Genetic diversity and polymorphism

基因多樣性與基因多型性

高雄醫學大學
生物醫學暨環境生物學系
助理教授 張學偉
Email: changhw@kmu.edu.tw

http://genomed.dlearn.kmu.edu.tw/diversity.htm
What is genetic diversity?

http://cnx.org/content/m12158/latest/

- **Genetic diversity** refers to any variation in the nucleotides, genes, chromosomes, or whole genomes of organisms.
**Genetic diversity:** refers to any variation in the nucleotides, genes, chromosomes, or whole genomes of organisms

**Genome:** the entire complement of DNA within the cells or organelles of the organism

**Genotype:** the genetic constitution of an organism that results from the arrangement of the DNA within the cell or organelles.

**Genotypic Variation:** the variation that exists between the genetic constitution of different individuals.

**Phenotype:** the physical constitution of an organism that results from its genetic constitution (genotype), and the action of the environment on the expression of the genes

**Phenotypic Variation:** the variation of the physical traits, or phenotypic characters of the organism, such as differences in anatomical, physiological, biochemical, or behavioral characteristics.
Why is genetic diversity important?

- Genotypes partly determine organisms’ physical form and function (*why partly?*)

- Genetic diversity helps organisms cope with current environmental variability

- Diversity within populations reduces potentially deleterious effects of breeding among close relatives

- Genetic diversity is the primary basis for adaptation to future environmental uncertainty.
Introduction to Genetics
1. Genetic diversity within cells- nuclei & mitochondria
2. Phenotypic diversity within cells

Measurements of Diversity
1. Individual and Population Genetics

Agriculture, Forensics, Field Research and Genetic Diversity
1. Genetic Diversity in Agriculture
2. Genetic Diversity in Forensics and Field Research
3. Examples of lacking genetic diversity

Diseases/Cancers and Polymorphisms
Introduction to Genetics

Part I – Genetic diversity within cells

Genetic diversity within cells - nuclei

Genetic diversity within cells - mitochondria

Note: Cell Differentiation: phenotypic diversity

Environment & gene regulation & phenotype
## Genetic diversity within cells-nuclei

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>TTT Phenylalanine</td>
<td>TCT Serine</td>
<td>TAT Tyrosine</td>
<td>TGT Cysteine</td>
</tr>
<tr>
<td></td>
<td>TTC Phenylalanine</td>
<td>TCC Serine</td>
<td>TAC Tyrosine</td>
<td>TGC Cysteine</td>
</tr>
<tr>
<td></td>
<td>TTA Leucine</td>
<td>TCA Serine</td>
<td>TAA Stop</td>
<td>TGA Stop</td>
</tr>
<tr>
<td></td>
<td>TTG Leucine</td>
<td>TCG Serine</td>
<td>TAG Stop</td>
<td>TGG Tryptophan</td>
</tr>
<tr>
<td>C</td>
<td>CTT Leucine</td>
<td>CCT Proline</td>
<td>CAT Histidine</td>
<td>CGT Arginine</td>
</tr>
<tr>
<td></td>
<td>CTC Leucine</td>
<td>CCC Proline</td>
<td>CAC Histidine</td>
<td>CGC Arginine</td>
</tr>
<tr>
<td></td>
<td>CTA Leucine</td>
<td>CCA Proline</td>
<td>CAA Glutamine</td>
<td>CGA Arginine</td>
</tr>
<tr>
<td></td>
<td>CTG Leucine</td>
<td>CCG Proline</td>
<td>CAG Glutamine</td>
<td>CGG Arginine</td>
</tr>
<tr>
<td>A</td>
<td>ATT Isoleucine</td>
<td>ACT Threonine</td>
<td>AAT Asparagine</td>
<td>AGT Serineine</td>
</tr>
<tr>
<td></td>
<td>ATC Isoleucine</td>
<td>ACC Threonine</td>
<td>AAC Asparagine</td>
<td>AGC Serineine</td>
</tr>
<tr>
<td></td>
<td>ATA Isoleucine</td>
<td>ACA Threonine</td>
<td>AAA Lysine</td>
<td>AGA Arginine</td>
</tr>
<tr>
<td></td>
<td>ATG Methionine</td>
<td>ACG Threonine</td>
<td>AAG Lysine</td>
<td>AGG Arginine</td>
</tr>
<tr>
<td>G</td>
<td>GTT Valine</td>
<td>GCT Alanine</td>
<td>GAT Aspartate</td>
<td>GGT Glycine</td>
</tr>
<tr>
<td></td>
<td>GTC Valine</td>
<td>GCC Alanine</td>
<td>GAC Aspartate</td>
<td>GGC Glycine</td>
</tr>
<tr>
<td></td>
<td>GTA Valine</td>
<td>GCA Alanine</td>
<td>GAA Glutamate</td>
<td>GGA Glycine</td>
</tr>
<tr>
<td></td>
<td>GTG Valine</td>
<td>GCG Alanine</td>
<td>GAG Glutamate</td>
<td>GGG Glycine</td>
</tr>
</tbody>
</table>

*The Genetic Code: The universal matching of codon to amino acids, used by all organisms. The first codon (TTT- phenylalanine) was identified by Nirenberg and Matthaei in 1961.*
Genetic diversity within cells-mitochondria

<table>
<thead>
<tr>
<th>Phe</th>
<th>UUU UUC</th>
<th>Thr</th>
<th>ACU ACC ACA ACG</th>
<th>Asp</th>
<th>GAU GAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu (1)</td>
<td>UUA UUG</td>
<td>Ala</td>
<td>GCU GCC GCA GCG</td>
<td>Glu</td>
<td>GAA GAG</td>
</tr>
<tr>
<td>Leu (2)</td>
<td>CUU CUC CU A CUG</td>
<td>Tyr</td>
<td>UAU UAC</td>
<td>Cys</td>
<td>UGU UGC</td>
</tr>
<tr>
<td>Ile</td>
<td>AUU AUC</td>
<td>Ter</td>
<td>UAA UAG</td>
<td>Trp</td>
<td>UGA UGG</td>
</tr>
<tr>
<td>Met</td>
<td>AUA AUG</td>
<td>His</td>
<td>CAU CAC</td>
<td>Arg</td>
<td>CGU CGC CGA CGG</td>
</tr>
<tr>
<td>Val</td>
<td>GUU GUC GUA GUG</td>
<td>Gin</td>
<td>CAA CAG</td>
<td>Ser (2)</td>
<td>AGU AGC</td>
</tr>
<tr>
<td>Ser (1)</td>
<td>UCU UCC UCA UCG</td>
<td>Asn</td>
<td>AAU AAC</td>
<td>Ter</td>
<td>AGA AGG</td>
</tr>
<tr>
<td>Pro</td>
<td>CCU CCC CCA CCG</td>
<td>Lys</td>
<td>AAA AAG</td>
<td>Gly</td>
<td>GGU GGC GGA GGG</td>
</tr>
</tbody>
</table>

Note that, unlike the universal code, **UGA** codes for tryptophan instead of termination, **AUA** codes for methionine instead of isoleucine, and **AGA** and **AGG** are terminators instead of coding for arginine.

* **AUU** codes for isoleucine during elongation but can code for methionine for initiation ([ND2](#)) See Fearnley & Walker (1987) and Peabody (1989).
OGRe: a relational database for comparative analysis of mitochondrial genomes

Welcome to the Organellar Genome Retrieval system at McMaster University.

OGRe is a searchable relational database which currently contains the complete mitochondrial genome sequences of 1015 metazoan organisms.

In addition to gene sequences, OGRe also contains information on gene order and codon usage.

To get started with OGRe, see the help page.
選擇不同的codon usage table

getorf程式是尋找基因的工具，主要功能在尋找序列中可能的Open Reading Frame(ORF, 開放讀架)位置。

使用方法：

1. 於GoTo處填入getorf或是由選單NUCLEIC==>GENE FINDING中選取getorf程式。
2. 在"input section"中填入欲尋找的序列。
3. 按下"Advanced Options"可得一選單，其中可設定參數舉例如下：

   Code to use：可選擇不同的codon usage table，包含有：
   1. Standard
   2. Standard (with alternative initiation codons)
   3. Vertebrate Mitochondrial
   4. Yeast Mitochondrial
   5. Mold, Protozoan, Coelenterate Mitochondrial and Mycoplasma/Spiroplasma
   6. Invertebrate Mitochonrial
   7. Ciliate Macronuclear and Dasyycladacean
   8. Echinoderm Mitochonrial
   9. Euplotid Nuclear
   10. Bacterial
   11. Alternative Yeast Nuclear
   12. Ascidian Mitochonrial
   13. Flatworm Mitochonrial
   14. Blepharisma Macronuclear
   15. Chlorophycean Mitochonrial
   16. Trematode Mitochonrial
   17. Scenedesmus obliquus
   18. Thraustochytrium Mitochonrial
MITOMAP  http://www.mitomap.org/
A human mitochondrial genome database
A compendium of polymorphisms and mutations of the human mitochondrial DNA

MtDNA Polymorphisms:
Control Region (DLoop) Polymorphisms
Coding Region (non-DLoop) Polymorphisms
Collection of Unpublished Polymorphisms
Animal mitochondrial DNA (mtDNA) is a small (15-20 kb) circular molecule, composed of about 37 genes coding for 22 tRNAs, two rRNAs and 13 mRNAs, the latter coding for proteins mainly involved in the electron transport and oxidative phosphorylation of the mitochondria.

The mitochondrial genome is arranged very efficiently. It lacks introns, has small intergenic spacers where the reading frames even sometimes overlap.

The control region is the primary non-coding region, and is responsible for the regulation of heavy (H) and light (L) strand transcription and of H-strand replication.
Mitochondrial genomes

phenotypic diversity-
e.g., Cell Differentiation

Multi-cellular organisms develop from a single cell into a variety of cell types by regulating gene expression - on and off the transcription and translation.
Gene regulation

→ guides differentiation (respond to *internal* environment).

→ allows an organism to respond to a changing *external* environment (→ mutation → genetic diversity?)
Environment & gene regulation & phenotype

*Exogenous* expression represses the *endogenous* expression.

e.g., sex gland: methyl-testosterone & testosterone
→ Loss reproduction?
DNA Fingerprinting*

• *DNA Fingerprinting* is actually a patented process, but the term has been adopted to describe the analysis of repetitive sequences.

→ performed *either by* probing DNA with markers that contain the repetitive sequences,

→ *or by* using PCR to amplify specific repeat regions within the genome.
DNA Fingerprinting*

Two types of DNA fingerprinting techniques are described below:

1. Variable Number Tandem Repeat (VNTR) analysis. VNTR = minisatellite marker

2. PCR-based methods of fingerprinting.
   e.g., randomly amplified polymorphic DNA (RAPD), arbitrarily primed PCR (AP-PCR), amplified fragment of length polymorphism (AFLP), and restriction enzyme of length polymorphism (RFLP).
中草藥藥材鑑定和品質管制


- 為中草藥新藥開發中最基礎關鍵的一環。
- 利用DNA的指紋圖譜的品種鑑別技術已相當成熟，較常用的技術有：
  1. 限制切割片段長度多型性（RFLP）
  2. 隨機核酸增幅多型性（RAPD）
  3. 增殖片段長度多型性（AFLP）

- 其中RFLP需人力及時間較多，多用於遺傳分析上，而RAPD與AFLP可以較短時間內得到結果，已用於品種鑑定的實例包括：蘋果、黑莓、大麥、大豆、棉花、馬鈴薯、燕麥、水稻、番茄等多種植物。
## Technologies

Numbers of ProbeDB entries by probe technology.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illumina Bead Array (genotyping)</td>
<td>2472542</td>
</tr>
<tr>
<td>Microarray (genotyping)</td>
<td>4796642</td>
</tr>
<tr>
<td>Resequencing (SNP discovery)</td>
<td>429931</td>
</tr>
<tr>
<td>RNAi (gene silencing)</td>
<td>274357</td>
</tr>
<tr>
<td>Morpholinos (gene silencing)</td>
<td>0</td>
</tr>
<tr>
<td>in situ hybridization (gene expression)</td>
<td>3763</td>
</tr>
</tbody>
</table>
• **Random Amplified Polymorphic DNA (RAPD)** markers are DNA fragments from **PCR** amplification of random segments of genomic DNA with *single primer* of *arbitrary nucleotide sequence*.

• **Restriction Fragment Length Polymorphism (RFLP)** is a difference in homologous DNA sequences that can be detected by the presence of fragments of different lengths *after digestion* of the DNA samples in question with specific **restriction endonucleases**.

• **Amplified Fragment Length Polymorphisms (AFLPs)** are differences in restriction fragment lengths *caused by SNPs* or **INDELs** that *create or abolish* **restriction endonuclease** recognition sites.

  The AFLP technique is based on the selective **PCR** amplification of restriction fragments *from a total digest of genomic DNA*. 
SNP-RFLPing analysis

The restriction fragment length polymorphism (RFLP) is a common laboratory method for the genotyping of single nucleotide polymorphisms (SNPs). Here we describe a web-based software SNP-RFLPing, provides the restriction enzyme for RFLP assays on a batch if SNPs and genes from the human, rat, and mouse genomes.

In the system, three user-friendly inputs are included: 1) NCBI dbSNP "rs" or "ss" IDs; 2) NCBI Entrez gene ID and HUGO gene name; 3) any formats of SNP-in-sequence, and are allowed to perform the SNP-RFLPing assay. These inputs are auto-programmed to SNP containing sequences and their complementary sequences for the selection of restriction enzyme. All RFLP available SNP of each input genes are provided even if many SNPs exist. The SNP-RFLPing analysis provides the SNP contig position, heterozygosity, function, protein residue, and amino acid position for cSNPs as well as commercial and noncommercial restriction enzymes.

It is time-saving and user-friendly to use the SNP-RFLPing for association studies in personalized medicine.
A. The components of the nuclear human genome
<table>
<thead>
<tr>
<th>Repeat name</th>
<th>Repeat size</th>
<th>Total size</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Satellite</td>
<td>5–200 bp</td>
<td>Up to several million bp</td>
<td>Found in heterochromatin and centromeres, Not transcribed</td>
</tr>
<tr>
<td>2 Minisatellite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Hypervariable family</td>
<td>10–60 bp</td>
<td>1000–20 000 bp</td>
<td>Share a common core sequence (motif), GGGCAGGG (where N is any base), dispersed, VNTRs</td>
</tr>
<tr>
<td>(b) Telomeric family</td>
<td>6 bp</td>
<td>1000–20 000 bp</td>
<td>Usually TTAGGG and repeated about a thousand times, Protects chromosome ends</td>
</tr>
<tr>
<td>3 Microsatellite</td>
<td>1–4 bp</td>
<td>Less than 1000 bp</td>
<td>Repeats of A and CA are the most common, Dispersed throughout genome</td>
</tr>
</tbody>
</table>

bp, base pairs; VNTRs, variable numbers of tandem repeats.
Genetic diversity within a Population: Polymorphism

- A working definition for polymorphism:
  - the less common allele has a population frequency >1%, (>5% is an alternative, equally arbitrary cutoff).

- Originally defined by E. B. Ford (a famous ecological geneticist) in the 1940s as the co-occurrence in the same locality, of two or more discontinuous forms of a species, in such proportions that the rarest of them cannot be maintained by recurrent mutation.
Polymorphisms are mutations that have become common

• Many mutations affect *pigmentation* in humans,
  – *albino* mouse (OCA1 albinism in humans),
  – *pink-eyed dilute* (OCA2 albinism in humans),
  – *brown* (OCA3 albinism).
  – *extension* (MC1R, red hair in humans)

• *Albinism* remains at low frequency in human populations, 
  and most cases are due to recurrent mutations in genes 
  in pigmentation pathway.

• *Red hair* in humans has become a common polymorphism in some human populations. Most cases 
  due to inheritance of MC1R loss-of-function alleles from an ancestor.
Genetic mapping & resolution

• RFLP (coverage 1/10 cM)  
  → map single gene disorder.

• Microsatellite markers (1/10 kb)  
  → map more complex diseases

• SNP (1/1~1.5 kb)  
  → the third generation map &  
    map functional variation in gene
<table>
<thead>
<tr>
<th>Class</th>
<th>Size of Locus</th>
<th>Number of Alleles</th>
<th>Number of Loci in Population</th>
<th>Rate of Mutation</th>
<th>Use</th>
<th>Method of Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>Single base pair</td>
<td>2</td>
<td>100 million</td>
<td>$10^{-9}$</td>
<td>Linkage mapping</td>
<td>PCR followed by ASO hybridization or primer extension</td>
</tr>
<tr>
<td>Microsatellite</td>
<td>30–300 bp</td>
<td>2–10</td>
<td>200,000</td>
<td>$10^{-3}$</td>
<td>Linkage mapping</td>
<td>PCR and gel electrophoresis</td>
</tr>
<tr>
<td>Multilocus Minisatellite</td>
<td>1–20 kb</td>
<td>2–10</td>
<td>30,000</td>
<td>$10^{-3}$</td>
<td>DNA fingerprinting</td>
<td>Southern blot and hybridization</td>
</tr>
<tr>
<td>Small Changes in DNA Content (deletions and duplications)</td>
<td>1–100 bp</td>
<td>2</td>
<td>N/A</td>
<td>$&lt;10^{-9}$</td>
<td>Linkage mapping</td>
<td>PCR and gel electrophoresis</td>
</tr>
</tbody>
</table>
Measurements of Diversity

Part I – Individual and Population Genetics
**Part I – Individual and Population Genetics**

*Individual genetics* is the study of the frequency of occurrence of alleles *within and between individuals*.

*Population genetics* is the study of the frequency of occurrence of alleles *within and between populations*.
Individual Variation

- **Sources of Individual Variation** in sexually reproducing organisms:
  - mutation and recombination.

- If it occurs during **meiosis**, the mutation will be heritable.

- If the mutation occurs during **mitosis**, the mutation cannot be inherited by the next generation.
diploid and alleles

Every **diploid** individual has two copies (two alleles) of each gene, one inherited from each parent.

**alleles** - are different versions of the same gene that are expressed as different phenotypes.
Effect of new alleles

New alleles appear in a population by the random and natural process of mutation.

The frequency of occurrence of an allele changes regularly as a result of mutation, genetic drift, and selection.
If an individual has two different versions of a particular gene, the individual is said to be **heterozygous** for that gene.

If the two alleles are the same, the individual is **homozygous**.
allele form in population- adj

• A gene with only one type of allele across a population is **monomorphic** (single form).

• A gene with more than one allelic variant across a population is said to be **polymorphic** (many forms).
Heterozygosity
- The two alleles at a locus are different.
- The proportion of individuals in a population that are heterozygous at a locus.

Homozygosity
- The two alleles at a locus are the same.
- The proportion of individuals in a population that are heterozygous at a locus.
Determining Variability of Individual

[Mendelian genetics]
predict the probability of the appearance (phenotype) of a particular allele in an offspring when the alleles of each parent are known.

punnett square

<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th>Ab</th>
<th>aB</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>AABB</td>
<td>AAbB</td>
<td>aABb</td>
<td>aAbB</td>
</tr>
<tr>
<td>Ab</td>
<td>AABb</td>
<td>AAbb</td>
<td>aABb</td>
<td>aAbb</td>
</tr>
<tr>
<td>aB</td>
<td>AaBB</td>
<td>AaBb</td>
<td>aaBB</td>
<td>aabB</td>
</tr>
<tr>
<td>ab</td>
<td>AaBb</td>
<td>Aabb</td>
<td>aaBb</td>
<td>aabb</td>
</tr>
</tbody>
</table>
The Importance of **Heterozygosity**

- A population needs variation.

- The measure of the amount of heterozygosity across all genes
  → used as a *general indicator* of the amount of genetic variability and genetic health of a population.
observed heterozygosity
• A measure of genetic variation in a population calculated as the mean frequency of heterozygotes over all loci.

expected heterozygosity (ALSO: gene diversity)
• A measure of genetic variability in a population: the mean expected heterozygosity per locus in a population.
Determining Variability of Population- 1

Similar predictions can be made about the frequencies of alleles in the next generation of an entire population.

By comparing the predicted or "expected" frequencies with the actual or "observed" frequencies in a real population, scientists can infer a number of possible external factors that may be influencing the genetic structure of the population (such as inbreeding or selection).
### Determining Variability of Population- 2

<table>
<thead>
<tr>
<th>Observed Number Individuals with each Genotype</th>
<th>Observed Genotype Frequencies</th>
<th>Expected Genotype Frequency</th>
<th>Note that: $p^2 + 2pq + q^2 = 1$; Therefore: Homozygotes = $p^2 + q^2$; Heterozygotes = $2pq$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA = 10</td>
<td>$A = \frac{(2 \times 10) + 10}{2 \times 80} = 0.1875$</td>
<td>$p^2 = 0.0352$</td>
<td></td>
</tr>
<tr>
<td>Aa = 10</td>
<td>$a = 1 - 0.1875 = 0.8125$</td>
<td>$2pq = 0.3046$</td>
<td></td>
</tr>
<tr>
<td>aa = 60</td>
<td></td>
<td>$q^2 = 0.6002$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N (individual)</th>
<th>A allele #</th>
<th>a allele #</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Aa</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>aa</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>total</td>
<td>80</td>
<td>30</td>
</tr>
</tbody>
</table>

- $P + q = 1$;
- $f(A) = p$, $f(a) = q$;

Exp[$f(AA)] = p^2$;
Exp[$f(Aa)] = 2pq$;
Exp[$f(aa)] = q^2$
## Determine genotype and allele frequency

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>17 (25.4%)</td>
<td>2 (6.5%)</td>
</tr>
<tr>
<td>TC</td>
<td>37 (55.2%)</td>
<td>16 (51.6%)</td>
</tr>
<tr>
<td>CC</td>
<td>13 (19.4%)</td>
<td>13 (41.9%)</td>
</tr>
</tbody>
</table>

### Allele frequency

<table>
<thead>
<tr>
<th>Allele</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>71 (53.0%)</td>
<td>20 (32.3%)</td>
</tr>
<tr>
<td>C</td>
<td>63 (47.0%)</td>
<td>42 (67.7%)</td>
</tr>
</tbody>
</table>

1 individual = 2 alleles
Hardy-Weinberg Principle- 1

- Based on *Mendel's principles of inheritance*, two men, G.H. Hardy and Wilhelm Weinberg, independently developed the concept that is known today as the Hardy-Weinberg Principle.
Hardy-Weinberg Principle- 2

• "In a large, randomly breeding (diploid) population, allelic frequencies will remain the same from generation to generation; assuming no unbalanced mutation, gene migration, selection or genetic drift.

• " When a population meets all of the Hardy- Weinberg conditions it is said to be in Hardy-Weinberg equilibrium.
This equilibrium can be mathematically expressed based on simple binomial (for two alleles) or multinomial (multiple allele) distribution of the gene frequencies as above:

\[ p^2 + 2pq + q^2 = 1 \]
Testing for Hardy-Weinberg Equilibrium

O is the observed number of individuals with the specific genotype.

E is the expected number of individuals based on the Hardy-Weinberg Equilibrium (i.e., for AA, the expected number of individuals is: \( P^2 \times N \); \( N \) = individual number)

- \( k \) is the number of genotypes

\[
\chi^2 = \sum_{i=1}^{K} \frac{(O-E)^2}{E}
\]
X² is used to determine the probability that the observed number differs from the expected number due to chance alone.

- Standardized statistical charts have been developed which correlate the X² value and degrees of freedom (the number of independent variables) with probability values (p).
- In this example, the X² value of 27.77 correlates to a greater than 99% chance that the difference between the observed and the expected is NOT due to chance alone.
- This high probability indicates that some external factor (i.e., migration, selection, inbreeding, or drift) is influencing the frequencies of alleles.
**Inbreeding effect** of genotype frequency

- Increase frequency of homozygote genotypes.
- Decrease frequency of heterozygote genotypes.
- **Not** in Hardy-Weinberg Equilibrium.
Part I - Genetic Diversity in Agriculture

Part II - Genetic Diversity in Forensics and Field Research

Part III - Examples of lacking genetic diversity
Introduction to Genetic Diversity

• the diversity (variation) of genes within and among populations of a species. This is the lowest level of biological diversity.

• refers to the variation at the level of individual genes (polymorphism), and provides a mechanism for populations to adapt to their ever-changing environment.
Inter-dependence between biological and genetic diversity

Changes in biodiversity result in changes in the environment, requiring subsequent adaptation of the remaining species.

Changes in genetic diversity, particularly loss of diversity through loss of species, results in a loss of biological diversity.
Domestication

farmers learned to *selectively breed* individuals for desirable traits (or to avoid breeding for undesirable traits).

The process of **domestication** resulted in breeds and varieties that are largely dependent on human inputs for their survival.
Monocultures

modern agriculture evolved into the practice of raising monocultures of crops and livestock, in which most of the gene variants (alleles) are the same in every individual of a particular variety or breed.

Present-day monocultures are highly productive, but their reduced genetic variability leaves them with a diminished capacity to deal with new diseases, pests, and other changes in environmental conditions.
In situ Conservation of Agricultural Genetic Resources

• Genetic diversity is preserved through a variety of in situ (in position or in-field) agricultural practices described above.

• In addition, there are a number of organizations that enlist teams of local farmers to grow native varieties, particularly those that are threatened by extinction due to lack of modern-day use.
Ex situ Conservation of Agricultural Genetic Resources

- There are also local, national and international efforts to preserve agricultural genetic resources through *ex situ* (off-site) methods such as seed and sperm banks.
Genetic diversity for Forensics and Field Research

Forensic Evidence in Genetic Variation:
   polymorphism and 911 accidents.

Application of DNA Forensics:
   Paternity
   Trade behavior
      - forensics laboratories for wildlife genetics issues.
      - 鮑魚包子?
Human Genetic Diversity (ethics view)-e.g., SNP
Illegal Trade of *Endangered* Species

The illegal trade in animals and animal parts (both domestic and international) is one of the primary causes driving over-exploitation of threatened and endangered species, including:

- **Food**: e.g., caviar from Caspian Sea sturgeon, freshwater and marine turtles; 鮑魚包子
- **Traditional medicine**: e.g., rhinoceros horns, bear gallbladders, and various plants;
- **Pets**: many species of exotic amphibians, reptiles and birds;
- **Timber**: rainforest hardwood trees such as mahogany and teak;
- **Animal furs and skins**: e.g., the trade in crocodile and alligator skins;
- **Tourism**: products sold as souvenirs such as figurines made from illegal ivory or marine turtle shells, or jewelry made of coral.
DNA analysis of species
hair, skin, tissue, blood, ascites, saliva, urine, stool.

1. 数位化基因多型性分析
数位化基因多型性分析可以分别计算出父亲或母亲对偶基因的基因多型性型态，即使有正常细胞的操
纵污染，仍然可以定量地分析对偶基因不平衡。这个新的技术，使用基因多型性的分子聚合酶链来诊
断，目的在侦测早期癌症发生。目前可成功使用检体包括血脅、脤水、组织、石癌红斑、口腔黏
膜、尿液、唾液、头发样本。

2. 非侵入性生物检体DNA分析
可成功使用尿液或唾液、毛髮或粪便…等检体萃取DNA，作为各种基因检测开发或疾病、癌症侦测。

3. 基因選擇性分析找出候選腫瘤基因
基因選擇性分析可以找出候選腫瘤或癌因基因，并用同步化定量RT-PCR加以快速篩選。目前本實驗
室進行有關卵巢癌与口腔癌的研究，以及腫瘤候選基因的分子選篩與蛋白質表現。

4. 基因多型性基因定序、單核型分析與其相關性研究
舉凡頭髮毛髮、血液、口腔黏膜、尿液或唾液…等DNA，均可作基因多型性的基因定序分析。選取
適當的標記，即可分析基因多型性與其致病的危險因子相關性。單核型SNP相關性研究是主要方向。
目前方向包括骨質疏鬆症與綿羊病、口腔癌。

5. 分子性别鑑定与演化分析
目前成功利用鸟类(如鸟类-白尾鵲、大冠鵑、黑背信天翁)、台湾鲑鱼与尼加的糞便、血液、組織或
羽毛，可以做分子性别鑑定、演化或同源鑑定分析。

6. 生物資訊軟體研發
與電腦軟體專家合作，研發與更新生物資訊相關軟體與視覺化。目前已有成熟的SNP相關拷貝軟體研
發。如: SNP-RFLPing, V-MitoSNP, SNP ID-info, Seq-SNPing, Primer design。
Field Research

• DNA analysis has been used in concert with behavioral data to confirm, substantiate or dispute conclusions based on observations of wild populations.

• 黑嘴端鳳頭燕鷗&白眉燕鷗 的故事
生態環境観察與紀錄

→

燕鷗検体(粪便、羽毛)取得

→

DNA 萃取

→

CHD-PCR

→

性別鑑定

(雌雄比例監測)

→

微衛星DNA 基因座分析

→

檢驗検體之個別重複率

• 評估族群多樣性

→

統線體 DNA 定序

→

物種辨別
Molecular Sexing of Bird
- feather, tissue, stool

- 圖1-A
- 圖1-B
- 圖1-C

- 圖A與圖B：為雄性金斑鶯的羽毛所萃取出DNA，由 primer所做出的結果。
- 圖C：由左至右為雌性紅尾伯勞、雄性山鵯及雌性紅嘴黑鶥，從這三種鳥類組織所萃取出DNA所做出的結果。
白眉燕鸥（编号88号）的CHD-W与CHD-Z序列概图。
Part III. Examples of lacking genetic diversity

Why Does The Cheetah Lack Genetic Diversity?
The cheetah, *Acinonyx jubatus*, is the sole member of its genus. Twenty thousand years ago, cheetahs roamed throughout the savannas and plains of four continents: Africa, Asia, Europe, and North America.

About 10,000 years ago - because of climate changes - all but one species of the cheetah, *jubatus*, became extinct. With the drastic reduction in their numbers, close relatives were forced to breed, and the cheetah became genetically inbred, meaning all cheetahs are closely related. Inbreeding occurs when members of the same family or close relatives breed only among themselves. For example, when you look around, you see different hair colors, eye colors, and heights. If you took blood from everybody in the room, and looked at the proteins in the blood, you would see proteins also vary between each person, just like hair color. When you look at the proteins in the blood of cheetahs, they are very similar; it looks as if they are identical twins of one another, meaning they are closely related.

The study of biological inheritance is called "genetic research." Genes, which are composed of DNA, store the information that an individual inherits from his or her parents. Genes in one animal vary from the same genes in another animal of the same species. By looking at the amount of variation existing in genes, scientists, called "geneticists" can begin to understand the relationships of animals within population, and how infectious diseases may affect that population. Also, by comparing the amount of variation between different species, geneticists can help us understand the evolutionary process.

When geneticists looked at the amount of variation within the genes of the cheetah, they found that cheetahs exhibit much lower levels of variation than other mammals. In most species, related individuals share about 80 percent of the same genes. With cheetahs, this figure goes to approximately 20 percent. The genetic inbreeding in cheetahs has led to lower variation.
樱花钩吻鲑基因均质化的危机

曾晴賢、王昱人
新竹清华大学生命科学系

摘要
本研究除了利用同功異構酶電泳法分析臺灣樱花鈎吻鲑的族群遺傳結構，另外，使用 DNA 定序的方法分析台灣樱花鈎吻鲑的遺傳多樣性。

在同功異構酶電泳分析中，23 個基因座只有一個有遺傳變異，平均観測(Ho)與平均理論質異度(He)為 0 與 0.0115，表示其基因多樣性偏低。利用 F-統計分析結果，族群之近交係數(FIS)為 1，顯示台灣樱花鈎吻鲑傾向於近親交配，族群間變異指數(FST)為 0.081，表示族群幾乎沒有分化。

在粒線體 DNA 序列分析方面，共定序了 12 尾台灣樱花鈎吻鲑粒線體 DNA 的控制區(D-Loop)。結果發現，12 尾台灣樱花鈎吻鲑只分成兩個基因型，而且兩種基因型中只有一個鹼基對的變異，遺傳距離僅有 0.001，顯示台灣樱花鈎吻鲑遺傳多樣性已經相當貧乏。這種基因多樣性貧乏的原因和樱花鈎吻鲑族群數量稀少，現有棲息的七家灣溪過多敗塲切割棲息地導致魚群無交流機會，同時因為大部分河域水溫過高而大多數族群無法繁殖成功所導致。
（二）、實驗方法：

1、同功異構酶電泳分析：
（1）、酵素萃取。
（2）、電泳膠的備製及電泳實驗。
（3）、切膠及酵素染色。
（4）、資料分析。

2、粒線體 DNA 序列分析：
（1）DNA 萃取
（2）DNA 引子之設計
（3）聚合酶連鎖反應
（4）DNA 定序
（5）DNA 序列分析
再論台灣鮭魚身世之謎

周以正1、鍾郁涵2、張學偉2、廖林彥3、林永發3、郭金泉4

1中華醫事學院; 2. 高雄醫學大學; 3. 雪霸國家公園管理處; 4. 台灣國立海洋大學

欖鰤家族（Onchorhynchus masu complex）僅分佈於西太平洋環日本海地區，包括日本全島、庫頁島、斯堪地半島南部、東西伯利亞與韓國東部的溪流與沿海。台灣則是唯一分佈該種魚類的亞熱帶國家。根據外部型態分類，台灣欖鰤（Onchorhynchus masou formosanus （Jordan et Oshima））與欖鰤（Onchorhynchus masou masou；masu salmon，其陸封型稱為山女 Yamame）、石川欖鰤（Onchorhynchus masou ishikawae，其陸封型稱為雨子 Amago）及琵琶欖（Biwa salmon；Onchorhynchus masou subsp. 屬陸封型）是同屬於俗名「欖鰤」(cherry salmon; Sakura salmon)的欖鰤家族的四個亞種。一般而言，台灣欖鰤與日本幾個欖鰤家族親戚中，與欖鰤的陸封型山女魚在體型與外觀上最为接近，但在一些細部特徵上仍然有所不同。例如：台灣欖鰤的鳍條數較少、體高較高、體側小黑點較少，而且腹面有明顯的橢圓斑紋。另外，台灣欖鰤魚頭部前端附近的形狀與日本欖鰤稍有差異。然而由於外表的差異甚難分辨不同亞種的欖鰤，因此我們研究以分子標記的技術來分辨三個亞種（台灣欖鰤、陸封型欖鰤 Yamame、與陸封型石川欖鰤Amago），同時推論其等之親緣性。
我們發現：

1. 台灣鮭魚與虹鱒（rainbow trout；Onchorhynchus mykiss）的 GH 1 基因之 intron 3 序列有極大的差異，此外虹鱒的背鰭和尾鰭有許多黑點，這特徵也可以用來區別虹鱒和櫻鰤。而且兩者魚卵的孵化積算溫度不同，在自然狀況下一般兩魚種並不會配對交配。況且即使交配，其雜交種亦無法存活。虹鱒的染色體數目（58～60）和櫻鰤的染色體數目（66）相差 6 到 8 條染色體。

2. 台灣鮭魚（雌雄兩性）沒有如一些櫻鰤家族成員（櫻鰤，琵琶鮭）一樣有 pseudogene—GH-ϕ 的存在。我們無法以pseudogene—GH-ϕ 的有無來判斷台灣鮭魚的性別。

3. 台灣鮭魚在 GH 1 之 intron 3 （或 intron c）的部分有一段 GCT 三個核酸鹼基對的缺失（deletion），而其他櫻鰤家族的成員（日本櫻鰤與石川氏鮭魚）基因上均無此現象。台灣鮭魚是歷經數十萬到數百萬年演化後落地生根於台灣的物種，不可能是 100 年前日本人殖民佔領臺灣時從日本拿日本櫻鰤來台灣放流的。

4. 台灣鮭魚和石川鮭魚陸封型的 Amago皆無 pseudogene—GH-ϕ 的存在。加上 GH1 基因之 intron 3 的分子演化學證據，我們推論在櫻鰤家族的各亞種中，台灣鮭魚和陸封型的石川鮭魚（amago）的親緣關係最近。這個論點應該比僅根據外觀型態推論「台灣鮭魚與日本幾個鮭魚親戚中，與櫻鰤的陸封型山女魚體型與外觀上最為接近」的說法更有說服力。

5. 調查櫻鰤家族（Onchorhynchus masu complex）中的三亞種：台灣鮭魚、櫻鰤和石川鮭魚粒腺體 DNA 基因體的 16652 配對，再畫出親緣樹狀圖。我們推論約 20 萬年前台灣鮭魚由櫻鰤家族的祖先分支出獨立演化，比日本櫻鰤成員早。
Diseases/Cancers and Polymorphisms

NCBI PubMed

Search for polymorphism cancer

History has expired.
Display Summary

   Associations between Androgen and Vitamin D Receptor Microsatellites and Postmenopausal Breast Cancer.
   Cancer Epidemiol Biomarkers Prev. 2007 Sep;16(9):1773-83. PMID: 17835696 [PubMed - in process]

2. Pande M, Chen J, Amos CI, Lynch PM, Broadaus R, Frazier ML.
   Influence of Methylenetetrahydrofolate Reductase Gene Polymorphisms C677T and A1298C on Age-Associated Risk for
   Colorectal Cancer in a Caucasian Lynch Syndrome Population.

   GLIOGENE an International Consortium to Understand Familial Glioma.
   Cancer Epidemiol Biomarkers Prev. 2007 Sep;16(9):1730-4. PMID: 17855690 [PubMed - in process]

   The CYP3A4(*)1B polymorphism and prostate cancer susceptibility in a Portuguese population.
GLIOGENE, an International Consortium to Understand Familial Glioma.
Cancer Epidemiol Biomarkers Prev. 2007 Sep;16(9):1730-4. PMID: 17855690 [PubMed - in process]

Conformation of MHC class II I-Ag7 is sensitive to the P9 anchor amino acid in bound peptide.

Lack of association of the 463 G/A myeloperoxidase promoter polymorphism with Behcet's disease in Italian patients.

Activation of polymorphonuclear neutrophils in patients with impaired left ventricular function.
Effect of Genetic diversity

Variation among individuals for some heritable trait.

Intra-specific genetic diversity:
1. genetic diversity within populations.
2. genetic diversity among populations.

Variations in their chromosomes makes the community in general more resistant to diseases or to changing ecological conditions, sometimes.

The heritable variation within and among populations which is created, enhanced or maintained by evolutionary forces.
Genetic polymorphism

• State in which individuals in a population have more than one genotype at a locus.

• One of several forms of a genetic characteristic at a locus in a population.

• The presence in a population of two or more relatively common forms of a gene, chromosome, or genetically determined trait.

• Difference in DNA sequence among individuals. Applied to many situations ranging from genetic traits or disorders in a population to the variation in the sequence of DNA or proteins.

• The occurrence of two or more alleles at a locus in a population.

• **Difference in DNA sequence among individuals that occurs in 1% or more of a population.**
A SNP is a *single-base variation* that occurs about every 1,000 bases along the three billion base pairs of the human genome.

The *most common SNP* is a change from cytosine to thymine (C — T) on one strand of DNA, with a change from guanine to adenine (G — A) on the complementary strand.