Genotyping by sequencing for footprints of selection in *Fusarium graminearum*

Chris Toomajian, Wei Yue & John F. Leslie

Department of Plant Pathology

KANSAS STATE UNIVERSITY

National Fusarium Head Blight Forum

December 9, 2014

Pathogen evolution

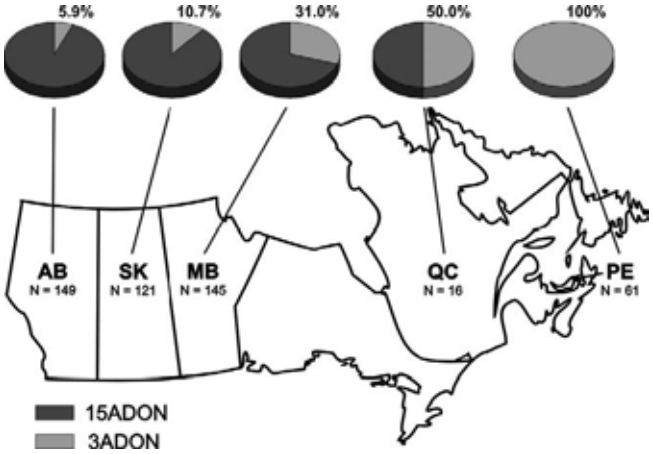
- Spread of novel mutations
  - fungicide resistance
  - host resistance breaking
- Shifts in use of alternate pathogenicity factors
- Adaptation to environmental conditions
  - abiotic
  - biotic
- Genetic loci involved in evolutionary process will be connected to pathogen fitness
- Thus, targeting fitness-related genes may offer a means of pathogen control by affecting pathogen fitness at any stage of their life cycle

F. graminearum population shifts

Trichothecene chemotypes:
3-acetyl deoxynivalenol (3-ADON)
15-acetyl deoxynivalenol (15-ADON)
Nivalenol (NIV)

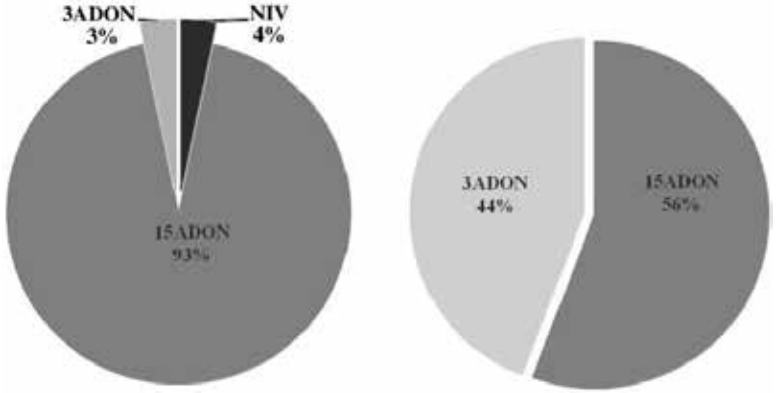
3-ADON chemotype has increased in frequency in some regions

492 *F. graminearum* isolates from wheat between 1984 and 2004



from Ward *et al*. (2008) *Fungal Genetics & Biology* **45**: 473

Frequency of chemotypes from North Dakota *F. graminearum*

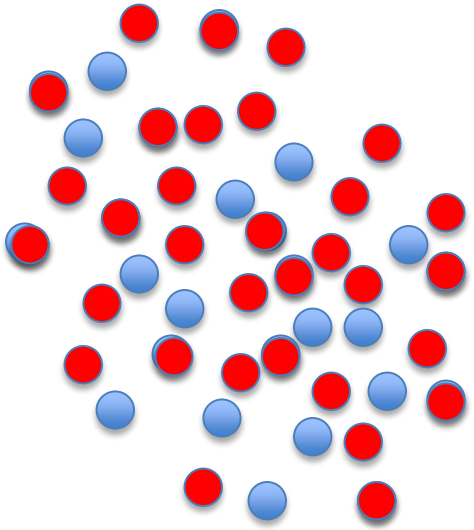


1980-2000 (*n*=120) 2008 (*n*=43)

from Puri and Zhong (2010) *Phytopathology* **100**: 1007

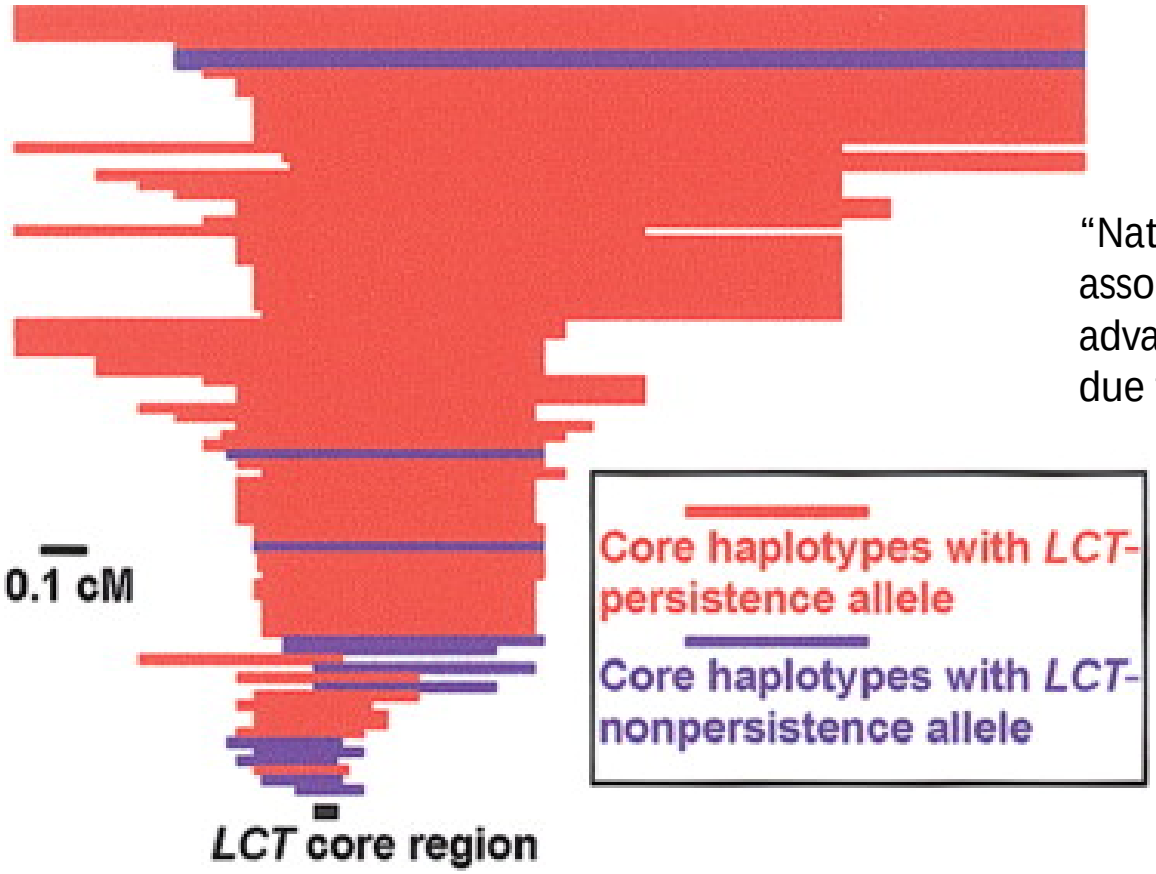
Approach and rationale

- Population genomic analyses
  - Regions of high genetic differentiation
  - Signatures of recent selection (decreased variability, increased linkage disequilibrium)
  - Allele frequency changes



- Unbiased genome scans can identify genes important in pathogen evolution we could not have predicted based on current annotation (and would be missed by candidate gene approaches).
- Purpose is not to find all mutations or rare polymorphisms in sample. Common alleles will tend to be associated with other linked alleles, so full genome resequencing is not required.

Example footprint of selection in human genetic variation



"Natural linkage drag" associated with an advantageous allele's increase due to natural selection

from Bersaglieri *et al*. (2004) *Am J Hum Genet* **74**: 1111

FY14 USWBSI Project

PI: Toomajian, coPI: J. F. Leslie, PhD student: Wei Yue

- **Goal**: Infer which genomic loci played role in spread of 3-ADON chemotype with aim of targeting them for pathogen control
  - **Aim 1**: Genotype > 500 US *Fg* isolates using genotyping by sequencing (GBS) approach
  - **Aim 2**: Analyze population structure, explore relationship between clusters and sampling location, trichothecene chemotype
  - **Aim 3**: Scan genome for footprints of natural selection, identify possible targets for control
  - **Aim 4**: Measure linkage disequilibrium (LD) and assess potential for genome-wide association (GWAS)

Acknowledgments

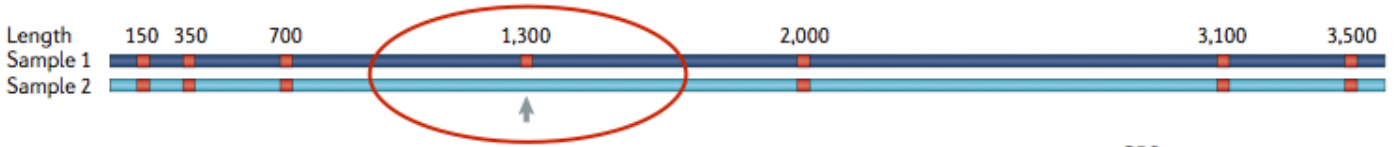
Sharing isolates:
Gary Bergstrom, Cornell
Shaobin Zhong, NDSU
Corby Kistler, Cereal Disease Lab

K-State lab assistance:
Amy Beyer
Wes Strouts
Brandi Worster

F. graminearum reference genome:
Broad Institute

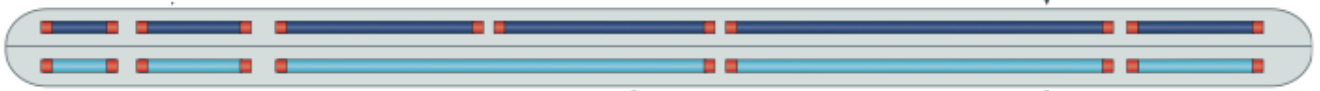
Funding:
USWBSI
Kansas State Research and Extension

Marker Discovery
Genotyping by sequencing - GBS

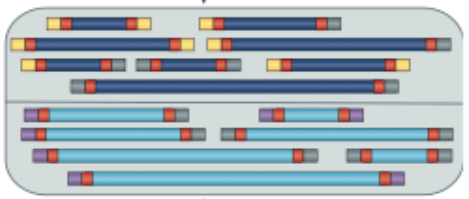


GBS

Restriction digest of each sample

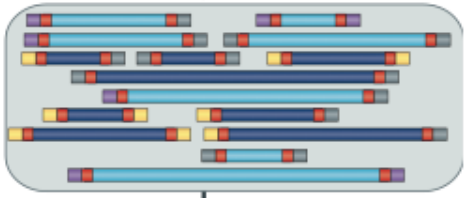


Ligate barcoded adapters



Our GBS protocol follows Poland *et al.* (2012) *PLoS One* 7: e32253

Pool samples

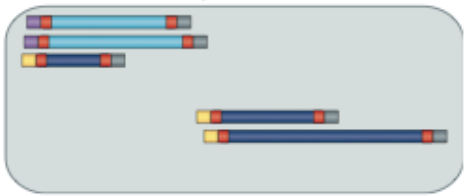


Two types of polymorphisms detected:

1) Presence/absence

2) Sequence difference

PCR & Sequence (HiSeq 100 cycle)



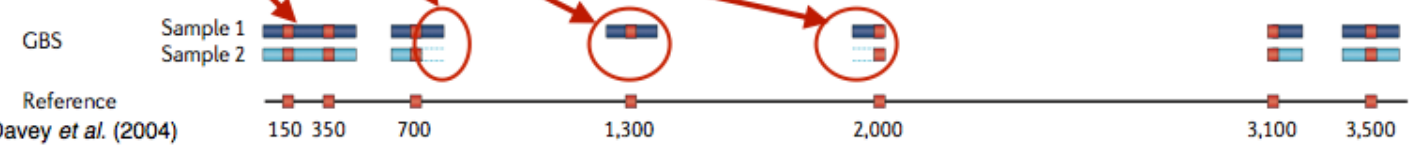
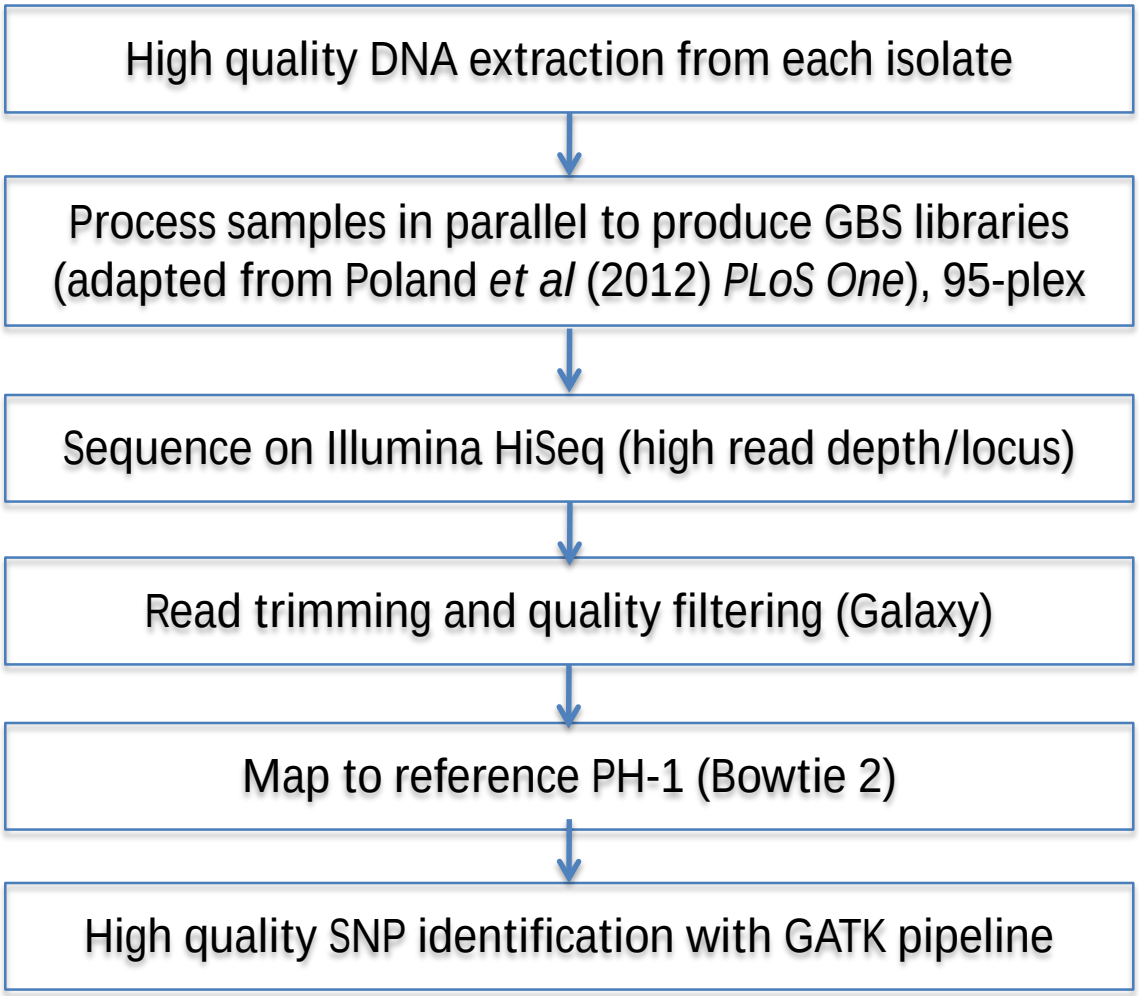


figure adapted from Davey *et al.* (2004) *Nature Reviews Genetics* **12**:499

Methods: from DNA to GBS SNPs



Sample composition

Current:

Location	# 3-ADON	# 15-ADON	years	Refs
Montana	0	22	2000	Zeller et al 2004 Molecular Ecology
North Dakota	14	3	2008	Puri and Zhong 2010 Phytopathology
New York	68	125	1998-2013	Zeller et al 2004 Molecular Ecology; Spolti et al 2014 Phytopathology

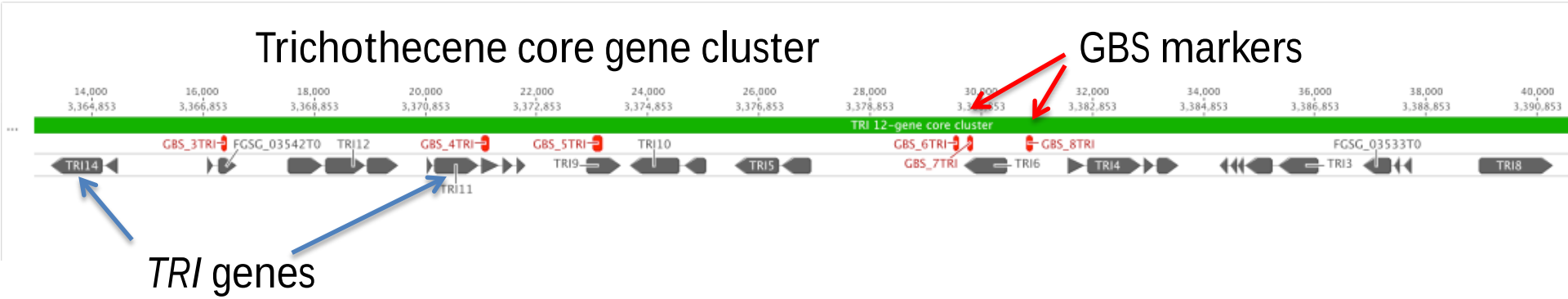
Future:

Additional samples from:

ND, MN from 1999 and 2000

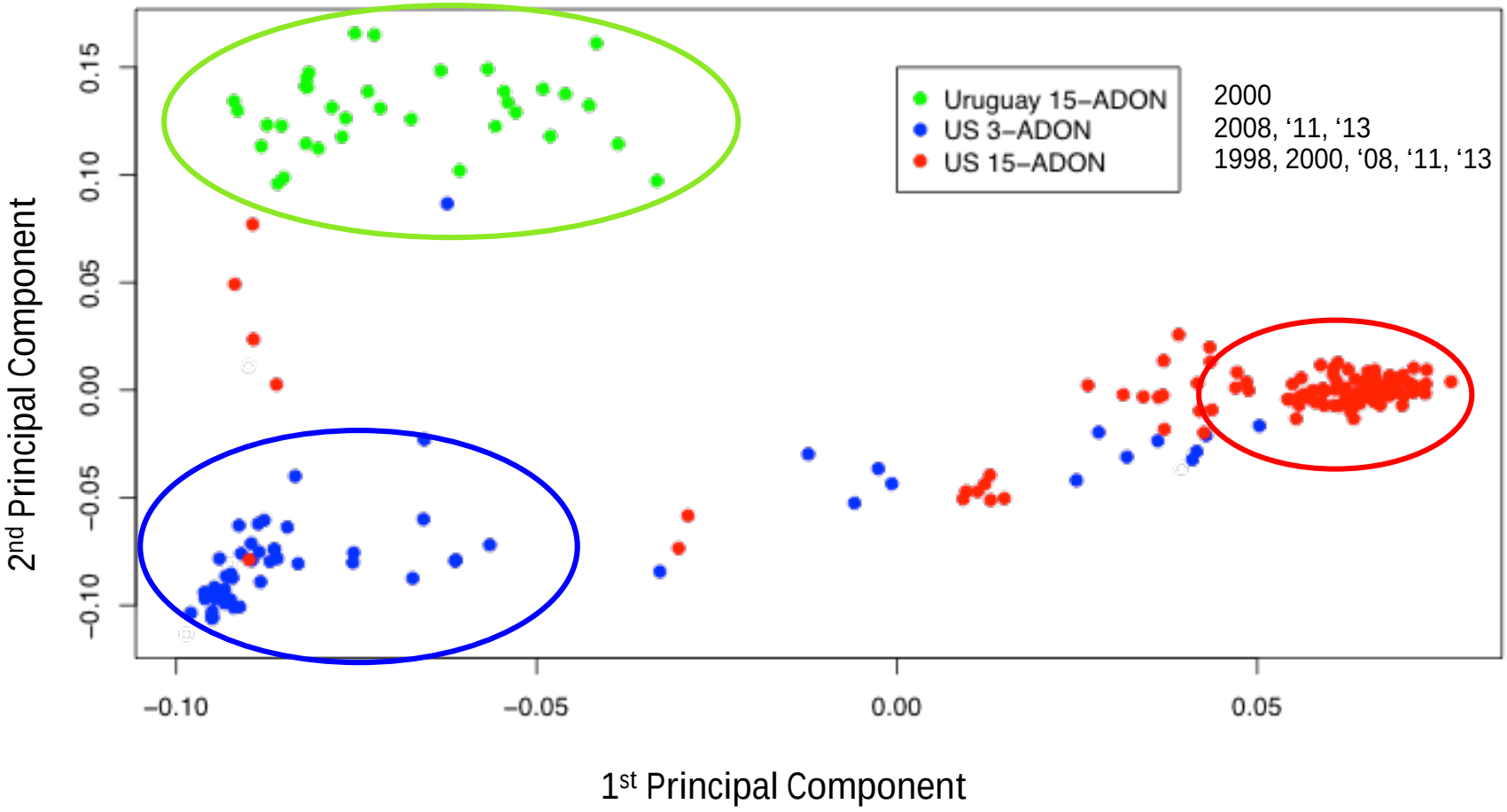
Southeast US (including NIV chemotype)

GBS marker summary

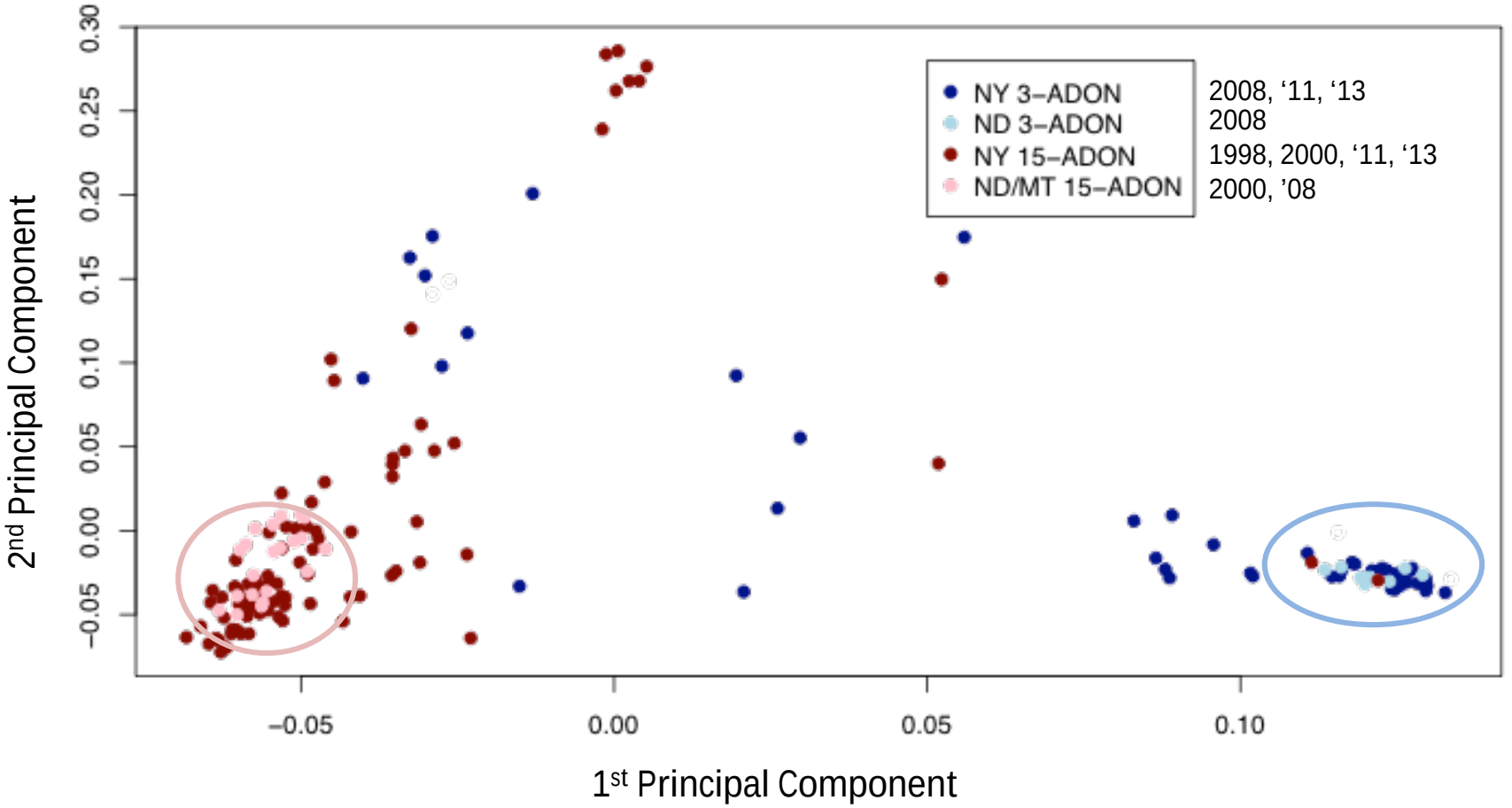


- Over 30,000 loci (< 100 bp) sequenced, and ~20,000 include polymorphisms
- Many loci have missing data in fraction of our sample
- ~16,000 SNPs have genotype data from at least half of the isolates
- Many SNPs are in high LD with neighboring SNPs – 6600 SNPs remain if we filter out these largely 'redundant' SNPs

Aim 2 - Genetic clustering of isolates

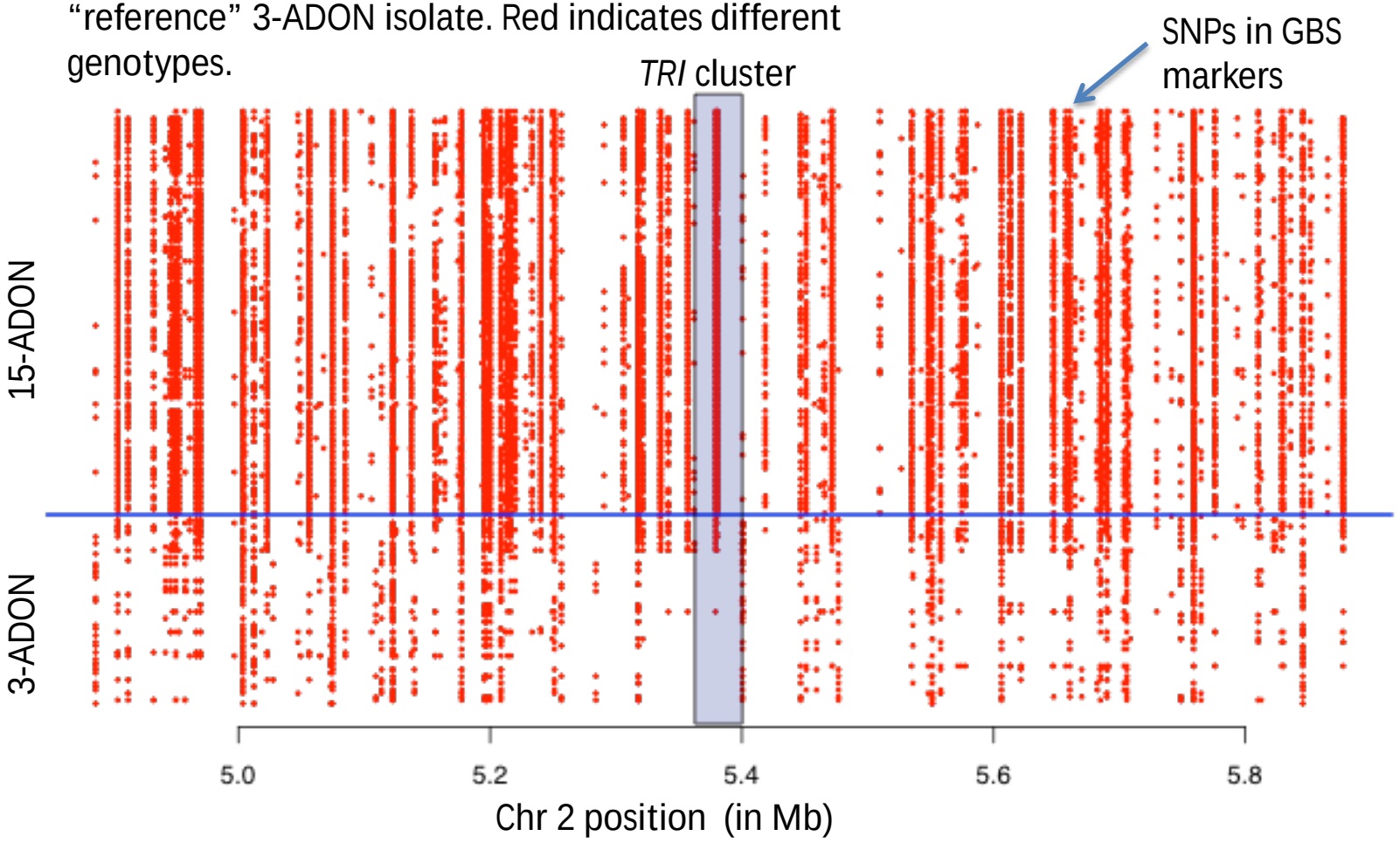


Genetic clustering of US isolates



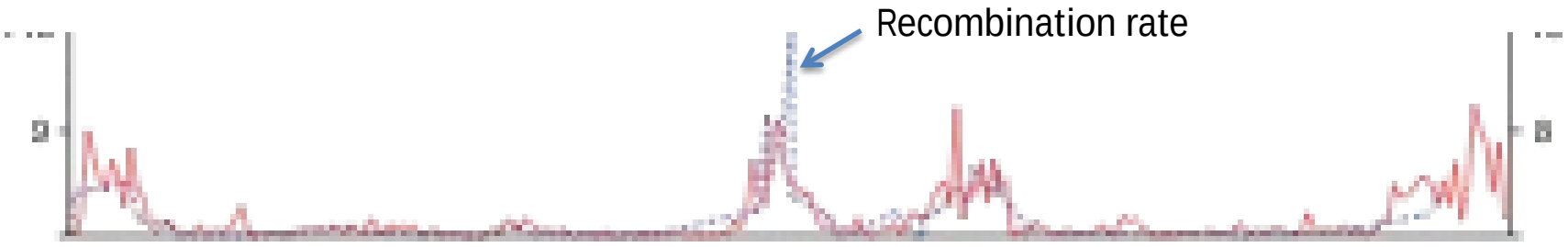
Aim 3 – Genome scans for selection

Each row represents the comparison of marker genotypes from a new isolate (3-ADON or 15-ADON) with a constant "reference" 3-ADON isolate. Red indicates different genotypes.



Aim 4 –Linkage disequilibrium measurement

- Association of alleles between GBS loci
- Genetic basis of traits can be inferred via statistical association between traits and markers (GWAS) - amount of association (LD) between GBS alleles will determine GWAS resolution
- Rates of LD decay with physical distance along chromosomes will depend on local rates of genetic recombination



from Cuomo *et al*. (2007) *Science* **317**: 1400

Preliminary conclusions

- GBS protocol produces 1000's of loci relatively evenly spaced, though polymorphisms enriched in regions of high recombination
- Provides high resolution examination of population structure
- Genome scans offer promise for identifying unexpected fitness-related genes that may be targets for pathogen control
- Trait measurements from our genotyped samples can lead to further gene identification via GWAS

Request for additional samples

- Field-collected isolates, especially if you have measured traits for them or performed preliminary genotyping
- Conversely, if you are looking for isolates from which to measure traits, our samples will have dense genotype data available
- And, suggestions about which samples we target are also appreciated