MATHEMATICAL THEORY OF POPULATION GENETICS

Fall 2012

Department of mathematics and statistics

University of Helsinki

KREITMAN, M. 1983. NUCLEOTIDE POLYMORPHISM AT THE ALCOHOL DEHYDROGENASE LOCUS OF *DROSOPHILA MELANOGAITER*.. NATURE 304: 412-417

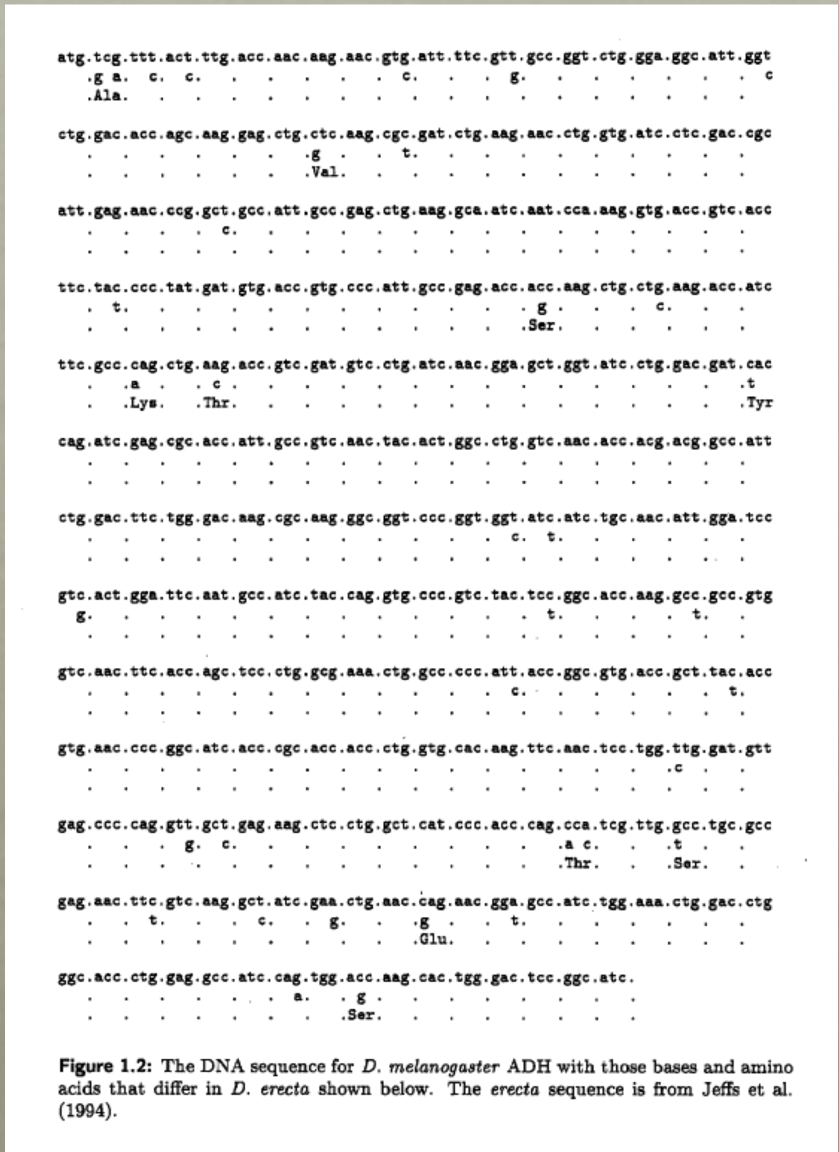
```
1                                                 g
atg.tcg.ttt.act.ttg.acc.aac.aag.aac.gtg.att.ttc.gtt.gcc.ggt.ctg.gga.ggc.att.ggt
Met.Ser.Phe.Thr.Leu.Thr.Asn.Lys.Asn.Val.Ile.Phe.Val.Ala.Gly.Leu.Gly.Gly.Ile.Gly
61
ctg.gac.acc.agc.aag.gag.ctg.ctc.aag.cgc.gat.ctg.aag.aac.ctg.gtg.atc.ctc.gac.cgc
Leu.Asp.Thr.Ser.Lys.Glu.Leu.Leu.Lys.Arg.Asp.Leu.Lys.Asn.Leu.Val.Ile.Leu.Asp.Arg
121
att.gag.aac.ccg.gct.gcc.att.gcc.gag.ctg.aag.gca.atc.aat.cca.aag.gtg.acc.gtc.acc
Ile.Glu.Asn.Pro.Ala.Ala.Ile.Ala.Glu.Leu.Lys.Ala.Ile.Asn.Pro.Lys.Val.Thr.Val.Thr
181                                                         t
ttc.tac.ccc.tat.gat.gtg.acc.gtg.ccc.att.gcc.gag.acc.acc.aag.ctg.ctg.aag.acc.atc
Phe.Tyr.Pro.Tyr.Asp.Val.Thr.Val.Pro.Ile.Ala.Glu.Thr.Thr.Lys.Leu.Leu.Lys.Thr.Ile
241
ttc.gcc.cag.ctg.aag.acc.gtc.gat.gtc.ctg.atc.aac.gga.gct.ggt.atc.ctg.gac.gat.cac
Phe.Ala.Gln.Leu.Lys.Thr.Val.Asp.Val.Leu.Ile.Asn.Gly.Ala.Gly.Ile.Leu.Asp.Asp.His
301
cag.atc.gag.cgc.acc.att.gcc.gtc.aac.tac.act.ggc.ctg.gtc.aac.acc.acg.acg.gcc.att
Gln.Ile.Glu.Arg.Thr.Ile.Ala.Val.Asn.Tyr.Thr.Gly.Leu.Val.Asn.Thr.Thr.Thr.Ala.Ile
361                               t       a
ctg.gac.ttc.tgg.gac.aag.cgc.aag.ggc.ggt.ccc.ggt.ggt.atc.atc.tgc.aac.att.gga.tcc
Leu.Asp.Phe.Trp.Asp.Lys.Arg.Lys.Gly.Gly.Pro.Gly.Gly.Ile.Ile.Cys.Asn.Ile.Gly.Ser
421                       a
gtc.act.gga.ttc.aat.gcc.atc.tac.cag.gtg.ccc.gtc.tac.tcc.ggc.acc.aag.gcc.gcc.gtg
Val.Thr.Gly.Phe.Asn.Ala.Ile.Tyr.Gln.Val.Pro.Val.Tyr.Ser.Gly.Thr.Lys.Ala.Ala.Val
481                                       a       c               g           t
gtc.aac.ttc.acc.agc.tcc.ctg.gcg.aaa.ctg.gcc.ccc.att.acc.ggc.gtg.acc.gct.tac.acc
Val.Asn.Phe.Thr.Ser.Ser.Leu.Ala.Lys.Leu.Ala.Pro.Ile.Thr.Gly.Val.Thr.Ala.Tyr.Thr
541                                              c
gtg.aac.ccc.ggc.atc.acc.cgc.acc.acc.ctg.gtg.cac.aag.ttc.aac.tcc.tgg.ttg.gat.gtt
Val.Asn.Pro.Gly.Ile.Thr.Arg.Thr.Thr.Leu.Val.His.Lys.Phe.Asn.Ser.Trp.Leu.Asp.Val
601   t           c                                       c
gag.ccc.cag.gtt.gct.gag.aag.ctc.ctg.gct.cat.ccc.acc.cag.cca.tcg.ttg.gcc.tgc.gcc
Glu.Pro.Gln.Val.Ala.Glu.Lys.Leu.Leu.Ala.His.Pro.Thr.Gln.Pro.Ser.Leu.Ala.Cys.Ala
661                           a
gag.aac.ttc.gtc.aag.gct.atc.gag.ctg.aac.cag.aac.gga.gcc.atc.tgg.aaa.ctg.gac.ctg
Glu.Asn.Phe.Val.Lys.Ala.Ile.Glu.Leu.Asn.Gln.Asn.Gly.Ala.Ile.Trp.Lys.Leu.Asp.Leu
721
ggc.acc.ctg.gag.gcc.atc.cag.tgg.acc.aag.cac.tgg.gac.tcc.ggc.atc.
Gly.Thr.Leu.Glu.Ala.Ile.Gln.Trp.Thr.Lys.His.Trp.Asp.Ser.Gly.Ile.
```

Figure 1.1: The DNA sequence for the coding region of the reference allele from the alcohol dehydrogenase locus of *Drosophila melanogaster*. The translation, given below the DNA sequence, uses the three-letter codes for amino acids. The letters over certain bases indicate the variants for those nucleotides found in a sample from nature. The variant at position 578 changes the amino acid of its codon from lysine to threonine.

Allele	39	226	387	393	441	513	519	531	540	578	606	615	645	684
Reference	T	C	C	C	C	C	T	C	C	A	C	T	A	G
Wa-S	.	T	T	.	A	A	C	.	.	.	.	.	.	.
Fl-1S	.	T	T	.	A	A	C	.	.	.	.	.	.	.
Af-S	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Fr-S	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Fl-2S	G	.	.	.	.	.	.	.	.	.	.	.	.	.
Ja-S	G	.	.	.	.	.	.	.	T	.	T	.	C	A
Fl-F	G	.	.	.	.	.	.	G	T	C	T	C	C	.
Fr-F	G	.	.	.	.	.	.	G	T	C	T	C	C	.
Wa-F	G	.	.	.	.	.	.	G	T	C	T	C	C	.
Af-F	G	.	.	.	.	.	.	G	T	C	T	C	C	.
Ja-F	G	.	.	A	.	.	.	G	T	C	T	C	C	.

Table 1.1: The 11 *ADH* alleles. A dot is placed when a nucleotide is the same as the nucleotide in the reference sequence. The numbers refer to the position in the coding sequence where the 14 variant nucleotides are found (see Figure 1.1). The first two letters of the allele name identify the place of origin. The S alleles have a lysine at position 192 of the protein; the F alleles have a threonine.

First letter	Second letter: U		C		A		G		
U	UUU	Phenyl-alanine (Phe)	UCU	Serine (Ser)	UAU	Tyrosine (Tyr)	UGU	Cysteine (Cys)	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine (Leu)	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Tryptophane (Tryp)	G
C	CUU	Leucine (Leu)	CCU	Proline (Pro)	CAU	Histidine (His)	CGU	Arginine (Arg)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Glutamine (GluN)	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Isoleucine (Ileu)	ACU	Threonine (Thr)	AAU	Asparagine (AspN)	AGU	Serine (Ser)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lysine (Lys)	AGA	Arginine (Arg)	A
	AUG	Methionine (Met)	ACG		AAG		AGG		G
G	GUU	Valine (Val)	GCU	Alanine (Ala)	GAU	Aspartic acid (Asp)	GGU	Glycine (Gly)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamic acid (Glu)	GGA		A
	GUG		GCG		GAG		GGG		G



```
atg.tcg.ttt.act.ttg.acc.aac.aag.aac.gtg.att.ttc.gtt.gcc.ggt.ctg.gga.ggc.att.ggt
   .g a.  c.  c.   .   .   .   .   .  c.   .   .  g.   .   .   .   .   .   .  c
   .Ala.   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

ctg.gac.acc.agc.aag.gag.ctg.ctc.aag.cgc.gat.ctg.aag.aac.ctg.gtg.atc.ctc.gac.cgc
   .   .   .   .   .   .   .g  .   .  t.   .   .   .   .   .   .   .   .   .
   .   .   .   .   .   .   .Val.   .   .   .   .   .   .   .   .   .   .   .

att.gag.aac.ccg.gct.gcc.att.gcc.gag.ctg.aag.gca.atc.aat.cca.aag.gtg.acc.gtc.acc
   .   .   .   .  c.   .   .   .   .   .   .   .   .   .   .   .   .   .   .
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

ttc.tac.ccc.tat.gat.gtg.acc.gtg.ccc.att.gcc.gag.acc.acc.aag.ctg.ctg.aag.acc.atc
   .  t.   .   .   .   .   .   .   .   .   .   .   . g .   .   .  c.   .   .
   .   .   .   .   .   .   .   .   .   .   .   .   .Ser.   .   .   .   .   .

ttc.gcc.cag.ctg.aag.acc.gtc.gat.gtc.ctg.atc.aac.gga.gct.ggt.atc.ctg.gac.gat.cac
   .   .a  .   . c .   .   .   .   .   .   .   .   .   .   .   .   .   .   .t
   .   .Lys.   .Thr.   .   .   .   .   .   .   .   .   .   .   .   .   .   .Tyr

cag.atc.gag.cgc.acc.att.gcc.gtc.aac.tac.act.ggc.ctg.gtc.aac.acc.acg.acg.gcc.att
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

ctg.gac.ttc.tgg.gac.aag.cgc.aag.ggc.ggt.ccc.ggt.ggt.atc.atc.tgc.aac.att.gga.tcc
   .   .   .   .   .   .   .   .   .   .   .   .  c.  t.   .   .   .   .   .
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

gtc.act.gga.ttc.aat.gcc.atc.tac.cag.gtg.ccc.gtc.tac.tcc.ggc.acc.aag.gcc.gcc.gtg
  g.   .   .   .   .   .   .   .   .   .   .   .   .  t.   .   .   .  t.   .
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

gtc.aac.ttc.acc.agc.tcc.ctg.gcg.aaa.ctg.gcc.ccc.att.acc.ggc.gtg.acc.gct.tac.acc
   .   .   .   .   .   .   .   .   .   .   .   .  c.   .   .   .   .   .  t.
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

gtg.aac.ccc.ggc.atc.acc.cgc.acc.acc.ctg.gtg.cac.aag.ttc.aac.tcc.tgg.ttg.gat.gtt
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .c  .   .
   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

gag.ccc.cag.gtt.gct.gag.aag.ctc.ctg.gct.cat.ccc.acc.cag.cca.tcg.ttg.gcc.tgc.gcc
   .   .   .  g.  c.   .   .   .   .   .   .   .   .   .a c.   .   .t  .   .
   .   .   .   .   .   .   .   .   .   .   .   .   .   .Thr.   .   .Ser.   .

gag.aac.ttc.gtc.aag.gct.atc.gaa.ctg.aac.cag.aac.gga.gcc.atc.tgg.aaa.ctg.gac.ctg
   .   .  t.   .   .  c.   .  g.   .   .g  .   .  t.   .   .   .   .   .   .
   .   .   .   .   .   .   .   .   .   .Glu.   .   .   .   .   .   .   .   .

ggc.acc.ctg.gag.gcc.atc.cag.tgg.acc.aag.cac.tgg.gac.tcc.ggc.atc.
   .   .   .   .   .   .  a.   . g .   .   .   .   .   .   .   .
   .   .   .   .   .   .   .   .Ser.   .   .   .   .   .   .   .
```

Figure 1.2: The DNA sequence for *D. melanogaster* ADH with those bases and amino acids that differ in *D. erecta* shown below. The *erecta* sequence is from Jeffs et al. (1994).

What is Population genetics?

- *Population genetics studies how genetic composition of the population changes over time under the influence of various "forces":*
  - Natural selection, mutations, recombination, migration, non-random mating etc.
- *Genetic basis of evolution*
  - Understand the diversity: why and how
  - Predict change
- *What is a population?*
  - Population is a collection of organisms/individuals…

ORGANISMS

- *Prokaryote vs. eukaryote*
  - Prokaryotes: Bacteria, Archae
  - Eukaryotes: all animals, plants, fungi
- *Ploidy (# of sets of chromosomes in the cell)*
  - haploid (e.g. gametes, some ants)
  - diploid (most higher eukaryotic cells, e.g. almost all mammals)
  - polyploid (triploid: seedless watermelon; hexaploid: kiwi, wheat; Extremeploid: ophioglossum, 1260 chromosomes, 84-ploid?)

- *Sexual vs. asexual*
  - Asexual (e.g. prokaryotes, many plants and fungi, some scorpions, some reptiles)
  - Sexual (almost all animals and plants)

LIFE CYCLE
SEXUALLY REPRODUCING DIPLOID ORGANISMS

Mitosis & Meiosis

MENDEL'S LAWS

- *Principle of Segregation*
  - Each gamete (reproductive cell) contains only one of the two alleles AND
  - ..each gamete is equally likely to contain either one
- *Principle of Independent Assortment*
  - When two or more pairs of genes segregate simultaneously, they do so independently
    - e.g. loci on different chromosomes

Allele and Genotype Frequencies