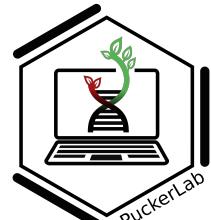


# PBPM2:

## Plant Genomics & Bioinformatics

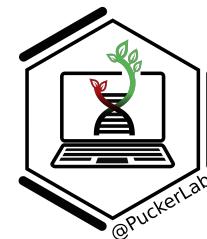
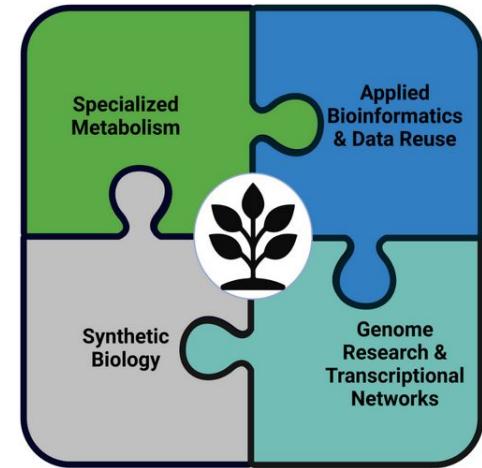


@puckerlab

Prof. Dr. Boas Pucker



- Biochemistry at HHU Düsseldorf
- (Systems) Biology at Bielefeld University
- Doctoral student (CeBiTec, Bielefeld University)
  - Genomics & Bioinformatics; synthetic biology (iGEM)
- Post doc (Ruhr-University Bochum)
- Post doc (Department of Plant Sciences, Cambridge, UK)
- Plant Biotechnology & Bioinformatics, TU Braunschweig (2021-2025)
- Plant Biotechnology & Bioinformatics, Uni Bonn (since 2025)
  - Evolution of specialized plant metabolism
  - Applied bioinformatics
  - Genomics/transcriptomics



# Availability of slides

- All materials are freely available (CC BY) - after the lectures:
  - eCampus: PBPM2 - Plant Genomics and Bioinformatics
  - GitHub: <https://github.com/bpucker/teaching/PBPM2>
- Questions: Feel free to ask at any time
- Feedback, comments, or questions: pucker[a]uni-bonn.de



My figures and content can be re-used in accordance with CC BY 4.0, but this might not apply to all images/logos. Some figure were constructed using bioRender.com.

# Discovery of DNA

- Isolation of DNA by Friedrich Miescher in 1869
- Working in the castle in Tübingen



Friedrich Miescher  
(1844-1895)

# Discovery of DNA structure

- WATSON, J., CRICK, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953). <https://doi.org/10.1038/171737a0>

## MOLECULAR STRUCTURE OF NUCLEIC ACIDS

### A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.



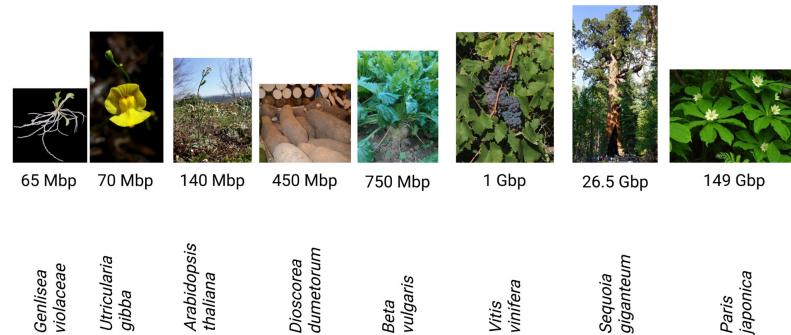
# What is the genome of a plant?

# Major types of DNA in plants

- gDNA from the nucleus
- mtDNA from the mitochondria (chondrome)
- cpDNA from the chloroplast (plastome)
- pDNA (plasmids, only in biotechnological applications)

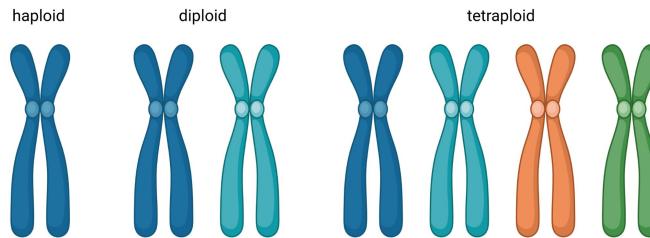
# Plant genome sizes

- Genome size: variation from 65 Mbp to 149 Gbp
- Ploidy: haploid/diploid genomes are much easier to analyze than polyploid genomes

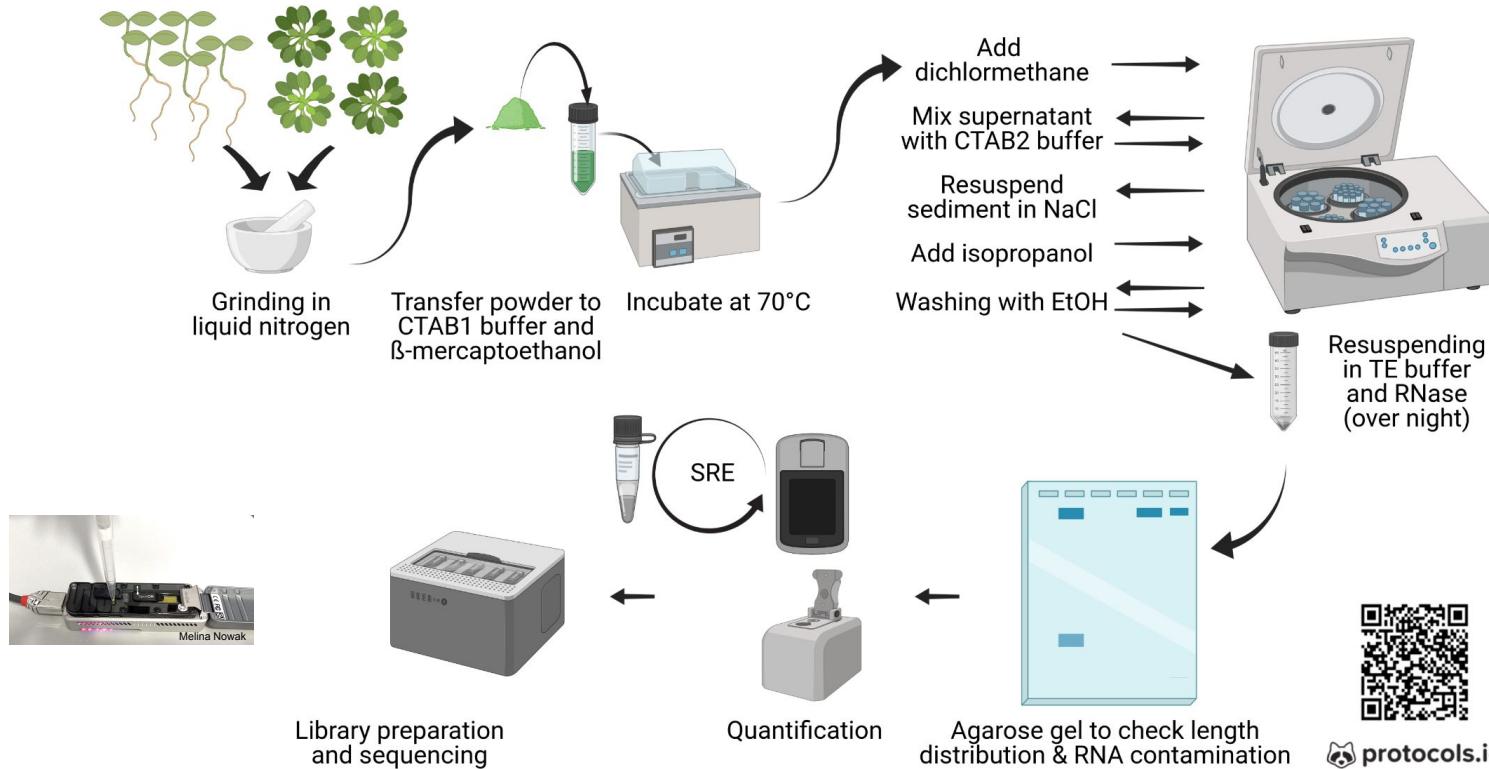


*Genlisea violacea*  
*Utricularia gibba*  
*Arabidopsis thaliana*  
*Dioscorea dumetorum*  
*Beta vulgaris*  
*Vitis vinifera*  
*Sequoia giganteum*  
*Paris japonica*

*Genlisea violacea*: [https://commons.wikimedia.org/wiki/File:Genlisea\\_violacea\\_giant.jpg](https://commons.wikimedia.org/wiki/File:Genlisea_violacea_giant.jpg)  
*Utricularia gibba*: [https://commons.wikimedia.org/wiki/File:Utricularia\\_gibba\\_1.jpg](https://commons.wikimedia.org/wiki/File:Utricularia_gibba_1.jpg)  
*Arabidopsis thaliana*: [https://commons.wikimedia.org/wiki/File:Arabidopsis\\_thaliana\\_st18.jpg](https://commons.wikimedia.org/wiki/File:Arabidopsis_thaliana_st18.jpg)  
*Dioscorea dumetorum*: <https://commons.wikimedia.org/wiki/File:YamsatBrixtonMarket.jpg>  
*Beta vulgaris*: <https://de.wikipedia.org/wiki/Zucker%C3%BCbe#/media/File:Zucker%C3%BCbe.jpg>  
*Vitis vinifera*: [https://commons.wikimedia.org/wiki/File:Vitis\\_vinifera\\_201205.jpg](https://commons.wikimedia.org/wiki/File:Vitis_vinifera_201205.jpg)  
*Sequoia giganteum*: [https://commons.wikimedia.org/wiki/File:Giant\\_Mariposa\\_Grove.jpg](https://commons.wikimedia.org/wiki/File:Giant_Mariposa_Grove.jpg)  
*Paris japonica*: [https://commons.wikimedia.org/wiki/File:Paris\\_japonica\\_Kinugasasou\\_in\\_Hakusan\\_2003\\_7\\_27.jpg](https://commons.wikimedia.org/wiki/File:Paris_japonica_Kinugasasou_in_Hakusan_2003_7_27.jpg)



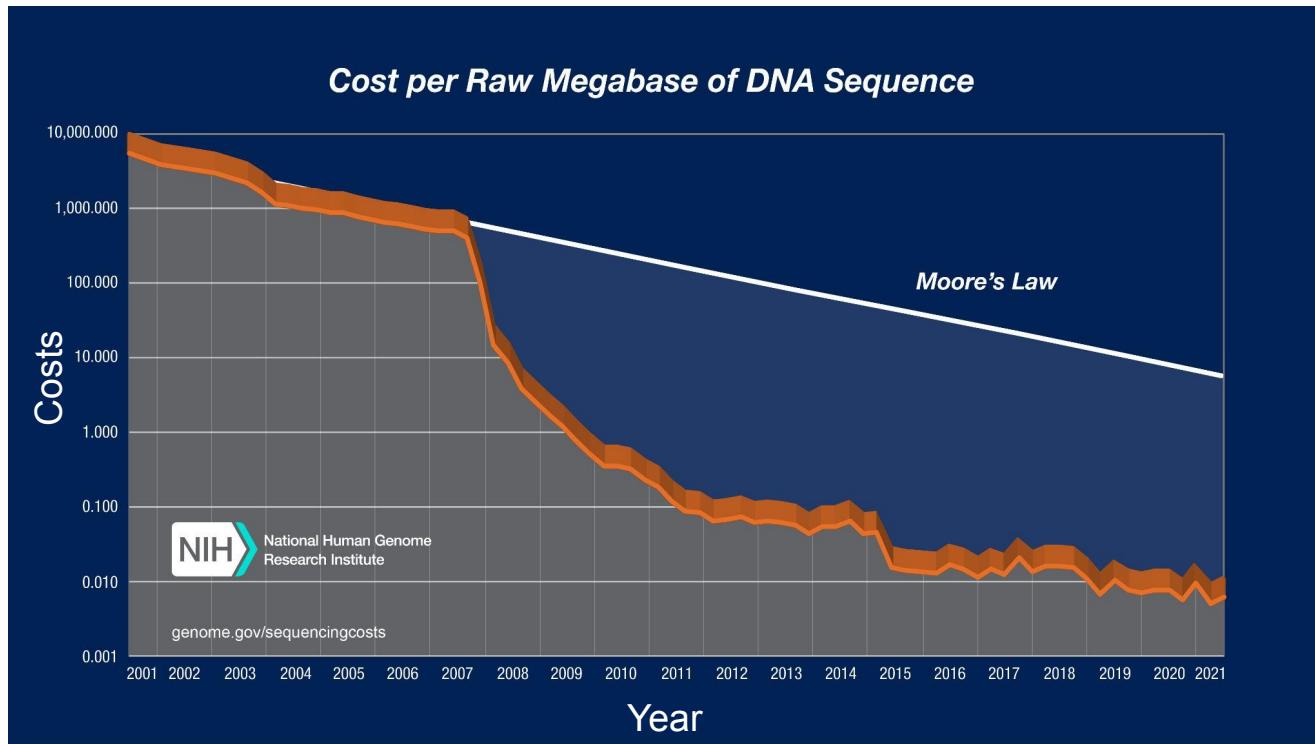
# CTAB-based DNA extraction



# Overview of sequencing technologies

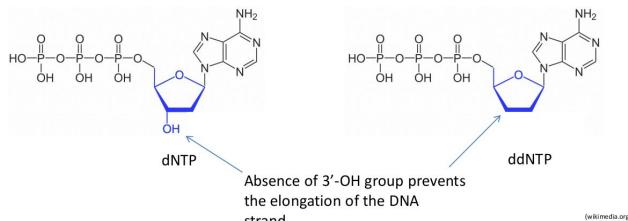
- Generation 1:
  - **Sanger sequencing**
  - Max-Gilbert sequencing
- Generation 2 (massive parallel sequencing):
  - 454 pyrosequencing
  - **Solexa/Illumina sequencing**
  - SOLID
  - Ion Torrent
  - BGI-seq
  - Synthetic long reads
- Generation 3 (long reads):
  - **Pacific Biosciences (PacBio)**
  - **Oxford Nanopore Technologies (ONT)**
- Generation 4:
  - What is next?

# Development of sequencing capacity

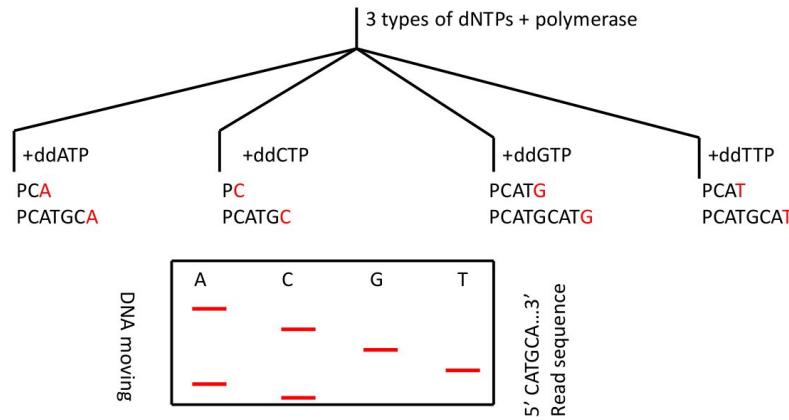


# Sanger sequencing

# Concept of Sanger sequencing

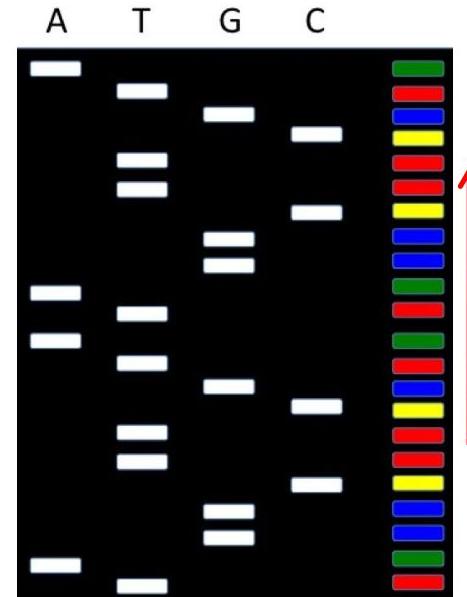


Primer (P): 5' –TGCATGGCATGATGCATG–3'  
 Template: 3' –ACGTACCGTACTACGTACGTACGTACGTCTAGGT–5'



# Sanger sequencing - original version

Two persons analyze the gel: one is calling the base ('basecaller') and the other person is writing down the bases

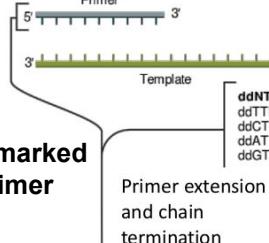


(modified from wikipedia.org)

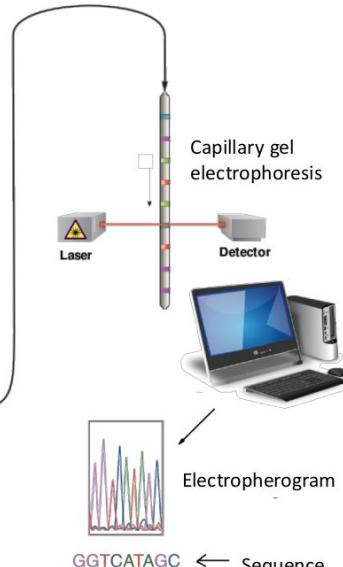
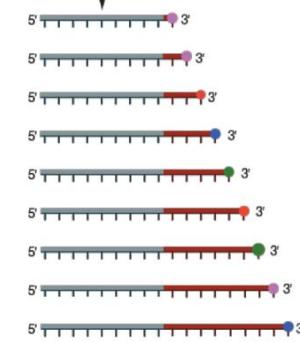
# Sanger sequencing - today

**Only one reaction!**

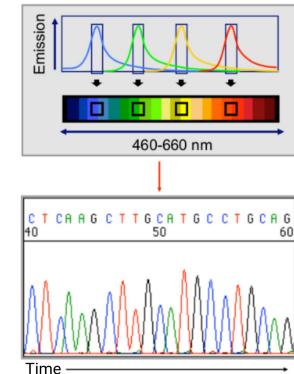
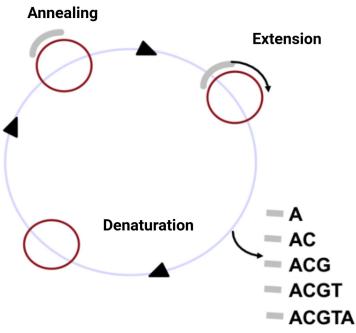
**ddNTPs are marked instead of primer**



**Low input required due to cycle sequencing**



**Capillary avoids interference of adjacent lanes on a gel!**



# FASTA

- There are two types of lines: header and sequence
- Header line starts with '>'; can contain name and information about sequence
- Example:

```
>seq1 len=5
```

```
ACGTA
```

```
>seq2 len=10
```

```
ACGTA
```

```
ACGTA
```

```
>seq len=1
```

```
A
```

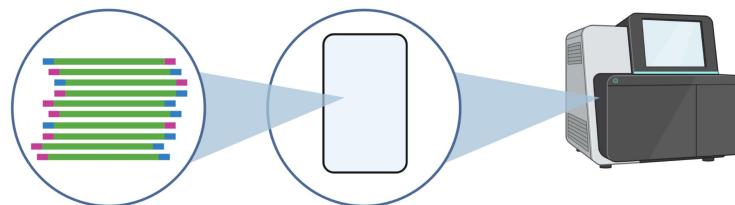
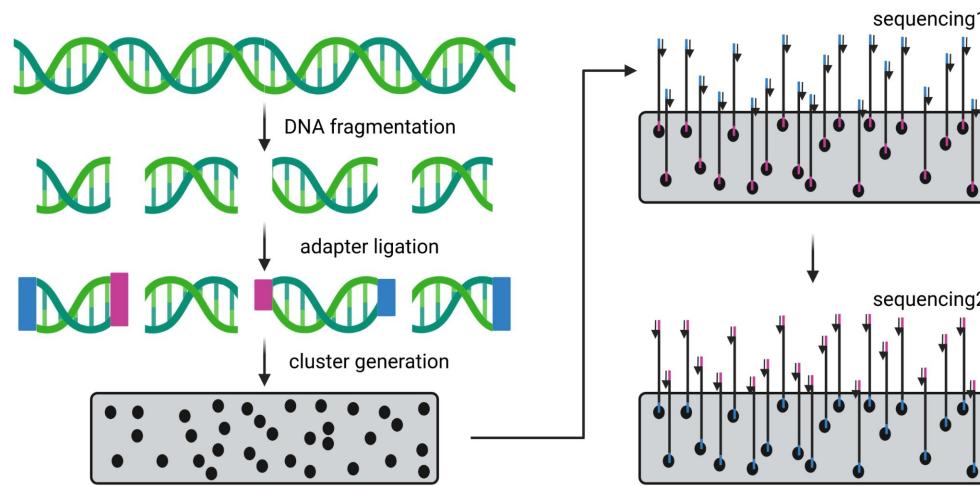
# Illumina sequencing

## Phred-Score

- Negative logarithm of the error probability for given position in read
- Multiplication by 10 to avoid floats

Phred quality score	Error probability	Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

# Illumina sequencing (overview)

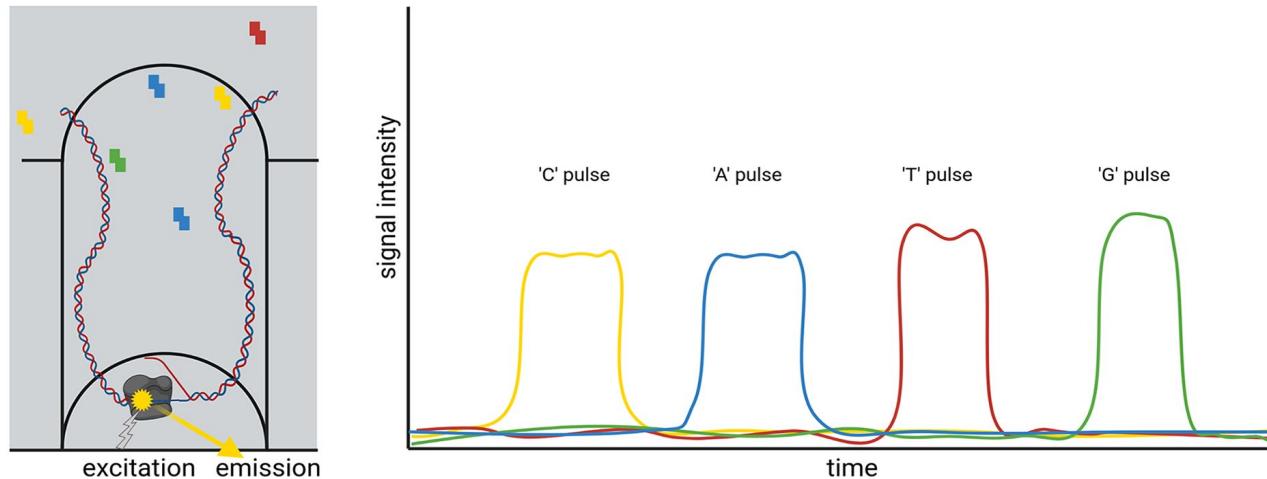


- Standard format for sequences with associated quality information
- Four lines per entry:
  - Header starts with @ (title + description)
  - Sequence
  - + (optional repetition of header)
  - Quality (phred encoded in ASCII character)
- Different versions exist that use different quality values offsets
- Example:

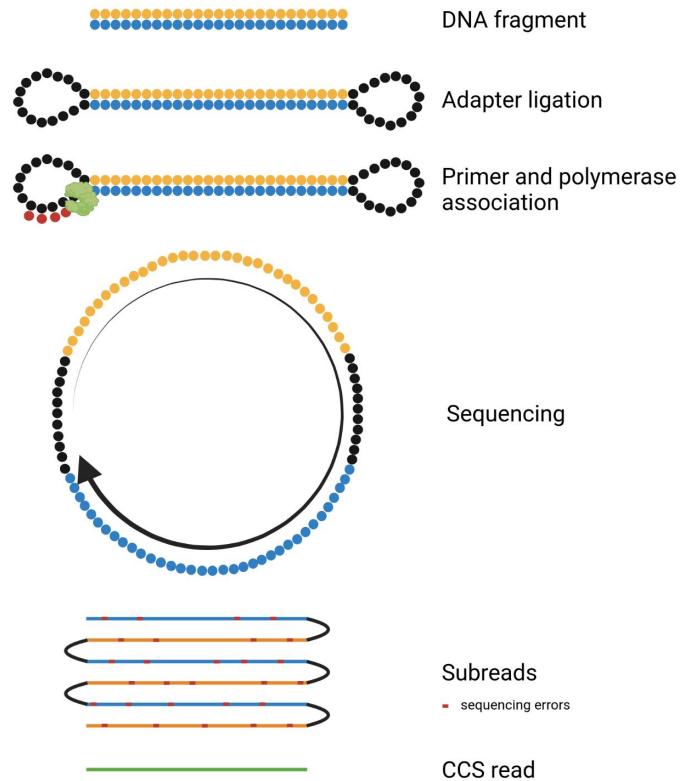
```
@seq1
ACGTACGTACGT
+
"""?CB"":DC"
```

# PacBio sequencing

- Polymerase located at bottom of well
- ZMWs = Zero Mode Waveguide



# PacBio - HiFi

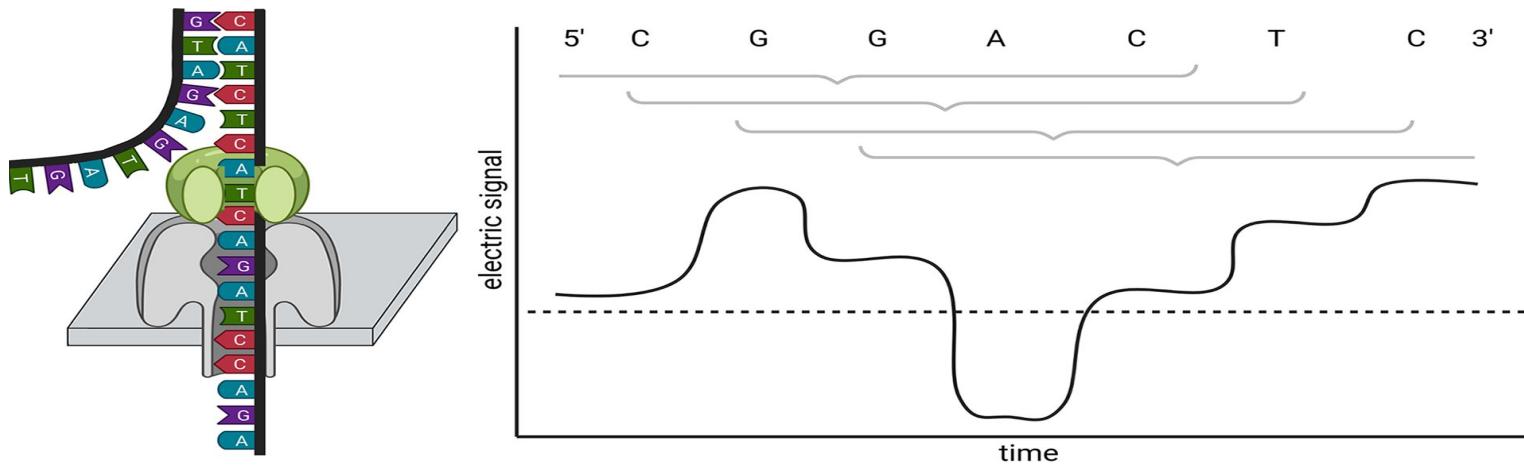


# ONT sequencing

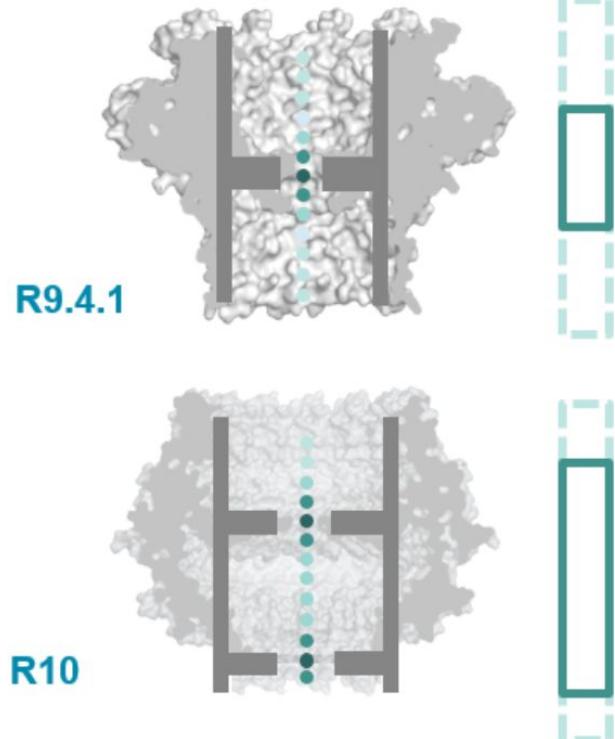
# Oxford Nanopore Technologies (ONT)

# Under development since the 1980s

“Analysis of anything, anywhere by anyone” ... not restricted to DNA sequencing



# Nanopores



ATCGGAAAAAAATCACGCCACGTCCAAA

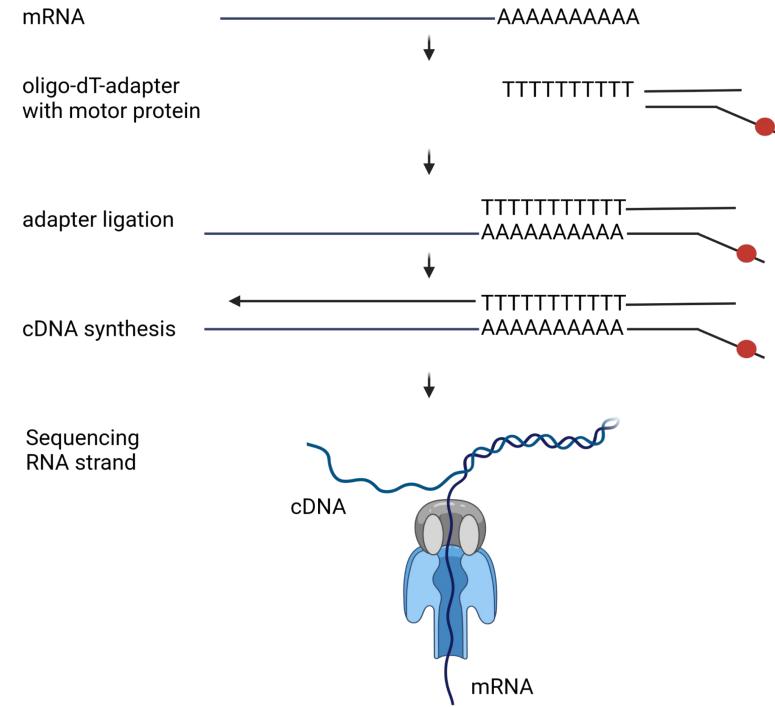


# ONT sequencing workflow

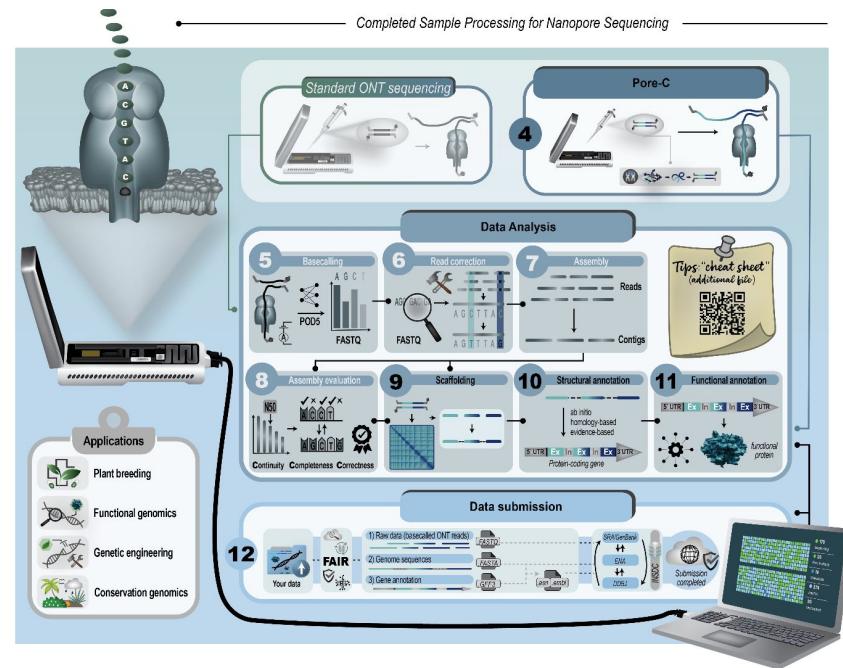
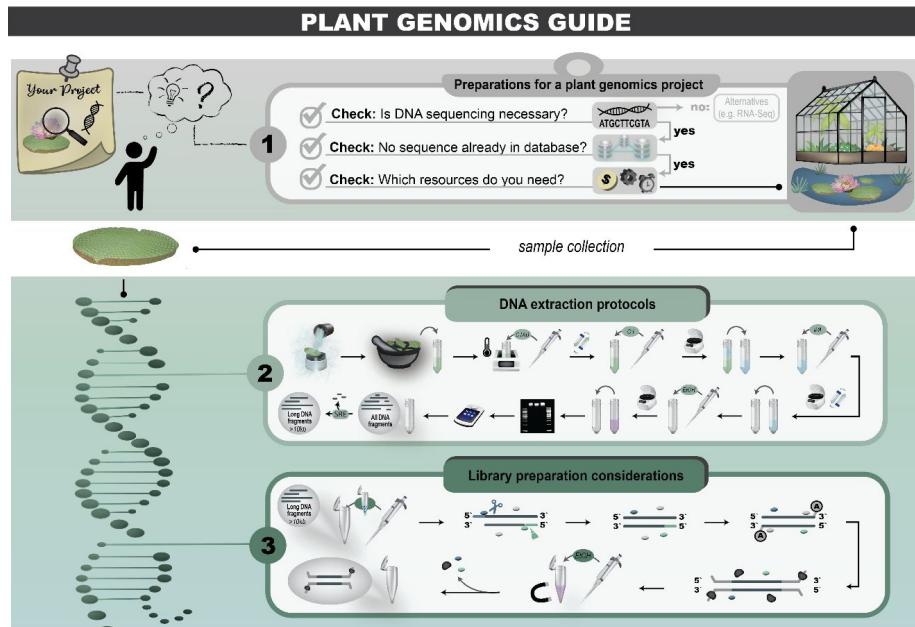
	task	consumed time	hands-on time	equipment	estimated costs of consumables	estimated costs of lab equipment
A		plant incubation in darkness ↓	2-3d	1h		
B		non-destructive sampling ↓	-	1h		
C		DNA extraction ↓	1d	8h	waterbath, centrifuge	\$50
D		quality control ↓	1h	1h	NanoDrop, Qubit	\$20
E		short fragment depletion ↓	2h	1h	centrifuge	\$50
F		quality control ↓	1h	1h	NanoDrop, Qubit	\$20
G		library preparation & sequencing ↓	1-5d	4-16h	centrifuge, magnetic rack, sequencer	\$3000
H		basecalling ↓	1d	1h	computer with GPU	\$250 \$1000
I		assembly ↓	1-15d	1h		\$3000
J		polishing ↓	1-5d	1h	compute cluster / cloud	
K		annotation ↓	1-5d	1h		
L		data submission	2h	2h	fast internet connection	

# Direct RNA sequencing

- Only sequencing technology to analyze RNA directly at high throughput
- RNA sequencing requires adjusted data processing
- Full length sequences of RNAs are generated



# Plant Genomics Workflow



**Supplementary file with all commands!**

- de Oliveira, J. A. V. S.; Choudhary, N.; Meckoni, S. N.; Nowak, M. S.; Hagedorn, M.; Pucker, B. (2025). Cookbook for Plant Genome Sequences. doi: [10.20944/preprints202508.1176.v2](https://doi.org/10.20944/preprints202508.1176.v2).
- Wolff, K.; Friedhoff, R.; Schwarzer, F.; Pucker, B. (2023). Data Literacy in Genome Research. *Journal of Integrative Bioinformatics*, 2023, pp. 20230033. doi: [10.1515/jib-2023-0033](https://doi.org/10.1515/jib-2023-0033).
- Pucker B, Irisarri I, de Vries J and Xu B (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quantitative Plant Biology*, 3, E5. doi: [10.1017/qpb.2021.18](https://doi.org/10.1017/qpb.2021.18).

# What should a publication presentation cover?

- Who are the authors?
- What is interesting about the species? What was the motivation?
- Which approach was used to sequence the plant genome?
- Quality of genome sequence?
- Data available?
- What is the purpose of the tool?
- Was proper benchmarking conducted?
- Presentation length: 10 minutes (strictly!)

## Technical considerations

- You can present with your own device
- Compatibility with HDMI needed (adpater?!)
- Upload of presentation into Sciebo/eCampus as PDF
- Presentation laptop can be used (only supports PDF)

- Friday, 5pm (c.t.) in SR2.007
- Up to 5 presentations per day
- 31st October
- 7th November
- 21st November
- 28th November
- 5th December

## Picking a topic

- Only one person should pick a topic/publication (be fast!)
- Please register your selected topic (publication) here:
  - only use your first name or initials (data protection!)
  - submit decision today before 10 pm
  - also send topic and preferred dates via email



<https://nuudel.digitalcourage.de/KoupJscZfTIK8ecf>

# Literature options I

- 1 = Hakim, S.E., Choudhary, N., Malhotra, K. et al. Phylogenomics and metabolic engineering reveal a conserved gene cluster in Solanaceae plants for withanolide biosynthesis. *Nat Commun* 16, 6367 (2025). doi: [10.1038/s41467-025-61686-1](https://doi.org/10.1038/s41467-025-61686-1).
- 2 = Karbstein K., Choudhary N., Xie T., Tomasello S., Wagner N. D., Barke B. H., Paetzold C., Bradican J. P., Preick M., Himmelbach A., Stein N., Papantonis A., Irisarri I., de Vries J., Pucker B., Hoerndl E. (2025). Assembling genomes of non-model plants: A case study with evolutionary insights from *Ranunculus* (Ranunculaceae). *Plant J.* 123: e70390. doi: [10.1111/tpj.70390](https://doi.org/10.1111/tpj.70390).
- 3 = Nowak M. S.\* , Harder B.\* , Meckoni S. N.\* , Friedhoff R., Wolff K., Pucker B. (2025). Genome sequence and RNA-seq analysis reveal genetic basis of flower coloration in the giant water lily *Victoria cruziana*. *bioRxiv* 2024.06.15.599162; doi: [10.1101/2024.06.15.599162](https://doi.org/10.1101/2024.06.15.599162).
- 4 = Lyu, Y., Sun, H., Yan, R. et al. Chromosome-level genome assembly of *Ulmus parvifolia* identifies genomic signatures and transcriptional profiling underlying Dutch elm disease resistance. *BMC Plant Biol* 25, 1130 (2025). <https://doi.org/10.1186/s12870-025-07100-w>.
- 5 = Wang, Y., Hu, H., Jin, J. et al. Comparative genomic analyses reveal different genetic basis of two types of fruit in Maloideae. *Nat Commun* 16, 7463 (2025). <https://doi.org/10.1038/s41467-025-62850-3>.
- 6= Zhang, RG., Zhao, H., Conover, J.L. et al. Reticulate allopolyploidy and subsequent dysploidy drive evolution and diversification in the cotton family. *Nat Commun* 16, 7480 (2025). <https://doi.org/10.1038/s41467-025-62644-7>.

## Literature options II

- 7 = Lv, Z., Wang, J., Zhou, L. et al. Near telomere-to-telomere genome assembly of *Camellia pitardii*. *Sci Data* 12, 1422 (2025). <https://doi.org/10.1038/s41597-025-05764-5>.
- 8 = Jesús Martínez-Gómez, Merly Escalona, Jack M Colicchio, Lauren N Hamm, Mohan P A Marimuthu, Oanh Nguyen, Noravit Chumchim, William Seligmann, Rachel S Meyer, Jason P Sexton, Benjamin K Blackman, The reference genome of a Sierra Nevada endemic, the cut-leaved monkeyflower, *Mimulus lacinatus* (syn. *Erythranthe lacinata*), *Journal of Heredity*, 2025;, esaf059, <https://doi.org/10.1093/jhered/esaf059>.
- 9 =Tian-Wen Xiao, Xin-Feng Wang, Zheng-Feng Wang, Hai-Fei Yan, Chromosome-scale genome assembly of *Sauvagesia rhodoleuca* (Ochnaceae) provides insights into its genome evolution and demographic history, *DNA Research*, Volume 32, Issue 5, October 2025, dsaf022, [https://doi.org/10.1093/dnaries/dsaf022](https://doi.org/10.1093/dnares/dsaf022).
- 10 = Luo, Z., Barrett, S.C.H., Tu, T., Zhao, Z., Jia, S., Gu, S., Duan, T., Zhang, Y., Xu, B., Gu, L., Deng, X., Jiang, L., Shi, M. and Zhang, D. (2025), Genetic architecture of the S-locus supergene revealed in a tetraploid distylous species. *New Phytol.* <https://doi.org/10.1111/nph.70521>.
- 11 = Wang, L., Jiang, X., Jiao, W. et al. Pangenome analysis provides insights into legume evolution and breeding. *Nat Genet* 57, 2052–2061 (2025). <https://doi.org/10.1038/s41588-025-02280-5>.
- 12 = Wang, D., Li, M., Yang, W. et al. Ancient allopolyploidy and specific subgenomic evolution drove the radiation of poplars and willows. *Nat Commun* 16, 6881 (2025). <https://doi.org/10.1038/s41467-025-62178-y>.

## Literature options III

- 13 = Zervakis, P.-I., Postel, Z., Losvik, A., Fracassetti, M., Solér, L., Proux-Wéra, E., Bunikis, I., Churcher, A. and Slotte, T. (2025), Genomic studies in Linum shed light on the evolution of the distyly supergene and the molecular basis of convergent floral evolution. *New Phytol*, 247: 2964-2981. <https://doi.org/10.1111/nph.70392>.
- 14 = Li, Z., Wang, J., Zhang, J. et al. Multi-omics profiling of *Nanhaia speciosa* and *Nanhaia fordii*: insights into lectin dynamics, nodulation, and triterpenoid saponin biosynthesis. *BMC Biol* 23, 169 (2025).  
<https://doi.org/10.1186/s12915-025-02278-9>.
- 15 = Wang, F., Jiang, Z., Gao, J. et al. Analysis of the *Rehmannia chingii* genome identifies RcCYP72H7 as an epoxidase in iridoid glycoside biosynthesis. *Nat Commun* 16, 6035 (2025). <https://doi.org/10.1038/s41467-025-60909-9>.
- 16 = He, W., Hu, D., Guo, M., Nie, B., Zhang, G., Jia, Y., Hou, Z., Shu, S., Shao, Y., Simonsen, H.T., Twamley, A., Li, C. and Wang, L. (2025), The telomere-to-telomere genome of *Sanicula chinensis* unveils genetic underpinnings of low furanocoumarin diversity and content in one basal lineage of Apiaceae. *Plant J*, 123: e70311.  
<https://doi.org/10.1111/tpj.70311>.
- 17 = Liwen Wu, Yuqing Cai, Chenggang Jiang, Xiang Shi, Shifa Xiong, Yicun Chen, Yangdong Wang, Uncovering the genetic basis for enhanced mushroom flavor in *Quercus fabri* through genome sequencing and metabolic profiling, *Horticulture Research*, Volume 12, Issue 9, September 2025, uhaf156, <https://doi.org/10.1093/hr/uhaf156>.
- 18 = Saei, A., Hunter, D., Hilario, E. et al. Chromosome-level genome assembly and annotation of *Petunia hybrida*. *Sci Data* 12, 1262 (2025). <https://doi.org/10.1038/s41597-025-05553-0>.

## Literature options IV

- 19 = Emiliano Mora-Carrera, Narjes Yousefi, Giacomo Potente, Rebecca Lynn Stubbs, Barbara Keller, Étienne Léveillé-Bourret, Stefan Grob, Ferhat Celep, Giorgi Tedoradze, Elena Conti, Genomic Patterns of Loss of Distyly and Polyploidization in Primroses, *Molecular Biology and Evolution*, Volume 42, Issue 8, August 2025, msaf162, <https://doi.org/10.1093/molbev/msaf162>.
- 20 = Paul B Frandsen, Abigail Borgmeier, Sam Bratsman, Brian J Cox, Sarah J Gottfredson, Robert Hadfield, Garrett Harding, Andrea L Kokkonen, Ying Fei Lin, Jackson Linde, Teagan Mulford, Andrew Parker, Shane Smith, Kaitlin Torres, Lauren Young, Hayley Mangelson, Eric N Jellen, Peter J Maughan, David E Jarvis, A chromosome-scale genome assembly and annotation of the tetraploid herb “epazote” (*Dysphania ambrosioides*), *G3 Genes|Genomes|Genetics*, 2025;, jkaf191, <https://doi.org/10.1093/g3journal/jkaf191>.
- 21 = Sielemann, K.\*, Pucker, B.\*., Orsini, E. et al. Genomic characterization of a nematode tolerance locus in sugar beet. *BMC Genomics* 24, 748 (2023). doi: 10.1186/s12864-023-09823-2.
- 22 = Feng, T.\*., Pucker, B.\*., Kuang, T.\*., Song B.\*., Yang Y., Lin N., Zhang H., Moore M.J., Brockington S.F., Wang Q., Deng T., Wang H. and Sun H. (2023). The genome of the glasshouse plant noble rhubarb (*Rheum nobile*) provides a window into alpine adaptation. *Commun Biol* 6, 706 (2023). doi: 10.1038/s42003-023-05044-1.
- 23 = Sielemann K., Schmidt N., Guzik J., Kalina N., Pucker B., Viehöver P., Breitenbach S., Weisshaar B., Heitkam T., Holtgräwe D. (2023). Pangenome of cultivated beet and crop wild relatives reveals parental relationships of a tetraploid wild beet. *bioRxiv* 2023.06.28.546919; doi: 10.1101/2023.06.28.546919.
- 24 = Ou, S., Su, W., Liao, Y. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* 20, 275 (2019). <https://doi.org/10.1186/s13059-019-1905-y>.
- 25 = Qian, J., Xue, H., Ou, S. et al. TEtrimmer: a tool to automate the manual curation of transposable elements. *Nat Commun* 16, 8429 (2025). <https://doi.org/10.1038/s41467-025-63889-y>.

# Time for questions!