*Database*, 2019, 1–7
doi: 10.1093/database/baz033
Database update

Database update

# Update on cpnDB: a reference database of chaperonin sequences

**Sarah J. Vancuren and Janet E. Hill***

Department of Veterinary Microbiology, University of Saskatchewan, 52 Campus Drive, Saskatoon SK, Canada S7N 5B4

*Corresponding author: Tel: (306) 966-7242; Fax: (306) 966-7244; Email: Janet.Hill@usask.ca

Citation details: Vancuren,S.J. and Hill,J.E. Update on cpnDB: a reference database of chaperonin sequences. *Database* (2019) Vol. 2019: article ID baz033; doi:10.1093/database/baz033

Received 22 January 2019; Revised 14 February 2019; Accepted 16 February 2019

## Abstract

cpnDB was established in 2004 to provide a manually curated database of type I (60 kDa chaperonin, CPN60, also known as GroEL or HSP60) and type II (CCT, TRiC, thermosome) chaperonin sequences and to support chaperonin sequence-based applications including microbial species identification, detection and quantification, phylogenetic investigations and microbial community profiling. Since its establishment, cpnDB has grown to over 25 000 sequence records including over 4 000 records from bacterial type strains. The updated cpnDB webpage (www.cpndb.ca) provides tools for text- or sequence-based searches and links to protocols, and selected reference data sets are available for download. Here we present an updated description of the contents and taxonomic coverage of cpnDB and an analysis of *cpn*60 sequence diversity.

**Database URL:** http://www.cpndb.ca/

## Introduction

Chaperonins are a diverse group of molecular chaperones found in virtually all prokaryotes and eukaryotes (1). The type I chaperonins (60 kDa chaperonin, CPN60, also known as GroEL or HSP60) are found in bacteria and a few archaea and in the mitochondria and chloroplasts of eukaryotes. Type II chaperonins include the eukaryote cytoplasmic CCT (for chaperonin containing TCP1, also called TRiC for TCP1 ring complex) and the archaeal homolog, thermosome. Their conservation across all domains of life makes the genes encoding chaperonin proteins attractive targets for use in phylogenetic investigations.

The discovery that a 549–567 bp region of the cpn60 gene termed the 'universal target' (UT; corresponding to nucleotides 271–825 of the *Escherichia coli* cpn60 gene) could be polymerase chain reaction (PCR) amplified with degenerate primers provided the foundation for development of numerous applications exploiting cpn60 sequences (2). Since that time, cpn60 UT sequences have been used as evidence for definition of new bacterial species (3, 4), provided targets for hybridization, PCR and sequencing-based diagnostics for bacteria (5–11) and for amplicon-based profiling of complex microbial communities (12–16). In addition, cpn60 UT sequences have proven to be excellent indicators of whole genome sequence similarities among bacteria (17–19). These applications are made possible by the sequence diversity of the UT region, especially among closely related taxa. Using criteria established by the

International Barcode of Life project (20), the cpn60 UT has been demonstrated to be a preferred barcode for bacteria, compared to the 16S rRNA gene (21). The relatively large 'barcode gap' between inter- and intra-specific distances facilitates resolution of taxa, often at subspecies levels (10, 22–24). More recently, 'universal' primers for amplification of partial thermosome gene sequences were developed for archaea. As with the bacterial type I chaperonin, evidence to date suggests that these type II chaperonin sequences provide higher resolution than 16S rRNA gene sequences (25).

The development of chaperonin sequence-based methods inspired the original development of a reference database of chaperonin sequences. cpnDB was released in 2004 (26) and has been continuously maintained and updated since then. The database has been hosted at the University of Saskatchewan (Saskatoon, Canada) since 2015. cpnDB was designed to provide users with the ability to query records by sequence similarity search, or based on keywords, and to allow users to download sequences of interest for offline analyses. At its inception, cpnDB contained approximately 2000 type I and archaeal type II chaperonin sequence records representing 240 genera and has now grown to more than 25000 records (including eukaryotic type II chaperonin sequences) representing more than 1800 genera. cpnDB records are manually curated to ensure high quality entries. The original goal was to include all published type I and II chaperonin sequences, but since whole genome sequencing became a common activity and the volume of new sequence data increased exponentially, the emphasis has shifted to providing high quality records with broad taxonomic coverage, and with a particular emphasis on inclusion of type strains since they are the most useful landmarks for microbial species identification. A complete list of citations of cpnDB is maintained on the database webpage (http://www.cpndb.ca/publications.php).

The purpose of this update is to describe changes to the form and content of cpnDB since its original publication (26) and to provide a description of the taxonomic and sequence diversity represented in the database.

## Database access, structure and web interface

cpnDB is freely available to users through the web interface (http://www.cpndb.ca). There is no requirement for creation of an account, and no password is needed to access the database.

cpnDB was constructed with MySQL and the web interface is implemented with PHP. The web interface of cpnDB supports searching by text using terms such as genus and species names, culture collection catalog numbers or hierarchical taxonomy terms from the National Center for Biotechnology Information (NCBI) Taxonomy database (https://www.ncbi.nlm.nih.gov/guide/taxonomy/). Sequence searches can be done using BLASTP, primer blast (27) or FASTA (28). Users can choose to limit their search to type I (UT or full-length sequences) or type II chaperonin sequences. Individual records or groups of records for download can be selected from search results. The homepage provides information about chaperonins and links to publications citing cpnDB to provide information about applications of chaperonin sequences (Figure 1).

cpnDB records are assigned a unique identifier (Chaperonin ID). Each record contains the full-length chaperonin gene and protein sequences when available and the corresponding UT regions for type I sequences (Figure 2). Records also contain the taxonomic lineage of the source organism based on the NCBI Taxonomy ID. Strain name synonyms are provided for bacterial type strain records, using information extracted from the List of Prokaryotic Names with Standing in Nomenclature (http://www.bacterio.net) or relevant culture collection catalogs. Nucleotide and peptide Genbank accession numbers are provided, with external links. cpnDB is a LinkOut provider (https://www.ncbi.nlm.nih.gov/projects/linkout/), and so reciprocal links from Genbank records to the corresponding entry in cpnDB can be accessed from relevant Genbank records. cpnDB records for sequences generated from microbial community studies or clinical specimens include information about their closest match in the reference database (updated with each new addition).

BLAST and FASTA databases are updated with each addition of new records to cpnDB, and the entire sequence collection can be downloaded in FASTA format. A non-redundant version of the database, cpnDB_nr, is similarly updated with each new addition. This subset of the database is limited to type I chaperonins (bacteria, archaea and eukaryotes) and includes only a single representative sequence from each species, with priority given to the type strain when available. Sequence-based queries of cpnDB_nr may be desirable when a broad view of the relationship of the query to chaperonin sequences from other taxa is wanted, since results will include more different species rather than multiple strains of one or a few closely related species. cpnDB_nr can also be downloaded (nucleotide or peptide UT sequences) for use in offline analyses, and versions are identified on the download page by the creation date.

## Sources of data and curation methods

All records in cpnDB are manually curated to ensure they contain complete sequences with no ambiguous positions

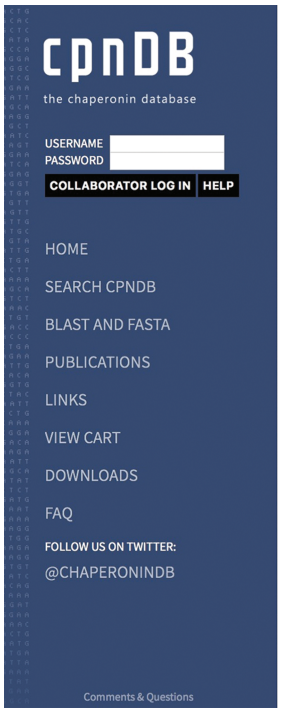
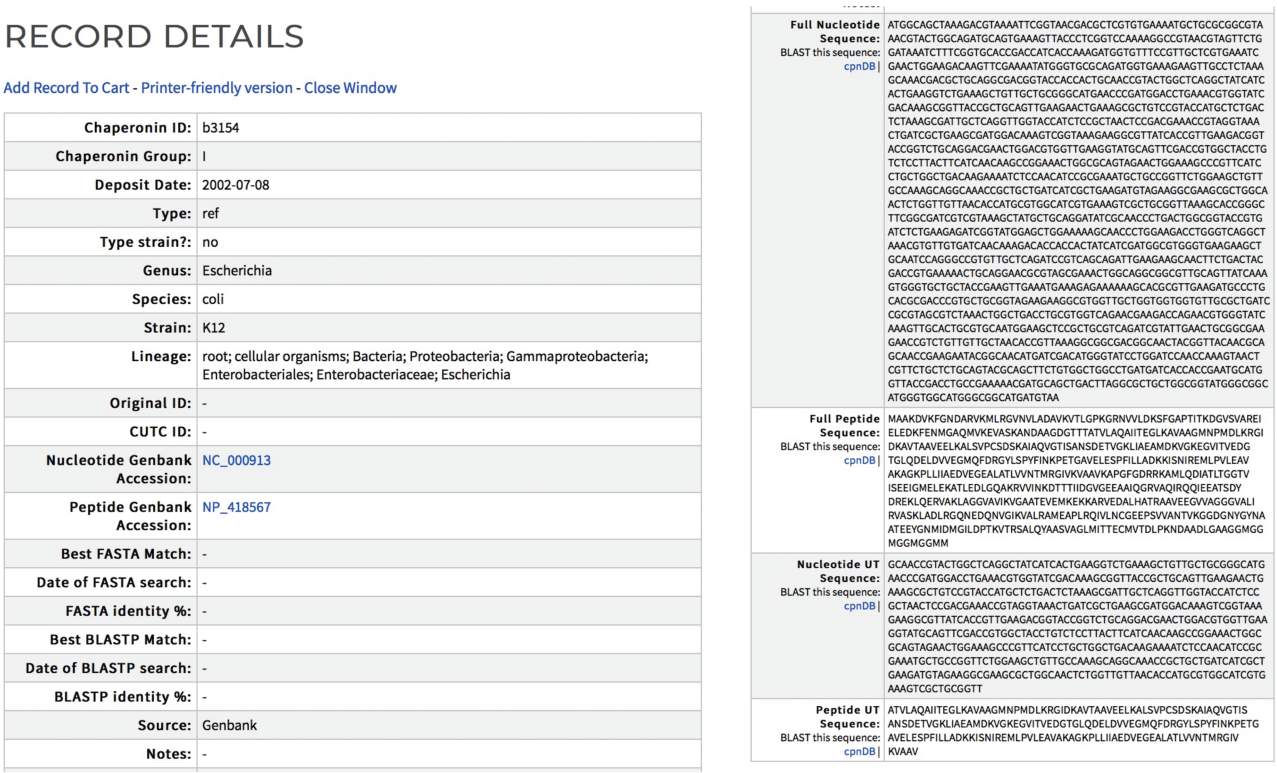



**Figure 1.** Homepage of cpnDB.



RECORD DETAILS

Add Record To Cart - Printer-friendly version - Close Window

| | |
|---|---|
| Chaperonin ID: | b3154 |
| Chaperonin Group: | I |
| Deposit Date: | 2002-07-08 |
| Type: | ref |
| Type strain?: | no |
| Genus: | Escherichia |
| Species: | coli |
| Strain: | K12 |
| Lineage: | root; cellular organisms; Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia |
| Original ID: | - |
| CUTC ID: | - |
| Nucleotide Genbank Accession: | NC_000913 |
| Peptide Genbank Accession: | NP_418567 |
| Best FASTA Match: | - |
| Date of FASTA search: | - |
| FASTA identity %: | - |
| Best BLASTP Match: | - |
| Date of BLASTP search: | - |
| BLASTP identity %: | - |
| Source: | Genbank |
| Notes: | - |

| | |
|---|---|
| Full Nucleotide Sequence: BLAST this sequence: cpnDB \| | ATGGCAGCTAAAGACGTAAAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTA AACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAAAAGGCCGTAACGTAGTTCTG GATAAATCTTTTCGGTGCACCGACCATCACCAAAGATGGTGTTTCCGTTGCTCGTGAAATC GAACTGGAAGACAAGTTCGAAAATATGGGTGCGCAGATGGTGAAAGAAGTTGCCTCTAAA GCAAACGACGCTGCAGGCGACGGTACCACCACTGCAACCGTACTGGCTCAGGCTATCATC ACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATGAACCCGATGGACCTGAAACGTGGTATC GACAAAGCGGTTACCGCTGCAGTTGAAGAACTGAAAGCGCTGTCCGTACCATGCTCTGAC TCTAAAGCGATTGCTCAGGTTGGTACCATCTCCGCTAACTCCGACGAAACCGTAGGTAAA CTGATCGCTGAAGCGATGGACAAAGTCGGTAAAGAAGGCGTTATCACCGTTGAAGACGGT ACCGGTCTGCAGGACGAACTGGACGTGGTTGAAGGTATGCAGTTCGACCGTGGCTACCTG TCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAACTGGAAAGCCCGTTCATC CTGCTGGCTGACAAGAAAATCTCCAACATCCGCGAAATGCTGCCGGTTCTGGAAGCTGTT GCCAAAGCAGGCAAACCGCTGCTGATCATCGCTGAAGATGTAGAAGGCGAAGCGCTGGCA ACTCTGGTTGTTAACACCATGCGTGGCATCGTGAAAGTCGCTGCGGTTAAAGCACCGGGC TTCGGCGATCGTCGTAAAGCTATGCTGCAGGATATCGCAACCCTGACTGGCGGTACCGTG ATCTCTGAAGAGATCGGTATGGAGCTGGAAAAAGCAACCCTGGAAGACCTGGGTCAGGCT AAACGTGTTGTGATCAACAAAGACACCACCACTATCATCGATGGCGTGGGTGAAGAAGCT GCAATCCAGGGCCGTGTTGCTCAGATCCGTCAGCAGATTGAAGAAGCAACTTCTGACTAC GACCGTGAAAAACTGCAGGAACGCGTAGCGAAACTGGCAGGCGGCGTTGCAGTTATCAAA GTGGGTGCTGCTACCGAAGTTGAAATGAAAGAGAAAAAAGCACGCGTTGAAGATGCCCTG CACGCGACCCGTGCTGCGGTAGAAGAAGGCGTGGTTGCTGGTGGTGGTGTTGCGCTGATC CGCGTAGCGTCTAAACTGGCTGACCTGCGTGGTCAGAACGAAGACCAGAACGTGGGTATC AAAGTTGCACTGCGTGCAATGGAAGCTCCGCTGCGTCAGATCGTATTGAACTGCGGCGAA GAACCGTCTGTTGTTGCTAACACCGTTAAAGGCGGCGACGGCAACTACGGTTACAACGCA GCAACCGAAGAATACGGCAACATGATCGACATGGGTATCCTGGATCCAACCAAAGTAACT CGTTCTGCTCTGCAGTACGCAGCTTCTGTGGCTGGCCTGATGATCACCACCGAATGCATG GTTACCGACCTGCCGAAAAACGATGCAGCTGACTTAGGCGCTGCTGGCGGTATGGGCGGC ATGGGTGGCATGGGCGGCATGATGTAA |
| Full Peptide Sequence: BLAST this sequence: cpnDB \| | MAAKDVKFGNDARVKMLRGVNVLADAVKVTLGPKGRNVVLDKSFGAPTITKDGVSVAREI ELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGI DKAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDG TGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFILLADKKISNIREMLPVLEAV AKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTV ISEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDY DREKLQERVAKLAGGVAIVKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALI RVASKLADLRGQNEDQNVGIKVALRAMEAPLRQIVLNCGEEPSVVANTVKGGDGNYGYNA ATEEYGNMIDMGILDPTKVTRSALQYAASVAGLMITTECMVTDLPKNDAADLGAAGGMGG MGGMGGMM |
| Nucleotide UT Sequence: BLAST this sequence: cpnDB \| | GCAACCGTACTGGCTCAGGCTATCATCACTGAAGGTCTGAAAGCTGTTGCTGCGGGCATG AACCCGATGGACCTGAAACGTGGTATCGACAAAGCGGTTACCGCTGCAGTTGAAGAACTG AAAGCGCTGTCCGTACCATGCTCTGACTCTAAAGCGATTGCTCAGGTTGGTACCATCTCC GCTAACTCCGACGAAACCGTAGGTAAACTGATCGCTGAAGCGATGGACAAAGTCGGTAAA GAAGGCGTTATCACCGTTGAAGACGGTACCGGTCTGCAGGACGAACTGGACGTGGTTGAA GGTATGCAGTTCGACCGTGGCTACCTGTCTCCTTACTTCATCAACAAGCCGGAAACTGGC GCAGTAGAACTGGAAAGCCCGTTCATCCTGCTGGCTGACAAGAAAATCTCCAACATCCGC GAAATGCTGCCGGTTCTGGAAGCTGTTGCCAAAGCAGGCAAACCGCTGCTGATCATCGCT GAAGATGTAGAAGGCGAAGCGCTGGCAACTCTGGTTGTTAACACCATGCGTGGCATCGTG AAAGTCGCTGCGGTT |
| Peptide UT Sequence: BLAST this sequence: cpnDB \| | ATVLAQAIITEGLKAVAAGMNPMDLKRGIDKAVTAAVEELKALSVPCSDSKAIAQVGTIS ANSDETVGKLIAEAMDKVGKEGVITVEDGTGLQDELDVVEGMQFDRGYLSPYFINKPETG AVELESPFILLADKKISNIREMLPVLEAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIV KVAAV |

**Figure 2.** Sample cpnDB record for *E. coli* strain K12. External links to the corresponding Genbank record are included.

and are correctly annotated. UT sequences from full-length type I chaperonin (cpn60) sequences are identified based on location of sequence regions corresponding to the predicted annealing sites of the degenerate PCR primers that are used to amplify the UT region, H279 and H280 (2). Direct submissions to cpnDB are not

**Table 1.** Taxonomic categories represented by Type I and II chaperonin sequences in cpnDB

| Domain; (phylum, group or superphylum) [a] | Number of Genera | Number of species |
|---|---|---|
| Type I | | |
| Archaea; Euryarchaeota | 10 | 13 |
| Bacteria; Acidobacteria | 7 | 9 |
| Bacteria; Actinobacteria | 162 | 725 |
| Bacteria; Aquificae | 8 | 13 |
| Bacteria; Armatimonadetes | 1 | 1 |
| Bacteria; FCB group | 139 | 379 |
| Bacteria; PVC group | 32 | 51 |
| Bacteria; Chloroflexi | 12 | 15 |
| Bacteria; Chrysiogenetes | 1 | 1 |
| Bacteria; Cyanobacteria | 44 | 64 |
| Bacteria; Deferribacteres | 5 | 5 |
| Bacteria; Deinococcus–Thermus | 6 | 23 |
| Bacteria; Dictyoglomi | 1 | 2 |
| Bacteria; Elusimicrobia | 1 | 1 |
| Bacteria; Firmicutes; Clostridia | 107 | 286 |
| Bacteria; Firmicutes; Bacilli | 92 | 545 |
| Bacteria; Firmicutes; Mollicutes | 4 | 17 |
| Bacteria; Firmicutes; Erysipelotrichia | 9 | 10 |
| Bacteria; Fusobacteria | 7 | 20 |
| Bacteria; Gemmatimonadetes | 1 | 1 |
| Bacteria; Nitrospirae | 3 | 4 |
| Bacteria; Proteobacteria; Alphaproteobacteria | 194 | 491 |
| Bacteria; Proteobacteria; Betaproteobacteria | 89 | 233 |
| Bacteria; Proteobacteria; Gammaproteobacteria | 214 | 764 |
| Bacteria; Proteobacteria; Delta/Epsilon subdivisions | 63 | 151 |
| Bacteria; Spirochaetes | 11 | 78 |
| Bacteria; Synergistetes | 8 | 9 |
| Bacteria; Tenericutes | 5 | 23 |
| Bacteria; Thermodesulfobacteria | 2 | 3 |
| Bacteria; Thermotogae | 8 | 15 |
| Eukaryota; Archaeplastida | 67 | 79 |
| Eukaryota; SAR supergroup | 32 | 50 |
| Eukaryota; Excavata | 9 | 20 |
| Eukaryota; Amoebozoa | 4 | 10 |
| Eukaryota; Opisthokonta; Metazoa | 83 | 94 |
| Eukaryota; Opisthokonta; Fungi | 57 | 132 |
| Type II | | |
| Archaea; Euryarchaeota | 66 | 158 |
| Archaea; DPANN | 1 | 1 |
| Archaea; Proteoarchaeota | 27 | 46 |
| Eukaryota; Archaeplastida | 25 | 28 |
| Eukaryota; SAR supergroup | 21 | 46 |
| Eukaryota; Excavata | 5 | 12 |
| Eukaryota; Amoebozoa | 5 | 8 |
| Eukaryota; Opisthokonta; Metazoa | 89 | 106 |
| Eukaryota; Opisthokonta; Fungi | 53 | 77 |
| Eukaryota; Hacrobia | 4 | 4 |

[a] Based on NCBI Taxonomy

accepted, but curators encourage users to deposit sequences to public databases, which are surveyed weekly with PubCrawler (29). Sequences from newly identified genera and species, especially bacterial type strains or reference strains with thorough annotation, are prioritized for inclusion.

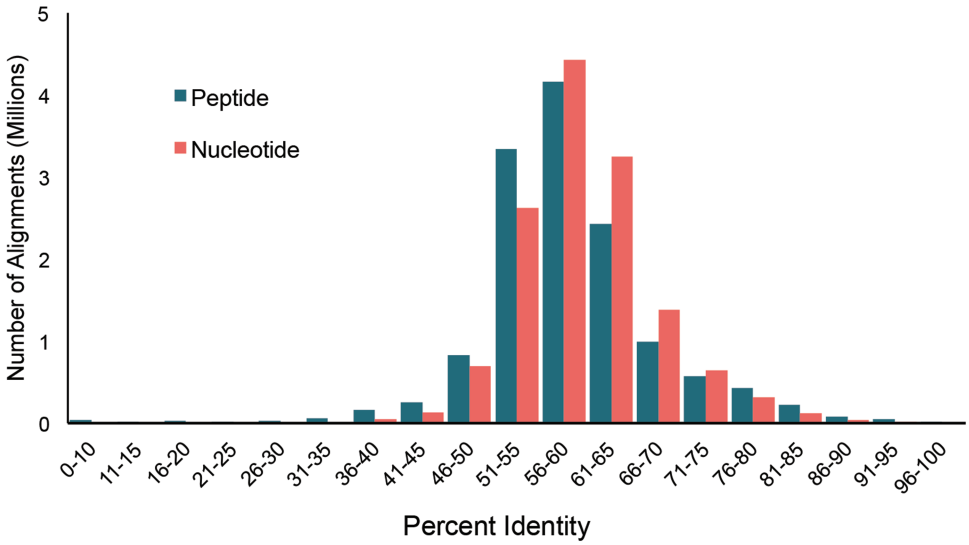



**Figure 3.** Distribution of pairwise percent identities for bacterial cpn60 UT sequences in cpnDB_nr (version 11 May 2018; 5235 species).

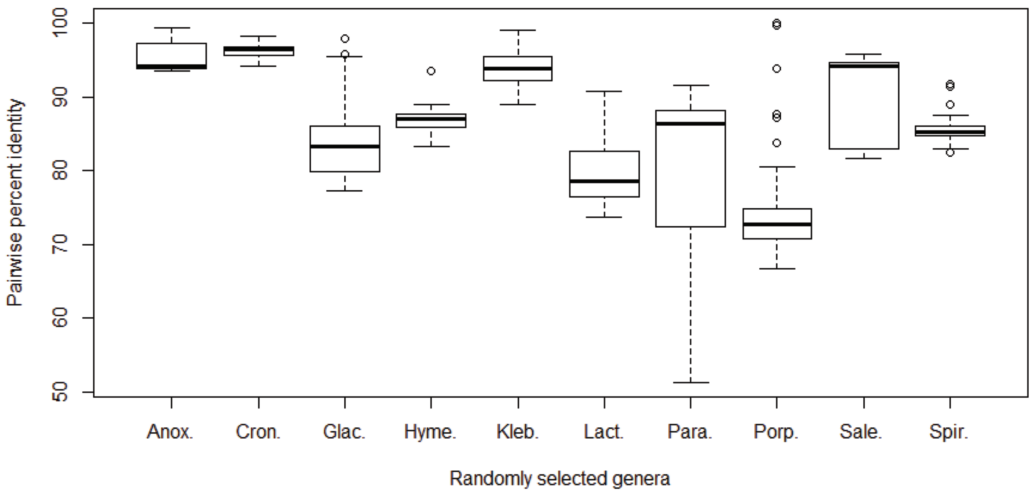



**Figure 4.** Pairwise percent identities of the *cpn*60 gene among species. The *cpn*60 sequences of species from 10 randomly selected bacterial genera with at least 6 unique species records were aligned to calculate pairwise identities. Genera from left to right: *Anoxybacillus* (6 species), *Cronobacter* (6 species), *Glaciecola* (10 species), *Hymenobacter* (6 species), *Klebsiella* (9 species), *Lactococcus* (7 species), *Paracoccus* (9 species), *Porphyromonas* (17 species), *Salegentibacter* (6 species) and *Spirosoma* (7 species).

## Database contents and scope

Since the release of cpnDB in 2004 (26), the database has grown from approximately 2000 records representing 247 genera to over 25 000 records representing 1802 genera, with a corresponding increase in taxonomic coverage (Table 1). Not surprisingly, since cpnDB records are primarily drawn from public databases, the highest numbers of entries are from organisms in the most well-studied bacterial phyla, with 1070/1802 genera belonging to phyla Firmicutes, Actinobacteria, Proteobacteria and Bacteroidetes. The number of species per genus ranges from 1 to over 200 (*Streptomyces* spp.).

Bacterial type I chaperonin (cpn60) sequences are the most widely exploited for research and diagnostics. To investigate sequence diversity among bacterial type I chaperonin sequences in cpnDB, pairwise percent identities were calculated for the type I bacterial sequences from cpnDB_nr (version 25 November 2018) nucleotide and peptide UT sequences (5235 sequences for each) using Clustalw2 (30). In both cases, normal distributions of pairwise percent identity values were observed, with median values of 59.8% and 58.1% for nucleotide and peptide sequences, respectively (Figure 3). The pattern is strikingly similar to the distribution of nucleotide and peptide UT sequence identities

calculated for representatives of the original 247 genera included in cpnDB (26), indicating that while the depth and scope of cpnDB have increased, the spectrum of chaperonin sequence diversity has not expanded substantially beyond what was initially described in 2004.

The 16S rRNA gene sequences are commonly used markers for describing bacterial diversity and an ~98.7–99% identity recommended for demarcation of species based on 16S rRNA sequences (31). More rapidly evolving protein-coding genes offer superior resolution (17, 32, 33), which supports their exploitation in diagnostics since lower sequence identities between closely related taxa makes their discrimination more obvious and technically tractable. To investigate inter-species cpn60 UT sequence identities, we randomly selected the following 10 bacterial genera for which there were at least 6 species available in cpnDB: *Anoxybacillus* (6 species), *Cronobacter* (6 species), *Glaciecola* (10 species), *Hymenobacter* (6 species), *Klebsiella* (9 species), *Lactococcus* (7 species), *Paracoccus* (9 species), *Porphyromonas* (17 species), *Salegentibacter* (6 species) and *Spirosoma* (7 species). cpn60 UT nucleotide sequences for each genus were aligned and pairwise identities were calculated (Figure 4). Median identities were highly variable among genera, ranging from 72.8% for *Porphyromonas* to 96.4% for *Cronobacter*. There was also a wide of range of values observed within genera, and in three cases (*Glaciecola*, *Paracoccus* and *Porphyromonas*), the range exceeded 20%. This wide distribution of identities within genera is consistent with previously published observations of *Campylobacter* (71–92% pairwise identity among 15 species; 23), *Parabacteroides* (83–97% among 5 species; 34), *Prevotella* (68–94% among 38 species; 34) and *Enterococcus* (78%–88% among 18 species; 9). A previous study comparing cpn60 UT and 16S rRNA gene sequence diversity within a set of 983 bacterial species from 21 phyla similarly concluded that cpn60 UT intra- and inter-specific identities were more broadly distributed and almost always lower than corresponding 16S rRNA gene sequence identities (21).

## Conclusions and future directions

cpnDB has supported chaperonin sequence-based research and diagnostics since 2004 and will continue to be updated as new genomes are sequenced and species are discovered to ensure its continuation as a source for reference sequences of chaperonin genes representing the broadest possible taxonomic range. Curation will continue to focus on reference sequence data from microbial taxa, but additional complementary resources focused on curation and analysis of data generated in microbiome studies using cpn60 amplicon sequencing are also being developed.

cpnDB remains one of very few sources of sequence barcode information, along with the Ribosomal Database Project (ribosomal RNA encoding genes for bacteria, archaea and fungi; 35) and the Barcode of Life Data System (cytochrome oxidase I genes for plants, animals, protists and fungi; 36). The accessibility of chaperonin UT sequences with degenerate primer PCR coupled with the superior resolution of these usually single-copy sequences will continue to make chaperonin sequences attractive barcodes for detection, identification and quantification of organisms in isolation or in complex microbial communities.

## Funding

Natural Sciences and Engineering Research Council of Canada (Discovery Grant to support for the development and curation of cpnDB to J.E.H. and Undergraduate Student Research Award to S.J.V.).

*Conflict of interest*. None declared.

## References

1. Horwich,A.L., Fenton,W.A., Chapman,E. *et al.* (2007) Two families of chaperonin: physiology and mechanism. *Annu. Rev. Cell Dev. Biol.*, **23**, 115–145.
2. Goh,S.H., Potter,S., Wood,J.O. *et al.* (1996) HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J. Clin. Microbiol.*, **34**, 818–823.
3. Sakamoto,M., Suzuki,N. and Okamoto,M. (2010) *Prevotella aurantiaca* sp. nov., isolated from the human oral cavity. *Int. J. Syst. Evol. Microbiol.*, **60**, 500–503.
4. Sakamoto,M., Ikeyama,N., Kunihiro,T. *et al.* (2018) *Mesosutterella multiformis* gen. nov., sp. nov., a member of the family *Sutterellaceae* and *Sutterella megalosphaeroides* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.*, **68**, 3942–3950.
5. Dumonceaux,T.J., Schellenberg,J., Goleski,V. *et al.* (2009) Multiplex detection of bacteria associated with normal microbiota and with bacterial vaginosis in vaginal swabs using oligonucleotide-coupled fluorescent microspheres. *J. Clin. Microbiol.*, **47**, 4067–4077.
6. Chaban,B., Musil,K.M., Himsworth,C.G. *et al.* (2009) Development of cpn60-based real-time quantitative PCR assays for the detection of 14 *Campylobacter* species and application to screening canine fecal samples. *Appl. Environ. Microbiol.*, **75**, 3055–3061.
7. Hill,J.E., Paccagnella,A., Law,K. *et al.* (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.
8. Goh,S.H., Santucci,Z., Kloos,W.E. *et al.* (1997) Identification of *Staphylococcus* species and subspecies by the chaperonin 60 gene identification method and reverse checkerboard hybridization. *J. Clin. Microbiol.*, **35**, 3116–3121.

9. Goh,S.H., Facklam,R.R., Chang,M. *et al.* (2000) Identification of *Enterococcus* species and phenotypically similar *Lactococcus* and *Vagococcus* species by reverse checkerboard hybridization to chaperonin 60 gene sequences. *J. Clin. Microbiol.*, **38**, 3953–3959.
10. Brousseau,R., Hill,J.E., Prefontaine,G. *et al.* (2001) *Streptococcus suis* serotypes characterized by analysis of chaperonin 60 gene sequences. *Appl. Environ. Microbiol.*, **67**, 4828–4833.
11. Masson,L., Maynard,C., Brousseau,R. *et al.* (2006) Identification of pathogenic *Helicobacter* species by chaperonin-60 differentiation on plastic DNA arrays. *Genomics*, **87**, 104–112.
12. Albert,A.Y., Chaban,B., Wagner,E.C. *et al.* (2015) A study of the vaginal microbiome in healthy Canadian women utilizing cpn60-based molecular profiling reveals distinct *Gardnerella* subgroup community state types. *PLoS One*, **10**, e0135620.
13. Pratt,D.L., Dumonceaux,T.J., Links,M.G. *et al.* (2012) Influence of mass burial of animal carcasses on the types and quantities of microorganisms within a burial site. *Trans ASABE*, **55**, 2195–2212.
14. Oliver,K.L., Hamelin,R.C. and Hintz,W.E. (2008) Effects of transgenic hybrid aspen overexpressing polyphenol oxidase on rhizosphere diversity. *Appl. Environ. Microbiol.*, **74**, 5340–5348.
15. Bondici,V.F., Lawrence,J.R., Khan,N.H. *et al.* (2013) Microbial communities in low permeability, high pH uranium mine tailings: characterization and potential effects. *J. Appl. Microbiol.*, **114**, 1671–1686.
16. Freitas,A.C., Chaban,B., Bocking,A. *et al.* (2017) The vaginal microbiome of healthy pregnant women is less rich and diverse with lower prevalence of Mollicutes compared to healthy non-pregnant women. *Sci. Rep.*, **7**, 9212.
17. Verbeke,T.J., Sparling,R., Hill,J.E. *et al.* (2011) Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (*cpn*60 UT): application to *Thermoanaerobacter* species. *Syst. Appl. Microbiol.*, **34**, 171–179.
18. Katyal,I., Chaban,B. and Hill,J.E. (2015) Comparative genomics of *cpn*60 defined *Enterococcus hirae* ecotypes and relationship of gene content differences to competitive fitness. *Microb. Ecol.*, **72**, 917–930.
19. Schellenberg,J.J., Paramel Jayaprakash,T., Withana Gamage,N. *et al.* (2016) *Gardnerella vaginalis* subgroups defined by cpn60 sequencing and sialidase activity in isolates from Canada, Belgium and Kenya. *PLoS One*, **11**, e0146510.
20. Hebert,P.D., Cywinska,A., Ball,S.L. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.*, **270**, 313–321.
21. Links,M.G., Dumonceaux,T.J., Hemmingsen,S.M. *et al.* (2012) The chaperonin-60 universal target is a barcode for bacteria that enables *de novo* assembly of metagenomic sequence data. *PLoS One*, **7**, e49755.
22. Vermette,C.J., Russell,A.H., Desai,A.R. *et al.* (2009) Resolution of phenotypically distinct strains of *Enterococcus* spp. in a complex microbial community using cpn60 universal target sequencing. *Microb. Ecol.*, **59**, 14–24.
23. Hill,J.E., Paccagnella,A., Law,K. *et al.* (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.
24. Paramel Jayaprakash,T., Schellenberg,J.J. and Hill,J.E. (2012) Resolution and characterization of distinct cpn60-based subgroups of *Gardnerella vaginalis* in the vaginal microbiota. *PLoS One*, **7**, e43009.
25. Chaban,B. and Hill,J.E. (2011) A 'universal' type II chaperonin PCR detection system for the investigation of Archaea in complex microbial communities. *ISME J.*, **6**, 430–439.
26. Hill,J.E., Penny,S.L., Crowell,K.G. *et al.* (2004) cpnDB: a chaperonin sequence database. *Genome Res.*, **14**, 1669–1675.
27. Altschul,S.F., Madden,T.L., Schaffer,A.A. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
28. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 2444–2448.
29. Hokamp,K. and Wolfe,K.H. (2004) PubCrawler: keeping up comfortably with PubMed and GenBank. *Nucleic Acids Res.*, **32**, W16–W19.
30. Larkin,M.A., Blackshields,G., Brown,N.P. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
31. Stackebrandt,E. and Ebers,J. (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today*, **33**, 152–155.
32. Case,R.J., Boucher,Y., Dahllof,I. *et al.* (2007) Use of 16S rRNA and *rpo*B genes as molecular markers for microbial ecology studies. *Appl. Environ. Microbiol.*, **73**, 278–288.
33. Zeigler,D.R. (2003) Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int. J. Syst. Evol. Microbiol.*, **53**, 1893–1900.
34. Sakamoto,M. and Ohkuma,M. (2010) Usefulness of the hsp60 gene for the identification and classification of Gram-negative anaerobic rods. *J. Med. Microbiol.*, **59**, 1293–1302.
35. Cole,J.R., Wang,Q., Fish,J.A. *et al.* (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.*, **42**, D633–D642.
36. Ratnasingham,S. and Hebert,P.D. (2007) bold: the barcode of life data system (http://www.barcodinglife.org). *Mol. Ecol. Notes*, **7**, 355–364.