# Using iPyrad for ddRAD data
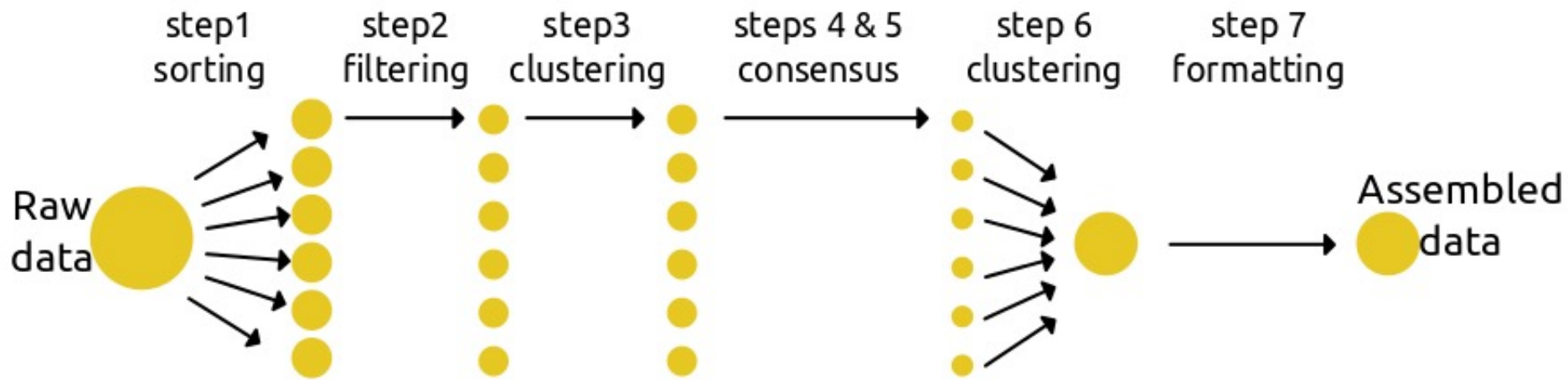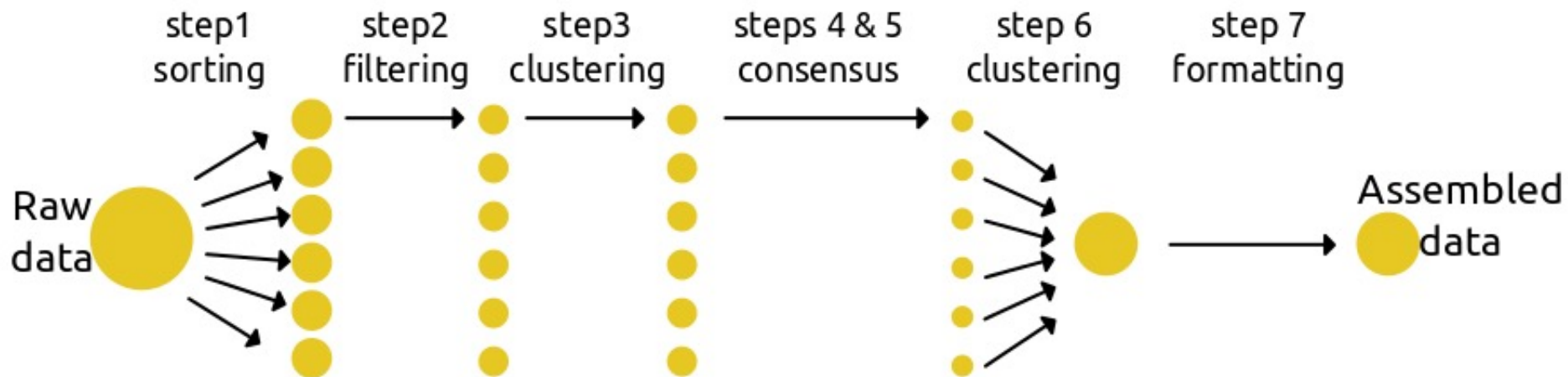
# Library preparation: break down into problems
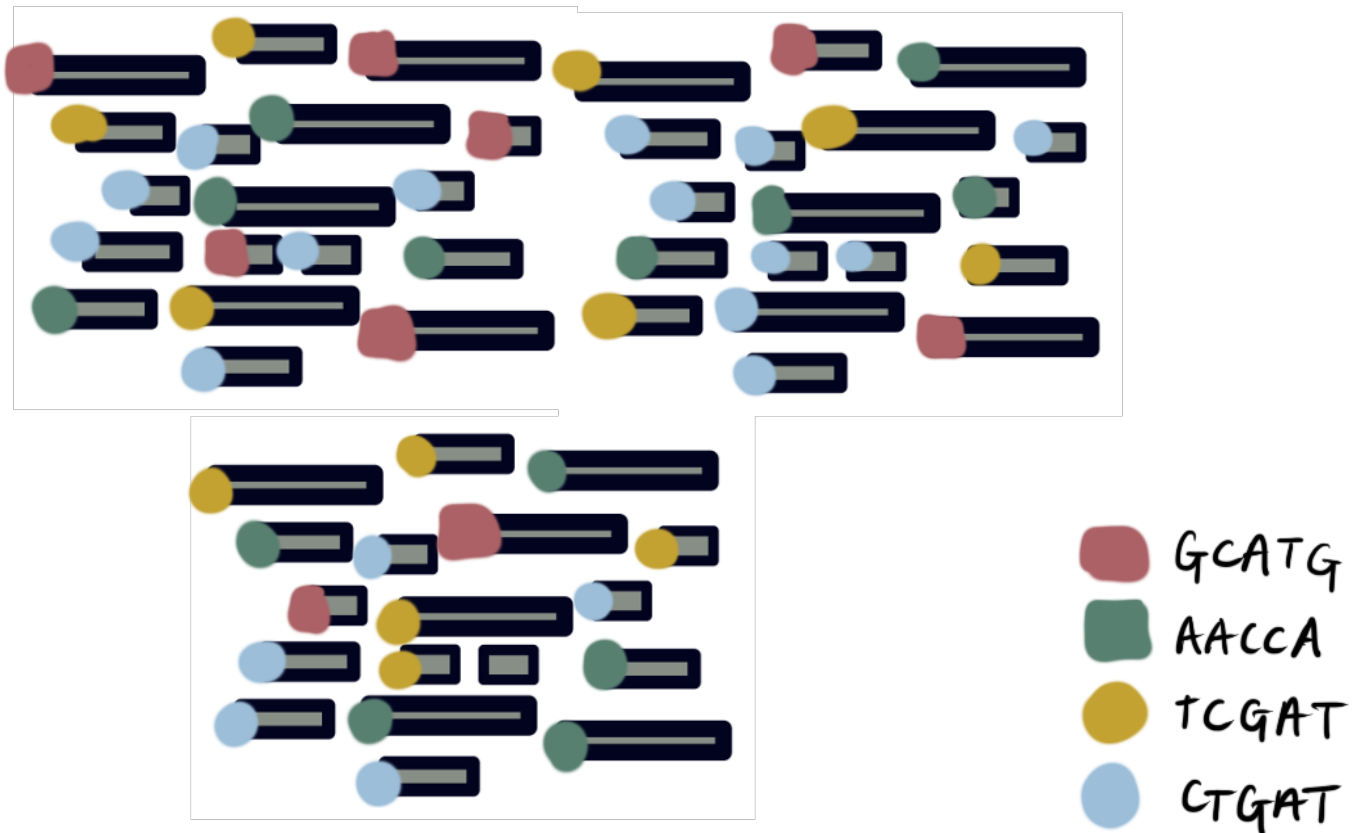
Genomic DNA

Cut up genome into bits (only ~2% genome)

Only want fragments of a certain length

Need unique identifiers for hundreds of samples

Sequencer needs to recognize fragments and assign to individuals
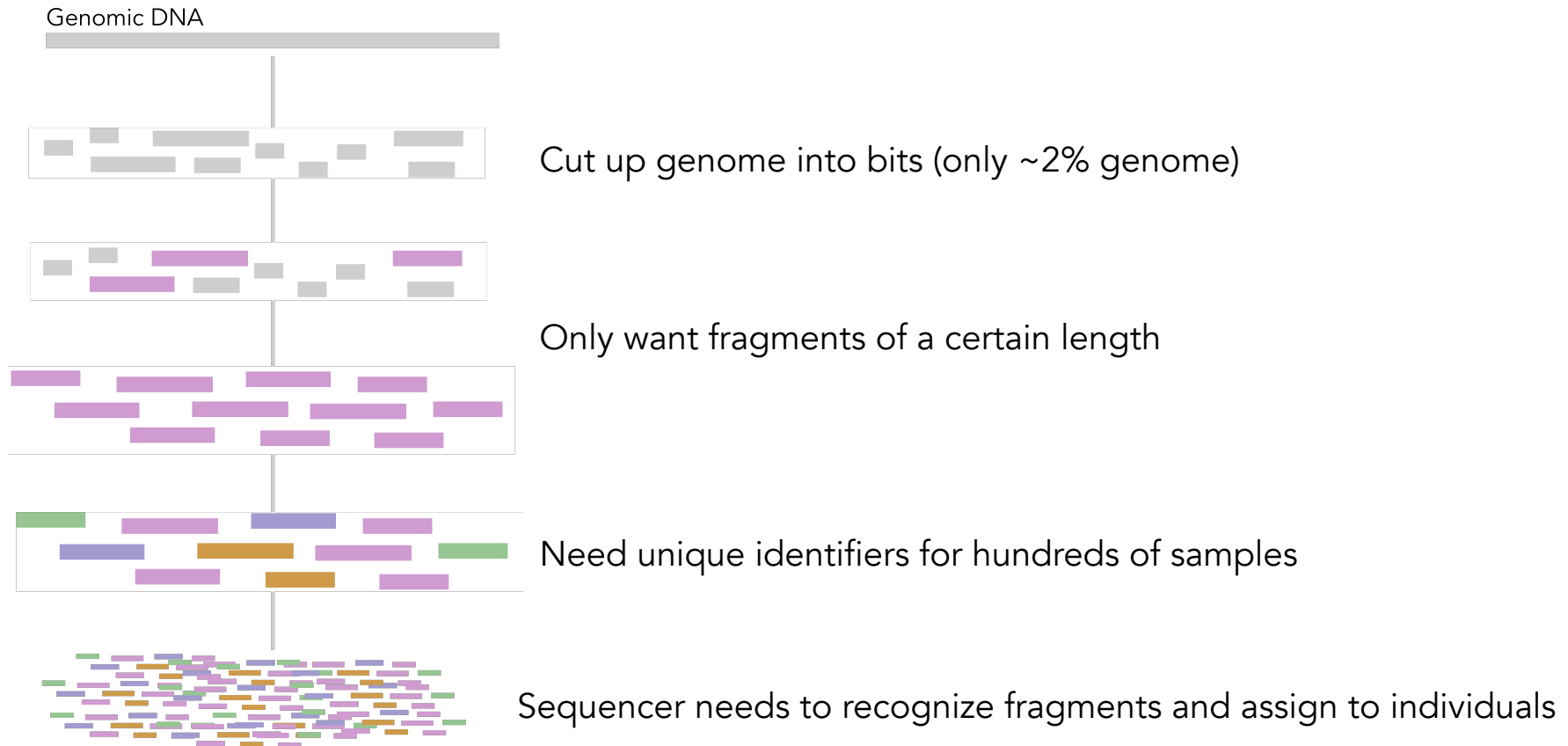
# iPyrad: basic workflow

Seven sequential steps



step1 sorting • step2 filtering • step3 clustering • steps 4 & 5 consensus • step 6 clustering • step 7 formatting

Raw data → Assembled data

After sequencing is done, we have a bunch of reads from a bunch of fragments



GCATG
AACCA
TCGAT
CTGAT

Step 1: sorting and demultiplexing

After sequencing is done, we have a bunch of reads from a bunch of fragments



individual 1

Individual 2

ind 4
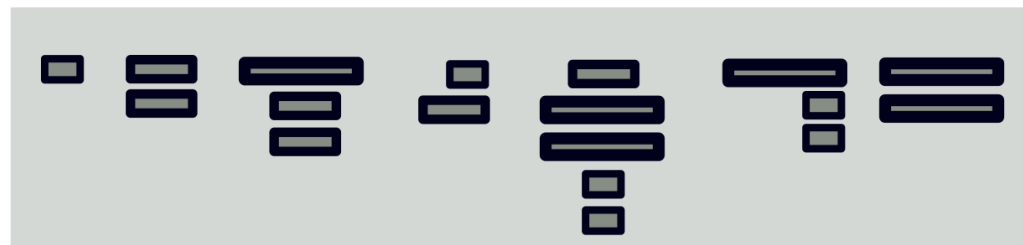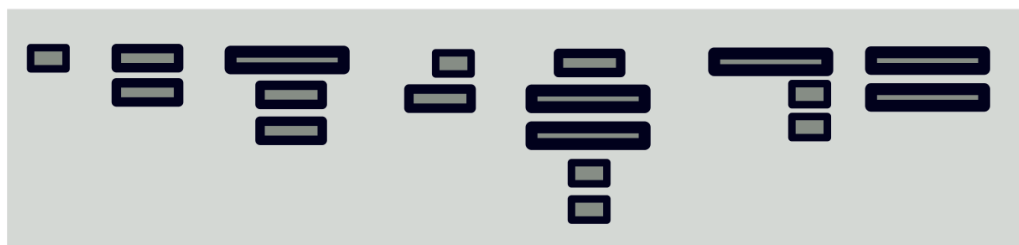
ind. 3

GCATG
AACCA
TCGAT
CTGAT
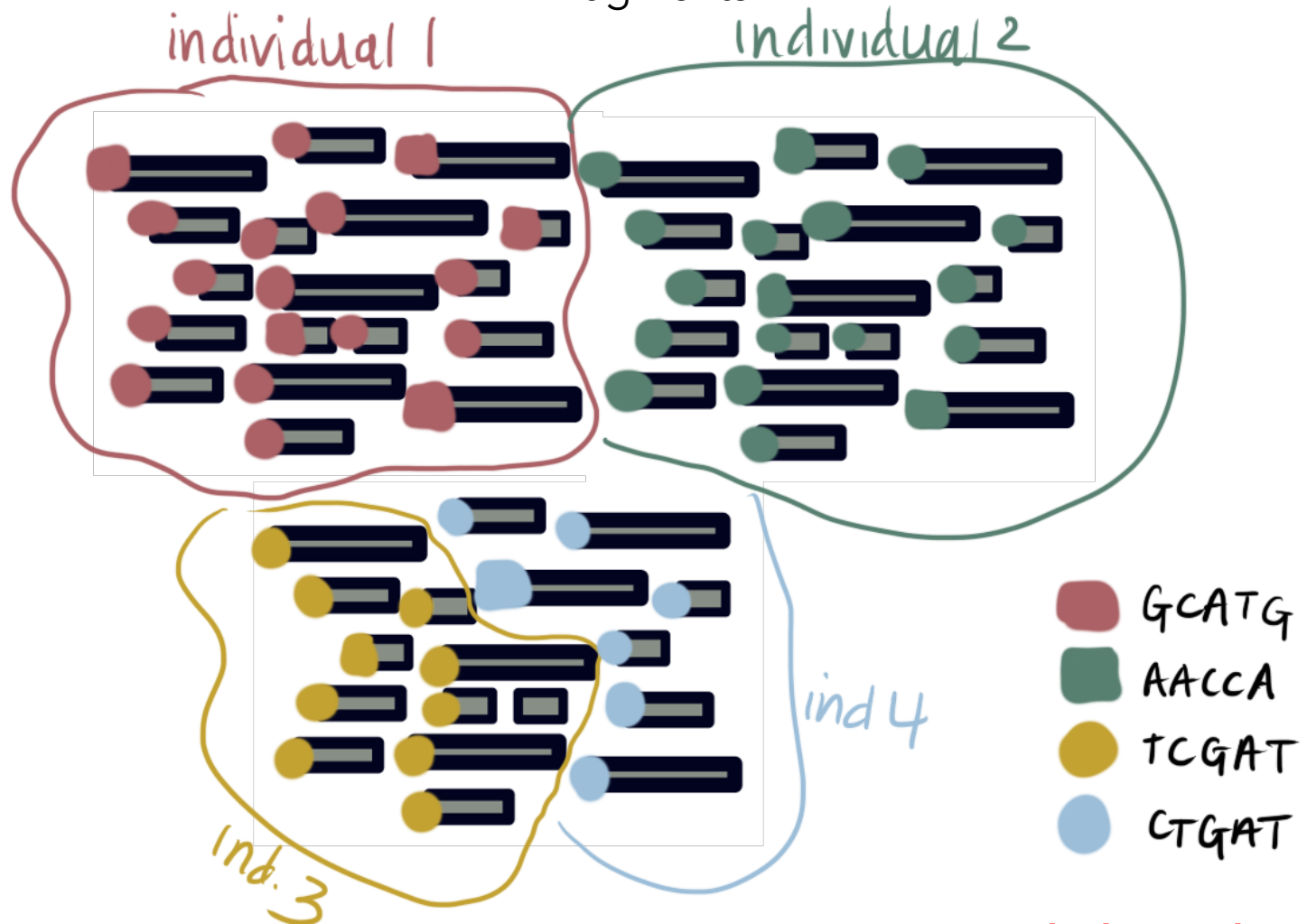
Step 1: sorting and demultiplexing
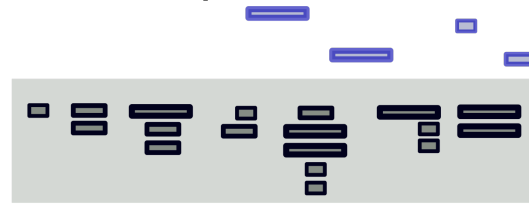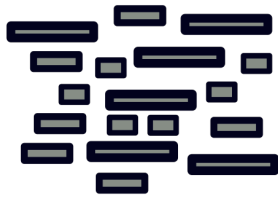
# Step 1 (sorting)

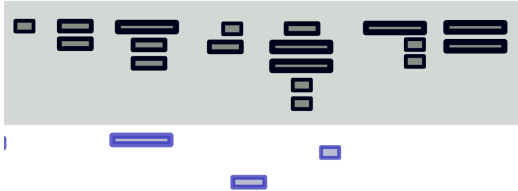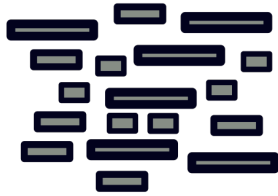species 1

species 2

species 1

species 2

# Step 1 (sorting)

# Step 2 (filtering)

species 1

species 2

Individual 1 reads

95%

clust.
th.

cluster 1

# Step 3: clustering *within* inds (with clustering threshold)
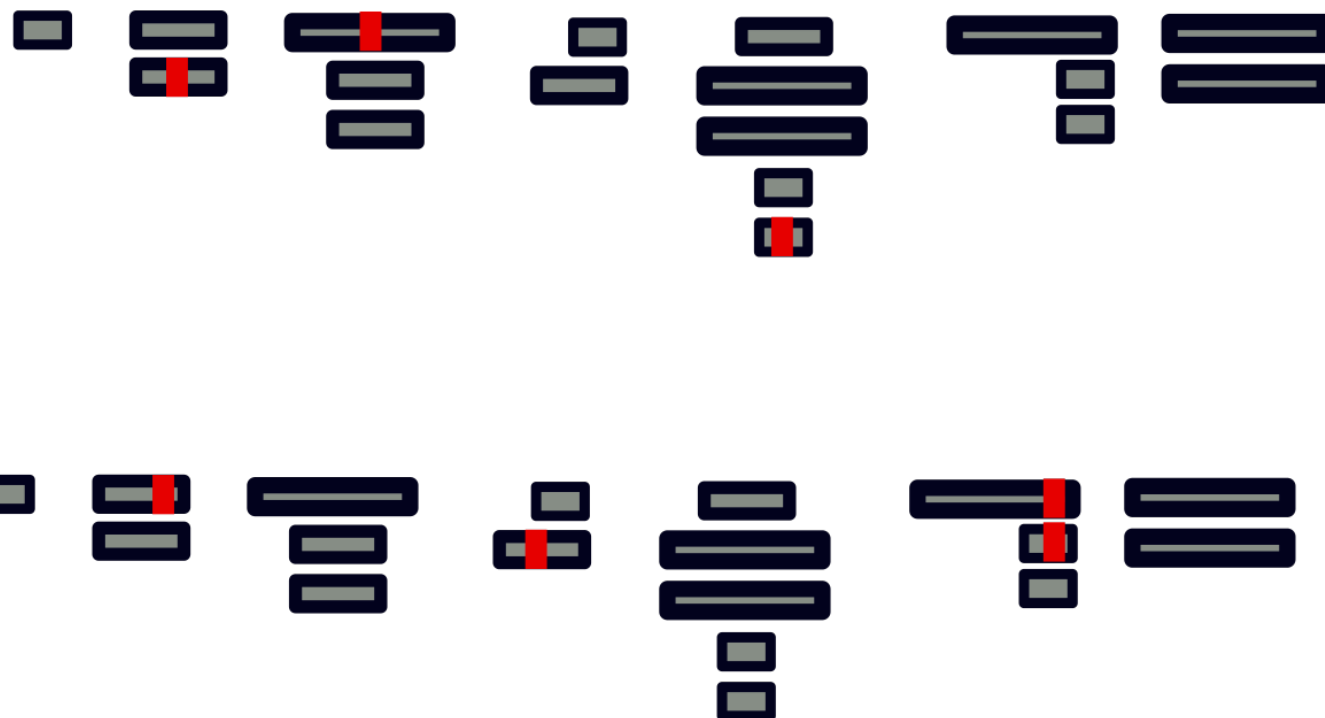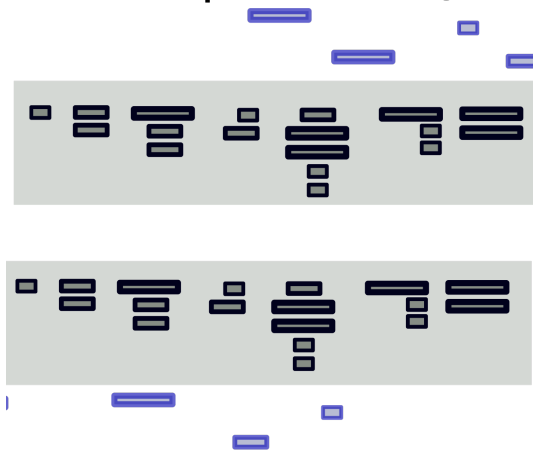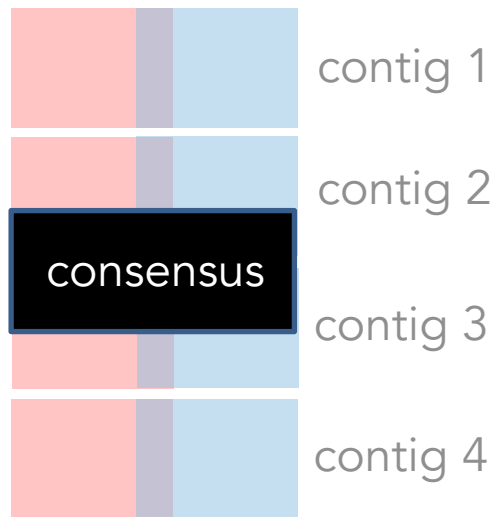
Step 1 (sorting)

Step 2 (filtering)

species 1

species 2

species 1

species 2

Step 3: clustering *within* inds (with clustering threshold)

Step 1 (sorting) — Step 2 (filtering) — Step 3 (clustering)

species 1

species 2

contig 1

contig 2

consensus

contig 3

contig 4
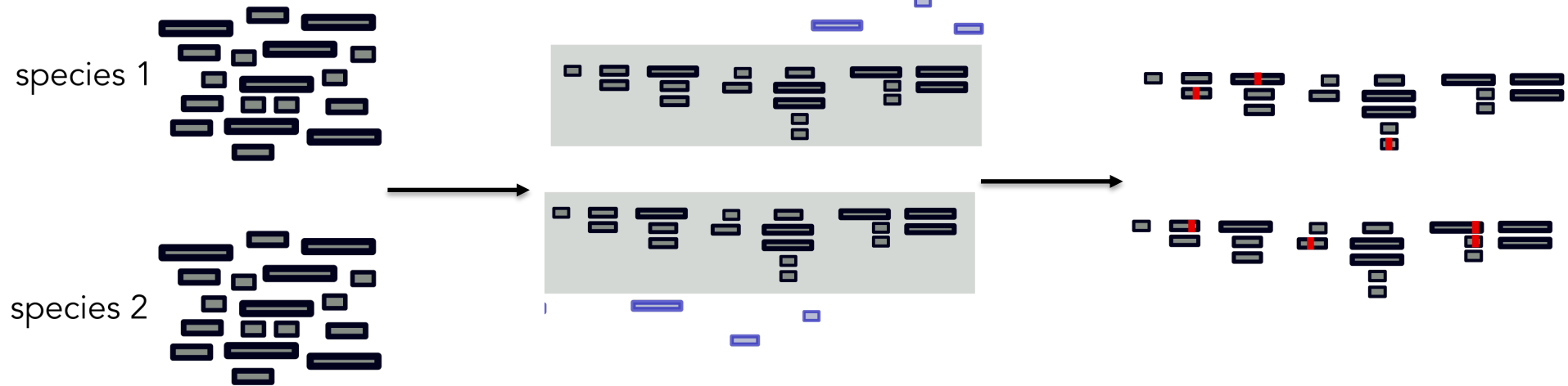
Step 4&5: obtaining consensus reads (heterozygotes and error)

Step 1 (sorting)  Step 2 (filtering)  Step 3 (clustering)  Steps 4&5 (consensus)

species 1

species 2

contig 1
contig 2
consensus
contig 3
contig 4

Individual 1: cluster 1

Individual 2: cluster 1

Cluster 1

85%

ind 1

ind 2

Step 6: Clustering among individuals

Step 1 (sorting)   Step 2 (filtering)   Step 3 (clustering)   Steps 4&5 (consensus)

species 1

species 2

contig 1
contig 2
consensus
contig 3
contig 4

species 1

species 2
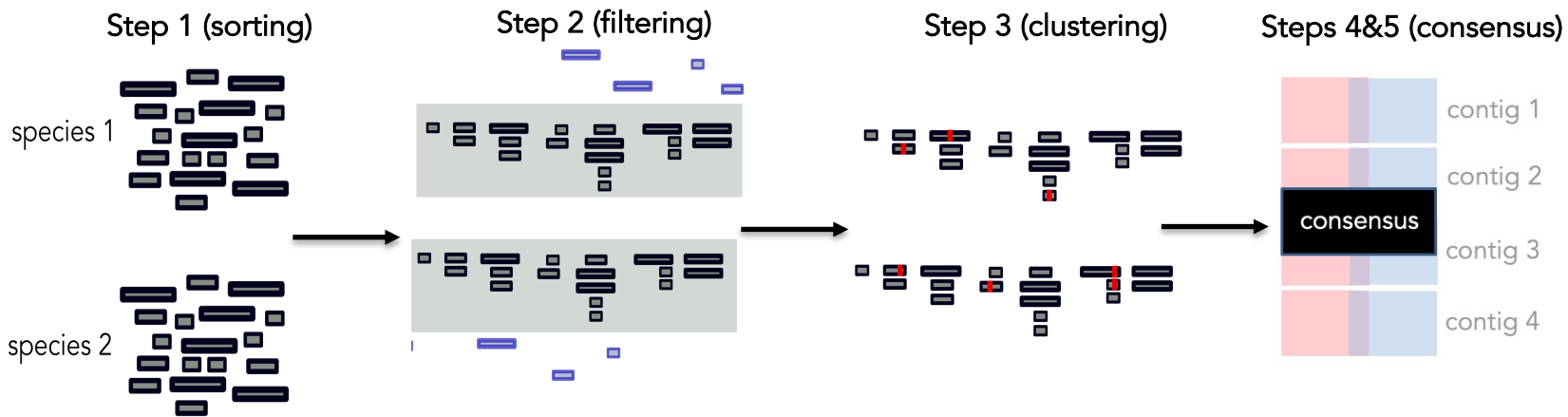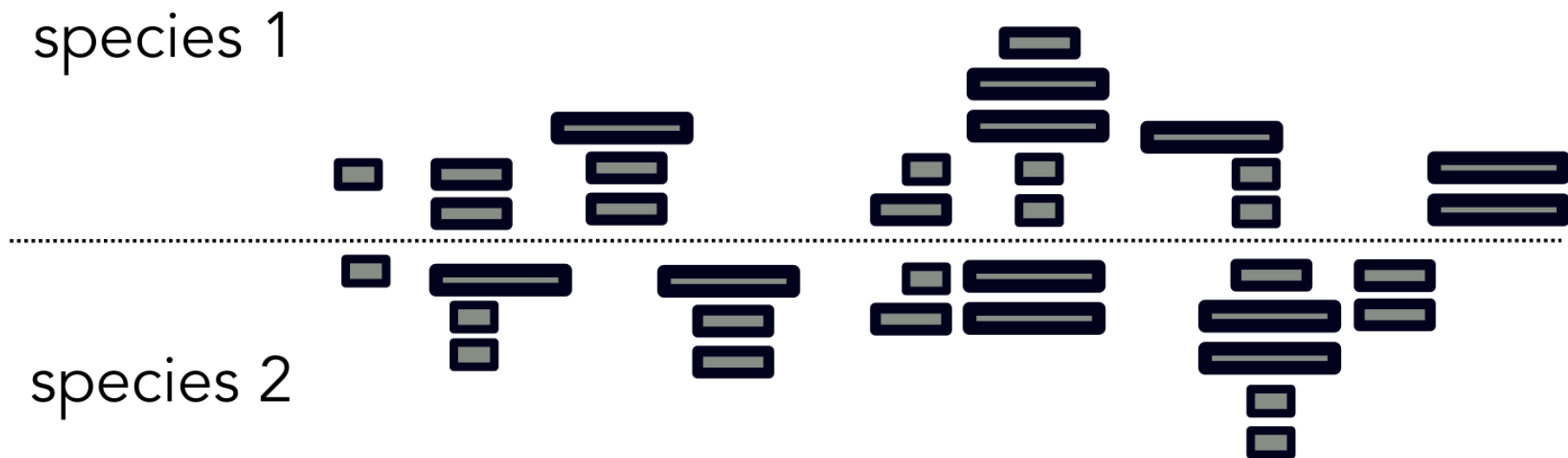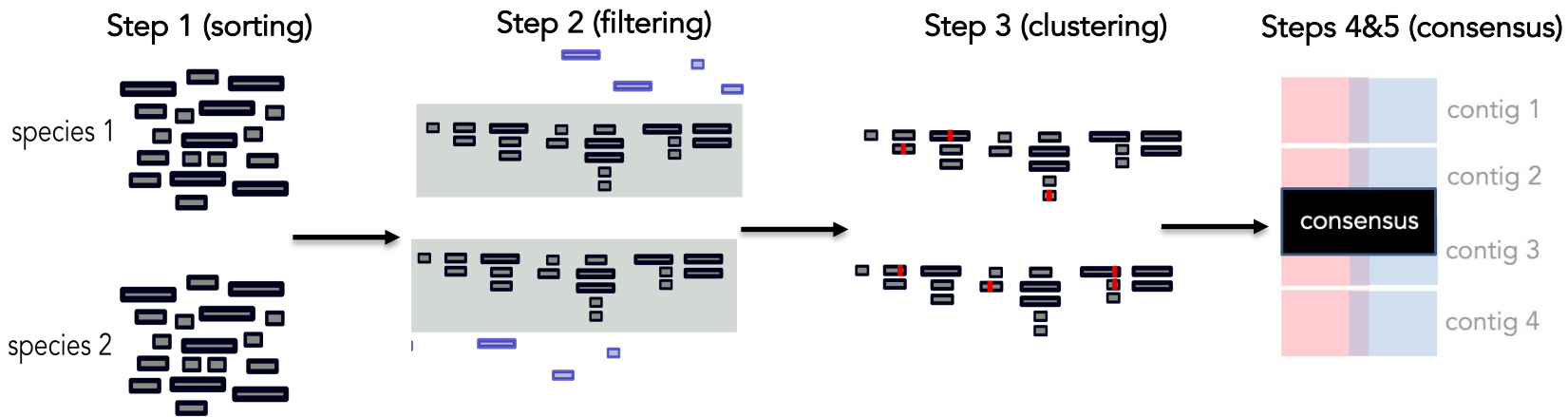
Step 6: Clustering among individuals

# iPyrad relies on a number of programs

- iPyrad (along with other pipelines for NGS data) is just a string of executable programs that have different functions

- e.g., **vsearch** (de novo clustering), **muscle** (alignment), **samtools** (reference mapping)
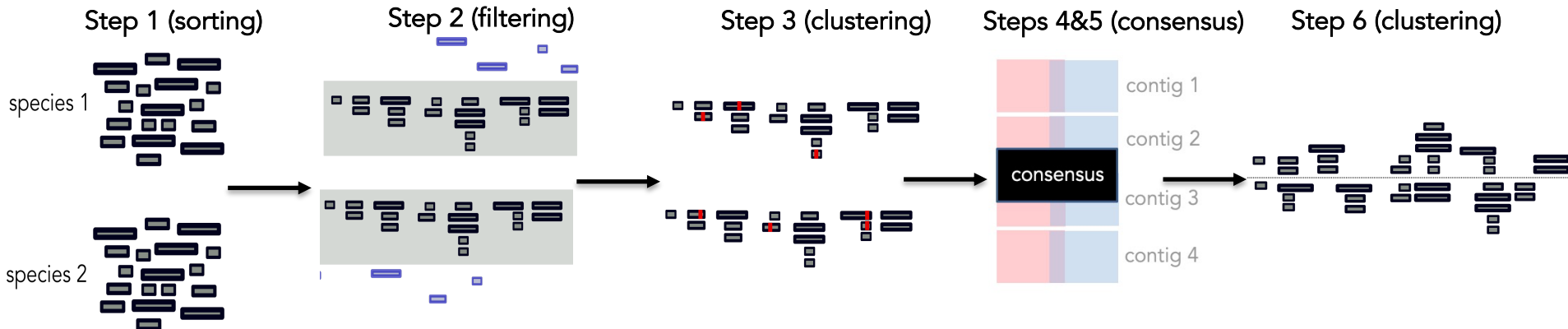


Step 1 (sorting)   Step 2 (filtering)   Step 3 (clustering)   Steps 4&5 (consensus)   Step 6 (clustering)

species 1

species 2

consensus

contig 1
contig 2
contig 3
contig 4

# What will we be analyzing in this course?
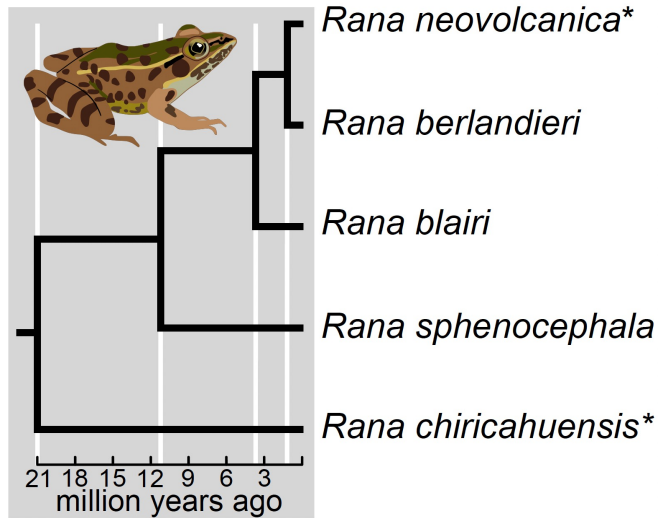


*Rana neovolcanica**

*Rana berlandieri*

*Rana blairi*

*Rana sphenocephala*

*Rana chiricahuensis**

21 18 15 12 9 6 3
million years ago
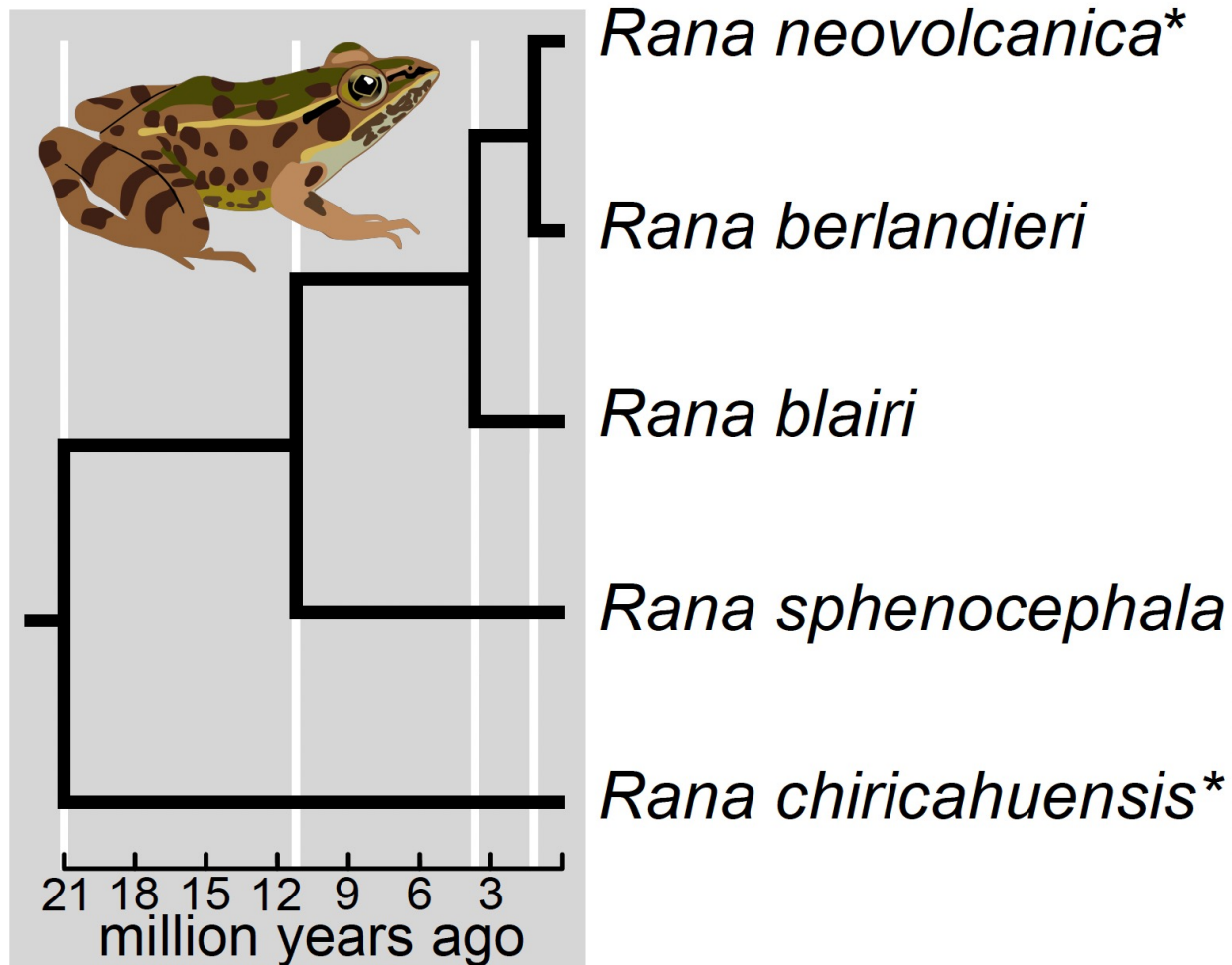
# What will we be analyzing in this course?



| Species | Locality |
|---|---|
| *Rana blairi* | South Dakota, USA |
| *Rana blairi* | South Dakota, USA |
| *Rana neovolcanica* | Jalisco, Mexico |
| *Rana neovolcanica* | Jalisco, Mexico |
| *Rana berlandieri* | Tamaulipas, Mexico |
| *Rana berlandieri* | Tamaulipas, Mexico |
| *Rana chiricahuensis* | Arizona, USA |
| *Rana chiricahuensis* | Arizona, USA |
| *Rana sphenocephala* | Texas, USA |
| *Rana sphenocephala* | Texas, USA |