



1-1-2021

The Dynamic Codon Biases: calculating prokaryotic codon usage biases

Brian Dehlinger
Loyola University Chicago

Jared Jurss
Loyola University Chicago

Karson Lychuk
Loyola University Chicago

Catherine Putonti
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/biology_facpubs

Recommended Citation

Dehlinger, Brian; Jurss, Jared; Lychuk, Karson; and Putonti, Catherine. The Dynamic Codon Biases: calculating prokaryotic codon usage biases. *Microbial Genomics*, 7, 10: 1-6, 2021. Retrieved from Loyola eCommons, Biology: Faculty Publications and Other Works, <http://dx.doi.org/10.1099/MGEN.0.000663>

This Article is brought to you for free and open access by the Faculty Publications and Other Works by Department at Loyola eCommons. It has been accepted for inclusion in Biology: Faculty Publications and Other Works by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

The Dynamic Codon Biased: calculating prokaryotic codon usage biases

Brian Dehlinger¹, Jared Jurss¹, Karson Lychuk¹ and Catherine Putonti^{1,2,3,4,*}

Abstract:

Bacterial genomes often reflect a bias in the usage of codons. These biases are often most notable within highly expressed genes. While deviations in codon usage can be attributed to selection or mutational biases, they can also be functional, for example controlling gene expression or guiding protein structure. Several different metrics have been developed to identify biases in codon usage. Previously we released a database, CBDB: The Codon Bias Database, in which users could retrieve pre-calculated codon bias data for bacterial RefSeq genomes. With the increase of bacterial genome sequence data since its release a new tool was needed. Here we present the Dynamic Codon Biased (DCB) tool, a web application that dynamically calculates the codon usage bias statistics of prokaryotic genomes. DCB bases these calculations on 40 different highly expressed genes (HEGs) that are highly conserved across different prokaryotic species. A user can either specify an NCBI accession number or upload their own sequence. DCB returns both the bias statistics and the genome's HEG sequences. These calculations have several downstream applications, such as evolutionary studies and phage-host predictions. The source code is freely available, and the website is hosted at www.cbdb.info.

DATA SUMMARY

The DCB webservice is available at www.cbdb.info and source code is available at <https://github.com/BrianDehlinger/DCB-Dynamic-Codon-Biased>.

INTRODUCTION

In many bacterial genomes there is a preferential usage of certain codons over other synonymous codons for the same amino acid. Processes such as mutational bias and translational selection often cause these biases (see reviews [1, 2]). Bacteria have been shown to have varying degrees of codon usage bias, suggesting that there are varying amounts of translational selection among different bacteria. In fact, genes that undergo more translational selection often have a greater codon usage bias [3, 4], and codon usage has been optimized by prokaryotic species over time to improve their translation [5]. Further bolstering this theory, evidence shows that high

codon usage bias correlates with high gene expression [6]. Transcript structure and codon usage have been found to have significant effects on both protein production, mRNA abundance and stability, and also bacterial growth rate [7–14]. Biases in di-codon usage also have been observed [15, 16] and usage differs between the coding sequences of highly and lowly abundant proteins [17]. Furthermore, codon usage optimization is well documented within bacteriophage species: phages frequently reflect the codon usage of their bacterial host [18–22]. While similarities in subsequence usage, including codon bias, have been used to predict a phage's host species [23, 24], alone it has limited success [24]. This prompted our prior development of CBDB: the Codon Bias Database [25], which contained precomputed calculations of codon bias usage within the highly expressed genes (HEGs) of bacterial Reference sequence (RefSeq) genomes [26].

A number of different metrics have been proposed to quantify codon bias, including relative synonymous codon usage

Received 31 May 2020; Accepted 02 August 2021; Published 26 October 2021

Author affiliations: ¹Bioinformatics Program, Loyola University Chicago, Chicago, IL 60660, USA; ²Department of Biology, Loyola University Chicago, Chicago, IL 60660, USA; ³Department of Computer Science, Loyola University Chicago, Chicago, IL 60660, USA; ⁴Department of Microbiology and Immunology, Loyola University Chicago, Stritch School of Medicine, Maywood, IL 60153, USA.

*Correspondence: Catherine Putonti, cputonti@luc.edu

Keywords: codon usage; codon bias; Dynamic Codon Biased; prokaryotes.

Abbreviations: CAI, codon adaptation index; DCB, Dynamic Codon Biased; ENC, effective number of codons; FB, frequency bias; GUI, graphical user interface; HEG, highly expressed gene; NRSCU, normalized relative synonymous codon usage; rCAI, relative codon adaptation index; RefSeq, reference sequence; RSCU, relative synonymous codon usage; SSCI, self-consistent codon index; tAI, tRNA adaptation index.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary table and one supplementary file are available with the online version of this article.

000663 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

(RSCU) [27], the codon adaptation index (CAI) [3], the self-consistent codon index (SSCI) [28] and relative codon adaptation index (rCAI) [29]. Rather than looking at the codons themselves to ascertain biases, a second approach exists in which biases are assessed relative to individual tRNA abundances, the tRNA adaptation index (tAI) [30]. Several resources for examining codon usage bias already exist (Table 1), including five web resources: CoCoPUTs [31], CAIcal [32], HEG-DB [33], SMS and COUSIN [34]. These web tools can be categorized as either a database (CoCoPUTs and HEG-DB) or interactive analysis for user-supplied sequences (CAIcal, SMS and COUSIN). The CoCoPUTs database provides a graphical user interface (GUI) that allows a user to specify a taxonomic id or scientific name and returns codon usage tables, the effective number of codons (ENC), and codon pair usage calculated from NCBI complete genome sequences [35]. HEG-DB gives the CAI values for HEGs of 200 bacterial genomes [32, 33]. Our previous database, CBDB, included codon usage metrics for HEGs for hundreds of bacterial RefSeq genomes [25], although with the introduction of our new tool presented here, it is no longer available. While the other tools listed in Table 1 are also capable of calculating codon metrics for HEGs, they require users to identify HEG sequences and supply these gene sequences (and/or codon usage tables).

Given the rate at which bacterial genomes are now being produced daily, a static resource has limited utility. Although CoCoPUTs integrates all complete genomes in GenBank, it does not include bacterial assemblies, the largest growing collection of bacterial genomic sequences. Furthermore, these databases must be updated by the tool's team. The Dynamic Codon Biased (DCB) was developed to facilitate statistical analysis of codon usage bias across different bacterial genomes using the 40 different highly conserved and highly expressed

Impact Statement

Codon usage bias is a key feature of many genomes. Within bacterial genomes, codon usage can emerge as a result of mutational bias and translational selection and has been associated with gene expression levels. Highly expressed genes often represent the strength of the bias within a given genome. We have developed a web tool – Dynamic Codon Biased (DCB) – for users to calculate codon usage bias for any publicly available genome sequence (complete or draft) or upload their own sequence. Currently no tool exists for users to analyse individual strains or unpublished sequences. Thus, researchers can consider this important metric when analysing a bacterial genome. The source code for this tool is also publicly available.

genes described by Sharp *et al.* [4]. This web application is dynamic and will identify HEG sequences and calculate results in real time for both publicly available genomes (draft or complete), by querying NCBI's GenBank directly, as well as user-supplied genome sequences.

IMPLEMENTATION

The DCB was developed in Python 3 utilizing completely open source tools and software libraries including Prodigal [36], DIAMOND [37], Biopython [38], BeautifulSoup4 (<https://www.crummy.com/software/BeautifulSoup/>) and Flask (<http://flask.pocoo.org/>). DIAMOND and Prodigal identify HEGs and annotate genome sequences, respectively. While there are many homology tools available, DIAMOND

Table 1. Available tools for calculating codon usage metrics

Tool	Functionality	Availability	URL (citation)
CAIcal	Calculates CAI for provided gene sequences and codon usage tables	Web	http://genomes.urv.es/CAIcal/ [32]
CoCoPUTs	Database of codon-pair and dinucleotide statistics for all genomes in GenBank	Web	https://hive.biochemistry.gwu.edu/review/codon2 [31]
CodonW	Calculates codon metrics for user-selected gene set and correspondence analysis	Local installation	https://sourceforge.net/projects/codonw/
coRdon	Calculates codon bias statistics	R package	https://www.bioconductor.org/packages/develop/bioc/vignettes/coRdon/inst/doc/coRdon.html
COUSIN	Calculates codon usage for user-supplied sequences	Web or install	http://cousin.ird.fr/index.php [34]
EncPrime	Calculates ENC metric	Local installation	https://github.com/jnovembre/ENCprime
GCUA	Calculates codon metrics for user-selected gene set and correspondence analysis	Local installation	http://mcinerneylab.com/software/gcua/
HEG-DB	Database of CAI index of HEGs for 200 genomes	Web	http://genomes.urv.cat/HEG-DB/ [33]
SMS	Calculates codon metrics for user-supplied sequences	Web	https://www.bioinformatics.org/sms2/codon_usage.html

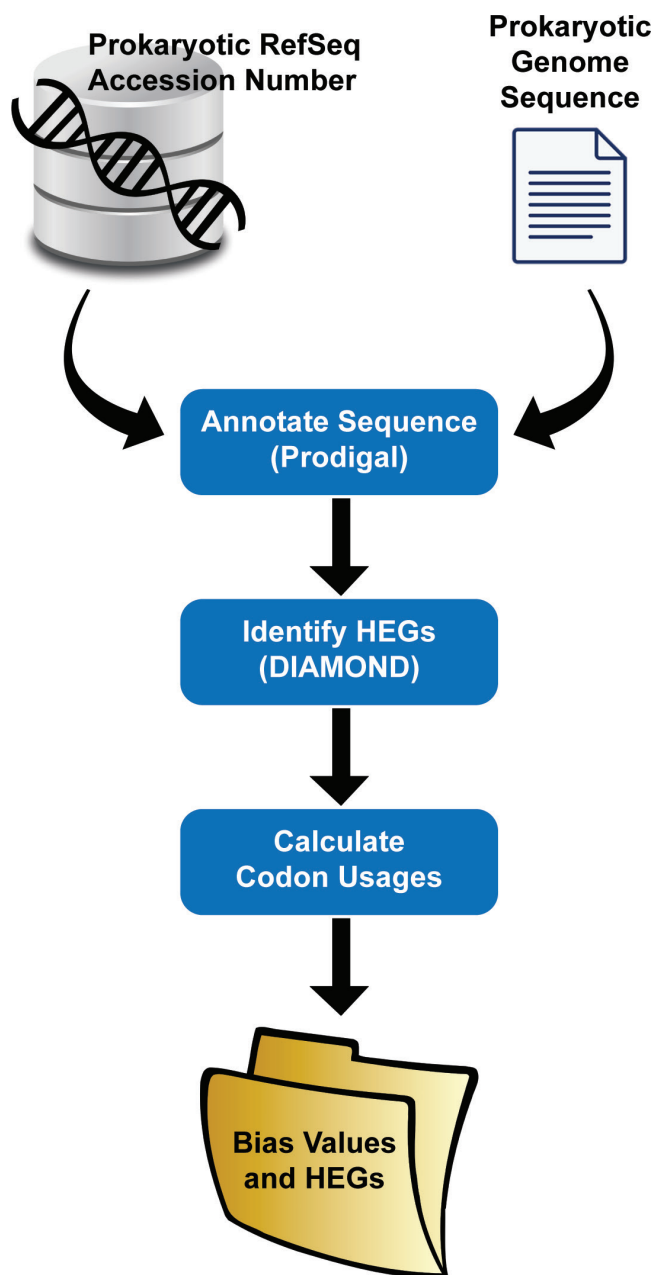


Fig. 1. Basic workflows for DCB.

was selected for its speed. Python flask was chosen as the web framework as it requires minimal support and is lightweight and scalable. The DCB webserver is available at www.cbdb.info and source code is available at <https://github.com/Brian-Dehlinger/DCB-Dynamic-Codon-Biaser>.

Fig. 1 outlines the process for analysis using either publicly available sequences from NCBI (left) or user-supplied genome sequences (right). For NCBI sequences, the web application accepts a RefSeq accession number as input. This can be either a complete genome or draft genome assembly. Beautifulsoup4 is used to navigate NCBI's ftp

back end and retrieve the organism's assembly and annotated coding sequences from the genome (*.fna files). If the genome sequence is not annotated, Prodigal is run to identify coding regions. DIAMOND then queries coding sequences (from either NCBI's annotation or Prodigal predictions) against a local protein database containing representatives of the 40 HEGs described by Sharp *et al.* [4]. This database was constructed by utilizing the Identical Protein Groups tool on NCBI and filtering to only include prokaryotes from the UniProtKB/Swiss-Prot source database [39]. The final database size included 1186 different sequences, representative of the phylogenetic diversity of sequenced prokaryotic species. This database can be retrieved via our GitHub repository, https://github.com/BrianDehlinger/DCB-Dynamic-Codon-Biaser/blob/master/testApp/protein_databasefasta. The top DIAMOND hit for each of the 40 HEGs is identified and codon usage is calculated using a modified version of Biopython's CodonUsage module [38]. DCB reports three statistics related to codon usage: the relative synonymous codon usage (RSCU), the normalized relative synonymous codon usage (NRSCU) and frequency bias (HEG FB). The web application returns a zip file containing the statistics in addition to the HEG file. Statistics are written in comma-separated value (csv) files to facilitate analysis via Excel or Python or R.

Computer code providing the same functionality available through the web is also available through our GitHub repository. Users can install and run the tool locally or set up their own web service. This code was developed using Python, HTML, CSS and Javascript for Linux Ubuntu 14.04 or higher. Dependencies include Prodigal and the following Python modules: Biopython, BeautifulSoup4, Pandas, Numpy, flask, requests and flask-bootstrap.

RESULTS

To evaluate the utility of DCB, three investigations were conducted. First, we evaluated the codon usage of *Escherichia coli* str. K-12 substr. MG1655 (accession NZ_CP032667.1). Entering this accession number into the web interface, the results are quickly generated and automatically downloaded in a zipped folder. These results include two files: (1) a comma-separated value file listing the codon usage statistics within the HEGs and (2) a fasta format file listing the HEG sequences used to compute the codon usage statistics. Note, these usage statistics could be used to evaluate all coding sequences within the genome; such codon usage tables are required for several of the tools listed in Table 1, such as CAIcal. This genome was selected as it is also one of the 200 genomes available in HEG-DB, which identifies codon usage biases in ribosomal proteins and predicts gene expression of all genes in the genome based upon their usage of these biases. Furthermore, this strain is an available option in CoCoPUTs. While DCB provides codon usage metrics based upon HEGs, CoCoPUTs provides raw counts of usage for the entire sequence. In this first proof-of-concept example, we selected a complete genome

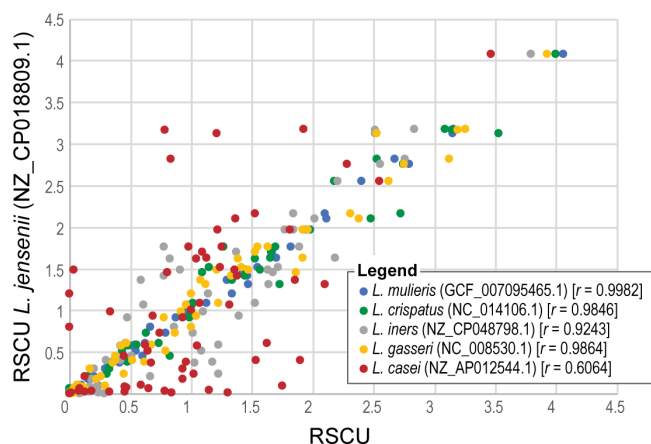


Fig. 2. Comparison of codon usage biases in *Lactobacillus* species. Five genomes are compared to *L. jensenii* (NZ_CP018809.1): *L. mulieris* (GCF_007095465.1), *L. crispatus* (NC_014106.1), *L. iners* (NZ_CP048798.1), *L. gasserii* (NC_008530.1) and *L. casei* (NZ_AP012544.1).

sequence. In the event that we wanted to evaluate a draft assembly or a user-supplied sequence, neither HEG-DB nor CoCoPUTs would be capable of conducting this analysis.

In our second investigation, we focused on comparison of several draft and complete assemblies. In March 2020 a new species of lactobacilli was identified, *Lactobacillus mulieris* [40], which also resulted in taxonomic reclassification of several *Lactobacillus jensenii* strains [40, 41]. Using DCB, we wanted to examine the codon usage profiles of *L. mulieris* and *L. jensenii* as well as other *Lactobacillus* species of the urogenital tract and lactobacilli of the gut. As expected, *L. mulieris* and *L. jensenii* have the most similar codon usage of the species examined (Fig. 2). Furthermore, *L. jensenii* has a codon usage more similar to other members of the *Lactobacillus delbrueckii* group, which are also found in the urogenital microbiota, than the *Lactobacillus casei* group (represented by *L. casei* from the gut microbiota). Generally speaking, the observed variation in codon usage mirrors other phylogenetic markers for the genus [42]. An open question, however, is how does the different environments in which these strains are found (urogenital vs. gut) shape codon usage. Previous metagenomic studies have noted such adaptations of entire microbial communities to their environments [43].

In our third proof-of-concept example, we returned to our original motivation behind exploring codon usage in bacteria – phages. Phage Pbnalikevirus phiHabibi was isolated from Lake Michigan and found to be able to lyse both *Pseudomonas aeruginosa* ATCC 15692 and *E. coli* C [44]. Codon usage biases were calculated by DCB for these two hosts, accession numbers NC_010468 and NZ_CP017149, respectively. RSCU values for each host were compared to RSCU values calculated for the phage coding regions individually and collectively (accession number KT254132). Python code to perform this calculation is provided in File S1 (available in the online version of this article). phiHabibi had a codon usage bias more similar to the *P.*

aeruginosa HEGs ($r=0.8205$) than the *E. coli* HEGs ($r=0.5500$), suggesting that *P. aeruginosa* is more likely to be the native or frequent host of this isolated phage. In total, 85 of the 90 protein coding genes similarly exhibited a codon usage more similar to *P. aeruginosa* HEGs (Table S1). Most notable is the similarity in codon usage between the *P. aeruginosa* HEGs and phiHabibi structural proteins. Previous bioinformatic analyses of phages and hosts have made similar observations [20]. Conducting such codon usage comparisons can also be useful for engineering phages; previous work has shown that codon optimization and deoptimization can increase and decrease, respectively, phage fitness [45, 46]. Given the recent renewed interest in phages for therapeutic use, codon usage is a promising avenue for phage engineering [47, 48].

DCB was specifically designed such that it has access to the latest publicly available complete and draft genomes; it is not dependent upon database updates, but rather it directly retrieves data from NCBI. Furthermore, it is flexible, allowing the user to upload a FASTA format file directly and does not require the user to supply HEGs; they are automatically detected for unannotated sequences, and HEG sequences are returned for all searches. For the calculation of a single genome, DCB uses around 70–75 MB of HDD space and around 72 MB of RAM. These run-time and memory usage statistics are informative for those users interested in running the tool locally rather than via the webserver, which is hosted as an EC2 instance. The source code is written to utilize one CPU core but allows for concurrent requests if a concurrent WSGI server is used as a wrapper for the flask server.

CONCLUSION

DCB provides codon usage bias analysis for all publicly available or user-supplied prokaryotic genomes. The program is available as a web application with source code available for those users interested in running it locally. The program is also modular and can be modified and expanded upon to meet different use cases. Because calculations can be generated in a matter of seconds and new prokaryotic genomes are being deposited in GenBank daily, a dynamic web service provides a better solution than static databases. Data generated from DCB analyses can easily be integrated into, for example, evolutionary studies of prokaryotes, comparative genomic studies and phage–host investigations.

Funding information

This work was supported through funding from the US National Science Foundation (CP, Award no.1661357). The funders played no role in the study or in the preparation of the article or decision to publish.

Acknowledgements

This work was conducted by the authors as part of Loyola University Chicago's Computational Biology course; the authors would like to thank their classmates for feedback throughout design and development.

Author contributions

B.D., J.J. and K.L. developed software and wrote the original draft. C.P. conceptualized and supervised the study and with B.D. assisted with writing review and editing.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Sharp PM, Stenico M, Peden JF, Lloyd AT. Codon usage: mutational bias, translational selection, or both? *Biochem Soc Trans* 1993;21:835–841.
- Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell* 2015;59:149–161.
- Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987;15:1281–1295.
- Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res* 2005;33:1141–1153.
- López JL, Lozano MJ, Fabre ML, Lagares A. Codon usage optimization in the prokaryotic tree of life: how synonymous codons are differentially selected in sequence domains with different expression levels and degrees of conservation. *mBio* 2020;11:e00766–20.
- Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* 1982;10:7055–7074.
- Cambay G, Guimaraes JC, Arkin AP. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat Biotechnol* 2018;36:1005–1015.
- Ceroni F, Algar R, Stan G, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat Methods* 2015;12:415–418.
- Frumkin I, Schirman D, Rotman A, Li F, Zahavi L. Gene architectures that minimize cost of gene expression. *Mol Cell* 2017;65:142–153.
- Guimaraes JC, Rocha M, Arkin AP. Transcript level and sequence determinants of protein abundance and noise in *Escherichia coli*. *Nucleic Acids Res* 2014;42:4791–4799.
- Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F. Synonymous codons direct cotranslational folding toward different protein conformations. *Mol Cell* 2016;61:341–351.
- Presnyak V, Alhusaini N, Chen Y, H, Martin S, Morris N. Codon optimality is a major determinant of mRNA stability. *Cell* 2015;160:1111–1124.
- Kanaya S, Yamada Y, Kudo Y, Ikemura T. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 1999;238:143–155.
- Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* 2018;19:20–30.
- Guo F-B, Ye Y-N, Zhao H-L, Lin D, Wei W. Universal pattern and diverse strengths of successive synonymous codon bias in three domains of life, particularly among prokaryotic genomes. *DNA Res* 2012;19:477–485.
- Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* 2016;166:679–690.
- Diambra LA. Differential bicodon usage in lowly and highly abundant proteins. *PeerJ* 2017;5:e3081.
- Carbone A. Codon bias is a major factor explaining phage evolution in translationally biased hosts. *J Mol Evol* 2008;66:210–223.
- Cardinale DJ, Duffy S. Single-stranded genomic architecture constrains optimal codon usage. *Bacteriophage* 2011;1:219–224.
- Lucks JB, Nelson DR, Kudla GR, Plotkin JB. Genome landscapes and bacteriophage codon usage. *PLoS Comput Biol* 2008;4:e1000001.
- Pride DT, Wassenaar TM, Ghose C, Blaser MJ. Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* 2006;7:8.
- Kunisawa T, Kanaya S, Kutter E. Comparison of synonymous codon distribution patterns of bacteriophage and host genomes. *DNA Res* 1998;5:319–326.
- Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* 2015;4.
- Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiol Rev* 2016;40:258–272.
- Hilterbrand A, Saelens J, Putonti C. CBDB: the codon bias database. *BMC Bioinformatics* 2012;13:62.
- O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;44:D733–745.
- Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 1986;14:5125–5143.
- Carbone A, Zinovyev A, Képès F. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 2003;19:2005–2015.
- Lee S, Weon S, Lee S, Kang C. Relative codon adaptation index, a sensitive measure of codon usage bias. *Evol Bioinform Online* 2010;6:47–55.
- dos Reis M, Wernisch L, Savva R. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res* 2003;31:6976–6985.
- Alexaki A, Kames J, Holcomb DD, Athey J, Santana-Quintero LV. Codon and Codon-Pair Usage Tables (CoCoPUTs): facilitating genetic variation analyses and recombinant gene design. *J Mol Biol* 2019;431:2434–2441.
- Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* 2008;3:38.
- Puigbò P, Romeu A, Garcia-Vallve S. HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. *Nucleic Acids Res* 2008;36:D524–527.
- Bourret J, Alizon S, Bravo IG. COUSIN (COdon Usage Similarity INdex): A Normalized Measure of Codon Usage Preferences. *Genome Biol Evol* 2019;11:3523–3528.
- Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV. A new and updated resource for codon usage tables. *BMC Bioinformatics* 2017;18:391.
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:119.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60.
- Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;25:1422–1423.
- UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47:D506–D515.
- Rocha J, Botelho J, Ksiezarek M, Perovic SU, Machado M. *Lactobacillus mulieris* sp. nov., a new species of *Lactobacillus delbrueckii* group. *Int J Syst Evol Microbiol* 2020;70:1522–1527.
- Putonti C, Shapiro JW, Ene A, Tsibere O, Wolfe AJ. Comparative genomic study of *Lactobacillus jensenii* and the newly defined *Lactobacillus mulieris* species identifies species-specific functionality. *mSphere* 2020;5:e00560–20.
- Salvetti E, Harris HMB, Felis GE, O’Toole PW. Comparative genomics of the genus *Lactobacillus* reveals robust phylogroups that provide the basis for reclassification. *Appl Environ Microbiol* 2018;84:e00993–18.

43. Roller M, Lucić V, Nagy I, Perica T, Vlahovicek K. Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic Acids Res* 2013;41:8842–8852.
44. Malki K, Kula A, Bruder K, Sible E, Hatzopoulos T. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. *Virology* 2015;12:164.
45. Jack BR, Boutz DR, Paff ML, Smith BL, Bull JJ. Reduced protein expression in a virus attenuated by codon deoptimization. *G3 (Bethesda)* 2017;7:2957–2968.
46. Kula A, Saelens J, Cox J, Schubert AM, Travisano M. The evolution of molecular compatibility between bacteriophage Φ X174 and its host. *Sci Rep* 2018;8:8350.
47. Ranjan A, Vidyarthi AS, Poddar R. Evaluation of codon bias perspectives in phage therapy of *Mycobacterium tuberculosis* by multivariate analysis. *In Silico Biol* 2007;7:423–431.
48. Chen Y, Batra H, Dong J, Chen C, Rao VB. Genetic engineering of bacteriophages against infectious diseases. *Front Microbiol* 2019;10:954.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.