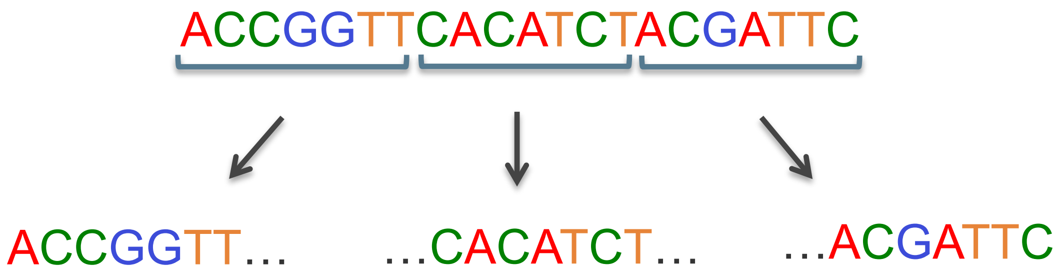
JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

Resequencing at the JGI

Nicole Johnson
JGI User Meeting
March 23, 2015

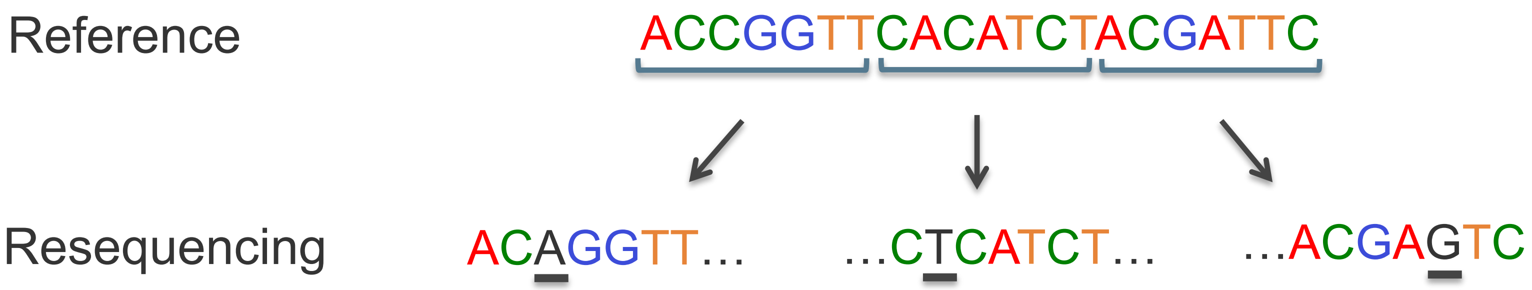
What is resequencing?



Reference ACCGGTTCACATCTACGATTC

Resequencing ACCGGTT… …CACATCT… …ACGATTC

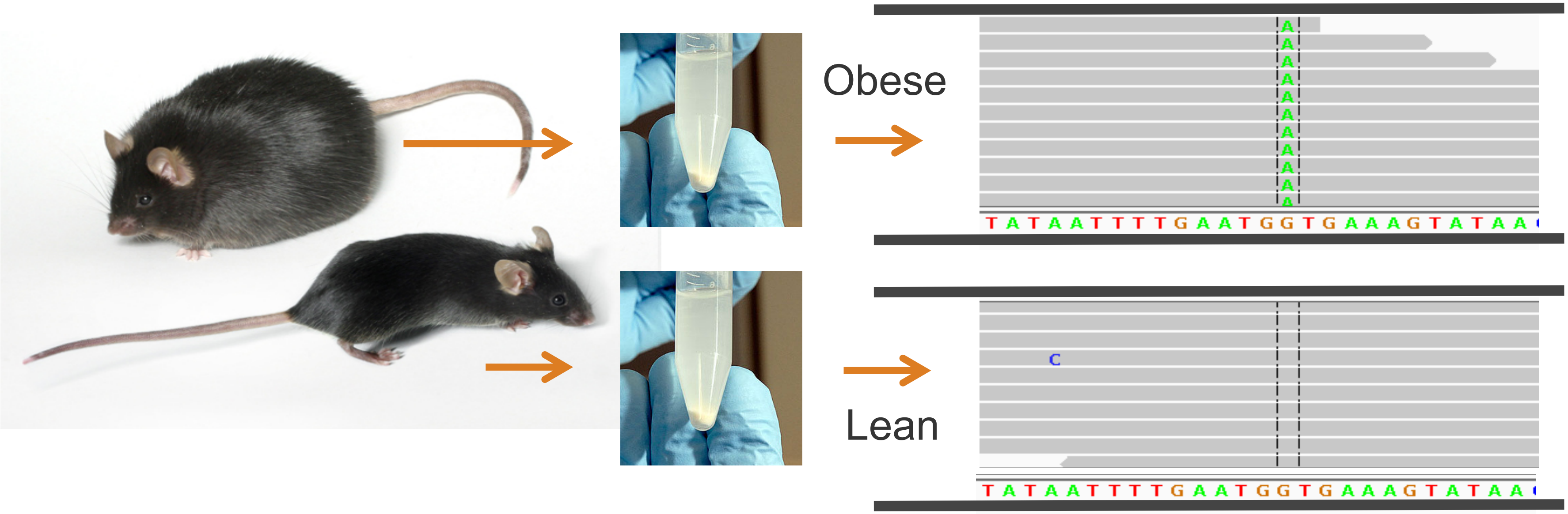
What is resequencing?



Reference ACCGGTTCACATCTACGATTC

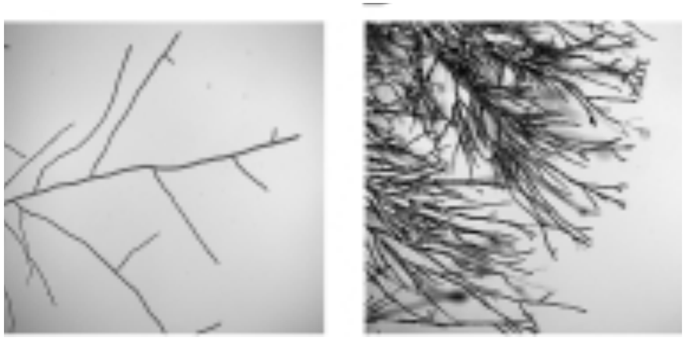
Resequencing ACAGGTT… …CTCATCT… …ACGAGTC

genotype -> phenotype



Resequencing projects at the JGI

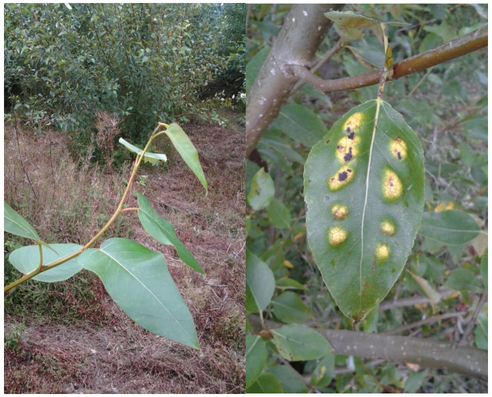
Causative mutations contributing to unique phenotypes



Wild type Tangerine

Neurospora crassa classical mutants

Genetic diversity relative to disease



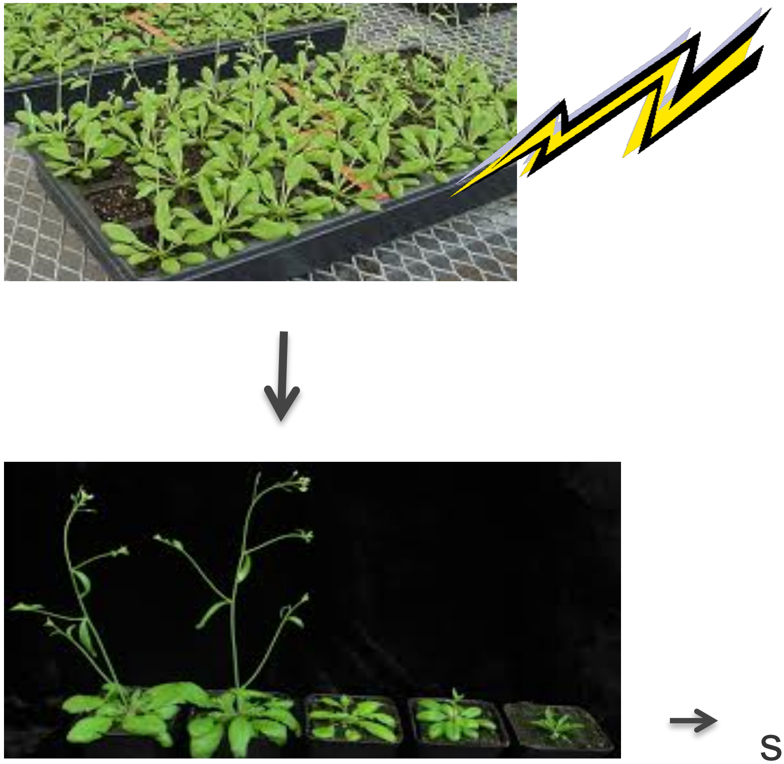
Wild poplar

Impact of mutations on biological processes

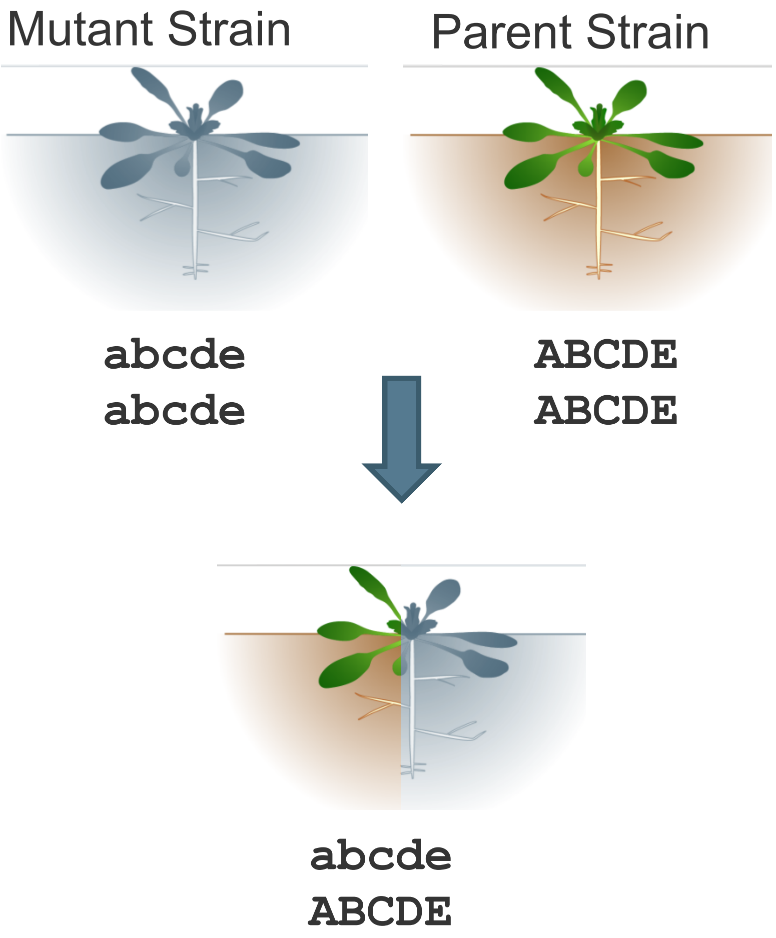
Arabidopsis thaliana

Resequencing of *A. thaliana*

Goal: To better understand the composition of plant cell walls to uncover the utility of Arabidopsis plants as a biofuel source



Mutant selection via phenotype



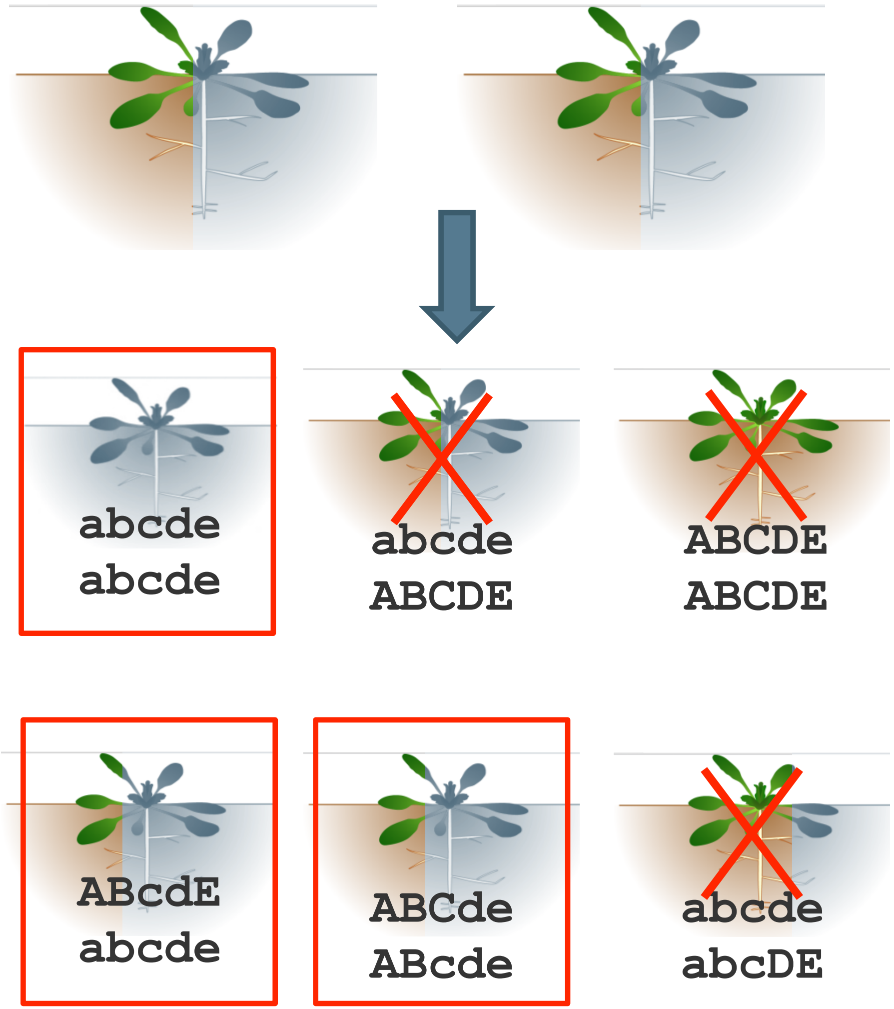
Mutant Strain

abcde
abcde

Parent Strain

ABCDE
ABCDE

abcde
ABCDE



abcde
abcde

abcde
ABCDE

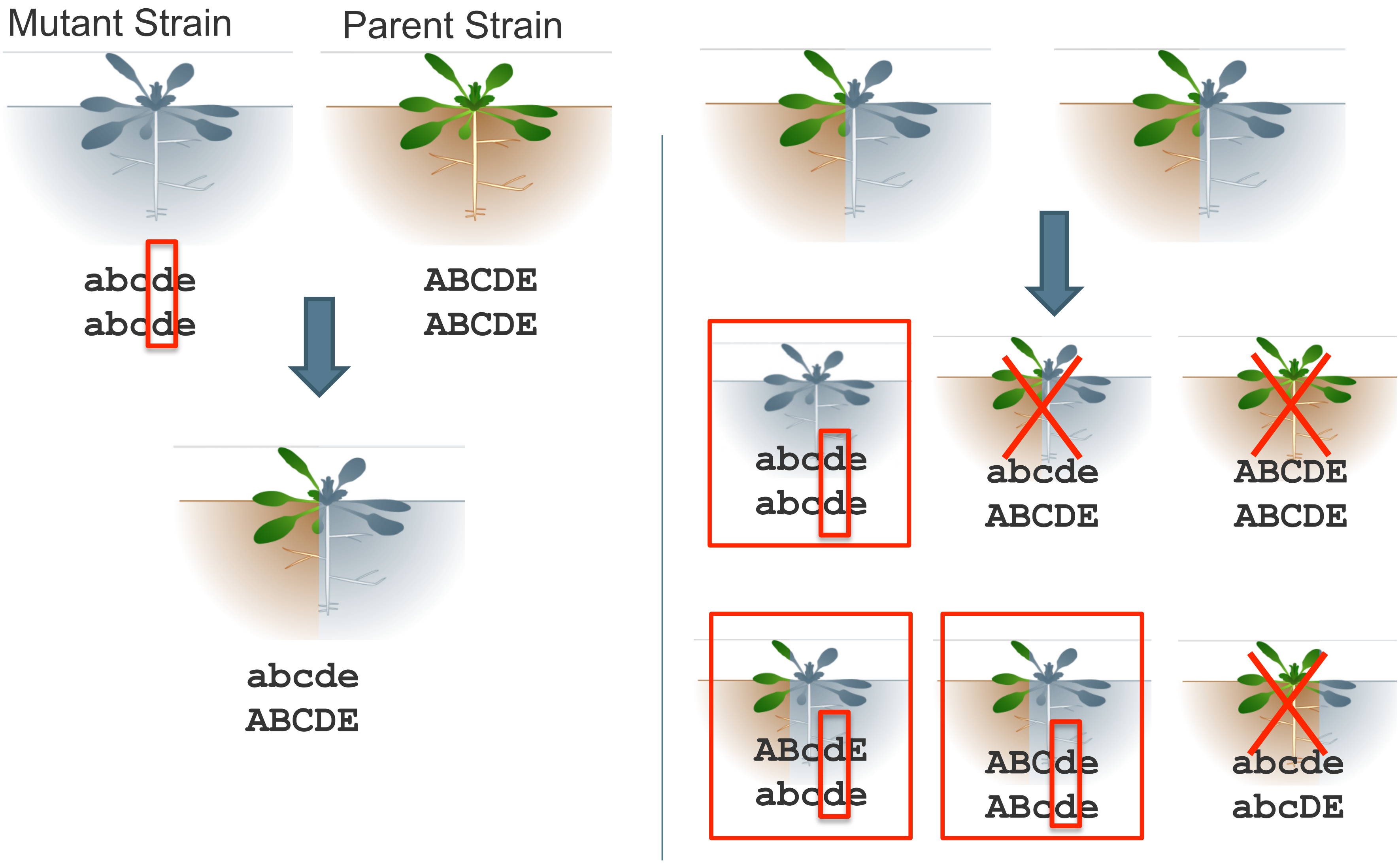
ABCDE
ABCDE

ABcdE
abcde

ABCde
ABcde

abcde
abcDE

Mutant phenotype reflected in the genotype



Mutant Strain

abcde
abcde

Parent Strain

ABCDE
ABCDE

abcde
ABCDE

abcde
abcde

abcde
ABCDE

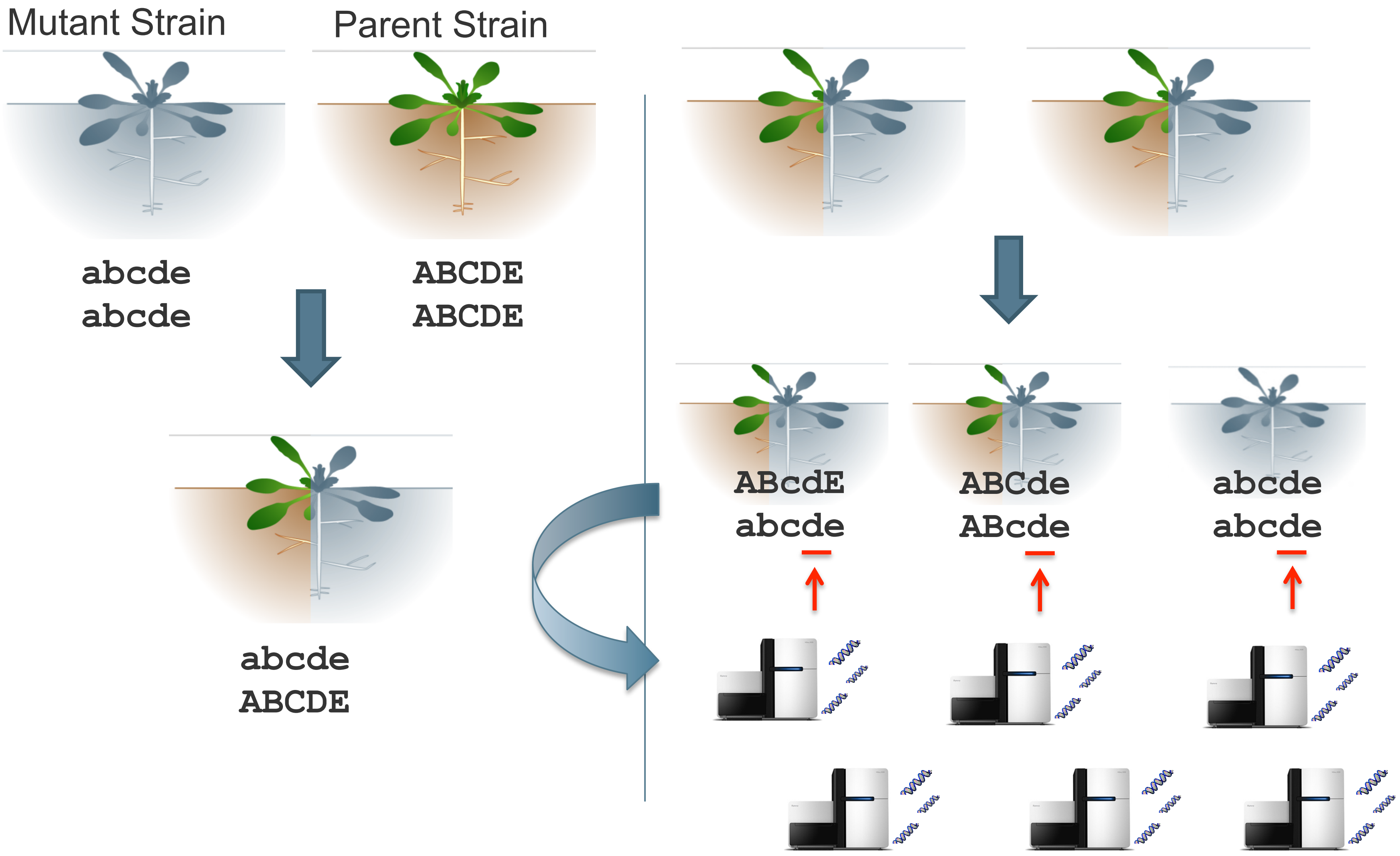
ABCDE
ABCDE

ABcdE
abcde

ABCde
ABcde

abcde
abcDE

Resequencing of mutant progeny



Mutant Strain

abcde
abcde

Parent Strain

ABCDE
ABCDE

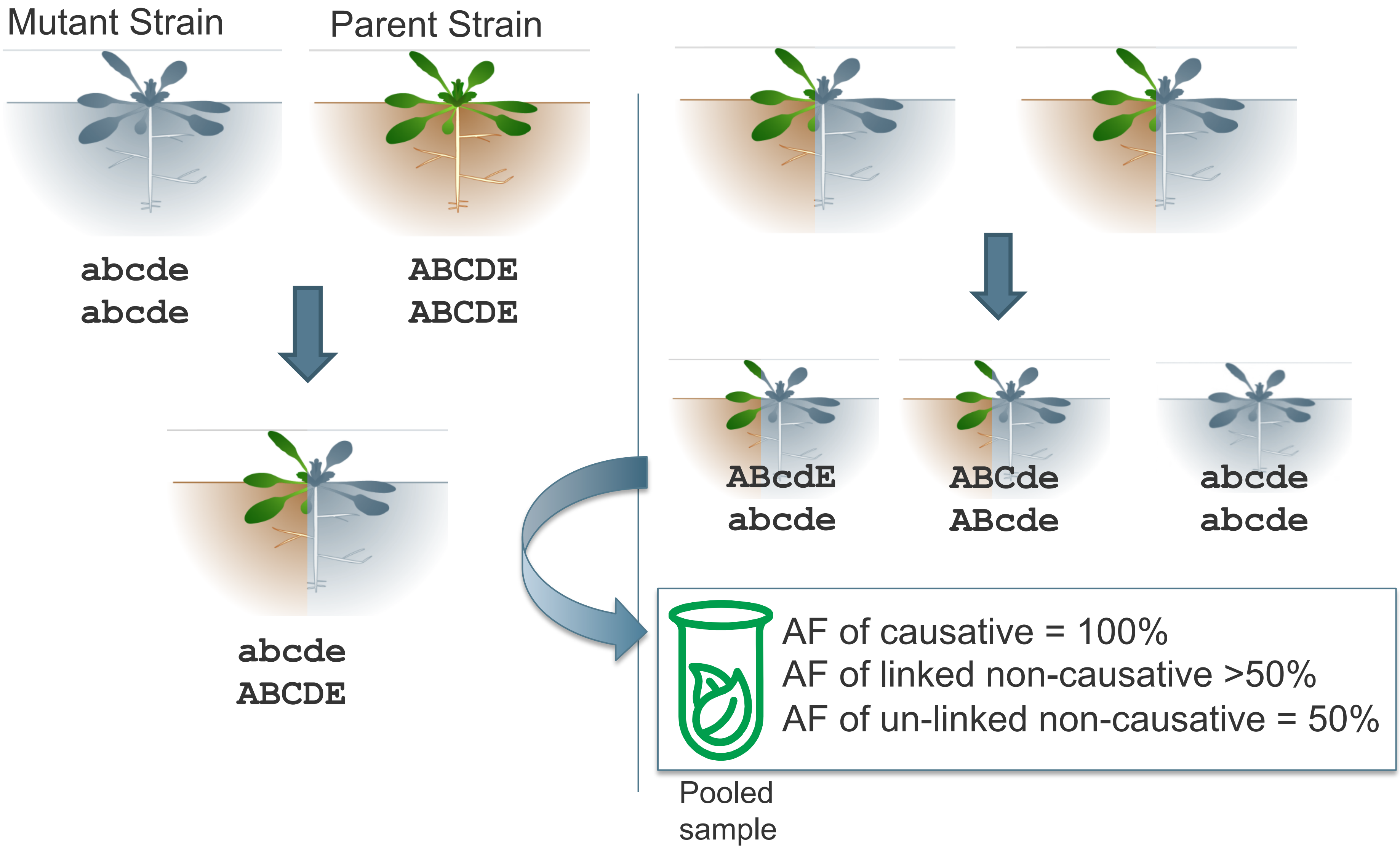
abcde
ABCDE

ABcdE
abcde

ABCde
ABcde

abcde
abcde

Resequencing of mutant progeny



AF of causative = 100%
AF of linked non-causative >50%
AF of un-linked non-causative = 50%

Identification of causative mutations

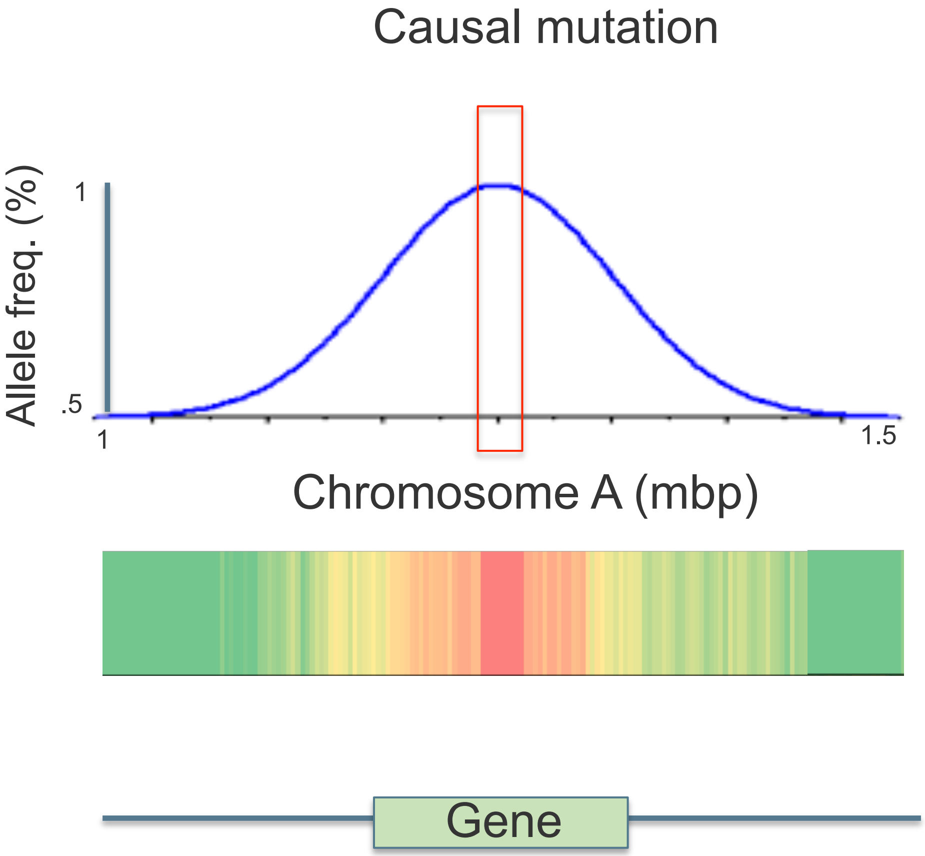
Pooled mutant
DNA (leaf)

1,000’s of mutations

How do we find the needle in the haystack?

Identification of causative mutations

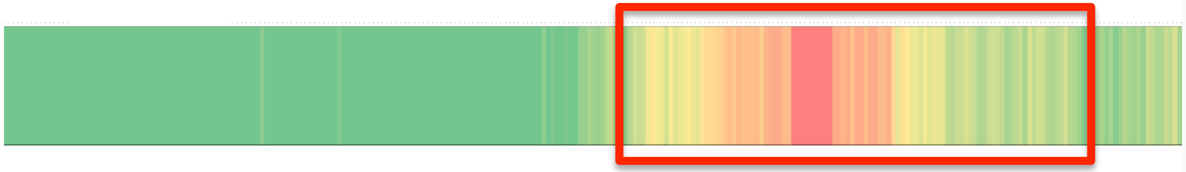
Sample pool



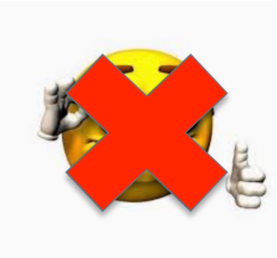
First attempt at finding the mutation…

Chromosome A

Sample Pool 1

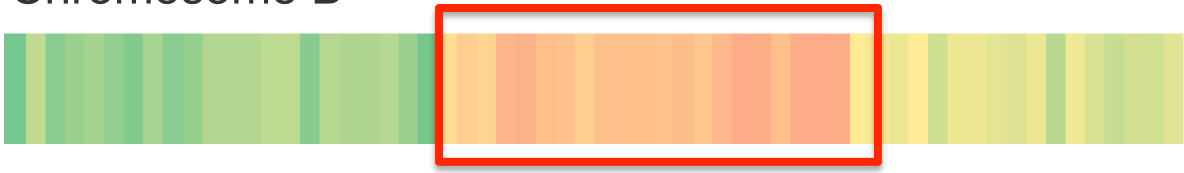


Max AF = ~100%



Chromosome B

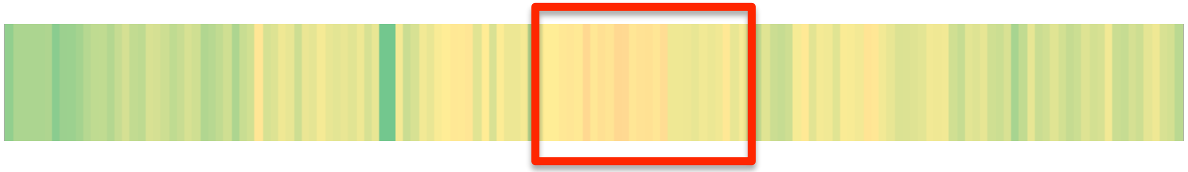
Sample Pool 2



Max AF = ~90%

Chromosome C

Sample Pool 3



Max AF = ~80%

Problem solving…

JGI?

Analysis?

Group lead

Program broken?

Analyst

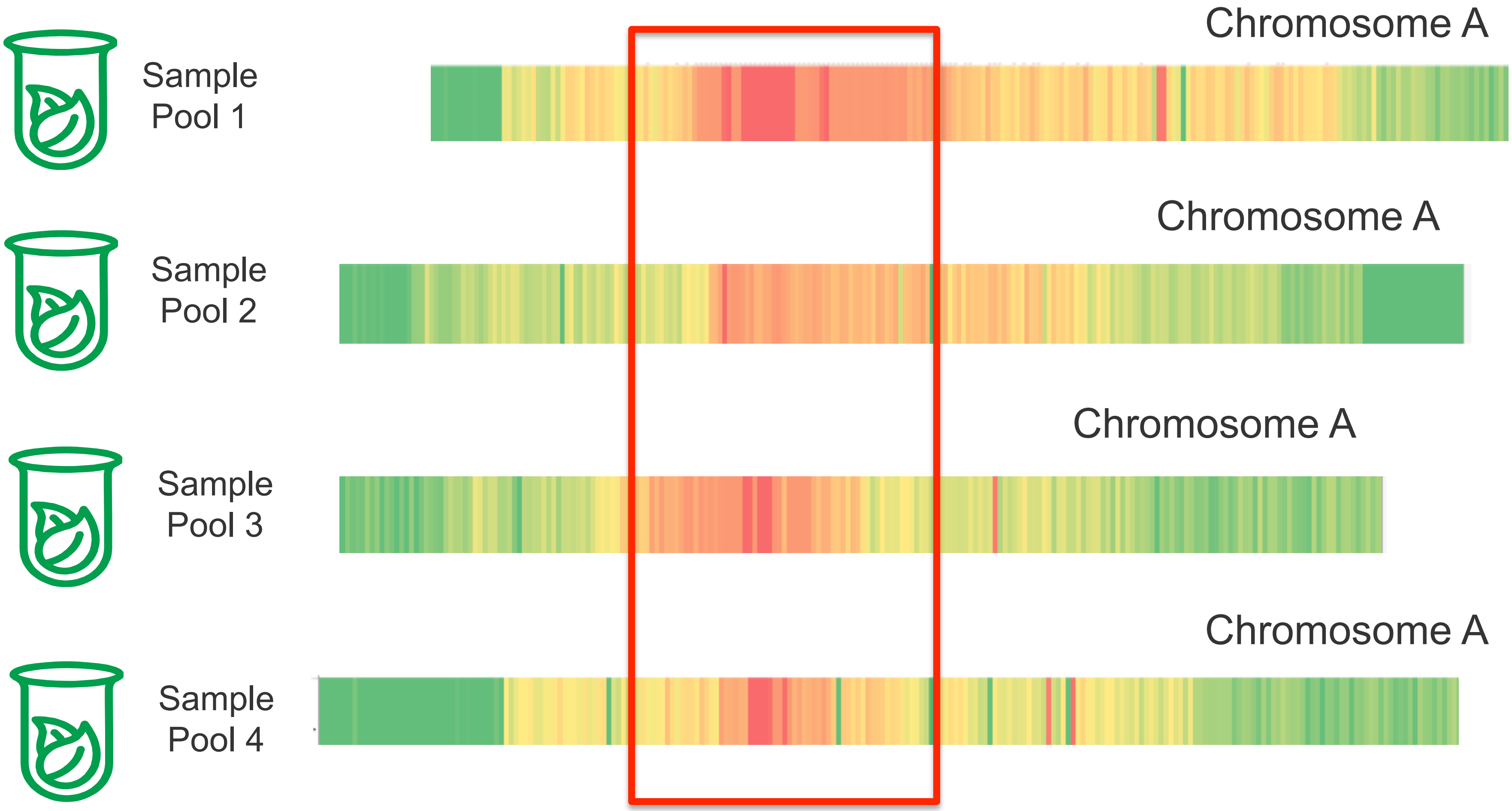
We did have trouble phenotyping some of the strains

PI

Post-doc

Graduate student

Second attempt to find the mutation…



Sample Pool 1
Chromosome A

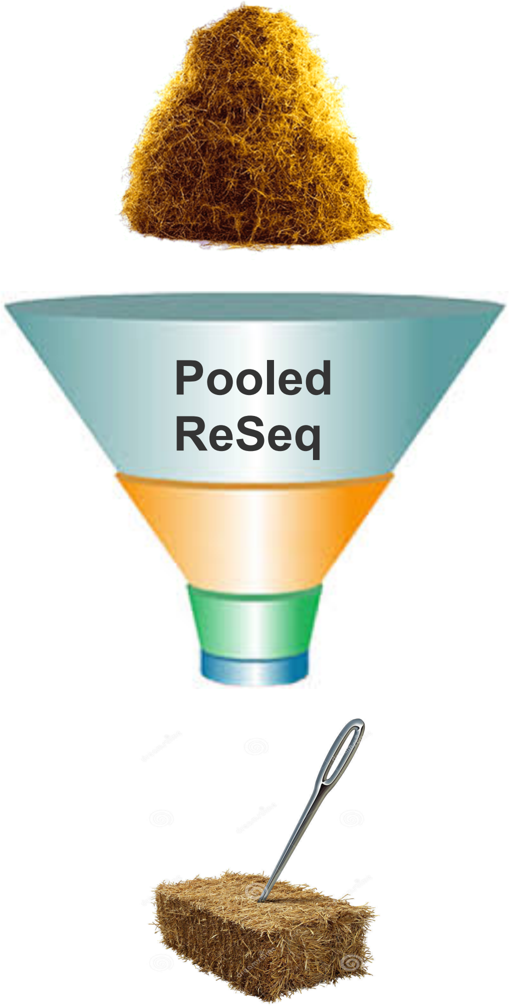
Sample Pool 2
Chromosome A

Sample Pool 3
Chromosome A

Sample Pool 4
Chromosome A

3 out of 4 causative mutations hit the same gene!

Summary



- Pooled resequencing is a time and cost effective way to detect causative variants
- Even with a small number of sample pools, it is possible to detect causative mutations
- Given the data, it can confirm or refute the phenotyping of the individuals

Acknowledgements

- **JGI ReSeq team**
  - Wendy Schackwitz
  - Joel Martin
  - Anna Lipzen
- **Collaborators**
  - Henrik Scheller lab

- **Want to discuss your existing or future data?**
  wsschackwitz@lbl.gov
  nicolevjohnson@lbl.gov

- **Or come find us!**