

# SCIENTIFIC REPORTS



OPEN

## Novel insights into the pathogenicity of epidemic *Aeromonas hydrophila* ST251 clones from comparative genomics

Received: 29 December 2014

Accepted: 13 March 2015

Published: 27 May 2015

Maoda Pang<sup>1,\*</sup>, Jingwei Jiang<sup>2,\*</sup>, Xing Xie<sup>1</sup>, Yafeng Wu<sup>1</sup>, Yuhao Dong<sup>1</sup>, Amy H. Y. Kwok<sup>2</sup>, Wei Zhang<sup>1</sup>, Huochun Yao<sup>1</sup>, Chengping Lu<sup>1</sup>, Frederick C. Leung<sup>2,3</sup> & Yongjie Liu<sup>1</sup>

Outbreaks in fish of motile *Aeromonad* septicemia (MAS) caused by *Aeromonas hydrophila* have caused a great concern worldwide. Here, for the first time, we provide two complete genomes of epidemic *A. hydrophila* strains isolated in China. To gain an insight into the pathogenicity of epidemic *A. hydrophila*, we performed comparative genomic analyses of five epidemic strains belonging to sequence type (ST) 251, together with the environmental strain ATCC 7966<sup>T</sup>. We found that the known virulence factors, including a type III secretion system, a type VI secretion system and lateral flagella, are not required for the high virulence of the ST251 clonal group. Additionally, our work identifies three utilization pathways for *myo*-inositol, sialic acid and L-fucose providing clues regarding the factors that underlie the epidemic and virulent nature of ST251 *A. hydrophila*. Based on the geographical distribution and biological resources of the ST251 clonal group, we conclude that ST251 is a high-risk clonal group of *A. hydrophila* which may be responsible for the MAS outbreaks in China and the southeastern United States.

*Aeromonas hydrophila* is ubiquitous in various aquatic environments, and has been considered as a pathogen of fish, amphibians, reptiles and mammals<sup>1</sup>. Being an opportunistic pathogen, *A. hydrophila* could still cause the outbreaks of motile *Aeromonad* septicemia (MAS) in fish. As has been reported, MAS has frequently caused huge economic losses in the cyprinid fish industry throughout China since 1989<sup>2</sup>. From the summer of 2009, MAS outbreaks have also occurred in the southeastern United States, resulting in industry-wide losses of channel catfish<sup>3,4</sup>. This disease has become an increasingly prominent problem for the rapid development of aquaculture industry.

The pathogenicity of *A. hydrophila* is usually considered to be multifactorial. Over the last 30 years, a number of virulence factors, including secretion systems, motility and adhesins, toxins, enzymes, quorum systems, iron acquisition and antibiotic resistance, have been identified<sup>1</sup>. To date, most of the research on *A. hydrophila* has still focused on the so-called virulence factors. However, recently, increasing reports<sup>5–7</sup> have proposed that animal environments that pathogens colonize have likely driven the evolution of new metabolic adaptations to maximize these new nutritional opportunities, and these adaptations may link with bacterial virulence. These suggest the known virulence factors may not be the only players in the bacterial infection process.

In this study, we present the first two complete genomes of epidemic *A. hydrophila* strains isolated in China, NJ-35<sup>8</sup> and J-1<sup>9</sup>, which were responsible for the MAS outbreaks in Jiangsu Province, in 2010 and 1989, respectively. To address the causes of the epidemic and highly virulent properties of *A. hydrophila*,

<sup>1</sup>College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, 210095, China. <sup>2</sup>Bioinformatics Center, Nanjing Agricultural University, Nanjing, 210095, China. <sup>3</sup>School of Biological Sciences, University of Hong Kong, Hong Kong SAR, 999077, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.J.L. (email: liuyongjie@njau.edu.cn) or F.C.L. (email: fcleung.njau@gmail.com)

Strain	NJ-35	J-1	ML09-119	AL09-71	pc104A	ATCC 7966 <sup>T</sup>
Accession number	CP006870	CP006883	NC_021290	CP007566	CP007576	NC_008570
Status	Complete	Complete	Complete	Complete	Complete	Complete
Date of isolation	2010	1989	2009	2009	2010	Undetermined
Country of isolation	China	China	USA	USA	USA	USA
Source	Diseased carp	Diseased carp	Diseased catfish	Channel catfish	Pond soil	Fishy milk
Genome size (bp)	5,279,644	5,000,814	5,024,500	5,023,861	5,023,829	4,744,448
G + C content (%)	60.51	60.90	60.80	60.80	60.80	61.50
No. of genes	4,716	4,462	4,577	4,492	4,493	4,284
No. of CDS	4,526	4,268	4,434	4,297	4,300	4,119
No. of rRNAs	31	31	31	31	31	31
No. of tRNAs	102	110	112	111	111	128

**Table 1.** General features of *A. hydrophila* genomes.

and to clarify the evolution relationships among the epidemic strains isolated from China and the United States, we performed comparative genomic analyses of five epidemic strains isolated from both countries, together with the environmental strain ATCC 7966<sup>T</sup>. Based on comparative genomic and functional analyses, we provide novel insights into the pathogenicity of epidemic *A. hydrophila* by identifying putative factors that may be involved in the bacterial pathogenicity, and we also determined that the ST251 clonal group of *A. hydrophila* is likely responsible for the ongoing MAS outbreaks in China and the southeastern United States.

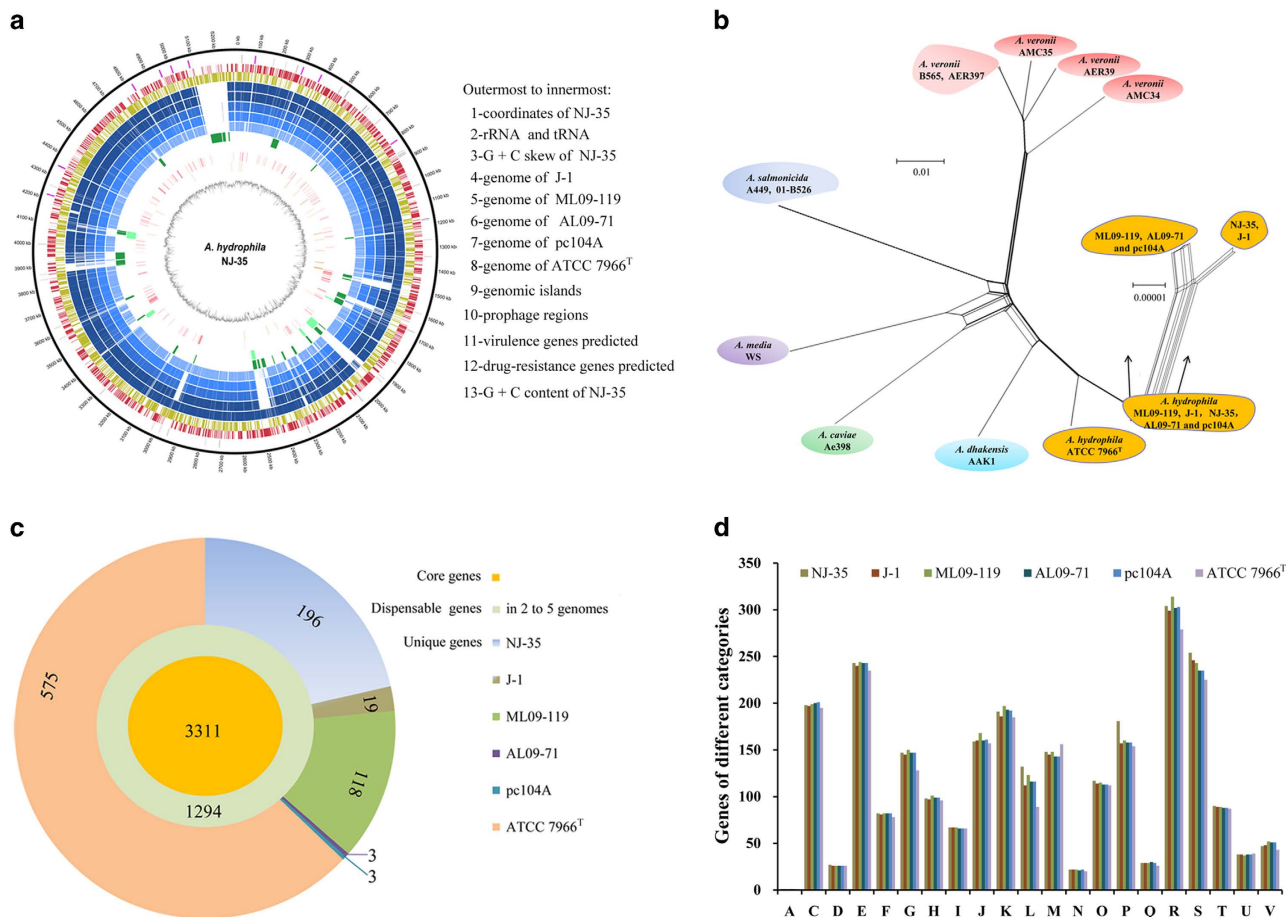
## Results

**General features of six *A. hydrophila* genomes.** The genomes of *A. hydrophila* strains NJ-35 and J-1 were sequenced and submitted to GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>). A whole-genome overview is shown in Table 1 and Fig. 1(a). Among the six strains studied, *A. hydrophila* NJ-35<sup>8</sup>, J-1<sup>9</sup>, ML09-119<sup>3,10</sup> and AL09-71<sup>4,11,12</sup> were epidemic piscine strains, whereas ATCC 7966<sup>T13</sup> was an environmental strain whose genome has been investigated extensively. Strain pc104A isolated from the soil of a catfish pond was from western Alabama, which has experienced an epidemic outbreak of MAS since 2009<sup>14</sup>. The genomes of six strains comprised a single circular chromosome which range in size from 4.74 Mb to 5.28 Mb. The G + C% content of each genome ranged from 60.51% to 61.50%, and the number of predicted CDS varied from 4119 to 4526. There were 10 rRNA operons encoding 31 rRNAs in each genome. Unlike rRNAs, the number of tRNAs in strain ATCC 7966<sup>T</sup> was obviously higher than in the other five strains, and the increasing number of tRNAs may be useful for increasing the rate of protein synthesis<sup>15</sup>.

**Virulence factors in *A. hydrophila*.** In previous studies, strains NJ-35<sup>8</sup>, J-1<sup>9</sup>, ML09-119<sup>3,10</sup> and AL09-71<sup>4,11,12</sup> were reported to be highly virulent strains, and pc104A<sup>14</sup> was a moderately virulent strain, whereas the virulence of the environmental strain ATCC 7966<sup>T</sup> was found to be significantly lower than those of the clinical isolates<sup>16</sup>. To determine which known virulence factors may contribute to the pathogenicity of the epidemic strains, we listed nearly all of the known virulence factors and queried them in six *A. hydrophila* genomes (Table 2). Surprisingly, genes encoding all the known virulence factors, with the exception of type III secretion system (T3SS), type VI secretion system (T6SS) and lateral flagella, were present in each *A. hydrophila* genome. T3SS and lateral flagella were absent not only in the environmental strain but also in the epidemic strains, whereas T6SS was only absent in the American strains ML09-119, AL09-71 and pc104A.

Moreover, to identify whether there were additional antibiotic resistance genes present in epidemic strains, each genome were annotated based on the Antibiotic Resistance Genes Database (ARDB, version 1.1; <http://ardb.cbcb.umd.edu/>), and 23 potential drug-resistance types were identified (Table S2). Although the six strains were isolated in different locations and at different times, antibiotic resistance genes were distributed similarly among these strains. The only exception is that a gene encoding aminoglycoside resistance has been not found in American strains ML09-119, AL09-71 and pc104A. While these *A. hydrophila* strains did not exhibit all drug-resistance phenotypes, the potential for antibiotic resistance in *A. hydrophila* merits increased attention.

Finally, to rule out the possibility that some virulence factors were not queried due to a careless omission, we also annotated the six *A. hydrophila* genomes using the Virulence Factors Database (VFDB, updated: 2012; <http://www.mgc.ac.cn/VFs/v3index.htm>). Intriguingly, the virulent strains NJ-35, J-1, ML09-119, AL09-71 and pc104A contained 193, 193, 191, 181 and 181 genes, respectively, which were less than ATCC 7966<sup>T</sup> (195 genes). However, most dispensable virulence genes were located in O-antigen



**Figure 1.** (a) Circular representation of six *A. hydrophila* genomes. The genomes were aligned with *dnaA* in the initial position and sequences moving clockwise. (b) Phylogenetic network of *A. hydrophila* and closely related *Aeromonas* species. Strains of the same species are displayed using identical colors. (c) Core genes and dispensable genes of six *A. hydrophila* genomes. (d) Comparison of COG categories among six *A. hydrophila* genomes. The ordinate axis indicates the number of genes in each COG functional category assigned to each of the six *A. hydrophila* genomes.

gene clusters and T6SS, and there were few findings except for an LPS-modification gene cluster specific to strain ATCC 7966<sup>T</sup> (Dataset S1).

The above analyses showed that the known virulence factors had similar distributions among the epidemic strains and the environmental strain. This finding raised the question of whether the six *A. hydrophila* strains really differed in virulence. Thus, we reassessed the virulence of these strains using zebrafish and ICR mice. Because strain ML09-119, AL09-71 and pc104A was absent in our laboratory, their virulence was not assessed. As shown in Fig. S1a, the LD<sub>50</sub>s in zebrafish of strains NJ-35, J-1 and ATCC 7966<sup>T</sup> were  $1.30 \times 10^2$ ,  $1.94 \times 10^3$ , and  $1.05 \times 10^6$  CFU/fish, respectively. According to our previous study<sup>8</sup>, NJ-35 and J-1 could be classified as virulent strains (LD<sub>50</sub> <  $1.0 \times 10^6$  CFU/fish), whereas ATCC 7966<sup>T</sup> is better classified as an avirulent strain (LD<sub>50</sub> >  $1.0 \times 10^6$  CFU/fish). Similarly, the virulence of strains NJ-35 and J-1 were also significantly greater than that of ATCC 7966<sup>T</sup> in ICR mice (Fig. S1b). Therefore, we speculated that there might be undiscovered virulence factors that play key roles in the pathogenicity of *A. hydrophila*.

**Whole-genome comparisons. Phylogenetic network analysis.** To determine the evolutionary relationships among the six *A. hydrophila* genomes, a phylogenetic network was constructed using 905 concatenated genes with a total length of 1.01 Mb. Ten genomes from other *Aeromonas* species were used as references. As shown in Fig. 1b, unlike strain ATCC 7966<sup>T</sup>, the other five virulent *A. hydrophila* strains fell into the same subclade. Further, the two Chinese epidemic strains (NJ-35 and J-1) clustered together, whereas the three American strains (ML09-119, AL09-71 and pc104A) formed a separate lineage. In addition, we calculated the average nucleotide identity (ANI) using JSpecies (version 1.2.1; <http://imedea.uib-csic.es/jspecies/about.html>). As shown in Table S3, the ANI values among the five virulent strains ranged between 99.81–100%, whereas the values between them and ATCC 7966<sup>T</sup> ranged between 96.74–96.79%. These results suggested that strain pc104A should belong to the clonal group of epidemic

Virulence function	Locus tag						Reference
	NJ-35	J-1	ML09-119	AL09-71	pc104A	ATCC 7966 <sup>T</sup>	
Secretion system							
T2SS <sup>a</sup>	02520–02525	02535–02540	20110–20105	20950–20945	20980–20975	3786–3785	24
T3SS	NF	NF	NF	NF	NF	NF	21
T6SS	13500–13610	13090–13200	NF	NF	NF	1848–1826	22
Motility and adhesion							
Polar flagellum <sup>b</sup>	07230–07305	07255–07330	15330–15255	15855–15780	15890–15815	2847–2832	25
Lateral flagella	NF	NF	NF	NF	NF	NF	32
Flp type IV pilus	15530–15590	15125–15185	08135–08075	08365–08305	08370–08310	1462–1450	33
Msh type IV pilus	21860–21950	21205–21295	01980–01890	02115–02025	02115–02025	0399–0383	34
Tap type IV pilus <sup>c</sup>	02085–02095	02100–02110	20515–20500	21385–21375	21415–21405	3871–3868	33
Type IV pilus	21135–21170	20480–20515	02695–02660	02840–02805	02840–02805	0526–0518	13
O-antigen gene clusters	06980–07085	07000–07110	15485–15595	16000–16095	16035–16130	2908–2877	10
Toxin							
cytotoxic enterotoxin, AerA	21570	20915	02265	02405	02405	0438	35
heat-stable cytotoxic enterotoxin, Ast	19690	19030	04100	04285	04285	0804	36
extracellular hemolysin, AHH1	15265	14860	08400	08630	08635	1512	37
hemolysin, HlyA	07430	07455	15145	15655	15690	2809	38
hemolysin III	03995	04025	18530	19245	19275	3493	13
thermostable hemolysin (TH)	05350	05370	17235	17890	17925	3217	39
RtxA	16300	15900	07370	07595	07600	1359	40
Enzyme							
AroA	09965	09595	13050	13505	13525	1979	41
DNA adenine methyltransferase, Dam	05505	05525	17080	17735	17770	3186	42
elastase, AhpB	19440	18780	04340	04535	04535	0851	43
enolase; Eno	19605	18945	04185	04370	04370	0821	44
extracellular protease, EprA1	08175	08205	14405	14915	14940	2713	45
glucose-inhibited division protein A, GidA	24435	23000	22245	23235	23270	4273	46
phospholipase A1, PLA	23360	22715	00550	00605	00605	0104	47
phospholipase C, PLC	20565	19910	03265	03410	03410	0635	47
exoribonuclease R, VacB	20220	19565	03605	03755	03755	0702	48
serine protease, SerA	08315	08330	14260	14760	14780	2687	49
ToxR-regulated lipoprotein, TagA	18525	17900	05230	05450	05455	0978	50
UDP N-acetylgalactosamine 4-epimerase, Gne	17585	17180	05955	06170	06170	1113	51
UDP-galactose-4-epimerase, GalE	00915	00920	21615	22575	22605	4103	51
UDP-glucose pyrophosphorylase, GalU	03965	03995	18560	19275	19305	3499	52
Quorum system							
AI-1 QS system	20975–20980	20320–20325	02850–02845	03000–02995	03000–02995	0557–0556	53
AI-2 QS system	20230	19575	03595	03745	03745	0700	54
QseBC QS system	05320–05325	05340–05345	17265–17260	17920–17915	17955–17950	3223–3222	55
Iron acquisition							
amonabactin synthesis and uptake	09855–09885	09485–09515	13155–13125	13615–13585	13635–13605	2479–2473	56
ferric uptake regulator, Fur	15170	14765	08495	08725	08730	1530	57
heme uptake	18575–18595	17950–17970	05180–05160	05400–05380	05405–05385	0968–0964	58
putative heme receptor	18555	17930	05200	05420	05425	0972	56
siderophore synthesis	05035–05040	05055–05060	17550–17545	18205–18200	18240–18235	3282–3281	13

Continued

Virulence function	Locus tag						Reference
	NJ-35	J-1	ML09-119	AL09-71	pc104A	ATCC 7966 <sup>T</sup>	
Antibiotic resistance							
β-lactamase, AmpC	05765	05785	16865	17475	17510	3135	59
metallo-β-lactamase, CphA	20015	19360	03800	03960	03960	0740	60

**Table 2.** Putative virulence factors of *A. hydrophila* genomes. <sup>a,b,c,d</sup>Only partial genes of these virulence factors were shown, and whole genes of them were shown in Table S1. NF represents “not found”. The beginning letters of locus tags (U876\_, V469\_, AHML\_, V428\_, V429\_ and AHA\_) are omitted.

*A. hydrophila* isolates which caused large-scale MAS outbreaks in the United States. The five epidemic strains were more similar to each other than to strain ATCC 7966<sup>T</sup>, indicating that some genetic traits distinctive to the environmental strains may appear in the epidemic strains.

**Core and dispensable genomes.** To find the essence and diversity of the six *A. hydrophila* strains, we analyzed their core and dispensable genomes. The pan-genome shared by the six strains consisted of 5519 genes, including 3311 core genes and 2208 dispensable genes consisting of 914 unique genes and another 1294 genes present in two to five strains (Fig. 1c). We found that ATCC 7966<sup>T</sup> has 575 unique genes, while epidemic strains NJ-35, J-1, ML09-119, AL09-71 and pc104A have 196, 19, 118, 3 and 3 unique genes, respectively. The smaller amount of unique genes in strains J-1, AL09-71 and pc104A occurs because most of their genes are contained by NJ-35 or ML09-119. The diversity among the six genomes provides genetic clues about the mechanisms contributing to bacterial virulence. However, relatively high proportions of dispensable genes, especially unique genes, are predicted to encode hypothetical proteins, which create challenges for further analysis.

**COG functional categories.** To identify which function of the epidemic strains may lead to the virulence diversity, we annotated the six *A. hydrophila* genomes using the protein database of Clusters of Orthologous Groups (COGs, updated: 2014; <http://www.ncbi.nlm.nih.gov/COG/>). As depicted in Fig. 1d, the gene numbers of the five epidemic strains categorized as COG G, L, P, R and S were relatively more than that of strain ATCC 7966<sup>T</sup>. This finding suggested that the genes responsible for various functions, such as carbohydrate metabolism and transport (COG G), replication, recombination and repair (COG L), inorganic ion transport and metabolism (COG P) and poorly characterized function (COG R and S), may contribute to the diversity in bacterial pathogenicity. To further analyze the genes causing the various gene numbers of COG G, L, P, R and S, we aligned these genes using strain NJ-35 as the reference (Dataset S2). Considering that COGs R and S were poorly characterized, we mainly focused on the genes within COGs G, L and P. Intriguingly, among the genes classified as COG G, some genes involved in the utilization of *myo*-inositol, sialic acid and L-fucose were found to be specific to all five epidemic strains. Among the COG L genes, plentiful genes encoding integrase and transposase were present in the five epidemic strains but absent in ATCC 7966<sup>T</sup>. Their presence suggested that more HGT (horizontal gene transfer) or transposition may have occurred in epidemic strains. Among the COG P genes, the accumulation of ABC-type transporters was an important factor affecting the number of genes categorized as COG P genes in strain NJ-35.

**Regions of genomic plasticity among the *A. hydrophila* strains.** To identify the pathogenicity-related regions, we identified 27 regions of genomic plasticity (RGPs). Of the 27 RGPs, 24 were predicted to be genomic islands using IslandViewer software (updated Sep 3, 2014; <http://pathogenomics.sfu.ca/islandviewer>). The epidemic strains NJ-35, J-1, ML09-119, AL09-71 and pc104A contained 18, 13, 17, 17 and 17 RGPs, respectively, which were significantly more than ATCC 7966<sup>T</sup> (6 RGPs). By analyzing the gene content of the 27 RGPs, we identified three O-antigen regions, 10 prophage regions and 15 other functional RGPs (Table 3, Dataset S3).

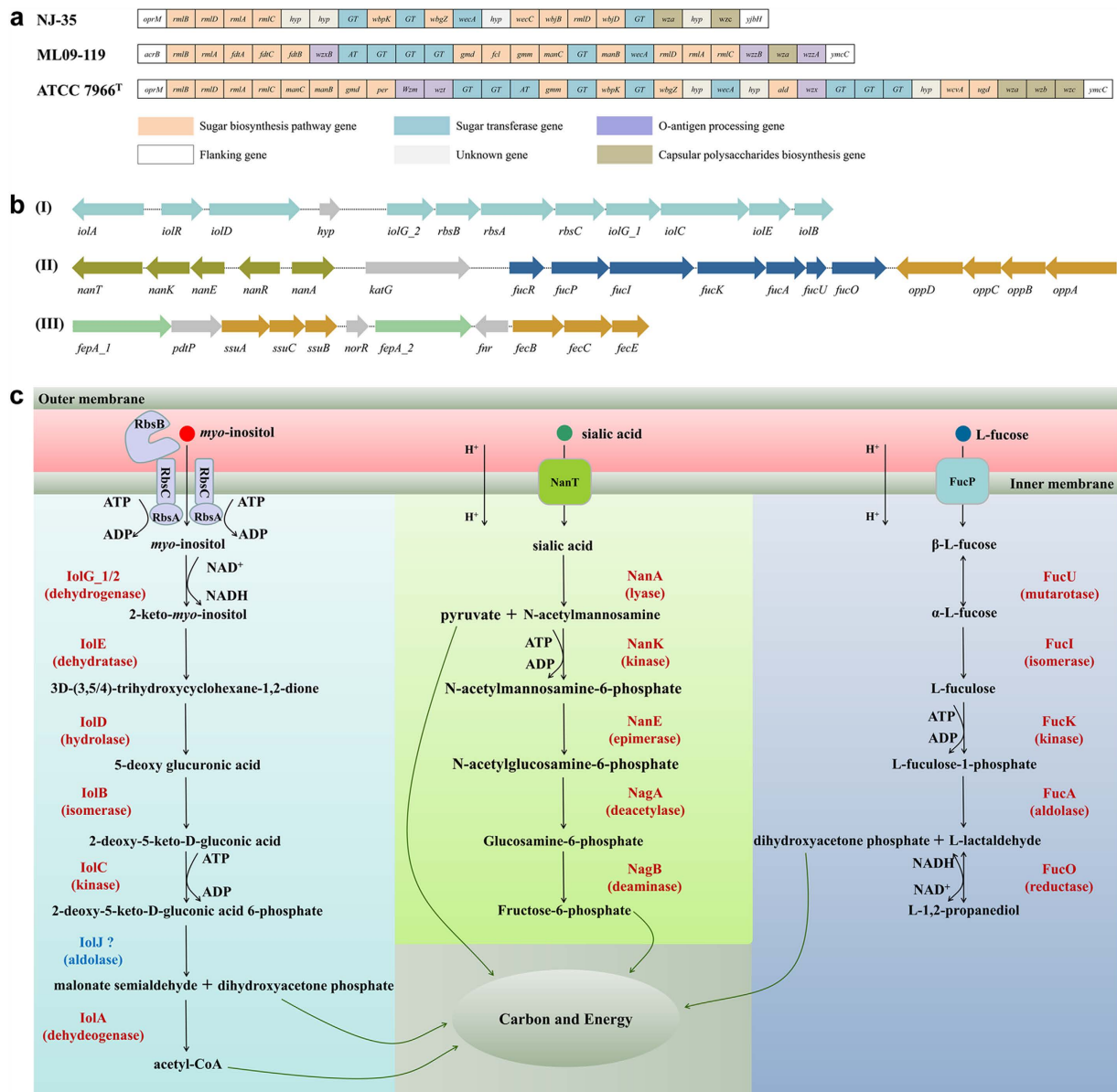
**(I)O-antigen regions.** RGP1, RGP21 and RGP26 carry diverse O-antigen gene clusters (Fig. 2a, Table 3, Table S4). The O-antigen gene clusters in strains NJ-35 and J-1 were identical and have not been observed previously in *A. hydrophila*, while strains ML09-119, AL09-71 and pc104A share the identical O-antigen gene clusters. All six O-antigen gene clusters contained the genes *rmlA*, *rmlB*, *rmlC* and *rmlD*, which encode proteins responsible for the biosynthesis of dTDP-L-rhamnose from glucose-1-phosphate. However, the organization and homology of the genes of the unique O-antigen gene cluster type varied substantially. In strains ML09-119, AL09-71, pc104A and ATCC 7966<sup>T</sup>, the *manB* and *manC* genes were located in O-antigen gene clusters, whereas *manA* was found elsewhere in the genomes. These three genes are required for the synthesis of GDP-mannose from fructose-6-phosphate<sup>10</sup>; however, they were not found in strains NJ-35 and J-1. In addition, *fdtA*, *fdtB* and *fdtC* which are required for the synthesis of the dTDP-sugar 3-acetamido-3,6-dideoxy-D-galactose<sup>10</sup>, were found to be specific to strains ML09-119, AL09-71 and pc104A.

RGP	Coordinate	Length (kbp)	Total CDSs	Genomic island	Trait	Strain					
						NJ-35	J-1	ML09-119	AL09-71	pc104A	ATCC7966 <sup>T</sup>
RGP1	U876_06980-07095	40.1	21	Yes	O-antigen						
RGP2	U876_07580-07700	34.2	18	Yes	Unknown						
RGP3	U876_07765-07910	32.8	24	Yes	Prophage-1; Metabolism						
RGP4	U876_08775-09165	59.8	70	Yes	Prophage-2; Prophage-3						
RGP5	U876_09485-09550	12.1	8	Yes	Prophage-4						
RGP6	U876_10025-10085	16.6	10	Yes	Unknown						
RGP7	U876_10380-10495	25.1	18	Yes	Uncharacterized						
RGP8	U876_11095-11220	18.6	21	Yes	Unknown						
RGP9	U876_11250-11595	61.1	67	Yes	Prophage-5						
RGP10	U876_13490-13610	35.3	24	Yes	T6SS						
RGP11	U876_14195-14250	12.7	11	Yes	Uncharacterized						
RGP12	U876_15635-15880	44.3	49	Yes	Prophage-6						
RGP13	U876_16480-16595	30.9	22	Yes	Metabolism						
RGP14	U876_17720-17915	36.0	34	Yes	Unknown						
RGP15	U876_17935-18165	59.6	40	Yes	Transport						
RGP16	U876_18670-18940	52.6	52	Yes	Prophage-7						
RGP17	U876_19120-19170	14.6	11	No	Transport						
RGP18	U876_23545-24335	154.0	140	Yes	Integrative plasmid						
RGP19	AHML_00260-00320	20.4	13	Yes	Uncharacterized						
RGP20	AHML_06930-07080	31.7	32	Yes	Unknown						
RGP21	AHML_15465-15595	30.4	27	Yes	O-antigen						
RGP22	AHML_15725-15885	30.7	33	Yes	<i>torCAD</i> operon						
RGP23	AHML_19460-19670	33.8	43	Yes	Prophage-8						
RGP24	AHA_1056-1087	44.3	30	Yes	Prophage-9						
RGP25	AHA_2018-2055	32.9	38	No	Prophage-10						
RGP26	AHA_2874-2908	40.8	35	Yes	O-antigen						
RGP27	AHA_4150-4188	38.9	39	No	LPS modification						

**Table 3.** Regions of genomic plasticity identified in *A. hydrophila* genomes. Alignment coverage [0–10%](#) [11–30%](#) [31–60%](#) [61–90%](#) [91–100%](#).

(II) *Prophage regions.* A total of eight prophage regions (prophage–1 to prophage–8) were identified by PHAST (updated Sep. 30, 2014; <http://phast.wishartlab.com>); their features are listed in Table S5, S6 and Fig. S2. Prophage–9 and prophage–10 were found in strain ATCC 7966<sup>T</sup> in previous studies<sup>13,16</sup>, although they were not identified by PHAST. Among the 10 known prophages, prophage–3, –6 and –8 were predicted to be intact prophages. Interestingly, all of the prophages identified in the five epidemic strains were absent in strain ATCC 7966<sup>T</sup>, and prophages–1, –4, –6 and –7 were conserved in all five epidemic strains. Prophages are thought to be important for bacteria because they often confer fitness and/or virulence factors. However, our analyses showed that none of the known virulence factors are present in the prophage regions (Fig. 1a). Nevertheless, the vast number of hypothetical proteins contained in the prophage regions may play roles in the evolution and virulence of *A. hydrophila*, and they therefore merit further functional characterization.

(III) *Other RGPs.* In addition to the O-antigen and prophage regions, 15 other RGPs were analyzed (Table 3). RGP2, RGP6, RGP8, RGP14, and RGP20 were classified as unknown RGPs because most of the proteins encoded by these regions were predicted to be hypothetical proteins and their functions were poorly understood. RGP7, RGP11, and RGP19 were classified as uncharacterized RGPs because the functions of some of their genes had been predicted but no representative gene clusters had been identified. The remaining seven RGPs were relatively well characterized or contained representative gene clusters. RGP10 consisted of a T6SS gene cluster that was absent in the three American strains. RGP13 contained two gene clusters encoding sialic acid and L-fucose utilization pathways. RGP15 contained two ABC-type transport systems, SsuACB and FecBCE. RGP17 also contained an ABC-type transport



**Figure 2.** (a) Genetic organization of the O-antigen gene clusters of *A. hydrophila* strains. *GT* represents the glycosyltransferase gene, *AT* represents the acetyltransferase gene, *hyp* represents the gene encoding the hypothetical protein. (b) Genetic organization of partial gene clusters of RGP3, RGP13 and RGP15. (I) depicts the *myo*-inositol utilization gene clusters; (II) depicts the sialic acid and L-fucose utilization gene clusters and *oppABCD* gene clusters; and (III) depicts the *ssuABC* and *fecBCE* gene clusters. *Hyp* represents the gene encoding the hypothetical protein. (c) Proposed utilization pathways for *myo*-inositol, sialic acid and L-fucose in *A. hydrophila*. The three proposed utilization pathways were modified according to previous studies<sup>18,19,26</sup> and the genome annotation in KEGG pathway database. *IolJ* was not identified in *A. hydrophila*.

system, which is involved in the transport of DL-methionine. RGP18 specific to strain NJ-35 was an integrative plasmid containing 140 genes. This region was found to contain numerous genes involved in DNA replication, transcription and repair. However, no virulence factors were observed in this integrative plasmid. RGP22 specific to American epidemic strains contained the *torCAD* operon encoding a trimethylamine N-oxide (TMAO) reductase respiratory system<sup>17</sup>. RGP27 specific to ATCC 7966<sup>T</sup> consisted of abundant genes involved in LPS modification. In addition, a *myo*-inositol utilization pathway was observed adjacent to prophage-1 in RGP3.

(IV) *RGPs specific to epidemic strains.* To elucidate the key determinants of pathogenicity of *A. hydrophila*, we focused on the RGPs that were specific to the five epidemic strains. Among these RGPs, only

RGP3, RGP13, and RGP15 were characterized and therefore further analysed using strain NJ-35 as the reference (see more details in Table S7).

The gene clusters in RGP3 encoded 11 proteins that were associated with the *myo*-inositol utilization pathway (Fig. 2b). As shown in Fig. 2c, once *myo*-inositol would be transported into the cytoplasm, *myo*-inositol 2-dehydrogenase (IolG\_1/IolG\_2) would initiate the degradation of *myo*-inositol, after which IolEIBCJA would be required for the degradation step. The net result of the inositol metabolic pathway in *A. hydrophila*, thus, would be the conversion of *myo*-inositol to a mixture of dihydroxyacetone phosphate and acetyl-CoA. Dihydroxyacetone phosphate can enter the glycolysis metabolic pathway, while acetyl-CoA conveys carbon atoms within the acetyl group to the citric acid cycle, to be oxidized for energy production. Similarly to the findings of a previous study<sup>10</sup>, IolJ (2-deoxy-5-keto-D-gluconic acid 6-phosphate aldolase) was not identified in each *A. hydrophila* genome. However, other aldolases, such as fructose/tagatose biphosphate aldolase (U876\_19685), which shares similar amino acid sequences with IolJ, could potentially be a substitute for IolJ.

RGP13 contained gene clusters encoding two metabolic pathways in which sialic acid and L-fucose are used (Fig. 2b). The sialic acid utilization pathway was encoded by the *nanTKERA* operon and other two genes—*nagA* and *nagB*—which are located outside of RGP13 (Fig. 2c). Briefly, sialic acid from the surrounding environment is transported into the cytoplasm by a symporter (NanT) through the inner membrane. A pyruvate group is first removed from the sialic acid with the help of N-acetylneuraminidase (NanA), after which N-acetylmannosamine is catalyzed by NanK, NanE, NagA and NagB, thereby generating fructose-6-phosphate, which may then enter the glycolysis metabolic pathway<sup>18</sup>. The pyruvate produced in this process is at a crossroads of central metabolic pathways, and it can yield acetyl-CoA to provide energy under aerobic conditions. The L-fucose utilization pathway was encoded by the *fucRPIKAUO* operon in the same orientation (Fig. 2b). L-fucose is transported into cytoplasm by a symporter (FucP); L-fucose isomerase (FucI) then converts  $\alpha$ -L-fucose to L-fucose with the help of fucose mutarotase (FucU), which can catalyze natural mutarotation between  $\alpha$ -L-fucose and  $\beta$ -L-fucose. Ultimately, L-fucose can be degraded into dihydroxyacetone phosphate, which lies in the glycolysis metabolic pathway, and L-1,2-propanediol, which is often secreted from cells<sup>19</sup> (Fig. 2c). Furthermore, a proposed ABC-type dipeptide/oligopeptide/nickel transport systems (OppABCD) was identified (Fig. 2b).

In RGP15, another two ABC-type transport systems were also identified (Fig. 2b), one of which was encoded by *ssuACB* and predicted to transport nitrate/sulfonate/bicarbonate and the other of which was encoded by *fecBCE* and predicted to transport  $\text{Fe}^{3+}$ -siderophores. In addition, two TonB-dependent receptors (FecA\_1 and FecA\_2) were identified near the transport systems. These ABC-type transport systems may contribute to the fitness of *A. hydrophila* by acquiring essential ions or nutrients efficiently in a nutrient-limiting host environment.

**MLST analysis and the distribution of putative virulence factors.** To determine whether there exists a link among ST, virulence factors and bacterial virulence, we performed the multilocus sequence typing (MLST) of 30 *A. hydrophila* strains, including the six *A. hydrophila* strains analyzed and another 24 strains with known virulence, as determined in our previous study<sup>8</sup> and in this study (Fig. S3). A total of 13 STs were identified in the 30 strains, and 17 virulent strains including the five epidemic strains analyzed all belonged to ST251 (Fig. 3). ST251 and ST328 belonged to the same clonal group (ST251 clonal group) because they shared five alleles out among the six housekeeping genes<sup>20</sup> (*gyrB*, *groL*, *gltA*, *metG*, *ppsA* and *recA*). We also constructed a phylogenetic tree based on the six housekeeping genes. As shown in Fig. 3, the strains analyzed in the phylogenetic tree could be classified into clade I (ST251 and ST328), clade II (ST322, ST327, ST355 and ST360) and clade III (ST1, ST321, ST323, ST324, ST325, ST326 and ST359). Intriguingly, a high correlation between the genetic phylogeny and  $\text{LD}_{50}$ s of *A. hydrophila* strains in zebrafish was observed. All strains placed in clades I could be classified as virulent strains ( $\text{LD}_{50} < 1.0 \times 10^6 \text{CFU/fish}$ ), whereas strains placed in clade III were all considered avirulent strains ( $\text{LD}_{50} > 1.0 \times 10^6 \text{CFU/fish}$ ).

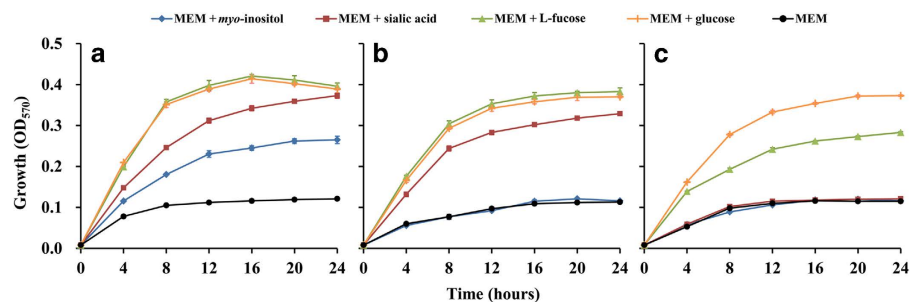
The most important finding was the distribution of virulent factors among *A. hydrophila* strains. With respect to the secretion systems, type II secretion system (T2SS) was observed in all 30 strains, T6SS was observed in 27 strains (all strains except the three American epidemic strains), and T3SS was not observed in any strain. A polar flagellum was present in each strain, whereas lateral flagella were absent in all strains. Among the detected genes encoding toxins and enzymes, only *aerA* was absent in all four avirulent strains; the other genes were all present in both virulent and avirulent strains. Intriguingly, the three metabolic pathways appeared only in strains belonging to the ST251 clonal group, ST322 and ST327, and all of these strains were virulent. Moreover, the ATCC 7966<sup>T</sup>, CH-14-1, XH14-1, ML14-14 and CS-34 strains, which lack the three metabolic pathways found in the virulent strains, were avirulent strains, even though they contained *aerA*. These findings suggested that the utilization pathways for *myo*-inositol, sialic acid, and L-fucose may be linked with the full virulence of epidemic *A. hydrophila*.

**The utilization of *myo*-inositol, sialic acid and L-fucose.** All strains containing the *myo*-inositol utilization, sialic acid and L-fucose utilization pathways could utilize the corresponding chemicals as carbon sources for growth. However, the level of bacterial growth on the different chemicals was diverse. As shown in Fig. 4, the growth of strains NJ-35, NJ-37 and ATCC 7966<sup>T</sup> was examined in comparison with bacterial growth on glucose. The growth of strain NJ-35 on L-fucose was similar to that on glucose but





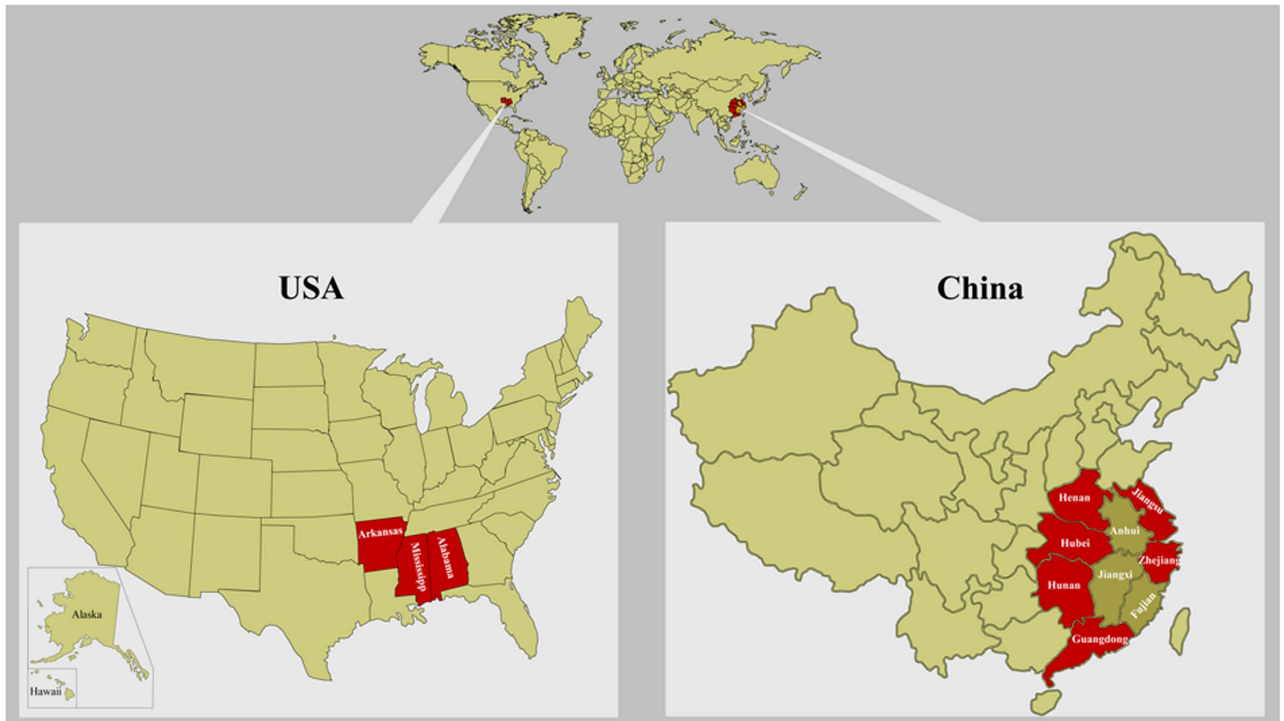
**Figure 3.** MLST analysis and the distribution of putative virulence factors in *A. hydrophila*. <sup>a,b</sup>P-fla and L-fla represent the polar flagellum and lateral flagella, respectively; <sup>c,d,e</sup>Ino, Sia and Fuc represent the utilization pathways for *myo*-inositol, sialic acid and L-fucose, respectively.



**Figure 4.** Growth of *A. hydrophila* strains in MEM supplemented with different chemicals. The growth of strains NJ-35 (a), NJ-37 (b) and ATCC 7966<sup>T</sup> (c) is depicted.

was faster than the growth on *myo*-inositol and sialic acid (Fig. 4a). This phenomenon was also observed in the other tested strains, suggesting that *A. hydrophila* could utilize L-fucose to grow more efficiently. As depicted in Fig. 4b, strain NJ-37, which lacks the *myo*-inositol utilization pathway, could not utilize *myo*-inositol to grow, but it could grow using sialic acid and L-fucose as carbon sources. Surprisingly, strain ATCC 7966<sup>T</sup>, without the predicted L-fucose utilization pathway, also had the ability to grow on L-fucose (Fig. 4c). This phenomenon was also observed in strain NJ-3, which also lacks the L-fucose utilization pathway. However, both ATCC 7966<sup>T</sup> and NJ-3 grew much more slowly on L-fucose than on glucose. This result indicated that the ATCC 7966<sup>T</sup> and NJ-3 strains may have an alternative metabolic pathway capable of utilizing L-fucose and that this pathway cannot utilize L-fucose as effectively as the classical L-fucose utilization pathway.

**Geographical distribution of the *A. hydrophila* ST251 clonal group.** Since all the epidemic strains analyzed were belong to ST251 clonal group, we investigated the geographical distribution of this clonal group through querying related sequences on the *Aeromonas* spp. MLST website (<http://pubmlst.org/Aeromonas>) and in GenBank. As summarized in Fig. 5 and Table S8, the ST251 clonal group of *A. hydrophila* has been found in the Jiangsu, Henan, Hubei, Hunan, Guangdong and Zhejiang provinces



**Figure 5.** Geographical distribution of the ST251 clonal group *A. hydrophila*. The regions filled in with red represent the distribution of the ST251 clonal group. This map was modified based on the maps obtained from PowerPoint Toolkit (<http://ppt-toolkit.com/>).

of China and in Alabama, Mississippi and Arkansas in the United States. Unfortunately, the nucleotide sequences of the six housekeeping genes of most of the isolated *A. hydrophila* strains have not been analyzed, which makes it difficult to evaluate the distribution of the ST251 clonal group satisfactorily. Considering most of the provinces in southern China including Anhui, Jiangxi and Fujian all experienced epidemic outbreaks of MAS between 1989 and 1991<sup>2</sup>, it is likely that the ST251 clonal group can also be found in these districts.

## Discussion

In this study, we try to clarify the factors that give *A. hydrophila* its epidemic and highly virulent nature. *A. hydrophila* possesses a variety of virulence genes, and some of them, such as the genes in T3SS and lateral flagella, are recognized as virulence markers<sup>13</sup>. However, the known virulence factors distributed similarly among the virulent strains and the avirulent strains examined in this study, which were queried about the roles of these factors in the pathogenicity of the epidemic strains.

Proteins in *A. hydrophila* are thought to be transported across the two membranes by either a one-step process (T3SS and T6SS) or a two-step mechanism (T2SS)<sup>1</sup>. T3SS has been demonstrated to be essential to the pathogenicity of *A. hydrophila* strains AH3<sup>21</sup>. Contrary to expectations, T3SS was not observed in any of the strains used in our study. Therefore, we speculated that T3SS may play important roles in T3SS-positive strains; however, this process is not essential to the pathogenicity of all *A. hydrophila* strains. For this reason, careful attention should be paid when T3SS is used as an indicator of virulence. T6SS was characterized as a virulence factor in *A. hydrophila* SSU<sup>22</sup>, however, T6SS was found in 27 out of 30 strains analyzed but was surprisingly absent in the American epidemic strains ML09-119, AL09-71 and pc104A. Recently, an increasing number of studies have proposed that although T6SS can, in rare instances, be linked to host interactions and virulence, the broader physiological importance of this system is likely to provide a defense against simple eukaryotic cells and other bacteria in the environment<sup>23</sup>. The functions of T6SS have been investigated extensively in relation to the pathogenicity of *A. hydrophila*; however, it would also be useful to examine its role in microbial communities, as well. Unlike T3SS and T6SS, T2SS was conserved in all *A. hydrophila* strains used in this study. In *A. hydrophila*, T2SS is essential for the translocation of multiple proteins, including aerolysin, phospholipase, proteases and DNase, from the periplasm into the extracellular milieu<sup>24</sup>. In general, these proteins are associated with the destruction of various tissues, thus contributing to cell damage or disease. In view of the fact that the epidemic *A. hydrophila* strains examined in this study lacked both T3SS and T6SS but were still highly virulent, we hypothesized that T2SS is both necessary and sufficient for the pathogenicity of these strains.

It has been proposed that with the possible exception of the motor genes, the polar flagellum and lateral flagella rely on entirely distinct sets of genes<sup>25</sup>. The appearance of the polar flagellum in all 30 strains examined suggests that the polar flagellum is conserved in *A. hydrophila*. In contrast, our finding suggests that lateral flagella may not be species-specific but may instead be strain-specific, consistent with the results of a previous study<sup>13</sup>. In addition, we also detected a number of genes encoding toxins and enzymes; however, six such genes, but not *aerA*, were present in each of the 30 strains examined. Therefore, although these toxins and enzymes may be essential for the full virulence of *A. hydrophila*, it would not be reliable to determine the bacterial virulence based solely on the presence of several toxins or enzymes.

Because the known virulence factors are not primarily responsible for the substantial virulence of the epidemic ST251 clonal group, we searched for new factors may involve in their pathogenicity. The result of an analysis based on COG functional categories indicated that additional genes involved in functions such as carbohydrate metabolism and transport were present in the epidemic strains. Thus, we identified three metabolic pathways, utilizing *myo*-inositol, sialic acid and L-fucose, and we determined that *A. hydrophila* strains containing specific metabolic pathways could utilize chemicals metabolized in these pathways to grow. Except for the *myo*-inositol utilization pathway, the other two were first found in *A. hydrophila*. Metabolic versatility under nutrient constraints is a key determinant for bacterial survival and proliferation in different niches and compartments. In pathogenic *Salmonella*, *Vibrio* and *Helicobacter*, it has been shown that metabolic genes are as important for a successful infection as are the classic virulence genes<sup>6</sup>. *myo*-Inositol is an essential component of various metabolites, such as phosphatidylinositol-based phospholipids, that are abundant in animal and plant cells<sup>5</sup>. The metabolism of inositol compounds is not only linked to successful plant-bacterium interactions in members of the *Rhizobiaceae* but also to the survival of the animal pathogen *Brucella* inside its host<sup>26</sup>. *In vivo* screening approaches have indicated that the *myo*-inositol utilization pathway plays roles in *Salmonella typhimurium* virulence in mice, pigs, chicken and calves<sup>5</sup>. Sialic acids constitute a family of related sugar moieties that are found throughout the body as components of glycoproteins, gangliosides, and other sialoglycoconjugates<sup>5</sup>. It has been demonstrated that *Escherichia coli*, *Clostridium perfringens* and *Bacteroides fragilis* are able to utilize sialic acid as a carbon and energy source via the sialic acid utilization pathway<sup>5</sup>. In addition, mutants of both *S. typhimurium* and *Clostridium difficile* that are unable to catabolize sialic acid exhibit impaired expansion<sup>27</sup>. L-fucose is one component of the carbohydrate moiety of the mucosal glycoconjugates, and it is also used by a variety of microorganisms as a carbon and energy source<sup>28</sup>. An earlier study demonstrated that L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage and that strains lacking the capability to use fucose are attenuated *in vivo*<sup>19</sup>. In addition, the genomes of *Salmonella enterica*, pathogenic *E. coli* and *C. perfringens* are also equipped with an L-fucose utilization pathway<sup>7</sup>.

Genetic links between the new metabolic capacities and virulence factors indicate that metabolic pathways are acquired as part of the pathogen evolution toward colonizing new niches with new food sources<sup>6</sup>. In our study, three metabolic pathways, utilizing *myo*-inositol, sialic acid and L-fucose, were identified in RGP's predicted as genomic island. Therefore, it seemed that the genetic determinants encoding metabolic functions were acquired by HGT to expand the metabolite target spectrum of epidemic *A. hydrophila*. Furthermore, our results showed that the *A. hydrophila* strains equipped with the three metabolic pathways were all considered virulent strains and had the ability to exploit the corresponding chemicals. These data suggest that the metabolic pathways presented in epidemic *A. hydrophila* may help the bacteria to overcome nutritional limitations *in vivo*, thereby increasing their fitness during infection and being the key factors causing their epidemic. Although interesting results are emerging, thus far they can only be considered preliminary. We can not exclude the possibility that the large numbers of hypothetical proteins present in the epidemic strains may also be responsible for the bacterial virulence. To demonstrate that the genes of these metabolic pathways indeed play a role in virulence, it is necessary to construct deletion mutants for these functions and test their virulence. Further study of the deletion mutants will be the fucous in our future work.

Considering the geographical distribution and biological resources of the ST251 clonal group, we speculated this clonal group is likely to be responsible for the outbreaks of *A. hydrophila* in China and the United States. Based on phylogenetic network and ANI analyses, we found that all the five epidemic ST251 strains were highly similar to each other. Moreover, the fact that 10 RGP's were common to the five epidemic strains also corroborates their similarity. These findings suggest that five epidemic strains analyzed may share a common ancestor. The *A. hydrophila* strain J-1 was isolated in Jiangsu Province, China, in 1989 and was used as a vaccine strain in China. Additionally, strain NJ-35 was isolated in Jiangsu Province in 2010, and this strain contains almost all of the gene content of strain J-1. Therefore, it is reasonable to speculate that the NJ-35 strain may have evolved from the J-1 strain. Hossain *et al.*<sup>3</sup> proposed that ST251 strains in the United States may have been introduced from China, as *A. hydrophila* ST251 strains have been identified in China and their occurrence documented since the late 1980s but this clonal group was not observed in the United States until 2004. Although we tend to agree with this opinion, more evidence is needed to support this speculation. In particular, the complete genome information of the ST251 clonal group strains isolated from both countries will likely shed light on this issue. In China, ST251 strains have primarily been isolated from carp, whereas in the United States, they have been isolated from channel catfish. This finding could be attributed to differences in the species of

farmed fish in these two countries. Of note, we showed that strains of the ST251 clonal group were also highly virulent in zebrafish, which indicated that these strains may be virulent in diverse fish species. Equally important is that *A. hydrophila* ST251 strains were still being isolated from diseased fish in our most recent survey, suggesting that ST251 is a high-risk type in fish farms. Furthermore, because of the globalization of the aquaculture and ornamental fish trade industries, the ST251 clonal group could be introduced into non-epidemic districts. Therefore, more work is urgently needed to prevent further epidemic outbreaks of ST251 *A. hydrophila* infection.

In conclusion, this report reevaluated the importance of known virulence factors, identified metabolic strategies developed by the epidemic strains, and thereby provides a comprehensive understanding of the pathogenicity of *A. hydrophila*. These findings suggest that links between metabolic adaptations of *A. hydrophila* and their ability to become epidemic and virulent should definitely receive more attention. Furthermore, the *A. hydrophila* strains of the ST251 clonal group are likely to be responsible for the MAS outbreaks in both China and the southeastern United States.

**Experimental procedures.** Ethics statement. All animal experiments were carried out according to animal welfare standards and were approved by the Ethical Committee for Animal Experiments of Nanjing Agricultural University, China.

**Strains and genomes.** Six *A. hydrophila* strains NJ-35, J-1, ML09-119, AL09-71, pc104A and ATCC 7966<sup>T</sup> were used in this study to perform the comparative genomic analyses (Table 1). The ten genomes (*Aeromonas veronii* B565, AER39, AER397, AMC34 and AMC35; *Aeromonas salmonicida* A449 and 01-B526; *Aeromonas dhakensis* AAK1; *Aeromonas media* WS; *Aeromonas caviae* Ae398) used to perform the phylogenetic network analysis were deposited in GenBank with accession numbers NC\_015424, NZ\_AGWT00000000, NZ\_AGWV00000000, NZ\_AGWU00000000, NZ\_AGWW00000000, NC\_009348, NZ\_AGVO00000000, NZ\_BAFL00000000, NZ\_ALJZ00000000 and NZ\_CACP00000000, respectively. The 24 *A. hydrophila* strains BSK-10, CS-34, CS-43, JH-17, JH-19, NJ-1, NJ-28, NJ-3, NJ-34, NJ-37, XX-14, XX-22, XX-49, XX-52, XX-58, XX-62, CH14-1, DW14-1, JD14-1, ML14-9, ML14-14, SQ13-11, XH14-1 and XS14-1 were used to examine the distribution of putative virulence genes and to perform MLST analysis. The sequences of housekeeping genes (*gyrB*, *groL*, *gltA*, *metG*, *ppsA* and *recA*) of the 24 strains were deposited in GenBank with accession numbers KP162270 to KP162274, KM235114 to KM235188, and KP781915 to KP781962.

**Sequencing, assembly and annotation.** The genomes of NJ-35 and J-1 were sequenced using 454 GS Junior pyrosequencing with 23-fold and 29-fold coverage, respectively. One shotgun run and one 8 kb-library span paired-end run were conducted, the raw data were assembled by Newbler (version 2.7; Roche/454 Life Science) with default parameters. In addition, gap-filling primers were designed and the Sanger sequencing results were compiled with the contigs into the complete genomic sequence via SeqMan (version 7.1.0; Larsegene). Then, the complete genomes of strains NJ-35 and J-1 were obtained and annotated using the Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (version 2.4; [http://www.ncbi.nlm.nih.gov/genome/annotation\\_prok/](http://www.ncbi.nlm.nih.gov/genome/annotation_prok/)).

**Virulence factors in *A. hydrophila*.** To summarize the known virulence genes of *A. hydrophila*, we read almost all of the articles discussing the pathogenicity of *A. hydrophila* that had been published between 1979 and 2014. The summarized virulence genes were then queried using BLASTp implemented in BLAST+ (version 2.2.29; <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/>) against the six *A. hydrophila* genomes. A given virulence gene was deemed to be present in the genome once a gene shared more than 80% identity and covered at least 90% with it. To annotate the antibiotic genes and virulence genes of *A. hydrophila*, BLASTp was applied to align the amino acid sequences against ARDB and VFDB, respectively. Amino acid sequences with alignment lengths over 90% of their own lengths and with over 40% match identity were chosen, and the description of the best hit was assigned as the annotation of the predicted gene.

**Pathogenicity assays.** The LD<sub>50</sub>s of *A. hydrophila* with zebrafish was performed according to our previous study<sup>8</sup>. For each *A. hydrophila* strain, eight groups of ten fish were intraperitoneally injected with 0.02 mL of serially tenfold-diluted bacterial suspensions containing 10<sup>1</sup>–10<sup>8</sup> CFU. Members of the control group were injected intraperitoneally with 0.02 mL sterile PBS only. The zebrafish were observed until one week post-infection. For each determination, three separate experiments were performed. Survival data were analyzed according to a method devised by Reed and Muench<sup>29</sup> for the calculation of LD<sub>50</sub> values. ICR mice were used as a mammalian model to determine the virulence of *A. hydrophila*. Briefly, for each *A. hydrophila* strain, groups of ten female ICR mice were injected intraperitoneally with doses of 5.0 × 10<sup>7</sup> CFU and 1.0 × 10<sup>8</sup> CFU. Deaths were recorded daily for 14 days post-infection.

**Whole-genome comparisons.** The phylogenetic network analysis was performed using 905 concatenated genes from the 16 *Aeromonas* genomes through the NeighborNet function in SplitsTree4 (version 4.13.1; <http://www.splitstree.org/>) on a set of 1000 bootstrap replicates to investigate the support for different branches in the network. To identify the core genome and the dispensable genome of *A.*

*hydrophila*, preliminary analyses using the reciprocal BLASTp result file filtered with parameters of 90% minimum alignment length and 20–90% match identity were run. Then, a 90% minimum alignment length and an 80% minimum match identity were used to define orthology between two CDSs<sup>30</sup>. Finally, the core genome and dispensable genome among *Aeromonas* strains were calculated using in-house Perl scripts based on these criteria. It should be noted that the high-identity-sharing CDSs belonged to one genome and could be defined as single genes based on these criteria. To classify the functions encoded in the six *A. hydrophila* genomes, BLASTp was used to align the amino acid sequences against the COGs database. Amino acid sequences with alignment lengths of over 90% of their own length and over 40% match identity were chosen, and the description of the best hit was assigned as the annotation of the predicted genes. All annotated genes were then classified based on their COG classes.

**The identification of regions of genomic plasticity.** RGP were defined as regions that were longer than 10 kb or that contained at least four contiguous genes but that were absent at least in one genome analyzed<sup>31</sup>. The 27 RGPs used as the references were then queried using BLASTn implemented in BLAST+ against the six *A. hydrophila* genomes, the alignment coverage was then summarized in Table 3 when the sequences shared more than 80% identity. In addition, the *A. hydrophila* genomes were submitted to the KEGG pathway database (updated Oct. 2, 2014; <http://www.genome.jp/kegg/pathway.html>) to analyze the metabolic pathways.

**MLST analysis and the distribution of putative virulence factors.** MLST was performed as previously described<sup>20</sup>. Briefly, six housekeeping genes, *gyrB*, *groL*, *gltA*, *metG*, *ppsA* and *recA*, from each strain were submitted to the *Aeromonas* spp. MLST website to determine their identity against existing alleles. Each gene fragment was translated into a distinct allele, and each strain was classified into its sequence type by the combination of the alleles of the six housekeeping genes. New alleles that differed from the pre-existing alleles were assigned a new allele designation. For the phylogenetic analysis, six genes were aligned and concatenated using MEGA6 (version 6.0; <http://www.megasoftware.net/>). Then, the phylogenetic tree was constructed using the neighbor-joining method on a set of 1000 bootstrap replicates. Each of the virulence factors was detected by one to four primer pairs. All primers used in this study are listed in Table S9.

**Growth on myo-inositol, sialic acid and L-fucose.** MEM medium (Yocan, CM0201) containing glutamine, but not phenol red, *myo*-inositol and glucose, was used as the carbon-restricted medium in this study. All chemicals were acquired from Sigma. Carbon sources such as *myo*-inositol, sialic acid, L-fucose and glucose were added to MEM to a final concentration of 0.1%. MEM containing glucose was used as a reference. To test the ability of *A. hydrophila* strains to utilize different carbon sources, overnight cultures grown in Luria broth (LB) were washed twice in MEM and inoculated into MEM containing different carbon sources at an OD<sub>600</sub> of 0.01. Aliquots of these cultures (200 μL) were injected into 96-well plates, as were blanks containing only MEM. Each growth condition was replicated in 4 wells. Cultures were grown on a shaking incubator (80 rpm) at 28 °C, and the growth of each strain was detected by measuring the absorbance of the cultures at 570 nm using the Bio-Rad Model 680 microplate reader. Growth experiments for each strain were repeated on four separate occasions.

## References

1. Tomas, J. M. The main *Aeromonas* pathogenic factors. *ISRN Microbiol.* **2012**, 256261 (2012).
2. Zhang, X., Yang, W., Wu, H., Gong, X. & Li, A. Multilocus sequence typing revealed a clonal lineage of *Aeromonas hydrophila* caused motile *Aeromonas* septicemia outbreaks in pond-cultured cyprinid fish in an epidemic area in central China. *Aquaculture* **432**, 1–6 (2014).
3. Hossain, M. J. *et al.* An Asian origin of virulent *Aeromonas hydrophila* responsible for disease epidemics in United States-farmed catfish. *mBio.* **5**, e00848-14 (2014).
4. Pridgeon, J. W. & Klesius, P. H. Molecular identification and virulence of three *Aeromonas hydrophila* isolates cultured from infected channel catfish during a disease outbreak in west Alabama (USA) in 2009. *Dis. Aquat. Organ* **94**, 249–53 (2011).
5. Reynolds, T. B. Strategies for acquiring the phospholipid metabolite inositol in pathogenic bacteria, fungi and protozoa: making it and taking it. *Microbiology* **155**, 1386–96 (2009).
6. Rohmer, L., Hocquet, D. & Miller, S. I. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol.* **19**, 341–8 (2011).
7. Staib, L. & Fuchs, T. M. From food to cell: nutrient exploitation strategies of enteropathogens. *Microbiology* **160**, 1020–39 (2014).
8. Pang, M. D. *et al.* *Tetrahymena*: an alternative model host for evaluating virulence of *Aeromonas* strains. *PLoS One* **7**, e48922 (2012).
9. Chen, H. Q. & Lu, C. P. Study on the pathogen of epidemic septicemia occurred in cultured cyprinoid fishes in southern China. *Journal of Nanjing Agricultural University* **14**, 87–91 (1991).
10. Hossain, M. J. *et al.* Implication of lateral genetic transfer in the emergence of *Aeromonas hydrophila* isolates of epidemic outbreaks in channel catfish. *PLoS One* **8**, e80943 (2013).
11. Pridgeon, J. W. & Klesius, P. H. Virulence of *Aeromonas hydrophila* to channel catfish *Ictalurus punctatus* fingerlings in the presence and absence of bacterial extracellular products. *Dis. Aquat. Organ* **95**, 209–15 (2011).
12. Pridgeon, J. W., Zhang, D. & Zhang, L. Complete genome sequence of the highly virulent *Aeromonas hydrophila* AL09-71 isolated from diseased channel catfish in west Alabama. *Genome. Announc.* **2**, e00450-14 (2014).
13. Seshadri, R. *et al.* Genome sequence of *Aeromonas hydrophila* ATCC 7966T: jack of all trades. *J. Bacteriol.* **188**, 8272–82 (2006).
14. Pridgeon, J. W., Zhang, D. & Zhang, L. Complete genome sequence of a moderately virulent *Aeromonas hydrophila* strain, pc104A, isolated from soil of a catfish pond in west Alabama. *Genome. Announc.* **2**, e00554-14 (2014).

15. Lee, Z. M., Bussema, C., 3rd & Schmidt, T. M. *rrnDB*: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic. Acids Res.* **37**, D489–93 (2009).
16. Grim, C. J. *et al.* Characterization of *Aeromonas hydrophila* wound pathotypes by comparative genomic and functional analyses of virulence genes. *mBio.* **4**, e00064-13 (2013).
17. Ansaldo, M., Simon, G., Lepelletier, M. & Mejean, V. The TorR high-affinity binding site plays a key role in both *torR* autoregulation and *torCAD* operon expression in *Escherichia coli*. *J. Bacteriol.* **182**, 961–6 (2000).
18. Anba-Mondoloni, J., Chaillou, S., Zagorec, M. & Champomier-Verges, M.C. Catabolism of N-acetylneuraminic acid, a fitness function of the food-borne lactic acid bacterium *Lactobacillus sakei*, involves two newly characterized proteins. *Appl. Environ. Microbiol.* **79**, 2012–2018 (2013).
19. Stahl, M. *et al.* L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc. Natl. Acad. Sci. U S A* **108**, 7194–9 (2011).
20. Martino, M. E., Fasolato, L., Montemurro, F., Novelli, E. & Cardazzo, B. *Aeromonas* spp.: ubiquitous or specialized bugs? *Environ. Microbiol.* **16**, 1005–18 (2014).
21. Vilches, S. *et al.* Complete type III secretion system of a mesophilic *Aeromonas hydrophila* strain. *Appl. Environ. Microbiol.* **70**, 6914–9 (2004).
22. Suarez, G. *et al.* Molecular characterization of a functional type VI secretion system from a clinical isolate of *Aeromonas hydrophila*. *Microb. Pathog.* **44**, 344–61 (2008).
23. Schwarz, S., Hood, R. D. & Mougous, J. D. What is type VI secretion doing in all those bugs? *Trends Microbiol.* **18**, 531–7 (2010).
24. Sandkvist, M. Type II secretion and pathogenesis. *Infect Immun.* **69**, 3523–35 (2001).
25. Canals, R. *et al.* Polar flagellum biogenesis in *Aeromonas hydrophila*. *J. Bacteriol.* **188**, 542–55 (2006).
26. Kohler, P. R., Choong, E. L. & Rossbach, S. The RpiR-like repressor IolR regulates inositol catabolism in *Sinorhizobium meliloti*. *J. Bacteriol.* **193**, 5155–63 (2011).
27. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–9 (2013).
28. Sugihara, J., Sun, L., Yan, N. & Kaback, H. R. Dynamics of the L-fucose/H<sup>+</sup> symporter revealed by fluorescence spectroscopy. *Proc. Natl. Acad. Sci. U S A* **109**, 14847–51 (2012).
29. Reed, L. J. & Muench, L. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).
30. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. U S A* **102**, 13950–5 (2005).
31. Mathee, K. *et al.* Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl. Acad. Sci. U S A* **105**, 3100–5 (2008).
32. Canals, R. *et al.* Analysis of the lateral flagellar gene system of *Aeromonas hydrophila* AH-3. *J. Bacteriol.* **188**, 852–62 (2006).
33. Boyd, J. M. *et al.* Contribution of type IV pili to the virulence of *Aeromonas salmonicida* subsp. *salmonicida* in Atlantic salmon (*Salmo salar* L.). *Infect. Immun.* **76**, 1445–55 (2008).
34. Hadi, N. *et al.* Bundle-forming pilus locus of *Aeromonas veronii* bv. *Sobria*. *Infect. Immun.* **80**, 1351–60 (2012).
35. Howard, S. P., Garland, W. J., Green, M. J. & Buckley, J. T. Nucleotide sequence of the gene for the hole-forming toxin aerolysin of *Aeromonas hydrophila*. *J. Bacteriol.* **169**, 2869–71 (1987).
36. Sha, J., Kozlova, E. V. & Chopra, A. K. Role of various enterotoxins in *Aeromonas hydrophila*-induced gastroenteritis: generation of enterotoxin gene-deficient mutants and evaluation of their enterotoxic activity. *Infect. Immun.* **70**, 1924–35 (2002).
37. Hirono, I. & Aoki, T. Nucleotide sequence and expression of an extracellular hemolysin gene of *Aeromonas hydrophila*. *Microb. Pathog.* **11**, 189–97 (1991).
38. Erova, T. E. *et al.* Identification of a new hemolysin from diarrheal isolate SSU of *Aeromonas hydrophila*. *FEMS Microbiol. Lett.* **275**, 301–11 (2007).
39. Singh, V., Mani, I., Chaudhary, D. K. & Somvanshi, P. Molecular detection and cloning of thermostable hemolysin gene from *Aeromonas hydrophila*. *Mol. Biol. (Mosk)* **45**, 601–9 (2011).
40. Suarez, G. *et al.* Actin cross-linking domain of *Aeromonas hydrophila* repeat in toxin A (RtxA) induces host cell rounding and apoptosis. *Gene* **506**, 369–76 (2012).
41. Hernanz Moral, C. *et al.* Molecular characterization of the *Aeromonas hydrophila* *aroA* gene and potential use of an auxotrophic *aroA* mutant as a live attenuated vaccine. *Infect. Immun.* **66**, 1813–21 (1998).
42. Erova, T. E. *et al.* DNA adenine methyltransferase influences the virulence of *Aeromonas hydrophila*. *Infect. Immun.* **74**, 410–24 (2006).
43. Cascon, A. *et al.* A major secreted elastase is essential for pathogenicity of *Aeromonas hydrophila*. *Infect. Immun.* **68**, 3233–41 (2000).
44. Sha, J. *et al.* Surface-expressed enolase contributes to the pathogenesis of clinical isolate SSU of *Aeromonas hydrophila*. *J. Bacteriol.* **191**, 3095–107 (2009).
45. Chang, T. M., Liu, C. C. & Chang, M. C. Cloning and sequence analysis of the gene (*eprA1*) encoding an extracellular protease from *Aeromonas hydrophila*. *Gene* **199**, 225–9 (1997).
46. Sha, J. *et al.* Molecular characterization of a glucose-inhibited division gene, *gidA*, that regulates cytotoxic enterotoxin of *Aeromonas hydrophila*. *Infect. Immun.* **72**, 1084–95 (2004).
47. Merino, S. *et al.* Cloning, sequencing, and role in virulence of two phospholipases (A1 and C) from mesophilic *Aeromonas* sp. serogroup O:34. *Infect. Immun.* **67**, 4008–13 (1999).
48. Erova, T. E. *et al.* Cold shock exoribonuclease R (VacB) is involved in *Aeromonas hydrophila* pathogenesis. *J. Bacteriol.* **190**, 3467–74 (2008).
49. Yu, H. B. *et al.* Identification and characterization of putative virulence genes and gene clusters in *Aeromonas hydrophila* PPD134/91. *Appl. Environ. Microbiol.* **71**, 4469–77 (2005).
50. Pillai, L. *et al.* Molecular and functional characterization of a ToxR-regulated lipoprotein from a clinical isolate of *Aeromonas hydrophila*. *Infect. Immun.* **74**, 3742–55 (2006).
51. Canals, R. *et al.* Role of Gne and GalE in the virulence of *Aeromonas hydrophila* serotype O34. *J. Bacteriol.* **189**, 540–50 (2007).
52. Vilches, S. *et al.* Mesophilic *Aeromonas* UDP-glucose pyrophosphorylase (GalU) mutants show two types of lipopolysaccharide structures and reduced virulence. *Microbiology* **153**, 2393–404 (2007).
53. Swift, S. *et al.* Quorum sensing in *Aeromonas hydrophila* and *Aeromonas salmonicida*: identification of the LuxRI homologs AhyRI and AsaRI and their cognate N-acylhomoserine lactone signal molecules. *J. Bacteriol.* **179**, 5271–81 (1997).
54. Kozlova, E. V. *et al.* Mutation in the S-ribosylhomocysteinase (*luxS*) gene involved in quorum sensing affects biofilm formation and virulence in a clinical isolate of *Aeromonas hydrophila*. *Microb. Pathog.* **45**, 343–54 (2008).
55. Khajanchi, B. K., Kozlova, E. V., Sha, J., Popov, V. L. & Chopra, A. K. The two-component QseBC signalling system regulates *in vitro* and *in vivo* virulence of *Aeromonas hydrophila*. *Microbiology* **158**, 259–71 (2012).
56. Ebanks, R. O., Dacanay, A., Goguen, M., Pinto, D. M. & Ross, N. W. Differential proteomic analysis of *Aeromonas salmonicida* outer membrane proteins in response to low iron and *in vivo* growth conditions. *Proteomics* **4**, 1074–85 (2004).
57. Sha, J., Lu, M. & Chopra, A. K. Regulation of the cytotoxic enterotoxin gene in *Aeromonas hydrophila*: characterization of an iron uptake regulator. *Infect. Immun.* **69**, 6370–81 (2001).

58. Najimi, M., Lemos, M. L. & Osorio, C. R. Identification of heme uptake genes in the fish pathogen *Aeromonas salmonicida* subsp. *salmonicida*. *Arch. Microbiol.* **190**, 439–49 (2008).
59. Avison, M. B., Niumsup, P., Walsh, T. R. & Bennett, P. M. *Aeromonas hydrophila* AmpH and CepH beta-lactamases: derepressed expression in mutants of *Escherichia coli* lacking *creB*. *J. Antimicrob. Chemother.* **46**, 695–702 (2000).
60. Massidda, O., Rossolini, G. M. & Satta, G. The *Aeromonas hydrophila* *cphA* gene: molecular heterogeneity among class B metallo-beta-lactamases. *J. Bacteriol.* **173**, 4611–7 (1991).

### Acknowledgements

This work was supported by the National Natural Science Foundation of China (31372454, 31072151), the Three New Aquatic Projects in Jiangsu Province (D2013–5–4), the ‘333’ Project Talent Training Fund of Jiangsu Province, and the Priority Academic Program Development of Jiangsu Higher Educational Institutions (PAPD).

### Author Contributions

Y.J.L. initiated the genome project; Y.J.L., M.D.P. and X.X. conceived the study and drafted the paper; F.C.L., J.W.J., M.D.P. and A.H.K. performed bioinformatics analysis; M.D.P., X.X., Y.F.W. and Y.H.D. performed the experiments; C.P.L., H.C.Y. and W.Z. contributed the materials and provided valuable suggestions.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Pang, M. *et al.* Novel insights into the pathogenicity of epidemic *Aeromonas hydrophila* ST251 clones from comparative genomics. *Sci. Rep.* **5**, 9833; doi: 10.1038/srep09833 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>