One genotype, many phenotypes: Testing hypothesized variability mutants in *Arabidopsis thaliana*

Rhea Nagpal, Jerica Tan, Hilary Callahan, Clare Kohler, Lu Gomezdelatorre, Amanda McLamb, Amita Wanar

Callahan Laboratory, Department of Biology, Barnard College

Introduction

- *Arabidopsis thaliana* is a genomic model in the study of phenotypic variability and applied plant sciences
- The SALK Institute Genome Analysis Laboratory has a library of over 300,000 T-DNA insertion knockout mutants (covering about 67% of all genes)
- The **unPAK** (Undergraduate Phenotyping of Arabidopsis Knockouts) project studies these mutants, investigating:
  - Fitness-related traits such as timing of reproduction and number of fruits
  - How often and which mutations result in mutant phenotypes
  - The role of the environment in shaping phenotypes
- In this experiment, we re-examine SALK mutants with known **unusual phenotypic variability**.
  - For a single trait, replicated individuals of a given genotype may display variable phenotypes, despite having identical genotypes and experiencing the same macro-environment.
- Phenotypic variability is an essential and heritable feature of organisms
  - Genes and environment jointly influence the scope of variability
  - Variability itself may be subject to natural selection

Methods

- We examine mutants with unusual variability for the **timing of bolting** and **fruit length**.
- In all unPAK experiments, SALK mutants are grown alongside and compared to wild-types and phytometers, different ecotypes of Arabidopsis designed to control for environmental variation between and within experiments.
- We quantify mutants' variability for traits using the coefficient of variation (CV) of each trait, comparing with CVs of wild-types and phytometers.
- **Increased replication** is being used to improve precision of CV estimation for these unstable mutants (CV > ± 2 standard deviations of wild-type CVs) and improve comparison with a set of wild phytometers and mutants with normal CVs.
  - 18 replicates of 20 mutant genotypes were randomly assigned to locations in a 6-flat experiment. Each flat has have 3 replicates of each mutant alongside 12 phytometers.

Selecting Experimental Genotypes

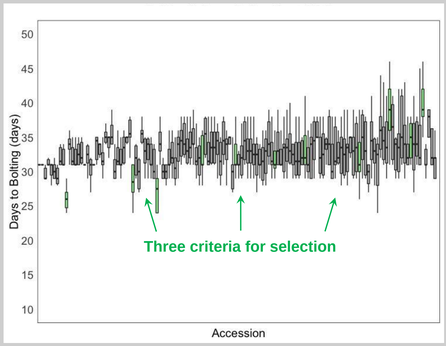

Figure 1a: Box plot of all lines in Experiment 3- Part 5 ranked from smallest to largest variance (L to R) of days to bolting, with KO lines unshaded and 11 phytometer lines in green.

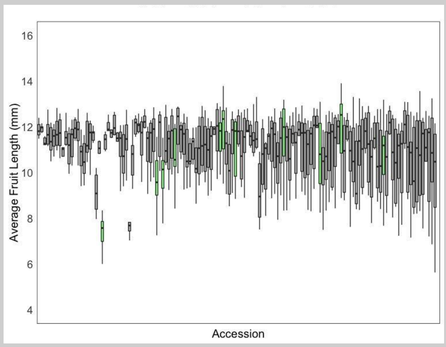

Figure 1b: Box plot of all lines in Experiment 3- Part 5 ranked from smallest to largest variance (L to R) of average fruit length, with KO lines unshaded and 11 phytometer lines in green.

Mutants with low, medium, and high variability were selected for an additional experiment with increased replication.

Accession	Trait	Variability	CV	Locus	Function
SALK_001197C	Bolt Days	High	21.68	AT5G13840	Protein expressed in heterotrimeric G-protein complex and nucleus; functions in in protein binding, has signal transducter activity
				AT5G13850	Protein expressed in Golgi apparatus, cytosol, cytosolic ribosome, nucleus, and plasmodesma; involved in response to salt stress
				AT5G13845	Involved in translational elongation; has triplet codon-amino acid adaptor activity
SALK_020283C	Bolt Days	High	23.02	AT5G18960	Protein expressed in nucleus; involved in response to red or far red light
SALK_061516C	Bolt Days	High	20.64	AT4G30890	Protein expressed in nucleus; has thiol-dependent ubiquitin-specific protease activity
				AT4G30900	Protein expressed in nucleus and plasma membrane
SALK_003448C	Bolt Days	Low	4.44	AT3G27000	Protein expressed in Arp2/3 protein complex and cytoplasm; involved in actin filament-based processes, cell morphogenesis, multidimensional cell growth, trichome; functions as structural constituent of cytoskeleton
SALK_062440C	Bolt Days	Low	4.47	AT1G71860	Protein expressed in cytoplasm, intracellular, nucleus, and plasma membrane; has protein tyrosine phosphatase activity
SALK_066558C	Bolt Days	Low	4.19	AT5G11180	Protein expressed in extracellular region; involved in cellular calcium ion homeostasis, response to light stimulus; has intracellular ligand-gated ion channel activity
SALK_020993C	Bolt Days	Average	15.86	AT1G18740	Protein expressed in chloroplasts
SALK_056546C	Bolt Days	Average	15.95	AT4G33680	Protein expressed in chloroplasts; involved in lysine biosynthetic process via diaminopimelate, response to cytokinin, systemic acquired resistance, salicylic acid mediated signaling pathway; functions in copper ion binding
				AT4G33690	Protein expressed in chloroplasts
SALK_013974C	Fruit Length	High	25.97	AT3G03160	Protein expressed in mitochrondria
SALK_019272C	Fruit Length	High	25.48	AT4G28950	Protein expressed in extracellular region, plasma membrane, functions in GTP binding
SALK_019794C	Fruit Length	High	26.19	AT2G45960	Protein expressed in plasma membrane, chloroplasts, plasmodesma, and vacuole; involved in carbon dioxide transport, cellular water homeostasis, ion transmembrane transport, response to salt stress
SALK_020296C	Fruit Length	High	25.6	AT2G13370	Protein expressed in nucleus and plasmodesmata; involved in chromatin assembly or disassembly, ATP binding, and DNA binding; has helicase activity
SALK_010581C	Fruit Length	Low	5.42	AT1G28790	Protein involved in translation, translational elongation; has triplet codon-amino acid adaptor activity
				AT1G28780	Protein involved in Involved in translation, translational elongation; has triplet codon-amino acid adaptor activity
SALK_020686C	Fruit Length	Low	4.74	AT3G52110	Protein expressed in nucleus
SALK_065091C	Fruit Length	Low	4.63	AT5G58710	Protein expressed in in chloroplast, endoplasmic reticulum, plasmodesma; involved in protein folding, root development
SALK_031510C	Fruit Length	Average	21.44	AT5G23920	Protein expressed in in mitochondrion, plasma membrane, and vacuole

* SALK_011196C, SALK_056651C, SALK_011196C, and SALK_053563C are not included as their chromosomal loci and functions have not yet been identified.

Table 1: SALK KO mutants used in the experiment and their respective trait-specific level of variability, CV, locus, and function at that locus (from The Arabidopsis Information Resource, TAIR).

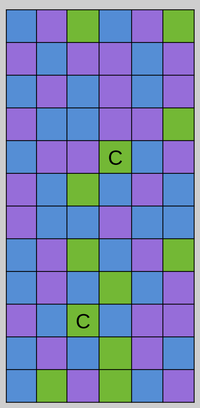

Figure 2: Sample randomization scheme of a single flat. Blue squares represent "fruit length" KOs, purple represent "days to bolting" KOs, and green represent phytometers (**C** represents Columbia wildtypes).

Fruit Length
- Larger fruits indicates more seeds and therefore greater reproductive fitness.

Plant Bolting
- Earlier bolting day indicates greater fitness. Control of bolting date has agricultural implications, as farmers require the ability to adjust depending on georgrahical region and growing season.

Take-homes and Next Steps

The ways in which genes and the environment interact in creating variation are an area of active research with model organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*, making our analyses and experiments early steps in developing Arabidopsis as a research system in this field. Part of unPAK involves addressing the phenotypic effects of macroenvironmental factors (nutrient availability, water stress). Our research expands this focus to how more subtle microenvironmental factors (such as light differences between pots on the same flat) contribute to phenotypic variability.

Although microenvironmental variation cannot be completely eliminated, we can distinguish variability with solely environmental sources from variability with heritable sources and associated candidate genes. We can then further characterize these genes. For example, genes coding for heat-shock proteins or proteins involved in proteostasis could result in increased variability if knocked out. In contrast, knocking out genes involved in amplifying environmental signals could result in less variable phenotypes. UnPAK's phenotyping strategy complements other methods such as genome-wide association studies that have also identified to such genes in other model systems.

Acknowledgements

We would like to thank our collaborators at twelve institutions, especially the College of Charleston, as well as the National Science Foundation for funding unPAK.