BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(10)

Genome sequence and comparative analysis of *Avibacterium paragallinarum*

David Requena[1, 2], Ana Chumbe[1], Michael Torres[1, 2], Ofelia Alzamora[1], Manuel Ramirez[1, 2], Hugo Valdivia-Olarte[1, 2], Andres Hazaet Gutierrez[1, 2, 3], Ray Izquierdo-Lara[1], Luis Enrique Saravia[1], Milagros Zavaleta[1], Luis Tataje-Lavanda[1], Ivan Best[1], Manolo Fernández-Sánchez[1], Eliana Icochea[1, 4], Mirko Zimic[1, 2] & Manolo Fernández-Díaz[1]* - FARVET Research Group

[1]FARVET S.A.C. Carretera Panamericana Sur Nº 766 Km 198.5, Chincha Alta. Ica – Peru; [2]Laboratorio de Bioinformática y Biología Molecular, Laboratorios de Investigación y Desarrollo, Facultad de Ciencias y Filosofía. Universidad Peruana Cayetano Heredia. Av. Honorio Delgado 430, San Martin de Porres Lima –Peru; [3]Institute for Immunology and Informatics, University of Rhode Island, 292 Morris Ave. Providence, RI – USA, ZIP: 02906-2611; [4]Laboratorio de Patología Aviar. Facultad de Medicina Veterinaria, Universidad Nacional Mayor de San Marcos. Lima, Peru. Av. Circunvalación Cdra. 28 s/n, San Borja. Lima - Perú; Manolo Fernández-Díaz– Email: farvet@farvet.com; Phone: (+511) 3441419; *Corresponding author

Received May 18, 2013; Accepted May 27, 2013; Published June 08, 2013

Abstract:
Background: *Avibacterium paragallinarum*, the causative agent of infectious coryza, is a highly contagious respiratory acute disease of poultry, which affects commercial chickens, laying hens and broilers worldwide.
Methodology: In this study, we performed the whole genome sequencing, assembly and annotation of a Peruvian isolate of *A. paragallinarum*. Genome was sequenced in a 454 GS FLX Titanium system. *De novo* assembly was performed and annotation was completed with GS De Novo Assembler 2.6 using the *H. influenzae* str. F3031 gene model. Manual curation of the genome was performed with Artemis. Putative function of genes was predicted with Blast2GO. Virulence factors were identified by comparison with the Virulence Factor Database.
Results: The genome obtained has a length of 2.47 Mb with 40.66% of GC content. Seventy five large contigs (>500 nt) were obtained, which comprised 1,204 predicted genes. All the contigs are available in Genbank [GenBank: PRJNA64665]. A total of 103 virulence factors, reported in the Virulence Factor Database, were found in *A. paragallinarum*. Forty four of them are present in 7 species of Haemophilus, which are related with pathogenesis, virulence and host immune system evasion. A tetracycline-resistance associated transposon (Tn10), was found in *A. paragallinarum*, possibly acting as a defense mechanism.
Discussion and conclusion: The availability of *A. paragallinarum* genome represents an important source of information for the development of diagnostic tests, genotyping, and novel antigens for potential vaccines against infectious coryza. Identification of virulence factors contributes to better understanding the pathogenesis, and planning efforts for prevention and control of the disease.

Key words: Infectious coryza, genome sequencing, virulence factors, syntenic homology, Tn10 transposon.

Background:
Infectious coryza is a highly contagious acute respiratory disease of layers and chickens, caused by the gram negative bacterium *Avibacterium paragallinarum*. This organism was initially classified as *Haemophilus paragallinarum*, but later re-classified as *A. paragallinarum* **[1].** The disease has a worldwide distribution, being endemic in commercial laying hens. Broilers may also be severely affected with depression, decreased feed and water consumption that negatively affect their growth. This lead to important economic losses associated to poor

productivity, increased mortality up to 48% and a reduction up to 75% of egg production. Although this disease is rarely seen in broilers, an outbreak in Panama caused mortality and 45% of production losses **[2].**

A. paragallinarum produces an acute catarrhal inflammation of mucous membranes and sinus passages, as well as catarrhal and subcutaneous edema of face and wattles. It is characterized by nasal discharge, watery eyes, facial swelling, anorexia, diarrhea and swelling of wattles. *A. paragallinarum* infection was also reported in non-respiratory organs such as liver, kidney and tarsus **[3].**

Nowadays, the use of inactivated *A. paragallinarum* vaccines against infectious coryza formulated from local strains is the best way to control the disease **[4, 5]**. However, these vaccines have the disadvantage of inducing protection only against the serotypes included in the vaccine, but not to other strains **[6].** The lack of effective vaccines for *A. paragallinarum* requires new efforts and the use of novel approaches. The *A. paragallinarum* genome sequence is an important source of information for a better understanding of the biology of this pathogen, in particular in the development of vaccines and more accurate methods of genotyping.

In this study we present the genome sequence and its annotation of a circulating pathogenic strain of *A. paragallinarum* isolated from a broiler outbreak in Ica, a city in the central coast of Peru. This strain was previously identified as serovar C. This genome was compared with related organisms, focusing on virulence factors.

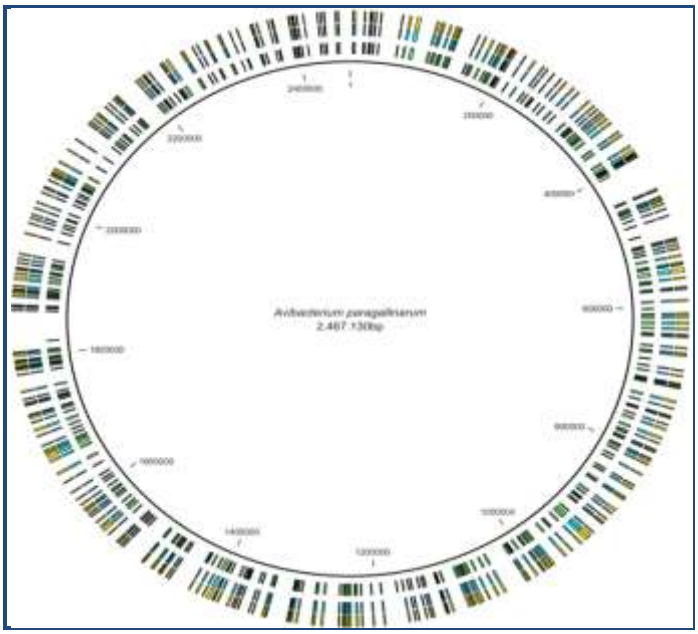

Figure 1: Circular representation of *Avibacterium paragallinarum* genome. Contigs obtained from whole genome sequencing was joined in a contiguous pseudo molecule, which was used to predict genes and estimate the GC content. This circular plot shows the distribution of coding DNA sequences, genes and mRNAs obtained in the *A. paragallinarum* genome. Each bar in the internal circle represents the mRNAs, in the middle circle represents the genes, and in the outer circle represents the coding DNA sequences, respectively.

Methodology:

Microbiological culture

A Peruvian local isolate of *A. paragallinarum* was obtained by direct culture of the infraorbital sinuses of a broiler from an infectious coryza outbreak in a local farm (Ica, Peru). The isolate was cultured in chocolate agar **[7]** with factor X-V and incubated in a microenvironment with 5% CO_2 at 40°C for 48 hours. After three passages, the colonies were collected in a modified BHI culture with 50% glycerol and stored at -80°C.

DNA library and sequencing

Genomic DNA was extracted from reactivated bacteria using DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) with slightly modification that includes a lysozyme treatment at 37°C for 1h followed by incubation with Proteinase K at 56°C overnight. The quantity and quality of the eluted DNA was tested with picogreen kit (Invitrogen, Carlsbad, CA, USA) and Biophotometer Plus (Eppendorf, Hamburg, Germany) respectively.

A 454-FLX shotgun library was prepared with 500 ng of genomic DNA using a GS FLX Titanium Rapid Library Preparation Kit (Roche, Branford, CT, USA). Quality assessment was performed with the Agilent Bioanalyzer using High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA). The obtained library was clonally amplified within a water-in-oil emulsion (EmPCR). EmPCR and titration by enrichment were made using GS FLX Titanium SV emPCR Kit Lib-L (Roche, Branford, CT, USA). Then, the DNA beads were sequenced in a GS FLX Titanium PicoTiterPlate 70x75 (Roche, Branford, CT, USA) on the GS FLX+ Sequencing System.

Data processing and assembly

The raw signal data were processed with the software GS Run Processor to obtain the reads. Given the lack of a reference genome for mapping assembly, *de novo* assembly was conducted using GS De Novo Assembler 2.6. The chicken genome [GenBank: PRJNA10808] was filtered out of the assembly. Quality control was performed using GS De Novo Assembler and CLC Main Workbench 6.7. We discarded chimeric sequences and homopolymeric errors originated by the pyrosequencing process itself.

Genome annotation

A pseudomolecule was built by arbitrarily joining the 75 largest contigs, adding 100 ambiguous nucleotides (represented with N) between each pair of contigs. The pseudomolecule was tested for presence of coding DNA sequences using Glimmer v3.02 **[8]** SNAP **[9],** and Augustus **[10]** with a gene model trained with the annotated genome of *H. influenzae* str. F3031 [GenBank:PRJNA62123]. Genome annotation was performed using the semi-automatic annotation pipeline Maker-GMOD **[11].** Specific analysis included ab-initio gene prediction from SNAP, Augustus, Fgenesh **[12]** and GeneMark **[13]**; protein alignments from EXONERATE **[14]** and BLASTx **[15],** and repetitive regions identified by RepeatMasker **[16].** Gene function and metabolic pathways predictions were obtained with the Blast2GO annotation pipeline **[17].** A manual curation of the genome annotation was performed using Artemis **[18].** This procedure included the verification of Open Reading Frames, stop/start codon of coding sequences and indels.

Virulence factors analysis

Local BLAST **[15]** was performed between the *A. paragallinarum* genome and the Virulence Factor Database (VFDB) [19], using e-value 1.10^{-3} and 60% of identity as cutoff. The predicted virulence factors in *A. paragallinarum* were compared with the virulence factors compiled in a comparative table indexed in VFDB for all *Haemophilus* species available in this database, which are: *H. ducreyi* 35000HP [GenBank:NC_002940], *H. influenzae* Rd KW20 [GenBank:NC_000907], *H. influenzae* 86-028NP [GenBank:NC_007146], *H. influenzae* PittEE [GenBank:NC_009566], *H. influenzae* PittGG [GenBank:NC_009567], *H. somnus* 129PT [GenBank:NC_008309] and *H. somnus* 2336 [GenBank:NC_010519]. We compare with Haemophilus because it was the closest organism with information available in VFDB.

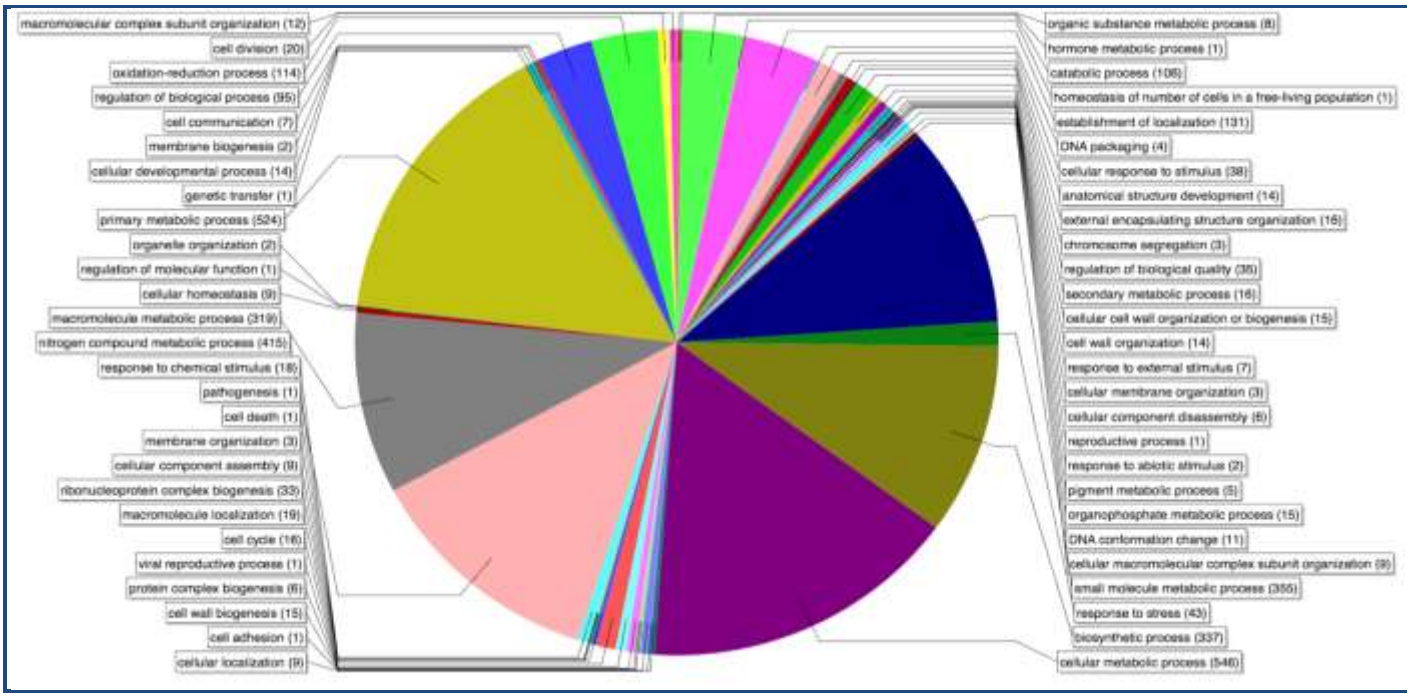

Figure 2: Functional classification of genes. Predicted genes from annotation process were processed in Blast2GO for functional annotation. This graphics shows the relative abundance of biological functions assigned.

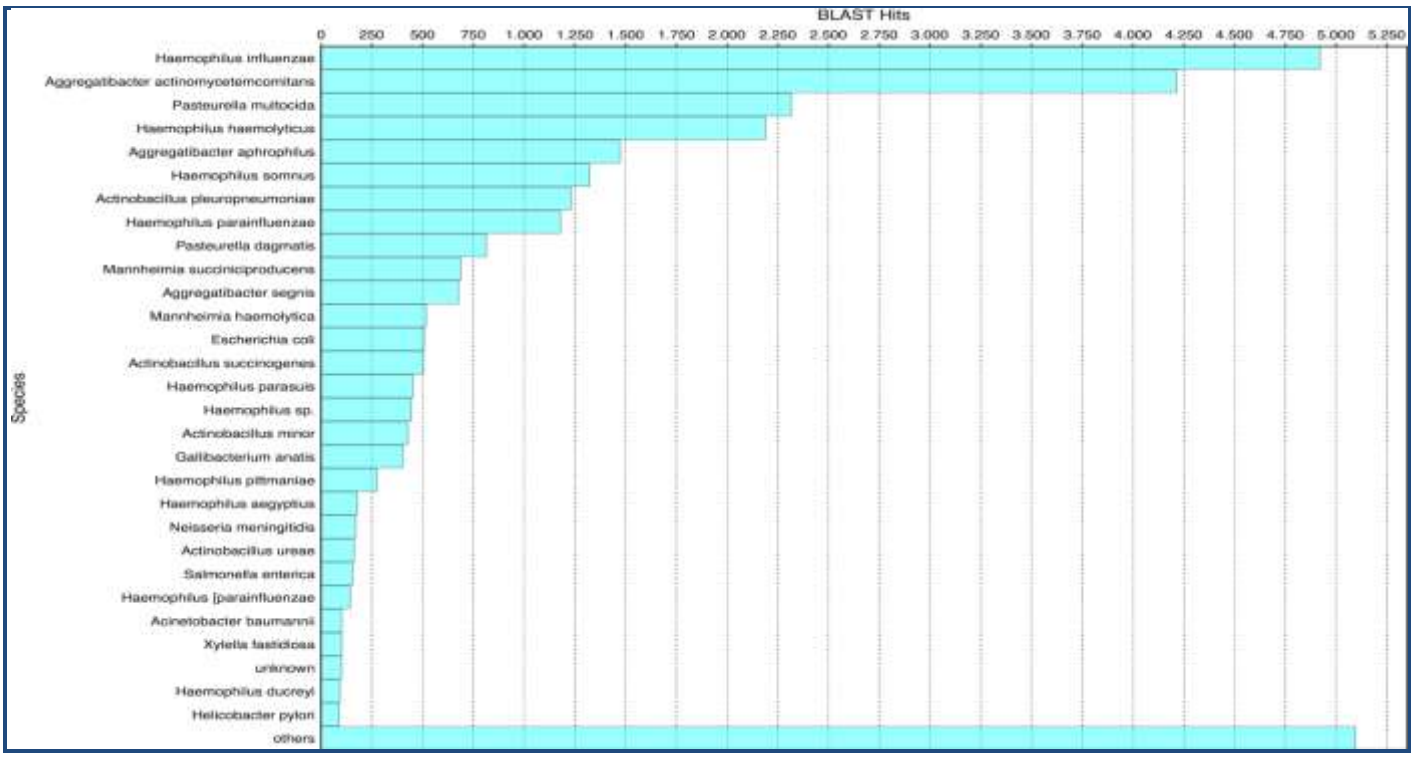

Figure 3: Amount of sequences matching with *Avibacterium paragallinarum*, by organism. Distribution of the number sequences which matches with the predicted ORFs of *A. paragallinarum*, by organism, in the Blast2GO functional annotation. Horizontal bars

corresponds to the number of significant BLAST hits between *A. paragallinarum* predicted ORFs and the non-redundant database from GenBank, for each organism listed. *H. influenzae*, *A. actinomycetemcomitans*, *P. multocida*, and *H. haemolyticus* were the mayor contributors for *A. paragallinarum* genome annotation.

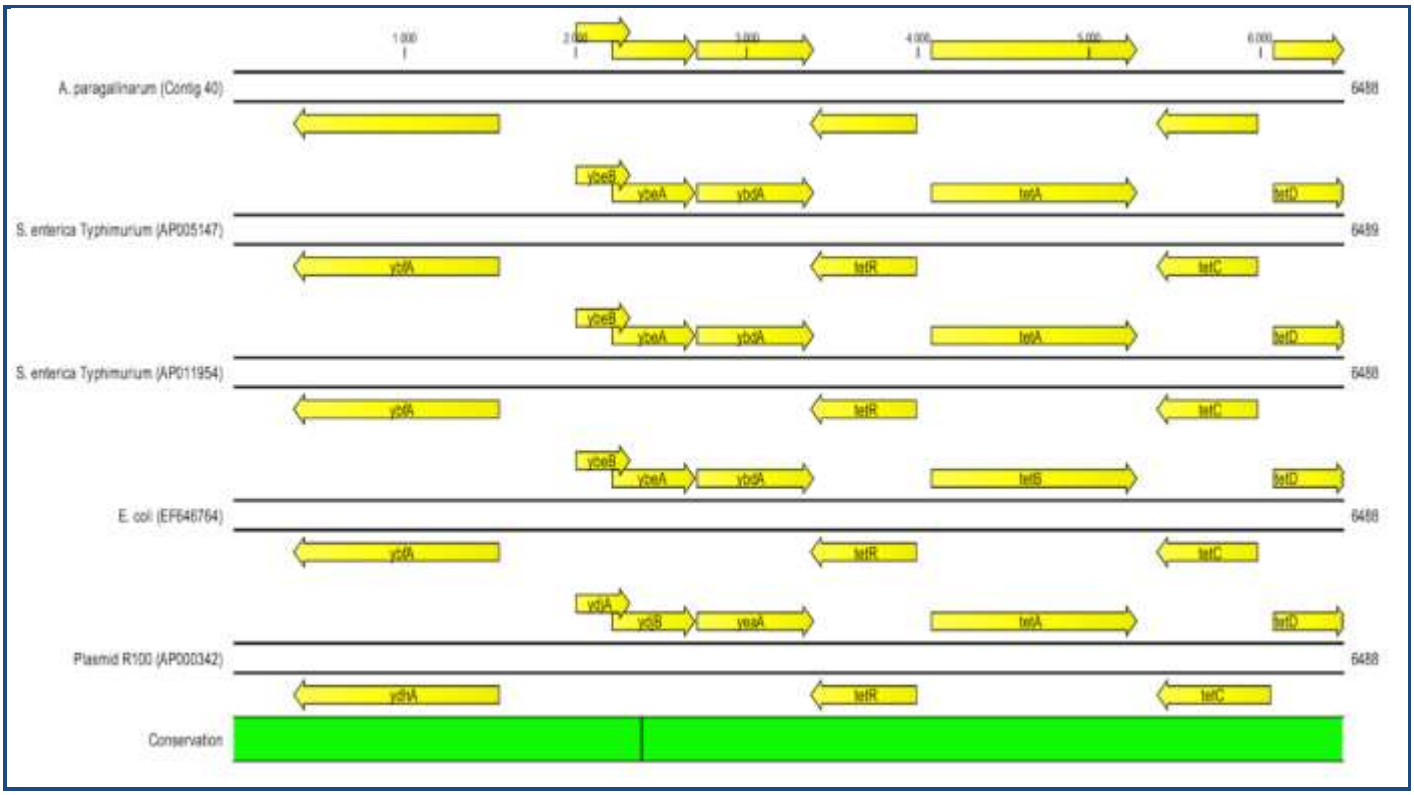

Figure 4: Tn10 transposon and tetracycline-resistance genes from *Avibacterium paragallinarum*. Sequence comparison and gene annotations (arrows) in a region of Tn10 transposon found in *A. paragallinarum* genome sequencing (Contig40). This transposon was reported in plasmid sequences of chicken pathogens: *Salmonella enterica* subsp. enterica serovar Typhimurium plasmid R64 [GenBank: AP005147], *Salmonella enterica* subsp. enterica serovar Typhimurium plasmid R621a [GenBank:AP011954], Plasmid R100 [GenBank:AP000342] and *Escherichia coli* transposon Tn10 [GenBank:EF646764]. Conservation bar (in green) shows there are few variations (black lines inside) between the compared sequences. This region was found with 99.9% of identity with the organisms listed above.

Results:

Sequencing, assembly and annotation

The average fragment length was 600-900 nt. The whole shotgun sequencing reached a mean depth of 23X, producing 183,434 reads (62'190,061 nt). 98.12% of the total reads formed contigs, obtaining 93 contigs (2'465,440 nt) with a N50 of 113,569 nt. The 75 largest contigs (>500 nt) comprised 2'459,730 nt. 99.70% of these showed a quality greater than Q40. The largest contig size was 439,531 nt and the average contig size was 32,796 nt. This assembly produced an estimated genome size of 2.47 Mb with 40.66% of GC content **(Figure 1).** All the contigs are available in Genbank [GenBank: PRJNA64665].

A total of 1,204 genes were predicted from the pseudomolecule. All of these were assigned with a putative function using Blast2GO **Table 1 (see supplementary material).** Eight major categories and 49 sub-categories from gene ontology were identified **Table 2 (see supplementary material)).** The distribution of orthologous genes clusters is presented in **(Figure 2).** *H. influenzae*, followed by *Aggregatibacter actinomycetemcomitans* and *P. multocida* were the organisms that showed the highest number of homologue genes with *A. paragallinarum* **(Figure 3).**

Virulence factors analysis

One hundred and three virulence factors from the VFDB were found in *A. paragallinarum* **Table 3 (see supplementary material),** and 44 of them were found in common with the 7 *Haemophilus* compared in the database **[19] Table 4 (Available with authors).** From these results, we found an IgA protease, adherence-related factors (ompP5 and type IV pili proteins); and a region of 6,488 nt highly identical (>99%) to the transposon Tn10, containing four tetracycline resistance genes **Table 4 (Available with authors)** with **(Figure 4)**.

Discussion & Conclusion:

The present study presents for first time a draft genome sequence of *A. paragallinarum,* its annotation and comparison with *Haemophilus*, identifying potential virulence factors. Interestingly, Tn10 transposon was found partially in *A. paragallinarum* sequencing data. This transposon was found in plasmids from several chicken pathogens, including *Escherichia coli*, and *Salmonella enterica* serovar Typhimurium. Tn10 has been used to induce mutagenesis to study the effect of mutations in the fitness **[20],** and for the construction of a tagged mini-Tn10 plasmid bank to attenuate the pathogen virulence, which could be used as live attenuated vaccine **[21].** Tn10 is a transposon of 9,147 nt, comprising four genes

associated to tetracycline resistance (tetR/A/C/D) **[22]**. These genes may potentially cause tetracycline resistance in *A. paragallinarum*, which needs to be further studied. Tn10 typically has two insertion sequences (IS10-L and IS10-R) and two transposases (ydgA and yedA) flanking them. However, these sequences were not found in the *A. paragallinarum* corresponding contig (6,488 nt), probably due to lack of coverage. IgA protease was found in the *A. paragallinarum* genome, suggesting that this specie may be able to hydrolyze chicken IgA-like inmunoglobulins. IgA proteases were reported in related species as *H. influenzae* **[23],** *Neisseria meningitidis* **[24]**. This protease is known to cleave host secreted IgA immunoglobulin enabling to circumvent host mucosal defense mechanisms; enhancing the ability to infect respiratory tract **[25].**

Studies in bacterial pathogens have shown that the profile of virulence genes are associated with disease **[26]**. Therefore, genomic comparison analysis provides the basis for understanding pathogenicity and for rational vaccine design and immunoassays development. It was interesting to find OmpP5 and Type IV pili virulence factors. It is known that OmpP5 is an outer membrane protein homologue to *E. coli* OmpA **[27],** the major protective antigen responsible for the integrity of the outer membrane, which induces strong antibody response in chickens **[28].** Type IV pili are involved in a variety of bacterial functions, including cell adhesion **[29],** bacteriophage adsorption, plasmid transfer **[30],** and twitching motility, a form of flagellum-independent locomotion **[31].** Contamination with chicken DNA was reduced from the assembly by filtering the chicken genome. Assembly errors, contig ordering and genome closure were not made since no paired-end library was produced and no reference genome was available yet. Therefore, it is important to perform complementary studies to build the complete chromosome in order to define the genetic structure and perform more accurate comparisons with related organisms.

The availability of *A. paragallinarum* genome is an important achievement for poultry industry, which would facilitate the development of useful tools against infectious coryza. Furthermore, the identification of virulence factors and immunogenic and antibiotic-resistant factors contributes to understanding the pathogenesis, and contribute to efforts for prevention and control of the disease.

Authors' contributions:
Conception and design: DR, AC and MFD. Acquisition, analysis and interpretation of data: DR, AC, MT, OA, MR, MZi. Drafting of the manuscript and revising it critically: DR, AC, MT, OA, MR, HV, AHG, RIL, LES, MZa, LTL, IB, MFS, EI, MZi and MFD. Final approval of the version to be published: DR, MZi, MFD.

Acknowledgement:
To Rocío Quispe, Isabel García, David Tarazona and Lenin Maturrano for their participation in the course of this project.

References:
[1] Blackall PJ *et al. Int J Syst Evol Microbiol*. 2005 **55:** 353 [PMID: 15653900]
[2] Calderon EN *et al. Avian Dis*. 2010 **54:** 1095 [PMID: 20945795]
[3] Droual R *et al. Avian Dis*. 1990 **34:** 1009 [PMID: 2282006]
[4] Wambura PN, *Trop Anim Health Prod*. 2010 **42:** 483 [PMID: 19731063]
[5] Chukiatsiri K *et al. Avian Dis*. 2009 **53:** 382 [PMID: 19848076]
[6] Wu JR *et al. Vaccine*. 2011 **29:** 660 [PMID: 21115052]
[7] Gunn BA, *J Clin Microbiol*. 1984 **20:** 822 [PMID: 6490866]
[8] Salzberg SL *et al. Nucleic Acids Res*. 1998 **26:** 544 [PMID: 9421513]
[9] Korf I, *BMC Bioinformatics*. 2004 **5:** 59 [PMID: 15144565]
[10] Keller O *et al. Bioinformatics*. 2011 **27:** 757 [PMID: 21216780]
[11] Cantarel BL *et al. Genome Res*. 2008 **18:** 188 [PMID: 18025269]
[12] Tyson GW *et al. Nature*. 2004 **428:** 37 [PMID: 14961025]
[13] Besemer J *et al. Nucleic Acids Res*. 2001 **29:** 2607 [PMID: 11410670]
[14] Slater GS & Birney E, *BMC Bioinformatics*. 2005 **6:** 31 [PMID: 15713233]
[15] Altschul SF *et al. Nucleic Acids Res*. 1997 **25:** 3389 [PMID: 9254694]
[16] Bao Z & Eddy SR, *Genome Res*. 2002 **12:** 1269 [PMID: 12176934]
[17] Conesa A *et al. Bioinformatics*. 2005 **21:** 3674 [PMID: 16081474]
[18] Rutherford K *et al. Bioinformatics*. 2000 **16:** 944 [PMID: 11120685]
[19] Chen L *et al. Nucleic Acids Res*. 2012 **40:** D641 [PMID: 22067448]
[20] Tamura M *et al. J Bacteriol*. 2012 **194:** 1919 [PMID: 22328678]
[21] Fuller TE *et al. Microb Pathog*. 2000 **29:** 25 [PMID: 10873488]
[22] Haniford DB, *Crit Rev Biochem Mol Biol*. 2006 **41:** 407 [PMID: 17092825]
[23] Mistry DV & Stockley RA, *Virulence*. 2011 **2:** 103 [PMID: 21422810]
[24] Vitovski S & Sayers JR, *Infect Immun*. 2007 **75:** 2875 [PMID: 17353288]
[25] Mistry D & Stockley RA, *Int J Biochem Cell Biol*. 2006 **38:** 1244 [PMID: 16293440]
[26] Sassetti C & Rubin EJ, *Curr Opin Microbiol*. 2002 **5:** 27 [PMID: 11834365]
[27] Sugawara E & Nikaido H, *J Biol Chem*. 1992 **267:** 2507 [PMID: 1370823]
[28] Fadl AA et *al. J Appl Microbiol*. 2002 **92:** 180 [PMID: 11849343]
[29] Manning PA & Meyer TF, *Gene*. 1997 **192:** 1 [PMID: 9244776]
[30] Kim SR & Komano T, *J Bacteriol*. 1997 **179:** 3594 [PMID: 9171405]
[31] Merz AJ *et al. Nature*. 2000 **407:** 98-102 [PMID: 10993081]

Edited by P Kangueane
Citation: **Requena *et al.*** Bioinformation 9(10): 528-536 (2013)

BIOINFORMATION

open access

Supplementary material:

Table 1: Genome annotation statistics. The overall statistical prediction shows the nucleotidic composition of the genes and the pseudomecule built for the annotation process. A total of 1,204 genes were predicted, with an average length of 1,037 nt.

	Genome annotation statistics	
	Genes predicted (1204)	**Pseudomolecule**
Number of bases	1249098	2467130
A composition	29.58%	29.39%
C composition	20.39%	20.92%
G composition	22.90%	19.73%
T composition	27.11%	29.64%
N composition	0.00%	0.29%
GC percentage	43.30%	40.65%
Gene density in pseudomolecule	0.488 genes per kb	
Average gene length	1037	

Table 2: Functional annotation of *Avibacterium paragallinarum*. This table shows 8 major categories and 49 sub-categories assigned to the annotated genes. These categories correspond to the general Cluster of Orthologous Genes functional categories.

	Functional annotation of *Avibacterium paragallinarum*	
Major Functions	**Categories by Gene Ontology**	**Number of sequences**
Biological process	metabolic process	585
	cellular process	557
	localization	126
	response to stimulus	58
	cellular component organization or biogenesis	59
	biological regulation	85
	carbon utilization	3
	signaling	4
	multi-organism process	3
	developmental process	14
	biological adhesion	1
Cellular metabolic process	phosphorus metabolic process	59
	heterocycle metabolic process	120
	generation of precursor metabolites and energy	33
	cofactor metabolic process	50
	cellular aromatic compound metabolic process	35
	sulfur compound metabolic process	13
	photosynthesis	2
	peptide metabolic process	3
Metabolic process	primary metabolic process	475
	oxidation-reduction process	103
	macromolecule metabolic process	282
	biosynthetic process	294
	small molecule metabolic process	276
	nitrogen compound metabolic process	340

	catabolic process	78
	organophosphate metabolic process	14
Cellular amino acid metabolic process	glutamine family amino acid metabolic process	18
	branched chain family amino acid metabolic process	8
	aspartate family amino acid metabolic process	12
	histidine family amino acid metabolic process	10
	amino acid activation	13
	serine family amino acid metabolic process	5
	cellular modified amino acid metabolic process	7
tRNA aminoacylation for protein translation	aspartyl-tRNA aminoacylation	3
	phenylalanyl-tRNA aminoacylation	2
	lysyl-tRNA aminoacylation	2
	glutamyl-tRNA aminoacylation	1
	glutaminyl-tRNA aminoacylation	1
	arginyl-tRNA aminoacylation	1
	asparaginyl-tRNA aminoacylation	1
Transport	organic substance transport	40
	ion transport	29
	nitrogen compound transport	10
	cofactor transport	3
	secretion	4
Nucleotide metabolic process	ribonucleotide metabolic process	43
Amino acid transport	aromatic amino acid transport	3
	branched-chain aliphatic amino acid transport	1

Table 3: Virulence factors of *Avibacterium paragallinarum*. Full list of virulence factors (103) found in *Avibacterium paragallinarum*, with at least 60% of identity with VFDB. A short description of their functions is provided, according the information obtained by the annotation process.

Virulence factors of *Avibacterium paragallinarum*	
Virulence Factor	**Short Description**
algU	alginate biosynthesis protein AlgZ/FimS
bexA	ATP-dependent polysaccharide export protein
bexB	Capsular polisaccharide export protein
bexC	Capsular polisaccharide export protein
bexD	Capsular polisaccharide export protein
clpE	ATP-dependent protease
clpP	ATP-dependent Clp protease proteolytic subunit
clpV1	Required for secretion of hcp1 probably by providing the energy source for its translocation
comE/pilQ	competence protein E, type IV pilus secretin PilQ
cpsG	phosphomannomutase
csrA	Binds to mRNA to regulate post-transcriptional activity. Regulates glycogen synthesis and cell size and surface properties
ctrB	Membrane fusion protein, capsular polisaccharide export
ctrC	capsule polysaccharide export inner-membrane protein
ctrD	capsule polysaccharide export ATP-binding protein
cyaB	Involved in the export of calmodulin-sensitive adenylate cyclase-hemolysin (cyclolysin)
eno	phosphopyruvate hydratase, putative enolase
fcl	putative fucose synthetase
fur	transcriptional represor of iron-responsive genes (Fur family) (ferric uptake regulator)
galE	UDP-glucose 4-epimerase

galU	UTP--glucose-1-phosphate uridylyltransferase, carbon storage regulator, argininosuccinate lyase
glnA1	glutamine synthetase
gmd	GDP-D-mannose dehydratase
gmhA/lpcA	phosphoheptose isomerase
hasB	UDP-glucose 6-dehydrogenase
hemA	glutamyl-tRNA reductase
hemB	porphobilinogen synthase
hemC	hydroxymethylbilane synthase, porphobilinogen deaminase
hemE	uroporphyrinogen decarboxylase
hemH	ferrochelatase
hemL	glutamate-1-semialdehyde-2,1-aminomutase, aminotransferase
hemN	oxygen-independent coproporphyrinogen III oxidase
hemX	phosphate-starvation-inducible protein PsiE, putative uroporphyrin-III C-methyltransferase
hemY	porphyrin biosynthesis protein
hlyB	hemolysin transport-secretion protein, alpha-hemolysin translocation ATP-binding protein
hlyD	HlyD protein
hscA	Transport of secretory b polysaccharide accors the outer membrane
hscB	Transport of secretory b polysaccharide accors the outer membrane
htpB	Hsp60, 60kDa heat shock protein, chaperonin
htrB	lipid A biosynthesis lauroyl acyltransferase
hxuC	heme-hemopexin utilization protein C
iga	IgA1 protease, specific serine endopeptidase, specific metalloendopeptidase
intT1	Tn21 integrase IntI1
IS1016	Insertion sequence
kdsA	2-dehydro-3-deoxyphosphooctonate aldolase
kdtA	3-deoxy-D-manno-octulosonic-acid transferase
kdtB	lipopolysaccharide core biosynthesis protein
kfiD	putative UDP-glucose 6-dehydrogenase
kpsF	arabinose-5-phosphate isomerase
lap	alcohol-acetaldehyde dehydrogenase
lgtF	UDP-glucose--lipooligosaccharide glucosyltransferase
lipA	capsule polysaccharide modification protein LipA
lipB	capsule polysaccharide modification protein
lpxA	acyl-(acyl-carrier-protein)--UDP-N-acetylglucosamine O-acyltransferase
lpxB	lipid-A-disaccharide synthase
lpxC	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
lpxD	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
lpxH	UDP-2,3-diacylglucosamine hydrolase
lsgF	putative UDP-galactose--lipooligosaccharide galactosyltransferase
manA	mannose-6-phosphate isomerase
mrsA/glmM	phosphoglucosamine mutase
msbA	lipid A, export ABC transporter ATP-binding protein
msbB	lipid A biosynthesis (KDO)2-(lauroyl)-lipid IVA acyltransferase
mutS	methyl-directed mismatch repair, recognize exocyclic adducts of guanosine
nanE	Sialic acid catabolism, which confers advantage in mucus intestine
neuB	N-acetyl neuramic acid synthetase
oapA	opacity associated protein
ompA/ompP5	major outer membrane protein homolog, OmpA2, hemagglutinin antigen
opsX/rfaC	ADP-heptose--lipooligosaccharide heptosyltransferase 1
orf 4 E. coli 536	Involved in horizontal gene transfer
orf 7 E. coli 536	Involved in horizontal gene transfer
orf16 E. coli 536	Putative F17-like fimbrial isher
orf17 S. flexneri R27	methyltransferase homologue to ybeA/ydjA
orf18 S. flexneri R27	transcriptional regulator, homologue to ybdA/yeaA
orf45 E. coli 536	putative lysil-tRNA synthetase LysU
orf70 E. coli 536	Involved in horizontal gene transfer
orfM	non-canonical purine NTP pyrophosphatase, rdgB/HAM1 family
pgi	glucose-6-phosphate isomerase
pilB	putative type IV fimbrial biogenesis secretion protein
radC	DNA repair protein RadC

relA	GTP pyrophosphokinase
rfaD	ADP-L-glycero-D-mannoheptose-6-epimerase
rfaE	ADP-heptose synthase, bifunctional heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase
rfaF	ADP-heptose-LPS heptosyltransferase II
rtxB	RTX toxin transporter, ABC-type bacteriocin/lantibiotic exporter
sigA	Mycobacterium tuberculosis H37Rv
sitA	Sallmonella iron transporter fur regulated
sitB	Sallmonella iron transporter fur regulated
sitC	Sallmonella iron transporter fur regulated
ssb	ssDNA-binding protein controls activity of RecBCD nuclease
tetA(B)	tetracycline resistance protein TetA(B)
tetC	putative transcriptional regulator
tetD	putative transcriptional regulator
tetR	tet repressor
tuf	elongation factor Tu
VC1777	sialic acid-specific subfamily of TRAP transporters
VC1779	sialic acid-specific subfamily of TRAP transporters
waaC	heptosyltransferase I
wbfV/wcvB	Predicted UDP-glucose 6-dehydrogenase
wecA	PII uridylyl-transferase, undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase
y1079	DNA-polymerase repair proteins
ybeB/ydjA	Ribosomal silencing factor RsfS
ybfA/ydhA	sodium/glutamate simporter
yhxB/manB	phosphomannomutase