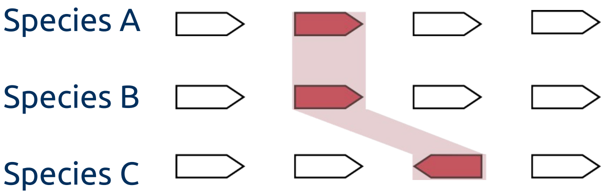
Maria Schreiber, M. Sc.

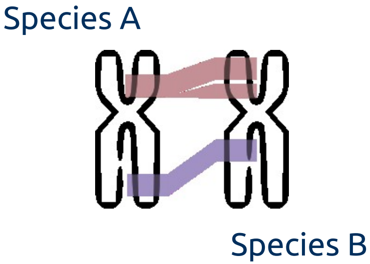
SweetSynteny
Unraveling Microsynteny Patterns

40th TBI Winterseminar in Bled, 2025

Definitions of synteny

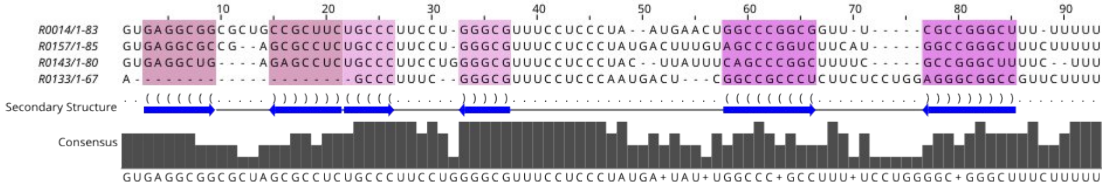
Microsynteny	*Macrosynteny*
Conservation of gene order and orientation within a small genomic region.	*Conservation of gene order and orientation across large chromosomal segments or even entire chromosomes.*
Local	*Global*
Preservation of gene clusters	*Preservation of chromosomal segments*





Why microsynteny?

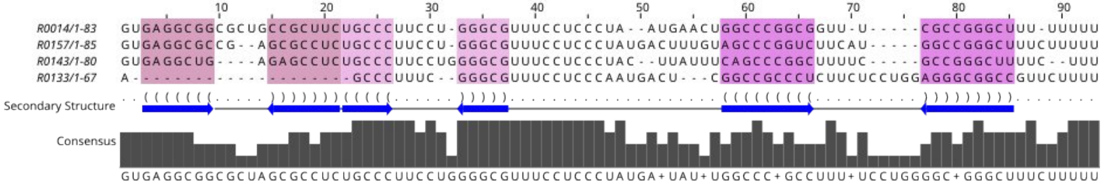
Characterization of sibling sRNAs in *Caulobacter crescentus*



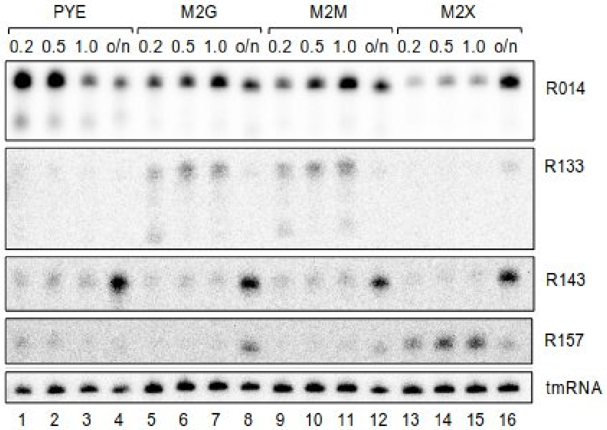
Dr. Kathrin Fröhlich
Junior Research Group
Bacterial RNA Biology

Why microsynteny?

Characterization of sibling sRNAs in *Caulobacter crescentus*



```
                 10            20            30            40            50            60            70            80            90
R0014/1-83  GUGAGGCGGCGCUGCCGCUUCUGCCCUUCCU-GGGCGUUUCCUCCCUA--AUGAACUGGCCCGGCGGUU-U-----CGCCGGGCUUU-UUUUU
R0157/1-85  GUGAGGCGCCG--AGCGCCUCUGCCCUUCCU-GGGCGUUUCCUCCCUAUGACUUUGUAGCCCGGUCUUCAU-----GGCCGGGCUUUCUUUUU
R0143/1-80  GUGAGGCUG----AGAGCCUCUGCCCUUCCUGGGGCGUUUCCUCCCUAC--UUAUUUCAGCCCGGCUUUUC-----GCCGGGCUUUUC--UUU
R0133/1-67  A---------------------GCCCUUUC--GGGCGUUUCCUCCCAAUGACU---CGGCCGCCCUCUUCUCCUGGAGGGCGGCCGUUCUUUU
Secondary Structure
            ..(((((((.....)))))))(((((......))))).....................(((((((((..........))))))))).........
Consensus
            GUGAGGCGGCGCUAGCGCCUCUGCCCUUCCUGGGGCGUUUCCUCCCUAUGA+UAU+UGGCCC+GCCUUU+UCCUGGGGC+GGGCUUUCUUUUU
```

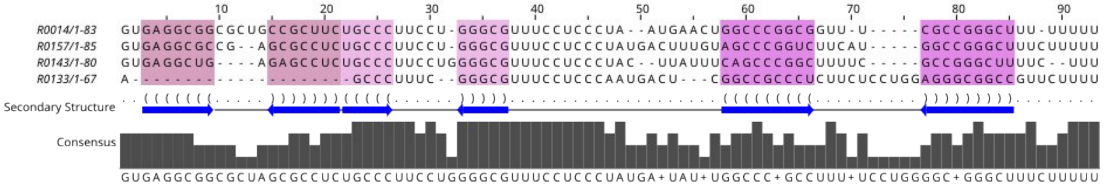


Northern blot analysis of *C. crescentus* wildtype grown in PYE, M2G (0.2% glucose), M2M (0.3% maltose), or M2X (0.2% xylose). Samples were taken at OD660 0.2, 0.5, 1.0, and overnight. Probes: R014, R133, R143, R157, loading control: tmRNA.

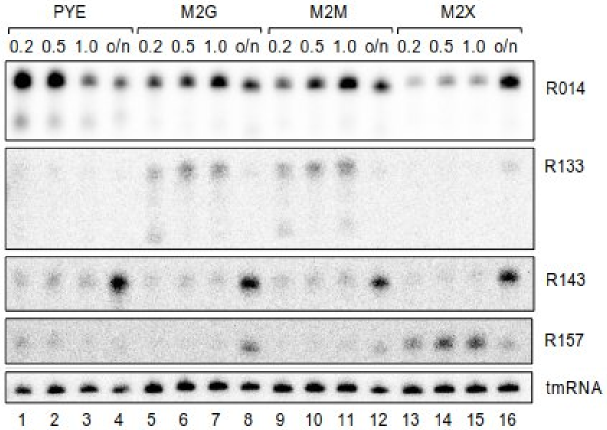
Dr. Kathrin Fröhlich
Junior Research Group
Bacterial RNA Biology

Why microsynteny?

Characterization of sibling sRNAs in *Caulobacter crescentus*

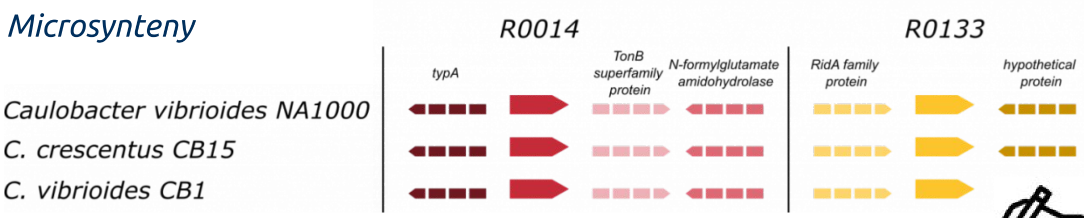


```
               10        20        30        40        50        60        70        80        90
R0014/1-83  GUGAGGCGGCGCUGCCGCUUCUGCCCUUCCU-GGGCGUUUCCUCCCUA--AUGAACUGGCCCGGCGGUU-U-----CGCCGGGCUUU-UUUUU
R0157/1-85  GUGAGGCGCCG--AGCGCCUCUGCCCUUCCU-GGGCGUUUCCUCCCUAUGACUUUGUAGCCCGGUCUUCAU-----GGCCGGGCUUUCUUUUU
R0143/1-80  GUGAGGCUG----AGAGCCUCUGCCCUUCCUGGGGCGUUUCCUCCCUAC--UUAUUUCAGCCCGGCUUUUC-----GCCGGGCUUUUC--UUU
R0133/1-67  A---------------------GCCCUUUC--GGGCGUUUCCUCCCAAUGACU---CGGCCGCCCUCUUCUCCUGGAGGGCGGCCGUUCUUUU
Secondary Structure
            ..(((((((.....)))))))(((((......))))).....................(((((((((..........)))))))))........
Consensus
            GUGAGGCGGCGCUAGCGCCUCUGCCCUUCCUGGGGCGUUUCCUCCCUAUGA+UAU+UGGCCC+GCCUUU+UCCUGGGGC+GGGCUUUCUUUUU
```



Northern blot analysis of *C. crescentus* wildtype grown in PYE, M2G (0.2% glucose), M2M (0.3% maltose), or M2X (0.2% xylose). Samples were taken at OD660 0.2, 0.5, 1.0, and overnight. Probes: R014, R133, R143, R157, loading control: tmRNA.

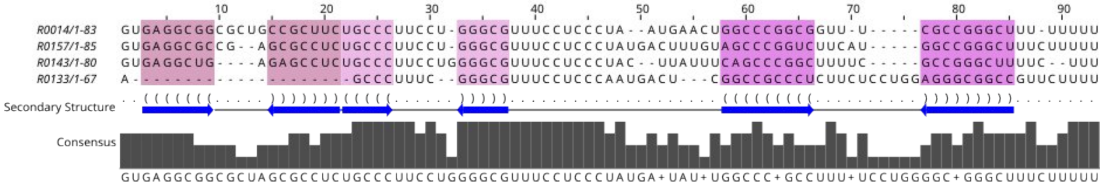
Microsynteny



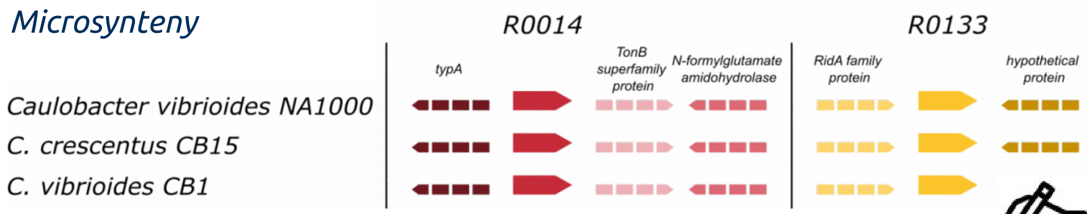
Dr. Kathrin Fröhlich
Junior Research Group
Bacterial RNA Biology

Why microsynteny?

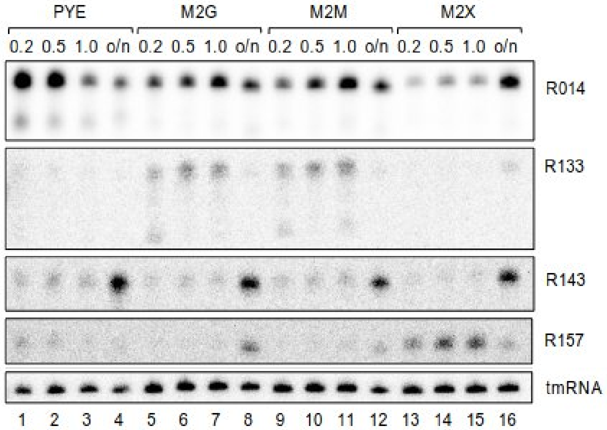
Characterization of sibling sRNAs in *Caulobacter crescentus*



Microsynteny



NOW: ***Alphaproteobacteria***
Only reference genomes: *#3097*



Northern blot analysis of *C. crescentus* wildtype grown in PYE, M2G (0.2% glucose), M2M (0.3% maltose), or M2X (0.2% xylose). Samples were taken at OD660 0.2, 0.5, 1.0, and overnight. Probes: R014, R133, R143, R157, loading control: tmRNA.

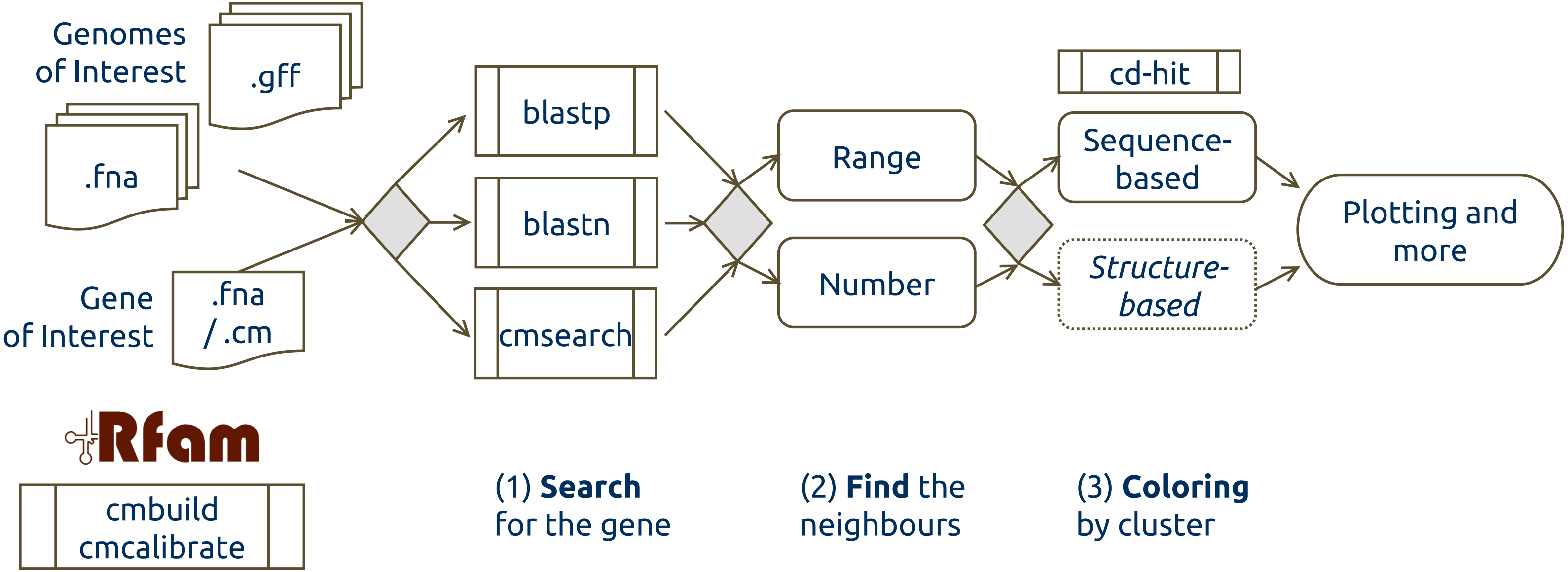
Dr. Kathrin Fröhlich
Junior Research Group
Bacterial RNA Biology

What is out there?

	Tool	**Why not?**
Web-based tools	MicroSyn	Hmm. We're having trouble finding that site
	SimpleSynteny	max 10 genomes with 1-10 contigs per file, no .svg
	Synteny Portal	pre defined species, no bacteria
	GeCoViz	only proteins
	genomicus	only 200 species, no bacteria
R package	gggenes, geneviewer, …	visualisation only
Perl tools	SynBlast	tblastn
	SMORE	repetitive elements, but genomic location figure
Python tools	GCsnap	blastp
	clinker	no search, protein level

Pipeline.nf overview

Tom Eulenfeld
PostDoc Programmer



Genomes
of Interest

.gff

.fna

Gene
of Interest

.fna
/ .cm

Rfam

cmbuild
cmcalibrate

blastp

blastn

cmsearch

Range

Number

cd-hit

Sequence-based

Structure-based

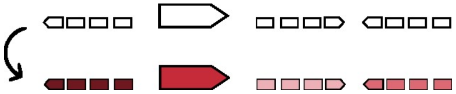
Plotting and more

(1) **Search**
for the gene

(2) **Find** the
neighbours

(3) **Coloring**
by cluster

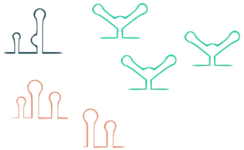
Coloring by cluster


Before

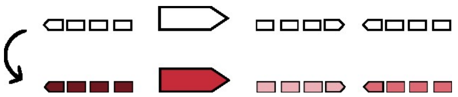
After

Coding RNA Sequence-based approach

Non coding RNA Sequence-based approach VS Structure-based approach



Coloring by cluster

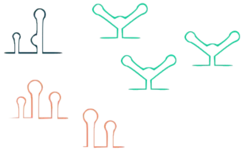


Before

After

Coding RNA Sequence-based approach

Non coding RNA Sequence-based approach vs Structure-based approach



Idea:
(1) RNAclust
(2) GraphClust
(3) cmscan

#target name	...	E-value	inc	description of target
suhB	...	$1.7*10^{-11}$	!	Makes More Granules Regulator RNA (mmgR)
anti-hemB	...	$1.2*10^{-05}$	!	Burkholderia RNA 7

Results for *Alphaproteobacteria*

```
Wolbachia             273
Sphingomonas          162
Rhizobium              91
Paracoccus             81
Bradyrhizobium         81
                     ...
Caulobacter           12
                     ...
Salinihabitans          1
Gimibacter              1
Aestuariibius           1
Aquidulcibacter         1
Chenggangzhangella      1

                      #3097
```

Results for *Alphaproteobacteria*

Genomes of Interest

.gff

.fna

Wolbachia	273
Sphingomonas	162
Rhizobium	91
Paracoccus	81
Bradyrhizobium	81
	...
Caulobacter	12
	...
Salinihabitans	1
Gimibacter	1
Aestuariibius	1
Aquidulcibacter	1
Chenggangzhangella	1
	#3097

Gene of Interest

new.cm

cmsearch

1786 genomes with 4112 significant* hits

~2.3 siblings per genome

(*E-value < 0.01)

Number 2,2

total seq: 19419

longest and shortest : 7749 and 50 nt

Sequence-based

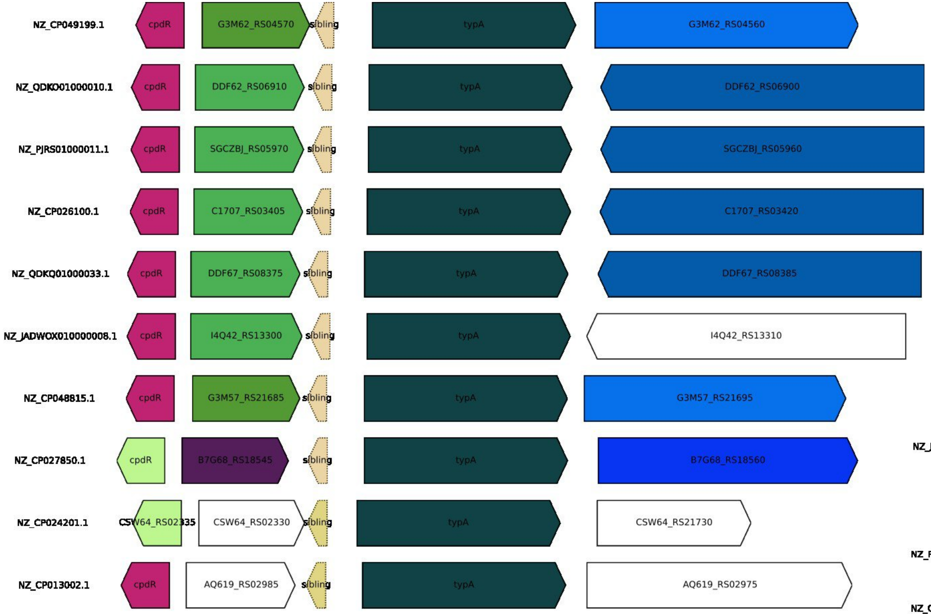
9833 #cluster

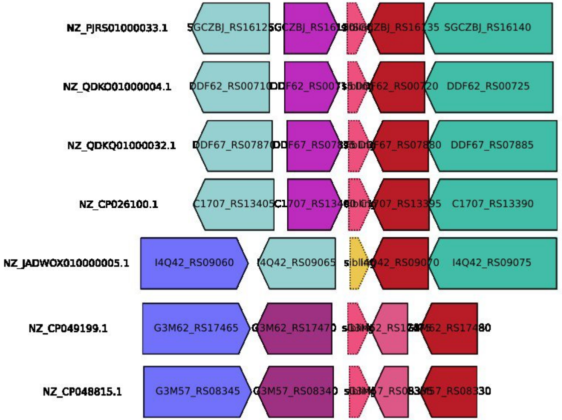
2139 #cluster with >1 sequence

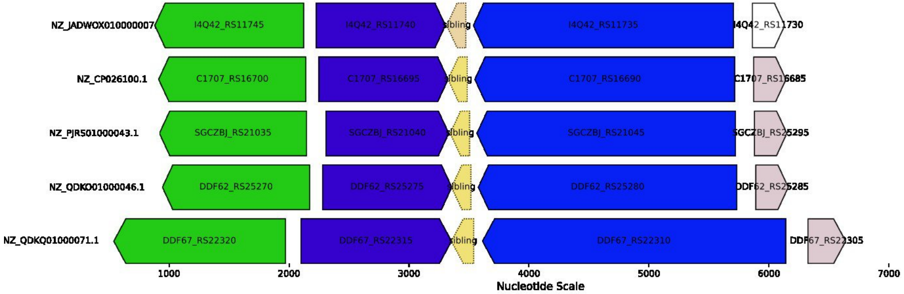
Nt similarity > 80%

#284 tRNAs
#4112 siblings

4112 genome context …





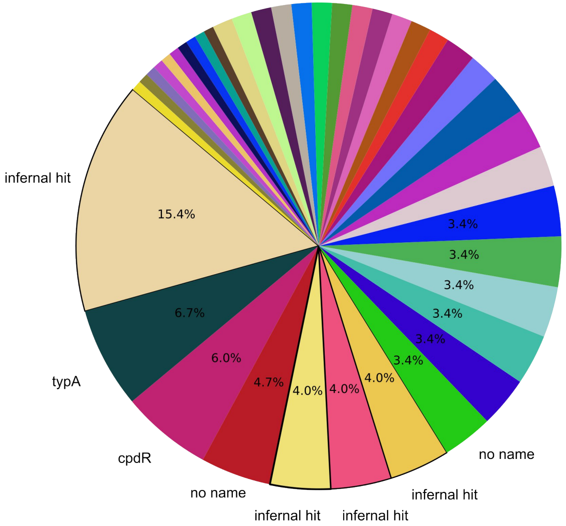


… and so on …

… No. But …

WORK IN PROGRESS

(1) General Statistics

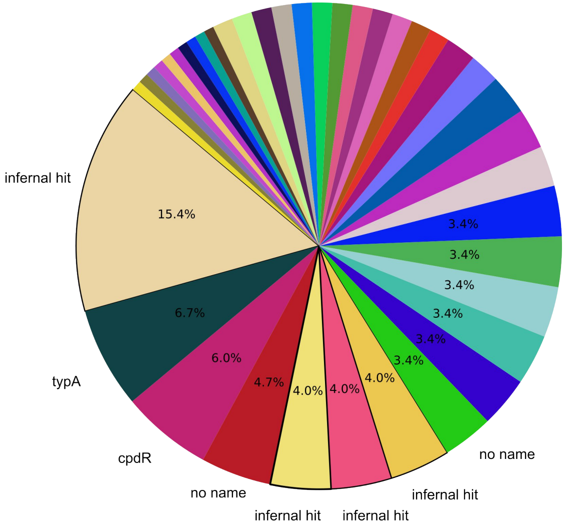


Clusters of neighbouring genes that show 80 % identity at the nt level in the *Caulobacter* reference genomes.

... No. But ...

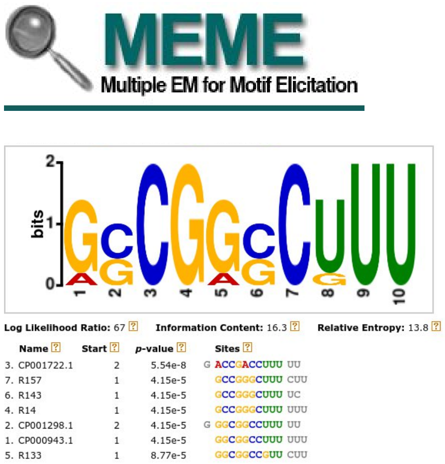
WORK IN PROGRESS

(1) General statistics



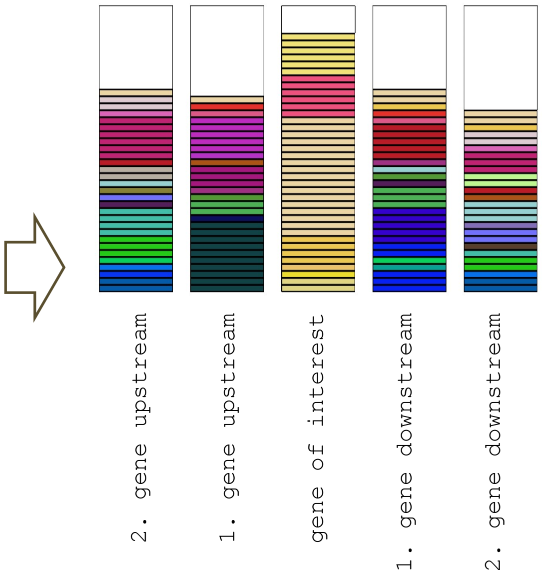
Clusters of neighbouring genes that show 80 % identity at the nt level in the *Caulobacter* reference genomes.

(2) Microsynteny motif



Log Likelihood Ratio: 67 ? **Information Content:** 16.3 ? **Relative Entropy:** 13.8 ?

Name ?	Start ?	*p*-value ?	Sites ?
3. CP001722.1	2	5.54e-8	G ACCGACCUUU UU
7. R157	1	4.15e-5	GCCGGGCUUU CUU
6. R143	1	4.15e-5	GCCGGGCUUU UC
4. R14	1	4.15e-5	GCCGGGCUUU UUU
2. CP001298.1	2	4.15e-5	G GGCGGCCUUU UU
1. CP000943.1	1	4.15e-5	GGCGGCCUUU UUU
5. R133	1	8.77e-5	GGCGGCCGUU CUU



Microsynteny motif in *Caulobacter* reference genomes

Limits and outlook

- ϟ Annotation / Genome dependency
- ϟ CM dependency

[...] More features like phylogenetic tree.

[...] Testing for eukaryotes.

[...] Investigation of microsynteny for toxin-antitoxin systems.

[...] Audience feedback.

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

RNA
BIOINFORMATICS & HIGH-THROUGHPUT ANALYSIS

So, as you can see, with SweetSynteny, your synteny analysis will be, well... *sweet*!

😅

Thank you for your attention!

Questions?