Additional content of the second state of the

COURSE on "METAGENOMICS" December 9-13, 2013 Campus de San Juan de la Universidad Miguel Hernández, Alicante, Spain

Marine microorganisms:

cultivation methods for improving their biotechnological applications

Topics covered

(1) Why doing Transcriptomics and Metatranscriptomics?

(2) How to do it?

(3) What kind of things can we find?

The organisms: *Trichodesmium* global distribution



Capone et al., 1997

Trichodesmium is an enigmatic organism: a major contributor of ,new' nitrogen, multicellular, forming large blooms, capabable of apoptosis and coordinated behavior [Rubin, Berman-Frank & Shaked "Dust- and mineral-iron utilization by the marine dinitrogen-fixer Trichodesmium" Nature Geoscience 4, 529–534]

picture: B. Bergman

December 13, 2011 vol. 108 no. 50 pp. 19837-20272

Proceedings of the National Academy of Sciences of the United States of America

Differentiation in cyanobacteria

Cytoprotective signaling in endothelial cells Ecology of traditional rice–fish agriculture Fertilization by acrosome-reacted sperm Meeting global food demands

Anabaena PCC7120



Variable pigmentation in cyanobacteria



Acaryochloris (Chlorophyll d)

Synechococcus (Phycoerythrin) Prochlorococcus (DV-Chlorophylls *a*+*b*)

Synechocystis PCC6803









Dienst et al. (2008) Microbiology 154, 3134-3143

(1) Why doing Transcriptomics and Metatranscriptomics?

Methods that provide comprehensive information about the status of gene expression: under different conditions



Synechocystis Complete Transcriptome Microarray sll0217-sll0219: flv4_flv2 operon, encoding an electron valve for photosystem II

Methods that provide comprehensive information about the status of gene expression in mutant lines



insertion mutant in crhR

Methods that provide comprehensive information about the status of gene expression in mutant lines





og2 norm. expresionvalue

Transcriptomic methods provide exact information about the status of gene expression under different conditions and in mutant lines and about the actually used TSS (promoters).



terminal differentiation, ~24 h

 NATURE REVIEWS
 MICROBIOLOGY

 Flores & Herrero
 VOLUME 8
 JANUARY 2010

dRNA-seq of the cyanobacterium Anabaena 7120



**alr3808 encodes a DpsA homologue with known Ndependent regulation (Ow et al., 2008).





dRNA-seq provides exact information about the TSS (promoters) & allows the inference of binding sites for regulatory proteins

		₁NtcAլ		, -10 ₁			+ tran	scription
2058877f	AGTTTAT	GTA ACCTATAAG <mark>AC</mark>	ATTTTATTTGATACCTCATACTC	ТААААТ	CAAGT <u>A</u>	nTSS		
2460393r	ТААААА	GTA ACTCTTGA <mark>TAC</mark>	ATACGCTTATGAAAACCGCATA	TACCAT	TGAAAAA <u>A</u>	gTSS	as12052	
580293f	ТААААСТ	GTAGCAATGCAG <mark>AC</mark>	TGTTGTTAGGAACAGTTATTAG	GAGAAT	GCGCCT <u>G</u>	gTSS	asr0485	(pipX)
1273249r	TCTTTTG	G GTACAAGATA <mark>TAC</mark>	AAAATAATATTGAGGAATTAGGC	TATCTT	CATATC <u>T</u>	gTSS	all1087	
5547631f	GTTTTTT	GTTGCGTGCTAG <mark>AC</mark>	ATAACCAGACGGGTGTTTTGATC	САААСТ	CCTGTA <u>A</u>	aTSS	all4644	
2059119f	TTATTT	GTATTTAACGGG <mark>AC</mark>	AGTTCTTACTTATCTAGTTAAGT	TTAAAT	AACAATC <u>A</u>	gTSS	alr1713	
2837125f	GTAGATA	G ATATCCACAA <mark>TAC</mark>	GGAAGTGTCAGTCTGATACTGG	CAGGCT	AAATT <u>A</u>	gTSS	alr2355	
5731963f	GTTTGTT	GGCGCAACGGC <mark>TAC</mark>	AGTTTGCTGGCGAGAGACAGGG	GATGAT	GGATTA <u>G</u>	aTSS	all4813	
4907756f	GCAAACT	GAATTGTTTGA <mark>TAC</mark>	GGCAGGATGTGCAGTTTTCTCT	TACCCT	GAGCAA <u>G</u>	gTSS	alr4077	
1693413r	АААААТ	GTAATCACGCTGAC	AGAACTATCGTCTGATTAGGAGG	ТАТААА	GTGATC <u>A</u>	gTSS	all1432	
3953418f	TGAGTTA	GTCGCTAAAGCTAC	ATTTTGGCTAACAGTATCCGACT	TATTAT	GAGATTT <u>A</u>	aTSS	all3278	
2400767r	GTTGCTC	GTA TATTTCAAC <mark>AC</mark>	GAATTTGATCATTTAGATGGTG	TACTGT	TTATAG <u>A</u>	gTSS	all2006	
519953f	ACATAAC	GT GTTTTCAGT <mark>TAC</mark>	AGTTATGCCAGATGCAATTAAGC	CACAAT	GTTGATT <u>A</u>	gTSS	alr0440	
105428r	CATTATG	GTATGAAATAG <mark>TAC</mark>	AGTTTAAAATTAGTGTTTGCGT	CATCAT	TACGAG <u>A</u>	gTSS	all7614	(beta)
1657401r	GAGAGTC	GTA GCATAACAC <mark>AC</mark>	TAAAACTTCTGGAAACAGTAGGT	TAGGCT	tgcctt <u>a</u>	gTSS	all1395	
3346518f	АТАААСТ	G ATAGTTATAA <mark>TAC</mark>	TGTTCTCAGAAACGAAAAACTA	TATATT	GAGCAT <u>A</u>	nTSS		
5248514r	TGTTTTT	GCGATCGGCGA <mark>TAC</mark>	AATTTACACGGGGCAAAAGCTG	GAATAT	GAAGGA <u>A</u>	iTSS	all4379	
5167792r	GGCTAGA	GTAACAAAGACTAC	AAAACCTTGGGCATGGGCTTGT	TACTTT	GAAATTC <u>A</u>	gTSS	all4312	(nrrA)
5407066f	CTCAGCA	ATTTGTTCAACCTGA	GCATTTTTCCCATTTGCAACTTGA	TACAAA	TATTTTT <u>A</u>	gTSS	asr4517	(nblA)
2793917r	CTTCCTC	ACTGCTCATACAGA	GCAGATACGGTTAAAAAAAGTTGC	AATTCT	CATAAGT <u>G</u>	gTSS	all2319	(glnB)

dRNA-seq of the cyanobacterium Anabaena 7120

Mitschke et al. (2011b) Proc. Natl. Acad. Sci. USA, 108, 20130-20135.

Metatranscriptomics & the typical question of microbial ecology:

Who is there and what are they doing?

Methods that provide comprehensive information about the status of gene expression in populations

16S rDNA sequencing for bacteria and archaea RNA seq / dRNA seq **SW Pacific**



16S rDNA sequencing for bacteria dRNA seq / RNA seq **Red Sea**



Red Sea sampling 2012



16S rRNA-based diversity

objective: nitrogen cycle

T12/13 24h sampling M t0 t1 t2 t3 t4 t5



VAHINE total RNA

SW Pacific 2013



collaboration with Sophie Bonnet

VAHINE sampling 2013





16S rRNA-based diversity

objective: nitrogen cycle

T12/13 24h sampling M t0 t1 t2 t3 t4 t5



VAHINE total RNA

SW Pacific 2013



collaboration with Sophie Bonnet

Methods that provide comprehensive information about the status of gene expression: finding new players



Synechocystis Complete Transcriptome Microarray

PsrR1 (SyR1), Photosynthesis regulatory RNA1

An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803

ian Mitschke^h, Jens Georg^h, Ingeborg Scholz^a, Cynthia M. Sharma^b, Dennis Dienst⁶, Jens Bantscheff⁴, Björn Voß⁴, Claudia Steglich⁴, Annegret Wilde⁴, Jörg Vogel⁴, and Wolfgang R. Hess^{4,6,2}

Methods that provide comprehensive information about the status of gene expression: finding new players



Synechocystis Complete Transcriptome Microarray

PsrR1 (SyR1), Photosynthesis regulatory RNA1

An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803

lan Mitschke^h1, Jens Georg^{h,1}, Ingeborg Scholz^a, Cynthia M. Sharma^b, Dennis Dienst⁶, Jens Bantscheff⁴, Björn Voß⁴, Claudia Steglich⁴, Annegret Wilde⁴, Jörg Vogel⁴, and Wolfgang R. Heis^{6,6,2}

Methods that provide comprehensive information about the status of gene expression: finding new players by dRNA-seq



Synechocystis Double-comparative dRNA-seq Analysis of ten different conditions

PsrR1 (SyR1), Photosynthesis regulatory RNA1

Are these new players relevant at all?



Pigmentation phenotype of PsrR1 overexpression in *Synechocystis*



An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803

ian Mitschke^h1, Jens Georg^{h1}, Ingeborg Scholz^a, Cynthia M. Sharma^b, Dennis Dienst⁴, Jens Bantscheff⁴, Björn Voß⁴, Claudia Steglich⁴, Annegret Wilde⁴, Jörg Vogel⁴, and Wolfgang R. Hess^{4-6,2}

Are these new players relevant at all?

Judged by

- their numbers (2/3 of all promoters give rise to non-mRNA transcripts),
- their regulation (explosite in the interview),
- and the phenotypic effects when they are mutated or overexpressed (exp., PsrR1 overexpression causes disturbed pigmentation and photosynthesis).

(2) How to do it?

2.1 Computational predictions

2.2 Microarrays

2.3 RNA-seq



2.1 Computational predictions and validation in experiments

These RNA players are abundant.



Why are they not just annotated during normal genome annotation?

Why are they not just annotated during normal genome annotation?



Scheme of a bacterial protein-coding gene

Why are they not just annotated during normal genome annotation?



Scheme of a bacterial non-protein-coding gene

Alignment of PsrR1 from 28 different cyanobacteria



Usually, the sequence conservation of non-coding transcripts is not very good,

but the secondary structure conservation is better and might be utilized:

Complementary mutations indicate non-coding RNAs



Comparison of 4 Prochlorococcus and 4 Synechococcus



Sequence/structure model for Yfr1 of 31 cyanobacteria.

Voss et al., (2007) BMC Genomics 8:375,

(R: A or G; Y: C or U; M: A or C; S: G or C; B: G, U or C; V: G, C or A; D: G, T or A). Colors: number of different base pairs at this position (red = 1, yellow = 2, green = 3 and blue = 4 or more). Shading: frequency of base pairing.



Syn. 7942	Tsyn. S elong.	Sync. 6803 7	Mic. 7806	Ana. 7120	Nos. punct.	Gloe. 7421
<u>Y</u> †	<u>r1 is</u>	<u>sa</u>	<u>rea</u>	<u>al s</u>	<u>KN</u>	<u>\:</u>
		-				
Yf	r1			1		



But, how to find out the function of a bacterial sRNA in a marine microbe that can't even be manipulated ?


Functional analysis of Yfr1 – target prediction and experimental verification

Synechococcus elongatus PCC 6301

- → homozygous mutants in S. elongatus PCC 6301 (Nakamura et al., 2007)
- \rightarrow reduced growth under various stress conditions, e.g. oxidative stress and high salt stress conditions
- → about 18,000 copies/cell (vs. 146,400 ribosomes/cell)
- \rightarrow putative target of Yfr1 in *Synechococcus* is *sbtA* (C transporter)

Synechocystis sp. PCC 6803 (work of Annegret Wilde Univ. Giessen)

- \rightarrow Yfr1 is essential in *Synechocystis* sp. PCC 6803
- \rightarrow mutants are not fully segregated
- \rightarrow strong growth effects in the mutants under light stress

Prochlorococcus

 \rightarrow ~ 100 molecules per cell (vs. ~ 2,000 ribosomes/cell)



Target prediction with INTARNA

http://rna.informatik.uni-freiburg.de:8080/IntaRNA/Input.jsp

 PMM1119 (som) mRNA
 energy: -13.4 kcal/mol

 * SD
 5' - ACUCAAAUUGUGUGAGGAUUUUUUAUGAAGCUUUUU...-3'

 []]]]]]]
 3' - UUUUUUCGGGCUAUUUAGCCCGCUAAACCACACACUCCUCAUACCCCAAAGGGGGUA-5'

Yfr1

Yfr1

PMM0538 mRNA energy: -8.1 kcal/mol * SD 5' - AAAUAUAACGGAGAUUAUUUUU<u>GAGG</u>AGU<u>UUG</u>CAAAUUUUU...-3' |||||| 3' - UUUUUUCGGGCUAUUUAGCCCGCUAAACCACACACUCCUCAUACCCCAAAGGGGGUA-5'

Yfr1

PMM1697 mRNA energy: -9.2 kcal/mol 5'-AAUCCACUUAAAGAGGCCAGG GUG^AUGGGGAUCCUU...-3' || ||| || 3'-UUUUUUCGGGCUAUUUAGCCCGCUAAACC_ACAC ACUCCUCAUACCCCAAAGGGGGUA-5'

Yfr1



GFP reporter system



Johannnes Urban & Jörg Vogel, NAR 2007





predicted structures of WT Yfr1 and mutated Yfr1





Synechocystis classical array (Eisenhut, et al. (2007) *Plant Physiol* <u>144</u>:1946-59)



Types of microarrays



Synechocystis PCC6803: "1/3 genome tiling array" (Georg et al., MSB 2009)



Types of microarrays: Prochlorococcus Affymetrix high density microarray





4 genomes (MED4, MIT9313, P-SSP7, P-SSM4) 25 base oligomers (PM,MM) Orfs,RNAs: 11 probes, or every 80 bases IG: 4 probes, or every 45 bases

In total: ~3000,000 probes



Most classical microarrays work with cDNA that may cause problems:





"Intergenic spacer only" array *Synechococcus* WH 7803 and direct RNA labeling (tiling factor =11, sense + antisense); no protein-coding genes; Format: 12K



cold stress RNA

control RNA

www.nature.com/ismej

ORIGINAL ARTICLE

Non-coding RNAs in marine *Synechococcus* and their regulation under environmentally relevant stress conditions

Gregor Gierga, Björn Voss and Wolfgang R Hess Faculty of Biology, University of Freiburg, Freiburg, Germany



A 105K Agilent tiling array covering 1/3 of the *Synechocystis* genome and with probes on both strands was hybridized against directly labelled total RNA



Synechocystis 6803 array

Example: *isiA* mRNA // IsrF antisense RNA

Other pitfalls:

RNA half life

RNA is not very stable as the gene as the gene expression profile is constantly changing in a cell.

You don't want the gene epxression on deck of the ship to be measured (or in the dark centrifuge either)





RNA half life - What do you guess is the median half life of mRNA in *Prochlorococcus*?





How to measure half-life time?

- rifampicin binds to beta subunit of RNA polymerase
- prevents initiation of new transcripts
- RNA stability can be measured



Affymetrix high density microarray



slide: C. Steglich



Shortest global half-life ever measured for any organism

global half-life of 2.4 min

cell doubling vs. half-life



slide: C. Steglich

2.3 RNA-seq (Isolation and sequence analysis of all small RNAs; "dRNA-seq", "RNomics", 454, Solexa)

Transcriptome sequencing



RNAseq mainly sequences cDNA of ribosomal RNA



dRNA-seq and rRNA depletion

Enzymatically - Terminator[™] Exonuclease (TEX):

- In the cell rRNA is processed from a primary transcript → rRNA maturation
- Primary transcripts: 5'-PPP
- Processed transcripts: 5'-P, 5'-OH (don't ligate)

он) 5′ ЛОЛОЛО - 3



Differential RNAseq (dRNA-seq) according to Sharma et al. (2010)

TEX (5'PPase)-treatment removes not only rRNA, but all processed/degraded RNAs.

⇒ Enrichment for primary transcripts
→ mapping of TSS

dRNA-seq:

• 2 libraries: TEX-treated vs. untreated



Difference between TEX/TAP-treated and untreated RNA (dRNA-seq versus RNA-seq) Window size: 3 secondary reads primary reads

Example: Total RNA from Synechocystis 6803 (high light cultures)
*
Transcriptional start site

NA

s1r2075

groELl

slr2076

(3) What kind of things can we find?

The primary transcriptome of Trichodesmium



non-coding potential





Tri<mark>chodesmium e</mark>rythraeum

highest proportion of non-coding nucleotides amongst all sequenced cyanobacteria to date;

some analogy to eukaryotic genomes

mapping of split transcriptome reads in Trichodesmium

36 selected bases on reverse strand: 7084..7119 = complement (7742990..7743025)



Interaction with bacteriophages: Identification of antiviral CRISPR expression (*Synechocystis*)



63

CRISPR in Synechocystis PCC6803



454-sequencing results



CRISPR-derived crRNAs are among the most highly expressed transcripts in the cell





Podoviridae: - short tails that are non contractile



- linear double stranded genome (~ 40 kbp, ~ 55 genes)

cyanophage P-SSP7, picture Bin Ni, Chisholm Lab

- Myoviridae: non enveloped viruses that consist of a head and a tail separated by a neck
 - linear double stranded genome (30-170 kbp, 200-300 genes)

cyanophage P-SSM4, picture Bin Ni, Chisholm Lab

Siphoviridae: - consist of a head and non contractile tail - linear double stranded genome (~ 50 kbp, ~ 70 genes)



cyanophage MIT9313-4, picture Matt Sullivan, Chisholm Lab



- asRNAs modulate gene expression through the protection of mRNAs from RNase E cleavage
- in T7 and T4 phage RNase E activity is regulated by phosphorylation of the C-terminus (not present in *Prochlorococcus* RNase E)
- indirect regulation of RNase E activity in *Prochlorococcus* by RNA duplex formation reflects an alternative mechanism of regulation of enzyme activity

Pigmentation phenotype of PsrR1 overexpression in *Synechocystis*



An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803

ian Mitschke^h1, Jens Georg^{h1}, Ingeborg Scholz^a, Cynthia M. Sharma^b, Dennis Dienst⁴, Jens Bantscheff⁴, Björn Voß⁴, Claudia Steglich⁴, Annegret Wilde⁴, Jörg Vogel⁴, and Wolfgang R. Hess^{4-6,2}

PsrR1 is a conserved sRNA

			interactio	n region	terminator hairpin		
Nostoc_azollae_0708 Anabaena_sp_90 Anabaena_variabilis_ATCC29413 Nostoc_sp_PCC7107 Calothrix_sp_PCC7507 Stanieria_variabilis_ATCC29413 Cyanothece_sp_PCC7807 Stanieria_vanosphaera_PCC7437 Cyanothece_sp_PCC8801 Anabaena_sp_PCC7120* Nostoc_punctiforme_PCC73102 Cylindrospermum_stagnale_PCC7417 Pleurocapsa_sp_PCC7327 Microcystis_aeruginosa_NIES843 Cyanothece_sp_PCC7824 Cyanothece_sp_PCC7824 Cyanothece_sp_PCC7824 Cyanothece_sp_PCC7822 Microcystis_aeruginosa_NIES843 Synechocystis_sp_PCC6714* Cyanothece_sp_PCC7822 Microcystis_aeruginosa_NIES843 Synechocccus_sp_PCC7822 Microcystis_aeruginosa_NIES843 Synechocccus_sp_PCC7822 Microcystis_aeruginosa_NIES843 Synechocccus_sp_PCC7822 Microcystis_aeruginosa_NIES843 Synechocccus_sp_PCC7822 Microcystis_aeruginosa_NIES843 Synechocccus_sp_PCC7828 Leptolyngbya_sp_PCC7376 Chroacoccidiopsis_thermalis_PCC72033 Geitlerinema_sp_PCC7407 Dactyloccccopsis_salina_PCC8305 Halothece_sp_PCC7418	wariable 5' region	AAATGI TTI CACCTTCGGTAT AGATITTI GC CTCGGTAT AGATITTI GC CTCGGTAT 	G	$ \begin{array}{c}$		-CCTCGGTCAGT T -CCTGGTCAGT T -CCTCGGTAGT T -CCTCGGTTAGT T -CCTCGGTTAGT -CCTCGGTTAGT -CCTGGTTA 	CACCGGGG TTTT CGCCAGGGTTTTT CGCCGGGGTTTTT CGCCGGGGTTTTT CGCCGGGGTTTTT



How bacterial sRNAs work: repression of gene expression



From: Corcoran *et al.* (2011), Hfq-associated regulatory small RNAs. In: Hess W.R. & Marchfelder A. (Editors) Regulatory RNAs in Prokaryotes. Springer-Verlag Wien New York.

Prediction of PsrR1 interaction sites within the mRNA targets



CopraRNA: Comparative Prediction Algorithm for sRNA Targets: PNAS Plus (2013) 110 (37), E3487-E3496.

The combination of different transcriptomic approaches allows to:

- measure gene expression in a comparative mode
- map precisely the suite of active promoters
- infer regulatory sequence elements
- derive excact information on transcript boundaries, i.e. define operons and transcript isoforms
- find highly transcribed regions of special importance, e.g.
 CRISPRs, the prokaryotic immune system
- identify previously unknown transripts, such as non-coding sRNAs, and
- PsrR1 is an sRNA regulator of *psaL* and likely of several more photosynthesis-associated genes in cyanobacteria




Special thanks to:

Jens Georg, Matthias Kopf, Steffen Lott, Jan Mitschke, Ulrike Pfreundt, Ingeborg Scholz, Damir Stazic, Claudia Steglich, Karsten Voigt, Björn Voss

More thanks to:

Agustín Vioque, Alicia Muro-Pastor (CSIC & Univ. Sevilla) Sophie Bonnet (MIO CNRS/Aix-Marseille University & IRD Noumea) Debbie Lindell (Technion, Israel) Ilana Berman-Frank (Bar-Ilan-Univ. Israel)

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 311957.

www.macumbaproject.eu





AUGUST 3 - 8, 2015 FREIBURG

15th international SYMPOSIUM on PHOTOTROPHIC PROKARYOTES