Phylogenetic analysis of the fish pathogen *Aeromonas salmonicida* underlines the dichotomy between European and Canadian strains for the *salmonicida* subspecies

Antony T. Vincent¹,²,³ and Steve J. Charette¹,²,³,#

1. Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Quebec City, QC, Canada, G1V 0A6
2. Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Quebec City, QC, Canada, G1V 0A6
3. Centre de recherche de l’Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ), Quebec City, QC, Canada, G1V 4G5

#To whom correspondence should be addressed: Institut de Biologie Intégrative et des Systèmes (IBIS), Pavillon Charles-Eugène-Marchand, 1030 avenue de la Médecine, Université Laval, Quebec City, QC, Canada, G1V 0A6

steve.charette@bcm.ulaval.ca; Telephone: 1-418-656-2131, ext. 6914; Fax: 1-418-656-7176

Running title: Phylogenetic analysis of *A. salmonicida*
Short Communication

The Gram-negative bacterium *Aeromonas salmonicida*, which belongs to the Gammaproteobacteria class, has a taxonomy containing five official subspecies (*pectinolytica*, *masoucida*, *achromogenes*, *smithia* and *salmonicida*) (Dallaire-Dufresne et al., 2014). There are also three strains from India (Y577, Y47 and Y567) (Vincent et al., 2016a), which have an unclear taxonomy and no subspecies assignment yet. This bacterium is of veterinary importance since many taxa belonging to the different subspecies are known to infect fish (Austin & Austin, 2012; Burr & Frey, 2007), thus causing important livestock losses worldwide (Dallaire-Dufresne et al., 2014). The most studied subspecies of this bacterium is *salmonicida*, which is sometimes qualified as “typical”, and is well known to cause furunculosis, a fish disease mainly affecting salmonids (Austin & Austin, 2012). The other psychrophilic subspecies are qualified as “atypical” and infect a wide range of hosts (Dallaire-Dufresne et al., 2014).

Early in the research field of *A. salmonicida* subsp. *salmonicida*, one study reported a genetic grouping of this bacterium based on geographical regions (Finland, Denmark, Sweden, Norway and Canada) using ribotyping and random amplified polymorphic DNA (RAPD) profiling (Hänninen et al., 1995). Many other studies reported that *A. salmonicida* subsp. *salmonicida* isolates are genetically highly homogeneous with a clonal population structure (Belland & Trust, 1988; García et al., 2000; McCormick et al., 1990; O’HICi et al., 2000; Umelo & Trust, 1998). However, these studies were based on low-resolution approaches, making it perilous to conclude on the genetic diversity of *A. salmonicida* subsp. *salmonicida* strains, which are evolutionarily close.
With the advances in sequencing technologies (Vincent et al., 2016c), some studies found differences in the mobilome between European and Canadian strains. For example, European strains have a greater tendency to lack small plasmids pAsa3 and pAsal1 compared to Canadian isolates (Attéré et al., 2015). However, one of the strongest discrepancies between the European and the Canadian strains is the genomic island AsaGEI. Many variants of this genomic island were reported (AsaGEIIa, 1b, 2a, 2b, and finally 2c) (Emond-Rheault et al., 2015a, b; Long et al., 2016). Each one of these variants was correlated with the geographical provenance of the isolates where it was found. Canadian isolates bear AsaGEIIa or 2a with only one isolate so far containing no AsaGEI (Emond-Rheault et al., 2015a). On their side, about 60% of the tested European isolates have no AsaGEI and the others carry AsaGEIIb or 2b (Emond-Rheault et al., 2015a, b). This dichotomy between European and Canadian strains of *A. salmonicida* subsp. *salmonicida* was also observed in a recent phylogenomic study (Vincent et al., 2016a). However this study contained a limited number of seven *A. salmonicida* subsp. *salmonicida* isolates.

Here, the genomes of 10 isolates (5 Europeans and 5 Canadians) were completely sequenced by next-generation sequencing. This represents a substantial increase in the number of available genomic sequences for this bacterium (Table 1). In addition to gain a clearer view of the diversity of *A. salmonicida* subsp. *salmonicida*, genomic comparisons were performed to test the hypothesis that a dichotomy between European and Canadian isolates of *A. salmonicida* subsp. *salmonicida* exists not only based on the mobilome (plasmids and genomic islands) but also on the core genome.
The pan-genome for all the 26 A. salmonicida isolates studied here (Table 1) was found using GET_HOMOLOGUES version 20160822 (Contreras-Moreira & Vinuesa, 2013), allowing to sort the genes in four categories based on orthologous gene cluster frequency distribution: cloud (genes present only in a few taxa), shell (genes present in several taxa), soft core (genes in 95% of the taxa) and core (genes in all taxa). The resulting distribution displayed a characteristic U-shape (Figure 1), as it was expected for a bacterial pan-genome analysis (Gordienko et al., 2013; Haegeman & Weitz, 2012).

Afterwards, it was interesting to evaluate if the pan-genome of A. salmonicida is “open” or “closed”, as defined elsewhere (Medini et al., 2005). The type of pan-genome reflects the lifestyle of a bacterium: an “open” pan-genome is found in sympatric species having a high capacity to exchange genetic material with other species, while a “closed” pan-genome is typical of allopatric species having a limited capacity to exchange genetic elements (Rouli et al., 2015). The pan-genome of A. salmonicida, at least for the current dataset, showed to be “open”, i.e. the number of new genes increases by adding new genomes in the dataset (Figure 2). This was not unexpected knowing, among others, the large number of plasmids found in the various strains of this bacterium [reviewed in (Piotrowska & Popowska, 2015)]. In fact, it was already proposed that Aeromonas sp. (Gordon et al., 2008) and more specifically A. salmonicida (Trudel et al., 2016; Vincent et al., 2014) could be important reservoirs of mobile genetic elements. The present study, by showing that A. salmonicida has an “open” pan-genome, supports this hypothesis. This information is of prime importance since many of the plasmids found in A. salmonicida so far harbor genes coding for antibiotic resistance (Piotrowska & Popowska, 2015; Vincent et al., 2014, 2016b).
Finally, orthologous genes of the soft core genome (3,333 clusters) were codon-aligned using PRANK version 150803 (Löytynoja, 2014) and filtered by BMGE version 1.12 (Criscuolo & Gribaldo, 2010). After removing the genes showing no variation or containing potential paralogous genes, a total of 3,257 clusters were concatenated into a matrix of 164,167 positions using SequenceMatrix version 1.8 (Vaidya et al., 2011). The best-fit model was chosen using jModelTest version 2.1.10 (Darriba et al., 2012). A phylogenetic analysis was performed by RAxML version 8.2.9 (Stamatakis, 2014) under the model GTR+Γ4 and 1,000 rapid bootstraps.

As already published elsewhere (Vincent et al., 2016a), the *A. salmonicida* subsp. *pectinolytica* 34mel clustered with mesophilic indian strains Y577, Y567 and Y47 while the psychophilic subspecies *smithia* and *achromogenes* clustered together with the subspecies *masoucida* being between the mesophilic and the psychophilic groups (Figure 3). The *salmonicida* subspecies isolates share a near common ancestor with the clade formed of *achromogenes* and *smithia*. Most interestingly, the tree presented in Figure 3 shows a clear and statistically well-supported split between the European and the Canadian isolates of *A. salmonicida* subsp. *salmonicida*. The tree is even robust enough to differentiate confidently between isolates from Quebec and New Brunswick, which are two neighboring Canadian provinces. Interestingly, isolates from New Brunswick cluster together and share a common ancestor with isolates from Quebec (QC-1). A second cluster also composed by isolates from Quebec (QC-2) is more basal. This suggests that isolates from New Brunswick might have a Quebecker origin. This scenario might be realistic by considering that infected fish from Quebec could have been transferred to New
Brunswick. Unfortunately, no information regarding neither the provenance nor the pedigree of the fish that could be implicated with the spread of *A. salmonicida* are available. Even if the 10 Canadian isolates were all isolated from farmed diseased fish from at least 8 different farms, larger sampling from both Quebecker clusters (QC-1 and QC-2) would be required to support this hypothesis.

As it was expected, there is a clear correlation between the type of the genomic island *AsaGEI* found in an isolate and its geographic provenance. However, the situation is less clear for the European isolates where there are two clades without any *AsaGEI*, one being the most basal clade for the European isolates. In all the cases, the bootstrap values are weaker for the nodes corresponding to the European isolates, meaning that it would not be rigorous to postulate evolutionary scenarios regarding these isolates. Finally, Canadian isolates bear *AsaGEI1a* and *2a*, with isolates bearing *AsaGEI2a* clustering together in the phylogenetic tree (New Brunswick + QC-1). Interestingly, the various phylogenetic clades are not correlated with the source of isolation (host species or seawater/freshwater), however, this should be validated with a larger and more equilibrated dataset. In light of this molecular phylogeny, our results support the earlier finding that *A. salmonicida* subsp. *salmonicida* does not have a clonal population structure and isolates from various geographical regions are different from each other.

In summary, the present study shows that the fish pathogen *A. salmonicida* has an “open” pan-genome, underlining the idea that this bacterium could be a hub for mobile genetic elements between various waterborne bacteria. This finding is worrying in a veterinary context since
many of these mobile elements are plasmids conferring drug resistance, such pSN254b and pAB5S9b, two plasmids causing resistance to all antibiotics approved by the Veterinary Drugs Directorate (VDD) of Health Canada to treat infected fish (Trudel et al., 2016). Given the confirmation here that *A. salmonicida* possesses an “open” pan-genome and the earlier finding concerning its ability to easily acquire drug resistance, research and development for alternative treatments against furunculosis in aquaculture should be urged. Moreover, the results presented here support that *A. salmonicida* subsp. *salmonicida* is more genetically heterogeneous than previously thought and that epidemiologic studies should be performed in this direction. This heterogeneity is not only due to mobile genetic elements but is also perceptible by the mutations within the coding sequences shared between the various strains. In fact, the correlation between the present phylogenetic analysis and the type of AsaGEI found in the isolates suggests a strong co-evolution between the core genome and the mobilome in concordance with the geographical region.

**Acknowledgments**

The authors thank Dr. Luca Freschi (U. Laval) for his critical reading of the manuscript. ATV received an Alexander Graham Bell Canada Graduate Scholarships from the Natural Sciences and Engineering Research Council of Canada (NSERC). This project was funded by an NSERC Discovery grant to SJC.
References


Sequence of the Melanin-Producing Extremophile *Aeromonas salmonicida* subsp. *pectinolytica* Strain 34melT. *Genome Announc* 1, e00675–13.


Umelo, E. & Trust, T. J. (1998). Physical map of the chromosome of *Aeromonas salmonicida*


Table 1. *A. salmonicida* isolates used in the present study.

<table>
<thead>
<tr>
<th>Isolates</th>
<th>Isolation year</th>
<th>Provenance</th>
<th>Host</th>
<th>GenBank</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-B522</td>
<td>2001</td>
<td>Canada (QC)</td>
<td><em>Salvelinus fontinalis</em></td>
<td>MIIM00000000</td>
<td>This study</td>
</tr>
<tr>
<td>01-B526</td>
<td>2001</td>
<td>Canada (QC)</td>
<td><em>Salvelinus fontinalis</em></td>
<td>NZ_AGVO01000000</td>
<td>(Charette et al., 2012)</td>
</tr>
<tr>
<td>2004-05MF26</td>
<td>2004</td>
<td>Canada (NB)</td>
<td>°</td>
<td>NZ_JRYW00000000</td>
<td>(Vincent et al., 2015)</td>
</tr>
<tr>
<td>09-0167</td>
<td>2009</td>
<td>Canada (QC)</td>
<td><em>Salmo salar</em></td>
<td>LMTK00000000</td>
<td>This study</td>
</tr>
<tr>
<td>2009-144K3</td>
<td>2009</td>
<td>Canada (NB)</td>
<td><em>Salvelinus fontinalis</em></td>
<td>NZ_JRYV00000000</td>
<td>(Vincent et al., 2015)</td>
</tr>
<tr>
<td>2009-157K5</td>
<td>2009</td>
<td>Canada (NB)</td>
<td><em>Salvelinus fontinalis</em></td>
<td>MIIQ00000000</td>
<td>This study</td>
</tr>
<tr>
<td>m23067-09</td>
<td>2009</td>
<td>Canada (QC)</td>
<td><em>Salvelinus fontinalis</em></td>
<td>MIIP00000000</td>
<td>This study</td>
</tr>
<tr>
<td>2010-47K18</td>
<td>2010</td>
<td>Canada (NB)</td>
<td><em>Salvelinus fontinalis</em></td>
<td>MIIR00000000</td>
<td>This study</td>
</tr>
<tr>
<td>J223</td>
<td>1999</td>
<td>USA</td>
<td><em>Salmo salar</em></td>
<td>NZ_LSGV00000000</td>
<td>-</td>
</tr>
<tr>
<td>CIP 103209</td>
<td>1959</td>
<td>UK</td>
<td><em>Salmo salar</em></td>
<td>NZ_CDDW00000000</td>
<td>(Colston et al., 2014; Smith, 1963)</td>
</tr>
<tr>
<td>170-68°c</td>
<td>1968</td>
<td>Norway</td>
<td>Troutd</td>
<td>MIIN00000000</td>
<td>This study</td>
</tr>
<tr>
<td>A449</td>
<td>1975</td>
<td>France</td>
<td><em>Salmo trutta</em></td>
<td>NC_009348</td>
<td>(Reith et al., 2008)</td>
</tr>
<tr>
<td>JF2267</td>
<td>1999</td>
<td>Switzerland</td>
<td><em>Salvelinus alpinus</em></td>
<td>MIIO00000000</td>
<td>This study</td>
</tr>
<tr>
<td>JF2506</td>
<td>2000</td>
<td>Norway</td>
<td><em>Salmo salar</em></td>
<td>MIIS00000000</td>
<td>This study</td>
</tr>
<tr>
<td>JF2507</td>
<td>2000</td>
<td>Scotland</td>
<td><em>Salmo salar</em></td>
<td>MIIT00000000</td>
<td>This study</td>
</tr>
<tr>
<td>JF3517</td>
<td>2003</td>
<td>Norway</td>
<td><em>Scophthalmus maximus</em></td>
<td>MIIPU00000000</td>
<td>This study</td>
</tr>
<tr>
<td>JF3224</td>
<td>2004</td>
<td>Switzerland</td>
<td><em>Salmo trutta</em></td>
<td>NZ_JXTA00000000</td>
<td>(Emond-Rheault et al., 2015b)</td>
</tr>
<tr>
<td>JF3791</td>
<td>2006</td>
<td>Switzerland</td>
<td><em>Salvelinus alpinus</em></td>
<td>NZ_JYFG00000000</td>
<td>(Vincent et al., 2016b)</td>
</tr>
<tr>
<td>RS 534</td>
<td>-</td>
<td>France</td>
<td><em>Salmo trutta</em></td>
<td>NZ_JYFF00000000</td>
<td>(Vincent et al., 2016a)</td>
</tr>
<tr>
<td>34mel (pectinolytica)</td>
<td>1988</td>
<td>Argentina</td>
<td>River</td>
<td>NZ_ARYZ00000000</td>
<td>(Pavan et al., 2013)</td>
</tr>
<tr>
<td>AS03 (achromogenes)</td>
<td>2006</td>
<td>Korea</td>
<td><em>Carassius carassius</em></td>
<td>NZ_AMQG00000000</td>
<td>(Han et al., 2013)</td>
</tr>
<tr>
<td>Y47</td>
<td>2006</td>
<td>India</td>
<td>Food</td>
<td>NZ_JZTF00000000</td>
<td>(Vincent et al., 2016a)</td>
</tr>
<tr>
<td>JF4097 (smithia)</td>
<td>2007</td>
<td>Austria</td>
<td><em>Salvelinus alpinus lepeschini</em></td>
<td>NZ_JZTI01000000</td>
<td>(Vincent et al., 2016a)</td>
</tr>
<tr>
<td>Y567</td>
<td>2008</td>
<td>India</td>
<td>Food</td>
<td>NZ_JZTG00000000</td>
<td>(Vincent et al., 2016a)</td>
</tr>
<tr>
<td>Y577</td>
<td>2008</td>
<td>India</td>
<td>Food</td>
<td>NZ_JZTH01000000</td>
<td>(Vincent et al., 2016a)</td>
</tr>
<tr>
<td>NBRC 13784 (masouida)</td>
<td>-</td>
<td>Japan</td>
<td><em>Oncorhynchus masou</em></td>
<td>NZ_BAWQ01000000</td>
<td>-</td>
</tr>
</tbody>
</table>

a: The subspecies is indicated between brackets when not *salmonicida* (i.e., *A. salmonicida* subsp. *salmonicida*).
b: The information is not available.
c: Also known as HER1085
d: The exact fish species is unknown.
Figure 1. Categorization of the pan-genome of the 26 studied genomes into four categories as defined and computed by the tool GET_HOMOLOGUES (Contreras-Moreira & Vinuesa, 2013): cloud (genes present only in a few taxa), shell (genes present in several taxa), soft core (genes in 95% of the taxa) and core (genes in all taxa). This categorization is performed by computing how many gene clusters (orthologous genes) are present in a given number of genomes (see Table 1). For example, the core genome (the white bar) is composed of 2509 gene clusters that are present in all 26 genomes.
Figure 2. Estimation of the pan-genome size (total number of genes) as a function of the number of genomes for the studied dataset (Table 1). To avoid bias, genomes were randomly sampled 10 times. Fitted curve follows function published elsewhere (Tettelin et al., 2005).
**Figure 3.** Cladogram of 26 *A. salmonicida* (Table 1). The subspecies others than *salmonicida* are represented by diamonds while the *salmonicida* subspecies isolates from Europe and Canada are represented by circles and squares, respectively. Finally, an isolate from USA (J223) is represented by a white square. The Canadian isolates are also separated by Canadian provinces (Quebec [QC] and New Brunswick [NB]). The bootstrap values are indicated for all the nodes. The type of *AsaGEI* (*1a*, *2a*, *1b* and *2b*) is indicated for each *A. salmonicida* subsp. *salmonicida* isolates. The tree was midpoint rooted using FigTree version 1.4.3 (https://github.com/rambaut/figtree).