Use of a four-tiered graph to parse the factors leading to phenotypic clustering in bacteria: a case study based on samples from the Aletsch Glacier

Miroslav Svercel¹§, Manuela Filippini¹*, Nicolas Perony²*, Valentina Rossetti¹, Homayoun C. Bagheri¹§

¹ Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, ² Chair of Systems Design, ETH Zurich, Zurich, Switzerland

* These three authors contributed equally to this work

§ Corresponding authors, E-mails: miro.svercel@gmail.com

bagheri@ieu.uzh.ch
Text S1: Supporting Materials and Methods

Sampling sites
In sampling site A, located at the edge of the Aletsch glacier, six different subsites were chosen for sampling (supporting Figure A of Text S1): “Blue lagoon” (BL) – small water pond situated 15-20 m east of the glacier; “Glacier drip” (GD) – drip water from glacier; “Glacier cave drip” (GCD) – water drip from the roof of a glacier cave (two samples taken labeled as GCDA, GCDB); “Glacier surface mud” (GSM) – mud particles covering the top part of glacier, “Mud under Glacier Drip” (MGD) – small mud ponds collecting under dripping water from edges of the glacier. At sampling site B, on a hilly ridge to the south of the glacier, two samples were collected and labeled “Grass reactor pond A and B” (GRPA, GRPB) – small ponds (approx. 16 m²) situated outside the glacial body (about 100 m) at an altitude of 2175 m above sea level, with floating grass debris and brownish water.

DNA extraction and PCR amplification
To extract DNA, a 30 µL aliquot of bacterial suspension from each cultivated isolate was transferred into 100 µL lysis buffer (50 mM KCl; 0.1% Tween 20; 10 mM Tris-HCl; pH 8.3) [9] in an ABgene® 96-well PCR plate (Thermo Fisher Scientific Inc., Waltham, USA). The suspension was centrifuged at 2880 x g for 5 min and incubated for 10 min at 99°C. The heat-lysed suspension was frozen at -20°C for 30 min. When thawed, 7 µL of the supernatant was taken for PCR. Amplification of 16S rRNA genes was performed using universal primers for bacteria 27F and 1492R [10] and carried out in 30 µL reaction mixtures containing 7 µL of lysed bacterial suspension, 12 µL of RNase-free water, 12 µL of TopTaq Master Mix (Qiagen, Hombrechtikon, Switzerland) and 1.5 µL of each primer (0.40 µM). The PCR was done using
the Techne TC 512 thermal cycler (Barloworld Scientific ltd, United Kingdom) with the following program: 5 min at 95°C, 37 cycles of denaturation at 95°C for 45 s, annealing at 58°C for 40 s and extension at 72°C for 120 s. The thermo cycling condition was concluded with a final extension at 72°C for 10 min.

**Restriction fragment length polymorphism and 16S rRNA gene sequencing of selected bacteria**

The PCR amplicons were digested with four fastDigest® restriction enzymes *Hae*III, *Hinf*I, *Msp*I and *Taq*I (LabForce AG, Nunningen, Switzerland) according to the supplier’s instructions. Restriction fragments were separated by electrophoresis in ethidium bromide stained 2% agarose gels. Isolates were selected for further investigation on the basis of different RFLP pattern profiles using the combination of above mentioned restriction enzymes. Prior to sequencing, PCR products were purified using the QuickStep™ 2 96-well PCR purification kit (EdgeBio, Gaithersburg, USA). The sequencing was performed in-house (ABI 3730 Sequencer) and confirmed by re-sequencing at an external facility (Microsynth, Balgach, Switzerland).

**Phylogenetic analysis**

Phylogenetic trees were constructed by using Neighbor-Joining (NJ), minimal evolution (ME), maximum parsimony (MP) and maximum likelihood (ML) algorithms using MEGA software package. Evolutionary distances were computed using Kimura-2 model and Maximum Composite Likelihood methods. Maximum likelihood was performed using Garli [S1]. The TIM3 +G model ( Transitional Model3, +G, gamma correction) was determined as the appropriate model of molecular evolution for our data by using jModeltest 0.1.1 [S2]. Analyses were performed using bootstrap values of 1000. The phylogenetic analysis was performed based
**TABLE A.** Carbon substrates used by all tested bacteria

<table>
<thead>
<tr>
<th>Amino Acids</th>
<th>Carboxylic Acids</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Aspartic Acid</td>
<td>Succinic Acid</td>
<td>Adenosine</td>
</tr>
<tr>
<td>L-Glutamic Acid</td>
<td>D-Gluconic Acid</td>
<td>Inosine</td>
</tr>
<tr>
<td>L-Asparagine</td>
<td>Fumaric Acid</td>
<td>Bromo Succinic Acid</td>
</tr>
<tr>
<td>L-Glutamine</td>
<td>D,L-Malic Acid</td>
<td></td>
</tr>
<tr>
<td>L-Alanine</td>
<td>L-Malic Acid</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE B.** Carbon substrates used by less than 10% of the tested bacteria

<table>
<thead>
<tr>
<th>Carbohydrates</th>
<th>Amino Acids</th>
<th>Carboxylic Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2 propanediol</td>
<td>D-Aspartic Acid</td>
<td>Alfa-hydroxy butyric acid</td>
</tr>
<tr>
<td></td>
<td>D-Threonine</td>
<td>Alfa-keto-butyric acid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tricarballylic Acid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M-Hydroxy Phenyl Acetic Acid</td>
</tr>
</tbody>
</table>

Classification was done according to Preston-Mafham et al. [S7].
Figure Legends of Figures in Text S1

Figure A of Text S1. Map of the sampling sites.
(a) Aerial view of the Aletsch Glacier (Canton Valais, Switzerland) showing the two sampling sites investigated in this study. Modified from International Space Station Imagery (ISS013-E-77377, 5 Sep. 2006). (b) Topographic map of the sampling region showing the location of the two sampling sites. Created using OpenStreetMap and Marble Desktop Globe version 0.8.5. (c) Sampling site codes and numbers of isolates tested.

Figure B of Text S1. Box plots of genetic variance.

Box plot of genetic variance data based on the whole data set of 16S rRNA gene sequences of *Pseudomonas* and *Serratia* genera. For box plots, the central mark is the median, and the edges of the box are the 25th (Q1) and 75th (Q3) percentiles. The maximum whisker length is determined by $w = 1.5$. Whiskers extend to the most extreme data points falling within the range $[Q1-w*(Q3-Q1), Q3+w*(Q3-Q1)]$. Data points outside this range are plotted individually as crosses. Significance of the difference in medians between the two genera was tested with the Wilcoxon test.

Figure C of Text S1. Neighbor-joining phylogenetic tree of 16S rRNA sequences of sampled $\gamma$-*Proteobacteria* representatives and their closest relatives.

Clade A represents the *Pseudomonas* genus whereas clade B is a subset of the *Enterobacteriaceae*. Codes in bold indicate strains isolated in the present study. Bootstrap values >50% are shown at the nodes as a percentage of 1000 replicates. Scale bar represents number of changes per nucleotide position. Accession numbers of 16S rRNA gene sequences are given in brackets. The tree was rooted using *Aquiflex pyrophilus* as the outgroup.
Many of our Gram-negative isolates were pseudomonads (clade A), and members of three subclades (A1, A2 and A3) matched with psychrophilic bacteria obtained from Arctic, Antarctic and glacial environments [S3,S4]. *Serratia* strains GDd8 and MGDd10 matched to psychrotrophic *S. proteamaculans* from food [S5]. A large group of *Serratia* was affiliated to strain C1 (subclade B2), which was found in the gut of the Tsetse fly [S6].

**Figure D of Text S1. Scatter plot analysis.**

(a) Scatter plot of pairwise distances between enzymes profiles (API ZYM) and carbon assimilation profiles (Biolog PM1), both based on normalized Hamming distance. (b) Scatter plot of pairwise distances between enzymes profiles and phylogenetic distance. (c) Scatter plot of pairwise distances between carbon assimilation profiles and phylogenetic distance (c). Coordinate of each point represents a variable pair measured for a given bacterium. The Pearson correlation coefficient was applied in order to establish correlation between datasets. The significance of the latter coefficient was tested through a t-test.
Figure A.

### Table A.2: Summary of isolates for RFLP

<table>
<thead>
<tr>
<th>Site</th>
<th>Subsite</th>
<th>Sample code</th>
<th>Isolates for RFLP</th>
<th>Sequenced isolates</th>
<th>Phenotypically tested isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Blue Lagoon</td>
<td>BL</td>
<td>96</td>
<td>12 (P-7; S-3) ¹</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Glacier Drip</td>
<td>GD</td>
<td>96</td>
<td>8 (P-4; S-4)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Glacier Cave Drip A</td>
<td>GCDA</td>
<td>96</td>
<td>16 (P-10, S-5)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Glacier Cave Drip B</td>
<td>GCDB</td>
<td>96</td>
<td>12 (P-11, S-0)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Glacier Surface Mud</td>
<td>GSM</td>
<td>96</td>
<td>5 (P-1; S-2)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Mud under Glacier Drip</td>
<td>MCD</td>
<td>96</td>
<td>41 (P-20, S-12)</td>
<td>8 (7) ²</td>
</tr>
<tr>
<td>B</td>
<td>Grass Reaction Pond A</td>
<td>GRPA</td>
<td>96</td>
<td>14 (P-3; S-11)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Grass Reaction Pond B</td>
<td>GRPB</td>
<td>96</td>
<td>10 (P-3, S-4)</td>
<td>4 (2)</td>
</tr>
<tr>
<td></td>
<td><strong>Total isolates</strong></td>
<td><strong>758</strong></td>
<td><strong>118</strong> ³(P-59, S-41)</td>
<td><strong>36</strong> (33)</td>
<td></td>
</tr>
</tbody>
</table>

¹ Numbers of *Pseudomonas* (P) and *Serratia* (S)

² Numbers of bacteria tested in Biolog (Gram-positive bacteria were not tested)

³ Beside P and S, additional genera (belonging to *Yersinia* sp., *Aeromonas* sp., *Janthinobacterium* sp. and to Gram-positive: *Bacillus* sp., *Viridibacillus* sp. and *Exiguobacterium* sp.) were isolated from samples
Figure B.

![Graph showing composition distance between Pseudomonas sp. and Serratia sp.](image)

- Composition distance range: 0.00 - 0.16
- Pseudomonas sp.
- Serratia sp.
- h = 1, p-value < 0.001
Figure D.

**Panel a:**
Distance between enzyme profiles

- Pearson corr coef $\rho = 0.75$, pval $< 0.001$

**Panel b:**
Distance between enzyme profiles

- Pearson corr coef $\rho = 0.35$, pval $< 0.001$

**Panel c:**
Distance between carbon profiles

- Pearson corr coef $\rho = 0.30$, pval $< 0.001$
References for supporting information


