

Genomsequenzierung

Genfunktionen effizient identifizieren mit *iPool-seq*

PHILIPP RINK, ARMIN DJAMEI

LEIBNIZ-INSTITUT FÜR PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG (IPK),
GATERSLEBEN

Classical gene characterisation using deletion mutants generated by homologous recombination is slow and can only be applied gene for gene. In our studies we established the *iPool-seq* method in the pathosystem of *Ustilago maydis* and *Zea mays*. This method uses parallel insertion sequencing to identify mutants with growth deficiencies under specific conditions. Furthermore, due to adaptability, the *iPool-seq* could be capable to identify novel gene functions not only in plant-pathogen-interactions but in a broad range of organisms.

DOI: 10.1007/s12268-020-1446-7
© Die Autoren 2020

■ Mit jedem weiteren sequenzierten Genom nähern wir uns einer postgenomischen Ära, bei der die Genomsequenzen bekannt sind und die Erforschung ihrer funktionellen Bedeutung immer weiter in den Mittelpunkt rückt. Die Herausforderung besteht darin, dass jenseits der üblichen *in silico*-Sequenzvergleiche echte experimentelle Daten zur Annotation der vielzähligen Genome unterschiedlichster Organismen häufig fehlen. Während die Generierung von neuen Genomsequenzen inzwischen weitestgehend ein technischer Prozess ist, ist die funktionelle Charakterisierung bisher nicht studierter Sequenzen in der Regel ein mühseliger, zeitaufwendiger und teurer Vorgang, der ein Heer von Wissenschaftlern in den Forschungslaboren beschäftigt hält. Um diese Herausforderung effizient zu lösen, müssen Methodiken weiterentwickelt werden, damit die funktionelle Rolle im Kontext der Biologie von Genen und regulatorischen Sequenzabschnitten im Genom parallelisiert und kostengünstig erfasst werden kann. Auch die Informationen, dass spezielle Sequenzen bei einem bestimmten Prozess keine biologische Rolle spielen, wäre bei genomweiten funktionellen Charakterisierungen von wissenschaftlichem Wert.

Identifizierung von Insertionskassetten durch parallele Sequenzierung

Um Genfunktionen mithilfe genetischer Methoden zu studieren, werden weitläufig zwei prinzipielle Ansätze verfolgt. Entweder wird das Gen inaktiviert, z. B. durch punktuelle Mutationen oder durch Insertion eines fremden DNA-Abschnitts, oder Gene von Interesse werden durch bekannte, starke regulatorische Elemente transkriptionell hochreguliert, um eine Überproduktion zu erreichen. In beiden Fällen werden über den Phänotyp Rückschlüsse auf die biologische Rolle des entsprechenden Genprodukts gezogen.

Die Mutationen von Genen können entweder gerichtet stattfinden, z. B. mithilfe der CRISPR-Cas9-Technologie, TALEN (*transcription activator-like effector nuclease*), homologen Rekombination, oder aber zufällig, z. B. mithilfe von mutagenen Stoffen, von Transposons (springende genetische Elemente) und anderen sich im Genom des Wirts spontan integrierenden Sequenzen. Der erste Fall eignet sich sehr gut für gezielte Studien, ist jedoch zeit- und kostenintensiv im Vergleich zum letzteren, ungerichteten Ansatz. Dieser wiederum ermöglicht keine direkten Schlussfolgerungen zwischen Genotyp und Phänotyp, da der Genotyp – bzw. die jeweils vorliegende Mutation – erst durch Sequenzierung nachträglich aufgedeckt werden muss.

Um genomweite, funktionelle Screens zur Identifikation von Genfunktionen durchzuführen, eignen sich insbesondere haploide, sich asexuell schnell vermehrende Organismen oder Zellsysteme, z. B. viele Pilze, einzellige Algen, Moose oder auch haploide tierische Zellkulturen. Dies liegt daran, dass eine Mutation im haploiden Genom direkte Auswirkungen auf den Phänotyp haben kann, während in diploiden oder polyploiden Organismen rezessive Mutationen durch funktionelle Sequenzen überdeckt sind und in der Regel nicht zu Phänotypen führen, bis sie homozygot vorliegen.

Funktionell genomische Ansätze wurden in den vergangenen Jahren entwickelt, um tausende von Loci eines Genoms auf deren Funktion zu testen. Bei dieser Methode wird zunächst eine Sammlung von Mutanten generiert, bei denen Insertionen im Genom herbeigeführt wurden. Dabei kann die Sammlung gezielt durch Rekombination oder Transposons, die zufällig in das Genom integrieren, hergestellt werden. Eine weitere Möglichkeit ist die Generierung von zufälligen Insertionen durch *Agrobacterium tumefaciens*-vermittelte Transformation von Pilzen. Dieses System wurde benutzt um Gene zu identifizieren, die an der Pathogenität des pflanzenpathogenen Pilzes *Magnaporthe oryzae* beteiligt sind. Vergleiche der Pathogenität von 21.070 Mutanten konnten zur Identifizierung von 202 neuen Loci beitragen, die an der Infektion von Reis beteiligt sind [1]. In der vorliegenden Studie wurden jedoch die Mutanten erst einzeln genotypisiert, um sie dann getrennt, aber parallelisiert, unter verschiedenen Stress- und Umweltbedingungen auf mögliche Phänotypen zu untersuchen.

Effektiver kann ein solcher Screen dadurch werden, dass die Experimente nicht nur mit den einzelnen Mutanten durchgeführt werden, sondern ein Mix der Mutantensammlung verwendet wird [2]. Voraussetzung dafür ist eine nachträgliche Genotypisierung der Mutanten, z. B. durch die massive parallele Sequenzierung von Insertionskassetten. Dieser Ansatz kann aufgrund seiner quantitativen und qualitativen Informationen sowohl für

positive als auch negative Selektionsscreens genutzt werden. Bei positiven Screens überleben die gesuchten Mutanten die entsprechende Selektion durch die Umwelt (**Abb. 1**). Bei negativen Selektionsexperimenten hingegen sterben die gesuchten Mutantengenotypen unter den Selektionsbedingungen. Diese negativen Screens sind um einiges komplexer, da aus der Zusammensetzung der überlebenden Mutanten Rückschlüsse über die Gestorbenen gezogen werden. Mutantenscreens zur Identifikation von Virulenzfaktoren bei Pathogen-Wirt-Interaktionen sind ein gutes Beispiel für solche negativen Selektionsprozesse, da beim Ausfall eines Virulenzfaktors die Fitness des Pathogens und damit sein Vermehrungserfolg im Wirt abnimmt (**Abb. 1**).

Identifikation von Virulenzfaktoren

Um Wirtsorganismen zu infizieren, benötigen Pathogene Virulenzfaktoren. Diese Virulenzfaktoren sind Moleküle, deren biologische Rolle es ist, die Infektion zu ermöglichen, die jedoch ansonsten für das Überleben des Pathogens außerhalb des Wirts nicht von Relevanz sind. Dabei können Virulenzfaktoren wichtig sein, um z. B. morphologische Voraussetzungen des Pathogens für die Infektion zu schaffen. Andere Virulenzfaktoren greifen direkt in die Interaktion mit der Wirtszelle ein, um ihre Anfälligkeit für das Pathogen zu erhöhen und dadurch die erfolgreiche Vermehrung des Pathogens voranzutreiben. Um Krankheitserreger effektiv bekämpfen zu können ist es wichtig, die Virulenzfaktoren zu kennen, da ihre z. B. chemische Inhibierung den Krankheitsverlauf entscheidend beeinflussen kann. Aufgrund ihrer pathogenspezifischen Natur sind bei der Inhibierung von Virulenzfaktoren auch weniger unerwünschte Nebeneffekte auf die Umwelt und den Wirtsorganismus zu erwarten.

Neben klimatischen Faktoren sind biotische Stresse wesentlichen Faktoren, welche die landwirtschaftliche Produktion und damit die Nahrungssicherheit bei einer wachsenden Weltbevölkerung beeinflussen. Die gezielte Erforschung der Virulenzfaktoren von Pflanzenpathogenen ist daher ein wichtiges Forschungsfeld.

iPool-seq als Methode zur Identifizierung von Genfunktionen

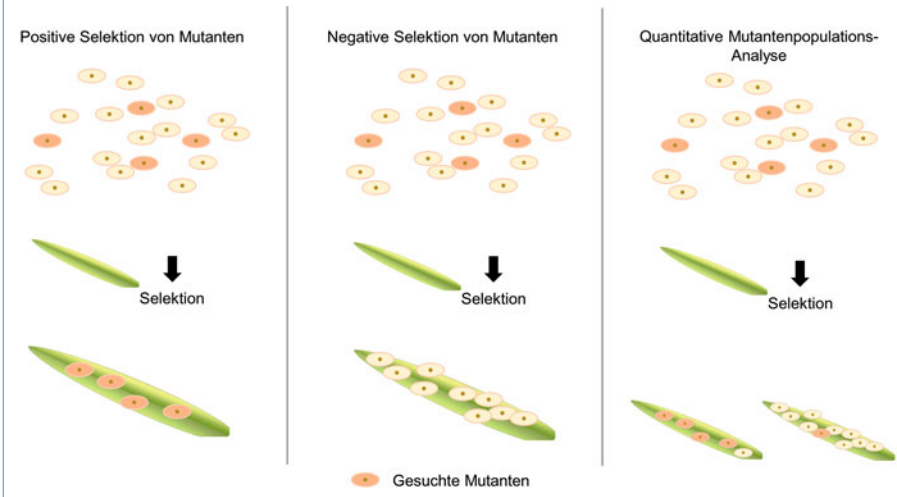
Um Virulenzfaktoren des Maisbrandpilzes *Ustilago maydis* in der Interaktion mit Mais effizient zu identifizieren, haben wir die

iPool-seq-Methode entwickelt [3]. Die besondere Herausforderung bestand darin, eine sensitive Methode zu entwickeln, die eine quantitative und qualitative Aussagen über die Zusammensetzung der Mutantengeneration des Pathogens im Wirtsgewebe der Pflanze ermöglicht. Die Methodenentwicklung wurde mit einer Sammlung von durch homologe Rekombination erstellten Deletionsmutanten von *Ustilago maydis* durchgeführt. *Ustilago maydis* ist ein biotropher Pilz, der die Maispflanze anregt, Gallen zu bilden und sich in der lebenden Maispflanze massiv vermehrt, um dann seine schwarzen Sporen aus den aufplatzen Gallen freizusetzen – was zum Namen Brandpilz führt. Um diese komplexe Manipulation seines Wirts zu bewerkstelligen, hat *Ustilago maydis* wahrscheinlich hunderte Effektorproteine evolviert, die er sekretiert, um seinen Wirt zum eigenen Wohl zu steuern. Als Virulenzfaktorkandidaten für den *ipool-Seq*-Ansatz wählten wir für die Methodenentwicklung daher rund 200 Effektorkandidatengene.

Als erstes wurde die Mutantensammlung aus ca. 200 gemischten Mutantenstämmen verwendet, um Maispflanzen zu infizieren. Die Infektion wurde in drei Replikaten mit jeweils mindestens 100 Maispflanzen durchgeführt. Im nächsten Schritt war es notwendig, die genomische DNA (gDNA) zu extrahieren (**Abb. 2A**). Hierbei ist zu beachten, dass der Großteil der genomischen DNA pflanzlichen Ursprungs ist und dass nur wenige Prozent der genomischen DNA aus den zu testenden Pilzmutanten stammen, welche sich wiederum nur im Bereich der Insertionskassette voneinander unterscheiden. Die gDNA wurde durch die Tagmentierung fragmentiert und ein Adapter angebracht. Dafür wurde eine hyperaktive Version der Transposase Tn5 verwendet. In einer Reaktion kann diese Transposase doppelsträngige DNA fragmentieren und einen kurzen DNA-Adapter, der einen variablen Sequenzabschnitt als Code am Ende enthält, an die gDNA-Fragmente anbringen (**Abb. 2B**). Zusätzlich zu den variablen Sequenzcode-Enden enthält der Adapter eine Primer-Bindestelle. Durch den einmaligen molekularen Marker (variablen Sequenzcode) ist nach der Sequenzierung eine Identifizierung der Häufigkeit von Kopien einzelner Mutanten möglich. Dies ist notwendig, um hervorgerufene Verzerrungen durch die nachfolgenden Polymerase-Kettenreaktionen (PCR) wieder herausrechnen zu können.

Hier steht eine Anzeige.

 Springer



▲ **Abb. 1:** Verschiedene Methoden zur Selektion von Mutanten. Bei einem positiven Screening sterben durch die Selektion (hier durch die infizierte Pflanze) die nicht gesuchten Mutanten und die gesuchten vermehren sich weiter. Bei einem negativen Screen sterben durch die Selektion die gesuchten Mutanten, während sich die nicht gesuchten weiter vermehren. Relative Veränderungen der Mutantenpopulation deuten auf schwächere Phänotypen hin, die einen Einfluss auf die Fitness und damit den Vermehrungserfolg einzelner Mutanten innerhalb der Population haben. Um diese Phänotypen sichtbar zu machen, muss die Anzahl der einzelnen Mutanten im Kontext der Populationszusammensetzung erfasst werden.

Nachdem die Adapter an die fragmentierte gDNA angebracht wurden, werden durch eine PCR nur zur Integrationskassette flankierende Bereiche exponentiell amplifiziert. Um dies zu erreichen, wird ein Primerpaar verwendet, welches an die Integrationskassette und an die Primer-Bindestelle des Adapters bindet (**Abb. 2C**). Dieser Schritt ermöglicht eine spezifische Amplifikation des gewünschten Bereichs. Nach dieser Amplifikation liegt ein Mix an Fragmenten vor, der aus den gDNA-Fragmenten und aus den amplifizierten Bereichen besteht. Um die

gewünschten DNA-Moleküle weiter anzureichern, wurden im vorherigen Schritt biotinylierte Integrationskassetten-Primer verwendet. Dadurch ist es nach der PCR möglich, über eine Streptavidin-Aufreinigung die gewünschten Insertionskassetten-Flanken-Übergangsfragmente anzureichern (**Abb. 2D**).

An die nun angereicherten Fragmente werden durch eine weitere PCR die Bindestellen für die Sequenzierung angebracht (**Abb. 2E**). Daraufhin wird die so generierte Sammlung durch Illumina-Sequenzierung sequenziert

(**Abb. 2F**). Durch die Sequenzierung wird entschlüsselt, welche Insertionen in der Sammlung noch vorhanden sind.

Die Ergebnisse der Sequenzierung der Maisinfektion werden anschließend bioinformatisch mit Ergebnissen der Sammlung vor der Infektion verglichen. Die Mutanten, die nun signifikant weniger im Vergleich zu der Ausgangssammlung zu finden sind, haben möglicherweise einen Defekt in ihrer Virulenz. Dementsprechend sind die deletierten Gene Kandidaten für Virulenzfaktoren. Durch mehrfaches unabhängiges Wiederholen dieser Mischinfektionen gefolgt von *iPool-seq*, werden robuste Virulenzdefekte einzelner Mutanten offenbart. Durch die *iPool-seq* konnten wir auf diese Weise unter den ca. 200 getesteten Mutanten 28 Virulenzfaktoren in *Ustilago maydis* identifizieren, von denen 23 bisher noch nicht bekannt waren.

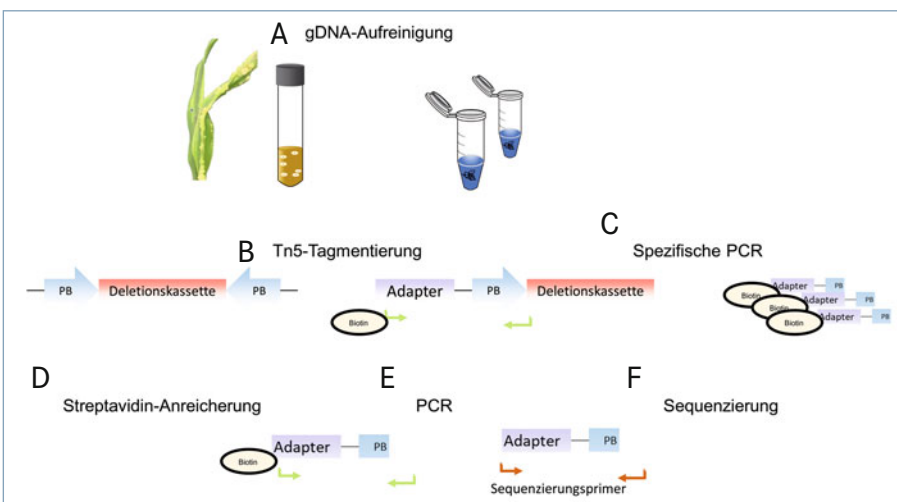
Limitationen des *iPool-seq*-Verfahrens

Durch das experimentelle Design dieses funktionell genomischen Verfahrens gibt es einige Limitationen. Wie für alle genetischen Screens sind haploide Organismen deutlich besser geeignet für diese Methodik, da es notwendig ist, durch eine eingeführte Mutation einen Phänotyp zu erhalten, um für oder gegen diesen zu selektieren.


Zu keinen Ergebnissen kommt es außerdem, wenn ein Gen Redundanz aufweist. Hierbei übernehmen mehrere Gene eine spezifische Funktion. Deshalb ist bei nur einer Deletion eines Gens kein Phänotyp zu identifizieren.

Bei unseren Experimenten konnte außerdem beobachtet werden, dass Pflanzen nur von einer limitierten Zahl an Mutanten infiziert werden. Durch diese Beschränkung ist es notwendig, eine ausreichend große Zahl an Pflanzen zu infizieren, um die stochastische Fluktuation des Nicht-Auftretens einzelner Mutanten nach der Infektion nicht falsch zu interpretieren. Zusätzlich konnten wir in unseren Versuchen erkennen, dass bekannte Mutationen mit schwächeren Effekten auf die Pathogenität nicht zuverlässig identifiziert werden konnten.

Es ist durch unsere Methode möglich, Veränderungen der quantitativen Häufigkeit der Mutanten festzustellen (**Abb. 3**). Jedoch ist es nicht möglich, Veränderungen in der Symptomatik der Infektion zu identifizieren, solange sich die Mutante in der Pflanze zu vermehren vermag. Bei mehr als einer Insertion pro Mutante ist die Methodik ebenfalls nicht mehr anwendbar, da die Information,



▲ **Abb. 2:** Arbeitsabläufe der *iPool-seq*-Methode. **A**, Im ersten Schritt wird die genomische DNA (gDNA) von infizierten Maispflanzen und der Ausgangssammlung aufgereinigt. **B**, Danach folgt die Tagmentierung der gDNA mit der Tn5 und den Adaptern. **C**, Daraufhin werden durch eine spezifische PCR die gewünschten Fragmente vervielfältigt und mit Biotin markiert. **D**, Die biotinylierten PCR-Produkte werden über Streptavidin angereichert. **E**, Abschließend wird eine weitere PCR durchgeführt, bei der die Bindestellen für die Sequenzierprimer an die Fragmente gebracht werden. **F**, Die Sequenzierung erfolgt.

	Klassische Gencharakterisierung	iPool-seq
	Charakterisierung einzelner Gene	Parallele Charakterisierung von Mutantensammlungen
	Zeitintensiv durch einzelne Experimente und Evaluation visueller Symptome	Reduzierung der Arbeitszeit durch Parallelisierung und Fokussierung auf Fitness aller Mutanten im Vergleich
	Identifizierung auch von schwachen Phänotypen	Schnelle Identifizierung von starken Phänotypen
	Kostenintensiv durch individuelle Identifikation	Geringere Kosten durch Parallelisierung
	Benötigt mehr Platz zum Kultivieren durch einzelne Versuche	Geringere Platzansprüche, da in einem Experiment mehrere Gene charakterisiert werden

▲ **Abb. 3:** Gegenüberstellung der klassischen Gencharakterisierung und der *iPool-seq*.

welche Insertionsmutationen in Kombination vorlagen, während der Isolation der genomischen DNA verloren geht.

Breites Anwendungsfeld durch *iPool-seq*

Die von uns etablierte Methode ermöglicht vor allem die Identifizierung von Phänotypen mit starken Fitnesskosten und entsprechender verringerter Vermehrungsrate in einer Population von Deletionsmutanten (**Abb. 3**). Die Planung eines auf *iPool-seq*-basierenden Mutantenscreens fällt und steht mit der Stringenz des Selektionskriteriums.

Dagegen ist die Methodik nicht auf Insertionsmutanten begrenzt, die durch homologe Rekombination erzeugt wurden. So könnten auch Transposon- oder T-DNA-Insertionsmutanten-Populationen dazu genutzt werden, ganze Genome zu screenen. Ähnliche Versuche wurde schon in mehreren Prokaryoten und Eukaryoten durchgeführt [2, 4–6]. Weiterhin könnten durch die Abänderung der Mutantensammlung verschiedene Fragen geklärt werden. So wäre es möglich, durch synthetische Letalität auch Gene zu identifizieren, die Redundanz aufweisen. Als Startpunkt würde in diesem Fall eine Mutante in einem spezifischen Gen vorliegen, welches durch funktionelle Redundanz bei Mutation keinen Phänotyp zeigt. In diesem Mutantenhintergrund würde dann eine Mutantensammlung generiert und Deletionsmutanten selektioniert werden, welche unter bestimmten Bedingungen nicht mehr überleben. Auf diese Weise wäre es möglich, gezielt Redundanz aufzuschlüsseln.

Auch ist es mit unserer Methode möglich, auf Phänotypen durch Überexpression zu suchen. Dafür wäre es nötig, starke regulatorische Elemente in die Nähe der Flanken des Transposons einzubauen und dieses dann im Genom des Organismus der Wahl springen

zu lassen [7]. Dazu würde ein Konstrukt generiert, welches durch einen starken, konstitutiven Promoter Gene aktiviert. Durch die zufällige Integration würde es so möglich werden, unbekannte Funktionen bisher nicht charakterisierten Genen zuzuordnen – auch in polyploiden Organismen.

Genfunktionsanalysen durch *iPool-seq* – ein Ausblick

Die hier von uns aufgezeigte Methodik, ähnliche Ansätze und deren Weiterentwicklung ermöglichen eine schnelle und kosteneffiziente Alternative zu klassischen experimentellen Ansätzen zur Gencharakterisierung – nicht nur im Maispathogen *Ustilago maydis*, sondern bei entsprechender Adaption weit darüber hinaus. Durch eine systematische Vorgehensweise bei der Auswahl der Selektionskriterien sollte es möglich sein, die wechselnden Umweltbedingungen, an die sich ein Studienobjekt evolutiv adaptiert hat, in Grundzügen abzubilden und dadurch eine Vielzahl von neuen Gen-Funktion-Korrelationen aufzudecken.

Dies sollte uns ermöglichen, die bis heute größte funktionelle Genkategorie in den meisten sequenzierten Genomen – *gene of unknown function* – schnell und kostengünstig zu minimieren und dadurch zur funktionellen Entschlüsselung der Genome beizutragen.

Literatur

- [1] Jeon J, Park SY, Chi MH et al. (2007) Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nat Genet* 39:561–565
- [2] Bronner IF, Otto TD, Zhang M et al. (2016) Quantitative insertion-site sequencing (Qlseq) for high throughput phenotyping of transposon mutants. *Genome Res* 26:980–989
- [3] Uhse S, Pflug FG, Stirnberg A et al. (2018) In vivo insertion pool sequencing identifies virulence factors in a complex fungal-host interaction. *PLoS Biol* 16:e2005129
- [4] Cole BJ, Feltcher ME, Waters RJ et al. (2017) Genome-wide identification of bacterial plant colonization genes. *PLoS Biol* 15:e2002860
- [5] Wang N, Ozer EA, Mandel MJ, Hauser AR (2014) Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung. *mBio* 5:e01163–14
- [6] van Opijnen T, Camilli A (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res* 22:2541–2551
- [7] Harb A, Pereira A (2013) Activation tagging using the maize *En-1* transposon system for the identification of abiotic stress resistance genes in *Arabidopsis*. *Methods Mol Biol* 1057:193–204

Funding Open Access funding provided 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. Armin Djamei
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)
OT Gatersleben
Corrensstraße 3
D-06466 Seeland
djamei@ipk-gatersleben.de

AUTOREN



Philipp Rink

Jahrgang 1993. Bis 2018 Master Studium an der Universität Düsseldorf. 2019–2020 Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben in der Gruppe von Dr. A. Djamei. Ab 2020 Angestellter bei der AG Personaldienstleistungen AG / Qiagen N.V.



Armin Djamei

Biologiestudium an der Universität Marburg. 2007 Promotion am Vienna Biocenter, Österreich. 2007–2013 Postdoc am Max-Planck-Institut für terrestrische Mikrobiologie, Marburg. 2013–2019 Arbeitsgruppenleiter am Gregor-Mendel-Institut Wien, Österreich. 2019–2020 Arbeitsgruppenleiter am Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Gatersleben. Ab Oktober 2020 Professor der Pflanzenpathologie an der Agrarwissenschaftlichen Fakultät der Universität Bonn.