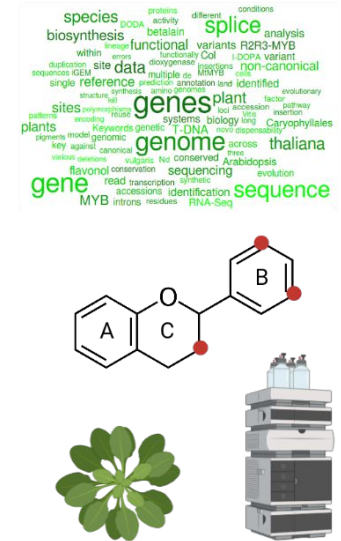
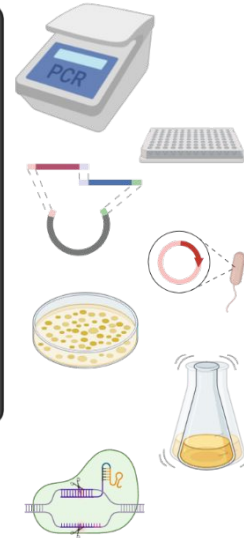
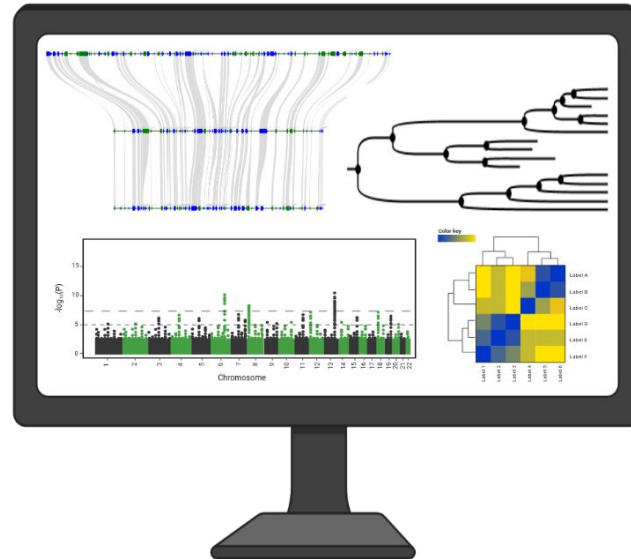
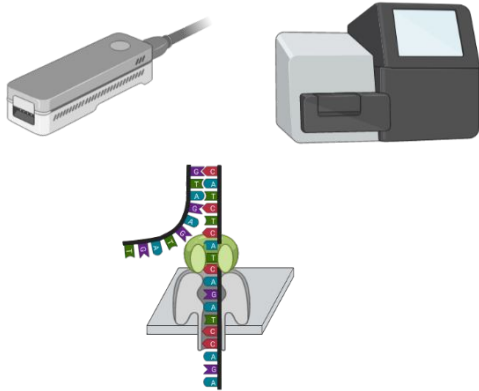




Technische  
Universität  
Braunschweig



Plant Biotechnology  
and Bioinformatics

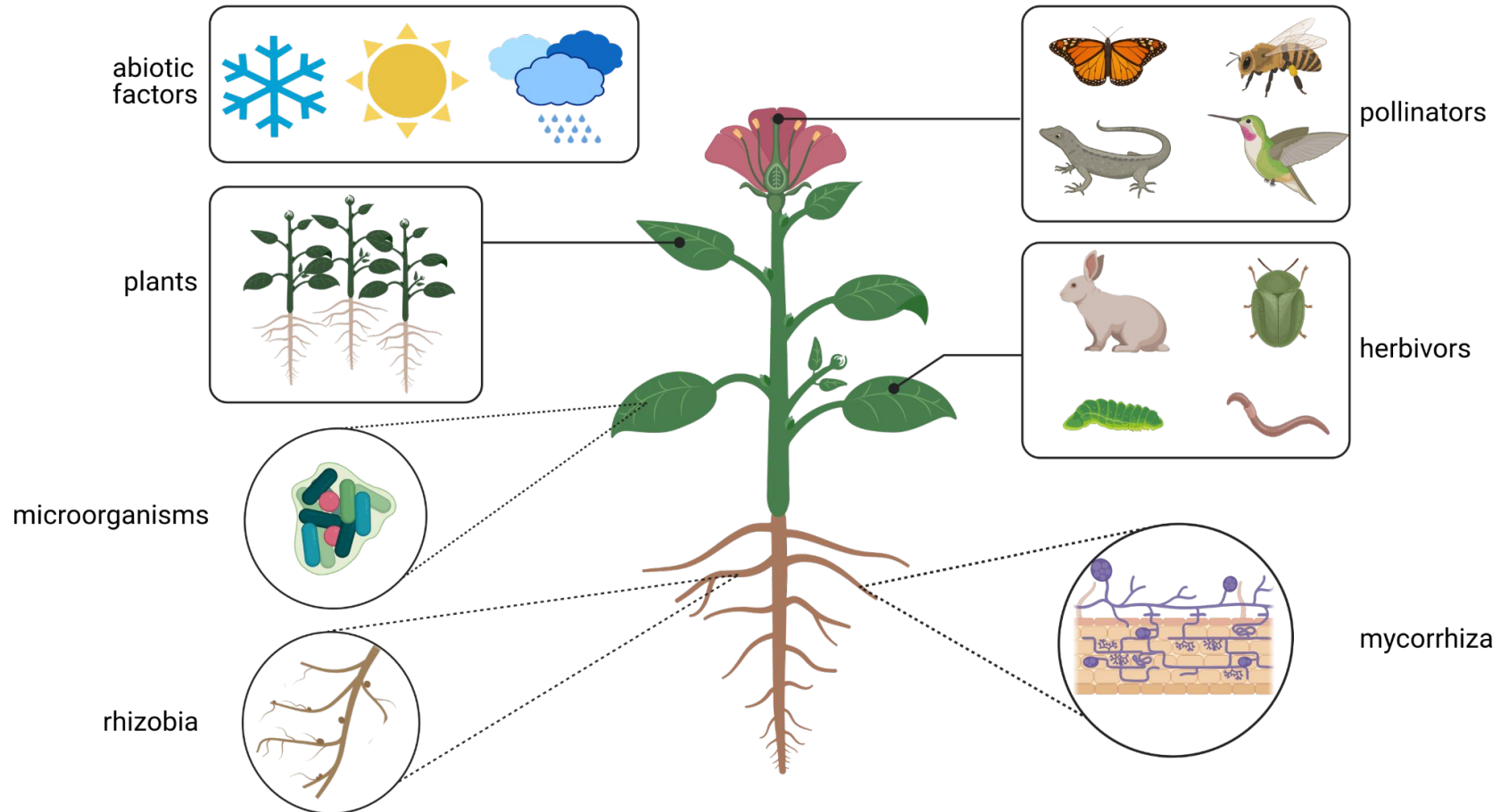
# Investigating plant specialized metabolism through genomics and applied bioinformatics

Boas Pucker (Plant Biotechnology and Bioinformatics)

# What are specialized plant metabolites?

- Definition:  
metabolites conferring evolutionary advantage,  
but not immediately required for survival
- *Papaver somniferum*: morphine (pain)
- *Taxus brevifolia*: paclitaxel (cancer)

# Why do plants produce specialized metabolites?



# Anthocyanins: colorful compounds with health benefits

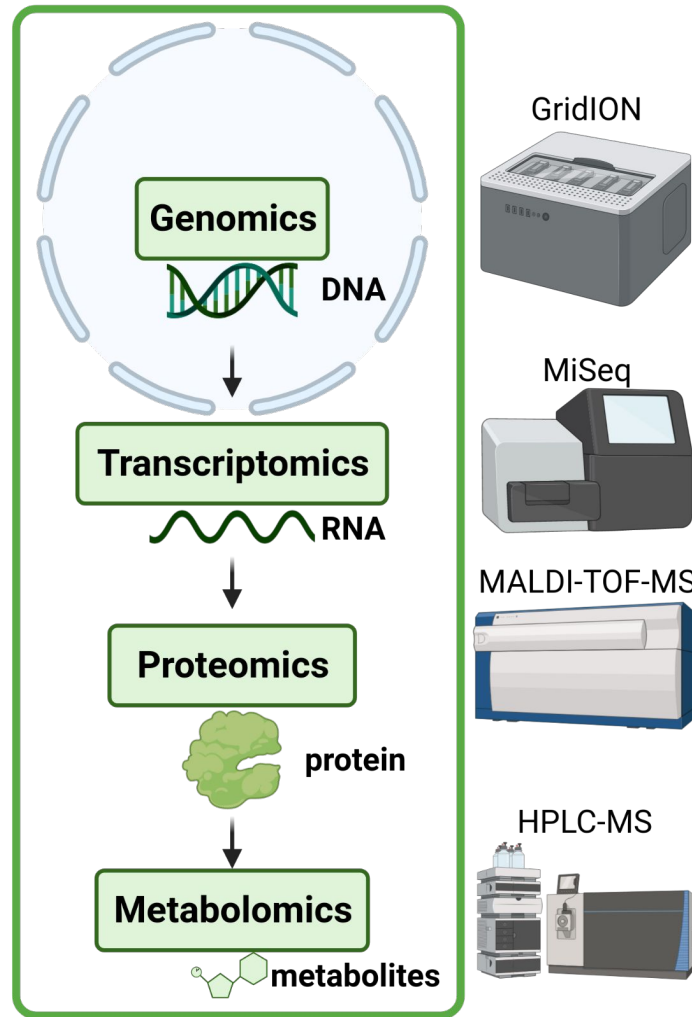
# How to understand a complex biosynthetic network?

data!

data!

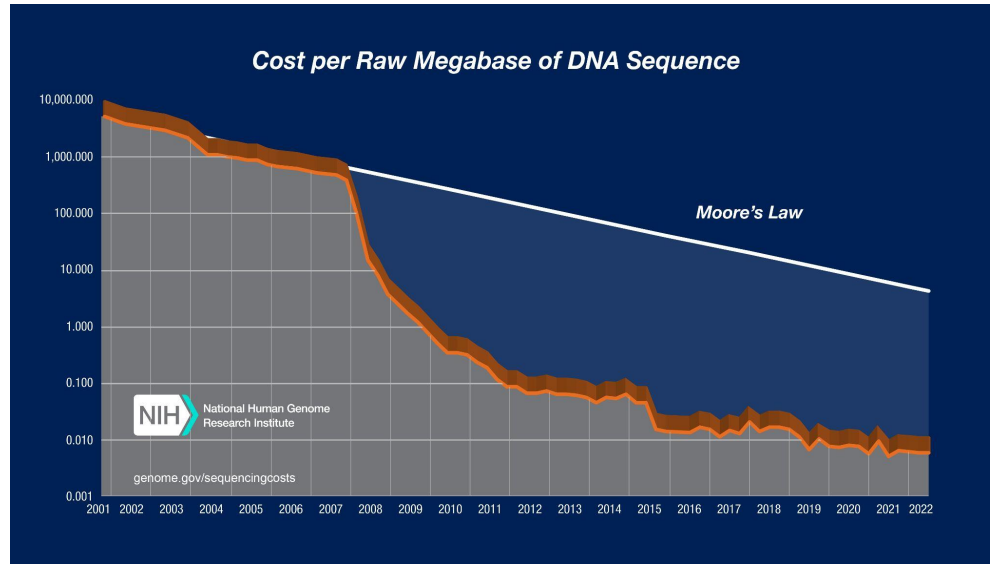
data!

data!

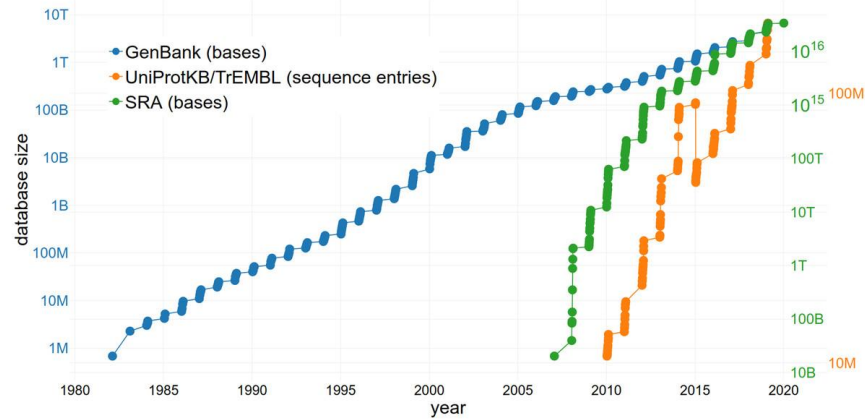


# Big data meets plant biology

Dropping costs  
for data  
generation



Increasing  
database sizes



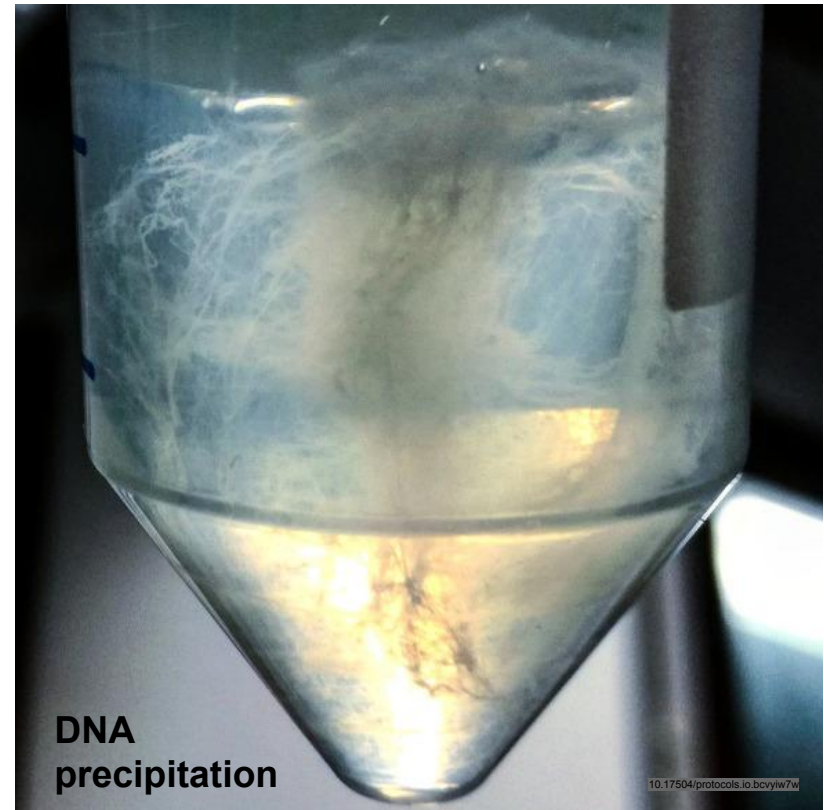
# How to deal with big data?

“We are drowning in information, while starving for wisdom.”

American biologist E. O. Wilson



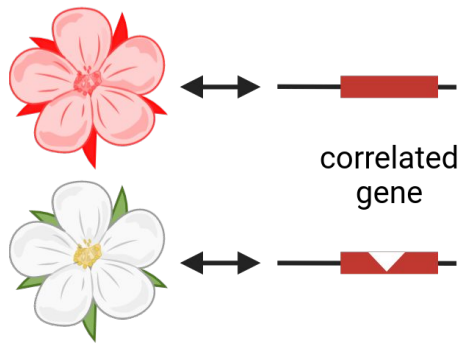
# Finding a gene is like finding a needle in a haystack



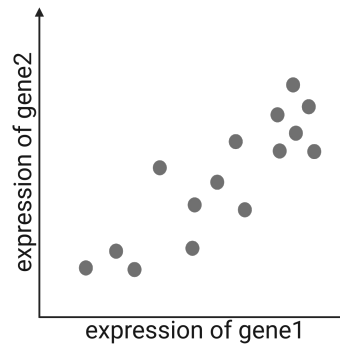


# Powerful methods for gene discovery & investigation

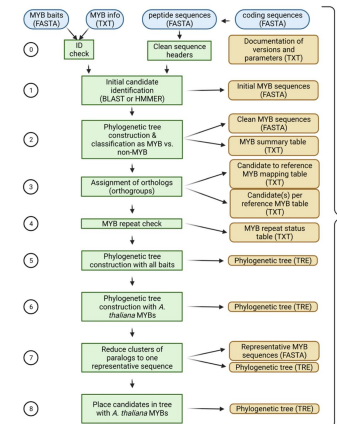
## Comparative genomics



## Coexpression



## Applied bioinformatics

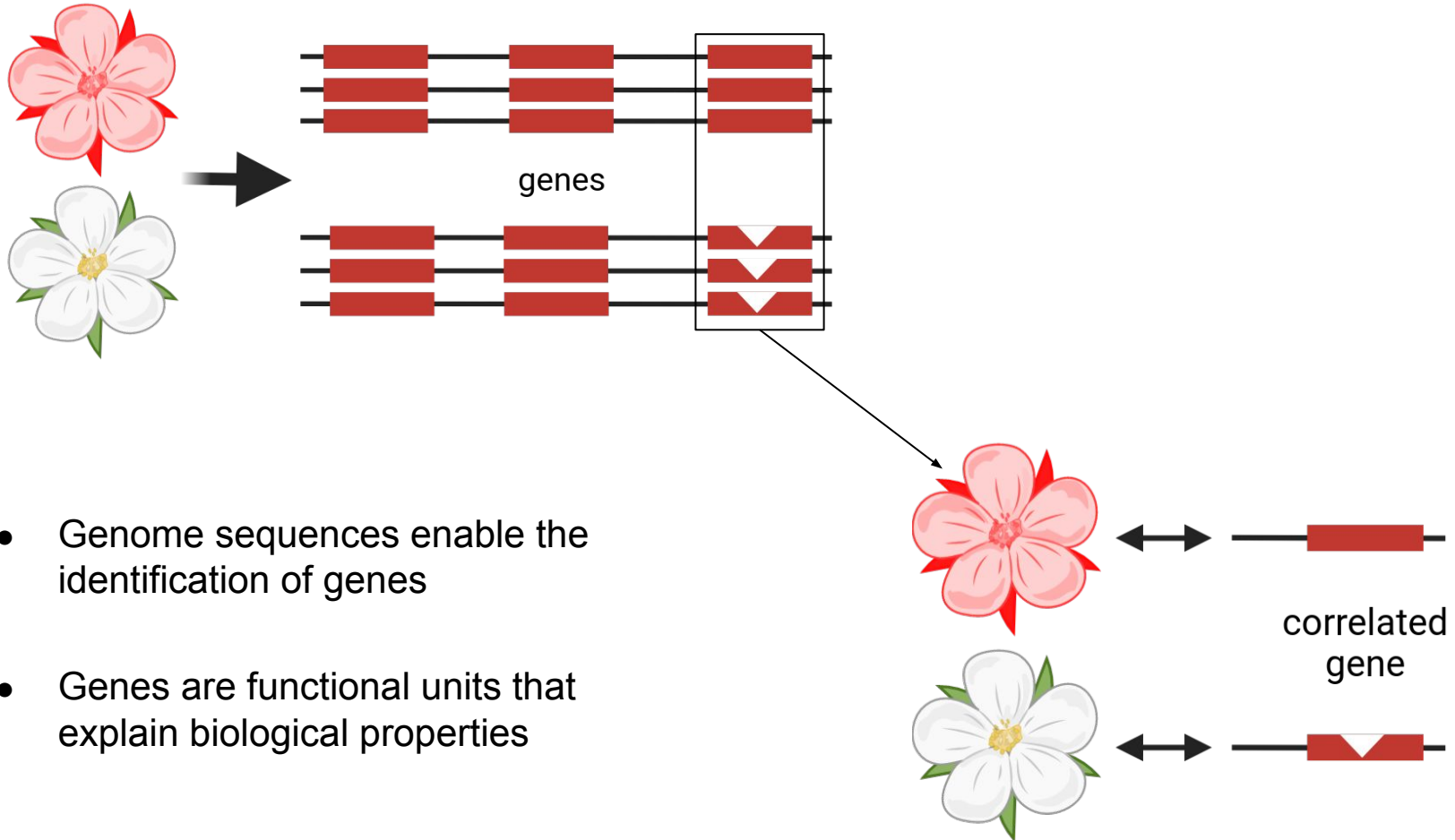


# Comparative genomics

# How does comparative genomics work?

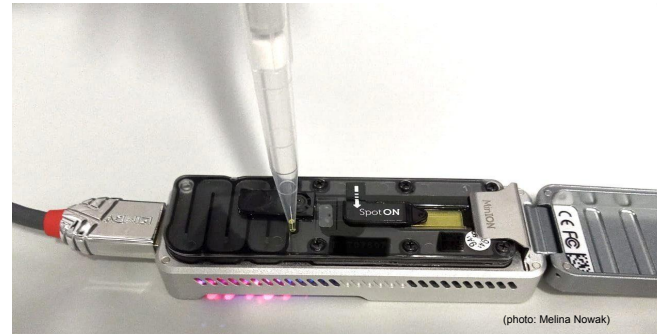
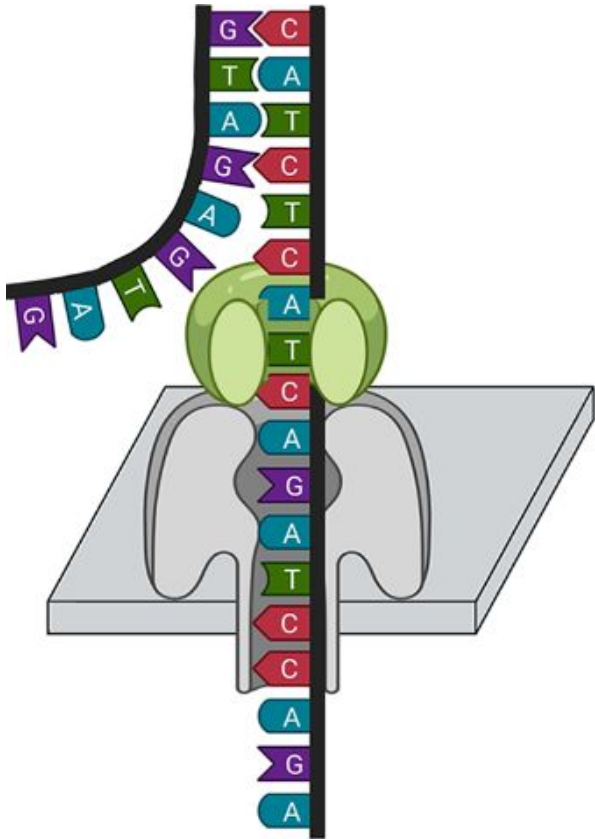
Observation

Genomic basis



- Genome sequences enable the identification of genes
- Genes are functional units that explain biological properties

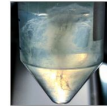
# Nanopore sequencing



# What does it take to sequence a plant genome?

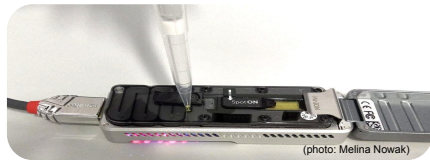


Fresh plant material without legal restrictions



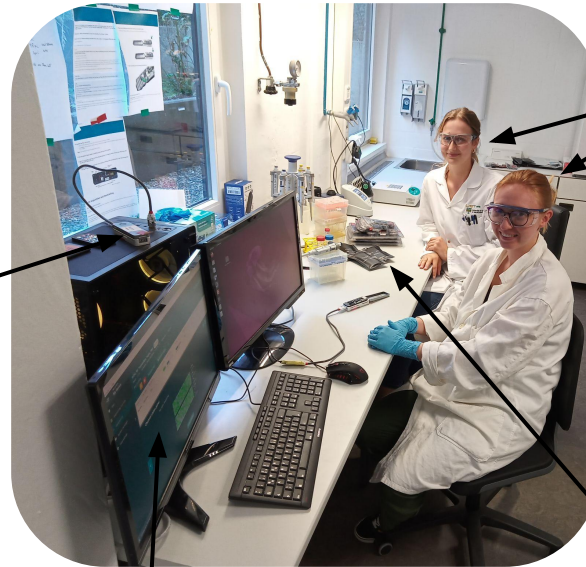
Plant DNA extraction and preparation for ONT sequencing  
DOI: [dx.doi.org/10.17504/protocols.io.bcvyiw7w](https://doi.org/10.17504/protocols.io.bcvyiw7w)  
Boas Pucker  
<sup>1</sup>Plant Biotechnology and Bioinformatics, Institute of Plant Biology, TU Braunschweig

Established DNA extraction protocol



(photo: Melina Nowak)

High tech devices



Highly skilled personnel



Sophisticated software



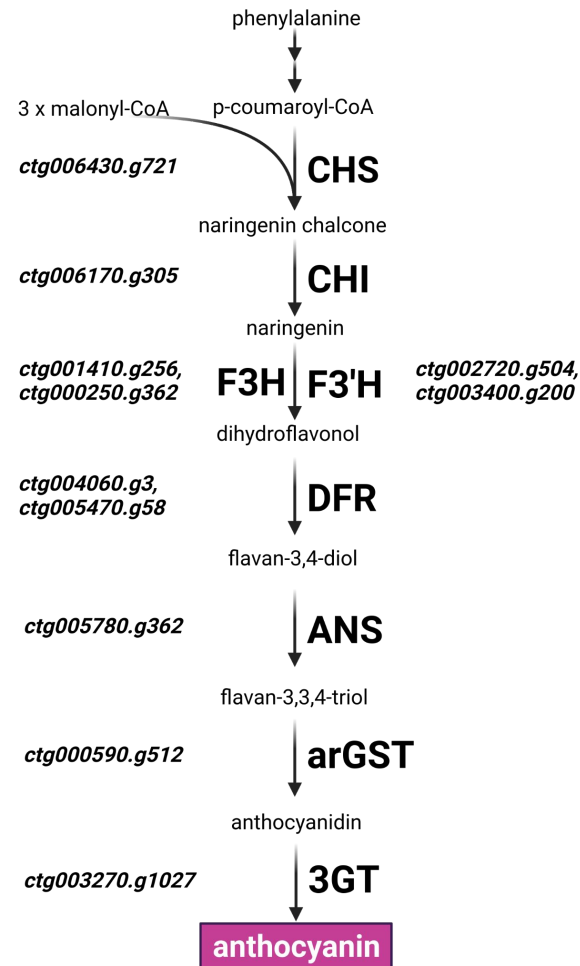
Extensive practice

# What causes pigmentation differences in common foxglove?

- Comparison of differently pigmented plants (pink vs. white)
- More pigmentation patterns



vs.



Katharina Wolff



Jakob Horz



Ronja Friedhoff

# Why is stinging nettle good for your health?

- Medicinal plant of the year 2022
- Many reported medicinal benefits
- Known for production of flavonoids (e.g. Rutin)



Katharina Wolff



Milan Borchert



Lena Fürstenberg



Bennet Garz

Naringenin chalcone

↓ Chalcone isomerase

Naringenin flavanone

↓ Flavanone 3-hydroxylase

Dihydrokaempferol

↓ Flavonoid 3'-hydroxylase

Dihydroquercetin

↓ Flavonol synthase

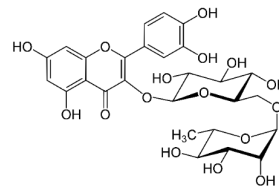
Quercetin

↓ UDP-glucose flavonoid  
3-O-glycosyltransferase

Isoquercitrin

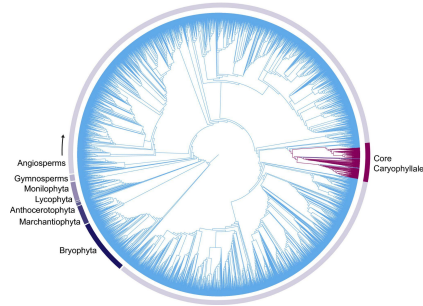
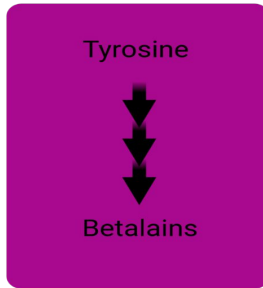
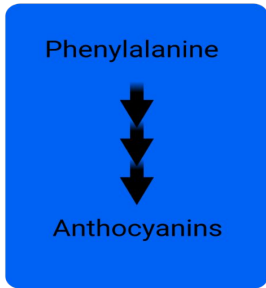
↓ flavonoid 3-O-glucoside  
L-rhamnosyltransferase

Rutin

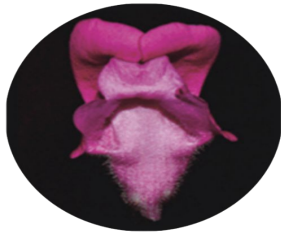


Collaboration with Till Beuerle

# Complex pigment evolution in the Caryophyllales



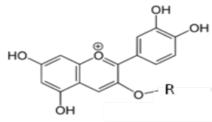
- Functional redundancy of both pigments
- At least four independent origins of betalain biosynthesis
- Mutual exclusion: anthocyanins and betalains were never observed in same (natural) plants



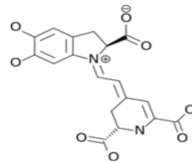
*Antirrhinum majus*



*Mirabilis jalapa*



Anthocyanin



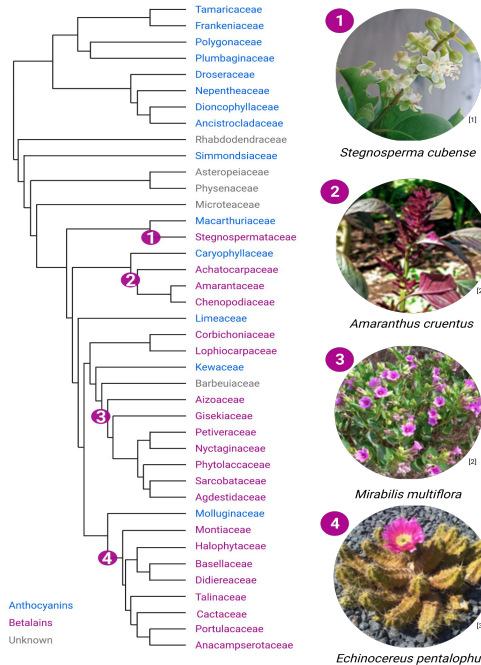
Betalain



Anthocyanin color range



Betalain color range



Brockington *et al.*, 2011: 10.1111/j.1469-8137.2011.03687.x

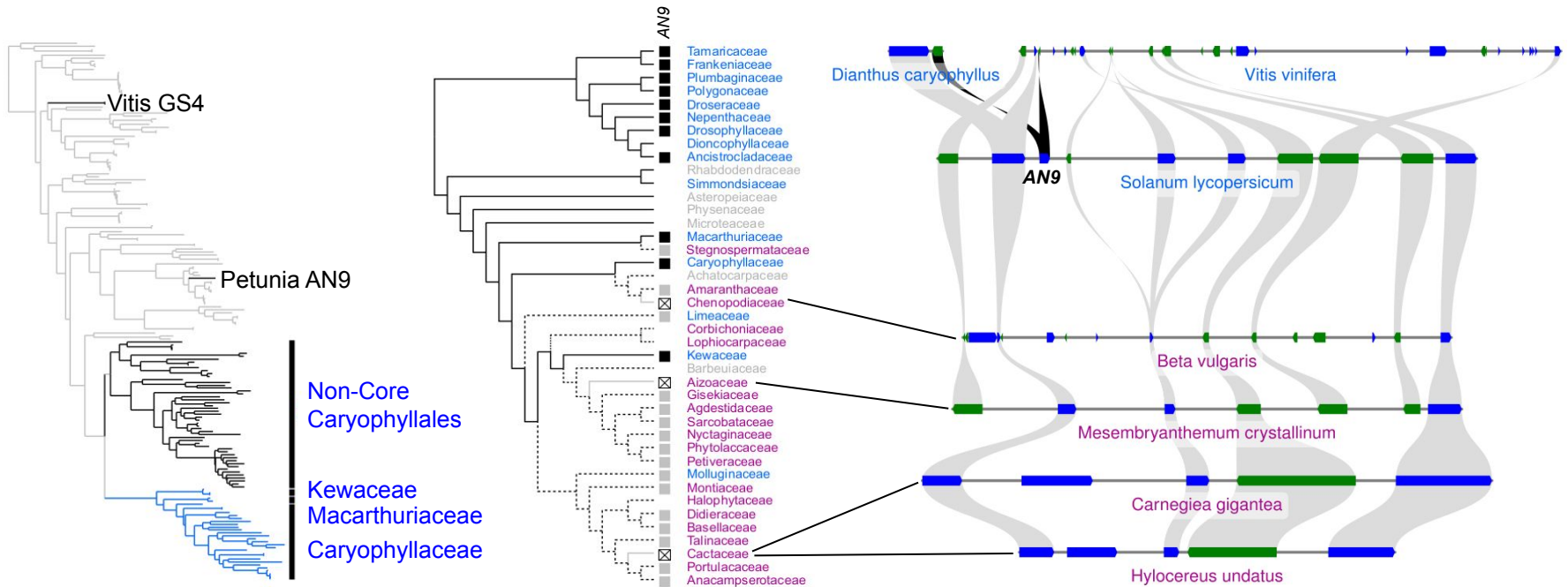
Timonedra *et al.*, 2019: 10.1111/nph.15980

Sheehan *et al.*, 2020: 10.1111/nph.16089

[1] Dick Culbert [2] Stan Shebs [3] Emöke Denes

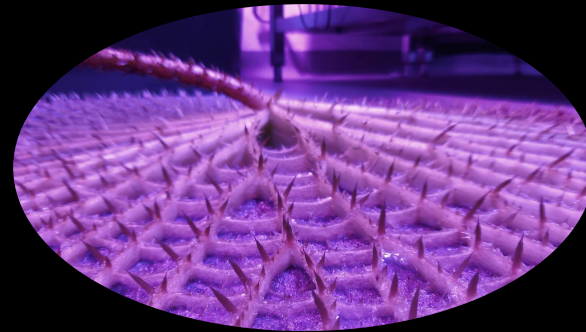


# Synteny analysis supports gene loss



# What causes flower color change in *Victoria cruziana*?

- Well known for huge floating leaves
- Flowers at night: white in first night and pink in second night



Melina Nowak



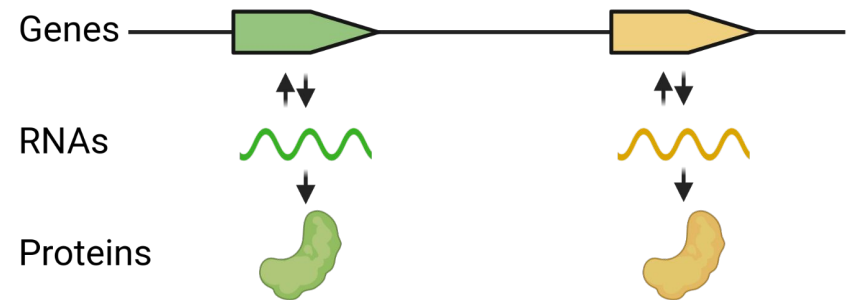
Benjamin Harder



Ronja Friedhoff

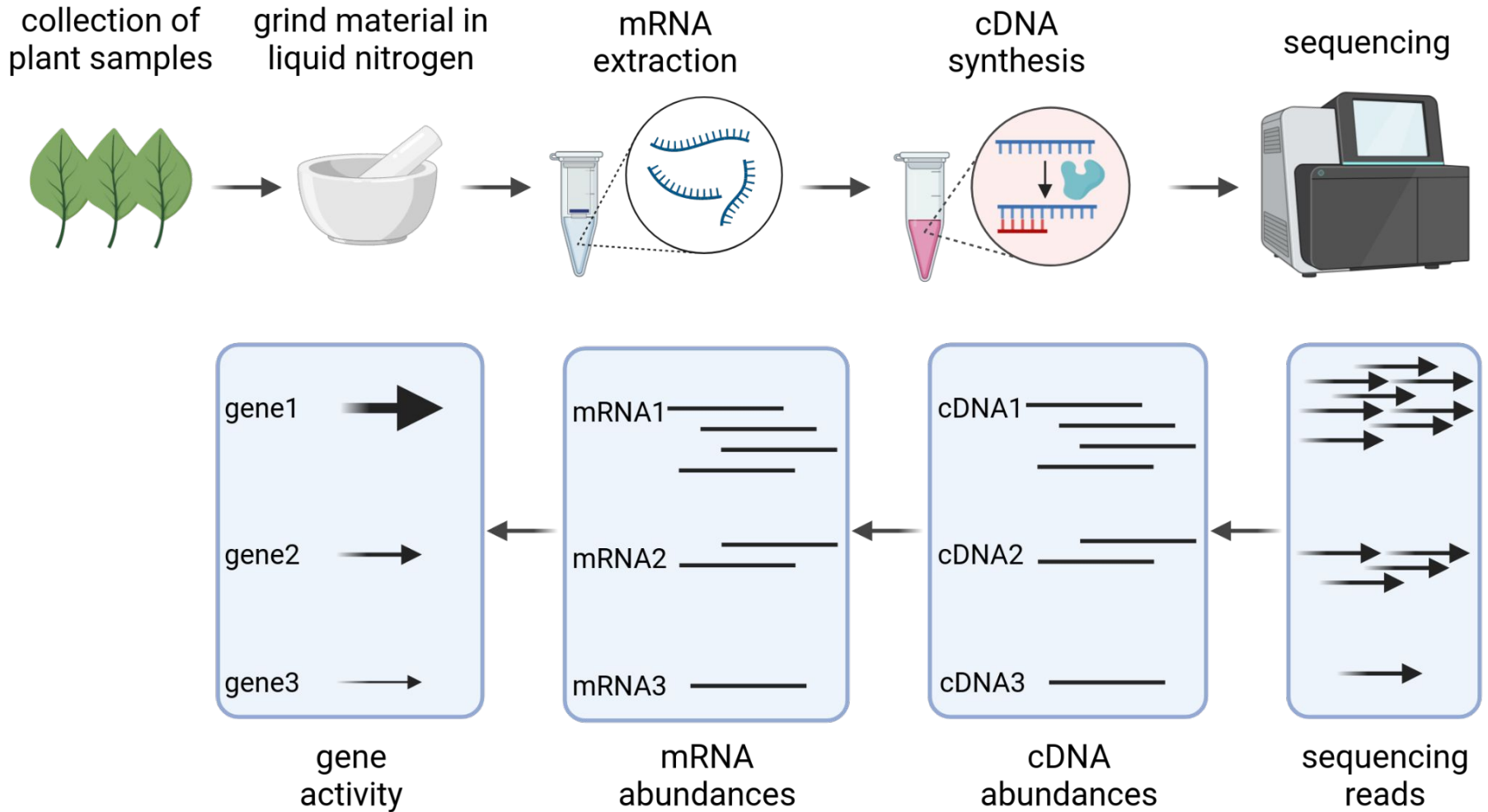
# Genes and gene activity

- Genomes are static
- Gene repertoire determines potential of an individual
- Expression of genes is biologically important for properties of the individual
- Gene expression = formation of the gene product (RNA, protein)



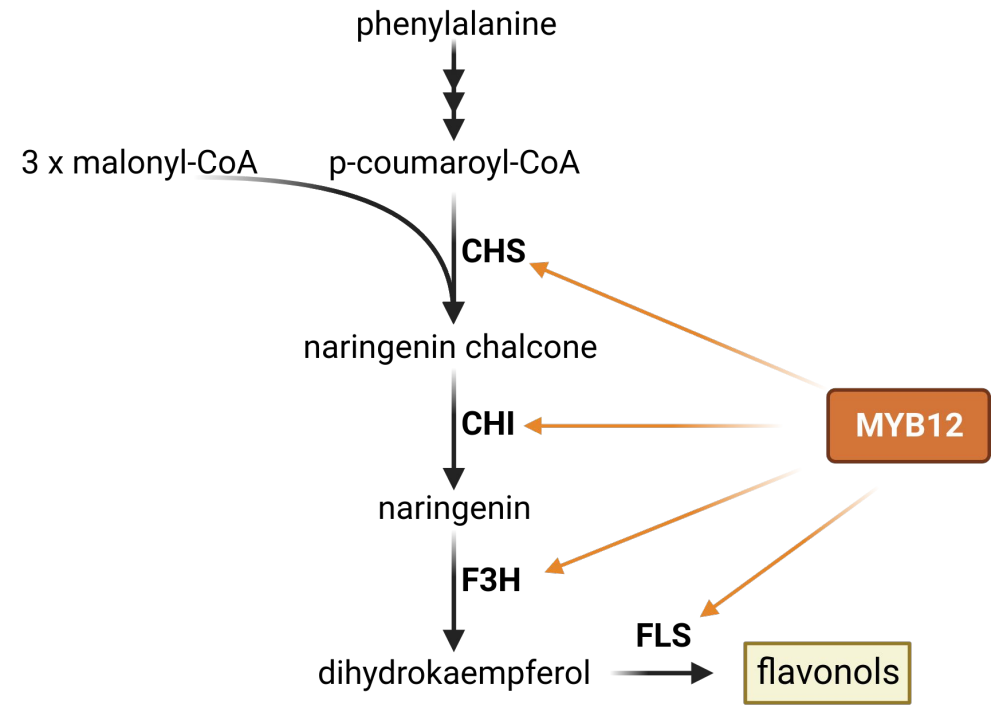
# Coexpression

# How to systematically measure activity of all genes?

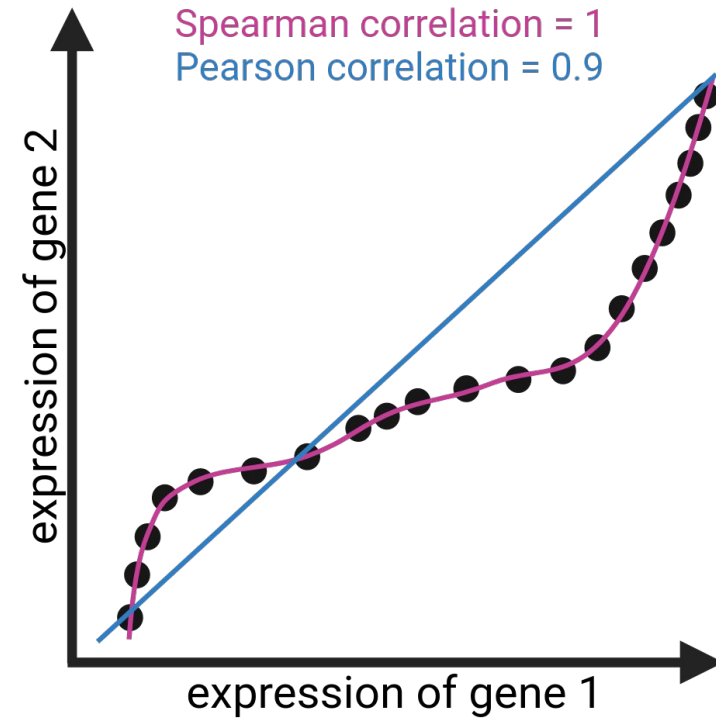
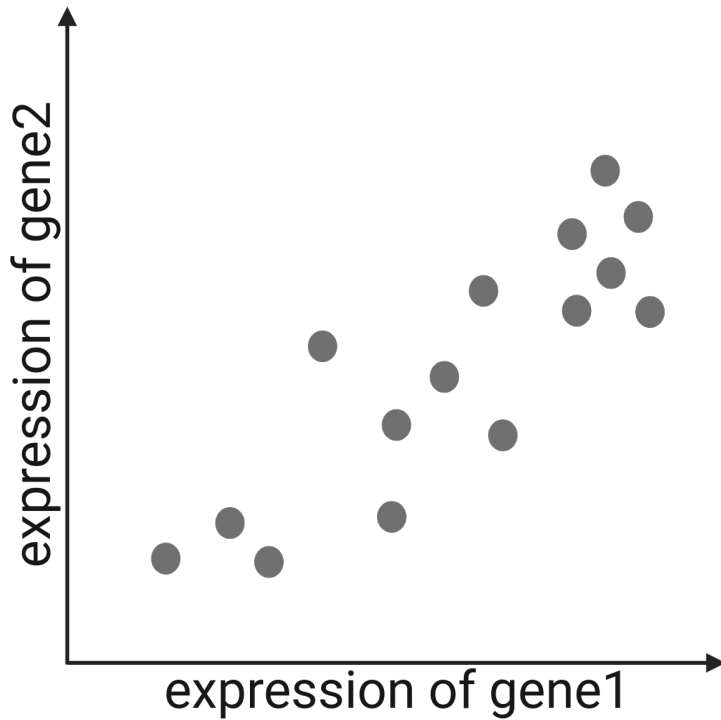


# Why do we expect coexpression?

- Genes in a pathway need to be active at the same time
- Identification of similar expression patterns can reveal connected genes
- One bait gene can help to find an entire biosynthesis pathway



# How can we analyze coexpression (pairwise)?



# Big transcriptomic data meets plant biology

- Large number of RNA-seq data sets available through databases
- Enormous potential for reuse of public data
- Downloading and processing of large data sets

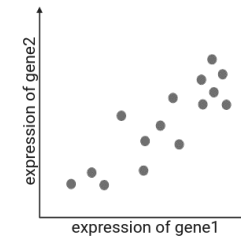


download data  
from SRA

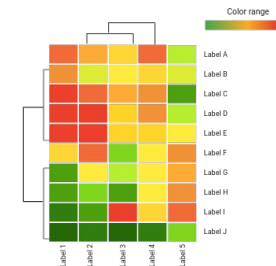
	s1	s2	s3	s4
g1	12	3	4	9
g2	32	3	0	5
g3	17	5	4	1
g4	41	2	0	9
g5	13	3	7	0

generation of  
count tables

co-expression  
analysis



construction  
of heatmaps



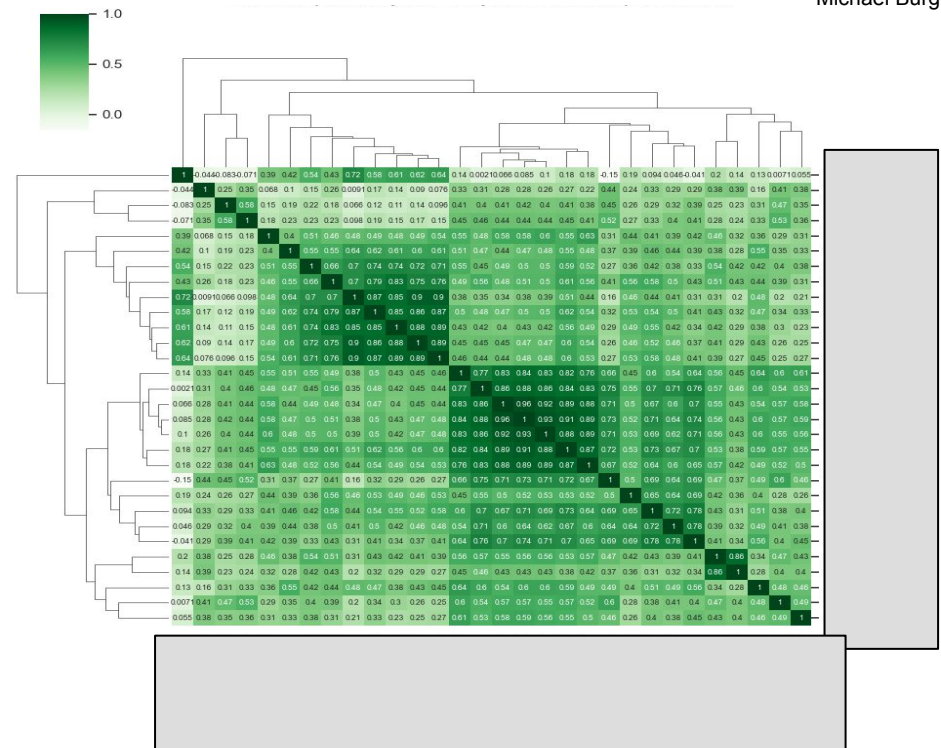


# How to find novel anthocyanin-associated genes with data?

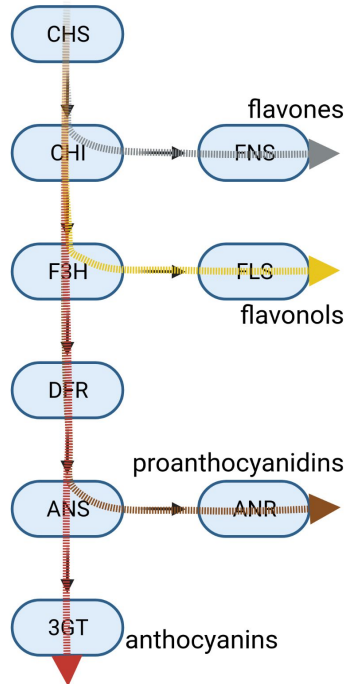
- Co-expression with anthocyanin biosynthesis genes as hints
- Anthocyanin-associated sugar transporter
- Additional modification enzymes



Michael Burgis



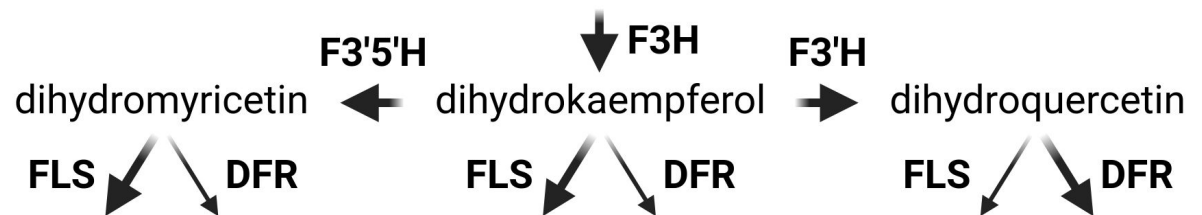
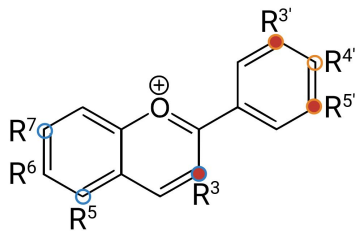
# Competing branches of the flavonoid biosynthesis



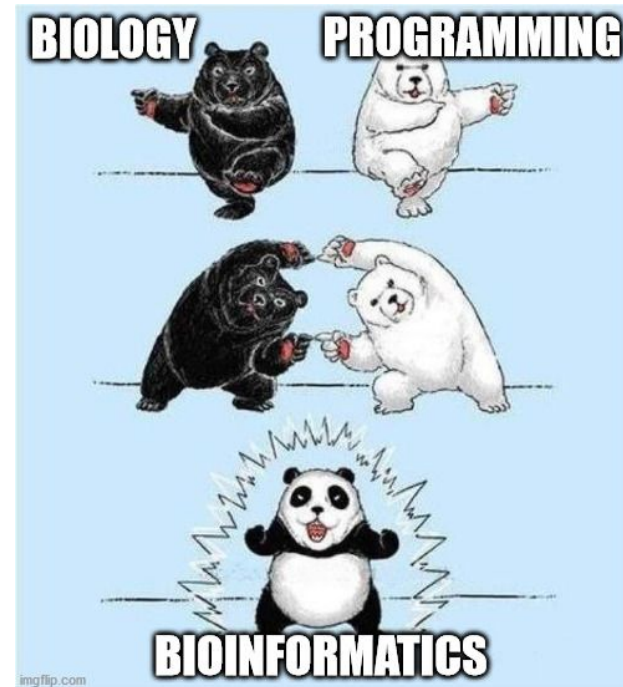
- Flavonol and anthocyanin biosynthesis are competing pathways
- FLS and DFR are crucial enzymes
- Hydroxylation pattern and substrate preferences might play a role
- Coexpression of genes can reveal activity of specific routes



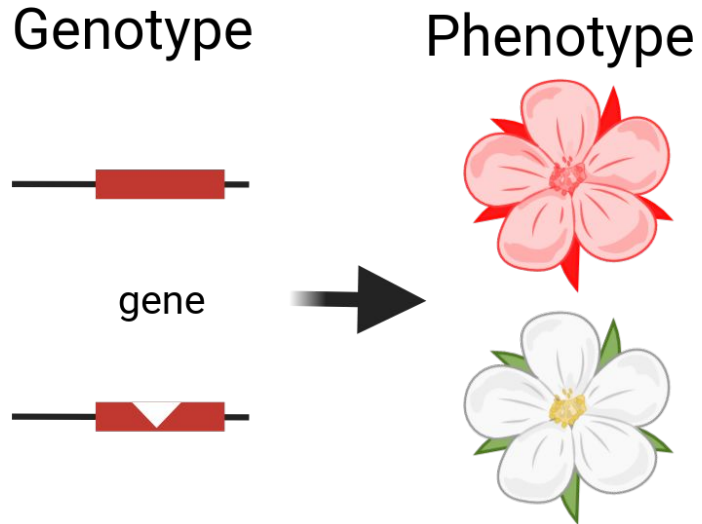
Nancy Choudhary



# Applied bioinformatics



# How can we identify a gene function?



# Studying knock-outs is not scalable

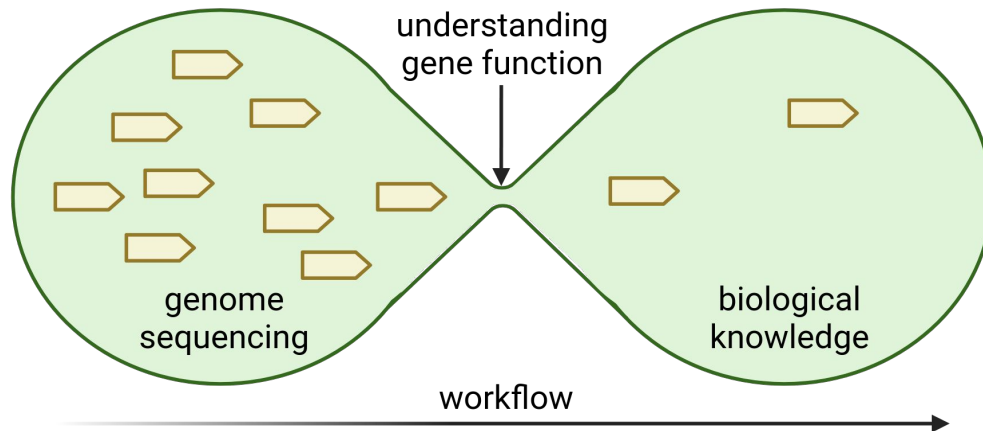
- Thousands of genes per plant species
- Only minority of *Arabidopsis thaliana* genes studied
- Not all species are genetically accessible



*Arabidopsis thaliana*

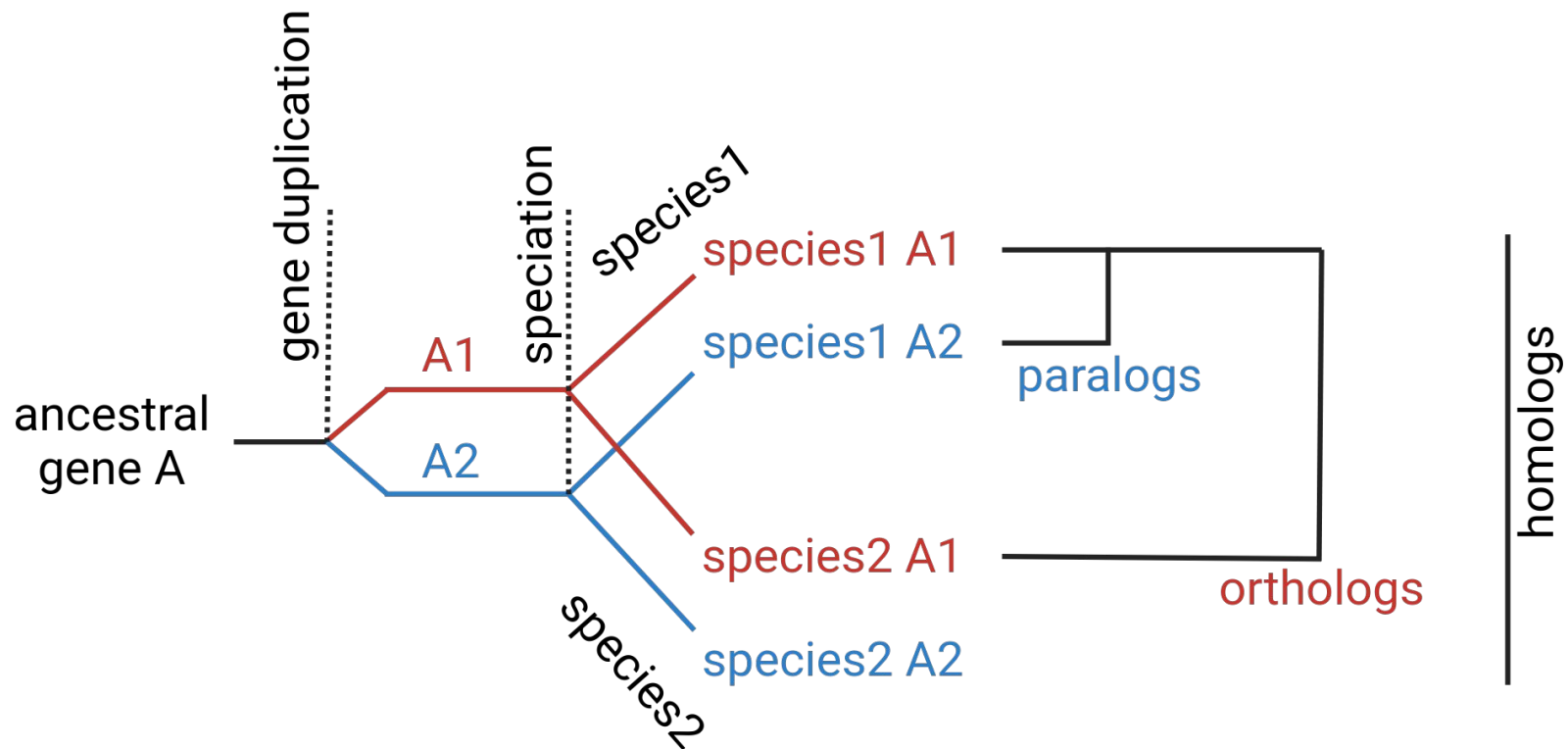


*Sequoiadendron giganteum*



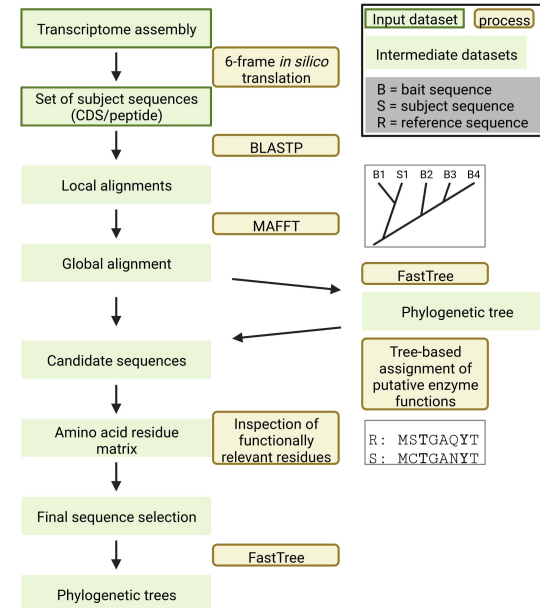
# Orthology-based annotation

- Orthologs are generally assumed to have the same function



# How can we annotate all genes in a pathway?

- KIPes = Knowledge-based Identification of Pathway Enzymes
- Developed for the annotation of the flavonoid biosynthesis
- Functionally relevant amino acid residues known
- Effective and automatic screening of large data sets possible
- Systematic comparisons of flavonoid biosynthesis across species



Andreas Rempel



Nancy Choudhary

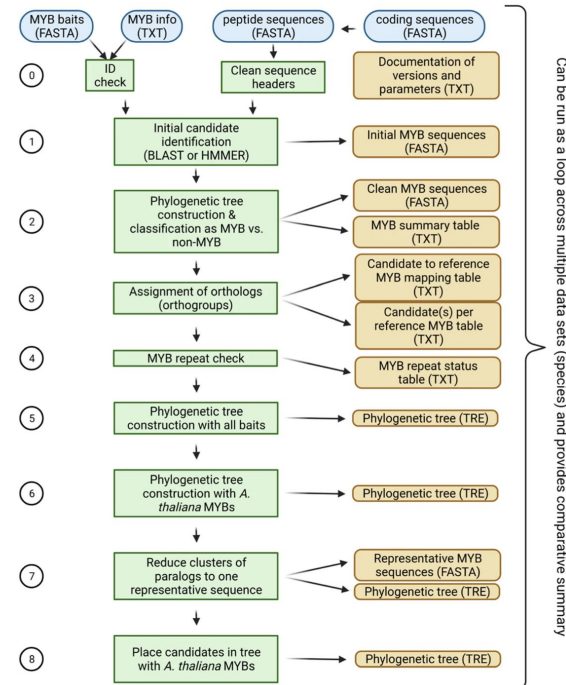


Hanna Marie Schilbert



# How can we find corresponding transcription factors?

- MYB and bHLH transcription factors are the largest gene families in plants
- Automatic identification and annotation
- Enables high-throughput comparative genomic analyses
- Concept generalized to allow annotation of other (transcription factor) gene families



Corinna Thoben





# Running a web server

- Web server enables life scientists without bioinformatic skills to run analyses
- Code for running the server is freely available

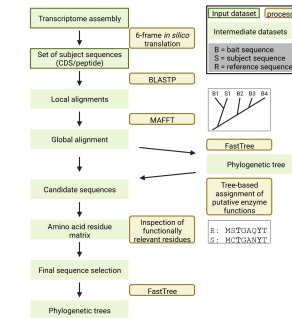
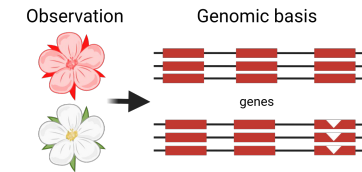
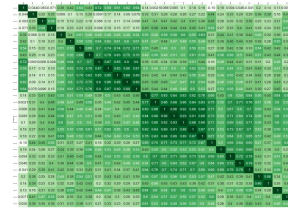
## BioInfToolServer

  
**bHLH\_annotator**  
automatic annotation of bHLH gene family in plants  
**CoExp**  
Co-Expression analysis tool  
**KIPes**  
Knowledge-based Identification of Pathway Enzymes (KIPes) performs an automatic annotation of the flavonoid biosynthesis steps in a new transcriptome of genome sequence assembly.  
**MYB\_annotator**  
This tool performs an automatic identification, annotation, and analysis of the MYB gene family in plants. It can be applied to new transcriptome of genome assemblies.

<https://pbb-tools.de/>

# Summary

- Plants are excellent chemists with a broad range of products
- Long read sequencing makes plant genomes accessible (functional & comparative genomics)
- Public RNA-seq data sets enable systematic analyses of gene activity (coexpression)
- Applied bioinformatics and data reuse are extremely powerful



# Acknowledgements

Plant Biotechnology and Bioinformatics

IfP & Botanical Garden: Thorsten Marschall, Michael Kraft

DSMZ: Stephan Winter, Samar Sheat & team

Supportive colleagues at TU Braunschweig

Previous colleagues:

Genetics & Genomics of Plants at CeBitec and Bielefeld University

Molecular Genetics and Physiology of Plants, Ruhr-University-Bochum

Evolution and Diversity, Plant Sciences, University of Cambridge

