

Esercitazione N° 1: La miostatina

Piero Rivoira
Istituto Agrario Penna – Asti
piero.rivoira@yahoo.it

Creiamo il nostro profilo su Ubuntu

questo è un commento!

lanciamo il terminale per aggiungere un nuovo profilo

ctrl+alt+t

\$ **sudo adduser nome_battesimo -uid 663**

\$ pw di labx

per eliminare un profilo utente (in caso di errore)

\$ **sudo deluser --remove-home nome_utente**

\$ pw di labx

per recuperare la passwd dimenticata

sostituire <user name> con il proprio nome utente

\$ **sudo passwd user name**

\$ pw di labx

\$ nuova pw

Creiamo il nostro profilo su Ubuntu

acquisiamo i privilegi dell'amministratore di sistema modificando il file di configurazione del Sistema Operativo (SO) /etc/sudoers

\$ **sudo visudo**

questo comando lancia l'editor di testo **nano** per creare ed aprire il file /etc/sudoers.tmp in modalità scrittura (per poterlo modificare); tale file è una copia di backup di /etc/sudoers

GNU nano 7.2 /etc/sudoers.tmp

Creiamo il nostro profilo su Ubuntu

```
# This file MUST be edited with the 'visudo' command as root.
#
# Please consider adding local content in /etc/sudoers.d/ instead of
# directly modifying this file.
#
# See the man page for details on how to write a sudoers file.
#
Defaults        env_reset
Defaults        mail_badpass
Defaults        secure_path="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/snap/bin"

# This fixes CVE-2005-4890 and possibly breaks some versions of kdesu
# (#1011624, https://bugs.kde.org/show_bug.cgi?id=452532)
Defaults        use_pty

# This preserves proxy settings from user environments of root
# equivalent users (group sudo)
#Defaults:%sudo env_keep += "http_proxy https_proxy ftp_proxy all_proxy no_proxy"

# This allows running arbitrary commands, but so does ALL, and it means
# different sudoers have their choice of editor respected.
#Defaults:%sudo env_keep += "EDITOR"

# Completely harmless preservation of a user preference.
#Defaults:%sudo env_keep += "GREP_COLOR"

# While you shouldn't normally run git as root, you need to with etckeeper
#Defaults:%sudo env_keep += "GIT_AUTHOR_* GIT_COMMITTER_"

# Per-user preferences; root won't have sensible values for them.
#Defaults:%sudo env_keep += "EMAIL DEBEMAIL DEBFULLNAME"

# "sudo scp" or "sudo rsync" should be able to use your SSH agent.
#Defaults:%sudo env_keep += "SSH_AGENT_PID SSH_AUTH_SOCK"

# Ditto for GPG agent
#Defaults:%sudo env_keep += "GPG_AGENT_INFO"

# Host alias specification

# User alias specification
```

Creiamo il nostro profilo su Ubuntu

GNU nano 7.2 /etc/sudoers.tmp

```

# This file specifies the proxy settings that is set in environment variables for the
# equivalent users (group sudo)
Defaults:%sudo env_keep += "http_proxy https_proxy ftp_proxy all_proxy no_proxy"

# This allows running arbitrary commands, but so does ALL, and it means
# different sudoers have their choice of editor respected.
Defaults:%sudo env_keep += "EDITOR"

# Completely harmless preservation of a user preference.
Defaults:%sudo env_keep += "GREP_COLOR"

# While you shouldn't normally run git as root, you need to with etckeeper
Defaults:%sudo env_keep += "GIT_AUTHOR_* GIT_COMMITTER_"

# Per-user preferences; root won't have sensible values for them.
Defaults:%sudo env_keep += "EMAIL DEBEMAIL DEBFULLNAME"

# "sudo scp" or "sudo rsync" should be able to use your SSH agent.
Defaults:%sudo env_keep += "SSH_AGENT_PID SSH_AUTH_SOCK"

# Ditto for GPG agent
Defaults:%sudo env_keep += "GPG_AGENT_INFO"

# Host alias specification

# User alias specification

# Cmnd alias specification

# User privilege specification
root    ALL=(ALL:ALL) ALL
piero   ALL=(ALL:ALL) ALL

# Members of the admin group may gain root privileges
%admin  ALL=(ALL) ALL

# Allow members of group sudo to execute any command
%sudo   ALL=(ALL:ALL) ALL

# See sudoers(5) for more information on "@include" directives:

@includedir /etc/sudoers.d

```

Elenco degli amministratori di sistema

portarsi con il cursore sull'ultima riga della lista

alt-6 # copia

ctrl-u # incolla l'intera riga

inserire il proprio nome utente

ctrl-o # per salvare

cancellare l'estensione .tmp

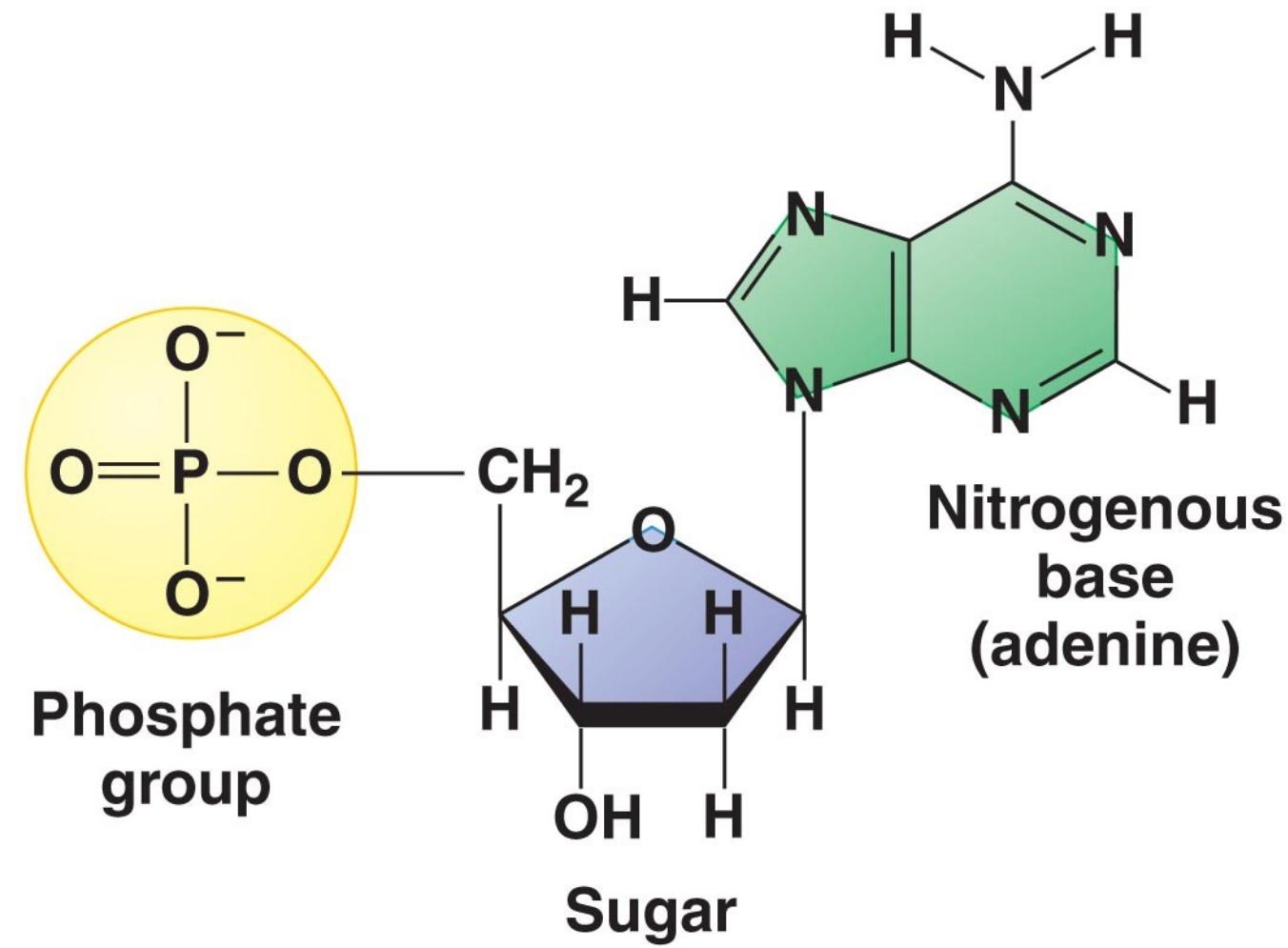
ctrl-x # per chiudere nano

alt-u # in caso di errore
(per annullare l'ultimo comando inserito)

Elenco degli amministratori di sistema

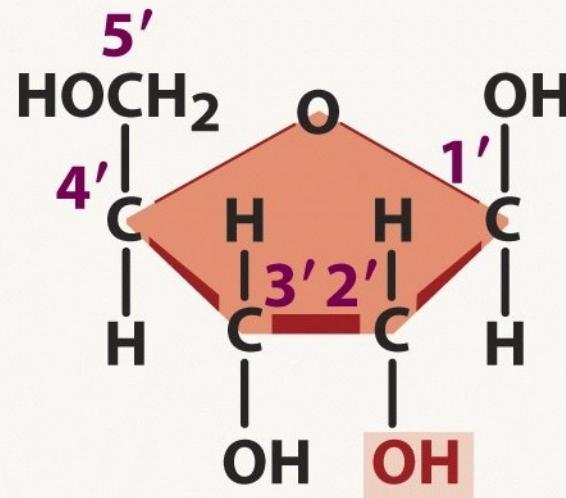
▲ Guida □ Salva □ Cerca □ Esegui □ Posizione □ Annnulla
■ Esci □ Inserisci □ Sostituisci □ Taolia □ Giustifica □ Vai a riga □ Ripeti □ Contrassegna □ Parentesi
■ Precedente □ Successiva □ Avanti

Struttura di un nucleotide

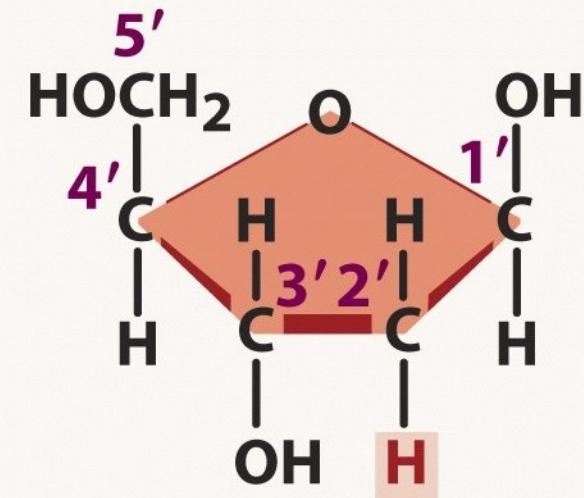


Zuccheri

Sugars



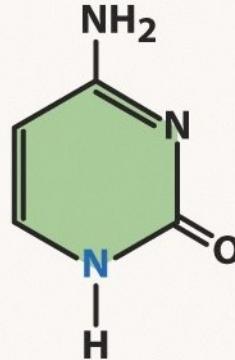
Ribose



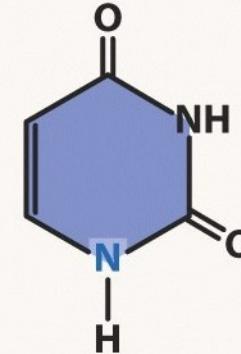
Deoxyribose

Basi pirimidiniche e puriniche

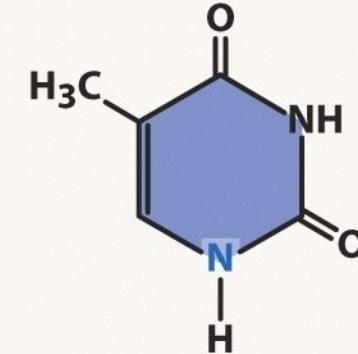
Nitrogen-containing bases



Cytosine (C)

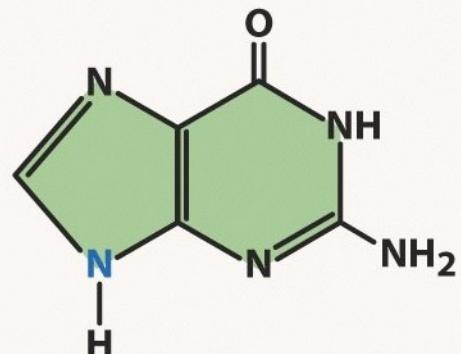


Uracil (U)

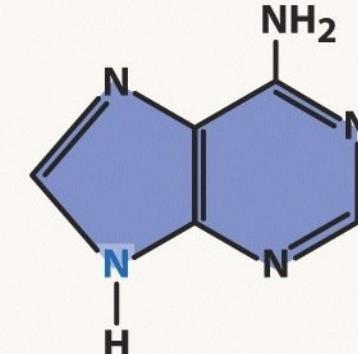


Thymine (T)

Pyrimidines



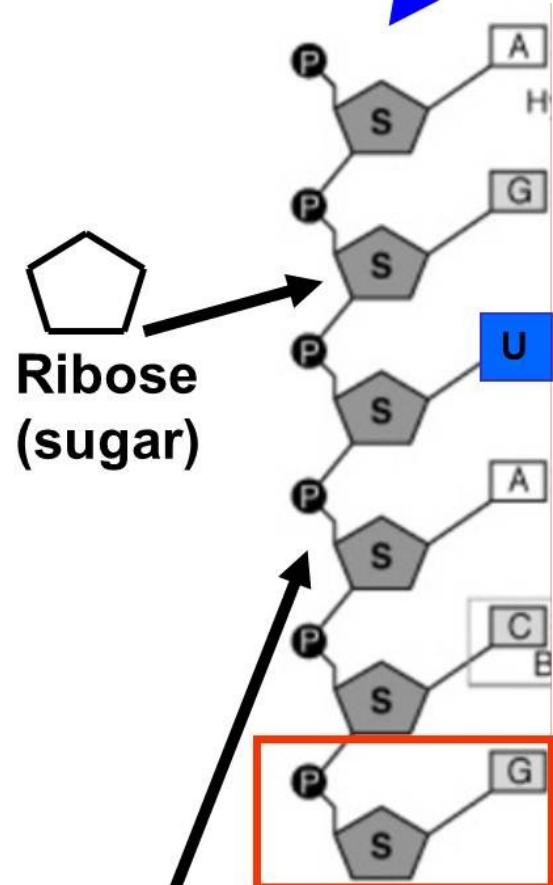
Guanine (G)



Adenine (A)

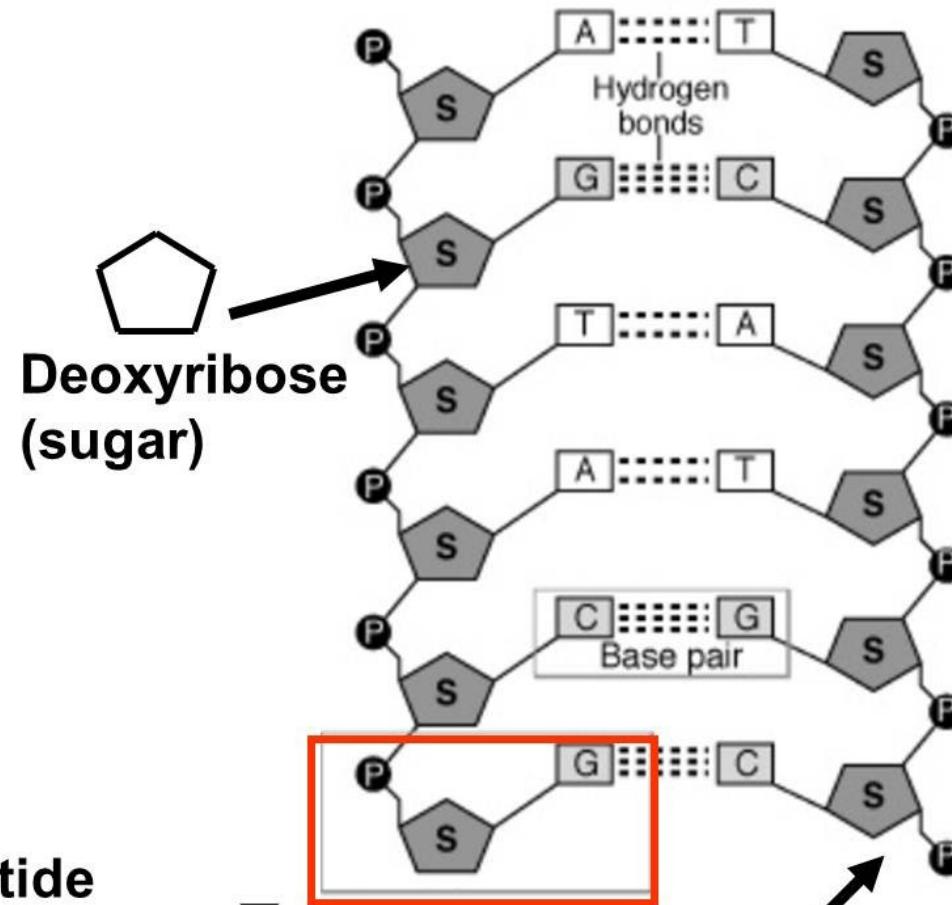
Purines

RNA
&
DNA
are chains of nucleotides



Ribose
(sugar)

Sugar-phosphate
backbone



Deoxyribose
(sugar)

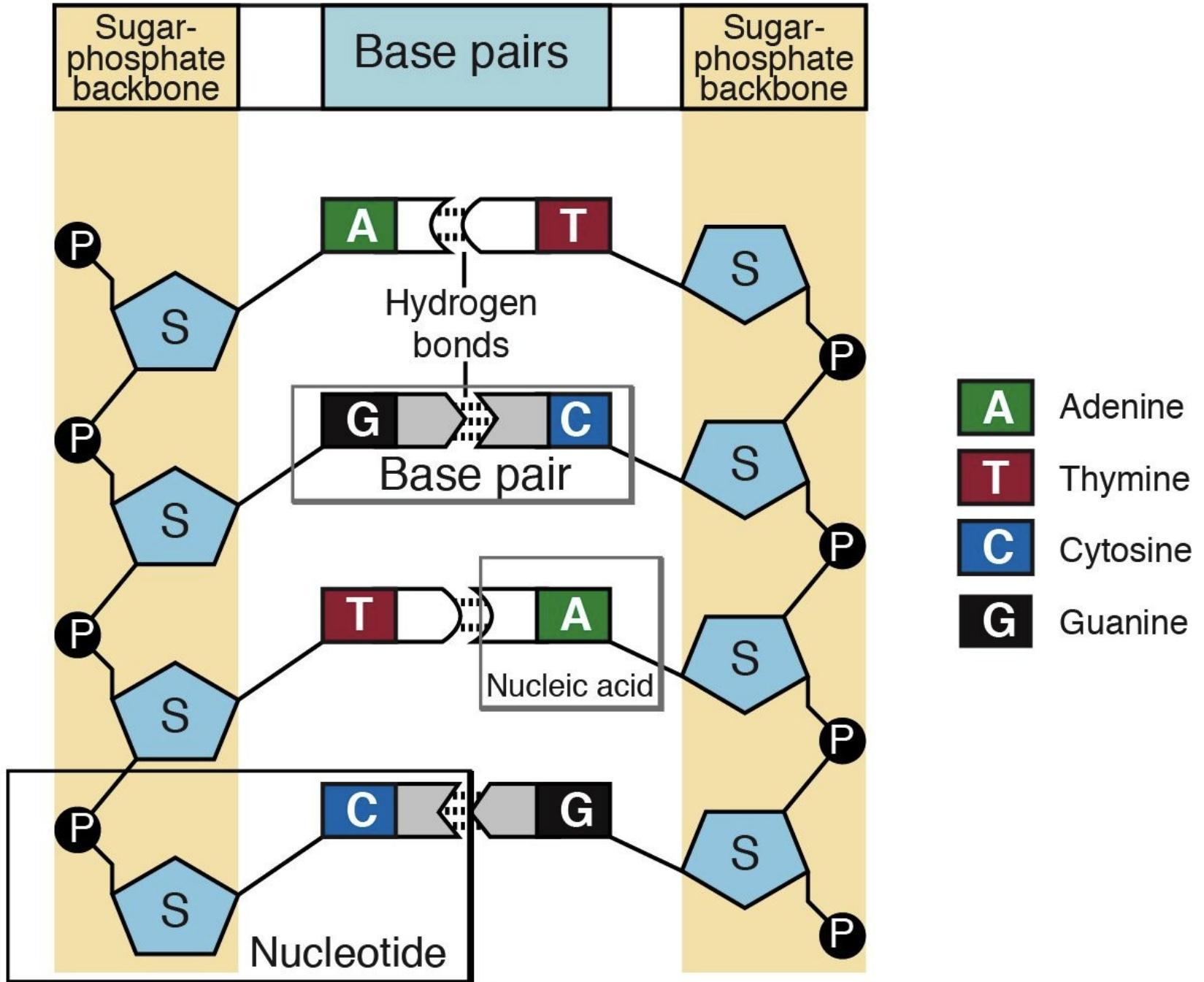
Nucleotide
Unit

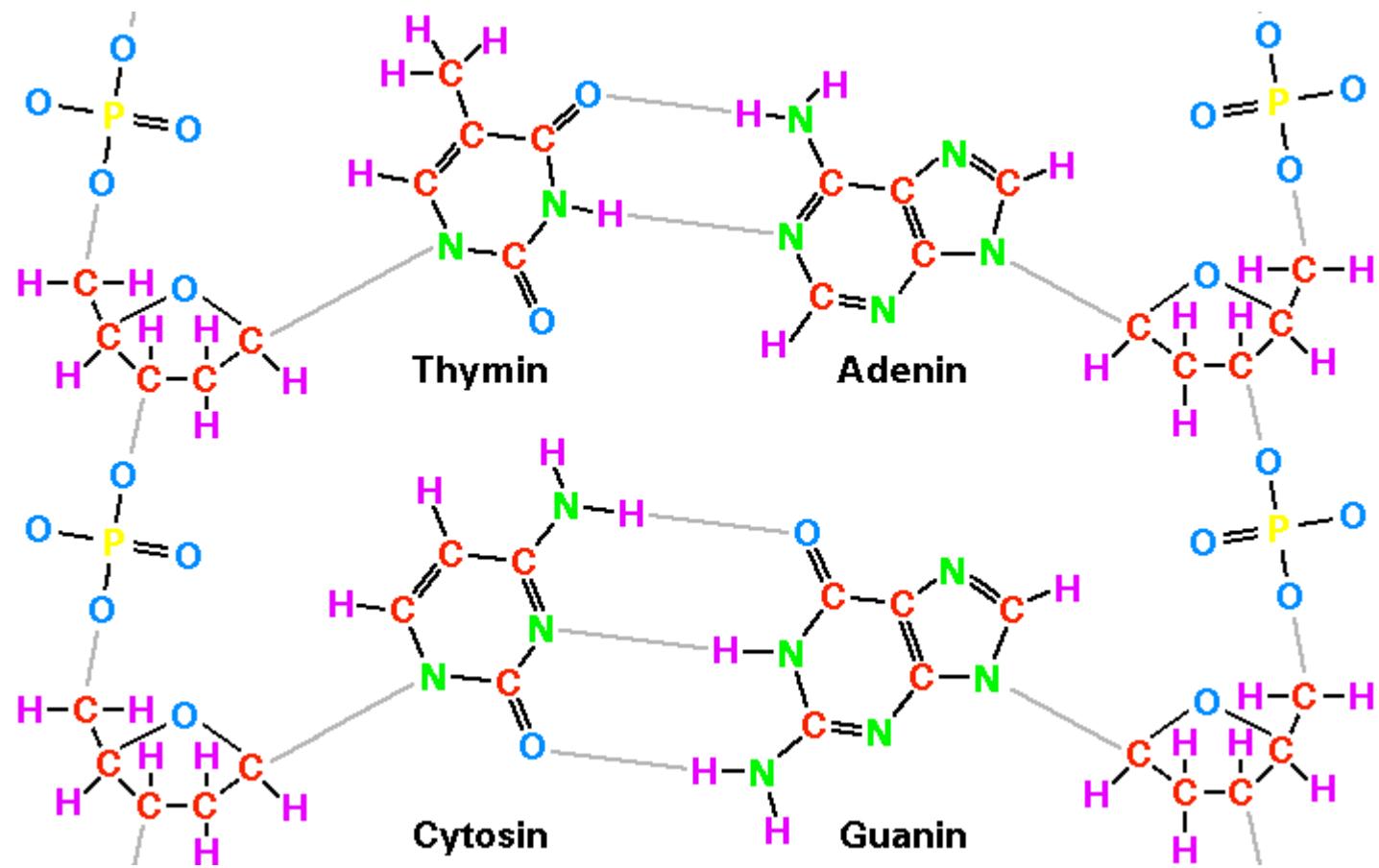
Base pair

Hydrogen
bonds

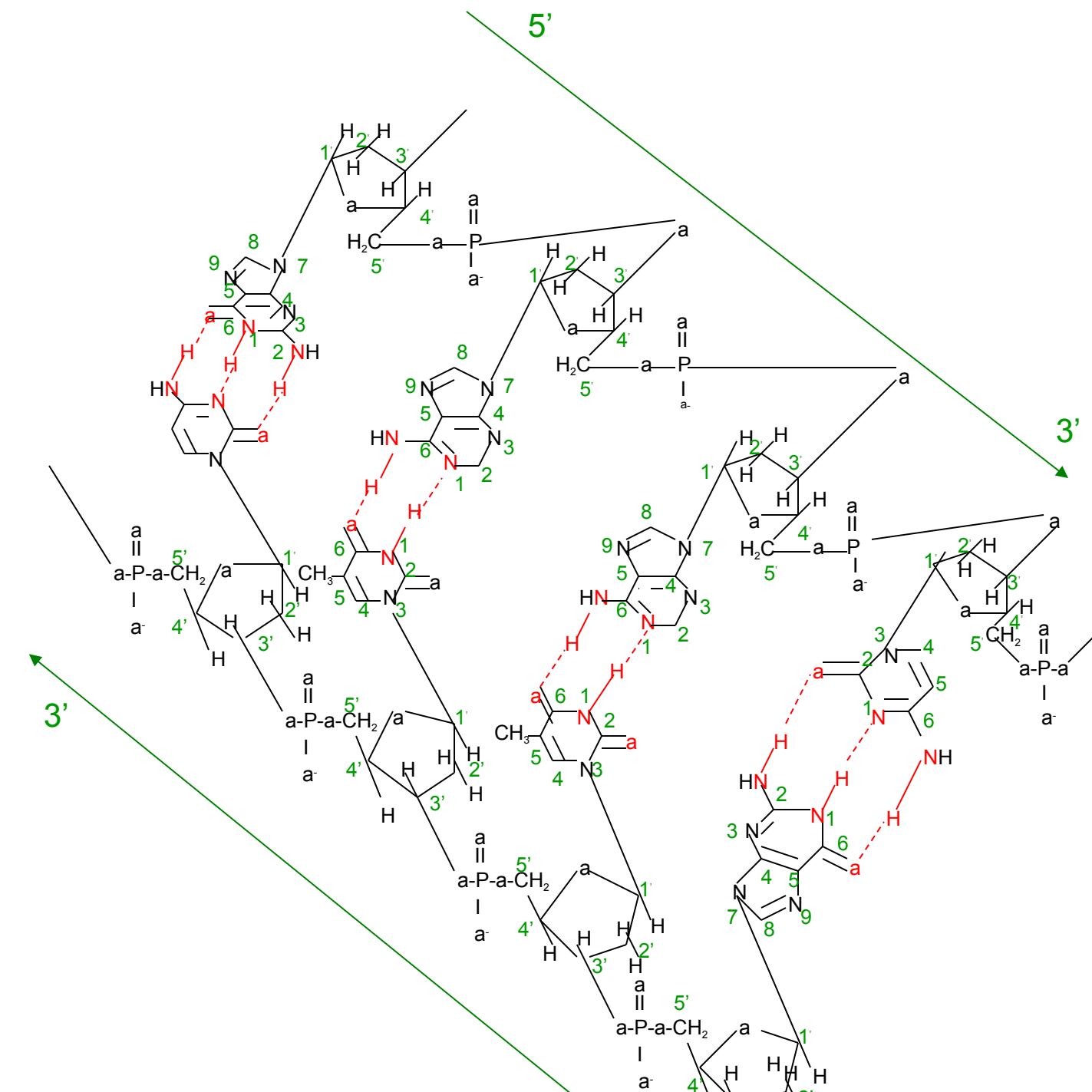
Sugar-phosphate
backbone

Deoxyribonucleic Acid (DNA)

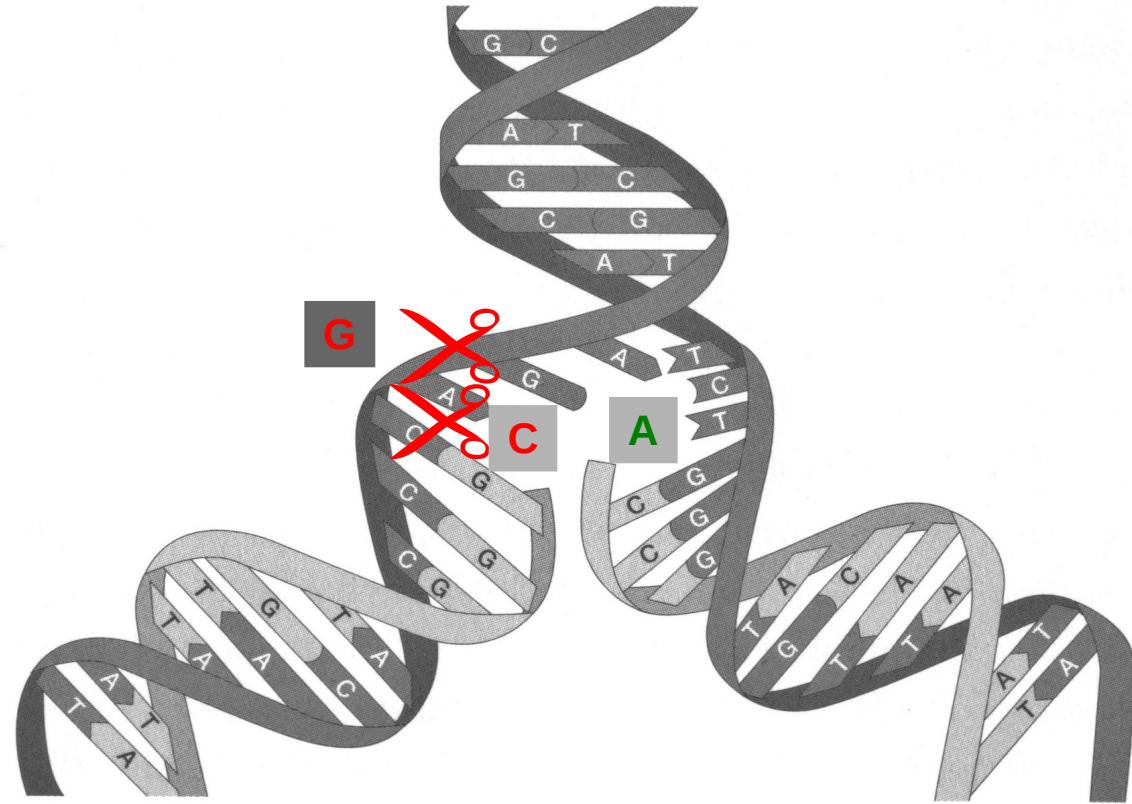




antiparallelismo



Tratto di DNA formato da 4 paia di basi (CTTG nella catena inferiore, GaaC nella catena antiparallela ad essa). Le frecce verdi indicano la direzione dei legami fra i nucleotidi, che è opposta nelle due catene.



L'incrocio industriale o F1 toro da carne X vacca da latte

- **Razza Belga** = razza incrociante più utilizzata in Piemonte, allevata in purezza solo in Belgio

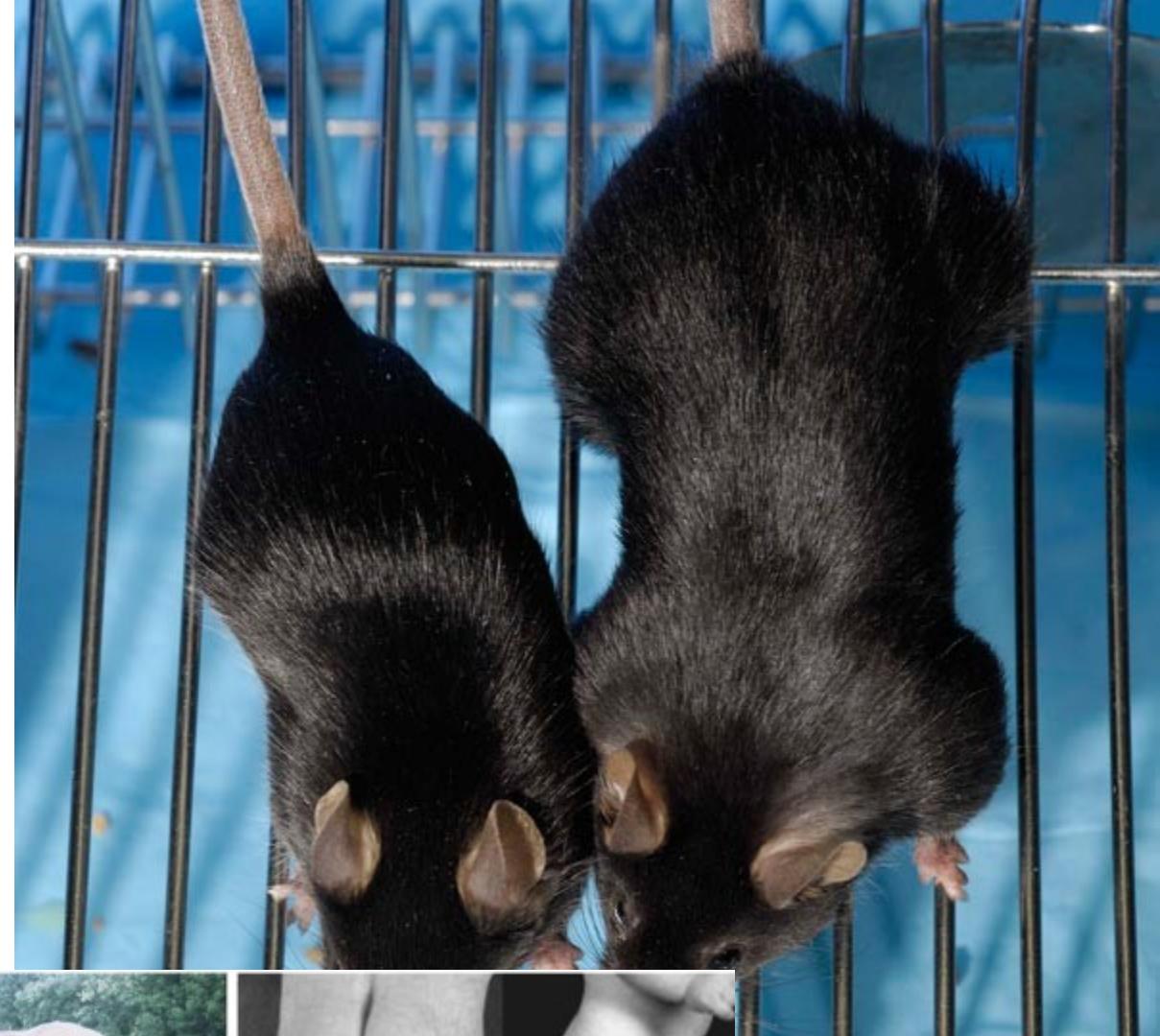


- Vitelli ibridi spuntano prezzi \leq 400 Euro
[Piemontese puro \sim 500 E, frisone puro (*gaiot*) 50 E]
- Quante bovine da latte posso inseminare con il Belga senza compromettere la **rimonta interna**?

La razza Belga e la miostatina

Phenotype of myostatin null vertebrates.

- a) upper forelimb muscles of wild-type mouse;
- B) upper forelimb of mouse that is myostatin null;
- C) bovine with myostatin mutation;
- D) and E) human with myostatin mutation as a neonate (D) and at 7 mo of age.



La razza Belga e la miostatina

A screenshot of a web browser showing the Google Scholar homepage. The URL in the address bar is https://scholar.google.com. The page features the Google Scholar logo at the top center. Below it is a search bar containing the query "Double muscling in cattle due to mutations in the myostatin gene". To the right of the search bar is a blue search button with a magnifying glass icon. Underneath the search bar are two radio button options: "Qualsiasi lingua" (selected) and "Pagine in Italiano". At the bottom of the page, there is a green link labeled "Sali sulle spalle dei giganti". The browser interface includes standard navigation buttons (back, forward, refresh), a zoom level indicator (133%), and various icons for saving and sharing.

Double muscling in cattle due to mutations in the myostatin gene

Qualsiasi lingua Pagine in Italiano

Sali sulle spalle dei giganti



RESEARCH ARTICLE | GENETICS |

f X in



Double muscling in cattle due to mutations in the myostatin gene

Alexandra C. McPherron and Se-Jin Lee [Authors Info & Affiliations](#)

November 11, 1997 | 94 (23) 12457-12461 | <https://doi.org/10.1073/pnas.94.23.12457>

La razza Belga e la miostatina

the GenBank dat... 1

GenBank

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 12457–12461, November 1997
Genetics

Check for updates

Double muscling in cattle due to mutations in the myostatin gene

ALEXANDRA C. MCPHERRON AND SE-JIN LEE*

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205

Communicated by Daniel Nathans, Johns Hopkins University School of Medicine, Baltimore, MD, August 26, 1997 (received for review August 12, 1997)

La razza Belga e la miostatina

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [baboon (accession no. AF019619), bovine (accession no. AF019620), chicken (accession no. AF019621), ovine (accession no. AF019622), porcine (accession no. AF019623), rat (accession no. AF019624), turkey (accession no. AF019625), zebrafish (accession no. AF019626), and human (accession no. AF019627)].

A commentary on this article begins on page 12249.

*To whom reprint requests should be addressed. e-mail: sejin_lee@qmail.bs.jhu.edu.

La razza Belga e la miostatina

lanciamo il terminale per usare R

ctrl+alt+t

\$ R

```
> install.packages("ape")
> library(ape)
> x = paste("AF0196", 19:27, sep = "")
```

ctrl+MAIUSC+v per incollare nel terminale o in R

```
> myostatin.numbers = c('NM_001001525.3', 'XM_010862903.1', 'XM_005907427.1', 'GQ184147.1', 'AF019620.1',
  'XM_061432883.1', 'BC134563.1', 'AY160688.1', 'XM_027555492.1', 'XM_019969486.1', 'AF019761.1',
  'NM_001290967.1', 'XM_055574715.1', 'KP057613.1', 'AY629304.1', 'XM_055397770.1', 'XM_061135030.1',
  'XM_043894881.1', 'XM_043488279.1', 'AY629307.1', 'XM_020900528.1', 'XM_052635569.1', 'EF629535.1',
  'NM_001009428.3', 'XM_040227852.1', 'JN662463.1', 'AM992883.1', 'AY629305.1', 'AF019622.1', 'KM371731.1',
  'JX840482.1', 'JN639006.1', 'HM462261.1', 'HM462259.1', 'NM_001285737.1', 'FM207636.1', 'AY629309.1',
  'AY629306.1', 'AY436347.1', 'AM992884.1', 'GU377303.1', 'XM_007104442.2', 'XM_022568276.1', 'XM_061203308.1',
  'XM_057551429.1', 'XM_036859085.1', 'AY629308.1', 'XM_029225345.1', 'XM_027114903.1', 'XM_024733638.1',
  'XM_004323907.2', 'XM_030847723.2', 'XM_060151388.1', 'XM_060015612.1', 'XM_060106598.1', 'XM_004276934.2',
  'XM_032637329.1', 'XM_059051852.1', 'XM_059928344.1', 'XM_057745280.1', 'XM_047770464.1', 'HM241657.1',
  'AF019623.1', 'JN630464.1', 'EF612791.1', 'NM_214435.2', 'XM_006209983.3', 'XM_006189223.3', 'XM_028125258.1',
  'XM_027590490.1', 'XM_025887402.1', 'XM_004426733.2', 'XM_010960036.2', 'XM_021703117.1', 'XM_004392403.2',
  'XM_046659551.1', 'XM_046650639.1', 'XM_014837576.2', 'XM_032417983.1', 'XM_010991955.2', 'XM_058548651.1',
  'XM_008516515.1', 'XM_036075506.1', 'NM_001081817.1', 'XM_006726961.2', 'XM_019626478.1', 'XM_054562700.1',
  'XM_036253439.1', 'XM_019740713.1', 'MZ169554.1', 'XM_035007537.1', 'XM_026490450.4', 'XM_045868351.2',
  'XM_045806119.1', 'XM_008687262.2', 'XM_029932956.1', 'XM_016215509.1', 'XM_036881965.2', 'XM_047086026.1',
  'XM_045480514.1')
```

La razza Belga e la miostatina

```
> x = paste("AF0196", 19:27, sep = "")  
> x = x[-c(1:8)]  
> x  
[1] "AF019627"  
> x = x[9]  
> x  
[1] "AF019627"  
>  
# perché l'output di questi due comandi è il medesimo?
```

La razza Belga e la miostatina

```
> y = c(x, myostatin.numbers)
> myostatin.seq = read.GenBank(x, as.character = TRUE)
> install.packages("seqinr")
> library(seqinr)

# nel terminale
$ mkdir DNA_SEQUENCES
$ ls
$ pwd

# in R
# alt gr+` → ~
> setwd("~/DNA_SEQUENCES")
> getwd()
> write.fasta(sequences = myostatin.seq, names = names(myostatin.seq), nbchar
=80, file.out = "myostatin.fas")

# nel terminale
$ cd DNA_SEQUENCES
$ gedit myostatin.fas
```

La razza Belga e la miostatina

dopo

>AF019620

Aggiungere

>bovine_belgian

atgcaaaaactgcaaatctctgttatatttacctatttatgctgattgttgcggccaggatggatctgaatgagaacag
cgagcagaaggaaaatgtggaaaaagaggggctgtgtaatgcattgcattgtggaggaaaacactacatcctcaagactag
aagccataaaaatccaaatcctcagtaaacttcgcctggaaacagtcctaaacatcagcaaagatgctatcagacaactt
ttgccaaggctccactcctggaaactgattgatcagttcgatgtccagagagatgccagcgtgacggctccttgg
agacgatgactaccacgccaggacggaaacaggcttacatgcggcggatctgtatcttctaaccgcaagtggaaaggaa
aacccaaatgtgtttaaatttagtctaagatacaataacttagtaaaggccaaactgtggatatatctg
aggcctgtcaagactcctgcacagtgtttgtcaaattcctgagactcatcaaaccatgaaagacggtacaaggatatac
tggaatccgatctctgaaacttgacatgaaaccaggcactggatttgcagacgttgcataagatggccatgttgcataacc
actggctcaaacaacctgaatccaacttaggcattgttgcataagatggccatgttgcataacc
ttcccagaaccaggagaagatggactgactccttttagaagtcaaggtacacagacacacacaaaagatctaggagaga
tttggcttgcattgtgacagaatctcgatgtcgatgttgcattgttgcataactgtggattttgaagctttggatggattgg
attattgcacccaaaagatataaggccaaattactgtctggagaatgtgaatttgttgcataactgttgcataacc
ccatcttgcaccaagccaaacccagggttcagccggccctgtgtactcctacaaagatgtctccaaattaatatgc
tatatttaatggcgaaggacaaataatcggaaagattccagccatggtagtagatcgctgtgggtgttgcata
>bovine_piedmontese

atgcaaaaactgcaaatctctgttatatttacctatttatgctgattgttgcggccaggatggatctgaatgagaacag
cgagcagaaggaaaatgtggaaaaagaggggctgtgtaatgcattgcattgtggaggaaaacactacatcctcaagactag
aagccataaaaatccaaatcctcagtaaacttcgcctggaaacagtcctaaacatcagcaaagatgctatcagacaactt
ttgccaaggctccactcctggaaactgattgatcagttcgatgtccagagagatgccagcgtgacggctccttgg
agacgatgactaccacgccaggacggaaacaggcttacatgcggcggatctgtatcttctaaccgcaagtggaaaggaa
aacccaaatgtgtttaaatttagtctaagatacaataacttagtaaaggccaaactgtggatatatctg
aggcctgtcaagactcctgcacagtgtttgtcaaattcctgagactcatcaaaccatgaaagacggtacaaggatatac
tggaatccgatctctgaaacttgacatgaaaccaggcactggatttgcagacgttgcataagatggccatgttgcataacc
actggctcaaacaacctgaatccaacttaggcattgttgcataagatggccaaattactgtctggagaatgttgcataacc
ttcccagaaccaggagaagatggactgactccttttagaagtcaaggtacacagacacacacaaaagatctaggagaga
tttggcttgcattgtgacactccacagaatctcgatgtcgatgttgcattgttgcataactgtggattttgaagctttg
gatgggattggattattgcacccaaaagatataaggccaaattactgtctggagaatgttgcataacc
tatcctcataccatcttgcaccaagccaaacccagggttcagccggccctgtgtactcctacaaagatgtctcc
aattaatatgtatatttaatggcgaaggacaaataatcggaaagattccagccatggtagtagatcgctgtgggt
gttgcata

La razza Belga e la miostatina

```
# ctrl+MAIUSC+s
```

```
# save as myostatin_mod.fas
```

```
$ clustalx
```

Mode: Multiple Alignment Mode Font: 10

Load Sequence File

Cancel Open

Nome Dimensione Tipo Data di modifica

Nome	Dimensione	Tipo	Data di modifica
allineamento_miostatina_blast.txt	480,1 kB	Testo	11 gen
database_completo_miostatina.txt	480,1 kB	Testo	13 gen
dexter.fasta	88,6 kB	Testo	3 set
link_Cervidae.txt	344 byte	Testo	10 gen
mcpheron-lee-1997-double-muscling-in-cattle-due-to-mutations-in-the-myostatin-ge...	537,5 kB	Documento	ven
MT_Cervidae	333,7 kB	Testo	10 gen
MT_Cervidae.aln	602,3 kB	Testo	11 gen
MT_Cervidae.dnd	627 byte	Testo	11 gen
MT_Cervidae.fasta	448,4 kB	Testo	11 gen
my_data.txt	98,9 kB	Testo	27 dic 2023
myostatin.fas	220,0 kB	Testo	10:36
myostatin.fasta	10,6 kB	Testo	Ieri
myostatin.RData	8,8 MB	R Data File	Ieri
MYOSTATIN_ACCESSION_NUMBERS.txt	1,7 kB	Testo	10:26
myostatin_all.fasta	215,1 kB	Testo	12 gen
MYOSTATIN_DATA.txt	17,3 kB	Testo	09:15
myostatin_mod.fas	222,9 kB	Testo	Ieri
myostatin_mod.fasta	740,2 kB	Testo	08:19
piero.RData	574,5 MB	R Data File	14 gen
wolves.csv	1,1 kB	Testo	25 dic 2023

All Files ▾

La razza Belga e la miostatina

National Library of Medicine
National Center for Biotechnology Information

Search NCBI

AF019620 X Search

Results found in 5 databases

NUCLEOTIDE SEQUENCE

Bos taurus myostatin (MSTN) mRNA, complete cds

Bos taurus

1,128 bp mRNA sequence

AF019620.1



[FASTA](#) [Gene](#)

[BLAST](#)

[Download](#)

FASTA

Send to:

Change region shown

Bos taurus myostatin (MSTN) mRNA, complete cds

GenBank: AF019620.1

[GenBank](#) [Graphics](#)

>AF019620.1 Bos taurus myostatin (MSTN) mRNA, complete cds

```
ATGCAAAACTGCAAATCTCTGTTATATTACCTATTATGCTGATTGTTGCTGGCCAGTGGATCTGA  
ATGAGAACAGCGAGCAGAAGGAAAATGTGGAAAAAGAGGGGCTGTATAATGCATGTTGTGGAGGGAAAA  
CACTACATCCTCAAGACTAGAACGCCATAAAATCCAATCCTCAGTAAACTTCGCCTGGAAACAGCTCCT  
AACATCAGCAAAGATGCTATCAGACAACCTTGCCCAAGGCTCCTCCACTCCTGGAACTGATTGATCAGT  
TCGATGTCCAGAGAGATGCCAGCAGTGACGGCTCCTTGGAAAGACGATGACTACCACGCCAGGACGGAAAC  
GGTCATTACCATGCCACGGAGTCTGATCTTAACGCAAGTGGAAAGGAAAACCCAAATGTTGCTTCTTT  
AAATTAGCTCTAACGATACAATACAATAAAACTAGTAAAGGCCAACTGTGGATATATCTGAGGCCTGTCA  
AGACTCCTGCGACAGTGTGCAAATCCTGAGACTCATCAAACCCATGAAAGACGGTACAAGGTATAC  
TGGAAATCCGATCTGAAACTTGACATGAACCCAGGCACTGGTATTGGCAGAGCATTGATGTGAAGACA  
GTGTTGCAGAACTGGCTCAAACACCTGAATCCAACCTTAGGCATTGAAATCAAAGCTTAGATGAGAATG  
GCCATGATCTGCTGTAACCTCCCAGAACCCAGGAGAAGATGGACTGACTCCTTTAGAAGTCAAGGT  
AACAGACACACAAAAAGATCTAGGAGAGATTTGGCTTGATTGTGATGAACACTCCACAGAACATCTGA  
TGCTGTCGTTACCCCTAACTGTGGATTTGAAGCTTTGGATGGATTGGATTATTGCACCTAAAGAT  
ATAAGGCCAATTACTGCTCTGGAGAATGTGAATTGTATTTGAAAGTATCCTCATACCCATCTGT  
GCACCAAGCAAACCCAGAGGTTAGCCGGCCCTGCTGTACTCCTACAAAGATGTCTCCAATTAATATG  
CTATATTAAATGGCGAAGGACAAATAATACGGGAAGATTCCAGCCATGGTAGTAGATCGCTGTGGGT  
GTTCATGA
```

Customize view

Analyze this sequence

[Run BLAST](#)[Pick Primers](#)

Articles about the MSTN gene

[Combined Transcriptome and Metabolome Analysis \[Int J Mol Sci. 2\]](#)[Myostatin site-directed mutation and simultaneous PPAF \[J Cell Physiol. 2\]](#)[Association between polymorphisms Myf5, MSTN and \[Anim Biotechnol. 2\]](#)

See

Reference sequence information

[RefSeq mRNA](#)

See reference mRNA sequence for the MSTN gene (NM_001001525.3)

Miostatina: allineamento multi-sequenza

```
> x = paste("AF0196", 19:27, sep = "")  
> setwd('~/DNASEQUENCES/')  
> load('piero.RData')  
> library(ape)  
> library(seqinr)  
> myostatin_x.seq = read.GenBank(x, as.character = TRUE)
```

usando la funzione `read.GenBank` del pacchetto APE (Analysis of Phylogenetics and Evolution), viene creata una lista delle sequenze nucleotidiche corrispondenti ai numeri di accesso a GenBank contenuti nel vettore `x`

GenBank Overview

What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research, 2013 Jan;41\(D1\):D36-42](#)). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#).

GenBank Resources

[GenBank Home](#)

[Submission Types](#)

[Submission Tools](#)

[Search GenBank](#)

[Update GenBank Records](#)

[Home](#)[Packages](#)[Datasets](#)[Task Views](#)[R resources](#)[Packages](#) ▾

Search...

[Home](#) > [Packages](#) > [ape](#)

ape package^{5.8-1}

Analyses of Phylogenetics and Evolution

```
install.packages("ape")
```

[Copy](#)[Functions](#)[Readme](#)[Datasets](#)[Dependencies](#)[Imports](#)[Versions](#)

subtreeplot

Zoom on a Portion of a Phylogeny by Successive Clicks

subtrees

All subtrees of a Phylogenetic Tree

summary.phylo

Print Summary of a Phylogeny

[⬇️ Download source package](#)[📘 Read PDF manual](#)

where

which.edge

write.dna

Functions for reading, writing, plotting, and manipulating phylogenetic trees, analyses of comparative data in a phylogenetic framework, ancestral character analyses

Miostatina: allineamento multi-sequenza

```
> str(myostatin_x.seq)
```

List of 9

```
$ AF019619: chr [1:1128] "a" "t" "g" "c" ...
$ AF019620: chr [1:1128] "a" "t" "g" "c" ...
$ AF019621: chr [1:1128] "a" "t" "g" "c" ...
$ AF019622: chr [1:1128] "a" "t" "g" "c" ...
$ AF019623: chr [1:1128] "a" "t" "g" "c" ...
$ AF019624: chr [1:1131] "a" "t" "g" "a" ...
$ AF019625: chr [1:1128] "a" "t" "g" "c" ...
$ AF019626: chr [1:1125] "a" "t" "g" "c" ...
$ AF019627: chr [1:1128] "a" "t" "g" "c" ...
- attr(*, "species")= chr [1:9] "Papio_hamadryas" "Bos_taurus"
"Gallus_gallus" "Ovis_aries" ...
> length(myostatin_x.seq)
```

```
[1] 9
```

Miostatina: allineamento multi-sequenza

```
> myostatin_x.seq[1]
$AF019619
 [1] "a" "t" "g" "c" "a" "a" "a" "a" "a" "c" "t" "g" "c" "a" "a" "c" "t" "c"
[19] "t" "g" "t" "g" "t" "t" "t" "a" "t" "a" "t" "t" "t" "a" "c" "c" "t" "g"
[37] "t" "t" "t" "a" "t" "g" "c" "t" "g" "a" "t" "t" "t" "g" "t" "t" "g" "c" "t"
...
[1099] "a" "t" "g" "g" "t" "a" "g" "t" "a" "g" "a" "c" "c" "g" "c" "t" "g" "c"
[1117] "g" "g" "g" "t" "g" "c" "t" "c" "a" "t" "g" "a"
```

>

Miostatina: allineamento multi-sequenza

```
> myostatin_x.seq = read.GenBank(x, as.character = FALSE)
> str(myostatin_x.seq)
List of 9
$ AF019619: raw [1:1128] 88 18 48 28 ...
$ AF019620: raw [1:1128] 88 18 48 28 ...
$ AF019621: raw [1:1128] 88 18 48 28 ...
$ AF019622: raw [1:1128] 88 18 48 28 ...
$ AF019623: raw [1:1128] 88 18 48 28 ...
$ AF019624: raw [1:1131] 88 18 48 88 ...
$ AF019625: raw [1:1128] 88 18 48 28 ...
$ AF019626: raw [1:1125] 88 18 48 28 ...
$ AF019627: raw [1:1128] 88 18 48 28 ...
- attr(*, "class")= chr "DNAbin"
- attr(*, "description")= chr [1:9] "AF019619.1 Papio hamadryas myostatin (MSTN) mRNA, complete cds" "AF019620.1 Bos taurus myostatin (MSTN) mRNA, complete cds" "AF019621.1 Gallus gallus myostatin (MSTN) mRNA, complete cds" "AF019622.1 Ovis aries myostatin (MSTN) mRNA, complete cds" ...
- attr(*, "species")= chr [1:9] "Papio_hamadryas" "Bos_taurus" "Gallus_gallus" "Ovis_aries" ...
>
```

Miostatina: allineamento multi-sequenza

```
> attributes(myostatin_x.seq)
$names
[1] "AF019619" "AF019620" "AF019621" "AF019622" "AF019623" "AF019624" "AF019625"
[8] "AF019626" "AF019627"

$class
[1] "DNAbin"

$description
[1] "AF019619.1 Papio hamadryas myostatin (MSTN) mRNA, complete cds"
[2] "AF019620.1 Bos taurus myostatin (MSTN) mRNA, complete cds"
[3] "AF019621.1 Gallus gallus myostatin (MSTN) mRNA, complete cds"
[4] "AF019622.1 Ovis aries myostatin (MSTN) mRNA, complete cds"
[5] "AF019623.1 Sus scrofa myostatin (MSTN) mRNA, complete cds"
[6] "AF019624.1 Rattus norvegicus myostatin (MSTN) mRNA, complete cds"
[7] "AF019625.1 Meleagris gallopavo myostatin (MSTN) mRNA, complete cds"
[8] "AF019626.1 Danio rerio myostatin (MSTN) mRNA, complete cds"
[9] "AF019627.1 Homo sapiens myostatin (MSTN) mRNA, complete cds"

$species
[1] "Papio_hamadryas"      "Bos_taurus"           "Gallus_gallus"
[4] "Ovis_aries"            "Sus_scrofa"           "Rattus_norvegicus"
[7] "Meleagris_gallopavo"   "Danio_rerio"           "Homo_sapiens"

>
```

Miostatina: allineamento multi-sequenza

```
> myostatin_x.seq = read.GenBank(x, as.character = TRUE)
> attributes(myostatin_x.seq)
$names
[1] "AF019619" "AF019620" "AF019621" "AF019622" "AF019623" "AF019624"
"AF019625"
[8] "AF019626" "AF019627"

$species
[1] "Papio_hamadryas"      "Bos_taurus"          "Gallus_gallus"
[4] "Ovis_aries"           "Sus_scrofa"          "Rattus_norvegicus"
[7] "Meleagris_gallopavo" "Danio_rerio"          "Homo_sapiens"

>
```

Miostatina: allineamento multi-sequenza

```
> library(seqinr)
> write.fasta(sequences = myostatin_x.seq, names = names(myostatin_x.seq),
nbchar =80, file.out = "myostatin_x.fas")
>
```

uso la funzione `write.fasta` del pacchetto SEQINR (Biological Sequences Retrieval and Analysis) per creare il file di testo in formato «fasta» `myostatin_x.fas`, contenente le sequenze elencate nella list `myostatin_x.seq`

Home > Packages > seqinr

seqinr package 42-36

Biological Sequences Retrieval and Analysis

```
install.packages("seqinr")
```

Copy

Functions

Datasets

Dependencies

Imports

Versions

a

Converts amino-acid three-letter code into the one-letter one

aaa

Converts amino-acid one-letter code into the three-letter one

AAstat

To Get Some Protein Statistics

⬇️ Download source package

📘 Read PDF manual

acnucopen

al2bp

allistranks

Exploratory data analysis and data visualization for biological sequence (DNA and protein) data. Seqinr includes utilities for sequence data management under the

Miostatina: allineamento multi-sequenza

nel terminale

```
$ locate myostatin_x.fas
/home/piero/myostatin_x.fas
/home/piero/DNASEQUENCES/myostatin_x.fas
# questo è il «percorso» (PATH) del file <myostatin_x.fas>
```

```
$ pwd
/home/piero
```

sono già nella home, quindi non è necessario specificare l'intero PATH

```
$ gedit DNASEQUENCES/myostatin_x.fas
```

oppure

```
piero@piero-XPS-9320:~$ cd DNASEQUENCES/
piero@piero-XPS-9320:~/DNASEQUENCES$ pwd
/home/piero/DNASEQUENCES
piero@piero-XPS-9320:~/DNASEQUENCES$ gedit myostatin_x.fas
```

1 >AF019619
2 atgcaaactcaactctgtttatattacacctgttgcattttggccaggatctaaatgagaacag
3 tgagaaaaaaaatgtggaaaaagaggggctgttaatgcattgtacttggagacaaaacactaaatctcaagaatag
4 aagccattaaaatacaaatcctcagtaacttcgtctggaaacagctctaaccatcagaaagatgtataagacaactt
5 ttacccaaagctctccactccggaaactgattgatcagtatgtatgtccagggatgacagcagcgtggcttttga
6 agatgacgattatcacgctacaacggaaacaatcattaccatgcctacagactgtttaatgcaagtggatggaa
7 aacccaaatgttgccttttaatttagtctaaaatacaataaaatggtaaaggccaactatggatataattt
8 agaccgtcgagactctacaacagtgtttgcataatcctgagactcatcaaccatgtaaagacggtacaaggatatac
9 tggaatccatctctggaaacttgacatgaacccaggcactgttgcagagcattgtatgtaaagacagtgttgc
10 attggctcaaacaacctgaatccaacttaggcattgaaataaaagctttagatgagaatggcatgtatgttgc
11 ttcccaggaccaggagaagatgggctgaatcccttttagaggtcaaggtacacagacacacaaaaagatccagaagg
12 ttttggcttactgtatgtgactcaacagaatcgcgatgtcgatcccttaactgtggatttgaagctctt
13 gatgggattggattatcgctctaaaagatataaggccaattactgtctggagagtgtaaattgtattttacaaaa
14 tatcctcataactcatctgtacaccaagcaaaccagggtcagcaggccctgttactcccacaagatgtctcc
15 aattaatatgttatatttaatggcaaaagaacaaataatataatggaaaattccagccatgttagatgaccgctgcgg
16 gctcatga
17 >AF019620
18 atgcaaaaactcaatctgtttatattacacctttatgttgcattttggccaggatctgtatgagaacag
19 cgaggcagaaggaaaatgtggaaaaagaggggctgttaatgcattttggggaaaacactacatcctcaagactag
20 aagccataaaaatccaaatcctcagtaacttcgtctggaaacagctctaaccatcagaaagatgtatcagacaactt
21 ttgccaaggctctccactccgttgcactgattgatcagtgcgtatgtccagggatgtccagcgtacgcgtcc
22 agacgtgactaccacgcccaggacggaaacggcattaccatgccccacggatgttgcattcaacgcgttgc
23 aacccaaatgttgccttttaatttagtctaaagatacaataaaacttagttaaaggccaactgtggatataatct
24 aggccgtcaagactctgcgacagtgtttgcataatcctgagactcatcaaccatgtaaagacggtacaaggatatac
25 tggaatccatctctggaaacttgacatgaacccaggcactgttgcagagcattgtatgtaaagacagtgttgc
26 actggctcaaacaacctgtatcttgcatt
27 ttcccagaaccaggagaagatggactgtatcccttttagaagtcaaggtacacagacacacacaaaaagatctagg
28 ttttggcttactgtatgtgacactccacagaatctcgatgtcgatcccttaactgtggatttgaagctttt
29 gatgggattggattatttgcacaaaagatataaggccaattactgtctggagaatgtgaatttttgc
30 tatcctcataccatcttgcaccaagcaaaccagggtcagcggccctgttactcctacaatggatgtctcc
31 aattaatatgttatatttaatggcaaggacaaataatatacgggaagattccagccatgttagatgaccgctgc
32 gttcatga
33 >AF019621
34 atgcaaaaactgtcgtatgtttatattacacctgttgcattttggccaggatctgtatgagaacag
35 tcagccccacagagaacgctggaaaaagacggactgtgcatttttttttttttttttttttttttttttttt
36 aagccataaaaatccaaatcctcagtaacttcgtctggaaacagcacctaaccatcagggacgttattaagcagc
37 ttacccaaagctctccactgcggaaactgattgatcagtatgtatgtccagggacgttgcgttgc
38 agacgtgactatcatgccacaacccggacgttgcatttgcattcaatgcctacggatgttgc
39 aacccaaatgttgccttttaatttagtctaaatatacaataaaatgttgcatttttttttttttttttttt
40 aggcaagtccaaaaccttacacccgttttgcatttttttttttttttttttttttttttttttttttt
41 tggaaattcgatcttgcatt
42 attggctcaaacacgcctgtatccatcttgcatttttttttttttttttttttttttttttttttttt
43 ttcccaggaccgggtgaagatggatt
44 ttttggcttactgtatgtgactcaacggaaatcccgatgttgcatttttttttttttttttttttttt
45 gatgggactggattatagcacaaaagatatacaacggcaattactgtctccggagaatgtcaatgttgc
46 taccggccacactcacccgttgcaccaagcaaatcccaacgggtcagcggccctgttgcacaccacca
47 tataaacatgttgcatt
48 gctcatga
49 >AF019622

Sostituire ogni numero di accesso con il nome scientifico della specie
**(ATTENZIONE:
NON DEVONO ESSERCI SPAZI!)**

La razza Belga e la miostatina

```
# dopo
>Bos_taurus
Aggiungere
>bovine_belgian
atgaaaaactgcaaatctgttatatttacctattatgctgattgtgctggcccagtggatctgaatgagaacag
cgagcagaaggaaaatgtggaaaaagaggggctgtgtaatgcattgtggagggaaaacactacatcctcaagactag
aagccataaaaatccaaatcctcagtaaacttcgcctggaaacagtcctaaacatcagcaaagatgctatcagacaactt
ttgccaaggctcctccactcctggaaactgattgatcagttcgatgtccagagagatgccagcgtgacggctccttggaa
agacgatgactaccacgccaggacggaaacggcattaccatgcccacggagtctgatcttctaaccgcaagtggaaaggaa
aacccaaatgtgtttaaatttagtctaagatacaataacttagtaaaggccaaactgtggatatatctg
aggcctgtcaagactcctgcacagtgttgcataatcctgagactcatcaaaccatgaaagacggtacaaggatac
tggaatccgatctgtgaaacttgacatgaaaccaggcactggatttgcagagcattgtgaaagacagtgatgtcaga
actggctcaaacaacctgaatccaacttaggcattgaaatcaaagcttagatgagaatggccatgatctgctgtacc
ttcccagaaccaggagaagatggactgactccttttagaagtcaaggtaacagacacacacaaaagatctaggagaga
tttggcttgattgtgacagaatctcgatgctgcgttaccctctaactgtggatttgaagctttggatggattgg
attattgcacccaaagatataaggccaaattactgctctggagaatgtgaatttgtattttgaaaagatcctcatac
ccatctgtgcaccaagccaaacccagggttcagccggccctgtgtactcctacaaagatgtctccaaattaatatgc
tatatttaatggcgaaggacaaataatcggaaagattccagccatggtagtagatcgctgtgggtgttcatga
>bovine_piedmontese
atgaaaaactgcaaatctgttatatttacctattatgctgattgtgctggcccagtggatctgaatgagaacag
cgagcagaaggaaaatgtggaaaaagaggggctgtgtaatgcattgtggagggaaaacactacatcctcaagactag
aagccataaaaatccaaatcctcagtaaacttcgcctggaaacagtcctaaacatcagcaaagatgctatcagacaactt
ttgccaaggctcctccactcctggaaactgattgatcagttcgatgtccagagagatgccagcgtgacggctccttggaa
agacgatgactaccacgccaggacggaaacggcattaccatgcccacggagtctgatcttctaaccgcaagtggaaaggaa
aacccaaatgtgtttaaatttagtctaagatacaataacttagtaaaggccaaactgtggatatatctg
aggcctgtcaagactcctgcacagtgttgcataatcctgagactcatcaaaccatgaaagacggtacaaggatac
tggaatccgatctgtgaaacttgacatgaaaccaggcactggatttgcagagcattgtgaaagacagtgatgtcaga
actggctcaaacaacctgaatccaacttaggcattgaaatcaaagcttagatgagaatggccatgatctgctgtacc
ttcccagaaccaggagaagatggactgactccttttagaagtcaaggtaacagacacacacaaaagatctaggagaga
tttggcttgattgtgatgaaacactccacagaatctcgatgctgcgttaccctctaactgtggatttgaagctttg
gatggattggattattgcacccaaagatataaggccaaattactgctctggagaatatgaaattgtattttgaaaag
tatcctcataccatctgtgcaccaagccaaacccagggatccagccatggtagtagatcgctgtgggt
atataatgtatatttaatggcgaaggacaaataatcggaaagattccagccatggtagtagatcgctgtgggt
gttcatga
```

La razza Belga e la miostatina

```
# ctrl+MAIUSC+s
# save as myostatin_x_mod.fas
```

Apri

Salva

myostatin_x.fas
~/DNA_SEQUENCES

```
1 >Papio_hamadryas
2 atgcaaaaactgcaactctgtt
3 tgagaaaaagaaaaatgtggaaa
4 aagccattaaatacaaattcctca
5 ttacccaaagtctccactccgg
6 agatgacgattatcacgctacaac
7 aacccaaatgttgcattttaaat
8 agaccgtcgagactcctacaacac
9 tggaatccatctctgaaacttgaa
10 attggctcaaacaacctgaatcca
11 ttcccaggaccaggagaagatgg
12 ttttgttcttgcattgtgatgacca
13 gatgggattggattatcgctccaa
14 tatttcataactcatctgttacac
15 aattaatatgttatatttaatgg
16 gctcatga
17 >Bos_taurus
18 atgcaaaaactgcaatctgtt
19 cgagcagaaggaaaatgtggaaa
20 aagccataaaaatccaaatcctca
21 ttgccaaggctccactcctg
22 agacatgactaccacgcccaggac
23 aacccaaatgttgcattttaaat
24 aggctgtcaagactcctgcgacac
25 tggaatccatctctgaaacttgaa
26 actggctcaaacaacctgaatcca
27 ttcccagaaccaggagaagatgg
28 ttttgttcttgcattgtgatgacca
29 gatgggattggattattgcaccta
30 tatccataccatctgtcace
31 aattaatatgttatatttaatgg
32 gttcatga
33 >bovine_belgian
34 atgcaaaaactgcaatctgtt
35 cgagcagaaggaaaatgtggaaa
36 aagccataaaaatccaaatcctca
37 ttgccaaggctccactcctg
38 agacatgactaccacgcccaggac
39 aacccaaatgttgcattttaaat
40 aggctgtcaagactcctgcgacac
41 tggaatccatctctgaaacttgaa
42 actggctcaaacaacctgaatcca
43 ttcccagaaccaggagaagatgg
44 ttttgttcttgcattgtgacgaaat
45 attattgcacaaaagatataa
46 ccatttgtgcaccaaggaaaccc
47 tatatttaatggcaaggacaaa
48 >bovine_piedmontese
49 atgcaaaaactgcaatctgtttatattacattatgttgcattgtgctggcccagtgatctgaaacag
```

Annula Nome myostatin_x_mod.fas Salva

Home piero DNA_SEQUENCES

Nome Dimensione Tipo Data di modifica

Bovidae_myostatin.fasta	8,6 kB	Testo	15 nov 2024
database_completo_miostatina.txt	480,1 kB	Testo	13 gen 2024
dexter.fasta	88,6 kB	Testo	3 set 2024
fcodons.R	210 byte	Testo	11 nov 2024
link_Cervidae.txt	344 byte	Testo	10 gen 2024
MT_Cervidae	333,7 kB	Testo	10 gen 2024
MT_Cervidae.fasta	448,4 kB	Testo	11 gen 2024
my_data.txt	98,9 kB	Testo	27 dic 2023
myostatin.dnd	3,7 kB	Testo	19 nov 2024
myostatin.fas	220,0 kB	Testo	gio
myostatin.fasta	714,6 kB	Testo	19 nov 2024
myostatin.nexus	1,2 MB	Testo	1 ott 2024
MYOSTATIN_ACCESSION_NUMBERS.txt	1,7 kB	Testo	29 set 2024
MYOSTATIN_DATA.txt	17,3 kB	Testo	29 set 2024
myostatin_mod.aln	0 byte	Testo	gio
myostatin_mod.dnd	3,8 kB	Testo	gio
myostatin_mod.fas	222,3 kB	Testo	gio
myostatin_mod.fasta	727,8 kB	Testo	gio
myostatin_species_names.dnd	4,3 kB	Testo	ven
myostatin_species_names.fas	222,9 kB	Testo	ven
myostatin_species_names.fasta	740,2 kB	Testo	ven
myostatin_x.fas	12,7 kB	Testo	15:16
myostatin_x_mod.dnd	359 byte	Testo	09:50
myostatin_x_mod.fasta	13,0 kB	Testo	09:50
Phocoena_sinus_myostatin.fas	3,0 kB	Testo	21 nov 2024
piero.fas	1,2 kB	Testo	6 dic 2024
Rhinolophus_sinicus_myostatin.fas	2,9 kB	Testo	28 nov 2024
Sus_scrofa_myostatin.fas	1,2 kB	Testo	21 nov 2024
sylvia.fas	31,5 kB	Testo	14 nov 2024
wolves.csv	1,1 kB	Testo	25 dic 2023

Codifica caratteri: Locale attuale (UTF-8) Fine riga: Unix/Linux

Tutti i file di testo

La razza Belga e la miostatina

nel terminale

```
$ clustalx
```

ClustalX 2.1

File Edit Alignment Trees Colors Quality Help

Mode: Multiple Alignment Mode ▾ Font: 10 ▾

Cancel

Load Sequence File

Open

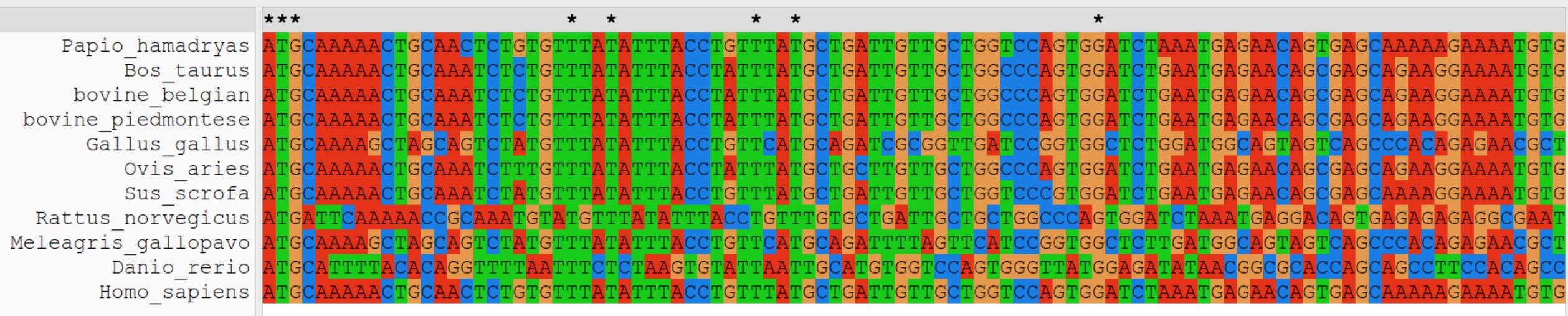
⌚ Recenti

-  Home
-  Documenti
-  Immagini
-  Musica
-  Scaricati
-  Video
-  DNA SEQUENCING

+ Altre posizioni

Nome	Dimensione	Tipo	Data di modifica
Bos_taurus_myostatin.fas	2,9 kB	Testo	15 nov 2024
Bovidae_myostatin.dnd	75 byte	Testo	15 nov 2024
Bovidae_myostatin.fas	8,6 kB	Testo	15 nov 2024
Bovidae_myostatin.fasta	8,6 kB	Testo	15 nov 2024
database_completo_miostatina.txt	480,1 kB	Testo	13 gen 2024
dexter.fasta	88,6 kB	Testo	3 set 2024
Fcodons.R	210 byte	Testo	11 nov 2024
link_Cervidae.txt	344 byte	Testo	10 gen 2024
mcpherron-lee-1997-double-muscling-in-cattle-due-to-mu...	537,5 kB	Documento	27 set 2024
MT_Cervidae	333,7 kB	Testo	10 gen 2024
MT_Cervidae.fasta	448,4 kB	Testo	11 gen 2024
my_data.txt	98,9 kB	Testo	27 dic 2023
myostatin.dnd	3,7 kB	Testo	19 nov 2024
myostatin.fas	220,0 kB	Testo	gio
myostatin.fasta	714,6 kB	Testo	19 nov 2024
myostatin.nexus	1,2 MB	Testo	1 ott 2024
myostatin.RData	8,8 MB	R Data File	28 set 2024
MYOSTATIN_ACCESSION_NUMBERS.txt	1,7 kB	Testo	29 set 2024
MYOSTATIN_DATA.txt	17,3 kB	Testo	29 set 2024
myostatin_mod.aln	0 byte	Testo	gio
myostatin_mod.dnd	3,8 kB	Testo	gio
myostatin_mod.fas	222,3 kB	Testo	gio
myostatin_mod.fasta	727,8 kB	Testo	gio
myostatin_species_names.dnd	4,3 kB	Testo	ven
myostatin_species_names.fas	222,9 kB	Testo	ven
myostatin_species_names.fasta	740,2 kB	Testo	ven
myostatin_x.fas	12,7 kB	Testo	15:16
myostatin_x_mod.dnd	359 byte	Testo	09:50
myostatin_x_mod.fas	12,7 kB	Testo	15:22
myostatin_x_mod.fasta	13,0 kB	Testo	09:50

La razza Belga e la miostatina



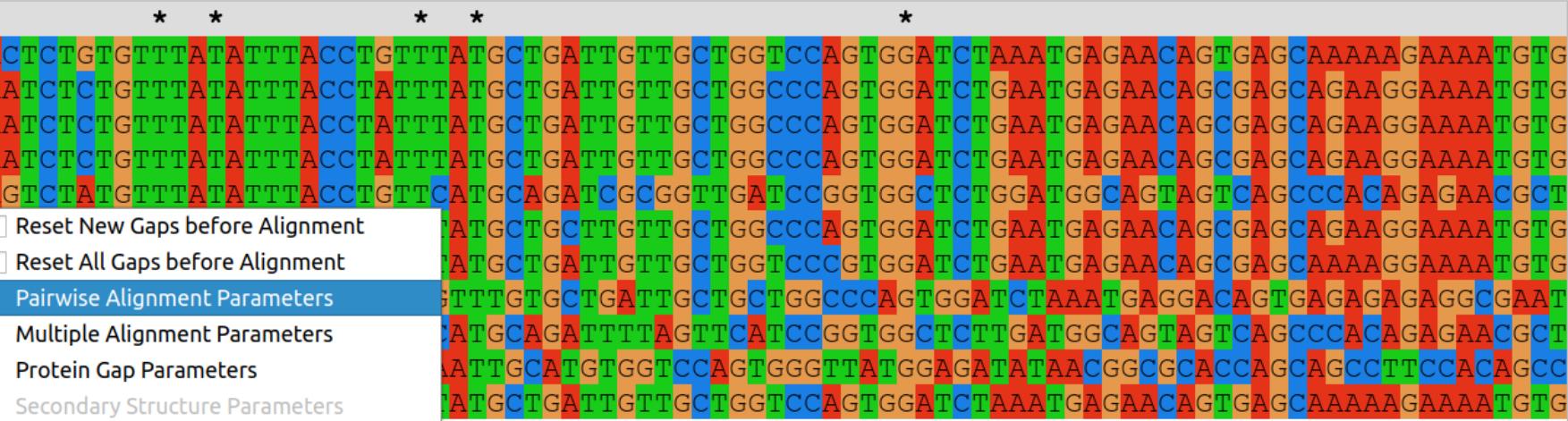
PAIRWISE ALIGNMENT PARAMETERS	GAOPEN	GAEXTENSION	DNA TRANSITION WEIGHT
10.0	0.2		
MULTIPLE ALIGNMENT PARAMETERS	10.0	0.1	0.5

File Edit Alignment Trees Colors Quality Help

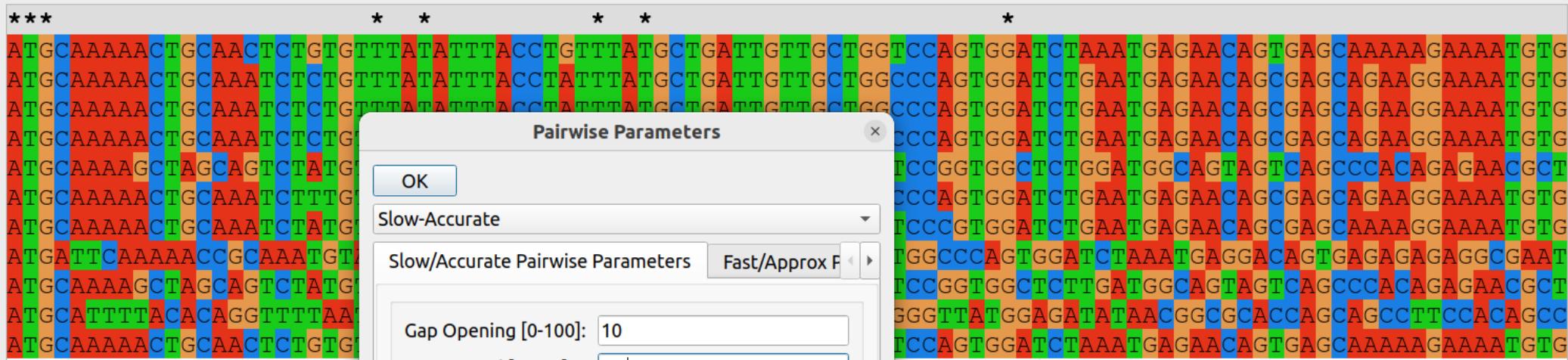
Mode: Multip

- Do Complete Alignment Ctrl+L
- Do Guide Tree Only Ctrl+G
- Do Alignment from Guide Tree
- Realign Selected Sequences
- Realign Selected Residue Range
- Align Profile 2 to Profile 1
- Align Profiles from Guide Trees
- Align Sequences to Profile 1
- Align Sequences to Profile 1 from Tree
- Alignment Parameters**
- Iteration
- Output Format Options
- Set All Parameters to default**

Danio rerio ATGCATTTACACAT
Homo sapiens ATGCAAAAACGTGCA



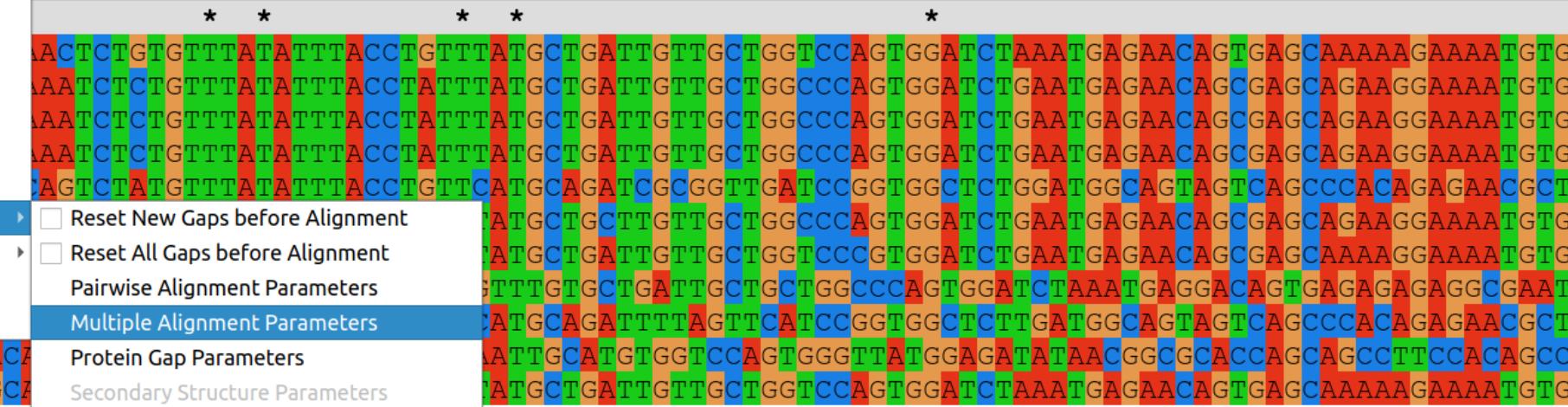
Papio_hamadryas
Bos_taurus
bovine_belgian
bovine_piedmontese
Gallus_gallus
Ovis_aries
Sus_scrofa
Rattus_norvegicus
Meleagris_gallolopavo
Danio_rerio
Homo_sapiens



Mode: Multip

- Do Complete Alignment Ctrl+L
- Do Guide Tree Only Ctrl+G
- Do Alignment from Guide Tree
- Realign Selected Sequences
- Realign Selected Residue Range
- Align Profile 2 to Profile 1
- Align Profiles from Guide Trees
- Align Sequences to Profile 1
- Align Sequences to Profile 1 from Tree
- Alignment Parameters**
- Iteration
- Output Format Options
- Set All Parameters to default**

Danio rerio	ATGCATTTACACAT
Homo sapiens	ATGAAAAAACTTGCA



Mode: **Multiple Alignment Mode** Font: **24**

Papio_hamadryas
Bos_taurus
bovine_belgian
bovine_piedmontese
Gallus_gallus
Ovis_aries
Sus_scrofa
Rattus_norvegicus
Meleagris_gallopavo
Danio_rerio
Homo_sapiens

The sequence logo displays the conservation of each nucleotide (A, T, C, G) across the 10 positions of the motif. The height of each bar indicates the probability of finding that nucleotide at a given position. The x-axis represents the position of the motif, and the y-axis represents the probability of each nucleotide.

Position	A	T	C	G
1	0.95	0.05	0.0	0.0
2	0.0	0.95	0.05	0.0
3	0.0	0.95	0.05	0.0
4	0.0	0.95	0.05	0.0
5	0.0	0.95	0.05	0.0
6	0.0	0.95	0.05	0.0
7	0.0	0.95	0.05	0.0
8	0.0	0.95	0.05	0.0
9	0.0	0.95	0.05	0.0
10	0.0	0.95	0.05	0.0

The screenshot shows the BioEdit software interface with two DNA sequences aligned. The top sequence has two asterisks (*). The bottom sequence has three asterisks (*). Below the sequences is a title "Alignment Parameters". A blue button labeled "OK" is visible. A section titled "Multiple Parameters" contains four input fields: "Gap Opening [0-100]" set to 10, "Gap Extension [0-100]" set to 0.1, "Delay Divergent Sequences(%):" set to 30, and "DNA Tolerance width [0-1]" set to 0.5.

A sequence logo visualization showing the conservation of each nucleotide (T, C, G, A) across 100 positions. The x-axis represents the position number (1 to 100). The y-axis lists the four nucleotides. Colored bars indicate the frequency of each nucleotide at each position: blue for T, green for C, red for G, and orange for A. A black asterisk (*) is positioned above the first column.

File Edit Alignment Trees Colors Quality Help

Mode: Multip

- Do Complete Alignment Ctrl+L
- Do Guide Tree Only Ctrl+G
- Do Alignment from Guide Tree
- Realign Selected Sequences
- Realign Selected Residue Range
- Align Profile 2 to Profile 1
- Align Profiles from Guide Trees
- Align Sequences to Profile 1
- Align Sequences to Profile 1 from Tree
- Alignment Parameters
- Iteration
- Output Format Options
- Set All Parameters to default**

bov

Align Profile 2 to Profile 1

Align Profiles from Guide Trees

Align Sequences to Profile 1

Align Sequences to Profile 1 from Tree

Alignment Parameters

Iteration

Output Format Options

Set All Parameters to default

Rat

Mele

Set All Parameters to default

Danio_rerio

ATGCATTTACACAGGTTAATTCTAAAGTGTATTAAATTGCATGGTCCAGTGGTTATGGAGATAAACGGCACCAGCAGCTTCCACAGCC

Homo_sapiens

ATGAAAAACTGCAACTCTGTGTTATATTACCTGTTATGCTGATTGTTGCTGGTCCAGTGGATCTAAATGAGAACAGTGAGCAAAAAGAAAATGTG



Mode: **Multiple Alignment Mode** Font: **24**

Papio_hamadryas
Bos_taurus
bovine_belgian
bovine_piedmontese
Gallus_gallus
Ovis_aries
Sus_scrofa
Rattus_norvegicus
Meleagris_gallopavo
Danio_rerio
Homo_sapiens

The sequence logo displays the probability of each nucleotide (A, T, C, G) at each position of the motif. The x-axis represents the positions, and the y-axis represents the nucleotides. The height of each bar indicates the probability of a nucleotide at a given position.

Position	A	T	C	G
1	0.0	1.0	0.0	0.0
2	0.0	1.0	0.0	0.0
3	0.0	1.0	0.0	0.0
4	0.0	1.0	0.0	0.0
5	0.0	1.0	0.0	0.0
6	0.0	1.0	0.0	0.0
7	0.0	1.0	0.0	0.0
8	0.0	1.0	0.0	0.0
9	0.0	1.0	0.0	0.0
10	0.0	1.0	0.0	0.0
11	0.0	1.0	0.0	0.0
12	0.0	1.0	0.0	0.0
13	0.0	1.0	0.0	0.0
14	0.0	1.0	0.0	0.0
15	0.0	1.0	0.0	0.0
16	0.0	1.0	0.0	0.0
17	0.0	1.0	0.0	0.0
18	0.0	1.0	0.0	0.0
19	0.0	1.0	0.0	0.0
20	0.0	1.0	0.0	0.0
21	0.0	1.0	0.0	0.0
22	0.0	1.0	0.0	0.0
23	0.0	1.0	0.0	0.0
24	0.0	1.0	0.0	0.0
25	0.0	1.0	0.0	0.0
26	0.0	1.0	0.0	0.0
27	0.0	1.0	0.0	0.0
28	0.0	1.0	0.0	0.0
29	0.0	1.0	0.0	0.0
30	0.0	1.0	0.0	0.0
31	0.0	1.0	0.0	0.0
32	0.0	1.0	0.0	0.0
33	0.0	1.0	0.0	0.0
34	0.0	1.0	0.0	0.0
35	0.0	1.0	0.0	0.0
36	0.0	1.0	0.0	0.0
37	0.0	1.0	0.0	0.0
38	0.0	1.0	0.0	0.0
39	0.0	1.0	0.0	0.0
40	0.0	1.0	0.0	0.0
41	0.0	1.0	0.0	0.0
42	0.0	1.0	0.0	0.0
43	0.0	1.0	0.0	0.0
44	0.0	1.0	0.0	0.0
45	0.0	1.0	0.0	0.0
46	0.0	1.0	0.0	0.0
47	0.0	1.0	0.0	0.0
48	0.0	1.0	0.0	0.0
49	0.0	1.0	0.0	0.0
50	0.0	1.0	0.0	0.0
51	0.0	1.0	0.0	0.0
52	0.0	1.0	0.0	0.0
53	0.0	1.0	0.0	0.0
54	0.0	1.0	0.0	0.0
55	0.0	1.0	0.0	0.0
56	0.0	1.0	0.0	0.0
57	0.0	1.0	0.0	0.0
58	0.0	1.0	0.0	0.0
59	0.0	1.0	0.0	0.0
60	0.0	1.0	0.0	0.0
61	0.0	1.0	0.0	0.0
62	0.0	1.0	0.0	0.0
63	0.0	1.0	0.0	0.0
64	0.0	1.0	0.0	0.0
65	0.0	1.0	0.0	0.0
66	0.0	1.0	0.0	0.0
67	0.0	1.0	0.0	0.0
68	0.0	1.0	0.0	0.0
69	0.0	1.0	0.0	0.0
70	0.0	1.0	0.0	0.0
71	0.0	1.0	0.0	0.0
72	0.0	1.0	0.0	0.0
73	0.0	1.0	0.0	0.0
74	0.0	1.0	0.0	0.0
75	0.0	1.0	0.0	0.0
76	0.0	1.0	0.0	0.0
77	0.0	1.0	0.0	0.0
78	0.0	1.0	0.0	0.0
79	0.0	1.0	0.0	0.0
80	0.0	1.0	0.0	0.0
81	0.0	1.0	0.0	0.0
82	0.0	1.0	0.0	0.0
83	0.0	1.0	0.0	0.0
84	0.0	1.0	0.0	0.0
85	0.0	1.0	0.0	0.0
86	0.0	1.0	0.0	0.0
87	0.0	1.0	0.0	0.0
88	0.0	1.0	0.0	0.0
89	0.0	1.0	0.0	0.0
90	0.0	1.0	0.0	0.0
91	0.0	1.0	0.0	0.0
92	0.0	1.0	0.0	0.0
93	0.0	1.0	0.0	0.0
94	0.0	1.0	0.0	0.0
95	0.0	1.0	0.0	0.0
96	0.0	1.0	0.0	0.0
97	0.0	1.0	0.0	0.0
98	0.0	1.0	0.0	0.0
99	0.0	1.0	0.0	0.0
100	0.0	1.0	0.0	0.0

Output Format Options

Output Files

<input type="checkbox"/> CLUSTAL format	<input type="checkbox"/> NBRF/PIR format
<input type="checkbox"/> GCG/MSF format	<input type="checkbox"/> PHYLIP format
<input type="checkbox"/> GDE format	<input type="checkbox"/> NEXUS format
<input checked="" type="checkbox"/> FASTA format	

GDE output case :

CLUSTALW sequence numbers :

Output order :

Parameters output :

The sequence logo displays the following conservation pattern across 10 DNA sequences:

- Position 1:** G (green)
- Position 2:** C (blue)
- Position 3:** G (green)
- Position 4:** G (green)
- Position 5:** A (green)
- Position 6:** T (red)
- Position 7:** C (blue)
- Position 8:** T (red)
- Position 9:** A (green)
- Position 10:** G (orange)

Asterisks (*) are placed above positions 1, 3, 5, 6, 7, 8, and 10, indicating high conservation of the indicated nucleotide at those specific positions.

File Edit Alignment Trees Colors Quality Help

Mode: Multip

- Do Complete Alignment Ctrl+L
- Do Guide Tree Only Ctrl+G
- Do Alignment from Guide Tree
- Realign Selected Sequences
- Realign Selected Residue Range
- Align Profile 2 to Profile 1
- Align Profiles from Guide Trees
- Align Sequences to Profile 1
- Align Sequences to Profile 1 from Tree
- Alignment Parameters
- Iteration
- Output Format Options
- Set All Parameters to default

Danio rerio ATGCATTTACACAGGTTAATTCTAAAGTGTATTAAATTGCATGGTCCAGTGGTTATGGAGATAAACGGCACCAGCAGCTTCCACAGCC
Homo sapiens ATGCAAAAACCTGCAACTCTGTGTTATATTACCTGTTATGCTGATTGTTGCTGGTCCAGTGGATCTAAATGAGAACAGTGAGCAAAAAGAAAAATGTG

Mode: **Multiple Alignment Mode** Font: **24**

Papio_hamadryas
Bos_taurus
bovine_belgian
bovine_piedmontese
Gallus_gallus
Ovis_aries
Sus_scrofa
Rattus_norvegicus
Meleagris_gallopavo
Danio_rerio
Homo_sapiens

*** * * * *

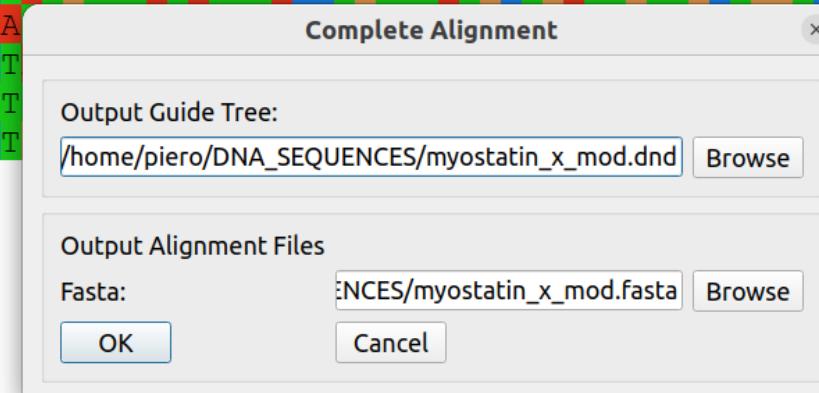
ATGCAAAAAC TGCAACTCTGTGTTATTTACCTGTTATGCTGATTGTTGCTGGTCCAGTGGATCTAAATGAGAACAGTGAGCAAAAGAAAAATGTG
ATGCAAAAAC TGCAAATCTCTGTGTTATTTACCTATTTATGCTGATTGTTGCTGGCCCAGTGGATCTGAATGAGAACAGCGAGCAGAAGGAAAATGTG
ATGCAAAAAC TGCAAATCTCTGTGTTATTTACCTATTTATGCTGATTGTTGCTGGCCCAGTGGATCTGAATGAGAACAGCGAGCAGAAGGAAAATGTG
ATGCAAAAAC TGCAAATCTCTGTGTTATTTACCTATTTATGCTGATTGTTGCTGGCCCAGTGGATCTGAATGAGAACAGCGAGCAGAAGGAAAATGTG
ATGCAAAAGCTAGCAGTCTATGTTATTTACCTGTTCATGCAGATCGCGGTTGATCCGGTGGCTCTGGATGGCAGTAGTCAGCCCCACAGAGAACGCT
ATGCAAAAAC TGCAAATCTTGTGTTATTTACCTATTTATGCTGCTTGTGCTGGCCCAGTGGATCTGAATGAGAACAGCGAGCAGAAGGAAAATGTG
ATGCAAAAAC TGCAAATCTATGTTATTTACCTGTTATGCTGATTGTTGCTGGTCCC GTGGATCTGAATGAGAACAGCGAGCAAAAGGAAAATGTG
ATGATTCAAAAAACCGGAAA
ATGCAAAAGCTAGCAGTCT
ATGCATTTACACAGGTTT
ATGCAAAAAC TGCAACTCT

Complete Alignment

Output Guide Tree:

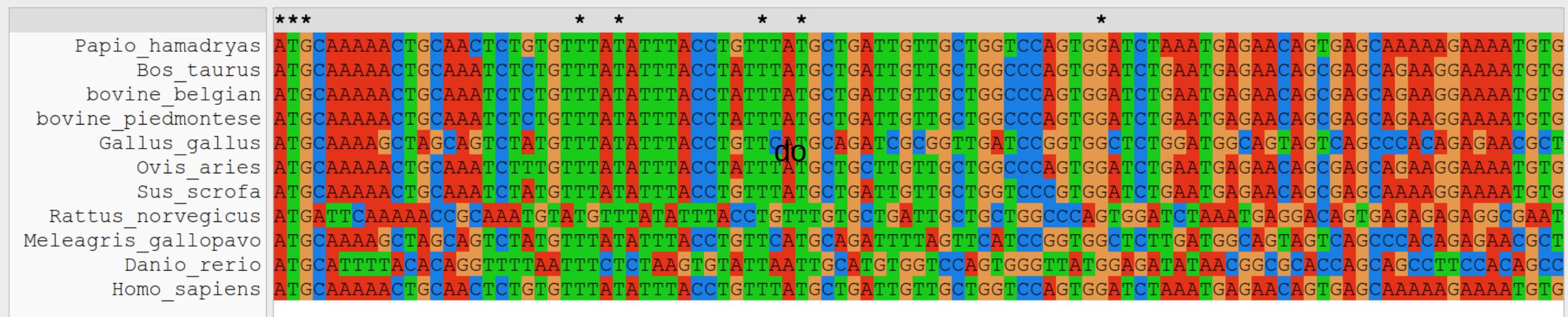
/home/piero/DNA SEQUENCES/mystatin_x.mod.dnd

Browse

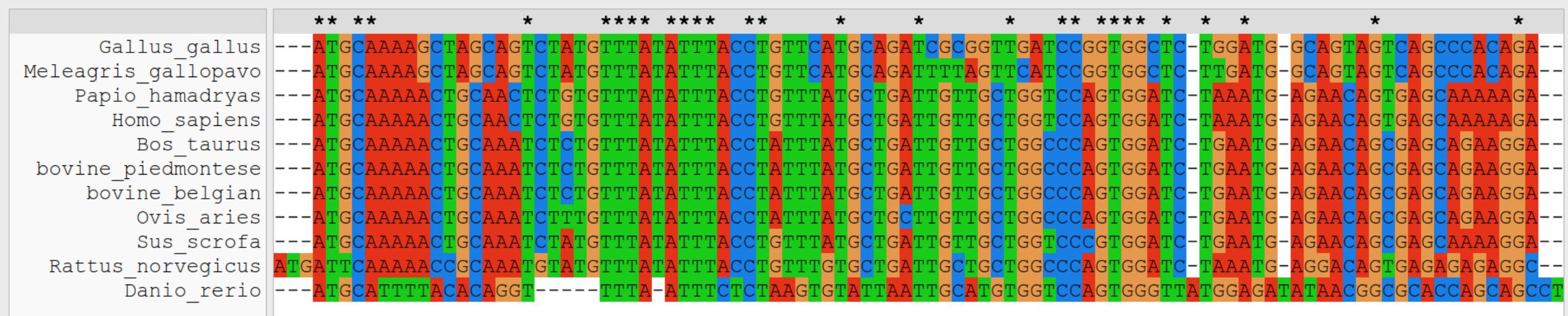


La razza Belga e la miostatina

Prima



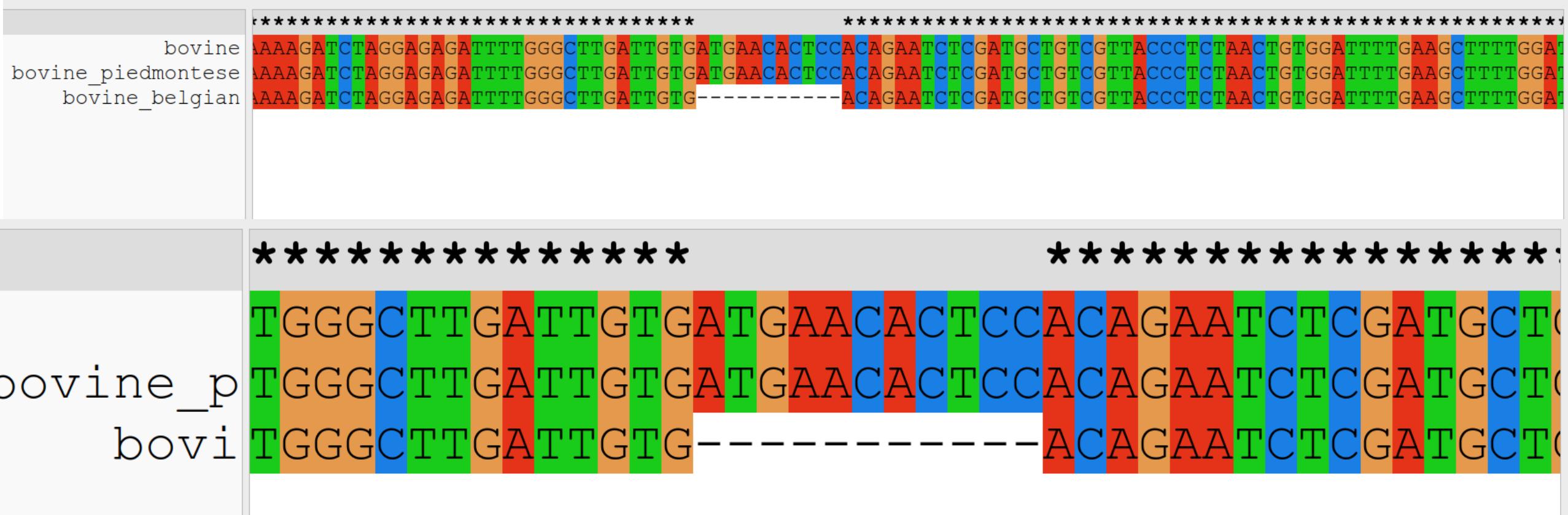
dopo



La razza Belga e la miostatina

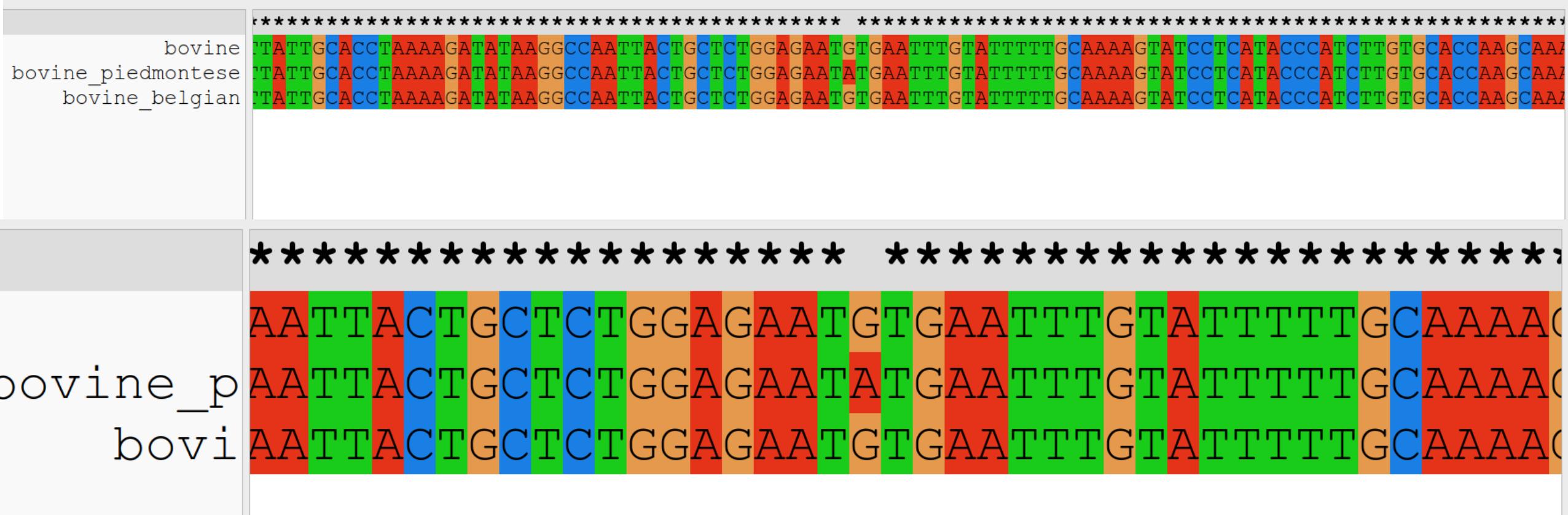
TGAACACTCCA

La razza Belga e la miostatina

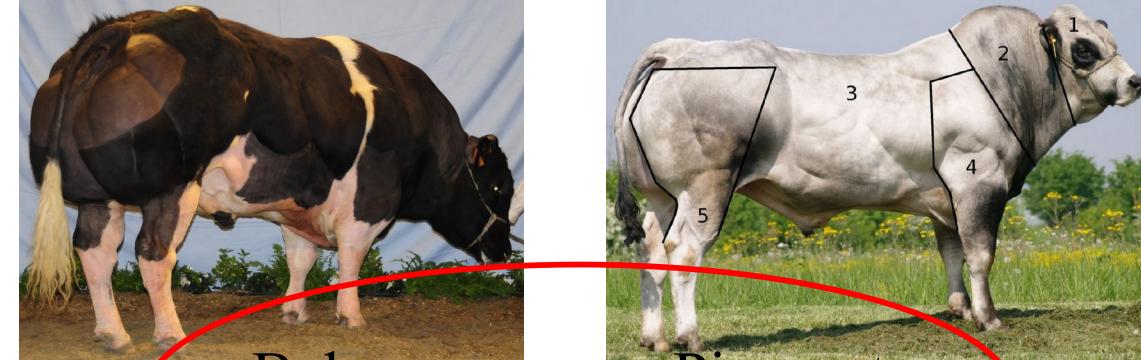


La razza Piemontese e la miostatina

Transizione G → A



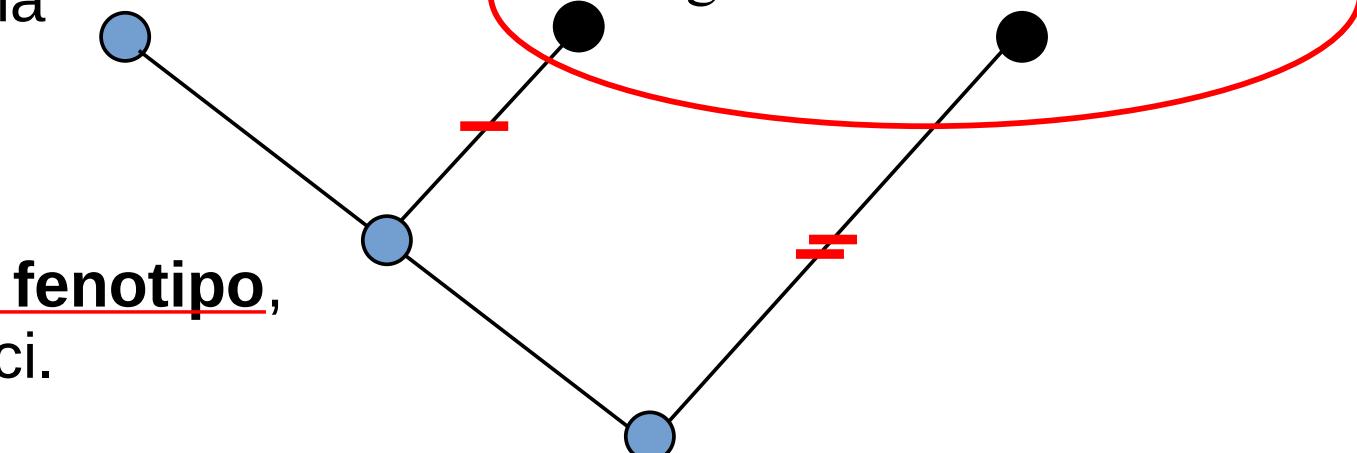
Le razze bovine Belga e Piemontese, se venissero classificate nella stessa categoria in base alla somiglianza fenotipica dovuta all'iperplasia muscolare congenita, formerebbero un **gruppo polifiletico**, ossia artificiale.



Belga

Piemontese

Ciò evidenzia i limiti della classificazione basata sul fenotipo, ossia sui caratteri morfologici.



Miostatina: allineamento multi-sequenza

gcnud.com/pierorivoira/ESERCITAZIONI/DIOD/main/myostatin_species_names.ras 200% Accedi

pierorivoira / ESERCITAZIONI

Code Issues Pull requests Actions Projects 1 Wiki ...

Files main ESERCITAZIONI / myostatin_species_names.fas ...

pierorivoira Add files via upload cdecaa6 · 45 minutes ago

2906 lines (2906 loc) · 218 KB Download raw file

Code Blame Raw ...

```
1 >baboon
2 atgcaaaaactgcaactctgtttatattacacctgttatgctgattgtgctggccagtggatctaaatgagaacag
3 ttagcaaaaagaaaaatgtggaaaaagagggcgtgtaatgcatgtacttggagacaaaacactaaatcttcaagaatag
4 aaqccattaaaatacaaattcctcagtaaacttcgtcqaaacaqctcctaacatcaqcaaqaqtqctataaqacaactt
```

Miostatina: allineamento multi-sequenza

Mode: Multiple Alignment Mode ▾ Font: 10 ▾ Cancel

Load Sequence File

Recenti

- Home
- Documenti
- Immagini
- Musica
- Scaricati
- Video
- DNA_SEQUENCES
- + Altre posizioni

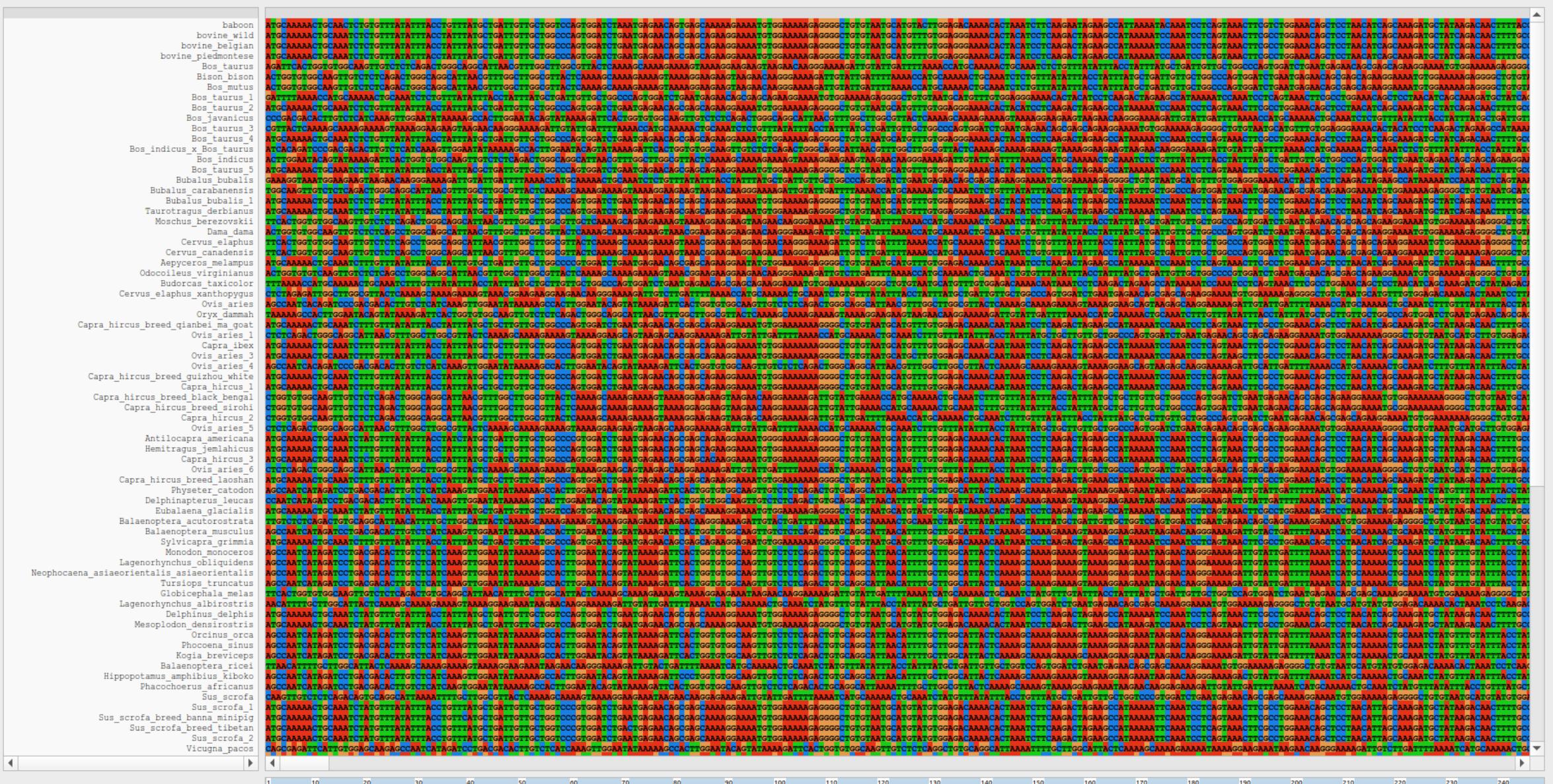
piero DNA_SEQUENCES

Nome	Dimensione	Tipo	Data di modifica
.RData	10,2 MB	R Data File	Ieri
.Rhistory	15,5 kB	Testo	Ieri
allineamento_miostatina_blast.txt	480,1 kB	Testo	11 gen 2024
Bos_taurus_myostatin.fas	2,9 kB	Testo	15 nov 2024
Bovidae_myostatin.dnd	75 byte	Testo	15 nov 2024
Bovidae_myostatin.fas	8,6 kB	Testo	15 nov 2024
Bovidae_myostatin.fasta	8,6 kB	Testo	15 nov 2024
database_completo_miostatina.txt	480,1 kB	Testo	13 gen 2024
dexter.fasta	88,6 kB	Testo	3 set 2024
Fcodons.R	210 byte	Testo	11 nov 2024
link_Cervidae.txt	344 byte	Testo	10 gen 2024
mcpherron-lee-1997-double-muscling-in-cattle-due-to-mu...	537,5 kB	Documento	27 set 2024
MT_Cervidae	333,7 kB	Testo	10 gen 2024
MT_Cervidae.fasta	448,4 kB	Testo	11 gen 2024
my_data.txt	98,9 kB	Testo	27 dic 2023
myostatin.dnd	3,7 kB	Testo	19 nov 2024
myostatin.fas	220,0 kB	Testo	Ieri
myostatin.fasta	714,6 kB	Testo	19 nov 2024
myostatin.nexus	1,2 MB	Testo	1 ott 2024
myostatin.RData	8,8 MB	R Data File	28 set 2024
MYOSTATIN_ACCESSION_NUMBERS.txt	1,7 kB	Testo	29 set 2024
MYOSTATIN_DATA.txt	17,3 kB	Testo	29 set 2024
myostatin_mod.aln	0 byte	Testo	Ieri
myostatin_mod.dnd	3,8 kB	Testo	Ieri
myostatin_mod.fas	222,3 kB	Testo	Ieri
myostatin_mod.fasta	727,8 kB	Testo	Ieri
myostatin_species_names.fas	222,9 kB	Testo	21:40
myostatin_x.fas	10,4 kB	Testo	17:03
myostatin_x_mod.dnd	265 byte	Testo	17:39
myostatin_x_mod.fas	11,5 kB	Testo	17:38

All Files ▾

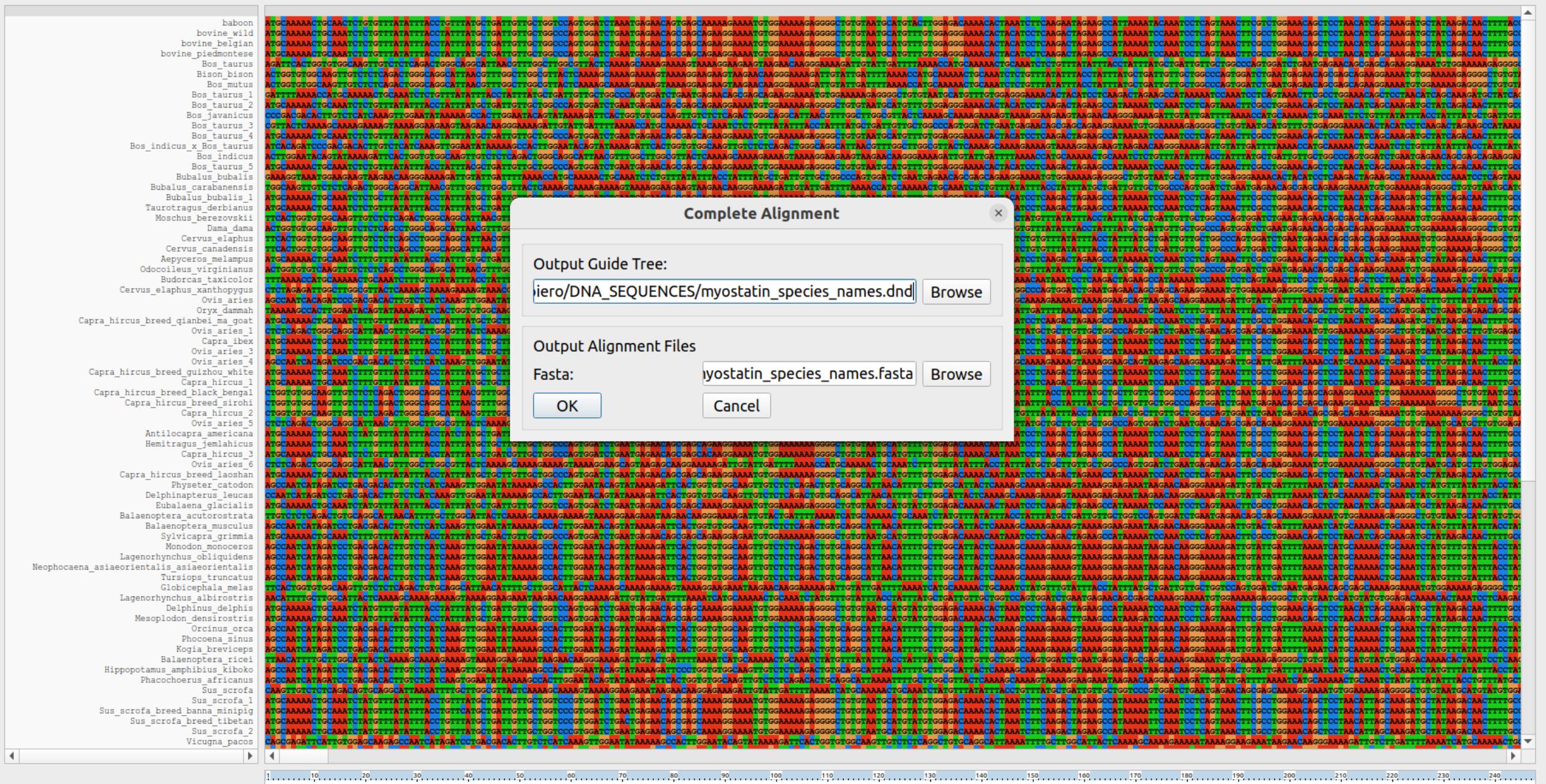
Miostatina: allineamento multi-sequenza

Mode: Multiple Alignment Mode ▾ Font: 10 ▾

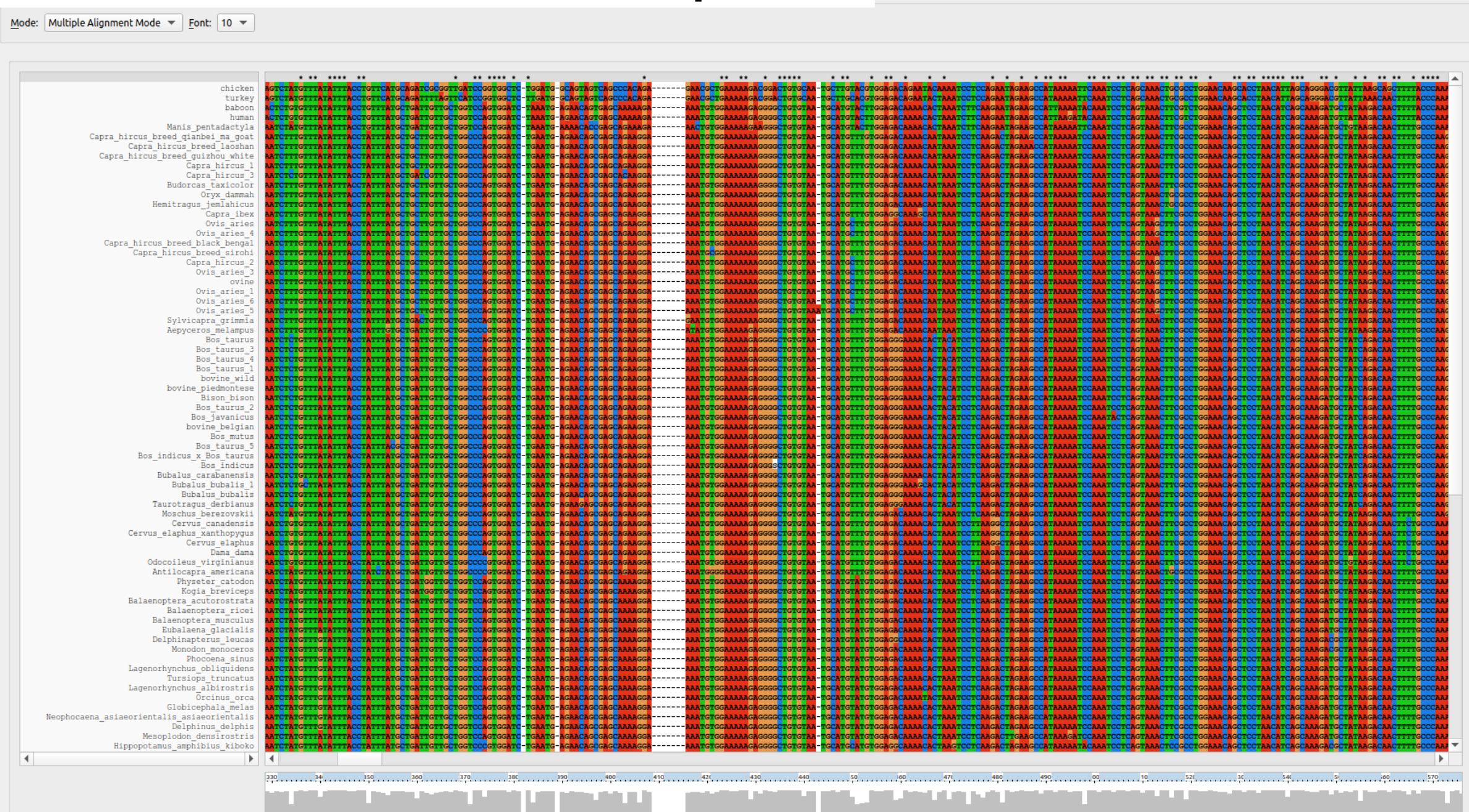


Miostatina: allineamento multi-sequenza

Mode: Multiple Alignment Mode ▾ Font: 10 ▾



Miostatina: allineamento multi-sequenza



La razza Belga e la miostatina

```
> setwd('~/DNASEQUENCES/')
> load('myostatin.RData')
# [sostituire <myostatin.RData> con il nome del file che avete creato (es.: <piero.Rdata>)]

> library(ape)
> library(seqinr)

> myostatin.fas = read.dna("myostatin_species_names.fas", format="fasta")
> myostatin.fasta = read.dna("myostatin_species_names.fasta", format="fasta")

> myostatin.fas

> myostatin.fasta
```

La razza Belga e la miostatina

> myostatin.fasta

111 DNA sequences in binary format stored in a matrix.

all sequences of same length: 6434

Labels:

chicken

turkey

baboon

human

Manis_pentadactyla

Capra_hircus_breed_qianbei_ma_goat

...

Base composition:

a	c	g	t
---	---	---	---

a.337	a.181	a.189	a.293
-------	-------	-------	-------

(Total: 714.17 kb)

La razza Belga e la miostatina

```
> str(myostatin.fasta)
```

```
'DNabin' raw [1:111, 1:6434] - - - - . . .
- attr(*, "dimnames")=List of 2
..$ : chr [1:111] "chicken" "turkey" "baboon" "human" ...
..$ : NULL
```

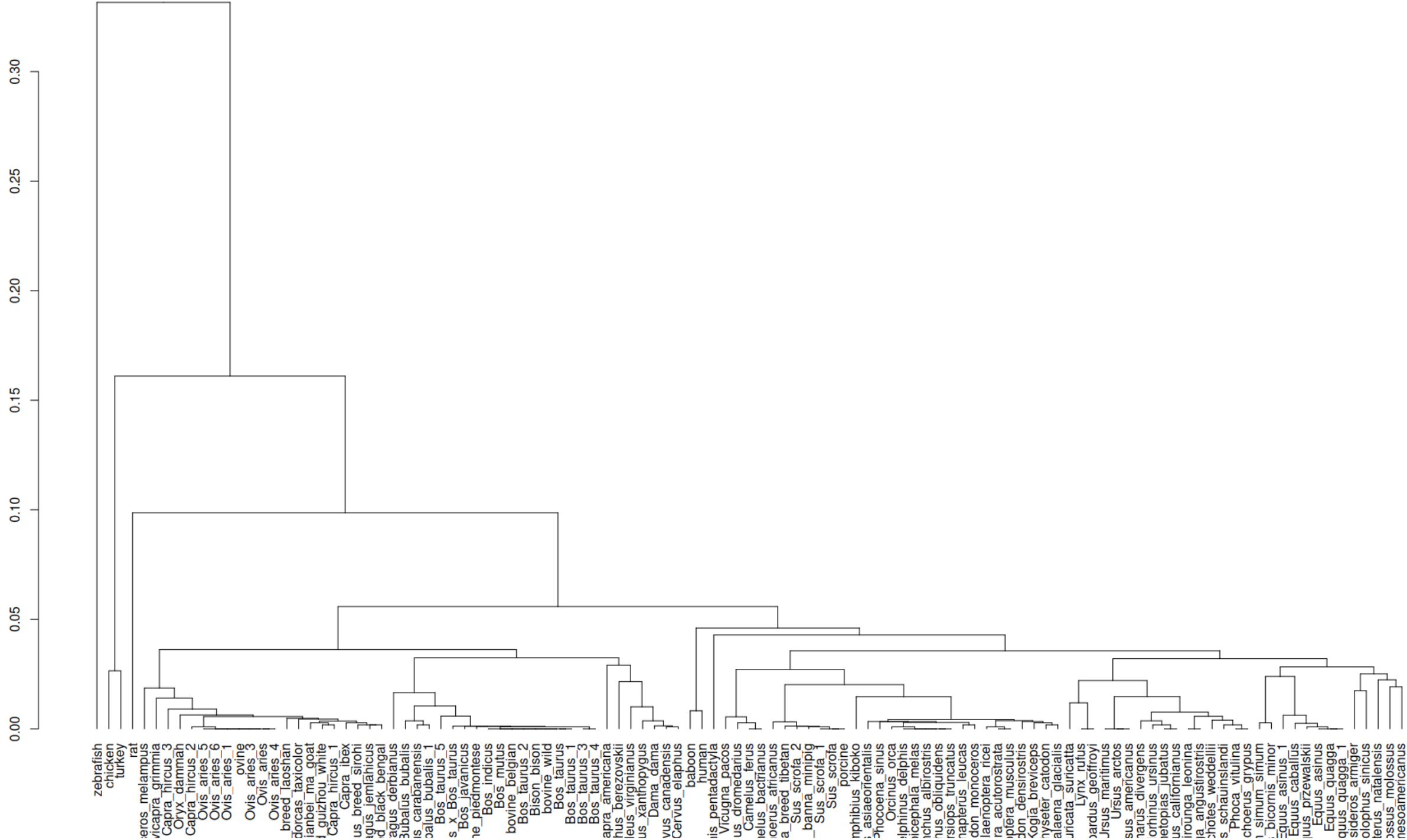
```
> myostatin.dist=dist.dna(myostatin.fasta, model="raw")
```

```
> myostatin.dist
```

turkey	chicken	turkey	baboon
baboon	a.a26484a18		
human	a.15a684932	a.143378995	
Manis_pentadactyla	a.149771689	a.142465753	a.aa8219178
Capra_hircus_breed_qianbei_ma_goat	a.1652968a4	a.153424658	a.a54794521
Capra_hircus_breed_laoshan	a.175342466	a.168a3653a	a.a76712329
Capra_hircus_breed_guizhou_white	a.1762557a8	a.169863a14	a.a78538813
Capra_hircus_1	a.1762557a8	a.168949772	a.a77625571
Capra_hircus_3	a.174429224	a.167123288	a.a75799a87
Budorcas_taxicolor	a.173515982	a.168a3653a	a.a76712329
	a.1762557a8	a.167123288	a.a739726a3

La razza Belga e la miostatina

```
> myostatin.tree = hclust(myostatin.dist,"average")
> plot(as.dendrogram(myostatin.tree))
```



La razza Belga e la miostatina

- Scegliere 20 specie, max 2 della stessa famiglia
- Uno dei taxa dev'essere **Bovine Belgian**, un altro **Zebra fish**
- Costruire l'albero UPGMa
- Come si distribuiscono i vari taxa sui rami dell'albero? Perchè?
(rispondere sul quaderno)

```
> library(ape)
> setwd('~/MATERIALE_DIDATTICO_A.S._2024-2025/ESERCITAZIONI/MIOSTATINA/')
> myostatin.fasta = read.dna("myostatin_5_species_100_bp.fasta",
format="fasta")
> X = myostatin.fasta
> myostatin.dist = dist.dna(X, model="raw")
```

raw: This is simply the proportion or the number of sites that differ between each pair of sequences.

```
> library(ape)
> setwd('~/MATERIALE_DIDATTICO_A.S._2024-2025/ESERCITAZIONI/MIOSTATINA/')
> X = read.dna("myostatin_5_species_100_bp.fasta", format="fasta",
as.character = FALSE)
> str(X)
'DNAbin' raw [1:5, 1:100] a a a a ...
- attr(*, "dimnames")=List of 2
 ..$ : chr [1:5] "baboon" "bovine" "chicken" "ovine" ...
 ..$ : NULL
> X
5 DNA sequences in binary format stored in a matrix.
```

All sequences of same length: 100

Labels:
baboon
bovine
chicken
ovine
zebra_fish

Base composition:

a	c	g	t
0.286	0.178	0.240	0.296
(Total: 500 bases)			

```
> myostatin.dist = dist.dna(X, model="raw")
```

```
> myostatin.dist
```

baboon bovine chicken ovine

bovine	0.08			
chicken	0.29	0.32		
ovine	0.09	0.02	0.33	
<u>zebra_fish</u>	0.67	0.64	0.70	0.63



1 2 3 4 5 6 7 8

	*****	*****	*****	*****	*****	*****	*****					
baboon	ATGCAAAA	ACTGCAAC	TCTGTGTT	TATTACCTG	TGCTGATTG	TGGTCCAG	TGGATCTAA	AATGAGAAC	AGTGA	GAGCAAAA	AGAAAATGT	GG
bovine	ATGCAAAA	ACTGCAAA	TCTGTGTT	TATTACCTG	TGCTGATTG	TGGTCCAG	TGGATCTGA	ATGAGAAC	AGCAGA	GAGAAGGAAA	ATGT	GG

1 234 5 6 7 8 9 1011 12 13 14 15 1617 1819 2021 22 2324 25 26 27 2829

	*****	***	*	*****	*****	*****	*****	*****	*	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
baboon	ATGCAAAA	ACTGCAAC	TCTGTGTT	TATTACCTG	TGCTGATTG	TGGTCCAG	TGGATCTAA	AATGAGAAC	AGTGA	GAGCAAAA	AGAAAATGT	GG											
chicken	ATGCAAAAG	CAGTCAGT	TATGTT	TATTACCTG	TGTTCATGC	AGATCGCGG	TGATCCGG	TGGCTCTGG	ATGGCAGT	AGTCAGCCC	ACAGAGAAC	GCTG											

1 2 3 4 5 6 7 8 9

	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
baboon	ATGCAAAA	ACTGCAAC	TCTGTGTT	TATTACCTG	TGCTGATTG	TGGTCCAG	TGGATCTAA	AATGAGAAC	AGTGA	GAGCAAAA	AGAAAATGT	GG										
ovine	ATGCAAAA	ACTGCAAAT	CTTGTT	TATTACCT	ATTGCTGCT	GGTCTGG	CCAGTGG	ATCTGAAT	GAGAACAGC	GAGCAGA	AGGAAAATGT	GG										

1234567 89 10 11 12 13 14 15 16 17 18 19 20 21 22 2324 2526 2728 29 30 3132

	*****	**	*	*	***	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*****	*	*
baboon	ATGCAAAA	ACTGCAAC	TCTGTGTT	TATTACCTG	TGCTGATTG	TGGTCCAG	TGGATCTAA	AATGAGAAC	AGTGA	GAGCAAAA	AGAAAATGT	GG											
ra_fish	ATGCATT	TTACACAGG	TTTAATTCT	CTAAAGTGT	ATTGCATGT	GGTCCAGT	GGTTATGG	AGATATAAC	GGCGCAC	AGCAGCCT	TCACAGCCA												

1 234 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 2324 2526 2728 29 30 3132

	*****	***	*	***	*****	*****	*****	*****	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
bovine	ATGCAAAA	ACTGCAAAT	TCTGTGTT	TATTACCT	ATTGCTG	ATTGCTG	GGCCCAGT	GGATCTGA	ATGAGAAC	AGCAGA	GAGCAGA	AGGAAAATGT	GG										
chicken	ATGCAAAAG	CAGTCAGT	TATGTT	TATTACCTG	TGTTCATGC	AGATCGCGG	TGATCCGG	TGGCTCTGG	ATGGCAGT	AGTCAGCCC	ACAGAGAAC	GCTG											

1

2

1234567 89 10 111213 14 1516171819 20 212223242526 27 28 29 3031323334353637 3839404142 43444546 4748495051 52 535455565758 5960 616263

1 234 5 6 7 8 9 10 11 1213 14 1516 17 18 19 2021 2223 2425 2627 2829 30 31 3233

123456 789 10 111213 14 15 1617181920 21 22 23242526272829 3031323334 3536373839404142 4344454647484950515253545556575859 606162 63646566 6768 6970

	*****	*	**	*	***	*	*	*	***	*	*	*	*	*	*****	*	*	**
chicken	ATGCAAAAGCTAGCAGTCTATGTTATATTACCTGTTCATGCAGATCGCGGTTGATCCGGTGGCTCTGGATGGCAGTAGTCAGCCCACAGAGAACGCTG																	
ra_fish	ATGCATTACACAGGTTAATTCTCTAAGTGTATTAAATTGCATGTGGTCCAGTGGGTTATGGAGATAAACGGCGACCAGCAGCCTCACAGCCA																	

1234567 89 10 111213 14 15 1617181920 21 222324252627 28 29 30 3132333435363738 3940414243 44454647 4849505152 53 545556575859 6061 626364

	*****	**	* *	****	*	*	*****	**	***	*	*	*	*	*	*	*	*	*****	*	*
ovine	ATGCAAAAAACTG	C	AAATCTTGT	TATTTACCT	ATTATGCT	GCTGGCCAGTG	GAATGAGAACAGC	GAGCAGAAGGAAAATGTGG												
ra_fish	ATGCATTACACAGG	TTAATTCTCTAAGT	GATGTATTAACTGGT	CCAGTGGGTTATGG	AGATAAACGGCGACC	CAGCCTTCACAGCCA														

```
> myostatin.UPGMA = hclust(myostatin.dist, "average")
# simple UPGMA
```

```
> myostatin.UPGMA
```

Call:

```
hclust(d = myostatin.dist, method = "average")
```

```
Cluster method : average
Distance       : raw
Number of objects: 5
```

La funzione hclust effettua una *cluster analysis* gerarchica utilizzando, in questo caso, l'algoritmo **UPGMA** (*Unweighted Pair Group Method with Arithmetic Mean*).

L'idea alla base della *cluster analysis* gerarchica è di rappresentare graficamente quali elementi (singoli esemplari di una stessa o di specie diverse, sequenze di DNA ecc.) di un certo campione sono più simili tra di loro, raggruppandoli a due a due sullo stesso ramo dell'albero.

Gli oggetti che sono più diversi fra loro vengono collocati in rami differenti.

```
> myostatin.dist
```

baboon bovine chicken ovine

	bovine	chicken	ovine	zebra_fish
bovine	0.08			
chicken	0.29	0.32		
ovine	0.09	0.02	0.33	
zebra_fish	0.67	0.64	0.70	0.63

Matrice **D1** di distanze genetiche
(n° di nucleotidi differenti su 100)
fra coppie di specie

Il bovino e la pecora sono le specie
tra le quali la D1 è <, quindi le
mettiamo insieme

Quindi calcoliamo la distanza media
tra bov-ovi e ciascuna delle altre
specie

$$D2((bov,ovi),bab) = [D1(bov,bab)+D1(ovi,bab)]/2 = (8+9)/2 = 8,5$$

$$D2((bov,ovi),gal) = [D1(bov,gal)+D1(ovi,gal)]/2 = (32+33)/2 = 32,5$$

$$D2((bov,ovi),zeb) = [D1(bov,zeb)+D1(ovi,zeb)]/2 = (64+63)/2 = 63,5$$

	bab	bov	gal	ovi	zeb
bab	0	8	29	9	67
bov	8	0	32	2	64
gal	29	32	0	33	70
ovi	9	2	33	0	63
zeb	67	64	70	63	0

	bab	bov-ovi	gal	zeb
bab	0	8,5	29	67
bov-ovi	8,5	0	32,5	63,5
gal	29	32,5	0	70
zeb	67	63,5	70	0

	(bov-ovi)bab	gal	zeb
(bov-ovi)bab	0	30,7 5	65,25
gal	30,75	0	70
zeb	65,25	70	0

$$D3(((bov,ovi),bab),gal) = [D2((bov,ovi),gal)+(D2(bab,gal)]/2 = (32,5+29)/2 = 30,75$$

$$D3(((bov,ovi),bab),zeb) = [D2((bov,ovi),zeb)+(D2(bab,zeb)]/2 = (63,5+67)/2 = 65,25$$

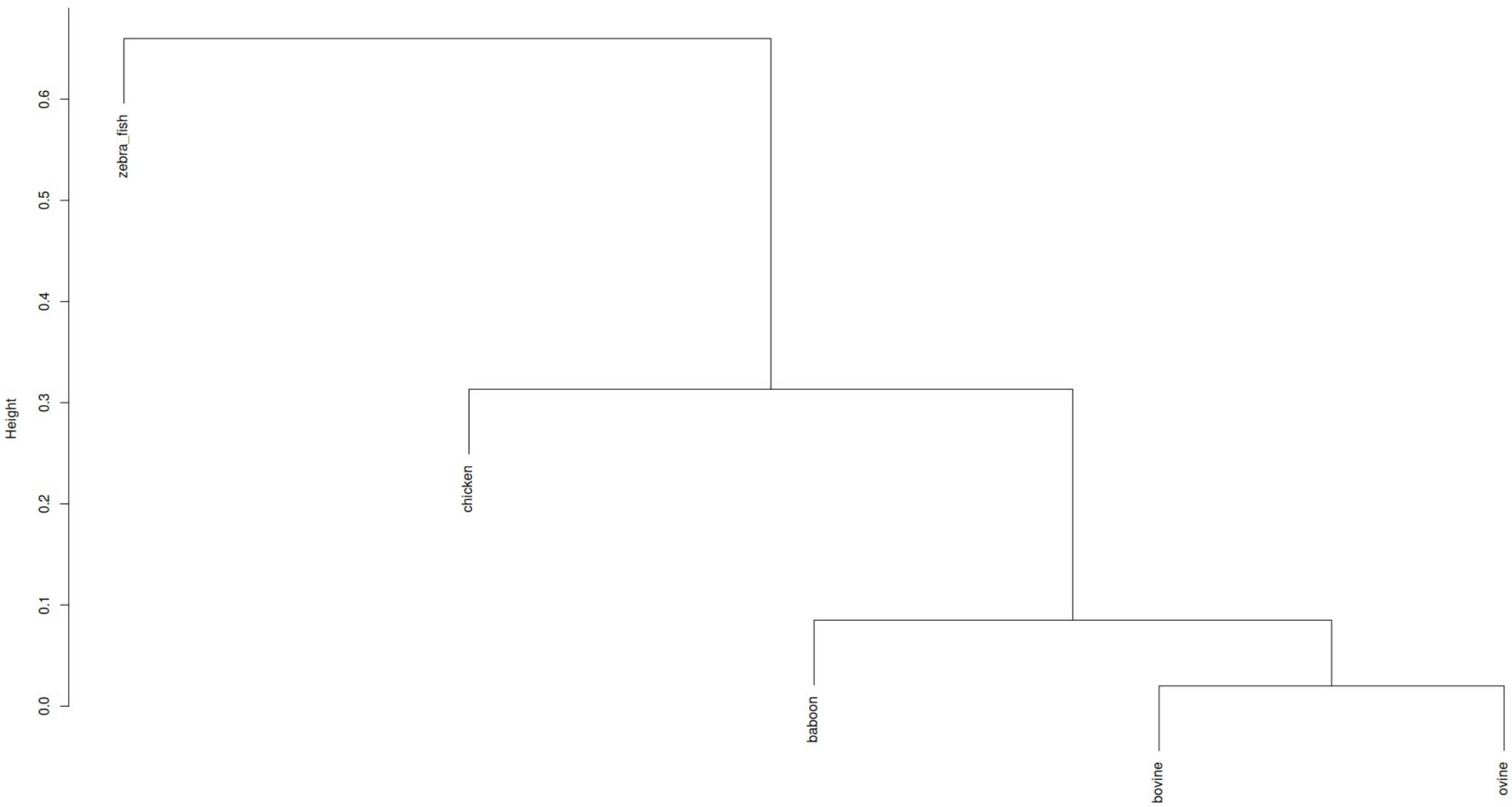
	(bov-ovi)bab	gal	zeb
(bov-ovi)bab	0	30,7 5	65,25
gal	30,75	0	70
zeb	65,25	70	0

$$D4(((bov,ovi),bab),gal),zeb) = [D3((bov,ovi),bab),zeb)+D3(gal,zeb)]/2 = (65,25+70)/2 = 67,625$$

	((bov-ovi)bab)ga	zeb
((bov-ovi)bab)gal	0	67,625
zeb	67,625	0

```
> plot(myostatin.UPGMA, main = 'UPGMA distances')
```

UPGMA distances

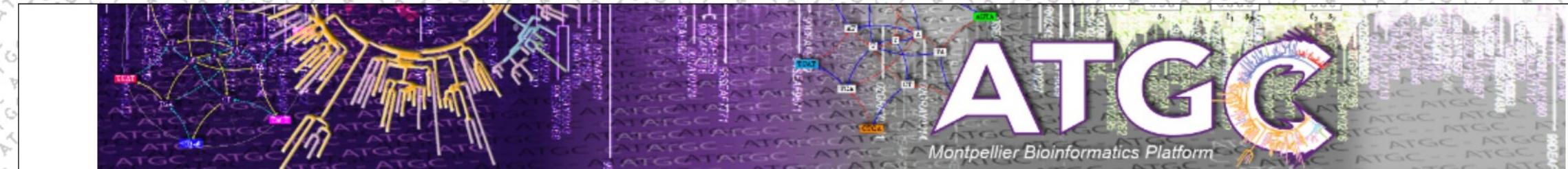


myostatin.dist
hclust (*, "average")

Scelta del modello di sostituzione nucleotidica

L'analisi filogenetica più semplice si basa sul calcolo delle distanze fra coppie di sequenze nucleotidiche, le quali diventano gli ingressi di una matrice che, tramite algoritmi di clusterizzazione, viene convertita in un albero filogenetico.
La distanza fra due sequenze è definita come numero atteso di sostituzioni nucleotidiche per sito (generalmente < 1) e può essere espressa come semplice proporzione dei siti variabili o distanza p.

```
> myostatin.dist  
            baboon bovine chicken ovine  
bovine          0.08  
chicken        0.29    0.32  
ovine          0.09    0.02    0.33  
zebra_fish     0.67    0.64    0.70   0.63
```



Home

Organization

Citations & Statistics

Partners

Online programs

PhyML

Benchmarks

Datasets

Forum & FAQ

Online execution

Source code on GitHub

Papers & contacts

PhyML versions

User guide

Binaries

Databases

Datasets

NGS



PhyML 3.0: new algorithms, methods and utilities

Please cite:

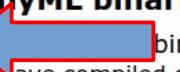
"New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0."

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O.

Systematic Biology, 59(3):307-21, 2010.

Get the latest PhyML releases from GitHub.

Download PhyML binaries

Click [here](#) to  binaries

The binaries you have compiled on various Mac, Windows or LINUX systems are much welcome ([Stéphane Guindon](#)) and will be displayed on this web page.

Contact: [Webmaster](#), LIRMM.

Il formato PHYLIP interleaved

```
5 100
baboon
atgaaaaactgcaactctgtttatattacctgttatgctgattgtgctggccagtggatctaaatgagaacagttagaaaaaaaatgtgg
bovine
atgaaaaactgcaaactctgtttatattacctatattatgctgattgtgctggcccagtggatctgaatgagaacagcgagcagaaggaaaatgtgg
chicken
atgaaaaagcttagcagtctatgttatattacctgttcatgcagatcgccgttgcgtccgtggatggcagtagtcagcccacagagaacgctg
ovine
atgaaaaactgcaaactttgttatattacctatattatgctgctgtgctggcccagtggatctgaatgagaacagcgagcagaaggaaaatgtgg
zebrafish
atgcatttacacaggttaattctctaagtgttattgcattgtggccagtggatggagatataacggcgcaccagcagcctccacagcca

> substitution_model_test = phymltest('myostatin_5_species_100_bp.phy', format
= "interleaved", itree = NULL, exclude = NULL, execname = '/home/piero/PhyML-
3.1/PhyML-3.1_linux64', append = TRUE)
```

--- PhyML 20120412 ---

<http://www.atgc-montpellier.fr/phym>

Copyright CNRS - Universite Montpellier II

Sequence filename: myostatin_5_species_100_bp.phy
Data set: #1
Tree topology search : NNIs
Initial tree: BioNJ
Model of nucleotides substitution: GTR
Number of taxa: 5
Log-likelihood: -402.43191
Unconstrained likelihood: -343.80642
Parsimony: 100
Tree size: 4.55423
Nucleotides frequencies:
- f(A) = 0.28600
- f(C) = 0.17800
- f(G) = 0.24000
- f(T) = 0.29600
GTR relative rate parameters :
A <-> C 11.85779
A <-> G 14.05957
A <-> T 0.00993
C <-> G 8.42143
C <-> T 13.46606
G <-> T 1.00000
Instantaneous rate matrix :
[A-----C-----G-----T-----]
-1.01259 0.38945 0.62260 0.00054
0.62574 -1.73412 0.37293 0.73546
0.74193 0.27659 -1.07313 0.05462
0.00052 0.44227 0.04428 -0.48708

L'output del test

```
> substitution_model_test
      nb.free.para    loglik      AIC
JC69                  1 -414.3124 830.6248
JC69+I                2 -414.1910 832.3821
JC69+G                2 -413.8407 831.6814
JC69+I+G              3 -413.8313 833.6625
K80                  2 -411.0496 826.0993
K80+I                3 -408.3180 822.6359
K80+G                3 -408.8849 823.7699
K80+I+G              4 -407.9957 823.9915
F81                  4 -414.1411 836.2823
F81+I                5 -414.0996 838.1992
F81+G                5 -413.7568 837.5136
F81+I+G              6 -413.7486 839.4972
F84                  5 -409.8965 829.7931
F84+I                6 -407.9673 827.9346
F84+G                6 -408.0924 828.1848
F84+I+G              7 -407.5650 829.1301
HKY85                5 -409.8806 829.7612
HKY85+I              6 -407.9661 827.9323
HKY85+G              6 -408.0732 828.1464
HKY85+I+G            7 -407.5730 829.1460
TN93                 6 -409.8804 831.7607
TN93+I               7 -407.9637 829.9274
TN93+G               7 -408.0700 830.1400
TN93+I+G             8 -407.5648 831.1297
GTR                  9 -402.4319 822.8638
GTR+I                10 -401.2392 822.4784
GTR+G                10 -400.2603 820.5205
GTR+I+G              11 -400.2379 822.4758
>
```

II file <myostatin_5_species_100_bp.phy_phyml_tree.txt>

(bovine:0.00000202, (ovine:0.00002963, zebrafish:9.99805710) 0.000000:0.02049504,
(baboon:0.00008628, chicken:0.50886609) 0.842000:0.09067509);
(baboon:0.01048862, (bovine:0.00000001, (zebrafish:1.37441488, ovine:0.00000508):0.02026739):0.07447686, chicken:0.35829670);
(baboon:0.00972686, chicken:0.36497752, (bovine:0.00000118, (ovine:0.00000029, zebrafish:1.54101437):0.02030498):0.07560156);
(baboon:0.00007584, chicken:0.44219573, (ovine:0.02056245, (bovine:0.00000001, zebrafish:3.54999086):0.00000001):0.08826723);
(baboon:0.00000910, chicken:0.45531080, (bovine:0.00000001, (zebrafish:4.16221975, ovine:0.00001874):0.02059383):0.08887950);
(bovine:0.00000099, (ovine:0.02028732, zebrafish:3.09237767):0.00003742, (baboon:0.01122146, chicken:0.36941408):0.07398107);
(zebrafish:8.82019763, (bovine:0.00000175, ovine:0.02039573):0.02277632, (chicken:0.39408351, baboon:0.01090682):0.05227845);
(bovine:0.00000099, (ovine:0.02044652, zebrafish:10.00000000):0.00002146, (baboon:0.00002629, chicken:0.50356626):0.08929782);
(chicken:0.43223177, baboon:0.00713230, (bovine:0.00000126, (zebrafish:8.89223736, ovine:0.01973263):0.00074093):0.08047846);
(baboon:0.01115834, chicken:0.36143405, (bovine:0.00000181, (zebrafish:1.42016394, ovine:0.00000268):0.02033915):0.07448850);
(bovine:0.00000091, (chicken:0.36579093, baboon:0.01074355):0.07520598, (zebrafish:1.52463974, ovine:0.00000001):0.02037077);
(bovine:0.00000117, (zebrafish:3.58786252, (chicken:0.44228493, baboon:0.00171289):0.08712725):0.00000872, ovine:0.02063859);
(baboon:0.00000824, chicken:0.46731497, (bovine:0.00000001, (zebrafish:4.82711651, ovine:0.00000756):0.02070614):0.08999625);
(bovine:0.00000101, (ovine:0.02042293, zebrafish:3.77229889):0.00003150, (baboon:0.01338516, chicken:0.37881952):0.07317730);
(zebrafish:8.98214792, (bovine:0.00000182, ovine:0.02045418):0.02675688, (chicken:0.39647730, baboon:0.01296603):0.04725382);
(bovine:0.00000204, (ovine:0.00002615, zebrafish:9.99812034):0.02054687, (baboon:0.00002661, chicken:0.50940308):0.09100487);
(chicken:0.44286719, zebrafish:9.08352240, (baboon:0.00942811, (bovine:0.00000132, ovine:0.02057252):0.07969971):0.00014779);
(bovine:0.00000099, (ovine:0.02039898, zebrafish:3.70535167):0.00003159, (baboon:0.01300483, chicken:0.37844433):0.07338886);
(zebrafish:8.98429384, (bovine:0.00000180, ovine:0.02042441):0.02673440, (chicken:0.39638088, baboon:0.01260347):0.04745109);
(bovine:0.00000202, (ovine:0.00002963, zebrafish:9.99805710):0.02049504, (baboon:0.00008628, chicken:0.50886609):0.09067509);
(chicken:0.44246753, zebrafish:9.08528623, (baboon:0.00898618, (bovine:0.00000130, ovine:0.02053515):0.07990025):0.00016885);
(bovine:0.00000099, (ovine:0.02039408, zebrafish:3.68224905):0.00003166, (baboon:0.01293915, chicken:0.37829051):0.07342438);
(zebrafish:8.98269671, (bovine:0.00000180, ovine:0.02044070):0.03061041, (chicken:0.39640554, baboon:0.01278578):0.04346390);
(bovine:0.00000205, (ovine:0.00003700, zebrafish:9.99683005):0.02045292, (baboon:0.00008704, chicken:0.50879952):0.09053445);
(chicken:0.44280285, zebrafish:9.08381557, (baboon:0.00936629, (bovine:0.00000130, ovine:0.02056699):0.07972586):0.00014726);
(bovine:0.00000001, (zebrafish:4.05071318, (baboon:0.00593667, chicken:0.39542302):0.08161689):0.00000004, ovine:0.02053827);
(zebrafish:4.34749500, (bovine:0.00000057, (chicken:0.40630281, baboon:0.00316606):0.08445039):0.02051856, ovine:0.00000251);
(zebrafish:10.00000000, (bovine:0.00000056, (baboon:0.00001018, chicken:0.54408882):0.09498755):0.02116598, ovine:0.00001086);
(baboon:0.0000993, chicken:0.53213267, (bovine:0.00000001, (zebrafish:9.76620951, ovine:0.00000001):0.02114857):0.09443561);

Scelta del modello di sostituzione nucleotidica

```
> r = '(baboon:0.00000993,chicken:0.53213267,(bovine:0.0000001,  
(zebrafish:9.76620951,ovine:0.0000001):0.02114857):0.09443561)';  
> albero = read.tree(text = r)  
> albero
```

Phylogenetic tree with 5 tips and 3 internal nodes.

Tip labels:

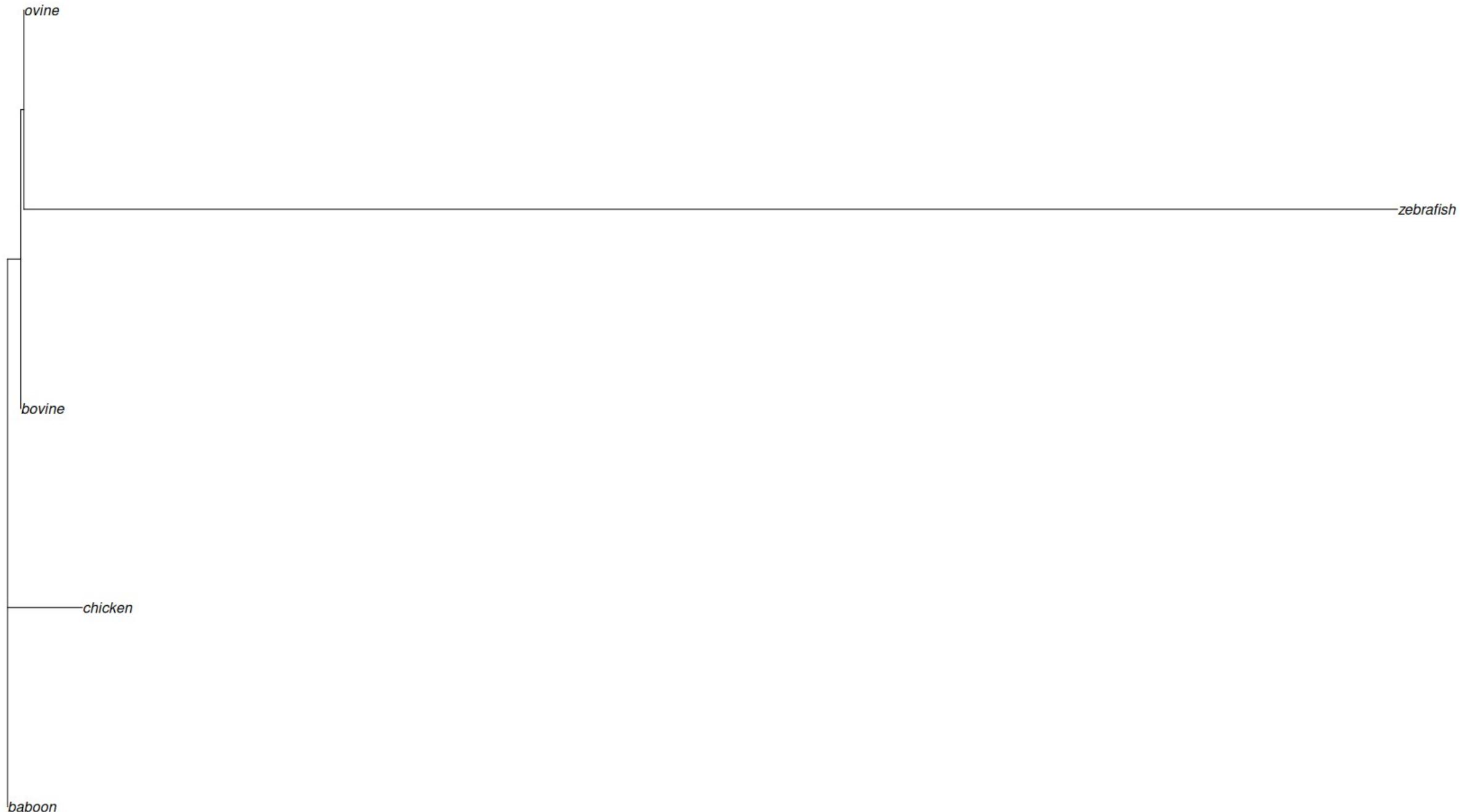
baboon, chicken, bovine, zebrafish, ovine

Unrooted; includes branch lengths.

```
> plot(albero, main = 'GTR model')
```

L'albero più verosimile

GTR model



```
> alb = read.tree(text = "((bovino:1, ovino:1):1, babbuino:1);")  
> alb
```

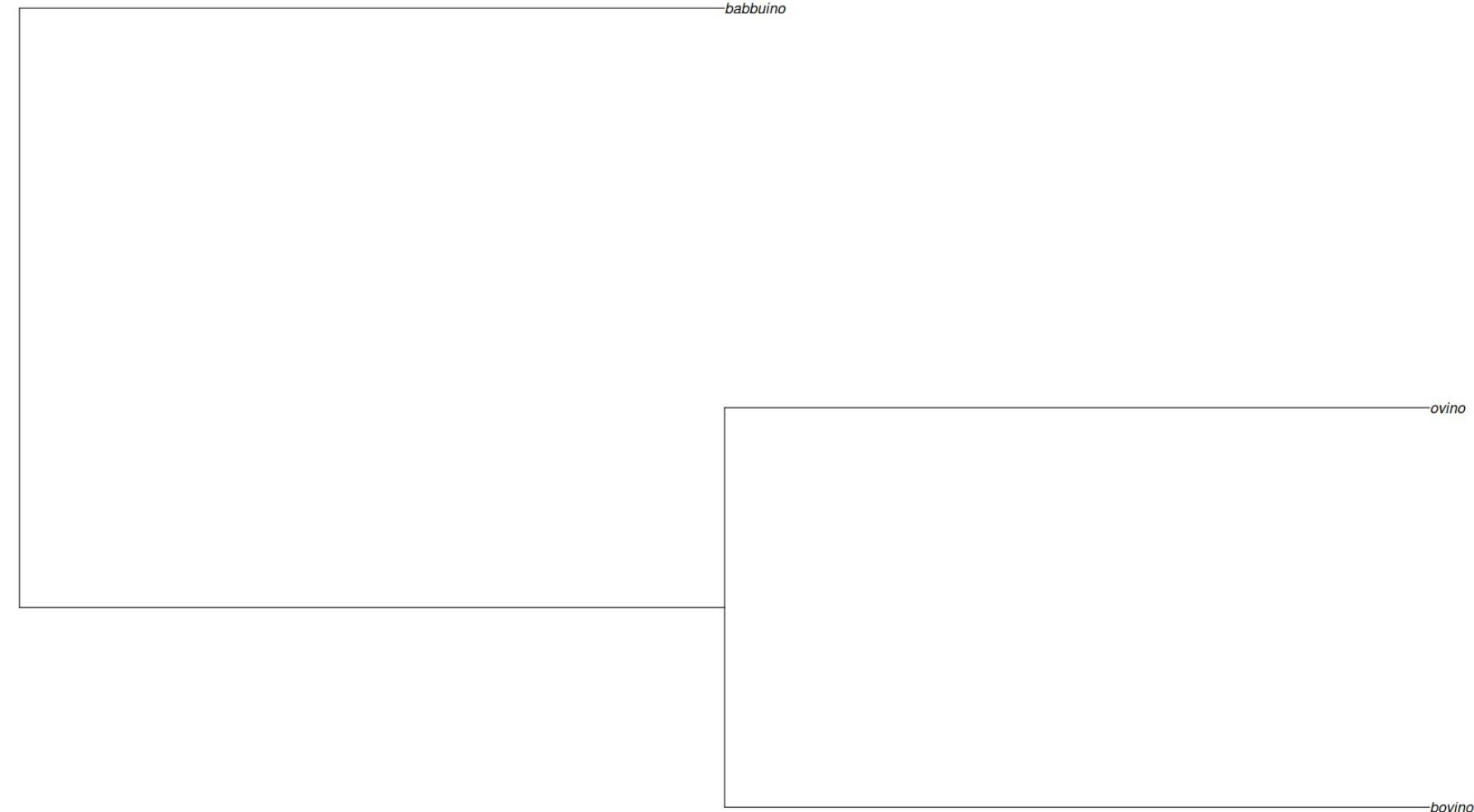
Phylogenetic tree with 3 tips and 2 internal nodes.

Tip labels:

bovino, ovino, babbuino

Rooted; includes branch lengths.

```
> plot(alb)
```



```
> alb = read.tree(text = "((bovino:1, ovino:1):1, babbuino:2);")  
> alb
```

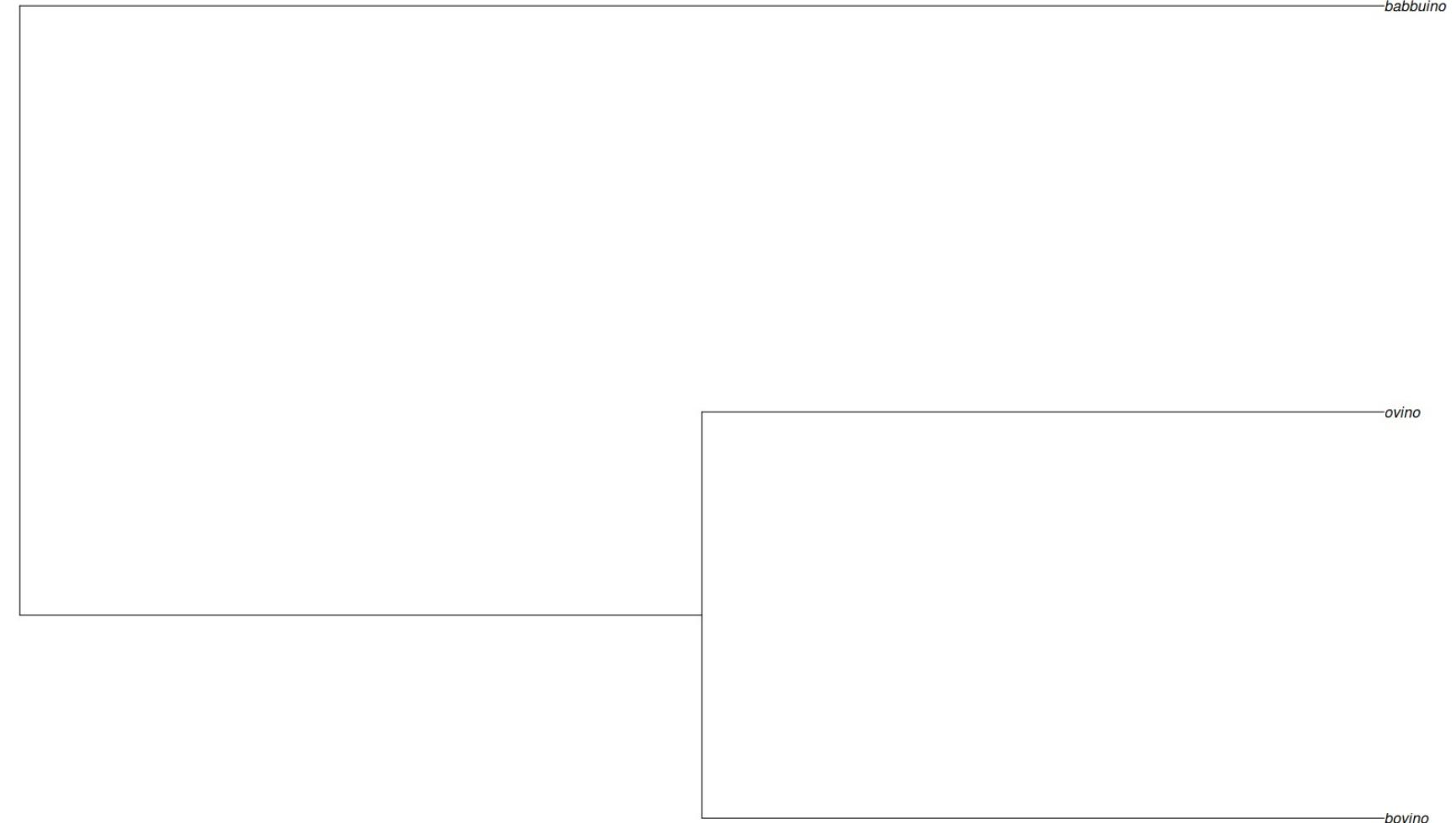
Phylogenetic tree with 3 tips and 2 internal nodes.

Tip labels:

bovino, ovino, babbuino

Rooted; includes branch lengths.

```
> plot(alb)
```



```
> alb = read.tree(text = "(((bovino:1, ovino:1):1, babbuino:2):2, gallo:3);")  
> alb
```

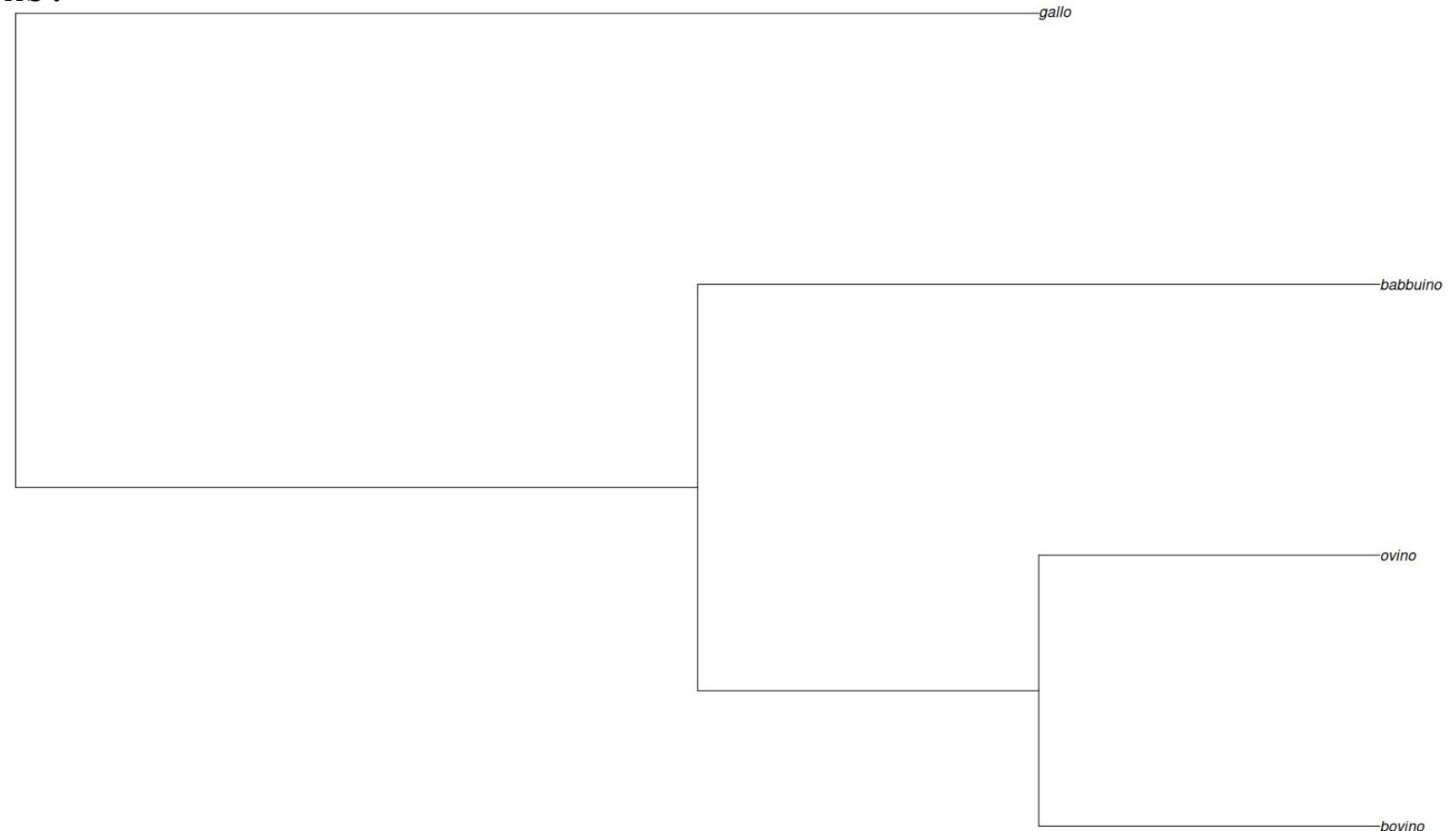
Phylogenetic tree with 4 tips and 3 internal nodes.

Tip labels:

bovino, ovino, babbuino, gallo

Rooted; includes branch lengths.

```
> plot(alb)
```



```
> alb = read.tree(text = "(((bovino:1, ovino:1):1, babbuino:2):2, gallo:4);")  
> alb
```

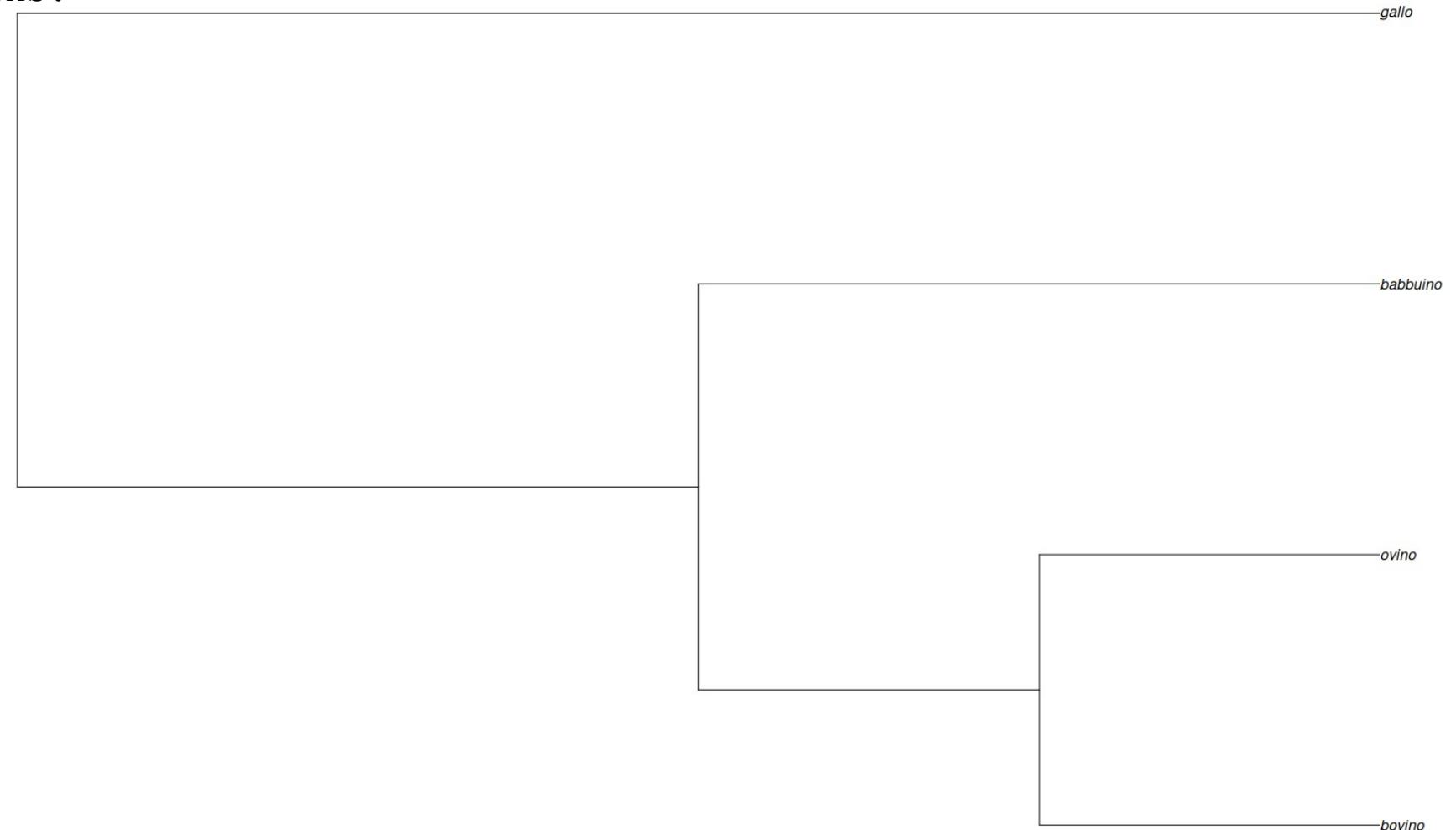
Phylogenetic tree with 4 tips and 3 internal nodes.

Tip labels:

bovino, ovino, babbuino, gallo

Rooted; includes branch lengths.

```
> plot(alb)
```



```
> alb = read.tree(text = "(((bovino:1, ovino:1):1, babbuino:2):2, gallo:4):3, pesce_zebra:7;")  
> alb
```

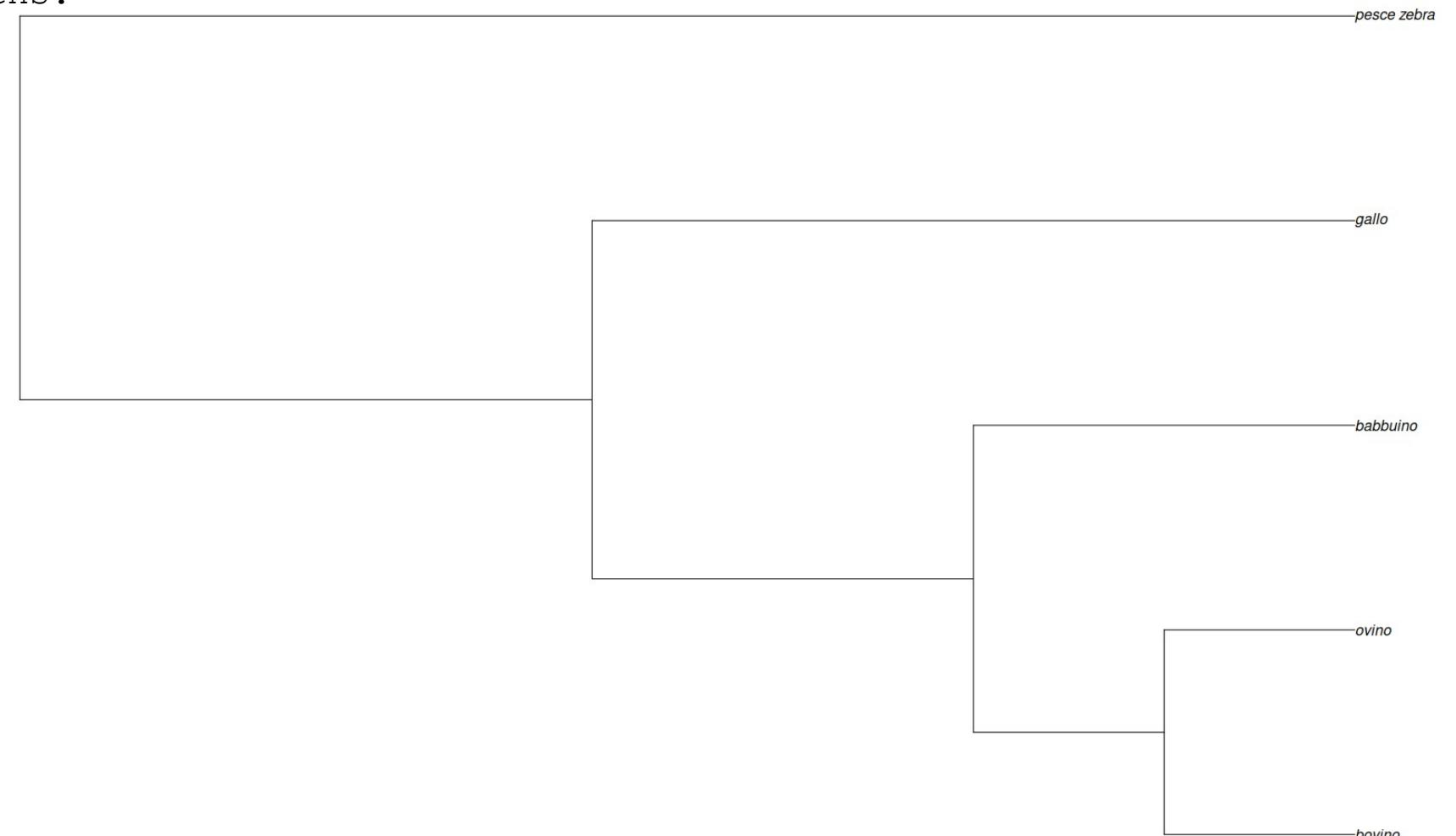
Phylogenetic tree with 5 tips and 4 internal nodes.

Tip labels:

bovino, ovino, babbuino, gallo, pesce_zebra

Rooted; includes branch lengths.

```
> plot(alb)
```



STATISTICA

L'effetto additivo dei fattori genetici ed ambientali

gene normale A -> effetto fenotipico = 8 p = 0,5

allele mutante A' -> effetto fenotipico = 4 q = 0,5

omozigoti AA -> valore fenotipico = 16 F = 0,25

omozigoti A'A' -> valore fenotipico = 8 F = 0,25

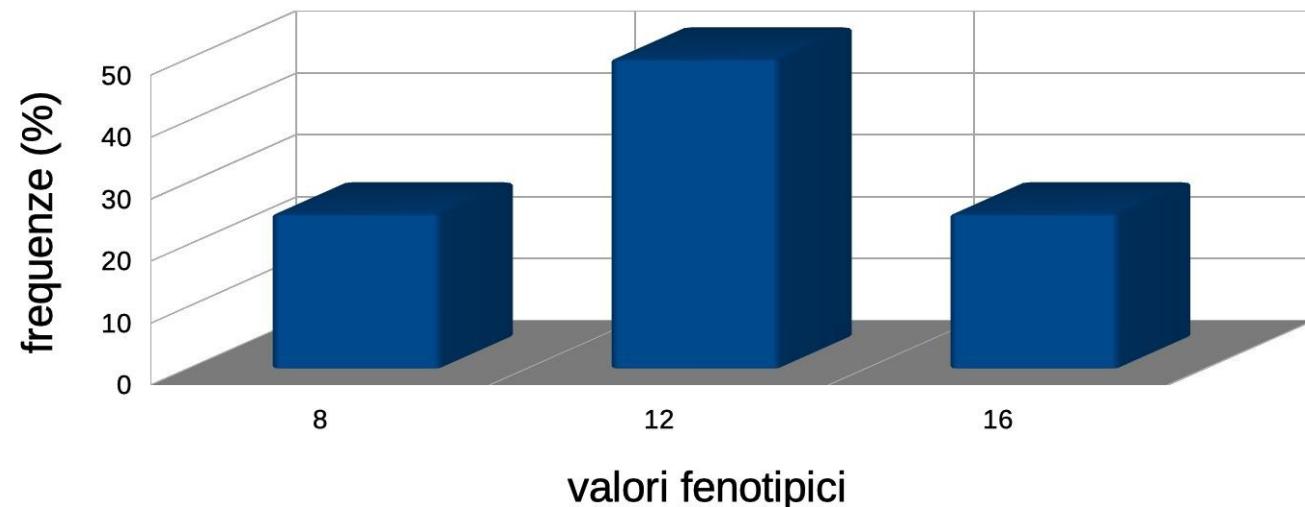
eterozigoti AA' -> valore fenotipico = 12 F = 0,25

eterozigoti A'A -> valore fenotipico = 12 F = 0,25

L'effetto additivo dei fattori genetici ed ambientali

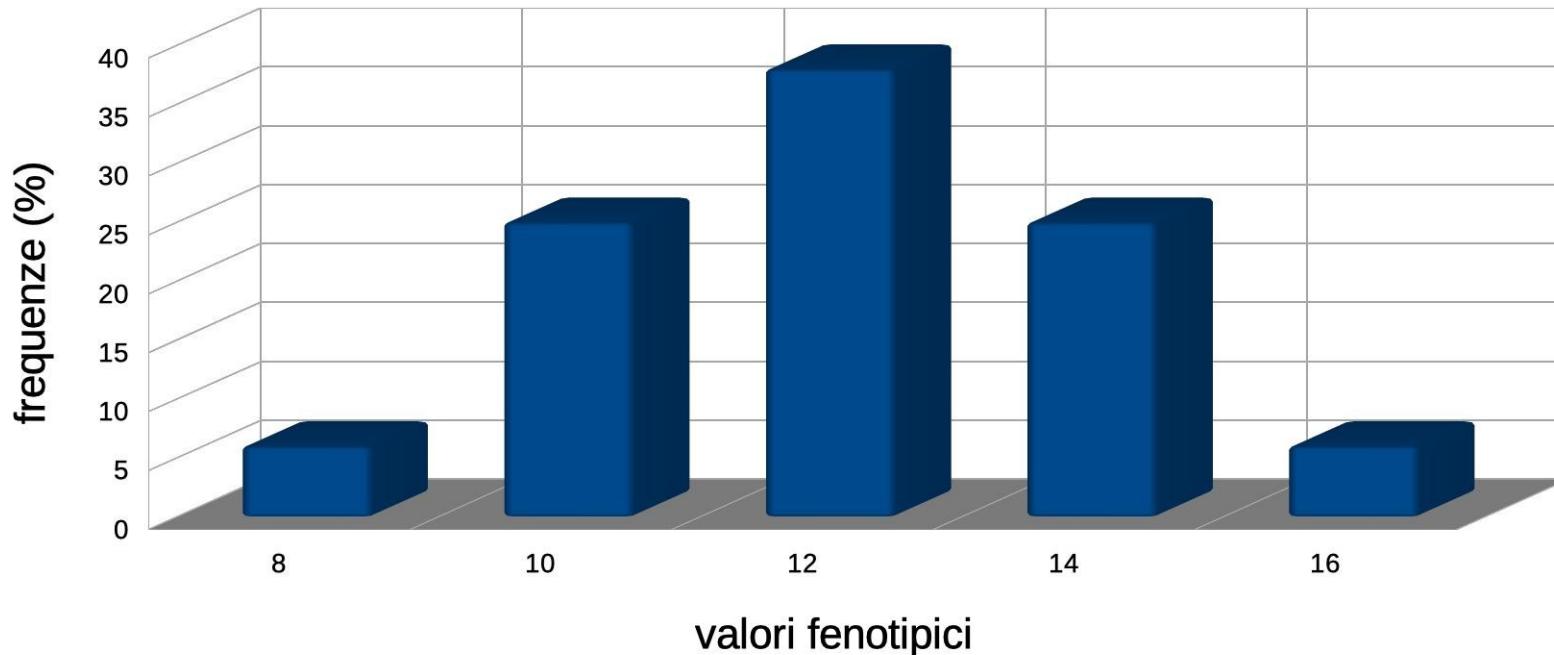
CLASSE GENOTIPICA	VALORE FENOTIPICO	FREQUENZA
AA	16	1/4 = 25%
AA' - A'A	12	2/4 = 50%
A'A'	8	1/4 = 25%

DISTRIBUZIONE DEI VALORI FENOTIPICI CON 1 COPPIA ALLELICA



Due loci

DISTRIBUZIONE DEI VALORI FENOTIPICI CON 2 COPPIE ALLELICHE



Valori e frequenze dei possibili genotipi nel caso in cui un carattere quantitativo è controllato da due loci («a» e «B»)

Tre loci

aumentando ulteriormente il numero di coppie alleliche interessate, aumenta in numero delle classi di valori ma diminuiscono le differenze fra i valori delle varie classi

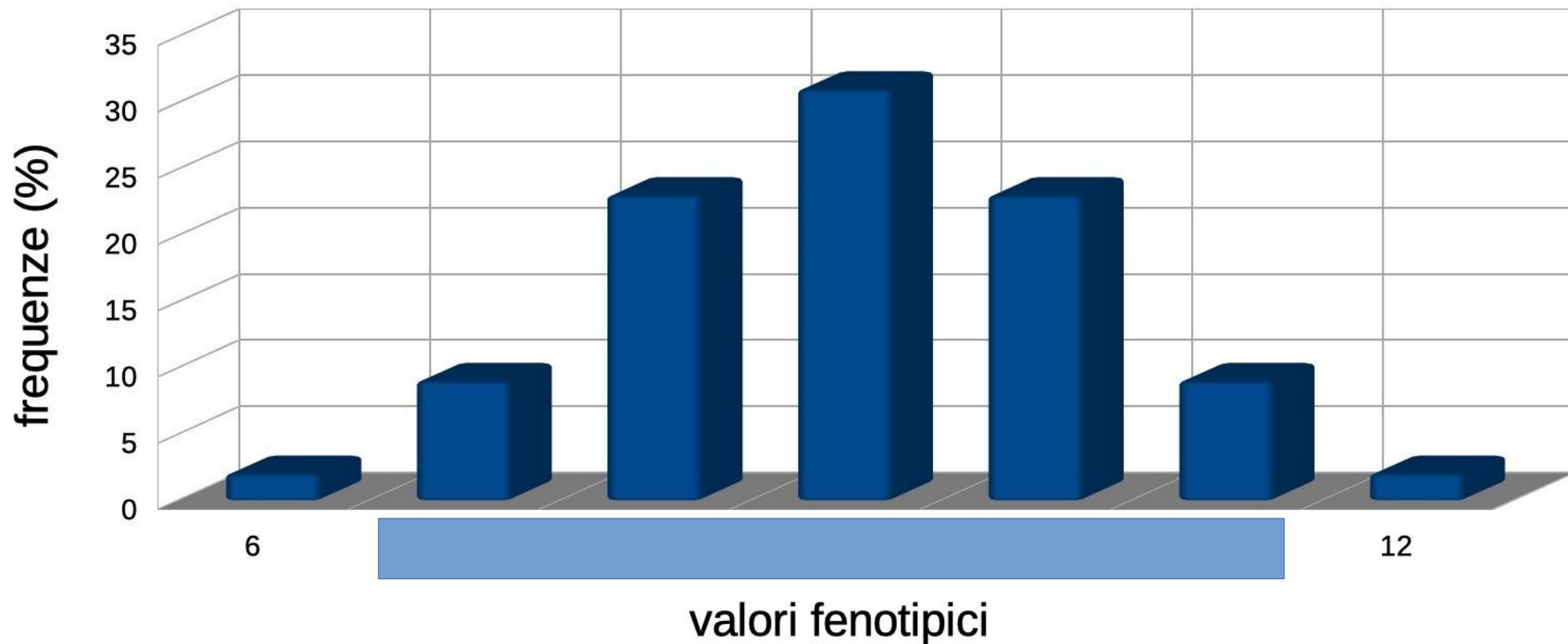
Se si considerano tre coppie di geni, a-a', B-B' e C-C', dove gli alleli «normali» a, B e C hanno effetto 2 e gli alleli «mutanti» a', B' e C' effetto 1, le classi genotipiche e i relativi valori fenotipici saranno:

Tre loci

Quali sono i valori fenotipici intermedi?

CLaSSE GENOTIPICa	VaLORE FENOTIPICO	FREQUENZA
aaBBCC	12	1/64 = 2%
aaBBCC' – aaBBC'C – aaBB'CC aaB'BCC – aa'BBCC – a'aBBCC		6/64 = 9%
aaBBC'C' – aaBB'CC' – aaB'BCC' aa'BBCC' – a'aBBCC' – aaBB'C'C aaB'BC'C – aa'BBC'C – a'aBBC'C aaB'B'CC – aa'BB'CC – a'aBB'CC aa'B'BCC – a'aB'BCC – a'a'BBCC		15/64 = 23%
aaBB'C'C' – aaB'BC'C' – aaB'B'CC' aaB'B'C'C – aa'BBC'C' – aa'BB'CC' aa'B'BCC' – aa'BB'C'C – aa'B'BC'C aa'B'B'CC – a'aBBC'C' – a'aBB'CC' a'aB'BCC' – a'aBB'C'C – a'aB'BC'C a'aB'B'CC – a'a'B'BCC – a'a'BB'CC, a'aBBC'C – a'a'BBCC'		2a/64 = 31%
aaB'B'C'C' – aa'BB'C'C' – a'aBB'C'C' a'aBBC'C' – aa'B'BC'C' – a'aB'BC'C' a'a'B'BCC' – aa'B'B'CC' – a'aB'B'CC' a'a'B'BC'C – aa'B'B'C'C – a'aB'B'C'C a'a'B'B'CC – a'a'BB'CC' – a'a'BB'C'C		15/64 = 23%
aa'B'B'C'C' – a'aB'B'C'C' – a'a'B'B'C'C a'a'B'B'CC' – a'a'BB'C'C' – a'a'B'BC'C'		6/64 = 9%
a'a'B'B'C'C'	6	1/64 = 2%

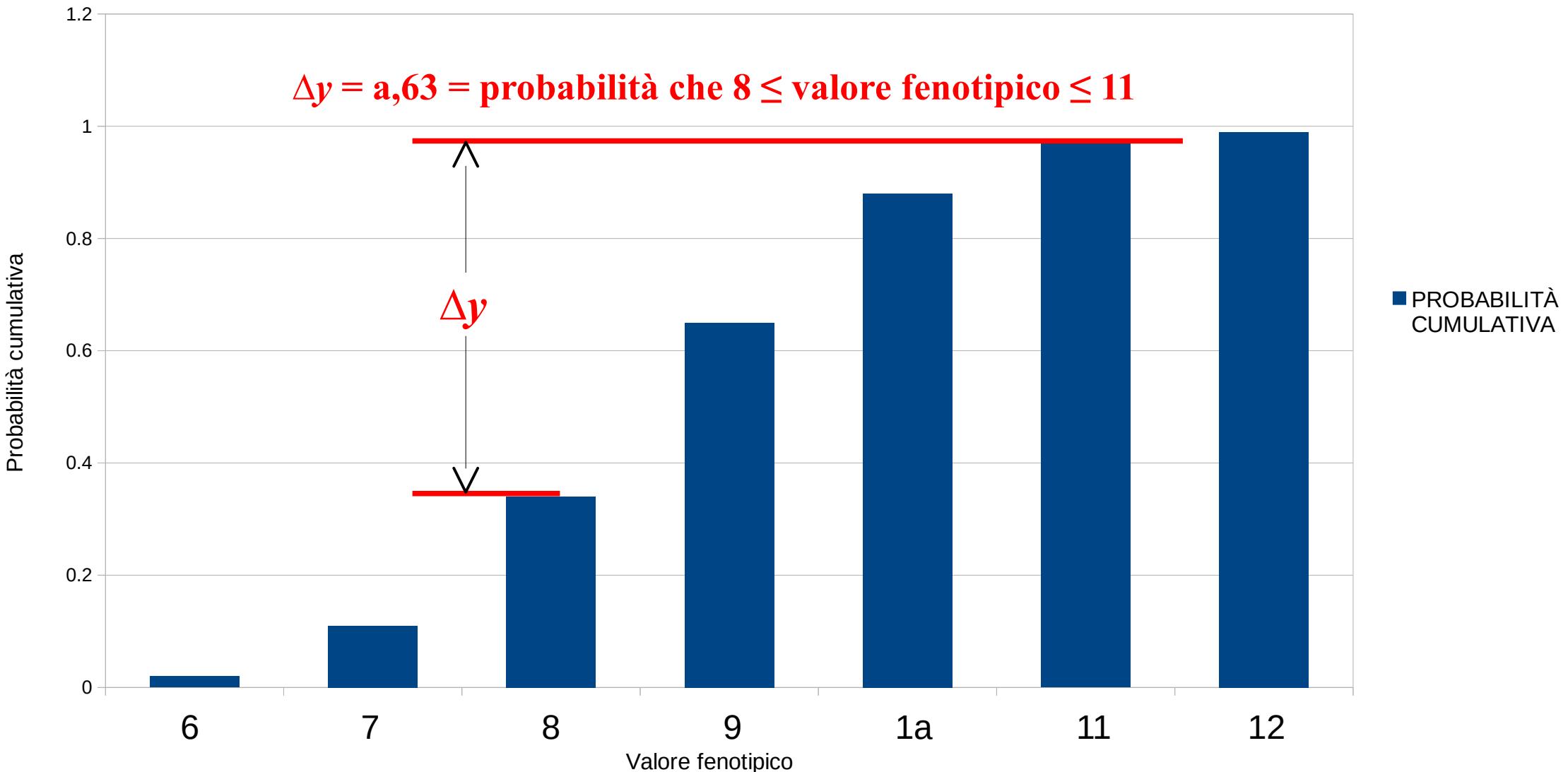
DISTRIBUZIONE DEI VALORI FENOTIPICI CON 3 COPPIE ALLELICHE



Tre loci

Distribuzione di probabilità cumulativa

Carattere quantitativo controllato da 3 loci

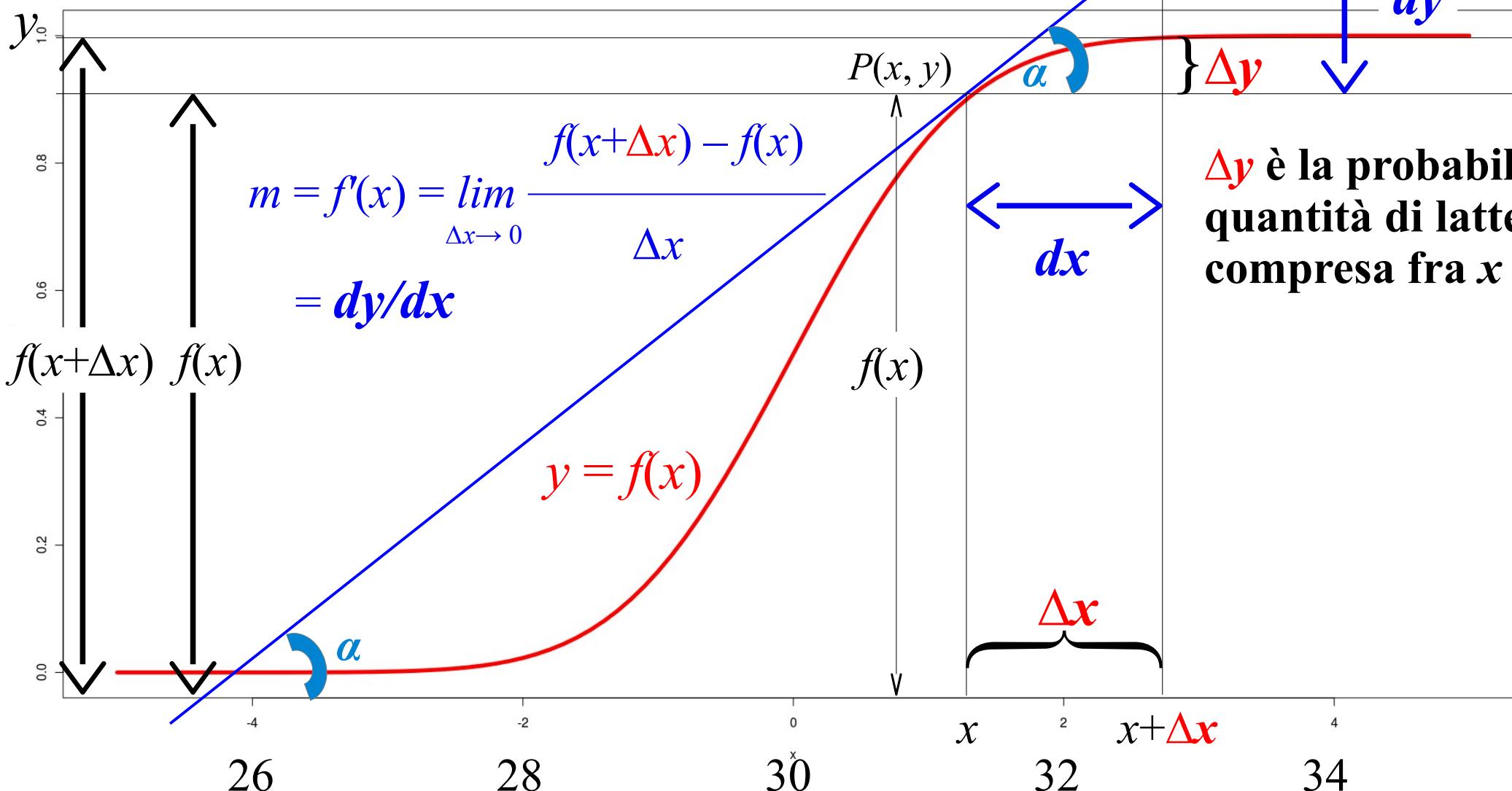


Un numero (quasi) infinito di fattori ambientali e genetici

$$y = mx + q$$

$y = f(x)$ → distribuzione di probabilità cumulativa

```
> curve(pnorm(x), -5, 5, col="red", lwd=5")
```



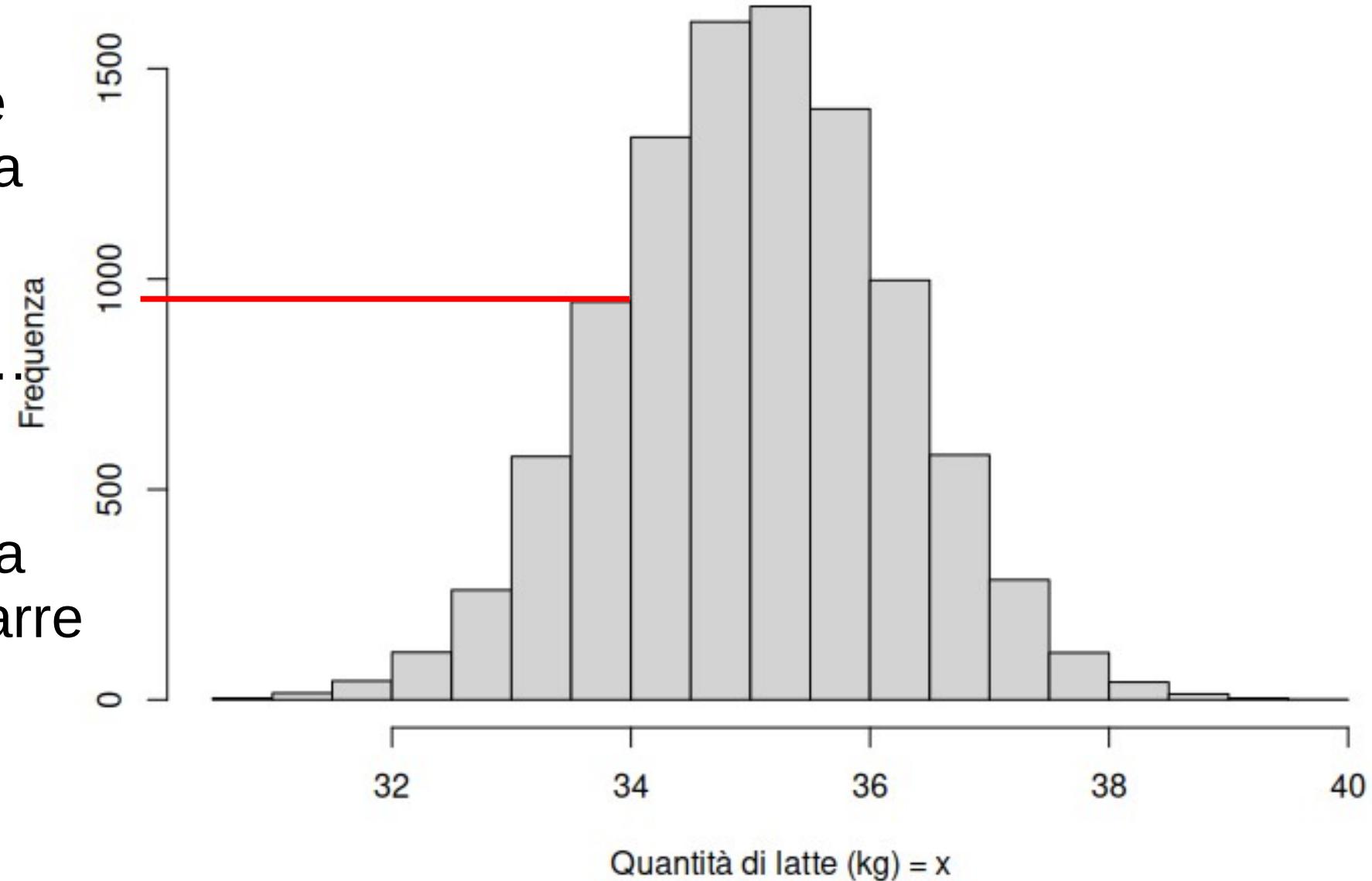
Δy è la probabilità che la quantità di latte sia compresa fra x e $x+\Delta x$

Un numero (quasi) infinito di fattori ambientali e genetici

Il N° di campioni compresi fra 33,5 e 34 è uguale a ≈ 95

Quindi l'h di ogni barra rappresenta...

Se ci sono 1aaaa campioni, la somma delle h di tutte le barre sarà uguale a...

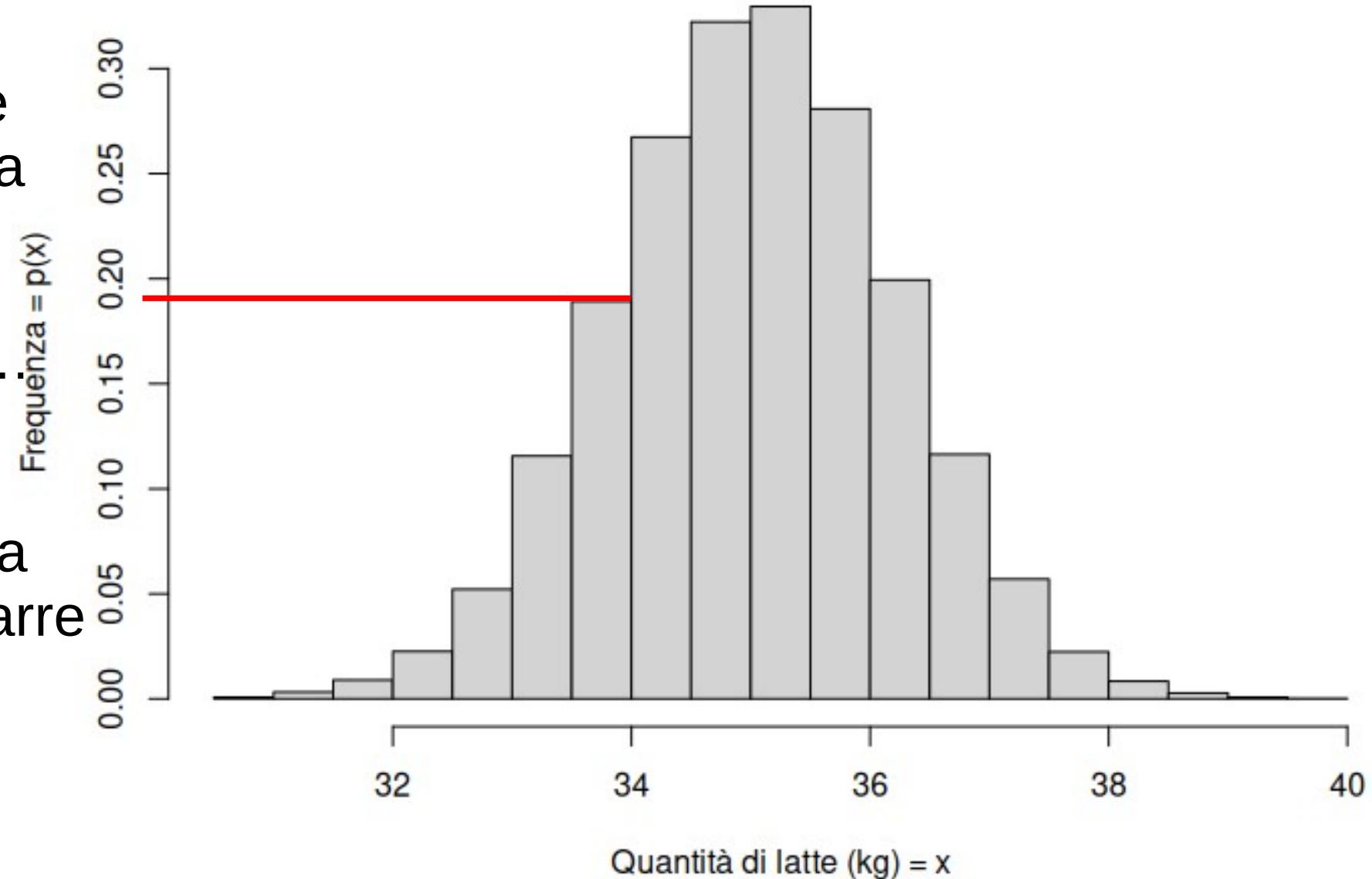


Un numero (quasi) infinito di fattori ambientali e genetici

Il N° di campioni compresi fra 33,5 e 34 è uguale a $\approx 95a$

Quindi l'h di ogni barra rappresenta...

Se ci sono 1aaaa campioni, la somma delle h di tutte le barre sarà uguale a...

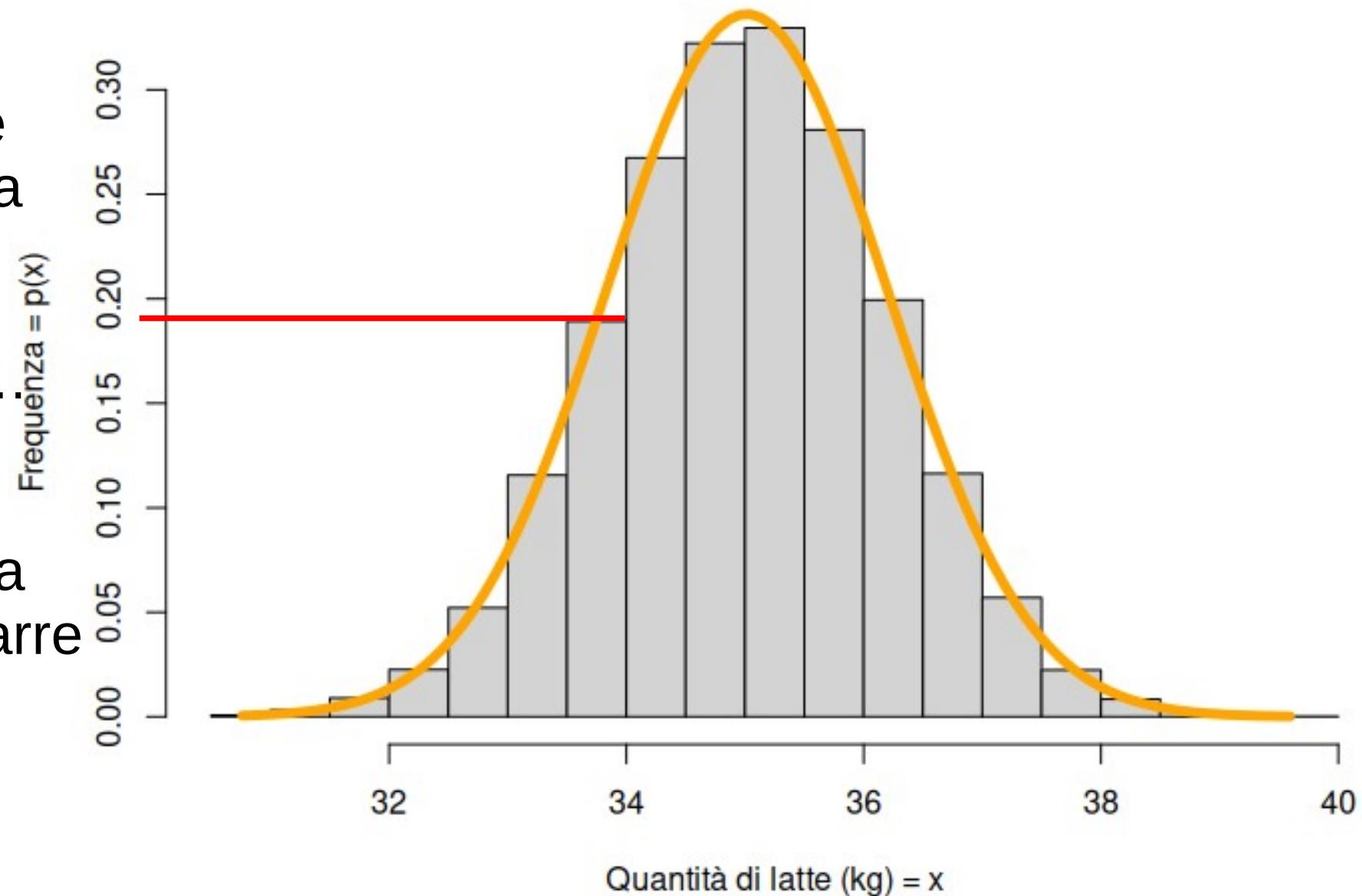


Un numero (quasi) infinito di fattori ambientali e genetici

Il N° di campioni compresi fra 33,5 e 34 è uguale a $\approx 95\alpha$

Quindi l'h di ogni barra rappresenta...

Se ci sono 1aaaa campioni, la somma delle h di tutte le barre sarà uguale a...



Un numero (quasi) infinito di fattori ambientali e genetici

per disegnare l'istogramma

```
> x = rnorm(1aaaa, mean = 35, sd = 1.2)
> hist(x, xlab = 'Quantità di latte (kg) = x', ylab = 'Frequenza')
> hist(x, prob = TRUE, xlab = 'Quantità di latte (kg) = x', ylab = 'Frequenza
= p(x)')
> x_values = seq(min(x), max(x), length = 1aa)
> fun = dnorm(x_values, mean = mean(x), sd = sd(x))
> lines(x_values, fun, col = 'orange', lwd = 5)
```

per disegnare la densità

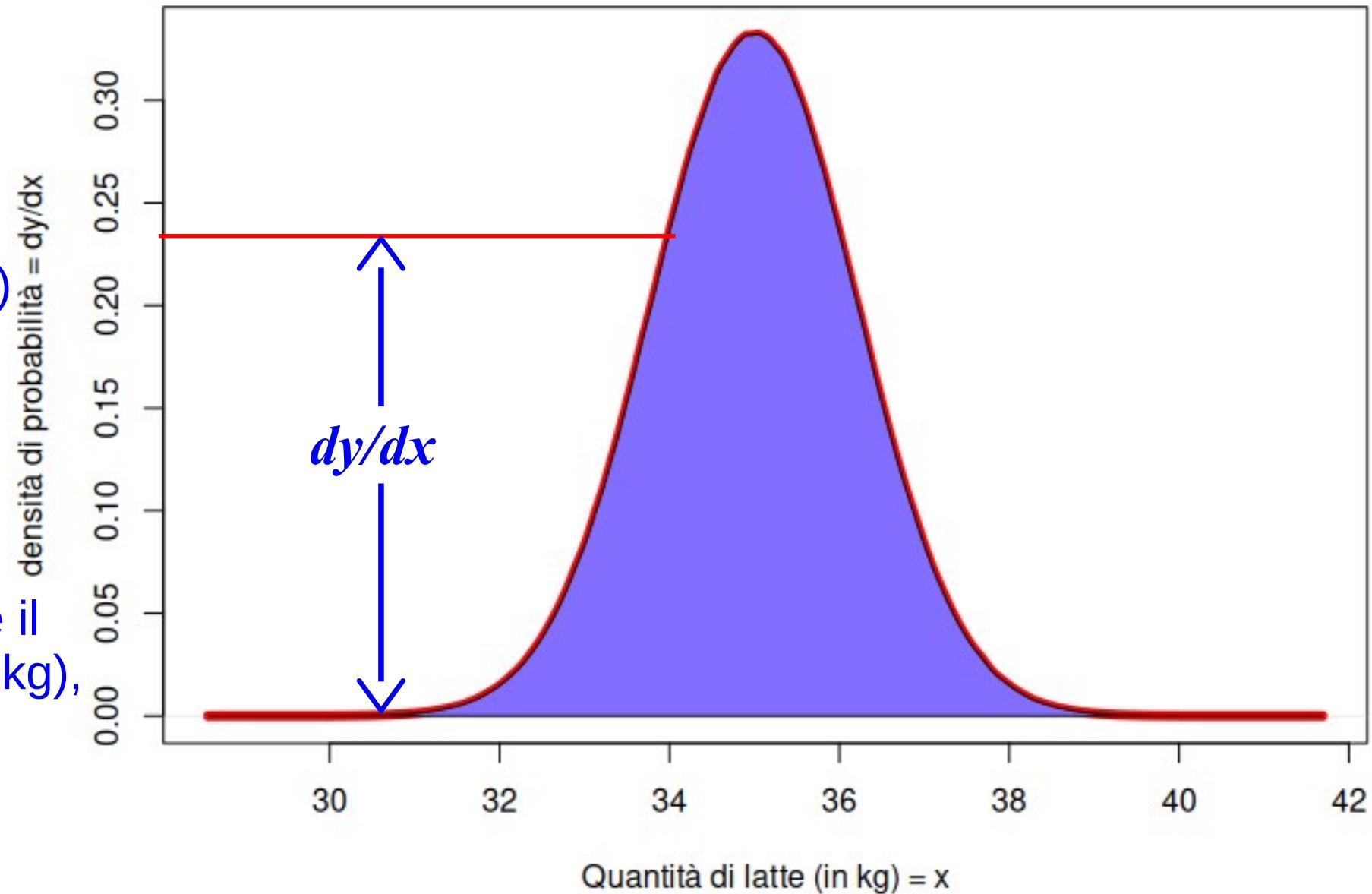
attenzione: x ha un significato diverso

```
> y = rnorm(1aaaaaaaa, mean = 35, sd = 1.2)
> den = density(y)
> plot(den, col = 'red', lwd = 5.a, xlab = 'Quantità di latte (in kg) = x',
ylab = 'densità di probabilità = dy/dx')
> polygon(den, col = 'slateblue1')
> value1 = 33
> value2 = 34
> l = min(which(den$x >= value1))
> h = max(which(den$x < value2))
> polygon(c(den$x[c(l, l:h, h)]), c(a, den$y[l:h], a), col = 'slateblue1')
```

Un numero (quasi) infinito di fattori ambientali e genetici

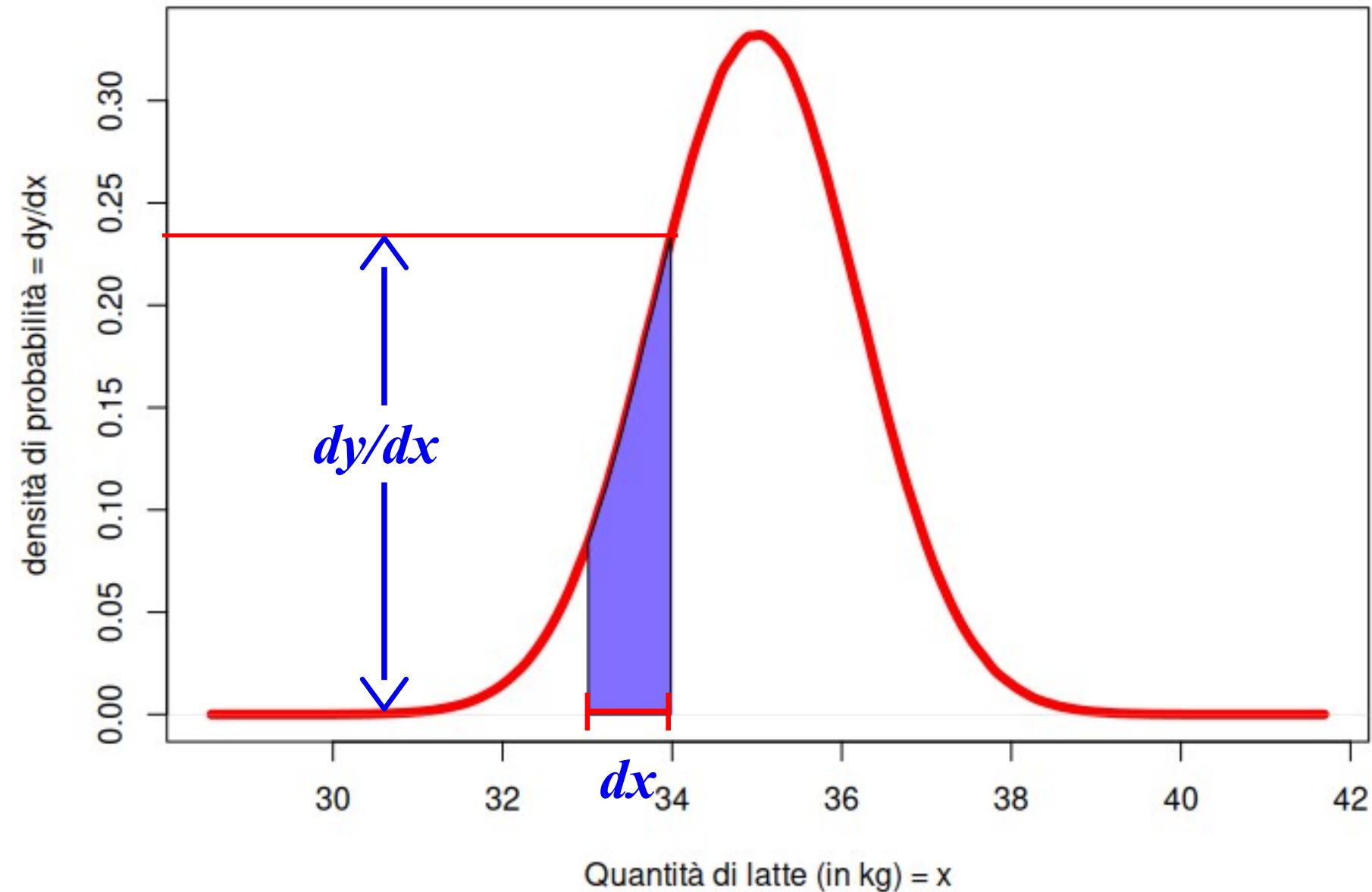
dy = probabilità che la quantità di latte sia compresa fra due valori, x e $x+dx$, in cui $dx \rightarrow 0$ (densità di probabilità)

L'area sotto la curva rappresenta la probabilità che x sia compreso tra il suo valore min ($\approx 3a$ kg) e il suo valore max ($\approx 4a$ kg), che equivale al 1aa% (evento certo).



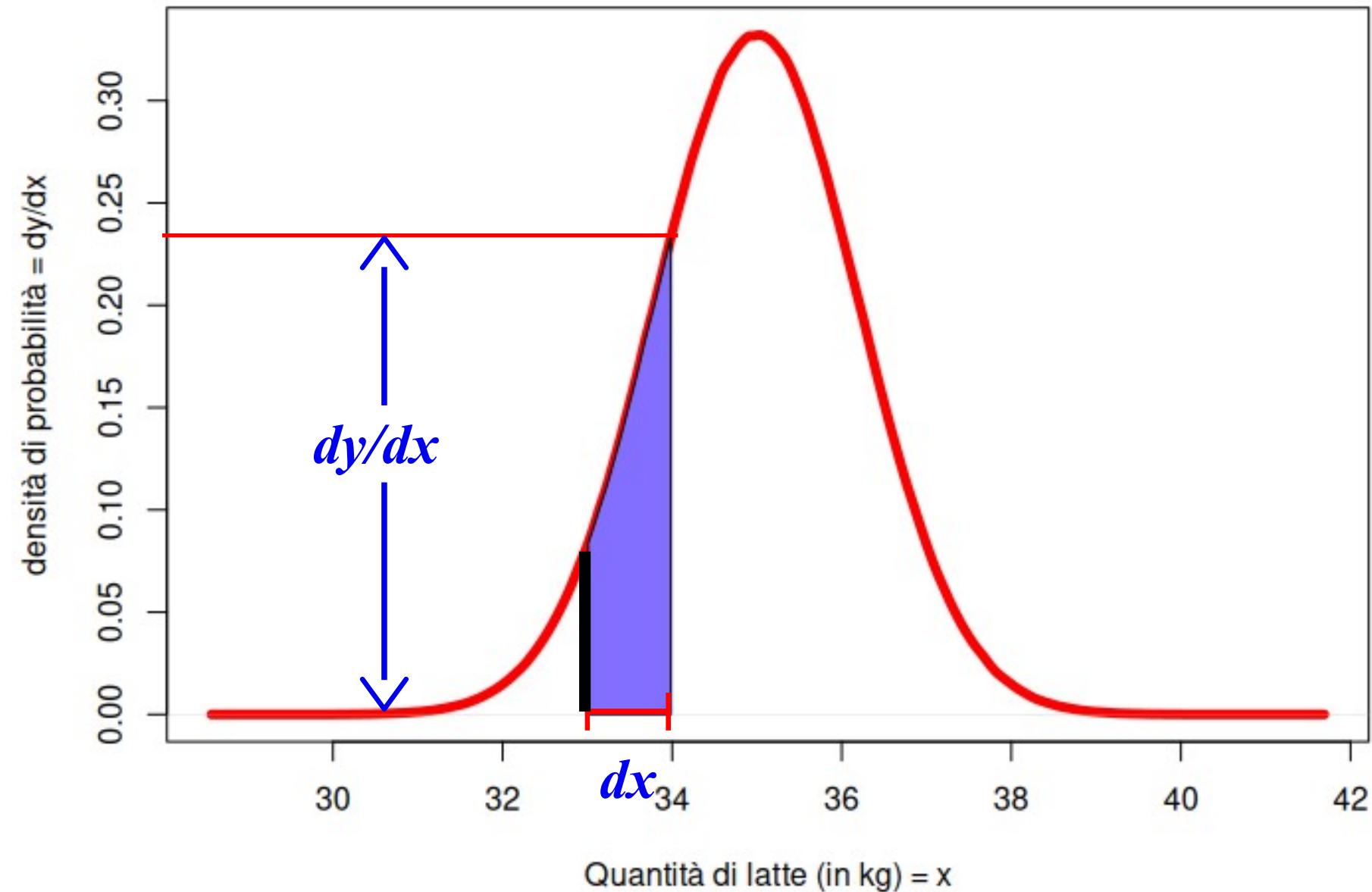
Un numero (quasi) infinito di fattori ambientali e genetici

La probabilità che la quantità di latte sia compresa fra 33 e 34 Kg è data dall'area blu, che può essere calcolata utilizzando un'operazione che prende il nome di «integrazione».



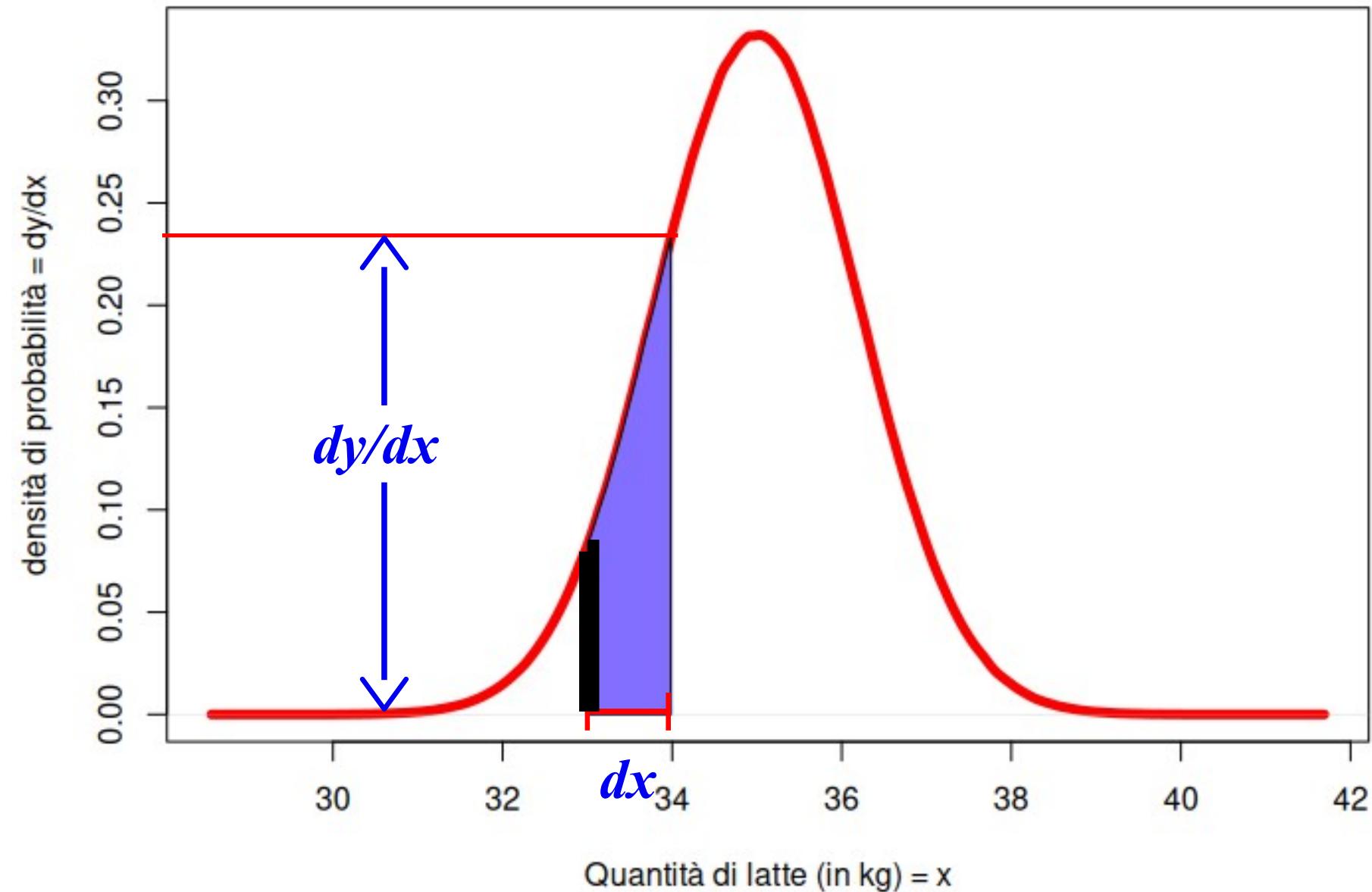
Un numero (quasi) infinito di fattori ambientali e genetici

La probabilità che la quantità di latte sia compresa fra 33 e 34 Kg è data dall'area blu, che può essere calcolata utilizzando un'operazione che prende il nome di «integrazione».



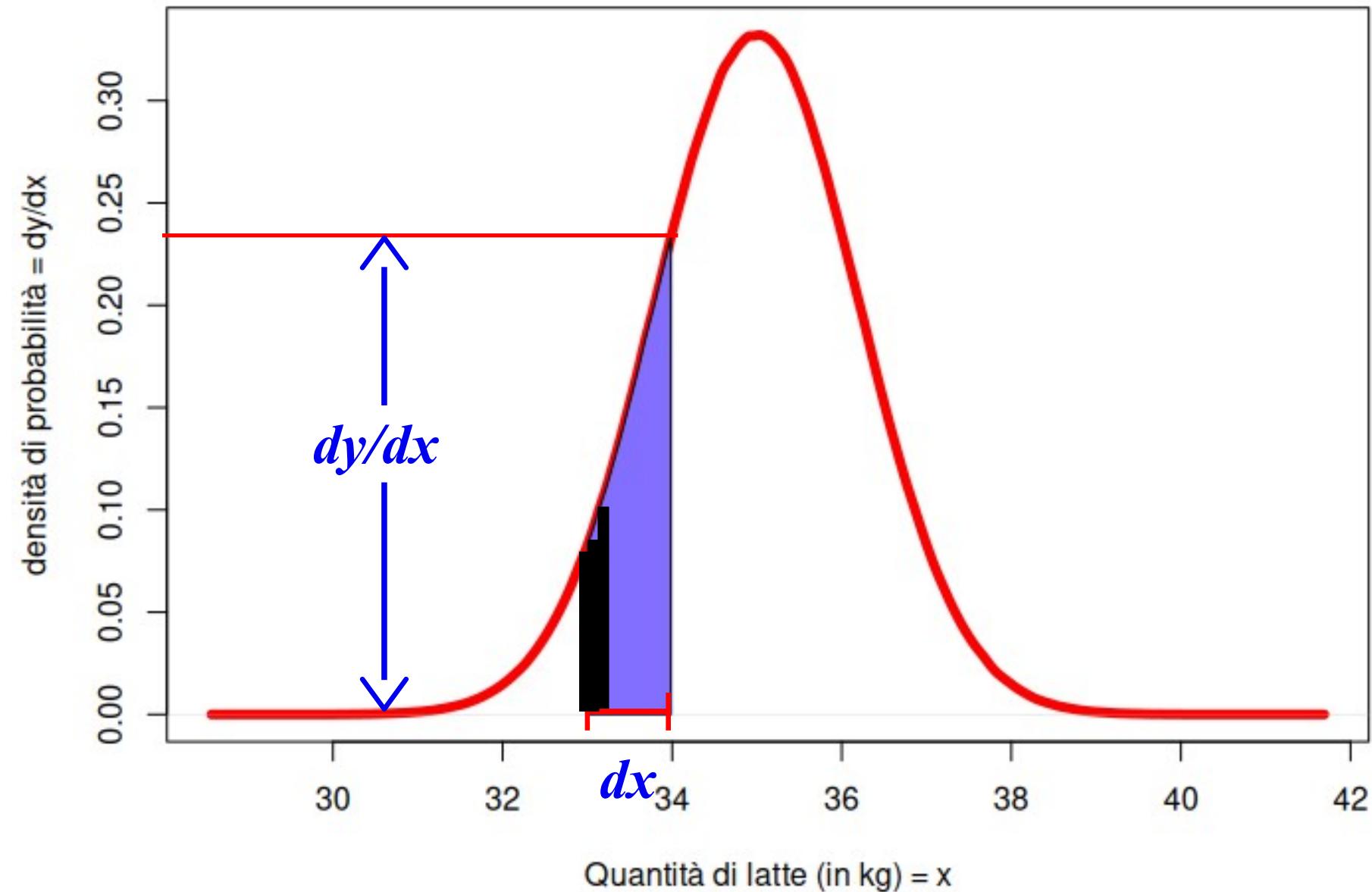
Un numero (quasi) infinito di fattori ambientali e genetici

La probabilità che la quantità di latte sia compresa fra 33 e 34 Kg è data dall'area blu, che può essere calcolata utilizzando un'operazione che prende il nome di «integrazione».



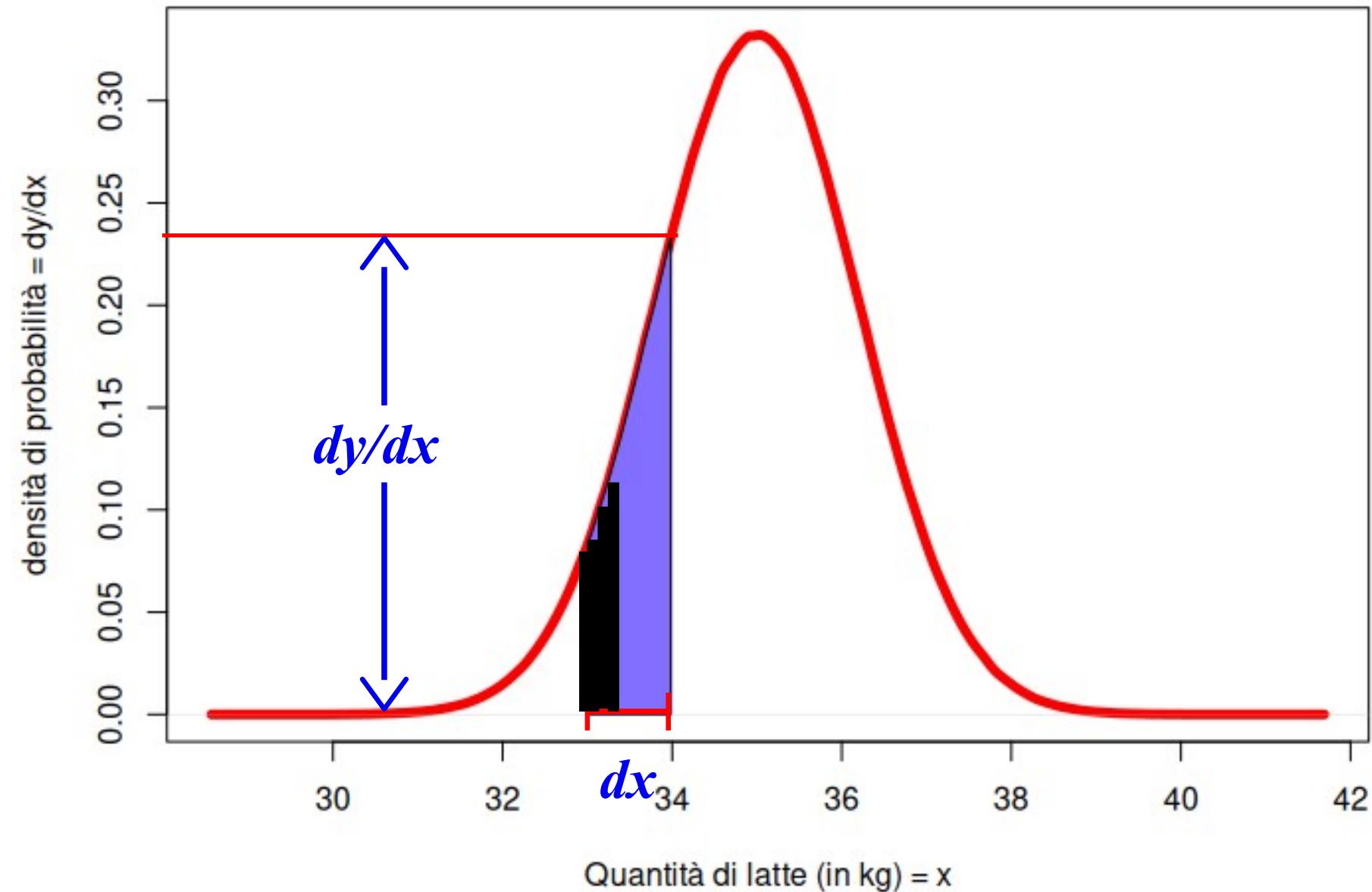
Un numero (quasi) infinito di fattori ambientali e genetici

La probabilità che la quantità di latte sia compresa fra 33 e 34 Kg è data dall'area blu, che può essere calcolata utilizzando un'operazione che prende il nome di «integrazione».



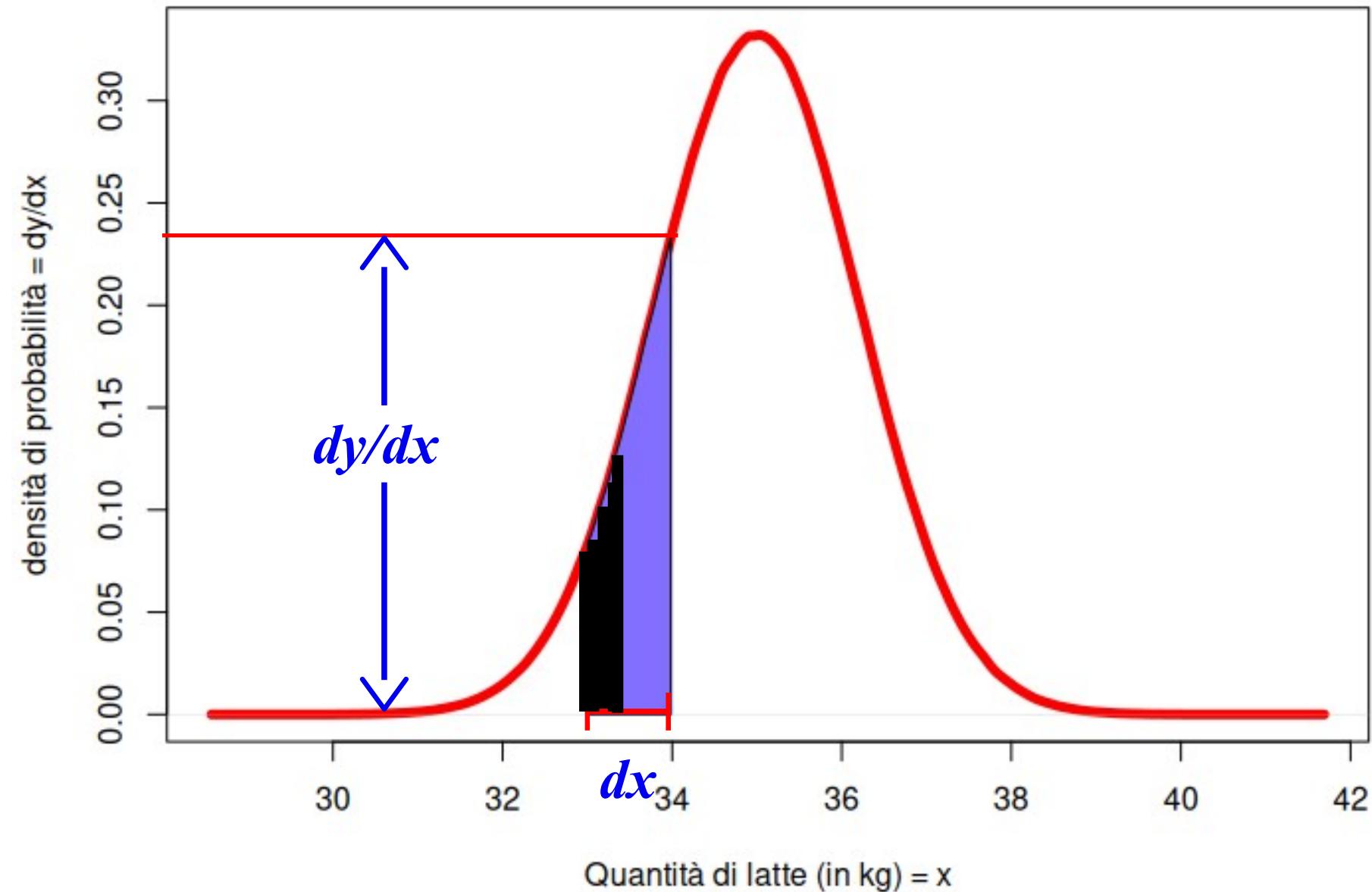
Un numero (quasi) infinito di fattori ambientali e genetici

La probabilità che la quantità di latte sia compresa fra 33 e 34 Kg è data dall'area blu, che può essere calcolata utilizzando un'operazione che prende il nome di «integrazione».



Un numero (quasi) infinito di fattori ambientali e genetici

La probabilità che la quantità di latte sia compresa fra 33 e 34 Kg è data dall'area blu, che può essere calcolata utilizzando un'operazione che prende il nome di «integrazione».



oooooooooooooooooooooooooooo

--- PhyML 20120412 ---

<http://www.atgc-montpellier.fr/phym>

Copyright CNRS - Universite Montpellier II

oooooooooooooooooooooooooooo

. Sequence filename: myostatin_5_species_100_bp.phy
. Data set: #1
. Tree topology search : NNIs
. Initial tree: BioNJ
. Model of nucleotides substitution: GTR
. Number of taxa: 5
. **Log-likelihood:** **-400.23791**
. Unconstrained likelihood: -343.80642
. Parsimony: 101
. Tree size: 10.41394
. Discrete gamma model:
- Number of categories: 4
- **Gamma shape parameter:** **0.997**
. Proportion of invariant: 0.023
. Nucleotides frequencies:
- $f(A) = 0.28600$
- $f(C) = 0.17800$
- $f(G) = 0.24000$
- $f(T) = 0.29600$
. GTR relative rate parameters :
A <-> C 345.16529
A <-> G 385.25348
A <-> T 0.29877
C <-> G 296.04247
C <-> T 390.59090
G <-> T 1.00000

oooooooooooooooooooooooooooooooooooo

--- PhyML 20120412 ---

<http://www.atgc-montpellier.fr/phym>

Copyright CNRS - Universite Montpellier II

oooooooooooooooooooooooooooo

. Instantaneous rate matrix :

[A-----C-----G-----T-----]

-0.99555	0.39721	0.59777	0.00057
0.63822	-1.84503	0.45935	0.74746
0.71234	0.34068	-1.05494	0.00191
0.00055	0.44949	0.00155	-0.45159

```
> x = seq(0, 10, by = 0.1)
> y = dgamma(x, shape = 0.997)
> plot(y, type = 'l', lwd = 5, col = 'red', xlab = 'Tasso di sostituzione nucleotidica', ylab = 'Densità di probabilità')
```

Densità di probabilità

0.2

0.0

0.8

0.6

0.4

0

20

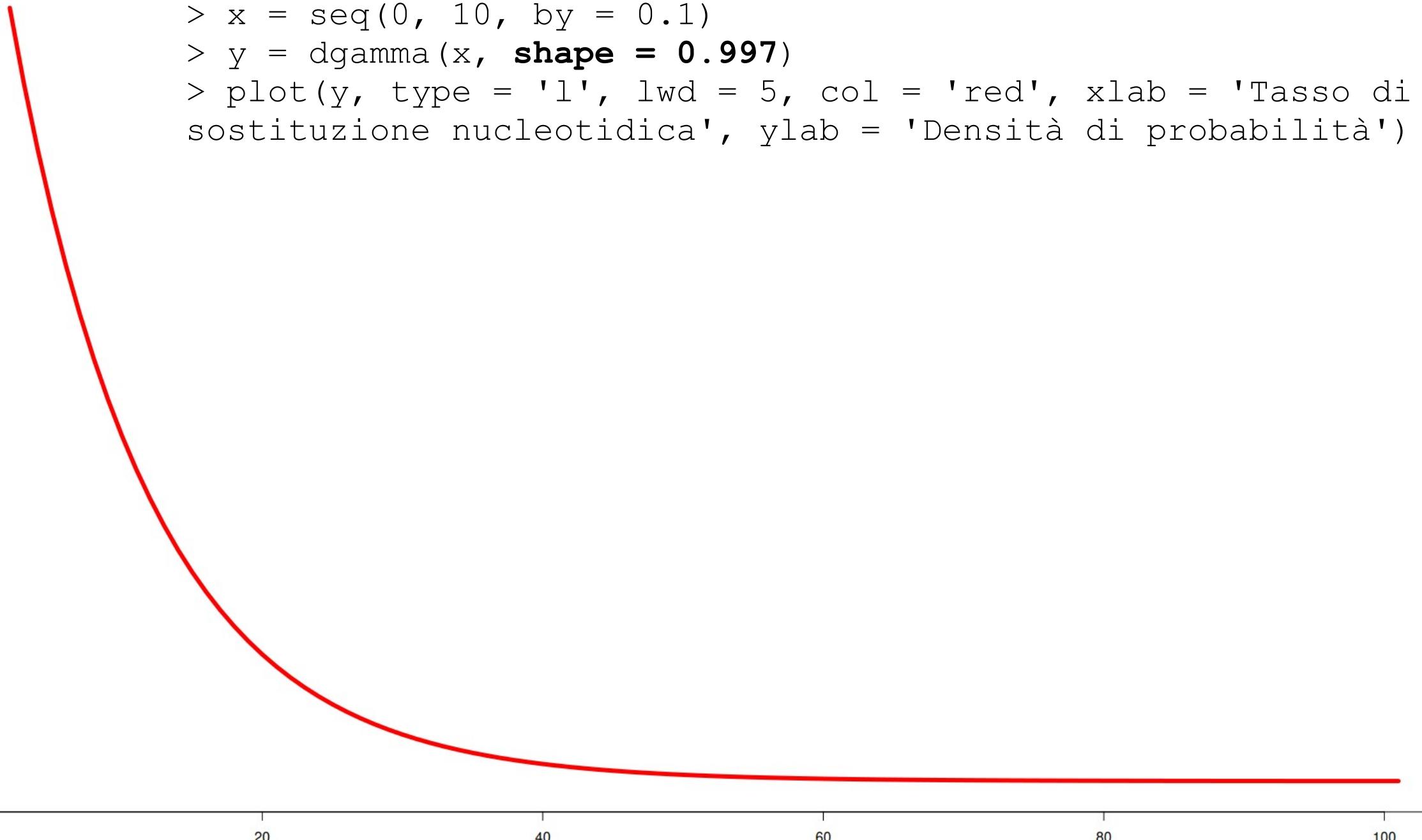
40

60

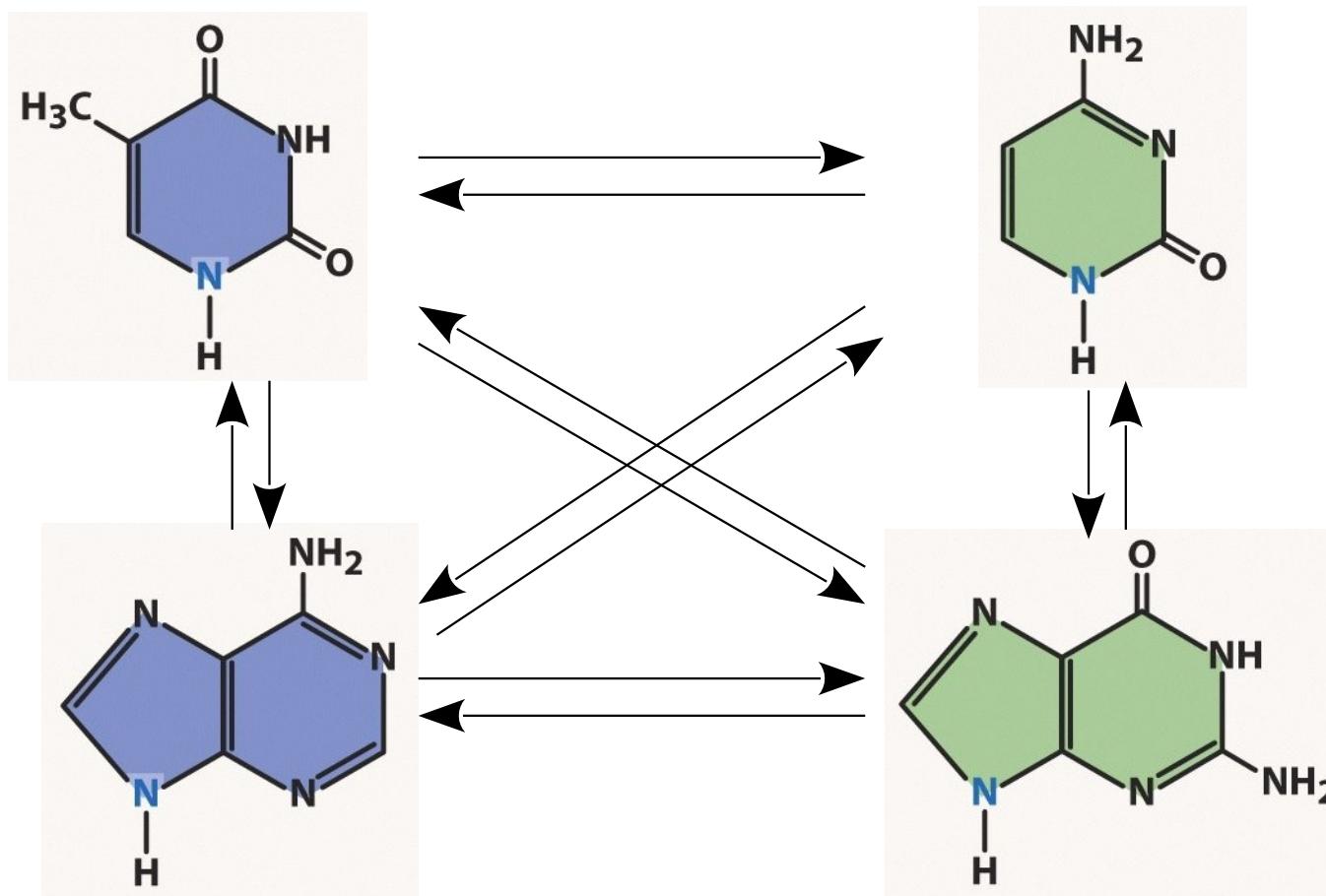
80

100

Tasso di sostituzione nucleotidica



Il modello JC69

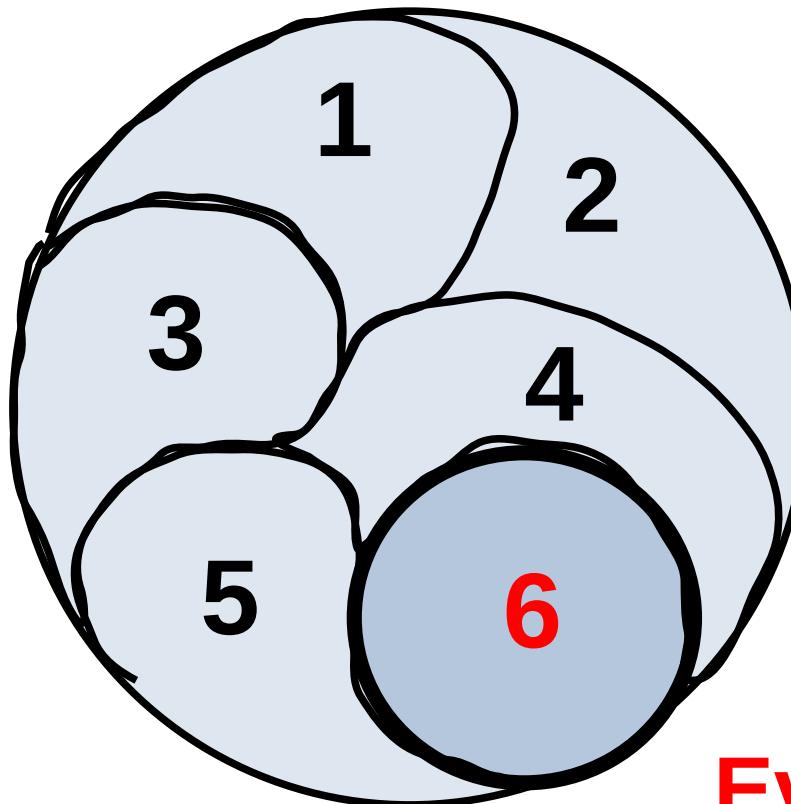


STATISTICA: concetti fondamentali

- **Esperimento:** processo che genera un'osservazione (per es., il lancio di un dado o di una moneta)
- **Evento:** ognuno dei possibili risultati di un esperimento (contrassegnati da lettere maiuscole)
- **Evento elementare:** evento che non può essere decomposto e che corrisponde ad un «punto campione» indivisibile

STATISTICA: spazio campione Ω

- Insieme di tutti i possibili risultati di un esperimento
Es. lancio di un dado
- 6 risultati equiprobabili (se il dado non è truccato)



$$\Omega = 1$$

Evento certo

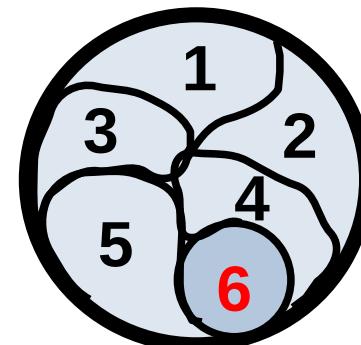
Evento elementare a

Cosa succede se lancio il dado una seconda volta?

- Vinco se esce di nuovo il numero 6: che probabilità ho di vincere?
- Lo spazio campione Ω diventa uguale ad 1/6
- Poiché si tratta di EVENTI INDIPENDENTI, la $P(6, 6) = 1/6 * 1/6 = 1/36$
- Probabilità congiunta di eventi indipendenti =
= prodotto delle probabilità dei singoli eventi

$$\Omega = 1/6$$

Evento certo



Evento elementare B

STATISTICA: probabilità di due eventi indipendenti

- a = evento elementare esce 6 al primo lancio
- B = evento elementare esce 6 al secondo lancio

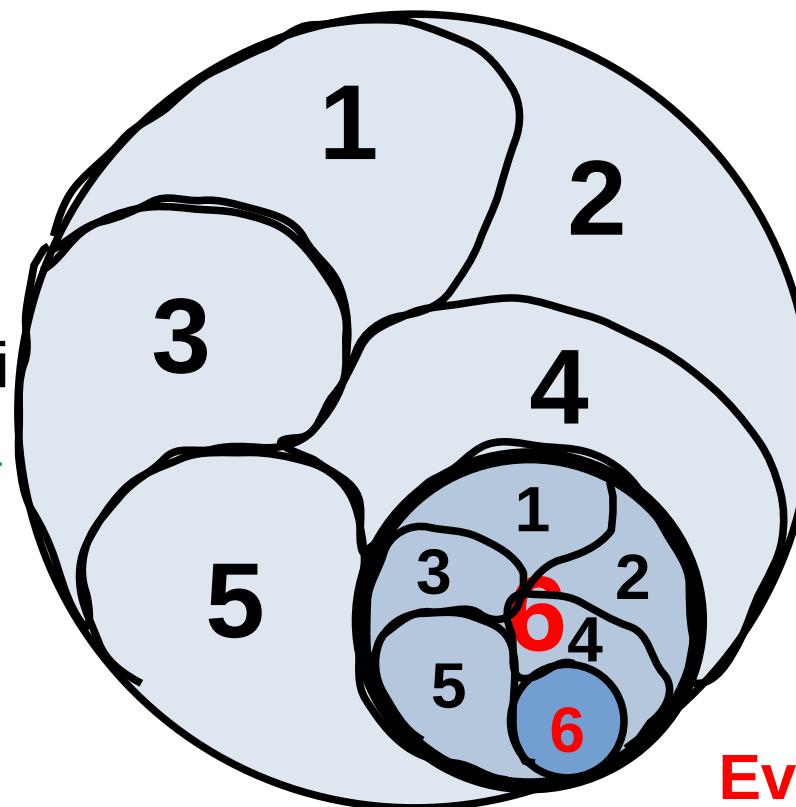
Quindi, noi vogliamo calcolare la Prob dell'intersezione fra il cerchio medio e il cerchio piccolo:

$$P(a \cap B) = P(a)*P(B|a)$$

Ma, poiché i due eventi sono indipendenti, $P(B|a) = P(B)$ e si applica la regola del prodotto

$$P(a \cap B) = P(a)*P(B)$$

$$P(a \cap B) = 1/6 * 1/6 = 1/36$$



$$\Omega = 1/6$$

Evento certo

Evento
elementare a

Evento elementare B

STATISTICA: la probabilità condizionata

Da

$$P(a \cap B) = P(a) * P(B|a)$$

si ricava:

$$P(B|a) = P(a \cap B) / P(a)$$

La sostituzione nucleotidica è un processo di Markov

Nel corso del tempo sono stati elaborati vari modelli di sostituzione nucleotidica, più complessi, i quali prevedono che i quattro differenti nucleotidi, che possono essere presenti ad ogni sito, siano gli stati di una catena di Markov.

Ogni nucleotide può essere sostituito da ogni altro ed anche da un nucleotide identico. Le probabilità di tali cambiamenti si chiamano **probabilità di transizione** e sono gli ingressi di una matrice quadrata $\mathbf{P}(t)$ di dimensioni 4×4 :

$$\mathbf{P} = \begin{matrix} & p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & & & & \\ p_{31} & & & & \\ p_{41} & & & & \end{matrix}$$
$$\begin{matrix} & p_{22} & p_{23} & p_{24} \\ p_{32} & & & \\ p_{42} & & & \end{matrix}$$
$$\begin{matrix} & p_{33} & p_{34} \\ p_{43} & & \end{matrix}$$
$$\begin{matrix} & p_{44} \end{matrix}$$

La sostituzione nucleotidica è un processo di Markov

I **processi di Markov**, chiamati così in onore del grande matematico russo Andrey Andreyevich Markov, sono catene di eventi che si presentano uno dopo l'altro in istanti successivi di tempo.

Nel nostro caso gli eventi sono le sostituzioni nucleotidiche casuali, in seguito alle quali lo stato X del carattere che si osserva ad un determinato sito nucleotidico, ossia la base azotata, può cambiare al trascorrere del tempo.

Esso dipende, però, soltanto dall'ultimo stato in cui il processo si trovava e non da quelli precedenti: la Prob che una certa base azotata venga sostituita da ciascuna delle altre tre dipende dalla base “attuale” e non da quelle che si trovavano in quella posizione della sequenza in passato.

$$P\{X_{t+1}=j \mid X_t=i, X_{t-1}=i-1, \dots, X_0=0\} = P\{X_{t+1}=j \mid X_t=i\} = P_{ij}$$

per tutti i valori di j , i , X_{t-1} , ..., X_0 e t . P_{ij} è costante nel tempo e non dipende da alcuno stato precedente; in altri termini, il processo non ha memoria degli stati precedenti a quello corrente, ad eccezione dell'ultimo (*proprietà di Markov*).

La sostituzione nucleotidica è un processo di Markov

$$P\{X_{t+1} = j \mid X_t = i, X_{t-1} = i-1, \dots, X_0 = 0\} = P\{X_{t+1} = j \mid X_t = i\} = P_{ij}$$

per tutti i valori di j , i , X_{t-1} , ..., X_0 e t .

P_{ij} = **probabilità di transizione** dallo stato i allo stato j (probabilità che il processo si sposti da uno stato ad un altro in un singolo passaggio) è costante nel tempo e non dipende da alcuno stato precedente; in altri termini, il processo non ha memoria degli stati precedenti a quello corrente, ad eccezione dell'ultimo (*proprietà di Markov*).

X_{t+1} = stato del processo nell'istante $t+1$

Ogni processo di Markov ha uno *spazio campione* Ω , definito come l'insieme, finito, dei possibili stati del processo (spazio degli stati):

$$\Omega = \{1, 2, \dots, n\}$$

con n stati possibili.

Stati transitori e stati definitivi

Gli **stati transitori**, dopo essere stati visitati, possono essere abbandonati mentre da quelli **definitivi** la catena, una volta entrata, non può più uscire. Gli stati definitivi, per cui $P_{ii} = 1$, si presentano nella matrice di transizione \mathbf{P} come altrettanti 1 (uno) collocati nella diagonale principale, mentre in tutti gli altri ingressi ci sono solo zeri.

Per gli stati transitori $P_{ii} < 1$ cosicché almeno qualche altro ingresso della i -esima riga è > 0 .

Per definizione, la catena deve trovarsi in uno degli stati possibili ad ogni *step*; per qualsiasi valore di i e j , $0 \leq P_{ij} \leq 1$ e

$$\sum_{j=1}^n P_{ij} = 1$$

per tutti gli i , ossia la somma delle probabilità che il processo si sposti dallo stato attuale i ad uno stato j qualsiasi (che può anche coincidere con i stesso) è uguale a 1.

La sostituzione nucleotidica è un processo di Markov

```
> mdat = matrix(c(2,1,3,4,1,3,1,5,7,1,1,1,4,2,1,3), nrow = 4, ncol = 4, byrow = TRUE)
> mdat
     [,1] [,2] [,3] [,4]
[1,]    2    1    3    4
[2,]    1    3    1    5
[3,]    7    1    1    1
[4,]    4    2    1    3
>
```

Come passare da i a j in due passaggi

p_{11} p_{12} p_{13} p_{14}

$$\mathbf{P} = \begin{matrix} & p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & & p_{32} & p_{33} & p_{34} \\ & p_{41} & p_{42} & p_{43} & p_{44} \end{matrix}$$

Le probabilità di transizione per $t = 1$ sono date dagli ingressi della matrice \mathbf{P} .

Per $t = 2$, invece, occorre tener conto del fatto che il processo può passare dallo stato i allo stato j attraversando un qualsivoglia stato intermedio k :

$$P_{ij}(2) = \sum_{k=1}^n P_{ik} P_{kj}$$

per tutti gli stati i, j .

Per ogni possibile stato intermedio k si calcola la probabilità congiunta che il processo passi da i a k e, successivamente, da k a j , quindi si sommano le singole probabilità congiunte, tante quanti sono i possibili stati k , ossia n

Come passare da i a j in due passaggi

p_{11} p_{12} p_{13} p_{14}

$\mathbf{P} =$ p_{21} p_{22} p_{23} p_{24}
 p_{31} p_{32} p_{33} p_{34}
 p_{41} p_{42} p_{43} p_{44}

Per calcolare la probabilità che il sistema passi dallo stato iniziale $i = ①$ allo stato finale $j = ③$ passando per un qualsivoglia stato intermedio k occorre applicare la regola della **moltiplicazione di due matrici**: riga X colonna

i					j				
k	1	2	3	4	k	1	2	3	4
①	P_{11}	P_{12}	P_{13}	P_{14}		P_{11}	P_{12}	P_{13}	P_{14}
2	P_{21}	P_{22}	P_{23}	P_{24}		P_{21}	P_{22}	P_{23}	P_{24}
3	P_{31}	P_{32}	P_{33}	P_{34}		P_{31}	P_{32}	P_{33}	P_{34}
4	P_{41}	P_{42}	P_{43}	P_{44}		P_{41}	P_{42}	P_{43}	P_{44}

Come passare da A a C in due passaggi

	A	T	C	G
	[,1]	[,2]	[,3]	[,4]
A[1,]	2	1	3	4
T[2,]	1	3	1	5
C[3,]	7	1	1	1
G[4,]	4	2	1	3

*

	A	T	C	G
	[,1]	[,2]	[,3]	[,4]
A[1,]	2	1	3	4
T[2,]	1	3	1	5
C[3,]	7	1	1	1
G[4,]	4	2	1	3

 $= Q_{AC}$

$$2*3 + 1*1 + 3*1 + 4*1 = \mathbf{14}$$

```
> mdat %*% mdat
```

	[,1]	[,2]	[,3]	[,4]
[1,]	42	16	14	28
[2,]	32	21	12	35
[3,]	26	13	24	37
[4,]	29	17	18	36

```
>
```

Come passare da A a C in due passaggi

```
piero@piero-XPS-9320:~$ octave
```

```
QSocketNotifier: Can only be used with threads started with QThread
```

```
GNU Octave, version 8.3.0
```

```
Copyright (C) 1993-2023 The Octave Project Developers.
```

```
This is free software; see the source code for copying conditions.
```

```
There is ABSOLUTELY NO WARRANTY; not even for MERCHANTABILITY or  
FITNESS FOR A PARTICULAR PURPOSE. For details, type 'warranty'.
```

```
Octave was configured for "x86_64-pc-linux-gnu".
```

```
Additional information about Octave is available at https://www.gnu.org.
```

```
Please contribute if you find this software useful.
```

```
For more information, visit https://www.gnu.org/get-involved.html
```

```
Read https://www.gnu.org/bugs.html to learn how to submit bug reports.
```

```
For information about changes from previous versions, type 'news'.
```

```
octave:1>
```

Come passare da A a C in due passaggi

```
octave:1> a = [2, 1, 3, 4; 1, 3, 1, 5; 7, 1, 1, 1; 4, 2, 1, 3]
a =
```

```
2   1   3   4
1   3   1   5
7   1   1   1
4   2   1   3
```

```
octave:2> a^2
ans =
```

```
42   16   14   28
32   21   12   35
26   13   24   37
29   17   18   36
```

```
octave:3>
```

Stati transitori e stati definitivi

Gli **stati transitori**, dopo essere stati visitati, possono essere abbandonati mentre da quelli **definitivi** la catena, una volta entrata, non può più uscire. Gli stati definitivi, per cui $P_{ii} = 1$, si presentano nella matrice di transizione \mathbf{P} come altrettanti 1 (uno) collocati nella diagonale principale, mentre in tutti gli altri ingressi ci sono solo zeri.

Per gli stati transitori $P_{ii} < 1$ cosicché almeno qualche altro ingresso della i -esima riga è > 0 .

Per definizione, la catena deve trovarsi in uno degli stati possibili ad ogni *step*; per qualsiasi valore di i e j , $0 \leq P_{ij} \leq 1$ e

$$\sum_{j=1}^n P_{ij} = 1$$

per tutti gli i , ossia la somma delle probabilità che il processo si sposti dallo stato attuale i ad uno stato j qualsiasi (che può anche coincidere con i stesso) è uguale a 1.

Il magico numero <e>

```
> 10*9*8*7*6*5*4*3*2*1  
[1] 3628800  
> factorial(10)  
[1] 3628800  
  
> Euler = function(n, z)  
+ {  
+   result = NULL  
+   numbers = seq(1:n)  
+   for (i in numbers)  
+   {  
+     result[i] = (z^i)/factorial(i)  
+   }  
+   return(1 + sum(result))  
+ }  
> Euler(100,1)  
[1] 2.718282  
> Euler(100,10)  
[1] 22026.47  
> exp(10)  
[1] 22026.47
```

$$e = 1 + 1/1! + 1/2! + 1/3! + 1/4! + \dots \cong 2.718282$$

$$e^z = 1 + z/1! + z^2/2! + z^3/3! + z^4/4! + \dots \cong 2.718282$$

Per $z < 1$:

$$e^z \cong 1 + z$$

Es.

$$\text{se } z = 0.1$$

$$e^{0.1} \cong 1 + 0.1 \cong 1.1$$

```
> exp(0.1)
```

```
[1] 1.105171
```

```
> Euler(10,0.1)
```

```
[1] 1.105171
```

Da
 $e^z = 1 + z/1! + z^2/2! + z^3/3! + z^4/4! + \dots \cong 1 + z$
 per $z < 1$ (al limite per z che tende a a)

$$\mathbf{P} = \begin{matrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{matrix}$$

Si ricava che:

$$e^A = 1 + A/1! + A^2/2! + A^3/3! + A^4/4! + \dots \cong (I + A) = \mathbf{P}$$

$$I + A = \mathbf{P}$$

$$1 \quad 0 \quad 0 \quad 0 \quad a_{11} \quad a_{12} \quad a_{13} \quad a_{14} \quad p_{11} \quad p_{12} \quad p_{13} \quad p_{14}$$

$$0 \quad 1 \quad 0 \quad 0 \quad a_{21} \quad a_{22} \quad a_{23} \quad a_{24} \quad p_{21} \quad p_{22} \quad p_{23} \quad p_{24}$$

$$+$$

$$0 \quad 0 \quad 1 \quad 0 \quad a_{31} \quad a_{32} \quad a_{33} \quad a_{34} \quad p_{31} \quad p_{32} \quad p_{33} \quad p_{34}$$

$$0 \quad 0 \quad 0 \quad 1 \quad a_{41} \quad a_{42} \quad a_{43} \quad a_{44} \quad p_{41} \quad p_{42} \quad p_{43} \quad p_{44}$$