

Genomische Methoden

Besser und günstiger: Volumenreduktion in der Einzelzellgenomik

MORGAN S. SOBOL, GUNNAR STURM, ANNE-KRISTIN KASTER
 INSTITUT FÜR BIOLOGISCHE GRENZFLÄCHEN 5,
 KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT)

Single cell genomics (SCG) is a complementary method to metagenomics for exploring the genomes of uncultivated microorganisms. However, due to the minute amounts of DNA in the individual microbial cell, an amplification step is required before sequencing. Unfortunately, this reaction is notoriously costly and does not amplify all genomic regions equally well, preventing high-throughput applications and leading to incomplete and biased genomes. Here, we show a simple volume reduction approach to make SCG more feasible.

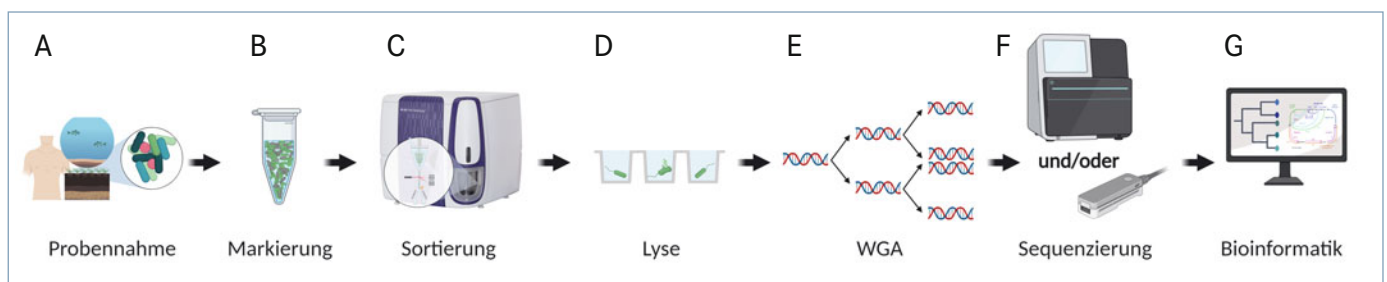
DOI: 10.1007/s12268-024-2126-9
 © Die Autorinnen und Autoren 2024

Die große Mehrheit prokaryotischer Spezies ist noch nicht ausreichend erforscht, da sie bisher nicht erfolgreich kultiviert werden können. Stoffwechsel, Funktionen in der Umwelt und biotechnologisches Potenzial sind daher oft unbekannt. Wir bezeichnen diese Mikroorganismen als mikrobielle dunkle Materie (*microbial dark matter*, MDM) [1]. Der Einsatz kultivierungsunabhängiger Methoden, wie z. B. der Metagenomik, hat unser Verständnis über diese Mikroorganismen erheblich verbessert. Aufgrund von Stammvariationen sowie der falschen Zuordnung von Sequenzen ist es mit bioinformatischen Methoden jedoch nach wie vor schwierig, individuelle Genome zuverlässig zusammensetzen. Darüber hinaus werden stark repetitive Sequenzen oft nicht präzise assembliert und 16S-rRNA-Sequenzen sowie

mobile genetische Elemente können häufig nicht zugeordnet werden [2]. Daher wurde die Einzelzellgenomik (*single cell genomics*, SCG) als ergänzende Methodik entwickelt, um die Analyse einzelner Zellen mikrobiologischer Gemeinschaften zu ermöglichen und so unser Wissen über MDM-Taxa zu erweitern (**Abb. 1**, [3]).

Bei der SCG ist allerdings ein Amplifikationsschritt erforderlich (*whole genome amplification*, WGA), um ausreichende Mengen an DNA für die Sequenzierung zu generieren, da eine typische mikrobielle Zelle nur wenige Femtogramm (fg) DNA enthält – eine Menge, die bislang noch nicht sequenziert werden kann. Im Laufe der Jahre wurden verschiedene WGA-Methoden entwickelt und verbessert (**Abb. 2**). Diese Methoden lassen sich in die Kategorien Polymerase-Ketten-

reaktion(PCR)-basierte Amplifikation, isotherme Amplifikation und Hybridverfahren, die beide Methoden kombinieren, einteilen. Die gängigste Methode zur Amplifikation von DNA aus einer einzelnen Zelle ist die *multiple displacement amplification* (MDA). Bei der MDA handelt es sich um eine isotherme Reaktion mit einer High-Fidelity-phi29-Polymerase. Im Vergleich zu den Standard-PCR-Polymerasen weist diese eine deutlich geringere Fehlerrate auf, verfügt über eine 3'→5'-Exonukleaseaktivität und ist in der Lage, Fragmente mit einer Länge von >10 kbp zu erzeugen. Leider werden Regionen mit hohem GC-Gehalt aufgrund der geringen Temperatur nicht gut amplifiziert, was zu einer ungleichmäßigen Genomabdeckung führt [4]. Darüber hinaus können Artefakte wie Chimären und unspezifische Produkte entstehen, von denen man annimmt, dass sie zufällig auftreten, da Sequenzen, die in einer MDA-Reaktion überrepräsentiert sind, in einer anderen unterrepräsentiert sein können. Zur besseren Amplifikation von Organismen mit hohem GC-Gehalt kann bei der WGA-X-Reaktion eine thermostabilere phi29-Polymerase verwendet werden [5]. Allerdings wird hier über eine geringere Genomabdeckung für Organismen mit niedrigem GC-Gehalt im Vergleich zur Standard-MDA berichtet. In jüngerer Zeit wurde außerdem die primäre Template-gesteuerte Amplifikation (PTA) entwickelt, bei der Exonuklease-resistente Terminatoren eingesetzt werden, um kleinere Amplikons zu erzeugen [6]. Dies begrenzt zum einen die Überrepräsentation



▲ Abb. 1: Pipeline für prokaryotische Einzelzellgenomik. **A**, Probenentnahme und -konservierung. **B**, spezifische oder unspezifische Zellfärbung. **C**, Zellsortierung. **D**, Zellylyse. **E**, Amplifikation (WGA). **F**, Genomsequenzierung. **G**, bioinformatische Analyse.

Methode	MDA	WGA-X™	PTA	MALBAC
Modus	Isotherm	Isotherm	Isotherm	Hybrid (isotherm & PCR)
Enzym	phi29	EquiPhi29™	phi29	Bst & Taq
Art der Amplifikation	Exponentiell	Exponentiell	Quasi-linear	Quasi-linear
Produkt Länge (bp)	>10,000	>10,000	250-1500	500-1500
Ø Genomabdeckung (<i>E. coli</i>)	~10-80%	36 ± 21%	≥92%	~80%
Reaktionsvolumen*	50 µl	10 µl	20 µl	65 µl
Kosten 1 µl Reaktion	0.48 \$	0.14 \$	1.5 \$	0.72 \$

▲ **Abb. 2:** Unterschiedliche WGA-Methoden. Schematische Darstellung und relevante Charakteristika der unterschiedlichen Reaktionen MDA (*multiple displacement amplification*), WGA-X™ (*whole genome amplification-X*), PTA (*primary template-directed amplification*) und MALBAC (*multiple annealing and looping based amplification cycles*).

zufälliger Positionen und verringert zum anderen die Fehlerfortpflanzung. Diese Methode scheint vielversprechend, um die Amplifikationsverzerrung zu reduzieren, befindet sich aber noch in der Testphase für Mikroorganismen und ist zudem teuer. Eine Hybridmethode, *multiple annealing and looping-based amplification cycles* (MALBAC), kombiniert PCR- und MDA-Methoden, um Amplifikationsverzerrungen erfolgreich zu verringern. Dennoch bleibt MALBAC in mikrobiellen SCG weitgehend ungenutzt, da die Bst- und Taq-Polymerasen aufgrund ihrer fehlenden Korrekturlesefähigkeit höhere Fehlerraten aufweisen [7]. Ein weiteres Problem ist, dass die empfohlenen Reaktionsvolumina der WGA-Methoden ihre Anwendung sehr kostspielig werden lassen, wenn sie auf SCG mit hohem Durchsatz angewendet werden. Außerdem bergen sie das Risiko für eine erhöhte Kontamination. Auch verhindert die begrenzte Anzahl von Zellen, die analysiert werden können, dass niedrig-abundante Taxa einer Probe mit SCG erfasst werden [8].

Eine methodisch einfache Lösung für dieses Problem besteht darin, das Reaktionsvolumen der WGA zu verringern. In früheren Studien wurde dies bei Volumina von weni-

ger als einem Nanoliter (nl) und einem Pikoliter (pl) in mikrofluidischen Geräten, Nanowells, auf planaren Oberflächen, sowie in Hydrogelen gezeigt. Viele dieser Ansätze blieben außerhalb ihrer jeweiligen Veröffentlichungen jedoch ungenutzt – wahrscheinlich, weil die meisten mikrofluidischen Setups und andere Plattformen nicht kommerziell erhältlich sind. Frühere Studien zeigen auch, dass die Volumenreduzierung die Spezifität der Polymerase durch „molekulares Crowding“ verbessert. Darüber hinaus verringern geringere Reaktionsvolumina die Oberfläche für die unspezifische Adsorption von Nukleinsäuren an den Wänden der Multiwell-Platten. Ein zu großes „Gedränge“ kann jedoch auch nachteilige Auswirkungen haben, indem es eine sterische Hinderung verursacht und den Zugang der Polymerase zur DNA erschwert [9].

In unserer Studie [10] haben wir einzelne *Escherichia coli*-Zellen in 384-Well-Platten sortiert, ihre DNA in verschiedenen Reaktionsvolumina amplifiziert und die Einzelzellgenome (*single amplified genomes*, SAGs) bioinformatisch verglichen (**Abb. 3**). Unsere Ergebnisse deuten darauf hin, dass ein MDA-Reaktionsvolumen von 1,25 µl der „Sweet-

Spot“ für eine signifikante Verringerung der Amplifikationsverzerrung und eine Erhöhung der Genomabdeckung auf fast 90 % ist. Die schlechteren Ergebnisse in den kleineren Reaktionsvolumina sind wahrscheinlich auf Verdunstung und/oder sterische Hinderung der Polymerase zurückzuführen. Um die gleichmäßige Verteilung der Abdeckung im gesamten Genom zu bewerten, wurden die Reads in 10-Kilobasen(kb)-Bins gemittelt und ihre Lesetiefen aufgetragen, um die Abdeckungstiefe für jedes Reaktionsvolumen zu visualisieren. Die Abdeckung des Genoms durch Reads ist in größeren Volumina deutlich unvollständiger. Zum Vergleich: WGA-X-Reaktion ermöglichte eine *E. coli*-Genomabdeckung von <60 Prozent, selbst bei ~5× mehr sequenzierten Reads. Geringere sowie höhere Volumina führen außerdem zu geringeren Sequenzieretiefen, da ein größerer Anteil Reads während der Qualitätskontrolle herausgefiltert wird, weshalb die resultierenden Assemblierungen eine geringere Qualität im Vergleich zur 1,25-µl-Reaktion zeigen. Bei 10 µl lagen unsere Assemblierungsabdeckungen im Bereich der Werte von WGA-X in 10-µl-Reaktionen publizierten Werte, was zeigt, dass WGA-X auch von einer weiteren

Volumenreduzierung profitieren könnte. Auch hier funktioniert die Reaktion im 1,25- μ l-Maßstab besser (unveröffentlichte Daten). Im Vergleich zu anderen Ansätzen zur Volumenreduktion lagen unsere Assemblierungsabdeckungen im Bereich der Reaktionen in pl-Tröpfchen (88–91%) und nl-Wells (88–94 %) bei ähnlichen Sequenztiefen.

Insgesamt zeigen diese Ergebnisse, dass die MDA in einem Reaktionsvolumen von 1,25 μ l erheblich verbessert wird, indem sie deutlich weniger verfälschte, weniger verunreinigte und vollständigere SAGs produziert als die größeren Standardreaktionsvolumina. Unsere Ergebnisse zeigen außerdem, dass eine weitere Volumenreduzierung in spezialisierten und komplexen Systemen (z. B. mikrofluidischen Chips) wahrscheinlich nicht notwendig ist, um mikrobielle Genome von höherer Qualität zu erhalten. Daher sollte jeder für sich selbst abschätzen, ob die

Zeit- und Kostenvorteile einer Volumenreduktion auf nl- und pl-Reaktionen im Rahmen einer Studie sinnvoll sind. Die Volumenreduzierung in Standard-384-Well-Platten und mit handelsüblichen Zellsortierern und Flüssigkeitsdispensern macht diesen Ansatz für andere Forscher und Forscherinnen leichter zugänglich und reduziert die Kosten drastisch um ~97,5 % gegenüber der Standardreaktion. Wir haben auch herausgefunden, dass bei unserem Ansatz eine 40-fache Sequenztiefe für qualitativ hochwertige Assemblierungen ausreicht, verglichen mit den üblichen >100-fachen Sequenztiefen, die im Allgemeinen bei mikrobiellen SCG verwendet werden. Letztendlich gehen wir davon aus, dass die hier vorgenommenen Verbesserungen auch für andere Einzelzellstudien von großem Interesse sein und daher den Einsatz von SCG erhöhen werden, insbesondere für Forschungsarbeiten, die

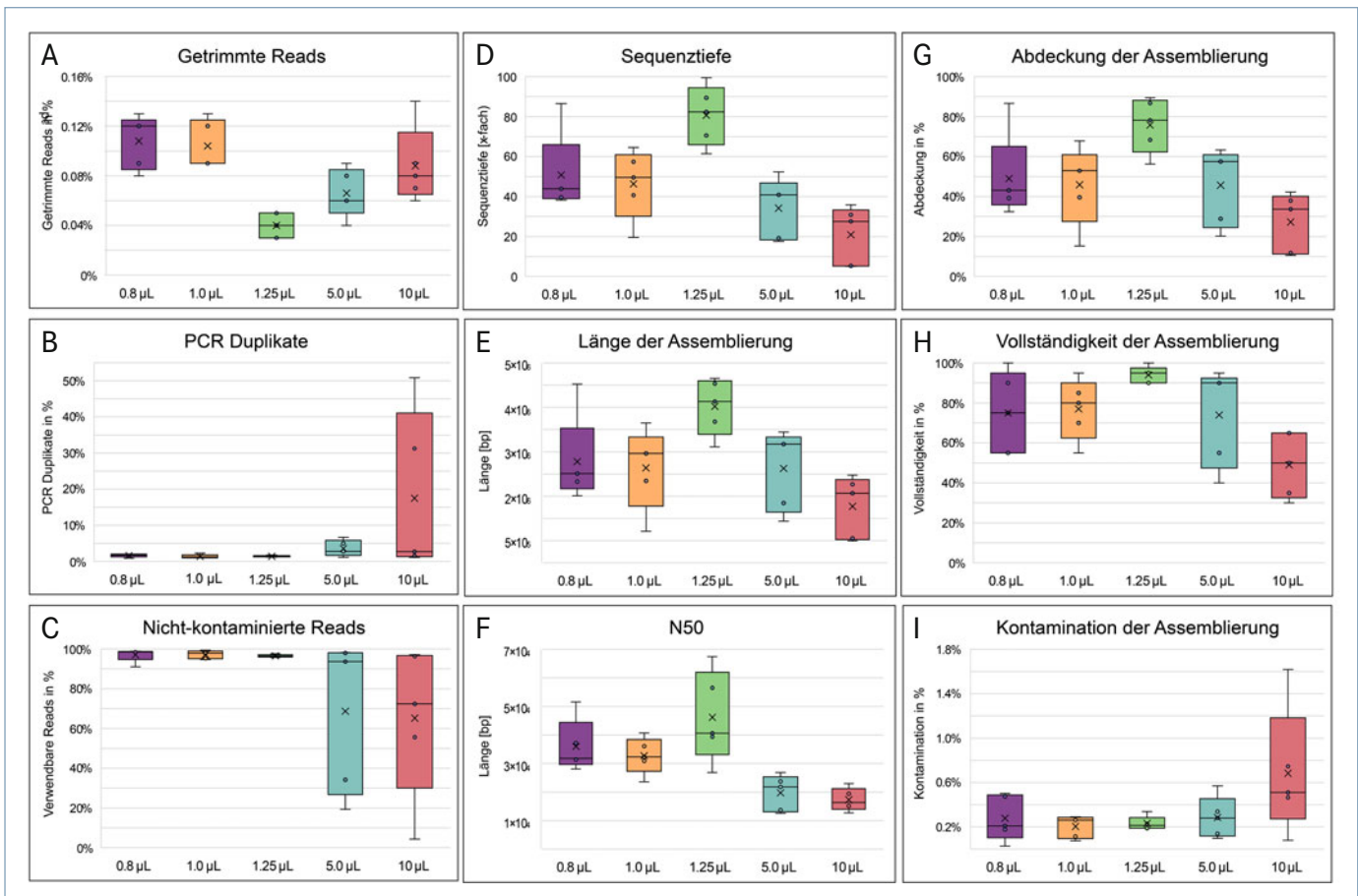
sich auf die Aufklärung des genomischen Potenzials seltener Taxa und/oder neuartiger MDM in Umweltproben konzentrieren. ■

Literatur

- [1] Hug LA, Baker BJ, Anantharaman K et al. (2016) A new view of the tree of life. *Nat Microbiol* 2016 15 1:1–6
- [2] Vollmers J, Wiegand S, Lenk F, Kaster AK (2022) How clear is our current view on microbial dark matter? (Re-) assessing public MAG & SAG datasets with MDMcleaner. *Nucleic Acids Res* 50: e76–e76
- [3] Kaster AK, Sobol MS (2020) Microbial single-cell omics: the crux of the matter. *Appl Microbiol Biotechnol* 104: 8209–8220
- [4] Sabina J, Leamon JH (2015) Bias in whole genome amplification: Causes and considerations. *Methods Mol Biol* 1347: 15–41
- [5] Stepanauskas R, Fergusson EA, Brown J et al. (2017) Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat Commun* 8: 1–10
- [6] Gonzalez-Pena V, Natarajan S, Xia Y et al. (2021) Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc Natl Acad Sci USA* 118: e2024176118
- [7] De Bourcy CFA, De Vlamincq I, Kanbar JN et al. (2014) A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. *PLoS One* 9: e105585

Hier steht eine Anzeige.

 Springer



▲ **Abb. 3:** Read-Prozessierungs- (A–C) und Assemblierungsstatistik (D–I) von *E. coli* SAGs. **A**, prozentualer Anteil der während der Qualitätskontrolle entfernten Reads. **B**, prozentualer Anteil der entfernten PCR-Duplikate. **C**, prozentualer Anteil der Reads nach dem Herausfiltern der Kontaminationen. **D**, finale Sequenztiefe. **E**, durchschnittliche Gesamtlänge der Assemblierungen. **F**, N50-Wert, Länge des kleinsten Contigs, der 50 % Abdeckung der Gesamtnukleotide der Assemblierung repräsentiert. **G**, prozentuale Abdeckung des Genoms von *E. coli*. **H**, Vollständigkeit der Assemblierungen. **I**, prozentuale Kontamination der Assemblierungen. Die Linie innerhalb der Boxen zeigt den Median, das x den Mittelwert aus 5 unabhängigen Replikaten.

[8] Dam HT, Vollmers J, Sobol MS et al. (2020) Targeted Cell Sorting Combined With Single Cell Genomics Captures Low Abundant Microbial Dark Matter With Higher Sensitivity Than Metagenomics. *Front Microbiol* 11: 511433

[9] Kuznetsova IM, Turoverov KK, Uversky VN (2014) What Macromolecular Crowding Can Do to a Protein. *Int J Mol Sci* 15: 23090–23140

[10] Sobol MS, Kaster AK (2023) Back to Basics: A Simplified Improvement to Multiple Displacement Amplification for Microbial Single-Cell Genomics. *Int J Mol Sci* 24: 4270

Funding note: Open Access funding enabled and organized by Projekt DEAL.
Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/ die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Korrespondenzadresse:

Prof. Dr. Anne-Kristin Kaster
 Institute für Biologische Grenzflächen 5
 Karlsruher Institute für Technology (KIT)
 Hermann von Helmholtz Platz 1
 D-76344 Eggenstein-Leopoldshafen
 kaster@kit.edu

AUTORINNEN UND AUTOREN



Morgan Sobol

2013–2016 B.Sc. in Biologie an der Texas A&M University, College Station, TX, USA. 2016–2018 M.Sc. in Meeresbiologie an der Texas A&M University. 2022 Promotion, Karlsruher Institute für Technology (KIT). Seit 2023 NASA Postdoctoral Fellow, University of Wisconsin, USA.



Gunnar Sturm

2009 Diplom Biologie, Universität Freiburg. 2014 Promotion, Karlsruher Institute für Technology (KIT). 2014–2022 PostDoc, KIT. Seit 2022 Gruppenleiter Institut für Biologische Grenzflächen IBG-5 am KIT.



Anne-Kristin Kaster

2007 Diplom in Chemie, Universität Marburg. 2011 Promotion, Max-Planck-Institut für terrestrische Mikrobiologie. 2011–2014 PostDoc, Stanford University, CA, USA. 2014–2017 Nachwuchsgruppenleiterin Leibniz-Institut DSMZ. Seit 2017 W3-Professorin und Institutsleiterin IBG-5 am Karlsruher Institute für Technology (KIT).