The analysis of foreign sequences in the Hydra genome

Thomas Weinmaier
Department of Computational Systems Biology

Thanks to

Thomas Rattei

Roland Arnold

Patrick Tischler

Charles David

Oleg Simakov

Prakash Balasubramanian

Jarrod Chapman

Uffe Hellsten

Rob Steele

Hans Bode

TUM

JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

THE UNIVERSITY OF CALIFORNIA
IRVINE

LMU

Outline

1. ***Hydra* basics**

2. **Identification of a bacterial genome in the *Hydra* assembly**

3. **Investigation of other foreign sequences in the *Hydra* genome**

The freshwater polyp *Hydra* - basics

- Size 10 ... 30 mm
- Reproduction: mostly asexual by budding
- One of the oldest and simplest animals
- Model organism in developmental biology
  - → Simple organization in three layers
  - → Enormous capacity of regeneration
- Genome size about 1 Gb on 15 chromosomes

The *Hydra* genome project

- Sequencing/Assembly: J. Craig Venter Institute / UC Berkeley
- 8-fold coverage
- ~ 900 Mb in 21.000 contigs
- ~ 163.000 Expressed Sequence Tags (ESTs)
- ~ 20.000 gene models
- ~ 29% G+C
- ~ 60% repeats
- Analysis: international consortium incl. LMU/TUM and the University of Vienna

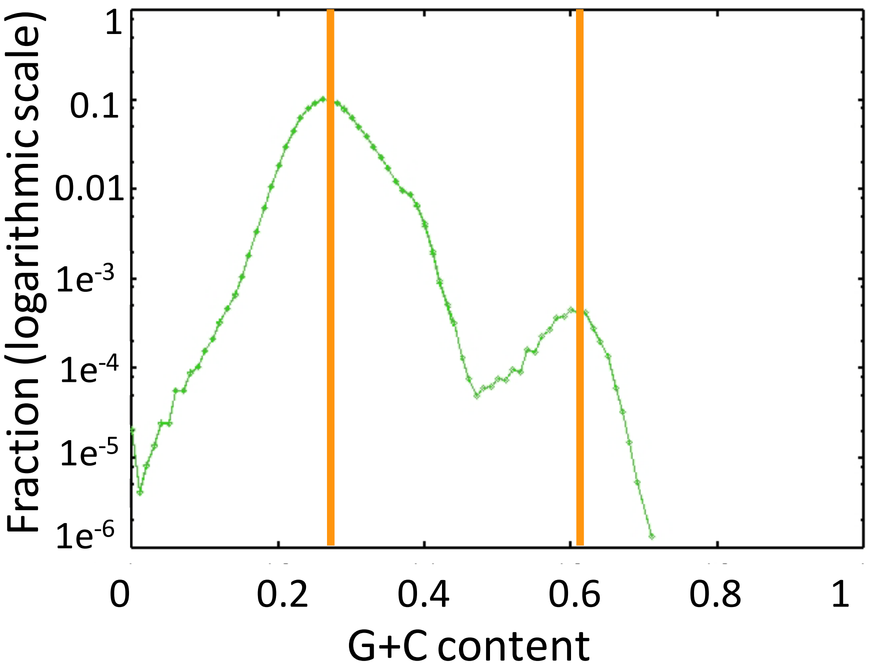
Outline

1. *Hydra* basics

2. **Identification of a bacterial genome in the *Hydra* assembly**

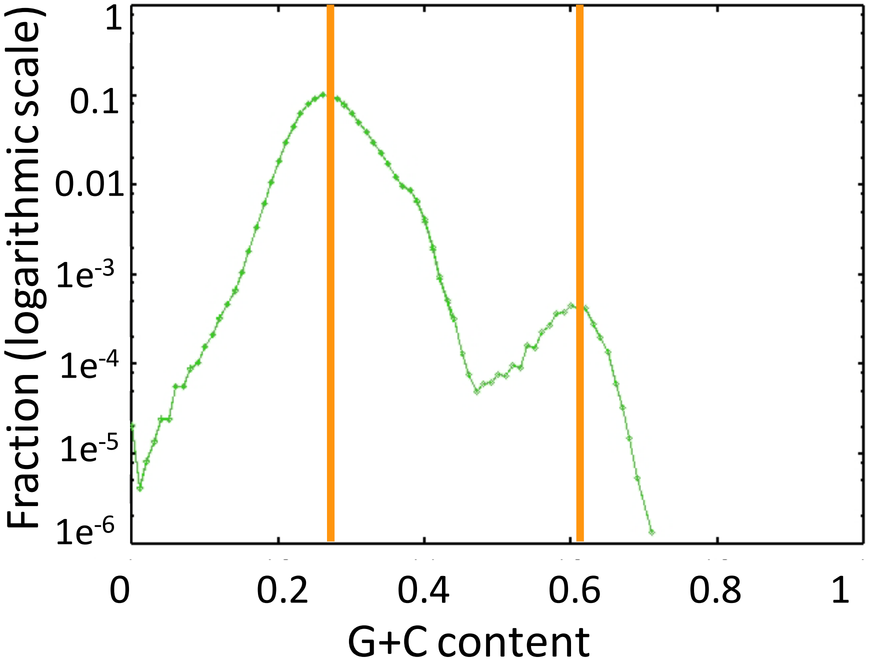
3. Investigation of other foreign sequences in the *Hydra* genome

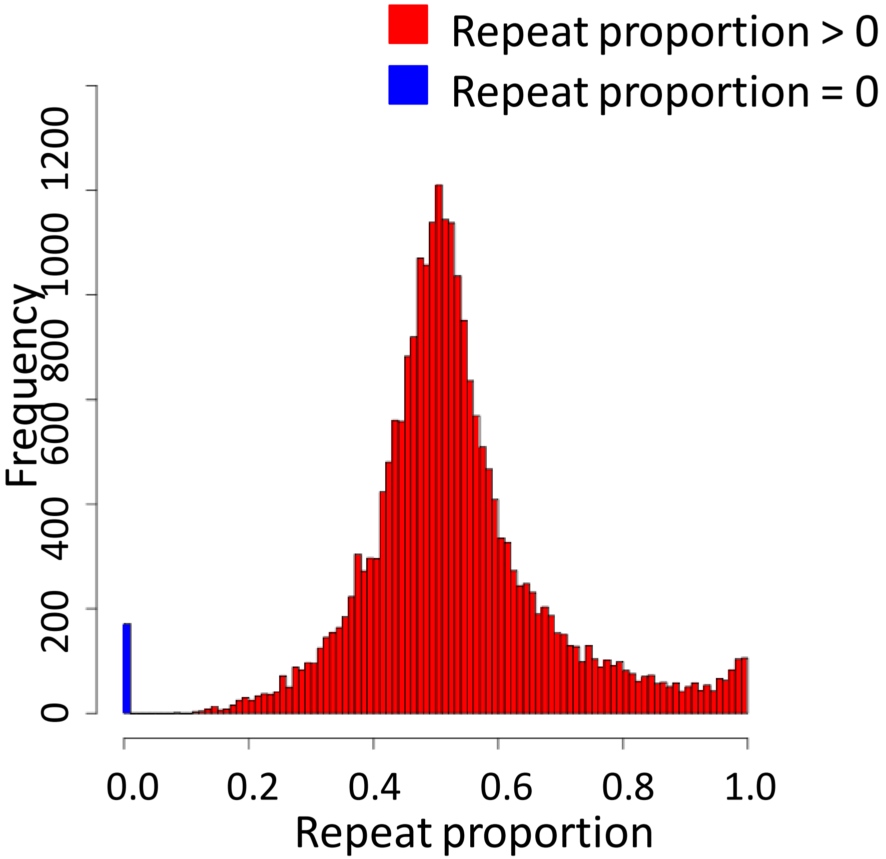
Signs of foreign sequences in the *Hydra* genome



- reads with *Hydra*-untypical G+C content

Signs of foreign sequences in the *Hydra* genome





- reads with *Hydra*-untypical G+C content
- contigs with no repetitive sequences

A nearly-complete bacterial genome in the *Hydra* sequences

Genome summary:

- 4 Mb in 8 contigs
- 3782 genes, avg. gene size 900 bp
- classified as *Curvibacter sp*. based on 16S rRNA
- encodes many more sugar transporters than close relatives

  → possible symbiotic relationship with *Hydra*
- the association with *Hydra* is stable

Outline

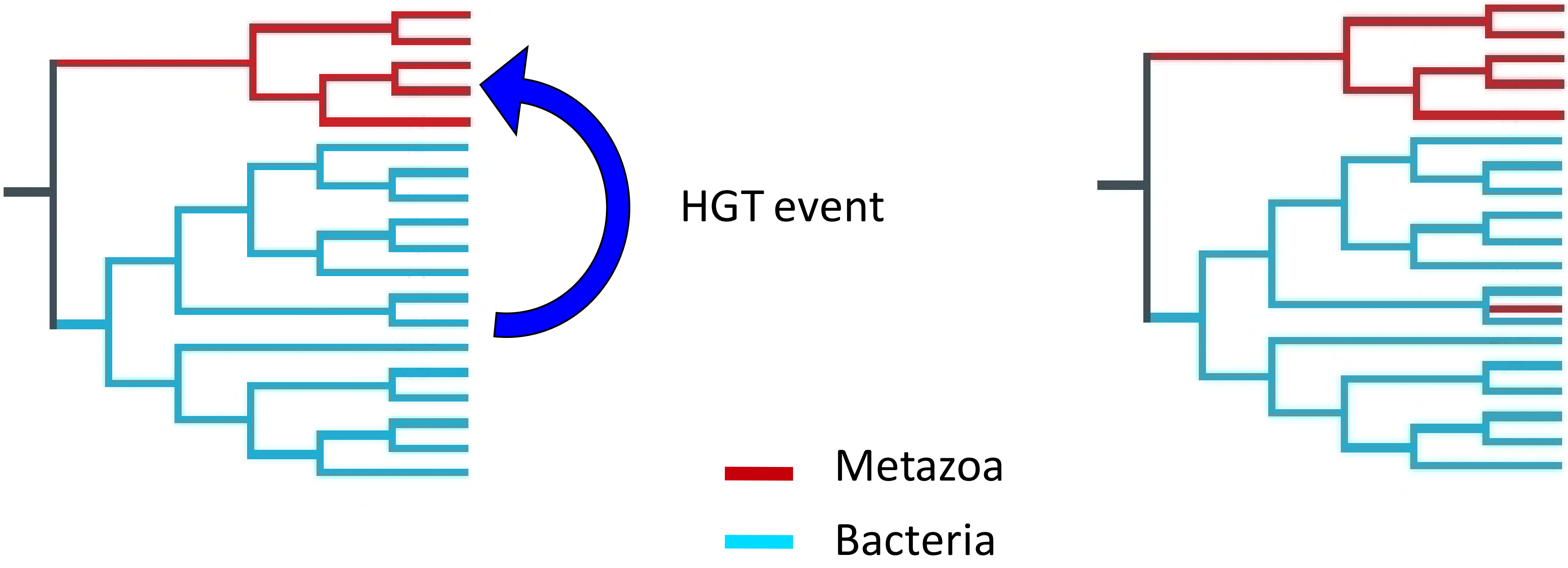
1. *Hydra* basics

2. Identification of a bacterial genome in the *Hydra* assembly

3. Investigation of other foreign sequences in the *Hydra* genome

Characteristic of horizontal gene transfers (HGTs)

Discrepancy between the species tree and the gene tree of a horizontally transferred gene



Species tree

Gene tree for horizontally transferred gene

HGT event

🟥 Metazoa

🟦 Bacteria

Procedure for the detection of HGT candidates

Identify all open reading frames (ORFs) in the *Hydra* genome

~ 158,000

Select only those with strong similarity to bacteria and weak similarity to metazoa (using SIMAP)

4,500

Select candidates that correspond to Hydra gene models (Hma2) and lack repeats

180

Calculate and inspect phylogenies

76

List of putative HGTs

Summary of HGT candidates

- 76 Hydra gene models / 48 different gene families
- 52 with functional annotation / 24 hypothetical proteins
- 56 have good EST support
- 18 have a spliced leader sequence
- 37 are single exon / 39 are multi exon
- 29 have a *Nematostella* homolog

Example for a known HGT : capA

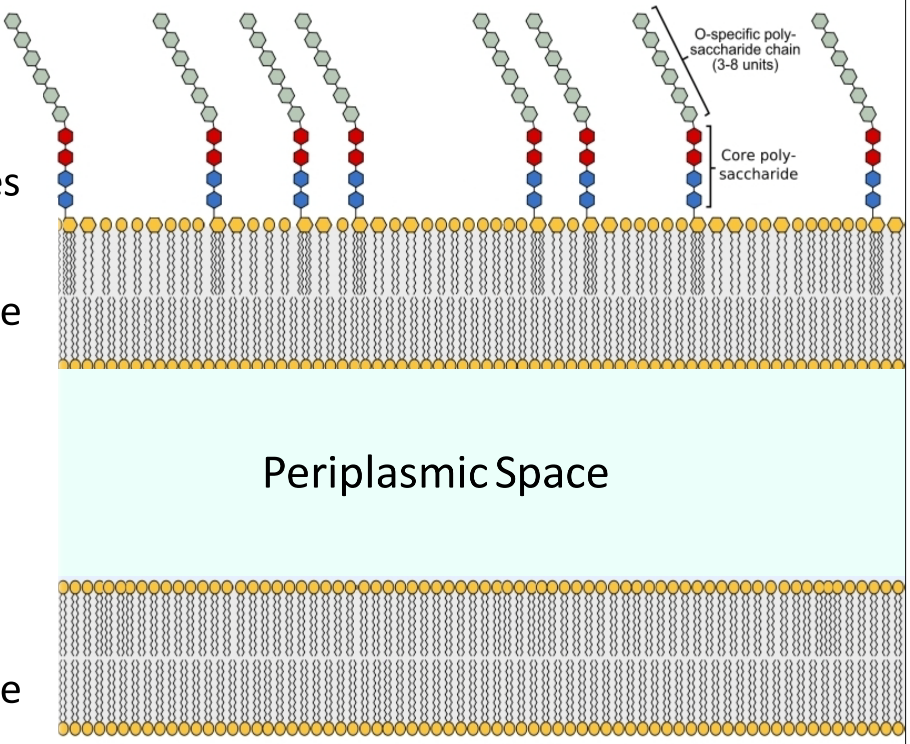
(Photo courtesy of H. Bode)

- involved in poly-gamma-glutamate (PG) synthase in bacteria

→ PG is present only in prokaryotes and nematocytes in cnidarians

- identified in *Clythia* and shown to be expressed in nematocytes (Denker et al, 2008)
- good EST support
- spliced leader sequence
- expressed in nematocytes in Hydra

Lipopolysaccharide (LPS) biosynthesis



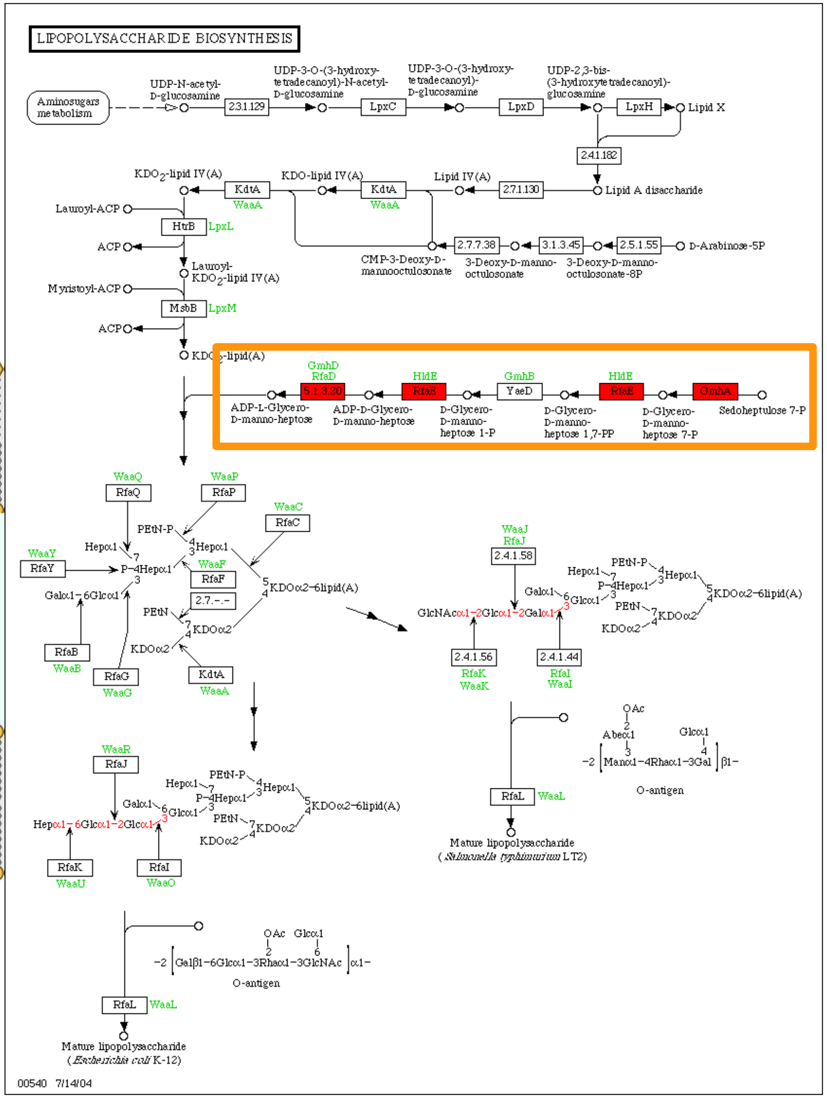
Lipopoly-saccharides

Outer Membrane

Periplasmic Space

Inner Membrane

Cytosol



Summary

- A nearly-complete bacterial genome in the *Hydra* sequences was identified

→ first sequenced genome of the species *Curvibacter*

→ genome annotation and comparative genomics give hints for symbiosis with *Hydra*

- 76 putative horizontally transferred genes in the *Hydra* genome were detected

  → three HGTs form a reaction chain in the LPS pathway