The EMBO Journal Vol.2 No.7 pp.1165–1168, 1983

The chromosomal arrangement of six soybean leghemoglobin genes

Kirsten Bojsen, Dorte Abildsten, Erik Ø. Jensen, Kirsten Paludan and Kjeld A. Marcker*

Department of Molecular Biology and Plant Physiology, University of Aarhus, C.F. Møllers Allé 130, 8000 Århus C, Denmark

Communicated by K.A. Marcker
Received on 28 March 1983

Clones containing six leghemoglobin (Lb) genes have been isolated from two genomic libraries of soybean. They encompass two independent DNA regions: a 40-kb region containing four genes in the order 5′ Lba-Lbc$_1$-ψLb-Lbc$_3$ 3′ with the same transcriptional polarity, and another 40-kb region containing two genes in the order 5′ Lbc$_4$-Lbc$_2$ 3′ with the same polarity. The order in which the Lb genes are arranged in the soybean genome imply that they are activated in the opposite order to which they are arranged on the chromosome. There is a close similarity between corresponding DNA regions outside the Lb genes in the two clusters. Thus, a moderately repetitive DNA element is present in corresponding positions in each cluster. In addition, at least two different non-Lb genes are linked to each Lb gene cluster in corresponding positions. These genes are apparently regulated in a way which differs from that of the Lb genes. The existence of two very similar Lb gene clusters in soybean suggest that soybean may have evolved from an ancestral form by genome duplication.

Key words: chromosomal arrangement/genome duplication/leghemoglobin genes

Introduction

Leghemoglobins (Lbs) are monomeric hemoproteins synthesized exclusively in the root nodules which develop through the symbiotic association of Rhizobia with leguminous plants. Soybean nodules contain four major species of Lbs called Lba, Lbc$_1$, Lbc$_2$ and Lbc$_3$ (Fuchsman and Appleby, 1979). In addition, several minor Lb components have been detected in the nodules, but it seems most likely that some of these components are derived by post-translational modifications of some of the major components. The differences in the amino acid sequences among the various Lb components are small and correspond to ~10 amino acids (Sievers *et al.*, 1978). Molecular hybridization experiments have shown that the Lbs are encoded in the soybean genome by a small family of genes (Sullivan *et al.*, 1981; Marcker *et al.*, 1981). Cloned Lb cDNA hybridizes to at least seven *Eco*RI genomic restriction fragments. Six hybridizing DNA fragments have been isolated from various soybean DNA libraries and DNA sequence analysis has been completed for the Lb genes located on these fragments. Four isolated genes code for Lba, Lbc$_1$, Lbc$_2$ and Lbc$_3$, respectively, while a fifth gene is probably non-functional and consequently is an Lb pseudo gene (ψLb) (Hyldig-Nielsen *et al.*, 1982; Wiborg *et al.*, 1982, 1983). Preliminary sequence analysis of a complete Lb gene (Lbc$_4$) encoded on a 10-kb *Eco*RI restriction fragment suggests that this particular gene is functional. In addition to the complete genes, two truncated genes have been identified. Thus, the Lb gene family in soybean consists of five functional genes, one pseudo gene and at least two truncated genes. We report here that the six complete Lb genes are arranged in two independent clusters in the soybean genome. Four genes are very closely linked in the order 5′ Lba-Lbc$_1$-ψLb-Lbc$_3$ 3′. The distance between the various genes in this cluster is ~2.5–3 kb. The two remaining genes are tightly linked in the order 5′ Lbc$_4$-Lbc$_2$ 3′ and are ~2 kb apart. No direct link between the two Lb gene clusters has been obtained so far. There are at least two different non-Lb genes closely linked to the four Lb gene cluster. These genes are also linked to the two Lb gene cluster in the same relative positions as for the four Lb gene cluster.

*To whom reprint requests should be sent.

Results and Discussion

Isolation of Lb genomic recombinants

An Lb cDNA clone was used to screen two independent soybean chromosomal libraries. Both libraries were constructed using bacteriophage λ as a vector. One library was constructed from a limited *Eco*RI digest of soybean DNA, while the other was constructed from a limited *Alu*I/ *Hae*III digest of soybean DNA. For chromosome walking, appropriate non-Lb gene fragments were subcloned into pBR322 and subsequently used for screening. In this way >20 different λ recombinants were obtained. Restriction enzyme maps for these phage DNAs were determined by analysing a combination of single and double restriction enzyme digests using the filter hybridization method of Southern (1975). By using these methods, complete sets of overlapping clones were obtained covering two regions of soybean DNA of ~40 kb each. Six Lb genes are encompassed within these two DNA regions.

The chromosomal arrangement of six soybean Lb genes

The six Lb genes are arranged in two independent clusters in the soybean genome as shown in Figure 1. Four genes constitute one cluster and are tightly linked in the order 5′ Lba-Lbc$_1$-ψLb-Lbc$_3$ 3′. The distances separating these genes are ~2.5–3 kb. The two remaining Lb genes constitute another cluster and are linked in the order 5′ Lbc$_4$-Lbc$_2$ 3′. The distance separating these genes is ~2 kb. The transcriptional orientations of the six Lb genes were determined by a combination of DNA sequence analysis and by analysis of single and double restriction enzyme digests of the appropriate clones using various subcloned exons as hybridization probes. So far no clone has been obtained which links the two Lb gene clusters and it is therefore not clear whether both clusters are on the same chromosome, or whether they are on separate chromosomes. In the former case the two clusters must be at least 25–30 kb apart.

The Lbc$_2$ and Lbc$_4$ genes are contained within a 10-kb *Eco*RI restriction fragment. During phage propagation, the Lbc$_4$ gene is very often deleted leaving a phage with a 6-kb fragment which contains the Lbc$_2$ gene. Similar deletions of globin sequences were also observed during cloning of human α-globin squences (Lauer *et al.*, 1980). In a previous study the Lbc$_2$ gene was isolated on a 6-kb *Eco*RI fragment (Wiborg *et*

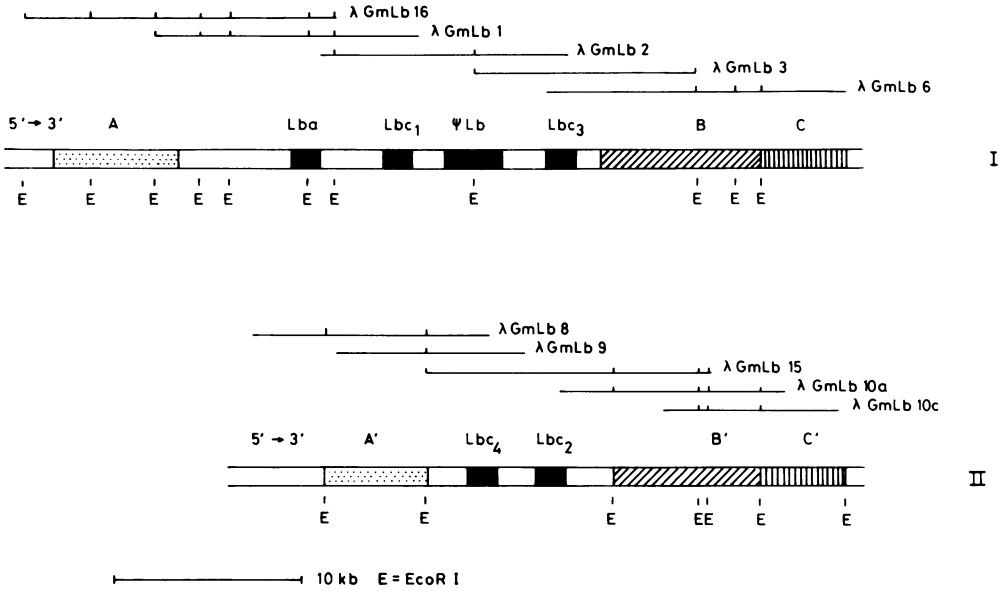

Fig. 1. Chromosomal arrangement of six soybean Lb genes. Solid boxes indicate the positions of the Lb genes (the three introns present in all Lb genes are not shown). Identical shadings (A and A', B and B', C and C') represent cross-hybridizing non-Lb regions. The transcriptional polarity of the Lb genes is indicated by the two arrows. The sizes of the soybean DNA inserts in both sets of overlapping λ clones are shown as horizontal lines below the two restriction maps. Genomic *Eco*RI(E) sites are indicated by vertical bars. In clones derived from the *AluI/HaeIII* library, the *Eco*RI sites generated by addition of *Eco*RI linkers are not indicated.

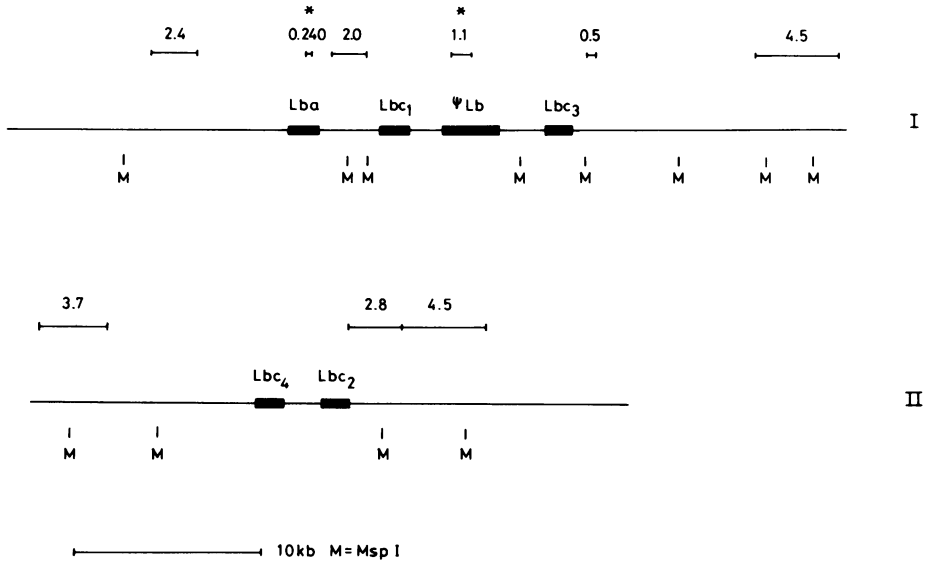

Fig. 2. *Msp*I restriction maps of the two genomic Lb gene clusters. Lb genes are indicated by solid boxes. The indicated *Msp*I(M) sites have been determined both in digests of total soybean DNA and in cloned fragments. The subcloned fragments (sizes given in kb) used as hybridization probes are illustrated by bars above the corresponding genomic regions. Asterisks indicate probes derived from intron sequences.

al., 1982). In genomic blots of soybean DNA, an Lb cDNA does hybridize to a 6-kb *Eco*RI fragment, indicating that a structural Lb sequence is present on such a fragment in soybean DNA. Despite the screening of several soybean DNA libraries we have never succeeded in cloning this particular Lb sequence. It is therefore not known whether the Lb sequence present on a 6-kb *Eco*RI fragment in soybean DNA represents a complete or a truncated gene. Brisson and Verma (1982) have isolated a truncated Lb gene consisting of exon 4 and the 3' non-coding end only. In addition, we have isolated a truncated Lb gene present within a 5.5-kb *Eco*RI fragment, which consists of exon 3, intron 3, exon 4 and the 3' non-coding end. Neither of these truncated genes, nor the structural sequence present within the 6-kb *Eco*RI fragment are contained within the regions of the soybean genome shown in Figure 1.

Correlation of cloned fragments with those in total genomic DNA

To determine whether the cloned Lb gene regions correlate with the corresponding regions in total genomic DNA, soybean DNA was digested with *Msp*I followed by filter hybridization analysis. Fragments from several λ clones were subcloned into pBR322 and used as hybridization probes. In addition, an Lb cDNA clone was used as probe. Figure 2 indicates the positions of the *Msp*I sites in the Lb gene regions to-

gether with the positions of the various DNA fragments used as hybridization probes. In this way, both Lb clusters were analysed. In all cases the sizes of the hybridizing restriction fragments in total soybean DNA corresponded to those predicted by the map presented in Figure 2. These results therefore clearly indicate that no detectable rearrangement occurred within the Lb gene regions during preparation, isolation or propagation of the clones examined except occasionally for the 10-kb *Eco*RI fragment previously referred to. Because of the limited sensitivity of this assay, however, we cannot rule out the possibility that small rearrangements may have occurred during the cloning procedure.

At least two different non-Lb genes are linked to each Lb gene cluster in corresponding positions

Molecular hybridization experiments revealed extensive cross-hybridization between corresponding DNA regions outside the Lb genes in the two clusters. Thus, a 5-kb DNA region denoted by A in cluster I cross-hybridizes strongly to a 5-kb region denoted by A′ in cluster II. Region A is ~5 kb upstream from the Lba gene while the corresponding region A′ is ~3 kb upstream from the Lbc_4 gene. Similarly, DNA regions B and C in cluster I cross-hybridize strongly to regions B′ and C′ in cluster II. These cross-hybridizing regions are located 3′ to the Lbc_2 and Lbc_3 genes, respectively. In eukaryotic genomes repetitive DNA elements are often closely linked to structural DNA sequences. Southern blotting experiments revealed that DNA sequences from regions B and B′ are repeated 10–20 times in the soybean genome, while the DNA sequences in regions A and C are not repeated elsewhere in the soybean genome.

Northern blotting analysis using subcloned DNA fragments from regions A and C as probes revealed the presence of hybridizing mRNAs in poly(A)$^+$ RNA extracted from soybean roots. These mRNAs are not amplified to any significant extent in poly(A)$^+$ RNA extracted from nodules. Consequently, the DNA regions in A, A′ and C, C′ most likely encode structural soybean genes. The function of these genes is unknown and it is also not clear whether the genes present in all four regions are expressed. However, at least two different genes present in these regions are expressed in uninfected soybean roots, and no amplification of expression seems to occur after the roots are infected with Rhizobia. Thus, the soybean Lb genes are closely linked to other soybean genes which are regulated in a manner different from that of the Lb genes.

Evolution of the soybean Lb genes

We have used molecular cloning and genomic blotting procedures to establish the chromosomal arrangement of six soybean Lb genes. The six Lb genes are arranged in two independent clusters. One consists of four genes and the other of two genes. At least two different non-Lb genes, which are closely linked to the Lb genes in both clusters, are present in corresponding positions in each cluster. In addition, a moderately repetitive DNA element is present in both clusters in corresponding positions. The close similarity between the two Lb gene clusters extending over a region of ~40 kb therefore suggests that the two clusters arose either by duplication of a large segment of a soybean chromosome or by genome duplication. Soybean is a tetraploid (Datta and Saha, 1973) and it is therefore plausible that the existence of two almost identical Lb clusters in soybean is due to genome duplication in an ancestral soybean species. The presence of two independent globin gene clusters in *Xenopus laevis* (Jeffreys *et al.*, 1980) is apparently also due to tetraploidization. Prior to the putative genome duplication in the soybean ancestor, it is reasonable to assume that an Lb gene duplication had already occurred generating a tandemly linked pair of Lb genes. After the assumed genome duplication, a tandem Lb gene duplication occurred in one of the clusters leading to a four Lb gene arrangement. The differences between the two clusters may therefore have arisen after genome duplication and may reflect recent divergence of the two clusters.

Vertebrate α and β globins are differentially expressed during development. Normally the globin genes are activated in the order in which they are arranged on the chromosome. An exception to this rule is the β-globin gene cluster in chicken which contains an embryonic gene on the 3′ side of an adult gene (Villeponteau and Martinson, 1981). The soybean Lb genes also appear to be induced at different times during nodule development (Fuchsman and Appleby, 1979; Verma *et al.*, 1979). Thus, it seems certain that the Lbc_3 gene is activated before the Lba gene. This, therefore, implies that the Lb genes are activated in the opposite order to which they are arranged on the chromosome. However, the detailed changes in the transcription pattern of the Lb genes during nodule development is unknown. Thus, it is still unclear at what stage the Lbc_1, Lbc_2 and Lbc_4 genes are activated.

Materials and methods

Screening of phage libraries

Two λ Charon 4A recombinant phage libraries containing either *Eco*RI or *Alu*I/*Hae*III partial digestion fragments of soybean DNA were used. About 6 x 10^5 recombinant phages from each library were screened with a ^{32}P-labelled Lb cDNA clone as described by Maniatis *et al.* (1978). For isolation of clones from regions outside the Lb genes, appropriate fragments were subcloned into pBR322 and used as hybridization probes. Phage corresponding to positive signals were purified and the DNA prepared according to Maniatis *et al.* (1982).

Preparation of soybean DNA

Chromosomal soybean DNA was prepared from 4 day old seedlings as previously described (Marcker *et al.*, 1981).

Restriction endonuclease analysis

Enzyme digestions were carried out according to the manufacturer's instructions. Samples were fractionated on agarose gels and transferred to nitrocellulose filters as described by Southern (1975).

Subcloning into pBR322

Recovery of DNA fragments from agarose gels for subcloning were performed according to Dretzen *et al.* (1981). The procedures used for subcloning and plasmid preparation were according to Lacy *et al.* (1979).

Hybridization

Hybridization of filters was conducted at 67°C for 16–20 h in 2 x SSC, 20 x Denhardt's solution (Denhardt, 1966), 0.1% SDS, 50 μg/ml denatured salmon sperm DNA, 100 μg/ml poly(A) and an appropriate amount of denatured ^{32}P-labelled probe. The filters were washed at 67°C in the following succession of solutions: 2 x SSC, 10 x Denhardt's, 0.1% SDS; 2 x SSC, 0.1% SDS; 1 x SSC, 0.1% SDS; 0.3 x SSC, 0.1% SDS. For analysis of soybean genomic DNA, dextran sulphate was added to the hybridization solution to a final concentration of 10%.

DNA sequencing

DNA sequencing was performed by the dideoxy chain termination method described by Sanger *et al.* (1980) using a synthetic dodecadenoxynucleotide as primer. Sequencing reaction products were electrophoresed on 6 or 8% 0.3 mm polyacrylamide-urea gels.

Northern blotting analysis

Total RNA was extracted from 3 week old nodules, roots or leaves by the guanidium rhodanide procedure (Chirgwin *et al.*, 1979) and subsequently fractionated on an oligo(dT) cellulose column. Poly(A)$^+$ RNA was separated by electrophoresis in 1% agarose gels, transferred to DBM paper and hybridized with appropriate ^{32}P-labelled clones (Alwine *et al.*, 1979).

Acknowledgements

We thank Drs. R. Goldberg and R. Fischer, UCLA, for providing the limited

*Eco*RI and AluI/*Hae*III soybean libraries. This research was supported by grants from De Danske Sukkerfabrikker and the Danish Natural Science Research Council.

References

Alwine,J.C., Kemp,D.J., Parker,B.A., Reiser,J., Renart,J., Stark,G.R. and Wahl,G.M. (1979) *Methods Enzymol.*, **68**, 220-242.

Brisson,N. and Verma,D.P.S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4055-4059.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.

Datta,P.C. and Saha,N. (1973) *Genét. Ibér.*, **25**, 37-62.

Denhardt,D.T. (1966) *Biochem. Biophys. Res. Commun.*, **23**, 641-646.

Dretzen,G., Bellard,M., Sassone-Corsi,P. and Chambon,P. (1981) *Anal. Biochem.*, **112**, 295-298.

Fuchsman,W.H. and Appleby,C.A. (1979) *Biochim. Biophys. Acta*, **579**, 314-324.

Hyldig-Nielsen,J.J., Jensen,E.Ø., Paludan,K., Wiborg,O., Garrett,R., Jørgensen,P. and Marcker,K.A. (1982) *Nucleic Acids Res.*, **10**, 689-701.

Jeffreys,A.J., Wilson,V., Wood,D., Simons,J.P., Kay,R.M. and Williams, J.G. (1980) *Cell*, **21**, 555-565.

Lacy,E., Hardison,R.C., Quon,D. and Maniatis,T. (1979) *Cell*, **18**, 1273-1283.

Lauer,J., Shen,C.-K.J. and Maniatis,T. (1980) *Cell*, **20**, 119-130.

Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687-701.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) in *Molecular Cloning,a Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY, pp.371-373.

Marcker,K.A., Gausing,K., Jochimsen,G., Jørgensen,P., Paludan,K. and Truelsen,E. (1981) in Panopoulos,N.J. (ed.), *Genetic Engineering in the Plant Sciences*, Praeger Publishers, pp. 63-71.

Sanger,R., Coulson,A.R., Barrell,B.Q., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.

Sievers,S.G., Huhtala,M.-L. and Ellfolk,N. (1978) *Acta Chem. Scand.*, **B32**, 380-386.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.

Sullivan,D., Brisson,N., Goodchild,B., Verma,D.P.S. and Thomas,D.Y. (1981) *Nature*, **289**, 516-518.

Verma,D.P.S., Ball,S., Guérin,C. and Wanamaker,L. (1979) *Biochemistry (Wash.)*, **18**, 476-483.

Villeponteau,B. and Martinson,H. (1981) *Nucleic Acids Res.*, **9**, 3731-3746.

Wiborg,O., Hyldig-Nielsen,J.J., Jensen,E.Ø., Paludan,K. and Marcker, K.A. (1982) *Nucleic Acids Res.*, **10**, 3487-3494.

Wiborg,O., Hyldig-Nielsen,J.J., Jensen,E.Ø., Paludan,K. and Marcker, K.A. (1983) *EMBO J.*, **2**, 449-452.

Note added in proof

Further DNA sequence analysis has shown that the Lbc_4 gene has a stop codon in exon 2 and therefore is a pseudo gene.