

Effect of Crude Oil and Corexit 9500 on Gene Expression in *Cinachyrella*

Polina Drozdova, Elizaveta Savochkina
Supervised by: Mike Rayko, Aleksey Komissarov

Bioinformatics Institute,
Theodosius Dobzhansky Center for Genome Bioinformatics

13 Dec 2014

Initial data

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

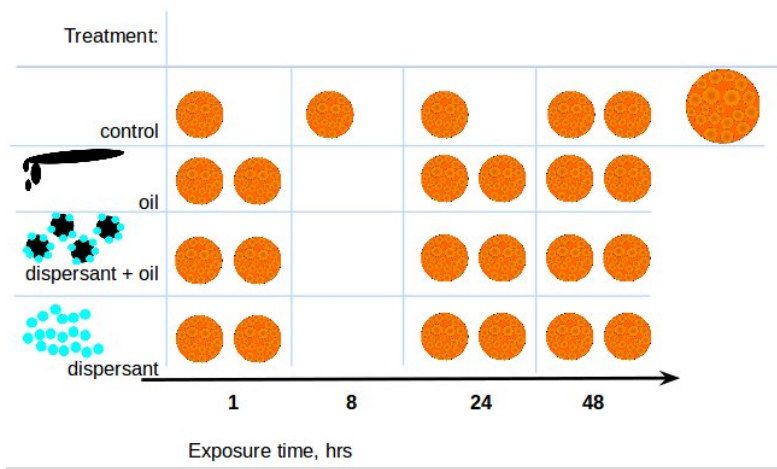
Introduction

Goals

Exploratory
Analysis

Results

Conclusions



Goals

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

Introduction

Goals

Exploratory
Analysis

Results

Conclusions

- Quality control of the data with FastQC.
- Filter out adapters and low quality reads with Cookiecutter.
- Find bacterial reads in transcriptomes with Kraken and filter them out of the reference transcriptome somehow.
- Reassemble the reference transcriptome with Trinity and annotate it with Trinotate.
- Align all samples to the assembly with RSEM.
- Find differentially expressed transcripts with edgeR.
- Annotate the differentially expressed transcripts with Trinotate.

QC: FastQC

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

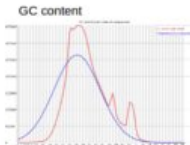
Introduction

Goals

Exploratory
Analysis

Results

Conclusions



Control	!	!	!	!	!	✓
Oil	✗	✗	!	!	✗	
Oil-Dispersant	✗	✗	✗	!	✗	
Dispersant	✗	✗	✓	!	✓	

Bacterial transcripts: Kraken

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

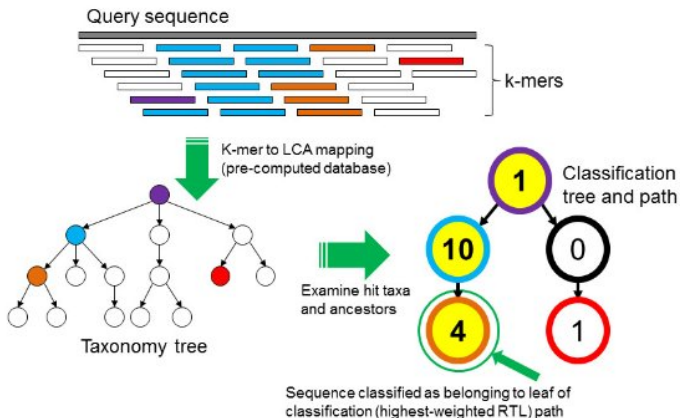
Introduction

Goals

Exploratory
Analysis

Results

Conclusions



Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15: R46. <http://genomebiology.com/2014/15/3/R46>

Bacterial transcripts: Kraken

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

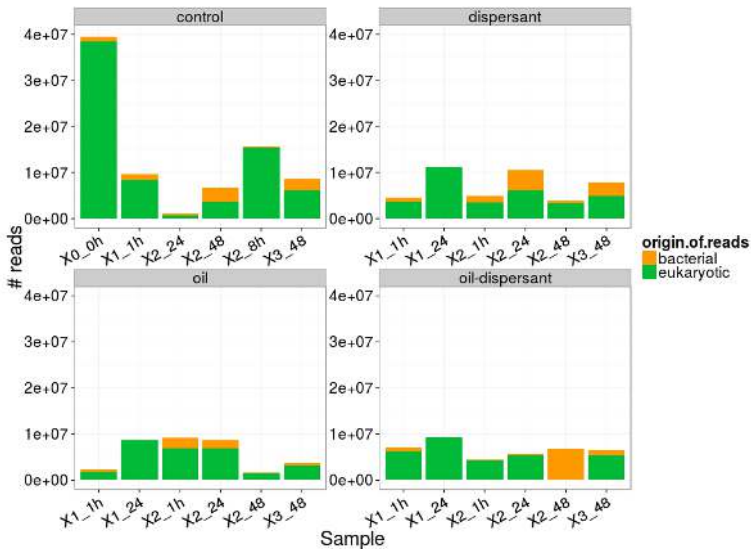
Introduction

Goals

Exploratory
Analysis

Results

Conclusions



Transcriptome assembly: Trinity

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

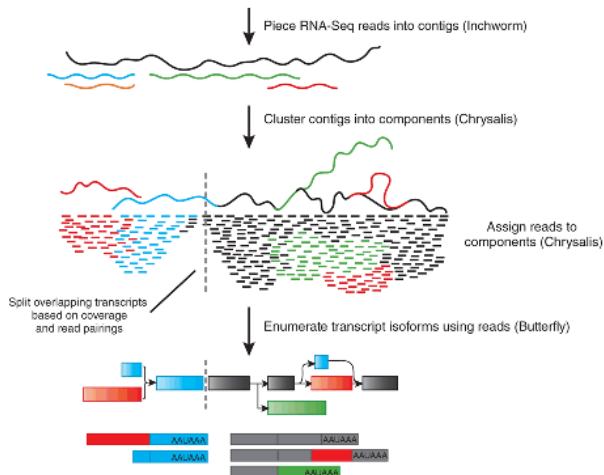
Introduction

Goals

Exploratory
Analysis

Results

Conclusions



Grabherr, Manfred G., et al. "Full-length transcriptome assembly from RNA-Seq data without a reference genome." *Nature biotechnology* 29.7 (2011): 644-652.

Evaluating Trinity Output: QCAST

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

Introduction

Goals

Exploratory
Analysis

Results

Conclusions

Statistics without reference	Trinity
# contigs	15643
# contigs (≥ 0 bp)	34226
# contigs (≥ 1000 bp)	7589
Largest contig	11691
Total length	19494471
Total length (≥ 0 bp)	25217219
Total length (≥ 1000 bp)	13783987
N50	1454
L50	4223
GC (%)	51.16
Predicted genes	
# predicted genes (unique)	18074
# predicted genes (≥ 300 bp)	15264
# predicted genes (≥ 1500 bp)	2645
# predicted genes (≥ 3000 bp)	384

Evaluating Trinity Output: CEGMA

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

Introduction

Goals

Exploratory
Analysis

Results

Conclusions

	#Prots	%Completeness	-	#Total	Average	%Ortho
Complete	184	74.19	-	370	2.01	58.15
Group 1	37	56.06	-	65	1.76	54.05
Group 2	36	64.29	-	89	2.47	66.67
Group 3	53	86.89	-	107	2.02	64.15
Group 4	58	89.23	-	109	1.88	50.00
Partial	210	84.68	-	423	2.01	56.67
Group 1	46	69.70	-	82	1.78	56.52
Group 2	45	80.36	-	102	2.27	55.56
Group 3	56	91.80	-	120	2.14	66.07
Group 4	63	96.92	-	119	1.89	49.21

Annotating the Transcriptome: Trinotate

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

Introduction

Goals

Exploratory
Analysis

Results

Conclusions

Total number of transcripts: 35,363

Number of known Eukaryotic transcripts: 16,365

Number of known transcripts of Porifera: 115

Alignment to Reference: RSEM

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

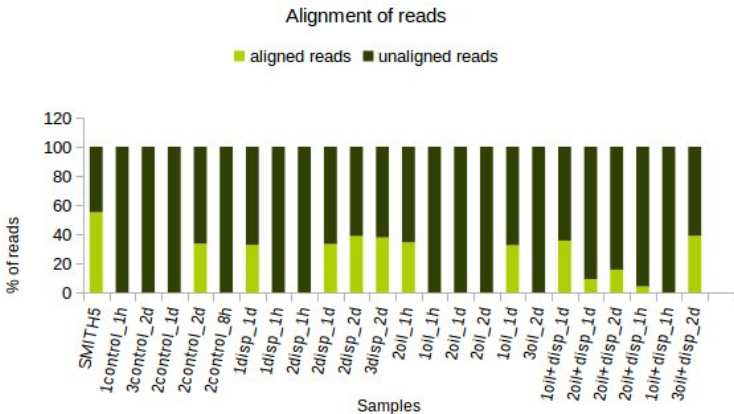
Introduction

Goals

Exploratory
Analysis

Results

Conclusions

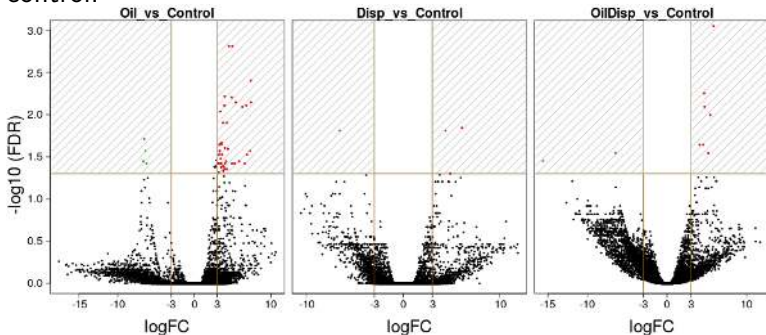


Differential Expression: edgeR

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

Chose 11 samples, fitted GLM and made 3 pairwise contrasts:
oil vs. control, dispersant vs. control, oil + dispersant vs.
control.



Introduction

Goals

Exploratory
Analysis

Results

Conclusions

Conclusions

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

Introduction

Goals

Exploratory
Analysis

Results

Conclusions

- Adapters and low quality bases were removed from reads.
- We found bacterial reads in transcriptomes with Kraken and filtered them out of the reference sample.
- The reference transcriptome was assembled with Trinity and annotated with Trinotate.
- All samples were aligned to the reference assembly with RSEM.
- Differentially expressed transcripts were found with edgeR, and Trinotate annotations were assigned to these transcripts.

Project overview & Distribution of work

Corexit and
Gene
Expression

Polina
Drozdova,
Elizaveta
Savochkina

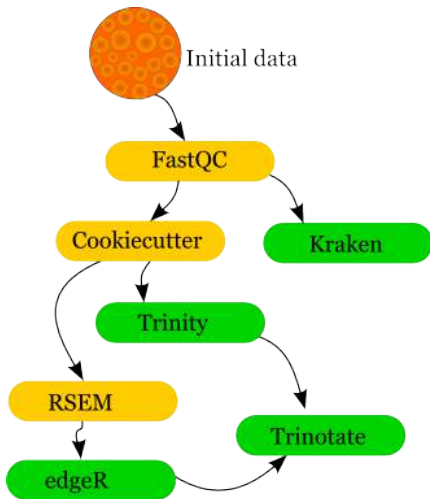
Introduction

Goals

Exploratory
Analysis

Results

Conclusions



Polina



Liza

Thank you for your attention!



© Nickelodeon