



Konferencija Srce DEI

# Kako **SUPEK** i **Padobran** pokreću arheogenetiku i genomiku goveda u Hrvatskoj

**Prezentira: izv. prof. dr. sc. Maja Ferenčaković**

Sveučilište u Zagrebu Agronomski fakultet

*Laboratorij za arheogenetiku*

Srce DEI 2026 · 28. travnja 2026. · Rektorat Sveučilišta u Zagrebu

**Srce DEI 2026**





# 15 tisuća godina genomike goveda.

*Od pleistocenskog tura do modernog holsteina — jedna nit, jedan HPC sustav.*



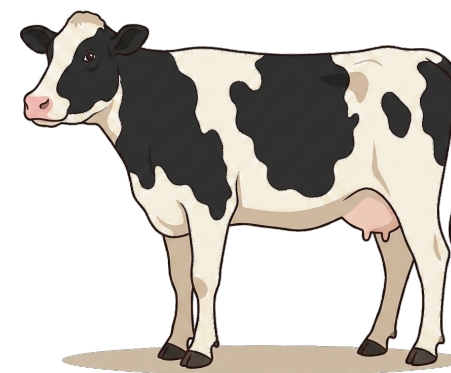
**tur** (*Bos primigenius*)

~14 500 cal BP · Šandalja, Istra



**15 000 godina**

evolucije · domestikacija · selekcija



**holstein**

danas · mliječna pasmina

Svaki genom u ovom rasponu — **SUPEK + Padobran.**



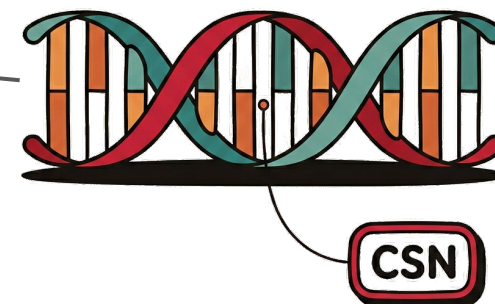
# Tri projekta — jedna infrastruktura.

**SUPEK + Padobran** — zajednički HPC temelj za sva tri projekta



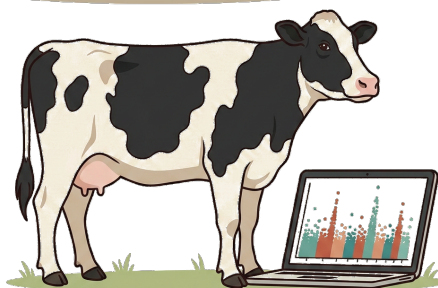
## GAbriidge

*aDNA · arheogenomika*  
HRZZ IP-2022-10-8926



## GenoPloV

*funkcionalna arheogenomika*  
EU 1-24-AGR17-581



## PhenoGeno

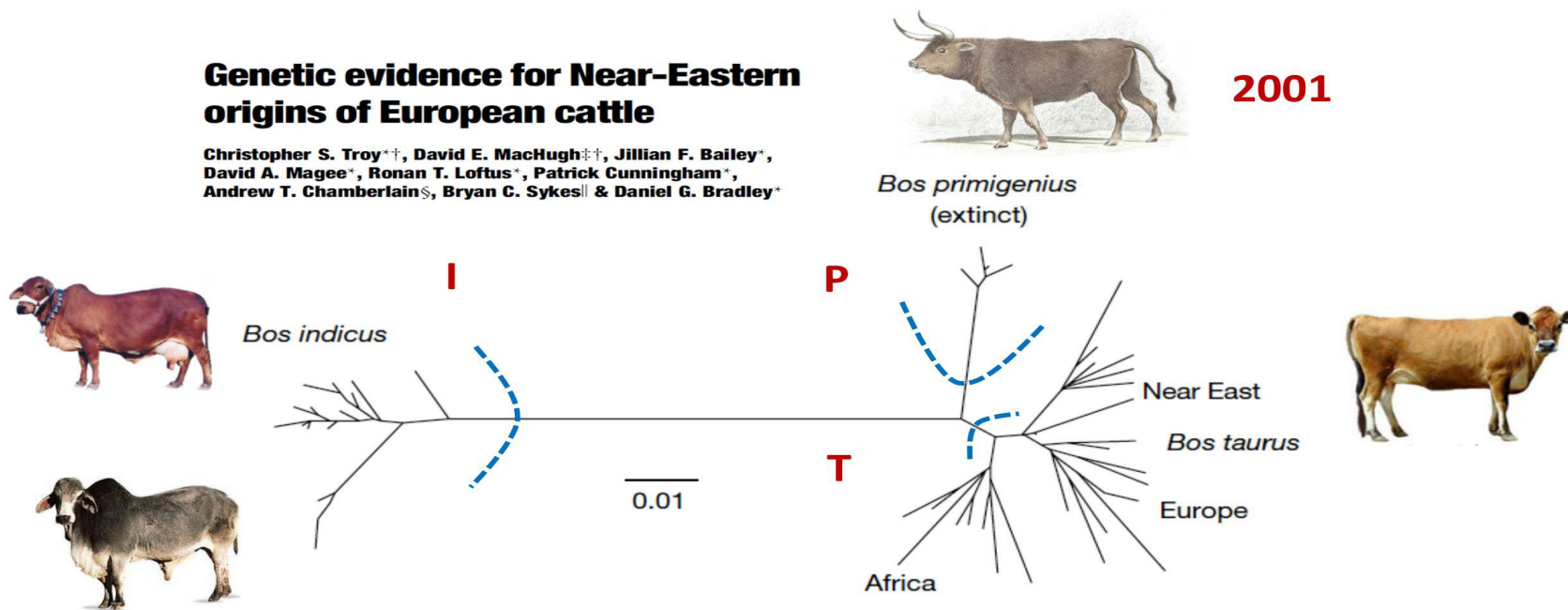
*moderno govedo · lpWGS, GWAS*  
HRZZ IP-2022-10-6914



# Dogma duže vrijeme (Troy et al. 2001).

Europska goveda = isključivo **near-eastern origin** (haplogrupe T, I, P iz Bliskog istoka).

## Dogma duže vrijeme, ali...



...pleistocenski europski tur (haplogrupa P)? **Nalazimo li ga u modernim pasminama?**



# GAbridge — Bridging the Disciplinary Gap.

HRZZ IP-2022-10-8926 · 20.12.2023 – 19.12.2027 · 198 888,39 € · PI V. Čubrić Čurik

**110+** drevnih uzoraka goveda  
*pleistocen + holocen*

**8** zemalja CEE/SEE

*Hrvatska · Austrija · Češka · Slovačka · Mađarska  
· Srbija · V. Britanija · Francuska · Danska*

**6** radnih paketa  
*arheogenetika × arheologija*

**WP1** Paleolithic aurochs diversity  
*pre-domestikacijska genetska raznolikost*

**WP2** Neolithic domestication  
*razumijevanje procesa domestikacije*

**WP3** Ancestral Recombination Graphs  
*kronologija širenja goveda*

**WP4** Povijesni hrvatski uzorci  
*srednjovjekovni goveđi genomi*

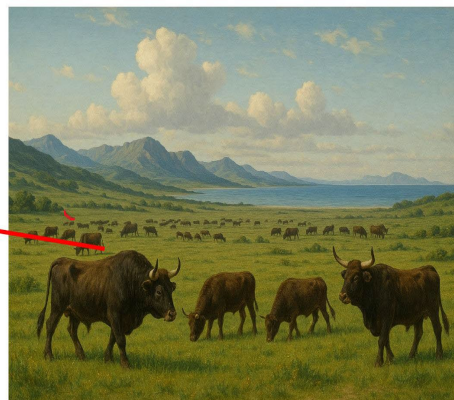
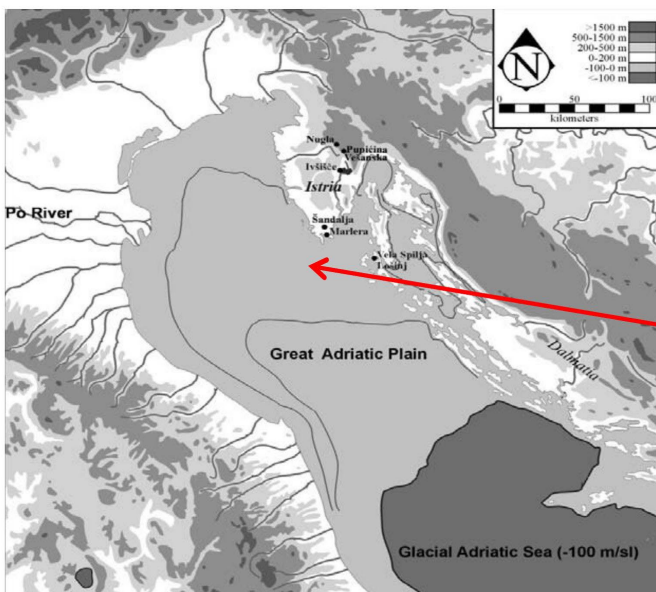
**WP5** Genetika × arheologija  
*integracija disciplinarnih dokaza*

**WP6** Diseminacija  
*management & outreach*



# Šandalja — pukotina u dogmi.

Adriatic plain - 'Garden of Eden' or 'no man's land'



## NALAZI

### Dva pleistocenska tura iz Šandalje

- ~14 500 cal BP · gornji paleolitik, post-LGM
- ~11 500 cal BP · kasni glacijal, terminalni pleistocen

Mitohondrijska haplogrupa **P** — nova „alternativna“ pod-haplogrupa  
*Klasteriraju s kasnopleistocenskim turovima sjeverne Italije → Velika jadranska nizina kao glacijalni refugij.*

Jadranska nizina — potopljeni koridor populacijske povijesti.



# Kako se grana posao — od jednog uzorka do 37 TB.

Stvarne brojke s /beegfs/group/IP-2022-10-8926 (22.04.2026).

## 1 DREVNI UZORAK

Bes2, 19× coverage · aDNA SRA

RAW

**175 GB**

12× FASTQ iz ENA (ERR3317xxx)

↓ *intermediate*

**1,1 TB**

workspace nuclear\_test/Bes2

↓ *final BAM*

**116 GB**

processed/Bes2

## GAbriidge PROJEKT

jedan HRZZ kod · 2 godine rada

drugi\_rad

**15 TB**

analize, 37 uzoraka workspace

mtDNA + PreSeq

**14 TB**

mtDNA\_novi + PreSequencing

UKUPNO

**37 TB**

na /beegfs

## 1 MODERNI UZORAK

PhenoGeno holstein · 4 lane NovaSeq

RAW

**11 GB**

8× FASTQ.gz · 4 lane × R1+R2

↓ *BWA + dedup*

**3–4 GB**

finalni BAM

× 1 200 uzoraka

**~13 TB**

planirana baza

Naš udio: **37 TB** od **384 TB** produkcijskog /beegfs (9,6 % kapaciteta).

Tih 37 TB vučemo naprijed-natrag kroz BWA → GATK → mapDamage → ANGSD — **brzi BeeGFS je nužan, ne luksuz.**



# Koliko HPC-a jede jedna aDNA analiza.

Stvarne brojke iz PBS logova — Rad 2, 37 drevnih uzoraka goveda, do sada 2 mjeseca rada.

**10 752**

**CPU-sati rezervirano**

za JEDAN distAngsd job · 64 jezgre × 168 h

**550**

**paralelnih PBS tasks**

NGSadmix array · 50 seed × K=1–11

**3 800 – 4 200**

**ukupno core-hours projekt**

kroz 2 mjeseca aktivnog rada

## SERIJSKI vs. PARALELNO

distAngsd 652 parova · 30 min/par

1 jezgra → **14 dana neprekidno**

6 workera (bigmem) → **2,25 dana**

NGSadmix array 550 tasks → **3 h wallclock**

(~40–80 simultano preko 50 nodova)

## NA LAPTOPU?

**280 GB RAM** po paru

MacBook (16 GB) → **OOM u 10 s**

**290 GB BAM data**

consumer SSD (500 MB/s) → **2 min/GB I/O**

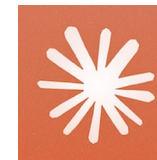
Padobran beegfs-fast → **3 GB/s** (6×)

Nije „brža verzija tvog laptopa“ — SRCE je **druga klasa infrastrukture.**



# AI nije spectator — partner u pipeline-u.

*Claude Code kao aktivni član tima*



## Debuggira

11 iteracija distAngsd-a:  
samtools container wrap,  
libgsl.so paths, --bind /tmp,  
node-RAM OOM, aptainer  
admin-kill, queue escalation  
— svaki korak AI debugging.

→ **job prošao na 1,8 TB  
bigmem, 652 para**

## Kodira

Iz rečenice opisa gotova  
skripta - 550 paralelnih  
ispitivanja ancestralne  
strukture: koje se genetske  
linije turopa prenose u  
moderna goveda, s kojim  
udjelima, s kolikom  
pouzdanošću.

→ **znanstveno pitanje u  
kod**

## Auditira metodologiju

aDNA - zamke u svakom  
koraku - damage pattern,  
mapiranje, autentikacija, rad s  
niskom pokrivenošću.  
Sistematski prolazi objavljene  
protokole i uspoređuje ih s  
našim pipeline-om.

→ **forsiran puni reset  
pipeline-a**

## Pamti

Memorija preko sesija: što  
smo dogovorili, koje metode  
su odbačene i zašto, gdje su  
BAM-ovi, što tko očekuje, sve  
odluke trajno zapisane.  
Kontinuitet kroz tjedne.

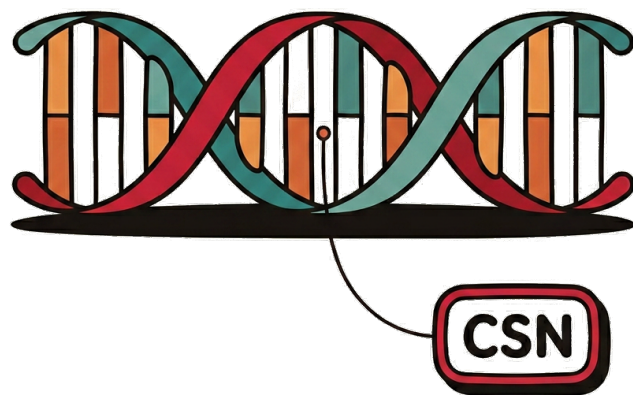
→ **ne gubi kontekst  
između sesija**

AI ne piše znanost — ali omogućava da **znanost ide paralelno** (pipeline na Padobranu, mi na novim WP-ovima).



# GenoPloV — funkcionalna arheogenomika.

EU 1-24-AGR17-581 · 2025 – 2029 · 120 k€ · PI Vlatka Čubrić Čurik



## CILJ

Rekonstruirati evolucijske promjene u funkcionalnim genima.

## FOKUSNI GENI

**CSN1S1 · CSN1S2 · CSN2 · CSN3** (*kazeini → mlijeko*)

**PLAG1** (*veličina tijela, morfologija kostura*)

## HPC POTROŠNJA

WP3 bioinformatika + WP4 katalog varijanti — **Padobran**

Od aDNA do funkcionalne varijante.



# GenoPloV — 6 radnih paketa, 4 godine.

EU 1-24-AGR17-581 · 1.10.2025 – 30.9.2029 · 120 k€ · priprema za ERC

## WP1

### Baza podataka

Izgradnja i integracija s digitalizacijom morfoloških, genetskih i arheoloških podataka.

## WP2

### Molekularna aDNA analiza

NGS sekvenciranje drevnih kosti goveda, capture enrichment, proteomika.

## WP3

### Rekonstrukcija haplotipova

Bioinformatika + haplotipska analiza funkcionalnih varijanti. → Padobran

## WP4

### Katalog funkcionalnih varijanti

Kazeini (CSN1S1, CSN1S2, CSN2, CSN3), PLAG1 — izgubljene varijante. → Padobran

## WP5

### Popularizacija i komercijalizacija

Model prijenosa znanja prema industriji i javnosti.

## WP6

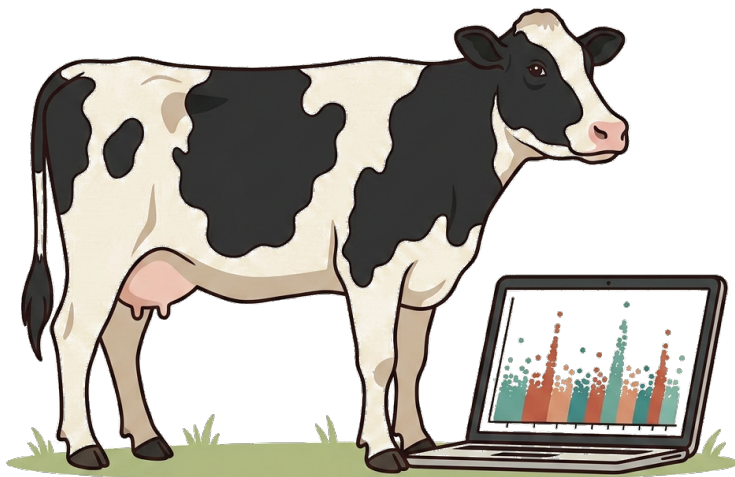
### Upravljanje projektom

Koordinacija, diseminacija, priprema ERC prijave.



# PhenoGeno — stotine genoma, jedan PBS queue.

HRZZ IP-2022-10-6914 · PI Ino Čurik · moderno govedo · lpWGS · GWAS



## METODA

### Low-pass whole-genome sequencing (lpWGS)

*zamjenjuje tradicionalne SNP čipove za genomsku selekciju*

## POPULACIJA

GWAS na hrvatskoj holstein populaciji

## HPC POTROŠNJA

**\$ qsub -J 1-512 pipeline.pbs**

*512 paralelnih uzoraka — ovdje paralelizacija radi dobro.*

Tamo gdje paralelizacija radi — PBS queue **vrti stotine jobova.**



# PhenoGeno — 8 radnih paketa, 1 200 holstein životinja.

HRZZ IP-2022-10-6914 · 29.12.2023 – 28.12.2027 · 197 989 € · PI Ino Čurik

<b>WP1</b>	<b>WP2</b> · HPC	<b>WP3</b> · HPC	<b>WP4</b> · HPC
<b>Moderna fenomika</b> FTIR spektrometrija za svojstva zgrušavanja mlijeka.	<b>Molekularne analize</b> IpWGS genotipizacija 1 200 holstein životinja.	<b>Mitogenomika</b> Haplotipovi + mikrobiota mlijeka.	<b>Računalne performanse</b> Baze podataka, bioinformatički workflow-i.
<b>WP5</b> · HPC	<b>WP6</b>	<b>WP7</b>	<b>WP8</b>
<b>GWAS + GBLUP</b> Asocijacijske studije, genetski parametri.	<b>Upravljanje raznolikošću</b> Inbreeding depresija, štetne mutacije.	<b>Diseminacija</b> Transfer znanja prema industriji.	<b>Management</b> Koordinacija i monitoring.

Crveno označeni paketi žive na Padobranu. **Partneri: Roslin Institute, BOKU, Czech Uni.**



# Ne jedan pipeline — tri različita opterećenja.

Tri različite računske logike u našim projektima. Svaka traži drugu vrstu čvora — jedan server nikad ne pokriva sve tri.

## MEMORIJSKA

*jedan veliki čvor · svi uzorci u istoj RAM-memoriji*

### distAngsd · GAbriage aDNA

pairwise udaljenost svih drevnih turova — ne može se podijeliti po kromosomu.

### 280 GB RAM po paru · 1,8 TB ukupno

na običnom serveru (200 GB) → OOM u nekoliko minuta.

→ **Padobran bigmem čvor (64 jezgre · 1,8 TB)**

## PARALELNA

*stotine/tisuće malih jobova · svaki uzorak zaseban job*

### IpWGS · PhenoGeno holstein

1 200 modernih uzoraka — svaki genom svoj zasebni posao.

### NGSadmix · GAbriage

550 jobova (50 random seeds × 11 K-vrijednosti) paralelno.

→ **Padobran PBS array (qsub -J 1–1200)**

## GPU-SERIJSKA

*jedan dugi lanac · brza GPU kartica*

### BEAST · GAbriage filogenija

800 milijuna MCMC iteracija × 207 taksona za datiranje turova.

### BEAGLE GPU knjižnica

10–30× ubrzanje nad čistim CPU-om; tjedan umjesto mjeseci.

→ **SUPEK GPU čvor**

Srce (Padobran + SUPEK) pokriva sve tri logike. **Lokalni server — nijednu u cijelosti.**



# Na lokalnom serveru: nemoguće. Na Srce HPC-u: izvedivo.

KORAK	LOKALNI SERVER 200 GB RAM · 64 jezgre	SRCE HPC Padobran + SUPEK	ZAŠTO TO ZAHTIJEVA
<b>distAngsd · PMR analiza</b> <i>pairwise genetska udaljenost svih drevnih turova (GABridge)</i>	<b>X OOM · exit 137</b> <i>6 krahova prije 1. uspjeha</i>	<b>✓ bigmem · 1,8 TB RAM</b> <i>64 jezgre · 168 h</i>	<i>Memorija nije dijeljiva — svi uzorci moraju stati u isti RAM.</i>
<b>BEAST MCMC · datiranje</b> <i>800 milijuna iteracija × 207 taksona (GABridge filogenija)</i>	<b>X mjesecima CPU-a</b> <i>serijski lanac — nema što paralelizirati</i>	<b>✓ SUPEK GPU · ~tjedan</b> <i>BEAGLE knjižnica · CUDA</i>	<i>GPU daje 10–30× ubrzanje u MCMC-u. CPU to ne može pratiti.</i>
<b>IpWGS PhenoGeno</b> <i>1 200 modernih goveda · 1 uzorak = 1 job</i>	<b>X mjesec sekvencijalno</b> <i>jedan po jedan uzorak</i>	<b>✓ PBS array 1–1 200</b> <i>~2 dana wallclock</i>	<i>Paralelizabilno po uzorku — PBS queue vrti sve odjednom.</i>

Tri koraka, tri razloga — **jedan server nije skalabilan ni memorijom, ni GPU-om, ni paralelizmom.**



# Brojke koje pokreću znanost.

Sveučilište u Zagrebu Agronomski fakultet na Srce HPC-u, 2025. — *javni Srce godišnji izvještaj.*

**684 840**

**CPU-sati**

*Padobran + Supek · ≈ 78  
CPU-godina u 2025.*

**34 417**

**PBS poslova**

*Padobran · ~95 poslova /  
dan, non-stop*

**2,35 %**

**cijelog Padobrana**

*7. mjesto među 44  
ustanova*

**1 392**

**GPU-sati**

*Supek BEAGLE · BEAST  
MCMC datiranje*

**37 TB**

**/beegfs aktivni**

*GAbridge workspace ·  
stalni I/O*

78 CPU-godina rada u 2025., kondenzirano u 12 mjeseci.

Bez Srce HPC-a — **ništa od ovoga.**



# Suradnici · partneri · financiranje.

Bez ovih ljudi i ovih projekata — ne bilo priče.

## HRVATSKA

### Agronomski fakultet · AnGen

Vlatka Čubrić Čurik · *PI GAbridge, GenoPloV*

Ino Čurik · *PI PhenoGeno*

Maja Ferenčaković

Ivana Keršić

Zvonimir Mijadžiković

### Arheologija · muzeji · veterina

Rajna Šošić Klindžić · *FF Zagreb*

Siniša Radović · *HAZU Kvaternik*

Dario Vujević · *Sveuč. u Zadru*

Hrvoje Vulić · *Muzej Vinkovci*

Maja Krznarić Škrivanko · *Muzej Vinkovci*

Goran Tomac

Maja Grgurić Srzentić

Dinko Novosel

## MEĐUNARODNI PARTNERI

### UK · Cambridge

Preston T. Miracle · Ana Prohaska

### Danska · Copenhagen (GeoGenetics)

Eske Willerslev · Lasse Vinner · Martin Sikora

### Francuska · INRAE GABI

Aurélien Capitan

### Austrija · NHM Beč

Ursula Göhlich · Erich Pucher · Konstantina Saliari

### Slovačka · SPU Nitra

Radovan Kasarda · Nina Moravčíková · Anna Đurišová · Zora Bjelihova

### Češka · CULS Prag

Luboš Vostrý

### Mađarska

László Daróczi-Szabó · Márta Daróczi-Szabó · Gáspárdy András ·

Maróti-Agóts Ákos

### Srbija

Sonja Vuković · Vesna Dimitrijević · Ivana Živaljević

## FINANCIRANJE · PROJEKTI

### GAbridge

HRZZ IP-2022-10-8926 · 2023–2027

### PhenoGeno

HRZZ IP-2022-10-6914 · 2023–2027

### GenoPloV

EU institucionalni 1-24-AGR17-581 · 2025–2029

### LAZAG

KK.01.1.1.04.0058 · 2021–2026

### MendTheGap

EU Horizon 692249 · 2016–2019

Srce HPC (SUPEK + Padobran) · [angen.agr.hr](http://angen.agr.hr) · arheogenetika i genomika goveda u Hrvatskoj



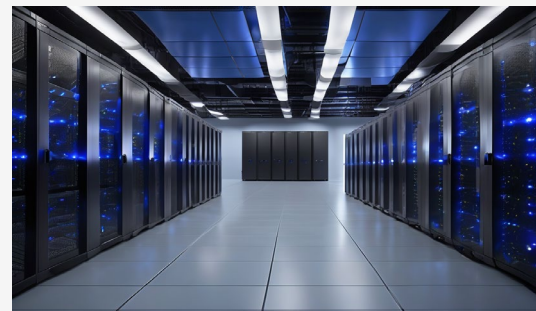
# Projekt pisan šest godina — stopiran pred javnu nabavu



## „Kakvi to kompjuteri koštaju 300.000 €?“

- 256 CPU jezgri
- 2× NVIDIA H200 GPU
- ~96 TB NVMe SSD
- Slingshot 200 Gb
- 5-god. HPE podrška

To nije laptop.



## „Zašto na SRCE?“

- 1.317 m<sup>2</sup>, 32 ormara
  - 1.100 kW priključak
  - 750 kVA UPS
  - 1.250 kVA dizel backup
  - 500 kW vodeni chiller
- HR-ZOO ZG2 već postoji.

466.987,78 kn javnog novca — BAČENO U SMEĆE.



# Hvala na pažnji.



Drago **Srce**,  
*sretan 55. rođendan!*




*S ljubavlju, AnGen tim*



55

*godina*



**P.S.** Nemojte stariti brzinom kojom mi trošimo CPU vrijeme!