

WORKSHOP ON
PHYLOGENOMICS

data



orthology inference



alignment



trimming/masking



analyses

data

WORKSHOP ON GENOMICS

DATABASES (Ensembl, UNIPROT, SRA...)

Download

Reduce
redundancy

*DE
NOVO*

Raw reads

Sanitize

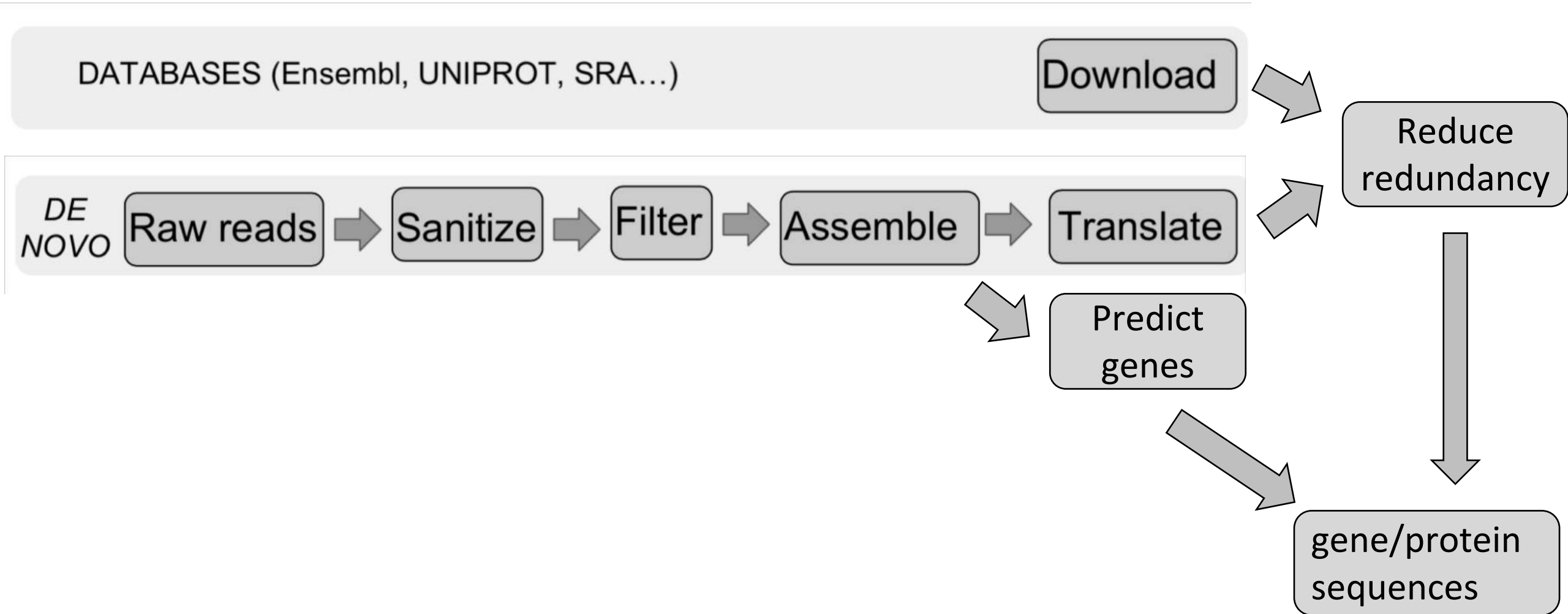
Filter

Assemble

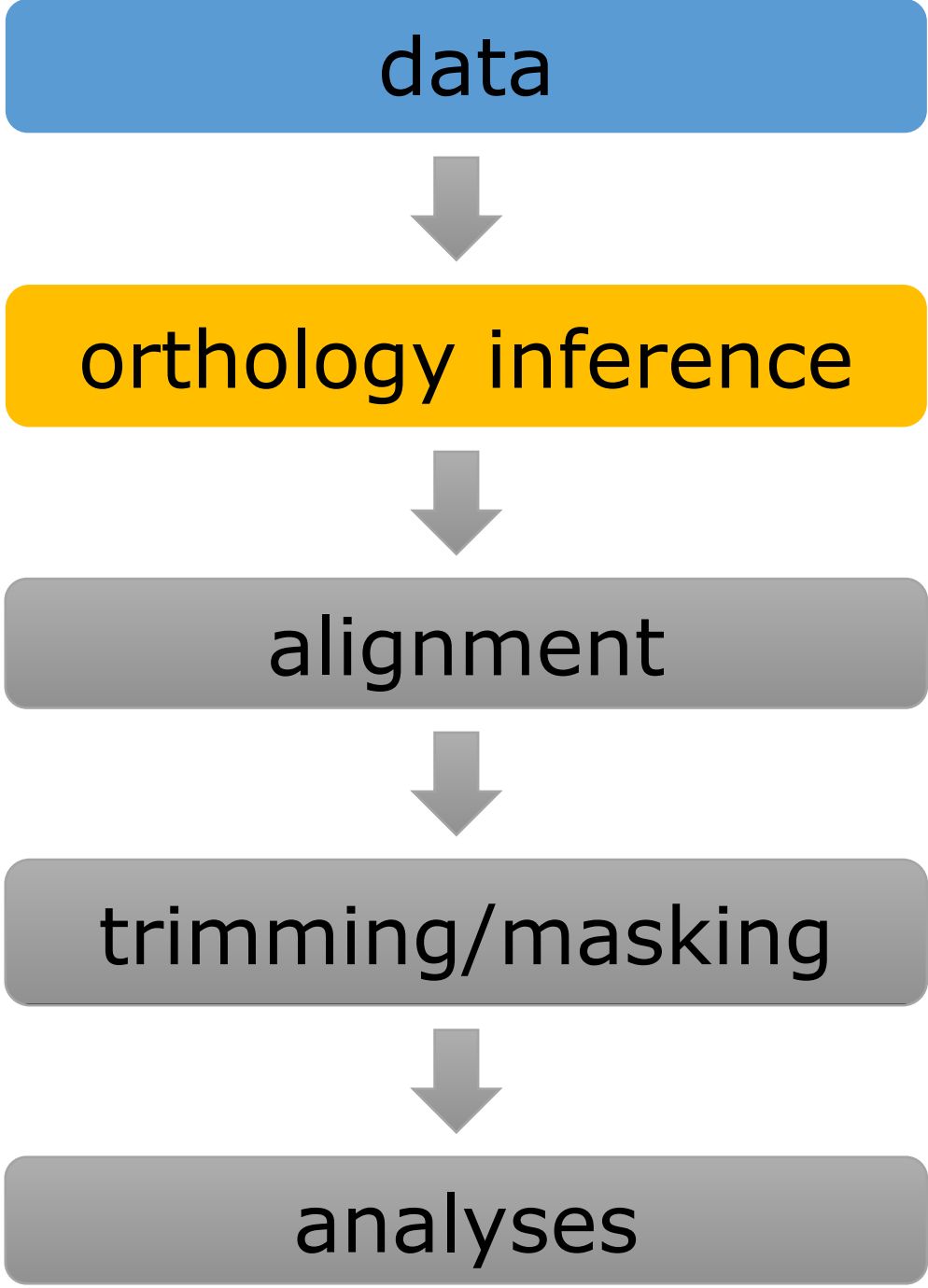
Translate

Predict
genes

gene/protein
sequences



data



```
graph TD; A[data] --> B[orthology inference]; B --> C[alignment]; C --> D[trimming/masking]; D --> E[analyses];
```

The diagram illustrates a five-step process for orthology inference. It begins with 'data' in a blue box, followed by 'orthology inference' in a yellow box, then 'alignment', 'trimming/masking', and finally 'analyses', all in grey boxes connected by downward arrows. To the right of the 'orthology inference' step, there are two grey boxes containing lists of methods and data sources.

orthology inference

alignment

trimming/masking

analyses

- Best reciprocal hits
- Inparanoid
- OrthoMCL
- PorthoMCL
- OrthoFinder
- Tree based predictions

- Use pre-calculated orthology predictions
- Use an existing dataset

data

```
graph TD; A[data] --> B[orthology inference]; B --> C[alignment]; C --> D[trimming/masking]; D --> E[analyses];
```

orthology inference

alignment

trimming/masking

analyses

- PASTA
- Bali-Phy
- MAFFT
- UPP
- Muscle

data



orthology inference



alignment



trimming/masking



analyses

- BMGE
- trimAl
- Gblocks
- Prequal

data



orthology inference



alignment



trimming/masking



analyses

analyses

1:1 orthologs



concatenation

- Manual
- Sequence Matrix
- CONCATENATOR
- catfasta2phym
- FASconCAT-G
- Phyutility



data matrix

analyses

1:1 orthologs

orthologs and/or paralogs

data matrix

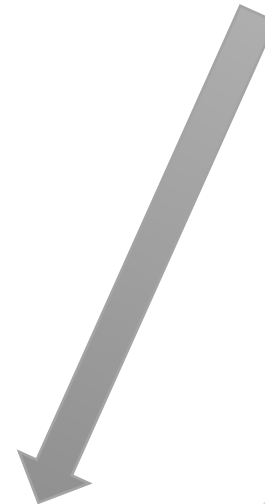
infer individual gene
trees

PartitionFinder

partitioning

model selection

- RAxML-NG
- IQ-tree
- ModelTest



analyses

1:1 orthologs



infer multigene
trees

orthologs and/or paralogs



infer individual gene
trees

- RAxML
- IQ-tree
- PhyML
- FastTree
- PhyloBayes
- MCMCTree
- SVDQ

analyses

```
graph TD; analyses[analyses] --> networks[networks]; analyses --> multitree[Infer multigene trees]; analyses --> indtree[infer individual gene trees]; networks --> intro[Introgression/hybridization events]; networks --> tools[• SplitsTree<br>• Dendroscope<br>• PhyloNet<br>• ALE];
```

networks

Infer multigene
trees

infer individual gene
trees

Introgression/hybridization
events

- SplitsTree
- Dendroscope
- PhyloNet
- ALE

analyses

1:1 orthologs



infer multigene
trees

orthologs and/or paralogs



infer individual gene
trees

ancestral state
reconstruction

horizontal gene
transfer detection

comparative
genomics

gene trees/species
tree reconciliation

evolutionary
relationships

gene family evolution

...

gene duplication/loss

visualization

visualization

- Phylo.IO
- iTOL
- EVOLVIEW
- FigTree
- DensiTree
- TreeGraph2
- ETE3
- jsPhyloSVG
- DiscoVista
- Dendroscope
- SplitsTree
- R (many repositories)

Open Lab questions for further discussion in groups

What kind of **data** I need to answer a specific question?

I don't have **data**

There is not enough reference **data**

Contaminated **data**

...