

Taxonomie

Moderne genombasierte Klassifikation von Bakterien und Archaeen

JAN P. MEIER-KOLTHOFF
LEIBNIZ-INSTITUT DSMZ, BRAUNSCHWEIG

Prokaryotic taxonomy is increasingly informed by genome sequence-based approaches, but the proper choice of input data and methods is not trivial. The recently introduced Type (Strain) Genome Server (TYGS) is a user-friendly high-throughput webserver for genome-based taxonomy of prokaryotes, connected to a large, growing database of type-strain genomes and related data on taxonomy and nomenclature. TYGS helps scientists to quickly address various taxonomic questions.

DOI: 10.1007/s12268-019-0211-2
© Springer-Verlag 2019

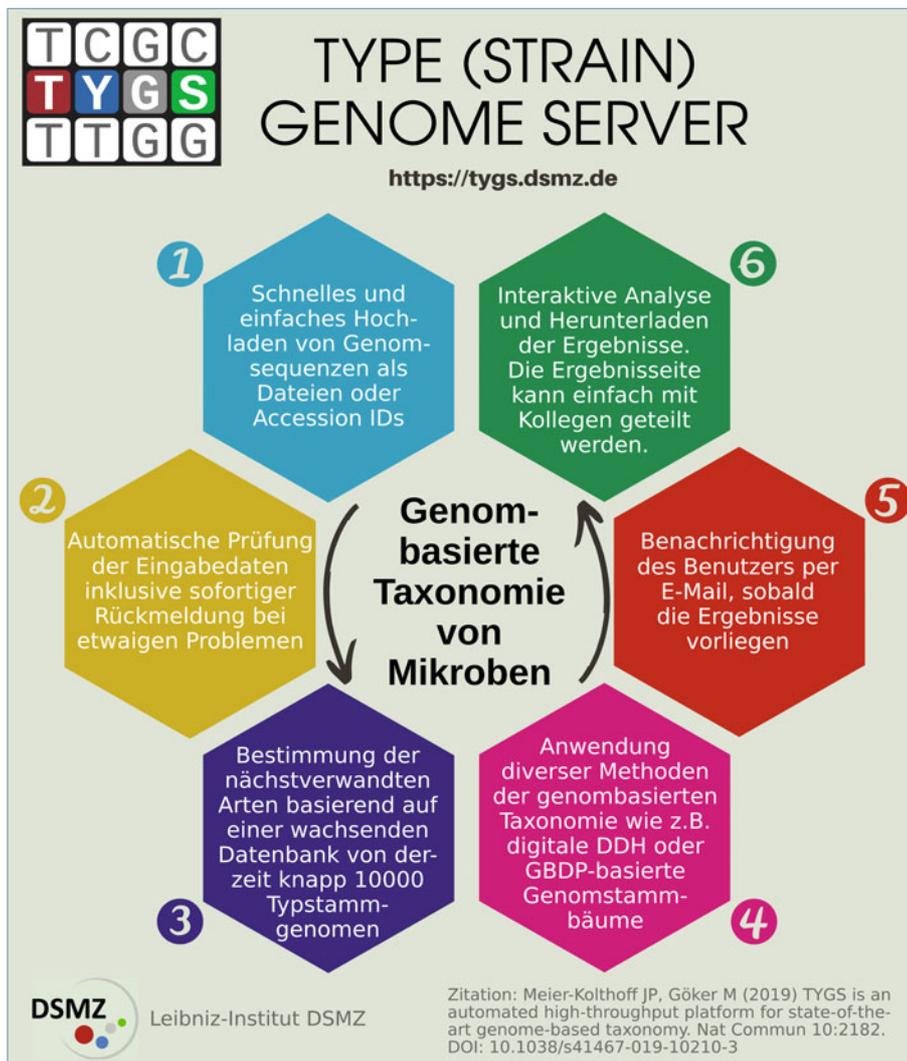
■ Mikroorganismen dominieren die Erde. Sie steuern und unterhalten das Leben auf vielfältige Weise [1]. Sie nehmen Schlüsselrollen in unterschiedlichen Stoffkreisläufen ein, im Leben von Pflanzen und Tieren, in der Landwirtschaft sowie im weltweiten Nahrungsnetz [1, 2]. Gerade in Zeiten des Klimawandels und der damit einhergehenden globalen Erwärmung ist ein genaues Verständnis sowohl der mikrobiellen Diversität als auch der ablaufenden und sich teils verändernden Prozesse von zentraler Bedeutung [2].

Entwicklung der bakteriellen Taxonomie

Das biologische Teilgebiet der Taxonomie befasst sich mit der Charakterisierung, Identifikation und Klassifikation von Organismen, mit dem Ziel, die natürlichen Verwandtschaftsverhältnisse möglichst akkurat abzubilden. Für die Taxonomie der Bakterien und Archaeen wurden bereits seit Ende des 19. Jahrhunderts unterschiedliche Laborverfahren entwickelt und im Laufe der Jahrzehnte immer weiter verfeinert [3]. Obwohl diese Methoden auf unterschiedliche Weise zur taxonomischen Positionierung einer Mikrobe beitragen, werden heutzutage zunehmend genomsequenzbasierte Methoden angewandt [4, 5]. Denn konventionelle Methoden, wie z. B. die DNA:DNA-Hybridisierung (DDH), sind für ihre Fehleranfälligkeit und dem resultierenden Arbeitsaufwand bekannt. Demgegenüber erlauben genomsequenzbasierte Alternativen eine hohe Standardisierung, da sie exakt reproduzierbar und schnell durchführbar sind sowie eine hohe methodische Genauigkeit mit sich bringen [4, 5].

Übergang zu genomsequenzbasierten Methoden

Die konventionelle DDH bestimmte über Jahrzehnte die prokaryotische Artabgrenzung im Rahmen des pragmatischen Artkonzepts, bei dem ein definierter Schwellenwert von 70 Pro-



◀ **Abb. 1:** Grundprinzip des Typstamm-Genomservers.

zent Ähnlichkeit über die Zugehörigkeit von Stämmen zur selben Art entscheidet [5]. Daher müssen neue genomsequenzbasierte Methoden hohe Korrelationen zu empirischen DDH-Ergebnissen aufweisen, um Konsistenz zwischen alter und jeweils neuer Methode sicherzustellen [6]. Die moderne, computergestützte digitale DDH (dDDH) vergleicht in Minutenschnelle Paare von Genomsequenzen mit entsprechend optimierten Algorithmen [5] und liefert wie die klassische DDH einen Ähnlichkeitswert, der ganz analog mittels des etablierten 70-Prozent-Grenzwerts zur Artabgrenzung genutzt werden kann [5, 7]. Ein alternatives Konzept, die *Average Nucleotide Identity* (ANI), basiert auf einem ähnlichen Grundprinzip; es fehlen aber eine Reihe wichtiger Funktionen und Optimierungen, die im dDDH-Ansatz enthalten sind [4, 5, 8]. Auch unterscheiden sich beide Ansätze in der jeweiligen Leistung [5, 8]. In Vergleichsstudien schnitt die dDDH-Methode durchweg besser ab als verschiedene ANI-Methoden, ohne dabei die technischen Nachteile der klassischen DDH zu kopieren [5, 8]. Viele neuere ANI-Varianten verzichten gänzlich auf eine Untersuchung der Korrelationen zu empirischen DDH-Ergebnissen [8], obwohl das zuständige taxonomische Komitee dies explizit forderte [6]. Die dDDH-Methode ist zudem als bisher einzige Methode für die Abgrenzung von Unterarten etabliert und stellt somit auch ein interessantes Werkzeug zur Untersuchung der Diversität innerhalb von Arten dar [7].

Darüber hinaus zeichnet sich die der dDDH zugrunde liegende *Genome BLAST Distance Phylogeny*-Methode (GBDP), die der Berechnung präziser Distanzen zwischen Genompaaren dient, durch verschiedene Optimierungen aus, die im ANI-Ansatz nicht enthalten sind, wie z. B. die Berücksichtigung von paralogenen Genen, von Sequenzbereichen geringerer Komplexität und von stark unterschiedlichen Genomgrößen [4, 5, 8]. Sie ist daher neben der Anwendung in der genom-basierten Art- und Unterartabgrenzung gerade für phylogénomische Analysen sehr gut geeignet, wie verschiedene Großstudien zeigten [9, 10].

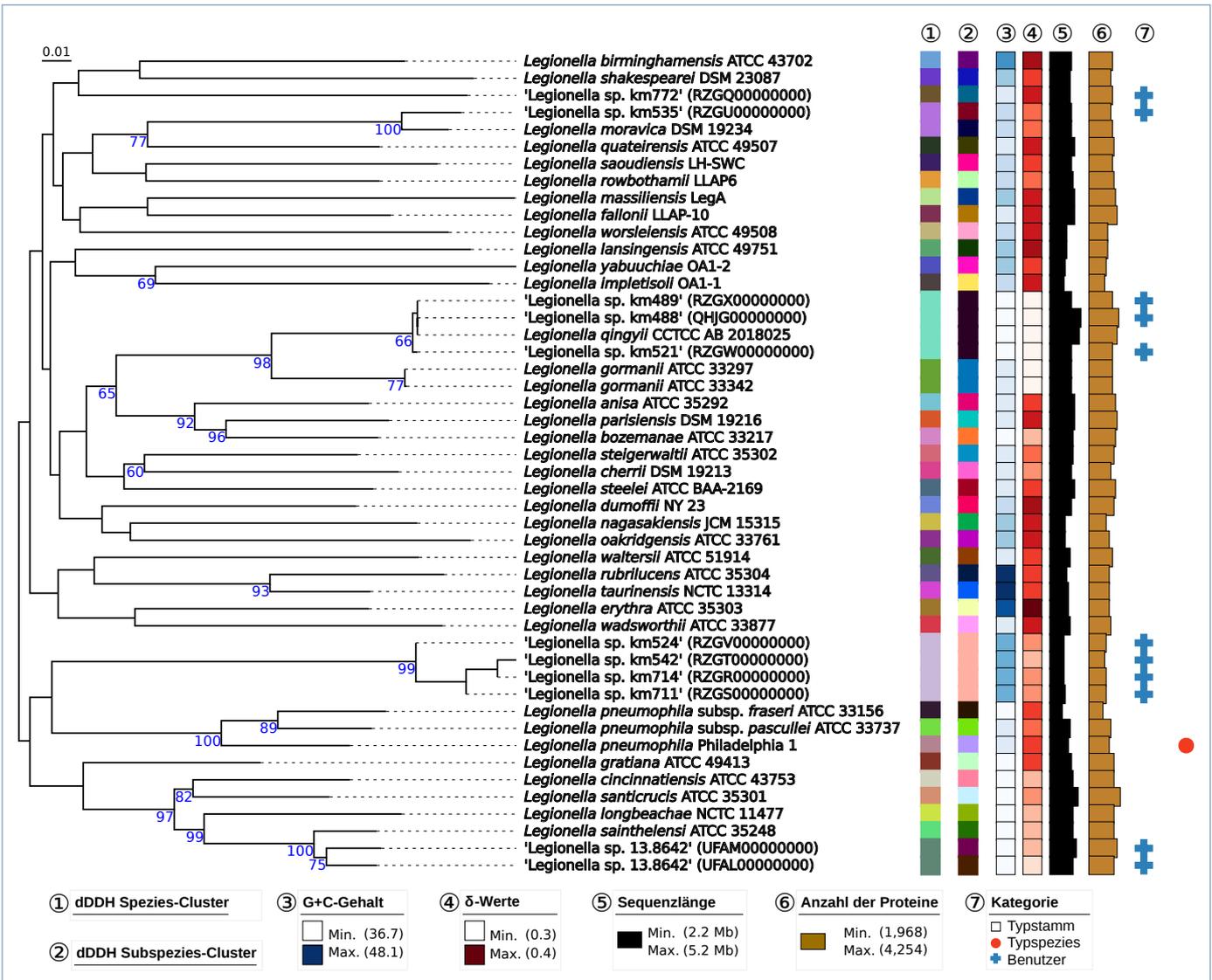
Herausforderungen und Lösungen

Neben der Wahl der passenden Methode ist in der prokaryotischen Taxonomie zudem die korrekte Wahl der Vergleichsstämme wichtig, die in Form von Typstämmen das Rückgrat der Systematik bilden und als nomenklatorische Typen von Arten und Unterarten fungieren. Ein Vergleich mit Typstämmen ist somit unerlässlich für die Identifikation und Klassifikation neuer Arten. Für genomsequenzbasierte Methoden bedeutet dies, dass die Genomsequenzen der jeweils relevanten Typstämme verfügbar sein müssen. Großsequenzierprojekte wie die *Genomic Encyclopedia of Bacteria and Archaea* (GEBA) konnten hier die Zahl der Genomsequenzierten Typstämme bereits deutlich erhöhen [10].

Die Synthese moderner Methoden mit einer umfangreichen Typstamm-Genomdatenbank würde daher mindestens zwei Probleme lösen: Genomsequenzen könnten so exakt einem bestimmten Typstamm zugeordnet werden, und eine angeschlossene bioinformatische Plattform würde Wissenschaftlern, selbst ohne technisches Spezialwissen, Zugriff auf verschiedene taxonomische Analysen ermöglichen.

Der Typstamm-Genomserver

Der von uns entwickelte Typstamm-Genomserver (TYGS; <https://tygs.dsmz.de>) stellt einen solchen integrierten Ansatz für genombasierte Taxonomie dar (**Abb. 1**, [8]). Er vereint verschiedene Komponenten, wie eine umfangreiche Typstamm-Genomdatenbank von Arten und Unterarten mit valide publizierten Namen, eine automatische Bestimmung der jeweils nächsten Verwandten zu benutzerdefinierten Stämmen und gesamtgenombasierte Methoden für Phylogenie und Klassifikation [8]. Dies ersetzt taxonomische Standardmethoden, wie die konventionelle DDH, Analysen des G+C-Gehalts, 16S-rRNA-Genanalysen sowie *multi locus sequence analysis* (MSLA), durch digitale Varianten und schafft eine Plattform, die einen deutlichen Mehrwert für Mikrobiologen, Taxonomen und Mediziner bietet [8].



▲ **Abb. 2:** Gesamtgenombasierter Stammbaum des *Legionella*-Datensatzes. Die Astlängen sind gemäß der Genome BLAST Distance Phylogeny (GBDP)-Distanzformel d_s skaliert. Die Zahlen oberhalb der Äste sind Unterstützungswerte, basierend auf 100 Replikaten. Die Blätter des Baums sind entsprechend der Legende annotiert. Die Zugriffsschlüssel (Accession IDs) sind in Klammern angegeben.

Analyse eines *Legionella*-Datensatzes

Eine übliche Frage in mikrobiologischen Forschungsprojekten ist, ob neu isolierte und Genom-sequenzierte Stämme neue Arten sind oder nicht. Dass dieser Aspekt und weitere einfach mit dem TYGS untersucht werden können, zeigt eine exemplarische Analyse eines Datensatzes der Gattung *Legionella*, die mit *Legionella pneumophila* auch den Erreger der Legionärskrankheit beinhaltet.

Drei der elf unbekannt Stämme gehören zu *L. qingyii*, ein Stamm entspricht *L. moravica*, während die restlichen sieben Stämme drei neuen *Legionella*-Arten angehören (**Abb. 2**). Dass es sich dabei um tatsächlich noch unbekannte Arten handelt, konnten wir mittels des TYGS anhand einer zusätzlichen

phylogenetischen Analyse eines erweiterten 16S-rRNA-Gen-Datensatzes sicherstellen [8].

Spannend ist zudem, dass laut TYGS die in der Klassifikation als Unterarten von *L. pneumophila* ausgewiesenen Stämme *L. pneumophila* subsp. *pascullei* und *L. pneumophila* subsp. *fraseri* gemäß der digitalen DDH eigene Arten darstellen (die dDDH-Werte liegen unter 55 Prozent) und somit entgegen der aktuellen Klassifikation keine Unterarten von *L. pneumophila* sind. Frühere Untersuchungen unterstützen diesen Fund mit Verweis auf die unterschiedliche Ökologie dieser Stämme [11].

Zusammenfassend bietet TYGS deutlich mehr als nur die Summe seiner Teile, indem verschiedene moderne Analysetechniken

kombiniert und vollautomatisch angewandt werden, Ergebnisse den Anwendern in aufbereiteter Form angeboten und einem weiten Nutzerkreis zugänglich gemacht werden. Die von TYGS bereitgestellten Ergebnisse können problemlos für weiterführende Analysen verwendet werden und bieten dem Anwender je nach Fragestellung ein großes Spektrum nachgelagerter Analysen. Die modulare Struktur des TYGS erlaubt einen zukünftigen Ausbau, wie beispielsweise die Extraktion und Annotation ausgewählter genomischer Merkmale.

Danksagung

Dank geht an die Deutsche Forschungsgemeinschaft, die diese Arbeit innerhalb des

Sonderforschungsbereichs TRR 51 unterstützt hat. ■

Literatur

- [1] Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive earth's biogeochemical cycles. *Science* 320:1034–1039
- [2] Cavicchioli R, Ripple WJ, Timmis KN et al. (2019) Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol* 17:569–586
- [3] Vandamme P, Pot B, Gillis M et al. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60:407–438
- [4] Meier-Kolthoff JP, Klenk H-P, Göker M (2014) Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* 64:352–356
- [5] Meier-Kolthoff JP, Auch AF, Klenk H-P et al. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60
- [6] Stackebrandt E, Frederiksen W, Garrity GM et al. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* 52:1043–1047
- [7] Meier-Kolthoff JP, Hahnke RL, Petersen J et al. (2014) Complete genome sequence of DSM 30083^T, the type strain (U5/41^T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 9:2
- [8] Meier-Kolthoff JP, Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182
- [9] Nouioui I, Carro L, García-López M et al. (2018) Genome-based taxonomic classification of the phylum *Actinobacteria*. *Front Microbiol* 9:1–119
- [10] Mukherjee S, Seshadri R, Varghese NJ et al. (2017) 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol* 35:676–683
- [11] David S, Sánchez-Busó L, Harris SR et al. (2017) Dynamics and impact of homologous recombination on the evolution of *Legionella pneumophila*. *PLoS Genet* 13:1–21

Korrespondenzadresse:

Dr. Jan P. Meier-Kolthoff
Leibniz-Institut DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
Abteilung Bioinformatik
Inhoffenstraße 7B
D-38124 Braunschweig
Tel.: 0531-2616-424
jan.meier-kolthoff@dsMZ.de
<https://jan.meier-kolthoff.com>

AUTOR



Jan P. Meier-Kolthoff

2001–2007 Bioinformatikstudium an der Universität Tübingen. 2007–2010 Bioinformatiker bei der GeneArt AG (Life Technologies) in Regensburg. 2010–2014 Promotion am Leibniz-Institut DSMZ in Braunschweig, dort seit 2014 Postdoc in der Abteilung Bioinformatik, AG Phylogenomik.