



75th EAAP Annual Meeting 2024
Florence, Italy - 1/5 September 2024



EAAP
European Federation
of Animal Science



A new version of buffalo SNP array to support Mediterranean breed genomic improvement

Milanesi M.¹, Pietrucci D.¹, Picarone L.¹, Vignali G.¹, Cosenza G.², Gomez M.³, Chillemi, G.¹

1 University of Tuscia, Department for Innovation in Biological Agro-Food and Forest Systems - DIBAF, Via San Camillo de Lellis, snc, 01100 Viterbo, Italy,

2 University of Naples "Federico II", Department of Agricultural Sciences, Piazza Carlo di Borbone, 1, 80055 Portici, Italy

3 Associazione Nazionale Allevatori Specie Bufalina - ANASB, Via Petrarca, 42-44, 81100 Caserta, Italy

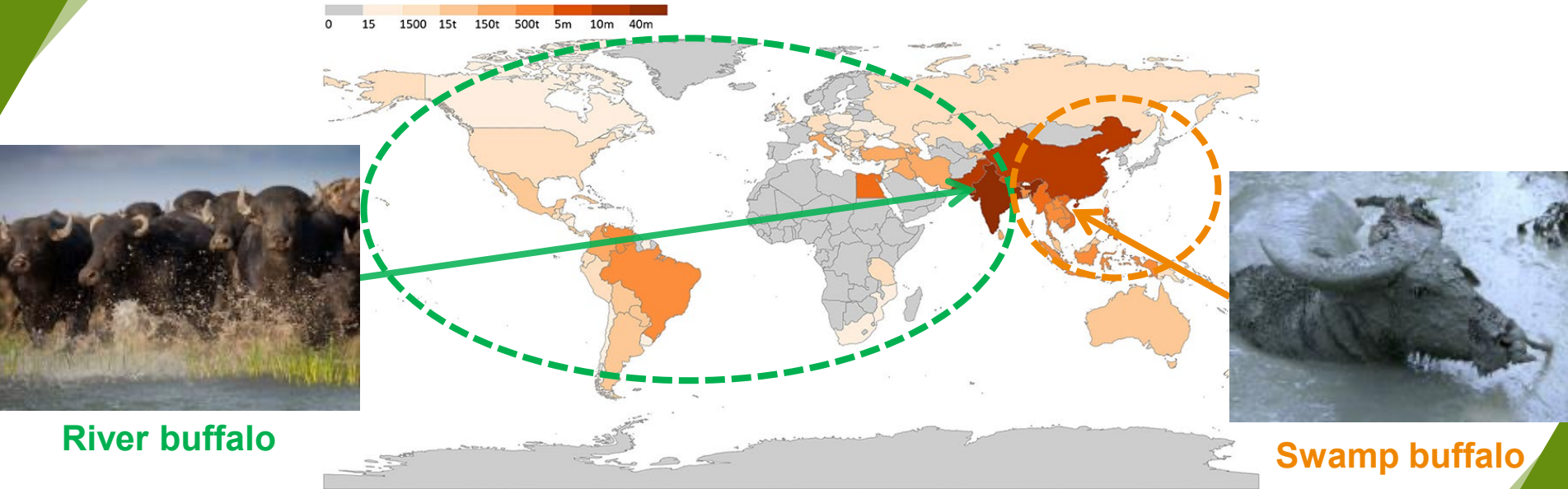


UNIVERSITÀ
DEGLI STUDI DELLA
TUSCIA

DIPARTIMENTO
PER LA INNOVAZIONE
NEI SISTEMI BIOLOGICI,
AGROALIMENTARI E FORESTALI

Theatre Session 33
Monday 2^o September, 17:30
Room Reggiana

Introduction

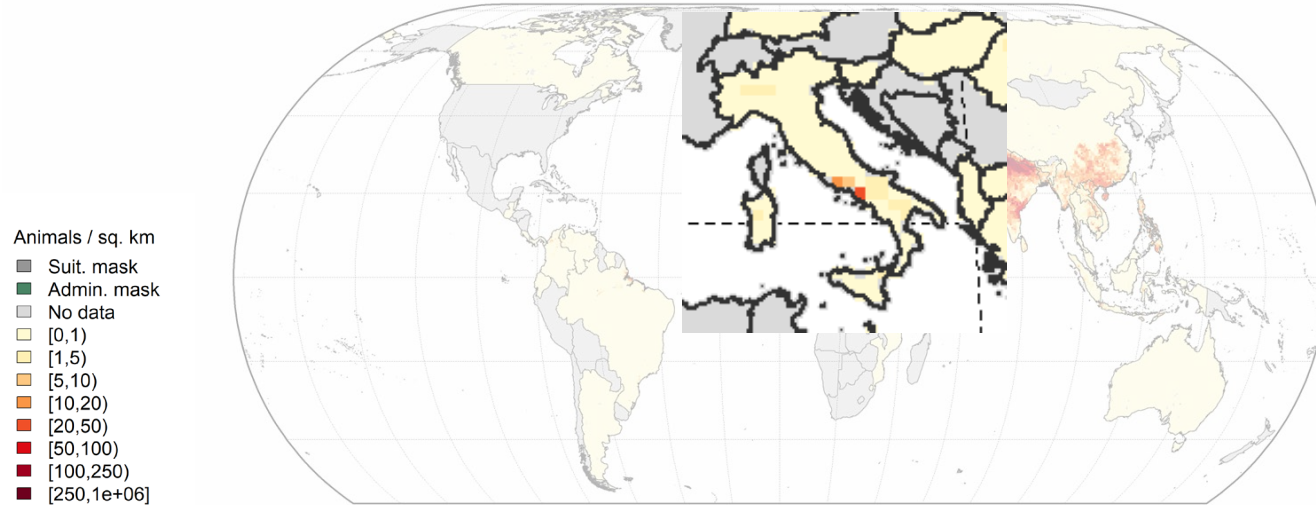


River buffalo

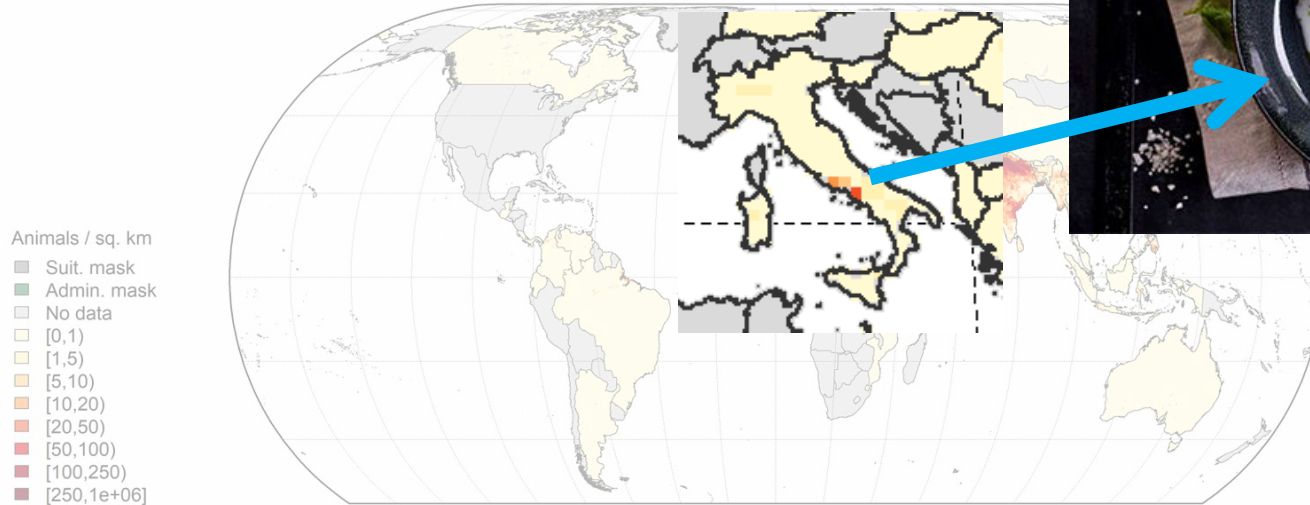
Swamp buffalo

- *Bubalus bubalis*, one species with two distinct ecotypes:
 - **Swamp type** (*B. bubalis carabanensis*), prefers marshlands and swamps; used for draught power
 - **River type** (*B. bubalis bubalis*), prefers clear running water; selected as a dairy animal

Introduction

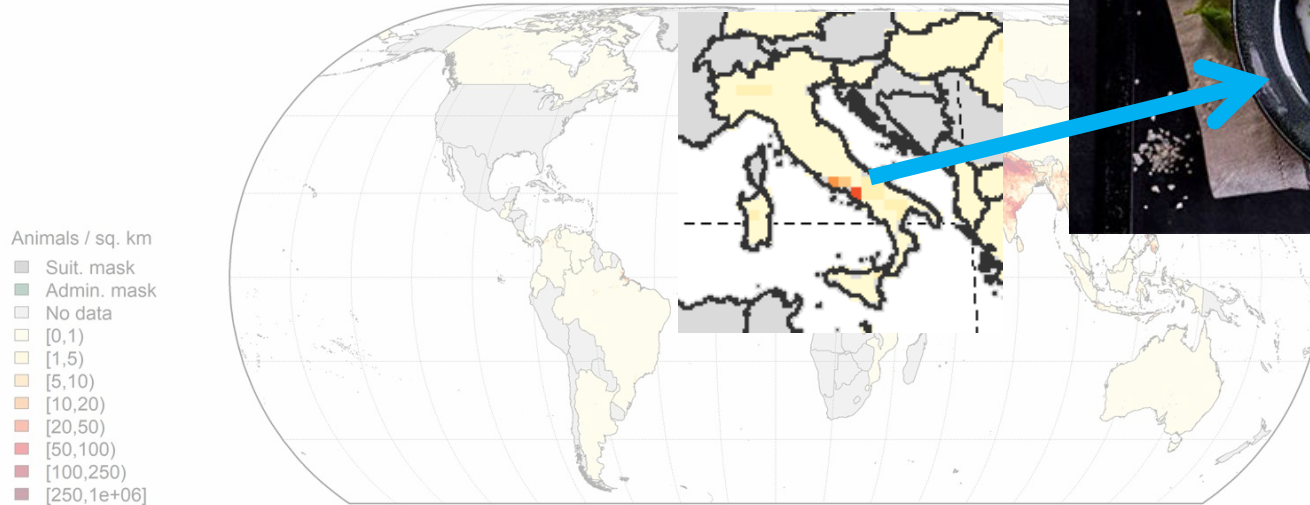


Introduction



- In Italy, the **Mediterranean breed** (river ecotype) is reared to produce “mozzarella di Bufala campana” PDO

Introduction



- In the last century, the number of animals increased rapidly
 - From almost 10,000 animals in the beginning of '900
 - To more than **430,000 head** at the end of 2022
- ANASB is involved in the improvement, also with **Genomic Selection**



*Associazione Nazionale
Allevatori Specie Bufalina*

Introduction

- A **medium density** array developed for **biodiversity** purpose
 - 90k SNPs and ~120k probes
 - No buffalo RefSeq at that time

(lamartino et al., 2017)

International Water Buffalo Genome Consortium

- A multi-species array with ~10k SNPs selected for buffalo, developed in the IMAGE project, also for **biodiversity** purpose

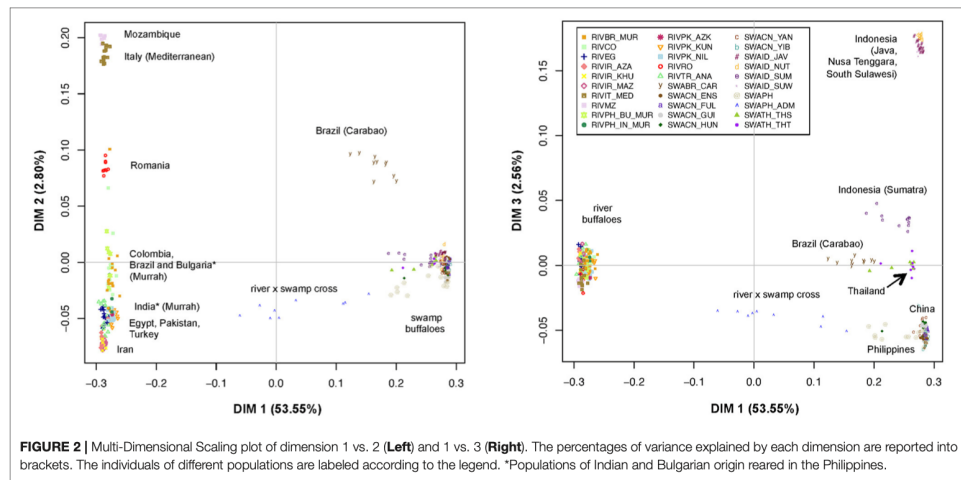


Introduction

- A **medium density** array developed for **biodiversity** purpose

- 90k SNPs and ~120k probes
- No buffalo RefSeq at that time

(Iamartino et al., 2017)



- However

- “The dataset version based on markers polymorphic **overall** contained **52,637** SNPs, 335 individuals and 31 populations.” (Colli et al., 2018)
- “... 529 individuals ... from 4 **Italian Mediterranean** buffalo farms. ... **66,534** high-quality (PolyHighRes) SNPs with MAF >1%. (Iamartino et al., 2017)

The GENOBU project



GENOBU

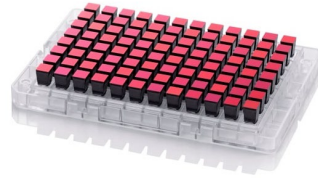


Objective

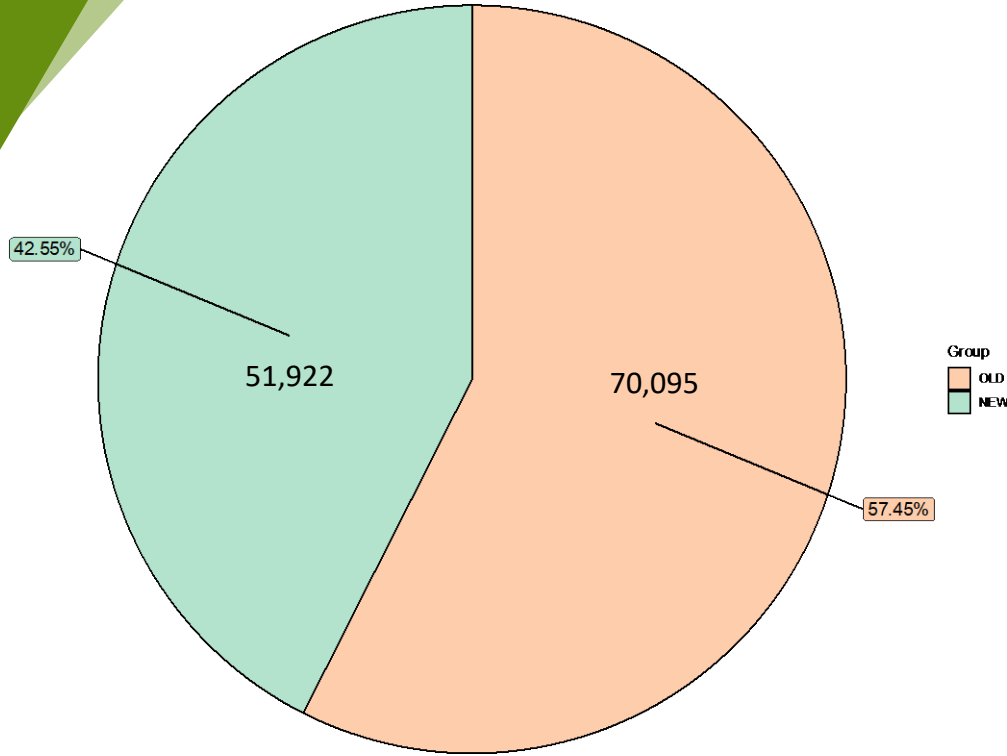
Develop a new version of buffalo SNP array

- **Specific for Mediterranean breed**
 - Mapping SNPs on most recent Mediterranean RefSeq (*UOA_WB_1*)
 - Optimized for Genomic Selection: Evenly spaced markers
- Fill the **~120k spots** in the array with SNPs
- Included markers in **Y chromosome** and **functional markers**
- Maximum **compatibility** with previous arrays

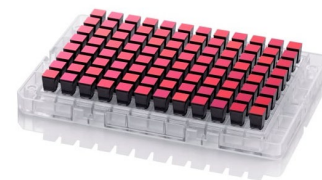
The new array



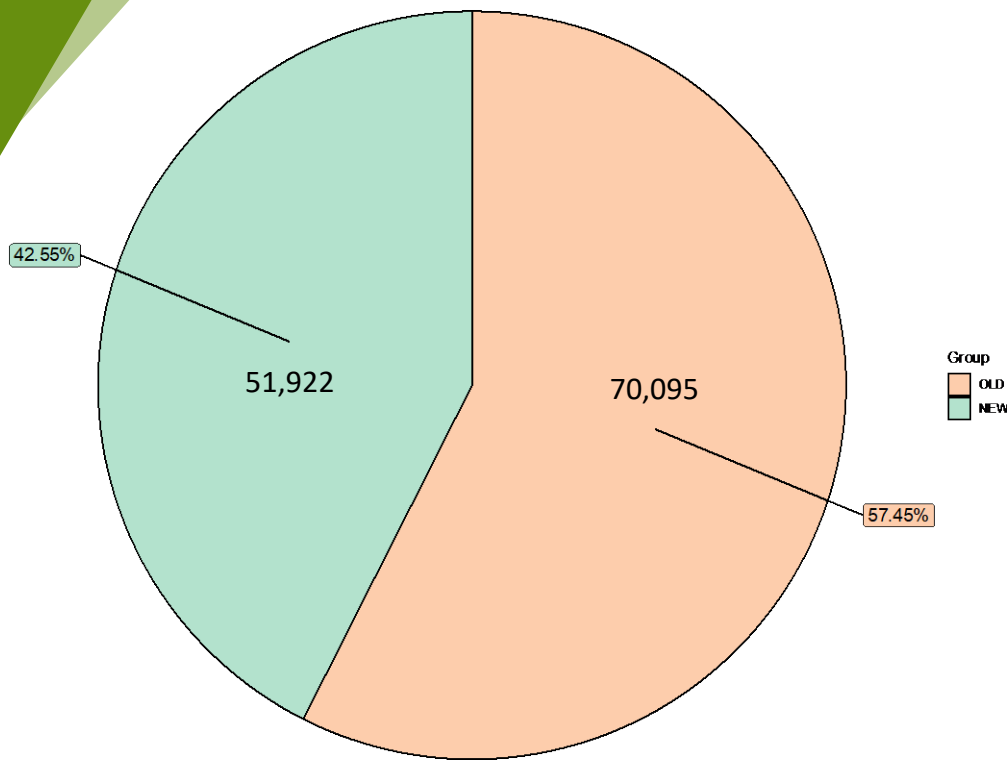
Axiom_BubMED2



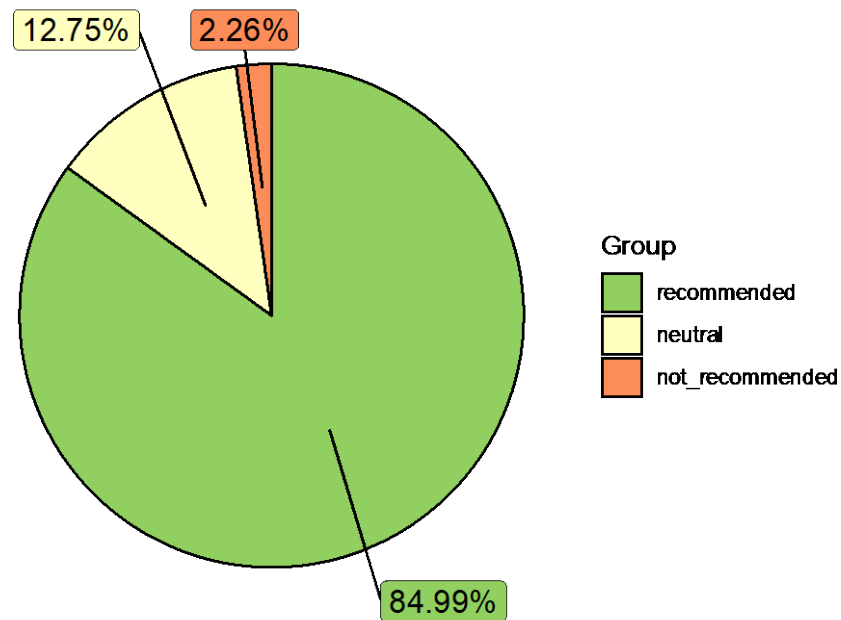
Total of 122,017 SNPs



Axiom_BubMED2

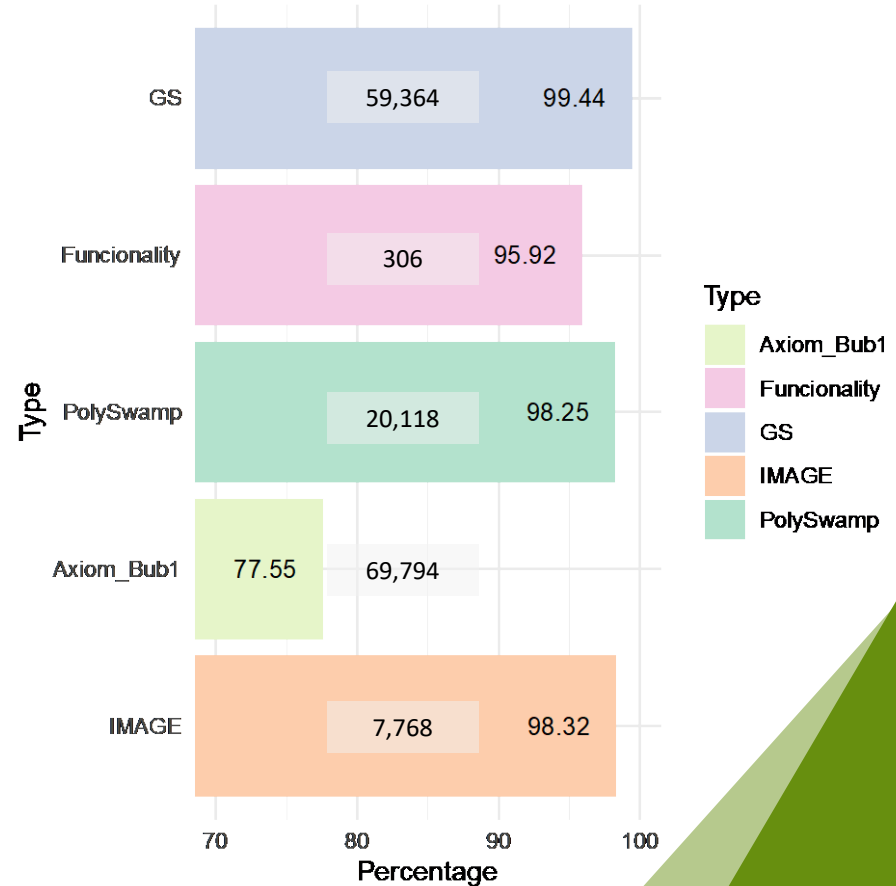


Total of 122,017 SNPs



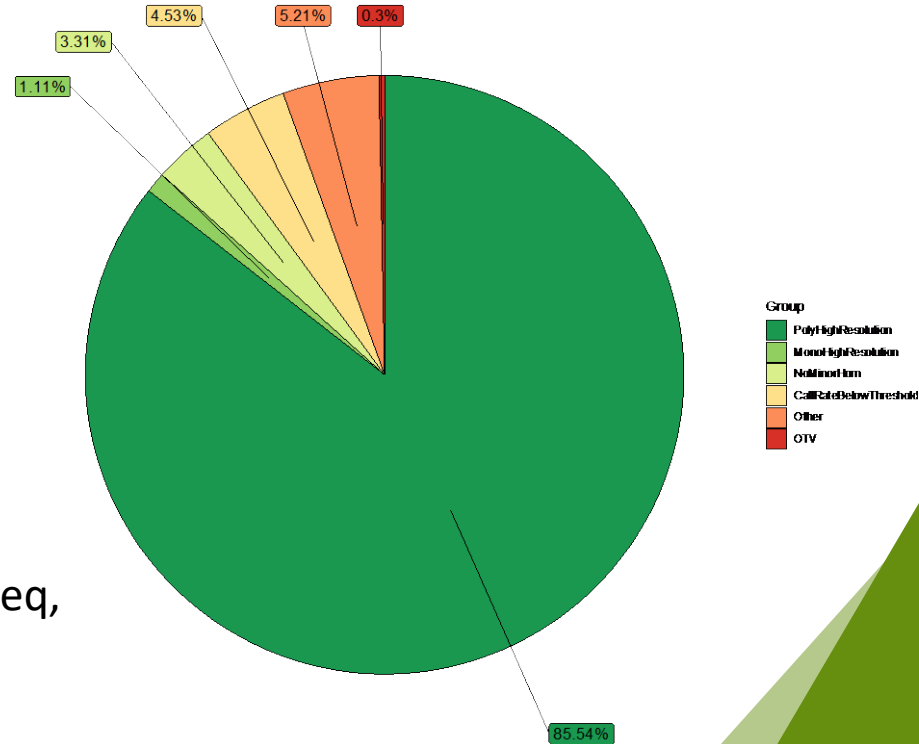
“Already known” SNPs

- Importance:
 - Already used in **Genomic Selection**
 - **Association** with important characteristics
 - **Biodiversity**
 - **Compatibility** with previous array



“Already known” SNPs

- Importance:
 - Already used in Genomic Selection
 - Association with important characteristics
 - Biodiversity
 - Compatibility with previous array
- SNP characteristics:
 - Good **mapping** in Mediterranean RefSeq, and in other buffalo RefSeqs
 - **Evaluate** Conversion Type, MAF, CR in ~500 Mediterranean animals (Iamartino et al., 2017)



“New” SNPs

- **WGS for diversity evaluation, from short reads sequencing**
 - 16 male Mediterranean [MED]
 - 5 male River (Murrah, Jaffrabadi and Nili-Ravi) [RIV]
 - 7 male Swamp (China's breeds) [SWA]
- **WGS for Mediterranean evaluation, from short reads sequencing**
 - 24 female MED
 - Extreme for IBMI index
 - In collaboration with CNR-IBBA
- **WGS from long reads**
 - 8 male MED
 - In collaboration with CEINGE

“New” SNPs

- **WGS for diversity evaluation, from short reads sequencing**

- 16 male Mediterranean [MED]
- 5 male River (Murrah, Jaffrabadi and Nili-Ravi) [RIV]
- 7 male Swamp (China's breeds) [SWA]

- **WGS for Mediterranean evaluation, from short reads sequencing**

- 24 female MED
 - Extreme for IBMI index
- In collaboration with CNR-IBBA

- **WGS from long reads**

- 8 male MED
- In collaboration with CEINGE

- *In-house* scripts, based on common best practice
- Aligned to *UOA_WB_1* reference genome
 - Added Y chromosome (from Yak)
- Variant calling, filtering and annotation:
 - Set missing genotypes with Genotype Quality (GQ)<30 and depth (DP)<5
 - Retained only bi-allelic SNPs
 - Annotation with *snpEff*

Pipeline for short reads: Alignment: BWA. Variant calling & filtering: BCFtools

Pipeline for long reads: Alignment: Minimap2. Variant calling & filtering: Clair3; GATK & Glnexus; BCFtools

“New” SNPs

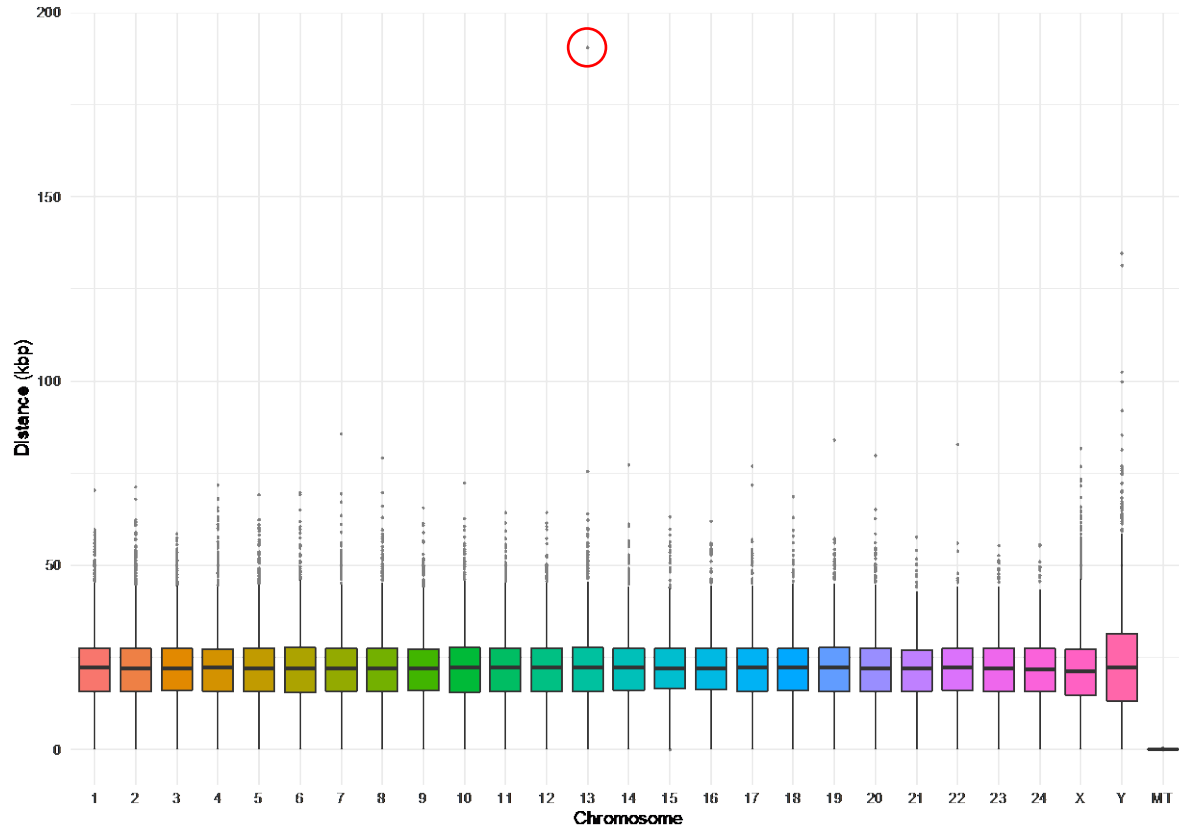
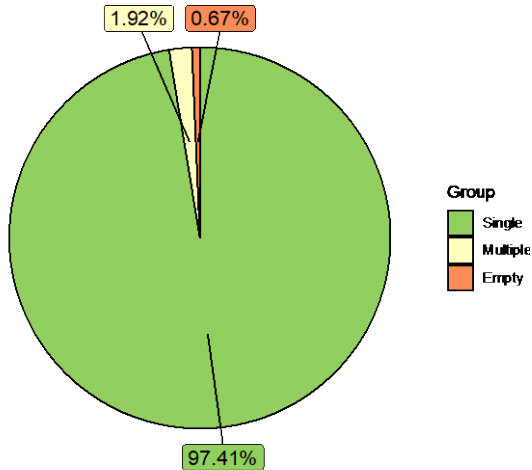
- Selection criteria:
 - Good **allele frequency and low missingness** in MED
 - If possible, the same in RIV and SWA
 - **Avoid ambiguous** SNPs
 - No variants in the flanking regions
- Results
 - Fill the gaps not covered by “already known” markers
 - +179 new functional markers
 - Milk production (casein cluster, *SCD*, *PRL*), *IGF2*, *GHR*, *GH*, etc...
 - + 1070 SNPs in Y chromosome



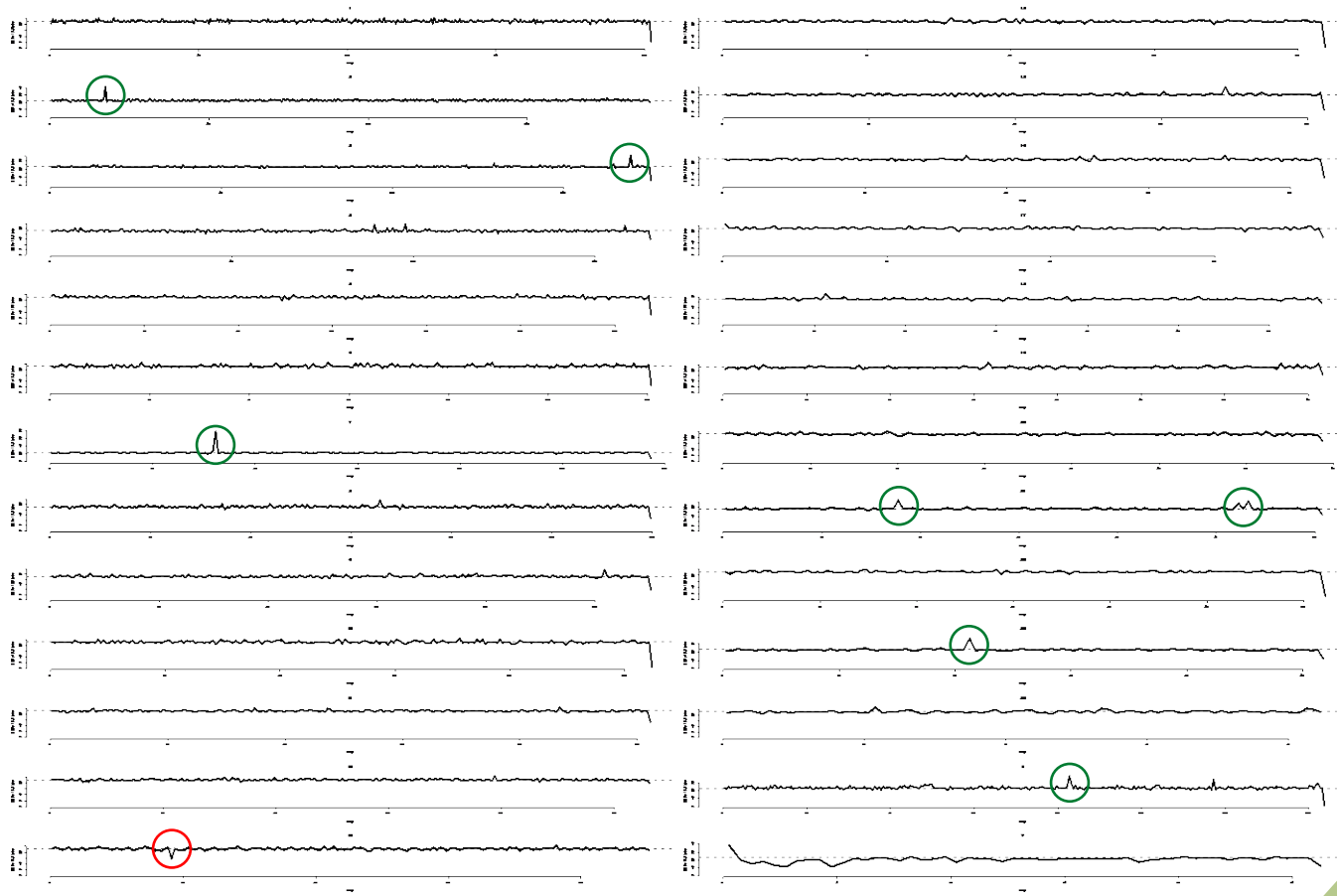
Results

SNPs distance

- Average: 21.8kbp
- Std. dev.: 8.8kbp
- Min.: 1bp
- Max.: 190.6kbp



SNP density



Validation

Dataset validation

- Testing the new SNP array
 - Ongoing; results before the end of the year
 - Testing also RIV and SWA animals

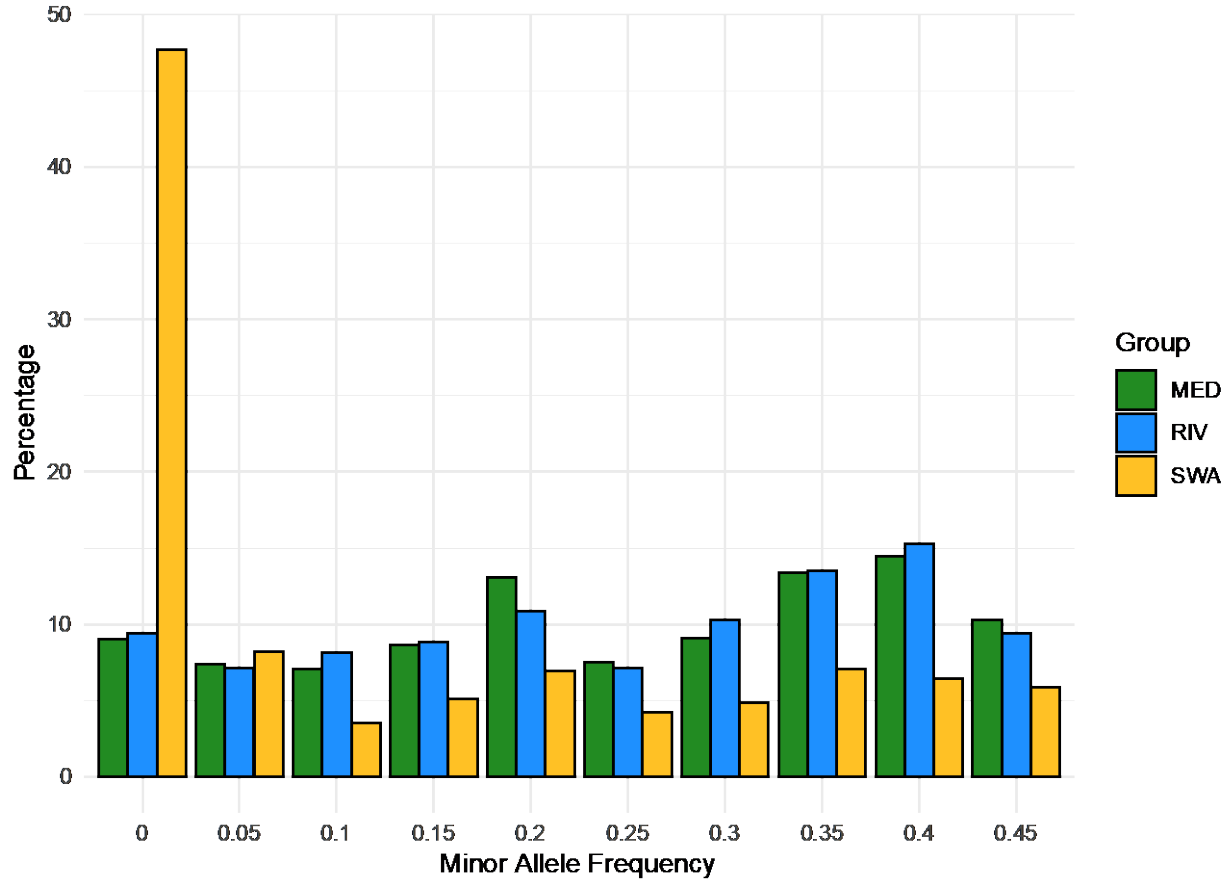


Dataset validation

- Testing the new SNP array
 - Ongoing; results before the end of the year
 - Testing also RIV and SWA animals
- *In silico* validation (using WGS data)
 - Dataset
 - 10 male MED
 - 10 male RIV (Jaffrabadi, Murrah, Nili-Ravi & from India)
 - 10 male SWA (from China)
 - Analyzed using the same pipeline previously reported

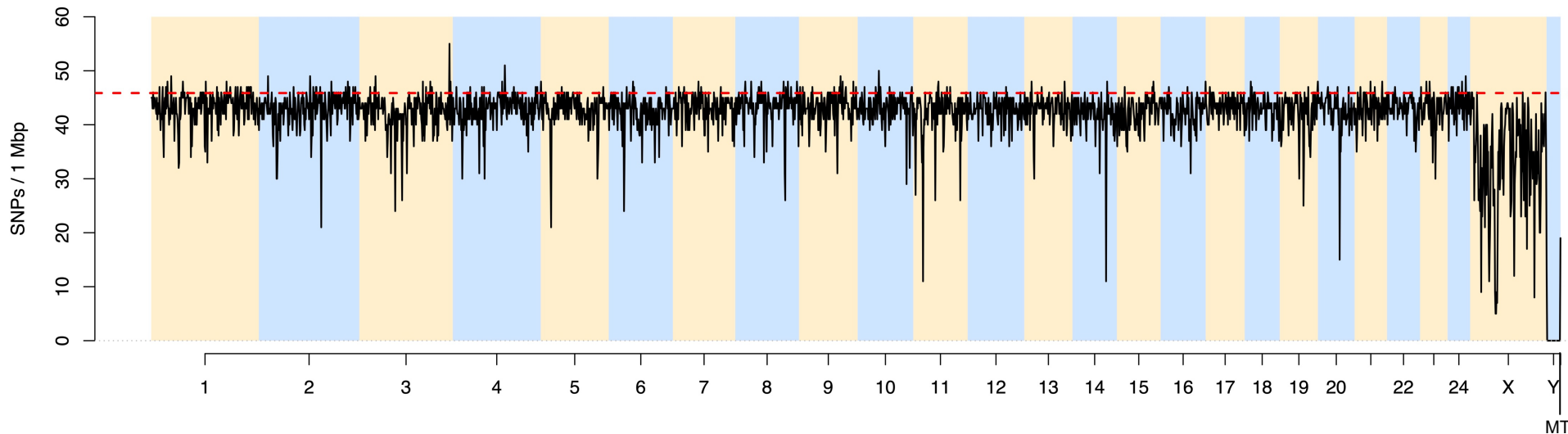


Allele frequency



“Working dataset” simulation

- Comparison with *Iamartino et al., 2017*
 - Keep only MED animals, SNPs with position in *UOA_WB_1*, MAF>0.01 and high CR
 - Total of **110,750 SNPs** kept



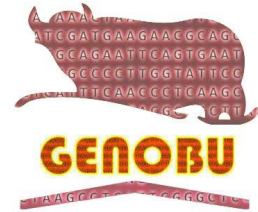
Conclusions

Take home messages

- **+36% SNPs** in the raw array, compared with the first version
 - High compatibility with previous arrays
 - Focus on functionality
- **+67% SNPs** in a “working dataset”, compared with previous analyses
 - Good genome coverage
- Release in the **end 2024/beginning 2025**

Acknowledgment

The research was founded by **PON01_00486** project, «SEQUENZIAMENTO DEL GENOMA BUFALINO PER IL MIGLIORAMENTO QUALI-QUANTITATIVO DELLE PRODUZIONI AGRO-ALIMENTARI», GENOBU



agritech

National Center for
Technology in Agriculture

Plan. Plant. Planet.

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This presentation reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

**Thank you
for your attention**

marco.milanesi@unitus.it