



# Discovering deviants in UNPAK's database of mutant phenotypes



Barnard UNPAK Team, 2015-16:

Amanda McLamb<sup>1</sup>, Lumanola Gomezdelatorre Clavel<sup>1</sup>, Amita Wanar<sup>1</sup>, Clare Kohler<sup>1,2</sup>  
<sup>1</sup>Barnard College, Columbia University; <sup>2</sup>College of Charleston; <sup>3</sup>Hampden-Sydney College

UNPAK Collaborating Investigators:

Hilary Callahan<sup>1</sup>, April Bisner<sup>2</sup>, Courtney Murren<sup>2</sup>, Matthew Rutter<sup>2</sup>  
<sup>1</sup>Barnard College, Columbia University; <sup>2</sup>College of Charleston; <sup>3</sup>Hampden-Sydney College

## ABOUT THE UNPAK PROJECT

### Goals and Central Questions:

#### Investigating Genotype-Phenotype Relationships

UNPAK is a comprehensive effort to phenotypically characterize the extensive SALK library of knockout mutants in the model plant for genomics, *Arabidopsis thaliana*, emphasizing fitness-related traits and evolutionary questions:

1. How often and which mutations result in no change, an increase, or a decrease in fitness-related traits? *This information is a significant gap in our understanding of gene and genome evolution.*
2. Do fitness-related phenotypes associate with variation in several gene attributes? *We can consider ontology, gene family size, many other attributes.*
3. Can we maximize undergraduate engagement in authentic genomics research? *Beyond involving students in plant care and data collection, we encourage them to design experiments, manage data, conceive and execute analyses, and collaborate on communicating progress.*
4. What is gained from such an open-ended, flexible and supportive framework? *Here we present a year-long exploration of novel hypotheses focusing on UNPAK's ability to identify mutations affecting traits' developmental stability.*

### Additional UNPAK questions:

#### Identifying and interpreting KO lines with unusual trait variances – "instability"

Designed features of UNPAK's screening protocols ...	... allow addressing multiple questions about the when, where and how of novel phenotypes
Replication across experiments/environments	Which mutants have unusual plasticity elicited by macro-environmental factors?
Replication within experiments	Which mutants have extreme variance elicited by micro-environmental factors?
Multiple traits	Do deviant phenotypes – whether for means or for variance – tend to be trait-specific, or in multiple traits? Which traits, if any, tend to be integrated?
Ecotypes as phytometers	How does deviation of a mutant lines compare to differentiation observed across natural ecotypes?

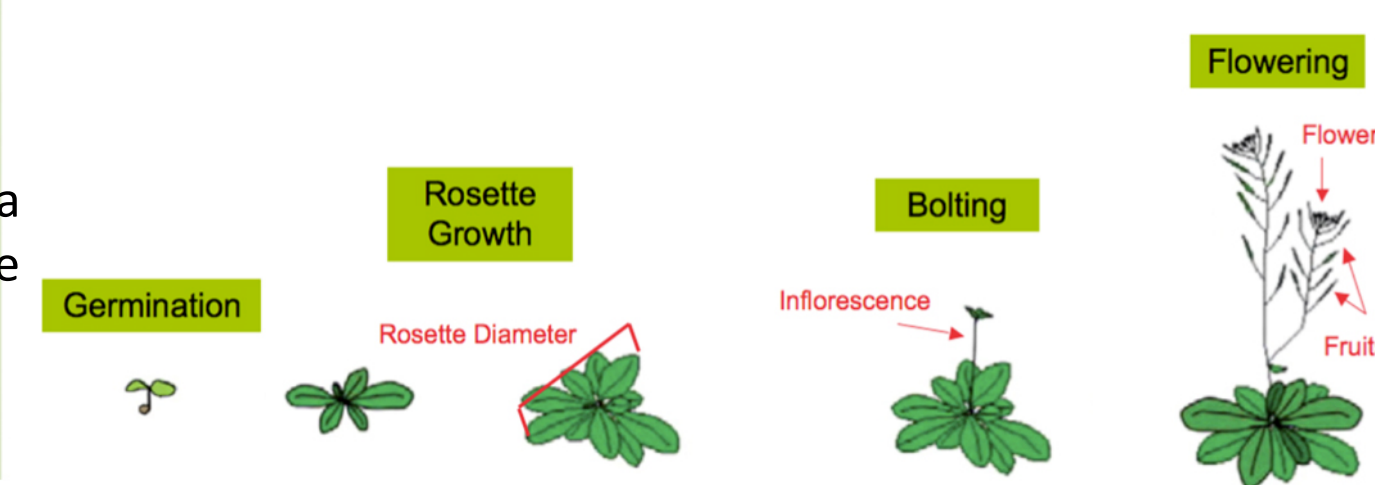
## SUMMARY OF METHODS



Photo by K. Hanslits

- Five experiments completed at Barnard from 3/2013 to 4/2016
- Stocks re-grown from ABRC seed source, to minimize maternal environment and seed age effects
- Preferentially included KO lines that UNPAK has verified as single-inserts
- 122 new KO lines (6 reps) per experiment, alongside the standard 11 ecotype lines (12 reps)
- Complete-block randomization across 12 standard flats in 72-pot inserts
- ProMixBX medium, 2 Percival AR-36L2 growth chambers at 22.5 C, 16L/8D, bottom watered regularly

A standard set of traits is scored by undergraduates who are trained via review of training videos and slide presentations, and one-on-one with a technician or senior student. Data are subject to extensive error-checking and outliers validation against sources such as dried plant samples, original data sheets, and lab notebooks. Stocks are maintained to permit follow-up studies.

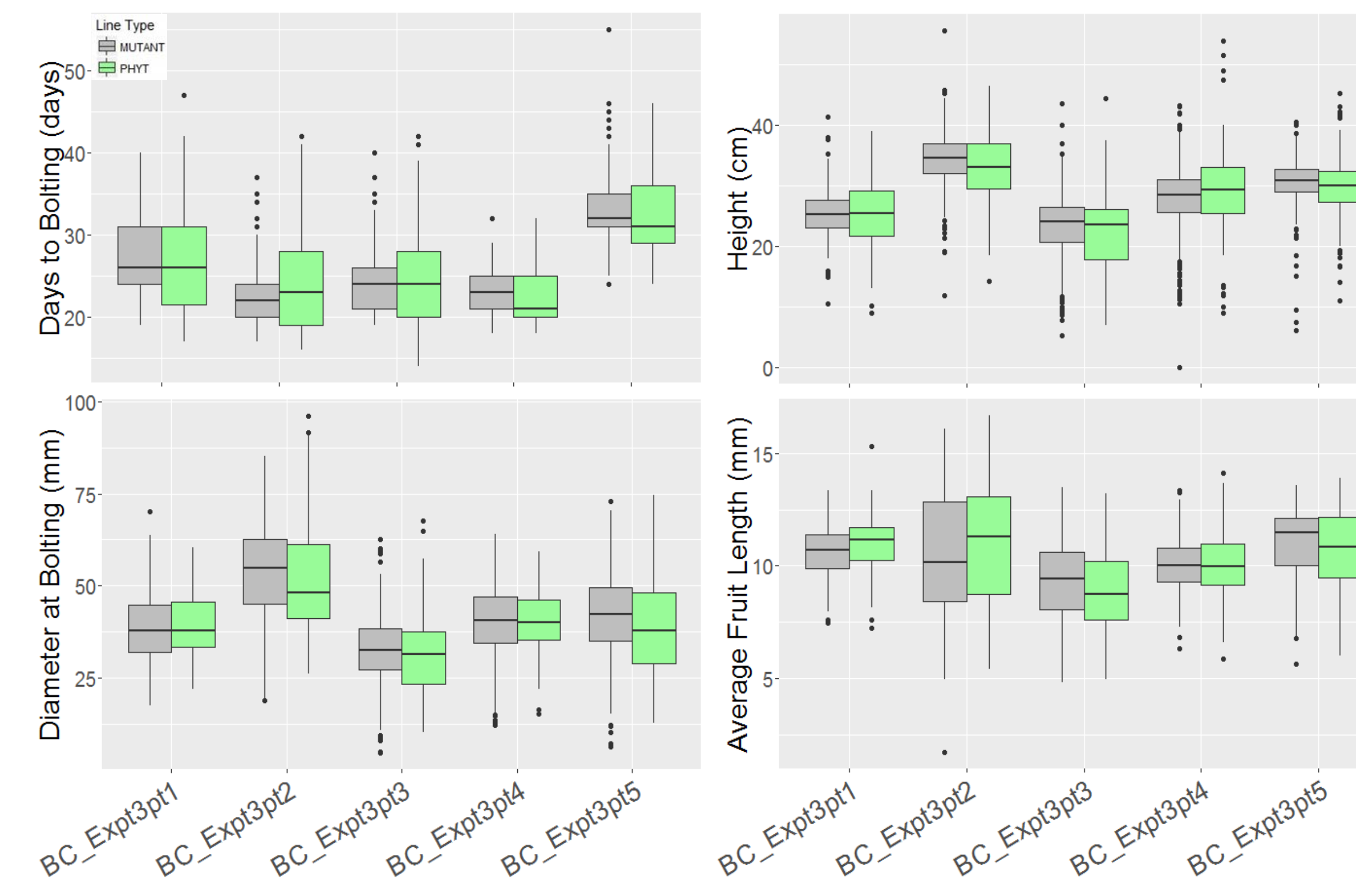


## EXPERIMENTAL DESIGN, DATA ANALYSES, RESULTS

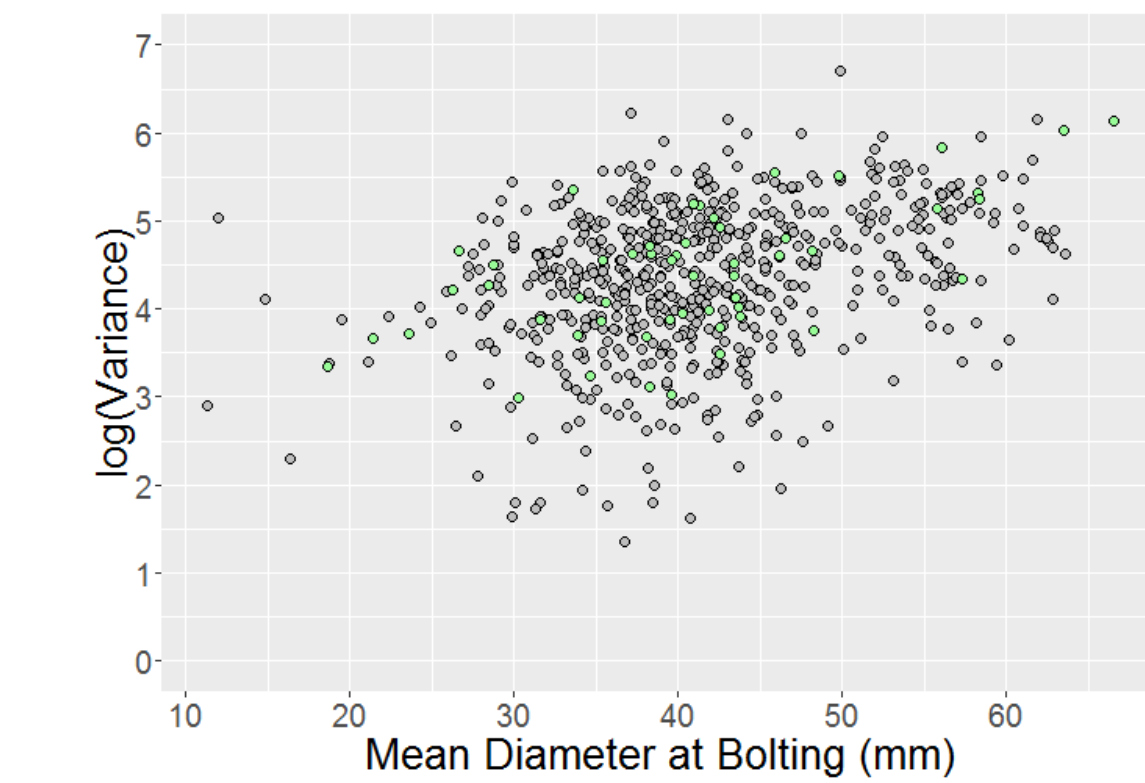
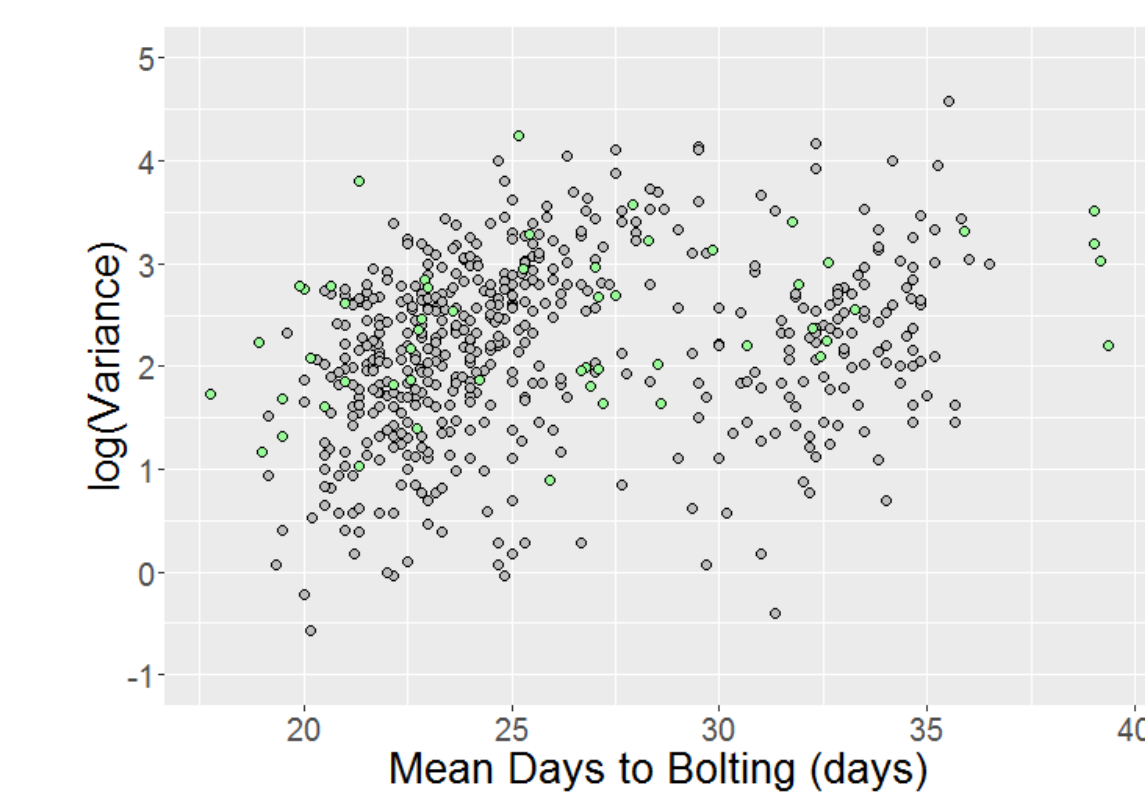
From KO mutant to phenotypic deviant: how UNPAK detects lines that have traits with an extreme mean or variance, relative to wild-type and to ecotypes

To date, five iteration of this experiment have been completed on Barnard's campus, phenotyping ~3500 individual plants for multiple traits. These plants represent ~600 SALK KO lines, all growing alongside wild-types and a standard set of ecotypes that served as "phytometers" that:

- match KO lines for trait means
- equal or exceed KO lines in trait variance
- gauge experiment-to-experiment variation
- are replicated 6-12 times per experiment, allowing estimation of line means and line variance; KO lines are similarly replicated

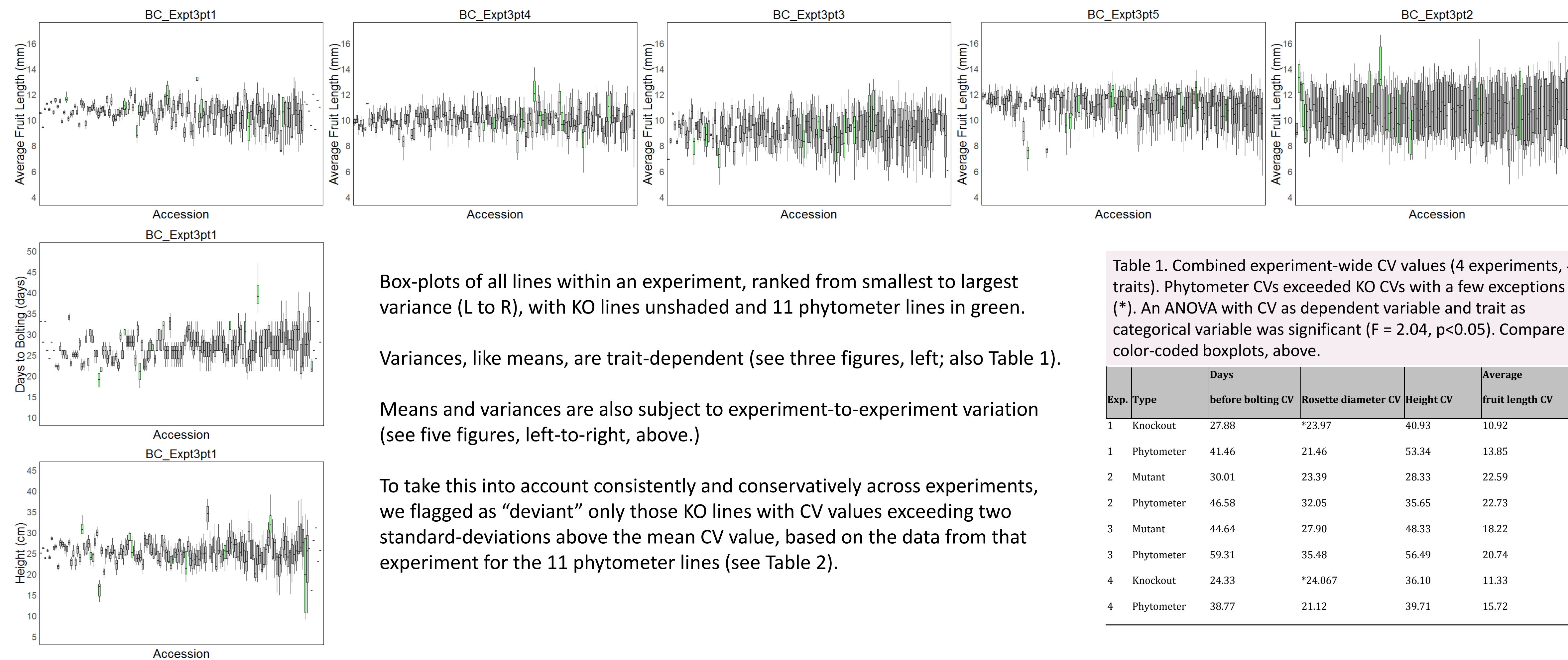


Extreme variance values are not generally correlated with extreme mean values



Variance values can be statistically problematic – non-normally distributed and scaling proportionally to mean values – making examination of these scatterplots a key step. Subsequently, as a standardized index of its "instability" or sensitivity to microenvironmental variation, we computed each KO line's Coefficients of Variation (CVs) for all traits.

UNPAK's framework accounts for line means and line variances being trait-specific and sensitive to experiment-to-experiment variation



Box-plots of all lines within an experiment, ranked from smallest to largest variance (L to R), with KO lines unshaded and 11 phytometer lines in green.

Variances, like means, are trait-dependent (see three figures, left; also Table 1).

Means and variances are also subject to experiment-to-experiment variation (see five figures, left-to-right, above.)

To take this into account consistently and conservatively across experiments, we flagged as "deviant" only those KO lines with CV values exceeding two standard-deviations above the mean CV value, based on the data from that experiment for the 11 phytometer lines (see Table 2).

Table 2. SALK KO mutants identified as potentially defective for genes regulating traits' developmental stability. Additional information obtained by cross-checking with TAIR and NCBI databases.

Accession #	CV Trait	Year	Experiment	Chromosome	Protein reported	GeneWeb Related
SALK_018462C	222064763	08	1	5	D2C/leaf/leaf family protein	Y
SALK_019364C	213054733	FA	1	2	Transmembrane protein	Y
SALK_020342C	254077863	FA	1	4	Tap1-1/booster/ankk/evanescent (GAF) like domain/COG2-like/Transmembrane protein	Y
SALK_020353C	270491802	R	1	1	leucylamylase/acetone 1	Y
SALK_020881C	215042426	FA	1	2	ethylen-activated protein kinase 17	Y
SALK_020882C	093004538	R	1	2	ethylen-activated protein kinase 17	Y
SALK_020883C	064010148	R	1	2	ethylen-activated protein kinase 17	Y
SALK_020884C	501003291	R	1	3	WFD5-like 1	Y
SALK_021774C	010006885	R	1	5	D-arylindoleamine (DAI) like family protein	Y
SALK_021984C	373504837	R	1	3	Ornithinecarbamoyl transferase/leaf family protein	Y
SALK_130711C	265203951	FA	1	1	RNA3/CPY/PRD class: Regor repeat/serpin protein	Y
SALK_151571C	383702786	FA	2	4	high alkylglycerol biosynthese 1/3 protein	Y
SALK_027777C	045428205	R	2	4	cellulose synthase protein 1	Y
SALK_044642C	020423269	R	2	5	Leaf-ectodermogenin-like/leucylamylase (LEAL) protein-related	Y
SALK_045947C	084508634	FA	2	5	non-classical PGL-like protein, putative	Y
SALK_045977C	067741989	FA	2	3	alpha-beta/tryptophan repeat family protein	Y
SALK_047979C	064072372	R	2	4	COP1-interacting protein	Y
SALK_050667C	184427423	FA	2	4	flavonoid-3-O-glucosyltransferase 1	Y
SALK_050668C	395387814	R	2	1	SRP22/serpin-like protein	Y
SALK_051011C	013308870	08	2	1	Hydroxyacid dehydratase regulator 1/serpin-like protein	Y
SALK_051047C	161512687	R	2	4	SRP-ABC domain-containing domain-receptor protein	Y
SALK_051052C	074737929	R	2	1	GLYCOXYLASE/RELATIVITY1	Y
SALK_051053C	265227634	08	2	3	Protein with RNA3-like and TIRAP-like domains	Y
SALK_051054C	143481401	08	2	4	Thrombospondin family protein	Y
SALK_051055C	316805892	FA	2	3	proteinase-like protein, putative	Y
SALK_051056C	094883326	R	2	5	unclassified P4H1 domain protein, putative	Y
SALK_051057C	061498768	R	1	2	SRP-24/serpin-like protein	Y
SALK_088702C	012082944	R	3	2	SRP-24/serpin-like protein	Y
SALK_100114C	042220944	R	3	1	serpinase-2	Y

Trait: BD (Days to bolting), R (Rosette diameter), FA (Average Fruit Length), H (Height), Growth-Related Expression: Y (Yes), N (No)

## ACKNOWLEDGEMENTS

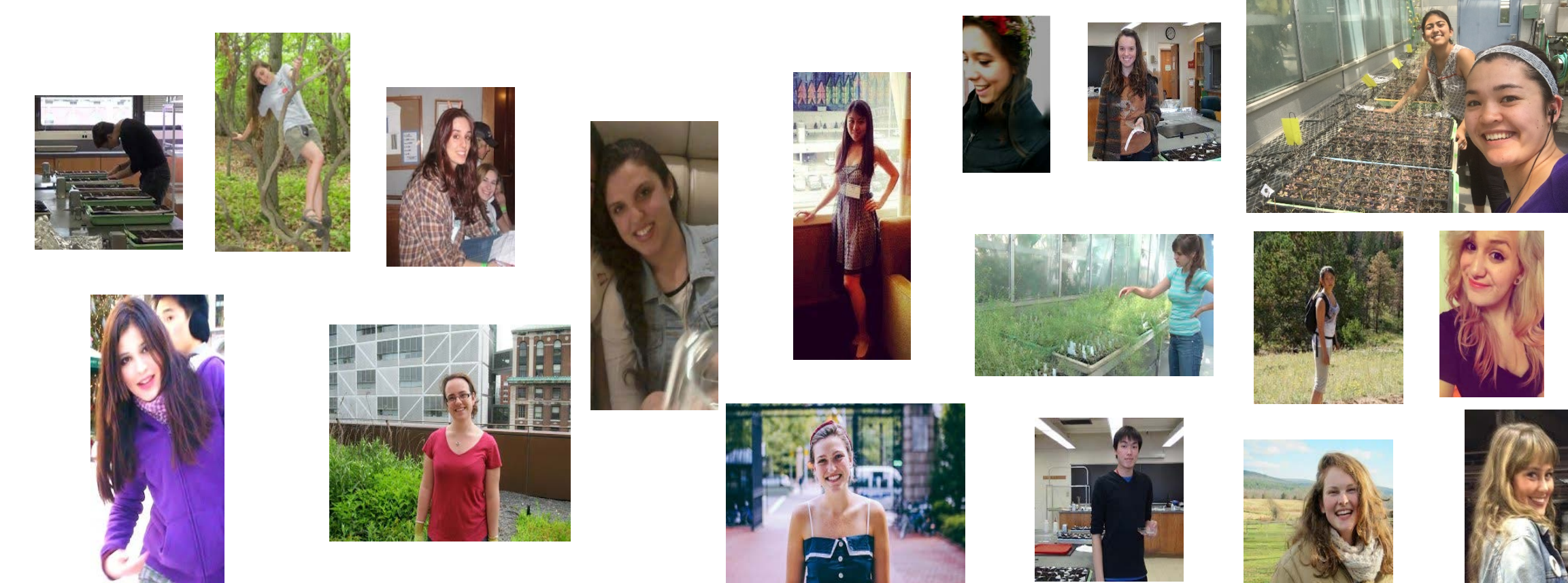
<http://arabidopsisunpak.org/>



### UNPAK: Plant genomics on the shoulders of undergraduates !!

More than 25 students from Barnard and CUNY's LaGuardia Community College have contributed to UNPAK's progress, including D. Cassidy, K. Cronin, L. Fletcher, L. Flynn, L. Gomezdelatorre Clavel, K. Hanslits, S. Kern, K. Lake, K. Lam, A. McLamb, J. Marrs, J. McCorkle, R. Nagpal, C. Nagy, M. Orozco, M. Pantalena, H. Peng, A. Sandoval, S. Schaeffer, M. Schwarz, J. Tan, N. Thulin, V. Varone, J. Wan, A. Wanar, G. Winship.

UNPAK's work has been supported by grants from the NSF's DOB-10S Program, #13-54603 and #10-52323, a Barnard College Presidential Research Award, and Barnard College's Hughes Science Pipeline Project, funded by the Howard Hughes Medical Institute.



Come to Matt Rutter's talk, Tues @ 9:00 AM, Session 6B

Lessons from 200,000 *Arabidopsis* phenotypes: the power of UNPAK