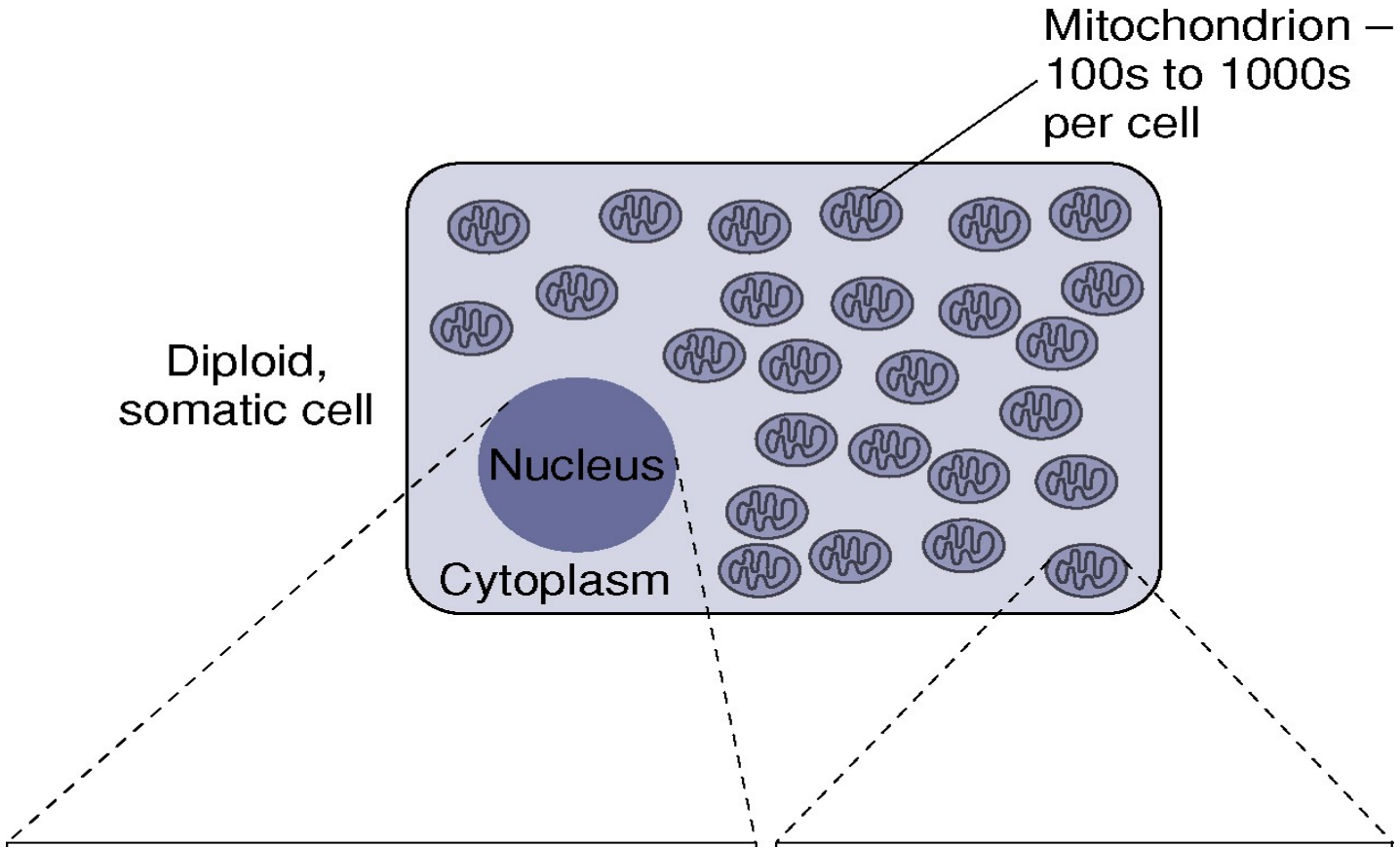
Is there evidence for selection in functional pathways across chromosomes?

Gyan Bhanot
Rutgers University
IAS Princeton
Cancer Institute of New Jersey
gyanbhanot@gmail.com

BIOLOGY INTRO = Longer talk

SKIP INTRO = Shorter talk

Structure of Diploid Somatic Cell



Nuclear genome

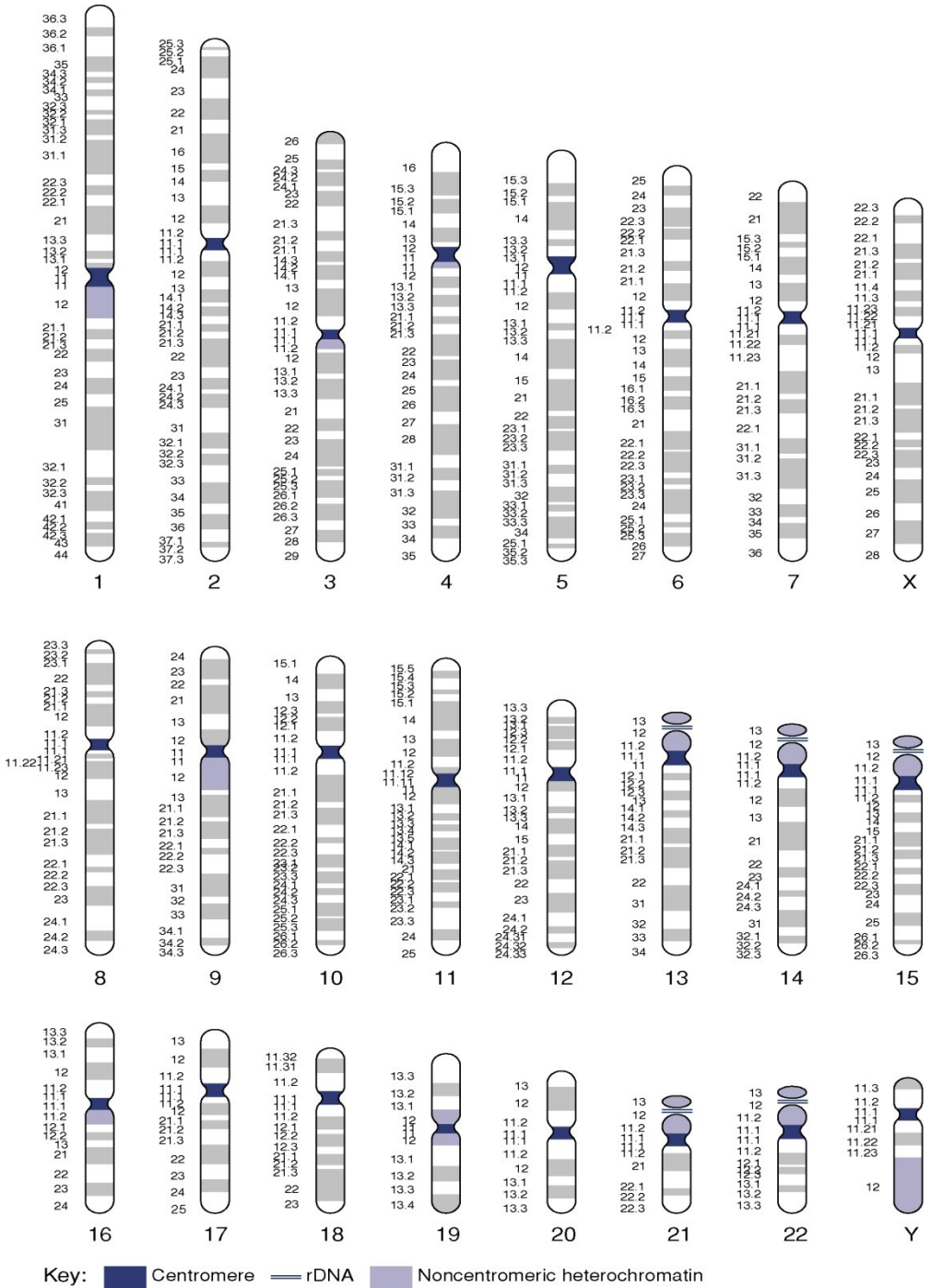
- 3200 Mbp (haploid genome size)
- 2 copies per diploid cell
- 46 linear molecules (chromosomes) per diploid cell

mtDNA

- 16569 bp
- 2–10 copies per mitochondrion, 1000s of copies per cell
- Circular molecule

Human Karotype

http://www.hsls.pitt.edu/about/news/hslsupdate/2005/april/ncbi_gen_bio_res/



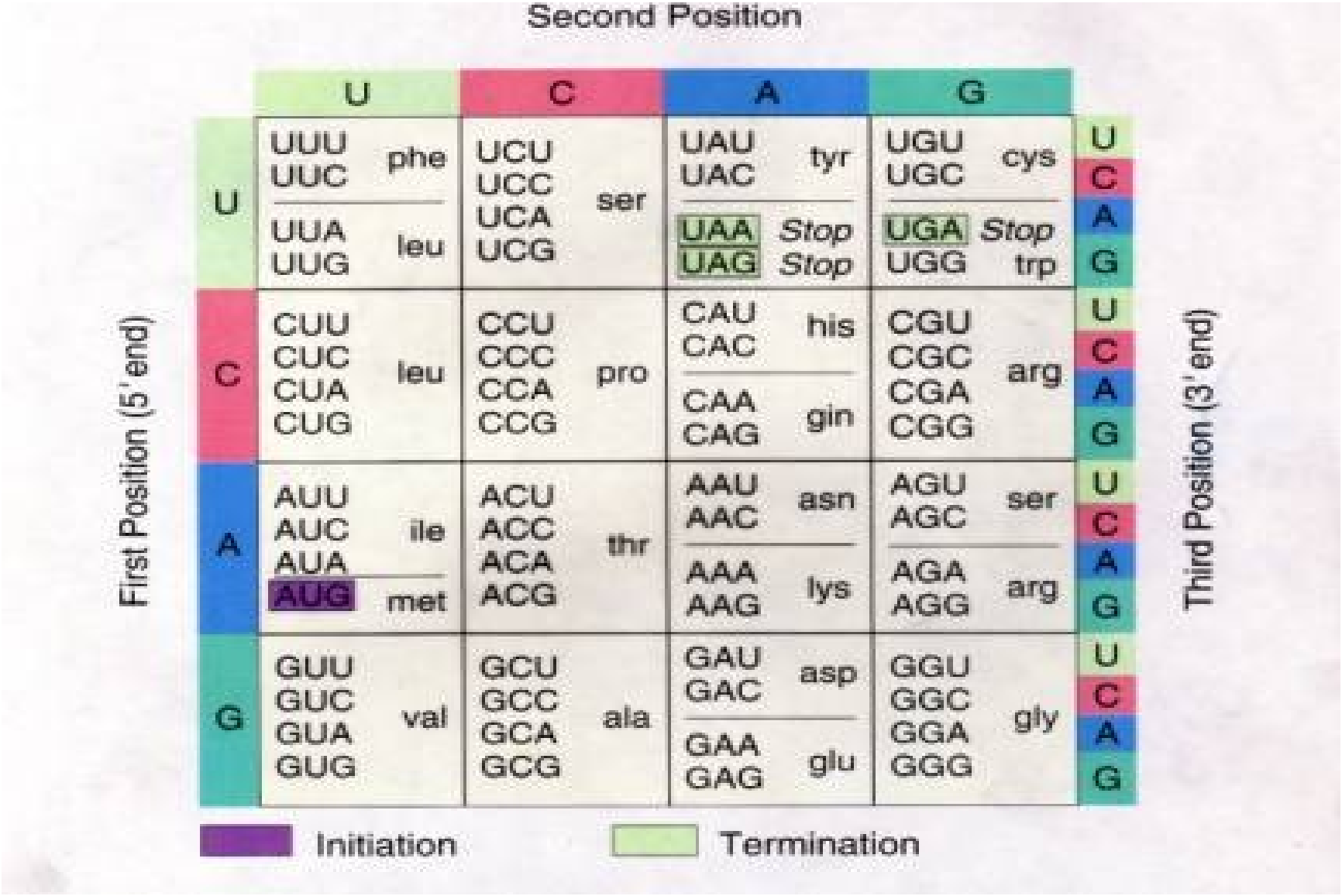
Properties of DNA

- DNA is a polymer consisting of units.
- Each unit has a sugar molecule (deoxyribose), a phosphate group and a base.
- The bases are monomeric subunits called nucleotides which are one of the following: adenine, guanine, cytosine and thymine or A, G, C, T
- A, G are double ringed and are called purines
- C, T are single ringed and called pyrimidines
- The sugar and phosphate link end to end to form a helical shape. Two helices inter-twine with the bases on the inside
- A preferentially binds to T and G to C

More on DNA Structure

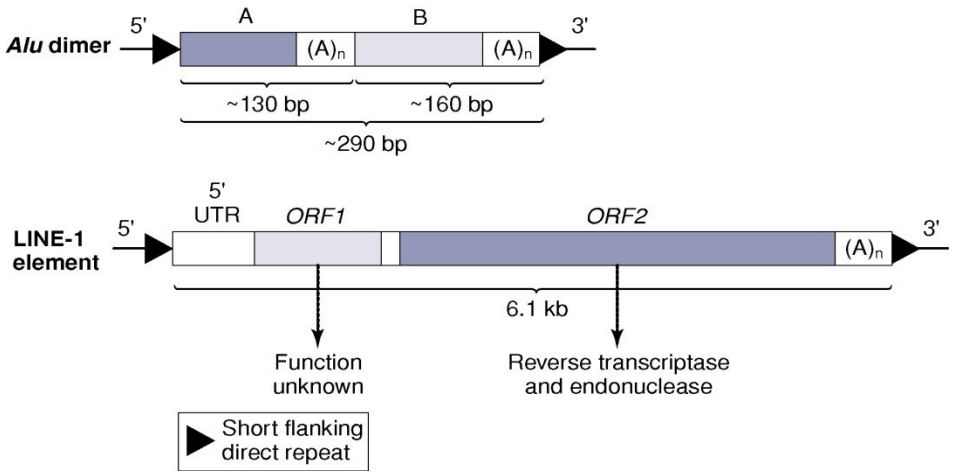
- The bonds that keep the strands together are covalent and hard to break (except at high temperature)
- The bonds between bases are hydrogen bonds. These break (denature) easily in vitro under
  - Brief heating to > 95^{o}C
  - Exposure to alkaline pH
  - Radiation
  - Active energetic processes within cells
- The sequence of the human genome is known and is approximately 3 billion bases long
- It contains approximately 30,000 genes which code for proteins
- The genes code for linear protein molecules consisting of a chain made of different orderings of amino acids
- **DNA is a universal code to make proteins.**

Universal Triplet Code



Other Properties of DNA

- 98.5% of the genome does not code for protein
- 70% of the genome is not transcribed
- Evolve by
  - Segmental duplication and subsequent modification
  - Mutation
- ~ 45% of genome contains highly repetitive sequences with copy numbers in 100s of thousands
  - LINES (Long Interspersed Linear Elements) ~ 7 kbp long
  - SINES (Short Interspersed Linear Elements) ~ few 100 bp long
-



Genes

- Segments of DNA with instructions to make proteins
- Consist of distinct regions containing sequences that act as promoters and enhancers followed by a series of “exons” and “introns”
- Exons are regions whose instructions are used in making proteins
- Introns are generally non functional and removed before translation (with some exceptions)
- The size of genes can vary:
  - SRY gene is only 612 bp and has one exon and no intron
  - Dystrophin gene is 14,000 bp long and has 79 exons.
- Genes are differentially expressed in cells and are modified by: alternate splicing, post transcriptional or , post translational modification

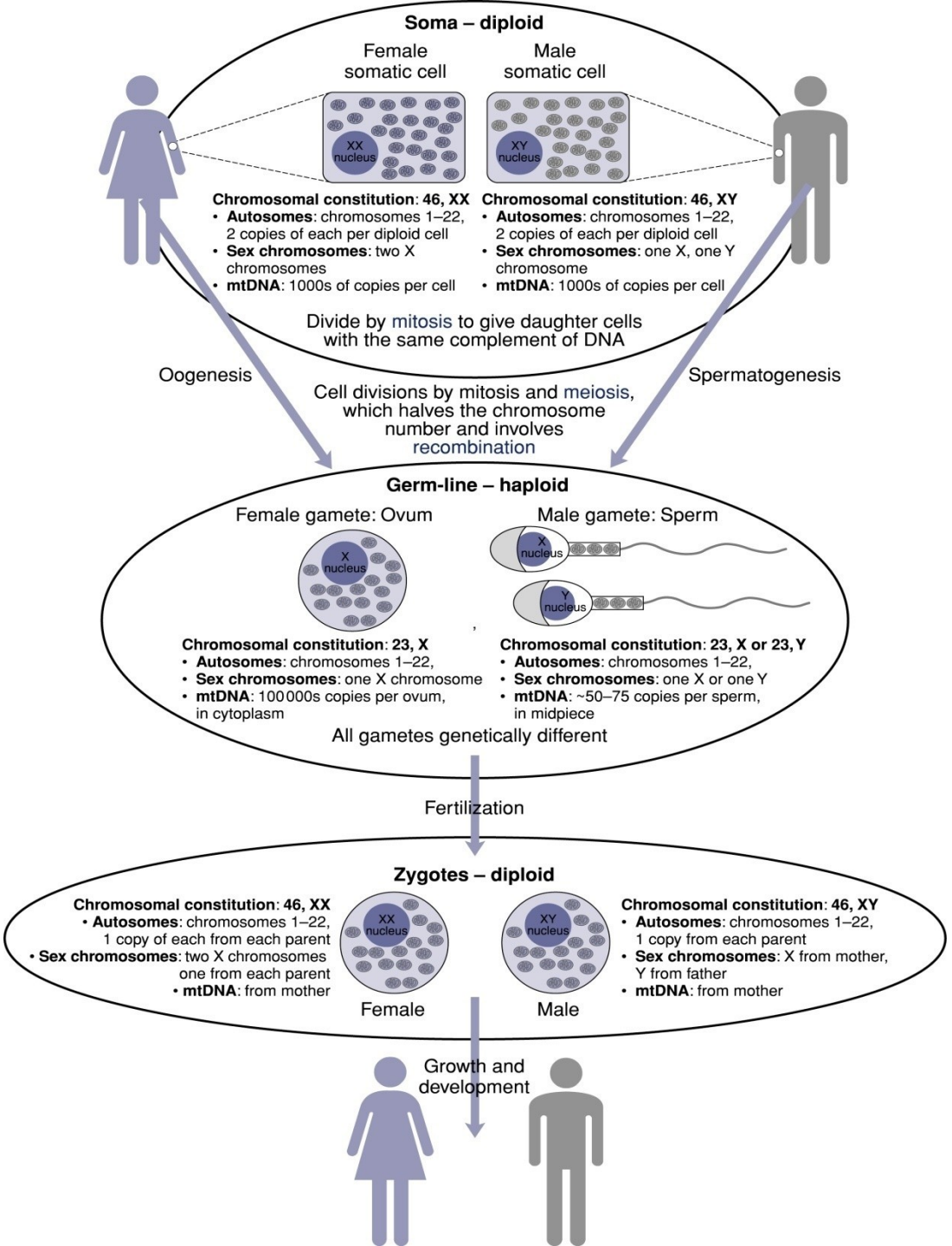
Mitosis and Meiosis

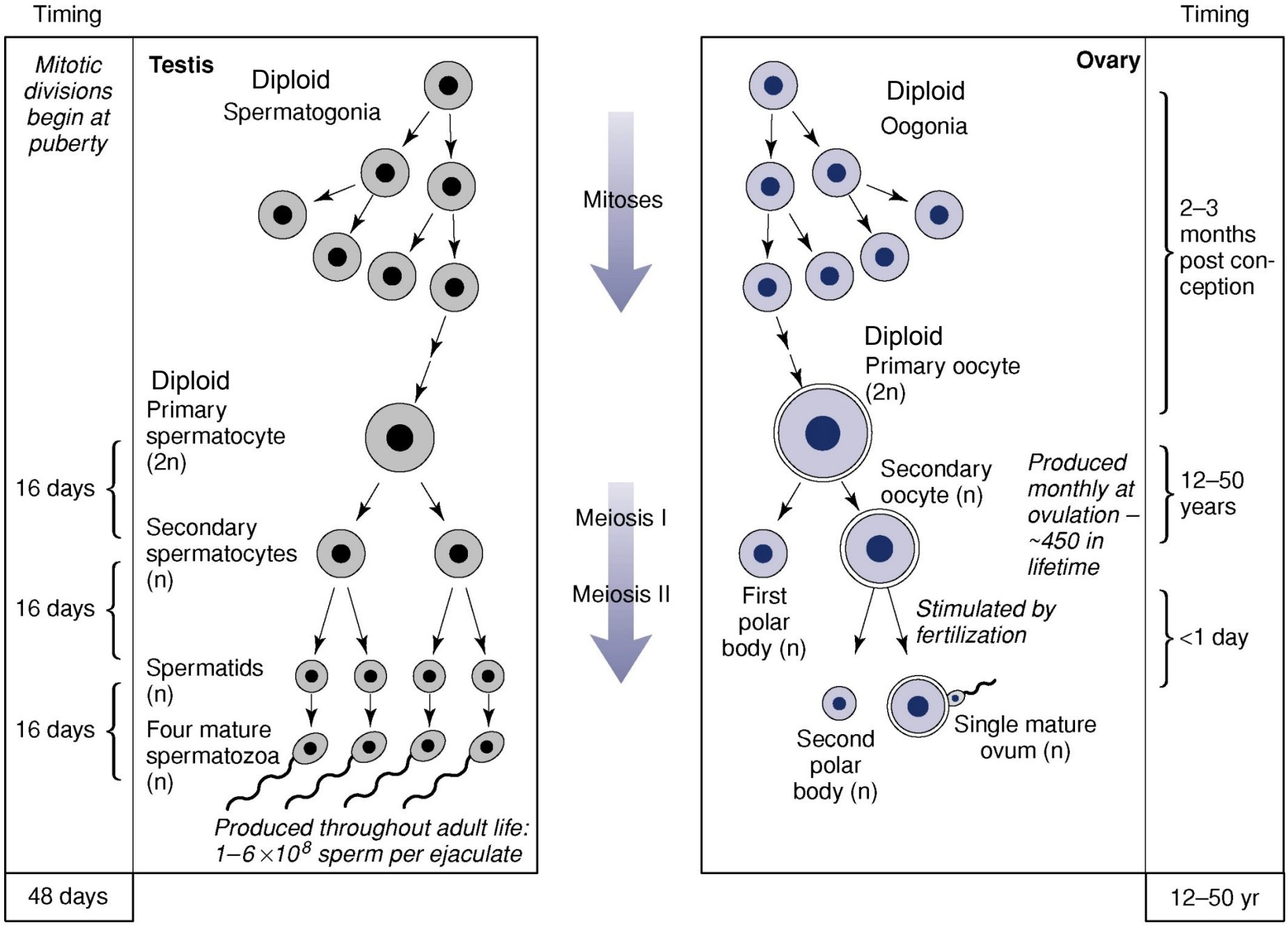
- All humans start life as a single cell which divides and differentiates to create ~ 10^{16} - 10^{17} cells
- Mitosis is cell division to create two diploid copies of cell.
  - One round of DNA replication and 2 daughter cells
  - Errors in mitosis (somatic mutations) may cause diseases eg. Cancers, Alzheimers (plaque).
  - Proceeds through distinct phases G0, G1, S, G2 and M
- Meiosis is process of creation of haploid germline cells.
  - One round of DNA replication and 4 daughter cells
  - Errors in meiosis cause genetically inherited diseases.
  - Includes a recombination step (~ 50 per event in males and 80 per event in females).
  - Recombination is more frequent near telomeres.

Mitosis or Cell Division

- Mitosis is nuclear division plus cytokinesis, producing two identical daughter cells
- Divided into interphase, prophase, prometaphase, metaphase, anaphase, and telophase. Interphase encompasses stages G1, S, and G2 of the cell cycle.
- Mitosis Animation

Inheritence of genetic information

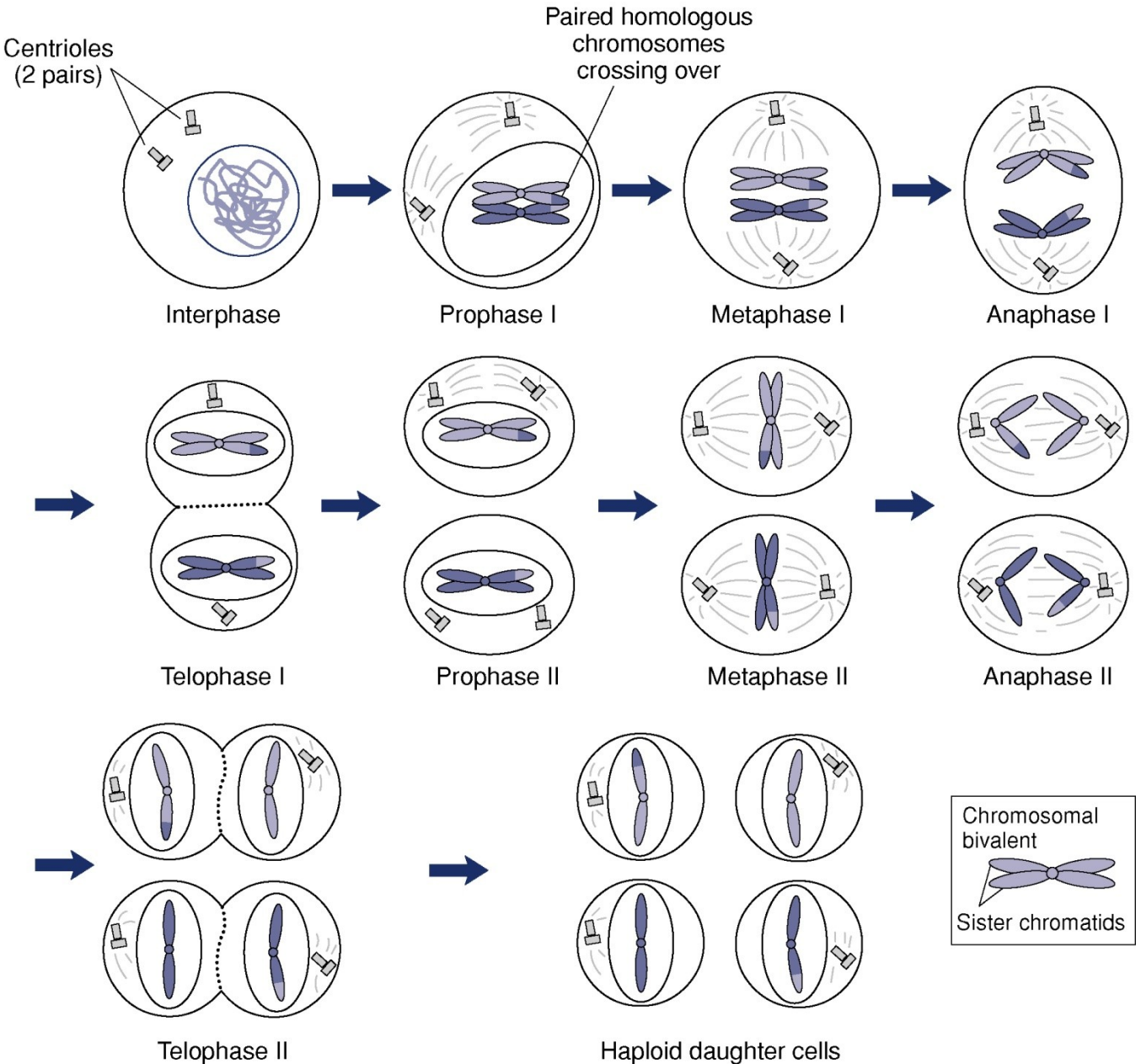



Timing
Timing
Mitotic divisions begin at puberty
Testis
Diploid
Spermatogonia
Mitoses
Diploid
Primary spermatocyte (2n)
16 days
Secondary spermatocytes (n)
16 days
Meiosis I
Meiosis II
Spermatids (n)
16 days
Four mature spermatozoa (n)
Produced throughout adult life: 1–6 ×10^8 sperm per ejaculate
48 days
Ovary
Diploid
Oogonia
2–3 months post con-ception
Diploid
Primary oocyte (2n)
Secondary oocyte (n)
Produced monthly at ovulation – ~450 in lifetime
12–50 years
First polar body (n)
Stimulated by fertilization
<1 day
Second polar body (n)
Single mature ovum (n)
12–50 yr

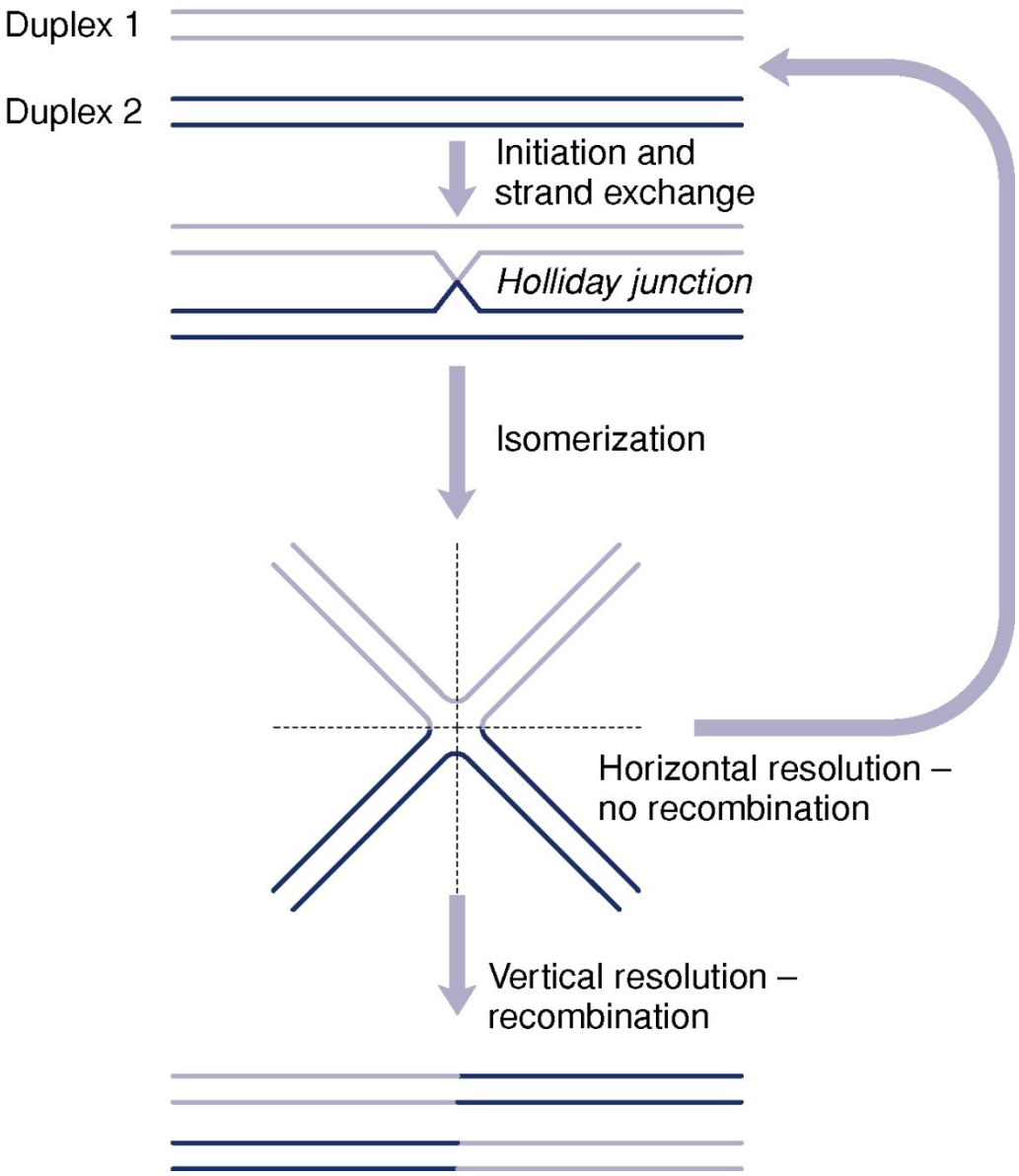
Meiosis

- Meiosis is the process by which a cell divides to produce four haploid germ line cells.
- Meiosis animation

Meiotic Recombination, the great shuffler



Molecular Model of Meiotic Recombination

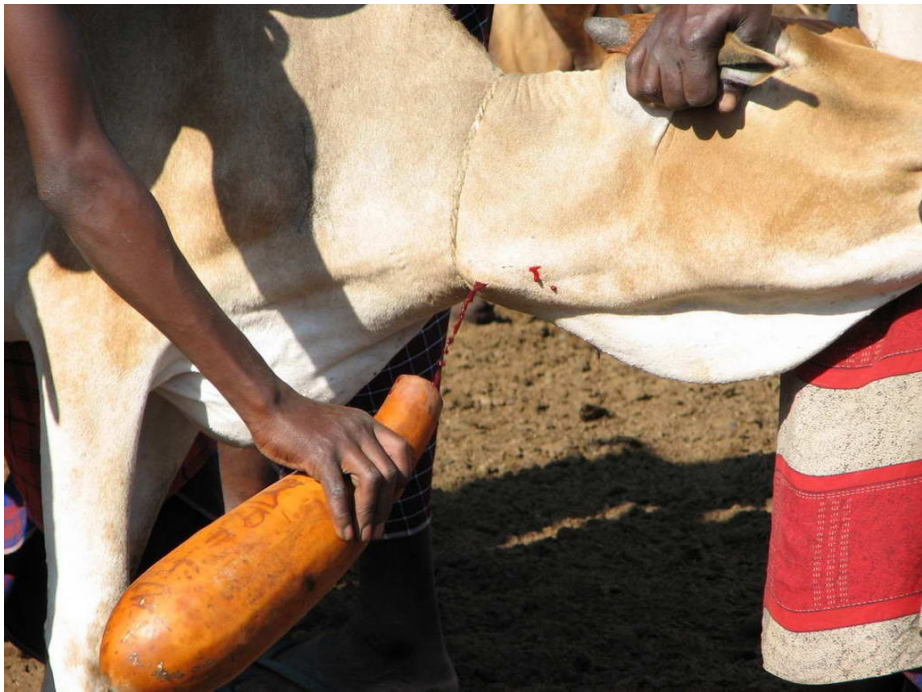


Definitions

- Genotype = diploid bi-allelic form at a specific locus or loci (inherited from both parents)
- Haplotype = allelic form *on the same chromosome* at specific loci (inherited from one parent).
- Phenotype = manifestation of underlying genetic or environmental characteristics (hair color, risk of diabetes, height etc.)

- Homozygous = same nucleotide on same locus in both chromosomes.
- Heterozygous = different nucleotide on the same locus in the two chromosomes.
- Wild-type = genotype of the majority of the population (depends on the definition of "population"). Thus the "wild type" blood group in the Basque is A^-.
- Polymorphism: varying nucleotides at a given locus. By convention a polymorphism is required to have a frequency between (0.05 and 0.95) – WHY?
- Homologs: Similar due to shared ancestry. Commonly used to denote regions of DNA similar by common descent.
- Paralogs: Similar because related by gene duplication event. (If due to whole genome duplication, they are called Ohnologs after Susumu Ohno)
- Orthologs: Homologous sequences separated by a speciation event.

The Maasai have protective polymorphisms against hyperlipidemia in spite of a high fat diet

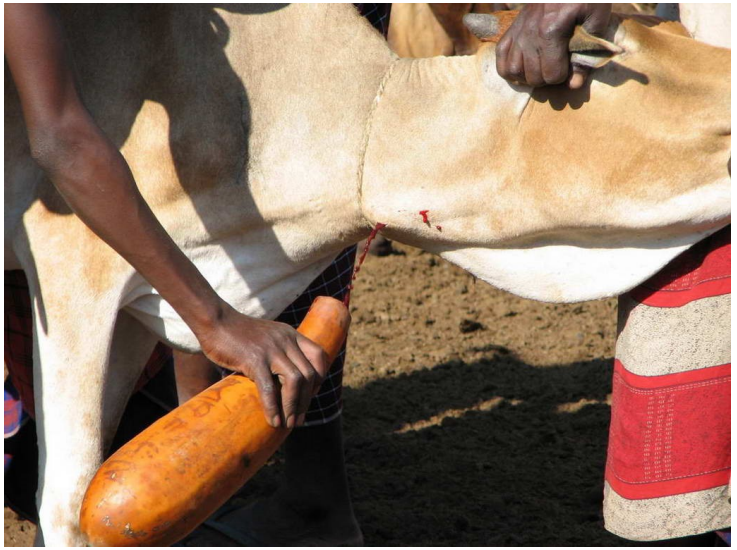


The Masai of East Africa: Some Unique Biological Characteristics

Arch Path—Vol 91, May 1971

Kang-Jey Ho, MD, PhD, Birmingham, Ala; Kurt Biss, MD, Chicago;
Belma Mikkelson, Birmingham, Ala;
Lena A. Lewis, PhD, Cleveland; and C. Bruce Taylor, MD, Birmingham, Ala

Dietary Habits.—Milk is their main staple, but they are also fond of fresh cow's blood and the meat of cattle, sheep, and goats. The cattle are milked directly into a gourd, and the milk is drunk either fresh or fermented, since bacterial fermentation occurs promptly in the gourd. When sufficient milk is available, the average Masai consumes from 3 to 5 qt of milk daily, usually as two meals.[6] During the dry season of four to five months, when the supply of milk is low, they bleed the cattle and mix the blood with milk or slaughter a sheep or goat and under extreme circumstances one of their beloved cows.



SOME UNIQUE BIOLOGIC CHARACTERISTICS OF THE MASAI OF EAST AFRICA*

KURT BISS, M.D., KANG-JEY HO, M.D., PH.D., BELMA MIKKELSON, B.S., LENA LEWIS, PH.D., AND C. BRUCE TAYLOR, M.D.

Abstract The Masai of East Africa exhibit some unique biologic characteristics. Despite their customary diet composed of 66 per cent calories as fat, they have persistent low serum cholesterol and beta-lipoprotein levels. Post-mortem examinations provided direct proof of a paucity of atherosclerosis. Metabolic studies revealed that the Masai absorbed large amounts of dietary cholesterol, but also possessed a highly efficient negative feedback control of endogenous cholesterol biosynthesis to compensate for the influx of dietary cholesterol. Two unusual serum-protein patterns were observed: the presence of a double $alpha_2$ band; and a high level of serum IgA that is apparent at an early age (four years). The high ratios of phospholipid to cholesterol and bile acid to cholesterol in their gallbladder bile explain the extreme rarity of cholesterol gallstones. All these characteristics may reflect a long-term biologic adaptation of the tribe.

ATHEROSCLEROTIC INDEX

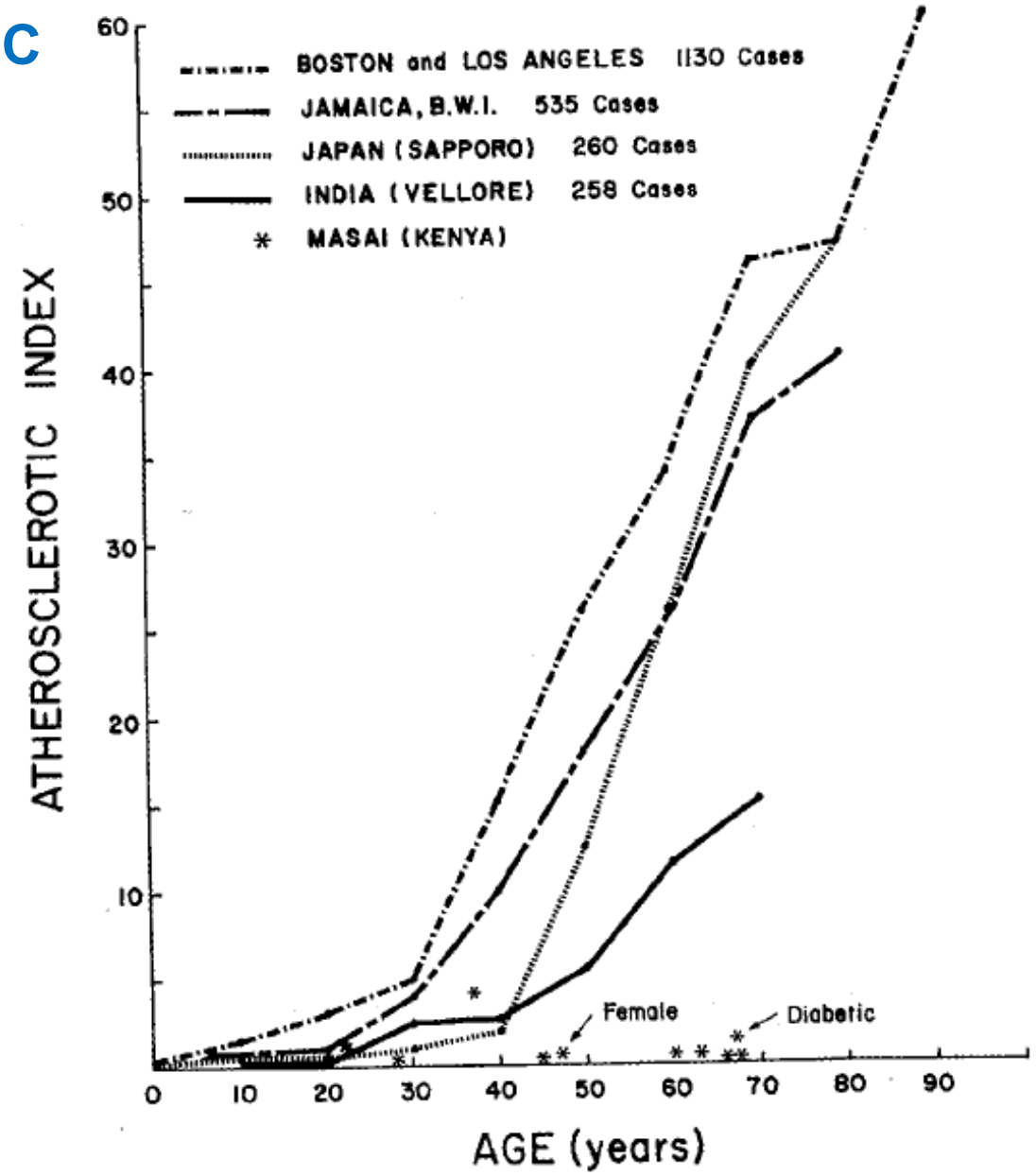


Fig 3.—Comparison of atherosclerotic indices of ten Masai aortas with studies by Gore and Tejada[10] on aortas from Boston and Los Angeles areas and on aortas from Jamaicans, Japanese, and Asian Indians.

CHOLESTEROL LEVELS COMPARED TO US MALES AND FEMALES

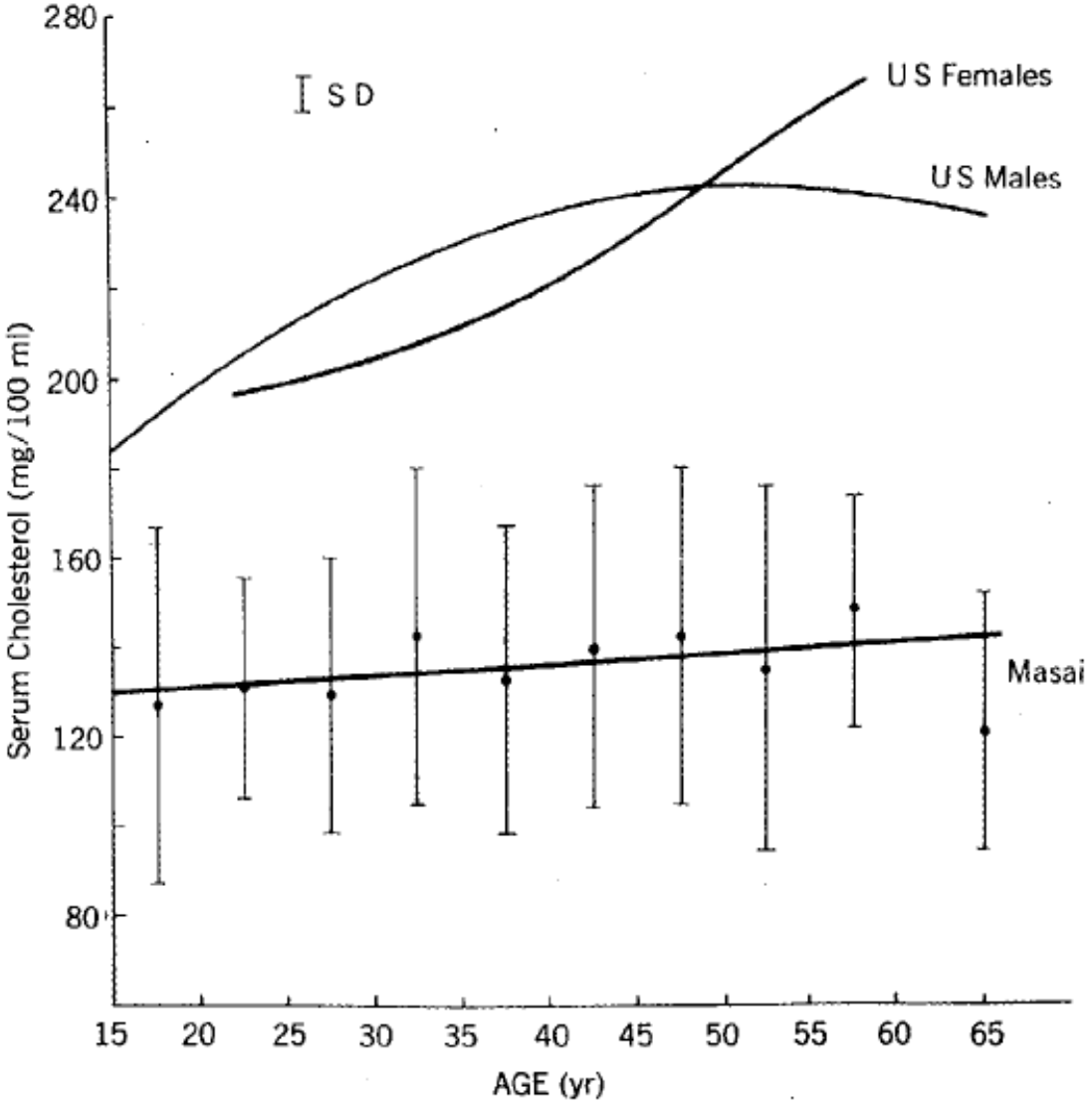
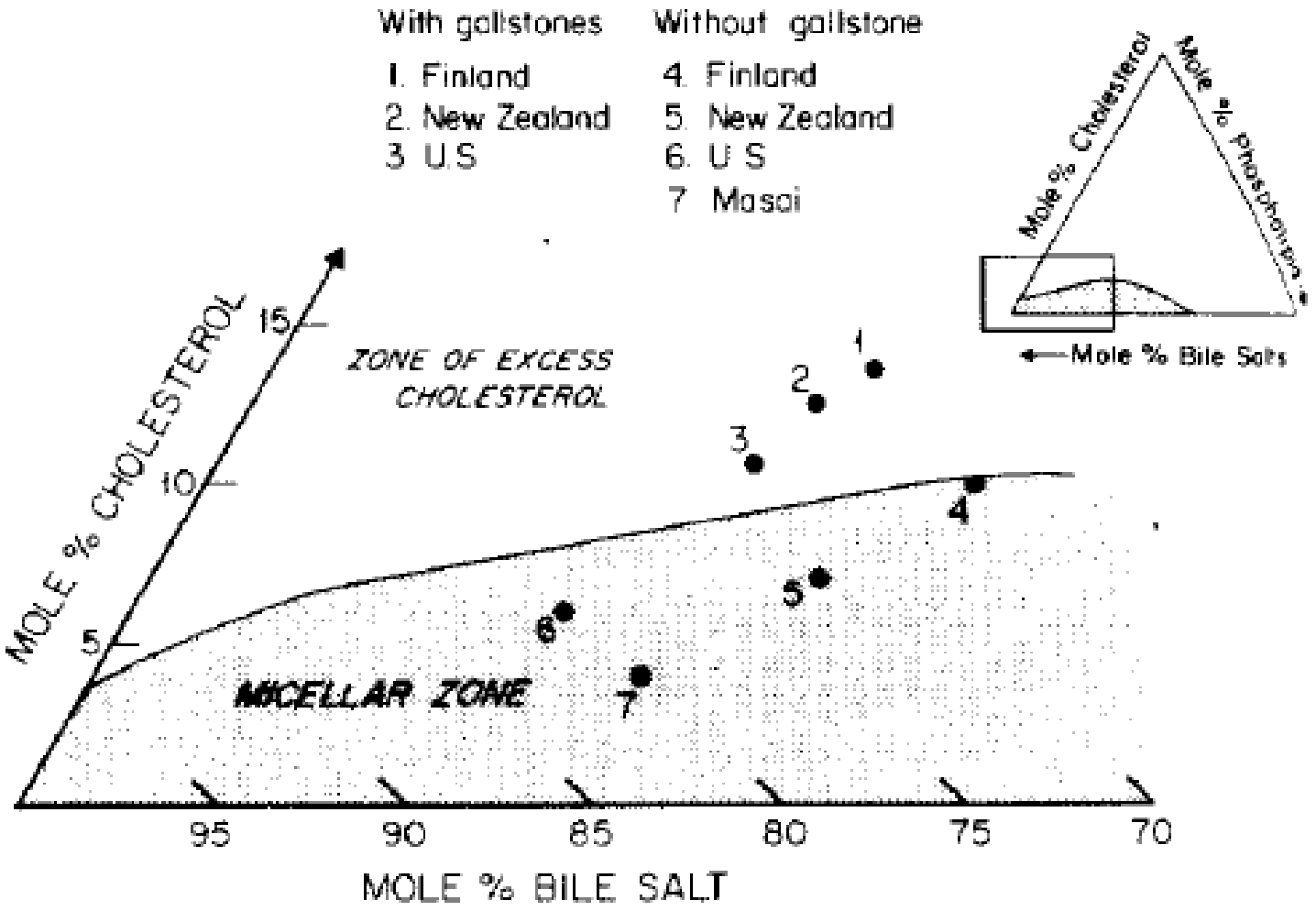


Fig 1.—Comparison of the serum cholesterol levels of the Masai and US populations at various ages.

Absence of Cholesterol Gallstones

Figure 2. Triangular Co-ordinate Plotting of Three Majo Gallbladder-Bile Components among Different Ethni Groups.



Unusual Serum IgA Levels

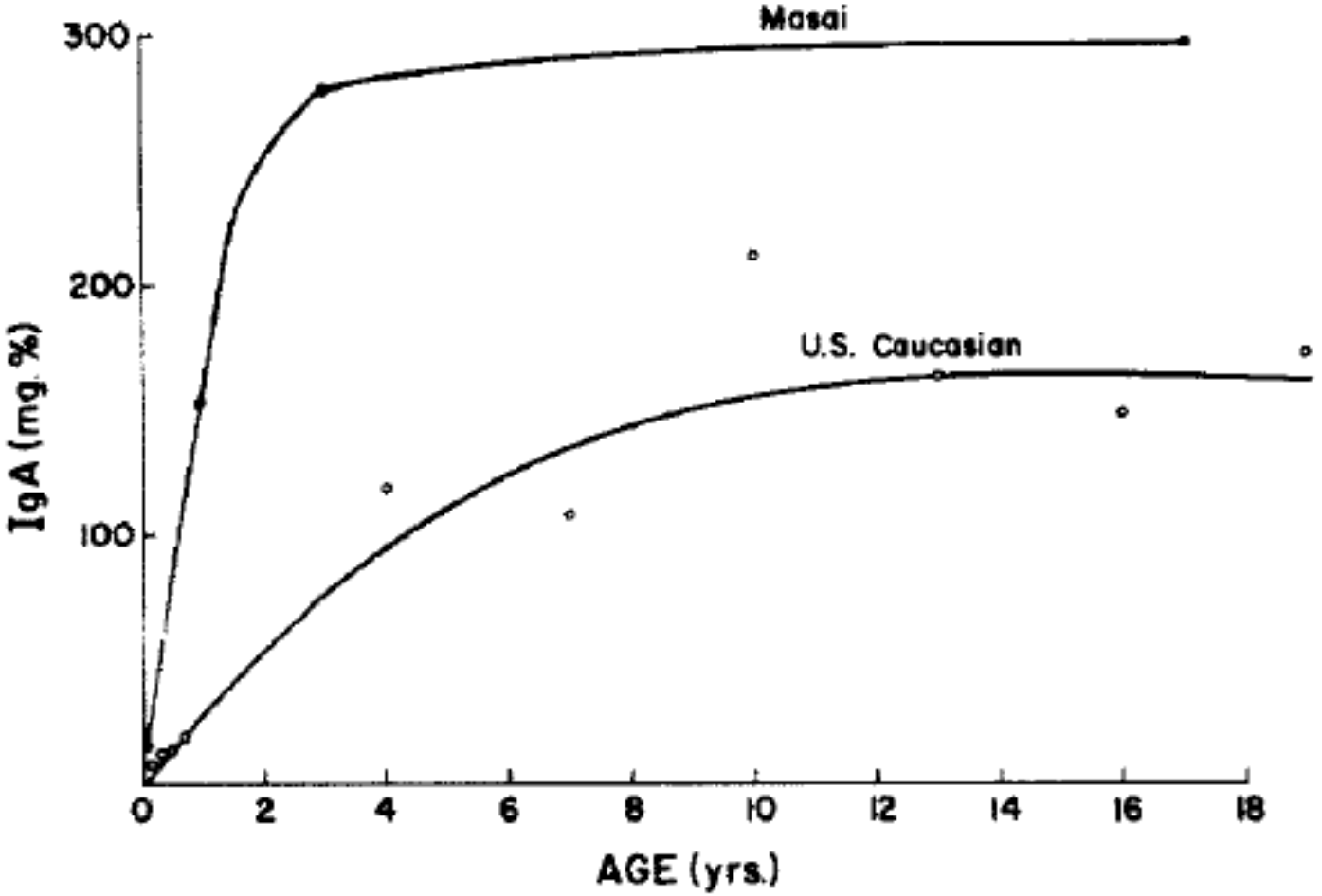


Figure 1. Development of Serum IgA in the Masai and in Whites in the United States.

12 Control and 11 Case. 8 week study with 6 month follow-up

Base diet = corn syrup solids, vegetable fat, corn, beans, sugar, Mazola

Case fed 2 g crystalline Cholesterol with a trace dose of Chol–4 C^{14}

Controls fed only trace Chol-4 C^{14} .

Serum and fecal sterols were determined weekly for 8 weeks.

6 month follow up gave rates of absorption, synthesis and turnover, size and turnover time of body cholesterol exchangable pool

Table 2. Various Aspects of Cholesterol Metabolism in the Control and Experimental Masai Groups.*

Group	No. of Subjects	Rate of Synthesis	Rate of Absorption	Turnover Rate	Turnover Time	Pool Size
		g/day	*g/day*	*g/day*	*days*	*g*
Control	12	1.37±0.15	–	1.37±0.15	60± 9	82.4±15.3
Experimental	11	0.65±0.18†	0.65±0.12†	1.30±0.19	63±12	81.8±19

*Mean ±SD.

†Significant difference from control

Summary: Maasai Characteristics

- Diet: Milk, Meat and Blood. 3000 calories/day – 66% fat (500-2000 mg Cholesterol)
- Paucity of Atherosclerosis
- Low Serum Cholesterol Levels not increasing with age: 135.4 +/- 33 mg/100 ml
- Homeostatic Control of Cholesterol Metabolism ! Liver Reduces Synthesis on excess Cholesterol in the diet.
- Early & Strong Immune System Activation

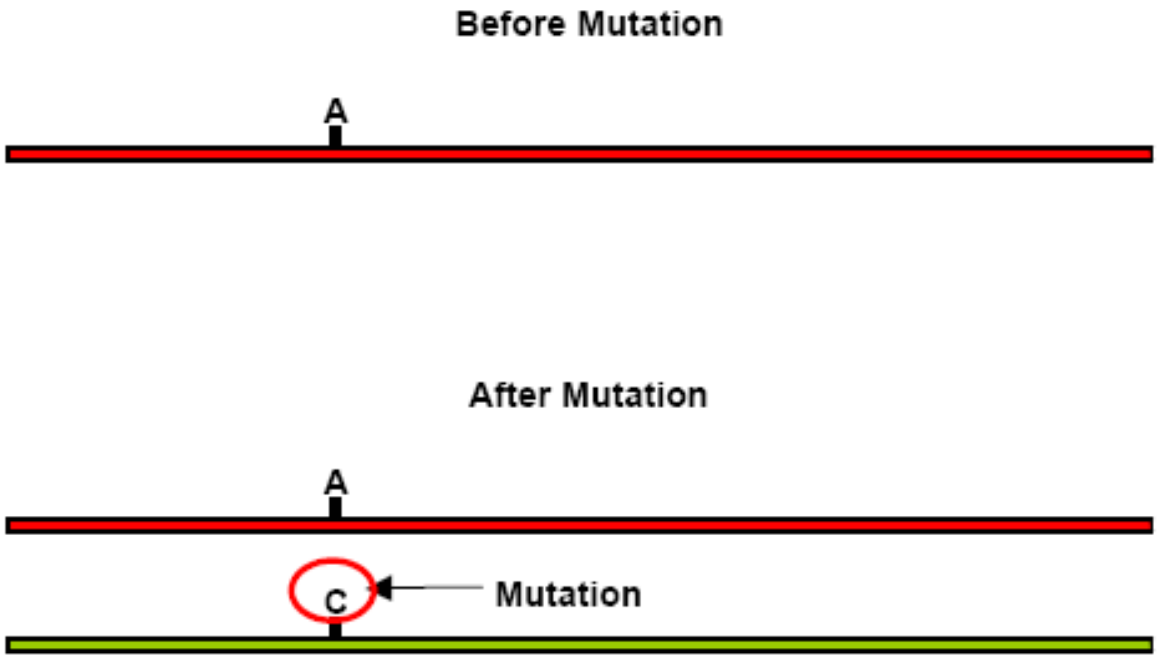
Overall Conclusion

Taylor, B.C., K.J. Ho, Studies on the Masai. Amer. J. of Clin. Nutr., 1971. 24: 1291-1293.

This leads us to believe, but without direct proof, that the Masai have some basically different genetic traits that result in their having superior biologic mechanisms for protection from hypercholesteremia and from many pathogenic organisms.

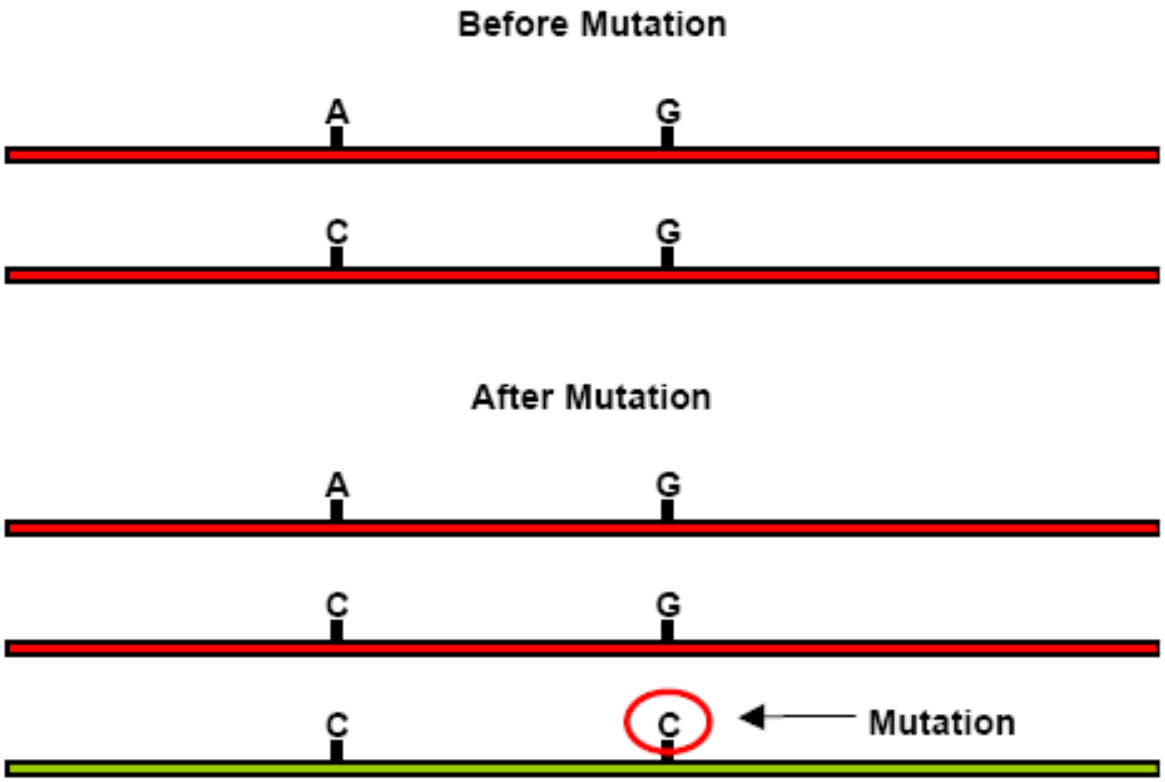
Evolution 101: "Standard Model" : Neutral Evolution

Alleles existing today arose from ancestral mutation events



Alleles occur successively in time

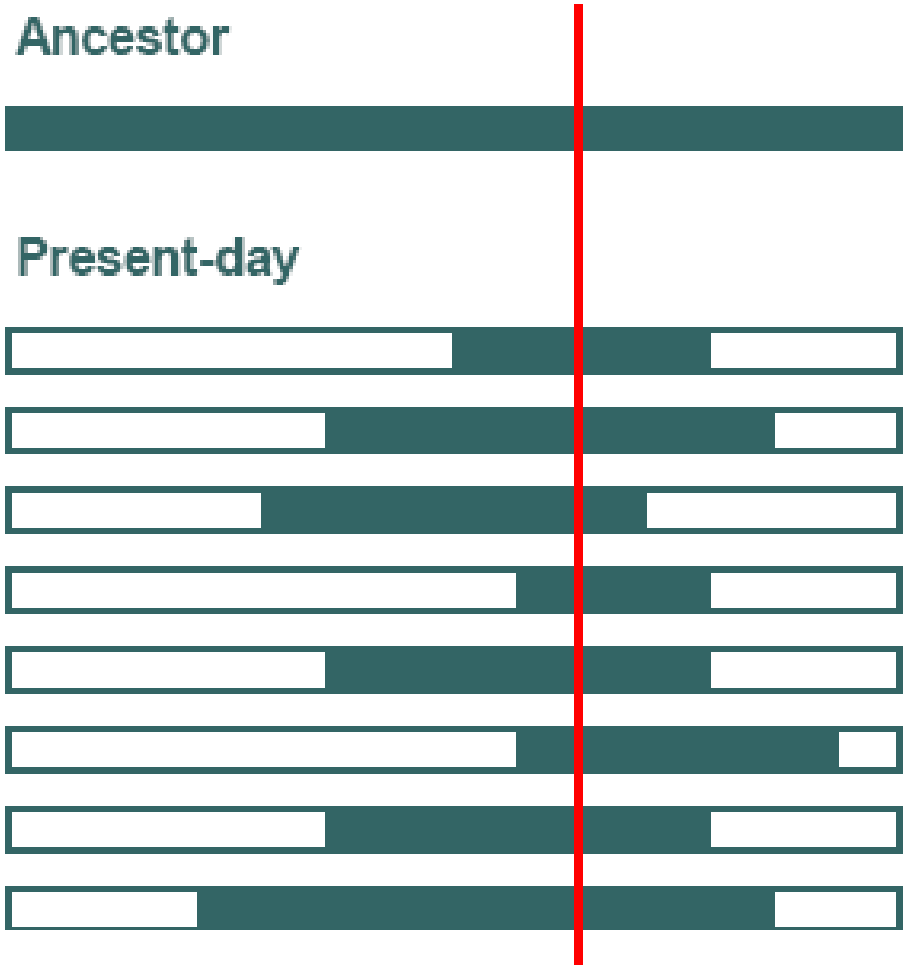
One allele first, then the other



Recombination creates new arrangements

Chromosomes in Populations

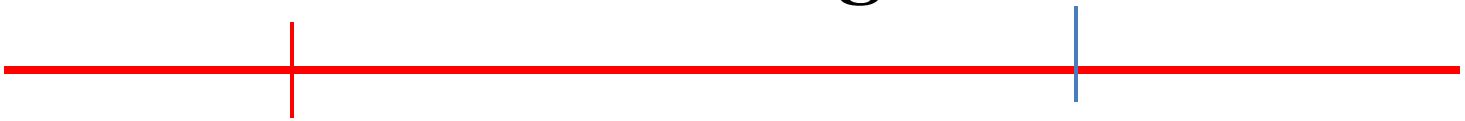
- Chromosome are mosaics
- Lengths and conservation of pieces depend on
  - *Mutation rate*
  - *Recombination Rate*
  - *Population Size*
  - *Selection*
- Allele groups close by on chromosomes reflect ancestral events



Haplotype Frequencies

		Locus B		Totals
		B	b	
Locus A	A	p_{AB}	p_{Ab}	p_A
	a	p_{aB}	p_{ab}	p_a
Totals		p_B	p_b	1.0

Linkage



Alleles: A or a

Alleles: B or b

$$p_{AB} = p_A p_B$$

$$p_{Ab} = p_A p_b = p_A(1 - p_B)$$

$$p_{aB} = p_a p_B = (1 - p_A) p_B$$

$$p_{ab} = p_a p_b = (1 - p_A)(1 - p_B)$$

← **Loci far apart:**

Independent events:

→ **Linkage Equilibrium**

Loci close together →

Not independent →

→ **Linkage Disequilibrium**

$$p_{AB} \neq p_A p_B$$

$$p_{Ab} \neq p_A p_b = p_A(1 - p_B)$$

$$p_{aB} \neq p_a p_B = (1 - p_A) p_B$$

$$p_{ab} \neq p_a p_b = (1 - p_A)(1 - p_B)$$

Linkage Disequilibrium Measures

$$D_{AB} = p_{AB} - p_A p_B$$

$$D'_{AB} = \begin{cases} \dfrac{D_{AB}}{\min(p_A p_B, p_a p_b)} & D_{AB} < 0 \\ \dfrac{D_{AB}}{\min(p_A p_b, p_a p_B)} & D_{AB} > 0 \end{cases}$$

$$\Delta^2 = \frac{D_{AB}^2}{p_A(1-p_A)p_B(1-p_B)}$$

Δ^2 is a Correlation Coefficient

- If we consider that the loci **A** and **B** are random variables X and Y respectively and give numeric values to their allelic forms
  - X = 1 if **A** allele is A; X = 0 if **A** allele is a
  - Y = 1 if **B** allele is B; Y = 0 if **B** allele is b
- Then it is easy to show that
- $\Delta^2 = \text{Cov}(X,Y)/(\sigma_x\sigma_y) = E[(X-\mu_x)(Y-\mu_y)]/(\sigma_x\sigma_y)$
- Δ^2 can be measured between chromosomes

Properties of Δ^2

- Ranges between 0 and 1
  - Δ^2 = 1 means the two markers are linked
  - Δ^2 = 0 means the two markers unlinked
- Neutral Evolution value is $1/2n$
- Influenced by Population Size, Mating patterns, Distance between Markers
- ***Decreases with greater recombination***
- ***Smaller in older populations: decays with time***

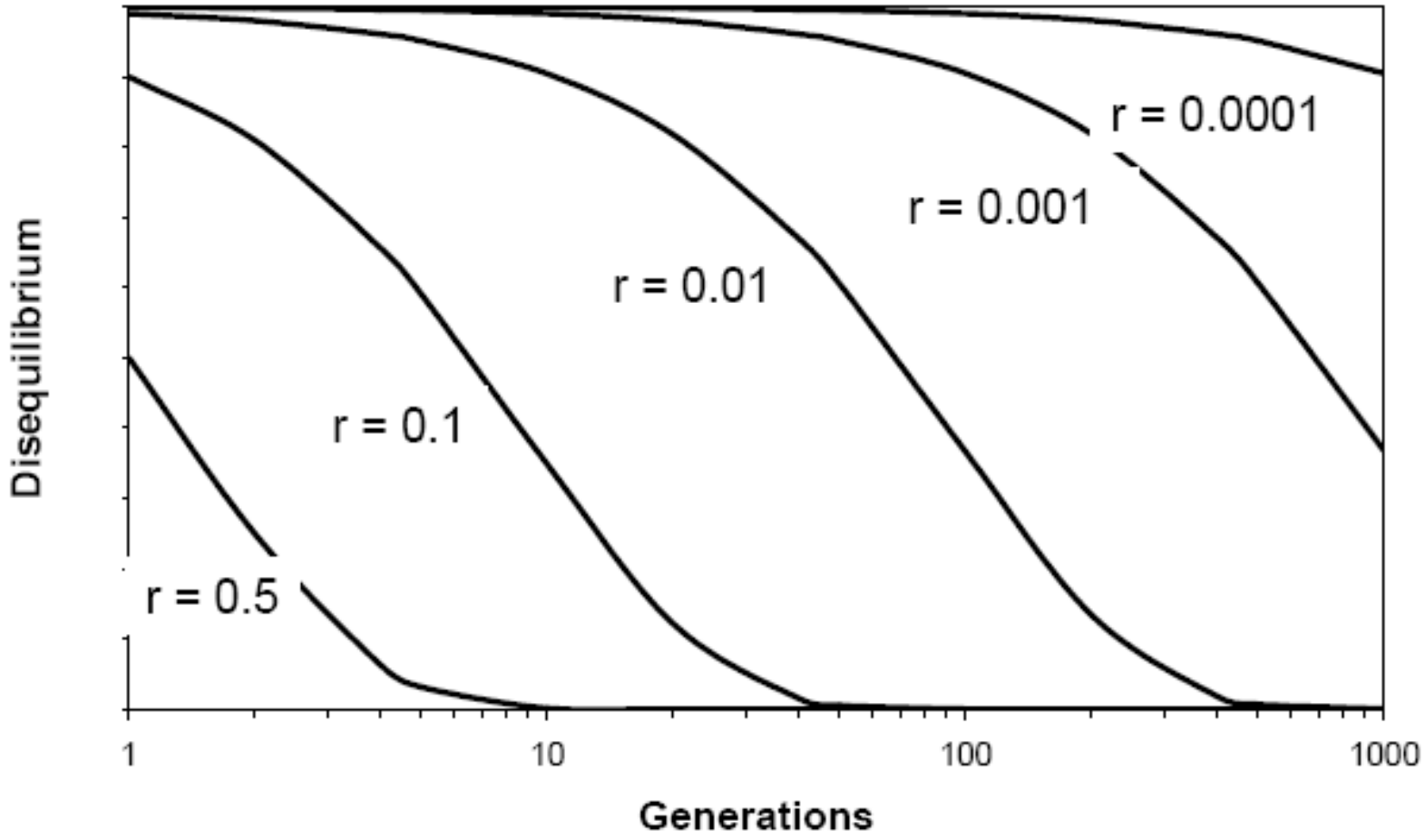
$$D_{AB}(t) = (1-\vartheta)^t D_{AB}(0)$$

Recombination Mixes up Haplotypes

- Initial value of D’ or Δ^2 will decrease with time
- Size of linkage block is a measure of time, mating patterns, bottlenecks, near-extinction events
  - Older populations have smaller linkage blocks
  - Random mating produces smaller blocks quickly
  - Bottlenecks and near extinction events increase linkage
- **BUT ONLY WITHOUT SELECTION !!!**

IN NEUTRAL MODEL

Decay of D with Time



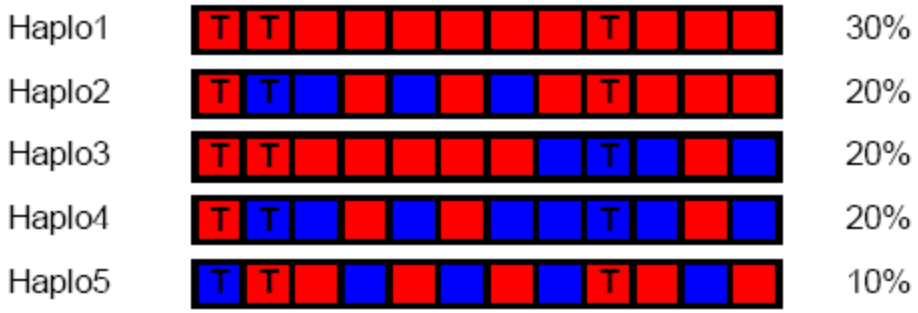
Association Studies and Linkage Disequilibrium

- If all polymorphisms were independent at the population level, association studies would have to examine every one of them…

- Linkage disequilibrium makes tightly linked variants strongly correlated producing cost savings for association studies

Tagging SNPs

- In a typical short chromosome segment, there are only a few distinct haplotypes
- Carefully selected SNPs can determine status of other SNPs

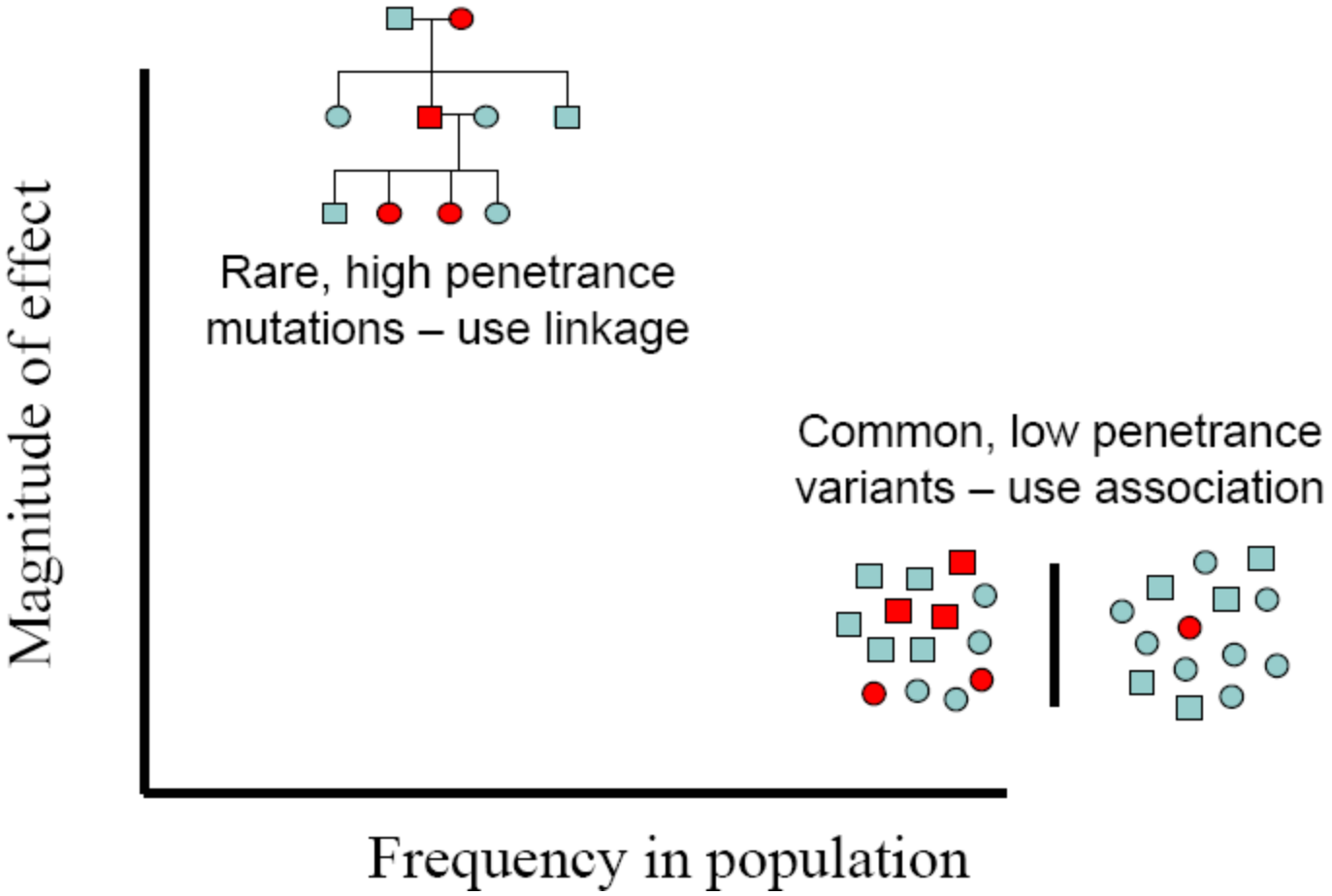


Linkage Disequilibrium Enables Genetic Association Studies

- In contrast to linkage studies, association studies can identify variants with relatively small individual contributions to disease risk
- However, they require detailed measurement of genetic variation and there are >10,000,000 catalogued genetic variants
- Until recently, studies limited to candidate genes or regions
  - A hit-and-miss approach…
- Because assay costs are decreasing and a modest number of variants can represent all others, genome-wide association studies are now possible.

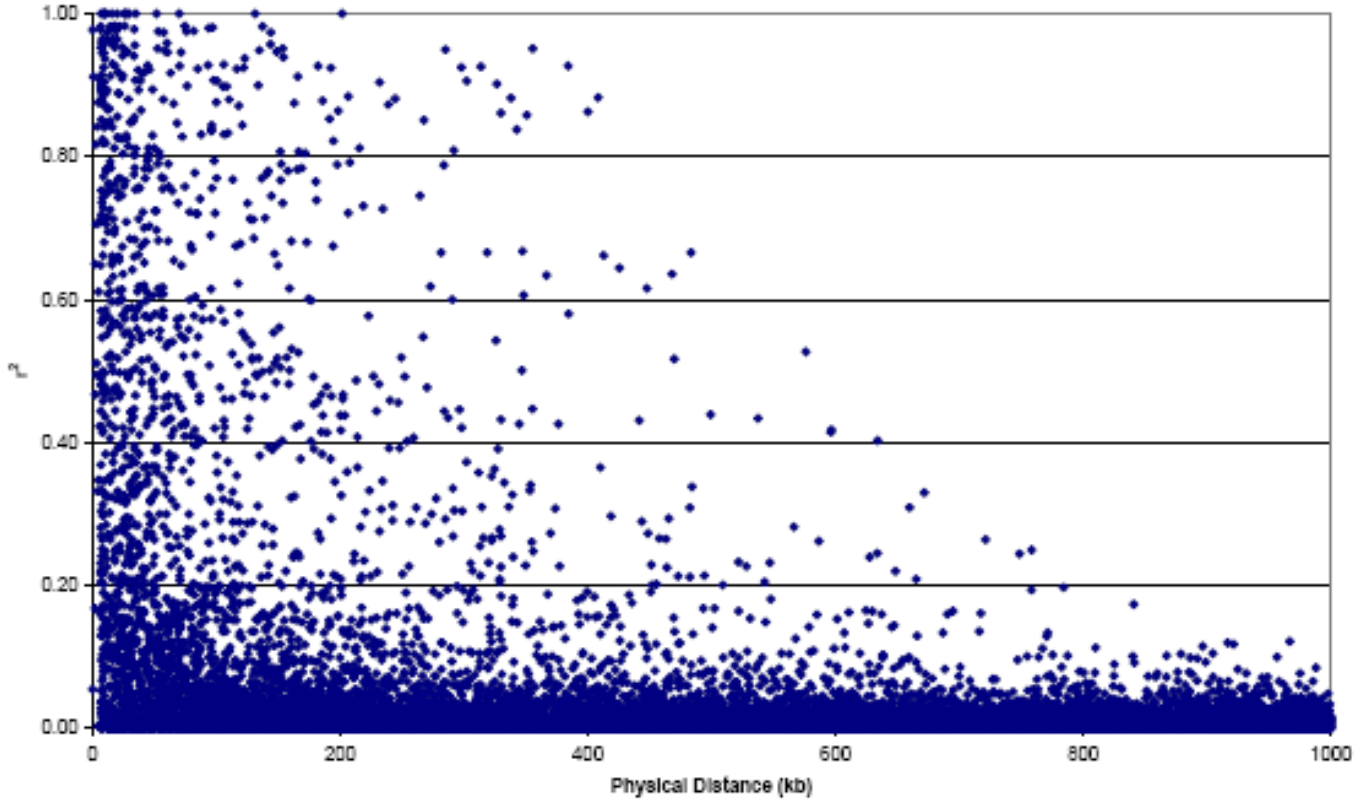
The Allelic Architecture of Disease

What is it and how do we discover it?

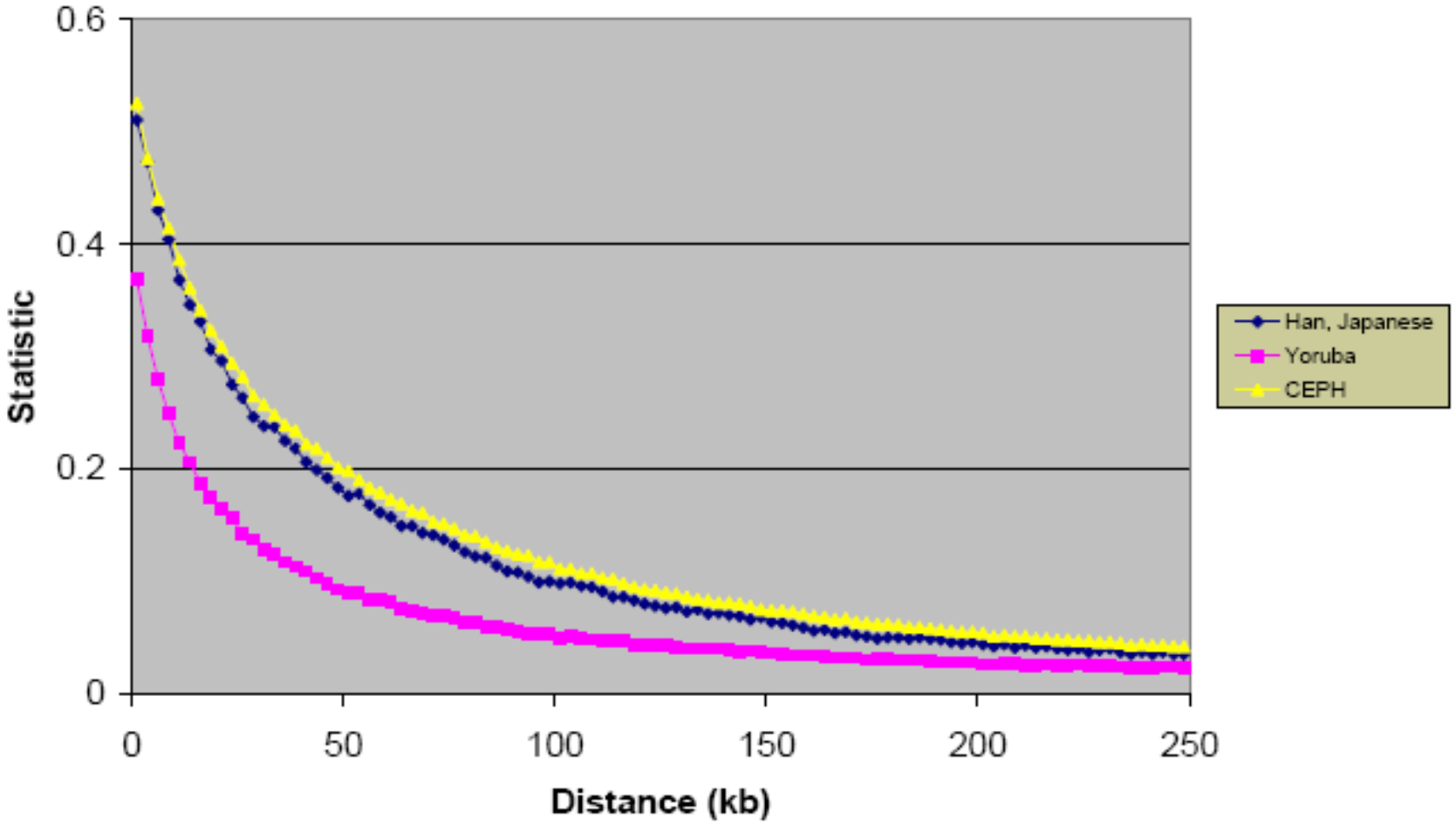


HapMap I and HapMap II
Yoruba, Han/Japanese and CEPH

Raw Δ^2 data from Chr22



HapMap I and HapMap II



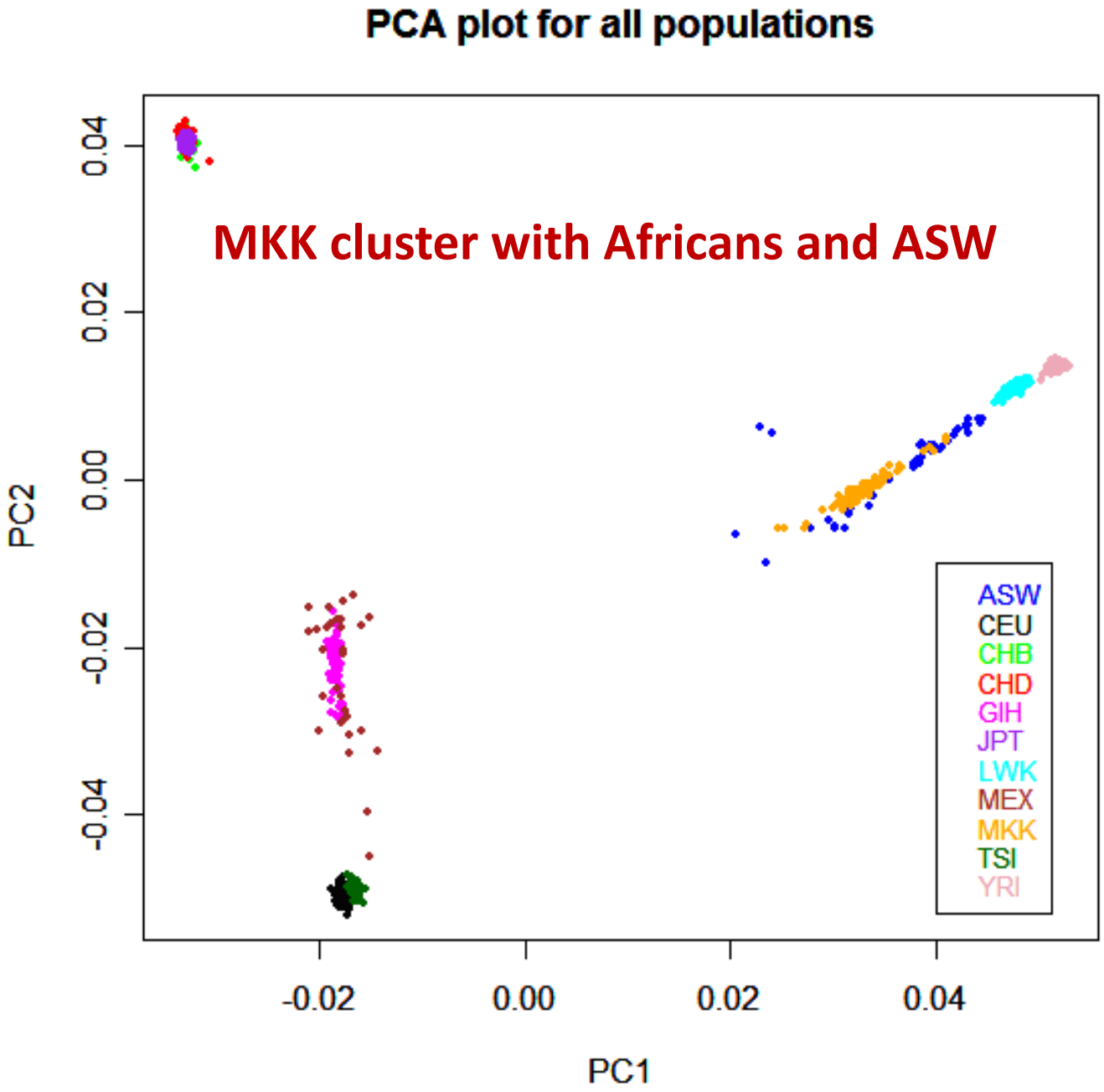
LD extends further in CEPH and the Han/Japanese than in the Yoruba

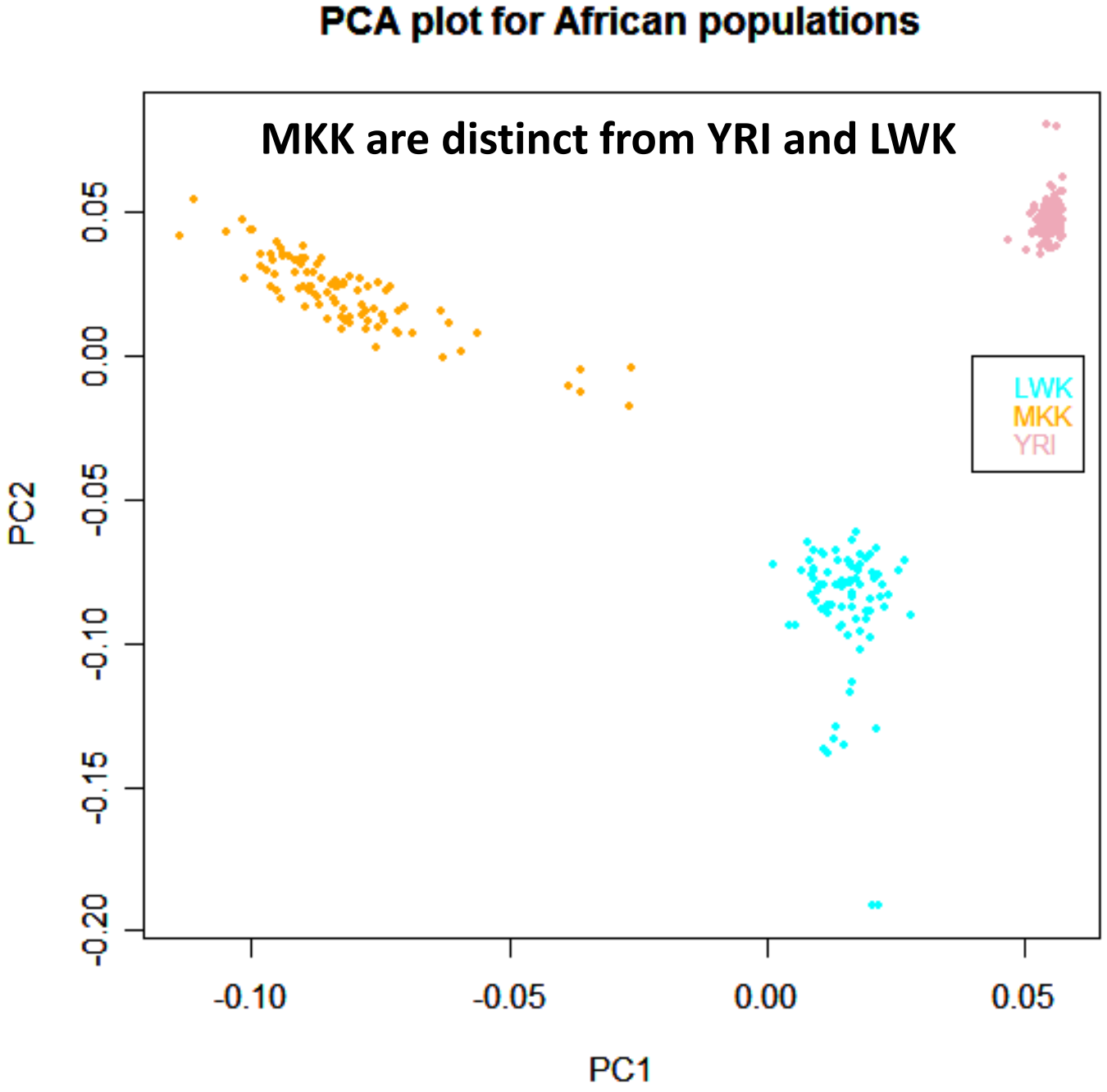
11 Populations of HapMaP III Project

- **CEU – Utah Residents (N or W Europe)**
- **TSI – Italians from Tuscany**
- **CHB – Han Chinese from Beijing**
- **CHD – Chinese in Metropolitan Denver**
- **JPT – Japanese from Tokyo**
- **GIH – Gujaratis from Houston (Asian Indian)**
- **MEX – Mexicans from Los Angeles**
- **ASW – African Americans from SW USA**
- **LWK – Luhya Tribe from Kenya**
- **YRI – Yoruba from Nigeria**
- **MKK – Maasai from Kenya**

HapMaP III data characteristics

- SNP Chip data (Wellcome Trust Sanger Institute; Broad Institute).Dec. 2008, Feb. 2009 releases.
- 1,440,616 SNPs from 1,184 individuals, 993 founders, 490 males and 503 females.
- 171 MKK (Maasai) with 144 founders.
- Cross platform genotype concordance = 0.993
- QC in population call rate > 0.998.


PCA plot for all populations
MKK cluster with Africans and ASW
PC2
PC1
0.04
0.02
0.00
-0.02
-0.04
-0.02
0.00
0.02
0.04
ASW
CEU
CHB
CHD
GIH
JPT
LWK
MEX
MKK
TSI
YRI

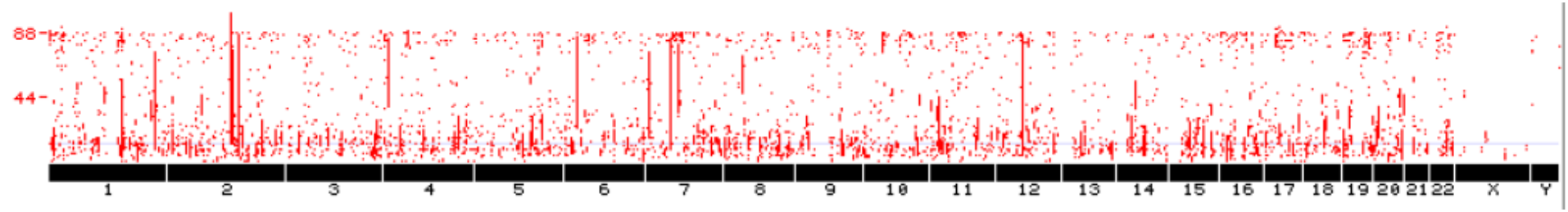

PCA plot for African populations
MKK are distinct from YRI and LWK
LWK
MKK
YRI
PC2
0.05
0.00
-0.05
-0.10
-0.15
-0.20
-0.10
-0.05
0.00
0.05
PC1

Our Analysis

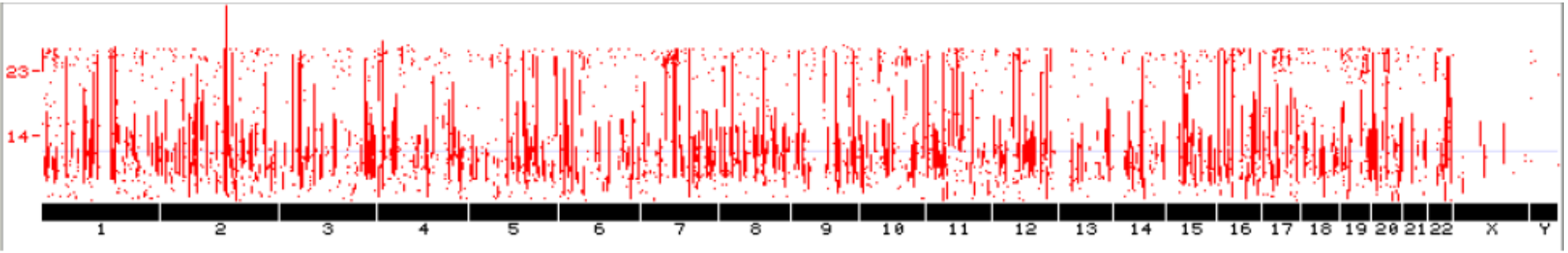
- 3 Association tests (Allelic Association Test, Cochran-Armitage Trend Test, Association Model Test
  & 12 bootstrap experiments on founder samples:
  - (i) Compare MKK to each of the other 10 populations,
  - (ii) Compare MKK to other Africans (YRI + LWE)
  - (iii) Compare MKK to union of all populations
- A SNP is MKK-associated if p-value < 0.00005 in at least 10 out the 12 experiments for each of the 3 tests
- Compute Δ^2 for all pairs of high association SNPs

5,173 MKK associated SNPs distributed over all chromosomes

MKK versus all populations

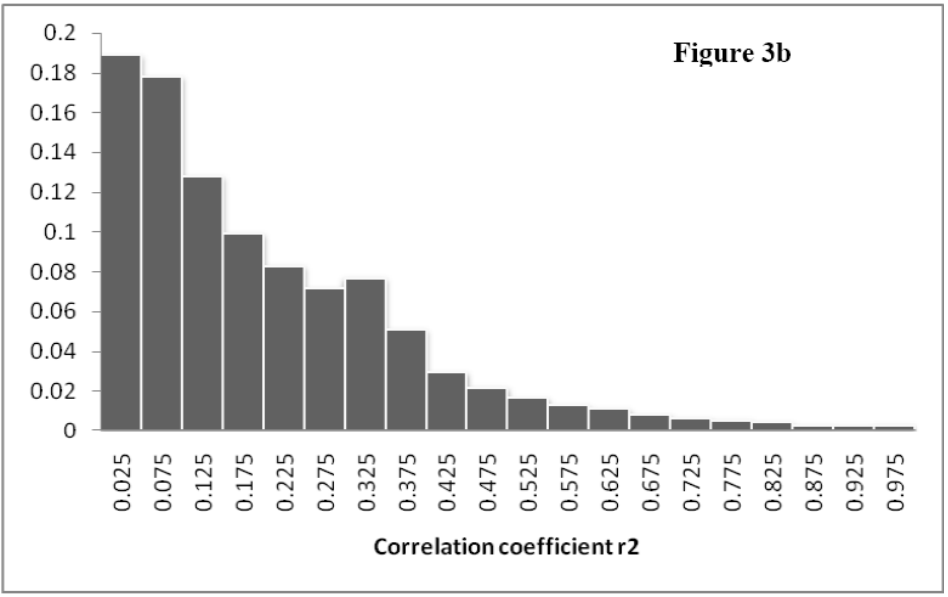


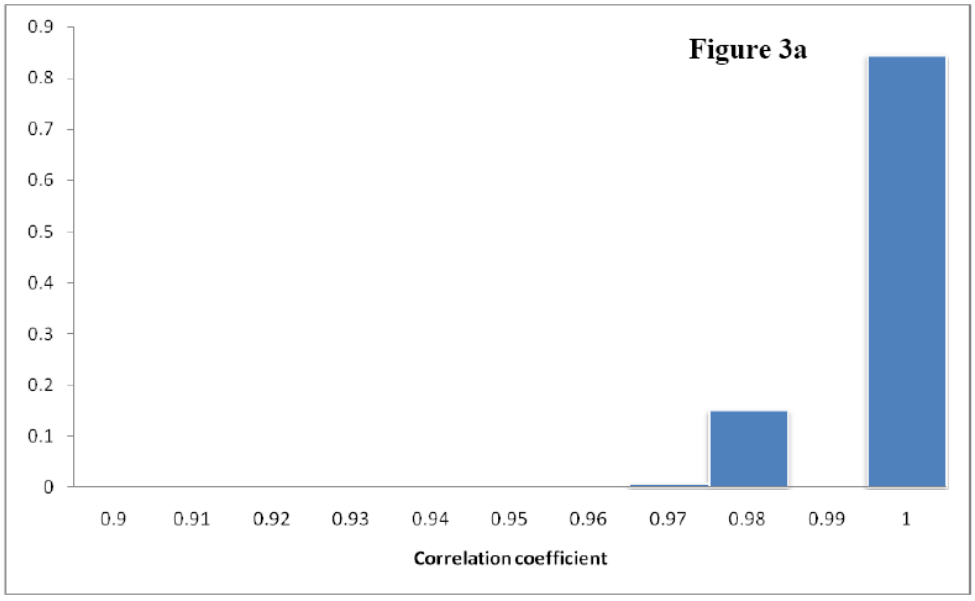
MKK versus Africans



More results

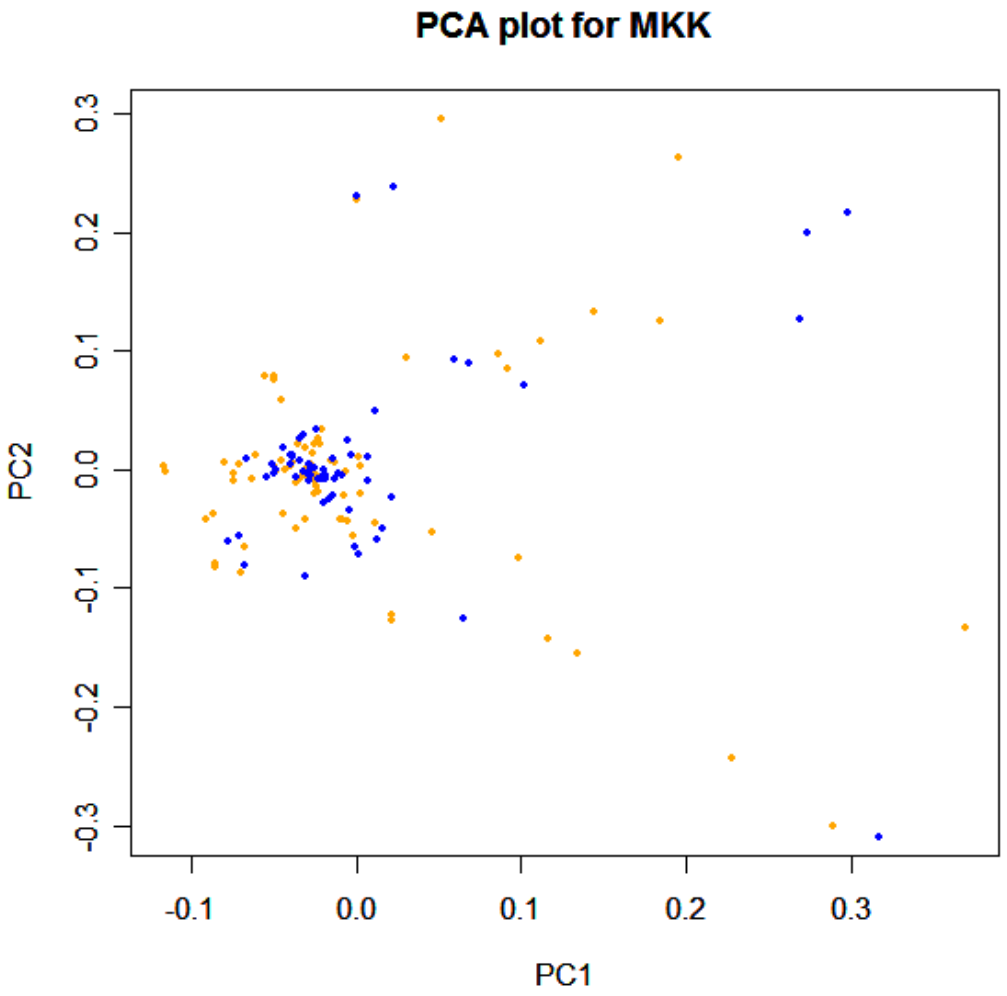
- 697 SNPs form a clique in 61/144 founders
  - Minor allele frequency = 0.23
  - Average Pairwise $\Delta^2 = 0.968$
- Null distribution (left) for Δ^2 from random SNP pairs
- Different from clique distribution (right) (Wilcoxon test p-value = 2.2×10^{-16})





- MKK samples with SNP clique well mixed with other MKK ➔ No obvious sampling bias

- Δ^2 distribution of non-associated SNPs is same as Δ^2 distribution of SNPs in other African populations (YRI and LWK).



Where are these 5173 SNPs?

- **Find pathway enrichment of genes within 100 kb.**
  - **Enrichment of lipid metabolism pathway:**
  - **Genes associated with: Hyperlipidemia, Cardiac Disorders, Hypercholesterolemia, Arteriosclerosis**

- **APoE, APoA2, APoC2, APoB, LPL, ACE, CETP, LDLR, LRP1, TNF, LCAT, LIPG, USF1, MPO, ACAT2, LPA, CYP family**

MKK Associated SNP	100Kb neighborhood	Kb from gene to SNP	chromosome	Function or disease association from GeneCard
rs11556510	APOE	-14.22	19	Binds to liver & peripheral cells. Essential for normal catabolism of triglyceride-rich lipoprotein constituents.
rs4233368	APOA2	5.162	1	Encodes apolipoprotein (apo-) A-II. May stabilize HDL by association with lipids.Defects linked to hypercholesterolemia.
rs11556510	APOC2	-54.42	19	Encoded plasma protein - component of VLDL Activates lipoprotein lipase, which hydrolyzes triglycerides.
rs666126	APOB	58.13	2	Main apolipoprotein of chylomicrons and LDL- mutations cause hypobetalipoproteinemia, hypobetalipoproteinemia, and hypercholesterolemia.
rs1534649	LPL	3.06	8	Functions as a triglyceride hydrolase and ligand/bridging factor for receptor-mediated lipoprotein uptake.
rs6504162	ACE	-17.03	17	Assists catalytic conversion of Angiotensin I into Angiotensin II, potent vasopressor controlling blood pressure.
rs33932458	CETP	-97.26	16	Transfers cholesteryl esters between lipoproteins. Variations in levels may affect susceptibility to arteriosclerosis.
rs11667019	LDLR	-95.06	19	Cell membrane protein involved in rate-limiting step in cholesterol synthesis
rs11172123	LRP1	56.58	12	Involved in cellular lipid homeostasis and plasma clearance of Chylomicron remnants and activated LRPAP1
rs805262	TNF	85.38	6	Multifunctional pro-inflammatory cytokine regulates cell proliferation, differentiation, apoptosis, lipid metabolism, coagulation.
rs7188449	LCAT	75.21	16	Encodes lecithin-cholesterol acyltransferase. Involved in cholesterol transport.
rs4447516	LIPG	-75.12	18	Involved in lipoprotein metabolism and vascular biology.
rs11265559	USF1	105.9	1	Linked to familial combined hyperlipidemia
rs916114	MPO	-0.128	17	Involved in LDL Oxidation, modification.
rs4354180	ACAT2	-51.05	6	Enzyme involved in lipid metabolism, encodes cytosolic acetoacetyl-CoA thiolase.
rs13194662	LPA	115.4	6	Main constituent of lipoprotein(a) (Lp(a)). Apo(a) fragments accumulate in atherosclerotic lesions.
rs382494	PLTP	-82.64	20	Transfers phospholipids from triglyceride-rich lipoproteins to HDL; regulates size of HDL particles.
rs430239	CYP11B2	-76.32	8	Encodes member of cytochrome P450 enzyme involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.
rs4646437	CYP3A4	10.48	7	Encodes member of cytochrome P450 enzyme involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.
rs11188098	CYP2C9	-83.06	10	Encodes member of cytochrome P450 enzyme involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.
rs10242455	CYP3A5	-5.637	7	Encodes member of cytochrome P450 enzyme involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.
rs9662359	CYP2J2	-1.193	1	Encodes member of cytochrome P450 enzyme which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.

81 mis-sense mutations

- KEGG Pathways, GAD (Hum. Gen. Assn. Database), HuGE_Genopedia, RGD_QTLs (Rat QTLs mapped to human coordinates), MGI_QTLs (Mouse QTLs mapped to human coordinates). (QTL = Quantitative Trait Locus)

- QTL Counts associated with phenotypes in these databases:
  - **Blood pressure = 72; Cardiac mass = 41**;
    Non-insulin dependent diabetes mellitus = 39
    Body weight = 39; Renal function = 37;
    Serum cholesterol level QTL = 30
    Stress response = 27.

Other Pathways?

- Enrichment of genes associated with **immune response**, **lactose intolerance**, and **protection against malaria**.
- 8 SNPs create or destroy a STOP codon within an exon.

A locus conferring resistance to diet-induced hypercholesterolemia and atherosclerosis on mouse chromosome 2

Journal of Lipid Research Volume 41, 2000 573

Aram Mouzeyan,* Junghee Choi,§§ Hooman Allayee,* Xuping Wang,† Janet Sinsheimer,§ Jack Phan,†† Lawrence W. Castellani,† Karen Reue,†,†† Aldons J. Lusis,*,†,** and Richard C. Davis[1,†]

Departments of Microbiology and Molecular Genetics,* Medicine/Cardiology Division,† Biomathematics,§ and Molecular Biology Institute,** University of California, Los Angeles, CA 90095; Lipid Research Laboratory,†† VA West Los Angeles Healthcare Center, Los Angeles, CA 90073; and Department of Food and Nutrition, §§ Seoul National University, Seoul, South Korea

Abstract Dietary cholesterol is known to raise total and low density lipoprotein cholesterol concentrations in humans and experimental animals, but the response among individuals varies greatly. Here we describe a mouse strain, C57BL/6ByJ (B6By), that is resistant to diet-induced hypercholesterolemia, in contrast to the phenotype seen in other common strains of mice including the closely related C57BL/6J (B6J) strain. Compared to B6J, B6By mice exhibit somewhat lower basal cholesterol levels on a chow diet, and show a relatively modest increase in absolute levels of total and LDL/VLDL cholesterol in response to an atherogenic diet containing 15% fat, 1.25% cholesterol, and 0.5% cholate. Correspondingly, B6By mice are also resistant to diet-induced aortic lesions, with less than 15% as many lesions as B6J. Food intake and cholesterol absorption are similar between B6By and B6J mice. To investigate the gene(s) underlying the resistant B6By phenotype, we performed genetic crosses with the unrelated mouse strain, A/J. A genome-wide scan revealed a locus, designated *Diet1*, on chromosome 2 near marker D2Mit117 showing highly significant linkage (lod = 9.6) between B6By alleles and hyporesponse to diet. Examination of known genes in this region suggested that this locus represents a novel gene affecting plasma lipids and atherogenesis in response to diet.—

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 277, No. 1, Issue of January 4, pp. 469–477, 2002
Printed in U.S.A.

The *Diet1* Locus Confers Protection against Hypercholesterolemia through Enhanced Bile Acid Metabolism*

Received for publication, July 26, 2001, and in revised form, October 25, 2001
Published, JBC Papers in Press, October 26, 2001, DOI 10.1074/jbc.M107107200

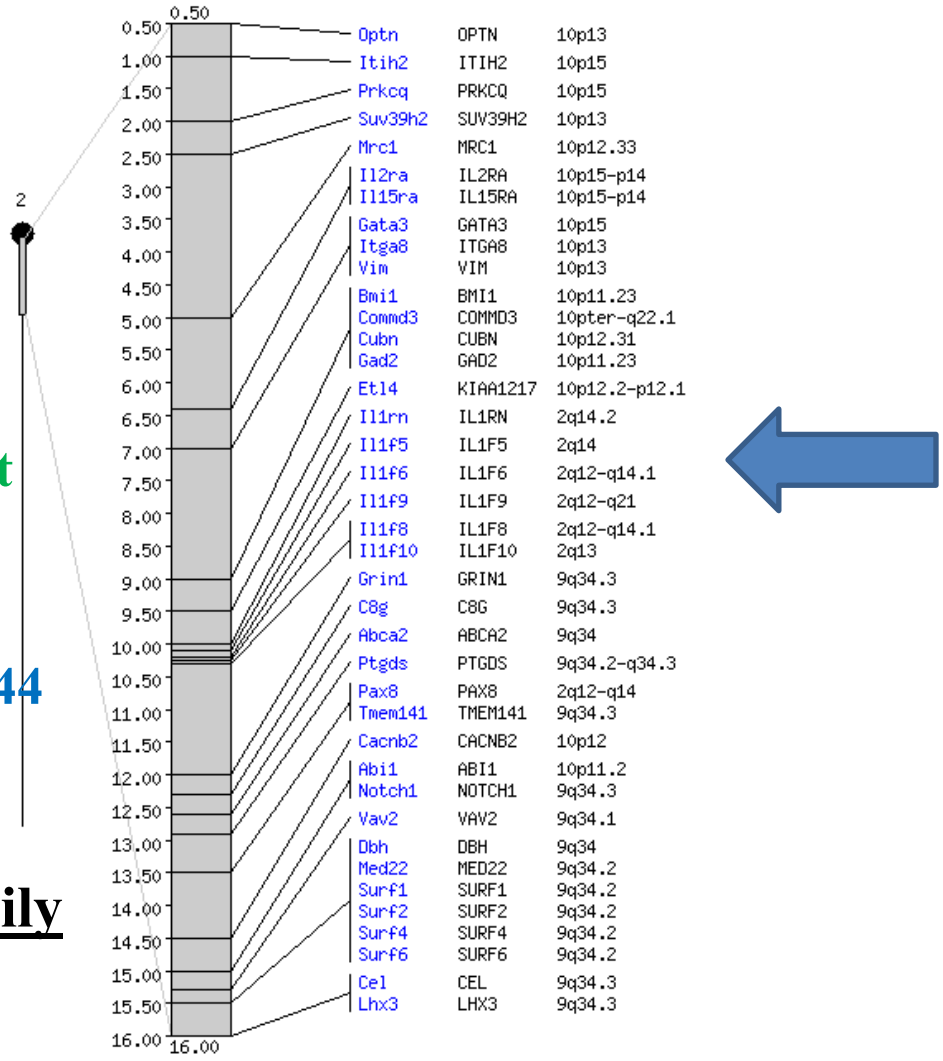
Jack Phan, Tina Pesaran, Richard C. Davis, and Karen Reue‡

From the Department of Medicine, UCLA and the Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, California 90073

function of *Diet1*, we compared mRNA expression profiles in the liver of B6By and B6J mice fed an atherogenic diet using a DNA microarray. These studies revealed elevated expression levels in B6By liver for key bile acid synthesis proteins, including cholesterol 7α-hydroxylase and sterol-27-hydroxylase, and the oxysterol nuclear receptor liver X receptor α. Expression levels for several other genes involved in bile acid metabolism were subsequently found to differ between B6By and B6J mice, including the bile acid receptor farnesoid X receptor, oxysterol 7α-hydroxylase, sterol-12α-hydroxylase, and hepatic bile acid transporters on both sinusoidal and canalicular membranes. The overall expression profile of the B6By strain suggests a higher rate of bile acid synthesis and transport in these mice. Consistent with this interpretation, fecal bile acid excretion is increased 2-fold in B6By mice, and bile acid levels in blood and urine are elevated 3- and 18-fold, respectively. Genetic analysis of serum bile acid levels revealed co-segregation with *Diet1*, indicating that this locus is likely responsible for both increased bile acid excretion and resistance to hypercholesterolemia in B6By mice.

Maasai have 127 SNPs in regions of high homology with the *Diet1 locus*

- **Diet1 locus on Mouse-Chr2**
- **Human Orthologous regions on**
  - **Chr2, Chr9 and Chr10**
- **All the HapMaP populations (except MKK) are wild type in these loci.**
- **In MKK: 9 SNPs with Δ^2 =1 in 61/144 samples; 127 SNPs with $\Delta^2 > 0.968$**
- **<u>Many SNPs in Interleukin 1 family</u>**



IL1 Pathway and Lipid Pathway are linked

- **Mouse studies link Interleukin 1 family to cholesterol levels**: *Devlin et al (2002),* ***Genetic alterations of IL-1 receptor antagonist in mice affect plasma cholesterol level and foam cell lesion size****. PNAS U S A. 99(9):6280-62855*
- **Interleukin 1 family also linked to Arteriosclerosis**: *Merhi-Soussi F et al (2005),* ***Interleukin-1 plays a major role in vascular inflammation and atherosclerosis in male apolipoprotein E-knockout mice****, Cardiovasc. Res.66(3):583-593).*
- **IL1 role in plaque formation in APOE deficient and LDLR deficient mice:** *Isoda K et al (2004)* ***Lack of interleukin-1 receptor antagonist modulates plaque composition in apolipoprotein E-deficient mice****, Arteri. Thromb Vasc Biol. 24(6):1068-1073; Babaev VR et al (2000),* ***Macrophage lipopoprotein lipase promotes foam cell formation and atherosclerosis in low density lipoprotein receptor-deficient mice****, Biol Chem. 2000 Aug 25;275(34):26293-26299.*

Why should there be any selection against Old Age diseases?

- Protection against old age diseases requires reproductive success into old age.
- Maasai Marriage customs may promote this:
  - Maasai practice "open marriage." Consensual extra marital sex is common and acceptable.
  - Older men (especially those with many cattle) often have children with younger women.
  - Children from such unions are ascribed to the "formal" husband (even if he is dead). Can hide high rates of infertility.
- Do Maasai have low incidence of Alzheimer's, Parkinson's, Osteoporosis & Stroke? Do Maasai women have late menopause?

What drives Evolution on time scales of 10,000 – 1000,000 years?

- Short term evolution may be driven by strong selection (over 2,000-10,000 years).
- On longer time scales, selection pressure disappears, modules may be unused (but available) or become lost and allele evolution seems neutral.
- Species phylogeny would agree with neutral model
- Climate Changes/Isolation/Pandemics would result in **rapid selection/speciation because genome has many modules in place which can be easily modified to adapt.**
- Examples of Recent selective sweeps:
  - Skin Color, Height, Adaptation to Malarial parasite, Resistance to Plague/HIV, diet adaptation.

But Many Issues Remain

- What creates and sustains linked SNP modules across chromosomes?
  - Inversion inhibits recombination on the same chromosome.
  - But how can SNPs remain linked across chromosomes?
- Perhaps germline selection/lethality?
- Segregation regulation by recognition markers (as in viral segregation)?
- → → CAN BE TESTED ← ←

Ongoing Next Steps

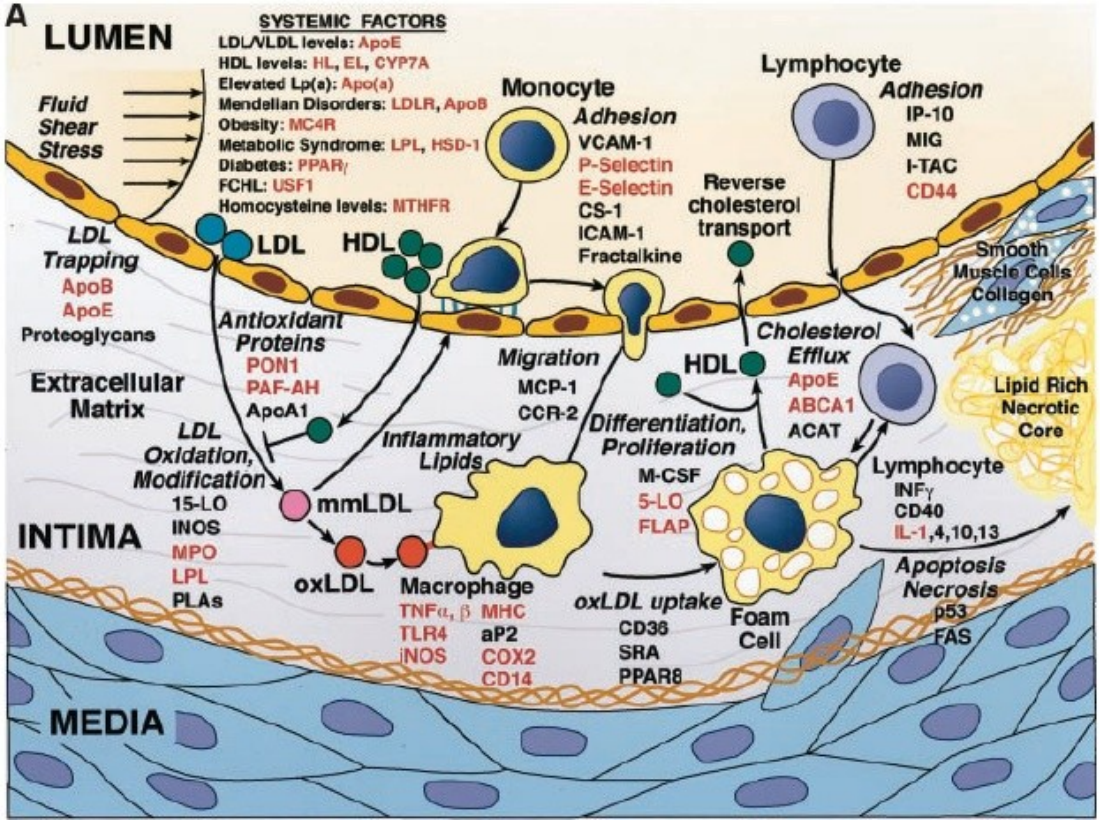
UNDERSTAND THE BIOLOGY OF IDENTIFIED SNPS

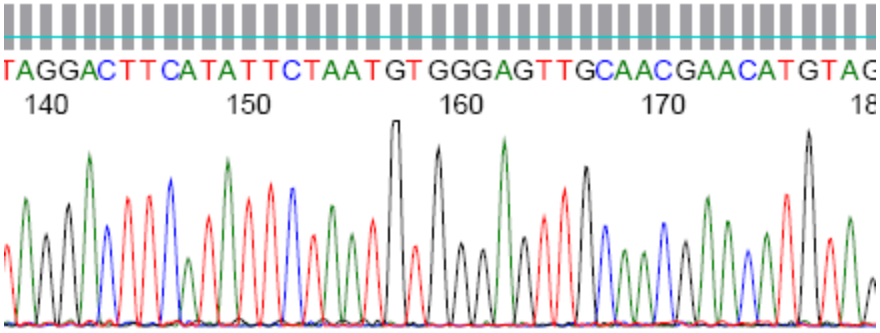
SEQUENCING TO LOOK AT FLANKING HAPLOTYPES

Gene Expression analysis of MKK derived cell lines.

Sequencing and large scale Sampling of MKK population

1000 Genomes project !?





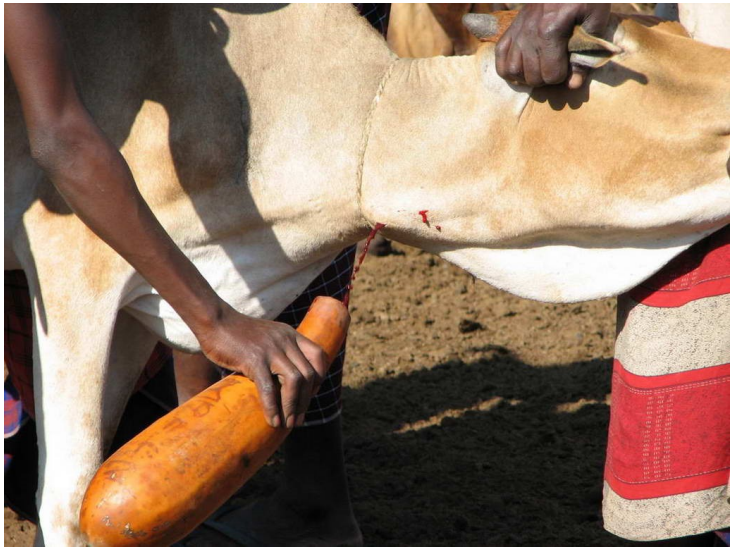
- **Gabriela Alexe: Broad Institute – now at BMS**
- **Anupama Reddy: BioMaPS Institute, Rutgers**
- **Rutgers**
  - **Todd Michael: Waksman (Sequencing); Lee Cronk: Anthropology (Maasai Social Customs); Lane McIntosh: Summer Intern (Analysis/Primers); Michael Seiler: BioMaPS (Clustering and PCA)**
- **CINJ**
  - **Shridar Ganesan, Ming Yao, Michael Boemo (Sanger Sequencing)**
- **KITP – UCSB**
  - **Boris Shraiman & Richard Neher (Simulations of r/s; SNP cliques)**
- **Mt Sinai School of Medicine**
  - **Ravi Sachidanandam and Ajish George (Sequencing, SNP Analysis)**
- **IAS**
  - **Arnie Levine (Scepticism, Encouragement)**

THANK YOU !

The Masai of East Africa: Some Unique Biological Characteristics

Arch Path—Vol 91, May 1971

Kang-Jey Ho, MD, PhD, Birmingham, Ala; Kurt Biss, MD, Chicago;
Belma Mikkelson, Birmingham, Ala;
Lena A. Lewis, PhD, Cleveland; and C. Bruce Taylor, MD, Birmingham, Ala



Vol. 284 No. 13

SOME UNIQUE BIOLOGIC CHARACTERISTICS OF THE MASAI OF EAST AFRICA*

KURT BISS, M.D., KANG-JEY HO, M.D., PH.D., BELMA MIKKELSON, B.S.,
LENA LEWIS, PH.D., AND C. BRUCE TAYLOR, M.D.

A locus conferring resistance to diet-induced hypercholesterolemia and atherosclerosis on mouse chromosome 2

Journal of Lipid Research Volume 41, 2000 573

Aram Mouzeyan,* Junghee Choi,§§ Hooman Allayee,* Xuping Wang,† Janet Sinsheimer,§ Jack Phan,†† Lawrence W. Castellani,† Karen Reue,†,†† Aldons J. Lusis,*,†, and Richard C. Davis[1,†]**

Departments of Microbiology and Molecular Genetics,* Medicine/Cardiology Division,† Biomathematics,§ and Molecular Biology Institute,** University of California, Los Angeles, CA 90095; Lipid Research Laboratory,†† VA West Los Angeles Healthcare Center, Los Angeles, CA 90073; and Department of Food and Nutrition, §§ Seoul National University, Seoul, South Korea

THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 277, No. 1, Issue of January 4, pp. 469–477, 2002
Printed in U.S.A.

The *Diet1* Locus Confers Protection against Hypercholesterolemia through Enhanced Bile Acid Metabolism*

Received for publication, July 26, 2001, and in revised form, October 25, 2001
Published, JBC Papers in Press, October 26, 2001, DOI 10.1074/jbc.M107107200

Jack Phan, Tina Pesaran, Richard C. Davis, and Karen Reue‡

From the Department of Medicine, UCLA and the Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, California 90073

Our Results

- 5,256 SNPs strongly associated with Maasai at an overall p-value < 10^{-12} on combining all tests.
- 27 SNPs removed because they did not have a dbSNP entry (Feb 2008) – merged or discarded
- 56 SNPs removed because of > 0.01 in minor allele frequency between Dec 2008 and Feb 2009 release
- **5,173 SNPs retained as significantly associated with MKK over all other populations. These SNPs are distributed across all chromosomes**

Linkage Equilibrium Expected for Distant Loci

$$p_{AB} = p_A p_B$$

$$p_{Ab} = p_A p_b = p_A(1 - p_B)$$

$$p_{aB} = p_a p_B = (1 - p_A) p_B$$

$$p_{ab} = p_a p_b = (1 - p_A)(1 - p_B)$$

Linkage Disequilibrium Expected for Nearby Loci

$$p_{AB} \neq p_A p_B$$

$$p_{Ab} \neq p_A p_b = p_A(1 - p_B)$$

$$p_{aB} \neq p_a p_B = (1 - p_A) p_B$$

$$p_{ab} \neq p_a p_b = (1 - p_A)(1 - p_B)$$

Linkage Disequilibrium Measures

$$D_{AB} = p_{AB} - p_A p_B$$

$$p_{AB} = p_A p_B + D_{AB}$$

$$p_{Ab} = p_A p_b - D_{AB}$$

$$p_{aB} = p_a p_B - D_{AB}$$

$$p_{ab} = p_a p_b + D_{AB}$$

D’ – A scaled version of D

$$D'_{AB} = \begin{cases} \dfrac{D_{AB}}{\min(p_A p_B, p_a p_b)} & D_{AB} < 0 \\ \dfrac{D_{AB}}{\min(p_A p_b, p_a p_B)} & D_{AB} > 0 \end{cases}$$

- Ranges between –1 and +1