

잣나무(*Pinus koraiensis*)의 cDNA library 제작 및 EST 분석

김준기¹ · 임수빈¹ · 최선희¹ · 이종석¹ · 노승문² · 임용표^{1*}

¹충남대학교 원예학과, ²미국 농무성 화훼분과

Construction of a full-length cDNA library from *Pinus koraiensis* and analysis of EST dataset

Joonki Kim¹, Subin Im¹, Sun Hee Choi¹, Jong-Suk Lee¹, Mark S. Roh², Yong Pyo Lim^{1*}

¹Department of Horticulture, Chungnam National University, Daejeon 305-764, Korea

²USDA, Agricultural Research Service, National Arboretum, Floral and Nursery Plants Research Unit, Beltsville, MD 20750, USA

Received on 11 January 2011, revised on 6 February 2011, accepted on 9 March 2011

Abstract : In this study, we report the generation and analysis of a total of 1,211 expressed sequence tags (ESTs) from *Pinus koraiensis*. A cDNA library was generated from the young leaf tissue and a total of 1,211 cDNA were partially sequenced. EST and unigene sequence quality were determined by computational filtering, manual review, and BLAST analyses. In all, 857 ESTs were acquired after the removal of the vector sequence and filtering over a minimum length 50 nucleotides. A total of 411 unigene, consisting of 89 contigs and 322 singletons, was identified after assembling. Also, we identified 77 new microsatellite-containing sequences from the unigenes and classified the structure according to their repeat unit. According to homology search with BLASTX against the NCBI database, 63.1% of ESTs were homologous with known function and 22.2% of ESTs were matched with putative or unknown function. The remaining 14.6% of ESTs showed no significant similarity to any protein sequences found in the public database. Gene ontology (GO) classification showed that the most abundant GO terms were transport, nucleotide binding, plastid, in terms biological process, molecular function and cellular component, respectively. The sequence data will be used to characterize potential roles of new genes in *Pinus* and provided for the useful tools as a genetic resource.

Key words : Contig, Gene ontology, Homology, *Pinus koraiensis*, Singleton, Unigene

I. 서론

소나무속에 속하는 거대 침엽수인 잣나무(*Pinus koraiensis*)는 러시아 한국 중국 일본의 해안지방에 자생하며 주로 해발 600 - 1000 m의 고도에서 성장한다(Mirov, 1967; Vidakovic, 1991). 성숙한 잣나무는 길이가 30 m에 달하며 직경은 약 1 m 내외이고 목질이 단단하여 오랜 기간 북아시아와 한국에서 양질의 가구 재료로 널리 사용되어 왔다. 또한 추위에 강한 특성으로 인해 캐나다와 미국 북동부 지방에서는 잣나무가 관상용으로 식재되고 있다. 잣나무의 종자는 구과의 열매에서 수확하며 양질의 단백질과 지방산, 아미노산, 탄수화물, 그리고 비타민을 포함하고 있

으며(Yoon, 1987; Han과 Hwang, 1990), 수피 추출물은 항암, 항산화, 노화방지, 돌연변이 억제 등의 효능을 갖는 등 경제적 가치가 매우 높은 수종이다(Asset 등, 1999; Bagchi 등, 2000; Li 등, 2007).

또한 잣나무는 병해충 및 환경오염에 대한 적응도가 다른 수종에 비해 높은 장점을 갖고 있으며, 실제로 William (1966)은 잣나무가 다른 소나무속의 150종 수목에 비해 병해충에 대한 저항성이 높은 것을 보고하였다. 최근에는 산업화에 따른 오염도가 높은 지역에서 잣나무와 리기다 소나무(*Pinus rigida*)의 성장률과 광합성량을 비교하여 잣나무가 상대적으로 공해에 대한 적응도가 높다는 결과가 보고되었다. 특히 잣나무는 중금속 중 망간(Mn)에 대한 저항성이 높은 것으로 보고되었다(Choi 등, 2006). 잣나무는 이처럼 다양한 장점을 지닌 경제적 수종으로 국내 주요 10

*Corresponding author: Tel: +82-42-821-5739

E-mail address: yplim@cnu.ac.kr

대 수목으로 선정되었고 근래 많은 국내외 연구자들에 의해 생태 생리 육종분야에서 폭넓은 연구가 이루어지고 있다. 최근에는 산림청의 임업연구원에서 117 kb 크기의 잣나무 엽록체 서열을 완전 해독하는 등 유전자원 분야로 연구가 확대되고 있다(Noh 등, 2007).

최근 희귀, 멸종위기 생물종 및 특산종에 대한 중요성과 유전자원으로서의 인식이 점차 증가되고 있는 상황이다(ISUN, 2001). 비록 계획적인 산림자원정책의 일환으로 소나무와 참나무의 점유면적증가로 잣나무가 멸종 위기종은 아니지만 생물 다양성 협약을 통해 유전자원의 다양성이 하나의 유형으로 인정되었고 유전자원의 이용에 의해 발생하는 이익배분에 대한 협약이 체결됨에 따라 경제적 가치가 높은 잣나무의 유전자원 선점이 중요하게 되었다. 본 논문은 지리산에 자생하는 국내 특산 수종인 잣나무의 유전자원을 발굴 분석하여 경제적 가치가 높은 주요 자생 식물의 유전자원 확보를 위해 진행되었다.

II. 재료 및 방법

1. 식물재료 및 RNA 추출

잣나무 시료는 2010년 4월 30일 경상북도 상주시 화북면 속리산(36° 34' 11.5" E 127° 52' 15.7" N)에서 흉고 직경 41 cm의 단일 개체에서 유엽을 채집하여 액체질소에 동결하고 -70°C에서 보관하였다. RNA의 추출은 Chang 등(1993)의 방법을 일부 변형하여 사용하였다. 잎 조직 1 g을 막자 사발과 액체질소를 사용하여 곱게 파쇄한 후 15 mL의 추출 용액[2% CTAB, 2% PVP, 100 mM Tris-Cl(pH 8), 25 mM EDTA, 2M NaCl, 0.05% spermidine, 2% β-mercaptoethanol, proteinase K]을 첨가하여 녹인 후 동일 부피의 chloroform을 처리하여 섞은 후 상온에서 10,000 rpm의 속도로 원심분리를 하여 상등액을 분리하였다. RNA 침전을 위해 1/4 부피의 10M LiCl를 첨가하고 4°C에서 12 시간 이상 보관한 후 10,000 rpm으로 20분 동안 원심분리하여 RNA 침전물을 얻었다. RNA 침전물은 500 uL의 SSTE 용액[1 M NaCl, 0.5% SDS, 10 mM Tris-Cl(pH 8), 1 mM EDTA (pH 8)]에 녹여 사용하였다. Total RNA로 부터 Poly (A+)RNA의 분리는 PolyATtract mRNA isolation system(Promega, USA)을 이용하여 제조사의 지침에 따라 수행하였다.

2. 전장 cDNA library의 제작 및 염기서열 분석

전장 cDNA enriched library의 제작을 위해 SMART™ 기술을 사용하였다(Okayama와 Berg, 1982). cDNA의 합성은 1 ug의 poly (A+)RNA를 In-Fusion^R SMARTer™ cDNA Library Construction Kit(BD Biosciences)를 사용하여 제조사의 지침에 따라 진행하였다. 제작된 library는 X-gal 선별 방법을 통해 선별하였고 이중 1,211개 clone을 무작위로 선별하여 염기서열을 결정 하였다. 선별된 각각의 clone은 96 well deep plate에서 배양 후 AccuPrep^R 96 Plasmid Extraction Kit를 사용하여 제조사의 지침에 따라 plasmid DNA를 분리하였다. 염기서열은 M13F primer를 사용하여 Big Dye Terminator Kit Ver 1.1 (Applied Biosystems, USA)로 반응시킨 후 ABI3100 Genetic Analyzer (Applied Biosystems, USA)이용하여 결정하였다.

3. EST 서열 분석 및 gene ontology 분석

분석된 염기서열 중 vector 서열은 CodonCode Aligner Ver.3.7.1를 사용하여 제거하였고, 남은 clone중 50 bp이하의 서열은 분석에서 제외하였다. 남은 서열의 assembling과 clustering은 CAP3 프로그램(<http://deepc2.psi.iastate.edu/aat/cap>)의 기본 설정을 이용하여 분석하였다. 분석된 서열 중 contig로 분류되지 않은 서열은 singlet으로 분류하였으며, 각각의 서열에 대한 상동성 탐색과 gene ontology의 결정은 Blast2GO(<http://blast2go.bioinfo.cipf.es/>) 프로그램을 사용하여 분석하였다(Conesa 등, 2005). BLASTX의 cut off 값은 10⁻⁵로 설정하여 검색하였고 mapping은 E-value 10⁻⁵, annotation cut off 45, GO weight 10으로 설정하여 검색하였다. Microsatellite는 SSR Finder(<http://www.csufresno.edu/ssrfinder/>)를 사용하여 분석하였다. Di nucleotide (nt)는 5번 이상 그리고 tri-nt 이상인 경우 4번 이상 반복된 것만 microsatellite로 분류하였다.

III. 결과 및 고찰

1. 잣나무 전장 cDNA library의 제작

속리산에서 채집한 잣나무의 유엽을 사용하여 total RNA를 추출하였다. 일부 목본류의 경우 조직내의 polyphenol

또는 polysaccharide의 함량이 높아 일반적으로 사용되는 Guanidinium thiocyanate, phenol/chloroform를 사용한 one-step extraction이 어렵다. 잣나무 역시 Trizol을 사용한 방법으로 total RNA추출이 되지 않아 Chang 등 (1993)에 의해 고안된 CTAB(hexadecyltrimethylammonium bromide) 방법을 사용하여 추출하였다. 추출한 RNA는 1% denaturing formaldehyde agarose를 사용하여 정성 분석을 진행하였다. 분리된 total RNA의 농도는 0.489 ug/uL 였고 A260/A280의 비율은 1.63(0.247/0.151)로 확인되었다. 전장 cDNA library의 제작은 SMART™ 기술을 사용하였으며(Okayama와 Berg, 1982), 0.5 ug의 poly(A+)mRNA를 사용하여 1st strand cDNA 합성, dC tailing, template switching을 진행하였다. 2nd cDNA 합성을 통해 만들어진 double strand(ds) cDNA는 CHROMA SPIN™+TE-1000 column을 사용하여 500 bp 이하의 cDNA를 제거하였다. 크기 선별을 완료한 ds cDNA는 in-fusion cloning 방법을 통해 pSMART2IF 벡터에 재조합을 유도한 후 *E. coli* strain *DH10B*에 electro-transformation 방법을 통해 형질전환 시켰다.

2. 염기서열 해독 및 분석

형질전환을 통해 얻어진 EST clone중 1,211개를 무작위로 선별하여 염기서열을 분석하였다. 평균적인 염기서열 분석 길이는 약 906 bp이며 전체 분석 서열길이는 7.8 X 10⁵ bp였다. 유전자의 단백질 암호화 서열의 크기 별 분포

는 0.7 - 0.8 kb가 27.7%(238개)로 가장 높게 나타났다 (Fig. 1). 이중 벡터 서열과 염기서열 분석 질이 낮은 354개를 제외한 857개의 서열에 대한 분석을 진행하였다. CAP3 프로그램을 사용하여 전체 411개의 EST cluster 중 contig 89개 singlet 322개를 확인하였다(Table 1). 89개의 contig 중 가장 많은 EST clone을 포함하는 것은 ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit로 96개의 clone이 중복되었고 이후 metallothionein-like gene 이 66개 그리고 photosystem II (PSII) 10 kDa PsbR subunit가 23개 빈도로 나타났다.

NCBI(National Center for Biotechnology Information) nr database의 BLASTX 검색은 Blast2GO 프로그램을 사용하여 최소 E-value 값을 10⁻⁴으로 설정하여 상동성을 검색하였다(Fig. 2). 이중 no hit는 127개(14.6%), unknown gene은 193개(22.2%) 그리고 known gene은 548개(63.1%)

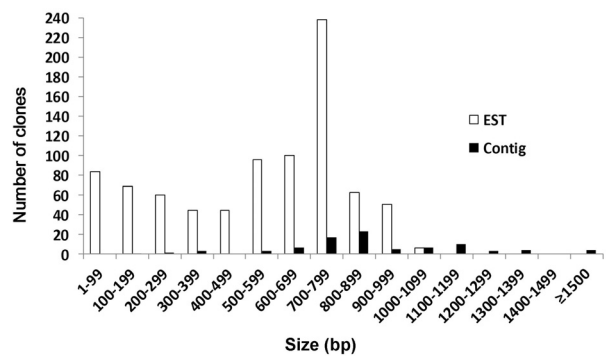


Fig. 1. Distribution of the sequence length of ESTs and contigs.

Table 1. Summary of assembly and characteristics of ESTs and unigenes.

EST sequencing	Number of cDNAs sequenced	1211
	Number of high quality sequence (>50bp)	857
Contig assembly results	Number of ESTs assembled	857
	Number of contigs	89
	Number of singletons	322
	Number of unigenes	411
Number of contigs containing	2-5 ESTs	76
	6-10 ESTs	8
	11-15 ESTs	1 ^a
	16-20 ESTs	1 ^b
	> 21 ESTs	3 ^c

^aPutative photosystem II oxygen-evolving complex protein.

^b26S ribosomal RNA.

^cRibulose-1,5-bisphosphate carboxylase/oxygenase small subunit, metallothionein-like gene and photosystem II (PSII) 10 kDa PsbR subunit.

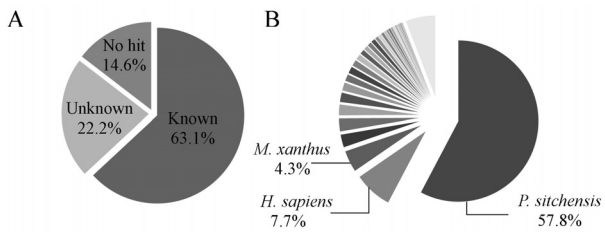


Fig. 2. Classification of all ESTs determined in this study. (A) functional annotation of ESTs with BLASTN. (B) distribution of the number of ESTs in all contigs.

로 확인되었다(Fig. 2A). No hit를 제외한 741개의 종별 분포는 *Picea sitchensis* 428개(57.8%), *Homo sapiens* 57개(7.7%), *Myxococcus xanthus* 32개(4.3%)의 순으로 나타났다. 이중 *Picea sitchensis*는 목본류로 계통분류학적 위치가 같지 않음에도 높은 상동성을 보여주었다(Fig. 2B). 이는 앞의 목본류가 목재로서의 상업적 가치가 높아 많은 연구가 선행되어 나타난 결과로 생각된다. 실제로 2011년 2월 1일기준으로 발표된 NCBI의 EST database(dbEST, <http://www.ncbi.nlm.nih.gov/projects/dbEST/>)에는 미국 서부산 침엽수인 *Picea sitchensis*의 등록된 EST가 186,637개로 다수의 EST가 확보되어있었다. 이러한 현상은 *Homo sapiens*에서 더욱 두드러지게 나타난다. 인간의 경우 보고된 전체 EST 중 12.3%(8,314,462 / 67,370,045)로 보고된 EST database중 가장 높은 비율을 차지하고 있어 계통분류학 위치와 상관없이 잣나무와의 일치도가 높게 나온 것으로 생각된다(Fig. 2B).

전체 1,211 clone에 대한 gene ontology는 Blast2GO 프로그램의 mapping 기능을 사용하여 molecular function, biological process, cellular component로 분류하고 Plant Go Slim 기능을 사용하여 중복되는 GO ID를 최소화 시켰다. 각각의 비율은 molecular function 17.6%, biological process 26.3%, cellular component 56.1%로 나타났다(Fig. 3). 전체 clone중 적어도 한 개 이상의 GO ID가 부여된 경우는 68.9%(599/869)로 나타났으며, clone당 평균적인 GO ID는 3.05개로 나왔고 전체 clone에 대한 총 GO ID는 2,653개로 확인됐다. 기능에 따른 분류에 있어 molecular function은 nucleotide binding, biological process는 transport, cellular component는 plastid가 가장 높은 비율로 나왔다. 분리된 411개의 unigene에서 microsatellite 서열을 분리하여 구조를 분석하였다(Table 2). 각각의 microsatellite는 반복된 서열의 구조에 따라 분류하였고

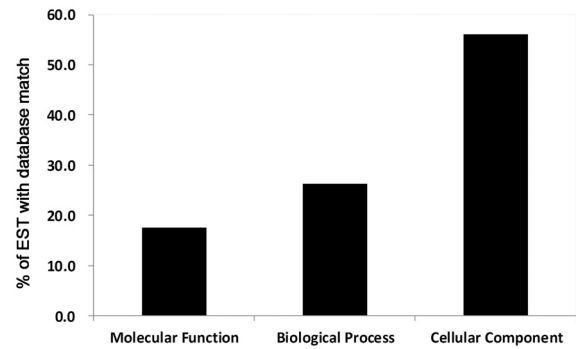


Fig. 3. Graphical representation of functional annotation and distribution of BLASTN top-hit species. (A) classification of all ESTs with BLASTN. (B) BLASTN top-hit species showing high homology with known genome sequences.

전체 77개의 microsatellite를 확인하였다(Table 1). Unigene 내의 microsatellite의 비율은 18.7%로 다른 목본류인 참나무(18.6%)에 비해 낮은 값이 나타났다(Durand 등, 2010). Di-, tri-, tetra-, complex repeat의 빈도는 각각 16.9%(22/77), 59.7%(46/77), 2.5%(2/77), 9%(7/77)로 나타났다. 전체 microsatellite중 (AG)_n과 (AGC)_n의 비율이 각각 12.9%(10/77)와 27.2%(21/77)로 다른 microsatellite에 비해 높은 비율이 확인됐으며, 이중 (AG)_n의 고빈도는 보리, 옥수수, 벼, 수수와 밀의 EST를 기반으로 한 microsatellite 탐색 결과에서도 일치되는 결과가 관찰되었다(Temnykh 등, 1999; Kantety 등, 2002). 그러나 분석에 사용된 전체 EST의 개수가 적어 분리해낸 microsatellite 전체의 비율을 반영한다고 보기 어렵고 이는 추가로 보다 많은 수의 EST 분석을 통해 통계적 유의성을 부여해야한다. 본 연구를 통해 분리해낸 잠재적인 77개의 microsatellite는 잣나무의 분자마커로서 육종 및 유전체연구에 있어 다양한 분야에 활용될 것으로 예상된다.

IV. 결론

본 논문에서는 속리산에서 자생하는 잣나무 유엽조직의 EST library를 제작하여 전체 1,211개의 EST clone을 분석하였다. EST와 unigene 서열의 분석은 컴퓨터를 기반으로 한 filtering과 수작업 그리고 NCBI의 BLAST 분석을 통해 수행하였다. 전체 89개의 contig와 322개의 singleton을 분리하여 총 411개의 unigene을 분리해냈다. 이중 BLASTX를 통해 상동성이 확인되지 않은 14.6%의 no hit 유전자는 잣나무 유엽 특이적 유전자일 가능성이 높을 것

Table 2. Characterization of the repeat structures for the identified microsatellite sequences.

Repeat unit	Repeat sequence	Synonym	Observed No.	Total No.
Di-	AT	TA	8	22
	AC	CA, GT, TG	3	
	AG	GA, CT, TC	10	
	CG	GC	1	
Tri-	AAC	ACA, CAA, GTT, TTG, TGT	4	46
	AAG	AGA, GAA, CTT, TTC, TCT	9	
	AAT	ATA, TAA, ATT, TTA, TAT	1	
	ACC	CCA, CAC, GGT, GTG, TGG	2	
	AGC	GCA, CAG, GCT, CTG, TGC	21	
	AGG	GGA, GAG, CCT, CTC, TCC	4	
	ATC	TCA, CAT, GAT, ATG, TGA	5	
Tetra-	GGAA	GAAG, AAGG, AGGA, TTCC, TCCT, CCTT, CTTC	1	2
	ACTA	CTAA, TAAC, AACT, TAGC, AGCT, GCTA, CTAG	1	
Complex repeat	(AC) ₂ (AT) ₅		1	7
	(AT) ₅ (TA) ₂		1	
	(AT) ₆ (GA) ₅		1	
	(AAT) ₄ (TC) ₂		1	
	(CTG) ₃ (TC) ₅		1	
	(TA) ₂ (AT) ₅ (AG) ₂		1	
	(CA) ₂ (GA) ₂ (AAT) ₅		1	
Total				77

으로 생각된다. 또한 전체 893개의 unigene의 서열 분석을 통해 77개의 microsatellite를 확인하였다. 본 연구를 통해 얻어진 잣나무의 유전자원들은 다양한 분야의 기초자료로 사용되어질 것으로 사료된다.

감사의 글

본 연구는 국립생물자원관의 “야생생물 유래 유용유전자 발굴 분석연구 사업”의 지원에 의해 수행되었으며, 저자 중 김준기·임수빈은 논문작성에 동일한 기여를 함.

참고 문헌

- Asset G, Staels B, Wolff RL, Bauge E, Madj Z, Fruchart JC, Dallongeville J. 1999. Effects of *Pinus pinaster* and *Pinus koraiensis* seed oil supplementation on lipoprotein metabolism in the rat. *Lipids* 34: 39-44.
- Bagchi D, Bagchi M, Stohs SJ, Das DK, Ray SD, Kuszynski CA, Joshi SS, Pruess HG. 2000. Free radicals and grape seed proanthocyanidin extract: Importance in human health and disease prevention. *Toxicology* 148: 187-197.
- Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11(2): 113-116.
- Choi DS, Kayama M, Jin HO, Lee CH, Izuta T, Koike T. 2006. Growth and photosynthetic responses of two pine species (*Pinus koraiensis* and *Pinus rigida*) in a polluted industrial region in Korea. *Environmental Pollution* 139: 421-432.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.
- Durand J, Catherine B, Chancerel E, Frigerio JM, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn HP, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herrm A, Ikaran Z, Caban C, Ueno S, Alberto F, Dumoulin PY, Guichoux E, Daruvar A, Kremer A, Plomion C. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* 11: 570-582.
- Han SS, Hwang BH. 1990. Analysis of amino acid, fatty acid and vitamin in Korean pine(*Pinus koraiensis*) seed. *Jour. Korean For. Soc.* 79(4): 345-351.
- ISUN. 2001. The ISUN red list categories (Version 3.1). ISUN Species Survival Commission. Gland, Switzerland.

9. Kantety RV, La Rota M, Matthews DE, Sorrells ME. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48: 501-510.
10. Li K, Qingwang L, Jian L, Zhang T, Han Z, Gao D, Zheng F. 2007. Antitumor activity of the procyanidins from *Pinus koraiensis* bark on mice bearing U14 cervical cancer. *Yakugaku Zasshi* 127: 1145-1151.
11. Mirov NT. 1967. *The Genus Pinus*, Ronald, New York.
12. Noh EW, Lee JS, Choi YI, Han MS, Lee H, Yi YS, Han SU. 2007. Hotspots in chloroplast genome as new tools to study relationship among pines. *Korean White Pine (Pinus koraiensis)II*: 41-50.
13. Okayama H, Berg P. 1982. High efficiency cloning of full-length cDNA. *Mol. Cell Biol.* 2: 161-170.
14. Temnykh S, Park WD, Ayers N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR. 1999. Mapping and genome organization of microsatellites in rice (*Oryza sativa*). *Theor. Appl. Genet.* 100: 697-712.
15. Vidakovic M. 1991. *Conifers: Morphology and Variation*, Crafickizavod Hrvatske, Croatia.
16. William BC. 1966. Geographic distribution on the pines of the world. 32: 5.
17. Yoon TH. 1987. Fatty acid composition of total lipids from seeds of *Pinus koraiensis*. *J. Korean Soc. Food Nutr.* 16(2): 93-97.