Low Coverage Sequencing of MAGIC

Richard Mott, Martha Imprialou

Genotyping By Sequencing

- High coverage reduced representation
  - RAD restriction digest
  - RNAseq

OPEN ACCESS Freely available online

PLoS one

Linkage Mapping and Comparative Genomics Using Next-Generation RAD Sequencing of a Non-Model Organism

Simon W. Baxter[1*☯], John W. Davey[2☯], J. Spencer Johnston[3], Anthony M. Shelton[4], David G. Heckel[5], Chris D. Jiggins[1], Mark L. Blaxter[2,6]

Genotyping by Low-coverage Sequencing

ARTICLES

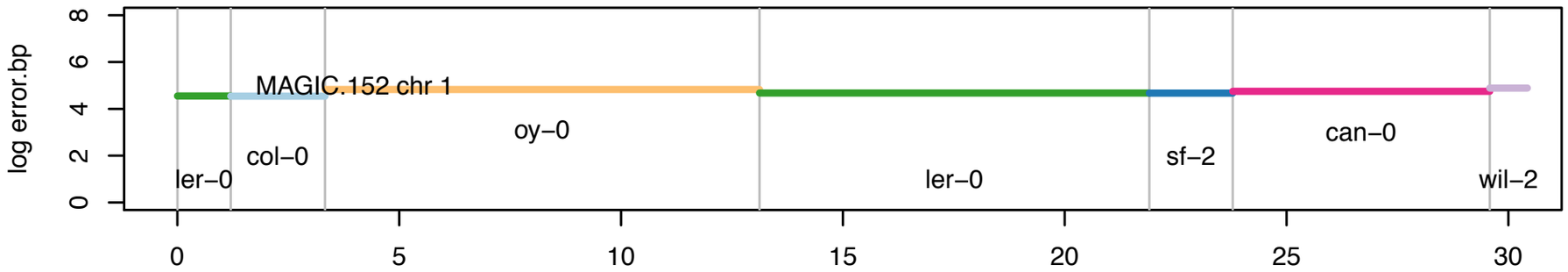
Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm

Xuehui Huang[1,2,5], Yan Zhao[1,2,5], Xinghua Wei[3,5], Canyang Li[1], Ahong Wang[1], Qiang Zhao[1], Wenjun Li[1], Yunli Guo[1], Liuwei Deng[1], Chuanrang Zhu[1], Danlin Fan[1], Yiqi Lu[1], Qijun Weng[1], Kunyan Liu[1], Taoying Zhou[1], Yufeng Jing[1], Lizhen Si[1], Guojun Dong[1,3], Tao Huang[1], Tingting Lu[1], Qi Feng[1], Qian Qian[3], Jiayang Li[4] & Bin Han[1,2]

Nature Genetics 2011

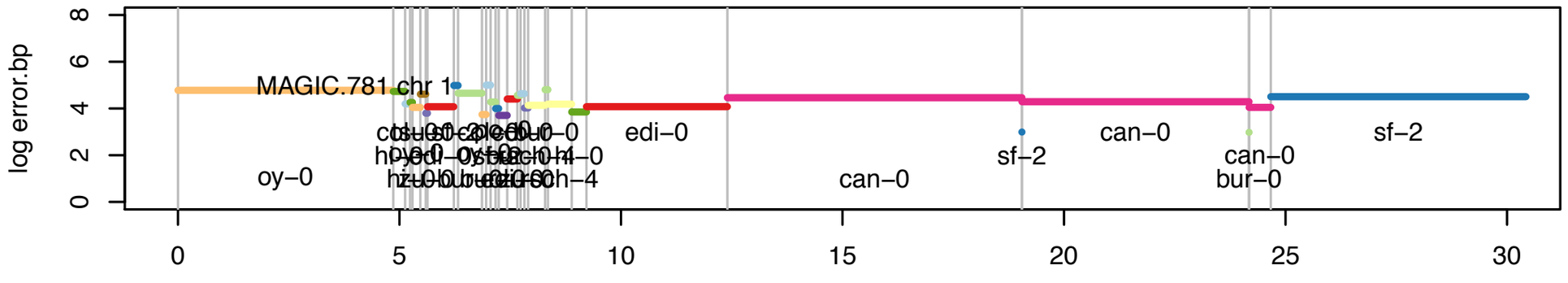
Genetic mosaics

- 476 MAGIC lines sequenced at $\sim$ 0.25x coverage
- Inferred genome mosaics of MAGIC lines from sequence data

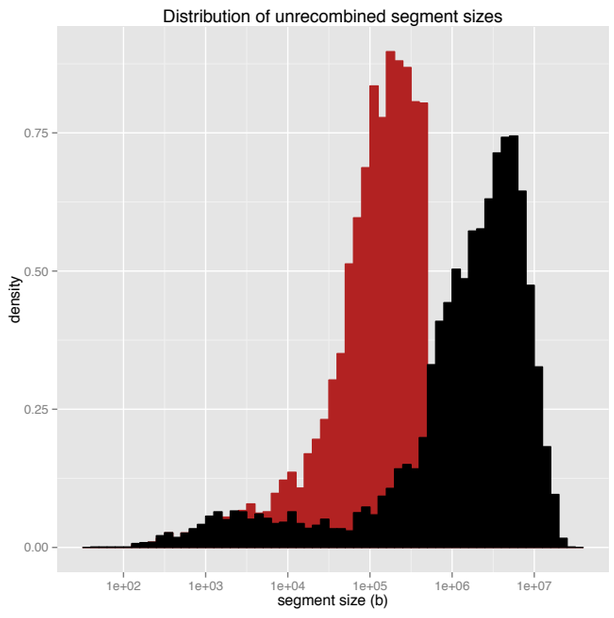


- 17,888 mosaic breakpoints in total
- Average segment size $\simeq$ 3.5Mb

Clusters of breakpoints

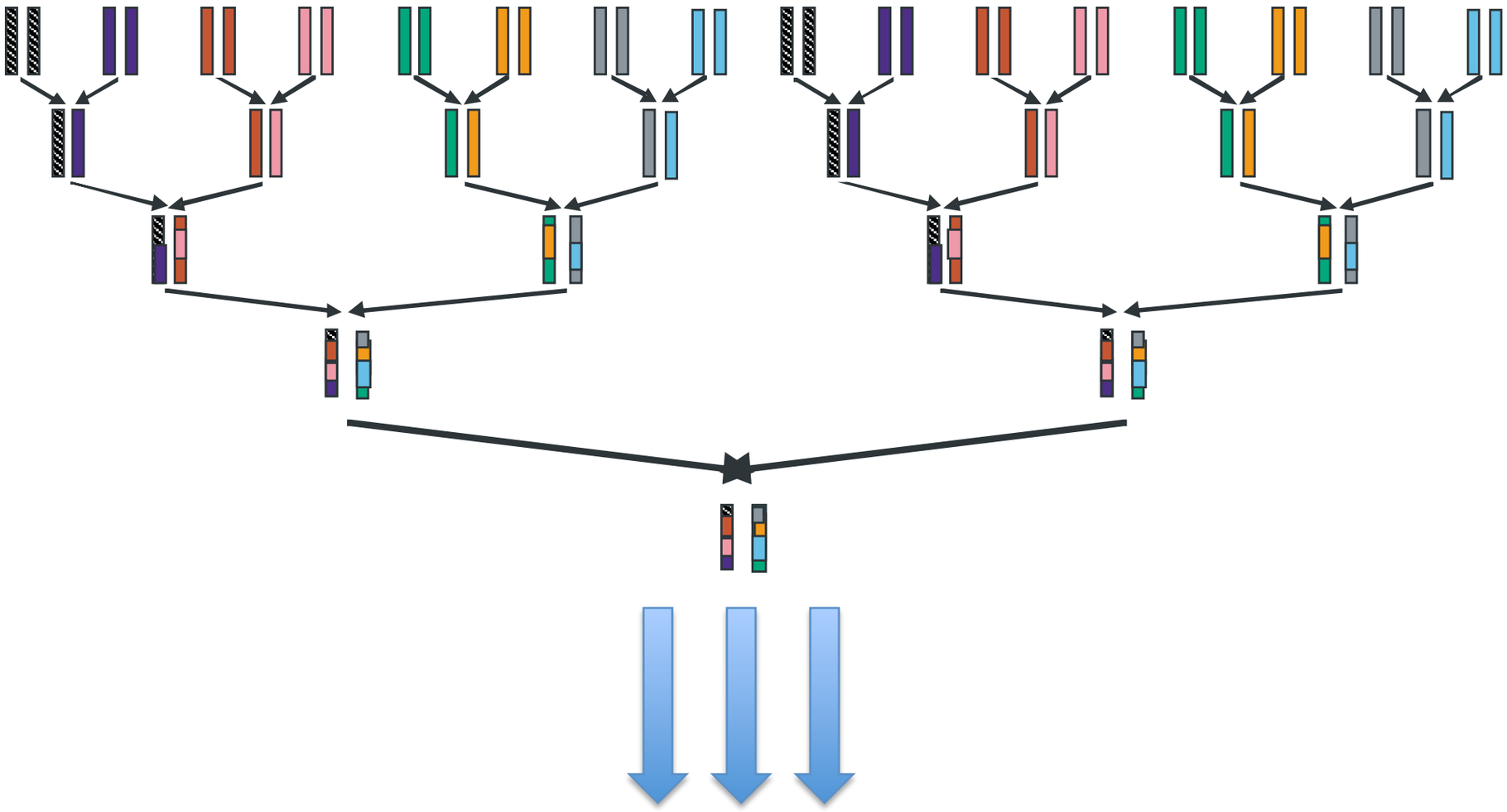


- Clusters in 76.5% of lines, involving 28.8% of breakpoints
- Average segment size $\simeq$ 3.5Mb, but in clusters $\simeq$ 200kb

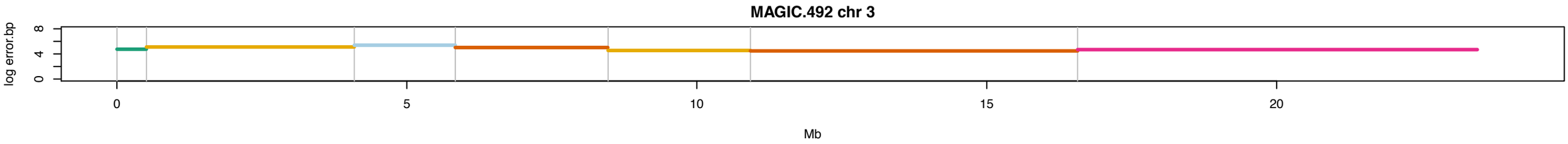


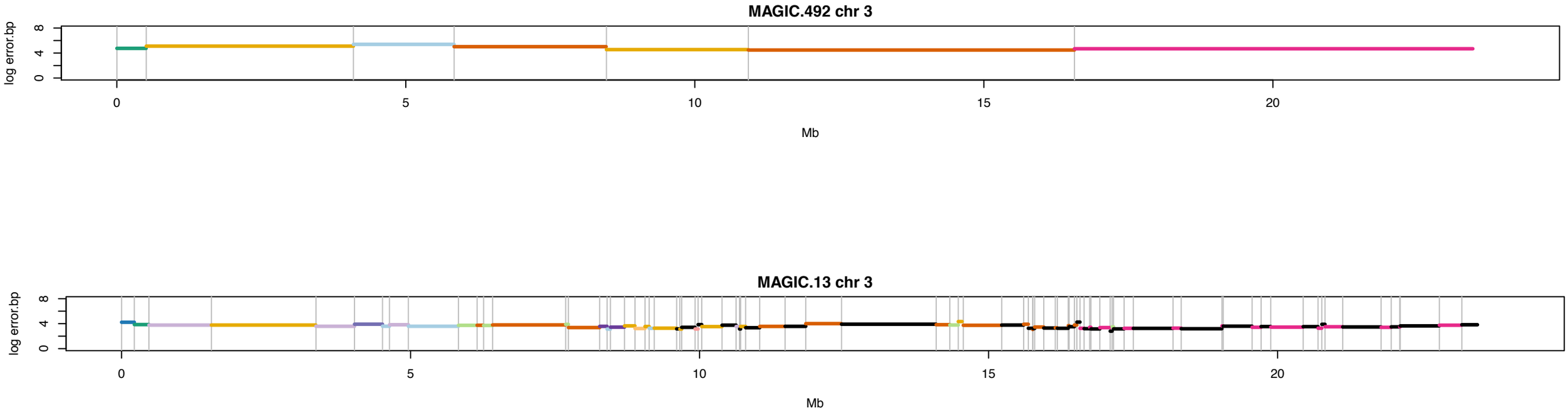
What are the clusters?

- Locally very high recombination rates in Mb-sized regions ?
- Gene conversion clusters ?
- Chromothripsis ?
- Artefacts ?
  - Introgression of novel genomes (novel sequence variants)
  - Translocations (read mapping errors, split contigs)
  - Heterozygosity
  - Repeats, transposons (read coverage)
- Resequenced 9 lines at 20x coverage, 100 bp paired-end reads.

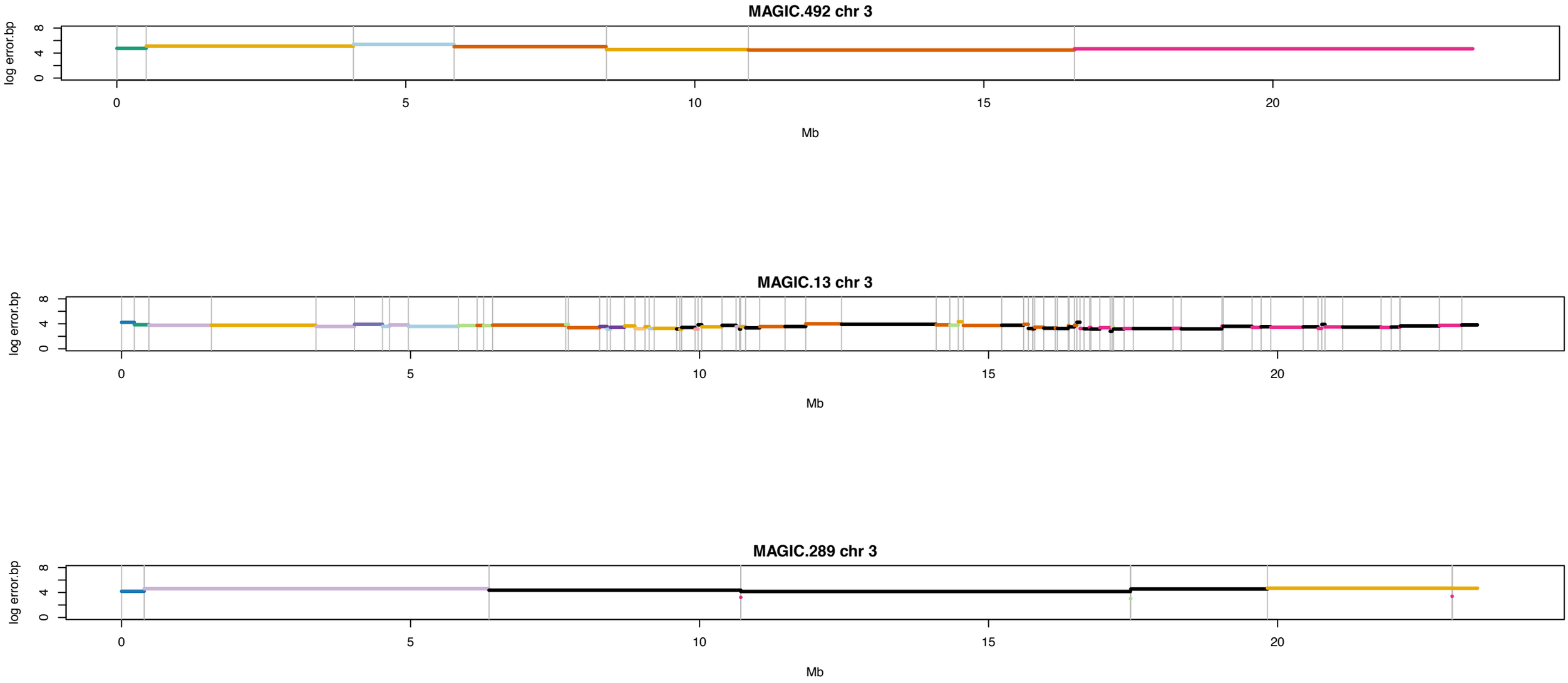


Cousin Lines
~25% similar at haplotype level


MAGIC.492 chr 3
log error.bp
0
4
8
0
5
10
15
20
Mb

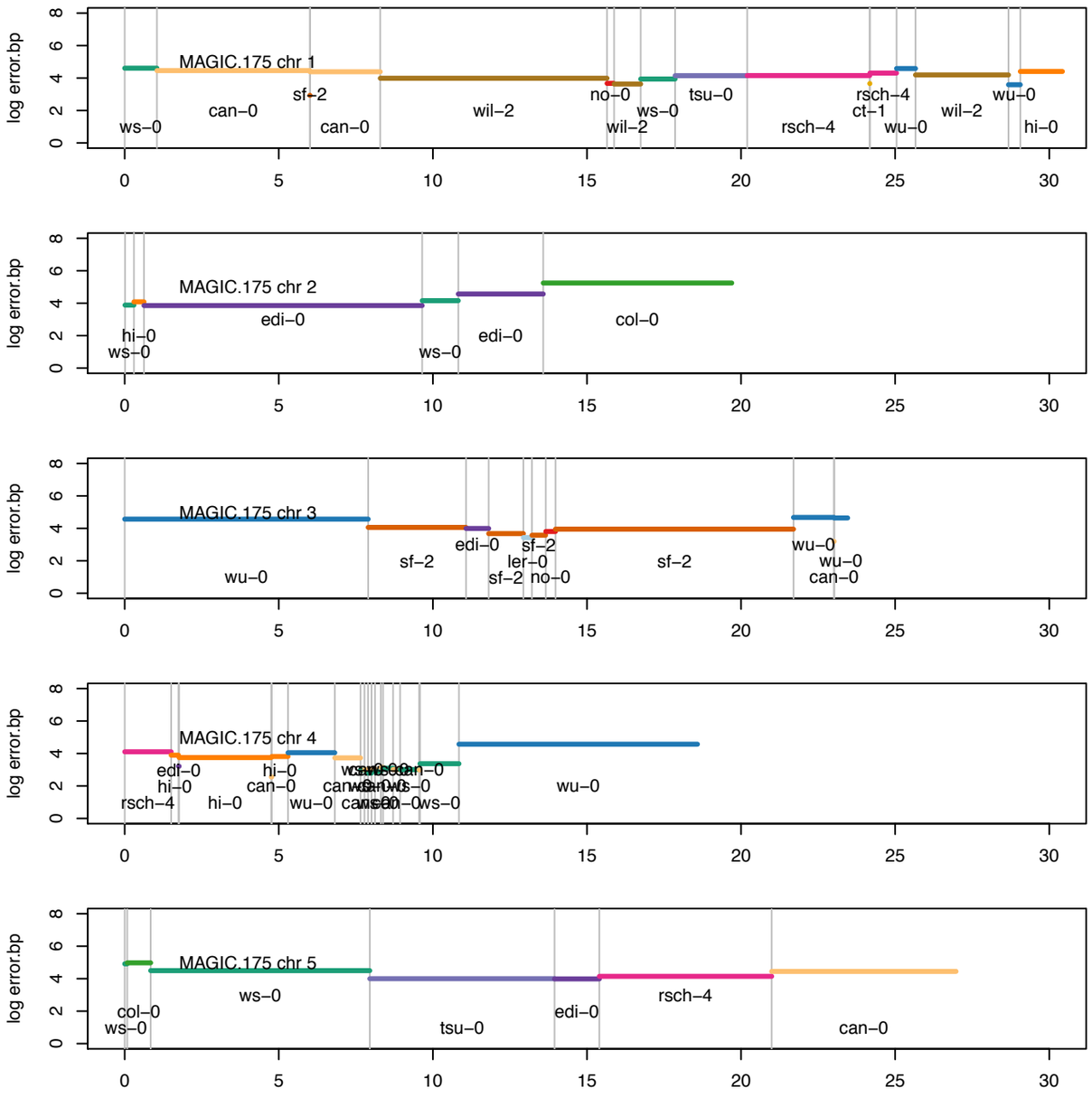

MAGIC.492 chr 3
log error.bp
0
4
8
0
5
10
15
20
Mb

MAGIC.13 chr 3
log error.bp
0
4
8
0
5
10
15
20
Mb

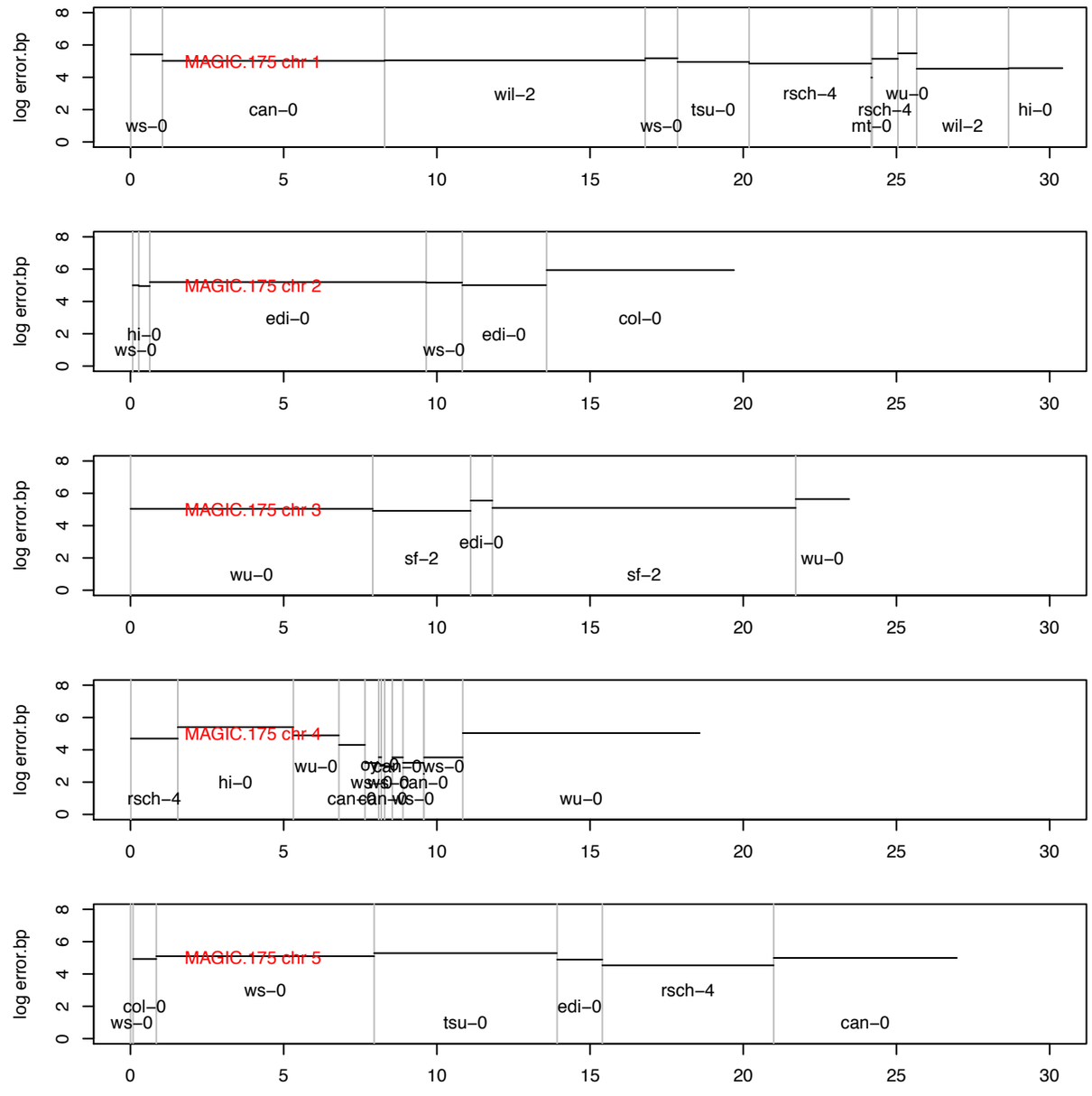

MAGIC.492 chr 3
log error.bp
0
4
8
0
5
10
15
20
Mb
MAGIC.13 chr 3
log error.bp
0
4
8
0
5
10
15
20
Mb
MAGIC.289 chr 3
log error.bp
0
4
8
0
5
10
15
20
Mb

Mosaics Reproduced by RNAseq

low-coverage genomic DNA (MAGIC.175)

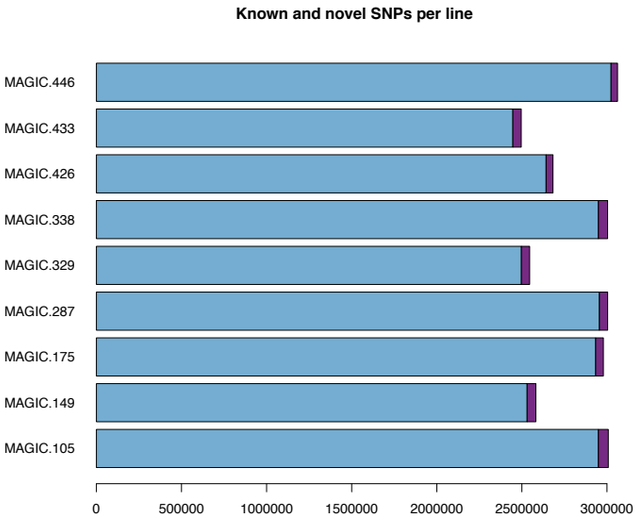


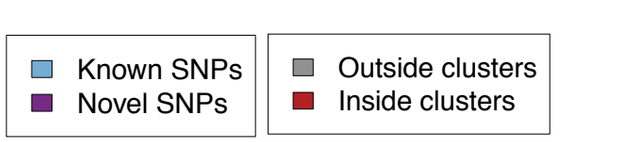
RNAseq (MAGIC.175 - different plant)

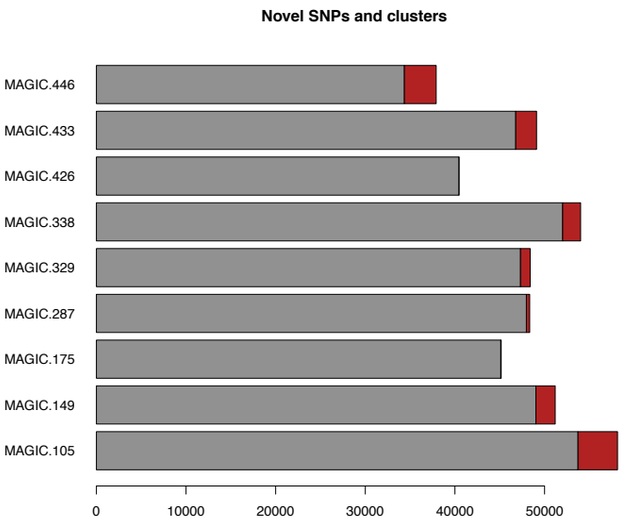


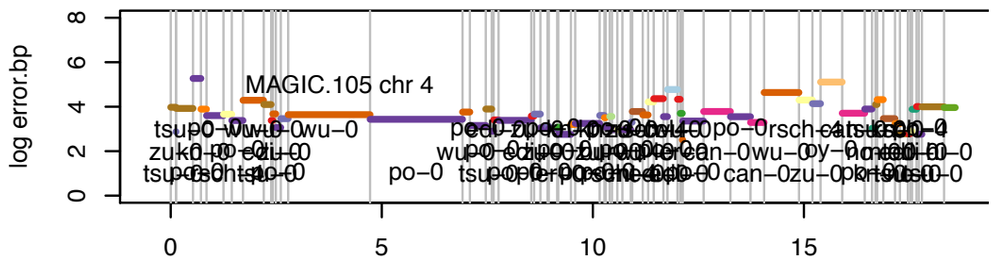
RNA from Richard Clark

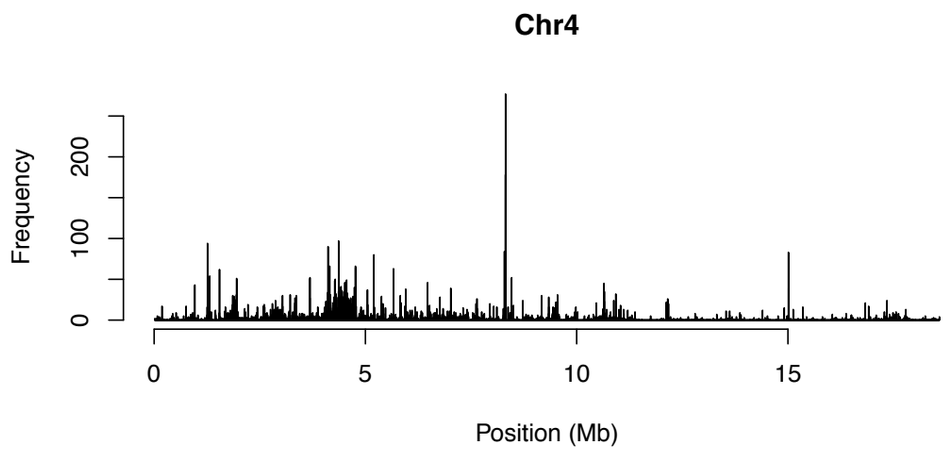
1% of SNPs are novel, 3% of novel SNPs in clusters





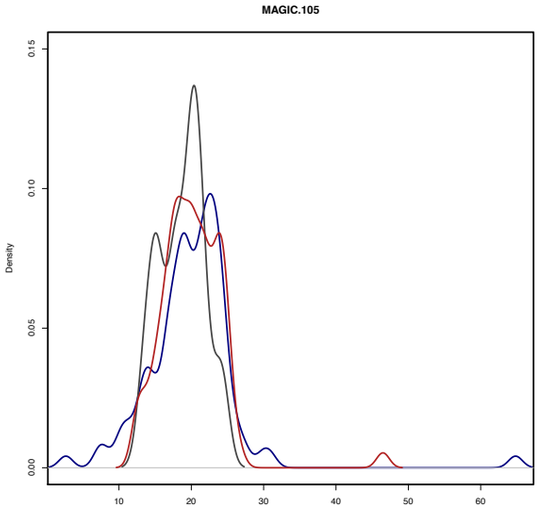




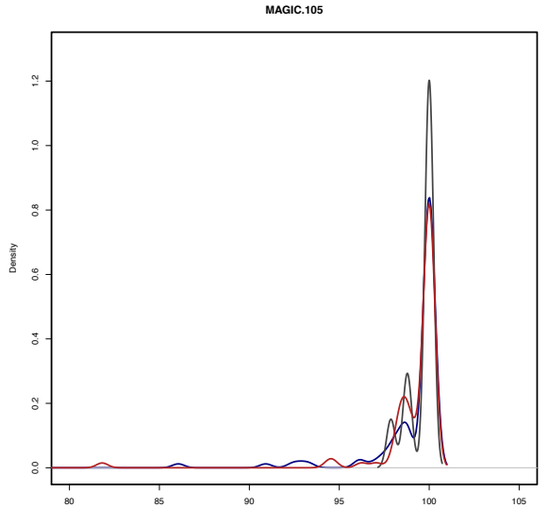


Cluster regions resemble non-cluster regions

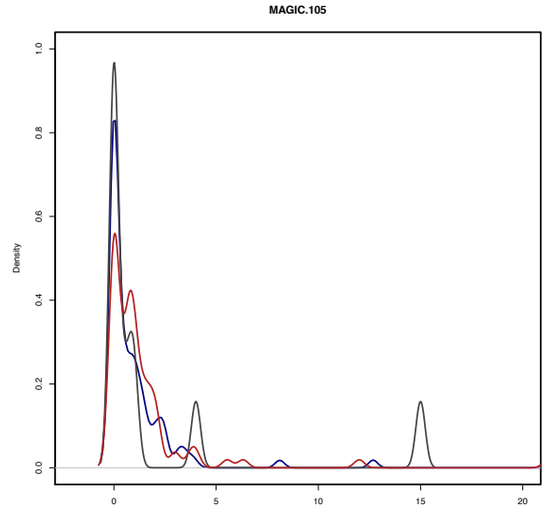
Mean coverage



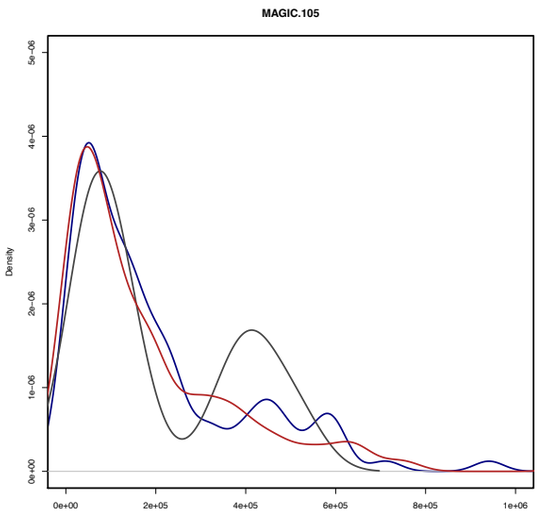
Properly paired reads



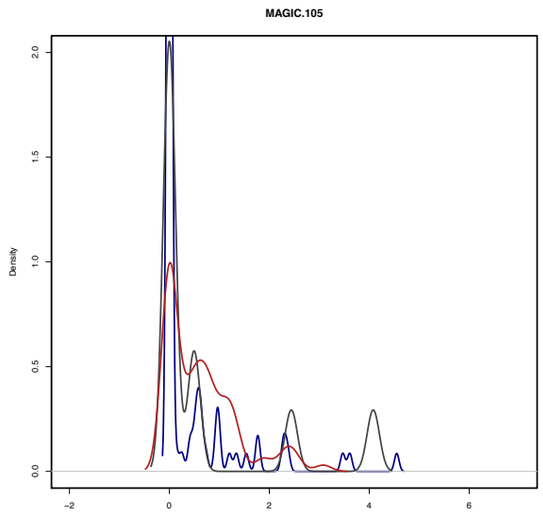
Read pairs on same strand

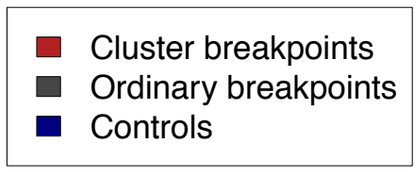


Nearest split contig



Heterozygosity

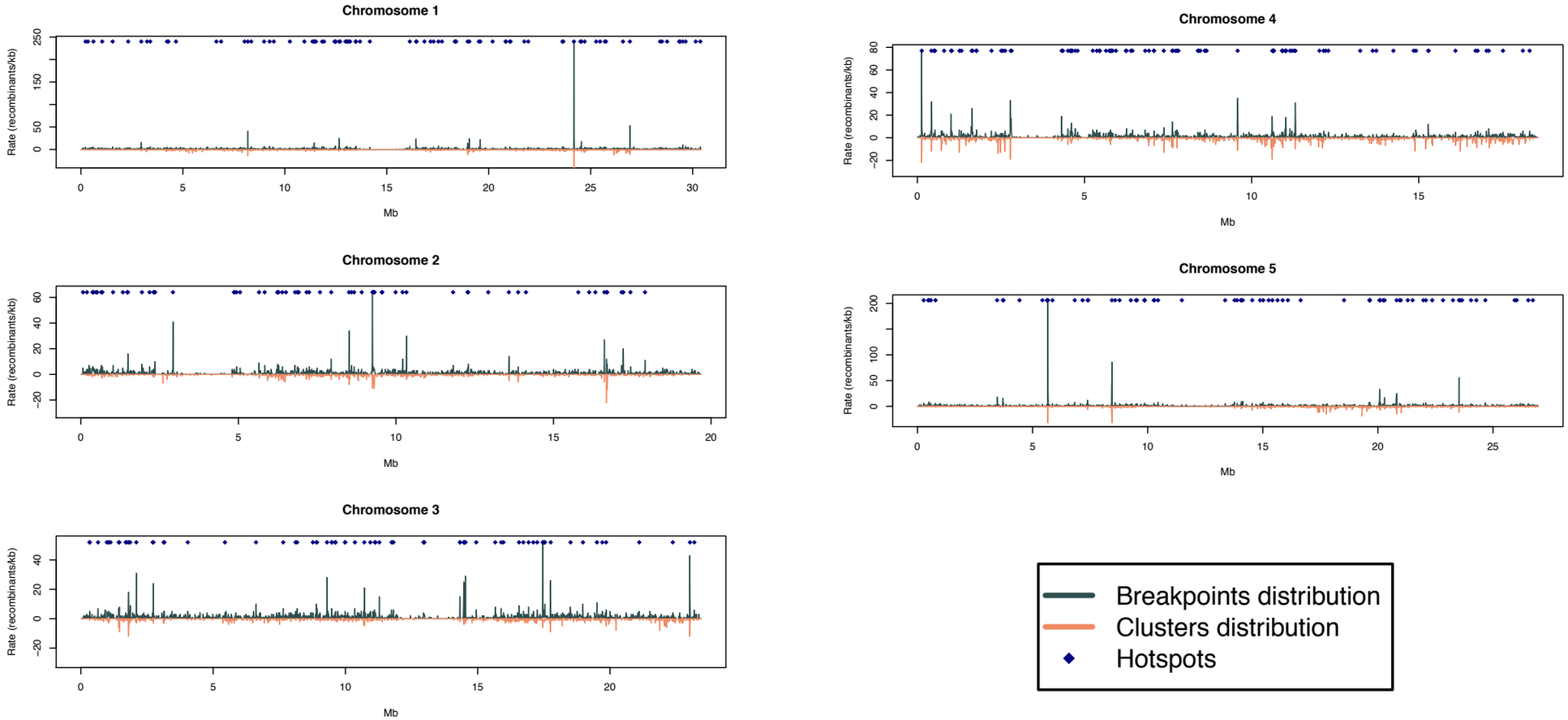




- >75% of breakpoints clusters have no abnormalities.
  - Remaining abnormalities are probably rearrangements.

Recombination hotspots

399 hotspots from 12,000 MAGIC breakpoints excluding clusters



Cluster breakpoints are half as likely to lie in hotspots as other breakpoints.

Conclusions

- Overall, it is likely that the clusters are recombination-related.
  - They are possibly recombinations or gene conversions.
  - About 35% are potentially gene conversions.
- Locally very high breakpoint rates in Mb-sized regions.
- The occurrence of clusters is lineage-specific.
- Clusters breakpoints distinct from recombination hotspots
  - Different mechanism ?

Could clusters improve mapping resolution?

- Suppose a QTL has been mapped in MAGIC to a 300kb locus 300kb
  - contains ~50 genes
- Cloning the gene(s) is still too slow
- Need more recombinants
- In 700 Arabidopsis MAGIC lines most of the genome covered by clusters
  - Too few lines to fine-map a locus by focussing on the clusters

- BUT - if we make and low-coverage sequence thousands of MAGIC lines
- THEN we phenotype those lines with clusters at the QTL to fine-map

Acknowledgements

- BBSRC

- Martha Imprialou
- Emma Huang
- Xiangchao Gan
- Paula Kover
- Richard Clark
- Gunnar Raetsch