An Oat (*Avena sativa* L.) cDNA Library Derived from Three Seed Development Stages

Aaron D. Beattie, Peter E. Eckstein, Tom Zatorski, Graham J. Scoles, Brian G. Rossnagel

Department of Plant Sciences/Crop Development Center, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5A8, CANADA

Introduction

Consumption of oat products can lower the risk of coronary heart disease which has led to its approval as a "heart-healthy" food by the FDA. The biologically active compound primarily responsible is soluble fibre (principally β-glucan). Oat also contains a number of antioxidants, such as vitamin E and the avenanthramides, which have been associated with preventing atherosclerosis and cholesterol accumulation.

To better understand the biosynthetic pathways controlling the production of these and other compounds of interest, three cDNA libraries were created from developing oat seeds. The ESTs derived from these libraries will be a useful resource suitable for microarray and TILLING studies to address questions related to the regulation and production of these compounds. Additionally, they can be mined for molecular markers (SSRs and SNPs). This library doubles the ~17,000 oat ESTs reported in the literature or NCBI database and represents the only source of seed-derived ESTs.

Materials and Methods

- Developing seeds from 'CDC Dancer' oat were collected at the watery, early milk and late milk/early dough stages (Zadoks scale 71, 73 and 77-83, respectively).
- Total RNA was extracted from four seeds (100mg), followed by poly(A)+ RNA purification using DynaBeads (Invitrogen).
- cDNA was synthesized with the Creator SMART cDNA Synthesis Kit (Clontech).
- Sequences were obtained from 5' single pass reads.
- EST data was handled with an in-house (NRC-PBI) database management and analysis system (FIESTA v.2). Poly(A)+ tails and vector were trimmed and the sequences clustered to produce a unigene set which was used to query the UniProt Plants protein database (Evalue cut-off of 1e-6). Annotation of unigene functions used Gene Ontology (GO).

Table 1. Summary of cDNA libraries created from developing oat seed.

Stage	ESTs	Quality Sequences	Unique Transcripts (Unigenes)			Discovery Rate
			Singletons	Contigs	Total	
Watery	4032	3249	1805	455	2260	70%
Early milk	4992	4601	1308	267	1575	34%
Late milk*	10656	9861	3626	1665	5291	55%
Combined	19680	17711	5644	2418	8062	46%

* Normalized

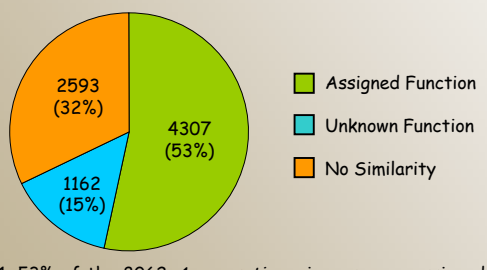

Figure 1. 53% of the 8062 *Avena sativa* unigenes were assigned a GO annotation, 15% matched a protein with unknown function while 32% showed no or poor alignment to proteins in the UniProt Plants database.

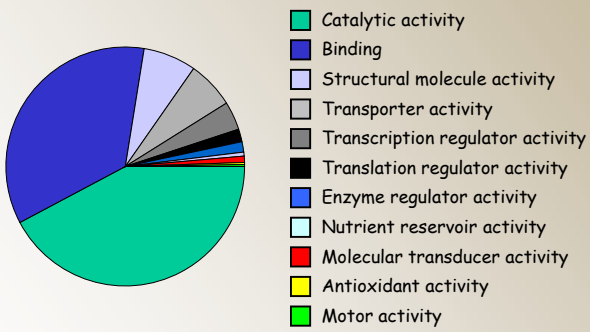

Figure 2. Molecular functions assigned to unigenes using the GO annotation system.

Table 2. Total number and percentage of the most abundant transcripts in the two non-normalized libraries.

BLAST Hit	UniProt Name	Early Milk		Watery	
		No.	%	No.	%
12 S globulin 2 precursor (*A. sativa*)	SSG2_AVESA	526	11.4	0	0.0
Puroindoline (*A. fatua*)	Q9XET6_AVEFA	423	9.2	4	0.12
12 S globulin precursor (*A. sativa*)	O49258_AVESA	388	8.4	0	0.0
12 S globulin 1 precursor (*A. sativa*)	SSG1_AVESA	251	5.5	0	0.0
Avenin-3 precursor (*A. sativa*)	AVE3_AVESA	206	4.5	2	0.06
Grain softness protein (*H. vulgare*)	Q0GIL5_HORVU	164	3.6	4	0.12
11 S globulin (*A. sativa*)	Q38780_AVESA	143	3.1	3	0.09
Lectin precursor (*H. vulgare*)	AGI_HORVU	73	1.6	19	0.58
Tryptophanin (*A. sativa*)	A7U440_AVESA	72	1.6	1	0.03
Grain softness protein (*H. vulgare*)	Q5ITH3_HORVD	67	1.5	0	0.0

Results and Discussion

- 19,680 cDNA clones were sequenced with 17,711 (90%) providing high quality sequence information. The ESTs were grouped into 2418 contigs and 5644 singletons (8062 unigenes) (Table 1).
- Putative biological functions could be assigned to 53% of the unigenes (Figure 1). The unigenes were categorized into 11 molecular function groups (Figure 2).
- The most abundant ESTs were seed storage proteins (globulins, avenins) and seed hardness proteins (puroindolines) (Table 2).
- Only 18 unigenes were previously identified in tame oat (*A. sativa*) and a further 5 in wild oat (*A. fatua*).
- Good representation of enzymes from the vitamin E and starch synthesis pathways, enzymes involved in β-glucan and fatty acid synthesis were also identified with some representation from the avenanthramide pathway.

Acknowledgements

This work was conducted as part of the NAPGEN (Natural Products Genomics Resource) initiative coordinated and funded by NRC-PBI (Canada). Additional funding was provided from the Crop Development Center oat variety pedigreed seed royalty funds.