

The nuclear ribosomal DNA intergenic spacers of wild and cultivated soybean have low variation and cryptic subrepeats

Daniel L. Nickrent and Jennifer A. Patrick

Abstract: The intergenic ribosomal DNA spacers (IGSs) from cultivated soybean (*Glycine max*) and wild soybean (*Glycine soja*) were sequenced and compared with six other legumes. These IGS sequences were 1821 bp in length in *G. soja* and *G. max* cultivars Arksoy, Ransom, and Tokyo, and 1823 bp long in the *G. max* cultivar Columbus. These represent the smallest published plant IGS sequences to date. Two clones from each of the above five cultivars were sequenced and only 22 sites (1.2%) were polymorphic, thereby supporting previous work that showed low genetic variation in cultivated soybean. The amount of variation observed between different clones derived from the same individual was equal to the amount seen between different cultivars. The soybean IGS sequence was aligned with six other published legume sequences and two homologous regions were identified. The first spans positions 706–1017 in the soybean IGS sequence and ends at a putative promoter site that appears conserved among all legumes. The second is located within the 5' external transcribed spacer, spans positions 1251–1823 in soybean, and includes sequences first identified as subrepeats IV-1 and IV-2 in *Vicia angustifolia*. Sequences homologous to these two subrepeats were identified among all legume species examined and are here designated "cryptic subrepeats" (CS-1, CS-2) given the range in similarity value (79–96% for CS-1 and 60–95% for CS-2). Comparisons of CS-1 and CS-2 sequences within individual species show that divergence (substitutional mutations, insertions, and deletions) is sufficiently high to obscure recognition of the repeat nature of these sequences by routine dot plot analytical methods. The lack of subrepeats in the 5' half of the soybean IGSs raises questions regarding the role they play in transcription termination or enhancement.

Key words: *Glycine*, Fabaceae, ribosomal DNA, promoter.

Résumé : Les espaceurs intergénomiques de l'ADN ribosomique (IGS) ont été séquencés chez le soya cultivé (*Glycine max*) et sauvage (*Glycine soja*) et ces séquences ont été comparées avec celles de six autres légumineuses. Ces séquences IGS mesuraient 1821 pb chez le *G. soja* de même que chez les cultivars Arksoy, Ransom et Tokyo du *G. max* alors qu'elles mesuraient 1823 pb chez le cultivar Columbus du *G. max*. Ces tailles sont les plus petites jamais publiées pour des séquences IGS d'origine végétale. Pour chacun des cinq cultivars, deux clones ont été séquencés et seuls 22 sites (1,2%) se sont avérés polymorphes. La variation observée entre différents clones dérivées du même individu était égale à celle observée entre différents cultivars, une observation qui vient appuyer les travaux antérieurs ayant montré un faible niveau de variation génétique chez le soya cultivé. Les séquences IGS de soya ont été alignées avec celles de six autres séquences publiées chez les légumineuses et ces comparaisons ont fait ressortir deux régions d'homologie. La première s'étend des positions 706 à 1017 de la séquence IGS du soya et se termine à l'endroit d'un possible promoteur qui semble être conservé chez toutes les légumineuses. La seconde région est située à l'intérieur de l'espaceur externe transcrit en 5', s'étend des positions 1251 à 1823 chez le soya et inclut des séquences appartenant à ce qui a été appelé les sous-unités répétées IV-1 et IV-2 chez le *Vicia angustifolia*. Des séquences homologues à ces sous-unités répétées ont été identifiées chez toutes les espèces de légumineuses examinées et elles sont appelées ici «sous-unités répétées cryptiques» (CS) compte tenu des similitudes observées (79–96% pour la région CS-1 et 60–95% pour la région CS-2). Des comparaisons entre les séquences CS-1 et CS-2 à l'intérieur des diverses espèces montrent que leur divergence (substitutions, insertions et délétions) est suffisamment grande pour rendre difficile à déceler la nature répétée de ces séquences par des méthodes analytiques conventionnelles («dot plot»). L'absence de sous-unités répétées dans la moitié 5' de l'IGS du soya soulève des interrogations sur le rôle que ces séquences répétées pourraient jouer dans la terminaison ou l'augmentation de la transcription.

Mots clés : *Glycine*, fabacées, ADN ribosomique, promoteur.

[Traduit par la Rédaction]

Corresponding Editor: R.J. Kemble.

Received August 11, 1997. Accepted December 1, 1997.

D.L. Nickrent¹ and J. A. Patrick. Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901–6509, U.S.A.

¹ Author to whom all correspondence should be addressed (e-mail: nickrent@plant.siu.edu).

