

**BPhyOG: An interactive server for
genome-wide inference of bacterial
phylogenies based on
overlapping genes**

Dong Chuan

- BPhyOG is an online interactive server for reconstructing the phylogenies of completely sequenced bacterial genomes on the basis of their shared overlapping genes. It provides two tree-reconstruction methods: Neighbor Joining (NJ) and Unweighted Pair-Group Method using Arithmetic averages (UPGMA). Users can apply the desired method to generate phylogenetic trees, which are based on an evolutionary distance matrix for the selected genomes.

- **Source of Dataset**
- **Methods**
- **Inferences of BPhyOG**
- **Discussions and Conclusion**

1.materials

- Bacterial genomes without plasmids were downloaded in the GenBank format from the NCBI ftp server in August 2004. A total of 79,855 overlapping gene pairs were extracted from the 177 acquired bacterial genomes using C/C++ scripts. The data are stored in a MySQL database.

- As we know, misannotated open reading frames (ORFs) may be included in the genome data. Therefore, to improve the accuracy of prediction, the author defined conserved homologous OG pairs as those for which the products of both overlapped genes are not annotated as "hypothetical" or "putative" or "unknown". However, the author still listed all the OG pairs of any given genome for other possible uses such as reannotation.

2. Methods

- 2.1. By searching bidirectional best hits with blast version 2.2.6 and applying thresholds of e-value < 10^{-4} and identity >40% to definition the orthologous overlapping genes
- 2.2 The distance between two genomes is defined as:

$$D_{ij} = 1 - \frac{x_{ij} + x_{ji}}{2 * \min(x_i, x_j)} \quad i, j = 1, 2, \dots, N$$

- where x_i is the number of OG pairs in genome i , N is the number of selected species and x_{ij} is the number of OG pairs in genome i with orthologs in genome j .
From this equation, an $N \times N$ distance matrix is produced, which indicates the evolutionary relationship among the selected N genomes based on the OGs

3. Inferences of BPhyOG

BPhyOG Bacterial Phylogenies Based on Overlapping Genes

Home | Phylogeny Inference | OGs in Genomes | Search by | Contact us

Search

Description:

- BPhyOG involves 177 prokaryotic genomes up to date (**Version: V-1.0**).
- See [documentation](#) for more information about BPhyOG.
- The article can be downloaded from [here](#).

3.1. Phylogenetic inference

- It provides two methods for reconstructing phylogenetic trees: NJ and UPGMA;

Inferring phylogeny for a set of genomes of interest

UPGMA <input checked="" type="radio"/>	NJ <input type="radio"/>	<input type="button" value="Submit"/>	<input type="button" value="Reset"/>
<input type="checkbox"/> <i>Acinetobacter</i> sp. ADP1		<input type="checkbox"/> <i>Aeropyrum pernix</i> K1	
<input type="checkbox"/> <i>Agrobacterium tumefaciens</i> str. C58		<input type="checkbox"/> <i>Aquifex aeolicus</i> VF5	
<input type="checkbox"/> <i>Archaeoglobus fulgidus</i> DSM 4304		<input type="checkbox"/> <i>Bacillus anthracis</i> A2012	
<input type="checkbox"/> <i>Bacillus anthracis</i> str. Ames		<input type="checkbox"/> <i>Bacillus anthracis</i> str. Ames Ancestor	
<input type="checkbox"/> <i>Bacillus anthracis</i> str. Sterne		<input type="checkbox"/> <i>Bacillus cereus</i> ATCC 10987	
<input type="checkbox"/> <i>Bacillus cereus</i> ATCC 14579		<input type="checkbox"/> <i>Bacillus halodurans</i> C-125	
<input type="checkbox"/> <i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		<input type="checkbox"/> <i>Bacillus thuringiensis</i> serovar <i>konkukian</i> str. 97-2	
<input type="checkbox"/> <i>Bacteroides thetaiotaomicron</i> VPI-5482		<input type="checkbox"/> <i>Bartonella henselae</i> str. Houston-1	
<input type="checkbox"/> <i>Bartonella quintana</i>		<input type="checkbox"/> <i>Bdellovibrio bacteriovorus</i> HD100	

3.2. OG pairs browse

- there are two ways for Users to browse all OG pairs in a genome:

Tid	Species name	GenBank accession number	Coding sequence size (bp)	No. of ORFs	Genome sequence size (bp)	No. of overlapping pairs
62977	Acinetobacter sp. ADP1	NC_005966	3163860	3325	3598620	317
272557	Aeropyrum pernix K1	NC_000854	1553222	1841	1669680	514
176299	Agrobacterium tumefaciens str. C58	NC_003062	2539614	2721	2841540	1583

- (1) . through hyperlinks by clicking the "OGs in Genomes" box displays a summary of the 177 bacterial genomes. This page gives basic features of the genomes including the species name, GenBank accession number, whole genome sequence size, coding sequence size, number of ORFs, and number of OG pairs

3.3. Searching interface

- Searching OGs by gene name and/or GenBank accession number, or by species name

Searching OGs by gene name and/or GenBank accession number

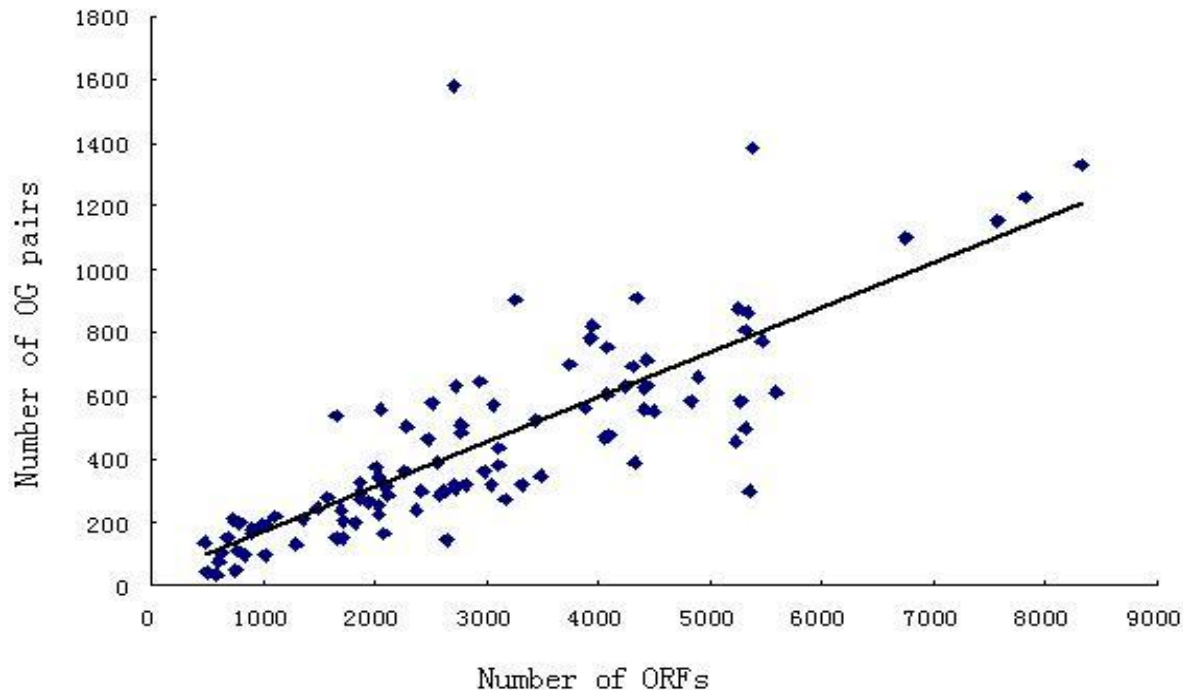
Gene name (e. g. ksgA):

GenBank accession number (e. g. NC_004431):

Submit

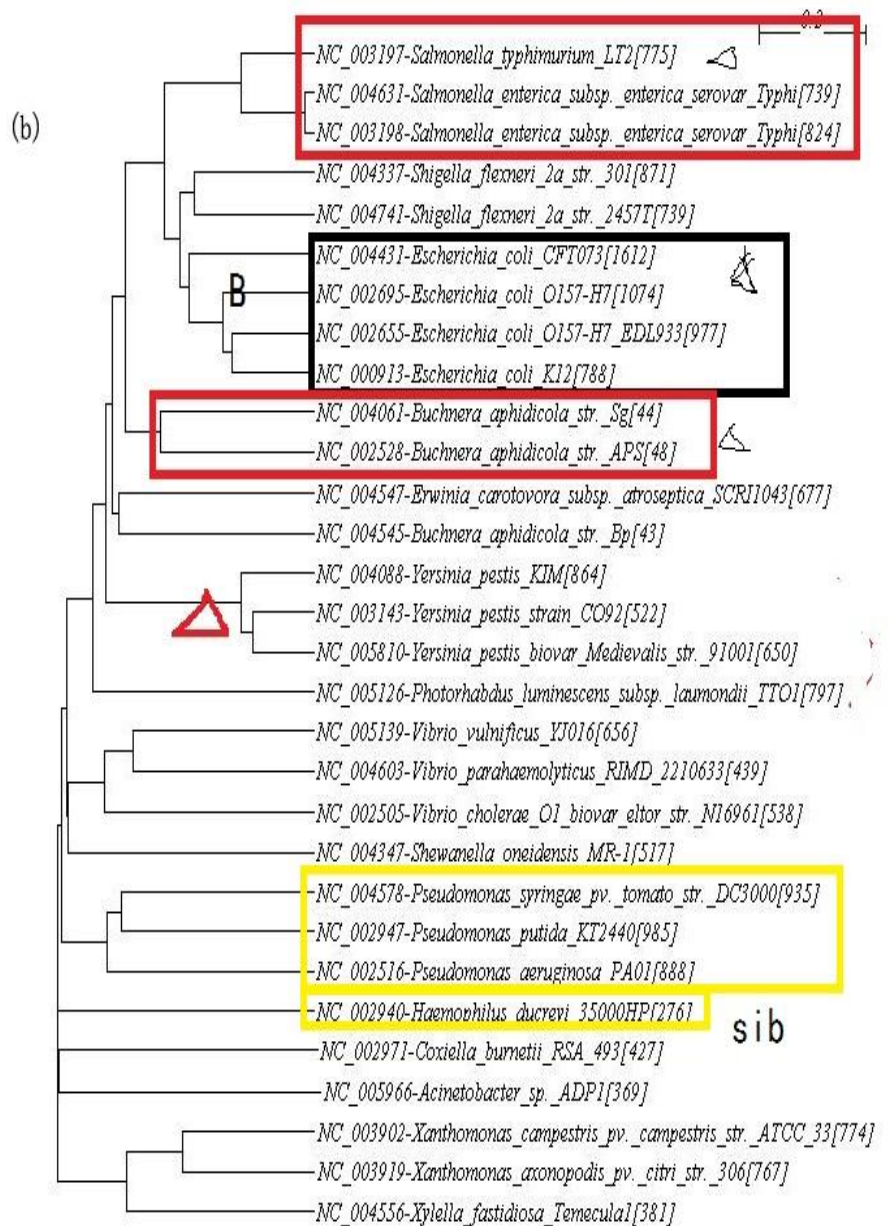
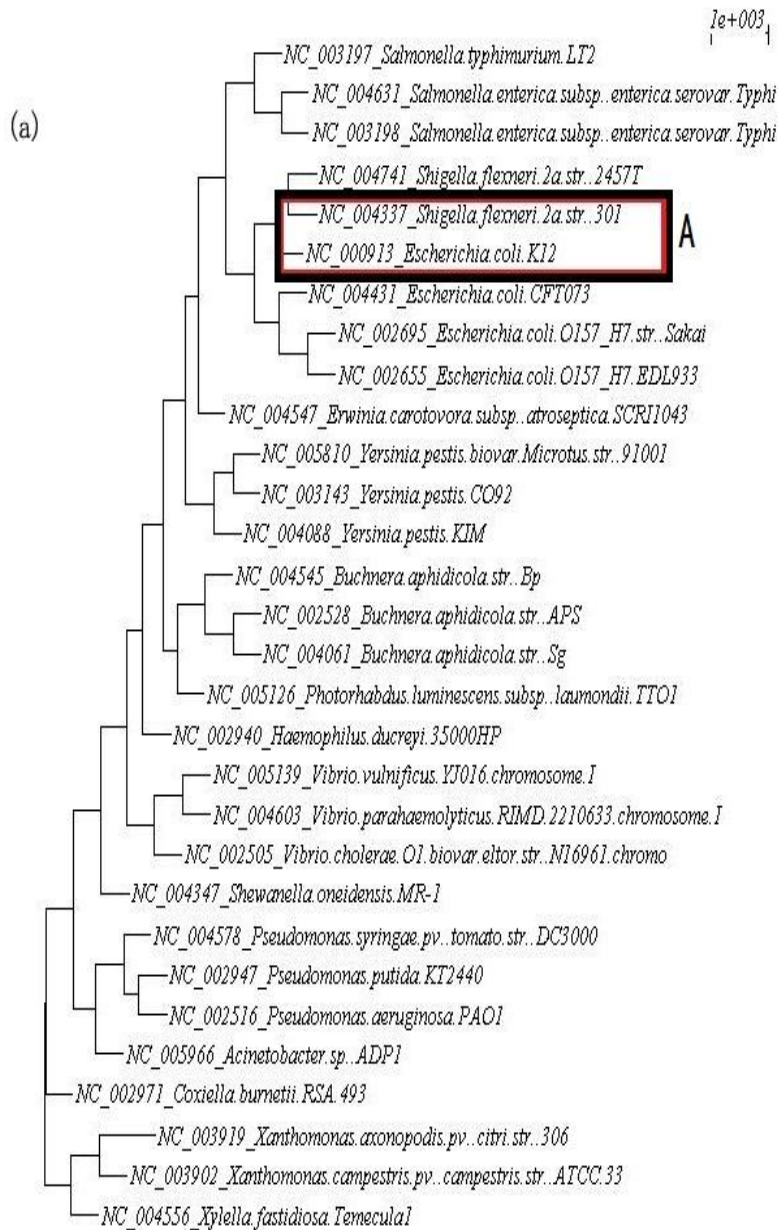
Reset

4. Dissuactions and main conclusions



The correlation between all ORFs and OG pairs in 177 genomes. As expected, the number of OG pairs in each genome is significantly correlated with its total number of ORFs (Pearson's correlation coefficient is 0.668(The same results in Ref1).

- 4.1. The author selected 30 gamma-proteobacteria genomes and constructed their phylogenetic tree as showing the below pictures using 16S rRNA sequences and OGS(UPGMA methods)



- BPhyOG also provides an easy-to-use interface for users to track the conservation of OG pairs by exploring homologous pairs in subtree-specific genomes.
- **Note:** a lot of reduction (gene loss) and gene transfer has probably occurred within these genomes during evolutionary history. Thus, the small numbers of overlapping genes that occur in these two genomes might affect the accuracy of their positions in our inferred phylogenies when compared with other genomes.

Thank you