Lecture Handout 21: Linkage disequilibrium and recombination.

Previous assumption: loci are inherited independently of each other, and the fitness of a genotype is independent of the fitness of a genotype at another locus. We now discard these simplifications:

Epistasis: the interaction of fitness values at different loci.

Linkage disequilibrium, LD (also called gametic disequilibrium): the nonrandom association of alleles at different loci into haplotypes (or gametes).

Linkage disequilibrium is generally a function of the rate of recombination.

D, for *deviation* from nonrandom association, is used as the measure of linkage disequilibrium. (This has nothing to do with the D of Nei's genetic distance.)

Linkage disequilibrium, for a 2-locus, 2-allele model:

TABLE 10.1 The gametes (or haplotypes), the alleles, and their frequencies for a two-locus, two-allele model.

Gamete (haplotype)	*Frequency*	*Allele*	*Frequency*
A_1B_1	x_{11}	A_1	$p_1 = x_{11} + x_{12}$
A_1B_2	x_{12}	A_2	$p_2 = x_{21} + x_{22}$
A_2B_1	x_{21}	B_1	$q_1 = x_{11} + x_{21}$
A_2B_2	x_{22}	B_2	$q_2 = x_{12} + x_{22}$

Important difference in nomenclature: in this model p_1 and p_2 represent the frequencies at locus A of alleles A_1 and A_2, so $p_1 + p_2 = 1$. While frequencies at locus B of alleles B_1 and B_2 are q_1 and q_2, with $q_1 + q_2 = 1$.

Linkage disequilibrium, for a 2-locus, 2-allele model:

TABLE 10.2 The frequencies of the four possible gametes when there are two alleles at each of two loci.

	A_1	A_2	*Total*
B_1	$x_{11} = p_1q_1 + D$	$x_{21} = p_2q_1 - D$	q_1
B_2	$x_{12} = p_1q_2 - D$	$x_{22} = p_2q_2 + D$	q_2
Total	p_1	p_2	1

D, for *deviation* from nonrandom association, is used as the measure of linkage disequilibrium, is equal to observed minus expected frequency of gamete A_1B_1.

$$D = x_{11} - p_1q_1$$

D, for *d*eviation from nonrandom association, is also equal to the product of the frequencies of the "**coupling**" gametes, minus the product of the frequencies of the "**repulsion**" gametes. Where "coupling" and "repulsion" refers to alleles given the same or different subscript notations, respectively:

$$D = x_{11}x_{22} - x_{12}x_{21}$$

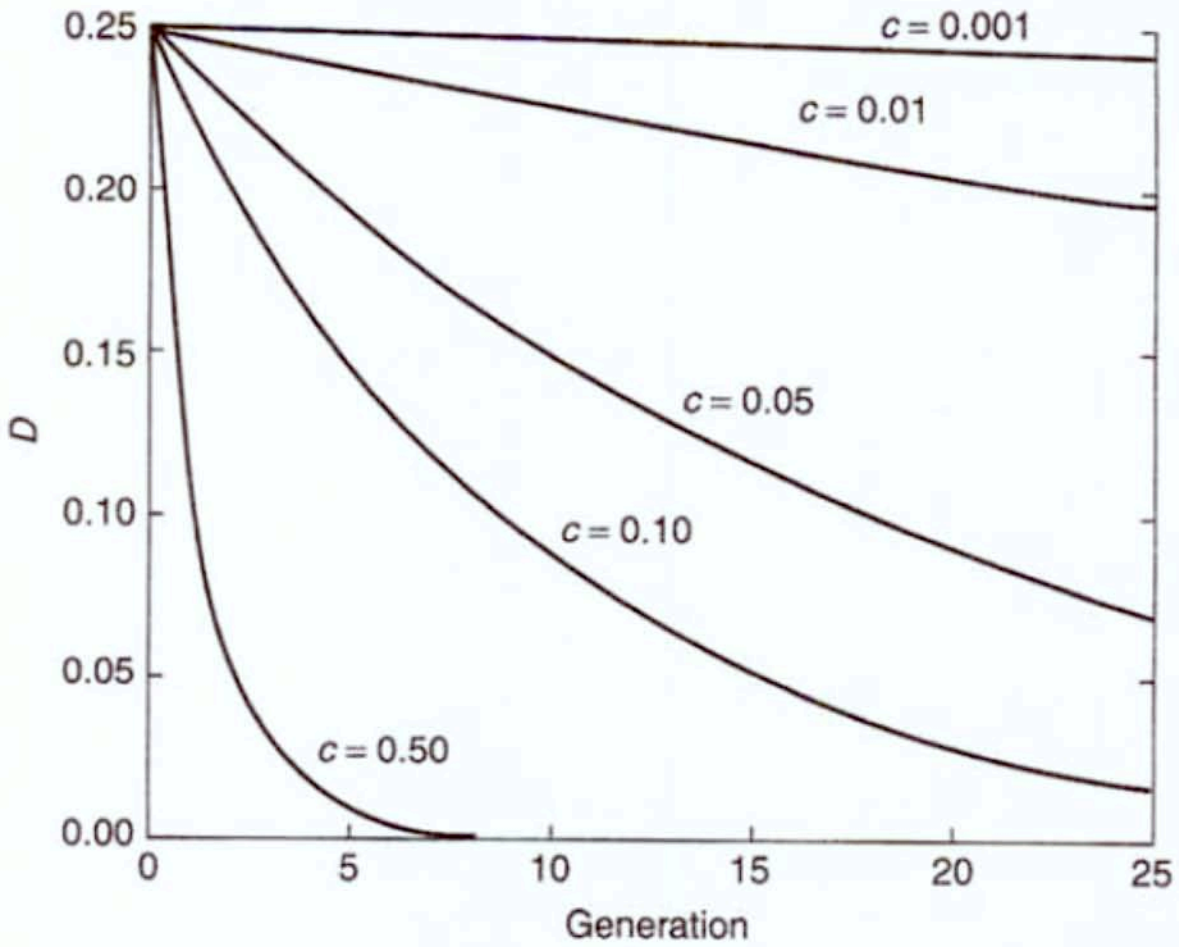
Rate of recombination, *c* (sometimes called r): ranges from zero when there is no recombination to 0.5 when there is independent assortment (since even with random assortment half the gametes will still be the parental type).

Recombination reduces linkage disequilibrium:

$$D_1 = (1 - c)D_0$$

Over multiple generations, recombination reduces linkage disequilibrium following a recursive relationship:

$$D_t = e^{-ct} D_0$$



$$t = \frac{\ln (D_t / D_0)}{\ln (1 - c)}$$

Figure 10.1. The decay of linkage disequilibrium D for different amounts of recombination (c).

Measured linkage disequilibrium varies with amount of recombination across a chromosome, eg, human chromosome 22:

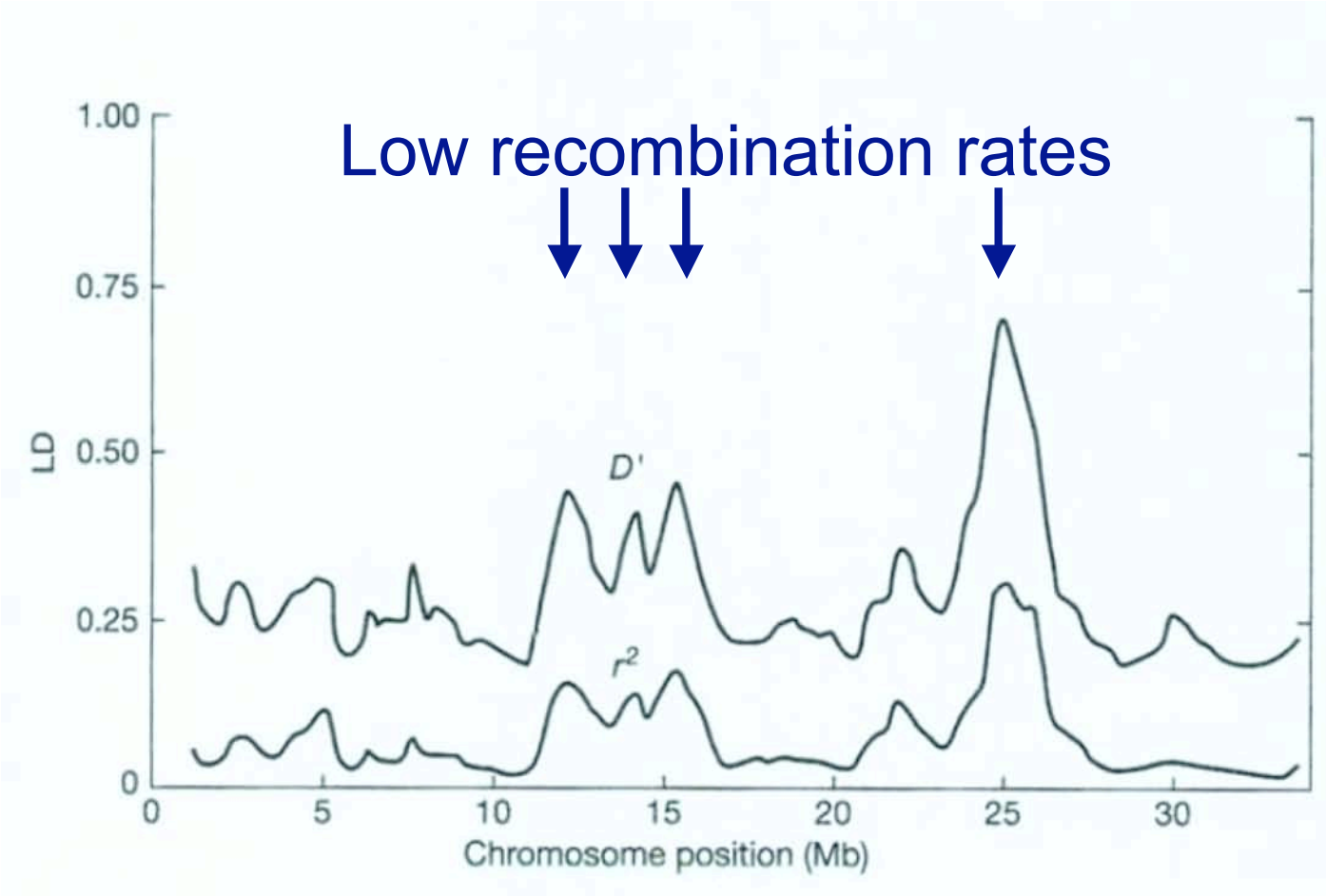
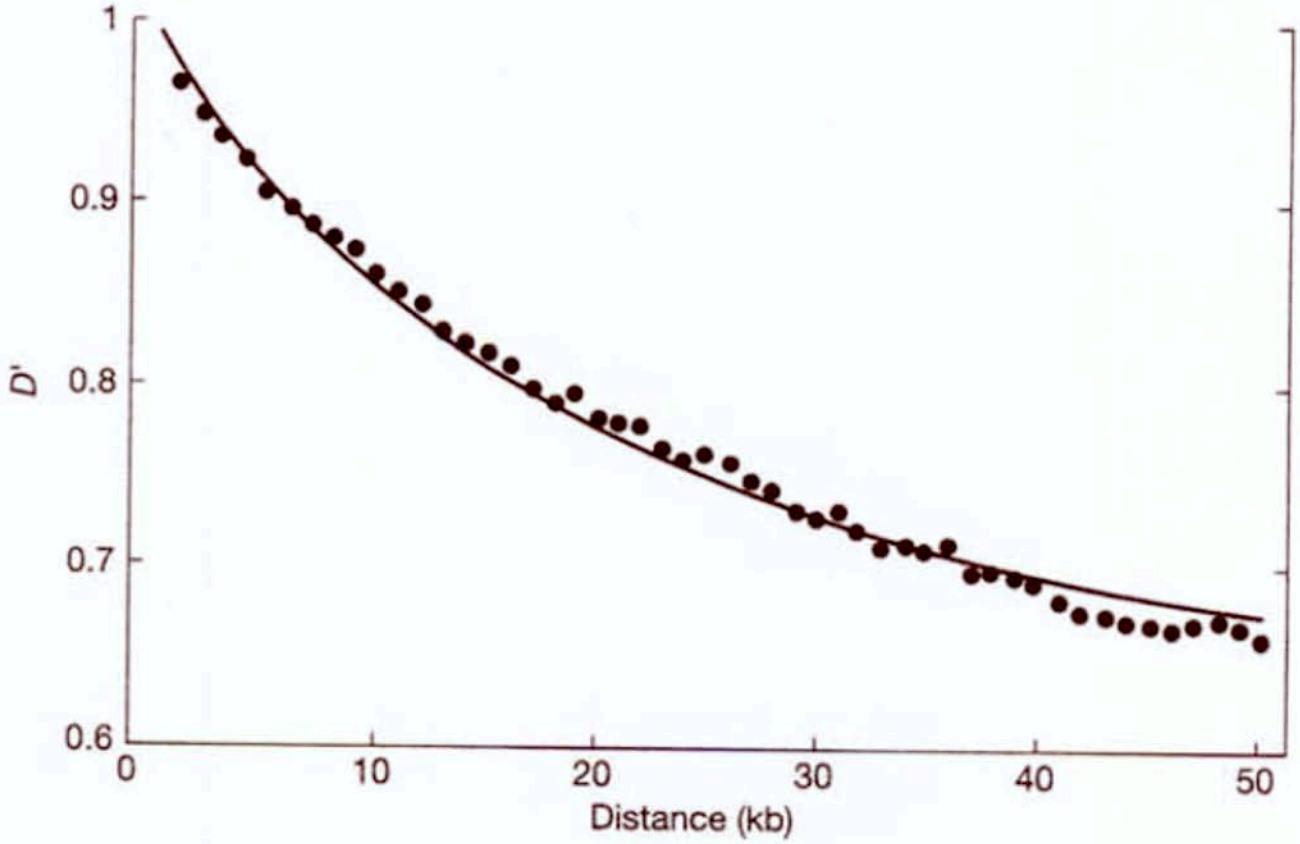


Figure 10.2. The extent of linkage disequilibrium between closely linked markers as a function of map position along human chromosome 22 as measured by D' (upper line) and r^2 (lower line) for the European ancestry CEPH sample. (Courtesy of *Nature*, 418: 2002, by Dawson, E., G.R. Abecasis, S. Bumpstead, *et al.* Reprinted with permission of Nature Publishing Group.)

Measured linkage disequilibrium drops with increasing distance between two loci on a chromosome; eg among 24000 SNPs in human chromosome 21:

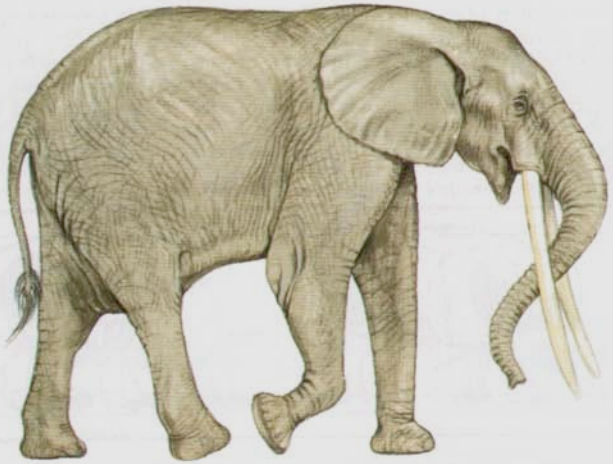
Figure 10.3. The decline of linkage disequilibrium as measured by D' between pairs of 24,056 SNPs on human chromosome 21 where the closed circles indicate the observed averages and the line shows the expected decline. (Courtesy of Innan, H., B. Padhukasahasram, and M. Nordborg. 2003. The pattern of polymorphism on human chromosome 21. *Genome Res.* 13:1158–1168.)



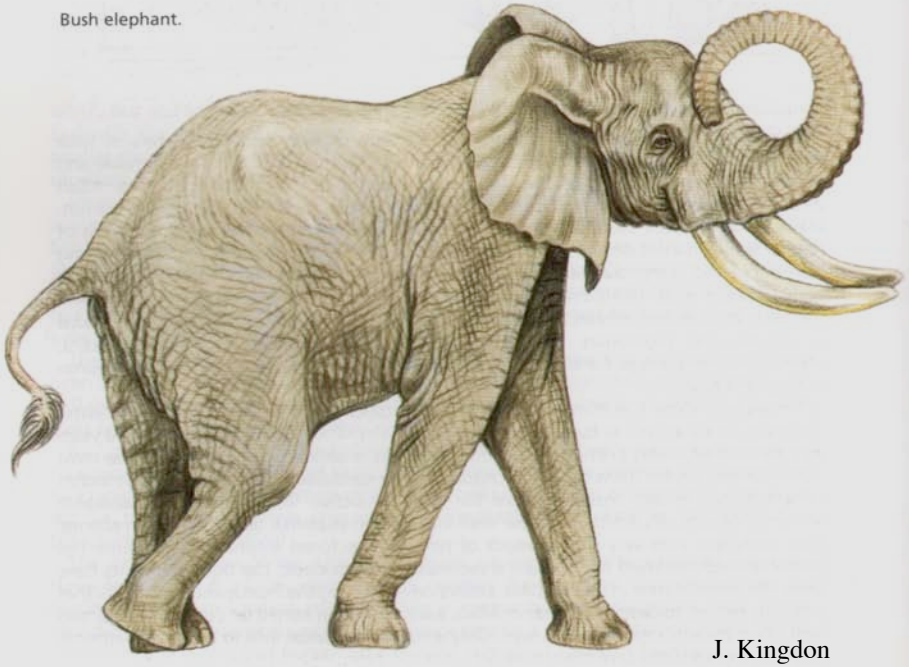
Cytonuclear disequilibrium: mtDNA haplotypes examined for association with nuclear markers. The 2 are completely unlinked, but comparison can be useful for determining direction of crosses in hybrids between species.

For example, if the only crosses are females of species 1 mating with males of species 2, the hybrids will have mtDNA only from species 1 and Y-chromosomes only from species 2, with intermediate levels of other markers.

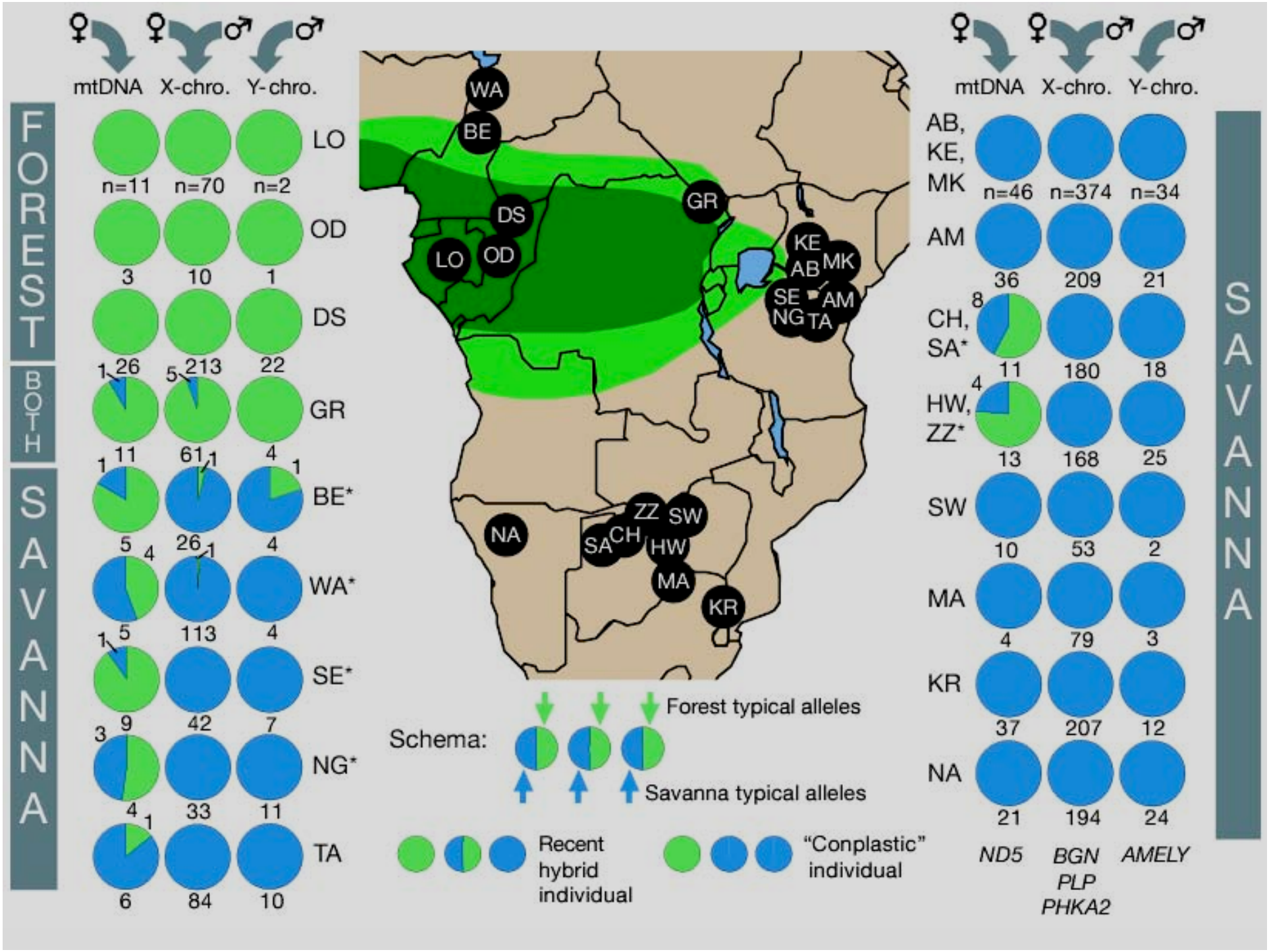
Forest elephant.



Bush elephant.



J. Kingdon


♀ ♀ ♂ ♂
mtDNA X-chro. Y-chro.
FOREST
BOTH
SAVANNA
LO
n=11 n=70 n=2
OD
3 10 1
DS
1 26 5 213 22
GR
1 11 61 1 4
BE*
1 5 4 26 1 4
WA*
4 1 113 4
SE*
1 5 9 42 7
NG*
3 4 33 11
TA
1 6 84 10
WA
BE
DS
LO
OD
GR
KE
AB
MK
SE
NG
TA
AM
NA
ZZ
SW
SA
CH
HW
MA
KR
Schema:
Forest typical alleles
Savanna typical alleles
Recent hybrid individual
"Conplastic" individual
♀ ♀ ♂ ♂
mtDNA X-chro. Y-chro.
SAVANNA
AB, KE, MK
n=46 n=374 n=34
AM
36 209 21
CH, SA*
8 11 180 18
HW, ZZ*
4 13 168 25
SW
10 53 2
MA
4 79 3
KR
37 207 12
NA
21 194 24
ND5
BGN PLP PHKA2
AMELY

Rate of recombination, c: estimated as the proportion of recombinant gametes produced from a parent.

Recombination may vary by sex, in which case the average of **c** for the two sexes may be estimated.

“hot spots” or “cold spots” exist in the genome, with regions of higher- or lower-than-average recombination.

Map units = number of centimorgans, cM:
a 1% chance that a marker at one locus on a chromosome will be separated from a marker at a second locus due to crossing over in a single generation.

Centimorgans or map units are the measure of recombinant frequency for measuring genetic linkage. It is often used to imply distance along a chromosome, on average, 1 million base pairs in humans is approximately one centimorgan.

TABLE 10.6 For each chromosome in the human genome, the physical length, the estimated genetic length in meioses from females, males, and the average over both sexes, and the ratio of the genetic to physical lengths (cM/Mb). The data are calculated from 1257 meiotic events from 146 Icelandic families using 5146 microsatellite markers (Kong *et al.*, 2002).

Chromosome	*Physical length (Mb)*	*Genetic length (cM)*			*cM/Mb (sex average)*
		Female	*Male*	*Sex average*	
1	282.6	345.4	195.1	270.3	0.96
2	252.5	325.4	189.6	257.5	1.02
3	224.5	275.6	160.7	218.2	0.97
4	205.4	259.1	146.5	202.8	0.99
5	199.2	260.2	151.2	205.7	1.03
6	190.9	241.6	137.6	189.6	0.99
7	168.5	230.3	128.4	179.3	1.06
8	158.1	209.9	107.9	158.9	1.01
9	150.2	198.2	117.2	157.7	1.05
10	145.6	218.1	133.9	176.0	1.21
11	153.0	195.5	109.4	152.4	1.00
12	153.4	206.6	135.5	171.1	1.12
13	100.4	155.9	101.3	128.6	1.28
14	87.1	142.4	94.6	119.5	1.36
15	87.2	155.0	102.6	128.8	1.48
16	106.4	149.6	108.1	128.9	1.21
17	89.4	161.5	108.6	135.0	1.51
18	89.4	142.57	98.6	120.6	1.35
19	69.4	126.8	92.6	109.7	1.58
20	59.4	122.0	74.7	98.4	1.66
21	30.0	76.4	47.3	61.9	2.06
22	31.2	82.8	49.0	65.9	2.11
X	156.8	179.9	—	179.0	1.14
Total or mean	3190.8	4460.0	2590.5	3614.7	1.13

Linkage disequilibrium can be affected by:

Genetic drift: non-random associations between alleles can result from stochastic effects in small populations

Mutation: generates new gamete types on one haplotype background

Inbreeding: reduces proportion of double heterozygotes (recombination in regions that are identical by descent does not lower LD).

Linkage disequilibrium can be affected by:

Gene flow: hybridization causes linkage disequilibrium if the allele frequencies at both loci are different between the two subpopulations; double heterozygotes are reduced within subpopulations due to the Wahlund effect.