



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA



**EGI Conference 2022**

19-23 September 2022,

Prague, Czech Republic

# Big data in livestock genomics can feed new concepts in One Health

**Samuele Bovo, Giuseppina Schiavo, Matteo Bolner,  
Mohamad Ballan, Anisa Ribani, Valeria Taurisano &  
Luca Fontanesi**

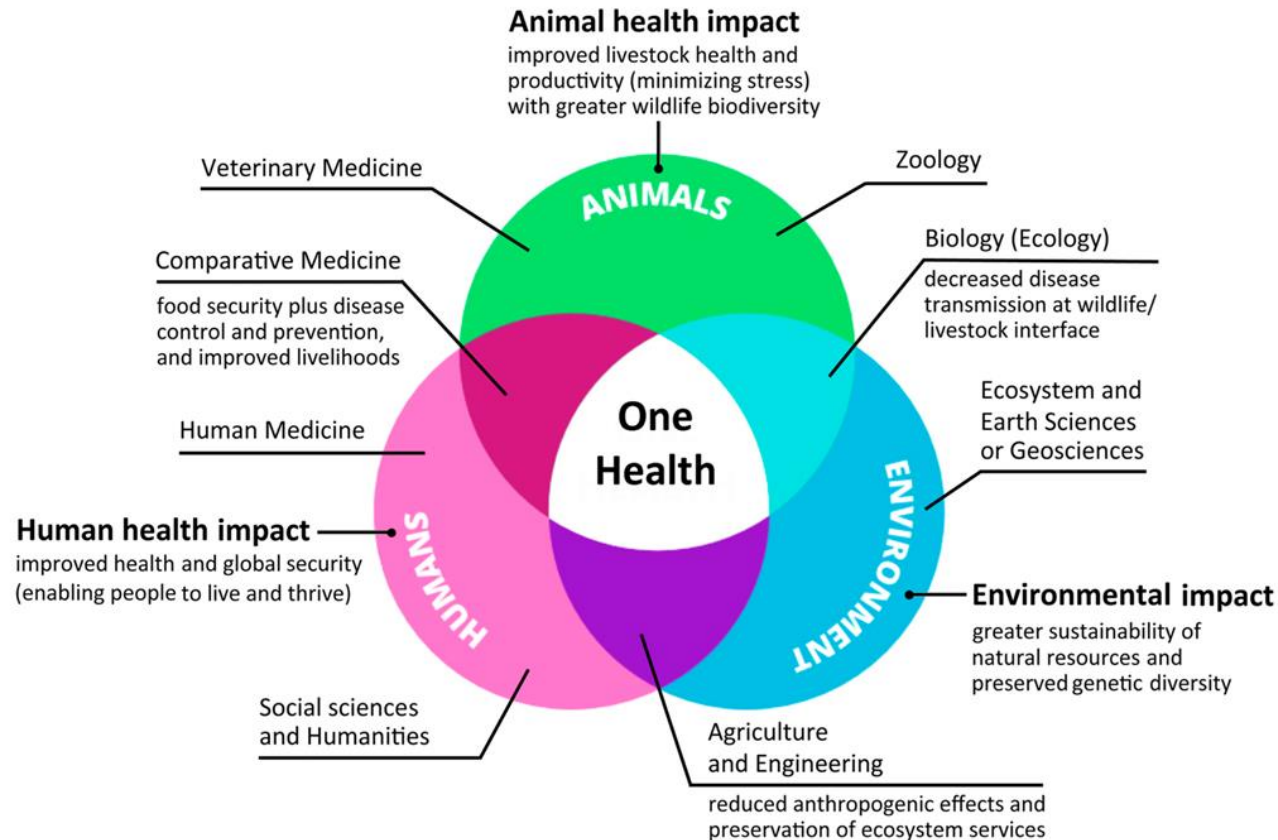
*Animal and Food Genomics Group,  
Department of Agricultural and Food Sciences (DISTAL)  
University of Bologna, Bologna, Italy (IT)*

[samuele.bovo@unibo.it](mailto:samuele.bovo@unibo.it)

[luca.fontanesi@unibo.it](mailto:luca.fontanesi@unibo.it)

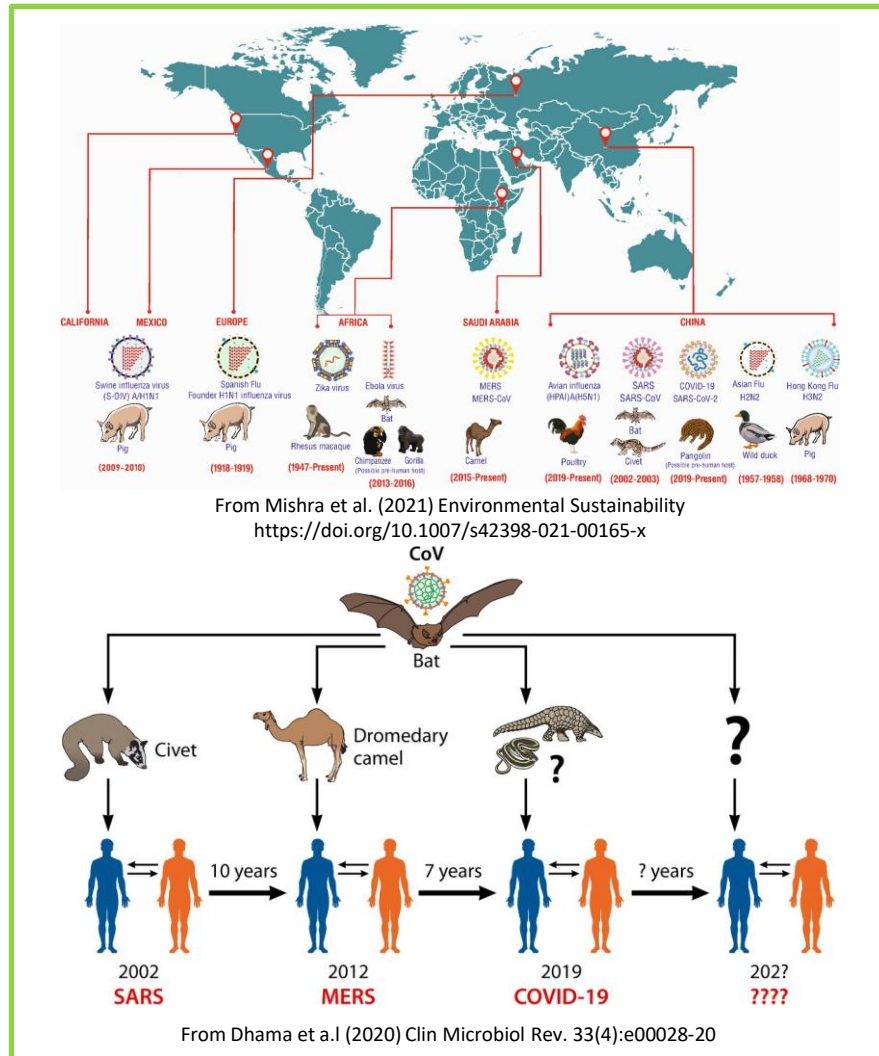
# The One Health approach: some definitions

One Health is a **multi-sectoral, interdisciplinary**, and collaborative approach to attain **optimal health** for animals, the environment, and humans (CDCP, 2020).



- Prevention
- Detection
- Response
- Recovery

# Pandemics and the One Health approach



The **COVID-19 pandemic** has strongly emphasized the importance of the **One Health approach**.

## Coronaviruses

- Causes mild to severe diseases;
- Pass cross species barriers

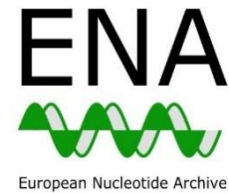
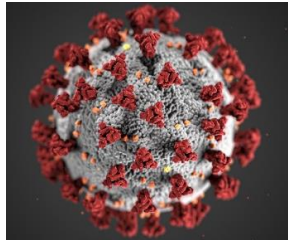
## What to do:

1. address the dynamics of disease emergence at the human, animal, agriculture and environment interface
2. Design interventions



# HOW?

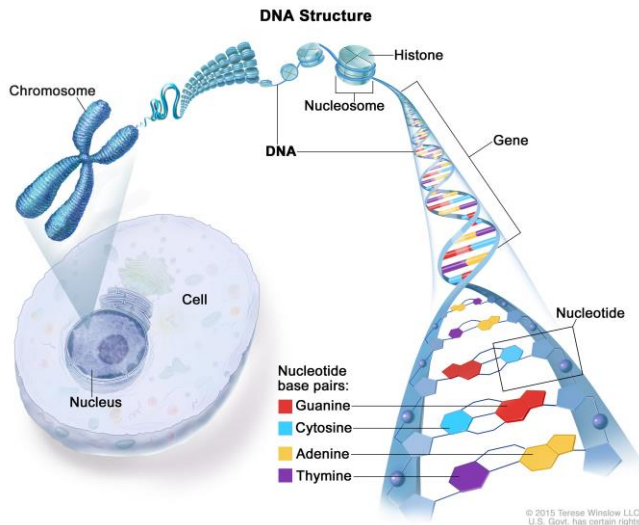
## A genomic approach!



# Genomics

**Genomics** = Study of genomes.

**Genome** = Entire genetic content/information of an organism.



```
ATCTAGCATCGATGCATGCTAG  
CACGAGCTAGCTGATCAGCTGA  
TGCTACTGATGCAGTCTAGCTG  
AATCTAGCATCGATGCATGCTA  
GCACGAGCTAGCTGATCAGCTG  
ATGCTACTGATGCAGTCTAGCT
```

High-throughput technologies

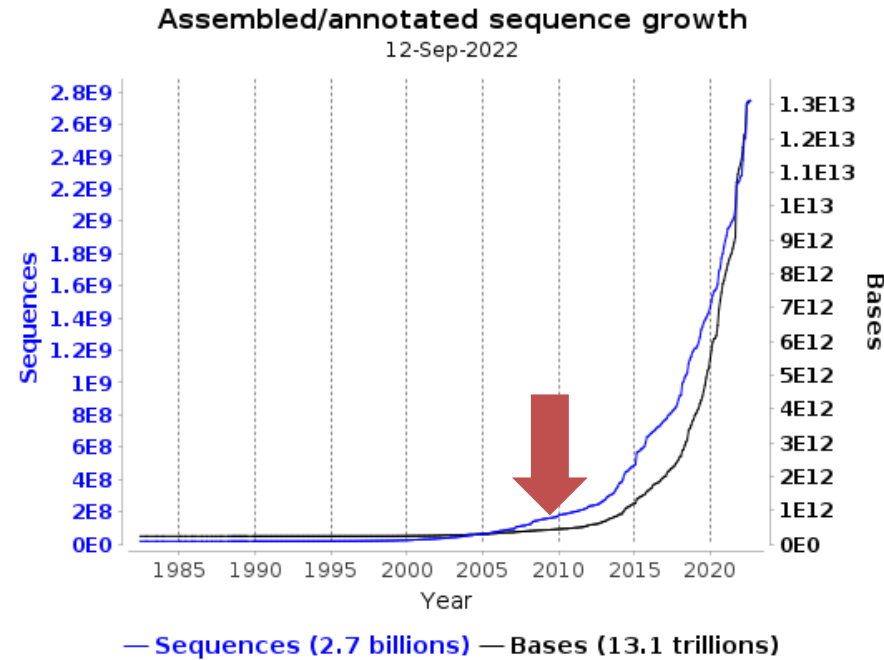
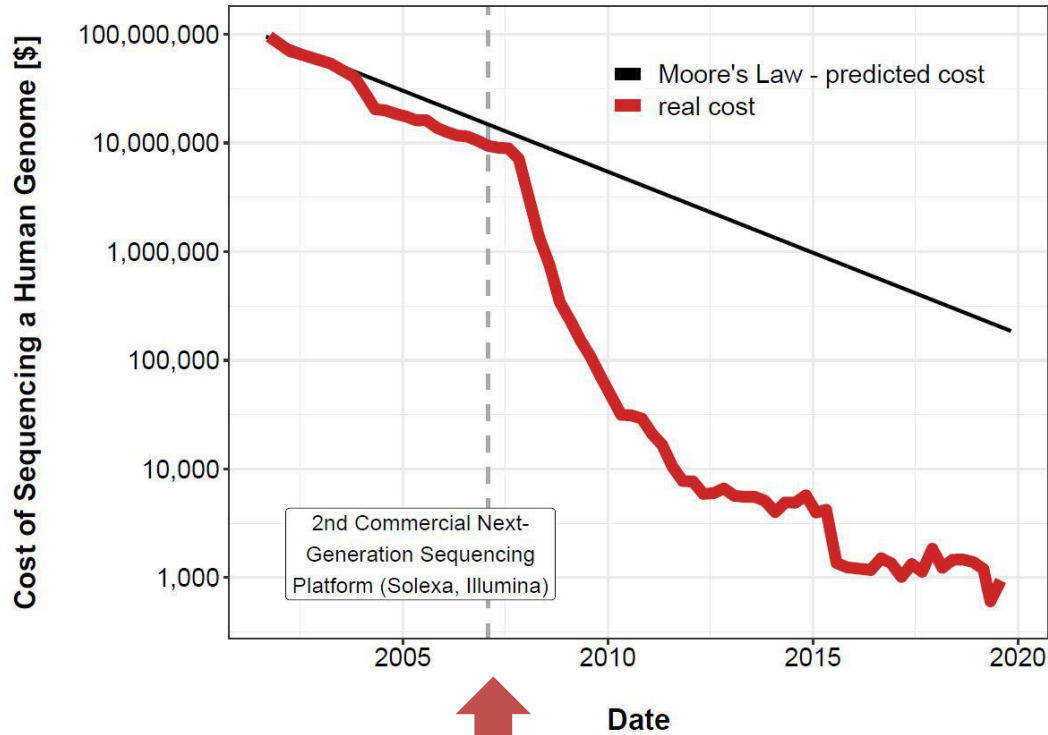
Animal genome  $\sim 3 \times 10^9$  Nucleotides  $\rightarrow$  3Gbp

Reliable sequencing 20X  $\rightarrow$  60Gbp

(Min 5X  $\rightarrow$  15Gbp)



# Genomics and Big Data




## Strategy

EpiGenomics (1,247)  
 Exome (725)  
 Genome (39,234)  
 RNASeq (295)  
 other (23,392)

## Data in Cloud

GS (64,790)  
 S3 (64,711)

## File Type

bam (3,266)   
 cram (75)  
 fastq (26,755)  
 illumina native (2)  
 sff (416)  
 srf (19)

## SRA query:

*Sus scrofa*





# Genomics and One Health



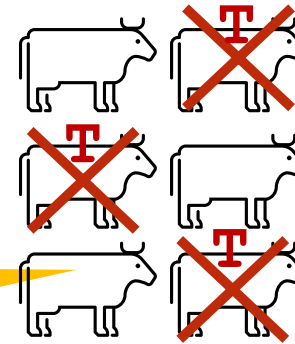
1

## DNA of the animal

ATC**T**AGCATCGATGCATGCTAG  
CACGAGCTAGCTGATCAGCTGA  
TGCTACTGATCC



RESISTANCE

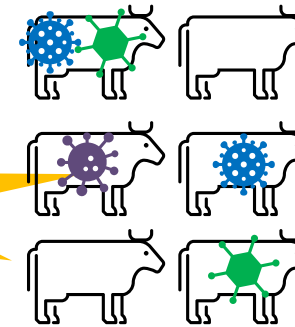


2

## DNA of viruses

TC**G**ATGCATGCTAGCACGAGCT  
AGCTGATCAGCTGATGCTACTG  
ATGCAGTCT**A**GCATCCGTCGA  
CGTAGCTAG

DETECTION



Genetic variants conferring resistance.  
→ help to design genetic improvement strategies

Prevalence of infections of some viruses.  
→ Establishment of a risk evaluation system.



# Genomics, Big Data and One Health

## Investigated animals

- Cattle, pigs, chickens, rabbits;
- Related wild populations.

## Metadata retrieval and sample selection

- ENA database: metadata via REST API;
- QC and filtering → Reliable datasets.



|                    | Cattle | Pig | Chicken | Rabbit |
|--------------------|--------|-----|---------|--------|
| No. of projects    | 9      | 27  | 43      | 5      |
| No. of datasets    | 503    | 464 | 437     | 67     |
| No. of populations | 35     | 52  | 15      | 31     |
| Stored data (TB)   | 13.0   | 9.5 | 5.8     | 1.8    |

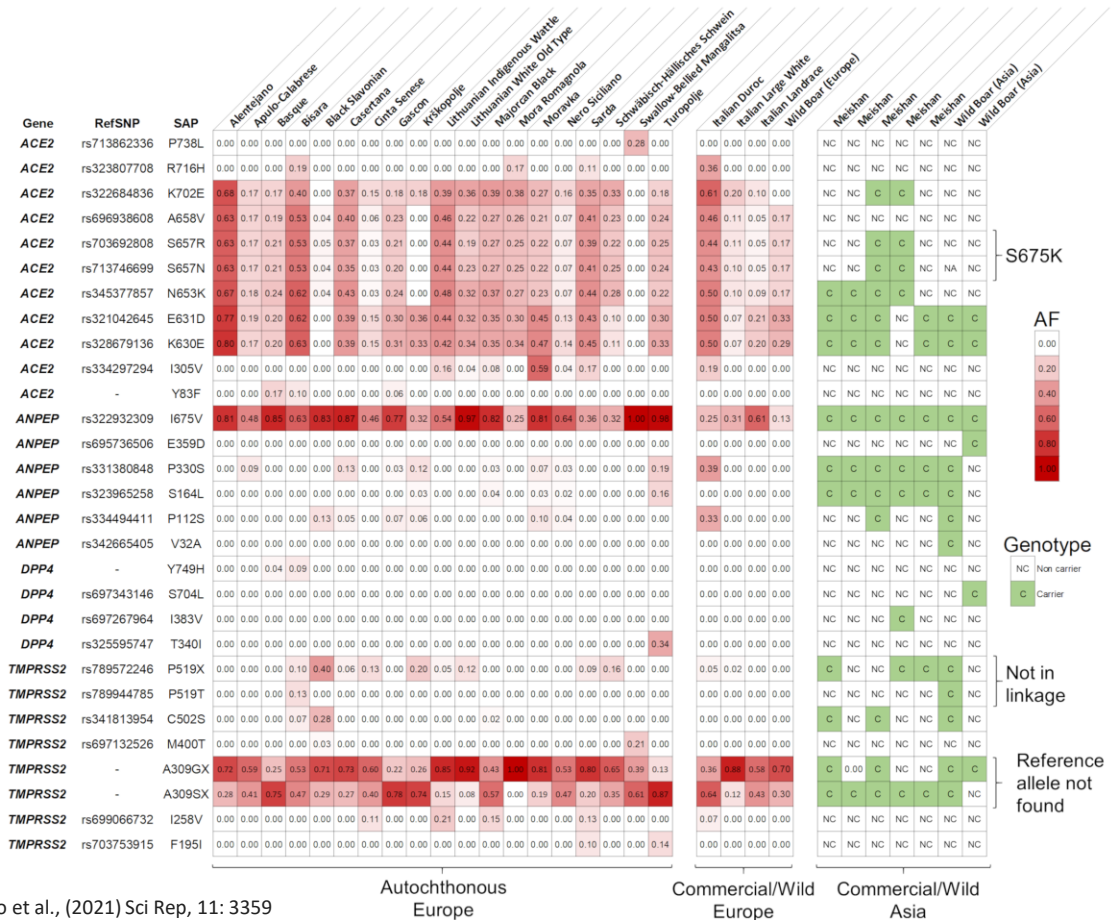
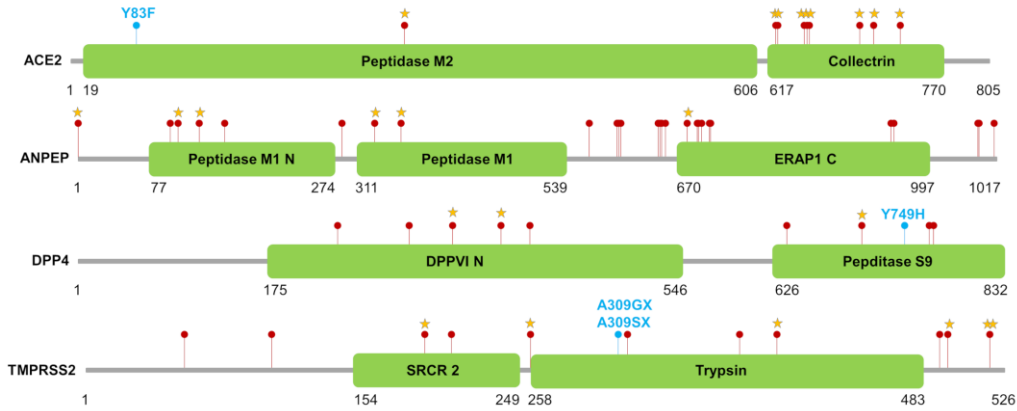
**1,471 datasets and >30 TB of sequencing data  
(>60TB considering also raw/source data)**





# 1

# Genetic variability affecting coronavirus related genes in pigs



**Few Novel Variants Functional Variants**

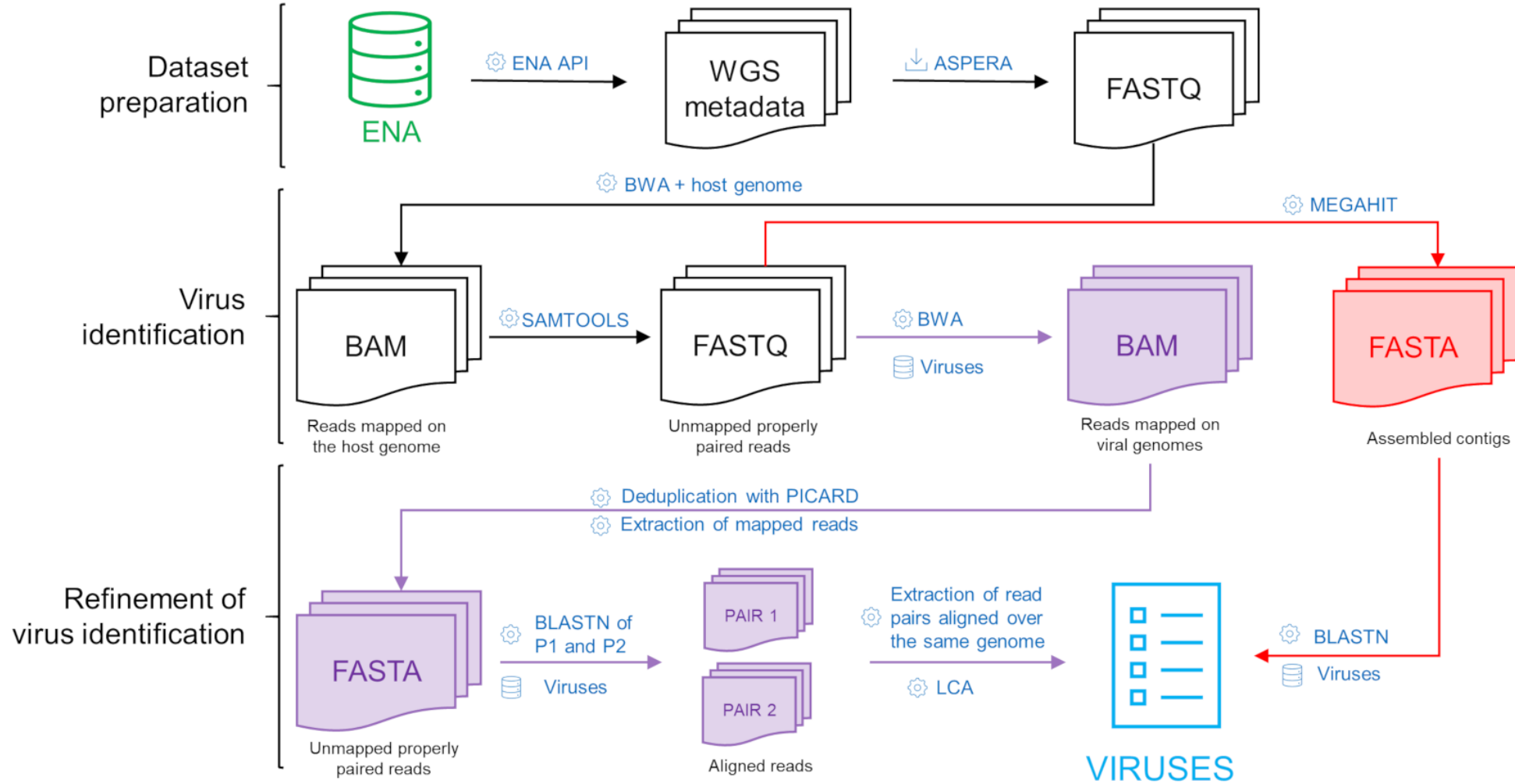
1. ACE2 p.Y83F → region for SARS-CoV-2 interaction;
2. DPP4 p.I383V → close to MERS-CoV receptor-binding domain;

**Establishment of selection programs designed to increase genetic resistance to coronaviruses**



# 2

## Evaluation of the animal virome



## 2 Evaluation of the animal VIROME: results (I)

- > 480 billion reads (1471 WGS datasets);
- < 2% of unmapped read on each sample;
- 510 matches with viral genomes:
  - **many viruses could be considered contaminants,**
  - **few (no. 38) classified as specific infecting agents**
- **Each sample contained 1-4 different viruses;**
- Some virus identified in more than one host dataset.

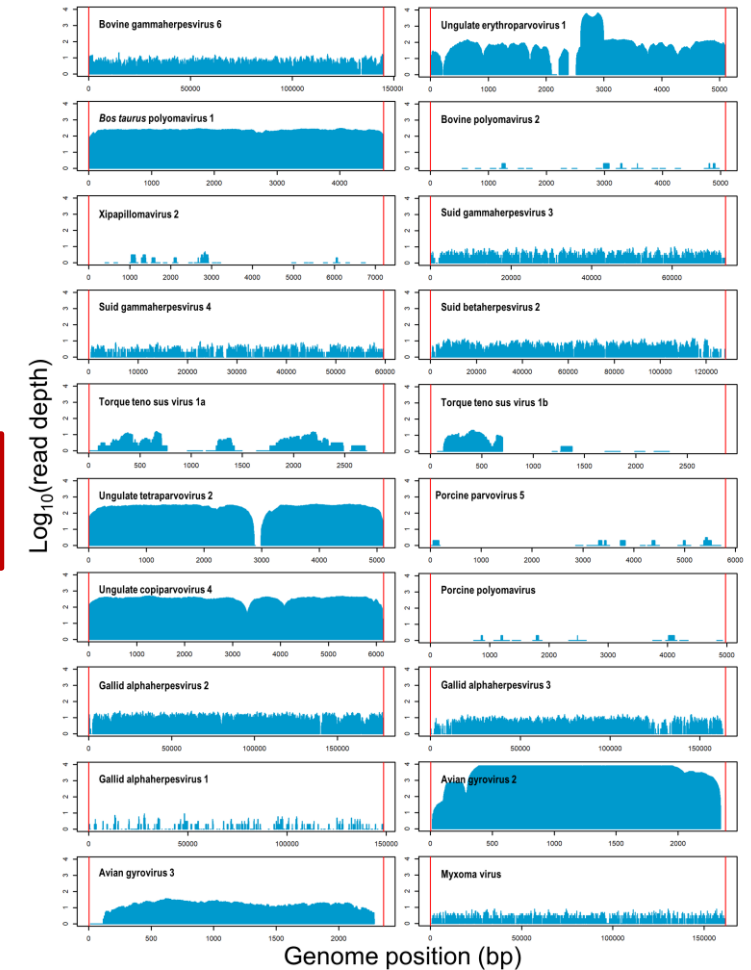
| Information on the datasets/detected viruses | Cattle    | Pig       | Chicken   | Rabbit   |
|--|-----------|-----------|-----------|----------|
| No. of projects                              | 9         | 27        | 43        | 5        |
| No. of datasets                              | 503       | 464       | 437       | 67       |
| No. of populations                           | 35        | 52        | 15        | 31       |
| Stored data (TB)                             | 13.0      | 9.5       | 5.8       | 1.8      |
| Total n. of viruses                          | 223       | 160       | 105       | 22       |
| <b>Number of host specific viruses</b>       | <b>14</b> | <b>11</b> | <b>12</b> | <b>1</b> |



## 2 Evaluation of the animal VIROME: results (II)

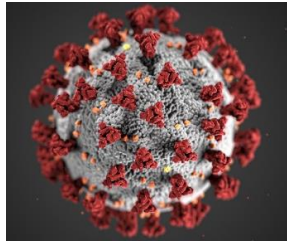
| Virus                               | No. of datasets | No. of projects | % of positive datasets |
|-------------------------------------|-----------------|-----------------|------------------------|
| Avian endogenous retrovirus, EAV-HP | 202             | 34              | 45.9                   |
| Gallid alphaherpesvirus 3           | 40              | 18              | 9.1                    |
| Gallid alphaherpesvirus 2           | 24              | 13              | 5.5                    |
| Avian leukosis virus                | 23              | 9               | 5.4                    |
| Avian gyrovirus 2                   | 8               | 4               | 1.8                    |
| Gyrovirus GyV3                      | 8               | 5               | 1.8                    |
| Gallid alphaherpesvirus 1           | 3               | 2               | 0.7                    |
| [...]                               | [...]           | [...]           | [...]                  |

Useful epidemiological hints to understand the distribution, diffusion and prevalence of viruses infecting livestock.

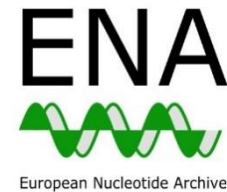


# Conclusions

**Genomics Big Data have revolutionized pathogen detection contributing to improve the safety of production chains.**



**Opportunistic mining of whole genome sequencing of animal genomes provide new information for epidemiological approaches of potential zoonosis.**





## Animal and Food Genomics Group



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA



**L. Fontanesi**  
Full Professor



**G. Schiavo**  
Assistant Professor



**S. Bovo**  
Assistant Professor



**M. Ballan**  
PhD student



**M. Bolner**  
PhD student



**A. Ribani**  
Assistant Professor



**V. Taurisano**  
PhD student

This work has received funding from:

- the **University of Bologna RFO** 2016-2019 programmes;
- the Italian **MIUR 2017 PigPhenomics** project;
- the **European Union's Horizon 2020** research and innovation programme under grant agreement No. 634476 for the project with acronym **TREASURE**;
- the **European Open Science Cloud** (EOSC) Secretariat, project "Application of animal genomics and data mining to predict and monitor novel coronavirus potential infections (**VirAnimalOne**)",
- the **EGI** call for COVID-19 research projects (**AnGen1H** project);
- The **Por Fesr Emilia-Romagna** 2014-2020 (actions 1.1.4 and 1.2.2—Bando per sostenere progetti di ricerca ed innovazione per lo sviluppo di soluzioni finalizzate al contrasto dell'epidemia da COVID-19—Project **LIVESTOCK-STOP-COVI**).

CIRI AGROALIMENTARE

