universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

Efficient screening for virulence factors with mutant pools in the *Ustilago maydis* – *Zea mays* pathosystem

verfasst von / submitted by

Simon Uhse, BSc, MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 794 685 490
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Molekular Biologie
Betreut von / Supervisor:	Dipl.-Biol. Dr. Armin Djamei, Privatdoz.

Table of Contents

Acknowledgement

The work presented was established in an iterative process with the help and input of many people that I would like to thank in the following lines.

To begin with, the initial idea behind iPool-Seq was presented to me by my supervisor and mentor Armin Djamei when I started. Today, the final protocol has not much in common with the first ideas. This indicates, that we went through many corrections and up and downs. The fascination for molecular biology and the passion to develop new tools for effector research that Armin sets as an example in the lab were essential for me. In my honest opinion, without his persistence and dedication I would probably have not succeeded in the estiblishment of this method. Thus, I would like to thank you sincerely, Armin, how you dealt with throwbacks during the process of establishing iPool-Seq and that you neither doubted me nor the ideas I came up with.

Moreover, I would like to thank my partners in crime Alexandra and Klaus, who were essential in creating the deletion library of *U. maydis* effectors together with me. It is needless to say that without you, the establishment of iPool-Seq would have not been possible.

The development of the technique had a critical turning point in 2016, when Kirsten, a postdoc from campus who attended my Monday seminar, gave me some helpful suggestions from literature. One of these papers made use of Tn5 fragmentation which I implemented in iPool-Seq and which improved the performance of the technique considerably. Thanks a lot for this!

Another essential player during the establishment of iPool-Seq was Florian. He was always available for Armin and me and implemented all adaptations to the protocol in the bioinformatic pipeline immediately. Even more, he contributed with his expertise about unique molecular identifiers and pushed our publication to another level. Thank you Florian, for your help and for a very fruitful collaboration.

I also would like to thank the whole Djamei group. We started as a small group with Armin, Alexandra, Franziska, Julia and Tilo, and established the group together at the GMI. Setting up a laboratory was not always easy, and Franziska and Alexandra really contributed a lot to this and improved many procedures. I appreciate your dedication and your help! Later, more people, like Andre, Fernando, Denise, Gesa and Martin joined the group and made this always a very pleasant working place with lots of positive interactions, personally and scientifically.

Finally, I would like to thank my family. First, thanks to my parents. They made it possible that I could study Biology and always support me. I do not take your love for granted and I am aware that I am a very lucky son. Eventually, I would like to thank my love and my wife Bianca, who joined me on the way from Göttingen to Vienna. We did not regret this decision! Thank you for your constant and emotional support during stressful times. Most of all, thank you for our little wonder Marvin that was born in May this year and makes both of us and everyone in the family so happy!

Preamble

Plants are sessile organisms and must handle various environmental stimuli during their life. These stimuli that plants perceive can be divided in abiotic and biotic signals. Abiotic stresses are triggered by fluctuations in e.g. temperature, salinity, light or humidity. In contrast, biotic signals constitute the exposure of plants to other organisms. Organisms that stimulate biotic stress can be competitors for nutrients or invaders of the plant and are comprised as plant pathogens. Due to their sessile nature, plants must be able to react appropriately and timely to all stimuli at all times. Thus, plants need to prioritize their reactions to the most dangerous stimulus if any of the above-mentioned threats occur simultaneously. To this end, plants have evolved sophisticated regulatory systems that are hard-wired in their genomes in each individual cell. There are two important differences in regulation of plant to animal immunity: first, plants do not possess an adaptive, but exclusively an innate immune system and second, plants do not have any mobile immune cells that are recruited to the site of infection but instead each cell can confer immunity (1, 2).

Plant immunity

Plant immunity can be described in a model with two layers, which was postulated by Dangl & Jones in 2006: The first layer of defense is called pathogen associated molecular and pattern-triggered immunity (PTI) and the second layer that relies on the recognition of secreted molecules from the attackers, is called effector-triggered immunity (ETI) (1, 3). PTI is the basal defense layer that can get triggered by physical contact between the plant and pathogens. To this end, plants recognize highly conserved elicitors of immunity that originate from intruders, called pathogen-associated molecular patterns (PAMPs) by membrane-associated plant pattern recognition receptors (PRRs). A bacterial PAMP is the flagellin-derived peptide flg22 and the corresponding plant receptor is FLS2 (4). In fungal pathogens, chitin is a major constituent of the cell wall and recognized by the PRRs CERK1 and LYK5 in *Arabidopsis thaliana* (5, 6). Moreover, danger signals that can derive from the plant membrane and are released after damage induced by pathogens, called damage-associated molecular patterns (DAMPs), are also perceived by PRRs. Examples of DAMPs are extracellular adenosine triphosphate (eATP) or peptides like AtPep1, which is recognized by the receptor PEPR1 (7, 8). PRRs are either classified as receptor kinases (RKs) or receptor like proteins (RLPs). Both contain an extracellular moiety that can bind a ligand, but an intracellular moiety that transduces the perception signal can only be found in RKs (8). Plants have different classes of RKs and RLPs. The biggest class of RKs contain leucine-rich repeats (LRR), including the above mentioned FLS2 and PEPR1. Interestingly, LRR-RKs are unique to plants and oomycetes and are much more diverse than the related animal cytoplasmic kinases involved in animal defense (9). After PAMP recognition, RKs can transduce the signal of ligand perception and elicit immune responses to counteract the pathogen attack. To do so, the kinase domains of RKs get phosphorylated and initiate signaling

cascades that eventually induce transcriptional changes resulting in immune responses. In contrast to RKs that include a kinase domain, RLPs require a co-receptor with a kinase domain to transmit the signal after ligand binding. Early PTI responses at the site of infection include alkalization of the adjacent apoplastic space, induction of ion fluxes across the membrane, the production of reactive oxygen species (ROS) and deposition of callose to fortify the cell wall (10, 11). To suppress these immune responses, biotrophic plant pathogens that rely on living host tissue secrete a plethora of small molecules, so called effectors, that can interfere and alter defense reactions on multiple levels. Plants on the other hand, have evolved the second layer of defense, ETI, that initiates a strong defense response upon detection of effectors that culminates in programmed cell death (PCD) (1). The detection of effectors is achieved by intracellular receptors, called nucleotide-binding leucine-rich repeat protein receptors (NLRs) that consist of a nucleotide-binding domain and a LRR domain (12). NLRs may either detect pathogenic effectors by direct binding, or observe the status of an effector target in the plant, a so called guardee (12). For instance, the plasma membrane-bound protein RIN4 is guarded by RPM1, which activates an ETI response upon recognition of an interaction between the effectors AvrRpm1 or AvrB and RIN4 (13). Genome sequencing has revealed that plant pathogens encode a large arsenal of predicted secreted molecules, which have likely evolved in an arms race between plants and pathogens (1). In the case of the smut fungus *Ustilago maydis*, 467 predicted secreted molecules are encoded on the genome (14). In contrast, plants harbor much less predicted NLRs, approximately 150 can be found in *Arabidopsis thaliana* (12). This suggests, that the majority of NLRs are guards of intrinsic plant proteins that could be targeted by several pathogenic effector molecules.

Another interesting aspect about plant immunity is that plants can “remember” attacks from biotrophic pathogens in the past and react faster and stronger to a secondary

attack. This defensive plant memory is mediated by the plant hormone salicylic acid (SA) and its regulatory elements, like NPR1, NPR3 and NPR4 (15). Even more remarkable is the fact, that SA confers immunity to a secondary attack not only at the site of the initial infection but also is involved in the regulation of a systemic immunity over the whole plant, called systemic acquired resistance (SAR) (16). However, the mobile signal of SAR itself remains unknown to date.

In summary, plants are sessile organisms that are constantly exposed to stress stimuli. For stress caused by biotrophic plant pathogens, plants possess an immune system that contains two layers: one that recognizes conserved patterns and one that triggers PCD upon detection of effector molecules.

Plant pathogens and effectors

Plant pathogens are manifold and can be found among insects, bacteria, viruses, and filamentous pathogens, like fungi and oomycetes. Especially fungal plant pathogens, like powdery mildew, rust fungi or smut fungi, cause substantial losses in agriculture every year (17). Due to this fundamental threat, the analysis of their virulence mechanism is of high importance in order to improve crop plant resistances and develop sustainable solutions.

Plant pathogens can adopt two different life-styles: they can either be necrotrophic or biotrophic. A plant pathogen that undergoes a switch of both life-styles is called hemi-biotrophic. Necrotrophic plant pathogens harbor a repository of cell wall degrading enzymes (CWDE) that destruct the host tissue during early infection stages (18). CWDEs facilitate colonization by many species and thus, necrotrophs possess a broad host range. In contrast to necrotrophic pathogens that feed on dead plant matter, biotrophic pathogens are reliant on living host tissue to fulfill their life cycle. Thus, biotrophs need to overcome plant immunity and reprogram the plant metabolism for their needs. During evolution, biotrophic pathogens have become specialized to their hosts resulting in a narrow host range (18). Whereas facultative biotrophs, like smut fungi, survive in the absence of the host, obligate biotrophs, like powdery mildew or rust fungi, are dependent on their host and cannot be cultivated *in vitro* in the laboratory. Due to the different life-styles of plant pathogens the plant immune system must distinguish between necrotrophic and biotrophic pathogens and react accordingly. For instance, PCD, the ultimate immune response during ETI, is very beneficial for the host plant against biotrophic pathogens but can be devastating for immunity against necrotrophic pathogens (19).

Most microbial plant pathogens are growing intercellularly in the apoplastic space of the plant, where they retrieve nutrients and multiply (2). To enter the apoplastic space,

microbes need to traverse or circumvent the plant epidermal cell layer, e.g. through natural openings like wounds or stomata. Alternatively, fungal and oomycete pathogens have evolved penetration mechanisms to enter the host, so called appressoria. Nutrient retrieval in many fungal and oomycete pathogens is achieved by specialized feeding organs, called haustoria (20). However, some fungal plant pathogens, like smut fungi, have not evolved haustoria and only contain appressoria. Nonetheless, the development of virulence is very similar: Appressoria and haustoria invaginate plant cells and form an apoplastic interaction zone between fungi and plants in which fungal effector proteins are secreted. Effectors either already reach their destination in the apoplasm or translocate in the plant cell across the plant plasma membrane, where they can target further sub-compartments. Most characterized effectors of eukaryotic plant pathogens are secreted by the conventional secretion machinery, which can be predicted by the presence of an N-terminal signal peptide (SP). Only in rare cases unconventional secretion of effectors without SPs have been reported (21). In contrast, the mechanism of translocation of effectors from the apoplasm to the plant cytoplasm is much less understood: in the case of oomycete pathogens, the short amino acid motif RXLR is probably involved in translocation (22). However, recent insights indicate, that the motif could be cleaved off prior to secretion (23). For fungal effectors, an effector translocation motif is not experimentally confirmed to date. Thus, translocation mechanisms remain enigmatic and await more investigation in future. On the contrary, the translocation of bacterial effectors to the plant cytoplasm is much better characterized. Bacterial plant pathogens use the well-studied type III secretion system (T3SS) to deliver effectors directly from the bacterial cytoplasm to the host plant cytoplasm (24).

In the last decade, genome analyses of filamentous plant pathogens have revealed a large number of putative effectors (25-28). However, the function of the majority of

these effectors remains unclear. The underlying function of an effector is hard to predict, because many effector sequences do not encode any annotated protein domains. Nevertheless, several plant pathogenic effectors lacking known domains have been investigated individually and were functionally characterized in the past. One important criterion is the contribution of an effector to the virulence of the pathogen. Essential effectors contribute indispensably to virulence and hence their mutation causes reduced symptoms during the infection, e.g. in the cases of Avr3a of the oomycete *Phytophthora infestans* or Pep1 of the smut fungus *U. maydis* (29, 30). However, it is likely that most effectors are not essential and will not show an obvious disease phenotype. Due to the long-lasting arms-race with host plants, pathogens have expanded their effector repertoire and many effectors have evolved paralogs on the genomic level, which likely confer virulence by congruent mechanisms (26). Moreover, functional redundancy of effectors, i.e. effectors converge functionally on the inhibition of the same resistance mechanism although their sequence is not related, may result in only subtle or unaltered phenotypes of a single effector knock-out in comparison to the progenitor strain.

Details about the mechanistic functions of plant pathogen effectors have been revealed in several example studies conducted with model organisms, but effector research is still at the beginning and most plant pathogenic effectors await characterization in future. In the following paragraph, some striking examples of effector functions are summarized. Firstly, effectors can help to avoid recognition of pathogens by the host defense system: the fungal effectors Avr4 and Ecp6 of *Cladosporum fulvum* or Slp1 of *Magnaporthe oryzae* sequester chitin to avoid PTI that is triggered through the perception of this cell wall component (31, 32). Interestingly, Avr4 is broadly conserved in many fungal species, including *Pseudocercospora fuligena*, suggesting that this effector is a core component for fungal virulence. The host plant of *C. fulvum* and

P. fuligena is tomato and encodes the receptor Cf-4 that recognizes Avr4 and elicits ETI upon perception. Recent structural analyses of Pf-Avr4 revealed that mutants without the ability to bind chitin can still be recognized by the receptor Cf-4, indicating that ligand binding is not crucial for recognition. Secondly, effectors block components of the defense machinery: a prominent example is the effector AvrPtoB of the bacterium *Pseudomonas syringae.* AvrPtoB adopted the function of an E3 ligase and plays a dual role for both PTI and ETI. On the one hand, AvrPtoB promotes the degradation of the plant PRR FLS2 (33). On the other hand, a recent study has shown that AvrPtoB also targets the SA master-regulator NPR1 for degradation (34). Thridly, several effectors interact with components of the host degradation machinery, indicating that it constitutes an important target for plant pathogen effectors (35). Fourthly, another function of plant pathogen effectors is an antagonistic effect to host proteases. For instance, effectors of the oomycete pathogen *P. infestans* either inactivate plant proteases by direct binding (36), or inhibit their secretion to the apoplasm (37). This is only an exert of known effector functions from single case studies. However, most predicted effectors, especially in filamentous plant pathogens, have not been characterized so far. Understanding effector functions is of outstanding importance, because their characterization will eventually provide insights about host plant targets and potential resistance genes (38). Based on this knowledge, plant geneticists could develop strategies to engineer resistant crop plant in future.

In summary, genome sequencing advances identified that plant-pathogens have evolved an immense effector repertoire. So far, only few mechanistic studies have shed light on the diverse functions of some effectors, but most effector strategies remain unknown.

Ustilago maydis – a filamentous plant pathogen model with a growing toolbox

U. maydis is a member of the smut fungi and known as the causative agent of corn smut disease. In contrast to most other studied pathogenic fungi that belong to the division of the ascomycota, the smut fungi belong to the basidiomycota. Besides the human pathogenic yeast pathogen *Cyrptococcus neoformans (39)*, *U. maydis* is the most prominent and best characterized pathogen from basidiomycetes (26). *U. maydis* is of special interest, because it infects the important crop plant maize. In addition, fungal pathogens have been estimated as one of the highest threats for agriculture and biodiversity (17). Thus, the analysis of virulence mechanisms of plant pathogens is key to the improvement of biocontrol of fungal pathogens. To this end, *U. maydis* has become an important model organism for research of fungal virulence mechanisms and is listed among the top 10 fungal pathogens as research objects (40).

The elucidation of the entire *U. maydis* genome enabled the prediction of putative effectors that follow conventional secretion through the endoplasmic reticulum (ER) (26). The study revealed that many of these putative effectors are clustered on the genome. Moreover, similar genome organizations were found in related smut fungi, like *Sporisorium reillianum* or more recently *Ustilago bromivora* (41, 42). The comparison of genomes from smut fungi also lead to the hypothesis, that effector genes in gene clusters evolve rapidly due to tandem gene duplications and enhanced transposable element activity (43). More recently, a study compared predicted secreted effectors of 12 related basidiomycetes including human and plant pathogens and discovered that conservation between those is rather low (14). In contrast, the 5 plant pathogenic smut species displayed a strong level of conservation and harbor a core effectome of approximately 100 genes (14). Until now, genomic analyses have not shed much light on the function of individual effectors and their contribution to

virulence, but rather serve as a pivotal dataset, i.e. to generate mutants of individual effectors. However, to decipher the functional diversity of effectors in smut fungi in detail, extensive genetic and biochemical studies are required in future.

The infection of maize with *U. maydis* causes symptoms on all aerial parts of the plant and mostly in close proximity to the site of infection. The most obvious symptom is the formation of globular galls. In contrast to obligate biotrophic pathogens, like rust fungi or powdery mildew, *U. maydis* is a facultative biotroph, enabling survival without its host and saprophytic cultivation *in vitro* (Fig. 1).

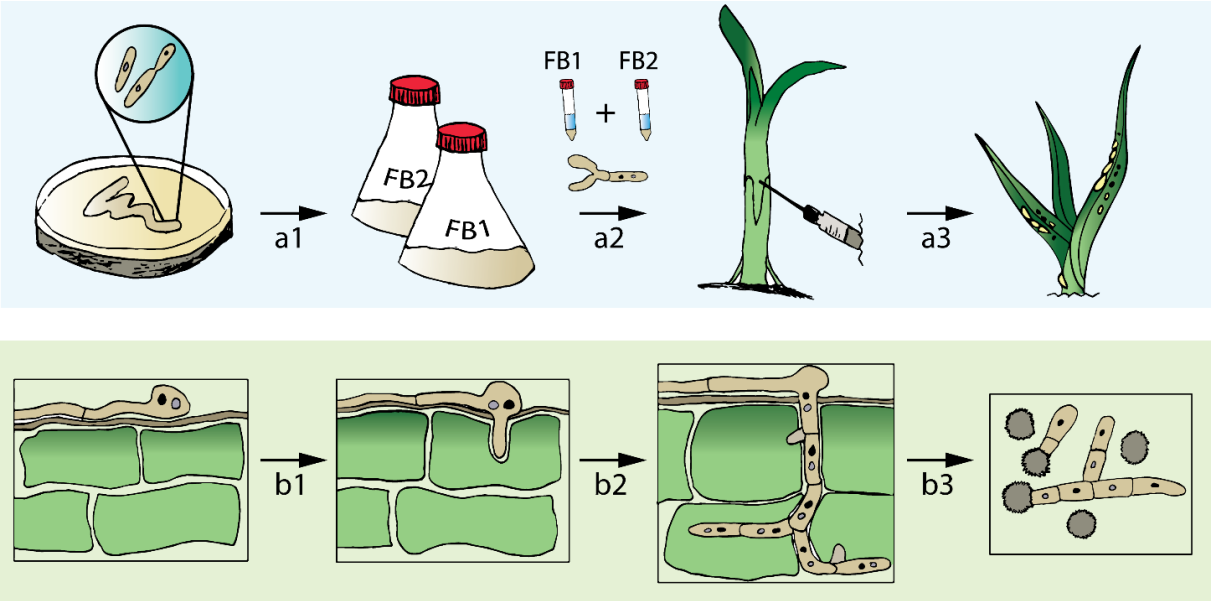


Figure 1. *U. maydis in vitro cultivation and the different colonization phases during maize infection.* **a1** *U. maydis* cultivation and saprophytic growth in vitro. Two compatible haploid mating partner strains FB1 and FB2 are grown separately, both on plate and in liquid culture. **a2** For the infection of the host plant *Z. mays*, both strains are pelleted at an OD_{600} of 0.6 – 1, resuspended in water and mixed in equal amounts to initiate mating and virulence under nutrient deprivation. In laboratory conditions, the strain mixture is injected in the center of a 7-day-old maize seedling. **a3** Symptoms become visible on all aerial parts of the infected plant after 3 - 5 days. The most characteristic symptom is the formation of smutted galls. **b1** After the injection and mating of the FB1-FB2-strain mixture on the host surface, diploid fungal cells contain nuclei from both mating partners and grow on the plant cuticle. **b2** The infection begins with penetration of the dikaryotic cell in epidermal maize cells. **b3** After successful penetration, *U. maydis* forms intercellular hyphae that start to spread throughout the host. Eventually, 4 days post infection gall formation is initiated. Inside galls, diploid spores develop. Once spores germinate, they give rise to haploid cells of the two different mating types, which can grow vegetative again. The figure is adapted from Kamper et al. (26).

Nonetheless, to fulfill its life cycle, *U. maydis* requires its host plant maize. Accordingly, two naturally occurring mating partners, the strains FB1 and FB2, can be cultivated separately on agar plates and in liquid media (Fig. 1a). Once the two mating partners get in close vicinity, they can sense each other with the help of a pheromone receptor system (44). Under nutrient deprivation conditions, the strains mate and filamentous hyphal growth is initiated, which is the major prerequisite for infection. In the first 24 hours, dikaryotic fungal hyphae react to plant surface cues and generate appressoria to penetrate the maize epidermal layer (Fig. 1b) (45). Next, the hyphal growth continues intercellularly between maize cells (Fig. 1b). Time-course RNA-sequencing analyses revealed that *U. maydis* upregulates the majority of effectors genes shortly after infection of the host (46). The upregulation and secretion of effector proteins during biotrophy is the main mechanism of *U. maydis* to establish virulence and overcome plant defense. In contrast to necrotrophic fungi, *U. maydis* does not require CWDE for its virulence (47). After the penetration, the fungus colonizes the sub-epidermal layers and initiates the production of diploid spores in aggregation cavities (47) (Fig 1b). These spores can undergo meiosis and give rise to haploid cells *de novo* that can grow saprophytically. Importantly, the *U. maydis – Zea mays* pathosystem allows reconstruction of the whole infection process under laboratory conditions. By mixing the strains in nutrient-free solution and injecting the solution in the center of a 7-day-old maize seedling the infection can be initiated and subsequently, the disease severity of the infection can be scored (Fig. 1a). To this end, infected plants are rated in different categories by the most severe symptom visible 7 days post infection (dpi) (26).

Genetic and virulence mechanisms of *U. maydis* have been under investigation for about three decades of scientific research. During that time, a variety of tools and resources were developed that allow for genetic and analytic research of disease

mechanisms. The possibility to grow *U. maydis* saprophytically *in vitro* and as a haploid allowed for the development of genetic tools and primarily paved the way for genome editing tools (Tab. 1). Transformation protocols were established early on and insertional mutagenesis via homologous recombination (HR) was used in several studies. Deletion mutants have been crucial to understand the virulence contribution of single genes and gene clusters. More recently, the CrispR/Cas9 system was implemented successfully in *U. maydis*, which might facilitate the generation of multiple gene mutations for the study of homology groups or functional groups of effectors (Tab. 1) (48).

Table 1. Genetic tools established for *U. maydis* research

Genetic Tool	**Application**	**References**
Haploid strain	Solo-pathogenicity and genome editing	(26, 49)
Transformation via homologous recombination	Mutant generation	(26, 29, 50, 51)
Flippase recombination	Selection marker recovery	(52)
Transposon mutagenesis	Generation of random mutants	(53)
Agrobacterium-mediated mutagenesis	Generation of random mutants	(54)
CRISPR-Cas9 system	Targeted multiplex genome editing	(48, 55)
Overexpression promoters	Strong and constitutive expression of heterologous genes	(56, 57)
Immuno-electron microscopy	Investigation of effector translocation	(50, 58)
BirA – translocation tagging	Investigation of effector translocation	(59)
Inducible promoters	Controlled expression of genes	(60, 61)
Minimal promoter	Generation of artificial promoters	(51)

To overcome diploidy and allow working with a haploid strains, a solopathogenic *U. maydis* strain was engineered (49) (Tab. 1). This FB1 strain harbors the *bW2*-gene of FB2 that confers pathogenicity on maize in the absence of the mating partner. Solo-pathogenic strains have facilitated the generation of effector mutants and have accelerated the analysis of mutant disease phenotypes. With the help of disease rating

assays and solo-pathogenic *U. maydis* strains, strong contribution to virulence was demonstrated for several effector genes. For instance, the effectors Pep1, Pit2 and Stp1 are essential and their deletion leads to a complete loss of gall formation (29, 62, 63) (Tab. 1). In addition, entire effector cluster deletions were analyzed, and especially the deletion of the largest cluster 19a displayed a strong impact on virulence of *U. maydis* (26, 64). Nonetheless, considering 467 potentially secreted effectors encoded by *U. maydis*, only few effector phenotypes have been demonstrated until now. Mainly, because the process of disease ratings with single deletion strains is still laborious and requires a large number of plants to get statistically significant results. This is especially true for mutations in effectors that have a minor effect in virulence, e.g. Cmu1, See1 or Tin2 (50, 58, 65). Therefore, new tools are required in future that offer less laborious workflows and that yield quantitative results to understand and decipher the virulence phenotypes of the *U. maydis* effectome more systematically.

Goal of the thesis

U. maydis mutants of potential effectors genes have been used in several studies to test an underlying defect in virulence in the absence of a growth phenotype in axenic culture. For instance, for the known effectors Pep1 and Stp1 (29, 62), knock-out mutants had a very severe effect that led to a dramatic reduction in virulence but showed unaltered growth phenotypes in vitro. However, the method of infection and disease rating becomes problematic for more subtle phenotypes, e.g. in the case of the effector Cmu1 (50). Moreover, disease ratings are based on the observation and analysis of qualitative traits, like gall-formation and chlorosis caused by *U. maydis*. Consequently, reproducibility has proven to be difficult for different group members, probably due to variations in infection success and observed differences in the qualitative analyses of disease severity. In addition, classical disease rating assays neglect the possibility that mutant strains can induce the same symptom severity on maize as the progenitor strain, but that the mutation causes reduced propagation in the host resulting in fewer gall formations. Therefore, classical disease rating assays would not be suitable for screening of a mutant library of a large subset of predicted *U. maydis* effectors, due to its low throughput, high work demand and qualitative and subjective characteristics.

In this study, a novel tool for the toolbox of *U. maydis* research was established, that allows for pooled infection assays with multiple *U. maydis* mutants simultaneously. The read-out of this method, called insertion pool-sequencing (iPool-Seq), is based on genome counts from next generation sequencing (NGS) reads rather than on infection symptoms. As a consequence, iPool-Seq delivers quantitative results that are much more precise than qualitative symptom scores. Moreover, iPool-Seq is suitable for high-throughput screening of hundreds of mutants at once and thus, is a method to save time and resources. Due to is high selectivity and sensitivity, iPool-Seq facilitates

the analysis of complex library samples, i.e. samples that are derived directly from infected host organisms. Importantly, iPool-Seq is not only useful for the plant-microbe interaction community but can be applied for any insertional mutant library.

Recent progress and genomic sequencing efforts have shown that disease mechanisms of biotrophic filamentous pathogens heavily rely on effectors (25-28). Effector biology has the potential to provide not only insights in microbial disease mechanisms, but also in host plant mechanisms, by deciphering plant effector targets and their roles during infection. Therefore, effector biology is an emerging field with an interdisciplinary character that has great potential to trigger breakthrough advances in microbiology, plant molecular biology, agriculture and pest control. In the second communication, we aim to provide a comprehensive overview about the latest research advances in effector biology, with a special focus on effectors of filamentous fungi.

List of publications

1. Uhse S, Pflug FG, Stirnberg A, Ehrlinger K, von Haeseler A, Djamei A. In vivo insertion pool sequencing identifies virulence factors in a complex fungal–host interaction. PLoS biology. 2018 Apr 23;16(4):e2005129.

   Contribution to the study:
   Leading role in resource generation; supervision; rationale and study design; generation of data; data acquisition; figure design; writing and reviewing of manuscript.

2. Uhse S, Djamei A. Effectors of plant-colonizing fungi and beyond. PLoS pathogens. 2018 Jun 7;14(6):e1006992.

   Contribution to the study:
   Figure design and implementation; critical discussion and reviewing of manuscript.

Publication 1

PLOS BIOLOGY

METHODS AND RESOURCES

In vivo insertion pool sequencing identifies virulence factors in a complex fungal–host interaction

Simon Uhse[1], Florian G. Pflug[2], Alexandra Stirnberg[1], Klaus Ehrlinger[1], Arndt von Haeseler[2,3], Armin Djamei[1]*

1 Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Vienna, Austria, **2** Center for Integrative Bioinformatics Vienna (CIBIV), Max F Perutz Laboratories (MFPL), University of Vienna, Medical University Vienna, Vienna, Austria, **3** Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria

* armin.djamei@gmi.oeaw.ac.at

OPEN ACCESS

Citation: Uhse S, Pflug FG, Stirnberg A, Ehrlinger K, von Haeseler A, Djamei A (2018) In vivo insertion pool sequencing identifies virulence factors in a complex fungal–host interaction. PLoS Biol 16(4): e2005129. https://doi.org/10.1371/journal.pbio.2005129

Academic Editor: Joseph Heitman, Duke University Medical Center, United States of America

Received: December 14, 2017

Accepted: March 20, 2018

Published: April 23, 2018

Data Availability Statement: Raw sequencing data has been deposited to ENA under accession number PRJEB23309 and processed data can be found in S1 Data and S2 Data.

Funding: European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) (grant number GA335691 „Effectomics"). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Austrian

Abstract

Large-scale insertional mutagenesis screens can be powerful genome-wide tools if they are streamlined with efficient downstream analysis, which is a serious bottleneck in complex biological systems. A major impediment to the success of next-generation sequencing (NGS)-based screens for virulence factors is that the genetic material of pathogens is often underrepresented within the eukaryotic host, making detection extremely challenging. We therefore established insertion Pool-Sequencing (iPool-Seq) on maize infected with the biotrophic fungus *U. maydis*. iPool-Seq features tagmentation, unique molecular barcodes, and affinity purification of pathogen insertion mutant DNA from in vivo-infected tissues. In a proof of concept using iPool-Seq, we identified 28 virulence factors, including 23 that were previously uncharacterized, from an initial pool of 195 candidate effector mutants. Because of its sensitivity and quantitative nature, iPool-Seq can be applied to any insertional mutagenesis library and is especially suitable for genetically complex setups like pooled infections of eukaryotic hosts.

Author summary

Insertion mutant screens are widely used to identify genotype–phenotype relationships. In negative selection screens, a major limitation is the efficient identification of mutants that are lost or retained after selection. To identify these mutants, the two genomic sequences flanking the insertion cassette must be found. However, pinpointing these insertion flanks within a genome is like looking for a needle in a haystack; a problem that becomes even worse when several organisms form a biotrophic interaction. To overcome this hurdle, we developed insertion Pool-Sequencing (iPool-Seq). With iPool-Seq, we were able to efficiently amplify and enrich insertion flanks from complex genomic DNA samples. This technique allows for the quantification of relative insertion mutant abundance before and after selection by next-generation sequencing (NGS). We demonstrate the power of iPool-Seq with a negative selection screen by infecting maize with 195 candidate effector mutants of

Science Fund (FWF) (grant number P27429-B22, P27818-B22, I 3033-B22). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ARS, autonomous replication sequence; Cm-medium, control Complete medium; dpi, days post infection; EGB, Early Golden Bantam; FDR, false discovery rate; gDNA, genomic DNA; hpt, hygromycin phosphotransferase; HITS, high-throughput insertion tracking by deep sequencing; iPool-Seq, insertion Pool-Sequencing; LB, left border; ME, mosaic end; NGS, next-generation sequencing; PE, paired-end; RB, right border; ROI, region of interest; SBS, sequencing primer binding site; UMI, unique molecular identifier; UPS, unique primer binding site; wt, wild-type.

the fungal pathogen *Ustilago maydis*. We identified 28 virulence factors, of which 23 have not been previously described. iPool-Seq is extremely sensitive, cost- and time-efficient, and promises to be a powerful tool for identifying target genes in large selection screens.

Introduction

Virulence factors are key for successful infections by pathogens. Their identification is of major interest because of the necessity to develop effective counter strategies. For instance, fungal virulence factors are typically identified by mutating single loci in fungi, followed by individual fungal mutant infections of host tissue and subsequent assessment of pathogen fitness [1–4]. Individual infection assays are not ideal for the genetic screening of a large number of pathogen mutants because they are laborious, cost-intensive, and—most importantly—assessment of infections is often subjective and qualitative rather than quantitative. An attractive alternative is infection with a pool of pathogen mutants allowing direct assessment of individual pathogen fitness in the same host tissue. However, using a pooled pathogen infection creates the challenge of identifying pathogens with reduced virulence within a complex mixture of genetic material extracted from infected host tissue.

Mutant collections can be efficiently generated using insertional mutagenesis. Insertional mutagenesis employs gene cassettes that commonly comprise a selectable marker under the control of a strong constitutive promoter. The detection of genome–cassette junctions can serve as a molecular identifier for each insertion mutant. During screening, insertional mutants before selection in the host are defined as the genetic input, whereas surviving insertional mutants after selection comprise the genetic output. Insertional mutagenesis can be achieved randomly through transposon insertion [5–8] or *Agrobacterium tumefaciens*-mediated transformation [3, 9], or specifically through site-specific insertion by homologous recombination [10, 11].

Over the last decade, several approaches were established that use massive parallel sequencing for the detection of inserted gene cassettes. These approaches were successfully used to track mutants from the small genomes of prokaryotic pathogens and allowed the identification of bacterial genes involved in virulence or host colonization after pooled infections [12–16]. However, only a few attempts were reported that identified virulence factors using pools of eukaryotic pathogens [17]. The main factors limiting the successful insertional mutagenesis of eukaryotic pathogens by pooled infections in complex host-pathogen systems are variable infection rates of individually mutated pathogens, the size ratio of host/pathogen genomes, the inability to sufficiently detect inserted gene cassettes from pathogenic material, and biases that arise through PCR-based amplification steps.

To enable successful and quantitative insertion mutant screen-based identification of virulence factors in complex biological systems, we developed insertion Pool-Sequencing (iPool-Seq). We determined the sensitivity and efficiency of iPool-Seq using an insertion mutant collection of 195 predicted virulence factors encoded by the maize pathogen *U. maydis*. The haploid *U. maydis* genome consists of approximately 20.5 megabases [18, 19], whereas the diploid genome of maize is 2.3 gigabases large [20]. This represents a 100-fold genome size difference, which is beside the proportion between fungal and host plant genome abundance as a limiting factor, making the robust detection of *U. maydis* sequence information in infected maize tissue necessary. The iPool-Seq workflow consists of Tn5 Transposase-mediated tagmentation of complex genomic DNA (gDNA) allowing efficient library preparation from low-input material [21, 22]. This is followed by the efficient enrichment of extremely rare insertion cassettes from fungal genomes using biotin-streptavidin affinity purification of PCR products. Amplification

biases are monitored through incorporated unique molecular identifiers (UMIs). Insertional mutant fitness within host tissues is directly measured through quantification of UMI counts present in infected output material compared to UMI counts from the input library.

iPool-Seq on *U. maydis* infections of maize confirmed the identity of 5 known fungal virulence factors that were included as positive controls in the screen. Importantly, 23 previously unreported virulence factors encoded by *U. maydis* were uncovered. Three of these factors were confirmed to be novel virulence factors of *U. maydis* after testing by individual infection. The combination of pooled insertion mutant infections and iPool-Seq technology represents a straightforward and cost-effective approach to map insertion mutants in complex host–pathogen systems with the potential to generate genome-wide virulence maps of relevant crop pathogens and beyond.

Results

iPool-sequencing design and library generation

We employed the smut fungus *U. maydis* as a model to establish iPool-Seq. We generated a Golden Gate cloning-compatible plasmid, which allows for recombination of multiple fragments in a single reaction [23]. To this end, we combined a hygromycin resistance cassette that is flanked by unique primer binding sites (UPSs) with the chromosomal up- and downstream regions (1,000 bp) of 195 predicted *U. maydis* effector genes (Fig 1A; S1 Table). Plasmids were linearized and transformed into *U. maydis* SG200 protoplasts for deletion of the putative virulence factors by homologous recombination (Fig 1B). For each of the insertion mutant constructs, we isolated 3 independent transformants and analyzed deletion events using PCR primers directed against the effector genes sequences. Absence of PCR products

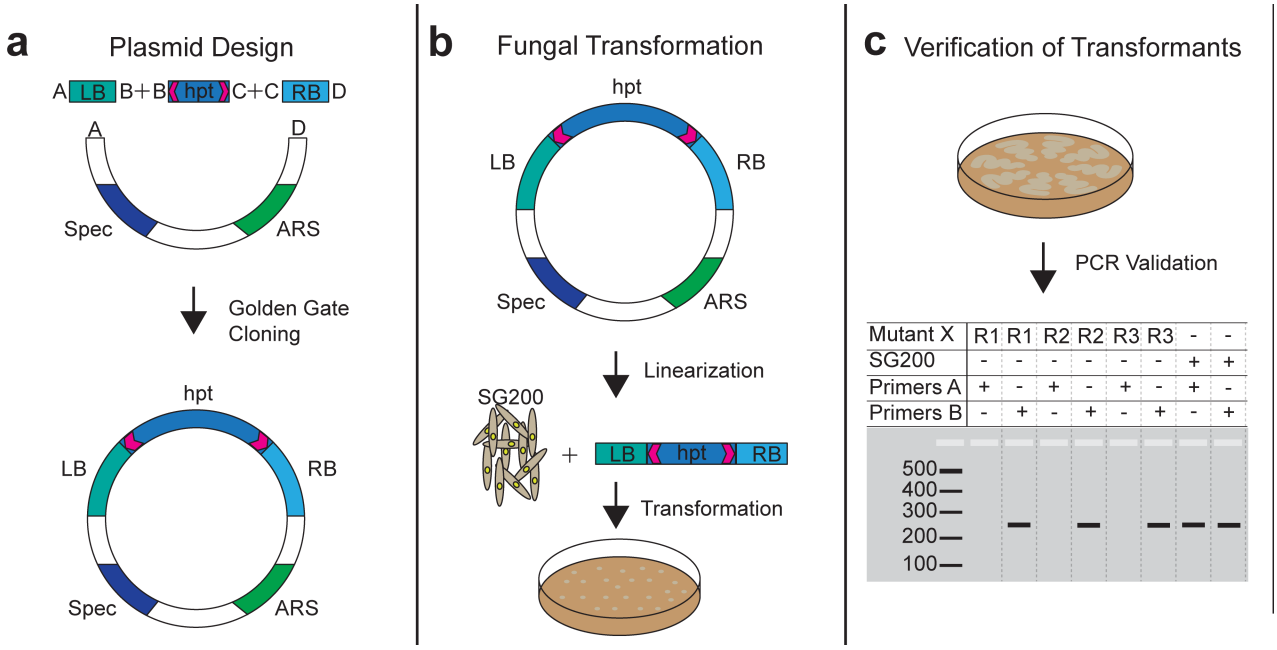


Fig 1. Design of deletion constructs and *U. maydis* insertional mutants. (**a**) Plasmid backbones containing a Spec and an ARS were combined with an hpt resistance cassette and specific borders (LB & RB) via Golden Gate Cloning [23]. The hpt resistance cassette is flanked by UPSs (magenta arrows). (**b**) Plasmids were linearized with *Asc*I and combined with haploid SG200 protoplasts. Transformants were selected on plates supplemented with hygromycin. (**c**) Schematic overview of PCR verification of transformants. Three independent fungal transformants were verified for each mutant locus via PCR. PCR products from primer-pair A targeting insertional mutant X was absent in positive transformants and detectable in SG200 control strains. A control primer-pair B gave a product in both insertional mutant X and SG200. ARS, autonomous replication sequence; hpt, hygromycin phosphotransferase; LB, left border; RB, right border; Spec, Spectinomycin resistance cassette; UPS, unique primer binding site.

https://doi.org/10.1371/journal.pbio.2005129.g001

indicated successful deletions (Fig 1C). For each successful deletion, 3 independent transformant replicates were verified and stored separately, allowing for individual propagation to avoid growth competition prior to pooled infections. We performed 2 independent infections with pools containing the entire collection of 195 insertional mutants and established the iPool-Seq library preparation protocol (S2 Fig).

For later comparison of mutant material abundance within the collection, iPool-Seq libraries were prepared from gDNA representing the mutant pool before infection (the input) and from infected tissues containing both maize and *U. maydis* genomes (the output, Fig 2A). To minimize the number of library preparation steps and conserve material, we replaced mechanical shearing of gDNA (requiring DNA-end repair, tailing, and adapter ligation steps) with Tn5-mediated tagmentation (Fig 2B) [21]. Although this approach yields a wider size range of DNA fragments, simultaneous DNA fragmentation and adapter ligation makes Tn5-mediated tagmentation preferable to DNA shearing approaches. We produced recombinant Tn5 transposase and adapted the published protocol to large gDNA inputs (S3 Fig) [21]. Furthermore, customized adapters for Tn5-mediated tagmentation were designed containing 12 bp unique molecular identifiers (UMIs) followed by a sequencing primer binding site (SBS; Fig 2B; S2 Table), which enables sequencing of UMIs using a custom-made first strand sequencing primer. Fragmented gDNA from pooled fungal infections of maize are not only highly diverse but fungal DNA content will certainly be underrepresented, making it necessary to efficiently enrich for insertion cassette junctions with genomic regions. To enrich for such junctions, the tagmentation-derived DNA fragments were amplified using specific adapter primers and biotinylated primers that bind to

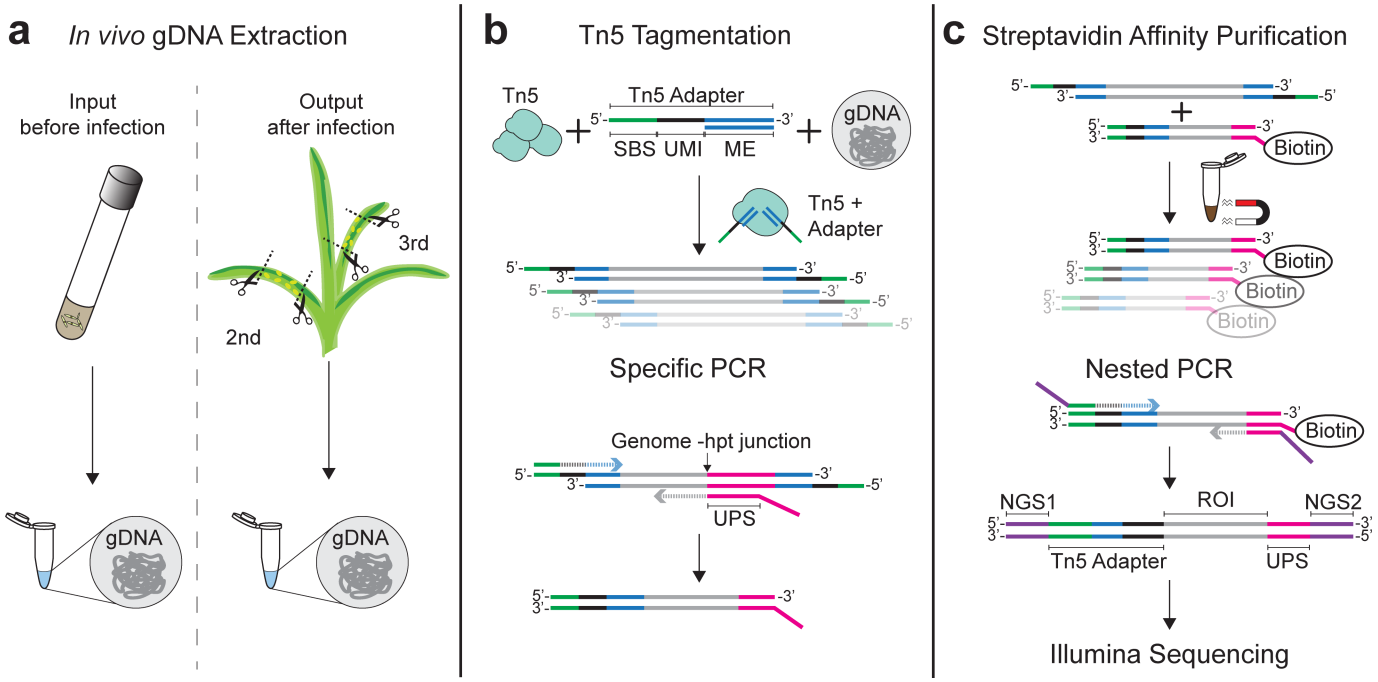


Fig 2. iPool-Seq library preparation workflow features tagmentation and UMIs. (**a**) Library preparation was carried out for the input mutant collection and for the output after infection. For the output, we harvested infected areas of the second and third maize leaves and isolated gDNA. (**b**) Extracted gDNA was fragmented with Tn5 Transposase loaded with custom adapters containing an SBS (green), 12-bp UMI, and Tn5 hyperactive MEs (blue). Genome–hpt resistance cassette junctions were PCR-amplified with biotinylated primers directed against UPSs (magenta) and adapter-specific primers directed at the SBS. (**c**) Biotinylated PCR products were streptavidin-affinity–purified and Illumina-compatible P5 (purple; NGS1) and P7 (purple; NGS2) ends were introduced by nested PCR. Final products were subjected to Illumina PE sequencing on a MiSeq platform. gDNA, genomic DNA; hpt, hygromycin phosphotransferase; iPool-Seq, insertion Pool-Sequencing ME, mosaic end; PE, paired-end; ROI, region of interest; SBS, sequencing primer binding site; UMI, unique molecular identifier; UPS, unique primer binding site.

https://doi.org/10.1371/journal.pbio.2005129.g002

unique sequences at the distal ends of deletion cassettes (Fig 2B; S2 Table). Consequently, both genomic junctions of individual insertion cassettes were amplified, yielding biotinylated PCR products from all insertional mutants. Biotinylated PCR products were isolated using streptavidin-based affinity purification (Fig 2C) and Illumina-compatible adapters were introduced via nested PCR (S2 Table). Sequencing was performed on an Illumina MiSeq platform. In conclusion, we designed iPool-Seq to benefit from tagmentation, specific amplification, and streptavidin purification for efficient enrichment of ultra-rare genome deletion cassette junctions out of a highly diverse gDNA mixture.

iPool-Seq facilitates the identification of fungal virulence factors

We infected maize in two independent experiments with three biological replicates of a pool of 195 verified insertional *U. maydis* mutants (S1 Table), resulting in six input and output libraries. The libraries were prepared as described above and sequenced on an Illumina MiSeq platform with paired-end (PE) sequencing. After read validation and read mapping, 87.7% ± 1.7% and 85.3% ± 1.6% of the obtained sequencing reads (input versus output, respectively) were mapped to *U. maydis* insertional mutation loci (Fig 3A; S1 Supporting methods).

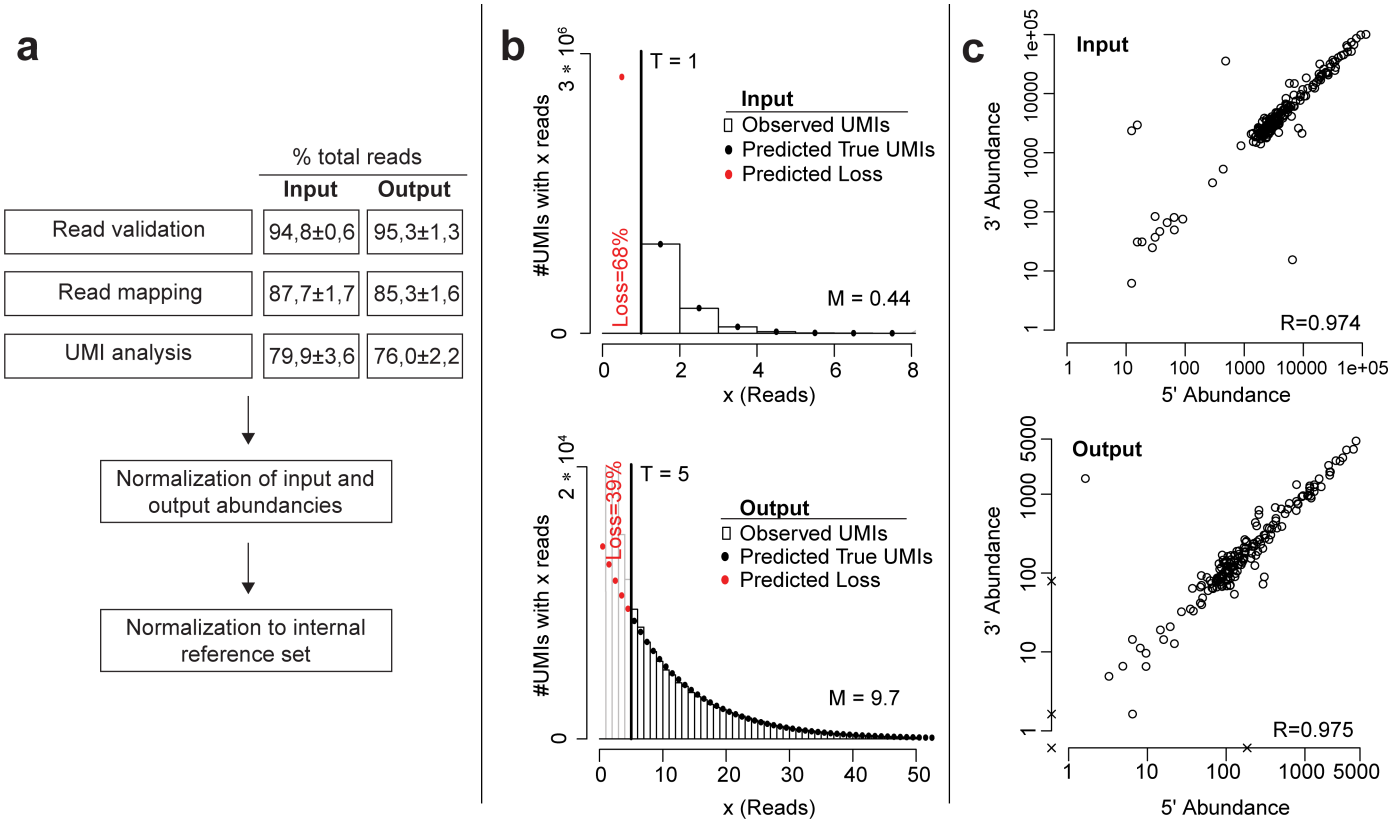


Fig 3. Quality control of iPool-Seq library. (**a**) Bioinformatic workflow of iPool-Seq analysis. Input and output read percentage after validation, mapping, and UMI analysis shows the mean ± SEM of 3 biological replicates and 2 independent infections. (**b**) Distribution of reads per individual UMI (bars) and model prediction (dots) over all insertional mutants of 1 representative replicate for input and output. Here, the error-correction threshold was set to 1 for the input and 5 for the output. Predicted true and lost UMIs are indicated. (**c**) Correlation plot of UMI counts for 5′- and 3′- genomic junctions of the hpt resistance cassette. One representative replicate of input and output is depicted. Each circle represents an insertional mutant. Missing up or downstream reads are marked with x. hpt, hygromycin phosphotransferase; iPool-Seq, insertion Pool-Sequencing; M, mean number of reads per UMI in the predicted distribution; R, correlation value; T, threshold; UMI, unique molecular identifier.

https://doi.org/10.1371/journal.pbio.2005129.g003

To remove reads produced by PCR bias and which would affect quantitative evaluation of input and output reads, we collapsed all reads with highly similar UMIs to a single UMI count after sequencing. Based on the observed distribution of reads per UMI and comparison to a model prediction, we then set a library-specific read count threshold, removed UMIs with fewer reads than the threshold as likely PCR and sequencing artifacts, and corrected the number of remaining UMIs for the estimated loss of real UMIs (Fig 3B, S1 Supporting methods). After this UMI analysis, we retained 79.9% ± 3.6% and 76.0% ± 2.2% of initial reads from input and output for downstream analyses, respectively (Fig 3A).

The sequencing results indicated that three-fourths of all iPool-Seq reads were informational for insertion mutant abundance. Moreover, iPool-Seq generated similar amounts of valid reads from input- and output-derived gDNAs, indicating that yield performance was not diminished using gDNA derived from two organisms.

Since each inserted mutagenesis cassette has two junctions with neighbouring genomic regions, an unbiased library preparation should produce similar read numbers for up- and downstream junctions. We observed high correlation values (R) for all insertion mutants for the input and output samples, indicating that iPool-Seq is not suffering from considerable PCR biases during exponential amplification of DNA fragments containing mutagenesis cassette–genome junctions (Fig 3C).

To identify *U. maydis* virulence factors, we analyzed input and output reads for significantly depleted sequences from the pool of 195 insertion mutants. First, the read output of all insertional mutants was normalized to the corresponding input reads. Second, we defined an internal reference set of *U. maydis* mutant strains that do not have virulence phenotypes [18, 24] and whose output and input reads showed a neutral and linear relationship (Fig 4A, neutral; Fig 4B; S3 Table). Our collection contains additional mutants that were previously reported to be neutral. In these communications, neutral mutants formed symptoms with the same severity as the progenitor strain SG200. However, these observations did not provide any distinct information about quantitative growth defects of these mutants. Therefore, we constrained the neutral reference set to five mutants that displayed a reproducible neutral behavior in the iPool-Seq data (S1 Supporting methods).

We then calculated, for each mutant, the level of depletion from the output sample compared to the input and determined significance through normalization to the internal reference set. This resulted in the identification of a substantial proportion of sequences that were significantly depleted from the output libraries (Fig 4B, red circles; S1 Data). We analyzed this depleted sequence set for known virulence factors and identified Pep1, Pit2, and Stp1 (UMAG_01987, UMAG_01375, and UMAG_02475) [25–27] as known essential virulence factors of *U. maydis* (Fig 4A, lost virulence). In addition, we found the previously described virulence factors ApB73 (UMAG_02011) [28] and Fer1 (UMAG_00105) [29] among the less depleted and reduced candidate sequences (Fig 4A, reduced). Two other mutants (UMAG_06223 and UMAG_02239), for which minor defects in disease symptom induction had been reported previously, were not significantly depleted in the iPool-seq results and one mutant (UMAG_12313) previously reported to be unaffected in virulence showed a weak but significant reduction in our iPool-seq approach (S4 Table) [24]. In summary, iPool-Seq results largely overlap with previously reported symptom scoring data for characterized virulence factors (S4 Table). It is also sensitive, as not only apathogenic but also reduced virulence factor mutants were identified. Importantly, analysis of the depleted sequence set yielded 23 fungal mutants that are potential novel virulence factors of *U. maydis* (Fig 4C; S4 Table).

In contrast to the identification of depleted mutant sequences, we did not identify sequences that were reproducibly enriched in all biological replicates, indicating that none of the fungal mutants tested conferred enhanced virulence to *U. maydis* on the tested host accession Early Golden Bantam (EGB; Fig 4C).

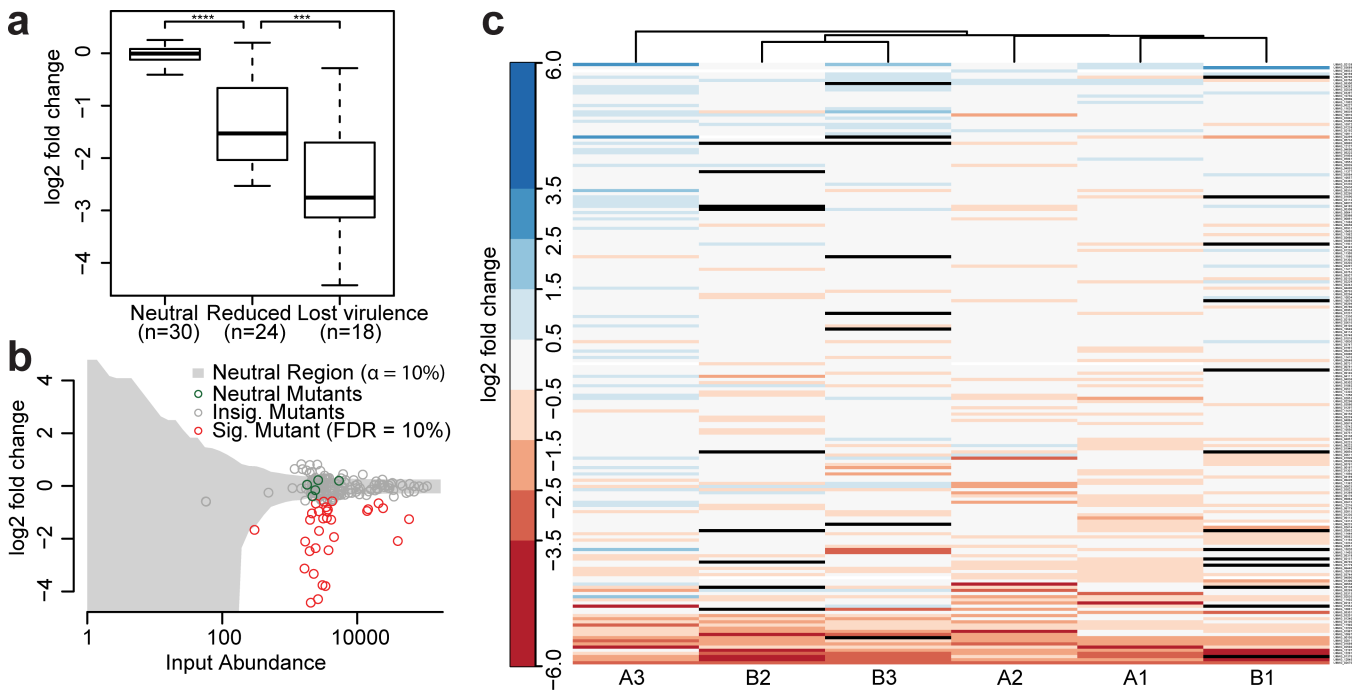


Fig 4. iPool-Seq identifies significantly depleted mutants after pooled infection. (**a**) Log_2-fold changes between normalized output abundances and internal reference set for mutants with known phenotypes. *p*-Values were calculated with Mann–Whitney U tests. $p = 5e^{-9}$ for neutral versus reduced and $p = 3e^{-4}$ for reduced versus lost virulence with $^{***}p < 0.001$; $^{****}p < 0.0001$ (S3 Table). (**b**) Log_2-fold change of output over input abundances for 1 representative replicate. Each circle represents 1 insertion mutant. Internal references are marked in green, significantly depleted in red (tested against reference set using negative binomial test; S1 Data; S1 Supporting methods), unaffected mutants in gray; Insig. area is also highlighted in gray. (**c**) Heatmap of log_2-fold changes of input normalized UMI counts of all insertional mutants sorted by mean level of abundance. Infection A and B are two independent experiments and 1, 2, and 3 are three biological replicates, which were clustered according to similarity. Mutants without detectable reads in output libraries are displayed in black (S1 Data; S1 Supporting methods). FDR, false discovery rate; Insig., insignificant; iPool-Seq, insertion Pool-Sequencing; Sig., significant; UMI, unique molecular identifier.

https://doi.org/10.1371/journal.pbio.2005129.g004

We next modeled the performance of iPool-Seq on a high-throughput mutant library of *U. maydis* (S9 Fig, S1 Supporting methods). To this end, we used the following parameters: 1) 20,000 insertion mutants were chosen cover the approximately 20-MB genome of *U. maydis* with approximately 1,000 bp average distance of insertion sites. 2) During maize colonization, approximately 1,500 of the approximately 6,900 *U. maydis* genes are transcriptionally up-regulated—and we showed that about 14% of all mutants from our library contributed to virulence (Fig 4C; S4 Table) [18, 30]. Based on these observations, we extrapolate that approximately 3% of all *U. maydis* genes are likely to be involved in virulence. 3) We showed with iPool-Seq that known reduced virulence factors of *U. maydis* had a mean logarithmic fold change of −1.53 and known essential virulence factors of −2.75 in comparison to the neutral reference set, respectively (Fig 4A). Due to a lack of data, the model does not take into account the number of unsuccessful infection events on the host plant but assumes 100% infection rate for each individual of a neutral mutant strain.

The model resulted in 40 (for essential virulence factors) and, respectively, 100 (for weak virulence factors) detected individuals necessary for each mutant in the input samples to identify virulence factors with 99% sensitivity. Based on observed average of approximately 10 reads per UMI (Fig 3B) and due to the insertion flank sequencing efficiency of at least 75% (Fig 3A), the required sequencing depth would be 26 Mio reads (20,000·100·10·1,33 = 26,600,000) per library. This suggests that the iPool-Seq technology can be used for large scale mutant screens in *U. maydis* and similar systems.

Validation of novel essential *U. maydis* virulence factors

To validate the 23 potential virulence factors identified by iPool-Seq, we chose three top candidates and tested their effects on virulence using individual infection assays. We observed a severe loss of *U. maydis* virulence upon infection of plants with fungi carrying these mutations. Whereas the wild-type progenitor strain SG200 produced galls on infected maize, all three mutant strains failed to form galls, indicating that they are essential for fungal virulence (Fig 5A). This effect was specifically due to virulence, as growth assays under stress-inducing conditions showed no difference between these mutant strains and SG2000 (Fig 5B). Using confocal microscopy on infected plants, we observed that mutant strains were severely impaired in colonizing maize leaf tissues (Fig 5C). Our combined results show that iPool-Seq facilitates the identification of essential factors for *U. maydis* virulence. Furthermore, the streamlined library preparation of iPool-Seq should make the method widely applicable for identifying unknown virulence factors in complex biological systems, such as in vivo infected tissues.

Discussion

Pooled mutant screens have proven to be very powerful tools to uncover individual genes affecting particular phenotypes in a time- and cost-effective fashion. Positive selection screens usually lead to limited numbers of individual surviving cells that are easily identifiable by a

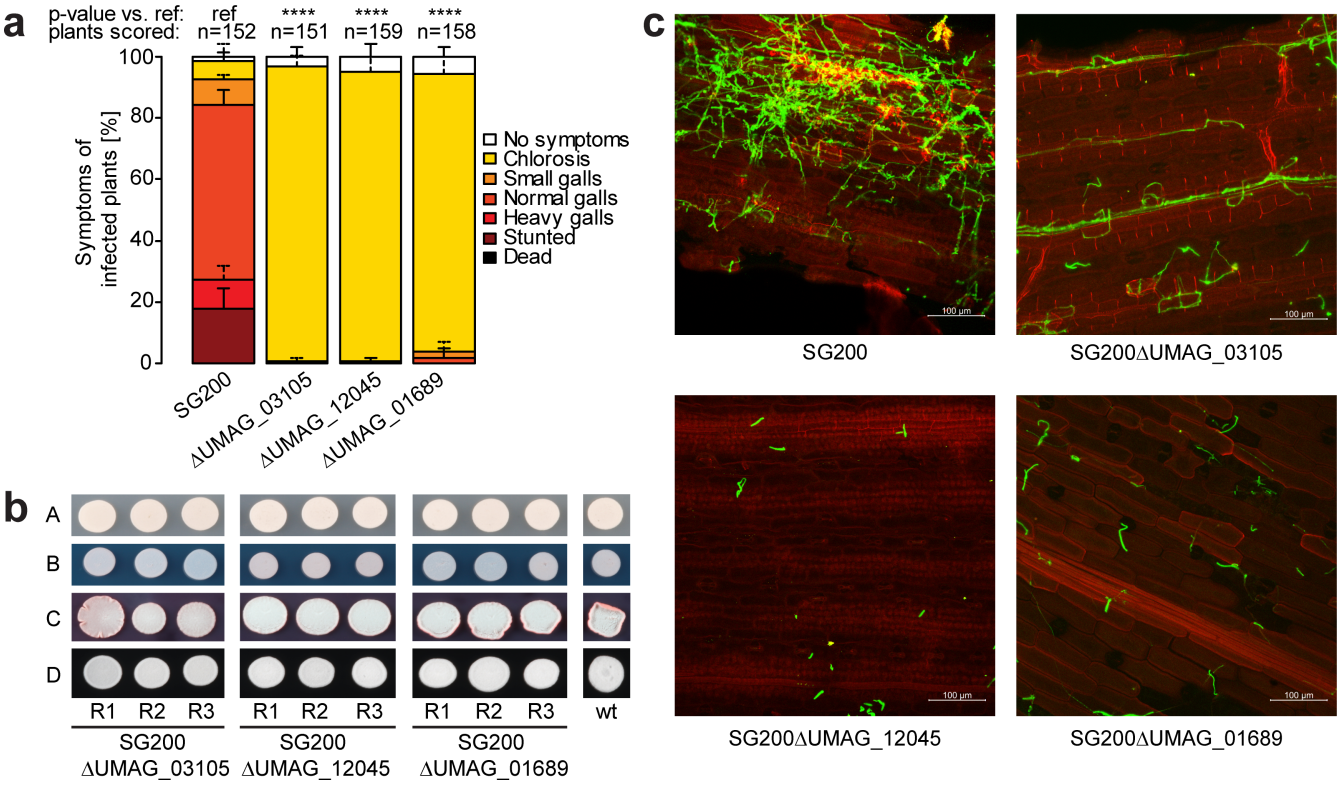


Fig 5. Virulence factor mutants identified by iPool-Seq cause reduced disease symptoms on maize. (**a**) Disease rating of insertional mutant strains 7 dpi. Mean standard deviation of relative counts from 3 replicates are displayed. Only positive error bars are shown. *p*-Values were calculated by Fishers exact test. Multiple testing correction was done by Benjamini-Hochberg algorithm. **** $p < 0.0001$. (S2 Data) (**b**) Growth assay of insertion mutants on (A) Cm-medium, or Cm-medium supplemented with (B) 75 μg/mL Calcofluor (cell wall stress), (C) 45 μg/mL Congo red (cell wall stress), and (D) Charcoal (b-filament inducing). (**c**) Confocal microscopy of maize infected with indicated insertional mutant strains 7 dpi. Infected plant tissue was stained with propidium iodide (red) and fungal hyphae with lectin binding WGA-AF488 (green). One representative picture of 9 infected plants is shown. Cm-medium, control Complete medium; dpi, days post infection; iPool-Seq, insertion Pool-Sequencing; ref, reference; wt, wild-type.

https://doi.org/10.1371/journal.pbio.2005129.g005

combination of restriction enzyme digests, inverse PCR, and sequencing. Negative selection screens rely on the survival of most analyzed cells, making it necessary to devise methodology that allows comparing the presence/absence of genetic information before and after selection. To tackle the later challenge, several insertional mutagenesis approaches have been developed [31]. Although successful in bacterial systems for the elucidation of virulence factors [5, 13, 32], such insertion mutant approaches were not widely used in eukaryotic systems, mainly because of unresolved technical issues such as low sensitivity and system-intrinsic limitations (for example, genome ploidy, lifestyle of the investigated model system).

Here, we introduce iPool-Seq as a versatile and highly sensitive method for the analysis of insertion mutant pools before and after selection, enabling both negative and positive mutant selection screens in complex eukaryotic systems including the analysis of host–pathogen interactions. We used iPool-Seq to examine virulence factors from a defined set of mutants of the crop fungus *U. maydis*, both confirming known factors and identifying novel ones. From the predicted mutant collection we used, most mutants were not significantly depleted from the output reads, indicating no function in virulence for the underlying genes. However, the role of some factors could be difficult to decipher, for example, because their action could be covered by functional redundancy of other virulence factors. Although we infected insertion mutants in dense pools, depleted insertional mutants appeared not to be affected by in trans complementation, by using the secreted factors of neighbouring fungal cells for example. Nevertheless, it cannot be excluded that, for certain gene products, in trans complementation could occur and mask the virulence defect of the respective mutant in a pooled infection setup. In conclusion, negative depletion screens have limitations to decipher redundancy and potential in trans complementation of virulence factors. In addition, we did not identify significantly enriched mutants in the iPool-Seq analysis of the mutant collection. A significant enrichment of output reads would indicate the loss of a negative regulator of virulence. A possible reason that we did not find enrichment could be our choice of the maize accession, EGB, which is highly susceptible to the *U. maydis* strain SG200.

Microscopy of *U. maydis* strain SG200 infecting maize tissue implies that many cells fail to penetrate the host [28]. In very complex insertion mutant libraries, this large individual failure rate could lead to the loss of mutants that lack any real defect in virulence. Therefore, for genome-wide virulence maps of *U. maydis* and similar biotrophic pathogens, the size of the insertion mutant pool must be individually adapted to the infection rate of the respective pathogen. To overcome this problem, genome-wide screens might need to be performed in subpools, as it has been done in a previous study with the fungal pathogen *Cryptococcus neoformans* [11].

iPool-Seq uses insertion cassette–specific primers to amplify the genomic insertion junctions from a mutant pool [17]. Additionally, iPool-Seq enriches for PCR products by using biotin/streptavidin interaction, an approach that has previously been used in bacterial transposon integration site identification methods such as high-throughput insertion tracking by deep sequencing (HITS) [5]. Importantly, UMIs in the adapter primer allow in silico elimination of PCR biases. The unique barcode identifiers additionally overcome cluster position identification problems during Illumina sequencing that would otherwise occur when the first bases from the insertion flank would otherwise be identical for all mutant loci. Dark cycle sequencing, as used in Quantitative insertion-site sequencing (QIseq) for example, is therefore unnecessary [17].

iPool-Seq was established using a defined insertion mutant collection of *U. maydis*. However, the technology can be adapted to any insertion mutant collection, such as transposon or *A. tumefaciens*-derived T-DNA libraries [33, 34]. The modeling of the iPool-Seq sensitivity indicates that iPool-Seq meets all premises to work for high-throughput. Therefore, iPool-Seq

promises to be a versatile technology for reanalysis of existing knock-in, activation-tagging, or transposon-insertion libraries, dramatically reducing labor costs for selection screens when compared to classical scoring approaches. Additionally, the relatively low costs of iPool-Seq for broad screens could also foster research in less funded emerging model systems. Due to the strong enrichment of insertion gene cassettes, the sequencing depth and costs of iPool-Seq are low. Thus, this technology will enable researchers to test diverse new selection criteria to efficiently build genotype–phenotype relationships. This will help to fill the knowledge gap that is currently still hampering research as exemplified for the well annotated *U. maydis* genome with 6,786 protein-encoding genes, of which 41.5% are in the category unknown [35]. Moreover, even if genes are annotated, their involvement in various biological processes might, simply, not yet be known.

From the candidate virulence factors that we identified with iPool-Seq, we chose 3 for verification and confirmed their virulence defect by classical scoring of disease symptoms. However, the assessment of disease symptoms is indirect, and discrepancies between the two methods might occur for other novel virulence factors. We speculate that the *U. maydis* genome encodes virulence factors whose mutants show reduced proliferation but still cause full disease symptoms based on qualitative measures. In line with this, the iPool-Seq data did not show significant depletions for two mutants that were previously reported with mild defects in symptom induction [24]. In contrast to these disease ratings, iPool-Seq has the potential to identify virulence factors that do not have an obvious effect on symptom formation on a genome-wide level.

In summary, we have demonstrated the functional genomic technology iPool-Seq by identifying both known and novel virulence factors from pooled infection assays of a biotrophic fungus within a complex host background. iPool-Seq is therefore a sensitive in vivo tool for researchers to help fill the genotype–phenotype gap in the post-genomic era.

Methods

Vector construction and insertional mutant generation

For all DNA manipulation we used *Escherichia coli* Mach1 (Thermo Fisher Scientific). The vector backbone for the generation of the mutant collection is based on pGBKT7 (Clontech Laboratories). We replaced kanamycin resistance with a spectinomycin resistance cassette and removed internal *Sap*I, *Bsa*I, *BsmB*I, and *Bbs*I restriction sites by direct mutagenesis from a derivative of the original vector, respectively [36]. The hygromycin resistance marker originates from vector pHwtFRT [37]; and *Sap*I, *Bsa*I, *BsmB*I, and *Bbs*I restriction sites were removed by site-directed mutagenesis. Moreover, we elongated the hygromycin cassette with a UPS on the 5′- and 3′-end (5′-TCGCCACAGGATACCACAGGACATCTGGGATATC and 3′-GCCACTCACGCCACAGGATACCACAGGACATCTGGGATATC; UPS is underlined). In detail, for each mutant locus we amplified 1,000 bp up- and downstream borders from *U. maydis* gDNA with standard molecular cloning procedures [38] and combined them with the modified hygromycin-selectable marker cassette flanked with UPS (Fig 2; S2 Table) and the plasmid backbone. Depending on the occurrence of internal restriction sites, we used either *Sap*I, *Bsa*I, *BsmB*I, or *Bbs*I restriction sites (ordered by priority of choice) for Golden Gate cloning [23]. Constructs were verified by Sanger sequencing and subsequently transformed into the haploid solopathogenic strain SG200 of *U. maydis* as previously described [18, 39, 40]. Transformants were verified by direct PCR: single mutants were grown in YepsLight (0.4% yeast extract, 0.4% peptone and 2% sucrose) liquid medium at 28˚C with shaking at 200 rpm in 48-well plates overnight. The next day, 100 μL overnight culture was pelleted and resuspended in 20 μL 0.02 M NaOH. 1 μL was then utilized for a direct PCR reaction with a primer pair directed against the replaced gene. As a positive control, a primer pair binding to another mutant locus was used.

Subsequently, we isolated gDNA from at least 4 PCR positive strains and repeated the direct PCR using 1 μL of 1:10 diluted gDNA as a template. All primer pairs used for the verification of deletion strains produced PCR products from a gDNA template from the progenitor strain SG200. Three independently verified *U. maydis* insertional mutants were preserved at −80˚C in PD liquid supplemented with 50% glycerol.

Growth conditions and pooled infection

For each mutant collection pool replicate we infected at least 100 plants of maize variety EGB (Olds Seeds, Madison, WI, USA). Seedlings were grown under a 14-hour/10-hour light/dark cycle at 28˚C/20˚C in plant growth chambers and infected 7 days after potting. *U. maydis* mutant strains were grown individually on selective PD plates supplemented with 200 μg/mL hygromycin for 2–3 days at 28˚C. Subsequently, for each mutant strain, 1 mL YepsLight (0.4% yeast extract, 0.4% peptone and 2% sucrose) liquid preculture was inoculated in 48-well plates and grown at 28˚C overnight with shaking at 200 rpm. For main cultures, precultures were diluted 1:2,000 in 3 mL YepsLight in test tubes and grown at 28˚C with shaking at 200 rpm overnight. After 14–16 hours, the main cultures of all mutants were adjusted to an OD_{600} of 3 and mixed in equal amounts. The mutant pool was pelleted at 2,000 x g for 10 minutes and resuspended in sterile water. 250 μL of the mutant pool was infected in each maize seedling with a syringe. After 7 days, infected areas from the second and third leaves were harvested, ground to a fine powder in liquid nitrogen, and preserved at −80˚C until iPool-Seq library preparation.

iPool-Seq library preparation

For output gDNA extraction, 0.75–1 g of infected plant powder was supplemented with 2 mLLysis buffer (10 mM Tris, pH 8; 100 mM NaCl; 1 mM EDTA; 2% Triton X 100 [v/v]; 1% SDS [w/v]), 2.5 mL TE-buffer equilibrated phenol, chloroform, and isoamyl alcohol (25:24:1, pH 7.5–8, Carl Roth) and 100 μL sterile glass beads (450–600 μM, B.Braun) in a 7-mL Precellys tube. The material was processed for 20 seconds at 4,500 rpm with a Precellys evolution bead mill (Bertin). The debris was pelleted at 17,000 x g for 15 minutes, and 2 mL supernatant was added to 2.2 mL Isopropanol. The precipitated gDNA was washed with 1 mL 80% EtOH and eluted in 150 μL or 200 μL TE supplemented with RNAse A (20 μg/mL, Thermo Fisher Scientific). For input gDNA extraction, gDNA was extracted from 2 mL of insertional mutant pool as previously described [41]. gDNA concentrations were determined with PicoGreen (Thermo Fisher Scientific). Tn5 fragmentation of a total of 10 μg gDNA for output and 1 μg gDNA for the input was adapted from [20], and performed as follows [21]: We combined 1 μg gDNA per 20 μL reaction with Tn5 transposase (150 ng/μL f.c.) preloaded with 25-μM adapters in 1x TAPS buffer (50 mM TAPS-NaOH, 25 mM MgCl2, 50% v/v DMF, pH 8.5 at 25˚C) and incubated the reaction mix in a thermocycler at 55˚C for 10 minutes. We purified each reaction mix with a 1:1 ratio of Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's protocol and performed PCR with Phusion polymerase (New England Biolabs) using an adapter specific forward primer and a biotinylated insertion specific primer from 250 ng fragmented gDNA (denaturation for 15 seconds at 95˚C, annealing for 15 seconds at 65˚C, elongation for 30 seconds at 72˚C; repeated for 15 cycles; 1 minute final elongation). We pooled all PCRs of the same sample and purified 1/5 with Agencourt AMPure XP beads (ratio 1:1; Beckman Coulter). The PCR amplicons eluted from each sample were split into 4 PCR reactions and amplified with nested primers to add Illumina compatible P5 and P7 ends (15 cycles, with 65˚C annealing temperature and 30 seconds elongation at 72˚C). The final PCR products were purified with Agencourt AMPure XP beads in a 1:1 ratio. The average fragment

size was measured on a fragment analyzer (Advanced Analytical Technologies, Inc.) and library quality was controlled with qPCR. Illumina Sequencing was performed on a MiSeq platform with 75 PE conditions. We used a custom designed forward sequencing primer and the standard Illumina primers for reverse and index sequencing (S2 Table).

Confirmation of iPool-Seq candidate virulence factors

We confirmed the results of iPool-Seq for 3 candidate genes with individual infection assays, microscopy, and in vitro growth assays. The infection assay was performed as previously described [18]. In summary, for each insertional mutant, 3 replicates of *U. maydis* were grown overnight in YepsLight liquid medium (0.4% yeast extract, 0.4% peptone and 2% sucrose) with 200 rpm agitation to an OD_{600} of 0.6–1 and adjusted to an OD_{600} of 1 in sterile water. We syringe-infected 7-day-old maize seedlings of the variety EGB with approximately 250 μL fungal suspension per plant. Symptoms were scored 7 days post infection (dpi) according to the published protocol [18]. Fungal leaf colonization was assessed 7 dpi via microscopy. Fungal hyphae were stained with WGA-AF488 (Thermo Fisher Scientific) and plant cell walls with propidium iodide (Sigma-Aldrich) as previously described [28]. Confocal microscopy was performed with the following settings: We utilized an LSM780 Axio Observer confocal laser scanning microscope with an LD LCI Plan-Apochromat 25x/0.8 Imm Corr DIC M27 objective (Zeiss, Jena, Germany). WGA-AF488 was excited at 488 nm and detected at 500–540 nm; propidium iodide was excited at 561 nm and detected at 580–660 nm.

Bioinformatic analysis

For each sequenced library, adapter read-throughs were removed from the raw Illumina reads, UMIs were extracted and stored separately, and the reads (lacking UMIs) were mapped to the *U. maydis* reference genome [18] using NextGenMap [42]. The reads mapping to each flank (5' and 3') of each insertional mutant were grouped by UMI, and highly similar UMIs were merged to correct for sequencing errors [43]. UMIs with fewer reads than the error-correction threshold were removed as likely artifacts, and the number of surviving (and thus likely true) UMIs for each gene and flank were counted. To correct for biases caused by different detection losses (i.e., # undetected genomes/# total genomes) between mutants and flanks, the mutant- and flank-specific losses were estimated from the observed mutant- and flank-specific distributions of reads per UMI (S1 Supporting methods) using the TRUmiCount algorithm (see S1 Supporting methods for details) [44]. To discern stochastic fluctuations from knockout phenotypes, the number of true UMIs detected in the output pool for neutral insertional mutants were assumed to follow a negative binomial distribution with mean $\mu_m = \lambda \cdot n_m^{\text{in}} \cdot \left(1 - \ell_m^{\text{out}}\right)/\left(1 - \ell_m^{\text{in}}\right)$ and (inverse) overdispersion parameter $r_m = n_m^{\text{in}}/\left(1 + d \cdot n_m^{\text{in}}\right)$. Briefly, a neutral mutant m's expected UMI count in the output pool thus depends on (1) the number n_m^{in} of detected UMIs in the input pool, (2) the estimated losses ℓ_m^{out} and ℓ_m^{in} for the output and input pool, and (3) a mutant-independent normalization factor λ to account for differences in total genome count between input and output samples. The sources of overdispersion of the output counts are (4) the (Poissonian) sampling uncertainty of the input pool counts n_m^{in}, and (5) random fluctuations of fungus proliferation accounted for by the mutant-independent parameter d. For each output pool, parameters λ and d were estimated (see S1 Supporting methods for details) by fitting the model to a reference set of presumed neutral mutants (S3 Table), 2 one-sided p-values for the significance of depletion (respectively, enrichment) compared to the reference set were computed for each insertional mutant and transformed to q-values to control for the false discovery rate (FDR) [45]. Undetected insertional mutants (i.e., with zero UMIs) in input pools were

excluded from the analysis of the corresponding output pools. Undetected insertional mutants in output pools were not assigned *p*- or q-values.

To quantify the change in virulence of an insertional mutant, its abundance in the output was first normalized to its abundance in the input (thus assuming independent fates of the individuals in the input). Then, the $\log_2$-fold change between its normalized output abundance and the normalized output abundance of the internal reference set was computed (see S1 Supporting methods for details). Further details on the modeling can be found in S1 Supporting methods.

Supporting information

S1 Data. q-Values of *U. maydis* mutant strains.
(XLSX)

S2 Data. Symptom rating of mutant strains.
(XLSX)

S1 Fig. Workflow of pooled infection of maize. For each replicate of the *U. maydis* mutant collection, at least 100 maize plants of the accession EGB were potted. Mutants were grown on selective plates for 2–3 days. From plates, precultures were inoculated and grown ON. The precultures were used for inoculation of the main cultures to avoid dead material in the infection pool. All main cultures were pooled with equal amounts that were adjusted to the same optical density and infected in 7-day old maize seedlings with a syringe. Infected areas of the second and third leaf of each plant were harvested 7 days after the infection. All 3 biological replicates of the mutant collection were processed in 14 days. EGB, Early Golden Bantam; ON, overnight.
(TIF)

S2 Fig. Tn5 fragmentation of gDNA with modified adapters. Recombinantly produced hyperactive Tn5 was tested with standard Tn5-ME-A and custom UMI-ME-A on 1 μg gDNA of *U. maydis*-infected maize tissue with indicated concentrations. gDNA; genomic DNA; In, Input; M, Marker 1 kb-ladder (Thermo Scientific); ME, mosaic end; Tn5-ME-A, Tn5-ME-A-dapter; UMI-ME-A, UMI-ME-adapter.
(TIF)

S3 Fig. Sensitivity of iPool-Seq. Estimated sensitivity of iPool-Seq for a genome-wide library of *U. maydis* mutants. Model shows for different (1 up to 100) mutant copies detected in the input sample for the sensitivity of virulence factor detection. Depicted model curves are given assuming 3% of all mutants have a reduced virulence of log2(FC) −1.53 and log2(FC) of −2.75, respectively, and the other 97% are neutral in respect to virulence. The sensitivity reaches 99% at 40 detected mutants (lost virulence) and 100 detected mutants (reduced virulence), respectively. FC, fold change; iPool-Seq, insertion Pool-Sequence.
(TIF)

S1 Software. iPool-Seq analysis pipeline. iPool-Seq, insertion Pool-Sequencing.
(TGZ)

S1 Supporting methods. iPool-Seq analysis pipeline description. iPool-Seq, insertion Pool-Sequencing.
(PDF)

S1 Table. *U. maydis* genes targeted for insertional mutagenesis.
(XLSX)

S2 Table. Key primers used in this study.
(XLSX)

S3 Table. *U. maydis* mutants used for the internal reference set.
(XLSX)

S4 Table. Significantly depleted *U. maydis* mutants identified by iPool-Seq. iPool-Seq, insertion Pool-Sequencing.
(XLSX)

Acknowledgments

We thank the Molecular Biology service at the Vienna BioCenter for their excellent Sanger Sequencing support. Special thanks go to the NGS Core Facility (VBCF) for the adaptation of sequencing protocols. Our acknowledgement goes to the Vienna BioCenter Core Facility (VBCF) plant facilities for their support with plant growth. We thank Dr. Kirsten Senti for technical advice. We would like to thank Dr. Matthias Schäfer for helpful advice on the manuscript. Special thanks go to Dr. Youssef Belkhadir and Dr. Yasin Dagdas for critical reading of the manuscript and constructive suggestions. We would also like to thank all the members of the Djamei lab and of the CIBIV (Center for Integrative Bioinformatics Vienna), particularly Dr. Angelika Czedik-Eysenberg, Denise Seitner, Luis Paulin-Paz and Celine Prakash, for valuable feedback on the project and manuscript, and Dr. J. Matthew Watson for proofreading.

Author Contributions

Conceptualization: Armin Djamei.

Data curation: Florian G. Pflug.

Funding acquisition: Arndt von Haeseler, Armin Djamei.

Investigation: Simon Uhse, Alexandra Stirnberg, Klaus Ehrlinger.

Methodology: Simon Uhse, Florian G. Pflug, Armin Djamei.

Project administration: Arndt von Haeseler, Armin Djamei.

Resources: Simon Uhse, Alexandra Stirnberg, Klaus Ehrlinger, Armin Djamei.

Software: Florian G. Pflug.

Supervision: Arndt von Haeseler, Armin Djamei.

Validation: Simon Uhse, Florian G. Pflug.

Visualization: Simon Uhse, Florian G. Pflug.

Writing – original draft: Simon Uhse.

Writing – review & editing: Armin Djamei.

References

1. Idnurm A, Reedy JL, Nussbaum JC, Heitman J. Cryptococcus neoformans virulence gene discovery through insertional mutagenesis. Eukaryot Cell. 2004; 3(2):420–9. https://doi.org/10.1128/EC.3.2.420-429.2004 PubMed PMID: WOS:000220945300018. PMID: 15075272
2. Wamatu J, White D, Chen W. Insertional mutagenesis of Sclerotinia sclerotiorum through Agrobacterium tumefaciens-mediated transformation. Phytopathology. 2005; 95(6):S108–S. PubMed PMID: WOS:000202991401123.

3. Jeon J, Park SY, Chi MH, Choi J, Park J, Rho HS, et al. Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. Nat Genet. 2007; 39(4):561–5. https://doi.org/10.1038/ng2002 PMID: 17353894.
4. Michielse CB, van Wijk R, Reijnen L, Cornelissen BJC, Rep M. Insight into the molecular requirements for pathogenicity of Fusarium oxysporum f. sp lycopersici through large-scale insertional mutagenesis. Genome Biol. 2009; 10(1). doi: ARTN R4 https://doi.org/10.1186/gb-2009-10-1-r4 PubMed PMID: WOS:000263823200010. PMID: 19134172
5. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. P Natl Acad Sci USA. 2009; 106(38):16422–7. https://doi.org/10.1073/pnas.0906627106 PubMed PMID: WOS:000270071600076. PMID: 19805314
6. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. 2009; 6(10):767–U21. https://doi.org/10.1038/nmeth.1377 PubMed PMID: WOS:000270355200023. PMID: 19767758
7. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 2009; 19(12):2308–16. https://doi.org/10.1101/gr.097097.109 PubMed PMID: WOS:000272273400015. PMID: 19826075
8. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. Cell Host Microbe. 2009; 6(3):279–89. https://doi.org/10.1016/j.chom.2009.08.003 PubMed PMID: WOS:000270290700011. PMID: 19748469
9. Michielse CB, Hooykaas PJJ, van den Hondel CAMJJ, Ram AFJ. Agrobacterium-mediated transformation as a tool for functional genomics in fungi. Curr Genet. 2005; 48(1):1–17. https://doi.org/10.1007/s00294-005-0578-0 PubMed PMID: WOS:000230624600001. PMID: 15889258
10. Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, et al. A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. P Natl Acad Sci USA. 2006; 103(27):10352–7. https://doi.org/10.1073/pnas.0601456103 PubMed PMID: WOS:000239069400037. PMID: 16801547
11. Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell. 2008; 135(1):174–88. https://doi.org/10.1016/j.cell.2008.07.046 PMID: 18854164; PubMed Central PMCID: PMCPMC2628477.
12. Crimmins GT, Mohammadi S, Green ER, Bergman MA, Isberg RR, Mecsas J. Identification of MrtAB, an ABC Transporter Specifically Required for Yersinia pseudotuberculosis to Colonize the Mesenteric Lymph Nodes. PLoS Pathog. 2012; 8(8). doi: ARTN e1002828 https://doi.org/10.1371/journal.ppat.1002828 PubMed PMID: WOS:000308558000009. PMID: 22876175
13. van Opijnen T, Camilli A. A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res. 2012; 22(12):2541–51. https://doi.org/10.1101/gr.137430.112 PubMed PMID: WOS:000311895500022. PMID: 22826510
14. Wang ND, Ozer EA, Mandel MJ, Hauser AR. Genome-Wide Identification of Acinetobacter baumannii Genes Necessary for Persistence in the Lung. Mbio. 2014; 5(3). doi: ARTN e01163-14 https://doi.org/10.1128/mBio.01163-14 PubMed PMID: WOS:000338875900069. PMID: 24895306
15. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, et al. Genome-wide identification of bacterial plant colonization genes. PLoS Biol. 2017; 15(9):e2002860. https://doi.org/10.1371/journal.pbio.2002860 PMID: 28938018.
16. Le Breton Y, Belew AT, Freiberg JA, Sundar GS, Islam E, Lieberman J, et al. Genome-wide discovery of novel M1T1 group A streptococcal determinants important for fitness and virulence during soft-tissue infection. PLoS Pathog. 2017; 13(8):e1006584. https://doi.org/10.1371/journal.ppat.1006584 PMID: 28832676.
17. Bronner IF, Otto TD, Zhang M, Udenze K, Wang CQ, Quail MA, et al. Quantitative insertion-site sequencing (QIseq) for high throughput phenotyping of transposon mutants. Genome Res. 2016; 26(7):980–9. https://doi.org/10.1101/gr.200279.115 PubMed PMID: WOS:000378986000011. PMID: 27197223
18. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature. 2006; 444(7115):97–101. https://doi.org/10.1038/nature05248 PubMed PMID: WOS:000241701500053. PMID: 17080091
19. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomic. 2015; 15(2):141–61. https://doi.org/10.1007/s10142-015-0433-4 PubMed PMID: WOS:000351397700003. PMID: 25722247
20. Schnable PS, Ware D, Fulton RS, Stein JC, Wei FS, Pasternak S, et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science. 2009; 326(5956):1112–5. https://doi.org/10.1126/science.1178534 PubMed PMID: WOS:000271951000044. PMID: 19965430

21. Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014; 24(12):2033–40. https://doi.org/10.1101/gr.177881.114 PubMed PMID: WOS:000345810600011. PMID: 25079858

22. Stern DL. Tagmentation-Based Mapping (TagMap) of Mobile DNA Genomic Insertion Sites. bioRxiv. 2017. https://doi.org/10.1101/037762

23. Engler C, Gruetzner R, Kandzia R, Marillonnet S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS ONE. 2009; 4(5). doi: ARTN e5553 https://doi.org/10.1371/journal.pone.0005553 PubMed PMID: WOS:000266107300017. PMID: 19436741

24. Schilling L, Matei A, Redkar A, Walbot V, Doehlemann G. Virulence of the maize smut Ustilago maydis is shaped by organ-specific effectors. Mol Plant Pathol. 2014; 15(8):780–9. https://doi.org/10.1111/mpp.12133 PubMed PMID: WOS:000342131900003. PMID: 25346968

25. Schipper K, Brefort T, Doehlemann G, Djamei A, Muench K, Kahmann R. The secreted protein Stp1 is cruical for establishment of the biotrophic interaction of the smut fungus Ustilago maydis with its host plant maize. Eur J Cell Biol. 2008; 87:29–. PubMed PMID: WOS:000255316100068.

26. Doehlemann G, van der Linde K, Amann D, Schwammbach D, Hof A, Mohanty A, et al. Pep1, a Secreted Effector Protein of Ustilago maydis, Is Required for Successful Invasion of Plant Cells. PLoS Pathog. 2009; 5(2). doi: ARTN e1000290 https://doi.org/10.1371/journal.ppat.1000290 PubMed PMID: WOS:000263928000034. PMID: 19197359

27. Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the Ustilago maydis-Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. PLoS Pathog. 2013; 9(2). doi: ARTN e1003177 https://doi.org/10.1371/journal.ppat.1003177 PubMed PMID: WOS:000315648900027. PMID: 23459172

28. Stirnberg A, Djamei A. Characterization of ApB73, a virulence factor important for colonization of Zea mays by the smut Ustilago maydis. Mol Plant Pathol. 2016; 17(9):1467–79. https://doi.org/10.1111/mpp.12442 PubMed PMID: WOS:000389134900013. PMID: 27279632

29. Eichhorn H, Lessing F, Winterberg B, Schirawski J, Kamper J, Muller P, et al. A ferroxidation/permeation iron uptake system is required for virulence in Ustilago maydis. Plant Cell. 2006; 18(11):3332–45. https://doi.org/10.1105/tpc.106.043588 PubMed PMID: WOS:000243093700035. PMID: 17138696

30. Lanver D, Muller AN, Happel P, Schweizer G, Haas FB, Franitza M, et al. The biotrophic development of Ustilago maydis studied by RNAseq analysis. Plant Cell. 2018. https://doi.org/10.1105/tpc.17.00764 PMID: 29371439.

31. Gray AN, Koo BM, Shiver AL, Peters JM, Osadnik H, Gross CA. High-throughput bacterial functional genomics in the sequencing era. Curr Opin Microbiol. 2015; 27:86–95. https://doi.org/10.1016/j.mib.2015.07.012 PubMed PMID: WOS:000365065400014. PMID: 26336012

32. Armbruster CE, Forsyth-DeOrnellas V, Johnson AO, Smith SN, Zhao L, Wu W, et al. Genome-wide transposon mutagenesis of Proteus mirabilis: Essential genes, fitness factors for catheter-associated urinary tract infection, and the impact of polymicrobial infection on fitness requirements. PLoS Pathog. 2017; 13(6):e1006434. https://doi.org/10.1371/journal.ppat.1006434 PMID: 28614382; PubMed Central PMCID: PMCPMC5484520.

33. Kemppainen M, Duplessis S, Martin F, Pardo AG. T-DNA insertion, plasmid rescue and integration analysis in the model mycorrhizal fungus Laccaria bicolor. Microb Biotechnol. 2008; 1(3):258–69. https://doi.org/10.1111/j.1751-7915.2008.00029.x PMID: 21261845; PubMed Central PMCID: PMCPMC3815887.

34. Honkanen S, Jones VAS, Morieri G, Champion C, Hetherington AJ, Kelly S, et al. The Mechanism Forming the Cell Surface of Tip-Growing Rooting Cells Is Conserved among Land Plants. Current Biology. 2016; 26(23):3238–44. https://doi.org/10.1016/j.cub.2016.09.062 PubMed PMID: WOS:000389590500032. PMID: 27866889

35. Walter MC, Rattei T, Arnold R, Guldener U, Munsterkotter M, Nenova K, et al. PEDANT covers all complete RefSeq genomes. Nucleic Acids Research. 2009; 37:D408–D11. https://doi.org/10.1093/nar/gkn749 PubMed PMID: WOS:000261906200074. PMID: 18940859

36. Rabe F, Seitner D, Bauer L, Navarrete F, Czedik-Eysenberg A, Rabanal FA, et al. Phytohormone sensing in the biotrophic fungus Ustilago maydis—the dual role of the transcription factor Rss1. Mol Microbiol. 2016; 102(2):290–305. https://doi.org/10.1111/mmi.13460 PubMed PMID: WOS:000386101000009. PMID: 27387604

37. Khrunyk Y, Munch K, Schipper K, Lupas AN, Kahmann R. The use of FLP-mediated recombination for the functional analysis of an effector gene family in the biotrophic smut fungus Ustilago maydis. New Phytol. 2010; 187(4):957–68. https://doi.org/10.1111/j.1469-8137.2010.03413.x PubMed PMID: WOS:000280998600012. PMID: 20673282

38. Sambrook J, Russell DW, Sambrook J. The condensed protocols from Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2006. v, 800 p. p.

39. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74(12):5463–7. PMID: 271968; PubMed Central PMCID: PMCPMC431765.

40. Kamper J. A PCR-based system for highly efficient generation of gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004; 271(1):103–10. https://doi.org/10.1007/s00438-003-0962-8 PMID: 14673645.

41. Hoffman CS, Winston F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. Gene. 1987; 57(2–3):267–72. PMID: 3319781.

42. Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics. 2013; 29(21):2790–1. https://doi.org/10.1093/bioinformatics/btt468 PMID: 23975764.

43. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017; 27(3):491–9. https://doi.org/10.1101/gr.209601.116 PubMed PMID: WOS:000395694000015. PMID: 28100584

44. Pflug FG, von Haeseler A. TRUmiCount: Correctly counting absolute numbers of molecules using unique molecular identifiers. bioRxiv. 2017. https://doi.org/10.1101/217778

45. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57(1):289–300. http://www.jstor.org/stable/2346101?seq=1#page_scan_tab_contents

Publication 2

PLOS | PATHOGENS

PEARLS

Effectors of plant-colonizing fungi and beyond

Simon Uhse, Armin Djamei*

Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Vienna, Austria

* armin.djamei@gmi.oeaw.ac.at

Plant–microbe interactions have evolved over hundreds of millions of years, generating a diversity of interactions covering a broad continuum from pathogenic to mutualistic coexistence. Although these different lifestyles have different needs, they all bear in common the use of secreted molecules, termed "effectors", that enable microbes to interact with their hosts and to influence the outcome of the interaction. Effectors are not distinguished by sharing similar chemical properties but are instead defined by their function within the biological context of an interaction. To understand effectors, one needs to understand the coevolutionary forces that shape them. The host defense system is a major selection force that eradicates pathogens with a nonadapted effector repertoire. Reciprocally, host plants only survive the evolutionary race if they have been selected to recognize and defend against invading pathogens. This ongoing coevolution creates complex interdependencies between the effector repertoire of microbes, their effectome, and the host susceptibility machinery and defense system of their host plants. This review will summarize recent advances made in the field of effector studies in filamentous plant-colonizing microbes.

Citation: Uhse S, Djamei A (2018) Effectors of plant-colonizing fungi and beyond. PLoS Pathog 14(6): e1006992. https://doi.org/10.1371/journal.ppat.1006992

Editor: Cyril Zipfel, THE SAINSBURY LABORATORY, UNITED KINGDOM

Published: June 7, 2018

Funding: Funding for effector research in my laboratory is provided by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) / ERC grant agreement n° [ERC grant agreement n° [GA335691 „Effectomics"], the Austrian Science Fund (FWF): [P27429-B22, P27818-B22, I 3033-B22], and the Austrian Academy of Science (OEAW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Effector gene expression—Being in the right place at the right time

Each produced effector can be considered as an investment that needs to pay off by giving a selective advantage to the invader, at least from time to time across generations, to be kept in the population. As many effectors are tools that redirect host metabolism and development, their dosage and timing should be controlled to achieve an optimal, balanced result, especially in the case of biotrophs, which need to retain the viability of their host. Evidence for the tight control of effector synthesis and their place and mode of secretion has been provided from various filamentous pathogens [1–4]. Lifestyle switches, e.g., from biotrophic to necrotrophic, or host switches require profound changes in the applied effector cocktail [5]. The same is true when changing environments within the host, e.g., by moving between organs, as exemplified for the biotrophic maize pathogen *Ustilago maydis* [6]. Growing evidence supports the view that adapting the composition of produced effectors to external cues and developmental requirements is a general feature of interspecies interactions. Infection-phase–specific expression of putative effectors has been demonstrated by transcriptomic time-course experiments, among others, in the obligate biotrophic poplar leaf rust *Melampsora larici-populina* [7]; the hemibiotrophic fungus *Colletotrichum higginsianum*, which causes anthracnose during *Arabidopsis thaliana* infection [8]; the obligate biotrophic barley fungus *Blumeria graminis* [9]; the root mutualistic fungus *Serendipita indica* (former *Piriformospora indica*) [5]; and the maize-infecting biotroph *U. maydis* [10]. Adaptation of effector secretion and/or expression may even be cell-type-specific, although this hypothesis lacks experimental support, likely because of technical challenges. An emerging concept is that adaptation of effector expression is not limited to developmental programs of the pathogen or infection strategies in different hosts or plant organs but also occurs when the host plant

is challenged by abiotic stresses. Transcriptomic studies on rice under mild drought stress showed that the hemibiotrophic fungus *Magnaporthe oryzae* transcriptionally downregulates the majority of its putative effectors despite being more successful in colonizing the stressed plants [11]. All these examples of adapted effector expression imply that specific environmental signals must be perceived during colonization by the invading microbes. On the pathogen side, very little is known about what these external signals are and how they are perceived, especially after infection [12, 13]. As misregulation of effectors has been shown to reduce pathogenicity in various pathogens, manipulating effector expression via these external cues could be an elegant way to interfere with pathogen infections [4, 14]. Studying the underlying regulatory networks controlling effector expression is an important future research direction.

Enigmatic effector translocation and place of action

A common hallmark of effectors is that they are, in one way or the other, secreted. Their place of action is therefore either in the interphase between the microbe and the host cell (apoplastic effectors) or inside the host cell (translocated/symplastic effectors). The term "symplastic effector" embodies the idea that translocated effectors might not be restricted to a single cell and includes all possible places of action within plant cells. Similar to the spreading of effectors within the symplast, effectors might diffuse within the apoplast and therefore act on several cells. Within these two compartments, further subcompartments can be delimited. Within the apoplast, effectors have been identified that bind fungal cell wall components, potentially to protect their degradation or recognition by plant pattern-recognition receptors [15, 16]. Other effectors act in the biotrophic interphase, e.g., as inhibitors of apoplastic proteases or to bind pathogen-associated molecular patterns (PAMPs) to reduce recognition [17, 18].

We are not aware of any effector being identified with targets associated with the host plasma membrane from the apoplastic side, and only a few have been identified acting from the cytosolic side at the membrane, likely because of technical limitations in identifying these interactions [19–21].

Type III secretion signals from bacteria and RXLR-dEER or LXLFLAK motifs from oomycetes are predicted to be translocation signals (although in case of RXLR-dEER, its role in uptake is under debate [22]), which make the prediction of symplastic effectors possible in these systems [23, 24]. For fungi, RXLR-like signals leading to translocation of fungal effectors have been controversially discussed but have not been confirmed [25, 26]. Experimental evidence for translocation has been generated either directly by fusing fluorescent proteins to effectors [27] or through immunoelectron-microscopy approaches [28, 29] or are inferred by cytosolic resistance gene (R-gene)–based recognition of avirulence (Avr) effectors [30]. Experimental results for the rust symplastic effector AvrM indicate a host-cell autonomous translocation [29, 31], which implies that AvrM harbors intrinsic biochemical properties mediating its translocation. In contrast to this, the effector Avr2 of *Fusarium oxysporum* does not show such properties but instead requires a pathogen-derived trigger for translocation [32]. The differences observed between pathosystems make it likely that the mechanisms of translocation into the host cell might differ between fungal species and potentially even between different symplastic effectors within a species [27–29, 33]. After translocation into the host cell, symplastic effectors might target specific host compartments. Transgenic production of effector proteins without signal peptides in plant cells have indicated specific localization for effectors in the nucleus, nucleoli, chloroplasts, mitochondria, and discrete cellular bodies [34, 35].

Effector functions—Avoid the alarm, activate what serves, and inhibit what harms

The functions that need to be covered by an effectome reflect the challenges presented by the host immune machinery and mirror the specific needs of the pathogen and its lifestyle. While effectors of biotrophs often function in suppression of host immunity, the necrotrophic fungus *Cochliobolus victoriae* targets a defense-associated thioredoxin TRX-h5 guarded by the NB-LRR protein LOV1 via the toxin effector victorin. The LRR recognition leads to host defense responses, conferring disease susceptibility to the necrotroph [36].

Looking at so-far-identified effector functions, one can identify different modes of action serving the strategies for successful host invasion illustrated in Fig 1.

The self-binder and self-modifier

Effectors with a defensive mode of action either sequester potential microbe-associated molecular patterns (MAMPs) or modify their cell walls upon penetration to minimize recognition. Examples include the chitin-oligomer-chelating LysM effectors Ecp6 of *Cladosporium fulvum* or the Slp1 LysM effector of *M. oryzae* [18, 37]. Another effector passively protects from antimicrobial counter attack [16, 38].

The inhibitor

Many effectors have classic inhibitory activities, e.g., against immune-related proteases, glucanases, or peroxidases, but also against intracellular signaling components to interfere with defense-related signaling processes [39–42]. Inhibition of the Jasmonic-acid–triggered degradation of PtJAZ6 by the MiSSP7 *Laccaria bicolor* effector is an example of signaling suppression by a mutualistic fungus [43].

The activator

Only a few effectors have been identified that clearly fall into the activator category, probably as evolution of inhibitory activity is more likely. The NUDIX hydrolase effector Avr3b of *Phytophthora sojae* and the deregulated, secreted chorismate mutase Cmu1 of *U. maydis* are examples [28, 44]. Some activating effectors function by interfering with the deactivation or degradation of their interacting host protein, thereby acting positively, although they are basically an inhibitor type of effector. One example is the *U. maydis* effector Tin2, which stabilizes the maize kinase TKK1 by inhibiting its degradation [33].

Most effector functions are usually inferred via the host interaction partners, as many effectors show low conservation on the sequence level because of high selection pressure to evade host recognition. One conceptional restriction is that effectors might interact with host molecules either to target and manipulate them or to use them as part of the host cellular machinery to reach their final destination. For example, an effector with a nuclear localization signal might interact with Importin α to enter the host nucleus, but its ultimate target might be the inhibition of a specific host transcription factor. Some effectors have a broader target spectrum, as exemplified by EPIC2B, a cystatin-domain-containing, protease-inhibiting effector from *Phytophtora infestans* [45]. Other effectors show a high degree of specificity even when they target members of expanded protein families, as is the case for the *M. oryzae* effector AvrPii, which targets specific vesicle-tethering Exo70 subunits involved in host immune responses, or the *P. infestans* effector PexRD54, which targets a specific autophagy-modulating ubiquitin-like ATG8 family member [46, 47].

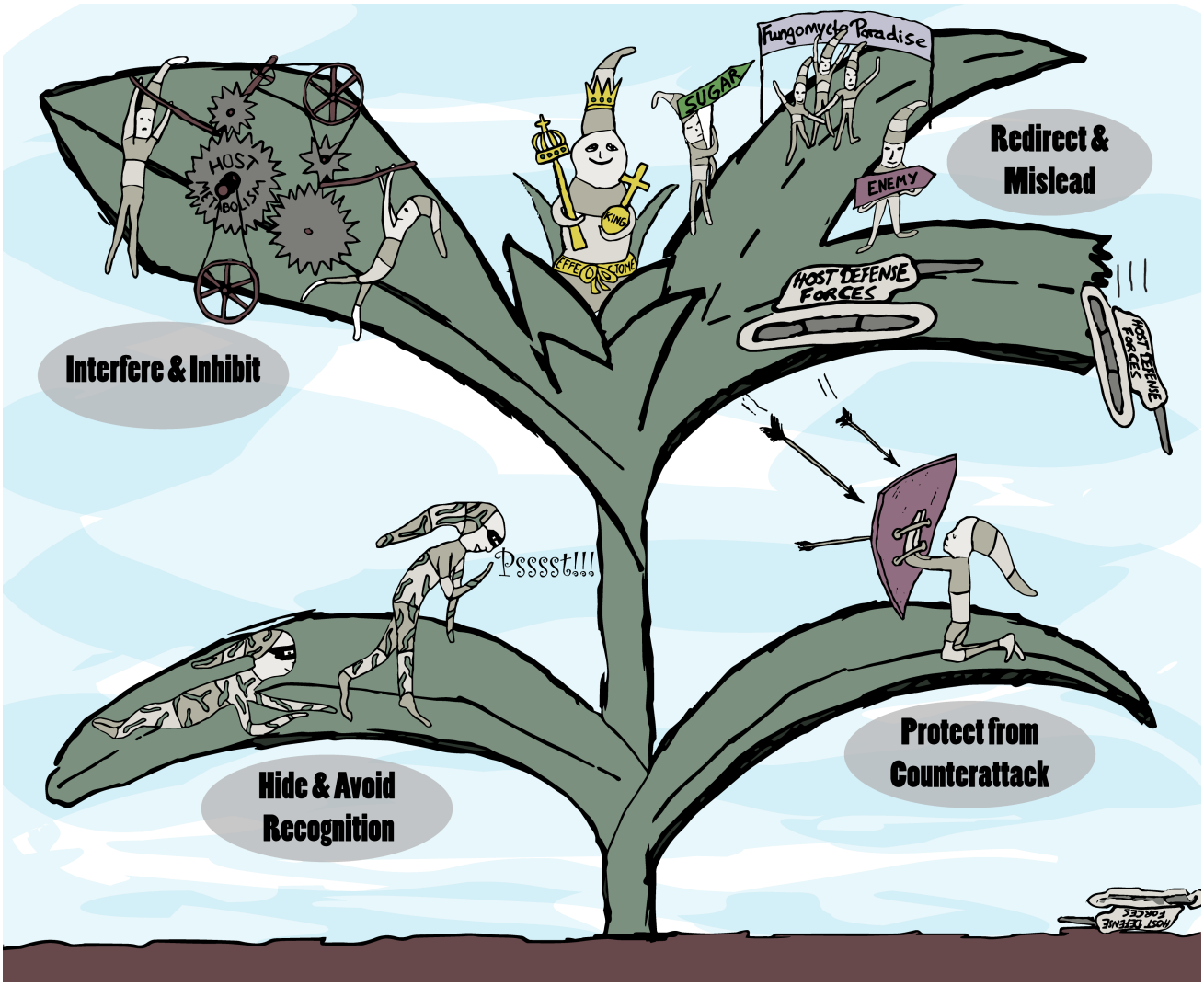


Fig 1. Strategies for successful host invasion. Plant-colonizing microbes employ effectors fulfilling various functions during the host invasion, which are visualized symbolically in this cartoon. Different modes of action (self-binding and self-modifying, activating or inhibiting activities) of effectors described in the text may be applied to serve the listed strategies (text on grey oval background).

https://doi.org/10.1371/journal.ppat.1006992.g001

Large-scale effector/host ORFeome interaction screens demonstrated that effector targets are usually well-connected cellular hubs [48, 49]. Furthermore, these and other studies revealed that effectors often converge on the same host targets [50]. This goes hand in hand with independent observations that many effector deletion strains do not show any observable virulence defect, potentially a reflection of functional redundancy [51]. Functional redundancy likely provides robustness to host-colonization success and could be considered a sign that the target is of specific importance for a successful interaction. This is supported by a correlation between converging effector-target–deletion plants often showing altered immune-response phenotypes [49].

The decoy-domain fusions found in many nucleotide binding domain and leucine-rich repeat receptor (NLR) proteins might represent effector-target mimics. This, among others, has been experimentally validated for the WRKY domain containing NLR RRS1-R [52]. Therefore, sensor domains fused to NLRs might serve as an informative way to preselect

common effector targets [53]. While effectors also target directly defense components, they more commonly target defense modulators, e.g., by exploiting antagonistic hormone pathways that promote both growth and development, thereby inhibiting immunity [48]. This could be a coevolutionary consequence of the host immune system being less able to detect manipulation of modulators that are involved in various processes beyond immunity.

Outlook

Within the context of the host metabolism, effectors act as alien molecules, overrunning feedback control systems that usually maintain homeostasis [33]. For this reason, they are valuable dominant acting molecular tools. Effectors teach us not only about the molecular defense machinery of the host but often disclose the wiring between immunity, growth, and developmental host pathways. Like a molecular language, effectors coevolve with the host population the invader needs to communicate with. Our understanding of this language is still in the early stages, and thousands of effectomes await to be understood. However, being able to translate this language will likely reward us with immense payback both in strategies for preventing pathogen infections and tools for understanding plant biology.

Acknowledgments

First, I would like to acknowledge the concepts and ideas I have picked up during discussions with my valuable colleagues, at conferences, or in the literature that I was unable to cite because of space constraints. Special thanks go to Sophien Kamoun for mentorship and for scientific discussions on this review. Many thanks also to J. Matthew Watson and Gesa Hoffmann for proof-reading and editing. AD wrote the manuscript and designed the cartoon, SU rearranged and digitalized the cartoon.

References

1. Bielska E, Higuchi Y, Schuster M, Steinberg N, Kilaru S, Talbot NJ, et al. Long-distance endosome trafficking drives fungal effector production during plant infection. Nat Commun. 2014; 5:5097. https://doi.org/10.1038/ncomms6097 PMID: 25283249.
2. Giraldo MC, Dagdas YF, Gupta YK, Mentlak TA, Yi M, Martinez-Rocha AL, et al. Two distinct secretion systems facilitate tissue invasion by the rice blast fungus Magnaporthe oryzae. Nat Commun. 2013; 4. https://doi.org/10.1038/ncomms2996 PMID: 23774898
3. Wang S, Boevink PC, Welsh L, Zhang R, Whisson SC, Birch PRJ. Delivery of cytoplasmic and apoplastic effectors from Phytophthora infestans haustoria by distinct secretion pathways. New Phytol. 2017; 216(1):205–15. https://doi.org/10.1111/nph.14696 PMID: 28758684.
4. Tollot M, Assmann D, Becker C, Altmuller J, Dutheil JY, Wegner CE, et al. The WOPR Protein Ros1 Is a Master Regulator of Sporogenesis and Late Effector Gene Expression in the Maize Pathogen Ustilago maydis. PLoS Pathog. 2016; 12(6). https://doi.org/10.1371/journal.ppat.1005697 PMID: 27332891
5. Lahrmann U, Ding Y, Banhara A, Rath M, Hajirezaei MR, Dohlemann S, et al. Host-related metabolic cues affect colonization strategies of a root endophyte. Proc Natl Acad Sci USA. 2013; 110(34):13965–70. https://doi.org/10.1073/pnas.1301653110 PMID: 23918389.
6. Skibbe DS, Doehlemann G, Fernandes J, Walbot V. Maize tumors caused by Ustilago maydis require organ-specific genes in host and pathogen. Science. 2010; 328(5974):89–92. Epub 2010/04/03. https://doi.org/10.1126/science.1185775 PMID: 20360107.
7. Duplessis S, Hacquard S, Delaruelle C, Tisserant E, Frey P, Martin F, et al. Melampsora larici-populina Transcript Profiling During Germination and Timecourse Infection of Poplar Leaves Reveals Dynamic Expression Patterns Associated with Virulence and Biotrophy. Mol Plant Microbe In. 2011; 24(7):808–18. https://doi.org/10.1094/Mpmi-01-11-0006 PMID: 21644839
8. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, et al. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. Nat Genet. 2012; 44(9):1060–5. https://doi.org/10.1038/ng.2372 PMID: 22885923.
9. Hacquard S, Kracher B, Maekawa T, Vernaldi S, Schulze-Lefert P, Ver Loren van Themaat E. Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs

in divergent hosts. Proc Natl Acad Sci USA. 2013; 110(24):E2219–28. https://doi.org/10.1073/pnas.1306807110 PMID: 23696672.

10. Lanver D, Muller AN, Happel P, Schweizer G, Haas FB, Franitza M, et al. The biotrophic development of Ustilago maydis studied by RNAseq analysis. Plant Cell. 2018. https://doi.org/10.1105/tpc.17.00764 PMID: 29371439.

11. Bidzinski P, Ballini E, Ducasse A, Michel C, Zuluaga P, Genga A, et al. Transcriptional Basis of Drought-Induced Susceptibility to the Rice Blast Fungus Magnaporthe oryzae. Front Plant Sci. 2016; 7:1558. https://doi.org/10.3389/fpls.2016.01558 PMID: 27833621.

12. Mendoza-Mendoza A, Berndt P, Djamei A, Weise C, Linne U, Marahiel M, et al. Physical-chemical plant-derived signals induce differentiation in Ustilago maydis. Mol Microbiol. 2009; 71(4):895–911. Epub 2009/01/28. https://doi.org/10.1111/j.1365-2958.2008.06567.x PMID: 19170880.

13. Lanver D, Mendoza-Mendoza A, Brachmann A, Kahmann R. Sho1 and Msb2-related proteins regulate appressorium development in the smut fungus Ustilago maydis. Plant Cell. 2010; 22(6):2085–101. Epub 2010/07/01. https://doi.org/10.1105/tpc.109.073734 PMID: 20587773.

14. Michielse CB, van Wijk R, Reijnen L, Manders EMM, Boas S, Olivain C, et al. The Nuclear Protein Sge1 of Fusarium oxysporum Is Required for Parasitic Growth. PLoS Pathog. 2009; 5(10). https://doi.org/10.1371/journal.ppat.1000637 PMID: 19851506

15. Wawra S, Fesel P, Widmer H, Timm M, Seibel J, Leson L, et al. The fungal-specific beta-glucan-binding lectin FGB1 alters cell-wall composition and suppresses glucan-triggered immunity in plants. Nat Commun. 2016; 7. https://doi.org/10.1038/ncomms13188 PMID: 27786272

16. van den Burg HA, Harrison SJ, Joosten MHAJ, Vervoort J, de Wit PJGM. Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. Mol Plant Microbe In. 2006; 19(12):1420–30. https://doi.org/10.1094/Mpmi-19-1420 PMID: 17153926

17. Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the Ustilago maydis-Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. PLoS Pathog. 2013; 9(2). https://doi.org/10.1371/journal.ppat.1003177 PMID: 23459172

18. de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, et al. Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science. 2010; 329(5994):953–5. https://doi.org/10.1126/science.1190859 PMID: 20724636.

19. Wu J, van der Burgh AM, Bi G, Zhang L, Alfano JR, Martin GB, et al. The Bacterial Effector AvrPto Targets the Regulatory Coreceptor SOBIR1 and Suppresses Defense Signaling Mediated by the Receptor-Like Protein Cf-4. Mol Plant Microbe Interact. 2018; 31(1):75–85. https://doi.org/10.1094/MPMI-08-17-0203-FI PMID: 28876174.

20. Lu D, He P, Shan L. Bacterial effectors target BAK1-associated receptor complexes: One stone two birds. Commun Integr Biol. 2010; 3(2):80–3. PMID: 20585495.

21. Cao L, Blekemolen MC, Tintor N, Cornelissen BJC, Takken FLW. The Fusarium oxysporum Avr2-Six5 Effector Pair Alters Plasmodesmatal Exclusion Selectivity Facilitating Cell-to-Cell Movement of Avr2. Mol Plant. 2018; 11(5):691–705. https://doi.org/10.1016/j.molp.2018.02.011 PMID: 29481865.

22. Wawra S, Trusch F, Matena A, Apostolakis K, Linne U, Zhukov I, et al. The RxLR Motif of the Host Targeting Effector AVR3a of Phytophthora infestans Is Cleaved before Secretion. Plant Cell. 2017; 29(6):1184–95. https://doi.org/10.1105/tpc.16.00552 PMID: 28522546

23. Birch PR, Rehmany AP, Pritchard L, Kamoun S, Beynon JL. Trafficking arms: oomycete effectors enter host plant cells. Trends Microbiol. 2006; 14(1):8–11. Epub 2005/12/17. https://doi.org/10.1016/j.tim.2005.11.007 PMID: 16356717.

24. Cornelis GR, Van Gijsegem F. Assembly and function of type III secretory systems. Annu Rev Microbiol. 2000; 54:735–74. https://doi.org/10.1146/annurev.micro.54.1.735 PMID: 11018143.

25. Kale SD, Gu BA, Capelluto DGS, Dou DL, Feldman E, Rumore A, et al. External Lipid PI3P Mediates Entry of Eukaryotic Pathogen Effectors into Plant and Animal Host Cells. Cell. 2010; 142(2):284–95. https://doi.org/10.1016/j.cell.2010.06.008 PMID: 20655469

26. Wawra S, Djamei A, Albert I, Nurnberger T, Kahmann R, van West P. In vitro translocation experiments with RxLR-reporter fusion proteins of Avr1b from Phytophthora sojae and AVR3a from Phytophthora infestans fail to demonstrate specific autonomous uptake in plant and animal cells. Mol Plant Microbe Interact. 2013; 26(5):528–36. Epub 2013/04/04. https://doi.org/10.1094/MPMI-08-12-0200-R PMID: 23547905.

27. Khang CH, Berruyer R, Giraldo MC, Kankanala P, Park SY, Czymmek K, et al. Translocation of Magnaporthe oryzae Effectors into Rice Cells and Their Subsequent Cell-to-Cell Movement. Plant Cell. 2010; 22(4):1388–403. https://doi.org/10.1105/tpc.109.069666 PMID: 20435900

28. Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, et al. Metabolic priming by a secreted fungal effector. Nature. 2011; 478(7369):395–8. Epub 2011/10/07. https://doi.org/10.1038/nature10454 PMID: 21976020.

29. Rafiqi M, Gan PH, Ravensdale M, Lawrence GJ, Ellis JG, Jones DA, et al. Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the absence of the pathogen. Plant Cell. 2010; 22(6):2017–32. Epub 2010/06/08. https://doi.org/10.1105/tpc.109.072983 PMID: 20525849.

30. Dodds PN, Lawrence GJ, Catanzariti AM, Teh T, Wang CI, Ayliffe MA, et al. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc Natl Acad Sci USA. 2006; 103(23):8888–93. https://doi.org/10.1073/pnas.0602577103 PMID: 16731621.

31. Ve T, Williams SJ, Catanzariti AM, Rafiqi M, Rahman M, Ellis JG, et al. Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. P Natl Acad Sci USA. 2013; 110(43):17594–9. https://doi.org/10.1073/pnas.1307614110 PMID: 24101475

32. Di X, Gomila J, Ma L, van den Burg HA, Takken FL. Uptake of the Fusarium Effector Avr2 by Tomato Is Not a Cell Autonomous Event. Front Plant Sci. 2016; 7:1915. https://doi.org/10.3389/fpls.2016.01915 PMID: 28066471.

33. Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, et al. A secreted Ustilago maydis effector promotes virulence by targeting anthocyanin biosynthesis in maize. Elife. 2014; 3: e01355. https://doi.org/10.7554/eLife.01355 PMID: 24473076

34. Petre B, Saunders DGO, Sklenar J, Lorrain C, Win J, Duplessis S, et al. Candidate Effector Proteins of the Rust Pathogen Melampsora larici-populina Target Diverse Plant Cell Compartments. Mol Plant Microbe In. 2015; 28(6):689–700. https://doi.org/10.1094/Mpmi-01-15-0003-R PMID: 25650830

35. Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep-Uk. 2017; 7. https://doi.org/10.1038/srep44598 PMID: 28300209

36. Lorang J, Kidarsa T, Bradford CS, Gilbert B, Curtis M, Tzeng SC, et al. Tricking the Guard: Exploiting Plant Defense for Disease Susceptibility. Science. 2012; 338(6107):659–62. https://doi.org/10.1126/science.1226743 PMID: 23087001

37. Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, et al. Effector-Mediated Suppression of Chitin-Triggered Immunity by Magnaporthe oryzae Is Necessary for Rice Blast Disease. Plant Cell. 2012; 24(1):322–35. https://doi.org/10.1105/tpc.111.092957 PMID: 22267486

38. van Esse HP, Bolton MD, Stergiopoulos I, de Wit PJGM, Thomma BPHJ. The chitin-binding Cladosporium fulvum effector protein Avr4 is a virulence factor. Mol Plant Microbe In. 2007; 20(9):1092–101. https://doi.org/10.1094/Mpmi-20-9-1092 PMID: 17849712

39. Shabab M, Shindo T, Gu C, Kaschani F, Pansuriya T, Chintha R, et al. Fungal effector protein AVR2 targets diversifying defense-related Cys proteases of tomato. Plant Cell. 2008; 20(4):1169–83. https://doi.org/10.1105/tpc.107.056325 PMID: 18451324

40. Rose JKC, Ham KS, Darvill AG, Albersheim P. Molecular cloning and characterization of glucanase inhibitor proteins: Coevolution of a counterdefense mechanism by plant pathogens. Plant Cell. 2002; 14(6):1329–45. https://doi.org/10.1105/tpc.002253 PMID: 12084830

41. Yamaguchi K, Yamada K, Ishikawa K, Yoshimura S, Hayashi N, Uchihashi K, et al. A Receptor-like Cytoplasmic Kinase Targeted by a Plant Pathogen Effector Is Directly Phosphorylated by the Chitin Receptor and Mediates Rice Immunity. Cell Host & Microbe. 2013; 13(3):347–57. https://doi.org/10.1016/j.chom.2013.02.007 PMID: 23498959

42. Hemetsberger C, Herrberger C, Zechmann B, Hillmer M, Doehlemann G. The Ustilago maydis effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. PLoS Pathog. 2012; 8(5):e1002684. Epub 2012/05/17. https://doi.org/10.1371/journal.ppat.1002684 PMID: 22589719.

43. Plett JM, Daguerre Y, Wittulsky S, Vayssieres A, Deveau A, Melton SJ, et al. Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. P Natl Acad Sci USA. 2014; 111(22):8299–304. https://doi.org/10.1073/pnas.1322671111 PMID: 24847068

44. Dong SM, Yin WX, Kong GH, Yang XY, Qutob D, Chen QH, et al. Phytophthora sojae Avirulence Effector Avr3b is a Secreted NADH and ADP-ribose Pyrophosphorylase that Modulates Plant Immunity. PLoS Pathog. 2011; 7(11): e1002353. https://doi.org/10.1371/journal.ppat.1002353 PMID: 22102810

45. Tian MY, Win J, Song J, van der Hoorn R, van der Knaap E, Kamoun S. A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiology. 2007; 143(1):364–77. https://doi.org/10.1104/pp.106.090050 PMID: 17085509

46. Dagdas YF, Belhaj K, Maqbool A, Chaparro-Garcia A, Pandey P, Petre B, et al. An effector of the Irish potato famine pathogen antagonizes a host autophagy cargo receptor. Elife. 2016; 5: e10856. https://doi.org/10.7554/eLife.10856 PMID: 26765567

47. Fujisaki K, Abe Y, Ito A, Saitoh H, Yoshida K, Kanzaki H, et al. Rice Exo70 interacts with a fungal effector, AVR-Pii, and is required for AVR-Pii-triggered immunity. Plant Journal. 2015; 83(5):875–87. https://doi.org/10.1111/tpj.12934 PMID: 26186703

48. Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, Moore J, et al. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science. 2011; 333 (6042):596–601. Epub 2011/07/30. https://doi.org/10.1126/science.1203659 PMID: 21798943.

49. Wessling R, Epple P, Altmann S, He YJ, Yang L, Henz SR, et al. Convergent Targeting of a Common Host Protein-Network by Pathogen Effectors from Three Kingdoms of Life. Cell Host & Microbe. 2014; 16(3):364–75. https://doi.org/10.1016/j.chom.2014.08.004 PMID: 25211078

50. Song J, Win J, Tian MY, Schornack S, Kaschani F, Ilyas M, et al. Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. P Natl Acad Sci USA. 2009; 106(5):1654–9. https://doi.org/10.1073/pnas.0809201106 PMID: 19171904

51. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature. 2006; 444(7115):97–101. Epub 2006/11/03. https://doi.org/10.1038/nature05248 PMID: 17080091.

52. Le Roux C, Huet G, Jauneau A, Camborde L, Tremousaygue D, Kraut A, et al. A Receptor Pair with an Integrated Decoy Converts Pathogen Disabling of Transcription Factors to Immunity. Cell. 2015; 161 (5):1074–88. https://doi.org/10.1016/j.cell.2015.04.025 PMID: 26000483

53. Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva KV. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. BMC Biol. 2016; 14. https://doi.org/10.1186/s12915-016-0228-7 PMID: 26891798

Discussion

In the first part of this work, a novel technique to elucidate virulence factors in an efficient and high-throughput approach was presented. The technique, called iPool-Seq, is very selective and specific, as it allows for isolation of flanking sequences of insertion cassettes directly from *in vivo* infected host material. iPool-Seq offers such a high efficiency by combining powerful fragmentation and adapter ligation features of Tn5 transposase with specific exponential PCR amplifications of flanking sequences originating from individual mutants. Moreover, the method overcomes PCR biases that can arise from exponential amplifications by implementation of unique molecular identifiers (UMIs) in the individual adapters that are incorporated at the beginning of the library preparation by the Tn5 transposase. Therefore, the method remains quantitative at all times, because all copies of individual UMIs can get merged during data analysis of NGS reads. Here, iPool-Seq was successfully tested for the first time on a pooled infection experiment with 195 *U. maydis* insertional mutants and yielded reproducibly 23 novel *U. maydis* virulence factors important for maize infection. This new technique is not only part of the toolbox for the analysis of smut fungi but also allows the analysis of any other insertional mutagenesis library, e.g. prepared with transposon mutagenesis or agrobacterium-mediated insertional mutagenesis.

Insertion mutagenesis library generation

Insertional mutagenesis is a powerful and widely used approach in genetics to elucidate phenotypes of targeted or random genes. Targeted insertional mutagenesis is most commonly accomplished by homologous recombination. To this end, a genetic marker gene including a promoter and terminator, like the bacterial hygromycin phosphotransferase gene (hph) cassette, which confers resistance against the antibiotic hygromycin, is flanked by sequences that are homologous to an integration locus of interest on the genome. On the one hand, integration of insertional cassettes by homologous recombination offers a high degree of precision and flexibility, i.e. any locus on the genome can be targeted. On the other hand, homologous recombination is a highly conserved genetic repair mechanism and hence, many biological model organisms can be transformed with homologous DNA sequences, including the filamentous fungus *U. maydis* (51, 66). In addition, the efficiency of transformation with homologous recombining constructs in *U. maydis* is high, yielding routinely more than 50% positive transformants. In this work, an insertional mutagenesis library with the solopathogenic haploid *U. maydis* strain SG200 was established by homologous recombination. The hph cassette was used for selection of the deletion mutants, consisting of hph gene under control of the heat shock protein 70 (hsp70) promoter and followed by the nos terminator (nosT), which was successfully introduced in *U. maydis* in the past (57). Despite its high efficiency and precision, homologous recombination is laborious and not ideal for fast generation of genome wide deletion libraries and only few holistic approaches were published in the past (67-69). Nonetheless, the deletion strain library of *U. maydis* was generated via homologous generation, because the method is well established in *U. maydis* and suitable for a subset of genes. In this respect, genes which are predicted to be secreted, have a short amino-acid sequence, have no known domains and which are upregulated during

biotrophy were selected for the library (46, 70). For all 195 genes of the final library two or more of these criteria were met. Therefore, genes investigated in this study are likely to encode effector proteins and to have a function in virulence. This set of 195 mutants was analyzed in pooled infections in a reverse genetics approach to establish iPool-Seq.

Alternative to targeted insertional mutagenesis by homologous recombination, high-throughput random mutagenesis approaches were developed for *in vivo* use in the last two decades. Most prominent are agrobacterium-mediated transformation (ATM) and transposon mutagenesis, which enable fast and efficient generation of large and genome-wide insertion libraries. ATM was implemented successfully in filamentous fungi, e.g. *Magnaporthe oryzae* or *Fusarium oxysporium* (71, 72). In both studies, the ATM libraries were used for negative depletion screens to identify virulence factors, yielding 202 and 111 pathogenicity loci, respectively. However, ATM comes with a clear disadvantage in comparison to the well-controlled homologous recombination: Higher transformation efficiencies increase the risk of multiple insertions of the transfer-DNA (T-DNA) in each individual cell. This danger is immanent and unavoidable, complicating the analysis of mutants with a phenotype. The study with *F. oxysporum* provided an estimation of multiple integration frequency based on few Southern blot analysis but did not provide the information for all mutants that displayed a virulence phenotype (72). The second study provided an extensive phenotypic analysis of mutants but showed exclusively PCR results of single T-DNA integration sites and did not address the problem of mutants with multiple integrations (71).

In contrast to ATM, transposon mutagenesis, or transposition, does not require another organism that confers transformation, but is based on natural occurring class II transposable elements that insert themselves in the genome following a cut-and-paste mechanism. *In vivo* transposition requires the action of a transposase in the nucleus

on the DNA transposon. The transposase gene can be encoded between the inverted repeats of the transposon or independently on the genome or a plasmid. The transposase mediates transposition by recognition of inverted repeat sequences flanking the transposon, followed by an induction of double-strand breaks and eventually, reintegration at a different locus (73). For heterologous transposition systems it is of importance that the transposase is capable to perform excision and integration of the transposon without any intrinsic co-factors. In contrast to ATM, transposition can be designed in a way that multiple insertions of the transposon per genome can be avoided. Transposition has been used successfully *in vivo* in various microbes, mainly bacteria, to identify essential genes (74-77) and was implemented recently in the plant symbiont *Pseudomonas simiae* in order to screen for genes that are required for host colonization (78). However, to date, there is no strikingly successful study using transposon insertion mutagenesis in filamentous fungi. Possibly, filamentous fungi are not accessible for heterologous transposons and evolved defense mechanisms, e.g. by RNA-silencing as shown in example study in animals (79). *U. maydis* is an exception and lacks most genes of the RNA-silencing machinery and thus could tolerate heterologous transposition *in vivo* (26). However, the successful generation of transposon insertion libraries in animals indicates, that filamentous fungi might have evolved another, yet unknown transposon defense mechanism (80, 81). A promising transposition system offers the piggyBac transposon, which was optimized for high activities in mammalian cell lines (82). Most importantly, the piggyBac transposon generally does not leave a footprint after its excision, that could result in a frame shift of an open reading frame of a protein coding gene, and exhibits an integration bias towards transcribed genes in mice (83). Both features are desirable, because the highly active transposase most likely induces several jumps of the transposon per genome and the main target in a genome-wide insertion screen are

protein-coding genes. Therefore, a forward-genetics screen with the piggyBac transposon offers an interesting alternative to homologous recombination for iPool-Seq.

Both insertional mutagenesis strategies, ATM and transposition, were tested in *U. maydis* in the past, however, with limited success (53, 54). ATM in *U. maydis* yielded a library of approximately 5000 mutants. However, downstream-analysis did not provide a detailed phenotypic characterization of the mutant library, resulting in two genes potentially involved in sexual reproduction of the fungus. Although, ATM itself seems to work efficiently in *U. maydis*, its beforementioned disadvantage of multiple integrations preclude the method most likely from future studies. In contrast, transposition offers more promising characteristics, but an attempt to generate a *U. maydis* mutant library with the *Caenorhabditis elegans* transposon Tc1 did not work efficiently (53). The nitrate reductase 1 (Nar1) locus of *U. maydis* was designed as a transposon trap. Transposon insertion in the Nar1 gene would confer chlorate resistance to the strain. However, none of the chlorate resistant strains identified displayed a Tc1 transposon integration in the *nar1* locus. The authors speculated, that transposition in *U. maydis* is not efficient, because the genome generally does not harbor any intact transposable elements and likely has evolved mechanisms to inhibit DNA-transposon propagation (26, 53). However, they did not provide any direct proof for the failure of the heterologous transposition system. Depending on transposition efficiency, transposable mutagenesis could be the method of choice to generate large mutant libraries to mutate all the 6,902 predicted protein-coding genes of *U. maydis* (26). A library of 20500 mutants would allow for an average mutation distance of 1000 bp in the 20.5-million-base pair genome and could offer minimal library size for the downstream analysis of mutants with iPool-Seq. However, such a library is not

available to date and the efficiency of transposable elements and corresponding transposase systems awaits further optimization in *U. maydis*.

More recently, novel methods for genome-wide mutagenesis have been developed based on Clustered Regularly Interspaced Short Palindromic Repeats (CrispR) - Cas9 system from *Streptococcus pyogenes*. CrispR-Cas9 methods are still under optimization and require the generation of genome-wide gRNA-libraries for all target sites. The first depletion screens with the system were carried out with human cell lines to identify new cancer targets and made use of the non-homologous end joining (NHEJ) repair mechanism that can incorporate insertions or deletions (Indels), possibly resulting in frame-shifts and loss of functions of targeted genes (84, 85). More recently, targeted methods using repair templates were established in bacteria and yeast allowing for precise targeted mutagenesis (86, 87). As an advantage over transposon mutagenesis, Cas9 does not require selection markers for mutant generation and is therefore less invasive. Furthermore, the mutagenesis targets can be edited specifically, even down to the nucleotide level using homologous repair cassettes. However, the CrispR-Cas9 system also comes with disadvantages, e.g. its dependency on PAM-sequences in case of *S. pyogenes* Cas9, and the possibility of off-target effects that can superimpose effects of the targeted gene. Nonetheless, CrispR-Cas9 is potentially the most promising future method for negative depletion screens. With decreasing plasmid library generation costs it is likely that the system will be implemented in plant pathogen models soon. Moreover, gene editing with *S. pyogenes* CrispR-Cas9 has proven to work efficiently in *U. maydis* in recent studies and thus, the foundation for first trial screens is fulfilled (48, 55). As outlined before, iPool-Seq can only be used for insertional mutagenesis libraries. Thus, to combine iPool-Seq with the CrispR-technology, gene editing would require next to gRNAs

complementary repair templates that contain a specific insertional sequence flanked by homologous repair flanks that would be introduced at each gRNA cutting site.

iPool-Seq paves the way for *in vivo* analyses of colonized host material

Insertional mutagenesis libraries of microbes enable two major possibilities for downstream analyses: Firstly, a saturated library covering all genes can be analyzed for attenuated strains in *in vitro* growth assays. Strains depleted from *in vitro* growth in rich-media possess most likely mutations in genes that are essential for the organism, e.g. recently analyzed in depth in yeast (88). Secondly, pathogens or symbiont mutant libraries can be screened for genes that are required for virulence or colonization during the interaction with the host, as shown in this study or for bacterial the symbiont *P. simiae* (78). For the analysis of insertion-cassette genome-junctions resulting in the identification of the mutated genomic locus a multitude of methods were developed in the past. Prior to the development of tools for mutant pools, negative depletion screens were conducted with single mutants (71, 89). Although these studies provided valuable data resources, the studies were not only laborious, but also error prone and expensive due to the extensive labor demand. Therefore, the development of methods for pooled mutant analysis would offer a great advance to elucidate gene functions. The first technique that enabled screens with pools of mutants was signature tagged mutagenesis (77). In this approach, mutants were generated with the Tn5 transposon tagged with a unique barcode in each mutant. The barcodes were used in the end for hybridization against an array and the mutant propagation success in the host was read out by radioactive labeling intensities. The approach was very innovative and provided a multitude of insights in bacterial virulence genes and several advancements of the technique were published in the following decade (90). However, with the advent of NGS other techniques superseded signature tagged mutagenesis rapidly. In 2009, several techniques using transposition libraries coupled with NGS were published (91-94). NGS propelled not only the complexity of mutagenesis pools that could be

analyzed, but also enabled the quantification of the results, in case protein-coding genes were covered by several insertions. Transposon insertion sequencing (Tn-Seq) became afterwards the most popular method and was used in a multitude of studies with bacteria (95-97). Tn-Seq makes use of the Mariner-Himar1-two component transposon system, that inserts the transposon cassette by chance in TA recognition sites (94). The library preparation of Tn-Seq for NGS begins with an elegant step based on the type II restriction enzyme *Mme*I that cuts 20 base pairs downstream of its recognition site. To this end, *Mme*I digestion results in 20bp long genomic overhangs at both flanking genome-transposon junctions which eventually will reveal the genomic integration site. Illumina NGS compatible overhangs are afterwards added by adapters and via PCR with a transposon specific primer. Therefore, the library preparation is straight forward and easy to master. However, the fragmentation of genomic DNA with a restriction enzyme has a drawback in comparison to random shearing methods: *Mme*I recognition sites close to genome-transposon junctions can hamper the sequencing results simply by removing the informative flanking sequences. Moreover, Tn-Seq can only yield quantitative and statistically analyzable data, if single genes are covered by several insertions. This requires high density of mutations along the genome, which were obtained in studies with small bacterial genomes but get more complicated in case of larger genomes, e.g. from eukaryotic microbes like filamentous fungi. In addition, Tn-Seq has not proven to be sensitive enough to identify integration sites efficiently from infected material, i.e., that the mutants need to be separated from the host after selection, for instance in an infection, before genomic DNA (gDNA) extraction and library preparation (78). Alternative library preparation methods, like HITS, fragmentize the gDNA by mechanical shearing with ultrasound (91). In comparison to the *Mme*I restriction digest of Tn-Seq, mechanical shearing is more laborious and requires end-repair, A-tailing and adapter ligation with multiple clean-up

steps in between. This can lead to considerable gDNA losses during the library preparation. Furthermore, the HITS library preparation protocol deviates also in another step (91): HITS uses biotinylated primers during the specific PCR that allow for affinity purification of PCR products obtained from transposon-genome junctions. This elegant purification technique can improve the signal-to-noise ratio during NGS and most likely improves sensitivity in comparison to Tn-Seq. However, the methods were never compared to each other on the same experimental conditions and thus, sensitivity improvements of HITS are not confirmed. Remarkably, HITS as well as Tn-Seq were only used on isolated bacterial pools, but a library preparation was never shown to be functional directly from complex infected or colonized tissue (78, 91).

In this study, the tool iPool-Seq was established with an explicit focus on high sensitivity and selectivity. iPool-Seq is designed to facilitate the isolation of integration-cassette genome-junctions directly from *in vivo* material. Thus, with iPool-Seq it should not be necessary to isolate microbial mutants from the host after infection and prior to NGS library preparation. To this end, iPool-Seq starts with the Tn5 transposon system for gDNA fragmentation instead of mechanical shearing. Next to mutant library generation, transposition is also a useful tool to randomly shear gDNA and insert adapters simultaneously. It is advantageous over mechanical shearing, because it minimizes gDNA losses during the library preparation by avoiding end-repair, A-tailing and adapter ligation. It is worth mentioning, that Tn5 transposition results in a much broader size range of gDNA fragments in comparison to mechanical shearing. However, the advantages of Tn5 transposition over mechanical shearing compensate for this minor disadvantage. Optionally, a size selection by agarose gel electrophoresis or with solid phase reversible immobilization (SPRI) magnetic beads could be applied to remove large and small fragments prior to PCR amplification. As recently described, functional hyperactive Tn5 transposase can be produced in large quantities in *E. coli*

and can be loaded with customized adapters, that are inserted at TA dinucleotide sites in the gDNA (98, 99). To compensate for the low ratio of fungal to plant gDNA, the Tn5 fragmentation protocol was successfully adapted for large gDNA quantities up to 1 µg. Subsequently, the iPool-Seq protocol continues with an exponential PCR step with 15 cycles, similar as in the HITS library preparation protocol but with 3 cycles less (91). The primer that is complementary to the unique sequence at the border of the integration cassette carries a Biotin-Teg-modification at its 5`-end that does not influence the DNA-polymerase. The second PCR primer will yield exponential amplification by annealing to the adapter that was inserted by the Tn5 transpose. As described in (98), the custom adapter is designed with a single stranded overhang, resulting in amplification only after elongation of the anti-sense strand along the specific biotinylated primer by the DNA-polymerase. This elegant design was adopted to iPool-Seq to ensure that annealing of the primer directed against the adapter only occurs at products that originate from genome-cassette junctions. This enables specific exponential amplification of flanking sequences of insertional cassettes and avoids unnecessary amplification of the residual adapter ligated fragments. Subsequently, PCR fragments are affinity purified with Streptavidin-coated beads, like the library protocol of HITS (91). However, iPool-Seq finishes eventually with a PCR with 15 cycles on streptavidin-affinity-purified fragments. This final PCR step is essential, because it removes the biotin-overhang, adds the Illumina sequencing compatible overhangs and generates fragments ready for NGS. Moreover, it adds a second multiplication step in case of low concentrations of the purified specific products. The reduced cycle number in comparison to HITS in the first PCR reduces the risk of PCR biases, that could occur due to the broad fragment size range resulting from Tn5 fragmentation. Still, iPool-Seq library preparation is slightly longer than other methods published in the past due to the second PCR amplification step (91-94).

Negative depletion screens aim to identify underrepresented mutants in a mutant pool. Consequently, NGS library protocols must conserve the ratios between single molecules that entered the analysis initially. This is especially difficult for low input sequences that must be isolated from host tissue and therefore, amplified and affinity purified. To ensure quantitative molecule ratios, iPool-Seq has a second novel feature in comparison to former published insertional mutagenesis protocols: It makes use of UMIs that are incorporated in every adapter in advance to PCR amplifications. UMIs can facilitate the analysis of NGS reads and help to remove biases that are most probably introduced during PCR amplification cycles (100, 101). To this end, PCR amplicons that share UMIs are grouped together to represent individual molecules derived from the gDNA. To improve UMI analyses eventually, the recently published TRUmiCount algorithm was implemented (102). TRUmiCount increases the identification of underrepresented amplicons and reduces the number of false UMIs that potentially emerge through sequencing errors or PCR artifacts. Thus, TRUmiCount improves the signal to noise ratio by removing low abundant UMIs that likely contain many false UMIs originating from late PCR cycles and correcting for marginally covered molecules. The UMIs in iPool-Seq adapters are 12 nucleotides long, following the design of Duplex sequencing, a method to detect DNA errors with high accuracy (103). Providing random nucleotide selection during primer synthesis, millions of distinct UMIs can be generated allowing for high input molecule quantities and sequencing depths. Samples prepared after the iPool-Seq protocol were sequenced on the Illumina MiSeq platform. As a site note, the iPool-Seq adapter design is fully compatible with any Illumina sequencing flow cell but requires the usage of a custom forward sequencing primer. In future, this design could be improved to allow sequencing with standard Illumina sequencing primer mixes to enable multiplexing with other samples, e.g. on the Illumina HiSeq platform that has a better

cost effectiveness than the smaller MiSeq platform, albeit the NGS read number received by the MiSeq platform was sufficient for the purposes of this study.

In conclusion, the above described advancements make iPool-Seq potentially more sensitive than Tn-Seq or HITS. This is underlined by the fact that this study yielded very high percentages of informative NGS reads directly from in vivo infected maize tissue. Especially, the implementation of gDNA fragmentation with Tn5 and UMI-count analysis are two novel procedures that have not been implemented in insertional mutagenesis library preparation tools in the past. Future benchmark tests of iPool-Seq against Tn-Seq and HITS need to be conducted to confirm superiority of iPool-Seq. Regardless from such benchmark tests, iPool-Seq provides innovative possibilities for researchers: Especially the analysis of insertional mutant pools directly from gDNA isolated from *in vivo* colonized or infected host tissue is enabled by iPool-Seq. Due to its high sensitivity it is also conceivable that iPool-Seq can facilitate the analysis of insertional mutant pools in even more complex mixtures of organisms, e.g. in vertebrate gut analyses.

A novel resource of 195 *U. maydis* effector insertional mutants

In this work, iPool-Seq was established with an insertional mutagenesis library of 195 *U. maydis* mutants. The technique is most likely advantageous over established protocols like Tn-Seq and enables screening of pooled mutants directly from *in vivo* infected maize material, as discussed extensively before. iPool-Seq does not only make fast screenings of large mutant pools possible, but it also relies on NGS reads and therefore brings about results with a fundamentally different basis to classical disease ratings that are routinely conducted in the *U. maydis* community. In contrast to disease ratings, iPool-Seq data will not provide any information about the symptoms that arise from the infection of a single mutant. Instead, the data supplies a computation of the individual mutant growth in comparison to reference mutants without altered growth, i.e. less reads than the references denote reduced growth and indicate reduced virulence or hypovirulence, whereas more reads denote enhanced growth and potentially hypervirulence. On the other hand, classical disease ratings are based on symptom observations. Thus, classical disease ratings give a qualitative read out, while iPool-Seq delivers a quantitative read out. Both outputs provide important information about the consequences of a mutation of a gene in respect to virulence. A reduced growth phenotype observed by iPool-Seq may result in reduced symptoms, as observed for three novel candidates that were tested with classical disease ratings in the study. Yet, it is also conceivable that mutants with a growth phenotype have no obvious symptom phenotype, because the mutation has no direct effect on gall formation but results in reduced gall incidence. Thus, iPool-Seq offers a new phenotyping tool for the fungal pathogen community that could decipher less severe mutant phenotypes that were potentially neglected in the past, because they lagged an obvious reduction of symptoms.

Upon infection of the mutant pool in maize, 28 mutants were reproducibly and significantly depleted from resulting NGS reads. Among the depleted 28 mutants are five well-characterized effectors with a confirmed virulence defect in classical disease ratings, reiniforcing that screening results are bona fide (29, 62, 104-106). In addition, three novel candidates were analyzed in classical disease ratings and displayed on top of the growth phenotype observed in the NGS data also strikingly decreased disease symptoms, further underlining the fact that iPool-Seq yields plausible results. Moreover, the iPool-Seq data revealed several reproducibly, significantly depleted mutants that had a much lower fold change over the reference mutant set than for instance the strong apathogenic mutants Pep1 and Stp1 (29, 62). It will be interesting to examine all these mutants in classical disease ratings and analyze the qualitative phenotype in future. Most likely, some of those mutants will lack a qualitative phenotype indicating that these effectors have no function in gall formation.

An enrichment of core effectors was analyzed next among mutants that displayed a significant depletion. It was hypothesized that core effectors might be of higher importance during virulence than unconserved effectors. Core effectors are defined by their high level of conservation in related fungal species. Recently the core effectome of *U. maydis* was restricted to pathogenic smut fungi whereas more distant or non-pathogenic fungi were excluded (14). However, it remains unclear if the classification of effectors in core and orphan, i.e. species specific, is indeed of importance and has an impact on the degree of virulence of an effector. In fact, iPool-Seq yielded not an enrichment of core effectors among depleted mutants, demonstrating that conservation of effector genes does not necessarily correlate with a higher importance on virulence. However, a genome-wide insertional screen needs to be conducted to elucidate the importance of core effectors globally.

Obviously, the majority of mutants was not significantly depleted after infection. This could be due to two reasons: Firstly, pooled infections have the disadvantage that mutants in close proximity during infection could functionally complement each other *in trans*. The phenomenon is more conceivable for effectors that have a rather systemic effect on the plant and whose function is not restricted to the local site of infection. Moreover, effectors that have a function during late stages could be more prone to *in trans* complementation, because the fungal biomass is already much higher at these stages. Possibly, a reduction of the inoculum density could reduce the risk of *in trans* complementation. However, *in trans* complementation is hard to prove and cannot be completely avoided in mutant pool infections. Secondly, most effectors probably have a function during the infection, but pathogens evolved functional redundancy to strengthen robustness of virulence. This theory seems logic, but is very hard to prove, because functional redundancy must not correlate with conservation on the sequence level. For instance, pathogenic effectors could act on several levels of the same signaling cascade in the host to be functional redundant and therefore their mutual deletion could have a negative epistatic effect on the pathogen. Functional redundancy of effectors is not a disadvantage of pooled infections *per se* but will also appear in qualitative classical disease ratings. Functional redundancy can only be elucidated efficiently with effector screens that build up on functional questions or treatments, but likely not by mutant phenotype screens. Once a functional redundant group of effector genes is spotted, CrispR-Cas9 targeted gene-editing can serve as a valuable tool to test for negative epistasis (14).

Surprisingly, hypervirulent mutants of *U. maydis* were not identified in the iPool-Seq data. Indeed, deletion of single effector genes has not revealed any significantly enhanced virulence phenotypes in *U. maydis* to date. It has been proposed that effectors of *U. maydis* could act as avirulence factors that are recognized by the plant

defense machinery (107). Possibly, the effects of single gene deletions are rather weak and only become significant upon multiple avirulence gene deletions. Classical disease ratings of effector cluster deletions of *U. maydis* have resulted in significantly enhanced virulence phenotypes in the past (26). In addition, iPool-Seq was used on maize accession Early Golden Bantam (EGB), which is highly susceptible to *U. maydis* infections and displays stronger symptoms than other maize accessions. In future, it would be interesting to test iPool-Seq on mutant pools infected in less susceptible maize accessions, like B73, to possibly identify hypervirulent mutants.

In summary, in this work a novel resource of 195 *U. maydis* deletion mutants was generated which is freely available for further analysis in the *Ustilago* community. Subsequently, the iPool-Seq NGS library protocol was developed and tested in a proof-of-principle analysis with this deletion mutant library, which resulted in 23 novel virulence mutants, that await further functional characterization in future.

iPool-Seq: Opportunities and future applications

iPool-Seq is a promising technique that still has potential for optimization: For example, it would be very intriguing to test iPool-Seq and a transposon insertion library of *U. maydis* infected in maize. To do so, three major challenges need to be solved in the order specified: 1) Transposition efficiencies in *U. maydis* need optimization (53). 2) After successful transposition, genes essential for growth and filamentation need to be identified, like recently shown in a study with yeast (88). 3) Pooled infections need optimization concerning maximal complexity of mutant library and number of maize plants that are required to identify all mutants without virulence defects. Once these three steps are established, iPool-Seq offers novel screening possibilities on a global view of insertional mutants: It can be used to assess the impact of conditional changes systematically on the mutant pool and the infection output, e.g. during abiotic stresses. It would also be intriguing, to harvest and analyze material from different stages of the infection as well as different tissues to gain insights on the spatial and developmental importance of single effectors. It has been recently shown, that some *U. maydis* effectors are required during late stages of the infection, e.g. for sporogenesis, whereas others are upregulated in the early stages, indicating for a role in immune suppression (46, 108). All above-mentioned applications of iPool-Seq examine the effects of single gene deletions. With insertional mutagenesis coupled with Cas9 it would be possible to engineer multiple insertional mutations per strain (86, 87). To this end, it would be interesting to delete effector paralogs or co-regulated effectors and examine pooled infections with iPool-Seq. iPool-Seq could shed new light on these exciting questions by adding another layer of information about the virulence contribution of single or even co-operative genes in concert.

Effector biology – functional characterization as an outstanding challenge

In a second publication, the latest insights in fungal effector biology were described and discussed. The main topics of the review comprise effector expression, translocation and, most importantly, effector function. The model in the review proposed three different effector function strategies that could mediate virulence of filamentous pathogens: Firstly, effectors may act on the filamentous pathogen itself and alter processes or structures to improve the infection success. For instance, an alteration of the fungal cell wall can help to avoid host recognition (31). Secondly, effectors can act as inhibitors, e.g. by deactivating host proteins through direct or indirect interactions. Thirdly, effectors can have an activating function on host processes to aid virulence, for instance by stimulating antagonistic pathways of the biotrophic defense. The latter two effector strategies can be involved in pathogenic defense against host counter-attacks and can allow for reprogramming of the host metabolism to redirect nutrient fluxes or shut down host defense responses. The functional analysis of effectors can be challenging. Depending on the accessibility and the molecular toolbox that is established in a pathogen, several aspects can be studied which are important for functional characterization of effectors to give first insights in their mode of action. Important questions are the contribution to virulence of a single effector, its host target identification and its localization *in planta*. As discussed before, most characterized effectors of *U. maydis* display a strong phenotype upon mutation. Their functional characterization was prioritized, because of their essential character during virulence. As indicated by the first results of iPool-Seq with *U. maydis* it becomes more and more obvious that the majority of effectors have weak or no obvious phenotypes upon deletion. Likely, functional conservation that is not based on sequence conservation has evolved during evolution. Thus, other effectors can

functionally compensate the loss of a certain effector upon its deletion. This is probably an advantageous strategy for plant pathogens in general, because the recognition or inactivation of an effector by the host defense machinery, *e.g.* by R genes, can be counteracted by other effectors that functionally converge on the same host mechanism (109).

It will be very interesting to see how future approaches and novel techniques, like CrispR-Cas9 deletion or iPool-Seq screens, will broaden our knowledge about virulence factor contribution on a genome wide as well as multi-deletion level in pathogens that are accessible to genome editing. Screens with random insertion libraries with these tools might also identify more effectors that were overlooked so far, because they lack a signal peptide and might be secreted unconventionally. Depending on the success of these tools, it is conceivable that virulence contribution can become a defining component for an effector in future, like transcriptional upregulation or secretion peptide prediction.

Although the beforementioned tools can give insights about phenotypes of effector mutants, further mechanistic work to decipher effector functions is necessary. One possibility to elucidate effector functions is the identification of potential proteinous host plant targets. This can be achieved by Yeast-two-hybrid screens of effectors against complementary DNA (cDNA) libraries (65, 105), or potentially vice-versa by plant proteins against effector libraries. In addition, co-immunoprecipitations (Co-IPs) of effector-tag fusions coupled with mass spectrometry analysis has proven to be an effective alternative, for instance by transient expression in *Nicotiana benthamiana* (110). The identification of the host target may have disparate outcomes: On the one hand, a substantial fraction of plant proteins contains functional annotations, mainly from orthologs originating from research with *Arabidopsis thaliana.* Functional annotations of plant targets might indicate the mechanism that an interacting

pathogenic effector could be involved in and potentially lead to downstream analyses, like enzymatic assays in the case of Pep1 (29). On the other hand, effectors may interact with host proteins without functional annotations, and thus, their functional characterization cannot be deduced from the host target. However, the functional analysis of such an effector can also provide a better understanding of the underlying mechanism of its plant target. Therefore, effector biology is a multifaceted research field that may generate novel insights in plant molecular biology by the characterization of host plant targets through the interacting effector.

Moreover, effector function can be further delineated by the analysis of its localization *in planta*. The localization to a sub-compartment can support our understanding of the effector function, for instance by differentiating plant target proteins due to their localization *in planta*. In the best case, this is pursued in the endogenous pathosystem, e.g. in maize with *U. maydis*. In this particular example maize is not easily accessible for genome editing tools. Thus, localization data are often derived from heterologous expression, like *A. thaliana* or *N. benthamiana,* which also delivered promising results in rust fungi (111). To circumvent this problem for *U. maydis*, it is possible to investigate the localization in a related pathosystems that harbors the ortholog of the effector of interest and that offers a more accessible host. An example is the pathosystem of the related smut fungus *U. bromivora*, which infects the genetically more accessible host *Brachypodium distachyon* (42). Thus, in future it is conceivable to provide localization data of conserved *U. maydis* effectors with fungal orthologs from *U. bromivora* transformed in *B. distachyon*.

The systematic analysis of effector functions by identification of their protein targets, *in planta* localization and contribution to virulence remains an outstanding challenge for future effector research. In the case of *U. maydis*, it may result in functional redundant effector groups and reveal major plant target hubs that several effectors might

converge on (112). Moreover, the research conducted with *U. maydis* might help to decipher conserved effector strategies that are found in various filamentous pathogens. The decoding of such crucial players in the arms race between plants and pathogens may eventually facilitate the development of sustainable and more resistant crop plants.

References

1. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444(7117):323-9.
2. Kwon C, Bednarek P, Schulze-Lefert P. Secretory pathways in plant immune responses. Plant Physiol. 2008;147(4):1575-83.
3. Boller T, He SY. Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. Science. 2009;324(5928):742-4.
4. Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, et al. Bacterial disease resistance in Arabidopsis through flagellin perception. Nature. 2004;428(6984):764-7.
5. Petutschnig EK, Jones AM, Serazetdinova L, Lipka U, Lipka V. The lysin motif receptor-like kinase (LysM-RLK) CERK1 is a major chitin-binding protein in Arabidopsis thaliana and subject to chitin-induced phosphorylation. J Biol Chem. 2010;285(37):28902-11.
6. Cao YR, Liang Y, Tanaka K, Nguyen CT, Jedrzejczak RP, Joachimiak A, et al. The kinase LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with related kinase CERK1. Elife. 2014;3.
7. Yamaguchi Y, Pearce G, Ryan CA. The cell surface leucine-rich repeat receptor for AtPep1, an endogenous peptide elicitor in Arabidopsis, is functional in transgenic tobacco cells. Proc Natl Acad Sci U S A. 2006;103(26):10104-9.
8. Zipfel C, Oldroyd GED. Plant signalling in symbiosis and immunity. Nature. 2017;543(7645):328-36.
9. Dievart A, Gilbert N, Droc G, Attard A, Gourgues M, Guiderdoni E, et al. Leucine-rich repeat receptor kinases are sporadically distributed in eukaryotic genomes. BMC Evol Biol. 2011;11:367.
10. Luna E, Pastor V, Robert J, Flors V, Mauch-Mani B, Ton J. Callose deposition: a multifaceted plant defense response. Mol Plant Microbe Interact. 2011;24(2):183-93.
11. Wu S, Shan L, He P. Microbial signature-triggered plant defense responses and early signaling mechanisms. Plant Sci. 2014;228:118-26.
12. Jones JDG, Vance RE, Dangl JL. Intracellular innate immune surveillance devices in plants and animals. Science. 2016;354(6316).
13. Chung EH, da Cunha L, Wu AJ, Gao Z, Cherkis K, Afzal AJ, et al. Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants. Cell Host Microbe. 2011;9(2):125-36.
14. Schuster M, Schweizer G, Kahmann R. Comparative analyses of secreted proteins in plant pathogenic smut fungi and related basidiomycetes. Fungal Genet Biol. 2017.
15. Yan SP, Dong XN. Perception of the plant immune signal salicylic acid. Curr Opin Plant Biol. 2014;20:64-8.
16. Fu ZQ, Dong XN. Systemic Acquired Resistance: Turning Local Infection into Global Defense. Annual Review of Plant Biology, Vol 64. 2013;64:839-63.
17. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al. Emerging fungal threats to animal, plant and ecosystem health. Nature. 2012;484(7393):186-94.
18. Oliver RP, Ipcho SVS. Arabidopsis pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens. Mol Plant Pathol. 2004;5(4):347-52.
19. Coll NS, Epple P, Dangl JL. Programmed cell death in the plant immune system. Cell Death Differ. 2011;18(8):1247-56.
20. Garnica DP, Nemri A, Upadhyaya NM, Rathjen JP, Dodds PN. The Ins and Outs of Rust Haustoria. Plos Pathog. 2014;10(9).
21. Liu TL, Song TQ, Zhang X, Yuan HB, Su LM, Li WL, et al. Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. Nat Commun. 2014;5.
22. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, et al. A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature. 2007;450(7166):115-+.
23. Wawra S, Trusch F, Matena A, Apostolakis K, Linne U, Zhukov I, et al. The RxLR Motif of the Host Targeting Effector AVR3a of Phytophthora infestans Is Cleaved before Secretion. Plant Cell. 2017;29(6):1184-95.

24. Galan JE, Lara-Tejero M, Marlovits TC, Wagner S. Bacterial Type III Secretion Systems: Specialized Nanomachines for Protein Delivery into Target Cells. Annu Rev Microbiol. 2014;68:415-38.
25. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus Magnaporthe grisea. Nature. 2005;434(7036):980-6.
26. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature. 2006;444(7115):97-101.
27. Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, et al. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature. 2009;461(7262):393-8.
28. Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A. 2011;108(22):9166-71.
29. Doehlemann G, van der Linde K, Amann D, Schwammbach D, Hof A, Mohanty A, et al. Pep1, a Secreted Effector Protein of Ustilago maydis, Is Required for Successful Invasion of Plant Cells. Plos Pathog. 2009;5(2).
30. Bos JIB, Armstrong MR, Gilroy EM, Boevink PC, Hein I, Taylor RM, et al. Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. P Natl Acad Sci USA. 2010;107(21):9909-14.
31. de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, et al. Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science. 2010;329(5994):953-5.
32. Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, et al. Effector-Mediated Suppression of Chitin-Triggered Immunity by Magnaporthe oryzae Is Necessary for Rice Blast Disease. Plant Cell. 2012;24(1):322-35.
33. Goehre V, Spallek T, Haeweker H, Mersmann S, Mentzel T, Boller T, et al. Plant Pattern-Recognition Receptor FLS2 Is Directed for Degradation by the Bacterial Ubiquitin Ligase AvrPtoB. Curr Biol. 2008;18(23):1824-32.
34. Chen H, Chen J, Li M, Chang M, Xu K, Shang Z, et al. A Bacterial Type III Effector Targets the Master Regulator of Salicylic Acid Signaling, NPR1, to Subvert Plant Immunity. Cell Host Microbe. 2017;22(6):777-88 e7.
35. Ustun S, Sheikh A, Gimenez-Ibanez S, Jones A, Ntoukakis V, Bornke F. The Proteasome Acts as a Hub for Plant Immunity and Is Targeted by Pseudomonas Type III Effectors. Plant Physiol. 2016;172(3):1941-58.
36. Song J, Win J, Tian M, Schornack S, Kaschani F, Ilyas M, et al. Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. Proc Natl Acad Sci U S A. 2009;106(5):1654-9.
37. Bozkurt TO, Schornack S, Win J, Shindo T, Ilyas M, Oliva R, et al. Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. P Natl Acad Sci USA. 2011;108(51):20832-7.
38. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the plant immune system from dissection to deployment. Science. 2013;341(6147):746-51.
39. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, et al. The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. Science. 2005;307(5713):1321-4.
40. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol. 2012;13(4):414-30.
41. Schirawski J, Mannhaupt G, Munch K, Brefort T, Schipper K, Doehlemann G, et al. Pathogenicity Determinants in Smut Fungi Revealed by Genome Comparison. Science. 2010;330(6010):1546-8.
42. Rabe F, Bosch J, Stirnberg A, Guse T, Bauer L, Seitner D, et al. A complete toolset for the study of Ustilago bromivora and Brachypodium sp. as a fungal-temperate grass pathosystem. Elife. 2016;5.
43. Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Munsterkotter M, Guldener U, et al. A Tale of Genome Compartmentalization: The Evolution of Virulence Clusters in Smut Fungi. Genome Biol Evol. 2016;8(3):681-704.

44. Hartmann HA, Kahmann R, Bolker M. The pheromone response factor coordinates filamentous growth and pathogenicity in Ustilago maydis. EMBO J. 1996;15(7):1632-41.
45. Lanver D, Berndt P, Tollot M, Naik V, Vranes M, Warmann T, et al. Plant surface cues prime Ustilago maydis for biotrophic development. Plos Pathog. 2014;10(7):e1004272.
46. Lanver D, Muller AN, Happel P, Schweizer G, Haas FB, Franitza M, et al. The Biotrophic Development of Ustilago maydis Studied by RNA-Seq Analysis. Plant Cell. 2018;30(2):300-23.
47. Doehlemann G, Wahl R, Vranes M, de Vries RP, Kamper J, Kahmann R. Establishment of compatibility in the Ustilago maydis/maize pathosystem. J Plant Physiol. 2008;165(1):29-40.
48. Schuster M, Schweizer G, Reissmann S, Kahmann R. Genome editing in Ustilago maydis using the CRISPR-Cas system. Fungal Genet Biol. 2016;89:3-9.
49. Bölker M, Genin S, Lehmler C, Kahmann R. Genetic regulation of mating and dimorphism in Ustilago maydis. Canadian Journal of Botany. 1995;73(S1):320-5.
50. Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, et al. Metabolic priming by a secreted fungal effector. Nature. 2011;478(7369):395-+.
51. Brachmann A, Konig J, Julius C, Feldbrugge M. A reverse genetic approach for generating gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004;272(2):216-26.
52. Khrunyk Y, Munch K, Schipper K, Lupas AN, Kahmann R. The use of FLP-mediated recombination for the functional analysis of an effector gene family in the biotrophic smut fungus Ustilago maydis. New Phytol. 2010;187(4):957-68.
53. Ladendorf O, Brachmann A, Kamper J. Heterologous transposition in Ustilago maydis. Mol Genet Genomics. 2003;269(3):395-405.
54. Ji LH, Jiang ZD, Liu YB, Koh CMJ, Zhang LH. A Simplified and efficient method for transformation and gene tagging of Ustilago maydis using frozen cells. Fungal Genet Biol. 2010;47(4):279-87.
55. Schuster M, Schweizer G, Kahmann R. Comparative analyses of secreted proteins in plant pathogenic smut fungi and related basidiomycetes. Fungal Genet Biol. 2018;112:21-30.
56. Kinal H, Tao JS, Bruenn JA. An expression vector for the phytopathogenic fungus, Ustilago maydis. Gene. 1991;98(1):129-34.
57. Spellig T, Bottin A, Kahmann R. Green fluorescent protein (GFP) as a new vital marker in the phytopathogenic fungus Ustilago maydis. Mol Gen Genet. 1996;252(5):503-9.
58. Redkar A, Hoser R, Schilling L, Zechmann B, Krzymowska M, Walbot V, et al. A Secreted Effector Protein of Ustilago maydis Guides Maize Leaf Cells to Form Tumors. Plant Cell. 2015;27(4):1332-51.
59. Lo Presti L, Kahmann R. How filamentous plant pathogen effectors are translocated to host cells. Curr Opin Plant Biol. 2017;38:19-24.
60. Zarnack K, Maurer S, Kaffarnik F, Ladendorf O, Brachmann A, Kamper J, et al. Tetracycline-regulated gene expression in the pathogen Ustilago maydis. Fungal Genet Biol. 2006;43(11):727-38.
61. Banks GR, Shelton PA, Kanuga N, Holden DW, Spanos A. The Ustilago maydis nar1 gene encoding nitrate reductase activity: sequence and transcriptional regulation. Gene. 1993;131(1):69-78.
62. Schipper K, Brefort T, Doehlemann G, Djamei A, Muench K, Kahmann R. The secreted protein Stp1 is cruical for establishment of the biotrophic interaction of the smut fungus Ustilago maydis with its host plant maize. Eur J Cell Biol. 2008;87:29-.
63. Hemetsberger C, Herrberger C, Zechmann B, Hillmer M, Doehlemann G. The Ustilago maydis Effector Pep1 Suppresses Plant Immunity by Inhibition of Host Peroxidase Activity. Plos Pathog. 2012;8(5).
64. Brefort T, Tanaka S, Neidig N, Doehlemann G, Vincon V, Kahmann R. Characterization of the Largest Effector Gene Cluster of Ustilago maydis. Plos Pathog. 2014;10(7).
65. Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, et al. A secreted Ustilago maydis effector promotes virulence by targeting anthocyanin biosynthesis in maize. Elife. 2014;3.
66. Wang J, Holden DW, Leong SA. Gene transfer system for the phytopathogenic fungus Ustilago maydis. Proc Natl Acad Sci U S A. 1988;85(3):865-9.

67. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science. 1999;285(5429):901-6.
68. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002;418(6896):387-91.
69. Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell. 2008;135(1):174-88.
70. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785-6.
71. Jeon J, Park SY, Chi MH, Choi J, Park J, Rho HS, et al. Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. Nat Genet. 2007;39(4):561-5.
72. Michielse CB, van Wijk R, Reijnen L, Cornelissen BJC, Rep M. Insight into the molecular requirements for pathogenicity of Fusarium oxysporum f. sp lycopersici through large-scale insertional mutagenesis. Genome Biol. 2009;10(1).
73. Craig NL. Target site selection in transposition. Annu Rev Biochem. 1997;66:437-74.
74. Kang Y, Durfee T, Glasner JD, Qiu Y, Frisch D, Winterberg KM, et al. Systematic mutagenesis of the Escherichia coli genome. J Bacteriol. 2004;186(15):4921-30.
75. Salama NR, Shepherd B, Falkow S. Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. J Bacteriol. 2004;186(23):7926-35.
76. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. 2009;6(10):767-72.
77. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. Simultaneous identification of bacterial virulence genes by negative selection. Science. 1995;269(5222):400-3.
78. Cole BJ, Feltcher ME, Waters RJ, Wetmore KM, Mucyn TS, Ryan EM, et al. Genome-wide identification of bacterial plant colonization genes. Plos Biol. 2017;15(9):e2002860.
79. Rozhkov NV, Hammell M, Hannon GJ. Multiple roles for Piwi in silencing Drosophila transposons. Gene Dev. 2013;27(4):400-12.
80. Amsterdam A, Burgess S, Golling G, Chen W, Sun Z, Townsend K, et al. A large-scale insertional mutagenesis screen in zebrafish. Genes Dev. 1999;13(20):2713-24.
81. Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, et al. Multiple independent defective suppressor-mutator transposon insertions in Arabidopsis: a tool for functional genomics. Plant Cell. 1999;11(10):1841-52.
82. Yusa K, Zhou L, Li MA, Bradley A, Craig NL. A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A. 2011;108(4):1531-6.
83. Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell. 2005;122(3):473-83.
84. Shi JW, Wang E, Milazzo JP, Wang ZH, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol. 2015;33(6):661-+.
85. Munoz DM, Cassiani PJ, Li L, Billy E, Korn JM, Jones MD, et al. CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. Cancer Discov. 2016;6(8):900-13.
86. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu RM, et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat Biotechnol. 2017;35(1):48-55.
87. Guo X, Chavez A, Tung A, Chan Y, Kaas C, Yin Y, et al. High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. Nat Biotechnol. 2018.
88. Michel AH, Hatakeyama R, Kimmig P, Arter M, Peter M, Matos J, et al. Functional mapping of yeast genomes by saturated transposition. Elife. 2017;6.
89. Liu OW, Chun CD, Chow ED, Chen CB, Madhani HD, Noble SM. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell. 2008;135(1):174-88.
90. Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. Nat Rev Genet. 2006;7(12):929-39.

91. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. P Natl Acad Sci USA. 2009;106(38):16422-7.
92. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. Cell Host & Microbe. 2009;6(3):279-89.
93. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 2009;19(12):2308-16.
94. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. 2009;6(10):767-U21.
95. Gallagher LA, Shendure J, Manoil C. Genome-Scale Identification of Resistance Functions in Pseudomonas aeruginosa Using Tn-seq. Mbio. 2011;2(1).
96. Fu Y, Waldor MK, Mekalanos JJ. Tn-Seq analysis of Vibrio cholerae intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. Cell Host Microbe. 2013;14(6):652-63.
97. Duong DA, Jensen RV, Stevens AM. Discovery of Pantoea stewartii ssp. stewartii genes important for survival in corn xylem through a Tn-Seq analysis. Mol Plant Pathol. 2018.
98. Stern DL. Tagmentation-Based Mapping (TagMap) of Mobile DNA Genomic Insertion Sites. bioRxiv. 2017.
99. Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014;24(12):2033-40.
100. Hug H, Schuler R. Measurement of the number of molecules of a single mRNA species in a complex mRNA preparation. J Theor Biol. 2003;221(4):615-24.
101. Kivioja T, Vaharautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2012;9(1):72-U183.
102. Pflug FG, von Haeseler A. TRUmiCount: Correctly counting absolute numbers of molecules using unique molecular identifiers. Bioinformatics. 2018.
103. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc. 2014;9(11):2586-606.
104. Eichhorn H, Lessing F, Winterberg B, Schirawski J, Kamper J, Muller P, et al. A ferroxidation/permeation iron uptake system is required for virulence in Ustilago maydis. Plant Cell. 2006;18(11):3332-45.
105. Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the Ustilago maydis-Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. Plos Pathog. 2013;9(2).
106. Stirnberg A, Djamei A. Characterization of ApB73, a virulence factor important for colonization of Zea mays by the smut Ustilago maydis. Mol Plant Pathol. 2016;17(9):1467-79.
107. Schirawski J, Mannhaupt G, Munch K, Brefort T, Schipper K, Doehlemann G, et al. Pathogenicity determinants in smut fungi revealed by genome comparison. Science. 2010;330(6010):1546-8.
108. Tollot M, Assmann D, Becker C, Altmuller J, Dutheil JY, Wegner CE, et al. The WOPR Protein Ros1 Is a Master Regulator of Sporogenesis and Late Effector Gene Expression in the Maize Pathogen Ustilago maydis. Plos Pathog. 2016;12(6):e1005697.
109. Birch PR, Boevink PC, Gilroy EM, Hein I, Pritchard L, Whisson SC. Oomycete RXLR effectors: delivery, functional redundancy and durable disease resistance. Curr Opin Plant Biol. 2008;11(4):373-9.
110. Petre B, Saunders DG, Sklenar J, Lorrain C, Win J, Duplessis S, et al. Candidate Effector Proteins of the Rust Pathogen Melampsora larici-populina Target Diverse Plant Cell Compartments. Mol Plant Microbe Interact. 2015;28(6):689-700.
111. Lorrain C, Petre B, Duplessis S. Show me the way: rust effector targets in heterologous plant systems. Curr Opin Microbiol. 2018;46:19-25.

112. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal Effectors and Plant Susceptibility. Annual Review of Plant Biology, Vol 66. 2015;66:513-45.

Abstract

Biotrophic, filamentous plant pathogens are a substantial threat to plant yield and cause immense annual losses in agriculture. Their virulence is promoted by small, secreted molecules, commonly known as effectors. Genomic analysis revealed that filamentous pathogens have large arsenals of effectors, mostly lacking known domains that could indicate their function. Effectors that have a strong impact on virulence likely display a reduced virulence phenotype upon genomic deletion. To test this efficiently and in high-throughput, developed insertion Pool-Sequencing (iPool-Seq) was developed, a tool that allows for the analysis of insertional mutant pool infections by extensive parallel Illumina sequencing. iPool-Seq is extremely sensitive, enabling genomic DNA extractions coupled with efficient enrichment of genome-insertion site junctions directly from in vivo infected host material. iPool-Seq was tested on an insertional mutant library of the maize-pathogen *Ustilago maydis,* yielding highly reproducible and quantitative results. Among the identified virulence factors, iPool-Seq confirmed five well characterized mutants and identified 23 unknown virulence factors. The iPool-Seq protocol is compatible with any existing insertional mutant library and is a promising tool that is not restricted to effector biology but has the potential to elucidate essential genes of various microbes. Moreover, a functional categorization was proposed, wherein effectors can act as self-modifiers, or either as suppressors, or activators of plant host targets. In future, it is of outstanding interest to decipher effector functions on a genome wide level with high precision. These functional analyses comprise effector host target identification, in planta subcellular localization and contribution of effectors to virulence. This knowledge might foster engineering of more resistant crop varieties in future.

Zusammenfassung

Biotrophe und filamentöse Pathogene stellen eine substanzielle Gefahr für pflanzliche Erträge dar und verursachen der Landwirtschaft jährlich beträchtliche Verluste. Die Virulenz der Pathogene beruht auf kleinen, sekretierten Molekülen, weitestgehend bekannt als Effektoren. In genomischen Analysen konnte gezeigt werden, dass filamentöse Pathogene große Arsenale an Effektoren haben, von denen die meisten keine bekannten Proteindomänen besitzen, die Aufschluss auf die Funktion der Effektoren geben könnten. Deletionsmutanten von Effektoren, die einen starken Einfluss auf die Virulenz des Pathogens haben, weisen meist auch einen verminderten Virulenzphänotyp auf. Im Rahmen dieser Dissertation wurde die Technik namens „insertion Pool-Sequencing“ (iPool-Seq) entwickelt, die es ermöglicht, effizient und im Hochdurchsatz Effektoren zu identifizieren, die einen Beitrag zur Virulenz leisten. Die Technik ermöglicht die Analyse von Infektionen mit mehreren Pathogenmutanten gleichzeitig und basiert letztendlich auf Hochdurchsatzsequenzierungen von Mutanten Genomen. iPool-Seq ist besonders effizient, wodurch erfolgreich die in der genomischen DNS enthaltenen Sequenzflanken der Insertionskassetten direkt aus dem infizierten Pflanzenmaterial angereichert werden können. In dieser Arbeit wurde iPool-Seq an einer Mutantensammlung des Maispathogens *Ustilago maydis* getestet und reproduzierbare sowie quantitative auswertbare Sequenzen erhalten. Unter den 28 identifizierten Mutanten, die eine signifikant verminderte Virulenz aufzeigten, konnten 5 bekannte Mutanten verifiziert werden. Das Verfahren ist mit jeder Sammlung von Insertionsmutanten kompatibel und könnte zum Beispiel auch für die Entschlüsselung von essentiellen Genen von Mikroben verwendet werden. Des Weiteren wurde im Rahmen dieser Doktorarbeit eine Kategorisierung von möglichen Funktionsweisen von Effektoren vorgeschlagen: Effektoren können auf das Pathogen selbst Auswirkungen haben, oder eine unterdrückende oder aktivierende Funktion

innerhalb der Pflanze einnehmen. Die systematische und eindeutige Entschlüsselung von Funktionsweisen der Effektoren in filamentösen Pathogenen ist eines der Hauptherausforderungen im Feld der Pflanzen-Mikroben Interaktionen. Diese funktionellen Analysen geben Aufschluss über potentielle pflanzliche Interaktionspartner, die subzelluläre Lokalisation in der Pflanze und den Beitrag der Virulenz eines Effektors. Bei erfolgreicher und umfassender Analyse der Funktionen von Effektoren kann dieses Wissen zu einer Weiterentwicklung von resistenten Nutzpflanzen eingesetzt werden.

Appendix: Supplements of publication 1

S1 Table. *U. maydis* genes targeted for insertional mutagenesis

	U. maydis Gene ID	Secretion prediction[1]	Genomic organization[2]	Prediction of *in planta* Localization wo. signal peptide sequence[3]	Apoplastic prediction wo. signal peptide sequence[4]	Core effector[5]
1	UMAG_00054	Yes	-	Nucleus	Non-apoplastic	Yes
2	UMAG_00081	Yes	-	-	Apoplastic	No
3	UMAG_00105	Yes	-	-	Non-apoplastic	No
4	UMAG_00159	Yes	-	-	Non-apoplastic	Yes
5	UMAG_00187	Yes	-	-	Non-apoplastic	No
6	UMAG_00558	Yes	-	-	Apoplastic	No
7	UMAG_00781	Yes	-	-	Non-apoplastic	No
8	UMAG_00792	Yes	-	Chloroplast	Apoplastic	No
9	UMAG_00793	Yes	-	Chloroplast	Non-apoplastic	No
10	UMAG_00795	Yes	-	Chloroplast	Non-apoplastic	No
11	UMAG_00823	Yes	-	-	Apoplastic	No
12	UMAG_00885	Yes	-	-	Apoplastic	No
13	UMAG_01018	Yes	-	-	Non-apoplastic	No
14	UMAG_01082	Yes	-	Chloroplast	Apoplastic	No
15	UMAG_01130	Yes	-	Chloroplast, Mitochondria	Non-apoplastic	No
16	UMAG_01235	Yes	Cluster 2A	-	Non-apoplastic	Yes
17	UMAG_01236	Yes	Cluster 2A	Chloroplast	Non-apoplastic	Yes
18	UMAG_01237	Yes	Cluster 2A	Nucleus	Non-apoplastic	Yes
19	UMAG_01238	Yes	Cluster 2A	-	Non-apoplastic	Yes
20	UMAG_01239	Yes	Cluster 2A	-	Non-apoplastic	Yes
21	UMAG_01240	Yes	Cluster 2A	-	Non-apoplastic	Yes
22	UMAG_01289	Yes	-	Chloroplast	Non-apoplastic	No
23	UMAG_01297	Yes	Cluster 2B	-	Apoplastic	No
24	UMAG_01300	Yes	Cluster 2B	Chloroplast	Apoplastic	No
25	UMAG_01301	Yes	Cluster 2B	-	Apoplastic	No
26	UMAG_01302	Yes	Cluster 2B	-	Apoplastic	No
27	UMAG_01375	Yes	-	Nucleus	Apoplastic	No
28	UMAG_01553	Yes	-	Chloroplast, Nucleus	Non-apoplastic	No
29	UMAG_01689	Yes	-	-	Non-apoplastic	No
30	UMAG_01690	Yes	-	-	Non-apoplastic	No
31	UMAG_01779	Yes	-	-	Apoplastic	Yes
32	UMAG_01820	Yes	-	-	Non-apoplastic	Yes
33	UMAG_01854	Yes	-	Nucleus	Non-apoplastic	Yes
34	UMAG_01858	Yes	-	Nucleus	Non-apoplastic	No
35	UMAG_01940	Yes	-	-	Non-apoplastic	Yes
36	UMAG_01977	Yes	-	-	Non-apoplastic	No
37	UMAG_01987	Yes	-	-	Apoplastic	Yes
38	UMAG_01997	Yes	-	Chloroplast	Apoplastic	No
39	UMAG_02006	Yes	-	-	Apoplastic	No

40	UMAG_02011	Yes	-	Chloroplast, Nucleus	Non-apoplastic	Yes
41	UMAG_02119	Yes	-	-	Non-apoplastic	No
42	UMAG_02135	Yes	-	Nucleus	Non-apoplastic	Yes
43	UMAG_02137	Yes	-	Nucleus	Non-apoplastic	Yes
44	UMAG_02138	Yes	-	-	Non-apoplastic	Yes
45	UMAG_02139	Yes	-	Nucleus	Non-apoplastic	Yes
46	UMAG_02141	Yes	-	Nucleus	Non-apoplastic	Yes
47	UMAG_02192	Yes	Cluster 5A	Chloroplast, Mitochondria, Nucleus	Non-apoplastic	No
48	UMAG_02193	Yes	Cluster 5A	Nucleus	Non-apoplastic	No
49	UMAG_02229	Yes	-	Nucleus	Non-apoplastic	No
50	UMAG_02239	Yes	-	-	Non-apoplastic	No
51	UMAG_02243	Yes	-	Mitochondria, Nucleus	Non-apoplastic	No
52	UMAG_02294	Yes	-	-	Non-apoplastic	No
53	UMAG_02295	Yes	-	-	Non-apoplastic	No
54	UMAG_02297	Yes	-	-	Non-apoplastic	No
55	UMAG_02298	Yes	-	Chloroplast	Non-apoplastic	No
56	UMAG_02299	Yes	-	Mitochondria	Non-apoplastic	Yes
57	UMAG_02430	Yes	-	Chloroplast	Apoplastic	No
58	UMAG_02466	Yes	-	Nucleus	Non-apoplastic	No
59	UMAG_02473	Yes	Cluster 5B	-	Non-apoplastic	No
60	UMAG_02474	Yes	Cluster 5B	-	Non-apoplastic	No
61	UMAG_02475	Yes	Cluster 5B	Mitochondria	Apoplastic	No
62	UMAG_02533	Yes	Cluster 6A	Chloroplast, Nucleus	Non-apoplastic	No
63	UMAG_02535	Yes	Cluster 6A	Nucleus	Non-apoplastic	No
64	UMAG_02537	Yes	Cluster 6A	Nucleus	Non-apoplastic	No
65	UMAG_02538	Yes	Cluster 6A	Mitochondria	Non-apoplastic	No
66	UMAG_02560	Yes	-	-	Non-apoplastic	No
67	UMAG_02611	Yes	-	-	Apoplastic	Yes
68	UMAG_02813	Yes	-	Nucleus	Non-apoplastic	Yes
69	UMAG_02826	Yes	-	-	Non-apoplastic	Yes
70	UMAG_02851	Yes	-	Nucleus	Non-apoplastic	No
71	UMAG_02852	Yes	-	-	Non-apoplastic	No
72	UMAG_02853	Yes	-	-	Non-apoplastic	No
73	UMAG_03023	Yes	-	Chloroplast	Apoplastic	Yes
74	UMAG_03046	Yes	-	-	Non-apoplastic	No
75	UMAG_03065	Yes	-	Mitochondria	Non-apoplastic	No
76	UMAG_03105	Yes	-	-	Non-apoplastic	No
77	UMAG_03112	Yes	-	-	Apoplastic	No
78	UMAG_03138	Yes	-	Chloroplast, Nucleus	Non-apoplastic	No
79	UMAG_03201	Yes	Cluster 8A	-	Non-apoplastic	No
80	UMAG_03202	Yes	Cluster 8A	-	Non-apoplastic	No
81	UMAG_03223	Yes	-	-	Non-apoplastic	No
82	UMAG_03313	Yes	-	Nucleus	Non-apoplastic	Yes
83	UMAG_03382	Yes	-	-	Non-apoplastic	Yes
84	UMAG_03397	Yes	-	Chloroplast	Non-apoplastic	Yes

85	UMAG_03564	Yes	-	Chloroplast	Apoplastic	No
86	UMAG_03586	Yes	-	Nucleus	Non-apoplastic	No
87	UMAG_03615	Yes	Cluster 9A	-	Non-apoplastic	Yes
88	UMAG_03650	Yes	-	-	Non-apoplastic	No
89	UMAG_03689	Yes	-	-	Non-apoplastic	No
90	UMAG_03744	Yes	Cluster 10A	Nucleus	Non-apoplastic	No
91	UMAG_03745	Yes	Cluster 10A	-	Non-apoplastic	No
92	UMAG_03747	Yes	Cluster 10A	-	Non-apoplastic	No
93	UMAG_03748	Yes	Cluster 10A	-	Non-apoplastic	No
94	UMAG_03750	Yes	Cluster 10A	Chloroplast	Non-apoplastic	Yes
95	UMAG_03751	Yes	Cluster 10A	-	Non-apoplastic	Yes
96	UMAG_03753	Yes	Cluster 10A	-	Non-apoplastic	Yes
97	UMAG_03880	Yes	-	-	Apoplastic	No
98	UMAG_04033	Yes	-	Chloroplast, Mitochondria	Non-apoplastic	No
99	UMAG_04038	Yes	-	Nucleus	Non-apoplastic	No
100	UMAG_04039	Yes	-	-	Non-apoplastic	No
101	UMAG_04057	Yes	-	-	Non-apoplastic	No
102	UMAG_04084	Yes	-	-	Apoplastic	Yes
103	UMAG_04096	Yes	-	Chloroplast	Non-apoplastic	No
104	UMAG_04104	Yes	-	-	Non-apoplastic	No
105	UMAG_04111	Yes	-	Mitochondria, Nucleus	Apoplastic	Yes
106	UMAG_04114	Yes	-	-	Non-apoplastic	No
107	UMAG_04145	Yes	-	-	Apoplastic	No
108	UMAG_04185	Yes	-	-	Apoplastic	No
109	UMAG_04189	No	-	-	-	-
110	UMAG_04282	Yes	-	-	Apoplastic	Yes
111	UMAG_04400	Yes	-	Nucleus	Apoplastic	Yes
112	UMAG_04696	Yes	-	-	Non-apoplastic	No
113	UMAG_04815	Yes	-	-	Non-apoplastic	No
114	UMAG_04893	Yes	-	-	Apoplastic	No
115	UMAG_05222	Yes	-	-	Non-apoplastic	No
116	UMAG_05227	Yes	-	-	Non-apoplastic	Yes
117	UMAG_05294	Yes	Cluster 19A	-	Non-apoplastic	No
118	UMAG_05299	No	-	-	-	-
119	UMAG_05300	Yes	Cluster 19A	Chloroplast, Mitochondria, Nucleus	Non-apoplastic	Yes
120	UMAG_05301	Yes	Cluster 19A	Mitochondria	Non-apoplastic	Yes
121	UMAG_05302	Yes	Cluster 19A	-	Non-apoplastic	No
122	UMAG_05308	Yes	Cluster 19A	Chloroplast, Mitochondria	Non-apoplastic	No
123	UMAG_05310	Yes	Cluster 19A	Mitochondria	Non-apoplastic	No
124	UMAG_05319	Yes	Cluster 19A	-	Non-apoplastic	No
125	UMAG_05439	Yes	-	Nucleus	Apoplastic	No
126	UMAG_05548	Yes	-	-	Non-apoplastic	Yes
127	UMAG_05562	Yes	-	-	Non-apoplastic	Yes

128	UMAG_05641	Yes	-	-	Non-apoplastic	No
129	UMAG_05731	Yes	-	-	Non-apoplastic	Yes
130	UMAG_05733	Yes	-	Mitochondria	Apoplastic	No
131	UMAG_05780	Yes	-	Chloroplast	Non-apoplastic	Yes
132	UMAG_05781	Yes	-	Nucleus	Non-apoplastic	Yes
133	UMAG_05819	Yes	-	Chloroplast, Mitochondria	Non-apoplastic	No
134	UMAG_05861	Yes	-	-	Non-apoplastic	Yes
135	UMAG_05926	Yes	-	Chloroplast, Mitochondria, Nucleus	Non-apoplastic	Yes
136	UMAG_05927	Yes	-	-	Non-apoplastic	Yes
137	UMAG_05931	Yes	-	-	Non-apoplastic	No
138	UMAG_05953	Yes	-	-	Apoplastic	No
139	UMAG_05988	Yes	-	-	Non-apoplastic	Yes
140	UMAG_06064	Yes	-	Chloroplast	Apoplastic	No
141	UMAG_06113	Yes	-	Nucleus	Non-apoplastic	No
142	UMAG_06146	Yes	-	-	Apoplastic	No
143	UMAG_06158	Yes	-	-	Non-apoplastic	Yes
144	UMAG_06178	Yes	-	Mitochondria	Non-apoplastic	No
145	UMAG_06179	Yes	-	-	Non-apoplastic	No
146	UMAG_06222	Yes	Cluster 22A	-	Non-apoplastic	No
147	UMAG_06223	Yes	Cluster 22A	-	Non-apoplastic	No
148	UMAG_06428	Yes	-	Chloroplast	Apoplastic	Yes
149	UMAG_06440	Yes	-	-	Apoplastic	Yes
150	UMAG_10024	Yes	-	-	Non-apoplastic	No
151	UMAG_10030	Yes	-	Nucleus	Non-apoplastic	Yes
152	UMAG_10067	Yes	-	-	Apoplastic	Yes
153	UMAG_10076	Yes	-	-	Non-apoplastic	No
154	UMAG_10403	Yes	Cluster 8A	-	Non-apoplastic	No
155	UMAG_10553	Yes	Cluster 19A	Chloroplast, Mitochondria	Non-apoplastic	No
156	UMAG_10555	Yes	Cluster 19A	Chloroplast, Mitochondria	Non-apoplastic	No
157	UMAG_10557	Yes	Cluster 19A	-	Non-apoplastic	No
158	UMAG_10640	Yes	-	-	Apoplastic	Yes
159	UMAG_10742	Yes	-	-	Non-apoplastic	No
160	UMAG_10756	Yes	-	-	Apoplastic	No
161	UMAG_10811	Yes	-	-	Non-apoplastic	Yes
162	UMAG_10816	Yes	-	-	Apoplastic	No
163	UMAG_10881	Yes	-	-	Apoplastic	Yes
164	UMAG_10972	Yes	-	-	Non-apoplastic	No
165	UMAG_10975	Yes	-	Nucleus	Non-apoplastic	Yes
166	UMAG_11002	Yes	-	-	Apoplastic	No
167	UMAG_11062	Yes	-	Nucleus	Non-apoplastic	No
168	UMAG_11094	Yes	-	Nucleus	Non-apoplastic	No
169	UMAG_11193	Yes	-	Mitochondria	Non-apoplastic	No

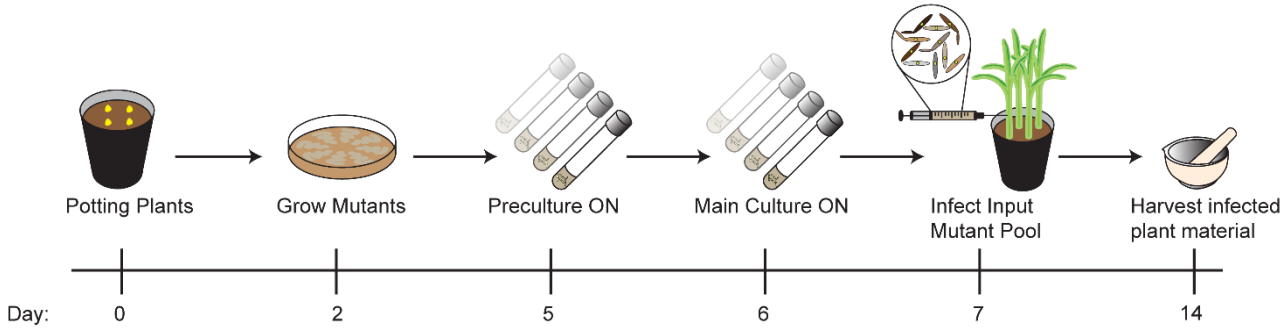
170	UMAG_11250	Yes	-	-	Non-apoplastic	No
171	UMAG_11305	Yes	-	Mitochondria	Non-apoplastic	No
172	UMAG_11362	Yes	-	-	Apoplastic	No
173	UMAG_11377	Yes	-	Nucleus	Non-apoplastic	No
174	UMAG_11402	No	-	-	-	-
175	UMAG_11403	Yes	-	-	Non-apoplastic	Yes
176	UMAG_11415	Yes	Cluster 6A	-	Non-apoplastic	No
177	UMAG_11416	Yes	Cluster 6A	-	Non-apoplastic	No
178	UMAG_11417	Yes	Cluster 6A	-	Non-apoplastic	No
179	UMAG_11444	Yes	-	-	Apoplastic	No
180	UMAG_11464	Yes	-	-	Non-apoplastic	Yes
181	UMAG_11586	Yes	-	Chloroplast, Mitochondria, Nucleus	Non-apoplastic	No
182	UMAG_11639	Yes	-	-	Non-apoplastic	No
183	UMAG_11931	Yes	-	-	Apoplastic	No
184	UMAG_11940	Yes	-	-	Non-apoplastic	No
185	UMAG_12045	Yes	-	Nucleus	Non-apoplastic	Yes
186	UMAG_12127	Yes	-	Chloroplast	Non-apoplastic	No
187	UMAG_12197	Yes	-	-	Apoplastic	No
188	UMAG_12216	Yes	-	Nucleus	Non-apoplastic	No
189	UMAG_12226	Yes	-	Nucleus	Non-apoplastic	No
190	UMAG_12233	Yes	-	-	Non-apoplastic	Yes
191	UMAG_12281	Yes	-	-	Non-apoplastic	No
192	UMAG_12302	Yes	-	Mitochondria	Non-apoplastic	No
193	UMAG_12313	Yes	-	-	Apoplastic	Yes
194	UMAG_12330	Yes	-	-	Non-apoplastic	Yes
195	UMAG_15020	Yes	-	Chloroplast	Apoplastic	No
		98.5% are predicted to be secreted	23.1% are found in effector clusters	40.5 % are predicted to localize to a compartment *in planta*	73.8% are predicted to be nonapoplastic	31.7% are putative core effectors
		1.5 % are predicted to be not secreted	76.9% are found outside of effector clusters	59.5% are predicted to no compartment *in planta*	26.2% are predicted to be apoplastic	68.3% are not core effectors

References

1. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785-6. doi: 10.1038/nmeth.1701. PubMed PMID: WOS:000295358000004.

2. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature. 2006;444(7115):97-101. doi: 10.1038/nature05248. PubMed PMID: WOS:000241701500053.
3. Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep-Uk. 2017;7. doi: ARTN 44598 10.1038/srep44598. PubMed PMID: WOS:000396543100002.
4. Sperschneider J, Dodds PN, Singh KB, Taylor JM. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. New Phytol. 2017. doi: 10.1111/nph.14946. PubMed PMID: 29243824.
5. Schuster M, Schweizer G, Kahmann R. Comparative analyses of secreted proteins in plant pathogenic smut fungi and related basidiomycetes. Fungal Genet Biol. 2017. Doi: 10.1016/j.fgb.2016.12.003. PubMed PMID: 28089076.

S1 Figure. Workflow of Pooled infection of maize



S1 Fig. Workflow of pooled infection of maize. For each replicate of the *U. maydis* mutant collection at least 100 maize plants of the accession EGB were potted. Mutants were grown on selective plates for 2-3 days. From plates precultures were inoculated and grown overnight (ON). The precultures were used for inoculation of the main cultures to avoid dead material in the infection pool. All main cultures were pooled with equal amounts that were adjusted to the same optical density and infected in 7-day old maize seedlings with a syringe. Infected areas of the 2nd and 3rd leaf of each plant were harvested 7 days after the infection. All three biological replicates of the mutant collection were processed in 14 days.

S1 Supporting methods. iPool-Seq analysis pipeline description

S1 SUPPORTING METHODS:
IPOOL-SEQ ANALYSIS PIPELINE DESCRIPTION

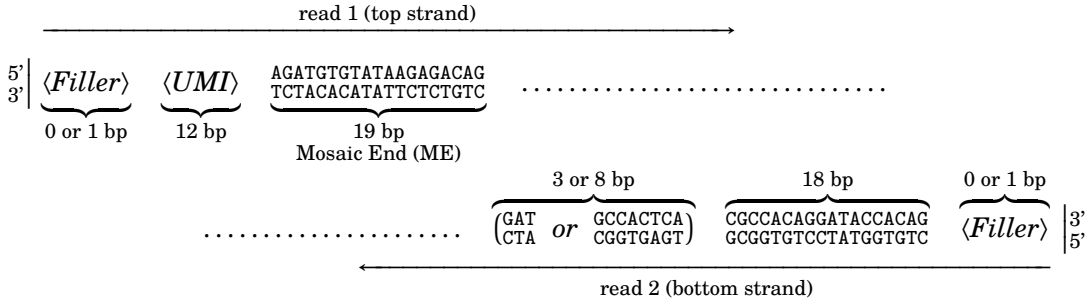
1. Read validation & mapping

Demultiplexing. The 12 libraries (one input and one output library for each of three replicates in experiments A & B) were sequenced (paired-end, 75 bp reads from both fragment ends) on two Illumina MiSeq flowcells (one per experiment). The runs were demultiplexed using *deML* [1] (pre-release, commit 80a491), and separate BAM files for each library are available in the *european nucleotide archive* (ENA), accession PRJEB23309.

Read-through removal. Read-throughs into the sequencing adapter on the other end (for short fragments) were removed using *Trimmomatic* [2] (version 0.33) in `PE` (paired-end) mode using commands `ILLUMINACLIP:adpaters.fa:2:24:15:1:true` and `MINLEN:40`, with `adapters.fa` containing the following two sequencing adapters:

```
>PrefixPE/1
CACGACGCTCTTCCGATCT
>PrefixPE/2
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
```

UMI extraction & technical sequence removal (`trim.tag.py`)**.** From the construction of the 195 (single-gene) insertional mutants of *U. maydis* and the library preparation protocol used, we expected the double-stranded fragments subjected to sequencing to have the following layout (both strands shown):



The part denoted "…" is a genomic *U. maydis* sequence, more specifically a sequence from the 3' or 5' flank of one the 195 studied genes. Our custom script `trim.tag.py` matched the sequenced read pairs against this expected pattern, allowing up to 4 mismatches (not counting `N`s) within the fixed part of each mate. Our script then stored the UMIs as part of the read names, and stripped all technical sequences (i.e. everything except the "…" part) from the reads. If the two mates of a pair overlapped (i.e. for fragments shorter than $2 \cdot 75 = 150$ bp), a technical sequence from one mate possibly appeared reverse-completed on the other mate as well. We detected this by checking whether a gap-less ends-free alignment of the two reads had an identity $\geq 90\%$, and then used the alignment to locate and remove the corresponding part of the complementary mate as well.

Assignment to mutants (`assign_to_features.py`)**.** To assign the reads to genes (and hence to insertional mutants), we mapped the paired-end reads (after UMI extraction and technical sequence removal) to the *U. maydis* genome, *GeneBank* accession GCF_000328475.2 [3], using *NextGenMap* [4] (version v0.4.13) with parameters `--end-to-end` `--pair-score-cutoff 0.5` `--sensitivity 0.3` `--kmer 13` `--kmer-skip 0`.

Proper read pairs (read pairs where one mate maps in the forward direction, the other in the reverse direction, and the mates point "towards" one another) were assigned to a particular gene if either mate's first sequenced base mapped to within ±10 bp of one of the genes flanks, and the rest of that read continued "away" from the gene.

Improper read pairs (non-proper read pairs where nevertheless both mates were mapped) were ignored.

Singleton reads (i.e. reads whose mate could not be mapped) were assigned to a particular gene if their first sequenced base mapped to within a 1000 bp window on either side of the gene and they continued "towards" the gene.

Read pairs assigned to no or multiple genes were ignored.

2. UMI analysis & abundance estimation

Correcting UMIs for sequencing errors (`umicounts.tag.py`)**.** To count *U. maydis* insertional mutant genomes (i.e. cells), we counted the number of (sufficiently distinct, to protect against sequencing errors) combinations of UMI and mapping position within the reads mapping to a particular flank (3' or 5') of a particular gene. For the sake of brevity, *UMI* in the following denotes a *combination* of a particular 12 bp molecular barcode (so far called UMI) and the two mate's mapping positions.

To merge similar UMIs (which likely stem from the same cell), we used a variation of the algorithm of Smith *et al.* [5]. We started with the raw list of unique UMIs. We then marked an UMI p as *mergeable* into UMI q if the molecular barcodes disagreed at most at a single position, the mapping positions by no more than ±3 bases, and p was found in fewer reads than q. The UMIs not marked as mergeable were then assumed to be error-free. The read counts of UMIs that were mergeable (directly or indirectly) with a single error-free UMI were added to the error-free UMI's read count. UMIs marked (directly or indirectly) as mergeable with multiple error-free candidates were discarded as being ambiguous.

This produced, for both flanks of every gene, a separate list of assumedly error-free UMIs and per-UMI read counts.

Correcting for artifacts and lost UMIs to estimate abundance (`counts2results.R`)**.** We then further processed the per-flank UMIs using the algorithm of Pflug & von Haeseler [6], i.e. we removed all UMIs with a read count below a manually set read-count threshold ($T = 1$, except $T = 5$ for Experiment B R1 & R2 Output, and $T = 9$ for Experiment B R3 Output), and then estimated (for both flanks of every gene separately) the percentage ℓ of UMIs lost during sequencing and data filtering.

This yielded, separately for both flanks of every gene, a number n^{obs} of observed UMIs (after all filtering steps) and a loss estimate ℓ. Given these two, a (flank-specific) estimate of true mutant abundance is $n^{\text{obs}}/(1-\ell)$.

3. Statistical Analysis

Modelling growth of neutral mutants (`model.R`)**.** Given an insertional mutant m's *true* (unknown) abundances A_m^{in} and A_m^{out} in a particular pair of input and output libraries, and given the respective losses (i.e. fraction of unobserved or filtered UMIs) ℓ_{mf}^{in} and ℓ_{mf}^{out} for flank f (3' or 5'), we assumed that the observed

number of per-flank UMIs (after filtering) is Poisson distributed with mean $A_m^{\text{in}} \cdot (1-\ell_{mf}^{\text{in}})$ respectively $A_m^{\text{out}} \cdot (1-\ell_{mf}^{\text{out}})$. For the *sum* N_m^{in} respectively N_m^{out} of UMIs on the two flanks (5' and 3') of the mutamt m in the input respectively output library, it follows that

$$(1)\quad N_m^{\text{in}} \,\big|\, A_m^{\text{in}} \sim \text{Poisson}\big(A_m^{\text{in}} \cdot (1-\bar{\ell}_m^{\text{in}})\big), \qquad N_m^{\text{out}} \,\big|\, A_m^{\text{out}} \sim \text{Poisson}\big(A_m^{\text{out}} \cdot (1-\bar{\ell}_m^{\text{out}})\big)$$
$$\text{where}\quad \bar{\ell}_m^{\text{in}} = \ell_{m,5'}^{\text{in}} + \ell_{m,3'}^{\text{in}}, \qquad \bar{\ell}_m^{\text{out}} = \ell_{m,5'}^{\text{out}} + \ell_{m,3'}^{\text{out}}.$$

We then further assumed that for neutral mutants the *expected* true input and output abundances are proportional (with the same factor λ for all neutral mutants in a particular pair of input and output libraries), but that the output abundances have additional dispersion d due to random fluctuations of mutant growth, i.e. that

$$(2)\qquad \mathbb{E}A_m^{\text{out}} = \lambda \cdot \mathbb{E}A_m^{\text{in}}, \qquad \mathbb{V}A_m^{\text{out}} = \lambda^2 \cdot \mathbb{V}A_m^{\text{in}} + d \cdot \big(\mathbb{E}A_m^{\text{out}}\big)^2.$$

To find the *null distribution* (i.e. assuming mutant m is neutral) for the output UMI count N_m^{out} given observed input count n_m^{in}, we computed the posterior $A_m^{\text{in}} \mid N_m^{\text{in}}$ (using degenerate prior Gamma(0,0)), added dispersion d to get $A_m^{\text{out}} \mid N_m^{\text{in}}$, and combined with $N_m^{\text{out}} \,\big|\, A_m^{\text{out}}$. The resulting *negative binomial* distribution depends on two mutant-independent parameters, proportionality factor λ and dispersion d,

$$(3)\qquad N_m^{\text{out}} \,\big|\, n_m^{\text{in}} \sim \text{NegBin}\left(\mu_m := \lambda \cdot n_m^{\text{in}} \cdot \tfrac{1-\bar{\ell}_m^{\text{out}}}{1-\bar{\ell}_m^{\text{in}}},\ r_m := \frac{n_m^{\text{in}}}{1+d\cdot n_m^{\text{in}}}\right).$$

Computing p-values, q-values and effect sizes (`r4896.Rmd`, `r5157.Rmd`). For each of the 6 pairs of input and output libraries, we estimated λ and d by maximizing the likelihood of the negative binomial model (3) over a reference set of neutral mutants (see below for how those were selected). Given λ and d, we then computed (one-sided) p-values p_m^{low} (sig. of depletion in output) and p_m^{high} (sig. enrichment in output), for each mutant m detected in both output and input, as

$$(4)\qquad p_m^{\text{low}} = \mathbb{P}\big(N_m^{\text{out}} \le n_m^{\text{out}}\big), \quad p_m^{\text{high}} = \mathbb{P}\big(N_m^{\text{out}} \ge n_m^{\text{out}}\big) \quad \text{if } n_m^{\text{in}}, n_m^{\text{out}} \ge 1.$$

To control the *false discovery rate* (FDR), we applied the Benjamini-Hochberg (BH) procedure [7] (separately) to the collection of low and high p-values computed for a particular pair of input and output libraries, and set the FDR target to 10%.

To quantify the effect size, we also computed the $\log_2$ fold change (lfc_m) between each mutant m's observed output UMI count and the expected value for neutral mutants,

$$(5)\qquad \text{lfc}_m = \log_2 \frac{n_m^{\text{out}} \cdot (1-\bar{\ell}_m^{\text{in}})}{\lambda \cdot n_m^{\text{in}} \cdot (1-\bar{\ell}_m^{\text{out}})}.$$

Selecting the neutral reference set. We started with a candidate list of 13 insertional mutants described as neutral in the literature (UMAG_01297, UMAG_01300, UMAG_01302, UMAG_02192, UMAG_02193, UMAG_03046, UMAG_03201, UMAG_03202, UMAG_03615, UMAG_06222, UMAG_10403, UMAG_10553, UMAG_12313), estimated λ and d for all 6 input-output pairs, and computed these mutants' $\log_2$ fold changes. Suspecting that not all of these mutants are truly neutral, we looked for outliers (defined as for boxplots in R, values more than 1.5 IQR larger/smaller than the 75%/25% quantile) amongst these $\log_2$ fold changes and discarded them. We repeated this procedure for the remaining 8 candidates (UMAG_01302, UMAG_02192, UMAG_02193, UMAG_03046, UMAG_03202, UMAG_03615, UMAG_10403, UMAG_10553), and found 3 additional outliers. The remaining 5 candidate mutants (UMAG_01302, UMAG_02193, UMAG_03202, UMAG_10403, UMAG_10553) were then used as the final neutral reference set, and all p-values, q-values and $\log_2$ fold changes were re-computed based on this set.

Sensitivity of a genome-wide screen. To estimate the sensitivity of a genome-wide screen, we simulated experiments containing $m = 20{,}000$ distinct mutants using the statistical model from equation (1), but assuming a negative binomial distribution for N_m^{out} to account for the additional dispersion d of the output abundances (see also equation 2). We assumed the input abundances to be identical for all mutants (i.e $A_1^{\text{in}} = \ldots = A_{20.000}^{\text{in}} = A^{\text{in}}$), the output abundances of k mutants to show a virulence phenotype and hence to be reduced $2^{-\rho}$-fold (i.e. $A_1^{\text{out}} = \ldots = A_k^{\text{out}} = A^{\text{in}} \cdot 2^{\rho}$), and the other $m - k$ mutants to be neutral ($A_{k+1}^{\text{out}} = \ldots = A_{20{,}000}^{\text{out}} = A^{\text{in}}$). Based on $\approx 14\%$ of mutants in our screen showing a reproducible phenotype, and supplemental table 5 of Lanver *et al.* [8] showing $\approx 22\%$ of genes to be upregulated during infection, we set $k = 20{,}000 \cdot 0.14 \cdot 0.22 = 600$ (i.e. $\approx 3\%$ of mutants have a virulence phenotype). We set the additional dispersion d to the highest value observed in our 6 experiments (0.0126), and simulated 100 experiments for each input abundance $A^{\text{in}} = 1, 2, \ldots, 100$, once with $\log_2$ fold change of $\rho = -1.53$ (corresponding to the "Reduced" group in figure 4a) and once with $\rho = -2.75$ (corresponding to the "Lost virulence" group). For each simulated experiment we computed q-values as described above (see *Computing p-values, q-values and effect sizes*), determined the percentage of significant mutants within the ones with a virulence phenotype, and averaged these percentages over the 100 experiments to compute the efficiencies shown in figure S3.

4. Running the pipeline

Required software in addition to cited. *GNU Bash* (4.2.53). *GNU Make* (4.0). *Picard* (1.141). *samtools* (1.3.1). *gzip* (1.6). *python* (2.7.5). Python libraries: *recordtype* (1.1), *distance* (0.1.3), *regex* (2016.4.15), *pysam* (0.12.0.1), *bcbio-gff* (0.6.2), *biopython* (1.66). *R* (3.2.1). R libraries: *data.table* (1.10.4), *parallel* (3.2.1), *rmarkdown* (1.8). R Bioconductor Libraries: *rtracklayer* (1.30.4). Other R libraries: *gwpcR*[a] (0.9.9).

Running "abundance estimation" (incl. prerequisite steps). The pipeline (see S1 Software *iPool-Seq Analysis Pipeline*) uses separate subdirectories under `data/` for each library, e.g. `data/r4896.in1` for the input library of replicate 1 of experiment A. These directories contains various file controlling the pipeline (`tom.cfg`, `ngm.cfg`, `ref.fasta`, `features.gff`, `ngm.results.cfg`). To repeat our analyses, download the BAM files belonging to 12 libraries from `ftp://ftp.sra.ebi.ac.uk/vol1/ERA112/ERA1125781/bam/`, and store the file named `r`*<experiment_id>*`/`*<library>*`.bam` as `data/r`*<experiment_id>*`.`*<library>*`/raw.bam`.

The pipeline produces for each library two R data files as output, `ngm.results.rda` and `ngm.stats.rda`. For each subdirectory of `data/` run:

```
make data/<subdir>/ngm.results.rda data/<subdir>/ngm.stats.rda
```

Running "Statistical Analysis". The pipeline contains two R notebooks, `r4896.Rmd` (experiment A) and `r5157.Rmd` (experiment B). In R, run them with:

```
library(rmarkdown)
render("<experiment_id>.Rmd", output_format="pdf_document")
```

This produces a PDF report (`r`*<experiment_id>*`.pdf`) and table (`r`*<experiment_id>*`.abundance.csv`) listing for each mutant the raw and loss corrected input and output abundances, p- and q-values for significant depletion and enrichment, and the $\log_2$ fold change. It also produces two tables summarizing the significantly depleted (`r`*<experiment_id>*`.low.csv`) respectively enriched (`r`*<experiment_id>*`.high.csv`) mutants, and a R data file (`r`*<experiment_id>*`.model.rda`) containing the parameters of the null distributions.

[a] `http://github.com/Cibiv/gwpcR`, see also Pflug & von Haeseler [6]

References

1. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics. 2015;31(5):770–772. doi:10.1093/bioinformatics/btu719.
2. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
3. Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature. 2006;444(7115):97–101. doi:10.1038/nature05248.
4. Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics. 2013;29(21):2790–2791. doi:10.1093/bioinformatics/btt468.
5. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017;27(3):491–499. doi:10.1101/gr.209601.116.
6. Pflug FG, von Haeseler A. TRUmiCount: Correctly counting absolute numbers of molecules using unique molecular identifiers; 2017. Preprint. Available from: `http://www.biorxiv.org/content/early/2017/11/13/217778`. Cited 13 November 2017. doi:10.1101/217778.
7. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289–300. doi:10.2307/2346101.
8. Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitza M, et al. The biotrophic development of Ustilago maydis studied by RNAseq analysis. Plant Cell. 2018;30(2):300–323. doi:10.1105/tpc.17.00764.

S1 Data. q-Values of *U. maydis* mutant strains

S1 Data. q-Values of *U. maydis* mutant strains

Experiment A

No	Gene ID	R1,log2fc	R1,qval	R1	R2,log2fc	R2,qval	R2	R3,log2fc	R3,qval	R3	Meanof log2fc	No. Significant	No significant or Zero
1	UMAG_00054	-0,206903	0,66931184	-	-1,07947503	0,0027139	!	0,363377836	1	-	-0,307666642	1	1
2	UMAG_00081	0,204544	1	-	-0,5292777	0,0007196	!	0,909605112	1	-	0,194957146	1	1
3	UMAG_00105	-1,77101	6,4642E-17	!	-1,86017911	4,984E-31	!	-1,59125914	5,263E-20	!	-1,740816078	3	3
4	UMAG_00159	-0,298131	0,61549867	-	-0,17210555	0,0068762	!	0,193223801	1	-	-0,092337557	1	1
5	UMAG_00187	-0,409309	0,02585885	!	-0,33989908	0,016946	!	1,144453155	1	-	0,131748522	2	2
6	UMAG_00558	0,150821	1	-	-3,635576	1,866E-51	!	0,72708932	1	-	-0,919221883	1	1
7	UMAG_00781	0,1776085	1	-	-0,11265072	0,3265497	-	-0,35269935	0,0250248	!	-0,095913855	1	1
8	UMAG_00792	-1,173289	6,2466E-09	!	-0,78417451	0,007379	!	0,091994348	1	-	-0,621822915	2	2
9	UMAG_00793	-0,283929	0,10076145	-	0,000989159	0,7290453	-	0,103620461	1	-	-0,059773134	0	0
10	UMAG_00795	-0,530777	0,11123326	-	0,410132518	1	-	0,615215862	1	-	0,164857148	0	0
11	UMAG_00823	0,0961623	0,95980203	-	-2,37774836	8,827E-36	!	-1,08442585	1,014E-10	!	-1,122003975	2	2
12	UMAG_00885	0,1814594	1	-	-0,79374122	0,0364668	!	1,201630377	1	-	0,196449522	1	1
13	UMAG_01018	-0,333047	0,17181603	-	-0,18653494	0,3704658	-	0,462744471	1	-	-0,018945911	0	0
14	UMAG_01082	-0,554145	0,00814935	!	-0,27825711	0,0708695	!	0,677962925	1	-	-0,05147958	2	2
15	UMAG_01130	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0
16	UMAG_01235	-0,130177	0,4460056	-	0,070122593	0,9805174	-	0,131042731	1	-	0,023662769	0	0
17	UMAG_01236	-0,559917	0,00475213	!	-0,28937489	0,0637696	!	-0,47677767	0,0045322	!	-0,442023049	3	3
18	UMAG_01237	-1,278675	2,3363E-10	!	0,116110363	1	-	0,125076787	1	-	-0,345829285	1	1
19	UMAG_01238	-0,638524	0,00105633	!	-0,03304627	0,6308748	-	-0,1658937	0,3962545	-	-0,279154601	1	1
20	UMAG_01239	-0,277071	0,18063639	-	-0,32577868	0,0286444	!	0,178965755	1	-	-0,141294788	1	1
21	UMAG_01240	-0,145455	0,50414337	-	0,001498508	0,7604142	-	-2,37577244	3,753E-37	!	-0,839909781	1	1
22	UMAG_01289	0,3486777	1	-	-1,84580611	1,103E-23	!	-0,31389821	0,0463839	!	-0,603675552	2	2
23	UMAG_01297	0,2565057	1	-	-1,36896018	1,775E-15	!	-0,03338768	1	-	-0,381947401	1	1
24	UMAG_01300	-1,40366	2,6868E-12	!	-0,27291311	0,0443098	!	0,033292239	1	-	-0,547760444	2	2
25	UMAG_01301	-0,829403	2,4034E-05	!	0,248248759	1	-	-0,19237114	0,3049033	-	-0,25784189	1	1
26	UMAG_01302	-0,185466	0,32399024	-	-0,06932944	0,3623971	-	-0,10090499	0,8064095	-	-0,118566925	0	0
27	UMAG_01375	-2,776285	1,1987E-33	!	-1,29375444	3,387E-17	!	-2,46301962	4,043E-43	!	-2,177686197	3	3
28	UMAG_01553	-1,081372	0,23692345	-	0,158725569	1	-	-3,99402365	7,938E-44	!	-1,638889976	1	1
29	UMAG_01689	-1,502188	4,6738E-13	!	-2,54621062	1,866E-51	!	-2,18357554	2,367E-32	!	-2,07732477	3	3
30	UMAG_01690	-0,528831	0,51073545	-	-0,05570737	0,5743052	-	0,541074865	1	-	-0,014487776	0	0
31	UMAG_01779	-0,754391	0,23829217	-	-1,41713296	8,115E-17	!	0,271825777	1	-	-0,633232568	1	1
32	UMAG_01820	0,1430131	1	-	-2,99692388	1,349E-39	!	1,092244506	1	-	-0,587222083	1	1
33	UMAG_01854	-0,068153	0,61549867	-	-0,02471302	0,7252528	-	0,464819872	1	-	0,123984656	0	0
34	UMAG_01858	-0,025174	0,71251382	-	0,355448197	1	-	0,870512189	1	-	0,400262017	0	0
35	UMAG_01940	-0,669651	0,00040392	!	-0,17149962	0,0883662	!	-0,35403561	0,023751	!	-0,398395481	3	3
36	UMAG_01977	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0
37	UMAG_01987	-0,283737	0,09352972	!	-3,80016454	6,173E-62	!	-1,26714916	5,072E-15	!	-1,783683406	3	3
38	UMAG_01997	-0,565578	0,00411616	!	-0,0105536	0,7603772	-	0,45255288	1	-	-0,0411928	1	1
39	UMAG_02006	0,2835474	1	-	0,264279463	1	-	0,291779693	1	-	0,279868861	0	0
40	UMAG_02011	-2,075561	3,5115E-22	!	-2,03158881	1,852E-35	!	-2,53035327	1,976E-40	!	-2,212501118	3	3
41	UMAG_02119	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0
42	UMAG_02135	0,1713792	1	-	-0,1057134	0,3701394	-	0,252421857	1	-	0,106029211	0	0
43	UMAG_02137	-1,206903	0,35778276	-	-0,12010491	0,4317692	-	-0,60866132	0,0005144	!	-0,645222986	1	1
44	UMAG_02138	-0,132902	0,62354012	-	-0,86738554	0,0002402	!	-0,95557921	2,568E-09	!	-0,65195563	2	2
45	UMAG_02139	1,0732052	1	-	0,114492724	1	-	2,744443	1	-	1,310713639	0	0
46	UMAG_02141	-4,105722	3,9894E-42	!	-2,03626228	5,404E-19	!	0,263907013	1	-	-1,959358932	2	2
47	UMAG_02192	0,5419349	1	-	0,55338216	1	-	-0,05206314	1	-	0,347751317	0	0
48	UMAG_02193	0,0816059	0,92560348	-	-0,00253478	0,7588107	-	-0,03780093	0,9940587	-	0,013756736	0	0
49	UMAG_02229	-0,514468	0,00668901	!	-0,3216834	0,0193527	!	0,004623754	1	-	-0,277175904	2	2
50	UMAG_02239	-0,404067	0,02585885	!	-0,11983243	0,2982654	-	0,012173133	1	-	-0,170575593	1	1
51	UMAG_02243	0,0840894	0,92560348	-	0,003084053	0,7603772	-	0,290702377	1	-	0,125958601	0	0
52	UMAG_02294	0,2582079	1	-	-0,18756467	0,152089	-	0,39714782	1	-	0,155930339	0	0
53	UMAG_02295	-0,32107	0,08961862	!	-0,06680537	0,501189	-	0,285278435	1	-	-0,034198939	1	1
54	UMAG_02297	-0,187479	0,10836539	-	-0,85698212	1,303E-08	!	-0,22891482	0,2851758	-	-0,424458652	1	1
55	UMAG_02298	-0,554504	0,00441441	!	-0,09910157	0,3888783	-	0,067187557	1	-	-0,195472665	1	1
56	UMAG_02299	-0,523749	0,00749423	!	-0,42961049	0,1639695	-	3,11973202	1	-	0,722124307	1	1
57	UMAG_02430	-0,158151	0,32399024	-	0,09395237	1	-	0,474529129	1	-	0,136776753	0	0
58	UMAG_02466	0,3265429	1	-	-0,10474867	0,2045014	-	0,147070483	1	-	0,1229549	0	0
59	UMAG_02473	-0,941859	0,15644317	-	-1,75122375	2,208E-16	!	0,552274847	1	-	-0,713602755	1	1
60	UMAG_02474	-0,680315	0,00099	!	0,137938884	1	-	0,205458116	1	-	-0,112306007	1	1
61	UMAG_02475	-2,790559	2,5805E-34	!	-2,73093199	2,189E-51	!	-2,59010001	6,015E-42	!	-2,703863693	3	3
62	UMAG_02533	-0,247545	0,40291182	-	-0,43634191	0,0052497	!	0,38569014	1	-	-0,099398826	1	1
63	UMAG_02535	-1,477071	7,8148E-14	!	-1,99642299	6,204E-38	!	2,244804832	1	-	-0,409562982	2	2
64	UMAG_02537	-0,038707	0,67895717	-	-0,27597809	0,061808	!	-0,31077816	0,0209747	!	-0,20848784	2	2
65	UMAG_02538	-0,217976	0,2757205	-	0,269152837	1	-	0,83666933	1	-	0,295948787	0	0
66	UMAG_02560	-5,865205	2,5273E-51	!	-1,35528957	5,159E-21	!	-4,47402353	8,449E-67	!	-3,898172665	3	3
67	UMAG_02611	-0,013512	0,71251382	-	0,638789827	1	-	-0,41527796	0,0171301	!	0,070000104	1	1
68	UMAG_02813	-0,009355	0,71352066	-	0,010295871	0,7604142	-	-0,57824983	0,0001978	!	-0,192436224	1	1
69	UMAG_02826	0,1461474	1	-	-0,98970704	9,044E-11	!	0,950477348	1	-	0,035639235	1	1
70	UMAG_02851	-0,212919	0,24212209	-	-0,54573334	6,434E-05	!	0,139489433	1	-	-0,206387613	1	1
71	UMAG_02852	-0,74456	0,08720047	!	-0,48896697	0,0239521	!	-0,03550907	1	-	-0,42301185	2	2
72	UMAG_02853	-2,234261	4,0345E-05	!	-1,04487416	6,275E-12	!	0,927351286	1	-	-0,783928061	2	2
73	UMAG_03023	-0,240341	0,18063639	-	-0,26472768	0,0636657	!	0,636583092	1	-	0,043838063	1	1
74	UMAG_03046	0,0471966	0,8136933	-	0,098557336	0,8629571	-	0,122223927	1	-	0,089325943	0	0
75	UMAG_03065	-0,070748	0,53537671	-	-0,09263198	0,3556632	-	0,123287365	1	-	-0,013364278	0	0
76	UMAG_03105	-0,784019	1,6281E-05	!	-2,24305994	1,541E-40	!	-1,21133256	3,035E-12	!	-1,412803823	3	3
77	UMAG_03112	0,0111974	0,76645173	-	-0,28117266	0,0116431	!	1,159902615	1	-	0,296642445	1	1
78	UMAG_03138	0,0065775	0,76346681	-	-0,21154171	0,1421778	-	-0,39867398	0,0211402	!	-0,201212722	1	1
79	UMAG_03201	-0,188301	0,27979724	-	-2,09873464	3,425E-42	!	-1,45655158	9,626E-20	!	-1,247862357	2	2
80	UMAG_03202	-0,217719	0,18063639	-	-0,14621006	0,2540204	-	0,184535355	1	-	-0,059797845	0	0
81	UMAG_03223	-1,012033	1,6092E-07	!	-0,2097868	0,146362	-	0,144931021	1	-	-0,358962843	1	1
82	UMAG_03313	-3,404338	2,34E-30	!	-0,12492473	0,3701394	-	0,407197248	1	-	-1,040688603	1	1
83	UMAG_03382	-0,016363	0,88034042	-	0,142909771	1	-	0,13081028	1	-	0,085785642	0	0
84	UMAG_03397	-0,007298	0,71251382	-	-0,03694275	0,8008645	-	0,818556889	1	-	0,258105489	0	0
85	UMAG_03564	-0,040978	0,69444331	-	0,039355587	0,948675	-	0,090672556	1	-	0,029683374	0	0
86	UMAG_03586	-0,080844	0,5472811	-	-0,31784409	0,0203839	!	-0,13745215	0,5346073	-	-0,178713289	1	1
87	UMAG_03615	0,3827854	1	-	-0,46811432	0,0318701	!	0,009535787	1	-	-0,025264385	1	1
88	UMAG_03650	0,2861682	1	-	0,205562232	1	-	0,182888452	1	-	0,224872957	0	0
89	UMAG_03689	1,0090889	1	-	-0,20800479	0,2384281	-	0,200816423	1	-	0,333966843	0	0
90	UMAG_03744	-0,818835	1,5221E-05	!	-0,61010021	6,007E-06	!	-0,5457345	0,0004456	!	-0,658223293	3	3
91	UMAG_03745	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0
92	UMAG_03747	-0,357994	0,05695691	!	-0,07968958	0,444644	-	-0,13658233	0,5346073	-	-0,191421864	1	1
93	UMAG_03748	-0,444088	0,02967189	!	-0,16144731	0,2391985	-	0,096998077	1	-	-0,169512272	1	1
94	UMAG_03750	0,5236603	1	-	0,698283254	1	-	0,145928188	1	-	0,455957264	0	0
95	UMAG_03751	0,0810858	1	-	-0,07043021	0,7604142	-	-0,11638426	0,8910212	-	-0,035242877	0	0

96	UMAG_03753	0,3104444	1	-	-0,34580575	0,0158656	!	0,065821299	1	-	0,010153308	1	1
97	UMAG_03880	-0,145801	0,43236761	-	-0,37707474	0,0083662	!	-0,03173102	1	-	-0,184868858	1	1
98	UMAG_04033	NA	NA	?	0,52256098	1	-	NA	NA	?	NA	0	0
99	UMAG_04038	-0,608148	0,00155038	!	-0,62691562	6,822E-06	!	0,491258112	1	-	-0,247935008	2	2
100	UMAG_04039	-0,38068	0,04010026	!	-0,20607753	0,146362	-	1,127831313	1	-	0,180357971	1	1
101	UMAG_04057	-1,094613	1,7576E-08	!	0,002593752	0,7604142	-	0,402230672	1	-	-0,229929598	1	1
102	UMAG_04084	-0,204433	0,31933789	-	-0,54296784	0,0001701	!	0,187635151	1	-	-0,186588517	1	1
103	UMAG_04096	-0,528831	0,62983649	-	-1,37790335	0,0026342	!	0,228518583	1	-	-0,559405195	1	1
104	UMAG_04104	-0,316927	0,08720047	!	-0,18957199	0,207587	-	0,887273223	1	-	0,126924889	1	1
105	UMAG_04111	-0,268342	0,17742966	-	0,126391656	1	-	1,228595804	1	-	0,362215145	0	0
106	UMAG_04114	0,0967521	0,93649615	-	-0,12215466	0,3175203	-	-0,2306533	0,1891018	-	-0,08535194	0	0
107	UMAG_04145	0,0820601	0,91797497	-	0,165544944	1	-	0,058386643	1	-	0,101997239	0	0
108	UMAG_04185	0,3973542	1	-	-1,17795762	2,045E-06	!	0,562704843	1	-	-0,072632864	1	1
109	UMAG_04189	-0,355563	0,05407034	!	-0,20293216	0,1192852	-	-0,3835412	0,0208511	!	-0,3140122	2	2
110	UMAG_04282	0,1795752	1	-	0,170851605	1	-	1,138267814	1	-	0,496231544	0	0
111	UMAG_04400	0,0891396	0,92560348	-	-0,07239993	0,4504544	-	0,555904259	1	-	0,190881296	0	0
112	UMAG_04696	0,0689882	0,92560348	-	0,181491814	1	-	1,06745604	1	-	0,43931201	0	0
113	UMAG_04815	-0,440728	0,01737333	!	0,134538322	1	-	0,885441161	1	-	0,193083807	1	1
114	UMAG_04893	0,0501211	0,87089169	-	-0,00248744	0,7604142	-	0,37600265	1	-	0,141212104	0	0
115	UMAG_05222	-0,064811	0,58410573	-	0,250245397	1	-	0,519206989	1	-	0,234880433	0	0
116	UMAG_05227	0,0965362	0,93649615	-	0,294804238	1	-	0,613522258	1	-	0,334954246	0	0
117	UMAG_05294	-0,244523	0,19603594	-	-0,31028128	0,0263143	!	-0,35808978	0,0387002	!	-0,304298102	2	2
118	UMAG_05299	-0,043003	0,62641397	-	-2,94877346	4,99E-26	!	-2,31989789	1,545E-20	!	-1,770558011	2	2
119	UMAG_05300	0,7765492	1	-	0,518575393	1	-	0,108312148	1	-	0,467812236	0	0
120	UMAG_05301	-2,34855	1,7031E-28	!	-0,22251119	0,0883662	!	0,172824395	1	-	-0,799412406	2	2
121	UMAG_05302	-0,073638	0,59098686	-	-0,375979	0,0137858	!	-0,04170309	1	-	-0,163773206	1	1
122	UMAG_05308	0,2449018	1	-	-0,84899788	0,0004972	!	0,150736831	1	-	-0,151119763	1	1
123	UMAG_05310	-0,560218	0,00565662	!	0,241889432	1	-	1,655348344	1	-	0,445673364	1	1
124	UMAG_05319	-0,025865	0,62597876	-	-0,9784589	3,151E-11	!	-0,53679098	0,0002028	!	-0,513704826	2	2
125	UMAG_05439	-0,928588	1,2306E-05	!	-0,35511522	0,0462346	!	0,655084389	1	-	-0,209539533	2	2
126	UMAG_05548	-0,699716	0,00164597	!	0,140734077	1	-	0,49000703	1	-	-0,022991521	1	1
127	UMAG_05562	-0,957281	6,3643E-07	!	-0,59207269	3,587E-05	!	0,006722833	1	-	-0,514210434	2	2
128	UMAG_05641	-0,098129	0,40291182	-	-0,33736635	0,0221459	!	0,884978137	1	-	0,149827659	1	1
129	UMAG_05731	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0
130	UMAG_05733	0,2983925	1	-	0,31642527	1	-	0,314704686	1	-	0,309840814	0	0
131	UMAG_05780	-0,3713	0,05184107	!	0,436254989	1	-	0,190712489	1	-	0,085222489	1	1
132	UMAG_05781	-0,375265	0,05462662	!	0,149256278	1	-	0,181426561	1	-	-0,014860575	1	1
133	UMAG_05819	-0,560946	0,00441441	!	-0,28709046	0,0520257	!	-0,46889564	0,0047349	!	-0,438977325	3	3
134	UMAG_05861	1,0154897	1	-	-0,28184818	0,0637696	!	0,147490615	1	-	0,293710709	1	1
135	UMAG_05926	-0,080608	0,5472811	-	0,061383355	0,8339599	-	-0,32376334	0,0141751	!	-0,114329185	1	1
136	UMAG_05927	-0,181	0,21868119	-	-0,1453266	0,2957806	-	0,071215312	1	-	-0,085037098	0	0
137	UMAG_05931	-0,153578	0,31632068	-	-0,09543298	0,3175203	-	0,623578771	1	-	0,124855973	0	0
138	UMAG_05953	-0,350209	0,18063639	-	-0,04955766	0,7604142	-	-0,12535076	1	-	-0,175039282	0	0
139	UMAG_05988	-0,052705	0,61549867	-	0,030622005	0,8295892	-	-0,12282519	0,5729636	-	-0,048302839	0	0
140	UMAG_06064	-0,562194	0,00333258	!	-0,43246937	0,0026342	!	-0,30964837	0,0593143	!	-0,434770459	3	3
141	UMAG_06113	-1,711244	8,5699E-17	!	-0,45797463	0,0017786	!	0,346356528	1	-	-0,607620828	2	2
142	UMAG_06146	-0,703729	0,00030939	!	0,197043995	1	-	0,113964223	1	-	-0,130906769	1	1
143	UMAG_06158	-0,627226	0,00123296	!	-0,61569153	2,226E-05	!	-0,24637711	0,1681368	-	-0,496431666	2	2
144	UMAG_06178	-0,522059	0,00814935	!	-1,43614926	2,898E-22	!	0,088481238	1	-	-0,623242273	2	2
145	UMAG_06179	-0,093469	0,69402476	-	-0,34781944	0,0077946	!	-0,19835584	0,642962	-	-0,213214893	1	1
146	UMAG_06222	-0,930633	3,4281E-06	!	0,772849219	1	-	0,391251906	1	-	0,077822849	1	1
147	UMAG_06223	-1,073998	6,8478E-09	!	-0,48173965	0,0004471	!	-0,10523289	0,5570772	-	-0,553656703	2	2
148	UMAG_06428	-0,762054	5,2688E-05	!	-0,12995679	0,2947463	-	-0,51045491	0,0011451	!	-0,467488571	2	2
149	UMAG_06440	-0,591397	0,00220149	!	-0,53692872	0,0001485	!	-0,72005195	2,87E-06	!	-0,61612585	3	3
150	UMAG_10024	-0,250233	0,17181603	-	-0,29662599	0,0364668	!	-0,34498264	0,0250248	!	-0,297280577	2	2
151	UMAG_10030	0,0956738	0,94559172	-	0,206641758	1	-	-0,73926015	6,137E-06	!	-0,145648189	1	1
152	UMAG_10067	0,1704759	1	-	-2,42531592	5,53E-47	!	-1,30652106	3,035E-14	!	-1,187120343	2	2
153	UMAG_10076	0,0347331	0,81341567	-	-0,90273608	0,001192	!	0,142926074	1	-	-0,241692286	1	1
154	UMAG_10403	0,018597	0,93649615	-	0,133785675	1	-	0,076431616	1	-	0,076271423	0	0
155	UMAG_10553	0,1751888	1	-	0,057372204	0,7604142	-	-0,19120814	0,3254188	-	0,013784283	0	0
156	UMAG_10555	-0,22816	0,21541256	-	-0,26349839	0,0654136	!	-0,18544694	0,3197443	-	-0,225701633	1	1
157	UMAG_10557	0,2139101	1	-	-0,3745321	0,0103725	!	0,37142557	1	-	0,070267865	1	1
158	UMAG_10640	-0,205274	0,20542648	-	0,230081753	1	-	-0,44576608	0,0024945	!	-0,140319384	1	1
159	UMAG_10742	-0,280697	0,15971105	-	-0,13609884	0,3263415	-	-0,13269964	0,5729636	-	-0,183165267	0	0
160	UMAG_10756	0,5901413	1	-	0,326445627	1	-	-0,24653964	0,1891018	-	0,223349111	0	0
161	UMAG_10811	-0,204329	0,28345535	-	0,254624941	1	-	0,135544384	1	-	0,061946842	0	0
162	UMAG_10816	0,4656694	1	-	-1,5290426	3,827E-12	!	1,09383915	1	-	0,010155314	1	1
163	UMAG_10881	0,1946705	1	-	-1,7570879	1,435E-31	!	0,065999783	1	-	-0,498805884	1	1
164	UMAG_10972	-0,174165	0,31156243	-	0,178974684	1	-	-0,1039682	0,6673446	-	-0,033052718	0	0
165	UMAG_10975	-1,036614	5,8648E-08	!	-0,3438465	0,0107393	!	-0,67082953	2,479E-05	!	-0,683763272	3	3
166	UMAG_11002	0,6759294	1	-	0,376837934	1	-	0,422338172	1	-	0,49170184	0	0
167	UMAG_11062	-0,118665	0,4460056	-	-0,04938411	0,5854676	-	0,499317115	1	-	0,110422738	0	0
168	UMAG_11094	-0,550707	0,01627942	!	-0,35997784	0,0349277	!	0,806726713	1	-	-0,034652588	2	2
169	UMAG_11193	-0,018529	0,69069937	-	-0,01115208	0,7199773	-	-0,61324849	2,698E-05	!	-0,214309818	1	1
170	UMAG_11250	-0,19954	0,08293528	!	-0,94475069	3,339E-12	!	0,953003478	1	-	-0,063762559	2	2
171	UMAG_11305	-0,421974	0,05407034	!	-0,18403535	0,3402124	-	0,082484037	1	-	-0,174508411	1	1
172	UMAG_11362	-0,520853	0,00716941	!	-1,50898248	1,858E-22	!	0,139470557	1	-	-0,630121565	2	2
173	UMAG_11377	-0,103139	0,47695038	-	0,140674771	1	-	0,332593893	1	-	0,123376615	0	0
174	UMAG_11402	-1,221668	4,6756E-06	!	-0,62352946	1,235E-05	!	-0,52077831	0,0006514	!	-0,788658448	3	3
175	UMAG_11403	-0,169243	0,31823561	-	-0,05871974	0,5043645	-	0,075077997	1	-	-0,050961494	0	0
176	UMAG_11415	-0,220167	0,21979569	-	-0,28815647	0,0500036	!	0,943183844	1	-	0,144953432	1	1
177	UMAG_11416	0,0172272	0,78349846	-	-0,01047652	0,7604142	-	-0,68416136	2,593E-05	!	-0,22580356	1	1
178	UMAG_11417	0,3959739	1	-	0,125168453	1	-	0,205890348	1	-	0,242344248	0	0
179	UMAG_11444	-0,175351	0,31870649	-	0,092010335	1	-	-0,75122113	2,906E-06	!	-0,278187186	1	1
180	UMAG_11464	-0,052373	0,61708027	-	-0,03448341	0,6128877	-	-0,14475673	0,5421004	-	-0,077204425	0	0
181	UMAG_11586	0,1830076	1	-	0,318009381	1	-	-0,71350915	4,623E-06	!	-0,070830717	1	1
182	UMAG_11639	0,3239982	1	-	0,097895554	1	-	0,333393243	1	-	0,251762318	0	0
183	UMAG_11931	-1,014258	0,37626489	-	0,256625644	1	-	-0,08678712	0,642962	-	-0,281473042	0	0
184	UMAG_11940	-0,628649	0,00223929	!	-1,02417756	1,372E-13	!	-3,11227355	1,813E-52	!	-1,588366858	3	3
185	UMAG_12045	-2,090559	5,6101E-22	!	-1,54531032	3,266E-23	!	-1,67322557	8,861E-12	!	-1,769698229	3	3
186	UMAG_12127	0,0076911	0,78349846	-	0,331167981	1	-	-0,00724201	1	-	0,110539032	0	0
187	UMAG_12197	-2,418656	4,833E-30	!	-2,16785824	1,895E-40	!	-0,48917258	0,0028923	!	-1,69189554	3	3
188	UMAG_12216	-0,769399	4,0288E-05	!	-0,37818079	0,0081347	!	0,996208714	1	-	-0,050456935	2	2
189	UMAG_12226	-1,200798	5,2688E-10	!	-1,2557052	3,542E-18	!	-0,81793351	5,767E-07	!	-1,091478855	3	3
190	UMAG_12233	-0,430986	0,01588665	!	-0,35688777	0,0052283	!	-0,28143609	0,0554927	!	-0,356436707	3	3
191	UMAG_12281	-2,597077	1,7682E-33	!	-1,81808434	3,85E-29	!	-0,83797358	2,573E-07	!	-1,751045072	3	3
192	UMAG_12302	0,1696971	1	-	-0,26062932	0,06566	!	0,110720958	1	-	0,006596233	1	1
193	UMAG_12313	-0,956836	7,6831E-07	!	-0,42325237	0,0027336	!	0,315960577	1	-	-0,354709174	2	2
194	UMAG_12330	-0,177641	0,30133843	-	-0,10414044	0,3701394	-	0,512267526	1	-	0,076828854	0	0
195	UMAG_15020	-0,93142	0,00411616	!	0,004990615	0,7603772	-	1,769768255	1	-	0,281113072	1	1

Experiment B

No	Gene ID	R1,log2fc	R1,qval	R1	R2,log2fc	R2,qval	R2	R3,log2fc	R3,qval	R3	Mean of log2fc	No. Significant	No. significant or Zero	Mean of log2fc of Experimenta A and B
1	UMAG_00054	NA	NA	0	NA	NA	0	-0,0276871	0,7703805	-	#VALUE!	0	2	NA
2	UMAG_00081	-0,396557	0,189911	-	-0,1928316	0,529158	-	0,06863466	0,9947147	-	-0,173584495	0	0	0,010686326
3	UMAG_00105	-2,041722	2,62E-19	!	-2,436836	3,43E-13	!	NA	NA	0	NA	2	3	NA
4	UMAG_00159	1,2490273	1	-	0,0356459	0,511684	-	0,98816919	1	-	0,757614133	0	0	0,332638288
5	UMAG_00187	-0,038206	0,326042	-	-0,350713	0,277034	-	-1,6066251	0,0003959	!	-0,665181362	1	1	-0,26671642
6	UMAG_00558	-0,94699	7,88E-06	!	0,1039908	0,869407	-	0,03748399	0,9665492	-	-0,268504986	1	1	-0,593863435
7	UMAG_00781	-0,24486	0,046625	!	-0,3758761	0,304731	-	0,00356186	0,805095	-	-0,205724818	1	1	-0,150819337
8	UMAG_00792	-0,462856	0,042321	!	NA	NA	0	-0,3055051	0,2745409	-	NA	1	2	NA
9	UMAG_00793	0,2079339	1	-	0,1745728	0,962913	-	-0,6636098	0,0129327	!	-0,093701001	1	1	-0,076737068
10	UMAG_00795	NA	NA	0	0,4509261	0,999956	-	0,55629466	1	-	NA	0	1	NA
11	UMAG_00823	0,4072626	1	-	0,5897754	0,999956	-	0,63296904	1	-	0,543335685	0	0	-0,289334145
12	UMAG_00885	-0,117006	0,442796	-	NA	NA	0	NA	NA	0	NA	0	2	NA
13	UMAG_01018	-0,469537	9,25E-24	!	-0,4043161	0,156739	-	0,18513096	1	-	-0,229574124	1	1	-0,124260018
14	UMAG_01082	-0,31932	7,41E-06	!	0,645151	0,999956	-	-1,1946718	0,7026497	-	-0,289613585	1	1	-0,170546583
15	UMAG_01130	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0	NA
16	UMAG_01235	-0,290599	0,085493	!	-0,1362131	0,583998	-	0,70168369	1	-	0,091623834	1	1	0,057643302
17	UMAG_01236	0,0208961	0,222309	-	-0,0860878	0,623077	-	-0,7738356	2,54E-05	!	-0,279675762	1	1	-0,360849406
18	UMAG_01237	0,243134	1	-	0,262472	0,995566	-	NA	NA	0	NA	0	1	NA
19	UMAG_01238	0,6417829	1	-	-0,157501	0,57107	-	0,18121963	1	-	0,221833872	0	0	-0,028660364
20	UMAG_01239	0,3376454	1	-	0,3041853	0,998297	-	0,69828124	1	-	0,446703973	0	0	0,152704593
21	UMAG_01240	-1,263026	6,05E-08	!	-1,2853107	0,000187	!	-1,8702101	1,33E-08	!	-1,472849012	3	3	-1,156379396
22	UMAG_01289	-0,778868	2,99E-05	!	0,6378062	0,999956	-	0,15224013	1	-	0,003725956	1	1	-0,299974798
23	UMAG_01297	-0,269374	0,03384	!	0,1740889	0,903823	-	-0,0169522	0,7703805	-	-0,03741232	1	1	-0,20967986
24	UMAG_01300	-1,915812	3,60E-27	!	0,0183252	0,807175	-	-0,0222563	0,7703805	-	-0,639914259	1	1	-0,593837352
25	UMAG_01301	-0,455811	0,010144	!	-0,2952111	0,428137	-	-0,0865326	0,2810777	-	-0,279184978	1	1	-0,268513434
26	UMAG_01302	0,0618652	1	-	0,2004577	0,968765	-	-0,1216704	0,5175705	-	0,046884162	0	0	-0,035841381
27	UMAG_01375	NA	NA	0	-4,4262326	2,33E-18	!	-2,9517716	1,18E-41	!	NA	2	3	NA
28	UMAG_01553	NA	NA	0	-0,205311	0,456281	-	0,65515382	1	-	NA	0	1	NA
29	UMAG_01689	-2,076729	7,88E-06	!	-2,3581821	1,05E-10	!	-2,3047901	8,28E-22	!	-2,24656714	3	3	-2,161945955
30	UMAG_01690	NA	NA	0	-0,1782472	0,57107	-	0,46226531	1	-	NA	0	1	NA
31	UMAG_01779	NA	NA	0	-0,4094525	0,241325	-	-0,3448653	0,2367278	-	NA	0	1	NA
32	UMAG_01820	-0,836759	8,17E-06	!	0,2320311	0,97673	-	0,85789701	1	-	0,08438984	1	1	-0,251416121
33	UMAG_01854	-0,050224	0,189533	-	-0,0352267	0,74958	-	0,32749813	1	-	0,080682471	0	0	0,102333563
34	UMAG_01858	0,4900857	1	-	-0,3558184	0,28138	-	-0,3917612	0,0354739	!	-0,085831293	1	1	0,157215362
35	UMAG_01940	-0,423948	4,38E-16	!	0,0684252	0,853248	-	0,14021974	1	-	-0,071767758	1	1	-0,23508162
36	UMAG_01977	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0	NA
37	UMAG_01987	-1,473564	1,63E-14	!	-1,7046938	2,51E-07	!	-1,9637009	2,97E-09	!	-1,713986061	3	3	-1,748834733
38	UMAG_01997	-0,472096	0,000102	!	0,0493768	0,830417	-	-0,2548662	0,5175705	-	-0,225861958	1	1	-0,133527379
39	UMAG_02006	-0,10723	0,439517	-	0,2432709	0,97673	-	0,48017819	1	-	0,205406387	0	0	0,242637624
40	UMAG_02011	-1,945418	8,72E-20	!	-2,4714622	2,11E-10	!	-1,3608316	0,0042673	!	-1,925903923	3	3	-2,06920252
41	UMAG_02119	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0	NA
42	UMAG_02135	-0,77531	0,000243	!	0,0572842	0,830417	-	0,01150201	0,8866598	-	-0,235507998	1	1	-0,064739393
43	UMAG_02137	NA	NA	0	-0,2540866	0,57107	-	-0,3583636	0,4914937	-	NA	0	1	NA
44	UMAG_02138	NA	NA	0	NA	NA	0	-0,540652	0,002121	!	NA	1	3	NA
45	UMAG_02139	NA	NA	?	-0,2654532	0,376555	-	1,52502685	1	-	NA	0	0	NA
46	UMAG_02141	0,4910673	1	-	0,2886417	0,995566	-	0,32104943	1	-	0,36691947	0	0	-0,796219731
47	UMAG_02192	-0,097649	0,519643	-	0,0116835	0,807175	-	-0,0791238	0,7703805	-	-0,055029674	0	0	0,146360821
48	UMAG_02193	-0,408912	0,035131	!	-0,159428	0,559979	-	-0,1181615	0,7703805	-	-0,22883395	1	1	-0,107538607
49	UMAG_02229	0,5357384	1	-	-0,0130343	0,780447	-	-0,0841395	0,3358182	-	0,146188225	0	0	-0,065493839
50	UMAG_02239	-0,41361	0,000174	!	-0,5961416	0,807175	-	0,241157	1	-	-0,256198079	1	1	-0,213386836
51	UMAG_02243	-0,486365	0,002953	!	-0,297883	0,351283	-	-0,009279	0,752688	-	-0,264509094	1	1	-0,069275247
52	UMAG_02294	-0,334276	0,218935	-	-0,5788573	0,06145	!	-0,0215824	0,7988799	-	-0,311571796	1	1	-0,077820728
53	UMAG_02295	-0,113786	0,394039	-	0,4325746	0,999956	-	0,07822441	1	-	0,132337547	0	0	0,049069304
54	UMAG_02297	0,7718344	1	-	0,3924473	0,999956	-	-0,1525882	0,5175705	-	0,337231186	0	0	-0,043613733
55	UMAG_02298	-0,302069	0,084851	!	-0,2453803	0,456281	-	-0,0889025	0,6493109	-	-0,212117338	1	1	-0,203795001
56	UMAG_02299	-1,62996	0,008992	!	NA	NA	?	NA	NA	0	NA	1	2	NA
57	UMAG_02430	-0,196123	0,006234	!	-0,1694506	0,511684	-	0,2623849	1	-	-0,034396367	1	1	0,051190193
58	UMAG_02466	-0,118098	0,353869	-	-0,1670569	0,560342	-	-0,1797453	0,3260636	-	-0,154966827	0	0	-0,016005963
59	UMAG_02473	-0,173932	0,097464	!	0,0457192	0,807175	-	0,33433455	1	-	0,068707376	1	1	-0,32244769
60	UMAG_02474	-1,603378	3,60E-13	!	-0,2649293	0,376555	-	-0,0800511	0,6230358	-	-0,649452648	1	1	-0,380879328
61	UMAG_02475	-3,340682	1,15E-29	!	-3,1319386	3,95E-12	!	-2,9841707	1,28E-08	!	-3,152263877	3	3	-2,928063785
62	UMAG_02533	NA	NA	0	-0,2804261	0,442245	-	-0,2268953	0,0189046	!	NA	1	2	NA
63	UMAG_02535	-1,428833	8,00E-86	!	-0,9496266	0,000303	!	-0,5337796	0,7703805	-	-0,970746459	2	2	-0,690154721
64	UMAG_02537	-0,146147	0,429804	-	-0,3486467	0,305012	-	0,07609265	1	-	-0,139566865	0	0	-0,174027353
65	UMAG_02538	0,2745602	1	-	-0,3145811	0,363538	-	0,66971956	1	-	0,209899544	0	0	0,252924165
66	UMAG_02560	-0,752769	0,000249	!	-0,8042031	0,008301	!	-0,6655452	0,0433016	!	-0,74083903	3	3	-2,319505847
67	UMAG_02611	-1,037033	7,53E-08	!	-0,0730535	0,692492	-	-0,57427	0,1401402	-	-0,561452036	1	1	-0,245725966
68	UMAG_02813	-1,045605	2,65E-07	!	0,1568728	0,938281	-	-0,6847466	3,56E-49	!	-0,524492828	2	2	-0,358464526
69	UMAG_02826	0,1129554	1	-	0,5719793	0,999956	-	-0,2403279	0,5175705	-	0,14820228	0	0	0,091920758
70	UMAG_02851	-0,426547	2,32E-07	!	-0,2665246	0,351283	-	-1,4575466	0,0002495	!	-0,7168727	2	2	-0,461630156
71	UMAG_02852	NA	NA	0	NA	NA	0	-0,3032726	0,0025153	!	NA	1	3	NA
72	UMAG_02853	0,0927118	1	-	0,0138455	0,807175	-	1,09534923	1	-	0,400635502	0	0	-0,191646279
73	UMAG_03023	-0,005393	0,611821	-	-1,0404077	0,001573	!	-0,8707624	0,021816	!	-0,638854394	2	2	-0,297508166
74	UMAG_03046	0,2839626	1	-	-0,0714334	0,67564	-	0,50720875	1	-	0,239912666	0	0	0,164619304
75	UMAG_03065	0,1055498	1	-	-0,0851168	0,67564	-	-0,087296	0,496114	-	-0,022287667	0	0	-0,017825973
76	UMAG_03105	-0,238644	0,223624	-	-2,1079323	3,34E-08	!	-1,0004422	0,0058837	!	-1,115672855	2	2	-1,264238339
77	UMAG_03112	-0,096303	0,146112	-	-0,1956352	0,511684	-	-0,3527796	0,268207	-	-0,214905852	0	0	0,040868297
78	UMAG_03138	-0,331138	0,042321	!	-0,2193058	0,428137	-	-0,1784416	1,02E-06	!	-0,242961679	2	2	-0,222087201
79	UMAG_03201	0,1058074	1	-	-2,0884543	4,99E-15	!	-0,9371541	1,60E-51	!	-0,973267011	2	2	-1,110564684
80	UMAG_03202	-0,075666	0,512568	-	0,0482981	0,830417	-	-0,0467438	0,7703805	-	-0,024703905	0	0	-0,042250875
81	UMAG_03223	-0,431336	4,53E-09	!	-0,0950324	0,67564	-	0,21767616	1	-	-0,102897367	1	1	-0,230930105
82	UMAG_03313	-0,375743	0,034867	!	-0,212951	0,442245	-	-0,1369323	0,7088152	-	-0,241875539	1	1	-0,641282071
83	UMAG_03382	-0,015555	0,099029	!	0,2920997	0,995566	-	-0,1540166	0,1566806	-	0,040842775	1	1	0,063314208
84	UMAG_03397	0,5259413	1	-	-0,1335861	0,57107	-	0,29358693	1	-	0,228647362	0	0	0,243376426
85	UMAG_03564	0,7401381	1	-	0,0602941	0,853248	-	-0,4561336	0,1999109	-	0,11476619	0	0	0,072224782
86	UMAG_03586	-0,531344	2,43E-05	!	-0,3354523	0,293464	-	0,17836138	1	-	-0,229478422	1	1	-0,204095856
87	UMAG_03615	-0,10705	0,442796	-	-0,0813819	0,67564	-	-0,4172248	0,3258196	-	-0,201885537	0	0	-0,113574961
88	UMAG_03650	-0,50268	0,000576	!	-0,503358	0,102078	-	0,31140288	1	-	-0,231545023	1	1	-0,003336033
89	UMAG_03689	3,2633022	1	-	0,100273	0,853248	-	0,25280832	1	-	1,205461188	0	0	0,769714016
90	UMAG_03744	-0,537497	3,36E-18	!	-0,6561113	0,012521	!	-0,1874826	0,2012578	-	-0,46036371	2	2	-0,559293502
91	UMAG_03745	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0	NA
92	UMAG_03747	-0,154085	0,257199	-	-0,1560658	0,511684	-	0,10647543	1	-	-0,067891805	0	0	-0,129656834
93	UMAG_03748	0,1290447	1	-	-0,1272186	0,560342	-	-0,2186514	0,2991396	-	-0,072275114	0	0	-0,120893693
94	UMAG_03750	-1,189314	2,99E-05	!	0,8253598	0,999956	-	0,71677768	1	-	0,117607909	1	1	0,286782587
95	UMAG_03751	-0,17696	0,042321	!	-0,5852883	0,807175	-	-0,4613793	0,217987	-	-0,407875721	1	1	-0,221559299

96	UMAG_03753	-0,314977	0,061006	!	-0,0113483	0,770595	-	-0,0070062	0,8074408	-	-0,111110591	1	1	-0,050478642
97	UMAG_03880	-0,368062	0,02665	!	0,004827	0,807175	-	0,08967937	1	-	-0,091185226	1	1	-0,138027042
98	UMAG_04033	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0	NA
99	UMAG_04038	-0,072822	0,180097	-	-0,087375	0,650847	-	-0,0863256	0,7286861	-	-0,08217418	0	0	-0,165054594
100	UMAG_04039	-0,457039	3,01E-09	!	-0,5225903	0,102078	-	1,57335018	1	-	0,19790705	1	1	0,189132511
101	UMAG_04057	-0,854011	1,25E-11	!	-0,380993	0,23198	-	0,57714588	1	-	-0,219286065	1	1	-0,224607831
102	UMAG_04084	0,165802	1	-	-0,5934925	0,056391	!	-0,2995835	0,3310206	-	-0,242424663	1	1	-0,21450659
103	UMAG_04096	NA	NA	?	NA	NA	?	NA	NA	?	NA	0	0	NA
104	UMAG_04104	-0,442422	2,67E-19	!	0,2270399	0,97673	-	-0,8543804	0,021816	!	-0,356587466	2	2	-0,114831288
105	UMAG_04111	-0,279746	6,06E-12	!	-1,6660175	0,032124	!	-0,1263207	0,6800493	-	-0,690694843	2	2	-0,164239849
106	UMAG_04114	-0,067677	0,343973	-	-0,1699225	0,511684	-	-0,2139918	0,004817	!	-0,15053045	1	1	-0,117941195
107	UMAG_04145	-0,327657	0,018505	!	0,2420492	0,995566	-	-0,3896751	0,2320324	-	-0,158427714	1	1	-0,028215237
108	UMAG_04185	0,5227842	1	-	NA	NA	0	-0,1642755	0,5555673	-	NA	0	1	NA
109	UMAG_04189	-0,586407	4,75E-12	!	-0,0751258	0,67564	-	-0,0850577	0,6493109	-	-0,248863539	1	1	-0,28143787
110	UMAG_04282	-0,441368	0,000249	!	0,0489359	0,807175	-	0,53510517	1	-	0,047557827	1	1	0,271894686
111	UMAG_04400	-0,615999	0,001314	!	0,0460681	0,830417	-	-0,4275089	0,0190928	!	-0,332479823	2	2	-0,070799263
112	UMAG_04696	-0,118994	0,009772	!	-0,0703984	0,67564	-	-0,4574953	0,0718403	!	-0,215629411	2	2	0,111841299
113	UMAG_04815	-0,103137	0,141095	-	-0,3356515	0,30547	-	0,05735788	1	-	-0,127143692	0	0	0,032970057
114	UMAG_04893	0,3001425	1	-	0,1024139	0,853248	-	-0,2942727	0,1104398	-	0,036094536	0	0	0,08865332
115	UMAG_05222	-0,192291	0,1811	-	-0,0747169	0,69799	-	0,19666967	1	-	-0,023445953	0	0	0,10571724
116	UMAG_05227	0,1663923	1	-	-0,2603391	0,428137	-	0,27261389	1	-	0,059555692	0	0	0,197254969
117	UMAG_05294	-0,037279	0,663147	-	0,0567309	0,830417	-	0,35449377	1	-	0,124648647	0	0	-0,089824727
118	UMAG_05299	-0,262845	0,203429	-	0,5311346	0,999956	-	1,23693632	1	-	0,50174196	0	0	-0,634408026
119	UMAG_05300	0,3892122	1	-	-0,3955451	0,277034	-	NA	NA	0	NA	0	1	NA
120	UMAG_05301	-2,856158	8,47E-42	!	-0,2239671	0,456262	-	-0,0052991	0,7703805	-	-1,028474619	1	1	-0,913943513
121	UMAG_05302	0,2021347	1	-	-0,5073576	0,106068	-	-0,2263541	0,5555673	-	-0,177192338	0	0	-0,170482772
122	UMAG_05308	-0,032778	0,512568	-	NA	NA	0	0,56200607	1	-	NA	0	1	NA
123	UMAG_05310	0,0231823	0,807996	-	-0,177813	0,511684	-	-0,8850115	0,0377318	!	-0,346547406	1	1	0,049562979
124	UMAG_05319	-0,069422	0,417776	-	-1,2350009	6,78E-05	!	-0,1700822	0,0384462	!	-0,49150169	2	2	-0,502603258
125	UMAG_05439	-0,163689	1,01E-05	!	0,0260523	0,807175	-	-0,0602968	0,7446771	-	-0,065977841	1	1	-0,137758687
126	UMAG_05548	-0,56099	0,000249	!	-0,0660024	0,650847	-	-0,1687806	0,5468528	-	-0,265257836	1	1	-0,144124679
127	UMAG_05562	-1,070107	7,05E-39	!	0,109916	0,869407	-	-0,0197353	0,7703805	-	-0,326642107	1	1	-0,42042627
128	UMAG_05641	-0,085539	0,141095	-	-0,0762916	0,733946	-	-0,2108411	0,2320324	-	-0,124223909	0	0	0,012801875
129	UMAG_05731	0,3078379	1	-	-0,5116554	0,102078	-	-0,2462573	0,0603326	!	-0,150024932	1	1	NA
130	UMAG_05733	-0,315311	0,03384	!	0,1178213	0,869407	-	0,01587569	0,890269	-	-0,060537864	1	1	0,124651475
131	UMAG_05780	-0,588026	1,12E-05	!	0,2222529	0,97673	-	-0,4700604	0,1878854	-	-0,278611153	1	1	-0,096694332
132	UMAG_05781	-0,699853	5,65E-14	!	-0,3495848	0,277034	-	-0,5002895	0,1104398	-	-0,516575805	1	1	-0,26571819
133	UMAG_05819	-0,066035	0,517694	-	0,0990851	0,857344	-	-0,0054253	0,8760015	-	0,009208385	0	0	-0,21488447
134	UMAG_05861	0,0842444	0,982545	-	-0,1444377	0,57107	-	-0,2133322	0,3412581	-	-0,091175153	0	0	0,101267778
135	UMAG_05926	-0,515844	0,010465	!	-0,2565106	0,456262	-	-0,4375313	7,55E-08	!	-0,403295168	2	2	-0,258812176
136	UMAG_05927	0,4518457	1	-	-0,2783872	0,428137	-	-0,3004984	0,2905259	-	-0,042346662	0	0	-0,06369188
137	UMAG_05931	-0,464355	4,13E-15	!	0,0258115	0,807175	-	0,0311076	0,8897957	-	-0,135811952	1	1	-0,00547799
138	UMAG_05953	0,3580587	1	-	-0,0763479	0,564161	-	-0,3742795	0,0851067	!	-0,030856222	1	1	-0,102947752
139	UMAG_05988	-0,064397	0,58003	-	-0,0650068	0,675609	-	0,34201244	1	-	0,070869506	0	0	0,011283333
140	UMAG_06064	0,1539009	1	-	-0,3182	0,30547	-	-0,4056891	1,18E-05	!	-0,18999604	1	1	-0,312383249
141	UMAG_06113	-0,278084	0,037124	!	0,0420645	0,807175	-	-0,1580066	0,565509	-	-0,131341954	1	1	-0,369481391
142	UMAG_06146	0,1022406	1	-	0,0794302	0,853248	-	-0,7612256	0,0384462	!	-0,193184933	1	1	-0,162045851
143	UMAG_06158	-0,129733	0,241802	-	0,0607614	0,853248	-	0,28397613	1	-	0,071668193	0	0	-0,212381737
144	UMAG_06178	-0,207053	0,008786	!	0,2170508	0,968765	-	0,01206692	0,8760015	-	0,007355039	1	1	-0,307943617
145	UMAG_06179	-0,27185	0,000548	!	-1,2795498	6,89E-06	!	0,08424719	0,9947147	-	-0,489050878	2	2	-0,351132886
146	UMAG_06222	-0,796505	2,61E-06	!	-0,3205091	0,350141	-	-0,509227	0,1673702	-	-0,542080302	1	1	-0,232128727
147	UMAG_06223	NA	NA	?	0,1400462	0,923253	-	NA	NA	0	NA	0	1	NA
148	UMAG_06428	0,0367979	1	-	-0,3285449	0,293464	-	-0,0027494	0,752688	-	-0,098165472	0	0	-0,282827022
149	UMAG_06440	-0,119706	0,000432	!	-0,667034	0,046111	!	-0,6189552	4,03E-19	!	-0,468565105	3	3	-0,542345478
150	UMAG_10024	0,8991185	1	-	-0,55679	0,128639	-	0,05751028	1	-	0,133279595	0	0	-0,082000491
151	UMAG_10030	0,5084194	1	-	-0,2111822	0,511684	-	-0,6317788	0,1299646	-	-0,111513869	0	0	-0,128581029
152	UMAG_10067	-0,156253	0,208811	-	-3,7596213	1,95E-19	!	-3,0550174	4,96E-16	!	-2,323630458	2	2	-1,7553754
153	UMAG_10076	NA	NA	0	-0,1031433	0,67564	-	0,39524771	1	-	NA	0	1	NA
154	UMAG_10403	0,1162202	0,915893	-	-0,3904599	0,241325	-	-0,0070476	0,8335081	-	-0,093762431	0	0	-0,008745504
155	UMAG_10553	0,0467013	0,841861	-	0,2202619	0,968765	-	0,2526698	1	-	0,173210976	0	0	0,093497629
156	UMAG_10555	0,3579621	1	-	-0,9599628	0,001223	!	-0,0358749	0,3769802	-	-0,212625201	1	1	-0,219163417
157	UMAG_10557	0,2616308	1	-	-0,1830238	0,62018	-	0,12394182	1	-	0,067516276	0	0	0,06889207
158	UMAG_10640	-0,275696	0,014472	!	0,114714	0,881087	-	NA	NA	0	NA	1	2	NA
159	UMAG_10742	-0,032763	0,611821	-	-0,4410582	0,170119	-	-0,2738726	0,1999109	-	-0,24923143	0	0	-0,216198348
160	UMAG_10756	0,1680264	1	-	0,2940597	0,995566	-	0,32624673	1	-	0,262777615	0	0	0,243063363
161	UMAG_10811	0,0332098	0,783544	-	0,1409827	0,901629	-	0,50379415	1	-	0,225995544	0	0	0,143971193
162	UMAG_10816	-0,049394	0,596408	-	0,7942473	0,999956	-	0,2837939	1	-	0,342882467	0	0	0,17651889
163	UMAG_10881	-0,223001	0,146112	-	NA	NA	0	-2,7539894	2,69E-13	!	NA	1	2	NA
164	UMAG_10972	-0,519195	0,000292	!	0,8122354	0,999956	-	0,73058883	1	-	0,341209693	1	1	0,154078488
165	UMAG_10975	-0,464086	0,019778	!	0,0168017	0,830417	-	-0,8533457	1,94E-08	!	-0,433543186	2	2	-0,558653229
166	UMAG_11002	0,1465229	1	-	-0,2220865	0,456262	-	-0,1931515	0,3561209	-	-0,089571684	0	0	0,201065078
167	UMAG_11062	-0,641573	3,03E-34	!	0,013468	0,807175	-	0,20490155	1	-	-0,141067937	1	1	-0,0153226
168	UMAG_11094	0,3798189	1	-	-0,1285592	0,494998	-	-1,7789223	0,0008149	!	-0,509220887	1	1	-0,271936738
169	UMAG_11193	-1,392726	0,110245	-	-0,0099878	0,807175	-	-0,486304	1,98E-11	!	-0,62967247	1	1	-0,421991144
170	UMAG_11250	-0,175149	0,429804	-	-1,2163399	6,32E-05	!	0,44344457	1	-	-0,316014672	1	1	-0,189888615
171	UMAG_11305	0,3743269	1	-	-0,3368079	0,304731	-	0,30674484	1	-	0,114754609	0	0	-0,029876901
172	UMAG_11362	0,2093598	1	-	-0,5366336	0,102078	-	0,5101606	1	-	0,060962256	0	0	-0,284579655
173	UMAG_11377	-0,307339	0,083399	!	NA	NA	0	0,30641492	1	-	NA	1	2	NA
174	UMAG_11402	-0,799659	3,23E-05	!	-0,8975244	0,002141	!	-0,4638228	2,44E-15	!	-0,720335429	3	3	-0,754496938
175	UMAG_11403	-0,114799	5,22E-06	!	0,0796867	0,853248	-	-2,778328	1,95E-13	!	-0,937813593	2	2	-0,494387543
176	UMAG_11415	-0,388207	1,41E-43	!	-0,0451066	0,67564	-	-0,8338664	8,56E-05	!	-0,422393442	2	2	-0,138720005
177	UMAG_11416	-0,060593	0,611821	-	-0,1680295	0,511684	-	-0,3675718	0,4143428	-	-0,198731555	0	0	-0,212267558
178	UMAG_11417	-0,314933	0,010528	!	-0,959693	0,001969	!	0,25460388	1	-	-0,340007454	2	2	-0,048831603
179	UMAG_11444	-0,519719	0,00063	!	-0,1332594	0,563587	-	-0,9987581	0,7548659	-	-0,550578968	1	1	-0,414383077
180	UMAG_11464	-0,173615	0,159162	-	0,2915659	0,995566	-	0,17598235	1	-	0,097977584	0	0	0,01038658
181	UMAG_11586	0,3654326	1	-	-0,3238049	0,351283	-	NA	NA	0	NA	0	1	NA
182	UMAG_11639	-0,477714	0,001314	!	0,309451	0,995566	-	0,54980418	1	-	0,127180356	1	1	0,189471337
183	UMAG_11931	NA	NA	0	0,5030926	0,999956	-	0,21072137	1	-	NA	0	1	NA
184	UMAG_11940	-0,51913	0,001987	!	-1,259413	6,16E-07	!	-1,0847486	5,41E-06	!	-0,954430437	3	3	-1,271398648
185	UMAG_12045	-4,501169	5,86E-30	!	-4,2928222	6,00E-20	!	-2,8894954	5,93E-08	!	-3,894495543	3	3	-2,832096886
186	UMAG_12127	-0,037125	0,596408	-	0,0822024	0,853248	-	0,30947647	1	-	0,118184506	0	0	0,114361769
187	UMAG_12197	-3,822788	0	!	-3,7956787	4,48E-20	!	-1,8585797	2,01E-11	!	-3,159015476	3	3	-2,425455508
188	UMAG_12216	-0,807401	1,79E-64	!	-0,8820729	0,000794	!	-0,1441566	0,5555673	-	-0,611210278	2	2	-0,330833606
189	UMAG_12226	-1,467882	8,00E-86	!	-1,9365471	2,74E-10	!	-1,0574146	0,0095072	!	-1,487281335	3	3	-1,289380095
190	UMAG_12233	-1,338749	5,36E-20	!	-0,3478455	0,277034	-	0,06591772	1	-	-0,540225596	1	1	-0,448331152
191	UMAG_12281	-3,748321	1,30E-188	!	-3,3334932	3,91E-15	!	-3,4112679	2,55E-40	!	-3,497693995	3	3	-2,624369533
192	UMAG_12302	-0,122127	0,319921	-	-0,8430745	0,000829	!	-0,1570398	0,3943877	-	-0,37408053	1	1	-0,183742148
193	UMAG_12313	-0,612849	3,13E-12	!	-0,2105798	0,494998	-	-0,3909241	0,0030968	!	-0,404784216	2	2	-0,379746695
194	UMAG_12330	-0,385502	4,97E-69	!	-0,0080893	0,807175	-	-0,4819063	0,0003405	!	-0,291832643	2	2	-0,107501894
195	UMAG_15020	NA	NA	0	-0,0106147	0,791503	-	-3,229704	1,81E-18	!	NA	1	2	NA

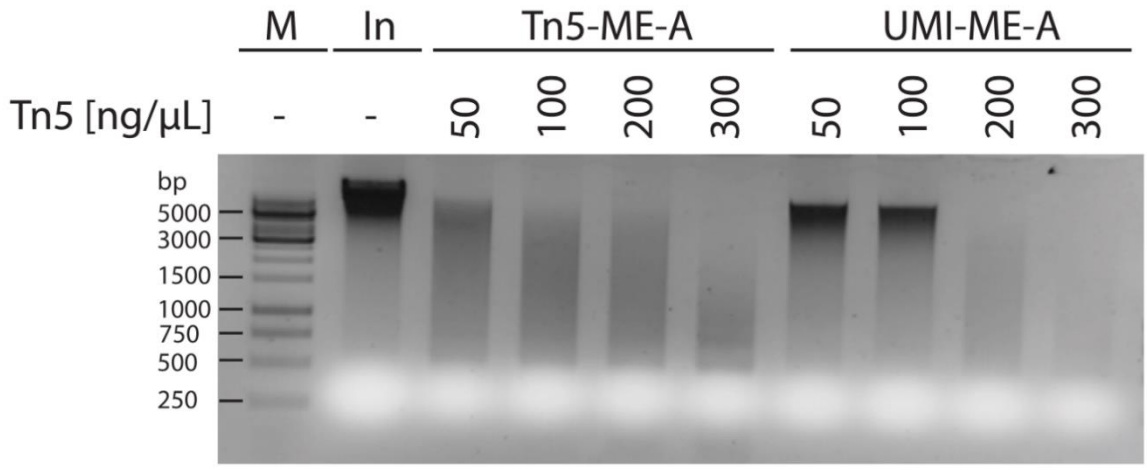
S2 Data. Symptom rating of mutant strains

Replicate 1	SG200_1	UMAG_03015	UMAG_12045	UMAG_01689
No symptoms	1	3	4	2
Chlorosis	0	45	51	46
Ligula Swelling	0	0	0	0
Small tumors	3	0	0	3
Normal tumors	22	0	0	3
Heavy tumors	6	0	0	0
Stunted	11	0	1	0
Dead	0	0	0	0

Replicate 2	SG200_2	UMAG_03015	UMAG_12045	UMAG_01689
No symptoms	1	0	0	2
Chlorosis	6	49	49	47
Ligula Swelling	0	0	0	0
Small tumors	5	0	0	0
Normal tumors	34	0	0	0
Heavy tumors	3	0	0	0
Stunted	8	0	0	0
Dead	0	0	0	0

Replicate 3	SG200_3	UMAG_03015	UMAG_12045	UMAG_01689
No symptoms	0	2	4	5
Chlorosis	4	51	50	50
Ligula Swelling	0	0	0	0
Small tumors	5	1	0	0
Normal tumors	31	0	0	0
Heavy tumors	5	0	0	0
Stunted	7	0	0	0
Dead	0	0	0	0

S2 Figure. Tn5 fragmentation of gDNA with modified adapters



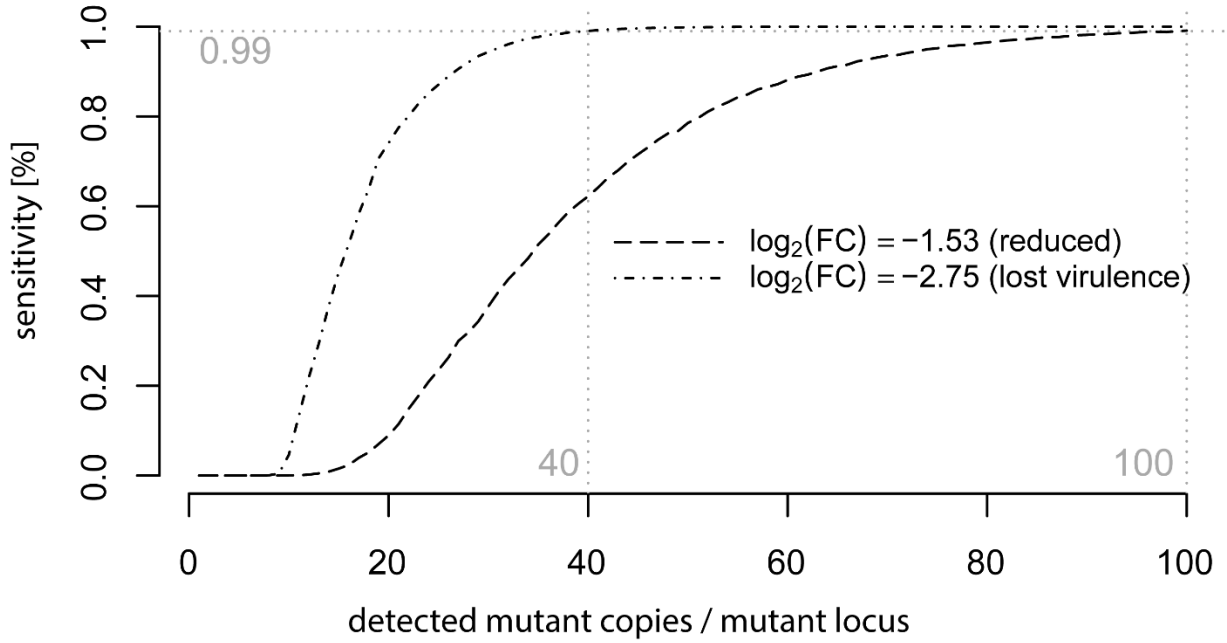
S2 Fig. Tn5 fragmentation of gDNA with modified adapters. Recombinantly produced hyperactive Tn5 was tested with standard Tn5-ME-Adapter (Tn5-ME-A) and custom UMI-ME-adapter (UMI-ME-A) on 1µg gDNA of *U. maydis* infected maize tissue with indicated concentrations. M = Marker 1 kb-ladder (Thermo Scientific); In = Input; ME = mosaic end.

S2 Table. Key primers used in this study

Name	Sequence
Tn5ME-A	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′
Tn5ME-rev	5′-[phos]CTGTCTCTTATACACATC[3InvdT]-3′
UMI-ME-A	5′-CACGACGCTCTTCCGATCT**NNNNNNNNNNNN**AGATGTGTATAAGAGACAG-3′
PCR1-Bio-rev	5′-[BioTEG]CCAGATGTCCTGTGGTATCCTGTG-3′
PCR1-A	5′-GAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC-3′
PCR2-P5	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACAC-3′
PCR2-P3Indexrev	5′-CAAGCAGAAGACGGCATACGAGATNNNNNNCTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTCCTGTGGTATCCTGTGGCG-3′
Seq-Fw	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′

Comment	Reference
Tn5 Mosaic end underlined	Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24, 2033-2040 (2014)
Modified 3InvdT to 3′-End to prevent elongation	Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24, 2033-2040 (2014)
Tn5 Mosaic end underlined, 12N-UMI in bold	this study
Biotinylated with TEG-linker	this study
Adapter specific	this study
Illumina P5 sequence	this study
Illumina P3 sequence, Multiplex Index Bar	this study
Custom Primer for Read 1 Sequencing	PE Read 1 Sequencing Primer (Illumina, San Diego, California)

S3 Figure. Sensitivity of iPool-Seq



S3 Fig. Sensitivity of iPool-Seq. Estimated sensitivity of iPool-Seq for a genome wide library of U. maydis mutants. Model shows for different (1 up to 100) mutant copies detected in the input sample the sensitivity of virulence factor detection. Depicted model curves are given assuming 3% of all mutants have a reduced virulence of log2(FC) -1.53 respectively log2(FC) of -2,75, and the other 97% are neutral in respect to virulence. The sensitivity reaches 99% at 40 detected mutants (lost virulence) and 100 detected mutants (reduced virulence), respectively.

S3 Table. *U. maydis* mutants used for the internal reference set

A) Neutral Reference set

Gene ID	Cluster	Reference
UMAG 02193	Cluster 5A	Kamper J. A PCR-based system for highly efficient generation of gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004;271(1):103-10. doi: 10.1007/s00438-003-0962-8. PubMed PMID: 14673645.
UMAG 03202	Cluster 8A	Kamper J. A PCR-based system for highly efficient generation of gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004;271(1):103-10. doi: 10.1007/s00438-003-0962-8. PubMed PMID: 14673645.
UMAG 01302	Cluster 2B	Kamper J. A PCR-based system for highly efficient generation of gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004;271(1):103-10. doi: 10.1007/s00438-003-0962-8. PubMed PMID: 14673645.
UMAG 10403	Cluster 8A	Kamper J. A PCR-based system for highly efficient generation of gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004;271(1):103-10. doi: 10.1007/s00438-003-0962-8. PubMed PMID: 14673645.
UMAG 10553	Cluster 19A	Schilling L, Matei A, Redkar A, Walbot V, Doehlemann G. Virulence of the maize smut Ustilago maydis is shaped by organ-specific effectors. Mol Plant Pathol. 2014;15(8):780-9. doi: 10.1111/mpp.12133. PubMed PMID: WOS:000342131900003.

B) Reduced reference set

Gene ID	Cluster	Reference
UMAG 00105	NA	Eichhorn H, Lessing F, Winterberg B, Schirawski J, Kamper J, Muller P, et al. A ferroxidation/permeation iron uptake system is required for virulence in Ustilago maydis. Plant Cell. 2006;18(11):3332-45. doi: 10.1105/tpc.106.043588. PubMed PMID: WOS:000243093700035.
UMAG 02011	NA	Stirnberg A, Djamei A. Characterization of ApB73, a virulence factor important for colonization of Zea mays by the smut Ustilago maydis. Mol Plant Pathol. 2016;17(9):1467-79. doi: 10.1111/mpp.12442. PubMed PMID: WOS:000389134900013.

UMAG 05302	Cluster 19A	Kamper J. A PCR-based system for highly efficient generation of gene replacement mutants in Ustilago maydis. Mol Genet Genomics. 2004;271(1):103-10. doi: 10.1007/s00438-003-0962-8. PubMed PMID: 14673645.
UMAG 12226	NA	This study

C) Lost virulence reference set

Gene ID	**Cluster**	**Reference**
UMAG 01375	NA	Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the Ustilago maydis-Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. Plos Pathog. 2013;9(2). doi: ARTN e100317710.1371/journal.ppat.1003177. PubMed PMID: WOS:000315648900027.
UMAG 01987	NA	Doehlemann G, van der Linde K, Amann D, Schwammbach D, Hof A, Mohanty A, et al. Pep1, a Secreted Effector Protein of Ustilago maydis, Is Required for Successful Invasion of Plant Cells. Plos Pathog. 2009;5(2). doi: ARTN e100029010.1371/journal.ppat.1000290. PubMed PMID: WOS:000263928000034.
UMAG 02475	Cluster 5B	Schipper K, Brefort T, Doehlemann G, Djamei A, Muench K, Kahmann R. The secreted protein Stp1 is cruical for establishment of the biotrophic interaction of the smut fungus Ustilago maydis with its host plant maize. Eur J Cell Biol. 2008;87:29-. PubMed PMID: WOS:000255316100068.

S4 Table. Significantly depleted *U. maydis* mutants identified by iPool-Seq

S4 Table. Significantly depleted *U. maydis* mutants identified by iPool-Seq

Green	Significantly depleted mutants found in this study.
Yellow	Significantly depleted mutants found in this study, which are known to cause reduced symptoms on maize.
Red	Significantly depleted mutants found in this study, which are known to be apathogenic on maize.

No.	Mutant ID	log2fc of experiment A	No. of replicates with significant depletion - experiment A	log2fc of experiment B	No. of replicates with significant depletion - experiment B	Average of log2fc	Average of significant replicates	Known phenotype of mutant	Reference	
1	UMAG_01689	-2,08	3	-2,25	3	-2,16	3	NA		Insertional mutants with significant depletion.
2	UMAG_01987	-1,78	3	-1,71	3	-1,75	3	apathogenic, lost virulence	1	
3	UMAG_02011	-2,21	3	-1,93	3	-2,07	3	reduced	2	
4	UMAG_02475	-2,70	3	-3,15	3	-2,93	3	apathogenic, lost virulence	3	
5	UMAG_02560	-3,90	3	-0,74	3	-2,32	3	NA		
6	UMAG_06440	-0,62	3	-0,47	3	-0,54	3	NA		
7	UMAG_11402	-0,79	3	-0,72	3	-0,75	3	NA		
8	UMAG_11940	-1,59	3	-0,95	3	-1,27	3	NA		
9	UMAG_12045	-1,77	3	-3,89	3	-2,83	3	NA		
10	UMAG_12197	-1,69	3	-3,16	3	-2,43	3	NA		
11	UMAG_12226	-1,09	3	-1,49	3	-1,29	3	NA		
12	UMAG_12281	-1,75	3	-3,50	3	-2,62	3	NA		
13	UMAG_00105	-1,74	3	-2,24	2	-1,99	2,5	reduced	4	
14	UMAG_01375	-2,18	3	-3,69	2	-2,93	2,5	apathogenic, lost virulence	5	
15	UMAG_03105	-1,41	3	-1,12	2	-1,26	2,5	NA		
16	UMAG_03744	-0,66	3	-0,46	2	-0,56	2,5	NA		
17	UMAG_10975	-0,68	3	-0,43	2	-0,56	2,5	NA		
18	UMAG_01236	-0,44	3	-0,28	1	-0,36	2	NA		
19	UMAG_01240	-0,84	1	-1,47	3	-1,16	2	NA		
20	UMAG_01940	-0,40	3	-0,07	1	-0,24	2	NA		
21	UMAG_02535	-0,41	2	-0,97	2	-0,69	2	NA		
22	UMAG_03201	-1,25	2	-0,97	2	-1,11	2	NA		
23	UMAG_05319	-0,51	2	-0,49	2	-0,50	2	NA		
24	UMAG_06064	-0,43	3	-0,19	1	-0,31	2	NA		
25	UMAG_10067	-1,19	2	-2,32	2	-1,76	2	NA		
26	UMAG_12216	-0,05	2	-0,61	2	-0,33	2	NA		
27	UMAG_12233	-0,36	3	-0,54	1	-0,45	2	NA		
28	UMAG_12313	-0,35	2	-0,40	2	-0,38	2	no effect on virulence	6	
29	UMAG_00187	0,13	2	-0,67	1	-0,27	1,5	NA		
30	UMAG_00792	-0,62	2	-0,38	1	-0,50	1,5	NA		
31	UMAG_01082	-0,05	2	-0,29	1	-0,17	1,5	NA		
32	UMAG_01289	-0,60	2	0,00	1	-0,30	1,5	NA		
33	UMAG_01300	-0,55	2	-0,64	1	-0,59	1,5	NA		
34	UMAG_02138	-0,65	2	-0,54	1	-0,60	1,5	NA		
35	UMAG_02813	-0,19	1	-0,52	2	-0,36	1,5	NA		
36	UMAG_02851	-0,21	1	-0,72	2	-0,46	1,5	NA		
37	UMAG_02852	-0,42	2	-0,30	1	-0,36	1,5	NA		
38	UMAG_03023	0,04	1	-0,64	2	-0,30	1,5	NA		
39	UMAG_03138	-0,20	1	-0,24	2	-0,22	1,5	NA		
40	UMAG_04104	0,13	1	-0,36	2	-0,11	1,5	NA		
41	UMAG_04189	-0,31	2	-0,25	1	-0,28	1,5	NA		
42	UMAG_05301	-0,80	2	-1,03	1	-0,91	1,5	NA		
43	UMAG_05439	-0,21	2	-0,07	1	-0,14	1,5	seedling, but reduced in	6	
44	UMAG_05562	-0,51	2	-0,33	1	-0,42	1,5	NA		
45	UMAG_05819	-0,44	3	0,01	0	-0,21	1,5	NA		
46	UMAG_05926	-0,11	1	-0,40	2	-0,26	1,5	NA		
47	UMAG_06113	-0,61	2	-0,13	1	-0,37	1,5	NA		
48	UMAG_06178	-0,62	2	0,01	1	-0,31	1,5	NA		
49	UMAG_06179	-0,21	1	-0,49	2	-0,35	1,5	NA		
50	UMAG_11094	-0,03	2	-0,51	1	-0,27	1,5	NA		
51	UMAG_11250	-0,06	2	-0,32	1	-0,19	1,5	NA		
52	UMAG_11415	0,14	1	-0,42	2	-0,14	1,5	NA		
53	UMAG_00558	-0,92	1	-0,27	1	-0,59	1	NA		
54	UMAG_00781	-0,10	1	-0,21	1	-0,15	1	NA		
55	UMAG_00823	-1,12	2	0,54	0	-0,29	1	NA		
56	UMAG_01297	-0,38	1	-0,04	1	-0,21	1	NA		
57	UMAG_01301	-0,26	1	-0,28	1	-0,27	1	NA		
58	UMAG_01820	-0,59	1	0,08	1	-0,25	1	NA		
59	UMAG_01997	-0,04	1	-0,23	1	-0,13	1	NA		
60	UMAG_02141	-1,96	2	0,37	0	-0,80	1	NA		
61	UMAG_02229	-0,28	2	0,15	0	-0,07	1	NA		
62	UMAG_02239	-0,17	1	-0,26	1	-0,21	1	reduced	6	

63	UMAG_02298	-0,20	1	-0,21	1	-0,20	1	NA	
64	UMAG_02299	0,72	1	-1,63	1	-0,45	1	NA	
65	UMAG_02473	-0,71	1	0,07	1	-0,32	1	NA	
66	UMAG_02474	-0,11	1	-0,65	1	-0,38	1	NA	
67	UMAG_02533	-0,10	1	-0,25	1	-0,18	1	NA	
68	UMAG_02537	-0,21	2	-0,14	0	-0,17	1	NA	
69	UMAG_02611	0,07	1	-0,56	1	-0,25	1	NA	
70	UMAG_02853	-0,78	2	0,40	0	-0,19	1	NA	
71	UMAG_03223	-0,36	1	-0,10	1	-0,23	1	NA	
72	UMAG_03313	-1,04	1	-0,24	1	-0,64	1	NA	
73	UMAG_03586	-0,18	1	-0,23	1	-0,20	1	NA	
74	UMAG_03753	0,01	1	-0,11	1	-0,05	1	NA	
75	UMAG_03880	-0,18	1	-0,09	1	-0,14	1	NA	
76	UMAG_04038	-0,25	2	-0,08	0	-0,17	1	NA	
77	UMAG_04039	0,18	1	0,20	1	0,19	1	NA	
78	UMAG_04057	-0,23	1	-0,22	1	-0,22	1	NA	
79	UMAG_04084	-0,19	1	-0,24	1	-0,21	1	NA	
80	UMAG_04111	0,36	0	-0,69	2	-0,16	1	NA	
81	UMAG_04400	0,19	0	-0,33	2	-0,07	1	NA	
82	UMAG_04696	0,44	0	-0,22	2	0,11	1	NA	
83	UMAG_05294	-0,30	2	0,12	0	-0,09	1	NA	
84	UMAG_05299	-1,77	2	0,50	0	-0,63	1	NA	
85	UMAG_05310	0,45	1	-0,35	1	0,05	1	NA	
86	UMAG_05548	-0,02	1	-0,27	1	-0,14	1	NA	
87	UMAG_05780	0,09	1	-0,28	1	-0,10	1	NA	
88	UMAG_05781	-0,01	1	-0,52	1	-0,27	1	NA	
89	UMAG_06146	-0,13	1	-0,19	1	-0,16	1	NA	
90	UMAG_06158	-0,50	2	0,07	0	-0,21	1	NA	
91	UMAG_06222	0,08	1	-0,54	1	-0,23	1	no effect on virulence	6
92	UMAG_06223	-0,55	2	0,14	0	-0,21	1	reduced	6
93	UMAG_06428	-0,47	2	-0,10	0	-0,28	1	NA	
94	UMAG_10024	-0,30	2	0,13	0	-0,08	1	NA	
95	UMAG_10555	-0,23	1	-0,21	1	-0,22	1	NA	
96	UMAG_10640	-0,14	1	-0,08	1	-0,11	1	NA	
97	UMAG_10881	-0,50	1	-1,49	1	-0,99	1	NA	
98	UMAG_11193	-0,21	1	-0,63	1	-0,42	1	NA	
99	UMAG_11362	-0,63	2	0,06	0	-0,28	1	NA	
100	UMAG_11403	-0,05	0	-0,94	2	-0,49	1	NA	
101	UMAG_11417	0,24	0	-0,34	2	-0,05	1	NA	
102	UMAG_11444	-0,28	1	-0,55	1	-0,41	1	NA	
103	UMAG_12302	0,01	1	-0,37	1	-0,18	1	NA	
104	UMAG_12330	0,08	0	-0,29	2	-0,11	1	NA	
105	UMAG_15020	0,28	1	-1,08	1	-0,40	1	NA	
106	UMAG_00054	-0,31	1	-0,03	0	-0,17	0,5	NA	
107	UMAG_00081	0,19	1	-0,17	0	0,01	0,5	NA	
108	UMAG_00159	-0,09	1	0,76	0	0,33	0,5	NA	
109	UMAG_00793	-0,06	0	-0,09	1	-0,08	0,5	NA	
110	UMAG_00885	0,20	1	-0,12	0	0,04	0,5	NA	
111	UMAG_01018	-0,02	0	-0,23	1	-0,12	0,5	NA	
112	UMAG_01235	0,02	0	0,09	1	0,06	0,5	NA	
113	UMAG_01237	-0,35	1	0,25	0	-0,05	0,5	NA	
114	UMAG_01238	-0,28	1	0,22	0	-0,03	0,5	NA	
115	UMAG_01239	-0,14	1	0,45	0	0,15	0,5	NA	
116	UMAG_01553	-1,64	1	0,22	0	-0,71	0,5	NA	
117	UMAG_01779	-0,63	1	-0,38	0	-0,51	0,5	NA	
118	UMAG_01858	0,40	0	-0,09	1	0,16	0,5	NA	
119	UMAG_02135	0,11	0	-0,24	1	-0,06	0,5	NA	
120	UMAG_02137	-0,65	1	-0,31	0	-0,48	0,5	NA	
121	UMAG_02193	0,01	0	-0,23	1	-0,11	0,5	NA	
122	UMAG_02243	0,13	0	-0,26	1	-0,07	0,5	NA	
123	UMAG_02294	0,16	0	-0,31	1	-0,08	0,5	NA	
124	UMAG_02295	-0,03	1	0,13	0	0,05	0,5	NA	
125	UMAG_02297	-0,42	1	0,34	0	-0,04	0,5	NA	
126	UMAG_02430	0,14	0	-0,03	1	0,05	0,5	NA	
127	UMAG_02826	0,04	1	0,15	0	0,09	0,5	NA	
128	UMAG_03112	0,30	1	-0,21	0	0,04	0,5	NA	
129	UMAG_03382	0,09	0	0,04	1	0,06	0,5	NA	
130	UMAG_03615	-0,03	1	-0,20	0	-0,11	0,5	NA	
131	UMAG_03650	0,22	0	-0,23	1	0,00	0,5	seedling, but reduced in	6
132	UMAG_03747	-0,19	1	-0,07	0	-0,13	0,5	NA	
133	UMAG_03748	-0,17	1	-0,07	0	-0,12	0,5	NA	
134	UMAG_03750	0,46	0	0,12	1	0,29	0,5	NA	
135	UMAG_03751	-0,04	0	-0,41	1	-0,22	0,5	NA	
136	UMAG_04096	-0,56	1	NA	0	-0,56	0,5	NA	
137	UMAG_04114	-0,09	0	-0,15	1	-0,12	0,5	NA	

Insertional mutants with unreproducable significant depletion.

138	UMAG_04145	0,10	0	-0,16	1	-0,03	0,5	NA	
139	UMAG_04185	-0,07	1	0,18	0	0,05	0,5	NA	
140	UMAG_04282	0,50	0	0,05	1	0,27	0,5	NA	
141	UMAG_04815	0,19	1	-0,13	0	0,03	0,5	NA	
142	UMAG_05302	-0,16	1	-0,18	0	-0,17	0,5	NA	
143	UMAG_05308	-0,15	1	0,26	0	0,06	0,5	NA	
144	UMAG_05641	0,15	1	-0,12	0	0,01	0,5	NA	
145	UMAG_05731	NA	0	-0,15	1	-0,15	0,5	NA	
146	UMAG_05733	0,31	0	-0,06	1	0,12	0,5	NA	
147	UMAG_05861	0,29	1	-0,09	0	0,10	0,5	NA	
148	UMAG_05931	0,12	0	-0,14	1	-0,01	0,5	NA	
149	UMAG_05953	-0,18	0	-0,03	1	-0,10	0,5	NA	
150	UMAG_10030	-0,15	1	-0,11	0	-0,13	0,5	NA	
151	UMAG_10076	-0,24	1	0,15	0	-0,05	0,5	NA	
152	UMAG_10557	0,07	1	0,07	0	0,07	0,5	NA	
153	UMAG_10816	0,01	1	0,34	0	0,18	0,5	NA	
154	UMAG_10972	-0,03	0	0,34	1	0,15	0,5	NA	
155	UMAG_11062	0,11	0	-0,14	1	-0,02	0,5	NA	
156	UMAG_11305	-0,17	1	0,11	0	-0,03	0,5	NA	
157	UMAG_11377	0,12	0	0,00	1	0,06	0,5	NA	
158	UMAG_11416	-0,23	1	-0,20	0	-0,21	0,5	NA	
159	UMAG_11586	-0,07	1	0,02	0	-0,03	0,5	NA	
160	UMAG_11639	0,25	0	0,13	1	0,19	0,5	NA	

References:

1. Doehlemann G, van der Linde K, Amann D, Schwammbach D, Hof A, Mohanty A, et al. Pep1, a Secreted Effector Protein of Ustilago maydis, Is Required for Successful Invasion of Plant Cells. Plos Pathog. 2009;5(2). doi: ARTN e1000290 10.1371/journal.ppat.1000290. PubMed PMID: WOS:000263928000034.

2. Stirnberg A, Djamei A. Characterization of ApB73, a virulence factor important for colonization of Zea mays by the smut Ustilago maydis. Mol Plant Pathol. 2016;17(9):1467-79. doi: 10.1111/mpp.12442. PubMed PMID: WOS:000389134900013.

3. Schipper K, Brefort T, Doehlemann G, Djamei A, Muench K, Kahmann R. The secreted protein Stp1 is cruical for establishment of the biotrophic interaction of the smut fungus Ustilago maydis with its host plant maize. Eur J Cell Biol. 2008;87:29-. PubMed PMID: WOS:000255316100068.

4. Eichhorn H, Lessing F, Winterberg B, Schirawski J, Kamper J, Muller P, et al. A ferroxidation/permeation iron uptake system is required for virulence in Ustilago maydis. Plant Cell. 2006;18(11):3332-45. doi: 10.1105/tpc.106.043588. PubMed PMID: WOS:000243093700035.

5. Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the Ustilago maydis-Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. Plos Pathog. 2013;9(2). doi: ARTN e100317710.1371/journal.ppat.1003177. PubMed PMID: WOS:000315648900027.

6. Schilling L, Matei A, Redkar A, Walbot V, Doehlemann G. Virulence of the maize smut Ustilago maydis is shaped by organ-specific effectors. Mol Plant Pathol. 2014;15(8):780-9. doi: 10.1111/mpp.12133. PubMed PMID: WOS:000342131900003.