MASC cDNAs and Clone-Based Functional Proteomics (ORFeomics)- 2006 Subcommittee Report
Prepared by Pierre Hilson (Chair)

Clone-based functional genomics is gradually picking up among *Arabidopsis* scientists (reviewed in Hilson, 2006). Large-scale screens based on the systematic introduction of constructs designed for ectopic expression or silencing of *Arabidopsis* genes are underway. The first limited yeast two-hybrid protein-protein interaction matrices have been published, and several groups have reported *in vivo* analyses of subcellular location for hundreds of proteins carrying fluorescent tags. Biochemical assays and protein arrays have been described that take advantage of ORF and cDNA clone collections. But, in comparison to other eukaryotic model species, no *Arabidopsis* genome-scale dataset has yet been published in this area of research.

The steady increase in the number of cloned gene sequences is for the most part due to large-scale initiatives, but also to a few projects focusing on the systematic analysis of particular mechanisms that have generously donated their materials to stock centers (Summary Table). Importantly, the publicly available resources enable systematic approaches but are also used by scientists interested in the functional characterization of only a few genes. Material dissemination via stock centers is beneficial because it prevents duplication of efforts and promotes the use of well-documented reference materials across multiple experiments conducted by independent laboratories. Any functional genomics projects generating clone resources should be encouraged by funding bodies to donate their materials for public release or to include long term dissemination plans in their activity. Forward looking, our community must understand the need to record the results obtained with these shared clone collections using standard procedures and formats that facilitate the integration of diverse data types. When possible, these standards should be borrowed from well-established initiatives, for example BIND (Biomolecular Interaction Network Database) for protein-protein interaction.

Arabidopsis cloning efforts have been increasingly coordinated with the publication of target sequence lists at early stages of planning. Such advance notice should be implemented whenever possible. But, there is still substantial overlap between certain collections, generally resulting from different format choice or validation policies.

Today, various entities distribute *Arabidopsis* clones (ABRC, NASC, RZPD, BRC, CNRGV and individual laboratories) and there is unfortunately no straightforward procedure to interrogate at once all relevant databases for constructs of interest. A system should be implemented to display the location(s) of available clones based on queries by sequence homology searches or by AGI code names. Several groups are currently investigating how to create a unique information system for *Arabidopsis* clones that will unite dispersed databases using webservices. Ideally, such a system should be extended beyond cDNA, ORF and silencing clones, to include repertoires with promoter and other non-protein coding sequence collections, as well as recipient plasmids (*e.g.* Gateway destination vectors) designed for particular functional assays. Interestingly, the first large sets of ORF entry clones from which the native stop codon was removed – and therefore compatible with carboxyl-terminal fusions - as well as expression clones have been made available recently (Summary Table).

The *Arabidopsis* community has done an excellent job at promoting stock centers, supported by long-term funding, that curate and distribute clone collections for very low fees and with no material transfer agreements. These centers are key to the efficient use of resources and should be maintained for the benefit of all plant researchers.

Reference

Hilson, P. (2006) Cloned sequence repertoires for small- and large-scale biology. *Trends Plant Sci*. 11, 133-141.

Table 1. Publicly available *Arabidopsis* clone collections

Creator	Format	Focus	Validation	Count	URL	Stock center
ORF clones						
SSP consortium & Salk Institute	Univector pUNI51 and Gateway entry	Random	Full sequence	13,854	signal.salk.edu	ABRC
Peking-Yale Joint Center	Gateway entry	Transcription factors	5' and 3' end seq.	1,283		ABRC
TIGR	Gateway entry	Hypothetical genes	Full sequence	768	www.tigr.org/tdb/hypos/TargetGeneList.shtml	ABRC
J. Callis et al.	Gateway entry	Protein ubiquitination	Full sequence	111	plantsubq.genomics.purdue.edu	ABRC
CESG	Gateway entry [c]	Potential new fold	Full single pass seq.	~ 1,500	www.uwstructuralgenomics.org/cloning.htm	CESG
REGIA	Gateway entry	Transcription factors	5' and 3' end seq.	~ 1,000	www.gabi.rzpd.de/materials/	RZPD
Dinesh-Kumar et al.	Expression (from Peking-Yale JC)	TAP-tagged transcription factor		1,100		ABRC
Doonan et al.	Expression (from SSP)	GFP fusion for subcellular location		155		ABRC
ATOME 1 ATOME 2	Gateway entry Gateway entry, no stop	Random Random (from SSP)	5' and 3' end seq. 5' and 3' end seq.	~ 2,000 ~ 3,500	http://www.evry.inra.fr/public/projects/orfeome/orfeome.html	CNRGV
cDNA clones						
RIKEN	λ ZAP or λ PS	Random	Full sequence	16,913	http://www.brc.riken.go.jp/lab/epd/Eng/order/order.shtml	BRC
RIKEN	λ ZAP or λ PS	Random	Single pass	246,640	same	BRC
MPI-MG	Gateway expression	Random	5' end seq.	4,500	www.gabi.rzpd.de/materials/	RZPD
RNAi clones						
AGRIKOLA	Gateway entry	Random	PCR sized insert	21,903	www.agrikola.org	NASC
AGRIKOLA	hp RNA expression	Random	PCR sized insert	19,640	www.agrikola.org	NASC
CFGC	ds RNA expression	Chromatin remodel.	Single pass seq.	144	www.chromdb.org	ABRC