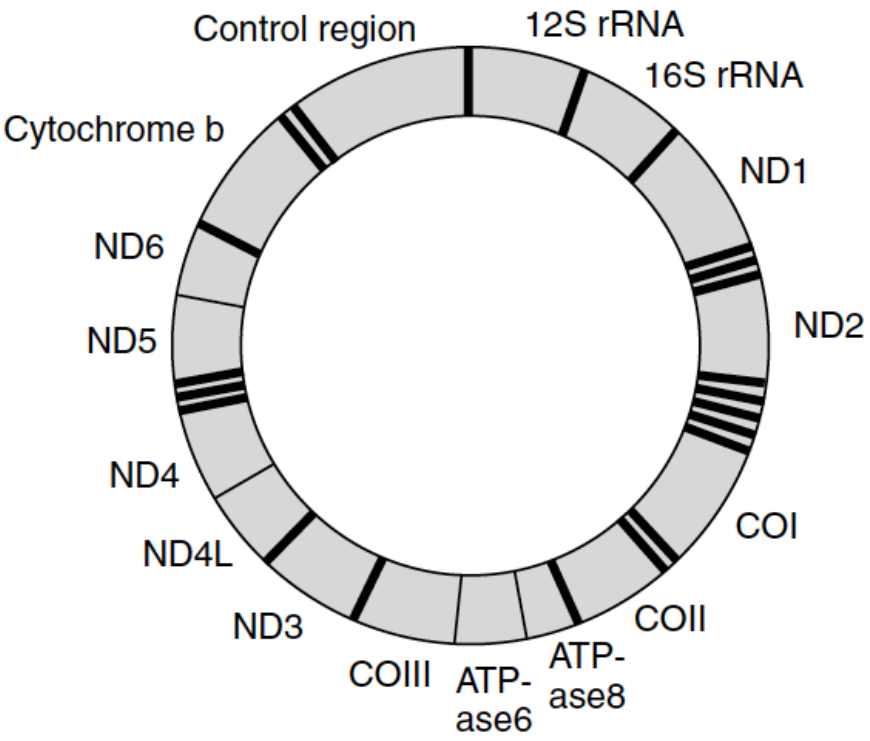
689 Special Topics in Ecological Genomics

Spring 2015

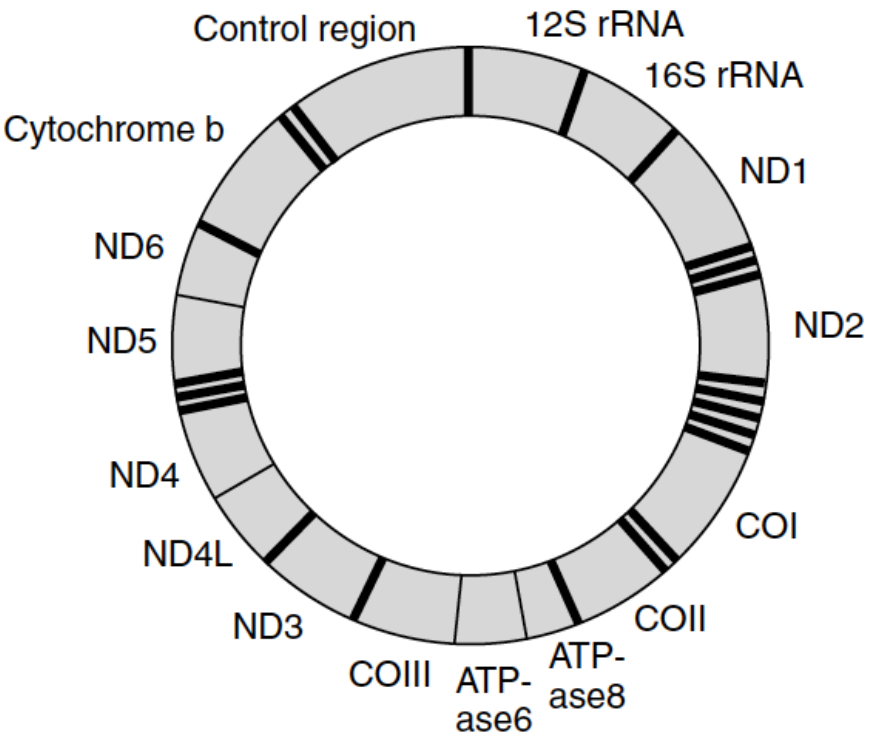
January 22, 2015

Animal mtDNA



Exceptions: heteroplasmy, paternal leakage, intra- and interspecific recombination

Animal mtDNA



Haploid and maternally inherited

Easy to work with (small size, conserved gene content and gene order)

Mutation rate is HIGH (~10x nuclear DNA in animals), especially in the D-loop

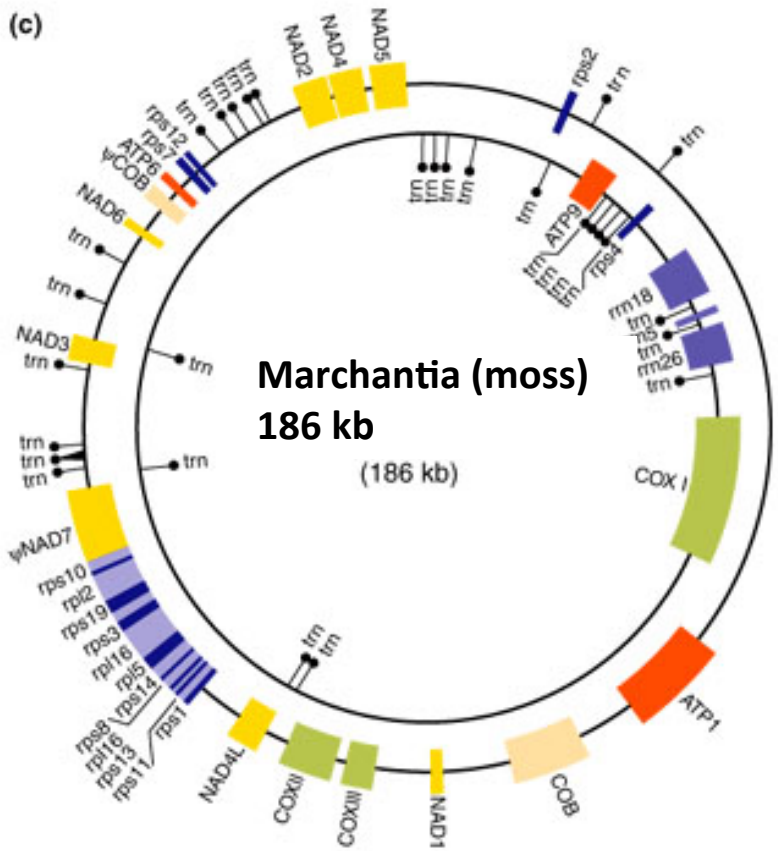
Lack of recombination

Evolutionary neutrality

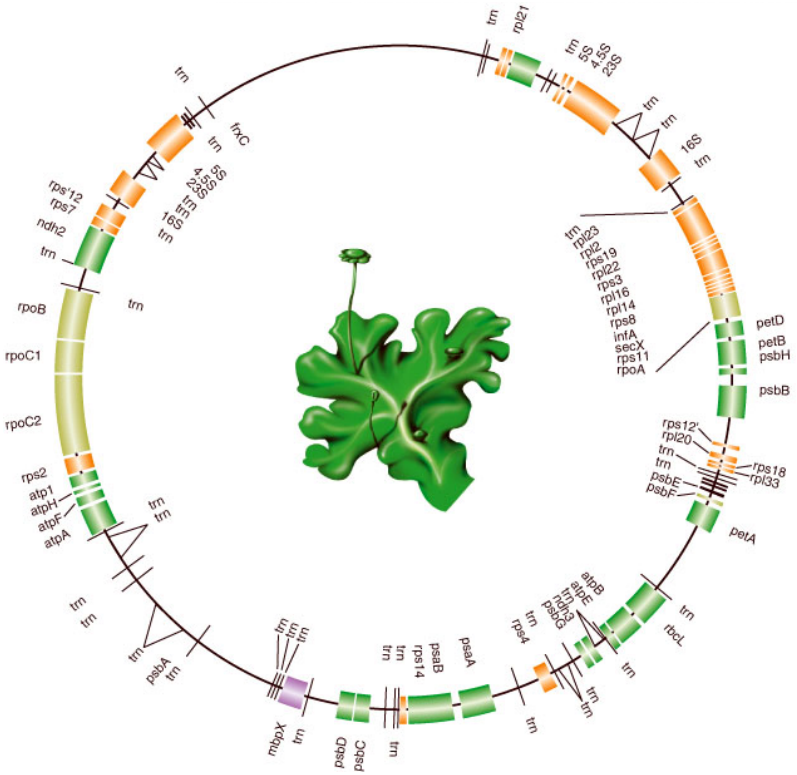
Good marker (supposedly) for Phyogeography and population ecology

Exceptions: heteroplasmy, paternal leakage, intra- and interspecific recombination

Plant mtDNA


(c)
Marchantia (moss)
186 kb
(186 kb)
NAD5
NAD4
NAD2
rps2
trn
ATP9
rps4
rrn18
rrn5
rrn26
COX I
ATP1
COB
NAD1
COXIII
COXII
NAD4L
rps1
rps11
rpl16
rps13
rps8
rpl5
rpl16
rps3
rps19
rpl2
rps10
ψNAD7
NAD3
NAD6
ATP6
ψCOB
rps12
rps7

Chloroplast genome (cpDNA)



Marchantia (moss)
CpDNA 121 kb

Haploid

Maternally inherited (except gymnosperms)

Relatively stable length and gene content/ gene order

Length variation due to changes in repeats length

Low levels of recombination

Mutation rate is 3X mtDNA, 4-5 lower than nuclear DNA

Microsats in cpDNA rapidly evolving

Common heteroplasmy

Useful marker for seed dispersal studies

What's missing from this table?

Table 2.1 Usual mode of inheritance of different genomic regions in sexually reproducing taxa

Genomic region	Typical mode of inheritance
Animals with XY chromosomes	
Autosomal chromosomes	Biparental
Mitochondrial DNA	Maternal in most animals
	Biparental in some bivalves
Y chromosome	Paternal
Higher plants	
Autosomal chromosomes	Biparental
Mitochondrial DNA	Usually maternal
Plastid DNA (including chloroplast DNA)	Maternal in most angiosperms
	Paternal in most gymnosperms
	Biparental in some plants
Y chromosome	Paternal in some dioecious plants

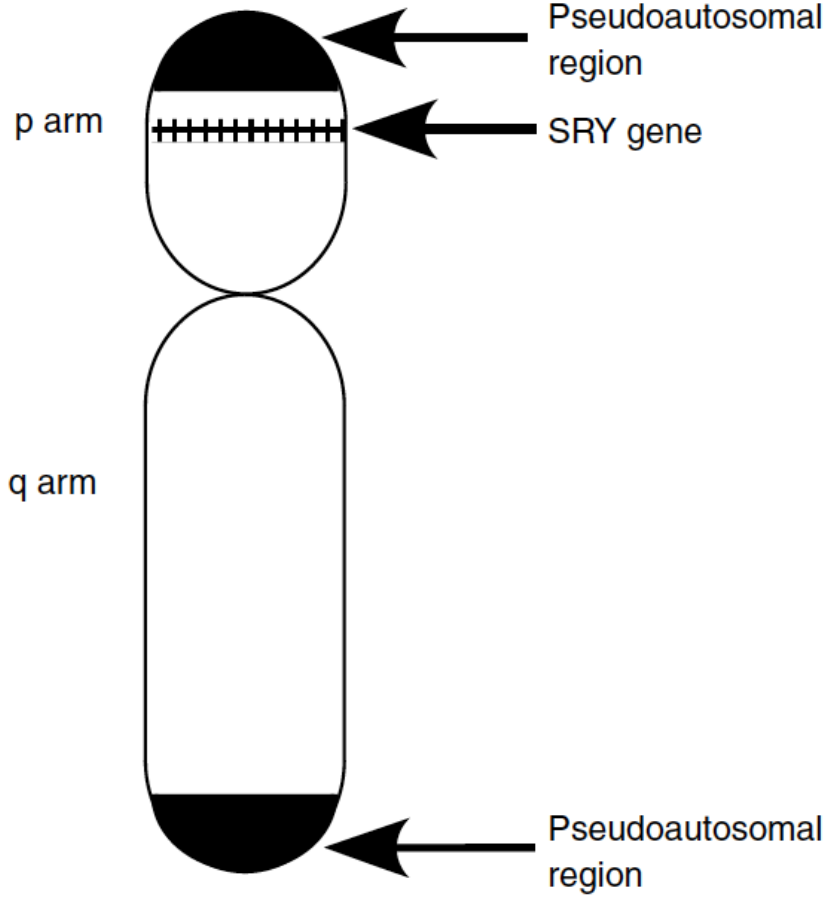
What's missing from this table?

Table 2.1 Usual mode of inheritance of different genomic regions in sexually reproducing taxa

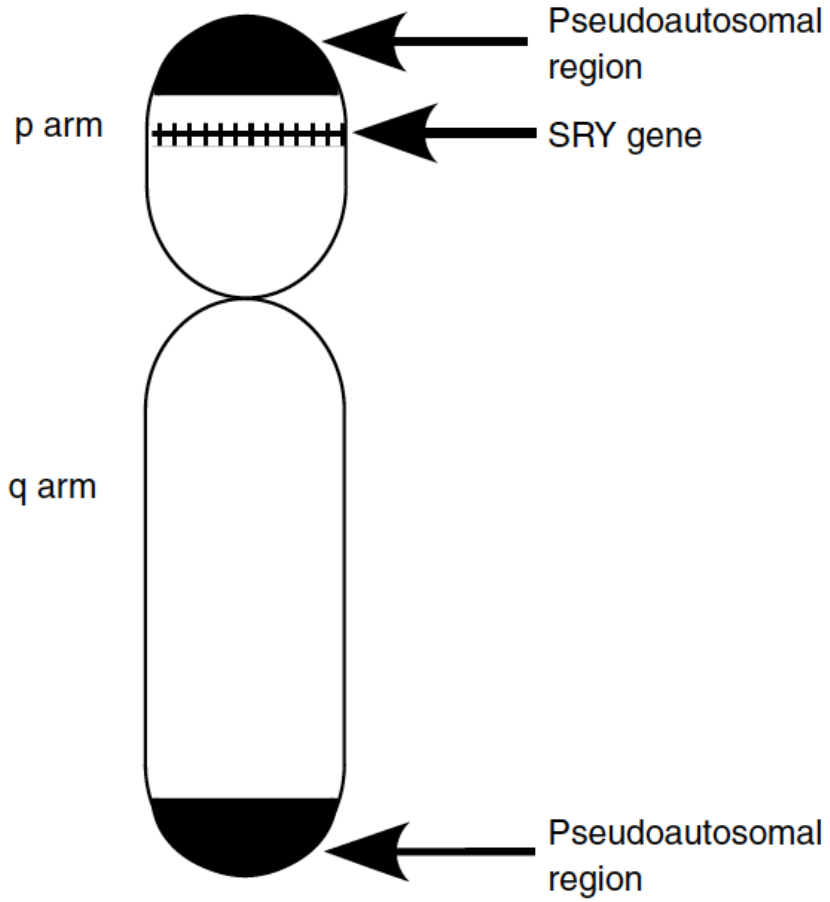
Genomic region	Typical mode of inheritance
Animals with XY chromosomes	
Autosomal chromosomes	Biparental
Mitochondrial DNA	Maternal in most animals
	Biparental in some bivalves
Y chromosome	Paternal
Higher plants	
Autosomal chromosomes	Biparental
Mitochondrial DNA	Usually maternal
Plastid DNA (including chloroplast DNA)	Maternal in most angiosperms
	Paternal in most gymnosperms
	Biparental in some plants
Y chromosome	Paternal in some dioecious plants

ZW sex chromosomes in animals and plants

Y chromosome in mammals



Y chromosome in mammals



Only in males, haploid

Non-recombining with X, except pseudoautosomal regions

Mutation rate is high in non-coding DNA, low in genes (~20 genes total)

Useful marker for paternal genotypes

Canyon tree frog study

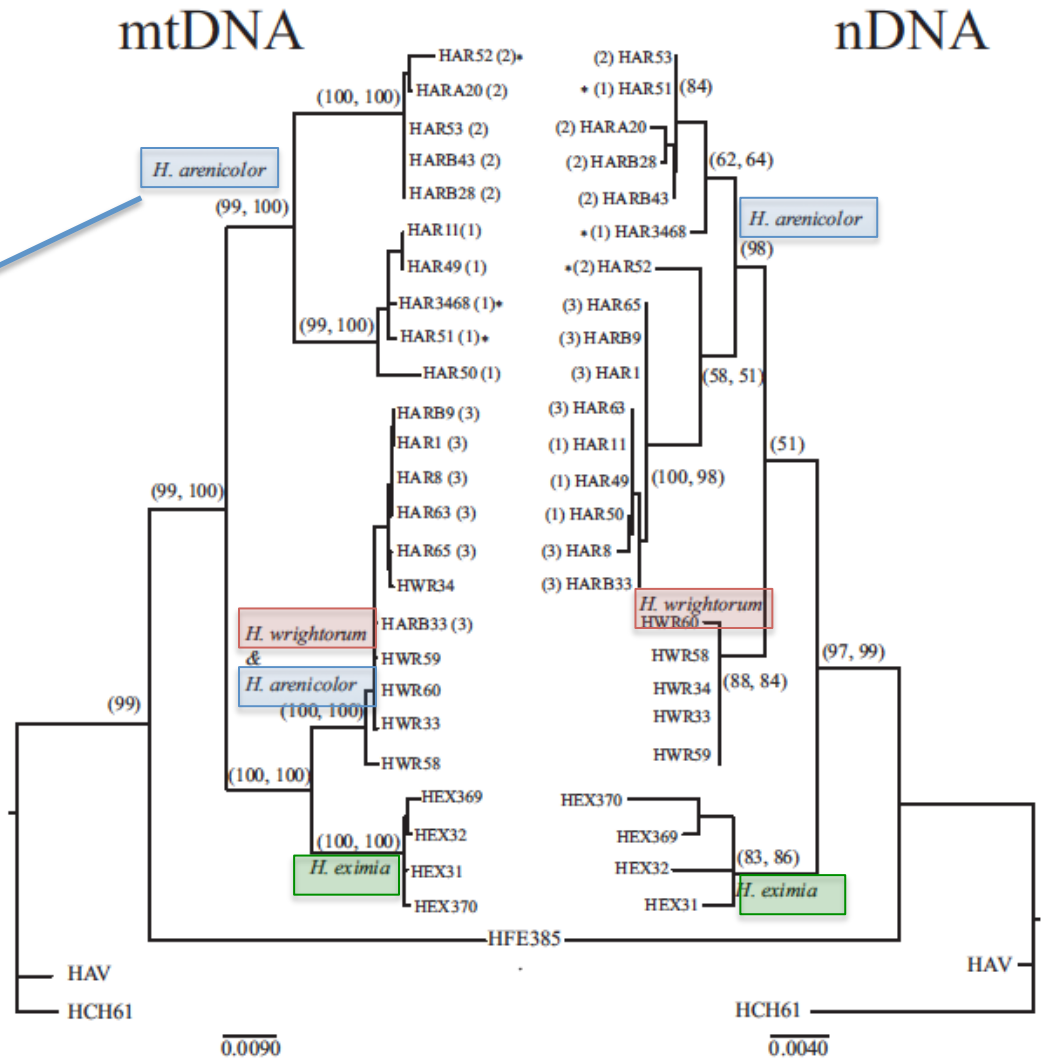
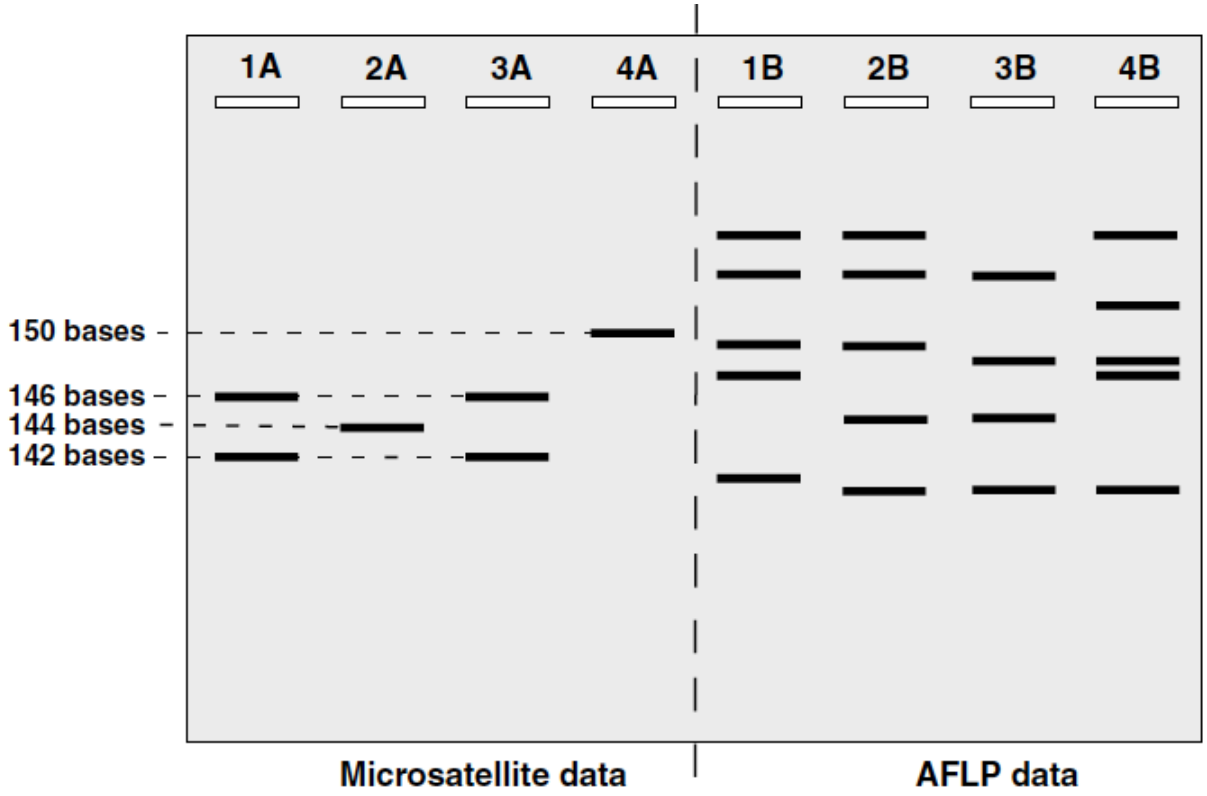


Fig. 4. Klymus et al, Mol. Ecol. 2010,


1A
2A
3A
4A
1B
2B
3B
4B
150 bases
146 bases
144 bases
142 bases
Microsatellite data
AFLP data

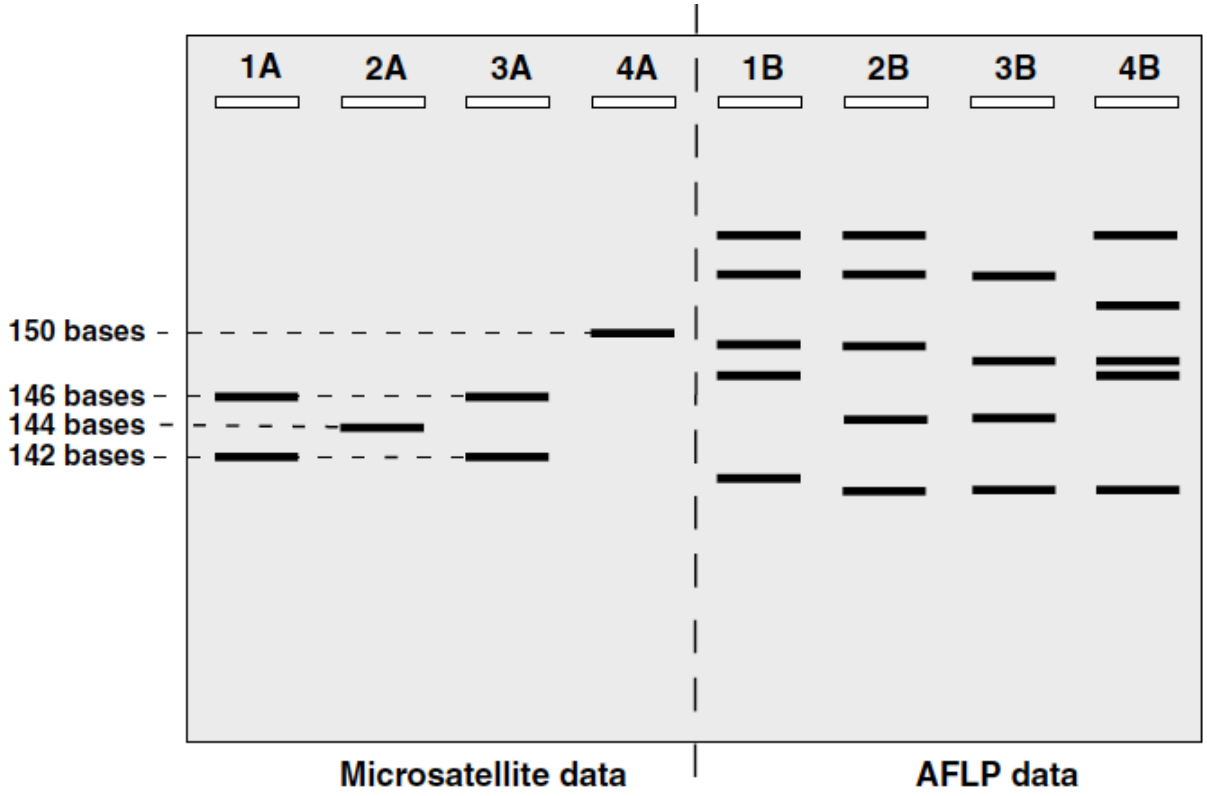


Figure 2.7 A gel showing the genotypes of four individuals based on one microsatellite (co-dominant) locus (1A–4A), and several AFLP (dominant) loci (1B–4B). According to the microsatellite locus, individuals 1 and 3 are heterozygous for alleles that are 142 and 146 bases long, whereas individuals 2 and 4 are homozygous for alleles that are 144 and 150 bases, respectively. Since there are two of each allele in this sample of eight alleles, the frequency of each microsatellite allele is 0.25. According to the AFLP marker, which screens multiple loci, all four individuals are genetically distinct, but we cannot identify homozygotes and heterozygotes, nor can we readily calculate allele frequencies.

Co-dominant markers

Used to identify all of the alleles present at a particular locus: Allow distinguishing between homozygous and heterozygous individuals

More time-consuming

Multiple loci usually sampled

Co-dominant markers

Allozymes

Co-dominant markers

Allozymes

- Relatively easy method; samples collection not always feasible
- Conservative estimates of genome variation (no synonymous sites)
- Sample only part of the genome (coding sequences)
- Not always neutral: good for some studies of adaptation

RFLPs

Co-dominant markers

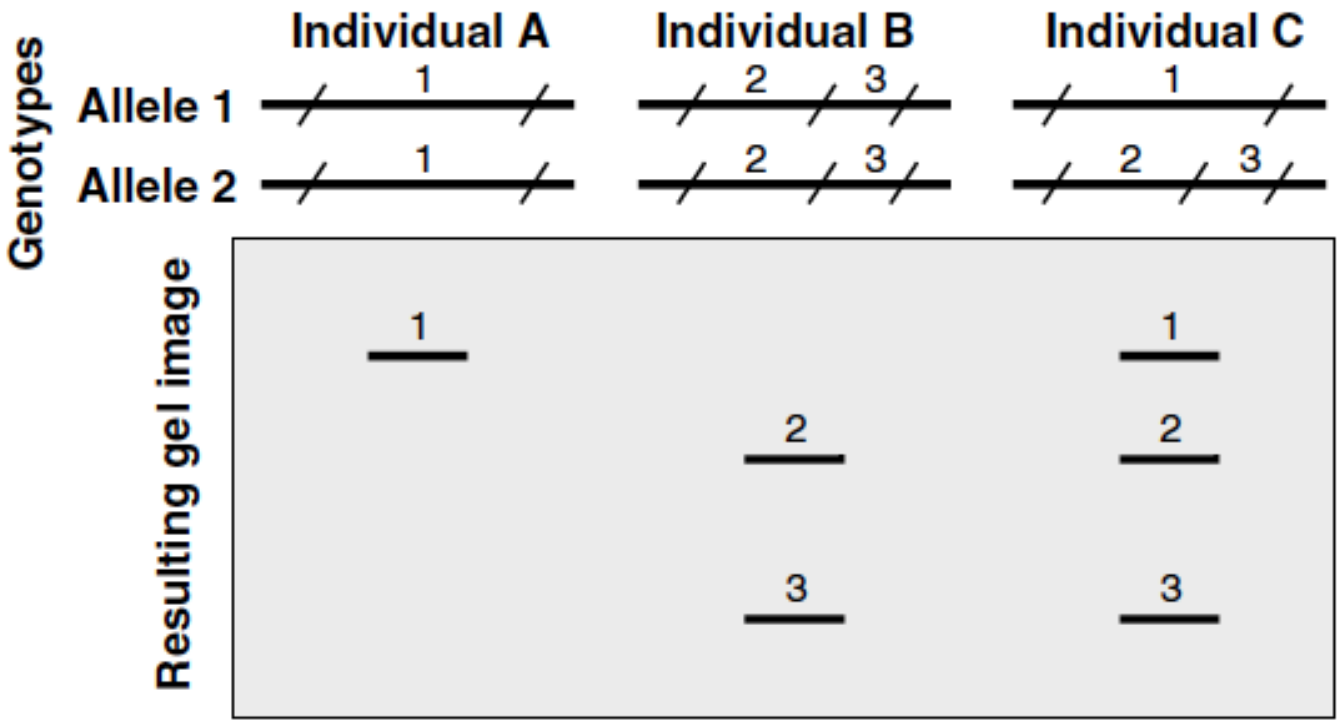
Allozymes

- Relatively easy method; samples collection not always feasible
- Conservative estimates of genome variation (no synonymous sites)
- Sample only part of the genome (coding sequences)
- Not always neutral: good for some studies of adaptation

RFLPs

- Restriction fragment length polymorphisms
- Conservative estimates of genome variation (no synonymous sites)
- Sample only part of the genome (coding sequences)
- Not always neutral: good for some studies of adaptation

Microsatellites


Genotypes
Individual A
Individual B
Individual C
Allele 1
Allele 2
1
2
3
Resulting gel image

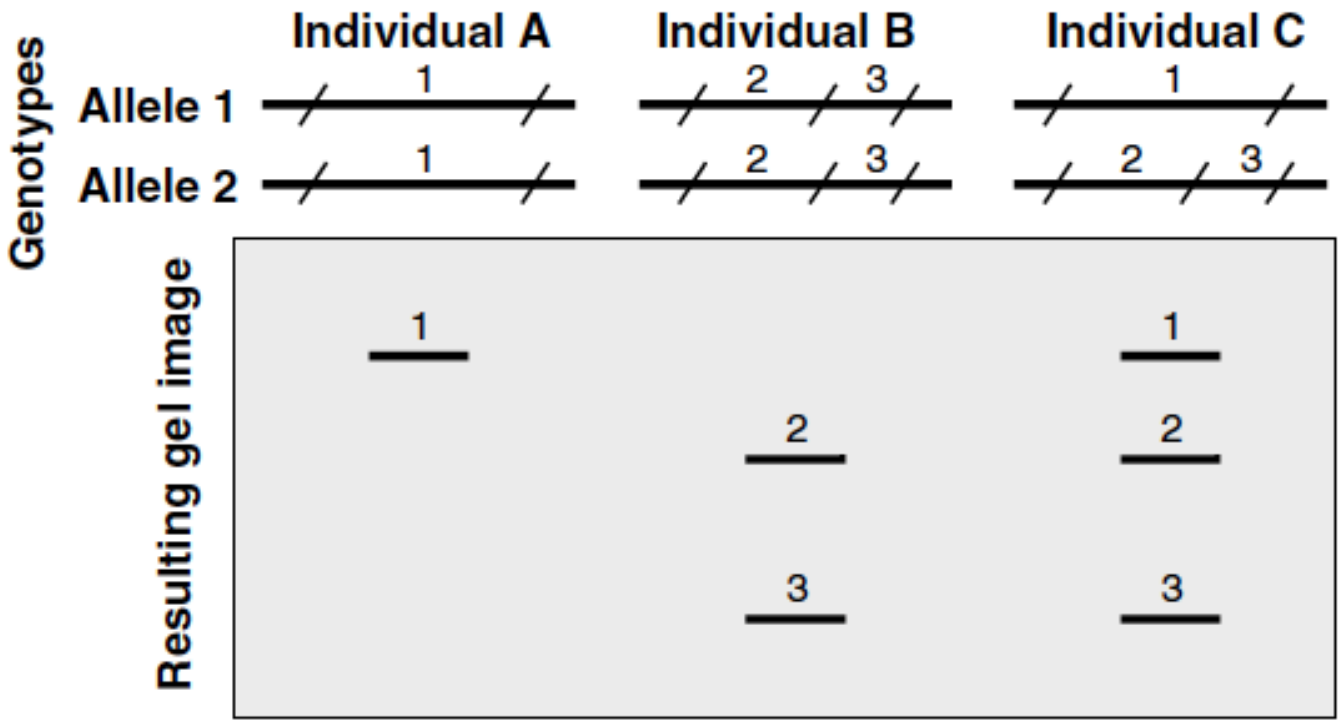


Figure 2.9 Three different RFLP genotypes result from sequence differences that affect the restriction enzyme recognition sites (designated as/). At this locus, individuals A and B are homozygous for alleles that have two and three restriction sites, respectively. Individual C is heterozygous, with two restriction sites at one allele and three restriction sites at the other allele. The numbers of bands that would be generated by the RFLP profiles are shown in the resulting gel image.

Dominant markers

RAPD

- Random amplified polymorphic DNA: generates random bands
- PCR-based
- Largely abandoned due to lack in reproducibility and availability of better methods

AFLPs

- Amplified fragment length polymorphisms: generates random bands
- Restriction enzymes-based: two different Res are used
- High reproducibility

Microsatellites

- Nuclear and chloroplast genomes, some mitochondrial genomes
- High mutation rates ~10^{-4}-10^{-6}
- Variable mutation rates across microsatellite loci
- Size homoplasy is common

Co-dominant markers

Allozymes

- Relatively easy method; samples collection not always feasible
- Conservative estimates of genome variation (no synonymous sites)
- Sample only part of the genome (coding sequences)
- Not always neutral: good for some studies of adaptation

RFLPs

- Restriction fragment length polymorphisms
- Conservative estimates of genome variation (no synonymous sites)
- Sample only part of the genome (coding sequences)
- Not always neutral: good for some studies of adaptation

Microsatellites

- Nuclear and chloroplast genomes, some mitochondrial genomes
- High mutation rates ~10^{-4}-10^{-6}
- Variable mutation rates across microsatellite loci
- Size homoplasy is common

Dominant markers

Find only a single dominant allele per locus

Less time-consuming

Multiple loci tested

Good for population studies (polymorphisms), bad for distant evolutionary relationships analyses

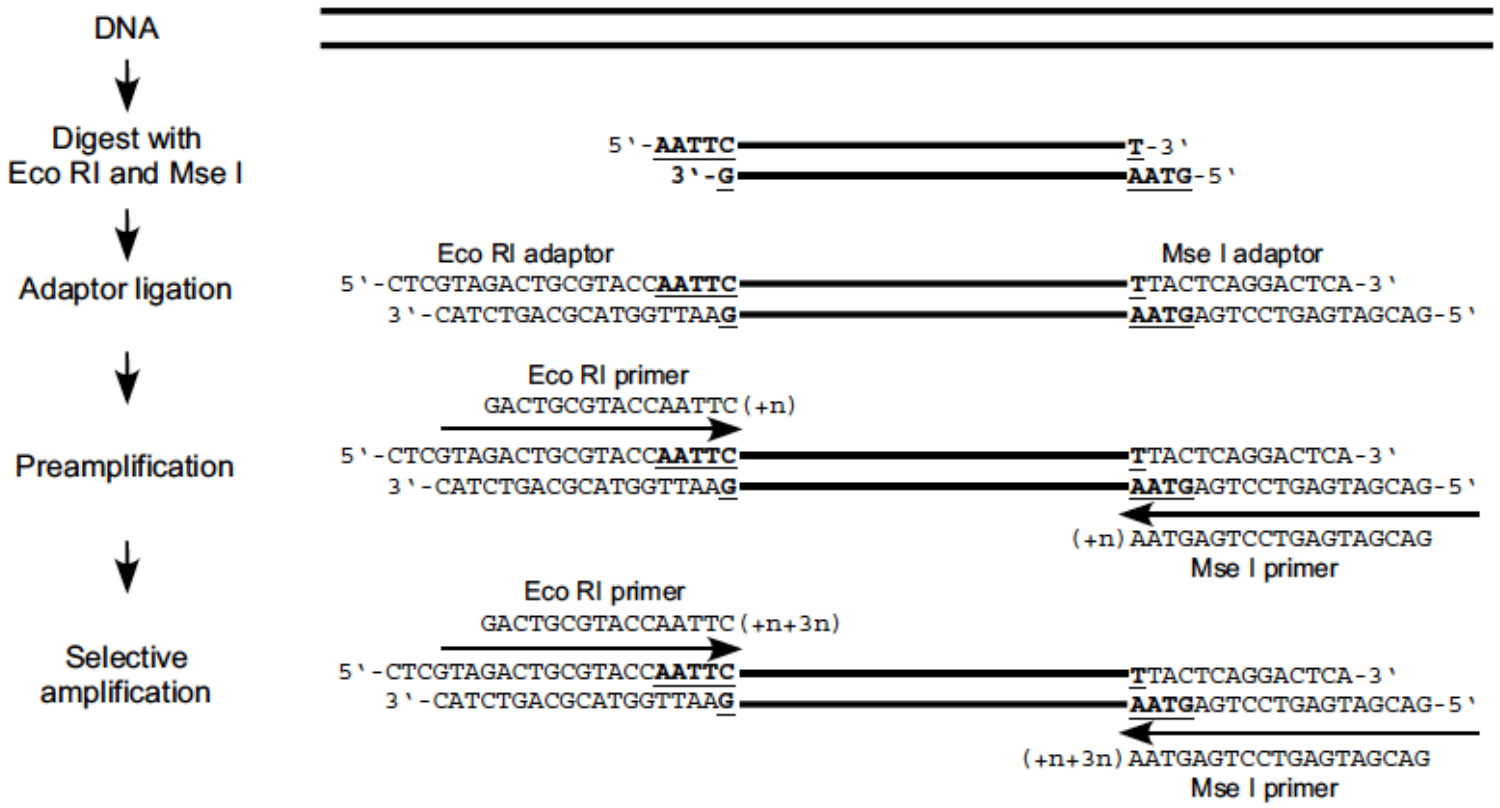


Figure 2.15 A schematic diagram showing how AFLP genotypes are generated. Digestion with two restriction enzymes produces sticky ends to which linkers can be ligated. During preamplification, the addition of a single base to the 30 end of each primer will reduce the number of amplified fragments to 1/16 of the number of fragments that would otherwise be amplified. The addition of three more bases to the 30 primer ends during selective amplification further reduces the chance of a perfect match between primers and target sequences, and as a result only 1/65 536 of the original set of fragments will be amplified.

Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals

Eric Bazin, Sylvain Glemin, Nicolas Galtier—*Science*, 2006

Questions: is mtDNA really neutrally evolving in animals? Can we use mtDNA markers to estimate population size, which is crucial for example for conservation purposes?

"For a neutral locus, **the expected polymorphism at mutation-drift equilibrium is proportional to the effective population size**, the equivalent number of breeders in an ideal, panmictic population."
Other factors that affect polymorphisms are: population structure, bottlenecks, natural selection, life cycle, mating system.

“Population size […] presumably varies by several orders of magnitude between species and taxa, so that one would typically predict that abundant species should be, on average, more polymorphic than scarce ones despite the noise introduced by other evolutionary forces.”

This is true according to meta-analysis of allozymes in invertebrates. mtDNA-based markers are supposedly better than allozymes at capturing the effective population size

In this study they used 3 datasets (hundreds of species in each dataset):
1: allozymes
2: nuclear markers
3: mtDNA markers

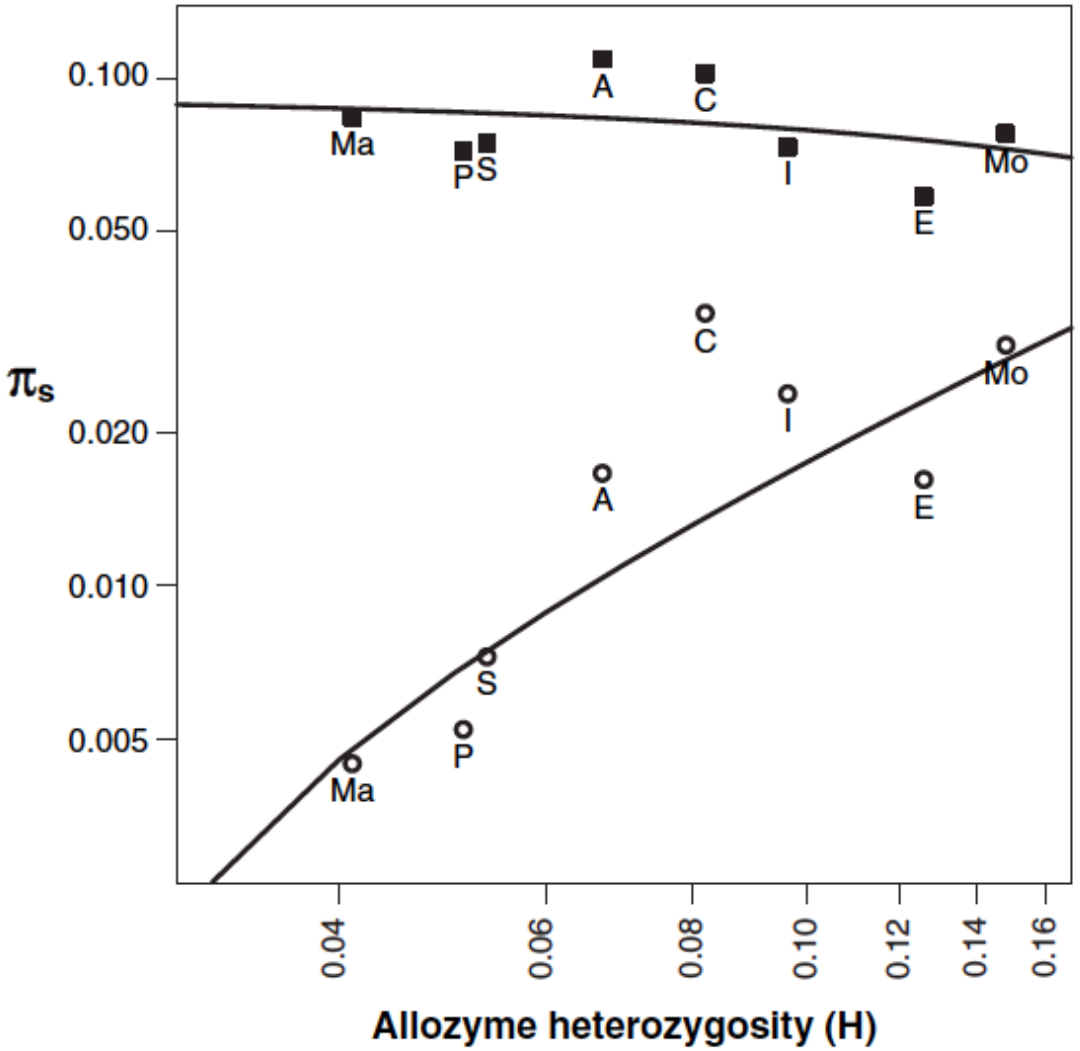


Fig. 1. Average allozymic, nuclear DNA, and mtDNA diversity in eight animal taxa. *x* axis: allozyme average heterozygosity. *y* axis: circles, nuclear DNA average synonymous diversity (kendall test: $\tau = 0.87$, $P < 0.05$); squares, mtDNA average synonymous diversity (kendall test: $\tau = -0.14$, not significant). Ma: Mammalia (allozymes: 184 species; nuclear: 30 species; mtDNA: 350 species); S: Sauropsida (reptiles and birds: 116, 20, 378); A: Amphibia (61, 4, 96); P: Pisces (bony fish and cartilaginous fish: 183, 22, 270); I: Insecta (156, 73, 511); C: Crustacea (122, 2, 78); E: Echinodermata (sea stars and urchins: 15, 14, 47); and Mo: Mollusca (46, 9, 125). The nuclear averages of the little-represented Amphibia (four species) and Crustacea (two species) are shown but were not used for the statistical test.

Table 1. Ecological determinism of allozyme and mtDNA genetic diversity. The numbers of species used are shown in parentheses.

Taxon		Allozymes (H, %)	mtDNA (π_s, %)
Fish	Freshwater	4.7 (71)	8.7** (123)
	Marine	6.1* (65)	3.7 (51)
Crustaceans	Large benthic	4.6 (81)	10.1 (26)
	Small planktonic	21.0* (8)	5.8 (6)
Mollusks	Terrestrial	7.4 (23)	7.8 (8)
	Marine	30.0** (17)	5.6 (34)

*$P < 0.05$ (Student's t test). **$P < 0.01$ (Student's t test).

Allozyme data agree with expectations about population sizes. For example, "mollusks, the terrestrial pulmonates were substantially less polymorphic than marine bivalves or gastropods, consistent with the enormous dispersal potential of the latter"; among crustaceans, the microscopic, planktonic branchiopods (e.g., Artemia and Daphnia) appeared much more diverse than the larger decapods (shrimps, lobsters, and crabs). [..] The mtDNA diversity, in contrast, failed to reflect these differences in average population size."

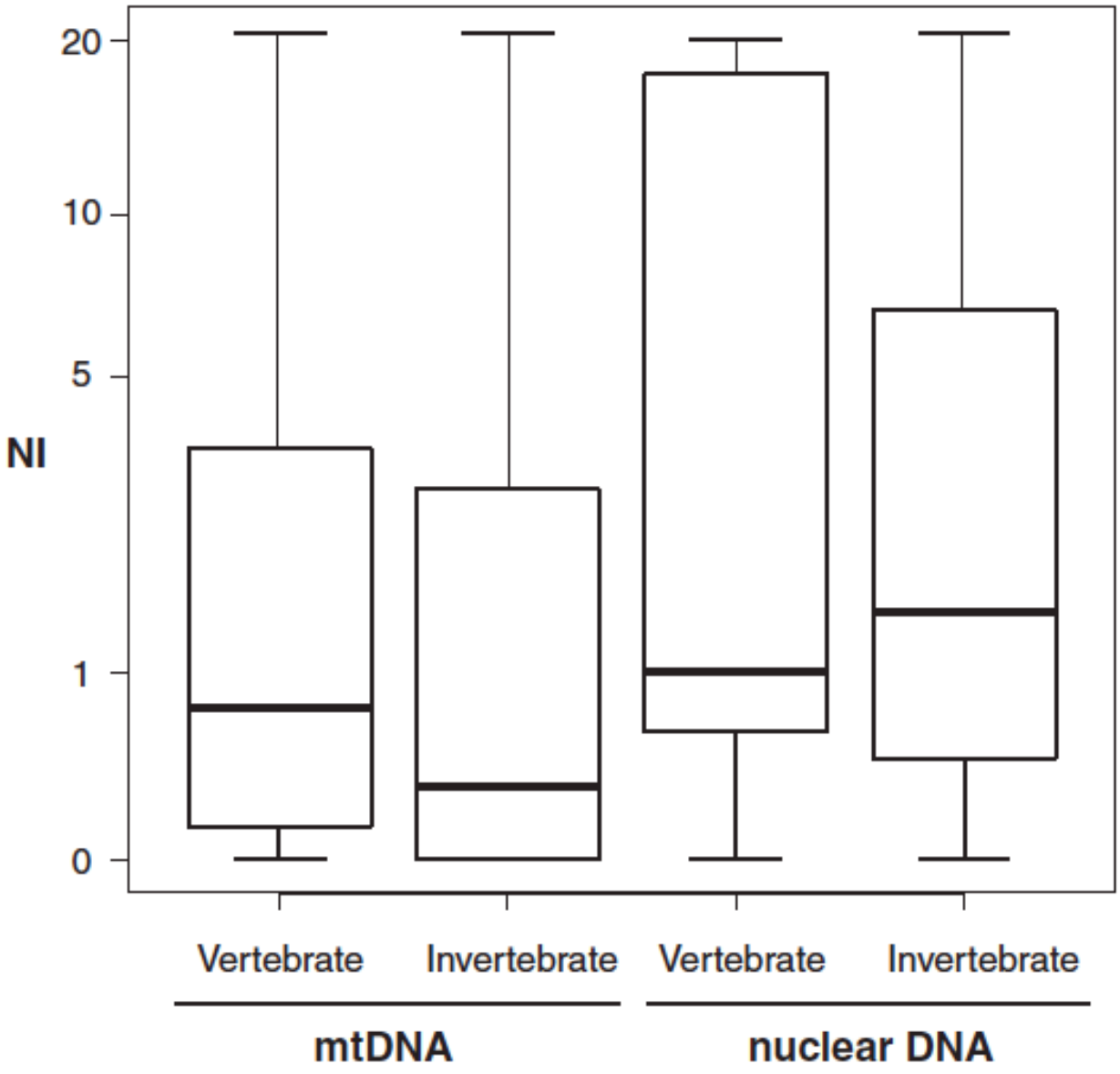
"The mtDNA pattern, however, appears to be in good agreement with the hypothesis of recurrent fixation of advantageous mutations leading to frequent loss of variability at linked loci, a process recently named "genetic draft" by Gillespie." J. H. Gillespie, Genetics 155, 909 (2000).
"The gene-dense, non-recombining context of the animal mitochondrial genome maximizes the potential impact of the genetic draft, as compared with that of the nuclear genome"

TEST: neutrality index (NI)

NI: ratio of nonsynonymous (amino acid–changing) to synonymous (silent) changes within species (π_N/π_S) and between species (dN/dS):

NI = 1 means evolution is neutral
NI > 1 under purifying selection
NI < 1 in case of adaptation

Fig. 2. Neutrality index (NI) distributions (logarithmic scale). Medians are indicated by thick horizontal bars. Boxes include 50% of the distributions. The invertebrate mtDNA median NI (0.42) is significantly lower than the vertebrate one (0.88; $P < 10^{-3}$, Mann-Whitney test). NI values greater than 20 were forced to 20 for clarity. Low-frequency (<0.125) polymorphic sites were excluded from the analysis.



CONCLUSIONS

"Natural selection acting on mtDNA contributes to homogenization of the average diversity among groups, in agreement with the genetic draft theory.
mtDNA appears to be anything but a neutral marker and probably undergoes frequent adaptive evolution"