# Gene gain and loss in bacteria transitioning between aquatic biomes

#### Krzysztof T. Jurdziński<sup>1</sup>, Maliheh Mehrshad<sup>2</sup>, Ziling Deng<sup>1</sup>, Stefan Bertilsson<sup>2</sup>, Anders F. Andersson<sup>1</sup>

<sup>1</sup> Department of Gene Technology, Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden; <sup>2</sup> Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden

# **Background:**

- Bacteria rarely undergo cross-biome transitions.  $\bullet$
- High rates of **horizontal gene transfer** in bacteria are  $\bullet$ maintained due to rare gains of genes key to adaptation.
- Strong selective pressure restricting bacterial genome sizes leads to frequent losses of redundant genes.

# **Conclusions:**

Identify

MSG 1

(freshwater)

**Specific genes** are **gained/lost** throughout transitions between aquatic biomes by diverse bacterial lineages. These genes can be:

- directly connected to adaptation to different concentrations of inorganic ions and osmotic pressure;
- connected to other genomic and physiological changes accompanying transitions;

Transitions as pairs of

monobiomic sister

groups (MSGs)

MSG 2

(brackish)

exposure to different mobile genetic elements.

## **Data analysis:**

#### Publicly available metagenome assembled genomes (MAGs)

11,276 MAGS ≥75% completeness ≤5% contamination Metagenomes: 416 freshwater,

85 brackish, 243 marine

### **Results:**

#### **Freshwater** ↔ **Marine**

3575

~species

K03549: KUP system potassium uptake protein K08974: putative membrane protein

Phylogeny

reconstruction

- K03499: trk/ktr system potassium uptake protein
- K07301: Na+/Ca2+ antiporter

Names of KOs differentially present for more than one transition type are

KEGG

#, \$, ^ or \* hoforo KO numbor

Gained/lost genes

**Presence/absence of** 

**Pairwise comparison** 

across transitions.

KEGG Orthology (KO) Groups.

Present in	-20 K07. -20 K160	052: MscS family membrane prote	ein highlighted.	before KO number
Freshwater Marine	Freshwate	55	Varine	Groups of KOs co- annotated (>50% of
Freshwater ← 20 12 12 12 12 12 12 12 12 12 12	<ul> <li>Brackish</li> <li>g2+ transporter</li> <li>A dioxygenase extradiol</li> <li>chuate 4,5-dioxygenase, alpha chain</li> <li>tem potassium uptake protein</li> <li>stem potassium uptake protein</li> <li>antiporter</li> <li>6-phosphate synthase/phosphatase</li> <li>membrane protein</li> </ul>	So Brackish	Brackish ↔ Marine K01480: agmatinase K04759: ferrous iron trans K00809: deoxyhypusine s K01585: arginine decarbo K03782: catalase-peroxic K21498: antitoxin HigA-1 K19159: antitoxin YefM K07064: uncharacterized #K05844: ribosomal prote #K18310: beta-citryloluta	cases) to the same genes sport protein B synthase oxylase dase protein ein S6L-glutamate ligase
<ol> <li>Level C KO annotation</li> <li>Alanine, aspartate and glutam</li> <li>Arginine and proline metabolis</li> <li>Betalain biosynthesis</li> <li>Bacterial motility proteins</li> <li>Cell motility</li> <li>Drug metabolism – other enzy</li> <li>Function unknown</li> <li>General function prediction on</li> <li>Glycosyltransferases</li> </ol>	ate metabolism M 13 Prokaryotic defense system 14 Protein processing 15 Replication and repair 16 Ribosome biogenesis 17 Secondary bile acid biosynthesis 18 Secretion system 19 Structural proteins 20 Transporters	Digital copy of the poster and more details here:	<ul> <li>#K14940: gamma-F420-</li> <li>K03284: magnesium trans</li> <li>K03282: large conductand</li> <li>K02237: competence prot</li> <li>K06987: uncharacterized</li> <li>K03321: sulfate permease</li> <li>K02036: phosphate trans</li> <li>K02040: phosphate trans</li> <li>K02038: phosphate trans</li> <li>K02037: phosphate trans</li> <li>K01673: carbonic anhydra</li> <li>K07007: 3-dehydro-bile a</li> </ul>	-2:alpha-L-glutamate ligase sporter ce mechanosensitive channel tein ComEA protein e, SulP family port system ATP-binding protein port system substrate-binding protein sport system permease protein sport system permease protein ase acid Delta4,6-reductase

Nitrogen metabolism



Polycyclic aromatic hydrocarbon degradation

#### **KEGG orthologs (KOs) significantly (FDR < 0.1)** differentially present between pairs of MSGs:

- The numbers of transitions (MSG pairs) between each two lacksquareof the biomes and the results of gene gain/loss analysis are given by the arrows connecting the two compared biomes.
- KOs were clustered by similarity in presence/absence patterns across all the MSG pairs.

K07290: AsmA family protein K07115: 23S rRNA (adenine2030–N6)–methyltransferase 21 K02482: NtrC family, sensor kinase K07343: DNA transformation protein and related proteins K03630: DNA repair protein RadC K03413: chemotaxis family, chemotaxis protein CheY 21 K20974: two-component system, sensor histidine kinase ^K07497: putative transposase ^K07483: transposase K02238: competence protein ComEC K07460: putative endonuclease K07391: magnesium chelatase family protein

