DNA RESEARCH 5, 95–101 (1998)

Isolation and Mapping of a Family of Putative Zinc-finger Protein cDNAs from Rice

Jianyu SONG,[1] Kimiko YAMAMOTO,[1] Ayahiko SHOMURA,[1] Hiraku ITADANI,[1] Hui Sun ZHONG,[1] Masahiro YANO,[2] and Takuji SASAKI[2,*]

Rice Genome Research Program, Institute of the Society for Techno-Innovation of Agriculture, Forestry and Fisheries, 446-1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305, Japan[1] *and National Institute of Agrobiological Resources, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305, Japan*[2]

(Received 19 January 1998; revised 31 March 1998)

Abstract

To understand the functions of rice homologues of the *Arabidopsis* flowering-time gene *CONSTANS* (*CO*) and salt-tolerance gene *STO*, we performed a similarity search of the single-run sequence data of cDNA clones accumulated by the Rice Genome Research Program, and isolated seven rice cDNA clones (S3574, C60910, S12569, R2931, R1479, R1577, and E10707) coding for proteins containing one or two zinc-finger-like motifs. Comparison of the deduced amino acid sequences between these cDNAs and the *CO* gene revealed significant similarities (46%-61%) in the region of zinc-finger motifs. A domain having a high content of basic amino acids at the C-terminus of the CO protein was found in the corresponding region of proteins predicted from cDNAs S3574, C60910, and S12569. Two amino acid sequences, "CCADEAAL" and "FCV(L)EDRA," which were present inside each zinc-finger in the *Arabidopsis* regulatory protein STO, were also found in each of the two zinc-finger regions of proteins predicted from cDNAs R2931, R1479, R1577, and E10707. Using restriction fragment length polymorphism (RFLP) linkage analysis, we determined the chromosomal location of the seven cDNA clones. The position of R2931 on the RFLP linkage map was closely linked to *Hd-3*, one of the putative quantitative trait loci (QTL) controlling heading date in rice.

Key words: zinc-finger protein; transcription; RFLP mapping; rice; *Oryza sativa* L.

1. Introduction

Zinc-finger proteins contain finger structures organized around zinc (II) metal ions, which are required to maintain the folded structure of zinc-finger peptides. Many zinc-finger proteins are known to be involved in transcriptional regulation and developmental control. The GATA-1 protein family is one of the families of Cys_2/Cys_2-type zinc-finger transcription factors; a "$Cys\text{-}x_2\text{-}Cys\text{-}x_{17}\text{-}Cys\text{-}x_2\text{-}Cys$" sequence has been defined as the DNA-binding domain of the GATA-1 family.[1,2] A member of the GATA-1 family, NTL1 from tobacco, is the first reported case in plants.[3] The predicted protein sequence of NTL1 shows similarity to NIT2, a fungal transcription factor that has been characterized as one of the regulatory elements of the nitrate assimilation pathway.[4]

The *CONSTANS* (*CO*) gene in *Arabidopsis* was isolated by using a map-based cloning strategy. Its protein product has been identified as a putative transcriptional factor promoting flowering.[5] The CO protein contains two zinc-finger motifs that show sequence similarities to those of members of the GATA-1 family. The *COL2* gene of *Arabidopsis* has been identified as a homologue of *CO*.[6] The *STO* gene of *Arabidopsis* has been implicated in salt tolerance in a yeast calcineurin mutant.[7] The similarity of the putative zinc-finger motifs in the STO protein to the zinc-finger motifs in the CO protein indicates that STO may be a GATA-1-like protein.[8]

Rice has been become a good model plant for genome research of cereals and for isolation of agronomically important genes, owing to its relatively small genome size (430 Mb). In the Rice Genome Research Program (RGP), we have conducted large-scale sequencing of cDNAs randomly selected from various kinds of libraries with the aim of cataloguing all expressed rice genes. In this paper, we describe the isolation of seven kinds of cDNAs that show similarities to the *Arabidopsis* flowering-time gene *CO* and salt-tolerance gene *STO*. We also compare the results of linkage analyses of the seven genes with those of QTL analyses for heading date[9] and salt tolerance.[10]

Communicated by Mituru Takanami

[*] To whom correspondence should be addressed. Tel. +81-298-38-7441, Fax. +81-298-38-7468, E-mail: tsasaki@abr.affrc.go.jp

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022

2. Materials and Methods

2.1. *Identification and sequencing of cDNAs coding for putative zinc-finger proteins*

cDNA clones of the *CO* and *STO* homologues were obtained by similarity search of partial-sequence data accumulated through large-scale cDNA analysis against the CO and STO protein sequences using the Basic Local Alignment Search Tool (BLAST) algorithm.[11] Based on the sequences of 3′-untranslated regions, the cDNAs were classified into seven independent groups, and the full nucleotide sequences of the longest cDNAs in each group were determined by the shotgun method. DNA fragments of these cDNAs were produced by sonication and subcloned into pBluescript II SK(+). The complete nucleotide sequences were determined for both strands with a thermal cycle sequencing kit (Perkin Elmer, CA, USA) by the use of a combination of a chemical robot, a CATALYST (Perkin Elmer), and an automated sequencer, Model 373A (Perkin Elmer).

2.2. *Restriction fragment length polymorphism mapping*

Genomic DNA was prepared from leaves of 186 rice F_2 plants and their parent lines, *japonica* variety Nipponbare and *indica* variety Kasalath, by the cetyltrimethylammonium bromide (CTAB) method.[12] DNA from the parent plants was digested with one of eight restriction enzymes: *Bam*HI, *Bgl* II, *Eco*RV, *Hin*dIII, *Apa* I, *Dra* I, *Eco*RI, or *Kpn* I. Each digested DNA (2 μg per lane) was electrophoresed on a 0.6% agarose gel and transferred onto a positively charged nylon membrane (Boehringer, Mannheim, Germany) as described in detail by Kurata et al.[13] Southern hybridization and detection were performed according to the protocol of the Enhanced Chemiluminescence (ECL) detection system kit (Amersham, Buckinghamshire, UK). All probes were amplified by polymerase chain reaction (PCR) by using the gene-specific 3′ untranslated region of the cDNA sequences as templates. F_2 DNAs were digested with the restriction enzyme corresponding to the restriction fragment length polymorphism (RFLP) detected between the two parent lines, and were analyzed by Southern hybridization. The linkage analyses were performed with the MAPMAKER/EXP 3.0 computer package,[14] based on a high-density linkage map, which included 1300 DNA markers, constructed by Kurata et al.[13]

3. Results and Discussion

We analyzed the similarity of the partial sequence data, which has been accumulated through large-scale cDNA analysis, against *Arabidopsis* CO and STO, and obtained 18 cDNA clones encoding for putative zinc-finger proteins. Gene-specific 3′ untranslated sequences of these clones were further analyzed, and the clones were classified into seven independent groups. The following seven clones, which were the longest cDNA sequences in each group, were selected for more detailed study: S3574, derived from etiolated seedlings; S12569, derived from green seedlings; R2931, R1479 and R1577, derived from the 9-day-old root cDNA library; C60910, derived from the cDNA library of heat-shocked callus; and E10707, derived from the cDNA library of panicles at the ripening stage.

The complete nucleotide sequences of the seven cDNAs were determined, and the predicted protein sequences were deduced from the sequences (Fig. 1). The nucleotide sequence data have been submitted to the EMBL, GenBank and DDBJ nucleotide sequence databases under accession numbers AB001882 (C60910), AB001883 (E10707), AB001884 (R1479), AB001885 (R1577), AB001886 (R2931), AB001887 (S12569), and AB001888 (S3574).

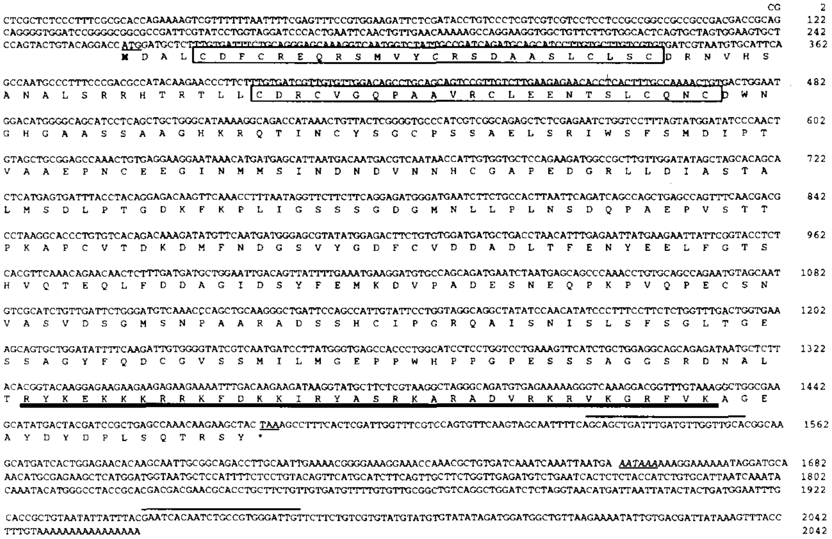
Translation was postulated to start at the first methionine in each cDNA sequence. In-frame stop codons preceding the translations were found in C60910, R2931, R1577, and E10707. Thus, it is likely that these clones are full-length cDNAs. The typical polyadenylation signal "AATAAA" was found only in the 3′-untranslated regions of S3574 and R2931. Analysis of these predicted protein sequences revealed that two cysteine cores in each finger were interrupted by 15 or 16 amino acid residues with a consensus sequence of "Cys-x_2-Cys-$x_{16(15)}$-Cys-x_2-Cys." It has been reported that the zinc-finger domain of GATA-1 transcription factors is in a "Cys-x_2-Cys-x_{17}-Cys-x_2-Cys" arrangement.[2] This implies that the seven cDNAs in this study comprise a GATA-1-like gene family in rice.

The predicted protein sequence of C60910 contained a glutamic acid-rich domain "EVEVVEEEEE" in the N-terminus and an alanine-rich domain "AAAATAAA" following the zinc-finger motifs (Fig. 1b). In S12569, the zinc-finger domain was followed by an acidic domain "DYDDDDADAAGEEDEE" (Fig. 1c).

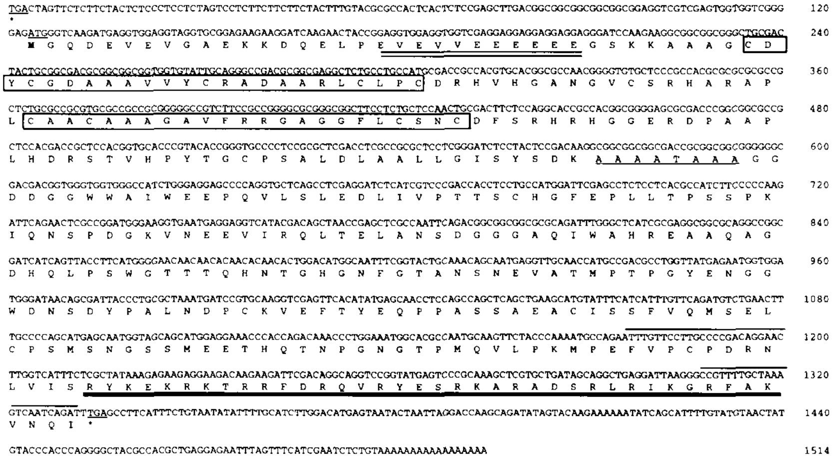
Figure 1. Nucleotide sequences and deduced amino acid sequences of rice cDNA clones encoding putative zinc-finger proteins. The nucleotides are numbered on the right. The polyadenylation signal "AATAAA" is underlined and in *italic* type. The deduced amino acid sequences are shown below the nucleotide sequences in one letter codes. Potential initiating methionines are in **bold** type. The in-frame stop codons preceding the translations are indicated by asterisks (*). The one or two Cys_2/Cys_2-type zinc-finger motifs in each protein are boxed. The acidic and basic regions are indicated by double underlines and bars, respectively. Underlined regions are explained in the text. Lines over the nucleotide sequences indicate the PCR primer sequence used to generate gene-specific sequences for Southern hybridization. The 5′ primer is the sequence indicated by a line above the nucleotides; the 3′ primer is the complement of that sequence.

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022

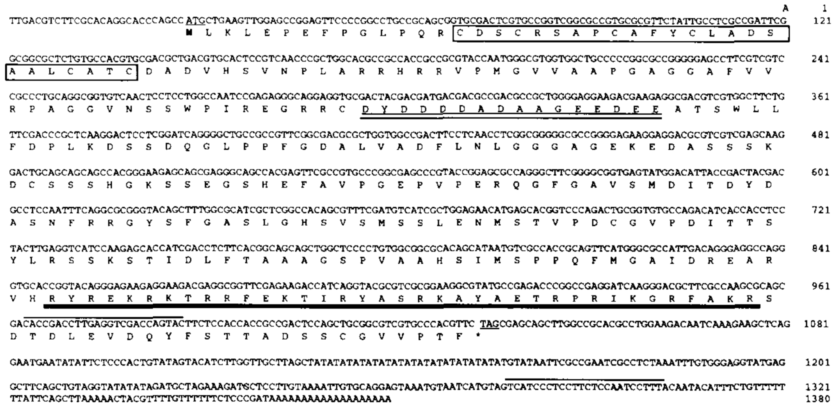
(a). S3574



(b). C60910



(c). S12569



Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022

(d). R2931

```
                                                                                                                     G   1
AGGGAGCGATCCAATCGCGTCGCTAAAACCCTAGCGGCTGGC TGATTTGGGGATTGCTCGTCGGAGAGATGAAGATCCAGTGCAACGCGTGCGGCGCGGCGGAGGCGCGGGTGCTGTGC  121
                                           *                         M  K  I  Q  C  N  A  C  G  A  A  E  A  R  V  L  C

TGCGCCGACGAGGCAGCGCTCTGCACGGCGTGCGACGAGGAGGTGCACGCCGCCAACAAGCTCGCCGGGAAGCACCAGCGGGTGCCGCTGCTCTCCGACGACGGCGGCGCCGCGCCCGCC  241
C  A  D  E  A  A  L  C  T  A  C  D  E  E  V  H  A  A  N  K  L  A  G  K  H  Q  R  V  P  L  L  S  D  D  G  G  A  A  P  A

GCCGCCGCCCCGGCCGTGCCCAAGTGCGACATCTGCCAGGAGGCTTCTGGATACTTCTTCTGCCTGGAGGACCGTGCACTTCTTTGCAGAGATTGTGATGTTTCTATACACACAGTAAAC  361
A  A  A  P  A  V  P  K  C  D  I  C  Q  E  A  S  G  Y  F  F  C  L  E  D  R  A  L  L  C  R  D  C  D  V  S  I  H  T  V  N

TCCTTTGTTTCAGTACACCAAAGATTCCTACTAACAGGTGTTCAAGTTGGCCTTGATCCTGCTGATCCAGTTCCACCTGTTGCTGACAAGCATGTTAAGAGTGCTGGTGGTTCAGTGGAT  481
S  F  V  S  V  H  Q  R  F  L  L  T  G  V  Q  V  G  L  D  P  A  D  P  V  P  P  V  A  D  K  H  V  K  S  A  G  G  S  V  D

TCAGCAACTAAACATTTGCAAAGGAATCCTACAGACTTATCTGGTGAAAACAGTGCATCTTTGCCCAGCCAAAATGTAATCAATGGTAATTATTCTAGGCAGAGTTCTGTTACAATGGCC  601
S  A  T  K  H  L  Q  R  N  P  T  D  L  S  G  E  N  S  A  S  L  P  S  Q  N  V  I  N  G  N  Y  S  R  Q  S  S  V  T  M  A

AAGACAGGACAGGTCAATTGGACTATGAGCAACAACACAATTAGATCAATAGACCCTCCACCCAAGTATTCATCAGAGGAAAGTCCAGCACTTCTGCTAGCTAGCCACACTAGCACCATG  721
K  T  G  Q  V  N  W  T  M  S  N  N  T  I  R  S  I  D  P  P  P  K  Y  S  S  E  E  S  P  A  L  L  L  A  S  H  T  S  T  M

GCAGCGTACTCCAGTCAAATCAGTAAGGATAGTGATCGGATCTACAACTTACCATTCACAGGTGGTAATGGGTCAGATAGTCTACATGATTGGCATGTTGATGAGTTCTTTAGTAACTCA  841
A  A  Y  S  S  Q  I  S  K  D  S  D  R  I  Y  N  L  P  F  T  G  G  N  G  S  D  S  L  H  D  W  H  V  D  E  F  F  S  N  S

GAATTTGGCTTTGCTGAGCATGGTTCTTCTAAGGGTGACAACGCTAAGCCAGGGAGTGCTGGTGGATCTCCGCAGTGCCGTCTGGCTGAAGGCCTGTTTGTCGAAGGACTTCTAGGTCAA  961
E  F  G  F  A  E  H  G  S  S  K  G  D  N  A  K  P  G  S  A  G  G  S  P  Q  C  R  L  A  E  G  L  F  V  E  G  L  L  G  Q

GTGCCTGACAATCCATGGACAGTGCCTGAGGTCCCCTCGCCACCGACAGCCTCTGGTCTCTATTGGCAAAATAATTTGCTTTGCCCTTCGTACGACAGCACCATGTTCGTCCCTGAGATT  1081
V  P  D  N  P  W  T  V  P  E  V  P  S  P  P  T  A  S  G  L  Y  W  Q  N  N  L  L  C  P  S  Y  D  S  T  M  F  V  P  E  I

TCCTCCTTGGAGAACTCTCAGAACAACTTCACTGTATCTGCTGGTTTGAAGCGCCGAAGGAGGCAGTTT TGATCTGCCGATAGCATCAGACACCATTCGGTTTAGTTGGCAGGTCACGTT  1201
S  S  L  E  N  S  Q  N  N  F  T  V  S  A  G  L  K  R  R  R  R  Q  F   *

GAGAGGAGGGCTTCTTGCCTGCCTGACCAAAGCCACAAAACACAGGGATGGATGATCATCCAGTGGTCAGTCTAAAAATATGTTGTAATGTTTGTGTTTCGACTATACGTTTGTTGCTGA  1321

AGTTTAAAGGCGAGGGAATTGTTGTGAGATCTGGGATGACAGATTTGTTGTTACCAATTTTGTGAATAGTGAA AATAAAAGAGCATGGATAAGATCAAAAAAAAAAAAAA  1433
```

(e). R1479

```
                                                                                                                        GC   2
CTAGCTCGTCGATCGCCGGCCATGCGGGTGCAGTGCGACGTCTGCGCCGCCGAGCCGGCCGCGGTGCTCTGCTGCGCCGACGAGGCCGCGCTCTGCTCCGCCTGCGACCGCCGCGTCCAC  122
                     M  R  V  Q  C  D  V  C  A  A  E  P  A  A  V  L  C  C  A  D  E  A  A  L  C  S  A  C  D  R  R  V  H

CGCGCCAACCGCCTCGCCAGCAAGCACCGCCGCCTCCCGCTCGTCCACCCGTCCTCCTCCTCCGGCGACGGTGGCGCCGCCGCCGCGCCGCTGTGCGACGTGTGCAGGGAGAAGAGG  242
R  A  N  R  L  A  S  K  H  R  R  L  P  L  V  H  P  S  S  S  S  S  G  D  G  G  A  A  A  A  P  L  C  D  V  C  R  E  K  R

GGCCTCGTGTTCTGCGTCGAGGACCGCGCCATCCTGTGCGCCGACTGCGACGAGCCCATCCACTCCGCCAACGACCTCACCGCCAAGCACACCCGCTTCCTCCTCGTCGGCGCCAAGCTC  362
G  L  V  F  C  V  E  D  R  A  I  L  C  A  D  C  D  E  P  I  H  S  A  N  D  L  T  A  K  H  T  R  F  L  L  V  G  A  K  L

TCCCCCGCCGCGCTCGCCGAACAGCCGCTCCCGTCCTCCGACTGCAGCTCCGACGACGACGCCGCCGCCGCCGCCACCGAGGAGGAGTACCACTCCTCCGCCGCATCCACCGGCGCCGCA  482
S  P  A  A  L  A  E  Q  P  L  P  S  S  D  C  S  S  D  D  D  A  A  A  A  A  T  E  E  E  Y  H  S  S  A  A  S  T  G  A  A

GTGAGCGCGCCTCTTGACGCCTCCAGCAACGGCGCCGGTGGCGGAGGAGGAGTAGGAGGGAGCAGCATCTCCGACTACCTCACCACCATCTGCCCCGGCTGGCGCGTCGAGGACCTCCTC  602
V  S  A  P  L  D  A  S  S  N  G  A  G  G  G  G  G  V  G  G  S  S  I  S  D  Y  L  T  T  I  C  P  G  W  R  V  E  D  L  L

CCCGACGACGACGCCTTCGCCGCCGCCGCCGCGCAGGCGGGGAAAGAGAAGGACGAGCGCGTGCCGTTCCTCGACGCCGACCTGTTCGACGTGGTCGCCGGCCGGCCGGAGAAGAAGGGC  722
P  D  D  D  A  F  A  A  A  A  A  Q  A  G  K  E  K  D  E  R  V  P  F  L  D  A  D  L  F  D  V  V  A  G  R  P  E  K  K  G

GGCGCGTGGGCGCCGCACGTGCCGCACCTGCCGGCGTGGTGCCTCGACGAGGTGCCGGTCGTCGTCGCCGCGTCGGCGGCGCCAGCGGCGACACCTGTAAAGGCGAAGCAGGGACACGTG  842
G  A  W  A  P  H  V  P  H  L  P  A  W  C  L  D  E  V  P  V  V  V  A  A  S  A  A  P  A  A  T  P  V  K  A  K  Q  G  H  V

CGGGACAGCCACTGGAGCGACAGCGACGCGTTCGCCGTGCCGGAGTTCTCGCCGCCGCCGCCGCCGGCCAAGAGGGCGCGGCCCAGCTCGCAGTTCTGGTGCTTC TGAACGCCGCCGCCG  962
R  D  S  H  W  S  D  S  D  A  F  A  V  P  E  F  S  P  P  P  P  P  A  K  R  A  R  P  S  S  Q  F  W  C  F   *

CCGTACAATCTGGCCGCCGGCGAGCTCGACGAACACCTCGATGGATCGATCGAACACCTGTGTAATCACTCAGCTGGTCACTACTACAAGTAGCAATCAGCAAGCTTGATGTGCCACTGT  1082

TACTAGCTGCTGCTGTTTAATTAGTGAGTACTAAGCTACTACTAGTTAATTTTGGTGAAAACTATGCTACTTTGTTACTATATAATGAAGTATGTGTATCAAACTAAACTGAGATGTTGG  1202
TAATTAATTATCGTATTTTGTAAAAAAAAAAAAAAAAAAAA                                                                                   1242
```

(f). R1577

```
                                                                                                                        CC   2
TTCTTCTGCCGCTCGTTTTCCATTGCCGACGCGCACACGCAGCTGCGCAGCCACACCTTTGCGGCTGTCTTATCTTGTTCT TGAGTTGTGTTCTTGGTTGGGGAGGAGGAGGAGGAG ATG  122
                                                                                   *                                     M

AAGATCCAGTGCGACGCGTGCGAGAGCGCGGCGGCGGCGGTGGTGTGCTGCGCGGACGAGGCGGCGCTGTGCGCGGCGTGCGACGTGGAGGTGCACGCGGCGAACAAGCTGGCCGGGAAG  242
K  I  Q  C  D  A  C  E  S  A  A  A  A  V  V  C  C  A  D  E  A  A  L  C  A  A  C  D  V  E  V  H  A  A  N  K  L  A  G  K

CACCAGCGGCTGCCGCTGGAGGCGCTCTCGGCGAGGCTCCCGCGCTGCGACGTGTGCCAGGAGAAGGCGGCGTTCATCTTCTGCGTGGAGGACCGCGCGCTCTTCTGCCGCGACTGCGAC  362
H  Q  R  L  P  L  E  A  L  S  A  R  L  P  R  C  D  V  C  Q  E  K  A  A  F  I  F  C  V  E  D  R  A  L  F  C  R  D  C  D

GAGCCCATCCACGTCCCCGGCACGCTCTCCGGCAACCACCAGCGCTACCTCGCCACCGGCATCCGCGTCGGCTTCGCCTCCGCCTCGCCCTGCGACGGCGGCAGCGACGCCCATGACTCC  482
E  P  I  H  V  P  G  T  L  S  G  N  H  Q  R  Y  L  A  T  G  I  R  V  G  F  A  S  A  S  P  C  D  G  G  S  D  A  H  D  S

GACCACCACGCCCCGCCCATGGGCTCCTCCGAGCATCATCACCATCATCAGCAGCCGGCCCCGACCGTCGCCGTCGACACGCCCTCGCCGCAGTTCCTGCCGCAGGGCTGGGCCGTCGAC  602
D  H  H  A  P  P  M  G  S  S  E  H  H  H  H  H  Q  Q  P  A  P  T  V  A  V  D  T  P  S  P  Q  F  L  P  Q  G  W  A  V  D

GAGCTCCTCCAGTTCTCCGACTACGAGACCGGCGACAAGCTGCAGAAGGAGTCGTCGCCGCCGCTCGGGTTCCAGGAGCTGGAGTGGTTCGCCGACATCGACCTGTTCCACAACCAGGCG  722
E  L  L  Q  F  S  D  Y  E  T  G  D  K  L  Q  K  E  S  S  P  P  L  G  F  Q  E  L  E  W  F  A  D  I  D  L  F  H  N  Q  A

CCCAAGGGCGGCGCCGCCGCCGGCCGGACGACGGCGGAGGTCCCCGAGCTCTTCGCTTCGCAGGCGGCCAACGACGTGGCGTACTACAGGCCGCCGACCAGGACCGCCGCCGCCGCCTTC  842
P  K  G  G  A  A  A  G  R  T  T  A  E  V  P  E  L  F  A  S  Q  A  A  N  D  V  A  Y  Y  R  P  P  T  R  T  A  A  A  A  F

ACCGCGGCCACCGGCTTCCGCCAGAGCAAGAAGGCCCGCGTCGAGCTCCCCGACGACGAGGAGGATTACCTCATCGTCCCTGATCTTGGT TGAATCTTGAAGAGGAGTTCGTTCCAGAGG  962
T  A  A  T  G  F  R  Q  S  K  K  A  R  V  E  L  P  D  D  E  E  D  Y  L  I  V  P  D  L  G   *

AGAAGATCCTCCTCCTGTGTTCTTGGGTTTTGCCATGTCCTTGTGTTTTCTGCTTGTATGCTAGCTTTTGTGTTTAACGTTAGTCCTGCTCAACTGTGACATATTTGAGCCACTCTGCTT  1082

CCGTCCCATGCCATCAGTGATCAATCTGATATAAGTTTTTCTGAAAATGAGCTCCAGGCTCTGCTCTTATGTTTAATCTTCCGAGATTACCAGTACAGAAAGAGATTACCAGTACAAAAA  1202
AAAAAAAAAA                                                                                                                  1212
```

(g). E10707

```
GTTTTGAGGGGAGTACATCGGCATGAGGATCCAGTGCGACGCGTGCGAGGCCGCGGCGGCCACGGTGGTGTGCTGCGCGGACGAGGCGGCGCTGTGCGCGCGCTGCGACGTCGAGATCCA  120
     *                M  R  I  Q  C  D  A  C  E  A  A  A  A  T  V  V  C  C  A  D  E  A  A  L  C  A  R  C  D  V  E  I  H

CGCCGCCAACAAGCTCGCCAGCAAGCACCAGCGCCTCCCGCTCGACGCCGCGCTCCCCGCCGCCCTCCCGCGCTGCGACGTCTGCCAGGAGAAGGCGGCGTTCATCTTCTGCGTGGAGGA  240
 A  A  N  K  L  A  S  K  H  Q  R  L  P  L  D  A  A  L  P  A  A  L  P  R  C  D  V  C  Q  E  K  A  A  F  I  F  C  V  E  D

CAGGGCGCTCTTCTGCCGGGACTGCGACGAGCCCATCCACGTCCCGGGGACGCTCTCCGGCAACCACCAGCGCTACCTCACCACCGGCATCCGCGTCGGGTTCAGCTCCGTCTGTAGCGC  360
 R  A  L  F  C  R  D  C  D  E  P  I  H  V  P  G  T  L  S  G  N  H  Q  R  Y  L  T  T  G  I  R  V  G  F  S  S  V  C  S  A

CAACGCCGACCACCTCCCGCCGCCAGCGCCCAAGGGGAACTCCAAGCCGCCGGCAAGCGGCATCGCTGCTGCTGCTGCTCCCAAGCCGGCCGTGTCCGCGGCGGCGCAGGAGGTGCCGTC  480
 N  A  D  H  L  P  P  P  A  P  K  G  N  S  K  P  P  A  S  G  I  A  A  A  A  A  P  K  P  A  V  S  A  A  A  Q  E  V  P  S

GTCACCGTTCTTGCCGCCGTCGGGCTGGGCCGTCGAGGATCTCCTGCAGCTCTCCGACTACGAGTCCAGCGACAAGAAGGGCTCTCCTATTGGGTTCAAGGATCTGGAGTGGCTCGATGA  600
 S  P  F  L  P  P  S  G  W  A  V  E  D  L  L  Q  L  S  D  Y  E  S  S  D  K  K  G  S  P  I  G  F  K  D  L  E  W  L  D  D

CATCGACCTGTTCCATGTCCAGTCGCCGGCCAAGGGAGGCAGCACGGCGGCGGAGGTGCCTGAGCTCTTCGCCTCGCCGCAGCCAGCGAGCAACATGGGGCTCTACAAGGCGAGCGGTGC  720
 I  D  L  F  H  V  Q  S  P  A  K  G  G  S  T  A  A  E  V  P  E  L  F  A  S  P  Q  P  A  S  N  M  G  L  Y  K  A  S  G  A

ACGCCAAAGCAAGAAGCCACGGGTGGAGATACCCGATGACGACGAGGACTTCTTCATCGTTCCTGATCTTGGA TGAGATGTTGTTATGTAACAGTAATGCTATGCAAGTGATGTAATTTT  840
 R  Q  S  K  K  P  R  V  E  I  P  D  D  D  E  D  F  F  I  V  P  D  L  G   *

GTATGTATGCAAAGAACACTGCCATCTCGGTTGTACTAAGTGTGCTAAGTACATCTTACTGTGGGCGTATGTAATTGTAAGACAATTTGGGCCCAATACTTATCAGCATGGTGTTGGGCA  960

TGAACTATGATGTACTATGACATGACGAAAAAAAAAAAAAAAAA                                                                                1004
```

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022

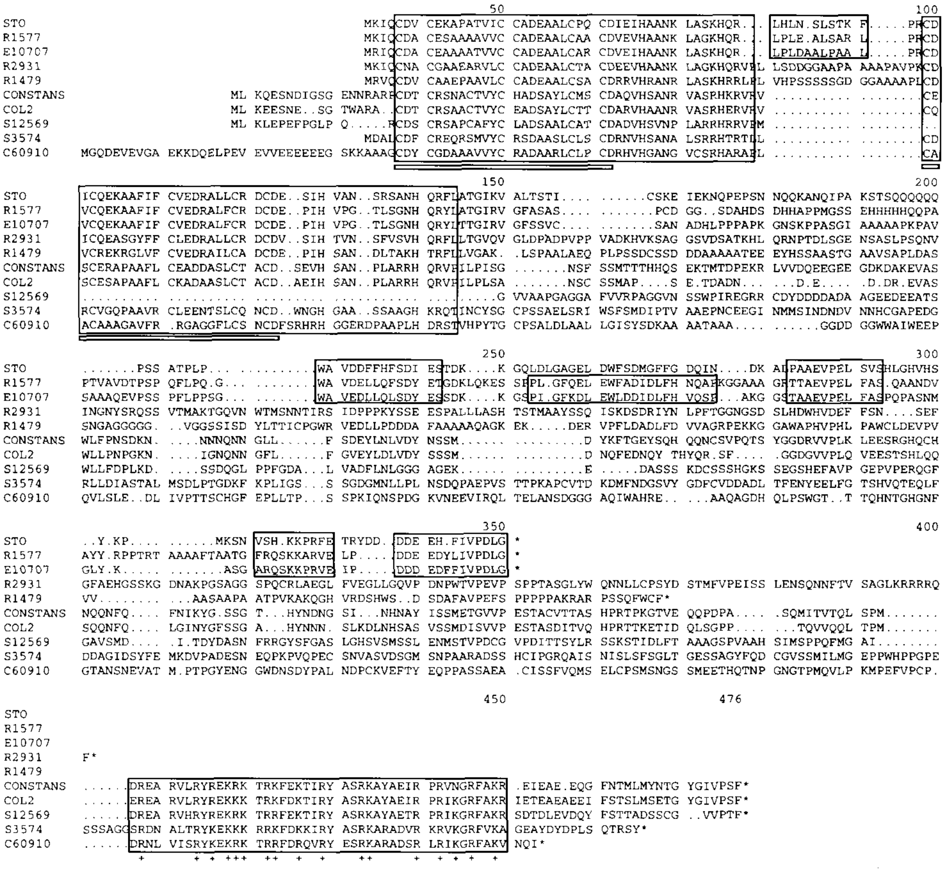


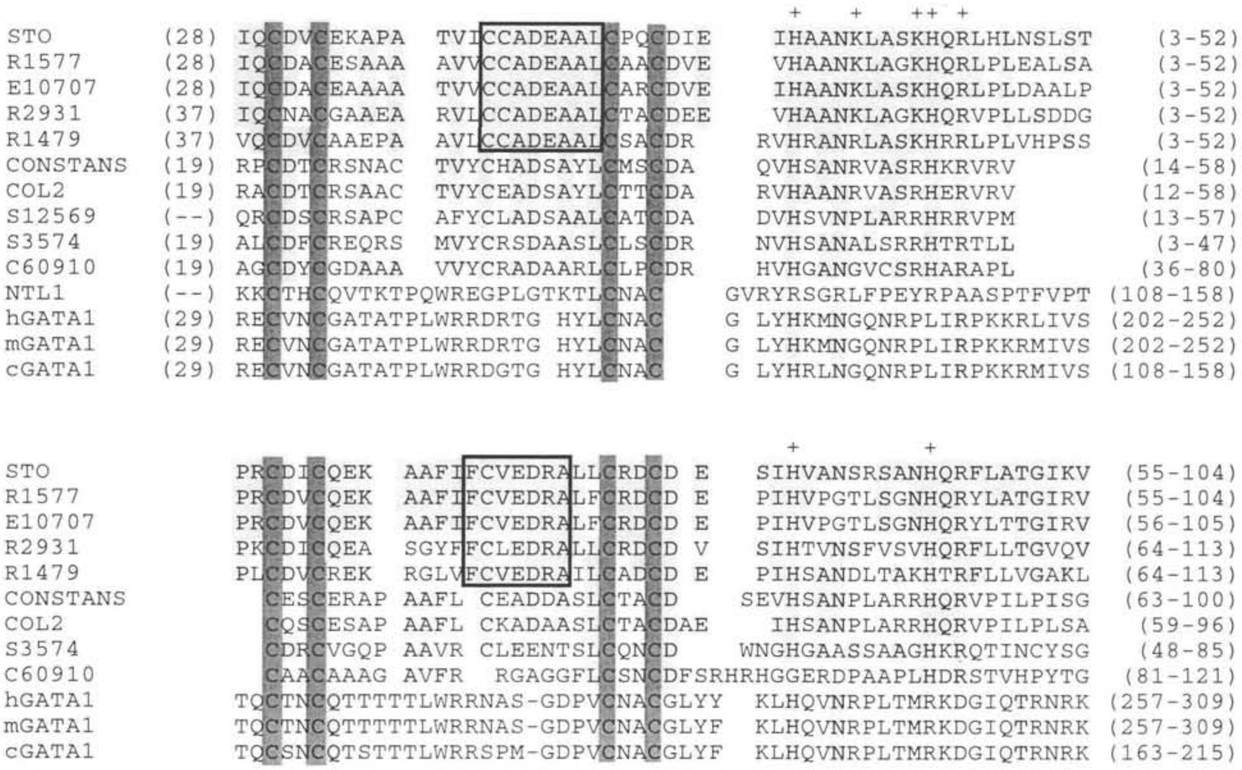
Figure 2. A comparison of 10 amino acid sequences of Cys_2/Cys_2-type zinc-finger proteins with pairwise alignment using the PILEUP program of the GCG package.[18] The protein sequences of S3574, C60910, S12569, R2931, R1479, R1577, and E10707 (this paper) were predicted from the cDNA sequences starting at the first methionine. Those of STO, CONSTANS, and COL2 were obtained from GenBank (accession numbers X94937, L81119, and X95572, respectively). Boxed amino acid sequences indicate small regions of homology. Dots are used to optimize alignment. The zinc-finger motifs are indicated by open bars. Positively charged amino acids in the C-terminus are marked with a "+."

The predicted protein sequence of R1479 contained glycine and alanine patches in the middle, and had a proline-rich domain, which is suggested to be a transcriptional activation domain for transcription factors,[15] in the C-terminus (Fig. 1e). The predicted protein sequence of R1577 contained an opa-like domain "HHHHHQQ" in the C-terminal side of the fingers (Fig. 1f). The opa-like domain has been reported to be present in a number of plant transcription factors.[16]

The amino acid sequences of seven putative zinc-finger proteins in rice and three known zinc-finger proteins in *Arabidopsis* (CO, STO, and COL2) were compared (Fig. 2). The amino acid sequences of the zinc-finger regions of the seven rice proteins are similar to those of the proteins CO (46%-61% identical) and STO (38%-86% identical). A basic region was found in the C-terminus of CO and STO, but there is no amino acid sequence similarity between the two proteins in this region. A basic region was also observed near the C-terminus of all seven predicted rice proteins. A high sequence similarity (58%-88% identical) in the basic region was found among the S3574, C60910, S12569, and CO proteins. The amino acid sequences of R1577, E10707, and STO also had significant similarity within their C-terminal basic regions. We suspect that the basic region could be a functional region related to flowering-time and salt tolerance.

A comparison of the seven predicted protein sequences with members of the GATA-1 and GATA-1-like families (Fig. 3) revealed that significant sequence similarity exists in the zinc-finger domains. There were 19 to 37 amino acid residues between the two fingers, except in S12569 and NTL1, which have only one zinc-finger motif. The zinc-finger domain of the GATA-1 transcription factors is followed by a region having a high content of basic residues; this region is required for the interaction of the protein with DNA.[2] The basic region following the zinc-finger domain was also found in the seven predicted proteins reported here. In addition, among the amino

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022



```
                                                   +   +   ++ +
STO       (28) IQCDVCEKAPA  TVICCADEAALCPQCDIE    IHAANKLASKHQRLHLNSLST    (3-52)
R1577     (28) IQCDACESAAA  AVVCCADEAALCAACDVE    VHAANKLAGKHQRLPLEALSA    (3-52)
E10707    (28) IQCDACEAAAA  TVVCCADEAALCARCDVE    IHAANKLASKHQRLPLDAALP    (3-52)
R2931     (37) IQCNACGAAEA  RVLCCADEAALCTACDEE    VHAANKLAGKHQRVPLLSDDG    (3-52)
R1479     (37) VQCDVCAAEPA  AVLCCADEAALCSACDR    RVHRANRLASKHRRLPLVHPSS    (3-52)
CONSTANS  (19) RPCDTCRSNAC  TVYCHADSAYLCMSCDA    QVHSANRVASRHKRVRV        (14-58)
COL2      (19) RACDTCRSAAC  TVYCEADSAYLCTTCDA    RVHAANRVASRHERVRV        (12-58)
S12569    (--) QRCDSCRSAPC  AFYCLADSAALCATCDA    DVHSVNPLARRHRRVPM        (13-57)
S3574     (19) ALCDFCREQRS  MVYCRSDAASLCLSCDR    NVHSANALSRRHTRTLL         (3-47)
C60910    (19) AGCDYCGDAAA  VVYCRADAARLCLPCDR    HVHGANGVCSRHARAPL        (36-80)
NTL1      (--) KKCTHCQVTKTPQWREGPLGTKTLCNAC    GVRYRSGRLFPEYRPAASPTFVPT (108-158)
hGATA1    (29) RECVNCGATATPLWRRDRTG HYLCNAC    G LYHKMNGQNRPLIRPKKRLIVS (202-252)
mGATA1    (29) RECVNCGATATPLWRRDRTG HYLCNAC    G LYHKMNGQNRPLIRPKKRMIVS (202-252)
cGATA1    (29) RECVNCGATATPLWRRDGTG HYLCNAC    G LYHRLNGQNRPLIRPKKRMIVS (108-158)

                                                   +        +
STO            PRCDICQEK  AAFIFCVEDRALLCRDCD E    SIHVANSRSANHQRFLATGIKV  (55-104)
R1577          PRCDVCQEK  AAFIFCVEDRALFCRDCD E    PIHVPGTLSGNHQRYLATGIRV  (55-104)
E10707         PRCDVCQEK  AAFIFCVEDRALFCRDCD E    PIHVPGTLSGNHQRYLTTGIRV  (56-105)
R2931          PKCDICQEA  SGYFFCLEDRALLCRDCD V    SIHTVNSFVSVHQRFLLTGVQV  (64-113)
R1479          PLCDVCREK  RGLVFCVEDRAILCADCD E    PIHSANDLTAKHTRFLLVGAKL  (64-113)
CONSTANS         CESCERAP AAFL CEADDASLCTACD    SEVHSANPLARRHQRVPILPISG  (63-100)
COL2             CQSCESAP AAFL CKADAASLCTACDAE    IHSANPLARRHQRVPILPLSA   (59-96)
S3574            CDRCVGQP AAVR CLEENTSLCQNCD    WNGHGAASSAAGHKRQTINCYSG   (48-85)
C60910           CAACAAAG AVFR  RGAGGFLCSNCDFSRHRHGGERDPAAPLHDRSTVHPYTG  (81-121)
hGATA1         TQCTNCQTTTTTLWRRNAS-GDPVCNACGLYY  KLHQVNRPLTMRKDGIQTRNRK (257-309)
mGATA1         TQCTNCQTTTTTLWRRNAS-GDPVCNACGLYF  KLHQVNRPLTMRKDGIQTRNRK (257-309)
cGATA1         TQCSNCQTSTTTLWRRSPM-GDPVCNACGLYF  KLHQVNRPLTMRKDGIQTRNRK (163-215)
```

Figure 3. Alignment of 14 amino acid sequences containing zinc-finger motifs and their flanking regions. The seven deduced amino acid sequences of the zinc-finger region are compared with the corresponding domains of CONSTANS,[5] COL2,[6] and STO[7] of *Arabidopsis*, NTL1 of tobacco,[3] hGATA-1 of human,[19] mGATA-1 of mouse,[20] and cGATA-1 of chicken.[21] Numbers in parentheses on the right are the position of the finger motif counted from the amino terminus of each protein. Amino acids identical to STO are indicated by light shading. The cysteines that form the finger structures are indicated by heavy shading. A "+" means that an amino acid is positively charged. Two well-conserved sequences in each finger of 5 sequences are boxed. Numbers in parentheses on the left indicate the number of amino acids between the 2 fingers in each protein.

acid sequences of R1577, E10707, R1479 and R2931, conservation was most apparent in each of the zinc-fingers, with consensus sequences "CCADEAAL" and "FCV(L)EDRA." These two consensus sequences were also found in zinc-fingers in the STO protein.[7] Therefore, we speculate that these four predicted rice proteins might comprise a subfamily of GATA-1-like proteins.

We performed a linkage analysis based on our high-density rice RFLP linkage map.[13] The results showed that the seven cDNA clones encoding putative zinc-finger proteins were located on five different rice chromosomes (Fig. 4). Clone S3574 was linked to RFLP markers L629 and C92 on chromosome 2; R1577 was linked to C436 and C953A on chromosome 2; C60910 was linked to C198 and C1135 on chromosome 3; E10707 was linked to C335 and R896 on chromosome 4; R2931 was linked to R1952 and R2749 on chromosome 6; R1479 was linked to R1028 and R1888 on chromosome 6; and S12569 was linked to C152 and L776 on chromosome 9.

The position of R2931 on the map is noteworthy. R2931 was found to be closely linked to *Hd-3*, one of the putative QTLs controlling heading date in rice.[9] It could therefore be useful for fine-scale genetic mapping of *Hd-3* as a step toward map-based cloning. Two putative QTLs conferring salt tolerance have been located on chromosomes 3 and 8 in rice,[10] and the putative proteins encoded by R1577 and E10707 showed relative high similarities (49% and 53%, respectively) to the STO protein, which is implicated in salt tolerance in *Arabidopsis*. However, R1577 and E10707 are not potential QTLs, as they were mapped to chromosomes 2 and 4.

Comparative mapping in *Arabidopsis* and *Brassica*, using the cloned *CO* candidate gene, revealed that a homologous region may also be regulating flowering time in *B. nigra*.[17] In this study, we report seven cDNAs as rice homologues of *CO* and *STO* genes. We consider that these clones may be a powerful tool for the identification of candidate genes conferring heading date and salt tolerance in rice.

Acknowledgments: We are grateful to all the members of the Rice Genome Research Program (RGP) for their contributions. RGP is supported by funds from the Japan Racing Association (JRA) and from the Ministry of Agriculture, Forestry and Fisheries of Japan (MAFF).

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022

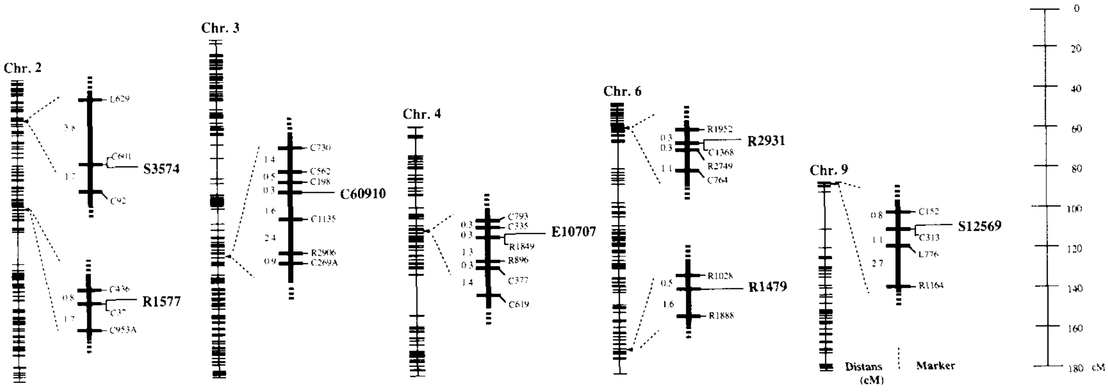


Figure 4. Chromosomal locations of the seven cDNAs encoding putative zinc-finger proteins. Designations of markers are as follows: C, cDNA clones derived from the callus library; R, cDNA clones derived from the young root library; S, cDNA clones derived from the shoot library; E, cDNA clones derived from the panicle library; L, *Not* I linking clone. Further details are given by Kurata et al.[13]

References

1. Pevny, L., Simon, M. C., Robertson, E. et al. 1991, Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1, *Nature*, **349**, 257–260.
2. Omichinski, J. G. 1993, NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1, *Science*, **261**, 438–446.
3. Daniel-Vedele, F. and Caboche, M. 1993, A tobacco cDNA clone encoding a GATA-1 zinc-finger protein homologous to regulators of nitrogen metabolism in fungi, *Mol. Gen. Genet.*, **240**, 365–373.
4. Feng, B., Xiao, X., and Marzluf, G. A. 1993, Recognition of specific nucleotide bases and cooperative DNA binding by the *trans*-acting nitrogen regulatory protein NIT2 of *Neurospora crassa*, *Nucleic Acids Res.*, **21**, 3989–3996.
5. Putterill, J., Robson, F., Lee, K. et al. 1995, The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors, *Cell*, **80**, 847–857.
6. Ledger, S. E., Dare, A. P., and Putterill, J. 1996, *COL2* is a homologue of the *Arabidopsis* flowering-time gene *CONSTANS*, *Plant Physiol.*, **112**, 862.
7. Lippuner, V., Cyert, M. S., and Gasser, C. S. 1996, Two classes of plant cDNA clones differentially complement yeast calcineurin mutants and increase salt tolerance of wild-type yeast, *J. Biol. Chem.*, **271**, 12859–12866.
8. Takatsuji, H. 1998, Zinc-finger transcription factors in plants, *Cell. and Mol. Life Sci.* (in press).
9. Yano, M., Harushima, Y., Nagamura, Y. et al. 1997, Identification of quantitative trait loci controlling heading date of rice using a high-density linkage map, *Theor. Appl. Genet.*, **95**, 1025-1032.
10. Teng, S. 1994, Gene tagging for salt tolerance in rice (*Oryza sativa* L.) via RFLP, Ph.D. Dissertation, Univ. Philippines at Los Baños.
11. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990, Bacic Local Alignment Search Tool, *J. Mol. Biol.*, **215**, 403-410.
12. Murray, M. G. and Thompson, W. F. 1980, Rapid isolation of high molecular weight plant DNA, *Nucleic Acids Res.*, **8**, 4321–4325.
13. Kurata, N., Nagamura, Y., Yamamoto, K. et al. 1994, A 300 kilobase interval genetic map of rice including 883 expressed sequences, *Nature Genetics*, **8**, 365–372.
14. Lander, E. S., Green, P., Abrahamson, J. et al. 1987, MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations, *Genomics*, **1**, 174–181.
15. Gerber, H. P., Speipel, K., Georgive, O. et al. 1994, Transcriptional activation modulated by homopolymeric glutamine and proline stretches, *Science*, **263**, 808–811.
16. Tague, B. W. and Goodman, H. M. 1995, Characterization of a family of *Arabidopsis* zinc-finger protein cDNAs, *Plant Mol. Biol.*, **28**, 267–279.
17. Lagercrantz, U., Putterill, J., Coupland, C. et al. 1996, Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering-time, *Plant J.*, **9**, 13–20.
18. Devereux, J., Haeberli, P., and Smithies, O. 1984, A comprehensive set of sequence analysis programs for the VAX, *Nucleic Acids Res.*, **12**, 387–395.
19. Trainor, C. D., Evans, T., Felsenfeld, G., and Boguski, M. S. 1990, Structure and evolution of a human erythroid transcription factor, *Nature*, **343**, 92–96.
20. Tsai, S. F., Martin, D. I. K., Zorn, L. I. et al. 1989, Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells, *Nature*, **339**, 446-451.
21. Evans, T. and Felsenfeld, G. 1989, The erythroid-specific transcription factor Eryf 1: a new zinc finger protein, *Cell*, **58**, 877-885.

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022

Downloaded from https://academic.oup.com/dnaresearch/article/5/2/95/613806 by guest on 21 August 2022