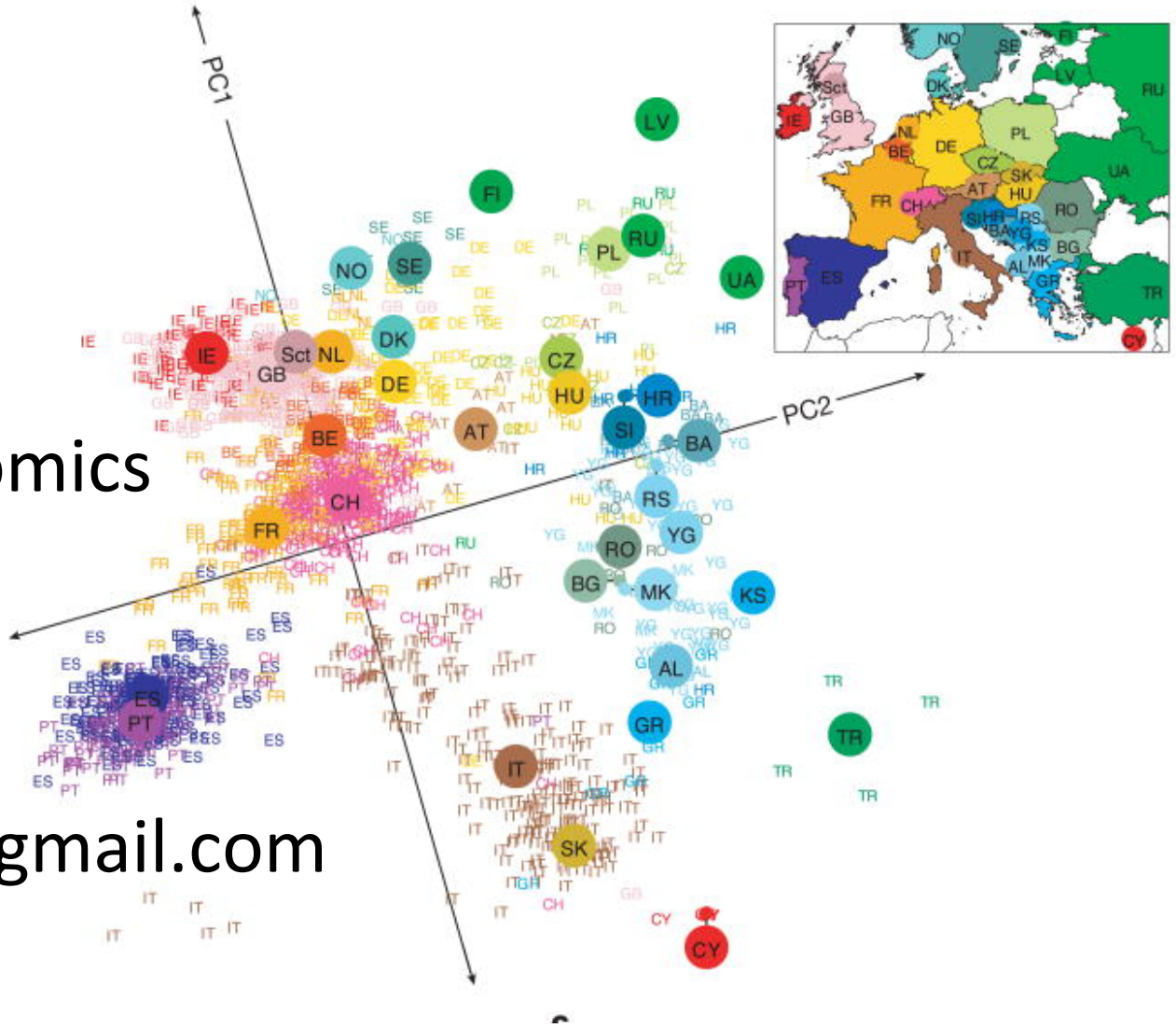
Population Genomics



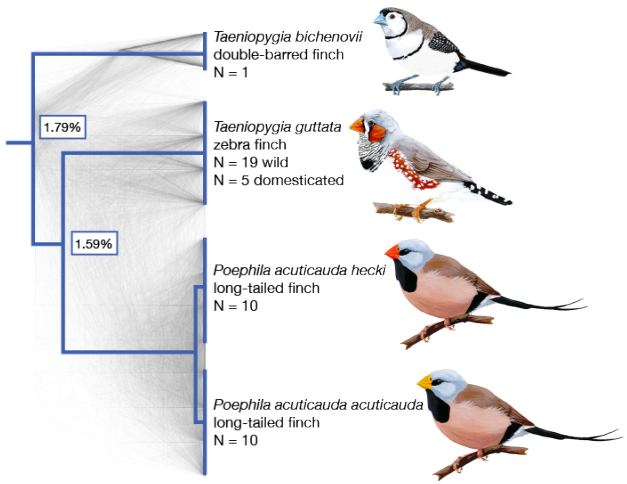
Sonal Singhal
sonal.singhal1@gmail.com

Rabosky Lab
University of Michigan

Novembre et al 2008
10.1038/nature07331

My Background

- Demography, hybrid zones & introgression
- Fine-scale recombination rate estimates
- Population genetics & macroevolution



Today's Plan

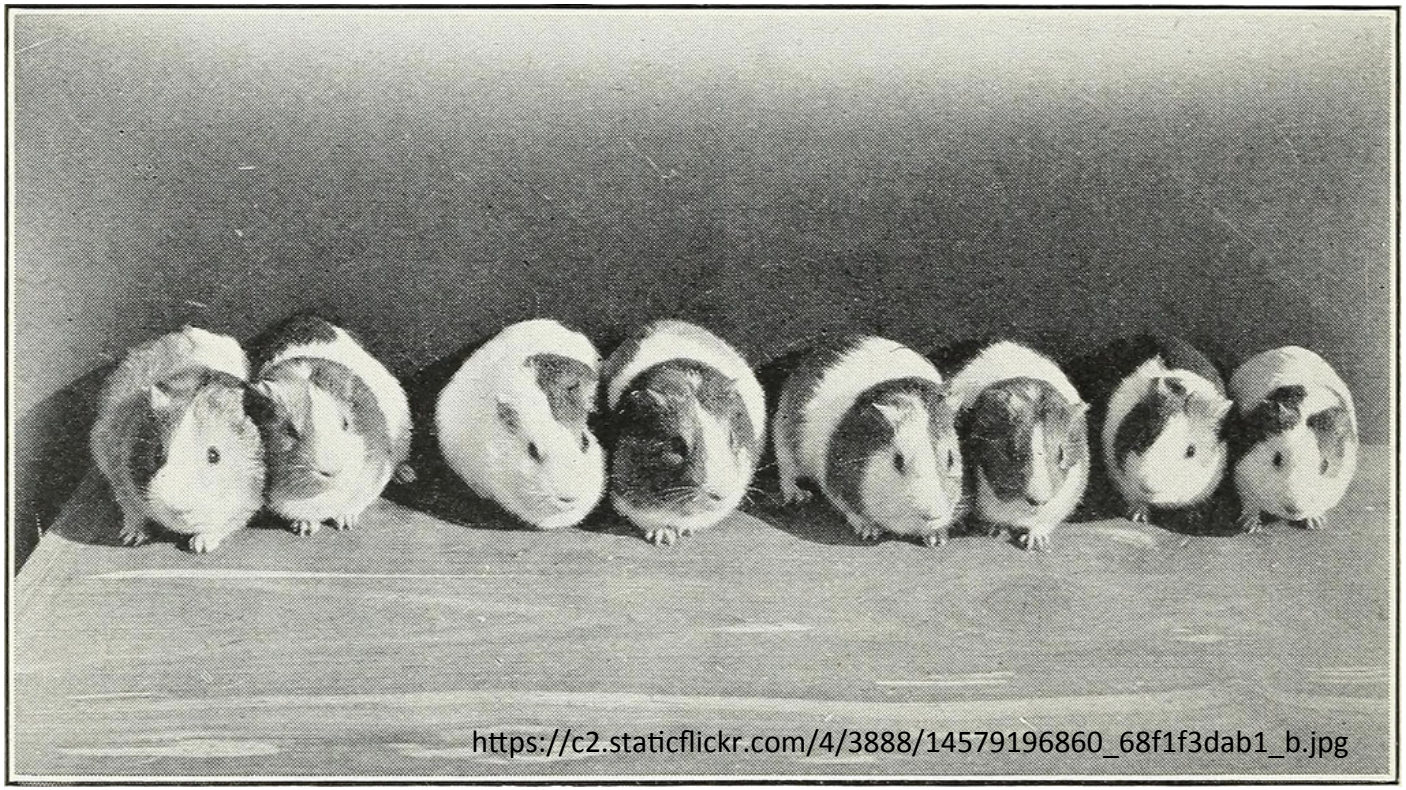
- What is population genomics?
- What are the possibilities and pitfalls?
- Sampling design
- Thought exercises
- Hands-on exercises
- Resource list

What is population genomics?

Population genomics has become a bit of an umbrella term.

- Genetic diversity (genotype calling)
- Linkage disequilibrium (phasing)
- Genetic differentiation across landscapes and the genome
- Population structure
- Population demography: parentage, inbreeding, etc.
- Reconstructing demographic history, i.e., expansions
- Molecular evolution
- Identifying targets of selection
- Often, phylogeography gets lumped in here

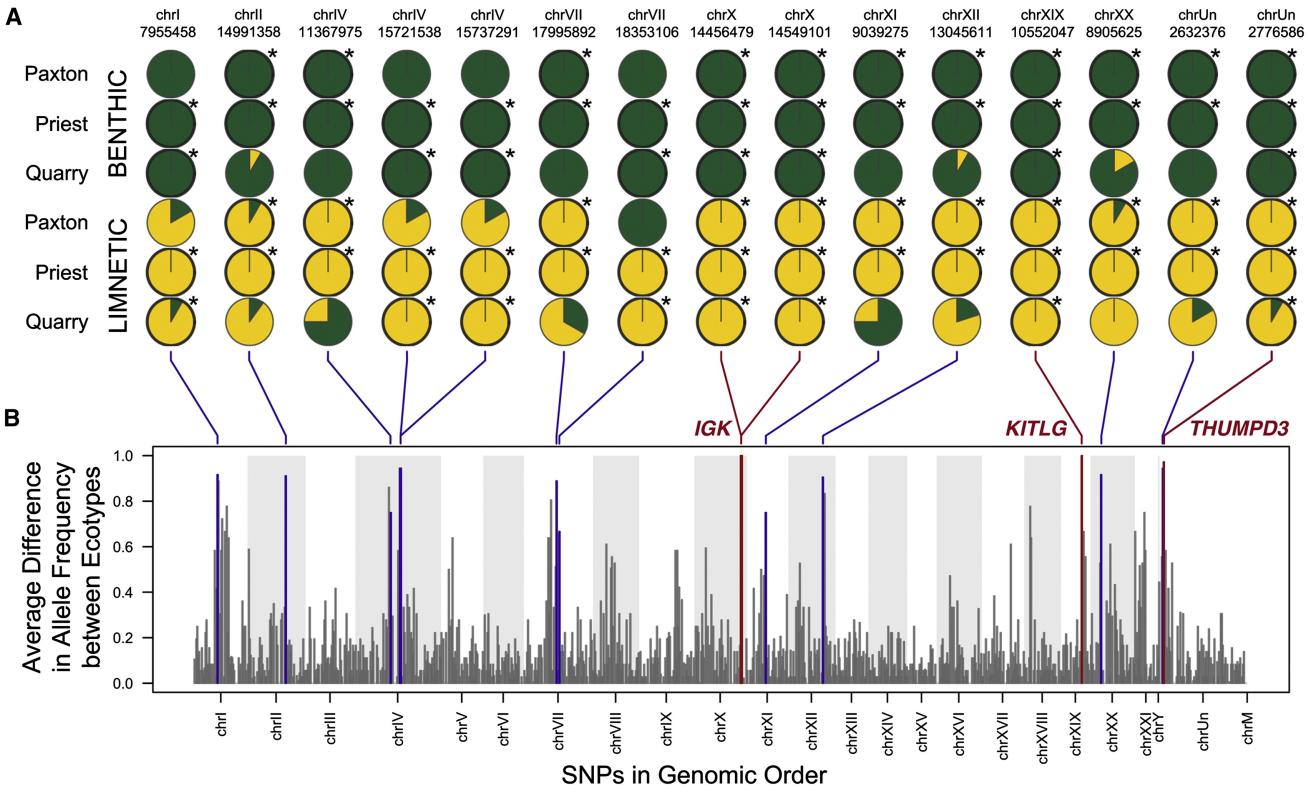
Don’t forget the roots of population genomics: population genetics.



https://c2.staticflickr.com/4/3888/14579196860_68f1f3dab1_b.jpg

Benefits of Population Genomics

Get genome wide patterns! So fun. So interesting.



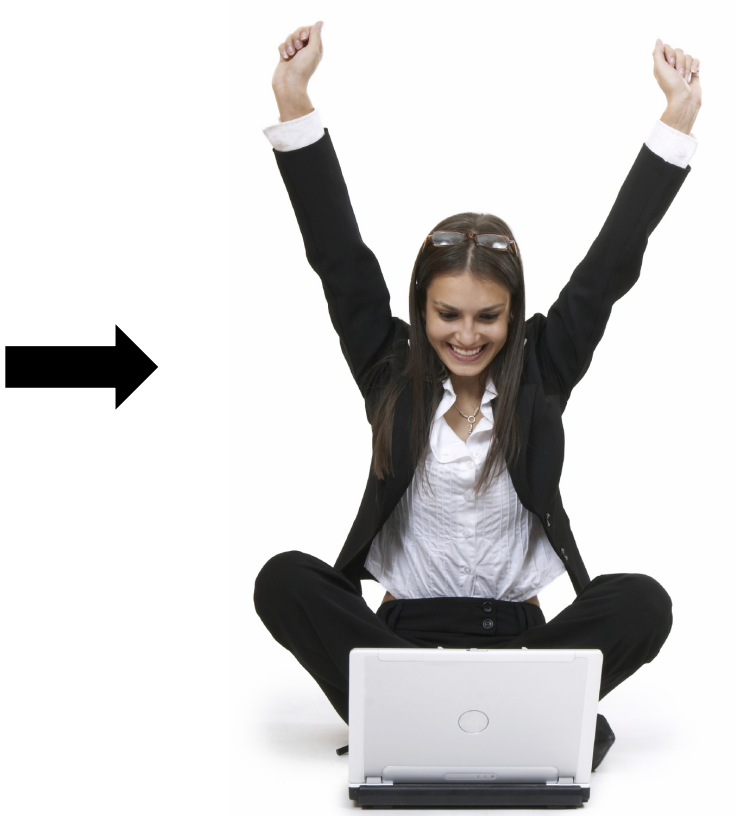
Jones *et al.*, 2012; 10.1016/j.cub.2011.11.045

Benefits of Population Genomics

The cost margin is negligible compared to traditional approaches.

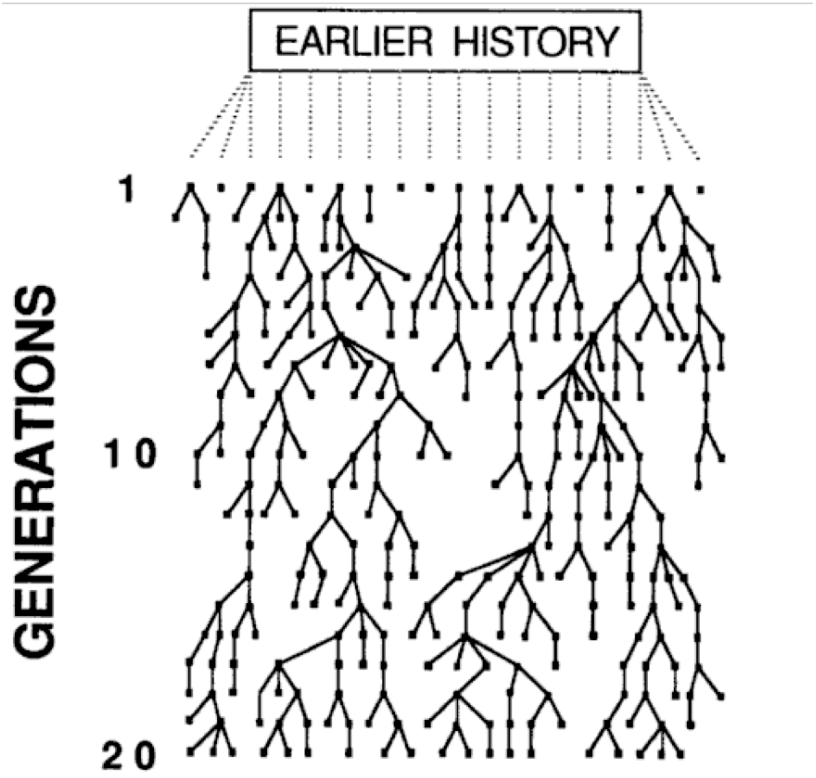
Benefits of Population Genomics

Move from spending time on collecting data to analyzing data.



Benefits of Population Genomics

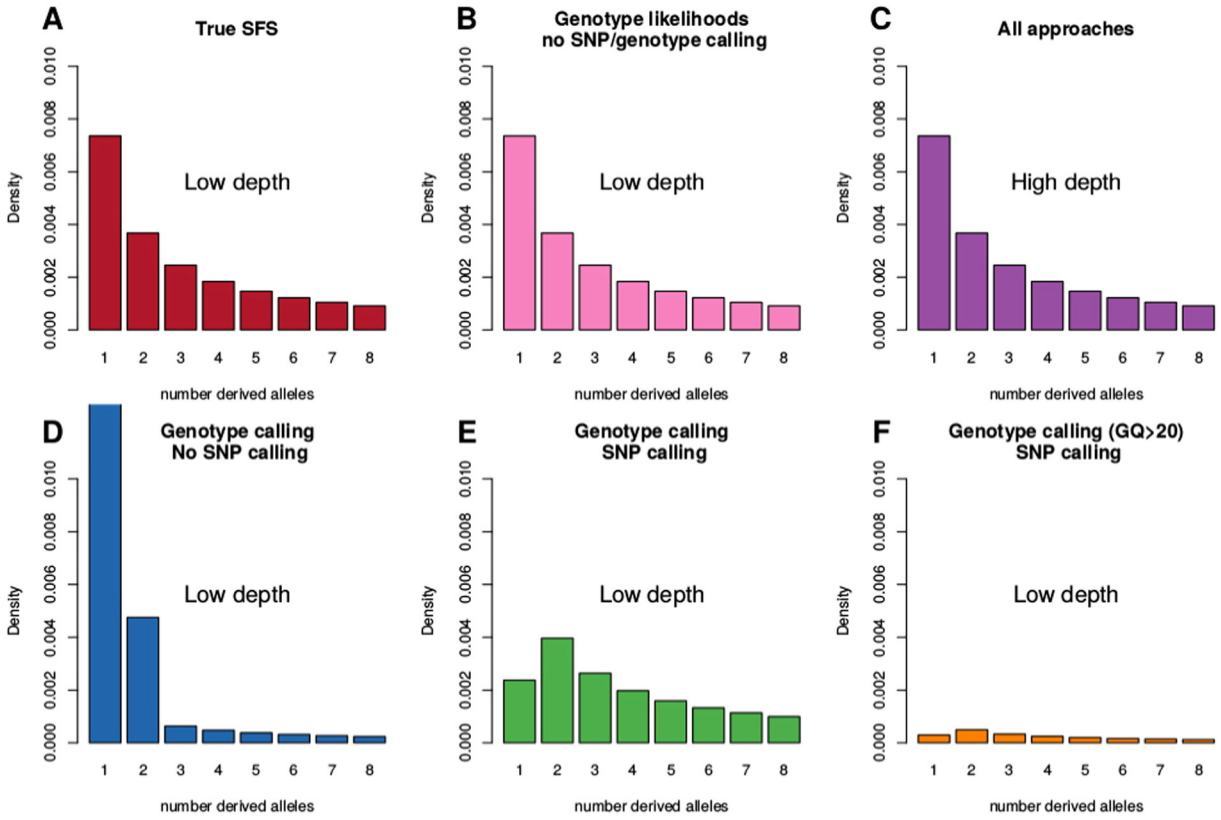
Every individual sampled is equivalent to sampling a population.



http://wallace.genetics.uga.edu/groups/evol3000/wiki/b1495/Coalescent_Theory.html

Challenges of Population Genomics

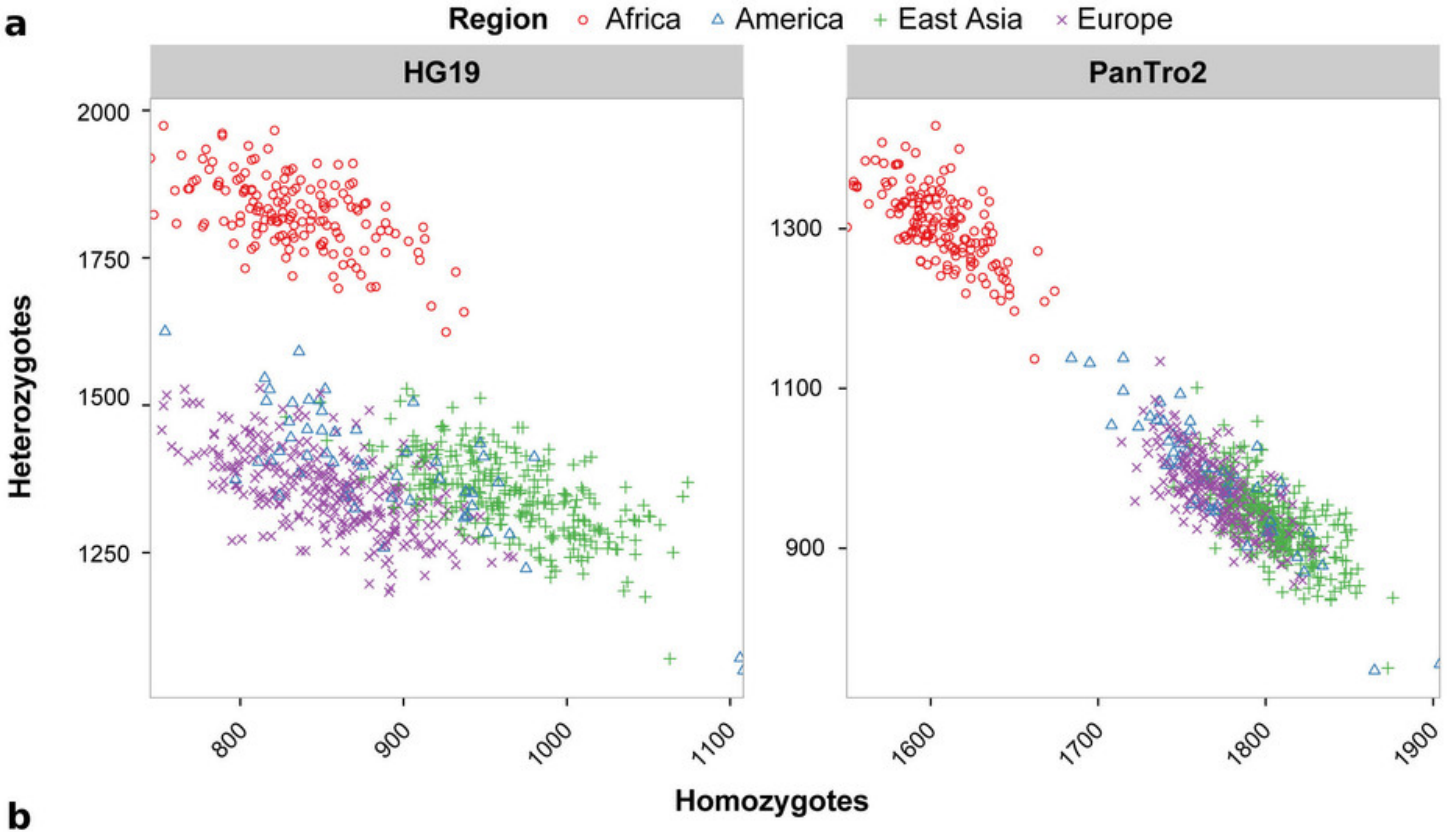
Accuracy of genotype calling: how do you filter?



da Fonesca *et al.*, 2016

Challenges of Population Genomics

Bias, especially ascertainment bias and reference bias.



Scott *et al.*, 2016; 10.1038/ng.3592

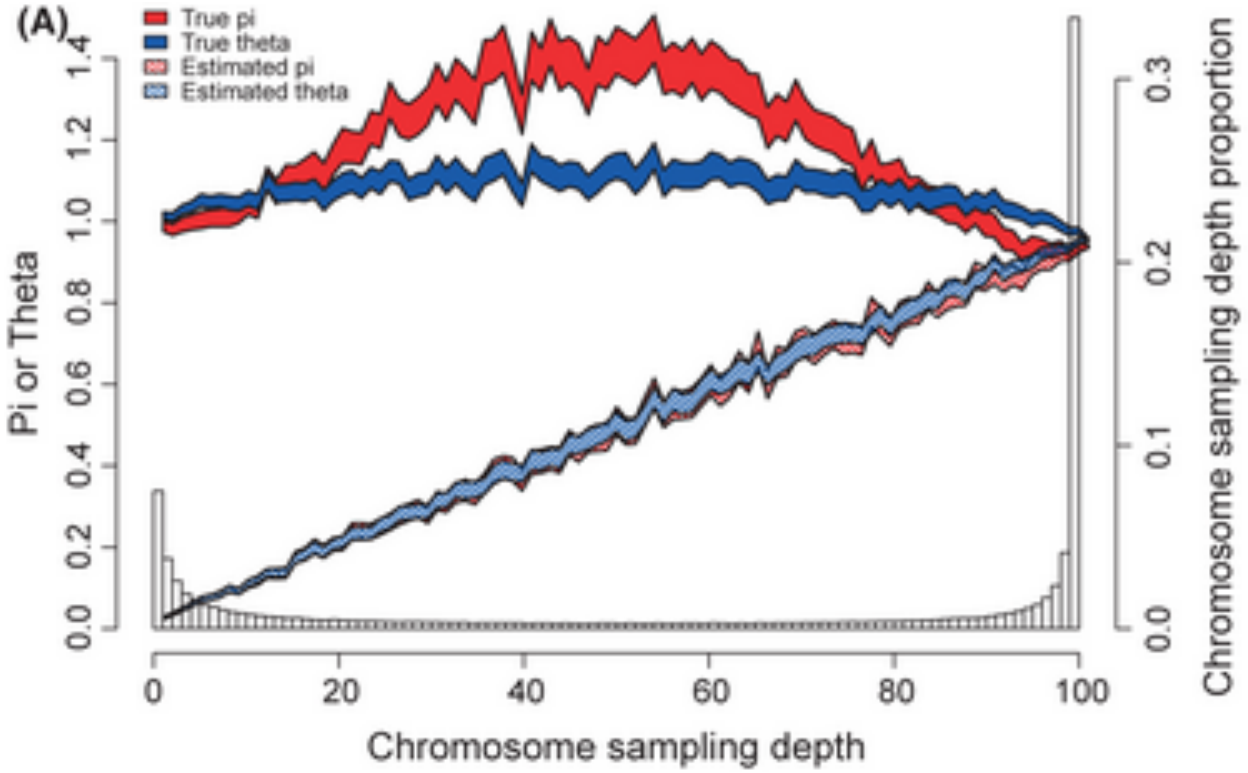
Challenges of Population Genomics

What should your reference be?

http://www.ikmb.uni-kiel.de/pibase/pibase_filtering.html

Challenges of Population Genomics

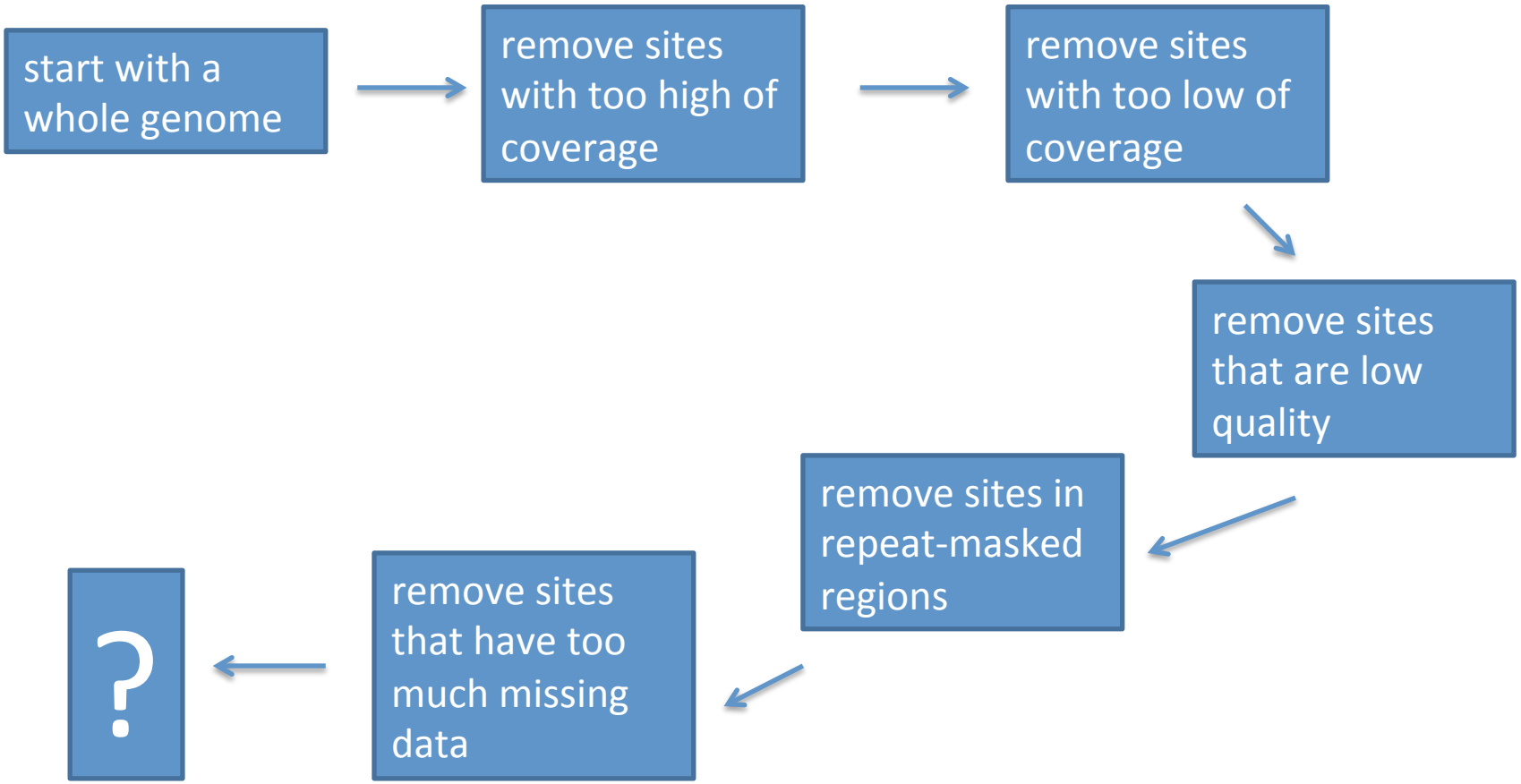
How do you handle missing data?



Arnold *et al*., 2013

Challenges of Population Genomics

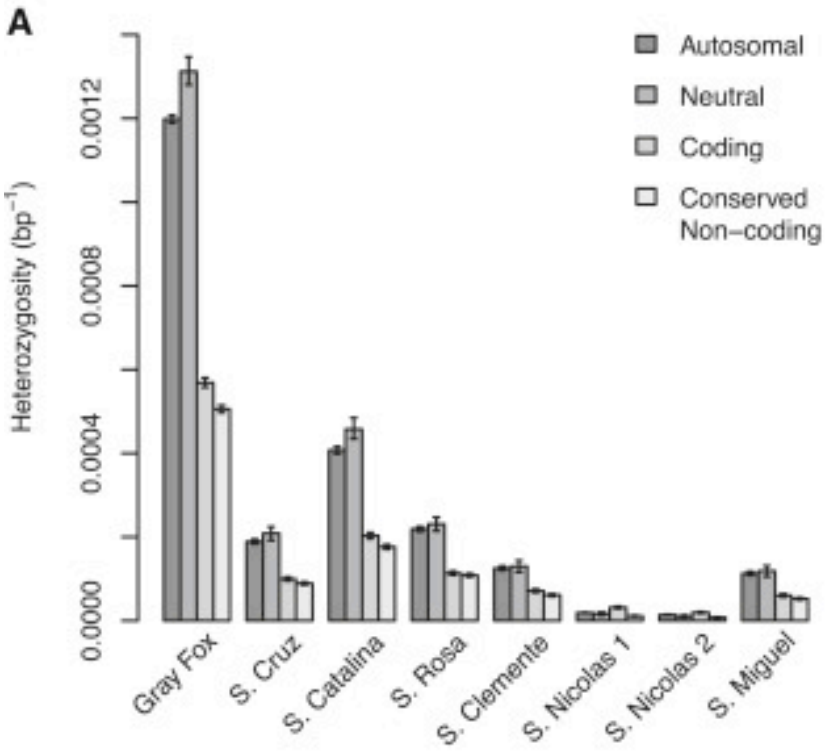
What's your denominator?



da Fonesca *et al.*, 2016

Challenges of Population Genomics

How do you annotate your variants?



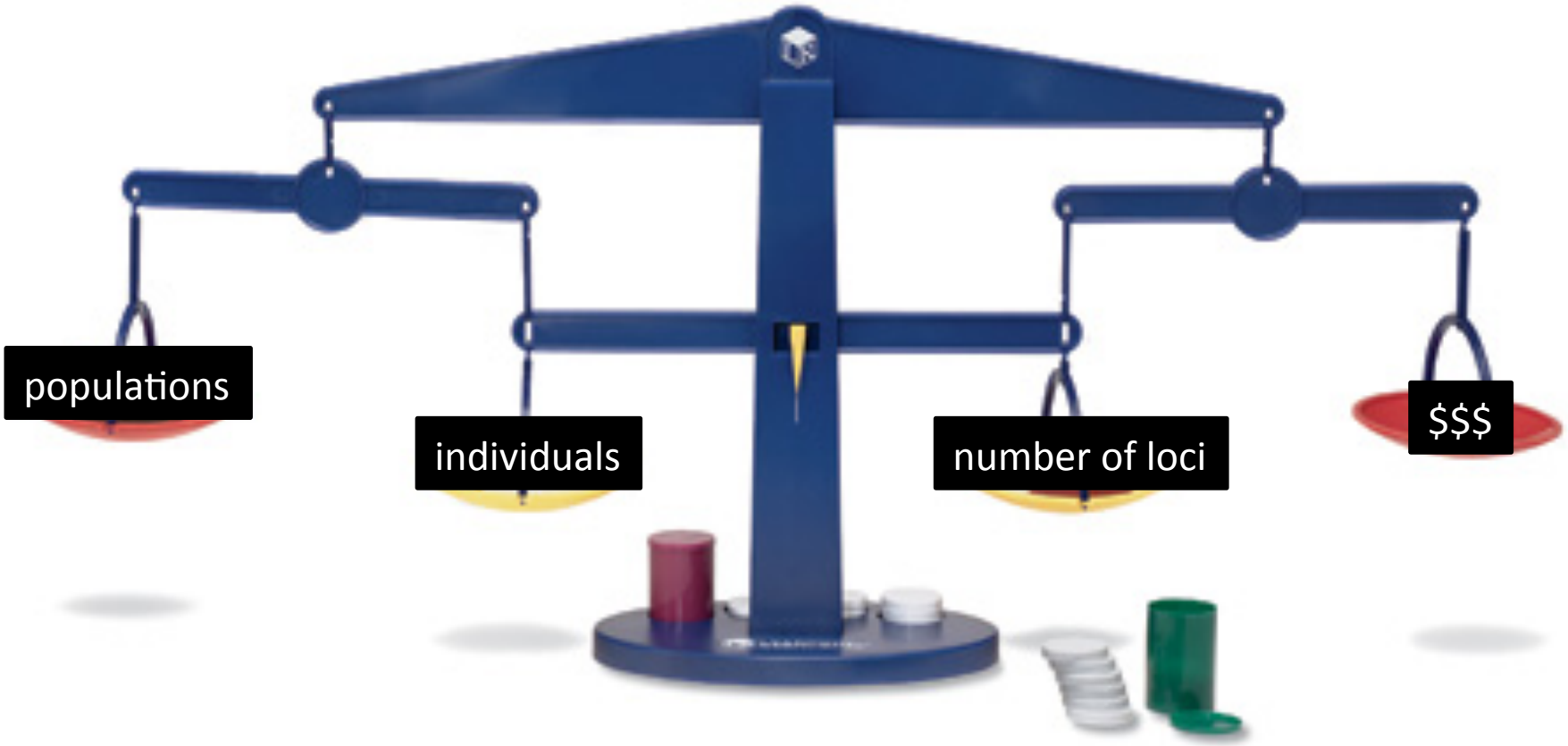
Robinson *et al*., 2016; 10.1016/j.cub.2016.02.062

Sampling Design

- How many individuals?
- How many markers?
- What kind of markers are needed?
- What coverage is needed?
- Can anonymous pooling work?
- What is your phylogenetic scale?
- How much money do you have?

… because how you design your study depends on the answers to these questions …

Sampling design



https://www.learningresources.com

Types of Loci

- whole-genome sequencing
- transcriptomes
- reduced representation sequencing
  - restriction-digest based approaches (RADtag, ddRADtag)
  - targeted capture
    - in-solution capture of exomes, conserved elements, RAD loci (Nimblegen, Roche, MycoArray)
    - home-brewed PCR probe-based
- genotyping SNPs using genotypers

Thought Experiment 1

You discovered a new endangered mammal species and want to explore how much variation is contained in it. How do you do it?

Thought Experiment 2

You are working on a widespread lizard species that you think likely consists of several cryptic lineages. You want to understand how many lineages there are and their demographic & divergence history. How do you do it?

Thought Experiment 3

You are working on an invasive plant species and want to know what the genetic structure is in the native vs. introduced range. How do you do it?

http://www.dcr.virginia.gov/natural-heritage/image/invspinfo-03.jpg

Thought Experiment 4

You are working on a fish that has a wide latitudinal gradient. You suspect that there has been local adaptation across the range. How do you test for this and identify the possible loci under selection?

www.marinebio.org

Thought Experiment 5

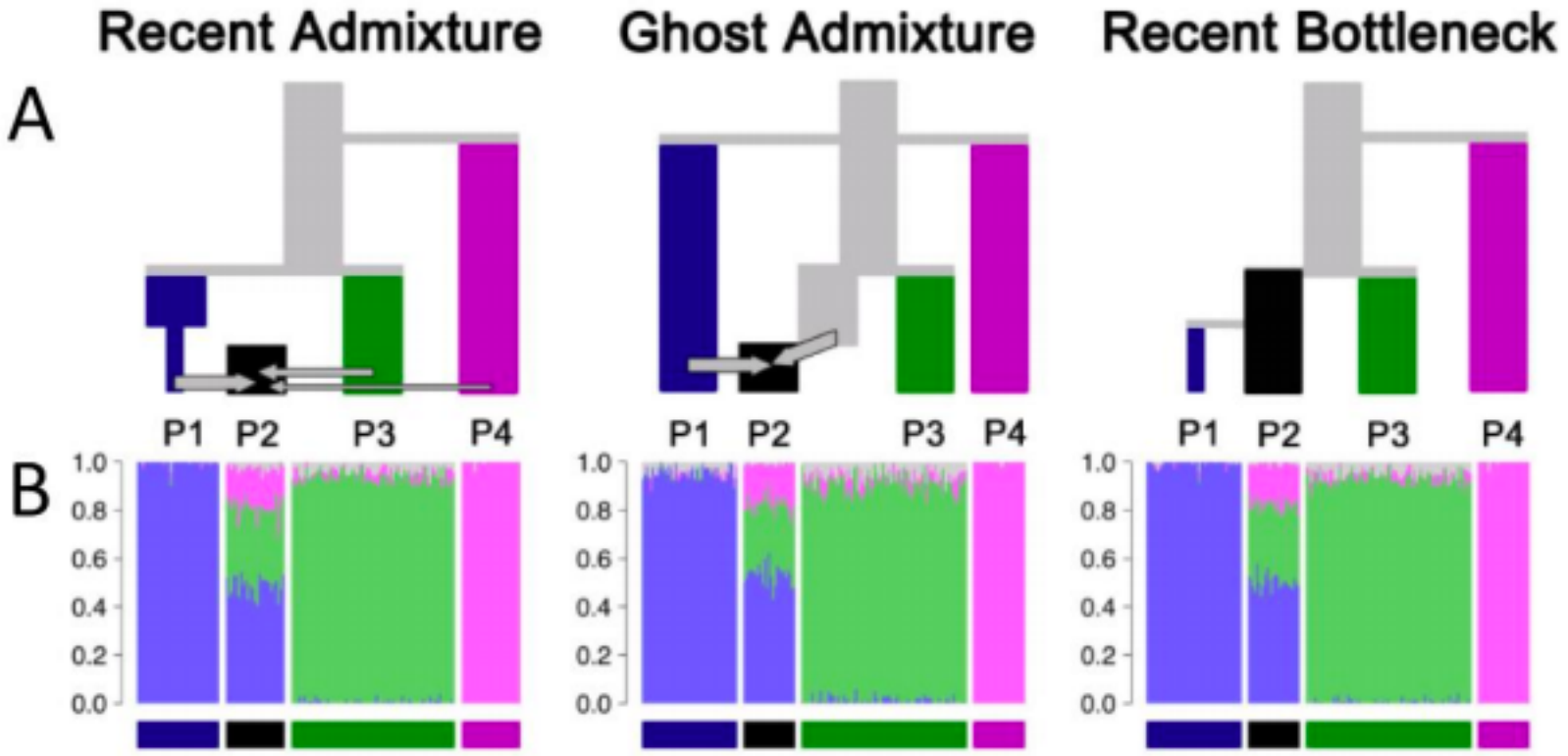
You are working on an insect that has experienced recent massive population declines due to pesticide residues in the environment. Some of its native range remains healthy, and a few susceptible populations have rebounded. How do both test for the genetic effects of these population declines and look for evidence of selective sweeps?

Thought Experiments

1. You discovered a new endangered mammal species and want to explore how much variation is contained in it. How do you do it?
2. You are working on a widespread lizard species that you think likely consists of several cryptic lineages. You want to understand how many lineages there are and their demographic & divergence history. How do you do it?
3. You are working on an invasive plant species and want to know what the genetic structure is in the native vs. introduced range. How do you do it?
4. You are working on a fish that has a wide latitudinal gradient. You suspect that there has been local adaptation across the range. How do you test for this and identify the possible loci under selection?
5. You are working on an insect that has experienced recent massive population declines due to pesticide residues in the environment. Some of its native range remains healthy, and a few susceptible populations have rebounded. How do both test for the genetic effects of these population declines and look for evidence of selective sweeps?

Thoughts on Population Genomics

Distinguish between programs that are process-based vs. descriptive.



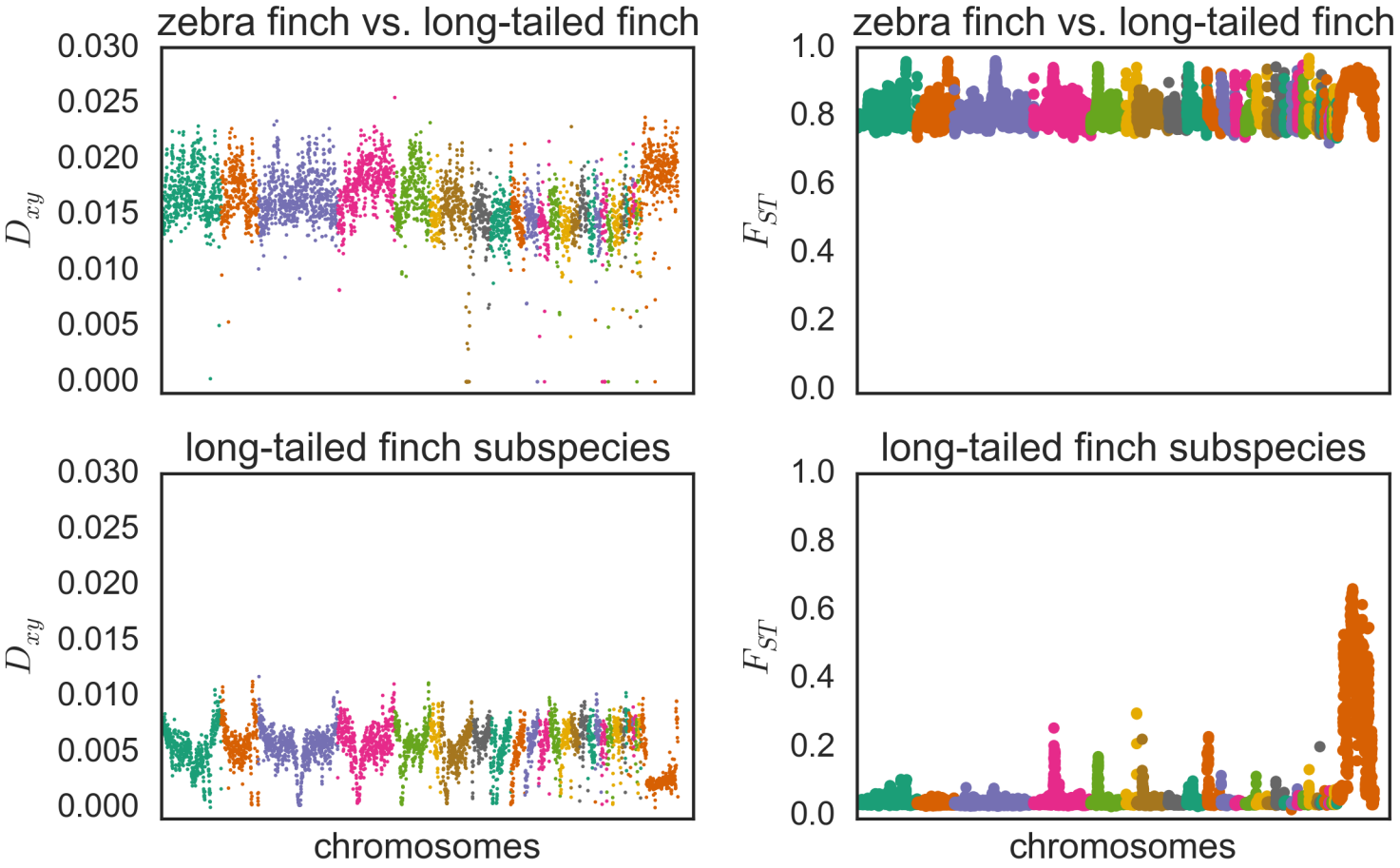
Falush *et al.*, 2016; 10.1101/066431

Thoughts on Population Genomics

You will spend most of your time moving between file formats.

Thoughts on Population Genomics

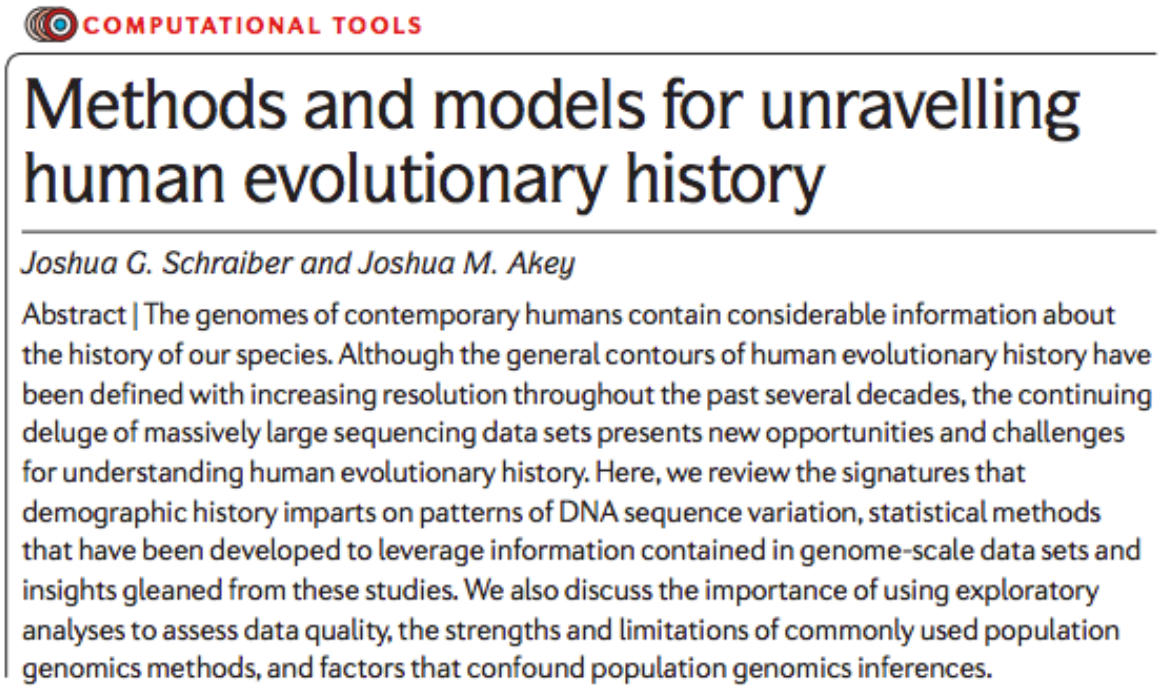
Get familiar with your data! Plot, plot, plot.



Singhal *et al.*, 2015

Thoughts on Population Genomics

Almost all methodological innovations start in the human population genetic literature. Read it.



COMPUTATIONAL TOOLS

Methods and models for unravelling human evolutionary history

Joshua G. Schraiber and Joshua M. Akey

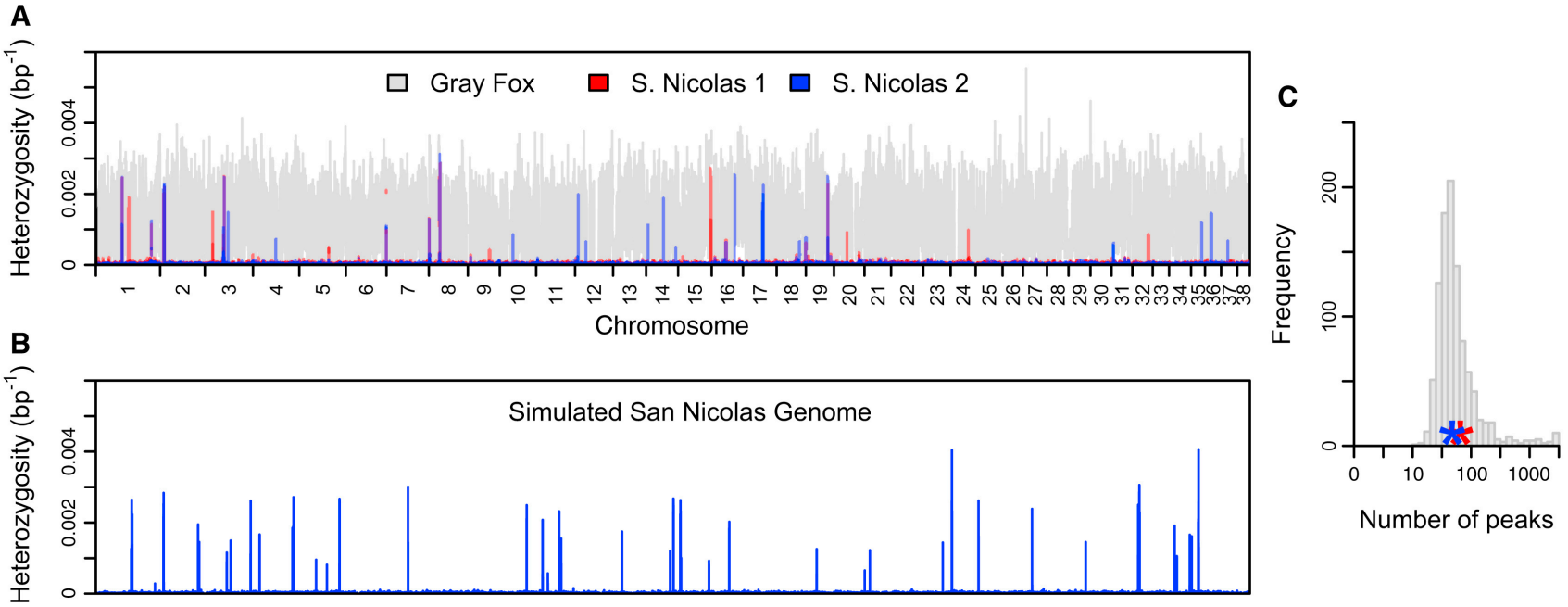
Abstract | The genomes of contemporary humans contain considerable information about the history of our species. Although the general contours of human evolutionary history have been defined with increasing resolution throughout the past several decades, the continuing deluge of massively large sequencing data sets presents new opportunities and challenges for understanding human evolutionary history. Here, we review the signatures that demographic history imparts on patterns of DNA sequence variation, statistical methods that have been developed to leverage information contained in genome-scale data sets and insights gleaned from these studies. We also discuss the importance of using exploratory analyses to assess data quality, the strengths and limitations of commonly used population genomics methods, and factors that confound population genomics inferences.

Schraiber and Akey 2015; 10.1038/nrg4005

Thoughts on Population Genomics

How do you bootstrap results?

Simulations and non-parametric statistics help.



Robinson *et al.*, 2016; 10.1016/j.cub.2016.02.062