

Chapter 5.7

INFORMATION TRANSFER: MAPPING AND CLONING IN OTHER LEGUMES

Lene Heegaard Madsen and Jens Stougaard

Laboratory of Gene Expression, Department of Molecular Biology, University of Aarhus, DENMARK.

Email: lhm@mb.au.dk

Phone: +45 8942 5011 Fax: +45 8620 1222

Keywords: Information transfer, legume, genetic markers

One of the big advantages of a model legume is the relative ease with which information from gene isolation and characterization of gene function can be transferred to other plants, including crop plants such as pea, bean, soybean, and peanut. This transfer of information includes both the cloning of small specific DNA regions of known map position in the model, for positioning on the crop genetic map, and the cloning of entire genes encoding functions of scientific or more applied interest. One important achievement from this type of work will be transfer of genetic markers from the model's high-density map to linkage maps of crop legumes. This will ease genetic map building, improve the density of the crop genetic map, and help in identification of molecular markers for breeding purposes (see also Chapter 5.6 of this Handbook). Taking advantage of the fast progress of Lotus genome sequencing, a limited number of genetic markers can provide invaluable information on genes located between the same markers in syntenic regions within crop genomes.

INTRODUCTION

Information transfer work involving comparative analysis of important control genes such as *Nin*, *SymRK*, *Nfr1*, and *Nfr5* from different legumes with different nodulation types and specificities towards diverse rhizobial strains and lipochitin-oligosaccharide Nod-factors may answer many questions. For instance, this approach may give insight into genome evolution, levels of general or regional syntony between model and crop genomes, and the biodiversity of symbiosis (or of any other field of interest). Here, information transfer from a model plant to a crop plant is illustrated by describing the isolation of a pea gene using information from Lotus. A number of different approaches are described and resources available in public domain databases included.

GENE CLONING

Once a gene with an interesting and important function is cloned in *Lotus*, one of the obvious applications is comparative analysis of structure and function after isolation of the corresponding gene from another legume of interest, in our case often pea. The choice of pea is based on several criteria. Pea belongs to a different cross-inoculation group than *Lotus*. The microsymbiont of *Lotus* is *Mesorhizobium loti* and *Lotus* forms determinate nodules. In contrast pea develops indeterminate nodules in symbiosis with *Rhizobium leguminosarum* bv. *viciae*. This biological diversity clearly adds to the potential information value of analysing the same nodulation gene in two different legumes. Furthermore, a very large collection of pea mutants exist and many of these are available from the pea germplasm collection at JIC Norwich, UK (www.jic.bbsrc.ac.uk/germplas/pisum/index.htm) or from researchers in the field. These mutants are invaluable when it comes to proving that a gene isolated from pea is actually the orthologue of the *Lotus* gene. If a pea mutant with a phenotype comparable to that of the *Lotus* mutant from which the gene was originally isolated is available, formal proof of functional homology of the genes from *Lotus* and pea can be obtained by showing that the pea mutants carry mutations in this particular gene. Another possibility is complementation of *Lotus* mutant(s) with the pea gene. This approach will however only succeed if the gene function is well conserved.

Direct hybridization strategy for gene isolation

If genomic libraries or cDNA libraries from appropriate tissue is available for the legume of interest, a direct hybridization approach using a probe derived from the *Lotus* sequence can be used. However, some complications caused by non-specific hybridization signals may be encountered because lowered stringency of hybridization would be necessary to ensure detection of signals from target DNA. Furthermore, the gene region used as a probe should be chosen with care, especially if the gene in question contains regions coding for well-conserved domains. For example, kinase domains or homeodomains, which are very well conserved between proteins of quite different function, should be avoided. If possible a region encoding an amino acid segment specific for this particular gene should be chosen and regions encoding e.g. signal peptides or transit peptide should be avoided since these are not nearly as well conserved as the rest of the coding sequence.

If sequence information is available only from the model, a good test for hybridization specificity and the stringency conditions that allow detection of target without cross hybridization to more distantly related sequences is to perform a genomic Southern blot using the model probe on a number of different digests of target plant genomic DNA. Enzymes that give good results are standard 6 base pair cutters like *HindIII*, *EcoRI*, *EcoRV*, *XbaI*, *DraI*, to mention a few. Hybridization at lowered temperature (e.g. 55°C) could be tried as a first approach, however the exact conditions must be determined experimentally and depends on the relatedness of the sequences in question. When using a *Lotus* probe to detect target sequences from pea we most often use hybridization at normal stringency (65°C, standard hybridization buffer, (see below)) followed by washes using 2xSSC, 0.1% SDS and

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 243-248.
<http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4>

then, if filters have high background when checked with a monitor continued washing using 1xSSC, 0.1% SDS at 65°C.

After exposure, complexity of the hybridization signal will show whether the probe is specific for the desired target or whether conditions more stringent (lower salt, higher temperature) are required. If this hybridization test shows a low complexity of the signal, one may choose to continue using the heterologous probe for cDNA or gene isolation. However, if the pattern is complex, or if difficulties from cross hybridizing related sequences such as members of gene families must be avoided, an alternative strategy is to develop PCR primers to amplify a fragment of the homologous gene from the DNA of the target organism. This will then allow hybridization to be carried out at high stringency.

PCR strategy to isolate a gene fragment from the DNA of the target plant

Comparison of Lotus gene sequences to available gene or EST (expressed sequence tag) sequences from different legumes or even Arabidopsis (Figure 1) often leads to identification of regions of sufficient DNA sequence conservation for design of only slightly degenerate PCR primers. Such primers can then be used in PCR reactions with low annealing temperature to amplify similar sequences from the target DNA. If cDNA sequences are available from other legumes, these are most conveniently presented in the TIGR databases (The Institute for Genomic Research, TIGR: www.tigr.org/tdb/tgi/). In this database, sequences are searchable on an organism basis, and the result often quickly gives a good overview of the complexity of a gene family (and therefore the risk of amplifying the "wrong" homolog), and gives information about the tissues from which the relevant ESTs were isolated. This may also give a hint whether the sequence is relevant or not, although expression in unexpected tissues is not *per se* a reason to exclude a sequence from analysis.

However, mere sequence conservation between expressed sequences such as ESTs is not sufficient information to design functional primers. Another parameter is the position of introns, which might interrupt the conserved regions making them unsuitable for primer design. One such example is shown in Figure 2. By simple comparison of three legume ESTs for which no gene structure information was available seven well-conserved regions for primer design were identified. However, by superimposing the intron/exon structure of the corresponding Arabidopsis gene only two of the primers remained, while the others were likely to span intron/exon junctions. One might argue that the intron/exon structure may not be conserved between legumes and Arabidopsis, however in our experience this is very often the case, and therefore comparison to Arabidopsis gene structure is a valid tool in this kind of analysis.

Similarity region I	Identities = 30/33 (90%)
<i>Nin</i> 692 TGGATCTTTAG CCCTTCCTGTCTTCGAAAGAGG 724	
at4g35270 795 TGGATCATTAGCCCTTCCTGTGTTTGAAGAGG 827	
Similarity region II	Identities = 36/40 (90%)
<i>Nin</i> 1275 TTTGTTTT GGAGTTTTTCCTTCCAAAAGATTGCCATGACA 1314	
at4g35270 1426 TTTGTGTTGGAGTTTTCTTTCCAAAAGCTTGCCCTTGACA 1465	
Similarity region III	Identities = 52/57 (91%)
<i>Nin</i> 1843 AAGTTCTA AGACAGTACTTTGCAGGAAGC ...	
at4g35270 1988 AAGTTCTTCGACAATACTTTGCAGGAAGC...	
<i>Nin</i> ... CTAAAAGATGCAGCAAAGAGCATTGGTG 1899	
at4g35270 ...CTCAAAGATGCAGCCAAGAGCATTGGTG 2044	

Figure 1. Regional matches between Lotus Nin mRNA and Arabidopsis at4g35270 mRNA. Although the overall similarity between the two sequences is low (less than 40%), three regions with high nucleotide similarity between Lotus Nin mRNA and the Arabidopsis homolog at4g25270 were detected in a blastn search of the NCBI non-redundant database (www.ncbi.nlm.nih.gov/blast) using the Lotus Nin mRNA as the query. These regions had sufficient sequence conservation to design slightly degenerate primers based on the sequences shown in bold. The forward primer 5'-GCCCTTCCTGTSTTYGAAAGAGG-3' derived from region I, and the reverse primer 5'-AGGCTTCCTGCAAAGTAYTGTC-3' derived from region III can be used in a PCR reaction at ~50°C to produce a ~1500 bp fragment of the pea Nin homolog, while combination of the same forward primer and the reverse primer 5'-GGCAAKCTTTTGAARGAAAACTCC-3' derived from region II gives an ~850 bp fragment (Borisov et al. 2003).

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 243-248.
<http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4>

Overview of primer site identification strategy

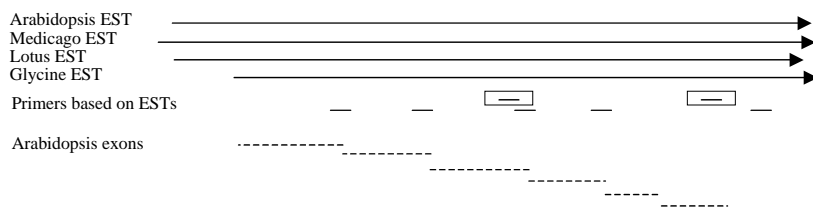


Figure 2. Simple sequence alignment of homologous EST sequences from Medicago, Lotus, and Glycine identified 7 well conserved regions suitable for primer design. After superimposing the exon/intron structure of the corresponding Arabidopsis gene, only two remained (boxed)

5'-RACE and 3'-RACE

If no libraries are available for the organism of interest (or if the desired clone is not found) 5'- and 3'-RACE (rapid amplification of cDNA ends) and subsequent cloning and sequencing is a good alternative to obtain the full coding sequence of the gene. To optimize this approach it is important to know the expression pattern of the gene in order to isolate RNA in which the transcript is present. This is often known from studies performed in the model plant. In our experience, the SMART-RACE kit from Clontech (K1811-1) is very efficient for both 5'- and 3'-RACE and this has given the best results for genes of low expression level. However, also the procedure described in the 5'- 3'- RACE kit from Roche works quite well (Cat. No. 1 734 792).

After cDNA isolation and sequence determination, the sequence of the corresponding gene can be determined using primers for the cDNA to amplify the genomic region. Primers that accidentally span exon-exon junctions will not be productive, but if introns are not too long, amplification of the whole coding region of the gene is a relatively straightforward task and the exon/intron structure can then be derived. To obtain information of non-transcribed regions such as the promoter, isolation of a genomic clone is required. Alternatively, more sophisticated PCR methods, such as Inverse PCR (Sambrook et al. 1989) or TAIL PCR (Liu and Whittier 1995) can be applied.

Identification of homologous mutants

To provide formal proof that the gene isolated from the crop plants is the functional homolog of the gene from the model, several different approaches can be envisioned. One is the functional complementation of mutants identified in the model by transfer of the crop plant gene. This approach is limited to genes encoding proteins in which all functional domains are well conserved. Another approach involves the identification, in the crop plant, of mutants with the same phenotype as the mutants defining the model plant gene and subsequent sequencing of the crop plant mutant alleles.

A.J. Márquez (Editorial Director). 2005. *Lotus japonicus* Handbook. pp. 243-248.
<http://www.springer.com/life+sci/plant+sciences/book/978-1-4020-3734-4>

Transfer of molecular markers for genetic map anchor points

One of the long term applied goals of mapping efforts in model plants is to unify the genetic systems of the models with that of the crops. To this end, we are currently trying to correlate the genetic maps of *Lotus* and *Arachis* (peanut) and we hope to extend this analysis to other crop legumes such as bean (see chapter 5.6 of this Handbook). In short, our aim is to transfer map position information of so-called anchor markers between the two maps. Assuming a limited number of rearrangements during legume genome evolution, we expect to find colinearity/synteny between marker positions on corresponding chromosomes or chromosome regions of the two maps. Exploiting the wealth of information accumulating from the *Lotus* sequencing project it will be possible to make predictions about gene content and gene identity between anchor markers in the crop. Furthermore, if a gene or trait of interest is pinpointed between two markers in the crop it may be possible to derive new, closer markers and finally identify candidate genes simply by inspecting the corresponding genomic region of *Lotus*, as has been elegantly demonstrated between barley and the monocot model rice (Collins et al. 2003).

REFERENCES

- Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, Sandal N, Mortensen A, Schauser L, Ellis N, Tikhonovich IA, and Stougaard J. (2003) **The *Sym35* Gene Required for Root Nodule Development in Pea Is an Ortholog of *Nin* from *Lotus japonicus***. *Plant Physiology* 131, 1009–1017.
- Collins NC, Thordal-Christensen H, Lipka V, Bau S, Kombrink E, Qiu JL, Huckelhoven R, Stein M, Freialdenhoven A, Somerville SC, and Schulze-Lefert P. (2003) **SNARE-protein-mediated disease resistance at the plant cell wall**. *Nature* 425, 973-977.
- Liu YG, and Whittier RF. (1995) **Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking**. *Genome* 25, 674-681.
- Sambrook J, Frisch EF, and Maniatis T. (1989) **Molecular Cloning, a laboratory manual**. Second edition, Cold Spring Harbor Laboratory Press.