Nuclear Localization of Proteins and Genome Editing in the Oomycete *Phytophthora sojae*

Yufeng Fang

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Genetics, Bioinformatics and Computational Biology

Brett M. Tyler, Committee Chair
Michael Freitag
Christopher B. Lawrence
John M. McDowell

October, 26, 2016
Blacksburg, VA

Keywords: Oomycetes, *Phytophthora sojae*, nuclear localization signals, CRISPR/Cas9, genome editing

Nuclear Localization of Proteins and Genome Editing in The Oomycete *Phytophthora sojae*

Yufeng Fang

Abstract (academic)

Oomycetes are fungi-like eukaryotic microorganisms, which are actually phylogenetic relatives of diatoms and brown algae, within the kingdom Stramenopila. Many oomycete species, mainly in the genera *Phytophthora*, *Pythium* and downy mildews, are devastating plant pathogens that cause multibillion-dollar losses to agriculture annually in the world. Some oomycetes are also animal pathogens, causing severe losses in aquaculture and fisheries, and occasionally causing dangerous infections of humans. *Phytophthora* species, represented by the Irish Potato Famine pathogen *P. infestans* and the soybean pathogen *P. sojae*, are arguably the most destructive pathogens of dicotyledonous plants among the oomycete species and thus have been extensively studied. This dissertation focuses on the model oomycete pathogen *P. sojae* to investigate specific aspects of its molecular biology and establish an efficient genetic manipulation tool.

Specifically, in Chapter 1, I briefly introduce the basic concepts of oomycete biology and pathology, and summarize the experimental techniques used for studies of oomycete genetics over the past two decades. Because the approach to studying fungi and oomycetes are similar (indeed they were incorrectly placed in the same taxonomic group until recently), a special section reviews the emerging genome editing technology CRISPR/Cas system in these organisms together.

Chapter 2 and Chapter 3 focus on one of the most important intracellular activities, nuclear localization of proteins, and describe the characterization of nuclear localization signals (NLSs) in *P. sojae*. This focus stemmed from my early work on genome editing in *P. sojae*, when I discovered that conventional NLS signals from SV40 used to target the TAL effector nuclease (TALEN) to the nucleus worked poorly in *P. sojae*. In the first part of this work (Chapter 2), I used confocal microscopy to identify features of nuclear localization in oomycetes that differ from animals, plants and fungi, based on characterization of two classes of nuclear localization signals, cNLS and PY-NLS, and on characterization of several conserved nuclear proteins. In the second part (Chapter 3), I determined that the nuclear localization of the *P. sojae* bZIP1 transcription factor is mediated by multiple weak nuclear targeting motifs acting together.

In Chapter 4 and Chapter 5, I describe my implementation of nuclease-based technology for genetic modification and control of *P. sojae*. In Chapter 4, I describe the first use of the CRISPR system in an oomycete, including its use to validate the function of a host specificity gene. This is of particular importance because molecular techniques such as gene knockouts and gene replacements, widely used in other organisms, were not previously possible in oomycetes. The successful implementation of CRISPR provides a major new research capability to the oomycete community. Following up on the studies described in Chapter 4, in Chapter 5, I describe the generalization and simplification of the CRISPR/Cas9 expression strategy in *P. sojae* as well as methods for mutant screening. I also describe several optimized methodologies for *P. sojae* manipulation based on my 5 years of experience with *P. sojae*.

Nuclear Localization of Proteins and Genome Editing in The Oomycete *Phytophthora sojae*

Yufeng Fang

Abstract (public)

Oomycetes (water molds) are eukaryotic microorganisms that resemble filamentous fungi (molds), but are actually relatives of diatoms and brown algae, within a different kingdom of life named Stramenopila. The functional relationship between oomycetes and fungi is similar to that between fish and dolphins, which also acquired similar functions via different evolutionary paths. Many families of oomycetes are devastating plant pathogens that cause multibillion-dollar losses to agriculture annually in the world. Other families of oomycetes are animal pathogens, causing severe losses in aquaculture and fisheries, and occasionally causing dangerous infections of humans. *Phytophthora* species, represented by the Irish Potato Famine pathogen *P. infestans* and the soybean pathogen *P. sojae*, are among the most destructive oomycete pathogens of plants and thus have been extensively studied. This dissertation is focused on the model oomycete pathogen *P. sojae*. It investigates specific aspects of its molecular biology and establishes an efficient genetic manipulation tool. All complex organisms (eukaryotes) package their genetic material in nuclei, which contain proteins as well as DNA. In the first part of my research (Chapter 2 and Chapter 3), I focused on the mechanisms used by *P. sojae* to target nuclear proteins into the nucleus, particularly the tags (called nuclear localization signals, or NLSs) that are identify the proteins that must travel to the nucleus. I showed that nuclear targeting mechanisms in oomycetes differ in distinct ways from well-studied

eukaryotes such as humans. In particular, the nuclear targeting signals in *P. sojae* proteins are diffused over multiple sites on the proteins, whereas in human proteins there's usually just a single signal. For one particular oomycete protein, a transcription factor, nuclear targeting involves four weak signals that cooperate synergistically. Two of these four weak signals define a new class of nuclear localization signal. In the second part of my research (Chapter 4 and Chapter 5), I implemented and further optimized a genome editing technology for genetic modification and control of *P. sojae*. This technology is based on the CRISPR system that has revolutionized genome editing in plants and animals over the last three years. This is of particular importance because genome editing techniques were not previously possible in oomycetes. The successful implementation of CRISPR technology in *P. sojae* has provided a major new research capability to the oomycete community. In Chapter 5, I also describe several optimized methodologies for *P. sojae* genetic manipulation based on my 5 years of experience with *P. sojae*.

To my parents Fāng Yuè Wén and Wǎng Guì Lán, my wife Zhào Yīng and my little son Jesse Ying-Xin Fang

Where there's a will, there's a way...

ACKNOWLEDGMENTS

This dissertation would have not been possible without the help, support and guidance of my committee, friends, family and many other people. It is with great pleasure and gratitude that I acknowledge their efforts.

I would like to thank my committee chair and my advisor Dr. Brett M. Tyler for the extraordinary supports during my entire Ph.D. study. In the past six and half years, he has been not only an adviser for training me how to be a good scientist, but also a great helper for personal matters. I thank him for providing me many opportunities to attend high-level conferences which broaden my perspective and benefit my career. I respect him for his passion in science, his knowledge as well as his research altitude. His intellectual insights on fungal and oomycete genetics and broad knowledge on life sciences have been major driving forces throughout my doctoral training. I also would like to thank my committee member Dr. John M. McDowell, who has been supporting me since I started my first Ph.D. project with his intellectual richness. He generously shared his lab with me during my transition from Virginia Tech to Oregon State University, and has been a great onsite helper at Virginia Tech. I thank my committee member Dr. Christopher B. Lawrence, who has been generously sharing his experience on scientific research and offered me priceless advice on future careers of fungal genetics. I thank my committee member Michael Freitag, who has been providing useful discussion on my research and future career, and also provided me opportunities to talk with many intelligent fungal geneticists.

I would like to thank the Genetics, Bioinformatics and Computational Biology (GBCB) program and thank Dr. David Bevan, and Ms. Dennie Munson, who have being tremendous help for me in this program, from application to graduation, and also provided me graduate fellowship.

I thank all members from the Tyler research group for helpful discussions and accompany. In particular, I thank Dulani Wellappili for research assistance; Mr. Felipe Arredondo, Drs. Linkai Cui and Biao Gu for collaboration on the CIRSPR/Cas9 protocol; Dr. Stephanie R. Bollmann for collaboration on the *P. sojae* RNAi machinery project; Drs. Brent Kronmiller and Danyu Shen for assistance on bioinformatics analysis. I thank all the colleagues in the Center for Genome Research and Biocomputing (CGRB) at Oregon State University for providing me great help and services throughout these years. In particular, Mrs. Rosa Hill who have been providing me tremendous help in my life.

I thank my friends Drs. Tian Hong, Gregory Watson, Hyo Sang Jang and Pinyi Lu for their assistance and advice on my research and life.

Last but not least, I would like to express my deepest love, gratitude and appreciations to my family, my parents Fang Yue Wen and Wang Gui Lian, and my wife Zhao Ying for their unconditional love and endless support.

ATTRIBUTIONS

Most of the chapters in this dissertation are from my manuscripts that will be or have been published.

Chapter 2 is based on my work "Distinctive nuclear localization signals in the oomycete *Phytophthora sojae*", which has been submitted to Molecular Microbiology. Drs. Hyo Sang Jang (a postdoc in Department of Environmental & Molecular Toxicology at Oregon State University) and Gregory W. Watson (who was a Ph.D. student in Molecular and Cellular Biology Program and School of Biological and Population Health Sciences at Oregon State University, and currently is postdoctoral fellow in Moffitt Cancer Center, Florida) contributed to human cell transfection. Dr. Gregory W. Watson also helped to edit the manuscript. Miss Dulani P. Wellapilli (an undergraduate majored in Microbiology and an undergraduate assistant at the Tyler Lab at Oregon State University) contributed to technical assistance. Dr. Brett M. Tyler (a professor in the Center for Genome Research and Biocomputing and Department of Botany and Plant Pathology at Oregon State University, and an adjunct professor in the Interdisciplinary Ph.D. program in Genetics, Bioinformatics & Computational Biology and the Department of Plant Physiology, Pathology and Weed Science at Virginia Tech) helped to analyze the data and edited the manuscript.

Chapter 3 includes my work "Nuclear localization of a putative *Phytophthora sojae* bZIP1 transcription factor is mediated by multiple targeting motifs". This paper has been submitted to Molecular Microbiology, accompanying with the manuscript involved in Chapter 2. Dr. Brett M. Tyler is the only coauthor who contributed to analyzing the data and editing the manuscript.

Chapter 4 is from my work “Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9”, which has been published in 2016 in Molecular Plant Pathology 17.1 (2016): 127-139. Dr. Brett M. Tyler is the only coauthor who contributed to conceiving the study, analyzing the data and manuscript writing.

Chapter 5 is based on my work “Efficient genome editing in the oomycete *Phytophthora sojae* using CRISPR/Cas9”. This paper was invited by Current Protocols in Microbiology, and now it is under review. Drs. Linkai Cui (a visiting scholar at the Tyler Lab at Oregon State University) and Biao Gu (a visiting scholar at the Tyler Lab at Oregon State University) contributed to construction of the generalized CRISPR/Cas9 plasmid and its quality control assays. Mr. Felipe Arredondo (a lab manager at the Tyler Lab at Oregon State University) contributed to organizing and providing methods related to *Phytophthora* manipulation. Dr. Brett M. Tyler contributed to manuscript revision.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Oomycete biology and pathology

Oomycetes are infamous but newly recognized microorganisms. The oomycete species, *Phytophthora infestans* − notorious for the Great Irish Famine of the mid 1840s− was the first microorganism proven to be responsible for disease (Large, 1940). They are also newly recognized, because for decades, oomycete species were defined as "phycomycetes having oospores" and were classified as oomycota within the fungi kingdom (Lévesque, 2011). Since the early 2000s however, the classification of oomycetes has been demonstrated to be independent from fungi and they were grouped into a different kingdom Stramenopila (also called Stramenipila, Stramenopile or Stramenopiles) (Dick, 2001).

Physiologically and morphologically, oomycetes closely resemble filamentous fungi. This is the reason that these species were traditionally classified in the fungal kingdom. With the development of molecular techniques, such as 18S rDNA sequencing and whole genome sequencing, phylogenetic analyses of modern taxa have indicated that the lineages of fungi and oomycetes diverged before fungi split from plants and animals (Raffaele & Kamoun, 2012). Some earlier studies also found significant differences between fungi and oomycetes (reviewed in Tyler, 2001; Judelson & Blanco, 2005; Raffaele & Kamoun, 2012). Oomycetes share the kingdom Stramenopila with brown algae and diatoms, and are thought to have evolved from phototrophic ancestors (Cavalier-Smith, 1986; Gunderson *et al.*, 1987; Dick, 2001; Tyler *et al.*, 2006). For instance, oomycete cell walls mainly consist of β-glucan and cellulose rather than β-glucan and chitin as in the fungi. Oomycete hyphae are rarely septate, unlike those of higher

fungi. Oomycetes are always diploid and lack a free haploid stage, while fungi are often haploid or dikaryotic.

Several genera of oomycetes are known as notorious plant destroyers, which mainly occur within the class Peronosporomycetidae, in the orders Peronosporales (*Phytophthora* species and downy mildews), Pythiales (*Pythium* species), and Albuginales (*Albugo* and other white rusts) (Jiang & Tyler, 2012). The most destructive plant pathogens are members of the genus *Phytophthora*, which cause a wide range of diseases, affecting agricultural, ornamental, and natural ecosystems (Erwin *et al.*, 1983; Erwin & Ribeiro, 1996). It has been estimated that *Phytophthora* species causes multibillion dollar losses to crops in the United States annually (Erwin & Ribeiro, 1996). To date, over 140 species have been reported (Érsek & Ribeiro, 2010; Kroon *et al.*, 2012) and have been classified into 10 clades based their morphological and physiological characters as well as molecular markers (Blair *et al.*, 2008). The potato and tomato pathogen *P. infestans* and soybean pathogen *P. sojae* are two well-known examples, because of their historical and economic impacts.

Some oomycete species (mostly in the class Saprolegniomycetidae) also can infect animals and humans. For example, *Saprolegnia parasitica* is a reemerging parasite that can form lesions on fish, including catfish, salmon and trout species (van West, 2006). The closely related *Aphanomyces astaci* is most serious pathogen of freshwater crayfish (crayfish plague, Oidtmann *et al.*, 2002). *Lagenidium* species, which are widely known as insect pathogens, were recently reported to infect small animals (lagenidiosis) (Grooters, 2003; Mendoza & Vilela, 2013). While most *Pythium* species are destructive plant pathogens, *Pythium insidiosum* can cause life-threatening infections (pythiosis) to farm animals, pets and humans (Gaastra *et al.*, 2010; De Cock *et al.*, 1987; Mendoza & Vilela, 2013).

1.2 Nuclear localization in eukaryotes, including oomycetes

Oomycetes are eukaryotic microbes. As such, their genetic materials and transcriptional machineries are separated (in the nucleus) from the translational machineries and metabolic systems (in the cytoplasm) (Lange *et al.*, 2007). Nuclear proteins, such as core histones and transcription factors are translated in the cytoplasm but function in the nucleus. Hence, an effective nuclear trafficking system is necessary to maintain the basic physiological function of cells.

Generally, molecules that smaller than 5 kDa can cross the nuclear envelope by passive diffusion, whereas molecules larger than 5 kDa need a specific structure called nuclear pore complexes (NPCs) to support their nucleocytoplasmic trafficking (Lange *et al.*, 2007). The NPC is a large multi-protein complex, which allows passive diffusion of molecules smaller than 40–60 kDa, but requires nucleocytoplasmic transporters (Karyopherins) to promote active transport of molecules larger than 40–60 kDa (Marfori *et al*., 2011; Xu, *et al*., 2010). During nuclear import, karyopherins in the cytoplasm bind specific cargo proteins via recognition of NLSs, promoting translocation of the cargoes into the nucleus (Marfori *et al*., 2011; Xu, et al., 2010; Chook & Suel, 2011). Inside the nucleus, RanGTP binds karyopherin-cargo complexes, allowing the release of cargoes into the nucleus (Marfori *et al*., 2011; Xu, et al., 2010; Chook & Suel, 2011). To date, 11 of the 19 known human karyopherin-βs and 10 of the 14 *S. cerevisiae* karyopherin-βs are found to mediate nuclear import (Chook & Suel, 2011). However, most of them have not been well-characterized. The best characterized transport signal is the classical NLS (cNLS) for nuclear protein import, which consists of either one (monopartite) or two (bipartite) stretches of basic amino acids (Lange *et al.*, 2007). Additional studies found cNLSs have relatively loose consensus sequences, for example, R (K/R)X(K/R) for the monopartite type, exemplified by the

SV40 large T antigen NLS (126-PKKKRKV-132)(Kalderon *et al*. 1984); and $KRX_{10–12}KRRK$ for the bipartite type, exemplified by the nucleoplasmin NLS (155-KRPAATKKAGQAKKKK-170) (Lange *et al.*, 2007). Another well-studied NLS class is the Proline-Tyrosine NLS (PY-NLS), which is associated with Kapβ2-mediated nuclear import (Chook & Suel, 2011; Lee, *et al*., 2006). Based on the crystal structure of the Kapβ2-M9-NLS model, Lee and coworkers summarized the following three features of PY-NLSs, which can be used to predict this NLS class: NLSs are structurally disordered in the free proteins, they have overall basic character, and possess a central hydrophobic or basic motif followed by a C-terminal $R/H/KX_{(2-5)}PY$ consensus sequence (Lee, *et al*., 2006). However, these rules appeared to vary in other organisms (Suel *et al.*, 2008) and different cell types (Mallet & Bachand, 2013).

In fact, however, it is very hard to accurately predict NLS-transport, because of the high diversity and complexity of the transport machinery: (1) One karyopherin type can carry many different cargos. For instance, Kapα/Kapβ1 is thought to account for half of the nuclear import trafficking (Lange *et al.*, 2007). (2) Cargos sharing the same karyopherin type usually have no sequence similarity, suggesting that one karyopherin type can recognize different NLSs (Chook & Suel, 2011). (3) One cargo can be imported by various karyopherin types. For example, human H2B can be imported by Kapβ2, Importin-5, -7, -9, or Importin-α/β (Baake *et al*. 2011; Muhlhausser *et al*., 2001). (4) Only a few substrates have been identified for most Kapβs (Chook & Suel, 2011). (5) Organism variety. For example, a proline-tyrosine nuclear localization signal (PY-NLS) is required for the nuclear import of fission yeast PAB2 but not for human PABPN1 (Mallet *et al.,* 2013). Overall, large sequence diversity among various cargoes has prevented the identification of NLSs for most Kapβs, and it remains extremely difficult to predict NLSs in candidate import substrates.

Many reports have shown that transcription factors are critical for growth, development or pathogenicity of oomycetes (Blanco & Judelson, 2005; Xiang & Judelson, 2010; Gamboa-Melendez *et al.*, 2013; Xiang & Judelson, 2014; Wang *et al.*, 2009; Ye *et al.*, 2013; Zhang, 2012). However, there have been no studies that have characterized the nuclear localization mechanisms of those transcription factors. Establishment of fluorescent protein labeling techniques in oomycetes has provided a strategy to answer those questions through fluorescent tagging of nuclear proteins in oomycete transformants (Ah-Fong & Judelson, 2011). As part of my thesis, I and my coworkers found that several well-known NLSs derived from mammalian and yeast cells, such as classical NLSs (cNLS) and proline tyrosine NLSs (PY-NLS), functioned poorly in the oomycete *P. sojae* (See Chapter 2). By focusing on a *P. sojae* bZIP transcription factor (PsbZIP1), I deciphered that the nuclear translocation the protein was mediated by multiple targeting motifs (See Chapter 3). These data further suggest that in oomycetes, basic eukaryotic functions such as nuclear localization may operate somewhat differently than in models such as yeast and humans.

1.3 Experimental techniques for genetic studies of oomycete pathogens

Despite their notable status in scientific history and economic importance, oomycetes have been chronically understudied at the molecular level, compared to their fungal counterparts. One of the important reasons for this has been the limited availability of genetic tools. Techniques such as DNA transformation and heterologous gene expression developed for ascomycetes and basidiomycetes could not initially be transferred to oomycetes (Judelson, 1997). For instance, fungal promoters never showed effectiveness in oomycetes (Judelson *et al.*, 1991; Judelson *et al.*, 1992). Molecular genetic studies of oomycetes became possible in the early 1990s once

oomycete-specific promoters were cloned (Judelson & Michelmore, 1991) and transformation systems were established using them (Judelson *et al.*, 1991; Judelson *et al.*, 1993) (Fig. 1.1). From then on, various techniques including gene overexpression, and gene silencing (Kamoun *et al.*, 1998; van West *et al.*, 1999; Whisson, *et al*., 2005; Ah-Fong *et al.*, 2008; Wang *et al.*, 2011) became prevalent for functionally testing the roles of genes found in the genome (summarized in Fig. 1.1). In 2006, Lamour and colleagues described a reverse-genetic strategy called targeting induced local lesions in genomes (TILLING) to isolate mutants in *P. sojae* (Lamour et al., 2006) (Fig. 1.1). However, TILLING is very laborious and has not been widely used for oomycete genetic study. Due to the historical and economic impacts of *P. infestans* and *P. sojae*, molecular genetic and genomics studies are most advanced in these two model oomycete pathogens. However with more and more oomycete genome sequences becoming available, studies in other oomycete species have also developed rapidly in the recent years, for example in *P. capsici*.

While gene silencing has facilitated functional analysis of interesting genes, there are many genes that are difficult to silence (Tyler & Gijzen, 2014). In addition, gene silencing is generally incomplete and varies among gene targets and experiments, and may be very laborious. For decades, direct gene editing such as gene deletion could not be achieved in oomycetes. One of the biggest problems has been that insertion of transgenes in oomycetes occurs exclusively by non-homologous end joining, which makes gene replacement by homologous-recombination not feasible (Tyler, 2007; Tyler & Gijzen, 2014). Other strategies, for example based on customized TALEN nucleases, did not prove useful in oomycetes, due to poor gene expression (Fang & Tyler, 2016).

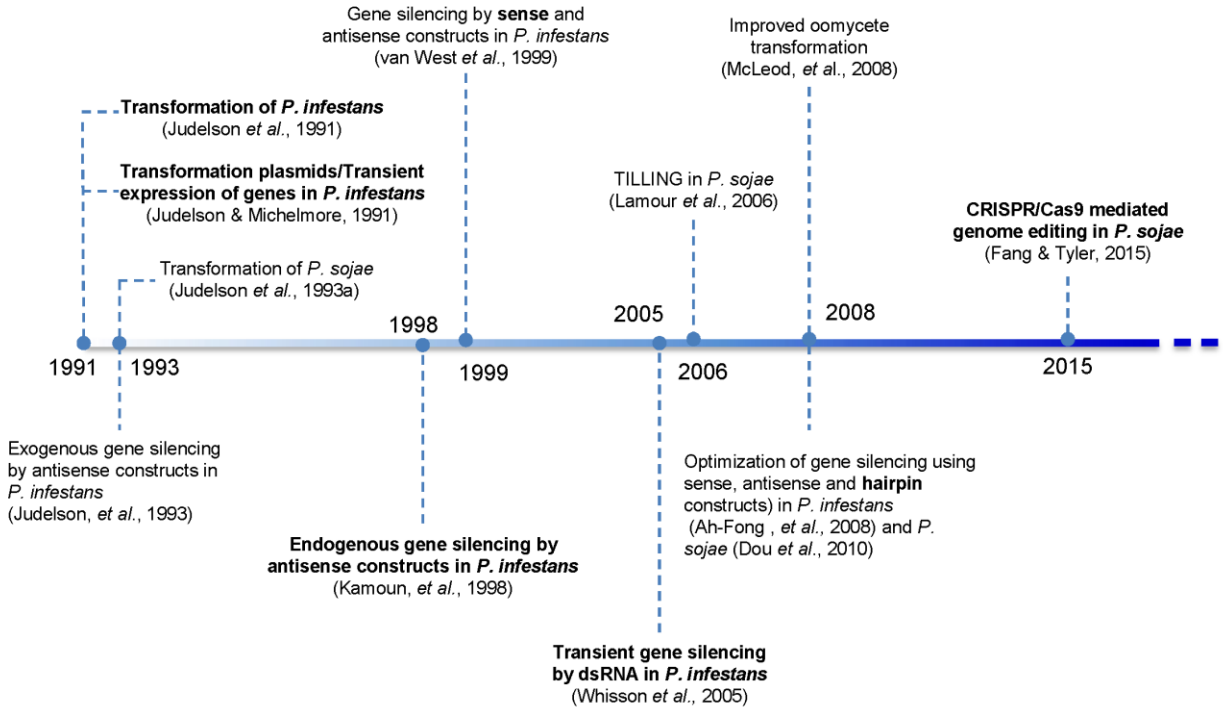


Fig. 1.1 Development of experimental techniques for genetic studies of oomycete pathogens. Key technical innovations in the development are in bold.

Recently however, I successfully applied the newly emerging technique called CRISPR/Cas9 (Clustered regularly interspaced short palindromic repeats and its associated protein, Cas9) to *P. sojae*, enabling genome editing in this pathogen for the first time (Fig. 1.l; Fang & Tyler, 2016)). Using *Avr4/6*, an endogenous gene involved in *P. sojae* infection, as a target, Fang & Tyler (2016) demonstrated that *P. sojae*-optimized CRIPSR/Cas9 could be used to introduce mutations both by non-homologous end-joining (NHEJ) and by homology-directed repair (HDR). Since then this technique has been widely adopted in the oomycete community for studying gene functions. According to a survey that I conducted up, till June 1, 2016, nine researchers in four different labs had succeeded in editing 26 genes. Notably, the *P. sojae* CRISPR/Cas9 system also mediated efficient genome editing in other *Phytophthora* species,

such as *P. capsici*. Therefore, the CRISPR/Cas9 system established in the model oomycete *P. sojae* could be applicable to all culturable oomycete species.

1.4 CRISPR/Cas promotes reverse genetic studies of fungi and oomycetes

The ability to modify specific genetic loci (i.e. reverse genetics) is a powerful tool in basic and applied microbiology. Such targeted gene manipulation (e.g. gene replacement) is classically achieved through the host cell's homologous recombination machinery, the efficiency of which varies greatly between organisms. Among the fungi, for example, gene replacement is readily achieved in *Saccharomyces cerevisiae* with only 50 bp of homologous sequence on either end of a selectable marker; in other species, however, targeting either requires much larger constructions (1 kb of homology) or effectively does not occur at all due to a highly efficient non-homologous end joining (NHEJ) mechanism. Fungi for which powerful gene knockout studies have been lacking include important human pathogens (e.g. *Histoplasma*), animal pathogens (e.g. *Chytrids*) and physiological models (e.g. *Phycomyces*). Furthermore, in many species, a paucity of antibiotic resistance markers, coupled with elevated ploidy, makes knockouts of a single function challenging due to the need for multiple gene disruptions. In the case of oomycetes, as mentioned above, reverse genetics was long blocked due to low rates of homologous recombination.

A major game changer in this regard has been the development of CRISPR/Cas9-based genome editing, which was derived from a system of adaptive immunity in bacteria and archaea (Cong *et al.*, 2013; Mali *et al.*, 2013). The harnessed CRISPR/Cas9 genome editing system contains two components, i.e. the nuclease Cas9 that can make a double-strand DNA break (DSB), and a 20-nucleotide RNA molecule called a single guide RNA (sgRNA) that can guide

Cas9 to a target DNA via Watson-Crick base pairing (Cong *et al*., 2013; Mali *et al*., 2013). By triggering repair of the DSB, the rate of gene editing is increased and can bring about a desired genetic change (Miller *et al.*, 2011). To implement this new technique efficiently in eukaryotic microbes, necessary modifications have been made to adjust to their different intracellular machineries including transcription and translation. Because the approach to studying fungi and oomycetes are similar, here I summarize and compare those optimizations that have been made to apply CRISPR/Cas9 systems successfully in these organisms together.

To date, almost all the CRISPR/Cas9 mediated genome editing systems in fungi and oomycetes have been based on the Cas9 derived from *Streptococcus pyogenes* (in the following context, Cas9 means *Sp*Cas9). Basically, these modifications include three aspects: efficient expression of the nuclease Cas9, transcription of the sgRNA and delivery of Cas9/sgRNA (summarized in Table 1.1).

1.4.1 Efficient expression of Cas9

To efficiently translate Cas9, species-specific codon-optimized Cas9's have been widely used (Nødvig *et al.*, 2015; Weber *et al.*, 2016; Katayama *et al.*, 2015; Arazoe *et al.*, 2015; Liu *et al.*, 2015; Schwartz *et al.*, 2016; Schuster *et al.*, 2015; Vyas *et al.*, 2015), but the human codon-optimized (Fuller *et al.*, 2015; Zhang *et al.*, 2016; Matsu-ura *et al.*, 2015; DiCarlo *et al.*, 2013; Jacobs *et al.*, 2014; Wang *et al.*, 2016; Fang & Tyler, 2016) or plant codon-optimized (Pohl *et al.*, 2016) Cas9's work as well in many species to produce efficient gene cleavage. For example, human- and fungal-codon optimized Cas9's mediated gene mutation equally in *Aspergillus fumigatus* (Fuller *et al.*, 2015; Zhang *et al.*, 2016). Ryan and colleagues even used a non-optimized Cas9 codon from *S. pyogenes* (Ryan *et al.*, 2014). One exception is *Trichoderma*

reesei, in which human codon-optimized Cas9 did not show function (Liu *et al.*, 2015), albeit the author did not mention which version of human codon optimized Cas9 was used. Another exception is *C. albicans*, where species-specific codon optimization is necessary, because the leucine CUG codon in *C. albicans* is predominantly translated as serine (Vyas *et al.*, 2015).

Meanwhile, to regulate gene expression conditionally and to lower the off-target rate, various inducible promoters also have been used for transcription of Cas9 (Zhang *et al.*, 2016; Weber *et al.*, 2016; Pohl *et al.*, 2016). Inducible promoters also have been used to examine or limit the potential toxicity of CRISPR-Cas activity in *Schizosaccharomyces pombe* (Jacobs *et al*., 2014) and *Saccharomyces cerevisiae* (DiCarlo *et al*., 2013).

In addition, because the transcriptional machinery is separated from the translational machinery in eukaryotic cells, strong nuclear localization signals (NLSs) are necessary to translocate Cas9 into the nucleus. In fungal cells, classical NLSs from SV40 large T- antigen and nucleoplasmin (NPL) are usually used together to transport Cas9 into the nucleus, while an oomycete-derived NLS was reported to be more efficient *in P. sojae* (Fang & Tyler, 2016).

1.4.2 Efficient transcription of sgRNA

The abundance of sgRNA has been reported to correlate with the efficiency of Cas9-mediated genome engineering (Hsu *et al.*, 2013). However, the transcription of sgRNA in fungi and oomycetes presents many challenges, mainly because the DNA-dependent RNA polymerase III (Pol III) promoters, such as the commonly used U6 small nuclear RNA (snRNA) promoters, have not been well-defined in those organisms. Pol III promoters have been reported to retain effectiveness across different species within the Ascomycota, for example, the *S. cerevisiae* pol

III *SNR52* promoter also showed activity in *A. fumigatus* (Fuller *et al.*, 2015), *Candida albicans* (Vyas, *et al*., 2015) and *Neurospora crassa* (Matsu-ura *et al.*, 2015). However, Pol III promoters are generally not compatible across phyla. For instance, the human U6 promoter and the yeast *SNR52* promoter did not show activity in *C. neoformans*, a species belonging to *Basidiomycota* (Wang *et al.*, 2016).

Given that the U6 snRNA genes are highly conserved among different organisms, a variety of putative endogenous fungal U6 promoters have been identified based on bioinformatics analysis such as ortholog sequence alignment and RNAseq (Zhang *et al.*, 2016; Katayama *et al.*, 2015; Arazoe *et al.*, 2015; Pohl *et al.*, 2016; Wang *et al.*, 2016; Schuster *et al.*, 2015). However, the sizes of U6 promoters used by those studies are variable, from 273 nt to 826 nt, indicating that those cloned U6 promoter may include extra fragments. In fact, cloning a functional U6 promoter has been a challenge in many fungal and oomycete species. One of the biggest problems is that U6 genes usually are present in multiple copies in the genome. For example *P. infestans* has 127 U6 orthologs (Fang & Tyler, 2016), which creates a challenge to identify a functional one. Zhang and colleagues tested three putative U6 promoters in *A. fumigatus* based on bioinformatics analysis, but only one showed significant activity (Zhang *et al.*, 2016), while the two putative U6 promoters cloned from *Magnaporthe oryae* both showed high activity (Arazoe *et al.*, 2015). In contrast, I was not able to find a functional *Phytophthora* U6 promoter after testing several U6 orthologs from different oomycete genomes (Fang & Tyler, 2016).

To overcome the challenge of identifying an active Pol III promoter, many alternative approaches have been used to drive sgRNA transcription in fungi and oomycetes. DNA-dependent RNA polymerase II (Pol II) has been used directly for transcription of sgRNA in *P.*

chrysogenum, *M. oryzae*, and *P. sojae*. While sgRNA synthesized by Pol II could mediate gene editing without further processing in *M. oryzae* (weakly, based on a pigment assay) (Arazoe *et al.*, 2015) and sufficiently in *Penicillium chrysogenum* (based on diagnostic PCR of gene replacement events) (Pohl *et al.*, 2016), it did not produce significant activity in *P. sojae* (based on PCR and restriction enzyme assay of NHEJ mutagenesis) (Fang & Tyler, 2016). It should be noted that the assayed mutations in *M. oryzae* and *P. chrysogenum* were mediated by homology-directed repair (HDR), whereas in *P. sojae* the assayed mutations were mediated by NHEJ. It has been reported in *A. fumigatus* (Zhang *et al.*, 2016) and *C. albicans* (Vyas *et al.*, 2015) that HDR-mediated mutation is more efficient than that of NHEJ when a complementary repair template is provided. Therefore, it is possible that a low level of DNA double-strand breaks (DBDs) can be detected by HDR-based diagnostic PCR assays but not by NHEJ assays. However, a more efficient way of using Pol II is to add ribozymes surrounding the sgRNA (Nødvig *et al.*, 2015; Fang & Tyler, 2016; Weber *et al.*, 2016). Ribozymes have self-cleavage activity, so that mRNA-specific structures, such as the 5′-cap and poly (A), can be removed to prevent the delivery of the sgRNA into the cytoplasm.

Various tRNA promoters have also been used to produce mature sgRNA efficiently in *S. cerevisiae* (Ryan *et al.*, 2014), and the filamentous fungi *P. chrysogenum* (Pohl *et al.*, 2016) and *Yarrowia lipolytica* (Schwartz *et al*., 2016). Interestingly, Ryan and collaborators added an HDV ribozyme structure between the tRNA and sgRNA, hypothesizing that the structured ribozyme would protect the 5′ end of the sgRNA from 5′ exonucleases. Schwartz and colleagues showed that hybrid promoters combining the truncated Pol III promoters RPR1, SCR1 and SNR52 with $tRNA^{Gly}$ produced the highest efficiency of genome editing (Schwartz *et al*., 2016). Transformation of *in vitro* transcribed sgRNA also has been used as an alternative strategy to

generate sgRNA in two studies (Liu *et al.*, 2015; Pohl *et al.*, 2016). The details will be discussed in section 1.3.3.

Schwartz and colleagues (Schwartz *et al.*, 2016) compared the efficiency of gene editing using different promoters, including Pol II (with HH and HDV ribozymes flanking the sgRNA) and a variety of native and synthetic Pol III promoters. In this study, Pol II combined with ribozymes showed the lowest efficiency (~10%). However, qPCR examination indicated that transcription of the sgRNA was also very low. Therefore, it could be possible that the Pol II promoter used in that study was not efficient for transcription. As noted above, their study showed a synthetic Pol III promoter consisting of a truncated SCR1 promoter and a glycine tRNA promoter yielded the highest efficiency.

Table 1.1 Summary of CRISPR/Cas9 mediated genome editing in fungi and oomycetes

Species		Strategy for sgRNA processing	Expression of Cas9[b]	Delivery of Cas9/sgRNA[c]	References
Ascomycetes					
Aspergillus	*A. aculeatus; A. brasiliensis; A. carbonarius; A. luchuensis; A. nidulans; A. niger*	*A. nidulans gpdA* promoter (Pol II) + HH/HDV ribozymes	≡ *A. niger* codon-optimized	*In vivo*, <Cas9+sgRNA, AMA1>	(Nødvig *et al.*, 2015)
	A. fumigatus	*S. cerevisae SNR52* promoter (Pol III)	≡ Human codon-optimized (Mali *et al.*, 2013)	*In vivo*, <Cas9+sgRNA> (Cas9)+|DsgRNA|	(Fuller *et al.*, 2015)
		A. fumigatus U6 promoters (Pol III) *In vivo* transcribed sgRNA	≡ & ≈P*niiA* Human codon-optimized (Cong *et al.*, 2013)	*In vivo*, <Cas9+sgRNA> (Cas9)+ *in vitro* RsgRNA	(Zhang *et al.*, 2016)
		A. nidulans gpdA promoter (Pol II) + HH/HDV ribozymes	≈ P*tet*^{ON} *A. niger* codon-optimized	(Cas9)+DsgRNA|	(Weber *et al.*, 2016)
	A. oryzae	*A. oryzae* U6 promoter (Pol III)	≡ *A. oryzae* codon-optimized	*In vivo*, |Cas9+sgRNA|	(Katayama *et al.*, 2015)
Candida albicans		*S. cerevisae SNR52* promoter (Pol III)	≡ *C. albicans* codon-optimized	(Cas9+sgRNA)	(Vyas *et al.*, 2015)
Magnaporthe oryzae		*M. oryzae* U6 *(Pol III)* *TrpC* promoter (Pol II) (less efficient)	≡ *M. oryzae* codon-optimized	*In vivo*, |Cas9+sgRNA|	(Arazoe *et al.*, 2015)
Neurospora crassa		*S. cerevisae SNR52* promoter (Pol III)	≡ Human codon-optimized (Mali *et al.*, 2013)	*In vivo*, |Cas9+sgRNA|	(Matsu-ura *et al.*, 2015)
Penicillium chrysogenum		*P. chrysogenum tRNA*^{Met} (Pol III); *P. chrysogenum tRNA*^{Leu} (Pol III); *P. chrysogenum* U6 (Pol III); *P. chrysogenum utp25* (Pol II); *In vivo* transcribed sgRNA	≈ P*xlnA* Plant codon-optimized (Nekrasov *et al.*, 2013)	*In vivo*, <Cas9+sgRNA, AMA1|; *In vivo* <Cas9, AMA1> RsgRNA|; |Cas9+RsgRNA|; RNP	(Pohl *et al.*, 2016)
Saccharomyces cerevisiae[a]		*S. cerevisae SNR52* promoter (Pol III)	≈ PGal-L Human codon-optimized (Mali *et al.*, 2013)		(DiCarlo *et al.*, 2013)
		SNR52 (Pol III, not working in strain ATCC4124) *tRNA*^{Tyr} (Pol III) +HDV ribozyme *tRNA*^{pro} (Pol III) +HDV ribozyme	≡ Original codon (Jinek *et al.*, 2012)	*In vivo*, |Cas9+sgRNA|	(Ryan *et al.*, 2014)
Schizosaccharomyces pombe		*rrk1* promoter (Pol II) +leader RNA + HH ribozyme	≡ Human codon-optimized (Mali *et al.*, 2013)	*In vivo*, |Cas9+sgRNA|	(Jacobs *et al.*, 2014)

Table 1.1 (continued)

Trichoderma reesei	*In vitro* transcribed sgRNA	≡ & ≈ P*cbh1 T. reesei* codon-optimized	(Cas9) +|RsgRNA|	(Liu *et al.*, 2015)
Yarrowia lipolytica	TEF promoter (Pol II) +HH/HDV ribozymes SNR52 (Pol III) $tRNA^{Gly}$(Pol III) truncated *RPR1* promoter + $tRNA^{Gly}$ truncated *SCR1* promoter + $tRNA^{Gly}$ truncated *SNR52* promoter + $tRNA^{Gly}$	≡ *Y. lipolytica* codon-optimized	*In vivo*, |Cas9+sgRNA|	(Schwartz *et al.*, 2016)
Basidiomycetes				
Cryptococcus neoformans	*C. neoformans* U6 promoter	≡ Human codon-optimized (Cong *et al.*, 2013)	*In vivo*, |Cas9+sgRNA|	(Wang *et al.*, 2016)
Ustilago maydis	*U. maydis* U6 promoter (Pol III)	≡ *U. maydis* codon-optimized	*In vivo*, <Cas9+sgRNA+ARS>	(Schuster *et al.*, 2015)
Oomycetes				
Phytophthora sojae	*P. sojae RPL41 promoter (Pol II, no function)* *P. sojae RPL41 promoter (Pol II) + HH/HDV ribozymes*	≡ Human codon-optimized (Cong *et al.*, 2013)	|Cas9+sgRNA|	(Fang & Tyler, 2016)

a. Genome editing of *S. cerevisiae* utilizing CRISPR/Cas9 has been reported by several groups (Jakociunas *et al.*, 2015; Horwitz *et al.*, 2015; Bao *et al.*, 2014; Ryan *et al.*, 2014; DiCarlo *et al.*, 2013). Here, we only show details of the first published genome editing in the laboratory strain *S. cerevisiae* S288c, together with a representative modification of the CRISPR/Cas9 system reported in Ryan *et al.*, 2014.

b. Expression of Cas9, ≡, constitutive expression, ≈ inducible expression. The inducible promoter is shown on the right beginning with 'P'.

c. (gene) indicates the gene was integrated into the genome; <gene> indicated the gene is episomally expressed; |gene| indicates the gene is introduced on a non-replicating plasmid and integration of the gene into the genome is uncertain; RNP, assembly of purified Cas9 protein and *in vitro* transcribed sgRNA; DsgRNA, expression of sgRNA *in vivo* based on a plasmid; RsgRNA, *in vitro* synthesized sgRNA; AMA1, autonomous maintenance in *Aspergillus* (a plasmid replicator) assembled with in the Cas9/sgRNA expression plasmids.

1.4.3 Effective delivery of Cas9/sgRNA

To date, there have been two principal ways of delivering Cas9/sgRNA components into fungi and oomycetes, namely *in vivo* expression from plasmid(s) harboring Cas9 and sgRNA genes, and delivery of *in vitro* expressed Cas9 and/or *in vitro* transcribed RNA. Since these two pathways have been combined, three approaches for delivery of Cas9/sgRNA have emerged, namely (1) *in vivo* expressed Cas9 + *in vivo* transcribed sgRNA; (2) *in vivo* expressed Cas9 + *in vitro* transcribed sgRNA; (3) *in vitro* expressed Cas9 + *in vitro* transcribed sgRNA (also called ribonucleoprotein, RNP) (Table 1.1).

Expression of both Cas9 and sgRNA *in vivo* is the simplest and most economical way, and thus has been widely used in various fungal and oomycete systems. In this strategy, Cas9 can either be integrated into the genome and expressed stably *in vivo*, or expressed episomally and thus transiently. To minimize off-target effects, autonomous replicating sequences (ARS) have been used to minimize integration of plasmids into the genome (Nødvig *et al*., 2015; Pohl *et al*., 2016; Schuster *et al*., 2015), and because ARS plasmids are readily lost without selection (Aleksenko & Clutterbuck, 1997; Khrunyk *et al*., 2010). On the other hand, several groups have used stable Cas9-expressing strains as recipient strains for sgRNA transformation without detection of any off-target effects (Fuller *et al.*, 2015; Liu *et al.*, 2015; Zhang *et al.*, 2016). However, the *in vivo* expression approach requires an established transformation and plasmid expression system, which may not be feasible to some fungal or oomycetes species.

In vitro synthesized sgRNA can be used for co-transformation with DNA encoding Cas9 (plasmid or genome-integrated). This strategy has been effectively used in *T. reesei* (Liu *et al*., 2015), *A .fumigatus* (Zhang *et al.*, 2016) and *P. chrysogenum* (Pohl *et al*., 2016). The biggest advantage of this approach is that no promoter is needed to transcribe sgRNA. However, naked

sgRNA may not be stable during transformation and inside the cell, and its usage may not be compatible with all transformation methods (Pohl *et al.*, 2016).

Pohl and coworkers also demonstrated rapid genome editing in *P. chrysogenum* by transformation with RNP directly (Pohl *et al.*, 2016). Since RNP-based genome editing is transient (RNPs are degraded eventually), the approach may avoid ectopic expression of foreign DNA and minimize the chance of off-target events. In addition, plasmid-free methods may also a good choice for those fungal and oomycete species that lack transformation and overexpression systems, albeit an RNP-delivery method is still needed. However, mutation screening may be a problem when using this strategy, due to the inability to first select for transformants, unless genes with a clear phenotype (such as pigment) can be selected a target. In addition, transformation with proteins has been rarely reported in fungi and oomycete, which may need further optimization. Finally, if DNA transformation is not available, the RNP approach is limited to NHEJ-mediated mutations and deletions.

Overall, the versatile CRISPR/Cas9 strategies described above provide multiple choices for genome editing in fungi and oomycetes.

1.5 Aims of the dissertation

The starting-point of my dissertation was to explore a nuclease based genome editing strategy to modify the *P. sojae* genome. This project was initialized with a class of customized nucleases called TALENs, (Transcription Activator-Like effector nuclease). However, my various attempts on this project were not successful. One important reason was that TALEN construct was silenced during transcription. Meanwhile, while trouble-shooting the system, I discovered that

the NLS (derived from SV40 large T-antigen) commonly used for nuclear import of heterologous proteins such as TALENs did not work efficiently in *P. sojae*. This observation motivated the investigation of distinctive mechanisms of nuclear localization in *P. sojae*. At the same time, with the emergence of a new nuclease-mediated genome editing technique, CRISPR/Cas9, I was able to continue to explore this new system to achieve genome editing in *P. sojae*.

1.6 ACKNOWLEDGEMENTS

Section 1.4 will be published as a section in a CRISPR review paper. I thank Dr. Kevin K. Full (Department of Genetics, Geisel School of Medicine at Dartmouth College) who contributed to the writing of the first paragraph in 1.4, and thank Dr. Brett. M Tyler who helped to revise the writing.

1.7 REFERENCES

Ah-Fong, A.M. & H.S. Judelson, (2011) Vectors for fluorescent protein tagging in *Phytophthora*: tools for functional genomics and cell biology. *Fungal Biol* **115**: 882-890.

Ah-Fong, A.M., C.A. Bormann-Chung & H.S. Judelson, (2008) Optimization of transgene-mediated silencing in *Phytophthora infestans* and its association with small-interfering RNAs. *Fungal Genet. Biol.* **45**: 1197-1205.

Aleksenko, A. & A. Clutterbuck, (1997) Autonomous plasmid replication in *Aspergillus nidulans*: AMA1 and MATE elements. *Fungal Genet. Biol.* **21**: 373-387.

Arazoe, T., K. Miyoshi, T. Yamato, T. Ogawa, S. Ohsato, T. Arie & S. Kuwata, (2015) Tailor-made CRISPR/Cas system for highly efficient targeted gene replacement in the rice blast fungus. *Biotechnol. Bioeng.* **112**: 2543-2549.

Baake, M., M. Bauerle, D. Doenecke & W. Albig, (2001) Core histones and linker histones are imported into the nucleus by different pathways. *Eur. J. Cell Biol.* **80**: 669-677.

Bao, Z., H. Xiao, J. Liang, L. Zhang, X. Xiong, N. Sun, T. Si & H. Zhao, (2014) Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS synthetic biology* **4**: 585-594.

Blair, J.E., M.D. Coffey, S.-Y. Park, D.M. Geiser & S. Kang, (2008) A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. *Fungal Genet. Biol.* **45**: 266-277.

Cavalier-Smith, T., (1986) The kingdom Chromista: origin and systematics. *Progress in phycological research* **4**: 309-347.

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini & F. Zhang, (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

Dou, D., S.D. Kale, T. Liu, Q. Tang, X. Wang, F.D. Arredondo, S. Basnayake, S. Whisson, A. Drenth & D. Maclean, (2010) Different domains of *Phytophthora sojae* effector Avr4/6 are recognized by soybean resistance genes Rps 4 and Rps 6. *Mol. Plant-Microbe Interact.* **23**: 425-435.

De Cock, A.W., L. Mendoza, A.A. Padhye, L. Ajello & L. Kaufman, (1987) *Pythium insidiosum* sp. nov., the etiologic agent of pythiosis. *J. Clin. Microbiol.* **25**: 344-349.

DiCarlo, J.E., J.E. Norville, P. Mali, X. Rios, J. Aach & G.M. Church, (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**: 4336-4343.

Dick, M.W., (2001) Straminipilous Fungi: Systematics of the *Peronosporomycetes* including accounts of the marine straminipilous protists, the *Plasmodiophorids* and similar organisms. Springer.

Érsek, T. & O. Ribeiro, (2010) Mini review article: an annotated list of new *Phytophthora* species described post 1996. *Acta Phytopathol. Entomol. Hung.* **45**: 251-266.

Erwin, D.C., S. Bartnicki-Garcia & P.H.-t. Tsao, (1983) *Phytophthora:* its biology, ecology and pathology. American Phytopathological Society.

Erwin, D.C. & O.K. Ribeiro, (1996) *Phytophthora* diseases worldwide. American Phytopathological Society (APS Press).

Fang, Y. & B.M. Tyler, (2016) Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* **17**: 127-139.

Fuller, K.K., S. Chen, J.J. Loros & J.C. Dunlap, (2015) Development of the CRISPR/Cas9 system for targeted gene disruption in *Aspergillus fumigatus*. *Eukaryot. Cell* **14**: 1073-1080.

Gaastra, W., L.J. Lipman, A.W. De Cock, T.K. Exel, R.B. Pegge, J. Scheurwater, R. Vilela & L. Mendoza, (2010) *Pythium insidiosum*: an overview. *Vet. Microbiol.* **146**: 1-16.

Gamboa-Melendez, H., A.I. Huerta & H.S. Judelson, (2013) bZIP transcription factors in the oomycete *phytophthora infestans* with novel DNA-binding domains are involved in defense against oxidative stress. *Eukaryot. Cell* **12**: 1403-1412.

Grooters, A.M., (2003) Pythiosis, lagenidiosis, and zygomycosis in small animals. *Vet Clin North Am Small Anim Pract* **33**: 695-720, v.

Gunderson, J.H., H. Elwood, A. Ingold, K. Kindle & M.L. Sogin, (1987) Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. *Proceedings of the National Academy of Sciences* **84**: 5823-5827.

Horwitz, A.A., J.M. Walter, M.G. Schubert, S.H. Kung, K. Hawkins, D.M. Platt, A.D. Hernday, T. Mahatdejkul-Meadows, W. Szeto, S.S. Chandran & J.D. Newman, (2015) Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Syst* **1**: 88-96.

Hsu, P.D., E.S. Lander & F. Zhang, (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262-1278.

Hsu, P.D., D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, E.J. Fine, X. Wu, O. Shalem, T.J. Cradick, L.A. Marraffini, G. Bao & F. Zhang, (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**: 827-832.

Jacobs, J.Z., K.M. Ciccaglione, V. Tournier & M. Zaratiegui, (2014) Implementation of the CRISPR-Cas9 system in fission yeast. *Nat Commun* **5**: 5344.

Jakociunas, T., I. Bonde, M. Herrgard, S.J. Harrison, M. Kristensen, L.E. Pedersen, M.K. Jensen & J.D. Keasling, (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab. Eng.* **28**: 213-222.

Jiang, R.H. & B.M. Tyler, (2012) Mechanisms and evolution of virulence in oomycetes. *Annu. Rev. Phytopathol.* **50**: 295-318.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna & E. Charpentier, (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

Judelson, H.S., (1997) The genetics and biology of *Phytophthora infestans*: Modern approaches to a historical challenge. *Fungal Genet. Biol.* **22**: 65-76.

Judelson, H.S. & F.A. Blanco, (2005) The spores of *Phytophthora*: weapons of the plant destroyer. *Nat. Rev. Microbiol.* **3**: 47-58.

Judelson, H.S., M.D. Coffey, F.R. Arredondo & B.M. Tyler, (1993) Transformation of the oomycete pathogen *Phytophthora-megasperma* f-sp glycinea occurs by dna integration into single or multiple chromosomes. *Curr. Genet.* **23**: 211-218.

Judelson, H.S. & R.W. Michelmore, (1991) Transient expression of genes in the oomycete *Phytophthora infestans* using *Bremia lactucae* regulatory sequences. *Curr. Genet.* **19**: 453-459.

Judelson, H.S., B.M. Tyler & R.W. Michelmore, (1991) Transformation of the oomycete pathogen, *Phytophthora infestans*. *Mol Plant Microbe Interact* **4**: 602-607.

Judelson, H.S., B.M. Tyler & R.W. Michelmore, (1992) Regulatory sequences for expressing genes in oomycete fungi. *Molecular and General Genetics MGG* **234**: 138-146.

Kamoun, S., P. van West, V.G. Vleeshouwers, K.E. de Groot & F. Govers, (1998) Resistance of Nicotiana benthamiana to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. *The Plant Cell* **10**: 1413-1425.

Kalderon, D., W.D. Richardson, A.F. Markham, and A.E. Smith (1984) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**: 33 - 38

Katayama, T., Y. Tanaka, T. Okabe, H. Nakamura, W. Fujii, K. Kitamoto & J.-i. Maruyama, (2015) Development of a genome editing technique using the CRISPR/Cas9 system in the industrial filamentous fungus *Aspergillus oryzae*. *Biotechnol. Lett.*: 1-6.

Khrunyk, Y., K. Münch, K. Schipper, A.N. Lupas & R. Kahmann, (2010) The use of FLP-mediated recombination for the functional analysis of an effector gene family in the biotrophic smut fungus *Ustilago maydis*. *New Phytol.* **187**: 957-968.

Kroon, L.P., H. Brouwer, A.W. de Cock & F. Govers, (2012) The genus *Phytophthora* anno 2012. *Phytopathology* **102**: 348-364.

Lamour, K.H., L. Finley, O. Hurtado-Gonzales, D. Gobena, M. Tierney & H.J. Meijer, (2006) Targeted gene mutation in *Phytophthora* spp. *Mol. Plant-Microbe Interact.* **19**: 1359-1367.

Lange, A., R.E. Mills, C.J. Lange, M. Stewart, S.E. Devine & A.H. Corbett, (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**: 5101-5105.

Large, E.C., (1940) The advance of the fungi. *The Advance of the Fungi.*

Lee, B.J., A.E. Cansizoglu, K.E. Süel, T.H. Louis, Z. Zhang, Y.M. Chook (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **126**: 543-558.

Lévesque, C.A., (2011) Fifty years of oomycetes—from consolidation to evolutionary and genomic exploration. *Fungal Diversity* **50**: 35-46.

Liu, R., L. Chen, Y. Jiang, Z. Zhou & G. Zou, (2015) Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. *Cell Discovery* **1**: 15007.

Mali, P., L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville & G.M. Church, (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.

Mallet, P.L. & F. Bachand, (2013) A proline-tyrosine nuclear localization signal (PY-NLS) is required for the nuclear import of fission yeast PAB2, but not of human PABPN1. *Traffic* **14**: 282-294.

Matsu-ura, T., M. Baek, J. Kwon & C. Hong, (2015) Efficient gene editing in *Neurospora crassa* with CRISPR technology. *Fungal Biology and Biotechnology* **2**: 1.

Mosammaparast, N., K.R. Jackson, Y.R. Guo, C.J. Brame, J. Shabanowitz, D.F. Hunt & L.F. Pemberton, (2001) Nuclear import of histone H2A and H2B is mediated by a network of karyopherins. *J. Cell Biol.* **153**: 251-262.

Mendoza, L. & R. Vilela, (2013) The Mammalian pathogenic oomycetes. *Current Fungal Infection Reports* **7**: 198-208.

Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, D.E. Paschon, E. Leung, S.J. Hinkley, G.P. Dulay, K.L. Hua, I. Ankoudinova, G.J. Cost, F.D. Urnov, H.S. Zhang, M.C. Holmes, L. Zhang, P.D. Gregory & E.J. Rebar, (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**: 143-148.

Nekrasov, V., B. Staskawicz, D. Weigel, J.D. Jones & S. Kamoun, (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**: 691-693.

Nødvig, C.S., J.B. Nielsen, M.E. Kogle & U.H. Mortensen, (2015) A CRISPR-Cas9 system for genetic engineering of filamentous fungi. *PloS one* **10**: e0133085.

Oidtmann, B., S. Bausewein, L. Hölzle, R. Hoffmann & M. Wittenbrink, (2002) Identification of the crayfish plague fungus *Aphanomyces astaci* by polymerase chain reaction and restriction enzyme analysis. *Vet. Microbiol.* **85**: 183-194.

Pohl, C., J.A. Kiel, A.J. Driessen, R.A. Bovenberg & Y. Nygard, (2016) CRISPR/Cas9 based genome editing of *Penicillium chrysogenum*. *ACS Synth Biol* **5**: 754-764.

Raffaele, S. & S. Kamoun, (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* **10**: 417-430.

Ryan, O.W., J.M. Skerker, M.J. Maurer, X. Li, J.C. Tsai, S. Poddar, M.E. Lee, W. DeLoache, J.E. Dueber, A.P. Arkin & J.H. Cate, (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife* **3**.

Süel K.E., H. Gu, Y.M. Chook (2008) Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals. *PLoS Biol.* **6**: e137

Schuster, M., G. Schweizer, S. Reissmann & R. Kahmann, (2015) Genome editing in *Ustilago maydis* using the CRISPR–Cas system. *Fungal Genet. Biol.*

Schwartz, C.M., M.S. Hussain, M. Blenner & I. Wheeldon, (2016) Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR–Cas9-mediated genome editing in *Yarrowia lipolytica*. *ACS synthetic biology* **5**: 356-359.

Tyler, B.M., (2001) Genetics and genomics of the oomycete–host interface. *Trends Genet.* **17**: 611-614.

Tyler, B.M., (2007) P*hytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* **8**: 1-8.

Tyler, B.M. & M. Gijzen, (2014) The *Phytophthora sojae* genome sequence: Foundation for a revolution. in: Genomics of plant-associated fungi and oomycetes: Dicot pathogens. Springer, pp. 133-157.

Tyler, B.M., S. Tripathy, X. Zhang, P. Dehal, R.H. Jiang, A. Aerts, F.D. Arredondo, L. Baxter, D. Bensasson, J.L. Beynon, J. Chapman, C.M. Damasceno, A.E. Dorrance, D. Dou, A.W. Dickerman, I.L. Dubchak, M. Garbelotto, M. Gijzen, S.G. Gordon, F. Govers, N.J. Grunwald, W. Huang, K.L. Ivors, R.W. Jones, S. Kamoun, K. Krampis, K.H. Lamour, M.K. Lee, W.H. McDonald, M. Medina, H.J. Meijer, E.K. Nordberg, D.J. Maclean, M.D. Ospina-Giraldo, P.F. Morris, V. Phuntumart, N.H. Putnam, S. Rash, J.K. Rose, Y. Sakihama, A.A. Salamov, A. Savidor, C.F. Scheuring, B.M. Smith, B.W. Sobral, A. Terry, T.A. Torto-Alalibo, J. Win, Z. Xu, H. Zhang, I.V. Grigoriev, D.S. Rokhsar & J.L. Boore, (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**: 1261-1266.

van West, P., (2006) *Saprolegnia parasitica*, an oomycete pathogen with a fishy appetite: new challenges for an old problem. *Mycologist* **20**: 99-104.

van West, P., S. Kamoun, J.W. van't Klooster & F. Govers, (1999) Internuclear gene silencing in *Phytophthora infestans*. *Mol. Cell* **3**: 339-348.

Vyas, V.K., M.I. Barrasa & G.R. Fink, (2015) A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. *Science advances* **1**: e1500248.

Wang, Q., C. Han, A.O. Ferreira, X. Yu, W. Ye, S. Tripathy, S.D. Kale, B. Gu, Y. Sheng, Y. Sui, X. Wang, Z. Zhang, B. Cheng, S. Dong, W. Shan, X. Zheng, D. Dou, B.M. Tyler & Y. Wang, (2011) Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. *Plant Cell* **23**: 2064-2086.

Wang, Y., D. Dou, X. Wang, A. Li, Y. Sheng, C. Hua, B. Cheng, X. Chen, X. Zheng & Y. Wang, (2009) The PsCZF1 gene encoding a C2H2 zinc finger protein is required for growth, development and pathogenesis in *Phytophthora sojae*. *Microb. Pathog.* **47**: 78-86.

Wang, Y., D. Wei, X. Zhu, J. Pan, P. Zhang, L. Huo & X. Zhu, (2016) A 'suicide' CRISPR-Cas9 system to promote gene deletion and restoration by electroporation in *Cryptococcus neoformans*. *Sci Rep* **6**: 31145.

Weber, J., V. Valiante, C.S. Nodvig, D.J. Mattern, R.A. Slotkowski, U.H. Mortensen & A.A. Brakhage, (2016) Functional Reconstitution of a Fungal Natural Product Gene Cluster by Advanced Genome Editing. *ACS Synth Biol*.

Whisson, S.C., A.O. Avrova, P. Van West & J.T. Jones, (2005) A method for double-stranded RNA-mediated transient gene silencing in Phytophthora infestans. *Mol. Plant Pathol.* **6**: 153-163.

Xiang, Q. & H.S. Judelson, (2010) Myb transcription factors in the oomycete *Phytophthora* with novel diversified DNA-binding domains and developmental stage-specific expression. *Gene* **453**: 1-8.

Xiang, Q. & H.S. Judelson, (2014) Myb transcription factors and light regulate sporulation in the oomycete *Phytophthora infestans*. *PLoS One* **9**: e92086.

Zhang, M., J. Lu, K. Tao, W. Ye, A. Li, X. Liu, L. Kong, S. Dong, X. Zheng & Y. Wang, (2012) A Myb transcription factor of *Phytophthora sojae*, regulated by MAP kinase PsSAK1, is required for zoospore development. *PLoS One* **7**: e40246

Zhang, C., X. Meng, X. Wei & L. Lu, (2016) Highly efficient CRISPR mutagenesis by microhomology-mediated end joining in *Aspergillus fumigatus*. *Fungal Genet. Biol.* **86**: 47-57.

Chapter 2

Distinctive nuclear localization signals in the oomycete *Phytophthora sojae*

Yufeng Fang[1, 2], Hyo Sang Jang[3], Gregory W. Watson[4, 5], Dulani P. Wellapilli[2], and Brett M. Tyler[1, 2*]

[1]Interdisciplinary Ph.D. program in Genetics, Bioinformatics & Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA

[2]Center for Genome Research and Biocomputing and Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

[3]Department of Environmental & Molecular Toxicology, Oregon State University, Corvallis, OR 97331, USA

[4] Molecular and Cellular Biology Program, Oregon State University, Corvallis, OR 97331, USA

[5]Biological and Population Health Sciences, Oregon State University, Corvallis, OR 97331, USA

* Corresponding author: Brett.Tyler@oregonstate.edu

This Chapter includes a research article submitted to Molecular Microbiology as "Distinctive nuclear localization signals in the oomycete Phytophthora sojae". I contributed 90% of the work described in this chapter. Hyo Sang Jang and Gregory W. Watson contributed the data shown in Fig. 2.3. Gregory W. Watson also helped to edit the manuscript. Dulani P. Wellapilli contributed technical assistance. Brett M. Tyler helped to analyze the data and edited the manuscript.

2.1 ABSTRACT

To date, nuclear localization signals (NLSs) that target proteins to nuclei in oomycetes have not been defined, but have been assumed to be the same as in higher eukaryotes. Here, we use the soybean pathogen *Phytophthora sojae* as a model to investigate these sequences in oomycetes. By establishing a reliable *in vivo* NLS assay based on confocal microscopy, we found that many canonical monopartite and bipartite classical NLSs (cNLSs) mediated nuclear import poorly in *P. sojae*. We found that efficient localization of *P. sojae* nuclear proteins by cNLSs requires additional basic amino acids at distal sites or collaboration with other NLSs. We found that several representatives of another well-characterized NLS, proline-tyrosine NLS (PY-NLS) also functioned poorly in *P. sojae*. To characterize PY-NLSs in *P. sojae*, we experimentally defined the residues required by functional PY-NLSs in three *P. sojae* nuclear-localized proteins. These results showed that functional *P. sojae* PY-NLSs include an additional cluster of basic residues for efficient nuclear import. Finally, analysis of several highly conserved *P. sojae* nuclear proteins including ribosomal proteins and core histones revealed that these proteins exhibit a similar but stronger set of sequence requirements for nuclear targeting compared with their orthologs in mammals or yeast.

2.2 INTRODUCTION

In eukaryotes, many proteins such as core histones, transcription factors and ribosomal proteins must be transported into the nucleus to accomplish their functions. Transport of those proteins into the nucleus occurs through large, proteinaceous structures called nuclear pore complexes (NPCs). Generally, NPCs allow passive diffusion of molecules smaller than 40-60 kDa, but

require an appropriate sorting signal for passage of larger proteins (Lange *et al.*, 2007). The sorting signals carried by those proteins are called nuclear localization signals (NLSs), that generally are short stretches of amino acids recognized by nucleo-cytoplasmic transporters (karyopherins) that promote active transport of proteins into the nucleus (Xu *et al.*, 2010; Marfori *et al.*, 2011). To date, two principal types of NLSs have been defined: the classical NLS (cNLS) and the proline-tyrosine NLS (PY-NLS). cNLSs are the best-characterized nuclear targeting signals, and are recognized by karyopherin-β (Importin-β, Kapβ) through direct binding to an adaptor protein karyopherin-α (Importin-α, Kapα) (Lange *et al.*, 2007; Marfori *et al.*, 2011). According to the numbers of basic amino acid clusters within them, cNLSs are further divided into two subclasses, monopartite and bipartite cNLSs. Monopartite cNLSs contain a single stretch of basic amino acids which may consist of at least four consecutive basic amino acids, exemplified by SV40 large T antigen NLS (PKKKRKV) (Kalderon *et al.*, 1983; Kalderon *et al.*, 1984). Alternatively, three non-consecutive basic amino acids may suffice, exemplified by the *c-Myc* proto-oncoprotein NLS (PAAKRVKLD) (Makkerh *et al.*, 1996). The monopartite cNLS has a consensus of K(K/R)X(K/R) (Lange *et al.*, 2007). Bipartite cNLSs have two stretches of basic amino acids separated by 10-12 amino acids (Lange *et al.*, 2007). They were first found in *Xenopus* nucleoplasmin (KRPAATKKAGQAKKKK) (Dingwall *et al.*, 1982) and are represented by the consensus sequence $(K/R)(K/R)X_{10–12}(K/R)_{3/5}$ (X is any amino acid and $(K/R)_{3/5}$ represents three lysine or arginine residues out of five consecutive amino acids) (Dingwall & Laskey, 1991).

The PY-NLS is recognized by karyopherin-β2 (Kapβ2) for nuclear import (Lee *et al.*, 2006). Compared to the cNLS, fewer PY-NLS proteins have been characterized experimentally (~42 through 2015, Soniat & Chook, 2015). PY-NLSs are generally longer (15-30 residues) and

more variable than cNLSs, making it more difficult to clearly define their common features (Xu *et al.*, 2010; Chook & Suel, 2011). M9NLS is the best-characterized PY-NLS. It consists of a 38-residue domain from a splicing factor, heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Bonifaci *et al.*, 1997; Truant *et al.*, 1999). On the basis of the crystal structure of human Kapβ2 bound to M9NLS, Lee *et al.* (2006) proposed three rules for PY-NLSs: (1) structurally disordered in free substrates; (2) overall basic character; and (3) possesses a central hydrophobic or basic motif (epitope 1) followed by the motif R/H/K-X_{2-5}-PY (the R/H/K- and PY-motifs are defined as epitopes 2 and 3 respectively). According to the composition of the amino acids in epitope 1, PY-NLSs have been further classified into hydrophobic and basic PY-NLSs (hPY- and bPY-NLSs) (Lee *et al.*, 2006). Later, the rules were updated to accommodate additional features identified in *Saccharomyces cerevisiae*: the Kapβ2 ortholog Kap104 only recognizes the basic but not the hydrophobic PY-NLS (Suel *et al.*, 2008); and the tyrosine in the C-terminal PY epitope (epitope 3) displays degeneracy; not only PY but also some other motifs like PL could be recognized by yeast Kap104 (Suel *et al.*, 2008). *In silico* predictions suggest that Kapβ2-mediated import accounts for a substantial fraction of substrates involved in RNA processing and transcription factors (Suel & Chook, 2009; Chook & Suel, 2011).

Phytophthora sojae is a destructive oomycete pathogen that infects soybean seedlings as well as established plants (Tyler, 2007). Although oomycetes physiologically and morphologically resemble fungi, molecular taxonomy has shown that oomycetes are phylogenetically close to algae and diatoms (Tyler, 2007; Kamoun *et al.*, 2015). Oomycetes are diploid and lack a free haploid life stage. The genomes of oomycetes (50-250 Mb) are also generally larger than those of true fungi (10-40 Mb) (reviewed in Judelson & Blanco, 2005). Most *Phytophthora* species are plant pathogens and together damage a huge range of

agriculturally and ornamentally important plants (Erwin & Ribeiro, 1996). For instance, *P. infestans*, which causes the potato late-blight disease, resulted in the Irish potato famine, and continues to be a problem for potato and tomato crops (Judelson & Blanco, 2005). *P. sojae* causes around $1-2 billion in losses per year to the soybean crop (Tyler, 2007). Because of its economic impact, *P. sojae*, along with *P. infestans*, has been developed as a model species for the study of oomycete plant pathogens (Tyler, 2007).

To date, most nuclear transport studies have been carried out in model organisms, and no NLS sequences have yet been defined in oomycetes. Here we find that many eukaryotic NLS sequences function poorly if at all in *P. sojae*. We demonstrate that efficient localization of *P. sojae* nuclear proteins by cNLSs requires additional basic amino acids at distal sites or collaboration with other NLSs. Furthermore, we show that a fully functional PY-NLS requires additional basic residues either within the motif itself or adjacent to the motif. Finally, comparison of the nuclear localization activities of NLS sequences from *P. sojae* ribosomal proteins and core histones with those from other eukaryotes reveals that *P. sojae* may use modified nuclear import mechanisms for those highly conserved nuclear proteins.

2.3 RESULTS

2.3.1 Establishment of reliable fluorescent labeling of *P. sojae* nuclei for assay of nuclear localization

To assess the activity of NLSs, we implemented a classic *in vivo* NLS assay, in which a candidate NLS was fused to GFP or to two fused GFP moieties (2XGFP) and then expressed in *P. sojae* transformants. The fusion of two GFP molecules created a protein (~55 kDa) larger than

the NPC threshold for passive diffusion. Subcellular localization was visualized by live-cell imaging using confocal laser scanning microscopy. To assist in verifying the nuclear localization of a protein, the commonly used nuclear dye DAPI (4', 6-diamidino-2-phenylindole) was initially employed, because it had been reported to label the nuclei of living (Hardham, 2001; Zhang *et al.*, 2012) or fixed (Gamboa-Melendez *et al.*, 2013) *Phytophthora* tissues. However, when living *P. sojae* hyphae were stained with DAPI, the NLS-fused GFPs were extensively distributed into the cytoplasm (Fig. S2.1A). This contrasted with the clear nuclear localization observed in hyphae that were not stained with DAPI (Fig. S2.1A). In particular, when using a 2XGFP reporter, fused to a strong synthetic NLS, PsNLS (Fang & Tyler, 2016), we noticed that regions of hyphae with poor DAPI staining exhibited strong GFP nuclear localization while regions of hyphae well-stained with DAPI showed poor GFP nuclear localization (Fig. S2.1B). This suggested that DAPI staining caused mis-localization of nuclear localized proteins to the cytoplasm. Indeed, a time-lapse experiment tracking a hypha during DAPI staining showed that after 18 min staining, the nuclei disintegrated and the nuclear localized PsNLS-2XGFP was released into the cytoplasm (Fig. S2.1C and Supplemental Movie 2.1).

To identify a reliable strategy to label *P. sojae* nuclei, we tested other nuclear staining dyes, such as Hoechst 33342. However, *P. sojae* hyphae were not permeable to that dye (data not shown). We also tried to stain the nucleus of fixed hyphae using a protocol described by Gamboa-Melendez *et al.* (2013), but redistribution of nuclear GFP into the cytoplasm was still commonly observed. Therefore, we sought to label *P. sojae* nuclei by expressing a nuclear-targeted fluorescent protein. *P. sojae* core histone H2B fused to mCherry was selected according to Ah-Fong & Judelson (2011), and this fusion showed predominant nuclear localization in *P. sojae* hyphae (Fig. 2.1A). Expression of H2B-mCherry appeared to cause some toxicity to *P.*

sojae, because reduced numbers of *P. sojae* transformants were obtained with this construct. Similar results were observed with a *P. sojae* histone H1 fusion protein, and even worse with histone H3 and H4 fusions (data not shown). Thus, we settled on H2B-mCherry as a nuclear marker for co-expression with NLS-GFP reporter genes.

2.3.2 Monopartite cNLSs show weak nuclear targeting activity in *P. sojae*

To examine cNLS activity in *P. sojae*, three well-characterized monopartite cNLS sequences were individually fused to the N-terminus of 2XGFP and expressed in *P. sojae* transformants. To quantify the activities of different NLSs, the ratio of fluorescence intensity in nuclei compared to the cytoplasm (Nuc:Cyt) was measured within a population of ~ 30 hyphae for each fusion. Due to the wide range of nuclear-cytoplasmic ratios observed, we found it convenient to express these ratios as $\log_2$(Nuc:Cyt), or LNC. As expected, 2XGFP alone was extensively localized to the cytoplasm (LNC=0.52; Fig. 2.1A and D). The well-studied cNLS derived from SV40 large T antigen (SV40 NLS) produced incomplete nuclear localization in any fusion configuration (fusion at N- or C-terminus of 2XGFP) or copy number (LNC=1.43-2.35; Fig. 2.1A, B and D). In comparison, H2B-GFP exhibited an LNC of 7.94. Another monopartite cNLS prototype, the *c-Myc* NLS, produced similar results (LNC=1.55; Fig. 2.1A and D). To characterize cNLSs in native *P. sojae* nuclear proteins, we also analyzed protein fragments that contained putative cNLS motifs predicted by *pSORTII* (PSORT, 1997). As summarized in Table S2.1, we observed that most putative monopartite cNLSs were not sufficient for nuclear localization of GFP reporters.

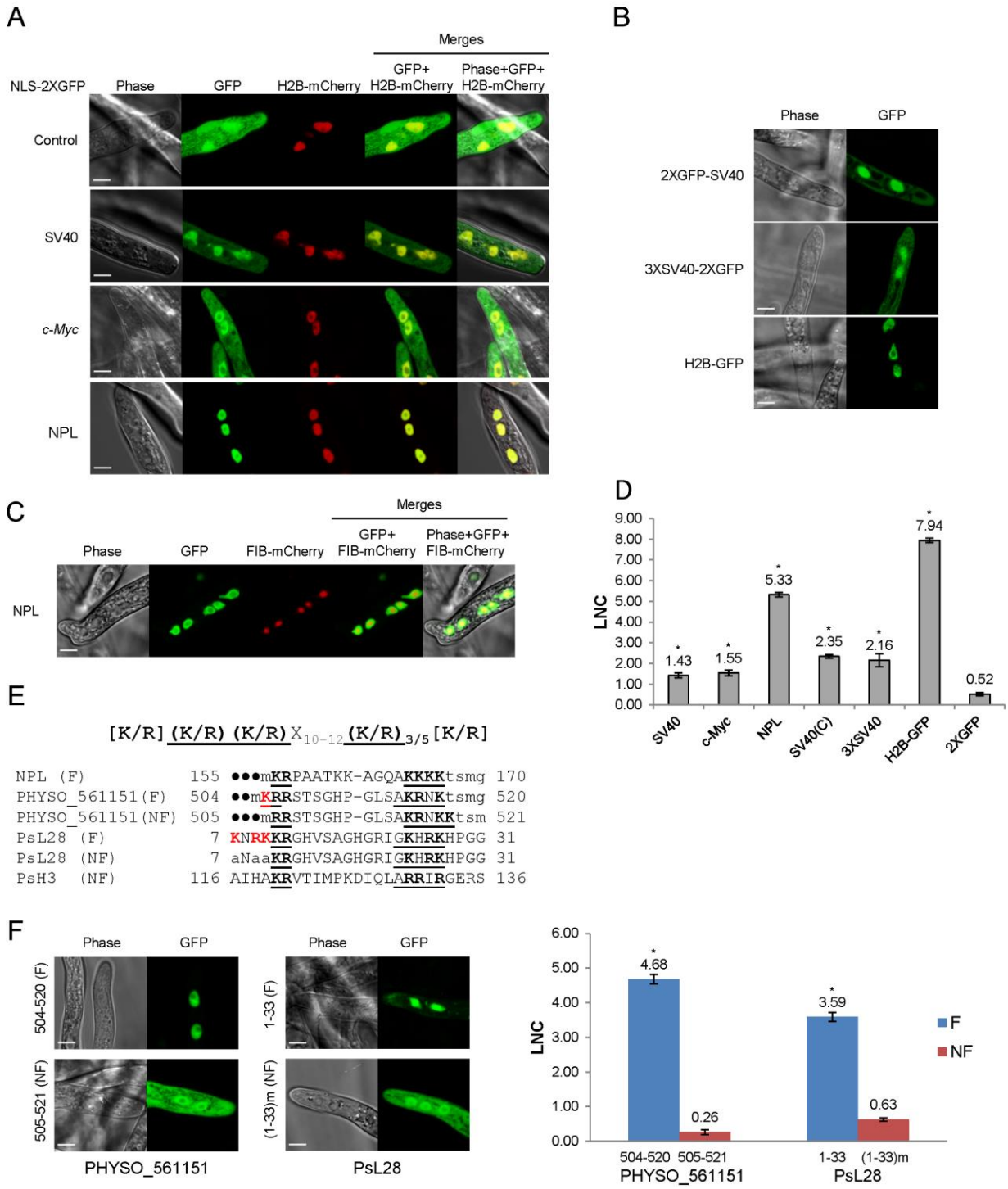


Fig. 2.1 Functional characterization of monopartite and bipartite cNLSs in P. sojae transformants.

A. Subcellular localization of N-terminal cNLS-2XGFP fusions. SV40, *c-Myc*, NPL, represent cNLSs identified in SV40 large T-antigen, human *c-Myc* proto-oncoprotein, and *Xenopus* nucleoplasmin respectively; H2B-mCherry = *P. sojae* histone H2B fused to the N-terminus of mCherry; Control =

2XGFP. "GFP" above the panel indicates visualization of the relevant GFP or 2xGFP fusion. Scale bar corresponds to 5 μm in this and all subsequent figures. Representative images are shown in A, B, C and F.

B. Subcellular localization of different fusions employing one or more copies of the SV40 cNLS. H2B-GFP used as a positive control.

C. NPL-2XGFP was excluded from the nucleolus, as confirmed by *P. capsici* fibrillarin fused to mCherry (FIB-mCherry).

D. Quantification of nuclear localization in A and B. For quantification in all figures, LNC indicates the mean $\log_2$-transformed nuclear fluorescence to cytoplasmic fluorescence ratio from ~30 randomly pairs of nuclear and adjacent cytoplasmic regions. SV40(C) = SV40NLS attached at the C-terminus of 2xGFP (i.e. 2XGFP-SV40NLS). Error bars, S.E. Asterisks, LNC values of NLS-GFP fusions that are significantly greater than 2XGFP ($p<0.01$), hereafter.

E. Sequence alignment of bipartite cNLSs tested in *P. sojae*. The proposed *P. sojae* bipartite cNLS consensus is shown on the top; [K/R] indicates the positions of additional positive residues required in *P. sojae*. In the sequences of each protein, the two elements of each bipartite cNLS predicted by *pSORTII* (PSORT, 1997) are underlined, and basic amino acids within each cluster are in bold. The additional basic amino acids demonstrated to contribute to nuclear localization are in bold red. The residues in lowercase indicate non-native residues flanking the candidate NLS in each construct. (F) = functional NLS; (NF) = non-functional NLS. For PHYSO_561151 and PsL28, both the full length functional cNLSs and the truncated or mutant non-functional cNLSs are shown.

F. Functional tests of *P. sojae* bipartite NLSs listed in E. Left, representative images showing subcellular localization of bipartite NLSs found in PHYSO_561151 and PsL28. Right, quantification of the localization. (1-33)m indicates mutation of the additional positively charged amino acids (R7A/K9A/R10A) that are not involved in the predicted bipartite cNLS in $Ps28_{1\text{-}33}$.

2.3.3 Functional bipartite cNLSs require additional basic amino acids compared to the conventional bipartite consensus

The bipartite cNLS of nucleoplasmin (NPL) produced strong nuclear localization (LNC=5.33; Fig. 2.1A and D). Small unstained regions in the centers of nuclei were validated as nucleoli by co-expression of the nucleolar marker, fibrillarin (from *Phytophthora capsici*; Genbank

accession XP_009521478.1) fused to mCherry (Fig. 2.1C). Most protein fragments predicted to have bipartite cNLSs (by *pSORTII*, 1997) were incapable of causing strong localization of GFP reporters into the *P. sojae* nucleus (Table S2.1). However, bipartite cNLSs found in two *P. sojae* nuclear-localized proteins did produce strong NLS activity, namely one located at the extreme C-terminus of PHYSO_561151 (LNC=4.68) and one at the N-terminus of PsL28 (LNC=3.59) (Fig. 2.1E and F). Sequence comparisons of the functional bipartite cNLSs in NPL, PsL28 and PHYSO_561151 revealed the presence of additional basic amino acids in one or other of the two basic amino acid clusters in each case, compared to the canonical consensus developed from mammalian and yeast proteins (Fig. 2.1E). Mutation of these additional basic amino acids from the PsL28 or PHYSO_561151 bipartite NLSs showed that these residues were essential for the activity of these NLSs (Fig. 2.1F).

2.3.4 Canonical PY-NLS motifs produce weak nuclear localization activity in *P. sojae*

Although a number of PY-NLS sequences have been characterized in human and yeast proteins, few of them have been reported in other organisms including oomycetes. To examine the activity of PY-NLS sequences in oomycetes, four well-characterized human and yeast PY-NLSs (two basic, bPY-NLS; two hydrophobic, hPY-NLS) were fused to the N-terminus of 2XGFP and their localization was examined in *P. sojae* transformants (Fig. 2.2A). Unexpectedly, none of these PY-NLSs were efficient in mediating nuclear accumulation of the 2XGFP reporter (Fig. 2.2B and C). Only the bPY-NLS in hnRNP M produced significantly more nuclear localization than the 2XGFP control (Fig. 2.2B and C). We also tested a chimeric peptide, M9M (Fig. 2.2A), which has high affinity to the Kapβ2 PY-NLS binding site in human and is usually used as a Kapβ2-specific inhibitor (Cansizoglu *et al.*, 2007). Interestingly, this peptide produced

significantly more nuclear localization than the four PY-NLS prototypes, although the localization was still weaker than that produced by the SV40 cNLS (Fig. 2.2B and C).

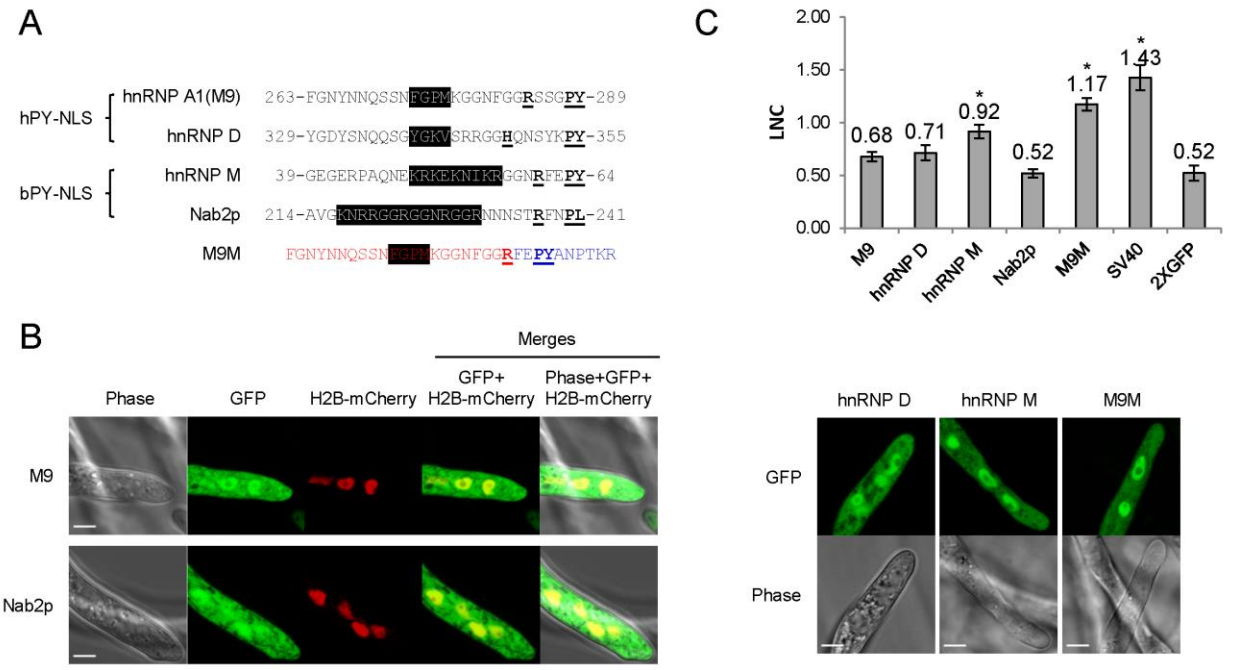


Fig. 2.2 PY-NLS prototypes exhibit weak nuclear targeting activities in P. sojae transformants**.**

A. Sequences of four well-characterized PY-NLSs derived from human or yeast proteins and the Kapβ2-specific nuclear import inhibitor, M9M peptide. Core residues that determine hydrophobic or basic PY-NLS type are shaded in black. The R/K/H-$X_{2\text{-}5}$-PY/L consensus residues are bold and underlined. Residues marked in red and blue in M9M are sequences originating from M9 and hnRNP M respectively.

B. Subcellular localization of the five PY-NLS described in (A). Representative images are shown.

C. Quantification of localization of fusions observed in (B). SV40 cNLS-2XGFP and 2XGFP alone indicate the controls (SV40 cNLS-2XGFP, positive control; 2XGFP, negative control) here and in subsequent figures; the same values were used in every figure.

To characterize native PY-NLS sequences in *P. sojae* nuclear localized proteins, we scanned the *P. sojae* proteome using the PY-NLS consensus sequence (the 'PL' rule was allowed for bPY-NLS, Fig. 2.3A), and then filtered for structural disorder and overall positive charge in

the NLS. This scan revealed 25 proteins containing candidate hPY-NLSs and 200 containing candidate bPY-NLSs (47 PY and 153 PL). Twelve candidates annotated as RNA-processing proteins or transcription factors were selected for further analysis.

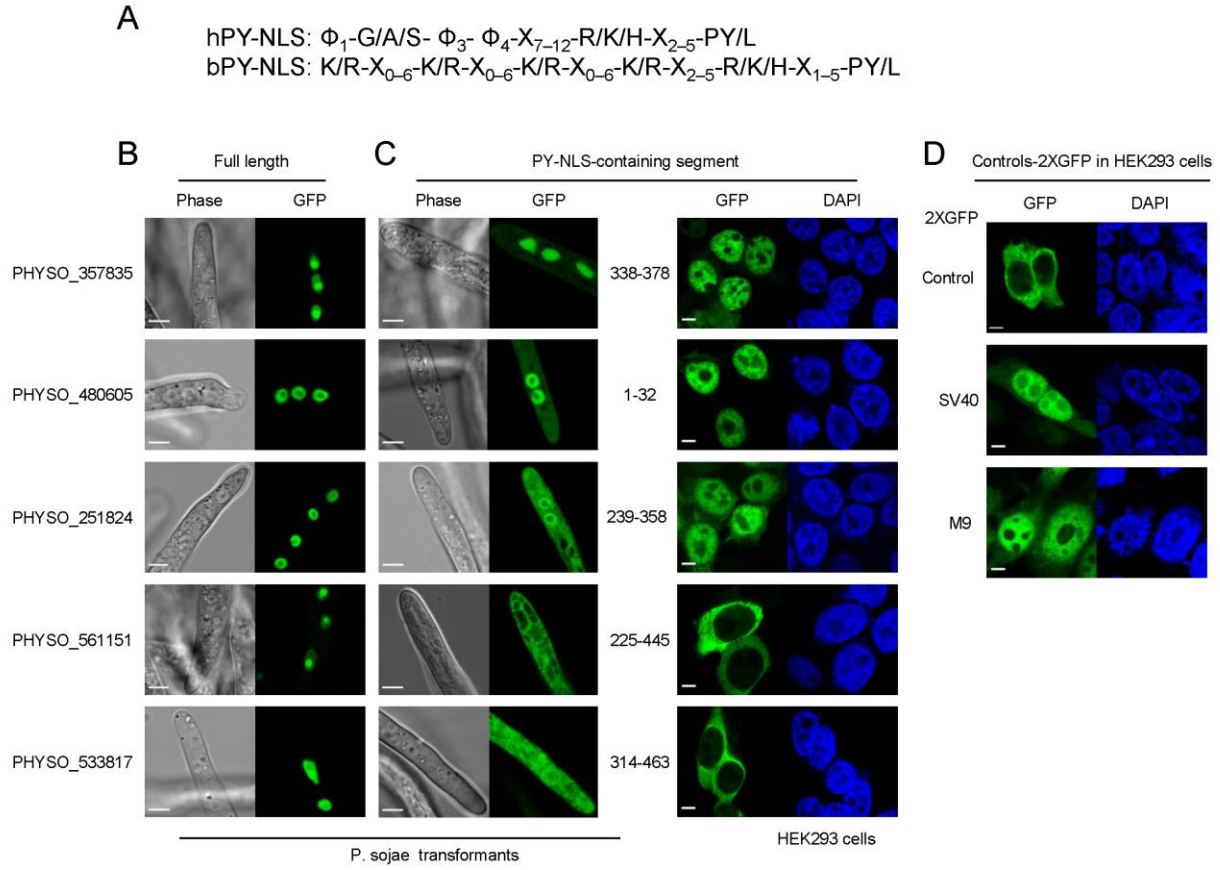


Fig. 2.3 Subcellular localization produced by five protein segments containing candidate PY-NLS motifs.

A. The modified consensus sequence used for searching for candidate PY-NLSs. Φ1 represents a hydrophobic residue (defined here as L, M, I, V, W, Y, H, A, P, or F) while Φ3 and Φ4 represent either hydrophobic residues or R or K. / indicates alternative residues. X_{m-n} indicates any number of unspecified residues from m to n in number.

B. Nuclear localization of the full-length proteins PHYSO_357835, PHYSO_480605, PHYSO_251824, PHYSO_561151 and PHYSO_533817 in *P. sojae* transformants. Representative images are shown in B-D.

C. Subcellular localization of PY-NLS-containing segments in *P. sojae* (left) and HEK 293 cells (right).

D. HEK 293 cells expressing 2XGFP alone, SV40 cNLS-2XGFP, and M9NLS-2XGFP, as controls.

It has been reported that the tyrosine in the C-terminal PY motif shows degeneracy in yeast (Suel *et al.*, 2008). However, since yeast Kap104 only recognizes bPY-NLS motifs, the degeneracy of tyrosine in hPY-NLS is still unclear. Thus, we reasoned that the 'PL' rule for the bPY-NLSs may also apply to hPY-NLS. By adding the 'PL-rule' in the NLS search (Fig. 2.3A), we obtained another 82 hPY-NLS candidates having C-terminal 'PL' motifs. One candidate (PHYSO_480605, annotated as an mRNA maturation protein) was selected for further analysis because the PY-NLS-like sequence was the only predicted NLS in the protein (Fig. 2.3B). Another candidate, PHYSO_251824, had already been selected for testing because it also contained a separate conventional PY-NLS motif.

To validate subcellular localization of the proteins, the 13 full-length proteins were tagged with GFP and expressed in *P. sojae* transformants. Five candidates showed strong nuclear localization in *P. sojae* hyphae at steady state, including PHYSO_480605 (Fig. 2.3B and Table 2.1). PHYSO_561151 and PHYSO_357835 also showed nucleolar localization (Fig. 2.3B). (Although initially selected as a PY-NLS candidate, PHYSO_561151 proved to contain a bipartite cNLS, as described above, and its predicted PY-NLS proved to be inactive, as described below). Other candidates either could not be PCR amplified, contained incorrect intron annotations causing frame shifts, or appeared toxic when overexpressed in *P. sojae* transformants (data not shown).

To determine whether the putative PY-NLSs from the nuclear-localized proteins were sufficient to mediate import of non-nuclear proteins, protein segments containing the motifs were fused to 2XGFP at either the N- or C-terminus, based on their positions in the native proteins. One PY-NLS-containing segment, PHYSO_357835$_{338\text{-}387}$, mediated very efficient nuclear targeting resulting in the reporter protein localized predominantly to the nucleus (Fig. 2.3C).

Other candidates, such as PHYSO_480605$_{1-32}$ and PHYSO_251824$_{239-358}$ showed incomplete nuclear accumulation with some remaining cytoplasmic signals (Fig. 2.3C). In contrast, the PY-NLS-containing segments PHYSO_561151$_{225-445}$ and PHYSO_533817$_{314-463}$ were not capable of transporting 2XGFP into the nucleus (Fig. 2.3C). As noted above, nuclear localization of PHYSO_561151 proved to be mediated by a bipartite cNLS at its C-terminus, while the nuclear localization of PHYSO_533817 was determined by an unidentified sequence between residues 172-314 (Fig. S2.3)

Table 2.1 Function of PY-NLSs predicted in *P. sojae* nuclear localized proteins.

FungiDB ID	Annotation	Core PY- NLS-like sequence	Type[2]	Nuc[3]
PHYSO_357835	U3 small nucleolar RNA –associated protein	359 *PAPA*DYTVATTRHK**R**IQ**PY** 377	h	√
PHYSO_480605	mRNA cleavage and polyadenylation specificity factor subunit 3	1 *MSKR*RLAEEAADER**H**IMRIM**PL** 22	h, v	√
PHYSO_251824	mRNA cleavage and polyadenylation factor I complex, subunit RNA15	259 *PAPA*PAKSGGTRWSA**R**PG**PL** 278 332 *RDPRRAGRDPR*LA**K**R**PY** 348	h, v b	√ √
PHYSO_561151	Homeodomain-like transcription factor	329 *RGVEQQLKKVAVRADPK*RK**K**ELADV**PY** 355	b	×
PHYSO_533817	C2H2 zinc finger protein	333 RTFKKEDARRQHQLAKHG**K**D**PL** 354	b, v	×

[1] Epitope 1 that determines the hydrophobic or basic subclasses of PY-NLSs is highlighted in italics. The R/K/H-X_{2-5}-PY/L motif is in bold and underlined.

[2] PY-NLS types: b, basic; h, hydrophobic; v, variant PL epitope

[3] √ contributes to nuclear localization; ×, does not contribute

Because four of the regions carrying candidate PY-NLSs showed weak or non-existent NLS activity, it was unclear if the problem was the general reliability of the three PY-NLS prediction rules or whether the *P. sojae* import machinery did not efficiently utilize the predicted

PY-NLS motifs. To address this question, parallel experiments were carried out in human embryonic kidney 293 cells (HEK 293) to determine the activity of the putative PY-NLSs in human cells. As shown in Fig. 2.3C and D, the subcellular localization of the PY-NLS-containing truncations in HEK 293 cells were well correlated with their localizations in *P. sojae*, except for segment PHYSO_480605$_{1\text{-}32}$ that showed much stronger nuclear targeting in the human cells.

Next, a series of truncations and mutations were made to more precisely define the roles of the predicted PY-NLS motifs in the nuclear localization of each of PHYSO_357835, PHYSO_480605 and PHYSO_251824. These analyses are detailed in the next three sections.

2.3.5 An augmented PY-NLS sequence in PHYSO_357835 is necessary and sufficient for nuclear import

The aforementioned experiments indicated that residues 338-387 of PHYSO_357835, containing a predicted PY-NLS sequence, were sufficient to mediate import of the 2XGFP reporter into the *P. sojae* nucleus. However, examination of the PHYSO_357835 protein sequence revealed the presence of another candidate NLS within this segment, namely a monopartite cNLS-like sequence (370-RHKR-373) overlapping with the epitope 2 of the PY-NLS motif (Fig. 2.4A). To test whether nuclear transport by PHYSO_357835$_{338\text{-}387}$ required the putative cNLS separately from the epitope 2 of the PY-NLS-like sequence, histidine at 371 and lysine at 372 were mutated to alanines, resulting in 370-RAAR-373 (the canonical PY-NLS consensus requires only a single basic residue at this position). Although 2XGFP-PHYSO_357835$_{338\text{-}387}$(H371A/K372A) appeared slightly more cytoplasmic than wild type

(LNC=1.82 compared to 2.42), the reporter remained primarily nuclear, suggesting that the nuclear localization of PHYSO_357835$_{338\text{-}387}$ was not primarily dependent on the putative cNLS.

To confirm whether the predicted PY-NLS sequence within PHYSO_357835$_{338\text{-}387}$ was the only NLS that determined the import of PHYSO_357835$_{338\text{-}387}$ into the nucleus, amino acid substitutions were further introduced at the key residues in the predicted PY-NLS sequence. As shown in Fig. 2.4B, mutation of the PY residues (P376A/Y377A) reduced but did not eliminate nuclear entry (LNC = 1.37). However, when the cNLS mutations H371A/K372A were combined with the PY mutations, nuclear accumulation by PHYSO_357835$_{338\text{-}387}$ was abolished (LNC=0.26; Fig. 2.4B), suggesting that the PY motif, in combination with the cNLS, is required for the nuclear import of PHYSO_357835$_{338\text{-}387}$. To further explore the role of the basic region (370-RHKR-373), all the four basic residues were converted to alanines. This mutation also abolished the nuclear accumulation of the GFP reporter (LNC=-0.02), indicating that the full set of four residues of this basic region is essential for nuclear import. Together these results suggest that both the predicted PY-NLS (defined by the residues PY in combination with at least 1 of RHKR) and the predicted cNLS (defined by all four of RHKR) are required for efficient nuclear localization of this *P. sojae* protein.

To test the role of this cNLS-augmented PY-NLS sequence in import of full length PHYSO_357835, mutations in the basic region (R370A/H371A/K372A/R373A) alone and combination with the PY dipeptide (R370A/H371A/K372A/R373A/P376A/Y377A) were introduced into full-length GFP-tagged protein. Mutation of the basic region resulted in substantial mis-localization of the full length protein into the cytoplasm (from LNC of 5.51 to 1.33), and additional mutation of the PY motif further decreased the LNC value (to 0.56, not significantly different than the 2XGFP control) (Fig. 2.4C). Together, these results indicate that

the augmented PY-NLS sequence of PHYSO_357835 is necessary as well as sufficient for the nuclear transport of this protein.

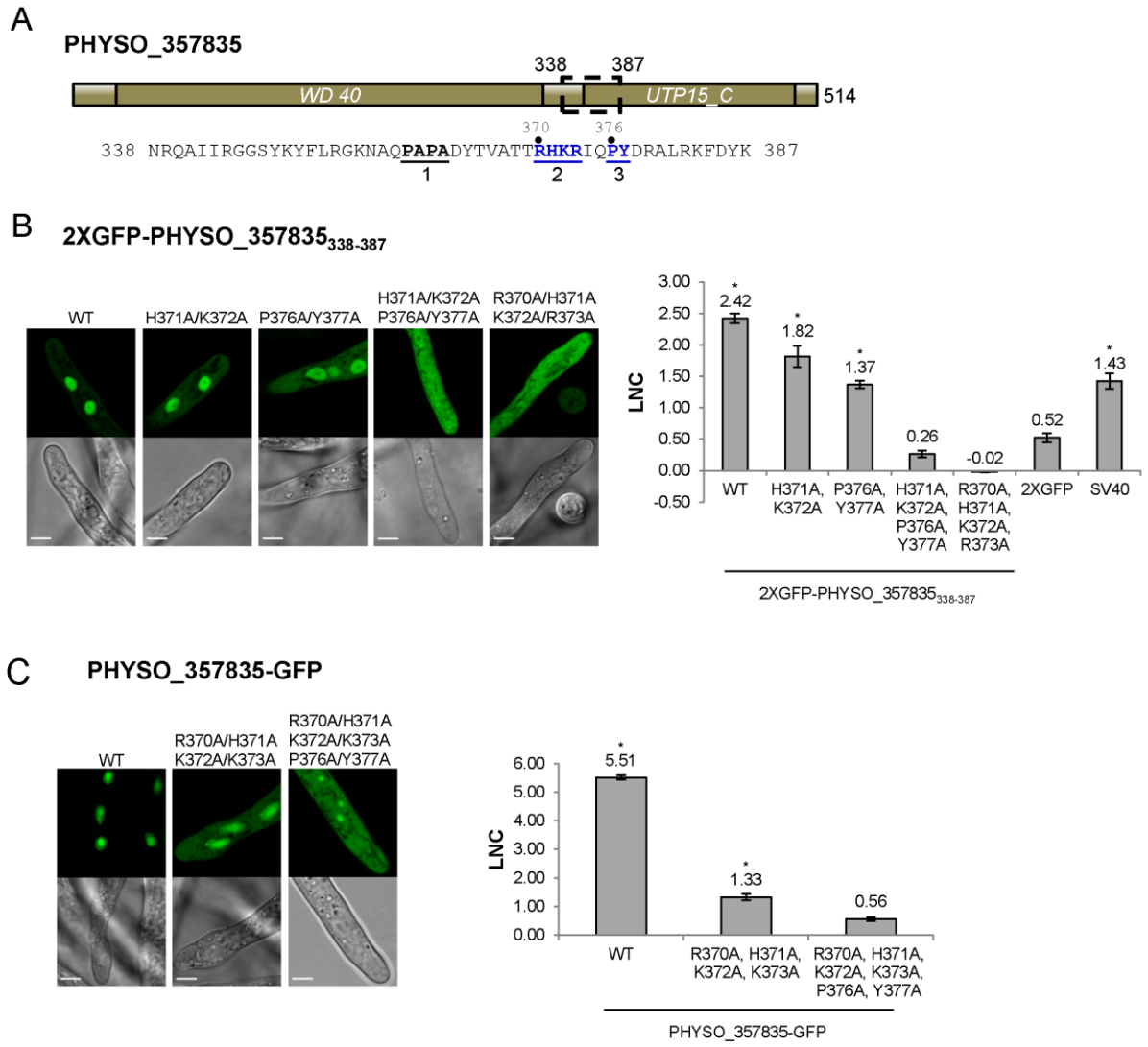


Fig. 2.4 Nuclear import of PHYSO_357835 is mediated by a PY-NLS that incorporates a cNLS.

A. Domain structure of PHYSO_357835. The position of the PY-NLS within PHYSO_357835 is indicated by a dotted rectangle, and the corresponding amino acid sequence is listed below. The three PY-NLS epitopes are underlined and numbered. The PY motif and the basic region corresponding to the predicted cNLS are in blue.

B. Subcellular localization of PHYSO_357835 mutants in the context of the C-terminal domain, 338-387. Left, representative images from *P. sojae* transformants expressing various mutations of 2XGFP-

PHYSO_357835$_{338-387}$. Right, quantification of the localization of 2XGFP-PHYSO_357835$_{338-387}$ fusion proteins.

C. Subcellular localization of PHYSO_357835 mutants in the context of full length PHYSO_357835-GFP. Left, representative images; right, quantitation. The dots observed in *P. sojae* hyphae expressing PHYSO_357835-GFP-(R370A/H371A/K372A/R373A/P376A/Y377A) may be nucleoli as the WT shows substantial nucleolar localization, but this was not verified. The LNC for this mutant was calculated assuming the dots were nucleoli.

2.3.6 Nuclear import of PHYSO_480605 is mediated by a variant PY-NLS

As noted above, the PY-NLS-like sequence located at 1-32 within PHYSO_480605 contains terminal PL residues rather than PY residues. To test if this variant motif is required for PHYSO_480605 nuclear localization, the protein was split at position 32 and each fragment was fused to 2XGFP and expressed in *P. sojae*. The N-terminal 32 residues fused to 2XGFP exhibited significant nuclear staining with some visible cytoplasmic signal (LNC=1.48; Figs. 2.3C and 2.5B), while the C-terminal fragment (residues 33-754) was exclusively distributed in the cytoplasm (Fig. 2.5B). These results indicated that the fragment containing the variant PY-NLS was necessary for the nuclear import of PHYSO_480605, but suggested that additional amino acids downstream of the motif may contribute to the strength of the nuclear targeting (LNC of the full length protein was 5.79). In fact, an expanded fragment (1-60) that includes a stretch of positively charged amino acids (36-KFKGK-40) showed significantly increased nuclear localization (from LNC = 1.48 to 2.18, Fig. 2.5B). However, localization was still much less than the full length protein (LNC=5.79), suggesting that additional downstream sequences might augment nuclear localization, despite being insufficient to independently direct localization.

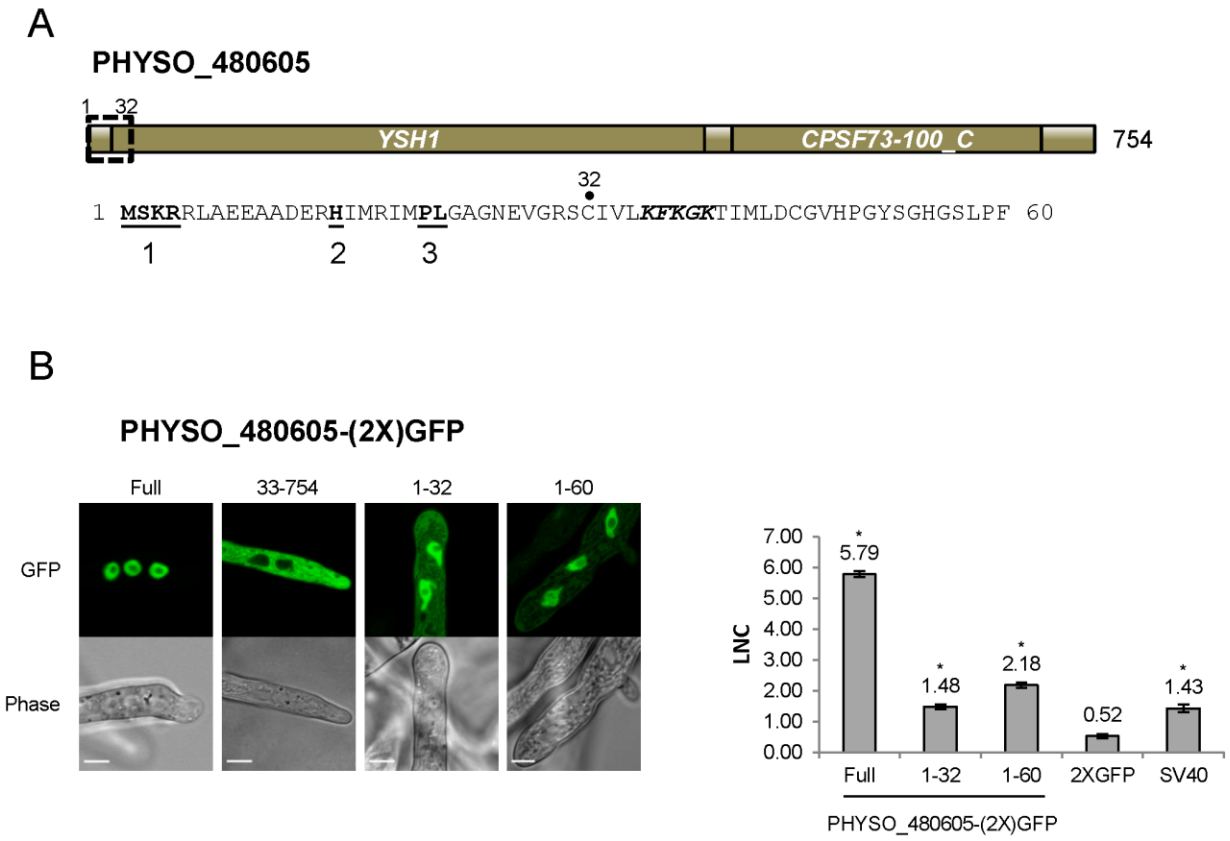


Fig. 2.5 Nuclear localization of PHYSO_480605 requires a region containing a PY-NLS with a variant PY motif.

A. Domain structure of PHYSO_480605. Position of the predicted PY-NLS is indicated by a dotted rectangle and the corresponding sequence is listed below. The three PY-NLS epitopes are underlined. The basic patch corresponding to a predicted cNLS is in bold and italics.

B. Subcellular localization of 2XGFP with full length PHYSO_480605 or fragments of it in *P. sojae* transformants. Left, representative images; right, quantification.

2.3.7 Nuclear localization of PHYSO_251824 requires collaboration of three distinct NLS-like sequences within the C-terminus

As noted above, residues 239-358 of PHYSO_251824 produced only weak nuclear localization (LNC=0.90; Figs. 2.3C and 2.6C). To better identify the sequences required for localization, we

generated fusions containing larger protein fragments of PHYSO_251824. The entire N-terminal (1-238) and C-terminal (239-419) segments of PHYSO_251824 were initially expressed as fusions with 2XGFP in *P. sojae* transformants. PHYSO_251824$_{1-238}$-2XGFP showed strictly cytoplasmic distribution (LNC = -0.52), despite possessing a putative monopartite cNLS (cNLS1, 100-RKRH-103) (Fig. 2.6A and B). In contrast, 2XGFP-PHYSO_251824$_{239-419}$ produced predominant nuclear localization (LNC= 5.30 compared to 5.38 for the full length protein). Comparison of the localization of 2XGFP-PHYSO_251824$_{239-419}$ (LNC=5.30) and 2XGFP-PHYSO_251824$_{239-358}$ (LNC=0.90) indicated that residues 359-419 must contribute to NLS function (Fig. 2.6C). Examination of those residues revealed another predicted monopartite cNLS (363-PSKRSKP-369, cNLS2) (Fig. 2.6A). To test whether cNLS2 was necessary for nuclear accumulation by PHYSO_251824$_{239-419}$, the basic amino acids in cNLS2 were all substituted to alanines (K365A/R366A/R368A); this mutation dramatically reduced the nuclear signal of 2XGFP-PHYSO_251824$_{239-419}$ (LNC=0.61 versus 5.30; Fig. 2.6B and C). To test whether cNLS2 was sufficient for nuclear localization, residues 363-369 were fused to 2XGFP; however these residues alone were not sufficient to direct 2XGFP into the nucleus (LNC= 0.31; Fig. 2.6B and C). Thus, cNLS2 was revealed to be a non-autonomous contributor to nuclear localization by PHYSO_251824$_{239-419}$.

In addition to cNLS2, residues 239-419 of PHYSO_251824 contain two predicted PY-NLS motifs. PY-NLS1 is located at positions 239-288 and has a variant PY motif (PL), while PY-NLS2 is a conventional PY-NLS located 309-358 (Fig. 2.6A). To test the contribution of PY-NLS1, residues 239-308 were deleted. The resulting segment 2XGFP-PHYSO_251824$_{309-419}$ showed a somewhat reduced localization (LNC=4.00) compared to 2XGFP-PHYSO_251824$_{239-419}$ (LNC=5.30; Fig. 2.6C). This finding suggested that PY-NLS1 may have weak NLS activity.

In support of this conclusion, extension of residues 239-308 onto PHYSO_251824$_{1-238}$-2XGFP resulted in reappearance of some nuclear signal (LNC=0.46 versus -0.36; Fig. 2.6B and C). To address the contribution of PY-NLS2, specific amino acids were converted into alanines in the PY-NLS2. In the context of residues 239-419, substitution of the PY residues of PY-NLS2 to alanines (P347A/Y348A) partially reduced the nuclear signal (from LNC=5.30 to 2.42), while additional substitutions in the basic epitope (R332A/R336A/R339A/P347A/Y348A) further reduced the nuclear accumulation to LNC of 0.77 (not significantly greater than the 2XGFP control; Fig. 2.6B and C). On the other hand, the PY-NLS2 alone (309-362) was not sufficient to direct 2XGFP into the nucleus (LNC= 0.66; Fig. 2.6B and C), suggesting that it is necessary but not sufficient for efficient nuclear localization.

Taken together, these findings indicate the nuclear accumulation of PHYSO_251824 is determined by the C-terminal region, in which two PY-NLS-like and one cNLS-like sequence operate synergistically to direct proteins into the nucleus.

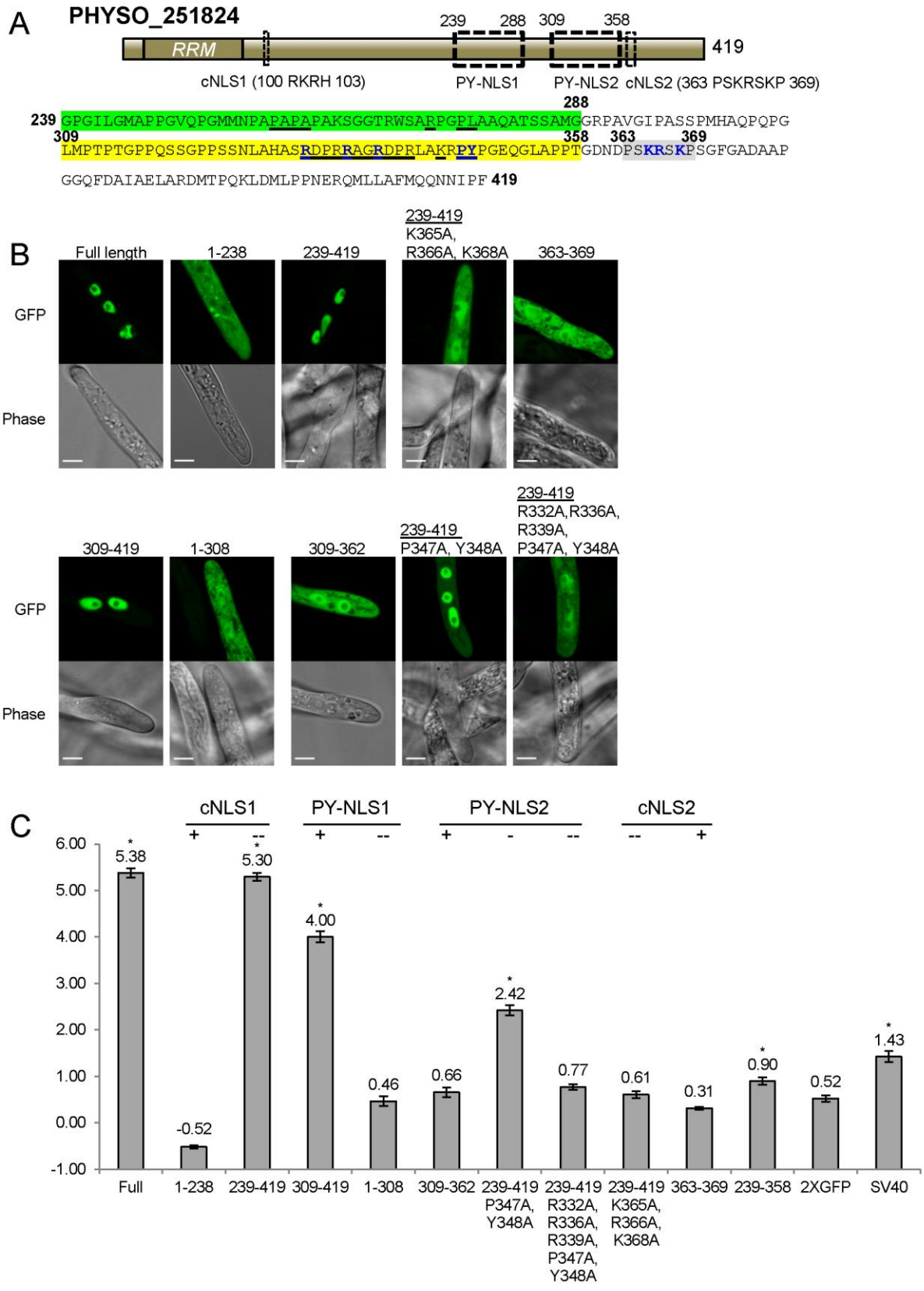

A
PHYSO_251824
239
288
309
358
RRM
419
cNLS1 (100 RKRH 103)
PY-NLS1
PY-NLS2
cNLS2 (363 PSKRSKP 369)
239 GPGILGMAPPGVQPGMMNPAPAPAPAKSGGTRWSARPGPLAAQATSSAMG 288
GRPAVGIPASSPMHAQPQPG
309
358
363
369
LMPTPTGPPQSSGPPSSNLAHASRDPRRAGRDPRLAKRPYPGEQGLAPPTGDNDPSKRSKPSGFGADAAP
GGQFDAIAELARDMTPQKLDMLPPNERQMLLAFMQQNNIPF 419
B
Full length
1-238
239-419
239-419
K365A,
R366A, K368A
363-369
GFP
Phase
309-419
1-308
309-362
239-419
P347A, Y348A
239-419
R332A,R336A,
R339A,
P347A, Y348A
GFP
Phase
C
cNLS1
PY-NLS1
PY-NLS2
cNLS2
+
--
+
--
+
-
--
--
+
6.00
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
5.38
-0.52
5.30
4.00
0.46
0.66
2.42
0.77
0.61
0.31
0.90
0.52
1.43
Full
1-238
239-419
309-419
1-308
309-362
239-419 P347A, Y348A
239-419 R332A, R336A, R339A, P347A, Y348A
239-419 K365A, R366A, K368A
363-369
239-358
2XGFP
SV40

Fig. 2.6 Nuclear accumulation of PHYSO_251824 requires contributions from two PY-NLS and one cNLS clustered within the C-terminus.

A. Domain structure of PHYSO_251824. Four predicted NLSs are indicated by dotted rectangles and their corresponding sequences are shown below. Predicted cNLS1 is inactive. The three epitopes of the two PY-NLS sequences are underlined. Amino acids subjected to mutational analysis are in blue.

B. Subcellular localization of PHYSO_251824-GFP and mutants fused to 2XGFP in *P. sojae* transformants. N-terminal truncations and cNLS2 were fused to 2XGFP at their C-termini, while various C-terminal PHYSO_251824 truncations were fused to 2XGFP at their N-termini. Representative images are shown.

C. Quantification of localization of PHYSO_251824 mutants. Images of PHYSO_251824$_{239\text{-}358}$ were shown in Fig. 2.3B. Mutational statuses of the various NLS candidates are labeled at the top. +, NLS candidate is the only one in the segment; -, NLS candidate is partially mutated; --, NLS candidate is completely mutated.

2.3.8 Highly conserved nuclear-localized proteins show different sequence requirements for nuclear import in *P. sojae* than in human and yeast counterparts

To examine whether *P. sojae* utilizes the same nuclear import sequences for transport of conserved nuclear-localized proteins, we examined ribosomal proteins and core histones. Newly synthesized ribosomal proteins are transported into the nucleus in order to assemble with rRNAs in the nucleolus (Lafontaine & Tollervey, 2001; Gerhardy *et al.*, 2014). Histones, including H2A, H2B, H3, H4, together with the linker histone H1, are essential components of chromatin (Baake *et al.*, 2001).

We first examined the *P. sojae* ribosomal protein PsL28 which is the ortholog of the yeast ribosomal protein L28 (ScL28, former name L29, Underwood & Fried, 1990; the naming system for *P. sojae* ribosomal proteins follows the nomenclature of *S. cerevisiae*; Mager *et al.*, 1997). Protein sequence alignment revealed that the N-termini of the L28 orthologs are highly

conserved among different organisms (Fig. 2.7A). ScL28 was reported to contain two NLSs: ScL28-NLS1, located at amino acid residues 7-13, and ScL28-NLS2 at 24-30 (Underwood & Fried, 1990). The sequence corresponding to NLS1 in PsL28 showed two amino acids differences from ScL28-NLS1, while the NLS2 sequence was exactly the same in all L28 orthologs (Fig. 2.7A). The minimal conserved region of $PsL28_{1-33}$ that contains both NLSs was expressed as a fusion with 2XGFP and showed strong nuclear localization (Fig. 2.7A). In contrast, when each predicted NLS was tested separately by fusion with 2XGFP, neither one produced nuclear localization in *P. sojae* transformants (Fig. 2.7A). However, as indicated above (Fig. 2.1E and F), the two sequences in PsL28 corresponding to NLS1 and NLS2 together constitute an extended bipartite NLS that includes a core bipartite NLS motif (residues 11-27) together with additional flanking positive residues at the N-terminus; this extended bipartite NLS is functional even though the core bipartite NLS ($PsL28_{1-33}$(K7A/R9A/K10A)) is not (Fig. 2.1E and F).

We also tested two other ribosomal proteins, S22a and L3. The conserved NLS containing fragments, residues 21-29 of *P. sojae* S22a (PsS22a) and residues 1-22 of *P. sojae* L3 (PsL3), showed nuclear accumulation in yeast but in *P. sojae* transformants produced only cytoplasmic accumulation (Fig. S2.4); addition of flanking sequences ($PsS22a_{1-34}$ and $PsL3_{1-36}$) had no effect on protein localization either. Thus other or additional sequences appear to be required for nuclear localization of these proteins in *P. sojae*.

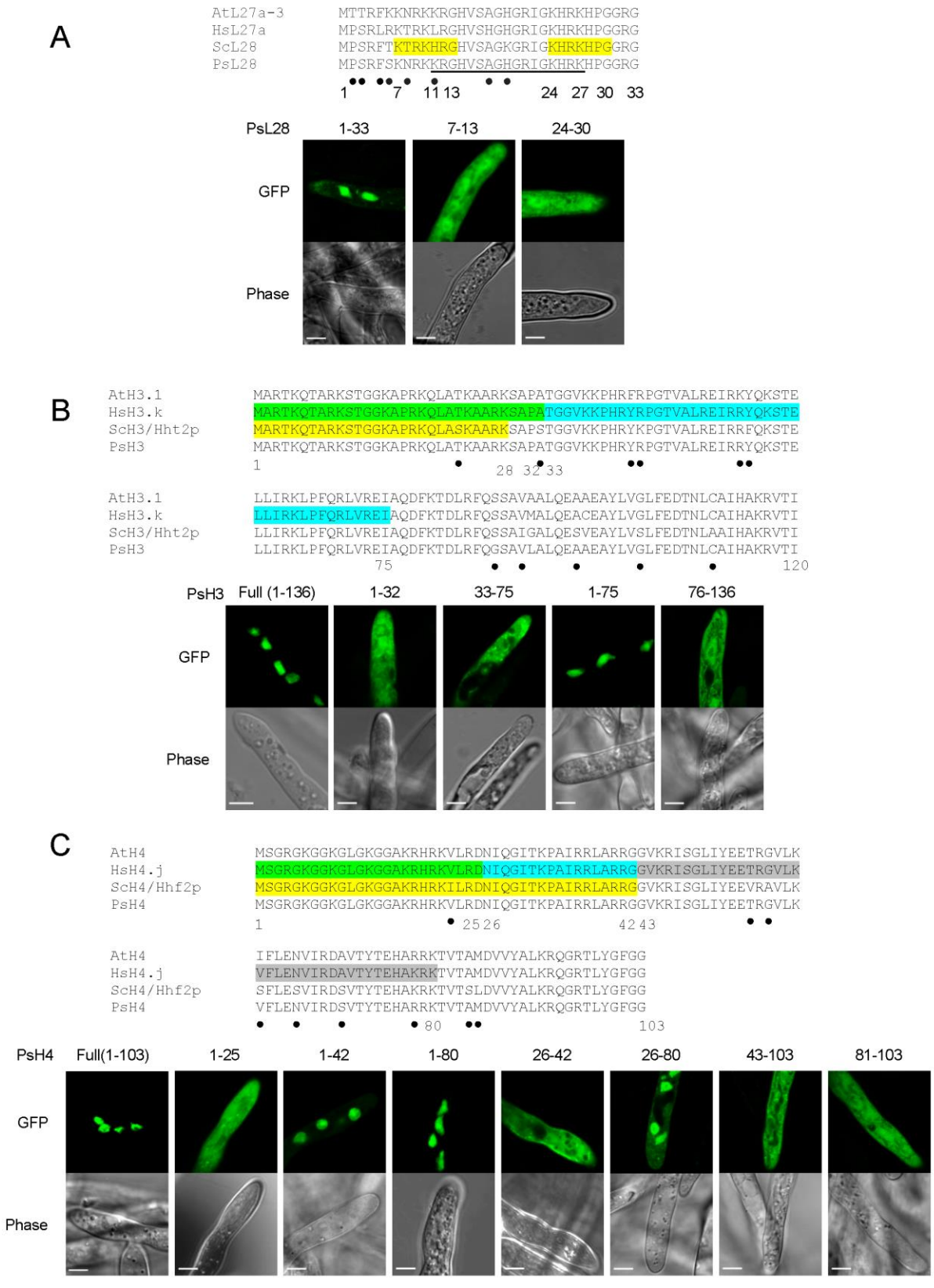

A
AtL27a-3 MTTRFKKNRKKRGHVSAGHGRIGKHRKHPGGRG
HsL27a MPSRLRKTRKLRGHVSHGHGRIGKHRKHPGGRG
ScL28 MPSRFTKTRKHRGHVSAGKGRIGKHRKHPGGRG
PsL28 MPSRFSKNRKKRGHVSAGHGRIGKHRKHPGGRG
1 7 11 13 24 27 30 33
PsL28 1-33 7-13 24-30
GFP
Phase
B
AtH3.1 MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRFRPGTVALREIRKYQKSTE
HsH3.k MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
ScH3/Hht2p MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGTVALREIRRFQKSTE
PsH3 MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
1 28 32 33
AtH3.1 LLIRKLPFQRLVREIAQDFKTDLRFQSSAVAALQEAAEAYLVGLFEDTNLCAIHAKRVTI
HsH3.k LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAKRVTI
ScH3/Hht2p LLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQESVEAYLVSLFEDTNLAAIHAKRVTI
PsH3 LLIRKLPFQRLVREIAQDFKTDLRFQGSAVLALQEAAEAYLVGLFEDTNLCAIHAKRVTI
75 120
PsH3 Full (1-136) 1-32 33-75 1-75 76-136
GFP
Phase
C
AtH4 MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK
HsH4.j MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGVKRISGLIYEETRGVLK
ScH4/Hhf2p MSGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRGVKRISGLIYEEVRAVLK
PsH4 MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK
1 25 26 42 43
AtH4 IFLENVIRDAVTYTEHARRKTVTAMDVVYALKRQGRTLYGFGG
HsH4.j VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG
ScH4/Hhf2p SFLESVIRDSVTYTEHAKRKTVTSLDVVYALKRQGRTLYGFGG
PsH4 VFLENVIRDSVTYTEHARRKTVTAMDVVYALKRQGRTLYGFGG
80 103
PsH4 Full(1-103) 1-25 1-42 1-80 26-42 26-80 43-103 81-103
GFP
Phase

Fig. 2.7 Combinatorial usage of NLSs for nuclear transport of *P. sojae* ribosomal protein L28 and core histones H3 and H4.

A-C, Upper panels, alignments of *P. sojae* ribosomal proteins L28 (PsL28, PHYSO_355737) and core histones H3 (PsH3, PHYSO_286415) and H4 (PsH4, PHYSO_285922) with their *Arabidopsis thaliana* (At), human (Hs) and *Saccharomyces cerevisiae* (Sc) orthologs. Dots, variable sites among L28, H3 or H4 orthologs. Regions individually conferring nuclear targeting in human core histones are highlighted in green, cyan, or gray while those required in yeast are highlighted in yellow. The bipartite cNLS predicted in PsL28 is underlined. Lower panels, subcellular localization of various PsL28, PsH3 or PsH4 truncations. Representative images are shown.

A similar strategy was carried out to determine the activities of NLSs in core histones H3 and H4, which are nearly identical in amino acid sequences in all eukaryotes (Fig. 2.7B and C). The NLSs in H3 or H4 have been experimentally characterized in human and yeast (Fig. 2.7B and C). There is only one NLS reported in each of yeast H3 (ScH3) and H4 (ScH4) (Mosammaparast *et al.,* 2002), while more than one NLS has been found in their human counterparts (Baake *et al.,* 2001). The NLSs in ScH3 and ScH4 somewhat overlapped with their human orthologs, suggesting a common origin.

A series of truncations were made of *P. sojae* H3 (PsH3) and H4 (PsH4) according to the NLS locations in the yeast and human orthologs. The resulting fragments were tested for nuclear accumulation as 2XGFP fusions. As expected, the full-length PsH3 and PsH4 reporter constructs produced predominantly nuclear localization (Fig. 2.7B and C). However, none of the H3 fragments responsible for nuclear localization in yeast or human individually exhibited the same NLS activities in *P. sojae* transformants. Region 1-75 of PsH3, which encompasses both of the two NLS regions of human H3, did efficiently transport 2XGFP into the nucleus (Fig. 2.7B). Expression of *P. sojae* H3 reporters that lacked the N-terminal domains, i.e. $H3_{(76\text{-}136)}$, did not

produce significant nuclear accumulation (Fig. 2.7B), indicating that no other sequences contribute to PsH3 nuclear localization.

In the case of PsH4, residues 1-25, which showed partial NLS activity in yeast and full activity in human cells, produced no observable nuclear localization in *P. sojae* (Fig. 2.7C). A larger N-terminal GFP fusion $PsH4_{1-42}$-2XGFP produced clear nuclear localization with little detectable fluorescence in the cytoplasm, and residues 1-80 of PsH4 produced complete nuclear localization in *P. sojae* (Fig. 2.7C). This result suggests that residues 26-42 of PsH4, which were identified as an NLS in human cells, contain determinants that may contribute to NLS function in *P. sojae*. However, $PsH4_{26-42}$ alone was insufficient to mediate nuclear targeting of the GFP reporter, indicating that these residues could not act as an independent NLS (Fig. 2.7C). $PsH4_{26-80}$, which spans two regions (26-42 and 43-80) with NLS activity in human, produced some GFP signal in the nucleus but less than 1-80 (Fig. 2.7C), whereas 43-103 showed no activity (Fig. 2.7C). Together, these results suggest that all three regions corresponding to NLSs in human H4, 1-25, 26-42 and 43-80, must work together in *P. sojae* to produce efficient nuclear localization. *P. sojae* H4 reporters that lacked the N-terminal domains ($H4_{81-103}$) produced no nuclear accumulation (Fig. 2.7C).

Taken together, the results above suggest that compared to their yeast or human counterparts, *P. sojae* has a similar but stronger set of sequence requirements for translocation of conserved nuclear proteins into the nucleus.

2.4 DISCUSSION

The diverse sequences and structures of NLSs have limited the reliable prediction of nuclear localized proteins based on amino acid sequence alone. Though consensus sequences have been

proposed for some NLS types, such as cNLS and PY-NLS, their application remains constrained by many factors, such as protein context and flanking sequences and by nuclear import regulation by mechanisms such as phosphorylation and protein interactions (Garcia-Bustos *et al.*, 1991). Moreover, the karyopherins that bind NLSs show differences in specificity from species to species (Marfori *et al.*, 2011), limiting the reliability of NLS consensus sequences when applied to widely divergent species. In this study, after eliminating artifacts caused by staining with DAPI, we have identified key differences in several classes of NLSs between the oomycete *P. sojae* on the one hand, and human and yeast on the other hand.

2.4.1 *P. sojae* classical NLSs exhibit distinct differences from human and yeast

The cNLSs are the best-characterized class of nuclear localization signals in human and yeast to date (Lange *et al.*, 2007). However, we found that monopartite cNLSs, represented by those from the SV40 large T-antigen and *c-Myc*, produced relatively weak nuclear localization in *P. sojae*. We showed that SV40 NLS fusions at the C-terminus of GFP, and multiple copies of the NLS increased nuclear accumulation compared to a single copy at the N-terminus of GFP (Fig. 2.1D), although some fluorescence was still visible in the cytoplasm (Fig. 2.1B). It has been reported that protein context may influence the activity of NLSs, and multiple copies of a NLS can produce faster, more pronounced nuclear localization (Garcia-Bustos *et al.*, 1991).

Based on assessment of several putative cNLSs predicted from the consensus sequence in *P. sojae* nuclear proteins, we found that cNLS-mediated nuclear import is conserved in *P. sojae* but that it exhibits some consistent differences: (1) most predicted monopartite cNLSs are not sufficient by themselves for efficient nuclear accumulation of GFP reporters. However, they may function to augment the activity of other NLSs or work collectively with other weak NLSs to

accomplish efficient nuclear import of proteins. For example, the monopartite cNLS embedded inside the PY-NLS of PHYSO_357835 was essential for efficient nuclear targeting (Fig. 2.4), while in PHYSO_251824 a C-terminal monopartite cNLS cooperated with two PY-NLSs to direct efficient nuclear localization (Fig. 2.6). (2) With regard to bipartite cNLSs, sequence comparisons and mutational analyses of several functional bipartite cNLSs, including nucleoplasmin, PsL28 and PHYSO_561151, revealed that a functional bipartite cNLS in *P. sojae* typically requires additional basic amino acids in either sub-motif, compared to the human/yeast consensus (Fig. 2.1E). In the case of PsL28, both of the two sub-motifs (KNRKKR and KHRKH) contain sufficient basic residues to be considered monopartite cNLSs by themselves, and actually function that way in yeast. But in *P. sojae*, the two must function together as a bipartite cNLS in order to direct nuclear targeting.

2.4.2 Efficient PY-NLS-mediated nuclear import requires additional clusters of basic amino acids

None of the human and yeast PY-NLSs we tested were capable of mediating efficient nuclear entry of GFP reporters in *P. sojae*, though weak NLS activity was observed with the hnRNP M and the chimeric M9M NLSs. Given that the M9M peptide is a hPY-NLS (Fig. 2.2A), our findings suggest that, in contrast to yeast, the *P. sojae* nuclear transport machinery may have evolved to transport both bPY-NLS and hPY-NLS cargos.

To characterize *P. sojae* PY-NLSs, we analyzed five native *P. sojae* nuclear localized proteins that harbored predicted PY-NLS-like motifs. Using deletion and substitution mutations, we identified PY-NLS-containing fragments from three of the proteins that could mediate different degrees of GFP translocation into the nuclei of both *P. sojae* and mammalian cells. In

each case however, additional basic amino acids were required, either in association with one of the three PY-NLS epitopes (as in PHYSO_357835), or C-terminal to the PY-NLS motif (PHYSO_480605 and PHYSO_251824).

In yeast, Suel *et al.* (2008) proposed the 'degeneracy rule' for the bPY-NLS in which PL is exchangeable with PY. However, yeast Kapβ2 only recognizes the basic class of PY-NLSs, so PY-motif degeneracy remains unclear for the hydrophobic class of PY-NLSs. Only a few PY-NLSs are reported to have a variant PY motif, such as the RNA-binding protein HuR which contains a PG motif instead of PY. In addition, several newly identified Kapβ2/Kap104 cargos were reported to lack a typical PY motif or even a recognizable PY-NLS (reviewed in Soniat & Chook, 2015). In our study, we found that the variant 'PL'-NLS in PHYSO_480605 produced substantial nuclear localization in both *P. sojae* and mammalian cells, suggesting the degeneracy rule may also exist for hPY-NLS.

The PY-NLS 'consensus sequence' is defined by a collection of three modular epitopes: an N-terminal hydrophobic or basic residue-enriched motif, a second R/K/H motif and the third P[Y/L] motif (Lee *et al.*, 2006; Suel *et al.*, 2008). Biochemical and biophysical analyses have shown that in different NLSs each epitope may contribute differently to Kapβ2 (or Kap104p) binding (Suel *et al.*, 2008; Lee *et al.*, 2006; Cansizoglu *et al.*, 2007). For instance, in hnRNP M and in the yeast mRNA processing protein Hrp1, epitope 3 contributes significantly more than the other two epitopes (Cansizoglu *et al.*, 2007; Suel *et al.*, 2008), In contrast, in the M9NLS and Nab2p, epitope 3 contributes weakly to Kapβ2 (or Kap104p) binding, while strong binding is conferred by epitope 1 or by multiple positions distributed across the three epitopes (Lee *et al.*, 2006; Suel *et al.*, 2008). In our study, we found that mutation of PY (epitope 3) in the PHYSO_357835 PY-NLS sequence reduced but did not abolish nuclear entry, whereas complete

abrogation of nuclear entry was observed upon mutation of the basic cluster overlapping epitope 2. In the case of the PY-NLSs in PHYSO_480605 and PHYSO_251824, an extra basic cluster downstream of the PY-NLS consensus markedly enhanced the nuclear entry, effectively constituting a fourth epitope (Fig. 2.5 and Fig. 2.6). These results suggest that extra positive charges may facilitate the binding of PY-NLSs to the presumptive *P. sojae* Kapβ2. Similar combinatorial organization of a PY-NLS was observed in the *Xenopus* Kapβ2 substrate ELYS (RRTRRRIIAKPVTRRKMR), in which a variant PY dipeptide (PV) is flanked by a C-terminal basic cluster (Lau *et al.*, 2009). Thus, our results may explain why the four PY-NLS prototypes from human and yeast functioned poorly in *P. sojae*, as they lack the additional basic residues required for efficient targeting to the *P. sojae* nucleus.

2.4.3 For nuclear import of highly conserved ribosomal and histone proteins, *P. sojae* requires combinations of NLSs that are autonomous in other eukaryotes

Like the nuclear entry mediated by the *P. sojae* cNLS and PY-NLS motifs, entry by highly conserved nuclear proteins also showed distinctive features in *P. sojae*. Amino acid sequences of many ribosomal proteins and core histones are highly conserved among different eukaryotes, though within an organism different conserved proteins are imported into the nucleus via different NLSs. This suggests that the responsible karyopherins from different organisms should show similar specificity. However in *P. sojae,* NLS motifs that could act autonomously in human and yeast were required to act together to deliver conserved proteins into the nucleus. For example, in yeast L28, two sequences (7-KTRKHRG-13 and 24-KHRKHPG-30) could individually serve as an NLS. However the corresponding sequences in *P. sojae* L28 were required together for efficient nuclear localization (thus constituting a bipartite NLS). In the case

of histone H3, where the amino acid sequences are nearly identical in all eukaryotes, two adjacent sequences that could each serve as an NLS in human (residues 1-32 and 33-75 of H3) were jointly required for NLS activity in *P. sojae*. Similarly, in H4 three adjacent sequences that could serve separately as NLSs in human were required together in *P. sojae* H4 for nuclear entry. The results suggest that the *P. sojae* karyopherins that presumably bind to these NLSs do so more weakly than in other characterized eukaryotes, and perhaps therefore that binding by multiple karyopherins is required to produce nuclear entry in *P. sojae*.

In summary, using *P. sojae* as a model, we have identified distinctive features of nuclear localization in oomycetes, based on characterization of multiple classes of NLS from 10 nuclear proteins. A consistent pattern has emerged in which individual NLS sequences that are sufficient for autonomous nuclear localization in other eukaryotes function weakly or not at all in *P. sojae*, but can collaborate with each other or with patches rich in basic residues to produce efficient localization. Because nuclear localization has not been dissected comprehensively in a wide diversity of eukaryotes or even in other oomycetes, it is currently unclear whether the differences we observe in *P. sojae* are part of a wider pattern of diversity in eukaryotes, or have a narrower functional significance, related perhaps to the pathogenic lifestyle of this organism.

2.5 MATERIALS AND METHODS

2.5.1 *P. sojae* strains and growth conditions

All of the NLS tests were carried out in the *P. sojae* reference isolate P6497 (race 2). Cultures were routinely grown and maintained in cleared V8 medium at 25 ℃ in the dark. *P. sojae*

transformants were incubated in 12-well plates containing V8 media supplemented with 50 μg ml^{-1} G418 (Geneticin) for 2-3 days before examination by confocal microscopy.

2.5.2 Sequence analysis

Phytophthora protein sequences and IDs were obtained from FungiDB (Stajich *et al.*, 2011). Their corresponding accession numbers in the GenBank database are listed in Table S2. Ribosomal protein and core histone sequences of *Arabidopsis*, yeast and human were obtained from the GenBank database: *Arabidopsis* L27a (accession number, NP_177217), H3.1 (NP_201339), H4, (NP_190179); human H3.k (P68431), H4.j (AAA52652), L27a (NP_000981); *Saccharomyces cerevisiae* L28 (NP_011412), H3 (or Hht2p, CAY82162); H4 (or Hhf2p, NP_014368). Protein sequence alignments were carried out using *Clustal Omega* (Sievers *et al.*, 2011). Additional sequence information for the proteins PHYSO_357835, PHYSO_480605, PHYSO_251824, PHYSO_561151, and PHYSO_533817 can be found in the *Supplemental Sequences*.

2.5.3 Construction of plasmids

All the primers used in this study are listed in Table S3 in the supplemental material. All the fusion protein reporter constructs in which the NLS was fused to the N-terminus of GFP or 2XGFP were based on the plasmid backbone pYF2-2XGFP (Fang & Tyler, 2016). To test reporters with the NLS fused to the C-terminus of GFP, a new plasmid backbone pYF3-2XGFP was generated by inserting an *eGFP* fragment with attached C-terminal multiple cloning sites and stop codon (*Bsr* GI-*Hpa* I-*Bsp* EI-*Mlu* I-TAA) into the *Afl* II and *Apa* I sites of the backbone pYF2-GFP. To clone NLS candidates efficiently and economically, NLSs derived from

exogenous genes smaller than ~80 bp were created by oligo annealing and inserted into the *Spe* I and *Sac* II sites of pYF2-2XGFP, or the *Bsr* GI and *Bsp* EI sites of pYF3-2XGFP. PCR products from *P. sojae* genes were inserted into the *Stu* I site of pYF2-2XGFP or the *Hpa* I site of pYF3-2XGFP by blunt ligation. The *P. sojae* H2B nuclear marker fusions (H2B-GFP and H2B-mCherry) and the *P. capsici* fibrillarin nucleolar marker fusion (FIB-mCherry) were created using the pGFPN or pMCherryN plasmids (Ah-Fong & Judelson, 2011). Point mutations and deletions of DNA sequences were made through QuikChange Lightning Multi Site-Directed Mutagenesis (Agilent Technologies).

To express 2XGFP fusions in mammalian cells, genes encoding 2XGFP and its SV40 cNLS and M9 NLS fusions were PCR-amplified from *P. sojae* expression plasmids using primers pYF2_GW_F and GFP_stop_GW_R, and integrated into pcDNA 3.2/V5-DEST by Gateway™ cloning (Thermo Scientific). Integration of 2XGFP into pcDNA 3.2/V5-DEST also generated a plasmid backbone pcDNA 3.2-2XGFP, in which a preserved *Eco* RV site originating from the pYF2 backbone was used for insertion of PY-NLS amplicons by blunt ligation.

PCR-amplification was conducted using Phusion® High-Fidelity DNA Polymerase (NEB). Standard molecular techniques were performed as described by Sambrook and Russell (2001) or according to instructions from kit manufacturers.

2.5.4 Transient expression assays in *P. sojae* transformants

To assess the localization of various protein fusions, we used a transient assay in which mixed cultures of transformed *P. sojae* hyphae were examined within days after the transformation procedure. Polyethylene glycol (PEG)-mediated protoplast transformations were conducted by an improved protocol as described previously (Fang & Tyler, 2016). In general, *P. sojae*

protoplasts were isolated by enzyme digestion using 0.5% Lysing Enzymes from *Trichoderma harzianum* (Sigma L1412) and 0.5% CELLULYSIN® Cellulase (Calbiochem 219466) in 0.4 M mannitol, 20 mM KCl, 20mM MES, pH 5.7, 10 mM $CaCl_2$. 1 ml of protoplasts at a density of $2x10^6$-$2X10^7$ ml^{-1} and 20-30 μg DNA were used for single plasmid transformations. For co-transformation experiments, 20-30 μg of the plasmid carrying the *NPT II* selectable marker gene was used, together with an equimolar ratio of any other DNAs included. Protoplasts were regenerated in pea broth containing 0.5 M mannitol and propagated in V8 liquid media containing 50 μg ml^{-1} G418 for 2~3 d at 25 ℃ prior to analysis.

2.5.5 Human cell line transfection and immunocytochemistry

HEK293 cells were maintained in DMEM media (Mediatech) supplemented with 10% fetal bovine serum (Tissue Culture Biologicals), penicillin-streptomycin solution (1 IU ml^{-1} penicillin and 100 ug ml^{-1} streptomycin, Mediatech). Transfection was done according to the manufacturer's protocol. Briefly, HEK293 cells were plated on a glass coverslip in a 24-well plate (Greiner Bio-One) at a density sufficient for 70% confluency at the time of transfection. Cells were transfected with 500 ng DNA and 1 μl of Lipofectamine 2000 (Invitrogen) per well for 48 h. For immunocytochemistry, the transfected cells were fixed with 3.7% (w/v) paraformaldehyde in PBS for 15 min at RT. The fixed cells were permeabilized with 0.1% Triton X-100 in PBS for 10 min at RT and incubated with 1% bovine serum albumin in PBS for 1 h at RT. Then, the cells were incubated with rabbit anti-GFP antibody conjugated with AlexaFluor 488 (Invitrogen) overnight at 4 ℃. Cells on the coverslip were mounted on a glass slide using ProLong Gold antifade with DAPI (Invitrogen) and dried overnight at RT in darkness.

2.5.6 Confocal microscopy imaging

Laser scanning confocal microscopy (Zeiss LSM 780 NLO) was used to monitor the subcellular localization of fluorescent protein fusions in *P. sojae* and human cell transformants. Clumps of 2-3 days old transformed *P. sojae* mycelia grown in liquid V8 media were removed with a toothpick, washed and maintained in modified Plich media (0.5 g KH_2PO_4, 0.25 g $MgSO_4 \bullet 7H_2O$, 1 g Asparagine, 1 mg Thiamine, 0.5 g Yeast extract, 10 mg β–sitosterol, 25 g Glucose dissolved in 1 liter water) before observation. For DAPI staining, mycelia samples were pre-treated with PBS containing 0.2 μg ml^{-1} DAPI for 25 min in dark, and washed twice with PBS according to Hardham (2001). Images were captured using a 63X oil objective with excitation/emission settings (in nm) of 405/410-490 for DAPI, 488/504-550 for GFP, and 561/605-650 for mCherry. For each sample, at least three independent *P. sojae* transformants were analyzed. Images were adjusted and quantitated using the microscope's built-in Zen 2012 software (Blue and/or Black edition according to different purposes). Images were cropped, and the tonal range was increased by adjusting highlights and shadows without altering the color balance. Nuclear and cytoplasmic intensities were collected by manually defining the nuclear and cytoplasmic regions in each image as described in Hunter *et al.* (2014). In most cases, nuclei were identified by the morphology of the GFP-stained region (a large, uniformly stained, irregular ovoid region, often with an unstained nucleolus), and were not confirmed by co-expression of H2B-mCherry. Where nuclei could not be identified due to very poor nuclear localization, H2B-mCherry was co-expressed to verify the nuclei. The intensity mean values of defined regions were generated by the Zen 2012 (Blue edition) 'measure' tool automatically, which ignored any adjustments of tonal range. The nuclear to cytoplasmic fluorescence ratio was calculated using the means of the $\log_2$ transformed ratios from 30 randomly paired nuclear and cytoplasmic regions, allowing the

standard error of the mean (S.E.) to be calculated. Significance was tested with a two-sample, unpaired t-test of the log ratios with a *P* value cutoff of 0.01, performed by the software GraphPad Prism 7.

2.6 ACKNOWLEDGEMENTS

We thank Anne-Marie Girard (Center for Genome Research and Biocomputing, Oregon State University) for assistance and training with confocal microscopy, Felipe Arredondo (Oregon State University) for assistance and training with *P. sojae* transformation, Danyu Shen (Nanjing Agriculture University) and Tian Hong (University of California, Irvine) for assistance with bioinformatics, and Howard Judelson (University of California, Riverside) for plasmids. This work was supported in part by USDA NIFA grant #2011-68004-30104 to B.M.T.

2.7 REFERENCES

Ah-Fong, A.M. & H.S. Judelson, (2011) Vectors for fluorescent protein tagging in *Phytophthora*: tools for functional genomics and cell biology. *Fungal Biol* **115**: 882-890.

Baake, M., M. Bauerle, D. Doenecke & W. Albig, (2001) Core histones and linker histones are imported into the nucleus by different pathways. *Eur. J. Cell Biol.* **80**: 669-677.

Bonifaci, N., J. Moroianu, A. Radu & G. Blobel, (1997) Karyopherin beta 2 mediates nuclear import of a mRNA binding protein. *Proc. Natl. Acad. Sci. USA* **94**: 5055-5060.

Cansizoglu, A.E., B.J. Lee, Z.C. Zhang, B.M. Fontoura & Y.M. Chook, (2007) Structure-based design of a pathway-specific nuclear import inhibitor. *Nat. Struct. Mol. Biol.* **14**: 452-454.

Chook, Y.M. & K.E. Suel, (2011) Nuclear import by karyopherin-betas: recognition and inhibition. *Biochim. Biophys. Acta* **1813**: 1593-1606.

Dingwall, C. & R.A. Laskey, (1991) Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.* **16**: 478-481.

Dingwall, C., S.V. Sharnick & R.A. Laskey, (1982) A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. *Cell* **30**: 449-458.

Erwin, D.C. & O.K. Ribeiro, (1996) *Phytophthora diseases worldwide*. American Phytopathological Society (APS Press, Minnesota).

Fang, Y. & B.M. Tyler, (2016) Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* **17**: 127-139.

Gamboa-Melendez, H., A.I. Huerta & H.S. Judelson, (2013) bZIP transcription factors in the oomycete *Phytophthora infestans* with novel DNA-binding domains are involved in defense against oxidative stress. *Eukaryot. Cell* **12**: 1403-1412.

Garcia-Bustos, J., J. Heitman & M.N. Hall, (1991) Nuclear protein localization. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1071**: 83-101.

Gerhardy, S., A.M. Menet, C. Pena, J.J. Petkowski & V.G. Panse, (2014) Assembly and nuclear export of pre-ribosomal particles in budding yeast. *Chromosoma* **123**: 327-344.

Hardham, A. R. (2001). Investigations of oomycete cell biology. Molecular and cell biology of filamentous fungi: a practical approach. Oxford University Press, pp. 127-155.

Hunter, C.C., K.S. Siebert, D.J. Downes, K.H. Wong, S.D. Kreutzberger, J.A. Fraser, D.F. Clarke, M.J. Hynes, M.A. Davis & R.B. Todd, (2014) Multiple nuclear localization signals mediate nuclear localization of the GATA transcription factor AreA. *Eukaryot. Cell* **13**: 527-538.

Judelson, H.S. & F.A. Blanco, (2005) The spores of *Phytophthora*: weapons of the plant destroyer. *Nat. Rev. Microbiol.* **3**: 47-58.

Kalderon, D., W.D. Richardson, A.F. Markham & A.E. Smith, (1983) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**: 33-38.

Kalderon, D., B.L. Roberts, W.D. Richardson & A.E. Smith, (1984) A short amino acid sequence able to specify nuclear location. *Cell* **39**: 499-509.

Kamoun, S., O. Furzer, J.D. Jones, H.S. Judelson, G.S. Ali, R.J. Dalio, S.G. Roy, L. Schena, A. Zambounis, F. Panabieres, D. Cahill, M. Ruocco, A. Figueiredo, X.R. Chen, J. Hulvey, R. Stam, K. Lamour, M. Gijzen, B.M. Tyler, N.J. Grunwald, M.S. Mukhtar, D.F. Tome, M. Tor, G. Van Den Ackerveken, J. McDowell, F. Daayf, W.E. Fry, H. Lindqvist-Kreuze,

H.J. Meijer, B. Petre, J. Ristaino, K. Yoshida, P.R. Birch & F. Govers, (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**: 413-434.

Lafontaine, D.L. & D. Tollervey, (2001) The function and synthesis of ribosomes. *Nat. Rev. Mol. Cell Biol.* **2**: 514-520.

Lange, A., R.E. Mills, C.J. Lange, M. Stewart, S.E. Devine & A.H. Corbett, (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**: 5101-5105.

Lau, C.K., V.A. Delmar, R.C. Chan, Q. Phung, C. Bernis, B. Fichtman, B.A. Rasala & D.J. Forbes, (2009) Transportin regulates major mitotic assembly events: from spindle to nuclear pore assembly. *Mol Biol Cell* **20**: 4043-4058.

Lee, B.J., A.E. Cansizoglu, K.E. Suel, T.H. Louis, Z. Zhang & Y.M. Chook, (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **126**: 543-558.

Mager, W.H., R.J. Planta, J.G. Ballesta, J.C. Lee, K. Mizuta, K. Suzuki, J.R. Warner & J. Woolford, (1997) A new nomenclature for the cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. *Nucleic Acids Res.* **25**: 4872-4875.

Makkerh, J.P., C. Dingwall & R.A. Laskey, (1996) Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.* **6**: 1025-1027.

Marfori, M., A. Mynott, J.J. Ellis, A.M. Mehdi, N.F. Saunders, P.M. Curmi, J.K. Forwood, M. Boden & B. Kobe, (2011) Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim. Biophys. Acta* **1813**: 1562-1577.

Mosammaparast, N., Y. Guo, J. Shabanowitz, D.F. Hunt & L.F. Pemberton, (2002) Pathways mediating the nuclear import of histones H3 and H4 in yeast. *J. Biol. Chem.* **277**: 862-868.

PSORT, I., (1997) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *J. Mol. Biol* **266**: 594-600.

Sambrook, J. & D.W. Russell, (2001) Molecular cloning: a Laboratory Manual 3rd edition. *Cold Spring-Harbour Laboratory Press, UK*.

Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert & J. Söding, (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**: 539.

Soniat, M. & Y.M. Chook, (2015) Nuclear localization signals for four distinct karyopherin-beta nuclear import systems. *Biochem. J.* **468**: 353-362.

Stajich, J.E., T. Harris, B.P. Brunk, J. Brestelli, S. Fischer, O.S. Harb, J.C. Kissinger, W. Li, V. Nayak & D.F. Pinney, (2011) FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res.*: gkr918.

Suel, K.E. & Y.M. Chook, (2009) Kap104p imports the PY-NLS-containing transcription factor Tfg2p into the nucleus. *J. Biol. Chem.* **284**: 15416-15424.

Suel, K.E., H. Gu & Y.M. Chook, (2008) Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals. *PLoS Biol.* **6**: e137.

Truant, R., Y. Kang & B.R. Cullen, (1999) The human tap nuclear RNA export factor contains a novel transportin-dependent nuclear localization signal that lacks nuclear export signal function. *J. Biol. Chem.* **274**: 32167-32171.

Tyler, B.M., (2007) *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* **8**: 1-8.

Underwood, M.R. & H.M. Fried, (1990) Characterization of nuclear localizing sequences derived from yeast ribosomal protein L29. *The EMBO journal* **9**: 91.

Xu, D., A. Farmer & Y.M. Chook, (2010) Recognition of nuclear targeting signals by Karyopherin-beta proteins. *Curr. Opin. Struct. Biol.* **20**: 782-790.

Zhang, M., J. Lu, K. Tao, W. Ye, A. Li, X. Liu, L. Kong, S. Dong, X. Zheng & Y. Wang, (2012) A Myb transcription factor of *Phytophthora sojae*, regulated by MAP kinase PsSAK1, is required for zoospore development. *PLoS One* **7**: e40246.

2.8 SUPPORTING INFORMATION

2.8.1 Fig. S2.1

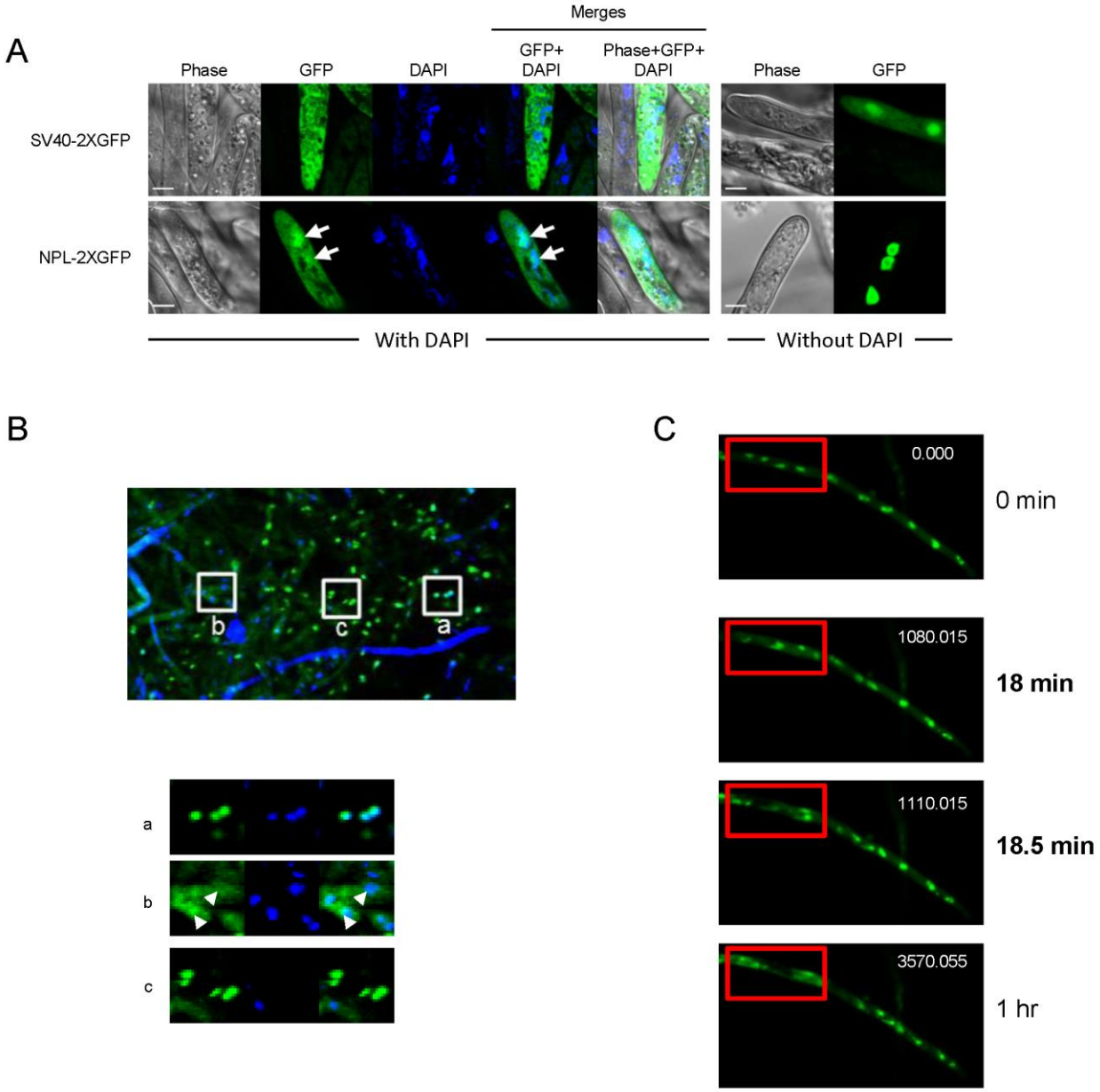


Fig. S2.1 Artifacts caused by DAPI (4', 6-diamidino-2-phenylindole) staining for live-cell imaging of P. sojae hyphae.

To label the *P. sojae* nuclei, *P. sojae* transformants expressing candidate NLSs fused to 2XGFP were stained with DAPI (final concentration, 0.2 μg ml^{-1}) for 25 min before imaging.

A. Different subcellular localizations of the NLS-tagged GFPs in *P. sojae* transformants with and without DAPI staining. Arrows indicate residual GFP staining of nuclei after DAPI treatment.

B. A representative image showing various localization patterns of 2XGFP fused to a synthetic NLS, PsNLS (Fang & Tyler, 2016). Region (a): GFP reporter showing nuclear localization in a small region of DAPI-stained hyphae. Region (b): Cytoplasmic localization in a larger region of DAPI-stained hyphae; arrowheads indicate artifactual cytoplasmic localization of NLS-GFP fusions. Region (c): Exclusive nuclear accumulation typically found in DAPI-free hyphae.

C. Time-lapse experiment tracking the process of mis-localization of PsNLS-2XGFP upon DAPI treatment. Red rectangles highlight changes occurring during DAPI incubation. In this example, the nuclear localized PsNLS-2XGFP was released into the cytoplasm following 18 to 18.5 minutes incubation with DAPI. Nuclear disintegration occurs at different rates in different regions of the hyphae possibly because newer hyphal regions absorb DAPI more slowly. Also see supplemental Movie 1.

2.8.2 Fig. S2.2

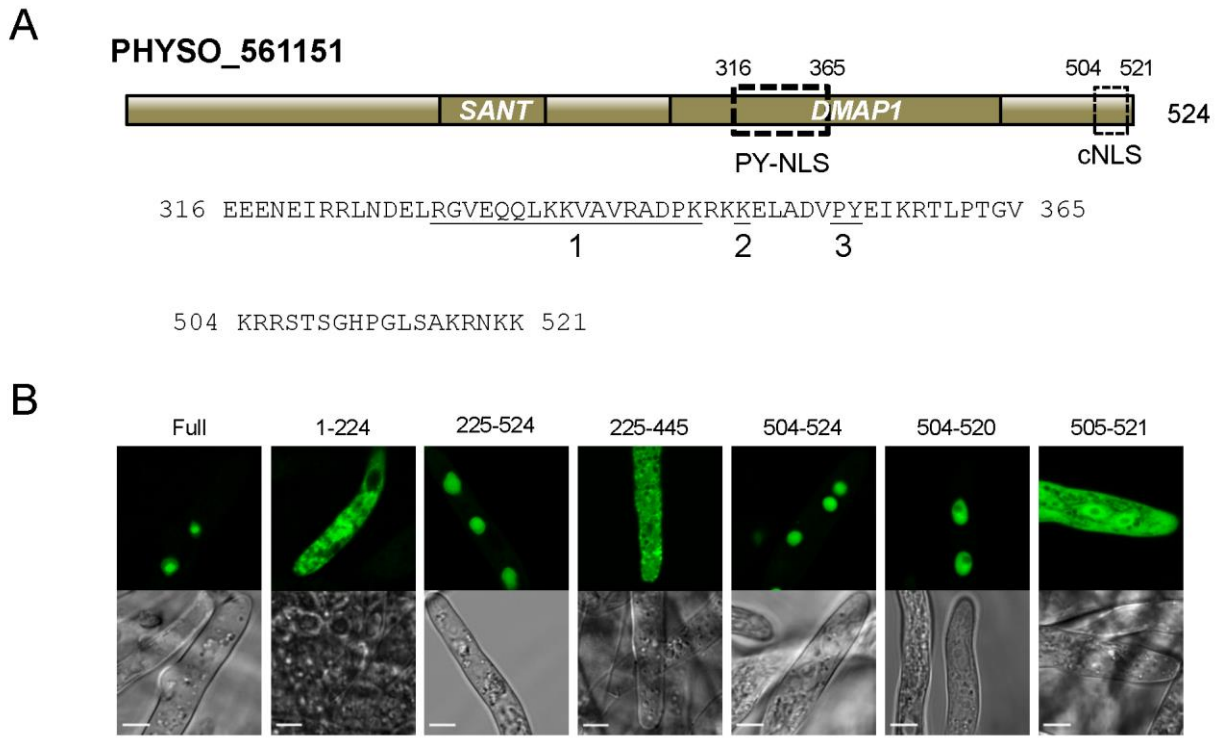


Fig. S2.2 Detailed mutational analysis of the PY-NLS candidate PHYSO_561151 reveals that an extended bipartite cNLS at the C-terminus is actually responsible for its nuclear localization.

A. Domain structure of PHYSO_561151. Positions of the candidate NLS sequences within PHYSO_561151 are indicated by rectangles. The corresponding amino acid sequences are listed below. Epitopes 1, 2, 3 of the putative PHYSO_561151 PY-NLS sequence are underlined.

B. Subcellular localization of PHYSO_561151 and mutants. Representative images are shown.

2.8.3 Fig. S2.3

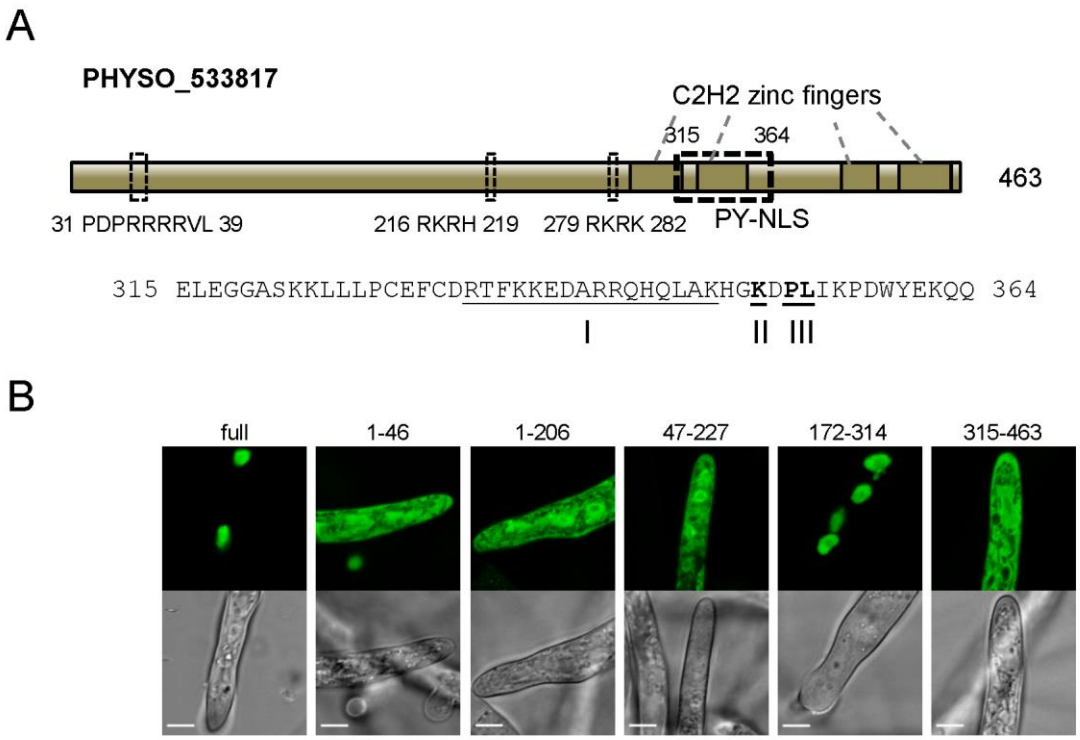


Fig. S2.3 Detailed mutational analysis of the PY-NLS candidate PHYSO_533817 reveals that residues 172-314 determine the nuclear accumulation.

A. Domain structure of PHYSO_533817. Positions of candidate NLS sequences within PHYSO_533817 are indicated by rectangles. The corresponding amino acid sequences are listed below. Epitopes 1, 2, 3 of the putative PY-NLS sequence are underlined.

B. Subcellular localization of PHYSO_533817 and mutants. Representative images are shown.

2.8.4 Fig. S2.4

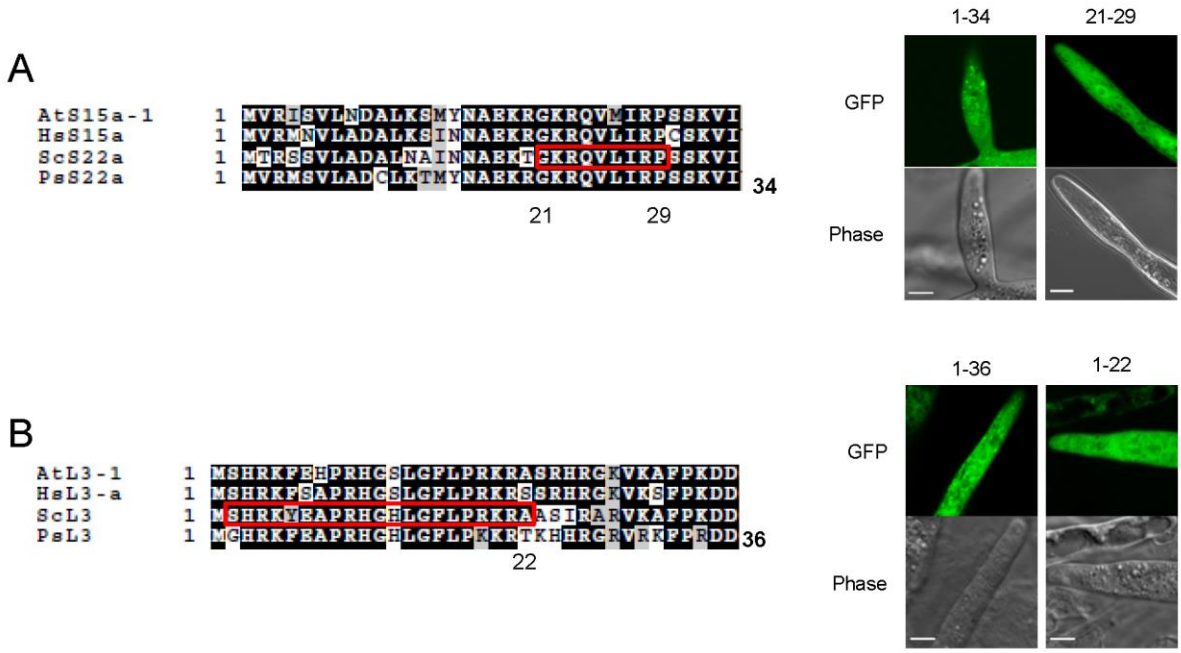


Fig. S2.4 Sequences used for nuclear import of ribosomal proteins S22a and L3 in yeast do not show the same activities in *P. sojae*.

A-B. Left panels, alignment of *P. sojae* ribosomal proteins S22a (PsS22a, PHYSO_287103) and L3 (PsL3, PHYSO_285779), with their orthologs in *Arabidopsis thaliana* (At), human (Hs) and *Saccharomyces cerevisiae* (Sc), respectively. Red rectangles, NLSs reported in yeast ribosomal proteins. Right panels, various fragments of ribosomal proteins, were expressed as fusions to 2XGFP in *P. sojae* transformants and visualized by confocal microscopy. Representative images are shown.

2.8.5 Supplemental sequences of the five putative PY-NLS-containing proteins

Notes:

Amino acids in lowercase and italics: incorrect intron annotation

Amino acids highlighted in yellow: sequences of candidate basic PY-NLS

Amino acids highlighted in green: sequences of candidate hydrophobic PY-NLS

Central basic or hydrophobic motifs (epitope I) are underlined and the R/K/H-PY motifs are in red bold

Amino acids in bold and italics: putative cNLS, or largest region covering multiple cNLSs.

The basic patch proximal to the core PY-NLS-like sequence in PHYSO_480605 is highlighted in dark.

> PHYSO_357835|U3 small nucleolar RNA-associated protein
MATSGEFKRLVLKQFPATTEVETAENTYWKKFHAPQELQQVGPVTHIDVSPVAPHQVAITSSTR
IHLYSTTTNEIVKTFSRFRDVVYSGTFRSDGKLLVAGGEAPYVQVLDINTRAILRSFKGHSAAI
RSTRFSADNVHVLSCSDDKTSRYWDLPTGKPLALLGEHSDYVRSSAANPSSHNVWATGSYDHTV
KLWDLRASDQTVSKSTMSLDHGAPVESCMIMPGGSLLLSAGGNSIKVWDILSGGRLLHSFSSHQ
KTITSLGLDGSGTRLMSGSLDGHLKIYDLKTYELAHGFKYKSGVLAFGMSPSNSHLFAGTVDGI
LAVRRRTVKRAEQTDAKNRQAIIRGGSYKYFLRGKNAQPAPADYTVATT***RHKR***IQ**PY**DRALRKF
DYKKALNEALDTRSPVVVASMLEELRLRVGLKRALGGRDEETLEPLLAFLIKYVTDPKYASLLI
HVCTIVCDLYAPKLSQSMLIDSLFVKLREKLNEELRVQKQVLGVVGMMDSVMAAQSNGTIGVDA
SS

>PHYSO_480605|mRNA cleavage and polyadenylation factor II
complex, BRR5 (CPSF subunit)
MSKRRLAEEAADER**H**IMRIM**PL**GAGNEVGRSCIVLKFKGKTIMLDCGVHPGYSGHGSLPFFDGV
EAEEIDLLLITHFHIDHVAALPHFTEKTNFKGRVFMTHPTKAVMQMMLRDFLRVSNISVDDQIY
DDKDLNNCVSKVEIIDFHQEIMHNGIKFTPYNAGHVLGACMYLIEIGGVKVLYTGDYSLENDRH
LMAAELPACSPDVLIVESTYGVQVHQSVVEREGRFTGQVEAVVRRGGRCLIPVFALGRTQELLL
ILDEHWRSHPDLQDIPIYFASKLAAKALRVYQTYINMMNDRIRKQIAISNPFQFEHISNLKSMD
DFDDSGPSVVMASPGMLQSGVSRQLFERWCSDKRNACLIPGYVVEGTLAKKILSEPTEIAALDG
RIIPMNCTVEYISFSAHADFVGTSGFVEKLTPPNIVLVHGEKNEMMRLKSALNKKFNDPKVYHP
SISTPANMQEIVLEFKGEKIAKAIGGLASDQPKNGKVISGLLVEVDSQTHLMDKEDLSTYTKLI
SGSITQKQHVPFEYNSFDVLITFIRQMYEDVVHLETENRVVVCKQVVVTRCPVAKGATEKLVVE
WTSAPTADMIADSVIALAMHAQASPASFKLSGQPTAACPHDHSKKEDESAHNEGAATKTEEEET
P*vmetediaenppsehdaekke*AESDLEKAARELGEADQDALNLLIVFRLLKDQYGDVDLDFET
NKIHVRTPSGIDAVVDHALQEIECKDAAFKLKLQTTVRRIEGALKPIATS

>PHYSO_251824|mRNA cleavage and polyadenylation factor I complex, subunit RNA15
MSNKSRASSAKERSVFVGNIPYDVTEDMLKEIFSEAGSVVNFRLVTDRETGKPKGYGFCEYADG
ATALSAMRNLNGYEINGRNLRVDFADGGDKSGGAD***RKRH***DNGSHARHGGTSGSNFRNGGSDGGP
PTMVTGEMAIHAIESAIARLGPVKLYDMLVQLKEHARQKPEVTKSILMANPALTHAIVQSFKTL
QIPIPSSTETQPVLLAPPPMMHQPPLMAGRMMPPPPRPLMGGHMGPGPGILGMAPPGVQPGMMN
PAPAPAPAKSGGTRWSARPGPLAAQATSSAMGGRPAVGIPASSPMHAQPQPGLMPTPTGPPQSS
GPPSSNLAHASRDPRRAGRDPRLA**K**R**PY**PGEQGLAPPTGDND***PSKRSKP***SGFGADAAPGGQFDA
IAELARDMTPQKLDMLPPNERQMLLAFMQQNNIPF

>PHYSO_561151|DNA methyltransferase 1-associated protein-1
MSDVAQILGLAGPKSGANGAAASELDQLKPTGASPAVRGKQSGGASKQKKLTGMQREVLELLES
NHRASHALYQGFGKTTLKQKWQERKKSPAVKWLRKSFRNPARAGLPGESGEEGLVLTHWGKAHV
EQPDYVFARFNVKCDTTSYTDEEYEAALANHLDPMMKWTKEETDLLLKLCQRFDLRWVVVTDKY
NSNPIAKSAPRSMEDIKYRYYEATRLLSEYRDKKTRGELEKKAA*vatpaatsagapategtpag*
*aataatpaaatve*TGGATSTPSTPVLDTPASSTSEHYRFNIAYEKQRKRQLDLTFSRTAEEENE
IRRLNDELRGVEQQLKKVAVRAD***PKRKKEL***ADV**PY**EIKRTLPTGVILRSSLLALPQQKHALSAK
LLKKLQLFLDEMGVPARPMPTKPVCETFDKLRQDAVGLLSLRKHLKSKQNEVQALRERYHALTG
KEYKPITTPVTLSERPGDAALPDGSSSAASAAGHAQSTISKGKTSKHSEKAIRAA**KRRSTSGHP**
GLSAKRNKKVPH

> PHYSO_533817|C2H2 zinc finger protein
MTQQEADTVPTPPSDAAPSEAKAATPSIPS***PDPRRRRVL***TVGDGNFSYSLALAKQHKSKGDDTR
PLQLTATSYDSYDELVAKYPESKRICAQLKELGASVLHRVDATNIRESLVAAGATSDKFHSVVF
NHPHCGEENVRRHQSLLSHFYASALEVLEKGEGDEVEQSEESGILLTLAEGQPERWQAVQRALS
AGLKLHRQVDNVDSDAKFGLEYE***RKRH***QNGKSFHQVTLHGERKKQASTLFIFRRQKAGEKVEAV
EAEVAPVvdaaavttgdgerkS**RKRK**AESELPLEFACTQCERSFKSAQGLRTHVHMVH**ELEGGA**
SKKLLLPCEFCDRTFKKEDARRQHQLAKHG**K**D**PL**IKPDWYEKQQAAAS*esaadagagdvakest*
*as*APSATTEAPATAEPQTCSICQLSFATAQEFDAHWQKLQPRSAAKRKCATCSREFDEERALRQ
HQNFCSQSKPAGSSS

2.8.6 Table S2.1

Table S2.1 Summary of cNLS-containing fragments that were tested throughout this study

Name	Fragment containing predicted NLS[1]	cNLS motif type[2]	Nuc[3]
SV40 cNLS	PKKKRKV	mono-pat7	I
c-myc	PAAKRVKLD	mono-pat4	I
NPL	KRPAATKKAGQAKKKK	bipartite	√
PsL28	1- MPSRFSKNRKKRGHVSAGHGRIGKHRKHPGGRG-33	Mix	√
	8-NRKKRG-13	mono-pat4	X
	24-KHRKHPG-30	mono-pat4	X
	11-KRGHVSAGHGRIGKHRK-27	bipartite	X
PsL3	1-MGHRKFEAPRHGHLGFLPKKRTKHHRGRVRKFPRDD-36	mono-pat4 and pat7	X
PsH3	76-...116KRVTIMPKDIQLARRIRGERS-136	bipartite	X
PsH4	1-MSGRGKGGKGLGKGGAKRHRKVLRD-25	mono-pat4	X
PHYSO_561151	225-...344PKRKKEL350...-455	mono-pat7	X
	504-KRRSTSGHPGLSAKRNK-520	bipartite	√
	505-RRSTSGHPGLSAKRNKK-521	bipartite	X
PHYSO_533817	1-...PDPRRRRVLTVGDGNF-46	mono-pat7	X
PHYSO_251824	1-...100RKRH103...-238	mono-pat4	X
	363-PSKRSKP-369	mono-pat7	X

[1.] Sequences predicted as NLSs in large fragments are underlined. Some sequences may have redundant cNLSs. Due to the limited space, long flanking sequences are abbreviated (indicated by dots).

[2.] cNLS patterns: mono-pat4, mono-pat7, monopartite cNLS having 4, or 7 consecutive amino acids. Prediction and classification are based on pSORT, 1997.

[3.] √ = the NLS-containing sequence showed clear nuclear targeting activity. I, incomplete nuclear localization (GFP signal is visible in the cytoplasm, approximately 1<LNC<3), X little nuclear localization.

2.8.7 Table S2.2

Table S2.2 *Phytophthora* proteins identifier in FungiDB and NCBI

Name or annotation*	FungiDB ID	NCBI accession #	Notes
Conserved WD40 repeat-containing protein	PHYSO_357835	XP_009538606.1	
mRNA cleavage and polyadenylation factor II complex, BRR5 (CPSF subunit)	PHYSO_480605	XP_009521134.1	Incorrect intron annotation
mRNA cleavage and polyadenylation factor I complex, subunit RNA15	PHYSO_251824	XP_009537811.1	
DNA methyltransferase 1-associated protein-1	PHYSO_561151	XP_009528012.1	Incorrect intron annotation
C2H2 zinc finger protein	PHYSO_533817	XP_009538325.1	Incorrect intron annotation
PsL28	PHYSO_355737	XP_009520762	
PsL3	PHYSO_285779	XP_009514874	
PsS15a	PHYSO_287103	XP_009523228	
PsH2B	PHYSO_474922	XP_009515672.1	
PsH3	PHYSO_286415	XP_009518876	
PsH4	PHYSO_285922	XP_009515675.1	
P. capsici fibrillarin (FIB)	PHYCA_506928	XP_009521478.1	

* The naming system for *P. sojae* ribosomal proteins follows the nomenclature of *S. cerevisiae*. All other protein annotations were adapted from FungiDB, except PHYSO_533817 which was annotated as a C2H2 transcription factor by fungal transcription factors database (FTFD, Park *et al*, 2008).

2.8.8 Table S2.3

Table S2. 3 Primers used in this study

Primer name	Sequence	Usage
pYF2_GW_F	5' GGGGACAAGTTTGTACAAAAAAGCAGGCTGCAGGAATTAA TTCGATATCAAGCTTATC 3'	PCR amplify 2XGFP and its flanking restriction sites from pYF2 plasmid and fuse to pcDNA3.2 using Gateway® technology
GFP_stop_GW_R	5' GGGGACCACTTTGTACAAGAAAGCTGGGTTCTACTTGTAGA GTTCATCCATGCCATG 3'	
GFP_AflII_F	5' CGG*CTTAAG*ATGGGCAAGGGCGAGGAAC 3'	Generate plasmid pYF3-2XGFP
GFP_CMCS_ApaI_R	5' TAGGGCCCTCAACGCGTTCCGGAGTTAACGGATTCTGTACA CTTGTAGAGTTCATCCATGCCATG 3'	
SV40_NLS_SacII_F	5' GGATGCCAAAGAAAAAGAGAAAGGTTA 3'	Oligo annealing, clone SV40 NLS to the N-terminus of 2XGFP
SV40_NLS_SpeI_R	5' CTAGTAACCTTTCTCTTTTTCTTTGGCATCCGC 3'	
SV40_NLS_BsrG1_F	5' GTACACCAAAGAAAAAGAGAAAGGTTT 3'	Oligo annealing, clone SV40 NLS to the C-terminus of 2XGFP
SV40_NLS_BspEI_R	5' CCGGAAACCTTTCTCTTTTTCTTTGGT 3'	
3XSV40_NLS_SacII_F	5' GGATGCCAAAGAAAAAGAGAAAGGTTCCAAAGAAAAAGAG AAAGGTTCCAAAGAAAAAGAGAAAGGTTA 3'	Oligo annealing, clone three copies of SV40 NLS to the N-terminus of 2XGFP
3XSV40_NLS_SpeI_R	5' CTAGTAACCTTTCTCTTTTTCTTTGGAACCTTTCTCTTTTTCT TTGGAACCTTTCTCTTTTTCTTTGGCATCCGC 3'	
c-Myc_NLS_SacII_F	5' GGATGCCGGCCGCCAAGCGCGTGAAGCTGGACA 3'	Oligo annealing, clone c-Myc NLS to the N-terminus of 2XGFP
c-Myc_NLS_SpeI_R	5' CTAGTGTCCAGCTTCACGCGCTTGGCGGCCGGCATCCGC 3'	
NLP_NLS_SacII_F	5' GGATGAAGCGCCCGGCCGCCACGAAGAAGGCCGGCCAGGC CAAGAAGAAGAAGA 3'	Oligo annealing, clone nucleoplasmin NLS to the N-terminus of 2XGFP
NLP_NLS_SpeI_RIV	5' CTAGTCTTCTTCTTCTTGGCCTGGCCGGCCTTCTTCGTGGCG GCCGGGCGCTTCATCCGC 3'	
PsH2B_PacI_F	5' CC*TTAATTAA*ATGGCGAAGACTCCCTCG 3'	Clone *P. sojae* H2B to pGPFN or pMCherryN
PsH2B_NheI_R	5' TT*GCTAGC*AGCGGACGTGAACTTGGT 3'	
PsFIB_AgeI_F	5' GGGACCGGTATGGCCGGTGGCGCCAAG 3'	Clone *P. capsici* fibrillarin to pMCherryN
PsFIB_PacI_R	5' CGCC*TTAATTAA*CTTCTTCTCCTTCTTGGGCACACGG 3'	
Ps561151_504_SacII_F	5' GGATGAAGCGACGCAGCACCAGTGGCCACCCAGGTCTCTCT GCCAAGCGCAACAAGA 3'	Oligo annealing, clone residues 504-520 of PHYSO_561151 to the N-

Ps561151_504_SpeI_R	5' CTAGTCTTGTTGCGCTTGGCAGAGAGACCTGGGTGGCCACTGGTGCTGCGTCGCTTCATCCGC 3'	terminus of 2XGFP
Ps561151_505_SacII_F	5' GGATGCGACGCAGCACCAGTGGCCACCCAGGTCTCTCTGCCAAGCGCAACAAGAAAA 3'	Oligo annealing, clone residues 505-521 of PHYSO_561151 to the N-terminus of 2XGFP
Ps561151_505_SpeI_R	5' CTAGTTTTCTTGTTGCGCTTGGCAGAGAGACCTGGGTGGCCACTGGTGCTGCGTCGCATCCGC 3'	
M9NLS_SacII_F	5' GGATGTTCGGCAACTACAACAACCAGTCGTCGAACTTCGGCCCGATGAAGGGCGGCAACTTCGGCGGCCGCTCGTCGGGCCCGTACA 3'	Oligo annealing, clone M9 NLS to the N-terminus of 2XGFP
M9NLS_SpeI_R	5' CTAGTGTACGGGCCCGACGAGCGGCCGCCGAAGTTGCCGCCCTTCATCGGGCCGAAGTTCGACGACTGGTTGTTGTAGTTGCCGAACATCCGC 3'	
hnRNPD_SacII_F	5' GGATGTACGGCGACTACTCGAACCAGCAGTCGGGCTACGGCAAGGTGTCGCGCCGCGGCGGCCACCAGAACTCGTACAAGCCGTACA 3'	Oligo annealing, clone hnRNPD NLS to the N-terminus of 2XGFP
hnRNPD_SpeI_R	5' CTAGTGTACGGCTTGTACGAGTTCTGGTGGCCGCCGCGGCGCGACACCTTGCCGTAGCCCGACTGCTGGTTCGAGTAGTCGCCGTACATCCGC 3'	
hnRNP-M_SacII_F	5' GGATGGAGCGCCCGGCCCAGAACGAGAAGCGCAAGGAGAAGAACATCAAGCGCGGCGGCAACCGCTTCGAGCCGTACGCCAACCCGACGAAGCGCA 3'	Oligo annealing, clone hnRNPM NLS to the N-terminus of 2XGFP
hnRNP-M_SpeI_R	5' CTAGTGCGCTTCGTCGGGTTGGCGTACGGCTCGAAGCGGTTGCCGCCGCGCTTGATGTTCTTCTCCTTGCGCTTCTCGTTCTGGGGCCGGGCGCTCCATCCGC 3'	
M9M_NheI_F	5' CGC*GCTAGC*ATGTTCGGCAACTACAACAACCAGTCG 3'	PCR amplify M9M using M9NLS as a template and clone it to the N-terminus of 2XGFP using *Stu I* of pYF2-2XGFP
M9M_AgeI_R	5' GCC*ACCGGT*GCGCTTCGTCGGGTTGGCGTACGGCTCGAAGCGGCCGCCGAAGTTGCC 3'	
Nab2p1_SacII_F	5' GGATGGACAACTCGCAGCGCTTCACGCAGCGCGGCGGCGGCGCCGTGGGCAAGAACCGCCGCGGCGGCCGCGGCGGCAAC 3'	Oligo annealing, clone Nab2p PY-NLS to the N-terminus of 2XGFP
Nab2p1_R	5' GCCGCGGTTGCCGCCGCGGCCGCCGCGGCGGTTCTTGCCCACGGCGCCGCCGCCGCGCTGCGTGAAGCGCTGCGAGTTGTCCATCCGC 3'	
Nab2p2_F	5' CGCGGCGGCCGCAACAACAACTCGACGCGCTTCAACCCGCTGGCCAAGGCCCTGGGCATGGCCGGCGAGTCGAACATGA 3'	
Nab2p2_SpeI_R	5' CTAGTCATGTTCGACTCGCCGGCCATGCCCAGGGCCTTGGCCAGCGGGTTGAAGCGCGTCGAGTTGTTGTTGCGGCC 3'	
Ps357835_F	5' ACCATGGCGACCAGCGGCGAG 3'	PCR amplification of full length PHYSO_357835
Ps357835_R	5' ACTGCTGGCATCCACACCGATGGTG 3'	
Ps357835_338_F	5' ATGAACCGTCAGGCGATCATCCG 3'	PCR amplification of residues

Ps357835_387_R	5' CTTGTAATCAAACTTCCGCAGCG 3'	338-387 in PHYSO_357835
Ps357835_H371A/K372A	5' ACGGCTGAATACGCGCGGCACGTGTGGTGGCCACGGTGT 3'	Mutate H371A/K372A in PHYSO_357835
Ps357835_P376A/Y377A	5' CGTCACAAGCGTATTCAGGCGGCCGACCGCGCGCTG 3'	Mutate P376A/Y377A in PHYSO_357835
Ps357835_H371A/K372A/P376A/Y377A	5' CAGCGCGCGGTCGGCCGCCTGAATACGCGCGGCACGTGTGGTGGCCAC 3'	Mutate H371A/K372A/P376A/Y377A in PHYSO_357835
Ps357835(H371A/K372A)_R370A/R373A	5' GTGGCCACCACAGCTGCCGCGGCTATTCAGGCGGC 3'	Further mutate R370A/R373A in PHYSO_357835(H371A/K372A)
Ps480605_F	5' ATGTCGAAGCGCAGGCTAGCCG 3'	PCR amplify full length PHYSO_480605 and its truncations
Ps480605_R	5' CGACGTAGCAATGGGCTTCAGTGC 3'	
Ps480605_32_R	5' CGCC*TTAATTAA*GCACGAGCGCCCCACCTCG 3'	
Ps480605_60_R	5' GAACGGCAGGCTGCCGTG 3'	
Ps480605_33_F	5' ATCGTGCTCAAGTTCAAGGGCAAGAC 3'	
Ps480605_810_F	5' CATCCCAATCTACTTTGCATCC 3'	Sequencing primer for PHYSO_480606
Ps480605_1539_F	5' GGGTAGTATCACCCAGAAGCA 3'	Sequencing primer for PHYSO_480607
Ps251824_F	5' ATGAGCAATAAGTCGCGCGCG 3'	PCR amplify full length PHYSO_251824 and its truncations
Ps251824_R	5' AAACGGAATGTTGTTCTGCTGCATG 3'	
Ps251824_238_R	5' GCC*CCGCGG*TTAAAACGGAATGTTGTTCTGCTGCATG 3'	
Ps251824_239_F	5' ATGGGGCCAGGCATCCTCGGC 3'	
Ps251824_309_F	5' ATGTTGATGCCAACTCCGACTGGTC 3'	
Ps251824_308_R	5' GCCCGGTTGAGGCTGAGCG 3'	
Ps251824_362_R	5' GTCATTGTCTCCTGTCGGCGGTG 3'	
Ps251824_cNLS2_F	5' GGATGCCGTCGAAGCGCTCGAAGCCGA 3'	Oligo annealing, clone residues 363-369 (the cNLS2 in PHYSO_251824) to the N-terminus of 2XGFP
Ps251824_cNLS2_R	5' CTAGTCGGCTTCGAGCGCTTCGACGGCATCCGC 3'	
Ps251824_cNLS_mut_R	5' GGAGACAATGACCCGTCCGCGGCGTCCGCGCCGAGCGGTTTTGGTGC 3'	Mutate cNLS2 in residues 239-419 of PHYSO_251824
Ps251824_P347A_Y348A	5' CCCTCGTTTGGCAAAGAGAGCGGCTCCCGGCGAG 3'	Mutate P347A/348A in context of PHYSO_251824(P347A/348A)
Ps251824_R332A/R336A/R339A	5' CACGCCAGCGCTGACCCTCGAGCTGCAGGCGCAGACCCTCG-3'	Mutate R332A/R336A/R339A in context of PHYSO_251824(P347A/348A)
Ps561151_F	5' GCCTTAATTAAATGAGCGACGTGGCGCAGATCTTGG 3'	PCR amplify full length PHYSO_561151 and its truncations
Ps561151 _R	5' ATGCGGGACTTTCTTGTTGC 3'	
Ps561151 _225_F	5' ATGAAGAAGACGAGAGGTGAACTTGAAA 3'	
Ps561151 _445_R	5' CGCGTGGTATCGTTCACGA 3'	
Ps561151_cNLS504_SacII_F	5' GGATGAAGCGACGCAGCACCAGTGGCCACCCAGGTCTCTCTGCCAAGCGCAACAAGA 3'	Oligo annealing, clone residues 504-520 of PHYSO_561151 to the N-terminus of 2XGFP
Ps561151_cNLS504_SpeI_R	5' CTAGTCTTGTTGCGCTTGGCAGAGAGACCTGGGTGGCCACTGGTGCTGCGTCGCTTCATCCGC 3'	

Ps561151_cNLS 505_SacII_F	5' GGATGCGACGCAGCACCAGTGGCCACCCAGGTCTCTCTGCC AAGCGCAACAAGAAAA 3'	Oligo annealing, clone residues 505-521 of PHYSO_561151 to the N-terminus of 2XGFP
Ps561151_cNLS 505_SpeI_R	5' CTAGTTTTCTTGTTGCGCTTGGCAGAGAGACCTGGGTGGCC ACTGGTGCTGCGTCGCATCCGC 3'	
Ps561151_del 225-403_R	5' TGGTGCTGCGTCGCTTCATCCTATCGATAAGC 3'	Delete residues 225-403 in constructs pYF2-2XGFP-PHYSO_561151$_{225\text{-}524}$ to generate residues 504-524
Ps561151_del 225-403_F	5' GCTTATCGATAGGATGAAGCGACGCAGCACCA 3'	
Ps533817_F	5' ATGACGCAGCAAGAAGCGGACAC 3'	PCR amplify full length PHYSO_ Ps533817 and its truncations
Ps533817_R	5' ACTACTGCTACCGGCAGGCTTACTCTG 3'	
Ps533817_46_R	5' GAAATTCCCGTCGCCCACAGTG 3'	
Ps533817_206_R	5' GCTGTCCACGTTGTCAACCTGTCG 3'	
Ps533817_47_F	5' ATGTCGTACTCATTGGCGTTGGCCAAG 3'	
Ps533817_227_R	5' CTGGTGGAACGACTTGCCGTTCTG 3'	
Ps533817_315_F	5' ATGCACGAGTTGGAGGGGGGCG 3'	
Ps533817_172_F	5' ATGATCCTGTTGACGCTGGCCG 3'	
Ps533817_314_R	5' CACCATGTGCACGTGCGTCCTC 3'	
Ps533817_315_F	5' ATGCACGAGTTGGAGGGGGGCG 3'	
PsL28_F	5' TGCCGTCCCGTTTTTCCAAGAAC 3'	PCR amplify residues 1-33 of PsL28
PsL28_R	5' ACCGCGACCACCCGGGTG 3'	
PsL28_NLS2-mus_R	5' GTGCTAGCACCGGCACCACCCGGGTGCTT 3'	Mutate R7A/K9A/R10A in PsL28$_{1\text{-}33}$
PsL28_NLS1-mus_R	5' CGCTGACGTGTCCGCGTTTCGCGGCGTTCGCGGAAAAACGG GACGGCATTTA-3'	
PsL28-NLS1_SacII_F	5' GGATGAACCGCAAGAAACGCGGAA 3'	Oligo annealing, clone NLS1 (residues 7-13) in PsL28 to N-terminus of 2XGFP
PsL28-NLS1_SpeI_R	5' CTAGTTCCGCGTTTCTTGCGGTTCATCCGC 3'	
PsL28-NLS2_SacII_F	5' GGATGAAGCACCGCAAGCACCCGGGTA 3'	Oligo annealing, clone NLS2 (residues 24-30) in PsL28 to the N-terminus of 2XGFP
PsL28-NLS2_SpeI_R	5' CTAGTACCCGGGTGCTTGCGGTGCTTCATCCGC 3'	
PsS22A_F	5' ATGGTTCGTATGAGCGTGCTGG 3'	PCR amplify residues 1-34 of PsS22A
PsS22A_34_R	5' GATCACCTTGGAGCTCGGGC 3'	
PsS22A_SacII_F	5' GGATGGGAAAGCGCCAGGTGCTCATCCGCCCGA	Oligo annealing, clone residues 21-29 of PsS22A to the N-terminus of 2XGFP
PsS22A_SpeI _R	CTAGTCGGGCGGATGAGCACCTGGCGCTTTCCCATCCGC	
PsL3_F	5' ATGGGTCACCGTAAGTTCGAGGC 3'	PCR amplify residues 1-36 of PsL3
PsL3_36_R	5' GTCATCGCGCGGGAACTTGC 3'	
PsL3_SacII_F	5' GGATGGGTCACCGTAAGTTCGAGGCCCCGCGCCACGGCCAC CTGGGTTTCCTGCCGAAGAAGCGCACCA 3'	Oligo annealing, clone residues 1-22 of PsL3 to the N-terminus of 2XGFP
PsL3_SpeI_R	5' CTAGTGGTGCGCTTCTTCGGCAGGAAACCCAGGTGGCCGTG GCGCGGGGCCTCGAACTTACGGTGACCCATCCGC 3'	
PsH3_F	5' ATGGCCCGTACCAAGCAGACCG 3'	PCR amplify full length PsH3 and its truncations
PsH3_R	5' CGAGCGCTCGCCACGGATG 3'	
PsH3_32_R	5' GGCGGGGGCCGACTTGC 3'	

PsH3_33_F	5' ATGACGGGCGGCGTCAAGAAAC 3'	
PsH3_75_R	5' GATCTCGCGCACCAGGCGC 3'	
PsH3_76_F	5' GCCCAGGACTTCAAGACGGAC 3'	
PsH4_F	5' ATGTCTGGACGCGGCAAAG 3'	PCR amplify full length PsH4 and its truncations
PsH4_R	5' TCCGCCGAAGCCGTAGAG 3'	
PsH4_25_R	5' GTCGCGAAGAACCTTGC 3'	
PsH4_42_R	5' ACCACGGCGAGCGAGAC 3'	
PsH4_80_R	5' CTTGCGGCGCGCGTG 3'	
PsH4_26_F	5' ATGAACATCCAGGGCATCACCAAG 3'	
PsH4_43_F	5' ATGGGAGTGAAGCGCATCTCCG 3'	
PsH4_81_F	5' ATGACCGTGACGGCCATGGAC 3'	

2.8.9 Additional references in the supporting information

Park, J., J. Park, S. Jang, S. Kim, S. Kong, J. Choi, K. Ahn, J. Kim, S. Lee & S. Kim, (2008) FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics* **24**: 1024-1025.

Chapter 3

Nuclear localization of a putative *Phytophthora sojae* bZIP1 transcription factor is mediated by multiple targeting motifs

Yufeng Fang[1, 2] and Brett M. Tyler[1, 2*]

[1] Interdisciplinary Ph.D. program in Genetics, Bioinformatics & Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA

[2] Center for Genome Research and Biocomputing and Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

* Corresponding author: Brett.Tyler@oregonstate.edu

This Chapter includes a research article submitted to Molecular Microbiology as "Nuclear localization of a putative Phytophthora sojae bZIP1 transcription factor is mediated by multiple targeting motifs". Brett M. Tyler helped to analyze the data and edited the manuscript.

3.1 ABSTRACT

Phytophthora sojae is an oomycete pathogen of soybean. In Chapter 2, we found that *P. sojae* uses nuclear localization signals (NLSs) for translocation of proteins into the nucleus that differ from conventional well-characterized NLSs. Most tested classical NLSs (cNLSs) and proline-tyrosine NLSs (PY-NLSs) from model organisms showed poor nuclear import activities in *P. sojae.* In comparison, functional nuclear localization sequences from *P. sojae* nuclear proteins resembled conventional NLSs but required additional basic residues, either within the NLS motif or nearby. To learn more about NLS-mediated transport of nuclear proteins in *P. sojae*, here we have characterized in depth the nuclear import mechanism of a *P. sojae* basic leucine zipper transcription factor, PsbZIP1. We found that the nuclear translocation of PsbZIP1 was determined by a central conserved region. Mutational analysis of this region indicated its nuclear translocation was mediated by four distinct motifs but was independent of conserved DNA binding residues. Three motifs showed autonomous NLS activity and the fourth motif served as a nuclear localization enhancer. Sequence comparison and mutational analysis of these nuclear localization motifs revealed a new form of bipartite NLS consisting of a triplet of basic residues followed by a tail of scattered basic amino acids.

3.2 INTRODUCTION

Phytophthora sojae is a fungus-like pathogen, but belongs to the oomycetes in the kingdom Stramenopila (Tyler, 2007). *Phytophthora* species are notorious as plant destroyers, causing multibillion-dollar damage to agriculture and natural ecosystems worldwide annually (Erwin & Ribeiro, 1996). For example, the potato late blight pathogen *P. infestans* caused the Irish potato famine in the 1840s. *P. sojae* mainly infects soybeans, causing ‘damping off’ of seedlings as well

as root and stem rot of mature plants (Tyler, 2007). Due to its economic importance, *P. sojae*, together with *P. infestans*, have become model species for the study of oomycete biology and pathology (Tyler, 2007).

Nuclear localization signals (NLSs) are specific sequences rich in positively charged amino acids that direct proteins from the cytoplasm to the nucleus in eukaryotes. Currently, two classes of NLSs have been defined: the classical NLS (cNLS) and the proline-tyrosine NLS (PY-NLS) (Lange *et al.*, 2007; Chook & Suel, 2011). Nuclear import of cNLSs is associated with specific karyopherins, namely a receptor, Kapβ1, and an adaptor protein, Kapα (Lange *et al.*, 2007). There are two cNLS subtypes, both of which interact with Kapα/Kapβ1 to mediate nuclear entry. Monopartite cNLSs consist of a single cluster of basic amino acids, the canonical example being the SV40 large T-antigen NLS (126-PKKKRKV-132). Bipartite cNLSs consist of two clusters of basic amino acids, the canonical example being the nucleoplasmin NLS (155-KRPAATKKAGQAKKKK-170) (Lange *et al.*, 2007). In contrast to cNLSs, PY-NLSs interact with Kapβ2 to mediate nuclear import. PY-NLSs are defined based on three rules: they are structurally disordered in the unbound state; they have overall basic character; and they possess a central hydrophobic or basic motif followed by the motif [R/H/K]-X_{2-5}-P[Y/L] (Lee *et al.*, 2006; Suel *et al.*, 2008). In Chapter 2, we found that many cNLSs and PY-NLSs that have been well-characterized in mammalian cells or *Saccharomyces cerevisiae* worked poorly in *P. sojae*. To accomplish nuclear import of proteins efficiently, fully functional *P. sojae* cNLSs and PY-NLSs required additional basic amino acids or collaboration with other weak NLSs.

Transcription factors (TFs) play crucial roles in almost all biological processes (Jakoby *et al.*, 2002). One large TF family features a basic leucine zipper domain called bZIP. This family is present in all eukaryotes and regulates many critical physiological activities, such as

development, metabolism, and responses to environmental changes (Hurst, 1993). Structurally, the bZIP domain comprises a basic region with a conserved 'N-X7-R/K' motif for DNA binding (also called the DNA binding domain, DBD), followed by C-terminal heptad repeats of leucine residues (called leucine zipper domain LZD) responsible for dimerization (Jakoby *et al.*, 2002; Marco Llorca *et al.*, 2014). A number of bZIPs have been identified in *P. infestans* and *P. sojae*, but only a few of them have been characterized (Blanco & Judelson, 2005; Gamboa-Melendez *et al.*, 2013; Ye *et al.*, 2013). Two closely similar *P. infestans* bZIP transcription factors, Pibzp1 and PITG_11668, were reported to constitutively localize in the nucleus, regardless of different development stages (Blanco & Judelson, 2005; Gamboa-Melendez *et al.*, 2013). However, the NLSs responsible for translocation of these bZIPs were not determined. In this study, we have used confocal microscopy to characterize the nuclear localization signals of a *P. sojae* bZIP transcription factor, PsbZIP1, which is a close homolog of Pibzp1 and PITG_11668. We found that the nuclear import of PsbZIP1 was mediated by four distinct motifs distributed within the conserved central region, which acted together to multiply the extent of nuclear localization. Two of these motifs defined a new form of bipartite NLS.

3.3 RESULTS

3.3.1 The nuclear accumulation of PsbZIP1 is determined by a central 134 amino acid domain

To identify homologs of Pibzp1 and PITG_11668, Pibzp1 was used as a query for a BLASTP search against the *P. sojae* P6497 proteome. Two predicted proteins, PHYSO_481435 and PHYSO_256931, were identified as top hits. However, when cloning the *P. sojae* bZIP cDNAs by RT-PCR, we found a predicted intron was retained in the cDNA of PHYSO_256931, causing

a frame shift in the predicted coding region (data not shown). On the other hand, we found that the gene model PHYSO_481435 was missing 333 bp at the 5´ end, based on analysis of RNAseq data and comparisons to homologs in other *Phytophthora* species (Fig. S3.1A and S3.1B). We defined the corrected PHYSO_481435 gene model as *PsbZIP1*. PsbZIP1 shares 47% amino acid identity with Pibzp1 and 59% identity with PITG_11668 (Fig. S3.1C).

To validate the subcellular localization of PsbZIP1, the full-length predicted protein was expressed as a fusion with GFP in *P. sojae* transformants. PsbZIP1-GFP showed strong nuclear accumulation in *P. sojae* hyphae (small unstained regions in the centers of nuclei were validated as nucleoli), whereas a fused dimeric GFP protein (2XGFP) showed diffuse staining in both the cytoplasm and nucleus (Fig. 3.1A). As expected, the two heterologous bZIPs, Pibzp1 and PITG_11668 also were exclusively localized in the *P. sojae* nuclei when fused to GFP and expressed in *P. sojae* transformants. However, both *P. infestans* TFs exhibited a heterogeneous distribution within the nuclei (Fig. 3.1B). Confirmation of nuclear localization by these two homologs suggested that the sequences determining the nuclear localization of PsbZIP1 were conserved, possibly within the highly conserved region from 113 to 246 (Fig. 3.1C). Indeed, residues 113 to 246, when fused to 2XGFP, produced strong nuclear localization, while the other two portions, residues 1-112 and residues 247-369 remained primarily in the cytoplasm (Fig. 3.1D).

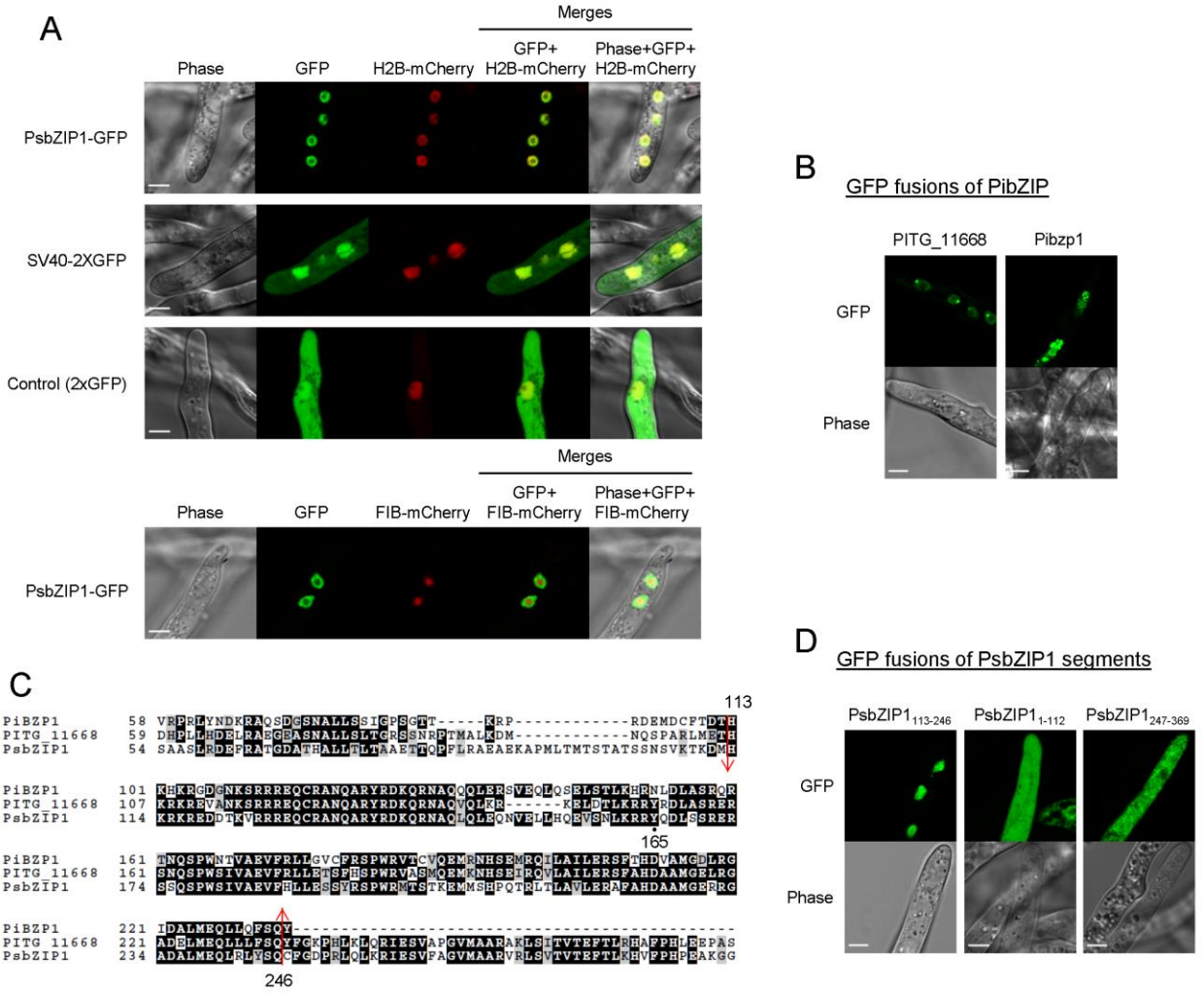


Fig. 3.1 Nuclear localization of PsbZIP1 is determined by a central 134 amino acid domain.

A. Subcellular localization of full length PsbZIP1 in *P. sojae* transformants. Top, PsbZIP1 was fused to a single copy of GFP. Positive control, SV40 (SV40 cNLS) fused to 2XGFP; negative control, 2XGFP alone. Histone H2B fused to mCherry was used as a nuclear marker (See Chapter 2). Bottom, PsbZIP1 was excluded from the nucleolus, as confirmed by the nucleolar marker FIB-mCherry (See Chapter 2).

B. Subcellular localization of *P. infestans* bZIP PITG_11668 and PibZP1 fused to GFP in *P. sojae* transformants.

C. Alignment of three *Phytophthora* bZIP sequences. Arrows indicate the left and right boundaries of the minimal conserved region.

D. Subcellular localization of N-terminal, central, and C-terminal segments of PsbZIP1.

3.3.2 The nuclear localization of PsbZIP1 is not dependent on DNA binding or protein dimerization motifs

Because PsbZIP1 is a putative transcription factor that contains a predicted DNA binding domain (DBD, residues 120-142) and a leucine zipper domain (LZD, 143-179) (Fig. 3.2A), it was possible that the nuclear accumulation produced by residues 113-246 was caused by DNA binding or protein interactions. Alternatively, residues 113-165 that are rich in highly conserved, basic amino acids, potentially constituted some kind of NLS (Fig. 3.1C). To test the role of this region, the first 56 amino acids were deleted from PsbZIP1$_{113-246}$-2XGFP and its subcellular distribution was analyzed. The distribution of 2XGFP between the nucleus and cytoplasm was measured in order to quantitate the degree of localization produced by each of the constructs. PsbZIP1$_{168-246}$-2XGFP appeared to be exclusively in the cytoplasm, with a $\log_2$-nuclear-to-cytoplasm ratio (LNC) of -0.77 (Fig. 3.2B and C), indicating that the residues 113-167 may be required for nuclear localization. To dissect this region further, we successively deleted smaller segments containing essential residues involved in DNA binding or/and protein dimerization. Partial deletion of the DBD, 113-246-Δ(127-141) or the LZD, 113-246-Δ(143-167) alone or together showed no visible effect on nuclear accumulation (Fig. 3.2B). Deletion of an even larger segment, 113-246-Δ(127-186), still displayed predominant nuclear staining (LNC=4.77) (Fig. 3.2B and C), suggesting that the nuclear localization of PsbZIP1 did not require the DNA binding or protein interaction motifs.

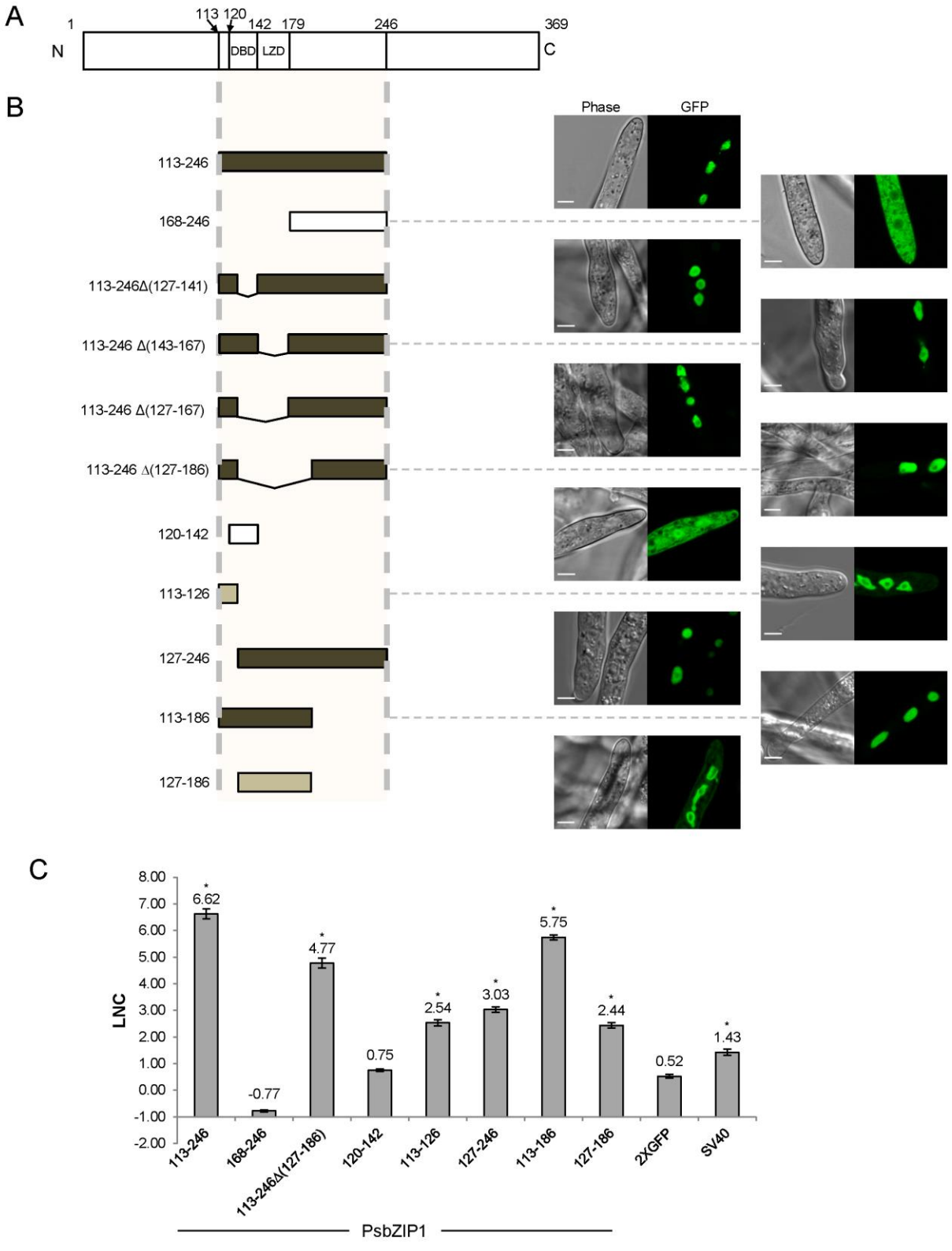


Fig. 3.2 Three distinct regions required for nuclear localization of PsbZIP1$_{113\text{-}246}$.

A. Schematic of PsbZIP1 domain structure. The boundaries of the DBD and LZD were as annotated by NCBI.

B. Mutational analysis of $bZIP1_{113-246}$. Left, boxes indicate truncated bZIP1 segments relative to the full length protein. Darkness of the shading of the boxes reflects the intensity of nuclear localization. Right, representative images of the subcellular localization of the segments fused to 2XGFP.

C. Quantification of nuclear to cytoplasmic ratios of key constructs displayed in B. 2XGFP, negative control; SV40, positive control; LNC, $\log_2$ transformed Nuc:Cyt ratio ($\log_2$[Nuc:Cyt]). In this and subsequent figures, the LNC values for 2XGFP and SV40-2xGFP are the same as in Chapter 2. *, LNC values significantly greater (P value < 0.01) than the 2XGFP negative control.

It was reported that the nuclear import of several mammalian and plant bZIPs was mediated by bipartite cNLSs within the conserved basic region (Miller, 2009). Examination of the PsbZIP1 protein sequence revealed a predicted bipartite cNLS (125-RREQCRANQARYRDKQR-141) within the DBD. However, when the entire DBD including this predicted cNLS (120-142) was fused to 2XGFP, the 2XGFP fusion was distributed throughout the cytoplasm as well as the nuclei (LNC= 0.75; not significantly greater than 2XGFP, LNC=0.52) (Fig. 3.2B and C), suggesting that the DBD and the predicted bipartite cNLS within it were not sufficient for efficient PsbZIP1 import into the *P. sojae* nucleus.

3.3.3 Three distinct regions required for nuclear import of $PsbZIP1_{113-246}$

Since $PsbZIP1_{113-246}\Delta$(127-186) retained strong nuclear localization activity, we tested whether the highly basic region 113-126 may be responsible. $PsbZIP1_{113-126}$-2XGFP mostly localized in the nucleus (LNC= 2.54) (Fig. 3.2B and C) indicating that it could contribute to nuclear localization, but localization was significantly weaker than $PsbZIP1_{113-246}\Delta$(127-186) (LNC=4.77). To test whether residues 113-126 were necessary for nuclear import of $PsbZIP1_{113-}$

$_{246}$, the segment was deleted, resulting in PsbZIP1$_{127\text{-}246}$-2XGFP. PsbZIP1$_{127\text{-}246}$-2XGFP produced clear nuclear localization (LNC= 3.03) although less than the full length PsbZIP1$_{113\text{-}246}$ (LNC= 6.62) (Fig. 3.2B and C). Together, these observations indicated that residues 113-126 contributed to the NLS activity of residues 113-246, but that additional NLS activity may exist within residues 127-246.

Since PsbZIP1$_{113\text{-}246}$Δ(127-186) (LNC= 4.77) produced more nuclear localization than PsbZIP1$_{113\text{-}126}$ (LNC=2.54) (Fig. 3.2C), it appeared that additional sequences between residue 187 and 246 could contribute to localization. However, as already shown above, residues 168-246 were not sufficient to import 2XGFP into the nucleus (LNC= -0.77), thus residues 187-246 should not constitute an autonomous NLS. To further examine whether residues 187-246 were necessary for the nuclear accumulation of PsbZIP1$_{113\text{-}246}$, those residues were trimmed from PsbZIP1$_{113\text{-}246}$-2XGFP, resulting in PsbZIP1$_{113\text{-}186}$-2XGFP. Deletion of residues 187-246 significantly reduced the nuclear staining (from LNC of 6.62 to 5.75), but the remaining nuclear accumulation was still strong, well above the activity of 113-126 (LNC=2. 54) (Fig. 3.2B and C). (Surprisingly, PsbZIP1$_{113\text{-}186}$-2XGFP produced primarily nucleolar localization in some transformants, shown in Fig. S3.2). Thus, residues 127-186 also appeared to contain sequences that contributed to nuclear localization. Indeed, PsbZIP1$_{127\text{-}186}$-2XGFP showed modest but significant nuclear accumulation (LNC=2.44) comparable to 113-126 (LNC=2.54) (Fig. 3.2B).

Taken together, these data suggested that the nuclear localization of bZIP1$_{113\text{-}246}$ was dependent on at least three distinct regions, namely residues 113-126, 127-186 and 187-246. Of these segments, residues 113-126 and 127-186 showed modest nuclear targeting activities by themselves, while residues 187-246 did not. Segment 113-126 combined with at least one of the

other two segments were necessary for substantial transport of PsbZIP1 into the *P. sojae* nucleus (LNC > 4.7), but all three were required for full activity (LNC= 6.62).

3.3.4 PsbZIP1$_{113-119}$ constitutes an independent NLS

To find the minimal region capable of conferring nuclear localization within residues 113-126, successive C-terminal deletions were assessed. Residues 113-123 (LNC=2.21) and 113-119 (LNC=2.23) produced similar nuclear localization as residues 113-126 (LNC=2.54, Fig. 3.3B), suggesting that the basic patch (124-RRR-126) made little contribution to the nuclear accumulation of PsbZIP1$_{113-126}$, and that residues 113-119 may be the minimal sequence within this region conferring NLS activity. Since segment 113-126 interacted synergistically with segments 127-167 and 187-246, it seemed possible that multiple copies of segment 113-119 might synergize to produce stronger localization. Indeed, we had previously shown (Fang & Tyler, 2016) that three tandem copies of PsbZIP1$_{113-126}$ (called PsNLS in that paper) could produce strong nuclear localization of a Cas9-GFP fusion. To confirm that no additional NLS was formed by the tandem copies, we attached two separated copies of segment 113-119 to 2XGFP, one at the N-terminus and one at the C-terminus of 2XGFP, creating PsbZIP1$_{113-119}$-2XGFP-PsbZIP1$_{113-119}$. This GFP fusion was strongly targeted to the nuclei (Fig. 3.3B). The LNC produced by this construct was 4.92, close to double the individual LNC (2.23 X 2 = 4.46), suggesting that each copy of segment 113-119 could multiply the activity of the other copy.

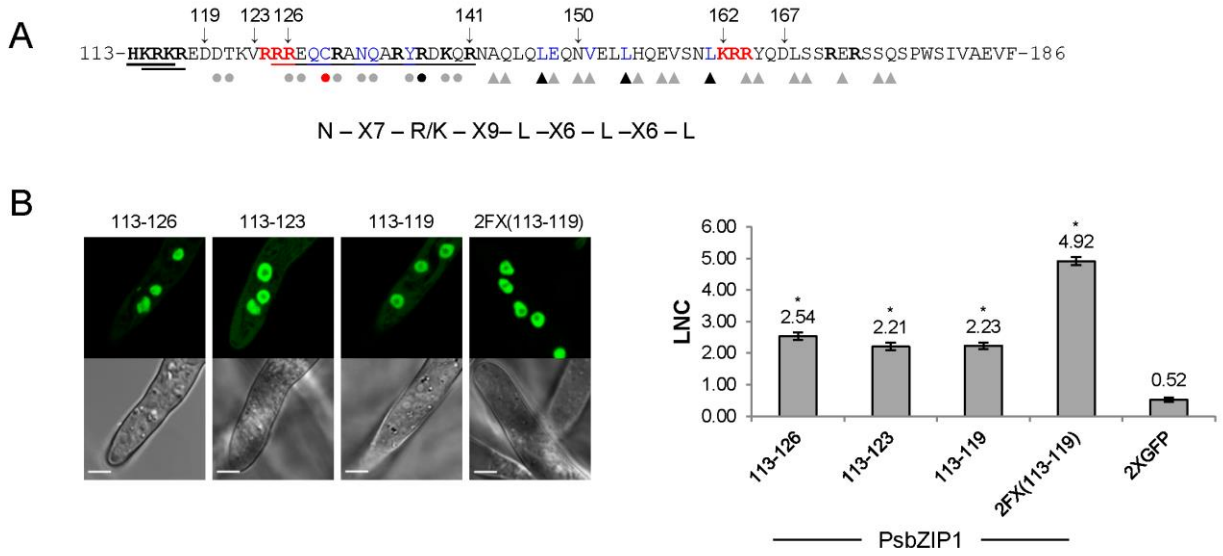


Fig. 3.3 Mutational analysis revealing segment PsbZIP1$_{113\text{-}119}$ can act as an independent NLS.

A. Amino acid sequence of PsbZIP1 residues 113-186. Dots and triangles denote predicted DNA binding residues and dimerization interface residues respectively, based on the annotation by NCBI. Residues corresponding to the bZIP TF consensus sequence shown below are highlighted in black dots and triangles. The novel substitution of residue N to C in PsbZIP1 is indicated by the red dot. Predicted mono- and bi-partite cNLS-like sequences are underlined. Amino acids subjected to mutation are in blue.

B. Subcellular localization of PsbZIP1$_{113\text{-}126}$ variants (left) and quantitation of their LNC ratio (right) as in Fig. 3.2. *, LNC values significantly greater (P value < 0.01) than the 2XGFP negative control. Here and in subsequent figures, 2FX indicates two copies of a fragment fused to the N-terminus and C-terminus of 2XGFP, respectively.

3.3.5 PsbZIP1$_{127\text{-}167}$ also constitutes an independent NLS

As already shown above, residues 127-186 mostly overlapped with a predicted bipartite cNLS which had little if any activity by itself (Fig. 3.3A).

To map the sequences responsible for the nuclear accumulation of PsbZIP1$_{127\text{-}186}$, a deletion series similar to that described for segment 113-126 was carried out. As shown in Fig. 3.4, residues 127-167 (LNC = 2.52) produced similar nuclear localization as residues 127-186

(LNC = 2.44), whereas residues 127-161 produced significantly less localization (LNC = 0.95), indicating that residues 162-167 contributed substantially to the activity of segment 127-167.

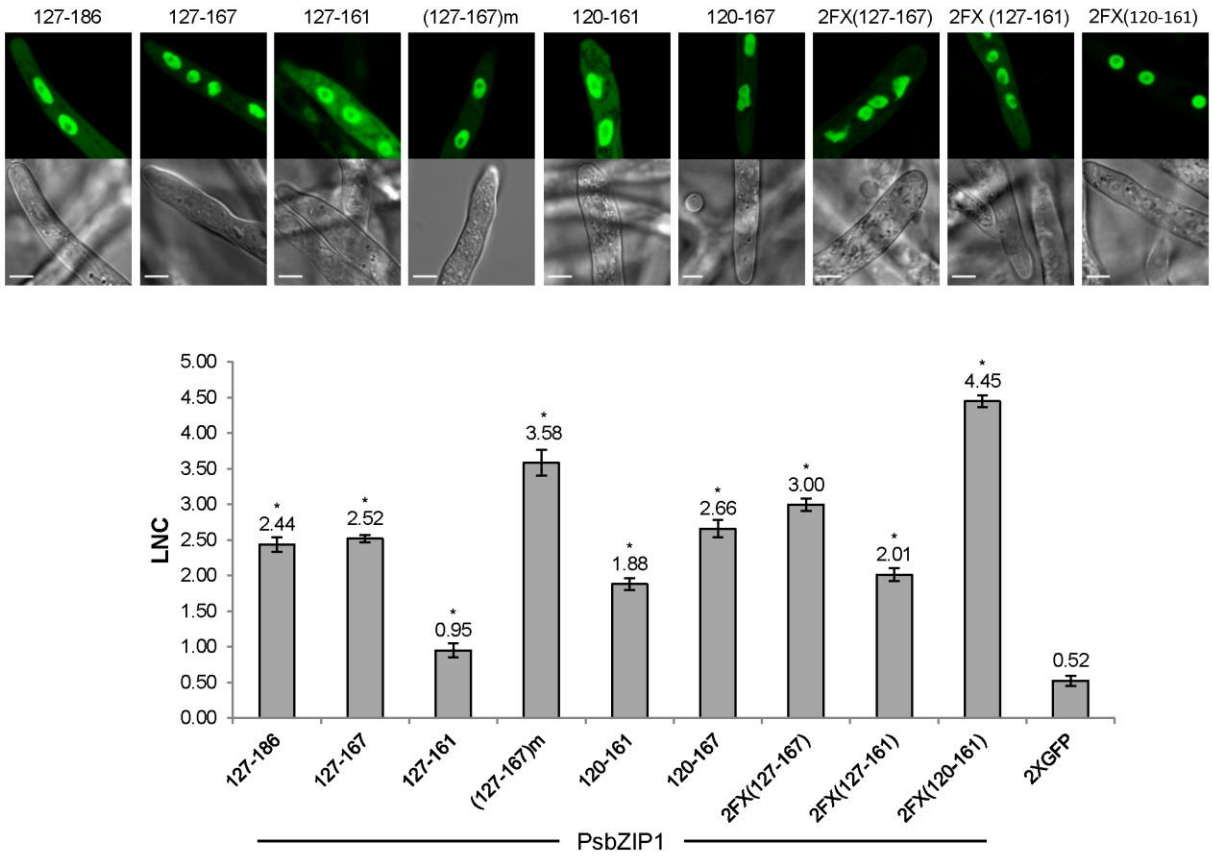


Fig. 3.4 Subcellular localization of various segments and mutants of PsbZIP1$_{120\text{-}186}$

Top, representative images; bottom, quantitation of nuclear localization as in Fig. 3.2.

(127-167)m denotes mutations Q17A/C18A/N21A/Q22A/Y25A/L36A/E37A/V40A/L43A/L50A within PsbZIP1$_{127\text{-}167}$.

*, LNC values significantly greater (P value < 0.01) than the 2XGFP negative control.

Because several residues within 127-167 were predicted to be involved in DNA binding and protein-protein interactions (Fig. 3.3A), we wanted to test their role in the nuclear retention of segment 127-167. We substituted most of the non-basic amino acids putatively involved in DNA binding or protein dimerization with alanines

(Q17A/C18A/N21A/Q22A/Y25A/L36A/E37A/V40A/L43A/L50A) (Fig. 3.3A). All these mutations combined within PsbZIP1$_{127\text{-}167}$ did not reduce the nuclear accumulation, and actually increased it slightly (Fig. 3.4).

To determine if segment 127-167 constituted an independent NLS (like segment 113-119), we again fused two copies of the segment to the two ends of 2XGFP, forming PsbZIP1$_{127\text{-}167}$-2XGFP- PsbZIP1$_{127\text{-}167}$. This double fusion was also strongly localized in the nuclei, but not much more than the single copy (LNC = 3.00 versus 2. 52) (Fig. 3.4). In contrast, two copies of the segment 127-161 (PsbZIP1$_{127\text{-}161}$ -2XGFP- PsbZIP1$_{127\text{-}161}$) produced around double the nuclear localization of the single copy (LNC, 2.01 versus 0.95) (Fig. 3.4).

3.3.6 124-RRR-126 and 162-KRR-164 independently enhance nuclear localization

In the experiments described above, segment 127-161 acted as an independent NLS, but produced less nuclear localization than segment 127-167, suggesting that the basic patch (162-KRR-164) could enhance nuclear accumulation directed by segment 127-161 (Fig. 3.4). A second basic patch (124-RRR-126) is located N-terminal to residues 127-161. To test if 124-RRR-126 also could enhance nuclear localization produced by segment 127-161, residues 120-161 were fused to 2XGFP; this fusion exhibited significantly stronger nuclear localization (LNC = 1.88) than residues 127-161 (LNC = 0.95), albeit a little lower than residues 127-167 (LNC = 2.52) (Fig. 3.4). Moreover, when present in two copies, as PsbZIP1$_{120\text{-}161}$-2XGFP-PsbZIP1$_{120\text{-}161}$, strong nuclear localization was observed (LNC = 4.45). This result suggested that both 124-RRR-126 and 162-KRR-164 could independently enhance localization produced by segment 127-161.

However, as noted above, 124-RRR-126 could only weakly enhance localization produced by segment 127-161. To further test the contribution of 124-RRR-126, we fused it to segment 127-167, producing PsbZIP1$_{120-167}$-2XGFP. However 120-167 did not show significantly increased localization compared to 127-167 (LNC of 2.66 versus 2.52). On the other hand, 120-167 did produce stronger nuclear localization than segment 120-161, which lacks 162-KRR-164 (LNC of 2.66 versus 1.88).

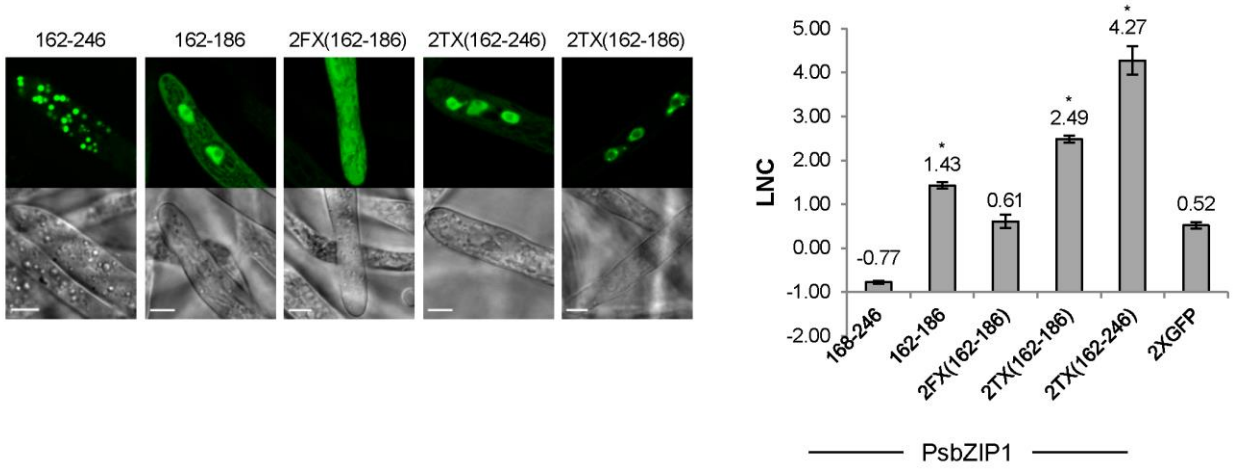


Fig. 3.5 Subcellular localization of various segments of PsbZIP1$_{165-186}$.

Top, representative images; bottom, quantitation of nuclear localization as in Fig. 3.2. 2TX, two copies of a fragment fused in tandem to the N-terminus of 2XGFP. *, LNC values significantly greater (P value < 0.01) than the 2XGFP negative control.

To further characterize how segment 162-KRR-164 contributes to nuclear localization, we compared the localization of PsbZIP1$_{162-246}$-2XGFP to that of PsbZIP1$_{168-246}$-2XGFP which alone produces exclusively cytoplasmic localization (LNC= -0.77) (Fig. 3.2 and Fig. 3.5). Unexpectedly, PsbZIP1$_{162-246}$-2XGFP was mislocalized into cytoplasmic speckles (Fig. 3.5). However, the shorter segment, PsbZIP1$_{162-186}$ fused to 2XGFP exhibited detectable nuclear accumulation (LNC=1.43) (Fig. 3.5). The double fusion construct, PsbZIP1$_{162-186}$-2XGFP-

PsbZIP1$_{162\text{-}186}$ showed even less nuclear localization than the single copy (Fig. 3.5), possibly because the basic patch (162-KRR-164) was masked when fused to the C-terminus of GFP. As an alternative test, we duplicated residues 162-186 in tandem. PsbZIP1$_{162\text{-}186}$-PsbZIP1$_{162\text{-}186}$ -2XGFP exhibited stronger nuclear localization (LNC=2.49) (Fig. 3.5). Interestingly, duplication of residues 162-246 in the same way diminished the 'speckle' localization and produced even stronger nuclear localization (LNC = 4.27). (Fig. 3.5)

3.3.7 New form of bipartite NLS in *P. sojae*

Comparison of the nuclear localized segments 127-167, 120-161 and 162-186 suggested a common sequence pattern of a three-basic-residues patch (positive head) combined with a certain upstream or downstream sequence containing scattered positive residues (tail) (Fig. 3.6). We named this pattern "positive head-tail" (PHT). Since the "head" can function either upstream or downstream of the "tail", this appears to be a form of bipartite NLS. As described above, the triplet of basic residues was required for nuclear accumulation in each case; deletion of the positive head resulted in compromised nuclear localization (Fig. 3.6A and C).

To further characterize the features of the tail sequence in the PHT model, we tested the role of the basic residues in the tail sequences; these basic residues are conserved in PibZIP1 and PITG_11668 (Fig. 3.1C). Conversion of all the basic residues in the tail sequences of PsbZIP1$_{120\text{-}161}$, PsbZIP1$_{127\text{-}167}$ and PsbZIP1$_{162\text{-}186}$ to alanines eliminated significant nuclear accumulation compared to 2XGFP (Fig. 3.6). However, segment 120-142 which contains five positive residues in the tail sequence showed much less nuclear localization than segment 120-161 (with the same number of positive residues) and 162-186 (only two positive residues), indicating that non-positively charged residues likely make important contributions to nuclear accumulation, also.

This characteristic distinguishes the *P. sojae* PHT bipartite NLS from the classical bipartite NLS.

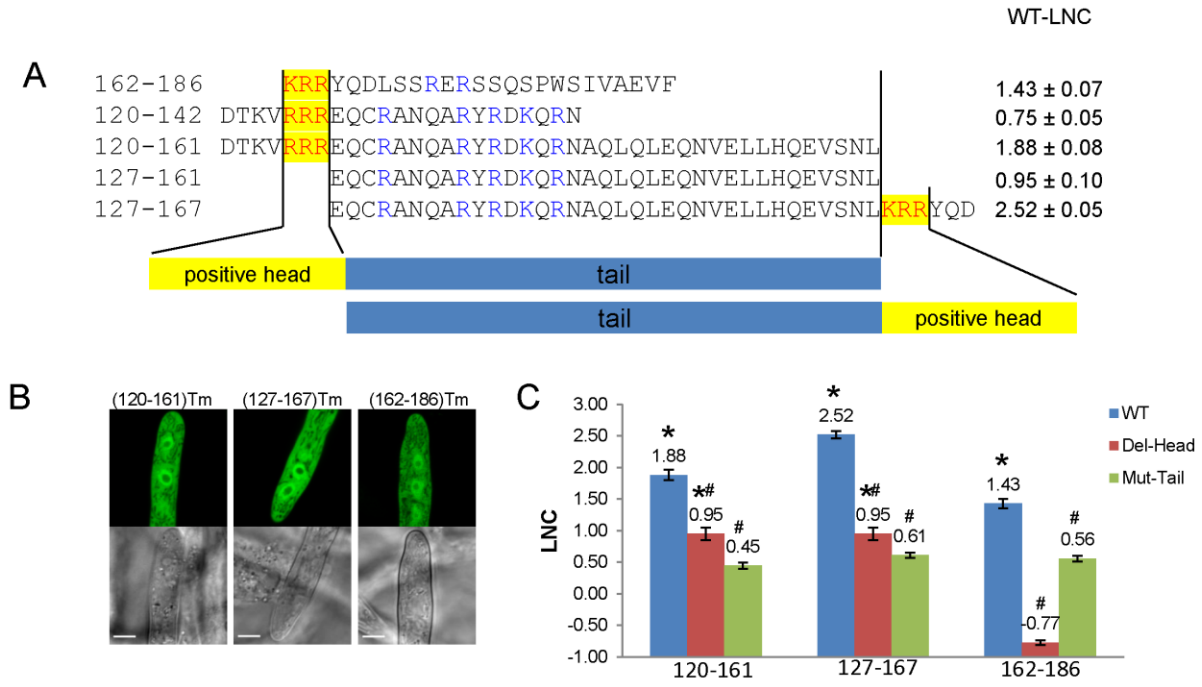


Fig. 3.6 'Positive head-tail' model (PHT) of novel *P. sojae* bipartite NLSs.

A. Sequence comparison of several tested segments highlighting the activities of the proposed "head" and "tail" sequences. Positive head residues are in red and positive tail residues are in blue.

B. Subcellular localization of segments containing alanine substitution mutations in positively charged tail residues (Tm; segments containing mutations in the residues highlighted in blue in A). Representative images are shown.

C. Quantitation of localization showing the contributions of 'Head' and 'Tail' sequences to the nuclear localization produced by segments 120-161, 127-167, and 162-186. WT, wild type; Del-Head, deletion of the positive 'head' sequence (127-161 for both 120-161 and 127-167; 168-246 for 162-186); Mut-Tail, mutants shown in B. *, LNC values significantly greater (P value < 0.01) than the 2XGFP negative control. # Mutants with significantly less (P value < 0.01) nuclear localization than the respective wild type.

3.3.8 Contribution of segment 187-246 to nuclear localization

Segment 165-246 did not produce nuclear localization by itself, indicating that no autonomous NLS elements occur in this region. However, four comparisons suggested that segment 187-246 could provide a modest enhancement of nuclear localization by upstream sequences. First, PsbZIP1$_{113-246}$-2XGFP (LNC = 6.62) exhibits significantly stronger nuclear localization than PsbZIP1$_{113-186}$-2XGFP (LNC = 5.75). Second, PsbZIP1$_{127-246}$-2XGFP (LNC = 3.03) exhibits significantly stronger nuclear localization than PsbZIP1$_{127-186}$-2XGFP (LNC = 2.44). Third, two copies of 162-246 (LNC = 4.27) showed significantly stronger nuclear localization than two copies of residues 162-186 (LNC = 2.49). Fourth, PsbZIP1$_{113-246}\Delta$(127-186)-2XGFP (LNC = 4.77) exhibits significantly stronger nuclear localization than PsbZIP1$_{113-126}$-2XGFP (LNC = 2.54). Thus segment 187-246, which contains 7 positively charged residues, appeared capable of contributing to nuclear localization, sufficient to increase LNC by an average of around 1.07.

3.4 DISCUSSION

bZIP TFs form a large class of transcriptional regulators that constitute the most conserved superfamily of TFs in eukaryotes (Miller, 2009; Amoutzias *et al.*, 2007; Marco Llorca *et al.*, 2014). In oomycetes, 38 bZIP TFs have been found in *P. infestans* (Gamboa-Melendez *et al*., 2013), and 71 in *P. sojae* (Ye *et al*., 2013), though these numbers may not be fully accurate due to errors in gene models such as we found for PHYSO_256931 and PHYSO_481435. Only a few *Phytophthora* bZIP TFs have been characterized, such as Pibzp1 and PITG_11668 (Blanco & Judelson, 2005; Gamboa-Melendez *et al.*, 2013). In *P. infestans*, GFP tagged PITG_11668 localized homogeneously in the nucleus (Gamboa-Melendez *et al.*, 2013), while this protein and

its paralog Pibzp1 displayed a punctate distribution in the *P. sojae* nucleus. Interestingly, the native PsbZIP1 also produced homogeneous nuclear accumulation. We don't know if the punctate localization was an artifact caused by overexpression, but the same pattern was observed for the HTLV bZIP TF, HBZ (Hivin *et al.*, 2005).

In Chapter 2, we found that many fully functional NLSs in *P. sojae* gain their functionality by combining an otherwise weak NLS with additional clusters of basic residues or with other weak NLSs. Here, through a comprehensive dissection of nuclear localization by *P. sojae* bZIP transcription factor PsbZIP1, we have shown that this factor achieves efficient nuclear localization via the action of at least four adjacent regions of the protein, of which three are individually sufficient for a moderate level of nuclear localization (Fig. 3.7). The first region (NLS1), encompassing amino acids 113 to 119, contains the five residues HKRKR, which are conserved in many other *Phytophthora* bZIP TFs. The second and third regions (NLS2 and NLS3), encompassing residues 120 to 161 and 162 to 186 respectively, have a common sequence pattern, namely a triplet of positively charged residues, followed by a longer region containing multiple non-contiguous positively charged residues. The fourth region, 187-246, cannot function as an independent NLS, but provides about a 2-fold enhancement of nuclear localization provided by the other NLS sequences – we named this the nuclear localization enhancer (NLE).

NLS1 resembles a canonical monopartite cNLS, with the addition of an extra basic residue. As noted in Chapter 2, a number of other *P. sojae* nuclear proteins use an extended monopartite cNLS for nuclear localization. NLS2 and NLS3 appear to represent a new form of NLS, not previously documented among known mammalian or yeast NLS sequences. Since this NLS consists of two elements whose order can be switched, namely a triplet of positively charged residues, followed by a longer region containing multiple, non-contiguous, positively-

charged residues, we have named this the "positive head-tail" (PHT) bipartite NLS. In the case of NLS2, the "tail region", segment 127-161, contains 5 positive residues, and exhibited modest independent NLS activity that was enhanced by inclusion of its "head sequence" 124-RRR-126. The NLS activity of 127-161 could also be enhanced by inclusion of the adjacent "head" sequence from NLS3, namely 162-KRR-164. NLS2 encompasses most of the DBD and LZP domains of PsbZIP1. However, mutation of all the key non-charged residues of the DBD and LZP domains within NLS2 did not reduce the NLS activity of NLS2, but actually increased it slightly (Fig. 3.4). The "tail" region of NLS3, segment 167-186, contains 2 positive residues, but did not exhibit independent NLS activity either by itself, or in combination with the adjacent enhancer region, segment 187-246. However, when combined with the head sequence, 162-KRR-164, it gained NLS activity. NLS3 was somewhat weaker than NLS2. The PHT NLSs may be functionally related to classical bipartite NLSs. However, segment 120-142, which contains all the positive residues of NLS2 as well as the predicted canonical bipartite cNLS, had only very weak NLS activity (LNC = 0.75), suggesting a role for uncharged or hydrophobic residues in the tail for its activity, thus distinguishing the PHT NLS from canonical bipartite cNLSs. Similarly, NLS3 only contains two positive residues in its tail, but many hydrophobic residues. The possible functional relationship between PHT and classical bipartite NLSs may eventually be resolved when the binding of PHT NLSs to *P. sojae* karyopherins is characterized.

Nuclear transport mediated by multiple NLSs has been documented in several types of DNA and RNA-binding proteins (Garcia-Bustos *et al.*, 1991; Dworetzky *et al.*, 1988). As observed here, this may increase the efficiency of localization by these proteins. Multiple NLS-mediated nuclear transport was also observed in a human bZIP TF, HBZ (Hivin *et al.*, 2005). At least two out of three distinct motifs (two basic regions and one DBD) that harbored weak NLS

activities were required to translocate HBZ into the nucleus, though localization was not quantitated in that study.

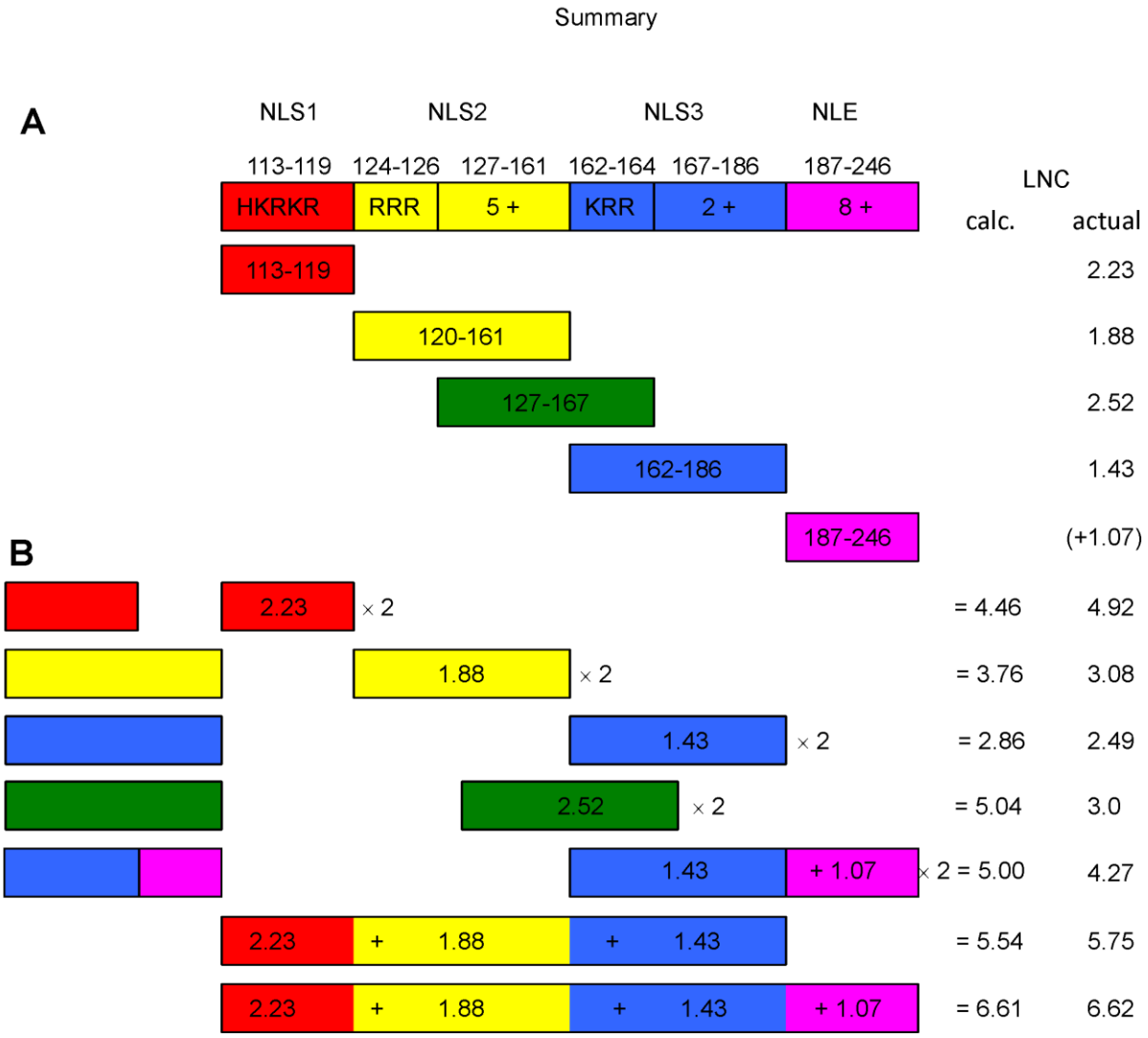


Fig. 3.7 Summary of the NLS sequences in PsbZIP1 and how they interact.

A. Summary of the three independent NLS sequences, including key sub-sequences, and the nuclear localization enhancer (NLE) region of PsbZIP1, as inferred from this study, together with the LNC conferred by each sequence individually. The green box shows a hybrid of NLS2 and NLS3 that also exhibits independent NLS activity. NLE does not confer autonomous localization but adds around 1.07 (averaged over four constructs) to the LNC of constructs containing it compared to those that lack it (see text). 5+, 2+, 8+, positively charged amino acids in each segment.

B. LNC conferred by combinations of NLS sequences (actual) compared to the sums of the constituent NLS or NLE sequences (calc.). The additivity of the LNC values (Pearson's correlation coefficient r^2 = 0.91 excluding 2X(127-164) suggests that the NLSs and NLE interact independently to multiply each other's contributions.

Here, we quantified the nuclear localization activity conferred by each protein segment as $\log_2$(nuclear-to-cytoplasmic ratio) (LNC). In doing so, we noticed that the LNC values of individual NLS elements were often additive when two or more NLS segments were combined into a single construct (summarized in Fig. 3.7B). This additive property was observed with different combinations of NLS1, NLS2 and NLS3, and also when two copies of each NLS were fused to 2xGFP. For example, when NLS1 (LNC = 2.23), NLS2 (LNC = 1.88), and NLS3 (LNC = 1.43) were combined in segment 113-186, the resultant LNC was 5.75, remarkably similar to the sum of the LNC values of the three constituent NLSs, (2.23+1.88+1.43= 5.54). Similarly, when two copies of NLS1 (LNC = 2.23) were combined, the fusion displayed an LNC of 4.92 (2.23X2=4.46). Furthermore, while residues 162-246 fused to 2XGFP were mislocalized into cytoplasmic speckles, half the LNC produced by its two copies (4.27/2 = 2.14) was similar to the combination of 162-186 and 187-246 (1.43+1.07= 2.50). This property suggests a simple model in which each NLS multiplies the contributions of the other NLSs in a fully independent manner. A simple mechanistic interpretation of this model might be that each NLS binds an additional karyopherin molecule. The NLE may also act in this manner, though its interactions with karyopherin might be too weak to confer independent NLS activity. Under our model, nuclear localization activity is represented as a quantitative property rather than a qualitative (in or out) property of proteins.

In our study, we found that a hybrid of NLS2 and NLS3 (127-167) also exhibited independent NLS activity. However, unlike the other three NLSs, two copies of the segment did

not produce a doubling of the LNC. Possibly, unfavorable protein folding when the segment was fused to the C-terminus of 2XGFP may have produced this result. Alternatively, the presumptive karyopherins responsible for import of segment 127-167 might be distinct from the karyopherins responsible for import of the other NLSs.

In mammalian and plant cells, many basic DBDs in bZIP TFs were found to be bifunctional, contributing to both DNA binding and nuclear translocation (Miller, 2009). Typically, the nuclear translocation function was modulated by a bipartite cNLS overlapping with the DBD (Miller, 2009; Marco Llorca *et al.*, 2014). The overlap of DBD and NLS was also observed in other TF families (LaCasse & Lefebvre, 1995; Cokol *et al.*, 2000). However, point-mutation analyses also indicated that the DNA binding and nuclear localization properties were not necessarily related in some DNA binding proteins (Matheny *et al.*, 1994; Varagona & Raikhel, 1994). In contrast, we found that a predicted canonical bipartite cNLS within the DBD of PsbZIP1 displayed only very weak nuclear localization activity by itself, but proved to be a part of a longer, more active NLS region, namely NLS2. This observation is in agreement with our accompanying study in Chapter 2 that canonical cNLSs generally work poorly in *P. sojae*. Comparison of the bipartite NLS motif predicted in the PsbZIP1 DBD (9-DTKVRRREQCRANQARYRDKQRN-31) with those functional bipartite cNLSs reported in Chapter 2 revealed that the C-terminal basic cluster of the DBD within PsbZIP1 lacks the consecutive lysine or arginine residues characteristic of bipartite NLSs that are strongly active in *P. sojae*.

The role of retention in nuclear localization has been controversial (LaCasse & Lefebvre, 1995). It has been reported that nuclear accumulation required a nuclear retention signal that mediates binding to other nuclear components, such as DNA (Suel & Chook, 2009) or non-

histone proteins (Spelsberg *et al.*, 1984), following nuclear entry. However, Garcia-Bustos *et al.* (1991) reported that nuclear localization of proteins is not associated with DNA binding. By selective mutational analysis, we found that the nuclear localization of PsbZIP1 is independent of the uncharged residues in its predicted DNA binding and protein dimerization motifs (Figs. 3.2 and 3.4).

In Chapter 2, we found that many fully functional NLSs in *P. sojae* consist of a weak NLS in combination with either other basic clusters or other weak NLSs. For example, in the *P. sojae* mRNA cleavage and polyadenylation factor I subunit, PHYSO_251824, efficient nuclear localization required three adjacent weak NLSs in the C-terminus, two of them PY-NLSs and one a monopartite cNLS. Similarly, in the *P. sojae* mRNA cleavage and polyadenylation factor II subunit, PHYSO_480605, nuclear localization was mediated by a PY-NLS at the N-terminus in combination with a nearby basic cluster, KFKGK, plus additional downstream sequences. Our finding that nuclear localization of PsbZIP1 is dispersed across four distinct adjacent regions is consistent with a general model for nuclear localization in *P. sojae* nuclear proteins in which NLS activity is dispersed across multiple individually weak elements. We speculate that this strategy facilitates the diversification of *P. sojae* nuclear proteins, which in turn facilitates the genetic diversification of this pathogen, which is locked into a co-evolutionary struggle with its hosts. Consistent with this idea, Gamboa-Melendez *et al.* (2013) and Ye *et al.* (2013) have found that nearly half of the *Phytophthora* bZIPs exhibit novel substitutions at some conserved residues in the DBD that are predicted to be important for DNA sequence recognition. A relaxed contribution to NLS activity by the DBD would provide more flexibility for the DBD to diversify.

3.5 MATERIALS AND METHODS

3.5.1 *P. sojae* strains, growth conditions, and transformation

The reference *P. sojae* strain P6497 (race 2) was used throughout the study. *P. sojae* strains were routinely cultivated on solid V8 agar medium at 25 ℃ in the dark. A transient gene expression assay based on an improved polyethylene glycol (PEG) mediated protoplast transformation protocol (Fang & Tyler, 2016) was used to assay nuclear localization (See Chapter 2). Briefly, *P. sojae* protoplasts were incubated with DNA in the presence of PEG 4000 (Sigma 81240) and re-generated overnight in liquid nutrient media. Regenerated protoplasts were collected and grown as colonies in solid media containing 50 μg ml^{-1} G418 (Geneticin) for 2 days. G418-resistant colonies were transferred to 12-well plates containing liquid V8 media supplemented with 50 μg ml^{-1} G418 and incubated for 2~3 days at 25 ℃ before confocal microscopy observation.

3.5.2 Sequence information

Phytophthora protein IDs and sequences were obtained from FungiDB (http://fungidb.org/fungidb/). Their corresponding accession numbers in Genbank are: PHYSO_256931 (XP_009520137), Pibzp1 (Q5BUB4.1), PITG_11668 (XP_002901207.1). The corrected PsbZIP1 gene model is being submitted to the GenBank database.

3.5.3 Construction of plasmids

All the primers used in this study are listed in Table S1 in the supplemental material.

To generate full-length bZIP homologs fused to GFP, the *bZIP* genes were PCR amplified from *P. infestans* genomic DNA or *P. sojae* cDNA, and inserted into the *Stu* I site of pYF2-GFP (Fang

& Tyler, 2016). To create N-terminal 2XGFP fusions, the plasmid backbone pYF2-2XGFP was used. Fragments smaller than 80 bp were inserted into the *Sac* II and *Spe* I sites by oligo-annealing; other fragments were PCR-amplified and inserted into the *Stu* I site. To construct the C-terminal 2XGFP fusions, the plasmid backbone pYF3-2XGFP described in Chapter 2 was used. Segments smaller than 80 bp were inserted into the *Bsr* GI and *Bsp* EI sites by oligo-annealing; otherwise, fragments were PCR-amplified and inserted into the *Hpa* I site. To insert two copies of an NLS flanking each terminus of 2XGFP in pYF3-2XGFP, one copy was inserted into the *Stu* I site at the N-terminus, and in a second step another copy was inserted into the *Hpa* I site at the C-terminus. To insert two tandem copies of an NLS at the N-terminus of 2XGFP, each NLS was cloned by sequential ligation into the *Stu I* site and into the *Sac* II-*Spe* I sites of pYF2-2XGFP.

Standard molecular techniques were performed as described in Molecular Cloning (Sambrook & Russell, 2001) or according to instructions from a kit manufacturer.

3.5.4 Confocal imaging of *P. sojae* transformants

Laser scanning confocal microscopy (Zeiss LSM 780 NLO) was used to examine the subcellular localization of fluorescent protein fusions in *P. sojae*. *P. sojae* specimens were collected from clumps of 2-3 days old mycelia grown in liquid V8 media, then washed and maintained in modified Plich media (0.5 g KH_2PO_4, 0.25 g $MgSO_4 \cdot 7H_2O$, 1 g asparagine, 1 mg thiamine, 0.5 g yeast extract, 10 mg β–sitosterol, 25 g glucose dissolved in 1 L water) before observation. Images were captured using a 63X oil objective with excitation/emission settings (in nm) 488/504-550 for GFP, and 561/605-650 for mCherry. For each sample, at least three independent *P. sojae* transformant colonies were analyzed. The microscope's built-in Zen 2012 software was

used for post-processing of images. To quantify the nuclear to cytoplasmic fluorescence ratio, ~30 nuclei and the corresponding cytoplasmic regions were randomly selected. In most cases, nuclei were identified by the morphology of the GFP-stained region (a large, uniformly stained, irregular ovoid region, often with an unstained nucleolus), and were not confirmed by co-expression of H2B-mCherry. Where nuclei could not be identified due to very poor nuclear localization, H2B-mCherry was co-expressed to verify the nuclei. The fluorescence intensities of the selected regions were measured using the 'measure' tool of the Zen 2012 Blue edition. All statistical analysis was performed using GraphPad Prism 7. Log_2(Nuc:Cyt) was calculated using the mean of the log_2-transformed ratios from pairs of nuclear and cytoplasmic regions. Statistical significance was tested with a two-sample, unpaired t-test of the log-transformed ratios with a *P* value cutoff of 0.01.

3.6 ACKNOWLEDGEMENT

We thank Howard Judelson (UC-Riverside) for helpful discussions and for plasmids, and Niklaus Grünwald (USDA ARS, Horticultural Crops Research Lab, Corvallis) for providing *P. infestans* genomic DNA. This work was supported in part by USDA NIFA grant #2011-68004-30104 to B.M.T.

3.7 REFERENCES

Amoutzias, G.D., A.S. Veron, J. Weiner, 3rd, M. Robinson-Rechavi, E. Bornberg-Bauer, S.G. Oliver & D.L. Robertson, (2007) One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Mol. Biol. Evol.* **24**: 827-835.

Blanco, F.A. & H.S. Judelson, (2005) A bZIP transcription factor from *Phytophthora* interacts with a protein kinase and is required for zoospore motility and plant infection. *Mol. Microbiol.* **56**: 638-648.

Chook, Y.M. & K.E. Suel, (2011) Nuclear import by karyopherin-betas: recognition and inhibition. *Biochim. Biophys. Acta* **1813**: 1593-1606.

Cokol, M., R. Nair & B. Rost, (2000) Finding nuclear localization signals. *EMBO Rep* **1**: 411-415.

Dworetzky, S.I., R.E. Lanford & C.M. Feldherr, (1988) The effects of variations in the number and sequence of targeting signals on nuclear uptake. *The J. of Cell Biology* **107**: 1279-1287.

Erwin, D.C. & O.K. Ribeiro, (1996) *Phytophthora diseases worldwide*. American Phytopathological Society (APS Press, Minneapolis).

Fang, Y. & B.M. Tyler, (2016) Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* **17**: 127-139.

Gamboa-Melendez, H., A.I. Huerta & H.S. Judelson, (2013) bZIP transcription factors in the oomycete *Phytophthora infestans* with novel DNA-binding domains are involved in defense against oxidative stress. *Eukaryot. Cell* **12**: 1403-1412.

Garcia-Bustos, J., J. Heitman & M.N. Hall, (1991) Nuclear protein localization. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1071**: 83-101.

Hivin, P., M. Frédéric, C. Arpin-André, J. Basbous, B. Gay, S. Thébault & J.-M. Mesnard, (2005) Nuclear localization of HTLV-I bZIP factor (HBZ) is mediated by three distinct motifs. *J. Cell Sci.* **118**: 1355-1362.

Hurst, H.C., (1993) Transcription factors. 1: bZIP proteins. *Protein profile* **1**: 123-168.

Jakoby, M., B. Weisshaar, W. Dröge-Laser, J. Vicente-Carbajosa, J. Tiedemann, T. Kroj & F. Parcy, (2002) bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* **7**: 106-111.

LaCasse, E.C. & Y.A. Lefebvre, (1995) Nuclear localization signals overlap DNA-or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res.* **23**: 1647.

Lange, A., R.E. Mills, C.J. Lange, M. Stewart, S.E. Devine & A.H. Corbett, (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282**: 5101-5105.

Lee, B.J., A.E. Cansizoglu, K.E. Suel, T.H. Louis, Z. Zhang & Y.M. Chook, (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **126**: 543-558.

Marco Llorca, C., M. Potschin & U. Zentgraf, (2014) bZIPs and WRKYs: two large transcription factor families executing two different functional strategies. *Frontiers in Plant Science* **5**.

Matheny, C., M.L. Day & J. Milbrandt, (1994) The nuclear localization signal of NGFI-A is located within the zinc finger DNA binding domain. *J. Biol. Chem.* **269**: 8176-8181.

Miller, M., (2009) The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Current Protein and Peptide Science* **10**: 244-269.

Sambrook, J. & D.W. Russell, (2001) Molecular cloning: a Laboratory Manual 3rd edition. Cold Spring-Harbour Laboratory Press, UK.

Spelsberg , T.C., B.J. Gosse, B.A. Littlefield, H. Toyoda & R. Seelke, (1984) Reconstitution of nativelike nuclear acceptor sites of the avian oviduct progesterone receptor: evidence for involvement of specific chromatin proteins and specific DNA sequences. *Biochemistry* **23**: 5103-5113.

Suel, K.E. & Y.M. Chook, (2009) Kap104p imports the PY-NLS-containing transcription factor Tfg2p into the nucleus. *J. Biol. Chem.* **284**: 15416-15424.

Suel, K.E., H. Gu & Y.M. Chook, (2008) Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals. *PLoS Biol.* **6**: e137.

Tyler, B.M., (2007) *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* **8**: 1-8.

Varagona, M.J. & N.V. Raikhel, (1994) The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. *Plant J.* **5**: 207-214.

Ye, W., Y. Wang, S. Dong, B.M. Tyler & Y. Wang, (2013) Phylogenetic and transcriptional analysis of an expanded bZIP transcription factor family in *Phytophthora sojae*. *BMC Genomics* **14**: 839.

3.8 SUPPORTING INFORMATION

3.8.1 Fig. S3.1

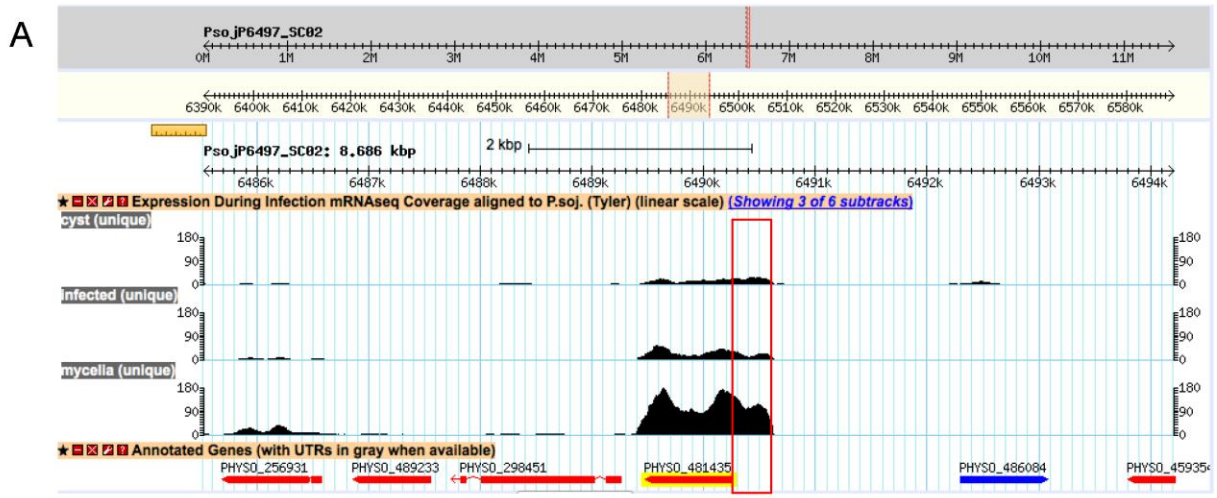


Fig. S3.1 Correction of the PsbZIP1 gene model.

A. RNAseq data indicating an additional transcribed region (in red rectangle) at the N-terminus of the annotated gene model.

B. Amino acid sequence of the corrected PsbZIP1

C. Screenshots showing the alignment of PsbZIP1 with its *P. infestans* orthologs PITG_11668 (left) and Pibzp1 (right)

3.8.2 Fig. S3.2

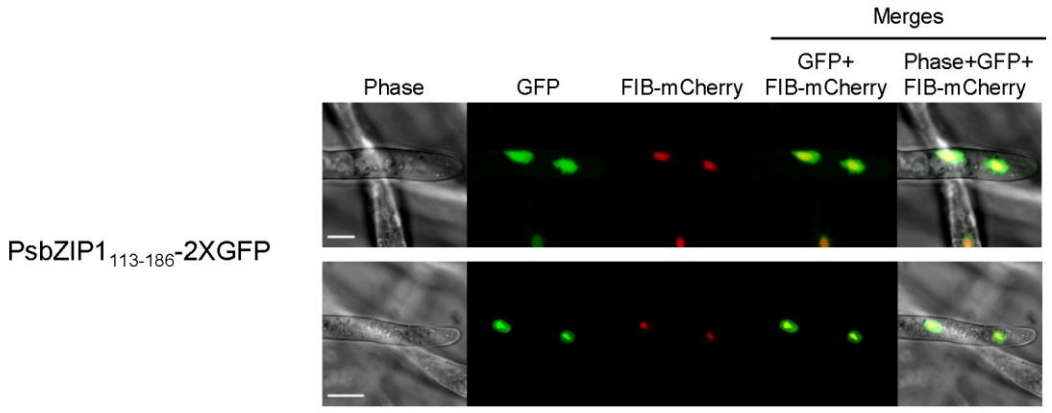


Fig. S3.2 PsbZIP1$_{113-186}$-2XGFP produced primarily nucleolar localization in some transformants.

Representative images from different transformants are shown. Nucleoli were confirmed by the nucleolar marker FIB-mCherry.

3.8.3 Table S3.1

Table S3.1 Primers used in the study

Name	Sequence	Usage
PsbZIP1_F	5' ATGAGCCGCTACCCGTTTGCC 3'	PCR amplify full length PsbZIP1 and its truncation series
PsbZIP1_113_F	5' ATGCACAAGCGCAAGCGCG 3'	
Ps479_120_F	5' ATGGACACCAAGGTCCGTCG 3'	
PsbZIP1_127_F	5' ATGGAGCAGTGTCGAGCCAACCAAGC 3'	
Ps479772_162_F	5' ATGAAGCGAAGGTACCAAGATCTGTCGTC 3'	
PsbZIP1_168_F	5' ATGCTGTCGTCGCGCGAGAGGAGC 3'	
PsbZIP1_247_F	5' TGCTTCGGCGACCCTCGC 3'	
PsbZIP1_112_R	5' CATGTCCTTGGTCTTCACTGAGTTTG 3'	
PsbZIP1_142_R	5' GTTGCGCTGCTTGTCTCGG 3'	
PsbZIP1_161_R	5' CAAGTTGCTGACTTCTTGGTGCAGTAGC 3'	
PsbZIP1_167_R	5' ATCTTGGTACCTTCGCTTCAAGTTGC 3'	
PsbZIP1_186_R	5' GAAGACCTCCGCCACGATGCTC 3'	
PsbZIP1_246_R	5' CTGCGAGTACAGTCGCAGTTGCTC 3'	
PsbZIP1_R	5' ATCAGTGAGACTACCGATGGTGCACTCC 3'	
PsbZIP1_162_SacII_F	5' C*CCCGCGG*AAGCGAAGGTACCAAGATCTGTCGTC 3'	PCR amplify PsbZIP1 truncations for cloning multiple copies of segments
PsbZIP1_246_SpeI_R	5' GG*ACTAGT*CTGCGAGTACAGTCGCAGTTGCTC 3'	
PsbZIP1_186_SpeI_R	5' GCGG*ACTAGT*GAAGACCTCCGCCACGATGCTC 3'	
PsbZIP1del127-141_F	5' GGTCCGTCGTCGCAACGCCCAGCTGC 3'	Delete residues 127-141 of $PsbZIP1_{113-246}$
PsbZIP1del127-141_R	5' GCAGCTGGGCGTTGCGACGACGGACC 3'	
PsbZIP1del143-167_F	5' CAAGCAGCGCAACCTGTCGTCGCGCG 3'	Delete residues 143-167 of $PsbZIP1_{113-246}$
PsbZIP1del143-167_R	5' CGCGCGACGACAGGTTGCGCTGCTTG 3'	
PsbZIP1del127-167_F	5' GTCCGTCGTCGCCTGTCGTCGCGC 3'	Delete residues 127-167 of $PsbZIP1_{113-246}$
PsbZIP1del127-167_R	5' GCGCGACGACAGGCGACGACGGAC 3'	
PsbZIP1del127-186_F	5' GGTCCGTCGTCGCCACCTGCTCGAGA 3'	Delete residues 127-186 of $PsbZIP1_{113-246}$
PsbZIP1del127-186_R	5' TCTCGAGCAGGTGGCGACGACGGACC 3'	
PsbZIP1_113-126_SacII_up	5' GGATGCACAAGCGCAAGCGCGAGGACGACACCAAGGTCCGTCGTCGCA 3'	Oligo annealing, clone residues 113-126 to the N-terminus of 2XGFP
PsbZIP1_113-126_SpeI_down	5' CTAGTGCGACGACGGACCTTGGTGTCGTCCTCGCGCTTGCGCTTGTGCATCCGC 3'	
PsbZIP1_113-123_SacII_up	5' GGATGCACAAGCGCAAGCGCGAGGACGACACCAAGGTCA 3'	Oligo annealing, clone residues 113-123 to the N-terminus of 2XGFP

PsbZIP1_113-123_SpeI_down	5' CTAGTGACCTTGGTGTCGTCCTCGCGCTTGCGCTTGTGCATCCGC 3'	
PsbZIP1_113-119_SacII_up	5' GGATGCACAAGCGCAAGCGCGAGGACA 3'	Oligo annealing, clone residues 113-119 to the N-terminus of 2XGFP
PsbZIP1_113-119_SpeI_down	5' CTAGTGTCCTCGCGCTTGCGCTTGTGCATCCGC 3'	
PsbZIP1_113-119_BsrG1_up	5' GTACAATGCACAAGCGCAAGCGCGAGGACT 3'	Oligo annealing, clone residues 113-119 to the C-terminus of 2XGFP
PsbZIP1_113-119_BspEI_down	5' CCGGAGTCCTCGCGCTTGCGCTTGTGCATT 3'	
PsbZIP1_127-167_m	5' CGTTGCGCTGCTTGTCTCGGGCTCGGGCAGCGGCGGCTCGAGCCGCCTCCATCCTATCGATAAGCT 3'	Mutate Q17A/C18A/N21A/Q22A/Y25A/L36A/E37A/V40A/L43A/L50A within $PsbZIP1_{127-167}$
PsbZIP1_127-161_Tm1_R1	5' CTGCAGCTGGGCGTTGGCCTGCGCGTCTGCGTATGCGGCTTGGTTGGCTCGA 3'	Mutate all basic amino acids (R135A/R137A/K139A/R141A) in residues 127-161
PsbZIP1_127-161_Tm1_R2	5' GCTTGGTTGGCTGCACACTGCTCCATCCTATCGATAA 3'	
PsbZIP1_162-186_Tm_F	5' GGACTCTGGCTGCTCGCCTCGGCCGACGACAGATCTTG 3'	Mutate all basic amino acids (R171A/R173A) in residues 162-186
PITG_11668_F	5' ATGAACCGCTACCCTTTTGCC 3'	PCR amplify the full length *P. infestans* bZIP TF, PITG_11668
PITG_11668_R	5' CAAGCTAGTGCAGTCGCTGCTTC 3'	
Pibzp1_F	5' ATGAATCTCAGCCGTTTCGCTCC 3'	PCR amplify the full length *P. infestans* bZIP TF, Pibzp1
Pibzp1_R	5' GTATTGCGAGAACTGCAGCAGCTG 3'	

Mutate R in PsbZIP1(51-75)

Chapter 4

Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9

Yufeng Fang[1, 2] and Brett M. Tyler[1, 2*]

[1]Interdisciplinary Ph.D. program in Genetics, Bioinformatics & Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA

[2] Center for Genome Research and Biocomputing and Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

* Corresponding author: Brett.Tyler@oregonstate.edu

This Chapter includes a research article published in Molecular Plant Pathology 17.1 (2016): 127-139 as "Efficient Disruption and Replacement of an Effector Gene in the Oomycete Phytophthora sojae using CRISPR/Cas9". Brett M. Tyler contributed to analyzing the data and editing the manuscript.

4.1 ABSTRACT

Phytophthora sojae is an oomycete pathogen of soybean. Due to its economic importance, *P. sojae* has become a model for the study of oomycete genetics, physiology and pathology. The lack of efficient techniques for targeted mutagenesis and gene replacement have long hampered genetic studies of pathogenicity in *Phytophthora* species. Here, we describe a CRISPR/Cas9 system enabling rapid and efficient genome editing in *P. sojae*. Using the RXLR effector gene *Avr4/6* as a target, we observed that in the absence of a homologous template, the repair of Cas9-induced DNA double-strand breaks (DSBs) in *P. sojae* was mediated by non-homologous end joining (NHEJ), primarily resulting in short indels. Most mutants were homozygous, presumably due to gene conversion triggered by Cas9-mediated cleavage of non-mutant alleles. When donor DNA was present, homology directed repair (HDR) was observed, which resulted in the replacement of *Avr4/6* with the *NPTII* gene. By testing the specific virulence of several NHEJ mutants and HDR -mediated gene replacements on soybeans, we have validated the contribution of Avr4/6 to recognition by soybean R gene loci, *Rps*4 and *Rps*6, but also uncovered additional contributions to resistance by these two loci. Our results establish a powerful tool for studying functional genomics in *Phytophthora*, which provides new avenues for better control of this pathogen.

4.2 INTRODUCTION

Phytophthora sojae causes “damping off” of soybean seedlings as well as stem and root rot of established plants (Tyler, 2007). Morphologically and physiologically, oomycetes such as *P. sojae* resemble filamentous fungi, but evolutionally they are classified in the kingdom

Stramenopila (Tyler, 2001). Most *Phytophthora* species are plant pathogens that can damage a huge range of agriculturally and ornamentally important plants (Erwin & Ribeiro, 1996). Because of its economic impact, *P. sojae,* along with *P. infestans,* has been developed as a model species for the study of oomycete plant pathogens (Tyler, 2007).

The first two genome sequences of oomycetes (*P. sojae and P. ramorum*) were published approximately nine years ago (Tyler *et al.*, 2006), but functional genomics studies have been hampered by the lack of efficient strategies for genome engineering. DNA transformation procedures have been developed (Judelson *et al.*, 1993a; Judelson *et al.*, 1993b), but gene knockouts and gene replacements have never been possible because insertion of transgenes occurs exclusively by non-homologous end-joining (NHEJ) (Judelson, 1997; Tyler & Gijzen, 2014). Alternative approaches for functional analysis have included TILLING (Lamour *et al.*, 2006) and gene silencing (Judelson *et al.*, 1993b; Whisson *et al.*, 2005; Ah-Fong *et al.*, 2008; Wang *et al.*, 2011). However, TILLING, which is based on random mutagenesis, is very laborious and requires long term storage of large pools of mutants, and has not proven very useful in oomycetes. Gene silencing (RNAi), triggered using hairpin, antisense, and sense RNA constructs (Ah-Fong *et al.*, 2008) or using dsRNA directly (Whisson *et al.*, 2005; Wang *et al.*, 2011) has proven useful. However, knockdown of genes in oomycetes by RNAi is incomplete, and varies among gene targets, experiments and laboratories. Also, selective silencing of closely related genes is difficult.

Recent advances in engineered nucleases that specifically cleave genomic sequences in living cells have provided valuable tools to create targeted mutations in numerous organisms, from vertebrates, insects, and plants (reviewed in Gaj *et al.*, 2013) to microbes including parasites (Shen *et al.*, 2014; Wagner *et al.*, 2014; Peng *et al.*, 2015; Zhang *et al.*, 2014) and fungi

(Jacobs *et al.*, 2014; Vyas *et al.*, 2015; Liu *et al.*, 2015). These nucleases, that include zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), and CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR associated), can generate a double-stranded break (DSB) at specific sites. By triggering repair of the DSB, either by error-prone NHEJ or homology directed repair (HDR), such methods can increase the rate of gene editing to levels that enable ready isolation of cells or organisms bearing a desired genetic change (Miller *et al.*, 2011). ZFNs and TALENs are engineered proteins containing a modular DNA recognition domain and a DNA cleavage domain.

Like ZFNs and TALENs, the type II CRISPR/Cas9 system derived from the adaptive immune system of *Streptococcus pyogenes* also has DNA recognition and cleavage functions (Cong *et al.*, 2013; Mali *et al.*, 2013). However, DNA recognition is mediated by a single guide RNA (sgRNA) rather than a fused DNA recognition protein domain. The specificity of this system relies on the sgRNA which can direct the nuclease Cas9 to the target DNA sequence (Cong *et al.*, 2013; Mali *et al.*, 2013).

Here we have implemented the CRISPR/Cas9 system in *P. sojae*, using the RXLR effector gene *Avr4/6* (Dou *et al*., 2010) as a target. RXLR effectors are a large superfamily of virulence proteins secreted by many oomycetes that have the ability to enter host cells in order to promote host susceptibility (Jiang & Tyler, 2012). The presence of some RXLR effectors, such as Avr4/6, can be recognized by intracellular receptors encoded by plant resistance genes, triggering vigorous defense responses (Jiang & Tyler, 2012). The presence of Avr4/6 is recognized by soybean R genes *Rps*4 and *Rps*6 (Whisson *et al.*, 1994; Gijzen *et al.*, 1996; Dou *et al.*, 2010); recognition by *Rps*4 requires the N-terminus of Avr4/6, while recognition by *Rps*6 requires the C-terminus (Dou *et al*., 2010). Our results demonstrate that CRISPR/Cas9-mediated

gene disruption and gene replacement is an efficient and useful strategy for testing the function of specific genes in *P. sojae* such as *Avr4/6*, which should be useful for all oomycetes.

4.3 RESULTS

4.3.1 Establishment of the CRISPR/Cas9 system for *P. sojae*

To establish a CRISPR/Cas9 system for *P. sojae*, we had to established efficient expression of Cas9, efficient expression of guide RNAs, and targeting of the Cas9 enzyme to the nucleus.

For expression in *P. sojae* we selected the *Streptococcus pyogenes* Cas9 encoded by a gene with human-optimized codons (*hSpCas9*), because this Cas9 version has been used in diverse organisms (Cong *et al.*, 2013; Zhang *et al.*, 2014; Peng *et al.*, 2015), and matches *P. sojae* codon usage relatively well. To test if this protein could be efficiently expressed in *P. sojae*, we fused GFP to the C-terminus of hSpCas9. Furthermore, to direct hSpCas9-GFP into the *P. sojae* nucleus, we used a strong synthetic NLS derived from a *P. sojae* bZIP transcription factor (Fang & Tyler, 2015), which we fused to the N-terminus of SpCas9 (Fig 1A). Preliminary experiments had shown that commonly used mammalian NLS signals did not work efficiently in *P. sojae.* Expression of the *P. sojae* NLS (PsNLS) fused hSpCas9-GFP construct in *P. sojae* transformants resulted in a bright GFP signal strongly localized within the nuclei of *P. sojae* hyphae (Fig. 4.1A). These results indicated that hSpCas9 was strongly expressed in *P. sojae* without further codon optimization, and that the bZIP-derived NLS efficiently targeted the fusion protein to *P. sojae* nuclei.

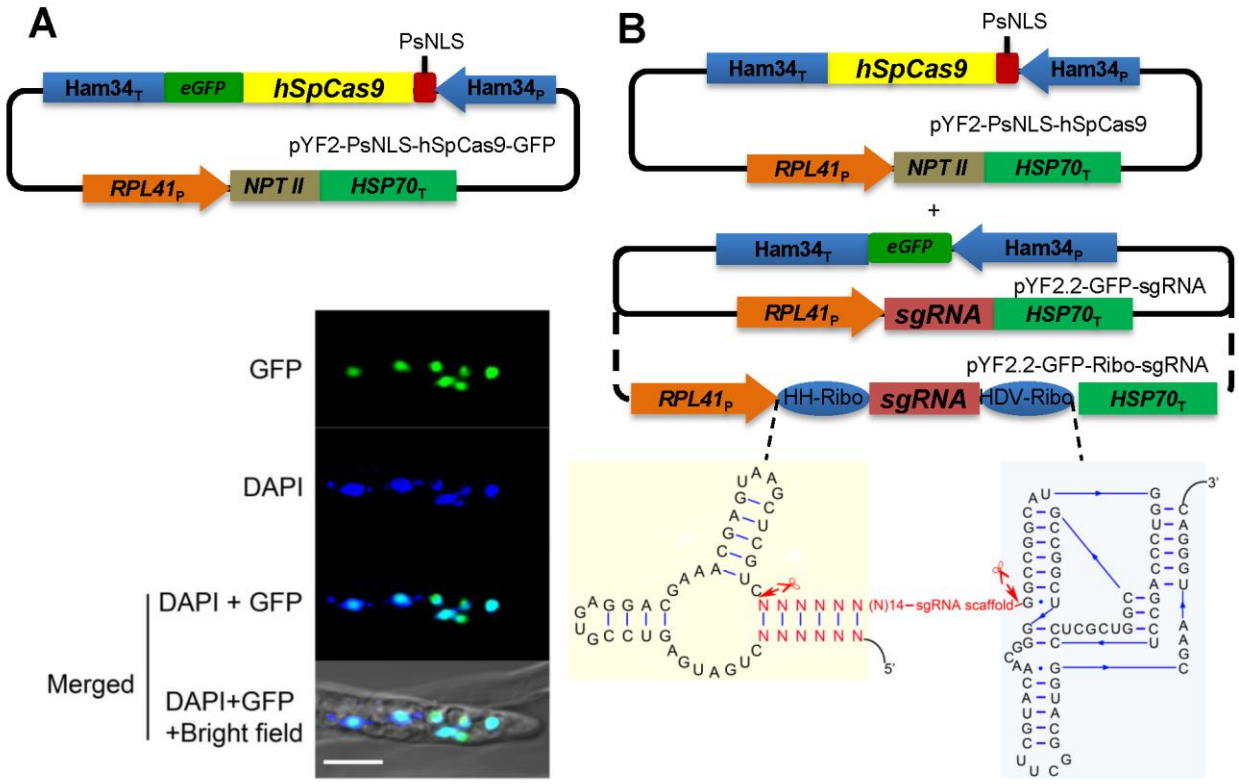


Fig. 4.1 Cas9 and guide RNA constructs for *P. sojae* genome editing.

A. Top: Plasmid for expression of *hSpCas9* fused to e*GFP* and an NLS in *P. sojae*. eGFP, enhanced green fluorescent protein. Bottom: *P. sojae* hyphae expressing PsNLS-hSpCas9-GFP from pYF2, counter-stained with DAPI; scale bars, 10 μm.

B. Top: Plasmids for expression of CRISPR constructs in *P. sojae.* PsNLS, a strong synthetic NLS derived from a *P. sojae* bZIP transcription factor; Cas9 expression is driven by the Ham34 promoter, on a plasmid with the selectable marker *NPTII* driven by the *P. sojae RPL*41 promoter. Transcription of sgRNA (including flanking ribozymes) is driven by the *RPL41* promoter on a plasmid with an eGFP expression cassette (used as a screening marker). Bottom: Double ribozyme construct for release of sgRNAs from the primary RNA polymerase II transcript.

In most systems, sgRNAs are synthesized by RNA polymerase III (RNA pol III), typically using a U6 small nuclear RNA (snRNA) promoter (Cong *et al.*, 2013; Mali *et al.*, 2013; Hwang *et al.*, 2013; Shen *et al.*, 2014; Zhang *et al.*, 2014). However, no RNA Polymerase III

promoters have yet been functionally defined in oomycetes. U6 gene sequences are highly conserved among different oomycetes (Fig. S4.1A), so we cloned the full length *P. sojae* and *P. infestans* U6 gene regions (Fig. S4.1B and C) and inserted a 150 bp fragment of the e*GFP* gene near the 3' end as a PCR reporter sequence (Fig. S4.1D). Surprisingly however, we did not detect any transcripts spanning the GFP reporter fragment by RT-PCR (data not shown).

Recently, the generation of sgRNAs from RNA polymerase II promoters was demonstrated in wheat (Upadhyay *et al.*, 2013), yeast (Gao & Zhao, 2014) and *Arabidopsis (*Gao *et al.*, 2015*).* The yeast and *Arabidopsis* systems used *cis*-acting ribozymes to trim flanking sequences from the sgRNAs, while the wheat system did not. Thus, we employed the constitutive *P. sojae RPL41* promoter (Dou *et al.*, 2008a; Dou *et al.*, 2008b) to direct the transcription of sgRNAs, and evaluated sgRNA constructs that either were or were not flanked on the 5′ side by a hammerhead (HH) ribozyme and on the 3′ side by a HDV ribozyme (Gao & Zhao, 2014)(Fig. 4.1B).

To simplify the generation and screening of *P. sojae* transformants, the hSpCas9 gene and a resistance selection marker (*NPT II*) were placed in one plasmid, while the sgRNA gene together with a GFP marker gene were placed in a second plasmid (Fig. 4.1B, Fig. S4.4).

4.3.2 Cas9-mediated mutagenesis of *Avr4/6*

To test the *P. sojae* sgRNA:Cas9 system, we selected as a target a *P. sojae* gene encoding an RXLR avirulence effector, Avr4/6 (GenBank: GU214064.1). *Avr4/6* is a single copy gene with no close paralogs. Furthermore, loss of *Avr4/6* function was expected to confer a phenotype that would not affect *in vitro* growth, namely the ability to successfully infect soybean cultivars containing resistance genes *Rps*4 or *Rps*6 (Dou *et al.*, 2010). sgRNAs targeting *Avr4/6* were

designed using the web tool, *sgRNA Designer* (Doench *et al.*, 2014). Then sgRNA candidates rated highly by the tool were further filtered by off-target analysis. Finally, two sgRNAs (sgRNA version A and B) were selected in which the respective Cas9 cleavage sites overlapped unique restriction enzymes sites (*Bst* UI and *Tsp* 45I respectively) that could be used to rapidly screen for mutations (Fig. 4.2A).

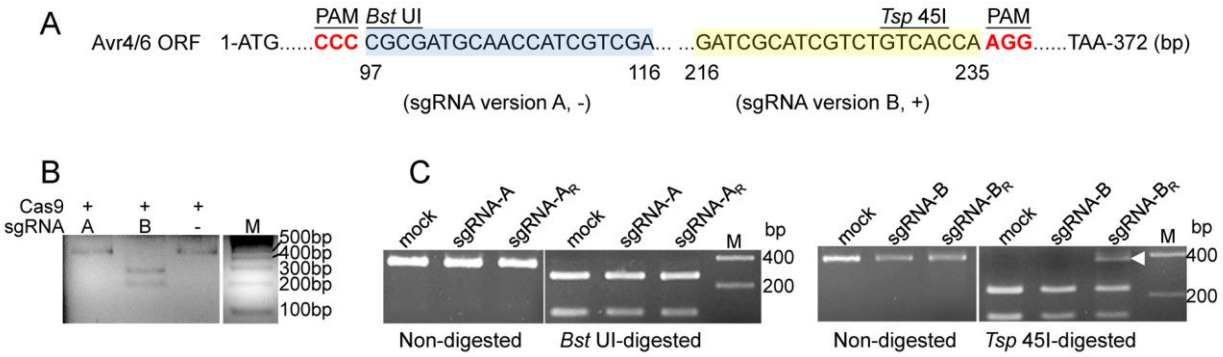


Fig. 4.2 sgRNAs for targeting of *Avr4/6.*

A. Two sgRNA target sites within the *Avr4/6* ORF. Target sites of sgRNA A (sgRNA-A) and B (sgRNA-B) are highlighted in blue and yellow, respectively. sgRNAs A and B target *Avr4/6* on the negative (-) and positive DNA strand (+), respectively. The sgRNA-A site overlaps with a *Bst* UI restriction enzyme site (CGCG) and sgRNA-B with a *Tsp* 45I site ($GT^{G}/_{C}AC$). PAM, Protospacer Adjacent Motif is in bold red.

B. *In vitro* cleavage assay indicating Avr4/6 sgRNA-B can direct Cas9 cleavage of target PCR products but sgRNA-A cannot. DNA template was amplified from pBS-Avr4/6 by using M13F and M13R (Supplemental material). The colors of the original gel are inverted for clarity.

C. PCR and restriction enzyme analysis of *P. sojae* pooled transient expression transformants indicating that only sgRNA-B flanked by ribozymes (sgRNA- B_R) produced amplicons resistant to restriction enzyme cleavage (arrowhead). Approximately 25-30% of the amplicon was resistant to *Tsp* 45I digestion. Experiment was performed in triplicate. sgRNA -A and -B, sgRNA lacking ribozymes; sgRNA-A_R, -B_R, sgRNAs flanked by ribozymes. Mock, *P. sojae* transformants only receiving Cas9 plasmid. In (B) and (C), all gel panels placed together were from the same gel; white dividers indicate lanes not adjacent in those gels.

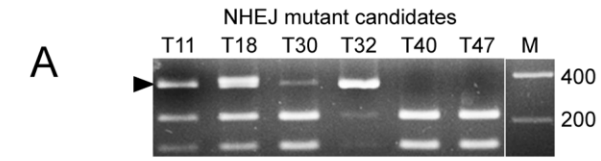
To examine the activity of the designed sgRNAs, Avr4/6-A and Avr4/6-B, we carried out a sgRNA-mediated *in vitro* cleavage assay of target DNA. In these assays, the sgRNAs were synthesized using T7 RNA polymerase, and purchased SpCas9 protein. Cas9/sgRNA-B completely cleaved the target DNA, whereas Cas9/sgRNA-A showed no activity *in vitro* (Fig. 4.2B).

In parallel with the *in vitro* assays, we used transient expression in *P. sojae* protoplasts to test whether the hSpCas9 and sgRNAs produced *in vivo* could modify the endogenous *Avr4/6* gene. The two *Avr4/6*-specific sgRNAs were assembled into the *P. sojae* expression plasmid under the control of the *RPL41* promoter, either flanked with ribozymes (Avr4/6-sgRNA-A_R, -B_R) or without ribozymes (Avr4/6-sgRNA-A, -B). The sgRNA constructs were co-transformed with the hSpCas9 expression plasmid into *P. sojae* strain P6497. Transformants were enriched by G418 selection 12 h after transformation when hyphae had regenerated. After 24 h, DNA was extracted from the culture containing the pooled transformants. *Avr4/6* sequences were amplified from the pool of genomic DNAs and screened for mutants resistant to the relevant restriction enzymes (*Bst* UI for A and *Tsp* 45I for B). We found that the *Avr4/6* amplicons from the two sgRNA-A transformations (constructs with and without ribozymes) were still fully subject to restriction enzyme cleavage, indicating failure of the Cas9-mediated mutagenesis (Fig. 4.2C). In contrast, the transformation utilizing the sgRNA version B flanked by ribozymes yielded restriction enzyme resistant amplicons (Fig. 4.2C). However the transformation utilizing the sgRNA version B without ribozymes did not yield restriction enzyme resistant amplicons (Fig. 4.2C). To validate the enzyme cleavage results, we sequenced the nested PCR products amplified from the enzyme digestion products from the sgRNA-B_R transformants. The sequence chromatograms showed pure sequences proximal to the target site, and mixed sequences distal to

the target site (data not shown), suggesting the presence of mutations at the target site. These observations indicated that in *P. sojae,* RNA polymerase II can be successfully used for generating sgRNA, provided that ribozymes are employed to remove the surrounding sequences from the transcripts. The failure of the sgRNA version A may result from strong self-complementarity that we subsequently discovered in its sequence, which could block its binding to target DNA.

To characterize CRISPR/Cas9-generated *Avr4/6* mutations in detail, the transformation with the ribozyme-containing Avr4/6-sgRNA-B_R construct was repeated. Individual G418-resistant transformants were isolated and screened for the presence of GFP indicating the presence of the sgRNA construct. Of 50 primary transformants screened, 6 exhibited green fluorescence. Of these, 4 yielded *Avr4/6* amplicons that were partially or fully resistant to *Tsp* 45I digestion (Fig. 4.3A), indicating the presence of *Avr4/6* mutations. Since *P. sojae* protoplasts and hyphae are multinucleate, and hence might be expected to harbor nuclei with a diversity of *Avr4/6* mutations, we isolated zoospores (which are mononucleate) from three of the primary transformants (T11, T18 and T32; the fourth, T30, did not produce zoospores). Ten single zoospore lines were isolated from each transformant, and the *Avr4/6* amplicons were screened by restriction enzyme digestion and by Sanger sequencing. T11, T18 and T32 yielded 7, 10 and 10 pure mutant lines respectively (summarized in Table 1). All 27 pure lines showed a homogeneous sequence profile (Fig. S4.2), indicating that all of them were already homozygous, and carried the same mutation in both alleles. All 10 of the T32 lines were homozygous for the same mutation. The lines derived from T18 included 9 lines homozygous for one *Avr4/6* mutation and one homozygous for a different mutation. The lines derived from T11 included two homozygous for one mutation (mut1) and five homozygous for a second mutation (mut2). The

three remaining lines were heterozygous and biallelic, containing a third mutation paired with mut1. (Fig. 4.3B and Fig. S4.2). No lines retained any wild type alleles, either homozygous or heterozygous.



B

Tsp 45I PAM

```
Avr4/6   GCTGCGATCGCATCGTCTGTCA CCAAGGAAGAAGAGGAACGC
           A  A  I  A  S  S  V   T  K  E  E  E  E  R
T11-1    GCTGCGATCGCATCG------A CCAAGGAAGAAGAGGAACGC
           A  A  I  A  S         T  K  E  E  E  E  R
T11-2-5  GCTGCGATCGCATCGTCTGTC- CCAAGGAAGAAGAGGAACGC
           A  A  I  A  S  S  V    P  R  K  K  R  N
T18-1    GCTGCGATCGCATCGTCTGT-- -aAAGGaagaagaggaacgc
           A  A  I  A  S  S  V      K  E  E  E  E  R
T18-2    GCTGCGATCGCATCGTCTGT-- -CAAGGAAGAAGAGGAACGC
           A  A  I  A  S  S  V      K  E  E  E  E  R
T32-3    GCTGCGATCGCATCGTCTGTCAaCCAAGGAAGAAGAGGAACGC
           A  A  I  A  S  S  V  N  Q  G  R  R  G  T
```

	Williams			Harosoy		
Mut type	*rps*	*Rps4*	*Rps6*	*rps*	*Rps4*	*Rps6*
Wild type	V	A	A	V	A	A
In-frame deletion	V	I/V	I/A	V	I/V	I/A
Deletion, frame shifted	V	A	I/A	V	A	I/A
In-frame del and C>A (rare codon mutation)	V	I	A	V	A	A
In-frame deletion	V	I/A	A	V	A	I/A
Insertion, frame shifted	V	I/V	I/A	V	I	I

Fig. 4.3 Characterization of individual NHEJ-mediated mutants.

A. *Tsp* 45I screening of *Avr4/6* amplicons from six individual transformants carrying *hSpCas9* and sgRNA-B_R plasmids. Arrowhead indicates *Tsp* 45I-resistant amplicons. Transformants are expected to carry a mixture of modified and non-modified *Avr4/6* genes.

B. Left: Sequences of *Avr4/6* mutant amplicons from single zoospore lines derived from transformants T11, T18 and T32 (T30 produced no zoospores). Target sites are highlighted in yellow and the PAM sequences are in bold red. (Right) Summary of virulence assays of *Avr4/6* mutants on Rps4- and Rps6-containing cultivars. V, virulent; A, avirulent; I, intermediate; I/A intermediate to avirulent; I/V intermediate to virulent.

Each of the mutations consisted of a short indel, located specifically at the Cas9 cleavage site, i.e. between the 17th and 18th nucleotide of the sgRNA target site. Deletions of one, three

and six bp were observed, one contained a one bp insertion, and one combined a three bp deletion with a one bp replacement (Fig. 4.3B and Table 1).

We also tested the stability of the mutants by sub-culturing each of the single zoospore lines for at least three generations on media without G418 selection. All of the mutated sites examined remained the same as the first generation, based on sequencing of the *Avr4/6* amplicons. Interestingly, one transformant T47 which did not show obvious mutations in the first generation acquired the same single adenine insertion as T32-3 after sub-culturing of the unpurified transformant for one generation (Fig. S4.3), presumably because the sgRNA:Cas9 constructs were integrated into the genome and continued to actively cleave the target in each generation. Collectively, these results indicate that our CRISPR/Cas9 system can efficiently and specifically trigger the introduction of NHEJ mutations, typically short indels, into the *P. sojae* genome.

Table 4.1 *Avr4/6* CRISPR/Cas-induced NHEJ mutations

	Mutant patterns					
	Homozygote			Heterozygote		Biallele
Mutants	WT/WT[a]	mut1/mut1	mut2/mut2	WT/mut1	WT/mut2	mut1/mut2,3
T11[b]	0	2	5	0	0	3*
T18[c]	0	9	1	0	0	0
T32[d]	0	10	0	0	0	0

[a] WT, wild-type sequence

[b] T11, mut1, 6-bp deletion; mut2, 1-bp insertion; *heterozygous mut1/mut3; mut3 is a three bp deletion. The sequencing profiles of the heterozygotes were disambiguated using the web-tool TIDE (Brinkman *et al.*, 2014)

[c] T18, mut1, 3-bp deletion; mut2, 3-bp deletion plus 1-bp substitution

[d] T32, mut1, one bp insertion.

4.3.3 Homologous gene replacement stimulated by the CRISPR/Cas9 system

Donor DNA-mediated repair of sgRNA-guided Cas9 cuts has proven an efficient way to facilitate gene replacements via homology-directed repair (HDR) (Cong *et al.*, 2013; Mali *et al.*, 2013). To determine if sgRNA:Cas9 mediated DSB could stimulate homologous recombination in *P. sojae*, we co-transformed the CRISPR constructs that were successfully used for mutation of *Avr4/6*, along with uncut donor DNA plasmids that contained the entire *NPT II* ORF flanked by different lengths of the sequences surrounding the *Avr4/6* gene. An equimolar ratio of the three plasmids was used (Fig. 4.4A). Since preliminary experiments had shown that expression of the *NPT II* gene from the *Avr4/6* promoter was insufficient for G418 selection, the *NPT II* gene was included in the Cas9 plasmid for selection of transformants. We used homology arms consisting of three different lengths of 5′ and 3′ flanking sequences, namely, 250 bp, 500 bp, and 1 kb, to assess which would enable the highest recombination efficiency (Fig. 4.4B). The *NPT II* gene in the Cas9 expression plasmid served as a negative control, because it lacked any *Avr4/6* flanking sequences.

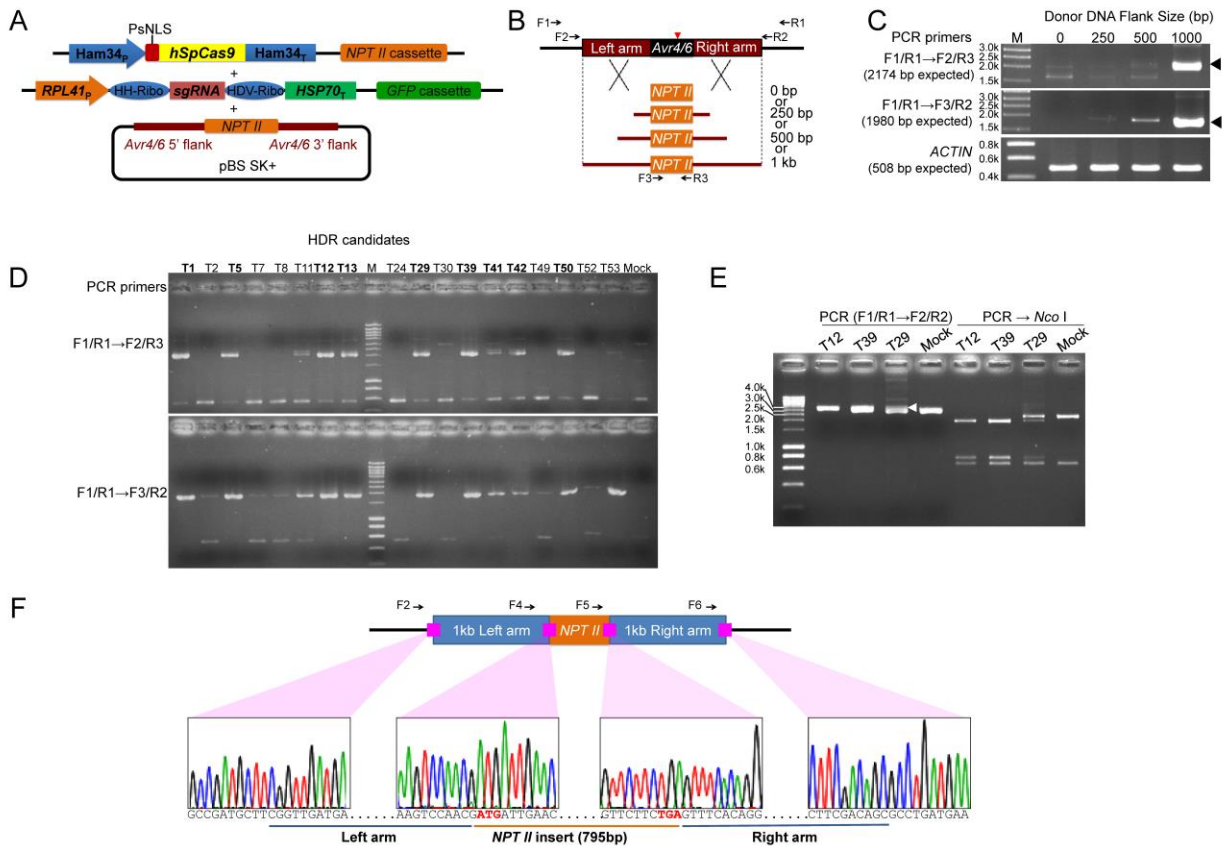


Fig. 4.4 HDR-mediated replacement of the *Avr4/6* ORF with an *NPT II* ORF.

A. Strategy used for gene replacement. Plasmids containing a homologous donor DNA (*NPT II* with *Avr4/6* flanking sequences) were co-transformed with the Avr4/6 sgRNA-B_R and hSpCas9 constructs.

B. Three different sizes of homology arms, 250 bp, 500 bp and 1 kb, flanking the *Avr4/6* locus were used. The *NPT II* gene in the hSpCas9 expression plasmid served as the control (0 bp homologous arm, Mock). Primers used to screen the HDR mutants and validate the replaced region are shown as arrows. Primer pairs F1/R1, and nested primer pairs F2/R3, F3/R2 and F2/R2 were used for HDR mutant screening. Red arrowhead indicates the CRISPR/Cas9 cleavage site (between 232 bp and 233 bp of *Avr4/6* ORF).

C. Analysis of genomic DNA from pooled transformants produced using the four sizes of flanking sequences, using nested PCR. Arrowheads indicate sizes expected if HDR has occurred. *ACTIN* = actin control for DNA quality.

D. Screening of individual HDR transformants generated with the 1 kb flanking sequence plasmid. The 9 positive HDR mutants are highlighted in bold.

E. PCR analysis of representative zoospore-purified lines of HDR mutants, demonstrating that T12 and T39 are homozygotes while T29 is a heterozygote. DNA sizes (bp) before restriction enzyme cleavage: WT, 2936; HDR mutants, 3327; after *Nco* I digestion: WT, 2273 + 663; HDR mutant, 1945 + 751 +663. The arrowhead indicates the fragment amplified from the *NPT II*-replaced allele. Primer F4= Avr4/6_up500bp_PhusF (Table S4.1).

F. Sanger sequencing traces of junction regions confirming that the *Avr4/6* ORF was cleanly replaced by the *NPT II* ORF in a representative zoospore purified clone (HDR-T12-1). Start and stop codons are in red bold.

Following co-transformation and G418 enrichment, the bulk transformants were subjected to genomic DNA extraction and PCR analysis. PCR amplifications using primers located outside the *Avr4/6* homology arms and within the *NPT II* gene were used to detect homologous recombination events. The results suggested that HDR had occurred, but the frequency was variable depending on the length of the flanking sequences in the donor DNA plasmids. The transformant population generated with the 1 kb flanking sequences showed the highest frequency of gene replacement. The population generated with the 500 bp flanking sequences showed a much lower frequency compared to the 1 kb population. The population from the 250 bp flanking sequences showed very low recombination frequencies (Fig. 4.4C).

Next, to characterize HDR events in detail, we generated single *P. sojae* transformants derived from the 1 kb arm donor. After screening 68 individual G418 resistant transformants for GFP production, we identified 18 transformants bearing the two CRISPR components. Then, using nested primers specific for HDR events (Fig. 4.4B), we found evidence for gene replacement events in 9 of the transformants (Fig. 4.4D). Sanger sequencing across the junctions of the flanking sequences and *NPT II* in the nested PCR products was also consistent with replacement of the *Avr4/6* ORF with *NPT II* gene (data not shown). Three HDR mutants, namely

HDR -T12, -T29 and -T39, that readily produced zoospores, were selected for functional tests. After zoospore isolation, the *Avr4/6* region of each single zoospore line was examined by PCR amplification using primers flanking the two homologous arms and cleavage of the amplicon by the restriction enzyme *Nco* I. We found that all of the 11 single zoospore lines obtained from HDR-T12, and all 8 obtained from HDR-T39 were homozygotes (Fig. 4.4E); this was further validated by Sanger sequencing (Fig. 4.4F). In contrast, all 20 of the single zoospore lines of T29 appeared to be heterozygotes that contained a HDR event in just one of the two *Avr4/6* alleles. This was further verified by PCR amplification using primers outside of homology arms and in *NPT II* gene. More detailed analysis of three of the HDR-T29 lines revealed that the non-HDR alleles possessed the same mutation, an adenine deletion, in every case (Fig. S4.3), presumably caused by NHEJ.

4.3.4 Modified recognition of *Avr4/6* mutants by soybeans carrying the *Rps*4 and *Rps*6 loci

In order to test the effects of the CRISPR/Cas9-induced *Avr4/6* mutations on *P. sojae* recognition by plants containing the *Rps*4 and *Rps*6 loci, the five homozygous NHEJ mutants and two homozygous HDR mutants were inoculated onto hypocotyls of soybean isolines containing *Rps*4 (L85-2352) or *Rps*6 (L89-1581) in a Williams background, as well as isolines containing *Rps*4 (HARO4272) or *Rps*6 (HARO6272) in a Harosoy background. 4 days after inoculation (dpi), the specific virulence of the different mutants was scored and analyzed by Fisher's exact test (Table 2). We observed that the frameshifted mutants T32-3 and T11-2-5 both showed increased killing of *Rps*4- or *Rps*6-containing soybean seedlings in both Williams and Harosoy backgrounds (Table2). The increased killing of both *Rps*4 and *Rps*6 plants by T32-3 was statistically significant ($p < 0.01$), while the increased killing by T11-2-5 of *Rps*4 but not

*Rps*6 plants was significant ($p<0.05$). On the other hand, the increased killing in every case was still significantly ($p < 0.05$) less than the killing of *rps* plants lacking *Rps*4 or *Rps*6 (Table 2). Thus T32-3 was scored as intermediate on *Rps*4 and *Rps*6 plants, while T11-2-5 was scored as avirulent and intermediate respectively. The other NHEJ mutants, containing in-frame deletions, also showed increased killing of *Rps*4- and *Rps*6-containing plants, but significantly ($p < 0.03$) less killing than observed with *rps* plants. The *Avr4/6* mutant having a two amino acid deletion (T11-1) showed an intermediate phenotype that was close to fully virulent on *Rps*4 plants while the two mutants a single amino acid deletion (18-1 and 18-2) showed intermediate to avirulent phenotypes (Fig. 4.5; Table 2).

Table 4.2 Characterization of the virulence of *P. sojae Avr4/6* NHEJ- and HDR-mutants on soybeans

Williams	**Ws (*rps*)**		**L85-2352 (*Rps*4)**				**L89-1581 (*Rps*6)**			
Strains	sv[a]	vir[b]	sv	p-value1[c]	p-value2[d]	vir	sv	p-value1	p-value2	vir
WT	0/60	V	39/58	-	-	A	30/46	-	-	A
T11-1	0/40	V	7/47	0.014	<0.0001	I/V	12/35	<0.0001	0.0074	I/A
T11-2-5 fs[e]	0/52	V	41/70	<0.0001	0.36	A	18/53	<0.0001	0.034	I/A
T18-1	1/40	V	18/53	0.0001	0.0006	I	15/32	<0.0001	0.16	A
T18-2	0/39	V	16/45	<0.0001	0.0016	I/A	12/26	<0.0001	0.21	A
T32-3 fs	0/54	V	6/60	0.029	<0.0001	I/V	18/51	<0.0001	0.0044	I/A
HDR-T12-1	0/20	V	9/28	0.0063	0.0027	I	8/32	0.0174	0.0006	I/V
HDR-T39-1	0/19	V	11/31	0.0035	0.0067	I	7/33	0.0390	0.0002	I/V
Harosoy	**(1-7)1 (*rps*)**		**4272 (*Rps*4)**				**6272 (*Rps*6)**			
Strains	sv	vir	sv	p-value1	p-value2	vir	sv	p-value1	p-value2	vir
WT	0/40	V	60/85	-	-	A	53/76	-	-	A
T11-1	0/38	V	5/37	0.025	<0.0001	I/V	15/39	<0.0001	0.0024	I/A
T11-2-5 fs	0/40	V	47/76	<0.0001	0.25	A	34/79	<0.0001	0.0011	I/A
T18-1	0/40	V	19/36	<0.0001	0.094	A	15/28	<0.0001	0.16	A
T18-2	0/39	V	20/35	<0.0001	0.20	A	16/40	<0.0001	0.0027	I/A
T32-3 fs	0/42	V	15/64	0.0004	<0.0001	I	25/64	<0.0002	0.0003	I
HDR-12-1	0/20	V	6/28	0.034	<0.0001	I/V	10/35	0.0090	<0.0001	I/V
HDR-39-1	0/21	V	10/33	0.0043	0.0001	I/V	9/31	0.0073	0.0002	I/V

[a] sv, soybean seedlings surviving after infection/total seedlings

[b] Vir, virulence of WT and mutants, V, virulent; A, avirulent; I, intermediate; I/A intermediate to avirulent; I/V intermediate to virulent

[c] p-value1, the difference between *Rps*4/*Rps*6 and the *rps* control

[d] p-value2, the difference between mutants and WT *P. sojae*

[e] fs = frameshift mutant

The two homozygous HDR mutants (T12-1 and T39-1), both showed significantly ($p < 0.01$) more killing of *Rps*4- or *Rps*6-containing soybean seedlings in both Williams and Harosoy backgrounds (Table 4.2 and Fig. 4.5), but the killing was significantly ($p < 0.05$) less than on *rps* plants. Thus, both mutants were scored as intermediate to virulent on *Rps*4 and *Rps*6.

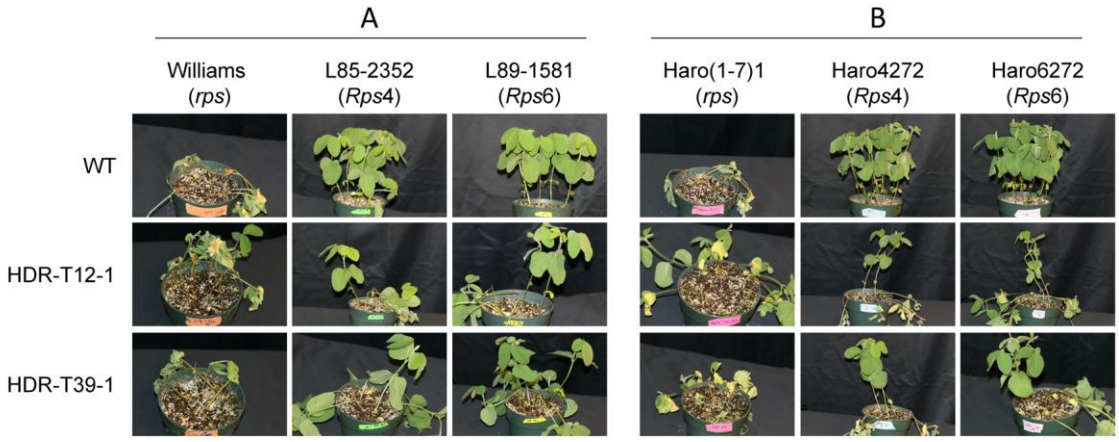


Fig. 4.5 Infection phenotypes of *Avr4/6* mutants.

Representative photos showing soybean seedlings inoculated on the hypocotyls with *P. sojae Avr4/6* HDR mutants. L85-2352 and HARO4272 contain *Rps*4 while L89-1581 and HARO6272 contain *Rps*6.

A. Williams isolines.

B. Harosoy isolines. Photographs were taken 4 days post-inoculation.

4.4 DISCUSSION

Substantial numbers of oomycete genomes have now been sequenced, and even larger numbers are underway (Jiang *et al.*, 2013; Jiang & Tyler, 2012; Kamoun *et al.*, 2015). These genomes contain 15,000 to 25,000 genes, of which approximately half in most species show the rapid sequence divergence expected of infection-related genes (Tyler *et al.*, 2006; Haas *et al.*, 2009; Baxter *et al.*, 2010; Jiang & Tyler, 2012; Jiang *et al.*, 2013). To date however, the tools available for assessing the functions of these genes have been limited to RNAi-mediated gene silencing and to expression of ectopic transgenes.

Here we have adapted the CRIPSR/Cas9 sequence-specific nuclease technology for use in *Phytophthora sojae.* This involved overcoming several technical hurdles. One of these was the fact that commonly used mammalian nuclear localization sequences (NLS) do not function in *P. sojae* and that *P. sojae* nuclear proteins use dispersed NLSs (See Chapter 2). The NLSs from several *P. sojae* nuclear proteins were delineated to the point that a small highly efficient NLS tag could be constructed (See Chapter 2). A second hurdle was that no RNA polymerase III promoters had been characterized in oomycetes, and that the U6 genes we attempted to use did not appear to be well transcribed under our transformation conditions. This we solved by using a strong RNA polymerase II promoter (*RPL41*) in conjunction with self-cleaving ribozymes to release the sgRNA from the RNA polymerase II transcript. A third hurdle was effective expression of the nuclease protein. Out initial attempts to use TALENs were blocked because the TALEN constructs were silenced extremely strongly in *P. sojae*, presumably due to their highly repetitive nature. The humanized SpCas9 protein was readily expressed, but unexpectedly the hSpCas9-GFP fusion protein we used to validate expression was not effective in generating mutations in *P. sojae*. despite the fact that Cas9-GFP fusions have been used in other organisms

such as *Toxoplasma gondii* (Shen *et al.*, 2014). We were successful only after we used the non-fused hSpCas9. A fourth hurdle was identifying effective sgRNAs. One of the two sgRNAs predicted using the *sgRNA Designer* web tool proved ineffective both *in vitro* and *in vivo*. This sgRNA, sgRNA-A, was found to be strongly self-complementary, potentially preventing hybridization with the target DNA. This observation, which is in agreement with Peng *et al*., (2015), underlines that sgRNAs with strong secondary structure predictions should be be eliminated. We also observed a discrepancy between *in vitro* and *in vivo* assays of Cas9/sgRNA activity, namely that the sgRNA lacking ribozymes was functional in the *in vitro* assay, whereas only the sgRNA flanked by ribozymes was effective *in vivo*. There clearly is room for further optimization of sgRNA design.

Using a single sgRNA, we were able to generate small indels at the site of the Cas9 cleavage. While these were useful, future attempts to disrupt genes with large deletions will likely require a pair of sgRNAs, or else the use of homology-directed repair to introduce specific mutations or to replace the gene entirely, as we did in replacing *Avr4/6* with *NPT II*. Of interest is the fact that from two of our transformant lines we were able to recover two or three different mutations. Several of these mutations were segregated into different single zoospore lines, indicating that the mutations had occurred in different nuclei of the regenerating protoplasts. One set of lines were biallelic indicating that two different mutations had occurred in the same nucleus. Also of interest is the observation that most of the mutations were recovered as homozygotes in this diploid organism. We speculate that once a mutation occurred in one allele that made it resistant to further Cas9 cleavage, cleavage of the remaining unmodified allele by Cas9 in most cases led to gene conversion of that allele to match the first allele. During sub-

culture and infection assays of all our Cas9-expressing mutants, we did not observe negative effects on *P. sojae* growth nor on overall virulence.

CRISPR-mediated gene disruptions and gene replacements will find numerous applications in *P. sojae* and other oomycetes. With careful design of sgRNAs, single members of closely related gene families could be eliminated, or else tags for transcriptional measurements could be introduced, to determine their individual contributions. Alternatively, entire gene clusters could be eliminated with a pair of sgRNAs. Gene replacements will enable mutations of all kinds, including promoter mutations and epitope and fluorescent protein tags, to be introduced into the endogenous gene where expression and phenotypes will not be confounded by position effects, over-expression artifacts, and contributions from the unmodified native gene. Targeted gene insertions are also expected to solve a longstanding problem in some oomycetes such as *P. sojae* where ectopic transgenes are invariably poorly expressed, even from the strongest promoter. Gene disruptions will also be useful for creating a much wider choice of selectable markers for transformation, through the creation of auxotrophic mutants. Gene deletions could also be used to remove integrated transgenes so that selectable markers such as *NPT II* can be recycled for repeated transformation experiments.

Some of the advantages of these CRISPR-enabled approaches are illustrated by insights gained from our manipulation of the *Avr4/6* gene. The elimination of the gene by replacement with *NPT II* confirms that *Avr4/6* makes a major contribution to recognition of the pathogen by plants containing the *Rps*4 and *Rps*6 loci, consistent with the findings of Dou *et al.*, (2010). However, the mutants did not kill the *Rps*4- and *Rps*6-containing plants as completely as they killed *rps* plants lacking *Rps*4 and *Rps*6. The *Rps*4 and *Rps*6-containing isolines were produced by introgression, not by transformation with individual *R* genes (Sandhu *et al.*, 2004). Since the

*Rps*4 and *Rps*6 loci, which are allelic, both contain many NB-LRR genes (Sandhu *et al.*, 2004), we speculate that additional NB-LRR genes at these loci (or even the *Rps*4 and *Rps*6 genes themselves) can recognize additional effectors produced by the *P. sojae* strain used in these studies (P6497). Dou *et al.* (2010) also observed that *P. sojae* strains silenced for Avr4/6 did not completely kill *Rps*4 and *Rps*6-containing lines, but those observations were ascribed to incomplete silencing of *Avr4/6*. With the availability of complete *Avr4/6* deletion mutants, we can now more confidently conclude that the *Rps*4 and *Rps*6 loci make additional contributions to resistance other than through recognition of Avr4/6.

The two frameshift mutants killed *Rps*6 plants nearly as well as the HDR mutants (62% combined killing versus 74%) indicating that these mutants were no longer recognized by *Rps*6-containing plants. T32-3 killed *Rps*4 plants better than the two HDR mutants (83% versus 74% killing), but T11-2-5 was clearly avirulent on the *Rps*4 plants (40% versus 74%; WT = 31%). The explanation for this difference may lie in the observation by Dou *et al* (2010) that the N-terminal domain of Avr4/6, up to and including the dEER motif, is sufficient for recognition by *Rps*4 plants, whereas recognition by *Rps*6 plants requires the C-terminus. Since the site of the CRISPR-induced NHEJ mutations is immediately upstream of the dEER motif, the +1 frameshift in T32-3 retained the N-terminal domain but eliminated the dEER motif, presumably abolishing effector entry into the plant (Dou *et al*., 2008b). The -1 frameshift in T11-2-5 also eliminated the dEER motif, but in its place created a highly positively charged sequence (RKKRNARSR). Thus we speculate that this positively charged sequence may act as a surrogate cell entry sequence (Snyder & Dowdy, 2004; Dou *et al.*, 2008b; Kale *et al.*, 2010; Milletti, 2012) delivering the N-terminal fragment for recognition by Rps4.

Of the three in-frame deletions, the two single amino acid deletion mutants scored as intermediate to avirulent on both *Rps*4 and *Rps*6 plants, suggesting that recognition was only slightly impaired. The two amino acid deletion mutant (T11-1) showed a stronger loss of Avr4/6 function, possibly due to disruption of the structure of Avr4/6.

In summary, the adaptation of CRISPR/Cas9-mediated gene targeting to oomycetes is expected to rapidly advance the functional analysis of these extremely destructive and important plant and animal pathogens.

4.5 MATERIAL AND METHODS

4.5.1 *Phytophthora sojae* strains and growth conditions

The reference *P. sojae* isolate P6497 (Race 2) used in this study was routinely grown and maintained in cleared V8 medium at 25 ℃ in the dark. Zoospores were induced and isolated from ~ 1-week old cultures grown on clarified V8 agar, as previously described (Judelson *et al.*, 1993a). Single *P. sojae* transformants were incubated in 12-well plates containing V8 media supplemented with 50 μg/mL G418 (Geneticin, AG Scientific, San Diego, California, USA) for 2~3 days before genomic DNA was isolated by small-scale genomic DNA extraction (gDNA miniprep).

4.5.2 sgRNA design

sgRNA target sites were selected according to the web-tool *sgRNA Designer* (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design; (Doench *et al.*, 2014). Potential off-target sites were checked using the FungiDB (www.fungidb.org) alignment search

tool (BLASTN) against the *P. sojae* genome and visual inspection of the results. Sequences that perfectly matched the final 12 nt of the target sequence and NGG PAM sequence were discarded (Cong *et al.*, 2013). Ribozymes were designed according to Gao, *et al*. (Gao & Zhao, 2014). The first six nucleotides of the Hammerhead (HH) ribozyme were designed to be the reverse complement of the first six nucleotides of the sgRNA target sequences.

4.5.3 Plasmid construction

All the primers used in this study are listed in the Table S4.1 in the supplemental information, and further details on plasmid construction can be found in the supplemental methods in the supplemental materialA map and sequence file for the plasmid backbones used for expressing Cas9 and the sgRNAs targeting the *Avr4/6* locus can be found in Fig. S4.4 in the supporting information.

4.5.4 sgRNA:Cas9 *in vitro* activity assay

To test the activity of the designed sgRNAs, an *in vitro* cleavage assay was carried out (Gao & Zhao, 2014). Briefly, sgRNA was *in vitro* transcribed through run-off reactions with T7 RNA polymerase using the MEGAshortscript^{TM} T7 kit (Ambion, Austin, TX, USA) according to the manufacturer's manual. Templates for sgRNA synthesis were generated by PCR amplification from the sgRNA expression plasmid pYF2.2-GFP-sgRNAs. The target DNA was amplified from pCR2.1-Avr4/6 using primer M13F and M13R (Supplemental material). SpCas9 nuclease was purchased from New England Biolabs Inc., Ipswich, MA, USA, and the cleavage assay was performed according to the product manual.

4.5.5 Improved transformation of *P. sojae*

Polyethylene glycol (PEG) mediated protoplast transformations were conducted using a modification of the previously described methods (Mcleod *et al.*, 2008; Dou *et al.*, 2008a). 2-4 days old *P. sojae* mycelial mats, cultured in nutrient pea broth, were harvested and pre-treated with 0.8 M mannitol for 10 min, then digested in 20 mL enzyme solution [0.4 M mannitol, 20 mM KCl, 20mM MES, pH 5.7, 10 mM $CaCl_2$, 0.5% Lysing Enzymes from *Trichoderma harzianum* ((Sigma L1412: St Louis, MO, USA), and 0.5% CELLULYSIN® Cellulase Calbiochem 219466: San Diego, California, USA)] for ~40 min at room temperature with gentle shaking. The mixture was filtered through a Falcon™ Nylon Mesh Cell Strainer (BD Biosciences: San Diego, CA, USA) and protoplasts were pelleted by centrifugation at 1,200g for 2 min in a Beckman Coulter Benchtop Centrifuge, Miami, FL, USA with swing buckets. After washing with 30 mL W5 solution (5 mM KCl, 125 mM $CaCl_2$, 154 mM NaCl, and 177 mM glucose), protoplasts were resuspended in 10 mL W5 solution and left on ice for 30 min. Protoplasts were collected by centrifugation at 1,200 g for 2 min in the Beckman centrifuge and resuspended at 10^6 /mL in MMg solution (0.4 M mannitol, 15 mM $MgCl_2$ and 4 mM MES, pH 5.7). DNA transformation was conducted in a 50 mL Falcon tube, where 1 mL protoplasts were well mixed with 20-30 µg DNA for single plasmid transformation. For co-transformation experiments, 20-30 µg of the plasmid carrying the *NPT II* selectable marker gene was used, together with an equimolar ratio of any other DNAs included. Then, three successive aliquots of 580 µl each of freshly made polyethylene glycol (PEG) solution (40% PEG 4000 v/v, 0.2 M mannitol and 0.1 M $CaCl_2$) were slowly pipetted into the protoplast suspension and gently mixed. After 20 min incubation on ice, 10 mL pea broth containing 0.5 M mannitol were added,

and the protoplasts were regenerated overnight at 18℃ in the dark. For production of stable transformants, the regenerated protoplasts were collected by centrifugation at 2,000 g for 2 min in the Beckman centrifuge, and then resuspended and evenly divided into three Falcon tubes containing 50 mL liquid pea broth containing 1% agar (42 ℃), 0.5 M mannitol and 50 μg/mL G418 (AG Scientific). The resuspended protoplasts were then poured into empty 60 mm × 15 mm petri dishes. Mycelial colonies could be observed after 2 d incubation at 25℃ in the dark. The visible transformants were transferred to V8 liquid media containing 50 μg/mL G418 and propagated for 2~3d at 25 ℃ prior to analysis. For transient expression, 50 μg/mL G418 was usually added into the regeneration medium after overnight recovery, to enrich the positive transformants. After 1 d incubation at 25 ℃ in the dark, hyphae were collected for genomic DNA extraction.

4.5.6 Detection and quantification of targeted mutagenesis

To detect the results of targeted mutagenesis in transformants, total genomic DNA (gDNA) was extracted from pooled or individual *P. sojae* transformants. For pooled transformants, 48 h after transformation 1 mL of the mycelial culture was pelleted, resuspended in 500 mL lysis buffer (200 mM Tris, pH 8.0, 200 mM NaCl, 25 mM EDTA, pH 8.0, 2% SDS, plus 0.1 mg/mL RNase A added prior to use) and broken by vortexing with 0.5 mm glass beads. For individual transformants, approximately a 7 mm diameter clump of *P. sojae* hyphae were blotted dry on Kimwipe™ paper, then frozen in liquid nitrogen and ground to a powder using a polypropylene pestle, then resuspended in 500 μL lysis buffer. Hyphal lysates were incubated at 37 ℃, 30 min for RNA digestion, then the DNA was recovered by phenol-chloroform extraction and isopropanol precipitation.

All PCR amplifications were conducted using Phusion® high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA) in order to exclude the possibility of mutations causing during PCR amplification. Generally ~10 ng gDNA was used as PCR template. Nested PCR was conducted if necessary, using 1:1000 diluted PCR products as a DNA template for the second round.

To detect NHEJ mutations, the entire 372 bp *Avr4/6* ORF was amplified and examined by digestion with the relevant restriction enzyme. For pooled transformants, a nested PCR was performed to enrich the mutated target before sequencing; this step was not needed for individual transformants, including single zoospore lines. To detect HDR events in pooled and individual transformants (other than zoospore lines), primers located outside the *Avr4/6* homology arms and in *NPT II* gene were used. For screening single zoospore lines, PCR was performed by only using primers outside the homology arms. In both cases, nested PCR was carried out for efficient amplification of the targets. PCR products were sequenced directly by the Sanger dideoxy method in the Oregon State University Center for Genome Research and Biocomputing.

4.5.7 Confocal Microscopy

Laser scanning confocal microscopy (Carl Zeiss LSM 780 NLO) was used to examine the expression and subcellular localization of hSPCas9 fused to the NLS and to GFP. Living hyphae were picked from liquid cultures after 2-3 days growth of transformants. Samples were stained with DAPI (4', 6-diamidino-2-phenylindole) for 20 min in the dark (Talbot, 2001) before microscopy examination. Images were captured using a 63X oil objective with excitation/emission settings (in nm) 405/410-490 for DAPI, and 488/510-535 for GFP.

4.5.8 Infection assays

The ability of *Avr4/6* mutants to infect soybean plants carrying *Rps*4 and *Rps*6 was evaluated by hypocotyl inoculation as previously described (Dou *et al.*, 2010). The wild type and mutant *P.sojae* strains were grown on V8 plates without G418 selection for ~5 days. Soybean cultivars HARO(1-7)1 (*rps*), HARO4272 (Harosoy background, *Rps*4, *Rps*7), HARO6272 (Harosoy background, *Rps*6, *Rps*7), Williams (*rps*), L85-2352 (Williams background, *Rps*4), and L89-1581 (Williams background, *Rps*6) were used. Each pathogenicity test was performed in triplicate, each replicate consisting of at least 19 seedlings. A strain was considered avirulent if the number of inoculated *Rps*4 or *Rps*6 seedlings surviving was significantly higher than among the seedlings lacking the *Rps* gene, as determined by Fisher's exact test (Sokal & Rohlf, 1995) and the number was not significantly different than the number of surviving seedlings inoculated with the unmodified control strain P6497. A strain was considered virulent if the surviving *Rps*4 or *Rps*6 seedlings were not significantly different than the *rps* seedlings, and were significantly fewer than the seedlings surviving P6497 inoculation. A strain was considered intermediate if the surviving *Rps*4 or *Rps*6 seedlings were not significantly different than the *rps* seedlings, and also were not significantly different than the seedlings surviving P6497 inoculation. A strain was also considered intermediate if the surviving *Rps*4 or *Rps*6 seedlings were significantly greater than the *rps* seedlings, and also significantly fewer than the seedlings surviving P6497 inoculation. Intermediate phenotypes were further designated intermediate/virulent or intermediate/avirulent if the p values indicating a difference from virulent or avirulent controls differed by more than 10 fold.

4.6 ACKNOWLEDGEMENTS

We thank F. Arredondo, S. Taylor and D. Wellappili (Oregon State University) for experimental assistance, M.A. Saghai-Maroof (Virginia Tech) for soybean seed, and H. Judelson (UC-Riverside) and members of the Tyler Laboratory for useful advice. We acknowledge the Sequencing and Confocal Microscopy Facilities of the Center for Genome Research and Biocomputing at Oregon State University. This work was supported in part by grant 2011-68004-30104 from the Agriculture and Food Research Initiative of the USDA National Institute for Food and Agriculture.

4.7 REFERENCES

Ah-Fong, A.M., C.A. Bormann-Chung & H.S. Judelson, (2008) Optimization of transgene-mediated silencing in *Phytophthora infestans* and its association with small-interfering RNAs. *Fungal Genet. Biol.* **45**: 1197-1205.

Baxter, L., S. Tripathy, N. Ishaque, N. Boot, A. Cabral, E. Kemen, M. Thines, A. Ah-Fong, R. Anderson, W. Badejoko, P. Bittner-Eddy, J.L. Boore, M.C. Chibucos, M. Coates, P. Dehal, K. Delehaunty, S. Dong, P. Downton, B. Dumas, G. Fabro, C. Fronick, S.I. Fuerstenberg, L. Fulton, E. Gaulin, F. Govers, L. Hughes, S. Humphray, R.H. Jiang, H. Judelson, S. Kamoun, K. Kyung, H. Meijer, P. Minx, P. Morris, J. Nelson, V. Phuntumart, D. Qutob, A. Rehmany, A. Rougon-Cardoso, P. Ryden, T. Torto-Alalibo, D. Studholme, Y. Wang, J. Win, J. Wood, S.W. Clifton, J. Rogers, G. Van den Ackerveken, J.D. Jones, J.M. McDowell, J. Beynon & B.M. Tyler, (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* **330**: 1549-1551.

Brinkman, E.K., Chen, T., Amendola, M. and van Steensel, B. (2014) Easy quantitativeassessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42, e168.

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini & F. Zhang, (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

Doench, J.G., E. Hartenian, D.B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B.L. Ebert, R.J. Xavier & D.E. Root, (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**: 1262-1267.

Dou, D., S.D. Kale, T. Liu, Q. Tang, X. Wang, F.D. Arredondo, S. Basnayake, S. Whisson, A. Drenth & D. Maclean, (2010) Different domains of *Phytophthora sojae* effector Avr4/6 are recognized by soybean resistance genes Rps 4 and Rps 6. *Mol. Plant-Microbe Interact.* **23**: 425-435.

Dou, D., S.D. Kale, X. Wang, Y. Chen, Q. Wang, X. Wang, R.H. Jiang, F.D. Arredondo, R.G. Anderson, P.B. Thakur, J.M. McDowell, Y. Wang & B.M. Tyler, (2008a) Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell* **20**: 1118-1133.

Dou, D., S.D. Kale, X. Wang, R.H. Jiang, N.A. Bruce, F.D. Arredondo, X. Zhang & B.M. Tyler, (2008b) RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. *Plant Cell* **20**: 1930-1947.

Erwin, D.C. & O.K. Ribeiro, (1996) *Phytophthora diseases worldwide*. American Phytopathological Society (APS Press, Minneapolis).

Fang, Y. and Tyler, B.M. (2015) The Oomycete *Phytophthora sojae* uses non-canonical nuclear localization signals to direct proteins into the nucleus. *Fung. Genet. Rep.* **61** (Suppl.), Abstract #165.

Gaj, T., Gersbach, C.A. and Barbas, C.F., 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405.

Gao, Y., Y. Zhang, D. Zhang, X. Dai, M. Estelle & Y. Zhao, (2015) Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 2275-2280.

Gao, Y. & Y. Zhao, (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and *in vivo* for CRISPR-mediated genome editing. *J Integr Plant Biol* **56**: 343-349.

Gijzen, M., H. Förster, M.D. Coffey & B. Tyler, (1996) Cosegregation of Avr4 and Avr6 in *Phytophthora sojae*. *Canadian journal of botany* **74**: 800-802.

Haas, B.J., S. Kamoun, M.C. Zody, R.H. Jiang, R.E. Handsaker, L.M. Cano, M. Grabherr, C.D. Kodira, S. Raffaele & T. Torto-Alalibo, (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* **461**: 393-398.

Hwang, W.Y., Y. Fu, D. Reyon, M.L. Maeder, S.Q. Tsai, J.D. Sander, R.T. Peterson, J.R. Yeh & J.K. Joung, (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**: 227-229.

Jacobs, J.Z., K.M. Ciccaglione, V. Tournier & M. Zaratiegui, (2014) Implementation of the CRISPR-Cas9 system in fission yeast. *Nat Commun* **5**: 5344.

Jiang, R.H., I. de Bruijn, B.J. Haas, R. Belmonte, L. Löbach, J. Christie, G. van den Ackerveken, A. Bottin, V. Bulone & S.M. Díaz-Moreno, (2013) Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen Saprolegnia parasitica. *PLoS Genet.* **9**.

Jiang, R.H. & B.M. Tyler, (2012) Mechanisms and evolution of virulence in oomycetes. *Annu. Rev. Phytopathol.* **50**: 295-318.

Judelson, H.S., (1997) The genetics and biology of *Phytophthora infestans*: modern approaches to a historical challenge. *Fungal Genet. Biol.* **22**: 65-76.

Judelson, H.S., M.D. Coffey, F.R. Arredondo & B.M. Tyler, (1993a) Transformation of the oomycete pathogen *Phytophthora-Megasperma* F-Sp *Glycinea* occurs by DNA integration into single or multiple chromosomes. *Curr. Genet.* **23**: 211-218.

Judelson, H.S., R. Dudler, C.J. Pieterse, S.E. Unkles & R.W. Michelmore, (1993b) Expression and antisense inhibition of transgenes in *Phytophthora infestons* is modulated by choice of promoter and position effects. *Gene* **133**: 63-69.

Kale, S.D., B. Gu, D.G. Capelluto, D. Dou, E. Feldman, A. Rumore, F.D. Arredondo, R. Hanlon, I. Fudal, T. Rouxel, C.B. Lawrence, W. Shan & B.M. Tyler, (2010) External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell* **142**: 284-295.

Kamoun, S., O. Furzer, J.D. Jones, H.S. Judelson, G.S. Ali, R.J. Dalio, S.G. Roy, L. Schena, A. Zambounis, F. Panabieres, D. Cahill, M. Ruocco, A. Figueiredo, X.R. Chen, J. Hulvey, R. Stam, K. Lamour, M. Gijzen, B.M. Tyler, N.J. Grunwald, M.S. Mukhtar, D.F. Tome, M. Tor, G. Van Den Ackerveken, J. McDowell, F. Daayf, W.E. Fry, H. Lindqvist-Kreuze, H.J. Meijer, B. Petre, J. Ristaino, K. Yoshida, P.R. Birch & F. Govers, (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**: 413-434.

Lamour, K.H., L. Finley, O. Hurtado-Gonzales, D. Gobena, M. Tierney & H.J. Meijer, (2006) Targeted gene mutation in *Phytophthora* spp. *Mol. Plant-Microbe Interact.* **19**: 1359-1367.

Liu, R., L. Chen, Y. Jiang, Z. Zhou & G. Zou, (2015) Efficient genome editing in filamentous fungus Trichoderma reesei using the CRISPR/Cas9 system. *Cell Discovery* **1**: 15007.

Mali, P., L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville & G.M. Church, (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.

Mcleod, A., B.A. Fry, A.P. Zuluaga, K.L. Myers & W.E. Fry, (2008) Toward improvements of oomycete transformation protocols. *J. Eukaryot. Microbiol.* **55**: 103-109.

Miller, J.C., S. Tan, G. Qiao, K.A. Barlow, J. Wang, D.F. Xia, X. Meng, D.E. Paschon, E. Leung, S.J. Hinkley, G.P. Dulay, K.L. Hua, I. Ankoudinova, G.J. Cost, F.D. Urnov, H.S. Zhang, M.C. Holmes, L. Zhang, P.D. Gregory & E.J. Rebar, (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**: 143-148.

Milletti, F., (2012) Cell-penetrating peptides: classes, origin, and current landscape. *Drug Discov. Today* **17**: 850-860.

Peng, D., S.P. Kurup, P.Y. Yao, T.A. Minning & R.L. Tarleton, (2015) CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. *MBio* **6**: e02097-02014.

Sandhu, D., H. Gao, S. Cianzio & M.K. Bhattacharyya, (2004) Deletion of a disease resistance nucleotide-binding-site leucine-rich-repeat-like sequence is associated with the loss of the *Phytophthora resistance* gene Rps4 in soybean. *Genetics* **168**: 2157-2167.

Shen, B., K.M. Brown, T.D. Lee & L.D. Sibley, (2014) Efficient gene disruption in diverse strains of *Toxoplasma gondii* using CRISPR/CAS9. *MBio* **5**: e01114-01114.

Snyder, E.L. & S.F. Dowdy, (2004) Cell penetrating peptides in drug delivery. *Pharmaceutical research* **21**: 389-393.

Sokal, R.R. & F.J. Rohlf, (1995) Biometry: the principles andpractice of statistics in biological research. *WH. Freeman& Co., San Francisco. SokalBiometry: the principles and practice of statistics in biological research1995*.

Talbot, N.J., (2001) *Molecular and cellular biology of filamentous fungi: a practical approach*. Oxford University Press.

Tyler, B.M., (2001) Genetics and genomics of the oomycete–host interface. *Trends Genet.* **17**: 611-614.

Tyler, B.M., (2007) *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* **8**: 1-8.

Tyler, B.M. & M. Gijzen, (2014) The *Phytophthora sojae* Genome Sequence: Foundation for a Revolution. In: Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens. Springer, pp. 133-157.

Tyler, B.M., S. Tripathy, X. Zhang, P. Dehal, R.H. Jiang, A. Aerts, F.D. Arredondo, L. Baxter, D. Bensasson & J.L. Beynon, (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**: 1261-1266.

Upadhyay, S.K., J. Kumar, A. Alok & R. Tuli, (2013) RNA-guided genome editing for target gene mutations in wheat. *G3 (Bethesda)* **3**: 2233-2238.

Vyas, V.K., M.I. Barrasa & G.R. Fink, (2015) A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. *Science advances* **1**: e1500248.

Wagner, J.C., R.J. Platt, S.J. Goldfless, F. Zhang & J.C. Niles, (2014) Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*. *Nat. Methods* **11**: 915-918.

Wang, Q., C. Han, A.O. Ferreira, X. Yu, W. Ye, S. Tripathy, S.D. Kale, B. Gu, Y. Sheng, Y. Sui, X. Wang, Z. Zhang, B. Cheng, S. Dong, W. Shan, X. Zheng, D. Dou, B.M. Tyler & Y. Wang, (2011) Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. *Plant Cell* **23**: 2064-2086.

Whisson, S., A. Drenth, D. Maclean & J. Irwin, (1994) Evidence for outcrossing in *Phytophthora sojae* and linkage of a DNA marker to two avirulence genes. *Curr. Genet.* **27**: 77-82.

Whisson, S.C., A.O. Avrova, P. Van West & J.T. Jones, (2005) A method for double-stranded RNA-mediated transient gene silencing in *Phytophthora infestans*. *Mol. Plant Pathol.* **6**: 153-163.

Zhang, C., B. Xiao, Y. Jiang, Y. Zhao, Z. Li, H. Gao, Y. Ling, J. Wei, S. Li, M. Lu, X.Z. Su, H. Cui & J. Yuan, (2014) Efficient editing of malaria parasite genome using the CRISPR/Cas9 system. *MBio* **5**: e01414-01414.

4.8 SUPPORTING INFORMATION

4.8.1 Fig. S4.1

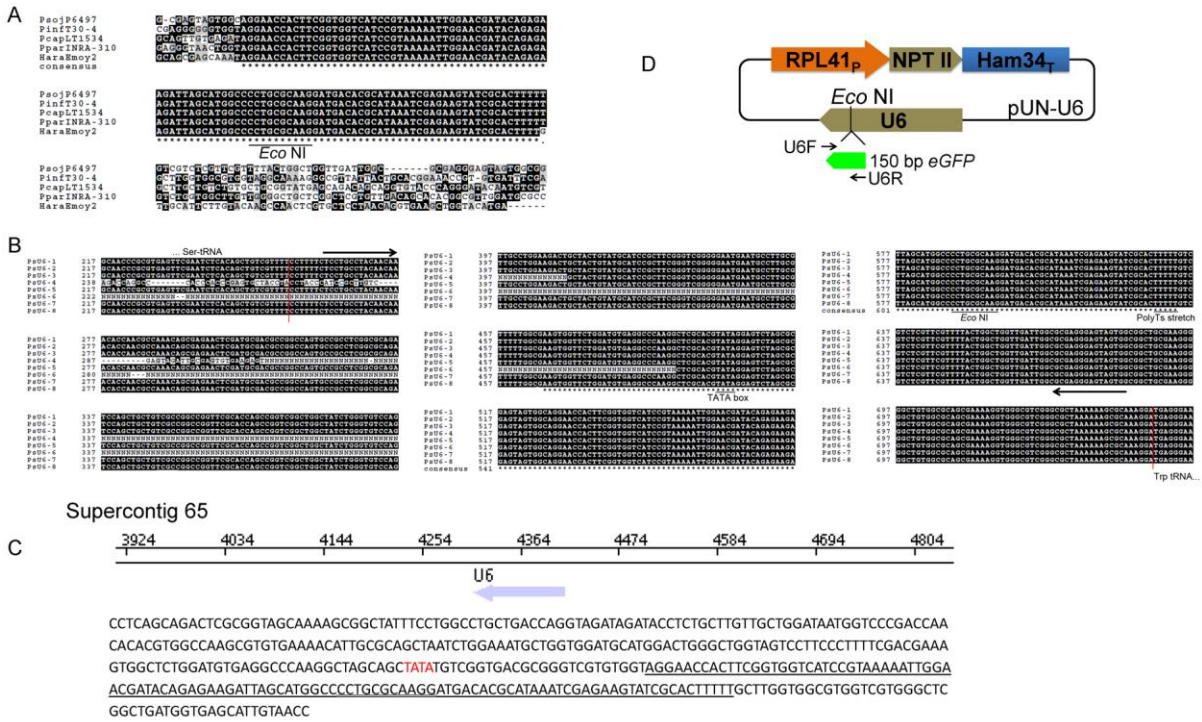


Fig. S4.1 *Phytophthora U6* promoter evaluation.

A. Alignment of selected oomycete *U6* genes, showing that *U6* transcripts are highly conserved. The numbers of *U6* genes were variable in different oomycete species, 8 in *P. sojae P6497* (*PsojP6497*), 127 in *P. infestans* (*PinfT30-4*), 5 in *capsici LT1534* (*PcapLT1534*), 1 *in P.* parasitica (Ppar INRA-310), and 1 in *H. arabidopsidis (HaraEmoy2)* respectively. Genome data are obtained from the fungidb.org website.

B. Alignment of the 8 annotated *P. sojae* U6 genes. PsU6-1 was used to test promoter activity. The red lines indicate the border of the upstream and downstream tRNAs repectively.

C. One of the 127 *P. infestans U6* genes cloned to test U6 promoter activity. (Top) Position of the PiU6 gene on *Phytophthora infestans* T30-4 Supercontig 65. (Bottom) PiU6 sequence used for promoter activity test. The putative U6 coding region is underlined; a putative TATA-box is in red.

D. The plasmid used for testing the functions of the PsU6-1 and PiU6 promoters in *P. sojae*. Residues 1-150 bp of *eGFP* was used as a transcription detection maker. Arrows indicate the primer pair U6GFP_F and U6GFP_R used for cloning the *eGFP* fragment and also for detection of *U6* transcripts by RT-PCR.

The *Eco* NI restriction enzyme site used for inserting the GFP detection marker is underlined in (A) and (B) and double underlined in (C).

4.8.2 Fig. S4.2

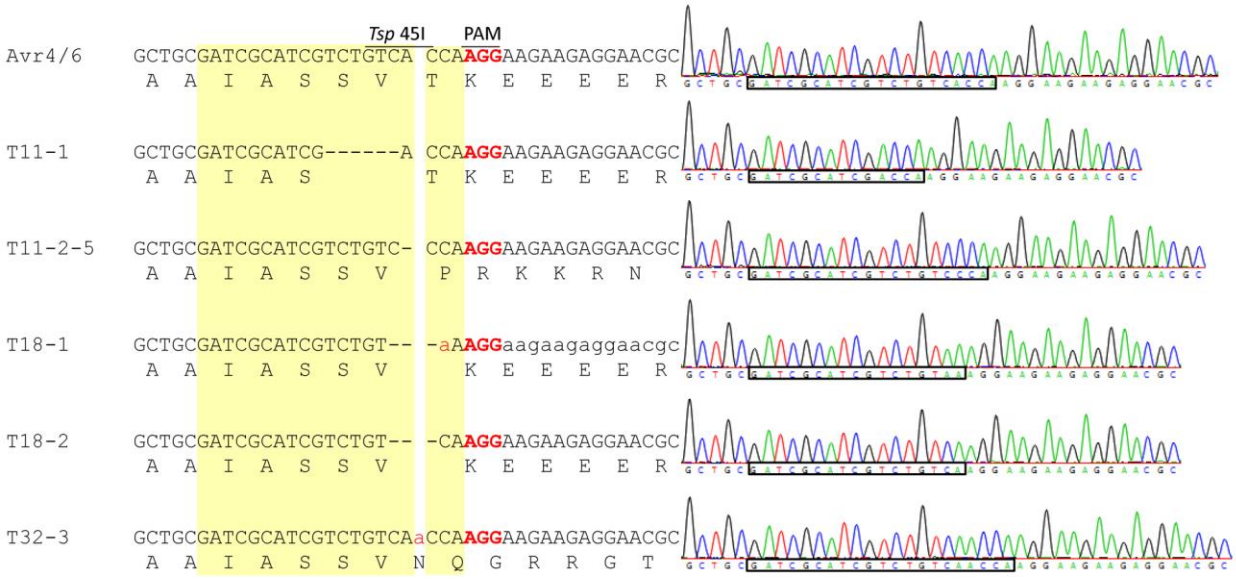


Fig. S4.2 Representative sequencing chromatograms of the *Avr4/6* mutations in the single zoospore-purified mutants.

Regions in box showing the sgRNA target sites within the *Avr4/6* gene. The unambiguous sequencing profiles indicate that these mutant lines are all homozygous.

4.8.3 Fig. S4.3

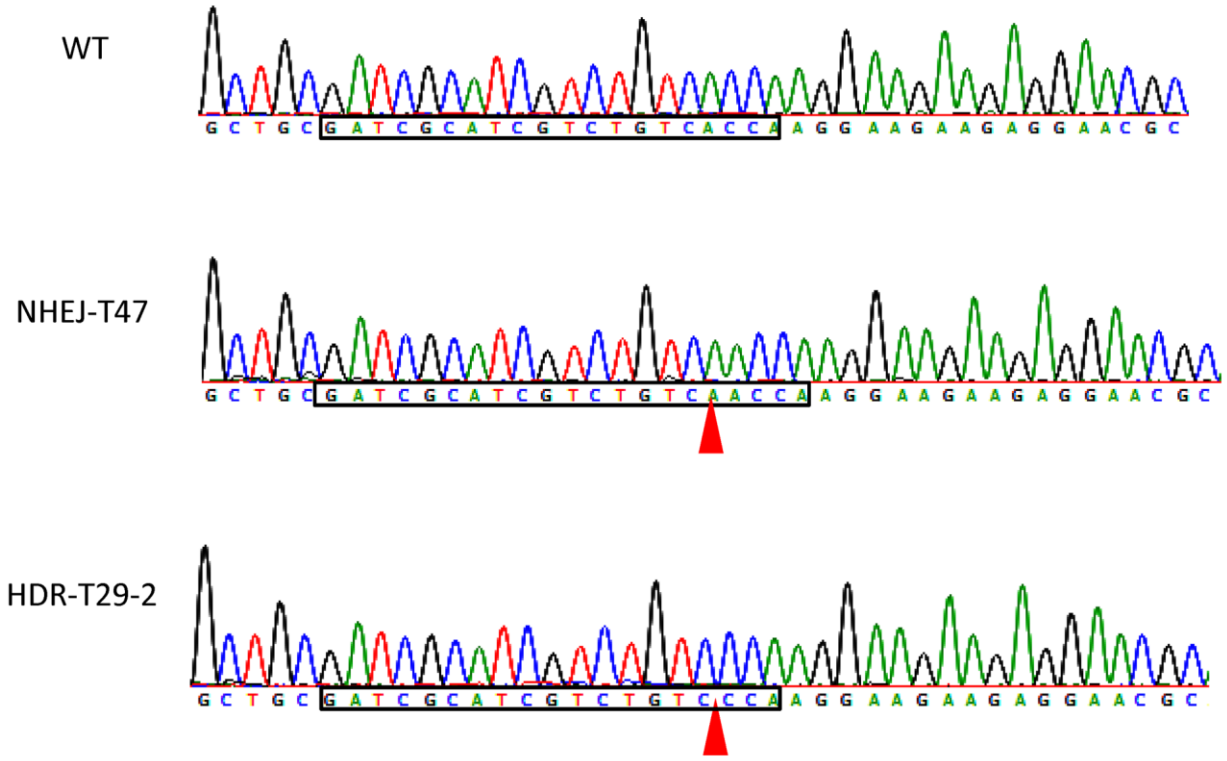
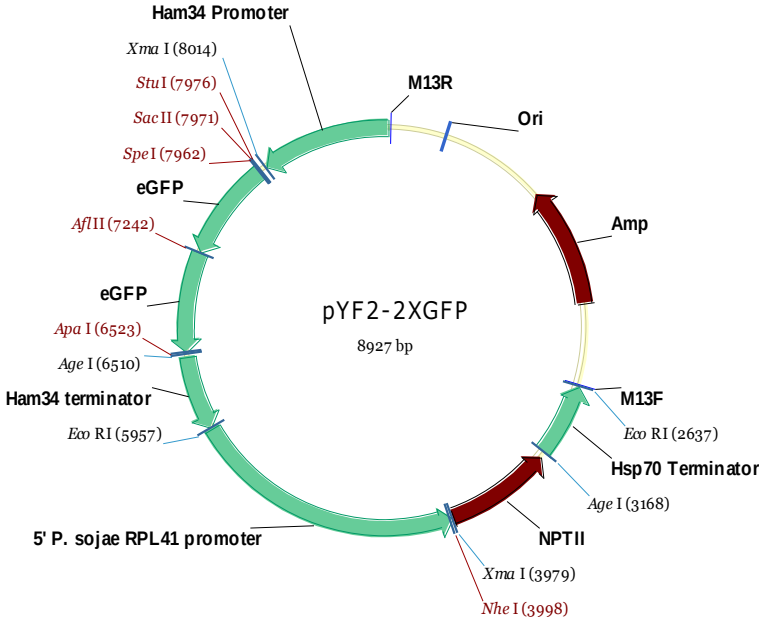


Fig. S4.3 Sanger sequencing profiles revealing that the sub-cultured Cas9:sgRNA transformant T47 (NHEJ-T47) and HDR mutant T29 (HDR-T29-2) had NHEJ mutations (one bp insertion and deletion respectively).

Red triangles indicate the differences between wild type (WT) and mutants.

4.8.4 Fig. S4.4

A



B

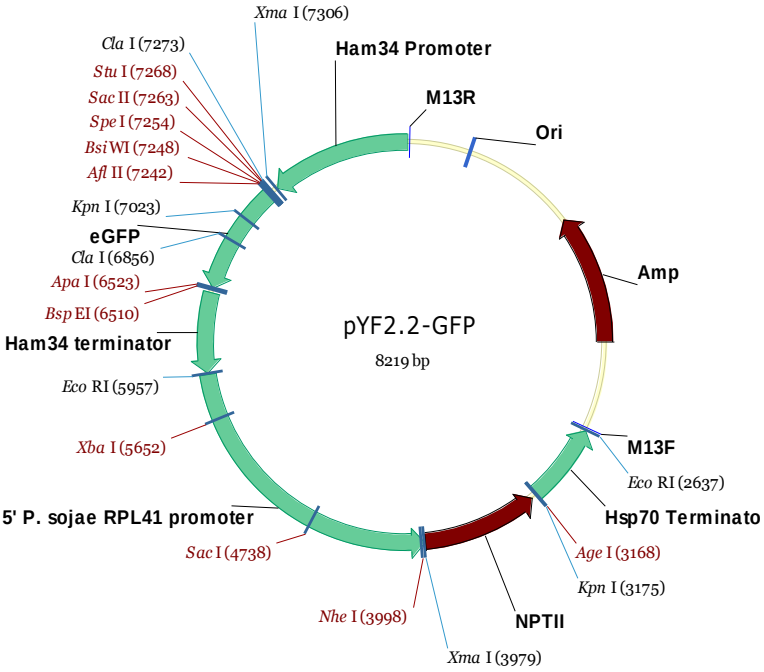


Fig. S4.4 Plasmid backbones used for expression of *hSpCas9* and sgRNA in *P. sojae*.

A. pYF2-2XGFP is used for tracking the subcellular localization and expression of PsNLS fused *hSpCas9*. PsNLS is inserted into *Sac* II and *Spe* I sites. *hSpCas9* is inserted into *Spe* I and *Afl* II sties for subcellular localization examination and *Spe* I and *Apa* I sites for CRISPR expression.

B. pYF2.2-GFP is used for expression of sgRNA (inserted into *Nhe* I and *Age* I sites).

4.8.5 Supplemental methods: generation of *P. sojae* CRISPR/Cas9 plasmids.

To express *hSpCas9* and sgRNAs effectively in *P. sojae*, we first created a new *Phytophthora* expression plasmid backbone pYF2 by combining elements from pHamT34 (Judelson *et al.*, 1991), pUN (Dou *et al.*, 2008) and pGFPN (Ah-Fong & Judelson, 2011) as follows. (i) The *HSP70* terminator was PCR amplified from pGFPN by using primers BlHSP70T_AgeI_F and BlHSP70T_NsiI_EcoRI_R which added an *Eco* RI site, and was then inserted into pUN using *Kpn* I and *Nsi* I sites, placing the *NPT II* gene under the control of the *RPL41* promoter and *HSP70* terminator. (ii) The entire *NPT II* cassette from (i) was then extracted by digestion with *Eco* RI and inserted into the *Eco* RI site of pHAMT34, resulting in pYF1. (iii) A synthetic multiple cloning site fragment containing the restriction enzyme sites *Xma I*-*Cla* I-*Stu* I-*Sac* II-*Spe* I-*Bsi* WI-*Afl* II-*Kpn* I was introduced between the *Xma* I and *Kpn* I sites of pYF1 by oligo annealing, creating pYF2. The plasmid pYF2-GFP was generated by adding an *eGFP* fragment amplified from pGFPN using primers GFP_AflII_F and GFP_ApaI_R into the restriction enzyme sites *Afl* II and *Apa* I of pYF2. The PsNLS, reported in Fang & Tyler (2015), was tested using the plasmid pYF2-PsNLS-2XGFP, which was constructed by two steps. (i) An extra *GFP* was amplified from the same plasmid pGFPN using primers GFP_SpeI_F and GFP_Afl II_R and inserted into the restriction sites *Spe* I and *Afl* II of the plasmid pYF2-GFP, generating pYF2-2XGFP. (ii) The PsNLS was inserted by annealing of two oligonucleotides encoding the NLS (MHKRKREDDTKVRRRMHKRKREDDTKVRRRMHKRKREDDTKVRRR).

To generate *hSpCas9*-expression plasmids pYF2-PsNLS-*hSpCas9*-GFP and pYF2-PsNLS-*hSpCas9*, the coding fragment of *hSpCas9* nuclease from the plasmid pSpCas9 (BB)-2A-GFP (PX458, Addgene plasmid # 48138) were subcloned into the *Spe* I and *Afl* II sites or *Spe* I

and *Apa* I sites (there is an *Apa* I site in *hSpCas9*, so the two digested fragments were cloned sequentially), respectively, of plasmid pYF2-PsNLS-2XGFP containing the *P. sojae* NLS.

To generate the construct for replacing the entire ORF of *Avr4/6*, the *NPT II* coding region together with 250 bp, 500 bp or 1 kb of 5′ and 3′ flanking regions outside the *Avr4/6* coding region were PCR amplified and cloned into the plasmid pBluescript II KS+ by In-Fusion® HD Cloning Kit (Clontech).

4.8.6 Supplemental sequences

DNA template used for CRISPR in vitro cleavage assay

The partial Avr4/6 sequence (in red) was TA-cloned into the plasmid pCR2 (Invitrogen). Targets of sgRNA-A and sgRNA-B are highlighted in blue and yellow respectively. Cleavage sites are indicated by arrows. PCR product (454bp) amplified by M13F/M13R was used as the DNA template for CRISPR *in vitro* cleavage assay .Expected fragment sizes (bp) after cleaved by Cas9, sgRNA-A: 136 and 318; sgRNA-B: 269 and 185.

>pCR2_Avr4/6 partial

M13R

CAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTG
CTGGAATTCGGCTTTTGTTGCCGCTCCAGCTGATGCGATCACAGATGAGTCTCAGCCCCGC↓GATGCAACCATCGTC
GATGCCCCACTCACTGGCAGGGGTGCCAATGCTCGGTATTTACGGACTAGCACATCGATCATCAAGGCCCCCGACG
CCCAGCTACCGAGTACAAAGGCTGCGATCGCATCGTCTGTCA↓CCAAGGAAGAAGAGGAACGCAAGATCTCGACC
GGTCTCAGCAAGCTCAGGCAGAAGCTGAGCAAGCGTTTTCACAAGCCGAATTCTGCAGATATCCATCACACTGGC
GGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTT
AC

M13F

Sequences of sgRNA and ribozyme flanked sgRNAs

Nucleotides highlighted in pink: the 20 nt target sequence

Nucleotides highlighted in yellow: 80 nt sgRNA scaffold

Nucleotides highlighted in green: hammerhead ribozyme (HH ribozyme)

Nucleotides highlighted in cyan: HDV ribozyme

Nucleotides underline: The first six nucleotides of the Hammerhead (HH) ribozyme must be complementary to the first six nucleotides of the target sequence.

↓: ribozyme cleavage sites

>Target A without Ribozyme

TCGACGATGGTTGCATCGCGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtg gcaccgagtcggtgctttt

>Target B without Ribozyme

GATCGCATCGTCTGTCACCAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtgg caccgagtcggtgctttt

>Target A with Ribozyme

cgtcgactgatgagtccgtgaggacgaaacgagtaagctcgtc↓TCGACGATGGTTGCATCGCGgttttagagctag aaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttt↓**ggccggcatggtcccagcctcctc gctggcgccggctgggcaacatgcttcggcatggcgaatgggac**

>Target B with Ribozyme

gcgatcctgatgagtccgtgaggacgaaacgagtaagctcgtc↓GATCGCATCGTCTGTCACCAgttttagagctag aaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttt↓**ggccggcatggtcccagcctcctc gctggcgccggctgggcaacatgcttcggcatggcgaatgggac**

>PsNLS-Cas9

The *P. sojae* NLS is in red. Restriction enzymes used for cloning are underlined. Start and stop codon are in bold.

ATGCACAAGCGCAAGCGCGAGGACGACACCAAGGTCCGTCGTCGCATGCACAAGCGCAAGCGCGAGGACGAC
ACCAAGGTCCGTCGTCGCATGCACAAGCGCAAGCGCGAGGACGACACCAAGGTCCGTCGTCGCACTAGTATGG
CCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGA
CATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTG
CTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAG
CCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGCTATCTGCA
AGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGG
AAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTA
CCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTG
GCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACG
TGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGG
CGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTG
CCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGA
GCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCT
GCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTG
AGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACG
AGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTC
TTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCA
TCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCG
GAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGG
CGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCC
CTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC

ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACT
TCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAAC
GAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGG
CCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAA
ATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACG
ATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGC
TGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTCGACGAC
AAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGC
ATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCAT
GCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGAT
AGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGG
TGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAA
CCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCT
GGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTAC
CTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACC
ATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGG
GGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACG
CCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAA
GGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCC
CGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCT
GGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACG
CCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGG
CGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAG
TACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCG
GCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCG
GAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAG
TCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCG
GCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACT
GAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTC
TGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTG
GAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCC
AAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAA
ACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAG
TGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGA
GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACA
CCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCAC
CGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACCTTAAG**TAA**GGGCCC

Sequence of plasmid harboring the donor DNA

NPT II gene is in gray. Nucleotides in cyan, pink and green are the borders of, 250 bp, 500 bp and 1 kb homologous arms. Restriction enzyme sites for donor DNA are underlined.

>pBS_KS_Avr4/6-1k

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGG
CCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACA
AGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTAC
GTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCC
CCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGG
CGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACA
GGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGC
CAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGT
AAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGAGGTC
GACGGTATCGATAAGCTTGATATCGAATTCCGGTTGATGACCCTGCTCGCGATATTGACACGCTGGCTGCCGACGA
GAGAGAAGCTTAGATAATTTCATGAATACACAAATTCATGAATACACCTACATGTATTGCCACTTGCAAATACTTT
GTCCTCTAAAATAATATGAGTGGACACGAGATAGGCGAGGGCTGTTTTAGCTTTGTGCTTTGGTGGAGGCCGCTTT
TAGGCCCAGAAGCTCGGTTTGACCAGTCTAGAATGGCATTTCTGCGTTTCCAAGCCGCAGAAAAGCGTTTTGGTCG
GGAGCACAGAGCCTCAAGCTTGGATGTTAGACCGTACTGGCGATACGTGTAAATCAGACCTAGGTCTGAATTAAG
AGCTGACTGGATTAAAATCGAGATTATTAACGAGATAATAAGCGTGCGTTAGCCAAAATATCTGCAAACGGCAAT
ACGTCGTTGCATTTGCATTAAACAGTTTTTTTCTGCGCGCTTACCCTGTGTATCTCGTGTACACCTAGGGTATCACG

TAGGAAATCTCAAATTCGGTATACCTACCATATTTAGTTATAATGGGGGTGGTACAAAATGGGAATGAGTACTTTC
CTACCTCGATCATCAATCACTTCATACAAACTTTTCATAAAGACTTCGCCCGTCTCCAAAATGGCTCCTTTTTCGGC
AACCCCATTATAACTAAACCCCCACAACAACTAAATAAATACGGTAAATGTTATAGGTGCCCTGCAGCTTGTCCGA
CATGTACATCGTTCATCATCCGTACGTTTTGGAGTTGGCTCGCCTGTTCTCATCAATGCCGTCGATCGTAGCAATTC
CTTACTCCGTTTTCTGGTGTGTACAGTATCGGGTCGTAAACACATGCGTAAACACATGCGTAATGCGATACGGCGC
GGTGTCTCAGAACCACCCGCCGATTTTGTACTGACAGCGTACCATTGACCTGTAGAGCCCACCCAGATCATAATTA
ATCTTCCAATAAGCAAGCGCCCTGCAATTCCAAGTCCAACGatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggct
attcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttcttttgtcaagaccgacctgtccggtgccctgaatgaact
gcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctgctattgggcgaagtgccgggg
caggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgc
atcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcgcatgcccga
cggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacat
agcgttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttctatcgccttcttgacgagt
tcttctgaGTTTCACAGGCGTTCTGGCTGGAAATCCCGCAATCGTGCCAATACGGGAAAGCACTCGTATTTGTATCCATA
AGCATATTCCTGACTCGATCAGTATCTGAGCCGCTAAAACAACTTCATAGCACCACATTTCTATAGAATTGTAAAC
CCTGAGAGCAAAGAATCGAACCTTCATTTAGCCTTTCTCTTTGAAAAAAGCGCCGGAGTCTGTGTGGCTCGTGCTG
GGACACTGGAACCCACTAGAACAGGCTGATGAGCATAAGGTTCACGCCACTCAGTATGACCGAGATATCCGGTAA
ACCCACTTGCAGCTTTTTCAAAACCCTTTTTAAGGCGCTCGGGGACCCTTTCGGCCTCTGGAGCAGCTTGCCAAGA
CCGCTCGATATCCCTCCCTCTTCGTCGACGCCTAGCGAAGTGTGCATCAGCTGGCTGACGGATGCCTTGGTCGAGT
CGTCACTCCGGAGTTGTCTCGCATCGGGACTAAGTAAGCGAGGCGCAGCTTCAATTCCGGTCGCACCATGGGGTAG
GGCCTTCTTAGAAATCGCCAGAGTAACGTCGCAGCAAAAGGCCAGACACCTTGCTGAATGAACACTGTAGGGTAG
GAATTAAACCAGTTCGCAGGTAACAAACCGCTCACTAGCACAGCAGACAAGTCTAGCATAGGTAGGTATAAGTTA
TAGGACGATTATAGCGGGACTCCTTTCCAAAAAAGCATGATGACCTTCATTTAGCCTTCTGTTCTGGTGCGTCGGTG
AATCTGCTCGCAACCAGCGCTTTGTCCAAGCCCAGATTTATTGTGGCTGGTACTGGAACACTGTGGTCTCGCACGC
CTCAACTATATTATGCTGAAGAGCACATGGTTCAAGTGATCAAGCATGGCCGAAGTATCCGGTAAACGCTCTTTCA
GCTTTTTGAGGAGCTTGACGAGAGCGCTTGGAGATTTTTTCAGCTTCTGGGCGAGTTTGCTGAGGCCGCTTCCGAG
CGCTCTTTCTTCGACAGCGGATCCACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTT
AGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT
CCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATT
GCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
GGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGG
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATG
TGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCC
CCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
GTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCC
TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGG
GCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAC
AGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTT
GCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTC
AGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAA
TTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT
GAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGAT
ACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCA
GCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATT
AATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC
CCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTAT
CACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAG
TACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATA
CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTT
ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCG
TTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA
CTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

4.8.7 Table S4.1

Table S4.1 Oligonucleotides used in this study

Primer name	Sequence	Usage
BlHSP70T_AgeI_F	5' GGCCACCGGTGTCGACTAGAATACTTAGCAATTGGTTAG 3'	PCR BIHSP70 terminator
BlHSP70T_NsiI_EcoRI_R	5' TGCATGCATGCAGGAATTCCCCCAATTCCCCGGATCGTC 3'	
ClaI-AflII instert_F	CGATAGGCCTCCGCGGACTAGTCGTACGC	Oligo annealing, to create pYF2
ClaI-AflII instert_R	TTAAGCGTACGACTAGTCCGCGGAGGCCTAT	
GFP_AflII_F	CGGCTTAAGATGGGCAAGGGCGAGGAAC	PCR *eGFP*, to create pYF2-GFP
GFP_ApaI_R	TAGGGCCCTCACTTGTAGAGTTCATCCATGCCATG	
a548t_t553a_F	5'-TGAGGGCCCGGTTCCGGAGCCTACTTCTTG-3'	Mutate *Age* I in the junction of *eGFP* and the Ham34 terminator to *Bsp* EI, to create pYF2.2-GFP
a548t_t553a_R	5'-CAAGAAGTAGGCTCCGGAACCGGGCCCTCA-3'	
GFP_SpeI_F	5' GGACTAGTATGGGCAAGGGCGAGGAAC 3'	PCR eGFP, to create pYF2.2-2XGFP
GFP_AflII_R	5' GACTTAAGCTTGTAGAGTTCATCCATGCCATGC 3'	
PsU6 _HindIII_F	5' CCCAAGCTTCTCCTGCCTACAACAAACACCAACG 3'	Clone *P. sojae* U6
PsU6_SpeI_R	5' CGGACTAGTCCACTACTCCCTCGCGCCAATC 3'	
PiU6_ HindIII_F	5' CCCAAGCTTCCTCAGCAGACTCGCGGTAGC 3'	Clone *P. infestans* U6
PiU6_SpeI_R	5' CGGACTAGTGGTTACAATGCTCACCATCAGCC 3'	
U6GFP_F	5' GAGCCTGCGCAAGGATGGGCAAGGGCGAGGAAC 3'	Insert 150 bp GFP fragment into U6 gene and detect the transcription of the GFP fragment
U6GFP_R	5' GTGCCTTGCGCAGGGGTAGTGCAGATGAACTTCAGGGTG 3'	
Cas9_SpeI_F	5' AGGACTAGTATGGCCCCAAAGAAGAAGCGG 3'	PCR *hSpCas9*
Cas9_Afl II_ApaI_R	5' TAGGGCCCTTACTTAAGGTCGCCTCCCAGCTGAGACAGG 3'	
Avr4/6_gRNA-A_NheI_F	5' CGGCTAGCTCGACGATGGTTGCATCGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAA 3'	Clone sgRNA-A, forward primer
Avr4/6_sgRNA-B_NheI_F	5' CGGCTAGCGATCGCATCGTCTGTCACCAGTTTTAGAGCTAGAAATAGCAAGTTAAAA 3'	Clone sgRNA-B, forward primer

sgRNA_AgeI_R	5' GCACCGGTAAAAGCACCGACTCGGTGC 3'	Clone sgRNA-A or -B, reverse primer
HH-Avr4/6_sgRNA-A_NheI_F	5' CGGCTAGCCGTCGACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTCGACGATGGTTGCATCGCG 3'	Insert HH ribozyme (the first 6 nt are underlined) to sgRNA-A, forward primer
HH-Avr4/6_sgRNA-B_NheI_F	5' CGGCGATCGCGATCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGATCGCATCGTCTGTCACCAGTTTTAG 3'	Insert HH ribozyme (the first 6 nt are underlined) to sgRNA-B, forward primer
HDV_sgRNA_AgeI_R	5' GCACCGGTGTCCCATTCGCCATGCCGAAGCATGTTGCCCAGCCGGCGCCAGCGAGGAGGCTGGGACCATGCCGGCCAAAAGCACCGACTCGGTGCCAC 3'	Insert HDV ribozyme to sgRNA-A or -B, reverse primer
sgRNA-A-T7P_EcoRI_F	5' CGGAATTCGAAATTAATACGACTCACTATAGGTCGACGATGGTTGCATCGCG 3'	Clone T7 promoter (underlined) fused sgRNA-A
sgRNA-B-T7P_EcoRI_F	5' CGGAATTCGAAATTAATACGACTCACTATAGGGATCGCATCGTCTGTCACCAGTTTTAG 3'	Clone T7 promoter (underlined) fused sgRNA-B
sgRNA-T7T_EcoRI_R	5' CGGAATTCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAAAAAGCACCGACTCGGTGCCAC 3'	Clone T7 terminator (underlined) fused sgRNA
3XPs479N15_Up1	GGATGCACAAGCGCAAGCGCGAGGACGACACCAAGGTCCGTCGTCGCATGCACAAGCGCAAGCGCGA	Oligo annealing to clone *P. sojae* NLS
3XPs479N15_Down1	TCGTCCTCGCGCTTGCGCTTGTGCATGCGACGACGGACCTTGGTGTCGTCCTCGCGCTTGCGCTTGTGCATCCGC	
3XPs479N15_Up2	GGACGACACCAAGGTCCGTCGTCGCATGCACAAGCGCAAGCGCGAGGACGACACCAAGGTCCGTCGTCGCA	
3XPs479N15_Down2	CTAGTGCGACGACGGACCTTGGTGTCGTCCTCGCGCTTGCGCTTGTGCATGCGACGACGGACCTTGGTG	
Avr4/6_up1kb_PhusF	GCTTGATATCGAATTCCGGTTGATGACCCTGCTC	PCR *Avr4/6* 1 kb left arm
Avr4/6_up_PhusR	ATCTTGTTCAATCATCGTTGGACTTGGAATTGCA	
Avr4/6_NPTII_PhusF	ATGATTGAACAAGATGGATTG	PCR *NPTII* ORF
Avr4/6_NPTII_PhusR	GAACGCCTGTGAAACTCAGAAGAACTCGTCAAGAA	
Avr4/6_down1kb_PhusF	GTTTCACAGGCGTTCTGG	PCR *Avr4/6* 1kb right arm
Avr4/6_down1kb_PhusR	TAGAACTAGTGGATCCGCTGTCGAAGAAAGAGCG	
Avr4/6_up250bp_PhusF	GCTTGATATCGAATTCCGTACGTTTTGGAGTTGGC	Clone *Avr4/6* 250 bp left arm
Avr4/6_down250bp_PhusR	TAGAACTAGTGGATCCGCCTGTTCTAGTGGGTTC	Clone *Avr4/6* 250 bp right arm
Avr4/6_up500bp_PhusF	GCTTGATATCGAATTCGTAGGAAATCTCAAATTCGG	Clone *Avr4/6* 500 bp left arm
Avr4/6_down500bp_PhusR	TAGAACTAGTGGATCCGCTGCGCCTCGCTTACTT	Clone *Avr4/6* 500 bp right arm
Avr4/6-F	5' ATGGGCCTCCACAAGGGCTTC 3'	Amplify and sequence *Avr4/6* target
Avr4/6-R	5' TTACGTTAGGTGGTGTAGTCCGACGG 3'	

Avr4/6-nested_F1	5' CTCCTCATTGTTGCCGCTCCAGC 3'	Nested PCR primers for Amplifying and sequencing Avr4/6 target
Avr4/6-nested_R1	5' GAAGCAGCCAGTCCGGGATATCG 3'	
F1	5' GACCTGCTGTAGTTAGTCGTGCTG 3'	Detect and sequence HDR mutant (Fig. 3B)
F2	5' CTGCTGGTTGCTGCTCGTGTG 3'	
F3	5' ATGACTGGGCACAACAGACAAT 3'	
F5	5' GCTGACCGCTTCCTCGTGCTT 3'	
F6	5' CTGTGGTCTCGCACGCCTCAAC 3'	
R1	5' CGAGGCGGAGAAAGACGA 3'	
R2	5' TCACCCTCACTCTTGTCAACTACGC 3'	
R3	5' CGGCGATACCGTAAAGCAC 3'	

Chapter 5

Efficient genome editing in the oomycete *Phytophthora sojae* using CRISPR/Cas9

Yufeng Fang[1, 2], Linkai Cui[2], Biao Gu[2], Felipe Arredondo[2] and Brett M. Tyler[1, 2*]

[1]Interdisciplinary Ph.D. program in Genetics, Bioinformatics & Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA

[2]Center for Genome Research and Biocomputing and Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

* Corresponding author: Brett.Tyler@oregonstate.edu

This Chapter is a protocol submitted to Current Protocols in Microbiology as "Efficient genome editing in the oomycete Phytophthora sojae using CRISPR/Cas9". I contributed 90% of the work described in this chapter. Linkai Cui and Biao Gu contributed to the construction of the "all-in-one" plasmid and its quality control test. Felipe Arredondo provided original protocols related to P. sojae transformation and zoospore isolation. Brett M. Tyler helped to edit the manuscript.

5.1 ABSTRACT

Phytophthora is a filamentous fungus-like microorganism, but belongs to the oomycetes, in the kingdom Stramenopila. *Phytophthora* species are notorious as plant destroyers, causing multibillion-dollar damage to agriculture and natural ecosystems worldwide annually. For a long time, genome editing has been unattainable in oomycetes, because of their extremely low rate of homologous recombination. The recent implementation of the CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated) system in the soybean pathogen *Phytophthora sojae*, an experimental model for oomycetes, has opened up a powerful new research capability for the oomycete community. Here, we describe a detailed protocol for CRISPR/Cas9-mediated genome editing in *P. sojae*, including single guide RNA (sgRNA) design and construction, efficient gene replacement, and mutant-screening strategies. This protocol should be generally applicable for most culturable oomycetes. We also describe an optimized transformation method which is useful for other *Phytophthora* spp. including *P. capsici* and *P. parasitica*.

5.2 INTRODUCTION

Phytophthora sojae is a destructive oomycete pathogen which causes “damping off” of soybean seedlings as well as stem and root rot of established plants (Tyler, 2007). Because of the morphological and physiological similarities between filamentous fungi and oomycetes, oomycetes were historically considered as members of the fungal kingdom. However, molecular analysis revealed that the lineages of fungi and oomycetes diverged before the split of fungi from plants and animals (Förster *et al*., 1990). Oomycetes are relatives of diatoms and brown algae,

classified in the kingdom Stramenopila (Tyler, 2001). Most *Phytophthora* species are plant pathogens that can damage a huge range of agriculturally and ornamentally important plants (Erwin & Ribeiro, 1996). For instance, *P. infestans*, which causes the potato late-blight disease, resulted in the Irish potato famine, and still is a major problem for potato and tomato crops (Judelson & Blanco, 2005). *P. sojae* causes around $1–2 billion in losses per year to soybean crops (Tyler, 2007). Due to its economic impact, *P. sojae*, along with *P. infestans*, has been developed as a model species for the study of oomycete plant pathogens (Tyler, 2007).

With continuing innovations in DNA sequencing technology, genome sequencing has become increasingly affordable. To date, more than 20 oomycete genomes have been sequenced. However functional genomics studies of oomycetes have been hampered by the lack of efficient strategies for genome engineering. DNA transformation procedures were developed for *P. infestans* and *P. sojae* more than 20 years ago (Judelson *et al*., 1993a; Judelson *et al*., 1991), but targeted gene mutations and gene replacements have not been possible because insertion of transgenes occurs exclusively by non-homologous end-joining (NHEJ) (Judelson, 1997; Tyler & Gijzen, 2014). Recently however, we successfully applied the new genome editing technology based on the CRISPR/Cas9 (Clustered regularly interspaced short palindromic repeats and its associated protein) system to *P. sojae* (Fang & Tyler, 2016).

CRISPR/Cas is naturally found in bacteria and archaea as a system of adaptive immunity against phage infection (Deveau *et al*., 2010; Garneau *et al*., 2010; Horvath and Barrangou, 2010). The genome editing technology is derived from the type II CRISPR/Cas system of the bacterium *Streptococcus pyogenes*, using its Cas9 enzyme as the nuclease (Cong *et al*., 2013; Mali *et al*., 2013). Basically, a CRISPR/Cas9 system contains two components, the nuclease Cas9 that can make a double-strand DNA break (DSB), and a 20-nucleotide RNA molecule

called a single guide RNA (sgRNA) that can guide Cas9 to a target DNA sequence via Watson-Crick base pairing (Hsu *et al*., 2014). Due to the repair of the DSB that is triggered, the frequency of targeted gene editing is increased, facilitating the creation of a desired genetic change (Miller *et al*., 2011) (Fig. 5.1). To implement this technique in *P. sojae*, Fang & Tyler (2016) created a *P. sojae*-compatible CRISPR/Cas9 system using an oomycete-specific nuclear localization signal and an oomycete-specific sgRNA transcription cassette (Fig. 5.2A). Using *Avr4/6*, an endogenous gene involved in *P. sojae* infection, as a target, Fang & Tyler (2016) demonstrated that CRIPSR/Cas9 could be used to introduce mutations both by non-homologous end-joining (NHEJ) and by homology-directed repair (HDR).

Here, we describe a generalized and simplified protocol for implementing the CRISPR/Cas9 technology in *P. sojae* to create heritable genome modifications via either the NHEJ or HDR pathways (Fig. 5.1). The overall protocol contains five basic protocols. Basic Protocol 1 addresses the steps of designing an efficient sgRNA. Basic Protocol 2 describes how to construct the sgRNA plasmid and introduces an 'all-in-one' CRISPR/Cas9 plasmid system utilizing a single plasmid encoding both Cas9 and the sgRNA. Basic Protocol 3 outlines generation of the repair template DNA if homologous recombination will be employed to replace a particular gene. Basic Protocol 4 details an optimized *P. sojae* transformation system. Basic Protocol 5 describes screening of *P. sojae* transformants to identify NHEJ- and HDR- mediated mutations.

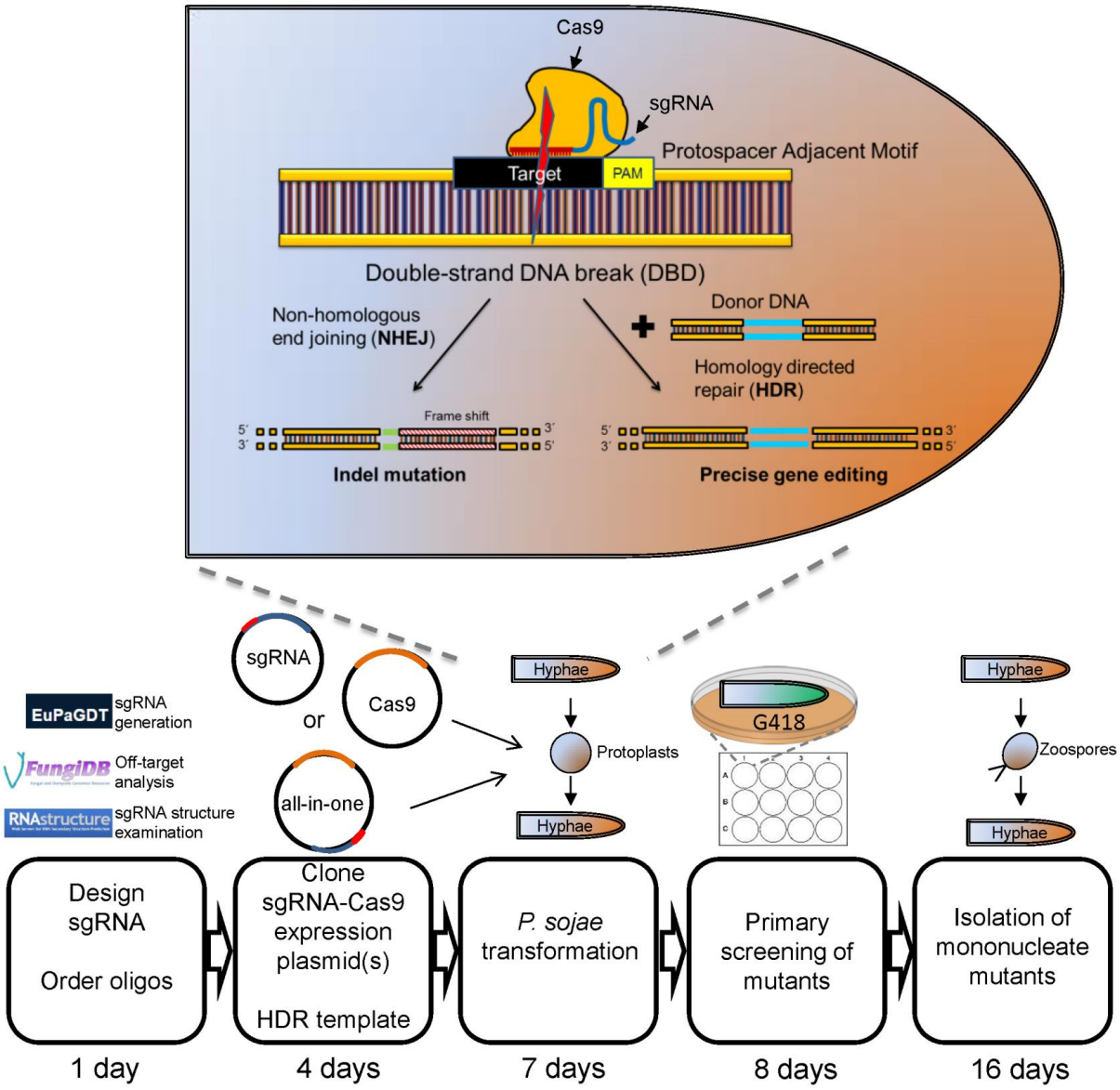


Fig. 5.1 Workflow and timeline of the CRISPR/Cas9-mediated genome editing pipeline in *Phytophthora sojae*.

P. sojae protoplasts are transformed with the Cas9 (gold) and the sgRNA (blue and red) genes located in two separate plasmids or in a single plasmid. The Cas9-sgRNA complex is guided by the sgRNA to the target gene through recognition of the 20-bp complementary DNA sequence as well as the downstream Protospacer Adjacent Motif (PAM) trinucleotides (5'-NGG), where it creates a double strand DNA break (DSB). The DSB induced by Cas9 can be repaired via error-prone NHEJ resulting in an indel mutation

which may cause a frame shift. Alternatively the break can be repaired by the HDR pathway, which enables precise gene editing.

5.3 BASIC PROTOCOL 1: sgRNA design

The specificity of the CRISPR/Cas9 system is mainly determined by the sgRNA through Watson-Crick base pairing with target DNA. Thus, designing a good sgRNA is important. In fact, our experience to date indicates that it is the single most important factor in a successful genome editing experiment. This basic protocol describes the web-tool resources and critical steps used for designing efficient sgRNAs for gene targeting in *P. sojae*. Our guidance here is that you may not be able to find a perfect sgRNA satisfying all parameters, but try to find the best ones. In some cases, it may be desirable to try several sgRNAs.

5.3.1 Materials

sgRNA design software, *EuPaGDT* (http://grna.ctegd.uga.edu/)

Sequence analysis, FungiDB (http://fungidb.org/)

RNA secondary structure prediction, *RNA structure*

(http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html).

5.3.2 Selection of a sgRNA target

In theory, one can manually select a sgRNA by identifying the 20 nt sequence directly upstream of any 5′-NGG sequence (also called protospacer adjacent motif, PAM). However, other factors such as sequence specificity and nucleotide composition may also affect the efficiency of sgRNA targeting. To help identify appropriate sgRNA targets, many useful web-tools have been

developed. We recommend *EuPaGDT* (http://grna.ctegd.uga.edu/) (Peng & Tarleton, 2015), because it includes many oomycete genomes imported from FungiDB (Stajich *et al*., 2011). It also includes very detailed instructions, or refer to Peng & Tarleton (2015). Here are a few additional notes related to design oomycete sgRNA using EuPaGDT:

At the homepage:

(A) Choose SpCas9 for 'RNA guided nuclease selection', because our Cas9 gene was derived from *Streptococcus pyogenes*.

(B) In 'Additional option settings', the server also looks for off-targets using alternative PAMs, such as NAG, NGA for SpCas9. The panel 'HDR repair template parameters' is not designed for oomycetes - see BASIC PROTOCOL 3 for design of HDR templates for oomycetes.

On the results page:

(C) We do not consider the flanking microhomology pair in the oomycete sgRNA design, so you can ignore the related options.

(D) 'Total scores' are calculated by unweighted averaging of the target score, activity-prediction score and microhomology-pair score (Peng & Tarleton, 2015). As the microhomology-mediated end joining (MMEJ) is not included in the oomycete sgRNA design, the total score may not reflect the real suitability for oomycetes. Our experience is that sgRNA candidates whose 'total scores' exceed 0.5 are all equally good, i.e. they do not exhibit any noticeable differences among each other.

(E) 'Efficiency score' was calculated by GC content and positional-specific nucleotide composition, based on a scoring matrix developed by Doench *et al*. (2014).

(F) In *P. sojae*, RNA polymerase II instead of polymerase III is used for transcription of the sgRNA, so the 5'-end of the sgRNA does not need to start with G.

Previously, we used the web-tool sgRNA created by Doench et al. (2014) to obtain sgRNA candidates ranked by scores and the desired positions (Fang & Tyler, 2016). However, in the latest version of sgRNA, one must select the 'target taxon' from a list that only contains human or mouse before proceeding with the sgRNA prediction. Therefore this website is no longer useful for oomycetes.

5.3.3 Off-target analysis

Originally, we manually checked for off-target sites using the alignment tool (BLASTN) within FungiDB (www.fungidb.org) to search the relevant oomycete genome sequence, followed by visual inspection of the results. Ideally, a desirable sgRNA should perfectly match the selected gene and should have no potential off-target sites in the genome (i.e. matches with two or less mismatched positions) (Wyvekens *et al*., 2015). The latest version of *EuPaGDT* facilitates this step. Once the sgRNA candidates are generated, their off-target analyses are done automatically. However, we recommend double-checking your best sgRNA candidates generated by *EuPaGDT* using the manual method described above, because *EuPaGDT* does not show the details of the sgRNA on-/off-target binding, and some off-targets might be misidentified. Furthermore, if mismatches occur near the 5' end of the sgRNA, the sgRNA may still be active.

5.3.4 Examination of sgRNA secondary structure

You should analyze the secondary structure of the candidate sgRNA sequence, as secondary structure can severely reduce the efficiency of a guide RNA. We recommend using *RNA structure* (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html).

The secondary structure of sgRNA influences the activity of CRISPR, so check and try to avoid self-complementarity in the sgRNAs, which potentially prevents hybridization with the target DNA. In practice, you may not always identify a sgRNA without any self-complementarity, so try to select the best one(s). Our experience suggests that no more than 3 consecutive paired bases generally is acceptable.

To increase the gene mutation rate, it may be desirable to design two targets for each gene.

To make mutant screening easier, we recommend if possible to select sgRNAs whose targets overlap with a restriction enzyme cleavage site, so that a restriction enzyme cleavage assay can be used to assess the efficiency of the sgRNAs using a transient expression assay and to identify mutants quickly.

5.4 BASIC PROTOCOL 2: Preparation of CRISPR/Cas9 plasmids

Plasmids containing Cas9 and the sgRNA are necessary for Cas9-mediated genome editing. The human codon-optimized *S. pyogenes* Cas9 (hSpCas9) plasmid was previously constructed (Fang & Tyler, 2016). If you work on other oomycete species, we suggest to check the expression of Cas9 *in situ* using the plasmid, pYF2-hSpCas9-GFP (Fang & Tyler, 2016) that expresses a GFP tagged hSpCas9 gene. In *P. sojae*, RNA polymerase II is used to transcribe the sgRNA from the *RPL41* promoter (derived from *P. sojae* ribosomal protein gene *RPL 41),* because no functional RNA polymerase III promoter has yet been identified in oomycetes (Fang & Tyler, 2016). However the sgRNA must be flanked by 5′ and 3′ ribozymes which remove the extra nucleotides surrounding the sgRNA (Fang & Tyler, 2016) (Fig. 5.2). The ribozymes were designed according to Gao & Zhao (2014). The first six nucleotides of the Hammerhead (HH) ribozyme

were designed to be the reverse complement of the first six nucleotides of the sgRNA target sequences. We have constructed a generalized sgRNA expression plasmid backbone (pYF2.3G-Ribo-sgRNA), which can be used for cloning all kinds of ribozyme-flanked sgRNAs using *Nhe* I and the type IIS restriction enzyme *Bsa* I together with an oligo annealing strategy (Fig. 5.2A and Fig. 5.3A). To facilitate screening of transformants for mutants, we also have constructed an 'all-in-one' CRISPR/Cas9 plasmid backbone expressing both the hSpCas9 gene and the sgRNA (Fig. 5.2B). This basic protocol describes the steps necessary to prepare CRISPR/Cas9 plasmids used for the original two-plasmid transformation system, as well as for a newly developed single-plasmid transformation system.

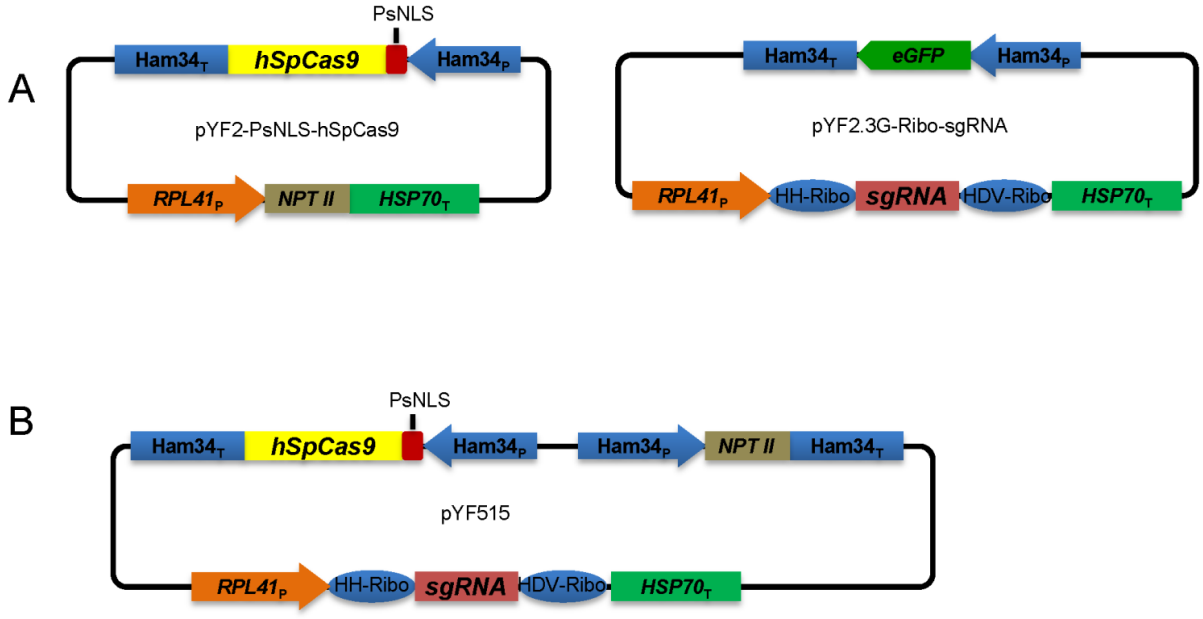


Fig. 5.2 Cas9 and guide RNA constructs for *P. sojae* genome editing.

A. Original Cas9-sgRNA expression plasmids. Cas9 and sgRNA are encoded on two separate plasmids. Cas9 expression is driven by the Ham34 promoter, on a plasmid with the selectable marker *NPT II* driven by the *P. sojae RPL41* promoter. Transcription of single guide RNA (sgRNA) (including flanking ribozymes) is driven by the *RPL41* promoter on a plasmid with an eGFP expression cassette (used as a screening marker).

B. The newer 'all-in-one' plasmid (pYF515) harboring both Cas9 and sgRNA cassettes.

5.4.1 Materials

hSpCas9 expression plasmid, pYF2-PsNLS-hSpCas9 (for two-plasmid transformation system)

Generalized sgRNA cloning plasmid pYF2.3G-Ribo-sgRNA (for two-plasmid transformation system)

pYF2-PsNLS-hSpCas9-GFP (Optional, for examination of Cas9 expression *in situ*)

pYF515 (‘all-in-one’ plasmid, for expression of both Cas9 and sgRNA)

The Phytophthora CRISPR plasmids can be obtained from the Tyler Lab at Oregon State University upon request.

Nhe I (NEB, cat. no. R0131S)

Bsa I and 10 × CutSmart Buffer (NEB, cat. no. R0535S)

T4 DNA Ligase and 10 × buffer (NEB, cat. no. M0202S)

M13 Forward universal primer (5'-GTTTTCCCAGTCACGACG-3')

RPL41_Pseq_F (5'-CAAGCCTCACTTTCTGCTGACTG-3')

TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0)

Green Taq DNA polymerase (Genscript)

LB agar plates containing 100 μg/ml ampicillin (See Current Protocols in Molecular Biology UNIT 1.1; Elbing and Brent, 2002)

LB liquid medium containing 100 μg/ml ampicillin (See Current Protocols in Molecular Biology UNIT 1.1; Elbing and Brent 2002)

SOC liquid medium (See Current Protocols in Molecular Biology, UNIT 1.8; Seidman *et al*., 1997)

Plasmid Maxi Kit (Qiagen)

PCR-grade sterile deionized water

Chemically Competent *E. coli* DH5α cells

PCR purification kit (Qiagen)

Zyppy plasmid Miniprep Kit (Zymo Research)

HiSpeed Plasmid Midi kit (Qiagen)

Maxiprep kit (Qiagen)

Sterile pipet tips for picking colonies from agar plates

37 ℃ incubator-shaker

Nanodrop microspectrophotometer (*http://www.nanodrop.com*)

42 ℃ water bath for heat-shocking cells

15-ml bacterial culture tubes

Access to a Sanger sequencing facility

Access to an oligonucleotide synthesis service

5.4.2 Prepare annealed insert sgRNA oligonucleotides

1. From an oligonucleotide synthesis service, order sense and antisense oligonucleotides that span the full length of the 43 nt HH ribozyme and the 20 nt sgRNA (Fig. 5.3A) as well as providing the 5'-*Nhe* I and 3'-*Bsa* I overhangs.

The guide RNA oligos contain overhangs for ligation into the Nhe I and Bsa I sites in pYF2.3G-Ribo-sgRNA or pYF515. When designing the ribozyme flanked sgRNA sequence, you will need to replace the "20 N" of the sgRNA template with your designed sgRNA sequence, and fill in the first six nucleotides of the HH-ribozyme (highlighted in cyan in Fig. 5.3A) with the reverse

complement of the first six nucleotides of the sgRNA target sequence (highlighted in cyan in Fig. 5.3A).

2. Suspend the oligonucleotides to 100 μM in TE.

3. Prepare the following mixture for phosphorylating and annealing the sgRNA oligonucleotides:

3 μl Sense oligo

3 μl Anti-sense oligo

3 μl 10X T4 DNA Ligase Buffer

2 μl T4 Polynucleotide Kinase (10 units/μl)

19 μl water

Incubate 30 min at 37 ℃

Add 4 μl 0.5M NaCl

Boil (100 ℃, can be heated by a thermal cycle machine) for 2 min

Cool slowly on bench to room temperature for 3~4 hr.

Dilute 1 μl into 499μl water

Phosphorylation by T4 Polynucleotide Kinase and incubation at 37 ℃ for 30 min are optional.

5.4.3 Prepare the plasmid backbone harboring the sgRNA expression cassette

4. Perform digestion of pYF2.3G-Ribo-sgRNA (pYF515, if single-plasmid transformation) with *Nhe* I and *Bsa* I:

10 μl 10 X CutSmart Buffer

3 μg pYF2.3G-Ribo-sgRNA DNA

1.5 μl *Nhe* I (10 u/μl)

1.5 μl *Bsa* I (10 u/μl)

Add H_2O for a total of 100 µl.

Incubate 3-5 hr at 37 ℃.

5. Clean up digested products using a Qiagen PCR purification kit or ethanol precipitation.

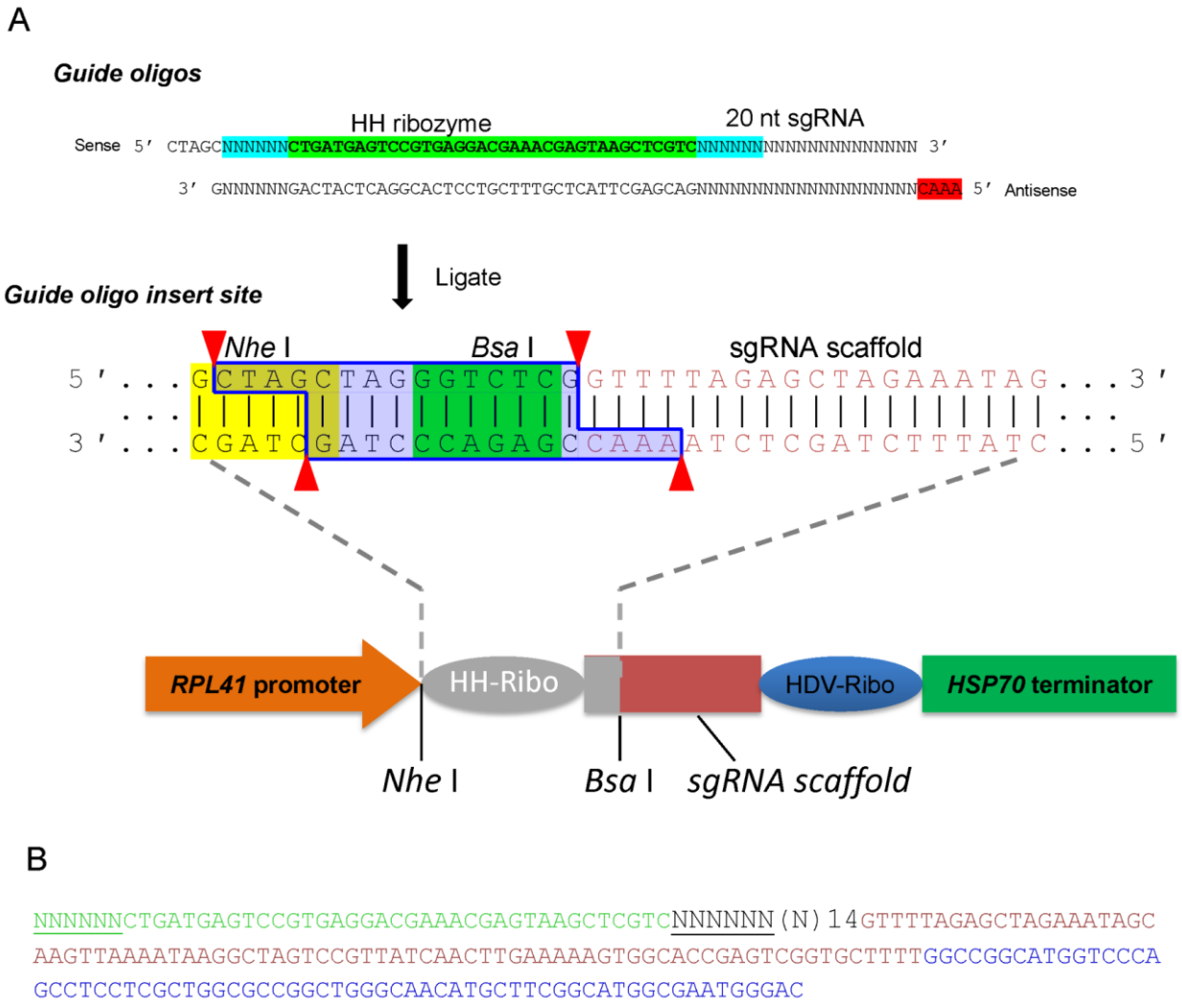


Fig. 5.3 Scheme for scarless cloning of the guide sequence oligonucleotides into a plasmid containing the sgRNA scaffold flanked by the HDV-ribozyme.

A. Steps showing cloning of a designed sgRNA-coding fragment into the generalized sgRNA expression plasmids (pYF2.3G-Ribo-sgRNA or pYF515). Any 20 nt-sgRNA coding sequence (together with the flanking HH-ribozyme) can be generated simply by the annealing of two oligos. The guide oligos contain overhangs for ligation into the *Nhe* I and *Bsa* I sites. When cloning ribozyme-flanked sgRNA, the "20 N" of sgRNA is should be replaced with the designed sgRNA sequence, and the first six nucleotides of the

HH-ribozyme (in cyan) should be filled in with the reverse complement of the first six nucleotides of the sgRNA target sequences (in cyan). Digestion of pYF2.3G-Ribo-sgRNA or pYF515 with *Nhe* I and *Bsa* I allows the direct insertion of annealed oligos into the cleaved plasmid.

B. Final sequence of the assembled sgRNA flanked by the two ribozymes. Nucleotides in larger font, the 20 nt target sequence; nucleotides in dark red, 80 nt sgRNA scaffold; nucleotides in green, hammerhead ribozyme (HH ribozyme); nucleotides in blue, HDV ribozyme; nucleotides underlined, the first six nucleotides of the Hammerhead (HH) ribozyme must be complementary to the first six nucleotides of the target sequence.

5.4.4 Ligate insert and plasmid DNA

6. Ligate the diluted annealed oligos (from step 3) with the purified digested plasmid (from step 5). Use the following reaction:

4 μl diluted annealed oligos

50 ng digested pYF2.3G-Ribo-sgRNA DNA (pYF515, if single-plasmid transformation)

2 μl 10X T4 Ligase Buffer

1 μl T4 DNA Ligase (400 u/μl)

Add H_2O for a total of 20 μl.

Incubate ligation reaction for 30 min at room temperature.

5.4.5 Transform ligation product

7. Thaw a 50-μl aliquot of chemically competent *E. coli* DH5α in the hand until the cells have almost fully defrosted and then put on ice.

8. Add 2 μl ligation products and incubate 30 min on ice.

9. Heat shock for 90 sec at 42 ℃, then incubate 2 min on ice.

10. Add 800 µl fresh SOC or LB medium and incubate the culture for 45 min at 37 °C with gentle shaking (<225 cycles/min).

11. Plate 200 µl bacteria onto LB agar containing 100 µg/ml ampicillin.

12. Incubate 12-16 hr at 37 ℃.

5.4.6 Verify the insertion of the sgRNA sequence into the plasmid by colony-PCR and Sanger sequencing

13. Set up PCR reaction in a thin-walled PCR strip-tube on ice. For each 20 µl reaction, you will need:

2 µl 10 × Green Taq buffer

1.6 µl dNTP (2.5 mM each)

0.4 µl M13F primer (10 µM)

0.4 µl sgRNA forward oligo (10 µM)

0.125 µl Green Taq DNA Polymerase (5 u/µl)

15.475 µl dH2O

A premix can be prepared depending on the numbers of colonies that will be examined.

14. Pick 3-6 individual colonies from step 12 using 10 µl sterile tips (or toothpicks); dip and twirl tips in the liquid in the PCR tubes, and then touch the same tips onto an agar plate to obtain replicas of the *E. coli* transformants, or directly drop the tips into 2 ml liquid LB supplied with 100 µg/ml ampicillin to create a backup culture.

15. Run PCR program using a thermocycler:

1 cycle 4 min at 94 ℃ (disruption of *E.coli* cells and denaturation of DNA)

32-35 cycles 30 sec at 94 ℃

30 sec at 55 ℃ (annealing)

1 min at 72 ℃ (extension)

1 cycle 5 min at 72 ℃ (final extension, optional)

16. Check PCR products by electrophoresing 8 µl of each product on a 1% agarose gel (See Current Protocols in Molecular Biology, UNIT 2.5A; Voytas, 2000). The expected DNA size is 700 bp.

17. Extract plasmids from the positive *E.coli* transformants using the Zyppy Miniprep kit (or your favorite miniprep method).

18. Send two positive plasmids for sequencing using primer RPL41_Pseq_F. The correct sgRNA sequence should be the same as the final sequence in Fig. 5.3B.

19. Perform DNA midipreps of the sequence-verified sgRNA expression plasmid (or 'all-in-one' plasmid) to obtain 200 µg of the plasmid DNA.

Our experience is that the annealing-oligo based cloning strategy is very efficient. With experience, you may not need to perform colony-PCR but just send two colonies for sequencing directly.

5.4.7 Prepare Cas9 plasmid (only for two-plasmid transformation system)

20. Transform *E. coli* DH5α cells with 10 ng plasmid pYF2-PsNLS-hSpCas9 and spread onto an LB agar plate containing 100 µg/ml ampicillin (see Basic Protocol 2, steps 7 to 11). Incubate plate overnight at 37 ℃.

21. Once colonies are visible, pick a single colony from the plate and inoculate into 1 ml liquid LB medium containing 100 µg/ml ampicillin. Grow 6-8 hr at 37 ℃ with shaking at 280 rpm.

22. Transfer the 1 ml inoculum into a flask for large-scale culture (100 ml to 250 ml, depending on how much plasmid DNA you want).

23. Extract plasmid DNA from the harvested *E. coli* cells using a plasmid Maxiprep kit. Use the Nanodrop microspectrophotometer to measure DNA concentration. Re-suspend DNA at 2-5 μg/μl in water. Use this product for *P. sojae* transformation

Occasionally, we found E. coli strains carrying pYF2-PsNLS-hSpCas9 cannot grow densely, possibly due to the toxicity of Cas9 weakly expressed from the oomycete promoter, but you still may obtain ~200ng/μl plasmid (in 50 μl water) from a relatively low density culture. To increase the yield of Cas9 plasmid, you may streak the E.coli culture onto an LB plate containing 100 μg/ml ampicillin to enrich the E.coli transformants harboring high copies of Cas9 plasmid, and then inoculate a colony into liquid medium again.

5.5 BASIC PROTOCOL 3: Preparation of homologous donor template for HDR-mediated mutation

Compared to the NHEJ-mediated repair pathway which produces an unpredictable variety of small indels, HDR produces precise gene edits that are more useful for making specific mutations. To accomplish HDR-mediated genome editing, a homologous donor template should be provided, in addition to the CRISPR components. Any circular plasmid can be used as a carrier of the HDR template. Our study indicated that HDR occurs efficiently in *P. sojae* when 1 kb of homologous flanking sequences (homology arms) on each side are included (Fang & Tyler, 2016). Here, we present an example of replacing a target gene with a donor DNA. We have used the common plasmid vector pBluescript SK II $^{+}$ (pBS-SK II^{+}) as a carrier of the HDR template

(Fig. 5.4) but any convenient plasmid vector will do. We often use the fluorescent protein gene *mCherry* as a donor DNA to replace the open reading frame (ORF) of the target gene so that we can use the transformants to monitor expression of the gene during infection. Assembly of the homologous donor template may be conveniently achieved using the In-Fusion PCR Cloning system (Clontech) but any strategy will do. A similar strategy can be applied to introduce small substitution mutations, but in that case the sgRNA recognition site should also be mutated (for example by substituting redundant codons) to avoid cleavage of the HDR donor. The same strategy can also be used to produce a clean deletion by joining the two homology arms directly together.

5.5.1 Materials

Plasmid pBluescript SK II $^{+}$

Donor DNA (for example, *mCherry* gene)

Phusion High-Fidelity DNA Polymerase (NEB, cat. no. M0530S)

In-Fusion HD Cloning Kit (Clontech)

PCR purification kit (Qiagen)

Zyppy plasmid Mini Kit (Zymo)

HiSpeed Plasmid Midi kit (Qiagen)

Sterile pipet tips (or toothpicks) for picking colonies from agar plates

37 ℃ incubator-shaker

Nanodrop microspectrophotometer (*http://www.nanodrop.com*)

42 ℃ water bath for heat-shocking cells

15-ml bacterial culture tubes

Access to a Sanger sequencing facility

Access to a DNA synthesis service

5.5.2 Construct homologous donor template

1. Design primers for In-Fusion PCR cloning with help of the online design tool (http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Cloning_Resources/Online_In-Fusion_Tools)

Two restriction enzyme sites within the MCS of pBS-SK II$^+$ will be needed, so decide which two enzyme sites are going to be used.

2. PCR-amplify (See Current protocols in toxicology, UNIT 15.1; Kramer and Coen, 2000) the upstream and downstream sequences of the gene of interest (1000 bp), as well as a donor DNA fragment (Fig. 5.4).

3. Clean up PCR products using Qiagen PCR purification kit.

4. Digest the plasmid carrier pBS-SK II$^+$ in the MCS using the two selected restriction enzymes, and clean up the products using Qiagen PCR purification kit.

5. Set up the In-Fusion cloning reaction:

2 μl	5X In-Fusion HD Enzyme Premix
50–100 ng	Linearized Vector
50–100 ng	Purified PCR Fragment1
50–100 ng	Purified PCR Fragment2
50–100 ng	Purified PCR Fragment3

Add H_2O to make a total 10 μl.

Incubate the reaction for 15 min at 50 ℃, then place on ice.

The latest version of the In-Fusion PCR cloning design tool also includes the In-Fusion® Molar Ratio Calculator to calculate the optimal amounts of vector and insert for the In-Fusion® Cloning reaction.

http://www.clontech.com/US/Support/xxclt_onlineToolsLoad.jsp?citemId=http://bioinfo.clontech.com/infusion/molarRatio.do&section=16260&xxheight=750

6. Transform 50 µl chemically competent *E. coli* DH5α cells with 2 µl of the ligation product and then spread the bacteria onto LB plates containing 100 µg/ml ampicillin (see Basic Protocol 2, steps 7 to 11). Incubate overnight at 37 ℃.

7. Verify the successful insertion of the three PCR amplicons by colony-PCR and Sanger sequencing.

The M13F primer in the plasmid pBS-SK II$^+$ and the reverse primer for PCR amplification of the right homology arm (or M13R and forward primer for PCR amplification of the left homology arm) can be used for colony-PCR screening for positive E. coli transformants.

Further verify correct assembly of donor DNA by sequencing the vector using primers M13F, M13R and two primers inside the donor DNA insert (Fig. 5.4).

8. Perform DNA midipreps for the sequence-verified donor plasmid DNA.

5.6 BASIC PROTOCOL 4: Optimized *P. sojae* transformation

Polyethylene Glycol (PEG) mediated protoplast transformations are used to introduce the plasmid DNAs into *P. sojae*. To achieve a higher transformation rate, we further optimized the previously described transformation methods (Dou *et al*., 2008; Mcleod *et al*., 2008). The optimized protocol is described below.

5.6.1 Materials

Midiprep of sgRNA expression plasmid (pYF2.3G-Ribo-sgRNA, for two-plasmid system; see Basic Protocol 2)

Maxiprep of Cas9 expression plasmid (pYF2-hSpCas9, for two-plasmid system; see Basic Protocol 2)

OR: Midiprep of 'all-in-one' Cas9-sgRNA expression plasmid (pYF515, for single-plasmid system; see Basic Protocol 2)

Midiprep of homologous donor plasmid (see Basic Protocol 3)

Phytophthora sojae race 2 (P6497)

V8 liquid and solid media (See *REAGENTS AND SOLUTIONS*)

Nutrient Pea Broth and agar media (See *REAGENTS AND SOLUTIONS*)

Regeneration media (See *REAGENTS AND SOLUTIONS*)

0.8 M mannitol (Sigma, cat. no. M1902), sterilize using a 0.45-mm filter

0.5M $CaCl_2$ (Sigma, cat. no. C3306), sterilize using a 0.45-μm filter

0.5 M 4-morpholinoethanesulfonic acid (MES; Sigma, cat. no. M2933). Adjust pH to 5.7 using 1M KOH; sterilize using a 0.45-μm filter

0.5 M KCl (Sigma, cat. no. P3911), sterilize using a 0.45-μm filter

Lysing Enzymes from *Trichoderma harzianum* (Sigma, cat. no. L1412)

CELLULYSIN® Cellulase (Calbiochem, cat. no. 219466)

Enzyme solution (See *REAGENTS AND SOLUTIONS)*

PEG4000 (Fluka, cat. no. 81240)

W5 solution (See *REAGENTS AND SOLUTIONS)*

MMG solution (See *REAGENTS AND SOLUTIONS)*

PEG–calcium transformation solution (See *REAGENTS AND SOLUTIONS)*

Ampicillin (Sigma)

G418 (Geneticin, AG Scientific)

50 ml Falcon tube

Beckman Coulter benchtop centrifuge with swing buckets

25 ℃ incubator

250 ml Erlenmeyer flasks with cotton stoppers

Aluminum foil

#3 Cork borer (6-mm diameter)

Scalpel

Cheese cloth

Miracloth (EMD Millipore, cat. no. 475855-1R)

70 μm Falcon Nylon Mesh Cell Strainer (BD Biosciences cat. no. 08-771-2)

Two Stainless steel lab spatulas

Light Microscope (inverted format is more convenient)

5.6.2 *P. sojae* growth

1. Start a fresh culture of *P. sojae* on a 90-mm cleared V8 agar plate.

2. After 3-7 days, inoculate *P. sojae* discs (punched by #3 cork borer, each 6-mm diameter) to 90-mm Nutrient Pea agar plates.

Optimized step: we suggest adding β-sitosterol into the nutrient pea liquid and solid media (see REAGENTS AND SOLUTIONS) as this enables P. sojae to grow more robustly.

3. After 3-7 days (as long as the young hyphae haven't grown to the edge of the plates), inoculate four 250 ml Erlenmeyer flasks containing 50 ml nutrient pea broth with 5 mycelia discs (6-mm diameter) from the edge of the mycelial colonies. Seal flasks with cotton stoppers and cover with aluminum foil.

Four flasks of P. sojae culture are optimal for 6-10 DNA samples.

4. Grow for 2.5-4 days at 25 ℃ in the dark without shaking.

5.6.3 Protoplast isolation

5. Collect the 2.5-4 days old *P. sojae* mycelia as a mat (the agar plugs need not be excluded) using autoclave-sterilized Miracloth, rinse once in with water and then with 0.8M mannitol.

6. Transfer mycelial mats into a tall (25 mm deep) petri dish and cover with 0.8 M mannitol for ~10 min for plasmolysis.

Critical step: to treat the hyphae sufficiently with mannitol, it's important to disperse the mycelia uniformly with spatulas after placing them in the mannitol solution.

7. Prepare the enzyme solution during the break, 20 ml for 4 flasks of mycelia.

8. Collect the mycelia as a mat again using Miracloth and transfer washed mycelia quickly into the prepared enzyme solution.

Critical step: to enable the hyphae to be digested by the enzymes sufficiently, it's important to disperse the mycelia uniformly with spatulas after placing them in the enzyme solution.

9. Digest for 40 min-1 hr at room temperature with gentle shaking.

Gentle shaking can be done on a rotary shaker with ~50 rpm. Digestion efficiency can be tracked by checking for the release of protoplasts in the solution under the microscope. Digested P. sojae protoplasts are round, approximately 10-30 μm in size.

10. Filter the digestion products through a 70 μm Falcon™ Nylon Mesh Cell Strainer (BD Biosciences) to remove mycelial debris, collecting the flow-through.

11. Centrifuge the flow-through at 1, 200 g to pellet the protoplasts in a 50 ml Falcon tube for 1–2 min in a Beckman Coulter benchtop centrifuge with swing buckets.

12. Wash the pellet with 30 ml W5 solution, spin 1,200 g, 2 min.

To re-suspend protoplasts more easily, a small volume of W5 (~15 ml) can be added first. Once the protoplasts are re-suspended, add W5 to 30 ml.

13. Pour off the supernatant and re-suspend the protoplasts in 10 ml W5 solution. Rest the protoplasts by keeping them on ice for 20 min.

Convenience step: protoplasts can actually be kept in W5 solution on ice for up to 60 hr without obvious influence on the transformation efficiency

Optional: re-suspend protoplasts at $2x10^6$-$2X10^7$ /ml in W5 solution after counting the protoplasts under the microscope using a hemacytometer. Our experience is that usually 10 ml W5 is fine.

14. Centrifuge 2 min at 1,200 g, room temperature, and remove as much of the W5 solution as possible. Re-suspend the protoplasts in MMG solution. Keep at room temperature for 10 min.

If you count protoplasts in step 13, you can add the same volume of MMG as W5 to get $2x10^6$-$2X10^7$/ml protoplasts. Our experience is that adding 6-10 ml MMG solution (for 6-10 DNA samples) provides highest efficiency.

During the waiting time, you can start to prepare DNA.

5.6.4 DNA-PEG-calcium transformation

For single plasmid transformation

15a. Add <20 μl DNA (30 μg of plasmid DNA) to a 50 ml Falcon tube.

For co-transformation

15b. Add <20 µl DNA (Use 30 µg of the plasmid carrying the *NPT II* selectable marker gene, together with an equimolar ratio of any other DNAs included) to a 50 ml Falcon tube.

For example, for production of HDR mutants, add

30 µg Cas9 plasmid (~12 kb)

20 µg sgRNA plasmid (~7.6 kb)

15 µg donor DNA (pBS-HDR-mCherry, ~5.7kb)

16. Add 1 ml protoplasts to each DNA sample, mix completely by gently tapping the tube, and keep samples on ice for 5-20min.

17. Add 1.74 ml freshly prepared PEG solution (at room temperature) to each tube in three aliquots of 580 µl each. Let the PEG solution slide against the tube wall into the protoplast solution, and then gently rotate the tube to mix the samples. .

18. Incubate the transformation mixture for ~15 min on ice.

19. Add 2 ml of cold (4 ℃) regeneration media to each tube. Gently invert tube once and replace on ice for 2 min.

20. Add 8 ml more of cold regeneration media to each tube, gently invert tube once, and replace on ice for another 2 min.

21. Pour protoplasts in each Falcon tube into a tall Petri dish containing 10 ml of cold regeneration media. Add ampicillin (final concentration, 50 µg/ml).

Ampicillin (50 µg/ml) can be added during the protoplast incubation after DNA transfection if bacterial growth is a concern.

If it is important to recover multiple transformants that are guaranteed to be independent, several aliquots of the protoplasts should be poured into different petri dishes at this step. The different protoplast-derived cultures should then be maintained separately through the remainder of this and subsequent protocols.

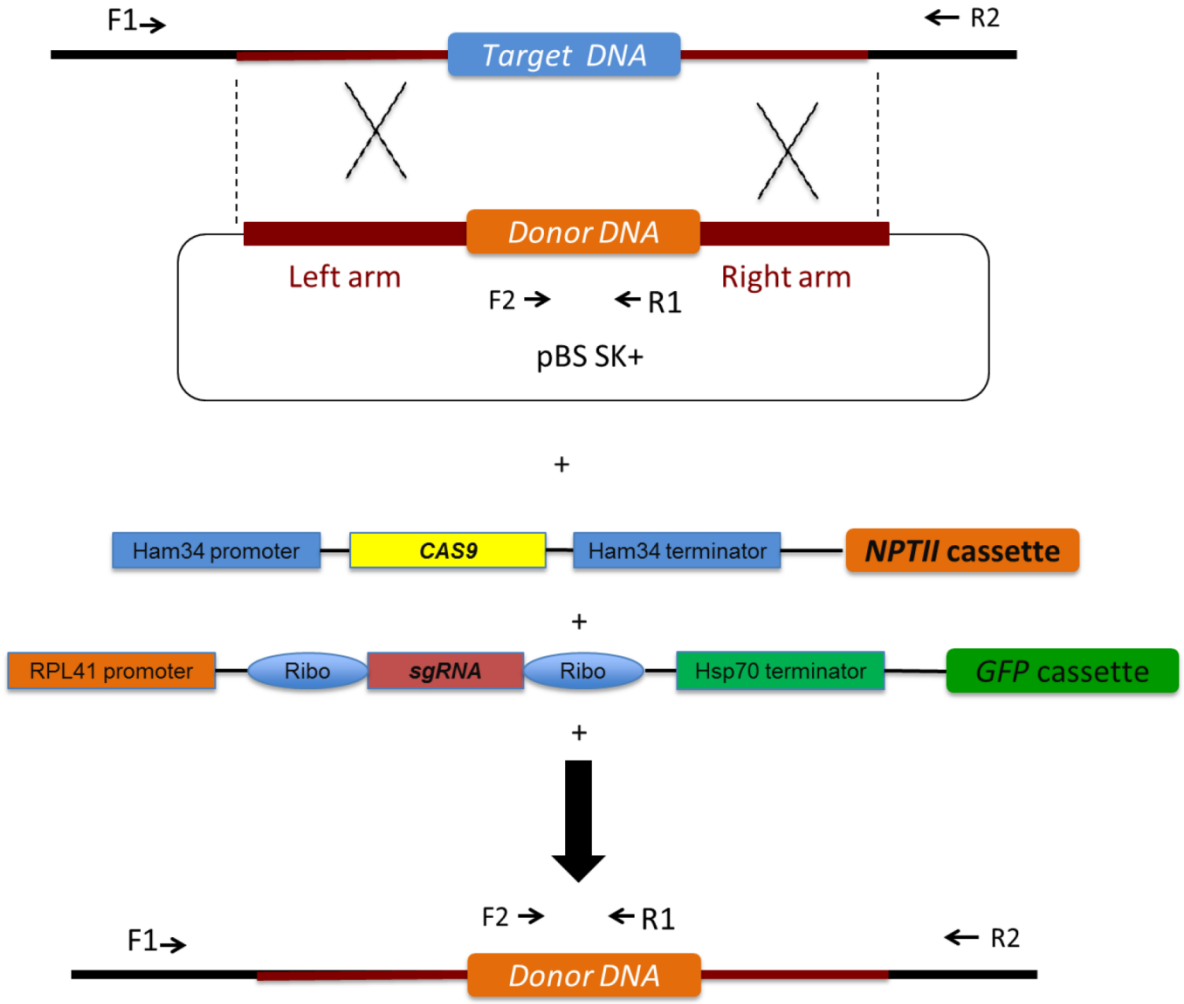


Fig. 5.4 Schematic of HDR-mediated modification of the target gene, stimulated by Cas9-induced DSB.

The all-in-one plasmid pYF515 can be used in place of the individual Cas9 and sgRNA plasmids. Primer pairs, F1/R1 and F2/R2 can be used for screening HDR mutants.

5.6.5 Regeneration and harvesting of hyphae

22. Incubate protoplasts without shaking for 12-18 hr at room temperature (~22 ℃) in the dark.

For transient expression assays

23a. Add G418 into the regenerated hyphae mix (final concentration 50 μg/ml) and incubate 1 d at 25 ℃ in the dark.

This step is optional but recommended; incubation of the regenerated hyphae in the presence of G418 can enrich the positive transformants.

24a. Re-suspend the hyphae in the petri dish then harvest them by centrifugation at 2,000 g for 2 min in the Beckman centrifuge.

25a. Pour off the supernatant until ~4 ml media left.

26a. Re-suspend the hyphae in the ~4 ml media remaining and transfer 1 ml of the suspension into a 1.5 ml Eppendorf tube.

The remaining 3 ml of mycelial culture can be further used for production of stable transformants.

27a. Pellet the 1 ml mycelial culture at 12,000 g.

28a. Discard the supernatant as much as possible and store samples for further genotype analysis at -80 ℃.

For production of stable transformants

23b. Re-suspend and harvest the regenerated hyphae by centrifugation at 2,000 g for 2 min in the Beckman centrifuge.

24b. Discard the supernatant until ~3 ml media left.

25b. Re-suspend and evenly divide mix into three 50 ml Falcon tubes.

26b. Add 45 ml liquid (42 ℃) regeneration media containing 1% agar and 50 μg/ml G418 to each Falcon tube, mix by inverting the tubes.

Critical step: the temperature for the liquid agar is very important, higher temperatures can kill the regenerated hyphae.

27b. Immediately pour each 45 ml agar mix into three empty 90 mm × 15 mm petri dishes.

28b. Incubate at 25 ℃ in the dark.

Mycelial colonies should be observable after 2 days' incubation.

5.7 BASIC PROTOCOL 5: Screening for *P. sojae* HDR-mediated mutants

After transformation, the targeting efficiency can be quickly assessed by genotyping the pooled transformants (transient expression assay). To obtain stable *P. sojae* mutants, it is necessary to screen transformants expressing both Cas9 and sgRNA. For the two-plasmid transformation system (Cas9 and sgRNA carried on two separate plasmids), a positive transformant should show both G418-resistance (a marker for the presence of the Cas9 plasmid) and GFP signal (a marker for the presence of the sgRNA plasmid). In contrast, a positive transformant generated by the 'all-in-one' transformation strategy need only show resistance to G418. For NHEJ transformants, mutations may be identified by PCR amplification combined with restriction enzyme digestion if there is a restriction site at the target site. Indels or point mutations caused by NHEJ will normally eliminate the restriction site and this can be useful for screening purposes. Alternatively the more general T7 endonuclease (T7EI) assay (Wyvekens *et al*., 2015) could be used. Since precise gene editing is more useful for most purposes, in this Basic Protocol, we describe the procedure for screening for *P. sojae* mutants harboring targeted gene replacements.

5.7.1 Materials

P. sojae lysis buffer (See REAGENTS AND SOLUTIONS)

Phenol-chloroform-isoamyl alcohol (25:24:1, saturated with 10 mM Tris, pH 8.0, 1 mM EDTA, Sigma, Cat. no. P3803)

Acid treated 0.5-mm glass beads (Sigma)

Liquid nitrogen

Isopropanol

70% ethanol

10 mg/ml RNase A, from bovine pancreas (Sigma, cat. no. R4875)

Pellet Pestle (disposable polypropylene Pestles)

Kimwipe paper

Phusion High-Fidelity DNA Polymerase (NEB, cat. no. M0530S)

PCR purification kit (Qiagen)

Benchtop Centrifuge

Refrigerated Benchtop Centrifuge

Vacuum concentrator

Access to a Sanger sequencing facility

Access to an oligonucleotide synthesis service

Light Microscope (inverted is more convenient)

5.7.2 Screening for stable *P. sojae* transformants expressing Cas9/sgRNA

1. Transfer ~60 visible *P. sojae* transformants growing on Nutrient Pea agar (Basic Protocol 4, step 28) each to 2 ml V8 liquid media supplemented with 50 μg/ml G418 in a well of a 12-well tissue culture plate. Propagate for 2-3 d at 25 ℃ in dark without shaking.

2. Check GFP expression of the G418-resistant *P. sojae* transformants with a fluorescence microscope.

Step 2 is only necessary for the two-plasmid transformation system, not for the 'all-in-one' transformation system.

If a fluorescence microscope is not accessible, genomic DNA extraction and PCR amplification can also be used to verify the presence of the GFP or sgRNA gene.

5.7.3 Isolation of genomic DNA (gDNA)

For pooled transformants

3a. Begin with the hyphae mix from Basic Protocol 4, step 28a.

For individual transformants

3b. Collect *P. sojae* hyphae from positive transformants

Don't pick all of the hyphae from the wells, leave some as backups.

4. Obtain *P. sojae* gDNA by Miniprep (See 5.8 SUPPORT PROTOCOL 5)

5.7.4 Detection and primary screening for gene replacement by PCR amplification and Sanger sequencing

Regular PCR combined with restriction enzyme cleavage generally allows for effective screening of *P. sojae* HDR mutants. In the case of the transient expression assay, as the background caused by un-transformed *P. sojae* protoplasts may be high, gene replacements in transformants can be detected using one primer located inside the donor DNA and the other one located outside the homology arm (e.g. primer pairs F1/R1 or F2/R2 in Fig. 5.4). In the case of the stable transformants, use primers both located outside of the repair template sequence region (e.g. primer pair F1/R2 in Fig. 5.4) so that both the target sequence and the donor sequence can be detected. Restriction enzymes unique to the respective target DNA and donor DNA sequences can be selected to verify the presence of the replacement in the PCR products, as well as the homokaryosis and homozygosity of the transformants. If the size of the target DNA sequence shows sufficient (usually >200 bp) difference from the donor DNA sequence, homokaryosis and homozygosity may also be assessed by agarose gel electrophoresis without the use of restriction enzymes.

5. Set up the following PCR mix:

10 μl 5 X Phusion HF

1 μl 10 mM dNTP mix

1.5 μl 10 μM forward primer

1.5 μl 10 μM reverse primer

1-3 μl *P. sojae* gDNA

1.5 μl DMSO

0.5 µl Phusion DNA Polymerase (2u/µl)

H_2O to 50 µl.

6. Run reactions in a thermal cycler as described in the manufacturer's instructions for NEB Phusion High-Fidelity DNA Polymerase.

7. Electrophoresis analysis of the PCR products.

8. Purify PCR products using Qiagen PCR purification kit.

9. Digest PCR products using a restriction enzyme that is unique to the donor DNA and another one specific to target DNA (or the same enzyme if different digestion patterns are expected from the donor and target sequences).

10. Analyze digestion products by electrophoresis.

Replaced target sequences should only be digested by the enzymes specific to the donor DNA, but should be resistant to the enzymes unique to target DNA, and vice versa.

11. Confirm gene replacements by Sanger sequencing based on the different situations:

a. For transient expression, sequence the PCR products amplified from one primer located inside the donor DNA and the other one located outside the homology arm (e.g. primer pair F1/R1, or F2/R2 in Fig. 5.4)

b. For a stable transformant, if the size of target DNA is close to the donor DNA (cannot be clearly separated on an agarose gel by electrophoresis), sequence PCR products amplified using the same set of primers as in situation a.

c. For a stable transformant, if the size of target DNA shows >200 bp difference from the donor DNA (can be clearly separated on an agarose gel). Excise and sequence the expected band.

5.7.5 Single zoospore isolation of homokaryotic mutants

As *P. sojae* protoplasts and hyphae are multinucleate, a *P. sojae* transformant might be heterokaryotic, harboring nuclei with a diversity of genotypes. Isolation of single zoospores, which are >95% mononucleate, should be used to obtain homokaryotic mycelial colonies carrying the desired genotype.

12. Transfer sequence-verified mutants (from the backup mycelia at step 3b) to V8 agar supplemented with 50 μg/ml G418, and then grow at 25 °C in dark for 7 to 10 days.

Seven to ten days old mycelia will produce the greatest amount of zoospores (This observation is for P6497, R2). The mycelia should not have reached the edge of the plate. Thin plates (15 ml for a 90 mm plate) of V8 agar is preferred, because this allows quicker depletion of nutrients in medium when flooding with water and results in greater sporangia production.

13. Flood plates with sterile room temperature dH_2O every ~ 30 min.

Make sure the entire mycelial colony is covered. Zoospore release usually begins after 4 hr of washing the plates.

The generation of sporangia and release of zoospores can be examined by a regular microscope. After 4-5 hours of washing/flooding a very few zoospores swimming should be visible, this is a good indication that there has been enough washing.

14. Seal the plates with parafilm and incubate 12 to 16 hr at 14 ℃ in the dark.

15. Count zoospores by using a hemocytometer.

For P6497, usually 1 x 10^4 to 1 x 10^5 zoospores/ml can be obtained from a 90 mm plate.

16. Make a series of dilutions with sterile dH_2O and plate ~ 15 zoospores onto a 90-mm V8 plate. Grow at 25 ℃ in dark for ~2 days.

17. Check mycelia under the microscope, transfer ~6 mycelial colonies for each mutant into 2 ml V8 liquid media in a well of a 12-well tissue culture plate. Grow at 25 ℃ in the dark for ~2 days without shaking.

18. Genotype single zoospore colonies by repeating steps 3-11.

19. Once the homokaryotic *P. sojae* mutants are verified, save them on plain V8 agar (without G418 selection).

Conveniently, ectopic CRISPR/Cas9 plasmids can often be lost or silenced after several generations of subculture without G418 selection. Mutants lacking G418 resistance can be used for future complementation experiments or introduction of additional mutations.

20. Choose 2-3 mycelial colonies (from different transformants) containing the desired mutations for downstream experiments.

We recommend to place verified mutant cultures into liquid nitrogen storage for long-term preservation.

5.8 SUPPORT PROTOCOL 5: SMALL SCALE EXTRACTION OF *P. sojae* GENOMIC DNA (*P. sojae* gDNA MINIPREP)

A traditional phenol-chloroform method is used for *P. sojae* gDNA minipreps; this produces stable and high quality gDNA.

5.8.1 Materials

P. sojae lysis buffer (See REAGENTS AND SOLUTIONS)

Phenol-chloroform-isoamyl alcohol (25:24:1, saturated with 10 mM Tris, pH 8.0, 1 mM EDTA, Sigma, Cat. no. P3803)

Liquid nitrogen

Isopropanol

70% ethanol

10 mg/ml RNase A, from bovine pancreas (Sigma, cat. no. R4875)

Acid treated 0.5-mm glass beads (Sigma)

Pellet Pestle (disposable polypropylene Pestles)

Benchtop Centrifuge

Refrigerated Benchtop Centrifuge

Vacuum concentrator

5.8.2 Breakage of *P. sojae* cell walls

For pooled transformants

1a. Collect media containing the hyphae mixture.

2-ml Eppendorf tubes with a round bottom are easier for beads to break samples sufficiently.

Samples can be stored for long-term storage at -80 ℃.

2a. Centrifuge 15 min at 12, 000 g, room temperature, discard supernatant as much as possible.

3a. Add a roughly equal volume of sterile 0.5 mm glass beads.

4a. Add 500 µl lysis buffer (add 0.1 mg/ml RNase A prior to use).

5a. Vortex samples for 1 min, then put on ice for 1 min. Repeat 3 times.

For individual transformants

1b. Pick a fingertip size (3-7 mm diameter) clump of *P. sojae* hyphae using sterile tips or toothpicks, remove excess liquid on Kimwipe paper

Remove as much liquid from the hyphae clump as possible using the Kimwipes; this makes subsequent grinding of the frozen mycelia easier.

2b. Transfer into an empty 1.5 ml Eppendorf tube, then freeze in liquid nitrogen.

Frozen samples can be stored for long-term storage at -80 ℃.

3b. Grind frozen hyphae to powder using pellet pestle

4b. Re-suspend hyphal powder in 500 µl lysis buffer (contains 0.1 mg/ml RNase A).

5b. Mix by vortexing for 30 s.

6. Incubated hyphal lysates for 30 min at 37 ℃ for RNA digestion.

5.8.3 Purification of gDNA

7. Add an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1; saturated with TE). Mix by inverting the tubes 5 times.

8. Centrifuge 15 min at 12,000 g, room temperature.

9. Transfer the upper clear layer of aqueous supernatant to a new tube.

10. Add an equal volume of chloroform. Mix by inverting the tubes 5 times.

11. Centrifuge 15 min at 12,000 g, room temperature.

12. Transfer the aqueous phase to a new tube.

13. Precipitate DNA by adding 0.5 volume of isopropanol, mix thoroughly and place the tubes in -20 ℃ for 15 min.

14. Centrifuge 15 min at 12, 000 g, 4 ℃.

15. Carefully decant the supernatant.

16. Wash the DNA pellet with 1 ml room temperature 70% ethanol.

17. Centrifuge 5 min at 12, 000 g, 4 ℃.

18. Carefully decant supernatant.

19. Dry gDNA in a vacuum concentrator (~5min at 50 ℃).

20. Dissolve gDNA in ~40 µl water.

21. Examination of gDNA quality and quantity by agarose gel electrophoresis.

As these P. sojae gDNA solutions contain lots of polysaccharides, DNA concentration measurements by Nanodrop are not accurate. Instead, the quantity and quality of gDNA can be examined by electrophoresis analysis.

5.9 REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps.

Enzyme solution (20 ml for 4 flasks of P. sojae culture)

10 ml 0.8 M Mannitol (final 0.4 M)

0.8 ml 0.5 M KCl (final 20 mM)

0.8 ml 0.5 M MES, pH 5.7 (final 20 mM)

0.4 ml 0.5 M $CaCl_2$ (final 10 mM)

0.1-0.15 g Lysing Enzymes from *Trichoderma harzianum* (Sigma, cat. no. L1412, final 0.5%-0.75%, w/v)

0.1-0.15 g CELLULYSIN Cellulase, *Trichoderma viride* (Calbiochem, cat. no. 219466, final 0.5%- 0.75%, w/v)

H_2O to 20 ml

The best results are obtained with these two enzyme providers.

Recommendation: Prepare the enzyme solution right before P. sojae transformation.

MMG solution

18.22 g Mannitol (final 0.4 M)

0.76 g $MgCl_2$ $6H_2O$ (final 15 mM)

2 ml 0.5 M MES, pH 5.7 (final 4 mM)

H_2O to 250 ml

Store up to 6 months at 4 ℃.

Nutrient Pea Broth and Agar Media

1. Add following components

1.0 g KH_2PO_4 (final 0.1%, w/v)

1.0 g K_2HPO_4 (final 0.1%, w/v)

3.0 g KNO_3 (final 0.3%, w/v)

0.5 g $MgSO_4$ (final 0.05%, w/v)

0.1 g $CaCl_2$ (final 0.01%, w/v)

2.0 g $CaCO_3$ (final 0.2%, w/v)

5.0 g D-sorbitol (final 0.5%, w/v)

5.0 g D-mannitol (final 0.5%, w/v)

5.0 g D-glucose (final 0.5%, w/v)

2.0 g Yeast Extract (final 0.2%, w/v)

Pea Broth to 1 liter

2. Mix for ~30 minutes using a magnetic stir plate and stir bar to dissolve dry ingredients (not all of the $CaCO_3$ will dissolve).

3. Centrifuge the solution for 10 minutes at 5000 rpm, room temperature, to remove green pea debris and undissolved $CaCO_3$. Collect the supernatant carefully and bring the volume to 1 liter with Pea Broth.

Aliquot each one liter of medium into four 1-liter bottles (each bottle, 250 ml) is recommended.

4. Add β-sitosterol (final concentration 100 mg/ml)

Recommendation: β-sitosterol can be dissolved in acetone. A stock solution may be prepared with a concentration of 25 mg/ml. It may become precipitated at room temperature, but can be re-dissolved at 37 ℃ with agitation.

5. Add 1% agar for solid media,

6. Autoclave for 40 min.

7. After the media cools down, add the following components under sterile conditions.

2 ml Vitamin Stock

2 ml Trace Elements

Add vitamins and trace elements right before inoculation.

The Nutrient Pea Broth and Agar Media without Vitamin Stock and Trace Elements can be stored up to 3 months at 4 ℃.

Pea Broth (prepared for nutrient pea medium and regeneration medium)

1. Weigh 120g frozen green peas (e.g. Wal Mart) and bring volume up to 1 liter with distilled water.

Recommendation: Divide 120 g of frozen peas in four 1 L bottle, i.e. add 30 g peas, 250 ml water

2. Autoclave for 20 minutes.

3. Filter through 4 layers of cheesecloth. Squeeze the cheesecloth gently to remove residual broth from the peas.

Save at room temperature until needed for the media described below. Don't leave the pea broth at room temperature for more than 1 day as it is not sterile.

PEG-calcium transformation solution (40% w/v)

12 g	PEG 4000 (Fluka, cat. no. 81240, final 40% w/v)
6 ml	H_2O
7.5 ml	0.8 M Mannitol (final 0.2 M)
6 ml	0.5 M $CaCl_2$ (final 0.1 M)

Final volume is ~30 ml sufficient for 10 DNA samples.

Mix and sterilize using a 0.45-μm filter.

Store up to 1 week at 4 ℃.

***P. sojae* lysis buffer**

20 ml	1 M Tris-Cl, pH 8.0 (final 200 mM)
5 ml	0.5 M EDTA, pH 8.0 (final 25 mM)
1.2 g	NaCl (final 200 mM)
2.0 g	SDS (final 2%, w/v)

H_2O to 100 ml

Sterilization by autoclaving

Store up to 1 year at room temperature.

Note: add RNase A (final concentration 0.1 mg/ml) prior to use)

V8 medium (5X)

340 ml V8 vegetable juice

5 g CaCO3

Mix for 30 min

5000 rpm for 10 minutes, collect supernatant

Dilute 5 times with deionized water for 1X V8 liquid medium

Add 1% Agar for the solid medium

Sterilize by autoclaving

Store up to 3 months at 4 ℃.

Vitamin Stock (prepared for nutrient pea medium)

10 μl	0.02%, w/v Biotin (final 6.7×10^{-7} g/ml)
10 μl	0.02%, w/v Folic Acid (final 6.7×10^{-7} g/ml)
0.012 g	1-inositol (final 4×10^{-5} g/ml)
0.06 g	Nicotinic acid (final 2×10^{-4} g/ml)
0.18 g	Pyridoxine-HCl (final 6×10^{-4} g/ml)
0.015 g	Riboflavin (final 5×10^{-5} g/ml)
0.38 g	Thiamine-HCl (final 1.3×10^{-3} g/ml)

H_2O to 300 ml

Sterilize using a 0.45-μm filter

Store up to 1 year at 4 ℃.

W5 solution

0.093 g KCl (final 5 mM)

4.6 g $CaCl_2$ (final 125 mM)

2.25 g NaCl (final 154 mM)

7.8 g Glucose (final 177 mM)

H_2O to 250 ml

Store up to 6 months at 4 ℃.

Trace Elements (prepared for nutrient pea medium)

0.215 g $FeC_6H_5O_7 \cdot 3H_2O$ (final 5.4×10^{-4} g/ml)

0.15 g $ZnSO_4 \cdot 7H_2O$ (final 3.8×10^{-4} g/ml)

0.03 g $CuSO_4 \cdot 5H_2O$ (final 7.5×10^{-4} g/ml)

0.015 g $MgSO_4 \cdot H_2O$ (final 3.8×10^{-5} g/ml)

0.01 g H_3BO_3 (final 2.5×10^{-5} g/ml)

10 µl 0.7%, w/v MoO_3 (final 1.8×10^{-5} g/ml)

H_2O to 400 ml

Sterilize using a 0.45-µm filter

Store up to 1 year at 4 ℃.

5.10 COMMENTARY

5.10.1 Background information

The difficulty in producing gene disruptions and gene replacements in oomycetes results from the extremely low rate of homologous recombination with exogenous DNA observed in these organisms. For two decades, approaches for functional analysis have been limited to TILLING (Lamour *et al*., 2006) and gene silencing (Ah-Fong *et al*., 2008; Judelson *et al*., 1993b; Wang *et al*., 2011; Whisson *et al*., 2005). However, TILLING, which is based on random mutagenesis, is very laborious and requires long term storage of large pools of mutants, and has not proven very useful in oomycetes. Gene silencing (RNAi), triggered using hairpin, antisense, and sense RNA constructs (Ah-Fong *et al*., 2008) or using dsRNA directly (Wang *et al*., 2011; Whisson *et al*., 2005) has proven useful. However, knockdown of genes in oomycetes by RNAi is incomplete, and varies among gene targets, experiments and laboratories. Also, selective silencing of closely related genes is difficult. Nucleases like Cas9 cause DSBs and therefore function to increase the rate of targeted gene mutations. The successful application of CRISPR/Cas9-mediated genome editing in *P. sojae* thus brings a critical advance to the genetic toolbox for oomycetes.

The sgRNA expression plasmid for the *P. sojae* CRISPR/Cas9 system was initially constructed for a specific gene target in *P. sojae* (the original test case, *Avr4/6*). To adapt this technology to different DNA targets of interest and different oomycete species, we have created a generalized sgRNA plasmid that can be easily used for cloning any sgRNA by oligo-annealing and ligation. The original *P. sojae* CRISPR/Cas9 technology was based on a two-plasmid transformation system in which the sgRNA and Cas9 were carried on different plasmids. It was necessary to check both G418 resistance and a GFP signal to identify a positive *P. sojae*

transformant carrying both plasmids, which is relatively time-consuming. In this protocol, we have added an updated 'all-in-one' plasmid system which enables the same cloning strategy as the generalized sgRNA expression plasmid but is much more convenient for screening for positive transformants.

5.10.2 Critical parameters and troubleshooting

sgRNA design

To design an efficient sgRNA for *P. sojae*, one should follow three steps sequentially, i.e. (1) Identification of sgRNAs (by software, such as *EupaGDT*); (2) Selection of candidate sgRNAs and analysis of their off-target effects; (3) Examination of self-complementarity of the passing candidate sgRNAs (by the secondary structure web server, *RNAstructure*). It is important to keep in mind that sgRNA design software is just a bioinformatics tool that helps identify possible sgRNAs by considering a series of criteria. It may not identify the very best sgRNA, but it can exclude bad sgRNA designs such as sgRNAs having off-targets. Our experience is that sgRNA candidates whose 'total scores' exceed 0.5 on the web server *EupaGDT* actually do not exhibit any noticeable differences among each other. Thus you may consider sgRNA candidates with those scores (above 0.5) as equally top priority.

It has been reported that the last 10-12 nucleotides of sgRNA sequence next to PAM (defined as the 'seed sequence') were generally more important than pairing in the rest of the guide region (Cong *et al*., 2013; Jinek *et al*., 2012; Mali *et al*., 2013). By default, *EuPaGDT* defines the last 12 nt of sgRNA combined with the 3 nt PAM as a seed sequence (15 nt), and classifies any non-target genomic DNA sequences having <3 nt mismatches within the 15 nt seed sequence as off-targets. However, large variations have been observed across target sites, cell

types and species with respect to the importance of base pairing at each position (Hsu *et al*., 2013; Pattanayak *et al*., 2013; Wu *et al*., 2014), which make it challenging to establish reliable rules for sgRNA design. Therefore, it may be necessary to use different strategy such as deep sequencing to verify the absence of off-target mutations. Alternatively, to verify that a mutation introduced by CRISPR/Cas9 is truly responsible for an observed phenotype (and not an off-target mutation), a complementation experiment (reintroducing the targeted gene back into the mutant) may be used.

Examination of potential secondary structures in the selected sgRNAs is essential, as severe sgRNA hairpin structures can result in failure of targeting (Fang & Tyler, 2016). We don't have direct evidence connecting the numbers of RNA base-pairs and the efficiency of targeting, but we suggest limiting the consecutive base-pairing no more than three. Alternatively, to evaluate the activity of the designed sgRNA, you may conduct an *in vitro* cleavage assay using an *in vitro* transcribed sgRNA and a purified Cas9 protein together with a PCR-amplified template (Fang & Tyler, 2016). Transient expression assays in *P. sojae* protoplasts can also be used to rapidly evaluate the effectiveness of an sgRNA (Fang & Tyler, 2016).

Preparation of sgRNA expression plasmid

Oligo annealing based cloning is very sensitive. It is critical to add the *Nhe* I- and *Bsa* I-compatible overhangs to the ends of the two oligos correctly; any tiny error may cause failure of ligation to the backbone plasmid. In addition, heat the two oligos sufficiently and let them anneal naturally at room temperature.

Assembly of homologous repair template

Assembly of a homologous repair template is relatively complicated, because four fragments (linearized plasmid carrier, left homology-arm, right homology-arm and donor DNA) must be combined with one another. We have found that the In-fusion cloning kit (Clontech) is a good option to facilitate the cloning, but it is important to follow the instructions provided by the manufacturer. Occasionally, In-fusion has proved problematic and we've used sequential ligation of each fragment into the plasmid backbone.

Precise gene editing mediated by HDR can also be used to make point-mutations in the targeted sequence. However, it is essential to provide a HDR template in which the sgRNA recognition site has been modified. Replacement of degenerate codons is the easiest strategy to introduce the mutations needed to avoid cleavage of the HDR donor sequence, and also offers the opportunity to create (or remove) a restriction site useful for screening.

P. sojae transformation

Protoplast isolation and hyphal regeneration are the two most important factors for *P. sojae* transformation. This basic protocol has been successfully used for transformations of other *Phytophthora* species such as *P. capsici*, *P. parasitica,* and *P. megakarya.* If you plan to try the protocol for other oomycete species, you may need to consider these two factors. For instance, the *P. sojae* cell wall lysis buffer also worked for *P. megakarya*, but the regeneration media was not suitable (V8/0.5 M mannitol media was preferred, instead of the nutrient pea/0.5 M mannitol). It is valuable to check the quantity and quality of protoplasts and regeneration by microscopy. In the case of co-transformation experiments, it is important to control the proportion of DNA samples (keep each DNA sample at the same molar ratio) to obtain a high rate of positive transformants.

Screening for P. sojae mutants

Transient expression assays using pooled *P. sojae* transformants is a fast way to detect the efficiency of designed sgRNAs. High-fidelity DNA polymerase must be used to avoid mutations caused during PCR amplification. The background of gDNA from pooled transformants is often very high and can mask mutations. You may design a nested PCR strategy to enrich the mutated sequences.

5.10.3 Anticipated results

The genome editing efficiency strongly depends on the transformation efficiency as well as the sgRNA design. For the two plasmid transformation system, positive transformants (showing both the G418 resistance and GFP signal) generally constitute 10%-20% of the G418-resistant mycelial colonies. Among those positive transformants, >80% may be mutants (for both NHEJ- and HDR-mediated mutagenesis). In the case of the 'all-in-one' plasmid system, our initial quality control tests using the *Avr4/6* target showed that >90% colonies growing in V8 supplemented with 50 μg/ml G418 were transformants. Among those G418-resistant transformants, ~80% harbored a NHEJ mutation (~50% were homozygous). Similar results are expected in the HDR transformants.

5.10.4 Time Considerations

See Fig. 5.1 for a description of the time estimated for the protocol described in this unit.

5.4.5 Acknowledgements

We are grateful for many members of the Tyler lab at Oregon State University and of the Oomycete Molecular Genetics community for constructive feedback regarding use of our *P. sojae* CRISPR system. We also thank Duo Peng at the University of Georgia for useful discussions about the usage of EuPathGDT. This work was supported in part by USDA NIFA grant #2011-68004-30104 to B.M.T. L.C. was supported by China Scholarship Council (No. 201508410357). BG was partially supported by a China Scholarship Council fellowship (No. 201506305046) via Northwest A&F University, the People's Republic of China.

5.11 REFERENCES

Ah-Fong, A.M., Bormann-Chung, C.A., and Judelson, H.S. 2008. Optimization of transgene-mediated silencing in *Phytophthora infestans* and its association with small-interfering RNAs. *Fungal Genet. Biol.* 45:1197-1205.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819-823.

Deveau, H., Garneau, J.E., and Moineau, S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* 64:475-493.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32:1262-1267.

Dou, D., Kale, S.D., Wang, X., Chen, Y., Wang, Q., Wang, X., Jiang, R.H., Arredondo, F.D., Anderson, R.G., Thakur, P.B., McDowell, J.M., Wang, Y., and Tyler, B.M. 2008. Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell* 20:1118-1133.

Elbing, K. and Brent, R. 2002. Media preparation and bacteriological tools. *Current Protocols in Molecular Biology* 1.1. 1-1.1. 7.

Erwin, D.C. and Ribeiro, O.K. 1996. *Phytophthora* diseases worldwide. American Phytopathological Society Press, St. Paul, M. N..

Fang, Y. and Tyler, B.M. 2016. Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* 17:127-139.

Förster, H., Coffey, M.D., Elwood, H., and Sogin, M.L. 1990. Sequence analysis of the small subunit ribosomal RNAs of three zoosporic fungi and implications for fungal evolution. *Mycologia* 306-312.

Gao, Y. and Zhao, Y. 2014. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol* 56:343-349.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67-71.

Horvath, P. and Barrangou, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167-170.

Hsu, P.D., Lander, E.S., and Zhang, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157:1262-1278.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., and Zhang, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31:827-832.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816-821.

Judelson, H.S. 1997. The Genetics and Biology of *Phytophthora infestans*: Modern approaches to a historical challenge. *Fungal Genet. Biol.* 22:65-76.

Judelson, H.S. and Blanco, F.A. 2005. The spores of *Phytophthora*: weapons of the plant destroyer. *Nat. Rev. Microbiol.* 3:47-58.

Judelson, H.S., Coffey, M.D., Arredondo, F.R., and Tyler, B.M. 1993a. Transformation of the oomycete pathogen *Phytophthora-Megasperma* F-Sp Glycinea occurs by DNA integration into single or multiple chromosomes. *Curr. Genet.* 23:211-218.

Judelson, H.S., Dudler, R., Pieterse, C.J., Unkles, S.E., and Michelmore, R.W. 1993b. Expression and antisense inhibition of transgenes in Phytophthora infestons is modulated by choice of promoter and position effects. *Gene* 133:63-69.

Judelson, H.S., Tyler, B.M., and Michelmore, R.W. 1991. Transformation of the oomycete pathogen, *Phytophthora infestans*. *Mol. Plant-Microbe Interact* 4:602-607.

Kramer, M.F. and Coen, D.M. 2001. Enzymatic amplification of DNA by PCR: standard procedures and optimization. *Current protocols in toxicology* A. 3C. 1-A. 3C. 14.

Lamour, K.H., Finley, L., Hurtado-Gonzales, O., Gobena, D., Tierney, M., and Meijer, H.J. 2006. Targeted gene mutation in *Phytophthora* spp. *Mol. Plant-Microbe Interact.* 19:1359-1367.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. 2013. RNA-guided human genome engineering via Cas9. *Science* 339:823-826.

Mcleod, A., Fry, B.A., Zuluaga, A.P., Myers, K.L., and Fry, W.E. 2008. Toward improvements of oomycete transformation protocols. *J. Eukaryot. Microbiol.* 55:103-109.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D., and Rebar, E.J. 2011. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29:143-148.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31:839-843.

Peng, D. and Tarleton, R. 2015. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microbial Genomics* 1.

Seidman, C.E., Struhl, K., Sheen, J., and Jessen, T. 1997. Introduction of plasmid DNA into cells. *Current protocols in molecular biology* 1.8. 1-1.8. 10.

Stajich, J.E., Harris, T., Brunk, B.P., Brestelli, J., Fischer, S., Harb, O.S., Kissinger, J.C., Li, W., Nayak, V., and Pinney, D.F. 2011. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res.* gkr918.

Tyler, B.M. 2001. Genetics and genomics of the oomycete–host interface. *Trends Genet.* 17:611-614.

Tyler, B.M. 2007. *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* 8:1-8.

Tyler, B.M. and Gijzen, M. 2014. The *Phytophthora sojae* genome sequence: Foundation for a revolution. *In* Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens pp. 133-157. Springer.

Wang, Q., Han, C., Ferreira, A.O., Yu, X., Ye, W., Tripathy, S., Kale, S.D., Gu, B., Sheng, Y., Sui, Y., Wang, X., Zhang, Z., Cheng, B., Dong, S., Shan, W., Zheng, X., Dou, D., Tyler, B.M., and Wang, Y. 2011. Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. *Plant Cell* 23:2064-2086.

Whisson, S.C., Avrova, A.O., Van West, P., and Jones, J.T. 2005. A method for double-stranded RNA-mediated transient gene silencing in *Phytophthora infestans*. *Mol. Plant Pathol.* 6:153-163.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., and Chen, S. 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32:670-676.

Wyvekens, N., Tsai, S.Q., and Joung, J.K. 2015. Genome editing in human cells using CRISPR/Cas nucleases. *Current Protocols in Molecular Biology* 31.33. 31-31.33. 18.

Chapter 6 Conclusions

Phytophthora species are plant killers. They include many destructive plant pathogens that damage a huge range of agriculturally and ornamentally important plants (Tyler, 2007; Jiang & Tyler, 2012; Kamoun *et al.*, 2015). Although *Phytophthora* species resemble filamentous fungi, they are oomycetes, in the kingdom Stramenopila. Because of their complexity (diploid) and difficulty in manipulation (lack of efficient genetic tools), oomycetes have remained understudied at the molecular level over the past decades. However, increased awareness of the economic importance of this group of pathogens and the advent of genomics resources has driven oomycete genetics research recently (Kamoun *et al.*, 2015). In this dissertation I have described my progress in dissecting the mechanisms of nuclear import of proteins in the model oomycete *Phytophthora sojae*, and in using that information to devise an efficient genome editing strategy in oomycetes.

To date, most of the intracellular trafficking studies have been carried out in model systems, such as *Saccharomyces cerevisiae* and mammalian cells. Nuclear localization signals (NLSs) that target proteins into nuclei have not been defined in oomycetes. As described in Chapter 2, using a reliable *in vivo* NLS assay based on confocal microscopy, I found that many canonical monopartite and bipartite classical NLSs (cNLSs) mediated nuclear import poorly in *P. sojae*. I found that efficient localization of *P. sojae* nuclear proteins by cNLSs requires additional basic amino acids at distal sites or collaboration with other NLSs. I also found that several prototypes of another well-characterized NLS, proline-tyrosine NLS (PY-NLS) showed weak nuclear targeting activity in *P. sojae*. To characterize PY-NLSs in *P. sojae*, I experimentally defined the residues required by functional PY-NLSs in three *P. sojae* nuclear-

localized proteins. These results showed that functional *P. sojae* PY-NLSs need additional clusters of basic residues either internally or externally, or require collaboration with other weak NLSs for efficient nuclear import. Finally, investigation of several highly conserved *P. sojae* nuclear proteins including ribosomal proteins and core histones revealed that these proteins exhibit a similar but stronger set of sequence requirements for nuclear targeting compared with their orthologs in mammals or yeast.

To learn more about NLS-mediated transport of nuclear proteins in *P. sojae* I have characterized in depth the nuclear import mechanism of a *P. sojae* basic leucine zipper transcription factor, PsbZIP1 (Chapter 3). I found that the nuclear translocation of PsbZIP1 was determined by a central conserved region. Mutational analysis of this region indicated its ability to confer nuclear localization was mediated by four distinct motifs but was independent of the conserved DNA binding residues. Three motifs showed autonomous NLS activity and the fourth motif served as a nuclear localization enhancer. Sequence comparisons and mutational analysis of these nuclear localization motifs revealed a new form of bipartite NLS consisting of a triplet of basic residues followed by a tail of scattered basic amino acids.

To overcome many of the challenges that have impeded genetic manipulation in oomycetes, I developed a CRISPR/Cas9 system enabling rapid and efficient genome editing in *P. sojae* (described in Chapter 4). Using the RXLR effector gene *Avr4/6* as target, I observed that in the absence of a homologous template, the repair of Cas9-induced double-strand breaks (DSBs) in *P. sojae* was mediated by non-homologous end joining (NHEJ), primarily resulting in short indels. Most mutants were homozygous, presumably due to gene conversion triggered by Cas9-mediated cleavage of non-mutant alleles. When donor DNA was present, homology directed repair (HDR) was observed at high frequency, which resulted in the replacement of the

target gene with the donor DNA. By testing the specific virulence of *P. sojae* strains carrying several NHEJ-induced mutations and HDR-mediated gene replacements on soybeans, I have validated the contribution of Avr4/6 to recognition by the soybean *R* gene loci, *Rps*4 and *Rps*6, but also uncovered additional contributions to resistance by these two loci.

After establishing the CRISPR/Cas9 system in *P. sojae*, I further optimized the system by generalizing the sgRNA cloning strategy to all kinds of sgRNA target genes (described in Chapter 5). I also upgraded my original Cas9/sgRNA two-plasmid system to an all-in-one system that harbors Cas9 and sgRNA expression cassettes on a single plasmid. Those modifications make genome editing in *P. sojae* simpler and faster. In this chapter, I also describe a detailed protocol for CRISPR/Cas9-mediated genome editing in *P. sojae*, including single guide RNA (sgRNA) design and construction, an optimized transformation method, efficient gene replacement, and mutant-screening strategies. This protocol should be generally applicable for most culturable oomycetes.

Overall, the studies focusing on nuclear localization have unveiled a distinctive nuclear trafficking mechanism in oomycetes, which will improve our understanding of the biology of these distinctive microorganisms. The efficient genome editing strategy described here will open a new era of oomycete genetic studies.

References

Jiang, R.H. & B.M. Tyler, (2012) Mechanisms and evolution of virulence in oomycetes. *Annu. Rev. Phytopathol.* **50**: 295-318.

Kamoun, S., O. Furzer, J.D. Jones, H.S. Judelson, G.S. Ali, R.J. Dalio, S.G. Roy, L. Schena, A. Zambounis, F. Panabieres, D. Cahill, M. Ruocco, A. Figueiredo, X.R. Chen, J. Hulvey, R. Stam, K. Lamour, M. Gijzen, B.M. Tyler, N.J. Grunwald, M.S. Mukhtar, D.F. Tome, M. Tor, G. Van Den Ackerveken, J. McDowell, F. Daayf, W.E. Fry, H. Lindqvist-Kreuze, H.J. Meijer, B. Petre, J.

Ristaino, K. Yoshida, P.R. Birch & F. Govers, (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**: 413-434.

Tyler, B.M., (2007) *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. *Mol. Plant Pathol.* **8**: 1-8.