Target identification and primer design

Boas Pucker

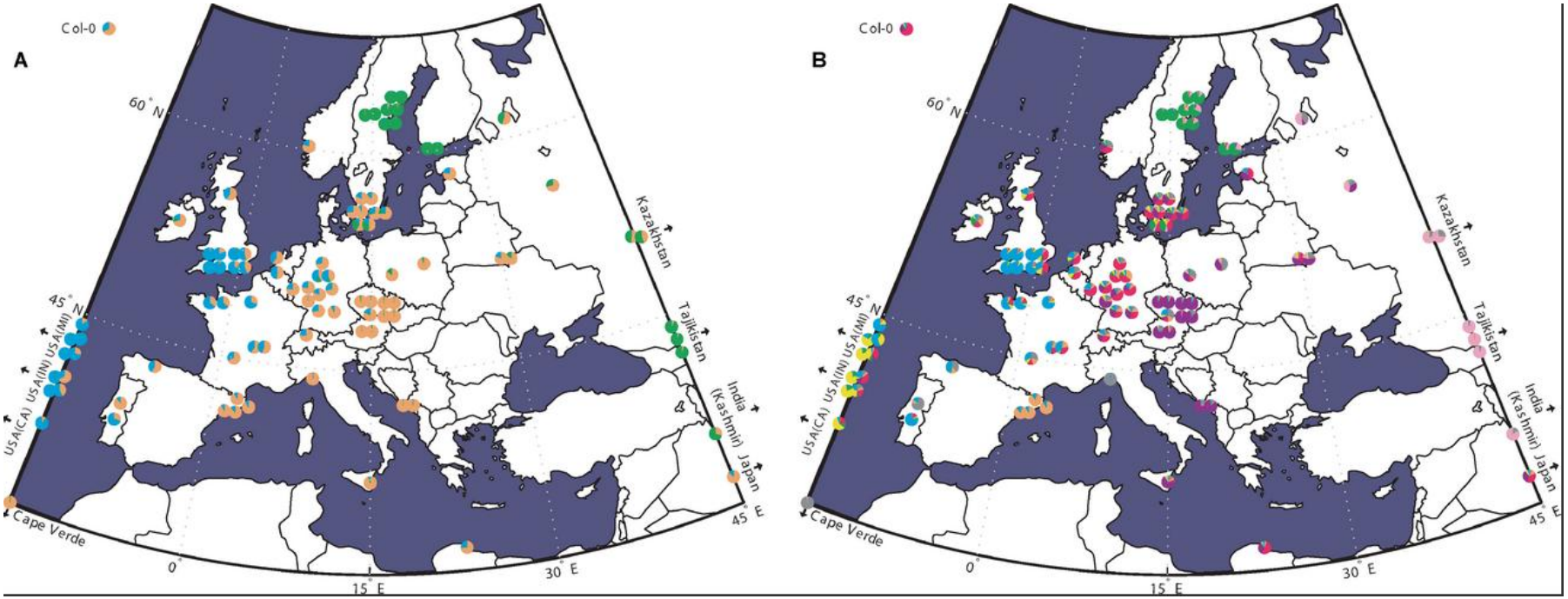
Arabidopsis thaliana

- THE model plant
- Genome sequence available since 2000
- 120 Mbp sequenced
- expected: 120-150 Mbp
- High number of seeds (10k)

- Over 1,000 different accessions “sequenced”

[picture removed]

Nordborg-Collection



https://doi.org/10.1371/journal.pbio.0030196

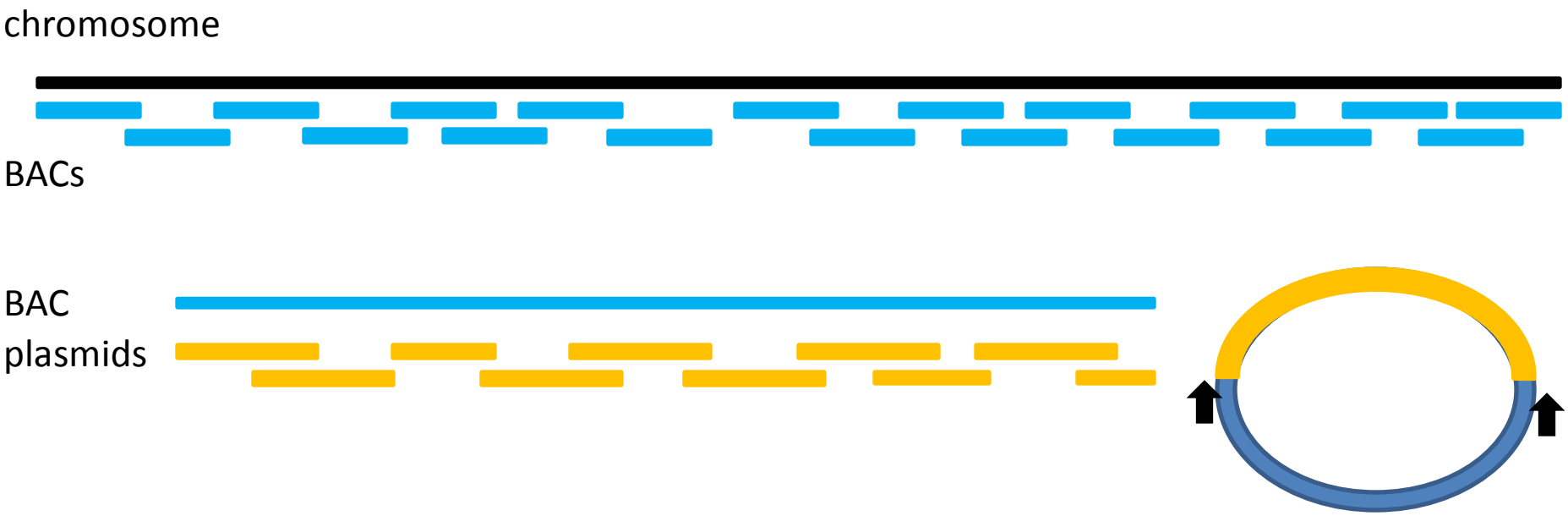
Nordborg-Collection in Bielefeld

- 27 different accessions are growing in the glasshouse
- Columbia-0 (Col-0), Niederzenz-1 (Nd-1), and Landsberg *erecta* (L*er*)
- Random assignment of accessions to participants
  - 4 different accessions per person

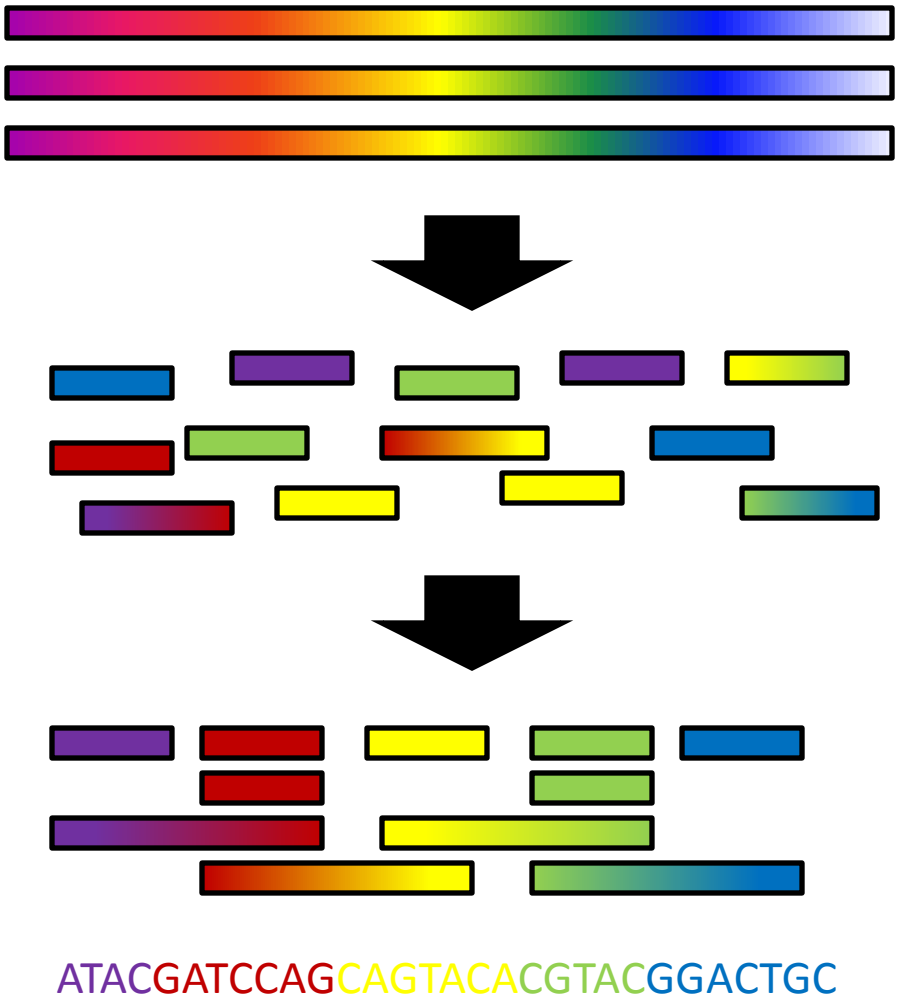
- Objective: investigate genetic/genomic differences!

BAC-by-BAC genome sequences

- *A. thaliana* Columbia-0 was sequences BAC-by-BAC
- Very expensive, but high quality
- BAC = bacterial artificial chromosome (insert<=300kb)

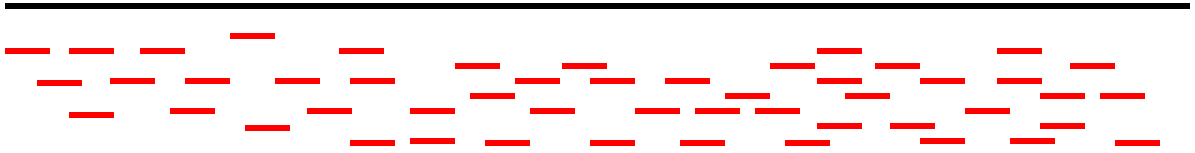


Whole Genome Shotgun (WGS) sequencing



Assembly problem

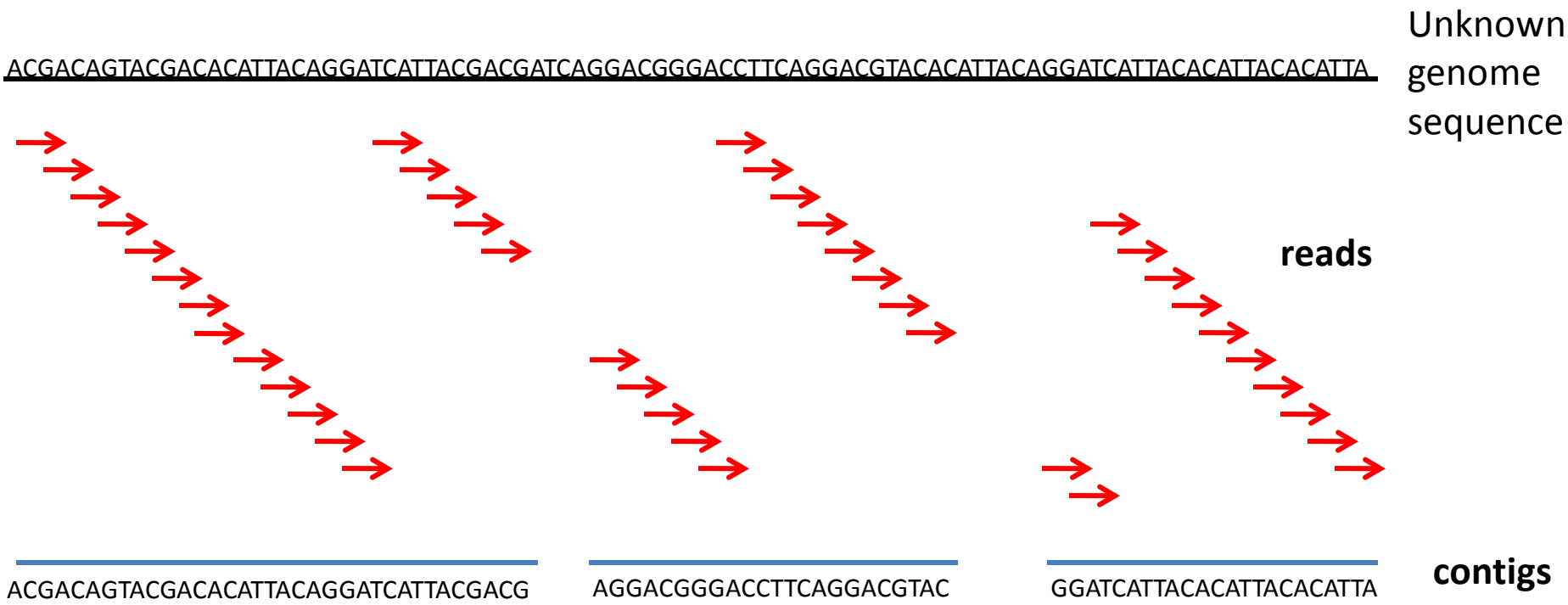
- Genome sequence length exceeds read length!



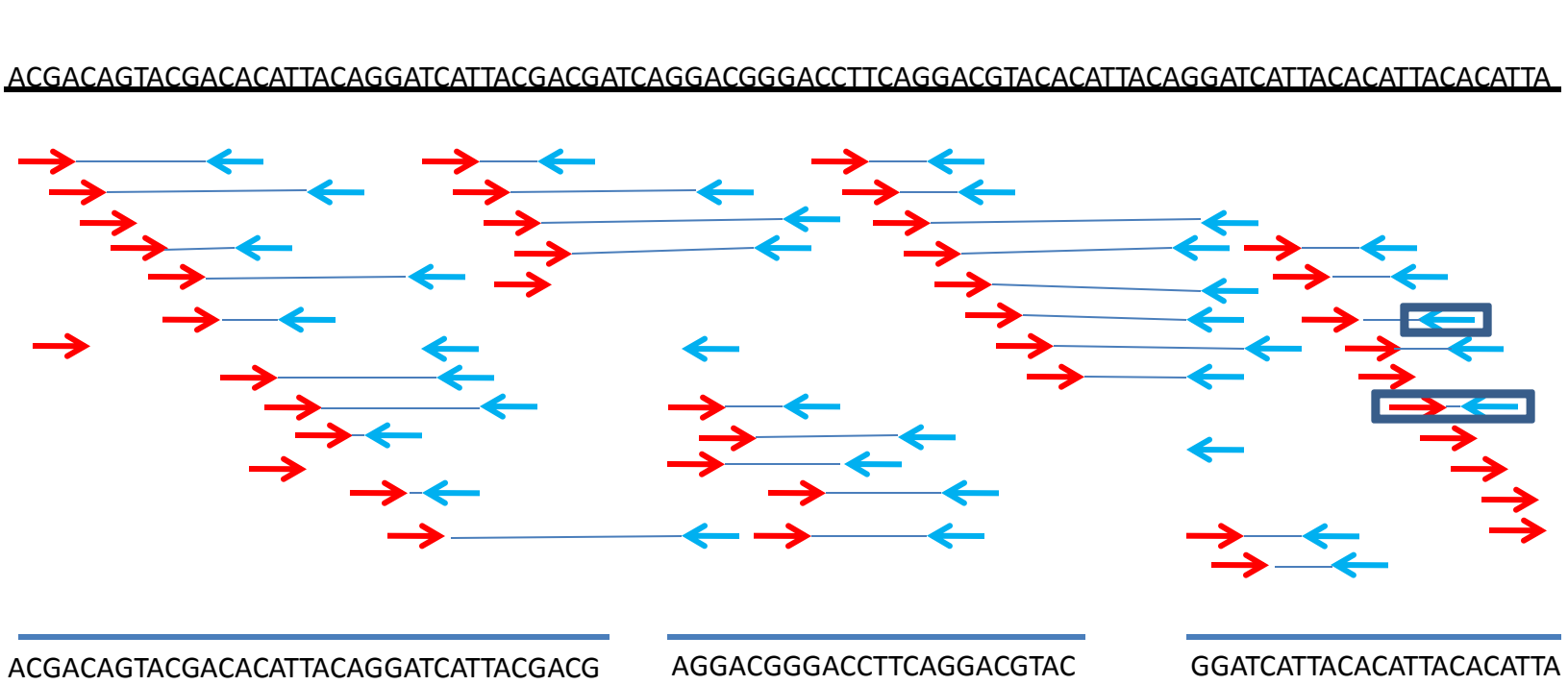
Genome sequence
(e.g. 125 Mbp *A. thaliana*)

Illumina Reads (32-300 nt)

Assembly

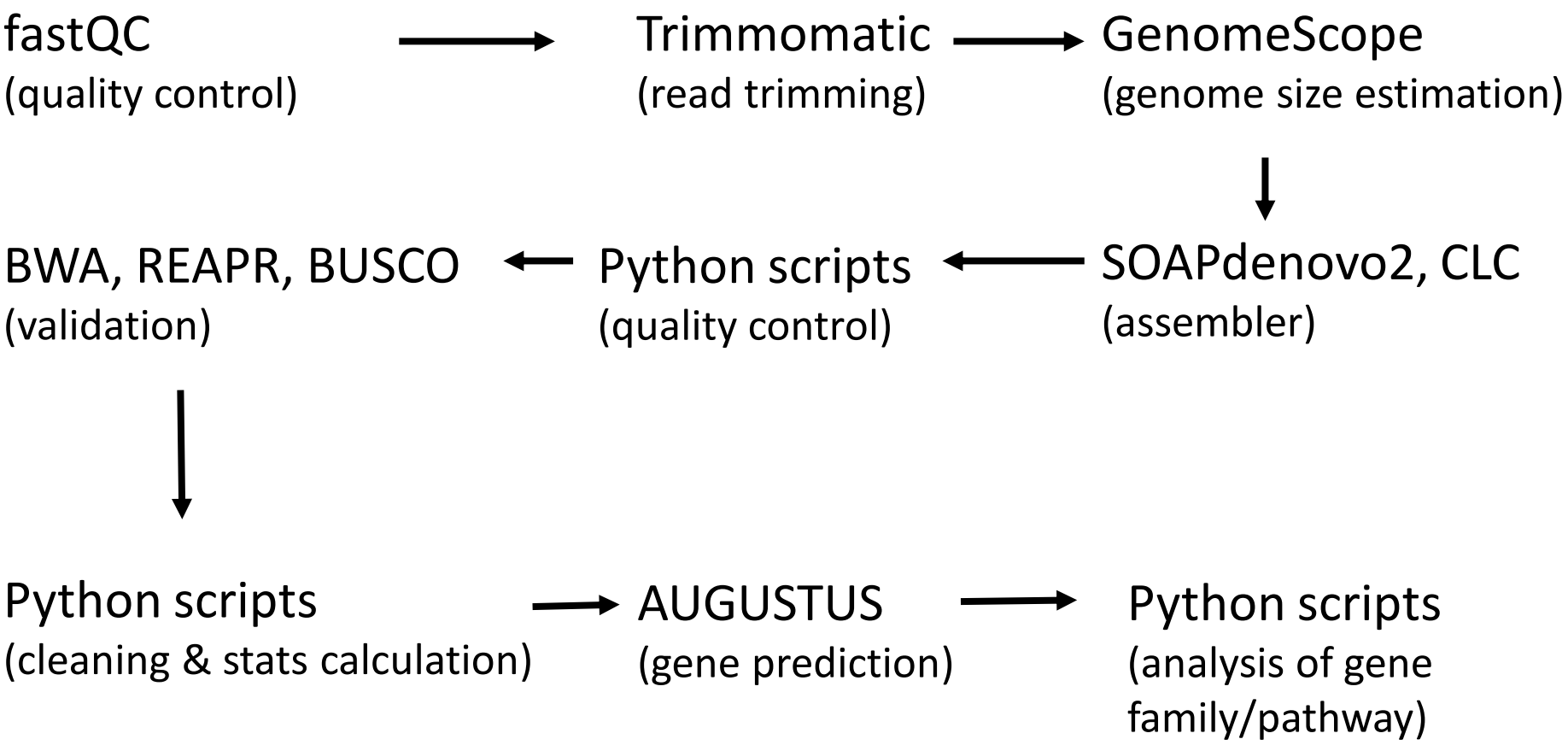

ACGACAGTACGACACATTACAGGATCATTACGACGATCAGGACGGGACCTTCAGGACGTACACATTACAGGATCATTACACATTACACATTA
Unknown genome sequence
reads
ACGACAGTACGACACATTACAGGATCATTACGACG
AGGACGGGACCTTCAGGACGTAC
GGATCATTACACATTACACATTA
contigs

Assembly

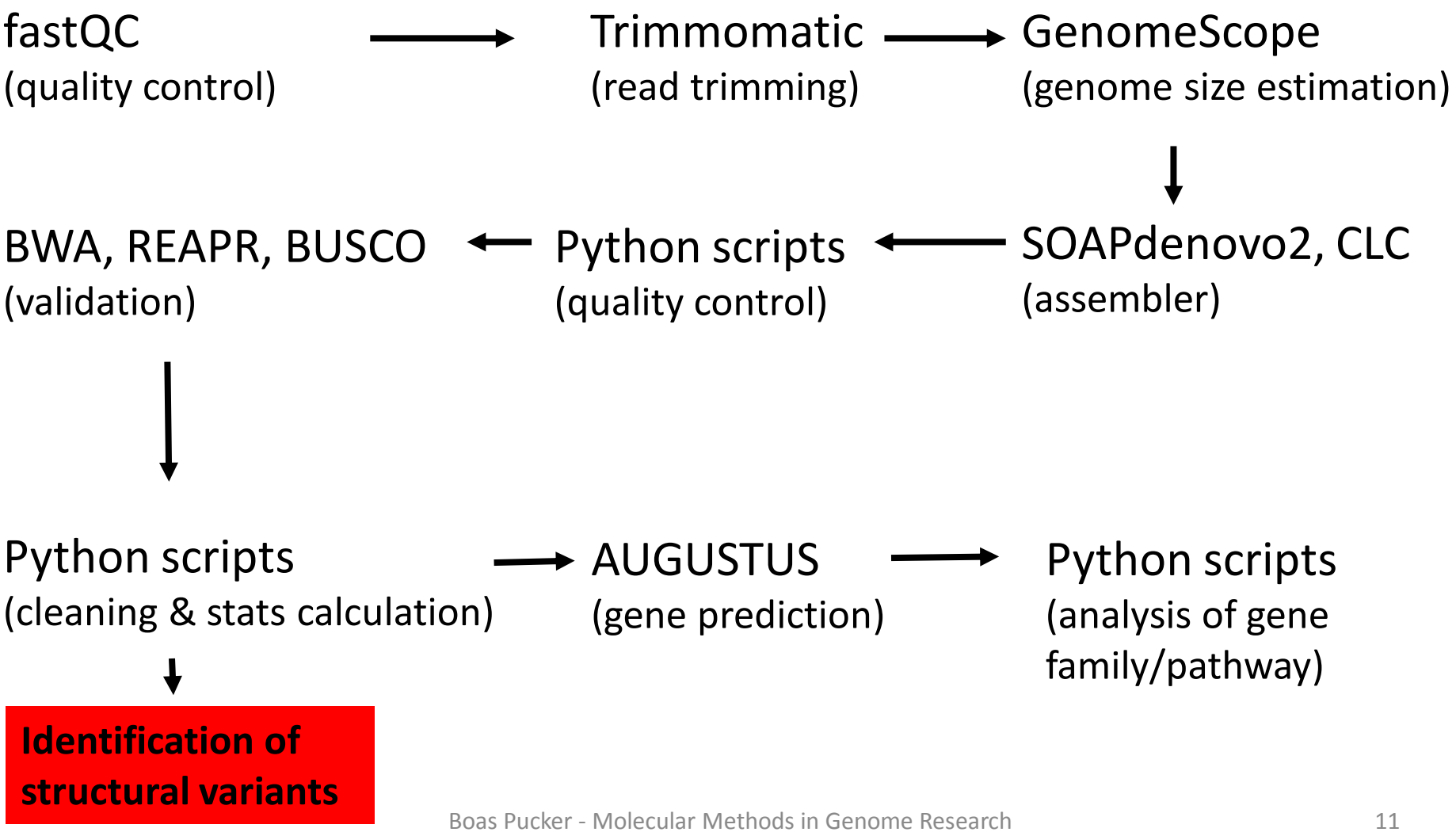


Contigs are connected by spanning fragments into scaffold: approximate size of gaps is known, but sequence remains unknown!

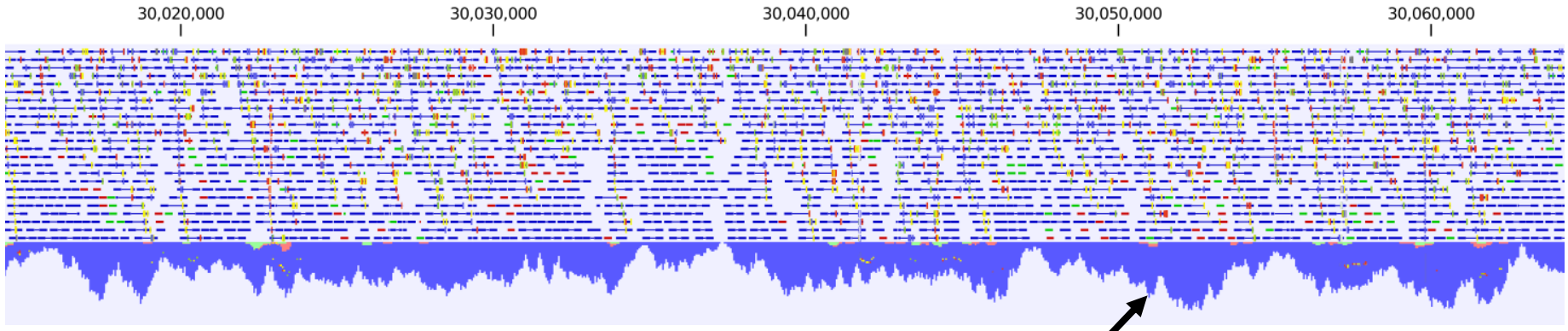
Bioinformatics workflow



Bioinformatics workflow



Assembly evaluation - coverage



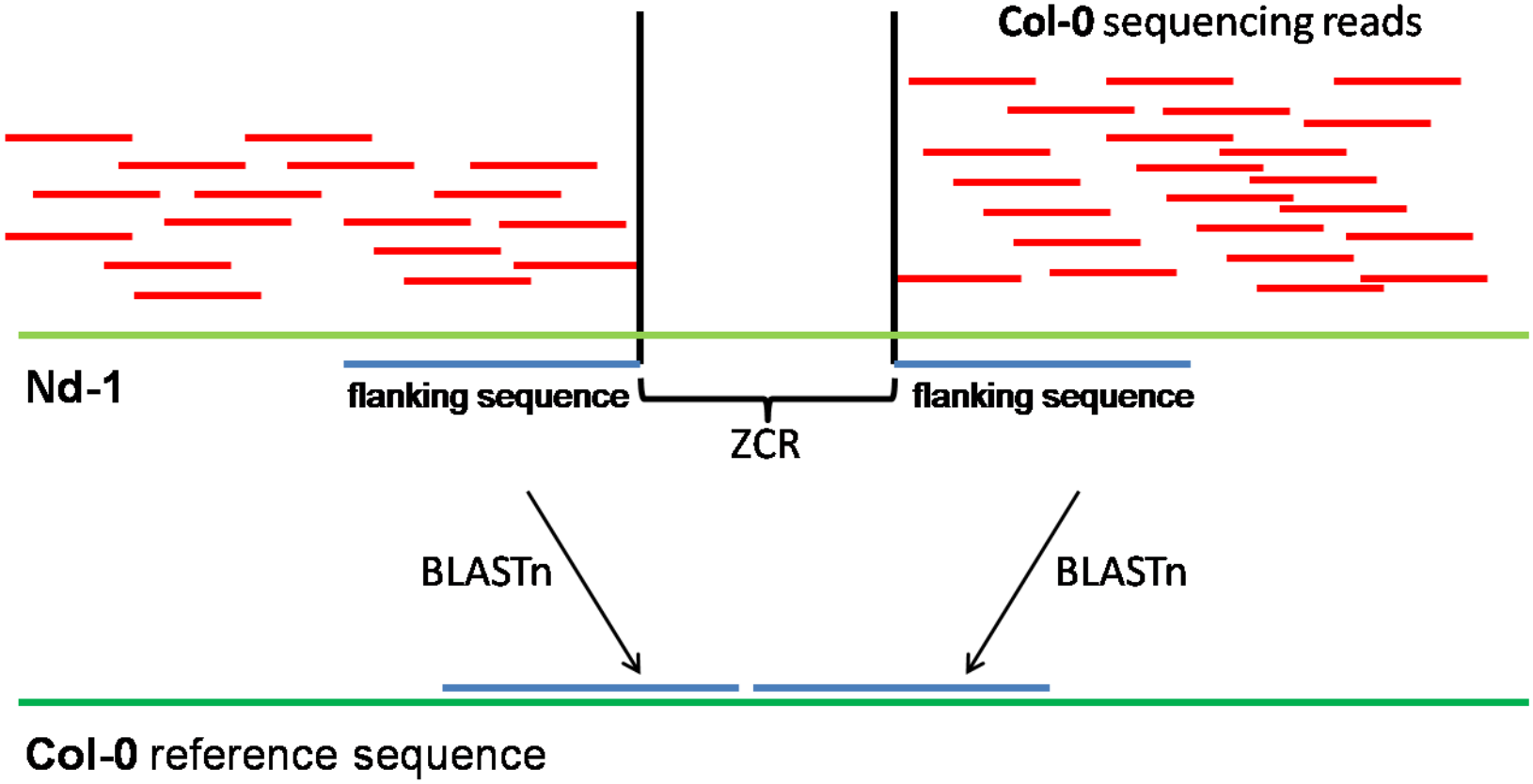
Read mapping of paired-end sequenced fragments (blue) to assembly

Coverage is too high to show all individual fragments at some positions

Read coverage depth:

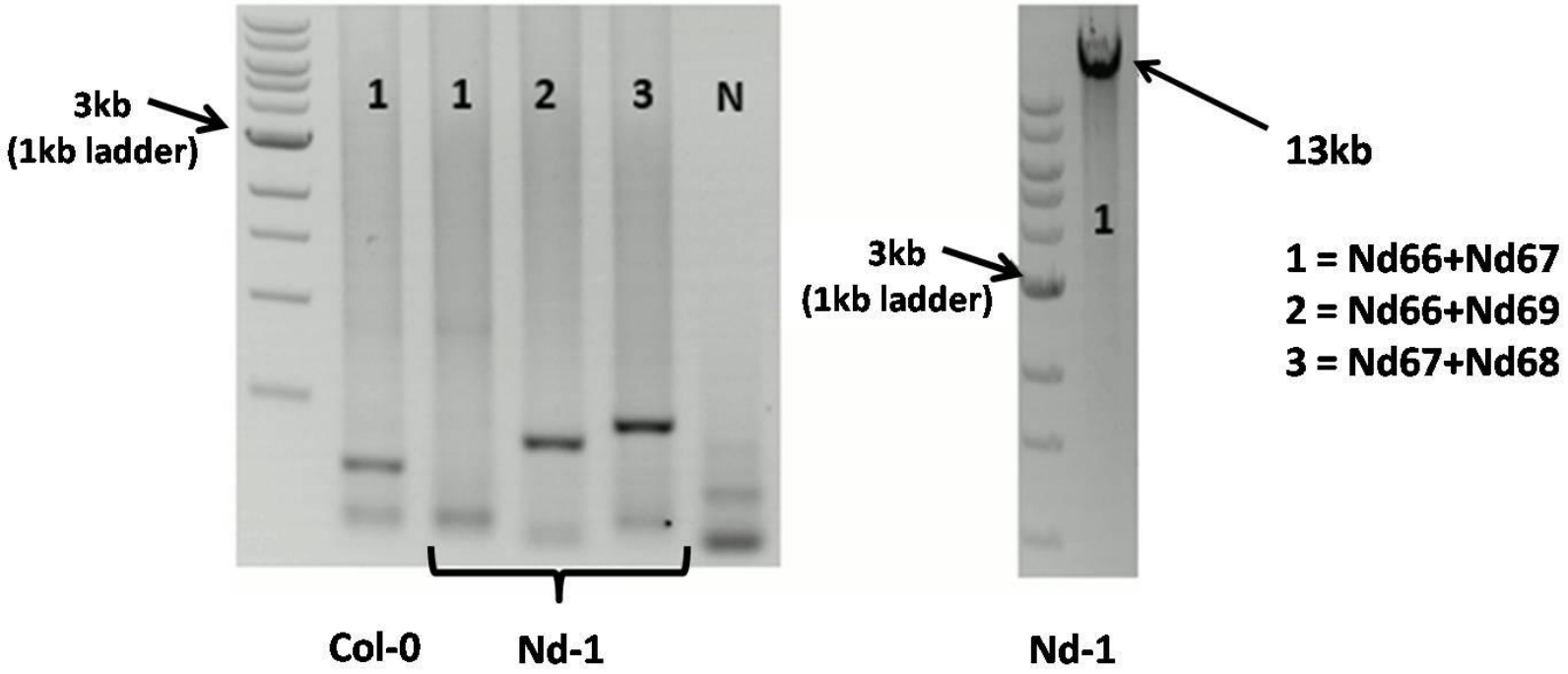
- number of times a certain base was sequenced
- Number of reads at one position

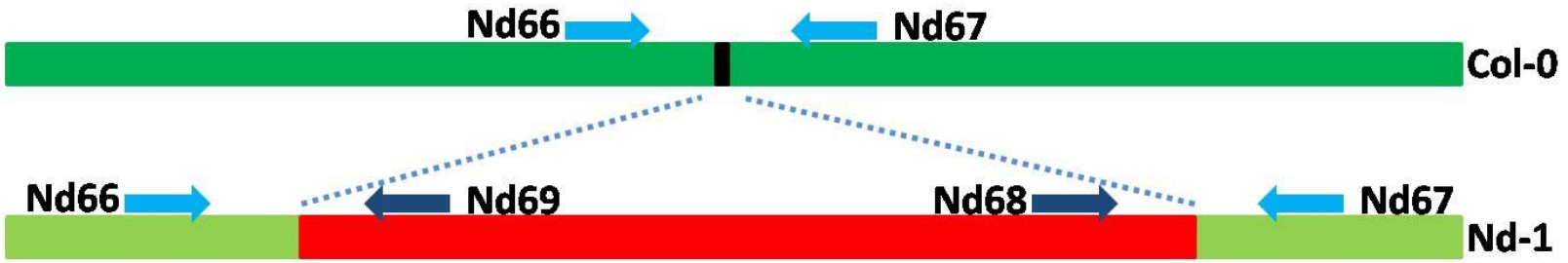
Structural variants



(Pucker et al., 2016)

Experimental validation





(Pucker et al., 2016)

How to proceed?

- Large number of TE-associated differences was previously identified:
- 1502 loci will be assigned to participants
- Sequence extraction:
  - seqex.py at https://github.com/bpucker/MolecularMethodsInGenomeResearch
- Primer design (manually)
- Primer validation via script:
  - check_primer_pairs.py at https://github.com/bpucker/MolecularMethodsInGenomeResearch
- Experiments: DNA extraction, PCR, gel electrophoresis

Primer design rules

- Length: 18-25 nucleotides (length determines costs)
- Annealing temperature: 50-72°C (optimal: 55-65°C)
- GC content: not relevant directly, but determines annealing temperature
- Avoid homopolymers (e.g. AAAA)
- Avoid complementary bases at start and end of your primer
- GC clamp: G or C at 3' end to enhance stable binding (no more than 3 G/C)
- Unique BLAST hit against reference genome sequence (if available)

- Similar annealing temperature for primers in pairs (1-2°C difference ok)

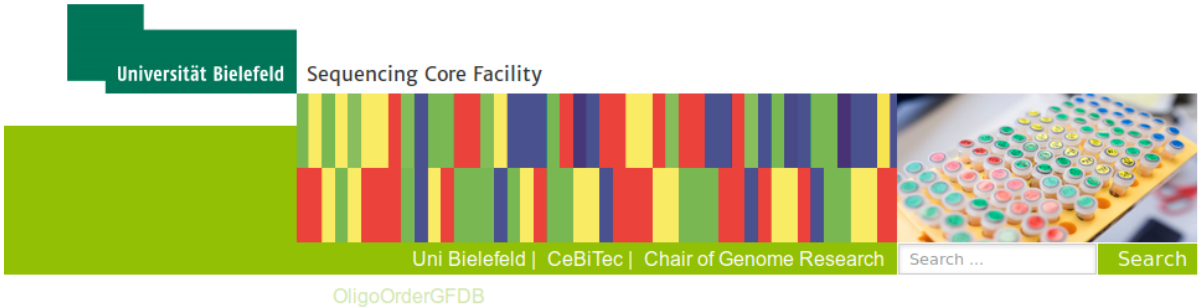
Primer validation

- Located on same sequence (same chromosome)
- Facing towards each other (proper orientation)
- Distance is matching the expectation
- No variants between accessions in primer binding sites

Designed primers need to be stored in FASTA format for validation:

```
>name1
sequence1
>name2
sequence2
>name3
sequence3
…
```

Ordering primers via OligoOrder



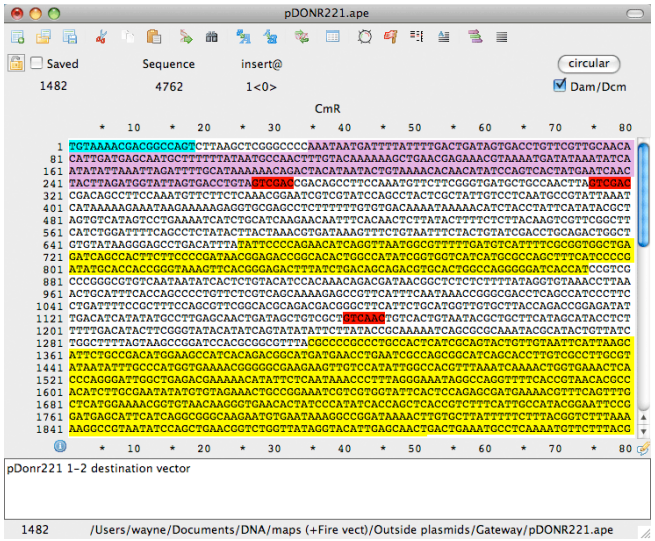
Ordering primers via OligoOrder

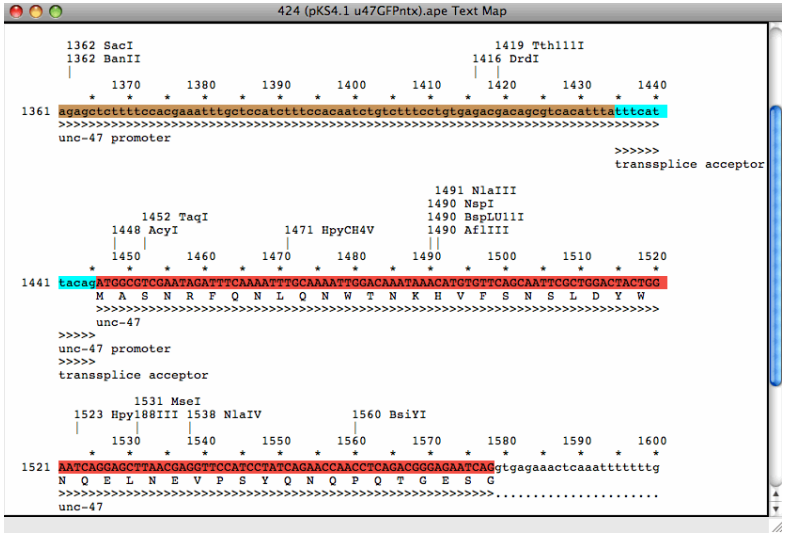
[picture removed]

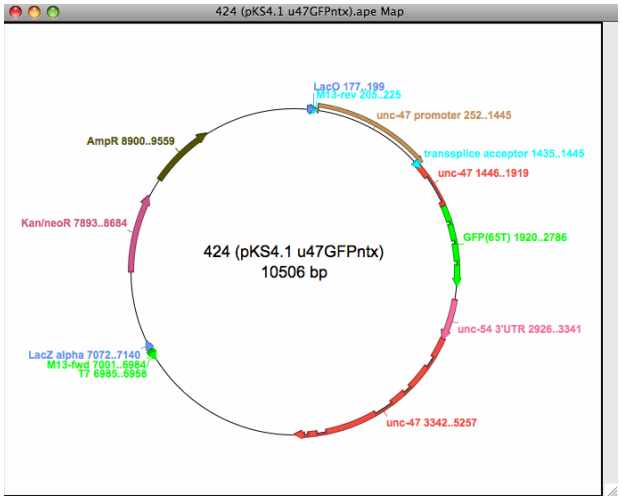
Required information per Primer: Please write it TAB-separated in your FASTA file headers

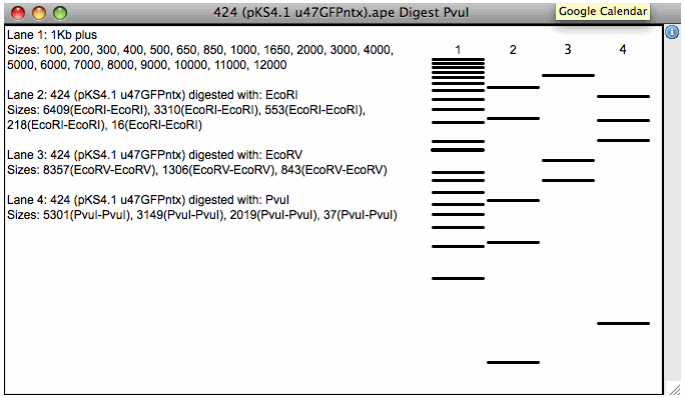
1) ID (will be assigned during process)
2) Name (10 characters)
3) Sequence length
4) Comment about intended usage
5) Sequence (this goes into a separate line!)

ApE – A plasmid editor









http://jorgensen.biology.utah.edu/wayned/ape/

ThinLinc

- IGLE or client on own computer
- Download ThinLinc client:
  - https://www.cendio.com/thinlinc/download
- LogIn:
  - Server: thinlinc.xxx.uni-bielefeld.de
  - ID + PW

[picture removed]

File manager

- Click on “homes”
  => Opens your personal “homes”

- Type“/xxx/xxx/xxx/” into file manager and hit ENTER

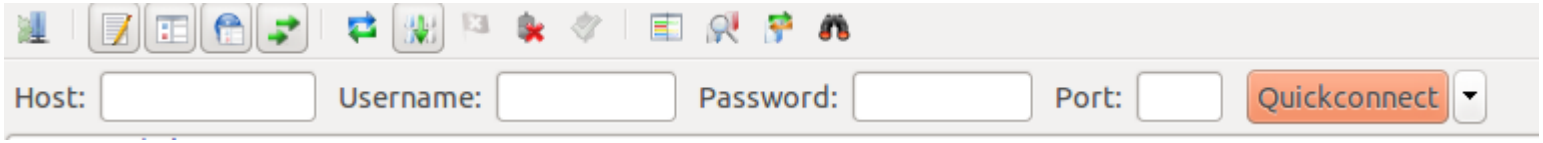
- right click and create a new subfolder: <UNIX-Name>

- Store all your files there

[picture removed]

Filezilla

- Download current version
- Start Filezilla



[picture removed]

Exercise: primer design

1. Extract your target sequences!
2. Design primer pairs for at least 10 regions! (multiple pairs per region possible)
3. Check primer pairs via script!
4. Select your best 3 primer pairs for ordering! (repeat 1-3 if necessary)
5. Ask another participant to check your primers again!
6. Send final primer pairs via E-mail to boas.pucker[at]uni-bielefeld.de!