Variants of a genomic island in *Aeromonas salmonicida* subsp. *salmonicida* link isolates with their geographical origins

Jean-Guillaume Emond-Rheault[1,2,3], Antony T. Vincent[1,2,3], Mélanie V. Trudel[1,2,3], Francis Brochu[1.4], Brian Boyle[1], Katherine H. Tanaka[1,2,3], Sabrina A. Attéré[1,2,3], Éric Jubinville[1,2,3], Thomas P. Loch[5], Andrew D. Winters[5], Mohamed Faisal[5,6], Michel Frenette[2,7], Nicolas Derome[1,4], and Steve J. Charette[1,2,3,#]

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, Quebec City, QC, Canada, G1V 4G5
4. Département de biologie, Faculté des sciences et de génie, Université Laval, Quebec City, Canada, G1V 0A6
5. Department of Pathobiology and Diagnostic Investigation, College of Veterinary Medicine, Michigan State University, East Lansing, MI 48824, USA
6. Department of Fisheries and Wildlife, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI 48824, USA
7. Groupe de Recherche en Écologie Buccale (GREB), Faculté de médecine dentaire, Université Laval, Quebec City, QC, Canada, G1V 0A6

* Author for Correspondence: Steve J. Charette, Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Quebec City, QC, Canada, G1V 0A6, telephone: 418-656-2131, ext. 6914, fax: 418-656-7176, email: steve.charette@bcm.ulaval.ca

Running title: *AsaGEI*s in *A. salmonicida* subsp. *salmonicida*

Keywords: bacteria, genomics/functional genomics/comparative genomics; evolution/evolutionary processes/gene transfer/mutation; furunculosis; *Aeromonas salmonicida*; fish pathogen, genomic island; prophage; geographical distribution

ABSTRACT

Aeromonas salmonicida subsp. *salmonicida* is a fish pathogen. Analysis of its genomic characteristics is required to determine the worldwide distribution of the various populations of this bacterium. Genomic alignments between the 01-B526 pathogenic strain and the A449 reference strain have revealed a 51-kb chromosomal insertion in 01-B526. This insertion (*AsaGEI1a*) has been identified as a new genomic island (GEI) bearing prophage genes. PCR assays were used to detect this GEI in a collection of 139 *A. salmonicida* subsp. *salmonicida* isolates. Three forms of this GEI (*AsaGEI1a*, *AsaGEI1b*, *AsaGEI2a*) are now known based on this analysis and the sequencing of the genomes of seven additional isolates. A new prophage (prophage 3) associated with *AsaGEI2a* was also discovered. Each GEI appeared to be strongly associated with a specific geographic region. *AsaGEI1a* and *AsaGEI2a* were exclusively found in North American isolates, except for one European isolate bearing *AsaGEI2a*. The majority of the isolates bearing *AsaGEI1b* or no GEI were from Europe. Prophage 3 has also a particular geographic distribution and was found only in North American isolates. We demonstrated that *A. salmonicida* subsp. *salmonicida* possesses unsuspected elements of genomic heterogeneity that could be used as indicators to determine the geographic origins of isolates of this bacterium.

INTRODUCTION

Mobile DNA elements acquired by horizontal transfer contribute to genomic plasticity in bacteria. These elements include insertion sequences (ISs), miniature inverted-repeat transposable elements (MITEs), bacterial interspersed mosaic elements (BIMEs), repetitive extragenic palindromic sequences (REPs), integrons, transposons, transposable bacteriophages, and genomic islands (GEIs) (Bellanger et al., 2013; Darmon and Leach, 2014; Juhas et al., 2009).

GEIs are chromosomal regions ranging from 10 to 200 kb in size, are acquired by horizontal transfer, and are flanked by direct repeated sequences. They are often inserted into tRNA genes. However, these genetic elements are not present in all strains of the same species. GEIs are subdivided into pathogenicity, fitness, metabolic, resistance, symbiosis, saprophytic, ecological, or defense islands based on the functional advantages they provide (Bellanger et al., 2013; Darmon and Leach, 2014; Juhas et al., 2009).

The Gram-negative *Aeromonas salmonicida* species, a non-motile psychrophilic bacterium, is divided into five subspecies (*salmonicida, achromogenes, masoucida, smithia*, and *pectinolytica*) (Beaz-Hidalgo and Figueras, 2013). *Aeromonas salmonicida* subsp. *salmonicida* is the etiological agent of furunculosis, a disease that causes septicemia and necrosis, especially in salmonids (Dallaire-Dufresne et al., 2014). This disease has negative impacts worldwide and causes important economic losses for fish farms. For example, in the province of Quebec (Canada) between 1999 and 2010, furunculosis was diagnosed as being responsible for between 30% and 60% of all infections reported for brook trout (Morin,

2010), the most common aquaculture fish species in Quebec. Epidemiological tools would thus be very valuable for tracking this disease, especially since global production of salmon and trout has greatly increased over the last decade (Food and Agriculture Organization of the United Nations, 2014). More particularly, in Canada and Europe, aquaculture production of salmon and trout made up 66.4% and 28.3% of total aquaculture production in 2011, respectively (European Commission, 2014; Fisheries and Oceans Canada, 2013).

The strains from *Aeromonas salmonicida* subsp. *salmonicida* are usually considered genetically homogeneous based on low-resolution approaches such as restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD) analyses (Belland and Trust, 1988; García et al., 2000; Hänninen et al., 1995; Miyata et al., 1995; O'hIci et al., 2000; Studer et al., 2013). The genomic sequences of the A449 and 01-B526 strains of *A. salmonicida* subsp. *salmonicida* are available (Charette et al., 2012; Reith et al., 2008). Genomic analyses of the *A. salmonicida* subsp. *salmonicida* A449 strain showed that the chromosome bears two prophages, 108 complete or partial ISs, and five plasmids (pAsa1, pAsa2, pAsa3, pAsa4, and pAsa5). However, no GEIs have been reported to date in *A. salmonicida* subsp. *salmonicida* (Boyd et al., 2003; Reith et al., 2008; Studer et al., 2013).

In the present study, the genomic analysis of the 01-B526 strain revealed that its chromosome contains an additional continuous 51-kb element. Bioinformatics analyses indicated that this chromosomal insertion is a putative GEI with an unknown function. PCR genotyping of 139 *A. salmonicida* subsp. *salmonicida* isolates plus the sequencing of six additional genomes revealed the existence of two other forms of this GEI. These three new

GEIs were associated with the specific geographical area from which the isolates were recovered. Fish trading, which is an integral part of the fishery industry, can also disseminate fish pathogens (Hedrick, 1996). Since *AsaGEI*s are associated with specific geographical regions, they are promising indicators for tracking isolates responsible for furunculosis outbreaks and their propagation and, as such, provide a better understanding of furunculosis progression worldwide.

MATERIAL AND METHODS

Bacterial isolates and growth conditions

The *A. salmonicida* subsp. *salmonicida* isolates used in the present study are described in Supplementary Table 1. They were grown on furunculosis agar (10 g of Bacto-Tryptone, 5 g of yeast extract, 1 g of L-tyrosine, 2.5 g of NaCl, and 15 g of agar per liter of distilled water) (Hanninen and Hirveri-koski, 1997) for two or three days at 18°C.

PCR analyses

The DNA templates, PCR mixtures, and program cycles were as previously described (Trudel et al., 2013). The PCR assays were performed at least twice and the appropriate positive and negative control were included with each assay. The PCR primers are listed in Supplementary Table 2.

DNA extraction and genomic sequencing

The total genomic DNA of six isolates (01-B522, M15879-11, HER1085, 09-0167, 2009-144 K3, and m23067-09) was extracted using the DNeasy Blood and Tissue kit (Qiagen, Canada) and was pyrosequenced at the Plateforme d'Analyse Génomique of the Institut de Biologie Intégrative et des Systèmes (IBIS; Université Laval) using various fragmentation sizes and mate-pair libraries. The detailed protocol for making mate-pair libraries can be found in the Paired End Rapid Library Preparation Method Manual (Roche, USA). The 454 pyrosequencing reads were assembled *de novo* using Newbler version 2.5.3 (Margulies et al., 2005) with default parameters.

130 The DNA of the ATCC 33658 strain was extracted as previously described (Diamanka et al.,
131 2013). The next-generation sequencing of this strain was done using an Illumina MiSeq
132 system at the Michigan State University Research Technology Support Facility, and the
133 sequencing library was prepared using a standard Illumina TruSeq DNA sample prep kit v2.
134 The resulting reads were *de novo* assembled with the A5 pipeline (Tritt et al., 2012).
135
136 **Sequence analyses**
137 The annotated sequences for the three *AsaGEI*s were deposited in GenBank under accession
138 numbers KJ626178, KJ626179, and KJ626180 for *AsaGEI1a*, *AsaGEI1b*, and *AsaGEI2a*,
139 respectively. The open reading frames (ORFs) identified in the three genomic islands were
140 predicted by Prodigal (version 1.20) (Hyatt et al., 2010), confirmed by GeneMark (version
141 2.5) (Besemer et al., 2001), and visualized in Artemis (version 16.0.0) (Carver et al., 2012).
142 Each ORF was verified manually. BLASTn and BLASTp searches (National Center for
143 Biotechnology Information, NCBI) were conducted to determine the functions of the ORFs
144 (Altschul et al., 1990). The non-redundant (nr/nt) and whole-genome shotgun (wgs)
145 nucleotide databases were used for BLASTn. The non-redundant protein sequence (nr)
146 database was used for BLASTp.
147
148 A Perl script was used to find identities between *AsaGEI1a* and genomic regions of selected
149 bacterial strains. This script scans the region of interest with a window size of 350 bp and
150 uses the heuristic algorithm of fasta36 (Pearson and Lipman, 1988) to search for identity
151 between sequences. The cutoff was set at a minimum of 60% identity over 80% of the length.
152

153 Putative prophage ORFs were determined using Prophage Finder (Bose and Barber, 2006),
154 and tRNA was found using the tRNAscan-SE webserver (Lowe and Eddy, 1997). The
155 PipMaker (Schwartz et al., 2000) webserver was used for dotplot alignment analyses. Easyfig
156 software (Sullivan et al., 2011) was used to align the genomic elements for the construction
157 of Figure 3 and Supplementary Figures 1, 2, 3, and 4.
158
159 The six bacterial chromosomal regions used for the comparative genomics were extracted
160 from the genomes of *A. hydrophila* strain RB-AH (GenBank Bioproject PRJNA253773), *A.*
161 *jandaei* Riv2 (GenBank Bioproject PRJNA237126), *A. salmonicida* subsp. *achromogenes*
162 AS03 (Han et al., 2013), *A. veronii* AMC35 (GenBank BioProject PRJNA71519), *A.*
163 *hydrophila* subsp. *hydrophila* ATCC 7966 (Seshadri et al., 2006), and *V. cholerae* 1587
164 (GenBank BioProject PRJNA18265).
165
166 The genome of the *A. salmonicida* subsp. *salmonicida* A449 strain was used as a reference to
167 establish the insertion sites of the three GEIs (GenBank Accession CP000644.1).
168

RESULTS AND DISCUSSION

A genomic island was found in the 01-B526 strain.

Genomic comparisons between the 01-B526 pathogenic strain and the A449 reference strain of *A. salmonicida* subsp. *salmonicida* resulted in the discovery of a 51-kb insertion in the 01-B526 chromosomal genome. Automatic annotation and manual curation revealed that this element has a G+C content of 57,69% and contains 63 predicted protein-encoding ORFs, 19 (30%) of which encoded putative prophage proteins (Figure 1A). Only 13 (21%) of the ORFs had a known function. See Supplementary Table 3 for the complete list of ORFs found in the 51-kb element.

The absence of this element in A449 genome suggests that it was likely acquired by horizontal gene transfer. The element was flanked by direct 19-bp repeated sequences, which suggested that the entire element is inserted into the chromosome through the action of an integrase. ORFs for phage integrase (ORF1), phage excisionase (ORF2), and two phage recombinases (ORF14 and ORF24) were found in the 51-kb element. Although this element contained 19 out of 63 genes coding for phage proteins, it is likely not a prophage because it lacked, among others, genes coding for capsid proteins. The 51-kb element in 01-B526 thus displayed features of a GEI (Dobrindt et al., 2004) and was named *Aeromonas salmonicida Genomic Island 1a* (*AsaGEI1a*). To our knowledge, this is the first time that a GEI of this nature has been described in *A. salmonicida* subsp. *salmonicida*.

To determine the potential origin of *AsaGEI1a*, its nucleotide sequence was compared by BLASTn analyses against complete and draft genome sequence NCBI databases. A more in-

192 depth investigation was then performed using more focused alignments against precise
193 genomic regions of the six bacteria sharing the highest identities with *AsaGEI1a* (Figure 1B).
194
195 Blast analyses of the wgs NCBI database using *AsaGEI1a* as a query revealed a genomic
196 element of approximately 52.8 kb with 92% sequence identity over 67% of the length in *A.*
197 *hydrophila* RB-AH (957,063 bp to 1,009,948 bp in Contig_1). A second genomic element
198 with 85% identity over 58% of the length was found in *A. jandaei* Riv2 using *AsaGEI1a*.
199 This element in *A. jandaei* Riv2 was located between 320,688 bp to 373,431 bp in contig 3.
200 These two elements in the RB-AH and Riv2 strains have many conserved regions, with more
201 than 65% of identity between each other in addition to their identity with *AsaGEI1a*
202 (Supplementary Figure 1). It is interesting to note that the element in *A. hydrophila* RB-AH
203 is located at the same chromosomal position as *AsaGEI1a* in *A. salmonicida* subsp.
204 *salmonicida* (data not shown).
205
206 *AsaGEI1a* also displayed 90% sequence identity over 35% of the length with a genomic
207 region of *A. salmonicida* subsp. *achromogenes* AS03 (138,843 bp to 195,462 bp in Contig38)
208 (Han et al., 2013). Three genes encoding phage proteins, including the portal protein, a virion
209 structural protein, and partial tail fiber proteins, have been predicted by bioinformatics
210 analyses in this region of *A. salmonicida* subsp. *achromogenes* AS03 (data not shown),
211 suggesting that it is a putative prophage. The regions encoding the three structural phage
212 proteins in the AS03 strain are not conserved in *AsaGEI1a* (Supplementary Figure 2).
213

The presence of the two other genomic elements in *A. hydrophila* RB-AH and *A. jandaei* Riv2 displaying high identity with *AsaGEI1a* indicate that this GEI may transfer from one species to another. Since the three genetic elements share high sequence identity and a similar structure, we propose that these elements may be part of the same GEI group. Nevertheless, given the nature of the genes in *AsaGEI1a* and the level of identity shared by this genomic element and the putative prophage in *A. salmonicida* subsp. *achromogenes* AS03, it is possible that *AsaGEI1a* and other related GEIs are derivatives of a former prophage that has lost several genes while acquiring others.

Some parts of *AsaGEI1a* exhibit a high level of identity with several genes in bacterial human pathogen genomes such as *A. hydrophila* subsp. *hydrophila*, *A. veronii*, and *V. cholerae* (see Supplementary Table 4 for the square matrix of the genomic alignments of the genomic elements sharing similarities with *AsaGEI1a*, and Supplementary Figure 3 for the alignment of *AsaGEI1a* with the six bacterial genomic regions discussed above).

Most of the time, sequences in *Aeromonas* genomes similar to *AsaGEI1a* were also similar to those of other bacteria of the same genus, while distinct regions of *AsaGEI1a* were similar to *V. cholerae* 1587 (Figure 1B). Although several *AsaGEI1a* ORFs displaying high levels of identity with orthologs from various pathogenic bacteria have been identified, sequence analyses using tools that detect genes encoding potential virulence factors gave no significant results. On the other hand, the vast majority of the sequences that are common to *AsaGEI1a* and the pathogens contain genes encoding unknown proteins.

Different variants of *AsaGEI* in *A. salmonicida* subsp. *salmonicida*

A PCR genotyping approach was used to screen for the presence of *AsaGEI1a* in other *A. salmonicida* subsp. *salmonicida* isolates. Nine primer pairs were consequently designed. Seven targeted specific ORFs (*orf11*, *orf19*, *orf33*, *orf54*, *orf57*, *orf67,* and *orf78*) while two others were designed to amplify the insertion sites of *AsaGEI1a* in the chromosome (*ins1* and *ins2*). See Supplementary Table 2 for a list of the primers used in this study. The nine primer pairs were used to screen 139 *A. salmonicida* subsp. *salmonicida* isolates in order to identify those bearing *AsaGEI1a*. See Supplementary Table 1 for the genotyping results. These isolates came from various European and North American regions. The screening showed that 29 of the isolates were positive for all the targets tested by the PCR assay, meaning that they likely contained an intact *AsaGEI1a* (Figure 2A). On the other hand, 15 isolates, as well as A449, were not amplified by the PCR assay (Figure 2B), indicating that some of the isolates do not bear *AsaGEI1a*.

The most surprising result was that a subset of PCR products was detected in many isolates. Detailed analyses detected two genotyping profiles that differed from the profile obtained with the 01-B526 strain. The first is illustrated in Figure 2C, where the isolate (HER1085) gave amplicons for all the targets except for *orf11*, *orf19,* and *orf33*. Twelve isolates had this profile. The second profile corresponded to amplicons for *orf57*, *orf67*, and *orf78* (see example of isolate 09-0167, Figure 2D). This profile was found in the majority of the isolates (82 of 139). Based on the genotyping PCR, it appeared that three different forms of *AsaGEI* exist in *A. salmonicida* subsp. *salmonicida*. Since the second *AsaGEI* was positive for the same insertion sites as *AsaGEI1a* (Figure 2C), it was named *AsaGEI1b* while the third was

named *AsaGEI2a* because it was inserted at a different location in the chromosome based on the negative PCR results for *ins1* and *ins2* (Figure 2D).

Structure of *AsaGEI1b* and *AsaGEI2a* and their relationship with *AsaGEI1a*

In order to analyze the structures of the *AsaGEI* variants, we sequenced the genomes of seven additional *A. salmonicida* subsp. *salmonicida* isolates: two bearing *AsaGEI1a* (01-B522 and M15879-11), two bearing *AsaGEI1b* (HER1085 and ATCC 33658), and three bearing *AsaGEI2a* (09-0167, m23067-09, and 2009-144 K3). The sequence analyses showed that the three *AsaGEI1a* sequences shared a high degree of identity to each other (99%) (Supplementary Table 5). It was the same for the three *AsaGEI2a* sequences compared to each other (99%) and for the identity of the two *AsaGEI1b* (99%).

Figures 3A, B, and C show dotplot analyses of the three *AsaGEI*s. The three alignments show that the gene orders were conserved among the GEIs. No inversions or recombinations were observed. The genomic alignments of *AsaGEI1a* and *AsaGEI1b* were punctuated by numerous insertions/deletions mostly concentrated in the first half of the two GEIs (Figure 3A). The alignments also revealed the presence of regions that differ between *AsaGEI1a* and *AsaGEI1b* even though both have similar lengths. The insertions/deletions did not seem to be limited to the first half in the *AsaGEI2a* alignments with the other *AsaGEI*s (Figures 3B and C). Here again, some regions in the first half of *AsaGEI2a* were completely different from those in the first halves of the other *AsaGEI*s. Interestingly, the sequences in the last 40% of the "right-end" regions of the three GEIs shared high levels of identity.

A copy of *AsaGEI2a* from isolate 09-0167 as well as the copy of *AsaGEI1b* from strain HER1085 were annotated, and their detailed structures are shown in Figure 3D. This figure also presents the genomic alignments of the three *AsaGEI*s. See Supplementary Table 6 and Supplementary Table 7 for the annotated ORFs of *AsaGEI1b* and *AsaGEI2a*, respectively. The alignments of the three GEIs showed that the same integrase (ORF1) and excisionase (ORF2) encoding genes are present in *AsaGEI1a* and *AsaGEI1b*, while different integrase and excisionase genes are found in *AsaGEI2a* (Figure 3D-a). Given the specificity of the integrase (Fogg et al., 2014), this difference could explain why *AsaGEI1* and *AsaGEI2* are inserted at different sites on the chromosome (see below). The structures of the *AsaGEI*s can be divided into two parts, with a sudden shift of transcriptional orientation between *orf34* and *orf35*, a featured shared by the various forms of the *AsaGEI*s (green diamond in Figure 3D). Most of the first halves of the *AsaGEI*s were interspersed with variable regions that were located at the same positions (see example in Figure 3D-b) and were bordered by small conserved regions (see example in Figure 3D-c).

Different ORFs with similar functions were located at the same positions in the various GEIs. For example, the genes for *orf17* were different in the three *AsaGEI*s, but all encoded a putative phage methylase. The same is true of the genes for *orf34*, which were different, but all encoded a DNA-binding protein. The fact that the *orf17* and *orf34* gene products are different but have the same function suggests they are important for the functionality of the *AsaGEI*s. The annotation of *AsaGEI2a* revealed the presence of a $tRNA^{Pro}$ gene (indicated in blue in Figure 3D). The co-linearity comparison between *AsaGEI2a* and the genetic elements in *A. jandaei* Riv2 and *A. salmonicida* subsp. *achromogenes* AS03 (Supplementary Figure 4)

306 showed that $tRNA^{Pro}$ is present in the same location in the three elements (indicated by a
307 black diamond). However, further analyses will be required to determine the origin of this
308 tRNA and why only *AsaGEI2a* bears it. Overall, the GEIs in *A. salmonicida* subsp.
309 *salmonicida* were composed of two sections: one in the left portion that was highly variable
310 and was characterized by numerous small insertions/deletions, and a second section in the
311 left portion that mainly bore genes encoding hypothetical proteins that were conserved in the
312 three GEIs (Figure 3D-d and Figure 3D-e, respectively).
313
314 **Chromosome insertion sites of *AsaGEI*s**
315 The insertion sites of *AsaGEI1(a,b)* and *AsaGEI2a* are illustrated in Figure 4. The two types
316 of GEIs were integrated into different genomic sites. *AsaGEI1a* and *AsaGEI1b* were
317 integrated in the beginning of the *ASA_2666* gene that encodes a hypothetical conserved
318 protein (data not shown). *AsaGEI2a* was integrated downstream a $tRNA^{Leu}$ gene at a distance
319 of six genes from the prophage 1 (Figure 4B) (Reith et al., 2008). The genomic sequencing of
320 two of the three *A. salmonicida* subsp. *salmonicida* isolates harboring an *AsaGEI2a* revealed
321 the presence of another prophage that we called prophage 3. This prophage was integrated
322 between *AsaGEI2a* and the $tRNA^{Leu}$ gene (Figure 4C). Following the discovery of prophage
323 3, a primer pair was designed to screen the collection of *A. salmonicida* subsp. *salmonicida*
324 isolates for this prophage. The screening results showed that 77 of the 82 isolates bearing
325 *AsaGEI2a* also bear prophage 3. In five *AsaGEI2a*-positive isolates lacking prophage 3, four
326 were from Quebec (Canada) and one was from Europe. Of the four isolates from Quebec,
327 three were from the same fish region (data not shown). Prophage 3 was exclusively found in
328 isolates bearing *AsaGEI2a* in North America, with the exception of three isolates that bore

329 *AsaGEI1a* and one that bore *AsaGEI1b* instead of *AsaGEI2a*. These four isolates were from
330 the Great Lakes region (Supplementary Table 1).
331
332 **Correlation between the form of the *AsaGEI*s and the geographical origins of the**
333 **isolates**
334 The results of the GEI genotyping were grouped by the geographical origins of the *A.*
335 *salmonicida* subsp. *salmonicida* isolates (Table 1). The geographical origins of the 139
336 isolates and the GEI genotyping results are listed in Supplementary Table 1. Four trends
337 could be observed: (1) 100% (29/29) of the isolates bearing *AsaGEI1a* were isolated from the
338 Great Lakes-St. Lawrence River System region, (2) 87% (71/82) and 12% (10/82) of the
339 isolates bearing *AsaGEI2a* were isolated from hydrographic basin of the Great Lakes-St.
340 Lawrence River System and the Canadian Maritimes, respectively, meaning that 99% of the
341 isolates bearing *AsaGEI2a* were from North America, (3) 100% (10/10) of the isolates from
342 the Canadian Maritimes had an *AsaGEI2a* genotype, and (4) 83% (10/12) of the isolates
343 bearing an *AsaGEI1b* and 88% (14/16) of the isolates without an *AsaGEI* were from Europe.
344 Each *AsaGEI* appeared to be mainly found in a specific region: *AsaGEI1a and AsaGEI2a* in
345 eastern North America and *AsaGEI1b* and isolates without GEIs in Europe. Since both
346 *AsaGEI1*s were at the same site in the chromosome, an independent integration event on each
347 continent may have occurred that would explain the existence of these two similar but
348 distinct GEIs, or an ancestral GEI may have been integrated that subsequently evolved into
349 *AsaGEI1a* and *AsaGEI1b*. Another interesting point is that prophage 3 was only observed in
350 North American isolates.
351

Based on low-resolution methods, *A. salmonicida* subsp. *salmonicida* isolates have been generally considered as genetically homogeneous (Belland and Trust, 1988; García et al., 2000; Hänninen et al., 1995; Miyata et al., 1995; O'hIci et al., 2000; Studer et al., 2013). Next-generation sequencing revealed that *A. salmonicida* subsp. *salmonicida* isolates display a high level of genetic diversity. The present study showed for the first time that there is a significant chromosomal variation in this bacterium. The sequencing of the 01-B526 pathogenic strain resulted in the discovery of the first GEI other than a prophage in the *A. salmonicida* subsp. *salmonicida* genome. Lastly, our analyses revealed the existence of three GEIs and the presence of a new prophage (prophage 3) in *A. salmonicida* subsp. *salmonicida* isolates. BLASTn analyses revealed the presence of genetic elements in two other *Aeromonas* species (*A. hydrophila* and *A. jandaei*) that are highly similar to the three GEIs in *A. salmonicida* subsp. *salmonicida* isolates (i.e., *AsaGEI*s). This points to potential horizontal gene transfers between these *Aeromonas* species. The five genetic elements (the three *AsaGEI*s and the two in *A. hydrophila* and *A. jandaei*) appear to have evolved differently from an ancestor element. The PCR screening used to determine the distribution of the GEIs in 139 *A. salmonicida* subsp. *salmonicida* isolates suggested that each *AsaGEI* is specific to a given geographical region, as is also the case for prophage 3. During a furunculosis outbreak, the *AsaGEI* type of an isolate could be used to determine whether it was a contaminant from another continent transported by human activity. For example, this approach could be used in the province of Quebec to have an idea whether an isolate is from Europe by looking for the presence of *AsaGEI1b*. *AsaGEI*s as well as prophage 3 thus appear to be good examples of genetic elements that can be used as epidemiological indicators for *A. salmonicida* subsp. *salmonicida* isolates. To use *AsaGEI*s and prophage 3 as strong indicators of the geographic

375 origins of *A. salmonicida* subsp. *salmonicida* isolates, it will be necessary to increase the
376 number of isolates from the Canadian Maritimes and European countries. Moreover, it will
377 be interesting to determine whether *AsaGEI*s are present in *A. salmonicida* subsp.
378 *salmonicida* worldwide or whether they are restricted to North America and Europe. It is also
379 possible to imagine that other forms of *AsaGEI1* and *AsaGEI2* as well as other GEIs may
380 exist in *A. salmonicida* subsp. *salmonicida* and that they have had an impact on the biology
381 of this bacterium. Future analyses on the rest of the genomic sequences now available may
382 also reveal additional divergence between the isolates.
383

ACKNOWLEDGMENTS

We thank the Laboratoire de bactériologie clinique, Faculté de médecine vétérinaire, Université de Montréal (Montreal, QC, Canada), the Félix d'Hérelle Reference Center (Département de biochimie, de microbiologie et de bio-informatique, Université Laval, Quebec City, QC, Canada), the Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec, and the Aquatic Animal Health Department, Fisheries and Oceans Canada (New Brunswick, Canada) as well as J. Frey (University of Bern) and R.M.W. Stevenson (University of Guelph) for the *A. salmonicida* subsp. *salmonicida* isolates.

JGER, MVT, and SA received scholarships from the CREATE program of Ressources Aquatiques Quebec (RAQ) while KHT and FB received scholarships from the Undergraduate Student Research Awards Program of the Natural Sciences and Engineering Research Council of Canada (NSERC). JGER received also a scholarship from Fonds de la recherche québécois sur la nature et les technologies (FRQNT). KHT received an Alexander Graham Bell Canada Graduate Scholarship from NSERC.

This project was supported by an NSERC Discovery grant to SJC, a grant from the Société de recherche et de développement en aquaculture continentale (SORDAC) to SJC and ND, a grant from Resources Aquatiques Québec (RAQ) to M. Frenette, and a grant from the United States Department of Agriculture - Animal and Plant Health Inspection Service (USDA-APHIS; 10-9100-1293-GR) to M. Faisal.

406 **REFERENCES**

407 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment
408 search tool. J. Mol. Biol. 215, 403–10.

409 Beaz-Hidalgo, R., Figueras, M.J., 2013. *Aeromonas* spp. whole genomes and virulence
410 factors implicated in fish disease. J. Fish Dis. 36, 371–88.

411 Belland, R.J., Trust, T.J., 1988. DNA:DNA reassociation analysis of *Aeromonas*
412 *salmonicida*. J. Gen. Microbiol. 134, 307–15.

413 Bellanger, X., Payot, S., Leblond-Bourget, N., Guédon, G., 2013. Conjugative and
414 mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol. Rev.
415 38, 1–42.

416 Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for
417 prediction of gene starts in microbial genomes. Implications for finding sequence motifs
418 in regulatory regions. Nucleic Acids Res. 29, 2607–18.

419 Bose, M., Barber, R.D., 2006. Prophage Finder: a prophage loci prediction tool for
420 prokaryotic genome sequences. In Silico Biol. 6, 223–7.

421 Boyd, J., Williams, J., Curtis, B., Kozera, C., Singh, R., Reith, M., 2003. Three small, cryptic
422 plasmids from *Aeromonas salmonicida* subsp. *salmonicida* A449. Plasmid 50, 131–144.

423 Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J. a, 2012. Artemis: an
424 integrated platform for visualization and analysis of high-throughput sequence-based
425 experimental data. Bioinformatics 28, 464–9.

426 Charette, S.J., Brochu, F., Boyle, B., Filion, G., Tanaka, K.H., Derome, N., 2012. Draft
427 genome sequence of the virulent strain 01-B526 of the fish pathogen *Aeromonas*
428 *salmonicida*. J. Bacteriol. 194, 722–3.

429 Dallaire-Dufresne, S., Tanaka, K.H., Trudel, M. V, Lafaille, A., Charette, S.J., 2014.
430 Virulence, genomic features, and plasticity of *Aeromonas salmonicida* subsp.
431 *salmonicida*, the causative agent of fish furunculosis. Vet. Microbiol. 169, 1–7.

432 Darmon, E., Leach, D.R.F., 2014. Bacterial genome instability. Microbiol. Mol. Biol. Rev.
433 78, 1–39.

434 Diamanka, A., Loch, T.P., Cipriano, R.C., Faisal, M., 2013. Polyphasic characterization of
435 *Aeromonas salmonicida* isolates recovered from salmonid and non-salmonid fish. J.
436 Fish Dis. 36, 949–63.

437 Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J., 2004. Genomic islands in pathogenic
438 and environmental microorganisms. Nat. Rev. Microbiol. 2, 414–24.

439 European Commission, 2014. European aquaculture figures [WWW Document]. URL
440 http://ec.europa.eu/fisheries/cfp/aquaculture/facts/index_en.htm

441 Fisheries and Oceans Canada, 2013. 2011 Canadian Aquaculture Production Statistics
442 [WWW Document]. URL http://www.dfo-mpo.gc.ca/stats/aqua/aqua11-eng.htm

443 Fogg, P.C.M., Colloms, S., Rosser, S., Stark, M., Smith, M.C.M., 2014. New Applications
444 for Phage Integrases. J. Mol. Biol. 426, 2703–2716.

445 Food and Agriculture Organization of the United Nations, 2014. The State of World
446 Fisheries and Aquaculture 2014.

447 García, J.A., Larsen, J.L., Dalsgaard, I., Pedersen, K., 2000. Pulsed-field gel electrophoriesis
448 analyis of *Aeromonas salmonicida* ssp. *salmonicida*. FEMS Microbiol. Lett. 190, 163–6.

449 Han, J.E., Kim, H., Shin, P., Jun, W., Chai, Y., Park, C., 2013. Draft genome sequence of
450 *Aeromonas salmonicida* subsp. *achromogenes* AS03, an atypical strain isolated from
451 Crucian Carp (*Carassius carassius*) in the Republic of Korea. Genome Announc. 1,
452 2007–2008.

453 Hanninen, M., Hirveri-koski, V., 1997. Microbiology molecular and phenotypic methods for
454 the characterization of atypical *Aeromonas salmonicida*. Vet. Microbiol. 56, 147–158.

455 Hänninen, M.L., Ridell, J., Hirvelä-Koski, V., 1995. Phenotypic and molecular
456 characteristics of *Aeromonas salmonicida* subsp. *salmonicida* isolated in southern and
457 northern Finland. J. Appl. Bacteriol. 79, 12–21.

458 Hedrick, R.P., 1996. Movements of pathogens with the international trade of live fish:
459 problems and solutions. Rev. Sci. Tech. 15, 523–531.

460 Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010.
461 Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC
462 Bioinformatics 11, 119.

463 Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., Crook, D.W., 2009.
464 Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS
465 Microbiol. Rev. 33, 376–93.

466 Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer
467 RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–64.

468 Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L. a, Berka, J.,
469 Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.

V, Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K. a, Volkmer, G. a, Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–80.

Miyata, M., Aoki, T., Inglis, V., Yoshida, T., Endo, M., 1995. RAPD analysis of *Aeromonas salmonicida* and *Aeromonas hydrophila*. J. Appl. Bacteriol. 79, 181–5.

Morin, R., 2010. L'utilisation des antibiotiques pour combattre la furonculose chez l'omble de fontaine génère de l'antibiorésistance chez *Aeromonas salmonicida* L'aquicole. Le bulletin l'Association des aquaculteurs du Québec 15,2–6.

O'hIci, B., Olivier, G., Powell, R., 2000. Genetic diversity of the fish pathogen *Aeromonas salmonicida* demonstrated by random amplified polymorphic DNA and pulsed-field gel electrophoresis analyses. Dis. Aquat. Organ. 39, 109–19.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A. 85, 2444–8.

Reith, M.E., Singh, R.K., Curtis, B., Boyd, J.M., Bouevitch, A., Kimball, J., Munholland, J., Murphy, C., Sarty, D., Williams, J., Nash, J.H., Johnson, S.C., Brown, L.L., 2008. The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. BMC Genomics 9, 427.

Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., Miller, W., 2000. PipMaker--a web server for aligning two genomic DNA sequences. Genome Res. 10, 577–86.

Seshadri, R., Joseph, S.W., Chopra, A.K., Sha, J., Shaw, J., Graf, J., Haft, D., Wu, M., Ren, Q., Rosovitz, M.J., Madupu, R., Tallon, L., Kim, M., Jin, S., Vuong, H., Stine, O.C., Ali, A., Horneman, A.J., Heidelberg, J.F., 2006. Genome sequence of *Aeromonas hydrophila* ATCC 7966T: jack of all trades. J. Bacteriol. 188, 8272–82.

Studer, N., Frey, J., Vanden Bergh, P., 2013. Clustering subspecies of *Aeromonas salmonicida* using IS630 typing. BMC Microbiol. 13, 13–36.

Sullivan, M.J., Petty, N.K., Beatson, S. a, 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–10.

Tritt, A., Eisen, J. a, Facciotti, M.T., Darling, A.E., 2012. An integrated pipeline for de novo assembly of microbial genomes. PLoS One 7, e42304.

505 Trudel, M. V, Tanaka, K.H., Filion, G., Daher, R.K., Frenette, M., Charette, S.J., 2013.
506 Insertion sequence AS5 (IS*AS5*) is involved in the genomic plasticity of *Aeromonas*
507 *salmonicida*. Mob. Genet. Elements 5, 1–7.

508

509

FIGURE LEGENDS

Figure 1. Genetic features of *AsaGEI1a*. (A) Schematic representation of the GEI. Each arrow represents an ORF. The names of the ORFs are indicated above the arrows. The purple, red, cyan and yellow arrows represent ORFs encoding phage protein with predicted function, hypothetical phage proteins, protein with predicted function and hypothetical proteins, respectively. (B) Identity levels of genes from *AsaGEI1a* with sequences found in the genomes of other bacterial species. Identity values with sequences from the six bacterial species with the highest similarities are shown.

Figure 2. Genotyping of *A. salmonicida* subsp. *salmonicida* isolates for the presence of nine sequences found in *AsaGEI1a*. The chromosomal gene *tapA* was used as a positive control. Electrophoresis analysis of isolates bearing *AsaGEI1a* (A), with no GEI (B), bearing *AsaGEI1b* (C), and bearing *AsaGEI2a* (D). The name of each isolate used as an example is indicated on the image. The results of positive isolates of the 139 isolates analyzed for each genotyping profile are indicated on the right of the image.

Figure 3. Alignments of the three *AsaGEI*s. Dotplot similarity distributions of *AsaGEI1b* and *AsaGEI1a* (A), *AsaGEI1a* and *AsaGEI2a* (B), and *AsaGEI1b* and *AsaGEI2a* (C). D. Co-linearity comparison of the three *AsaGEI*s. The gray boxes represent regions sharing similarity with regions found by the Easyfig software. The green diamond indicates the site of transcriptional divergence. The same integrase (ORF1) and excisionase (ORF2) encoding genes are present in *AsaGEI1a* and *AsaGEI1b*, while different integrase and excisionase

genes are found in *AsaGEI2a* (highlighted by the line "a" in panel D). Most of the first halves of the *AsaGEI*s were interspersed with variable regions that were located at the same positions (line "b" in panel D as an example) and were bordered by small conserved regions (line "c" in panel D as an example). Lines "d" and "e" in panel D define the two sections in GEIs: one in the left portion (d) that is highly variable and a second section in the left portion (e) that mainly bore genes encoding hypothetical proteins that were conserved in the three GEIs.

Figure 4. Schematic representation of the insertion sites of the three *AsaGEI*s relative to the genome of the *A. salmonicida* subsp. *salmonicida* A449 strain. The black and white rectangles represent *AsaGEI1*(*a,b*) and *AsaGEI2a,* respectively. ORF1 and ORF82 are indicated for each GEI. Insertion sites of *AsaGEI1a* and *AsaGEI1b* (A). Insertion site of *AsaGEI2a* in association (C) or not (B) with prophage 3, which is found in some isolates. The proportions of the 82 *AsaGEI2a*-containing isolates bearing prophage 3 (C) or not (B) are indicated. In B and C, the six black lines represent the genes between prophage 1 and the insertion site of *AsaGEI2a*.

Table 1. Distribution of *A. salmonicida* subsp. *salmonicida* isolates based on geographical origin and the presence of *AsaGEI*s

	AsaGEI				**Total**
	1a	***2a***	***1b***	**none**	
Great Lakes-St. Lawrence River System	29	71	2	2	104
Canadian Maritimes	0	10	0	0	10
Europe [a]	0	1	10	14	25
Total	29	82	12	16	139

a: The isolates were from Switzerland (n=12), Norway (n=5), France (n=4), the United Kingdom (n=2), Denmark (n=1), and an unknown European country (n=1).

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Co-linearity comparison of *AsaGEI1a* with the genetic elements in *A. hydrophila* RB-AH and *A. jandaei* Riv2.

Supplementary Figure 2. Co-linearity comparison of *AsaGEI1a* with the genetic element in *A. salmonicida* subsp. *achromogenes* AS03.

Supplementary Figure 3. Co-linearity comparison of *AsaGEI1a* with the genetic elements in *V. cholera* 1587, *A. hydrophila* RB-AH, *A. jandaei* Riv2, *A. salmonicida* subsp. *achromogenes* AS03, *A. veronii* AMC35, and *A. hydrophila* subsp. *hydrophila* ATCC 7966.

Supplementary Figure 4. Co-linearity comparison of *AsaGEI2a* with the genetic elements in *A. jandaei* Riv2 and *A. salmonicida* subsp. *achromogenes* AS03.

LIST OF SUPPLEMENTARY TABLES

Supplementary Table 1. *A. salmonicida* subsp. *salmonicida* isolates used in this study

Supplementary Table 2. Primers used in this study

Supplementary Table 3. ORFs found in the 51-kb insertion in 01-B526 (*AsaGEI1a*)

Supplementary Table 4. Square matrix of the alignments of the genomic elements sharing similarities with *AsaGEI1a*

Supplementary Table 5. Square matrix of the genomic comparison of the three *AsaGEI*s in eight strains of *A. salmonicida* subsp. *salmonicida*

Supplementary Table 6. List of *AsaGEI1b* ORFs

Supplementary Table 7. List of *AsaGEI2a* ORFs

Figure 1

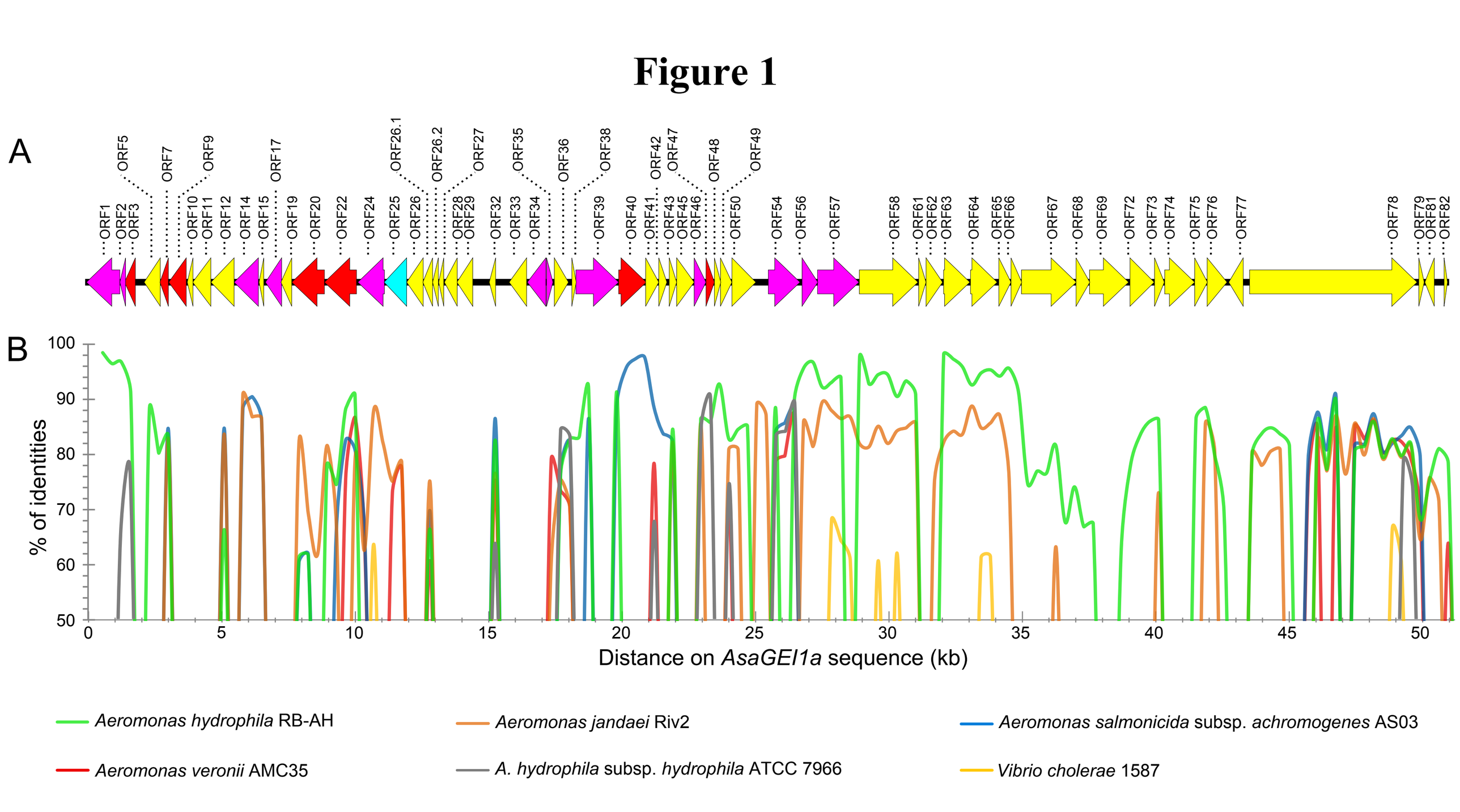

A
ORF1
ORF2
ORF3
ORF5
ORF7
ORF9
ORF10
ORF11
ORF12
ORF14
ORF15
ORF17
ORF19
ORF20
ORF22
ORF24
ORF25
ORF26
ORF26.1
ORF26.2
ORF27
ORF28
ORF29
ORF32
ORF33
ORF34
ORF35
ORF36
ORF38
ORF39
ORF40
ORF41
ORF42
ORF43
ORF45
ORF46
ORF47
ORF48
ORF49
ORF50
ORF54
ORF56
ORF57
ORF58
ORF61
ORF62
ORF63
ORF64
ORF65
ORF66
ORF67
ORF68
ORF69
ORF72
ORF73
ORF74
ORF75
ORF76
ORF77
ORF78
ORF79
ORF81
ORF82
B
% of identities
100
90
80
70
60
50
0
5
10
15
20
25
30
35
40
45
50
Distance on AsaGEI1a sequence (kb)
Aeromonas hydrophila RB-AH
Aeromonas jandaei Riv2
Aeromonas salmonicida subsp. achromogenes AS03
Aeromonas veronii AMC35
A. hydrophila subsp. hydrophila ATCC 7966
Vibrio cholerae 1587

Figure 2

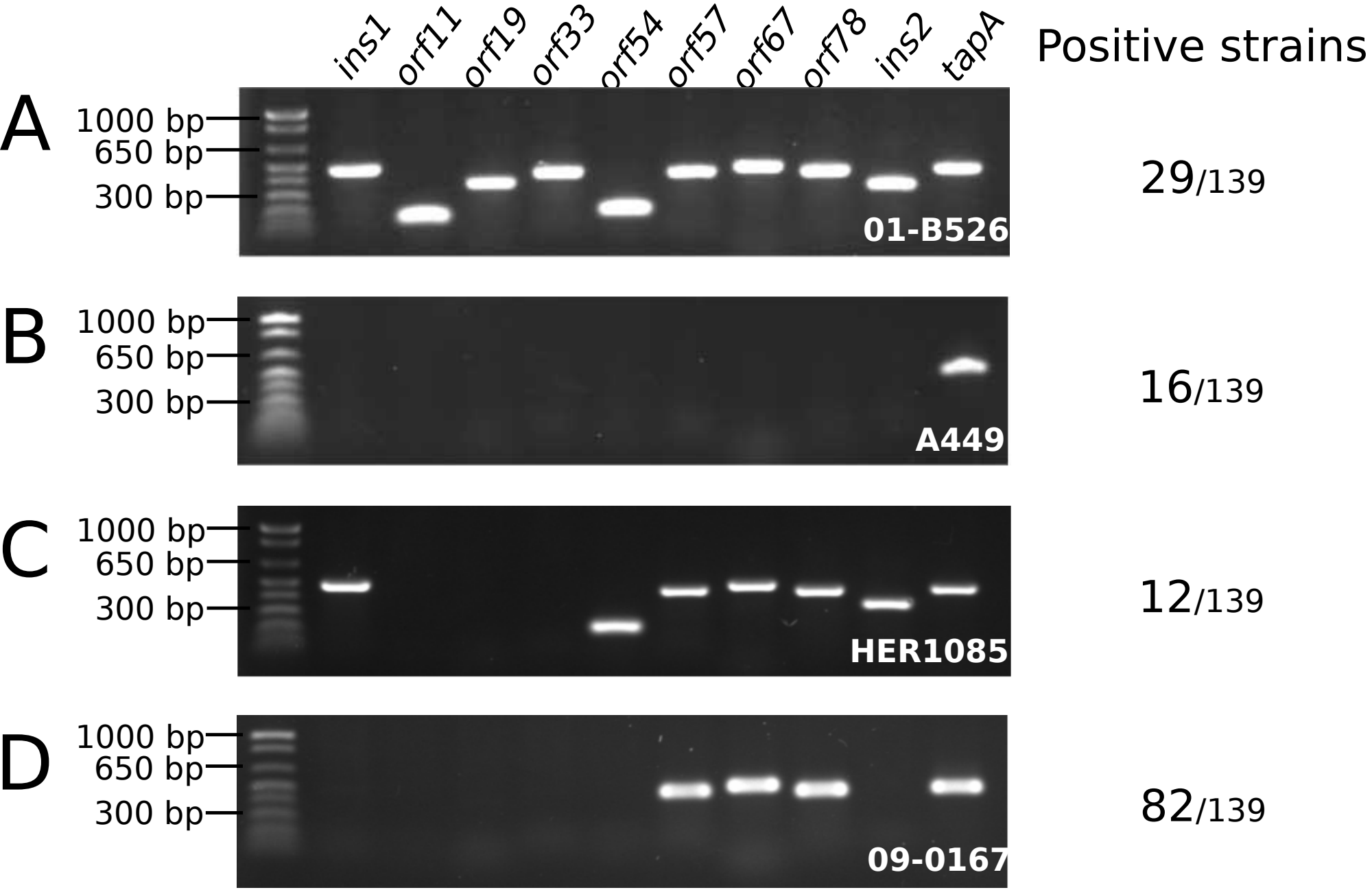

ins1
orf11
orf19
orf33
orf54
orf57
orf67
orf78
ins2
tapA
Positive strains
A
1000 bp
650 bp
300 bp
01-B526
29/139
B
1000 bp
650 bp
300 bp
A449
16/139
C
1000 bp
650 bp
300 bp
HER1085
12/139
D
1000 bp
650 bp
300 bp
09-0167
82/139

Figure 3

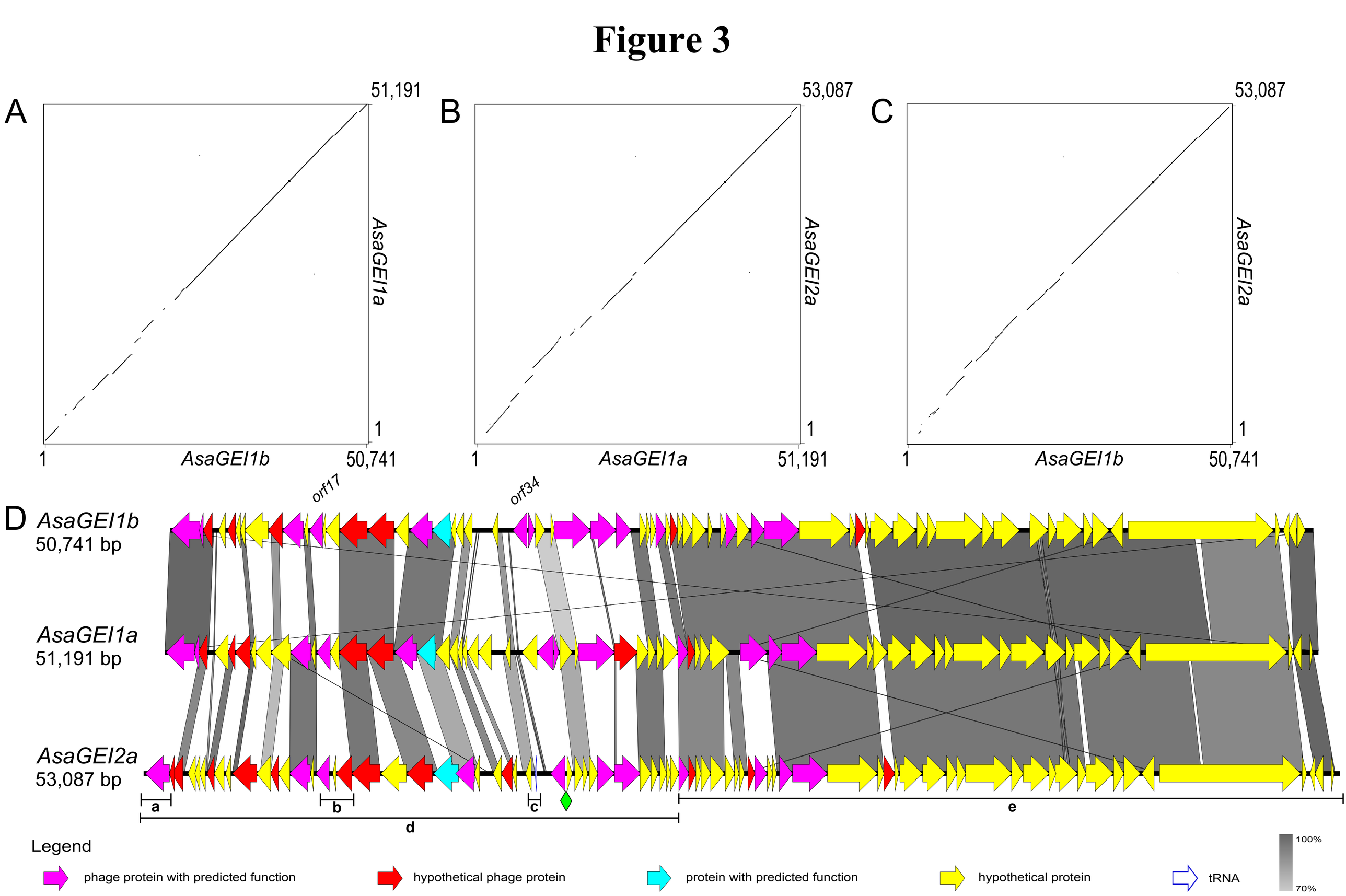
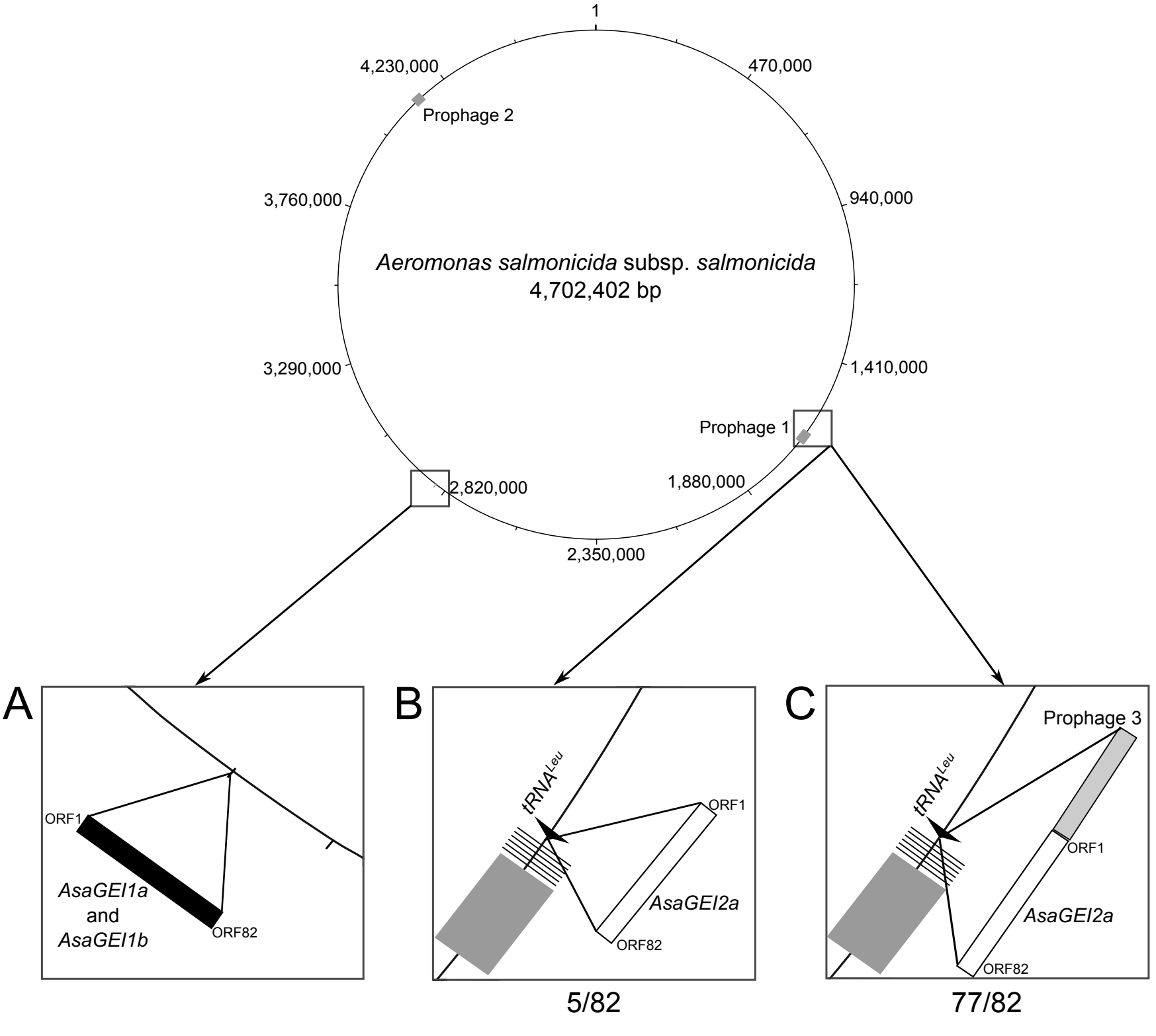
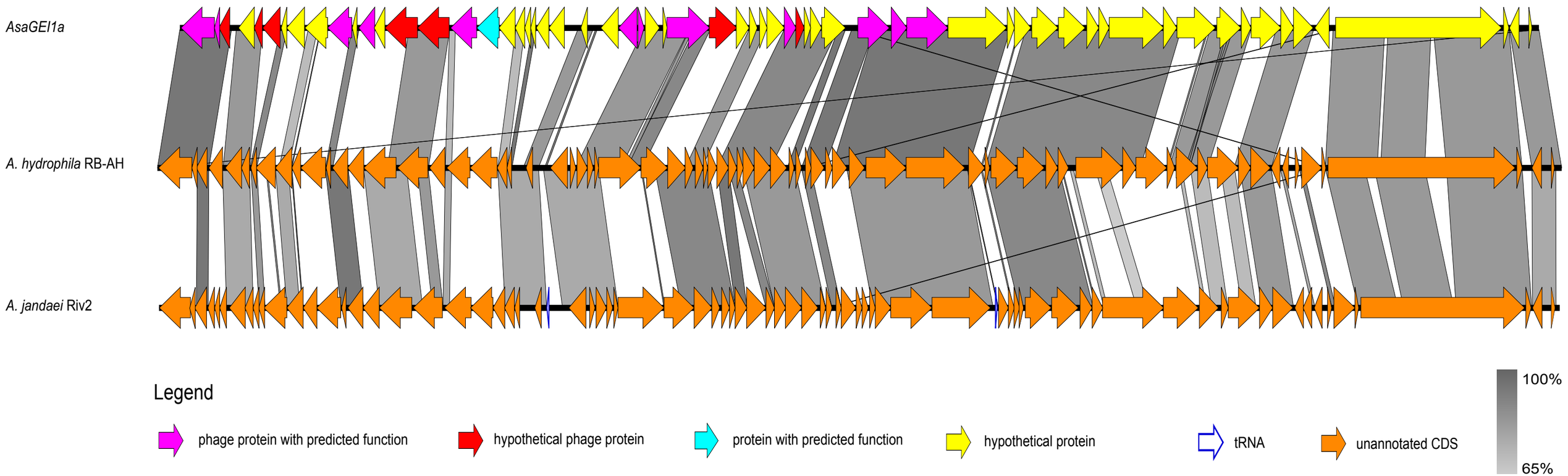

A
51,191
AsaGEI1a
1
1
AsaGEI1b
50,741
B
53,087
AsaGEI2a
1
1
AsaGEI1a
51,191
C
53,087
AsaGEI2a
1
1
AsaGEI1b
50,741
D
orf17
orf34
AsaGEI1b
50,741 bp
AsaGEI1a
51,191 bp
AsaGEI2a
53,087 bp
a
b
c
d
e
Legend
phage protein with predicted function
hypothetical phage protein
protein with predicted function
hypothetical protein
tRNA
100%
70%

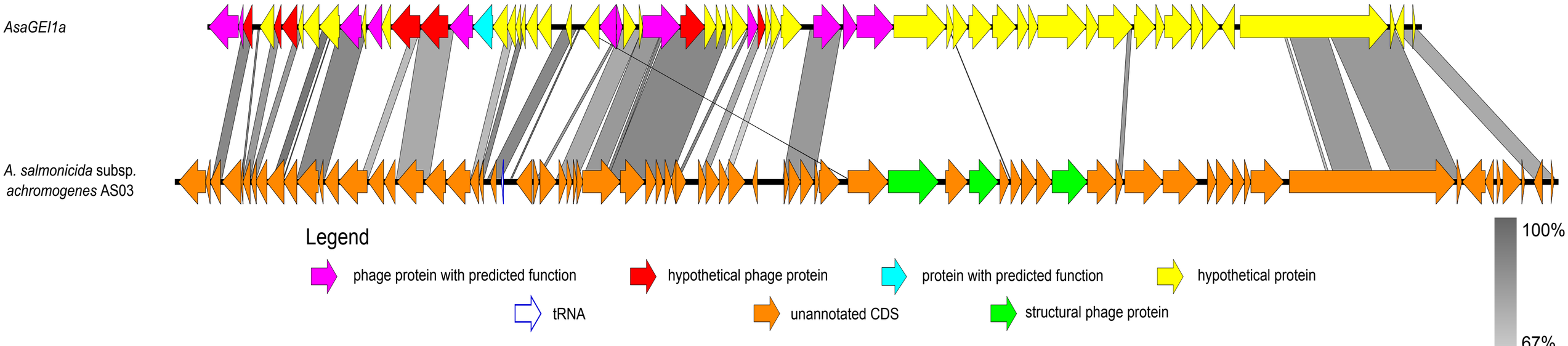
Figure 4


1
470,000
940,000
1,410,000
1,880,000
2,350,000
2,820,000
3,290,000
3,760,000
4,230,000
Prophage 2
Prophage 1
Aeromonas salmonicida subsp. salmonicida
4,702,402 bp
A
ORF1
AsaGEI1a
and
AsaGEI1b
ORF82
B
tRNALeu
ORF1
AsaGEI2a
ORF82
5/82
C
Prophage 3
tRNALeu
ORF1
AsaGEI2a
ORF82
77/82

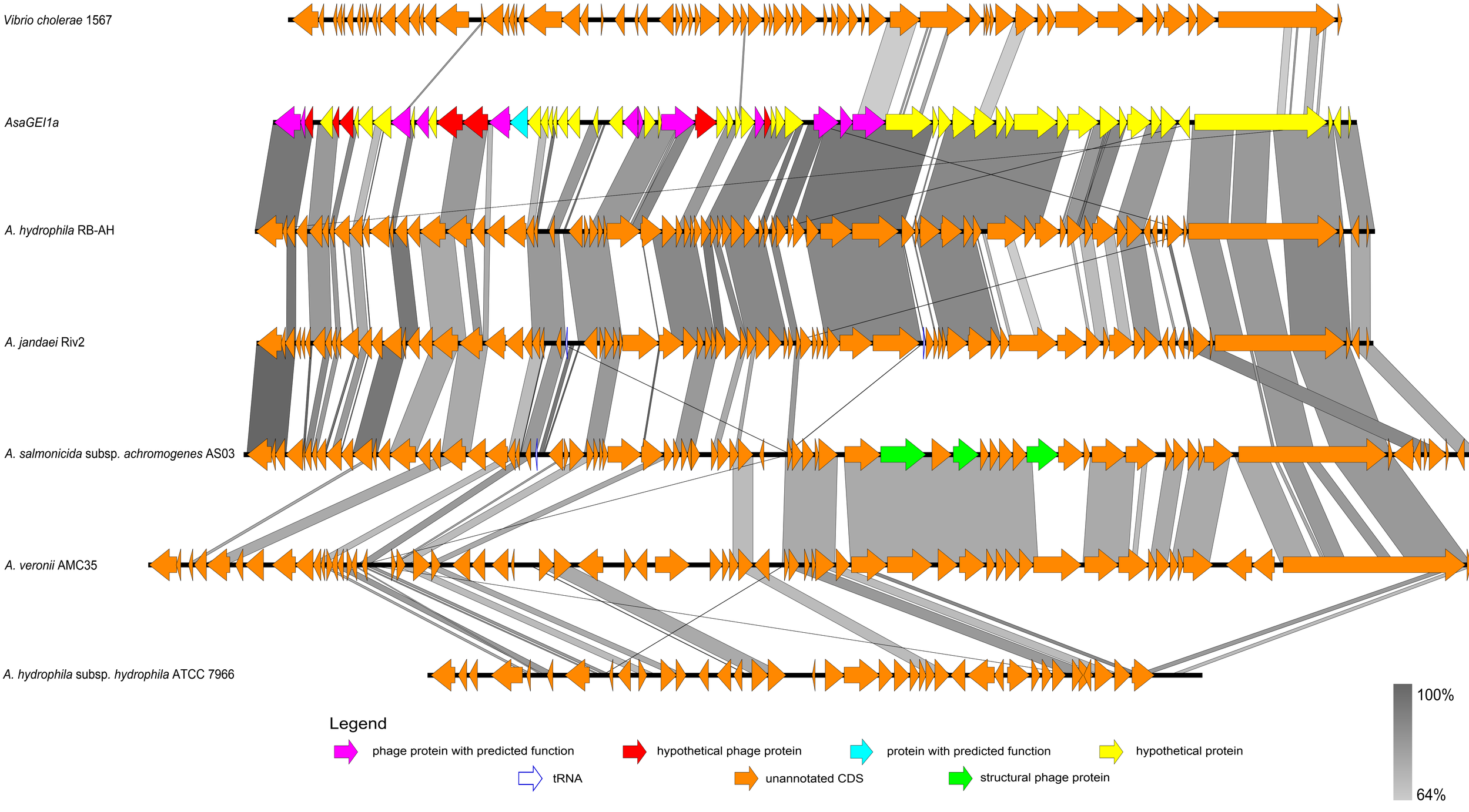
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4