BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **226,** 32–36 (1996)
ARTICLE NO. 1307

Ribosomal Protein RL44 Is Encoded by Two Subfamilies in Upland Cotton (*Gossypium hirsutum* L.)[1,2]

Gregory Hood,* Rickie B. Turley,† and James S. Steen*

**Department of Biology, Delta State University, Cleveland, Mississippi; and †USDA-ARS, Cotton Physiology and Genetics Research, P.O. Box 345, Stoneville, Mississippi 38776*

Received July 27, 1996

We have isolated 4 cDNA clones encoding the full-length sequence of the eukaryotic ribosomal protein RL44 from upland cotton (*Gossypium hirsutum* L.). Sequencing of these clones resulted in the classification of 2 subfamilies of RL44; these subfamilies had coding regions (315 bp) which were 92% identical. RL44-1 (454 bp) and RL44-2 (485 bp) constitute subfamily 1, whereas RL44-3 (913 bp) and RL44-5 (541 bp) constitute subfamily 2. The differences in nucleotide sequences, however, occurred only at third codon positions and the resulting amino acid sequences of the two RL44 subfamilies were identical. The ORF encoded a protein of 105 residues with a M_r = 12029. A bipartite nuclear targeting sequence was identified from residues 29 to 43.

Eukaryotic ribosomes are complex structures containing 3 to 4 rRNAs and 70 to 80 distinct proteins. The genomic organization and the regulation of genes encoding ribosomal proteins (eukaryotic) are well documented for many animal and yeast systems (1,2). Documentation of genes encoding ribosomal proteins in plants, however, has not been as forthcoming (3,4). This may be partially due to an increased complexity of ribosomal biogenesis. Plants produce one more set of ribosomes than other eukaryotes, i.e., cytoplasmic, mitochondrial, and plastidic (2,3,5). All of the ribosomal proteins for the cytoplasmic type, and approximately two-thirds or greater of the mitochondrial and plastidic types are encoded in the nucleus (6,7,8). Also, many plants, including cotton (*Gossypium hirsutum* L.), have elevated ploidy levels; cotton is an allotetraploid (AADD) derived from the polyploidization of an old world (AA genome) and new world (DD genome) diploids (9,10). Consequently, alloalleles in the cotton genome complicate the study of ribosomal biogenesis during plant development. Increased ploidy levels have been speculated to generate multiple forms of the same ribosomal proteins (5,11). However, polyploidy can not explain all multiple forms of ribosomal proteins in plant systems (4).

This report is part of an ongoing project to characterize ribosomal proteins from cotton. Three ribosomal proteins have already been characterized, i.e., RL41 (12), RS16 (13), and RS4e (14); each was encoded by a multigene family as determined by Southern blot analysis and sequencing. Our interest in ribosomal proteins is two-fold: first, characterization of ribosomal proteins is essential in elucidating ribosomal structure, gene expression, and genomic organization; and second, ribosomal biogenesis precedes or accompanies periods of cell growth and can therefore be used as an indicator of cellular activity (4,11,15,16). Therefore, our goal is to use these clones in multiple studies of cotton cell development, e.g., cotton fiber initiation/elongation and in characterizing differences between diploid and tetraploid cotton species.

[1] The nucleic acid sequences of in this paper of RL44-1 and RL44-5 have been submitted to GenBank under the Accession Numbers U64677 and U64678, respectively.

[2] Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

0006-291X/96 $18.00

MATERIALS AND METHODS

Materials. Deoxyadenosine 5′-(α-thio)triphosphate, [α-^{35}S] (1000-1500 Ci/mmol, 12.5 mCi/mL in 10 mM Tricine, pH 7.6) was purchased from NEN-DuPont (Boston, MA). Resolution was obtained from E. M. Corp. (Chestnut Hill, MA). Sequenase (version 2.0), DNA sequencing kit, random primer kit, urea, Tris, and acrylamide were purchased from United States Biochemical Corp. (Cleveland, OH). The AutoRead Sequencing Kit for nonradioactive sequencing was obtained from Pharmacia (Piscataway, NJ). Long Ranger Gel Solution (50% concentration) was purchased from J.T. Baker (Phillipsburg, NJ). All other chemicals were obtained from Sigma, Inc. (St. Louis, MO).

Screening of cDNA Libraries. A UNI-ZAP library constructed from poly(A)$^+$RNA isolated from cotyledons of 72-h dark grown cotton seedlings (17) was screened with a partial length RL44 probe. The probe was obtained from a cotton leaf library as an expressed sequence tag which demonstrated restriction fragment length polymorphisms. The probe did not encode a full length RL44 protein. Screening was performed on BioTrace NT nitrocellulose membranes as described by the manufacturer (GelmanSciences, Ann Arbor, MI) using a random primed RL44 insert (approximately 2×10^6 cpm). Phagmid was prepared as described by the manufacturer (Stratagene, La Jolla, CA).

DNA Sequence Determination and Amino Acid Analyses. DNA sequencing was performed by the dideoxy method (18) using double stranded DNA templates. Two methods of sequencing were used establishing consensus sequences for these clones. Both the Sequenase, Version 2.0, and the AutoRead kits were used to sequence representative RL44 clones in their entirety on both strands. All comparisons and analyses were performed with programs from PC/GENE (IntelliGenetics, Inc., Campbell, CA).

RESULTS

Four cDNAs encoding the full-length RL44 protein sequences from cotton are aligned and shown with the deduced amino acid sequence in Fig. 1. Clones were grouped into two subfamilies by sequence comparisons. RL44-1 (454 bp) and RL44-2 (485 bp) constitute subfamily 1, whereas, RL44-3 (913 bp) and RL44-5 (541 bp) constitute subfamily 2. The ORFs between the 2 subfamilies are 92% identical with all nucleotide differences occurring at the third codon position, and do not translate into amino acid differences. The 5′ and 3′ noncoding regions are more widely diverse. Differences in the ORF are marked with x's placed above the nucleotide sequence in Fig. 1. The ORF is 315 bp and encodes a protein with the molecular weight of 12,029. RL44 is basic in composition, containing 30 arginines and lysines with only 5 aspartates and glutamates (estimated pI of 11.03). A bipartite nuclear targeting sequence, i.e., residues 29-43, was determined and underlined in Fig. 1. A nuclear targeting signal would direct cytosolic ribosomal proteins to the nucleus; the site of biogenesis of ribosomal precursors (1). Also, the two subfamilies coexisted in a single library indicating they were concurrently expressed during seedling growth.

A comparison was performed between the deduced amino acid sequence of cotton and seven yeast RL44 proteins of *Candida maltosa* (19), *Candida tropicalis* (19), *Kluyveromyces lactis* (20), *Kluyveromyces marxianus* (19), *Pichia gilliermondii* (21), *Saccharomyces cerevisiae* (19), and *Schwanniomyces occidentalis* (22), and shown in Fig. 2. The identities between yeast and cotton clones range from 74-78% with an additional 8-10 % similarity. This high identity indicates that cotton RL44 is a homolog of the RL44 proteins from yeast.

DISCUSSION

Generally, genes encoding ribosomal proteins are present in only one or two copies per haploid genome in eukaryotic cells (23). In mammalian cells it has been reported that only one copy is actively expressed; one exception is RS4e which has two different forms located on the X and Y chromosomes of humans (24). Multigene families encoding ribosomal proteins are common in plant systems, and include cytoplasmic ribosomal proteins L2 (5), L3 (25), L25 (11), L34 (11), L41 (12), S4e (14), S11 (4), S14 (3), S16 (13), and chloroplast protein L12 (26). Multiple forms of L2 and L34 are believed to be derived from different subgenomes in allotetraploid tobacco (5,11). Similarly, different forms of RL44 and other cotton ribosomal proteins (12-14) may result from the expression of genes from two subgenomes in allotetraploid cotton. Multiple forms of ribosomal proteins, however, are not always the consequence of

```
                                                                       x
RL44-1  TTTT------------------------------CAGCCGCAGCCATGGTGAACGTACCTAAGACCAAG    39
RL44-2  ----------------------------------CAGCCGCAGCCATGGTGAACGTACCTAAGACCAAG    35
RL44-3  TTTACACCGAAACAGAGCTCGGGTCATCTCTCACCAGCCGCAACCATGGTGAACGTACCTAAGACGAAG    69
RL44-5  TTTACACCGAAACAGAGCTCGGGTCATCTCTCACCAGCCGCAACCATGGTGAACGTACCTAAGACGAAG    69
                                                     M  V  N  V  P  K  T  K     8

             x  x                       x                 x                 x
RL44-1  AAGACCTACTGCAAGAGCAAGGAGTGCAGGAAACACACTTTGCACAAGGTCACACAGTATAAGAAGGGC   108
RL44-2  AAGACCTACTGCAAGAGCAAGGAGTGCAGGAAACACACTTTGCACAAGGTCACACAGTATAAGAAGGGC   104
RL44-3  AAGACTTATTGCAAGAGCAAGGAGTGCAGGAAGCACACTTTGCACAAGGTTACACAGTATAAGAAGGGT   138
RL44-5  AAGACTTATTGCAAGAGCAAGGAGTGCAGGAAGCACACTTTGCACAAGGTTACACAGTATAAGAAGGGT   138
          K  T  Y  C  K  S  K  E  C  R  K  H  T  L  H  K  V  T  Q  Y  K  K  G    31

                         x  x        x        x  x        x     x
RL44-1  AAGGATAGTTTGGCTGCACAGGGGAAGCGTCGTTACGACCGGAAACAATCCGGTTACGGTGGTCAGACC   177
RL44-2  AAGGATAGTTTGGCTGCACAGGGGAAGCGTCGTTACGACCGGAAACAATCCGGTTACGGTGGTCAGACC   173
RL44-3  AAGGATAGTTTGGCTGCTCAAGGGAAGCGACGTTACGATCGCAAACAATCAGGTTATGGTGGTCAGACC   207
RL44-5  AAGGATAGTTTGGCTGCTCAAGGGAAGCGACGTTACGATCGCAAACAATCAGGTTATGGTGGTCAGACC   207
          K  D  S  L  A  A  Q  G  K  R  R  Y  D  R  K  Q  S  G  Y  G  G  Q  T    54

                x                                                     x     x
RL44-1  AAACCAGTGTTCCACAAGAAGGCAAAGACCACCAAGAAGATTGTGCTAAGGCTGCAATGCCAAGGTTGC   246
RL44-2  AAACCAGTGTTCCACAAGAAGGCAAAGACCACCAAGAAGATTGTGCTAAGGCTGCAATGCCAAGGTTGC   242
RL44-3  AAACCAGTCTTCCACAAGAAGGCAAAGACCACCAAGAAGATTGTGCTAAGGCTGCAATGCCAGGGTTGT   276
RL44-5  AAACCAGTCTTCCACAAGAAGGCAAAGACCACCAAGAAGATTGTGCTAAGGCTGCAATGCCAGGGTTGT   276
          K  P  V  F  H  K  K  A  K  T  T  K  K  I  V  L  R  L  Q  C  Q  G  C    77

             x  x     x                             x                       x
RL44-1  AAGCACGTTTCACAACATCCGATCAAGAGGTGCAAGCACTTTGAGATTGGTGGAGACAAGAAGGGGAAA   315
RL44-2  AAGCACGTTTCACAACATCCGATCAAGAGGTGCAAGCACTTTGAGATTGGTGGAGACAAGAAGGGGAAA   311
RL44-3  AAGCATGTCTCACAGCATCCGATCAAGAGGTGCAAGCACTTTGAAATTGGTGGAGACAAGAAGGGGAAG   345
RL44-5  AAGCATGTCTCACAGCATCCGATCAAGAGGTGCAAGCACTTTGAAATTGGTGGAGACAAGAAGGGGAAG   345
          K  H  V  S  Q  H  P  I  K  R  C  K  H  F  E  I  G  G  D  K  K  G  K   100

                x  x
RL44-1  GGAACATCTCTGTTTTAAACGTGTTATTCATGGTATCTTGTTATGTTT------------------GCT   366
RL44-2  GGAACATCTCTGTTTTAAACGTGTTATTCATGGTATCTTGTTATGTTT------------------GCT   362
RL44-3  GGAACATCACTTTTCTAGATGTGATGATATTCATGTCAAATTTTATGTTATGGGTTTTTACTTGGAACT   414
RL44-5  GGAACATCACTTTTCTAGATGTGATGATATTCATGTCAAATTTTATGTTATGGGTTTTTACTTGGAACT   414
          G  T  S  L  F                                                         105

RL44-1  TTACTTTATATGGAGT-------AGTTTG----------------------------------------   388
RL44-2  TTACTTTATATGGAGT-------AGTTTG----------------------------------------   384
RL44-3  TAACGCAGAATGGAGTTTTGTTAAGTTTGGTATTAGTATTCTAATGTTGTTTTTAAGAGGATATTGTGT   483
RL44-5  TAACGCAGAATGGAGTTTTGTTAAGTTTGGTATTAGTATTCTAATGTTGTTTTTAAGAGGATATTGTGT   483

RL44-1  ---------------------------------------------------------------------   388
RL44-2  ---------------------------------------------------------------------   384
RL44-3  TTTTTAAGGTTAATGGCATTGGTCACACTCATCGATAGCAATTGTCATTGGTCCTGTATATATCAATAT   552
RL44-5  TTTTTAAGGTTAATGGCATTGGTCACACTCATCGATAGCAA----------------------------   524

RL44-1  ------AACTTACAAGACTGTAATGGAATTTTGTTA---------------------------------   418
RL44-2  ------AACTTACAAGACTGTAATGGAATTTTGTTA---------------------------------   414
RL44-3  CTTGTTAGCTCGCATCACTTCCTTGAGATCTTGTAGGTCAACGCCCTAACGAAGAAACTTCATCGAATC   621
RL44-5  ---------------------------------------------------------------------   524

RL44-1  ---------------------------------------------------------------------   418
RL44-2  ---------------------------------------------------------------------   414
RL44-3  AATAGCAATAAGGCCTGTTTGAAGAGGCTCATAAACGGAACGTCTCGAAATAATACCTGGGGCAGAAGA   690
RL44-5  ---------------------------------------------------------------------   524

RL44-1  ---AATTTGCTTAGTTTCT--------------------------------------------------   434
RL44-2  ---AATTTGCTTAGTTTCT--------------------------------------------------   430
RL44-3  TTCAATTAACCGAGATTCAGAAGCTGAAATTTCACCCTGACCGTCAATAGGCTTAGCCAGGGCATTTAT   759
RL44-5  ---------------------------------------------------------------------   524

RL44-1  ---------------------------------TCT---------------------------------   437
RL44-2  ---------------------------------TCTGATGAAGTTTTGCAGTTTTCT------------   454
RL44-3  AACACGACCCGAAATAAGCCTCCCTCATTGGTATCTGAGTAAATTTTTCTTGTTGCTTTTACAGAAACT   828
RL44-5  ---------------------------------------------------------------------   524

RL44-1  ------------------------------------------------------------------AAA_17   454
RL44-2  -------------AACAAA-------------------------------------------AAAAAAAAA_17   485
RL44-3  TCCCTCTTGTATCATCAAACCGTCACCCATTAATATAACACCAACATTATTTGATTCCAAATTTAAAA_17   913
RL44-5  --------------------------------------------------------------------A_17   541
```

FIG. 1. Nucleotide and the deduced amino acid sequences of ribosomal protein RL44 from cotton. In the nucleotide sequence, the ATG start sites and the termination sites are underlined. In the amino acid sequence the bipartite nuclear targeting sequence is underlined.

```
Gh   MVNVPKTKKTYCKSKECRKHTLHKVTQYKKGKDSLAAQGKRRYDRKQSGYGGQTKPVFHKKAKTT   65
Cm   -..I...RN....G.G.....I.......S.RA..F...................Q.........   64
Ct   -..I...RN....G.G.....I.......A.RA..F...................Q.........   64
Kl   -......R.....G......AQ.......A..A..Y.............F.....QI........   64
Km   -......R.....G.A....SQ.......A..A..Y.............F.....QI........   64
Pg   -......RR....G.D.....Q.......A..A..F.............F...............   64
Sc   -......R.....G.T.....Q.......A..A..F.............F...............   64
So   -......R.....G.......Q.......A..A..F...........R.......QI........   64

                                                % Ident.   % Simil.   Tot. AA
Gh   KKIVLRLQCQGCKHVSQHPIKRCKHFEIGGDKKGKGTSL-F     100          0        105
Cm   .....K.E.TV..TKK.L.L.....I.L..E..Q..QA.Q.      74          9        105
Ct   .....K.E.TV..TKK.L.L.....I.L..E..Q..QA.Q.      74          9        105
Kl   ..V....E.MS..TKT.LAL.......L..E..Q..QA.Q.      75         10        105
Km   ..V....E.MS..TKT.LAL.......L..E..Q..QA.Q.      74         10        105
Pg   ..V....E.VV..TKA.LSL.......L.....Q..QA---      77          9        102
Sc   ..V....E.VK..TRA.LTL.......L..E..Q..QA.Q.      77          8        105
So   ..V....E.VV..TKA.LSL.......L.....Q..QA.Q.      76          8        105
```

FIG. 2. Comparison of ribosomal protein RL44 from cotton and yeast. Residues which are identical to the cotton sequence are denoted with dots ''.'' and gaps are denoted with ''–'' and inserted in the sequence to achieve maximum identity. Organisms from which the sequences were derived are listed: *Gh, Gossypium hirsutum; Cm, Candida maltosa; Ct, Candida tropicalis; Kl, Kluyveromyces lactis; Km, Kluyveromyces marxianus; Pg, Pichia gilliermondii; Sc, Saccharomyces cerevisiae;* and *So, Schwanniomyces occidentalis.* The percent identity, similarity and total proteins are also listed.

increased ploidy; ribosomal proteins L3 (25) and S11 (4) are two examples of multiple forms from a diploid plant. It is not understood how differences in primary structure of ribosomal proteins effect the biogenesis of ribosomes, or how structurally distinct ribosome populations differ in their ability to translate protein. Plants may regulate multiple forms of a ribosomal protein differently. Differential expression patterns were reported for ribosomal protein S14 in maize tissues (3), however, usually all three forms were expressed at a given time. The multiple forms of RL44 genes from cotton would not add to the formation of heterologous populations of ribosomes since the two subfamilies encode identical proteins. Further work is needed to characterize RL44 gene origins, along with their developmental regulation. Differences in ribosomal protein structure, e.g., RL44, between plants, yeast, and animals could give us new insights on the importance of specific regions of ribosomal proteins.

ACKNOWLEDGMENTS

We thank Shelia Meeks for her excellent technical assistance, and Bill Meredith and Alex Kahler for the production of the RL44 probe (RFLP).

REFERENCES

1. Wittmann-Liebold, B. (1985) Structure, Function, and Genetics of Ribosomes (Hardesty, B., and Kramer, G., Eds.), pp. 326–361. Springer-Verlag, NY.
2. Boynton, J. E., Gillham, N. W., and Lambowitz, A. M. (1979) Ribosomes: Structure, Function, and Genetics (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahal, and Nomura, M., Eds.), pp. 903–950. University Park Press, Baltimore, MD.
3. Larkin, J. C., Hunsperger, J. P., Culley, D., Rubenstein, I., and Silflow, C. D. (1989) *Genes Dev.* **3,** 500–509.
4. Lu, G., Wu, K., and Ferl, R. J. (1993) *Plant Physiol.* **102,** 695–696.
5. Marty, I., and Meyer, Y. (1992) *Nucleic Acids Res.* **20,** 1517–1522.
6. Aubert, D., Bisanz-Seyer, C., and Herzog, M. (1992) *Plant Mol. Biol.* **20,** 1169–1174.
7. Mache, R. (1990) *Plant Sci.* **72,** 1–12.
8. Gantt, J. S., and Thompson, M. D. (1990) *J. Biol. Chem.* **265,** 2763–2767.
9. Wendel, J. F. (1989) *Proc. Natl. Acad. Sci.* **86,** 4132–4136.
10. Percival, A. E., and Kohel, R. J. (1990) *Adv. Agron.* **44,** 225–256.
11. Gao, J., Kim, S. R., Chung, Y. Y., Lee, J. M., and An, G. (1994) *Plant Mol. Biol.* **25,** 761–770.
12. Turley, R. B., Ferguson, D. L., and Meredith, W. R., Jr. (1994) *Plant Physiol.* **105,** 1449–1450.

13. Turley, R. B., Ferguson, D. L., and Meredith, W. R., Jr. (1994) *Plant Physiol.* **106,** 1219–1220.
14. Turley, R. B., Ferguson, D. L., and Meredith, W. R., Jr. (1995) *Plant Physiol.* **108,** 431–432.
15. Mager, W. H. (1988) *Biochim. Biophys. Acta* **949,** 1–15.
16. Joanin, P., Gigot, C., and Philipps, G. (1993) *Plant Mol. Biol.* **21,** 701–704.
17. Ni, W., and Trelease, R. N. (1991) *Arch. Biochem. Biophys.* **289,** 237–243.
18. Sanger, F., Nicklin, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
19. Kawai, S., Murao, S., Mochizuki, M., Shibuya, I., Yano, K., and Takagi, M. (1992) *J. Bacteriol.* **174,** 254–262.
20. Dehoux, P., Davies, J., and Cannon, M. (1993) *Eur. J. Biochem.* **213,** 841–848.
21. Kawai, S. (1992) *EMBL/GENEBANK/DDBJ* data bank submission.
22. Del Pozo, L., Abarca, D., Hoenicka, J., and Jimenez, A. (1993) *Eur. J. Biochem.* **213,** 849–857.
23. Amaldi, F., Bozzoni, I., Beccari, E., and Pierandrei-Amadi, P. (1989) *TIBS* **14,** 175–178.
24. Fisher, E. M. C., Beer-Romero, P., Brown, L. G., Ridley, A., McNeil, J. A., Lawrence, J. B., Williard, H. F., Bieber, F. R. and Page, D. C. (1990) *Cell* **63,** 1205–1218.
25. Kim, Y., Zhang, H. and Scholl, R. L. (1990) *Gene* **93,** 177–182.
26. Schmidt, M., Pichl, L., Lepper, M., and Feierabend, J. (1993) *Biochim. Biophys. Acta* **1172,** 349–352.