

quickARSC: standalone package and web interface for profiling elemental stoichiometry of proteomes

Satoshi Nishino,^{1,2} Kento Tominaga,³ Yuki Nishimura,¹ Susumu Yoshizawa^{2,3}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Here, we present quickARSC, a command-line tool computing elemental composition per amino acid residue side chain (ARSC) from amino acid sequences. We also provide the web interface to enable users to access, filter, visualize, and download the pre-computed N-ARSC, C-ARSC, and S-ARSC of 143,614 prokaryotic species.

KEYWORDS ARSC, computational biology, python package, elemental stoichiometry, web interface

Microbial genomes and proteomes are shaped not only by functional requirements but also constrained by the elemental resource landscape (1, 2). These elemental constraints, commonly quantified using genomic guanine-cytosine proportion (GC content) and proteomic elemental composition per amino acid residue side chain (ARSC), serve as proxies for underlying elemental stoichiometry. For example, environmental resource availability drives shifts in genomic base composition to optimize nitrogen (N) investment (3, 4). Similarly, nutrient availability drives shifts in the N-ARSC and C-ARSC of proteomes (2, 5, 6). These stoichiometric traits enable a systematic quantification of nutrient investment across diverse microorganisms.

While GC content can be readily computed using many existing bioinformatics tools, such as SeqKit (7), the routine calculation of ARSC relies on custom in-house scripts (2, 5). The lack of a standardized, high-throughput tool limits the accessibility of stoichiometric profiling, creating a technical barrier to the broader adoption of these analyses into comparative (meta)genomics. To address this, we developed quickARSC, a standalone and platform-independent package that computes N-ARSC, C-ARSC, and S-ARSC from amino acid or nucleotide sequences. Calculations follow established definitions (1, 2, 5), computing average elemental counts per side chain from protein FASTA files (.faa and .faa.gz) or via integrated open reading frame prediction using Prodigal (8) for nucleotide files (e.g., .fna and .fna.gz). quickARSC is designed for high-throughput workflows, employing a multiprocessing module to distribute computations across CPU cores. With minimal core dependencies (only Biopython [9]) and a simple command-line interface, our tool can be readily incorporated into existing bioinformatics pipelines.

To demonstrate the utility of quickARSC, we computed ARSC for 143,614 representative prokaryotes from the Genome Taxonomy Database (GTDB) r226.0 (10) (Fig. 1A through F). Consistent with previous studies (2, 11), bacteria exhibited a positive correlation between genomic GC content and N-ARSC (Fig. 1A, Pearson's $r = 0.66$), and a negative correlation with C-ARSC (Fig. 1B, Pearson's $r = -0.93$). Bacteria with GC content below 30% (e.g., the genera *Shikimatogenerans*, *Stammera*, and *Spiroplasma*) exhibited elevated N-ARSC as previously reported (2). In contrast, archaea showed no overall correlation between GC content and N-ARSC (Fig. 1D, Pearson's $r = 0.16$). This decoupling is in line with the acidic proteome of halophilic archaea (12); for instance, the class *Halobacteria* shows high GC contents yet low N-ARSC (Fig. 1D). The S-ARSC peaked within the intermediate GC content range (Fig. 1C and F). This pattern is consistent

Editor Patrick D. Schloss, University of Michigan, Ann Arbor, Michigan, USA

Address correspondence to Satoshi Nishino, satoshi-nishino@ecc.u-tokyo.ac.jp, or Susumu Yoshizawa, yoshizawa@ecc.u-tokyo.ac.jp.

The authors declare no conflict of interest.

See the funding table on p. 3.

Ed. Note: This article was handled by Pat Schloss, who acted as a guest editor in consultation with the editor in chief, Irene Newton. Dr. Schloss's role as the guest editor of this article is outside of his duties as the chair of the Journals Committee.

Received 25 January 2026

Accepted 10 March 2026

Published 30 March 2026

Copyright © 2026 Nishino et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

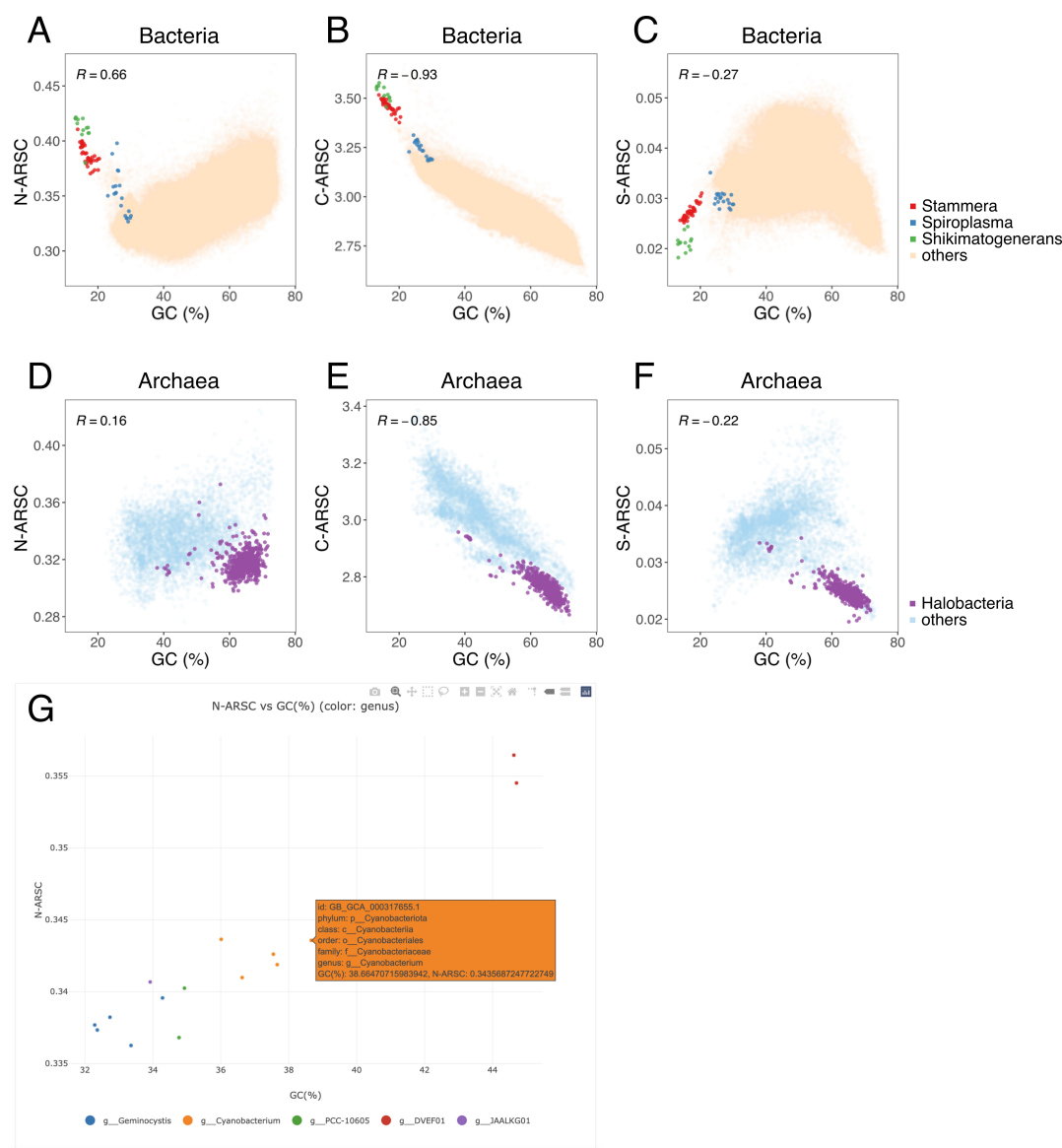


FIG 1 Relationship between genomic GC content and proteomic elemental stoichiometry across 143,614 prokaryotic species and an overview of the quickARSC web interface. Scatter plots display the average number of atoms per amino acid side chain (ARSC) against genomic GC content (%) for 136,646 bacteria and 6,968 archaea. (A) Bacterial N-ARSC, (B) bacterial C-ARSC, (C) bacterial S-ARSC, (D) archaeal N-ARSC, (E) archaeal C-ARSC, and (F) archaeal S-ARSC. Point colors are based on taxonomic information from the Genome Taxonomy Database release r226.0: the genera *Stammera* (red), *Spiroplasma* (blue), and *Shikimatogenerans* (green), and the class *Halobacteria* (purple). Other bacteria (light orange) and archaea (light blue) are shown with semi-transparency. Each panel shows Pearson's r ; all correlations are statistically significant ($P < 0.001$, calculated using the R package ggpubr). (G) The quickARSC web interface visualizing N-ARSC of the family *Cyanobacteriaceae*.

with expectations based on the standard genetic code (13); specifically, because the codons for Met (ATG) and Cys (TGT and TGC) contain both A/T and G/C bases, their frequencies are influenced by genomic base composition. These results illustrate that quickARSC reliably reproduces previously reported relationships between genomic base composition and proteomic elemental stoichiometry.

The pre-computed N-ARSC, C-ARSC, and S-ARSC are accessible through an interactive web interface (<https://stsnsn.github.io/quickARSC/>), enabling ARSC analysis of specified lineages (Fig. 1G). Additionally, it performs fast calculations based on a user input amino acid sequence and includes the ability to compare new results with pre-computed data.

Our package and web interface allow users to quantify the elemental composition of proteomes and provide insights into nutrient resource allocation.

ACKNOWLEDGMENTS

This work was supported by JST GteX grant number JPMJGX23B2 and JSPS KAKENHI grant numbers JP22H04925 and JP24KJ0979.

We acknowledge the use of GitHub Copilot for improving the readability of the code and Google Gemini for assisting in the grammar correction of the manuscript.

The authors take full responsibility for the final content and code.

AUTHOR AFFILIATIONS

¹Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

²Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba, Japan

³Department of Natural Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

AUTHOR ORCIDs

Satoshi Nishino  <http://orcid.org/0009-0007-2070-7678>

Kento Tominaga  <http://orcid.org/0000-0003-4814-5690>

Yuki Nishimura  <http://orcid.org/0000-0002-7990-1316>

Susumu Yoshizawa  <http://orcid.org/0000-0001-9667-6010>

FUNDING

Funder	Grant(s)	Author(s)
Japan Society for the Promotion of Science	JP24KJ0979	Satoshi Nishino
Japan Society for the Promotion of Science	JP22H04925	Yuki Nishimura
Japan Science and Technology Agency	JPMJGX23B2	Yuki Nishimura

AUTHOR CONTRIBUTIONS

Satoshi Nishino, Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review and editing | Kento Tominaga, Supervision, Writing – review and editing | Yuki Nishimura, Funding acquisition, Resources, Supervision, Writing – review and editing | Susumu Yoshizawa, Funding acquisition, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

quickARSC is openly available under the GNU General Public License (v2.0) on GitHub (<https://github.com/stsnsn/quickARSC/>) and was tested using Python 3.11.5 and Biopython 1.85. The package is distributed via PyPI (<https://pypi.org/project/arsc/>). Instructions and usage examples are provided in the GitHub repository. The pre-computed data set for 143,614 representative prokaryotes is available via our GitHub pages (<https://stsnsn.github.io/quickARSC/>). The visualization R code and the accompanying data set used in this study are archived in Zenodo (<https://doi.org/10.5281/zenodo.18758052>).

REFERENCES

1. Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D. 2001. Molecular evolution of protein atomic composition. *Science* 293:297–300. <https://doi.org/10.1126/science.1061052>
2. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, DeLong EF. 2017. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol* 2:1367–1373. <https://doi.org/10.1038/s41564-017-0008-3>

3. Rocha EPC, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. *Trends Genet* 18:291–294. [https://doi.org/10.1016/S0168-9525\(02\)02690-2](https://doi.org/10.1016/S0168-9525(02)02690-2)
4. Agashe D, Shankar N. 2014. The evolution of bacterial DNA base composition. *J Exp Zool Pt B* 322:517–528. <https://doi.org/10.1002/jez.b.22565>
5. Getz EW, Tithi SS, Zhang L, Aylward FO. 2018. Parallel evolution of genome streamlining and cellular bioenergetics across the marine radiation of a bacterial phylum. *mBio* 9:e01089-18. <https://doi.org/10.1128/mBio.01089-18>
6. Gralka M, Pollak S, Cordero OX. 2023. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nat Microbiol* 8:1799–1808. <https://doi.org/10.1038/s41564-023-01458-z>
7. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. <https://doi.org/10.1371/journal.pone.0163962>
8. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
9. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
10. Parks DH, Chaumeil P-A, Mussig AJ, Rinke C, Chuvpochina M, Hugenholtz P. 2026. GTDB release 10: a complete and systematic taxonomy for 715 230 bacterial and 17 245 archaeal genomes. *Nucleic Acids Res* 54:D743–D754. <https://doi.org/10.1093/nar/gkaf1040>
11. Baudouin-Cornu P, Schuerer K, Marlière P, Thomas D. 2004. Intimate evolution of proteins: proteome atomic content correlates with genome base composition. *J Biol Chem* 279:5421–5428. <https://doi.org/10.1074/jbc.M306415200>
12. Oren A. 2013. Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front Microbiol* 4:315. <https://doi.org/10.3389/fmicb.2013.00315>
13. Muto A, Osawa S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci USA* 84:166–169. <https://doi.org/10.1073/pnas.84.1.166>