

Development of reverse genetic tools for wheat

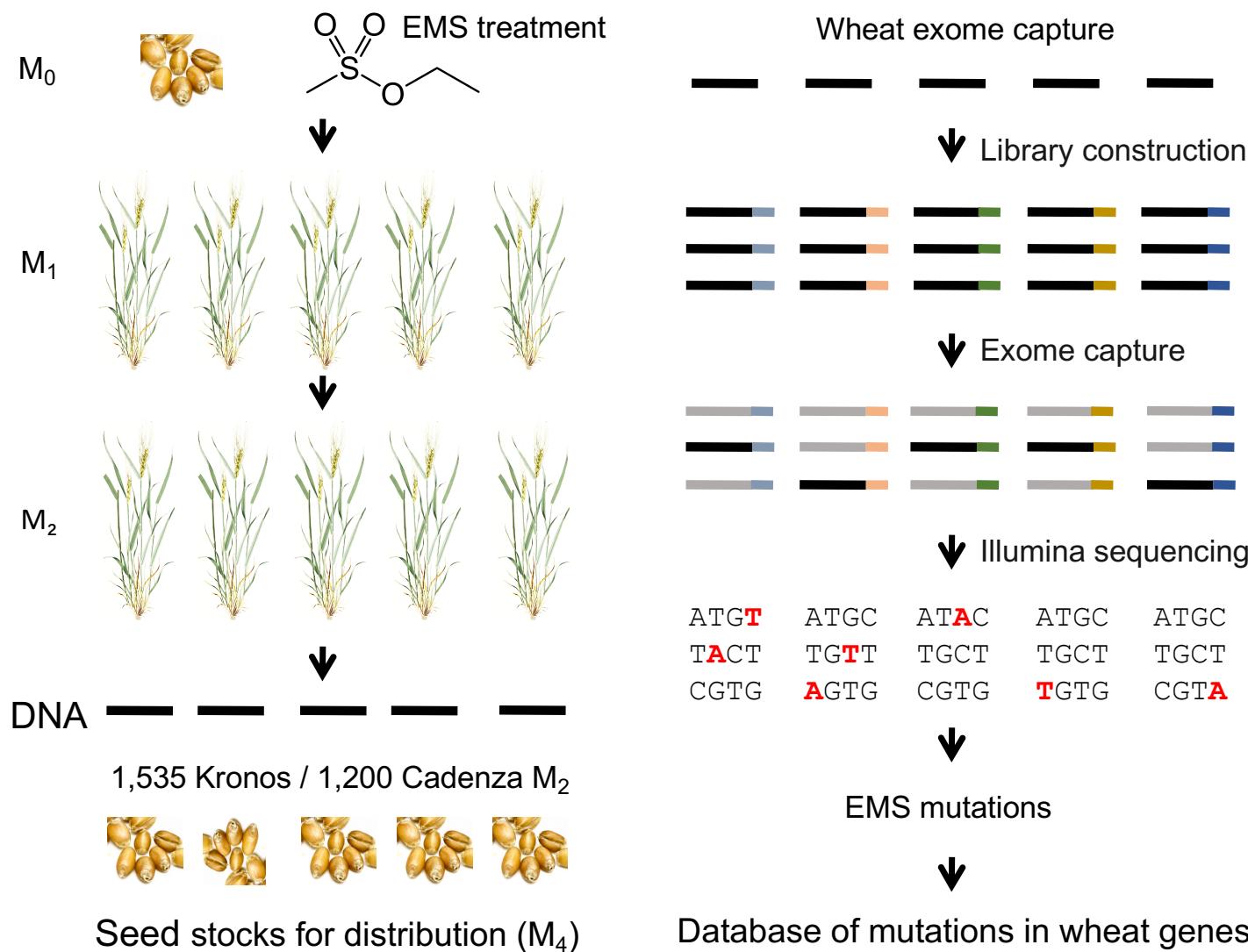
Ksenia Krasileva
Group Leader EI/TSL
[@kseniakrasileva](https://twitter.com/kseniakrasileva)

Cristobal Uauy
Group Leader JIC
[@CristobalUauy](https://twitter.com/CristobalUauy)

Wheat as a model organism, Dec 6th 2016

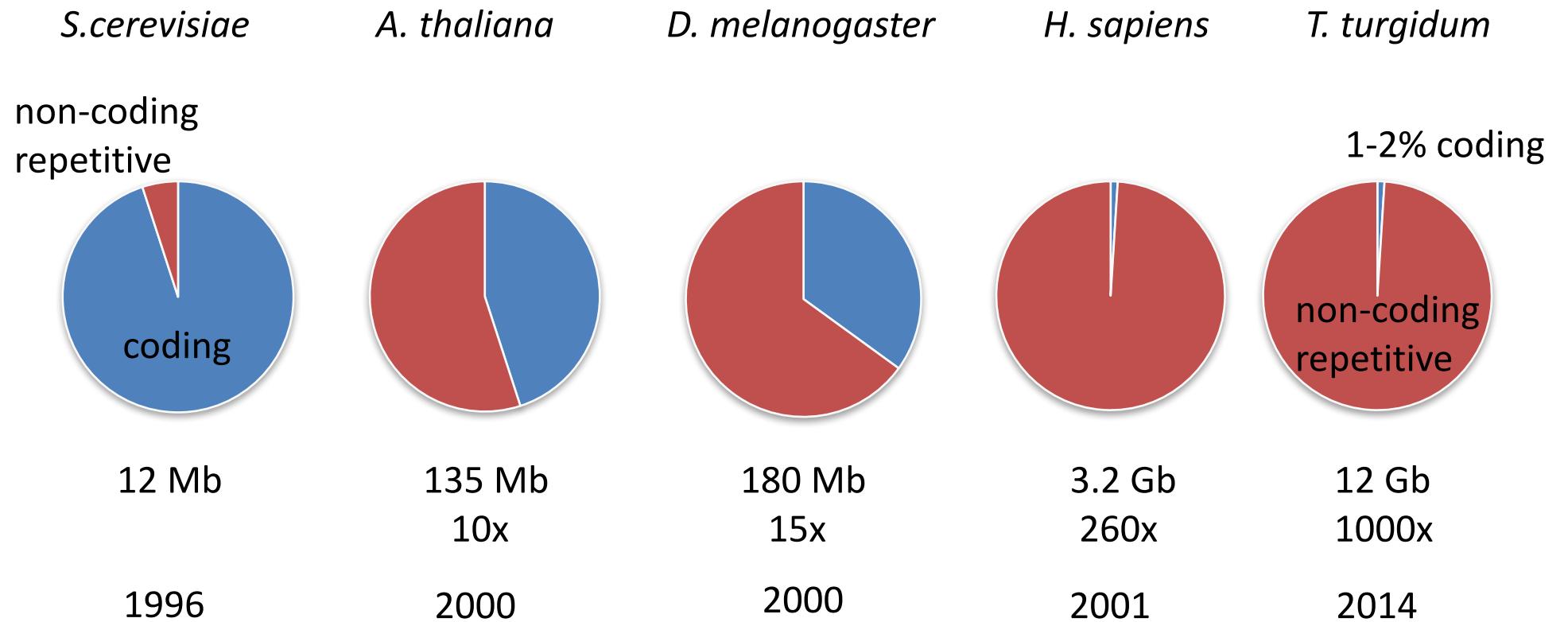


Reverse genetics in wheat: TILLING-by-sequencing

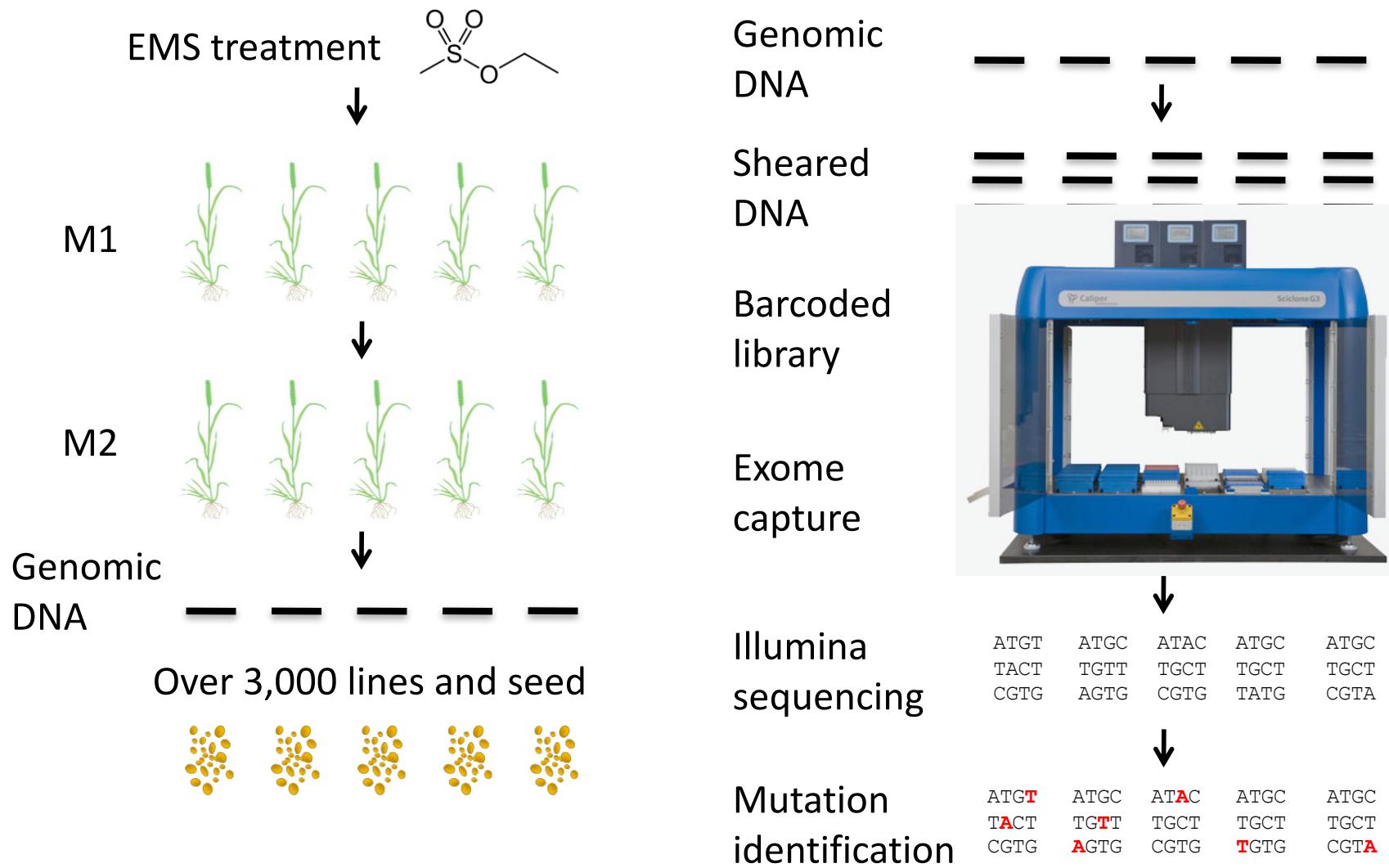


Krasileva et al, *in review* 2016

Most of the wheat genome is non-coding and repetitive



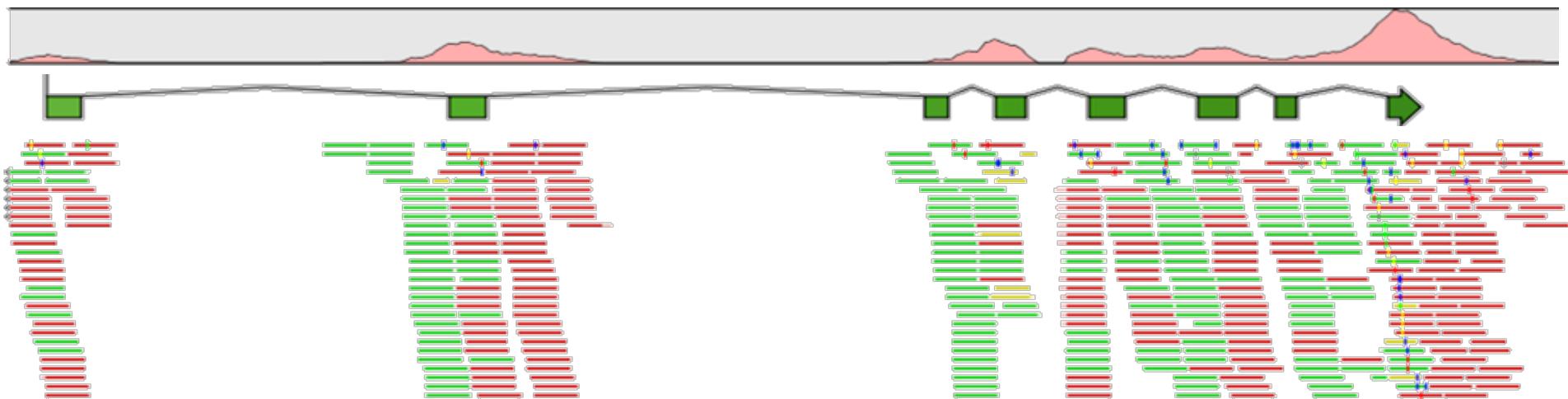
in silico TILLING workflow: Cadenza (6x) and Kronos (4x)



Krasileva et al *Genome Biology* 2013

Krasileva et al *under review*

Exome capture is extremely effective !

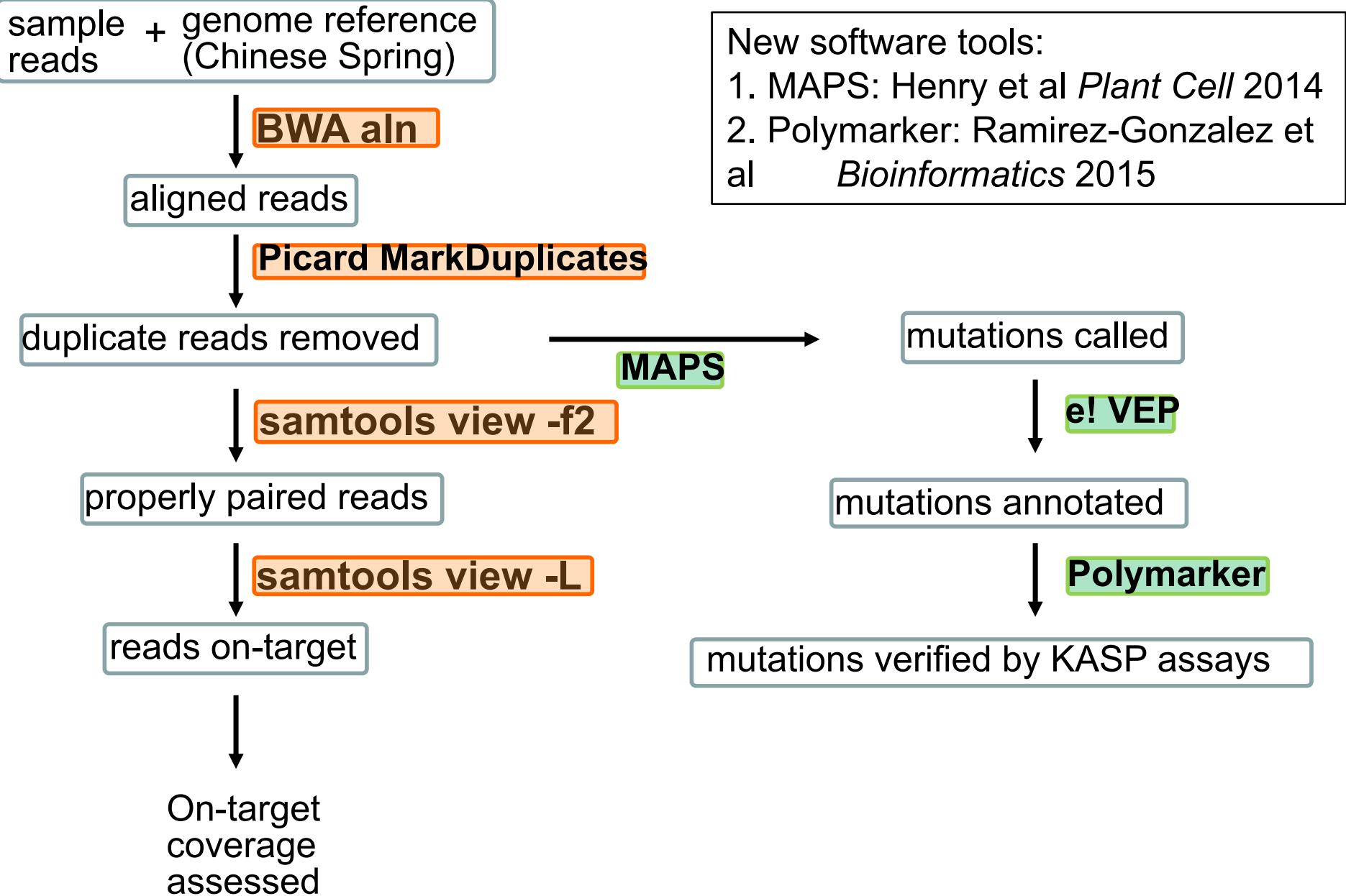


4-plex pooling = ~ 10 Gb per mutant ($0.625 \times$ whole genome coverage)

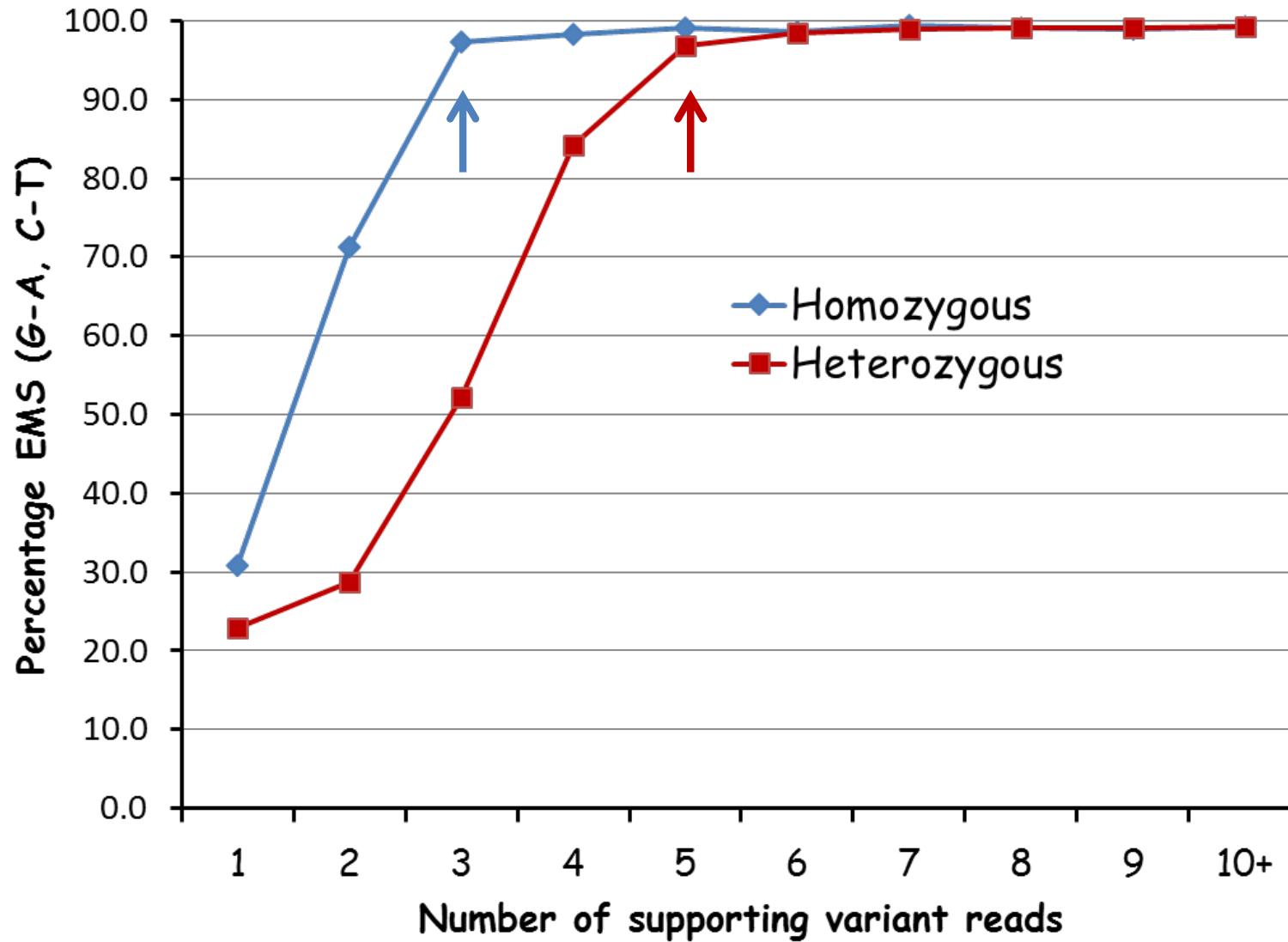
At least 60% of captured reads map are “on-target”

On-target Cov Per Base	% On-target Bases $\geq 1 \times$ Cov	% On-target Bases $\geq 6 \times$ Cov	% Genome Coverage Bases $\geq 1 \times$ Cov	% Genome Coverage Bases $\geq 6 \times$ Cov
45.51 x	93.1 %	85.8 %	8.7 %	1.6 %

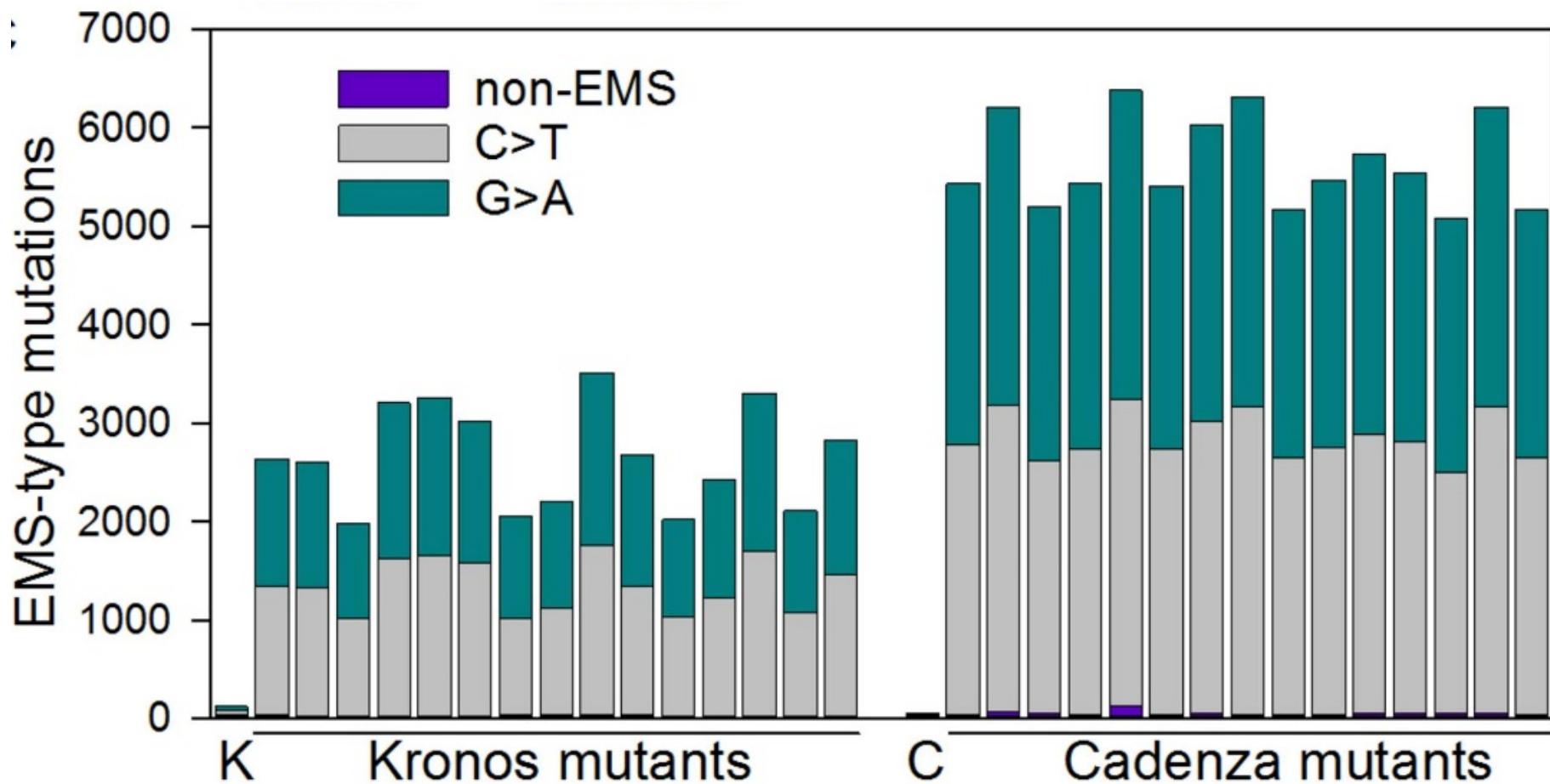
Read alignment and mutation calling pipeline



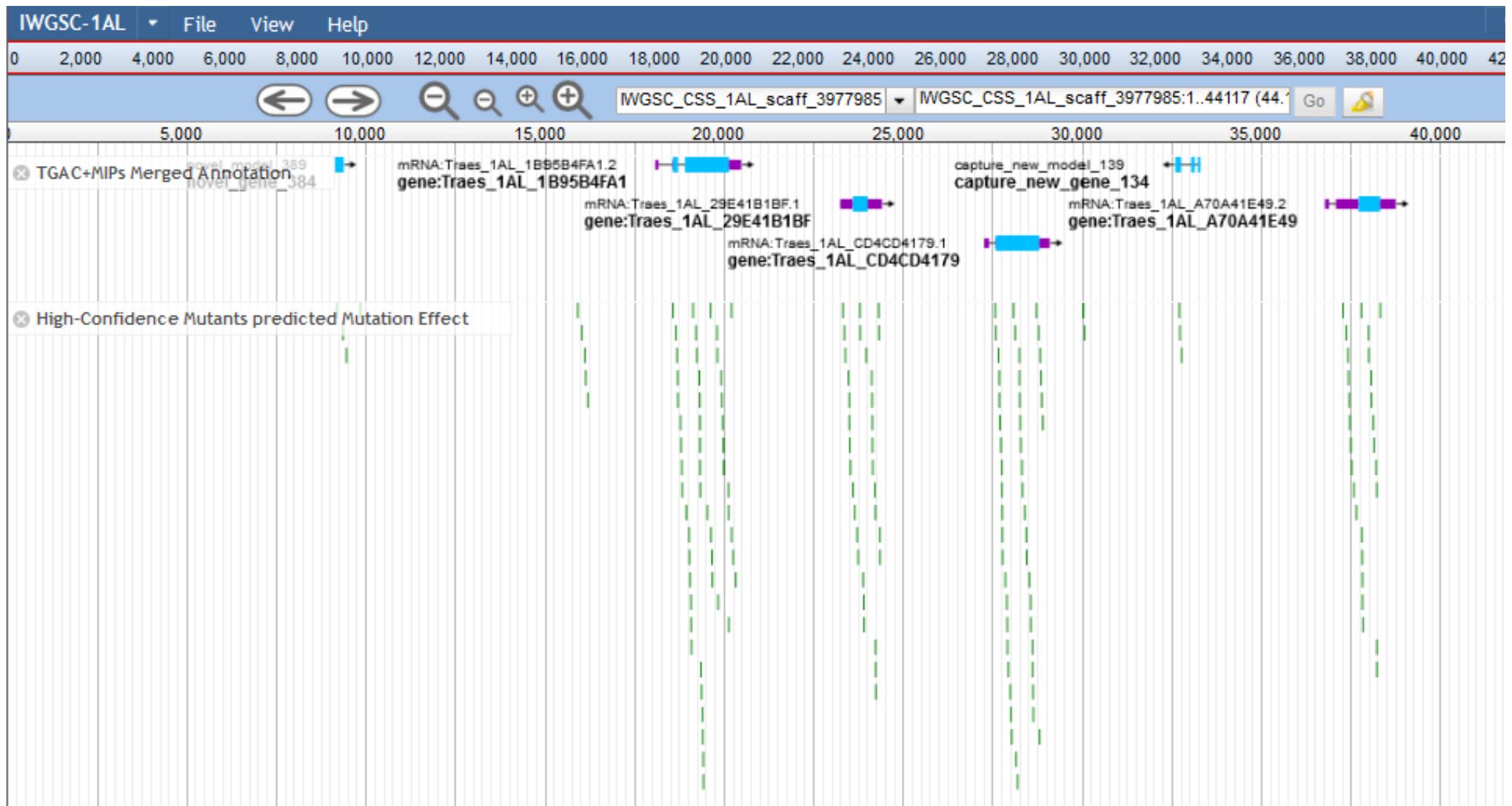
EMS mainly G>A and C>T transitions



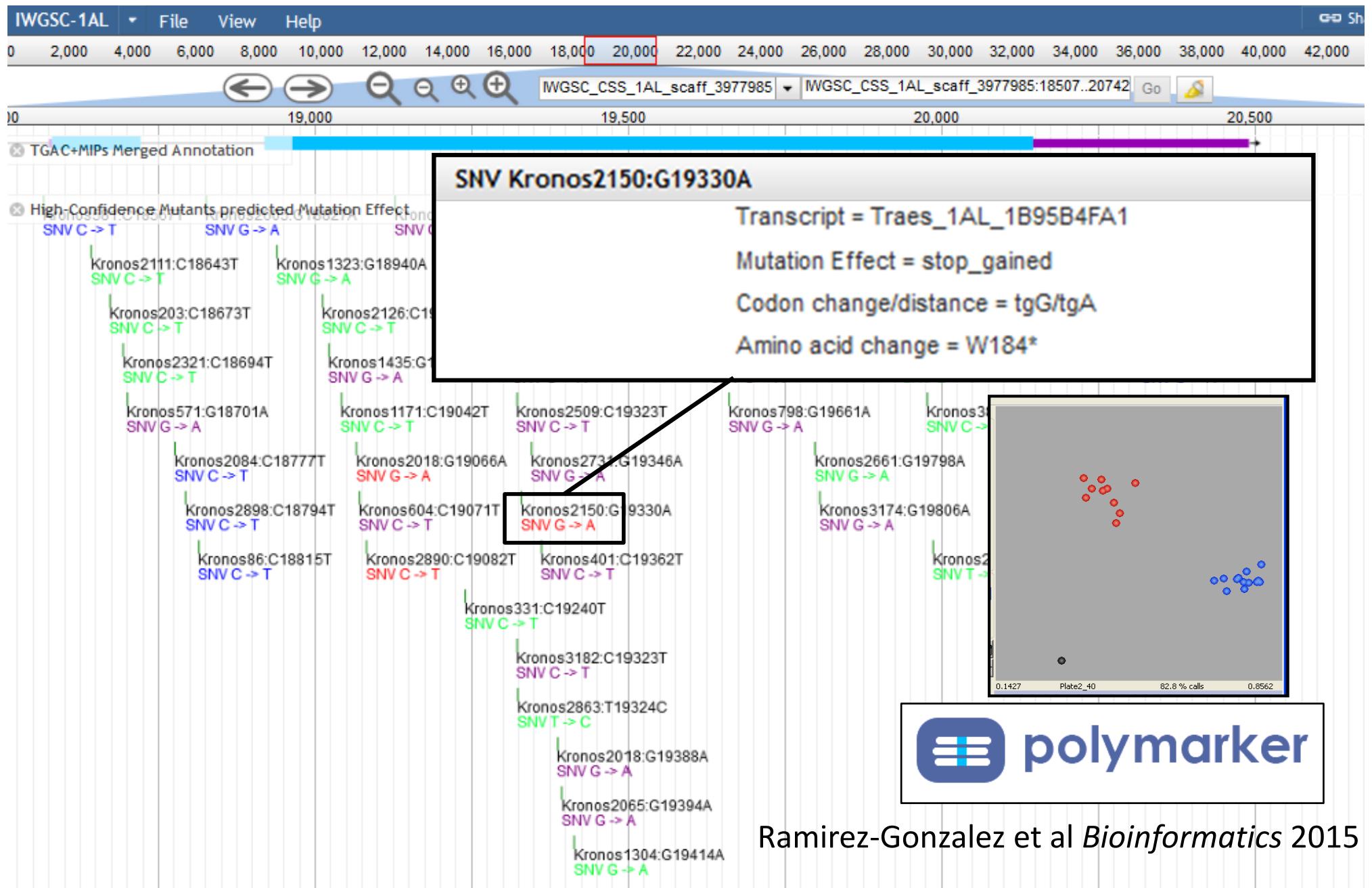
Distribution of mutations (subset)



Genome Browser www.dubcovskylab.ucdavis.edu



Genome Browser www.dubcovskylab.ucdavis.edu





Welcome to the *in silico* wheat Target Induced Local Lesions In Genome (TILLING) website

This resource consists of TILLING populations developed in tetraploid durum wheat cv 'Kronos' and hexaploid bread wheat cv 'Cadenza' as part of a joint project between the University of California Davis, Rothamsted Research, The Earlham Institute, and John Innes Centre.

We have re-sequenced the exome of 1,535 Kronos and 1,200 Cadenza mutants using Illumina next-generation sequencing, aligned this raw data to the IWGSC Chinese Spring chromosome arm survey sequence, identified mutations, and predicted their effects based on the protein annotation available at Ensembl Plants.

Search TILLING data

Population Cadenza Kronos Both

Type in list of search terms (scaffold, line or gene)

Search the database by: gene (eg. `Traes_1AL_9EC1E6F0C`;
`Traes_1AL_9EC1E6F0C.1`), scaffold (eg. `IWGSC_CSS_2AL_scaff_6343779`;
`2AL_6343779`) or, mutant line (eg. `Cadenza0250`)

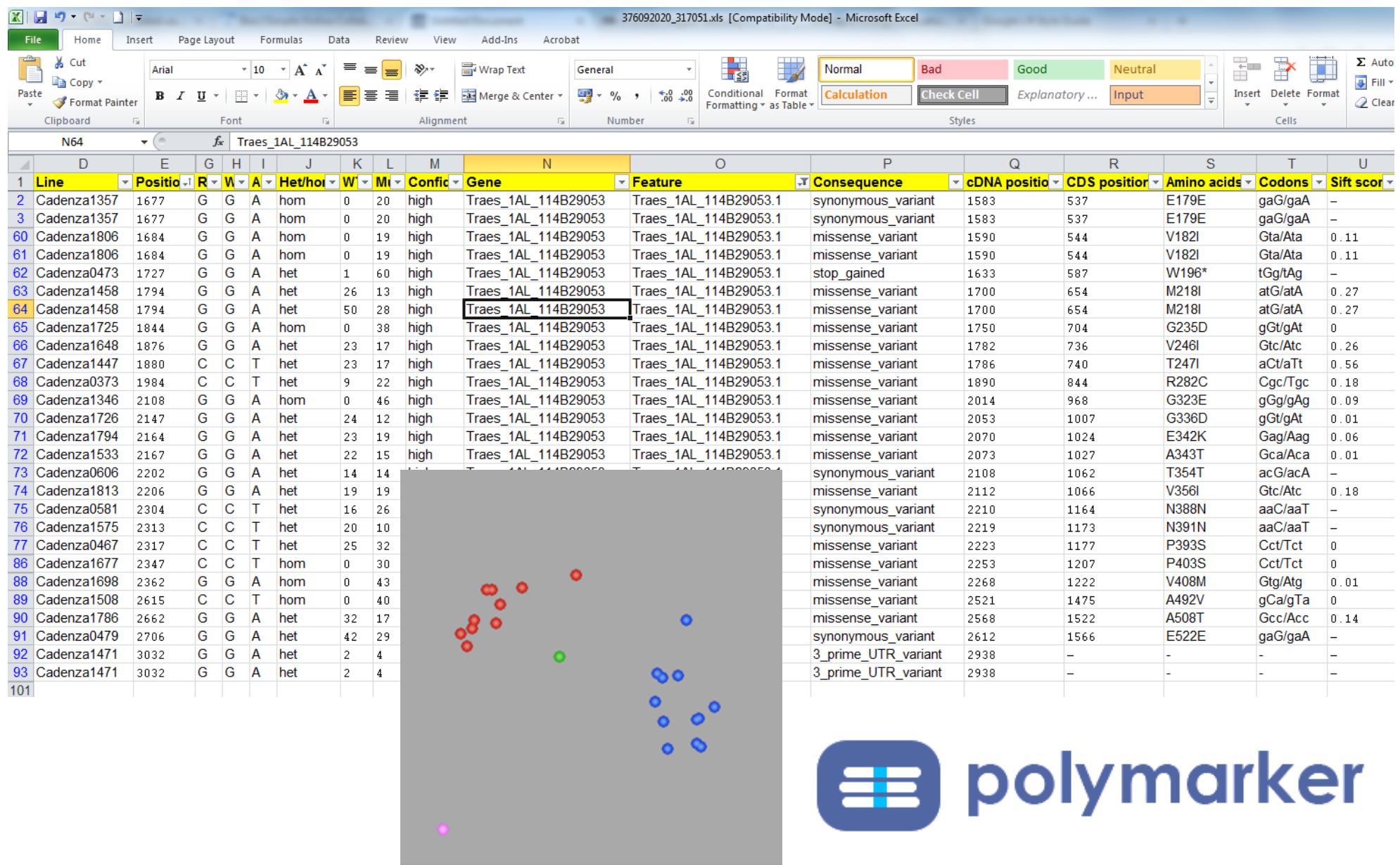
Examples: `Traes_1AL_9EC1E6F0C`, `Traes_1AL_9EC1E6F0C.1`, `IWGSC_CSS_2AL_scaff_6343779`, `2AL_6343779`,
`Cadenza0250`

BLAST Scaffold

Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...

User: Triticum
Password: FiatLux

Example output (based on gene)



Column name	Description
Scaffold	This is the name of the IWGSC CSS scaffold using the EnsemblPlants nomenclature. The name has a hyperlink to the actual FASTA sequence of the scaffold.
Chr	Chromosome number.
Allele	Allele ID.
Amino acid	Amino acid change due to the mutation.
Codons	Sequence of the wild type / mutant codon in the mutant line. The uppercase letter indicates the base which was mutated. For example, cCa/cTa corresponds to a mutation in the middle position of the CCA codon (P) which was mutated to a CTA codon (L).
Sift score	Sorting Intolerant From Tolerant (SIFT) probability score that that the amino acid change is tolerated. Scores <0.05 are considered deleterious whereas all others are considered tolerated. SIFT was implemented through Ensembl VEP as described in Krasileva et al; Materials and Methods section 3.6.
Primer type	The predicted specificity of the designed KASP assays according to its ability to preferentially amplify a single target genome (specific), discriminate against at least one alternative genome (semi-specific) or not able to discriminate and hence amplifying all genomes or paralogues of the gene (non-specific). Specific and semi-specific assays are preferred where possible for downstream validation of the mutation. Primer were designed using http://polymarker.tgac.ac.uk/ (Ramirez-Gonzalez et al 2015).
Orientation	Plus (+) or minus (-) according to the strand orientation of wild type and alternative primers on the IWGSC scaffold (column 1) sequence.
WT primer*	Sequence of the wild type primer. Uppercase bases within the primer sequence correspond to genome specific SNPs, whereas the 3' uppercase base corresponds to the target EMS SNP.
Alt primer*	Sequence of the alternative primer which amplifies the mutant allele.
Common	Common primer sequence for the KASP assay

Summary



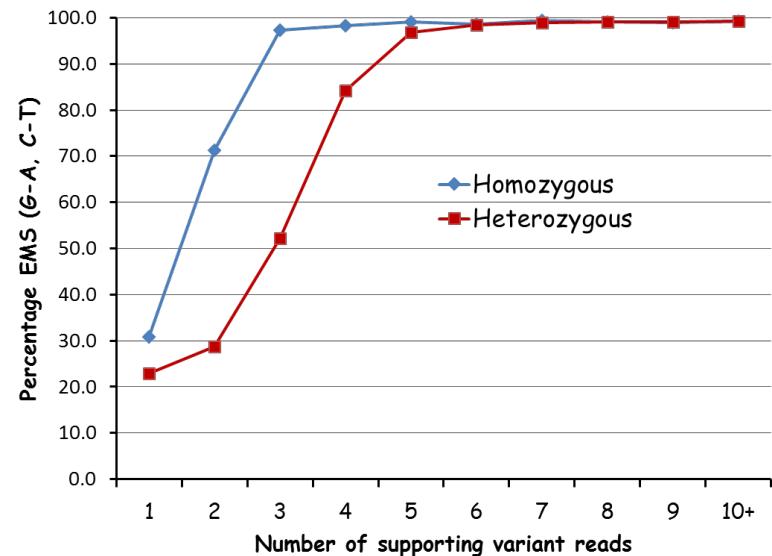
	Tetraploid Kronos	Hexaploid Cadenza
Uniquely mapped SNPs	4,189,561	6,470,733
Heterozygous / Homozygous ratio at M ₂	1.87	2.21
Uniquely mapped EMS-type mutations	4,152,707	6,421,522
Avg. EMS-type mutations/line	2,705	5,351
Avg. EMS-type mutations per kb (population)	34.8	39.5
% EMS-type	99.1%	99.2%

Summary



	Tetraploid	Hexaploid
	Kronos	Cadenza
Gene models with at least one mutation (GM ₁)	48,172	73,895
GM ₁ with at least one truncation	28,604 (59%)	45,311 (61%)
GM ₁ with at least one missense mutation	46,198 (96%)	69,543 (94%)
Avg. number of missense mutations per GM ₁	21.4	22.6
GM ₁ with truncation and/or deleterious missense	43,787 (91%)	67,830 (92%)
No. of unique genes eliminated in large deletions	832	6,657
<i>splice_donor_variant</i>	15,074	26,783
<i>splice_acceptor_variant</i>	14,624	20,889
<i>stop_gained</i>	46,580	85,985
<i>initiator_codon_variant</i>	943	1,953
<i>missense_variant</i>	1,030,287	1,668,693

Lower confidence



Coverage	# SNPs	Het/Hom	EMS SNP	Avg. EMS SNP / line	%EMS	non-EMS transitions	%EMS error
Kronos							
<i>HetMC3/HomMC2</i>	5,525,228	2.46	5,085,379	3,313	92.04	35,707	0.70
<i>HetMC4/HomMC3</i>	4,601,287	2.15	4,507,550	2,937	97.96	12,525	0.28
<i>HetMC5/HomMC3</i>	4,189,561	1.87	4,152,707	2,705	99.12	7,323	0.18
Cadenza							
<i>HetMC3/HomMC2</i>	8,599,721	2.85	8,083,066	6,736	93.99	108,261	1.34
<i>HetMC4/HomMC3</i>	7,203,110	2.58	7,054,109	5,878	97.93	29,019	0.41
<i>HetMC5/HomMC3</i>	6,470,733	2.21	6,421,522	5,351	99.24	10,569	0.16

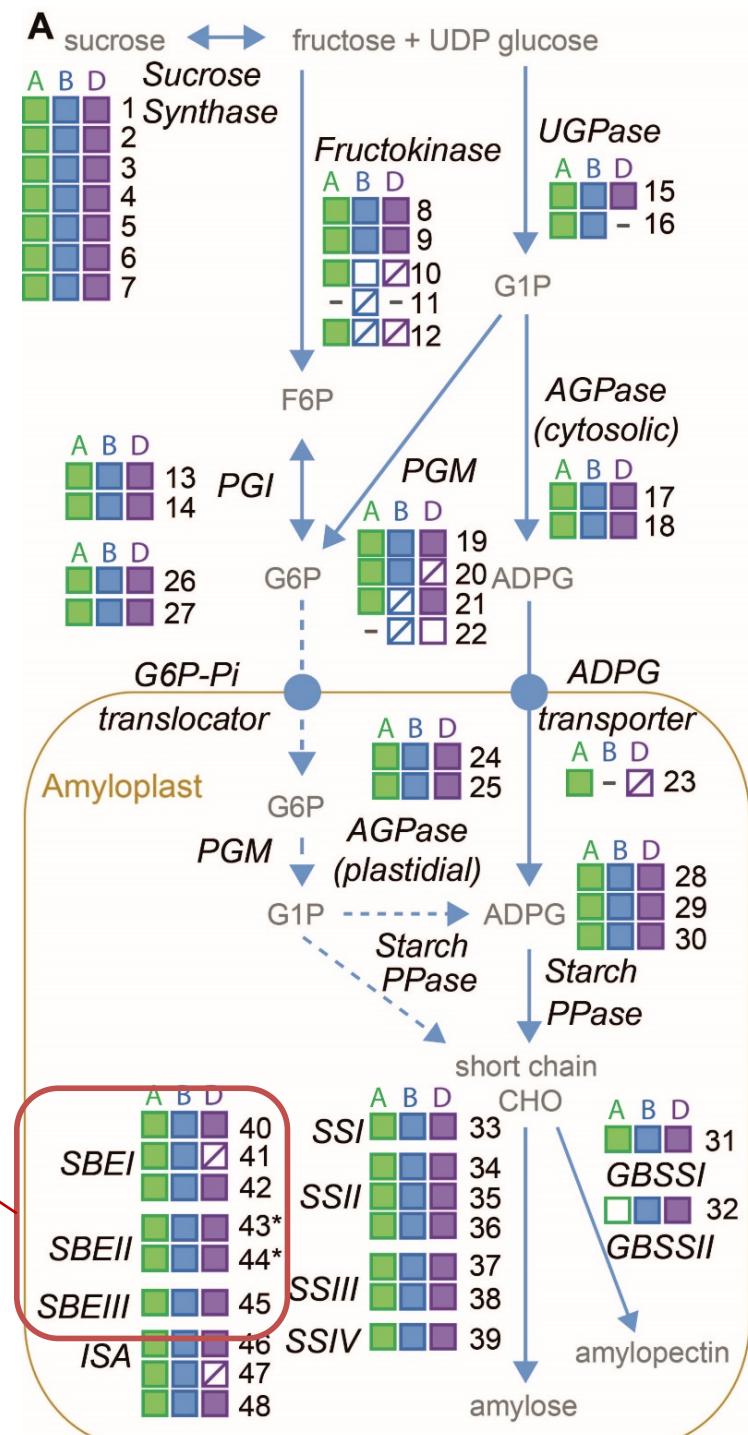
- Over 10M mutations in wheat
- Eliminate protein or alter function in >90% of wheat genes
- Open access searchable database
- Seeds free of IP

SeedStar



A	B	D	
40			
41			
42			
43*			
44*			
45			

Brittany Hazzard



Access to mutants



- Databases of mutations are available at
 - ✓ www.wheat-tilling.com
 - ✓ www.dubcovskylab.ucdavis.edu
 - ✓ <http://pre-test.plants.ensembl.org/index.html>
- We hold mirror collection of seeds at UC Davis and JIC Seed Store (seedstor.ac.uk)
- Mutants will be free from any IP for the mutations people find
 - £15 for non-commercial use
 - £200 for commercial use(freedom to operate)



<https://www.seedstor.ac.uk/>

SeedStor



The Cadenza TILLING Resource in the Field, May 2015.

Germplasm Resources Unit

..... a national capability supported by the BBSRC

Collection Recovery and MTA templates

MTA Name

MTA-Tilling-Research.rtf
MTA-Tilling-Commercial.rtf

MTA Method

Signed MTA before Dispatch
Signed MTA before Dispatch

Cost Description

Admin Fee
Maintenance Charge
Freedom to operate (Commercialisation Fee)
Non-EU clients: Phytosanitary certificate with no inspection or lab test

Amount

£10.00
£15.00
£200
£6.94

Method

Per Job
Per Line
Per Line
Per Job

Phytosanitary certification charge may be higher if additional tests are required to export to your country or if rates rise.

Shopping Cart

Cadenza and Kronos TILLING Resources

Please complete the fields below so that we can fulfil your request

Please enter the requested Lines

Please use the format Cadenza0000, Cadenza0001, ...

Cost Recovery

This request involves accessions that are Cost Recoverable.
An estimate of costs will be included when we acknowledge the job.

MTA for Material Use

Please pick the MTA you wish to operate under

Name

Organisation

Full Postal Address

Have you used wheat in your research before?

Yes No - First time user

Email

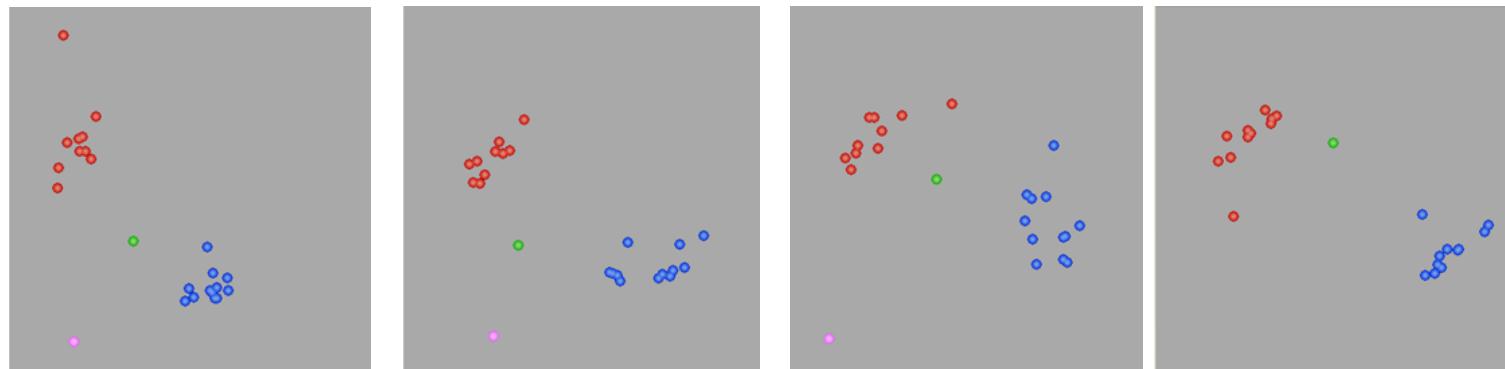
Standard phytosanitary also available. Useful for most countries

Validation of SNPs with KASP markers



PolyMarker assays worked first time for 85% of the SNPs.

	Kronos		Cadenza	
	No.	%	No.	%
Independent M ₄ families tested	67		19	
Valid KASP assays and Sanger sequence	133		147	
False positives	1	0.7%	1	0.7%
Mutations confirmed	132	99.2%	146	99.3%
Expected segregation (MAPS with correction ¹)	130	98.5%	139	95.2%
HOM mutation originally classified as HET	2	1.5%	2	1.4%
HET mutation originally classified as HOM ²	0	0.0%	5	3.4%



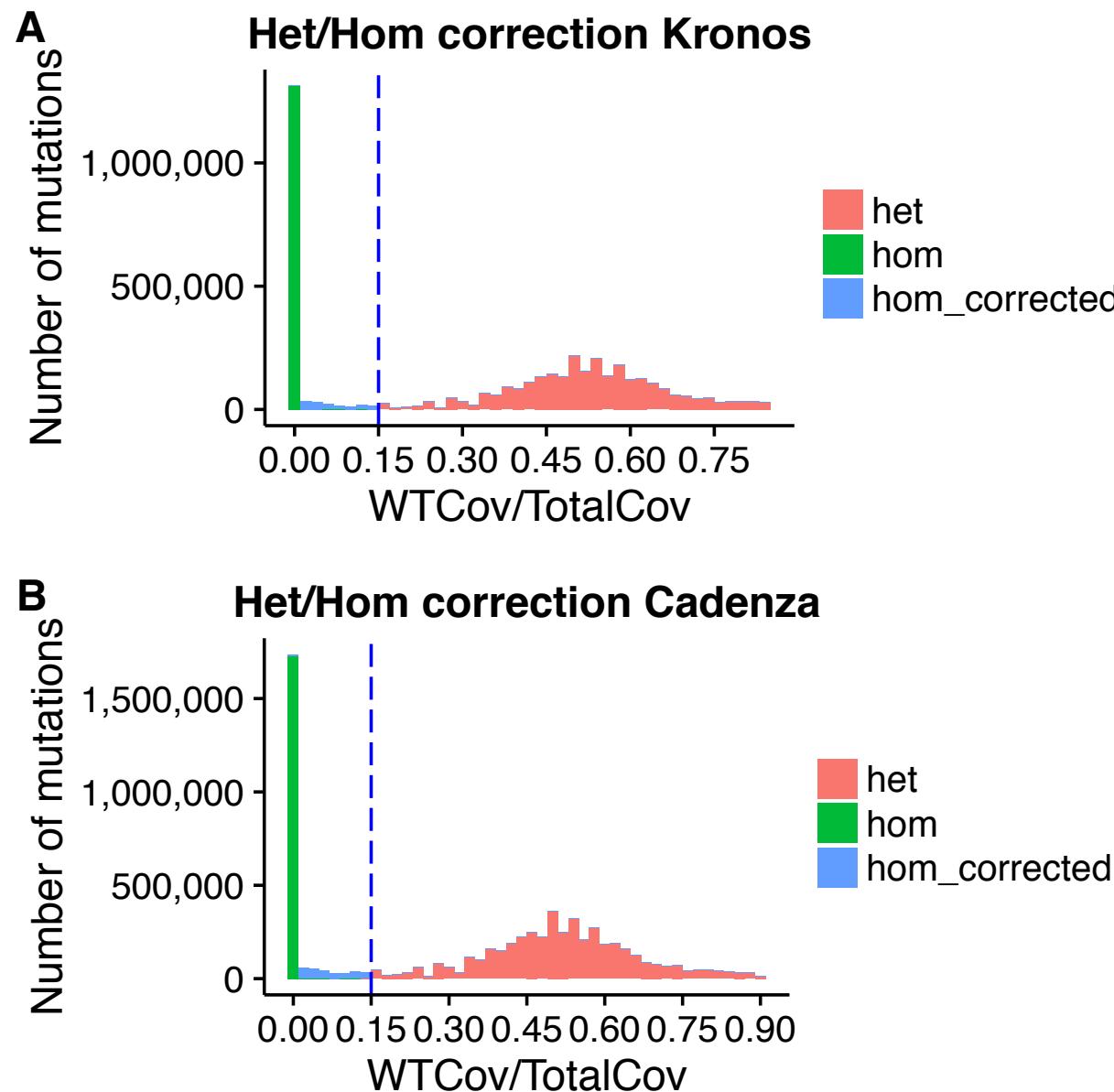
Potential issues

- Het:Hom calls
- Residual heterogeneity
- Blindspots
- Missing homoeologue
- Mis-annotation of gene model

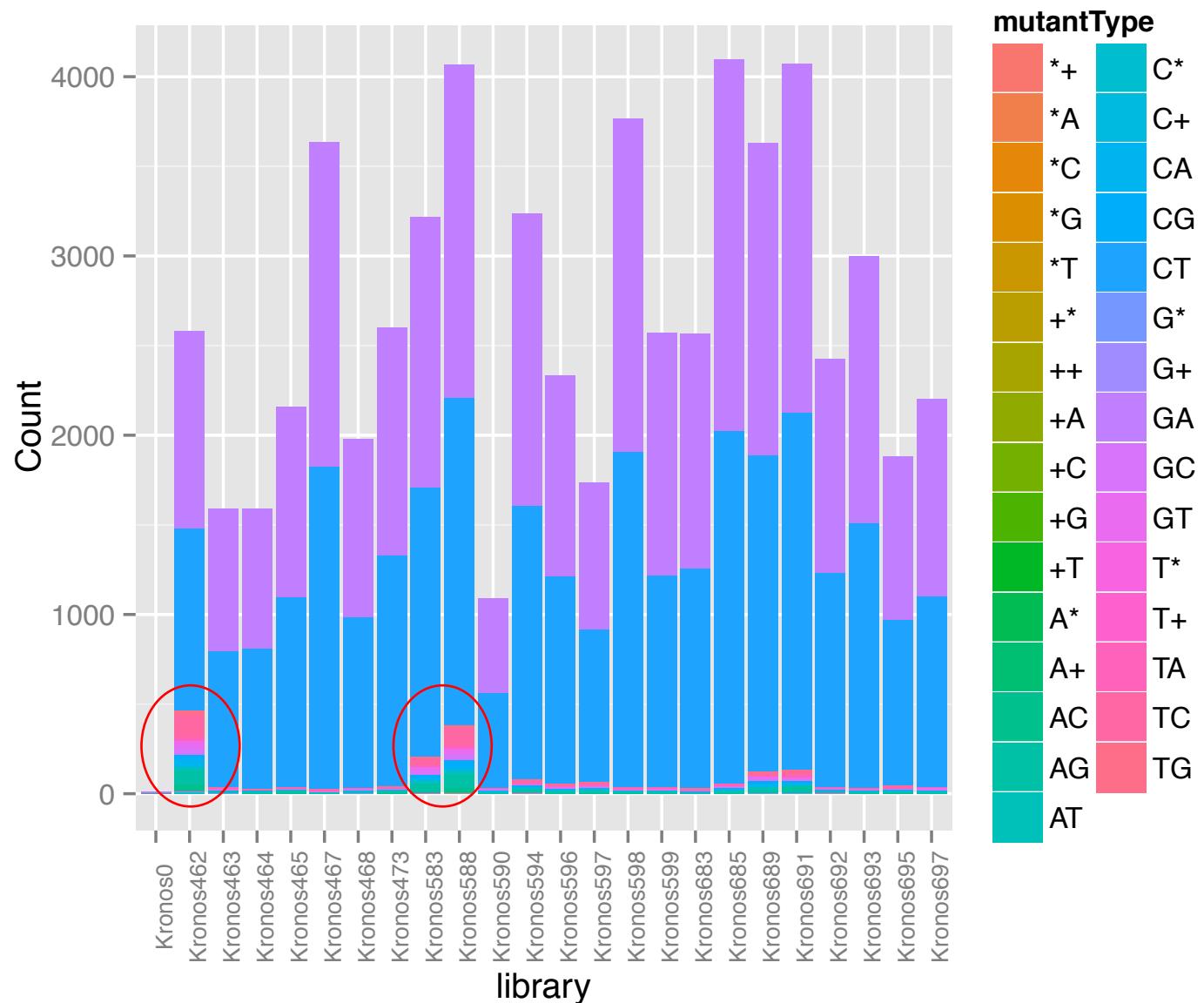


Wheat TILLING

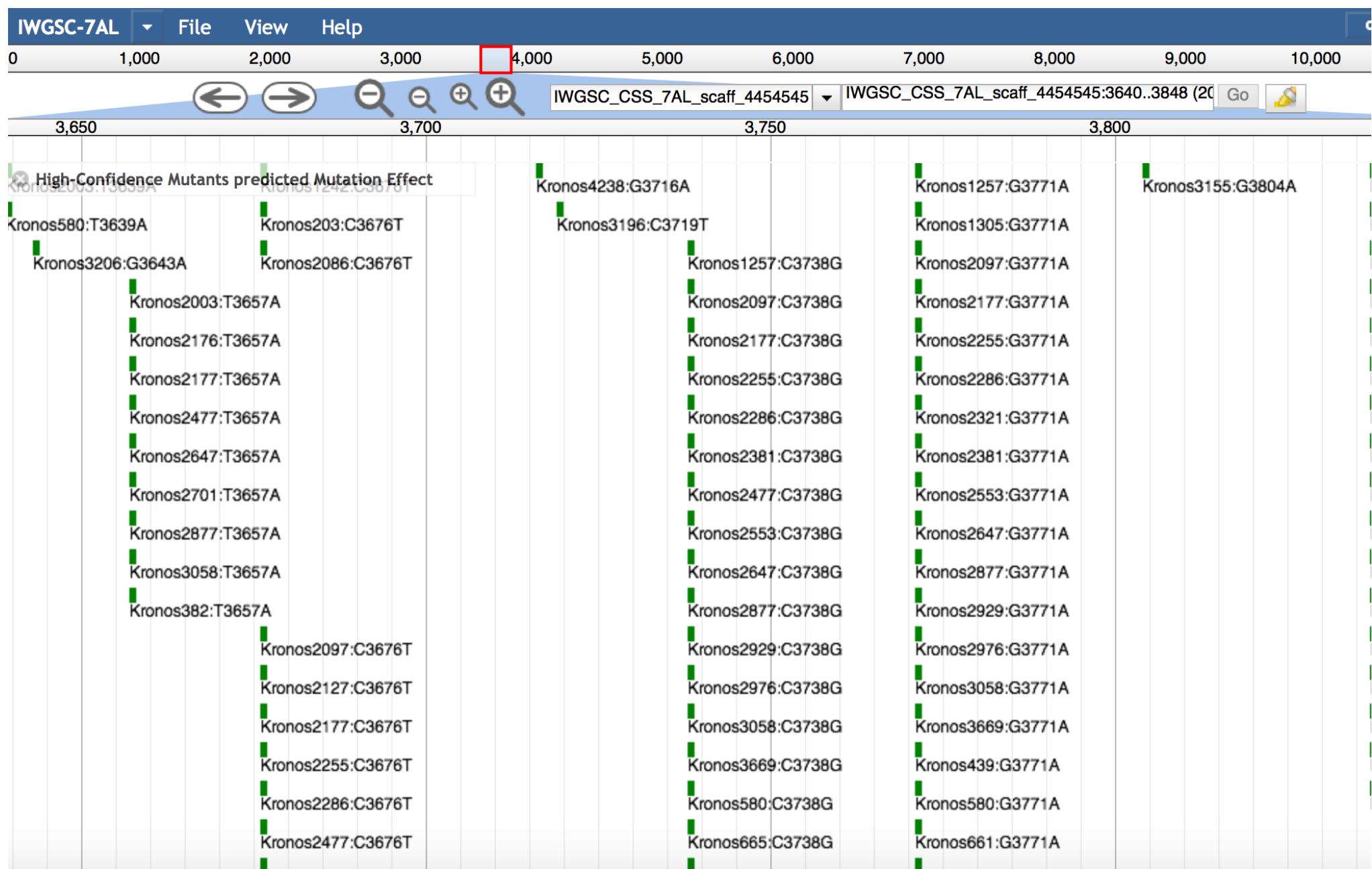
Het:Hom calls



In the full sight: We detect residual heterogeneity within populations

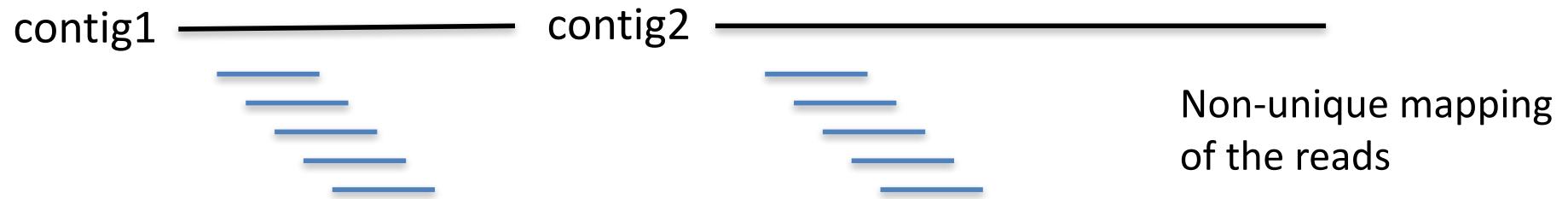


Residual heterogeneity in Kronos population



Blindspots: assembly redundancies and copy number variants

Contig1 is 100% identical to region in contig 2



Mapping quality = 0

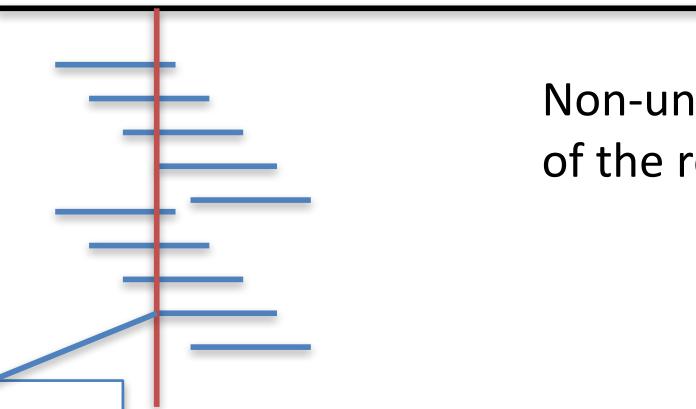
Our threshold for calling mutations is 20

Current solution: masking

Contig1 is 100% identical to region in contig 2

contig1 XXXXXXXXXXXXXXX

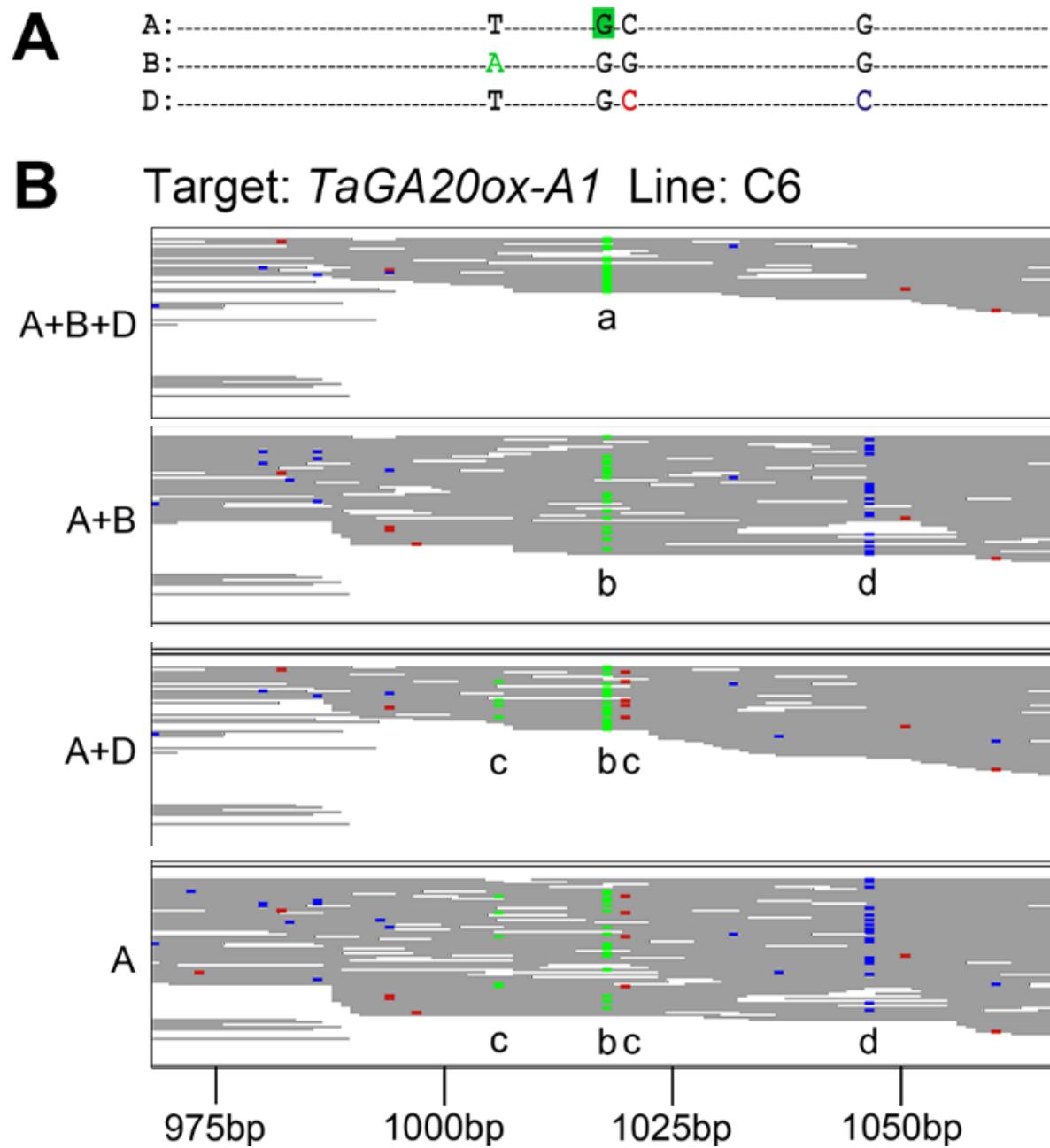
contig2



Non-unique mapping
of the reads

This mutation can also
be at contig1

Missing Homoeologue

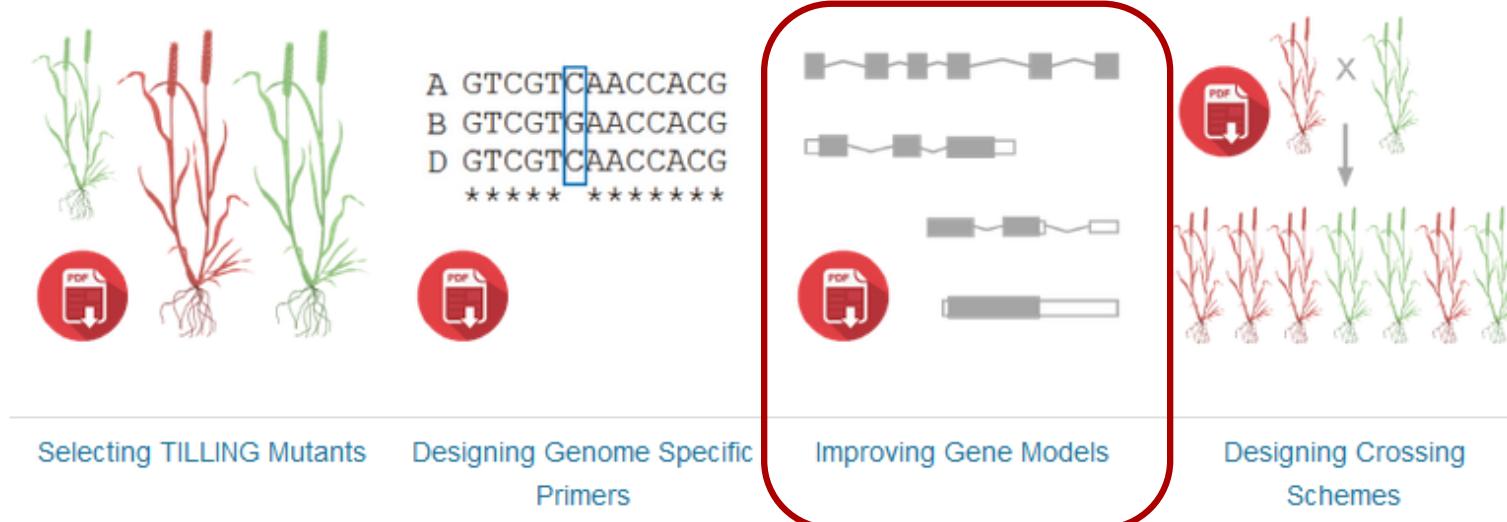


Mis-mapped reads from D dilute homozygous mutation in A

- Two possible mistakes:
 - HET v HOM
 - Incorrect genome

Mis-annotation of gene model

- Predictions on putative effects are based on gene model.
- If gene model missing: mutations called as inter-genic variants
- If gene model is incorrect: our variant effect prediction will also be incorrect.



<http://www.wheat-training.com/wp-content/uploads/TILLING/pdfs/Improving-Gene-models.pdf>

Selecting mutations to take forward

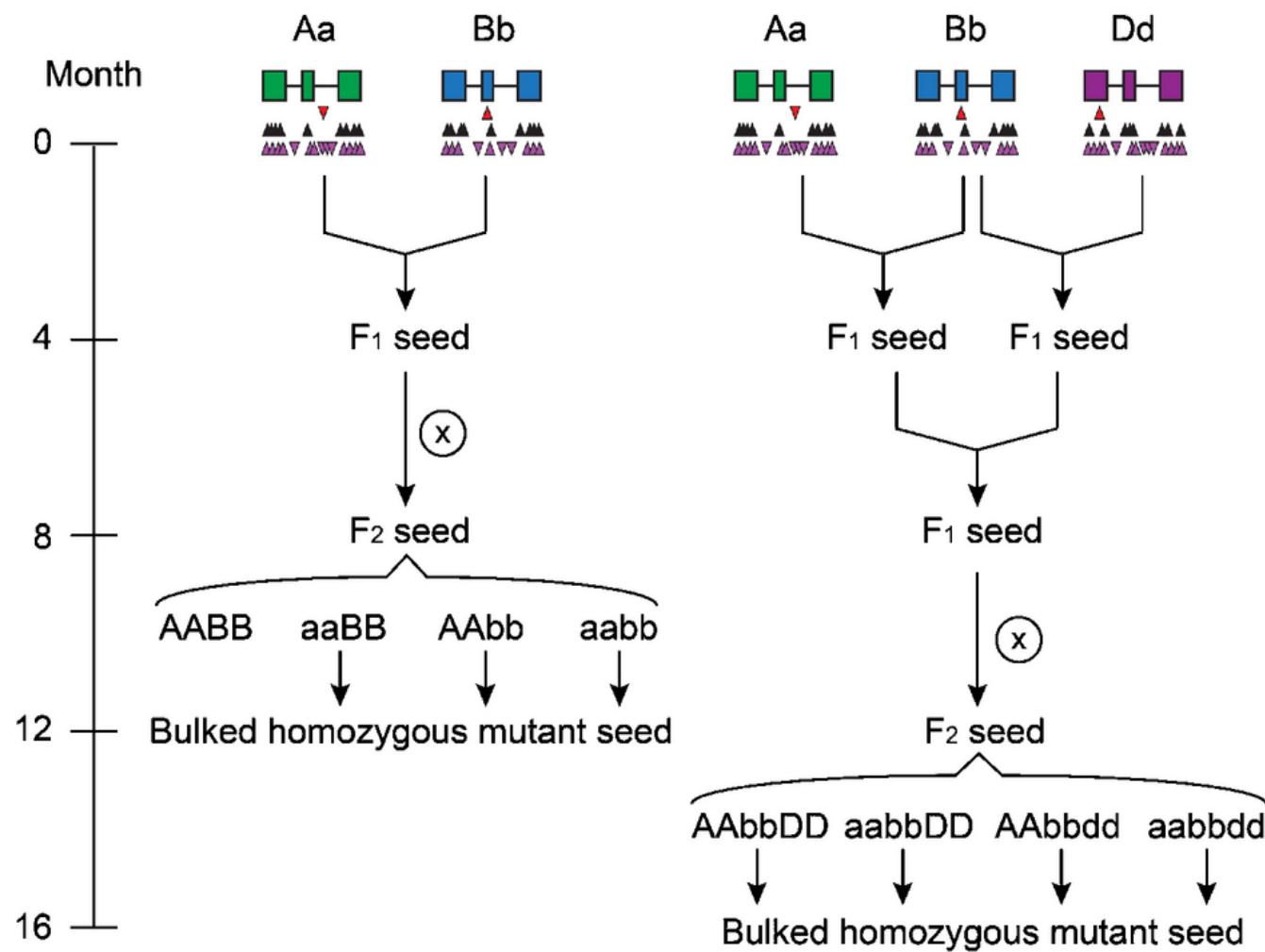
- Type of mutation (truncation, position, SIFT (!))
- Zygosity
- Population (4x vs 6x)
- Evidence (coverage)



<http://www.wheat-training.com/wp-content/uploads/TILLING/pdfs>Selecting-TILLING-mutants.pdf>

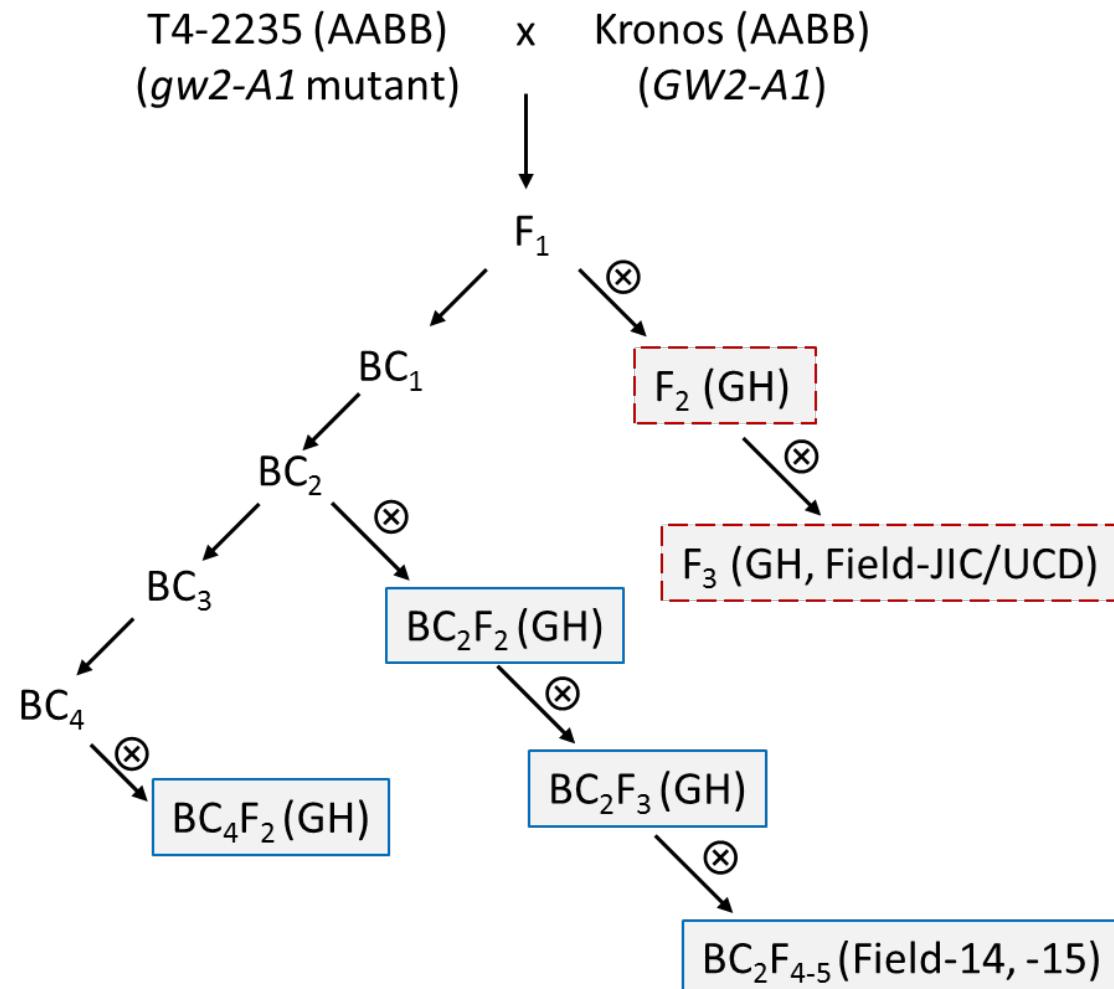
Combining mutations

- Simple scheme



Combining mutations

- More complex phenotype (backcrossing to dilute ~500,000 background mutations)



Combining mutations

- More details online



Selecting TILLING Mutants

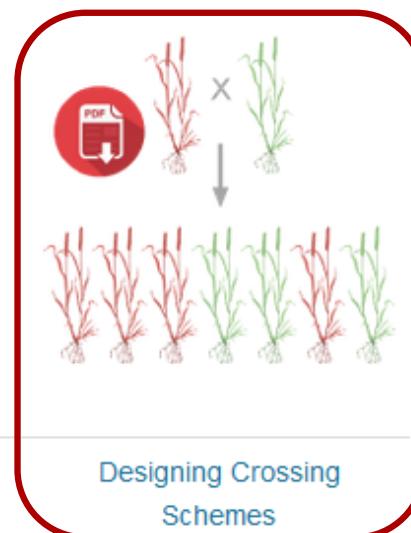
A GTCGTCAACCACG
B GTCGTGAACCACG
D GTCGTCAACCACG
***** * *****



Designing Genome Specific
Primers



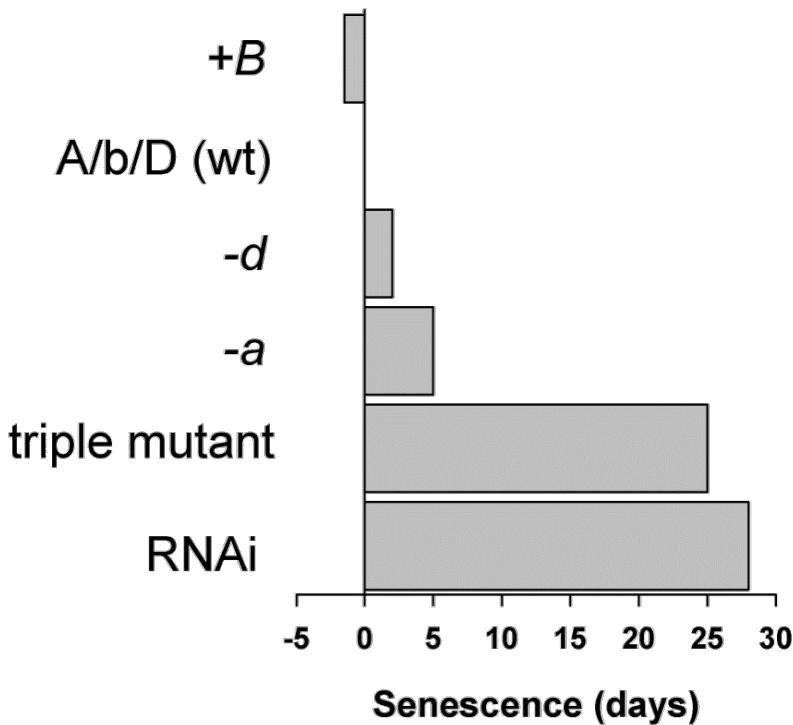
Improving Gene Models



Designing Crossing
Schemes

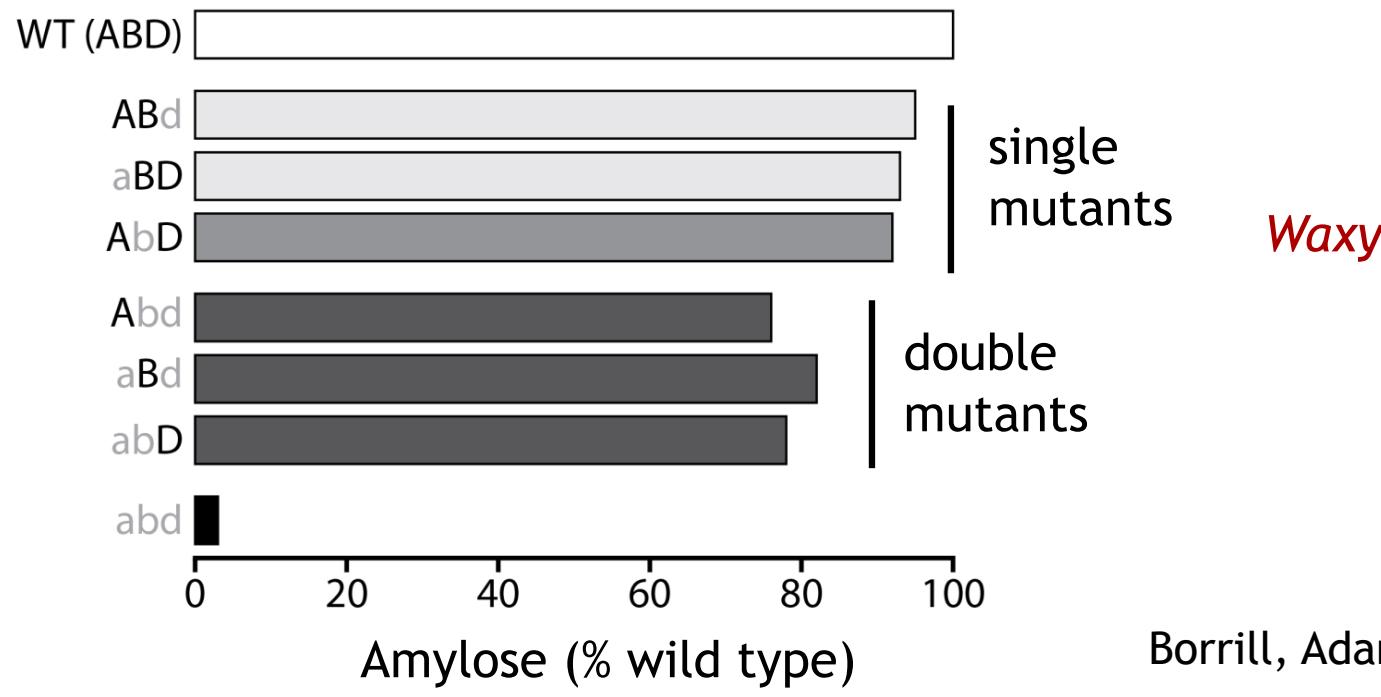
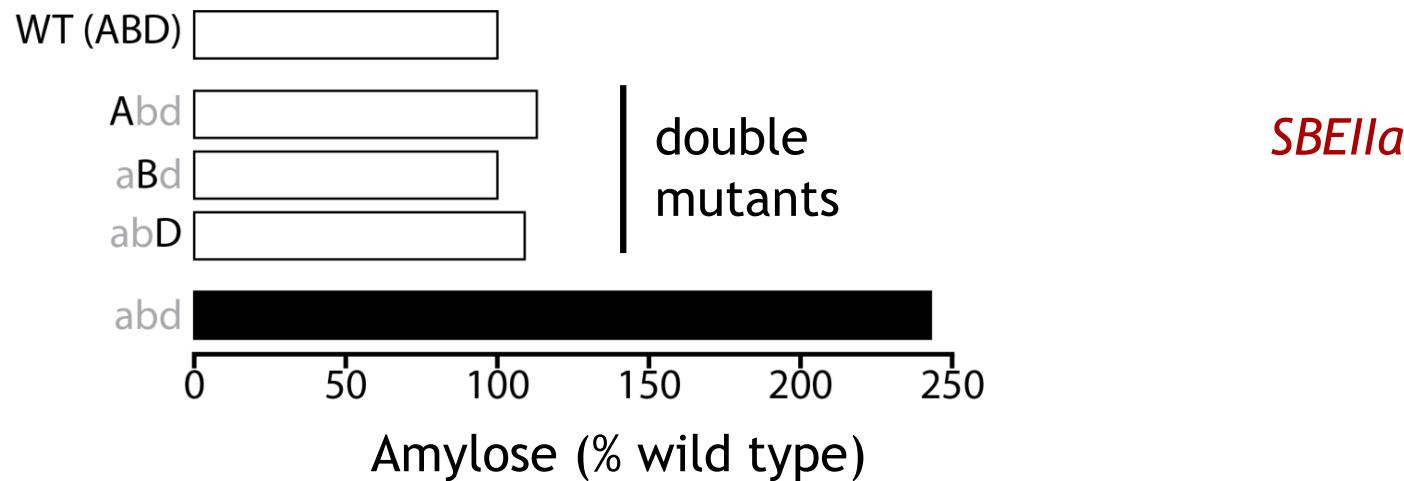
<http://www.wheat-training.com/wp-content/uploads/TILLING/pdfs/Designing-crossing-schemes.pdf>

Dosage of quantitative trait: Grain Protein Content (GPC) example



Uauy et al Science 2006; Avni et al Planta 2014; Uauy/Borrill *unpub.*

Dosage of quantitative trait: amylose content



Borrill, Adamski & Uauy (2015)
New Phytologist

Dominant gain-of-function mutations: disease resistance



Mutant
Resistant



WT
Partially
Resistant



Susceptible

Raats/Hegarty, Deatker, Schudoma, Dubcovsky and Krasileva, unpublished

The team

Krasileva Group (EI/TSL)

Paul Bailey (EI)

Dina Raats (EI)

Christian Schudoma (EI)

Francesca Stefanato (TSL)

Andrew Deatker (TSL)

EI

Platforms and pipelines

High Performance Computing

Leah Clissold

Christine Fosker

Sarah Ayling

John Innes Centre

Cristobal Uauy

James Simmonds

Philippa Borrill

Ricardo Ramirez-Gonzalez

Rothamsted

Andy Phillips

Robert King

EBI

Paul Kersey

Dan Bolser

Guy Naamati

International
Wheat
Genome
Consortium

Our funding



United States Department of Agriculture
National Institute of Food and Agriculture

