

*SPECIES SELECTION FOR LARGE-SCALE GENOME PROJECTS –
AN AUTOMATED PROCESS BASED ON
EXPLICITLY BOTTOM-UP DEFINED CRITERIA*

Torsten H. Struck

t.h.struck@nhm.uio.no

 @fez_nhm

www.frontiersinevolutionaryzoology.com

blog.annelida.de

Natural History Museum
University of Oslo

The Sampling and Sample Processing (SSP) committee

Coordinator

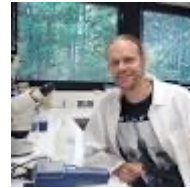


Rita
Monteiro

Steering committee



Katja
Reichel



Jaakko
Pohjoismäki



Astrid
Böhne



Jennifer
Leonard



Olga Vinnere
Pettersson

Contributors



Thomas
Marcussen



Rebekah
Oomen



Ana
Riesgo



Bernhard
Hausdorf



Christian
de Guttry



Alice
Minotto



Carmela
Gissi



Andrii
Tarieiev



Joao
Pimenta



Isabelle
Florent



Camila
Mazzoni



Rosa
Fernandez



Seanna
McTaggart



Jose Melo
Ferreira



samples@erga-biodiversity.eu



Why select species?

Large scale genome consortia



- More species to be sequenced than funding allows
- Develop criteria and strategies on how to select species
- ERGA/BGE choose a bottom-up approach involving SSP

How to select species?

Compile and assess individual selection criteria

- Definition and possible impact
- Decide on relevant ones
- Group into different stages and categories

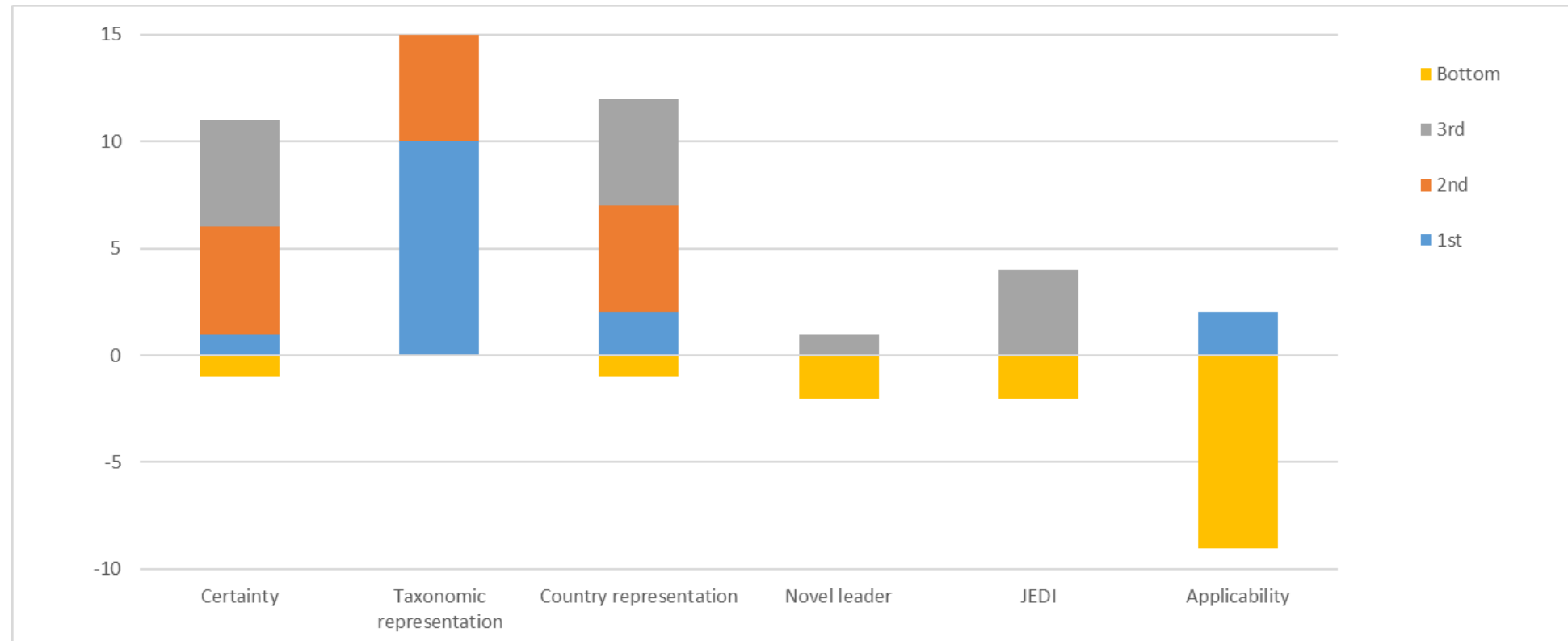
Four stages in the selection process:

- Exclusion (3 criteria)
- Prioritization (11 criteria in 6 categories)
- Feasibility (11 criteria)
- Permits (3 criteria)

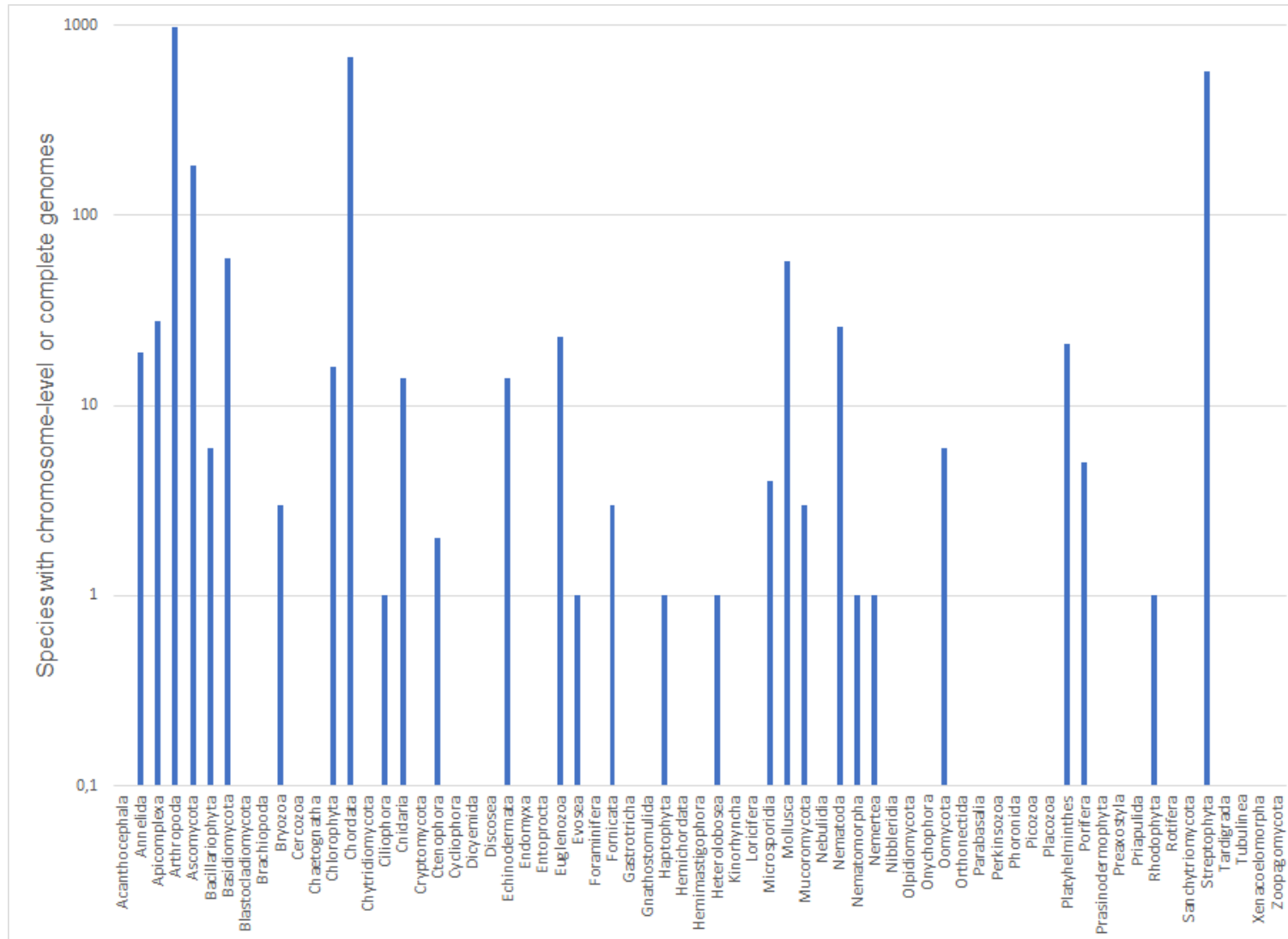
How to select species?

Develop different ranking processes (models)

- Determine importance of different categories

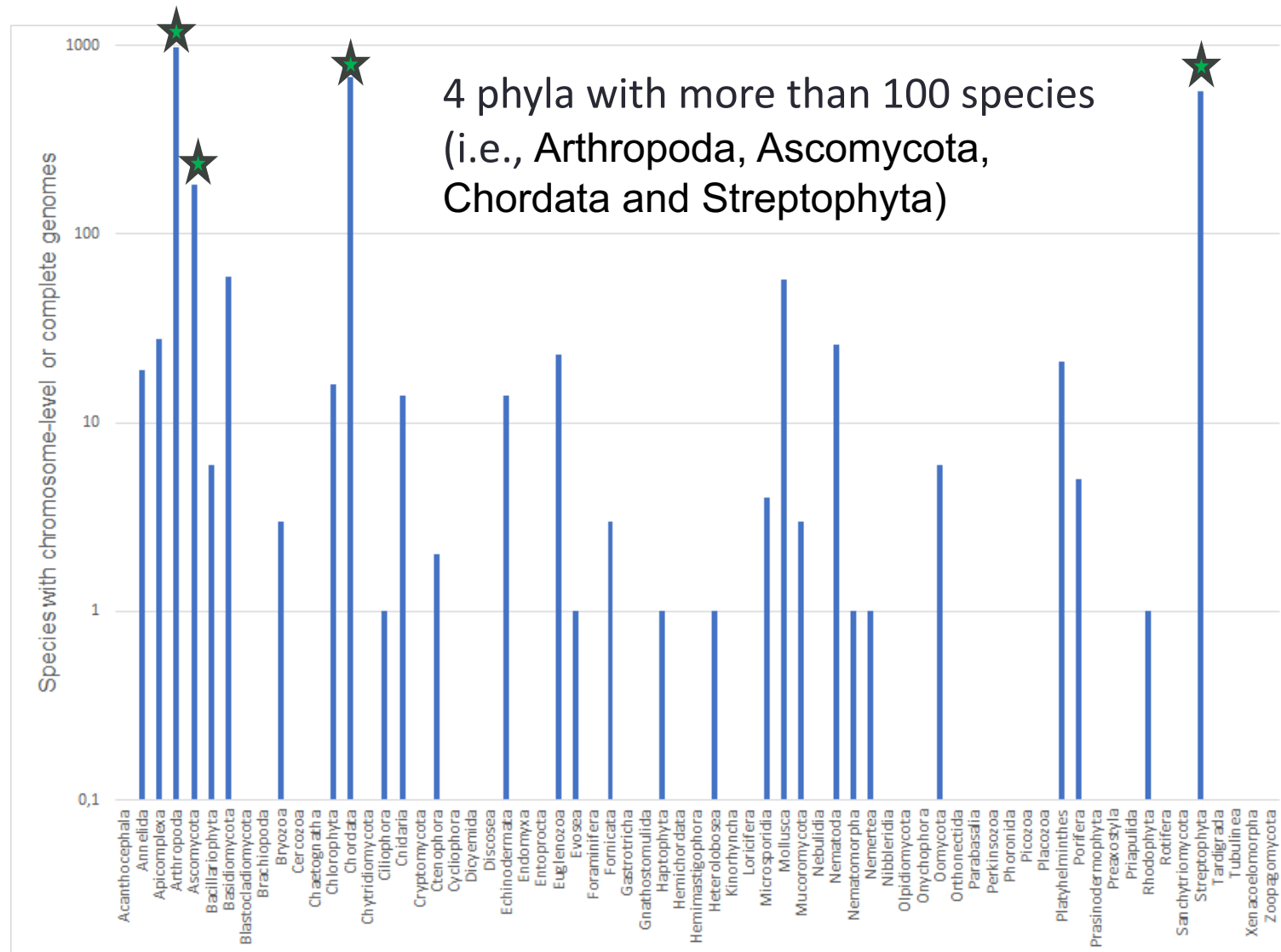


Reference or complete genomes per phylum



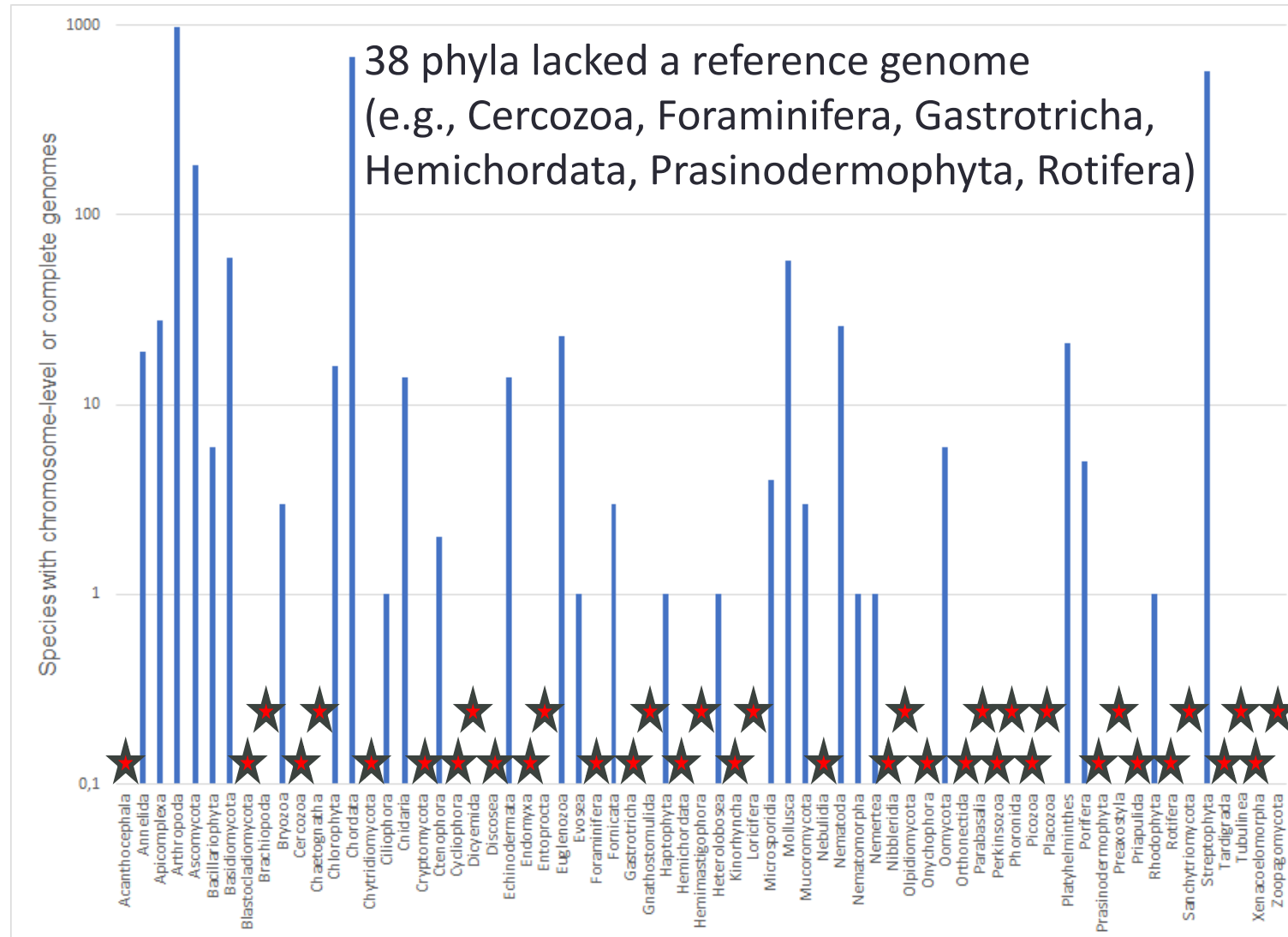
- Number of species with chromosome-level or complete genomes per phylum
- Y-axis is logarithmic
- Retrieved from GoAT on July 28th 2023.

Reference or complete genomes per phylum



- Number of species with chromosome-level or complete genomes per phylum
- Y-axis is logarithmic
- Retrieved from GoAT on July 28th 2023.

Reference or complete genomes per phylum



- Number of species with chromosome-level or complete genomes per phylum
- Y-axis is logarithmic
- Retrieved from GoAT on July 28th 2023.

How to select species?

Develop different ranking processes (models)

- Determine importance of different categories
- 8 different models applying different ranking orders, but same weights within each category

Level	Model #7
1 st	Taxonomic representation
2 nd	Country representation & Certainty
3 rd	JEDI
4 th	Applicability & Novel leader

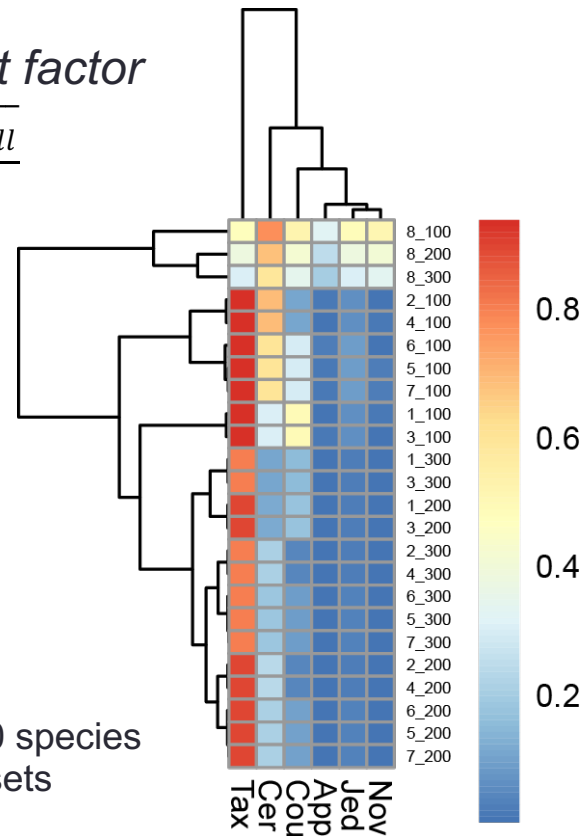
How to select species?

Test the different models using simulated and empirical data

- Assess if the desired effect occurs

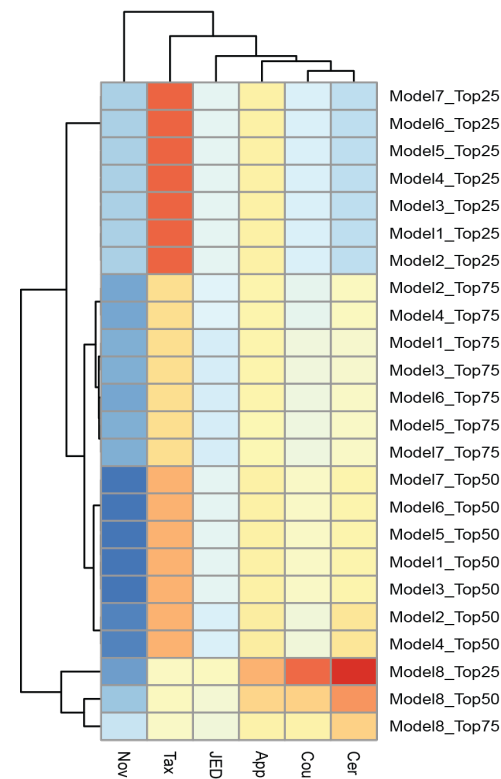
Relative enrichment factor

$$\frac{Weight_{Top} - Weight_{All}}{Max_{poss.} - Weight_{All}}$$

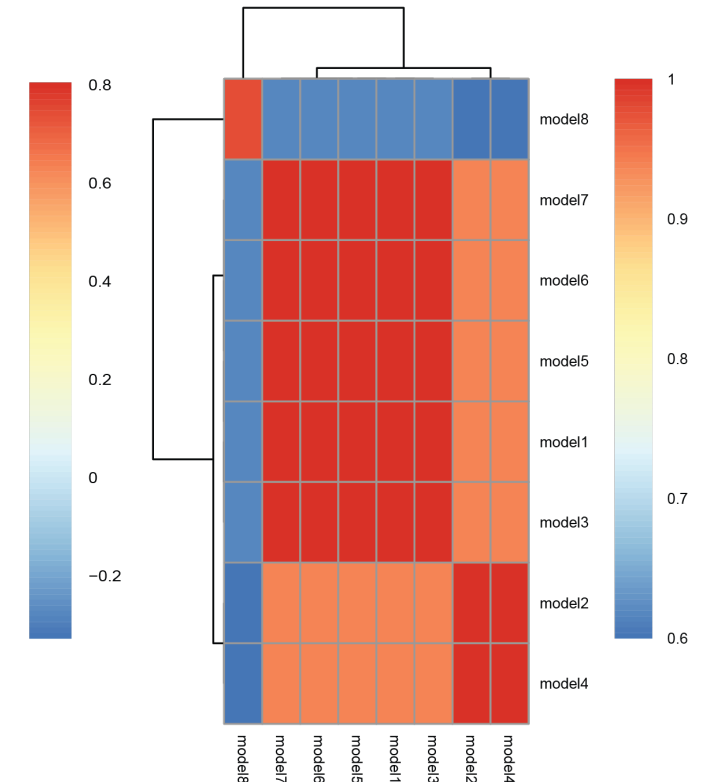


- Each dataset with 1000 species
- 72,900 simulated datasets

Enrichment



Intersection between 50 selected species

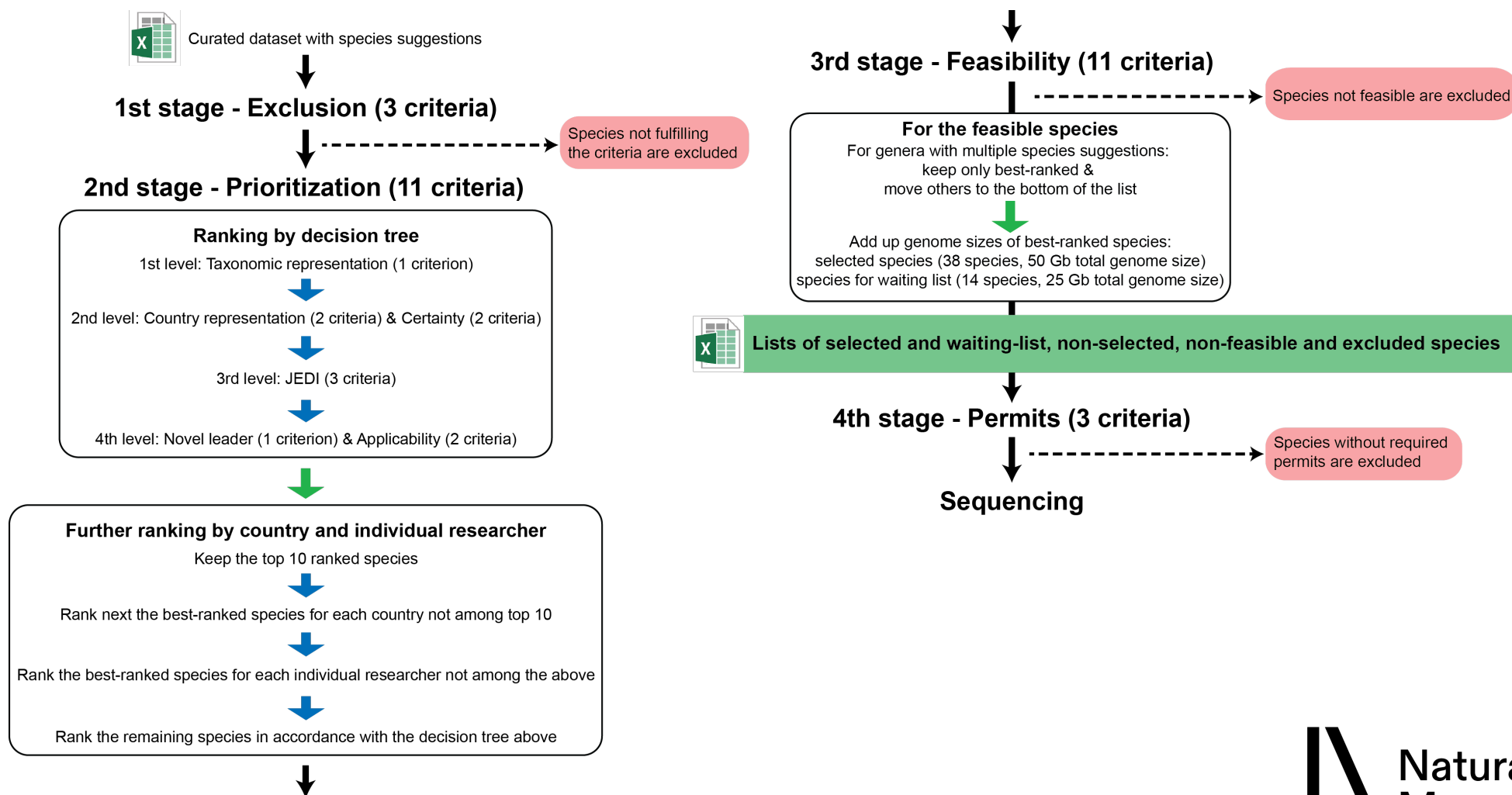


How to select species?

Presentation of selected models and results at ERGA council with representatives from each European country

- Vote on the preferred species prioritization process
- Ranking process #7 with 89.7%
- Implementing an additional ranking step favoring both countries and individual researchers with 43.6%
- Implement the additional ranking after stage 2 with 44,8%

The species selection process of BGE for community sequencing (2 rounds)



The selected species after stage 3 (round 1)

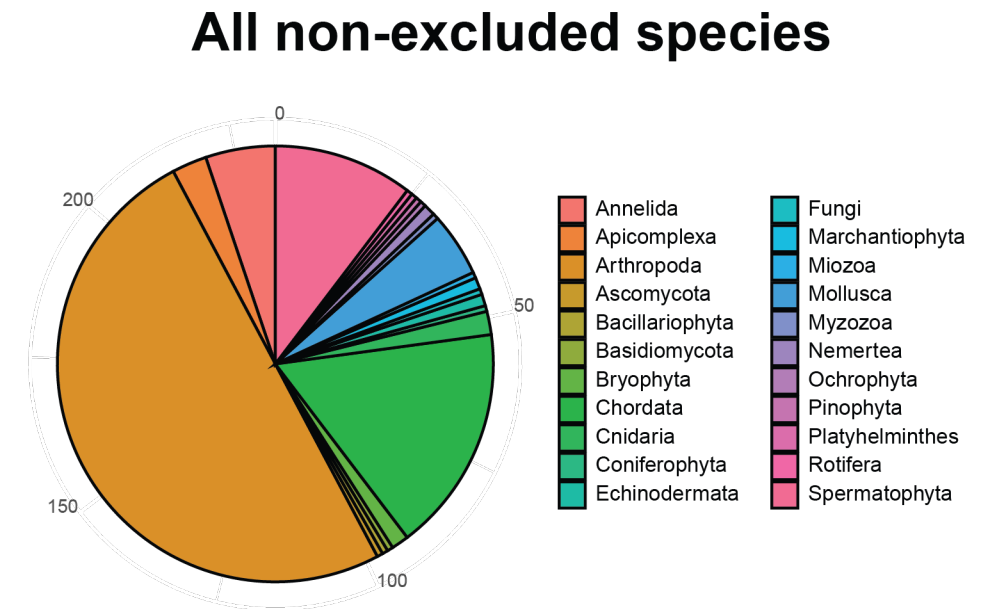
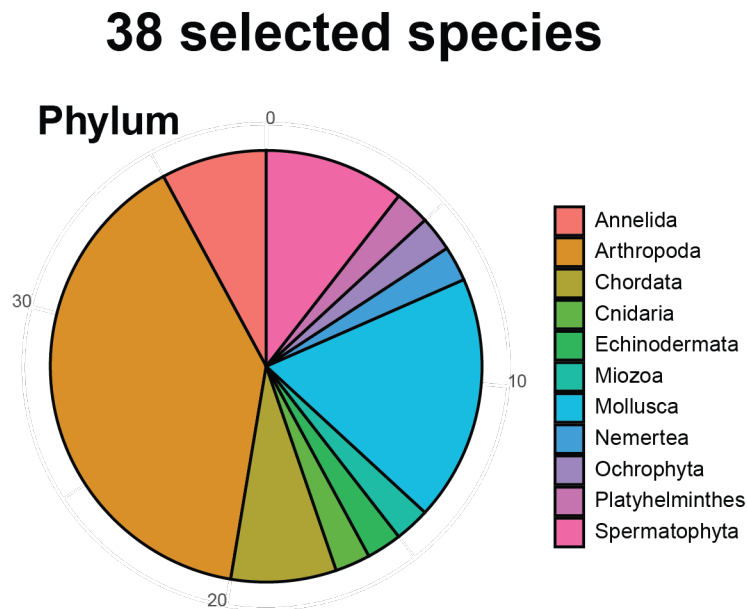
<i>Wirenia</i>	<i>argentea</i>	Mollusca	Norway
<i>Hydrurus</i>	<i>foetidus</i>	Ochrophyta	Norway
<i>Tetrademna</i>	<i>longissimum</i>	Nemertea	Spain
<i>Marifugia</i>	<i>cavatica</i>	Annelida	Croatia
<i>Proasellus</i>	<i>hercegovinensis</i>	Arthropoda	Croatia
<i>Diadema</i>	<i>setosum</i>	Echinodermata	Greece
<i>Alexandrium</i>	<i>minutum</i>	Miozoa	France
<i>Tritomurus</i>	<i>scutellatus</i>	Arthropoda	Slovenia
<i>Herpes</i>	<i>porcellus</i>	Arthropoda	Hungary
<i>Hydroglyphus</i>	<i>hamulatus</i>	Arthropoda	Sweden
<i>Omalisus</i>	<i>fontisbellaquei</i>	Arthropoda	Czechia
<i>Protonethes</i>	<i>ocellatus</i>	Arthropoda	Montenegro
<i>Hermodice</i>	<i>carunculata</i>	Annelida	Spain
<i>Phorcus</i>	<i>turbinatus</i>	Mollusca	Spain
<i>Alpioniscus</i>	<i>balthasari</i>	Arthropoda	Croatia
<i>Gammarus</i>	<i>oceanicus</i>	Arthropoda	Poland
<i>Gluvia</i>	<i>dorsalis</i>	Arthropoda	Spain
<i>Phyllidia</i>	<i>flava</i>	Mollusca	Spain
<i>Xylophaga</i>	<i>dorsalis</i>	Mollusca	Spain

<i>Eunicella</i>	<i>cavolini</i>	Cnidaria	France
<i>Trienophorus</i>	<i>nodulosus</i>	Platyhelminthes	Finland
<i>Distaplia</i>	<i>bermudensis</i>	Chordata	Italy
<i>Cistus</i>	<i>crispus</i>	Spermatophyta	France
<i>Mytilopsis</i>	<i>leucophaeata</i>	Mollusca	France
<i>Pyrrhula</i>	<i>murina</i>	Chordata	Portugal
<i>Androsace</i>	<i>saussurei</i>	Spermatophyta	France
<i>Arnica</i>	<i>montana</i>	Spermatophyta	Germany
<i>Graellsia</i>	<i>isabellae</i>	Arthropoda	Spain
<i>Ailoscolex</i>	<i>lacteospumousus</i>	Annelida	Spain
<i>Botrylloides</i>	<i>israeliense</i>	Chordata	Italy
<i>Erebia</i>	<i>palarica</i>	Arthropoda	Spain
<i>Anodonta</i>	<i>cygnea</i>	Mollusca	Switzerland
<i>Cedrorum</i>	<i>azoricus</i>	Arthropoda	Portugal
<i>Dysdera</i>	<i>ambulatenta</i>	Arthropoda	Spain
<i>Rosalia</i>	<i>alpina</i>	Arthropoda	Hungary
<i>Trachusa</i>	<i>byssina</i>	Arthropoda	Germany
<i>Alyssoides</i>	<i>utriculata</i>	Spermatophyta	Switzerland
<i>Anisus</i>	<i>vorticulus</i>	Mollusca	Germany

Effect of the selection process

- 11 (round 1) or 14 (round 2) of the originally 22 suggested phyla
- 10 out of excluded 11 due to the feasibility check (round 1)
- 3 excluded and 5 included due to the feasibility check (round 2)

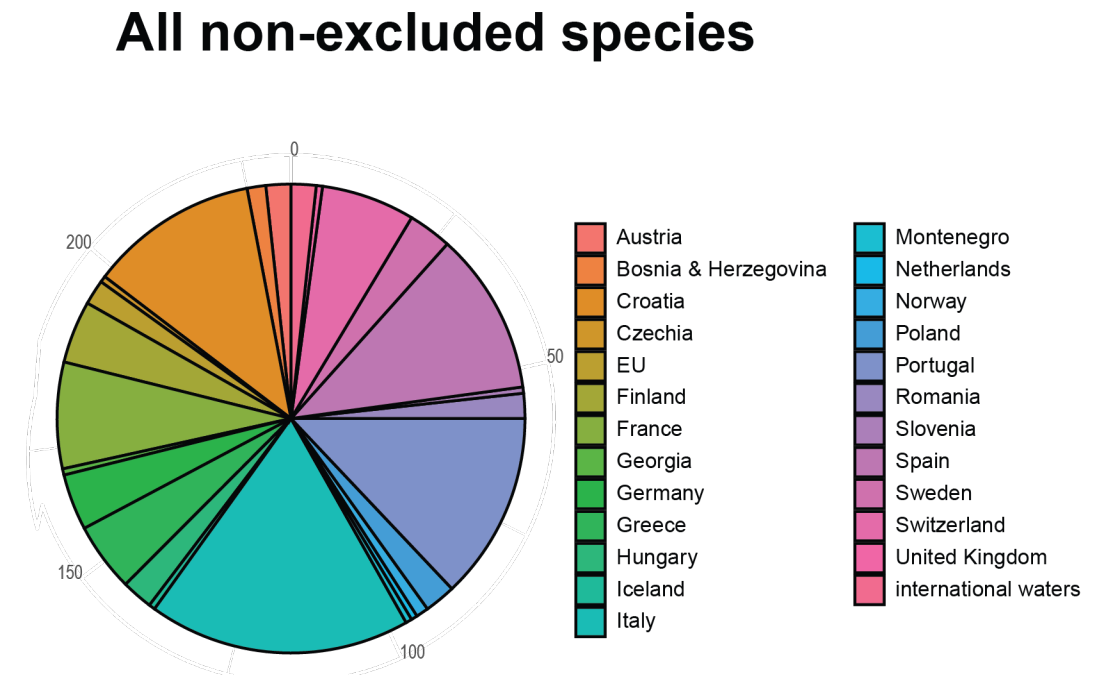
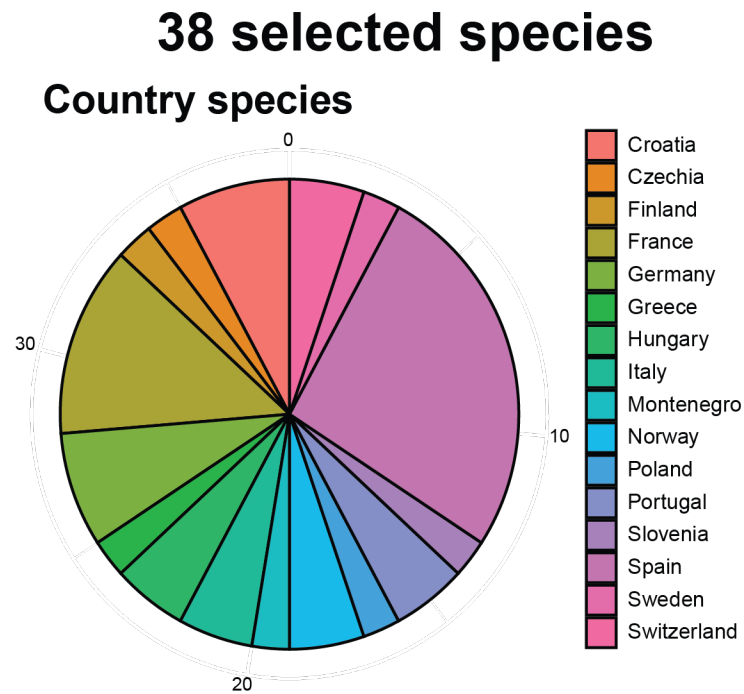
Round 1



Effect of the selection process

- 16 out of 25 countries (round 1), 18 out of 29 countries (round 2)
- Again, mostly due to the feasibility check

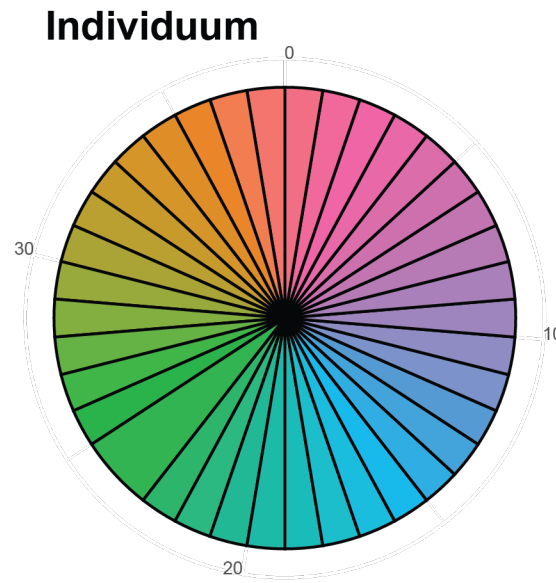
Round 1



Effect of the selection process

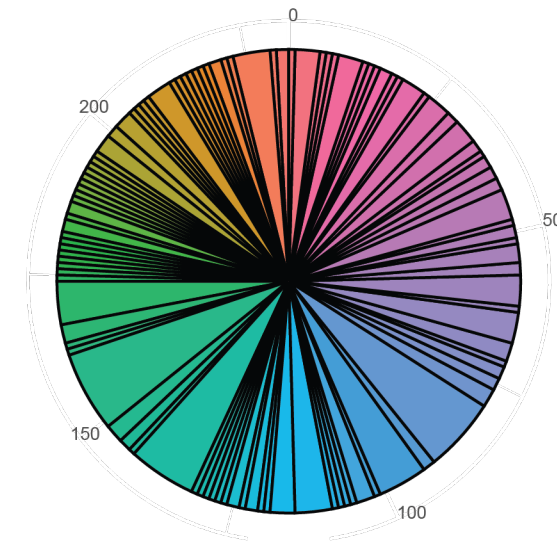
- Only one individual researcher with 2 species (round 1), one with 3 species, three with 2 species (round 2)
- One species – one researcher

38 selected species



Round 1

All non-excluded species



Effect of the feasibility check

Round 1: 110 of 230 species considered not-feasible (47.8%)

Round 2: 103 of 268 species considered not feasible (38.4%)

Both rounds: 213 of 426 species considered not feasible (50.0%)

Too small or too large

- 19.6% due to small body size or low numbers of nucleated cells
- 23.8% due to large genome size
- 36.2% due to one of the two or both

Too bad condition

- 31.2% not snap-frozen
- 25.0% not possible to preserve within 5 minutes of their death
- 25.4% not possible to maintain a strict cold chain at -70°C

Too challenging to collect

- 17.5% not already collected or not easy to obtain

Conclusions

- Species selection process developed bottom-up with community involvement
- Four stage process with a total of 28 objective criteria
- Tested using simulated and empirical data
- Automated selection process based on a R script
- **More research effort is needed to make the sequencing of reference genomes feasible for a much larger part of biodiversity**

General applicability

Khrono.no Friday 05.04.2024 - 15:49

OPPTAKSMELDINGEN

Nye opptaksregler skaper debatt om mer karakterpress

— *Men hvor stor mener dere førstevalgsgruppen bør være?*

— Det må vi se litt nærmere på. Vi kom ikke i mål med de beregningene nå. Det er også fordi det også er avhengig av hvor mange tilleggspong som blir stående igjen når Stortinget har gjort sine vedtak. Vi kunne hatt en mening om det hvis vi visste at det var vårt forslag som gikk gjennom, men hvis det blir justeringer i Stortinget vil kanskje vurderingene bli annerledes. Så dermed kommer vi tilbake til dette, sier Hoel.

- Essentially prioritization process
- Only one model, not several
- Not tested because parliament might make changes
- **Testing would provide better basis for judgement if desired political effects will be accomplished**

Acknowledgement



BIODIVERSITY GENOMICS EUROPE receives funding from the European Union's Horizon Europe Research and Innovation Action. <https://biodiversitygenomics.eu/>



Thanks for your attention!