Rynos summer project 2017**:**

**Seasonal patterns of gene expression in Baltic Sea bacterioplankton**

**Background**

Although bacterioplankton are the main drivers of many biogeochemical cycles in the oceans we still know relatively little about how their activities are regulated and distributed among taxonomic groups. In this project gene expression data will be analysed from surface water samples collected at different time points during three years (2012-2014) at an offshore station east of Öland (Station LMO) in the Baltic Sea. RNAseq (metatranscriptome) and metaproteome data have been mapped to the genomes of a set of 83 genomes that have been reconstructed from metagenomes from the same station. The dataset may potentially give insights into how gene expression in the different populations change with environmental conditions, as well as how functional activities varies between taxonomic groups. Moreover the data gives a unique chance to compare expression levels at the transcript and protein level in bacteria in nature.

**Examples of scientific questions that could be addressed**

* Are there any significant differences in gene expression of specific functional genes and/or groups of functionally related genes (e.g. of same metabolic pathway) between taxonomic groups (averaged over time)?
* Are there any significant correlations in gene expression of specific genes within MAGs and environmental parameters (nutrient levels, temperature, etc.)?
* Are there any seasonal patterns in gene expression within MAGs?
* Are there any correlations between RNA and protein expression levels?
  + What factors influence the strength of these correlations (absolute expression levels, etc.)?
* Are there any correlations in changes in RNA and changes in protein expression levels?

**Data and methodology**

Data

* Metatranscriptome data from (25 samples):
  + (Seq\_run1) 120516,120613,120712,120813,120927,121024,121220
  + (Seq\_run1) 130123,130226,130403,130416,130422
  + (Seq\_run2) 130507,130605,130705,130815,130905,131003
  + (Seq\_run2) 140408,140506,140604,140709,140820,140916,141013
* Metaproteome data from (17 samples, 12 overlap with MT):
  + (Proteomics\_run1) 120512,120613,120712,120813,120927,121024,121220
  + (Proteomics\_run2) 130422, 130507, 130705, 130815, 130905, 131003, 131126
  + (Proteomics\_run2) 140109,140220,140408
* The data has been mapped to 83 metagenome-assembled genomes (MAGs) of Hugerth et al 2015
* Genes have been annotated with COG, TIGRFAM, EC and PFAM numbers
* Tables have been generated with gene expression in every MAG per gene per sample
* Table with environmental data for every sample available

Methods

* Differential expression analysis using SAMseq and/or DESeq2 in Ror Python (not these can be run inside Python though)
  + Comparisons between (e.g. taxonomic) groups
  + Correlations to environmental parameters

**Expected output**

* Short report describing the findings illustrated with graphs and tables.

**Some background reading**

[Metagenome-assembled genomes uncover a global brackish microbiome.](https://www.ncbi.nlm.nih.gov/pubmed/26667648)

**Hugerth LW**, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, Andersson AF.

Genome Biol. 2015 Dec 14;16:279. doi: 10.1186/s13059-015-0834-7.

PMID:26667648

[Quantitative analysis of a deeply sequenced **marine** microbial **metatranscriptome**.](https://www.ncbi.nlm.nih.gov/pubmed/20844569)

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA.

ISME J. 2011 Mar;5(3):461-72. doi: 10.1038/ismej.2010.141. Epub 2010 Sep 16.

PMID:20844569

[Microbial community transcriptional networks are conserved in three domains at ocean basin scales.](https://www.ncbi.nlm.nih.gov/pubmed/25775583)

Aylward FO, Eppley JM, Smith JM, Chavez FP, Scholin CA, DeLong EF.

Proc Natl Acad Sci U S A. 2015 Apr 28;112(17):5443-8. doi: 10.1073/pnas.1502883112. Epub 2015 Mar 9.

PMID:25775583

[Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages.](https://www.ncbi.nlm.nih.gov/pubmed/25013074)

Ottesen EA, **Young** CR, Gifford SM, Eppley JM, Marin R 3rd, Schuster SC, Scholin CA, **DeLong** EF.

Science. 2014 Jul 11;345(6193):207-12. doi: 10.1126/science.1252476.

PMID:25013074

[Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom.](https://www.ncbi.nlm.nih.gov/pubmed/22556258)

**Teeling** H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R.

Science. 2012 May 4;336(6081):608-11. doi: 10.1126/science.1218344.

PMID:22556258

[Expression patterns of elemental cycling genes in the Amazon River Plume.](https://www.ncbi.nlm.nih.gov/pubmed/28387773)

Satinsky BM, Smith CB, Sharma S, Landa M, Medeiros PM, Coles VJ, Yager PL, Crump BC, **Moran MA**.

ISME J. 2017 Apr 7. doi: 10.1038/ismej.2017.46. [Epub ahead of print]

PMID:28387773

[**Gene-specific** **correlation** of **RNA** and **protein** **levels** in human cells and tissues.](https://www.ncbi.nlm.nih.gov/pubmed/27951527)

Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, Forsström B, Uhlén M.

Mol Syst Biol. 2016 Oct 20;12(10):883. doi: 10.15252/msb.20167144.

PMID:27951527

[Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data.](https://www.ncbi.nlm.nih.gov/pubmed/22127579)

**Li J**, **Tibshirani R**.

Stat Methods Med Res. 2013 Oct;22(5):519-36. doi: 10.1177/0962280211428386. Epub 2011 Nov 28.

PMID:22127579

[Moderated estimation of fold change and dispersion for RNA-seq data with **DESeq2**.](https://www.ncbi.nlm.nih.gov/pubmed/25516281)

Love MI, **Huber W**, Anders S.

Genome Biol. 2014;15(12):550.

PMID:25516281