Developing a molecular consensus between multiple phylogenies: a case study of the genus *Aeromonas*

Kamlesh Jangid, Bhagwan Rekadwad, GL Padmavathi, Shailesh Mantri

National Centre for Microbial Resource, National Centre for Cell Science, Pune, India

Abstract

Taxonomy of the genus *Aeromonas* has been controversial since its original description nearly 125 years ago. There is lack of congruence between 16S rRNA gene sequence similarity and DNA-DNA hybridization (DDH) values within *Aeromonas*. Single gene sequencing has only helped marginally and Multi Locus Sequence Typing (MLST) has therefore been proposed. However, there are at least four sets of MLST genes which have shown potential with other taxa that are currently in use for the genus *Aeromonas*. Because of lack of statistical and evolutionary evidence, results from these schemas have been incongruent, and there are several taxa within *Aeromonas* with pending inconsistencies. In order to resolve such inconsistencies, we use whole-genome sequences (WGS) to rationally arrive at a core-genome MLST for the genus *Aeromonas.* This was followed by narrowing down in an evidence-based manner to arrive at the least number of genes that have the same discriminating power as that of whole genome taxonomy. Our analysis yielded a set of 12 genes that were completely different from the existing MLST schemas. Of these, two genes, N-acetylglutamate synthase and transcriptional regulator ArgP, LysR family gene proved to have better statistics individually than any other existing genes used in *Aeromonas* MLST. Further analysis will allow us to test and validate these genes for their taxonomic resolution of inconsistent clusters within the genus *Aeromonas*.

Introduction

The Genus *Aeromonas:*

- Proposed by Kluyver and van Niel in 1936.
- A medically significant genus.
- Changing phylogenetic relationships.
- 35 distinct species are grouped into 14 phenospecies or 17 hybridization groups
- Need for more reliable and advanced molecular sleuths such as MLST analysis.

***Aeromonas* Taxonomy is STILL a BIG problem:**

- Lack of congruence.
- Limited strains for a few species.
- Genomically homogenous, but biochemically distinguishable HG1 to HG3.
- Correlation between different DNA Hybridization methods.

Methods

- Completely annotated WGS (Table 1)
- Core genome selection
- WG-MLST (Jironkin et al. 2016. BMC Genomics 17:964)
- Leave-one-out analysis of tree outputs:

$$D=S_t/S_{ref}$$

where, D is the similarity score between reference and target trees, S_{ref} is total number of nodes in a tree, and S_t is the total number of nodes such that children of these nodes are the same in reference and target tree.

- Statistical validation of genes

Table 1. Selected whole genomes of *Aeromonas* spp.

Aeromonas spp.	Accession No.
A. aquatica MX16A	NZ_CP018201
A. caviae 8LM	CP024198
A. dhakensis AAK1	NZ_CP023141
A. hydrophila subsp. *hydrophila* ATCC 7966	CP000462
A. media WS	NZ_CP007567
A. salmonicida subsp. *salmonicida* A449	CP000644
A. schubertii WL1483	NZ_CP013067
Aeromonas sp. CA23	NZ_CP023818
Aeromonas sp. CU5	NZ_CP023817
Aeromonas sp. O23A	NZ_CP021654
A. veronii B565	CP002607

Results

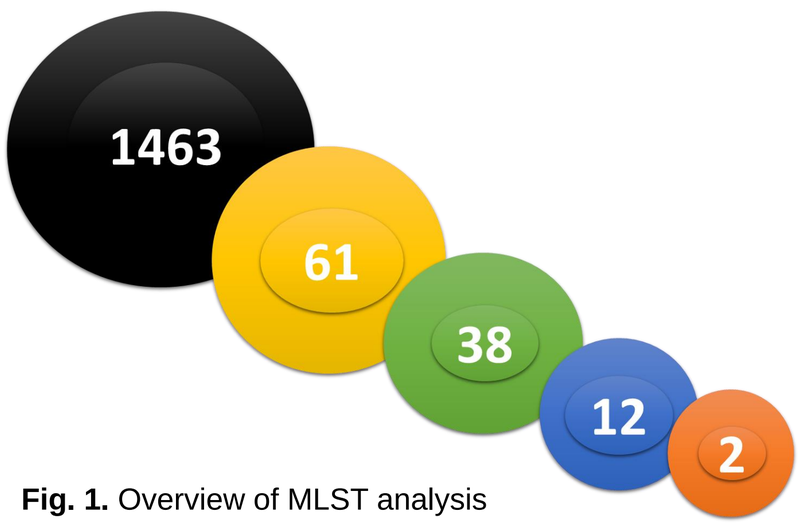

Fig. 1. Overview of MLST analysis

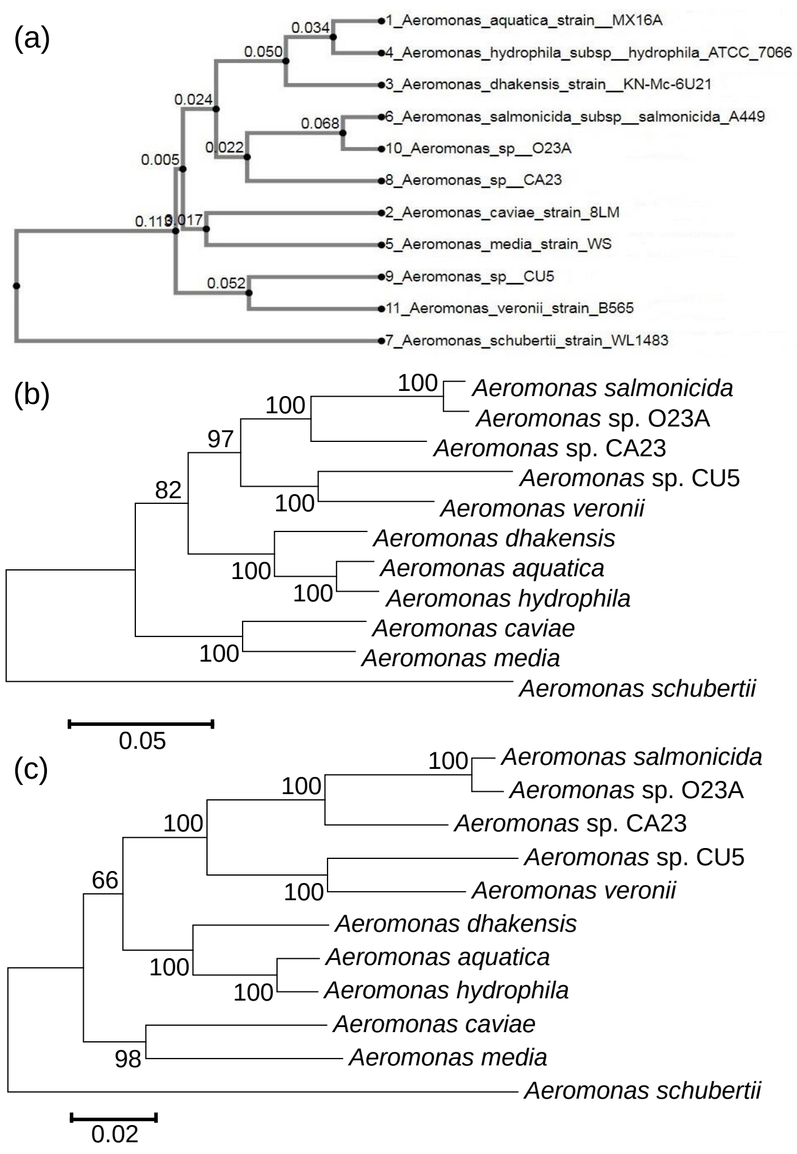

Fig. 2. *Aeromonas* MLST Trees prepared from (a) 1463 concatenated genes, (b) 16S rRNA genes, and (c) existing seven MLST genes (*atpD-dnaJ-gyrA-recA-rpoD-gyrB-dnaX*). The bootstrap consensus Maximum Likelihood trees were inferred from 1000 replicates.

List of 12 core genes for MLST of *Aeromonas*

1. (2E,6E)-farnesyl diphosphate synthase (EC 2.5.1.10)
2. Diaminopimelate decarboxylase (EC 4.1.1.20)
3. Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)
4. Dihydropteroate synthase (EC 2.5.1.15)
5. Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23)
6. N-acetylglutamate synthase (EC 2.3.1.1)
7. Outer-membrane-phospholipid-binding lipoprotein MlaA
8. Oxidoreductase, short-chain dehydrogenase/ reductase family
9. Ribose-phosphate pyrophosphokinase (EC 2.7.6.1)
10. Ribosomal large subunit pseudouridine synthase C (EC 5.4.99.24)
11. Transcriptional regulator ArgP, LysR family
12. tRNA-dihydrouridine(16) synthase

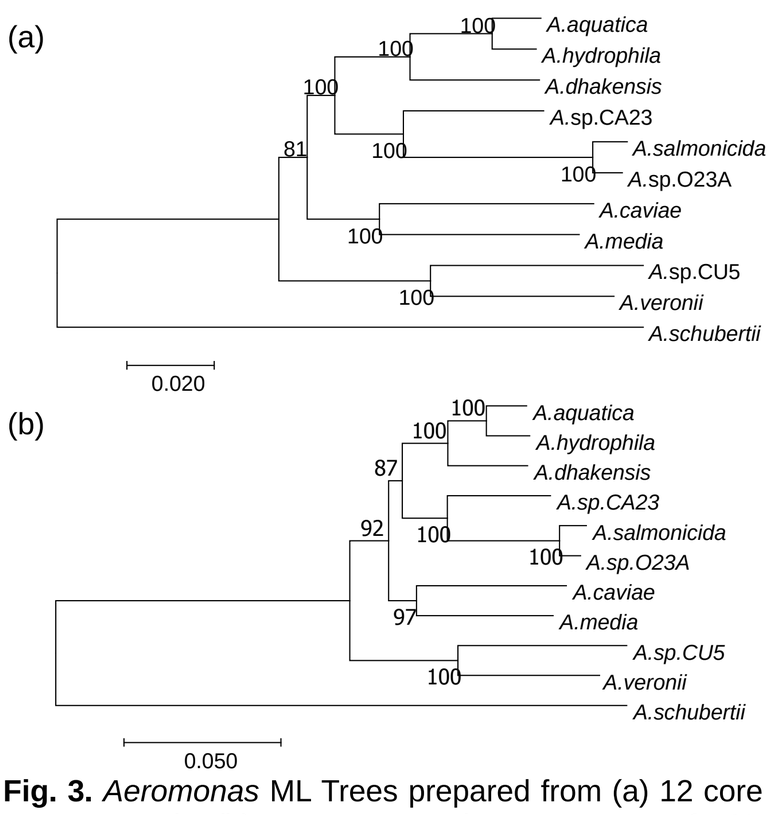

Fig. 3. *Aeromonas* ML Trees prepared from (a) 12 core genes, and (b) concatenated sequence of N-acetylglutamate synthase & Transcriptional regulator ArgP, LysR family genes. The bootstrap consensus Maximum Likelihood trees were inferred from 1000 replicates.

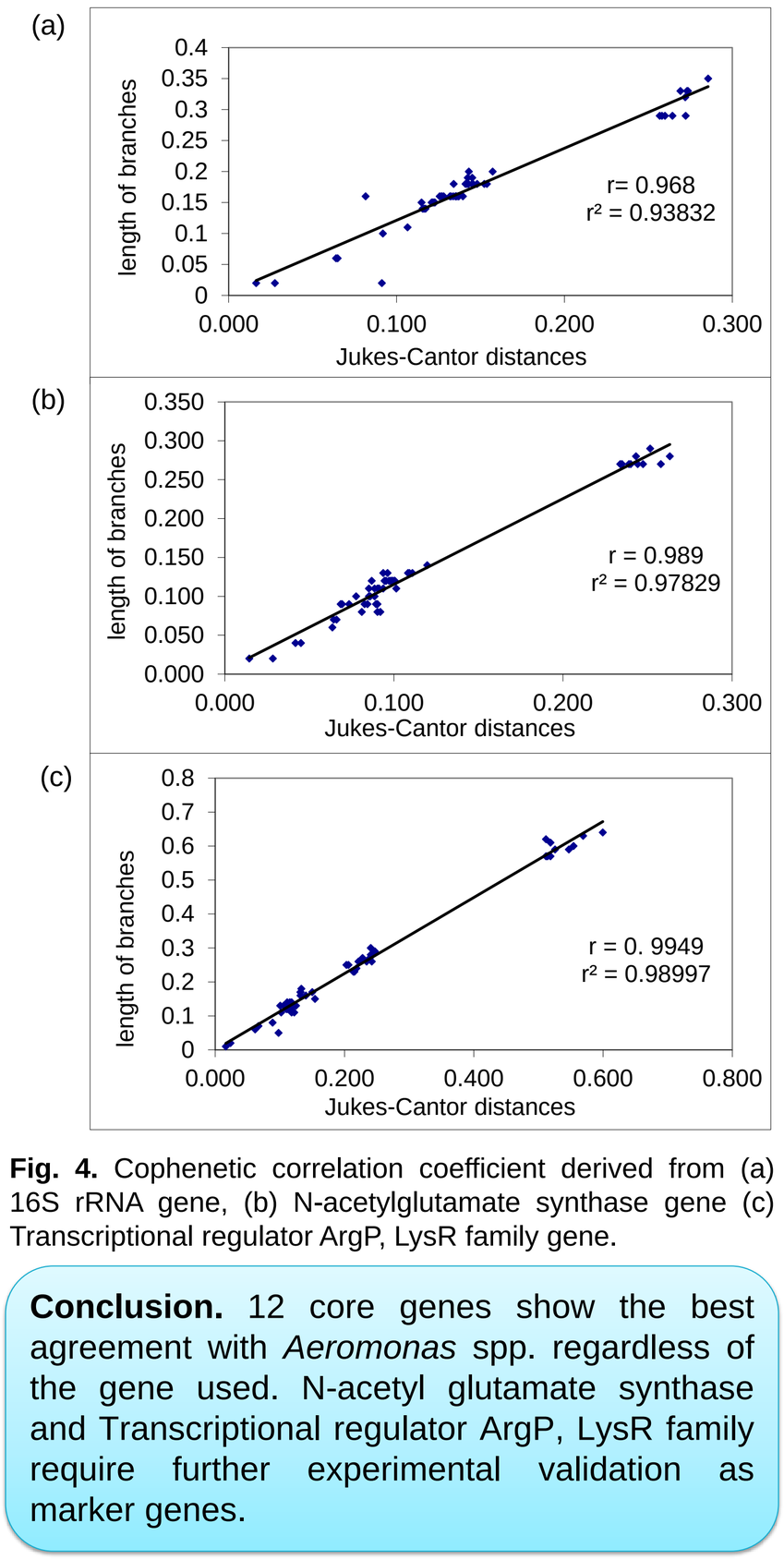

Fig. 4. Cophenetic correlation coefficient derived from (a) 16S rRNA gene, (b) N-acetylglutamate synthase gene (c) Transcriptional regulator ArgP, LysR family gene.

Conclusion. 12 core genes show the best agreement with *Aeromonas* spp. regardless of the gene used. N-acetyl glutamate synthase and Transcriptional regulator ArgP, LysR family require further experimental validation as marker genes.

Further communication

Kamlesh Jangid: jangidk@gmail.com

The authors acknowledge the funding provided by the following organizations for conducting and presenting this work at BISMiS 2018

DEPARTMENT OF BIOTECHNOLOGY
Ministry of Science & Technology

SERB DIA

NCMR राष्ट्रीय सूक्ष्मजीव संपदा केंद्र
NCCS

BISMiS