

READING AND ANALYSIS OF MICROSATELLITE DATA



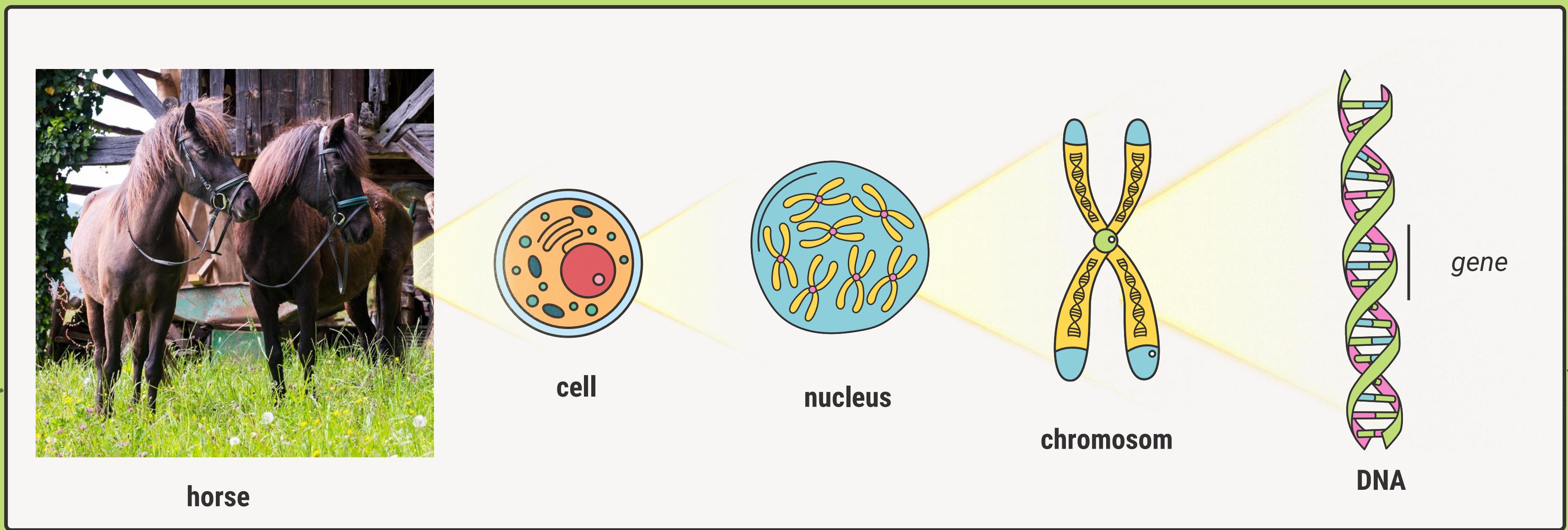
Summer school 2025

 BIOINFORMATICS APPROACHES IN ANIMAL BREEDING



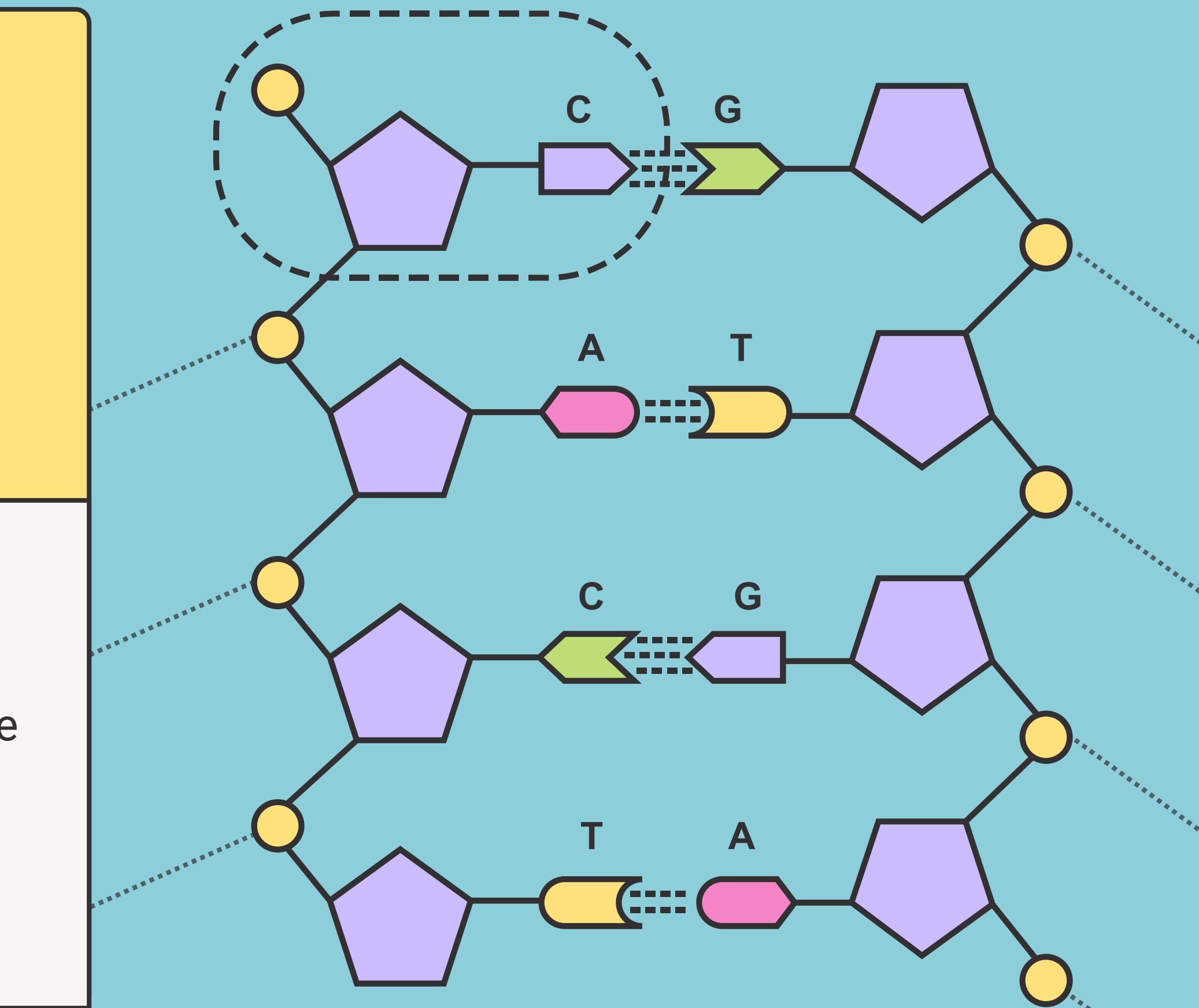
BASIC KNOWLEDGE

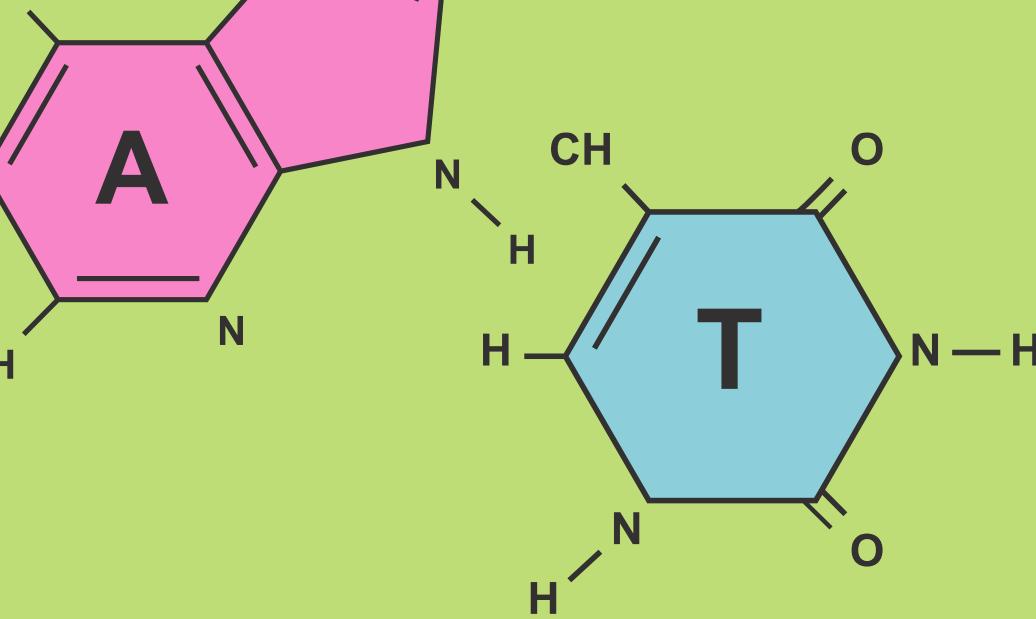
From animal to its genetic code



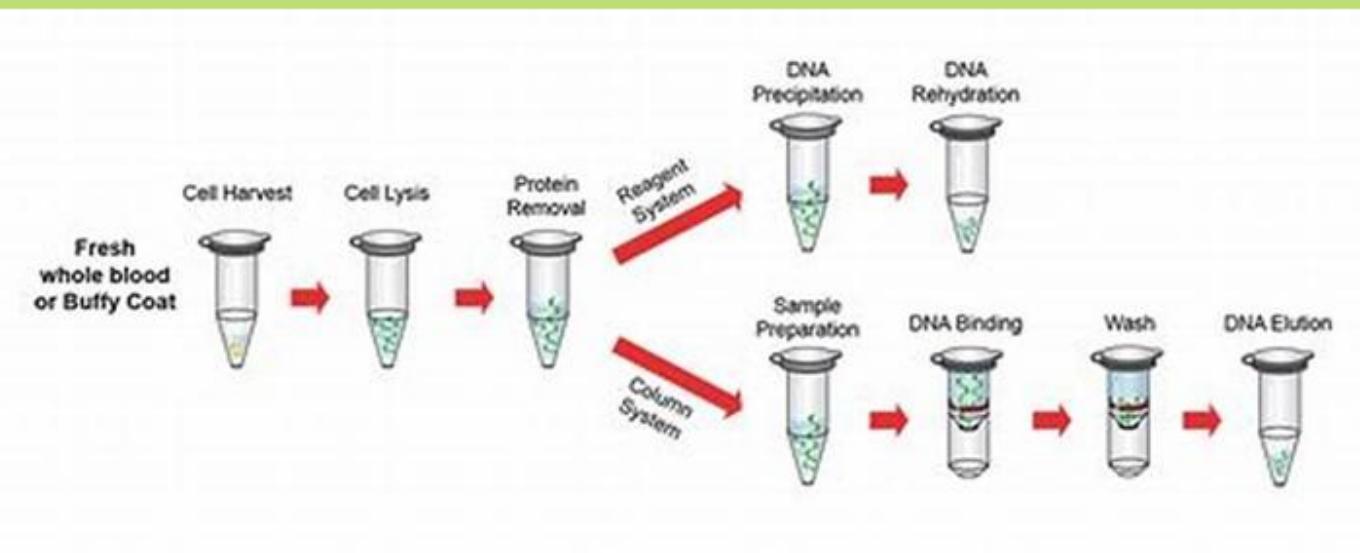
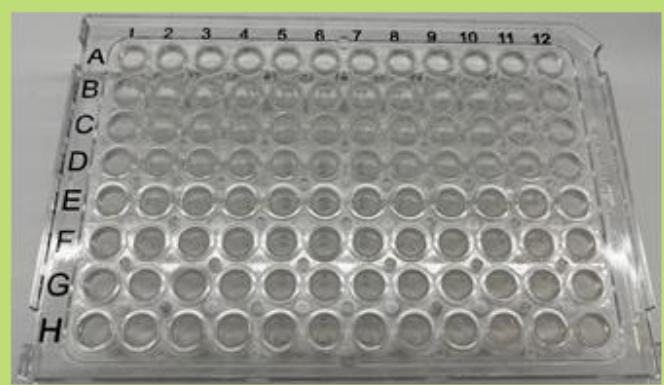
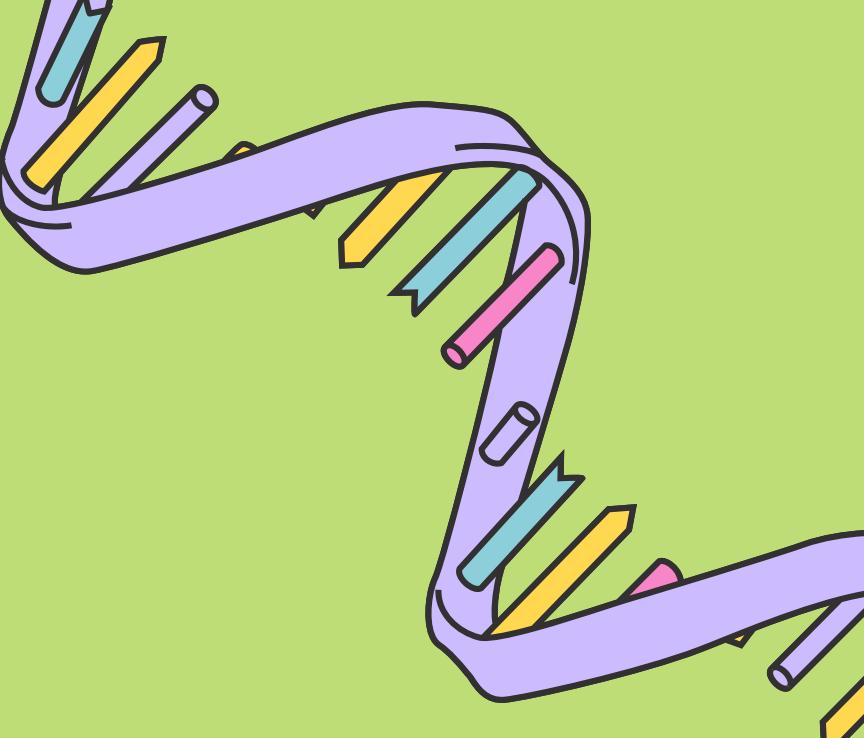
DNA ISOLATION

- separation of DNA from other cellular components in the nucleus
- blood, saliva, tissue, or hair sample
- several methods that differ in price, time requirements, and complexity of the procedure
- amount, purity, and degradation of DNA
- manual or automated procedure





HOW DO WE GET STR DATA?



MULTIPLEX PCR

It basically proceeds as a normal PCR reaction, but in multiplex PCR we amplify two or more sequences in the same reaction, for which several sets of initial oligonucleotides are used.



SELECTION OF INITIAL OLIGONUCLEOTIDES

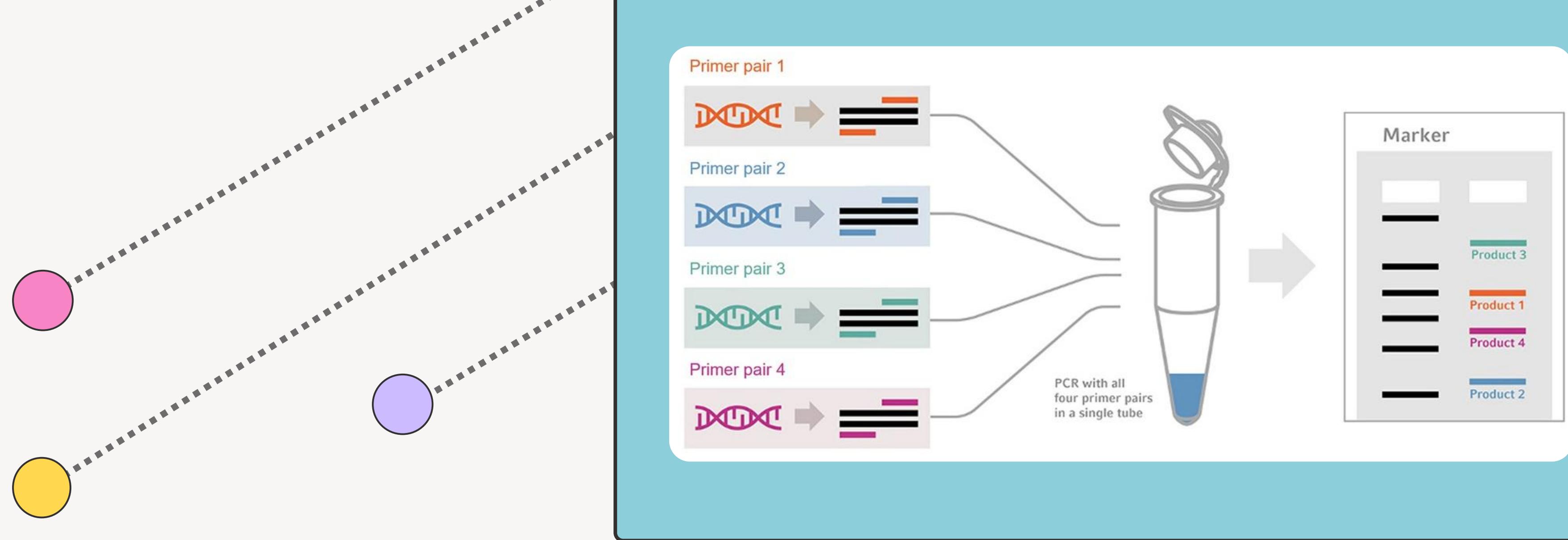
All examples must have similar T-fits and must not be homologous to each other (e.g., dimers). Selection of initial oligonucleotides

PCR REACTION PRODUCTS

PCR reaction products must be of the same size order (100–500 bp).

ADVENTAGES

By multiplying simultaneously, we can quickly analyse multiple target sequences in the same reaction, saving time and money for analysing the same number of samples.

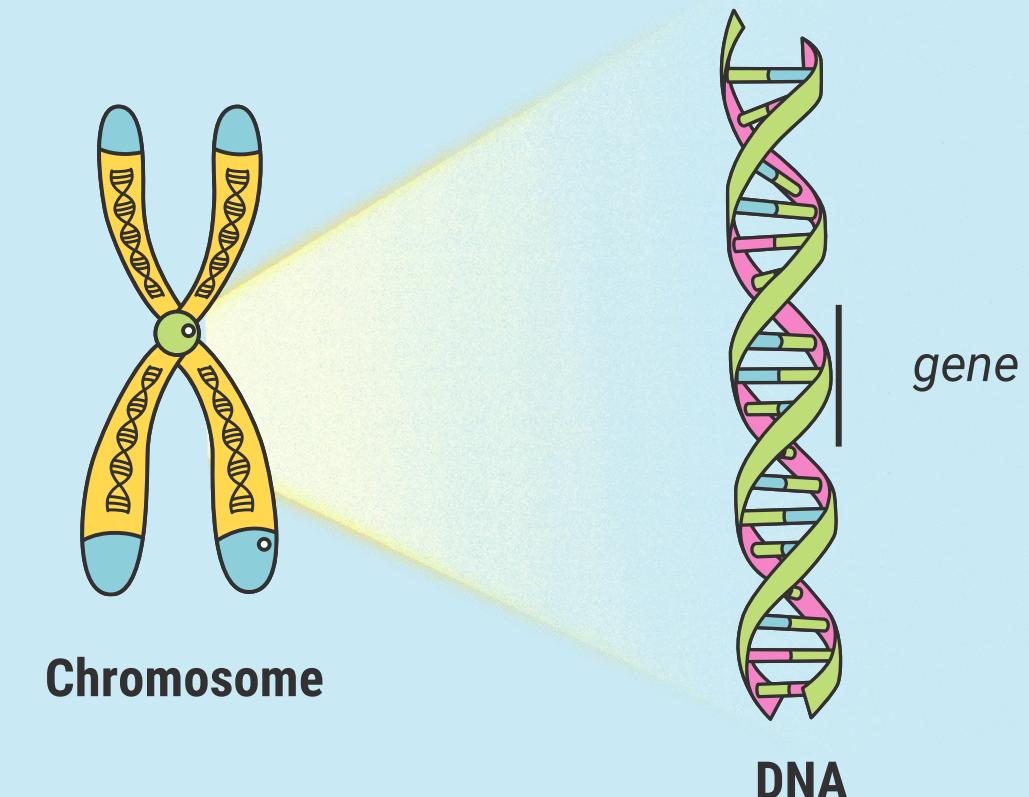


For each target sequence, specific primer oligonucleotides (F and R) are added to the reaction mixture. This allows simultaneous amplification of multiple DNA regions.

PRINCIPLE OF MULTIPLEX PCR

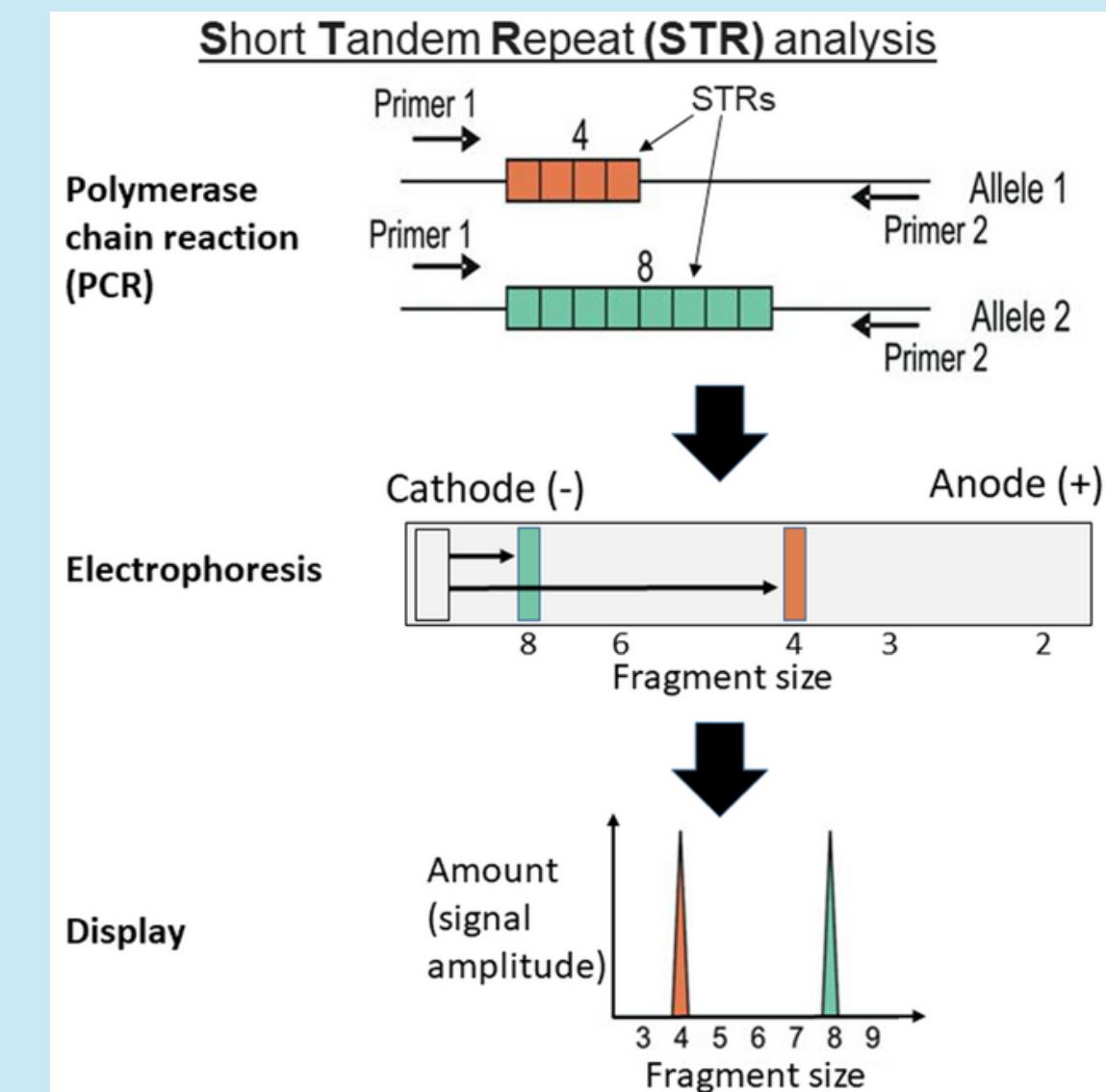
Each forward primer oligonucleotide is labeled with a color tag, which then allows us to analyze the resulting products. Products that are similar or identical in length must be labeled with different color tags.

PARENTAGE VERIFICATION



MICROSATELLITE MARKERS

- polymorphic DNA loci composed of repeating nucleotide sequences
- STR (short tandem repeats), usually between 2 and 7 bp in length, with 5–50 repeats
- are distributed throughout the genome
- are inherited according to Mendel's laws of inheritance
- may be specific to individual populations
- have specific sizes, which facilitates the interpretation of results



PARENTAGE VERIFICATION

SNP

- single-nucleotide polymorphisms
 - biallelic loci
- lower informativeness → 100 SNPs are needed to replace MS
- price reduction, gradual replacement of MS
- markers on the Y chromosome – for tracing paternal lines
- markers on mtDNA – tracing the maternal line

MICROSATELLITE MARKERS

- multiplex of 17 loci with 4 color markers
- use of letters instead of numbers for horses
 - high information content of 17 loci

 **BF** UNIVERZA V LJUBLJANI
Biotehniška fakulteta

 **GEN LAB**
Institutional member of ISAG

Preverjanje porekla z analizo DNK

Naročnik:
NEZNAN

Vzorec:

IME	Vzorec	Datum rojstva	Pasma	Št. transponderja	Življenska številka	Rejec
366	potomec	26.04.2014	L	705035000087000	705002142000366	
548	kandidatni oče	03.05.2006	L	nima	705002061000548	
375	mati	16.02.2005	L	191100000749000	191001050037505	

Genotipski podatki:

IME	vhl20	htg4	aht4	hms7	htg6	aht5	hms6	asb23	asb2	htg10	htg7	hms3	hms2	asb17	lex3	hms1	ca425
366	NR	KM	HM	LN	OO	JJ	LN	IU	IK	OR	KN	MM	HJ	NN	KP	MN	NN
548	MR	KK	HM	LN	OO	JO	LP	UU	KQ	OR	NO	MM	JM	NO	P	IM	NN
375	NP	LM	HJ	LN	IO	JM	LN	IU	IK	RT	KK	HM	HL	NR	KK	JN	NN

Ugotovitve:

Rezultati ne spodbijajo navedenega starševstva.

Rezultati izključujejo žrebec z imenom _____ kot možnega očeta.

Rezultati ne izključujejo žrebec z imenom 548 kot možnega očeta.

Genotipi zgoraj navedenih konj so shranjeni v podatkovni zbirki genetskega laboratorija na Oddelku za zootehniko.

Laboratorij ne razpolaga z genotipi enega ali obeh staršev.

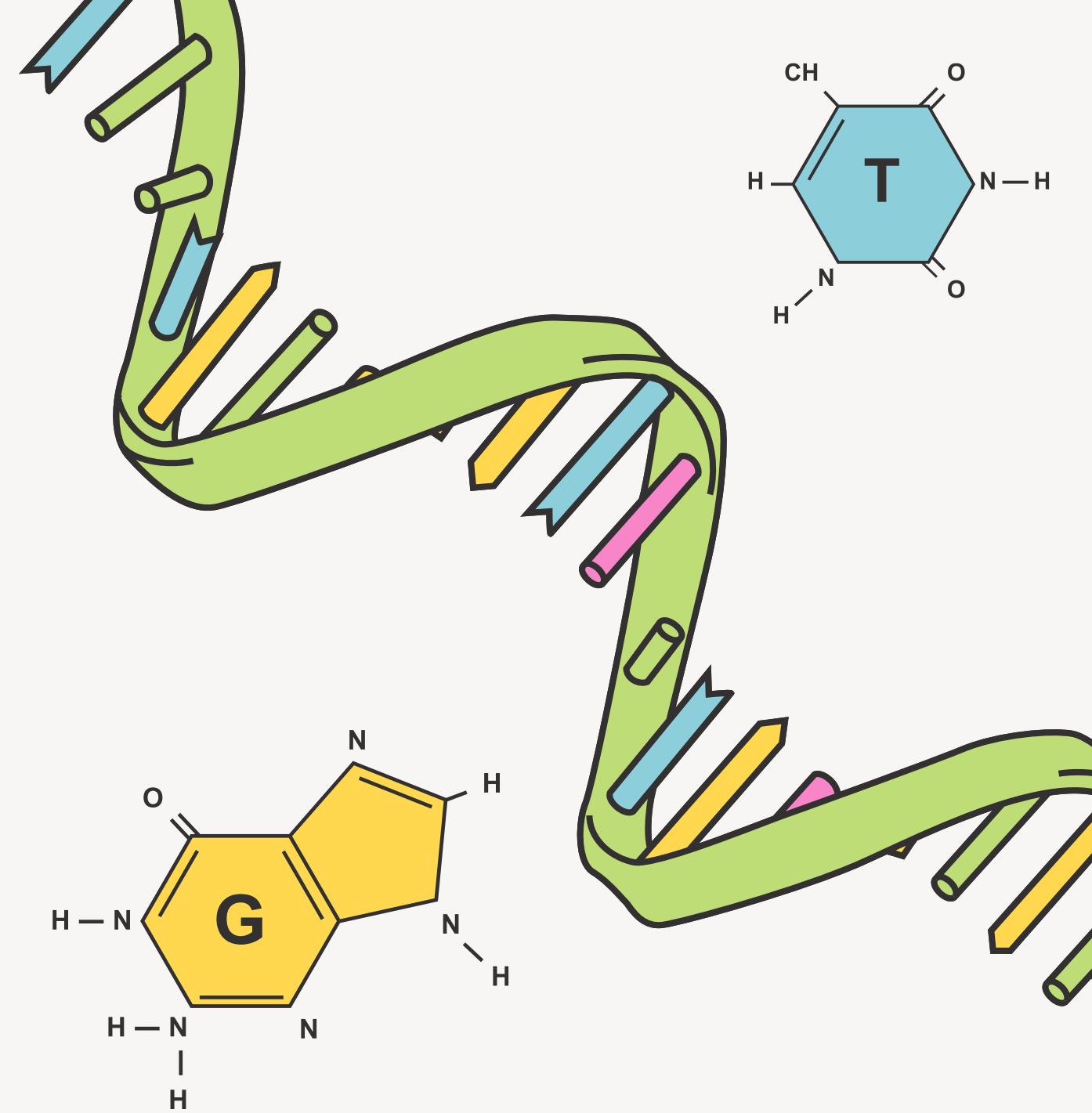
Žig in podpis:




Katedra za genetiko, animalno biotehnologijo in imunologijo, Oddelek za zootehniko
Grobje 3, SI-1230 Domžale, tel.: +386 1 3203 836, e-mail: peter.dovc@bf.uni-lj.si

GENOTYPE READING

- a specific form appears at individual loci
- many genotypes are read for accurate determination and ignoring false positives



WE DID THIS LAST YEAR, SO JUST TO REMEMBER

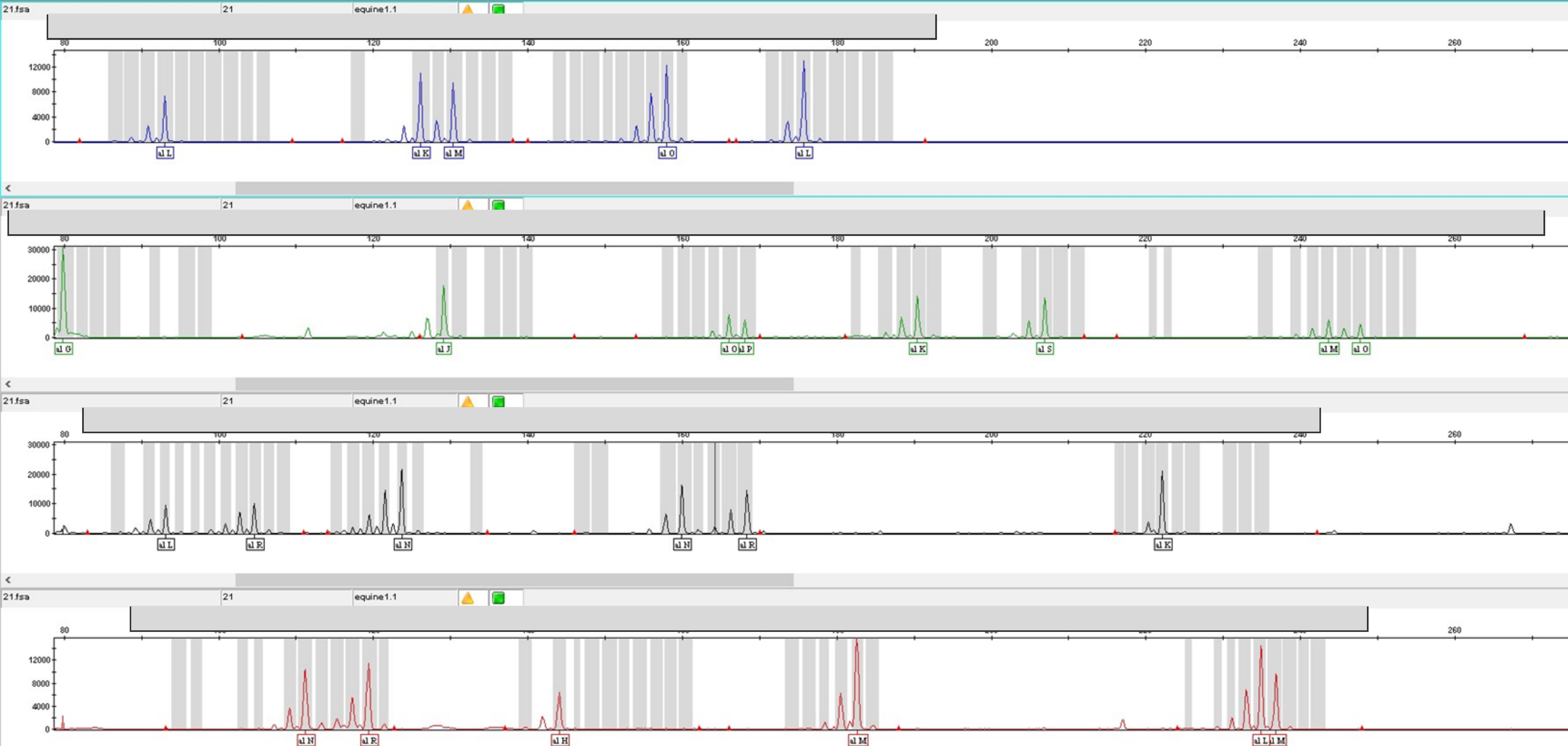


Samples Plot

File Edit View Tools Alleles Help

Plot Setting: Microsatellite Default | Panes: 4

Sample File	Sample Name	Panel	OS	SQ
-------------	-------------	-------	----	----



HOW ARE INITIAL OLIGONUCLEOTIDES LABELLED?

Colour badges

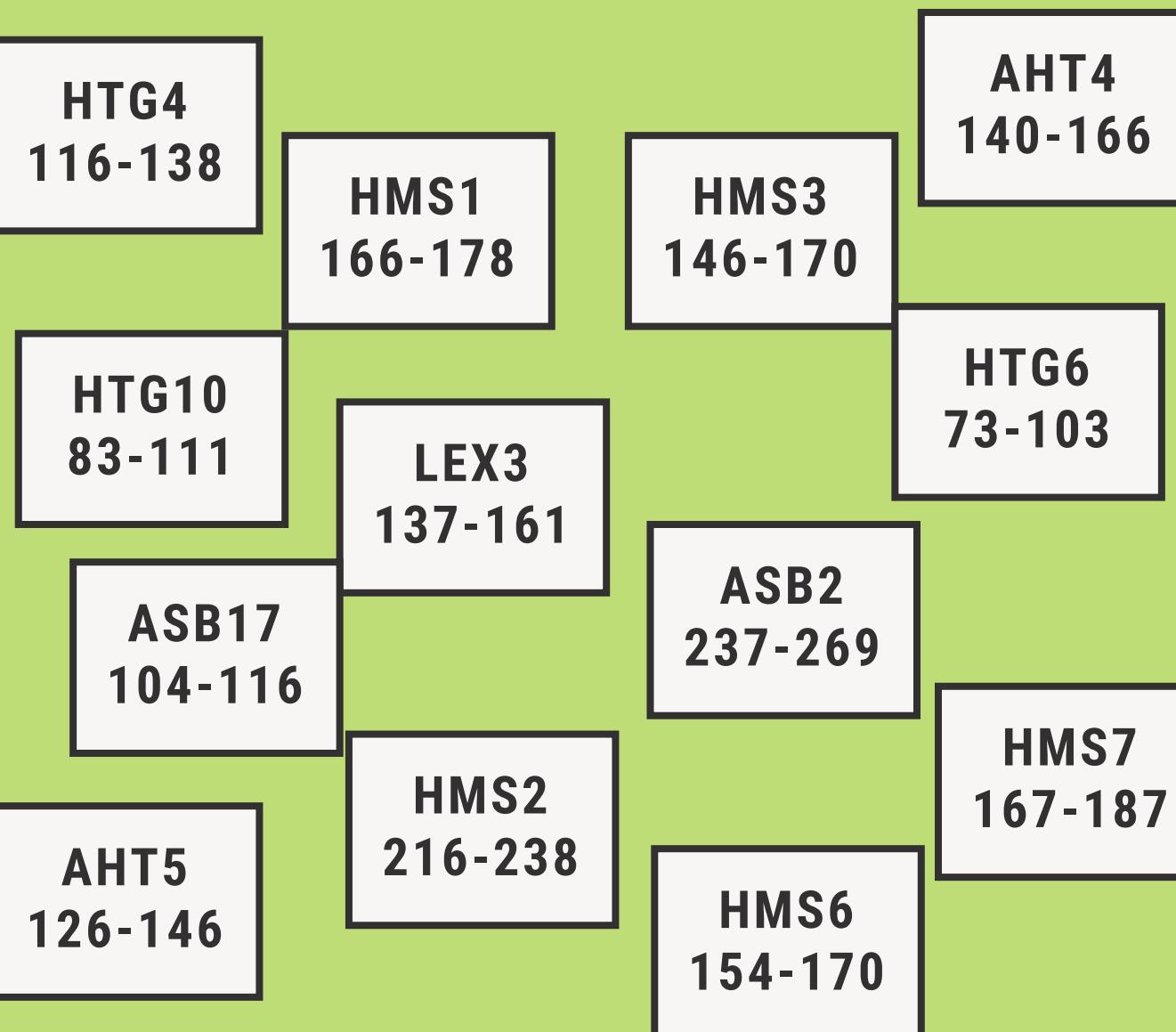
- 4 color codes
- distribution of loci according to the size of multiples



FAM HEX TAMRA ROX

Locus name	Chromosome	Repeat motif	Size range (bp)
VHL20	30	di	82–102
HTG4	9	di	116–138
AHT4	24	di	140–166
HMS7	1	di	167–187
HTG6	15	di	73–103
AHT5	8	di	126–146
HMS6	4	di	154–170
ASB23	3	di	176–212
ASB2	15	di	237–269
HTG10	21	di	83–111
HTG7	4	di	114–128
HMS3	9	di	146–170
HMS2	10	di	216–238
ASB17	2	di	104–116
LEX3	X	di	137–161
HMS1	15	di	166–178
CA425	28	di	224–248

ARRANGE THE LOCUSES AMONG THE COLOUR CODES



ARRANGE THE LOCUSES
AMONG THE COLOUR CODES

HTG4
116-138

HMS1
166-178

HTG10
83-111

HMS3
146-170

LEX3
137-161

ASB2
237-269

AHT4
140-166

HTG6
73-103

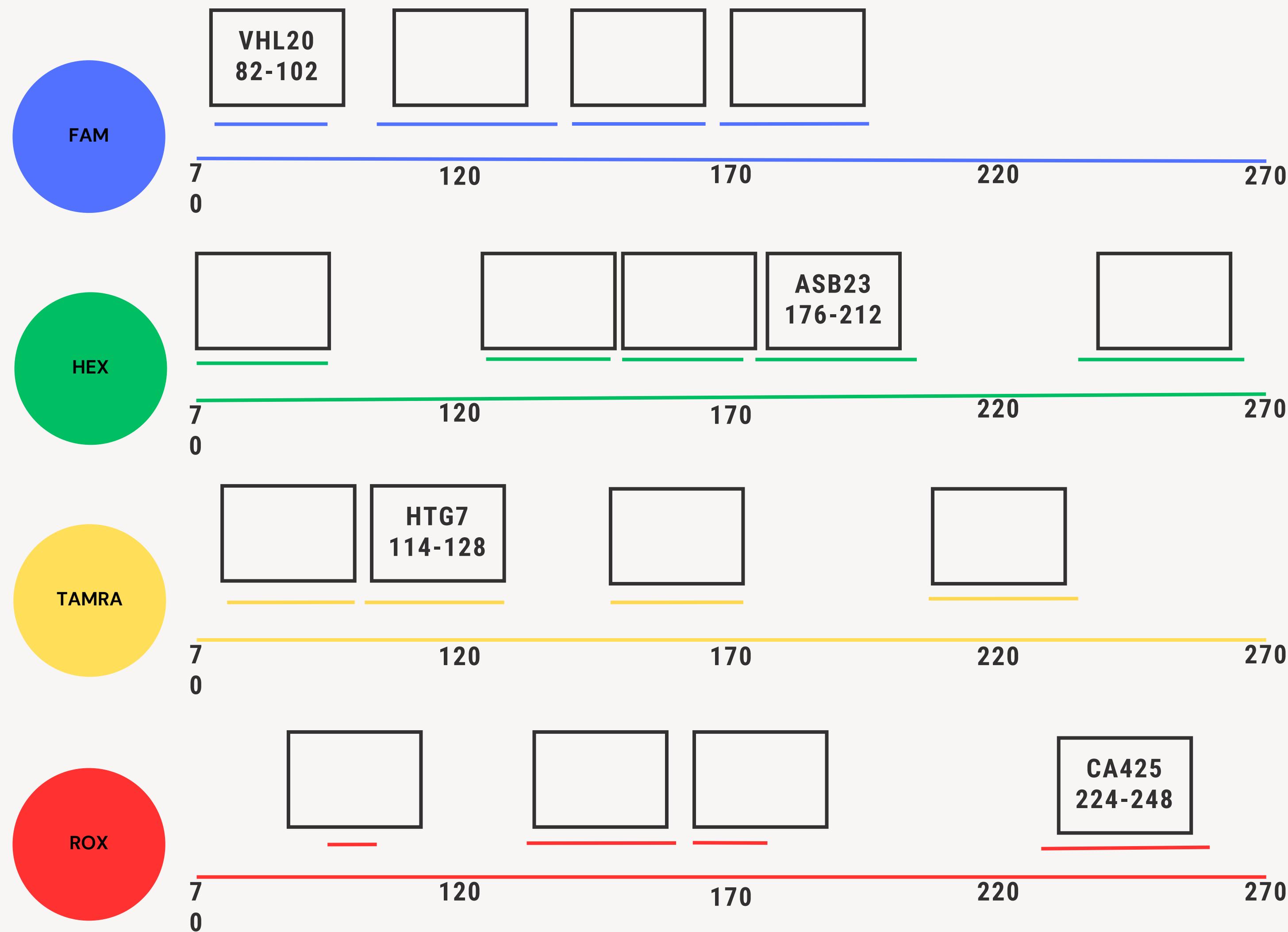
ASB17
104-116

HMS7
167-187

HMS2
216-238

HMS6
154-170

AHT5
126-146

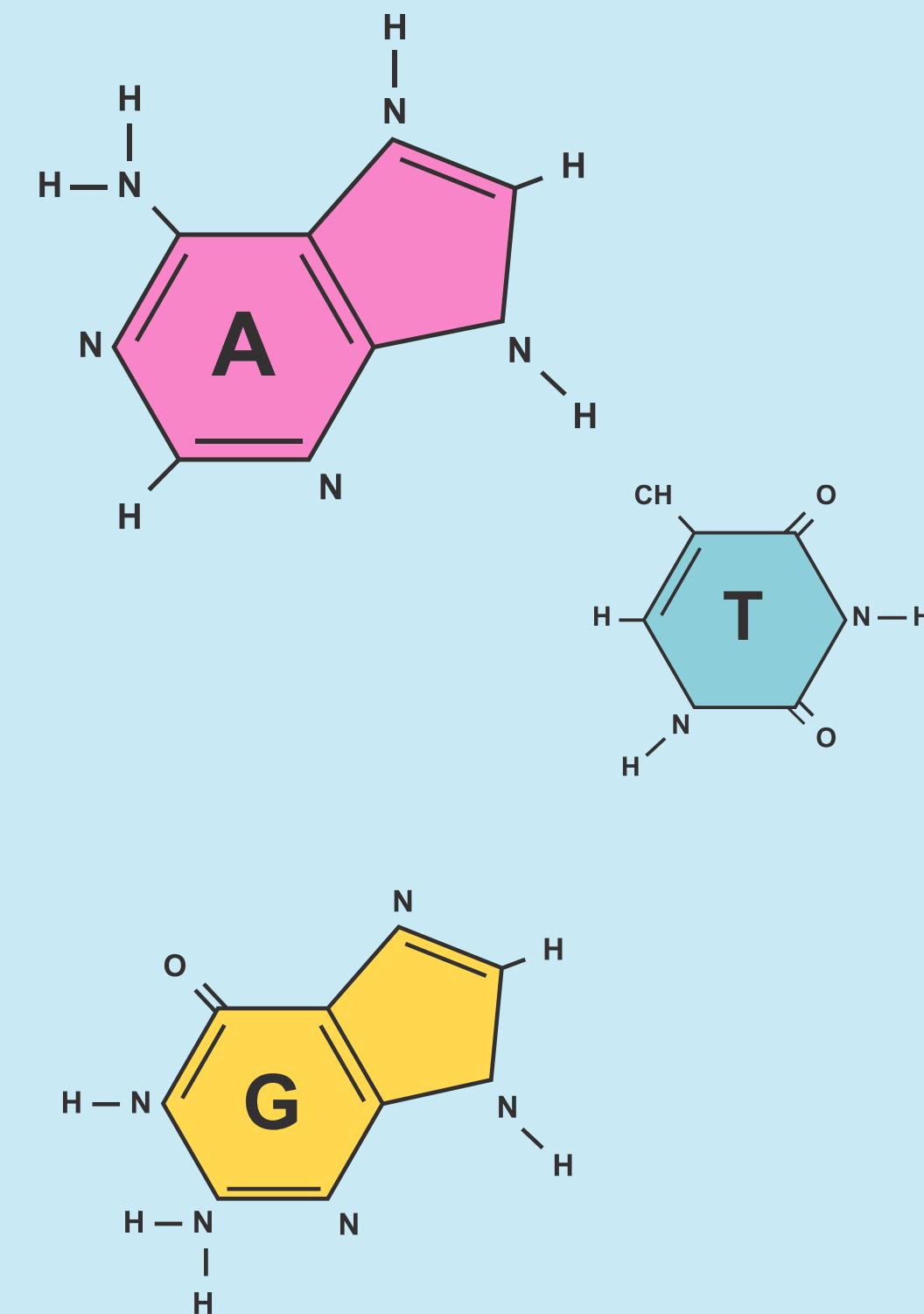


MICROSATELLITE ANALYSIS

PREPARE PROGRAM AND DATA

- download
 - Mikrosateli.str
 - Mikrosateli.xlsx
 - PCA_students.R
- Do you have R-studio?

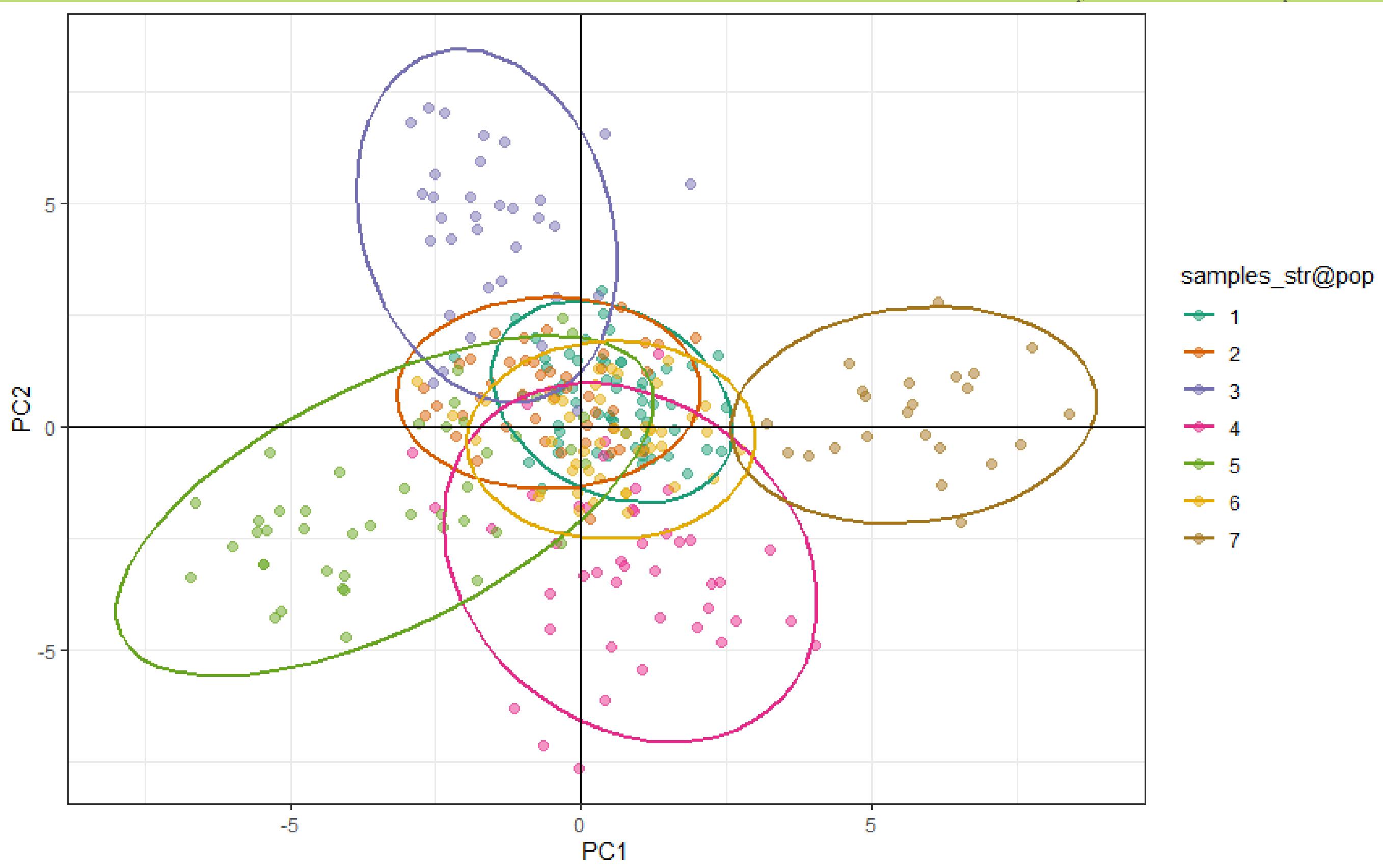
OPEN PCA_STUDENTS.R



COMPLETE THE SCRIPT

- Set working directory - write your path to data
 - `setwd("/Users/tferme/Documents/Vaje s študenti/Analiza MS")` - my path
- Upload data, write the number of individuals, the number of loci
 - `samples_str <-read.structure ("Mikrosateliti.str", n.ind=000, n.loc=00, col.lab=1, col.pop=2, col.others = NULL, row.marknames = 1, NA.char = "-9", ask = TRUE, quiet = FALSE)`

WHAT DO WE GET?



PCA

- Principal Component Analysis. It's a statistical technique used to reduce the dimensionality of data while retaining most of the important variation (information) in the dataset.

