

LA MIGRAZIONE E I POLIMORFISMI GENETICI IN GENETICA UMANA

LA MIGRAZIONE

Unione casuale	Inincrocio
Popolazione grande	Deriva
Mutazione trascurabile	Mutazione
Migrazione trascurabile	MIGRAZIONE
Mortalità indipendente dal genotipo	Selezione
Fertilità indipendente dal genotipo	Selezione

Se non vengono soddisfatte le condizioni dell' Equilibrio di Hardy-Weinberg:

*** What is migration?**
What types of migration exist?

Legal
illegal
irregular migrants
refugees
labour migration

* Migration - is the crossing of the boundary of a political or administrative unit for a certain minimum period of time. It includes the movement of refugees, displaced persons, uprooted people as well as economic migrants.

* Types:

Internal migration

International migration

- *Today 192 million people live outside their place of birth - it is about 3% of the world's population;
- *1 of every 35 persons in the world is a migrant;
- *Current annual growth rate of international migrants is about 2,9%;

- * Japan and all countries of Europe are expected to face declining population growth over the next 50 years.
- * Population of Italy in 2050 will decline from 57 to 41 million of people
- * Population of Japan in 2080 will decline from 127 to 105 million

- * Legal migrants
- * Illegal migrants
- * Irregular migrants
- * Refugees
- * Labour migration

* Types of migration

* According to the 1951 United Nations Convention Relating to the Status of Refugees, a refugee is a person who owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion, is outside the country of their nationality, and is unable to or, owing to such fear, is unwilling to avail him/herself of the protection of that country.

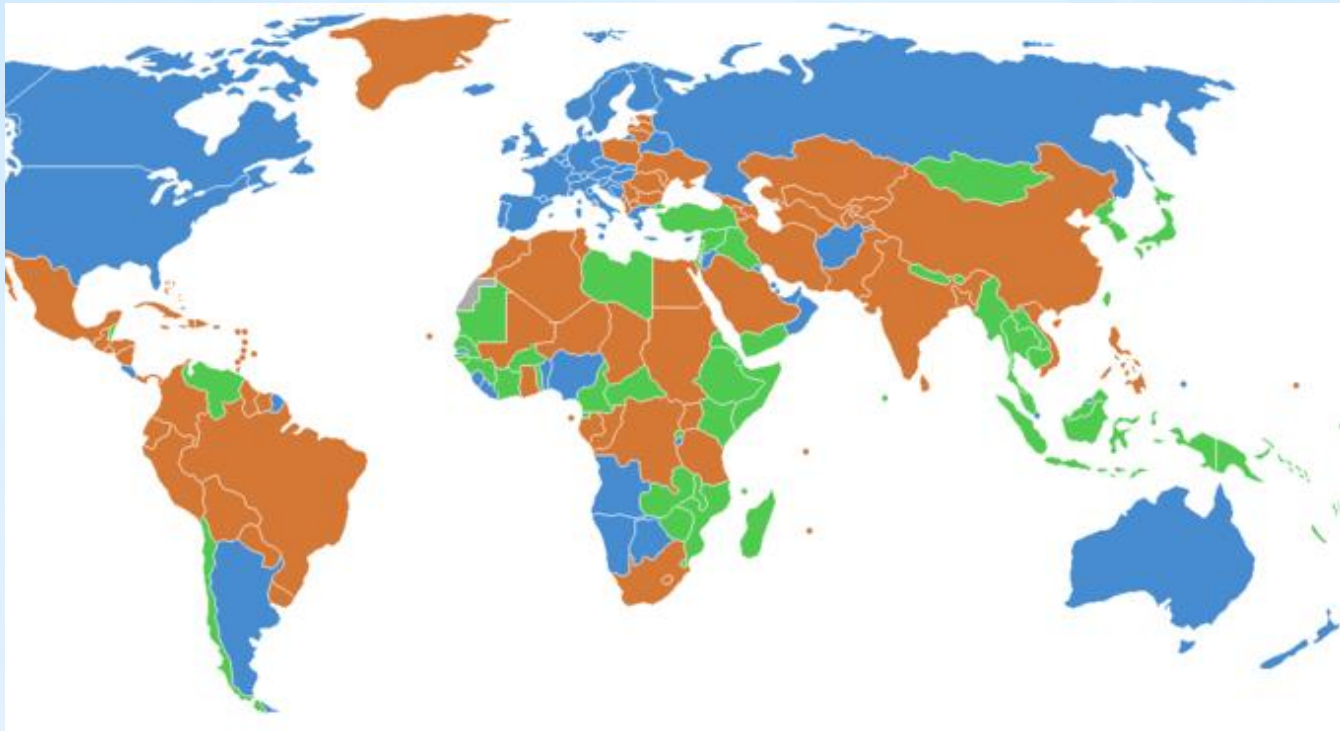
* **Refugees**

** An international migrant worker is defined by the 1990 United Nations (UN) International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families as “a person who is to be engaged, is engaged or has been engaged in remunerated activity in a State of which he or she is not a national.*

** Labour Migration*

- * **Temporary labour migrants** (also known as **guest workers** or **overseas contract workers**): People who migrate for a limited period of time in order to take up employment and send money home.
- * **Highly skilled and business migrants**: People with qualifications as managers, executives, professionals, technicians or similar, who move within the internal labour markets of trans-national corporations and international organizations, or who seek employment through international labour markets for scarce skills. Many countries welcome such migrants and have special 'skilled and business migration' programs to encourage them to come.

*Types of Labour Migration



* Net migration rate

Net migration rate is the difference of immigrants and emigrants of an area in a period of time.

MIGRAZIONE: GENETICA

Flusso di individui da una popolazione a un'altra (mescolamento di due popolazioni).

L'effetto della migrazione sulle frequenze geniche può essere studiato attraverso due modelli:

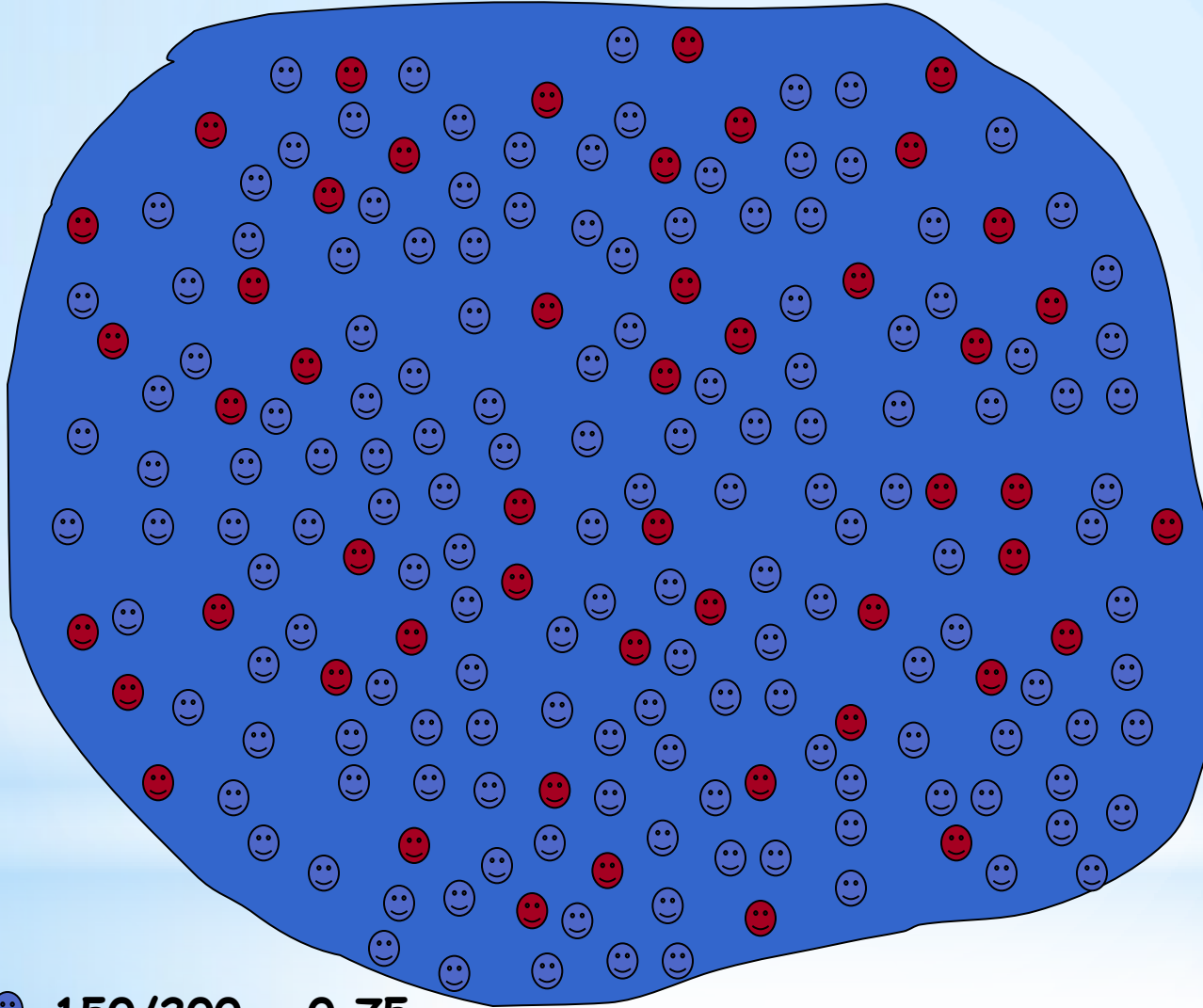
- migrazione regolare da una popolazione molto grande (freq. alleliche considerate stabili) verso una più piccola e viceversa:

• MODELLO DELL' ISOLA

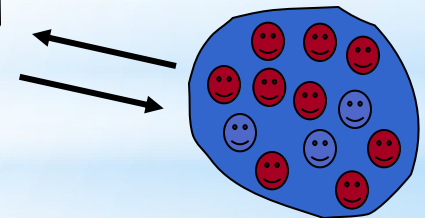
- un gruppo di popolazioni, parzialmente isolate, che si scambiano tra loro individui:

• MODELLO DELL' ARCIPELAGO

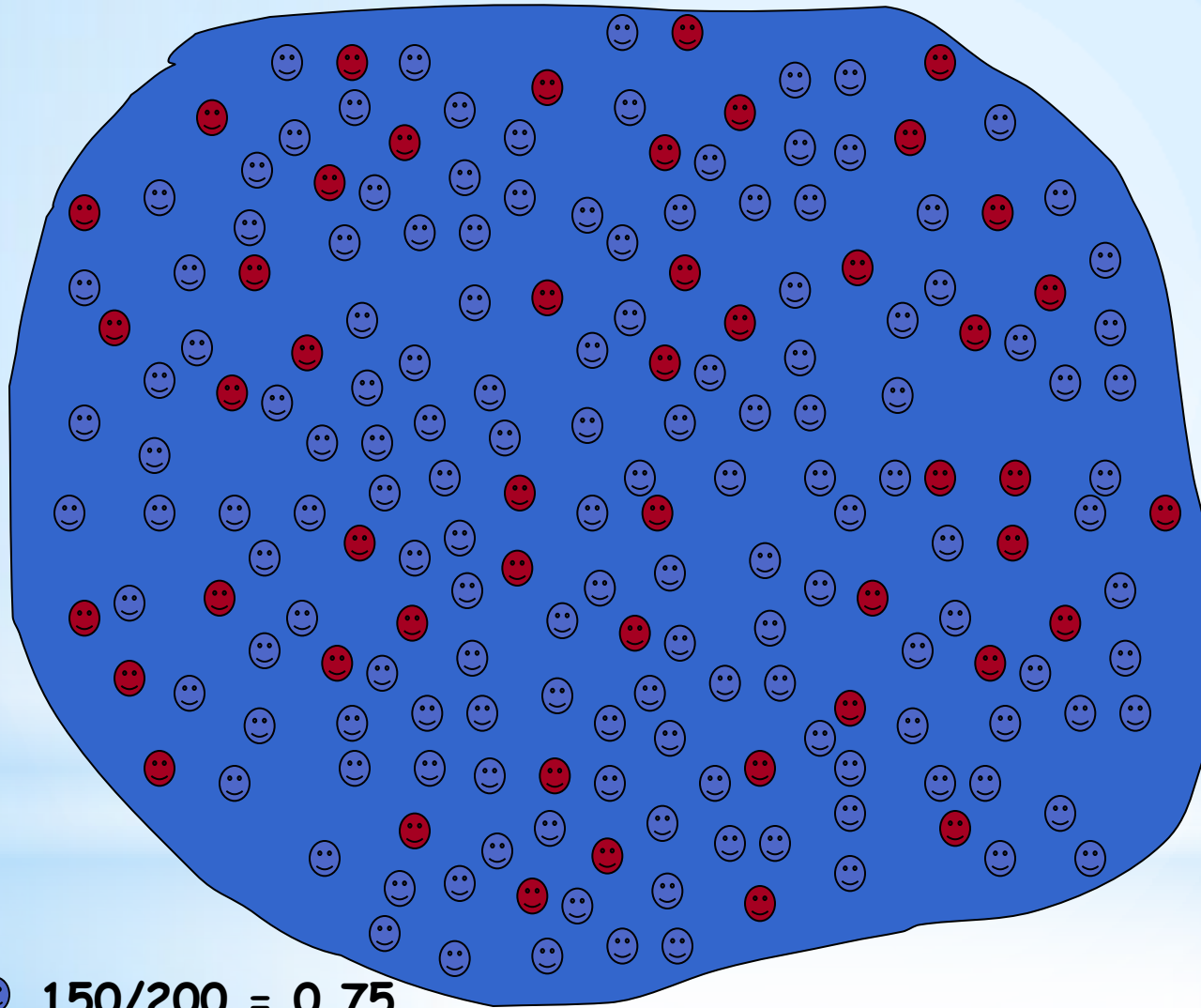
MODELLO DELL' ISOLA



● $150/200 = 0.75$
● $50/200 = 0.25$

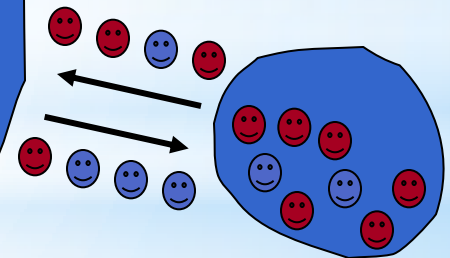


● $3/12 = 0.25$
● $9/12 = 0.75$

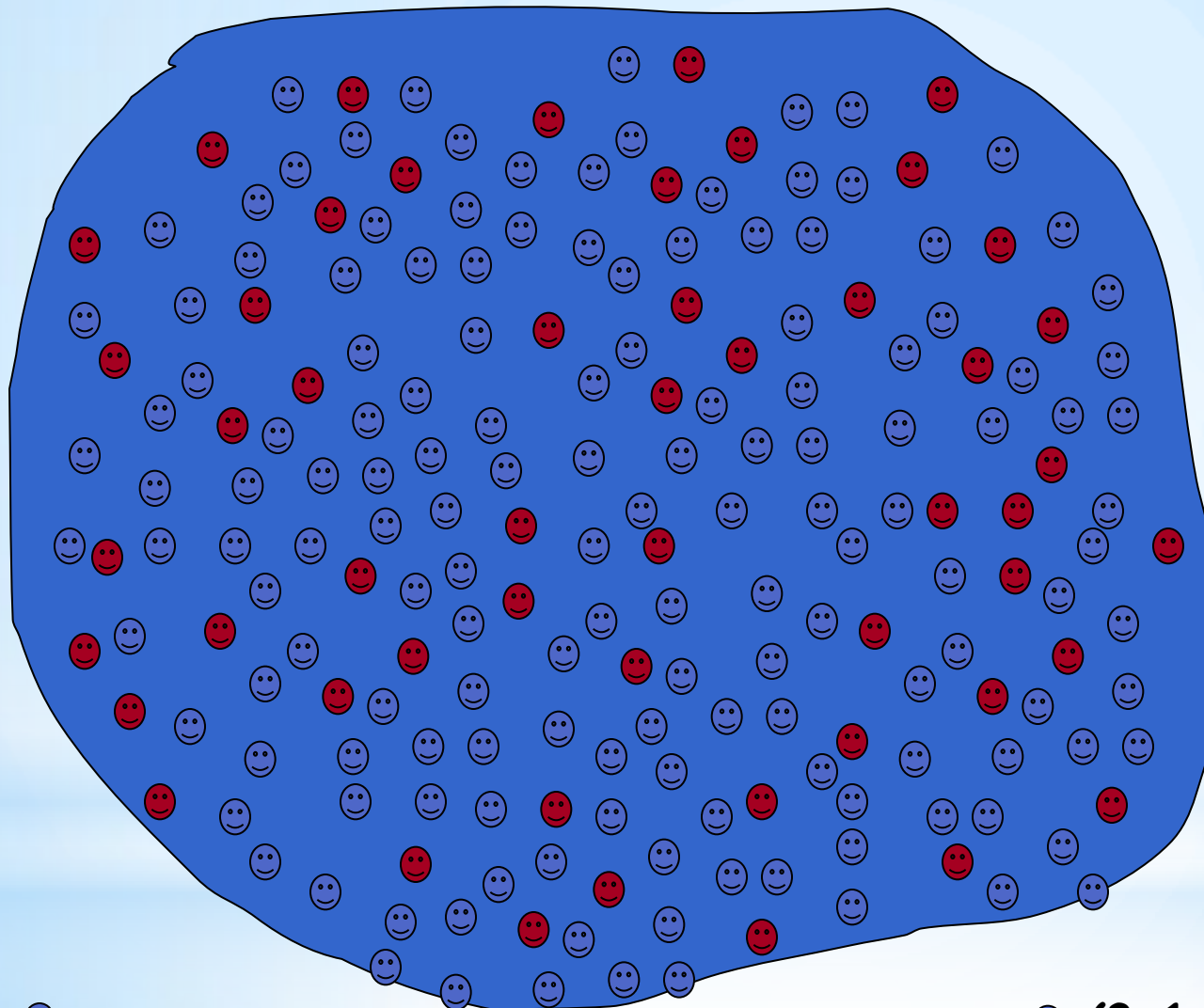


● $150/200 = 0.75$
● $50/200 = 0.25$

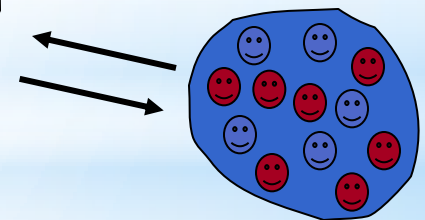
$m = 0.33$
4 alleli escono
4 alleli entrano
con stesse freq
della popolazione
di partenza



● $3/12 = 0.25$
● $9/12 = 0.75$



$$m = 0.33$$



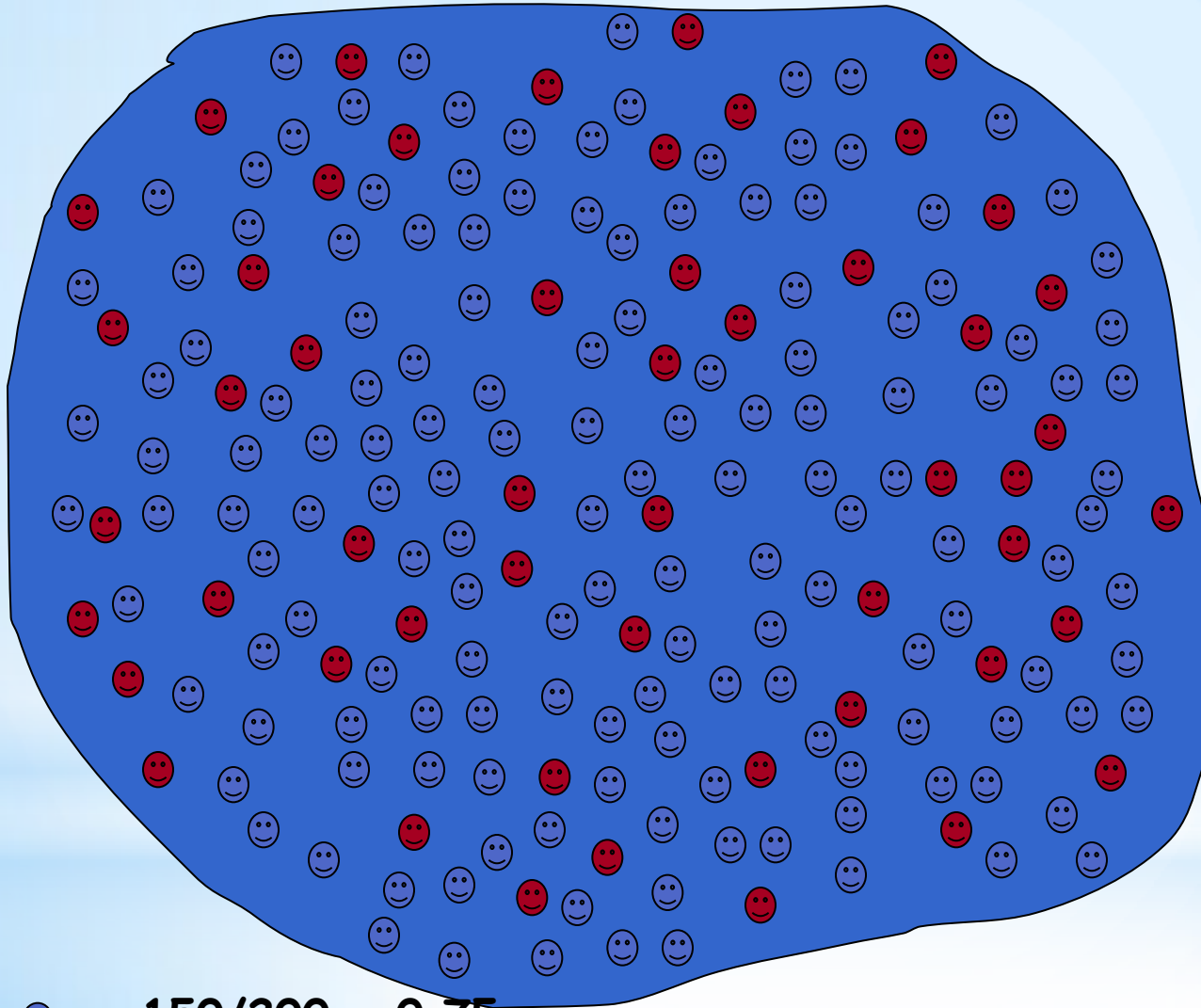
$$\text{blue smiley} \quad (150 - 3 + 1) / 200 = 148 / 200 \cong 0.75$$

$$\text{red smiley} \quad (50 - 1 + 3) / 200 = 52 / 200 \cong 0.25$$

$$\text{blue smiley} \quad (3 - 1 + 3) / 12 = 5 / 12 \cong 0.42$$

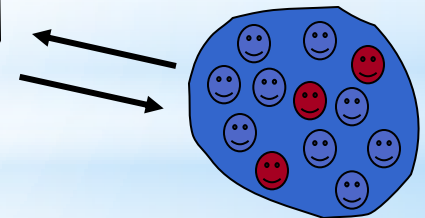
$$\text{red smiley} \quad (9 - 3 + 1) / 12 = 7 / 12 \cong 0.58$$

... dopo molte generazioni



☺ ~ $150/200 = 0.75$
☹ ~ $50/200 = 0.25$

$m = 0.33$



☺ $9/12 = 0.75$
☹ $3/12 = 0.25$

MIGRAZIONE: modello dell' isola

Limitazioni del modello:

- I flussi migratori da e per l' isola devono essere uguali $\rightarrow$ N cost.
- La migrazione deve essere costante nel corso delle generazioni.
- Non vi è migrazione differenziale per genotipo.

Isola $\rightarrow$ fr. (A_1) = p e fr. (A_2) = q

Continente $\rightarrow$ fr. (A_1) = $\overline{p}$ e fr. (A_2) = $\overline{q}$

Come cambia la frequenza di A_2 nell' isola e nel continente?

Isola:

$(1 - m)$ = individui che restano sull' isola

m = individui immigrati dalla popolazione generale

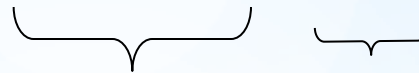
q = frequenza dell' allele A_2 negli individui che restano sull' isola

$\overline{q}$ = frequenza dell' allele A_2 negli individui immigrati dal continente

MIGRAZIONE: modello dell' isola

La frequenza dell' allele A_2 nell' isola dopo una generazione sarà:

$$q' = (1 - m) q + m \bar{q}$$



I termine

II termine

I termine: sono gli alleli degli individui che sono rimasti

II termine: sono i nuovi alleli introdotti dagli immigranti

MIGRAZIONE: modello dell' isola

Dopo n generazioni la deviazione tra la frequenza dell' isola e quella del continente sarà:

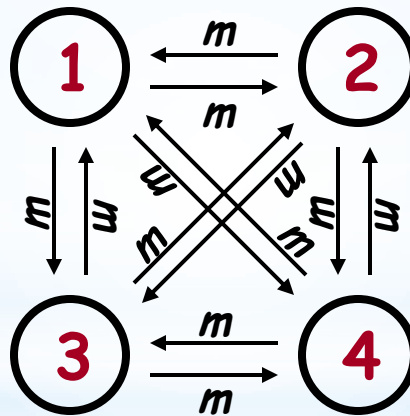
$$q_n - \bar{q} = (1 - m)^n (q_0 - \bar{q})$$

La deviazione tra la frequenza dell' isola e quella del continente diminuisce a ogni generazione di un fattore $(1 - m)$, quindi è chiaro che l'avvicinamento all'equilibrio (stesse frequenze tra isola e continente, cioè quelle del continente) è tanto più rapido quanto maggiore è il tasso di migrazione m .

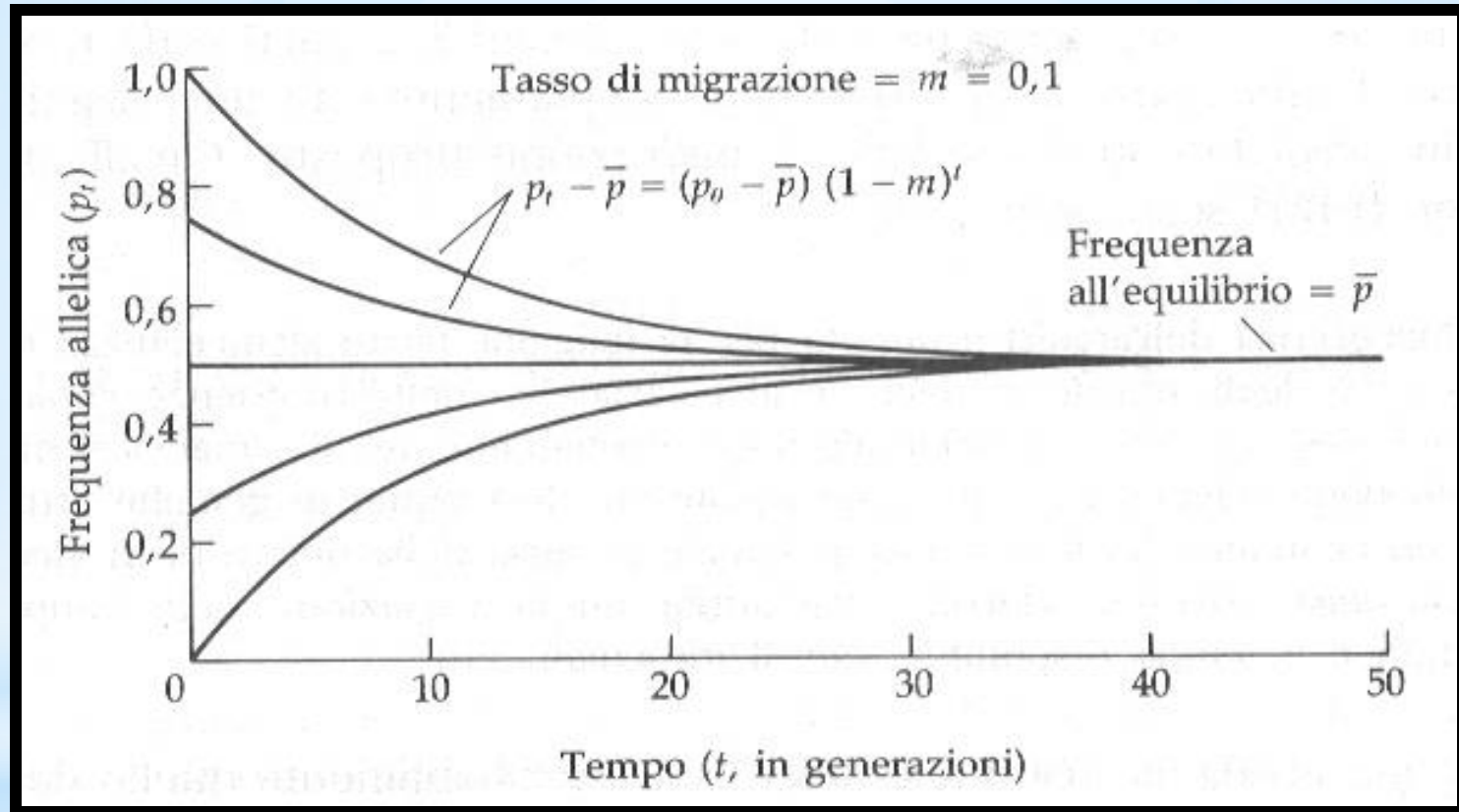
MIGRAZIONE: modello dell'arcipelago

Un gruppo di popolazioni, ben delimitate nello spazio, che si scambiano migranti a ogni generazione.

Per esempio: 4 popolazioni di eguali dimensioni, con eguale tasso di migrazione



MIGRAZIONE tra popolazioni



Le 4 popolazioni inizialmente differenziate convergono a p medio che si avvicinerà alla media ponderata delle frequenze

$n = 10$ (n. generazioni di immigrazione di africani negli USA a partire da circa 250-300 anni fa con la schiavitù)

$$\frac{0,42}{0,60} = (1 - m)^{10}$$

$$-0,3567 = 10 \ln (1 - m)$$

$$\ln (1 - m) = -0,03567$$

$$m = 0.035$$

E' il tasso di migrazione genica dagli europei agli africani nella popolazione afroamericana degli USA, cioè a ogni generazione il 3,5% di geni europei è entrato a far parte del pool di geni della popolazione degli afroamericani degli USA.

“Decrease of heterozygosity”

TABLE 4.1 HIERARCHICAL STRUCTURE OF *LINANTHUS PARRYAE*

Region	Subpopulations		Regions		Total	
	Allele frequency	Heterozygosity	Average allele frequency	Heterozygosity	Average allele frequency	Heterozygosity
W	0.573	0.4893	0.5153	0.4995		
	0.717	0.4058				
	0.504	0.5000				
	0.657	0.4507				
	0.302	0.4216				
	0.339	0.4482				
C	9 × 0.000	0.0000	0.0138	0.0272		
	0.032	0.0620				
	0.007	0.0139				
	0.008	0.0159				
	0.005	0.0100				
	0.009	0.0178				
	0.005	0.0100				
	0.010	0.0198				
	0.068	0.1268				
	0.002	0.0040				
	0.004	0.0080				
	0.126	0.2202				
E	0.106	0.1895	0.1888	0.3062	0.1374	0.2371
	0.224	0.3476				
	0.411	0.4842				
	0.014	0.0276				
Average heterozygosity		$H_S = 0.1424$	$H_R = 0.1589$		$H_T = 0.2371$	

Source: Data from Wright 1943a.

La variabilità genetica moderna riflette in buona parte l'effetto di migrazioni umane recenti



Early *Homo sapiens sapiens* in Africa

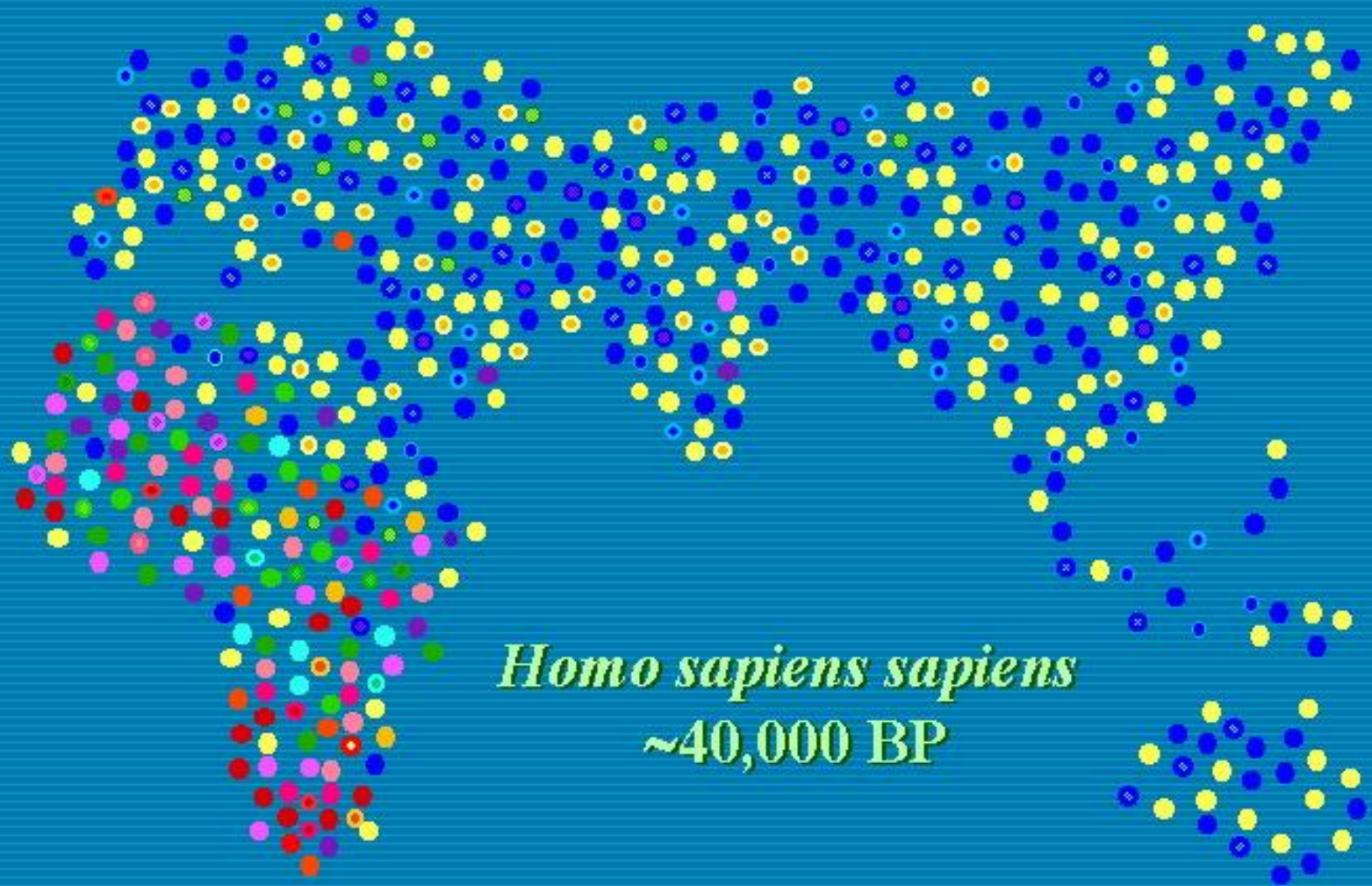


150,000 to 100,000 BP

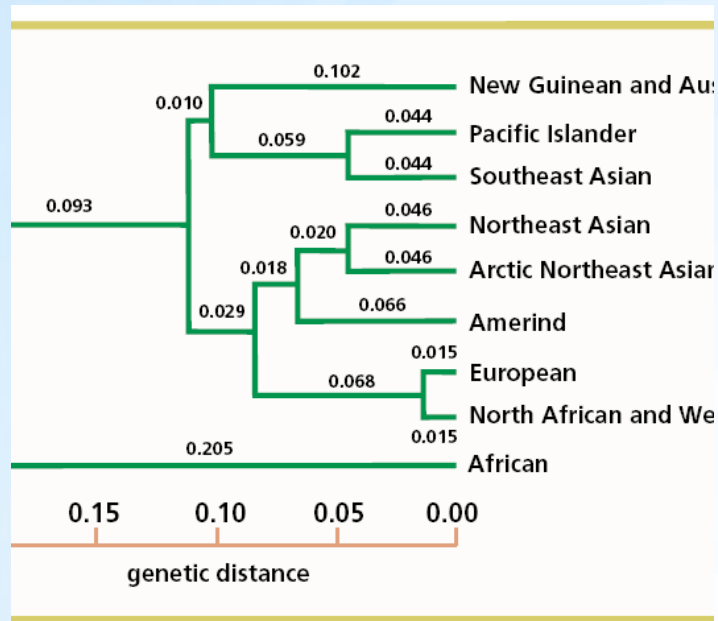


Homo sapiens sapiens
colonizing south west Asia

~100,000 BP



Homo sapiens sapiens
~40,000 BP



ree of world populations. Phylogenetic tree based on polymorphisms of 120 protein genes
continental sub-areas and F_{st} genetic distances¹⁴. Root placed assuming a constant rate of

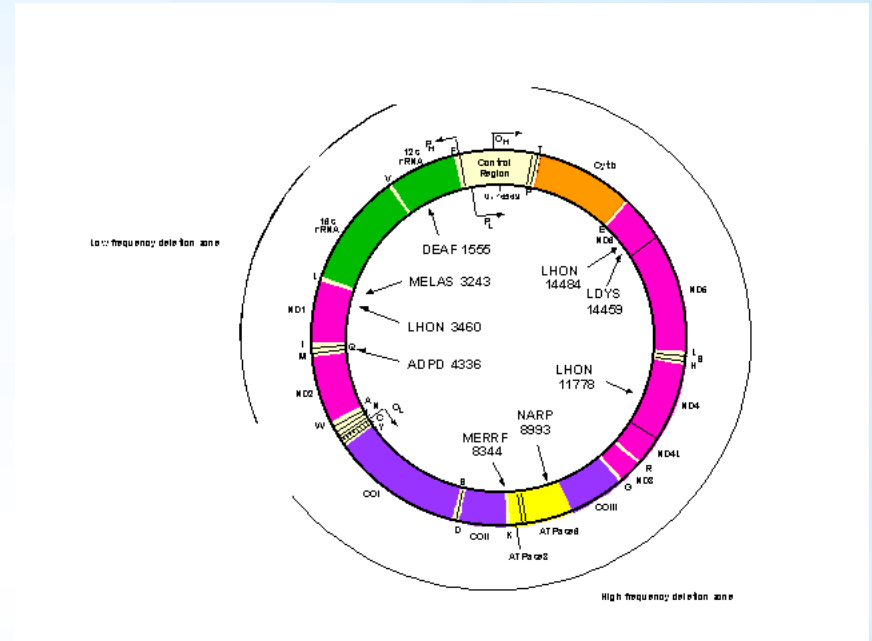
* Inference of Human Phylogenetic Tree

- * Archeological evidence
 - * origin in Africa 50-100kya
 - * spread to rest of world, 50-60kya
- * What does genetic evidence say?
- * What about the location?

* Time to Most Recent
Common Ancestor (TMRCA)

* Mitochondrial DNA

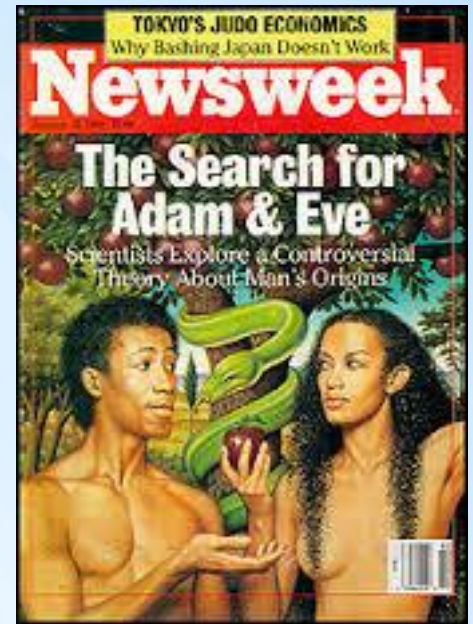
- * An organelle of the animal cell
- * Krebs' s Cycle (aerobic respiration) takes place here
- * Transmitted only along female lineage
- * Haploid genome, independent from human “host”
- * High mutation rate



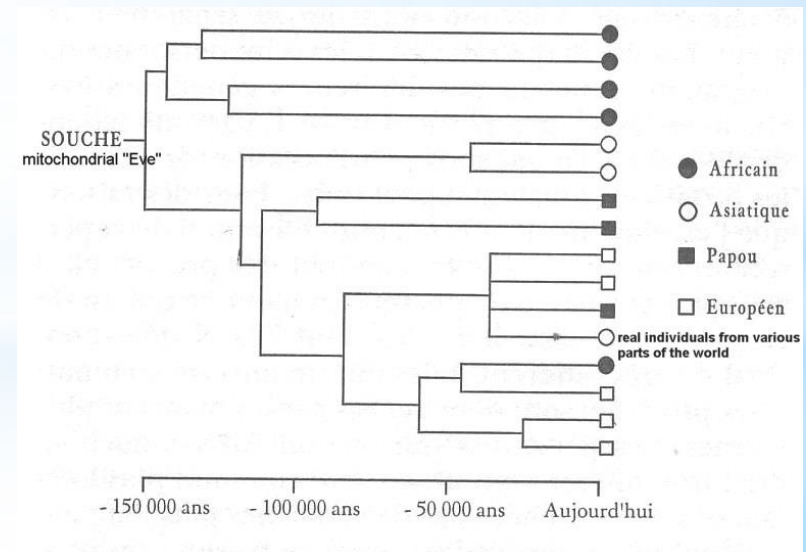
* Mitochondrial “Eye”

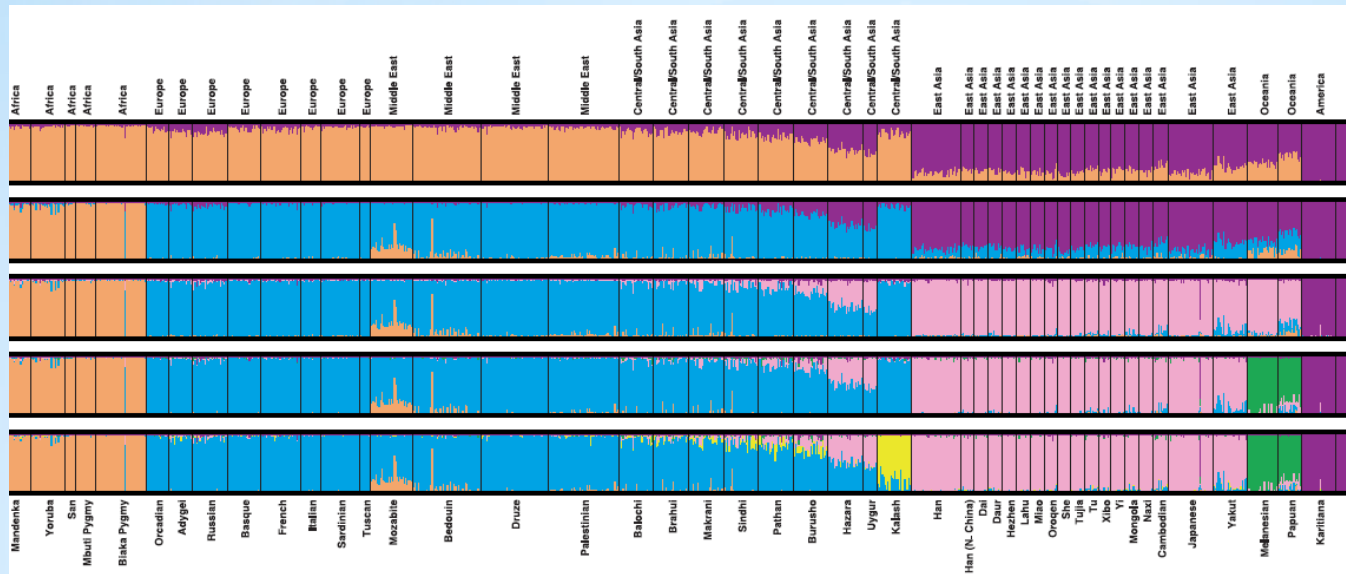
* Y-chromosome “Adam”

- * Part of the Y chromosome does not recombine
- * Hence we can do a similar trick
 - * However, only men (XY) carry the Y chromosome
 - * So we can only identify the most recent patrilineal common ancestor of all *men*



- * Why are mtDNA and Y chromosome TMRCA dates so different?
 - * lower N_E for males than for females?
 - * polygyny more frequent than polyandry?
 - * higher male mortality rates?
 - * higher male variability in reproductive success?
 - * patrilineal marriage more common than matrilineal?
 - * mtDNA mutation rates variable, causing error?





timated population structure. Each individual is represented by a vertical line, which is partitioned into K colored segments that represent individual's estimated membership fractions in K clusters. Black lines separate individuals of different populations. Populations are labeled below, with their regional affiliations above it. Ten *structure* runs at each

K produced nearly identical individual membership coefficients, however similarity coefficients above 0.97, with the exceptions of four runs at $K = 3$ that separated East Asia instead of Europe and one run at $K = 6$ that separated Karitiana instead of Kalash. The results shown for a given K is based on the highest probability run at that

* Human microsatellite data

- 1052 individuals; 52 populations; 377 autosomal microsatellite markers
- “microsatellite” or Short Tandem Repeat (STR) = 2-6 bases repeated several times
- e.g., TCTA TCTA TCTA TCTA TCTA TCTA TCTA TCTA
- “indigenous populations” only; all individuals’ grandparents lived in same place

*POLIMORFISMI

Qualsiasi carattere polimorfico mendeliano che può essere impiegato per seguire l' ereditarietà di un segmento cromosomico attraverso un albero genealogico

* POLIMORFISMO

I polimorfismi genetici sono variazioni nelle sequenze di DNA presenti in una popolazione con una frequenza maggiore dell' 1%. Quando la frequenza e' inferiore a tale valore arbitrario, si preferisce parlare di varianti genetiche rare, che in molti loci sono presenti in aggiunta ai polimorfismi.

* Polymorphism

* polymorphism= the inheritance of genes in different forms termed alleles

alleles have different DNA sequences

* polymorphic locus: the frequency of the most common allele is less than 99%.

1 allele in 100 alleles

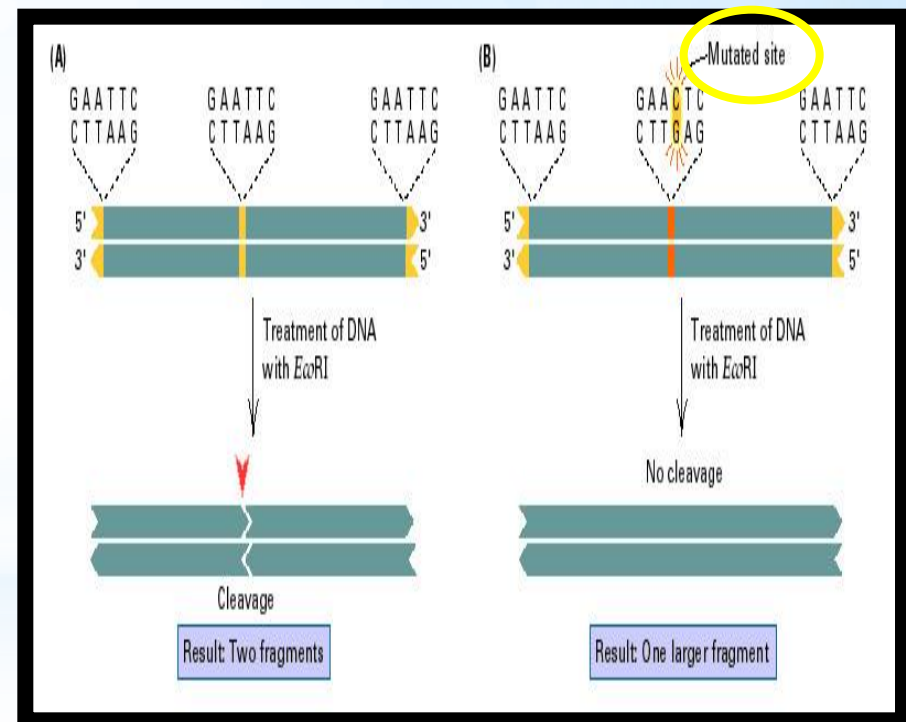
100 alleles =50 people

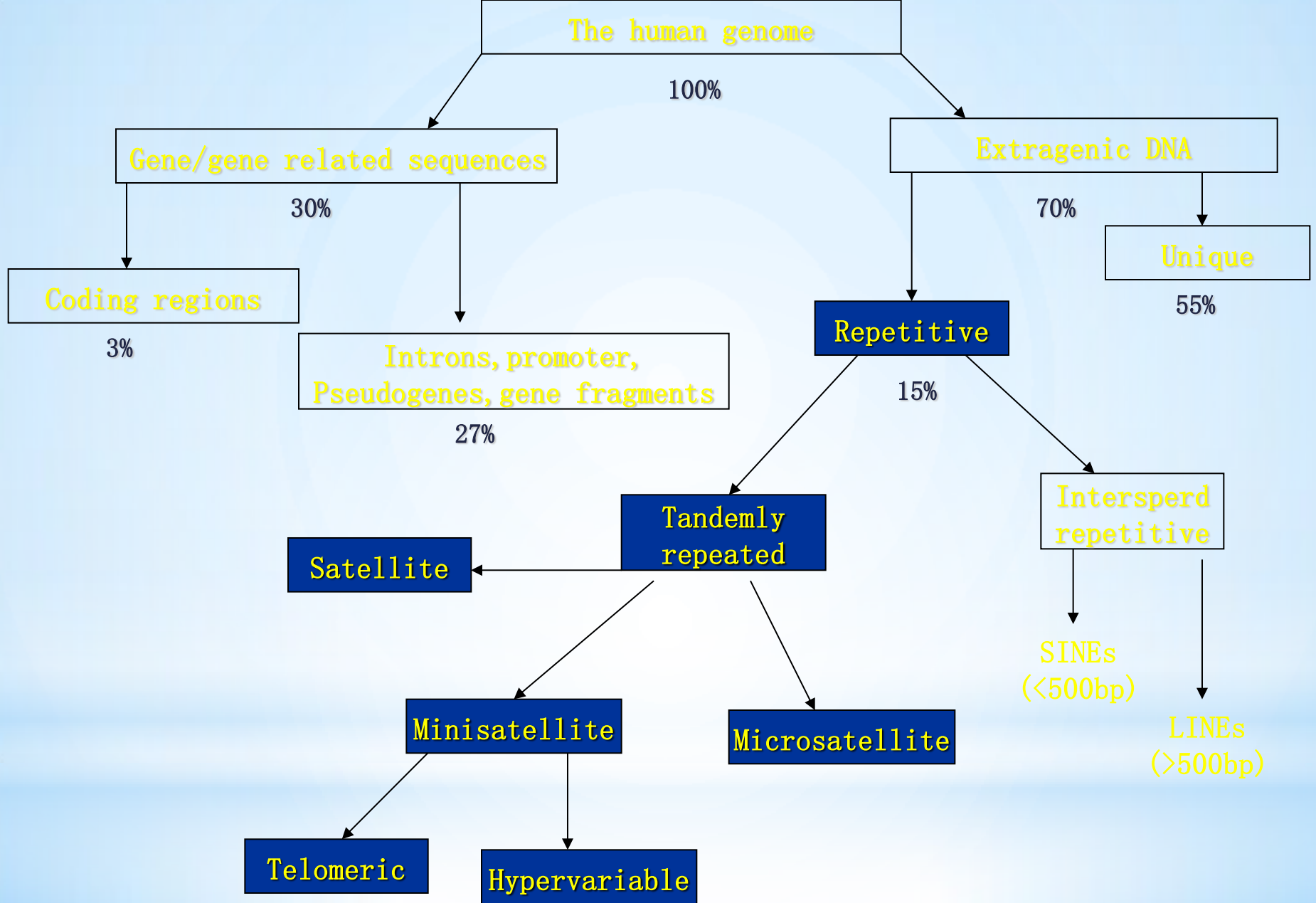
1 person in 50 (2%) is heterozygous

* RFLP

I primi marcatori molecolari ad essere studiati furono gli RFLP (Restriction Fragment Length Polymorphism): particolari tratti di DNA presenti nella popolazione e trasmessi in modo ereditario

La regione del genoma di interesse viene amplificata tramite PCR e i prodotti ottenuti vengono incubati con un enzima di restrizione in grado di riconoscere una sequenza specifica e di catalizzare una reazione di taglio al suo interno





Sequential breakdown of the genome into component DNA types (Bennett P: Demystified...Microsatellites. *J Clin Pathol:Mol Pathol* 2000,53:177-183)

* DNA ripetitivo intersperso

- * Le singole unità ripetute non sono raggruppate ma sparse in più punti del genoma. Gli esempi più comuni sono le sequenze SINEs (Short Interspersed Nuclear Element) e LINEs (Long Interspersed Nuclear Element)

LINEs

- * Sono associate prevalentemente a DNA genomico ricco in A/T perché tendono a posizionarsi in regioni del cromosoma povere di geni allo scopo di imporre il minimo impatto mutazionale al genoma.

SINEs

- * Sono associate prevalentemente a DNA genomico ricco in G/C. *Perché?* Sembra che tali sequenze svolgano una qualche funzione positiva per il genoma: esse sarebbero espresse in condizioni di stress ed i risultanti RNA legherebbero una particolare protein chinasi PKR e bloccherebbero la sua capacità di inattivare la traduzione.

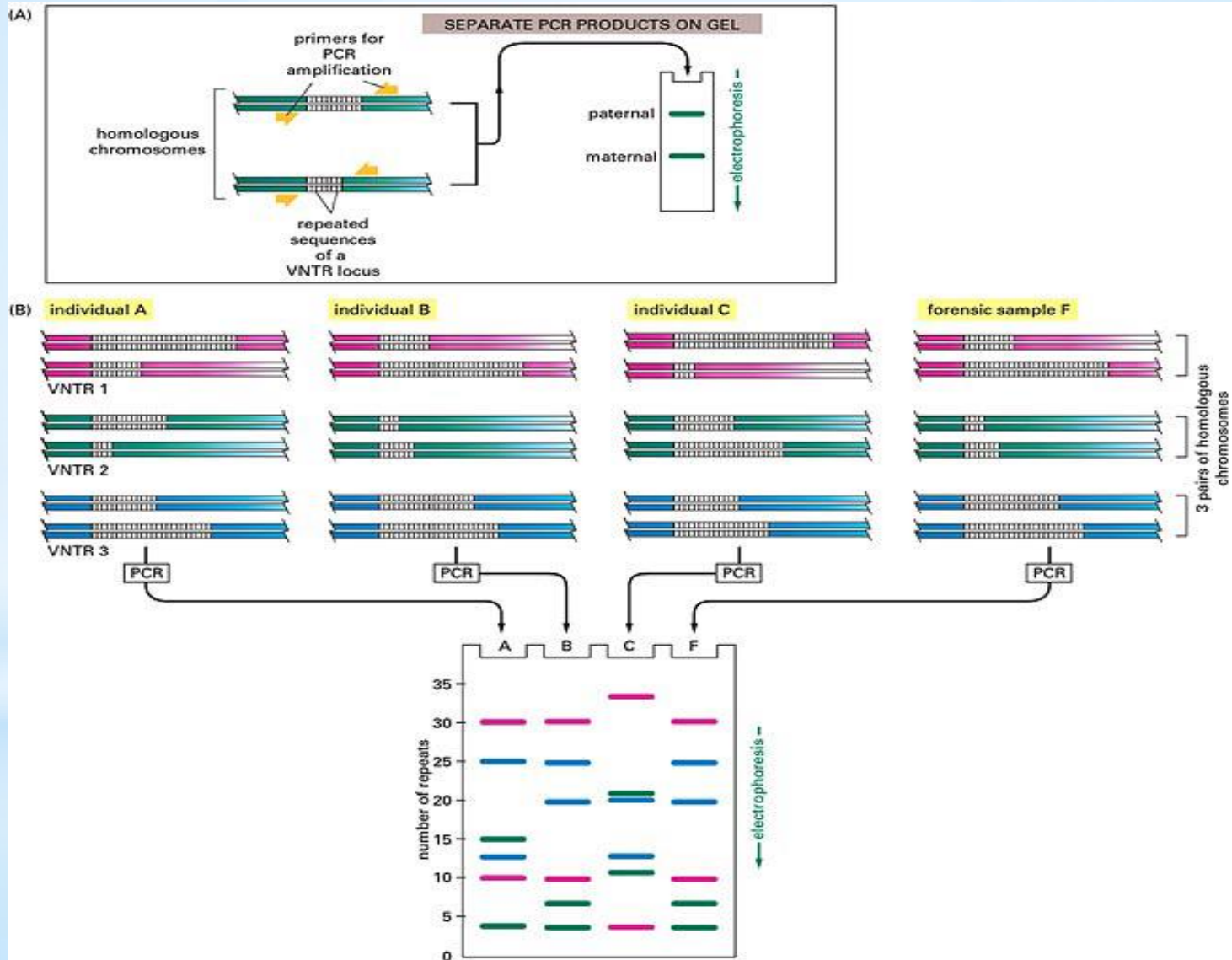
* DNA TELOMERICO

- * Questo tipo di DNA è localizzato all' estremità dei cromosomi, nei telomeri
- * Il costituente principale del DNA telomerico è rappresentato da 10-15kb di unità esanucleotidiche ripetute in tandem, in particolare TTAGGG, che vengono aggiunte da un enzima specializzato, la telomerasi
- * Tali ripetizioni sono responsabili della funzione dei telomeri in quanto agiscono come protezioni delle estremità dei cromosomi dalla degradazione e dalle perdite di materiale
- * Inoltre forniscono un meccanismo per la replicazione delle estremità lineari del DNA cromosomico

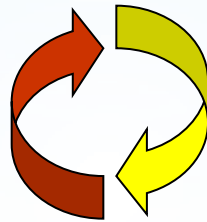
* DNA minisatellite ipervariabile o VNTR (Variable Number of Tandem Repeat)

- * Sono sequenze altamente polimorfiche ed organizzate in oltre 1000 gruppi (lunghi da 0.1 a 20 kb) di corte unità ripetute in tandem, che variano considerevolmente per dimensioni ma posseggono una sequenza comune centrale (*core*), GGGCAGGAXG
- * Molti di questi gruppi si trovano vicino ai telomeri
- * La maggior parte di queste sequenze *non* sono trascritte eccetto alcuni elementi all' interno di sequenze intrageniche non codificanti
- * Il significato non è ancora chiaro, ma indipendentemente dalla loro reale funzione nel genoma umano, esiste un utilizzo pratico di questi gruppi, ma in genere delle sequenze ripetute, nel *DNA fingerprinting* (impronta digitale del DNA)

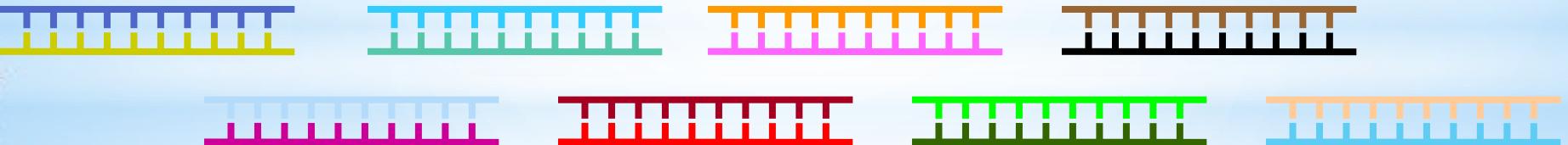
Minisatelliti: Riconosciuti su gel dopo amplificazione con PCR



PCR Copies DNA Exponentially through Multiple Thermal Cycles



Thermal cycle

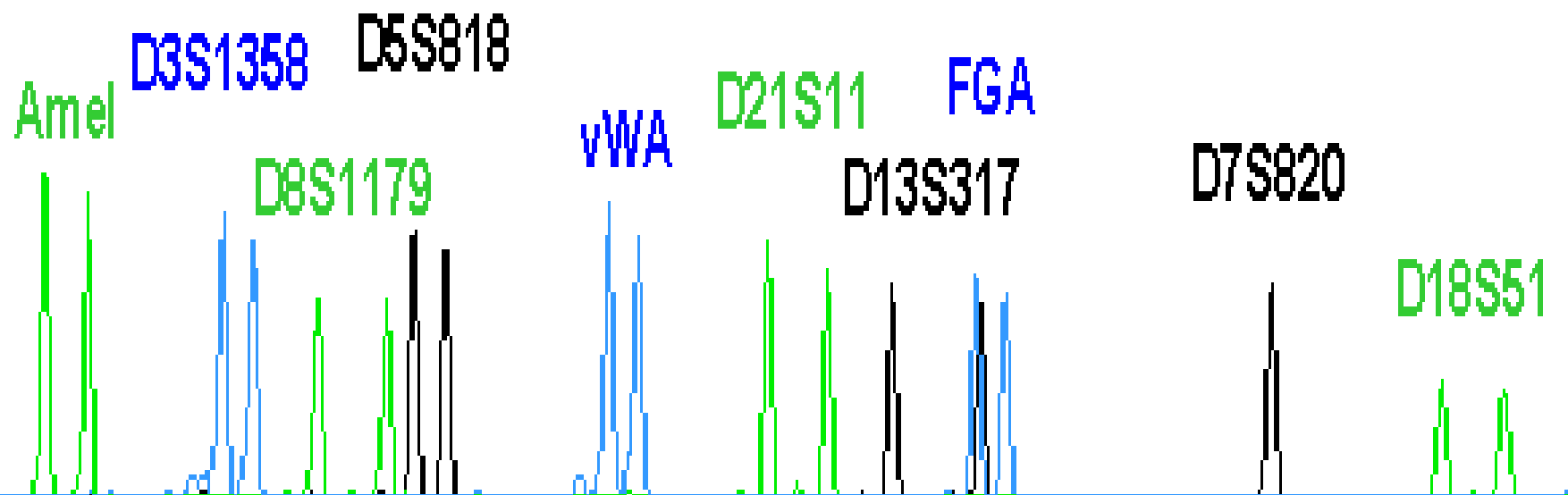


In 32 cycles at 100% efficiency, 1.07 billion copies of targeted DNA region are created

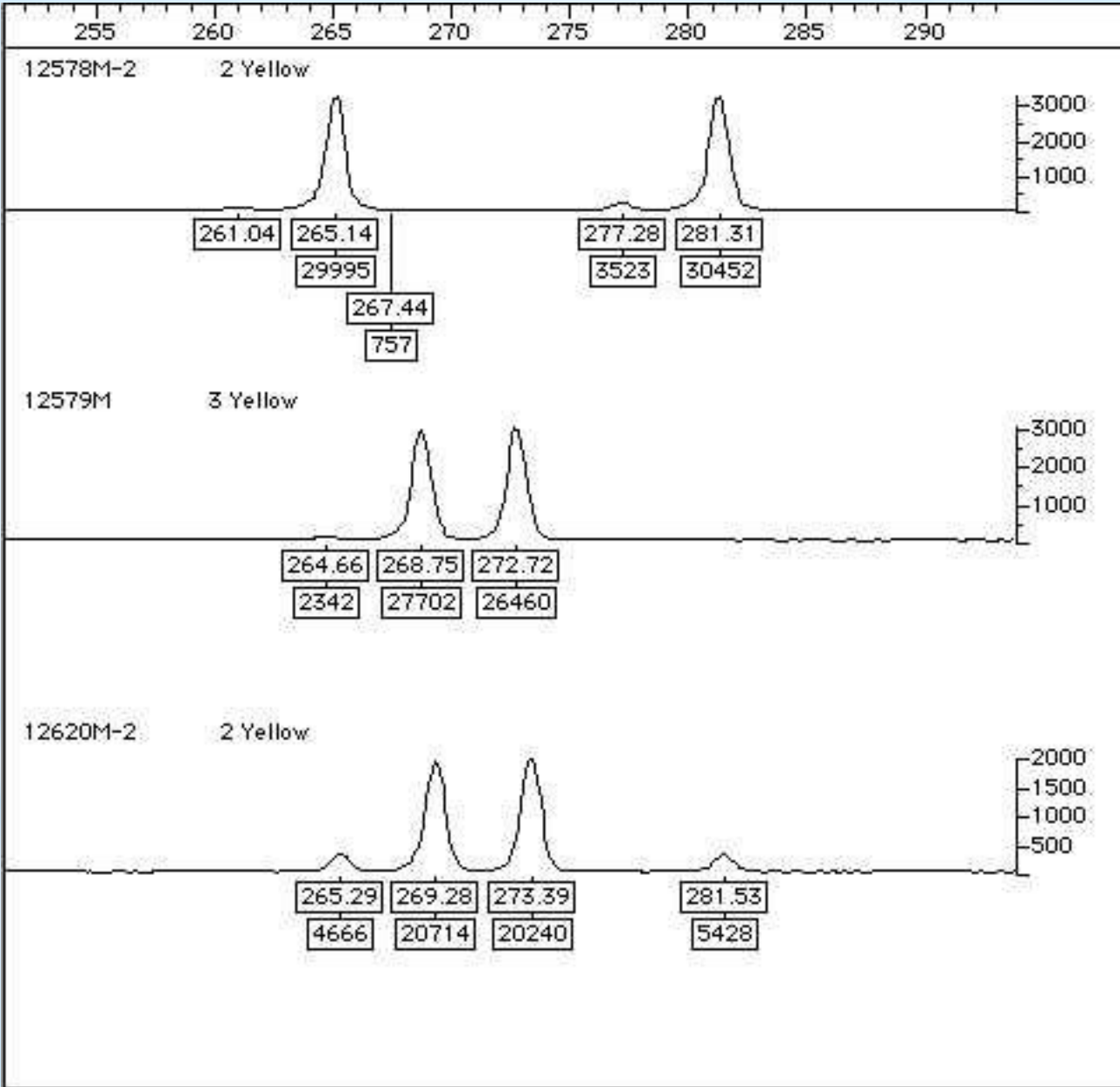
ELETTROFORETOGRAMMA

PCR Product Size (bp)

100 125 150 175 200 225 250 275 300



Esempio: caso I



(Single Nucleotide Polymorfism)

- * Gli SNPs, scoperti negli anni 80, sono variazioni di sequenza del DNA che si verificano quando è alterato un singolo nucleotide della sequenza genomica
- * Affinchè una variazione nella sequenza nucleotidica sia considerata uno SNP, deve essere presente in almeno l' 1% della popolazione per cui l' inserzione/delezione di una singola base non deve essere considerata uno SNP
- * A differenza delle VNTR e degli STRs, gli SNPs non sono sequenze ripetute e possono trovarsi sia nelle regioni codificanti che non-codificanti del genoma (Dwight H.O. et al. *Journal of Molecular Diagnostics*, 2000, 2(4):202-208)

- * Attualmente sono i marcatori molecolari più utilizzati perché il grande vantaggio nell'utilizzarli è dato dall'elevato numero di polimorfismi che possono essere genotipizzati e dalla loro elevata densità lungo tutto il genoma
- * Costituiscono ca il 90% di tutti i polimorfismi presenti nel genoma umano. A giugno del 2004 nell'uomo è stata stimata una frequenza per gli SNP pari a 1/700bp
- * Il recente progresso della genomica ha messo in luce come una parte rilevante della variabilità tra individui sia da attribuirsi a polimorfismi a singolo nucleotide

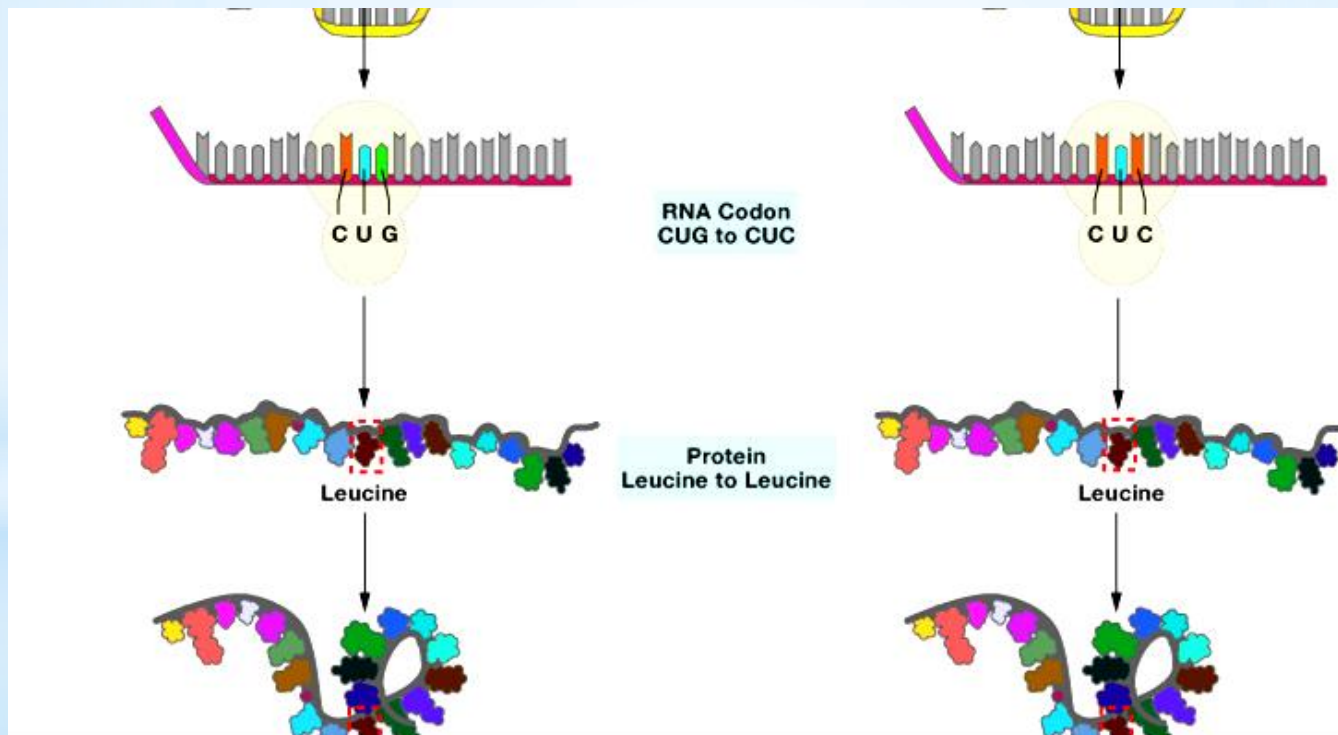
- * Gli SNP acquistano particolare rilevanza in campo biomedico quando possono essere messi in relazione a patologie che non presentano una trasmissione genetica semplice
- * Confrontando lo schema e le frequenze degli SNP su geni potenzialmente coinvolti in patologie e i fenotipi esibiti dai soggetti portatori, è possibile utilizzare tali sequenze come marcatori molecolari

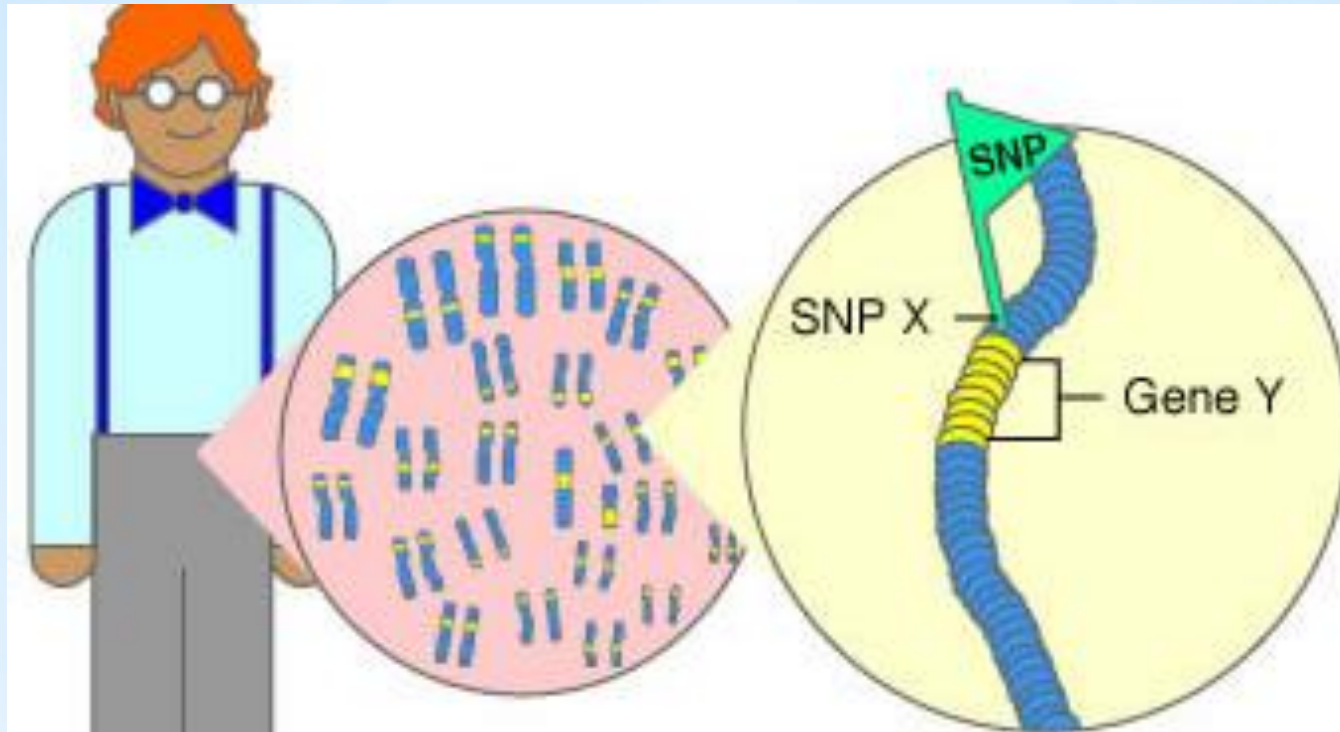
PROBLEMI

- * Attualmente sono poco conosciute le distribuzioni degli SNP all' interno di diverse popolazioni
- * Non essendo uguali tutti gli SNP, per capire il loro effetto sarà importante eseguire un' analisi computazionale prima di eseguire uno studio relativo al loro eventuale coinvolgimento in una patologia

- * SNPs are found in
 - * coding and (mostly) noncoding regions.
- * Occur with a very high frequency
 - * about 1 in 1000 bases to 1 in 100 to 300 bases.
- * The abundance of SNPs and the ease with which they can be measured make these genetic variations significant.
- * SNPs close to particular gene acts as a marker for that gene.
- * SNPs in coding regions may alter the protein structure made by that coding region.

* SNPs may / may not alter protein structure

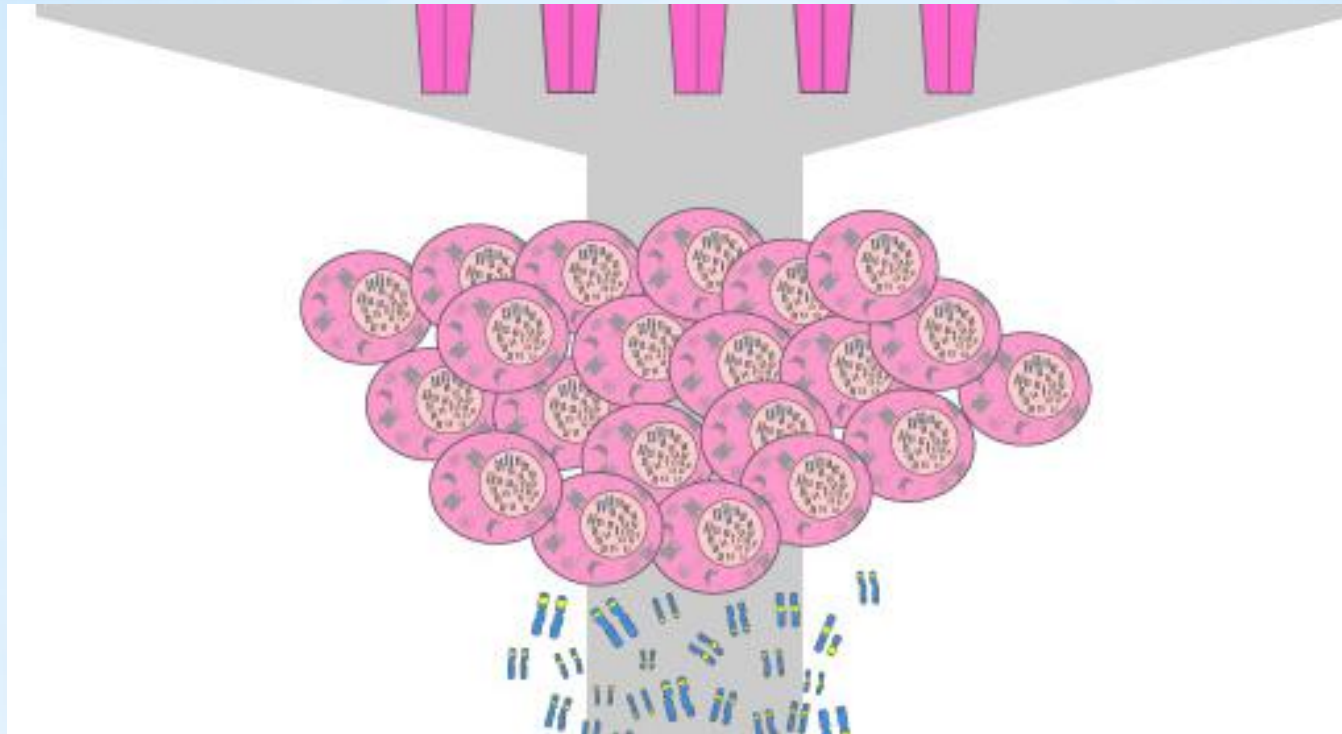




***SNPs act as gene markers**

- * Sequence genomes of a large number of people
- * Compare the base sequences to discover SNPs.
- * Generate a single map of the human genome containing all possible SNPs => SNP maps

*** SNP maps**



* **SNP Maps**

- * Genome of each individual contains distinct SNP pattern.
- * People can be grouped based on the SNP profile.
- * SNPs Profiles important for identifying response to Drug Therapy.
- * Correlations might emerge between certain SNP profiles and specific responses to treatment.

* SNP Profiles

- * Hybridization Techniques
 - * Micro arrays
 - * Real time PCR
- * Enzyme based Techniques
 - * Nucleotide extension
 - * Cleavage
 - * Ligation
 - * Reaction product detection and display
- * Comparison of Techniques used

* Techniques to detect known Polymorphisms

- * Direct Sequencing
- * Microarray
- * Cleavage / Ligation
- * Electrophoretic mobility assays
- * Comparison of Techniques used

* Techniques to detect unknown Polymorphisms

SIGNIFICANCE OF SNPs

- ❖ IN DISEASE DIAGNOSIS**
- ❖ IN FINDING PREDISPOSITION TO DISEASES**
- ❖ IN DRUG DISCOVERY & DEVELOPMENT**
- ❖ IN DRUG RESPONSES**
- ❖ INVESTIGATION OF MIGRATION PATTERNS**

ALL THESE ASPECT WILL HELP TO LOOK FOR MEDICATION & DIAGNOSIS AT INDIVIDUAL LEVEL

SNP genotyping methods for detecting genes contributing to susceptibility or resistance to multifactorial diseases, adverse drug reactions:

=> **case-control association analysis**

case

... . GCC**G**TTGAC... .
... . GCC**A**TTGAC... .

control

... . GCC**A**TTGAC... .
... . GCC**A**TTGAC... .



HAPLOTYPE

A set of closely linked genetic markers present on one chromosome which tend to be inherited together (not easily separable by recombination)

*SNP-Haplotype

Phenotype

BLACK EYE

BROWN EYE

BLACK EYE

BLUE EYE

BROWN EYE

BROWN EYE

6

5

4

3

2

1

SNP

SNP

GAT**A**TTTCGTAC**G**GA-T

GAT**G**TTTCGTACT**T**GAAT

GAT**A**TTTCGTAC**G**GA-T

GAT**A**TTTCGTAC**G**GAAT

GAT**G**TTTCGTACT**T**GAAT

GAT**G**TTTCGTACT**T**GAAT

Haplotypes

AG – 2/6 (BLACK EYE)

GTA 3/6 (BROWN EYE)

AGA 1/6 (BLUE EYE)

DNA Sequence



HAPLOTYPE CORRELATION WITH PHENOTYPE

- ❖ The “Haplotype centric” approach combines the information of adjacent SNPs into composite multilocus haplotypes.
- ❖ Haplotypes are not only more informative but also capture the regional LD information, which is assumed to be robust and powerful
- ❖ Association of haplotype frequencies with the presence of desired phenotypic frequencies in the population will help in utilizing the maximum potential of SNP as a marker.

ADVANTAGES:

1. **SNPs ARE THE MOST FREQUENT FORM OF DNA VARIATIONS**
2. **THEY ARE THE DISEASE CAUSING MUTATIONS IN MANY GENES**
3. **THEY ARE ABUNDANT & HAVE SLOW MUTATION RATES**
4. **EASY TO SCORE**
5. **MAY WORK AS THE NEXT GENERATION OF GENETIC MARKERS**

Some important SNP database Resources

1. dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>)

LocusLink (<http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi>)

2. TSC (<http://snp.cshl.org/>)

3. SNPper (<http://snpper.chip.org/bio/>)

4. JSNP (<http://snp.ims.u-tokyo.ac.jp/search.html>)

5. GeneSNPs (<http://www.genome.utah.edu/genesnps/>)

6. HGVbase (<http://hgibase.cgb.ki.se/>)

7. PolyPhen (<http://dove.embl-heidelberg.de/PolyPhen/>)

OMIM (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>)

8. Human SNP database

(<http://www-genome.wi.mit.edu/snp/human/>)