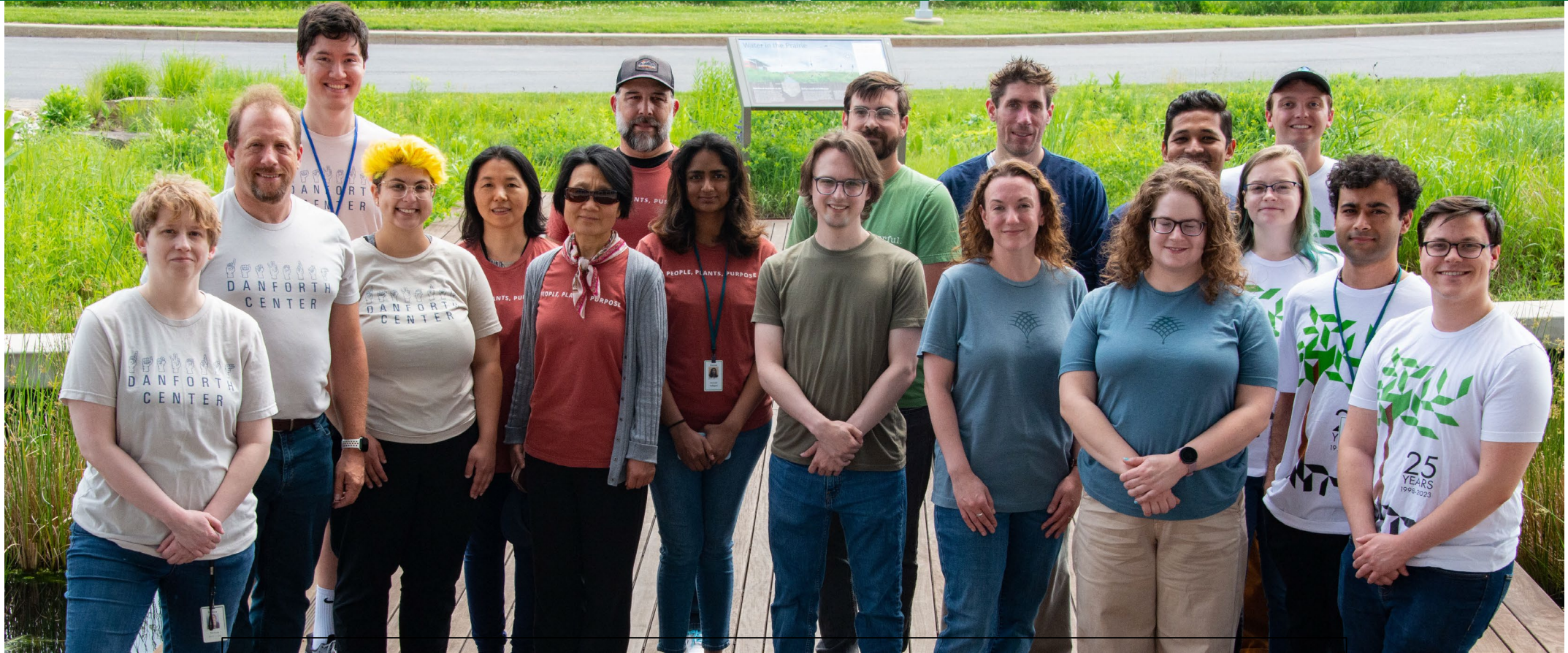


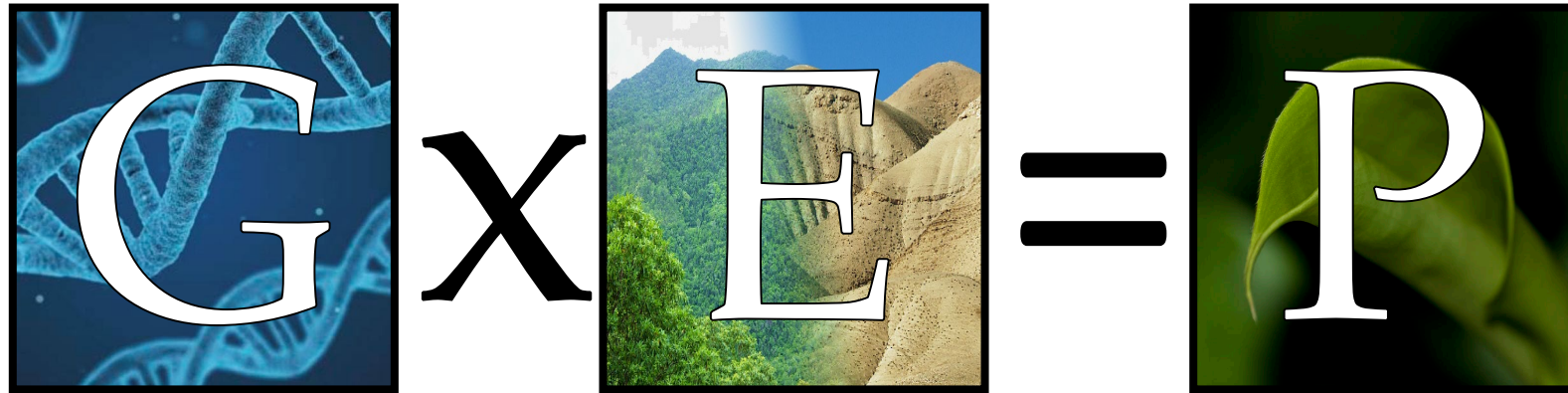
A Comparative Approach for Selecting Orthologous Candidate Genes Underlying Signal in Genome-Wide Association Studies across Multiple Species



Baxter Lab

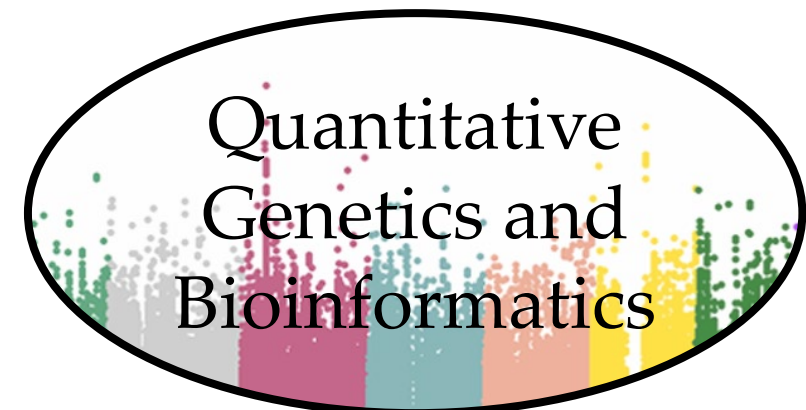
Donald Danforth Plant Science Center

The Baxter Lab



Genomics

Cr	25	26	27	28	29	30	31	32	33	34	Br
Chromium	Manganese	Iron	Cobalt	Nickel	Copper	Zinc	Ga	Ge	As	Selenium	Bromine
42	43	44	45	46	47	48	49	50	51	52	53
Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I
Molybdenum	Technetium	Ruthenium	Rhodium	Palladium	Silver	Cadmium	Indium	Tin	Antimony	Tellurium	Iodine
74	75	76	77	78	79	80	81	82	83	84	85
W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At
Tungsten	Rhenium	Osmium	Iridium	Platinum	Gold	Mercury	Thallium	Lead	Bismuth	Polonium	Astatine
107	108	109	110	111	112	113	114	115	116		
Lr	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv		
Lanthanum	Hassium	Moscovium	Darmstadtium	Roggenbergium	Copernicium	Nihonium	Flerovium	Moscovium	Livermorium		



The Ionome: What Is It and Why Do We Care?

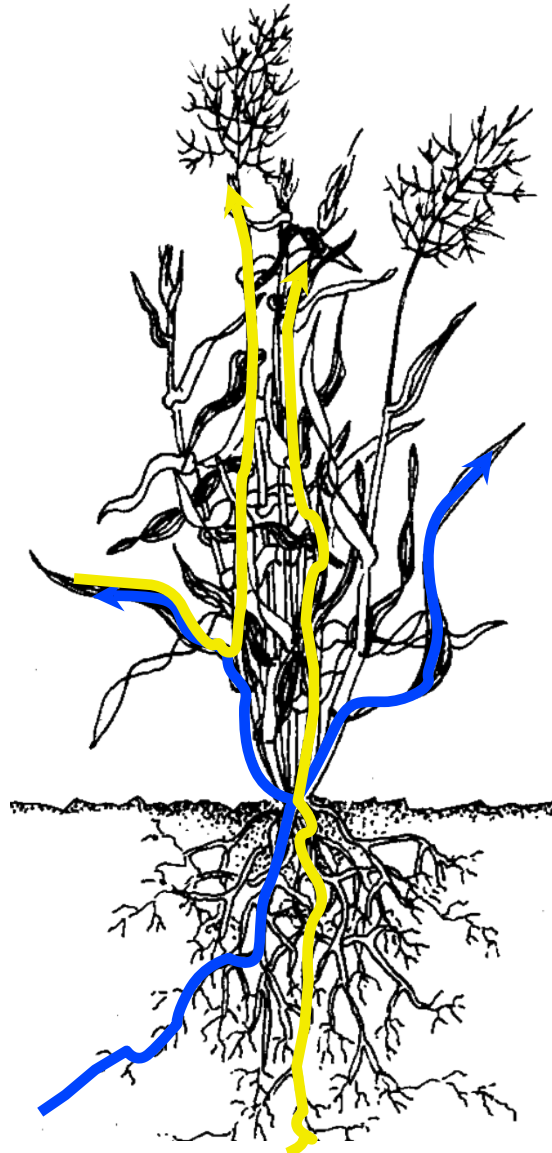
H ¹					
Li	Be ⁴				
Na	Mg				
K	Ca	Sc ²¹	Ti ²²	V	
Rb	Sr	Y ³⁹	Zr ⁴⁰	Ni	
Cs ⁵⁵	Ba ⁵⁶	La ⁵⁷	Hf ⁷²	Ta	
Fr ⁸⁷	Ra ⁸⁸	Ac ⁸⁹			



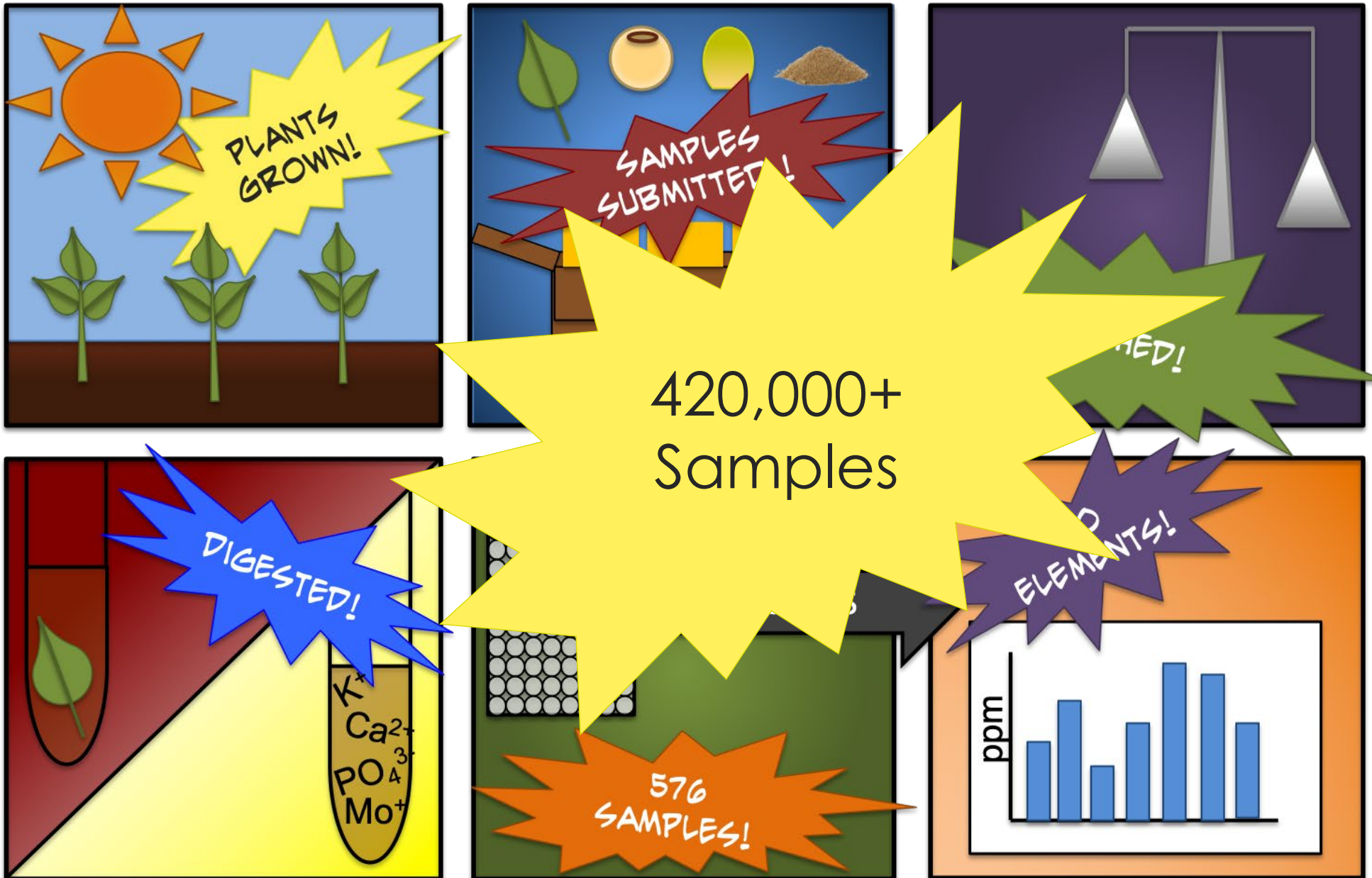
				He ²
6	N ⁷	O ⁸	F ⁹	Ne ¹⁰
14	P	S	Cl ¹⁷	Ar ¹⁸
32	As	Se	Br ³⁵	Kr ³⁶
50	Sb ⁵¹	Te ⁵²	I ⁵³	Xe ⁵⁴
82	Bi ⁸³	Po ⁸⁴	At ⁸⁵	Rn ⁸⁶

Ionome: “all” the inorganic ions within a cell, tissue, or organism

Elemental Accumulation Is Determined by Plant/Environment Interactions



The Ionomics Pipeline



Seed ionome GWAS datasets are available for a many species.

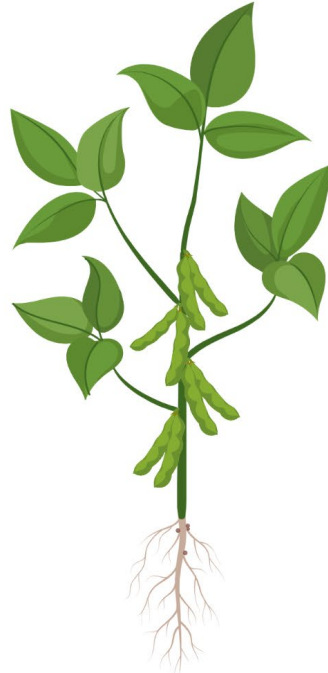
Arabidopsis thaliana



Single

1404

Glycine max
(soybean)



Multiple

7308

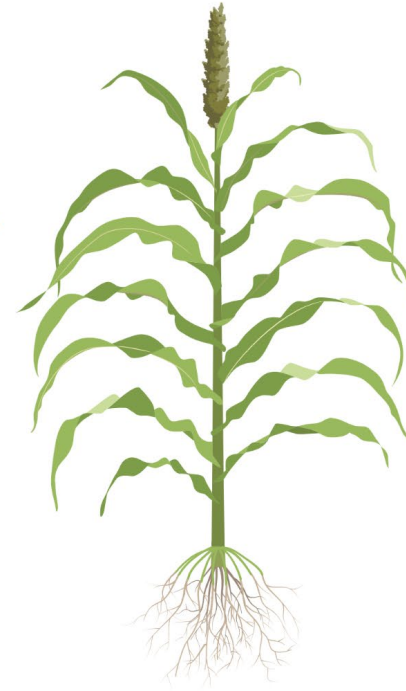
Oryza sativa
(rice)



Multiple

432

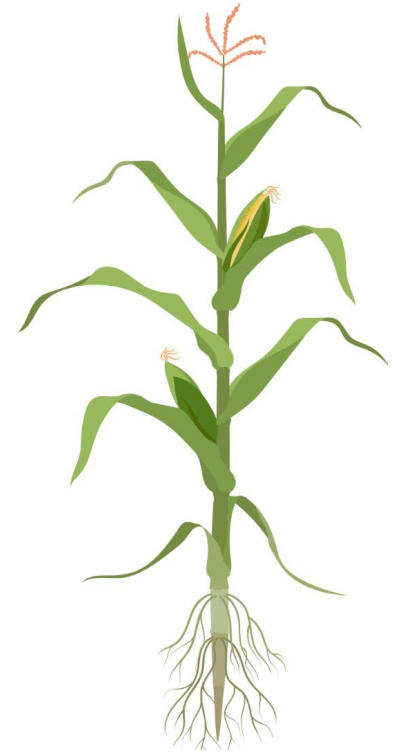
Sorghum bicolor



Multiple

1768

Zea mays
(corn)



Multiple

905

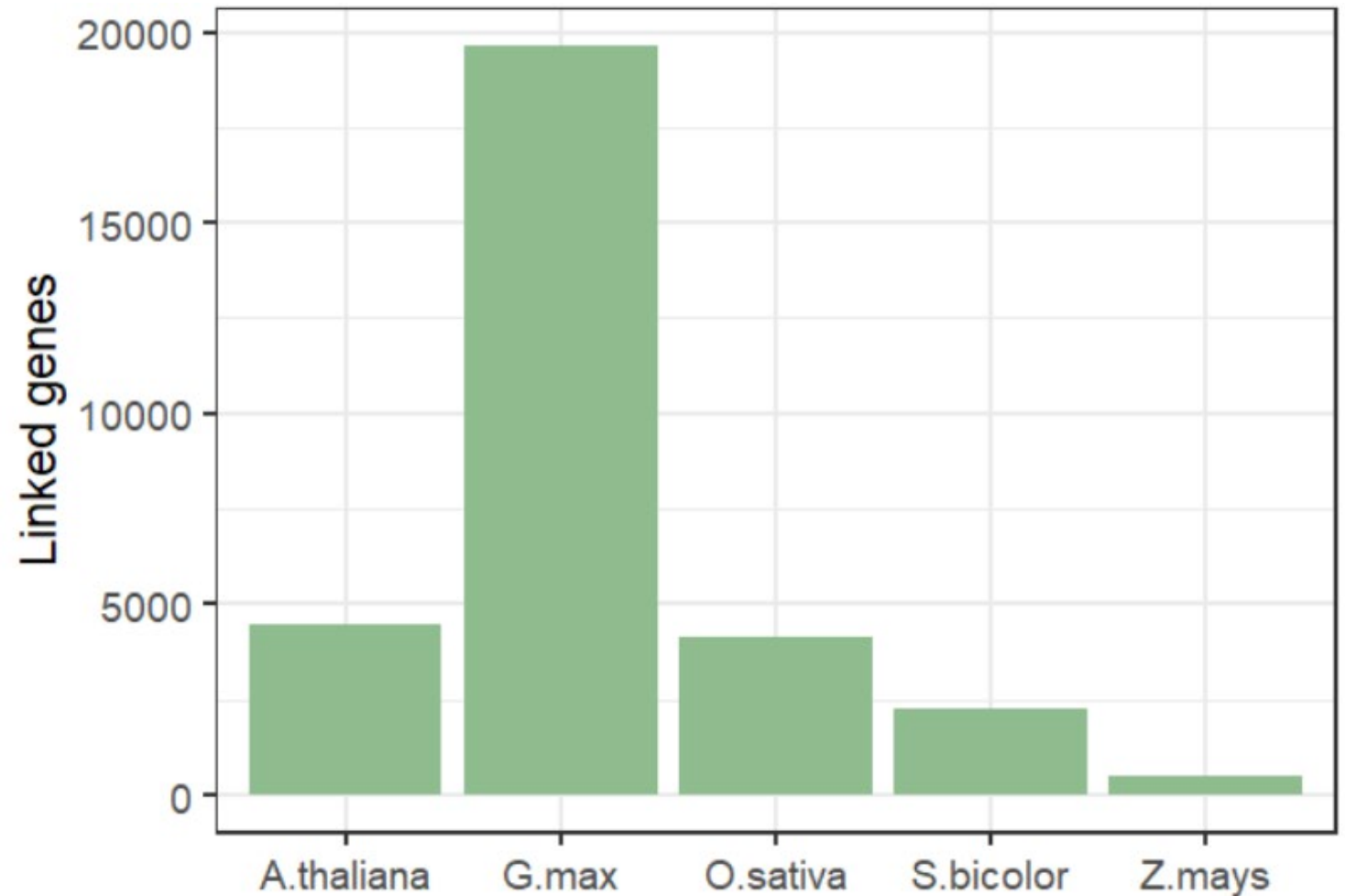
Environments

Loci across 19
traits

Gene lists from GWAS can be quite large, so how do we narrow them down?

Genes linked to molybdenum uptake loci

Species	# GWAS loci
<i>A. thaliana</i>	109
<i>G. max</i>	362
<i>O. sativa</i>	70
<i>S. bicolor</i>	115
<i>Z. mays</i>	73



What Are the Genes Underlying Elemental Accumulation?

New York Times



By **Carl Zimmer**

Sept. 18, 2018

MATTER

Why Your DNA Is Still Uncharted Territory

Scientists are focusing on a relatively small number of human genes and neglecting thousands of others. The reasons have more to do with professional survival than genetics.

Just 15 explanations mostly accounted for how many papers have been published on a particular gene. The reasons have more to do with the working lives of scientists than the genes themselves.

What Does an Ionomic Candidate Gene Look Like?

The image shows a screenshot of a PNAS (Proceedings of the National Academy of Sciences) article page. On the left, a Zscore plot displays several box plots representing different data series. The y-axis is labeled 'Zscore' and ranges from 0 to 5. The plot shows various distributions, with one prominent black box plot extending above the 5 mark. Below the plot are navigation links: 'ABOUT', 'LABS', and 'COMMUNITY'. The main article header includes the PNAS logo and the text 'Proceedings of the National Academy of Sciences of the United States of America'. A search bar is present with the placeholder text 'Keyword, Author, or DOI'. Navigation tabs include 'Home', 'Articles', 'Front Matter', 'News', 'Podcasts', and 'Authors'. Below the header, there are filters for 'NEW RESEARCH IN' with dropdown menus for 'Physical Sciences' and 'Social Sciences'. The main article title is 'Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root'. The authors listed are Prashant S. Hosmani, Takehiro Kamiya, John Danku, Sadaf Naseer, Niko Geldner, Mary Lou Guerinot, and David E. Salt. Below the article title, there are links for 'SUBMIT MY RESEARCH' and 'LOG IN/REGISTER'. At the bottom of the page, there is a section for 'CELL BIOLOGY, PLANT BIOLOGY' with a search icon and a download icon. The article title is 'A receptor-like kinase mutant with absent endodermal diffusion barrier displays selective nutrient homeostasis defects'. The authors listed are Alexandre Pfister, Marie Barberon, Julien Allassimone, Lothar Kalmbach, Yuree Lee, Joop EM Vermeer, Misako Yamazaki, Guowei Li, Christophe Maurel, Junpei Takano, Takehiro Kamiya, David E Salt, Daniele Roppolo, Niko Geldner, and a link to 'see less'. The affiliations are University of Lausanne, Switzerland; CNRS/INRA/SupAgro/Université Montpellier, France; Hokkaido University, Japan; University of Aberdeen, United Kingdom. Below the article title, there is a bar chart showing nutrient levels for Cu, Zn, As, Se, Mo, and Cd.

Zscore

PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or DOI

Home Articles Front Matter News Podcasts Authors

NEW RESEARCH IN Physical Sciences Social Sciences

Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root

Prashant S. Hosmani, Takehiro Kamiya, John Danku, Sadaf Naseer, Niko Geldner, Mary Lou Guerinot and David E. Salt

ABOUT LABS COMMUNITY SUBMIT MY RESEARCH LOG IN/REGISTER

eLIFE MENU HOME MAGAZINE

CELL BIOLOGY, PLANT BIOLOGY

A receptor-like kinase mutant with absent endodermal diffusion barrier displays selective nutrient homeostasis defects

Alexandre Pfister, Marie Barberon, Julien Allassimone, Lothar Kalmbach, Yuree Lee, Joop EM Vermeer, Misako Yamazaki, Guowei Li, Christophe Maurel, Junpei Takano, Takehiro Kamiya, David E Salt, Daniele Roppolo, Niko Geldner « see less

University of Lausanne, Switzerland; CNRS/INRA/SupAgro/Université Montpellier, France; Hokkaido University, Japan; University of Aberdeen, United Kingdom

Cu Zn As Se Mo Cd

esb1 mutant in two different trays
Baxter et al. *PLoS Genetics* 2009

Known Ionomeric Genes: What do we know?

Plant Direct

Open Access

ORIGINAL RESEARCH | Open Access |

A curated list of genes that affect the plant ionome

Lauren Whitt, Felipe Klein Ricachenevsky, Greg Ziegler Ziegler, Stephan Clemens, Elsbeth Walker, Frans J. M. Maathuis, Philip Kear, Ivan Baxter

First published: 21 October 2020 | <https://doi.org/10.1002/pld3.272>

216 primary KIGs – experimentally proven

2523 inferred KIGs – homologs to primary KIGs

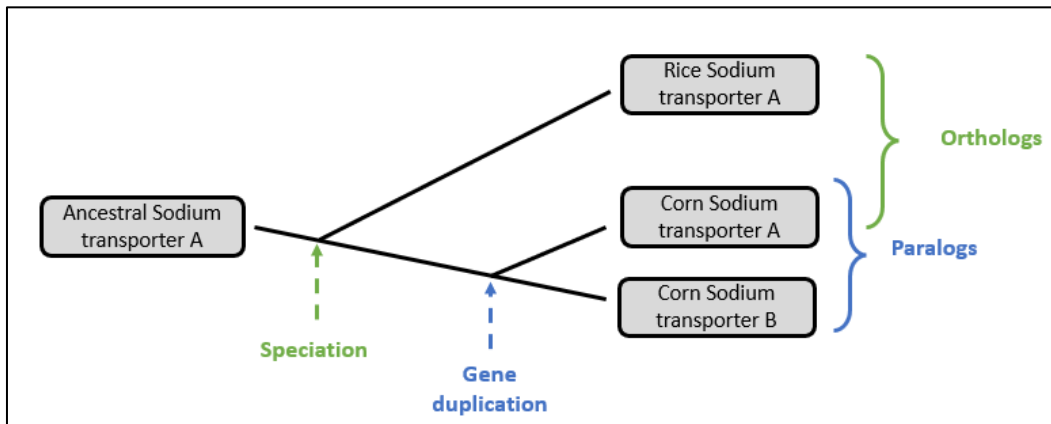
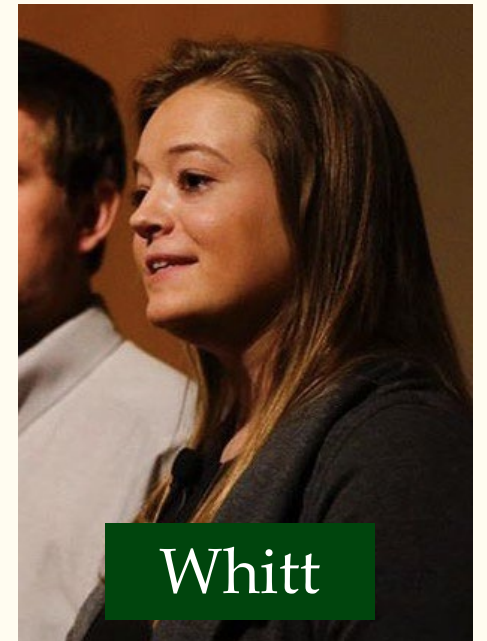


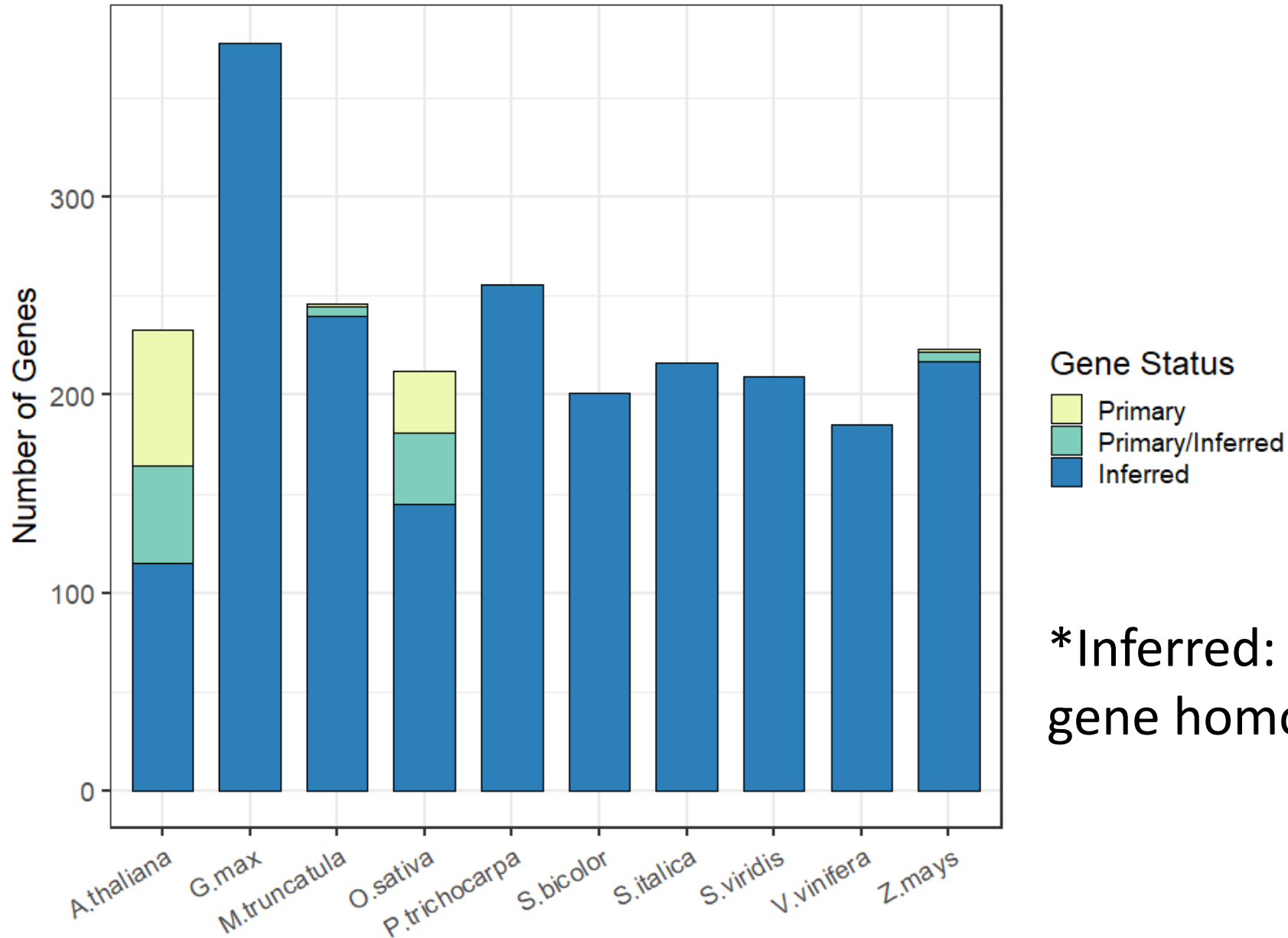
TABLE 1

Primary known ionome genes

Species	GeneID	GeneName	Elements	Tissue	Citation(s)
<i>A. thaliana</i>	AT1G01340	CNGC10	K, Ca, Mg	Roots, shoots	Guo et al. (2010)
<i>A. thaliana</i>	AT1G01580	FRO2			
<i>A. thaliana</i>	AT1G07600	MT1A			
<i>A. thaliana</i>	AT1G08490	CPNIFS			
<i>A. thaliana</i>	AT1G12640	LPCAT1			
<i>A. thaliana</i>	AT1G14040	PHO1;H3			
<i>A. thaliana</i>	AT1G14870	PCR2			
<i>A. thaliana</i>	AT1G18910	BTSL2			
<i>A. thaliana</i>	AT1G20110	FYVE1			
<i>A. thaliana</i>	AT1G30270	CIPK23			
<i>A. thaliana</i>	AT1G30400	ABCC1	Cd	Shoots	Park et al. (2012)
<i>A. thaliana</i>	AT1G30450	CCC	Ca, K, Na,S	seeds	McDowell et al. (2013)



We do Gene Discovery in Models



*Inferred: Primary gene homolog

KIG ortholog groups overlapping GWAS loci in 3+ species

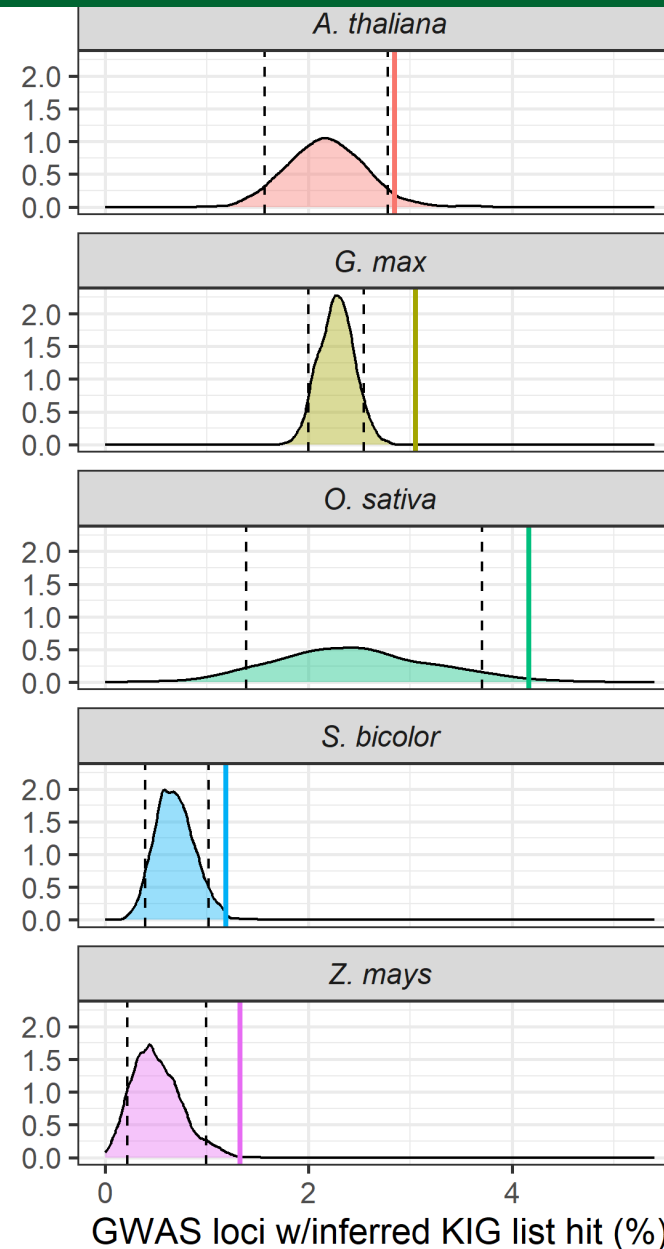
KIG family	Cited trait	GWAS trait	Orthologs
AtZIF1, AtZIFL2 OsZIFL12. ZmYS3	Cs, Zn, Fe	Fe	AtZIFL2 Glyma.06G313600 / 06G313700 Sobic.008G015700
AtPHT1;1, AtPHT1;4, AtPHT1;5 OsPT2	P, As, Se	P	AtPHT1;4 Glyma.10G186400 / 10G186500 LOC_Os08g45000
AtPEN3	Cd	Cd	AT3G16340 Glyma.07G015800 LOC_Os01g52560 / 06g36090
AtNRAMP3	Fe, Mn, Zn	Mn	AtNRAMP3 Glyma.04G044000 Sobic.001G170000 Zm00001d030846
AtMTP11	Mn	Mn	AtMTP11 Glyma.18G286100 Sobic.003G349200 Zm00001d042939
AtMOT2	Mo	Mo	AtMOT2 Glyma.08G208700 LOC_Os01g45830 Sobic.003G237000. Zm00001d044068
AtMOT1. MtMOT1.2, MtMOT1.3. OsMOT1;1	Mo	Mo	AtMOT1. Glyma.06G074100 / 17G203000 / 17G203200 / 17G203500. Zm00001d033053
AtBTS	Fe, Mn, Zn	Fe	Glyma.05G237500 / 08G044700 Sobic.009G222200 Zm00001d038916
AtESB1	Ca, Mn, Zn, Na, S, K, As, Se, Mo	Mn	AT1G07730, AT2G39430 Glyma.08G258300 / 12G067400 / 18G282600 Sobic.003G064800
AtHMA2, AtHMA3, AtHMA4 OsHMA3	Zn, Cd	Cd	AtHMA4, AtHMA2, AtHMA3 Glyma.09G055600 OsHMA3 Sobic.002G083000 / 002G083100 Zm00001d005189, Zm00001d005190

5/5 species



KIG/GWAS Overlap is Significant

Species	False discovery rate (FDR)
<i>A. thaliana</i>	0.73
<i>G. max</i>	0.64
<i>O. sativa</i>	0.53
<i>S. bicolor</i>	0.58
<i>Z. mays</i>	0.38



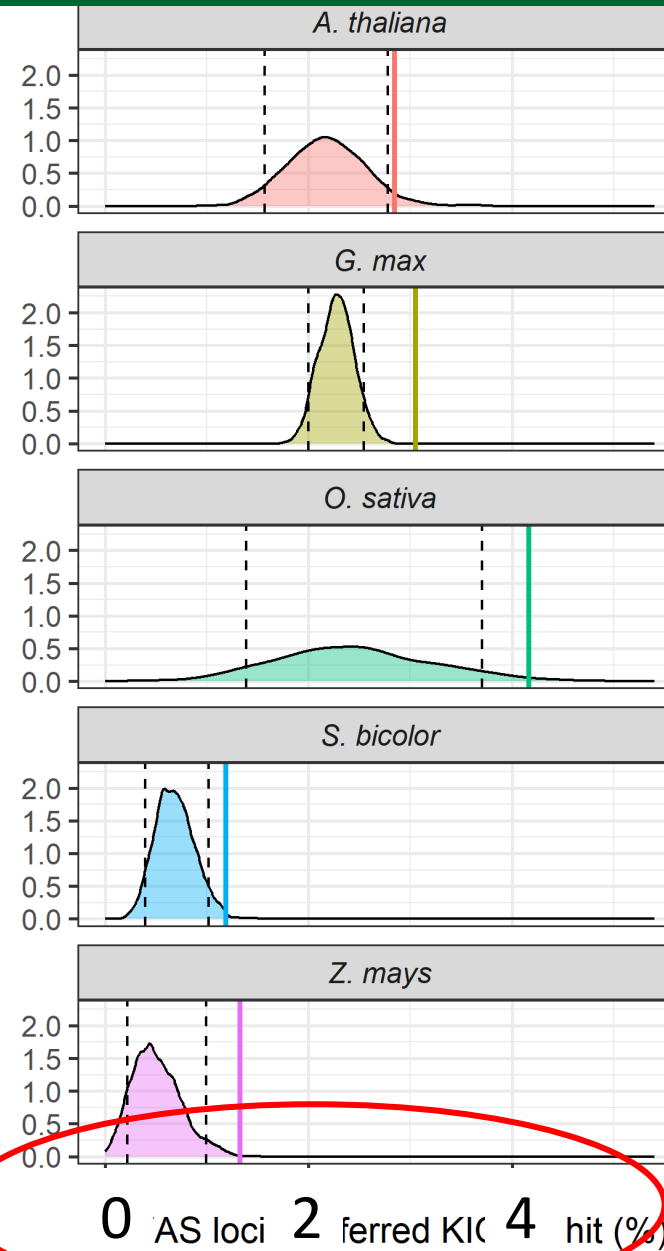
Solid lines mark the total number of loci

Curves show 1000 permutations of the data using random genome locations to replicate loci

Dashed lines mark the 5th/95th percentiles

KIG/GWAS Overlap is <5% of All Loci

Species	False discovery rate (FDR)
<i>A. thaliana</i>	0.73
<i>G. max</i>	0.64
<i>O. sativa</i>	0.53
<i>S. bicolor</i>	0.58
<i>Z. mays</i>	0.38



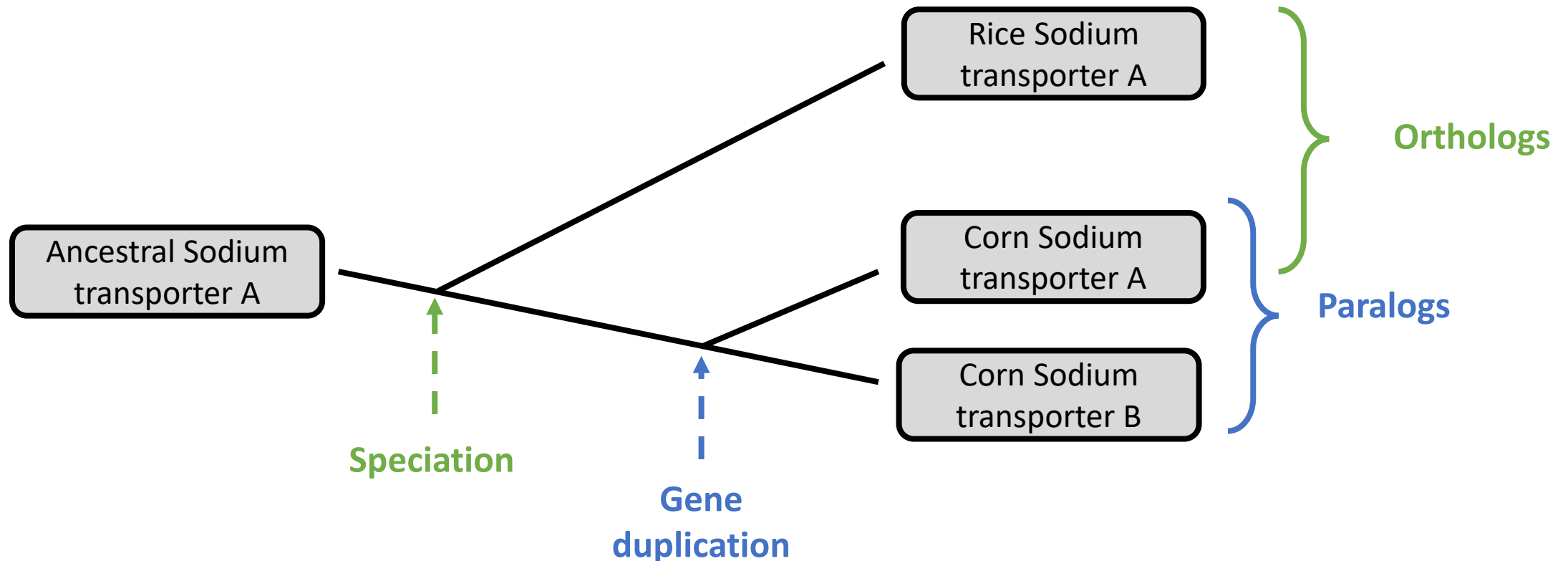
Solid lines mark the total number of loci

Curves show 1000 permutations of the data using random genome locations to replicate loci

Dashed lines mark the 5th/95th percentiles

How Do We Get to Unknown Genes?

If known genes are conserved under peaks,
the unknown (un-KIG) are likely to follow the same pattern.





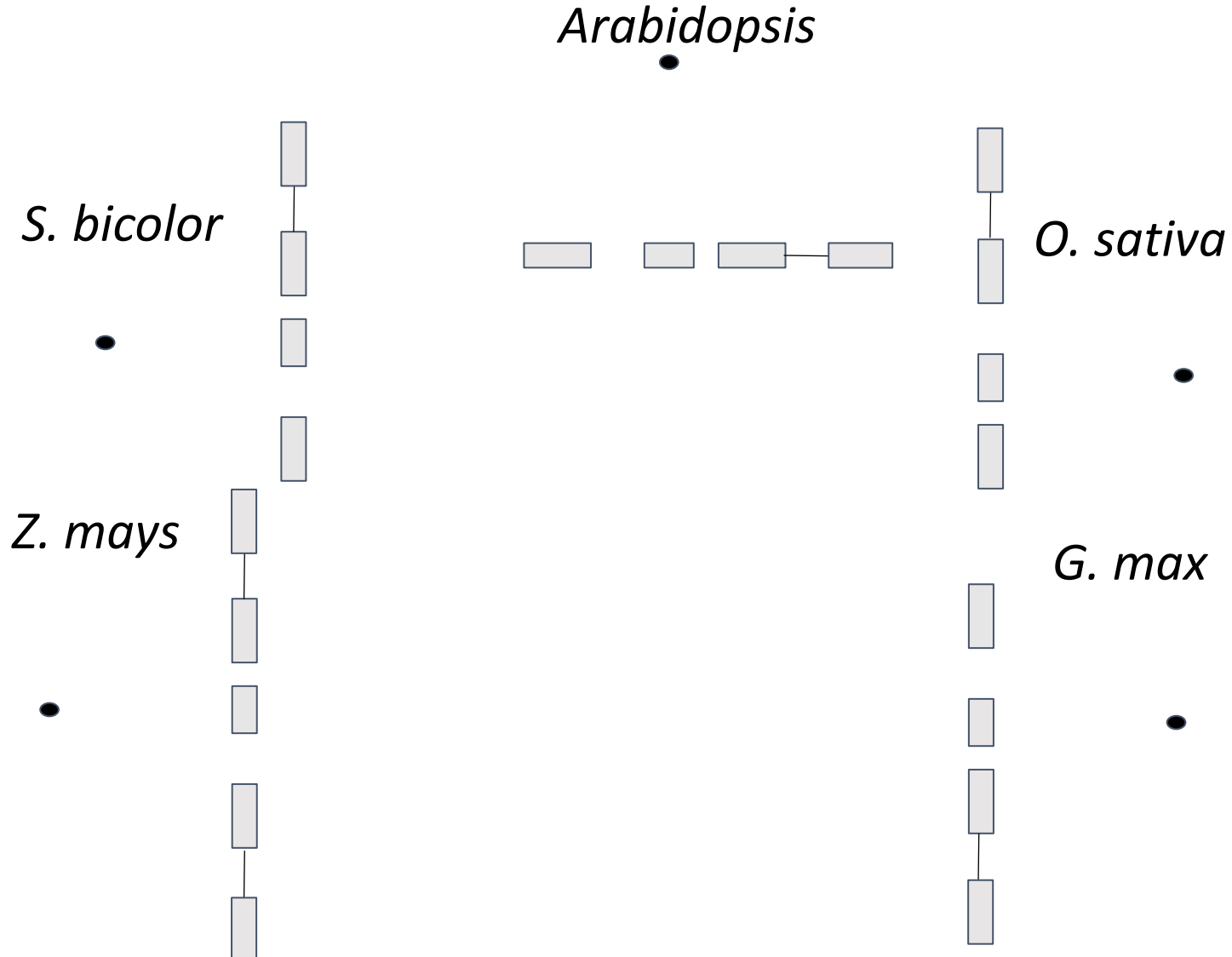
Filtering Results of Multispecies Analogous GWAS Experiments

FiReMAGE Scheme



Legend

- Marker/SNP
- ▭ Gene
- ◁ Linkage range
- ▭ Orthologous gene



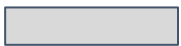
FiReMAGE Scheme



Legend



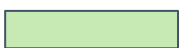
Marker/SNP



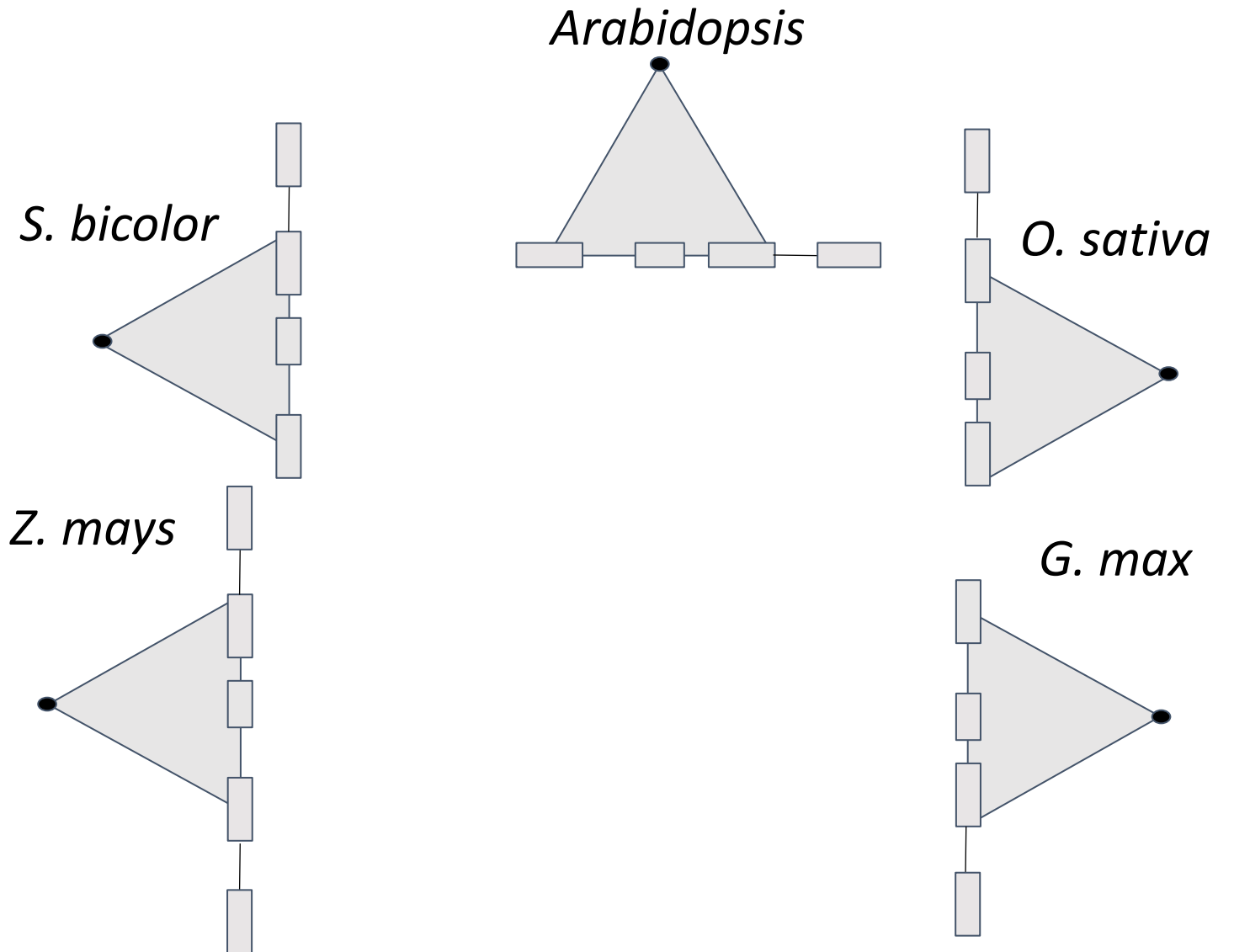
Gene



Linkage range



Orthologous gene



FiReMAGE Scheme



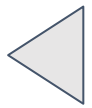
Legend



Marker/SNP



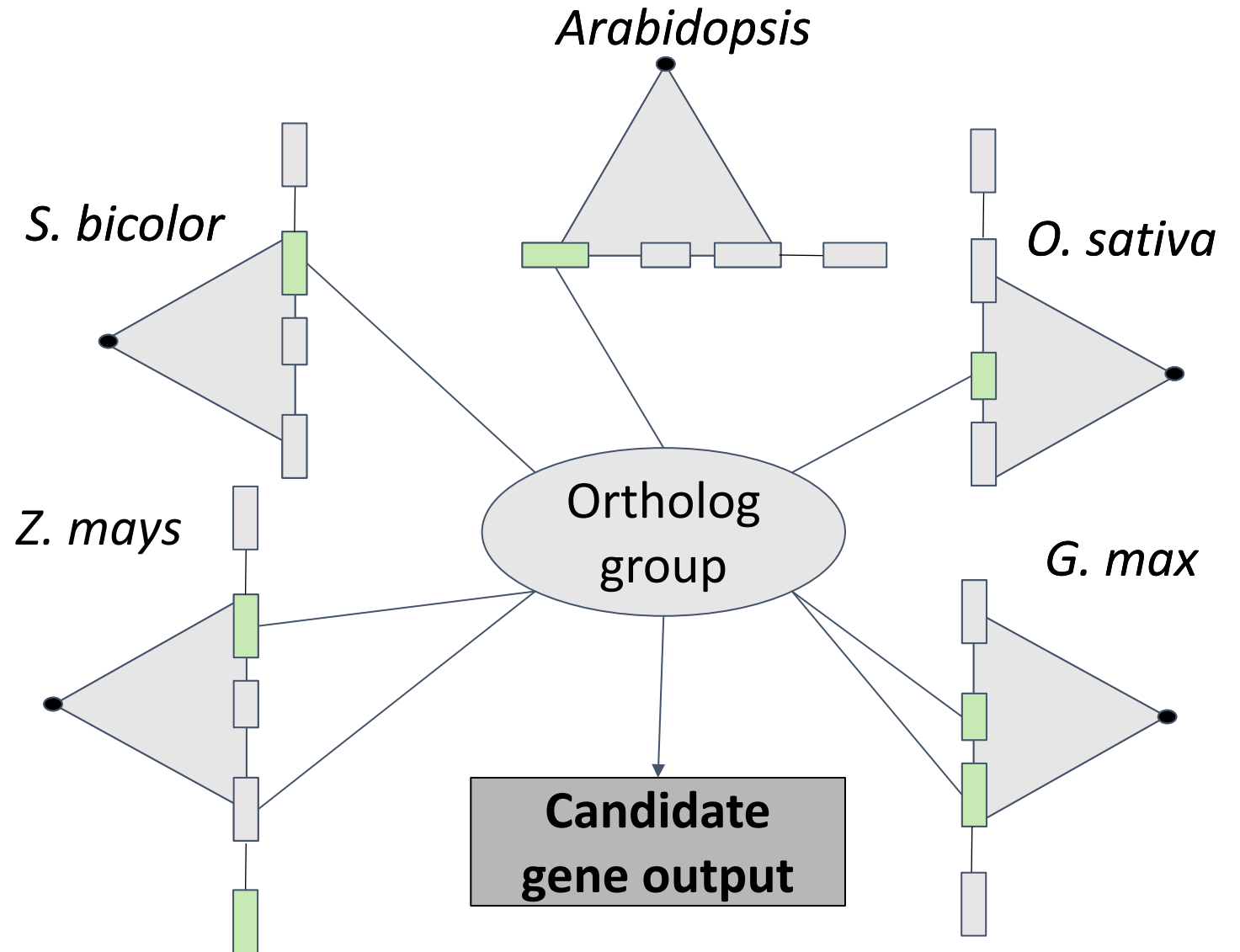
Gene



Linkage range



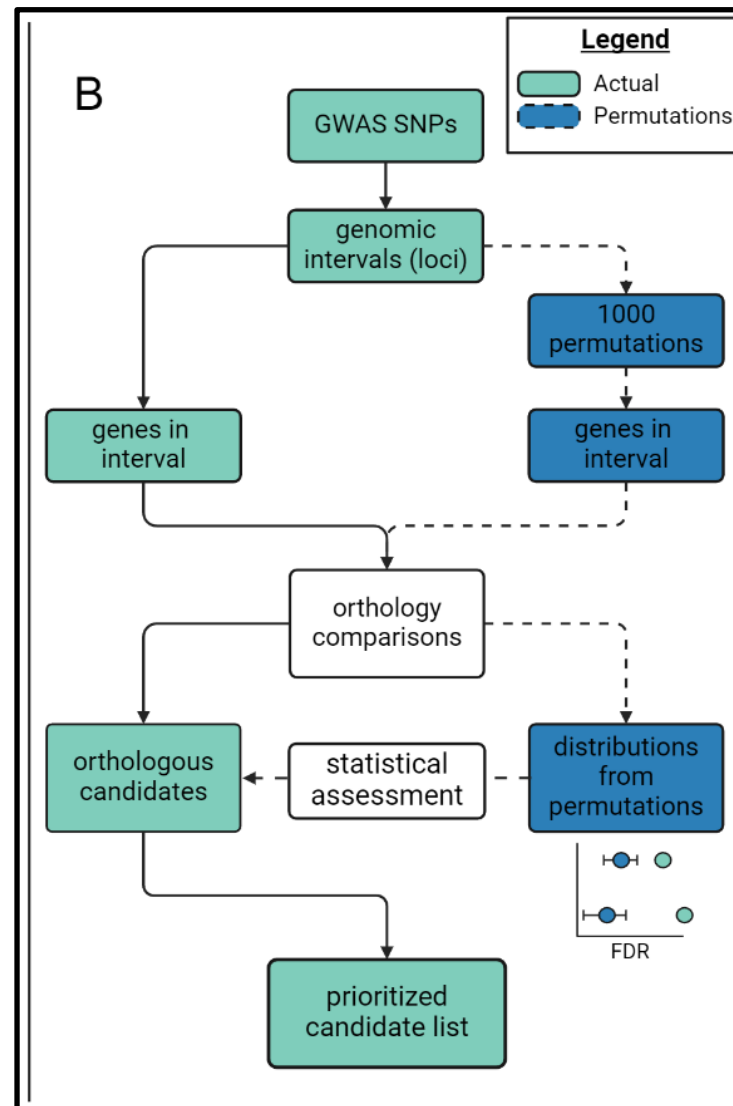
Orthologous gene



Permutations allow us to estimate likelihoods without knowing the “correct” choice for each variable and parameter.

Variables to consider include (but are not limited to):

- Trait measurement error
- GWAS statistical power
- Differences between GWAS
- Species in the comparison
- Linkage ranges
- P-value thresholds

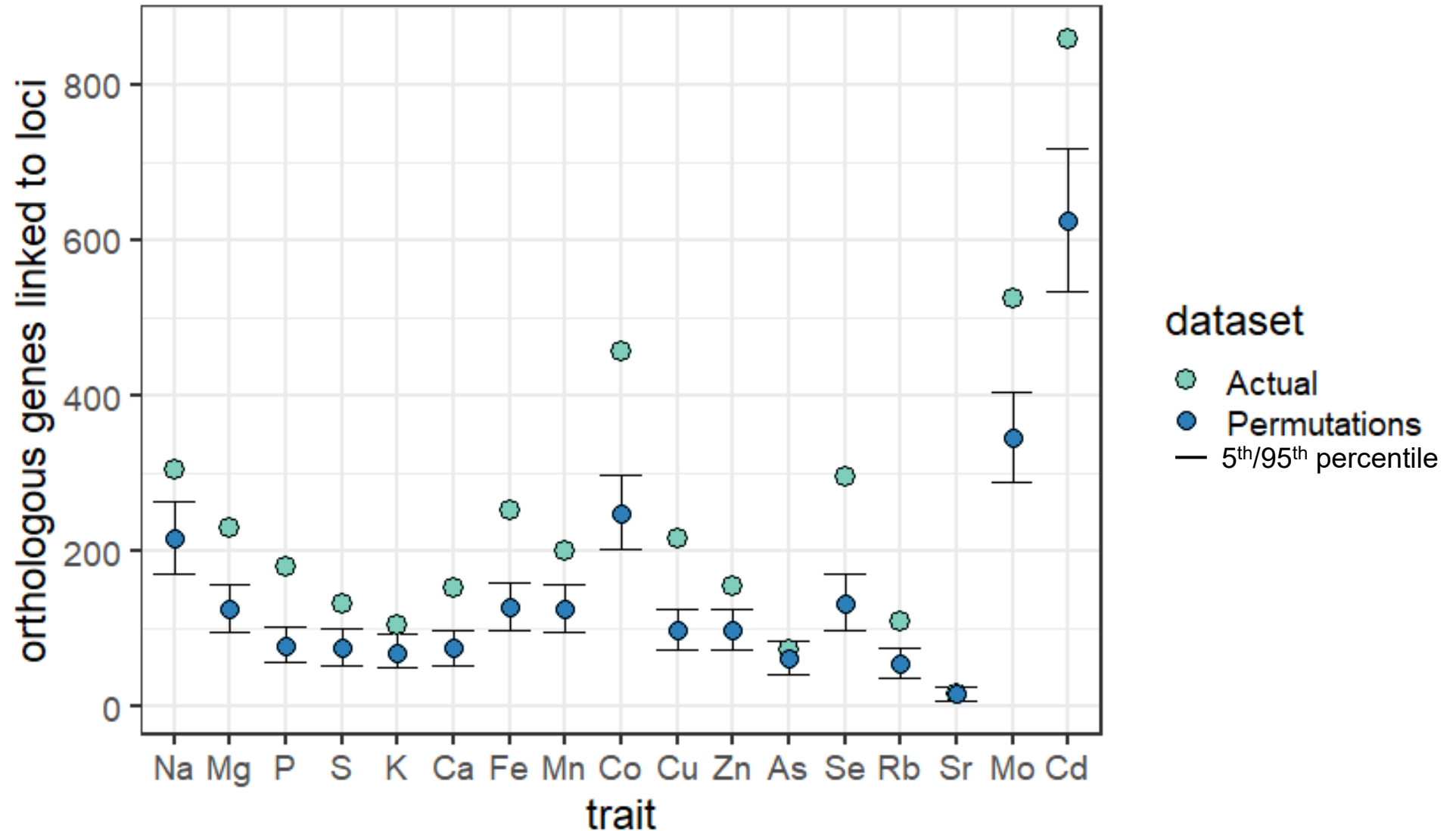


Permutations match the exact species and trait composition of the actual data

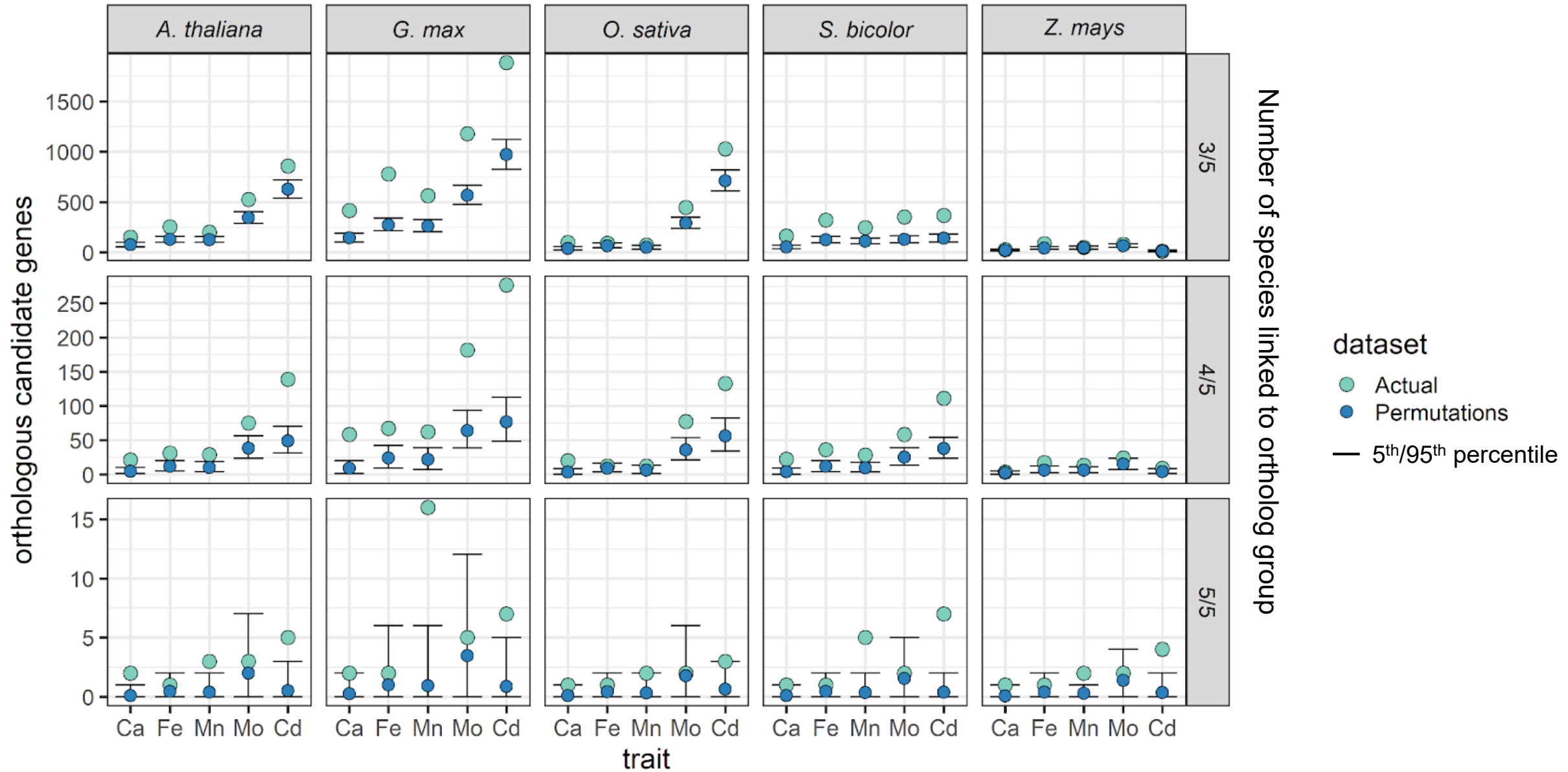
Loci locations are shuffled around randomly

Number of linked orthologs > than expected by random chance

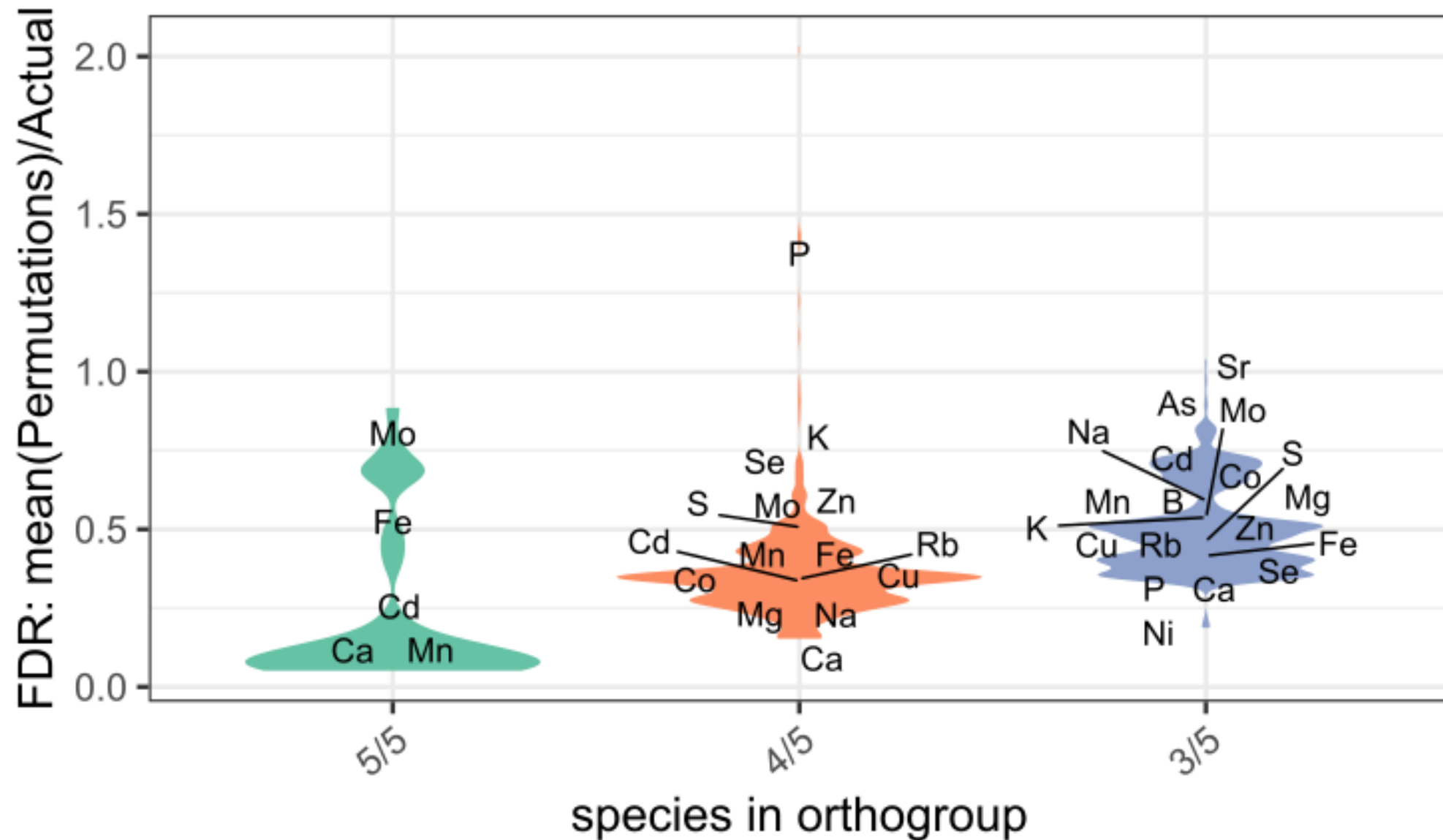
Arabidopsis candidate genes in ortholog groups linked to 3/5 species



Requiring ortholog groups to overlap loci in more species reduces gene list size



Requiring ortholog groups to overlap loci in more species reduces False Discovery Rate



5/5 Ortholog Groups: Known and Unknown

Ortholog group	trait	Phytozome description	Supporting Literature
OG0002041	Ca	Arabidopsis: RING/U-box superfamily	Currently unknown role
OG0005693	Fe	Mevalonate kinase	Currently unknown role
OG0000070	Mn	Histone proteins	Currently unknown role
OG0000306	Cd	RAB GTPase homolog A1	Currently unknown role
OG0002774	Cd	Cd/Zn heavy metal ATPases	KIG list
OG0013522	Mo	Molybdate transporter 2	KIG list
OG0000899	Cd	DUF567 – protein of unknown function	In a list of 129 <i>Arabidopsis halleri</i> genes associated with metal hyperaccumulation and expressed in roots (Filatov et al. 2006, Mol. Ecol.)
OG0002849	Mo	Adenine nucleotide transporters NTT1/NTT2	Molybdopterin is a Mo containing cofactor synthesized from nucleotide precursors (Hover et al. 2013, J. Amer. Chem. Soc.)
OG0000048	Mn	Cellulose synthase catalytic subunit	Mn ions are used to catalyze glycosidic bond formation during cellulose synthesis (Qiao et al. 2021, PNAS)

Baxter Lab

Lauren Whitt

Magdalena Janik

Liz Mahood

Greg Ziegler

Collin Luebbert

Jennie Barrett

Melissa Jurkowski

Slotkin Lab

Lauren McClain

Seth Edwards

Trey Klaas

Danforth Center

Noah Fahlgren

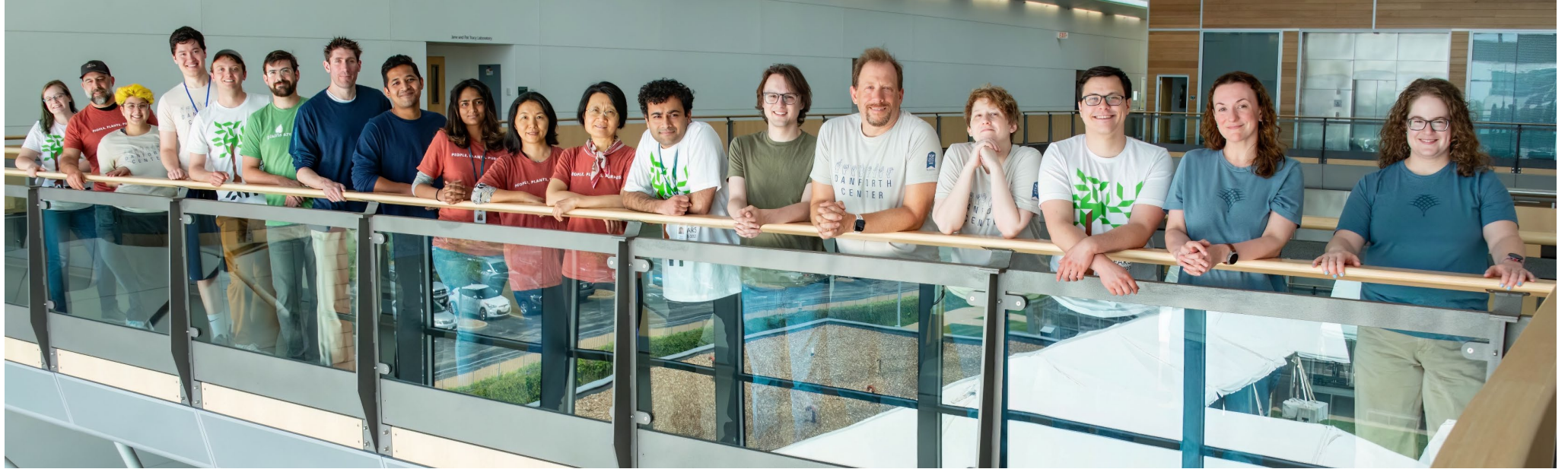
Josh Rothgapt

Armando Bravo

Toby Kellogg

Chris Topp

Topp Lab



Collaborators

Brian Dilkes, Purdue University

Jason Gillman, USDA-ARS / MU

David Salt, University of Nottingham

Grant Norton, University of Aberdeen



College of Agriculture,
Food & Natural Resources
University of Missouri



FFAR Fellows
Future Leaders for Food & Agriculture



DONALD DANFORTH
PLANT SCIENCE CENTER
DISCOVERY | COMMUNITY | IMPACT