |
| EDTA | EthyleneDiamineTetraAceticacid |
| EGTA | EthyleneGlycol-bis-(β-aminoethylether)-N,N,N´,N´-TetraAceticacid |
| HEPES | (N-[2-**H**ydroxy**e**thyl] **p**iperazine-N´-[2-**e**thane**s**ulfonic acid]) |
| LB | Luria Bertani medium, for bacterial cultures |
| MOPS | (3-[N-**Mo**rpholino] **p**ropane**s**ulfonic acid) |
| SDS | **S**odium**D**odecyl**S**ulphate |
| Tris | **Tris**-(hydroxymethyl)-aminomethane |
| TAE | **T**ris-**A**cetate-**E**DTA electrophoresis buffer |
| TEMED | N,N,N´,N´-**Te**tra**m**ethyl-**e**thylene**d**iamine |

## Genetics and Molecular biology

| | |
|---|---|
| aa | amino acids |
| AD | DNA **ad**apter (AIMS analysis) |
| AFLP | amplified fragment length polymorphism |
| AIMS | Amplification of Insertion Mutagenised Sites |
| ATP | **a**denosine 5´-**t**ri**p**hosphate |
| B73 | an inbred maize line with coloured kernel, |
| cDNA | complementary DNA |
| dCTP | 2´-**d**eoxy**c**ytidine 5´-**t**ri**p**hosphate |
| DAG | **d**ays **a**fter **g**ermination |
| DAP | **d**ays **a**fter **p**ollination |
| DNA | **d**eoxyribo**n**ucleic **a**cid |
| ddNTP | 2´-3´-**d**i**d**eoxyribo**n**ucleoside 5´-**t**ri**p**hosphate |
| dGTP | 2´-**d**eoxy**g**uanosine 5´-**t**ri**p**hosphate |
| dNTP | 2´-**d**eoxyribo**n**ucleoside 5´-**t**ri**p**hosphates |
| EST | Expressed Sequence Tag |
| *et1*/ET1 | *etched1*/ETCHED1 (mutation/ gene) |
| *et1-m* | *etched1-mutant* |
| *et1-R/et1-Ref* | reference allele of *etched1* mutation |
| indel | insersion/deletion; polymorphism observed in homologous allelic sequences |
| LC | Line C, a pure inbred colour converted derivative of the W22 maize line |

| | |
|---|---|
| MCS | Multiple Cloning Site / polylinker |
| mRNA | messanger RNA |
| NEP | **n**uclear-**e**ncoded RNA **p**olymerase |
| OEC | **O**xygen **E**volving **C**omplex |
| ORF | open reading frame |
| PAGE | **P**oly**A**crylamide **G**el **E**lectrophoresis |
| PCR | Polymerase Chain Reaction |
| PEP | **p**lastid-**e**ncoded RNA **p**olymerase |
| pfu | plaque forming unit(s) |
| Poly A+ mRNA | **p**oly**a**denylated mRNA |
| Polym. | Polymerase |
| Pt. Taq Polym. | Platinum Taq Polymerase |
| GTP | **G**uanosine 5´-**t**ri**p**hosphate |
| RAPD | random amplified polymorphic DNA |
| RFLP | restriction fragment length polymorphism |
| RNA | **r**ibo**n**ucleic **a**cid |
| RNAP | **RNA p**olymerase |
| RNAse | ribonuclease |
| rNTP | **r**ibo**n**ucleoside 5´-**t**ri**p**hosphate |
| RPB9 | ninth largest RNA polymerase II subunit |
| rRNA | ribosomal RNA |
| SNP | **s**ingle **n**ucleotide **p**olymorphism |
| TFIIS | **t**ranscription elongation **f**actor S-II |
| UTR | **U**n**t**raslated **R**egion (of a transcript) |
| wt | wild type |

## Standard Units and Physical Parameters

| | |
|---|---|
| A | optical Absorbance |
| bp | basepairs |
| °C | degree Centigrade/ Celsius |
| Ci | Curie, unit of radioactivity |
| conc. | concentration |
| g | gram |
| ×g | × acceleration due to gravity,<br>unit used for relative centrifugal force (rcf) |
| hr | hour(s) |
| J | joule(s) |
| kb(p) | kilobase (pairs) |
| kD | kilodalton |
| m | meter(s) |
| M | molar |
| min. | minute(s) |

| | |
|---|---|
| mix. | mixture |
| µl | microlitre(s) |
| mol | mole(s) |
| OD | Optical Density |
| rpm | revolutions per minute |
| RT | room temperature |
| s | second(s) |
| U | **u**nit (restriction enzyme) |
| UV | ultraviolet light |
| V | Volts |
| v/v | volume / volume |
| w/v | weight / volume |

## Standard prefixes used for Measurements

| | |
|---|---|
| c | centi ($10^{-2}$) |
| m | milli ($10^{-3}$) |
| µ | micro ($10^{-6}$) |
| n | nano ($10^{-9}$) |
| p | pico ($10^{-12}$) |

## Amino Acids

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **A** | Ala | Alanine | **H** | His | Histidine | **Q** | Gln | Glutamine |
| **B** | Asx | Asparagine or aspartic acid | **I** | Ile | Isoleucine | **R** | Arg | Arginine |
| **C** | Cys | Cysteine | **K** | Lys | Lysine | **S** | Ser | Serine |
| **D** | Asp | Aspartic Acid | **L** | Leu | Leucine | **T** | Thr | Threonine |
| **E** | Glu | Glutamic Acid | **M** | Met | Methionine | **V** | Val | Valine |
| **F** | Phe | Phenylalanine | **N** | Asn | Asparagine | **W** | Trp | Tryptophan |
| **G** | Gly | Glycine | **P** | Pro | Proline | **Y** | Tyr | Tyrosine |
| **Z** | Glx | Glutamine or glutamic acid | | | | | | |

# 1. Introduction

## 1.1. Maize as a crop plant

Maize, also known by the name 'Indian corn' or simply 'corn', is a member of the grass family *Gramineae* (*Poaceae*), to which all of the other important cereal crops like rice, wheat, barley, oats, millets and sugarcane belong. Like other cereals, maize became established as a food crop in its probable region of origin, Central America, because it provided a storable form of food energy, which could be planted and harvested periodically in a predictable manner, and was easily improved by mass selection (Watson, 1988). Today, it is one of the three major cereal crop plants of the world, along with rice and wheat and is grown in almost all the continents around the world (White, 1999).

Maize is estimated to have been domesticated as a food crop around 8000 years ago by America's indigenous people (Galinat, 1988), who selected the heritable changes leading to the emergence of the maize plant of today. However, the widespread use of maize as a crop plant took place as maize breeders in early-to-middle part of the 20th century introduced hybrid maize, which led the farmers around the world to adopt this crop (Peterson, 1999; White, 1999). The early decades of the 20$^{th}$ century saw a surge of activity in the field of genetics and inheritance of newly discovered traits, as Mendel's laws were rediscovered and "*the chromosome theory of heredity*" was proven. This led the attention of the researchers to the inheritance of newly uncovered traits, the observation of heterosis, the linkages among traits, and the chromosomal relationships of these genetic linkages, which established the solid foundation of maize genetics. Since then maize breeders have made significant progress in the improvement of maize over the last decades, which has contributed to about 50% of the increase in yield through the use of improved varieties; the other half derives from greater and efficient use of fertilisers and crop management (Peterson, 1999).

Today, maize is not only a major food and feed crop, but it is also important for industrial uses. Maize is used industrially for varied purposes, which include the

manufacture of products like corn starch, corn syrup, dextrose, corn oil, binding material, foaming agents in fire extinguishers, fermented beverages, distilled liquors, fuel alcohol (ethanol) and other industrial alcohols, antibiotics and plastics (including biodegradable plastics) as examples. Corn starch is used in the manufacture of varied industrial products like plastics, ceramics, dyes, paper, linoleum etc. Corn oil, besides for cooking purposes, is also used in the manufacture of soaps, varnishes, paints and other similar products. The reason for its wide industrial use is that it is a very cheap source of an industrial raw material. In addition, the large heritable diversity available in maize has been used by breeders for producing further hybrid varieties required for specialised purposes (Kochar, 1981; Watson, 1988).

## 1.2. Maize as a model system for genetic analysis

The maize plant and specifically the seed-bearing portion named the maize ear possess a number of varied qualities that make it an ideal candidate for research. Because of these traits, maize is one of the most characterised among all the crop plants leading to significant achievements in maize genetics and breeding.

As a plant, maize is easy to systematically culture on a small as well as large scale. It is a monoecious plant with the male (tassel) and female (ear) inflorescence present separately on the culm. The tassel terminates the stem and the ears arise terminally on shortened lateral branches. Maize is a cross-pollinating plant in nature, but which can be easily manipulated in a crossing program, including selfing. The robust stature of the plant, as also of its tassel, ear and kernels (with a conspicuous endosperm and embryo) contribute greatly to the ease with which observations as well as manipulations can be carried out in maize. In addition to the size, the large number of kernels available per ear is also advantageous, which can be as many as several hundreds, and easily stored for later use in genetic analysis. In a large-scale field-crossing program, the maize plant offers five days of available pollen to an experimental field researcher. Moreover, it is possible to pollinate as many as fifty plants from the pollen of a single plant. The maize plant has also been successfully adapted to different “hostile” environmental conditions, a characteristic also called “genetic flexibility” by geneticists (Motto et al., 1999; Peterson, 1999).

Moreover, the plant possesses bilateral symmetry so that the vegetative parts of the plant body are divided into developmental units, which make maize suitable for analysing clonal relationships in organogenesis, and for identifying genes controlling cell differentiation and morphogenesis, as well as pattern formation. This tractable genetic system of maize allows for efficient and precise cytological and cytogenetic experimentation, and the identification and systematic analysis of the tissues at different developmental stages. Earlier research on endosperm genetics, including cytogenetics, has provided a wealth of information and laid foundation for molecular biological research in maize. Moreover, maize has a vast and unparalleled number of mutants, whose analysis, involving the cloning and characterisation of genes engaged in important physiological processes in the plant and the endosperm, is also contributing greatly to the understanding of developmental processes in plants (Coe et al., 1988; Motto et al., 1999; Peterson, 1999).

The availability of a large number of well-characterised transposable elements in maize has played a very important role and has provided another valid reason for using maize as a model genetic-system for understanding plant processes. Transposable elements were first discovered in maize by Barbara McClintock in the mid-twentieth century, as these novel genetic factors were found to inhibit expression of neighbouring genes, and were not fixed to a permanent chromosomal location. Their ability to move from one position to another in the maize genome has been utilised in the field of reverse genetics to develop a number of molecular biological techniques and strategies for rapid isolation of novel genes, which are then used for further examination of their function. This has made possible the identification and discovery of genes involved in key regulatory mechanisms, such as spatial and temporal control of cell differentiation and pattern formation. Thus, vast opportunities have been opened for a better understanding of the complex regulation of plant growth and development, not only in maize, but also in other important plants (Peterson, 1999; Motto et al., 1999; Chomet, 1994).

## 1.3. Structure of maize kernel and endosperm

Botanically, the maize kernel is a caryopsis – a dry, one-seeded indehiscent fruit where the pericarp is fused with the seed coat forming a tough protective covering (Kochar, 1981). Morphologically, the mature kernel is made up of four major parts: the tip cap or pedicel (maternal sporophytic tissue); the 'hull' (sporophytic) constituted together by the pericarp, and remnants of nucellus and the seed coat; the endosperm and the embryo. The relative proportions of these components in a dent maize, on a dry weight basis, are endosperm, ~82%; embryo, ~12; hull, ~5% and pedicel, ~1% (Kochar, 1981; Kiesselbach, 1949). The details of the maize kernel structure are shown in Fig. 1.1.

The endosperm, which constitutes the bulk of the caryopsis, develops from the fertilised triploid central cell, following double fertilisation. During double fertilisation, the egg cell is fertilised by one of the two sperm nuclei to form the zygote, and the central cell, containing the two polar nuclei, is fertilised by the second sperm nucleus to form the endosperm. The polar nuclei undergo karyogamy with the second sperm nucleus to form a triploid primary endosperm nucleus. It, then, undergoes a series of free-nuclear divisions to form an initial coenocytic stage (also called syncytium). This is soon followed by cellularisation. Free cell wall formation in the anticlinal plane begins at the periphery of the endosperm and progresses centripetally to form tube like wall structures, called alveoli, surrounding each nucleus. From this point onwards, the normal cell division begins in the periclinal plane, with cell plate formation between daughter nuclei. This process continues and proceeds towards the central region of the endosperm, until it is completely cellularised. In maize, it is completed around 3 to 5 days after pollination (DAP). (Kranz et al., 1998; Becraft and Asuncion-Crabb, 2000).

As the endosperm grows further, mitotic cell divisions first occur in the central endosperm region until ~12 DAP, which then get localised to the periphery of the endosperm until 20 to 25 DAP (Kranz et al., 1998; Becraft and Asuncion-Crabb, 2000). The endosperm is composed of three major cell-types: the starchy endosperm (E), the aleurone layer (AL) and the basal endosperm transfer layer (BETL). Mutational studies as well as *in situ* hybridisation experiments have shown that proper development of all these three regions is important for successful germination

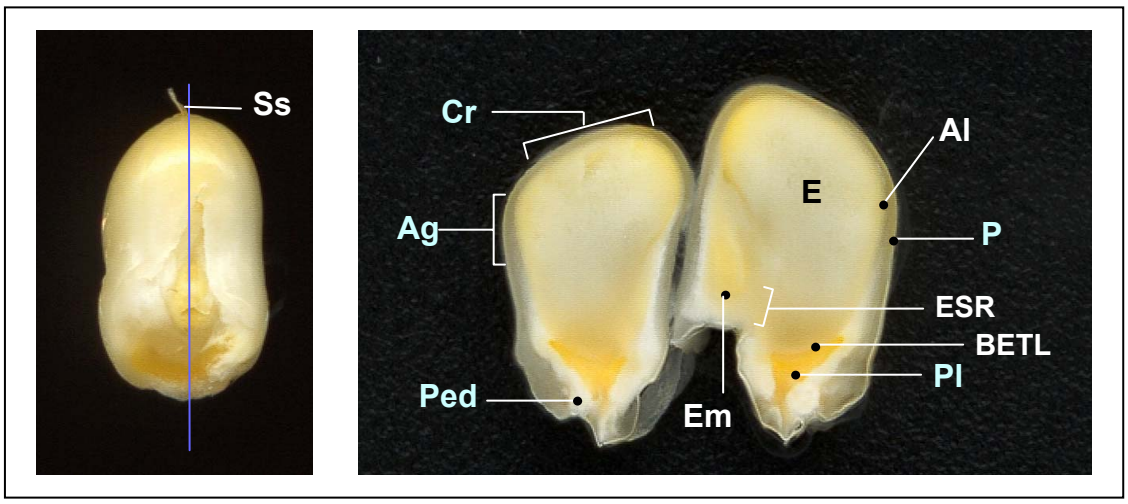



**Fig. 1.1: Longitudinal section of a 10 DAP maize kernel (caryopsis).** Left panel: A whole 10 DAP maize kernel, showing the silk scar (Ss), as viewed perpendicular to the plane of section (blue line). Right panel: Longitudinal section of a 10 DAP maize kernel, showing the various external and internal parts, including the maternal sporophytic tissue, the growing embryo or new sporophyte (Em) and endosperm (E). **Abbreviations:** External features: **Ss**, silk scar; **Cr**, crown region of the kernel; **Ag**, abgerminal region of the kernel. Sporophytic tissue: **Ped**, pedicel; **P**, pericarp; **Pl**, placental region. **Em**, embryo; Endosperm: **E**, starchy endosperm region; **Al**, aleurone layer; **BETL**, basal endosperm transfer layer; **ESR**, embryo surrounding region.

and survival of the sporophyte (embryo) in the next generation. Maize endosperm possesses a single peripheral layer of densely cytoplasmic aleurone cells, constituting the aleurone layer that surrounds the starchy endosperm, except in the region opposing the embryo. Upon kernel germination and stimulation by gibberelic acid, it secretes hydrolytic enzymes that break down the storage products in the starchy endosperm into simple sugars and free amino acids, which are made available for uptake by the growing seedling. The basal endosperm transfer layer consists of elongate cells with finger-like cell wall involutions to maximise the surface area of the cell-membrane. These cells reside over the pedicel and transfer nutrients from the maternal vasculature to the developing endosperm. The starchy endosperm cells are filled with amyloplasts, which contain prominent starch grains and protein bodies. The starchy endosperm shows the presence of two specialised regions: the sub-aleurone layer (SAL) and the embryo-surrounding region (ESR). The sub-aleurone layer is present at the periphery of the starchy endosperm, next to the aleurone layer. These cells are smaller than the rest of the endosperm and constitute the region of cell proliferation during later stages of endosperm development. The

embryo-surrounding region consists of small cells with a denser cytoplasm than other starchy endosperm cells (Olsen et al., 1999; Becraft and Asuncion-Crabb, 2000).

## 1.4. Plastid organisation

Plastids are sub-cellular, semi-autonomous compartments or organelles, found in eukaryotic cells of vascular plants and algae. Evolutionarily, plastids, as well as mitochondria, are believed to have originated from prokaryotic systems, which probably established an endosymbiotic relationship with the proto-eukaryotic "host" before they established themselves as the present-day eukaryotic cell organelle. This process involved several adaptive changes including transfer of a majority of their genes to the nucleus of the eukaryotic "host" cell (Schwartz and Dayhoff, 1978; Howe, 1996). Therefore, they rely on the nucleus for a majority of the structural proteins and regulatory factors that control the expression of their genes. Plastids still possess about hundred genes, most of which encode components of the photosynthetic electron transport machinery and constituents of transcriptional and translational apparatus (Stern et al., 1997). Plastids are enclosed by a double-membrane, called the envelope, and contain an inner membrane system, constituting the thylakoids – the sites of photosynthetic electron transport – in chloroplasts, and some similar structures in the chromoplasts.

Various specialised plastid types are known from different plant parts, whose differentiation from the proplastids also involves an active participation of the nuclear genes (Fig. 1.2). Four major types of plastids are known from plants. The chloroplasts, the chromoplasts, the leucoplasts and the amyloplasts. Matabolism of all these plastids is linked to the function of the plant tissue in which they are present. The chloroplasts usually occur in all the green parts of the plant, especially leaves, as they contain chlorophyll pigment-containing molecules of the photosynthetic electron transport machinery. They can be slightly variable in morphology, depending on their activity, age or location in the plant. They also sometimes store starch (transitory starch) and simultaneously photosynthesise. Older chloroplasts, in senescing leaves, are called gerontoplasts (Camara et al., 1995). Moreover, chloroplasts can also be converted into chromoplasts, as observed in ripening tomatoes (Hess, 1999; Joyard

et al., 1998). The chromoplasts, which usually occur in fruits and flowers and are responsible for their colouring, produce and accumulate different carotinoid pigments. The chromoplasts also show variable morphology and possess four broad categories of pigment containing structures called tubules, crystals, plastoglobules and membranes. The plastoglobules are chemically distinct from that of senescing chloroplasts. These structures can either be present simultaneously or separately at different stages of development (Ljubesic et al., 1991; Camara et al., 1995). Based on their morphology, they have been classified into four (Hess, 1999; Fig. 1.2) or by some others into five major types (Camara et al., 1995; in legend to Fig. 1.2).

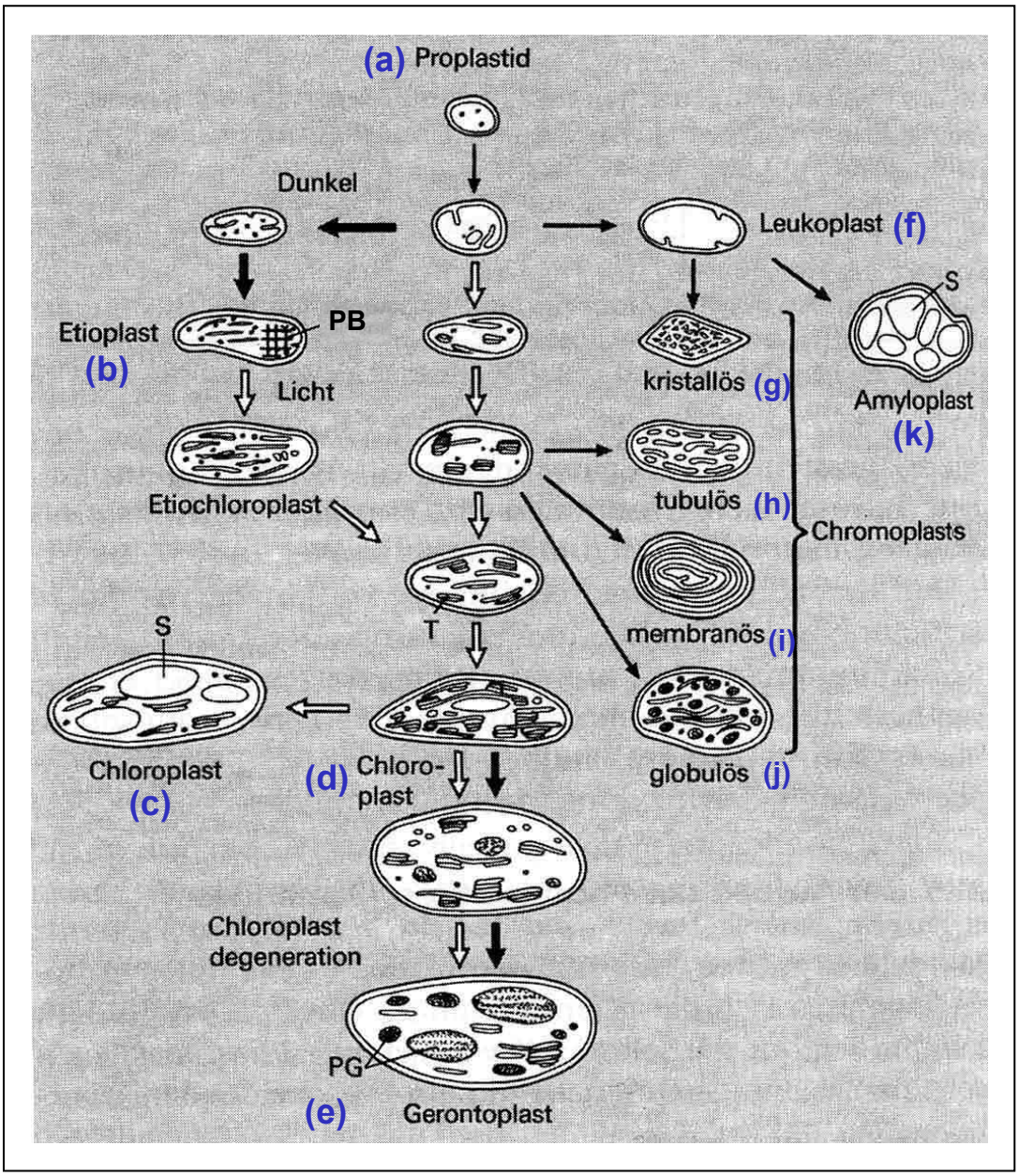



**Fig. 1.2:** Legend on the next page

**Fig. 1.2: General scheme of development of different types of plastids (modified from Hess, 1999).** Proplastids **(a)** are the plastid precursors, giving rise to all the different types of plastids, depending upon the function of the tissue where these develop. In green tissues, like leaves, they normally develop into chloroplasts **(d)** through different intermediate stages in a light dependent manner. In some tissues they are sometimes also found to contain starch grains **(c)**. In the absence of light, proplastids develop into etioplasts **(b)**, which can, however, develop into chloroplasts, when exposed to light. Older chloroplasts are called gerontoplasts **(e)**. In specialised tissues, like the oil glands of mint or in mustard seeds, leucoplasts **(f)** are the sites of monoterpene and fatty acid biosynthesis respectively. In other plant tissues, like fruits and flowers, the proplastids develop into chromoplasts. There are four major types of chromoplasts: cristalline **(g)**, tubular and fibrilar **(h)**, membranous **(i)** and globular **(j)**. A fifth type (not shown) is reticulo-tubular (Camara *et al.*, 1995). Amyloplasts **(k)** are the specialised starch-storing plastids of storage tissues, like roots, stems, fruits, and seeds.

**Abbreviations and translations :** Dunkel: dark; Licht: light; PB: prolamellar bodies; PG: plastoglobules; S: starch; T: thylakoid. The following symbols were used for showing the different development patterns of chloroplasts:

➡ **:** chloroplast development in dark; ⇨ **:** chloroplast development in light.

Leucoplasts are structurally comparatively simple. They do not possess inner thylakoid like membranes, are colourless, and have been identified to be structurally distinct from white chromoplasts and amyloplasts (Carde, 1984). They have been identified as the sites of initiation of monoterpene (terpinoids) biosynthesis in the oil gland secretory cells (glandular trichomes) in mint (*Mentha* sp.) during essential oil synthesis (Turner et al., 2000). Moreover, they have also been found to be the sites of fatty acid synthesis in the seeds of *Brassica campestris* (Gupta and Singh, 1996). Amyloplasts are the storage plastids, which synthesise and accumulate starch. They are usually present in starch storing tissues, like roots, stem, tubers, fruits and seeds.

In addition, another type of plastids, called etioplasts, develop in the leaves of etiolated seedlings, grown in the dark. These are sometimes found to contain the prolamellar bodies, which are formed by a fine grid-like network of tubular vesicles, initially destined to form the thylakoids. As soon as the seedlings are subjected to light, these bodies reorganise to form the thylakoids and the etioplasts transform into chloroplasts (Hess, 1999).

## 1.5. Kernel and seedling mutants of maize

As mentioned before, an unparalleled number of mutations are available from the maize germplasm, of which kernel as well as seedling mutations constitute an important part. A large number of the mutations affecting the two plant parts, directly or indirectly, are caused by a lack of normal plastid differentiation or development in the respective tissues. So far, only for a relatively small fraction of these mutations have the responsible genes been isolated and characterised. A number of the identified genes were found to be structural genes, coding for key enzymes in the plant metabolic pathways, partly known from earlier physiological and biochemical studies. More importantly, however, characterisation of regulatory genes disrupted in some of these maize mutants has shown the value of mutant analyses, which have made possible the identification of regulatory functions controlling plant growth and development. On the whole, these regulatory functions involve spatial and temporal co-ordination of various physiological processes, like cellular differentiation and pattern formation, as well as primary and secondary metabolic pathways. This involves a complex interplay between a large number of regulatory factors, which are also influenced by external abiotic factors.

Similarly, regulation of maize kernel differentiation and development also involves a number of regulatory genes, as indicated by the large number as well as variations of kernel mutants (phenotypes) available from maize. A large number of *defective kernel* mutants (*dek1- dek33*) are available (Neuffer et al., 1997), which seem to be affecting regulatory genes. All these mutants are generally lethal, as the embryos or, in some cases, the seedlings, are non-viable. Sometimes, only the roots are produced upon germination. The kernels are collapsed and floury, pitted, or just opaque. In a large number of the cases, no proper aleurone layer develops or only in patches. Similarly, the starchy endosperm is either floury or showing both floury and corneous patches. The seedlings, when present, show narrow leaves with white, pale-green or, sometimes, normal green colour, or white stripes (*dek5*). In some cases, the seedlings are viable in embryo cultures. This observation indicates that, in some mutants, the seedling development is hindered by the lack of available nutrients. This, in turn, probably happens due to reduced amount of the starchy endosperm or lack of aleurone cells, which provide the hydrolytic enzymes for

digesting the starch and proteins into usable nutrients, i.e., glucose and amino acids. However, in a number of cases, the embryo is also reduced in size, because of which the cultured embryo either does not germinate, or produces a weak seedling with one or two pale leaves, which soon dies (Neuffer et al., 1997).

Three *reduced endosperm* mutations (*ren1-3*) show small kernels with reduced endosperm and small embryo, which are usually lethal. However, larger kernels germinate to produce either fertile plants or small plants with sterile rudimentary tassel (Neuffer et al., 1997). There are other mutations affecting the starch quality of the endosperm. In some cases, normal starch metabolism is disturbed so that the kernel remains soft and sugary, and does not solidify or only partly, leading to shrinkage on desiccation. Some well-known examples of genes responsible for such mutations (many of which are also structural genes) are: *shrunken1* (*sh1*) – encoding sucrose synthase1 (SS1); *sugary1* (*su1*) – encoding starch debranching enzyme I; *sus1* – encoding sucrose synthase2 (SS2) (Neuffer et al., 1997). Some further *shrunken* mutations also show a pleiotropic effect with lethal seedlings. In still other mutants, the proportions of amylopectin and amylose in the endosperm can also vary, as detected by biochemical analysis, or simply by the coloration produced by IKI staining. The endosperm of mutants with high amounts of the highly branched amylopectin shows red coloration, whereas those rich in amylose straight chains show blue coloration. *Waxy1* (*wx1*) is one such example, with high amylopectin content (Neuffer et al., 1997). *Opaque2* (*o2*) is another example, which has been characterised molecularly using transposon tagging to be a b-ZIP transcription initiation factor, controlling the expression of zein and b-23 genes in the kernels (Meschi and Iwabuchi, 1995).

Other examples of kernel mutations include *albescent plant1* (*alb1*), where reduced endosperm pigmentation is observed. Associated with it is the presence of both green and albino or pale green leaves in the plant. Very often, the older leaves are normal green, with younger leaves showing the mutant phenotype. Sometimes they can also be cross-banded. This phenotype is due to pleiotropic effect of the mutation, just as in many of the *dek* mutants, where embryo is also affected along with the endosperm. The plants are, however, fairly viable. Other similar pleiotropic mutations are available, like *y10*, which shows pale endosperm and white, lethal seedlings; or

pale endosperm with pale to green seedlings, showing delayed maturity (*y9*). Some other pale endosperms show normal seedling growth with good viability (*y11*, *y12*) and are not pleiotropic in nature (Neuffer et al., 1997).

Similarly, a large number of mutants affecting seedling development are available, which appear to be genes of regulatory nature rather than structural genes. Some such examples are the *zebra crossbands1-8* (*zb1-8*) mutations, which show the presence of yellow crossbands on normal green leaf, either at the seedling stage or at a later stage on older leaves. These bands appear due to the absence of chlorophyll. *Zebra necrotic1* (*zn1)* and *zn2* are similar mutations where necrotic tissue appears between veins in regularly spaced crossbands on leaves of half-grown plants. These plants, however, are all viable. Apart from these, there are other large numbers of mutations causing virescent phenotype of the seedlings (*v1-29*), showing white, pale yellow or pale green leaves on germination, which turn green either fast or slowly. Some of these green in longitudinal streaks, with the newer leaves turning green with white streaks. Most of these mutations are not lethal.

## 1.6. The *etched1* mutation

*Etched1* is a mutation in maize (*Zea mays* L.) first described by Stadler in year 1940. It is a recessive pleiotropic mutation, which was observed in a population of maize plants obtained through pollination with X-ray irradiated pollen grains (Stadler, 1940). This mutant allele is called as *et1-Ref* (or *et1-R*) and represents the reference allele. This mutation has been mapped to the long arm of chromosome 3 [3L-161] (Neuffer et al., 1997). A second *etched* mutation is also known, which shows a slightly different phenotype than *et1* and is called *etched2* (*et2*). It is a lethal mutation, which shows pitted/etched endosperm and albino seedlings with slightly green tips. *Et1* was one of the few mutations to uncover a gene, which seemed to influence two distinct developmental processes, i.e., kernel and seedling development.

The homozygous *et1-Ref* mutation is phenotypically characterised by two distinct features (Neuffer et al., 1997):

(1) Appearance of depressions and crevices in the endosperm surface beneath the pericarp of the kernels.

(2) Virescent phenotype of the young seedlings, i.e., pale-green to albino seedlings up to about 10-15 days after germination (DAG).

The *etched1* mutation gets its name from the fissured or scarred appearance of the kernels. Biochemical and structural analysis of the *et1-Ref* allele (Sangeetha and Reddy, 1991) showed that the mutation caused a reduction in starch content of the endosperm tissue and that starchless cells were present around the indentations in the kernels. These depressions or crevices begin to appear after approximately 15 DAP on the kernel surface. The severity of the *etched1* phenotype in the *et1-Ref* kernels is quite variable, ranging from strongly “etched” to almost normal looking kernels. However, no correlation between the phenotypic variation and the genetic background could be observed.

Microscopic observations of these kernels revealed a deviation from the normal endosperm morphology (Ahrend, 1998; Fig. 1.3). The discontinuity of the aleurone layer or the abnormal aleurone development as a possible cause of the *etched1* phenotype was examined. The presence of a continuous aleurone layer over the storage tissue of the endosperm, even in the fissured areas, could be confirmed in the various microscopic observations. However, as can be observed in figure 1.3 E, the aleurone layer in the apparently fissured area, despite being continuous, was found to be distorted, as if it had been pulled down into the crevice and compressed between the starch cells of the endosperm, thus contributing to the typical *etched1* phenotype (Fig. 1.3).

Leaves of the *etched1* virescent seedlings show reduced levels of chlorophyll and carotenoid pigments (Ramesh et al., 1984, Fig.1.3 F, G). Chloroplast development is delayed in these seedlings (Sangeetha et al., 1986). Microscopic observations of the virescent leaves revealed malformed chloroplasts showing few and almost unorganised chloroplast membranes (Ahrend, 1998). Sometimes thylakoids with poorly organised grana were observed (Fig. 1.3 H, I, J). No difference between the abaxial and the adaxial sides of the leaves were observed.

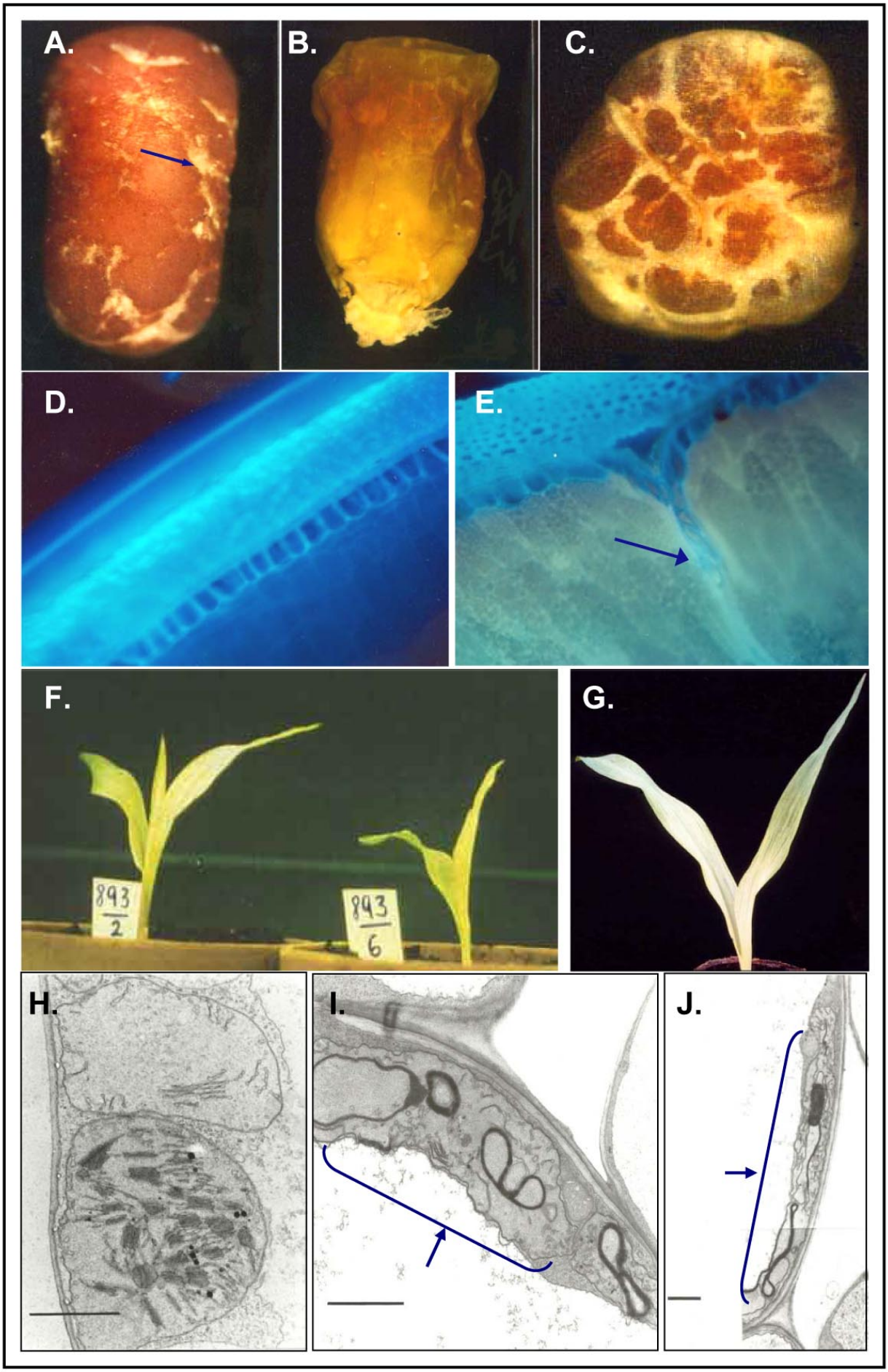



**Fig. 1.3:** Legend on the next page

**Fig. 1.3: *etched1* phenotype of mature kernels and young seedlings.** The *etched1* (*et1*) phenotype is characterised by two main features: Pitted endosperm and pale white to pale green seedlings up to 10-15 days after germination. **(A) to (C).** mature kernels of the reference allele, *et1-R* **(A)**; the *Mu*-induced *et1* mutants, *et1-m3* **(B)**, *et1-m16* **(C)**, showing a number of indentations (blue arrows) on the kernel surface. That these indentations represent pitted endosperm surface (blue arrows) can be observed in the image **(E)**, which is a cross-section of an *et1-R* mature kernel. It shows a continuous pericarp above the endosperm. The aleurone layer of the endosperm is also continuous, however, it has been pulled down into the pits in the starchy endosperm. On the left is an image of a cross-section of a wild-type endosperm **(D)**. The *et1* kernels on germination give rise to seedlings with white to pale green leaves up to 10 to 15 days after germination. Image in **(F)** shows two *et1-R* seedlings, and the image in **(G)** shows an *et1-m16* seedling. Electron microscopic micrographs of such seedlings showed the presence of underdeveloped or almost undifferentiated chloroplasts (blue arrows) in the leaves **(H)**, **(I)** and **(J)**.
**(D)**, **(E)**, **(H)**, **(I)** and **(J)** have been taken from the Diploma thesis (Ahrend, 1998).

## 1.7. Previous work on the isolation and cloning of the *et1* gene

In order to isolate the gene responsible for the *et1* mutation, new *et1* mutants were produced using the *Mutator* element in transposon tagging experiments (Wienand *et al.*, 1982). Thirteen allelic *et1-mutant* maize lines, tagged with the *Mu* tranposons, have been isolated and are being analysed for the e*t1* genomic structure (Scanlon *et al.*, 1994; P. Schnable, Iowa State University, unpublished). Some of the mutant alleles, which have been partially characterised, are *et1-m3* and *et1-m16* by O. da Costa e Silva and *et1-m10*, *et1-m12* and *et1-m15* by M. Wassman (da Costa e Silva *et al.*, 2001). Molecular analysis of some of these *et1* mutations was followed up using a PCR based method called AIMS (Amplification of Insertion Mutagenised Sites; Frey *et al.*, 1998).

A PCR-amplified, 171 bp *HpaG* genomic fragment, cosegregating with the *et1* phenotype (21 mutant individuals) and absent in the wild type plants (15 individuals), was identified from two populations segregating for the mutant (*et1-R/et-m3*) and the wild type (*et1-R/+*) phenotypes (da Costa e Silva *et al.*, 2001). The 171 bp *HpaG* AIMS fragment was used to isolate genomic clones from a lambda genomic library prepared from heterozygous *et1-m3/+* plants. One clone, *et1-m3* λ *5.1*, containing a

*Mu8* insert and three clones not containing a *Mutator* insertion, were isolated from the library, which represented the *et1* mutant allele and the wild type (wt) *Et1* allele respectively (da Costa e Silva *et al.*, 2001).

After carrying out restriction fragment analysis and Southern hybridisation with the different wild type lambda clones, a 2.5 kb (*XhoI/HindIII*) putative wt *et1* genomic fragment, which hybridised with the AIMS fragment, was subcloned and sequenced. This clone was used for further *et1* characterisation. Another 4 kb *Hind III* fragment present next to the 2.5 kb fragment was also subcloned and partly sequenced. The 2.5 kb *XhoI/HindIII* fragment identified from *et1-m3* was used in an RFLP mapping experiment and was mapped to the long arm of chromosome 3 (Ben Burr, personal communication), around the same region, where the *et1-R* mutation has also been mapped (Neuffer *et al.*, 1997).

In another experiment, using *et1*-specific primers the region containing the site of *Mu8* insertion in *et1-m3* was amplified from a green revertant sector of an otherwise pale *et1-m3* seedling leaf (da Costa e Silva *et al.*, 2001). Comparison of the amplified sequence to the corresponding sequences in the wild type and the *et1-m3* mutant clones revealed that the *Mu* element had excised out from the *et1-m3* allele in the green revertant sector. During excision, it left behind the target site duplication, but with a deleted base. This did not cause a frameshift in the ORF, as the region of *Mu* insertion was present upstream of the start codon (Fig. 1.4). This experiment was another strong indication that the clones isolated so far represented the *et1* gene.

| | |
|---|---|
| insertion site (wt) : | GTTGTCTGACAGGCAAAGCGGCTATG |
| *et1-m3 :: Mu8* : | GTTGTCTGACAGGC *Mu8* CTGACAGGCAAAGCGGCTATG |
| *et1-m3 Revertant-1* : | GTTGTCTGACAGG CTGACAGGCAAAGCGGCTATG |

**Fig. 1.4: A comparison of the region of *Mu* insertion (from top to bottom) in the wild type (B73), *et1-m3* mutant allele and the revertant sector of *et1-m3* respectively.** The *Mu8* transposon (cursive blue; in *et1-m3::Mu8*) left behind a target site duplication (underlined) in the revertant sector and while jumping out of the gene caused a basepair deletion. This probably restored the gene activity

in the revertant sector as its position of insertion in the gene was upstream of the translation start site (ATG, dark blue).

A cDNA library derived from developing kernels of a wildtype maize line, Line C (LC), was screened using the 2.5 kb wt *Et1* fragment as a probe (da Costa e Silva *et al.*, 2001), from which a cDNA clone, *c9.1*, containing an open reading frame (ORF) was isolated. An *et1-ref* genomic library was screened by M. Ahrend (Diploma thesis, 1998) using the 2.5 kb *et1* wt fragment probe. The clones isolated could be divided into two classes, based on their hybridisation pattern with the *et1* cDNA probe, *c9.1*. The sequence analysis of the first class and its comparison to the 2.5 kb wt *et1* sequence revealed that the complete *et1* gene was deleted from its locus and, therefore, did not hybridise with the *et1* cDNA probe, *c9.1*. The second class, which hybridised with the *Et1* cDNA probe, were not analysed in the study.

## 1.8. Aim of the study

The focus of this study was a more intensive analysis of a wild type *et1* allele for further molecular characterisation of the *et1* mutation. Before this work was started, a number of different studies were underway, in order to characterise the *et1* mutation in different mutant alleles. An *et1*-specific genomic fragment had been isolated from a *Mu* induced *et1* mutant using transposon tagging and AIMS methods, and a number of new *et1* mutant alleles were being analysed at the molecular level. A more intensive analysis of the wild type *et1* allele was required to further characterise the *et1* gene, which had to be undertaken in this study.

In this study, using a wild type *Et1* cDNA clone, isolated and sequenced from a Line C (LC) developing kernel cDNA library, further analyses were to be performed to confirm that this small sequence really represented the complete and correct ORF, coding for the ET1 protein. Further, isolation and analysis of the *et1* gene from a pure inbred line was required. It was important to characterise the *et1* expression pattern in the whole plant through Northern analyses, which would also indicate the correct

*et1* transcript size and give a better understanding of the possible role played by *et1* during different stages of plant development.

Once the correct *et1* transcript had been identified, the analysis of the ET1 protein with respect to its structure and function would also be required. In light of the available information on the putative ET1 protein structure, the possible regulatory role that ET1 might be playing during early plastid development in the seedlings and endosperm would need to be examined. Since the *et1* mutation seemed to disturb normal plastid development in two different tissues, the possibility that ET1 might be a plastid located protein would also need to be tested through chloroplast import experiments. The results of this experiment could give light to the mechanism of *et1* regulatory role.

Apart from the wild type *et1* clone, analysis of an *et1-Ref* mutant, containing the reference allele, was also to be continued in this study. Isolation and analysis of the genomic clones representing the reference allele previously had shown that the complete *et1* gene was deleted from the *et1* locus (Ahrend, 1998). However, a few genomic clones showed a hybridisation signal with the *et1* cDNA probe, which had to be analysed further in this study. There were two possible explanations for the hybridisation. One was that it was the outcome of a rearrangement that probably took place during *et1* mutation event. Another possibility could be that it was another gene – a second copy or a paralogue. Therefore, the basis of the homology had to be found out. If it were a homologous gene, it would not be unusual, as many other examples of gene duplication in maize were available. In that case, the function of this *et1* homologue, and its possible relation to ET1 would also need to be analysed.

# 2. Materials and Methods

## 2.1. Materials

### 2.1.1. Chemicals and enzymes

The chemicals, used in this work, were obtained from the companies Biomol (Hamburg), Biozyme (Oldenburg), Boehringer Mannheim (now Roche), Duchefa (Harlem, NL), Fluka (neu-Ulm), Gibco BRL/Life technologies (now called Invitrogen, Karlsruhe), Merck (Darmstedt), Perkin Elmer Applied Biosystems (Weiterstadt), Pharmacia Biotech (Freiburg), Promega (Mannheim), Qiagen (Hilden), Roth (Karlsruhe), Serva (Heidelberg), Sigma (Deisenhofen), and Stratagene (Heidelberg). All the chemicals unless specifically mentioned, were of p. A. purity grade. All the enzymes were from the companies Gibco BRL/Life technologies (now called Invitrogen, Karlsruhe), Boehringer Mannheim (now Roche), MBI Fermentas (St. Leon-Rot), Biolabs (Schwalbach) and Pharmacia Biotech (Freiburg). In all the experiments, unless specifically mentioned, double distilled $H_2O$ was used which was prepared with the equipment Millipore Milli-Q Plus Water system (Millipore, Bedford, MA, USA). For nucleic acid radiolabelling, $\alpha$-[$^{32}$P]-dCTP (3,000Ci/mM) and Prime-It® II kit from Pharmacia Biotech (Freiburg) were used. For radiolabelling proteins with $S^{35}$ –Cysteine, the Flexi® Rabbit Reticulocyte Lysate System and an *in vitro* translation kit, from Promega (Mannheim) were used.

### 2.1.2. Nucleic acid Markers

For the estimation of DNA fragment size and concentration in agarose gel the following markers were used:

| Marker | Source | Use |
|---|---|---|
| **1kb ladder** | Gibco BRL (Karlsruhe) | DNA mol. weight marker |
| **DNA Molecular Marker V** | Roche (Mannheim) | DNA mol. weight marker |
| **Smart ladder** | Eurogentec (Searing, BE) | DNA mol. weight and concentration estimation |

| | | |
|---|---|---|
| **Lambda EcoRI-HindIII ladder** | MBI Fermentas (St. Leon-Rot) | DNA concentration estimation |
| **0.24-9.5kb RNA Ladder** | Gibco BRL (Karlsruhe) | RNA mol. weight marker |
| **0.16-1.77kb RNA Ladder** | Gibco BRL (Karlsruhe) | RNA mol. weight marker |

### 2.1.3. Nucleic acid hybridisation membranes and X-ray films

For plasmid and phage-DNA Southern blots the nylon membrane Hybond™ N+ and for genomic DNA Southern experiments the nylon membrane Hybond™ NX from Pharmacia Biotech (Freiburg) were used. For the phage lift hybridisation experiments the nitrocellulose filter Hybond™ C-pure and for autoradiography experiments the X-ray film Hyperfilm™-MP from the same company were used.

### 2.1.4. Kits

| Kit | Use |
|---|---|
| Concert™ Rapid Plasmid miniprep kit, Gibco BRL (Karlsruhe) | plasmid DNA preparation |
| Dynabeads® mRNA Purification kit, Dynal (Oslo, Norway) | Poly A+ mRNA isolation |
| Flexi® Rabbit Reticulocyte Lysate System, Promega Corporation (Madison, WI, USA) | *In vitro* translation and protein radiolabelling |
| Gel band Isolation kit, Pharmacia Biotech (Freiburg) | DNA gel band isolation or purification of PCR fragments |
| Nucleospin® Plasmid DNA purification kit, Machery-Nagel (Düren) | plasmid DNA preparation |
| Prime-It® II, Stratagene (Heidelberg) | nucleic acid radiolabelling |
| Qiagen® Lambda phage DNA Midi prep., Qiagen (Hilden) | λ phage DNA isolation |
| Topo® TA Cloning kit, Invitrogen (Groeningen, NL) | cloning of PCR fragments |
| cDNA synthesis kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III Gold Cloning kit, Stratagene (Heidelberg) | cDNA library preparation |

### 2.1.5. Vectors

| Vector | Remarks | Source | Use |
|---|---|---|---|
| **Uni-ZAP® XR** | λ phage | Stratagene | cDNA cloning |
| **pBluescript SK -** | phagemid | Stratagene | Cloning and sequencing |
| **pZErO-II plasmid** | plasmid | Invitrogen | Cloning and sequencing |
| **PCR® 2.1-Topo** | plasmid | Invitrogen | Cloning of PCR fragments |
| **pBAT** | pBluescript KS+, modified | Annweiler et al. (1991) | *In vitro* transcription and translation |

### 2.1.6. Bacterial strains

Following *Escherichia coli* strains were used during this work :

| *Strain* | *Genotype* | *Source* |
|---|---|---|
| **XL1Blue** | recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB $lacI^q$ ZΔM15 Tn10 ($Tet^r$)] | Stratagene |
| **XL1Blue MRF´** | Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F´proAB $lacI^q$ ZΔM15 Tn10 ($Tet^r$)] | Stratagene |
| **XLOLR** | Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 thi-1 recA1 gyrA96 relA1 lac [F´proAB $lacI^q$ ZΔM15 Tn10 ($Tet^r$)] $Su^-$ (non-suppressing) $\lambda^R$ (lambda resistant) | Stratagene |
| **XL1BlueMRA(P2)** | Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 gyrA96 relA1 lac (P2 lysogen) | Stratagene |
| **TOP 10** | $F^-$ mcrA Δ(mrr-hsdRMS-mcrBC)ϕ80lacZΔM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7697 galU galK rpsL ($Str^R$) endA1 nupG | Invitrogen |
| **TOP 10 F´** | F´{$lacI^q$Tn10($Tet^R$)} mcrA Δ( mrr-hsdRMS-mcrBC)ϕ80lacZΔM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG | Invitrogen |

### 2.1.7. Plant material

The following maize (*Zea mays* L.) lines were used for the *et1* analysis:

| Maize Line | Remarks |
|---|---|
| LC | Line C; Colour converted W22 line |
| *et1-Ref* | all GH # originating from GH 935 and Ac 3479 (Agron 37) |
| *et1-m3* | Glasshouse no. GH 1178, 1182, 1179, 1181; B73 background |
| *et1-m15* | Glasshouse no. GH 1738, 1739 |

### 2.1.8. Clones available before begin of this work

The following clones were available before this work was started:

| Clone | Information | Source |
|---|---|---|
| *c 9.1* | Line C *Et1* cDNA | O. Costa e Silva |
| 2.5kb (XhoI/H3) B73 wt | *Et1* genomic subclone from B73 | O. Costa e Silva |
| 4 kb (H3) B73 wt | *Et1* genomic subclone from B73 | O. Costa e Silva |
| *λ et R 1.1* | *et1-Ref* λ genomic clone | M. Waßman |
| *λ et R 10.2* | *et1-Ref* λ genomic clone | M. Waßman |
| *λ et R 18.1* | *et1-Ref* λ genomic clone | M. Waßman |

### 2.1.9. Primers

All the oligodeoxynucleotide primers used and prepared during this work are listed in Appendix D.

### 2.1.10. General Buffers and stock solutions

| | |
|---|---|
| 10x MEN | 400 mM MOPS, 100 mM sodium acetate, 10 mM EDTA (pH 7.2, with NaOH) |
| Methylene blue stain | 300 mM Sodium acetate (pH 5.2), 0.05% methylene blue |
| 10× MOPS buffer | 50 mM sodium acetate pH7.0, 200 mM MOPS, 10 mM EDTA |

| | |
|---|---|
| 10× Orange G loading dye | 50% (v/v) Glycerine, 100 mM Tris-HCl (pH 7.5), 0.8% (w/v) Orange G |
| $^{32}$P Hybridisation solution | 1 M NaCl, 1% SDS, 10% Dextran sulphate (Na-salt), 100 μg/ml denatured salmon sperm DNA |
| RNA loading dye | 50% (v/v) Formamide, 7% (v/v) Formaldehyde, 1x MEN, 0.4% (w/v) bromophenol blue |
| 10× SSC | 1.5 M NaCl, 0.15 M Sodium citrate |
| SM buffer | 100 mM NaCl, 50 mM Tris-HCl (pH 7.5), 4g/l (16.2 mM) $MgSO_4.7H_2O$ |
| SM Top-Agar | 100 mM NaCl, 50 mM Tris-HCl (pH 7.5), 4g/l (16.2 mM) $MgSO_4.7H_2O$, 7% (w/v) Agarose |
| TE I | 10 mM Tris-HCl (pH 8), 1 mM EDTA |
| TE II | 50 mM Tris-HCl (pH 8), 10 mM EDTA |
| TELT | 5ml 1 M Tris-HCl (pH 7.5), 12.5 ml 0.5 M EDTA (pH 8), 50 ml 5 M LiCl, 0.4 ml Triton X –100, 32.1 ml $H_2O$ to total 100ml |

All the media, buffers and solutions, unless specifically mentioned, were prepared with double distilled $H_2O$ and sterilised through autoclaving as per the standard molecular biological procedures (Sambrook et al., 1989).

## 2.2. Methods

### 2.2.1. General methods

All the general molecular biological methods, unless specifically mentioned, were carried out under sterile conditions, using sterilised solutions prepared with double-distilled water (see 2.1.1) and based on Sambrook et al. (1989).

### 2.2.2. Plant growth

All the maize (*Zea mays* L.) lines used were grown in the greenhouse with 16 hours light at 24°C and 8 hours dark at 16°C, and under 55-95% relative humidity. The light intensity was maintained at 24,000-25,000 Lux.

For Northern analyses, some kernels were first germinated in the dark at 30 °C or RT, before placing them over wet paper towels and allowing them to grow in the greenhouse. These were then used for collecting root samples and partly also the leaves.

### 2.2.3. Bacterial cultures and bacterial glycerol stocks

All the bacteria were grown in LB medium as per Sambrook et al. (1989). Solid LB medium was prepared by adding 1.5% (w/v) agar. For selection of the plasmids, suitable antibiotics were added in the appropriate concentration. The optical density of a culture was measured using a Uvicon 922 spectrophotometer from Kontron, at 600 nm wavelength.

For long-term storage, a bacterial culture stock was prepared in 15% glycerol, shock-frozen in liquid $N_2$ and stored at –70°C to – 80°C.

### 2.2.4. Preparation of competent *Escherichia coli* cells

For transformation, the XL1-Blue *E. coli* strain was used to prepare competent cells with the $CaCl_2$ method. 100 ml LB broth was inoculated with 1ml overnight LB-tetracycline culture at 37°C in a shaker and allowed to grow up to an $OD_{600}$ = 1. These cells were then cooled for two hours on an ice-cold water bath, then centrifuged at 1500 × g for 10 min. at 4°C and the pellet thoroughly resuspended in 7.5 ml of a sterile-filtered transformation Buffer, TfB I (100mM RbCl, 45mM $MnCl_2$, 35mM KAc, 10mM $CaCl_2$, 0.5 mM LiCl, 15% Glycerine; pH with HCl to 5.8). The resuspended cells were put on an ice water bath for approximately 10 min. to cool. Again, the cells were centrifuged down as before and the pellet was gently resuspended in 4ml of a second sterile-filtered transformation Buffer, TfB II (10 mM MOPS, 10 mM RbCl, 10 mM $CaCl_2$, 15% Glycerine; pH with NaOH to 7.0), as mentioned before. Keeping the cell suspension at 4°C, 100 µl aliquots were placed into 1.5 ml microfuge tubes and shock frozen in liquid nitrogen. These competent cells were stored at –80°C until used. The competent cells were thawed on ice, just before transformation.

### 2.2.5. DNA Analysis

#### 2.2.5.1. Isolation of plasmid DNA from bacterial culture

Two different methods were used for mini-preparations of Plasmid DNA from bacteria. For general use like restriction digests, cloning etc, the TELT method was used. For sequencing, DNA purification was performed using a Nucleospin Plasmid DNA Purification kit or Concert™ Rapid Plasmid mini-prep kit (see 2.1.4).

#### 2.2.5.1.a. TELT DNA Mini-preparation :

Single colonies were cultured in 2 ml LB-antibiotic broth overnight at 37°C and 180 rpm in a shaker. 1.5 ml of each culture was pelleted by centrifuging in microfuge tubes for 2 min. at ≥12,000 rpm in a table centrifuge. The supernatant was removed and the pellet was resuspended in 200 µl TELT solution by vortexing. To lyse the bacteria, 20 µl Lysozyme solution (10 mg/ml in TE II) was added to the sample and allowed to shake for 5 min. at RT. To completely lyse the bacteria, the sample was placed at 95°C in a heating block for 3-5 min. followed by 5 min. on ice. The sample was then centrifuged at 15,000 rpm for 15 min. at 0-4°C. The pellet of lysed cell debris was removed with a sterile toothpick. To the supernatant, 100 µl isopropanol at RT was added. The plasmid DNA was pelleted by centrifuging at 15,000 rpm for 20 min. at 0°C. In order to remove excessive salts in the DNA pellet, it was carefully washed with 70% ethanol (–20°C), and then centrifuged for 10 min. The ethanol was removed and the pellet, following air-drying, was resuspended in 40 to 50 µl TE I.

#### 2.2.5.1.b. Plasmid Midi-preparation :

A 50 ml bacterial culture medium (LB + antibiotic) was allowed to shake at 37°C with 180 rpm overnight. It was then centrifuged in a sterile 50 ml tube in a swing out rotor at 3000 rpm for 10 min. at RT. The pellet was resuspended in 9 ml TE II. After adding 1ml lysozyme (10 mg/ml in TE II), the sample was placed on ice for 5 min. Then, 20 ml Solution II (0.2 N NaOH, 1% SDS) was added and it was allowed to cool on ice for 10 min. This step was repeated with 15 ml of Solution III (3M potassium acetate, 11.5% Glacial acetic acid). The sample was then centrifuged at 3,500 rpm for 15 min. at 4°C. In order to remove the cell debris, the supernatant was filtered through Miracloth® and divided into two 50 ml tubes with approximately 20 ml filtrate in each. In order to precipitate the plasmid DNA, 0.6 volume of isopropanol (12ml) was added

to each of the two samples and then put on ice for 10 min. The samples were centrifuged at 3,500 rpm for 15 min. at 4°C. The DNA pellets, after air-drying, were resuspended in 1ml TE II and were placed together in a 2ml microfuge tube. In order to remove RNA, 1 volume of 5M LiCl was added to the sample and the sample was placed on ice for 10 min. Then the sample was centrifuged for 15 min. at 15,000 rpm, at 4°C. The supernatant, with the DNA, was transferred into a 15 ml polypropylene tube. In order to precipitate the DNA sample, 2.5 volume of ethanol (5ml), at –20°C, was added to the sample and the sample was placed at –20°C in the freezer for 15-30 min. This precipitated DNA was centrifuged down in a polypropylene tube at 3,500 rpm for 30 min. at 4°C. The pellet was resuspended in 500 μl TE II and transferred to a 2ml microfuge tube. To the sample, 5 μl RNAse A (10 mg/ml) was added and the sample was then incubated at 37°C for 30 min. Afterwards, in order to remove contaminating proteins, the sample was extracted twice by adding 1 volume of 1:1 phenol:chloroform. For this 0.5 volume of phenol, equilibrated with Tris pH 7-7.8, was added to the sample followed by 0.5 volume of chloroform, mixing thoroughly after each addition, and was centrifuged at 12,000 rpm for 5 min. at RT. To remove traces of phenol from the sample, 1 volume of chloroform was added to the upper aqueous phase in a new tube and was centrifuged again at 12,000 rpm for 5 min. at RT. The upper aqueous phase, containing the DNA, was then precipitated with 0.1 volume of 3 M sodium acetate, pH 5 and 2.5 volume of ethanol(–20°C). This mixture was kept at –20°C for 15-30 min. and then centrifuged down at 15,000 rpm for 30 min. at 4°C. The pellet was washed with 70% ethanol and was centrifuged again at 15,000 rpm for 10 min. at RT. The pellet was air-dried and then resuspended in 100-500 μl of TE I, depending upon the size of the pellet, and stored at –20°C.

#### 2.2.5.2. Isolation of phage DNA

This method was used to isolate genomic Lambda phage clones selected from the Lambda Fix II library (2.2.5.13). For Lambda phage DNA isolation either a phage midi or maxi preparation was done, depending upon the amount of phage DNA needed.

#### 2.2.5.2.a. Phage DNA Midi-preparation

For phage DNA midi preparation, the Lambda Midi Kit (25) from Qiagen was used. The DNA was isolated based on the protocol accompanying the kit with one

additional step, after resuspension in L3 a Proteinase K digestion for 30 min. at 55°C was performed. The enzyme was added to a final concentration of 0.1-0.2 mg/ml.

#### 2.2.5.2.b. Phage DNA Maxi Preparation

For a phage DNA maxi preparation, bacterial host strain XLMRA(P2) was prepared. For this, 50 ml LB broth, supplemented with 0.2% Maltose and 10 mM $MgSO_4$, in a 250 ml conical flask was inoculated with a colony from a freshly streaked plate (XLMRA/P2). The culture was either incubated at 30°C overnight or at 37°c over the day (180 rpm) until an $OD_{600}$ = 1 was reached. The cells were centrifuged down at 3,000 rpm for 10 min. at RT, and the pellet was resuspended in equal volume of 10 mM $MgSO_4$, so that an $OD_{600}$ = 1 was maintained.

For the phage DNA maxi-preparation, a big lysate of the phage was prepared. For this, a sample of the phage mini lysate was diluted to 0.001. A 50 µl sample was added to 2.5 ml of the freshly prepared XLMRA(P2) cells in 10 mM $MgSO_4$ in a tube and was incubated at 37°C for 15 min. to allow infection of the bacteria to occur. The cells were then transferred to 500 ml of LB with 10 mM $MgSO_4$ in a 2 litre conical flask and allowed to incubate in a shaker (180 rpm) overnight at 37°C. The next day, 10 ml chloroform was added to lyse the remaining cells by shaking further for the next 15 min. at 37°C. Then 200 µl DNAse (10 mg/ml) and 100 µl RNAse (10 mg/ml) were added to the flask and it was allowed to shake further for 30 min. at 37°C (or 60 min. at RT) to degrade the bacterial chromosomal-DNA and RNA. Then, 30g of NaCl was mixed into the lysate (1M final concentration) and kept on ice for 60 min. in the cold room. The lysate was then centrifuged at 6,000 rpm in a Sorvall GS-A rotor for 30 min. at 4°C. The supernatant was then filtered through miracloth into a new conical flask. To this, 50 g PEG 6,000 was added and gently mixed in the shaker at RT to make a final concentration of 10% PEG. This was then incubated at 4°C for at least 2 hours, or preferably overnight, and then centrifuged at 6,000 rpm for 30 min. at 4°C. The supernatant was decanted without disturbing the pellet and discarded. The pellet was then gently resuspended in 5 ml of SM buffer and transferred to a 15 ml polypropylene tube. The solution was then extracted twice with 1 volume of chloroform by mixing into the solution and then centrifuging at 3,500 rpm for 5 min. at RT each time. To the upper phase, 0.75 g/ml Cesium chloride ($CsCl_2$) was added and mixed to make a uniform solution and centrifuged in a Centricon T 1170

ultracentrifuge (Kontron Instruments), with a SW-60 rotor at 33,000 rpm for 20 hours at RT (20°C). Afterwards, the phage band in the $CsCl_2$ gradient was isolated with the help of a 200 µl pipette or a pasture pipette. The phages were stored in a microfuge tube at 4°C.

For phage DNA extraction, 100 µl of the phages in $CsCl_2$ were added to 900 µl of 10 mM Tris pH 7.5 with 20 µl 10% SDS. This solution was then incubated with 50 µl Proteinase K (5mg/ml) at 37°C for 20 min. in a water bath, so that the phage protein coat was degraded to release the phage DNA. It was then placed at 52°C for 10 min. to denature the enzyme. It was extracted twice with 1 volume of 1:1 phenol:chloroform and once with 1 volume of chloroform (Sambrook et al., 1989; section 2.2.5.1.b). The upper phase containing the DNA was then precipitated with 2 volumes of ethanol (–20°C) in a corex glass tube (Since enough CsCl salt was already present in the solution, no further addition of salt was required for the precipitation). The precipitate was then centrifuged at 10,000 rpm for 30 min. in a Sorvall HB-6 rotor at 4°C. The pellet was then washed with 70% ethanol (–20°C) and centrifuged again for 10 min.. It was then air-dried and resuspended in 200 µl TE I pH8, from which a 2-5 µl sample was run on an agarose gel for a quality check.

#### 2.2.5.3. Isolation of maize genomic DNA

The genomic DNA from maize leaves was isolated based on the method of Dellaporta et al. (1983). As a variation, in some cases, 4g of plant material was extracted with 20 ml of extraction buffer (100 mM Tris, pH 8, 50 mM EDTA, pH 8, 500 mM NaCl ; 0.1% β-mercaptoethanol, added just before use) and 1.5 ml of 20% SDS. The volumes of the solutions in the following steps of the protocol were also adjusted to the extraction buffer volume.

#### 2.2.5.4. Estimation of nucleic acid concentration

For estimating genomic DNA and RNA concentrations, a spectrophotometric measurement was done using the Gene-Quant Spectrophotometer, DNA/RNA Calculator from Pharmacia Biotech (Sambrook et al., 1989). The calculation was based on the principle that an absorbance value for double stranded DNA at 260 nm, $A_{260}$ = 1, is equivalent to 50 µg/ml, whereas for RNA is 40 µg/ml. For estimating the amount of protein contamination, as a measure of the quality of the nucleic acid

preparation, the ($A_{260}/A_{280}$) quotient was calculated. The quotient had to lie between 1.8 and 2.0 for a good quality preparation.

For estimating concentrations of smaller amounts of DNA, like plasmid DNA, the intensity of fluorescence of ethidium bromide bound to DNA, in the presence of UV light, was compared to that of a standard DNA marker, like λ EcoRI/Hind3 marker (Roche) or the Smart ladder (Eurogentec) using a UV transilluminator.

#### 2.2.5.5. DNA restriction digest

Restriction endonucleases were used according to the manufacturer´s recommended protocol (see chapter 2.1.1). When using TELT plasmid DNA preparation, 5 µl RNAse A (10mg/ml) was added to the restriction digest. The restriction digest of genomic DNA was carried out with 2 to 5 U of a restriction enzyme per µg genomic DNA, or as per the suggestions of the manufacturer. In total, 10 µg genomic DNA was used and the restriction digest was carried out in 40 to 50 µl final volume for 3 to 5 hours at 37°C or as per the requirement of a particular restriction enzyme.

#### 2.2.5.6. DNA agarose gel electrophoresis

Agarose gels were used for horizontal gel electrophoresis, which was carried out to separate DNA fragments based on their molecular weight (Sambrook et al. 1989). Depending upon the need, the gels with 0.5 to1.5% agarose were prepared in 1× TAE and 0.5% Ethidium bromide. 0.1 volume of loading dye (50% glycerine, 50% 1× TAE and 0.25% w/v Orange G or a combination of 0.25% bromophenol blue and 0.25% xylenecynol as colour markers) was added to the DNA samples before loading on the gel. For estimating the size of the DNA fragment, 1 kb ladder and Smart DNA ladders (see 2.1.2) were used. The agarose gels were run at 30-120 Volts. The DNA was observed under a UV Transilluminator (wavelength 304 nm), where orange fluorescence of ethidium bromide, physically bound to the nucleic acid, made it visible. This could be documented with the help of photographs taken with a camera attached to the transilluminator unit (Herolab Company).

### 2.2.5.7. Isolation of DNA fragments from agarose gel

Two different methods were used to isolate DNA fragments from agarose gel. In one method, the gel piece with the DNA was placed on glasswool held at the bottom of a 1ml pipette tip or a 1.5 ml microfuge tube with a hole. This was again placed over an intact 1.5 ml microfuge tube. This unit was centrifuged at 12000 rpm for 5 min. at RT. The eluate was extracted twice with 1 volume of 1:1 phenol: chloroform and then once with 1 volume of chloroform. The DNA in the upper aqueous phase was precipitated with sodium acetate/ ethanol (Sambrook et al., 1989; see also 2.2.5.1.b). The pellet was resuspended in an appropriate amount of $H_2O$ or TE I.

In another method, the Gel band Isolation kit from Amersham Pharmacia Biotech was used. This method was also used for directly isolating PCR amplification products. The DNA fragments were isolated as per the accompanying protocol.

### 2.2.5.8. PCR amplification of DNA fragments

The following standard reaction mix was used for a 50 µl PCR reaction (Rybicki, 1996):

| | |
|---|---|
| Template DNA | 10-50 ng plasmid DNA, 100ng genomic DNA |
| 5´ Primer | 0.2 µg |
| 3´ Primer | 0.2 µg |
| dNTPs | 0.2 mM |
| PCR Buffer (10×, Gibco BRL) | 1× |
| $MgCl_2$ (50mM, Gibco BRL) | 1.5 mM |
| Taq DNA Polymerase | 1 U |
| $H_2O$ to | 50 µl |

The Polymerase chain reactions were carried out in a Trio-Thermoblock (Biometra). The general PCR cycles used for plasmid templates, depending on the case, were as follows (Rybicki, 1996):

| | | |
|---|---|---|
| 1. Hot start | 94°C, 3 min. | |
| 2. Addition of Taq Polym. | 85°C, ∝ | |
| 3. Denaturation | 94°C, 3 min. | 1 cycle, Only for genomic DNA as template |
| 4. Annealing | 55-70°C, 3 min. | |
| 5. Elongation | 72°C, 1-2 min. | |
| 6. Denaturation | 94°C, 30-45 s | 30 to 35 cycles |
| 7. Annealing | 55°C-65°C, 30-45 s | |
| 8. Elongation | 72°C, 1-2 min | |

(continued on the next page)

| | | |
|---|---|---|
| 9. | Elongation | 72°C, 10 min. |
| 10. | End | 4°C, ∝ |

The annealing temperature was selected in the range of 2-6°C higher or lower than the melting temperature, $T_m$. An approximate $T_m$ value was calculated as follows:

$$Tm = \{4\times (G+C)\}+\{2\times(A+T)\}\ °C, \text{ and}$$

$$T_{PCR} = (Tm-5)\ °C$$

The PCR amplifications were carried out in 0.5 ml microfuge tubes and were covered with a layer (2 drops) of mineral oil (Sigma) to avoid the water evaporation.

### 2.2.5.9. Cloning of DNA subfragments and PCR products

DNA fragments isolated and purified as described in 2.2.5.7 were either taken directly for cloning or, when needed, the ends were made blunt with the Klenow fragment of DNA Polymerase I by a fill-in reaction (Sambrook et al., 1989) and then cloned. PCR fragments had a 3´ A-overhang and did not contain the 5´phosphate group (Gibco BRL). Therefore, unless the Topo® cloning kit was used, the PCR fragments were first blunted by a fill-in reaction before cloning into a vector whose 3´ends had not been dephosphorylated.

Sticky end fragments were either cloned into vectors having the same sticky end sequence or they were made blunt. For the latter, a dephosphorylated blunt ended vector was used (Sambrook et al., 1989). Cloning into the pZErO-II vector was based on the instructions accompanying the vector. For selection, the plasmid vector contained the lethal gene, ccdB, which only allowed the growth of cells containing plasmids with an insert that disrupted the ccdB gene.

#### 2.2.5.9.a. Ligation of DNA fragments

For ligation, T4 DNA ligase was used. The approximate molar ratio of insert DNA to vector DNA was made to 3:1. The ligation was either carried out at room temperature for 2-4 hours or at 14°C overnight.

When using the Topo® TA cloning kit, the PCR product was purified on a GFX column and cloned into pCR2.1® vector as per the accompanying manual.

### 2.2.5.9.b. Transformation of *E.coli* cells

For transformation, the ligation reaction was added directly to 100 μl of competent cells (see 2.2.4) that had been thawed on ice. This mixture was kept on ice for at least 30 min. and then subjected to a heat shock, either at 37°C for 5 min. or at 42°C for 1 minute. The sample was then immediately placed on ice for up to 5 min. LB broth (250 μl) was added to the sample and it was allowed to shake for 1hour at 37°C. Afterwards, different volumes were plated on three different LB plates (containing the proper antibiotic). The inverted plates were allowed to grow overnight at 37°C.

The next day, 12-24 colonies were picked with a toothpick and, after streaking on a copy plate, were cultured in 2 ml LB with antibiotic overnight. These samples were first used for TELT-DNA minipreps (see 2.2.5.1) and analysed either through restriction digest or PCR using specific primers. The putative positive clones were selected for sequencing (see 2.2.5.10). Glycerol stocks were prepared from positive clones and stored at –80°C.

### 2.2.5.10. DNA sequencing

The DNA sequencing method was based on the Sanger Dideoxy method (Sanger et al., 1977). For sequencing flourescently labelled ddNTPs were used (Perkin Elmer/Applied Biosystems). The sequence reaction was prepared in 20 μl final volume as follows:

| | | |
|---|---|---|
| Template plasmid DNA | | 300-500 μg |
| Primer | | 15 pmol |
| Big Dye Terminator | | 3 μl |
| Half Term buffer | | 5 μl |
| $H_2O$ | to | 20 μl |

The sequence reaction cycle used is as follows:

| | | | |
|---|---|---|---|
| 1. | 96°C | 4 min. | |
| 2. | 96°C | 30 s | 25 cycles |
| 3. | 50°C | 15 s | |
| 4. | 60°C | 30 s | |
| 5. | 4°C | ∝ | |

After the reaction, the sample was precipitated with 35 μl ethanol (–20°C) by centrifuging at 12,000 rpm for 20-30 min. at RT. The pellet was allowed to air dry for

5 min. The sample was resuspended in appropriate loading dyes suitable for the sequence gel prepared. The nucleotide sequence of this sample was determined using the ABI Prism™ automated DNA Sequencer model 377A (Perkin Elmer/Applied Biosystems).

#### 2.2.5.11. *In silico* sequence analysis and database searches

Sequence data was analysed with the following programs:

1. Editseq (DNA STAR Vers.3.6, DNA Star Inc., USA)
2. Seqman (DNA STAR Vers.3.6, DNA Star Inc., USA)
3. Mac Vector™ 4.1.4 for Apple Macintosh computer (Scientific Imaging System, USA)
4. Clustal X (Thompson et al., 1997)
5. Clustal W (Thompson et al., 1994)
6. Genedoc Vers.2.5.000 (Nicholas et al., 1997)
7. MultAlin (Corpet, 1988)

The online BLAST® server, of the National Centre for Biotechnology Information (NCBI), was used to compare the sequence data to the GenBank, EMBL, DDBJ, dbEST, PIR, PRF, SWISS-PROT/TrEMBL and PDB databases.

The online ExPASy Molecular Biology server, of the Swiss Institute of Bioinformatics (SIB) at http://www.expasy.ch/ , was used to further characterise the *et1* and *zmzr1* genes as well as their predicted proteins using the proteomics tools.

The following programs were used to characterise the *et1* genomic and protein structure:

1. BLAST 2.0.10 - 2.2.1 (Altschul et al., 1997)
2. PlantCARE (Rombatus et al., 1999)
3. ChloroP 1.1 (Emanuelsson et al., 1999)
4. TargetP v1.01 (Emanuelsson et al., 2000; Nielsen et al., 1997)
5. PSORT- version 6.4 (Nakai et al., 1992)
6. Predotar -version 0.5 (Small et al., 2000)
7. PredictProtein (PP) (Rost, 1996)
8. ProDom (Corpet et al., 1998; Rost et al., 1994)
9. Profile Scan (PROSITE) (Hofmann et al., 1999; Bucher and Bairoch, 1994)

In addition, secondary protein structure analyses were carried out using different methods available at the ExPASy server under:
http://www.expasy.ch/tools/#secondary and http://www.expasy.ch/cgi-bin/protscale.pl

### 2.2.5.12. Radioactive labelling of DNA fragments

Radioactively labelled DNA probes were prepared using $\alpha^{32}$P-dCTP and 9-mer random primers. For this, the Prime-It® II Random primer Labelling Kit from Stratagene (Heidelberg) was used. For each labelling reaction, 25 ng DNA and 35 to 50 μCi $\alpha^{32}$P dCTP (3,000Ci/mmol) were used. Probes for hybridising plasmid blots were labelled for approximately an hour, whereas probes for Northern or genomic Southern blots were labelled for 3 to 5 hours, before hybridising to the blots. Unincorporated radioactive dCTPs were removed by one of the following methods. In one case, the sample was loaded over a Sephadex S-25 column and centrifuged for 5 min. at 2,000 rpm. Alternatively, the sample was loaded over a MicroSpin™ S-300 HR column from Pharmacia Biotech (Freiburg) and, as per the instructions accompanying the kit, the sample was centrifuged at approximately 730 ×g for 2 min. The sample was heated at 94°C for 10 min. to denature the labelled DNA, and then immediately added into the hybridisation solution in the bottle containing the blot(s) to be hybridised.

### 2.2.5.13. Southern blot transfer of DNA

The method of capillary transfer of DNA from a gel to a membrane was carried out according to Sambrook et al. (1989).

#### 2.2.5.13.a. Southern blot transfer of phage and plasmid DNA

After the separation of DNA fragments on an agarose gel (see 2.2.5.6), the gel was capillary blotted to a Hybond™ N+ nylon membrane using 0.4 M NaOH buffer. The set-up was made as per Sambrook et al. (1989). After blotting, the membrane was crosslinked with UV light at 120,000 μJ/cm$^2$ using the appliance UVC 1,000 (Hoefer, San Francisco). The membrane was then washed for 30 min. in 2× SSC and then used for hybridisation.

### 2.2.5.13.b. Southern blot transfer of genomic DNA

For blotting genomic DNA from an agarose gel, the DNA in the gel was first denatured with a denaturing solution, Southern I (0.5 M NaOH, 1.5 M NaCl), with gentle agitation in a tray for 45 min. After rinsing the gel with water once, the gel with the denatured DNA was treated with a neutralising solution, Southern II (0.5 M Tris-HCl, pH7.5, 1.5 M NaCl), by agitating gently for 1 hour. After approximately 30 min., the Southern II solution was changed. This gel was then taken for blotting onto Hybond™ NX membrane in 10× SSC buffer as per Sambrook et al. (1989). After cross-linking, the blot was rinsed with 2× SSC.

### 2.2.5.14. Genomic library screening

Before screening the genomic λ Fix II library, the titer of the library was determined in pfu/ml. At least a million clones were screened. For the first round, 24 large (150 mm) LB plates supplemented with 10 mM $MgSO_4$ were prepared. Fresh bacteria (strain XLMRA/P2) were prepared according to 2.2.5.2.b.

Approximately 50,000 clones were mixed with 600 µl of XL1 Blue MRA(P2) cells ($OD_{600}$ =1) for each plate and incubated in a water bath at 37°C for 15 min. Each sample was then mixed with 9 ml of SM Top agarose, pre-warmed at 50°C, and plated on the pre-warmed LB plates supplemented with 10 mM $MgSO_4$. After the top agarose had hardened, the plates were closed, inverted and incubated at 37°C overnight. Later on, the plates were stored at 4°C.

For performing plaque lifts, a nitrocellulose filter of the size of the plate was placed over each plate. It was allowed to incubate for 2 min. on the plate. During this, the filter was labelled with a ballpoint pen and for orientation, at three different positions near the periphery, one, two and three fine holes respectively were made through the filter into the plate using a needle. The filters were then placed with the plaque side up on Whatman 3mm paper soaked in Southern I (0.5 M NaOH, 1.5 M NaCl) for 2- 5 min. They were then transferred to a Whatman 3mm paper soaked in Southern II (0.5 M Tris-HCl, pH7.5, 1.5 M NaCl) for 5-10 min. Afterwards, the filters were submerged in 0.2× SSC solution for 10 min. These plaque lifts were then placed between 2 dry sheets of Whatman 3mm paper and baked in an oven at 80°C for 2 hours to crosslink the DNA to the filters. Before hybridisation, the filters were placed in a 0.2×

SSC/0.1% SDS solution at 65°C with gentle agitation for 30 min. and then hybridised (see 2.2.5.14).

After carrying out autoradiography, the radioactivity signals observed on the X-ray film were compared to their plates. The area on the plate, corresponding to the radioactive signal, was cored out with the help of the broad end of a Pasteur pipette and added to 1 ml of SM buffer in a microfuge tube. To this, 50 µl chloroform was added and it was stored at 4°C overnight or until needed.

For the second round screening, 100 mm plates were used and an aliquot from the first round pick in SM (~2,000 pfu) was mixed with 200 µl freshly prepared XLMRA(P2) cells using 4 ml SM top-agar. The rest of the procedure was the same as for the first round.

For the third round, the same procedure was carried out as for the second round, except that approximately 100 pfu/plate were plated, so that isolated pfu were observed, which were then cored with the smaller end of the Pasteur pipette. These were the pure isolates, from which phage DNA was prepared and analysed (2.2.5.2).

### 2.2.5.15. DNA Hybridisation and autoradiography methods

The membrane to be hybridised was first prehybridised for 2 to 3 hours at 65°C in a solution containing salmon sperm DNA as a blocking agent (see 2.1.10, $^{32}$P Hybridisation solution). Depending upon the size of the blot and the tube in which the hybridisation was carried out, approximately 10 ml /100 cm$^2$ of the pre-warmed hybridisation solution was taken. After prehybridisation, the probe was added to the same solution (see 2.2.5.11) and the hybridisation was carried out overnight in a hybridisation oven.

For hybridising plaque lifts, the lifts, after washing in 0.2× SSC/0.1% SDS solution (2.2.5.13), were stacked with the DNA side down in a round container with a lid, containing pre-warmed hybridisation solution. Air bubbles between filters were removed. The lifts were covered in the hybridisation solution and prehybridised for 2-3 hours at 65°C with gentle agitation in a water bath. Then, tilting the pan, the denatured probe was added into the solution. The pan was then gently swirled and closed. It was then placed back in the water bath and hybridised overnight.

The next day the hybridisation solution with the probe was decanted. The blot in the hybridisation bottle was rinsed once with 2× SSC/0.5% SDS solution at RT. The blot was then washed with pre-warmed 0.2x SSC/ 0.5% SDS for 15 min. at 65°C. This step was repeated with fresh pre-warmed solution. The same was done with the nitrocellulose filters in the water bath.

After the membranes had been washed, the specific activity on the membranes was measured with a Geiger counter. Nitrocellulose lifts were placed on older X-ray films for support and sealed with Saran Wrap. These were then exposed to X-ray films (Hyperfilm MP, Amersham, Braunschweig) in a cassette with intensifier screens in the dark at –80°C. Depending on the intensity of the counts, the blots were either exposed to X ray films or Phosphoimager plates (Fujifilm Imaging plate, BAS-MP 2025P, Rayset, Straubenhardt). The Phospho-imager plate data was read and analysed with the program PCBAS 2.09g (Rayset, Straubenhardt) on the computer.

Before hybridising a new probe to the blot, the older radioactive probe was removed from the membrane by adding boiling hot 1% SDS in a deep vessel containing the blot and rocking it for 30 min. at 80°C. After approximately 15 min., the specific activity on the membrane was checked and the solution was changed.

### 2.2.5.16. AIMS Analysis

A modified version (Lauert et al., 1999) of the AIMS analysis (Amplification of Insertion Mutagenised Sites; Frey et al., 1998) was used to analyse the *et1-Ref* genomic background and check the origin of the putative *et1-Ref λ 1.1* allele. For comparison, two samples from LC, four samples from *et1-m3*/*et1-R* and *et1-m15/et1-R* mutant lines each were used as controls along with four test samples from *et1-Ref*. In addition, the two genomic λ clones, *et1-Ref λ 1.1* and *et1-m3 λ 5.1* were used as positive controls.

A sample of genomic DNA from different plants from all the lines were diluted to 50 ng/µl and their concentrations tested on an agarose gel. 500 ng sample in 10 µl from each plant was then digested with *Hapa*II in total 25 µl volume at 37°C. This digested DNA was then taken for ligation to enzyme-specific adapters by adding 5 µl ligation mix to the digest. The ligation was carried out for 2 hours at 22°C. The digestions and ligations were carried out as shown in the tables below:

| (A) Digest (1hr, 37°C) | |
|---|---|
| DNA (500 ng) | 10 µl |
| Enzyme (Hpa II, 10U/µl) | 0.5 µl |
| 10 × Buffer (1) | 3.5 µl |
| $H_2O$ to | 25 µl |

| (B) Ligation (2hr, 22°C) | |
|---|---|
| Adapter (50 µM) | 0.8 µl |
| 5× ligase buffer | 1.6 µl |
| T4 DNA ligase (high conc.) | 0.2 µl |
| $H_2O$ | 2.4 µl |
| Digest | 25 µl |

Afterwards, the AIMS specific PCR reactions were performed (see tables below). For this the ligation reaction was diluted 1:2 and from this 15 µl sample was added to 10 µl of the PCR 1 mixture. For the PCR 2 amplification, only 2.8 µl sample of the 1:10 diluted PCR 1 was added to 25 µl of the PCR 2 mixture containing fluorescent adapter and *Mu* primers. PCR 2 (10%) was loaded on a vertical acrylamide sequence gel and analysed in the ABI Prism™ automatic Sequencer type 377A (Perkin Elmer/Applied Biosystems) with the program GeneScan™.

The PCR reactions were prepared as shown below:

| PCR 1 | | PCR 2 | |
|---|---|---|---|
| *Mu*-1 (10 µM) | 1.25 µl | *Mu*-2 (10 µM) | 2.8 µl |
| AD-1 (10 µM) | 1.25µl | AD-2 (10 µM) | 2.8 µl |
| 10× KCl buffer | 2.5 µl | 10× KCl buffer | 2.8 µl |
| dNTPs (2.5 mM) | 2.5 µl | dNTPs (2.5 mM) | 2.8 µl |
| Pt. Taq Polym. (10U/µl) | 0.1 µl | Pt. Taq Polym. (10U/µl) | 0.2 µl |
| $H_2O$ | 2.5µl | $H_2O$ | 13.7 µl |
| Ligation mix. (1:2 diluted) | 15 µl | PCR 1 mix. (1:10 diluted) | 2.8 µl |

The amplification cycles used for both the PCRs are as follows:

| PCR1 | PCR2 |
|---|---|
| 1. 94°C 3 min.<br>2. 94°C 1 min.<br>3. 65°C 30 s } 20 cycles<br>4. 72°C 1 min.<br>5. 72°C 3 min.<br>6. 4°C ∝ | 1. 94°C 3 min.<br>2. 94°C 1 min.<br>3. 65°C 30 s } 31 cycles<br>4. 72°C 1 min.<br>5. 72°C 5 min.<br>6. 4°C ∝ |

After analysing the amplification, a PCR 2 reaction was carried out for selected samples using chemiluminescent Digoxigenin-labelled *Mu*- and adapter-primers. From this, approximately 30% of the sample was run on an acrylamide gel along with

a DNA ladder, DNA Marker V (Roche, Mannheim), and transferred to a nylon membrane (PALL Biodyne® A) using two sheets of 3mm Whatman paper (for capillary action) and a moderate weight laid over the membrane for 1 hour. The membrane was crosslinked twice at 120,000 µJ/cm$^2$. The Dig detection was carried out using alkaline phosphatase conjugated Dig antibody, Anti-Dig Fab fragments, and CSPD® as the chemiluminescent substrate (Roche). For this, the membrane was first washed with B1 (0.1 M Malic acid, 0.15 M NaCl pH 7.5 with NaOH, 0.3% Tween) for 5 min., followed by incubation in B2 (B1+ 1% blocking agent) for 30 min., followed by B2+ Dig antibody (1:10,000) for 30 min. The membrane was then washed 3 times with B1 for 15 min. each, followed by incubation in B3 (0.1 M Tris pH 9.5, 0.1 M NaCl, 0.05 M $MgCl_2$) for 10 min. To this point, the procedure was carried out with gentle agitation on a rotary platform. Then, approximately 25 ml of CSPD/B3 (1:100) was applied to the membrane in a plastic bag, without air bubbles, and allowed to stand for 5 min. The solution was then decanted and the membrane, after sealing in a new plastic bag, was incubated at 37°C for 15 min. The membrane was then exposed to an X-ray film at RT.

After this primary analysis, the membrane was first hybridised to the LC cDNA radioactive probe, c9.1, followed by the 2.5 kb (*Xho* I/*Hind* III) *et1* probe.

### 2.2.6. RNA Analysis

#### 2.2.6.1. Isolation of total RNA

The plant RNA extraction was carried out using two different methods. The first one, based on Dunsmuir et al. (1988), was used for extracting RNA from larger amounts of tissue, whereas the other, trizol method, based on Chomecynski et al. (1987), was used for smaller quantities. Throughout the isolation procedure, only DEPC-treated $H_2O$ was used and all the solutions were prepared in DEPC-treated $H_2O$.

##### 2.2.6.1.a. RNA Isolation

The following protocol (based on Dunsmuir et al., 1988) describes the preparation for 5 g of plant material. For smaller amounts, all the reagents were proportionally reduced.

The plant material was shock-frozen in liquid nitrogen and then ground into a fine powder using a pestle and mortar. The still-frozen powder was transferred into a 50 ml tube. To this, 10 ml lysis buffer (100 mM Tris-HCl pH 9, 100 mM NaCl, 20 mM EGTA, 2% SDS) was added and the tissue was allowed to thaw. Then, 5 ml phenol and 100 µl β-mercaptoethanol was added and it was agitated gently for 5 min. on a wheel. After addition of 5 ml chloroform, it was agitated for further five min., followed by centrifugation at 3,500 rpm for 10 min. at RT. The upper phase was transferred into a new 15 ml tube and extracted two more times with phenol:chloroform as described above. However, this time it was carried out without β-mercaptoethanol. The upper phase was then taken in a new tube and precipitated with 0.1 volume 3 M sodium acetate (pH5.2) and 2 volumes ethanol by placing it at –80 °C for 30 min. The solution was centrifuged at 3,500rpm for 30 min. at 4°C. The pellet was then air-dried and resuspended in 5 ml $H_2O$. 1 volume (5 ml) of 4 M LiCl was added to precipitate the RNA. After incubating it on ice for at least 3 hours or overnight, the sample was centrifuged at 3,500 rpm for 30 min. at 4°C. The pellet was resuspended in 1 ml $H_2O$ and again precipitated with 0.1 volume 3 M sodium acetate (pH5.2) and 2 volumes ethanol by placing the precipitate at –80°C for at least 30 min. This was then centrifuged at 3,500rpm for 30 min., at 4°C. The pellet was washed with 70% ethanol and then, depending on the size of the pellet, resuspended in 100-500 µl $H_2O$. The RNA concentration was determined using a spectrophotometer (see 2.2.5.4).

#### 2.2.6.1.b. Trizol method of RNA Isolation (Chomecynski et al., 1987)

200 mg of shock-frozen plant material was ground into a fine powder and transferred to a 2 ml microfuge tube. This material was then thawed in 1 ml trizol by hand shaking. It was then incubated for 2-3 min. at RT and then centrifuged in a pre-cooled centrifuge at 14,000 rpm for 15 min. at 4°C. The supernatant was then transferred to a new 1.5 ml microfuge tube. Isopropanol (500 µl) was added to the sample and incubated at RT for 10 min. The sample was then centrifuged again for 10 min. at 14,000 rpm, at 4°C. The pellet was washed with 75% ethanol and centrifuged at 9,000 rpm for 5 min. at 4°C. The pellet, after air-drying, was resuspended in 100 µl $H_2O$. When required, the sample was placed at 55°C for a short while and then at RT. This was repeated until the pellet dissolved. This sample was then precipitated with 0.2 volume of sodium acetate (3 M, pH5.2) and 100 µl of ethanol on ice for at least 60 min. or overnight at 4°C. This was then centrifuged at 10,000 rpm for 10 min.

The supernatant was discarded and the pellet was washed with 80% ethanol. This was then centrifuged at 9,000 rpm for 5 min. at 4°C. The pellet was air-dried for 5-10 min. and then resuspended in 100 µl $H_2O$.

#### 2.2.6.2. RNA gel electrophoresis

In order to carry out Northern analysis, total RNA samples were separated on an agarose gel (Sambrook et al., 1989). The gel contained 1× MEN, 1% agarose and 7% (v/v) formaldehyde. Ethidium bromide was either added to the gel (1 µg/ml) or to the RNA sample (1 µg) before loading on the gel. The RNA sample was prepared by adding 15 µl RNA loading dye to 20 or 30 µg RNA and 1 µg (1 µg/µl) ethidium bromide, when not added directly to the formaldehyde gel. The samples were denatured before loading by incubating them at 65°C for 15 min. The gel was run overnight in 1× MEN, at 20-30 volts.

#### 2.2.6.3. Northern blot transfer of RNA

The method of Southern (1975) of capillary transfer was used to transfer RNA from a gel to a nylon membrane, Hybond NX, in 10× SSC buffer. After blotting, the membrane was crosslinked and then rinsed with 2× SSC once. This blot was then stained with methylene blue stain (0.05% methylene blue, 300 mM sodium acetate, pH 5.2) for 10 min. with gentle agitation on a rotary platform, followed by de-staining and washing in $H_2O$ until the RNA ladder and the rRNA bands in the samples became visible.

#### 2.2.6.4. Hybridisation and autoradiography of Northern blots

The Northern blot, after prehybridisation for 3 to 5 hours, was hybridised with a radioactively labelled DNA probe (see 2.2.5.11) at 68°C overnight. For this, the same hybridisation solution was used as for DNA hybridisation (see 2.2.5.14). After hybridisation, the solution was decanted and the blot was first rinsed with 2× SSC. Then it was washed with 1× SSC for 10 min. at 68°C, followed by 1× SSC/1% SDS for 15 min. at 68°C. The autoradiography procedures were the same as for the Southern blots (see 2.2.5.14).

### 2.2.7. complementary DNA library preparation

For preparing an *et1-Ref* cDNA library, developing kernel samples were collected from selfed *et1-Ref* cobs at 15, 18 and 22 DAP. First total RNA was isolated from each of these samples. Afterwards, mRNA was isolated from equal amounts of each of the total RNA samples, followed by complementary DNA (cDNA) synthesis. This was then cloned into a modified Lambda phagemid vector, which was then packaged into the Lambda phage protein coat to give a primary cDNA library. This library was then amplified and used for screening.

#### 2.2.7.1. Isolation of poly A+ mRNA from total RNA

For this method, the Dynabeads® mRNA Purification Kit (Dynal®, Hamburg) was used. The isolation of poly A+ mRNA was carried out using the principle of magnetic separation. Here the poly A+ mRNA was isolated from total RNA with the help of paramagnetic polymer beads containing covalently attached oligo $dT_{(25)}$ chains. Poly A+ residues at the 3´end of most mRNAs bind to the oligo $dT_{(25)}$ chains on the paramagnetic beads, which are then separated magnetically.

Equal amounts of the total RNA was taken to obtain mRNA from each of these samples. From each total RNA sample, 375 µg (5 x 75 µg) total RNA was taken to isolate mRNA with the Dynal protocol. For each sample, the mRNA was collected in a total of 50 µl DEPC-$H_2O$. The mRNA concentration was calculated by measuring the optical density of each mRNA sample in a spectrophotometer. From the 15, 18 and 22 DAP kernels 196.25 µg/ml, 197.5 µg/ml and 205 µg/ml mRNA could be obtained, respectively.

#### 2.2.7.2. cDNA library preparation

In total, 5 to 6 µg mRNA was taken from the *et1-Ref* developing kernels collected at 15, 18 and 22 days after pollination for preparing complementary DNA with the Zap® cDNA synthesis kit from Stratagene (Heidelberg). Some modifications were made in the protocol. No radioactivity was used during the synthesis. Ten percent of each of the $1^{st}$ strand synthesis (i.e., 5 µl) and $2^{nd}$ strand synthesis (i.e., 1 µl) reactions were run on an agarose gel to test the quality of the syntheses. The processing of cDNA fragments was done with the help of gel filtration combined with centrifugation. For this, the cDNA sample was purified on a Sephacryl 400 column (Sigma) made from a

disposable 1 ml plastic syringe with a cotton plug at the bottom. The column was first rinsed twice with 1× STE (100 mM NaCl, 20 mM Tris-HCl pH 7.5, 10 mM EDTA). For this, 300 µl 1× STE was loaded on the column, which was then placed inside a sterile 50 ml tube and centrifuged at 400 rpm in a Hereaus centrifuge with swing out rotor for 2 min. each time. The 60 µl cDNA sample was then loaded on the column, which was then placed over a microfuge tube and centrifuged in a 50 ml tube in the same way as before. The sample was collected in the microfuge tube and set aside. Then, 60 µl 1× STE was loaded over the column, which was then placed over a new microfuge tube in the 50 ml tube and centrifuged as before. The $2^{nd}$ sample was also collected and set aside. This was repeated until 10 such samples had been collected. Ten percent of each sample was run on a very thin 1% agarose gel. The samples showing good cDNA size and amount were selected for cDNA library preparation. The rest of the procedure for cDNA library preparation was done as per the kit protocol.

#### 2.2.7.3. cDNA library amplification

In order to amplify the library, the complete library was plated on 24 LB petri plates (150 mm) supplemented with 10 mM $MgSO_4$ were prepared. Fresh XL 1Blue MRF´ were prepared as described in 2.2.5.2.b.

The primary cDNA library was divided into 24 microfuge tubes. To each 600 µl of XL1 Blue MRF´ cells ($OD_{600}$ =1) were added and mixed. The samples were then incubated in a water bath at 37°C for 15 min. Each sample was then mixed with 9 ml of SM Top agarose at 50°C and plated on the pre-warmed LB plates supplemented with 10 mM $MgSO_4$. The plates, after the top agarose had hardened, were closed, inverted and incubated at 37°C overnight. The plates were then stored at 4°C for approximately an hour. To each of these plates 8-10 ml SM buffer was added and the plates were incubated at 4°C with gentle rocking overnight. Afterwards, the SM buffer from each plate was collected and pooled into a sterile polypropylene tube or glass bottle. The plates were rinsed with 2 ml SM buffer, which was also added to the amplified library in the polypropylene tube. To this 5% (v/v) chloroform was added and mixed by inverting a few times. This amplified library was stored at 4°C.

### 2.2.8. cDNA library screening

#### 2.2.8.1. Phage plating and performing plaque lifts

Two different cDNA libraries were screened during this study: First, the LC cDNA library made from developing kernels collected at 13, 17, 19, 20, 22, 25 and 28 days after pollination. Second, the *et1-Ref* cDNA library described in 2.2.7.2. Before plating the cDNA library, the titer of the library was determined in pfu/ml. For the first round of screening, 24 large LB plates (150 mm) supplemented with 10 mM $MgSO_4$ were prepared. An amount of the cDNA library equivalent to approximately 50,000 clones per plate was plated as described in 2.2.7.3. At least a million clones were screened. After the overnight incubation at 37°C, the plates were chilled at 4°C for at least 30 min.

Plaque lifts were performed as described for genomic library screening (2.2.5.13) except that here XL1 Blue MRF´ cells were used. After the third round, the single plaque isolates were taken directly for *in vivo* excision.

#### 2.2.8.2. Hybridisation and screening

The hybridisation and screening of the clones was performed as described for genomic library screening (2.2.5.14)

#### 2.2.8.3. *in vivo* excision of phagemids

The *in vivo* excision was carried out based on the protocol accompanying the kit. A single plaque was cored out from the agar plate and added to 200 µl freshly prepared XL1 Blue MRF´ cells in a test tube and incubated at 37°C for 15 min. 5 ml LB broth, containing 10 mM $MgSO_4$, was added and incubated at 37°C overnight in a shaker. The next day, by adding 500 µl chloroform to the lysate and incubating further at 37°C for 30 min., the remaining bacteria in the lysate were killed. The mini lysate was centrifuged in a 15 ml polypropylene tube at 3,500 rpm for 20 min. at RT and then decanted into a sterile new 15 ml polypropylene tube. From this, 100-200 µl lysate ($1\times 10^5$ pfu) was added to 1 µl Excessist (M13) helper phage ($\sim 1\times 10^6$ pfu/µl) with 200 µl XL1-Blue MRF´ cells ($OD_{600}$ =1) and incubated in 37°C water bath for 15 min. Then, 3 ml LB broth containing 10 mM $MgSO_4$ was added and it was shaken gently for 2-3 hours at 37°C. The tube was then placed at 70°C for 15 min. to kill the cells

followed by centrifugation at 3,500 rpm for 15 min. at RT to remove the cell debris. Then, 1-100 µl of the supernatant, containing the excised phagemid, was added to 200 µl of freshly prepared SLOLR cells in a microfuge tube and incubated at 37°C for 15 min. This sample was then plated on LB-ampicillin agar plates and incubated overnight at 37°C. The colonies were then taken for DNA mini-preparation (2.2.5.1) and sequence analysis. The positive phagemid clones were stored as bacterial glycerol stocks at –80°C.

### 2.2.8.4. PCR amplification from the cDNA-library phage suspension

The following PCR reaction was designed to amplify specific cDNA clones from a cDNA library. In order to get a good amount of PCR product, the PCR samples had to be re-amplified. The first amplification was, therefore, referred to as PCR I, and the re-amplification as PCR II.

**PCR conditions**

| | | |
|---|---|---|
| phage suspension | 1-5 | µl |
| 10×PCR buffer (Gibco BRL) | 5 | µl (1×) |
| 50 mM $MgSO_4$ (Gibco BRL) | 2 | µl (2 mM) |
| 2.5 mM dNTPs mix. (Gibco BRL) | 4 | µl (200 mM) |
| 10 µM 5´+ 3´ primers | 1+1 | µl (10 pmol each) |
| $H_2O$ to | 50 | µl |
| Taq DNA Polymerase (Gibco BRL) | 0.2 | µl (1U) |

**PCR Cycles** (PCR I and II):

| | | |
|---|---|---|
| 94°C | 4 min. | |
| 85°C | ∝ | |
| 94°C | 45 s | 32 cycles |
| 58-60°C | 45 s | |
| 72°C | 2 min. | |
| 72°C | 10 min. | |
| 4°C | ∝ | |

The Taq DNA Polymerase was added after the “Hot start” at 85°C.

From the *et1-Ref* amplified cDNA library only 1 µl sample was taken for PCR I, as the library had a good titre (more than $10^7$ pfu/ml), whereas from LC cDNA library with a lower titre a 5 µl aliquot was taken. After amplification, 10% sample of the PCR I was run on a gel. From PCR I, the following dilutions were used in PCR II to reamplify the weakly amplified PCR products: 5µl, 1µl, 0.1µl and 0.01µl of PCR I. The PCR

conditions as well as the amplification cycles were the same as for PCR I. In PCR II, stronger bands could be observed in all the dilutions.

### 2.2.9. ETCHED1 Protein analysis

#### 2.2.9.1. Cloning *et1* cDNA into pBAT

In order to carry out efficient *in vitro* transcription of the ET1 protein, the *et1* cDNA was cloned into the pBAT vector (Annweiler et al., 1991, Fig. 2.1) at the *Eco R*V restriction site in the polylinker downstream of the 5´β-Globin sequence. For this, an complete cDNA insert (*Eco R*I/*Xho*I) was isolated out of the original pBluescript SK (-) vector and blunted by a fill-in reaction (2.2.5.9; Sambrook et al., 1989). It was then ligated into the dephosphorylated *Eco R*V site of the pBAT vector and transformed into competent XL1-Blue cells. Through lacZ (blue-white) screening, the positive colonies, containing an insert in the vector (white colonies), were picked. In order to find a cDNA clone in the right orientation, colony PCR was performed. For this, a 5´ vector primer and a 3´ cDNA specific primer were used. The colonies indicating a PCR product of the right size were then taken for sequence analysis to confirm the positive clone.

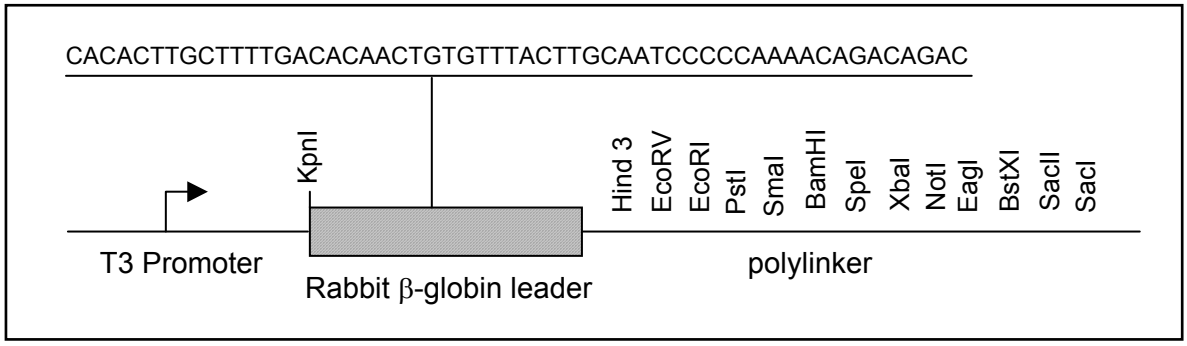



**Fig. 2.1: The multiple cloning site (MCS) of pBluescript KS+, modified to give the new vector pBAT (Annweiler et al., 1991).** The MCS contains the T3 promoter sequence at the 5´ end followed by a rabbit β-Globin leader sequence and then the polylinker. The β-Globin leader sequence enhances the transcription efficiency of a protein with its own start codon.

### 2.2.9.2. *In Vitro* transcription

About 20 µg of the plasmid was first linearised at the 3´end of the cDNA in the vector with Bam HI restriction enzyme and 0.2 µl and 1 µl samples were run along with 1 kb ladder on an agarose test gel:

| **Digest :** | | |
|---|---|---|
| | DNA | 20 µl (20 µg) |
| | 10× Buffer | 20 µl (1×) |
| | Bam HI | 3 µl |
| | $H_2O$ to | 200 µl |

Time: 2 hours; Temp.: 37°C

The rest was extracted once with 1 volume of 1:1 phenol:chloroform, and once with 1 volume of chloroform. This was then precipitated overnight with sodium acetate/ethanol at –20°C. Then, it was centrifuged at 13,000 rpm for 30 min., at 4°C and the pellet was washed once with 70% ethanol and resuspended in 200 µl $H_2O$. From this, a 1 µl sample was run on an agarose test gel.

The *in vitro* translation was based on Sambrook et al. (1989). Both the linearised as well as circular plasmids were used for *in vitro* transcription. Since T3 phage primer was present at the 5´end of the cDNA in the plasmid, T3 RNA polymerase was taken for transcription, which was carried out as follows:

| **Sample no. :** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| Vector | Linearised | Linearised | Circular | Circular | +ve control |
| Vector DNA (µl) | **2.0** | **4.0** | **0.4** | **0.8** | **2.0** |
| DEPC-$H_2O$ (µl) | 3.1 | 1.1 | 4.7 | 4.3 | 3.1 |
| 10× T3 buffer (µl) | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 |
| 5× rNTPs, red. rGTP, (µl) | 5.0 | 5.0 | 5.0 | 5.0 | 5.0 |
| 10 mM cap analogue (µl) | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
| 100 mM DTT (µl) | 7.5 | 7.5 | 7.5 | 7.5 | 7.5 |
| 35U/ µl RNAse Inhibitor (µl) | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 |
| 50U/ µl T3 Polymerase (µl) | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| Total (µl) | **25.0** | **25.0** | **25.0** | **25.0** | **25.0** |

All the reagents were mixed together on ice by stirring with a pipette tip and the reaction was carried out at 37°C for 30 min. Then 1 µl of 11.25 mM rGTP was included in the reaction and the reaction was carried out further for 30 min.

Finally, DEPC-$H_2O$ was added to each of the transcription reaction to make a final reaction volume of 100 µl, from which 4 µl sample were run on a 1% agarose gel along with 1kb ladder. The gel was made by adding 10 mg/ml agarose into 1× MOPS buffer containing 0.05% DEPC which had been incubated for at least 30 min. on a magnetic stirrer at RT. The agarose was then cooked and ethidium bromide was added to a final concentration of 1 µg/ml.

The rest of the reaction volume was precipitated with 5 µl of 4M ammonium acetate and 250 µl ethanol at –20°C, until needed.

#### 2.2.9.3. *In Vitro* translation and radioactive labelling

Aliquots (5 µl and 10 µl) of the precipitated *in vitro* transcription were taken in new microfuge tubes and centrifuged at 13,000 rpm for 30 min., at 4°C. To the pellets, after air-drying, 12.5 µl of the *in vitro* translation master mix was added. All the reagents were mixed gently by stirring to make the master mix.

The reagents were added in the order and amounts as shown below:

| **Master Mix** | |
|---|---|
| **$H_2O$** | 4.35 µl |
| **1 M KCl** | 0.90 µl |
| **100 mM DTT** | 0.25 µl |
| **Amino acid mixture (minus Cysteine)** | 0.25 µl |
| **Reticulocyte lysate** | 6.25 µl |
| **$S^{35}$-Cysteine** | 0.50 µl |
| **Total** | **12.5 µl** |

The reaction was allowed to take place at 30°C for 90 min. in a water bath. Afterwards, the reaction was placed for 2 min. in a boiling water bath, in order to denature the synthesised proteins, and then cooled on ice. For long storage periods, the samples were stored at –80°C.

### 2.2.9.4. SDS-PAGE

For analysing the *in vitro* translation protein samples on a gel, they were centrifuged at 13,000 rpm for 2 min. at RT. One volume of 1× Laemmli buffer (12.5 µl) was added to the sample and mixed by stirring. 12.5 µl of the sample was loaded on a mini SDS-polyacrylamide gel (Sambrook et al., 1989). The gel was run at 80-100 Volts for 1-2 hours.

The different chloroplast fractions after chloroplast import (2.2.10.3) were analysed on a 20 × 20 $cm^2$ resolving SDS-polyacrylamide gradient gel (Sambrook et al., 1989) consisting of 2 components: A resolving gel and a 5% stacking gel on top. The components of the gel were added in the order given in the following tables A and B:

**(A) Resolving gel:**

| **Acrylamide concentration** | **10%** | **17.5%** |
|---|---|---|
| $H_2O$ | 9.8 ml | – |
| 2M Tris-HCl (pH 8.8), | 4.5 ml | 4.5 ml |
| 80% Sucrose | 1.4 ml | 5.2 ml |
| 30% Acrylamide/Bisacrylamide (29:1) | 8.0 ml | 14.0 ml |
| 10% SDS | 240 µl | 240 µl |
| 10% Ammonium persulphate (APS) | 100 µl | 100 µl |
| TEMED | 10 µl | 10 µl |
| **Total** | **24 ml** | **24 ml** |

**(B) 5% Stacking Gel:**

| | |
|---|---|
| $H_2O$ | 6.95 ml |
| 1 M Tris-HCl pH 6.8 | 1.25 ml |
| 30% Acrylamide/Bisacrylamide (29:1) | 1.67 ml |
| 10% SDS | 100 µl |
| 10% Ammonium persulphate (APS) | 80 µl |
| TEMED | 8 µl |
| **Total** | **10 ml** |

For the gel frame, two glass plates were placed with one over the other with two spacers and pushed to the edges. They were held together with clamps. The two sides and the bottom were properly sealed with a tape resistant to the reagents like

SDS. The bottom edge was secured with an additional piece of tape to avoid any leakage. With the help of two clamps on each side the set-up was held in a vertical position.

The two different gradient concentrations of the resolving gel were prepared fresh. APS was added just before pouring to start the polymerisation. The two solutions were slowly mixed into each other in a gradient mixer and simultaneously poured between the glass plates of the gel frame. 17.5% acrylamide was present in the front chamber and flowed faster than the 10% acrylamide present in the back chamber, which mixed slowly into the front chamber. At the end, a uniform gradient was built in the gel. Then, approximately 1 ml of $H_2O$ was slowly pipetted at one edge of the gel along the glass plate. Being lighter, it distributed itself as a thin layer on top of the gel and removed all the air bubbles from top. The gel was allowed to polymerise. After approximately 1 hour, when the gel had polymerised, the water was poured out and the space between the plates dried with the help of the 3mm Whatman paper. The stacking gel was then poured over the resolving gel. After pouring, a comb was inserted on the top. It was then allowed to polymerise for an hour, and after removal of the comb, the wells were cleaned with water to remove pieces of gel.

The protein samples were allowed to separate on the gel by running it at 80-100 volts for 1-2 hours, until the bromophenol blue front was at the bottom of the gel.

#### 2.2.9.5. Gel staining and analysis

The gel plates were separated from the gel and the gel was placed first in Coomassie stain (0.15% w/v Coomassie Brilliant blue in 45% v/v methanol, 10% v/v glacial acetic acid) for 20 min. with gentle rocking, followed by the destaining solution (45% v/v methanol, 10% v/v glacial acetic acid) for approximately 1 hour. This gel was placed on a wet Whatman paper, lying over a dry Whatman paper, then covered on the top with Saran Wrap and then dried in a gel drier for 1 hour. This gel, dried on the Whatman paper, was then exposed to a phosphoimager plate in the dark for 2-3 days.

### 2.2.10. Chloroplast Import

The chloroplast import experiment was performed throughout at 4°C or on ice, unless specifically mentioned. Throughout the protocol, the chloroplasts were pipetted with 1ml pipette tips, whose tips had been cut off to broaden the opening and thus, cause minimal damage to the chloroplasts.

#### 2.2.10.1. Isolation of spinach chloroplasts

For one preparation, approximately 60 g fresh, deep green spinach leaves without the middle vein were cleaned an evening before and stored at approximately 6°C. All the materials, solutions and instruments to be used were pre-cooled to 4°C overnight. For isolating undamaged chloroplasts, a PBF-percoll stock solution (3% w/v PEG 6000, 1% w/v BSA, 1% w/v Ficoll 400 in Percoll) was prepared. The two layers of the percoll gradient, i.e., 85% PBF-percoll layer (85% v/v PBF-Percoll-stock solution, 2 mM EDTA, 1 mM $MgCl_2$, 50 mM HEPES-KOH pH 7,6) and the 45% layer (45% v/v PBF-Percoll-stock solution, 2 mM EDTA, 1 mM $MgCl_2$, 50 mM HEPES-KOH pH 7,6) were then prepared. The percoll gradient was prepared by stacking 5 ml of 45% PBF-percoll on top of 3 ml of 85% PBF-percoll by gently pouring the two into a slanting tube.

In a blender, 180 ml of homogenisation medium (2 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 50 mM HEPES-KOH pH 7,6) was added to the spinach leaves. To blend the spinach, 2-3 seconds long pulses were given 4 to 6 times and a final pulse of approximately 10 seconds. The material was mixed in between with a spatula. The homogenate was filtered through 2 layers of a nylon filter (100 µm pore size) or Miracloth and then centrifuged for 2 min. at 5,000 rpm in a rotor pre-cooled to 4°C. The supernatant was decanted and the pellet was gently resuspended in 3 ml homogenisation medium using a soft hair brush and pipetting 1-2 times with a 1 ml pipette tip. These chloroplasts were then placed over a 45%-85% Percoll gradient in a tube and centrifuged in a swing-out rotor at 10,500 rpm for 10 min., at 4°C. Two distinct chloroplast bands were observed in the gradient. The one at the top, which represented damaged chloroplasts, was discarded and the lower band was carefully taken up in a 1 ml pipette tip and transferred to two 15 ml corex glass tubes. These intact chloroplast bands were washed with 10 ml homogenisation medium each, by letting it run slowly along the tube wall. The chloroplasts were then centrifuged for 1

minute at 5,000 rpm, at 4°C (fast-stop on) in a fixed angle rotor. The chloroplast pellets were resuspended in 1-2 ml homogenisation medium and stored in a 2 ml microfuge tubes. The chlorophyll concentration was measured at wavelengths 663 nm and 645 nm in a spectrophotometer by diluting 10 µl sample with 990 µl 80% acetone, thoroughly vortexing followed by centrifugation for 5 min. at 13,000 rpm, at RT. The concentration of chloroplasts was calculated with the chlorophyll content, using the following formula:

$$C_{chlorophyll} = [(20.2 \times E_{645} + 8.02 \times E_{663}) \times \text{Dilution factor}]\ \mu g/\mu l$$

### 2.2.10.2. Chloroplast import experiment (Clausmeyer et al., 1993)

All the reagents and buffers for the chloroplast import reaction were kept on ice. The reaction mixture was also prepared on ice. To reduce the radioactive cysteine present in the translation reaction to a negligible amount, non-radioactive cysteine was added in much higher amounts in the chloroplast reaction, so that no unspecific labelling of the proteins synthesised in the chloroplasts during the import occurs and no unspecific background radioactivity is detected. For the import, fresh *in vitro*-translation reaction was prepared as shown in the following table.

| **Experiment :** | **A** | **B** | | | |
|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** |
| Chloroplast suspension (≡ 50µg), µl | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 |
| SRM, µl | 64.0 | 64.0 | 64.0 | 64.0 | 64.0 |
| 100 mM ATP, µl | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| 250 mM Cysteine, µl | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| *In vitro* translation reaction, µl | 12.5 | 12.5 | 12.5 | 12.5 | 12.5 |
| Competitor, µl (concentration) | **–** | – **(0 µM)** | 0.75 **(1 µM)** | 1.5 **(2 µM)** | 3.0 **(4 µM)** |
| 7 M Urea, µl | – | 3.0 | 2.25 | 1.5 | – |
| $H_2O$, µl | 19.0 | 16.0 | 16.0 | 16.0 | 16.0 |
| **Total** | **150 µl** | **150 µl** | **150 µl** | **150 µl** | **150 µl** |

The import reaction was carried out at 25°C for 20 min. in a well-lighted water bath. In another experiment done in parallel, the thylakoid lumen located 33kD protein

subunit (pOE33) of the oxygen evolving complex was used as an unlabelled competitor of the protein import into chloroplasts (Michl et al., 1994). This unlabelled protein solution was prepared in 7M urea. Four different concentrations of the competitor were used in a series of four reactions. In order to maintain the same concentration of urea in all samples, the appropriate amount of a 7M urea solution was added. The first reaction in the series was a control without the competitor, but with urea.

After 20 min., the reaction was stopped by adding 1 ml ice-cold SRM (50 mM HEPES-KOH, pH 8.0, 0.33 mM Sorbitol). The sample was then centrifuged at 6,000 rpm for 3 min. at 4°C. The chloroplast pellet was then resuspended in 250 µl SRM.

#### 2.2.10.3. Chloroplast lysis and isolation of sub-fractions

The radiolabelled protein fraction left in the suspension outside the chloroplasts was first degraded with 25 µl Thermolysin (2 mg/ml, 25mM $CaCl_2$) on ice for 30 to 40 min. The reaction was stopped by adding 50 µl SRME 50 (SRM in 50 mM EDTA). This chloroplast suspension was then placed over 500 µl 35% percoll cushion (35% v/v Percoll in 1× SRM) and centrifuged at 8,000 rpm for 8 min., at 4°C. Damaged chloroplasts present in the supernatant were removed with the supernatant, without disturbing the intact chloroplast pellet. The pellet was then washed twice with 900 µl SRME10 (SRM buffer in 10 mM EDTA), by gently resuspending the pellet with 1 ml pipette and then centrifuging at 6,000 rpm for 3 min.

To lyse the chloroplasts in the pellet, the pellet was resuspended in 75 µl HME 10 buffer (10 mM HEPES-KOH, pH 8.0, 5 mM $MgCl_2$, 10mM EDTA) and incubated on ice for 10 min. Then, the thylakoids were separated from the stroma by centrifuging at 13,000 rpm for 5 min and obtained as a pellet. The supernatant was transferred to a new 1.5 ml microfuge tube, to which 25 µl 4× Laemmli buffer was added, and placed on ice. This protein fraction was called the **Stromal fraction**.

To remove traces of the Stromal fraction from the thylakoids in the pellet, the pellet was washed twice with HM buffer (10 mM HEPES-KOH, pH 8.0, 5 mM $MgCl_2$) by resuspending it in 200 µl HM and centrifuging at 13,000 rpm for 5 min. The pellet was then resuspended in 200 µl HM and divided equally into 2 microfuge tubes. One sample was placed on ice and was called the **Minus fraction**. To the other sample,

10 µl Thermolysin was added to degrade proteins present on the outer side of the thylakoid membrane and it was incubated for 30 min. on ice. This fraction was called the **Plus fraction**. Both the fractions were centrifuged at 13,000 rpm for 5 min. and the pellets were resuspended in 100 µl 2× Laemmli buffer with EDTA (50% v/v 4x Laemmli buffer, 20mM EDTA) prepared from 4x Laemmli buffer (0.25 M Tris-HCl, pH 6.8, 8% w/v SDS, 40% w/v, glycerine, 20% w/v β-mercapto-ethanol, 0.016% w/v bromophenolblue).

Finally, all the 3 fractions were denatured in a boiling water bath (100°C) for 2 min., then cooled down on ice and centrifuged for 5 min. at 13,000 rpm.

Samples (25 µl) of the three different fractions from each chloroplast import reaction were loaded on an SDS-polyacrylamide gradient gel (see 2.2.9.4) along with a protein ladder, SDS-7L (Sigma).

# 3. Results

In this study, the *etched1* (*et1*) gene was characterised phenotypically as well as at the molecular level. In order to understand the physiological role of the *et1* gene during kernel development, the *et1* phenotype in the kernels was examined. The molecular characterisation of the *et1* gene was initiated with the isolation of cDNA and genomic clones and their analysis. In order to characterise the expression of *et1* in the whole plant, Northern experiments were performed. Further experiments, including chloroplast import experiments and *in silico* analyses that indicated the presence of functional domains in the protein, were performed to characterise the ET1 protein structure and function. Analysis of the protein's secondary structure, in view of the identified functional domain, was also carried out.

Southern experiments performed with *et1* probes on the maize wild type genome (Line C) and work done by M. Ahrend (1998) indicated the presence of a second copy of *et1* in maize. Additional experiments were performed, which involved examination of *et1-Ref* genomic clones. Sequence analysis of these clones revealed the presence of a new *et1* homologue, *zmzr1*. This gene was characterised using Southern analyses of genomic DNA isolated from Line C (LC) and *et1-Ref*, at the post-transcriptional level through Northern analysis, and through isolation and characterisation of cDNA clones.

In view of their homologies, the two genes were compared at the molecular level. The structure of the *et1* and *zmzr1* gene, including the upstream regions, as well as their cDNAs and the putative protein sequences were examined. Online EST databanks were searched to identify homologues to the two proteins in other plants.

## 3.1. The examination of the *etched1* phenotype in kernels

In order to better understand and visualise the *etched1* phenotype in kernels, hand sections of 20 DAP old *et1-Ref* coloured kernels were examined. Presence of depressions and crevices on the kernel surface were observed. Hand sections stained with IKI solution or unstained were examined (Fig. 3.1). In the pericarp and

the embryo development no abnormalities could be observed. IKI-stained *et1-Ref* kernel sections showed the presence of sectors of normal violet blue stained starchy cells in the endosperm alternating with cells, which did not stain violet-blue, indicating the absence of starch grains in these cells. In transverse sections, the sectors of normal starchy endosperm cells were found to be traversed by such starchless sectors radiating out from the central region. These sectors very often seemed to be present at regular intervals arranged in a radial pattern from the centre of the kernel (Fig. 3.1 E, F, and G). The central region of the endosperm was almost invariably found to be filled mainly with starchless cells.

In some kernels, the intensity of the surface phenotype appeared milder, as fewer or almost no depressions were observed on the crown. The depressions were mainly restricted to the lower side of the crown, on the abgerminal side as well as on the germinal side of the crown, below the silk scar. Hand sections of the kernels with milder phenotype showed fewer starchless cells below the crown, but the intensity at the centre and in the affected regions appeared in many cases to be the same as in other kernels. In contrast, in wild type *Et1-A188* and *Et1-Q2* kernels, a uniform violet-blue staining with no colourless bands was observed throughout the starchy endosperm.

Microscopic observations of these 20 DAP *et1-Ref* kernels showed the presence of a normal pericarp. The endosperm region was found to be fully cellularised. However, the normal starchy endosperm cells were found to be traversed by rows of cells lacking any starch grains. A continuous aleurone layer, which appeared to grow normally, including the regions where starchless storage cells were present below, was observed at the periphery of the endosperm. Below the sub-aleurone layer, the peripheral cells were elongated in the anticlinal plane, whereas in the central region, the cells were more rounded in shape. The normal looking starchy endosperm cells were filled with prominent starch grains. Although the starchless cells at the periphery had the form of a starchy endosperm cell, appropriate for that region, these cells were much thinner and smaller. At the centre, where a larger number of starchless cells were observed, the cells were partly interspersed with normal starchy cells, filled with starch grains. At the periphery of the kernel, in the regions where the starchless cells were present, the endosperm, including the outermost aleurone layer, was found to be pulled away from the maternal pericarp, giving rise to depressions on the

kernel surface. These depressions were responsible for the “etched” appearance of the kernels. Some close-up views of such kernel sections that were photographed or scanned are shown in Figure 3.1 (next page). Microscopic examination of 22 DAP wild type *Et1-A188* kernels revealed a uniform distribution of starch grains in the starchy endosperm. The sub-aleurone layer was found to contain fine starch grains, smaller than in the rest of the endosperm. The endosperm in the vicinity of the transfer layer, however, appeared to lack starch grains as no violet blue staining in this region was observed.

Based on this analysis, it could be concluded that the *et1* phenotype showing depressions and cracks on the endosperm surface on maturity was probably caused by a disturbed or lack of normal amyloplast development, also associated with starch synthesis, which, however, was restricted mainly to the affected starchy endosperm radial sectors. This also correlated with the microscopic examination carried out by M. Ahrend (Diploma thesis, 1998; see introduction, Fig. 1.3) of the *et1-Ref* leaves, where chloroplast differentiation is disturbed in the pale seedling leaves. This indicated an association of the pleiotropic phenotype with disruption of plastid development, as the normal role of the gene is probably affecting the plastid differentiation and development in different plant tissues.

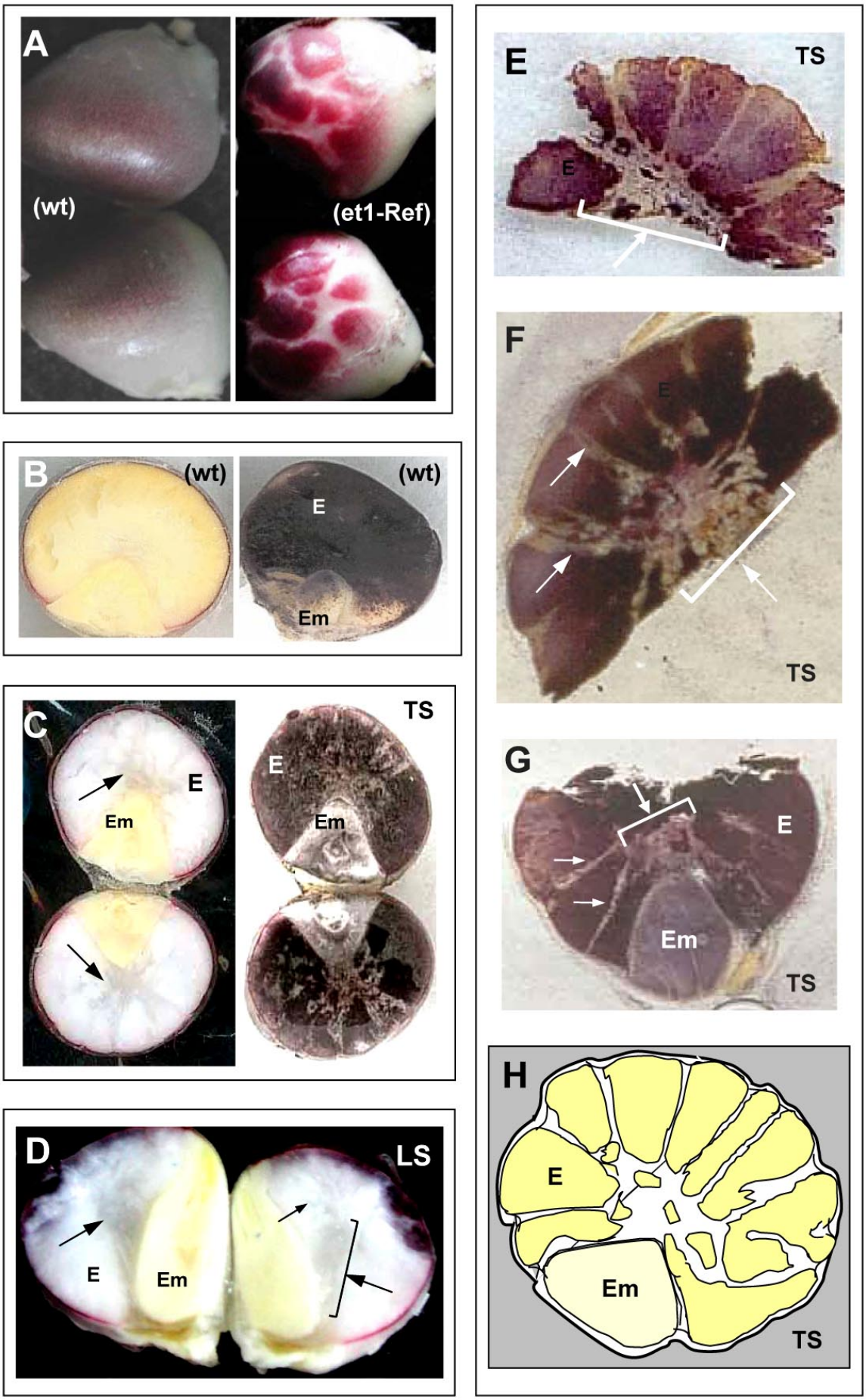



**Fig. 3.1:** continued on the next page

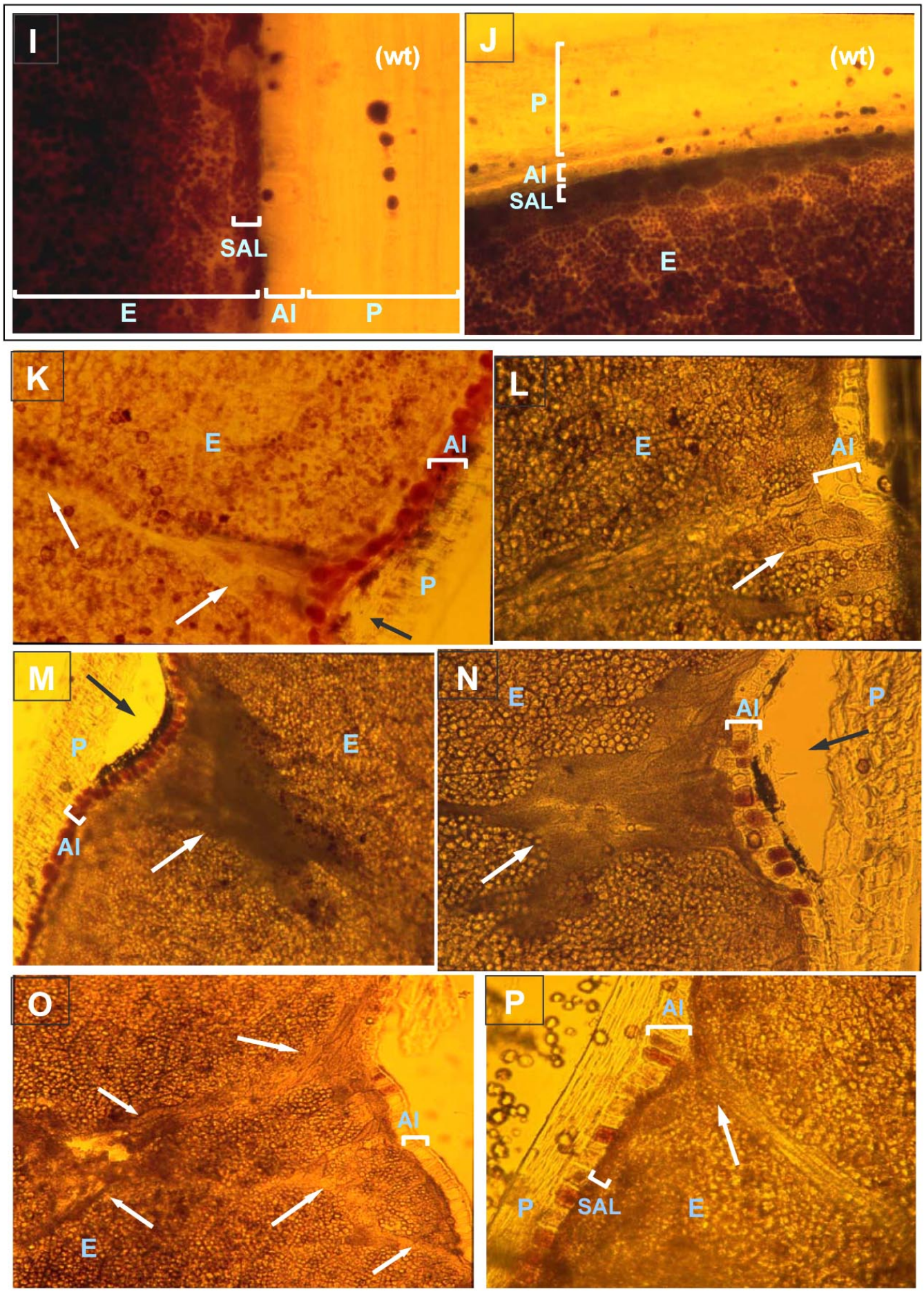



**Fig. 3.1:** continued on the next page

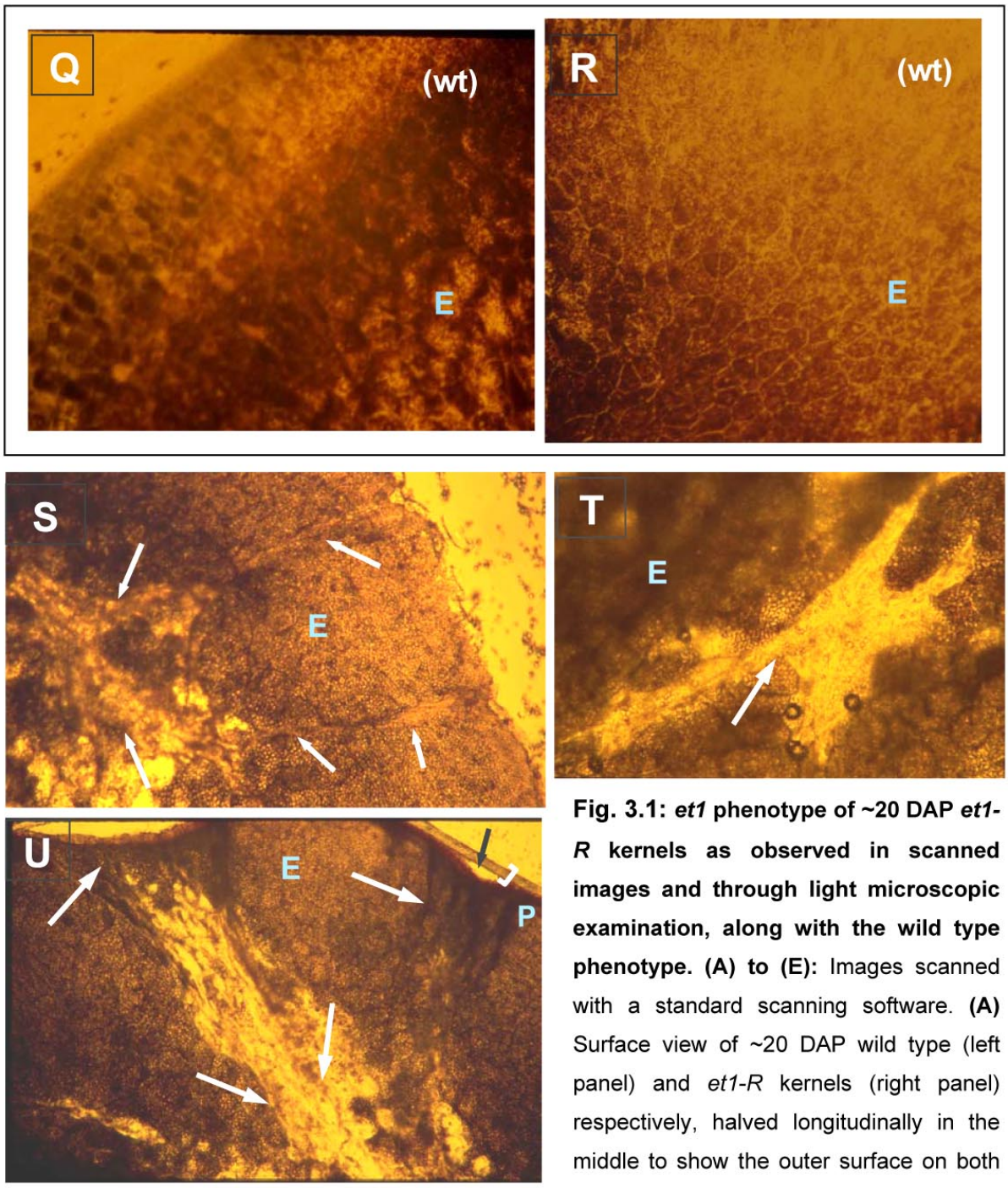



**Fig. 3.1:** ***et1*** **phenotype of ~20 DAP** ***et1-R*** **kernels as observed in scanned images and through light microscopic examination, along with the wild type phenotype. (A) to (E):** Images scanned with a standard scanning software. **(A)** Surface view of ~20 DAP wild type (left panel) and *et1-R* kernels (right panel) respectively, halved longitudinally in the middle to show the outer surface on both the sides.

**(B)** Transverse sections (TS) of two different 20-23 DAP wild type kernels. Left panel: A kernel showing the embryo and the starchy endosperm, whose outer layers have begun to desiccate. Right panel: The section has been stained with IKI solution showing uniform violet-blue staining in the endosperm (E), except at some places, due to uneven kernel surface. The embryo (Em) also showed weak blue staining. **(C)** ~20 DAP *et1-R* kernel transversely sectioned (TS) into two halves. Left panel: showing watery or translucent starchless cells (black arrows) radiating out from the centre. Radial sectors of starchy endosperm cells (white) in the starchy endosperm are also observed. Whereas the embryo appears yellowish, the endosperm lacks any coloration. Right panel: the same stained with IKI. The starchless cells are observed more clearly as white coloured radial streaks or bands. The starchy endosperm cells are stained violet-blue.

**Fig. 3.1:** continued on the next page

**(D)** Longitudinal section (LS) of a ~20 DAP *et1-R* kernel, showing both the halves from inside. The central endosperm region, without starch, appears translucent and greyish due to lack of starch grains, and is marked with black arrows. The starchy endosperm cells are white in colour. **(E), (F) and (G)** Thin transverse sections (TS) of ~20 DAP *et1-R* kernels stained with IKI. The starchless cells of the starchy endosperm are white (marked with white arrows), which radiate out from the central region up to the periphery and do not stain violet, like the starch containing endosperm cells. Starchy endosperm cells appear as radial sectors, with starchless cells in-between. **(H)** Schematic representation of a transverse section of an *et1-R* kernel, based on **(E), (F)** and **(G)**.

**(I) to (U):** Light microscopic images of transverse sections of wild type (wt) and *et1-R* kernels. White arrows indicate regions of starchless cells, and black arrows indicate pits on the endosperm surface giving rise to the *etched1* phenotype. **(I)** (40X) **& (J)** (20X), Transverse sections (TS) of IKI stained ~20 DAP wild type (A188) kernels showing the violet-blue stained starchy endosperm (E) and the sub-aleurone layer (SAL) of the starchy endosperm, the colourless aleurone layer (Al) and multi-layered pericarp (P). **(K) to (P):** Transverse sections of ~20 DAP *et1-R* kernels showing the peripheral starchless cells present below the aleurone layer. **(K), (L), (N) & (P)** 20X; **(M) & (O)** 10X. The section in image (**K)** has been lightly stained with diluted IKI. All the sections reveal a continuous and intact aleurone layer (Al), present above the starchy endosperm (E). In a few cases, the aleurone is found to be pulled away from the pericarp (P), as the starchless endosperm cells, present below the aleurone, do not keep pace with the growing normal starchy endosperm cells present around them, which gives rise to pits created between the pericarp and aleurone (shown with black arrows). As can be seen in all the cases, the starchless cells are not just restricted to the periphery, but appear to be showing a clonal distribution where the starchless cells appear to be penetrating deeper into the endosperm, as observed in the images **(K), (L), (O)** and **(P)**. Although, the distribution of the starchless cells appears to be more complex under the light microscope (**(L)** and **(O)**) and not as regular as apparent in the scanned images **(E)** to **(H)**. The starchless cells in **(M)** and **(N)** might be moving in another plane towards the centre, so that these are only a few layers deep in the cross-sections. Although, they reveal the broadness of the starchless sectors, also observed in the scanned images.

**(Q)** 5X **& (R)** 10X, Transverse sections of ~20 DAP wild type (A188) kernels, stained with IKI and showing peripheral and central regions of the starchy endosperm respectively. The cells of the starchy endosperm (E) are evenly stained with IKI, indicating uniform distribution of starch grains (amyloplasts) throughout. Moreover, a smooth and uniform surface at the periphery of the endosperm is observed in **(Q)**. **(S) to (U):** Transverse sections of ~20 DAP *et1-R* kernels showing *et1* phenotype in the peripheral and central regions. **(S) & (U)** 5X, a wide view of the *et1* phenotype under the light microscope showing radial sectors of normal starchy endosperm cells separated by starchless cells radiating out from the central region up to the periphery. **(T)** 10X, starchy endosperm showing central region of starchless cells surrounded by starchy sectors, which are again separated by starchless cells radiating out from the centre towards the periphery.

## 3.2. Molecular characterisation of the *etched1* gene

A number of *et1* mutant alleles, obtained from different *Mutator* lines, had been analysed genetically so far. One wild type allele originating from *et1-m3*/+(B73) had also been isolated and analysed (da Costa e Silva et al., 2001).

To characterise the *et1* gene, it was necessary to examine the gene in an inbred maize line wild type for *et1*. Line C (LC), a colour converted W22 line, being one such line, was chosen for studying *et1* genomic structure and all further analyses. Using a 2.5 kb *et1* fragment from B73 as a probe, a full-length *et1* cDNA, *c9.1*, was isolated from a LC developing kernel cDNA library (da Costa e Silva et al., 2001).

In this study, the cDNA clone, *c9.1*, was used to isolate further *et1* cDNA clones from a LC developing kernel cDNA library. The 2.5 kb *et1* fragment probe was used to isolate the *et1* gene from a Line C genomic library. In addition, the expression pattern of the *et1* gene in different maize tissues was analysed in order to characterise the gene function. Based on an *in silico* prediction, a chloroplast import experiment was carried out to show plastid localisation of the ET1 protein. The mature ET1 protein sequence, remaining after chloroplast import, was analysed *in silico* to identify a functional domain.

### 3.2.1. Isolation of new cDNA clones from the LC kernel cDNA library screening

The LC *et1* cDNA clone, *c9.1*, isolated by O. da Costa e Silva (da Costa e Silva et al., 2001), was 662 bp in length and contained a 12 bp long 5´untranslated region upstream of the putative translation start site. This 5´untranslated region and the cDNA itself appeared to be comparatively small. Therefore, an amplified dT primer cDNA library, prepared from LC developing kernels (13, 17, 19, 20, 22, 25 and 28 DAP), was screened with the aim of identifying possibly larger cDNA clones. For this, $1.2 \times 10^6$ phage clones were screened using *c9.1* as a probe.

Nine different clones were isolated, but all were smaller than *c9.1*. Sequence analysis of these clones showed that all of them were identical to *c9.1*.

```
                *         20         *         40         *         60
c9.1 : GGCAAAGCGGCTATGACGACGACGGCCGCCGGGCACGGCTGCTGCTGGGCGGGGATTCCG
PG2  : ..................GACGACGGCCGCCGGGCACGGCTGCTGCTGGGCGGGGATTCCG

                *         80         *        100         *        120
c9.1 : CCCTCCGCGTTGTTGCCGCGGATTCTCTCGACCGGCCGGGAGACTCCTCCTCCTCGCGCT
PG2  : CCCTTCGCGTTGTTGCCGCGGATTCTCTCGACCGGCCGGGAGACTCCTCCTCCTCGCGCT

                *        140         *        160         *        180
c9.1 : TCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCGCGGCTGAGAGTTTCGAAC
PG2  : TCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCGCGGCTGAGAGTTTCGAAC

                *        200         *        220         *        240
c9.1 : CGTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGAGGCCAATTCGGACGCGGTG
PG2  : CGTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGAGGCCAATTCGGACGCGGTG

                *        260         *        280         *        300
c9.1 : CCGTCGCCAACGGAAGCCGCTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAA
PG2  : CCGTCGCCAACGGAAGCCGCTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAA

                *        320         *        340         *        360
c9.1 : TTTACATGCAACGCATGTGGCGAAAGGACCAAGCGCTTGATAAACAGAGTAGCCTATGAA
PG2  : TTTACATGCAACGCATGTGGCGAAAGGACCAAGCGCTTGATAAACAGAGTAGCCTATGAA

                *        380         *        400         *        420
c9.1 : AGAGGCACAGTTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAAT
PG2  : AGAGGCACAGTTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAAT

                *        440         *        460         *        480
c9.1 : CTTGGGCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCG
PG2  : CTTGGGCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCG

                *        500         *        520         *        540
c9.1 : GTGGACACCAGTTCTGAAGATTGATCTGCTGTGAGAAGCGATGTTGGTATGCAAAACGCC
PG2  : GTGGACACCAGTTCTGAAGATTGATCTGCTGTGAGAAGCGATGTTGGTATGCAAAACGCC

                *        560         *        580         *        600
c9.1 : CTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCATAAAAATGGCACTTTT
PG2  : CTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCATAAAAATGGCACTTTT

                *        620         *        640         *        660
c9.1 : TAAAGTTGTTGCACATACTCATCTGAAATTCTGAATACAGCAGGGCCTACATTGTACTTC
PG2  : T.....................................................


c9.1 : TT
PG2  : ..
```

**Fig. 3.2:** (Legend on the next page)

**Fig. 3.2: Comparison of the sequences of the two LC *et1* cDNA clones, pg2 and *c9.*1, isolated from a LC developing kernel cDNA library.** The sequence of the newly isolated cDNA clone, PG2, was one of the nine different cDNA clones isolated from the cDNA library screening, all being identical in size and sequence. This class of cDNA clones differed from *c9.1*, isolated by O. da Costa e Silva (da Costa e Silva et al., 2001), in having a different polyadenylation site, which resulted in a smaller 3´untranslated region in these clones. In addition, these clones were incomplete at the 5´end and lacked both the 5´ UTR as well as the translation start codon, present in *c9.1*.

However, these clones were only partial derivatives of *c9.1*. They lacked the 5´ end of the cDNA including the 5´UTR and the translation start codon, which are present in *c9.1*. These clones also had a different polyadenylation site. This polyadenylation site was closer to the translation stop codon, thus resulting in a transcript with a smaller 3´ untranslated region (Fig. 3.2). This category of *et1* cDNA clones had also been obtained in earlier library screening experiments performed by O. da Costa e Silva (da Costa e Silva et al., 2001). Thus, although no new clones could be isolated, the doubt of having missed a possibly larger clone from the library could be removed. Further doubts about the size and completeness of the *et1* transcript were cleared partly characterised with the help of Northern analyses shown later in this chapter.

### 3.2.2. Southern Analysis of Line C genomic DNA

In order to analyse the genomic structure of the *et1* gene, the LC genomic DNA was subjected to Southern analysis (Fig. 3.3a and b). Six digestions of LC genomic DNA samples were performed with the restriction enzymes *Bam H*I*, Hind*III*, Hind*III*/Eco R*I*, Eco R*I*, Sal*I and *Xba*I respectively for the analysis. The Southern blot was first hybridised with *c9.1* followed by the 2.5 kb *et1* genomic fragment probes.

The Southern analysis with the *et1* cDNA revealed the presence of two sets of bands. The first set of bands gave a strong hybridisation signal and a second set, which showed a weaker signal. Considering only the stronger bands first, two hybridising bands were observed in a *Hind*III digestion with the *c9.1* probe. This would be expected based on the *et1* cDNA sequence data. The *et1* cDNAs contained

two *Hind*III restriction sites with a gap of 10 bp (at positions 271 bp and 282 bp from the start of *c9.1* sequence), which on digestion would divide the cDNA into two main parts, and a small 11 bp fragment. Since the 2.5 kb (*Xho*I/*Hind*III) *et1* genomic fragment only contained a part of the cDNA sequence, 5´ to the *Hind III* restriction site, a single hybridising band was seen in the *Hind III* digest with this probe (Fig.3.3b). A few weaker bands, corresponding in size to the stronger ones observed with the *et1* cDNA probe, were detected. These additional weaker bands probably indicate the presence of sequences with similarity to *et1* in the genome.

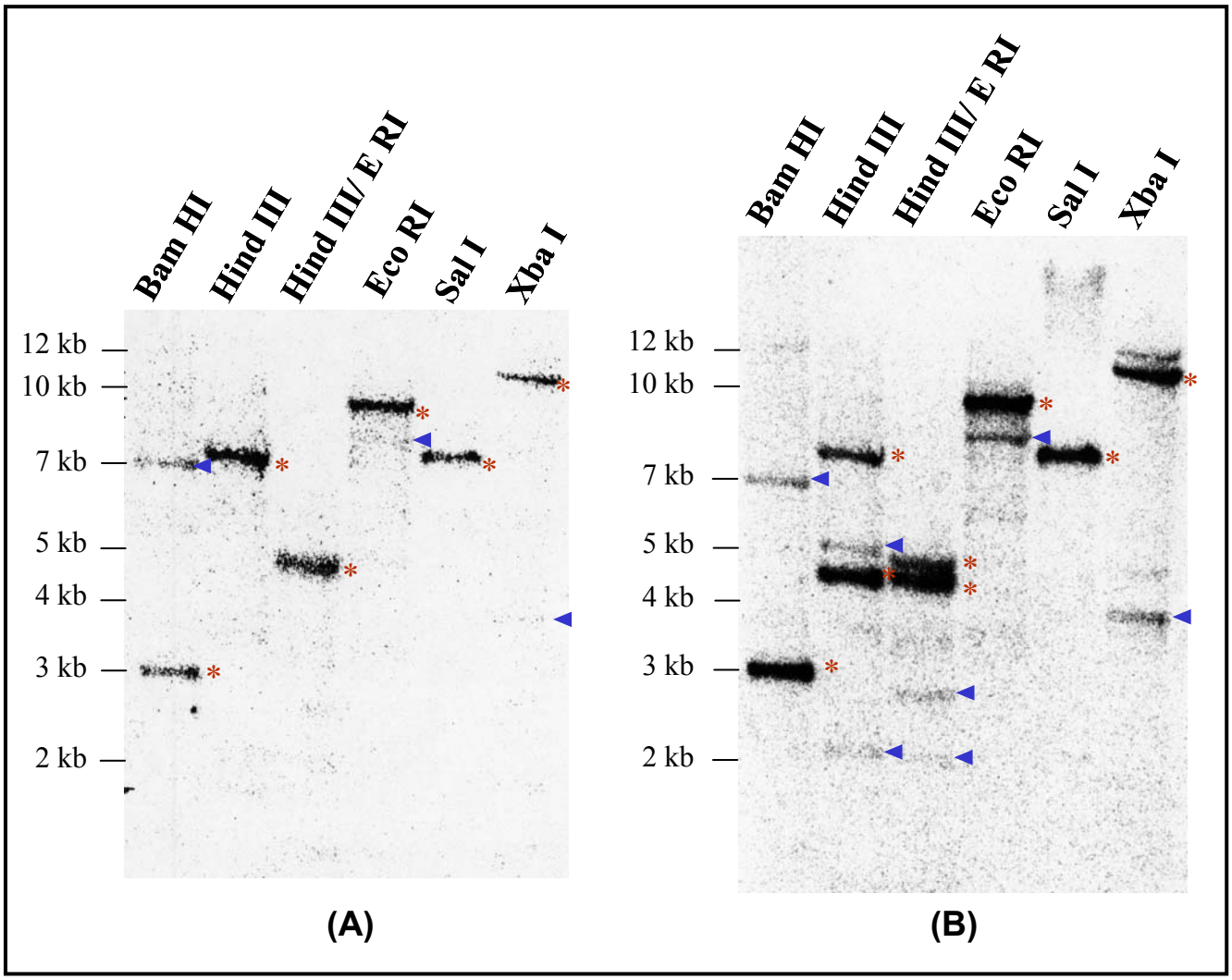



**Fig. 3.3: Southern analysis performed with LC genomic DNA**. 10 µg genomic DNA was taken for each restriction digest. (**A**) Southern hybridisation with the 2.5 kb wt *et1* fragment probe. (**B**) Southern hybridisation with the *et1* cDNA, *c9.1*, probe. The bands giving a stronger signal, indicated with asterisk, represent the bands corresponding to the *et1* genomic clone *Et1* λ *LC 11.1/ 7.2*. The set of bands indicated with arrows correspond to the *zmzr1* genomic clones (see section 3.3). Additional unidentified bands were also observed with the *et1* cDNA probe.

### 3.2.3. Molecular analysis of the *et1* genomic structure in Line C

In order to further characterise the *et1* genomic structure, an amplified LC genomic library was screened for genomic clones. In total 1.4 × $10^6$ λ phage clones of a Lambda Fix II library were analysed with the 2.5 kb *et1* fragment (B73) as a probe. From this screening twelve positive clones were isolated. After carrying out a restriction digest with eight of these clones, a Southern analysis was performed, in which, first the *et1* cDNA (*c9.1*) probe and then the 2.5 kb fragment probe were hybridised to the blot. Two identical clones LC λ *pgg 11.1* and LC λ *pgg 7.2* were selected for further analysis (referred as *LCg 11.1(7.2)* here onwards). Based on the hybridisation results, a restriction map was prepared for this clone. It contained the complete *et1* locus (Fig. 3.4).

In order to sequence the *et1* gene, some subcloning experiments had to be carried out with *LCg 11.1(7.2)*. A 7 kb *Sal*I sub-fragment (Fragment A, Fig. 3.4) was cloned into the pZErO-II plasmid vector. It contained the *et1* gene and the regions flanking it, to which the *et1* probes had hybridised. Subfragments of this clone were further subcloned (as shown in Table 3.1) and then sequenced.

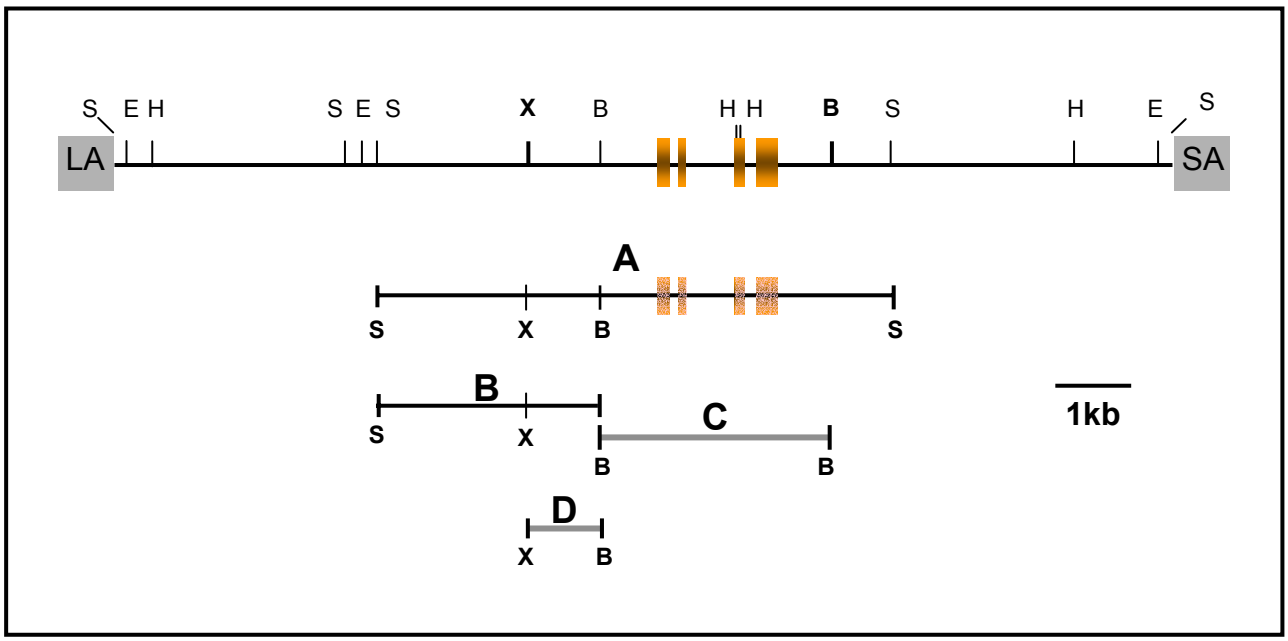



**Fig. 3.4: The restriction pattern of the *et1* genomic clone *Et1 λ LCg 11.1 (7.2)* obtained from Line C.** The subclones prepared from this genomic clone are indicated below the restriction map, shown in the 5´–> 3´ orientation of the gene. The coloured blocks represent the four exons in the gene. After cloning fragment A, the fragments B and C were further subcloned from it. The subcloned

fragments C and D were sequenced completely to reveal the *Et1* genomic structure. The scale bar on the right represents a length of 1kilobases. Abbreviations: LA : long arm, SA: short arm of the λ phage; The restriction enzymes indicated are: B: *Bam H*I; E*: Eco R*I; H: *Hind*III; S: *Sal*I ; X: *Xho*I.

The results of the Southern hybridisation experiments with LC genomic DNA were compared to the restriction map of the λ genomic clones (Fig. 3.4). One set of genomic DNA bands that showed a stronger hybridisation signal (asterisk, Fig. 3.3A and B) corresponded in size to the restriction pattern of the *et1* genomic clones. However, a number of additional, comparatively fainter bands were also observed in the Southern analysis, especially with the *c9.1* probe. As will be shown in section 3.4.3, one set of these additional bands (arrows, Fig. 3.3 A and B) could later be identified as belonging to an *et1* homologue, *zmzr1*.

| S.No. | Fragment | Origin | Length (Restriction Cuts) | Vector (Site of cloning) |
|---|---|---|---|---|
| 1. | Fragment A | *Et1 λ LCg 11.1/7.2* | ~ 7 kb (*Sal*I) | pZErO-II (*Xho*I) |
| 2. | Fragment B | Fragment A | ~ 3.2 kb (*Sal I/Bam H*I) | pZErO-II (*Xho*I/*BamH*I) |
| 3. | Fragment C | Fragment A | ~ 3 kb (*Bam H*I) | pZErO-II (*BamH*I) |
| 4. | Fragment D | Fragment B | ~ 1 kb (*Xho*I/*BamH*I) | pZErO-II (*Xho*I/*BamH*I) |

**Table 3.1: Fragments subcloned from the *Et1 λ LCg 11.1(7.2)* genomic clone for sequencing.**

### 3.2.4. Sequence Analysis of the wild type *et1* gene and the corresponding cDNA

From the *et1* λ genomic clone *LCg 11.1/7.2* the two subclones: 1kb (*Xho*I/*Bam H*I) and 3kb (*Bam H*I), which contained the complete *et1* locus (Fig. 3.4), were sequenced and analysed. The sequence data of these *et1* genomic clones was compared to those available from the B73 line (da Costa e Silva et al., 2001) and the LC cDNA, *c9.1*. The LC genomic sequence was compared to the cDNA sequence

and four exons and three introns were detected in the *et1* gene based on the 5´ and 3´ exon/intron splice site consensus for monocots (Simpson et al., 1996). The wild type genomic sequence from B73 background was found to have a high sequence similarity with LC *et1* (Fig. 3.5) and the same exon-intron structure. The two *et1* genomic clones showed 97% homology between the two alleles with minor differences present in the form of single nucleotide substitutions throughout the gene sequence.

In the coding region and the intervening sequences, only single nucleotide substitutions were observed. These substitutions were either transitions or transversions. In the exons, however, all of the substitutions except for one were transitions. These single nucleotide substitutions at three positions in the exons I and II changed the deduced protein sequences of the two alleles. However, as will be shown later, these differences did not cause any major change in the protein physiochemical properties. The remaining single nucleotide substitutions, when present in the exons, were synonymous substitutions, present at the wobble position, causing no consequential change in the deduced protein sequence of either of the genes.

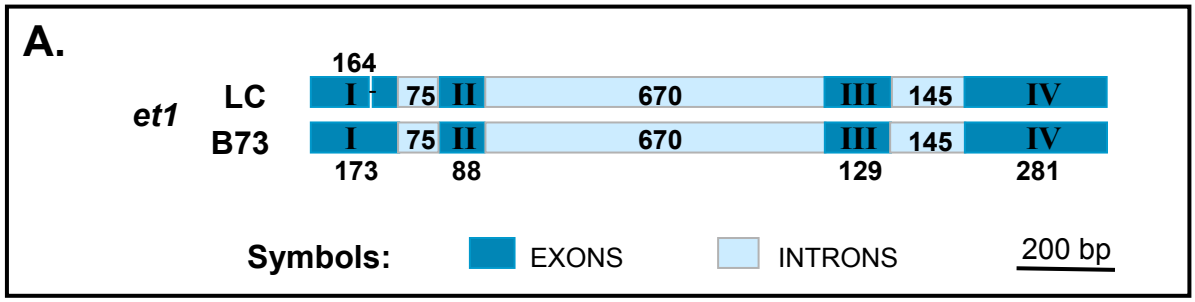
However, at one position in exon 1 (103 bp downstream of transcription start), the wild type *et1* sequence from B73 differed from that of LC in having three additional triplet bases. This difference gave the deduced protein sequence of the B73 allele three additional proline amino acids than the LC allele (Fig.3.5).

Considering all the single nucleotide substitutions as single nucleotide polymorphism (SNP) between the two *et1* alleles, for the complete *et1* gene (with a length of 1552bp, ignoring the 9 additional base pairs coding for 3 prolines in B73), 44 SNPs were observed between the two alleles. This made an average of 1 SNP every 35.3 bp between the LC and B73 alleles. From this, for the coding region, with a length of 662bp, only 6 SNPs were observed, which is an average of 1 in every 110 bp. Out of the 6 SNPs three were synonymous substitutions, causing no consequential change in the protein sequence, and the other three were nonsynonymous, causing a change in the protein sequence at the corresponding positions. The remaining SNPs represented noncoding or silent variations. On the whole, this pattern of single

nucleotide polymorphism is similar to the known SNP values for allelic loci in maize (Tenaillon et al., 2001).

Analysis of the upstream regions of both the wild type alleles, 1661 bp from LC and 1524 bp from B73 (Fig. 3.5), showed that they are largely homologous to each other. The sequence in the proximal region of the transcription start site, up to approximately 150 bp upstream, was almost identical between the two alleles. However, further upstream, the two sequences differed considerably at two positions. In one case, about 150 bp upstream of the start codon, approximately 200 bp of the two sequences showed less than 50% homology. At another position further upstream, the B73 sequence lacked approximately 200 bp of the corresponding LC sequence, which together with other smaller gaps between the two sequences constituted about 14% of the two gene upstream regions (Fig.3.5). Taking all these differences together, the two gene upstream sequences were approximately 78% homologous.

Similarly, downstream of the *et1* gene, sequence data up to approximately 400 bp was available from the B73 genomic clone, which showed high homology to that of LC. Apart from a few basepair substitutions and a small deletion of 13 bp in the LC allele (at about 350 bp downstream of the gene), no major changes were observed and the sequences showed 97% homology in this region.




**B.**

```
                  *         20         *         40         *         60         *         80
LC        : CTCGAGGGTATACATGTGGATGAAGAAAAGACTATCAATGATCCAGAACTTCTTATGGAGGTAATTTCATTAATCTTGGA
et1m3wtA  : CTCGAGGGTATACATGTGGATGAAGAAAAGACTATCAATGATCCAGAACTTCTTATGGAGGTAATTTCATTCATCTTGGA
et1m3wtB  : ................................................................................
LCcDNA    : ................................................................................
```

**Fig. 3.5:** (continued, legend on page 73)

```
                  *       100         *       120         *       140         *       160
LC       : AATTGGAATACACGTGCTGGCATGCAGTAAATATAACCATCCTTCTCATTTCATGTATCTTTTGATAACTGTGAATAACA
et1m3wtA : AATTGGAATACACGTGCTGGCATGCAGTAAATATAACCATCCTTCTCATTTCATGTATCTTTTGATAACTGTGAATAACA
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       180         *       200         *       220         *       240
LC       : TTGCAACGATTGGAA.........CACACATTCAATGTGTTAAACATCCTATGTAAAGTACATTAAAATGAATGAACTCTG
et1m3wtA : TTGCAACAATTGGAAGTTTGGAACACACATTCAATGTGTTAAACATCCTATGTAAAGTACATTAAAATGAATGAACTCTG
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       260         *       280         *       300         *       320
LC       : AATTGAGATGAATACAACAACACAAACAGTTTATTTATATACATCTTATTCCATTAACTTACACGAAATGAAGATGACAA
et1m3wtA : AATTGAGATGAATACAACAACACAAACGGTTTATTTATGTACATCTTATTCCATTAACTTACACGAAATGAAGATGACAA
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       340         *       360         *       380         *       400
LC       : TGTTCTCCGTAAATCAATAACAAATCGCTTGCAGATGATGGATATACGGGAAGCTGTCAGTGATGCCAGTGATTCTCAAA
et1m3wtA : TGTTCTCCATAAATCAATAACAAATCGCTTGCAGATGATGGAGATACGGGAAGCTGTCAGTGATGCCAGTGATTCTCAAA
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       420         *       440         *       460         *       480
LC       : CCCTGGAGAAGATCCAATCTCAGGTTCTGATCTCTGAATAAAGCTACACGTGCCATAACATAACCAGGGAGCATGATTGC
et1m3wtA : CCCTGGAGAAGATCCAATCTCAGGTTCTGATCTCTGAATAAAGCTACACGTGCCATAACATAACCAGGGAGCATGATTGC
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       500         *       520         *       540         *       560
LC       : GTCGCAATGGTTTCTGGTCTCATAGCCTTTTGTTCCCTTTTCCAGGTTAAGGCAAAGCTCGAAACATGGTCCGATTCCTT
et1m3wtA : GTCGCAATGGTTTCTGGTCTCATAGCCTTGTGTTCCCTTTTCCAGATTAAGGCAAAGCTCGAAACCTGGTCCGATTCCTT
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       580         *       600         *       620         *       640
LC       : CCAGGAGGCATTTGACAGGAAGGATTTTGACCGTGCAGTGGAAGCCACACAGAGAATGAGGTACTATGAACGTGCAGTGG
et1m3wtA : CCAGGAGGCATTTGACAGGAAGGATTTTGACCGTGCGGTGGAAGCCACACAGAGAATGAGGTACTATGAACGTGCAGTGG
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       660         *       680         *       700         *       720
LC       : AAGAAACAGTGAAGAAGCTCTGAGCCTCTGATCTTCTGTTACAGCTGCCAAAGGTGGTGAGGCTTGTGTCTTTTCGTACG
et1m3wtA : AAGAAACAGTGAAGAAGCTCTGAGCCTCTGATCTTCTGTTACAGCTGCCAAAGGTGGTGAGGCTTGTGTCTTTTCGTACG
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       740         *       760         *       780         *       800
LC       : .TTCCTTGCAGATCCTCCTGTGCTCACGCAGCTACATTGCAACAGTAACATAGTACAGAGTAGGTCAATTTTTTTACTCA
et1m3wtA : GTTCCTTGCAGATCCCCCTGTGCTCACGCAGCTACATTGCAACAGTAACATAGTACAGAGTAGGTCATTTTTTTTACTTA
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................

                  *       820         *       840         *       860         *       880
LC       : TACGTTGCAATGTATTATATGACTACATTGGAATATATTACTGTAGCTACATTGCAACAGTAATATAT..GACTACAGAA
et1m3wtA : TACGTTGCAATGTATTATATGACTACATTGGAATATGTTACTGTAGCTACATTGCAACAGTAATATATATGACTACACAA
et1m3wtB : ................................................................................
LCcDNA   : ................................................................................
```

**Fig. 3.5:** (continued, legend on page 73)

```
                 *       900         *       920         *       940         *       960
LC      : GAGCAGACGTTGTGATGCTAAATTCCAAATGTATCCATGGCTGGTGCAGAGTAGGTCATTTTTTTT.ACTTATACTTTCC
B73A    : GACCATACGTTGTAATGCTAAATTCCAAATGTATCCATGGCTGGTGCACCGCAGGTAATTTTTTTTTACTTATACTTCCC
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *       980         *      1000         *      1020         *      1040
LC      : ACAATTCCAAATGGATCCTTGGCTTTTCAATAAAACAAAACTTGGTGTGTTTGGTTGTGGGACAGCTAGAATAGGGACGT
B73A    : ACAATTCCAAATGCATCCTTGGCTTTCCAATAAAACAAAACT.....................................
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1060         *      1080         *      1100         *      1120
LC      : CCTCTCTCGTCTCTCCAATTTTTGAGATATAACTGGGACAACATTGGGATAGTCTTGTCTCAACCTTTGATTCTAAATCA
B73A    : ................................................................................
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1140         *      1160         *      1180         *      1200
LC      : AACAATCTTATTTGAGGGATCGTCATATCCCGTCCCATCGTCCTGTTACTGCAACCAAATGCATCATAAAAGAGTAGATA
B73A    : ............................................................AAAAGAGTAGATA
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1220         *      1240         *      1260         *      1280
LC      : GGGATTCAGAATAAGGCTTCGATGCCATATATTTATAGAGCGCAAACCTTTGTATTCTGGTGTGGCCGTGTGGGTTTGTT
B73A    : GGGATTCAGATTAAGGCTTTGATGCCATATATTTATAGAGCGCAAACCCTTGTATTCTGCTGTGGCT.TGTGGGTTTGTT
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1300         *      1320         *      1340         *      1360
LC      : GGAATGTACG.....TGGT.....GGATGTGATGCATGTTAAGCGAGGATGTTCTACCCGGATCAAAAGGCCGTTGATGA
B73A    : GGAATGTACCATTGCTGGTTACGTGGATGTGATGCATGTTATGTGAGGATGTTCTACCCGGATCAAAAGGCCGTTGATGA
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1380         *      1400         *      1420         *      1440
LC      : ..............GAGGGACAGACTCAGATACAGTTAGCGTCGGG.........CTCACTCAGGCCATGAT.G..AA
B73A    : CAATAATGTACGATGAGAGGGACAGACTCAGATAC......CGTCGGGTTCGAGGATGTGTTTGTTTGCGAGATAGCTAA
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1460         *      1480         *      1500         *      1520
LC      : TCACTGCCAGAAATACGAGTTTTTGTTTTTTTTGCCGTCCACAAGTCCCAGAGAGGAGACTCAGGAGAAATACCAGAATT
B73A    : GAACAGGCCTGGGGACAATATTGAGATAAATCTTATCTTAACATCTGACTCTGAAACAAACAACCTTATTTAAGGGAACT
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1540         *      1560         *      1580         *      1600
LC      : .TCTTCCGGCGGCCGCAAAAACCA..GCGAGAGCA.TGAGAACG.GAGAACCATCGCCGTCGGGACCGTGACGCTGGACG
B73A    : ATCTCGTTATATTCATTACCACCATAGCAGAAGCAGCGAGAGCACAAGAACCATCGCCGCCGGGACCGTGACACTGGACG
B73B    : ................................................................................
LCcDNA  : ................................................................................

                 *      1620         *      1640         *      1660         *      1680
LC      : GCCTAAAGCGGCTGCGGGGAGGTGTGTTAAACGGGCCTTTCCTGGGCTACGGTGGACAGCCCAGTTACTAAGTTCACCCT
B73A    : GCCTGAATCGGCTGCGGGGAGGGGTGTTAAACGGGCCTTTCCTGGGCTACGGTGGACAGCCCAGTTACTAAGTTAACCCT
B73B    : ................................................................................
LCcDNA  : ................................................................................
```

**Fig. 3.5:** (continued on the next page)

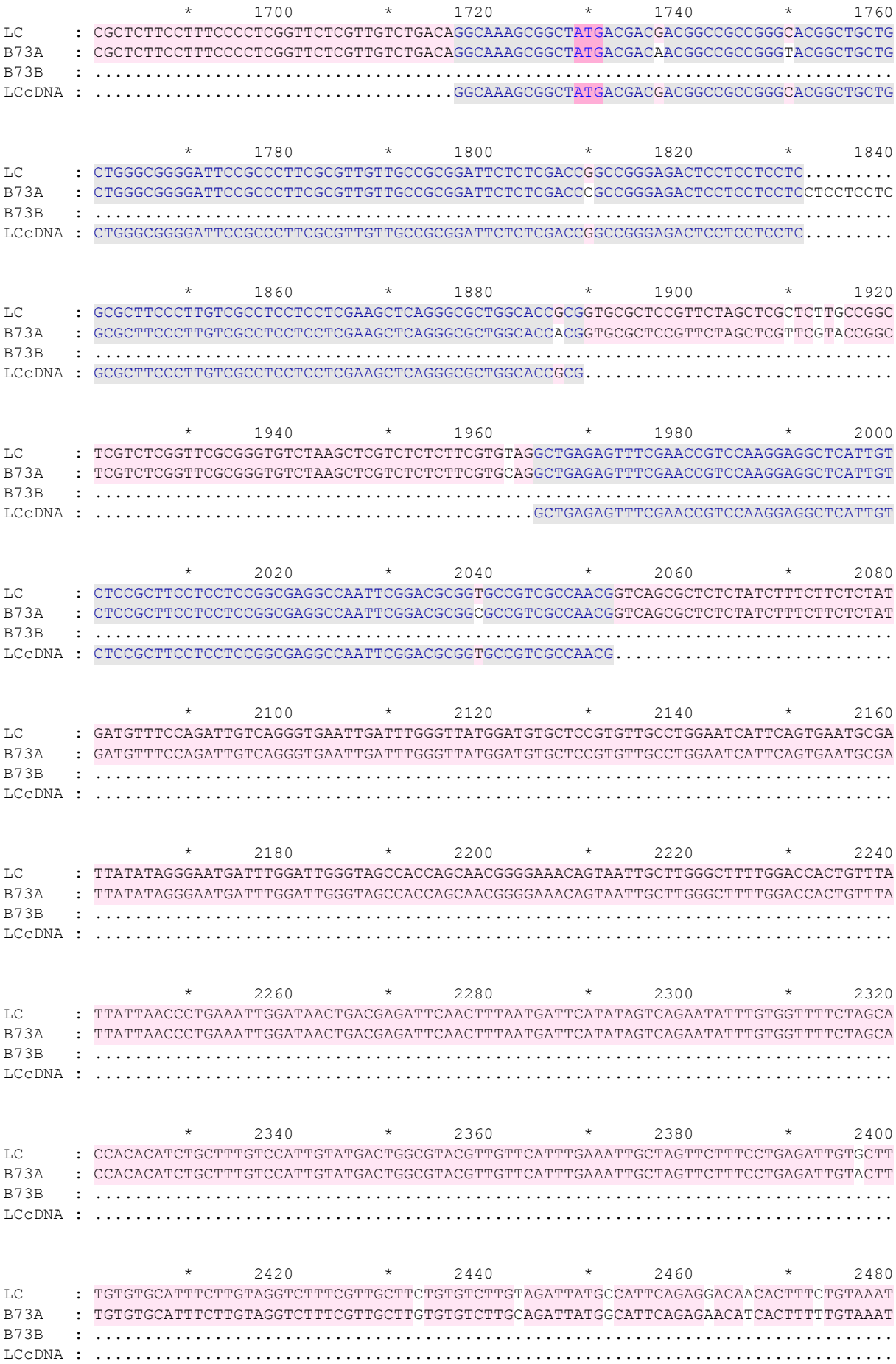



**Fig. 3.5:** (continued on the next page)

```
                *      2500         *      2520         *      2540         *      2560
LC      : TTGGCAGGTTGATCGCATACATATGTGAACATATGAAAAAGAAATGATATTATATAATTCCGCGGTTTATTGCATGAAAG
B73A    : TTGGCAGGTTGATCGCATACATATGTGAGCATATGAAAAAGAAATGATATTATATAATTCAGCGGTTTATTGCATGAAAG
B73B    : ................................................................................
LCcDNA  : ................................................................................

                *      2580         *      2600         *      2620         *      2640
LC      : ACAACATAACAGATTGCTTGCTGATTGAGGAACAAACTAGATTCAGAAATGAATAACAAGTGCTCAACTTTGCTGGTAGA
B73A    : ACAACATAACAGATTGCTTGCTGATTGAGGAACAAACTAGGTTCAGAAATGAATAACAAGTGCTCAACTTTGCTGGTAGA
B73B    : ................................................................................
LCcDNA  : ................................................................................

                *      2660         *      2680         *      2700         *      2720
LC      : TTGATTACTGAAGCGGCAAACTGCAGACTAGTTTATAGGCCGGAATCTCCTTAATGTACCTGTTATGCGTGTGATGTTTC
B73A    : TTGATTACTGAAGCGGCAAACTGCACACTAGTTTATAGGCCGGAATCTCCTTAATGTACATGTTATTCGTGTGATGTTTC
B73B    : ................................................................................
LCcDNA  : ................................................................................

                *      2740         *      2760         *      2780         *      2800
LC      : AGGAAGCCGCTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAATTTACATGCAACGCATGTGGCGAAAGGACC
B73A    : AGGAAGCCGCTATTGATATAAAGCTT.....................................................
B73B    : .................................AAGCTTGCTTGTTCAATTTACATGCAACGCATGTGGCGAAAGGACC
LCcDNA  : ..GAAGCCGCTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAATTTACATGCAACGCATGTGGCGAAAGGACC

                *      2820         *      2840         *      2860         *      2880
LC      : AAGCGCTTGATAAACAGAGTAGCCTATGAAAGAGGCACAGTTTTTCTTCAGGTAAAAACTTCTTTCCATGTAAATGCAAA
B73A    : ................................................................................
B73B    : AAGCGCTTGATAAACAGAGTAGCCTATGAAAGAGGCACAGTTTTTCTTCAGGTAAACACTTCTTTCCATGTAAATGCAAA
LCcDNA  : AAGCGCTTGATAAACAGAGTAGCCTATGAAAGAGGCACAGTTTTTCTTCAG...........................

                *      2900         *      2920         *      2940         *      2960
LC      : CTTCAGACCTACTGCAGCTTTGCTCATTCTGGTCCGCTGTATGAGAAAAGACCACATTCGTATGTTTGACGACACATACC
B73A    : ................................................................................
B73B    : ATCCAGACCTACTGCAGCTTTGTTTATTTTGGTCCGCTGTATGAGAAAAGCCCACATTCATATGTTTGACGACACATGCC
LCcDNA  : ................................................................................

                *      2980         *      3000         *      3020         *      3040
LC      : GTAAATTTATGTTCTAACATACACTGCACAAAACAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAATCTTGG
B73A    : ................................................................................
B73B    : GTAAATTTATGTTCTAACATACACTGCACAAAACAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAATCTTGG
LCcDNA  : ..................................TGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAATCTTGG

                *      3060         *      3080         *      3100         *      3120
LC      : GCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCGGTGGACACCAGTTCTGAAGATTGAT
B73A    : ................................................................................
B73B    : GCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCGGTGGACACCAGTTCTGAAGATTGAT
LCcDNA  : GCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCGGTGGACACCAGTTCTGAAGATTGAT

                *      3140         *      3160         *      3180         *      3200
LC      : CTGCTGTGAGAAGCGATGTTGGTATGCAAAACGCCCTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCAT
B73A    : ................................................................................
B73B    : CTGCTGTGAGAAGCGGTGTTGGTATGCAAAACGCCCTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCAT
LCcDNA  : CTGCTGTGAGAAGCGATGTTGGTATGCAAAACGCCCTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCAT

                *      3220         *      3240         *      3260         *      3280
LC      : AAAAATGGCACTTTTTAAAGTTGTTGCACATACTCATCTGAAATTCTGAATACAGCAGGGCCTACATTGTACTTCTTACA
B73A    : ................................................................................
B73B    : AAAAATGGCACTTTTTAAAGTTGTTGCACATACTCATCTGAAATTCTGAATACAGCAGGGCCTACATTGTACTTCTTACA
LCcDNA  : AAAAATGGCACTTTTTAAAGTTGTTGCACATACTCATCTGAAATTCTGAATACAGCAGGGCCTACATTGTACTTCTT...
```

**Fig. 3.5:** (continued on the next page)

```
                 *      3300         *      3320         *      3340         *      3360
LC      : ACGTGATATTTATATACTAGTCTATTAAGAATCTAATATAAGTACCTAGTGCTACCACGACTTCACCCTCTACGTTGATA
B73A    : ..........................................................................
B73B    : ACGTGATATTTATATACTACTCTATTAAGAACCTAATATAAGCACCTAGTGCTACCACGACTTCACCCTCTACGTTGATA
LCcDNA  : ..........................................................................

                 *      3380         *      3400         *      3420         *      3440
LC      : CTCTGGACCTGCCTCAGGCATCCCGCCCTAGGCACCGTTCCCTGTGGGCCCAGCGCCCACACAACGTTCTCTCAGCTACT
B73A    : ..........................................................................
B73B    : CTCTGGACCTGCCTCAGGCATCCCGCCCTAGGCACCGTTCCCTGTGGGCCCAGCGCCCACACAGCGTTCTCTCAGCTACT
LCcDNA  : ..........................................................................

                 *      3460         *      3480         *      3500         *      3520
LC      : TAATTGTGGACCCCCAGCATCCGCGCGCAGCCACCACACCACGCAAAAATAGTGCAGAACAAACACTTGAACTCACGCTT
B73A    : ..........................................................................
B73B    : TAATTGTGGACCCCCAGCATCCGCGCGCAGCCACCACACCACGCAAAAATAGTGCAGAACAAACACTTGAACTCACGCTT
LCcDNA  : ..........................................................................

                 *      3540         *      3560         *      3580         *      3600
LC      : CGTTGCTCTAGAAATTTAGGACTAACCACCAAACCACACATAATTTGGTGTTTAGAAATAAACAATAATTATATTTGAAT
B73A    : ..........................................................................
B73B    : CGTTGCTCTAGAAATTTAGGACTAACCACCAAACCACACATAATTTGGTGTTTAGAAATAAACAATAATTATATTTGAAT
LCcDNA  : ..........................................................................

                 *      3620         *      3640         *
LC      : AAATATCTTAA.............ATACAATTGG.TTAGTTTGGTTTTAAGAA.
B73A    : ..................................................
B73B    : AAATATCTTAATCCTATTTATTTGATACAATTGGGTTAGTTTGGTTTTAAGAA.
LCcDNA  : ..................................................
```

**Fig. 3.5: Comparison of the genomic sequence of the two *et1* genes from LC and B73.** (**A**) The two sequences represent the *et1* genes from LC and B73. The four exons (blue boxes) are represented with Roman numerals with their lengths in base pairs (bp) below. The introns are depicted with their lengths in bp. (**B**) The first sequence in the alignment represents the *et1* gene from LC. This is followed by the sequences of two genomic subclones, B73A and B73B, representing the *et1* gene from B73 background. The *et1* cDNA from LC is also depicted below the sequences at corresponding positions. The start and stop codons have been marked. As can be seen, the sequences are largely homologous, although a number of differences are also present at a few positions between the two genes. The gene upstream region also shows a few variations at two different positions between the two genomic sequences.

### 3.2.5. Characterisation of *et1* expression in LC and *et1-Ref* backgrounds

In order to analyse the expression pattern of the *et1* gene, Northern blots containing total RNA from different parts of the plant from LC and *et1-Ref* lines were prepared. In one experiment, total RNA from leaves and roots at different days after germination, tassels at a stage before anthesis, the unfertilised female inflorescence (cob), the unfertilised female flowers, kernels at different stages of development after

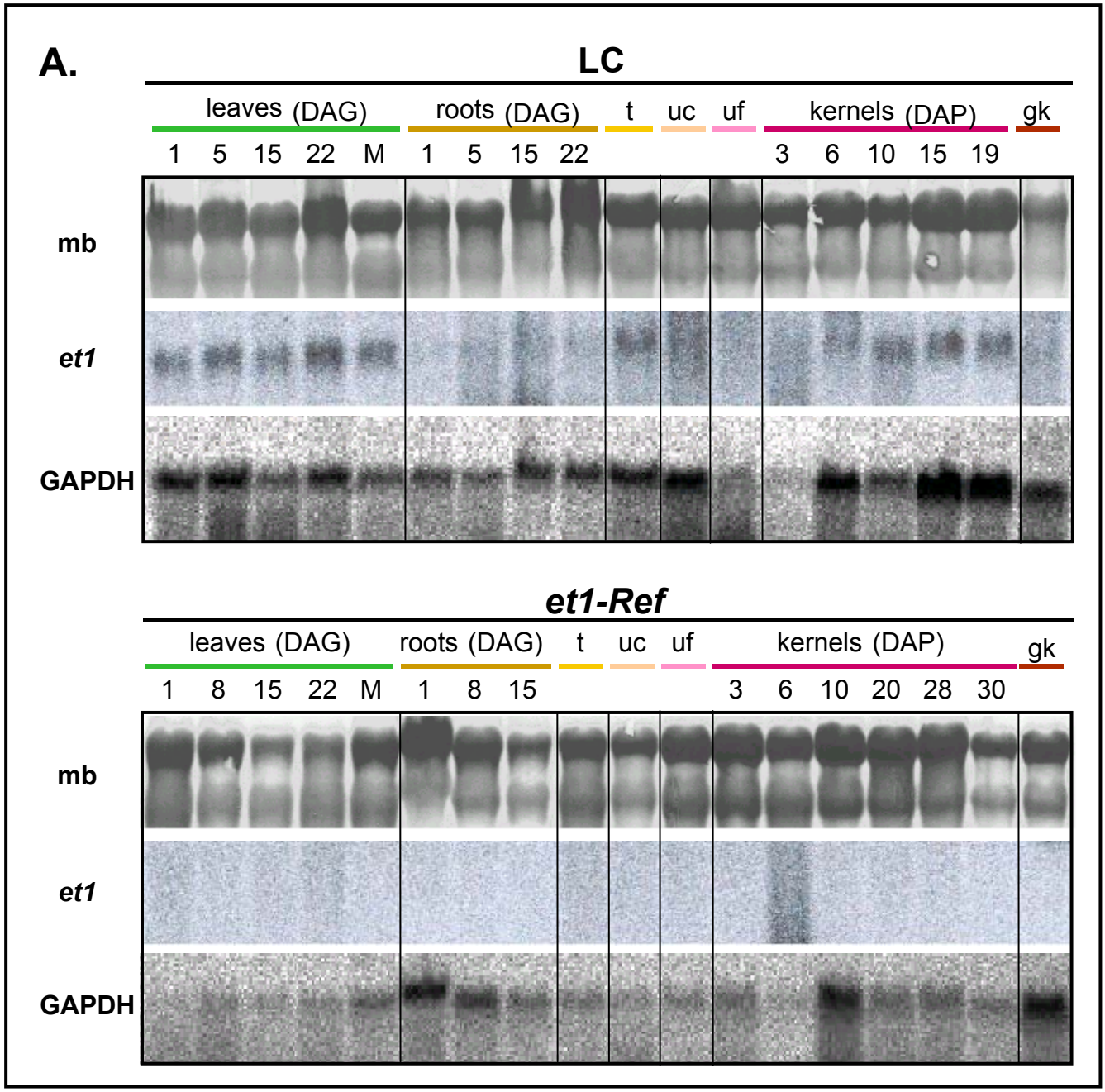
pollination, and germinating kernels from both LC and *et1-Ref* were used (Fig. 3.6 A). These blots were subjected to Northern analysis using *et1* specific probes. In order to ensure a gene specific hybridisation, only a sequence, which did not contain the ORF, was used as a probe. This was necessary, since it was known that a sequence with homology to the *et1* gene was present in the genome, and it was possible that this second sequence could be a functional gene. Therefore, a 300 bp fragment of the *Et1-LC* 3´ untranslated region, which also contained a small part of the genomic sequence downstream of the gene, was used as a probe.

In another experiment, using total RNA obtained from different tissues of LC plants, a Northern analysis was carried out to examine the *et1* expression pattern in more varied types of the plant tissues. However, the *et1* cDNA, *c9.1*, was used as a probe here (Fig.3.6 B).

As can be seen in Fig.3.6, the Northern analysis showed the presence of the *et1* transcript only in LC and not in *et1-Ref*. In addition, the length of the *et1* transcript was within the range expected, based on the sequence of the *et1* cDNA, *c9.1*. The length of the transcript was 700 to 800 bases, which included the poly A+ tail. Although, based on this analysis the actual length of the *et1* 5´untranslated region cannot be identified, it could be concluded that the ORF present in *c9.1* was probably complete.

In LC, the *et1* expression was predominantly observed in the leaves at all the stages examined, in tassels, in kernels and weakly in unfertilised female flowers and cobs. In contrast, only a very faint signal, indicating basal expression, was observed in the roots at all the stages, in 3 DAP kernels and in the germinating kernels.

The Northern analysis of *et1* expression in the second experiment (Fig. 3.6 B) revealed that in addition to leaves, developing kernels and tassels it was also expressed in the stem and coleoptile. That the *et1* expression in the kernels (caryopsis) was mainly restricted to the endosperm, was indicated by the last two samples in Fig. 3.6 B, where RNA from the embryo and the remaining caryopsis was analysed separately. Unlike the sample from embryo (E), where only basal *et1* expression was observable, in the sample from caryopsis without embryo (C-E) a stronger signal with the *et1* probe was detected.



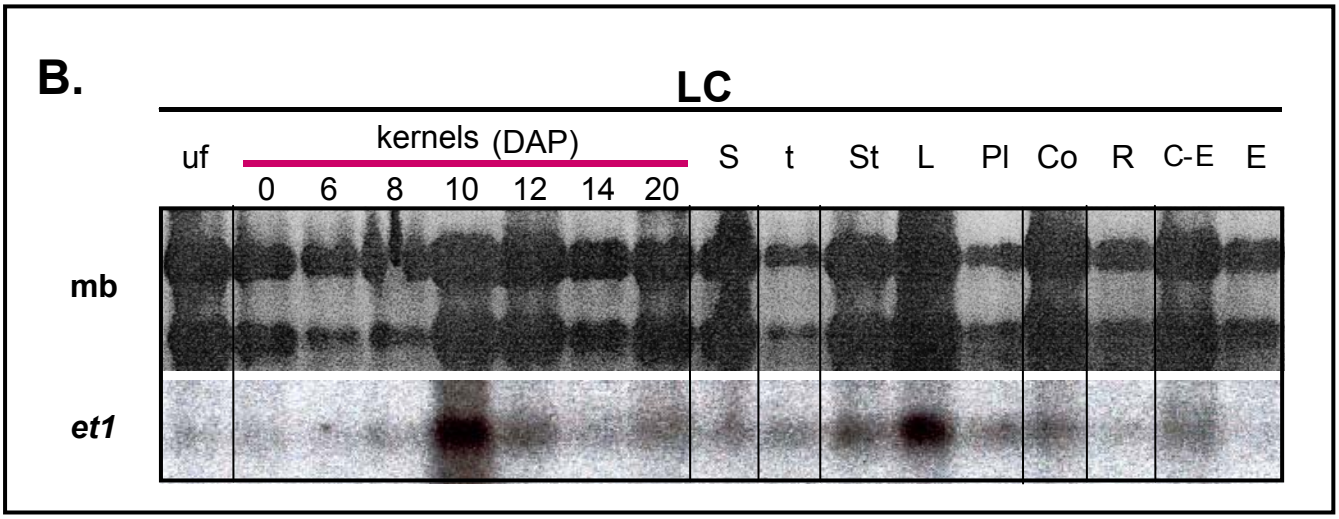




**Fig. 3.6:** (Legend on the next page)

**Fig. 3.6: Northern analysis performed with LC and *et1-Ref* using *et1* specific probes**. **(A)** The Northern blots contain approx. 30 µg total RNA per lane from different plant tissues. Both the blots, before hybridisation, were stained with methylene blue (mb) to stain the rRNA and the RNA marker (1st row). These were then hybridised with the 300 bp *Et1-LC* 3´UTR fragment probe (2nd row). Afterwards, a GAPDH probe was hybridised to the blots (3rd row).

The **LC** blot contains the following samples from left to right: the first four lanes are leaf samples, collected at 1, 5, 15 and 22 days after germination (DAG).The fifth lane is a leaf sample (**M**) mature leaves. The next four lanes represent roots (primary and secondary, when present) collected at 1, 5, 15, and 22 DAG. The next lanes are tassels (**t**) collected at a young stage before anthesis, young unfertilised cob (**uc**), unfertilised female flower (**uf**), kernel samples collected at 3, 6, 10, 15 and 19 days after pollination (DAP) and kernels germinated in the dark (**gk**).

Similarly, ***et1-Ref*** total RNA samples, in the lanes from left to right on the blot are: Leaves at 1, 8, 15 and 22 DAG, and mature leaf (**M**). Roots (primary and secondary) at 1, 8 and 15 DAG. Tassels (**t**) before anthesis, young unfertilised cob (**uc**) and unfertilised female flowers (**uf**). Kernels collected at 3, 6, 10, 20, 28, and 30 DAP, and germinating kernels (**gk**).

**(B)** Northern analysis with LC total RNA samples using the *et1* cDNA, c9.1 as a probe. The blot contained approx. 25 µg total RNA samples per lane. The blot was first stained with methylene blue (**mb**) before hybridising with GAPDH. The samples in the lanes from left to right on the blot are: Unfertilised female flower (**uf**); kernels at different days after pollination (DAP), viz., 0, 6, 8, 10, 12, 14 and 20 DAP; Silk (**S**) and tassel (**t**), both at a young stage; Stem (**St**); Mature leaf (**L**); whole plant (**Pl**) at 16 DAG; Coleoptile (**Co**); primary root (**R**); 20 DAP caryopsis without embryo (**C-E**); and 20 DAP embryo (**E**).

Based on these results, it could be concluded that the *et1* transcript is predominantly present in the leaves from a very young stage, in tassels, in stem and in kernels approximately 6 DAP onwards, with the expression restricted mainly to the endosperm. However, a weak signal indicating basal *et1* expression could be observed in all the other tissues analysed.

## 3.3. Characterisation of ETCHED1 protein structure and function

The isolation of *et1* cDNA from developing kernels and the Northern analyses revealed the expression pattern as well as the structure of the *et1* transcript. In order

to further analyse the correlation between the *et1* phenotype and the *et1* expression, characterisation of the ET1 protein was required.

```
  1  GGC AAA GCG GCT ATG ACG ACG ACG GCC GCC GGG CAC GGC TGC TGC   45
  1                  M   T   T   T   A   A   G   H   G   C   C    11

 46  TGG GCG GGG ATT CCG CCC TTC GCG TTG TTG CCG CGG ATT CTC TCG   90
 12   W   A   G   I   P   P   F   A   L   L   P   R   I   L   S   26

 91  ACC GGC CGG GAG ACT CCT CCT CCT CGC GCT TCC CTT GTC GCC TCC  135
 27   T   G   R   E   T   P   P   P   R   A   S   L   V   A   S   41

136  TCC TCG AAG CTC AGG GCG CTG GCA CCG CGG CTG AGA GTT TCG AAC  180
 42   S   S   K   L   R   A   L   A   P   R   L   R   V   S   N   56

181  CGT CCA AGG AGG CTC ATT GTC TCC GCT TCC TCC TCC GGC GAG GCC  225
 57   R   P   R   R   L   I   V   S   A   S   S   S   G   E   A   71

226  AAT TCG GAC GCG GTG CCG TCG CCA ACG GAA GCC GCT ATT GAT ATA  270
 72   N   S   D   A   V   P   S   P   T   E   A   A   I   D   I   86

271  AAG CTT CCT AGA AGA AGC TTG CTT GTT CAA TTT ACA TGC AAC GCA  315
 87   K   L   P   R   R   S   L   L   V   Q   F   T   C   N   A  101

316  TGT GGC GAA AGG ACC AAG CGC TTG ATA AAC AGA GTA GCC TAT GAA  360
102   C   G   E   R   T   K   R   L   I   N   R   V   A   Y   E  116

361  AGA GGC ACA GTT TTT CTT CAG TGT GCA GGG TGC CAG GTG TAC CAT  405
117   R   G   T   V   F   L   Q   C   A   G   C   Q   V   Y   H  131

406  AAG TTT GTT GAT AAT CTT GGG CTA GTT GTT GAG TAT GAT CTA CGA  450
132   K   F   V   D   N   L   G   L   V   V   E   Y   D   L   R  146

451  GAA GAA AAC GAG CTA CAA GGA GAA AAT GCG GTG GAC ACC AGT TCT  495
147   E   E   N   E   L   Q   G   E   N   A   V   D   T   S   S  161

496  GAA GAT TGA TCT GCT GTG AGA AGC GAT GTT GGT ATG CAA AAC GCC  540
162   E   D   -                                                 163

541  CTG TAC TCT GTA GGT TTT TGA CAA CAT TGG TTA TTT GTA TAG CAT  585
586  AAA AAT GGC ACT TTT TAA AGT TGT TGC ACA TAC TCA TCT GAA ATT  630
631  CTG AAT ACA GCA GGG CCT ACA TTG TAC TTC TT                   662
```

**Fig. 3.7: An alignment of the nucleotide sequence of *et1* cDNA from LC and the deduced ET1 protein sequence.** The cDNA sequence (first row in the alignment), which is 662 bp long, is followed by the ET1 protein sequence (163 aa long) below it, where the amino acid sequence corresponds to the triplet codons in the cDNA sequence above. The length of the respective sequences from the start is depicted both at the beginning as well as the end of each line. The protein sequence is written in bold. The translation stop codon is depicted with a dash in the protein sequence.

The deduced ET1 protein sequence (Fig. 3.7) was first analysed for cellular localisation using online databases, before beginning with the experimental analysis. This was followed by the comparison of the ET1 protein to online protein domain databases to identify a functional domain in the mature ET1 protein.

### 3.3.1. ETCHED1 cellular localisation analysis

Microscopic observations of the pleiotropic *et1* phenotype revealed two features: The presence of starchless cells in the endosperm of developing kernels (Fig. 1.3; Fig. 3.1), which suggested disturbed amyloplast development, as well as undifferentiated and malformed chloroplasts in the pale young seedling leaves (Fig.1.3). These observations indicated that the ET1 protein might be directly involved in plastid development. Based on these observations, the ET1 protein sequence was analysed for the presence of plastid localisation signals and the protein experimentally tested for plastid localisation.

#### 3.3.1.1. *In silico* analysis indicates the presence of an N-terminal plastid targeting signal sequence

In order to analyse the ET1 protein for the presence of cellular localisation signals, the different programs available at the ExPASy server, ChloroP, TargetP, Predotar and PSORT were used. Based on these analyses, the ET1 protein was found to show the highest possibility of being localised in the plastids as compared to other cellular organelles like the nucleus, mitochondria, vacuoles, microsomes, peroxisomes, and others. The presence of a signal peptide for entry into the secretory pathway (endoplasmic reticulum) was also not detected (TargetP, PSORT). Moreover, both TargetP and Predotar indicated the absence of a mitochondrial target peptide at the N-terminal region of ET1.

According to the ChloroP as well as TargetP prediction results, the ET1 proteins showed the highest probability for containing N-terminal plastid targeting signals as compared to that for other organelles, viz., mitochondria, secretory pathway and others locations (Emanuelsson et al., 2000; Nielsen et al., 1997). For the ET1 protein from LC, with a length of 163 amino acids, the highest score (neural network output) was 0.838 – indicating highest probability – for the presence of a plastid transit sequence, in a reliability class 2 (out of 5), which is the second highest reliability range below reliability class1 (The reliability class is calculated by subtracting the second highest score from the first highest score. The higher the resultant score, the higher is the reliability class and the greater is the certainty of the result). For the ET1 putative protein from B73, with a length of 166 amino acids, the highest score (neural

network output) was 0.799 for the presence of a plastid transit sequence, also in a reliability class 2.

These results indicated that ET1 contains an N-terminal transit sequence (peptide) for the plastids. The N-terminal transit sequence was predicted to be 64 aa long for ET1 from LC and 67 aa for ET1 from B73 respectively. The length of the mature ET1 protein, after the cleavage of the transit sequence, would then be 99 amino acids. According to Predotar analysis, where the possibility of a protein containing either mitochondrial or plastid targeting signals is tested, the ET1 protein was interpreted to be plastid localised with a 60% probability, while only a 0% probability for mitochondrial localisation was predicted.

In order to characterise this putative transit sequence, the N-terminal region of the ET1 protein was examined for the presence of the characteristic features of a plastid targeting transit sequence. Although no common consensus sequence or length of plastid transit sequences has been identified so far, a few characteristic features have been identified from a large number of the sequences (Fulgosi and Soll, 2001; Cline and Henry, 1996). The transit sequences ranging from 30 to 150 amino acids are rich in serine and threonine, and contain a three domain structure. The amino terminal part contains mostly uncharged residues. The middle part predominantly harbours hydroxylated and positively charged residues and is rich in serine, threonine, lysine and arginine. The carboxy-terminal region contains a loosely conserved consensus sequence, Ile/Val-x-Ala/Cys-Ala, for proteolytic processing (von Heijne et al., 1989; de Boer and Weisbeek, 1991; Bartling et al., 1990).

The ET1 N-terminal sequence, on analysis, seemed to have the three-domain structure of a plastid transit sequence. The amino terminal region was mostly devoid of charged residues (Fig. 3.8). The central region contained a large number of positively charged as well as hydroxylated residues. The carboxy terminal region also showed homology to the C-terminal consensus sequence mentioned above.

Apart from the three-domain structure, the transit sequence also contains sites of phosphorylation, which occurs at the serine or threonine residues. This site of phosphorylation is recognised by the hetero-oligomeric cytosolic guidance complex, consisting of 14-3-3 proteins, HSP 70 and probably other unknown components (May and Soll, 2000). A consensus motif for the site of phosphorylation has been

determined (Waegemann and Soll, 1996), which is recognised by a protein kinase: P/G-$X_n$-R/K-$X_n$-S*/T*-$X_n$-S*/T*, where n= 0-3 residues and S*/T* are the phosphate acceptors.

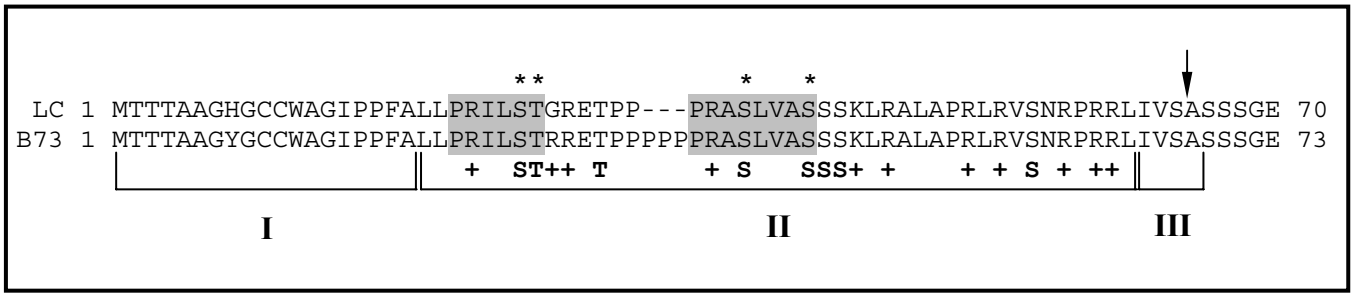



**Fig. 3.8: An alignment of the N-terminal regions of ET1 preproteins from LC and B73 showing the structure of the putative transit sequence.** The ET1 presequence was characterised based on the known general characteristic features of chloroplast transit sequences (see description in text). The first sequence is from the LC background with the B73 sequence below it. The length of the presequence from the start (1) is depicted on the right of the sequence. The predicted (TargetP) transit sequence length is 64 residues. The site of cleavage is depicted with an arrow above the alignment. The three domain structure of the transit sequence has been marked accordingly. The hydroxylated and positively charged residues have been marked below the respective residues in the central domain. The two putative phosphorylation site motifs are shaded grey, where the phosphate acceptors have been marked with an asterisk above.

### 3.3.1.2. Chloroplast Import experiment with the *in vitro* translated ET1 protein

The possibility of a protein being plastid localised can be tested experimentally through chloroplast import experiments. It is now widely accepted that all the different plastid types use the same protein import pathway through the plastid envelop (Soll et al., 1998; Klösgen and Weil, 1991; Klösgen et al., 1989). Hence, in order to test the possibility that the ET1 protein is plastid localised, chloroplast import experiments were carried out. Through this experiment it was possible to simultaneously test if the chloroplasts and the amyloplasts are possible sites of ET1 protein localisation, where, based on earlier observations, the ET1 protein was suspected to be playing a role during their differentiation and development.

The chloroplast import experiment was performed using chloroplasts isolated from young spinach (*Spinacia oleracea*) leaves. The ET1 protein was translated *in vitro*

from the *in vitro* transcribed *et1* mRNA using radioactively labelled $S^{35}$-Cysteine. The approximately 20 kD *in vitro* translated ET1 protein could be detected with a Phospho-Imager plate exposed to the protein sample run on an SDS-polyacrylamide mini-test gel.

The chloroplast import experiment and the following analysis carried out with the radiolabelled ET1 protein is described in chapter 2.2.10.2. After the import, the chloroplasts were lysed and after different treatments divided into three fractions. One fraction, containing the stroma and most of the chloroplast envelop membrane, was referred to as the Stromal fraction. The next two fractions, containing thylakoid membranes, were designated as the Plus and Minus fractions, depending on the protease treatment. The Plus fraction represented proteins present in the thylakoid lumen and the integral proteins of the thylakoid membrane. The Minus fraction included the proteins present on the exterior side of the thylakoid membrane in addition to the integral and lumenal ones. These two fractions also contain small amounts of the chloroplast envelop membrane. These three fractions were analysed through autoradiography of the SDS-PAGE carried out with the three samples (Fig. 3.9).

In order to further prove that the ET1 protein is imported through the same protein import machinery known from chloroplasts, another similar experiment was carried out in parallel. In this experiment, an unlabelled 33 kD protein subunit (pOE33) of the oxygen-evolving complex (OEC), present in the thylakoids, was used as a competitor to the ET1 protein in the chloroplast import reaction. In a series of four reactions, the unlabelled competitor was added in increasing amounts. The basis of the experiment was that the unlabelled protein, which is known to be located in the thylakoids, would compete with the radioactively labelled test protein for import through the chloroplast envelop. Increasing amounts of the unlabelled protein added to the labelled test protein would cause a decrease in the intensity of the signal. If the test protein is imported into chloroplasts would, therefore, be observed through the radioactive signal detected by a Phospho-Imager. In the second experiment a signal with diminishing strength would be observed in the series of samples. The strength of the signal would be inversely proportional to the amount of the unlabelled protein used for the corresponding import reactions.

As can be seen in Fig. 3.9 A, the labelled ET1 protein was detected in the Stromal fraction on the SDS-PAGE analysis, indicating that it was imported into the chloroplasts. After import, the ET1 protein was smaller with an estimated molecular weight in the range of 15-17 kD. This would be expected from the ET1 protein as the transit sequence of a chloroplast localised protein, necessary for the import process, is known to be cut off after import. This also indicated that the protein is localised in the chloroplast stroma and is absent in the thylakoid membrane fractions, as no radioactive signal was detected there. The probability that it might be associated with the chloroplast envelop membrane could not, however, be ruled out.

Similarly in the second experiment (Fig. 3.9 B), with increasing amounts of the unlabelled competitor, decreasing amounts of the labelled ET1 protein imported into the chloroplasts was detected in the Stromal fraction.

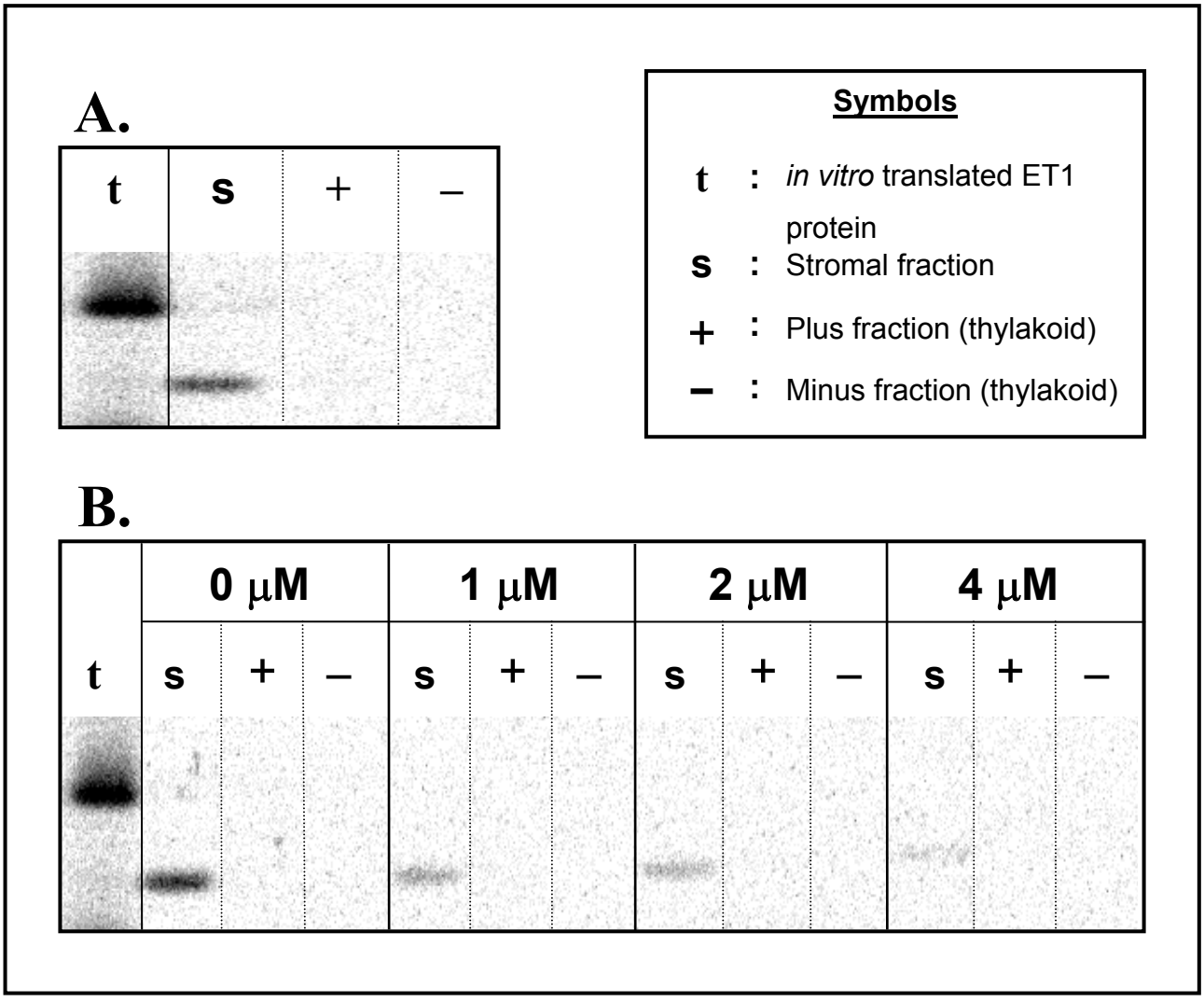



**Fig. 3.9:** (Legend on the next page)

**Fig. 3.9: Chloroplast Import experiment performed with the *in-vitro* translated ET1 protein.** The Chloroplast Import reaction with the $S^{35}$ labelled ET1 protein (t) was carried out at 25°C for 20 minutes in a well-lighted water bath. Afterwards, the chloroplasts were fractionated to give three fractions. These fractions were then analysed with SDS-PAGE followed by exposure to Phospho-Imager plates. The Stromal fraction (S) contained the stroma and most of the chloroplast envelop membrane. The Plus fraction (+) contained thylakoids treated with the protease, thermolysin, to degrade proteins present on their outer surface. Thus, it represented intrinsic proteins of the thylakoids. The Minus fraction (–) contained thylakoids, which were not treated with a protease. Therefore, it represented peripheral as well as intrinsic thylakoid proteins.

**(A)** The *in vitro* translated ($S^{35}$ labelled) ET1 protein (t) after import is detected in the Stromal fraction (+) of the chloroplasts. The ET1 preprotein (t) had a molecular weight of approximately 20 kD. After import, when the transit sequence is cut off, the mature protein (s) is approximately 15 kD in size.

**(B)** In a second experiment, during the chloroplast import, in addition to the $S^{35}$ labelled ET1 protein, a known unlabelled import competitor protein, pOE33, was added to the reaction. This 33 kD subunit of the oxygen evolving complex is thylakoid localised. In a series of four reactions, increasing amounts of the competitor pOE33 were added to the reaction (0, 1, 2 and 4 µM). The SDS-PAGE analysis showed that with increasing amounts of the competitor, decreasing amounts of the ET1 protein could be detected in the stromal fraction. This showed that the ET1 protein is imported through the same protein import pathway as is already known for the competitor protein, pOE33.

### 3.3.2. *In silico* characterisation of *ETCHED1* protein structure and function

After clear experimental evidence of the ET1 protein being chloroplast localised was available, it was important to analyse the function of the protein. The protein primary structure was compared to the online protein domain databases with the aim of identifying a functional domain in the ET1 mature protein. This analysis was extended to include a secondary structure analysis of the ET1 protein. The results of both the analyses were compared with each other to check their compatibility.

#### 3.3.2.1. Identification of a protein domain in ET1 based on homology to protein domain databases

The ET1 protein was compared to the different online protein databases using tBLASTN and BLASTX search tools (Altschul et al., 1997) at NCBI as well as using the advanced PredictProtein search tool (Rost, 1996) separately. The results of both

the analyses showed homology to a number of unknown or hypothetical proteins from plant as well as non-plant sources. In the ProDom data base (Prodom Release 2001.2; Corpet, et al., 1998; Rost et al., 1994), a protein domain, PD022385, was identified, which contained homologues from *A. thaliana*, fruit fly (*Drosophila melanogaster*), fission yeast (*Schizosaccharomyces pombe*), yeast (*Saccharomyces cerevisiae*), mouse (*Mus musculus*) and *Caenorhabditis elegans*. These were also identified in the BLASTX search. The homologous regions of all the proteins were aligned together with MultAlin (Corpet, 1988) at the ProDom server. The homology was found to be mainly restricted to a cysteine rich region in the middle of the 99 residues long ET1 mature protein. In addition to four cysteines, a number of other residues among all the proteins were found to be identical or conserved (Fig. 3.10).

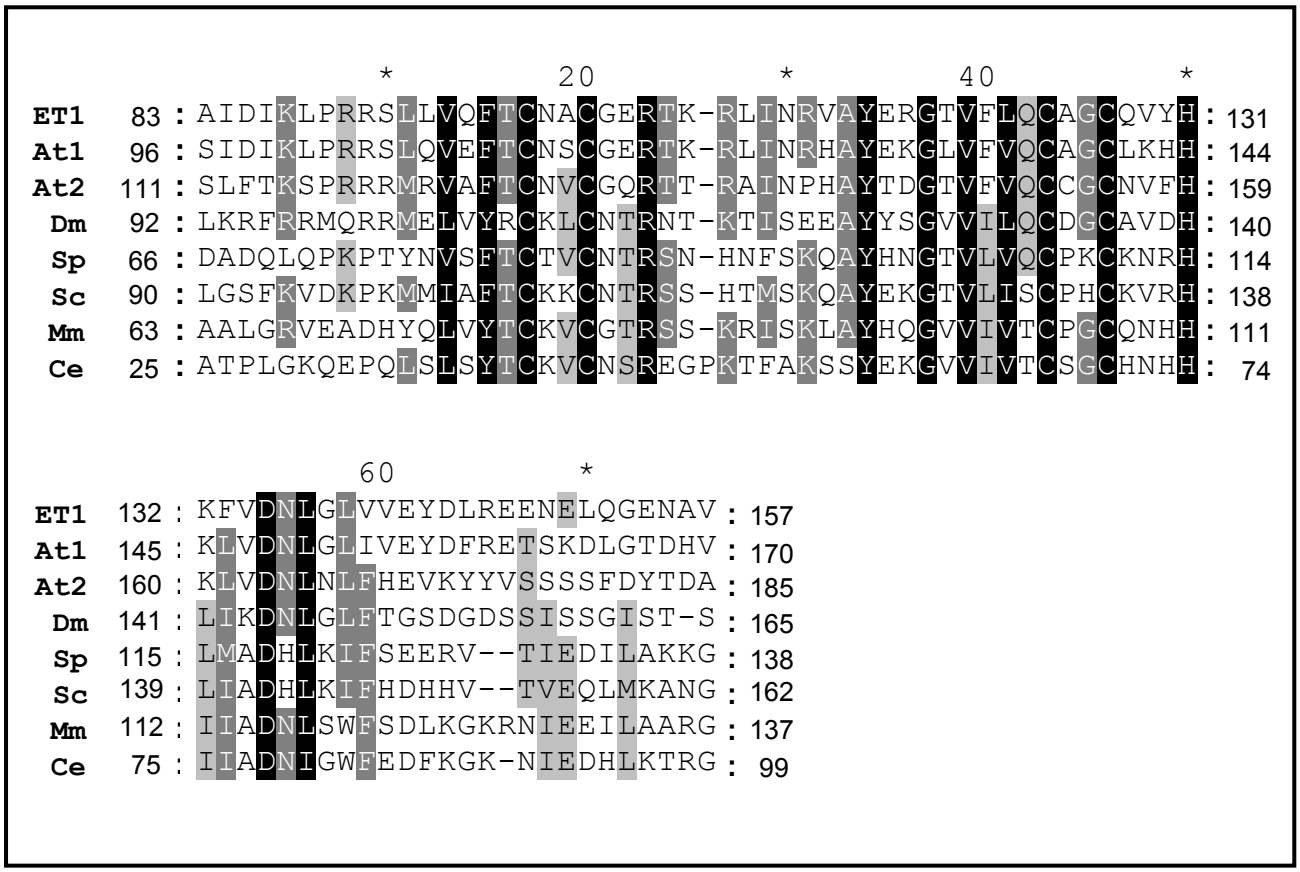



**Fig. 3.10: Alignment of homologous domains of different unidentified proteins in the protein domain, PD022385, to ET1.** Different unknown proteins, identified with BLAST as well as in the protein domain, PD022385, through ProDom database searches, have been aligned to show high homology of all the proteins. The sequences were aligned using MultAlin and GeneDoc. The conserved residues (shaded) have been shaded darker with increasing homology among all the sequences. The first sequence in the alignment is ET1, followed by proteins (from top to bottom) from *A. thaliana*, At1 (AAG52044.1, gi:12324143) and At2 (AAB61076.1, gi:2191191), fruitfly (Dm)

(AAF48501.1, gi:7293117), fission yeast (Sp) (Q09759, gi:1175433), yeast (Sc) (P42844, gi:1176588), mouse (Mm) (BAB23162.1, gi:12835133) and *C. elegens* (Ce) (AAF02170.1, gi:6042159) respectively.

In another analysis carried out to identify other possible protein domains, using the PROSITE tool, Profile Scan (Hofmann, et al., 1999; Bucher and Bairoch, 1994), three databases – Prosite pattern, Prosite profile and Pfam collection of hidden Markov Models – were searched. A homology to the zinc ribbon domain of the TFIIS transcription factor (Fig. 3.11) was found in the Pfam database (Bateman et al., 1999). Based on this analysis, further protein domain searches using protein databanks available at the NCBI server as well as at the PROSITE server were carried out either using the ET1 sequence or description and accession numbers of the identified proteins.

The different sequences obtained through these protein domain searches were compared to the ET1 protein. The analyses yielded some homology of ET1 to the zinc finger domains belonging to *Schizosaccharomyces pombe* (fission yeast), *Saccharomyces cerevisiae* (yeast), *Drosophila melanogaster* (fruitfly) and many mammalian, including human, transcription factors (Fig.3.11). These domains were $C_2$-$H_2$ type and $C_4$ type zinc-fingers, as well as $C_4$ type zinc ribbons. The most homologous of these domains was the zinc ribbon domain. The main features of a zinc ribbon domain, first identified from the eukaryotic transcription elongation factor TFIIS (Qian et al., 1993), are the four cysteine residues present in pairs, where the two cysteines of each pair are separated by two residues and the two pairs are present at a distance of 15-25 amino acids, and the presence of a few conserved aromatic residues around the four cysteines (Awrey et al., 1998). Another protein containing two zinc ribbon domains is Rpb9 (Hemming et al., 2000), a small subunit of RNA Polymerase II. From the NCBI database, a number of other protein entries could be identified, both of eukaryotic as well as archael origin, which contained putative zinc ribbon motifs. Among them were the homologues of Rpb9 present in RNA Polymerases I and III, and in archael RNA Polymerases, the archael transcription factor S (Hausner et al., 2000), enzymatic subunits binding to the RNA transcript during translation in archaea, DNA methylases and many other plant and non-plant proteins with an unidentified function.

However, based on sequence comparisons, only a few of these identified proteins showed reasonable homology and could be aligned together with ET1 (Fig. 3.11). Based on the alignment, homologies to three different proteins could be established. On the one hand, it showed some homology to the prokaryotic dksA/traR C4 type zinc finger proteins (EcC4ZF in Fig. 3.11), on the other hand, it also showed weak homology to the *Drosophila* hunchback zinc finger protein, where two cysteines each of the two tandem $C_2H_2$ zinc finger motifs were homologous to the four cysteines of ET1. However, ET1 showed highest homology to the zinc ribbon motif. It showed the presence of a number of conserved amino acid residues present in the zinc ribbon domains of TFIIS, Rpb9 and some archael Rpb9/TFIIS like proteins.

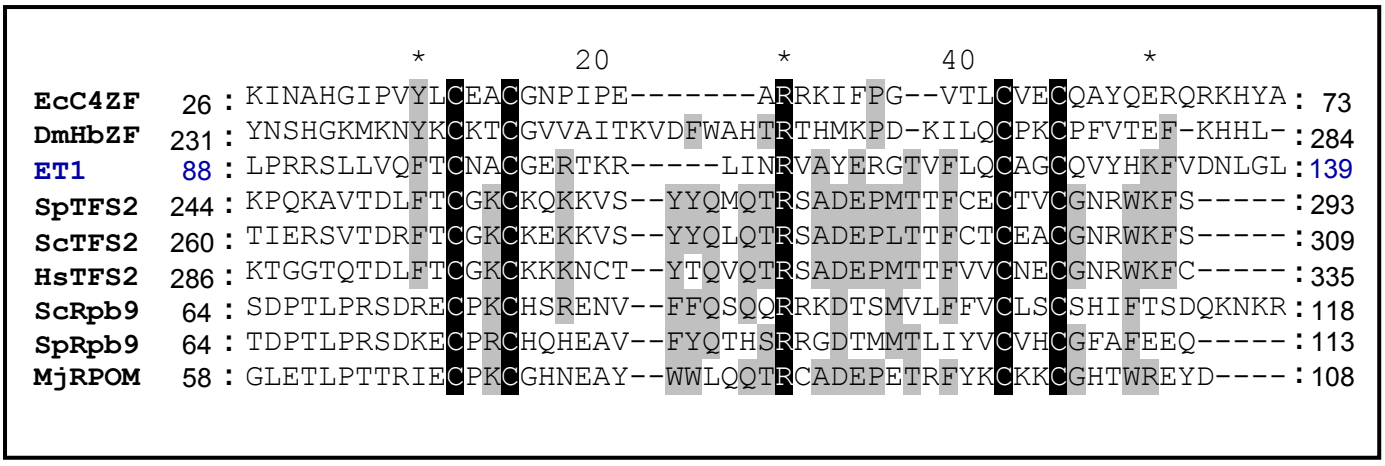



**Fig. 3.11: Alignment of ET1 to homologous zinc binding motifs of different proteins.** The sequences were aligned using GeneDoc. The length of the alignment is shown at the top. The position of the motif in the individual proteins is indicated at the beginning and at the end of each protein. The ET1 protein has been marked with blue. The dashes indicate gaps in a sequence. The conserved residues have been shaded darker with increasing homology among all the sequences in the alignment. The first sequence (EcC4ZF) represents a C4 Zinc finger domain of TraR gene from *E.coli* (gi:730987, P41065), showing some homology to the ET1 protein. Below this a hunchback protein (DmHbZF) with two tandem $C_2H_2$ zinc finger motifs from fruitfly (*Drosophila melanogaster*) (gi:12644098, P05084) is present, followed by the ET1 sequence with the zinc ribbon like domain. ET1 is followed by the zinc ribbon proteins, which are arranged from top to bottom with decreasing homology to ET1. These are: three TFIIS Zinc Ribbon domains from fission yeast (*Schizosaccharomyces pombe*) (gi:1351227, P49373), yeast (*Saccharomyces cerevisiae*) (gi:1729915, P07273), and humans (*Homo sapiens*) (gi:1174652, P23193) respectively, followed by the zinc ribbon domains of Rpb9 subunit of RNA Polymerase II from yeast (gi:6321368, NP_011445.1) and fission yeast (gi:3582119, AB007988.1) respectively and at the end the zinc ribbon domain of Archael (*Methanococcus jannaschii*) RNA Polymease sununit M (gi: 3024570, Q58548).

### 3.3.2.2. Characterisation of the ET1 protein secondary structure

The 99 residue long sequence of the ET1 mature protein, remaining after cleavage of the amino terminal transit sequence (see 3.3.1.1), was analysed for the presence of secondary structure elements. For this, fourteen different methods were used to predict the probabilities for the presence of $\alpha$ helices, $\beta$ strands and $\beta$ turns or loops (White et al., 1994; Stultz et al., 1997; Rost et al., 1996; Rost and Sander, 1993, 1994a; Kneller et al., 1990; Deleage and Roux, 1987; Chou and Fasman, 1978; Levitt, 1978; Garnier et al., 1996; Gibrat et al., 1987; Garnier et al., 1978; Frishman and Argos, 1995; Kabsch and Sander, 1983; Geourjon and Deleage, 1994; Geourjon and Deleage, 1995; Guermeur, 1997). Based on the predictions, a graphical view of the protein´s secondary structure corresponding to the sequence was prepared (Fig. 3.12). The secondary structure of a particular region in the sequence was accepted when the same secondary structure was predicted in at least eight out of the fourteen predictions. In those cases, where the rest of the six or seven methods predicted another secondary structure, it was also taken into consideration and depicted in the graphical view below the sequence (Fig. 3.12).

The amino acids in the sequence were analysed for antiparallel and parallel beta strand preferences (Lifson and Sander, 1979). The residues showed a tendency for both the types, except for the last beta strand at position 75-78, where the tendency was slightly higher for a parallel beta strand. The relative solvent accessibility of the amino acid residues in the ET1 mature protein sequence was calculated with two different methods, which predicted the buried and exposed residues in the sequence (Rost and Sander, 1994b; Janin, 1979). Based on these analyses, this ET1 protein secondary structure showed homology to that of TFIIS in the zinc ribbon domain region. In addition, from position 30 to 70 in the zinc ribbon like domain of ET1, a large number of buried residues could be predicted throughout with high certainty. However, in the region between positions 35 to 55 and between 65 to 70, the average percent accessibility was high, and correspondingly, the percent buried probability was low, indicating that a number of exposed residues could be present here. Outside the zinc ribbon like domain, a majority of the residues showed high percent accessibility and a low percent buried probability. Based on the first prediction analysis (Rost and Sander, 1994b), ET1 was also categorised as a compact globular protein.

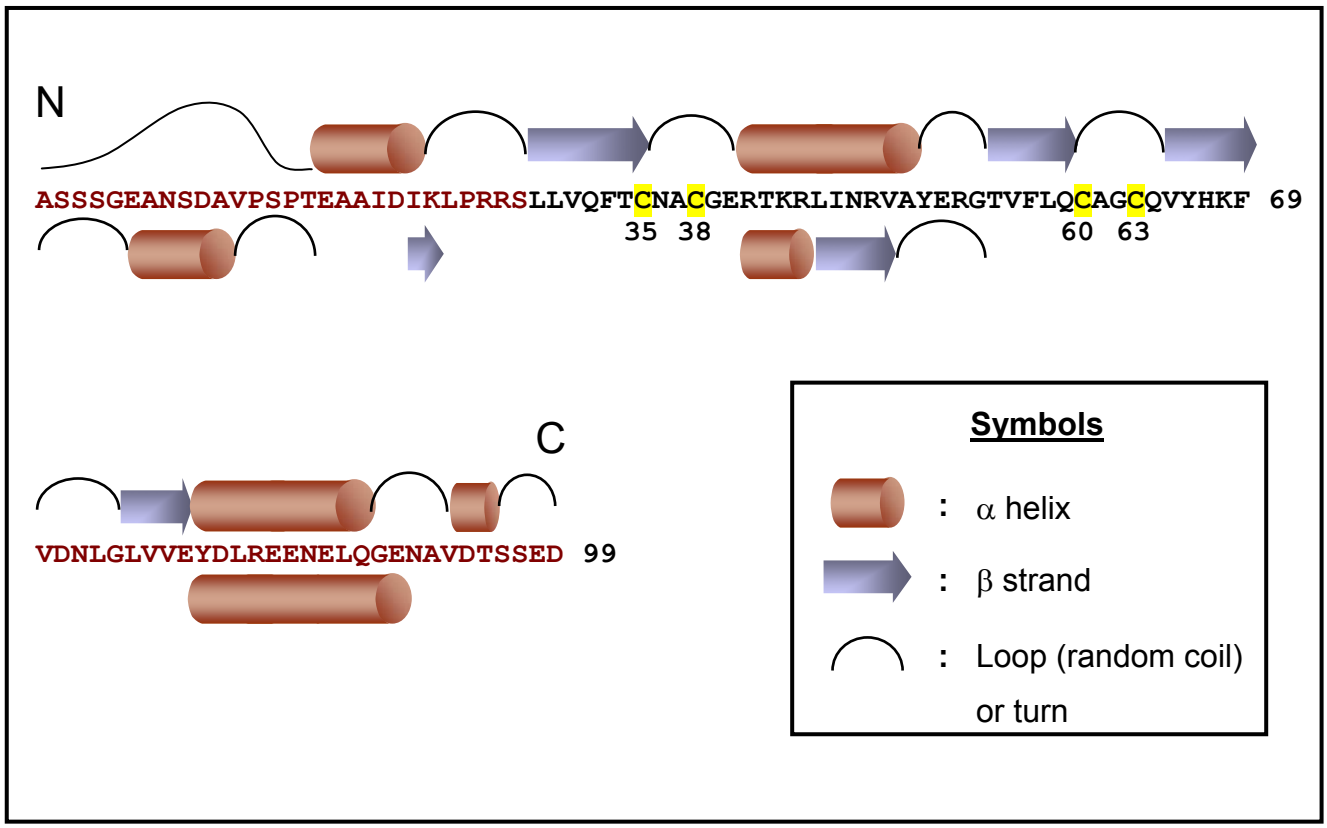



**Fig 3.12: Prediction of ET1 mature protein secondary structure.** The secondary structure of the putative mature ET1 protein, remaining after cleavage of the putative N-terminal transit sequence, was predicted using fourteen different programs available at the Expasy server (see text for details). The top row represents the secondary structure, corresponding to the protein sequence, predicted by a majority of the programs. In the protein sequence below, the region containing a putative zinc ribbon like domain is written in black, whereas the rest is in brown. The four conserved cysteine residues at positions 35, 38, 60 and 63 from the sequence start, which correspond to the four ligands of zinc ion in a zinc ribbon domain, are shaded yellow. A comparison of the secondary structure predictions showed that, in some cases, a high probability for a second alternative was present at some locations in the sequence, depicted below the line. Apart from α-helices and β-strands, the rest of the regions were categorised as either loop or turns.

Based on further analyses, no coiled-coil structures (Lupas et al., 1996) or helical transmembrane regions were detected in the sequence (Krogh et al., 2001; Rost et al., 1996; Hoffman et al., 1993). A weak helix-turn-helix DNA binding motif, 22 residues long, at position 13 from the start of the mature ET1 protein (Fig. 3.12), was detected, but predicted to be “not significant” (Dodd et al., 1990). Moreover, the

presence of a helix turn helix motif in this region could not be supported by any of the fourteen secondary structure prediction methods used for this analysis (Fig. 3.12).

## 3.4. Isolation of an *etched1* homologue, *zmzr1*

Gene duplication is a frequent phenomenon observed in maize. A large number of genes in maize, which are otherwise known to be single copy genes, have been found to maintain a second copy in the maize genome. The indication that there might be a second copy of e*t1* came from the Southern experiments carried out on LC genomic DNA (Fig. 3.3) as well as on *et1-Ref* (Ahrend, 1998), where a number of bands were detected with the *et1* cDNA probe that did not correspond to *et1*.

In order to examine if a second functional copy of *et1* existed, and if so, to study its effect on the *et1* gene, it was necessary to analyse its structure at the genomic as well as at the functional level. *et1-Ref* was one such line with an *et1* mutation, showing a deletion of the *et1* locus (Ahrend, 1998), which was found appropriate for this analysis.

### 3.4.1. Analysis of *et1-Ref* genomic clones

The *et1-Ref* genomic library had been screened by M. Ahrend (Diploma thesis, 1998) using the 2.5 kb *et1* genomic fragment probe, which lead to the identification of the *et1* locus. However, two different classes of clones were isolated from the library. The Class I contained fifteen clones, which on analysis were found to represent the *et1* locus. The complete sequence of the *et1* gene was found to be missing from this locus. The sequence of the *et1* probe hybridising to the *et1-Ref* genomic clones represented the gene upstream region.

In the second class, three clones were present, which were different from the Class I clones in their restriction pattern as well as pattern of hybridisation with the 2.5 kb *et1* probe. In addition, unlike the Class I clones, they also showed a hybridisation signal with the *et1* cDNA, *c9.1* (Ahrend, 1998). This information indicated that they belonged to another locus. One probable explanation for the hybridisation with *c9.1* was a deletion and translocation of the *et1* gene, complete or in part, to another

position during the X-ray irradiation event. On the other hand, the presence of a second copy or a gene homologue, which is often the case in *Zea mays*, was also possible. The latter possibility was also supported by the Southern analysis carried out on LC genomic DNA, where additional bands, not belonging to *et1*, were detected. Hence, these three genomic clones were further sequenced and analysed.

In order to characterise these three clones, viz., *et1-R 1.1, et1-R 10.2,* and *et1-R 18.1*, they were first analysed by new hybridisation experiments with the *et1* cDNA and the 2.5 kb *et1* genomic fragment probes. Using restriction fragment and hybridisation analyses, a restriction map of the clones was constructed. Based on their restriction maps, two clones, *et1-R 10.2* and *et1-R 18.1*, were found to overlap each other (Fig. 3.13, next page) and the third clone, *et1-R 1.1*, did not to belong together with the other two. Moreover, all three clones showed a restriction pattern different from that of *et1* from the wild type alleles. This indicated the presence of possibly two different loci in addition to the deleted *et1* locus. This, however, also supported the hypothesis that apart from the second copy of *et1*, a translocated *et1* gene fragment might also be present at another position in the *et1-Ref* genome. In order to clarify their identities, these clones were analysed further.

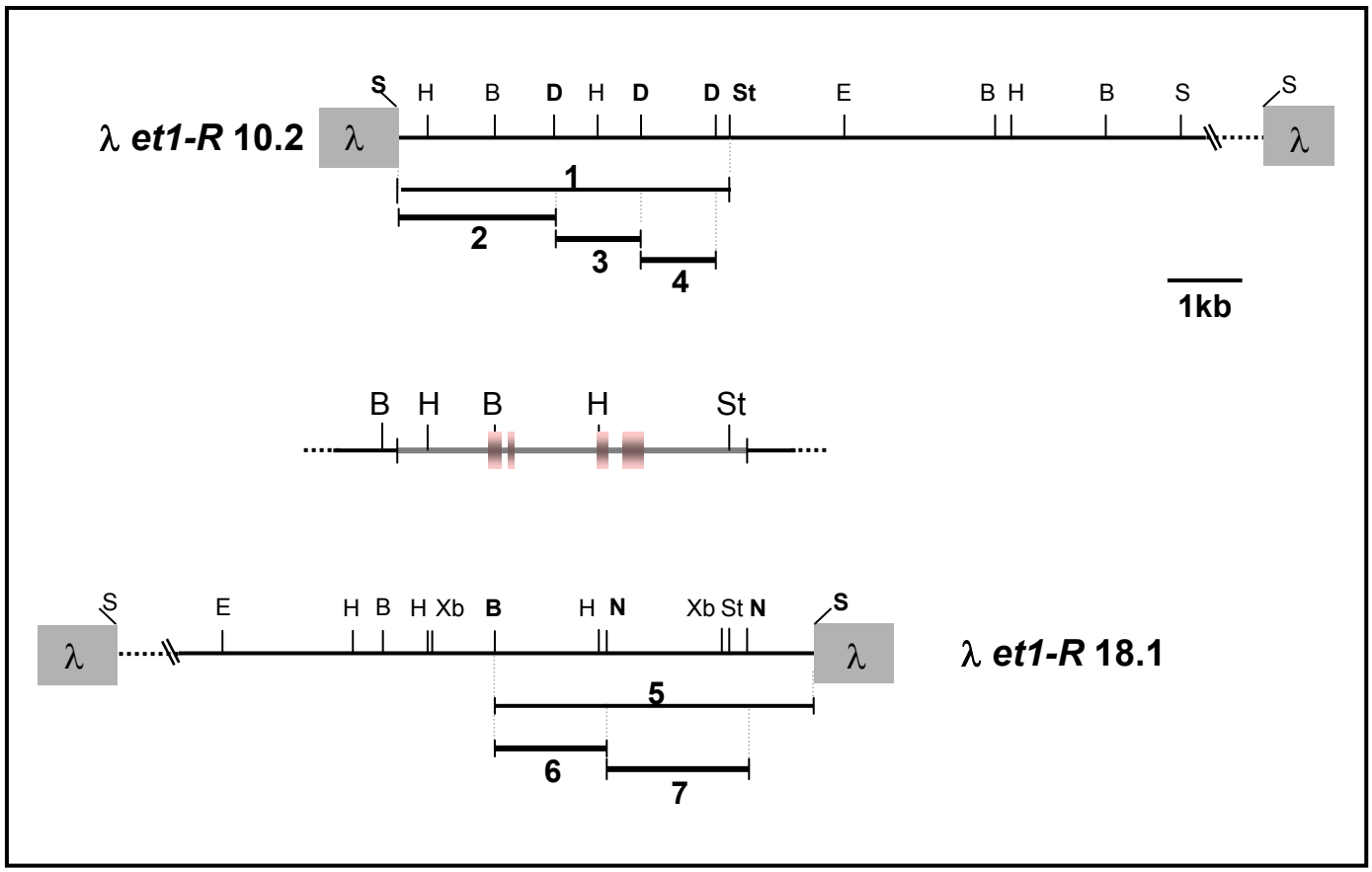



**Fig. 3.13: The restriction pattern of the Class II genomic λ clones, *et1-R 10.2* and *et1-R 18.1*, obtained from an *et1-Ref* line and representing a new gene, *zmzr1*.** The λ clones are shown in the 5´–> 3´ orientation of the gene. The λ arms are represented as grey boxes at the ends of the clones. From top to bottom, the first clone is λ *et1-R 10.2*, followed by a restriction map of the *zmzr1* gene obtained with the help of both the clones, and finally the clone λ *et1-R 18.1*. The subclones prepared from these two genomic clones are indicated below each of their restriction maps. After cloning fragment 1 from λ *etR 10.2*, the fragments 2, 3 and 4 were further subcloned from it, which were then sequenced completely. Similarly, after subcloning fragment 5 from λ *etR 18.1*, the fragments 6 and 7 were again subcloned from it and sequenced. The common restriction pattern in both the clones is represented in the middle of the two λ clones, showing the structure of the new gene *zmzr1*. The coloured blocks represent the four exons in the gene. The bold line represents the region sequenced. The scale bar on the right represents a length of 1kilobase. Abbreviations: λ, the phage λ arms; restriction enzymes: B *Bam H*I, D *Dra*II, E *Eco R*I, H *Hind*III, N *Nsi*I, S *Sal*I, St *Sst*I, X *Xho*I, Xb *Xba*I.

For sequencing the regions of hybridisation, first, some subfragments of the three clones were chosen for subcloning experiments (Table 3.2). These subfragments, as shown in Fig.3.13, were sequenced and evaluated for homology to *et1*.

| S.No. | Clone (fragment) | Length (Restriction Site) | Vector (Restriction Site) |
|---|---|---|---|
| 1. | *et1-R 1.1* (1) | ~ 2kb (*Sal*I) | pZErO-II (*Xho*I) |
| 2. | *et1-R 1.1* (2) | ~ 1.3 kb (*Sal*I/ *Pst*I) | pZErO-II (*Xho*I/ *Pst*I) |
| 3. | *et1-R 1.1* (3) | ~ 0.7 kb (*Sal*I / *Pst*I) | pZErO-II (*Xho*I/ *Pst*I) |
| 4. | *et1-R 10.2* | ~ 5.5 kb (*Sst*I/ *Sal*I) | pZErO-II (*Sst*I/ *Xho*I) |
| 5. | *et1-R 10.2* (1) | ~ 4.5 kb (*Sal*I/*Sst*I) | pZErO-II (*Xho*I/ *Sst*I) |
| 6. | *et1-R 10.2* (2) | ~ 2 kb (*Sal*I/ *Dra*II) | pZErO-II (*Xho*I/ *Dra*II) |
| 7. | *et1-R 10.2* (3) | ~ 1.2 kb (*Dra*II) | pZErO-II (*Dra*II) |
| 8. | *et1-R 10.2* (4) | ~ 1 kb (*Dra*II) | pZErO-II (*Dra*II) |
| 9. | *et1-R 18.1* (5) | ~ 4 kb (*Sal*I/ *Bam H*I) | pZErO-II (*Sal*I/ *Bam H*I) |
| 10. | *et1-R 18.1* (6) | ~ 2 kb (*Nsi*I) | pZErO-II (*Nsi*I) |
| 11. | *et1-R 18.1* (7) | ~ 1.5 kb (*Nsi*I) | pZErO-II (*Nsi*I) |

**Table3.2: Subcloning of the three *et1-Ref* λ genomic clones into the plasmid vector, pZErO-II for sequencing.** The first column represents the λ clones from which the fragments were subcloned. If they were given a number (Fig. 3.13), these are shown in parentheses. The second column shows the approximate size of the fragments and in parentheses their restriction digested ends. The third column shows the cloning vector and in parentheses the site of cloning.

Sequencing of the *et1-Ref* subclones revealed, as also indicated by their restriction maps, that the two genomic clones, *et1-R 10.2* and *et1-R 18.1*, contained overlapping regions (Fig. 3.13) which were completely homologous. Therefore, it indicated that these two clones belonged to one locus, while the genomic λ clone *et1-R 1.1* probably belonged to another locus. These two classes of clones were then analysed separately.

Sequence analysis of the *et1-R 1.1* λ clone revealed the presence of only a part of the *et1* gene. However, the λ clone also contained a *Mu8* element upstream of the start codon at the same position as the *et1-m3* mutant allele. A comparison of the 2kb *Sal I* sequence of the *et1-R 1.1* λ clone to that of *et1-m3*/+(B73) wild type alleles (da Costa e Silva et al., 2001) revealed that on leaving out the *Mu8* sequence from

*et1-R 1.1*, the sequence was completely identical to the wild type 2.5 kb *et1* sequence. The restriction map of one of the *et1-m3* mutant clones, λ 5.1 (da Costa e Silva et al., 2001), was also found to be similar to that of *et1-R 1.1* (Appendices A and C). In addition, on carrying out sequence analysis, the sequences of both their λ clone ends were found to be the same (data not shown).

In an attempt to isolate a λ clone belonging to λ *et1-R 1.1* locus, containing a complete *et1* sequence, an *et1-Ref* genomic library was screened using *c9.1* and *Mu8*-specific probes. However, no clones hybridising with both the probes could be isolated. The clones isolated with *c9.1* were found to belong to the same class as the other two *et1-Ref* clones, *et1-R 10.2* and *et1-R 18.1*.

Thus, in order to clarify the origin of λ *et1-R 1.1*, a *Mu*-primer based AIMS experiment was performed using genomic DNA from LC as a negative control and that from *et1-m3/et1-Ref* and *et1-m15/et1-Ref* mutants as positive controls. The homozygous *et1-Ref* and *et1-Ref/+*(B73) genomic DNA were taken as test samples. The AIMS blots were hybridised with the LC *et1* cDNA, *c9.1*, as well as the 2.5 kb *et1* fragment probes. No *Mu* associated amplified bands in *et1-Ref* samples hybridised to either of the probes (data not shown), whereby, if the λ *et1-R 1.1* clone had originated from *et1-Ref* genome, both the probes should have shown hybridisation signals with *et1-Ref* samples. Only the positive controls, *et1-m3/et1-Ref* and *et1-m15/et1-Ref*, showed two bands each of the expected size.

Based on these analyses, the *et1-R 1.1* clone was no longer regarded as an *et1-Ref* clone, but a contaminant, which might be coming from the *et1-m3* genomic library. Thus, the λ clones *et1-R 10.2* and *et1-R 18.1* were classified as Class II *et1-Ref* clones and will be referred to as such from here onwards.

The sequence analysis of these Class II *et1-Ref* clones, i.e., *et1-R 10.2* and *et1-R 18.1*, and their comparison to the available sequence data from the wild type *et1* genomic clones and the *et1* cDNA clones showed the presence of an ORF and a putative gene with high homology to the *et1* gene (Fig. 3.14). The homology was however, mainly restricted to the coding region. Just like *et1*, four exons and three introns could be identified from the sequence. Although the average value of homology was only 60% in this region, the exons were much more homologous than

the introns. Comparison of the exons of both the genes showed 78 to 79 % homology. The second intron in this gene was found to be much larger than that of *et1*, so that about 45% of *et1* intron sequence in this region was gapped. Although the remaining sequences were largely homologous, a number of small differences were observed between the two sequences throughout the coding region.

The upstream proximal region of the gene showed almost no homology to that of *et1* from either LC or B73. Similarly, only a little homology could be observed between the two downstream flanking sequences of the genes, which decreased further on moving farther downstream.

Therefore, comparison of the restriction patterns and the sequence analyses of the two genes indicated that they were two different genes showing high homology to each other. The gene represented by the Class II clones was named as *zmzr1* (*Zea mays zinc ribbon 1*) and will be referred to as such from here onwards. Whether the gene was functional still remained to be studied.

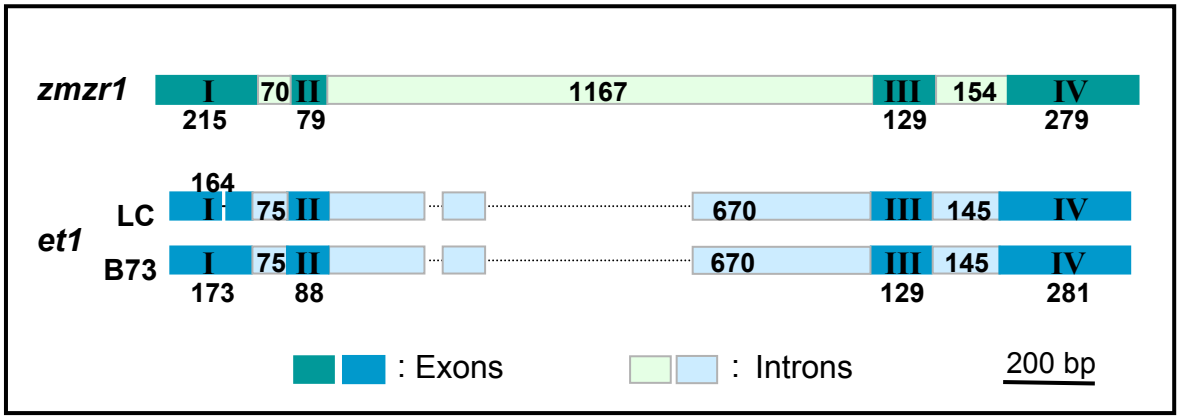



**Fig. 3.14: The genomic structure of the new gene, *zmzr1*, obtained through sequencing of Class II genomic λ clones *et1-R 10.2* and *et1-R 18.1* from *et1-Ref* line, and its comparison to *et1*.** The *zmzr1* gene is aligned with the *et1* gene from LC and B73 lines. The genes are represented in the 5´–> 3´ orientation. The exons are represented as darker boxes and intron boxes with lighter colours. Their lengths in base pairs (bp) have been shown either below and above the exons or within the intron boxes. The exons have been labelled using roman numerals within the boxes.

### 3.4.2. Expression analysis of the *zmzr1* gene

Based on the genomic sequence of *zmzr1*, the coding regions were found to be so homologous to *et1* that it could code for a protein that might be performing a similar function as ET1. Although, since they showed completely different promoter regions, it was necessary to check if the *zmzr1* gene was expressed and, if so, examine its expression pattern. Another reason for the analysis of *et1-Ref* line for *zmzr1* expression was that the *et1-Ref* phenotype was less severe as compared to that of some other *Mu* induced *et1-m* alleles (Fig. 1.3).

In order to carry out the expression analysis, a Northern experiment was performed with *et1-Ref* as well as with LC total RNA from different tissues of maize plants as described in section 3.2.5. Since *zmzr1* and *et1* are highly homologous, especially in the region containing the ORF, a probe had to be found which would hybridise specifically to the *zmzr1* transcript and thus, differentiate it from the *et1* transcript. Therefore, the putative 3´ untranslated region of the *zmzr1* gene was cloned through PCR amplification. This approx. 300 bp long 3´UTR fragment was then used as a *zmzr1* specific probe. However, despite carrying out two different trials, no *zmzr1* mRNA could be detected in the Northern experiments. However, it could still not be concluded that *zmzr1* expression does not take place in the plant samples analysed.

Therefore, in another experiment developing kernels were tested for the presence of the *zmzr1* transcript. The *et1* cDNA clone had also been isolated from developing kernels in the wild type line, LC. The LC cDNA library had already been screened twice with *et1* specific probes, once with the 2.5 kb *et1* genomic fragment containing a part of the ORF and once with the *et1* cDNA probe. However, all of the clones isolated were found to be *et1* cDNAs.

For the *zmzr1* transcript analysis in *et1-Ref* line, a λ ZAP developing kernel (15, 18 and 22 DAP) cDNA library was, therefore, prepared and approximately $1.3 \times 10^6$ pfu of the primary library were plated and screened with the *et1* cDNA and the putative *zmzr1* 3´UTR as probes. This screening also did not give any positive clones with either of the probes. One reason for this could be that the *et1* cDNA was not homologous enough to hybridise under those conditions with the *zmzr1* cDNA or that the 3´UTR probe was too small making it inappropriate for the Southern hybridisation experiments. With the 3´UTR probe similar problems were encountered in Southern

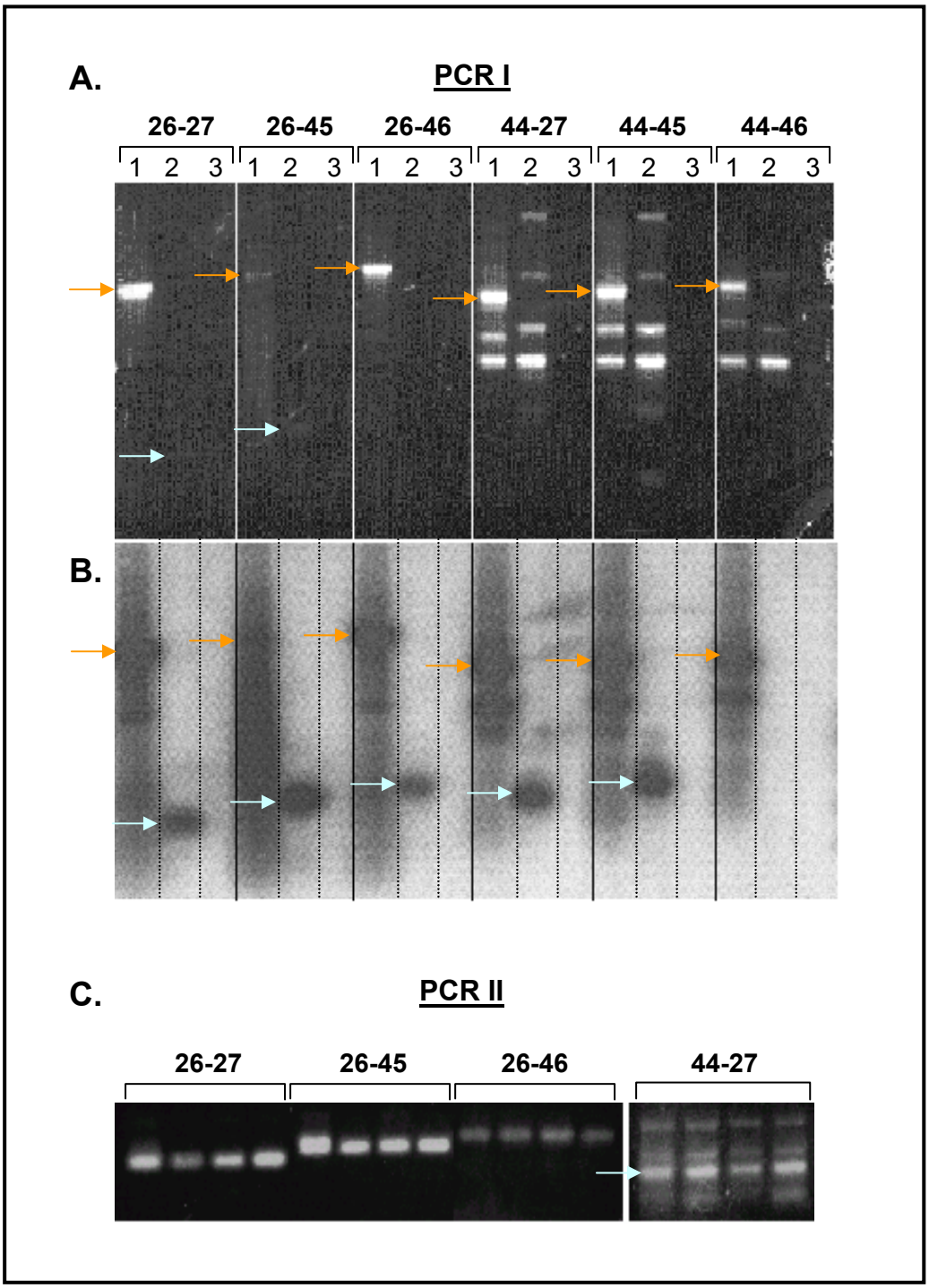
hybridisation experiments with LC and *et1-Ref* genomic DNA, where only a very weak signal was detected. Another possibility, also based on the Northern analysis, was weak expression of the *zmzr1* gene.

Therefore, based on the possibility that the *zmzr1* cDNA clone, if present, had somehow been missed in the cDNA library screening, another simpler but sensitive experiment was tried. Using *zmzr1* specific primers present in the 5´ and 3´ untranslated regions, with a melting temperature ranging between 65°C-70°C, and vector specific primers, the cDNA library was screened through PCR amplification experiments (Fig. 3.15). To select the *zmzr1* specific primers, the *zmzr1* transcript was conceptually translated from the *zmzr1* gene by comparison to the *et1* cDNA sequence. From the 5´ end of the deduced cDNA sequence, two gene specific primers were taken, one, Et 44, from the 5´ UTR and another, Et 26, at the start of the putative exon 2. From the 3´ end, three different primers, at three different positions in exon 4, viz., Et 27, Et 45 and Et 46, were taken (Table. 3.3). These experiments were carried out with both *et1-Ref* as well as LC cDNA libraries simultaneously.

After the first amplification of the samples from the *et1-Ref* developing kernel cDNA library, a number of strong bands could be observed on an ethidium bromide agarose gel, which, however, did not correspond in length to the bands expected from the *zmzr1* gene. The putative bands with lengths in the range expected from the *zmzr1* gene were either only faintly visible or, in many cases, were absent. Therefore, in order to check for the presence of the *zmzr1* specific cDNA clones, these amplifications were hybridised with the *et1* cDNA probe, containing sequences homologous to *zmzr1*. Only from the *et1-Ref* cDNA library were positive hybridisation signals in the expected size range detected. In the LC cDNA library no positive amplification products were detected.

In order to analyse the positive *et1-Ref* cDNA amplifications further, it was necessary to clone these sequences. For this, a larger amount of the PCR product, clearly visible on an ethidium bromide gel, was needed. Therefore, using small aliquots ranging from 0.01 µl to 5 µl from the first PCR, a reamplification (description in 2.2.8.4) of the cDNA clones was carried out using the same primers and the same PCR conditions. In the second round, the *zmzr1* specific amplifications were much

stronger in intensity than the other unspecific ones, a number of which were no longer visible in this round (Fig. 3.15).




→ : PCR amplifications with a *zmzr1* genomic clone used as a positive control for the reaction.

→ : PCR amplifications with an *et1-R* developing kernel cDNA library specific to *zmzr1*.

**Fig. 3.15:** (Legend on the next page)

**Fig. 3.15: The PCR amplification of *zmzr1* specific cDNA obtained from an *et1-Ref* developing kernel cDNA library.** In first round PCR, positive amplifications could be obtained from the cDNA library (lane 2) with the primer combinations: 26-27, 26-45, 26-46, 44-27, 44-45. For each combination, one positive control with the *zmzr1* λ genomic clone as template (lane 1) and one negative control with $H_2O$ as template (lane 3) were taken. The amplification as observed on the gel containing 10% of the PCR is shown in (**A**). The positive amplifications after hybridisation with the *et1* cDNA probe are shown in (**B**). Four different dilutions from each of the five positive PCRs were taken to reamplify (PCR II) the PCR products with the same primer combinations as in PCR I. In PCR II, shown in (**C**), the combinations 26-27, 26-45, 26-46 and 44-27 could be reamplified. These were isolated and cloned for further analysis.

Similar trials were also carried out using vector specific primers homologous to T3 and T7 promoter sequences, present on either side of the cDNA insert. These primers were used in combination with the *zmzr1* specific primers, in order to amplify both the 5´ and 3´ ends and get a complete cDNA sequence. One PCR product containing the 5´end of the cDNA, amplified with the primer pairs T3 and Et 27, was cloned and sequenced, (Fig. 3.16, Table 3.3 and 3.17). However, no 3´-end amplifications could be obtained.

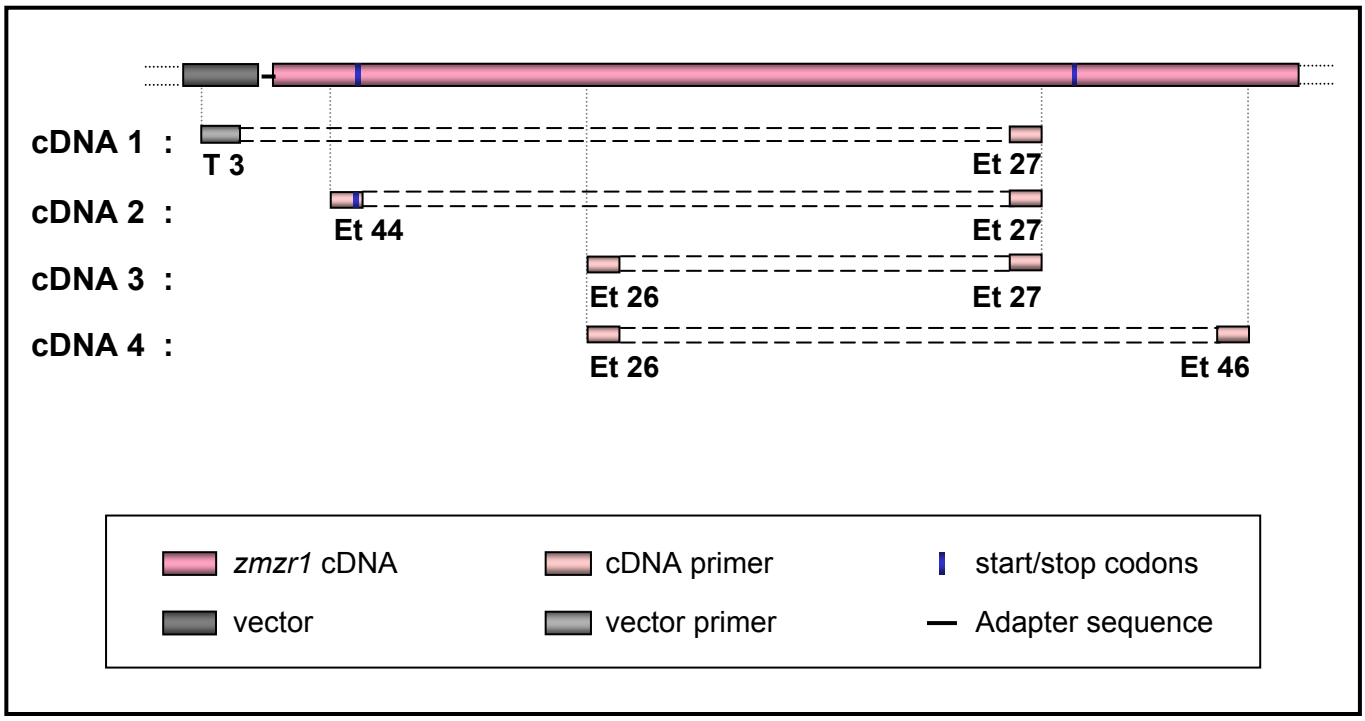



**Fig. 3.16:** (Legend on the next page)

**Fig. 3.16: A schematic representation of the amplified *zmzr1* cDNAs obtained from the *et1-Ref* developing kernel cDNA library**. The four amplified cDNA clones as described in Table 3.3 are represented schematically in the 5´–>3´ orientation. The topmost clone represents the putative cDNA present in the cDNA library, from which the four cDNA clones could be amplified using *zmzr1* specific primers. The names and positions of primers with respect to the cDNA are shown below each. The start and stop codons are shown as vertical blue bars.

| S.no. | cDNA clone | Primer Pair | Region of amplification | length of cDNA |
|---|---|---|---|---|
| 1. | cDNA1 | T3, Et 27 | 5´UTR (Exon 1) – Exon 4 | 433 bp |
| 2. | cDNA2 | Et 44, Et 27 | Exon 1 – Exon 4 | 480 bp |
| 3. | cDNA3 | Et 26, Et 27 | Exon 2 – Exon 4 | 297 bp |
| 4. | cDNA4 | Et 26, Et 46 | Exon 2 – 3´ UTR (Exon 4) | 439 bp |

**Table 3.3:** Four positive clones could be isolated from four independent amplification events of the *et1-Ref* dT primer developing kernel cDNA library. These were then cloned and, after sequencing, compared to the *zmzr1* genomic clones.

The sequence analysis of cDNA clones isolated from the *et1-Ref* developing kernel cDNA library revealed the presence of a *zmzr1* transcript. The cDNA contained the same open reading frame as that deduced from the comparison of the *zmzr1* gene to the *et1* cDNA. As can be seen in Fig. 3.17 (next page), the complete ORF could be amplified from the cDNA library. However, the length of the complete 3´ UTR could not be determined and the amplified 5´UTR of the cDNA could also possibly be incomplete. Moreover, from the PCR amplification of 5´UTR, obtained using the T3 and Et 27 primers, a truncated PCR product was obtained, where a small part of the ORF was absent. The amplified sequence was otherwise homologous to the other amplified cDNA sequences (Fig. 3.17, next page).

```
                    *        20         *        40         *        60
put.cDNA   : TGAGATCTCGCTCTTCCTTTATCCTCGCCTCCTCGGTCCAAGTAGGAGTCGGCGGCTATG
amplified  : TGAGATCTCGCTCTTCCTTTATCCTCGCCTCCTCGGTCCAAGTAGGAGTCGGCGGCTATG
cDNA1      : TGAGATCTCGCTCTTCCTTTATCCTCGCCTCCTCGGTCCAAGTAGGAGTCGGCGGCTATG
cDNA2      : ..........................................GAGTCGGCGGCTATG
cDNA3      : ............................................................
cDNA4      : ............................................................

                    *        80         *       100         *       120
put.cDNA   : GCGACGACGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTG
amplified  : GCGACGACGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTG
cDNA1      : GCGACGACGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTG
cDNA2      : GCGACGACGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTG
cDNA3      : ............................................................
cDNA4      : ............................................................

                    *       140         *       160         *       180
put.cDNA   : CCTGGGATCCTCTCGACCCGCCTGCGCCGGGAGCCTTCTCCTCCTCGCGTGGCCCTTGTG
amplified  : CCTGGGATCCTCTCGACCCGCCTGCGCCGGGAGCCTTCTCCTCCTCGCGTGGCCCTTGTG
cDNA1      : CCTGGGATCCTCTCGACCCGCCTGCGCCGGGAGCCTTCTCCTCCTCGCGTGGCCCTTGTG
cDNA2      : CCTGGGATCCTCTCGACCCGCCTGCGCCGGGAGCCTTCTCCTCCTCGCGTGGCCCTTGTG
cDNA3      : ............................................................
cDNA4      : ............................................................

                    *       200         *       220         *       240
put.cDNA   : GCCTCCTCCCCGAAGCTCAGGGCGCCGGCACCGCGGTTGAGAGTTTCGTGCGGTCCAAGG
amplified  : GCCTCCTCCCCGAAGCTCAGGGCGCCGGCACCGCGGTTGAGAGTTTCGTGCGGTCCAAGG
cDNA1      : GCCTCCTCCCCG................................................
cDNA2      : GCCTCCTCCCCGAAGCTCAGGGCGCCGGCACCGCGGTTGAGAGTTTCGTGCGGTCCAAGG
cDNA3      : ..............................................TGCGGTCCAAGG
cDNA4      : ..............................................TGCGGTCCAAGG

                    *       260         *       280         *       300
put.cDNA   : AGGCTCGCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGAAGCC
amplified  : AGGCTCGCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGAAGCC
cDNA1      : ..............................................TCGCCAACGGAAGCC
cDNA2      : AGGCTCGCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGAAGCC
cDNA3      : AGGCTCGCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCACCAACGGAAGCC
cDNA4      : AGGCTCGCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGAAGCC

                    *       320         *       340         *       360
put.cDNA   : ACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGT
amplified  : ACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGT
cDNA1      : ACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGT
cDNA2      : ACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGT
cDNA3      : ACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGT
cDNA4      : ACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGT

                    *       380         *       400         *       420
put.cDNA   : GGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAAAGAGGGACAATTTTTCTT
amplified  : GGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAAAGAGGGACAATTTTTCTT
cDNA1      : GGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAAAGAGGGGCAATTTTTCTT
cDNA2      : GGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAAAGAGGGACAATTTTTCTT
cDNA3      : GGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAAAGAGGGACAATTTTTCTT
cDNA4      : GGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAAAGAGGGACAATTTTTCTT
```

**Fig. 3.17:** (continued on the next page)

```
                    *       440         *       460         *       480
put.cDNA   : CAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAG
amplified  : CAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAG
cDNA1      : CAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAG
cDNA2      : CAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAG
cDNA3      : CAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAG
cDNA4      : CAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAG

                    *       500         *       520         *       540
put.cDNA   : TATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATTGATACTAATTCTGAA
amplified  : TATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATTGATACTAATTCTGAA
cDNA1      : TATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATT..............
cDNA2      : TATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATT..............
cDNA3      : TATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATT..............
cDNA4      : TATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATTGATACTAATTCTGAA

                    *       560         *       580         *       600
put.cDNA   : GATTGATCTGTTGTGAGAAGCAGTGTTGGTATGCAAAACATCCTGTTGTAGCCTGTTGTC
amplified  : GATTGATCTGTTGTGAGAAGCAGTGTTGGTATGCAAAACATCCTGTTGTAGCCTGTTGTC
cDNA1      : ............................................................
cDNA2      : ............................................................
cDNA3      : ............................................................
cDNA4      : GATTGATCTGTTGTGAGAAGCAGTGTTGGTATGCAAAACATCCTGTTGTAGCCTGTTGTC

                    *       620         *       640         *       660
put.cDNA   : TGAATGTCTGTAGGTTTTGACAATATTGATTATTTATATAAAATGACATTTTAGTTGCTG
amplified  : TGAATGTCTGTAGGTTTTGACAATATTGATTATTTATATAAAATGACATTTTAGTTGCTG
cDNA1      : ............................................................
cDNA2      : ............................................................
cDNA3      : ............................................................
cDNA4      : TGAATGTCTGTAGGTTTTGACAATATTGATTATTTATATAAAATGACATTTTAGTTGCTG

                    *       680         *       700
put.cDNA   : CATATACTCATCTGAATACATGGGGGCCTAGATGTATTTTAT
amplified  : CATATAC.................................
cDNA1      : ........................................
cDNA2      : ........................................
cDNA3      : ........................................
cDNA4      : CATATAC.................................
```

**Fig. 3.17: *zmzr1* cDNA clones amplified from the *et1-Ref* developing kernel cDNA library.** Four cDNA sequences could be amplified from the dT primer cDNA library. From top to bottom, the first sequence represents the putative cDNA sequence (`put.cDNA`) derived from the comparison of the *zmzr1* genomic sequence to *et1* cDNA sequence. The second sequence represents the amplified region of the *zmzr1* cDNA derived from the four *zmzr1* cDNA clones taken together. The last four sequences represent the four cDNA clones amplified from the *zmzr1* cDNA library (see Table 3.3, page 99, and Fig. 3.16, page 98). The start and stop codons as well as the exon junctions are shaded as pink and yellow on the sequences respectively. As can be seen, the *zmzr1* cDNA sequence is 667 bp in length and contains an open reading frame with four exons.

### 3.4.3. Characterisation of *zmzr1* in LC and its comparison to the *et1-Ref* allele

Additional support for *zmzr1* being a new gene, different from *et1*, was provided by the analysis of the *zmzr1* gene in the LC background. Southern analysis of the LC genomic DNA (Fig.3.3, page 64), performed using the *et1* specific probes *c9.1* and 2.5 kb *et1* fragment, had revealed the presence of additional sequences with homology to *et1*, which were represented in the form of weak bands in the Southern hybridisation experiment (Fig.3.3, blue arrows). On comparison with the restriction map of the Class II *et1-Ref* clones, these bands corresponded in size to that expected for *zmzr1*. This in itself indicated that this locus was not only present in the *et1-Ref* background but also in the LC genomic background, in addition to the *et1* locus.

In order to analyse *zmzr1* in more detail, its genomic structure was analysed in LC and then compared to that in *et1-Ref*. For this, Southern experiments were performed using a *zmzr1* specific promoter fragment and cDNA as probes on LC and *et1-Ref* genomic DNA.

First a *zmzr1* specific 1 kb promoter fragment, present directly upstream of the gene, was used as a probe on LC genomic DNA. The sizes of the bands obtained (Fig. 3.18) were compared to the respective restriction fragment lengths on the restriction map of *et1-Ref* genomic clones (Fig.3.13). Based on the comparison, a very similar genomic structure for *zmzr1* in both LC and *et1-Ref* could be predicted. However, this analysis of the *zmzr1* gene was restricted mainly to the gene upstream region and a part of the gene. The Southern analysis also indicated that this promoter region of the *zmzr1* gene was a single copy sequence specific to *zmzr1*, as no bands other than the ones expected from the *zmzr1* genomic clones were found to hybridise with the probe.

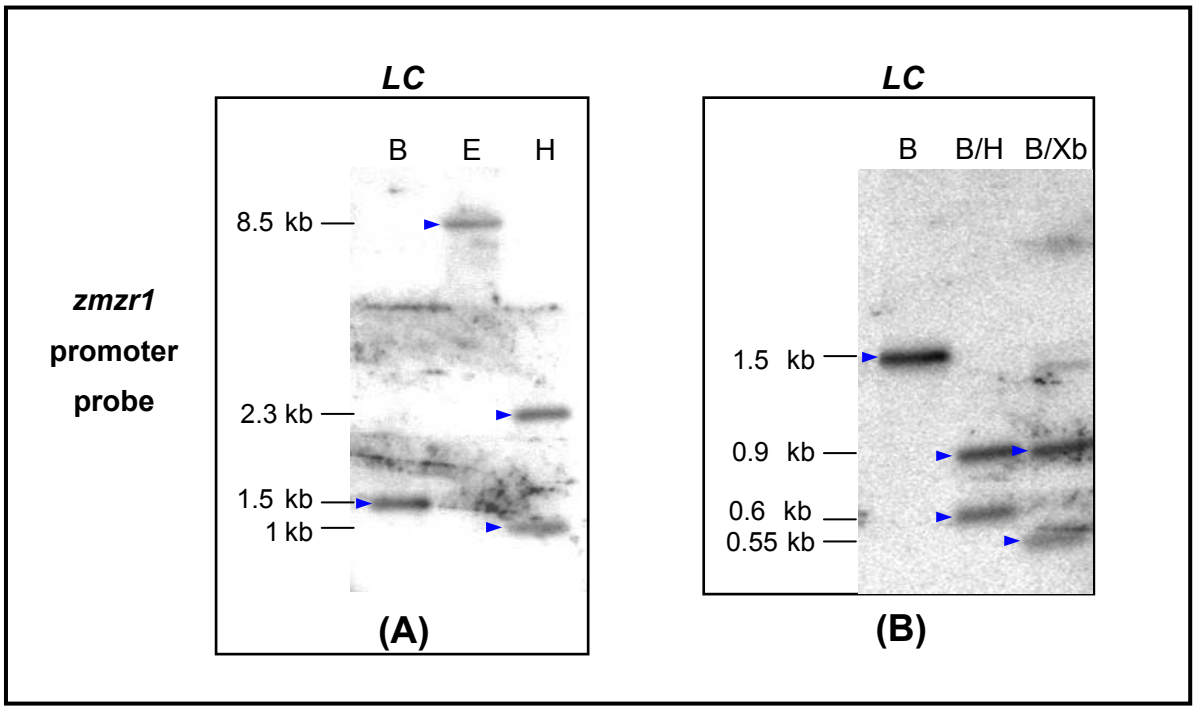



**Fig. 3.18: Southern analysis performed with LC genomic DNA using a *zmzr1* specific probe**. Both the Southern hybridisation experiments shown in (A) and (B) were carried out with LC genomic DNA, using a 1 kb *zmzr1* promoter fragment, present in the immediate vicinity of the gene, as a probe. The hybridising bands have been marked with an arrow. The experiment showed that this *zmzr1* sequence is also present in LC whose restirction pattern is the same as for the corresponding *zmzr1* gene in *et1-Ref*, and also that it is a single copy sequence. Ten µg genomic DNA was used for each sample (lane) restriction digest. Abbreviations: B *BamHI*, E *EcoRI*, H *Hind III*, B/H *BamHI/Hind III*, B/Xb *BamHI/Xba I*.

In order to analyse and compare the genomic structure of *zmzr1* in LC and *et1-Ref*, a further Southern experiment using the *zmzr1* cDNA probe was performed (Fig. 3.19). A set of bands corresponding to the restriction map of *zmzr1* from *et1-Ref* could be identified in both the lines (Fig. 3.19, arrows). Based on this analysis, the restriction pattern of *zmzr1* appeared to be very similar in both the lines. From LC another set of fainter bands, corresponding to the *et1* genomic structure, could also be identified (Fig. 3.19, asterisk).

Apart from these two sets of bands, a number of additional fainter bands were observed on both *et1-Ref* and LC Southern blots. This was also the case in the Southern experiment with the *c9.1* probe on LC genomic DNA (see section 3.2.3, Fig.

3.3). None of these fainter bands could be explained by either *et1* or *zmzr1* genomic structure.

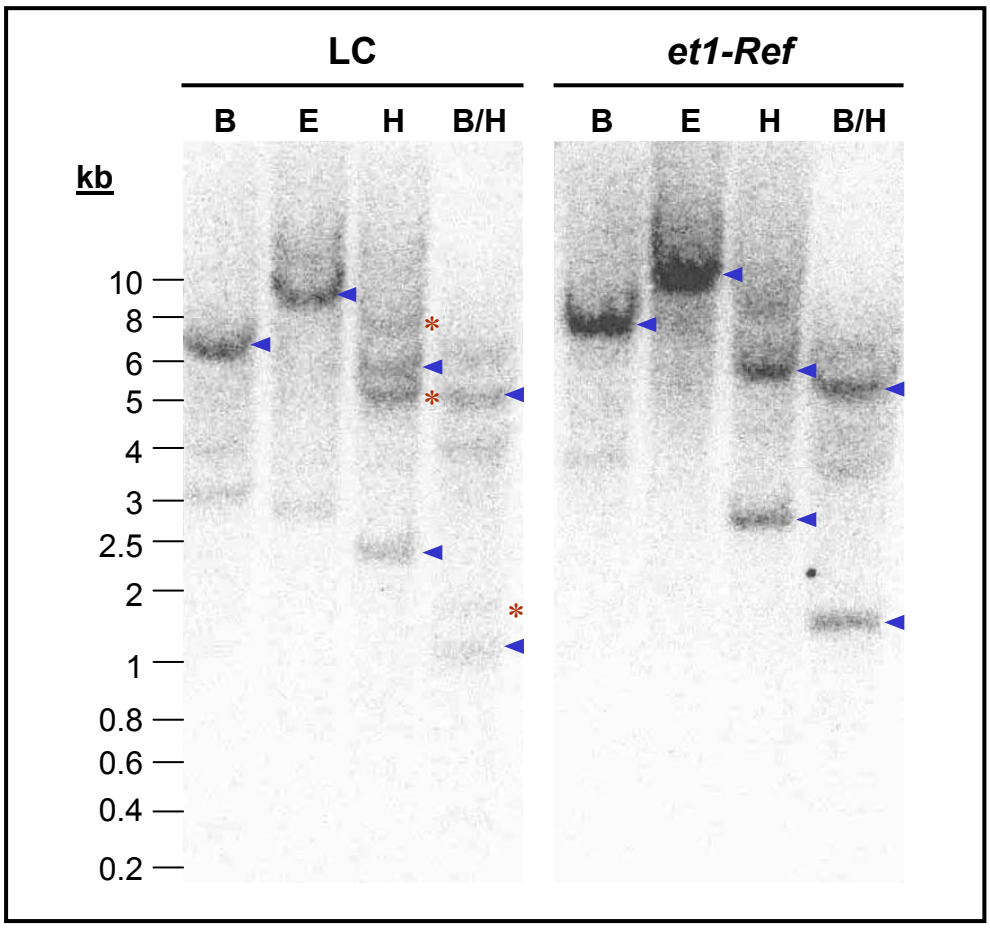



**Fig. 3.19: Southern experiment with LC and *et1-Ref* genomic DNA using *zmzr1* cDNA probe**. LC (left) and *et1-R*ef (right) genomic DNA were restriction digested each with: *BamH*I (B), *EcoR*I (E), *Hind*III (H), B/H (*BamH*I/*Hind*III). For each digest, ten µg genomic DNA was used and, after digestion, transferred to a nylon membrane. For this analysis the complete *zmzr1* cDNA clone, amplified from the *et1-Ref* cDNA library, was taken as a probe. The set of bands marked with arrows represent the *zmzr1* specific bands and the ones marked with asterisk represent the *et1* specific bands.

## 3.5. *In silico* characterisation of *et1* and *zmzr1* and their comparison

Comparisons of the genomic and protein sequences of both *et1* and *zmzr1* were carried out to find similarities and differences between the two and, thus, further characterise the two genes.

### 3.5.1. Sequence analysis of the promoter regions of *et1* and *zmzr1* and their comparison

Despite being highly homologous in their genomic structures, a comparison of the gene upstream regions of *et1* and *zmzr1* showed almost no homology between the two promoters (Appendix B3). Moreover, an examination of the expression pattern of the two genes through Northern experiments did not indicate any *zmzr1* expression in a wide variety of plant tissues examined, including the ones where *et1* expression was observed (see section 3.4.2). Although a very weak expression of the *zmzr1* transcript was observed in *et1-Ref* developing kernels through PCR amplifications from LC developing kernels no *zmzr1* transcript could be isolated or amplified. This gave an indication that its expression might be differently regulated than *et1* and may be expressed in response to different environmental factors. Therefore, it was necessary to examine the gene upstream regions of the two genes.

In order to characterise the gene upstream region of both the genes, these were compared with the promoter databank PlantCARE, available at the ExPASY server. For this analysis, the gene upstream sequences of both the LC and B73 alleles of *et1*, and the *et1-Ref* allele of *zmzr1* genes, starting upstream from the transcription start were used. Based on the comparisons, a number of putative cis regulatory elements could be identified from the upstream sequences of both the genes (Fig. 3.20; Fig. 3.21).

The analysis of the gene upstream regions of the two *et1* alleles from LC and B73 showed that a large number of elements were common to both and remained unaffected by the sequence differences between the two. However, the presence of a few regulatory elements in this region with differing sequences was detected in both the *et1* promoters. Comparison of the *et1* promoter region sequences showed that the *et1* and *zmzr1* genes largely possessed different gene upstream elements, apart from a few common regulatory elements. These common regulatory elements were, however, present at different positions in the two promoters, when measured from the transcription start. This was not unexpected because of the essentially non-homologous upstream regions of the two genes.

| (A) Light Responsive Elements | | | | | |
|---|---|---|---|---|---|
| S.no. | Name | Core motif | S.no. | Name | Core motif |
| 1. | **ACE** | ACGT | 7. | **GA motif** | AACC |
| 2. | **AE box** | AGAA | 8. | **GATA motif** | GATA |
| 3. | **ATC motif** | ATCC | 9. | **GT1 motif** | GGTTAA |
| 4. | **Box I** | TTTC | 10. | **I box** | GATAGGG |
| 5. | **Box II** | CCAC | 11. | **LAMP element** | AACC |
| 6. | **G box** | CACGTG | 12. | **TCCC motif** | TCTC |

| (B) Endosperm Specific Elements | | | | | |
|---|---|---|---|---|---|
| S.no. | Name | Core motif | S.no. | Name | Core motif |
| 1. | **A box (Os)** | TAAC | 3. | **Prolamin box** | AAAG |
| 2. | **GCN4 motif** | AGTCA | 4. | **Skn-1 motif** | GTCAT |

| (C) Other Elements | | | |
|---|---|---|---|
| S.no. | Name | Core motif | Role |
| 1. | **EIRE** | TTCGACC | Elicitor responsive element |
| 2. | **MeJA element** | TGACG | Methyl jasmonate responsiveness |
| 3. | **WUN motif** | CATT/AATT | Wound responsiveness |
| 4. | **LTR** | CCGAAA | Low temperature responsiveness |
| 5. | **A box (Pc)** | CCGTCC | Promoter element from parsley |

**Figure 3.20: Cis regulatory elements and their core motifs identified from the *et1 and zmzr1* promoter regions.** The elements based on their function have been grouped into three tables, designated A, B and C. Table A shows light responsive elements, table B endosperm specific elements and table C others, including elicitor responsive, MeJA responsive, wound responsive and low temperature responsive elements.

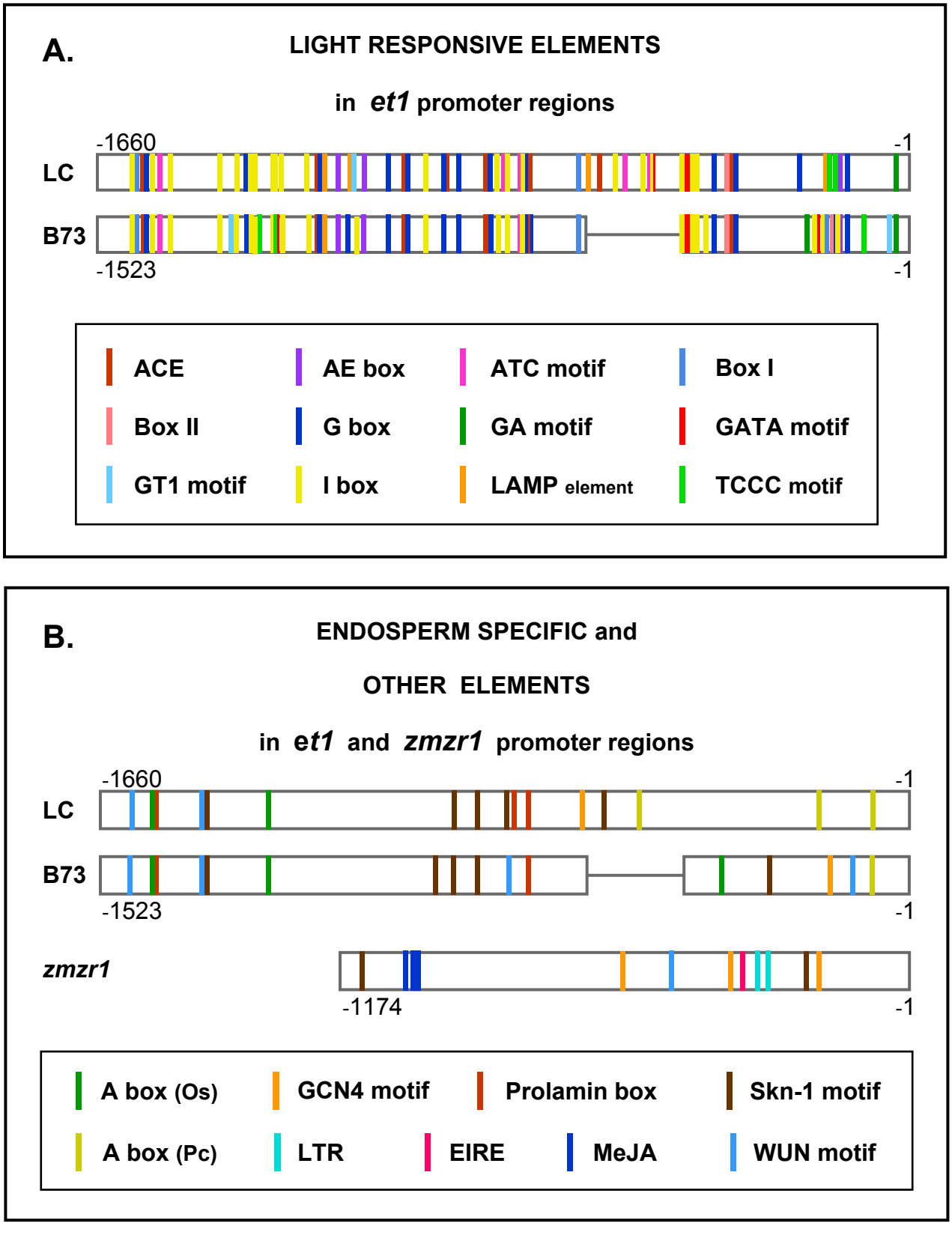



**Fig. 3.21: Promoter analysis of *et1* as well as *zmzr1* gene upstream regions.** The promoter regions of the two *et1* alleles from LC and B73, and from *zmzr1* were analysed for the presence of cis regulatory elements. The *et1* and *zmzr1* promoter regions have been placed together for comparison of their lengths and do not indicate homology. The LC and B73 *et1* alleles, however, are aligned

based on homology of the two sequences. The coloured vertical bars represent motifs showing homology to the core sequences of different cis regulatory elements. The lengths of the promoter regions are represented in basepairs starting from the putative transcription starts (from right to left) derived from their cDNA sequences. **(A)** The different light responsive elements identified from promoter regions of LC and B73 alleles. *zmzr1* promoter region is not depicted here. Although, a few light responsive elements were also found for it in the PlantCARE database. **(B)** The endosperm specific elements, viz., A box(Os) from *Oryza sativa*, GCN4, Prolamin box and Skn-1 motif ; and others like an unknown cis element, A box (Pc), from parsley; elicitor responsive (EIRE); methyl jasmonate responsive (MeJA); wound responsive (WUN motif) and low temperature responsive (LTR) elements were identified from the *et1* and *zmzr1* promoter regions (details in text and at: http://sphinx.rug.ac.be:8080/PlantCARE/cgi/Menu.html).

Only those motifs were selected from the promoter analysis results which showed 100% homology to the core sequence of the element. Apart from some light responsive and endosperm specific elements, some other putative elements with homology to the core sequences of other regulatory elements could also be identified from both the genes. The tables in Fig. 3.20 show the motifs of all the gene upstream putative cis regulatory elements identified upstream of *et1* in the LC and B73 alleles and the *zmzr1* gene in the *et1-Ref* allele. Some putative light responsive elements, which could be identified for *zmzr1* promoter region, are not depicted in Fig. 3.21.

### 3.5.2. Comparison of the *zmzr1* and *et1* cDNAs and their protein primary structures

A comparison of the *et1* cDNA sequences, including the LC cDNA as well as the deduced *et1* cDNA from the genomic B73 allele, to *zmzr1* cDNA was carried out using the GeneDoc program. For the comparison, putative 3´ UTR regions of *zmzr1* that could not be PCR amplified were also included. The sequence comparison showed a homology of about 87% between the *zmzr1* and *et1* sequences within the open reading frames (Fig. 3.22). The 5´ and the 3´ untranslated regions, however, showed a higher degree of dissimilarity between the two cDNA sequences. The 5´UTRs were about 61 % homologous, when the unamplified *et1* sequence upstream of the known transcription start was also taken into consideration. Whereas the 3´UTRs were only about 55% homologous between the two genes. In the open

reading frames of the two genes, all the nucleotide substitutions were transitions and a number of these substitutions were only synonymous codon substitutions observed at the wobble position, which did not affect their deduced protein sequences. Moreover, in a few cases, the differences between the triplet codons of the two transcripts only caused a minor physiochemical change, where an aliphatic amino acid was replaced with another aliphatic amino acid, and this way the amino acids remained conserved.

```
                    *        20         *        40         *        60
et1(LC)   : ----------------------------------------------GG-CAAAGCGGCTATG :  15
et1(B73)  : ----------------------------------------------GG-CAAAGCGGCTATG :  15
zmzr1     : TGAGATCTCGCTCTTCCTTTATCCTCGCCTCCTCGGTCCAAGTAGGAGTCGGCGGCTATG :  60

                    *        80         *       100         *       120
et1(LC)   : ACGACGACGGCCGCCGGGCACGGCTGCTGCTGGGCGGGGATTCCGCCCTTCGCGTTGTTG :  75
et1(B73)  : ACGACAACGGCCGCCGGGTACGGCTGCTGCTGGGCGGGGATTCCGCCCTTCGCGTTGTTG :  75
zmzr1     : GCGACGACGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTG : 120

                    *       140         *       160         *       180
et1(LC)   : CCGCGGATTCTCTCGACCGGCC------GGGAGACTCCTCCTCCTC---------GCGCT : 120
et1(B73)  : CCGCGGATTCTCTCGACCCGCC------GGGAGACTCCTCCTCCTCCTCCTCCTCGCGCT : 129
zmzr1     : CCTGGGATCCTCTCGACCCGCCTGCGCCGGGAGCCTTCTCCTCCTC---------GCGTG : 171

                    *       200         *       220         *       240
et1(LC)   : TCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCGCGGCTGAGAGTTTCGAAC : 180
et1(B73)  : TCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCACGGCTGAGAGTTTCGAAC : 189
zmzr1     : GCCCTTGTGGCCTCCTCCCCGAAGCTCAGGGCGCCGGCACCGCGGTTGAGAGTTTCGTGC : 231

                    *       260         *       280         *       300
et1(LC)   : CGTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGAGGCCAATTCGGACGCGGTG : 240
et1(B73)  : CGTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGAGGCCAATTCGGACGCGGCG : 249
zmzr1     : GGTCCAAGGAGGCTC---------GCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTG : 282

                    *       320         *       340         *       360
et1(LC)   : CCGTCGCCAACGGAAGCCGCTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAA : 300
et1(B73)  : CCGTCGCCAACGGAAGCCGCTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAA : 309
zmzr1     : CCGTCGCCAACGGAAGCCACTATTGATATAAAGCTTCCTAGAAGAAGTTTGCTTGTTCAA : 342

                    *       380         *       400         *       420
et1(LC)   : TTTACATGCAACGCATGTGGCGAAAGGACCAAGCGCTTGATAAACAGAGTAGCCTATGAA : 360
et1(B73)  : TTTACATGCAACGCATGTGGCGAAAGGACCAAGCGCTTGATAAACAGAGTAGCCTATGAA : 369
zmzr1     : TTTACATGCAATGCATGTGGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCCTATGAA : 402

                    *       440         *       460         *       480
et1(LC)   : AGAGGCACAGTTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAAT : 420
et1(B73)  : AGAGGCACAGTTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAAT : 429
zmzr1     : AGAGGGACAATTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAAT : 462
```

**Fig. 3.22:** (continued on the next page)

```
                    *       440         *       460         *       480
et1(LC)   : AGAGGCACAGTTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAAT : 420
et1(B73)  : AGAGGCACAGTTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAAT : 429
zmzr1     : AGAGGGACAATTTTTCTTCAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAAT : 462

                    *       500         *       520         *       540
et1(LC)   : CTTGGGCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCG : 480
et1(B73)  : CTTGGGCTAGTTGTTGAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCG : 489
zmzr1     : CTTGGTCTAGTTGTTGAGTATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTG : 522

                    *       560         *       580         *       600
et1(LC)   : GTGGACACCAGTTCTGAAGATTGATCTGCTGTGAGAAGCGATGTTGGTATGCAAAACGCC : 540
et1(B73)  : GTGGACACCAGTTCTGAAGATTGATCTGCTGTGAGAAGCGGTGTTGGTATGCAAAACGCC : 549
zmzr1     : ATTGATACTAATTCTGAAGATTGATCTGTTGTGAGAAGCAGTGTTGGTATGCAAAACATC : 582

                    *       620         *       640         *       660
et1(LC)   : CTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCATAAAAATGGCACTTTT : 600
et1(B73)  : CTGTACTCTGTAGGTTTTTGACAACATTGGTTATTTGTATAGCATAAAAATGGCACTTTT : 609
zmzr1     : CTGTTGTAGCCTGTTGTCTGA-ATGTCTGTAGGTTTTGACAATATTGATTATTTATATAA : 641

                    *       680         *       700         *       720
et1(LC)   : TAAAGTTGTTGCACATACTCATCTGAAATTCTGAATACAGCAGGGCCTACATTGTACTTC : 660
et1(B73)  : TAAAGTTGTTGCACATACTCATCTGAAATTCTGAATACAGCAGGGCCTACATTGTACTTC : 669
zmzr1     : AATGACATTTTAGTTGCTGCATATACtcatctgaatacatgggggcctagat-gtatttt : 700


et1(LC)   : TT : 662
et1(B73)  : TT : 671
zmzr1     : at : 702
```

**Fig. 3.22: Comparison of the two *et1* cDNAs with the *zmzr1* cDNA from *et1-Ref*.** The *et1* and *zmzr1* cDNAs were amplified from developing kernel cDNA libraries of LC and *et1-Ref* lines respectively. An *et1* cDNA deduced from the genomic sequence of the B73 allele is also depicted. As can be seen, the cDNAs show high homology among each other. The *zmzr1* cDNA differs from the *et1* cDNAs at a number of positions in the ORF as well as in the 5´and 3´ UTR regions. The non-homologous bases in the alignment are written in blue with a grey background, the exon junctions have been shaded yellow and the start and stop codons pink. The putative 3´UTR of *zmzr1*, derived from the genomic sequence, but not present in the amplified cDNA clone, is depicted in lower case.

In order to check the functional similarity between *zmzr1* and *et1*, the ZMZR1 protein sequence was examined for the presence of a transit sequence for plastid targeting with the same programs used for ET1. For the ZMZR1 protein, with a length of 162 amino acids, the certainty in the TargetP prediction results was also the highest for plastid localisation. The highest score was 0.743 for the presence of a plastid transit sequence. Although lower than ET1, it was also in the reliability class 2, which is the

second highest reliability class from five. The predicted length of the N-terminal transit sequence was 65 amino acids. The length of the mature ZMZR1 protein would then be 97 amino acids. The Predotar analysis revealed a high probability of 96% for the protein being plastid localised as compared to mitochondria, for which 0% probability was determined.

A comparison of the ET1 and ZMZR1 protein sequences, deduced from their cDNA clones, showed the amount of homology among the proteins and made their structural similarities and differences clearer (Fig. 3.23). About 82-83% of the two proteins were completely identical. However, when amino acids with similar properties were taken into consideration, the similarity increased to 86-87%. Sequence gaps between the two proteins constituted 3% of the total length. Therefore, only approx. 10 % of the two proteins were non-homologous.

However, major differences between the two proteins were mainly restricited to the first 85 amino acid residues, of which the first 65 residues showing differences were present in the transit sequence region. Some minor differences were also observed in the last fifteen C-terminal residues (Fig. 3.23). The region between the residues at position 90 and 150, representing the zinc ribbon like domain, was found to be completely identical except at one position, where the valine residue of ET1 was replaced with isoleucine (at position 124 in Fig. 3.23). These residues, despite being different, were physiochemically conserved, as they possessed the same physiochemical properties. The few differences did not necessarily cause any change in the functional property of the amino acid at that position.

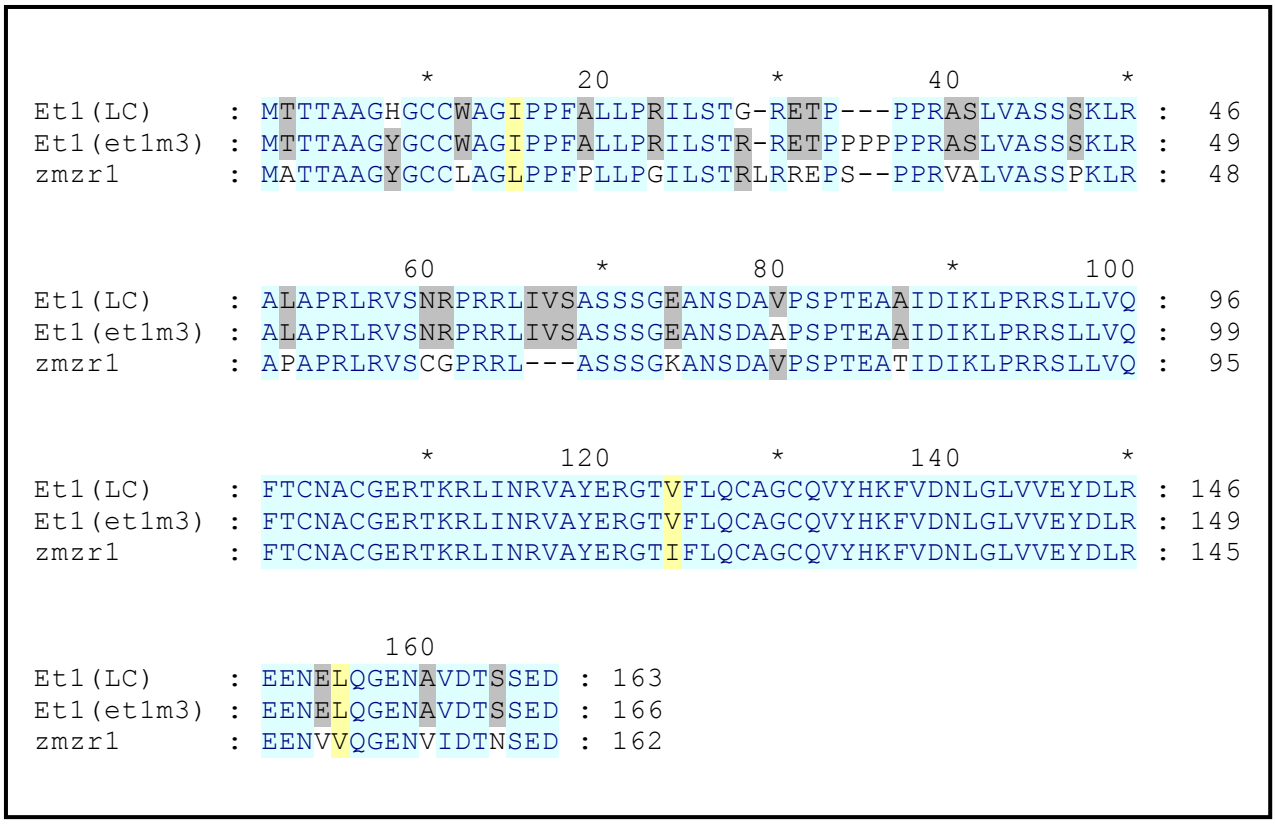



**Fig. 3.23: Comparison of the ET1 protein sequences from LC and B73(et1-m3) with ZMZR1 from *et1-Ref*.** The protein sequences were deduced from their respective cDNA sequences. All the amino acid differences can be observed here. The amino acids shaded in yellow are the ones having the same conserved physiochemical property. Other differences, representing different physiochemical properties or gaps, have been shaded grey.

### 3.5.3. Homology of ET1 and ZMZR1 to other translated plant ESTs

The deduced protein sequences of the *et1* cDNAs from LC and B73, and that of *zmzr1* cDNA were used as test proteins to search the online databases using tblastn at the BLAST server of NCBI. The data obtained was aligned and compared to the test sequences with the help of Clustal X, Clustal W and Genedoc programs. A number of ESTs as well as genomic sequence data from plant and non-plant sources were identified, whose conceptually translated proteins were aligned with the two ET1 and the ZMZR1 test proteins (Fig.3.24). Except for two proteins derived from *Arabidopsis thaliana* genomic sequences, all the other proteins were translated ESTs, obtained from *Arabidopsis thaliana* (thale cress), *Glycine max* (soybean), *Gossypium sp.* (cotton), *Hordeum vulgare* (barley), *Lycopersicon esculentum* (tomato), *Medicago truncatula*, *Mesembryanthemum crystallinum* (ice plant), *Oryza*

*sativa* (rice), *Solanum tuberosum* (potato), *Triticum aestivum* (wheat) and *Zea mays* (maize). In *Arabidopsis*, the proteins derived from incomplete ESTs were found to be identical with the one derived through conceptual translation of one of the genomic sequences. Therefore, only the complete conceptually translated protein sequence is depicted here.

The putative proteins depicted in Fig. 3.24 showed high homology among each other. In some cases different amino acids were present at one position, which, however, belonged to a common physiochemical group. This made the property of the residue at these positions unchanged and, thus conserved. These homologies were especially high near the C-terminal end of these proteins, where, in a length of approx. 60 amino acids, a number of conserved amino acid residues were observed. This was also the region showing homology to the zinc ribbon domain of TFIIS in ET1. Therefore, all the sequences were highly conserved in this zinc ribbon like domain. The ice plant protein, probably due to bad quality of the EST sequence, was only partly homologous to the conserved domain. Similarly one of the soyabean EST and the potato EST were incomplete and only contained a part of its zinc ribbon like domain (data not shown).

Taking the approximately 60 aa long conserved zinc ribbon domain among all the proteins as the reference point, the N-terminal as well as the C-terminal ends of a number of proteins were of variable lengths. A number of ESTs from sources like rice, maize, barley and tomato seemed to contain an incomplete N-terminal end of the sequence. This was concluded due to the absence of a complete ORF, the methionine residue at the start of the protein, and also through comparison with other complete ESTs. Despite this, all the incomplete sequences were long enough for analysing the homology among the different translated ESTs and ET1 and ZMZR1 proteins, as they all contained the conserved zinc ribbon like domain and seemed to contain the complete mature protein. All the putative proteins showed variable sequence homology among each other in the N- and C-terminal regions as well as variable sequence length, measured from the conserved zinc ribbon like domain (Fig. 3.24, next page).

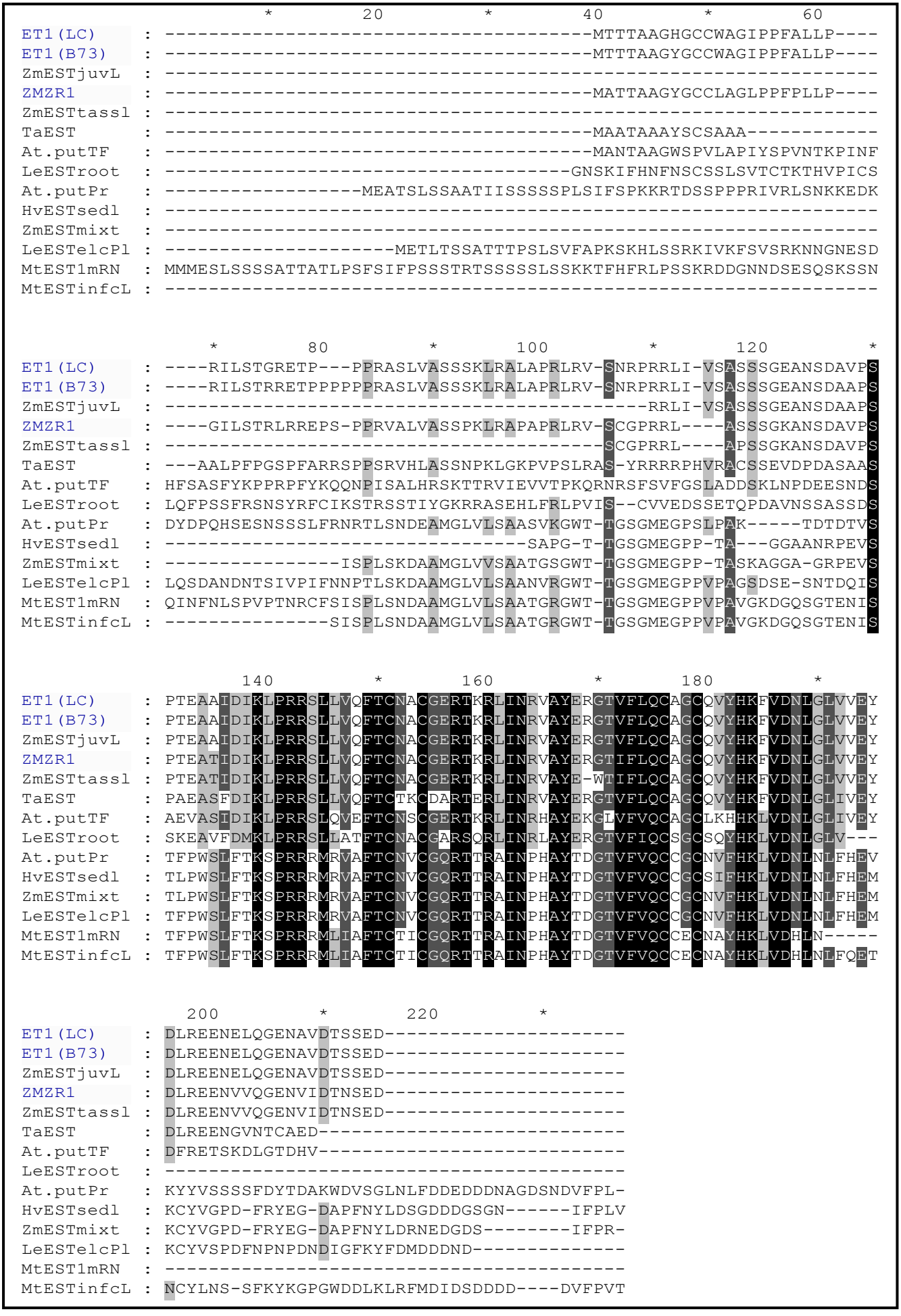

ET1(LC)
ET1(B73)
ZmESTjuvL
ZMZR1
ZmESTtassl
TaEST
At.putTF
LeESTroot
At.putPr
HvESTsedl
ZmESTmixt
LeESTelcPl
MtEST1mRN
MtESTinfcL

**Fig. 3.24: Comparison of the ET1 and ZMZR1 proteins with translations of ESTs and genomic sequences obtained from online databanks.** The names of sequences derived from the cloned ET1 and ZMZR1 cDNAs are written in blue. The other sequences in the alignment have been arranged based on their homology to ET1. The sequences from top to bottom are: ET1 from LC and B73 lines, ZmEST1: maize EST from juvenile leaf and shoot (cultivar: W64A) (gi:15313257, gb:BI478635.1), ZMZR1 from *et1-Ref* background, ZmEST2: maize EST from 1-3 mm tassel primordia (cultivar: OH43) (gi:9953030, gb:BE639613.1), TaEST: wheat flag leaf EST (gi:9846808, BE591735.1), At.putTF: *Arabidopsis* putative transcription factor (gi:6524186, gb: AAF15071.1), LeEST root: tomato root EST (gi:7333989, gb:AW622342.1), At.putPr: *Arabidopsis* putative protein (gi:7485299, T01795) and ESTs from mixed tissue (gi:239321, AA585805.1; gi:2413159, AA597736.1), HvESTsedl: barley EST from green seedling leaf (gi:11197727, BF266732), ZmEST3: maize EST from mixed adult tissue (cultivar: W23) (gi:6127241, gb:AW129887.1), LeESTelcPl: tomato EST from leaf inoculated with disease response elicitors (gi: 6061943, AW096348), MtEST1mRN: *Medicago* root nodule EST (gi: 10698688, BE998412) and MtESTinfcL: *Medicago* leaves after inoculation with *Colletotrichum trifolii* (gi: 11608735, BF520052).
Similar residues at a position in the alignment are shaded. The more common the homology among all the sequences, the darker is their shading. The common residues present in all the sequences are shaded black. The residues in the shaded column are sometimes not identical, but represent amino acids belonging to a common physiochemical group (Gene Doc: Nicholas et al., 1997). The homologies among all the proteins are the highest in the zinc ribbon domain region. At a number of positions, different residues, but with conserved physiochemical properties are observed.
Based on their homologies, two groups of sequences were identified from the alignment. The first eight sequences belong to the first group, where the sequences are more homologous to ET1 than to the remaining six ESTs in the second group. The last six sequences of the second group were more homologous to each other than to the ET1 group. Despite these small differences, all the sequences are highly homologous and contain a conserved zinc ribbon-like domain. Both the N-terminal and the C-terminal regions of the proteins around the zinc ribbon domain are variable in length as well as in percent homology between the two groups as well as among the sequences within each group.

Since the N-terminal sequences of ET1 and ZMZR1 proteins were known to be transit sequences for plastid localisation, the proteins obtained from the databanks were also analysed *in silico* for the presence of plastid targeting transit sequences. In TargetP analysis, all these plant proteins, including the partly incomplete ones were interpreted as plastid localised. The probability of being targeted to the plastids as compared to other cellular compartments was the highest. , In the case of proteins deduced from complete ORF sequences, the certainty for plastid localisation was higher than that obtained for the ET1 and ZMZR1 proteins (Appendix E). Based on

the analysis, the length of the transit sequences, for proteins with complete ORF, was variable among all the proteins, ranging from 35 to 80 aa residues (data not shown).

From all the homologous sequences obtained from the database, three different ESTs could be identified from *Zea mays* alone. From these, the translation of two incomplete ESTs showed more than 90% homology to ET1 and ZMZR1 protein sequences (Fig. 3.24). One of these clones, isolated from juvenile leaf and shoot tissue, was clearly more homologous to ET1, whereas the other, isolated from 1-3 mm tassel primordia, was more homologous to ZMZR1 protein. Their cDNA sequences also showed high homology to the *et1* and *zmzr1* cDNAs respectively. Another EST clone from *Zea mays*, isolated from mixed adult tissue (tassel, kernel, silk, husk, root, leaf in the ratio 4/2/1/1/1/1), was also found to be homologous to ET1, but was less homologous as compared to the other two *Z. mays* EST clones. Analysis of ESTs from other plant sources, like *Arabidopsis*, barley, *Medicago*, soybean and tomato also indicated the presence of two different ESTs from each plant. One of these two ESTs was more homologous to ET1 and ZMZR1, and the other more homologous to the third EST from *Z. mays*, so that they could be grouped into two different groups. From rice (immature leaf and apical meristem, gi:3763200, dbj:AU029952.1, not depicted) and wheat only one EST each was obtained, which was more homologous to the ET1 group of proteins. However, a few EST clones from other plant sources not depicted here, like ice plant (6 week old, gi:4464843, AI547355), potato (leaves, gi:13610149, BG592009.1), tomato seeds (gi:5894937, AW036095) and tomato red ripe fruits (without seeds and locules, gi:6976620, AW441369), were more homologous to the third EST clone from *Z. mays* (Fig. 3.24). Apart from these ESTs, one EST from potato (axilary buds representing developing stolons, gi: 9249551, BE340020.1; gi:9250329, BE340798.1) was found to be slightly different from both the groups of proteins in the zinc ribbon like domain.

Therefore, based on the homology observed in the alignment in Fig. 3.24, two main groups of sequences could be classified. The first group contained ESTs more homologous to ET1 protein and the second, containing ESTs more homologous to the third maize EST. The region of the protein sequences showing the differences between the two groups extended throughout the protein. However, the differences were clearer in the mature protein region. The transit sequences were less conserved among the proteins and showed more variability in the sequence length as well as

amino acid consensus. Despite this, the transit sequences of the individual groups were more homologous among each other than with those of the other group.

Based on the homology of ZMZR1 to ET1 and the protein structure analysis carried out with ET1 (see section 3.3.3), ZMZR1 was also found to contain a secondary structure like that of ET1 and was also homologous to the TFIIS/Rpb9 zinc ribbon domain.

# 4. Discussion

The aim of this study was the molecular characterisation of the *et1* gene responsible for a pleiotropic mutation showing abnormal endosperm development as well as pale seedling leaves during a period of 10-15 days after germination. This recessive mutation, significant due to the resulting reduced starch synthesis in the endosperm, was previously analysed largely through the mutant alleles of the *et1* locus. Therefore, this study focused on the analysis of the dominant wild type *et1* gene. The *et1* gene was analysed at the molecular level through isolation and sequencing of *et1* allele from Line C (LC) and its cDNA, and their comparison to the previously isolated *et1* allele from B73. The *et1* expression was further characterised through Northern analysis of different tissues of the maize plant. Based on the phenotypic indication that *et1* might be regulating plastid development, the *in vitro* translated ET1 protein was experimentally tested for chloroplast localisation and the deduced protein sequence was analysed *in silico* for the presence of functional domains or homologies to already characterised protein domains through comparisons with online databases.

Based on indications from Southern analyses with different lines that a sequence homologous to *et1* was present in the genome, the possibility that it might represent another gene was checked through analysis of genomic clones from the *et1-Ref* line. This revealed the presence of a novel *et1* homologue, *zmzr1*, which was characterised at the genomic level. Although Northern analyses with total RNA from different tissues did not reveal any *zmzr1* expression, a *zmzr1* cDNA was isolated from developing kernels; however, it represented very weak *zmzr1* expression in the kernels. Sequence comparisons of the two homologues showed a similar genomic as well as protein structure, but a number of differences in the non-coding regions of the gene were observed. The upstream regulatory region as well as the 3´ region of the gene showed almost no homology to *et1*. The *in silico* analysis of protein structure predicted a very similar secondary structure for the two mature proteins, after cleavage of the transit peptide.

## 4.1. Molecular characterisation of *et1* and its homology to *zmzr1*

### 4.1.1. Sequence Polymorphism in the *et1* gene – as revealed by the two alleles from LC and B73

Sequence analysis of the *et1* alleles from LC and B73 and their comparison reveal an overall similarity between the two. The *et1* gene, with its exons and introns, is about fifteen hundred base pairs in length and shows differences in sequence between the two alleles in the form of single nucleotide substitutions and a nine base pair long insertion/deletion (indel) in exon I. Comparison with the LC cDNA showed the same exon-intron structure in both the alleles. The splice sites also showed a high similarity to the monocot consensus for 5´ and 3´ splice sites (Simpson et al., 1996). In addition to the nine additional nucleotides present in the *et1* allele from B73, coding for three proline residues, single nucleotide substitutions were also observed in the coding sequence of the two alleles. Comparison of their deduced protein sequences, however, showed fewer differences between the two proteins, since a number of the single nucleotide substitutions in the ORF were synonymous codon substitutions present at the wobble position, and thus causing no consequential change in the protein sequence at that position (Fig. 3.22 and 3.23).

The gene upstream regions of the two *et1* alleles, however, showed some larger variations, like a number of indels (insertions/deletions) and sequence differences mainly localised to one region, about 200 bp long and 150 bp upstream of the transcription start. All of these variations together constituted about 14% of the promoter region and made them slightly different from each other (Fig. 3.21).

Altogether, this indicated a certain degree of genetic polymorphism between two alleles. This is, however, not unusual for maize. Genetic diversity in maize has been analysed with SNPs (Tenaillon et al., 2001), isozymes (Doebley et al., 1987), RFLPs (Dubreuil et al., 1999), RAPDs (Moeller et al., 1999), AFLPs (Lubberstedt et al., 2000), microsatellites (Senior et al., 1998) and chromosomal knobs (Smith et al., 1982), and all of these confirm that maize is highly variable within as well as among different maize populations.

Evidence through DNA sequence analyses also confirms this observation of sequence variability and polymorphism among alleles. On the whole, the observed

single nucleotide polymorphism (SNP) between the two *et1* alleles was found to be similar to the average SNP observed for 21 loci in 25 different maize individuals (Tenaillon et al., 2001). The sequence of the *glb1* locus was analysed from maize *(Zea mays* ssp. *may*s), its wild progenitor (*Z. mays* ssp. *parviglumi*s), a more distant species (*Z. luxurian*s), and *Trypsacum dactyloides* (Hilton and Gaut, 1998). It showed a high sequence variation including both nonsynonymous and synonymous polymorphism among all the maize and other *Zea* species as well as *T. dactyloides*. All nine examined maize alleles belonged to nine individual haplotypes, where each of these haplotypes represented one sequence variant. Similarly for *Adh1* locus, the same *Zea* species as above were analysed and showed a high degree of genetic diversity in maize (Gaut and Clegg, 1993; Walker et al., 1998). For four other maize genomic regions in seven maize cultivars (Shattuck-Eidens et al., 1990) and for the *Adh2* locus in maize (Goloubinoff et al., 1993), a high sequence variation, in the form of base substitutions as well as indels (insertions/deletions), was observed. For *Adh2* the sequence variation was observed to be higher in the non-coding region than in the coding region. Sequence diversity in the *terminal ear1* (*te1*) gene (White and Doebley, 1999), a regulatory locus in maize, was examined in twelve maize lines (*Z. mays* ssp. *mays*), five populations of *Z. mays* ssp. *parviglumi*s, six other *Zea* populations, and two *Tripsacum* species. All the species showed variable genetic diversity. Based on this analysis, the *te1* sequences from maize (*Z. mays* ssp. *mays*) could be grouped into three haplotypes, containing one, four and seven maize individuals respectively. The second haplotype also contained one teosinte individual.

Therefore, all of these examples indicate that sequence polymorphism observed for *et1* gene from the two inbred lines, LC and B73, is very common among homologous allelic regions of different maize lines, both in the form of indels and base substitutions – synonymous or nonsynonymous.

The sequence polymorphism observed between the two alleles was especially high in the gene upstream regulatory region. Despite that, the putative upstream gene regulatory elements identified through the *in silico* analysis were mostly present in the homologous regions of the two alleles (Fig. 3.21). Eight additional putative light responsive elements could be identified in the LC promoter region that was deleted in the B73 allele. On the other hand, in the region that differed from the B73 allele only

one light responsive element was identified in LC. Although from this region in the B73 allele were about eight light responsive elements identified. Similarly, other putative regulatory elements, such as endosperm specific elements, were also dispersed so that its absence in one allele at one position could be compensated by another position, which was then absent in the other allele. Moreover, Northern experiments carried out by O. da Costa e Silva (da Costa e Silva et al., 2001) show that *et1* expression in LC and B73 is comparably strong. This indicates that despite the sequence polymorphism observed in the two alleles, the gene shows normal expression and, therefore, is functional in both lines.

### 4.1.2. *zmzr1* is paralogous to *et1*, representing a gene duplication

Southern analysis with genomic DNA from different maize lines using *et1* specific probes revealed that an additional DNA sequence with homology to *et1* was present in the maize genome (Fig. 3.3; Fig. 3.19). Using *et1* specific probes against a genomic library the *zmzr1* gene was isolated. Sequence analysis and comparison to *et1* revealed a high sequence homology between the two genes, which was, however, mainly restricted to the coding region and portions of the introns (Fig. 3.14, Appendix B3). The upstream as well as downstream regions were found to be non-homologous. Although a similar exon-intron structure was observed for both the genes, the second intron in *zmzr1* was much larger (~500 bp larger) than that of *et1*. In all the intron sequences, including that from the second intron, a large number of nucleotide substitutions and indels (insertions/ deletions) were found. In the coding region as well as the 5´ and 3´ UTR, a number of indels and nucleotide substitutions were observed. Comparison of their deduced protein sequences showed that both the genes show both synonymous as well as nonsynonymous polymorphism and contain three indels. Based on their homology, it could, therefore, be concluded that *et1* and *zmzr1* are two paralogous genes, which probably arose due to gene duplication.

In maize, the presence of duplicated unlinked loci, which are found associated with chromosomal duplications, is often observed. A number of duplicated chromosomal segments have been detected. For example, extensive sequence similarity has been

observed between chromosomes 1S with 9L, 1L with 5S, 2S with 10L, 4S with 10S, 3 with 8, and the terminal segment of 6 with the central region of 8 (Helentjaris et al., 1988; Helentjaris, 1995). A number of genes, which are known to be present as single copy genes in other plant genomes within the grass family Poaceae (Clayton, 1987), like rice (Ahn and Tanskley, 1993) and sorghum (Whitkus et al., 1992; Pereira et al., 1993), have been found to be duplicated in maize. Through comparison of rice and maize genetic linkage maps, more than 72 % of the rice single copy genes were found to be present as duplicated loci in the maize genome (Ahn and Tanksley, 1993). In the case of sorghum, 44% of the RFLP markers detected more bands in maize than in sorghum (Pereira et al., 1993). Earlier isozyme studies (Goodman et al., 1980; McMillin et al., 1980; Davis et al., 1999) and RFLP mapping studies (Helentjaris et al., 1988; Davis et al., 1999) also indicated gene duplication in maize.

Such duplications were also indicated by early cytological observations showing chromosomal associations between nonhomologous chromosomes (McClintock, 1930). The reason for the presence of such a large number of duplicated genes is believed to represent a polyploidy event – very probably a segmental allotetraploidy – which lead to the duplication of the maize genome about 11 million years ago and whose genome later switched over to disomic inheritance (n = 10), while many other members of the tribe, *Andropogoneae* (Clayton, 1987), still maintained their original chromosome number (n = 5), (Gaut and Doebley, 1997; White and Doebley, 1998; Gaut et al., 2000). Segmental allotetraploidy involves hybridisation of species with only partially differentiated chromosome sets, so that they exhibit both bivalent and tetravalent formation during meiosis, until the shift to disomic inheritance, when further on only bivalent formation takes place during meiosis (Gaut and Doebley, 1997; Stebbins, 1971).

Comparison of the ET1 and ZMZR1 mature protein sequences revealed very similar primary and secondary structures. The differences in the amino acid sequence between the two proteins in the mature protein region did not change their putative secondary structures. As it appeared, all the critical amino acids necessary for correct protein folding and functioning were still conserved. Based on these analyses it could be concluded that the two proteins perform a similar function, even though they might be expressed differently (details in the next topic). Based on these analyses, the two genes were regarded as paralogous to each other.

### 4.1.3. e*t1* and *zmzr1* are expressed differently

Since the polyploidisation of the maize genome, it seems to have undergone a number of rearrangements as indicated by the observation that different segments of one chromosome show homology to segments of different chromosomes in maize (Helentjaris et al., 1988; Helentjaris, 1995). Moreover, the change to disomic inheritance also led to the independent evolution of these duplicated loci, causing a wide variation in their expression patterns. Examples of such differences are available from the duplicated loci *c2* and *whp* (Wienand et al., 1986; Franken et al., 1991), *a1* (Bernhardt et al., 1998), *gl8a* and *gl8b* (Dietrich et al., 2001), and *zmkcs-1* and *zmkcs2* (Frenzel, 2000).

In the case of *et1* and *zmzr1*, Northern analyses also revealed that *et1* and *zmzr1* are expressed differently. Comparison of their gene upstream regions showed that their sequences, which were more than 1kb long, did not show any homology to each other. In addition, comparison of their promoter regions to other online promoter databanks showed that the two genes possessed largely different regulatory elements. For *et1*, the expression pattern could be characterised (Fig. 3.6) and revealed that *et1* was expressed predominantly in the leaves and kernels. In the leaves it was expressed at all the stages examined, from 1 DAG to mature leaves. It was also expressed in stem, tassels before anthesis and 20 day old seedlings. The expression in kernels seemed to be restricted mainly to the endosperm, as separate examination of the embryo and kernels without embryo did not show any expression in the embryo. A weak basal-expression could, however, be observed in all the tissues examined.

The *in silico* promoter analysis (Fig. 3.21) revealed a large number of putative light responsive elements were present in the gene upstream region of *et1*. Another Northern experiment performed by O. da Costa e Silva (da Costa e Silva et al., 2001) with etiolated seedlings and seedlings grown in light revealed that only light grown seedlings showed *et1* expression, indicating that it might be light inducible. A number of regulatory elements involved in endosperm specific expression could also be identified in the upstream region.

However, for *zmzr1* no expression could be observed in the Northern analyses. Screening of a developing kernel cDNA library through PCR amplification indicated a weak expression of *zmzr1* in the kernels. An incomplete EST clone homologous to *zmzr1* was also detected in the 1-3 mm tassel primordia. This could, however, also indicate a basal expression, which could also be present in other tissues. On the other hand, other possibilities like expression induced in response to stress factors or other stimuli is also possible as indicated from the *in silico* promoter analysis of *zmzr1* (Fig. 3.21). The *in silico* promoter analysis suggested, among others, the presence of methyl jasmonate responsive elements, elicitor responsive elements, wound responsive elements and low temperature responsive cis regulatory elements in the gene upstream region of *zmzr1*.

### 4.1.4. Identification of orthologues and paralogues in other plant species, indicating the origin of ET1 and ZMZR1 as a duplication during maize evolution

Through online EST databank searches a number of ESTs, whose deduced protein sequences showed high homology to that of *et1* and *zmzr1*, could be identified. Analysis using TargetP showed that all of the ESTs had the highest probability to be localised in the plastids as compared to other cell compartments (Appendix E). The transit sequence length was analysed for ESTs (data not shown) containing complete ORFs and was found to be variable among all the deduced proteins. The transit sequences of all the proteins also showed a low degree of homology (Fig. 3.24). However, this showed that despite weakly conserved amino acid sequence of the transit sequences, all the proteins showed high probability for chloroplast localisation. This is not unusual for chloroplast located proteins, as all the chloroplast targeting transit sequences analysed so far do not show a common amino acid consensus or length (Fulgosi and Soll, 2001).

In comparison to the transit peptide region, all the proteins showed a much higher homology to *et1* and *zmzr1* in the mature protein region. With the help of the Genedoc program, quantitative analysis of the homologies among the sequences was carried out (Fig. 3.24). The proteins could be categorised into two main

homology groups. In one group, proteins showing higher homology to ET1 were present, which also included ZMZR1. In the other group proteins were less homologous to ET1, but more homologous between each other.

From maize, three protein sequences translated from ESTs could be identified. Two were present in the ET1 group, one of which was most homologous to ET1 and the other most homologous to ZMZR1 and, therefore, they apparently represent ET1 and ZMZR1 in their maize varieties respectively. The third maize protein sequence, ZmEST3, was present in the second homology group. From other plant sources like barley, *Arabidopsis*, tomato, *Medicago* and soybean, two paralogous protein sequences each were identified. However, one of these paralogues from each plant was more homologous to the ET1 group and the other to the second group, indicating that each of these two homology groups apparently contained orthologous sequences in different plant species.

The presence of putative orthologues in different plant species for the two groups of maize proteins indicated that they are not just the result of polymorphism often observed among different maize lines. Instead, they clearly point towards the conclusion that maize contains a family of ET1 like zinc ribbon proteins with at least three paralogous members, of which two were already characterised as ET1 and ZMZR1 and the third, ZmEST3, was identified from the EST search. The detection of ZmEST3 also explains the weak hybridising bands observed with the *et1* and *zmzr1* probes on southern blots with LC and *et1-Ref* genomic DNA, which were present in addition to those representing *et1* or *zmzr1* loci.

Since the other plant sources, both monocot and dicot, also contained two paralogous ET1 like proteins, representing orthologues of ET1/ZMZR1 and ZmEST3 respectively, it could be concluded that they represent an older polyploidy or duplication event which took place before the evolution of all these species. This also means that the third maize protein, ZmEST3, evolved from an ancestral sequence that existed before the evolution of maize. Whereas ET1 and ZMZR1 are recent paralogues. Very probably, these came into existence after the duplication of their progenitor sequence during a polyploidy event that gave rise to the *Zea* taxon (White and Doebley, 1998; Gaut et al., 2000).

## 4.2. Characterisation of ETCHED1 function

### 4.2.1. *et1* phenotype and the role of ET1 in endosperm development

As part of the functional characterisation of the *et1* gene, the *et1* phenotype was examined in 20 DAP *et1-Ref* kernels (Fig. 3.1). The endosperm region was found to be fully cellularised. A continuous aleurone monolayer, which appeared to grow normally, was observed at the periphery of the endosperm. However, in transverse kernel sections, the starchy endosperm was found to be divided into alternate sectors of starchy and starchless cells, where the broader starchy endosperm sectors were interspersed by starchless sectors radiating out from the central endosperm region. The central region was also found to be filled mainly with starchless cells. At the kernel periphery, in the regions with starchless cells, the endosperm – including the outermost aleurone layer – was found to be pulled away from the maternal pericarp, giving rise to depressions on the kernel surface. These depressions were found to be responsible for the “etched” appearance of these kernels. On maturity, the cross-sections of these kernels only showed traces of these starchless cells in the form of either empty spaces in these regions, especially in the central endosperm region, or traces of compressed cells or membranes in the starchy endosperm .

Based on these observations and those made earlier with mature kernels (Fig. 1.3), it can be concluded that on the onset of the desiccation process, which was observed around 20-22 DAP onwards, the smaller starchless endosperm cells begin to shrink together. As they are mainly composed of cytoplasmic fluids, they probably cannot maintain their structure and collapse due to the desiccation process, unlike the neighbouring starchy endosperm cells, filled with starch grains. Therefore, depending upon the degree of compression, pits or hollow spaces in the endosperm are apparently formed. When the aleurone layer is present above a larger sector of starchless cells, it is probably pulled down deeper into the starchy endosperm on the onset of desiccation and is observed at maturity as a compressed layer present between the starchy cells in the radial plane (Fig. 1.3).

Moreover, based on two observations it can also be assumed that the *et1* mutation does not affect the endosperm cellularisation process. First, the complete endosperm was found to be cellularised at 20 DAP and no acellular regions were observed.

Secondly, earlier examinations of endosperm development (Olsen et al., 1999; Opshal-Ferstad et al., 1997) have shown that complete cellularisation in maize takes place around 4 days after pollination (DAP), and based on the Northern analysis with *et1* probe (Fig. 3.6), only a basal *et1* expression was observed in the kernels during this period, which showed an increase only after 6 DAP. Based on these observations, *et1* does not seem to play any role in the endosperm cellularisation process.

The absence of starch grains in only some parts of the starchy endosperm (Fig. 3.1) indicates aberrant amyloplast development in these regions. However, this also points out that *et1* being a possible mutation in a structural gene involved in starch biosynthetic pathway is less probable than it being a mutated regulatory gene co-ordinating amyloplast development in the endosperm. Due to the presence of radial starchy sectors in the endosperm, it could be supposed that *et1* is also induced or regulated by spatial and temporal cues, which might even be originating from the surrounding maternal tissue.

The presence of starchy sectors in the endosperm (Fig. 1.3; Fig. 3.1) is similar to a pattern described in the clonal patterns of endosperm development (McClintock, 1978). There, the endosperm tends to first divide into right and left halves, then into quarters, and finally into cone like sectors emanating from the centre of the endosperm out to the surface (Becraft and Asuncion-Crabb, 2000; Lopes and Larkins, 1993). Therefore, *et1* might have a clonal basis of expression. This also indicates that *et1* is switched on at an early stage of endosperm development, probably as soon as cellularisation has taken place. It might be accompanied by another regulatory gene at a later stage of development, or in its absence either a second regulatory pathway or a second gene is switched on at a later stage of endosperm development, which would replace the *etched1* function in the endosperm. This would also explain the delayed development of the endosperm, indicated by the sector formation. In addition, the presence of such sectors indicates that it is probably temporal and/or spatial factor(s), which influence *et1* function during endosperm development. However, so far little is known about such factors or the developmental decisions (Lopes and Larkins, 1993) controlling cell lineages and their cellular interactions leading to pattern formation, which involve shapes, patterns and chemical compositions of the endosperm.

That ET1 might be playing an important role in early endosperm development is also indicated by the fact that in *et1-Ref* kernels a variable intensity of the *et1* phenotype is observed. A cause for this might be that as soon as the second gene regulating amyloplast development is switched on, it might be compensating for the *et1* function, therefore, partly effacing the affect of *et1* mutation and the lagging starch synthesis, and sometimes giving rise almost normal looking kernels.

Similar to endosperm development, during early seedling development, the first leaves after germination are pale and remain so up to about 10-15 days after pollination. Afterwards, all the new leaves are green and grow normally, and the older leaves also turn almost green. This clearly indicates that the *et1* function is replaced during development by other factors or regulatory systems in the chloroplast, just like in the amyloplasts of the endosperm. This further indicates that *et1* is involved in the regulation of early plastid development in the endosperm (amyloplasts) and chloroplasts.

The *et1* expression pattern also indicated that *et1* is involved in plastid development (Fig. 3.6). *Et1* showed expression in all the plant parts containing chloroplasts, i.e., leaves, right from 1 DAG onwards and up to mature leaves, in the stem, in tassels and 20 DAG seedlings. Among these tissues, the expression was the strongest in the leaves. Apart from that, *et1* showed strong expression in the kernels too. Its expression seems to begin by about 6 DAP and continues up to late kernel development. That this expression was in the endosperm region of the kernel was indicated by the Northern experiment, where on separation of 20 DAP kernels into embryo and the rest, the embryo only showed basal expression and the remaining kernel, mainly containing the endosperm, showed strong *et1* expression (Fig. 3.6).

### 4.2.2. ET1 contains a zinc ribbon like domain, showing homology to the transcription factor, TFIIS

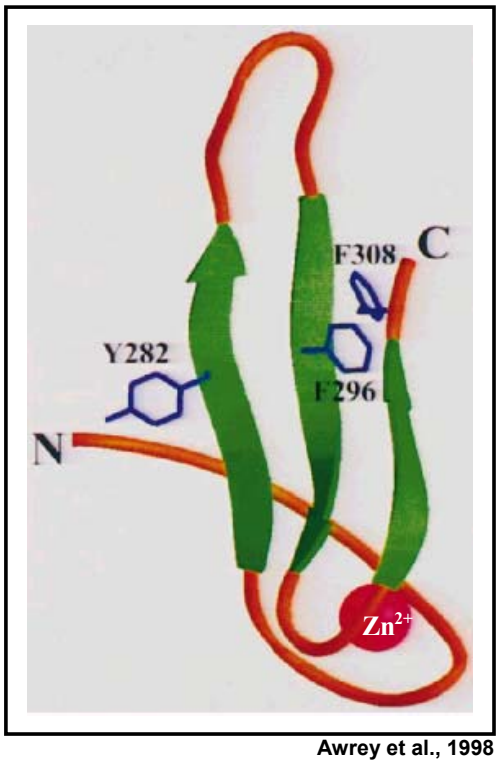
Comparison of ET1 to online databases revealed that it contains a protein domain with structural homology to the nucleic acid binding zinc ribbon domain of eukaryotic transcription elongation factor, TFIIS.

TFIIS is the general transcription elongation factor that helps the eukaryotic RNA Polymerase II (RNAP II) elongation complexes to read through blocks to transcription. RNAP II elongation factors can be categorised into three functional classes (Shilatifard, 1998). One class includes proteins regulating the rate of transcription elongation by allowing transcription through nucleosomes. A second class of proteins increase the catalytic rate of transcription by stimulating the rate of nucleotide incorporation. The third class includes transcription elongation factors stimulating RNAP II activity by enabling it to transcribe through hindrances to elongation such as DNA-binding proteins, DNA-binding drugs or specific DNA sequences that promote transcription arrest. TFIIS belongs to the last category and enables RNAP II to read through DNA-dependent blocks to transcription. It also has archaeal and viral homologues as well as bacterial functional homologues (Olmsted et al., 1998).

TFIIS consists of three structural domains termed I, II and III. The Domains I and II are involved in RNAP II binding. The Domain I consists of a 111 aa long four-helix bundle, which interacts directly with the RNAP II holoenzyme (Booth et al., 2000). The region containing domains II to III have been found to be sufficient for TFIIS elongation activity *in vitro* and is able to rescue TFIIS mutant phenotype in yeast cells (Nakanishi et al., 1995). Domain II, which consists of a 90 aa long N-terminal 3-helix bundle, has been shown to be necessary for TFIIS activity *in vitro*. Between the domains II and III is a 25 aa long linker region, which does not have any defined secondary structure but is necessary for TFIIS elongation (Awrey et al., 1998). Domain III consists of a small 46-residue zinc ribbon motif, which is supposed to be directly involved in transcript cleavage and resumption of transcript elongation by RNAP II. It is also this domain that shows homology to the ET1 protein.

A zinc ribbon domain, as characterised from TFIIS, contains a three-stranded anti-parallel β-sheet structure (Qian et al., 1993; Olmsted et al., 1998). At one end of the β-sheet, a $Zn^{2+}$ ion stabilises this globular mini-domain by holding together the two loops, between which the first two β-strands are present, which harbour the four cysteines that chelate the $Zn^{2+}$ ion (Fig. 4.1). At the opposite end, a flexible disordered loop, which is important for the TFIIS function, is present between the first two β-strands of the β-sheet. Mutational studies have shown that a mutation in the

negatively charged dipeptide Asp-Glu (D-E), present in this disordered loop, leads to inability of the RNAP II to carry out transcript cleavage and read-through (Jeon et al., 1994; Awrey et al., 1998). In addition, the positively charged residue present in this loop was also found to be indirectly important. Its replacement with either of the anionic residues (D, E) lead to total loss of TFIIS activity, whereas with other neutral hydrophobic residues, it did not show a complete loss. Apart from that, three aromatic phenylalanine residues, present at three different positions, including the last two β-strands, were also experimentally tested to be necessary for TFIIS activity (Awrey et al., 1998). Since mutations in all these amino acids lead to loss of TFIIS activity, without affecting the native structural confirmation of the TFIIS protein, it could be concluded that they are directly involved in transcription activity (Awrey et al., 1998).




Awrey et al., 1998

**Fig. 4.1: Ribbon diagram of the Zinc Ribbon domain of yeast TFIIS.** The β-strands are shown as green arrows, the zinc atom as a red point. The aromatic amino acids, with their side chains shown in blue, are the putative residues involved in nucleic acid interaction (Awrey et al., 1998).

Based on sequence homology, ET1 was found to contain the four cysteine residues of a zinc ribbon domain. It was also found to contain the three conserved phenylalanine residues at their respective positions (Fig. 4.2). The predicted secondary structure of ET1 also showed similarity to the zinc ribbon domain of TFIIS. All the different secondary structure programs predicted the last two of the three β-strands and the two turns or loops of a zinc ribbon for the ET1 protein domain. However, the region in ET1 corresponding to the first of the three β-strands was predicted by some methods as an α-helix, and by some others as a β-strand. The loop region, corresponding to the disordered loop – TRSADEP – of TFIIS, was, however, comparatively smaller in ET1. Moreover, another aromatic amino acid, tyrosine, present in the first β-strand of TFIIS and also postulated to be involved in

TFIIS activity was not found at its position. Instead, a tyrosine was present in the putative truncated disordered loop of ET1. The loop contained only one of the two conserved anionic residues (D-E) of TFIIS supposed to be necessary for its function, i.e., Glu (E).

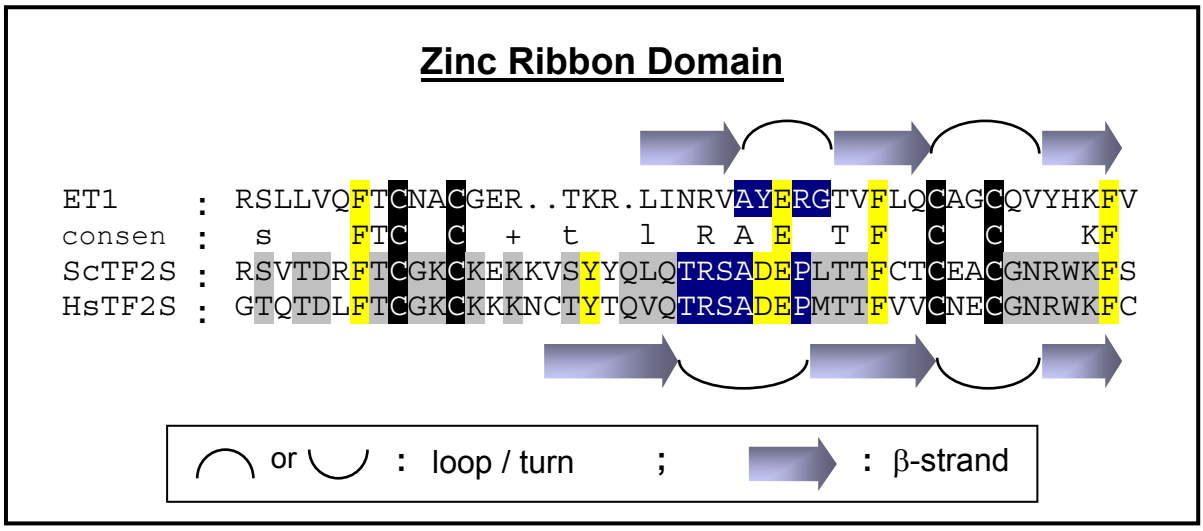



**Fig. 4.2: Comparison of the three-stranded β-sheet region of the zinc ribbon domain of TFIIS from yeast and humans to the corresponding ET1 protein secondary structures.** From top, the first sequence in the alignment represents the zinc-ribbon-like domain of ET1 (residues 91-134), followed by the consensus sequence (consen) with the zinc ribbon domain of TFIIS. Below it are the sequences of the zinc ribbon domains of yeast (ScTF2S, residues 263-309) and human (HsTF2S, residues 288-335) TFIIS respectively. One of the two secondary structure predictions for ET1 zinc ribbon like domain, showing higher homology to that of TFIIS is depicted above the alignment, whereas the NMR determined 3-stranded β-sheet secondary structure for yeast TFIIS is depicted below it. The four cysteines common to all the three sequences are shaded black. The disordered loop, supposed to be involved in TFIIS activity and its putative corresponding structure in ET1 is written in white and shaded dark blue. Those residues, known to be important for TFIIS activity and when present in ET1, are shaded yellow. The remaining conserved residues of the two TFIIS proteins are shaded grey. The consensus shows conserved residues, including those belonging to a common physiochemical group, among the protein sequences.

Rpb9, a small RNAP II sub-unit, also contains two zinc ribbon domains separated by 30-amino acid linker region. The linker region harbours a conserved sequence needed for binding to RNAP II (Hemming and Edwards, 2000). The N-terminal domain has been shown to be involved in transcription initiation (Hemming et al., 2000), but does not influence transcription elongation greatly (Hemming and

Edwards, 2000). The C-terminal zinc ribbon domain, based on experimental evidence, has been shown to be involved in transcript cleavage and read through during transcription elongation and is required for TFIIS activity (Awrey et al., 1997; Hemming et al., 2000). This zinc ribbon domain also possesses a 3-stranded β-sheet structure, same as that of TFIIS (Wang et al., 1998).

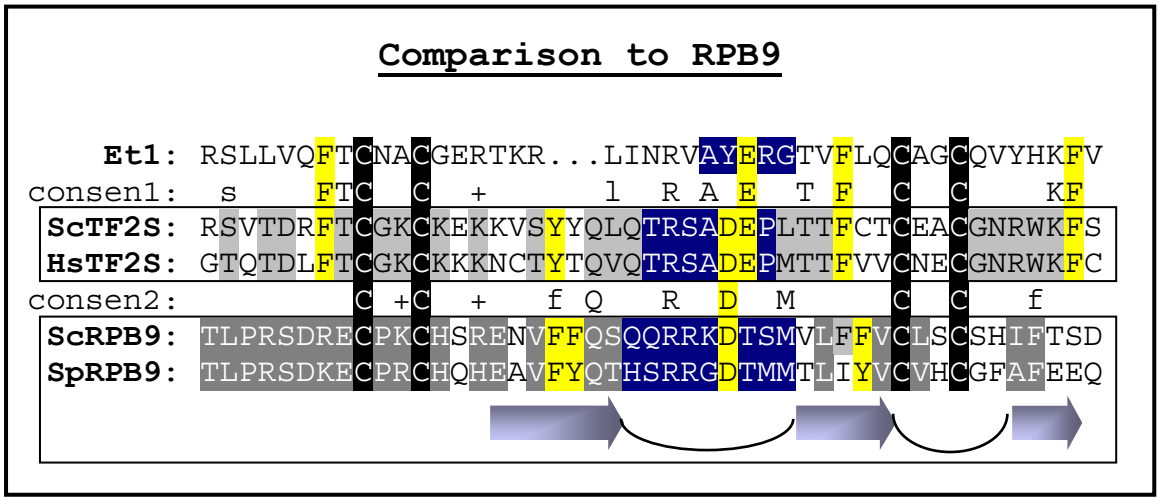



**Fig. 4.3: Alignment of the C-terminal zinc ribbon domain of RNAP II sub-unit, RPB9, with that of TFIIS, and the homologous region in ET1.** From top to bottom, the first sequence in the alignment represents the zinc-ribbon-like domain of ET1, followed by the consensus sequence (consen1) with the zinc ribbon domain of TFIIS, whose sequences from yeast (ScTF2S) and human (HsTF2S) respectively are present below it. At the bottom are the sequences of the two C-terminal RPB9 zinc ribbon domains from yeast (ScRPB9) and fission yeast (SpRPB9) respectively. The NMR determined β-sheet secondary structure for RPB9 (Wang et al., 1998) is depicted below its alignment (legend in Fig. 4.2). The consensus (consen2) between TFIIS and RPB9 zinc ribbon domains is present between their alignments. The four cysteines common to all the three different proteins are shaded black. The disordered loop, supposed to be involved in TFIIS activity is written in white and shaded dark blue. Those residues, known to be important for zinc ribbon domain activity and when present in ET1, are shaded yellow. The remaining residues conserved within the individual alignments of proteins are shaded grey. The consensus shows conserved residues, including those belonging to a common physiochemical group, among the protein sequences.

Nonetheless, its sequence as well as structural homology to TFIIS is variable and shows a number of differences. Just like ET1, it also contains only a single negatively charged residue in the charged disordered loop. Despite that, this region is functionally conserved. The length of this loop is also different from that of TFIIS (Fig. 4.3). The first of the three phenylalanines in TFIIS, suggested to be important for its

activity, does not have a counterpart in Rpb9. Moreover, architecturally, the Rpb9 protein also differs in the orientation of its $Zn^{2+}$ binding region from TFIIS (Wang et al., 1998). On the whole, both Rpb9 and ET1, based on its *in silico* secondary structure analysis, deviate from TFIIS in their architecture. Therefore, just like RPB9, which itself contains two tandem zinc ribbon motif variants, ET1 probably represents a novel zinc ribbon domain.

ET1 also shows high homology to a novel domain, PD022385, classified in the PROSITE database (Hofman et al., 1999), which is based on the sequence homology of the included proteins. It also includes, among others, the two ET1 homologues from *Arabidopsis thaliana*. The length, primary sequence and the physiochemical nature of the aligned residues in the sequence of these proteins, as well as ET1, are highly homologous. However, none of the proteins with the domain have been structurally analysed so far. Although, based on the structural analyses carried out *in silico* with some of the proteins containing this domain, as well as other plant homologues identified during this study, they all seemed to contain a very similar secondary structure as ET1 in the putative zinc ribbon domain (data not shown). For some proteins, the presence of a three-stranded β-sheet in the zinc ribbon domain was more clearly predicted than for ET1.

As mentioned in section 4.1.4, all the plant homologues identified from the database, including the different paralogues in the same plant species, were predicted to be plastid localised. On analysing the proteins containing the protein domain, PD022385, for sub-cellular localisation, all the non-plant proteins were found to be mitochondria-targeted. Only two of the sequences in the domain were from a plant source, *A. thaliana*, and were predicted to be plastid localised. It was not unexpected that the non-plant homologues of the plastid-located-proteins were located in mitochondria. Both mitochondria and plastids are two sub-cellular semi-autonomous systems that posses their own genomes and perform a number of genetic functions, like replication, transcription and translation. Both these compartments are supposed to be of prokaryotic origin, which evolved through endosymbiosis, and posses a number of structural and functional similarities (Schwartz and Dayhoff, 1978; Dalbey and Kahn, 2000). A number of these similarities, like the translocon proteins on their envelop membranes, needed for protein import, the intra-organeller translocation

systems, and the nuclear-encoded RNA polymerases, may have also arisen due to the evolution of these organelles together in a eukaryotic host cell.

On the other hand, it is very interesting to note that no mitochondria located plant paralogues have been identified so far. The lack of mitochondria located paralogues would indicate two possibilities. First, that due to competition with the second endosymbiotic organelle, the plant mitochondria have substituted for the requirement of this protein during evolution. Second would be that the protein itself underwent more intensive structural changes, which could have caused a change in the protein structure of the plant mitochondrial homologues, such that their requirement for a $Zn^{2+}$ as a component for stabilising their protein structure being replaced by some other unknown structural features. An example of such a change is available from the enzyme, Topoisomerase I, which is involved in DNA manipulation and generally contains a number of tandem zinc binding motifs containing four cysteine ligands each. However, in a number of prokaryotes, some of these domains have undergone structural changes. One such domain in *E. coli*, containing two tandem repeats of this motif, has lost some of the four cysteine residues involved in zinc binding, and consequently does not form a zinc binding motif. Despite that, this domain maintains its structural as well as functionally similarity to its homologous motifs in other Topoisomerase I (PDB id: 1yua; Yu et al., 1995). Similarly, in a Topoisomerase I from a hyperthermophilic eubacterium, *Thermotoga maritima*, the single zinc binding domain still possesses its tetracysteine sequence motif (Viard et al., 2001), but has seemingly undergone other changes in order to adapt itself to the extreme environmental conditions. It also does not seem to require the zinc ion for its correct structural confirmation, as replacement of one or more of the four cysteines does not affect the function of the protein.

### 4.2.3. Role of ET1 as a zinc binding domain: in nucleic acid or protein-protein interactions?

Zinc binding motifs have been identified from a large variety of eukaryotic as well as prokaryotic proteins. Although these proteins possess highly variable structures as well as functions, most of them, in general, are regulatory proteins that are all evolutionarily related. In most of these proteins, zinc is only indirectly involved in the

protein function, where it plays an important role in providing the correct domain confirmation (Markov et al., 1999; Yanagisawa, 1995). These domains, requiring zinc, have been found to be too small to fold into the right confirmation without the aid of the $Zn^{2+}$ ion. Zinc provides the structure-stabilising cross-links, without introducing undesirable chemical reactivity (Berg and Shi; 1996).

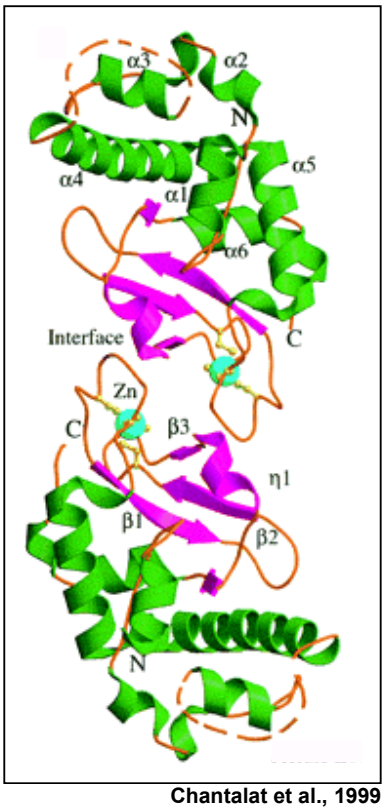
The ET1 zinc ribbon like domain also corresponds, in description, to such small domains. The protein itself is very small, consisting of only 99 amino acids, of which only about 45 central amino acids constitute the putative zinc binding domain with the zinc ribbon motif. The remaining regions around this domain seem to form random coils or helices. Moreover, based on the phenotypic analyses of *et1* mutants, it also appears to be a regulatory protein or a part of one.

Among zinc binding domains, another domain very similar to zinc ribbons is the zinc finger motif, which is very common and is present mostly as small multiple tandem repeats in eukaryotic proteins, like TFIIIA. They are known to be involved in DNA binding and contain two antiparallel $\beta$ strands, followed by an $\alpha$ helix (Lee et al., 1989). The tandem finger like structures of TFIIIA interact with the DNA duplex and help in transcription initiation (Berg and Shi, 1996). However, these motifs are slightly smaller than that of ET1. Usually, they contain 12-17 residues present between cysteine and hystidine pairs. Single zinc fingers as well as those involving four cysteines as zinc ligands are also known. Translation Initiation Factor, eIF-2$\beta$, and the putative pre-rRNA processing protein SRD1 are such proteins from yeast containing one zinc finger motif each. Both the proteins show similarity to each other and contain a zinc finger motif towards the C-terminal end. As mentioned before, small zinc binding domains are found in a large variety of proteins carrying out different functions. At the molecular level, this could involve nucleic acid binding, both with single stranded or double stranded, and either with RNA or DNA. On the other hand, this could also involve protein-protein interactions. The three-stranded $\beta$-sheet structure of a zinc ribbon domain has also been found to display both these types of interactions. The zinc ribbon motif of the transcription elongation factor TFIIS, which also has homologues in RNAP I and III, represents a domain interacting with the RNA being transcribed in a stalled ternary elongation complex with the RNA Polymerase II and the DNA template (Quian et al., 1993; Powell et al., 1996).

Similarly, the small subunit RPB9 of RNA Polymerase II, as well as its homologues in archaea, are known to contain two functionally different zinc ribbon motifs (Wang et al., 1998). The first domain of RPB9 plays a role in identification of the right transcription start site by RNAP II and the second domain is involved in transcript cleavage and elongation, along with TFIIS (Hemming et al., 2000; Hemming and Edwards, 2000). The TFIIS homologue in RNAP III has been found to be involved in transcript cleavage during transcription termination rather than transcription elongation and helps overcome a kinetic barrier to the termination process (Chédin et al., 1998). Another comparatively different zinc ribbon domain involving three cysteins and one histidine, and containing two β-strands, has been identified in the N-terminal domain of the transcription initiation factor, TFIIB from eukaryotes as well as archaea (Chen et al., 2000).

In general, β-sheet structures, especially three-stranded β-sheet structures, are very common among nucleic acid binding proteins, both DNA (Connolly et al., 2000) and RNA binding proteins, which might not necessarily contain a zinc binding site. Some examples include S1 RNA binding domain found in RNA associated proteins, like the S1 ribosomal protein (Bycroft et al., 1997), the RNA-binding domain of a small nuclear U1A spliceosomal protein (Oubridge et al., 1994), and the DNA binding domain of the replication protein A (Bochkarev et al., 1997).

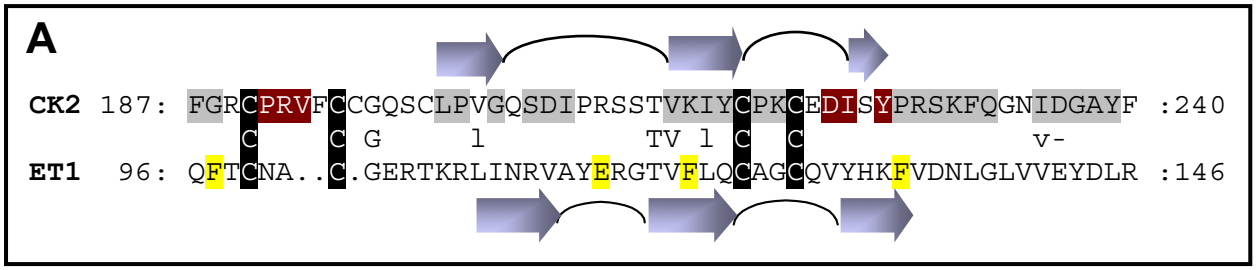
However, that zinc ribbon domains can also be involved in protein-protein interaction is also illustrated with the example of the eukaryotic Ser/Thr protein kinase, Casein Kinase II (CK2),

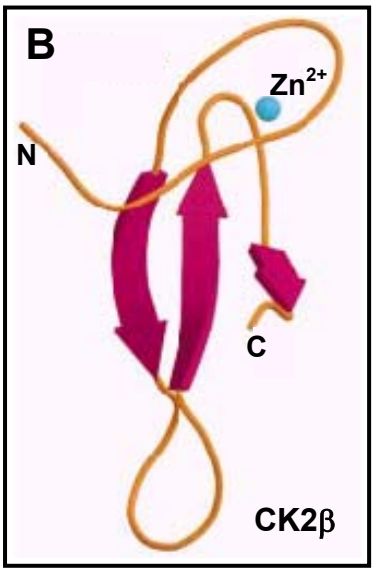



Chantalat et al., 1999

**Fig. 4.4: Ribbon diagram of the CK2β$^{\delta}$ dimer.** The β-strands in the zinc ribbon domain and a $3_{10}$ helix are shown in pink arrows and coils respectively. The zinc atom and the four cysteines are shown in ball and stick model in light blue and yellow respectively. The α-helices are shown in green. The α-helices and β-strands are numbered from N-terminal to C-terminal end of the monomer.

studied from animal, yeast and plant systems (Chantalat et al., 1999; Sugano et al., 1998). The human CK2 has been structurally analysed to show two α catalytic subunits and two β regulatory subunits (Fig. 4.4), of which the β subunit shows homodimerisation, mediated by the three-stranded β-sheet structure of the zinc ribbon domains of the opposing subunits. Four cysteines act as zinc ligands in each subunit. Although the zinc ribbon domain of CK2β subunit shows a remarkable similarity in its topology to that of TFIIS, a number of differences are also observed. Whereas the TFIIS zinc ribbon domain is a highly soluble nucleic acid binding domain, the CK2β zinc ribbon domain is lined by hydrophobic amino acids involved in hydrophobic interactions with that of the opposing monomer to form a dimeric interface. However, it is mainly the C-$X_4$-C region (Fig. 4.5) and the β3 strand, whose side chains are involved in the hydrophobic interactions. The remaining two β-strands, β1 and β2, are accessible for interaction with other molecules, and the loop between them is also accessible to other proteins and has been found to be exposed to the outer surface in a dimer (Fig. 4.5; Chantalat et al., 1999).




Chantalat et al., 1999

**Fig. 4.5: Structure of the zinc ribbon domain of CK2β subunit and its alignment to that ET1. (A)** The upper sequence in the alignment represents the zinc-ribbon domain of *Arabidopsis* CK2β subunit, followed by the consensus sequence with the zinc ribbon-like domain of ET1, whose sequences is present below it. The three-stranded β-sheet structure of both the protein sequences has been aligned to them respectively. The four cysteines common to both the protein sequences is shaded black. The conserved residues of CK2β are shaded grey and the residues involved in the hydrophobic interactions, which form the dimer interface, are shaded brown. The four residues in ET1, shaded yellow, were identified from the homology to TFIIS. These were involved in nucleic acid interaction in TFIIS.

(**Fig. 4.5:** continued on the next page)

**(B)** Ribbon diagram of the zinc ribbon domain of human CK2β with its three-stranded β-sheet structure (Chantalat et al., 1999). The zinc ion ($Zn^{2+}$) is shown as a blue ball, whereas the β strands in red. The remaining peptide chain is shown in brown. The N-terminal (N) and C-terminal (C) ends are indicated.

Therefore, both of these proteins, TFIIS and CK2, represent examples of two similar three-stranded β-sheet zinc ribbon domains involved in totally different interactions. Whereas, TFIIS is needed for transcript cleavage and read through during transcription elongation of RNAP II, the zinc ribbon domain of CK2β subunit is more complex and is involved in forming a dimeric interface as well as in other interactions. ET1 also shows some similarities with the CK2β zinc ribbon domain, but it might, just like with TFIIS, be of topological nature and not of any significance for Et1 function.

There are other functions, identified from putative zinc ribbon domains of archaeal proteins, like RNA binding during translation (NP 148639.1, GI: 14602215 ; NP 276810.1, GI: 15679901), or they constitute a part of certain DNA methylases (NP 071283.1, GI: 11500033). A eukaryotic protein family, SIR2, comprising a novel class of NAD-dependent protein deacetylase, function in transcriptional silencing, DNA repair and lifespan-extention in yeast. It contains a unique variant of the three-stranded β-sheet zinc ribbon domain. SIR2 has also been evolutionarily conserved in diverse organisms from bacteria to humans.

### 4.2.4. Role of ET1 in chloroplasts

Since the chloroplast import experiments show that ET1 is plastid localised and very probably in the stroma, its functional sphere would probably involve physiological activities being carried out in the stroma. Plastids are semi-autonomous organelles, which are believed to have originated from oxygenic photosynthetic prokaryotes (Howe, 1996). Therefore, the stroma represents an equivalent of the prokaryotic cytoplasm and carries out a number of functions independent of the cell. These functions also include transcription and translation. Based on its structural homology to known eukaryotic as well as prokaryotic zinc binding proteins, involved in

transcription and translation, it could be postulated that ET1 might also be taking part in these processes being carried out in the plastids.

The Northern analyses with *et1* probe have shown that apart from the leaves, tassels and endosperm, where a strong *et1* expression is observed, *et1* shows a basal expression level in all the remaining tissues examined. It is not unusual that a gene shows basal expression level in the absence of the requirement of its gene product. However, another explanation for this expression pattern could also be that ET1 influences the housekeeping functions of all the plastids, but its requirement during increased physiological activity in the plastids is also increased, as in the amyloplasts of the developing kernels, and the chloroplasts in the leaves. This increased expression is probably stimulated by upstream elements of the *et1* gene, which respond to external (light) or internal (endosperm development) stimulatory factors. Increased *et1* expression in response to light has been shown through Northern analysis (da Costa e Silva et al., 2001), where etiolated seedlings, grown in dark, or seedlings grown in light showed absence or strong expression of the *et1* transcripts respectively. All these possibilities also indicate that ET1 might be controlling plastid development either at the transcriptional level or at the post-transcriptional level.

Although a large majority of plastid genes are believed to have been transferred to the nucleus during evolution, it still maintains a large number of vital genes, most of which encode components of the photosynthetic electron transport machinery, and transcriptional and translational apparatus (Stern et al., 1997). Transcription of the remaining plastid genes is driven by at least three different plastid RNA polymerases. One transcription apparatus comprises a multi-subunit homologue of *E. coli* RNA polymerase, called PEP (**P**lastid **E**ncoded RNA **P**olymerase), which is the only plastid encoded polymerase (Howe, 1996; Isono et al., 1997; Tan and Troxler, 1999; Hu and Bogorad, 1990). Another transcriptional apparatus comprises a NEP (**N**uclear **E**ncoded RNA **P**olymerase), which appears to be a single subunit homologue of the T7/T3 phage RNA polymerase (Allison et al., 1996; Maliga, 1998; Chang et al., 1999). Various studies indicate the presence of a third nucleus encoded RNA polymerase (Bligny et al., 2000; Hedtke et al., 2000). In *Arabidopsis thaliana*, whose complete genome has now been sequenced, showed the presence of three different phage type RNA polymerase genes, of which two had already been characterised as encoding plastid and mitochondrial NEPs respectively. The third RNA polymerase

was found to show dual targeting to plastids as well as mitochondria (Hedtke et al., 2000).

All these polymerases seem to work independent of each other, under the control of different regulatory mechanisms, and show differential activity in different tissues as well as developmental stages. Moreover, analyses of plastid promoters also revealed that these systems use different promoters for transcription. Many plastid genes also seem to contain promoters for more than one plastid RNA polymerase (Pfannschmidt and Link, 1994; Stern et al., 1997; Hajdukiewicz et al., 1997; Serino and Maliga, 1998; Maliga, 1998; Silhavy and Maliga, 1998; Bligny et al., 2000). A NEP is supposed to be the first polymerase to become active during early plastid development, which transcribes the genes for ribosomal rRNAs and the PEP transcription machinery. It is then soon followed by the PEP, whose activity upon illumination increases greatly (Stern et al., 1997). Since Northern analyses indicate that ET1 might be active right from a very early leaf developmental stage, it could be involved in transcription by acting as a transcription factor for either a NEP or PEP, or as their subunit. Moreover, the leaves of *et1* mutants remain pale during early seedling development, showing the lack of chloroplast development, like the absence of organised thylakoids, as observed in *et1* mutant seedlings (EM micrographs, Fig. 1.3). Since it also influences amyloplast development, it is very likely not a part of the photosynthetic apparatus. However, the genes for photosystems I and II are transcribed by the PEP, and PEP encoding genes, in turn, are transcribed by an NEP. Therefore, ET1 could be interacting with either the PEP or any of the NEPs.

One explanation for the recovery of the *et1* seedlings from *et1* phenotype after about 15 DAG could be that ET1 functions as a transcription factor that is replaced by another transcription factor in its absence. On the other hand, a large number of plastid genes are known to possess more than one type of promoter elements and are transcribed by more than one type of transcriptional apparatus (Krupinska and Falk, 1994; Stern et al., 1997; Maliga, 1998; Silhavy and Maliga, 1998; Bligny et al., 2000). Therefore, it could be possible that in the absence of ET1, another transcription apparatus is activated, which replaces the ET1-associated transcription apparatus and transcribes those genes, thus bringing the plastid system back to normal. However, it is also possible that ET1, as a transcription factor, is crucial to the plastid system during the early seedling development. In its absence, the

seedling just manages to tide-over this early period of growth with the help of the storage reserves of the endosperm. Thus, providing it with the needed relief until the next transcriptional apparatus can become active to carry out its function.

However, based on its homology to the zinc ribbon domain, the ET1 protein could also be involved at the post-transcriptional or translational level, instead of during transcription. The relative transcription rates of most chloroplast genes have been found to be almost constant, which indicates that post-transcriptional and translational mechanisms play an important role in the differential expression of chloroplast genes (Stern et al., 1997). Analysis of nuclear mutations have revealed that a large number of nuclear-encoded factors are involved in plastid development, which principally influence and regulate plastid expression pattern at the post-transcriptional level (Rochaix, 1996). ET1 could be involved at various steps, which demand an interaction of the proteins with the transcribed RNA molecules, like in RNA stability, RNA processing, RNA splicing and translation. In eukaryotic systems, a number of zinc ribbon or zinc finger proteins are known to be involved in post-transcriptional processing and modification. For example, SRD1 is a yeast pre-rRNA processing protein and the translation initiation factor eIF-2$\beta$, containing a $C_2$-$C_2$ zinc finger domain, and have bacterial homologues (Hess et al., 1994).

One possible function of ET1 could be an involvement of the zinc ribbon domain in protein-protein interaction, like in CK2$\beta$. If ET1 does happen to possess a CK2$\beta$ type structure, and undergo dimerisation to constitute a subunit of a multimeric regulatory protein, then it could constitute a part of a CK2 like protein kinase or some other similar regulatory protein, whose catalytic domain, as also in CK2, is present in another subunit. Similar to CK2$\beta$ subunit, ET1 also possesses $\alpha$ helices at both its ends, which cluster together in case of CK2$\beta$. However, in ET1 the $\alpha$-helical region is much smaller. The indications of the presence of CK2-like and other protein kinases in chloroplasts is available from a number of studies. The PEP polymerase from mustard (*Sinapis alba*) has been found to be associated with a Ser/Thr protein kinase, called plastid transcription kinase (PTK) (Baginsky et al., 1999). Reversible phosphorylation of the photosystem II protein components has been observed associated with a CK2 like Ser/Thr protein kinase, which also indicates involvement in signal transduction (Testi et al., 1996).

## 4.3. Outlook

The preliminary analyses of the *et1* mutant phenotype, carried out earlier, had indicated that the mutation affects plastid development in the kernels (amyloplasts) and seedlings (chloroplasts). The molecular characterisation of the *et1* gene carried out in this study showed that ET1 protein is plastid localised and probably plays a regulatory role in plastid differentiation, which is crucial during early phases of endosperm as well as seedling development. Structural characterisation of the ET1 protein showed that its possesses a central domain with a zinc ribbon like motif, which could either be involved in nucleic acid binding and cleavage, or it could be playing a crucial role in protein dimerisation. Variations in function of zinc ribbon domains are known from different proteins and depend on their topological features, including the physiochemical nature of the residues present in the interacting sub-domain. Therefore, based on the present knowledge on the flexible structure of zinc ribbon motifs, no definite function can be assigned to the ET1 zinc ribbon like domain. In order to characterise the ET1 function, further experiments need to be performed to clarify its role as a regulatory protein in plastid development.

Since ET1 is a small protein, it is very likely that it is either a part of a multi-subunit protein complex or it works as a regulatory factor interacting temporarily with a protein or protein complex in the plastids to perform its function. One possibility to analyse its interaction with other proteins would be to use the ET1 mature protein as a bait in the yeast two-hybrid system against an appropriate cDNA library, like that from developing kernels or seedling leaves, and identify the gene encoding the protein interacting with ET1. In order to check if the ET1 protein function involves direct interaction with the plastid chromosomal DNA, gel-retardation assays with the radiolabelled ET1 protein and the plastid DNA could be performed. Another interesting possibility would be to use the *in vitro* translated radioactive ET1 preprotein in an *in organello* experiment to analyse ET1 interaction with other plastid proteins. Instead of radioactive labelling, an antibody against ET1 protein could possibly also be used. After performing chloroplast import and incubation of the intact chloroplasts with the ET1 protein, the organeller proteins could be isolated and separated through ultracentrifugation procedures, and analysed on a native gel. On the other hand, isolated plastid protein-extracts could also be analysed for interaction with the ET1 protein, either through the use of labelled ET1 protein or an antibody

against it. Comparison of run-on transcription assays with plastids from wild type and *et1* mutants could also reveal the influence of ET1 on plastid transcription and its possible role in plastids. *In situ* hybridisation experiments, using labelled ET1 antisense transcripts as well as with an antibody against it, would also help further characterise ET1 expression and function.

The molecular characterisation of the *et1* homologue, *zmzr1*, through southern analyses revealed a very similar structure of the gene in both *et1-Ref* and LC. Since so far, the *zmzr1* gene has only been cloned from the *et1-Ref* line, a molecular characterisation of the gene in the inbred line, LC could also be carried out. The expression analysis of *zmzr1* through Northern experiments with total RNA did not reveal any expression in all the tissues examined. However, a weak *zmzr1* expression in the kernels was indicated from the *et1-R* developing kernel cDNA library screening. Moreover, a maize cDNA clone from an EST databank, which is almost identical to *zmzr1* cDNA, was found to express in the tassel primordia. Performing Northern experiments using mRNA samples could be helpful, in case this transcript possesses a short half-life. *In silico* sub-cellular localisation analyses indicate a plastid localisation of ZMZR1. A chloroplast import experiment with the *in vitro* synthesised protein could be carried out to confirm this observation. Further on, similar experiments could be designed for the ZMZR1 protein as for ET1, to uncover its role in plastid development.

# 5. Summary

*Etched1* (*et1*) is a pleiotropic mutation of maize, which visibly affects plastid development in the kernels and seedlings. The mature kernels show a large number of indentations and crevices on the endosperm surface. The absence of starch grains, produced normally in the amyloplasts of the starchy-endosperm cells, around the crevices causes the “etched” appearance. In more extreme phenotypic variations, the kernels appear shrivelled and in some cases also fail to germinate. Phenotypic analysis carried out in this study on 20 DAP kernels showed that the starchless cells were also present in the central region of the endosperm radiating out up to the periphery and were, then, observed superficially as surface indentations in mature kernels. The seedlings were virescent, as the leaves remained white to pale green in the first 10 to 15 days after germination, due to delayed chloroplast development. The *et1* gene was chosen for further genetic and molecular analysis, as it affected two distinct developmental stages of the plant, which were also critical to the plant survival. Moreover, molecular examinations of the regulatory mechanisms involved in endosperm development of maize have long been of interest due to its economic importance.

Molecular observations indicated that the absence of the *et1* gene product was the cause of the mutant phenotype. The *et1-R* reference allele revealed a deletion of the complete *et1* gene from its locus (Ahrend, 1998), and in the Northern analyses with the *Mu*-transposon induced *et1* mutants, no *et1* transcript was detectable (da Costa e Silva et al., 2001). Therefore, this study focused on the analysis of the structure and function of the *et1* gene as well as its gene product, which was evaluated using the wild type inbred line, LC. Northern analyses revealed strong *et1* expression at all stages of development in the leaves, and from 6 DAP onwards in the kernels. Other tissues with good *et1* expression included the tassels, stem and coleoptile. Otherwise, a weak basal *et1* expression was observed in all the other tissues analysed.

Sequence analysis of the *et1* gene as well as its cDNA clone revealed that it codes for a small protein, about 163 amino acids long. However, chloroplast import experiments revealed that the protein is plastid localised. Therefore, based on the

sequence analysis as well as *in silico* analyses on the pre-protein, the mature protein, after cleavage of the transit sequence, was predicted to be only 99 amino acids long. Thus, it represented a comparatively small protein. Structural analyses of the protein and its comparison to structurally characterised proteins indicated that the ET1 protein probably contained a central zinc ribbon like motif, bordered on both sides by helical sub-domains. Zinc ribbon domains are characterised by the presence of a three-stranded antiparallel β-sheet, held together by a zinc ion, bound to four amino acid ligands in the domain. It is also characterised by the presence of other topological features such as the physiochemical nature of the amino acids, required for interaction with other molecules, which could either be nucleic acids or homologous protein domains, involved in dimerisation. Based on its comparison to the different functional as well as structural variations of the zinc ribbon domain from different proteins, all possible functions were attributable to the ET1 protein. On the other hand, it was possible that it represented another unique zinc ribbon domain. This aspect needs to be further researched. However, present knowledge about the plastid genetic system, as well as the various mechanisms controlling plastid gene expression, indicate that ET1 probably regulates plastid development at the transcriptional or post-transcriptional level, which includes translation.

Another analysis of the *et1-R* line (Ahrend, 1998) showed genomic clones hybridising with the *et1* cDNA, which did not correspond to the *et1* locus. Sequence analysis carried out in this study revealed that they represented another gene, *zmzr1*, with very high homology to *et1*. Southern analyses on LC indicated that *zmzr1* had a very similar structure in LC. Comparison of the upstream promoter region of *zmzr1* and *et1* showed that this was completely different between the two loci – a feature often observed among paralogous genes. Analysis of their promoter regions indicated the presence of different cis elements so that they were probably differently regulated. Northern analyses with *zmzr1* probes using total RNA blots did not reveal any expression. However, a cDNA clone could be amplified from an *et1-R* developing kernel cDNA library. Moreover, an EST clone expressed in tassels was also identified from online databanks. Comparison of the ZMZR1 protein to ET1 showed high homology between the two proteins. Most of the variation was present in the plastid transit sequence region. The mature proteins of both revealed similar secondary structure.

Comparison of the proteins to online EST databanks revealed a large number of plant homologues, with an additional paralogue identified from maize. Alignment of the protein sequences from all the ESTs indicated that it represented a third homologous transcript/gene in maize. Based on the homology among all the plant homologues, two groups of sequences could be identified from the alignment. In one group, sequences more homologous to ET1 and ZMZR1 were present, whereas the other contained sequences more homologous to the new maize paralogue of ET1. More importantly, two EST sequences each were obtained from different plant sources, including *Arabidopsis*, one of which showed higher homology to the ET1/ZMZR1 group and the other to the third maize EST. This indicated that these two groups of paralogous sequences probably arose very early, before the divergence of the dicots and monocots. On the other hand, ZMZR1, based on its high homology to the first group, represented a more recent duplication event – very probably within the maize genome, as the three different homologues were only found in maize.

The results presented in this thesis provide further evidence that it is the cloned gene analysed in this study, which, on mutation, is responsible for the *etched1* phenotype. Although, it cannot be ruled out that other independent or epistatic factors might also be influencing the *et1* phenotype, causing the variations observed among different maize lines. The putative *Et1* gene might be involved in regulating early plastid development. However, further analyses need to be carried out to determine the type of interactions in which the ET1 protein might be involved to bring about the regulatory affects in the plastids.

# 6. References

Ahn, S., Tanksley, S.D. (1993). Comparative Linkage Maps Of The Rice And Maize Genomes. *Proc. Natl. Acad. Sci. USA* **90 (17)**, 7980-84.

Ahrend, M. (1998). Molekulare Analyse der *etched1-Reference* Mutante von *Zea mays* L. Diploma thesis (Biochemie/Molekular Biologie), Hamburg Universität.

Allison, L.A., Simon, L.D., Maliga, P. (1996). Deletion of rpoB reveals a second distinct transcription sytem in plastids of higher plants. *EMBO J.* **15**, 2802-09.

Altschul, S. F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol*. **215**, 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Annweiler, A., Hipskind, A., Wirth, T. (1991). A strategy for efficient *in vitro* translation of cDNAs using the rabbit β-globin leader sequence. *Nucleic Acids Res.* **19**, 3750.

Appel R.D., Bairoch A., Hochstrasser D.F. (1994). A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem. Sci.* **19**, 258-260. http://www.expasy.ch/

Awrey, D.E., Shimasaki, N., Koth, C., Weilbaecher, R., Olmstedi, V., Kazanisi, S., Shani, X., Arellano, J., Arrowsmithi, C.H., Kane, C.M., Edwards, A.M. (1998). Yeast Transcript Elongation Factor (TFIIS), Structure and Function II: RNA POLYMERASE BINDING, TRANSCRIPT CLEAVAGE, AND READ-THROUGH. *J. Biol. Chem*. **273 (35)**, 22595–605.

Baginsky, S., Tiller, K., Pfannschmidt, T., Link, G. (1999). PTK, the chloroplast RNA polymerase-associated protein kinase from mustard (*Sinapsis alba*), mediates redox control of plastid in vitro transcription. *Plant Mol. Biol.* **39 (5)**, 1013-23.

Bairoch, A., Apweiler, R. (1998). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res.* **26 (1)**, 38-42. http://www.expasy.ch/

Bartling D., Clausmeyer, S., Oelmüller, R., Herrman, R.G. (1990). Towards epitope models for chloroplast transit sequences. *Bot. Mag.* **2**, 119-144.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn R.D., Sonnhammer, E.L.L. (1999). Pfam 3.1: 1313 multiple alignments match the majority of proteins. *Nucl. Acids Res.* **27**, 260-262.

Becraft, P.W., Asuncion-Crabb, Y. (2000). Positional cues specify and maintain aleurone cell fate in maize endosperm development. *Development* **127**, 4039-48.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., Wheeler, D.L. (2000). GenBank. *Nucleic Acids Res.* **28 (1)**, 15-18. http://www.ncbi.nlm.nih.gov/

Berg, J.M., Shi, Y. (1996). The Galvanisation of Biology: A Growing Appreciation for the Roles of Zinc. *Science* **271**, 1081-85.

Bernhardt, J., Stich, K., Schwarz-Sommer, Z., Saedler, H., Wienand, U. (1998). Molecular analysis of a second functional *A1* gene (dihydroflavonol-4-reductase) in *Zea mays. Plant J.* **14 (4)**, 483 – 488.

Bligny, M., Courtois, F., Thaminy, S., Chang, C.-C., Lagrange, T., Baruah-Wolff, J., Stern, D., Lerbs-Mache, S. (2000). Regulation of plastid rDNA transcription by interaction of CDF2 with two different RNA polymerases. *EMBO J.* **19 (8)**, 1851-60.

Bochkarev, A., Pfeutzner, R.A., Edwards, A.M., Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of the replication protein A bound to DNA. *Nature* **385**, 176-181.

Booth, V., Koth, C.M., Edwards, A.M., Arrowsmith, C.H. (2000). Strucuture of a Conserved Domain Common to Transcription Factors TFIIS, Elongin A, and CRSP70. *J. Biol. Chem.* **275 (40)**, 31266-68.

Bucher, P., Bairoch, A. (1994). A generalised profile syntex for biomolecular sequence motifs and its function in automatic sequence interpretation. In: ISMB-94; *Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology*. Altman, R., Brutlag, D., Karp, P., Lathrop, R., Searls, D. Eds., AAAI Press, Menlo park, pp 53-61. MEDLINE: 96039003.

Bycroft, M., Hubbard, J.P., Proctor, M., Freund, S.M.V., Murzin, A.G. (1997). The Solution Structure of the S1 RNA Binding Domain: A member of an Ancient Nucleic Acid-Binding Fold. *Cell* **88**, 235-242.

Camara, B., Hugueney, P., Bouvier, F., Kuntz, M., Moneger, R. (1995). Biochemistry and Molecular Biology of Chromoplast Development. *Int. Rev. Cytol.* **163**, 175-247.

Carde, J.P. (1984). Leucoplasts: a distinct kind of organelles lacking typical 70S ribosomes and free thylakoids. *Eur. J. Cell Biol.* **34 (1)**, 18-26.

Chang, C.-C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S., Stern, D.B. (1999). Functional Analysis of Two Maize cDNAs Encoding T7-like RNA Polymerases. *Plant Cell* **11**, 911-926.

Chantalat, L., Leroy, D., Filhol, O., Nueda, A., Benitez, M.J., Chambaz, E.M., Cochet, C., Dideberg, O. (1999). Crystal structure of the human protein kinase CK2 regulatory subunit reveals its zinc finger-mediated dimerization. *EMBO J.* **18 (11)**, 2930–40.

Chédin, S., Riva, M., Schultz, P., Sentenac, A., Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev.* **12**, 3857–71.

Chen, H.T., Legault, P., Glushka, J., Omichinski, J.G., Scott, R.A. (2000). Structure of a (Cys3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. *Protein Sci.* **9 (9)**, 1743-52.

Chomecynski, P., Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium Thiocyanate-Phenol-Chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Chomet, P.S. (1994). Transposon tagging with Mutator. In: *The Maize Handbook*, Freeling, M., Walbot, V. Eds. Springer-Verlag, New York, Inc., pp. 243-249.

Chou P.Y., Fasman G.D. (1978). Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**, 251-276.

Clausmeyer, S., Klosgen, R.B., Herrmann, R.G. (1993). Protein import into chloroplasts. The hydrophilic lumenal proteins exhibit unexpected import and sorting specificities in spite of structurally conserved transit peptides. *J. Biol. Chem.* **268 (19)**, 13869-76. PMID: 8314754.

Clayton, W.D. (1987). Andropogoneae. In: *Grass Systematics and Evolution.* Sonderstrom, T.R., Hilu, K.W., Campbell, C.S., Barkworth, M.E., eds., Smithsonian Institution press, Washington, DC, pp. 307-309.

Cline, K., Ralph, H. (1996). Import and routing of nucleus encoded chloroplast proteins. *Annu. Rev. Cell Dev. Biol*. **12**, 1-26.

Coe, Jr. E.H., Neuffer, M.G., Hoisington, D.A. (1988). The Genetics of Corn. In: *Corn and Corn Improvement* – Agronomy Monograph no. **18**, 3rd. Ed. Sprague, G.F., Dudley, J.W. eds., ASA, CSSA, SSSA Publishers. Madison. pp. 1-31.

Connolly, K.M., Ilangovan, U., Wojciack, J.M., Iwahara, M., Clubb, R.T. (2000). Major Groove Recognition by Three-stranded β-Sheets: Affinity Determinants and Conserved Structural Features. *J. Mol. Biol.* **300**, 841-856.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16 (22)**, 10881-90. http://prodes.toulouse.inra.fr/prodom/doc/prodom.html

Corpet, F., Gouzy, J., Kahn, D. (1998). The ProDom database of protein domain families. *Nucl. Acids Res.* **26**, 323-326. http://prodes.toulouse.inra.fr/prodom/doc/prodom.html

Corpet, F., Servant, F., Gouzy, J., Kahn, D. (2000). ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res.* **28**, 267-269. http://prodes.toulouse.inra.fr/prodom/doc/prodom.html

da Costa e Silva, O., Garg, P., Wassmann, M., Lorbieke, R., Lauert, P., Scanlon, M., Hsia, A.-P., Schnable, P.S., Wienand, U. (2002). The *Etched1* gene product of *Zea mays* L. encodes a zinc-finger protein with homology to the eukaryotic transcription elongation factor TFIIS. (*manuscript under preparation*).

Dalbey, R.E., Kahn, A., (2000). Evolutionarily Related Insertion Pathways of Bacterial, Mitochondrial, and Thylakoid Membrane Proteins. *Annu. Rev. Cell Biol.* **16**, 51-87.

Davis, G. L., McMullen, M. D., Baysdorfer, C., Musket, T., Grant, D., Staebell, M., Xu, G., Polacco, M., Koster, L., Melia-Hancock, S., Houchins, K., Chao, S., Coe Jr., E. H. (1999). A Maize Map Standard With Sequenced Core Markers, Grass Genome Reference Points and 932 Expressed Sequence Tagged Sites (ESTs) in a 1736-Locus Map. *Genetics* **152**, 1137–1172.

de Boer, A.D., Weisbeek, P.J. (1991). Chloroplast protein topogenesis: import, sorting and assembly. *Biochem. Biophys. Acta* **1071**, 221-253.

Deleage, G., Roux, B. (1987). An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng.* **1 (4)**, 289-94. PMID: 3508279.

Dellaporta, S.L., Wood, J., Hicks, J.B. (1983). A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep.* **1 (4)**, 19-22.

Dietrich, C.R., Mahapatabandige, P.A., Meely, R.B., Nikolau, B.J., Schnable, P.S. (2001). Beta-keto acyl reductase activity is essential for maize development. Talk T22 in 43rd *Annual Maize Genetics Conference*, 14-18 March, 2001. http://www.agron.missouri.edu/Coop/Conf/2001.html

Dodd, I.B., Egan, J.B. (1990). Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res.* **18**, 5019-26.

Doebley, J.F., Goodman, M.M., Stuber, C.W. (1987). Patterns of Isozyme Variation Between Maize And Mexican Annual Teosinte. *Econ. Bot.* **41**, 234-246.

Dubreuil, P., Charcosset, A. (1999). Relationships among maize inbred lines and populations from European and North-American origins as estimated using RFLP markers. *Theor. Appl. Genet.* **99**, 473-480.

Dunsmuir, P., Bond, D., Lee, K., Gidoni, D., Townsend, J. (1988). C1: Stability of introduced genes and stability in expression. In: *Plant Biology Manual.* Kluwer Academic Publishers, Dordrecht. pp. 1-17.

Emanuelsson, O., Nielsen, H., von Heijne, G. (1999). ChloroP, a neural network based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**, 978-984. http://www.cbs.dtu.dk/services/ChloroP/

Emanuelsson, O., Nielsen, H., Brunak S., von Heijne, G. (2000). Predicting subcellular localisation of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005-16. http://www.cbs.dtu.dk/services/TargetP/

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L., Gaut, B.S. (1998). Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. USA* **95**, 4441-46.

Franken, P, Niesbach-Klösgen, U., Weydemann, U., Maréchal-Drouard, L., Saedler, H., Wienand, U. (1991). The duplicated chalcone synthase genes *C2* and *Whp* (*white pollen*) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene *in*. *EMBO J.* **10 (9)**, 2605 – 2612.

Frey, M., Stettner, C., Gierl, A. (1998). A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). *Plant J.* **13**, 717-721.

Frenzel, K. (2000). Characterisierung von *ZmKCS-1* und *ZmKCS-2*; zwei mögliche β-Ketoacyl-CoA-Synthasegene aus dem Wachsstoffwechsel von Mais (*Zea mays L.*). Ph.D. thesis, Hamburg Universität.

Frishman, D. & Argos, P. (1995). Knowledge-based secondary structure assignment. *Proteins: structure, function and genetics* **23**, 566-579. http://www.embl-heidelberg.de/cgi/predator_serv.pl

Fulgosi, H., Soll, J. (2001). A gateway to chloroplasts – protein translocation and beyond. *J. Plant Physiol.* **158**, 273-284.

Galinat, W.C. (1988). The Origin of Corn. In: *Corn and Corn Improvement* – Agronomy Monograph no. **18**, 3rd. Ed. Sprague, G.F., Dudley, J.W. eds., ASA, CSSA, SSSA Publishers. Madison. pp. 1-31.

Garnier,J., Gibrat, J.F., Robson, B. (1996). GOR secondary structure prediction method version IV. In: *Methods in Enzymology*. Doolittle R.F. eds., vol **266**, 540-553. http://www.expasy.ch/tools/#secondary

Garnier, J., Osguthorpe, D.J., Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120 (1)**, 97-120. http://www.expasy.ch/tools/#secondary

Gaut, B.S., Clegg, M.T. (1993). Molecular Evolution of the Adh1 Locus in the Genus *Zea*. *Proc. Natl. Acad. Sci. USA* **90**, 5095-99.

Gaut, B.S., Doebley, J.F. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. USA* **94**, 6809-14.

Gaut, B.S., Le Thierry d'Ennequin, M., Peek, A.S., Sawkins, M.C. (2000). Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* **97(13)**, 7008-15.

Geourjon, C., Deleage, G. (1994). SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* **7 (2)**, 157-164. http://www.expasy.ch/tools/#secondary

Geourjon, C., Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11 (6)**, 681-684. http://www.expasy.ch/tools/#secondary

Gibrat, J.F., Garnier, J., Robson, B. (1987). Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.* **198 (3)**, 425-443.
http://www.expasy.ch/tools/#secondary

Gish, W., States, D.J. (1993). Identification of protein coding regions by database similarity search. *Nature Genet.* **3**, 266-272.

Goloubinoff, P., Paabo, S., Wison, A.C. (1993). Evolution of Maize Inferred from Sequence Diversity of an *Adh2* Gene Segment from Archaeological Specimens. *Proc. Natl. Acad. Sci. USA* **90**, 1997-2001.

Goodman, M.M., Stuber, C.W., Newton, K., Weissinger, H.H. (1980). Linkage Relationships Of 19 Enzyme Loci In Maize. *Genetics* **96 (3)**, 697-710.

Guermeur, Y. (1997). Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteines. Ph.D. Thesis.
http://npsa-pbil.ibcp.fr/NPSA/npsa_references.html#hnn

Gupta, R., Singh, R. (1996). Fatty acid synthesis by isolated leucoplasts from developing Brassica seeds: role of glycolytic intermediates as the source of carbon and energy. *Indian J. Biochem. Biophys*. **33 (6)**, 478-483.

Hajdukiewicz, P.T.J., Allison, L.A., Maliga, P. (1997). The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* **16 (13)**, 4041-48.

Hausner, W., Lange, U., Musfeldt, M. (2000). Transcription Factor S, a Cleavage Induction Factor of the Archael RNA Polymerase. *J. Biol. Chem.* **275 (17)**, 12393-99.

Hedtke, B., Börner, T., Weihe, A. (2000). One RNA polymerase serving two genomes. *EMBO reports* **1 (5)**, 435-440.

Helentjaris, T., Weber, D., Wright, S. (1988). Identification Of The Genomic Locations Of Duplicate Nucleotide-Sequences In Maize By Analysis Of Restriction Fragment Length Polymorphisms. *Genetics*, **118 (2),** 353-363.

Helentjaris, T. (1995). Atlas of duplicated sequences. *Maize Genet. Coop. Newsl.* **69**, 67-81.

Hemming, S.A., Edwards, A.M. (2000). Yeast RNA Polymerase II Subunit RPB9: MAPPING OF DOMAINS REQUIRED FOR TRANSCRIPTION ELONGATION. *J. Biol. Chem.* **275 (4)**, 2288–94.

Hemming, S.A., Jansma, D.B., Macgregor, P.F., Goryachev, A., Friesen, J.D., Edwards, A.M. (2000). RNA Polymerase II Subunit Rpb9 Regulates Transcription Elongation *in Vivo*. *J. Biol. Chem.* **275 (45)**, 35506–11.

Hess, D. (1999).Die unselbstständige vegetative phase. In *Pflanzenphysiologie – Molekulare und biochemische Grundlagen von Stoffwechsel und Entwicklung der Pflanzen*. 10. Auflage. UTB Für Wissenschaft, Verlag Eugen Ulmer Stuttgart. pp. 496-499.

Hess, S.M., Stanford, D.R., Hooper, A.K. (1994). SRD1, a *S. cereviciae* gene affecting pre-rRNA processing contains a $C_2/C_2$ zinc finger motif. *Nucl. Acids Res.* **22 (7)**, 1265-71.

Hilton, H., Gaut, B.S. (1998). Speciation and Domestication in Maize and Its Wild Relatives: Evidence From the *Globulin-1* Gene. *Genetics* **150**, 863–872.

Hofman, K., Bucher, P., Falquet, L., Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucl. Acids Res.* **27**, 215-219. MEDLINE : 9847184.

Hofmann, K., Stoffel, W. (1993). TM base - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**,166.
http://www.ch.embnet.org/software/TMPRED_form.html

Howe, C.J. (1996). RNA Polymerases and plastid evolution. *Trends Plant. Sci.* **1 (10)**, 323-324.

Hu, J., Bogorad, L. (1990). Maize chloroplast RNA polymerase: The 180-, 120-, 38-kilodalton polypeptides are encoded in chloroplast genes. *Proc. Natl. Acad. Sci. USA* **87**, 1531-35.

Isono, K., Shimizu, M., Yoshimoto, K., Niwa, Y., Satoh, K., Yokota, A., Kobayashi, H. (1997). Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of $\sigma^{70}$ factors of bacterial RNA polymerases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **94**, 14948-53.

Jeon, C., Yoon, H., Agarwal, K. (1994). The transcription factor TFIIS zinc ribbon dipeptide Asp-Glu is critical for stimulation of elongation and RNA cleavage by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **91 (19)**, 9106-10.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403-405.

Joyard, J., Teyssier, E., Miège, C., Berny-Seigneurin, D., Maréchal, E., Block, M.A., Dorne, A.-J., Rolland, N., Ajlani, G., Douce, R. (1998). The Biochemical Machinery of Plastid Envelope Membranes. *Plant Physiol.* **118**, 715–723.

Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.

Kiesselbach, T.A. (1949). *The Structure and Reproduction of Corn*. University of Nebraska Press, Lincoln and London.

Klösgen, R.B., Saedler, H., Weil, J.H. (1989). The amyloplast-targeting transit peptide of the waxy protein of maize also mediates protein transport in vitro into chloroplasts. *Mol. Gen. Genet.* **217 (1)**, 155-161. PMID: 2475752.

Klösgen, R.B., Weil, J.H. (1991). Subcellular location and expression level of a chimeric protein consisting of the maize waxy transit peptide and the beta-glucuronidase of *Escherichia coli* in transgenic potato plants. *Mol. Gen. Genet.* **225 (2)**, 297-304. PMID: 2005871.

Kneller, D.G., Cohen, F.E., Langridge, R. (1990). Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. *J. Mol. Biol.* **214**, 171-182. http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html

Kochar, S.L. (1981). Cereal Crops. In: *Economic Botany in the Tropics*. Macmillan India Ltd. Madras. pp. 88-104.

Kranz, E., von Wiegen, P., Quader, H., Lörz, H. (1998). Endosperm Development after Fusion of Isolated, Single Maize Sperm and Central Cells in Vitro. *Plant Cell* **10**, 511–524.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305 (3)**, 567-580.
http://www.cbs.dtu.dk/services/TMHMM-2.0/

Krupinska, K., Falk, J. (1994). Changes in RNA-Polymerase Activity during Biogenesis, Maturation and Senescence of Barley Chloroplasts. Comparitive Analysis of Transcripts Synthesized Either in Run-on Assays or by Transcriptionsally Active Chromosomes. *J. Plant Physiol.* **143**, 298-305.

Laemmli, U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lauert, P., da Costa e Silva, O., Wienand, U. (1999). The use of Thermomixer Comfort for AIMS. *Eppendorf Appl.* N.**15**.

Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A., Wright, P.E. (1989). Three-Dimentional Solution Structure of a Single Zinc Finger DNA-Binding Domain. *Science* **245**, 635-637.

Levitt M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry* **17 (20)**, 4277-85. PMID: 708713.

Lifson, S., Sander, C. (1979). Antiparallel and parallel beta-strands differ in amino acid residue preferences. *Nature* **282 (5734)**, 109-111. PMID: 503185

Lopes, M.A., Larkins, B.A. (1993). Endosperm Origin, Development, and Function. Plant Cell **5**, 1383-99.

Lubberstedt, T, Melchinger, A.E., Dussle, C., Vuylsteke, M., Kuiper, M. (2000). Relationships among early European maize inbreds: IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. *Crop. Sci.* **40,** 783-791.

Ljubesic, N., Wrischer, M., Devide, Z. (1991). Chromoplasts--the last stages in plastid development. *Int. J. Dev. Biol.* **35 (3)**, 251-258.

Lupas, A. (1996). Prediction and Analysis of Coiled-Coil Structures. *Methods in Enzymology* **266**, 513-525.

Maliga, P. (1998). Two plastid RNA polymerases of higher plants: an evolutionary story. *Trends Plant Sci.* **3 (1)**, 4-6.

Markov, D., Naryshkina, T., Mustaev, A., Severinov, K. (1999). A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. *Genes Dev.* **13**, 2439–48.

May, T., Soll, J. (2000). 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell* **12**, 53-63.

McClelland, J.L., Rumelhart, D.E. (1988). *Explorations in Parallel Distributed Processing*. Vol. **3**. MIT Press, Cambridge MA, pp. 318-362.
http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html

McClintock, B. (1978). Development of maize endosperm as revealed by clones. In: *The Clonal Basis of Development*. Eds. Subtelny, S., Sussex, I.M. Academic press, Inc. NY. pp. 217-237.

McClintock, B. (1930). A Cytological Demonstration of the Location of an Interchange Between Two Non-Homologous Chromosomes of Zea Mays. *Proc. Natl. Acad. Sci. USA* **16**, 791-796.

McMillin, D.E., Scandalios, J.G. (1980). Duplicated Cytosolic Malate-Dehydrogenase Genes In Zea-Mays. *Proc. Natl. Acad. Sci., USA -Biological Sciences* **77 (8)**, 4866-4870.

Meschi, T., Iwabuchi, M. (1995). Plant Transcription Factors. *Plant cell Phsyiol.* **36 (8)**, 1405-20.

Michl, D., Robinson, C., Shackleton, J.B., Herrmann, R.G., Klosgen, R.B. (1994). Targeting of proteins to the thylakoids by bipartite presequences: CFoII is imported by a novel, third pathway. *EMBO J.* **13 (6)**,1310-17. PMID: 8137815.

Motto, M., Bianchi, A., Salamini, F. (1999). Recent Developments in the Genetic and Molecular Biology of Maize Mutants. In: *Maize Genetics and Breeding in the 20th Century*. Peterson, P.A., Bianchi, A. eds. World Scientific.

Moeller, D., Schaal, B. (1999). Genetic relationships among native American maize accessions of the great plains assessed by RAPDs. *Theor. Appl. Genet.* **99**, 1061-1067.

Nakai, K., Kanehisa, M. (1992). A knowledge base for predicting protein localisation sites in eukaryotic cells. *Genomics* **14**, 897-911. http://psort.nibb.ac.jp/

Nakanishi, T., Shimoaraiso, M., Kubo, T., and Natori, S. (1995). Structure-Function Relationship of Yeast S-II in Terms of Stimualation of RNA Polymerase II, Arrest Relief, and Suppression of 6-Azauracil Sensitivity. *J. Biol. Chem.* **270**, 8991-95.

Neuffer, M.G., Coe, E.H., Wessler, S.R. (1997). Mutants of Maize. Cold Spring Harbour Lab. Press.

Nicholas, K.B., Nicholas H.B. Jr. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed at: http://www.psc.edu/biomed/genedoc/

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1-6. http://www.cbs.dtu.dk.services/TargetP/

Olmsted, V.K., Awrey, D.E., Koth, C., Shan, X., Morin, P.E., Kazanis, S., Edwards, A.M., Arrowsmith, C.H. (1998). Yeast Transcript Elongation Factor (TFIIS), Structure and Function I: NMR STRUCTURAL ANALYSIS OF THE MINIMAL TRANSCRIPTIONALLY ACTIVE REGION. *J. Biol. Chem.* **273 (35)**, 22589-94.

Olsen, O-A., Linnestad, C., Nichols, S.E. (1999). Developmental Biology of the cereal endosperm. *Trends Plant Sci.* **4**, 253-257.

Opshal-Ferstad, H.G., Le Deunff, E., Dumas, C., Rogowsky, P.M. (1997). *ZmESR*, a novel endosperm specific gene expressed in a restricted region around the maize embryo. *Plant J.* **12**, 235-246.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.

Pereira, M.G., Lee, M., Bramelcox, P., Woodman, W., Doebley, J., Whitkus, R. (1994). Construction of an RFLP Map in Sorghum and Comparative Mapping in Maize. *Genome* **37 (2)**, 236-243.

Peterson, P.A. (1999). Development of Maize Genetics and Breeding. In: *Maize Genetics and Breeding in the 20$^{th}$ Century*. Peterson, P.A., Bianchi, A. eds. World Scientific

Pfannschmidt, T., Link, G. (1994). Separation of two classes od plastid DNA-dependent RNA polymerase that are differentially expressed in mustard (*Sinapsis alba* L.) seedlings. *Plant Mol. Biol.* **25**, 69-81.

Powell, W., Bartholomew, B., Reines, D. (1996). Elongation factor SII Contacts the 3´-End of RNA Polymerase II Elongation Complex. *J. Biol. Chem.* **271 (37)**, 22301-04.

Qian, X., Jeon, C., Yoon, H.-S., Weiss, M.A. (1993). Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. *Nature* **365**, 277-279.

Ramesh, S.V., Santa Kumari, A.S., Reddy, A.R. (1984). Photoacoustic and physiological studies on a maize mutant: a delay in leaf pigment synthesis in virescent seedlings. *Biochem. Intl.* **9**, 121-128.

Rhoades, M.M. (1961). Distribution ratios due to semi-lethality of etched kernels. *Maize Genet. Coop. News Lett.* **35**, 48.

Robertson, D.S. (1978). Characterisation of a *mutator* system in maize. *Mut. Res.* **51**, 21-28.

Rochaix, J.-D. (1996). Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* **35**, 327-341.

Rombauts,S., Déhais, P., Van Montagu, M., Rouzé, P. (1999). Plant CARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.* **27 (1)**, 295-296. http://sphinx.rug.ac.be:8080/PlantCARE/cgi/Menu.html

Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. *Meth. Enzym.* **226**, 525-539. http://cubic.bioc.columbia.edu/predictprotein/

Rost, B., Fariselli, P., Casadio, R. (1996). Topology prediction for helical transmembrane proteins at 86% accuracy. *Prot. Science* **7**, 1704-18.

Rost, B., Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584-599. http://cubic.bioc.columbia.edu/predictprotein

Rost, B., Sander, C. (1994a). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55-77.

Rost, B., Sander, C. (1994b). Conservation and prediction of solvent accessibility in protein families. *Proteins* **20**, 216-226. http://cubic.bioc.columbia.edu/predictprotein

Rost, B., Sander, C., Schneider, R. (1994). Redefining the goals of protein secondary structure prediction. *J. Mol. Biol.* **235**, 13-26.

Rybicki E.P. (1996). PCR Primer Design and Reaction Optimisation. In: *Molecular Biology Techniques Manual, 3rd ed.* by Coyne, V. E., James, M. N., Ried, S. J., Rybicki, E. P. At website: http://www.uct.ac.za/microbiology/pcroptim.htm

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). *Molecular cloning: A Laboratory Manual. 2nd Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Sangeetha, H.G., Ramesh, S.V., Jayaram, Ch., Reddy, A.R. (1986). Studies on chlorophyll and protein accumulation in thylakoids and chloroplast development in virescent seedlings of maize mutant. *Biochem. Intl.* **12**, 181-188.

Sangeetha, H.G., Reddy, A.R. (1991). Genetical and biochemical analysis of the *etched* (*et*) mutant of *Zea mays* L. *Maydica* **36**, 343-354.

Sanger, F., Nickler, S, Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl. Acad. Sci.* **74**, 5463-67.

Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., Altschul, S.F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29 (14)**, 2994-3005.

Scanlon, M.J., James, M.G., Stinard, P.S., Myers, A.M, Robertson, D.S. (1994). Characterisation of 10 new mutations of the maize *Etched-1* locus. *Maydica* **39**, 301-308.

Schwartz, R.M., Dayhoff, M.O. (1978). Origin of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts. *Science* **199**, 395-403.

Senior, M.L., Murphy, J.P., Goodman, M.M., Stuber, C.W. (1998). Utility of SSRs for Determining Genetic Similarities and Relationships in Maize Using an Agarose Gel System. *Crop Sci.* **38**, 1088–98.

Serino, G., Maliga, P. (1998). RNA Polymerase Subunits Encoded by the Plastid *rpo* Genes Are Not Shared with the Nucleus-Encoded Plastid Enzyme. *Plant Physiol.* **117**, 1165-70.

Shattuck-Eidens, D.M., Bell, R.N., Neuhausen, S.L., Helentjaris, T. (1990). DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. *Genetics*, Sep., **126 (1)**, 207-217.

Shilatifard, A. (1998). Factors regulating the transcriptional elongation activity of RNA polymerase II. *FASEB J.* **12**, 1437-46.

Silhavy, D., Maliga, P. (1998). Plastid promoter utilization in rice embryonic cell culture. *Curr. Genet.* **34**, 67-70.

Simpson, G.G., Filipowicz, W. (1996). Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. *Plant Mol. Biol.* **32**, 1-41.

Small, I. *et al* (2000) Predotar (version 0.5): A prediction service for identifying putative mitochondrial and plastid targeting sequences. http://www.inra.fr/Internet/Produits/Predotar/

Smith, J.S.C., Goodman, M.M., Kato, T.A. (1982). Variation within teosinte 2. Numerical analysis of knob data. *Econ Bot.* **36**, 100-112.

Soll, J. and Tien, R. (1998), Protein translocation into and across the cholroplastic envelop membranes. *Plant Mol. Biol.* **38**, 191-207.

Sonnhammer, E.L.L., Eddy, S.R., Birney, E., Bateman, A., Durbin, R. (1998). Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucl. Acids Res.* **26**, 320-322.

Sonnhammer, E.L.L., Eddy, S.R., Durbin, R. (1997). Pfam: A Comprehensive Database of protein Families Based on Seed Alignments. *Proteins* **28**, 405-420.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-541.

Stadler, L.J. (1940). Gene list and linkage map of corn (maize). *Maize Genet. Coop. News Lett.* **14**, 26-27.

Stebbins, G.L. (1971). *Chromosomal evolution in higher plants*. Arnold, London.

Stern, D.B., Higgs, D.C., Yang, J. (1997). Transcription and translation in chloroplasts. *Trends Plant Sci.* **2 (8)**, 308-315.

Stultz, C.M., White, J.V., Smith, T.F. (1993). Structural Analysis Based on State-space Modelling. *Protein Science* **2**, 305-314. http://bmerc-www.bu.edu/psa/

Stultz, C.M., Nambudripad, R., Lathrop, R.H., White, J.V. (1997).”Predicting protein Structure with Probabilistic Models”. In: Protein Structural Biology in Bio-Medical Research, ed. By Allewell, N., Woodward, C., Vol. **22B**; In: *Advances in Molecular and Cell Biology*, ed. By Bittar, E.E., JAI Press, Greenwich, pp 447-506.

Sugano, S., Andronis, C., Green, R.M., Wang, Z.-Y., Tobin, E.M. (1998). Protein kinase CK2 interacts with and phosphorylates the *Arabidopsis* circadian clock-associated 1 protein. *Proc. Natl. Acad. Sci. USA* **95**, 11020–25.

Tan, S., Troxler, R.F. (1999). Characterization of two chloroplast RNA polymerase sigma factors from *Zea mays*: Photoregulation and differential expression. *Proc. Natl. Acad. Sci. USA* **96**, 5316–21.

Tenaillon, M.I., Sawkins, M:C., Long, A.D., Gaut, R.L., Doebly, J.F., Gaut, B.S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. USA* **98 (16)**, 9161-9166.

Testi, M.G., Croce, R., Polverino-De Laureto, P., Bassi, R. (1996). A CK2 site is reversibly phosphorylated in the photosystem II subunit CP29. *FEBS Lett.* **399 (3)**, 245-250.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997). The clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876-82.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-80. http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html

Viard, T., Lamour, V., Duguet, M., Bouthier, De La Tour C. (2001). Hyperthermophilic Topoisomerase I from Thermotoga maritima. A VERY EFFICIENT ENZYME THAT FUNCTIONS INDEPENDENTLY OF ZINC BINDING. *J. Biol. Chem*. **276 (49)**, 46495-503.

von Heijne G., Steppuhn J., Herrman R.G. (1989). Domian structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* **180**, 535-545.

Waegemann, K., Soll, J. (1996). Phosphorylation of the Transit Sequence of Chloroplast Precursor Proteins. *J. Biol. Chem.* **271 (11)**, 6545-54.

Wang, B., Jones, D., Kaine, B.P., Weiss, M.A. (1998). High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. S*tructure* **6 (5)**, 555-569.

Watson, S.A. (1988). Corn Marketing, Processing, and Utilization. In: *Corn and Corn Improvement* – Agronomy Monograph no. **18**, 3rd Ed. Sprague, G.F., Dudley, J.W. eds., ASA, CSSA, SSSA Publishers. Madison. pp. 1-31.

White, D.G. (1999). Compendium of Corn. 3rd. Ed. APS Press.

White, J.V., Stultz, C.M., Smith. (1994). Protein Classification by Stochastic Modelling and Optimal Filtering of Amino-Acid Sequences. *Mathematical Biosciences* **119**, 35-75.

White, S.E., Doebley, J.F. (1998). Of genes and genomes and the origin of maize. *Trends Genet.* **14 (8)**, 327-332.

White, S.E., Doebley, J.F. (1999). The Molecular Evolution of *terminal ear1*, a Regulatory Gene in the Genus *Zea. Genetics* **153**,1455–62.

Whitkus, R., Doebley, J. & Lee, M. (1992). Comparative genome mapping of Sorghum and maize. *Genetics* **132**, 1119-30.

Wienand, U., Sommer, H., Schwarz, Z.S., Shepherd, H., Saedler, H., Kreutzaler, F., Ragg, H., Fautz, E., Hahlbrock, K., Harrison, B., Peterson, P.A. (1982). A general method to identify plant structural genes among genomic DNA clones using transposable element-induced mutations. *Mol. Gen. Genet.* **187**, 195-201.

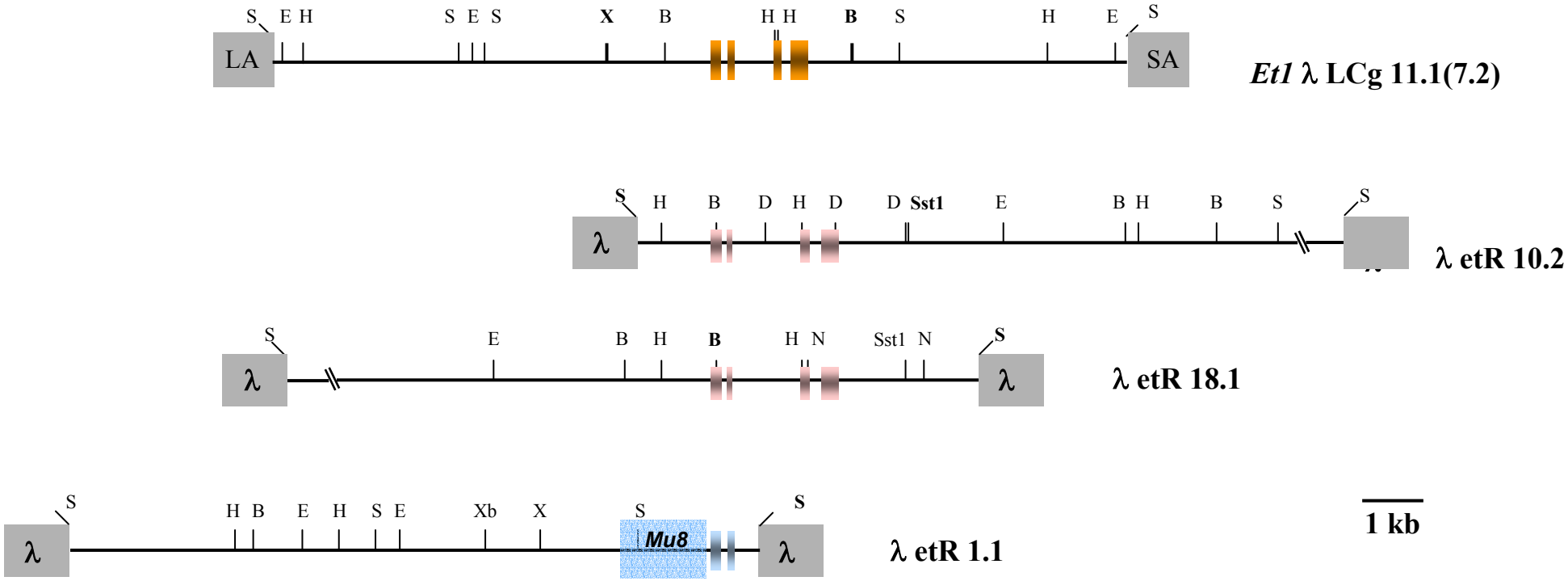
Wienand, U., Weydemann, U., Niesbach-Klösgen, U., Peterson, P.A., Saedler, H. (1986). Molecular cloning of the *C2* locus of *Zea Mays* L., the gene coding for chalcone synthase. *Mol. Gen. Genet.* **203**, 202 – 207.

Yanagisawa, S. (1995). A novel DNA-binding domain that may form a single zinc finger motif. *Nucleic Acids Res.* **23 (17)**, 3403-10.

Yu, L., Zhu, C.X., Tse-Dinh, Y.C., Fesik, S.W. (1995). Solution structure of the C-terminal single-stranded DNA-binding domain of *Escherichia coli* topoisomerase I. *Biochemistry* **34 (23)**, 7622-28.

# Appendices

**Appendix A:** Restriction maps of the genomic clones analysed in this study.



**Appendix B1:** Nucleic acid sequence of the *Et1* gene from Line C.

```
1       10        20        30        40        50        60       bp
CTCGAGGGTATACATGTGGATGAAGAAAAGACTATCAATGATCCAGAACTTCTTATGGAGGTAAT   65
TTCATTAATCTTGGAAATTGGAATACACGTGCTGGCATGCAGTAAATATAACCATCCTTCTCATT  130
TCATGTATCTTTTGATAACTGTGAATAACATTGCAACGATTGGAACACACATTCAATGTGTTAAA  195
CATCCTATGTAAAGTACATTAAAATGAATGAACTCTGAATTGAGATGAATACAACAACACAAACA  260
GTTTATTTATATACATCTTATTCCATTAACTTACACGAAATGAAGATGACAATGTTCTCCGTAAA  325
TCAATAACAAATCGCTTGCAGATGATGGATATACGGGAAGCTGTCAGTGATGCCAGTGATTCTCA  390
AACCCTGGAGAAGATCCAATCTCAGGTTCTGATCTCTGAATAAAGCTACACGTGCCATAACATAA  455
CCAGGGAGCATGATTGCGTCGCAATGGTTTCTGGTCTCATAGCCTTTTGTTCCCTTTTCCAGGTT  520
AAGGCAAAGCTCGAAACATGGTCCGATTCCTTCCAGGAGGCATTTGACAGGAAGGATTTTGACCG  585
TGCAGTGGAAGCCACACAGAGAATGAGGTACTATGAACGTGCAGTGGAAGAAACAGTGAAGAAGC  650
TCTGAGCCTCTGATCTTCTGTTACAGCTGCCAAAGGTGGTGAGGCTTGTGTCTTTTCGTACGTTC  715
CTTGCAGATCCTCCTGTGCTCACGCAGCTACATTGCAACAGTAACATAGTACAGAGTAGGTCAAT  780
TTTTTTACTCATACGTTGCAATGTATTATATGACTACATTGGAATATATTACTGTAGCTACATTG  845
CAACAGTAATATATGACTACAGAAGAGCAGACGTTGTGATGCTAAATTCCAAATGTATCCATGGC  910
TGGTGCAGAGTAGGTCATTTTTTTTACTTATACTTTCCACAATTCCAAATGGATCCTTGGCTTTT  975
CAATAAAACAAAACTTGGTGTGTTTGGTTGTGGGACAGCTAGAATAGGGACGTCCTCTCTCGTCT 1040
CTCCAATTTTTGAGATATAACTGGGACAACATTGGGATAGTCTTGTCTCAACCTTTGATTCTAAA 1105
TCAAACAATCTTATTTGAGGGATCGTCATATCCCGTCCCATCGTCCTGTTACTGCAACCAAATGC 1170
ATCATAAAAGAGTAGATAGGGATTCAGAATAAGGCTTCGATGCCATATATTTATAGAGCGCAAAC 1235
CTTTGTATTCTGGTGTGGCCGTGTGGGTTTGTTGGAATGTACGTGGTGGATGTGATGCATGTTAA 1300
GCGAGGATGTTCTACCCGGATCAAAAGGCCGTTGATGAGAGGGACAGACTCAGATACAGTTAGCG 1365
TCGGGCTCACTCAGGCCATGATGAATCACTGCCAGAAATACGAGTTTTTGTTTTTTTTGCCGTCC 1430
ACAAGTCCCAGAGAGGAGACTCAGGAGAAATACCAGAATTTCTTCCGGCGGCCGCAAAAACCAGC 1495
GAGAGCATGAGAACGGAGAACCATCGCCGTCGGGACCGTGACGCTGGACGGCCTAAAGCGGCTGC 1560
GGGGAGGTGTGTTAAACGGGCCTTTCCTGGGCTACGGTGGACAGCCCAGTTACTAAGTTCACCCT 1625
CGCTCTTCCTTTCCCCTCGGTTCTCGTTGTCTGACAGGCAAAGCGGCTATGACGACGACGGCCGC 1690
CGGGCACGGCTGCTGCTGGGCGGGGATTCCGCCCTTCGCGTTGTTGCCGCGGATTCTCTCGACCG 1755
GCCGGGAGACTCCTCCTCCTCGCGCTTCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCA 1820
CCGCGGTGCGCTCCGTTCTAGCTCGCTCTTGCCGGCTCGTCTCGGTTCGCGGGTGTCTAAGCTCG 1885
TCTCTCTTCGTGTAGGCTGAGAGTTTCGAACCGTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCT 1950
CCGGCGAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGTCAGCGCTCTCTATCTTTCTTCTCTA 2015
TGATGTTTCCAGATTGTCAGGGTGAATTGATTTGGGTTATGGATGTGCTCCGTGTTGCCTGGAAT 2080
CATTCAGTGAATGCGATTATATAGGGAATGATTTGGATTGGGTAGCCACCAGCAACGGGGAAACA 2145
GTAATTGCTTGGGCTTTTGGACCACTGTTTATTATTAACCCTGAAATTGGATAACTGACGAGATT 2210
CAACTTTAATGATTCATATAGTCAGAATATTTGTGGTTTTCTAGCACCACACATCTGCTTTGTCC 2275
ATTGTATGACTGGCGTACGTTGTTCATTTGAAATTGCTAGTTCTTTCCTGAGATTGTGCTTTGTG 2340
TGCATTTCTTGTAGGTCTTTCGTTGCTTCTGTGTCTTGTAGATTATGCCATTCAGAGGACAACAC 2405
TTTCTGTAAATTTGGCAGGTTGATCGCATACATATGTGAACATATGAAAAAGAAATGATATTATA 2470
TAATTCCGCGGTTTATTGCATGAAAGACAACATAACAGATTGCTTGCTGATTGAGGAACAAACTA 2535
GATTCAGAAATGAATAACAAGTGCTCAACTTTGCTGGTAGATTGATTACTGAAGCGGCAAACTGC 2600
AGACTAGTTTATAGGCCGGAATCTCCTTAATGTACCTGTTATGCGTGTGATGTTTCAGGAAGCCG 2665
CTATTGATATAAAGCTTCCTAGAAGAAGCTTGCTTGTTCAATTTACATGCAACGCATGTGGCGAA 2730
AGGACCAAGCGCTTGATAAACAGAGTAGCCTATGAAAGAGGCACAGTTTTTCTTCAGGTAAAAAC 2795
TTCTTTCCATGTAAATGCAAACTTCAGACCTACTGCAGCTTTGCTCATTCTGGTCCGCTGTATGA 2860
GAAAAGACCACATTCGTATGTTTGACGACACATACCGTAAATTTATGTTCTAACATACACTGCAC 2925
AAAACAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAATCTTGGGCTAGTTGTTGAGT 2990
ATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCGGTGGACACCAGTTCTGAAGATTGA 3055
TCTGCTGTGAGAAGCGATGTTGGTATGCAAAACGCCCTGTACTCTGTAGGTTTTTGACAACATTG 3120
GTTATTTGTATAGCATAAAAATGGCACTTTTTAAAGTTGTTGCACATACTCATCTGAAATTCTGA 3185
ATACAGCAGGGCCTACATTGTACTTCTTACAACGTGATATTTATATACTAGTCTATTAAGAATCT 3250
AATATAAGTACCTAGTGCTACCACGACTTCACCCTCTACGTTGATACTCTGGACCTGCCTCAGGC 3315
ATCCCGCCCTAGGCACCGTTCCCTGTGGGCCCAGCGCCCACACAACGTTCTCTCAGCTACTTAAT 3380
```

(continued on the next page)

```
1       10        20        30        40        50        60       bp
TGTGGACCCCCAGCATCCGCGCGCAGCCACCACACCACGCAAAAATAGTGCAGAACAAACACTTG 3445
AACTCACGCTTCGTTGCTCTAGAAATTTAGGACTAACCACCAAACCACACATAATTTGGTGTTTA 3510
GAAATAAACAATAATTATATTTGAATAAATATCTTAAATACAATTGGTTAGTTTGGTTTTAAGAA 3575
GACATCGAATGATCATAGCACCAGATTACACTGCTGTTCCAAAAAAAAAATGATATCTTATTAAAA 3640
TATGAAAACCATACATTGTCGGCAGGTTATACTGGCGCTACTGATATGGTGCTGATATTTGATAT 3705
CCACGCTACTCTGGCATCATTGAGATTTTGATCGACGACCGGGACATGTTTTTTTTTCTTTTTTC 3770
GTCGTCATCTCTTTCTCGTCCCCTTTGGTCTCACATGGTCCATTTTAACCTTGACTACTATCCTT 3835
GGCGTGGGCCTTCTTCCCATTTAGGAACTGGATGAGACCCTCAAGCGCGACGTCGGTGTCTTCGT 3900
TCCTCATGAGCACTTCAGCGACCTCTGCCGGGGTCACCATCACCTCCGTGATCAGCGCTTCTATC 3965
TGAGGGTACGTGTCGTGGTCGGTGATGGAGTGGTAGTTGGAGGCCAGGATCC               4017
```

**Legend**

bp : length in **b**ase **p**airs

ATGC : Introns, and regions 5´and 3´ of the gene

ATGC : Exons

ATG/TGA : Translation start or stop codons

**Appendix B2:** Nucleic acid sequence of the *zmzr1* gene from *et1-R* line.

```
1       10        20        30        40        50        60       bp
GATCACTGTGGCTGGGCAGGTTCACAAGTTCGTCGCCAATGACCAATCGCATCCTTGGACGTTAG   65
AGATCCTGGCGATGATGGAGGAGATCGCATGCAGGCTGAAGTCTGTGGGTTATTCTCCAGTCACG  130
TCAAGGATAGCAGTTGACGTTGACGAGGAAGAGAAAGAACAGGCACTTTTAGCACACAGTGAGAA  195
GATAGCCATTGCTTTTGGTCTCATCAGTCTTCCACCGAGCTTACCGATCCATATCGTGAAGAACC  260
TGAGAGTCTGTGAGGATTGTCACTCTGCAATCAAGTTAGTCTCTCAGCTTTGGAACCGGGAAATC  325
ATCGTTAGAGATCGGAGTAGGTTCCACCATTTCCGAGACGGAGCATGTTCTTGCAATGATTTTTG  390
GTGACTACCATTCTGCTAAGCTTCTAGTTCTAGAGGCCAGAAAACAGCCACATGGATACATTGAA  455
AGCACAAGTGCTTCATCTCGCAATTCGGCTTCCGGATTCAAATTTATGAATGCCCAAGGTACTAT  520
TTTCTTCCTGTTTCTTCATGCCATATACTTGCTCCGTTTTAAATTTTAAGTTTTTTGACTCTTTT  585
TTAGATATTTAACTTTCGATATGATACATAGTAAAAACTTAAAAAAAAAACAGAACGATTCACCAT  650
TTATTGTGAACGGAGGGAGTACTAAACGAAATGAGCTAACGACACGAATCTTCTGATGATAACAC  715
AAAGTTTTCTTGTGTGCTTTGCCAGGGGTAGGAAACCGGATGTCTACACGCAAAATCACCCCTTT  780
CGTGGCTGCTAAACGGTGTGCCACTGCCATGGCATTGTTGTTTCGACCACCCCTGTTTGGGAATA  845
GCCCAGTTTCGGGTTTTAACTTAGCCCAGTTTCGGGACAGTGATAAAAAGGCAACAAGCCCACAA  910
CCCAAAGGCTGACAGGGGTAAACACAGGAATGGGCTGGATAGGTCATCTAAATTGCTTAACGGGC  975
CAGCCAAGTCAATATGGACACAGGCGGAACCAAAGCCGGCCCACGCTGCTCGCCTGCTCCGGCGA 1040
CCGCCCTCCGCCTCCGCCCTGAGTGACATCCCCTGCGACCAGCCGCTCGCCGGCGGTGTTCGCGC 1105
TTCGCCGGCACGCCGGCGTGCGCCCTCGTGTGCCGTGGCTGCTGCGTGCCTTGTCCACCGAGGTG 1170
ACTGGTGAGATCTCGCTCTTCCTTTATCCTCGCCTCCTCGGTCCAAGTAGGAGTCGGCGGCTATG 1235
GCGACGACGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTGCCTGG 1300
GATCCTCTCGACCCGCCTGCGCCGGGAGCCTTCTCCTCCTCGCGTGGCCCTTGTGGCCTCCTCCC 1365
CGAAGCTCAGGGCGCCGGCACCGCGGTGCGCTCCGTTCTAGCCCCTTCTTCACGGCTCGGCTCGC 1430
GGGTGTGTAAGCCCGTTTCTCTTCGCGCAGGTTGAGAGTTTCGTGCGGTCCAAGGAGGCTCGCTT 1495
CCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGTTAGTTAGCGCTCTCTCTCC 1560
TTCCTCTCTATGATGTTTCCAGATCATCAGGGGTGAATGGTTTGGGTTATGGATGGTCCATGTTC 1625
CTTGGAATTATTTACTGGAGGCGATTTAGGGAATGGATTGGGTTAGGTAGCCACCAGCAATGGAG 1690
GAGACAGTAAATTGCTTGGGCTTTTGGACCACTGTTTATTGAAATTGGATAACCTGACAAGATTC 1755
AATCTGATATGCTGATGTAGGTTGCTAGCTATTCTATATCAACTTTAATGATTCTACCCGTGTTT 1820
TGCTACGGTTTTTATATATCATATAAATATGTATTGAGATATTTATTCAAATATAATTATTGTGT 1880
ATTTCTAAACTCTAAGGTACGTGTGGTTTATTGGTTAGCCCTAAATTTCTGTAGCAGGAGGCGAG 1950
CGTGTGTTCGAGTGCTCGCTTTGCACTATTTTTTATATAATTATTGTTTATTTCTAAACACTAAG 2015
TTACGTATGGTCTGGTGGTAGCTGCGCGGGCGGGGTCGAGTACTGGACAGGCGCTGCGGTCCACA 2080
GGGCAGTAGCTGAGAAAGCGCTGCGCGTGGGGCCCACAGGGCATGGCGCCAAGGCAGGGCGCGTG 2145
GAGCGGGCCCAGAGTGTCAGCGCGGAGGGTGAAGCCGTGGTAGTACCAGATGTCCACATTAGGTT 2210
TTTAATAGAGTAGTATAAATATTCTCAGAATAATTGTGTTTTTCTAGTATCCTATCTTCTTTGTC 2275
CATTGTCTGATTGGCTTTAGTAAACTATATGAATGCAATCCGAATGTTGTTCATTTGGAATTGCT 2340
AATTCTTTTCTGAGTTTGTACTTAGTGTACATTTCTTGTAGGTCTTCCATTGCTTGGGTGTCTTG 2405
CAGATTAAGGCATTCAAAAGACAACACAGTTTGTAAAGTGTAAATTTGGTAGGTTAATAGCATAA 2470
ATATGTGAAAACATGGAAAATAAACAATGATAAGGTCTTGCATTTTATGTTTGATCCAGTAGTTT 2535
ATTGCATGAAAGACAATGTAATAAGTTGCTTGCTGATTGAGGATCAAACTAGGTTCAGAAATAAA 2600
TAATAAATGCTCAATTTGTTGATAGATTGAAGTGAAGCGGCAACCAACAGACTAGTTTGTAGGCA 2665
GAGTGTACTGAATGTATTTTTTTGTTTGTGTGATGTTTCAGGAAGCCACTATTGATATAAAGCTT 2730
CCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGTGGTGAAAGGACCAAGCGCTTGAT 2795
AAATAGAGTAGCCTATGAAAGAGGGACAATTTTTCTTCAGGTAAACTTTGTTCCATGCCTTAATT 2860
TATGTGCAAATGTAATTTCCAAACCTAATGCTATGATCATTTTGGATCGCTGTATGAGACGATTC 2925
TTATGATTGACGTTTTCTTACACATGCTGTAAATATATTTTCTAACATACACTGCATAACACAGT 2990
GTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTTGAGTATGATCTA 3055
CGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATTGATACTAATTCTGAAGATTGATCTGTTGT 3120
GAGAAGCAGTGTTGGTATGCAAAACATCCTGTTGTAGCCTGTTGTCTGAATGTCTGTAGGTTTTG 3185
ACAATATTGATTATTTATATAAAATGACATTTTAGTTGCTGCATATACTCATCTGAATACATGGG 3250
GGCCTAGATGTATTTTATAGTGTGATATATGGTTTTATGAAGGAAACAAAGGATCACGATACTAG 3315
AGTACCCTGCATTTCCAAAAAAGACCTGTCATTAATATGGGACCAATACATTTTTAGCAGATTAT 3380
```

(continued on the next page)

```
1       10        20        30        40        50        60        bp
TCTGGCTCTTGTGATATGGTACTAATATTTGATATGCACACTACTTTGCACCATTATTGGCCTGC 3445
ATCATTGAGATTTTGATCAACTGGGGCAATTTTGTAGTCATCACCTTCGCTATCCTTGGCATCAC 3510
CCTTCTTTCCCTTGAGGAACCGAATAAGGCCCGGTGTCGTCATTCCTCATAAGCACTTCTGAGAT 3575
CTCTGTTGGAGTCAGCATCACCTCCTTAGGTCAATGCTTGTTTACCAGTACACGAGAAAATGGAA 3640
GCTGGATGCTTCTTTACCAATACACCAGCATTTACAGAACCACTAGGAGCAACAGCAGCATACTA 3705
TGTCAAAAACCAGTAAGTATTTGTATTCTTGACTAGTAGCATTTCTGTCATTAAGCACTGTCATA 3770
CTATGTAAAAAATGATCAGTACTACTGTACTACTTTACAATGATCAAACCATATACCACTAAAAA 3835
TATTCGAAGTATTGCTATTCGTAGGGATTCCTATGTAGGCCTGTAATGAGTTGGGTTGAAATGAC 3900
TTTAATGGGTTAGGTAGGTTCTCAATTGGGTACATTTTGGATGGGTGGGAATGGGTTGATTTACT 3965
TAGGCCTTTTTTGGATGTCCATTTATCGATCTTAATCCATGTGAAATTTAGTTTGAATTTCACAC 4030
CAACCCACTCCAATACATATAGGTTGAGGTCAACATATGAGCACCCAAACAAATTCTTAGCGGAT 4095
TGGGTTAGATTTGTTTTAATAGAAAAATATGTTGGATGGGATGGGATCAGGGTTTTTTAGGTATG 4160
AATGGGTTGAACCCTTGGGTTTGATGTTGGATTTTCAACATATTGCAGTATACATATATATTGGT 4225
TTGATTTGAGGTGGTATACTAACTTAAGGTGGGACCTGAACAAATTAAATTACAACACCGGTTAG 4290
TATAGAGTAGCTTTATGACATTTTAAATCGTACCCAAATAATTTTCAGTCAATTTCGTTAAAATT 4355
ATAAGGTGCGAATGTGGAGCGTTTATTTTGTCTAGAGCTCAGTTTGTGATGTGAGGCCCACATGT 4420
AGATGTAGATCTTGCTGCGCGCTGGCGCAGATGAGCTCTATGACAGATTTCAGTCAGTTTTGATA 4485
AAATCTCTGTGCTGCCTTTTTTGCTGTACAATGAGATGCCATTGCAGAGAGCACGGTCATGTCCA 4550
ACGTGGTTCTGGGTGGTTTGTGCAACCTGAAGCTGACTGGGGGACCTCCTTTGCTCCTTGTGCTA 4615
GACATGAGACGCCGGCATGGTTTGGACTTGAATGGTCTCTCCTACTGTTACGCGATGAAAGTATG 4680
CTATCAGGATGAAGAATGCAT                                             4701
```

**Legend**

bp : length in **b**ase **p**airs

ATGC : Introns, and regions 5´and 3´ of the gene

ATGC : Exons

ATG/TGA : Translation start or stop codons

## **Appendix B3**: Comparison of *et1* and *zmzr1* genomic sequences.

```
                *        20         *        40         *        60         *        80
zmzr1g : GATCACTG..............TGGC......TGGGCA.GGTTCACAAGTTCGTCGCCAATGACCAATCGCATCCTTGGA
LC     : GATCTCTGAATAAAGCTACACGTGCCATAACATAACCAGGGAGCATGATTGCGTCGCAATGGTTTC.TGGTCTCATAGCC
LCcDNA : ................................................................................

                *       100         *       120         *       140         *       160
zmzr1g : CGTTAGAGATCCTGGC.GATGATGGAGGAGATCGCATGCAGGCTGAAGTCTGTGGGTTATTCTCCAGTCACGTCAAGGAT
LC     : TTTTGTTCCCTTTTCCAGGTTAAGGCAAAGCTCGAAACATGGTCCGATTCCTTCCAGGAGGCATTTGACAGGAAGGATTT
LCcDNA : ................................................................................

                *       180         *       200         *       220         *       240
zmzr1g : AGCAGTTGACGTTGACGAGGAAGAGAAAGAACAGGCACTTTTAGCACACAGTGAGAAGATAGCCATTGCTTTTGGTCTCA
LC     : TGACCGTGCAGTGGAAGCCACACAGAGAATGAGGTACTATGAACGTGCAGTGGAAGAAACAGTGAAGAAGCTCTGAGCCT
LCcDNA : ................................................................................

                *       260         *       280         *       300         *       320
zmzr1g : TCAGTCTTCCACCGAGCTTACCGATCCATATCGTGAAGAACCTGAGAGTCTGTGAGGATTGTCACTCTGCAATCAAGTTA
LC     : CTGATCTTCTGTTACAGCTGCCAAAGGTGGTGAGGCTTGTGTCTTTTCGTACGTTCCTTGCAGATCCTCCTGTGCTCACG
LCcDNA : ................................................................................

                *       340         *       360         *       380         *       400
zmzr1g : GTCTCTCAGCTTTGGAACCGGGAAATCATCGTTAGAGATCGGAGTAGGTTCCACCATTTCCGAGACGGAGCATGTTCTTG
LC     : CAGCTACATTGCAACAGTAACATAGTACAGAGTAGGTCAATTTTTTTACTCATACGTTGCAATGTATTATATGACTACAT
LCcDNA : ................................................................................

                *       420         *       440         *       460         *       480
zmzr1g : CAATGATTTTTGGTGACTAC.CATTCTGCTAAGCTTCTAGTT..CTAGAGGCCAGAAAACAGCCACATGGATACATTGAA
LC     : TGGAATATATTACTGTAGCTACATT.GCAACAGTAATATATGACTACAGAAGAGCAGACGTTGTGATGC.TAAATTCCAA
LCcDNA : ................................................................................

                *       500         *       520         *       540         *       560
zmzr1g : AGCACAAGTGCTTCATCTCGCAATTCGGCTTCCGGATTCAAATTTATGAATGCC.CAAGGTACTATTTTCTTCCTGTTTC
LC     : ATGTATCCATGGCTGGTGCAGAGTAGGTCATTTTTTTTAC..TTATACTTTCCACAATTCCAAATGGATCCTTGGCTTTT
LCcDNA : ................................................................................

                *       580         *       600         *       620         *       640
zmzr1g : TTCATGCCATATACTTGCTCCGTTTTAAATTTTAAGTTTTTTGACTCTTTTTTAGATATTTAACTTTCGATATGATACAT
LC     : CAATAAAACAAAACTTGGTGTGTTTGGTTGTGGGACAGCTAGAATAGGGACGTCCTCTCTCGTCTCTCCAATTTTTGAGA
LCcDNA : ................................................................................

                *       660         *       680         *       700         *       720
zmzr1g : AGTAAAAACTTAAAAAAAAACAGAACGATTCACCATTTATTGTGAACGGAGGGAGTACTAAACGAAATGAGCTAACGACA
LC     : TATAACTGGGACAACATTGGGATAGTCTTGTCTCAACCTTTGATTCTAAATCAAACAATCTTATTTGAGGGATCGTCATA
LCcDNA : ................................................................................

                *       740         *       760         *       780         *       800
zmzr1g : CGAATCTTCTGATGATAACACAAAGTTTTCTTGTGTGCTTTGCCAGGGGTAGGAAACCGGATGTCTACACGCAAAATCAC
LC     : TCCCGTCCCATCGTCCTGTTACTGCAACCAAATGCATCATAAAAGAGTAGATAGGGATTCAGAATAAGGCTTCGATGCCA
LCcDNA : ................................................................................

                *       820         *       840         *       860         *       880
zmzr1g : CCCTTTCGTGGCTGCTAAACGGTGTGCCACTGCCATGGCATTGTTGTTTCGACCACCCCTGTT.TGGGAATAGCCCAGTT
LC     : TATATTTATAGAGCGCAAACCTTTGTATTCTGGTGTGGCCGTGTGGGTTTGTTGGAATGTACGTGGTGGATGTGATGCAT
LCcDNA : ................................................................................
```

(continued on the next page)

```
                *       900         *       920         *       940         *       960
zmzr1g : TCGGGTTTTAACTTAGCCCAGTTTCGGGACAGTGATAAAAAGGCAACAAGCCCACAACCCAAAGGCTGACAGGGGTAAAC
LC     : ...GTTAAGCGAGGATGTTCTACCCGGATC......AAAAGGCCGTTGATGAGAGGGACAGACTCAGATACAGTTAGCGT
LCcDNA : ..........................................................................

                *       980         *      1000         *      1020         *      1040
zmzr1g : ACAGGAATGGGCTGGATAGG.TCA.TCTAAATTGCTTAACGGGCCAGCCAAGTCAATATGGACACAGGCGGAACCAAAGC
LC     : CGGGCTCACTCAGGCCATGATGAATCACTGCCAGAAATACGAGTTTTTGTTTTTTTTGCCGTCCACAAGTCCCAGAGAGG
LCcDNA : ..........................................................................

                *      1060         *      1080         *      1100         *      1120
zmzr1g : CGGCCCACGCTGCTCGCCTGCTCCGGCGACCGCCCTCCGCCTCCGCC.CTGAG....TGACATCCCCTGCGACCAGCCGC
LC     : AGACTCAGGAGAAATACCAGAATTTCTTCCGGCGGCCGCAAAAACCAGCGAGAGCATGAGAACGGAGAACCATCGCCGTC
LCcDNA : ..........................................................................

                *      1140         *      1160         *      1180         *      1200
zmzr1g : TCGCCGGCGGTGTTCGCGCTTCGCCGGCACGCCGGCGTGCGCCCTCGTGTGCCGTGGCTGC.TGCGTGCCTTGTCCA.CC
LC     : GGGACCGTGACGCTGGACGGCCTAAAGCGGCTGCGGGGAGGTGTGTTAAACGGGCCTTTCCTGGGCTACGGTGGACAGCC
LCcDNA : ..........................................................................

                *      1220         *      1240         *      1260         *      1280
zmzr1g : GAGGTGACTGGTGAGATCTCGCTCTTCCTTTATCCTCGCCTCCTCG..GTCCAAGTAGGAGTCGGCGGCTATGGCGACGA
LC     : CAGTTACTAAGTTCACCCTCGCTCTTCCTTTCCCCTCGGTT.CTCGTTGTCTGAC.AGGCAAA.GCGGCTATGACGACGA
LCcDNA : ..................................................GGCAAA.GCGGCTATGACGACGA

                *      1300         *      1320         *      1340         *      1360
zmzr1g : CGGCCGCGGGGTACGGCTGTTGCTTGGCGGGGCTACCGCCCTTTCCGTTGTTGCCTGGGATCCTCTCGACCCGCCTGCGC
LC     : CGGCCGCCGGGCACGGCTGCTGCTGGGCGGGGATTCCGCCCTTCGCGTTGTTGCCGCGGATTCTCTCGACCGGCC.....
LCcDNA : CGGCCGCCGGGCACGGCTGCTGCTGGGCGGGGATTCCGCCCTTCGCGTTGTTGCCGCGGATTCTCTCGACCGGCC.....

                *      1380         *      1400         *      1420         *      1440
zmzr1g : CGGGAGCCTTCTCCTCCTCGCGTGGCCCTTGTGGCCTCCTCCCCGAAGCTCAGGGCGCCGGCACCGCGGTGCGCTCCGTT
LC     : .GGGAGACTCCTCCTCCTCGCGCTTCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCGCGGTGCGCTCCGTT
LCcDNA : .GGGAGACTCCTCCTCCTCGCGCTTCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCGCG............

                *      1460         *      1480         *      1500         *      1520
zmzr1g : CTAGCCCCTTCTTCACGGCTCGGCTCG.....CGGGTGTGTAAGCCCGTTTCTCTTCGCGCAGGTTGAGAGTTTCGTGCG
LC     : CTAGCTCGCTCTTGCCGGCTCGTCTCGGTTCGCGGGTGTCTAAGCTCGTCTCTCTTCGTGTAGGCTGAGAGTTTCGAACC
LCcDNA : ..........................................................GCTGAGAGTTTCGAACC

                *      1540         *      1560         *      1580         *      1600
zmzr1g : GTCCAAGGAGGCTC.........GCTTCCTCCTCTGGCAAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGTTAGTTAG
LC     : GTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGAGGCCAATTCGGACGCGGTGCCGTCGCCAACGGT....CAG
LCcDNA : GTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGAGGCCAATTCGGACGCGGTGCCGTCGCCAACG.........

                *      1620         *      1640         *      1660         *      1680
zmzr1g : CGCTCTCTCTCCTTCCTCTCTATGATGTTTCCAGATCATCAGGGGTGAATGGTTTGGGTTATGGATG.G.TCCATGTTCC
LC     : CGCTCTCTATCTTTCTTCTCTATGATGTTTCCAGATTGTCAGGGTGAATTGATTTGGGTTATGGATGTGCTCCGTGTTGC
LCcDNA : ..........................................................................

                *      1700         *      1720         *      1740         *      1760
zmzr1g : TTGGAATTATTTACTGGAGGCGATTT...AGGGAATGGATTGGGTTAGGTAGCCACCAGCAATGGAGGAGACAGTAAATT
LC     : CTGGAATCATTCAGTGAATGCGATTATATAGGGAATGATTTGGATTGGGTAGCCACCAGCAACGGG.GAAACAGT.AATT
LCcDNA : ..........................................................................
```

(continued on the next page)

```
                *      1780         *      1800         *      1820         *      1840
zmzr1g : GCTTGGGCTTTTGGACCACTGTTTATT.........GAAATTGGATAACCTGACAAGATTCAATCTGATATGCTGATGTA
LC     : GCTTGGGCTTTTGGACCACTGTTTATTATTAACCCTGAAATTGGATAA.CTGACGAGAT.....................
LCcDNA : ................................................................................

                *      1860         *      1880         *      1900         *      1920
zmzr1g : GGTTGCTAGCTATTCTATATCAACTTTAATGATTC.TACCCGTGTTTTGCTAC.GGTTTTTATATATCATATAAATATGT
LC     : ..................TCAACTTTAATGATTCATATAGTCAGAATATTTGTGGTTTTCTAGCACC.ACACATCTGCT
LCcDNA : ................................................................................

                *      1940         *      1960         *      1980         *      2000
zmzr1g : ATTGAGATATTTATTCAAATATAATTATTGTGTATTTCTAAACTCTAAGGTACGTGTGGTTTATTGGTTAGCCCTAAATT
LC     : TTGTCCATTGTATGACTGGCGTAC.....................................................
LCcDNA : ................................................................................

                *      2020         *      2040         *      2060         *      2080
zmzr1g : TCTGTAGCAGGAGGCGAGCGTGTGTTCGAGTGCTCGCTTTGCACTATTTTTTATATAATTATTGTTTATTTCTAAACACT
LC     : ................................................................................
LCcDNA : ................................................................................

                *      2100         *      2120         *      2140         *      2160
zmzr1g : AAGTTACGTATGGTCTGGTGGTAGCTGCGCGGGCGGGGTCGAGTACTGGACAGGCGCTGCGGTCCACAGGGCAGTAGCTG
LC     : ................................................................................
LCcDNA : ................................................................................

                *      2180         *      2200         *      2220         *      2240
zmzr1g : AGAAAGCGCTGCGCGTGGGGCCCACAGGGCATGGCGCCAAGGCAGGGCGCGTGGAGCGGGCCCAGAGTGTCAGCGCGGAG
LC     : ................................................................................
LCcDNA : ................................................................................

                *      2260         *      2280         *      2300         *      2320
zmzr1g : GGTGAAGCCGTGGTAGTACCAGATGTCCACATTAGGTTTTTAATAGAGTAGTATAAATATTCTCAGAATAATTGTGTTTT
LC     : ................................................................................
LCcDNA : ................................................................................

                *      2340         *      2360         *      2380         *      2400
zmzr1g : TCTAGTATCCTATCTTCTTTGTCCATTGTCTGATTGGCTTTAGTAAACTATATGAATGCAATCCGAATGTTGTTCATTTG
LC     : ............................................................GTTGTTCATTTG
LCcDNA : ................................................................................

                *      2420         *      2440         *      2460         *      2480
zmzr1g : GAATTGCTAATTCTTTTCTGAGTTTGTACTTAGTGTACATTTCTTGTAGGTCTTCCATTGCTTGGGTGTCTTGCAGATTA
LC     : AAATTGCTAGTTCTTTCCTGAGATTGTGCTTTGTGTGCATTTCTTGTAGGTCTTTCGTTGCTTCTGTGTCTTGTAGATTA
LCcDNA : ................................................................................

                *      2500         *      2520         *      2540         *      2560
zmzr1g : AGGCATTCAAAAGACAACACAGTTTGTAAAGTGTAAATTTGGTAGGTTAATAGCATAAATATGTGAAAACATGGAAAATA
LC     : TGCCATTCAGAGGACAACACTTTCTG.......TAAATTTGGCAGGTTGATCGCATACATATGTGAACATATGAAAAAGA
LCcDNA : ................................................................................

                *      2580         *      2600         *      2620         *      2640
zmzr1g : AACAATGATAAGGTCTTGCATTTTATGTTTGATCCAGTAGTTTATTGCATGAAAGACAATGTAATAAGTTGCTTGCTGAT
LC     : AATGATATTATATAATT.............CC.GCGGTTTATTGCATGAAAGACAACATAACAGATTGCTTGCTGAT
LCcDNA : ................................................................................
```

(continued on the next page)

```
                 *      2660         *      2680         *      2700         *      2720
zmzr1g : TGAGGATCAAACTAGGTTCAGAAATAAATAATAAATGCTCAA.TTTGTTGATAGATTGA..AGTGAAGCGGCAACCAACA
LC     : TGAGGAACAAACTAGATTCAGAAATGAATAACAAGTGCTCAACTTTGCTGGTAGATTGATTACTGAAGCGGCAAACTGCA
LCcDNA : ................................................................................

                 *      2740         *      2760         *      2780         *      2800
zmzr1g : GACTAGTTTGTAGGCAG.AGTGTACTGAATGTATTTTTTTGTTTGTGTGATGTTTCAGGAAGCCACTATTGATATAAAGC
LC     : GACTAGTTTATAGGCCGGAATCTCCTTAATGTACC.TGTTATGCGTGTGATGTTTCAGGAAGCCGCTATTGATATAAAGC
LCcDNA : ..........................................................GAAGCCGCTATTGATATAAAGC

                 *      2820         *      2840         *      2860         *      2880
zmzr1g : TTCCTAGAAGAAGTTTGCTTGTTCAATTTACATGCAATGCATGTGGTGAAAGGACCAAGCGCTTGATAAATAGAGTAGCC
LC     : TTCCTAGAAGAAGCTTGCTTGTTCAATTTACATGCAACGCATGTGGCGAAAGGACCAAGCGCTTGATAAACAGAGTAGCC
LCcDNA : TTCCTAGAAGAAGCTTGCTTGTTCAATTTACATGCAACGCATGTGGCGAAAGGACCAAGCGCTTGATAAACAGAGTAGCC

                 *      2900         *      2920         *      2940         *      2960
zmzr1g : TATGAAAGAGGGACAATTTTTCTTCAGGTAAA..CTTTGTTCCATGCCTTAATTTATGTGCAAATGTAATTTCCAAACCT
LC     : TATGAAAGAGGCACAGTTTTTCTTCAGGTAAAAACTTCTTTCCATGTA.AATGCAAACTTCAG.............ACCT
LCcDNA : TATGAAAGAGGCACAGTTTTTCTTCAG.....................................................

                 *      2980         *      3000         *      3020         *      3040
zmzr1g : AAT...GCTATGATCATTTTGGATCGCTGTATGAGACGATTCTTATGATTGACGTTTTCTTACACATGCTGTAAATATAT
LC     : ACTGCAGCTTTGCTCATTCTGGTCCGCTGTATGAGAAAAGACCACATTCGTATGTTTGACGACACATACCGTAAATTTAT
LCcDNA : ................................................................................

                 *      3060         *      3080         *      3100         *      3120
zmzr1g : TTTCTAACATACACTGCATAACACAGTGTGCAGGGTGCCAGGTGTACCACAAGTTTGTTGATAATCTTGGTCTAGTTGTT
LC     : GTTCTAACATACACTGCACAAAACAGTGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAATCTTGGGCTAGTTGTT
LCcDNA : ........................TGTGCAGGGTGCCAGGTGTACCATAAGTTTGTTGATAATCTTGGGCTAGTTGTT

                 *      3140         *      3160         *      3180         *      3200
zmzr1g : GAGTATGATCTACGAGAGGAAAATGTGGTACAAGGAGAAAATGTGATTGATACTAATTCTGAAGATTGATCTGTTGTGAG
LC     : GAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCGGTGGACACCAGTTCTGAAGATTGATCTGCTGTGAG
LCcDNA : GAGTATGATCTACGAGAAGAAAACGAGCTACAAGGAGAAAATGCGGTGGACACCAGTTCTGAAGATTGATCTGCTGTGAG

                 *      3220         *      3240         *      3260         *      3280
zmzr1g : AAGCAGTGTTGGTATGCAAAACATCCTGTTGTAGCCTGTTGTCTGAATGTCTGTAGGTTTT.GACAATATTGATTATTTA
LC     : AAGCGATGTTGGTATGCAAAACGCCCTGTAC..................TCTGTAGGTTTTTGACAACATTGGTTATTTG
LCcDNA : AAGCGATGTTGGTATGCAAAACGCCCTGTAC..................TCTGTAGGTTTTTGACAACATTGGTTATTTG

                 *      3300         *      3320         *      3340         *      3360
zmzr1g : TATAAAATGACA........TTTTA..GTTGCTGCATATACTCATCTGAA.TACATG........GGGGCCTAGA.TGTA
LC     : TATAGCATAAAAATGGCACTTTTTAAAGTTGTTGCACATACTCATCTGAAATTC.TGAATACAGCAGGGCCTACATTGTA
LCcDNA : TATAGCATAAAAATGGCACTTTTTAAAGTTGTTGCACATACTCATCTGAAATTC.TGAATACAGCAGGGCCTACATTGTA

                 *      3380         *      3400         *      3420         *      3440
zmzr1g : .TT.TTATAGTGTGATATATGGTT.....TTATGAAGGAAACAAAGGATCACGATACTAGAGTACCC..TGCATTTCCAA
LC     : CTTCTTACAACGTGATATTTATATACTAGTCTATTAAGAATCTAA.........TA.TA.AGTACCTAGTGC.TACCACG
LCcDNA : CTTCTT..........................................................................

                 *      3460         *      3480         *      3500         *      3520
zmzr1g : AAAAGACCTGTCA...TTAATA.TG.GGACCAATACATTTTTAG.CAGATTATTCT.GGCTC..TTGTGATATGGTACTA
LC     : ACTTCACCC.TCTACGTTGATACTCTGGACC..TGCCTC...AGGCATCCCGCCCTAGGCACCGTTCCCTGTGGGCCCAG
LCcDNA : ................................................................................
```

(continued on the next page)

```
                 *      3540         *      3560         *      3580         *      3600
zmzr1g : ATATTTGATATGCACACTACTTTGCACC.A.TTATTGGCCTGCATCATTGAGATTTTGATCAACTGGGGCAATTTTGTAG
LC     : CGCCC....ACACAACGTTCTCT.CAGCTACTTAATTGTGGACCCCCAGCATCCGCGCGCAGCCACCACACCACGCAAAA
LCcDNA : ........................................................................

                 *      3620         *      3640         *      3660         *      3680
zmzr1g : TCATCACCTTCGCTATCCTTGGCATCACCCTTCTTTCCCTTGAGGAACCGAATAAGGCCCGGTGTCGTCATTCCTCATAA
LC     : ATAGTGCAGAACAAACACTTGAACTCACGCTTCGTTGCTCTAGAAATTTAGGACTAACCACCAAACCACA..CATAATTT
LCcDNA : ........................................................................

                 *      3700         *      3720         *      3740         *      3760
zmzr1g : GCACTTCTGAGATCTCTGTTGGAGTCAGCATCACCTCCTTAGGTCAATGC.TTGTTTA.CCAGTACACGAGAAAATGGAA
LC     : GGTGTTTAGAAATAAACAATAATTATATTTGAATAAATATCTTAAATACAATTGGTTAGTTTGGTTTTAAGAAGAC..AT
LCcDNA : ........................................................................

                 *      3780         *      3800         *      3820         *      3840
zmzr1g : GCTGGATGCTTCTTTACCAATACACCAGCATTTACAGAACCACTAGGAGCAACAGCAGCATAC........TATGTCAAA
LC     : C..GAATG.......ATCATAGCACCAGAT..TACACTGCTGTTCCAAAAAAAAATGATATCTTATTAAAATATG...AA
LCcDNA : ........................................................................

                 *      3860         *      3880         *      3900         *      3920
zmzr1g : AACCAGTAAGTATTTGTATTCTTGACTAGTAGCATTTCTGTCATTAAGCACTGTCATACTATGTAAAAAATGATCAGTAC
LC     : AACCATACATTGTCGGCAGGTTATACTGGCGCTACTGATATGGTGCTGATATTTGATATCCACGCTACTCTGGCATCATT
LCcDNA : ........................................................................

                 *      3940         *      3960         *      3980         *      4000
zmzr1g : TACTGTACTACTTTACAATGATCAAACCATATACCACTAAAAATATTCGAAGTATTGCTATTCGTAGGGATTCCTATGTA
LC     : GAGATTTTGATCGACGACCGGGACATGTTTTTTTTTCTTTTTTCGTCGTC.ATCTCTTTCTCGTCCCCTTTGGTCTCACA
LCcDNA : ........................................................................

                 *      4020         *      4040         *      4060         *      4080
zmzr1g : GGCCTGTAATGAGTTGGGTTGAAATGACTTTAATGGGTTAGGTAGGTTCTCAATTGGGTACATTTTGGATGGGTGGGAAT
LC     : TGGTCCATTTTAACCTTGACTACTATCCTTGGCGTGGGCCTTC..TTCCCATTTAGGAACTGGATGAGACCCTCAAGCGC
LCcDNA : ........................................................................

                 *      4100         *      4120         *      4140         *      4160
zmzr1g : GGGTTGATTTACTTAGGCCTTTTTTGGATGTCCATTTATCGATCTTAATCCATGTGAAATTTAGTTTGAATTTCACACCA
LC     : GACGTCGGTGTCTTCGTTCCTCAT.GAGCAC..TTCAGCGACCTCTGCCGGGGTCACCATCACCTCCGTGATCAGCGCTT
LCcDNA : ........................................................................

                 *      4180         *      4200         *
zmzr1g : ACCCACTCCAATACATATAGGT.TGAGGTCAACATATGAGCACCCAAACAAATTCTT
LC     : CTATCTGAGGGTACGTGTCGTGGTCGGTGATGGAGTGGTAGTTGGAGGCCAGGATCC
LCcDNA : ..........................................................
```

**Legend**

**LC** : LC genomic sequence

**zmzr1g** : *zmzr1* genomic sequence

ATGC : conserved regions between the two genes

ATGC : conserved regions between *zmzr1* and LC cDNA

ATGC : translation start or stop codons

**Appendix C**: Sequence of the ~2kb *Sal* I genomic subclone of *et1-R1.1*.

```
1       10        20        30        40        50        60       bp
GTCGACATTGCAGGAGCAGGCGCGCAGCGCCAGGCCCTCCAGCGCCCGCAACTCCGTCAGCTCGC   65
CCATCATGACGCGCAACTCCTCCTACGGCCGCTCCCCCTACTCCCGCCGCCTCTCCGACTTCTCC  130
ACCTCCGACTTCACCCTCTCCATCCACGACCCGCACCACCACCACCGGACCATGGCCGACAAGCA  195
GCTGGCGTTCCGCGCCGGCGCCAGCTCCTTCCTGCGCCTCGCCAGGATGAACTCGCCCGAGTGGG  260
CCTACGCGCTCGCCGGCTCCATGGTCCGCGGCTCCTTCAGCGCCATCTTCGCCTACATCCTCAGC  325
GCCGTGCTCAGCGTGTACTACGCGCCGGACCCGCGGTACATGAAGCGCGAGATCGCCAAATACTG  390
CTACCTGCTCATCGGCATGTCCTCCGCGGCGCTGCTGTTCAACACGGTGCAGCACGTGTTCTGGG  455
ACACGGTGGGCGAGAACCTGACCAAGCGGGTGCGCGAGAAGATGTTCGCCGCCGTGCTCCGCAAC  520
GAGATCGCCTGGTTCGACACGGACGAGAACGCCAGCGCGCGCGTGGCCGCCAGGCTCGCGCTGGA  585
CGCCCAGAACGTGCGCTCCGCCATCGGGGACCGCATCTCCGTCATCGTCCAGAACTCGGCGCTGA  650
TGCTGGTGGCCTGCACCGCGGGGTTCGTCCTCCAGTGGCGCCTCGCGCTCGTGCTCCTCGCCGTG  715
TTCCCGCTCGTCGTGGGCGCCACCGTGCTGCAGAAGATGTTCATGAAGGGCTTCTCGGGGGACCT  780
GGAGGCCGCGCACGCCAGGGCCACGCAGATCGCGGGCGAGGCCGTGGCCAACCTGCGCACCGTGG  845
CCGCGTTCAACGCGGAGCGCAAGATCACGGGGCTGTTCGAGGCCAACCTGCGCGGCCCGCTCCGG  910
CGCTGCTTGTGTCAAAACAGGTATTGCGAGCAGGCCGACCCCGAGAGCAGCCGCAAGAAGAGACG  975
GCGTAGAGAGGAGTATGCCAGACGGCGTCTGCGCAGGCGAACGGTAAACGGGGACAGGAGACTAA 1040
TACCTGTTTACACAGCTCCAAGCGCTGTCTCCTTTCCGTTTTGGCGTTTGGCTGTCGTCTGCGTT 1105
TCCAGAACAGAGAAGCCAACGCCATCGCCTCCATTTCGTCGAATCCCGTCCGCTCTTCGTCTATA 1170
ATGACAATTATCTCCTGACAGGCAAAGCGGCTATGACGACAACGGCCGCCGGGTACGGCTGCTGC 1235
TGGGCGGGGATTCCGCCCTTCGCGTTGTTGCCGCGGATTCTCTCGACCCGCCGGGAGACTCCTCC 1300
TCCTCCTCCTCCTCGCGCTTCCCTTGTCGCCTCCTCCTCGAAGCTCAGGGCGCTGGCACCACGGT 1365
GCGCTCCGTTCTAGCTCGTTCGTACCGGCTCGTCTCGGTTCGCGGGTGTCTAAGCTCGTCTCTCT 1430
TCGTGCAGGCTGAGAGTTTCGAACCGTCCAAGGAGGCTCATTGTCTCCGCTTCCTCCTCCGGCGA 1495
GGCCAATTCGGACGCGGCGCCGTCGCCAACGGTCAGCGCTCTCTATCTTTCTTCTCTATGATGTT 1560
TCCAGATTGTCAGGGTGAATTGATTTGGGTTATGGATGTGCTCCGTGTTGCCTGGAATCATTCAG 1625
TGAATGCGATTATATAGGGAATGATTTGGATTGGGTAGCCACCAGCAACGGGGAAACAGTAATTG 1690
CTTGGGCTTTTGGACCACTGTTTATTATTAACCCTGAAATTGGATAACTGACGAGATTCAACTTT 1755
AATGATTCATATAGTCAGAATATTTGTGGTTTTCTAGCACCACACATCTGCTTTGTCCATTGTAT 1820
GACTGGCGTACGTTGTTCATTTGAAATTGCTAGTTCTTTCCTGAGATTGTACTTTGTGTGCATTT 1880
CTTGTAGGTCTTTCGTTGCTTGTGTGTCTTGCAGATTATGGCATTCAGAGAACATCACTTTTTGT 1950
AAATTTGGCAGGTTGATCGAGTCGAC                                          1976
```

**Legend**

| | | |
|---|---|---|
| bp | : | length in **b**ase **p**airs |
| ATGC | : | *Mu8* sequence |
| **GTCGAC** | : | *Sal*I restriction site present within *Mu8* |
| **ATGC** | : | Exons of the *et1* gene |
| ATGC | : | Introns and upstream region of the *et1* gene |

**Appendix D:** PCR primers used during this study.

| **(A)** ***Et1*** **primers : Primers from LC and B73 genomic sequences and LC cDNA clones** | | | | | |
|---|---|---|---|---|---|
| *Serial No.* | *Primer name* | *Primer sequence in 5´to 3´* | *Primer Origin and remarks* | **Primer Orientation* | *Primary Use* |
| 1. | Et 1 | GGCAAAGCGGCTATGACGAC | LC cDNA, c9.1, 5´utr | → | test PCR |
| 2. | Et 2 | TCAGAACTGGTGTCCACCGC | LC cDNA, c9.1, orf 3´ end | ← | test PCR |
| 3. | Et 7 | GCCCTGCTGTATTCAGAATTTC | LC cDNA, c9.1, exon 4, 3´utr 646-625 | ← | test PCR |
| 4. | Et 8 | CCAATGTTGTCAAAAACCTACAG | LC cDNA, c9.1, exon 4, 3´utr 570-548 | ← | test PCR |
| 5. | Et 9 | TGCTGTATTCAGAATTTCAGATGAG | LC cDNA, c9.1, exon 4, 3´utr 642-618 | ← | test PCR |
| 6. | Et 10 | TTCGAGGAGGAGGCGACAAG | LC cDNA, c9.1, exon 1 (144-124) | ← | test PCR |
| 7. | Et 11 | GCTTGGTCCTTTCGCCAC | LC cDNA, c9.1, exon 3 (334-317) | ← | test PCR |
| 8. | Et 12 | GTCGCCTCCTCCTCGAAG | LC cDNA, c9.1, exon 1 (127-144) | → | test PCR |
| 9. | Et 13 | CGCTTGATAAACAGAGTAGCCTATG | LC cDNA, c9.1, exon 3 (335-359) | → | test PCR |
| 10. | Et 16 | ATTGGGTAGCCACCAGCAACG | LC g BamHI clone, intron 2 | → | Sequencing |

(continued on the next page)

| *Serial No.* | *Primer name* | *Primer sequence in 5´to 3´* | *Primer Origin and remarks* | **Primer Orientation* | *Primary Use* |
|---|---|---|---|---|---|
| 11. | Et 17 | CCGGATCAAAAGGCCGTTGAT | LC g clone, promoter region | → | Sequencing |
| 12. | Et 18 | CATGTCCCGGTCGTCGATCA | LC g BamHI clone 3´end | ← | Sequencing |
| 13. | Et 34 | GTCGGGACCGTGACGCT | LC g clone, promoter region | → | Sequencing |
| 14. | Et 35 | CGCATTCACTGAATGATTCCA | LC g BamHI clone, intron 2 | ← | Sequencing |
| 15. | Et 36 | ACCTGAAGAAAAACTGTGCCTCT | LC g BamHI clone, intron 3/ exon 3 | ← | Sequencing |
| 16. | Et 42 | GTTGCAGTAACAGGACGATGG | LC g clone, 5´ promoter region | ← | Sequencing |
| 17. | Et 50 | ATTGATCTGCTGTGAGAAGCG | LC cDNA, c9.1, exon 4, 3´utr | → | *et1*- 3´UTR amplification |
| 18. | Et 51 | AGAACGTTGTGTGGGCGGTG | LC g BamHI clone 3´end | → | *et1*- 3´UTR amplification |
| 19. | Et 52 | GAATTCTCCGCTTCCTCCTCCGGCGAG | LC cDNA, c9.1(modified primer) | → | cDNA without tp |
| 20. | Et 53 | GAATTCAGATGAGTATGTGCAACAACTTTA | LC cDNA, c9.1(modified primer) | → | cDNA without tp |
| 21. | Et LC5´P | CATG*AAG*CTTTTGATAACTGTGAATAAC | LC g, 5´promoter end (modified primer) | → | Promoter amplification |
| 22. | Et LC3´P | GCC*CTG*CAGACAACGAGAACCGAG | LC g, 5´utr (modified primer) | ← | Promoter amplification |
| 23. | etM3P5´ | GAAG*C*TTGGAACACACATTCAATG | wt B73 g, 5´promoter end (modified primer) | → | Promoter amplification |
| 24. | etM3P3´a | GGAA*CTGC*AGAGCGAGGGTTAACTTAG | wt B73 g, promoter 3´end (modified primer) | ← | Promoter amplification |

(continued on the next page)

| *Serial No.* | *Primer name* | *Primer sequence in 5´to 3´* | *Primer Origin and remarks* | **Primer Orientation* | *Primary Use* |
|---|---|---|---|---|---|
| 25. | etM3P3´b | GAGAAC*TGCA*GGGAAAGGAAGAGCG | wt B73 g, promoter 3´end (modified primer) | ← | Promoter amplification |
| **(B) *zmzr1* primers : Primers from sequences of *et1-R* genomic and cDNA clones** | | | | | |
| 26. | Et 15 | CAAGTGCTTCATCTCGCAATTC | et1-R zmzr1 g, promoter region | → | Sequencing |
| 27. | Et 20 | ACCACACGTACCTTAGAGTT | et1-R zmzr1 g, intron 2 | ← | Sequencing |
| 28. | Et 22 | CATTCAGTACACTCTG*G*CTAC | et1-R zmzr1 g, intron 2 | ← | Sequencing |
| 29. | Et 23 | GTTGCTGCATATACTCATCTG | et1-R zmzr1 g, 3´utr | → | Sequencing |
| 30. | Et 24 | CTAAAGCCAATCAGACA*A*TGGA | et1-R zmzr1 g, intron 2 | ← | Sequencing |
| 31. | Et 26 | TGCGGTCCAAGGAGGCTCGCTTC | et1-R zmzr1 g, exon 2 | → | Sequencing |
| 32. | Et 27 | AATCACATTTTCTCCTTGTACCA | et1-R zmzr1 g, exon 4 | ← | Sequencing |
| 33. | Et 28 | ATCCTGTTGTAGCCTGTTGTC | et1-R zmzr1 g, exon 4, 3´utr | → | Sequencing |
| 34. | Et 29 | ACTATAAAATACATCTAGGCCC | et1-R zmzr1 g, exon 4, 3´utr | ← | Sequencing |
| 35. | Et 30 | GGCTTACACACCCGCGAGC | et1-R zmzr1 g, intron 1 | ← | Sequencing |
| 36. | Et 31 | CAATGATCAAACCATATACCAC | et1-R zmzr1 g, 500 bp downstream of exon 4 | → | Sequencing |
| 37. | Et 37 | CGACTCGATCACTGTGGCTGG | et1-R zmzr1 g, 5´end of promoter region | → | Sequencing |

(continued on the next page)

| *Serial No.* | *Primer name* | *Primer sequence in 5´to 3´* | *Primer Origin and remarks* | **Primer Orientation* | *Primary Use* |
|---|---|---|---|---|---|
| 38. | Et 38 | CTCACCAGTCACCTCGGTGG | et1-R zmzr1 g, promoter region | ← | Sequencing |
| 39. | Et 39 | CTGTTGTGAGAAGCAGTGTTG | et1-R zmzr1 g, 3´utr | → | 3´UTR amplification |
| 40. | Et 40 | GGTACTCTAGTATCGTGATCC | et1-R zmzr1 g, 3´utr | ← | 3´UTR amplification |
| 41. | Et 41 | GTGCAAAGTAGTGTGCATATC | et1-R zmzr1 g, 3´utr | ← | 3´UTR amplification |
| 42. | Et 43 | CTCGGTCCAAGTAGGAGTCG | et1-R zmzr1 g, 5´utr | → | cDNA amplification |
| 43. | Et 44 | GAGTCGGCGGCTATGGCG | et1-R zmzr1 g, 5´utr | → | cDNA amplification |
| 44. | Et 45 | CAGACAACAGGCTACAACAGG | et1-R zmzr1 g, 3´utr | ← | cDNA amplification |
| 45. | Et 46 | GTATATGCAGCAACTAAAATGTC | et1-R zmzr1 g, 3´utr | ← | cDNA amplification |
| 46. | Et 47 | CCACATGCATTGCATGTAAATTG | et1-R zmzr1 cDNA, exon 3 | ← | cDNA amplification |
| 47. | Et 48 | CATTGCATGTAAATTGAACAAGC | et1-R zmzr1 cDNA, exon 3 | ← | cDNA amplification |
| 48. | Et 49 | TTCAGTGTGCAGGGTGCCAG | et1-R zmzr1 cDNA, exon3/ exon 4 | → | cDNA amplification |
| 49. | zmzr1P5´ a | *AAGCTT*ATCACTGTGGCTGGGCAG | et1-R zmzr1 g, promoter (modified primer/ H3) | → | Promoter amplification |
| 50. | zmzr1P5´ b | GGCTGGG*AAGC*TTCACAAGTTCGTCG | et1-R zmzr1 g, promoter (modified primer/ H3) | → | Promoter amplification |
| 51. | zmzr1P3´ a | *CTGCAG*CCGAGGAGGCGAGGATAAAGG | et1-R zmzr1 g, promoter (modified primer/ PstI) | ← | Promoter amplification |

(continued on the next page)

| *Serial No.* | *Primer name* | *Primer sequence in 5´to 3´* | *Primer Origin and remarks* | **Primer Orientation* | *Primary Use* |
|---|---|---|---|---|---|
| 52. | zmzr1P3´ b | GATCTC*TG*CAGTCACCTCGGTGGACAAGG | et1-R zmzr1 g, promoter (modified primer/ PstI) | ← | Promoter amplification |
| **(C) *et1-R 1.1* (*et1-m3 λ5.1*) primers : Primers from the sequence of λ *et1-R 1.1* genomic clone** | | | | | |
| 53. | Et 14 | CTGTTTACACAGCTCCAAGC | λ etR1.1 (1.3 kb, SalI/ PstI) subclone | → | Sequencing |
| 54. | Et 19 | TAGCCACCAGCAACGGGGAA | λ etR1.1 (1.3 kb, SalI/ PstI) subclone | → | Sequencing |
| 55. | Et 21 | CGCTGGACGCCCAGAACGT | λ etR1.1 (2 kb, SalI) subclone | → | Sequencing |
| 56. | Et 25 | GGCAGGTTGATCGAGTCGAC | λ etR1.1 (2 kb, SalI) subclone | → | Sequencing |
| 57. | etR1.1rev | TTCTCTGAATGCCATAATCTGC | λ etR1.1 (2 kb, SalI) subclone | ← | test PCR |

∗ Key to symbols in “Primer Orientation”: → elongation in 5´ to 3´ orientation of the gene; ← elongation in 3´ to 5´ orientation of the gene.

# **Appendix E**: Cellular localisation signals for ET1 homologues

## 1. Target P V1.0 prediction results for plant ESTs /genomic sequences using PLANT networks (see Fig. 3.24 for details)

| Name | Length | cTP | Loc. | RC |
|---|---|---|---|---|
| ET1LC | 163 | 0.838 | C | 2 |
| ET1B73 | 166 | 0.799 | C | 2 |
| ZMZR1 | 162 | 0.743 | C | 2 |
| At.putTF | 170 | 0.786 | C | 3 |
| At.putPr | 212 | 0.937 | C | 1 |
| TaEST | 154 | 0.922 | C | 1 |
| HvESTsedl | 116 | 0.843 | C | 2 |
| LeESTroot | 153 | 0.916 | C | 1 |
| LeESTelcpl | 200 | 0.868 | C | 2 |
| MtESTinfcL | 151 | 0.760 | C | 3 |
| MtEST1mRN | 153 | 0.701 | C | 4 |

## 2. Target P V1.0 prediction results for non-plant proteins of the protein domain, PD022385 using NON-PLANT networks (see Fig. 3.10 for details)

| Name | Length | mTP | Loc. | RC |
|---|---|---|---|---|
| Dm | 191 | 0.815 | M | 3 |
| Ce | 119 | 0.894 | M | 2 |
| Mm | 177 | 0.797 | M | 2 |
| Sc* | 205 | 0.594 | _ | 5 |
| Sp | 175 | 0.875 | M | 2 |

**TargetP V1.0 Abbreviations**:

**cTP** : chloroplast transit peptide

**mTP** : mitochondrial transit peptide

**Loc.**: cellular location

**RC** : Reliability Class (1>2>3>4>5)

***** : the **Sc** protein from protein domain, PD022385 (see below)

## 3. MitoProt II 1.0a4 prediction results for mitochondrial localisation.

| Name | Length | Probability |
|---|---|---|
| Sc* | 205 | 0.9930 |

***** : the **Sc** protein from protein domain, PD022385

# Acknowledgements

I can hardly thank Prof. Dr. Udo Wienand and Dr. Oswaldo da Costa e Silva enough for giving me the opportunity to work on this exciting project. I also thank Prof. Dr. Udo Wienand for his supervision as well as guidance throughout my work.

I thank Dr. Oswaldo da Costa e Silva for introducing me to molecular biology in practice and for his guidance, which, although lasted for a short period, was very motivating. At this place, I must also thank Dr. René Lorbiecke for his invaluable guidance and for the many discussions during the later part of my work.

My thanks are also due to Dr. Patricia Lauert and Dr. Karsten Frenzel for their constant support and motivation throughout my time in the lab. I am thankful to Ulrike Peters, Edelgard Brinkmann and Sylke Wahlandt for the technical support.

I deeply thank Prof. Dr. Ralf Bernd Klösgen and Dr. Jürgen Berghöfer at the University of Halle for helping me with the Chloroplast Import experiments.

I am grateful to Dr. Brian Scheffler for agreeing to be the “native speaker” and for the onerous task of correcting this thesis. I also thank Dr. Ali Naseer Eddine and Dr. Patricia Lauert for helping me with the grammatical corrections in this thesis. I am thankful to Prof. Dr. Horst Lörz for agreeing to review this thesis.

My hearty thanks to all my seniors and colleagues at AMP I for the friendly and co-operative environment, which made these years a lifetime experience.

I thank my sisters for their encouragement, and my husband Ajit for his unfailing support and understanding throughout this work. I also thank my father-in-law Dr. O. P. Garg for his financial as well as moral support. Finally, I am deeply thankful to my parents for making my earlier studies possible and for the encouragement to carry on with my work.