

Erlernen der regulatorischen Grammatik von Pflanzen

Tobias Jores

Article - Version of Record



Suggested Citation:

Jores, T. (2024). Erlernen der regulatorischen Grammatik von Pflanzen. *Biospektrum*, 30(4), 390–392.
<https://doi.org/10.1007/s12268-024-2210-1>

Wissen, wo das Wissen ist.



UNIVERSITÄTS- UND
LANDESBIBLIOTHEK
DÜSSELDORF

This version is available at:

URN: <https://nbn-resolving.org/urn:nbn:de:hbz:061-20250225-123036-9>

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: <https://creativecommons.org/licenses/by/4.0>

Pflanzliche Genregulation

Erlernen der regulatorischen Grammatik von Pflanzen

TOBIAS JORES

INSTITUT FÜR SYNTHETISCHE BIOLOGIE UND EXZELLENZCLUSTER
FÜR PFLANZENWISSENSCHAFTEN (CEPLAS), UNIVERSITÄT DÜSSELDORF

Faced with accelerating climate change and rapid population growth, we need crops with higher yields and greater resilience to ensure food security. Crop genome engineering will likely play a major role in meeting future food needs. However, we do not understand plant gene regulation well enough to target engineering and achieve predictable outcomes. Therefore, we study regulatory DNA and its interactions – the regulatory grammar – in plants using high-throughput assays and computational approaches.

DOI: 10.1007/s12268-024-2210-1
© Der Autor 2024

Die präzise Steuerung der Genexpression ist von entscheidender Bedeutung für das Wachstum, die Entwicklung und die Anpassung an die Umwelt. Sowohl in Pflanzen als auch in Tieren wird die Genexpression durch *cis*-regulatorische Elemente kontrolliert, wie Promotoren, Enhancer, Silencer und Insulatoren. Die Genom-Editierung solcher Elemente hat das Potenzial, schnelle Verbesserungen von Nutzpflanzen zu ermöglichen und damit zur Ernährungssicherheit beizutragen [1]. Auch der Evolution und Domestikation moderner Nutzpflanzen lagen oft Veränderungen von regulatorischer DNA zugrunde [2]. Um die Genom-Editierung von regulatorischen Elementen routinemäßig zur Verbesserung von Nutzpflanzen einsetzen zu können, fehlt uns jedoch bisher ein ausreichendes Verständnis der pflanzlichen Genregulation.

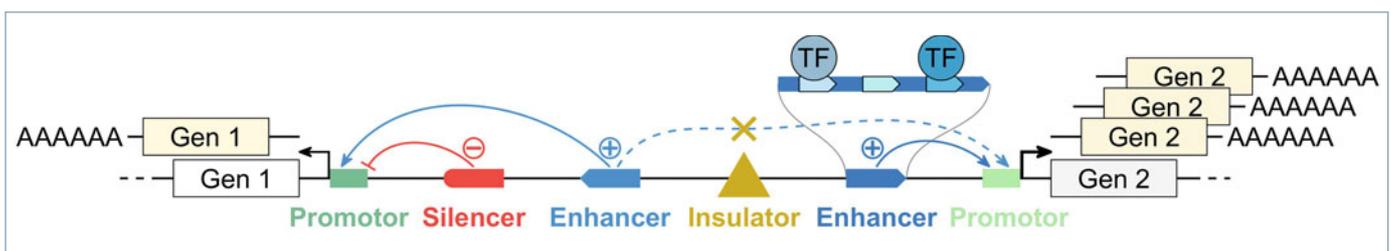
Genregulation ist komplex

Der genetische Code, welcher der Übersetzung von DNA- zu Proteinsequenzen zugrunde liegt, wurde bereits in den Sechzigerjahren entschlüsselt. Im Gegensatz dazu steht die Entschlüsselung des genregulatorischen Codes, der die Verbindung zwischen DNA-Sequenzen und Genexpression herstellt, noch aus. Das liegt vor allem an seiner hohen Komplexität (**Abb. 1**). Verschiedene *cis*-regulatorische Elemente beeinflussen die Genexpression auf unterschiedliche Weise [3]: Kern-Promotoren rekrutieren die Transkriptionsmaschinerie und etablieren ein Grundniveau der Transkription. Dieses Transkriptionsniveau kann durch Enhancer erhöht oder durch Silencer verringert werden. Insulatoren können das Zusammenspiel zwischen umgebenden *cis*-regulatorischen Elementen blockieren. Um das finale Transkrip-

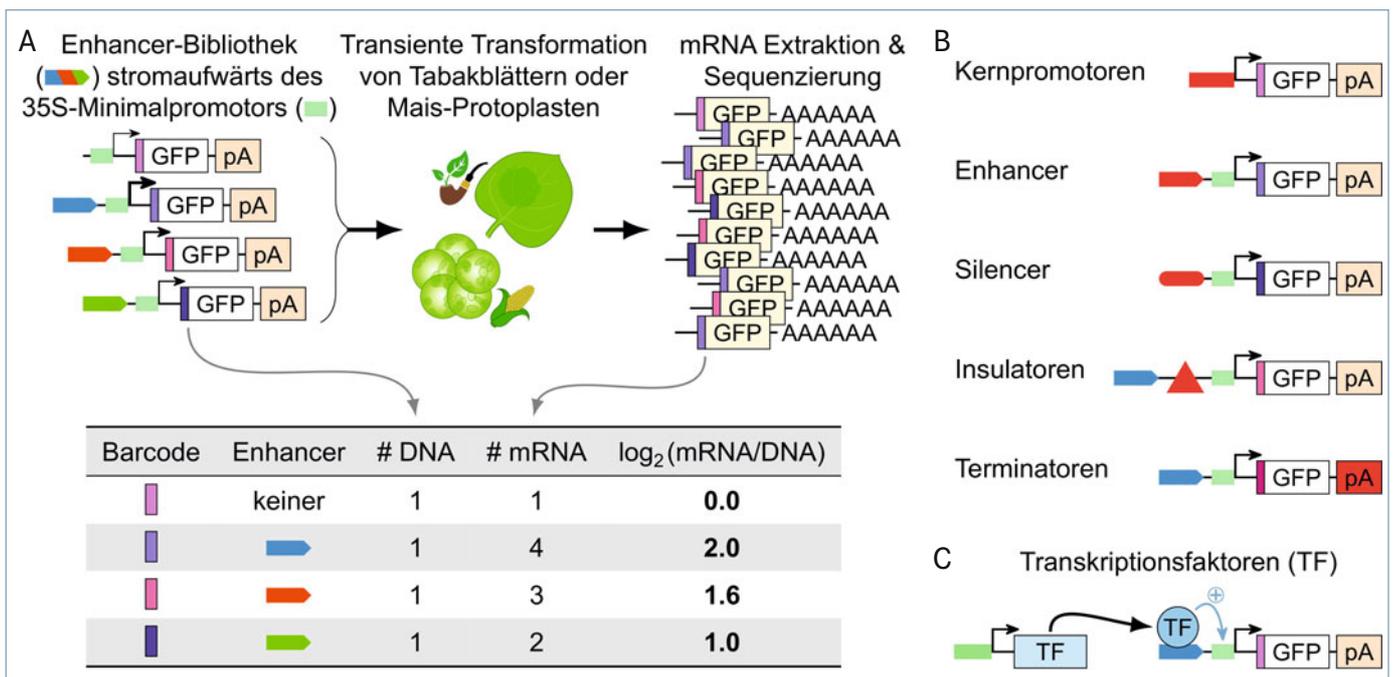
tionsniveau zu erreichen, müssen die Aktivitäten aller beteiligten Elemente integriert werden. Die einzelnen regulatorischen Elemente selbst bestehen wiederum aus mehreren funktionalen Einheiten, wie z. B. Bindestellen für Transkriptionsfaktoren. Ob und von welchem Transkriptionsfaktor die jeweiligen Bindestellen besetzt sind, hängt dabei u. a. vom Gewebe und den Umweltbedingungen ab. Im Gegensatz zum nahezu universell gültigen genetischen Code ist der genregulatorische Code also kontextabhängig. Zu guter Letzt ist die Genexpression eine von Natur aus quantitative Eigenschaft. Es spielt nicht nur eine Rolle, ob ein Gen überhaupt exprimiert wird, sondern auch wie stark. Selbst eine relative kleine Anpassung im Transkriptionsniveau eines Gens kann zu deutlichen phänotypischen Veränderungen führen. Aufgrund dieser Komplexität brauchen wir zur Untersuchung des genregulatorischen Codes Methoden, welche die regulatorische Aktivität vieler verschiedener Elemente, allein und in Kombination mit zusätzlichen Elementen oder Transkriptionsfaktoren, quantitativ messen können.

Plant STARR-seq: ein Hochdurchsatzverfahren zur Untersuchung regulatorischer DNA

Wir haben kürzlich Plant STARR-seq entwickelt, ein Hochdurchsatzverfahren mit dem die Aktivität von regulatorischer DNA gemessen werden kann [4]. Dazu wird eine Bibliothek von Enhancer-Kandidaten stromaufwärts von einem 35S-Minimalpromotor kloniert. Der Promotor kontrolliert die Expres-



▲ Abb. 1: Die Genexpression wird durch das komplexe Wechselwirkungsnetzwerk zwischen *cis*-regulatorischen Elementen (z. B. Promotoren, Enhancer, Silencer und Insulatoren) und *trans*-aktivierenden Proteinen (z. B. Transkriptionsfaktoren, TF) kontrolliert.



▲ **Abb. 2:** Plant STARR-seq – ein Hochdurchsatzverfahren zur Untersuchung regulatorischer DNA. **A**, Ein mit einem DNA-Barcode versehenes Reporter-Gen wird mit einer Enhancer-Bibliothek kombiniert. Per Hochdurchsatz-Sequenzierung wird bestimmt, wie oft die Barcodes in der DNA und mRNA von transformierten Pflanzenzellen vorkommen. Das mRNA/DNA-Verhältnis gibt Aufschluss über die Enhanceraktivität. **B**, Plant STARR-seq kann die Aktivität verschiedener regulatorischer Elemente (rot) messen. **C**, Die Aktivität von koexprimierten Transkriptionsfaktoren kann mit Plant STARR-seq bestimmt werden.

sion eines mit einem DNA-Barcode versehenen Reporter-Gens. Alle Reporter-Konstrukte werden gemischt und zur transienten Transformation von Tabakblättern oder Mais-Protoplasten verwendet. Nach ein bis zwei Tagen Inkubation in den Pflanzenzellen werden die mRNAs des Reporter-Gens extrahiert und zusammen mit der DNA der Reporter-Bibliothek sequenziert. Das Verhältnis der Häufigkeit eines Barcodes in der mRNA relativ zur DNA gibt Auskunft über die Aktivität des entsprechenden Enhancers (**Abb. 2A**). Durch die Verwendung von DNA-Barcodes und Hochdurchsatz-Sequenzierung können bis zu einer Million verschiedene Enhancer-Kandidaten gleichzeitig in einem Experiment analysiert werden. Neben Enhancern kann mit Plant STARR-seq auch die Aktivität von diversen anderen regulatorischen Elementen gemessen werden (**Abb. 2B**). Wir haben die Methode z. B. zur Bestimmung der Stärke von über 75.000 Kernpromotoren von *Arabidopsis*, Mais und Sorghum verwendet [5]. Selbst Sequenzen, die klassischerweise nicht zu den *cis*-regulatorischen Elementen gezählt werden, aber mRNA-Level beeinflussen können, wie beispielsweise Terminatoren, lassen sich mit Plant STARR-seq untersuchen [6].

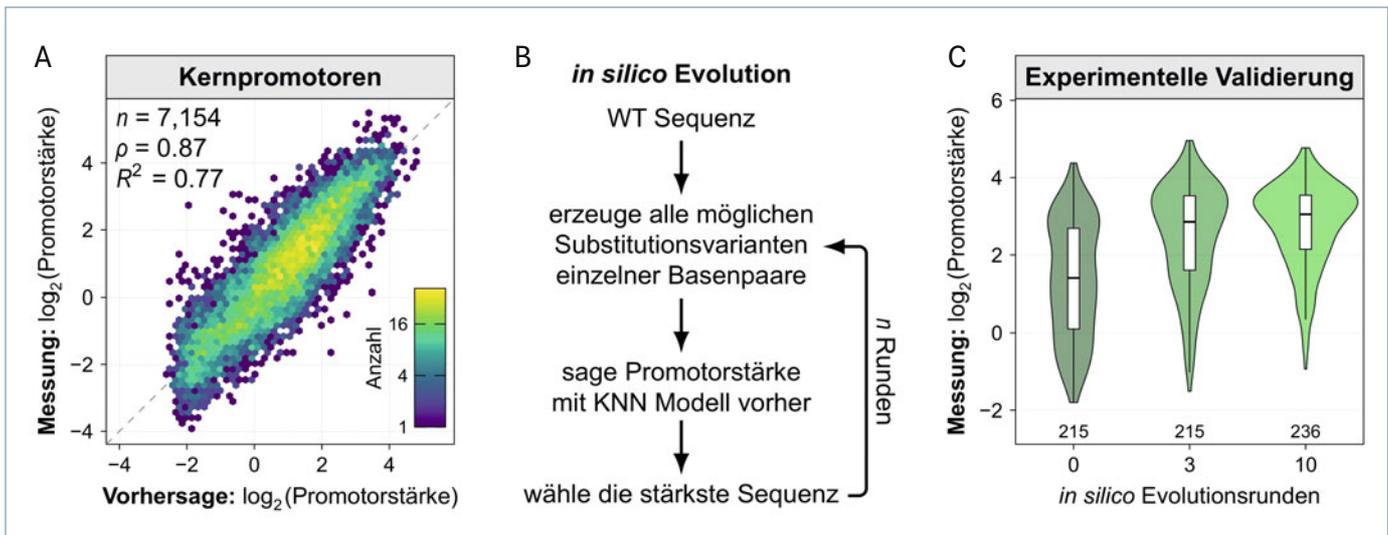
Statt zur Charakterisierung von vielen regulatorischen Elementen gleichzeitig kann

Plant STARR-seq auch verwendet werden, um einzelne Elemente im Detail zu studieren. Durch die Kombination von Plant STARR-seq mit Sättigungsmutagenese identifizierten wir in drei Enhancern von Photosynthese-assoziierten Genen die Regionen, die entscheidend zur Aktivität der Enhancer beitragen [7]. Ausgehend von diesen Plant STARR-seq-Daten haben wir Sequenzlogos erstellt, welche die quantitativen Beiträge aller Nukleotide innerhalb einer Region zu ihrer Gesamtaktivität beschreiben. Die Sequenzlogos konnten bekannten Motiven zur Bindung von Transkriptionsfaktoren zugeordnet werden; dabei war jedoch auffällig, dass die Enhancer-Sequenzen oft von den optimalen Bindemotiven abweichen. Solche suboptimalen Bindestellen sind vermutlich nötig, um sicherzustellen, dass volle Enhanceraktivität nur dann erreicht wird, wenn alle relevanten Transkriptionsfaktoren anwesend und aktiv sind. Kooperative Wechselwirkungen zwischen den Transkriptionsfaktoren führen dazu, dass diese, trotz reduzierter Affinität zu den jeweiligen Bindestellen, an den Enhancer binden können. Die Verwendung von suboptimalen Transkriptionsfaktor-Bindestellen zur Feinjustierung der Enhanceraktivität ist auch aus tierischen Systemen bekannt [8,9].

Rechnergestützte Verfahren unterstützen die Erforschung der pflanzlichen Genregulation

Aufgrund der Komplexität des genregulatorischen Codes bedarf es nicht nur neuartiger experimenteller Ansätze, sondern auch moderner computergestützter Verfahren zur Untersuchung der regulatorischen Grammatik [10]. Die Menge der Daten, die von Hochdurchsatzverfahren wie Plant STARR-seq erzeugt wird, erfordert fortschrittliche rechnergestützte Methoden für die Analyse, Interpretation und Visualisierung. Der Einsatz von Deep Learning, einer Spezialform des maschinellen Lernens mit künstlichen neuronalen Netzwerken, hat sich besonders für diesen Zweck bewährt. Neuronale Netzwerke, die mit Daten aus Plant STARR-seq trainiert wurden, können die Aktivität von neuen Sequenzen vorhersagen, die zugrunde liegenden funktionellen DNA-Motive identifizieren und die regulatorische Aktivität durch *in silico*-Evolution verbessern (**Abb. 3**; [5, 6]).

Deep-Learning-Modelle können auch mit genomischen Datensätzen trainiert werden, um beispielsweise Bindestellen für Transkriptionsfaktoren oder die Zugänglichkeit von Chromatin vorherzusagen. Auch bei der Vorhersage der Genexpression basierend auf



▲ Abb. 3: Plant STARR-seq + Deep Learning: eine leistungsfähige Kombination. **A,** Ein künstliches neuronales Netzwerk (KNN) wurde mit Plant STARR-seq-Daten für Kernpromotoren trainiert und kann die Stärke von neuen Promotoren vorhersagen. **B,** Schema der *in silico*-Evolution. **C,** Die Stärke von Kernpromotoren vor (0 Runden) und nach 3 oder 10 Runden *in silico*-Evolution wurde per Plant STARR-seq bestimmt. Daten aus [5].

der umliegenden DNA-Sequenz wurden jüngst deutliche Fortschritte erzielt [11]. In der Zukunft könnte die Kombination von genomischen Datensätzen mit Daten aus Hochdurchsatzexperimenten dazu beitragen noch genauere Modelle zu erstellen.

Zukunftsansichten

In den letzten Jahren wurden bedeutende Fortschritte bei der Charakterisierung regulatorischer DNA in Pflanzen gemacht. Bisher lag dabei der Fokus jedoch meist auf der Aktivität einzelner Elemente. In der Zukunft müssen vermehrt die Wechselwirkungen zwischen verschiedenen regulatorischen DNA-Abschnitten untersucht werden, um zu verstehen, wie die Aktivität aller beteiligten Elemente bei der Genexpression integriert wird. Die Erkenntnisse aus solchen Studien können dazu beitragen, mittels Optimierung entsprechender regulatorischer DNA ertragsreichere und widerstandsfähigere Nutzpflanzen zu erzeugen. ■

Literatur

- [1] Crisp PA, Bhatnagar-Mathur P, Hundley P et al. (2022) Beyond the gene: epigenetic and cis-regulatory targets offer new breeding potential for the future. *Curr Opin Biotechnol* 73: 88–94
- [2] Swinnen G, Goossens A, Pauwels L (2016) Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. *Trends Plant Sci* 21: 506–515

- [3] Schmitz RJ, Grotewold E, Stam M (2022) Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *Plant Cell* 34: 718–741
- [4] Jores T, Tonnies J, Dorrity MW et al. (2020) Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves. *Plant Cell* 32: 2120–2131
- [5] Jores T, Tonnies J, Wrightsman T et al. (2021) Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nat Plants* 7: 842–855
- [6] Gorjifard S, Jores T, Tonnies J et al. (2024) Arabidopsis and Maize Terminator Strength is Determined by GC Content, Polyadenylation Motifs and Cleavage Probability. *bioRxiv*, DOI: 10.1101/2023.06.16.545379
- [7] Jores T, Tonnies J, Mueh NA et al. (2024) Plant enhancers exhibit both cooperative and additive interactions among their functional elements. *Plant Cell: koae088*
- [8] Farley EK, Olson KM, Zhang W et al. (2015) Suboptimization of developmental enhancers. *Science* 350: 325–328.
- [9] Lim F, Solvason JJ, Ryan GE et al. (2024) Affinity-optimizing enhancer variants disrupt development. *Nature* 626: 151–159
- [10] Jores T, Hamm M, Cuperus JT et al. (2023) Frontiers and techniques in plant gene regulation. *Curr Opin Plant Biol* 75: 102403
- [11] Peleke FF, Zumkeller SM, Gültas M et al. (2024) Deep learning the cis-regulatory code for gene expression in selected model plants. *Nat Commun* 15: 3488

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

Dr. Tobias Jores
 Institut für Synthetische Biologie
 Heinrich-Heine-Universität Düsseldorf
 Universitätsstraße 1
 Gebäude 26.24, Etage/Raum 00.105
 D-40225 Düsseldorf
 tobias.jores@hhu.de

AUTOR



Tobias Jores

2008–2013 Studium der Biochemie an der Universität Tübingen; dort bis 2018 Promotion in Biochemie im Labor von Prof. Dr. D. Rapaport. 2019–2023 PostDoc an der University of Washington in Seattle, WA, USA, in den Laboren von Prof. Dr. S. Fields und Prof. Dr. C. Queitsch. Seit 2024 Emmy-Noether-Gruppenleiter im Institut für Synthetische Biologie der Universität Düsseldorf.