

Marker Development for Population Genetics

**James Pease
14 July 2014**

What is your research question?

Example:

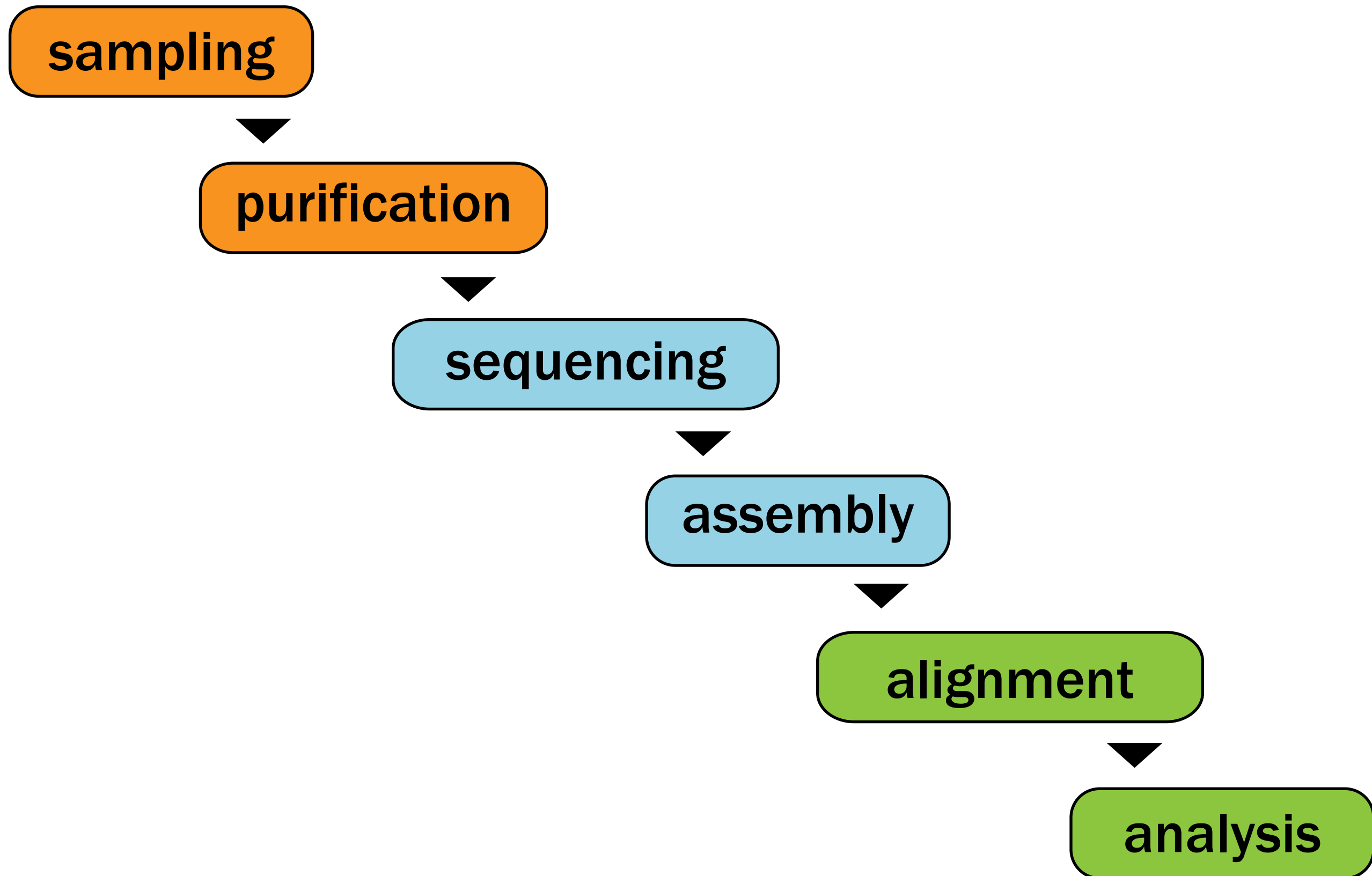
What is the geographic distribution of populations of a disruptive invasive species?

What is your research question?

Example:

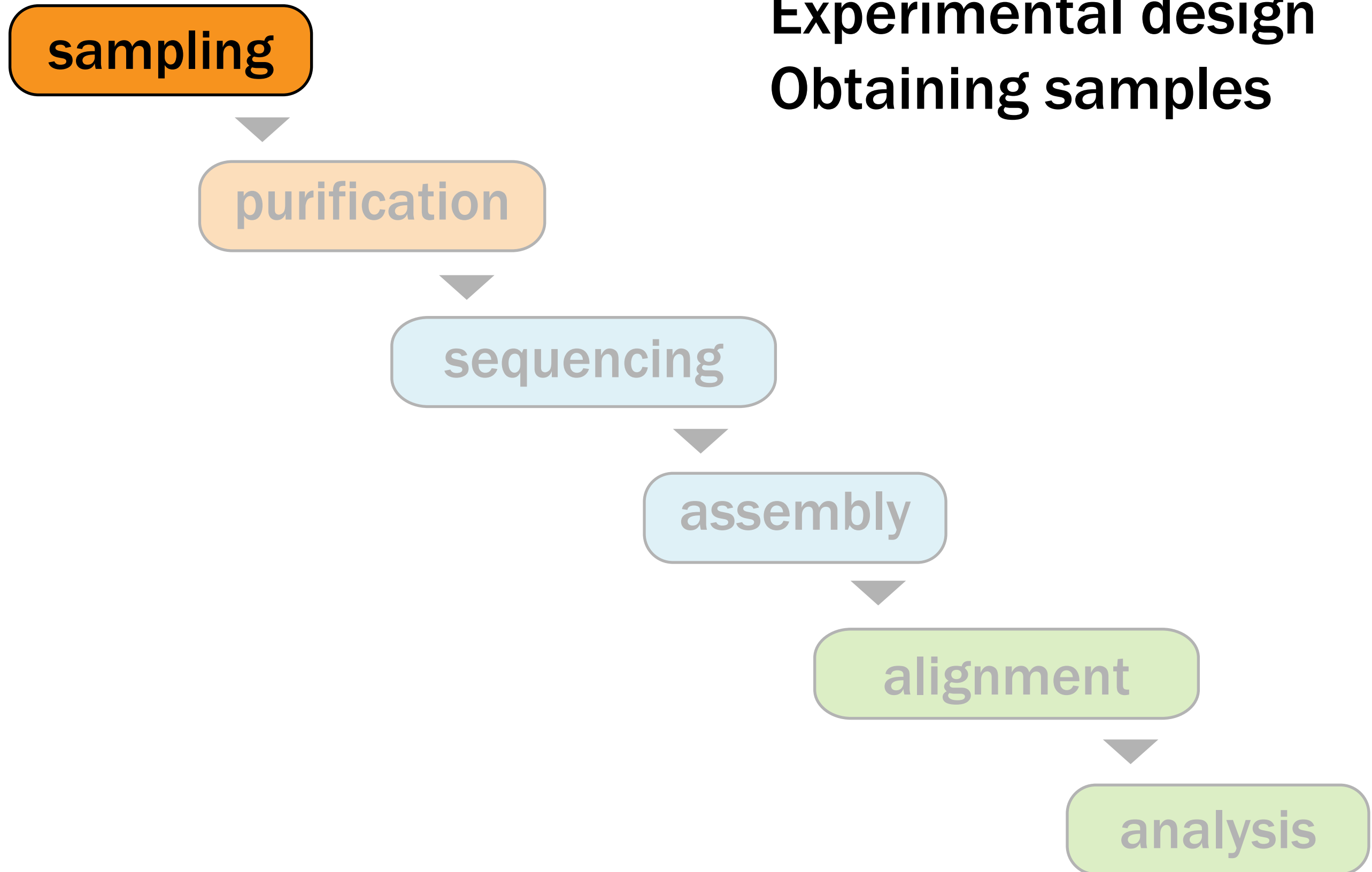
What is the geographic distribution of populations of a disruptive invasive species?





Experimental design

Obtaining samples



sampling

purification

sequencing

assembly

alignment

analysis

**DNA/RNA Extraction
DNA/RNA Purification
Sequencing Prep.**



sampling



purification



sequencing



assembly

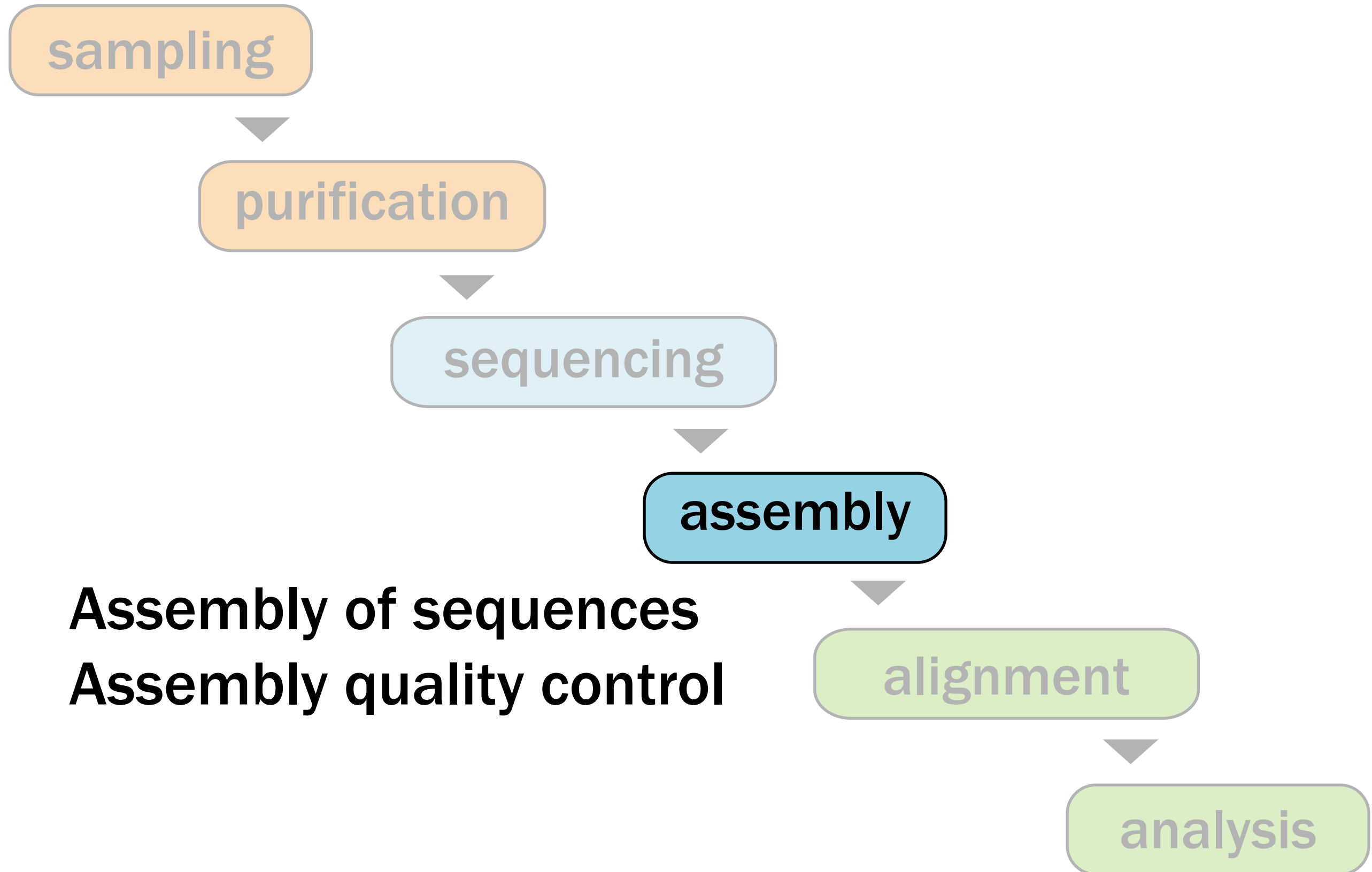


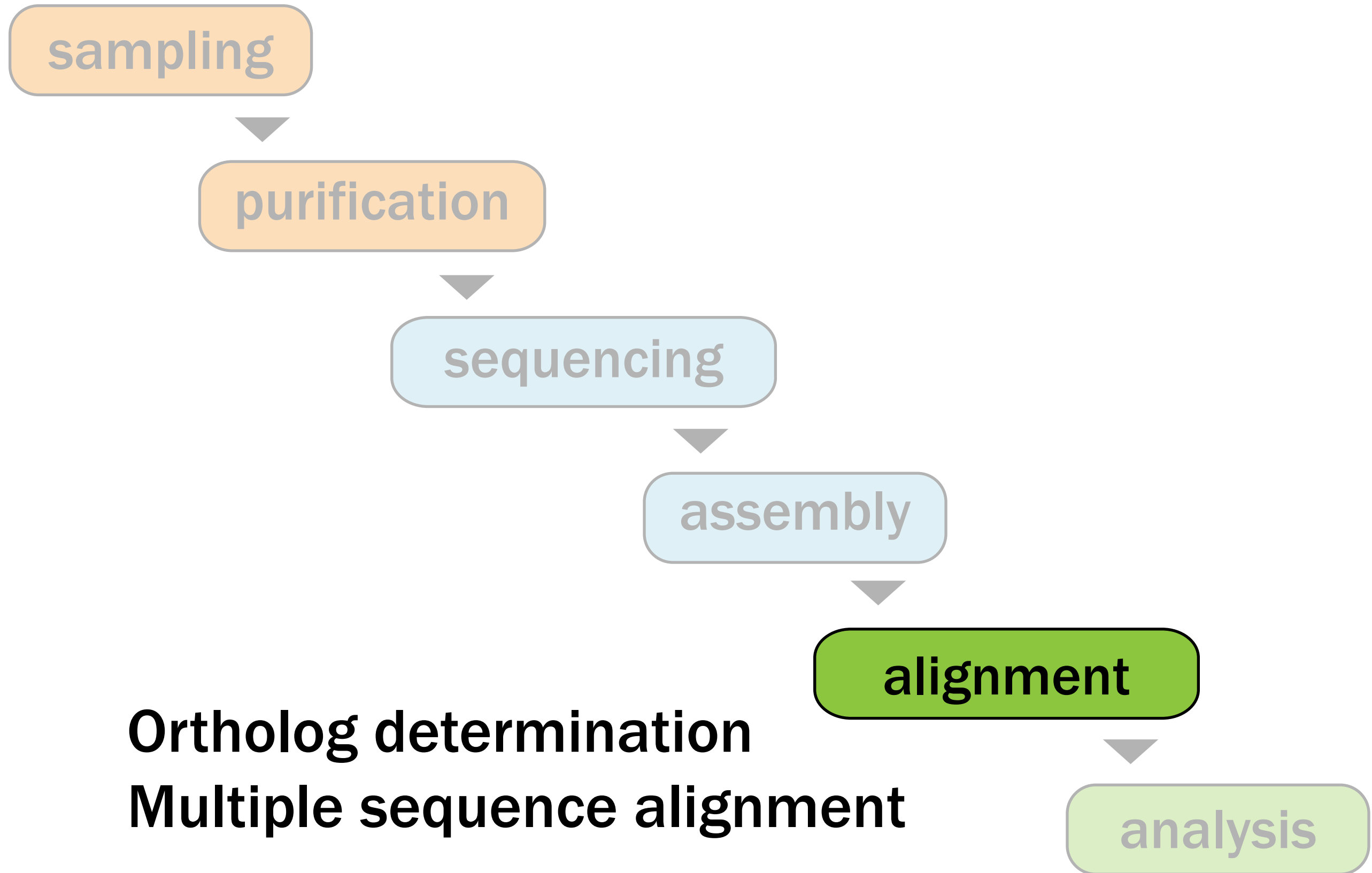
alignment

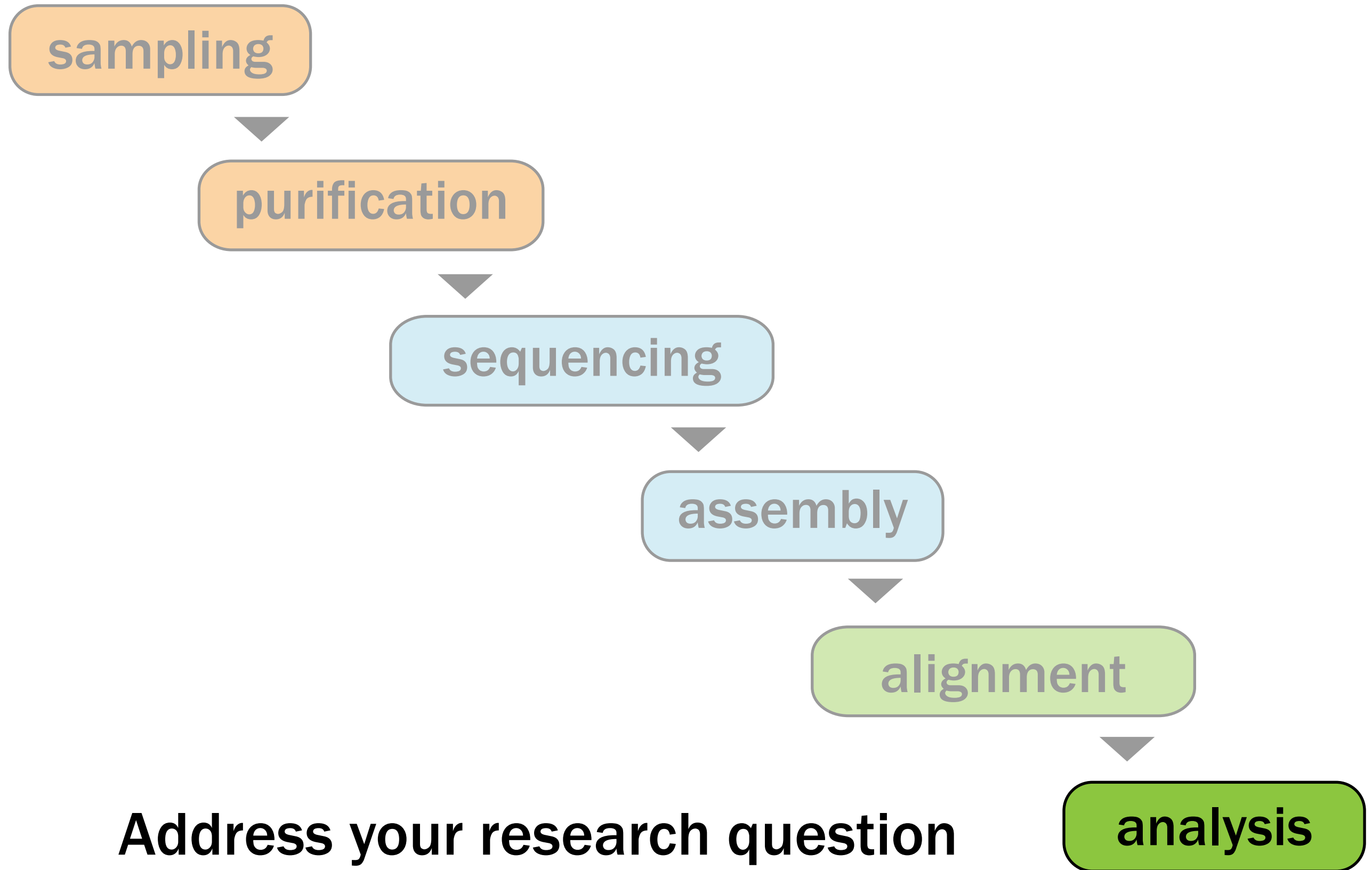


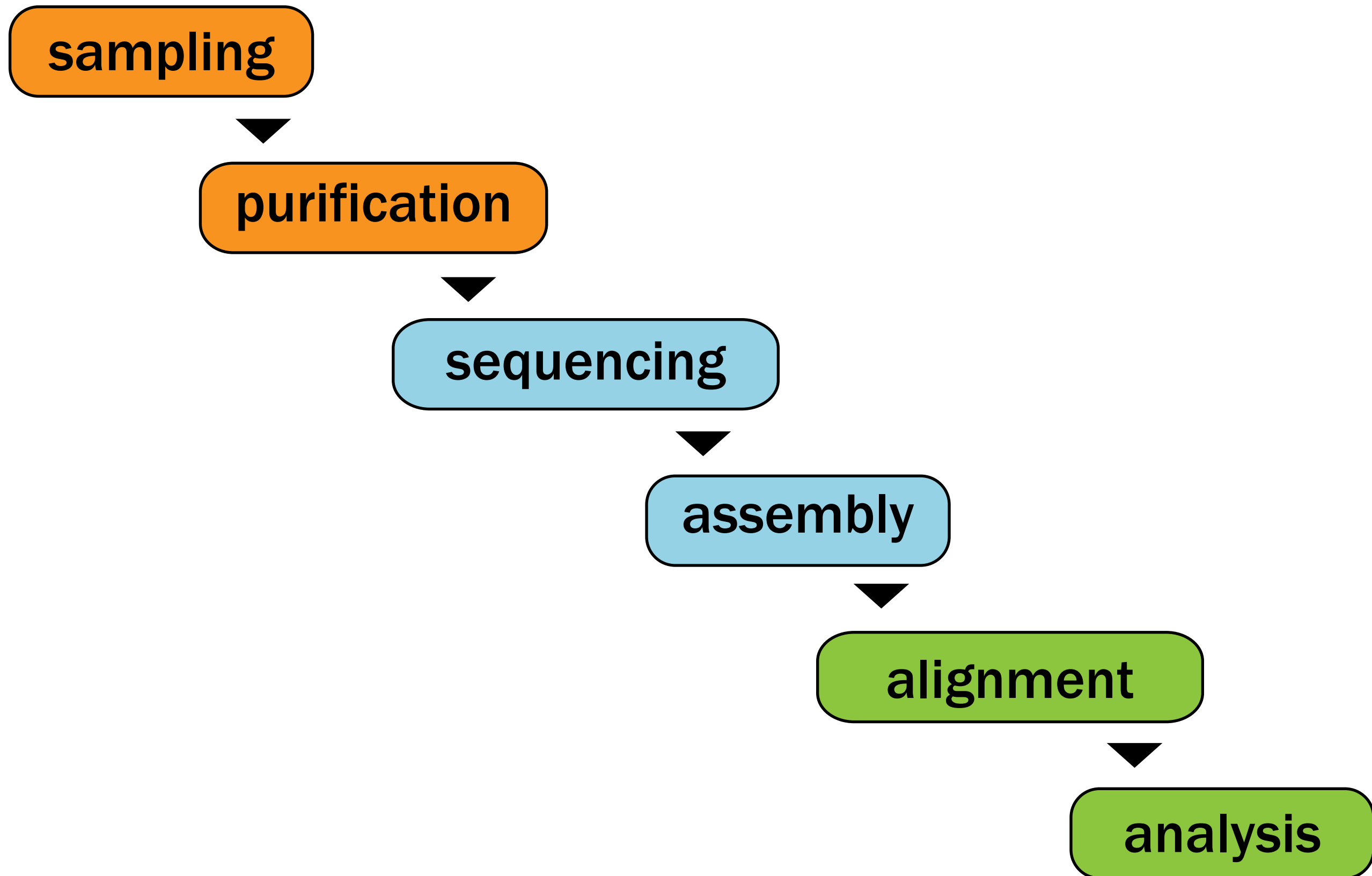
analysis

High-throughput sequencing
Sequencing quality control







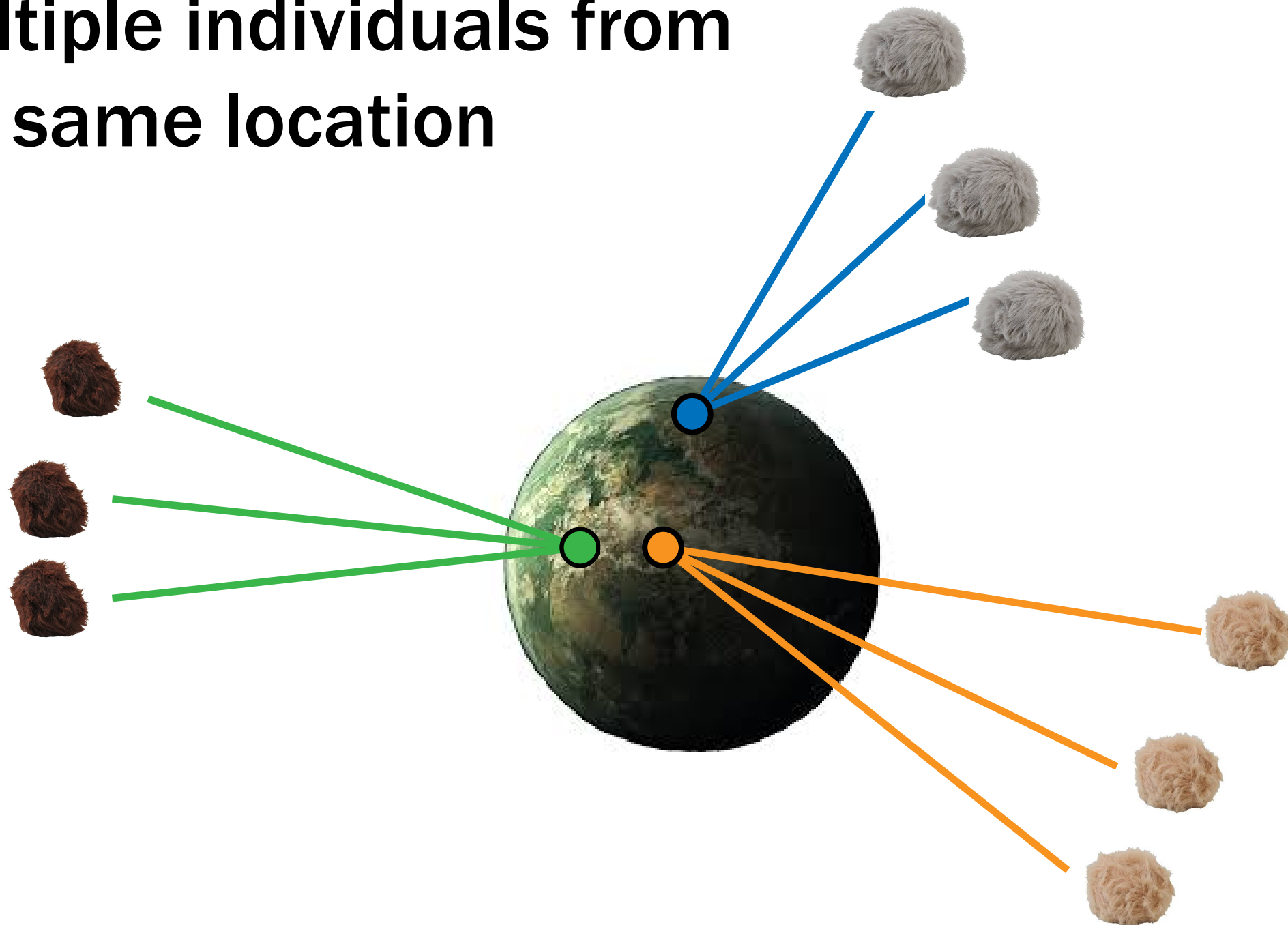


Experimental Design

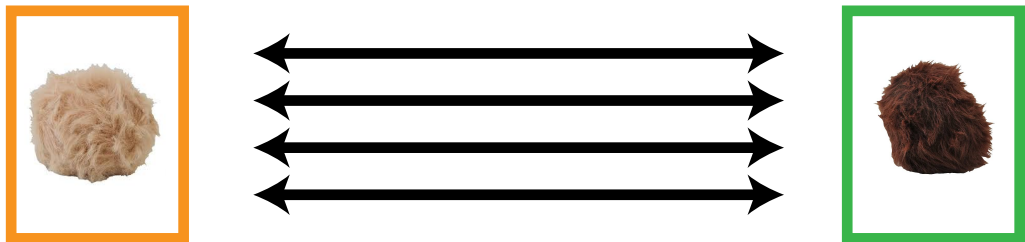
- **Replication**
- **Multiple Tissues**
- **Multiple Developmental Stages**
- **Multiple Treatment Groups**
(environmental, experimental condition)

Replication

**Sample and sequence
multiple individuals from
the same location**

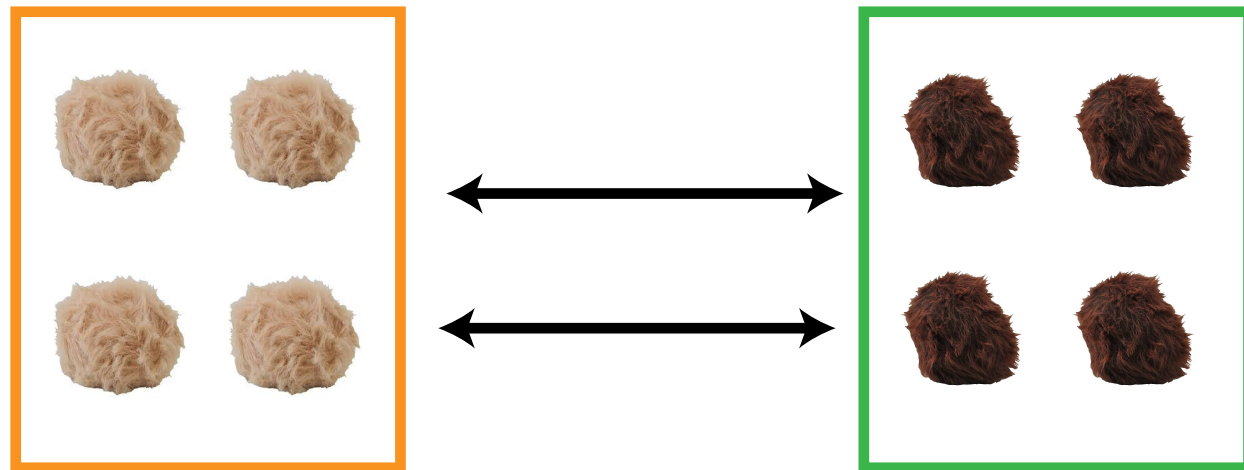


Replication



**There are many SNPs
between individuals**

but...



**fewer consistent SNPs
between populations**

Replication

ACGTATGACGTACACAGTGTA**T**GTAC

ACGTAT**A**ACGTAT**T**ACAGTGTA**C**GTAC

✓

x

x

ACGTAT**G**ACGTACACAGTGTT**T**CGTAC

ACGTAT**G**ACGTAT**T**ACAGTGTT**T**CGTAC

ACGTAT**G**ACGTAT**T**ACAGTGTT**T**CGTAC

ACGTAT**A**ACGTAT**T**ACAGTGTA**A**CGTAC

ACGTAT**A**ACGTAT**T**ACAGTGTA**C**CGTAC

ACGTAT**A**ACGTAT**T**ACAGTGTA**A**CGTAC

Replication

If you have low coverage, you may recover more erroneous SNPs

GATCCAGTACTGACATGCGTA**T**GTAG

GATCCAGTACT**A**ACATGCGTACGTAG

GATCCAGTACTGACATGCGTACGTAG

ACTGACATGCGTACGTAGATGGCCGC

TACGTAC**C**ATGGCCGCAGAGAG**G**ACAG**T**

GATCCAGTACTGACATGCGTACGTAGATGGCCGCAGAGAGGACAGT



Sampling/Purification

Possible sampling issues

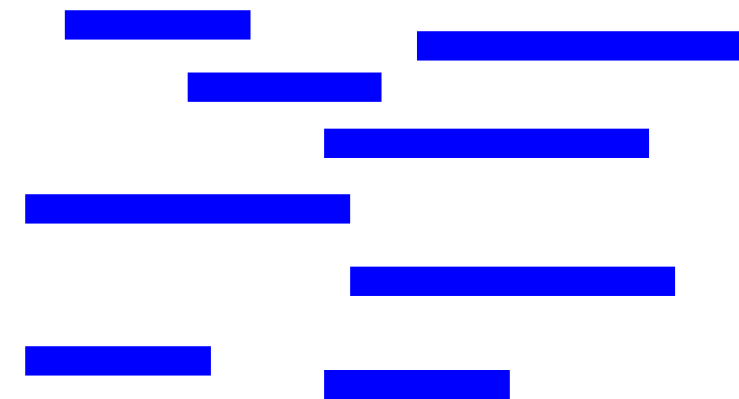
- Contamination between samples
- Contamination from other organisms
- Degradation of DNA/RNA
(especially RNA)

Sampling/Purification

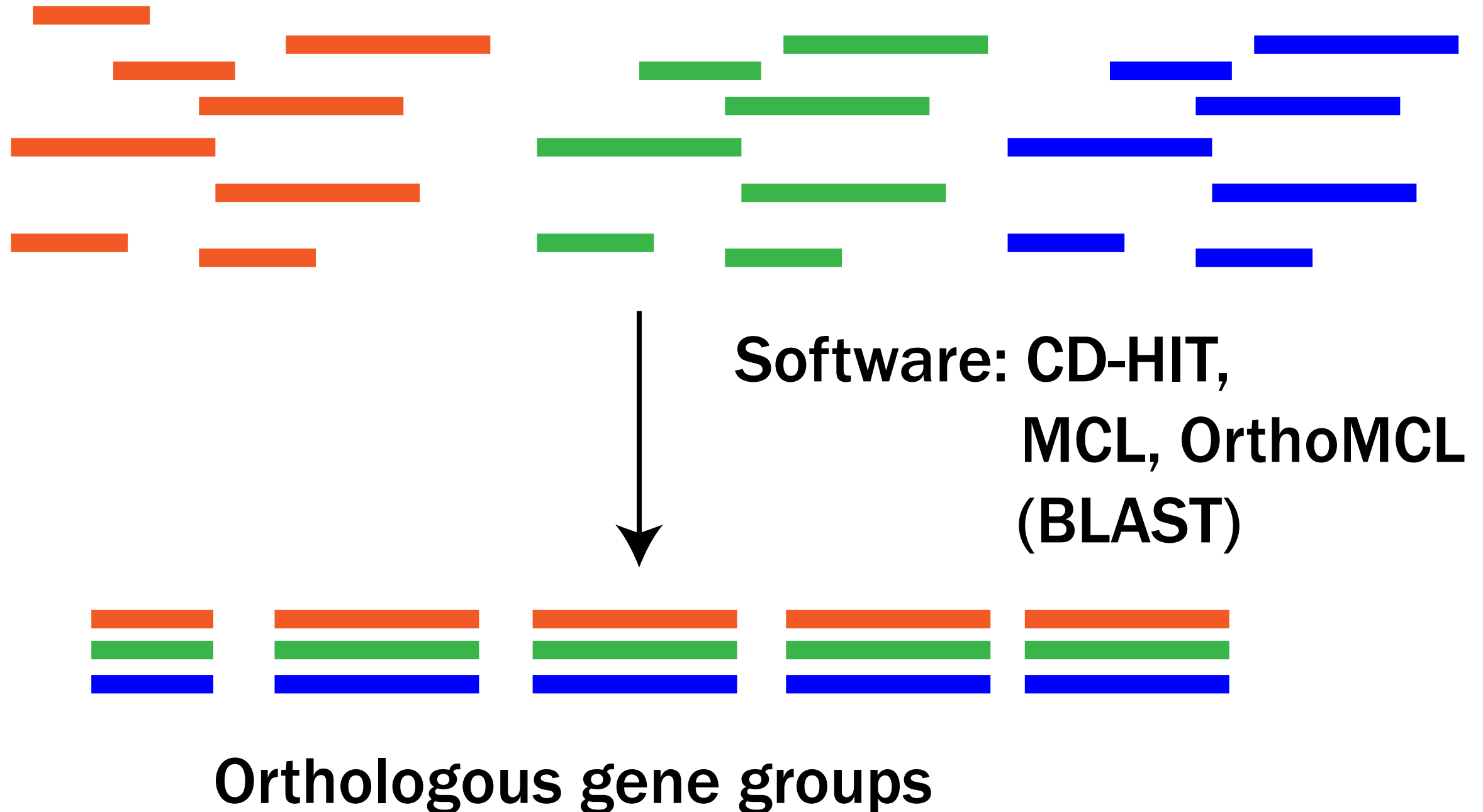
Possible sampling issues

- Contamination between samples
- Contamination from other organisms
- Degradation of DNA/RNA
(especially RNA)

Alignment: Ortholog Determination

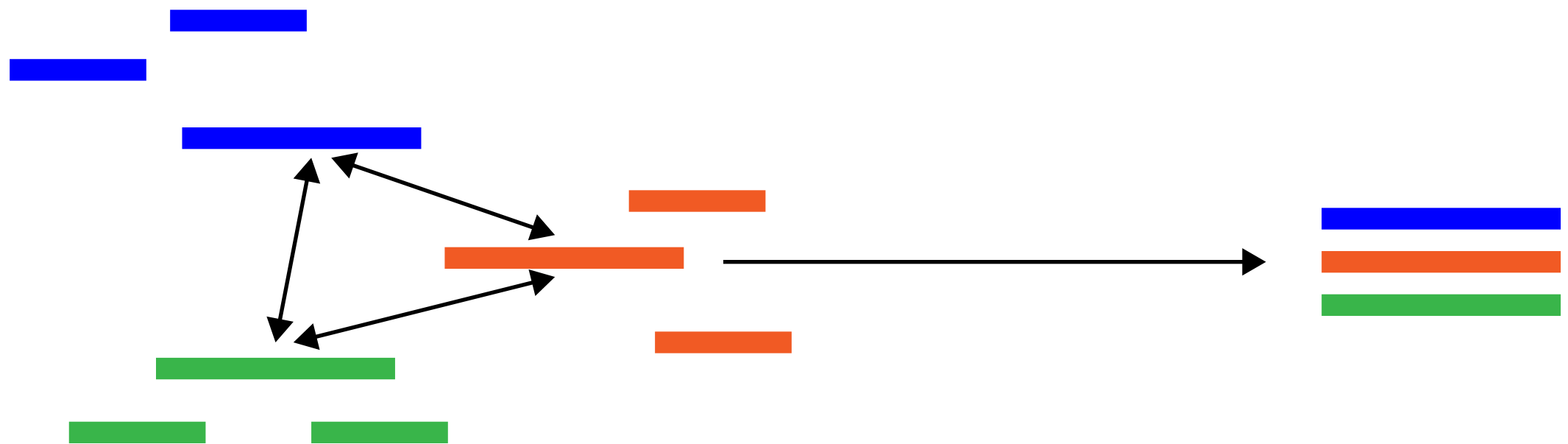


Alignment: Ortholog Determination



Alignment: Ortholog Determination

Groups are determined by reciprocally and transitively strong hits from BLAST



http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi

Alignment: Multiple Sequence Alignment

**TATTGCGACTCGGTATTACAGGCTAGAACACGAGGTAAAA
TATCGCGACTCGGTATTATAGGCTAGAACGCGAGGTAAAA
TATCGCGACTCGGTATTACAGGCTAGAACACGAGGTAAAA
TATCGCGACTCGGTACTACAGACTAGAACACGAGGCAAAA
TATCGCGACTCGGTACTACAGACTAGAACACGAGGCAAAA
TATCGCGACTCGGTACTACAGACTAGAACACGAGGCAAAA**

**Software: MUSCLE, PRANK, CLUSTAL-W,
MAFFT**

Alignment: Multiple Sequence Alignment

<http://www.ebi.ac.uk/goldman-srv/webprank/>

<http://mafft.cbrc.jp/alignment/server/>

Alignment: Multiple Sequence Alignment

SNP "Single Nucleotide Polymorphism"

```
TATTGCGACTCGGTATT
TATCGCGACTCGGTATT
TATCGCGACTCGGTATT
```

```
TATCGCGACTCAGTACT
TATCGCGACTCGGTACT
TATCGCGACTCGGTACT
```

SNPs within
an individual

```
AGGCTAGAACACAAAGGTAAAA
AGGCTAGAACGCAAGGTAAAA
AGGCTAGAACACAAAGGTAAAA
```

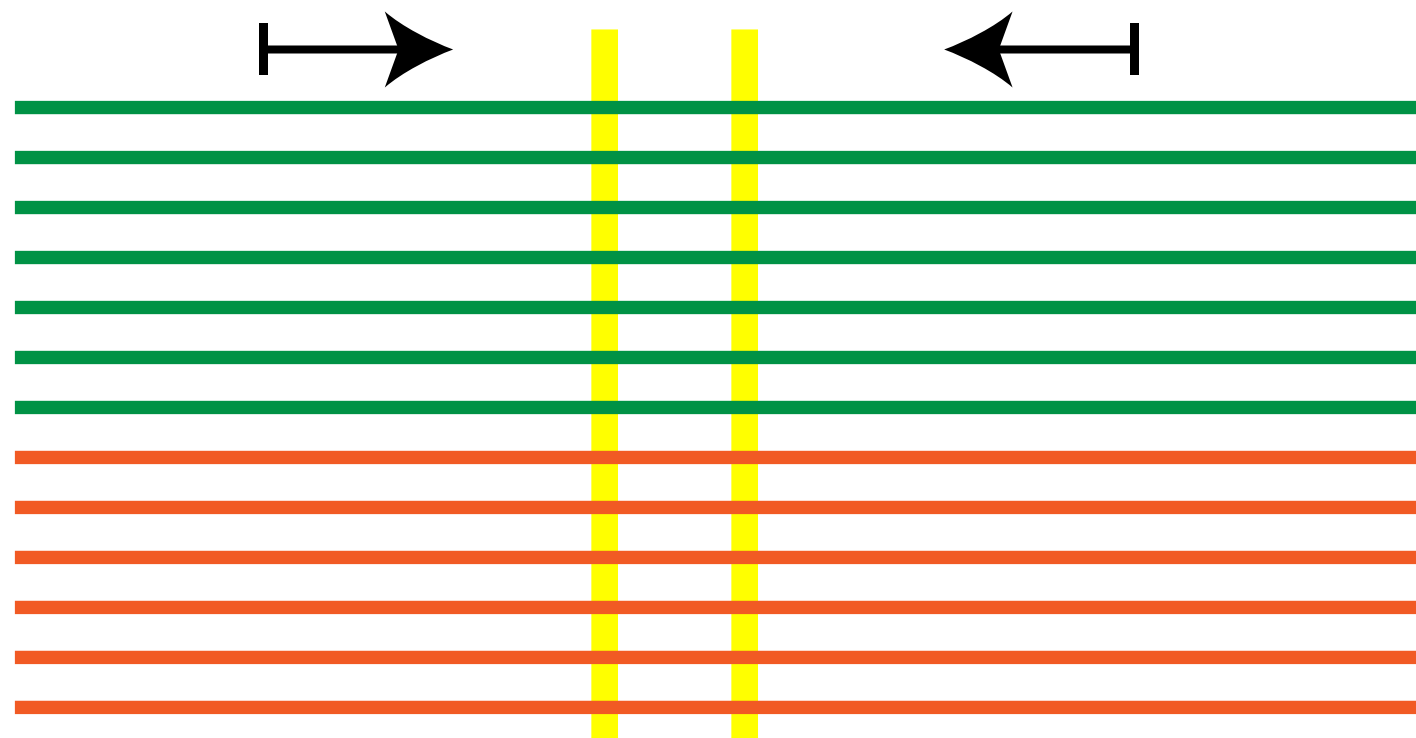
```
AGACTAGAACACGAGGCAAAAA
AGACTAGAACACGAGGCAAAAA
AGACTAGAACACGAGGCAAAAA
```

SNPs within
a population

Analysis: Gene Amplification

PCR (Polymerase Chain Reaction)

rapid amplification of sequences
between two primers



Analysis: Population Resampling

Select several genes with “diagnostic” SNPs for resampling/resequencing.

primer → ← primer

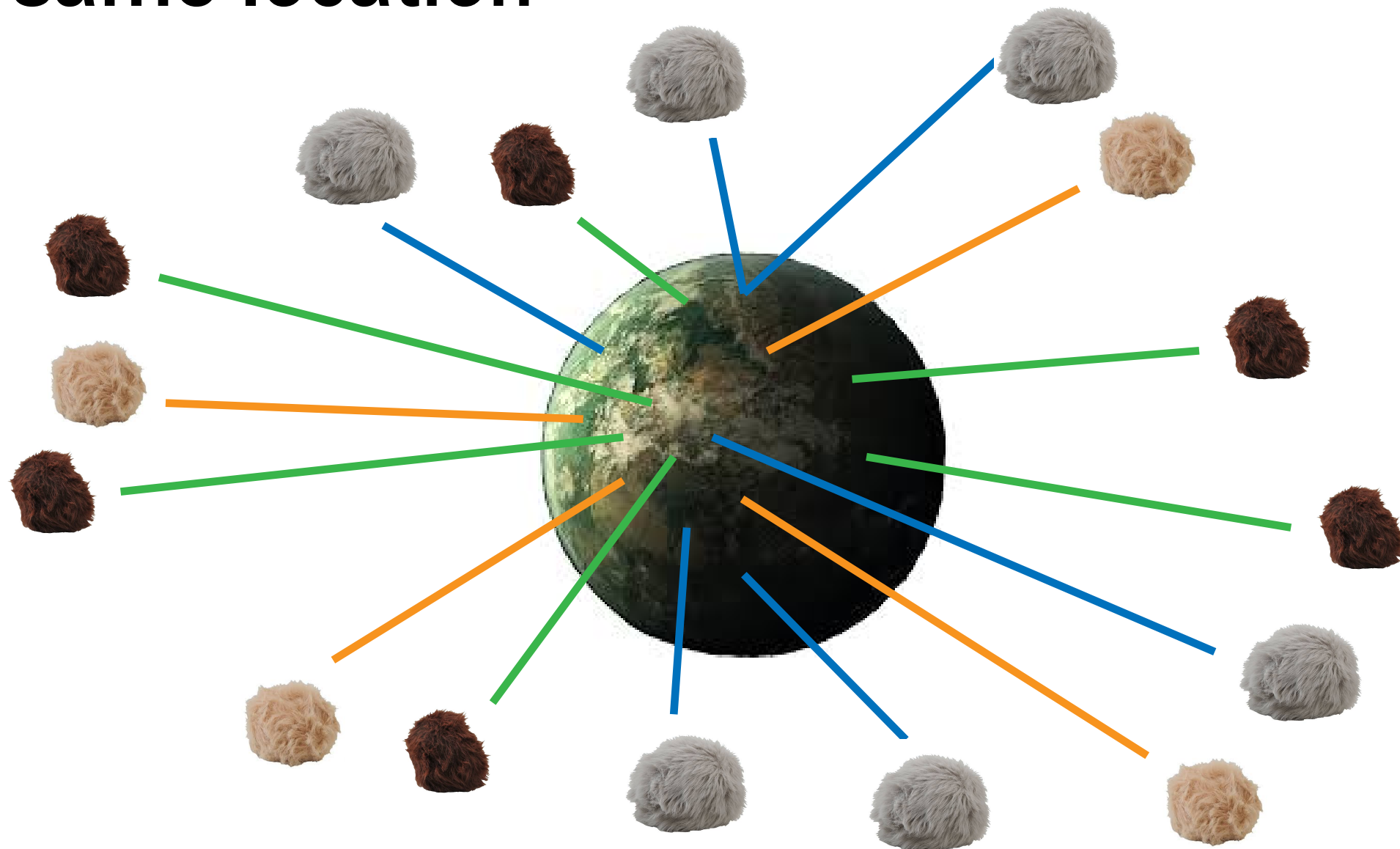
TATCGCGAAGGCTAGAACACAAGGTAAAATATCGCGA
TATCGCGAAGGCTAGAACGCAAGGTAAAATATCGCGA
TATCGCGAAGGCTAGAACACAAGGTAAAATATCGCGA

TATCGCGAAGACTAGAACACGAGGCCAAAATATCGCGA
TATCGCGAAGACTAGAACACGAGGCCAAAATATCGCGA
TATCGCGAAGACTAGAACACGAGGCCAAAATATCGCGA

Diagnostic population SNPs

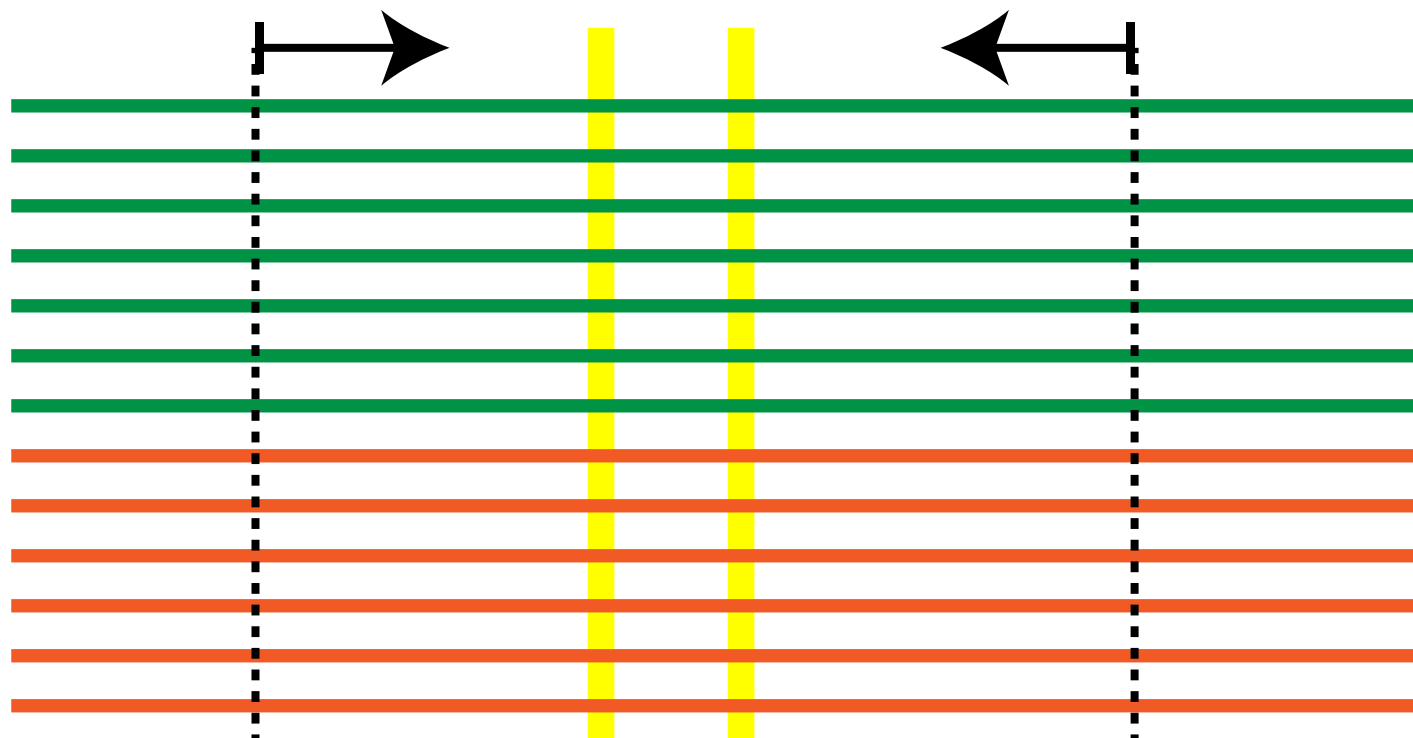
Analysis: Population Resampling

**Sample and sequence
multiple individuals from
the same location**



Analysis: Population Resampling

Now it's only necessary to examine the target gene set with the “diagnostic” SNPs to determine to which population an individual belongs.



Analysis: CAPS

**CAPS (Cleaved Amplified
Polymorphic Sequences)**

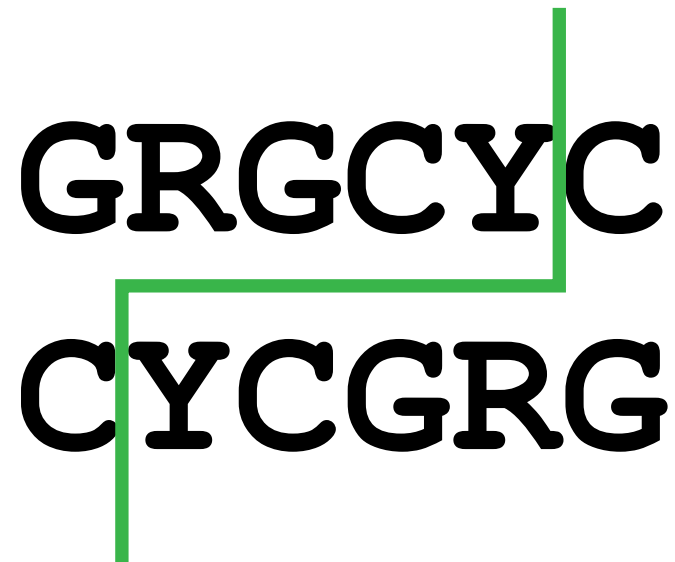
Analysis: CAPS

caagaGAGCTCctctgacat
caagaGAGCTCctctgacat
caagaGAGCTCctctgacat
caagaAAGCTCctctgacat
caagaAAGCTCctctgacat
caagaAAGCTCctctgacat

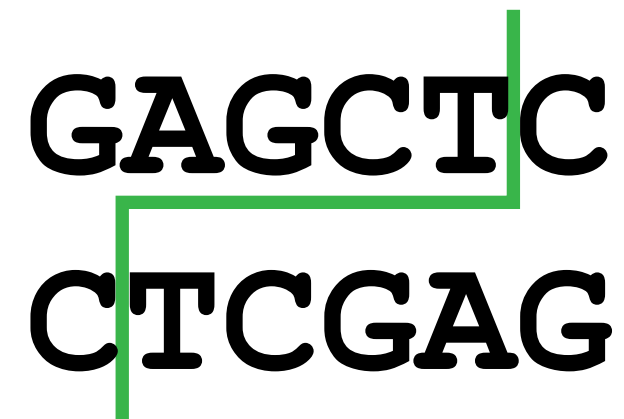
Analysis: CAPS

Restriction enzymes

For Example:
“BamI”



GRGCYC
CYCGRG



GAGCTC
CTCGAG



AAGCTC
TTCGAG

Analysis: CAPS

These sequences
are cut by the enzyme



These are not

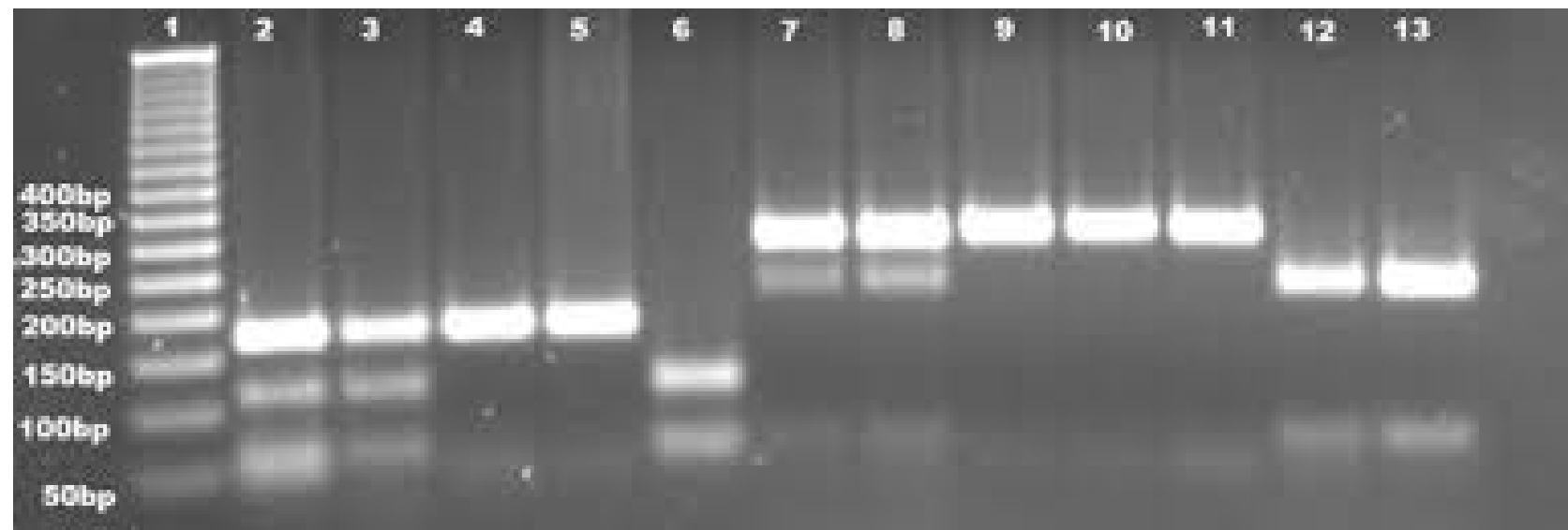


Analysis: CAPS

these are
cut

caagaGAGCTCctctgacat
caagaGAGCTCctctgacat
caagaGAGCTCctctgacat
caagaAAGCTCctctgacat
caagaAAGCTCctctgacat
caagaAAGCTCctctgacat

these are
not cut



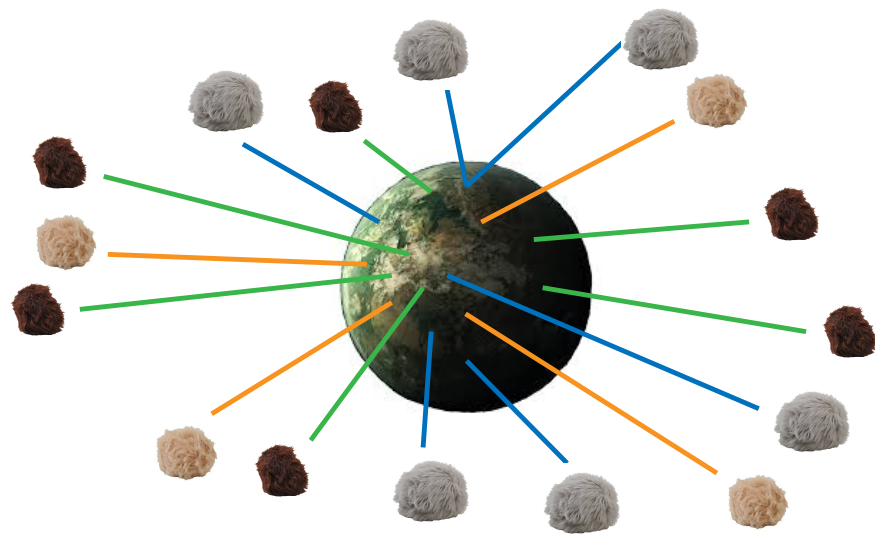
Analysis: CAPS

Online tool for CAPS enzyme determination

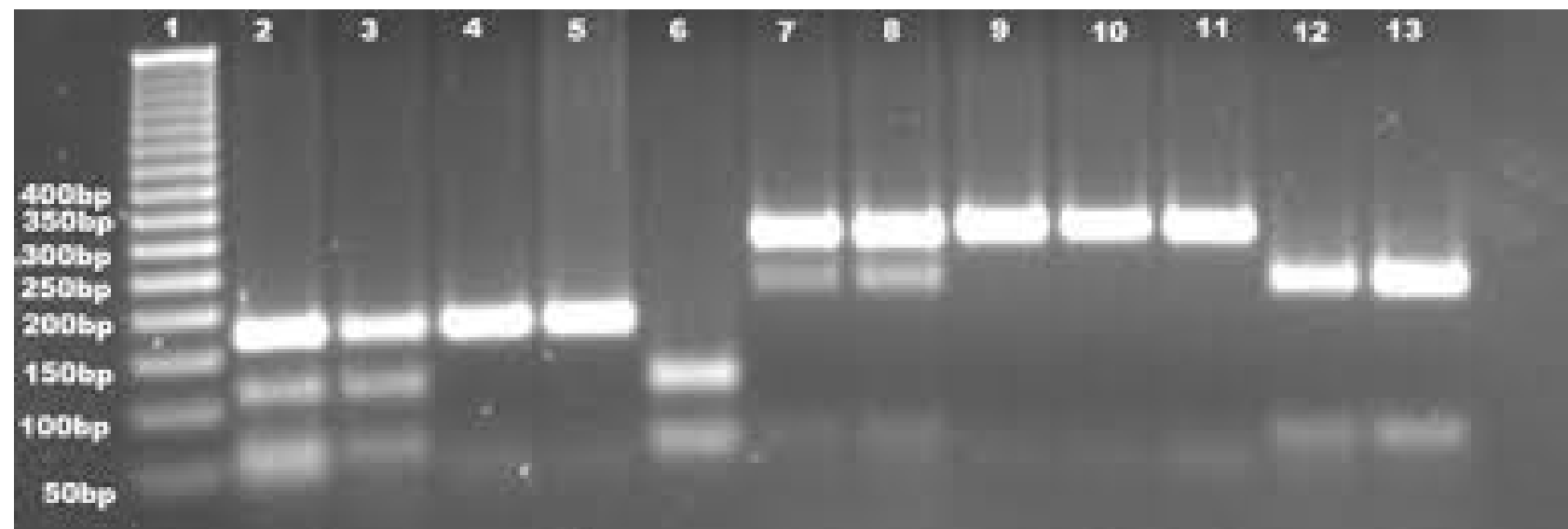
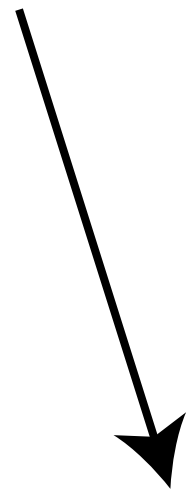
http://solgenomics.net/tools/caps_designer/caps_input.pl

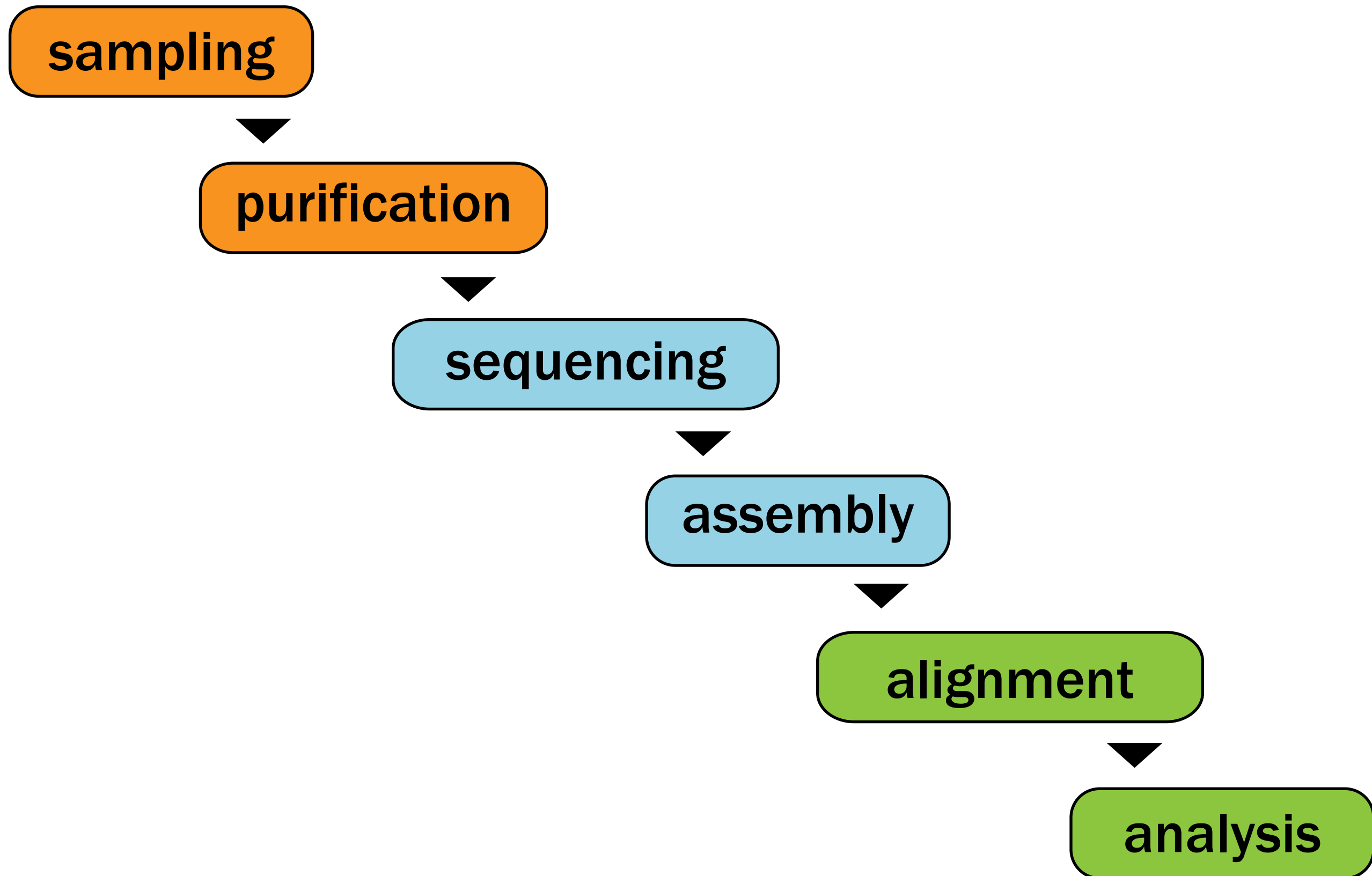
Alineamiento → Enzima de restricción

Analysis: CAPS

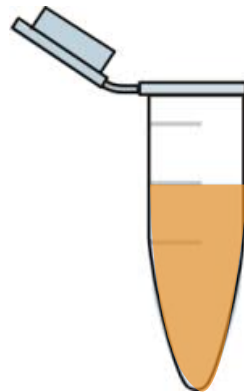


The population identity of each sample is determined by CAPS on diagnostic SNPs





Pooled Sequencing



**Why should I know all of this
if I'm a bioinformatician?**