

2009 Report: Clone-based Functional Genomics Resources (ORFeomics)

Prepared by Joe Ecker (Chair, ecker@salk.edu)

Activities during the last year represent a significant shift from the past. While some additional unique ORFs have been produced, large scale Open-Reading-Frame (ORF) clone production has transitioned from ORFeome production to the construction of large set of 'destination' or 'functional' clone sets (Table 1). Examples of functional clone collection may include clone sets for *in planta* overexpression, yeast two-hybrid interactome mapping, tagged ORF expression, etc. that are derived from the ORFeome collection.

Regarding unique ORF clone production (gene ORF clones not currently available in any form) and deposition during the past year, Salk Institute Genomic Analysis Laboratory (SIGnAL) deposited 192 additional ORF clones with ABRC. In addition, it is anticipated Gateway ORF clones produced from the ATOME project will soon be available in a French stock center (<http://urgv.evry.inra.fr/ATOME/index.cgi>) and that ~500 of the unique ORF clones will be deposited with ABRC. Finally, the RIKEN Plant Science Center (PSC) project to collect full-length cDNAs (clone with 5' and 3' UTRs) from *Arabidopsis thaliana* is now completed and all RAFL clones produced were deposited with the RIKEN Bio Resource Center (BRC). In addition, the RIKEN group has collected full-length cDNAs from the salt-tolerant *Arabidopsis* relative, *Thellungiella halophila* (1). These cDNA clones are available from RIKEN BRC. More recently, RIKEN PSC activities have shifted to the collection of full-length cDNAs from various crops and trees.

Regarding 'functional' *Arabidopsis thaliana* ORF clone collections, several new large "destination" vector clone sets were deposited in ABRC during the past year. Examples of large clone sets deposited in ABRC include: 588 ORF expression clones in a pLIC-C-TAP vector deposited by the Kumar/Snyder: Arabidopsis Protein Chip Project and 18,288 yeast two-hybrid expression clones (pDEST-AD: 9,152 ORFs and pDEST-DB: 9,106 ORFs) deposited by the Vidal/Hill/Ecker (CSSB/Salk) Plant Protein Interactome project.

Finally, there remain a significant number of annotated genes with no evidence of expression and/or no cDNA/ORF clone (see thermometers, page 17-18). The development of a new method called RNA-Seq (Lister et al. 2009) for deep, strand-specific transcriptome sequencing will likely allow the identification of transcripts for many of the remaining annotated genes along with novel spliced forms for other genes. For example, RNA-seq may be combined with flow sorted cell samples prepared from many distinct cell types to identify rare/low expressed transcripts. Future ORF clone production for the remaining genes with no ORF clone will likely utilize deep paired-end RNA-Seq information for *in silico* gene model construction followed by standard RT-PCR subcloning/sequencing approaches or, more likely, with ever decreasing cost, these ORFs may simply be chemically synthesized.

References:

1. Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523-36.
2. Taji T, Sakurai T, Mochida K, Ishiwata A, Kurotani A, Totoki Y, Toyoda A, Sakaki Y, Seki M, Ono H, Sakata Y, Tanaka S, Shinozaki K. (2008) Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, *Thellungiella halophila*. *BMC Plant Biol.* 8:115.

Table 1. Arabidopsis ORF and cDNA clone repertoires*

Creator	Format	Focus	Validation	Scale	URL	Stock center
ORF clones						
SSP consortium & Salk Institute	Univector pUNI51		Full sequence	14,398	signal.salk.edu/cdnastatus.html http://methylo.me.salk.edu/cgi-bin/clones.cgi	ABRC
Salk/Invitrogen	Gateway entry		Full sequence	12,114	signal.salk.edu/cdnastatus.html http://methylo.me.salk.edu/cgi-bin/clones.cgi	ABRC
CCSB/Salk	Y2H clones	Plant Interactome	Full sequence	18,258	http://interactome.dfc.harvard.edu/A	ABRC
TIGR	Gateway entry	Network Map Hypothetical genes	Full sequence	3,041	_thaliana/host.php www.tigr.org/tdb/hypos/	ABRC
Peking-Yale Joint Center	Gateway entry	Transcription factors	5' and 3' end seq.	1,282		ABRC
Dinesh-Kumar et al.	Gateway expression	TAP-tagged transcription factor	5' and 3' end seq.	1,281		ABRC
REGIA	Gateway entry	Transcription factors	5' and 3' end seq.	962	gabi.rzpd.de/materials/	GABI/RZPD
Dinesh-Kumar et al.	Gateway entry, no stop pLIC-CTAP	Plant protein chips	5' and 3' end seq.	7,300	plants.gersteinlab.org/	ABRC
ATOME 1	Gateway entry		5' and 3' end seq.	1,809	urgv.evry.inra.fr/orfeome/	CNRGV
ATOME 2	Gateway entry, no stop	Originates from SSP	5' and 3' end seq.	3,476	same	CNRGV
Doonan et al.	Gateway Expression	GFP fusion for subcellular location		155		ABRC
Callis et al.	Gateway entry	Protein ubiquitination	Full sequence	111	plantsubq.genomics.purdue.edu	ABRC
Sheen et al.	Expression	Epitope tagged MAPK	Full sequence	100	genetics.mgh.harvard.edu/sheenweb/category_genes.html	ABRC
cDNA clones						
RIKEN/SSP/Salk Insitute	λ ZAP or λ PS		Full sequence	22,614	www.brc.riken.go.jp/lab/epd/Eng/order/order.shtml	BRC
MPI-MG	Gateway expression		5' end seq.	4,500	gabi.rzpd.de/materials/	GABI/RZPD
Génoscope/LTI	Gateway entry		Full single pass seq.	28,866	www.genoscope.cns.fr/Arabidopsis	CNRGV

Stock centers distributing Arabidopsis clone repertoires:

- Arabidopsis Biological Resource Center (ABRC, USA), <http://www.biosci.ohio-state.edu/pcmb/Facilities/abrc/abrchome.htm>
- RIKEN BioResource Center (BRC, Japan), <http://www.brc.riken.jp/lab/epd/Eng/catalog/pDNA.shtml>
- GABI Primary Database (GABI/RZPD, Germany), <http://gabi.rzpd.de/>
- National Resources Centre for Plant Genomics (CNRGV, France), <http://cnrgv.toulouse.inra.fr/ENG/index.html>
- European Arabidopsis Stock Centre (NASC, United Kingdom), <http://arabidopsis.info/>
- BCCM/LMBP Plasmid and DNA library collection (BCCM/LMBP, Belgium), http://bccm.belspo.be/db/lmbp_gst_clones/
- Open Biosystems Inc., www.openbiosystems.com/

Clone collections from plant species other than Arabidopsis are also distributed via the RIKEN BioResource Center, the Rice Genome Resource Center (<http://www.rgrc.dna.affrc.go.jp/>), and the GABI primary database.