Taxonomic Study of Campylobacter Species using *hsp60* Protein

Mohammed Z. Saeed Al-khayyat

Department of Biology
College of Education
University of Mosul
Mosul City, Iraq
E-mail: mzalkhayat@yahoo.com

Received: December 30, 2012 **Accepted: April 5, 2013**

Published: April 15, 2013

Abstract: *hsp60 protein sequences of twenty two Campylobacter species were downloaded from NCBI (National center for biotechnology information). Alignments and phylogenetic tree were constructed using ClustalW2, and the protein sequences showed 100% similarities between C. curvus and C. concisus, and between C. lari and C. fetus subspecies. The phylogenetic tree showed that C. rectus and C. sputorum are of the same line of origin and the same applies to C. peloridis and C. volucris. Amino acid substitutions between C. helveticus and C. upsaliensis were in the positions 98 and 149 where N replaced S respectively. Only one amino acid substituted in the position 38 where N replaced A between C. peloridis and C. volucris while six substitutions between C. curvus and C. concisus in the positions 94, 98, 118, 150, 153 and 160 so E, N, A, I, V and T and V replaced Q, S, P, V, I and S respectively. Seventeen amino acid substitutions between C. rectus and C.sputorum in the amino acid positions 6, 23, 32, 34, 42, 44, 47, 66, 68, 70, 94, 118, 122, 126, 127, 150 and 153 that made the two species being divergent from each other.*

Keywords: *Campylobacter, hsp60, Phylogenetic tree, Alignments, ClaustalW2.*

Introduction

The genus *Campylobacter* comprises a taxonomically diverse group of Gram-negative microaerophilic and /or anaerobic bacteria that may be found in a variety of habitats. At present, there are sixteen species and six subspecies, and several species are associated with a range of diseases in humans and animals including diarrhea, abortion, infertility, septicemia and peritonitis. Both the identification and subtyping of *Campylobacter* species are known to be problematic. The taxonomic diversity and relative biochemical inactivity of these organisms has consistently made accurate species and defined subspecies level identification difficult [1].

The science of taxonomy comprises three principal areas: classification, identification and nomenclature, each linked together. Strains are ordered or classified into groups on the basis of some common features and traits allowing the group to be identified i.e. discriminated from similar taxa are defined. The genus *Campylobacter* taxonomic structure has changed extensively since its inception and the use of increasingly sophisticated molecular methods has revealed fascinating aspects of *Campylobacter* biodiversity. Nevertheless, certain proposals have proven controversial [2] e.g. *C. jejuni* and *C. coli* seems to lack polymorphisms in their 16S rRNA gene and phylogenetic analysis based on 16S rRNA sequences was not always sufficient for differentiation between them [3].

Gene content contains a strong phylogenetic signal and has helped to clarify several taxonomic uncertainties, if a complete genome cannot obtained, gene content can still be used

to address taxonomical questions by means of signature genes. The presence or absence of genes that are stable in evolution provides phylogenetic evidence that complements sequence-based information because gene content evolves at different levels (whole genes instead of residues) and signature genes specifically exploit those genes that do not have a very wide phylogenetic distribution [4] e.g. *hsp60, hsp65, hsp70, soda, recA, rpoB, rpoC, gyrA* and *gyrB* genes were used for that purpose [5-6].

The 60 KD heat shock proteins are ubiquitous abundant proteins of eubacterial genomes [7] *hsp60*, also known as *cpn60* or *groEL*, is a member of the heat-shock protein (Hsp) family. The *hsp60* gene has been shown to be more discriminative than the 16S rRNA gene for the identification of *Streptococcus suis* serotypes [8] and some *Prevotella* species [9]. The potential for the use of the *cpnDB* has been demonstrated by its use as a tool for the identification of *Campylobacter* spp. and their distinction from phenotypically similar *Helicobacter* and *Arcobacter* spp. [10]. More recently, the *hsp60* gene has been found to be an alternative phylogenetic marker for the classification of species of the genus *Bacteroides* [11].

Materials and methods

hsp60 protein sequences of twenty two Campylobacter species were downloaded from NCBI available at: http://www.ncbi.nlm.nih.gov/. Table 1 shows the strains with the accession number of the *hsp60* sequences used.

Table 1. Species of *Campylobacter* spp.
used in this study with their accession number

Strain No.	Strain	Accession No.
1	*Campylobacter jejuni* subsp. *jejuni*	AAX19046.1
2	*Campylobacter coli*	AAZ94813.1
3	*Campylobacter lari* subsp. *lari*	AAZ94811.1
4	*Campylobacter lari* subsp. *concheus*	CAP58760.1
5	*Campylobacter fetus* subsp. *fetus*	AAZ94791.1
6	*Campylobacter fetus* subsp. *venerealis*	AAZ94779.1
7	*Campylobacter curvus*	AAZ94793.1
8	*Campylobacter gracilis*	AAZ94803.1
9	*Campylobacter rectus*	AAZ94764.1
10	*Campylobacter concisus*	AAZ94801.1
11	*Campylobacter hyointestinalis*	AAZ94762.1
12	*Campylobacter mucosalis*	AAZ94792.1
13	*Campylobacter helveticus*	AAZ94809.1
14	*Campylobacter upsaliensis*	AAZ94808.1
15	*Campylobacter sputorum*	AAZ94795.1
16	*Campylobacter peloridis*	CAP58758.1
17	*Campylobacter subantarcticus*	CAP69868.1
18	*Campylobacter volucris*	CAT12841.1
19	*Campylobacter canadensis*	ABS17622.1
20	*Campylobacter insulaenigrae*	CAP58757.1
21	*Campylobacter ureolyticus*	AAX24021.1
22	*Campylobacter troglodytis*	CAZ65706.1

hsp60 protein of *C. jejuni* subsp. jejuni was subjected to basic local alignment of NCBI-BLAST available at: http://blast.ncbi.nlm.nih.gov/Blast.cgi to identify the related proteins.

Alignments were constructed using ClustalW2 (multiple sequence alignment (EBI, United Kingdom) available at: http://www.ebi.ac.uk/Tools/msa/clustalw2/. Phylogenetic tree of the species was drawn using ClaustalW2.

Results and discussion

hsp60 of *C. jejuni* subsp. *jejuni* was analyzed using PSI-BLAST of Blastp (where PSI-BLAST is the most sensitive program to find the very distantly related proteins) with the default parameters of e-value = 10, inclusion threshold = 0.005, maximum number of alignments = 1000 to retrieve possibly all statistically significant alignments of proteins, the sequences of all species and subspecies of *Campylobacter* understudy showed significant alignments (or hits) with e-values better than threshold with similarities of 88%-100%.

Gupta [5], studied molecular signatures of Campylobacterales, indicated that a protein was considered to be epsilon-proteobacteria specific i.e. belonging to Campylobacterales if all significant hits in PSI-BLAST with the query protein were from epsilon-proteobacteria species. In few cases other species also exhibited borderline significance, but there was a large jump or increase of e-values.

Karenlampi et al. [12] analyzed phylogenetic relationships of twelve *Campylobacter* species using *groEL* gene sequences and the phylogenetic tree was similar to the tree based on 16S rRNA, however *groEL* was found to provide a better resolution for *Campylobacter* species with lower interspecies similarities ranged between 65%-94% compared with those for 16S rRNA that ranged between 90%-99% and high intraspecies sequence similarities ranged between 95%-100%, so *C. jejuni* NCTC11168 has a similarity of 91%, 98%, 92% and 96% with *C. coli* CCUG11283, *C. lari* CCUG23949, *C. rectus* ATCC33238 and *C. upsaliensis* CCUG14913 respectively using 16S rRNA phylogenetic analysis (Table 2).

In addition, Korczak et al. [13] used *rpoB* gene in phylogenetic analysis and compared the results with 16S rRNA for a total of 59 strains, the branching tree was similar to that of 16S rRNA and the resolution provided by *rpoB* gene was generally higher resulting in good separation of most species and even some subspecies. The tree showed a close relationship of *C. rectus* with *C. showae* and between *C. curvus* and *C. gracilis* while *C. jejuni* and *C. coli* shared the same phylogenetic origin and were at the same level with *C. lari* and *C. upaliensis*.

Amino acid substitutions (Fig. 2) between *C. helveticus* and *C. upsaliensis* were in the positions 98 and 149 where N replaced S respectively. Only one amino acid substitution in the position 38 where N replaced A between *C. peloridis* and *C. volucris* while six substitutions between *C. curvus* and *C. concisus* in the positions 94, 98, 118, 150, 153 and 160 so E, N, A, I, V and T and V replaced Q, S, P, V, I and S respectively. Seventeen amino acid substitutions between *C. rectus* and *C. sputorum* in the amino acid positions 6, 23, 32, 34, 42, 44,47, 66, 68, 70, 94, 118, 122, 126,127,150 and 153 where H, V, Q, A, L, R, T, S, I, G, Q, A, Q, Q, S, I and T replaced L, I, V, E, M, K, K, E, V, S, N, S, T, S, N, V and S which made the two species being divergent from each other (Fig. 1 and Fig. 2).

Table 2. Similarities among *Campylobacter* species

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1		98	96	96	89	89	88	87	91	85	89	88	96	95	91	97	97	97	84	96	90	93
2			97	97	88	88	88	86	90	85	88	87	96	95	91	96	98	96	84	96	89	93
3				100	87	87	86	84	89	84	85	87	94	93	89	95	98	96	82	97	87	91
4					87	87	86	84	89	84	85	87	94	93	89	95	98	96	82	97	87	91
5						100	90	89	88	87	94	89	87	86	87	88	87	88	80	88	85	88
6							90	89	88	87	94	89	87	86	87	88	87	88	80	88	85	88
7								88	88	96	88	96	87	86	87	86	87	86	78	86	87	87
8									87	88	89	87	86	85	88	86	85	86	79	85	88	84
9										85	88	88	89	88	90	90	89	91	78	89	87	88
10											85	94	85	85	88	84	85	84	78	84	88	85
11												87	87	86	88	88	87	87	79	86	87	88
12													87	85	86	87	87	87	78	88	86	85
13														98	89	96	95	96	84	94	87	94
14															88	95	94	95	84	92	87	95
15																88	90	89	81	89	89	88
16																	95	99	83	96	88	92
17																		95	83	96	88	92
18																			83	96	88	92
19																				80	80	85
20																					88	91
21																						88

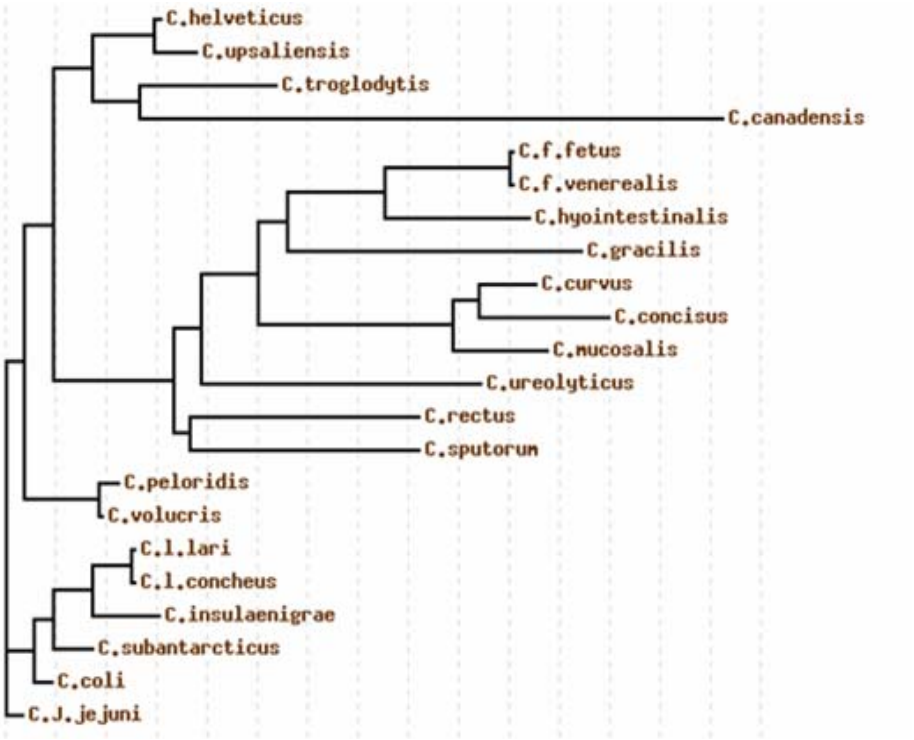

Fig. 1 Phylogenetic tree using ClaustalW2

Anita et al. [14] analyzed the phylogenetic relationships among 36 species and subspecies of the genus *Staphylococcus* using *hsp60* gene sequences, their results showed that using *hsp60* may offer advantages over both DNA-DNA hybridization and 16 S RNA sequencing in defining the taxonomy and phylogenetic relationships within that genus.

Jiang et al. [15] studied the taxonomy of *Sorangium* strains using *hsp60* and 16 rRNA also indicated that the phylogeny constructed by *groEL* gene was rather consistent with morphological characteristics including information from the phylogenetic analysis.

Moreover, Blaiotta et al. [16] on analyzing *Lactobacillus* strains diversity based on partial *hsp60* gene sequences indicated that *hsp60* could be considered an excellent molecular marker for inferring the taxonomy and phylogeny of the members of the genus.

More recently, gene targets providing more discriminating information, such as *omp50* and *cpn60* (also known as the 60 KD chaperonin, *groEL* or *hsp60*), have been found to provide the necessary inter-species discrimination within the genus Campylobacter [10].

Conclusion

The 60 KD heat shock proteins are *hsp60*, also known as Cpn60 or *groEL*, has been shown to be more discriminative than the 16S rRNA.

```
                1
60
C.helveticus    ATVLAHAIFKEGLRNITAGANPIEVKRGMDKACEAIVDELKKLSREVKD-KKEIAQVATI
C.upsaliensis   ATVLAHAIFKEGLRNITAGANPIEVKRGMDKACEAIVDELKKLSREVKD-KKEIAQVATI

C.curvus        ATVLAHAIFKEGLRNVTAGANPIEVKRGMDKEVAALIDELKNISKKVSG-SKEIAQIATI
C.concisus      ATVLAHAIFKEGLRNVTAGANPIEVKRGMDKEVAALIDALKNISKKVSG-SKEIAQIATI

C.rectus        ATVLAHSIFKEGLRNITAGANPVEVKRGMDKQAAAIIAELKNLSRKVTD-KKEIAQVATI
C.sputorum      ATVLALSIFKEGLRNITAGANPIEVKRGMDKVAEAIIAELKNMSKKVKD-KKEIAQVATI

C.peloridis     ATVLAHAIFKEGLRNITAGANPIEVKRGMDKACEAIVNELKKLSREVKD-KKEIAQVATI
C.volucris      ATVLAHAIFKEGLRNITAGANPIEVKRGMDKACEAIVAELKKLSREVKD-KKEIAQVATI

               61
120
C.helveticus    SANSDEKIGALIADAMERVGKDGVITVEEAKSINDELNVVEGMQFDRGYLSPYFITNADK
C.upsaliensis   SANSDEKIGALIADAMERVGKDGVITVEEAKSINDELSVVEGMQFDRGYLSPYFITNADK

C.curvus        SANSDESIGKLIADAMEKVGKDGVITVEEAKSIEDELNVVEGMQFDRGYLSPYFITNAEK
C.concisus      SANSDESIGKLIADAMEKVGKDGVITVEEAKSIQDELSVVEGMQFDRGYLSPYFITNPEK

C.rectus        SANSDSAIGGLIADAMEKVGKDGVITVEEAKSIQDELNVVEGMQFDRGYLSPYFITNAEK
C.sputorum      SANSDEAVGSLIADAMEKVGKDGVITVEEAKSINDELNVVEGMQFDRGYLSPYFITNSEK

C.peloridis     SANSDEKIGNLIADAMEKVGKDGVITVEEAKSINDELNVVEGMQFDRGYLSPYFITNADK
C.volucris      SANSDEKIGNLIADAMEKVGKDGVITVEEAKSINDELNVVEGMQFDRGYLSPYFITNADK

               121
180
C.helveticus    MIVELSNPYILLFDKKITNLKDLLPILEQIQKTGKPLLIIAEDIEGEALATLVVNKLRGV
C.upsaliensis   MIVELSNPYILLFDKKITSLKDLLPILEQIQKTGKPLLIIAEDIEGEALATLVVNKLRGV

C.curvus        MQVELSNPFILLFDKKITNLKDLLPVLEQIQKTGKPLLIVAEDIEGEALATLVVNKLRGV
C.concisus      MQVELSNPFILLFDKKITNLKDLLPVLEQVQKSGKPLLIIAEDIEGEALATLVVNKLRGV

C.rectus        MQVELQSPYILLFDKKITNLKDLLPVLEQIQKTGKPLLIIAEDIEGEALATLVVNKLRGV
C.sputorum      MTVELSNPYILLFDKKITNLKDLLPVLEQVQKSGKPLLIIAEDIEGEALATLVVNKLRGV

C.peloridis     MLVELQSPYILLFDKKITNLKDLLPILEQIQKTGKPLLIIAEDIEGEALATLVVNKLRGV
C.volucris      MLVELQSPYILLFDKKITNLKDLLPILEQIQKTGKPLLIIAEDIEGEALATLVVNKLRGV

              181
C.helveticus    LNISAV
C.upsaliensis   LNISAV

C.curvus        LNISAV
C.concisus      LNISAV

C.rectus        LNISAV
C.sputorum      LNISAV

C.peloridis     LNISAV
C.volucris      LNISAV
```

Fig. 2 Sequences of *hsp60* showing amino acid substitutions (red)

References

1. On S. L. W., C. S. Harrington (2000). Identification of Taxonomic and Epidemiological Relationships Among *Campylobacter* Species by Numerical Analysis of AFLP Profiles, FEMS Microbiol Letters, 193,161-169.
2. On S. L. W. (2001). Taxonomy of *Campylobacter*, *Arcobacter*, *Helicobacter* and Related Bacteria: Current Status, Future Prospects and Immediate Concerns, J Appl Microbiol, 90, 1-15.
3. Hansson I., M. Persson, L. Svensson, E. O. Engvall, K. E. Johansson (2008). Identification of Nine Sequence Types of 16S rRNA Genes of *Campylobacter jejuni* subsp. *jejuni* Isolated from Broilers, Acta Vet Scaninavica, 50(10), 1-10.
4. Dutilh B. E, B. Snel, T. J. G. Ettema, M. A. Huynen (2008). Signature Genes as Phylogenetic Tool, Mol Biol Evol, 25(8), 1659-1667.
5. Gupta R. S. (2006). Molecular Signatures (Unique Proteins and Conserved Indels that are Specific for the Epsilon Proteobacteria (Campylobacterales), BMC Genomics, 7, 167.
6. Adekambi T., M. Drancourt (2004). Dissection of Phylogenetic Relationships Among 19 Rapidly Growing *Mycobacterium* Species by 16S rRNA, *hsp65*, *soda*, *recA* and *rpoB* Gene Sequencing, Int J Sys Evol Microbiol, 54, 2095-2105.
7. Brocchieri L., S. Karlin (2000). Conservation Among *hsp60* Sequences in Relation to Structure, Function and Evolution, Protein Science, 9, 476-489.
8. Brousseau R., J. E. Hill, G. Prefontaine, S. H. Goh, J. Harel, S. M. Hemmingsen (2001). *Streptococcus suis* Serotypes Characterized by Analysis of Chaperonin 60 Gene Sequences, Appl Environ Microbiol, 67, 4828-4833.
9. Sakamoto M., N. Suzuki, M. Okamoto (2010). *Prevotella aurantiaca* sp. nov., Isolated from the Human Oral Cavity, Int J Syst Evol Microbiol, 60, 500-503.
10. Hill J. E., A. Paccagnella, K. Law, P. L. Melito, D. L. Woodward, L. Price, A. H. Leung, L. Ng, S. M. Hemmingsen, S. H. Goh (2006). Identification of *Campylobacter* spp. and Discrimination from *Helicobacter* and *Arcobacter* spp. by Direct Sequencing of PCR-Amplified cpn60 Sequences and Comparison to cpnDB, a Chaperonin Reference Sequence Database, J Med Microbiol, 55, 393-399.
11. Sakamoto M., N. Suzuki, Y. Benno (2010). *hsp60* and 16S rRNA Gene Sequence Relationships Among Species of the Genus *Bacteroides* with the Finding that *Bacteroides suis* and *Bacteroides tectus* are Heterotypic Synonyms of *Bacteroides pyogenes*, Int J Syst Evol Microbiol (in press).
12. Karenlampi R. I., T. P. Tolvanen, M.-L. Hanninen (2004). Phylogenetic Analysis and PCR-restriction Fragment Length Polymorphism Identification of *Campylobacter* Species Based on Partial *groEL* Gene Sequences, J Clin Microbiol, 24(12), 5731-5738.
13. Korczak B. M., R. Stieber, E. Stefan, A. P. Burnens, J. Frey, P. Kuhnert (2004). Genetic Relatedness within the Genus *Campylobacter* Inferred from *rpoB* Sequences, Int J Sys Evol Microbiol, 56, 937-945.
14. Anita Y, C. Kwok, S. Su, R. Reynolds, S. Bay, Y. Av-Gay, N. J. Dovichi, A.W. Chow (1999). Species Identification and Phylogenetic Relationships Based on Partial *hsp60* Gene Sequences within the Genus *Staphylococcus*, Int J Syst Bacteriol, 49, 1181-1192.
15. Jiang D., L. Zhao, C. Zhang, J. Li, Z. Xia, J. Wang, Z. Wu, Y. Li (2008). Taxonomic Analysis of *Sorangium* Strains Based on *hsp60* and 16 sRNA Gene Sequences and Morphology, Int J Syst Evol Microbiol, 58, 2654-2659.
16. Blaiotta G., V. Fusco, D. Ercolini, M. Aponte, O. Pepe, F. Villani (2008). Lactobacillus Strain Diversity Based on Partial *hsp60* Gene Sequences and Design of PCR-Restriction Fragment Length Polymorphism Assays for Species Identification and Differentiation, Appl Environ Microbiol, 74(1), 208-215.

Mohammed Z. Saeed Al-khayyat
E-mail: mzalkhayat@yahoo.com

M. Z. Saeed Al-khayyat was born in Baghdad, Iraq on 19 August 1972. He graduated from the University of Mosul, College of Science, Mosul, Iraq. He is a Member of the American Society of Microbiology and Member of British society of Cell biology. At present he is with University of Mosul, College of Science, Biology Department: teaching practical lab genetics. His current researches are: Compartive study of Urease genes in *Klebsiella pneumoniae* and *Proteus mirabilis* (accepted in Science and education Journal) and Codon usage and sequence analysis of *Shigella flexneri* genes (accepted in Al-Rafiddin Journal of Science).