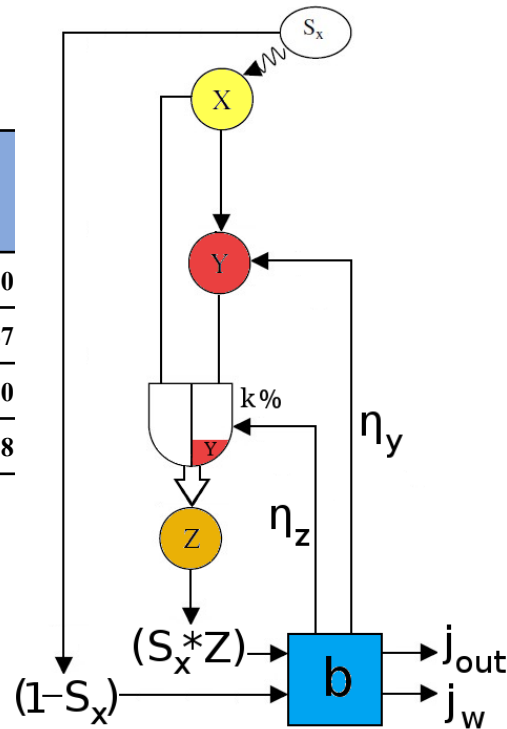
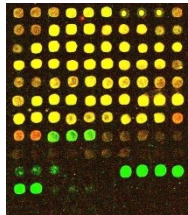
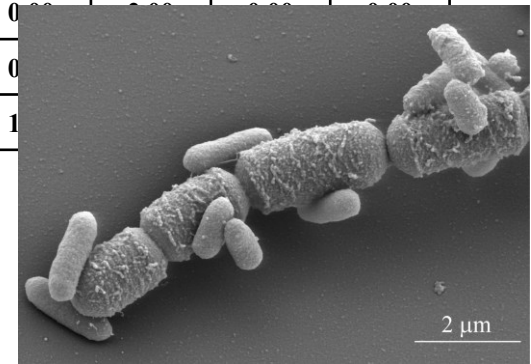


|   | 1    | 2    | 3    | 4    | 5    |
|---|------|------|------|------|------|
| A | 0.00 | 0.00 | 0.00 | 0.44 | 0.00 |
| C | 0.14 | 0.00 | 0.00 | 0.13 | 0.67 |
| G | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 |
| T | 1.42 | 2.00 | 0.00 | 0.13 | 0.38 |



| 10   | 11   | 12   | 13   | 14   |
|------|------|------|------|------|
| 0.67 | 0.14 | 0.00 | 2.00 | 1.08 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.38 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |



## Gene regulatory networks in bacteria

Dieter Jahn, Braunschweig Integrated Center of Systems Biology, TU Braunschweig  
[d.jahn@tu-bs.de](mailto:d.jahn@tu-bs.de), <https://www.tu-braunschweig.de/forschung/zentren/brics>

30 years TRANSFAC, 3. March 2018

.....it all started in Marburg..



Marburg  
1985

## ....first steps towards TRANSFAC



Marburg, 1985/86

# ....Transcription....*in vitro* transcription...

THE JOURNAL OF BIOLOGICAL CHEMISTRY  
© 1986 by The American Society of Biological Chemists, Inc.

Vol. 261, No. 3, Issue of January 25, pp. 1409-1413, 1986  
Printed in U.S.A.

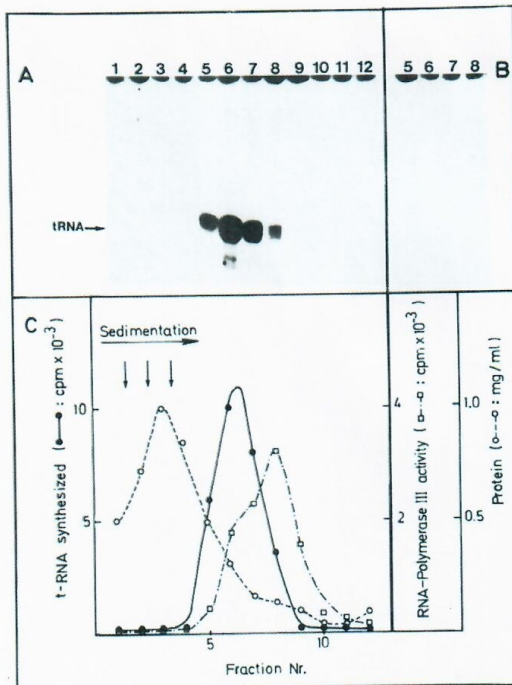
## Association of RNA Polymerase III with Transcription Factors in the Absence of DNA\*

(Received for publication, July 29, 1985)

Edgar Wingender, Dieter Jahn, and Klaus H. Seifart

From the Institut für Physiologische Chemie I, D-3550 Marburg/Lahnberge, Federal Republic of Germany

29. July 1986

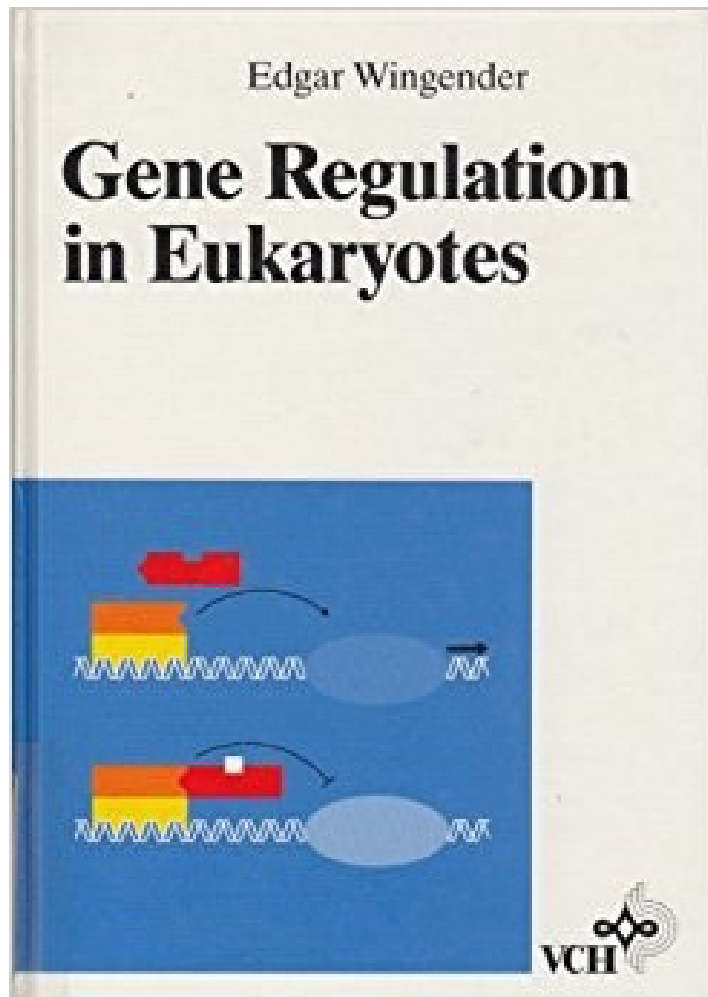


# .....our new partner: ATARI 520+



Made: 1985  
Chip: Motorola 68000 8MHz  
Memory: 512 kB  
System: TOS 1x

....“Atari 520+“ plus „Transcription“ plus „Edgar“ =



1993  
430 pages

# .....2000ff....the dancing genes – Inter-genomics..

INTERGENOMICS - Bioinformatics Competence Center Braunschweig



## INTERGENOMICS

Towards a bioinformatic framework for “InterSystems Biology”

### INTERGENOMICS – Structure of the Competence Center

#### Project Coordination:

**Prof. E. Wingender**  
(UKG / GBF)

#### Project Management:

**Dr. R.-J. Müller**  
(GBF)

#### Scientific Advisory Board (external)

Dr. R. Apweiler (EBI / Hinxton, UK)  
Prof. R. Hofstaedt (University of Bielefeld)  
Prof. D. Schomburg (Cologne University)  
Dr. T. Werner (Genomatix Software GmbH, Munich)

#### Partners

• GBF - German Research Centre for Biotechnology  
• University of Goettingen UKG

6 Groups  
Coordination:  
**Dr. H. Blöcker**

• Technical University of Braunschweig  
3 Groups  
(+2 being set up)

Coordination:  
**Prof. D. Jahn**

• University of Applied Sciences Braunschweig / Wolfenbuettel

1 Group

Coordination:  
**Prof. U. Klages**

• Biobase GmbH  
1 Group

Coordination:  
**Dr. H. Saxel**



Hans-Dieter Ehrich



Ulrich Klages



Reinhard Hehl



Frank Klawonn



Technische  
Universität  
Braunschweig

Dieter Jahn | 7. March 2018 | 30 years TRANSFAC | Seite 7



# Intergenomics: Our start into bioinformatics



Muench, R., Hiller, K., Barg, H., Heldt, D., Linz, S., **Wingender, E.** & Jahn, D. (2003)  
PRODORIC: Prokaryotic Database of Gene Regulation.

*Nucleic Acids Res.* **31**, 266-269

Grote, A., Klein, J., Retter, I., Haddad, I., Behling, S., Bunk, B., Biegler, I., Yarmolinetz, S., Jahn, D. & Muench, R. (2009) PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes. *Nucleic Acids Res.* **D61-5**

Hiller, K., Schobert, M., Hundertmark, C., Jahn, D. Muench, R. (2003)

JVirGel: Calculation of virtual two-dimensional protein gels.

*Nucleic Acids Res.* **31**, 3862-3865

Hiller, K., Grote, A., Maneck, M., Muench, R. & Jahn, D. (2006)

JVirGel 2.0: computational prediction of proteomes separated via two dimensional gel electrophoresis under consideration of membrane and secreted proteins.

*Bioinformatics* **22**, 2441-244

Hiller, K., Grote, A., Scheer, M., Muench, R. & Jahn, D. (2004)

PrediSi: prediction of signal peptides and their cleavage positions.

*Nucleic Acids Res.* **32**, 375-379

Grote, A., Hiller, K., Scheer, M., Muench, R., Noertemann, B., Hempel, D.C. & Jahn D. (2005)

JCat: a novel tool to adapt codon usage of a target gene to its potential expression host.

*Nucleic Acids Res.* **33**, 526-531

Scheer, M., Klawonn, F., Muench, R., Grote, A., Hiller, K., Choi, C., Koch, I., Schobert, M., Haertig, E., Klages, U. & Jahn, D. (2006) JProGO: A novel tool for the functional interpretation of prokaryotic microarray data using Gene Ontology information.

*Nucleic Acids Res.* **34**, 510-515

Bunk, B., Kucklick, M., Jonas, R., Munch, R., Schobert, M., Jahn, D. & Hiller, K. (2006)

MetaQuant: a tool for the automatic quantification of GC/MS based metabolome data.

*Bioinformatics.* **22**, 2962-2965

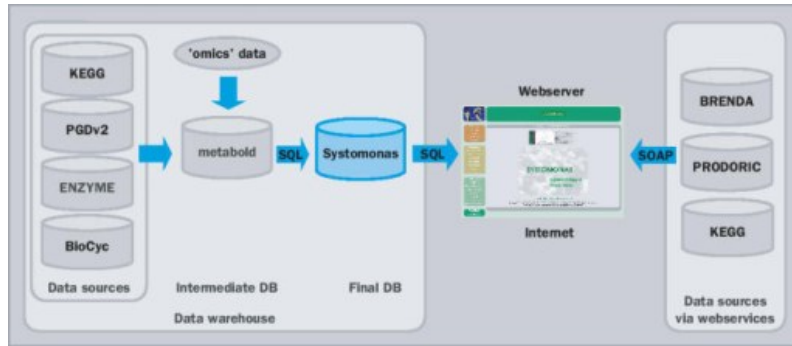


PrediSi  
**PREDiction of Signal peptides**



**MetaQuant**

# Intergenomics: Our start into bioinformatics



Choi, C., Muench, R., Leupold, S., Klein, J., Siegel, I., Thielen, B., Benkert, B., Kucklick, M., Schobert, M., Barthelmes, J., Ebeling, C., Haddad, I., Scheer, M., Grote, A., Hiller, K., Bunk, B., Schreiber, K., Retter, I., Schomburg, D. & Jahn, D. (2007) SYSTOMONAS - an integrated database for systems biology analysis of *Pseudomonas*. *Nucleic Acid Res.* **35**, D533-D537

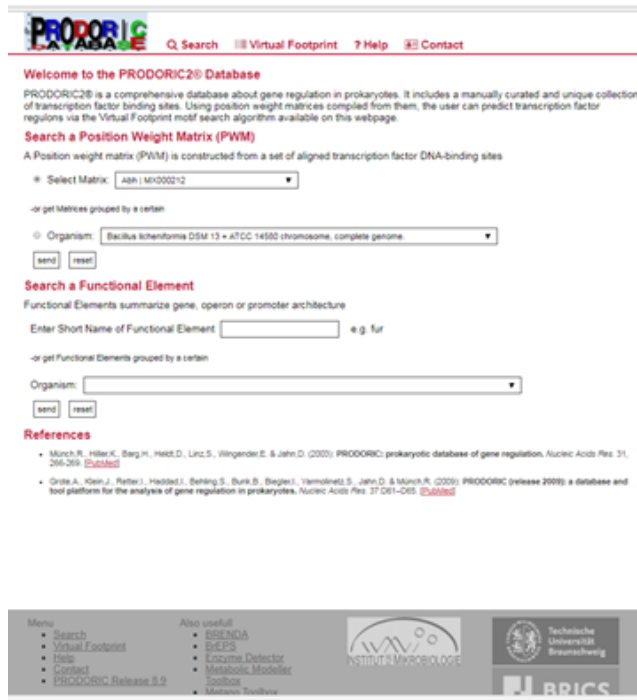
Choi, C., Münch, R., Bunk, B., Barthelmes, J., Ebeling, C., Schomburg, D., Schobert, M. & Jahn, D. (2007) Combination of a data warehouse concept with web services for the establishment of the *Pseudomonas* systems biology database SYSTOMONAS. *Journal of Integrative Bioinformatics* **4**:48

Pommerenke, C., Gabriel, I., Bunk, B., Münch, R., Haddad, I., Tielen, P., Wagner-Döbler I. & Jahn D. (2008) ROSY - a flexible and universal database and bioinformatics tool platform for *Roseobacter* related species. *In Silico Biol.* **8**:177-186

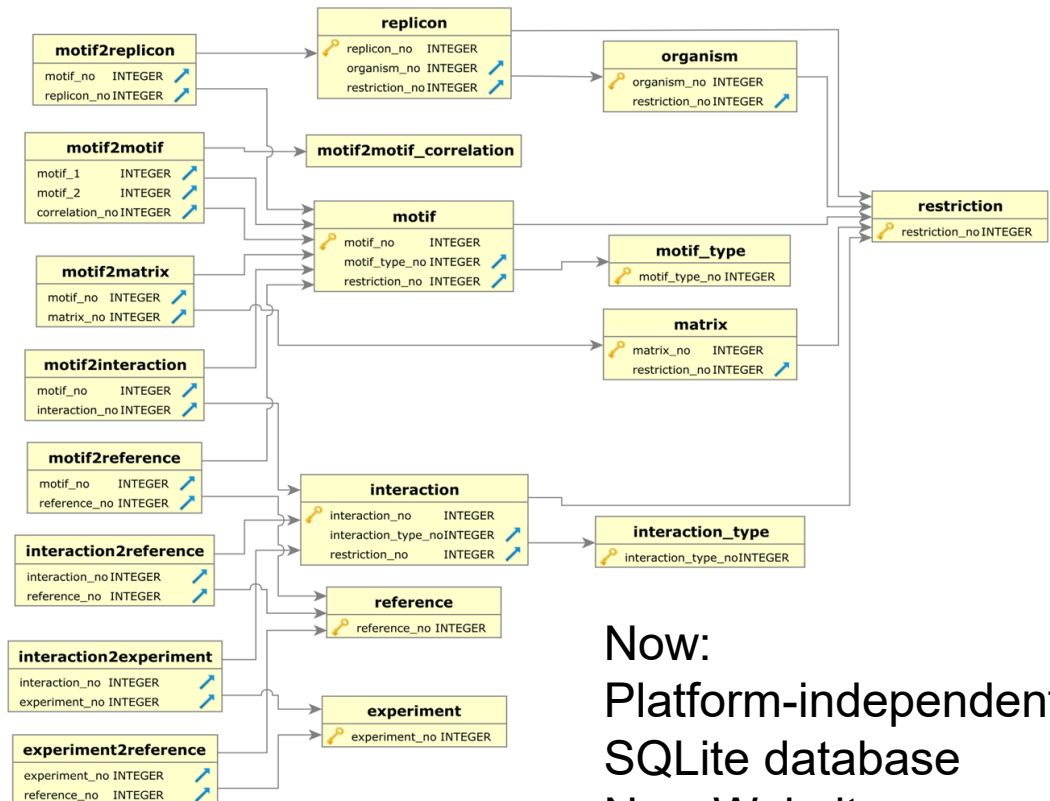
Klein, J., Münch, R., Biegler, I., Haddad, I., Retter, I. & Jahn, D. (2009) Strepto-DB, a database for comparative genomics of group A (GAS) and B (GBS) streptococci, implemented with the novel database platform 'Open Genome Resource' (OGeR) *Nucleic Acids Res.* **D494-8**

Shevshuk, O., Roselius, L., Günther, G., Klein, J., Jahn, D., Steinert, M. & Münch, R. (2012) InFiRe – a novel computational method for the identification of insertion sites in transposon mutagenized bacterial genomes *Bioinformatics*, **28**:306-310

# The new PRODORIC2 database (<http://www.prodoric2.de>)



Eckweiler, D., Dudek, C.-A., Hartlich, J., Brötje, D., & Jahn, D. (2017): **PRODORIC2: the bacterial gene regulation database in 2018.** *Nucleic Acids Res.* 46, D320-D326.



Now:  
Platform-independent  
SQLite database  
New Website

# The new PRODORIC2 database (<http://www.prodoric2.de>)

Table 1. Statistics of the PRODORIC2 content (September 2017)

| Organism                                         | TFBSs | PWMs <sup>a</sup> | Increase in % <sup>b</sup> |
|--------------------------------------------------|-------|-------------------|----------------------------|
| <i>Bacillus licheniformis</i>                    | 8     | 3                 | new                        |
| <i>Bacillus megaterium</i>                       | 7     | 2                 | new                        |
| <i>Bacillus subtilis</i>                         | 907   | 91(65)            | 28                         |
| <i>Bradyrhizobium japonicum</i>                  | 36    | 4(3)              | 33                         |
| <i>Campylobacter jejuni</i>                      | 5     | 2                 | 100                        |
| <i>Clostridium acetobutylicum</i>                | 23    | 3                 | new                        |
| <i>Clostridium beijerinckii</i>                  | 7     | 2                 | new                        |
| <i>Clostridium difficile</i> 630                 | 266   | 13(12)            | new                        |
| <i>Corynebacterium glutamicum</i>                | 59    | 6                 | 500                        |
| <i>Dinoroseobacter shibae</i>                    | 255   | 6                 | new                        |
| <i>Escherichia coli</i> CFT073                   | 4     | 2                 | 100                        |
| <i>Escherichia coli</i> str. K-12 substr. MG1655 | 1562  | 94(82)            | 7                          |
| <i>Escherichia coli</i> str. K-12 substr. W3110  | 76    | 6                 | 50                         |
| <i>Helicobacter pylori</i>                       | 24    | 3                 | 50                         |
| <i>Listeria monocytogenes</i>                    | 62    | 5                 | 150                        |
| <i>Mycobacterium tuberculosis</i>                | 73    | 5                 | 400                        |
| <i>Pseudomonas aeruginosa</i> PAO1               | 444   | 41(36)            | 86                         |
| <i>Rhodobacter sphaeroides</i>                   | 38    | 6                 | 50                         |
| <i>Salmonella enterica</i>                       | 6     | 1                 | -                          |
| <i>Staphylococcus aureus</i> COL                 | 34    | 2                 | 100                        |
| <i>Staphylococcus aureus</i> Newman              | 2     | 1                 | -                          |
| <i>Streptococcus agalactiae</i>                  | 5     | 1                 | -                          |
| <i>Streptococcus pyogenes</i> MGAS5005           | 4     | 1                 | -                          |
| <i>Streptococcus pyogenes</i> MGAS8232           | 10    | 2                 | 100                        |
| <i>Streptococcus pyogenes</i>                    | 3     | 1                 | -                          |
| <i>Synechococcus elongatus</i>                   | 11    | 1                 | -                          |
| <i>Synechocystis</i>                             | 16    | 3                 | 200                        |
| Sum                                              | 3947  | 307(261)          | 31                         |

<sup>a</sup>The non-redundant number of PWMs (in parentheses).

<sup>b</sup>Based on the total numbers.

# The new PRODORIC2 database (<http://www.prodoric2.de>)



[Browse](#)
[Virtual Footprint](#)
[? Help](#)
[Statistics](#)
[Contact](#)

## Details for MX000245

### General Information

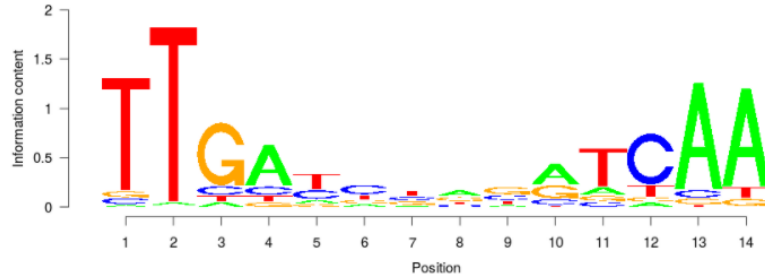
**Accession** MX000245  
**Factor** FnrL  
**Locus Tag** [dshi\\_0660](#)  
**Replicon** Dinoroseobacter shibae DFL 12 chromosome, complete genome.  
**External Links** [NCBI](#) [GenBank](#) [BacDive](#)

### Frequency Matrix

|           |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A         | 2  | 0  | 4  | 45 | 4  | 7  | 7  | 26 | 6  | 32 | 9  | 3  | 58 | 52 |
| T         | 54 | 62 | 2  | 5  | 33 | 12 | 24 | 9  | 12 | 1  | 45 | 8  | 1  | 6  |
| G         | 3  | 0  | 50 | 3  | 6  | 11 | 16 | 19 | 26 | 20 | 4  | 4  | 2  | 4  |
| C         | 3  | 0  | 6  | 9  | 19 | 32 | 15 | 8  | 18 | 9  | 4  | 47 | 1  | 0  |
| Consensus | T  | T  | G  | A  | T  | C  | T  | A  | G  | A  | T  | C  | A  | A  |

[download TRANSFAC](#)

### Sequence Logo



### Motif List

[download FASTA](#)

| Motif Acc                | Motif Name       | Binding Sequence | Start  | End    | Orientation | PubMed-ID                |
|--------------------------|------------------|------------------|--------|--------|-------------|--------------------------|
| <a href="#">MF102974</a> | eftB operon      | TTCCCTGCGCGTAA   | 215858 | 215845 | -           | <a href="#">28473807</a> |
| <a href="#">MF102975</a> | rpsU operon      | TTGCTGGACAACATA  | 244049 | 244036 | -           | <a href="#">28473807</a> |
| <a href="#">MF102976</a> | secE operon      | TTGAAATCTTCACAA  | 256209 | 256223 | +           | <a href="#">28473807</a> |
| <a href="#">MF102977</a> | Dshi_0424 operon | TTGATCTGCGTTAA   | 417031 | 417044 | +           | <a href="#">28473807</a> |
| <a href="#">MF102978</a> | hemN1 operon     | TTGATGAAGGTCAA   | 526350 | 526363 | +           | <a href="#">28473807</a> |
| <a href="#">MF102979</a> | hemN2 operon     | TTGATCTGGGTCAA   | 664358 | 664345 | -           | <a href="#">28473807</a> |

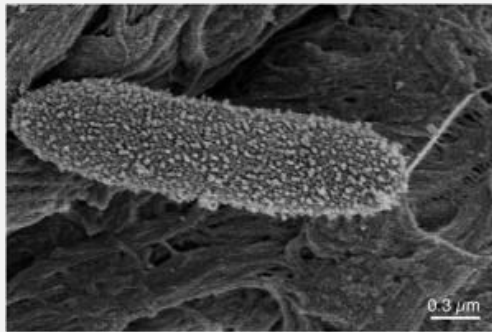
*Dinoroseobacter shibae* FnrL

Homolog of Anr of  
*Pseudomonas aeruginosa*

# Two cases

## 1. Infection

### *Pseudomonas aeruginosa*

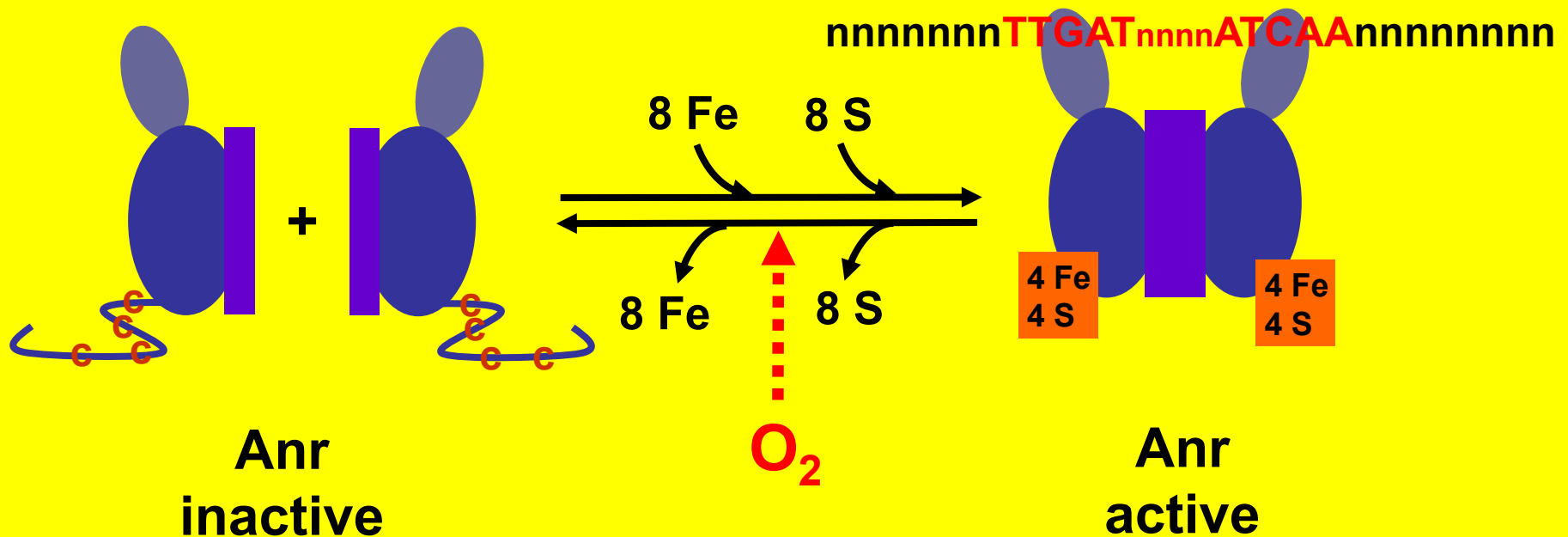


(Nelly Panté, University of British Columbia)

- Gram-negative, ~ 2  $\mu\text{m}$
- 6,3 Mbp, ~5600 genes
- Denitrification
- Opportunistic pathogen

# The oxygen regulator Anr regulon

## Model of Anr function





# Virtual Footprint

Transcription Factor Binding Sites  
(Training Set)

Position Weight Matrix

Hits with Homology Scores

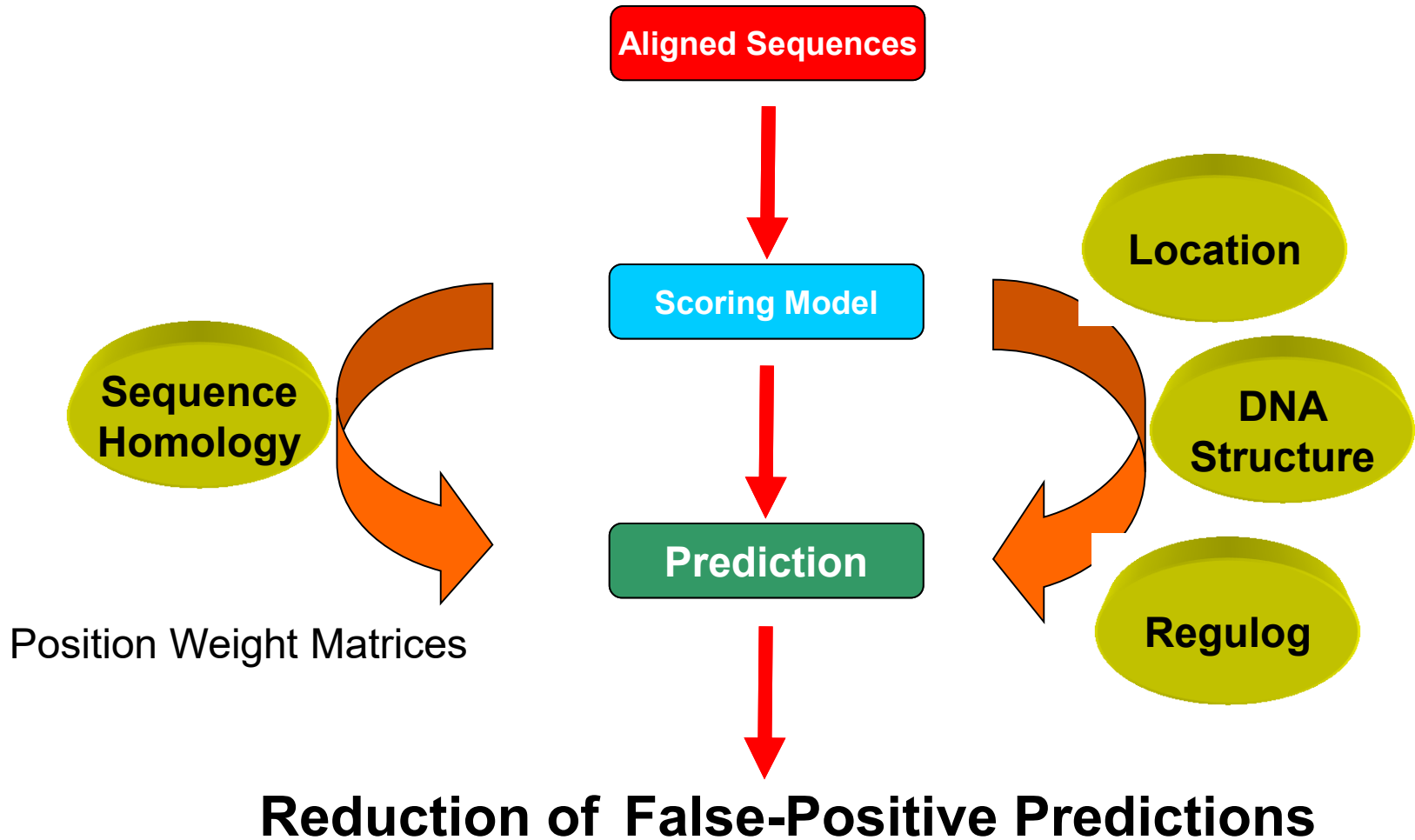
Aligned Sequences

Scoring Model

Prediction

TTGACGTGGATCAG  
TTGACCTGAATCAG  
CTGTCAITGGATCAA  
TTGATACAAATCAA  
TTGACGGCCGTCAA  
TTGATCGCGGTCAA  
TTGCCGTGCGTCAA  
TTGACCGGAATCAA  
TTGATTCCTATCAA  
TTGTCTCGCGACAA  
TTGCTCTGCATCAA

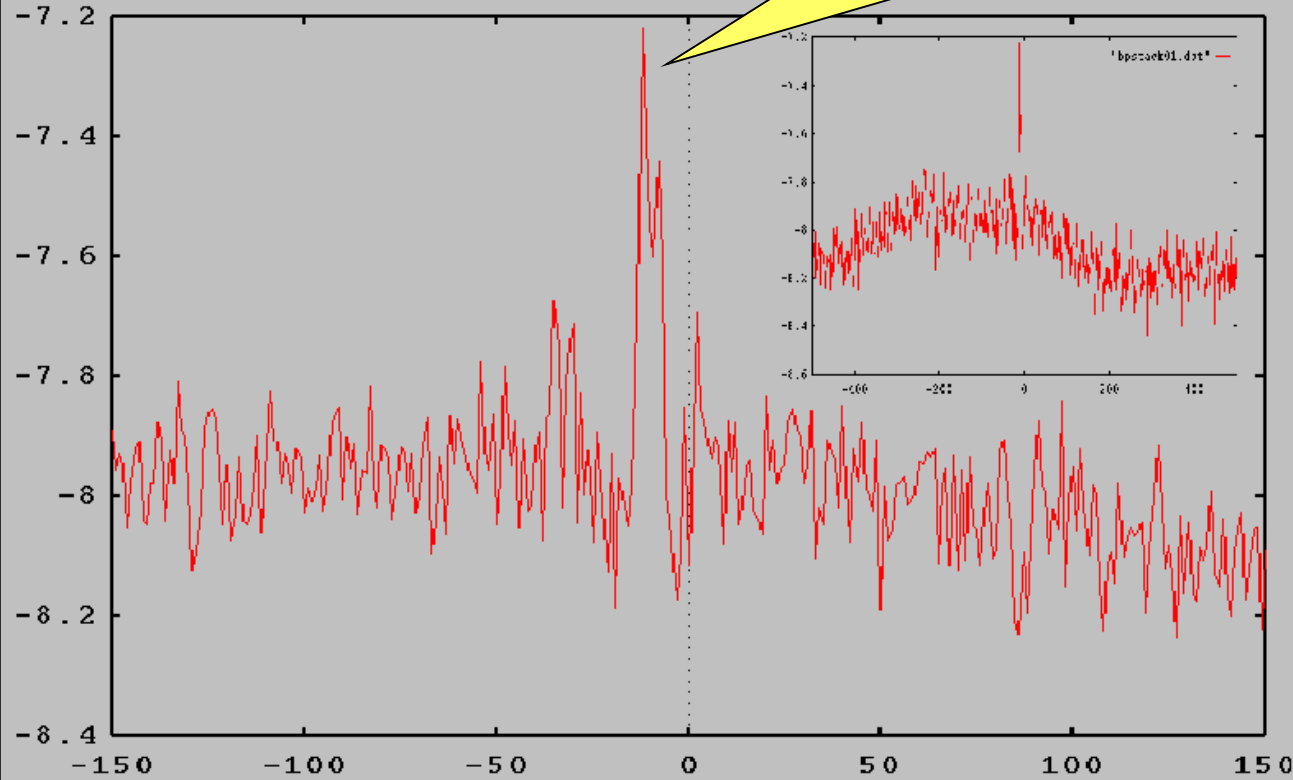
|   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| A | 0.00 | 0.00 | 0.00 | 0.44 | 0.00 | 0.01 | 0.00 | 0.07 | 0.04 | 0.67 | 0.14 | 0.00 | 2.00 | 1.08 |
| C | 0.14 | 0.00 | 0.00 | 0.13 | 0.67 | 0.02 | 0.13 | 0.21 | 0.05 | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 |
| G | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 | 0.02 | 0.13 | 0.48 | 0.04 | 0.38 | 0.00 | 0.00 | 0.00 | 0.24 |
| T | 1.42 | 2.00 | 0.00 | 0.13 | 0.38 | 0.01 | 0.21 | 0.00 | 0.01 | 0.00 | 1.42 | 0.00 | 0.00 | 0.00 |



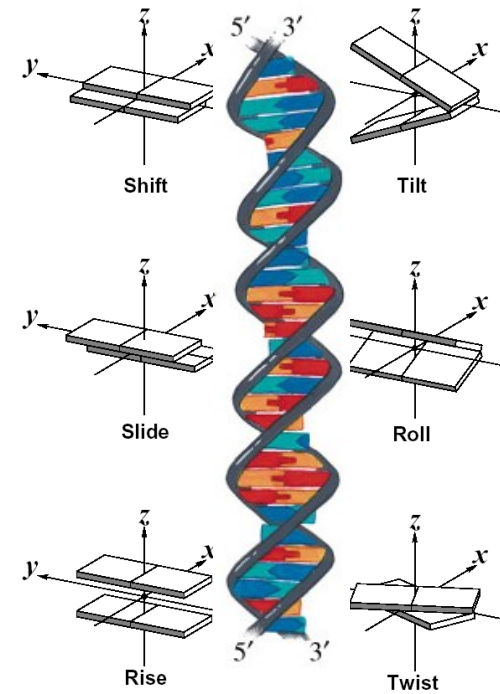


## Base-stacking

The average base-stacking energies for 1911 promoters, averaged over the transcription start site.



## DNA Structure



Base Stacking

# Regulog Analysis: *hemN*

Phylogenetic conservation of the Anr binding site upstream of other *hemN* genes:

Regulog

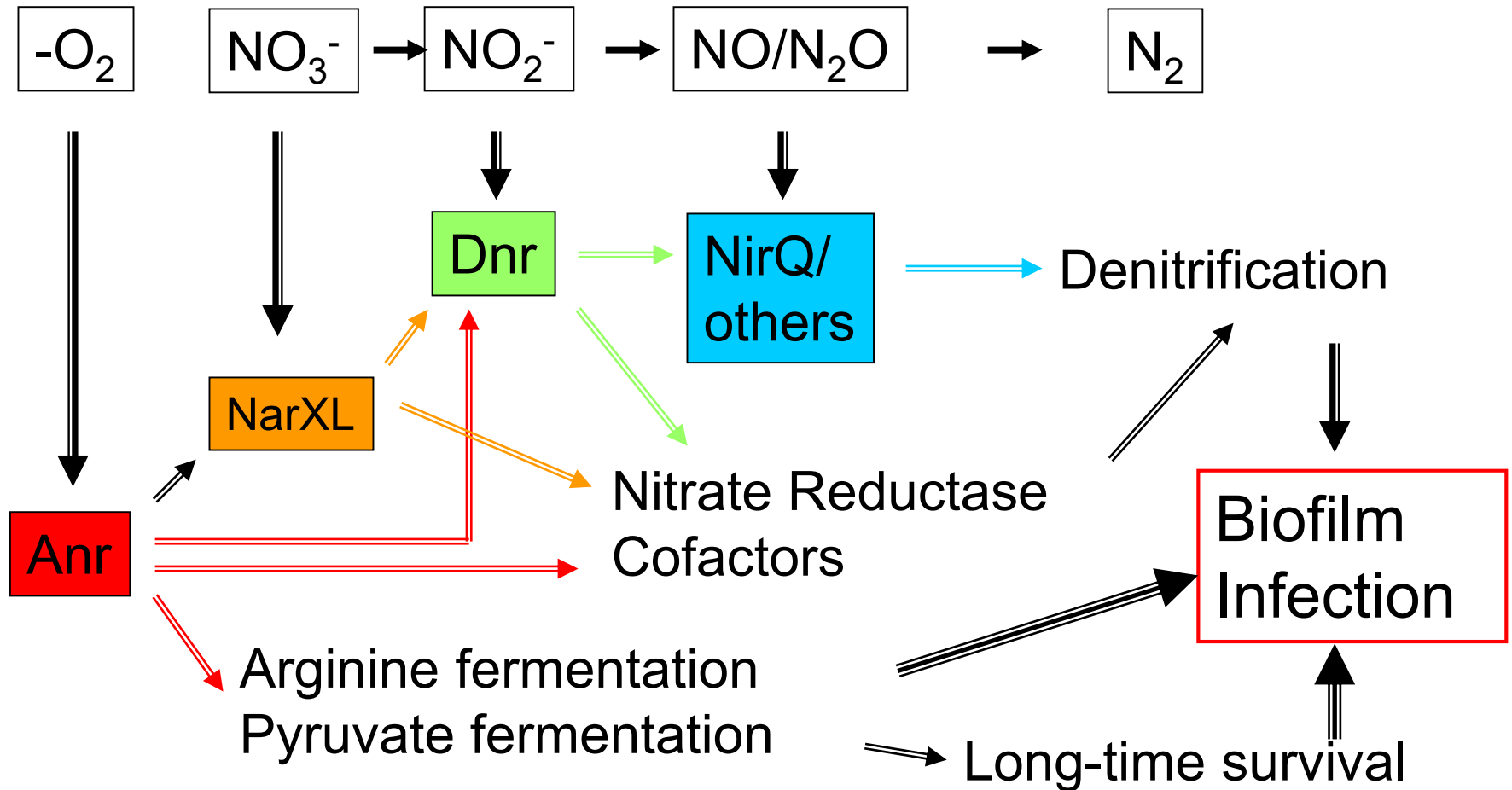
| Organism                         | Pos.    | Sequence                                                |
|----------------------------------|---------|---------------------------------------------------------|
| <i>Pseudomonas aeruginosa</i>    | -77/-64 | . . . GGAAGGCCTG <b>TTGATTTGTATCAAG</b> CGGGTGAAT . . . |
| <i>Pseudomonas putida</i>        | -77/-64 | . . . GGAAACTGTG <b>TTGATTTGTATCAAG</b> GGCATTGGG . . . |
| <i>Pseudomonas syringae</i>      | -76/-63 | . . . GTGGCGGGCC <b>TTGATGTGGATCAA</b> AAGGCGGGAA . . . |
| <i>Chromobacterium violaceum</i> | -79/-66 | . . . CCGGCGAAGC <b>TTGACCCTTATCAAG</b> GCGGAATTC . . . |
| <i>Neisseria meningitidis</i>    | -87/-74 | . . . TAAACCTGTC <b>TTGATATGTGTCAAT</b> AAGCACATA . . . |
| <i>Xanthomonas campestris</i>    | -48/-35 | . . . CAGAGGCGCC <b>TTGATCAGAATCAAT</b> CTATCGTGA . . . |
| <i>Agrobacterium tumefaciens</i> | -83/-70 | . . . GTTGCAACT <b>TTGACTTCAATCAA</b> ATTGCGCCAT . . .  |
| <i>Rhizobium loti</i>            | -70/-57 | . . . CGCCAATGCC <b>TTGATTTCGGTCAA</b> AGCCTTTGCA . . . |

# Experiment and prediction

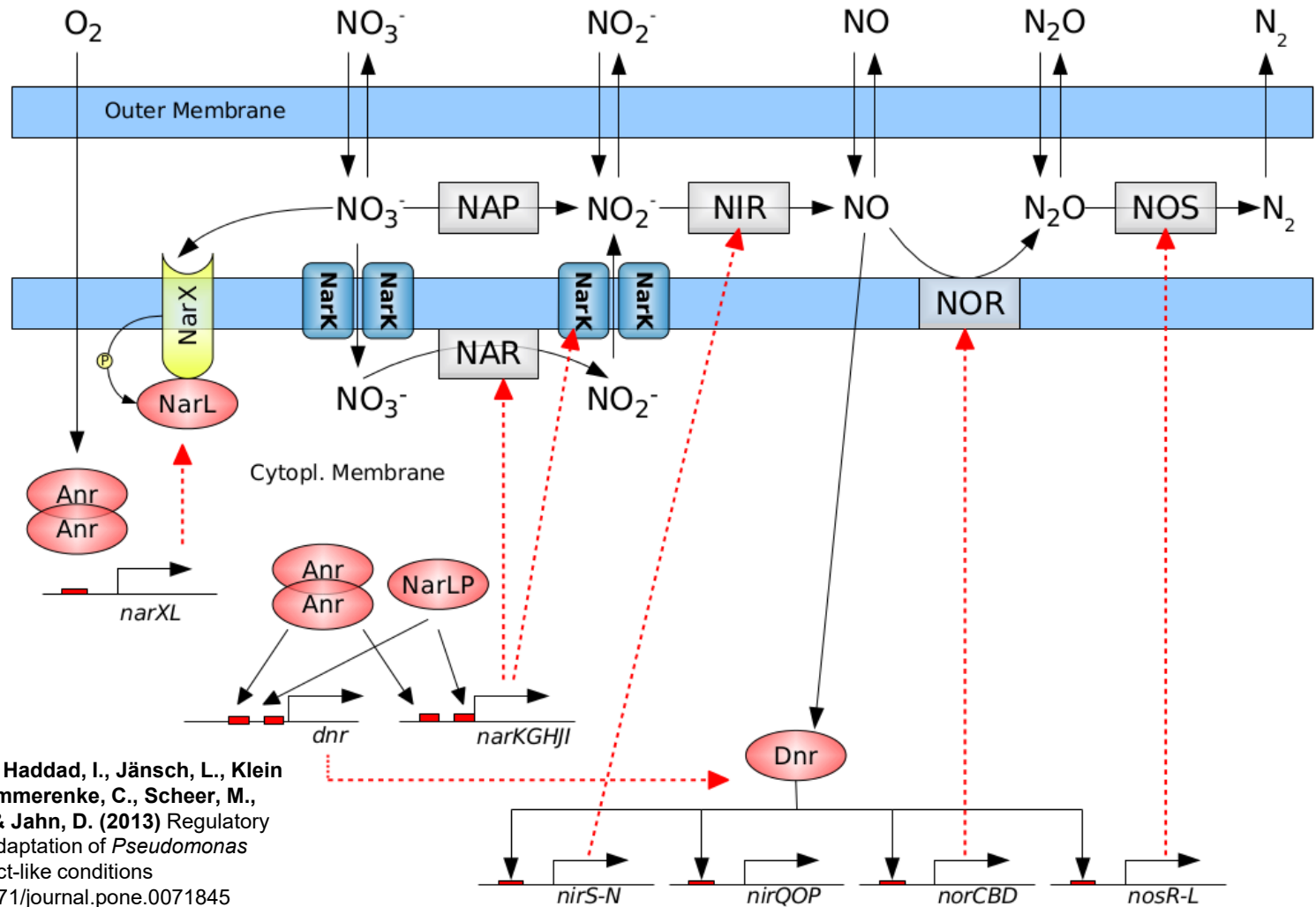
Predicted Anr regulon  
versus experimentally determined (DNA array)

| Prediction                  | Matches | Genes | Regulog | Overlap |
|-----------------------------|---------|-------|---------|---------|
| Whole genome                | 523     | 1000  | 27%     | 87      |
| Intergenic                  | 179     | 352   | 75%     | 79      |
| Intergenic, Stacking Energy | 105     | 212   | 80%     | 68      |

# *P. aeruginosa* oxygen regulatory network

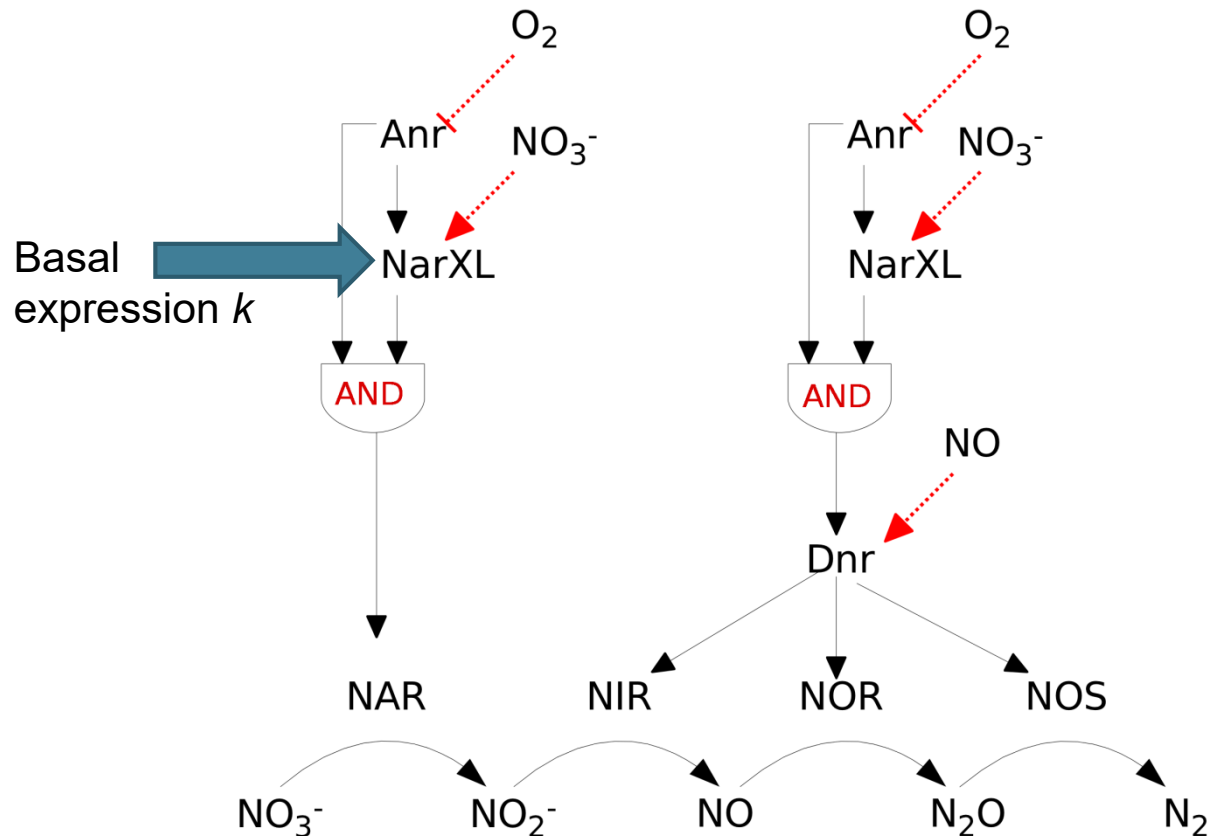


## Regulatory circuits for denitrification onset



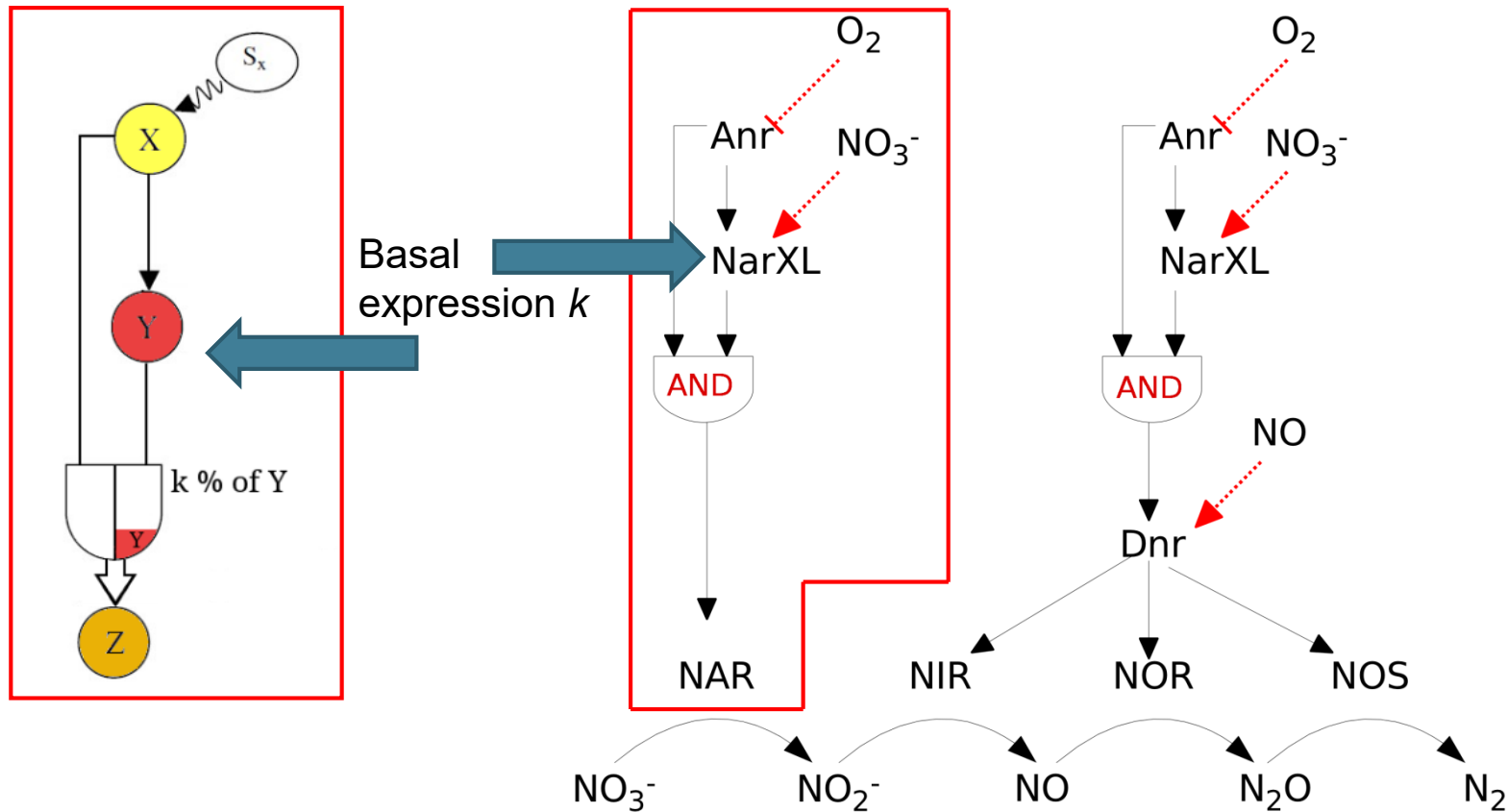
**Tielen, P., Rosin, N., Dohnt, K., Haddad, I., Jansch, L., Klein J., Meyer, A.-K., Narten, M., Pommerenke, C., Scheer, M., Schobert, M., Schomburg, D. & Jahn, D. (2013) Regulatory and metabolic networks for the adaptation of *Pseudomonas aeruginosa* biofilms to urinary tract-like conditions**  
*PLOS ONE* **8**: e71845. doi:10.1371/journal.pone.0071845

# Anr / NarXL feed-forward loops as digital logic schema



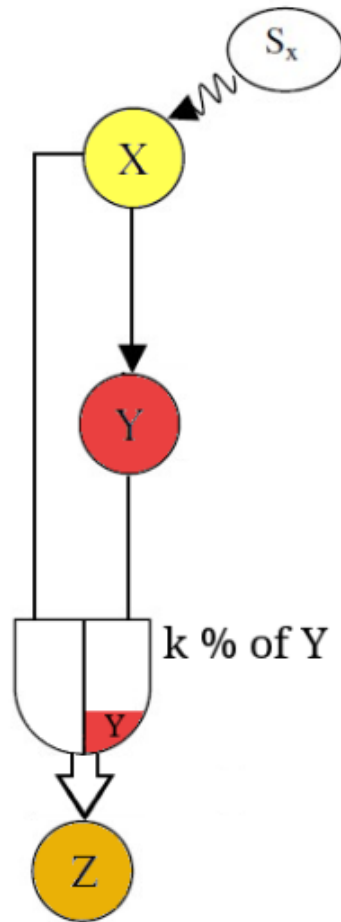
**Question:** What is the influence of basal expression in feed-forward loops on environmental adaptation in general?

# Anr / NarXL feed-forward loops as digital logic schema



**Question:** What is the influence of basal expression in feed-forward loops on environmental adaptation in general?

# Initial model with 3 differential equations



$$\frac{dX(t)}{dt} = S_x(t) - X(t)$$

$$\frac{dY(t)}{dt} = k + (1 - k)X(t) - Y(t)$$

$$\frac{dZ(t)}{dt} = X(t)Y(t) - Z(t)$$

Signal  $S_x$  (oxygen tension)

Regulator X

Regulator Y

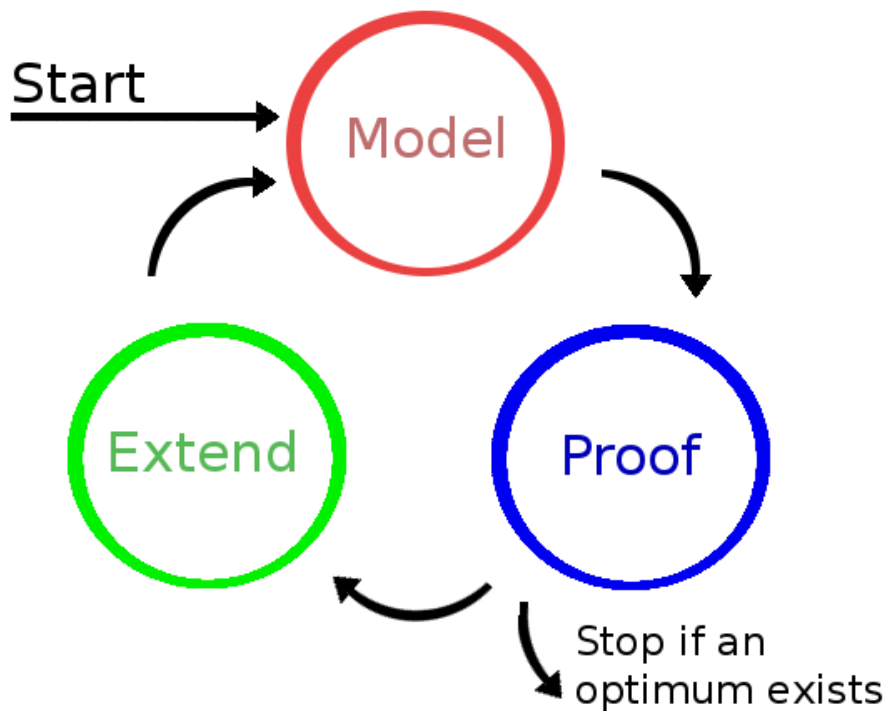
Signal (nitrate) for regulator Y = constant

Production of Z as a function of X and Y

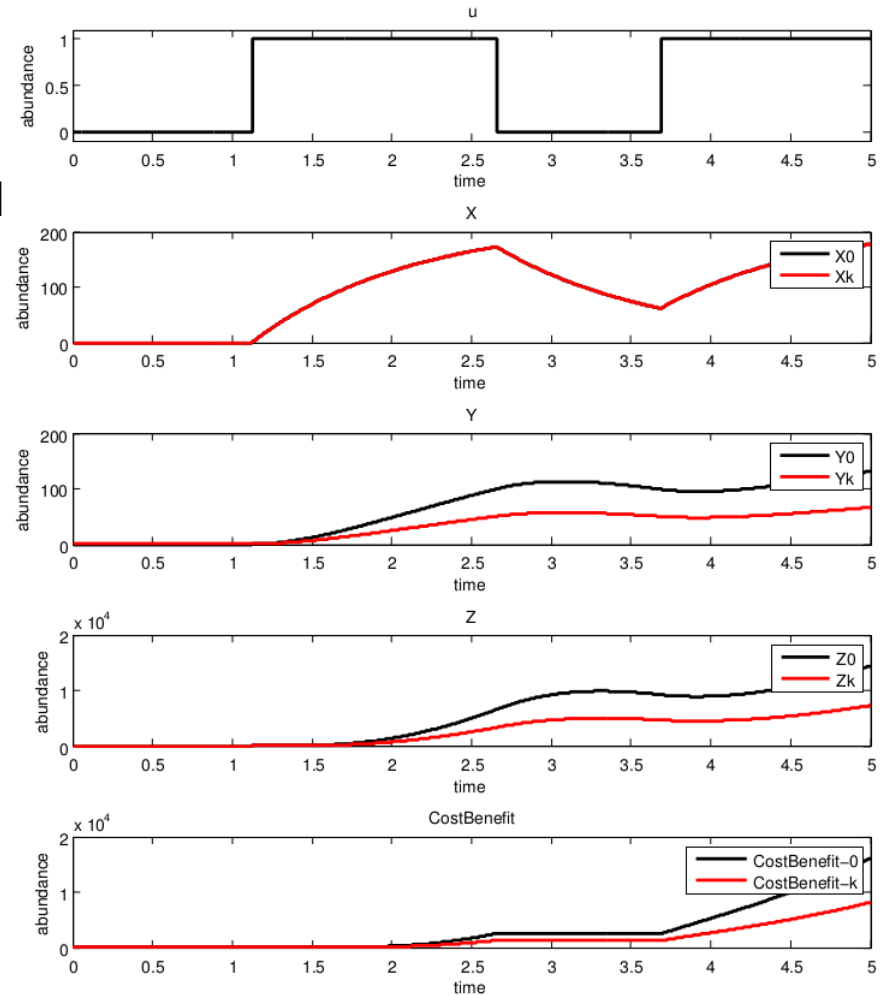
Basal expression of k?

# Working Procedure

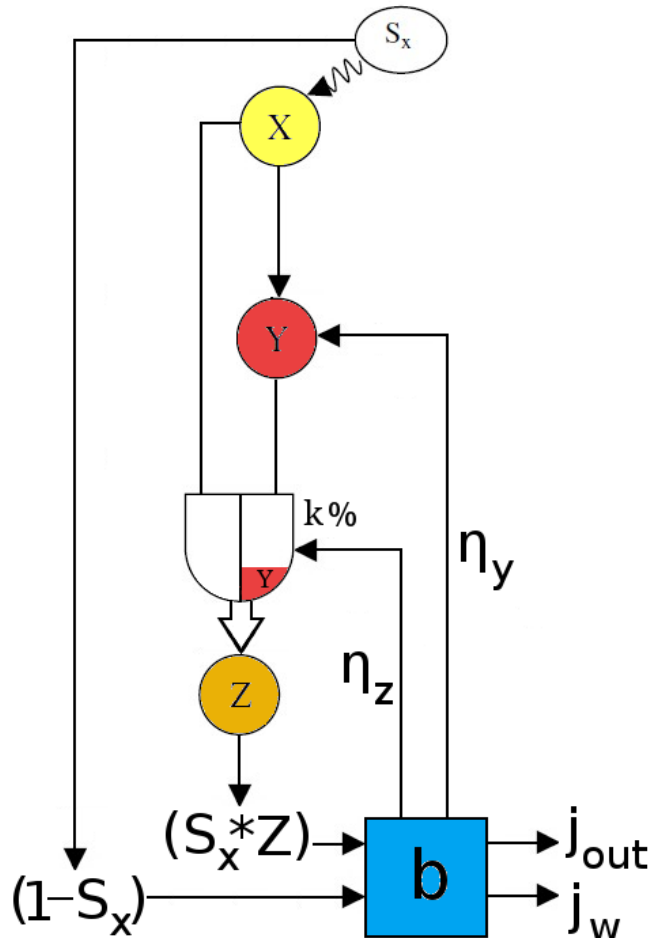
- Start with the simplest model
- Proof whether an optimal  $k$  exist via simulations
- If no optimal  $k$  exists, extend the model



## Simulations with the model



# Final model

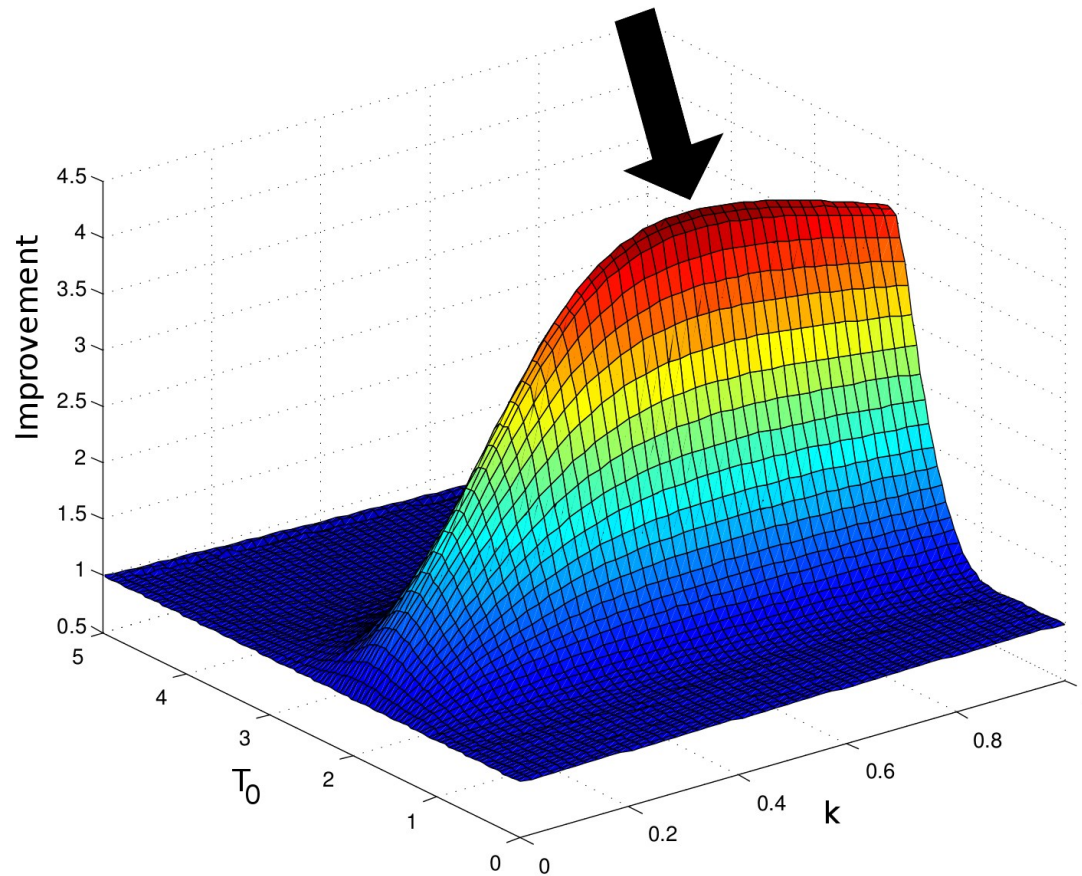


- $\frac{dX(t)}{dt} = S_x(t) - X(t)$
- $\frac{dY(t)}{dt} = k + (1 - k) \frac{(1+K_{xy})X(t)^h}{X(t)^h+K_{xy}} - Y(t)$
- $\frac{dZ(t)}{dt} = \frac{(1+K_{xz})X(t)^h}{X(t)^h+K_{xy}} \frac{(1+K_{yz})Y(t)^h}{Y(t)^h+K_{yz}} - Z(t)$
- $\frac{db(t)}{dt} = 1 - S_x(t) + S_x(t)Z(t) - j_{out} - \eta_y \left( k + (1 - k) \frac{(1+K_{xy})X(t)^h}{X(t)^h+K_{xy}} \right) - \eta_z \left( \frac{(1+K_{xz})X(t)^h}{X(t)^h+K_{xy}} \frac{(1+K_{yz})Y(t)^h}{Y(t)^h+K_{yz}} \right) - j_w$
- $\frac{dw(t)}{dt} = j_w$
- $j_w = B(1 - b(t))^+$

Two more state variables:

1. Energy store  $b(t)$  (ATP)
2. Growth rate of the population  $w(t)$

# Final Model

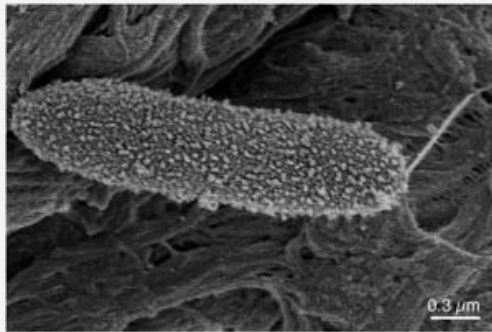


**Roselius, L., Langemann, D., Müller, J., Hense, B., Filges, S., Jahn, D. & Münch, R. (2014)**  
Modeling and analysis of a gene-regulatory feed-forward loop with basal expression of the second regulator  
*J Theoret Biol.*, **363**:290–299

# Two cases

## 1. Infection

### *Pseudomonas aeruginosa*

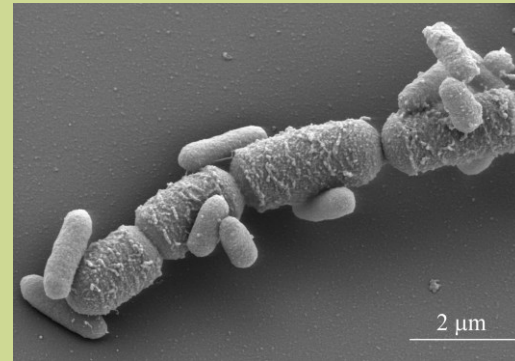


(Nelly Panté, University of British Columbia)

- Gram-negative, ~ 2 μm
- 6,3 Mbp, ~5600 genes
- Denitrification
- Opportunistic pathogen

## 2. Biotechnology

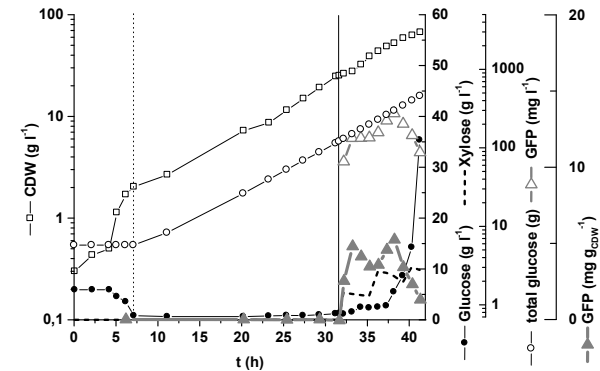
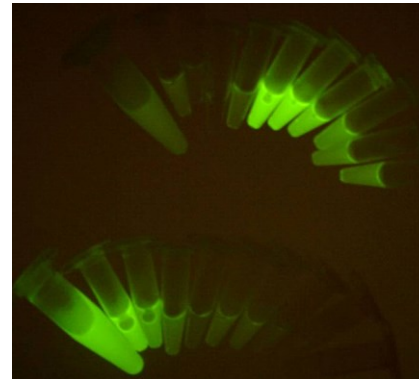
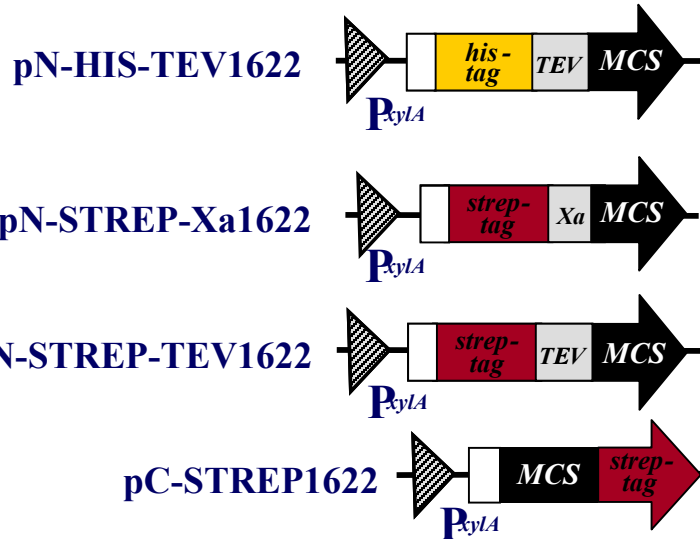
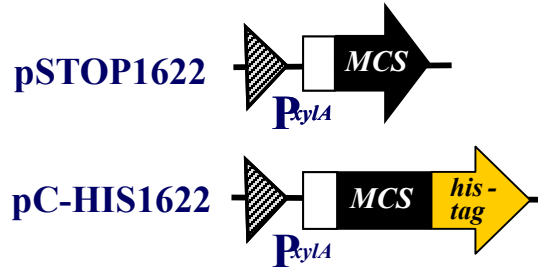
### *Bacillus megaterium*



(M. Rohde, HZI Braunschweig)

- Gram-positive, ~4-8 μm
- 5.1 Mbp, ~5300 genes
- Biotechnological applications (e.g. protein production)

# Bacillus megaterium biotechnology

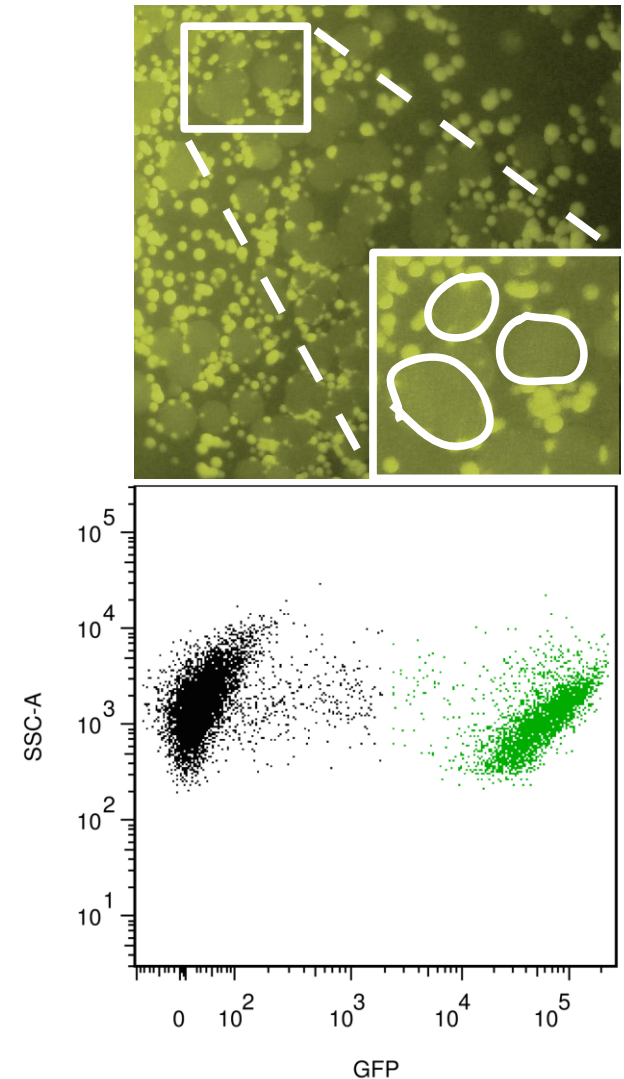


1.25 g recombinant protein per liter culture (intracellular)  
500 mg recombinant protein per liter culture (extracellular)

# Phenotypic heterogeneity during protein production

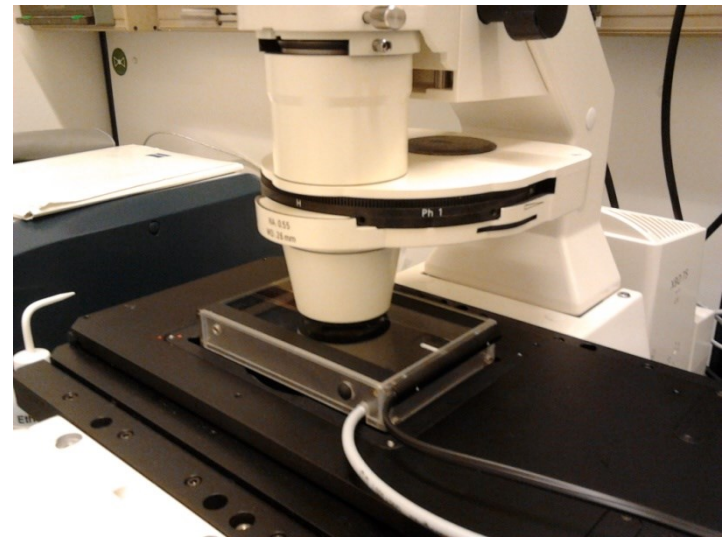
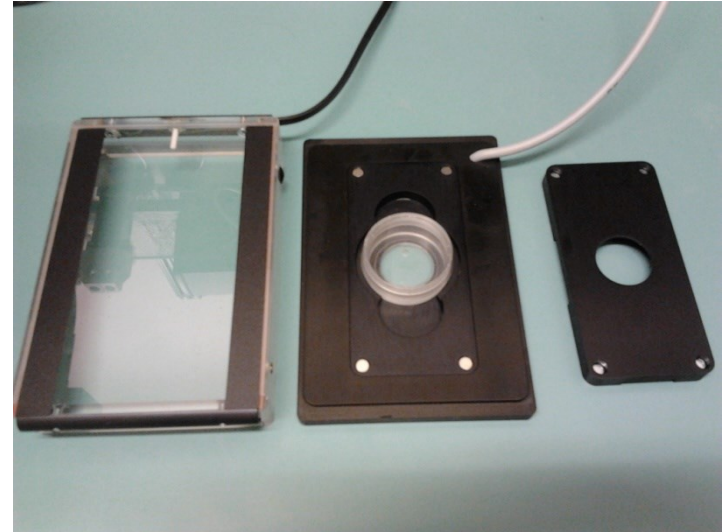
## Subpopulations of low and high-producing cells

- Small green high producing colonies (~70%)
- Large white low producing colonies
- Two distinct cell fractions in FACS analysis



# Time-lapse microscopy – experimental setup

- Inverse fluorescence microscope
- Heating system and controller



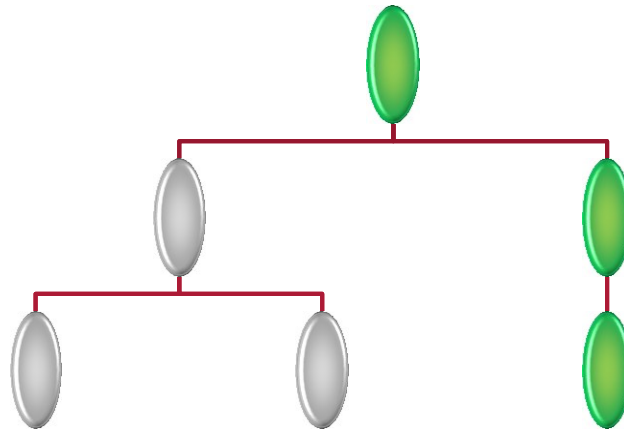
# Time lapse microscopy of GFP production by *B.megaterium*



# Culture heterogeneity in bacteria

## Definition:

Culture heterogeneity = genetically identical cells acquire distinct phenotypes.



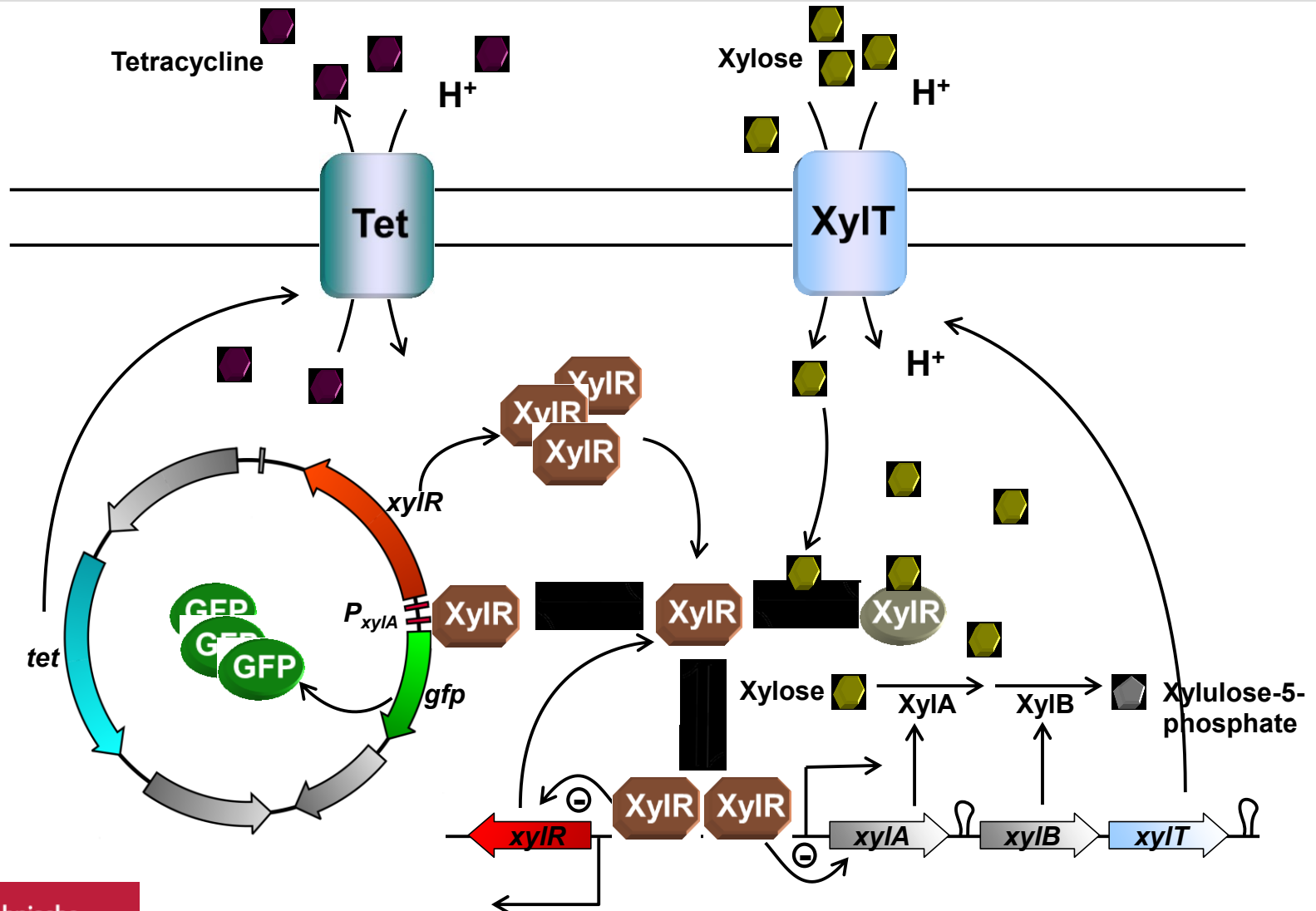
The stable formation of two subpopulations is also called **bistability**.

## Examples:

Adaptation to C-sources, competence, sporulation, persister cells, self-destructive cooperation, etc.

**Possible explanation:** Adaptation to fluctuating environments, „bet hedging“

# Xylose inducible protein production system

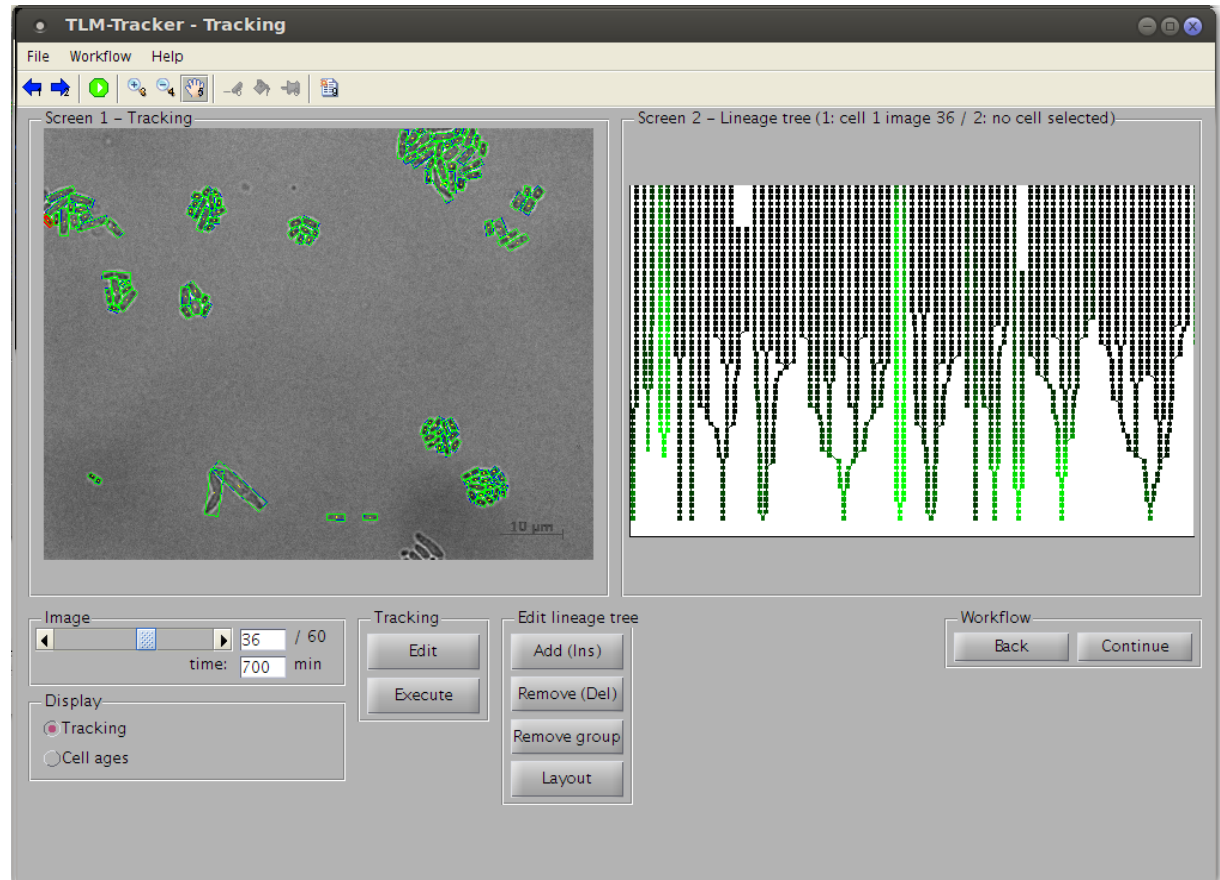


# TLM-Tracker: Image analysis software for time-lapse movies

**First step:**  
Segmentation of cells

**Second step:**  
Tracking of cells

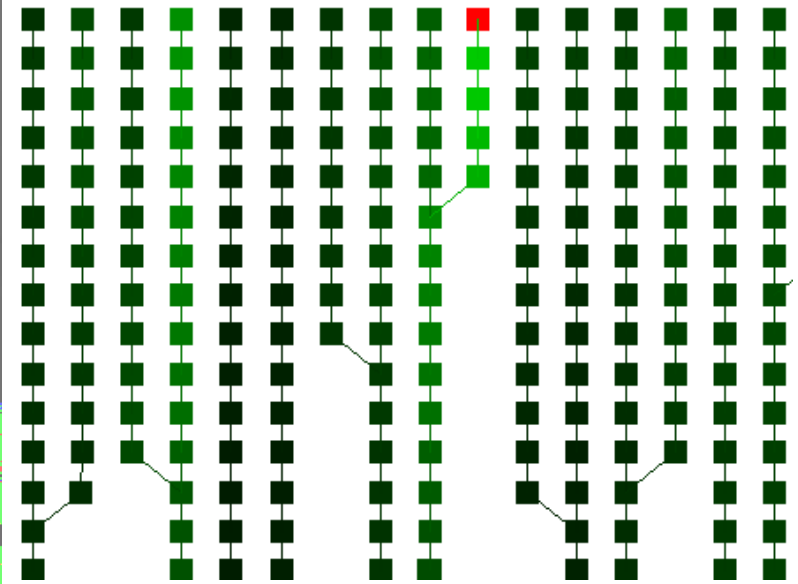
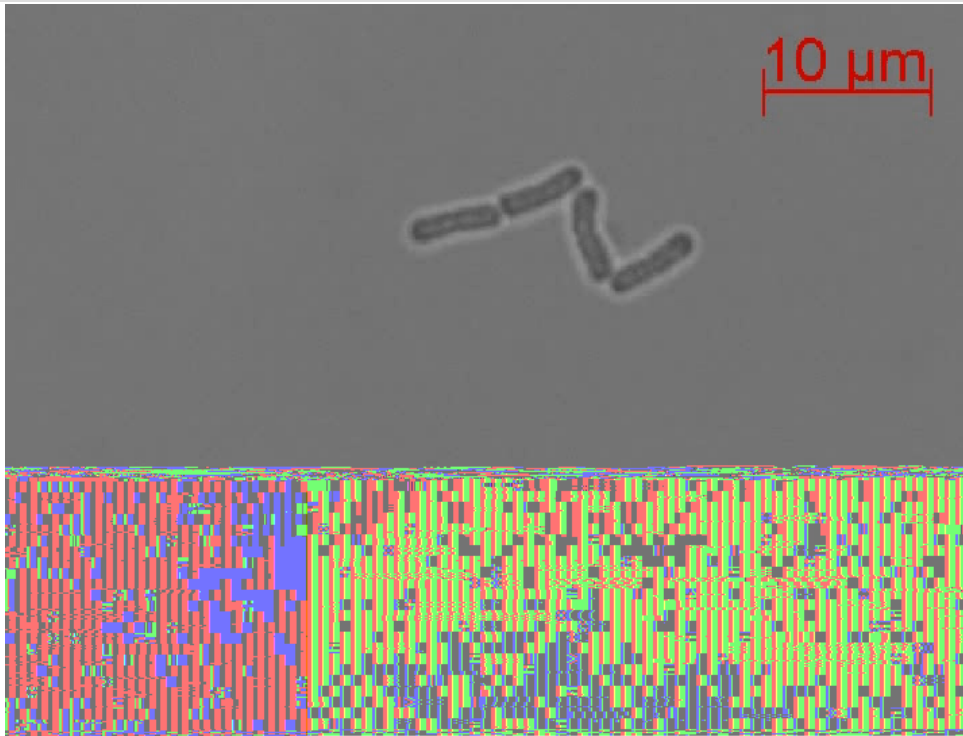
**Third step:**  
Data export and  
analysis



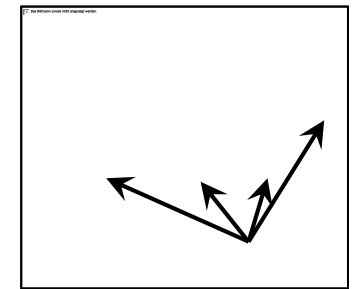
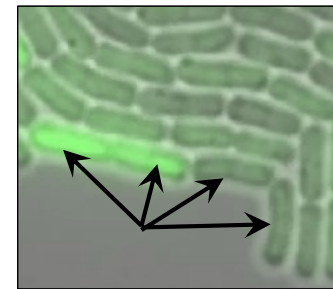
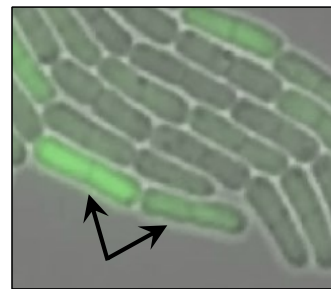
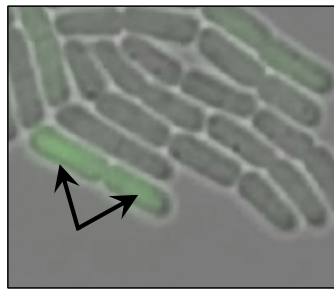
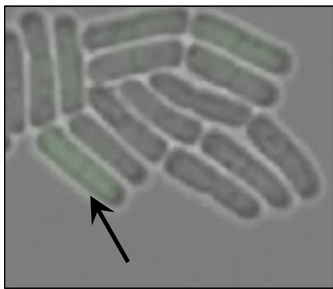
Klein J., Leupold S., Biegler I., Biedendieck R., Münch R. & Jahn D. (2012) TLM-Tracker: software for cell segmentation, tracking and lineage analysis in time-lapse microscopy movies. *Bioinformatics* **28**:2276-2277

**Free Download:** <http://www.tlmtracker.tu-bs.de/>

# GFP production in a growing microcolony

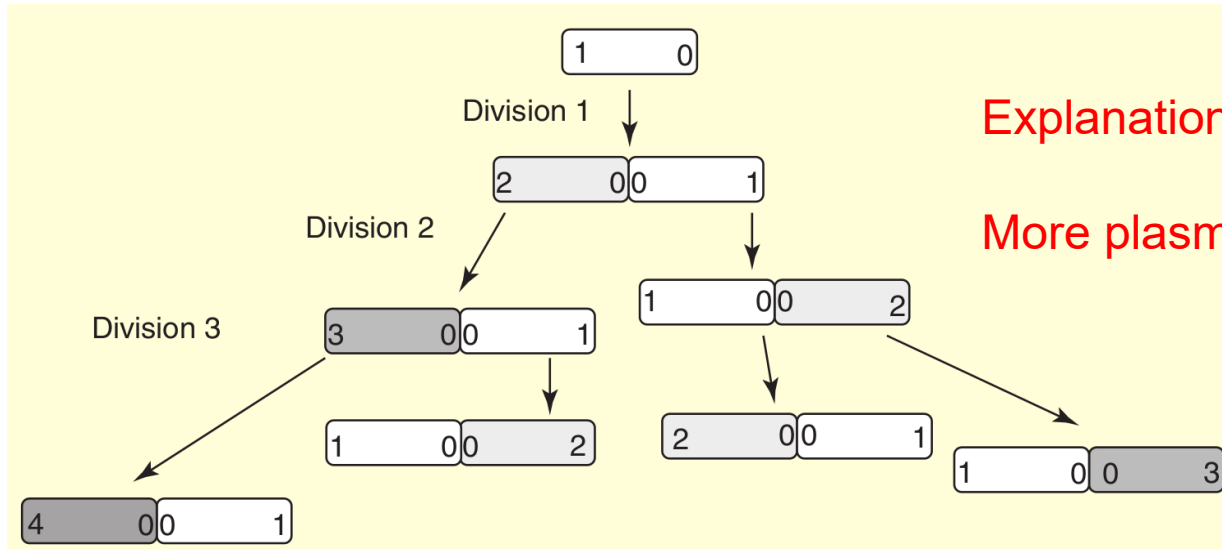


→ neighboring cells stop production



# Conclusion:

## Influence of cell aging



Explanation: Unequal plasmid distribution

More plasmids at the old cell pole

Münch, K., Müller, J., Wienecke, S., Bergmann, S., Heyber, S., Biedendieck, R., Münch, R. & Jahn, D. (2015)  
Polar fixation of plasmids during recombinant protein production in *Bacillus megaterium* results in population heterogeneity.  
*Appl Environ Microbiol.*, **81**:5976-5986



<https://www.tu-braunschweig.de/forschung/zentren/brics>

# Summary:

- How it all started with Edgar
- Some Bioinformatics from Braunschweig
- PRODORIC database
- Prediction of gene regulatory networks in bacteria
- Modeling of regulatory circuits
- Culture heterogeneity
- BRICS

# Thank you:

