Chapter 13 in: **Which DNA Marker for Which Purpose?** Final Compendium of the Research Project **Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees** in the European Union DGXII Biotechnology FW IV Research Programme **Molecular Tools for Biodiversity**. Gillet, E.M. (ed.). 1999. URL http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Isolation and sequence analysis of oak and spruce cDNA clones

M. Berenyi[1], S. Fluch[1], K. Hohl[1], K. Burg[1]*, R. Schubert[2], R. Riegel[2], G. Müller-Starck[2]

[1] Austrian Research Centers, Life Sciences, Department of Biotechnology, 2444 Seibersdorf, Austria

[2] Technical University of Munich, Faculty of Forest Sciences, Section of Forest Genetics, Am Hochanger 13, 85354 Freising, Germany

*Corresponding author: Email: kornel.burg@arcs.ac.at

Introduction

The adaptation of individuals as well as populations to the environment is not well understood because of the lack of knowledge of the genes involved in these processes. The highly developed and so far preferentially used marker systems, such as RAPDs, AFLPs, nuclear SSRs, and rDNA ITS, mainly represent repetitive regions of the nuclear genome, with the rDNA markers being further confined to the nucleolar organizer regions (NOR). These neutral markers are ideal for assessing the genetic diversity as well as evolutionary relationship of populations, since the distribution of their variation is presumably not influenced by selective forces. Close linkage of any important genes controlling adaptive characters with molecular markers currently available can only be a fortuitous and rare occurrence. There is therefore a need to develop additional DNA polymorphisms representing genes involved in adaptive processes; such non-neutral markers could indicate the differentiation of populations on the basis of selective and adaptive features. The recently developed DNA microarray technology could facilitate the discovery of genes modulated by environmental changes (Schena *et al.*, 1995), thus identifying genes involved in adaptation processes. However, this technology requires the availability of cDNA (complementary DNA) sequences to be tested for differential expression under various environmental conditions. Complementary DNA sequences are the representatives of the messenger RNA (mRNA) transcripts of the expressed genes. It is also known that useful sequence variation for polymorphic DNA markers can often be found at non-translated parts of mRNA sequences (Gil *et al.* 1997). Therefore, here we describe the isolation and sequence characterisation of cDNA clones randomly selected from cDNA libraries of oak and spruce.

Material and Methods

Tissue culture

Suspension cultures of a locally established *Quercus petraea* embryonic line were maintained in P24 medium. For osmotic treatment the medium was diluted four times with distilled water. Preliminary experiments showed that cell growth was not significantly influenced under such

nutrition conditions for several weeks (not shown).

Isolation of total RNA from plant cells

The modified method of Cathala *et al.* (1983) was used as follows: 9.2g oak tissue culture was homogenized in liquid N_2 until it became fine powder. 30ml of lysis buffer (5M guanidium monothiocyanate dissolved at 60°C, 10M EDTA, 50 mM Tris pH 7.5, 8% ß-Mercaptoethanol) were added to the frozen samples and carefully mixed during melting (the sample was divided into aliquots to be able to handle the volume). Then the samples were centrifuged at 10,000 rpm. for 20 min. Afterwards the supernatant was transferred into sterile 34ml Beckman centrifuge tubes, four volumes of 7M LiCl were added and incubated at 4°C overnight (or for a minimum of 12 hours). The samples were centrifuged 90 min. at 14,000 rpm at 4°C in a Sorvall S34 rotor. The pellets were resuspended in 5ml 3M LiCl and centrifuged again for 60 min. at 15,000 rpm. 3ml solubilisation buffer (0.1% SDS, 1mM EDTA, 10mM Tris pH7.5) were added to dissolve the pellets. The samples were frozen again and vortexed during melting. Then the RNA was extracted with an equal volume of phenol and then phenol-chloroform. Subsequently 0.1 volume of 3M NaAc (pH 4.9) was added and the chloroform extraction was repeated. The samples were precipitated with 2.5 volumes of ethanol (abs.) at -70°C for 2 hours, then centrifuged for 20 min. at 10,000 rpm and washed with 80% Ethanol (chilled to -20°). The pellets were dried under vacuum and dissolved in a total volume of 400 µl solubilisation buffer. The RNA concentration was determined by photometer at 230 and 260nm wavelengths, and finally 2µg was checked on a 1.5% sterile agarose gel in 1x TBE.

Poly A+ RNA purification

The poly A+ RNA was purified by Dynabeads (Dynabeads Oligo(dT)25) according to the manufacturer's specifications.

Library Construction:

The oak cDNA libraries were constructed with Clontech's SMART® PCR cDNA Library Construction Kit according to the instructions of the user manual. The parameters of the libraries are presented in Table 1.

The Norway spruce cDNA library originates from photomixotrophic suspension cells of *Picea abies* (L.) Karst. PolyA+ RNA was isolated after treatment with a fungal elicitor (Galliano *et al.*, 1993).

DNA sequencing

The selected clones were sequenced on an ABI 373XL sequencer, using the ABI Prism's BigDye® Terminator Cycle Sequencing Ready Reaction Kit with the following modification: Instead of 8µl Terminator Ready Reaction Mix only 4µl plus 4µl halfBigDye (Sigma) were taken, and 4 instead of 3.2 pmol of each primer were used. Nearly all identified clones were sequenced to full length on both DNA strands by an oligonucleotide walking strategy.

BLAST search

The obtained oak DNA sequences were compared directly to DNA sequence databases by the BLAST 2.0 search system (NCBI). They were also translated to putative amino acid sequences and then compared to the Swissprot database. The Norway spruce nucleotide sequences were compared

to all main public databases using the network WU-BLAST similarity search server of the Swiss Institute for Experimental Cancer Research.

Results and discussion

Oak

One untreated and two osmotic shock induced cDNA libraries have been established from *Quercus petraea* tissue culture cells. The initial clone number of the libraries was about 3 x 106 each (Table 1) which allows the isolation of cDNA clones representing low copy number mRNA species as well. The average length of inserts in the cDNA clones varied from 792 to 881 basepairs. In the present study fifty randomly selected clones were picked from each library and their insert size established. Clones containing inserts shorter than 500 bp were discarded. The remaining 82 clones were sequenced mostly to full length. Sixty out of the 82 clones proved to contain an appropriate 3' end of the mRNA represented by the presence of the poly A tail (Table 2). The analysed clones have an average insert size of 0.9 kb. The putative identity of 27 clones (45%) could be established by the BLAST sequence comparison system. As far as protein function is concerned, it was possible to identify six ribosomal protein sequences (three 60S /clones 6, 31, 70/ and 40S /clones 21, 75, 76/ each) possibly representing abundant mRNA species. Two heat shock (clones 43, 168) and two lipid transfer proteins could also be identified (clones 82, 92). The rest of the putative proteins represent single proteins of a different function (Table 2).

Norway spruce

One-hundred-thirty cDNA clones were sequenced from the fungal elicitor-induced spruce cDNA library. Based on significant homologies with known genes of other organisms, 30% of the spruce clones were identified as housekeeping genes and putative stress-related genes, encoding a broad spectrum of metabolic pathways. Using the sequence data, PCR primer pairs were designed in order to amplify expressed sequence tag (EST) sites in *Picea abies*. For the 18 trees tested, all primer pairs yielded PCR bands matching the size exactly predicted from our cDNA data. In the case of 11 PCR primer pairs, polymorphic amplification patterns were seen in diploid bud DNA extracts. Seven EST markers detected co-dominant inheritance by comparing the banding pattern obtained from the diploid bud-DNA extract to the corresponding haploid megagametophytes. The remaining markers revealed polymorphic bands in the megagametophyte samples, confirming the existence of multigene families. Our results indicate that the number of alleles, which were identified at each locus within a population of 100 trees, varies between two and five. Such markers are suitable tools for the verification of genetic variation within populations and corresponding forest reproductive material. Furthermore, such markers may be utilised in the monitoring of viability selection and genetic loads such as inbreeding.

Conclusions

- We could establish three cDNA libraries of oak representing approx. 3 x 10^6 clones each.
- So far we sequenced and characterised 60 oak cDNA clones, and the identity of 27 clones could be postulated. This represent 45% of the clones sequenced so far.
- One-hundred-thirty Norway spruce cDNA clones have been sequenced.
- Seven EST sites showed co-dominant inheritance in the 18 Norway spruce individuals tested.

References

Cathala G *et al.* (1983) A method for isolation of intact, translationally active ribonucleic acid. *DNA* 2(4): 329-339.

Galliano H, Cabane M, Eckerskorn C, Lottspeich F, Sandermann H, Ernst D (1993) Molecular cloning, sequence analysis and elicitor-ozone-induced accumulation of cinnamyl alcohol dehydrogenase from Norway spruce [*Picea abies* (L.) Karst.]. *Plant Mol. Biol.* 23: 145-156.

Gill RW, Hodgman TC, Littler CB, Oxer MD, Montgomery DS, Taylor S, Sanseau P (1997) A new dynamic tool to perform assembly of expressed sequence tags (ESTs). *Computer Applications in the Biosciences* 13(4): 453-457.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.

Library	Treatment	Average insert length bp	Initial clone number	Partially sequenced clones	Fully sequenced clones
Control	Non-treated	**864**	3.2×10^6	15	1
1H	Hypotonic 1 h	**792**	2.5×10^6	9	-
2H	Hypotonic 2 d	**881**	3.2×10^6	32	3

Table 1: Parameters of the oak cDNA libraries

			Insert length*	Sequenced length		
Library	Clone	Putative identity	kb	3’	5’	Total
Untreated	2	Anther-specific proline-rich protein	1.3	1420		1420
"	5	Initiation factor 5A-2	0.9	902		902
"	6	60S Acidic ribosomal protein P2	0.7	622		622
"	10	Enolase	1.7	525		525
"	11		0.6	511		511
"	12		1.8	386	531	917
"	16		0.5	390		390
"	21	40S Ribosomal protein s15A	0.5	302		302
"	22	Cyclin A	1.0	897		897
"	25		0.7	554		554
"	31	60S Ribosomal protein L7	0.9	901		901

"	39	Vacuolar ATP synthase	1.1	977		977
"	41		0.7	597		597
"	43	Mitochondrial heat shock 70 kd protein	1.1	1027		1027
"	47	G10 protein	1.0	507		507
"	49		0.6	502		502
1H	103		0.7	665		665
"	104		0.5	453		453
"	107		0.8	677		677
"	109		1.4	1286		1286
"	110	Aspartic proteinase precursor	0.9	799		799
"	118		0.8	636		636
"	120		0.5	497		497
"	121		1.1	521		521
"	124		0.5	458		458
2H	53	Biotin carboxyl carrier prot. (BCCP)	1.20	1164		1164
"	54		0.56	430		430
"	55	Small nuclear Ribonucleoprotein E homologue C29	0.65	569		569
"	63		1.40	501	531	1032
"	64		0.73	733		733
"	67		0.85	752		752
"	69		1.10	1040		1040
"	70	60S ribosomal protein L18A	0.80	752		752
"	71	Nuclear pore complex protein (NUP358)	0.75	664		664
"	73	Histon H4	0.72	617		617
"	75	40S Ribosomal prot.s19	0.93	846		846
"	76	40S Ribosomal prot.L22	0.78	715		715
"	77		0.93	875		875
"	79		0.78	671		671
"	80		0.55	430		430
"	81		0.85	794		794
"	82	Non-specific lipid-transfer prot.(LTP)	0.79	638		638
"	85		1.20	1139		1139
"	87		0.58	426		426
"	89		0.83	904		904
"	92	Non-specific lipid-transfer prot.1 (LTP1)	0.80	641		641

"	94		1.30	1273		1273
"	95	Cystationine gamma-lyase	0.62	520		520
"	96	Cysteine protease inhibitor	0.65	541		541
"	97	Protein kinase	1.00	923		923
"	98		0.80	679		679
"	99	RNA pol. II 13.2 kD polypeptide	0.75	635		635
"	100		1.10	227	533	760
"	164		0.65	551		551
"	166		0.80	695		695
"	168	18.5 kD class I heat shock prot.	0.80	641		641
"	172	Tubulin beta-2/beta3 chain	1.70	582	592	1174
"	173		0.85	854		854
"	174	Pyruvate dehydrogenase E1 component	1.00	1138		1138
"	175		0.90	894		894

Table 2: List of the *Quercus petraea* EST clones. * Approximate insert length of the cDNA clone established by agarose gel electrophoresis.