

Target identification and primer design

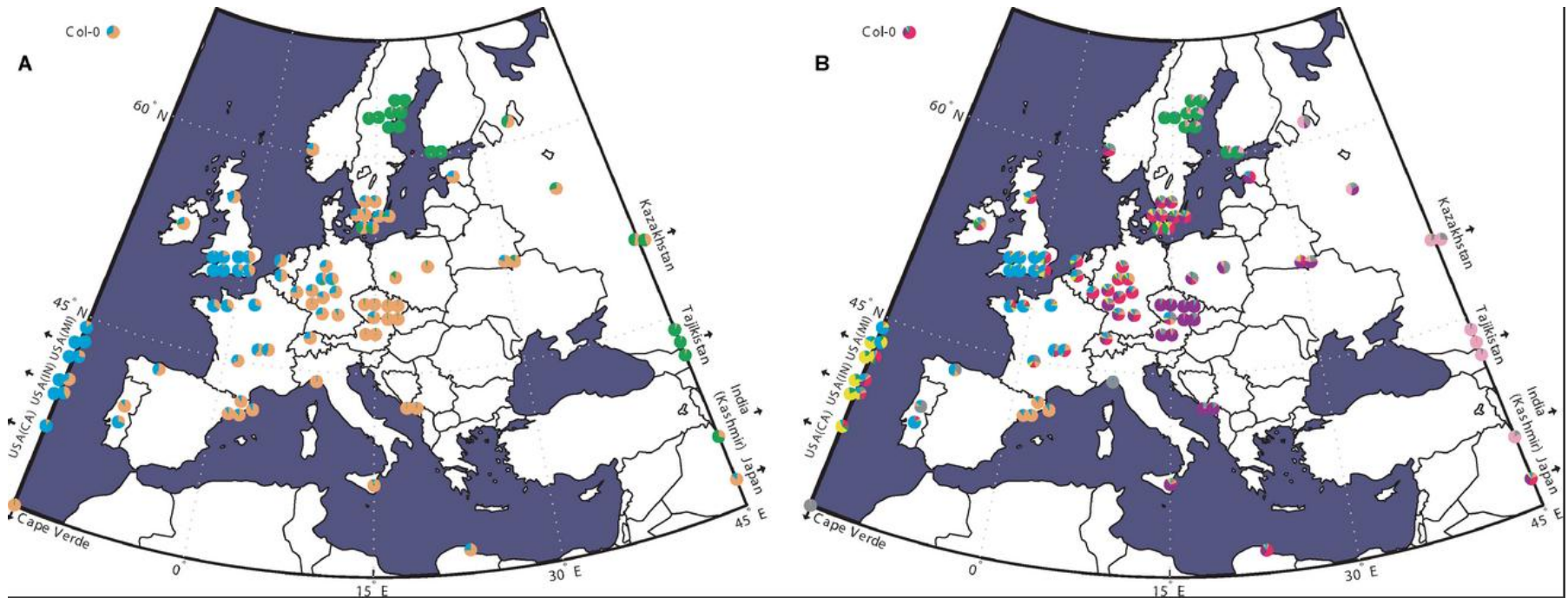
Boas Pucker

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* (Ler)
- 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 sequenced BAC-by-BAC
- Very expensive, but high quality
- BAC = bacterial artificial chromosome (insert $\leq 300\text{kb}$)

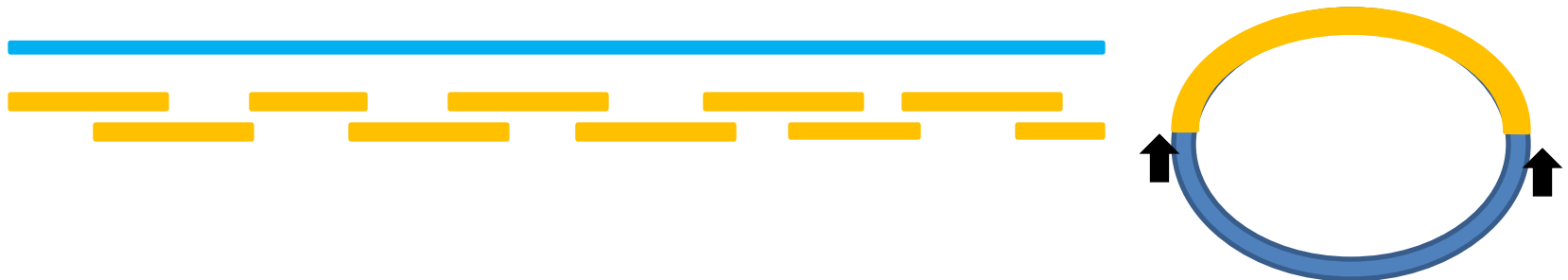
chromosome



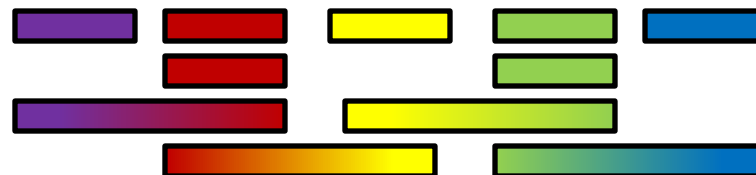
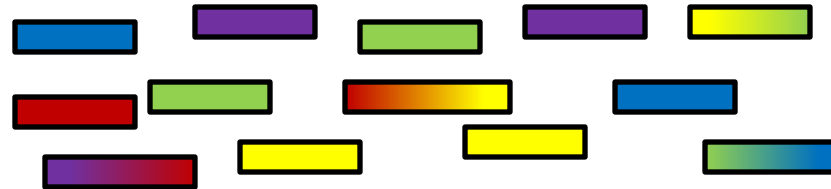
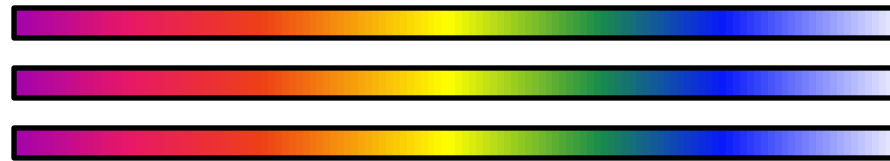
BACs

BAC

plasmids



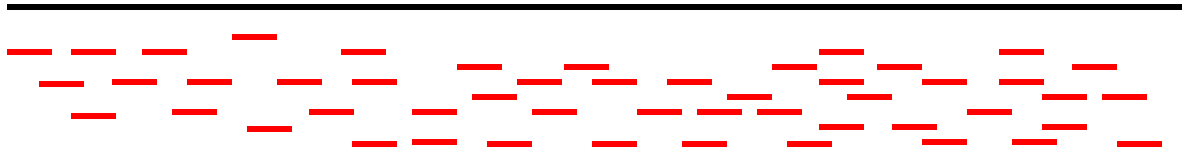
Whole Genome Shotgun (WGS) sequencing



ATACGATCCAGCAGTACACGTACGGACTGC

Assembly problem

- Genome sequence length exceeds read length!



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

Illumina Reads (32-300 nt)

Assembly

Unknown
genome
sequence

ACGACAGTACGACACATTACAGGATCATTACGACGATCAGGACGGGACCTTCAGGACGTACACATTACAGGATCATTACACATTACACATTA

reads

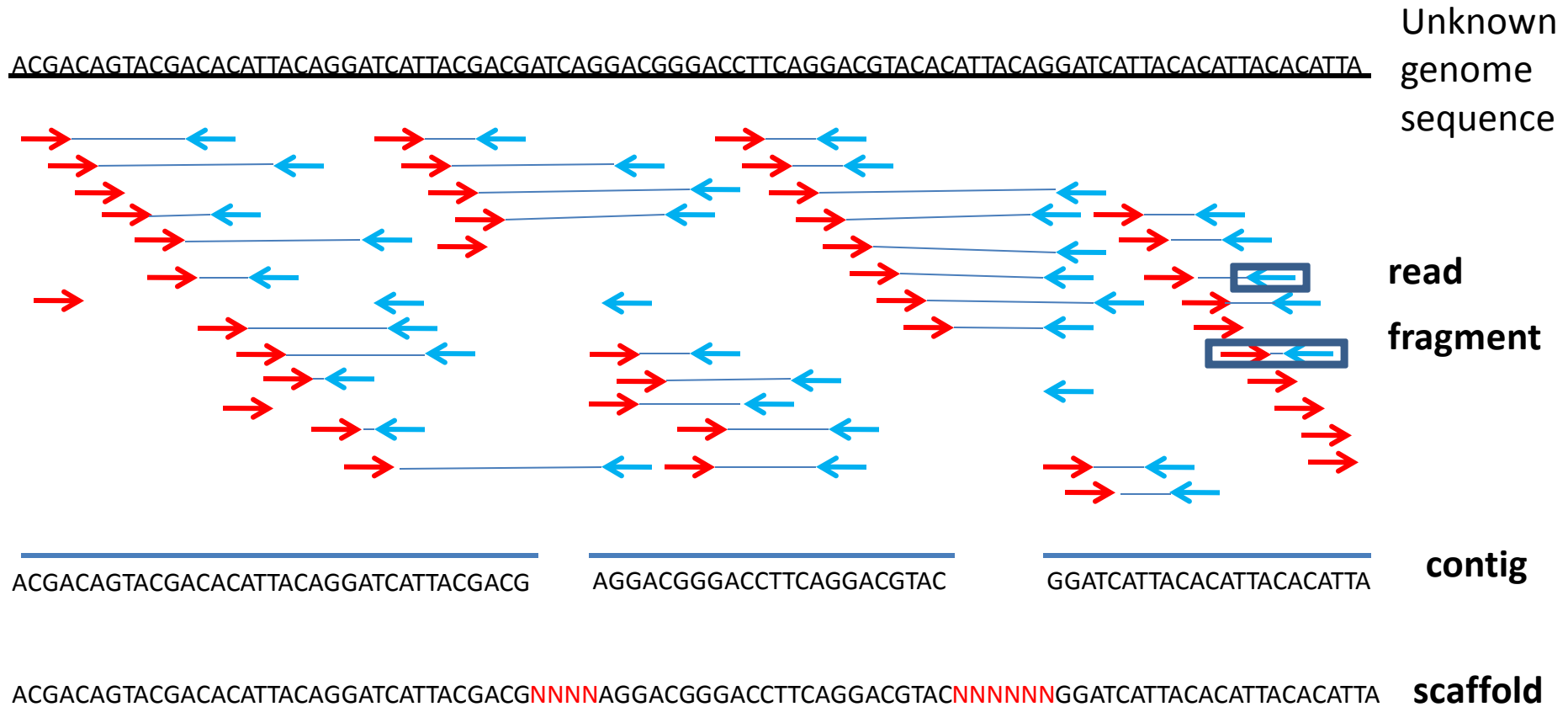
contigs

ACGACAGTACGACACATTACAGGATCATTACGACG

AGGACGGGACCTTCAGGACGTAC

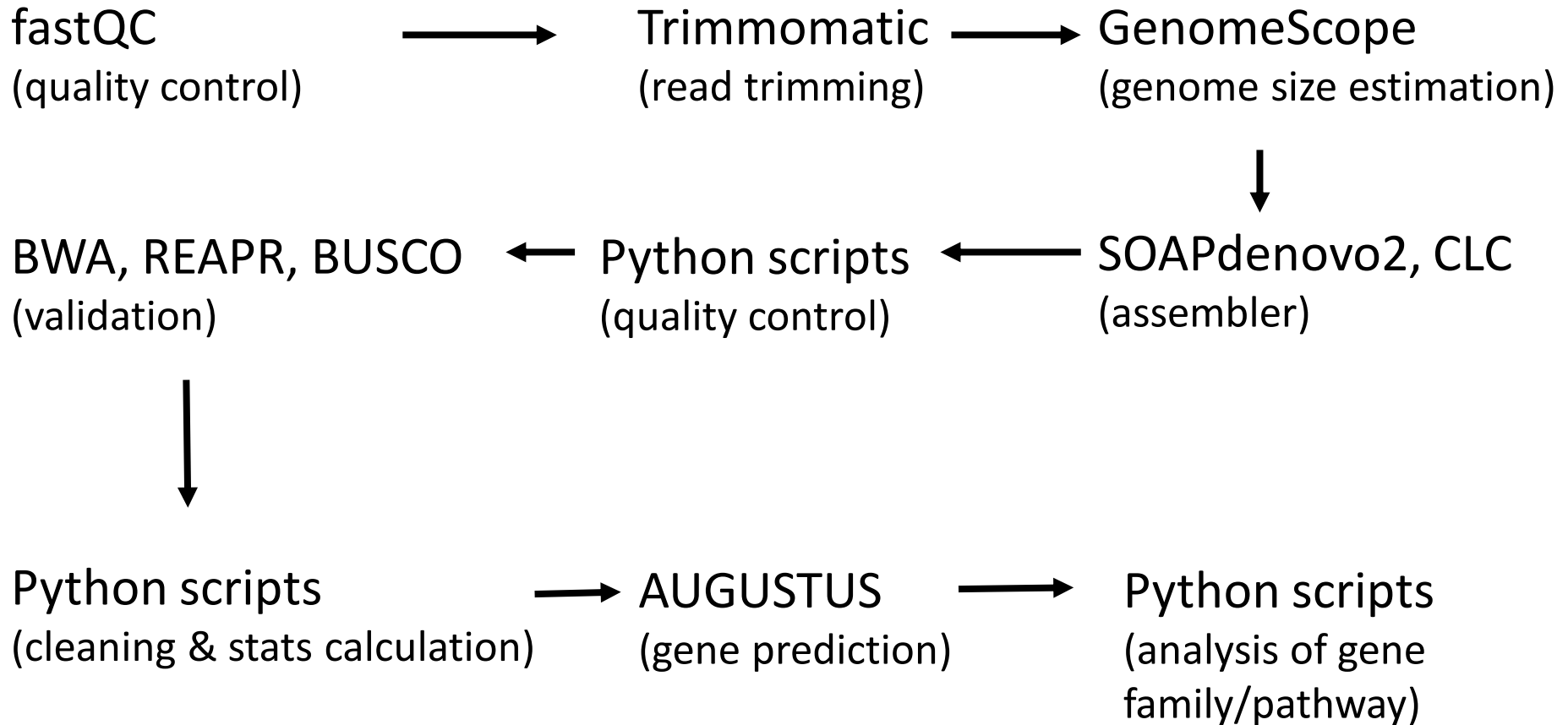
GGATCATTACACATTACACATTA

Assembly

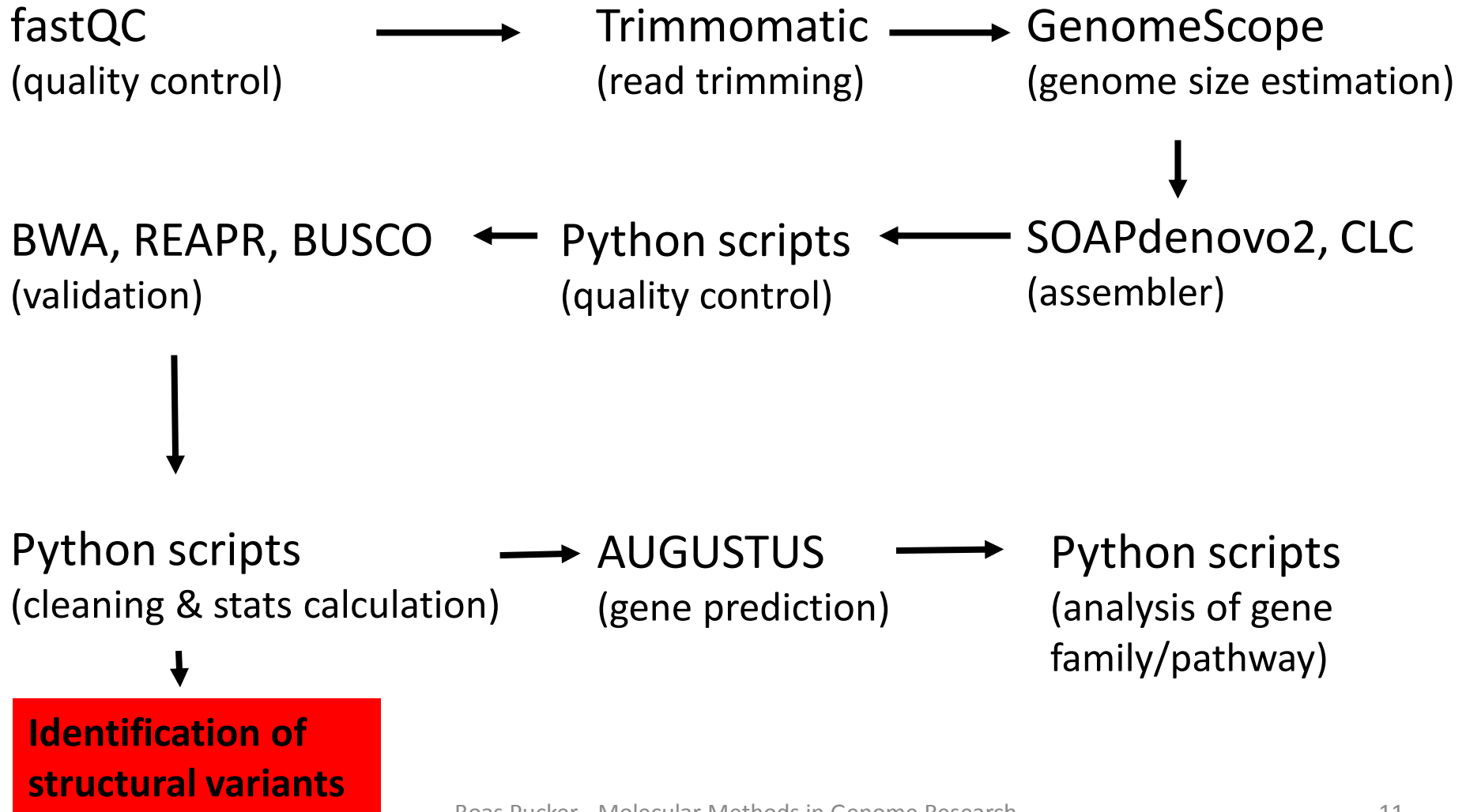


Contigs are connected by spanning fragments into scaffold: approximate size of gaps is known, but sequence remains unknown!

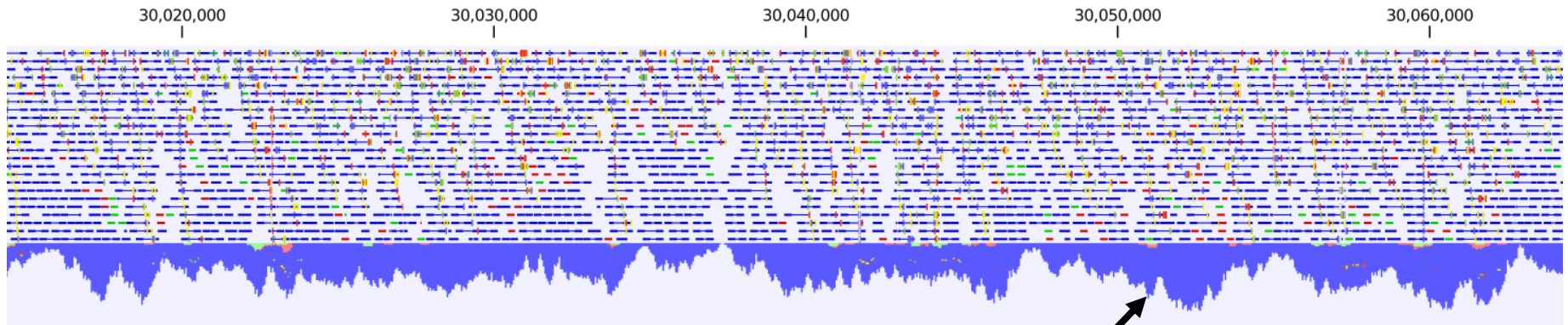
Bioinformatics workflow



Bioinformatics workflow



Assembly evaluation - coverage



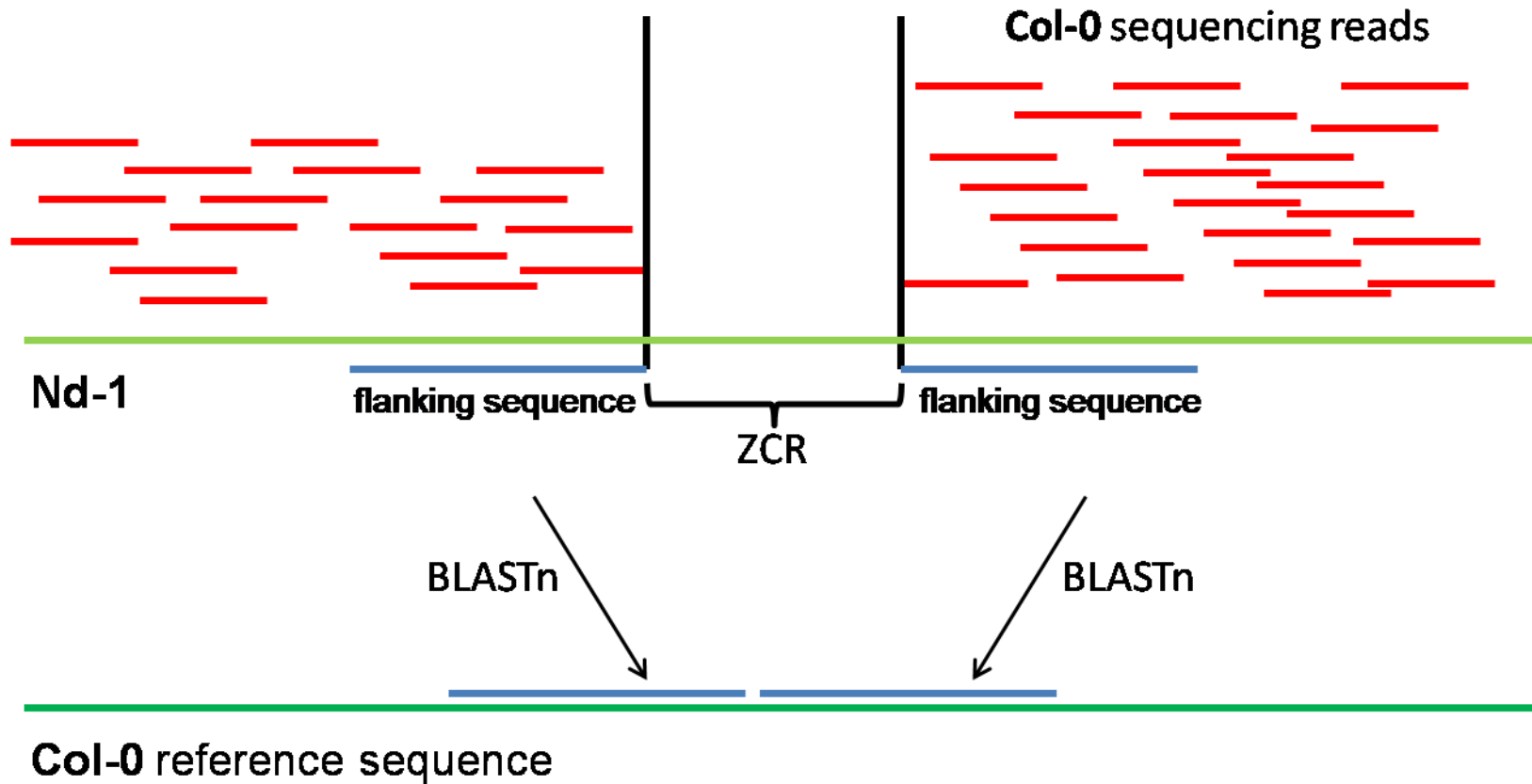
Read mapping of paired-end sequenced fragments (blue) to assembly

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

Read coverage depth:

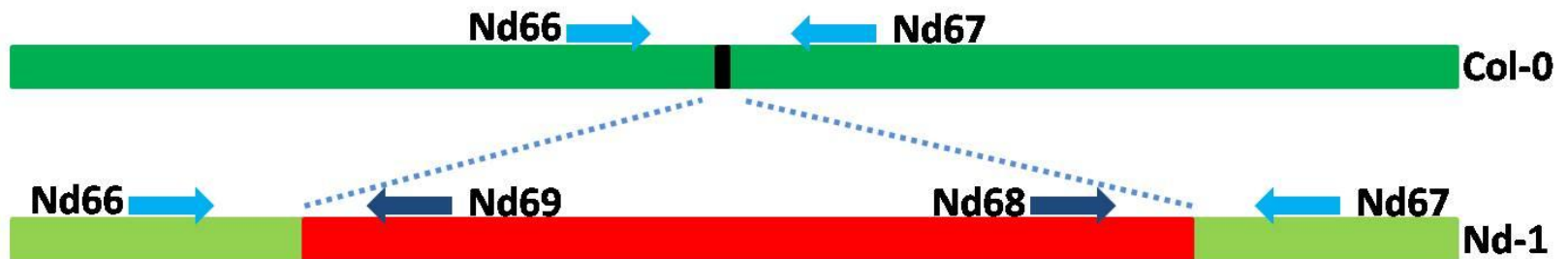
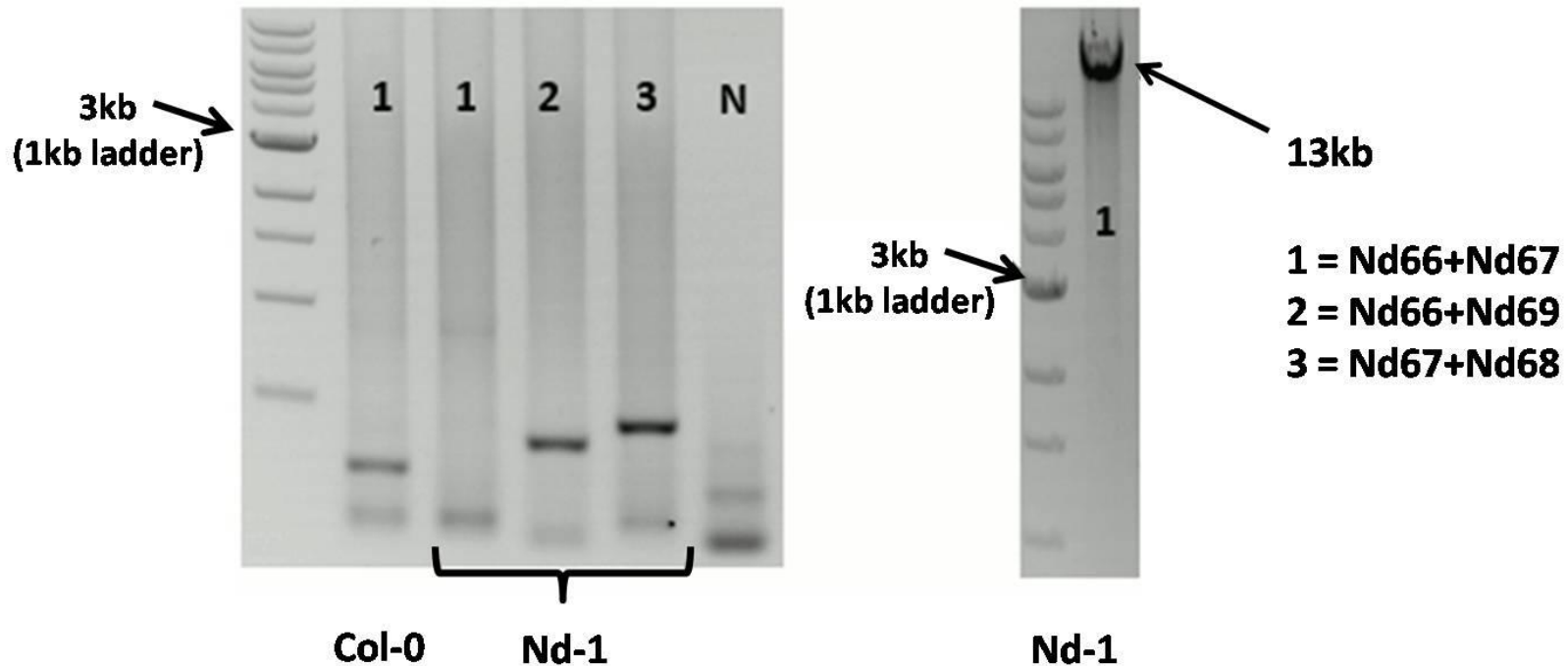
- number of times a certain base was sequenced
- Number of reads at one position

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

Universität Bielefeld

Sequencing Core Facility

Uni Bielefeld | CeBiTec | Chair of Genome Research

Search ...

Search

OligoOrderGFDB

Home

OligoOrderGFDB

Order oligos

Query OligoGFDB

Add oligo data

FAQ

Order FlowChart

SeqOrderGFDB

Kostenstellen

OligoOrderGFDB

Order oligos

YES, MY GROUP LEADER AGREES WITH MY OLIGO ORDER!

You need to make sure **before you order** that your group leader agrees with your order. Your group leader will be automatically informed about your order by Email. Price lists of the negotiated prices of the suppliers are included in OligoOrderGFDB. Please use the links in the header and footer of the pages after login.

Query OligoOrderGFDB or Search in OligoOrderGFDB

The interface should be self-explanatory. After searching, you can click the "edit icon" and add information to the "usage remarks" field. You can also download a list which contains the oligos you found in your search (search for your name and include the resulting list in M&M of your thesis ...).

If you have created links between sequence reactions in SeqOrderGFDB and oligos used as primers, you can find the sequencing reactions which have been created by using a given primer.

Approve pending oligo orders (department-specific)

Each oligo order must be approved ("bestätigt") by an additional person from the respective department. Access to this tool is restricted to selected persons. These are listed on the right side of the login page.

Add oligo data to OligoOrderGFDB

In addition to the automatic data collection when an oligo is ordered via the local WWW order form, you can add data to the database by using the oligo data entry form.

Listing of Kostenstellen available to OligoOrderGFDB (and SeqOrderGFDB)

This list should allow you to check if the Kostenstelle you would like to use for ordering is available. The terms are displayed exactly as accepted by the program.

OligoOrder GFDB - FAQ

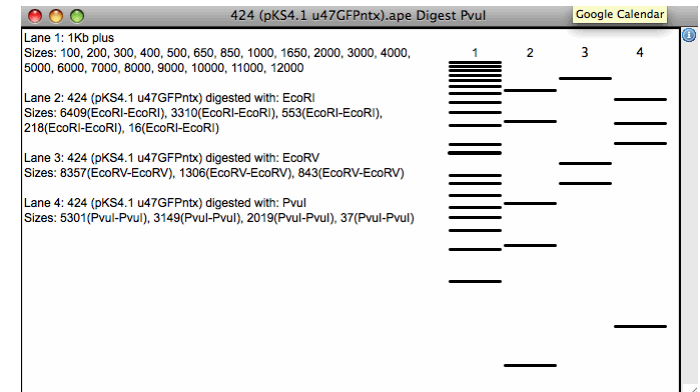
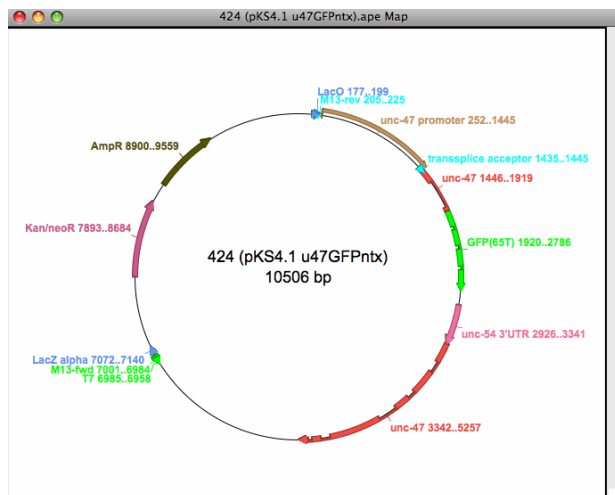
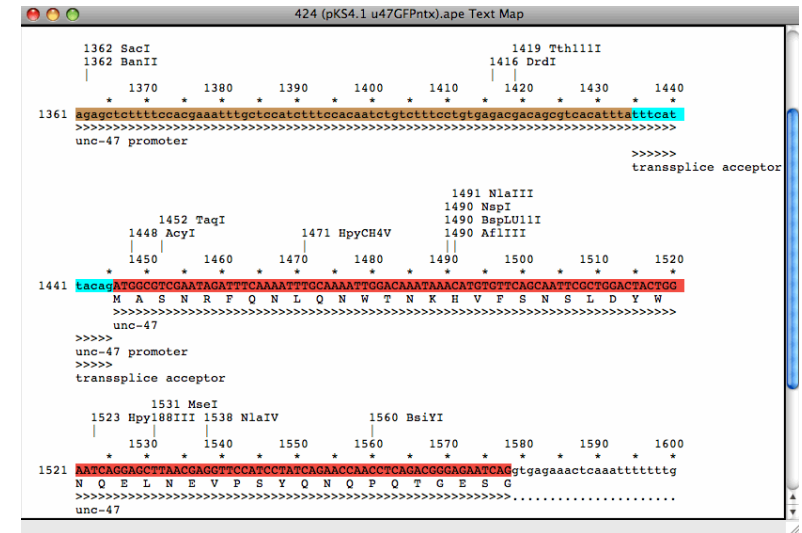
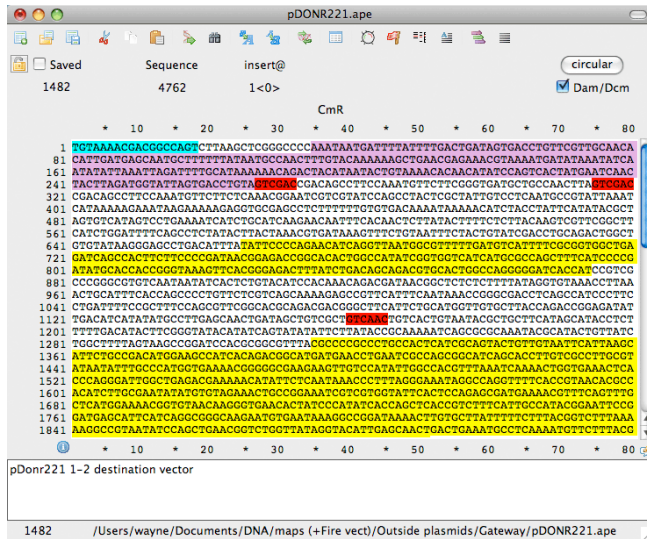
Ordering primers via OligoOrder

[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]

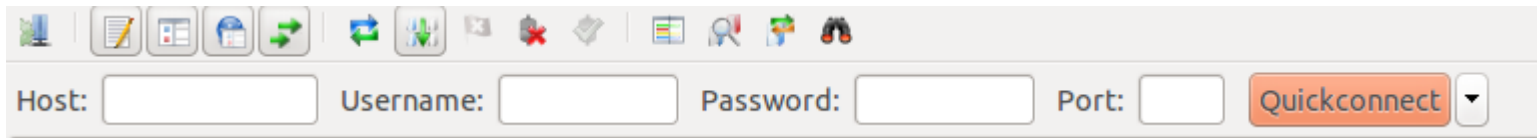
File manager

- 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](mailto:boas.pucker@uni-bielefeld.de)!