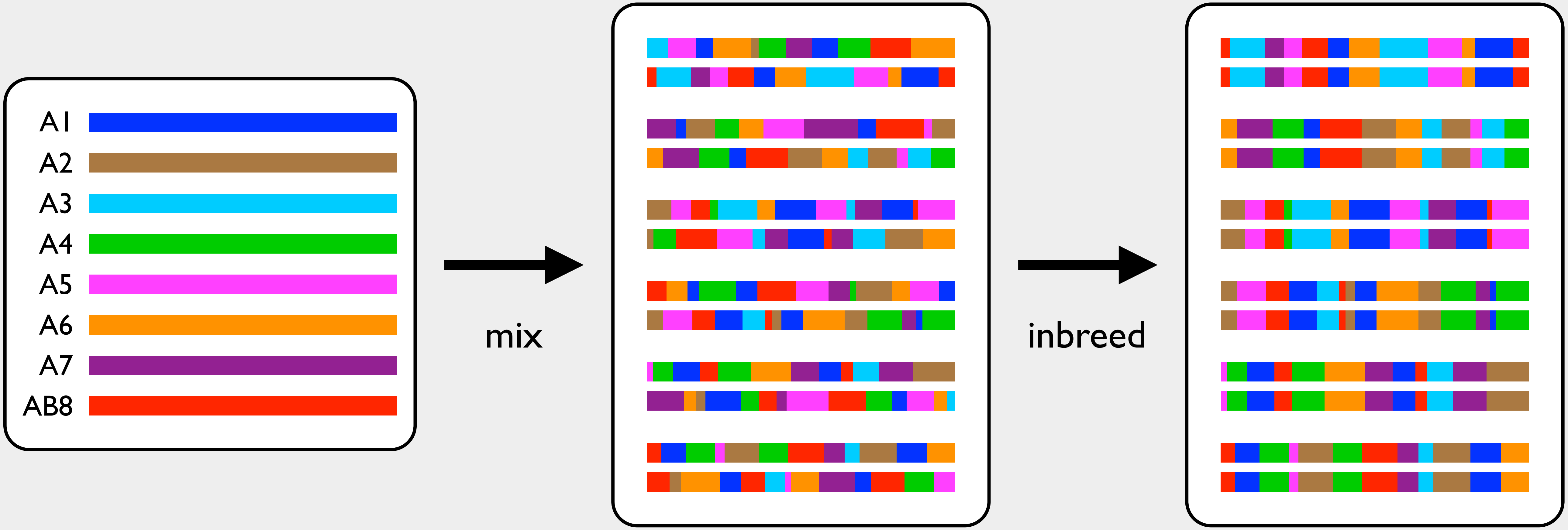
The *Drosophila* Synthetic Population Resource

Stuart J Macdonald

Department of Molecular Biosciences

THE UNIVERSITY OF KANSAS

Drosophila **S**ynthetic **P**opulation **R**esource



F_0

Founders

F_{50}

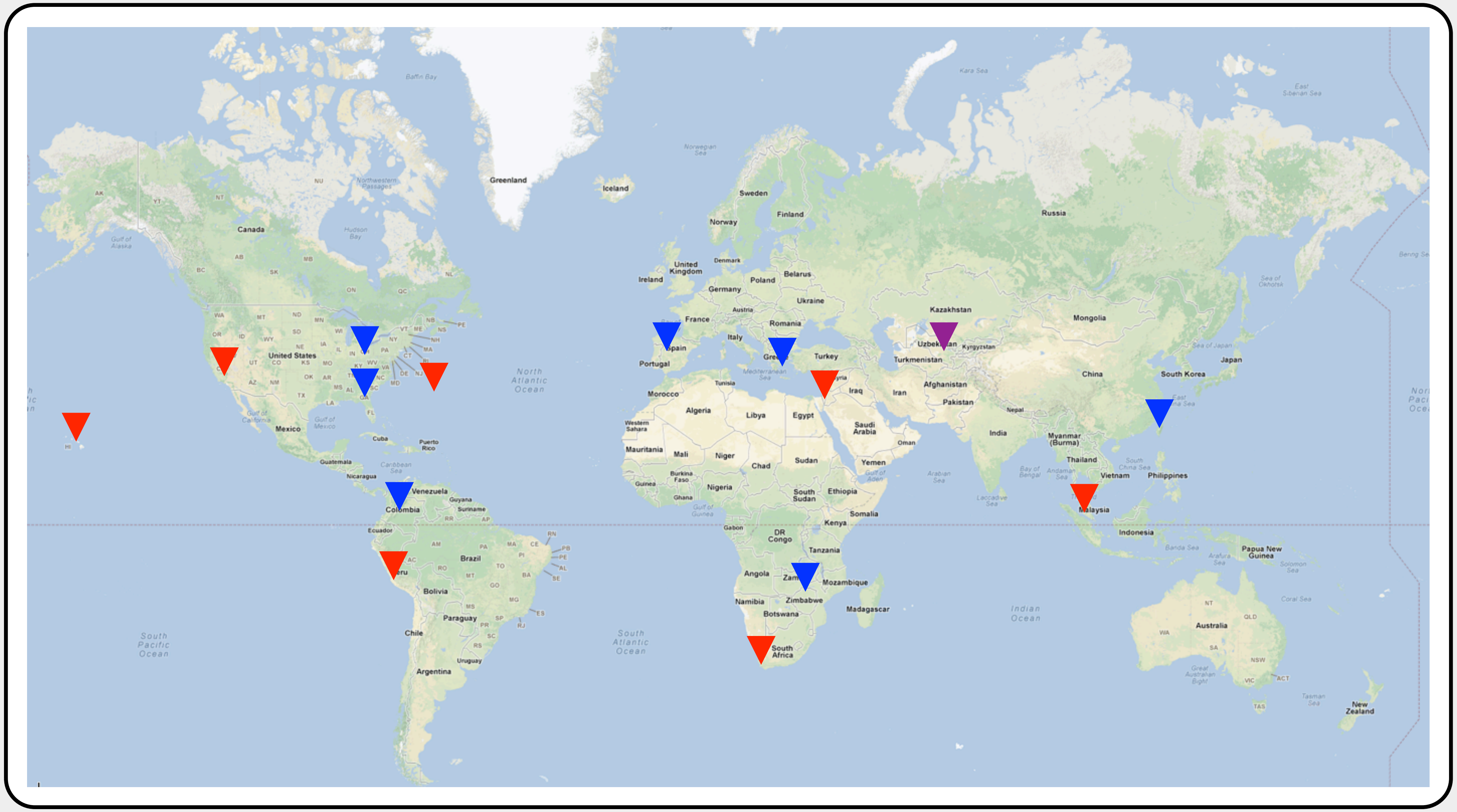
Synthetic recombinant population

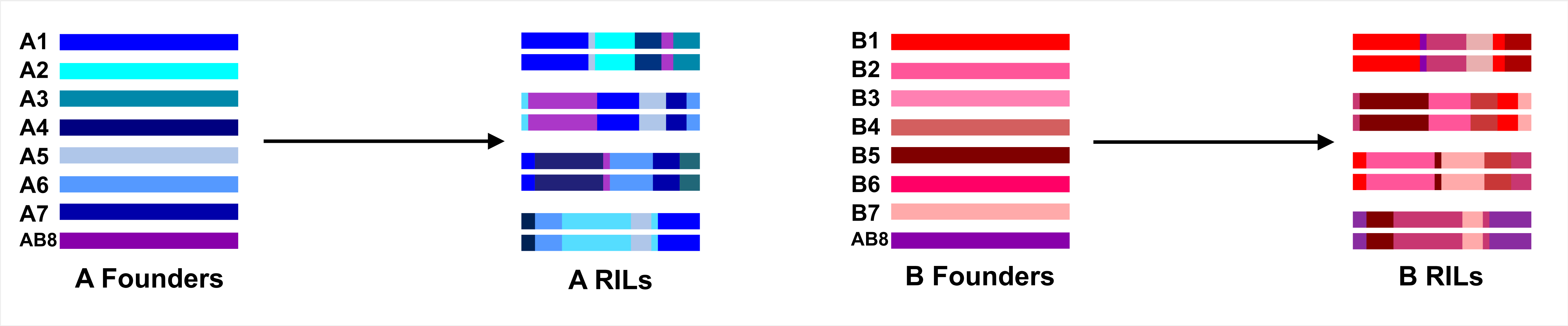
F_{75}

RILs

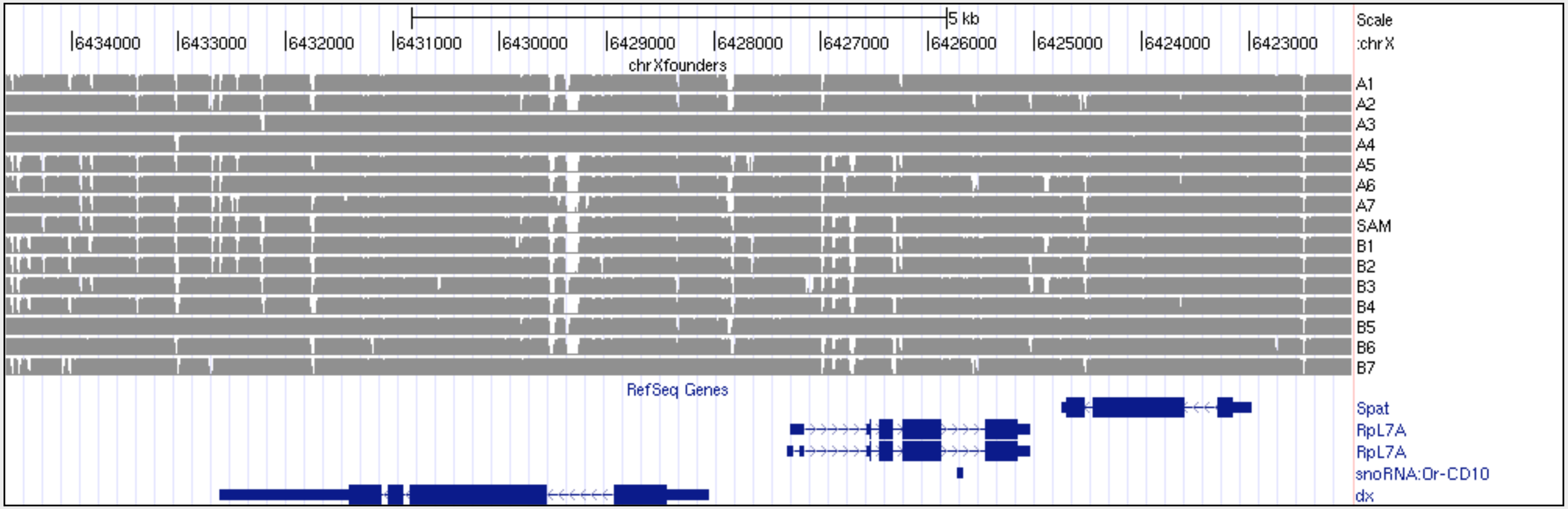
King *et al.*, Genome Res. 2012

DSPR Founders





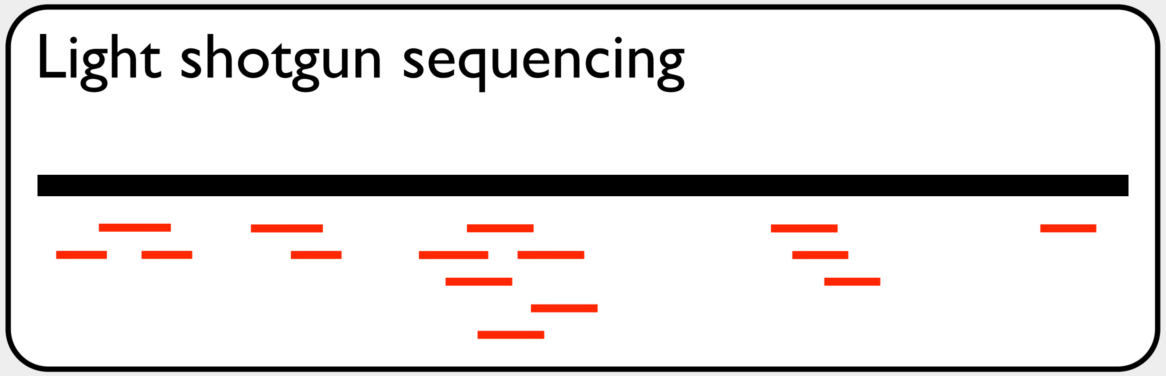
Founder Sequencing



UCSC Genome Browser, *D. melanogaster* chrX:6,422,043-6,434,615

- 50X coverage per founder
- 1.89 million SNPs

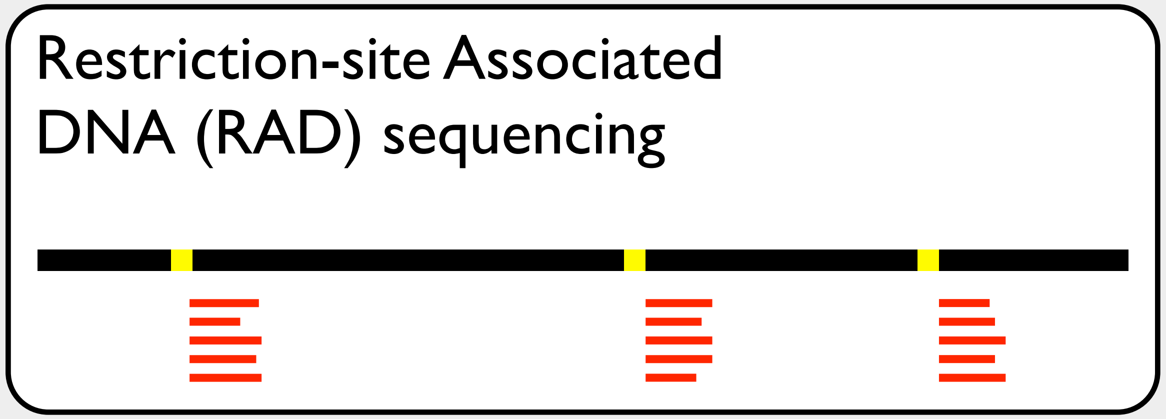
RIL Genotyping-By-Sequencing



Light shotgun sequencing

Low read coverage at any given position

Low confidence in SNP genotype call



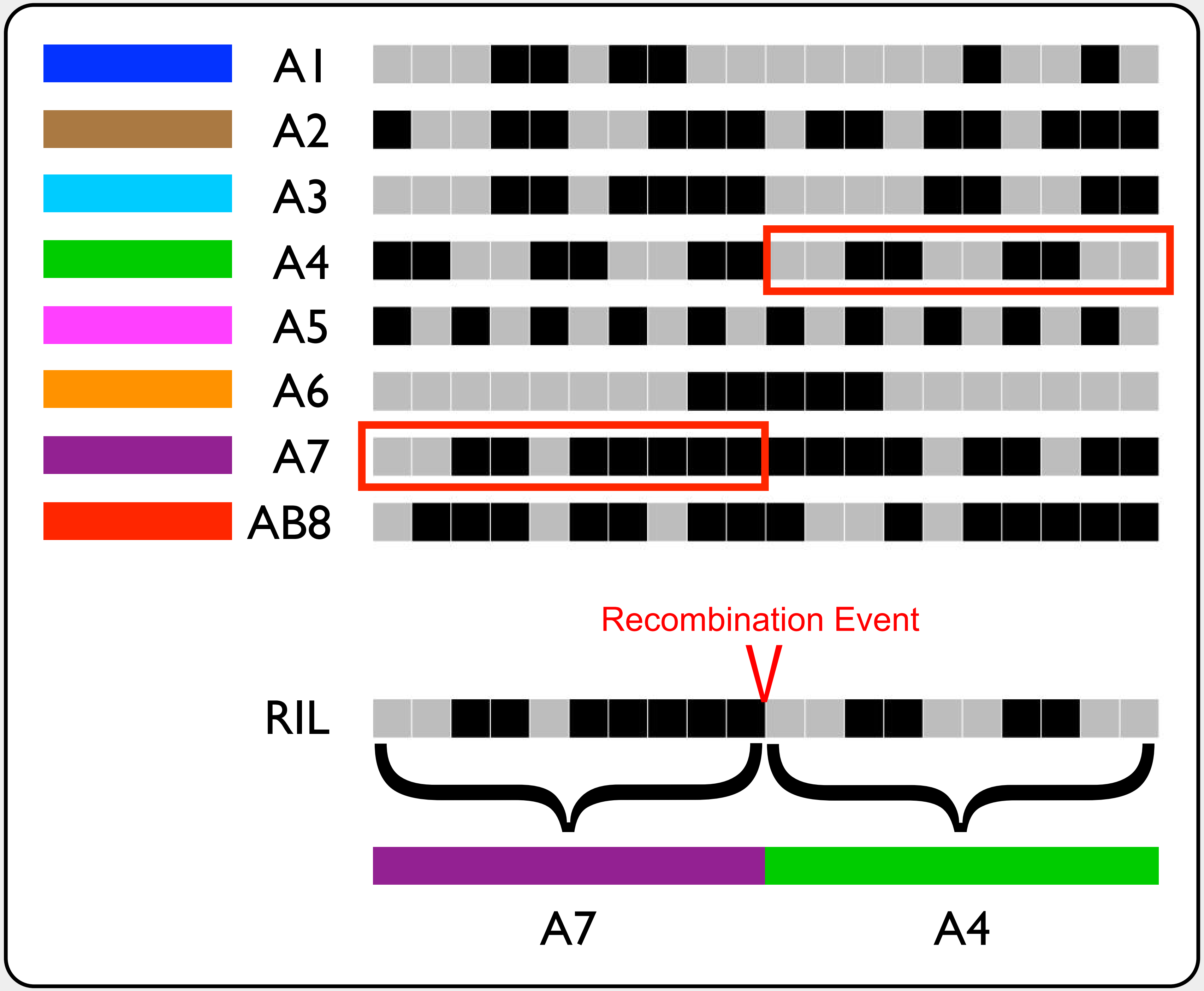
Restriction-site Associated DNA (RAD) sequencing

High read coverage adjacent to restriction sites

Confident genotype calls for SNPs within 100-bp

- 1,700 RILs genotyped at 96-plex
- Genotyped 10,275 SNPs

Uncover Mosaic RIL Structure

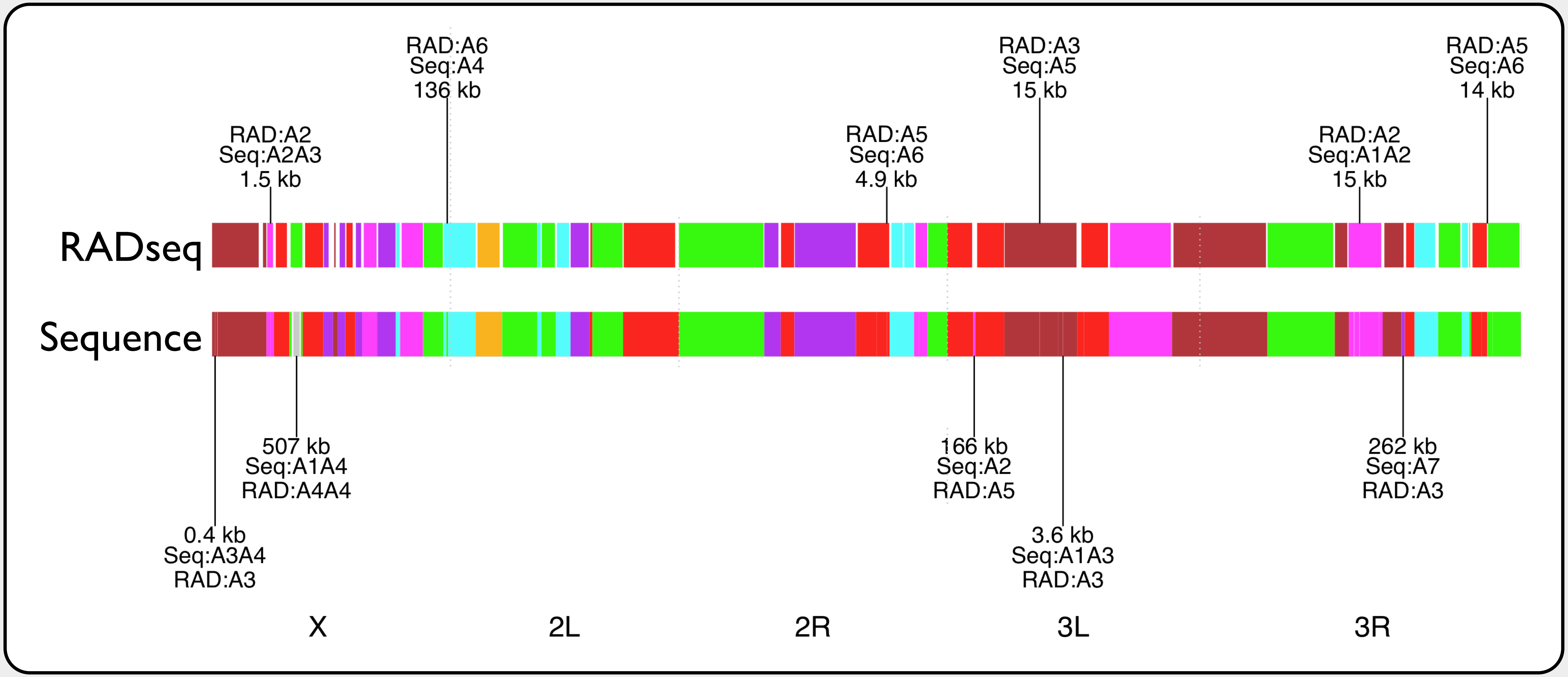


SNPs sequenced in founders
(50X coverage/line)

SNPs genotyped in RIL
(10K SNPs via RADseq)

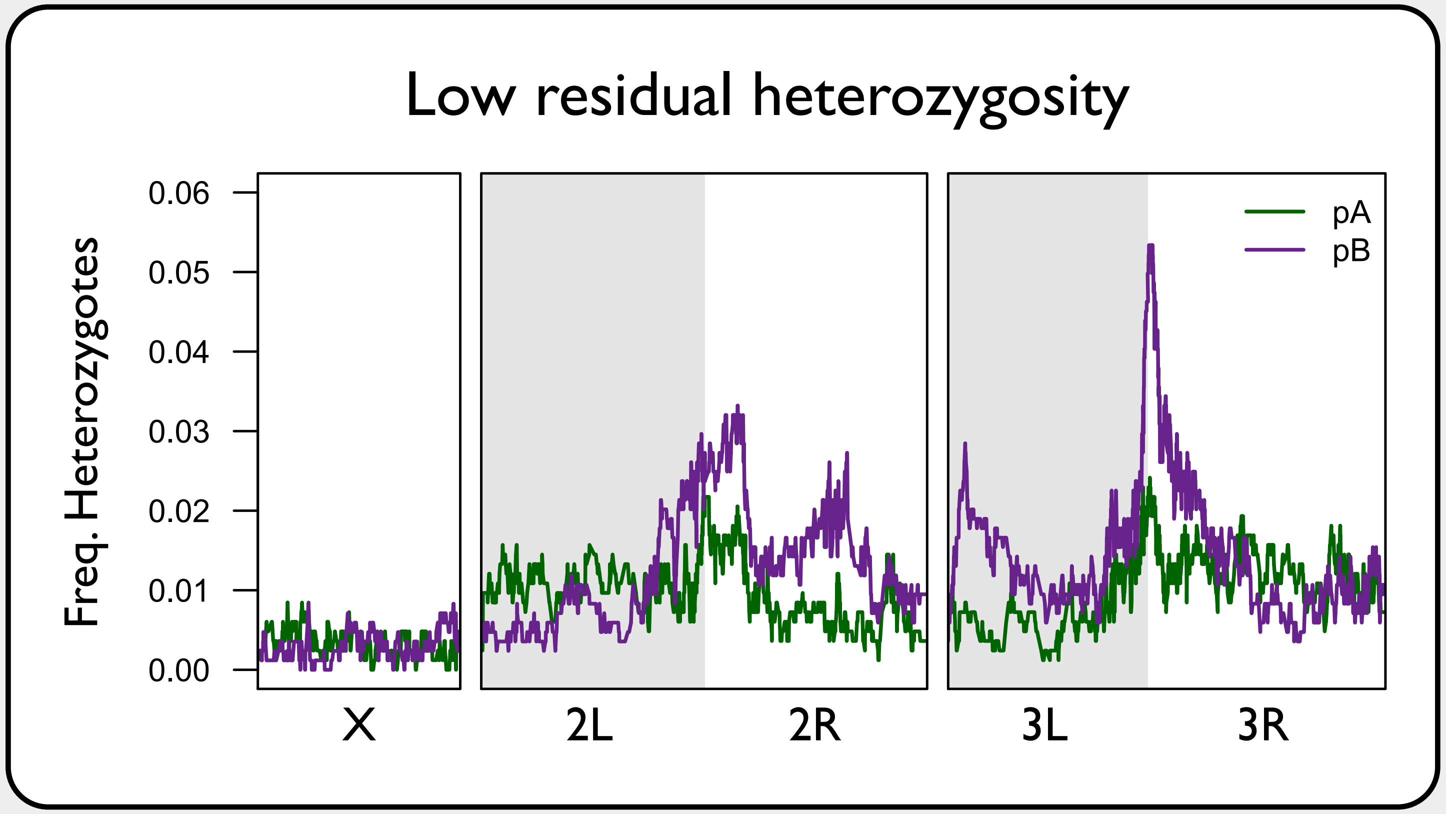
RIL haplotype structure
(HMM)

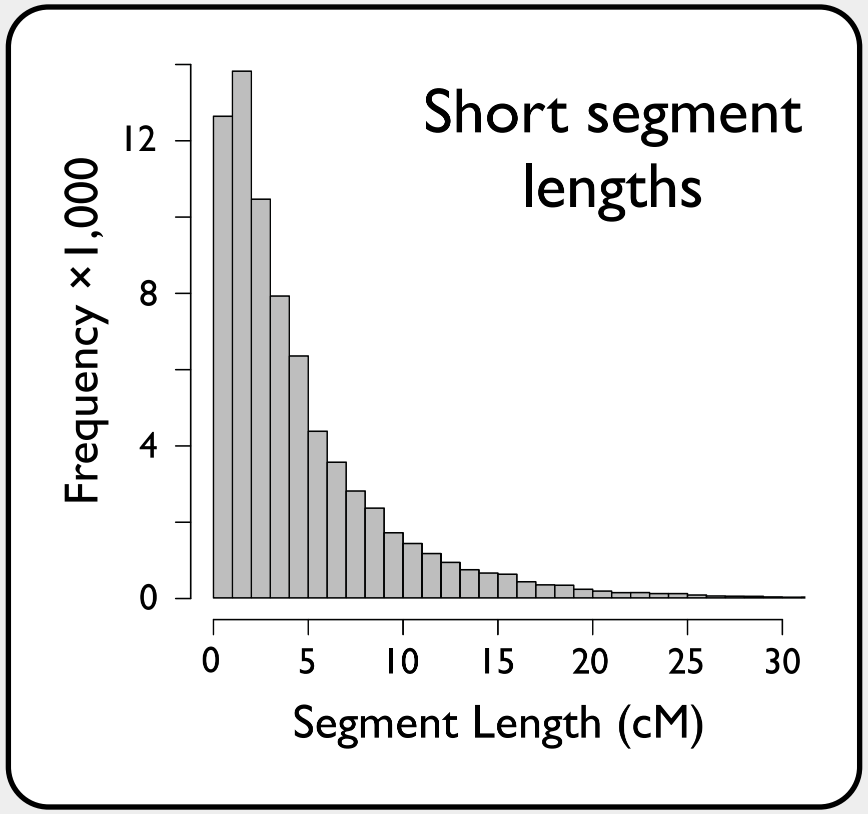
RADseq *vs* Full Resequencing

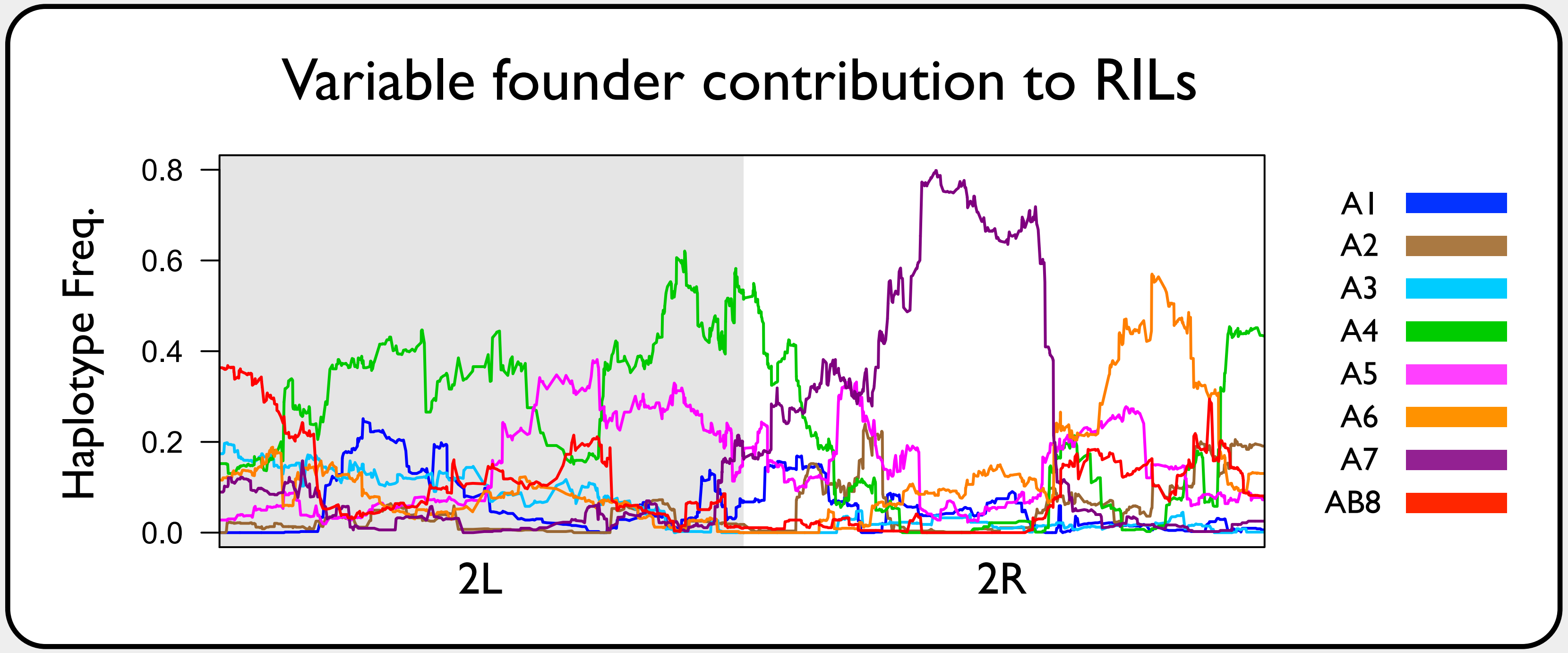


97% genomewide identity between RADseq-derived genotypes and full resequencing

Properties of DSPR RILs

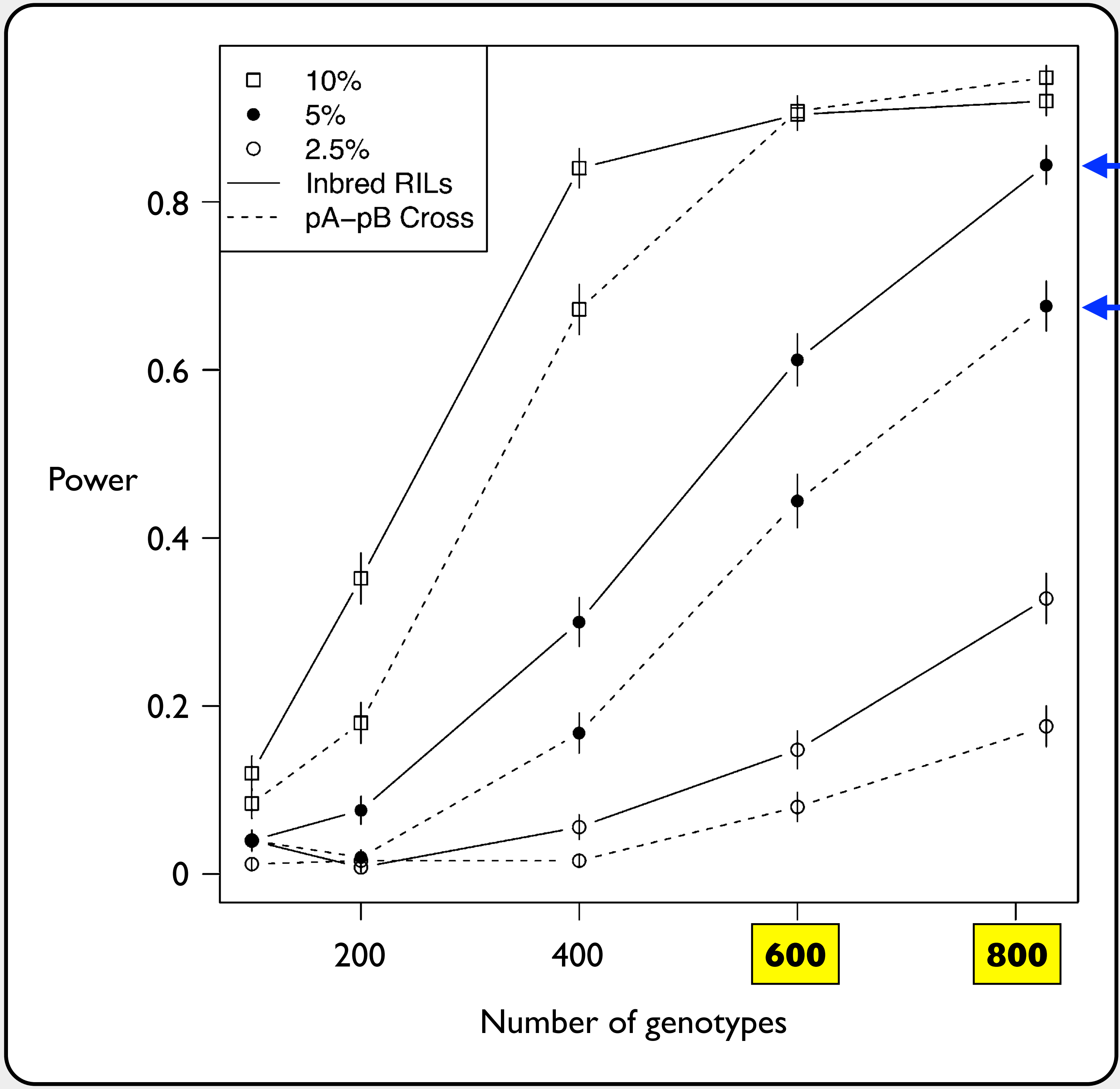

Low residual heterozygosity


Short segment lengths


Variable founder contribution to RILs

- Expect each founder to be at 1/8th frequency
- Variation is due to selection and drift during population maintenance

QTL Mapping Power



5% QTL in inbred RILs

5% QTL in pA × pB crosses

Slight reduction in power with pA × pB heterozygous mapping design

King *et al.*, Genetics 2012

Community Resource

DSPR: The Drosophila Synthetic Population Resource

A Powerful Tool for the Genetic Analysis of Complex Traits

What is the DSPR?

The Drosophila Synthetic Population Resource (DSPR) consists of a new panel of over 1700 recombinant inbred lines (RILs) of *Drosophila melanogaster*, derived from two highly recombined synthetic populations, each created by intercrossing a different set of 8 inbred founder lines (with one founder line common to both populations). Complete genome sequence data for the founder lines are available, and in addition, there is a high resolution genetic map for each RIL. The DSPR has been developed as a community resource for high-resolution QTL mapping and is intended to be used widely by the *Drosophila* community.

Here you can track the progress of the development of the DSPR and find information regarding: the DSPR in general, our analytical machinery, and how to obtain the RILs and/or genomic data.

Motivation

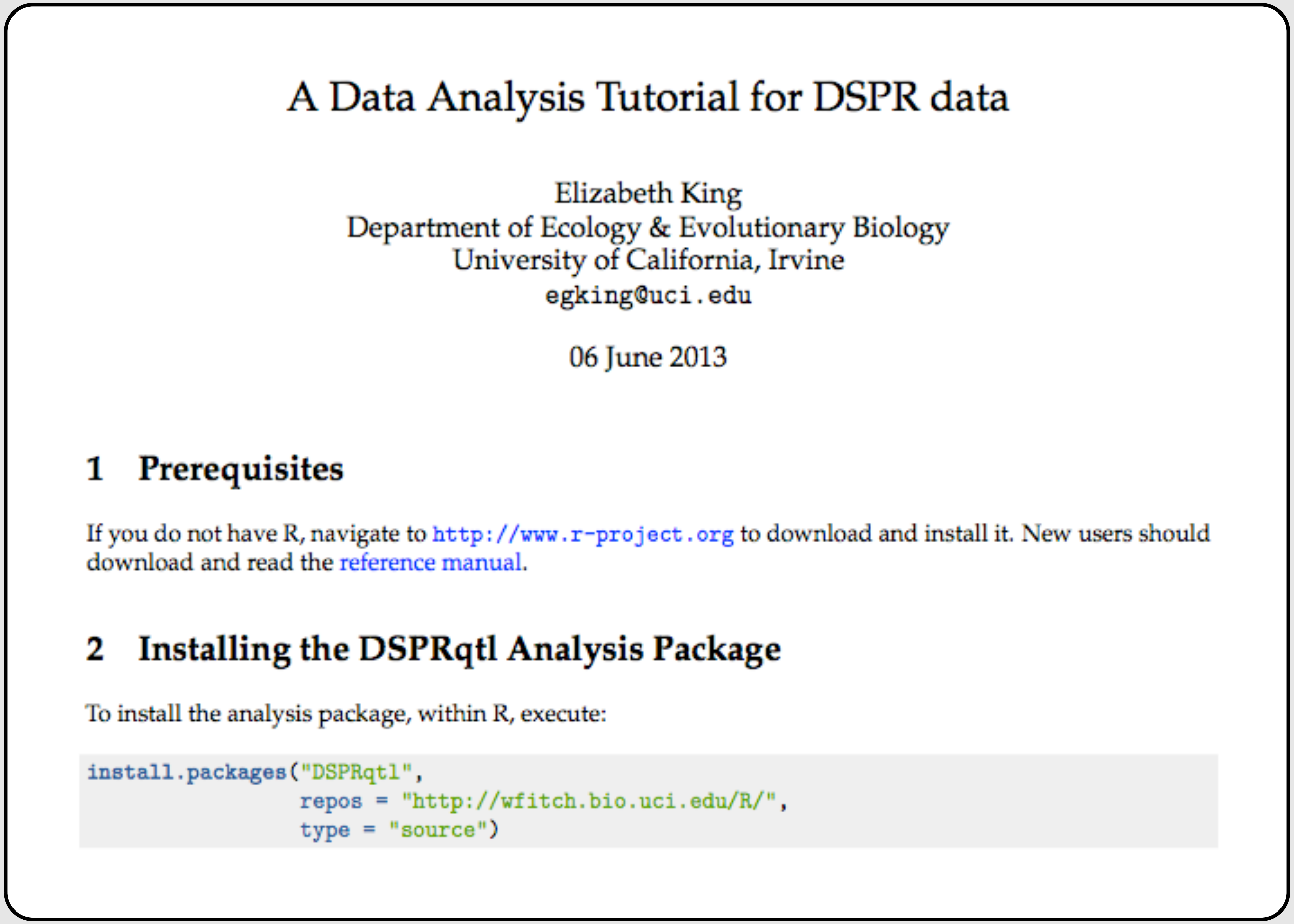
Researchers ultimately wish to dissect the genetic basis of complex phenotypic traits to the nucleotide level, and functionally characterize the effect of the causative allelic substitutions. The DSPR is the next step toward this important goal. Previous approaches aimed at mapping traits, namely traditional QTL mapping and genome-wide association mapping, have significant drawbacks that have prevented widespread success addressing this issue. In traditional QTL mapping, there are few recombination breakpoints, thus, QTL are typically mapped to broad intervals containing an unmanageable number of genes. Genome-wide association mapping utilizes the pattern of historical recombination in natural populations and therefore has the potential for very high resolution. Unfortunately, this approach requires very large sample sizes and high-density SNP maps to achieve high power, and even so is prone to false-positives unless severe statistical correction for multiple testing is applied. The DSPR panel, created from initially crossing multiple inbred lines to make a synthetic population that then undergoes multiple generations of recombination before formation of recombinant inbred lines, represents an idealogical midway between traditional QTL mapping and genome-wide association mapping and has several advantages over these methods. The set of resources available to researchers utilizing the DSPR (a large panel of genotyped 8-way RILs, fully sequenced founders, and easy to use analytical software) will allow researchers utilizing the DSPR to move rapidly between phenotype and a small set of putative causative loci. Ultimately, as the DSPR is used more widely by the Drosophila community, researchers can address broader questions that require data on multiple traits mapped using the same set of RILs.

Advantages of the DSPR

www.FlyRILs.org

- RILs
- Data
- Software

DSPRqtl



A Data Analysis Tutorial for DSPR data

Elizabeth King
Department of Ecology & Evolutionary Biology
University of California, Irvine
`egking@uci.edu`

06 June 2013

1 Prerequisites

If you do not have R, navigate to `http://www.r-project.org` to download and install it. New users should download and read the reference manual.

2 Installing the DSPRqtl Analysis Package

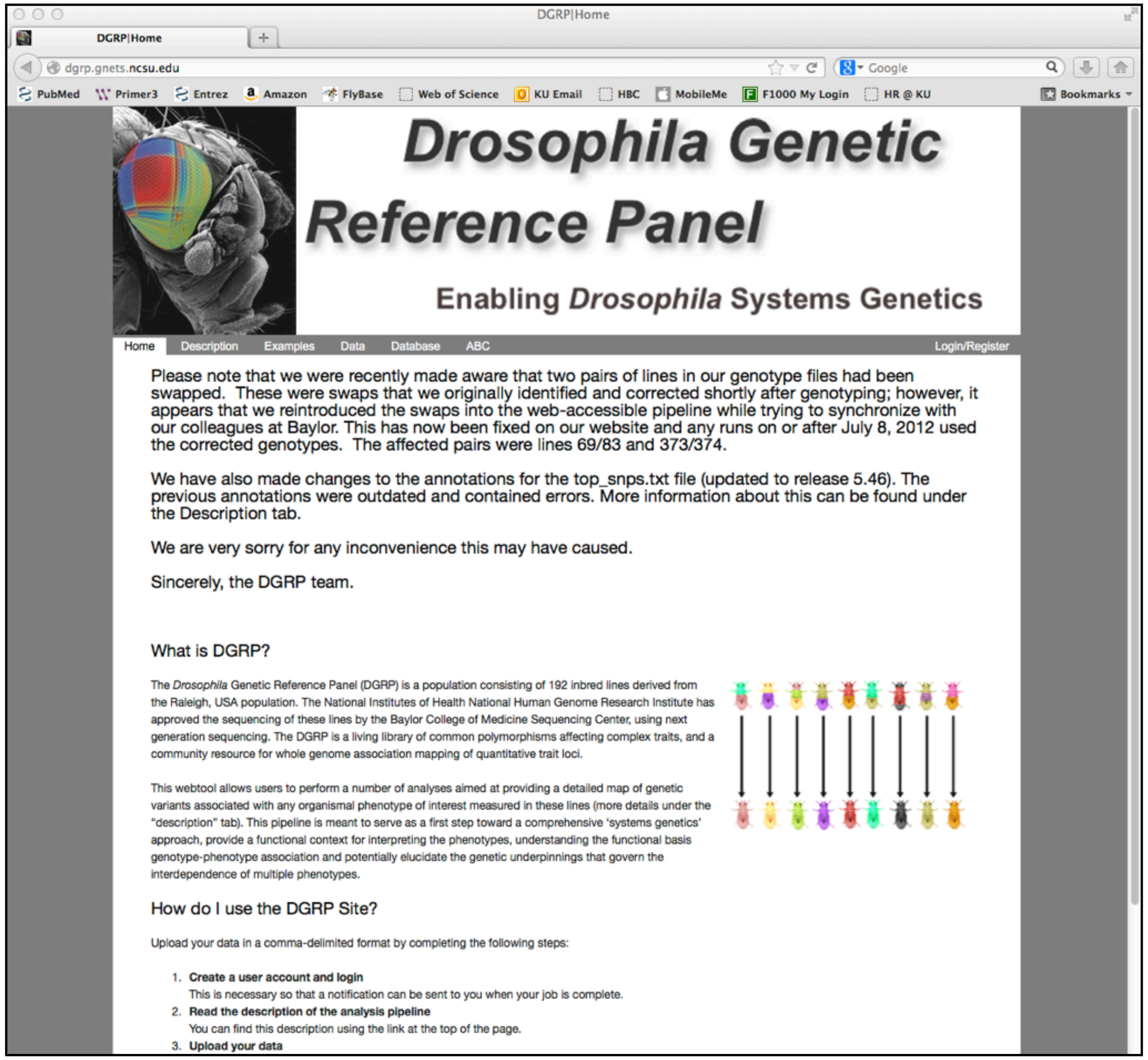
To install the analysis package, within R, execute:

```
install.packages("DSPRqtl",
                 repos = "http://wfitch.bio.uci.edu/R/",
                 type = "source")
```

Libby King

Complementary Resources

- DGRP
- *Drosophila* Genetic Reference Panel
- 168 naturally-derived inbred strains
- Sequenced to an average of ~19X



Drosophila Genetic Reference Panel

Enabling *Drosophila* Systems Genetics

Please note that we were recently made aware that two pairs of lines in our genotype files had been swapped. These were swaps that we originally identified and corrected shortly after genotyping; however, it appears that we reintroduced the swaps into the web-accessible pipeline while trying to synchronize with our colleagues at Baylor. This has now been fixed on our website and any runs on or after July 8, 2012 used the corrected genotypes. The affected pairs were lines 69/83 and 373/374.

We have also made changes to the annotations for the top_snps.txt file (updated to release 5.46). The previous annotations were outdated and contained errors. More information about this can be found under the Description tab.

We are very sorry for any inconvenience this may have caused.

Sincerely, the DGRP team.

What is DGRP?

The *Drosophila* Genetic Reference Panel (DGRP) is a population consisting of 192 inbred lines derived from the Raleigh, USA population. The National Institutes of Health National Human Genome Research Institute has approved the sequencing of these lines by the Baylor College of Medicine Sequencing Center, using next generation sequencing. The DGRP is a living library of common polymorphisms affecting complex traits, and a community resource for whole genome association mapping of quantitative trait loci.

This webtool allows users to perform a number of analyses aimed at providing a detailed map of genetic variants associated with any organismal phenotype of interest measured in these lines (more details under the "description" tab). This pipeline is meant to serve as a first step toward a comprehensive 'systems genetics' approach, provide a functional context for interpreting the phenotypes, understanding the functional basis genotype-phenotype association and potentially elucidate the genetic underpinnings that govern the interdependence of multiple phenotypes.

How do I use the DGRP Site?

Upload your data in a comma-delimited format by completing the following steps:

1. **Create a user account and login**
   This is necessary so that a notification can be sent to you when your job is complete.
2. **Read the description of the analysis pipeline**
   You can find this description using the link at the top of the page.
3. **Upload your data**

Mackay *et al.*, Nature 2012

Acknowledgements

Current Lab Members

Jenny Hackett
Chad Highfill
Sophia Loschky
Brittny Smith

Kristen Cloud*
Matthew Turner*
Kenna Whitley*

Former Lab Members

Steve Hoofer
Tara Marriage
Casey McNeil
Theresa Melhem
Chris Merkes
Michael Najarro
Brian Sanderson

Clint Bain*

Collaborators

Karl Broman (UW-M)
David Davido (KU)
Libby King (UCI)
Tony Long (UCI)
Erik Lundquist (KU)
Lynda Morrison (SLU)
Saunak Sen (UCSF)
Andrew Symons (KU Med)

National Center for Research Resources
NIH and the ARRA

NIH R01-OD010974
(co-I: Tony Long)

K-INBRE
Kansas IDeA Network of Biomedical Research Excellence

NIH P20-GM103418
(PI: Doug Wright)

National Institute of General Medical Sciences

NIH R01-GM085260

National Institute of Neurological Disorders and Stroke

NIH R21-NS070417
(PI: Erik Lundquist)