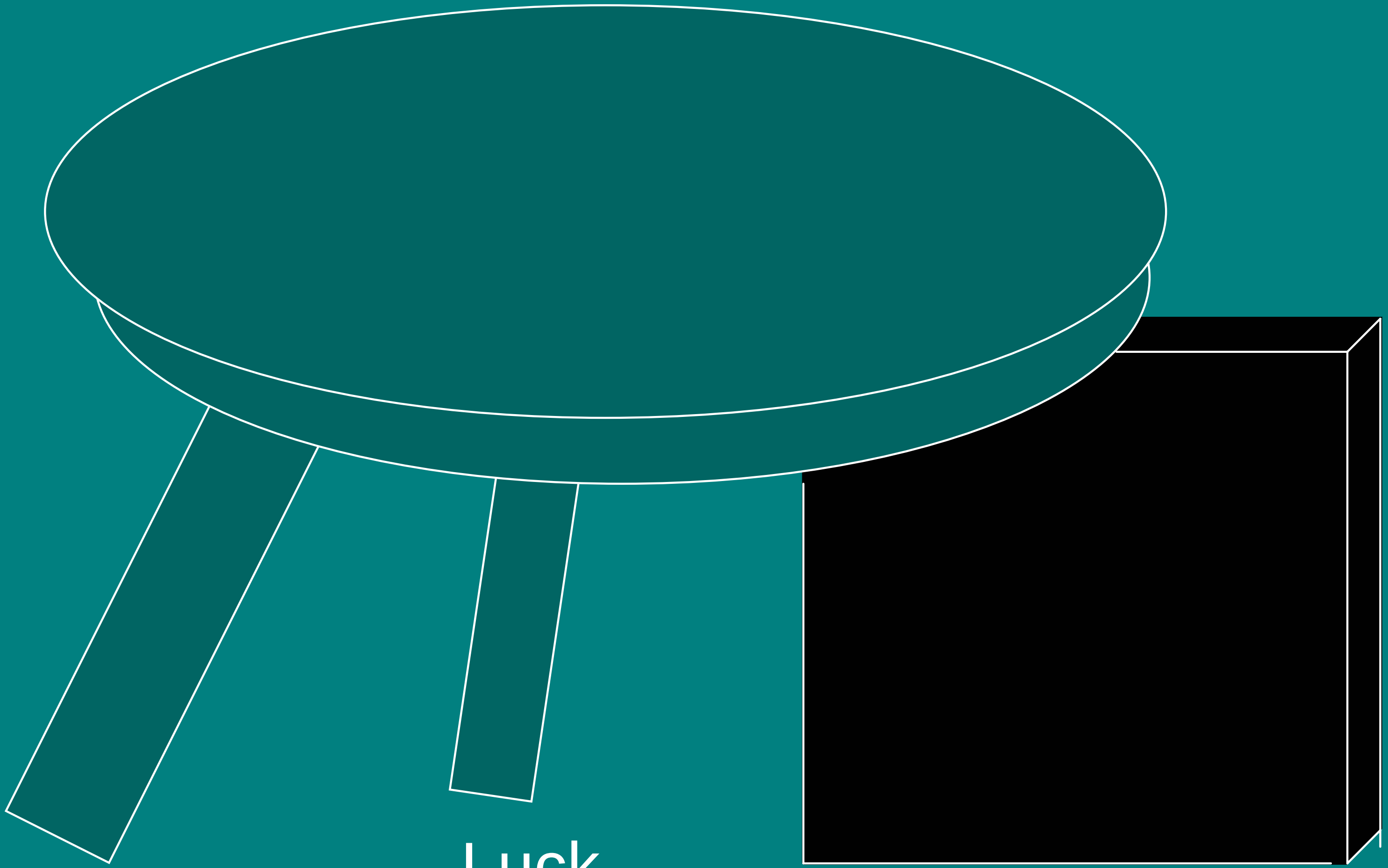
# EAAP
# 2007
# Dublin

# Applications of Horse Genomics

**Ernie Bailey**
**University of Kentucky**

# Three Legged Stool Problem



Luck

Genetics

Management

# Applications of Genetics/Genomics

1. Selection
2. Management
3. Products

# What are the areas of application?

Performance

Behavior

Immunology products… vaccines, adjuvants

Hereditary Diseases

Non-Hereditary Diseases

Appearance and Hair Color

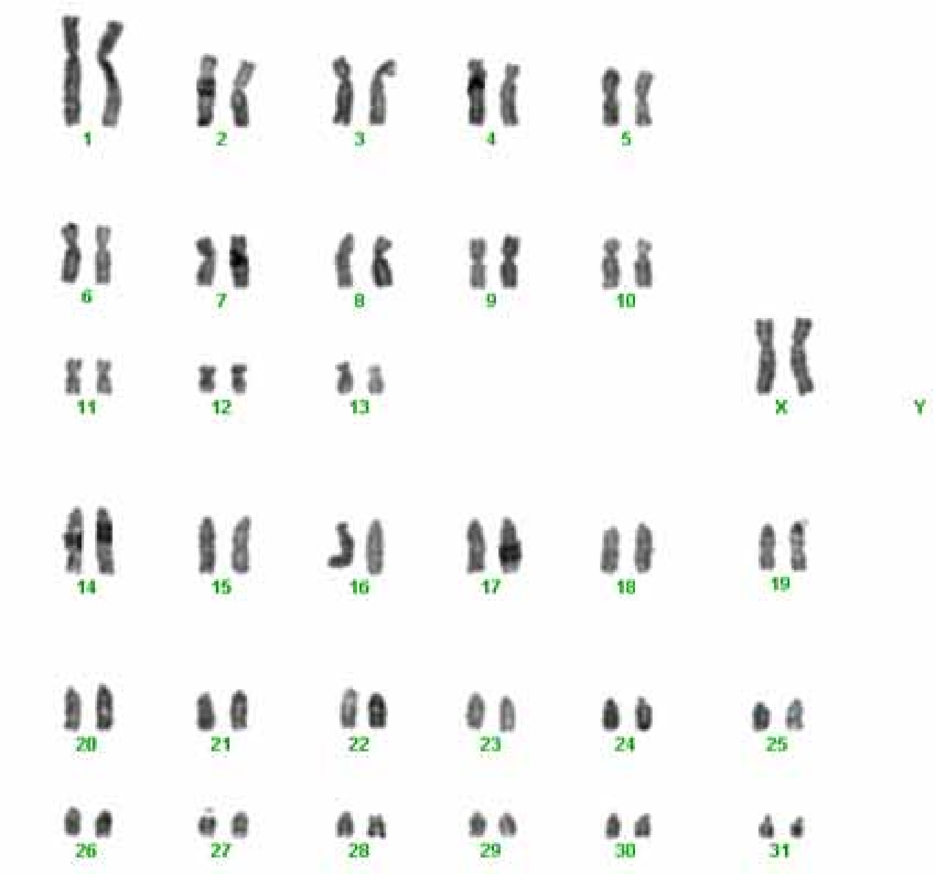
… ?

Source:http://www.biped.ri.cmu.edu

# 1959 - The First View of the Horse Genome

## Cytogenetics

Chromosome number for horse 2n=64



# 1960s – Horse Blood Groups

Effective parentage tests

# How did the Horse Genome Project Start?

1990: Plan for Human Genome Sequence

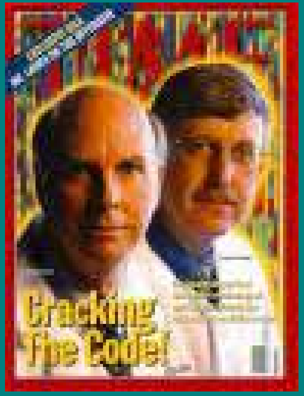
1992: USDA Animal Gene Mapping

1995: 1st Horse Genome Workshop

1996: Horse added to USDA Animal Gene Mapping Projects -- <u>as an International Program</u>

2003: Human Genome Sequence Completed

Source: www.horsepresence.com



# How has the Human Genome Project been Informative & of Benefit to People?

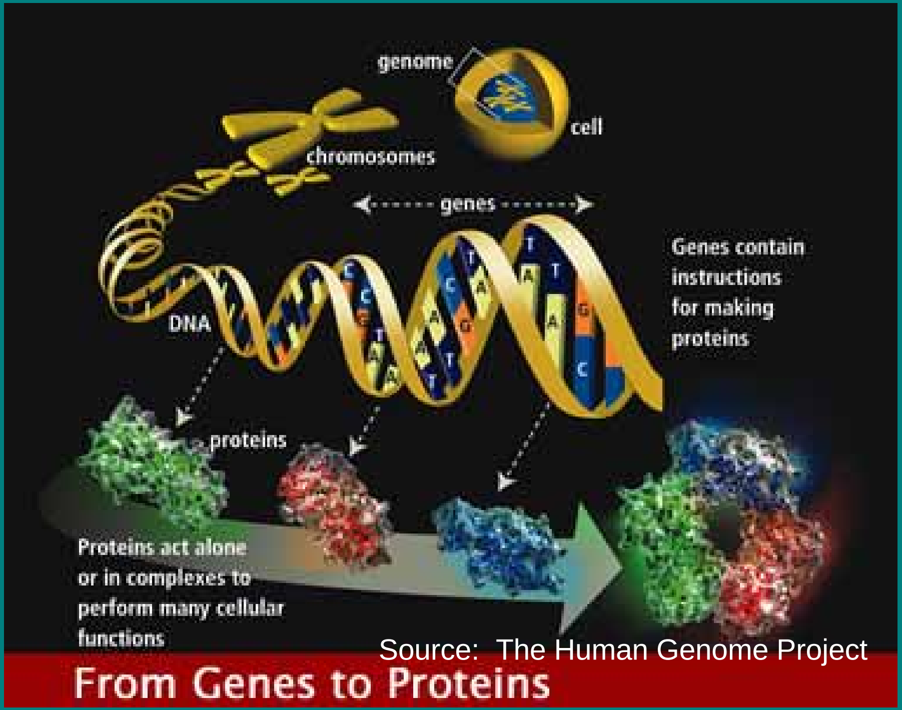
**Information:**

- 99.9% Same for All People
- 2% encodes proteins
- 20,000 genes
- Complexity due to gene regulation, not gene number!!

**Benefits:**

- Genetic Tests
- Individualized Drug Therapy

**Mutations** cause Muscular Dystrophy, Cystic Fibrosis, Neurofibromatosis, Sickle Cell Anemia, Cancers, etc.



Source: The Human Genome Project

Mapping the human genome? Easy.
Mapping downtown? That's another story.

*To learn more about the city that's been home to so much groundbreaking medical research, visit boston04.com* CELEBRAT

# Animals Protest NHGRI Priorities

By Martha J. Heil

BETHESDA, MD — Responding to an NHGRI report naming six "high-priority" organisms for genome sequencing, thousands of animals took to the NIH campus here today to protest.

The animals, including the *Xenopus* frog, pig, and duck-billed platypus, insisted that the selection of some organisms over others amounted to genetic discrimination. "It's bad enough that we are burdened with being the universal metaphor for everything filthy," said the pig with a snort.

Protesters sang, croaked, buzzed, squealed, and threw their own feces at researchers entering the NHGRI labs. "No wonder we're going mad," said the cow, threatening to hold a milk strike.

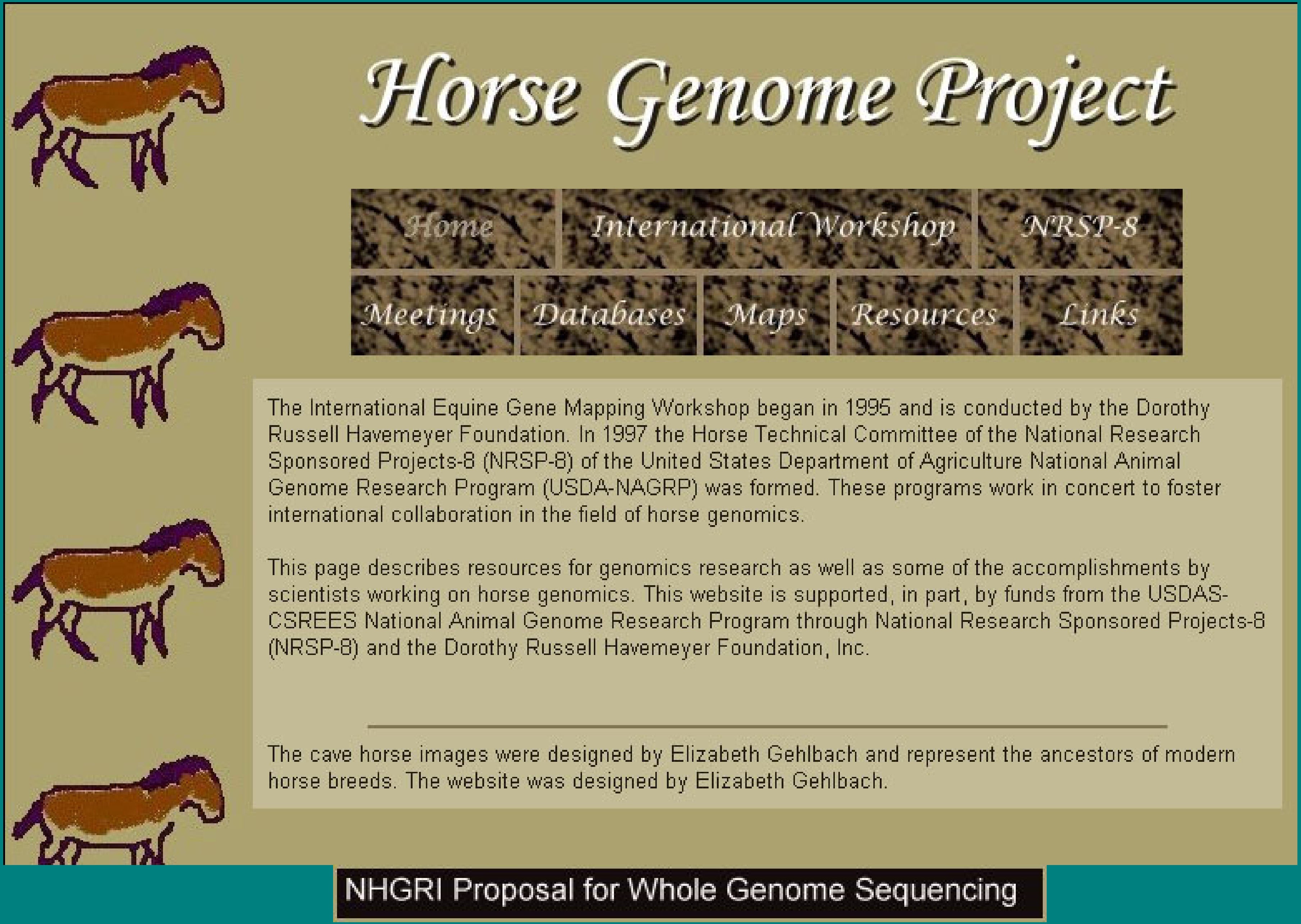
Why did NHGRI sequence the horse?

Why didn’t they sequence a zebra
or rhinoceros?

# Make the case for sequencing horses

Gerard Guerin, France;
Telhisa Hasegawa, Japan; Kamal Khazanehdari, Dubai;
Bhanu Chowdhary, Texas;
Ernie Bailey, Kentucky

# http://www.uky.edu/Ag/Horsemap



The International Equine Gene Mapping Workshop began in 1995 and is conducted by the Dorothy Russell Havemeyer Foundation. In 1997 the Horse Technical Committee of the National Research Sponsored Projects-8 (NRSP-8) of the United States Department of Agriculture National Animal Genome Research Program (USDA-NAGRP) was formed. These programs work in concert to foster international collaboration in the field of horse genomics.

This page describes resources for genomics research as well as some of the accomplishments by scientists working on horse genomics. This website is supported, in part, by funds from the USDAS-CSREES National Animal Genome Research Program through National Research Sponsored Projects-8 (NRSP-8) and the Dorothy Russell Havemeyer Foundation, Inc.

---

The cave horse images were designed by Elizabeth Gehlbach and represent the ancestors of modern horse breeds. The website was designed by Elizabeth Gehlbach.

# Molecular Tests before Sequencing

1. **Hyperkalemic Periodic Paralysis** - HYPP (Quarter Horses)
2. **Severe Combined Immunodeficiency** - SCID (Arabians)
3. **Coat Color** - Chestnut, Tobiano, Sabino, Palomino, Bay, Black, White, Grey
4. **Overo Lethal White Foal Disease** - LWF (Paints)
5. **Junctional Epidermolysis Bullosa** - JEB (Draft)
6. **Glycogen Branching Enzyme deficiency** - GBE (Muscle)
7. **Sex Determining Region Y** - SRY (Fertility)
8. **HERDA** – (Quarter Horses)
9. **Parentage Testing** - Microsatellite DNA markers

**(Low Hanging Fruit - Simple Genetics)**

# Traits with Complex Inherited Components

- ***Developmental Bone Diseases*** **– OCD, Wobblers**
- ***Congenital Disorders*** **– Contracted Foal Syndrome, Microopthalmia, Dwarfism, Parrot Mouth, EI**
- ***Muscle diseases*** **– "Tying Up"**
- ***Response to Infectious Pathogens*** **– EPM, Parasites, Strangles, West Nile, EVA carrier state, etc.**
- ***Understanding Aging*** **– Arthritis, Cushing's**
- ***Allergic Diseases*** **– Summer Eczema, COPD**
- ***Laminitis***
- ***Performance*** **– strength, speed, endurance jumping, athleticism**
- ***Behavior*** **– gaits, stable vices, demeanor**

Source: www.horsepresence.com

# Time Table for Horse Genome Sequencing

- Proposal submitted July 2005
- Broad asks for DNA Nov 2005
- Broad begins sequencing Feb 2006
- NHGRI announce July 19, 2006
- 6.8X sequence July 2006

- First Assembly February 2007
- Final Assembly Fall 2007

# Complete Sequence for inbred female Thoroughbred: 6.7X coverage

**Genetic Variants**: Spot sequence 7 breeds,

Thoroughbred, Standardbred, Quarter Horse, Arabian Horse, Akal-Teke, Icelandic and Andalusian,

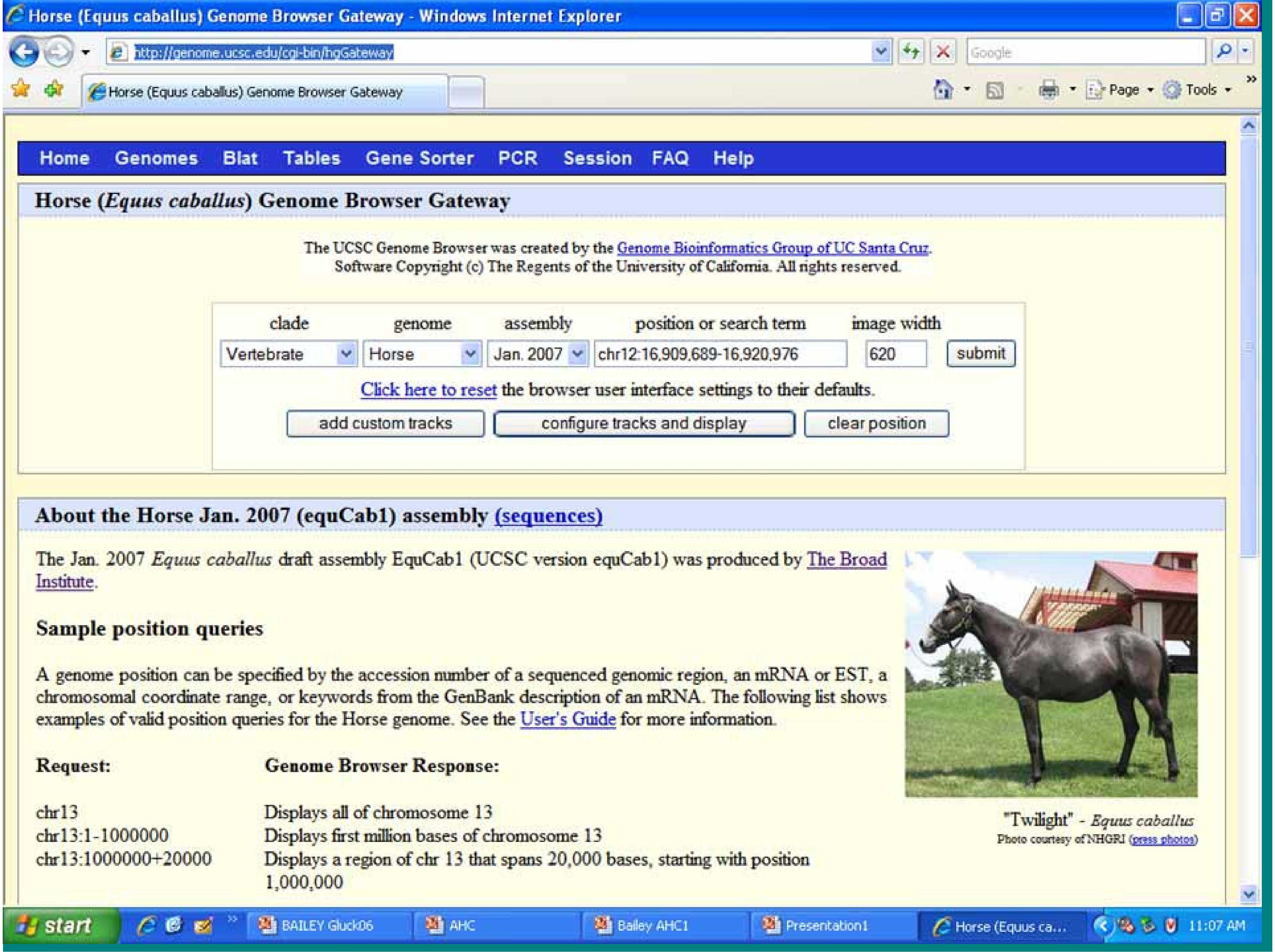
at 100,000 sites: 1.47 million single nucleotide Polymorphisms (SNPs)

# Do SNPs occur randomly or in blocks?

100 SNPs compared in 10 2MB regions
for 24 horses of 11 different breeds:

Thoroughbred, Standardbred, Quarter Horse, Arabian,
Andalusian, Icelandic, Hannoverian, Hokkaido,
French Trotter, Norwegian Fjord, American Saddlebred.

Saw average of 19 Blocks of SNPs for
10 regions…
but only 5 within a particular breed.



# Horse (*Equus caballus*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

| clade | genome | assembly | position or search term | image width |
|---|---|---|---|---|
| Vertebrate | Horse | Jan. 2007 | chr12:16,909,689-16,920,976 | 620 |

Click here to reset the browser user interface settings to their defaults.

## About the Horse Jan. 2007 (equCab1) assembly (sequences)

The Jan. 2007 *Equus caballus* draft assembly EquCab1 (UCSC version equCab1) was produced by The Broad Institute.

### Sample position queries

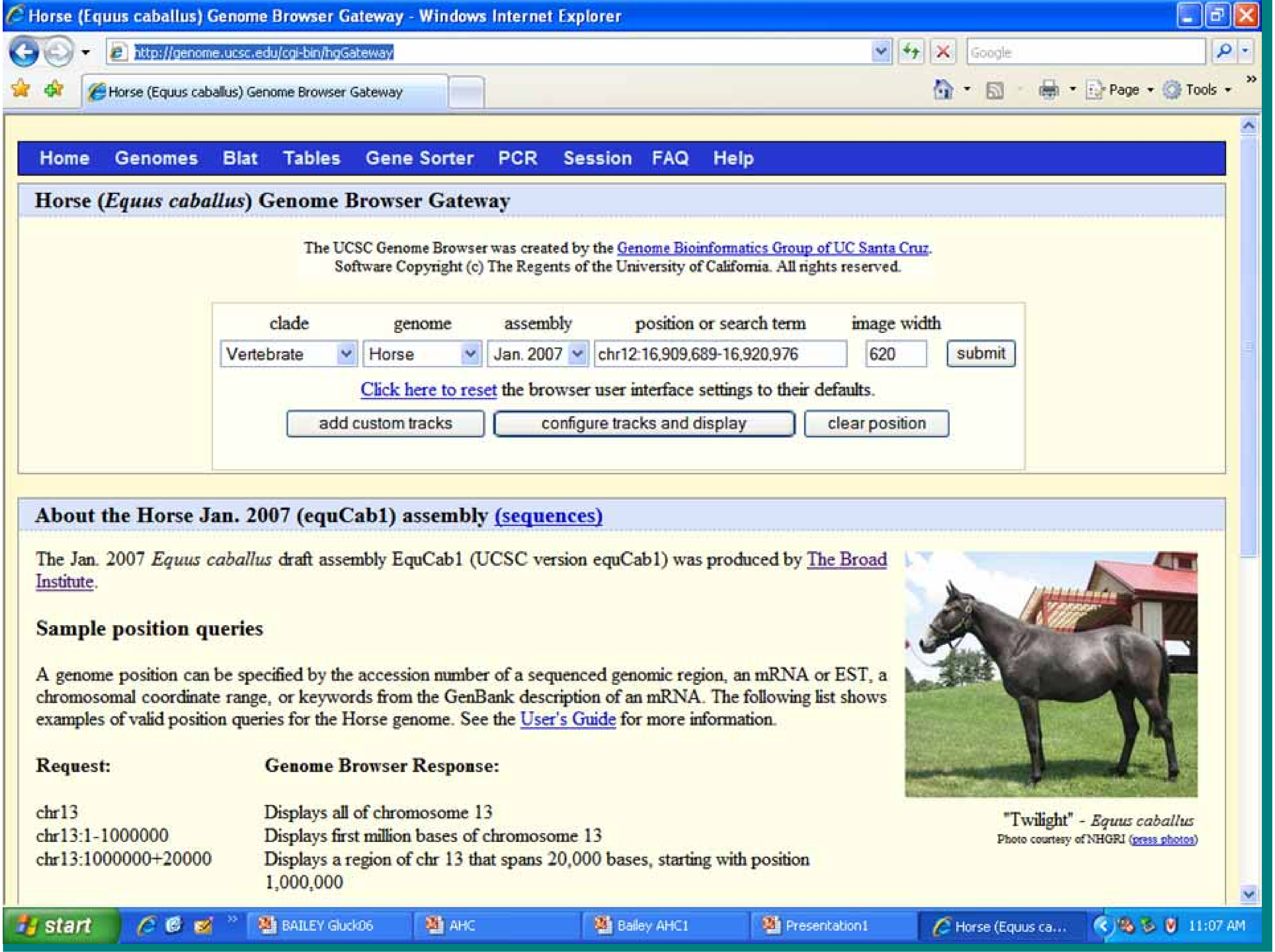
A genome position can be specified by the accession number of a sequenced genomic region, an mRNA or EST, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the Horse genome. See the User's Guide for more information.

| Request: | Genome Browser Response: |
|---|---|
| chr13 | Displays all of chromosome 13 |
| chr13:1-1000000 | Displays first million bases of chromosome 13 |
| chr13:1000000+20000 | Displays a region of chr 13 that spans 20,000 bases, starting with position 1,000,000 |

"Twilight" - *Equus caballus*
Photo courtesy of NHGRI (press photos)

# How to use the tools?

1. Database - Find DNA sequences
2. SNP Arrays
3. Expression Arrays



## Horse (*Equus caballus*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

| clade | genome | assembly | position or search term | image width |
|---|---|---|---|---|
| Vertebrate | Horse | Jan. 2007 | chr12:16,909,689-16,920,976 | 620 |

Click here to reset the browser user interface settings to their defaults.

## About the Horse Jan. 2007 (equCab1) assembly (sequences)

The Jan. 2007 *Equus caballus* draft assembly EquCab1 (UCSC version equCab1) was produced by The Broad Institute.

### Sample position queries

A genome position can be specified by the accession number of a sequenced genomic region, an mRNA or EST, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the Horse genome. See the User's Guide for more information.

| Request: | Genome Browser Response: |
|---|---|
| chr13 | Displays all of chromosome 13 |
| chr13:1-1000000 | Displays first million bases of chromosome 13 |
| chr13:1000000+20000 | Displays a region of chr 13 that spans 20,000 bases, starting with position 1,000,000 |

"Twilight" - *Equus caballus*
Photo courtesy of NHGRI (press photos)

# Genetic Studies: SNP Arrays

- 1.4 million base substitions identified

  (Single Nucleotide Polymorphism or SNP)

- 1 SNP found about every 1200 base pairs

- Use depends on

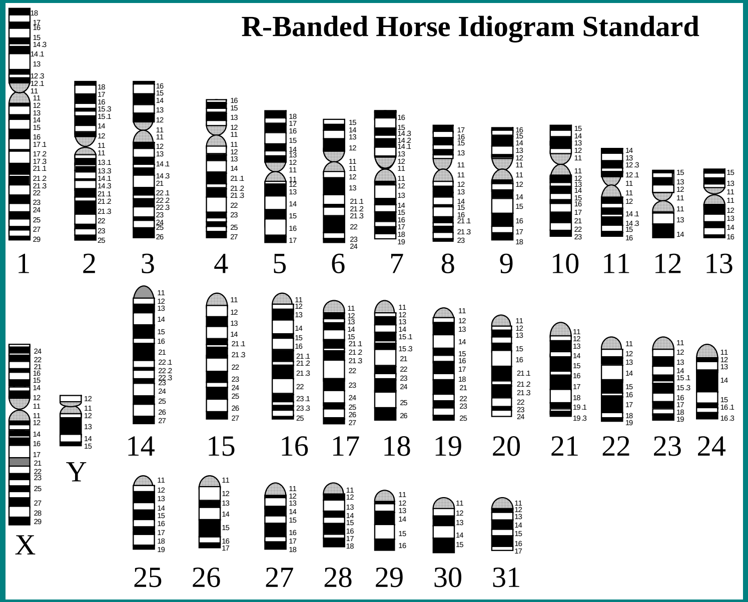
  "Linkage Disequilibrium" (LD)

## What is LD and how can we use it?

# Chromosomes: the primary units of linkage

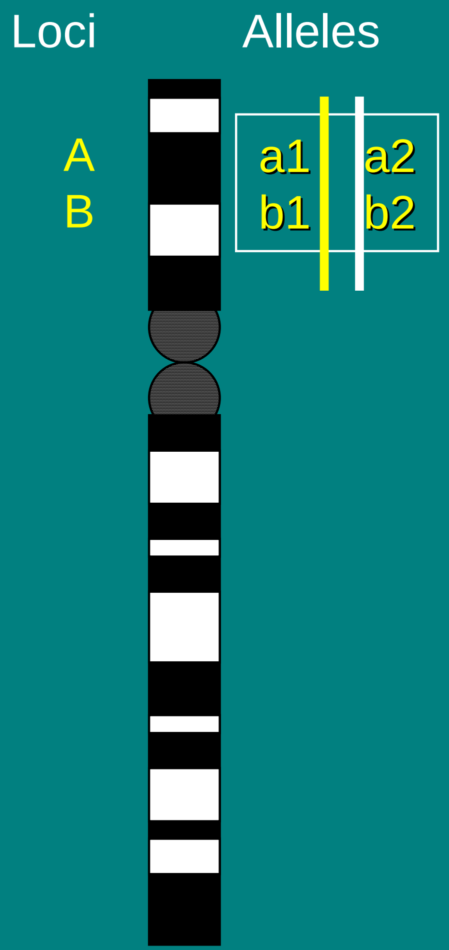
Meiotic recombination decreases the size of the linkage unit every generation.

The size of the linkage unit is called LD and varies between populations.



R-Banded Horse Idiogram Standard

# Consider two linked loci… A & B...
# -- Can the allele at A be used to predict the allele at B?

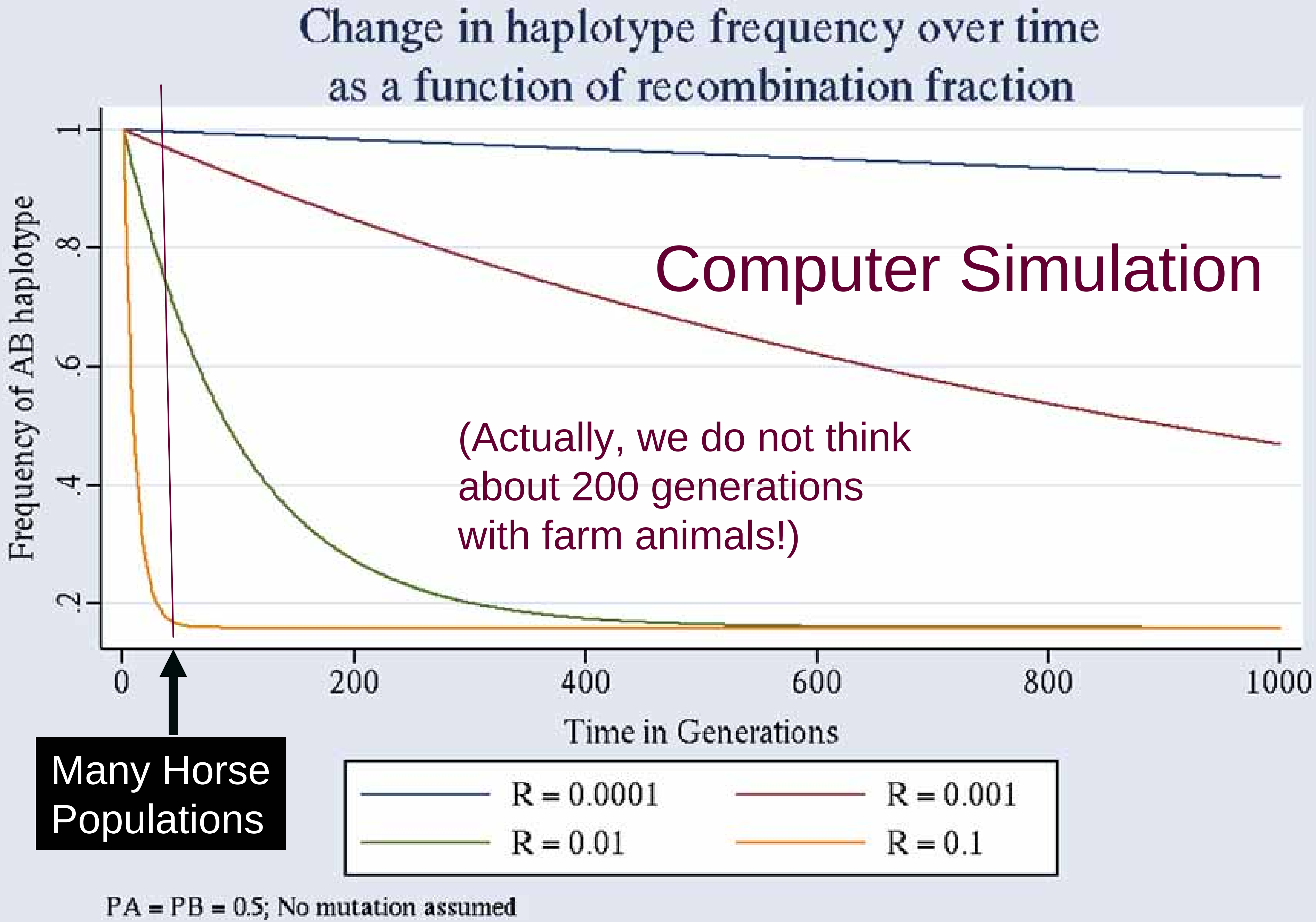


Loci

Alleles

A
B

| | |
|---|---|
| a1 | a2 |
| b1 | b2 |



Computer Simulation

(Actually, we do not think
about 200 generations
with farm animals!)

Many Horse
Populations

From http://slack.ser.man.ac.uk/theory/ld_gen.html

**Humans have small LD… large populations separated by many generations.**

**Dogs have large LD… small populations and separated by fewer generations**

# Horses…. Intermediate sized LD

- tougher to find genes than in dogs
- better model for humans than dogs

# For horse research…

We can afford 60,000 SNPS.

Therefore we need lots of horses for association studies…

- several hundred for simple traits
- perhaps a thousand for complex traits

# Association Studies to begin by Spring 2008

OCD, Recurrent Airway Disease,
Tying up, Lordosis, Dwarfism,
Contracted Tendons, Fractures
and other conditions.

# Physiology & Genetics:  Expression Arrays

- DNA chip with DNA sequences from all 20,000 genes.

- Isolate mRNA from tissue, convert to cDNA and assay presence, absence & amount

- Assays are quantitative and qualitative.

- Verify with qRT-PCR

# Expression Arrays

…by the math

- 20,000 genes
- 10,000 expressed in a cell at a time
- Tissues composed of many cell types
- Response to treatment may be variable for time of expression and amount.

**Is the measured response cause or effect?**

**Expression arrays are currently used to study cartilage function, performance, inflammation and response to infectious diseases.**

**This is the tool for nutrigenomics.**

# Many opportunities for intelligent choices!

**Combined strategy:**

1) Use SNP array study to identify chromosome region with QTL.

2) Use expression array to find gene in QTL region with differential expression.

-Learn how to use the information available online. http://www.genome.ucsc

-Think inside the black-box.

-If not for us…

for our students!

# Supporters of the Equine Genome Workshop

MORRIS ANIMAL FOUNDATION
Est. 1948

Havemeyer Foundation

USDA

# Acknowledgements

NHGRI

Kerstin Lindblad-Toh
Claire Wade
Bruce BIrren
Broad Institute Staff

Workshop

Cornell
UC Davis
TAMU
UMN
UGA
OSU
Kentucky
AHT-U.K.
RVC- U.K.
Japan
INRA-France
The Netherlands
Switzerland
Australia
China
Poland
New Zealand
South Africa
Ireland-Weatherbys
Ireland-UCD
Czech Republic
Germany
Italy