Identifying *Xanthomonas* recruitment genes in the soil-borne legacy

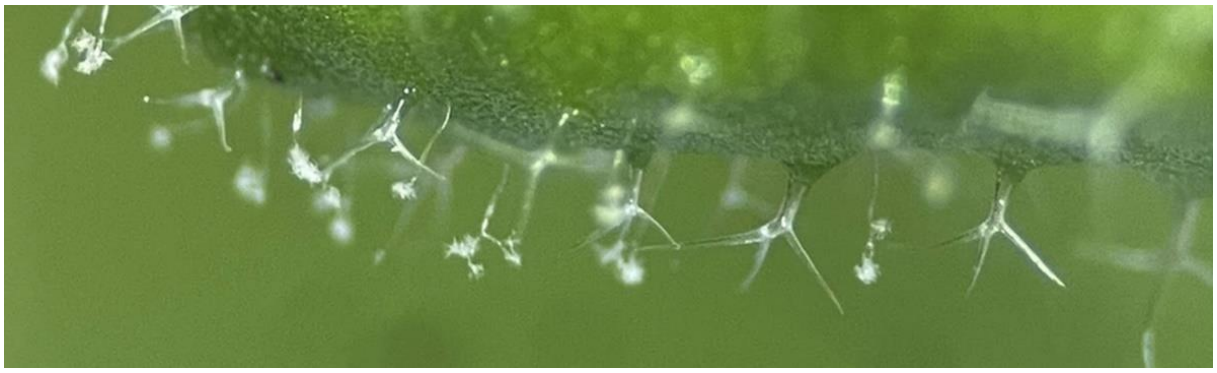
E. E. K. Klaasen

Plant-Microbe Interactions Group, Utrecht University

Daily supervisor: Dr. B. L. Ford

Examiner: Dr. R. L. Berendsen

Second examiner: Prof. dr. ir. C.M.J. Pieterse



©E.E.K.K.

Plain language summary

They are present in (almost) every environment, however, you cannot see them by eye: micro-organisms. One teaspoon of soil contains a billion microbes. It has been proven that plants can actively control which microbes live on and inside them. These beneficial microbes help the plant by facilitating nutrient uptake from the environment, increasing resilience to drought and heat stresses, but maybe most important, enhancing the resistance against pathogens. An example of a pathogen to which a plant uses bacteria to protect itself is the oomycete *Hyaloperonospora arabidopsidis* (Hpa), also known as downy mildew. Oomycetes resemble fungi, but they are from a different family. It has been proven that *Arabidopsis thaliana* can actively recruit a team of microbes known as the Hpa associated microbiota (HAM), which are able to suppress the Hpa infection. However, much is unknown about how these interactions work. It has been found that plants can use root secretions to select for beneficial microbes around their roots (the rhizosphere) and there is insight into which bacteria are recruited. For example, *Xanthomonas* (sp. WCS2014-23) is the most prevalent HAM member which significantly increases on the leaves and stems (phyllosphere) and rhizosphere in response to Hpa infection. But the processes inside the HAM are unknown. Therefore, with this research we aim to investigate which genes of *Xanthomonas* are important for its recruitment to the phyllosphere of *Arabidopsis* in response to aboveground infection with Hpa. To examine this, we made a mixture of thousands *Xanthomonas* bacteria with each a different, random mutation in its genome, also called a mutant library. This mutant library is used in multiple experiments to investigate which mutations impaired the recruitment to the plant during Hpa infection.

Our mutant library seems quite diverse since it has mutations in ~58% of the possible locations that we could mutate in the genome, which corresponds to ~50.6 to ~72.2% of the 4399 genes. We found 63 potential recruitment genes. The most promising candidates fall under three categories. The first two relate to the physical process of recruitment, which involves the processes of recognising a signal and moving to the shoots. The third category is more in terms of the interaction between *Xanthomonas* and Hpa (suppressing the infection and tolerance to Hpa in the phyllosphere). These genes fit a scenario where *Xanthomonas* perceives an Hpa-infected plant and starts a process inside the bacterium which causes its movement to the phyllosphere. Once arrived at the sites of infection, the bacterium uses certain tactics to fight Hpa and so reduce infection. With this research we provide insight into which genes of beneficial microbes could be involved in their recruitment to the plant in response to an infection in the shoot. Moreover, we provide leads for further research into beneficial plant-microbe interactions. This could contribute to the development of biological pesticides which use beneficial bacteria to control diseases instead of environmentally unfriendly chemicals.

Abstract

The plant microbiome can help the plant mitigate biotic stress. Hpa associated microbiota (HAM) are beneficial bacteria for which it is proven that *Arabidopsis thaliana* can actively recruit them in response to infection with the obligate biotrophic pathogen *Hyaloperonospora arabidopsidis* (Hpa). Furthermore, the HAM have been shown to suppress Hpa infection as part of the “soil-borne legacy” of generationally transmitted disease resistance. The mechanism behind this phenomenon is still poorly understood. On the plant side, secretion of root exudates such as coumarins has been identified as one mechanism for selection of beneficials. However, the mechanisms from the microbe side of the interaction are less understood, although it is known that *Xanthomonas* ASV a0e1a seems to be an important HAM member. A recent study by Goossens et al. (2023), showed that this is the most prevalent HAM ASV and that it responds strongly to Hpa infection. This study aims to investigate which genes in *Xanthomonas* sp. WCS2014-23 (representative of ASV a0e1a) are important for its recruitment to the phyllosphere of *Arabidopsis* during aboveground infection of Hpa. To investigate this, we created an INSeq mutant library in WCS2014-23 via transposon insertion mutagenesis. Subsequently, this library was utilised in multiple mutant screens involving recruitment experiments to identify candidate recruitment genes.

Our INSeq library had an insertion density of ~58%, which corresponds to insertions in ~50.6 to ~72.2% of the 4399 predicted coding sequences present in the reference genome. We identified 63 potential recruitment genes. The most promising candidates fall under three categories. The first two categories relate to the physical process of recruitment, “Perception & Regulation” and “Movement”. The last category, “Antagonism” relates to the interaction with Hpa which included suppressing the infection combined with co-occurring in the phyllosphere. These genes fit a scenario where *Xanthomonas* perceives an Hpa-infected plant and activates a process inside the bacterium which causes its movement to the phyllosphere. Once arrived at the sites of infection, the bacterium uses certain mechanisms to repress Hpa and reduce infection. With this research we provide insight into which genes of beneficial microbes could be involved in their recruitment to the plant in response to a foliar pathogen infection, while also providing leads for further research into beneficial plant-microbe interactions. Additionally, insight into how HAM are able to suppress Hpa could contribute to the development of new strategies for reducing crop losses due to Hpa infection. This could potentially include the development of bioinoculants with beneficial bacteria which could replace environmentally unfriendly, chemical pesticides.

Contents

PLAIN LANGUAGE SUMMARY 2
ABSTRACT 2
CONTENTS 4
INTRODUCTION 5
RESULTS 7
Creating a Xanthomonas INSeq mutant library *7*
Screening for recruitment genes *12*
How to beat the beads? – Troubleshooting for library preparation *20*
DISCUSSION 25
MATERIALS AND METHODS 28
Creating Xanthomonas mutant library *28*
Mutant screens *29*
DNA isolation, library preparation and sequencing *31*
INSeq data analysis *32*
REFERENCES 33
SUPPLEMENTAL 37
Primer sequences *37*
Scripts INSeq analysis *38*

Introduction

It is becoming increasingly clear that a plant's microbiome is important for its survival. For example, microorganisms in the rhizosphere can influence the ability of nutrient uptake, the resilience to abiotic stresses, but maybe the most important is that they can enhance the resistance against pathogens. A good example are beneficial microbes in disease-suppressive soils (Bakker et al., 2018). These soils typically develop in fields with monocultured crops after a severe disease outbreak. Since the altered microbiome composition in the soil also protects the next generation of plants grown on that soil, this phenomenon is called the soil-borne legacy (SBL). It has been shown that plants can assemble these protective microbes by a "cry for help" when attacked by a pathogen (Berendsen et al., 2018; Goossens, et al., 2023) (**Figure 1**). For example, *Arabidopsis thaliana* (from now abbreviated as *Arabidopsis*) selectively recruits specific bacteria to their rhizosphere and phyllosphere when infected with the downy mildew *Hyaloperonospora arabidopsidis* (from now abbreviated as Hpa) (Berendsen et al., 2018; Goossens et al., 2023). Hpa is a obligate biotrophic pathogen that causes severe crop losses, for example, up to 50% in cucumber and spinach (Correll et al., 1994; Keinath & de Figueiredo Silva, 2022).

The exact mechanisms of the soil-borne legacy and the cry for help are unknown, but some characteristics have been identified. For example, that *Arabidopsis* can select for beneficial microbes in the rhizosphere via the secretion of root exudates, especially coumarins (Vismans et al., 2022; Yuan et al., 2018). Furthermore, there are insights into which micro-organisms are recruited in response to certain pathogen attacks. For example, the HAM, Hpa-associated microbiota, are attracted by *Arabidopsis* during Hpa infection (Berendsen et al., 2018; Goossens et al., 2023). This recruitment takes place both to the rhizosphere and to the phyllosphere (Goossens et al., 2023). It has been proven that the application of three isolates from these HAM: *Xanthomonas* (sp. WCS2014-23), *Stenotrophomonas* and *Microbacterium*, activates induced systemic resistance (ISR) against Hpa, inhibits spore formation and promotes plant growth (Berendsen et al., 2018). Moreover, from the HAM, *Xanthomonas* ASV a0e1a seems to be an important member. A recent study by Goossens et al. (2023), showed that this is the most prevalent HAM and that it responds strongly to Hpa infection with increases in abundance of up to 17-fold in the phyllosphere and 2.7-fold in the rhizosphere. Nevertheless, the mechanisms inside these microbes (for example which bacterial genes are involved) are poorly understood.

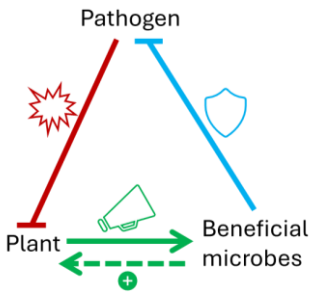


Figure 1. *Schematic overview of the cry for help hypothesis. In reaction to a pathogen attack, plants secrete signals, e.g. root exudates such as coumarins, to attract beneficial microbes which attack the pathogen. This reduces the infection. However, beneficial microbes may also have other advantages for the plant, for example growth promotion.*

Transposon sequencing (Tn-seq) is a powerful tool that has been employed in numerous studies to identify bacterial genes involved in various processes, including plant-microbe interactions. There are different variations of the technique, utilising different types of transposons. For instance, Cole et al., (2017) employed RB-Tn-seq (Randomly Barcoded transposon) to identify genes involved in the colonisation of *Arabidopsis* roots by *Pseudomonas simiae*. Similarly, Wheatley et al., (2020) used another form of Tn-seq called INSeq (insertion sequencing using an unbarcoded and modified mariner transposon) to identify which genes of *Rhizobium leguminosarum* are involved in four different symbiotic lifestyle stages with pea (*Pisum sativum*). Finally, Chen et al., (2020) used Tn5 based Tn-seq to identify which genes of *Pseudomonas taiwanensis* are responsible for its toxicity against fungi and the rice pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) among others.

The principle of Tn-seq is that a transposon mutant library is exposed to different conditions to select for genes which are essential in these specific circumstances, which is effectively a high-throughput mutant screen. This is determined based on comparing the relative abundances of mutations. Genes

that are required for the bacterial survival in a certain condition allow fewer mutations than non-essential or less essential genes. A transposon mutant library is a pool of bacteria, derived from the same strain, which all have a mutation in a random site of their genome due to the insertion of a transposon. These libraries are created by random transposon insertion mutagenesis, which uses the principle that transposons are mobile genetic elements which can integrate at (nearly) random places into the genome. Himar1 (mariner transposon) and Tn5 are the two most commonly used transposons for the construction of transposon mutant libraries (Ioerger, 2022). Tn5 can insert almost everywhere into the genome, Himar1 at random TA-sites (consecutive thymine (T) and adenine (A), 5'-3', **Figure 2a**) (Kwon et al., 2016). Although mariner transposons are limited to integrating into TA-sites, it has been shown that this restriction does not introduce a significant bias compared to non-mariner transposons like Tn5 (Kwon et al., 2016). Moreover, the preference for inserting in TA-sites is an advantage over Tn5, which is known to create several "hot spots" where most insertions occur, influencing the later analysis (Liu et al., 2013).

To compare which mutations are in which ratios present in a certain condition, Tn-seq involves a process in which the genomic flanks adjacent to the transposon are isolated, sequenced and mapped to the reference genome (**Figure 2b**). The places where reads map indicate where the mutations are and the relative numbers of reads give insight into the ratios in which the mutants are present (DeJesus et al., 2015). By comparing these relative abundances between different conditions, it is possible to investigate which genes are essential or important under that condition or for certain processes. The process of isolating these genomic flanks to send for sequencing (also called "library preparation") differ per type of transposon. In this study we used INSeq (transposon insertion sequencing), a variant of Tn-seq, which makes use of a mariner transposon from which the flanks are modified to contain MmeI recognition sites (Goodman et al., 2011). These sites are helpful for the library preparation. MmeI is a restriction enzyme that cuts ~20 bp downstream of its recognition sites, in this case in the genomic DNA next to the transposon (**Figure 2b**) (Skurnik et al., 2013). After removing the piece of transposon, this gives genomic fragments of ~16bp. Methods using other transposons, such as those employing BsmFI restriction, result in shorter fragments of 11 to 12 bp, which reduces the number of reads suitable for unambiguous mapping by half compared to 16 bp fragments (Kwon et al., 2016). Moreover, transposons without restriction sites at all which therefore require the shearing by sonification, give variable lengths of genomic flanks (Cain et al., 2020). This introduces bias during the library preparation and can damage the DNA with the change of miscalling bases (Giannoukos et al., 2018).

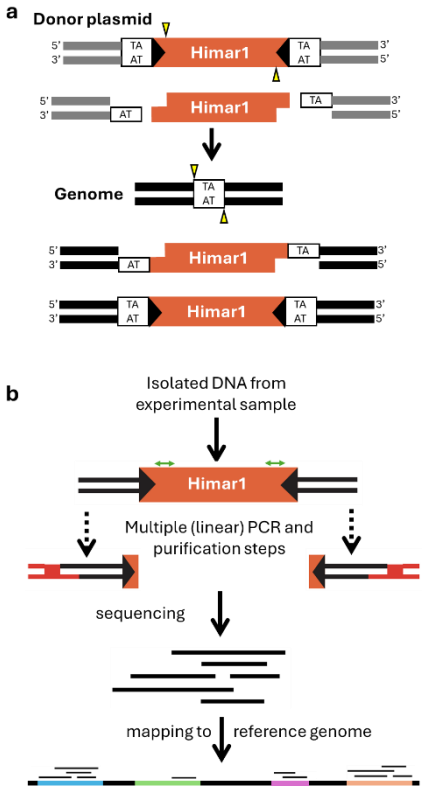


Figure 2. *Insertion of mariner transposon Himar1 from a donor plasmid into a TA-site in the genome via a cut and paste mechanism of transposition* ***a****. Overview of the different steps involved in INSeq, including the preparation of the library DNA for sequencing (also known as "library preparation")* ***b****. The genomic flanks next to the transposon are isolated with the help of multiple (linear) PCR and enzymatic steps. The flanks are mapped to the reference genome. Green arrows are MmeI recognition sites where MmeI can bind to cut ~20 bp downstream (adapted from Goodman et al., 2011).*

In this study we employed the powerful technique of INSeq to shed light on the mechanisms behind the recruitment of HAM to the plant (especially the phyllosphere) in response to Hpa infection from the bacterial perspective, as studies so far have focused mainly on the plant aspect of the interaction. To examine this, we investigated which genes in *Xanthomonas sp.* WCS2014-23 (a representative of the most abundant HAM ASV a0e1a) are important for its recruitment to the phyllosphere of

Arabidopsis during aboveground infection with Hpa. We did this in roughly two steps. First, an INSeq *Xanthomonas* mutant library was made in WCS2014-23 via transposon insertion mutagenesis. Subsequently, this library was be utilised in multiple mutant screens involving recruitment experiments to identify candidate recruitment genes. Shedding light on the genetic background of this specific interaction will provide new insights into beneficial plant-microbe interactions in general, but also contribute to the development of new (more) sustainable strategies to combat Hpa infection in agriculture.

Results

Creating a *Xanthomonas* INSeq mutant library

Identifying the right helper strain

To make the INSeq mutant library in *Xanthomonas sp.* WCS2014-23, we used transposon insertion mutagenesis via a triparental mating. This involves multiple conjugations in which a transposon donor plasmid is transferred from an *E. coli* host strain into the recipient strain, in this case *Xanthomonas*. This transfer is possible with the assistance of a helper plasmid from a second *E. coli* strain. Once in *Xanthomonas,* the transposon can integrate into the genome and potentially cause a mutation (**Figure 3**). To choose a suitable helper strain, we performed test matings using three different helper strains with the following helper plasmids: pRK2013, pRK2073 and pRK600. Furthermore, we did the matings with two different WCS2014-23 strains as recipient, the wildtype (XmR, resistant to rifampicin (Rif/R)) and an mCherry (mCh) fluorescent strain (XmRmCh, Rif and Tetracyclin (Tet) resistant). We determined the efficiency of these conjugations by CFU counts of serial dilutions of the conjugation mixes on selection plates with the corresponding antibiotics (**Figure 3**). No colonies were visible from any of the control matings, which were with only the helper and the recipient but without the donor pSAM_RI (data not shown). This indicates that the antibiotic resistance genes from the helper plasmids are not transferred to the recipient. The matings with helper plasmid pRK600 showed the highest efficiency (**Figure 3**). pRK2013 showed a similar efficiency as pRK600 when the recipient was XmRmCh, but not with XmR. pRK2073 showed a lower efficiency than pRK2013 and pRK600, but the efficiency was comparable between XmR and XmRmCh.

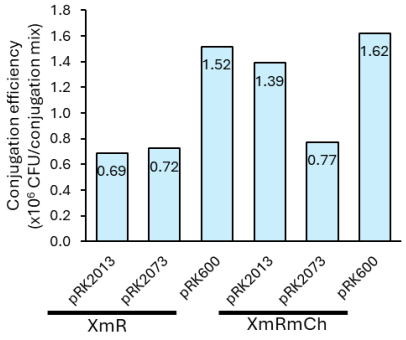


Figure 3. *Conjugation efficiencies matings with different helper strains (pRK2013, pRK2073 and pRK600). XmR and XmRCh are wildtype Xanthomonas (Xm) WCS2014-23 resistant to Rifampicin (R) and or fluorescent labelled with mCherry (mCh, also tetracyclin resistant)). No colonies were observed in control matings without pSAM_RI (data not shown). Values are based on CFU counts of one conjugation mix.*

However, for making a reliable transposon mutant library, besides a high conjugation efficiency, it is also important to check if the acquired resistance is really the cause of the integration of the transposon and not only integration of the kanamycin (Kan) resistance gene inside the transposon. Moreover, there is the possibility that the whole donor plasmid pSAM_RI (**Figure 4a**), which also contains the transposase gene, integrates into the genome of the recipient. When this happens, the transposon can be moved by the transposase inside the genome of the mutated *Xanthomonas* which makes it impossible to determine which mutation causes a certain phenotype. It is therefore important that this occurs as little as possible. To check for these two undesirable situations per helper strain, we performed multiple colony PCRs followed by gel electrophoresis with the transconjugants (recipients XmR or XmRCh from the conjugation mixes which acquired the kanamycin resistance and are therefore expected to have a mariner transposon integrated in their genome) (**Figure 4**). As a positive control, we did the same for pSAM_RI colonies and used colonies of the three different helpers (**Figure 4b**) or agar without bacterial growth (**Figure 4c-e**) as negative controls.

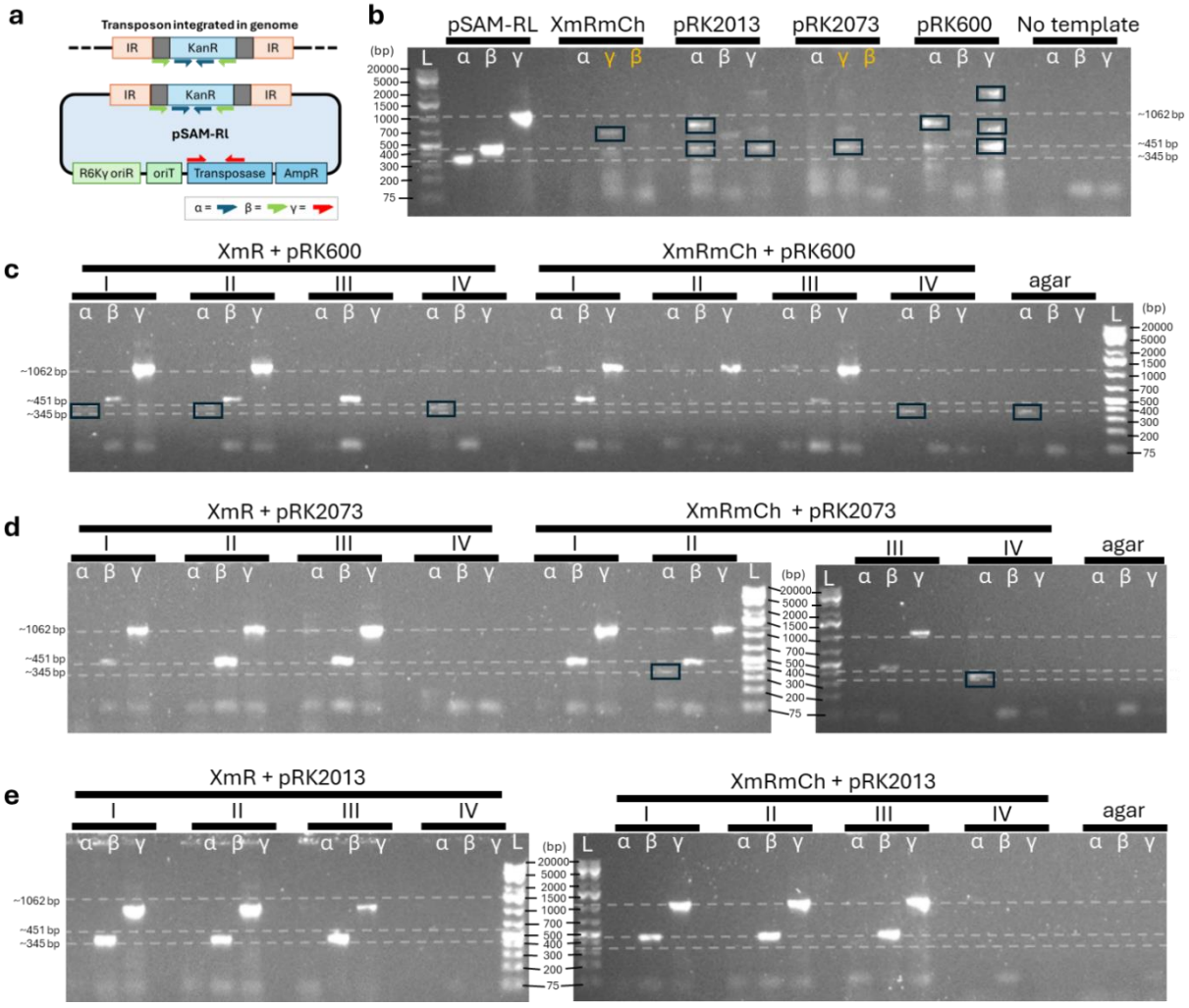


Figure 4. *Overview primer binding sites* ***(a)*** *and agarose gels with colony PCR products performed on pRK600, pRK2073 and pRK2013* ***(b)*** *and transconjugants (Xanthomonas XmR and XmRmCh) from test conjugations with the three helper strains* ***(c-e)****. α, β and γ indicate the used primer sets which amplifying respectively a part of the transposase gene (α, 345 bp), a part of the Kan-resistance gene (β, 451 bp) and a part of the transposon around the Kan-resistance gene (γ, 1062 bp). Lanes α, β and γ contain corresponding PCR products. Orange letters indicate when lanes were swapped compared to used pattern in the rest of the figure. Roman numbering indicates the size of the touched colony (I, II or III are large, IV is tiny). Lanes indicated with "agar" are agar without bacterial growth. Blue rectangles mark unexpected bands. "KanR" or "R" is kanamycin resistance gene.*

We used three different primer sets on four colonies per test conjugation. The first set amplifies a fragment of 345 bp of the transposase gene (α in **Figure 4a**) and gives an indication for the presence of this gene. As expected, there is a band visible in the lane of pSAM_Rl (**Figure 4b**). In most cases there is no band visible around this size in the transconjugants. However, there are exceptions, such as with the transconjugants mutated using helper plasmid pRK600 are bands present (3 and 2 in XmR and XmRmCh respectively) (**Figure 4c**). While for pRK2073 this is the case for two colonies (both of XmRmCh, **Figure 4d**) and for pRK2013 none (**Figure 4e**). The next set of primers amplifies a part of the kanamycin resistance gene which is present inside the transposon and gives a fragment of 451 bp (β in **Figure 4a**). Indeed, a band is visible in pSAM_Rl and most transconjugants, however, in a few colonies of XmRmCh transconjugants made using pRK600 they were missing (**Figure 4c**), while this was not the case for XmR. The last set of primers amplifies a part of the kanamycin resistance gene inside the transposon (γ in **Figure 4a**) and serve as an indicator for an inserted transposon. As expected, these gave a band from pSAM_Rl as well as from all colonies of the transconjugants except one colony of XmR mutated using pRK600 (**Figure 4c**). With one exception for primer set α (**Figure 4c**), none of the three different primer pairs showed amplification when only agar was touched. Interestingly, mainly

in pRK2013 and pRK600, bands were visible in the lanes of the primer sets that amplify the transposase (α) and transposon (γ) (**Figure 4b**). Moreover, they were not the expected size. This may be the result of non-specific binding of the primers to (partially) matching binding sites. Bands below 75 bp are probably the effect of non-specific binding of primers with each other, since these are also visible when no template is present. Finally, the fact that the fourth (IV) colonies of each group were more likely than other colonies to give no band or a faint band is probably due to the fact that these were very tiny colonies. Therefore, they can be more easily missed by picking or are completely used up after the first touch with the toothpick.

Based on these results, although pRK600 gave the highest conjugation efficiency in the test matings, we excluded pRK600 as candidate helper strain. The fact that both XmR and XmRCh transconjugants obtained by a mating using this helper contained the transposase gene made it too risk full for generating a library. Furthermore, however pRK2013 showed the second highest conjugation efficiency, we choose to make the actual library using pRK2073. Even though it was not observed in the controls of our test matings, it is known that pRK2013 has the risk for transferring its Kan-resistance without the insertion of the transposon into the genome of the recipient. As a result, wildtype bacteria could remain in the library that do not have a mutation but do have the Kan-resistance. This is possible because pRK2013 both contains a Kan-resistance gene and is able to transfer itself into the recipient since it is a conjugative plasmid. As the recipient, we chose XmR because colonies of transconjugants from this strain, unlike some from XmRCh, did not contain the fragment of the transposase gene when the mating was performed using pRK2073.

Optimising the conjugation efficiency

Now we had chosen the best suitable helper and recipient strain, we started to optimise the conjugation efficiency in order to get enough transconjugants to proceed in the protocol for generating the mutant library. The test matings were performed following a protocol which was used by two other studies for generating mutant libraries in *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) and *Cupriavidus* isolate Cv1 (Wheatley et al., 2017; Ford et al., 2024 unpublished data). The test mating with pRK2073 described above had an efficiency of 724,500 CFU per conjugation mix and repeating the conjugation with multiple biological replicates gave a more accurate estimation of approximately ~645,000 CFU/conjugation mix. A total of 4.5 million transconjugant is needed for the mutant library.

The first way in which we tried to generate more transconjugants was to scale up the number of mating plaques while following the same protocol as from the previous studies just mentioned. Combining all 20 matings gave enough transconjugants in total based on CFU counts. However, we observed a high variability in efficiencies, varying from 130,500 to 1.07 million CFU/conjugation mix (**Figure 5**). This poses a risk that the actual number of mutants used in the next step of the protocol could be lower, potentially resulting in a less diverse library. Therefore, we investigated if we could improve the efficiency by adjusting the protocol. We started with adjusting the OD to which the recipient is grown prior to the mating. So far, we used overnight growth according to the protocol, which for WCS2014-23 means a final OD of around 0.2-0.3 because of its slower growth rate than *Cupriavidus* and *Rhizobium.* However, we found that a previous study which made mutant libraries in two other *Xanthomonas* strains, *X. oryzae* pv*. oryzae* and *X. campestris* pv*. Campestris*, grew those recipient strains to the mid-log phase (OD 0.5) (Sun et al., 2003). WCS2014-23 grown to mid-log phase led to a 2.3 times increase in efficiency with ~1.5 million CFU/conjugation mix on average. Nevertheless, a huge variability in efficiency was still observed (0.5 to 2.1 million CFU/conjugation mix) (**Figure 5**).

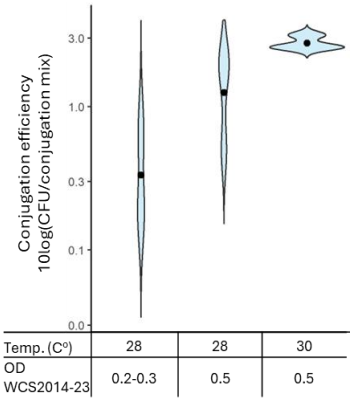


Figure 5. *Effect of protocol adjustments (conjugation temperature (Temp.) and OD of recipient (WCS2014-23)) on the conjugation efficiency. Dots*

Another option we looked into was the temperature at which the mating was performed, since it has been shown that this can influence the transposon insertion efficiency (Champie et al., 2023). For example, it has been shown that increasing the conjugation temperature from 28°C to 30°C for mutating *Pseudomonas simiae* WCS417 led to a doubling in conjugation efficiency (Goossens et al. unpublished data). Therefore, we performed a new set of matings at 30°C instead of 28°C but still with the recipient grown to mid-log phase prior to mating, since this showed an improvement in the previous test. This resulted in nearly double the conjugation efficiency on average compared to mating at 28°C (~2.8 million versus ~1.5 million CFU per conjugation mix, respectively) (**Figure 5**). This means that around two conjugation mixtures would be enough for the 4.5 million mutants needed to proceed in the protocol. Moreover, the efficiency was much more consistent, varying between 2.6 to 3.1 million CFU/conjugation mix (compared to 0.5-2.1 million CFU/conjugation mix at 28°C). Because of this higher consistency and efficiency, we decided to use these conjugation mixes for making the mutant library.

Library characterisation

To analyse the diversity of our library, we mapped the isolated genomic flanks next to the transposons to the reference genome of *Xanthomonas* sp. WCS2014-23. The sequencing yield was around the expected amount of around 20 million reads per sample (**Table 1**). Around 99.1% of the reads contained the mariner transposon (Himar1C9) end sequence and are therefore considered as 'true' TA-site-disrupted reads. This indicates the absence of the integrated donor vector pSAM_Rl in the mutant library. This is supported by the fact that mapping of the reads to the sequence of pSAM_Rl aligns only 6 to 44 reads, which is classified as an overall alignment rate of 0%. Around 64.9% of these reads mapped successfully to the reference genome (allowing for one mismatch). However, we noticed variation in TA hits. TA hits are the number of TA-sites in the reference genome that have at least one read mapping to them. As mentioned before, TA-sites are the places where it is possible for mariner transposons to insert into the genome (85,676 TA-sites in the case of WCS2014-23). The sample from the original library freezer stock (Xm) had a sequencing yield of 25 million and showed ~9,000 TA hits. However, one of the technical replicates of the input sample of the soil experiment (IPo2) gave 22% more TA hits (11,000) while having a lower sequencing yield of 18 million reads. Moreover, the second technical replicate of this input sample (IPo1) had only ~4,000 TA hits despite having a comparable sequencing yield to IPo2. Input samples are cultures grown out of a library freezer stock sample and are used for inoculating the plants in the mutant screens. The second time culturing gives a slight reduction in diversity, but they should not differ much from the freezer library stock. Therefore, it is unlikely that one of them (IPo2) would give a higher number of TA hits than the original library freezer stock.

Table 1. *Transposon-insertion sequencing (Tn-Seq) statistics of library freezer stock and input samples separately or combined*

Sample ‡	Sequencing yield †	No. of Tn- end containing reads	Read count (TA sites only)	TA hits	Insertion density (%)	Mean read count over non-zero TA
Xm	24,383,946	24,092,317	18,764,423	9,523	11.1	1,970.40
IPo1	18,836,264	18,711,114	9,615,780	3,932	4.6	2,445.50
IPo2	17,390,330	17,284,512	8,880,557	11,063	12.9	802.70
IPo1+2	36,226,594	35,995,626	18,496,271	13,342	15.6	1,386.30
IPo1+2+Xm	60,610,540	60,087,943	37,260,846	19,164	22.4	1,944.30
IP+Xm+Exp	392,105,888	388665260	249824349	49,691	58.0	5027.6

‡ IPo1 and 2 are input samples of the soil experiment. IP input samples of the soil and spray experiment. Xm is a freezer stock library used as inoculation. Exp are experimental samples (derived from mock and gnoHpa inoculated plants).
† Sequencing yield is total number of reads from Illumina paired-end sequencing (forward plus reverse).

Besides the difference in number of TA hits, we also noticed that there was a difference in which TA-sites were covered. Therefore, we combined the reads from the input samples with each other and or with the freezer library sample to get a more representative set of TA sites. We indeed saw that this combination increased the number of TA hits and so the insertion density (percentage of TA-sites that have insertions). We did the same for the experimental samples of the mutant screens (see **Table 2** in screening for recruitment genes below). Remarkably, for both mutant screens, one of the combined experimental samples showed a higher insertion density than the combined input samples. This is quite unlikely, since the harvested mutants from our in vivo plant experiments have gone through multiple experimental bottlenecks, each of which can be expected to cause a reduction in diversity compared to the input sample. Examples of such bottlenecks vary from possible technical bottlenecks during the inoculation to the colonisation and survival of the bacteria on the shoots. Therefore, this suggests that there is another bottleneck somewhere between the harvested bacteria (input or experimental) and the sequencing results. This could be somewhere in the library preparation (the isolation of the genomic flanks next to the inserted transposons), the enzymatic adapter ligation used by the sequencing company or maybe the sequencing method (Illumina) itself.

For a meaningful statistical analysis it is recommended that a transposon mutant library has insertions in at least ~35% of the TA sites (*TPP Overview — TRANSIT v1.1.2 Documentation*, n.d.). Based on the combined sample of the freezer stock sample and both the input samples (IPo1+2+Xm), the library seemed to have a saturation in TA-sites of ~22.4%. Nevertheless, we hypothesised that the actual saturation level was higher than initially estimated. This is because combining different samples resulted in an increased insertion density (**Table 1**), along with the higher insertion density observed in the experimental samples. We therefore estimated the library diversity based on the mapping of the combined reads of all available reads, so freezer stock, input and experimental samples. Mapping of all these reads together gave indeed an insertion density of 58%, considerably higher than the ~22.4% mentioned previously.

The number of different genes with insertions can be estimated by selecting for genes with a minimal mean of mapped reads per TA site. For regular data sets, a minimum average of 100 reads per gene is recommended (*Analysis Methods — TRANSIT v2.0.0 Documentation*, n.d.). Based on this criterium, we found that ~50.6% (2229 of the 4399 CDSs) of the total genes present in the reference genome contained insertions. However, this is based on the pooling of highly variable samples with a large number of reads in total, which probably pulled the average down. Therefore, we also did the same for a minimum mean of 50 reads, which gives an estimation 72.2% of the genes. However, it is important to note that insertion density provides a more accurate estimate of library diversity, as the number of distinct mutants is equivalent to the number of unique TA sites where insertions occur. Finally, we examined the distribution of transposon insertions by plotting the read counts for the different TA sites. The insertion distribution was quite evenly spread over the genome, however, in only a few TA-sites located between 395,000-397,000 bp extreme high read counts are observed (**Figure 6a**). However, when zoomed into the bottom part of figure 6a, it was apparent that the other insertions have expected read counts around 100 or sometimes higher (**Figure 6b**).

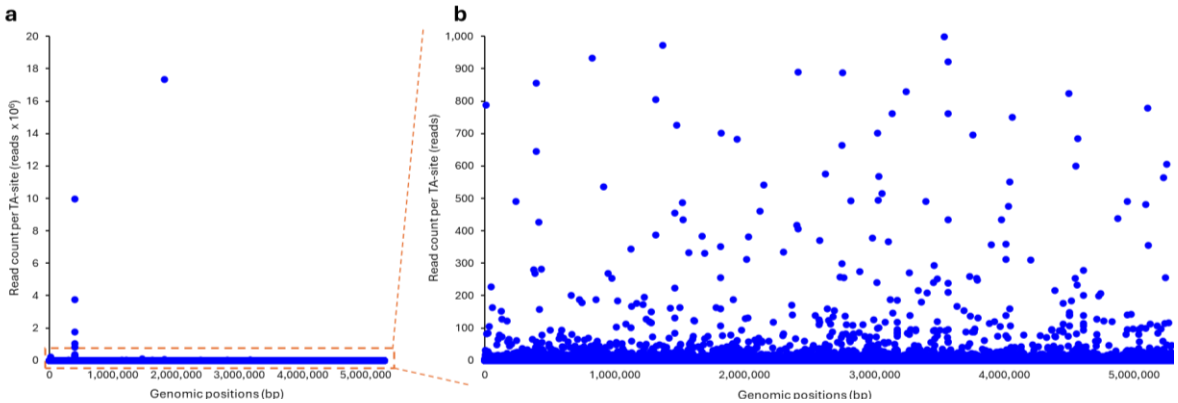


Figure 6. *Insertion distribution in TA-sites of mariner transposons across genome. X-axis shows the genomic position in base pairs (bp) based on arranging the 46 contigs of the reference genome in order from largest to smallest.* ***a,*** *shows the overview inclusive outliers,* ***b*** *is zoomed in on the bottom part of* ***a*** *with a max read count of 1,000 reads per TA-site. Each dot is the read count in number of reads per TA-site based on the mapping of the combined reads of two input samples (of the soil experiment, IPo1 and 2) and one freezer library sample.*

For analysing the essentiality of the genes in the genome of WCS2014-23, we used a Hidden Markov Model (HMM) (DeJesus et al., 2015). This algorithm uses statistics from the mapping of the genomic flanks to the reference genome to perform a four-state classification which categorises the genes. The first category are essential genes (ES) for survival. The second are non-essential (NE) genes which allow mutations. The last two categories are describing the mutant phenotypes in which the disruption of the gene causes a growth-defect (GD) or growth-advantage (GA). Unfortunately, the available reference genome of WCS2014-23 in the NCBI (National Center for Biotechnology Information) database consists of 46 contigs, which made the analysis with HMM difficult. Nevertheless, we solved this by combining the contigs in order from largest to smallest to mimic a single chromosome. Furthermore, it was re-annotated which gave 4399 coding sequences (CDS), from which 4 rRNAs, 53 tRNAs and 1 tmRNA (transfer messenger RNA). The earlier described skewness in the insertion distribution could affect the analysis of HMM, however we found a suitable normalisation method, Betageom, which was able to correct for this. Betageom fits the data to an "ideal" geometric distribution with a variable probability parameter p. (*Normalization — TRANSIT v3.3.3 Documentation*, n.d.). Using the HMM, we first identified 204-384 ES, 76-101 GD, 3885-4035 NE and ~3 GA genes in the initial library we made. From 51 genes classification was impossible since these do not contain TA-sites. 674 genes were classified with low-confidence or ambiguous. Nevertheless, it is good to keep in mind that these are ranges due to the lower insertion densities.

Screening for recruitment genes

For investigating which genes are important for the recruitment of *Xanthomonas*, we designed a set-up based on the soil-borne legacy (SBL) experimental method (Bakker et al., 2018). It is comparable to a one-generation SBL experiment, however, for the first screen we sowed sterilised *Arabidopsis* seeds on imitation Hpa-conditioned natural soil which was made by mixing *Xanthomonas* through natural Reijerscamp soil (10^7 CFU/g) (**Figure 7**). This is 10 times higher than the abundance of indigenous wildtype *Xanthomonas* in natural soil, previously estimated from sequencing data, in order to outnumber these (Unpublished data, Spooren et al., 2024). Moreover, the plants were infected with gnoHpa, which is Hpa without the presence of any other microorganisms (Goossens et al., 2023). To summarize: After two weeks of growth, half of the wildtype *Arabidopsis* plants were inoculated with gnoHpa, and recruitment was assessed one week after the gnoHpa inoculation **Figure 7**. The advantage of this non-sterile, *in vivo* plant system is that it is closer to natural conditions compared to *in vitro* plant systems. A possible disadvantage could be that the mixing through the soil could create a bottleneck effect to reduce mutant diversity between in the soil and in the phyllosphere. Therefore, we incorporated a second set-up in which *Xanthomonas* was sprayed directly onto the shoots just before gnoHpa inoculation ($3*10^8$ CFU/pot, $<10^7$ CFU/plant) (**Figure 7**). In this way we started with a maximal diversity in the phyllosphere and gave all mutants the opportunity to be recruited there in response to gnoHpa.

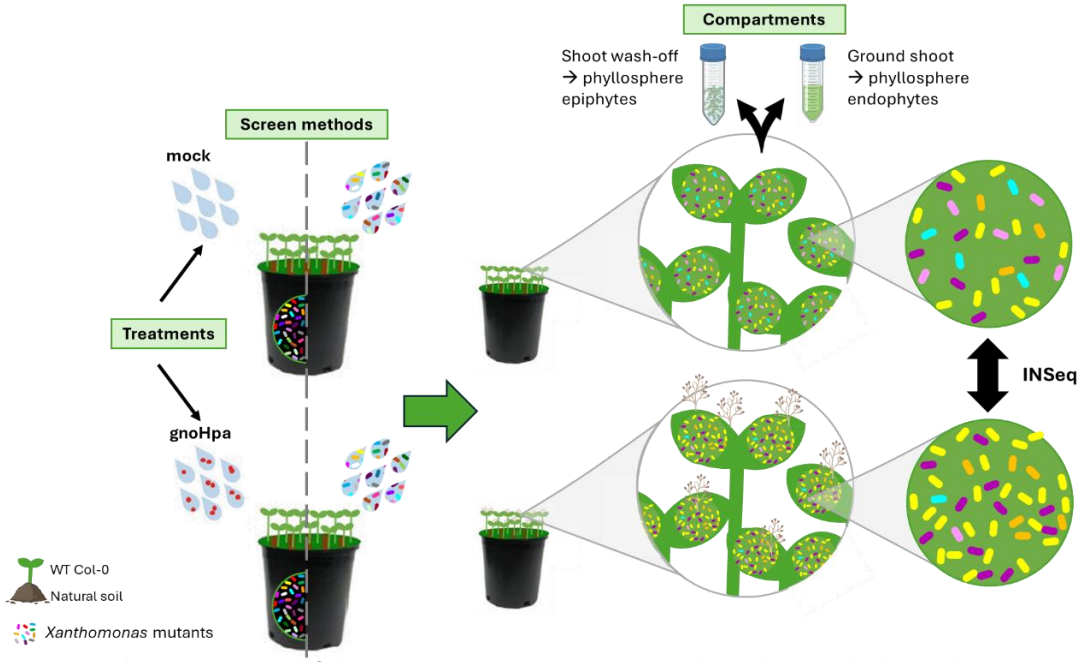


Figure 7. *Schematic overview of the mutant screens using spray- or soil-inoculation with the Xanthomonas mutants. Both are based on a one-generation soil borne legacy experiment in which half of the plants are inoculated with gnoHpa two weeks after sowing and harvested one week after gnoHpa(or mock)-inoculation. (Adapted from Vismans et al. (2021) and "downy mildew," by BioRender.com (2024). Retrieved from https://app.biorender.com/biorender-templates).*

System validation using wildtype Xanthomonas

To prove that the recruitment of *Xanthomonas* in response to gnoHpa infection can be determined in our systems described previously, we first performed the assays with the wildtype (WT) WCS2014-23. We did this by measuring the population densities by CFU counts from serial diluted wash-offs of one-week gnoHpa-infected and mock-inoculated plants. Around 15 and 16 times more *Xanthomonas* was present in the phyllosphere of gnoHpa-infected plants compared to uninfected plants following soil- or spray-inoculation, respectively (**Figure 8**). Moreover, we used these abundances to calculate the minimum required number of plants needed to recover sufficient bacteria for the mutant screens, ensuring that the statistics involving INSeq are valid. The maximum number of possible mutants is approximately 85,000 since this is the number of TA-sites in the reference genome of *Xanthomonas.* These sites are the places where a mariner transposon can insert into the DNA. The rule of thumb is that we need 100 bacteria per mutant, so in total 8.5 million bacteria are needed per treatment. Based on the CFU counts, that would correspond to a shoot fresh weight (SFW) of 0.615 grams, which corresponds to the feasible number of 4 to 7 pots with ~30 plants per pot. However, the amount of recovered mutants can also be affected by the number of colonisation events per plant, which is unknown for *Xanthomonas.* Therefore, we scaled the group sizes up to around 22 pots per treatment group per replicate (so ~44 pots total per biological replicate).

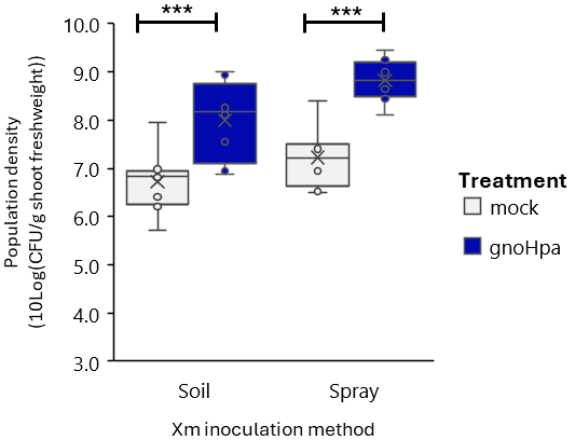


Figure 8. *Recruitment of wildtype Xanthomonas (Xm) sp. WCS2014-23 to the phyllosphere in response to gnoHpa infection. Boxplots showing WCS2014-23 population densities in the phyllosphere in 10-log transformed CFU/g shoot fresh weight. Asterisks indicate significance level in one-sided Student's t-test: *** is P<0.001. Crosses (x) are averages and dots represent biological replicates (n=8).*

Recruiting the mutants

We performed the mutant screens as described previously, but decided to harvest the phyllosphere in two different compartments. Besides making shoot wash-offs (which harvest the phyllosphere epiphytes), the washed shoots were also ground to harvest the bacteria inside the shoot tissue, the phyllosphere endophytes, as well (**Figure 7**). In this way, we made it possible to compare the differences in mutant populations in the separate compartments of the phyllosphere. For both mutant screens (soil- and spray-inoculation), we observed the expected significantly higher population density of *Xanthomonas* mutants in the phyllosphere epiphytes of gnoHpa-infected plants (**Figure 9a**), as demonstrated in the system validation (**Figure 8**). For the phyllosphere endophytes we only saw a higher abundance when mutants were soil-inoculated. However, it has not earlier been investigated if the population density of HAM among phyllosphere endophytes also increases in response to (gno)Hpa infection as they do among the epiphytes. We also determined the amount of harvested bacteria, which was above the minimum of 8.5 million bacteria needed for valid INSeq statistics in all three biological replicates for both compartments in both mutant screens (**Figure 9b**). Finally, no colonies were visible from any of the dilutions of samples derived from plants without *Xanthomonas* inoculation (data not shown). This confirms that there was no cross-contamination between samples and that there were no other rifampicin and kanamycin resistant fast-growing bacteria present in the systems.

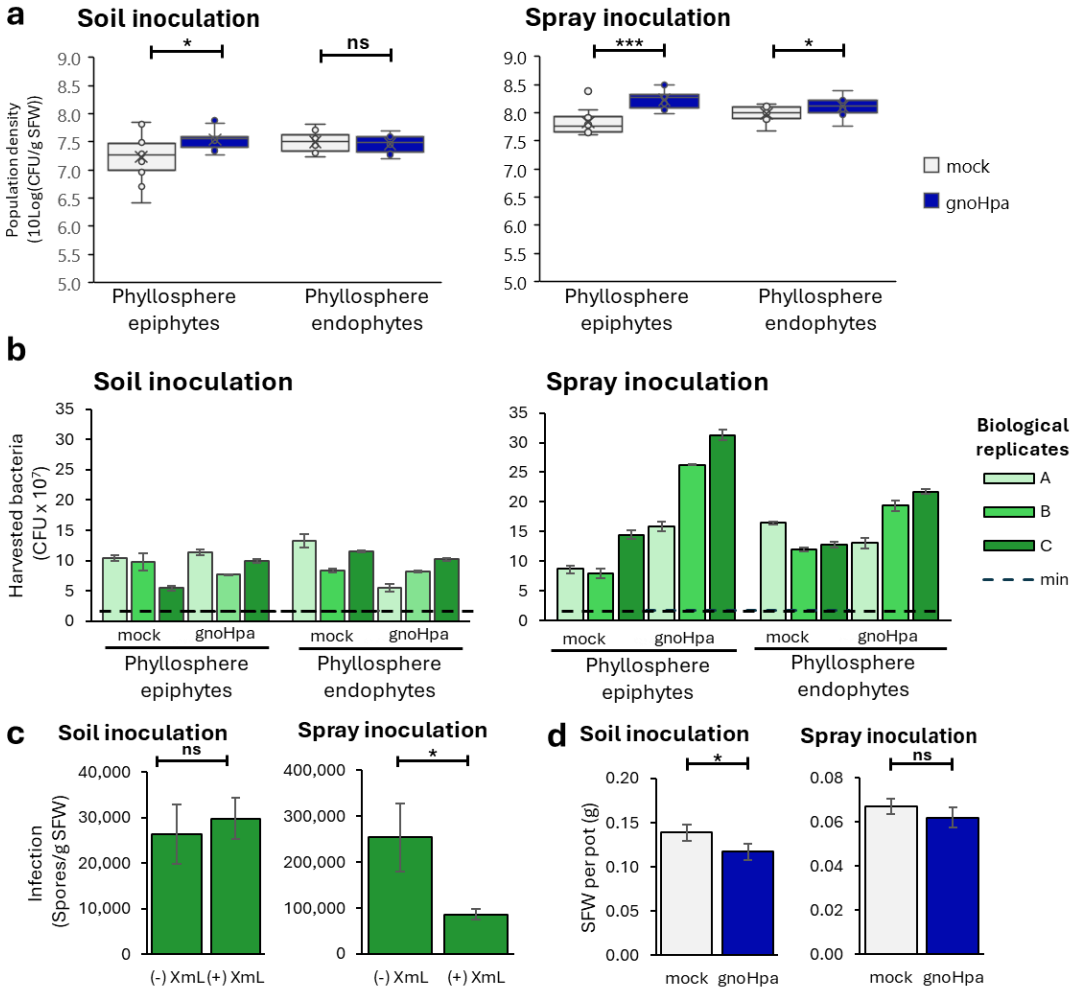


Figure 9. *Boxplots of the phyllosphere Xanthomonas mutants population densities in 10-log transformed CFU/g shoot fresh weight(SFW)* ***(a),*** *total harvested bacteria per biological replicate* ***(b),*** *average Hpa spore production* ***(c)*** *and average shoot freshweights per pot* ***(d)*** *of mock- and gnoHpa-inoculated Arabidopsis Col-0 when Xanthomonas mutants were soil- or spray-inoculated. All error bars represent Standard Errors of 8 biological replicates except for* ***b*** *(2-3 technical replicates). Asterisks indicate significance level in one-sided Student's t-test: * is P<0.05, *** is P<0.001. Crosses (x) are averages and dots represent biological replicates. Dashed line "min" in* ***b*** *is the minimum needed amount of 8.5 million*

Furthermore, we checked how well the plants were infected in the two different screens. For plants not inoculated with mutants, we saw in the experiment with spray-inoculation a 10 times stronger infection then for the soil-inoculation experiment (250,000 versus 25,000 spores per gram SFW respectively, **Figure 9c**). Moreover, we also compared the infections between plants inoculated with or without the *Xanthomonas* mutants (**Figure 9c**). However, it has not previously been proven that inoculation with WCS2014-23 alone can reduce the infection by Hpa in *Arabidopsis* or promote plant growth (Berendsen et al., 2018). In the spray-inoculation experiment, we saw significantly less infection for plants inoculated with *Xanthomonas* mutants compared to the plants without. However, this was not the case in the soil-inoculation experiment. Additionally, we compared the SFW since this is also an indicator for infection severity due to the growth-defence trade-off (**Figure 9d**). In the soil-inoculation experiment the SFW of gnoHpa-infected plants was indeed significantly lower than in mock-treated plants, but in the spray-inoculation experiment this was not the case. Nevertheless, this study was also with *Xanthomonas* in soil, so it was remarkable that spray-inoculation here with *Xanthomonas* mutants seemed to lead to a reduced infection while the soil-inoculation did not. Combined with the fact that an overall mutant library population behaves like the wild type because most mutations occur in non-essential (NE) genes (Voogdt et al., 2024), and that mutants in NE genes exhibit wildtype fitness levels (van Opijnen et al., 2009) (van Opijnen et al., 2009). Moreover, it is noteworthy that when the spores per gram SFW were significantly lower (when spray inoculated with *Xanthomonas*) the growth trade-off is not observed (**Figure 9c-d**).

Since we observed the significant higher abundance of Xanthomonas mutants in the phyllosphere epiphytes of gnoHpa-infected plants and harvested up to ten times more bacteria than the calculated minimum in each biological replicate, DNA of the experimental samples was sent for sequencing after INSeq DNA library preparation. Unfortunately, as mentioned before in the section library characterisation, we observed a variability in TA hits among samples. Therefore, to have a high enough insertion density for the HMM statistics, we had to pool the reads of three biological replicates per treatment group into one biological replicate (**Table 2**). Since the insertion densities of the combined samples were still below the recommended 35% (from 4 up to 26%), this limited the number of genes for which we could made a prediction, but there were enough to perform the four-state classification per condition.

Table 2. *Transposon-insertion sequencing (Tn-Seq) statistics of experimental samples.*

Sample ǂ	Sequencing yield †	No. of Tn- end containing reads	Read count (TA sites only)	TA hits	Insertion density (%)	Mean read count over non-zero TA
IPp1+2+Xm	34,763,188	34,535,517	16,868,355	12,882	15	1,310
SWp135M	44,083,612	43,820,357	21,314,200	21,941	26	971
SWp246H	37,434,806	37,209,949	20,575,979	10,325	12	1,993
IPo1+2+Xm	60,610,540	60,087,943	37,260,846	19,164	22	1,944
SoW135M	61,927,712	61,082,980	51,253,527	20,368	24	2,516
SoW246H	57,977,550	57,453,583	37,652,648	7,418	09	5,076
SoG135M	52,366,360	51,807,465	41,245,954	9,329	11	4,421
SoG246H	42,942,120	42,667,466	23,652,969	3,477	4	6,803

ǂ *IP are input samples. Xm is a freezer stock library used for inoculation input samples. Experimental samples, shoot (S) wash-off (W) or ground (G) shoot of the experiment with soil (o) or spray (p) inoculation. The numbers indicate the biological replicates from which the odd have been mock treated (M) and the even gnoHpa inoculated (H).* ***†*** *Sequencing yield is total number of reads from Illumina paired-end sequencing (Forward plus reverse).*

Recruitment genes were selected by filtering for genes classified non-essential (NE) in the input sample and in the presence in the phyllosphere of mock-treated plants, but became essential (ES) or gave a growth-defect (GD) when mutated in the phyllosphere of gnoHpa-infected plants. This indicates that a gene is not necessary for its survival neither in culture conditions nor for its presence in the phyllosphere, but that it is required for its presence or recruitment on gnoHpa-infected plants. Moreover, we selected for genes classified with a confidence score of 0.75 or higher (categorised as high by Gollapalli et al., (2022)) and an insertion density of at least 33% (in the middle of the range recommended by the HMM documentation) (*HMM — TRANSIT v3.3.3 Documentation*, n.d.). We identified 63 genes, 28 of which are already annotated while the remainder are predicted as hypothetical proteins (**Figure 10**, **Table 3-5**). Unfortunately, annotating all hypothetical proteins was not feasible within the time constraints of this Master's project. Therefore, we will focus on the annotated genes and highlight the most promising ones based on literature research. First the genes identified in the phyllosphere epiphytes, followed by those identified in the phyllosphere endophytes and the overlapping genes.

For the phyllosphere epiphytes we have examined the most promising candidate genes which are solely found in this group. To begin with, *clp,* coding for CRP-like protein, Clp (**Table 3**). It can interact with as well DNA, as a transcription factor, as with second messenger molecules (K. Xu et al., 2021). For *Lysobacter enzymogenes* (a gram-negative bacterium used as biocontrol agent against fungal and oomycete plant-pathogens) there is evidence that Clp regulates an anti-fungal attack. This includes the activation of the production of lytic enzymes combined with biofilm formation around hyphae. It was found that lytic enzymes involved in the anti-fungal attack are, for example, chitinase, β-1,3-glucanase and proteases, which can disrupt the fungal cell wall. However, despite oomycetes not containing significant chitin in their cell walls, it is expected that the other lytic enzymes such as glucanases are also involved in anti-oomycete attack, since oomycete cell walls also contain glucans (*Oomycetes - an Overview | ScienceDirect Topics*, n.d.). For the biofilm formation it was found that Clp regulates this by affecting the twitching motility of *L. enzymogenes* towards fungal hyphae. Moreover, for two pathogenic *Xanthomonas* species, *X. axonopodis* pv. citri and *X. campestris* pv. Campestris, it has been shown that Clp interacts with di-GMP, a second messenger in gram-negative bacteria (K. Xu et al., 2021). Combining this information, our hypothesis is therefore that Clp could have a similar function in *Xanthomonas sp.* WCS2014-23, which could potentially lead to the suppression of Hpa. This happens possibly by disrupting the cell walls via the production of lytic enzymes, and maybe also via biofilm formation around Hpa hyphae. This would be supported by our finding of a gene for a lytic enzyme, *bprV_1*, coding for an Extracellular basis protease (**Table 3**). In addition, in our screen we also discovered two genes which are linked to motility and chemotaxis which potentially could contribute to the biofilm formation process of an attack (The UniProt Consortium, 2023). These are *hrpB* coding for an ATP-dependent RNA helicase HrpB (Granato et al., 2016) and *cheB_2*, coding for Protein-glutamate methylesterase (**Table 3**). Moreover, the chemotaxis associated gene could also support a hypothesis in which *Xanthomonas* senses a chemoattractant from a Hpa-infected plant (or Hpa itself) which causes its recruitment to the plant or the site of infection.

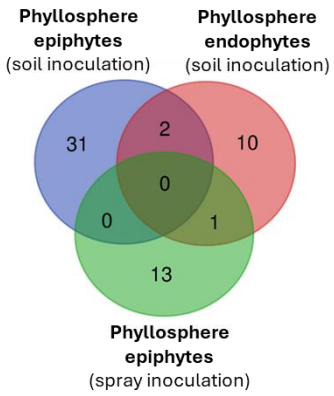


Figure 10. *Venn diagram of identified genes in mutants screens were Xanthomonas mutants were inoculated by mixing trough the soil of by spraying on the shoots.*

Table 3. *Candidate recruitment genes for phyllosphere epiphytes*

Phyllosphere epiphytes			
Based on soil-inoculation			
ORF	**Gene**	**Annotation**	**Classification ǂ**
XANT_00120	**ccmE_1**	Cytochrome c-type biogenesis protein CcmE	GD
XANT_00372	**rhaR_1**	HTH-type transcriptional activator RhaR	ES
XANT_00379	**asnB_2**	Asparagine synthetase B [glutamine-hydrolysing]	ES
XANT_00482	**-**	hypothetical protein	ES
XANT_01241	**-**	hypothetical protein	GD
XANT_01260	**bepF_2**	Efflux pump periplasmic linker BepF	ES
XANT_01344	**acdA_2**	Acyl-CoA dehydrogenase	GD
XANT_01372	**-**	hypothetical protein	ES
XANT_01650	**-**	hypothetical protein	ES
XANT_01835	**clp**	CRP-like protein Clp	ES
XANT_01972	**iolS_1**	Aldo-keto reductase IolS	ES
XANT_02058	**-**	hypothetical protein	ES
XANT_02075	**cheB_2**	Protein-glutamate methylesterase/protein-glutamine glutaminase	ES
XANT_02611	**-**	hypothetical protein	ES
XANT_02673	**-**	hypothetical protein	ES
XANT_02771	**-**	hypothetical protein	GD
XANT_02930	**-**	hypothetical protein	ES
XANT_03034	**-**	hypothetical protein	ES
XANT_03191	**btsR_2**	Transcriptional regulatory protein BtsR	ES
XANT_03271	**glaR**	HTH-type transcriptional repressor GlaR	GD
XANT_03322	**bprV_1**	Extracellular basic protease	GD
XANT_03370	**rsuA_2**	Ribosomal small subunit pseudouridine synthase A	GD
XANT_03534	**hchA**	Protein/nucleic acid deglycase HchA	ES
XANT_03607	**mprA_3**	Response regulator MprA	ES
XANT_03614	**ssuC_2**	Putative aliphatic sulfonates transport permease protein SsuC	GD
XANT_03658	**metF**	5%2C10-methylenetetrahydrofolate reductase	ES
XANT_03730	**-**	hypothetical protein	ES
XANT_03897	**-**	hypothetical protein	ES
XANT_03950	**-**	hypothetical protein	ES
XANT_03954	**-**	hypothetical protein	ES
XANT_04208	**-**	hypothetical protein	ES
XANT_04229	**-**	hypothetical protein	ES
XANT_04308	**-**	hypothetical protein	GD
Based on spray inoculation			
ORF	**Gene**	**Annotation**	**Classification ǂ**
XANT_00311	**hrcA**	Heat-inducible transcription repressor HrcA	ES
XANT_00490	**-**	hypothetical protein	ES

XANT_00570	**mscS**	Small-conductance mechanosensitive channel	ES
XANT_00602	**ugpC**	sn-glycerol-3-phosphate import ATP-binding protein UgpC	ES
XANT_00659	**-**	hypothetical protein	ES
XANT_01239	**thiG**	Thiazole synthase	ES
XANT_01317	**hrpB**	ATP-dependent RNA helicase HrpB	ES
XANT_01337	**-**	hypothetical protein	ES
XANT_01454	**-**	hypothetical protein	ES
XANT_01950	**-**	hypothetical protein	ES
XANT_02071	**-**	hypothetical protein	ES
XANT_02867	**-**	hypothetical protein	ES
XANT_03044	**adeP**	Adenine permease AdeP	ES
XANT_03893	**-**	hypothetical protein	ES

ǂ ES is essential, GD is growth-defect. Table is sorted by order of ORF (open reading frame) number.

Table 4. *Candidate recruitment genes for phyllosphere endophytes*

Phyllosphere endophytes			
Based on soil-inoculation			
ORF	**Gene**	**Annotation**	**Classification ǂ**
XANT_00311	**hrcA**	Heat-inducible transcription repressor HrcA	ES
XANT_00326	**fabG_2**	3-oxoacyl-[acyl-carrier-protein] reductase FabG	ES
XANT_00638	**-**	Putative pterin-4-alpha-carbinolamine dehydratase	ES
XANT_01286	**-**	hypothetical protein	ES
XANT_01650	**-**	hypothetical protein	ES
XANT_01898	**aroG**	Phospho-2-dehydro-3-deoxyheptonate aldolase%2C Phe-sensitive	GD
XANT_02227	**-**	hypothetical protein	ES
XANT_02485	**-**	hypothetical protein	ES
XANT_02848	**-**	hypothetical protein	ES
XANT_03244	**-**	hypothetical protein	ES
XANT_03252	**-**	hypothetical protein	ES
XANT_03534	**hchA**	Protein/nucleic acid deglycase HchA	ES
XANT_04408	**-**	hypothetical protein	ES

ǂ ES is essential, GD is growth-defect. Table is sorted by order of ORF (open reading frame) number.

Table 5. Candidate recruitment overlapping in phyllosphere epi- and endophytes

Epi- & Endo-phyllosphere			
ORF	**Gene**	**Annotation**	**Classification ǂ**
XANT_00311	**hrcA**	Heat-inducible transcription repressor HrcA	ES
XANT_01650	**-**	hypothetical protein	ES
XANT_03534	**hchA**	Protein/nucleic acid deglycase HchA	ES

ǂ ES is essential, GD is growth-defect. Table is sorted by order of ORF (open reading frame) number.

Furthermore, two genes coding for transcription regulators were found. The first one is *mprA_3*, coding for MprA, from which a homologue in *Mycobacterium* is involved in multiple stress-responses by regulating multiple genes under which genes involved in iron metabolism (**Table 3**) (*mprA - Response Regulator MprA - Mycobacterium Tuberculosis (Strain ATCC 25618 / H37Rv) | UniProtKB | UniProt*, n.d.). The next one is *btsR_2*, which has in other bacteria been found to be involved in regulating the balancing of the physiological state of cells within a population (**Table 3**) (Vilhena et al., 2018). Both would fit in the hypothesis where, for recruitment, a certain set of genes need to be activated in response to a perceived signal caused by the Hpa infection. For example, to be able to interact with other bacteria (*Xanthomonas* or other HAM), co-occur with Hpa in the phyllosphere or possibly to activate the production of substances which attack Hpa. Moreover, *bepF_2,* coding for efflux pump periplasmic linker BepF was found (**Table 3**). Efflux pumps are bacterial transport proteins which are used by bacteria to remove harmful substrates (e.g. antibiotics) from their interior to the environment (Sharma et al., 2019). It is possible that *Xanthomonas* uses the pump to be resistant to compounds secreted by Hpa (or the plant) and is able to co-occur with Hpa as a result. However, it could also be an important gene for just its presence on or in the plant.

The last gene solely found among the phyllosphere epiphytes was *thiG,* coding for a Thiazole synthase (**Table 3**). Thiazole together with pyrimidine are used for the production of thiamine (Jin et al., 2021). Thiamine is a vitamin for which oomycetes are auxotrophic. For example, it has been shown that the oomycete *Phytophthora infestans* is dependent on the tomato host for the production of this vitamin (Rodenburg et al., 2021). It is therefore possible that the presence of Hpa in the phyllosphere causes a thiamine limitation which makes it more important for *Xanthomonas* to produce its own thiamine, via thiazole. However, for the fungus *Beauveria bassiana* it has been shown that thiamine biosynthesis is also important for conidiation (Jin et al., 2021). If thiamine would have a same role for conidiation of oomycetes, another theory could be that *Xanthomonas* could suppresses Hpa infection by interfering in the biosynthesis of thiazole by *Arabidopsis*, for example, by utilising the resources itself. In a hypothetical next step *Xanthomonas,* or another HAM member, could scavenge thiazole thiamine production, which reduces the availability for the thiamine auxotrophic Hpa and so affect conidiation.

Most genes found among the *Xanthomonas* phyllosphere endophytes are classified as hypothetical proteins (**Table 4**). From the four annotated genes, two overlap with genes found among the phyllosphere epiphytes, namely *hrcA,* coding for Heat-inducible transcription repressor HrcA, and *hchA,* coding for a Protein/nucleic acid deglycase HchA (**Table 5**). Both could fit in the scenario where a signalling cascade in *Xanthomonas* is activated, on transcriptional level by transcription regulators or protein level by deglycosylation (removal of a sugar group) of a protein, since this is known to be a form of signal transfer within cells (*Deglycosylation - an Overview | ScienceDirect Topics*, n.d.). Despite that, it would be unlikely that HrcA would be activated by heat, so it is possible that there is something changed in *Xanthomonas* which makes it possible to activate this transcription repressor in absence of heat but in reaction to an Hpa-infected plant. This has been shown previously for a cold shock protein (Cspa5) in *R. leguminosarum* (Wheatley et al., 2020). Finally, one gene annotated as coding for a hypothetical protein (XANT_01650) overlaps between phyllosphere endophytes of the soil-inoculation experiment and phyllosphere epiphytes of the spray-inoculation experiment. Because this is the only gene which overlaps in the results of both mutant screens, we made the exception to blast this protein. The gene was found in multiple other *Xanthomonas* species and contains an alpha fold structure (**Figure 11**) (Altschul et al., 1997; D. Xu & Zhang, 2009). These helices are often found in transmembrane proteins such as receptors in the plasma membrane (Alberts et al., 2002). This could point into the direction of a receptor which could enable *Xanthomonas* to perceives a signal molecule produced by Hpa-infected plants.

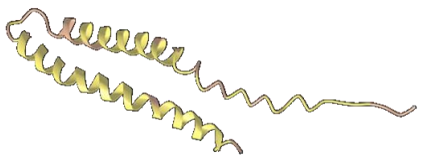


Figure 11. *3D structure of hypothetical protein XANT_01650 which occurs in both mutant screens (D. Xu & Zhang, 2009)*

In summary, the genes that were identified as ES or GD in the phyllosphere of gnoHpa-infected *Arabidopsis* plants could be involved in multiple processes which can roughly be divided in three categories (**Table 6**). However, some genes could be involved in multiple processes across the categories. The first two categories relate to the physical process of recruitment, "Perception & Regulation" and "Movement". This includes processes of signalling outside (hypothetical transmembrane protein) or inside the bacterial cells (transcription regulation or protein modifications), chemotaxis, motility and biofilm formation. The last category, "Antagonism", includes processes which enables *Xanthomonas* to repress, protect itself against and co-occur with Hpa in the phyllosphere (disrupting cell wall using lytic enzymes, competition on resources, maintaining physiological state or regulating stress responses).

Table 6. *Overview of processes with corresponding candidate genes which could be involved in the recruitment of Xanthomonas to the phyllosphere of Hpa-infected plants divided over three categories*

<table>
<tr><th>Categories</th><th>Signalling & Regulation</th><th>Movement</th><th>Antagonism</th></tr>
<tr><td rowspan="6">Process (genes)</td><td>Hypothetical transmembrane protein (XANT_01650)</td><td colspan="2">Biofilm formation (hrpB, clp)</td></tr>
<tr><td rowspan="3">Intracellular regulation (mprA_3, btsR_2, mprA_3, hrcA, clp and hchA)</td><td rowspan="3">Motility (hrpB, cheB_2, clp)</td><td>Co-occurrence:</td></tr>
<tr><td>Maintaining its physiological state (btsR_2 and bepF_2)</td></tr>
<tr><td>Regulating stress responses (mprA_3)</td></tr>
<tr><td colspan="2" rowspan="3">Chemotaxis (cheB_2)</td><td>Suppressing Hpa:</td></tr>
<tr><td>Disrupting cell wall (clp, bprV_1)</td></tr>
<tr><td>Competing for resources (thiG)</td></tr>
</table>

How to beat the beads? – Troubleshooting for library preparation

To be able to analyse which mutants are present in the different conditions, the genomic flanks adjacent to the integrated transposons need to be isolated from the rest of the genomic DNA before sequencing them for further analysis. Since the transposon insertions are at random places in the genome, for the isolation of the genomic flanks a protocol of three to four days is required (from now on called "INSeq DNA preparation (IDP) protocol"). Unfortunately, we encountered a number of challenges to get this protocol working in our lab. This is mainly because the protocol came from a different lab, the Rhizosphere Lab at Oxford University, in combination with the fact that it was an amalgamation of the methods used in several studies (Goodman et al., 2011; Perry & Yost, 2014; Wheatley et al., 2017). Moreover, the Qubit™ Fluorometer (using dsDNA HS/BR Assay Kit™, from now abbreviated as Qubit) in our lab turned out to measure unreliable DNA concentrations, which can make a significant difference for this IDP protocol (which will be discussed below). Furthermore, library preparation protocols like this are known to be very sensitive. First, we discuss the mechanism behind the IDP protocol, followed by the main problems and solutions we encountered.

In short, the IDP protocol consists of the following steps (**Figure 12a-h**) (Goodman et al., 2011). The principle is that you use a certain amount of genomic mutant DNA from which you amplify the genomic flanks adjacent to the transposons via a linear PCR utilising a biotinylated primer (BioSamA) (**Figure 12b**, **Table 8**). These biotin-labelled single stranded flanks are then separated from the genomic template DNA by binding them to streptavidin coated beads (**Figure 12c**). While bond to these beads, the second strand is synthesised, after which an enzymatic digestion using Mme1 follows (**Figure 12d-e**). A specific adapter (LibAdapt) is ligated onto the resulting sticky ends, which makes the amplification of the fragments in the final PCR possible (**Figure 12f-g**). Finally, this final PCR product can be sent for sequencing, in this case Illumina (**Figure 12h**).

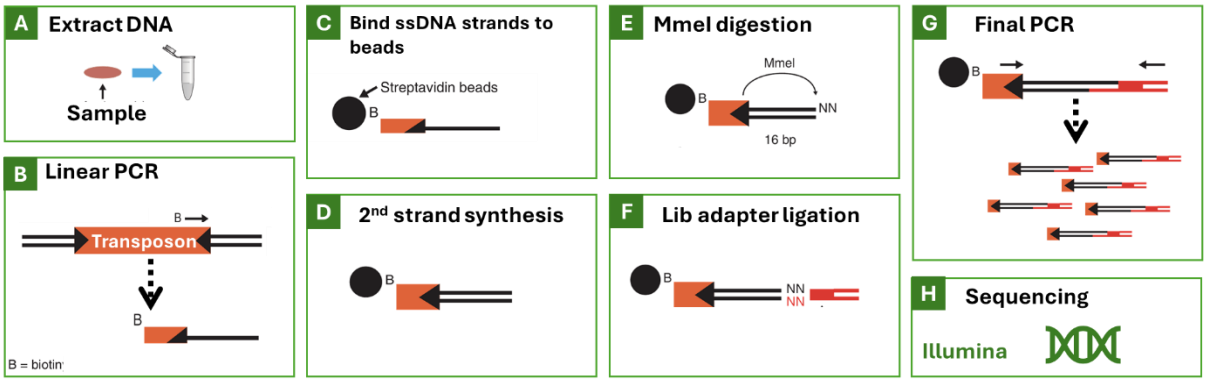


Figure 12. *Overview IDP protocol. It starts with DNA extraction* ***a,*** *followed by a linear PCR step using the biotinylated primer BioSamA (**Table 9**)* ***b,*** *binding the ssDNA to the beads* ***c****, second strand synthesis, MmeI digestion* ***e,*** *ligation of adapter LibAdap (**Table 9**) to the genomic flanks, a final PCR step to amplify the DNA fragments bound to the beads* ***g,*** *and eventually (illumina) sequencing* ***h*** *(Adapted from Goodman et al., 2011).*

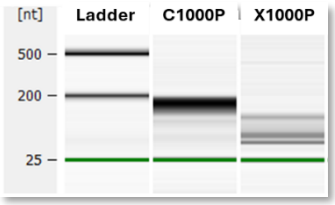


Figure 13. *Fragment of the BioAnalyzer RNA chip gel. X1000P is final PCR product of the IDP protocol with 1000 ng Xanthomonas INSeq library input DNA, C1000P is product when 1000 ng Cupriavidus library DNA was used as input.*

The first trials with the IDP protocol gave a yield at the final PCR of 50-200 ng, while 1500 ng/µl was expected based on the information of Oxford. A good possibility for this low yield was the fact that the streptavidin beads in our own lab turned out to be two years expired. However, the following trials with new beads from two different brands (NEB & Pierce) gave variable yields. Remarkable was that the NEB beads gave 0 to 400 ng, while the two with samples prepared with Pierce beads 560 and 1920 ng. Based on in-depth research into the specifications of the two brands, this was possibly explainable as the Pierce beads are 2.5 times more concentrated and appeared to have a 7-fold higher binding capacity compared to the NEB beads. To get an estimation if the final PCR products were from the right size, we analysed the samples using a BioAnalyser RNA chip (**Figure 13**), since DNA chips were not available. Strikingly, the band of the *Xanthomonas* sample, despite having a 4-fold higher measured DNA concentration, was much weaker than that of *Cupriavidus* which was also loaded on this chip. This suggested that the true DNA concentration differed from that measured by the Qubit fluorometer. For both *Xanthomonas* and *Cupriavidus,* were bands visible of around the right size (121 and 187bp respectively). However, for *Xanthomonas* also two other bands were visible. A possible explanation could be the fact that this was a chip intended for RNA instead of DNA.

To investigate the hypothesis that the Qubit gave incorrect quantities, we made several measurements on DNA standard solutions of which the DNA concentration is known (0, 10 or 100 ng/µl). These are the standards that are included with a Qubit kit and are used to calibrate the device. Indeed, it turned out that the Qubit values deviated drastically from the actual DNA concentrations, both when using the high sensitivity (HS) and the broad range (BR) kit. This ranged from indicating serious overestimations to underestimates. Good examples of overestimations were, where the BR kit measured a concentration of 33 ng/µl for the standard solution of 0 ng/µl or where 39.6 ng/µl was measured instead of 10 ng/µl by the HR kit. Underestimations were, for example, that only 55.6 ng/µl was measured for the standard solution of 100 ng/µl. Testing these standard solutions on the EzDrop™ did give an underestimation of the true concentration (~80% of the real concentration), but the measurements were reasonably consistent. So, for the next IDP protocol trials the input DNA amount was calculated based on EzDrop values instead of the Qubit.

Nevertheless, the IDP protocol states that 500 up to 2000ng DNA can be used as input for the linear PCR, we were wondering how the input amount influences the IDP protocol. Especially since the actual input concentrations so far were probably incorrect due to the mismeasurements by the Qubit.

Therefore, we tested three different input amounts (500, 1000 and 2000 ng) to examine their effects on the linear PCR (**Figure 14**). Moreover, since we noticed differences in the compositions of the PCR reaction mixtures used in different studies, we also included a test to assess the effect of these variations. Likewise, we examined the effect of a two degree higher or lower annealing temperature. As composition variations we tested the addition of extra $MgCl_2$ (Mg^{2+} being a cofactor of the used DNA polymerase, Q5 (Biolabs, n.d.)) and the withholding of the GC enhancer since this is optional for the reaction mixture. As a baseline for these variations we used 1000 ng DNA since this amount previously worked for the IDP protocol for *Cupriavidus*. As positive control plasmid DNA of pSAM_Rl was used, since this contains the transposon on which BioSamA should be able to bind. The linear PCR products were loaded on a gel for electrophoresis and visualisation (**Figure 14**). To begin with, when 500 or 1000 ng mutant DNA was used, we saw a smear containing multiple bands as expected form a linear PCR product (**Figure 14**). However, for 2000 ng input DNA, only a band of >20,000 bp was visible which was probably the genomic input DNA. This indicates that a high input amount of 2000 ng has an inhibitory effect on the PCR reaction. This is possibly caused by the large amount of negatively charged DNA sequestering the positively charged magnesium ions (Mg^{2+}) by which these became unavailable for Q5. Also, in the absence of the GC enhancer no linear PCR products were visible (**Figure 14**). This indicates that the addition of the enhancer is important for the reaction and should be kept in the reaction mixture. The addition of $MgCl_2$ did not completely stop the linear PCR, but the intensity of the bands was slightly lower compared to the reaction according to the protocol (**Figure 14**). So, the addition of extra $MgCl_2$ probably does not improve the linear PCR and might have an inhibitory effect.

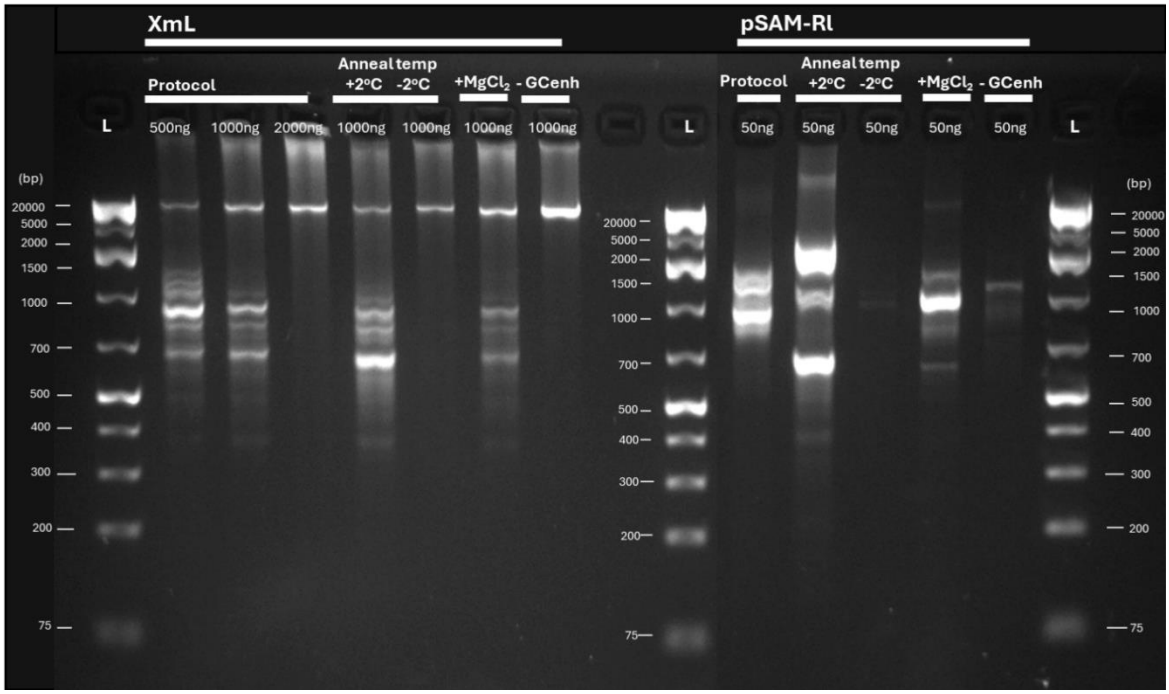


Figure 14. *Linear PCR test with multiple variations on the reaction mixture. GCenh is GC-enhancer, L is ladder, XmL is Xanthomonas library DNA as input, pSAM_Rl is the donor plasmid.*

Reducing the annealing temperature by 2°C gave no smear, while when increasing it by 2°C the smear was visible (**Figure 14**). However, the pattern of the smear seemed slightly different from the two (1000 and 500 ng) produced with an anneal temperature according to the protocol (58.6 °C). The brightest band in the smear was around 600 bp, while this was not observed for the 1000 ng with annealing at protocol temperature. Moreover, for 500 ng the brightest band was around 900 bp. This suggests that the increase in annealing temperature has an effect on the size of the produced single stranded DNA (ssDNA) molecules. For pSAM_Rl similar patterns were observed where no smears were visible when anneal temperature was decreased or when the GC-enhancer was left out of the reaction mixture. The bands were brighter, but this is explainable based on the characteristics of the input templates. The plasmid pSAM_Rl is around 1000 times smaller than the chromosome of *Xanthomonas* (4,600 bp vs

5.1 million bp). So the proportion of the transposon from the rest of the DNA is bigger in the plasmid than in the *Xanthomonas* chromosome. Therefore, the used plasmid DNA contains relatively more template (binding sites for the BioSam primer) for the PCR reaction than the DNA of *Xanthomonas* mutants.

Based on these results, we conclude that the linear PCR works with the in the protocol specified composition of the reaction mixture and annealing temperature. In addition, 500 or 1000 ng input DNA work, but the maximal input amount of 2000 ng in the protocol stated can disturb the linear PCR. These input amounts were calculated based on EzDrop-measured DNA concentrations. However, in all our trials so far, we used 500 or 1000 ng input DNA based on Qubit concentrations. Together, these findings support our hypothesis that an excessive DNA input by an underestimated DNA concentration by the Qubit can be problematic for the functioning of the protocol. However, the Qubit can also make moderate faults or overestimations. In these cases, the linear PCR could probably still produce some ssDNA molecules which can be bound to the beads and eventually serve as template in the final PCR. So, if the input amount would be the only problem, one would expect that the IDP protocol would give a moderate DNA yield more frequently than we observed so far. Therefore, it is possible that another factor is also causing the IDP protocol to not work as expected. There were two options that could lead to less template being available for the final PCR: first, the DNA does not bind or binds poorly to the beads, secondly, improper ligation of the adapters to the sticky ends of the genomic flanks.

To investigate if there is DNA bound to the beads at all, we designed a test primer to replace the normal reverse primer in the final PCR. This test primer does not bind to the adapter sequence, but ~25 bp downstream on the transposon (**Figure 15a**). We used the primer to perform the PCR on beads from a previous IDP protocol that gave a low yield and the products were again loaded on a gel for analysis (**Figure 15b**). As negative control we used a mixture without template (NT). As positive controls we used the DNA from donor plasmid pSAM_Rl, input library DNA from *Cupriavidus* (since the IDP protocol worked for this strain before), but also library input DNA from *Xanthomonas*. Indeed, bands of around the expected size were visible for PCR performed on the beads as well as on input DNA from *Xanthomonas* and *Cupriavidus* (67 and 131 bp respectively). Moreover, also pSAM_Rl showed a band of the right size (67 bp), and for NT as expected only a band around primer size was visible (~20 bp) (**Figure 15b**). This confirms that there is DNA bound to the beads. To see if this DNA also contains ligated adapters, we did the final PCR with the normal reverse primer as well (**Figure 15c**). Again, bands were visible around the expected size for *Xanthomonas* (121 bp), but they were less bright than those of the test primer. This indicates that at least a portion of the DNA had ligated adapters, but possibly less then expected.

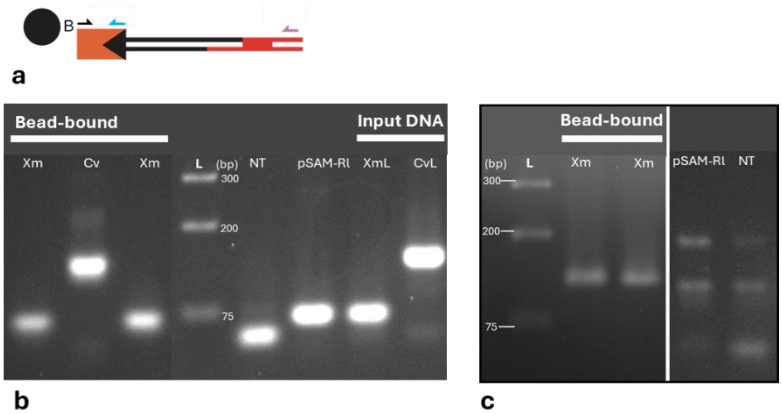


Figure 15. *Schematic overview primer binding locations on DNA bound to beads* ***(a)****. The back, blue and pink arrows represent the normal forward, test reverse and normal reverse primers, respectively. Adapter (LibAdap), in red. Agarose gel of the PCR products from the test with the test primer* ***(b)*** *or normal reverse primer* ***(c)*** *were loaded on an agarose gel for analysis. The reaction is tested on Xanthomonas (Xm) and Cupriavidus (Cv) on beads, the input DNA from the library (XmL, CvL) and on plasmid DNA of pSAM_Rl. L is ladder, bp is basepairs, NT is no-template. If products are loaded on different gels, they are separated by a white gap. With the test primer, a product of 67 and 131 bp is expected and with the normal primer set a product of 121 or 187bp for Xm (and pSAM_Rl for test primer) and Cv respectively.*

It was a good sign that the bands were around the expected size on the agarose gel, but the resolution of the available ladder is not high enough to conclude this with certainty. Therefore, the samples from a new IDP protocol test were loaded on a bioanalyzer DNA chip (**Figure 16**). All the eleven samples on

the gel of the bioanalyzer had a size of 121 bp, which confirms that the final PCR products are indeed of the right size. This shows that the right DNA is amplified during the final PCR. Surprisingly, pSAM_RI and NT showed some bands around 110 and 190 bp (**Figure 15c)**. Since these were made from a second master mix, it is possible that there has been some contamination causing the unexpected PCR products.

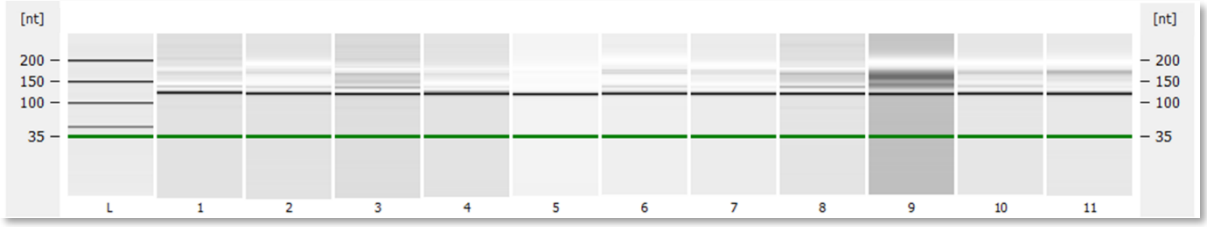


Figure 16. *Bioanalyzer DNA chip of final PCR products IDP protocol. All samples have a band at the expected size of 121bp. L is ladder. Numbers under lanes are biological replicates.*

After all the above mentioned trouble shooting, the yields of the final PCR's were acceptable. However, they were still lower than expected based on the information from Oxford. Moreover, the less intense bands of the final PCR products with the normal primers compared to the test primers suggested that the adapter ligation step was not functionating optimal (**Figure 12f**). Nevertheless, with combining the yields of multiple IDP protocol rounds it was possible to acquire enough DNA to send the first batch for sequencing. There were two final missing pieces for optimising the ligation efficiency. First, it turned out that that the T4 ligase used in Oxford was 400 times more concentrated than the one in Utrecht (2000 vs 5 U/µl respectively). The final PCR yields of the IDP protocols with the concentrated ligase were significantly higher than the previous ones with the less concentrated ligase (**Figure 17a**). Moreover, we discovered a correlation between the ssDNA yield of the linear PCR and the yield of the final PCR (**Figure 17b**) and that the yield of the IDP protocol could be improved when the products of two linear PCR reactions were used to bind to the beads instead of one. Secondly, we tested two different types of agitation methods for resuspending the beads during the incubation steps, mechanical shaking and hand flicking. We estimated the adapter ligation method by comparing the ratios of molecules produced in final PCRs with the test primer and the normal reverse primer as described in Figure 15a. If the ligation efficiency is 100%, you would expect a 50/50 ratio. We found

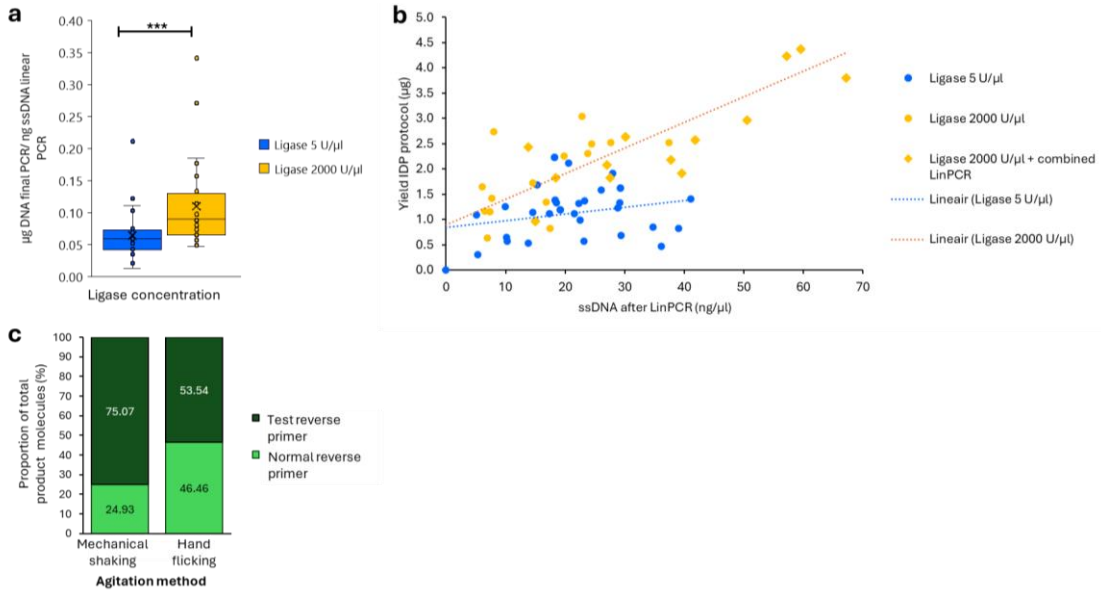


Figure 17*. Effect ligase concentration* ***(a)*** *and linear PCR (LinPCR) yield* ***(b)*** *on yield final PCR IDP protocol. Asterisks indicate significance level in one-sided Student's t-test of 30 (ligase 5U/µl) and 28 (ligase 5U/µl) replicates: *** is P<0.001. Crosses (x) are averages and dots represent individual samples. Effect of agitation method on Ligation efficiency* ***(c)*** *represented by percentage of molecules from total produced in the final PCR by using the test primer inside the transposon or the normal final reverse primer binding to the adapter (Figure 15a).*

that hand flicking gave a higher ligation efficiency of ~86% adapter ligation (~53:46 ratio, Test:Normal primers) than the mechanical shaking (~30%, corresponding to ~25:75, Test:Normal) (**Figure 17c**). By using the concentrated ligase combined with hand flicking, we got in most cases enough DNA to send for sequencing.

Discussion

Here, we investigated which genes of *Xanthomonas sp.* WCS2014-23 (a representative of the most abundant HAM ASV) are important for its recruitment to the phyllosphere of *Arabidopsis* during aboveground Hpa infection. To examine this, we created a *Xanthomonas* INSeq transposon mutant library and utilised this in multiple mutant screens to compare the present mutations between the phyllospheres of mock-treated and gnoHpa-infected plants. Our INSeq library had an insertion density of approximately 58%, which corresponds to insertions in ~50.6 to ~72.2% of the 4399 CDSs present in the reference genome. According to this insertion percentage, we cautiously conclude that our library has a high diversity since an insertion density of >50% is classified as ideal (Ioerger, 2022). Moreover, the insertion density is above the recommended 35% for statistical analysis (*TPP Overview — TRANSIT v1.1.2 Documentation*, n.d.). Due to the variability of TA hits among individual freezer stock library samples (and input samples), we were not able to make an exact estimate of the distribution of the CDSs over the four-state classification in the initial library. Nevertheless, we were able to estimate the approximate ranges. About 204-384 are predicted to be essential, 76-101 to give a growth-defect when disrupted and approximately three to give a growth-advantage. For 51 genes classification was impossible since these do not contain TA-sites. This is comparable to *Xanthomonas hortorum* pv. vitians LM16734, a plant pathogen of lettuce, which has ~370 essential, 50 growth-defect, 79 growth-advantage, 3885-4035 non-essential and 34 genes without classification due to the lack of TA-sites (Morinière et al., 2021).

We identified 63 candidate recruitment genes, 28 of which are already annotated while the remainder are predicted as hypothetical proteins (**Figure 10**, **Table 3-5**). For this Master's project we focused on the annotated genes and selected the ones with the most potential based on current literature. The genes that came through the selection can be divided in roughly three categories which include multiple processes. The first category, "Perception & Regulation", includes genes involved in signalling outside (hypothetical protein, XANT_01650) or inside the bacterial cells on level of transcription regulation (*mprA_3, btsR_2* and *hrcA*) or protein modifications (*hchA*). The second category, "Movement", includes genes involved in motility and biofilm formation (*hrpB*, *cheB_2* and *clp).* The last category, "Antagonism", includes genes which enables *Xanthomonas* to repress Hpa or to protect itself against it. Genes possibly involved in Hpa suppression, encode for example for lytic enzymes (*clp, bprV_1*) or possibly genes for competition on resources as thiazole (*thiG*). However, biofilm formation could also play a role in repressing Hpa (K. Xu et al., 2021). Genes potentially enabling the co-occurrence are for example encoding for proteins involved in maintaining its physiological state such as efflux pumps (*btsR_2* and *bepF_2*) or regulate stress responses (*mprA_3*). These genes could fit in a scenario where *Xanthomonas* perceives a Hpa-infected plant which activates a process inside the bacterium leading to the movement to the phyllosphere (**Figure 18**). Once arrived at the sites of infection, the bacterium would use certain mechanisms to repress Hpa and cause a reduced infection. The theory that *Xanthomonas* on its own is able to suppress Hpa could be supported by the significantly weaker infection of plants spray-inoculated with the *Xanthomonas* mutant library (**Figure 9**). Nonetheless, previous studies have not proved that *Xanthomonas* by itself, without other HAM members, can inhibit Hpa (Berendsen et al., 2018). Finally, it remains important to keep in mind that the described genes are candidate genes and that for definitive confirmation as recruitment genes it will be necessary to experimentally validate them.

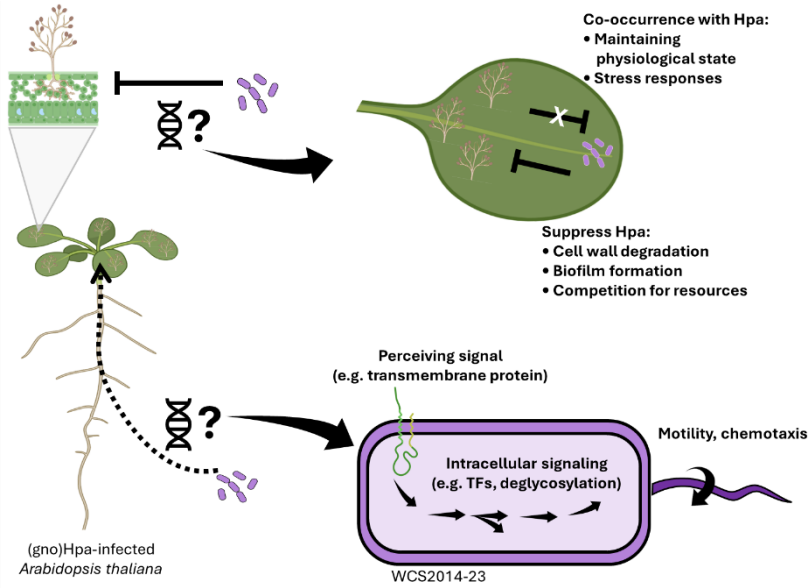


Figure 18. *Schematic overview processes in which the found candidate recruitment genes could be involved. The found genes point into the direction of a mechanism of recruitment in which Xanthomonas somehow observes a signal from an infected plant, which causes a process inside the bacterium which leads to chemotaxis and motility leading to movement towards the phyllosphere. Once arrived in the phyllosphere, Xanthomonas could have mechanisms in which it is able to suppress Hpa or which make it able to co-occur with Hpa. Adapted from BioRender.com (2024). Retrieved from https://app.biorender.com/biorender-templates*

Although the number of genes about which we could make a prediction in this study was somewhat limited, we are confident that the genes found could play a role in the recruitment of *Xanthomonas*. These mutants passed the entire selection process in the mutant screens and the two most important parameters for validation of our set-ups gave results as expected. Firstly, we saw a significant higher population density of *Xanthomonas* among the phyllosphere epiphytes of gnoHpa-infected plants, which confirms the recruitment in response to Hpa. Secondly, we harvested considerably more than the minimum number of mutants needed for the INSeq statistical analysis. Moreover, the selected genes all had confidence scores above the 75% (classified as a high certainty by Gollapalli et al., (2022)) and we used a normalisation method (Betageom) suitable for datasets which contain some skewness (*Normalization — TRANSIT v3.3.3 Documentation*, n.d.). The skewness in our data, wherein only a few TA sites had read counts in the millions, could suggest positive selection or an artifact resulting from PCR jackpotting (Ioerger, 2022). It would be quite remarkable if somehow Himar1C9 has a preference for inserting in these TA sites, since this is normally not observed with mariner transposon and is a phenomenon more reported among Tn5 transposons (Liu et al., 2013). A possible scenario could be that these locations in the genome of the wildtype *Xanthomonas* have (partial) homologies with the primer used in the linear PCR (BioSamA, **Table 9**. Primers INSeq library preparation), possibly due to the insertion of a transposon in the past. This could be checked by performing the IDP (INSeq DNA Preparation) protocol with wildtype DNA. However, BLASTing BioSamA to this part of reference genome did not give binding sites that matched well enough to support this hypothesis.

Nevertheless, there are improvements possible for this study. It would be valuable to examine which step(s) in the process between DNA isolation and the sequencing result is the bottle neck that causes the difference in TA hits between comparable samples. By solving this, it could be predicted for a larger number of genes whether they are a potential recruitment gene. Moreover, the danger of a difference in TA hits is the possibility that reads are absent in a sample due to chance (stochasticity), instead of due to the gene essentiality in that condition. The bottleneck could be somewhere in the IDP (INSeq DNA preparation) protocol, but since the protocol seems to be working properly after all the improvements and gives now the expected DNA yield, this seems less likely. A more probable direction to look at is the sequencing method, Illumina, combined with the preceding enzymatic adapter ligation without PCR amplification performed by the sequencing company. It would be worth testing if sending a higher amount of DNA for sequencing than recommended could reduce potential biases during the adapter ligation to give less variable TA hits. In addition, previous INSeq studies, by Perry & Yost, (2014) and Wheatley et al., (2017), used the alternative sequencing method of ion semiconductor sequencing

(Ion Torrent) instead of Illumina. Moreover, for this method the sequencing company does not perform the adapter ligation, but barcoded LibAdapters are used during the IDP protocol instead. Lastly, it would be possible to sequence multiple technical replicates per biological replicate and combine the reads for the INSeq analysis, since we observed that pooling reads from multiple samples increased the amount of TA hits. For our analysis, we pooled the reads from the three biological replicates within the same condition and compared these combined samples, rather than analysing them separately. Finally, for the shoot wash-off samples of the *Xanthomonas* spray-inoculation mutant screen, it would be better to repeat the IDP protocol using the more concentrated ligase, as this turned out to be a critical factor in the IDP protocol (see section *How to beat the beads? Troubleshooting for library preparation,* **Figure 17**). This could be the reason why fewer genes emerged from the selection in this mutant screen than in the soil-inoculation mutant screen (13 versus 31, respectively) (**Figure 10**).

One possible improvement for the sequencing analysis would be to use a more complete reference genome. A complete reference genome with a circular chromosome would enable to use the LOESS correction of the HMM algorithm which allows an extra normalisation based on genomic position and variations of insertion counts on a large scale across the genome (DeJesus et al., 2015). Moreover, it would be an improvement to implement a correction for gene duplications. This could, for example, be done by BLASTing the hypothetical proteins against the reference genome to see which ones overlap (Morinière et al., 2021). However, while the presence of duplicated genes probably does not result in false positives of potentially recruitment genes, it could give a better estimate of the number of essential, growth-defect or growth-advantage genes in the initial mutant library.

Previous studies using Tn-seq to investigate beneficial microbes have primarily focused on genes involved in root recruitment and colonisation, whereas our study focused on recruitment to the phyllosphere. Nonetheless, there is some overlap between the processes the genes identified in those studies are involved and our findings, for example genes involved in motility. This process emerged in our study but have also been shown to be essential for the colonisation of *Arabidopsis* roots by *P. simiae* and pea roots by *R. leguminosarum* (Cole et al., 2017; Wheatley et al., 2020). Moreover, Wheatley et al., (2020) found that certain metabolic adaptations are required for nodulation in pea roots, for example the synthesis of certain amino acids. Among our list of candidate genes there are also some involved in metabolic processes, for example the production of thiazole (*thiG*) as mentioned above, but also Cytochrome c-type biogenesis (*ccmE_1*) and asparagine synthesis (*asnB_2*) (**Table 3**). Finally, genes involved in the production of lytic enzymes and biofilm formation were earlier found to be involved in the attack of plant pathogenic oomycetes and fungi by beneficial microbes (Xu et al., 2021).

Future research would be needed to validate the candidate recruitment genes. This could be done by making knock-out mutants and testing their recruitment individually in a similar experimental set-up as was used for the mutant screens in this study. Furthermore, blasting the hypothetical proteins could help to make a bigger selection of genes for validation. For the validated genes, it would be interesting to investigate with the help of microscopy whether *Xanthomonas* localisation in the phyllosphere is also affected besides the recruitment. Moreover, for transcription factors it could be valuable to identify which genes they might regulate. This could be done, for example, using ChIP-Seq, but also by comparing expression patterns of knockout mutants with the wildtype using RNA sequencing. This could give a better understanding of which mechanisms they activate that lead to recruitment or the suppression of Hpa.

Besides the further investigation of the genes found in this research, the INSeq mutant library could also be utilised in additional mutant screens. A possibility would be to perform a mutant screen in a SBL experiment over multiple generations, since this study was similar to a one-generation SBL experiment. This would potentially give an even stronger selection for genes essential for recruitment

or the antagonism with Hpa. Another option would be to sample the mutants at different timepoints during the same setup as used in this study. This could give insight into which genes are important in the different stages of recruitment. Another aspect that could contribute to the interpretation of the results from our screens, as well as the design of future mutant screens, would be to examine the number of *Xanthomonas* colonization events per plant (or possibly per leaf) in the two phyllosphere compartments. This may be helpful to assess whether the lower number of selected genes among phyllosphere endophytes compared to epiphytes in the soil-inoculation mutant screen (10 versus 31, respectively) is due to fewer colonization events in the endosphere, or, for example, a lower concentration of mutant DNA in the sample because of the presence of plant DNA. In addition, it could help with a more accurate calculation of how many plants are needed per treatment group, which could potentially reduce the size of treatment groups in experiments, allowing the testing of more conditions simultaneously in the same experimental run. Finally, a worthy follow-up could be to create a mutant library in one of the enriched bacteria on the crop spinach when infected with downy mildew pathogen *Peronospora effusa* (Pe) (Goossens et al., 2024). For example, a representative of one of the ASVs overlapping with HAM. Utilising this library in a similar mutant screen could provide the opportunity to compare the candidate recruitment genes with our findings on *Xanthomonas* in *Arabidopsis*.

Taken all together, with this research we give insight into which genes of beneficial microbes may be involved in their recruitment to the phyllosphere in response to a foliar pathogen infection. It builds further on the study by Goossens et al., (2023) in which they provide evidence that plants actively recruit HAM from soil in response to Hpa and that HAM suppress Hpa infection. Our findings give leads for further research and could provide new insights into the mechanisms involved in these interactions from the microbial side of the interaction. Additionally, insight into how HAM are able to suppress Hpa could contribute to the development of new strategies for reducing crop losses due to downy mildew infection. This could potentially include the development of bioinoculants with beneficial bacteria which could replace environmentally unfriendly, chemical pesticides.

Materials and methods

Creating *Xanthomonas* mutant library

For making the mutant library in *Xanthomonas sp.* WCS2014-23, we used transposon insertion mutagenesis via a triparental mating. Rifampicin (Rif) resistant WCS2014-23 was grown shaking at 150 rpm at 28°C to mid-log phase (OD 0.5-0.6) prior to mating. The helper with the conjugative plasmid pRK2073 (some exceptions pRK2013 or pRK600) and the donor strain with plasmid pSAM_Rl (Figure 4a) containing the mariner transposon (himar1C9) were subcultured at 37°C 150 rpm, four hours in advance (**Table 7**). The cultures were spun down at 6000 rpm and resuspended in 1 ml LB. Subsequently, the three strains were mixed in ratio 4:4:2 (pSAM_Rl:WCS2014-23:pRK2073) and the mating was carried out overnight at 30°C.

To clean the library, we selected for mutants by growing the conjugation mix on 14 selection plates (25x25 cm, 300 000 CFU/plate) with LB + $Rif_{100}Kan_{50}Nys_{50}$ 2-3 days at 28°C. Subsequently, the grown colonies were scraped from the selection plates using sterile autoclaved razorblades and suspended in LB. This is used to expand the library, by setting up two litres of liquid culture of OD 0.1 in LB + $Rif_{100}Kan_{50}Nys_{50}$ and grew this at 150 rpm 27°C until OD 1.0. From this, the INSeq library stock is made by adding glycerol in final concentration 20% after which aliquots of 1 ml were flash frozen with liquid nitrogen and stored at -80°C.

Table 7. *Overview plasmids*

Plasmid	Role in conjugation	Strain	Reference
pSAM_Rl	Donor	*E. coli* SM10λpir	(Perry & Yost, 2014)
pRK2073	Helper	*E. coli* HB101	(*Addgene: Vector Database - pRK2073*, n.d.; Hall et al., 2023)
pRK2013	Helper	*E. coli* HB101	(Hall et al., 2023)
pRK600	Helper	*E. coli* HB101	(Kessler et al., 1992)

Colony PCRs for checking transposon insertions

For the colony PCRs, colonies were gently touched using sterilised toothpicks and used to mix the attached bacterial cells into the master mixes with different primer sets (**Table 8**). Touch-down PCR was used and the thermocycler have run as follow: 95 °C for 10 min, followed by 6 cycles in which the anneal temperature started at 60°C which each cycle was decreased by 2°C (denaturation 95 °C 30 sec, annealing 30 sec, extension 72°C 1 min), this was followed by 35 cycles with a constant anneal temperature of 52°C and a final extension at 72°C for 10 min. PCR products were mixed with loading dye (Invitrogen) and 5 µl was loaded onto a 1.2% agarose gel (TBE) and run at 110V for 10 min.

Mutant screens

Plant materials and growth conditions

In this study, we used the *Arabidopsis* accessions Col-0. For both mutant screens, the plants were grown in natural soil from the Reijerscamp nature reserve (52.0107° N, 5.7825° E) in the Netherlands excavated in April 2023 (Berendsen et al., 2018). The soil was air-dried and sieved as described in (Berendsen et al., 2018) . On the day of sowing, the soil was watered in a 1.25:10 v/w ratio. 60 ml pots were filled with 120 g of soil (±2.5 g) placed in 60 mm Petri dishes. Seeds were sterilised by vapour-phase sterilisation as described in Lindsey et al., (2017), suspended in sterile 0.2% water agar and stratified in the dark at 4°C for 3 days. Approximately 30 seeds were sown per pot by pipetting one to three seeds into each hole of a 16-hole plastic template as described by Goossens et al., (2023) and Vismans et al., (2022). After sowing, the plants were placed in trays with transparent lids and grown at 21 °C, light intensity 100 µmol m^{-2} s^{-1} 10/14h-light/dark and 70% relative humidity (Goossens et al., 2023; Vismans et al., 2021). In the first week, plants were watered three times, 3 ml on day of sowing, 5 ml the day after sowing and another 3ml after two days. After the first week of growth, transparent lids were changed to mesh lids. In the second week the plants were watered three times with 5 ml ½-strength Hoagland Supplementation medium (Vismans et al., 2022). After two weeks of growth, the pots were randomised and divided into biological replicates in such a way that each replicate consisted of an equal composition based on germination fraction, person who sowed and person who potted. Hereafter, half of the plants were inoculated with gnotobiotic Hpa as described below (Goossens et al., 2023; Vismans et al., 2022).

gnoHpa *inoculation and preparation of spore suspension*

gnoHpa was maintained on Col-0 and *eds1* plants as described in Goossens et al., (2023). *eds1* plants are *Arabidopsis* mutants hypersusceptible to Hpa infection (Parker et al., 1996), which makes them suitable for a higher spore production. The spore suspensions were prepared as follows. Shoot material of 14 days infected *eds1* plants was collected into autoclaved tap water and the tubes were thoroughly hand-shaken to release spores. To remove plant material, the suspension was filtered through Miracloth (22-25 µM pore size) after which the spore density was determined by counting 3 separate droplets of 1 µl using a transmitted-light microscope (Carl Zeiss Microscopy, Standard 25 International Classification for Standards, item number 450815.9902). The spore suspension was directly used to spray-inoculate two-week-old plants (~58 and ~125 spores/µL for soil and spray experiment, respectively) using an airbrush (Vismans et al., 2021, p. 20). Plants were sprayed two times to create

clear droplet formation on the leaves, were airdried for 1h and randomised with mock-treated pots (sprayed with sterilised tap water) in the trays. To increase the humidity to ensure a proper infection, lids were replaced by closed ones which were sprayed with water on the inside and lids were taped to the trays (Vismans et al., 2022).

Biological replicates

For both mutant screens, approximately 150 pots with ~30 WT *Arabidopsis* plants per pot were inoculated with the *Xanthomonas* mutants. These were divided in three biological replicates ~44 pots from which half were inoculated with gnoHpa as described above. Besides that, a control group of 15 pots without *Xanthomonas* was used from which 10 pots were infected with gnoHpa.

Soil inoculation with *Xanthomonas* mutant library

Two millilitres of *Xanthomonas* mutant library were used to inoculate a 400 ml culture of LB + $Rif_{100}Kan_{50}$ which was grown to OD 0.5. The bacteria were pelleted (centrifugating at 4500 rpm), washed 2 times with 5 ml 10mM $MgSO_4$ and resuspended in 400 ml 10 mM $MgSO_4$. From this resuspension, 50 ml was reserved as input sample, pelleted as described above and frozen at -20°C for later DNA extraction. The rest of the suspension is used to mix the mutants though the natural soil in a concentration of 10^7 CFU/g gram soil

Spray inoculation shoots with *Xanthomonas* mutant library

Again 2 ml of *Xanthomonas* mutant library were used to inoculate a 400 ml culture of LB + $Rif_{100}Kan_{50}$ which was grown to OD 0.4. The culture was washed and resuspended as described above. From this resuspension, 50 ml was reserved as the input sample, pelleted, and stored as described for later DNA extraction. The remaining was used to make a suspension of OD 0.3 in 10 mM $MgSO_4$ for inoculation. Per 10 pots, 10 ml suspension was sprayed from top on the shoots of two-week-old *Arabidopsis* plants using 10 ml "perfume spray" bottles. These bottles were sterilised before use by incubating in 10% bleach solution of 20 min followed by 5 washes in sterile miliQ.

Isolation of mutants from colonisation experiments

One week after infection, the present mutants were harvested from different compartments of the plants. For both mutant screens the shoot wash-offs and ground shoots were harvested (**Figure 19**). Additionally, for the experiment where the library was mixed trough the soil the rhizosphere/roots was harvested as well. For both experiments, the three biological replicates of ~44 pots each (22 mock, 22 gnoHpa-infected) were harvested in eight groups of 5 to 6 pots. On the day of harvesting, the shoots of each group were collected using 70% ethanol sterilised scissors and tweezers by cutting just above the soil surface and put in 50 ml falcon tubes containing 10 ml 10 mM $MgSO_4$ + 0.02% silwet solution. This is done carefully without touching the soil to prevent the contamination of these phyllosphere samples with mutants present in the soil. Subsequently the tubes were incubated for 1h at 15°C 150rpm to wash-off the mutants. After reserving 100 µl from each tube for making serial dilutions and spore counting, the wash-offs of the four tubes corresponding to the mock or Hpa treatment were filtered through sterilised Miracloth (22-25 µM pore size), pooled and toped-up to 50 ml with 10 mM $MgSO_4$. This is used to inoculate two liquid cultures by adding 25 µl of this wash-off solution. For the soil experiment, 100 ml LB1x +$Rif_{100}Kan_{50}Nys_{50}$ was used, but 75 ml LB1.33x+$Rif_{100}Kan_{50}Nys_{50}$ for the spray experiment. Subsequently, the shoot materials were transferred to "tissuelyser tubes" (50 ml falcon tubes containing 8 ml 10 mM $MgSO_4$ and ~2 ml of glass beads (3 mm)). They were ground using a paint shaker by two rounds of 9 minutes at 60 Hz. The roots were harvested by washing away the soil under a tap above a sieve and a GMO container. The roots were then per group added to tissuelyser tubes and ground using a paint shaker as described for the shoots above. For both ground shoots and roots, 100 µl was reserved for serial diluting, the samples were filtered, pooled and used for subculturing as described above. Subcultures were grown for 12h

or 16h at 28°C 150 rpm respectively for the soil and spray experiment, after which cultures were pelleted by centrifuging at 4500rpm and frozen at -20°C for later DNA extraction.

Determining Xanthomonas population densities and infection rates

To measure the mutant abundance, per harvested group of 5-6 pots, for the wash-off, ground shoots and ground roots, dilution series of 10^0-10^{-7} dilutions were prepared using 10 mM $MgSO_4$ and plated in droplets of 3 µl on LBA+$Rif_{100}Kan_{50}Nys_{50}$. With at 10^0 the using the above mentioned reserved 100 µl. A minimum of two technical replicates was used by plating on different plates (**Figure 19**). Bacterial CFU were quantified after 3 days growth at room temperature (~25 °C). To measure the infection rate, from the wash-off samples, the spore density was determined as described above.

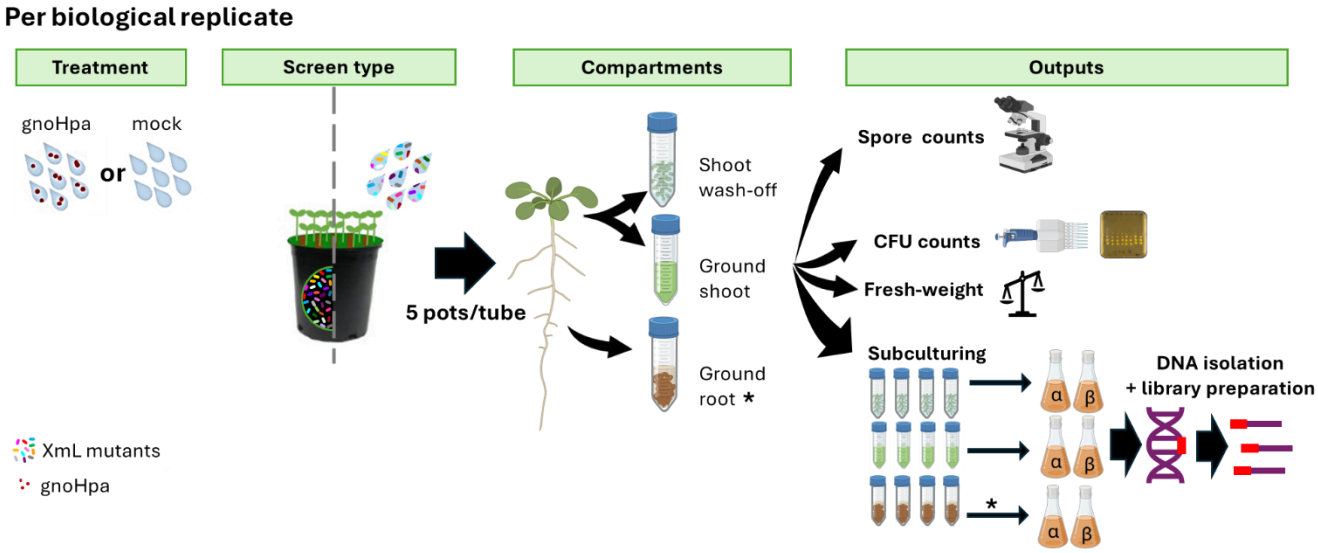


Figure 19. *Schematic overview of treatments and harvesting mutant screen experiments. * was only harvested from experiment which was soil inoculated, data from the roots is beyond the scope of this Master thesis. α and β are technical replicates subculturing. Adapted from Vismans et al. (2021) and made with BioRender.com (2024).*

DNA isolation, library preparation and sequencing

DNA from the frozen bacterial pellets described above was isolated according to the for gram-negative bacteria optimised protocol using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). However, the following modifications were made: pellet was resuspended in 360 µl ATL instead of 180µl, 2 µl or RNase A (100 mg/µl) was added prior to incubation at 56°C, after this incubation the mixture was added to ribolyzer tubes and ribolyzed at f=30Hz for a total of 90 seconds (in 30 sec intervals), ribolyzer tubes were centrifuged at 13 000 rpm for 5 min and supernatant was taken for DNA extraction following manufacturers protocol, DNA was eluted in 100 µl of Milli-Q water. DNA was quantified with a EzDrop 1000 Micro-Volume spectrophotometer (Blue-Ray Biotech, Taipei, Taiwan).

Mariner transposon insertion tags were prepared for DNA sequencing using a combination of the INSeq methods from Goodman et al. (2011) and Perry & Yost (2014). Linear PCR products were amplified using a biotinylated primer, BiosamA (**Table 9**), using 500 ng template DNA. Reaction mixture was split into 2 × 50 µl in PCR tubes. For ground shoot samples an extra linear PCR was performed. The thermocycler was runned as follow: 98 °C for 2 min, followed by 50 cycles of 98°C for 15 s and 58.6°C for 1 min (Perry & Yost, 2014). PCR products were pooled and purified using a Monarch PCR & DNA Cleanup Kit (Bioke, NEB) according to protocol specified for ssDNA of manufacturer, but elution was in 50 µl miliQ. The ssDNA was quantified as described above to check the yield before continuing the rest of the protocol. The binding of the biotinylated linear PCR products to Pierce Streptavidin Magnetic Beads (Thermo Scientific) and the following enzymatic library preparation steps were performed as described in Goodman et al. (2011). However, for the spray experiment wash-off samples a T4 DNA

ligase with a concentration of 5 U/µl was used. Moreover, for the LibAdapters and final PCR different primers were used (**Table 8**). The final PCR products were purified using a Monarch PCR & DNA Cleanup Kit (Bioke, NEB) according to protocol specified for dsDNA of manufacturer, but DNA was eluted in 30-50 µl Mili-Q. The generated 121 bp fragments consisted of ~16 bp genomic DNA (the tag) flanked by 79 bp transposon DNA and 26 bp LibAdapter (**Table 9**). Libraries were concentrated to 60 ng/µl in 20-30 µl (total 1.5~3 µg) and send for paired-end Illumina sequencing (Novogene Munich).

INSeq data analysis

For analysing the sequencing data, the 46 contigs of the reference were combined in order from biggest to smallest. The genome was annotated using PROKKA with genus *Xanthomonas*. Gff file was converted into prot_table using TRANSIT function gff_to_prottable. Reads were trimmed using Cutadapt using a minimal quality score of 20 leaving only 14-17 bp of genomic flanks plus 11 bp LibAdapter (**Table 9**) (see also the section *Supplemental, scripts INSeq analysis*). The second reads of the pairs were transformed in the reverse complement to make them in the same direction starting with the 11 bp LibAdapter. Subsequently, the read pairs were merged to one file. Reads were mapped to reference genome using bwa aln, a seedlength of 13 and by allowing one mismatch (default TPP) (see also the section *Supplemental, scripts INSeq analysis*). The wig file produced was used as input for TRANSIT HMM four-state classification which labels genes as essential (ES), non- essential (NE), growth-defect (GD) or growth-advantage (GA) (see also the section *Supplemental, scripts INSeq analysis*). TA-sites in 10% of the N- and C-terminus were ignored. Confidence scores for the gene classes were added using TRANSIT post processing script HMM_conf.py (*HMM — TRANSIT v3.3.3 Documentation*, n.d.) (see also the section *Supplemental, scripts INSeq analysis*). Recruitment genes were selected by: screening for genes with more than 3 TA-sites, NE in input and mock condition, but ES, GD or GA in gnoHpa-inoculated plants and with a confidence score of 0.75 or higher.

References

Addgene: Vector Database—pRK2073. (n.d.). Retrieved 19 January 2024, from https://www.addgene.org/vector-database/3941/

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Membrane Proteins. In *Molecular Biology of the Cell. 4th edition*. Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK26878/

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Analysis Methods—TRANSIT v2.0.0 documentation. (n.d.). Retrieved 16 August 2024, from https://transit.readthedocs.io/en/v2.0.0/transit_methods.html

Bakker, P. A. H. M., Pieterse, C. M. J., de Jonge, R., & Berendsen, R. L. (2018). The Soil-Borne Legacy. *Cell*, *172*(6), 1178–1180. https://doi.org/10.1016/j.cell.2018.02.024

Berendsen, R. L., Vismans, G., Yu, K., Song, Y., De Jonge, R., Burgman, W. P., Burmølle, M., Herschend, J., Bakker, P. A. H. M., & Pieterse, C. M. J. (2018). Disease-induced assemblage of a plant-beneficial bacterial consortium. *The ISME Journal*, *12*(6), 1496–1507. https://doi.org/10.1038/s41396-018-0093-1

Berendsen, R. L., Vismans, G., Yu, K., Song, Y., de Jonge, R., Burgman, W. P., Burmølle, M., Herschend, J., Bakker, P. A. H. M., & Pieterse, C. M. J. (2018). Disease-induced assemblage of a plant-beneficial bacterial consortium. *The ISME Journal*, *12*(6), 1496–1507. https://doi.org/10.1038/s41396-018-0093-1

Biolabs, N. E. (n.d.). *Four tips for PCR amplification of GC-rich sequences*. Retrieved 19 July 2024, from https://www.neb.com/en/nebinspired-blog/four-tips-for-pcr-amplification-of-gc-rich-sequences

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., & van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nature Reviews Genetics*, *21*(9), 526–540. https://doi.org/10.1038/s41576-020-0244-x

Champie, A., De Grandmaison, A., Jeanneau, S., Grenier, F., Jacques, P.-É., & Rodrigue, S. (2023). Enabling low-cost and robust essentiality studies with high-throughput transposon mutagenesis (HTTM). *PLOS ONE*, *18*(4), e0283990. https://doi.org/10.1371/journal.pone.0283990

Chen, P., Liu, J., Zeng, M., & Sang, H. (2020). Exploring the molecular mechanism of azole resistance in Aspergillus fumigatus. *Journal de Mycologie Médicale*, *30*(1), 100915. https://doi.org/10.1016/j.mycmed.2019.100915

Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., Malmstrom, R. R., Deutschbauer, A. M., Dangl, J. L., & Visel, A. (2017). Genome-wide identification of bacterial plant colonization genes. *PLoS Biology*, *15*(9), e2002860. https://doi.org/10.1371/journal.pbio.2002860

Deglycosylation—An overview | ScienceDirect Topics. (n.d.). Retrieved 26 August 2024, from https://www-sciencedirect-com.utrechtuniversity.idm.oclc.org/topics/biochemistry-genetics-and-molecular-biology/deglycosylation

DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C., & Ioerger, T. R. (2015). TRANSIT--A Software Tool for Himar1 TnSeq Analysis. *PLoS Computational Biology*, *11*(10), e1004401. https://doi.org/10.1371/journal.pcbi.1004401

Giannoukos, G., Ciulla, D. M., Marco, E., Abdulkerim, H. S., Barrera, L. A., Bothmer, A., Dhanapal, V., Gloskowski, S. W., Jayaram, H., Maeder, M. L., Skor, M. N., Wang, T., Myer, V. E., & Wilson, C. J. (2018). UDiTaS^{TM}, a genome editing detection method for indels and genome rearrangements. *BMC Genomics*, *19*(1), 212. https://doi.org/10.1186/s12864-018-4561-9

Gollapalli, P., Tamizh Selvan, G., Santoshkumar, H. S., & Ballamoole, K. K. (2022). Functional insights of antibiotic resistance mechanism in *Helicobacter pylori*: Driven by gene interaction network and centrality-based nodes essentiality analysis. *Microbial Pathogenesis*, *171*, 105737. https://doi.org/10.1016/j.micpath.2022.105737

Goodman, A. L., Wu, M., & Gordon, J. I. (2011). Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature Protocols*, *6*(12), 1969–1980. https://doi.org/10.1038/nprot.2011.417

Goossens, P., Baremans, K., Alderkamp, M., Boshoven, J. C., Ackerveken, G. van den, & Berendsen, R. L. (2024). *Selective enrichment of specific bacterial taxa in downy mildew-affected spinach: Comparative analysis in laboratory and field conditions* (p. 2024.08.23.609345). bioRxiv. https://doi.org/10.1101/2024.08.23.609345

Goossens, P., Spooren, J., Baremans, K. C. M., Andel, A., Lapin, D., Echobardo, N., Pieterse, C. M. J., Van Den Ackerveken, G., & Berendsen, R. L. (2023). *Congruent downy mildew-associated microbiomes reduce plant disease and function as transferable resistobiomes* [Preprint]. Microbiology. https://doi.org/10.1101/2023.03.14.532520

Goossens, P., Spooren, J., Baremans, K. C. M., Andel, A., Lapin, D., Echobardo, N., Pieterse, C. M. J., Van den Ackerveken, G., & Berendsen, R. L. (2023). Obligate biotroph downy mildew consistently induces near-identical protective microbiomes in Arabidopsis thaliana. *Nature Microbiology*, 1–16. https://doi.org/10.1038/s41564-023-01502-y

Granato, L. M., Picchi, S. C., Andrade, M. de O., Takita, M. A., de Souza, A. A., Wang, N., & Machado, M. A. (2016). The ATP-dependent RNA helicase HrpB plays an important role in motility and biofilm formation in Xanthomonas citri subsp. Citri. *BMC Microbiology*, *16*, 55. https://doi.org/10.1186/s12866-016-0655-1

Hall, A., Donohue, T., & Peters, J. (2023). Complete sequences of conjugal helper plasmids pRK2013 and pEVS104. *microPublication Biology*. https://doi.org/10.17912/micropub.biology.000882

HMM — TRANSIT v3.3.3 documentation. (n.d.). Retrieved 8 August 2024, from https://transit.readthedocs.io/en/latest/method_HMM.html

Ioerger, T. R. (2022). Analysis of Gene Essentiality from TnSeq Data Using Transit. *Methods in Molecular Biology (Clifton, N.J.)*, *2377*, 391–421. https://doi.org/10.1007/978-1-0716-1720-5_22

James C. Correll, Teddy E. Morelock, Mark C. Black, Steven T. Koike, Lynn P. Brandenberger, & Frank J. Dainello. (1994). *Economically Important Diseases of Spinach*. *Plant Disease*.

Jin, D., Sun, B., Zhao, W., Ma, J., Zhou, Q., Han, X., Mei, Y., Fan, Y., & Pei, Y. (2021). Thiamine-biosynthesis genes *Bbpyr* and *Bbthi* are required for conidial production and cell wall integrity of the entomopathogenic fungus *Beauveria bassiana*. *Journal of Invertebrate Pathology*, *184*, 107639. https://doi.org/10.1016/j.jip.2021.107639

Keinath, A. P., & de Figueiredo Silva, F. (2022). Economic impacts of reduced fungicide efficacy against downy mildew on slicing cucumber. *Crop Protection*, *155*, 105934. https://doi.org/10.1016/j.cropro.2022.105934

Kessler, B., de Lorenzo, V., & Timmis, K. N. (1992). A general system to integratelacZ fusions into the chromosomes of gram-negative eubacteria: Regulation of thePm promoter of theTOL

plasmid studied with all controlling elements in monocopy. *Molecular and General Genetics MGG*, *233*(1), 293–301. https://doi.org/10.1007/BF00587591

Kwon, Y. M., Ricke, S. C., & Mandal, R. K. (2016). Transposon sequencing: Methods and expanding applications. *Applied Microbiology and Biotechnology*, *100*(1), 31–43. https://doi.org/10.1007/s00253-015-7037-8

Lindsey, B. E., Rivero, L., Calhoun, C. S., Grotewold, E., & Brkljacic, J. (2017). Standardized Method for High-throughput Sterilization of Arabidopsis Seeds. *Journal of Visualized Experiments : JoVE*, *128*, 56587. https://doi.org/10.3791/56587

Liu, H., Bouillaut, L., Sonenshein, A. L., & Melville, S. B. (2013). Use of a Mariner-Based Transposon Mutagenesis System To Isolate Clostridium perfringens Mutants Deficient in Gliding Motility. *Journal of Bacteriology*, *195*(3), 629–636. https://doi.org/10.1128/JB.01288-12

Morinière, L., Lecomte, S., Gueguen, E., & Bertolla, F. (2021). In vitro exploration of the Xanthomonas hortorum pv. Vitians genome using transposon insertion sequencing and comparative genomics to discriminate between core and contextual essential genes. *Microbial Genomics*, *7*(6), 000546. https://doi.org/10.1099/mgen.0.000546

mprA - Response regulator MprA - Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) | UniProtKB | UniProt. (n.d.). Retrieved 21 August 2024, from https://www.uniprot.org/uniprotkb/P9WGM9/entry

Normalization—TRANSIT v3.3.3 documentation. (n.d.). Retrieved 14 August 2024, from https://transit.readthedocs.io/en/latest/method_normalization.html

Oomycetes—An overview | ScienceDirect Topics. (n.d.). Retrieved 21 August 2024, from https://www-sciencedirect-com.utrechtuniversity.idm.oclc.org/topics/agricultural-and-biological-sciences/oomycetes

Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., & Daniels, M. J. (1996). Characterization of eds1, a mutation in Arabidopsis suppressing resistance to Peronospora parasitica specified by several different RPP genes. *The Plant Cell*, *8*(11), 2033–2046. https://doi.org/10.1105/tpc.8.11.2033

Perry, B. J., & Yost, C. K. (2014). Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiology*, *14*, 298. https://doi.org/10.1186/s12866-014-0298-z

Rodenburg, S. Y. A., Seidl, M. F., de Ridder, D., & Govers, F. (2021). Uncovering the Role of Metabolism in Oomycete–Host Interactions Using Genome-Scale Metabolic Models. *Frontiers in Microbiology*, *12*, 748178. https://doi.org/10.3389/fmicb.2021.748178

Sharma, A., Gupta, V. K., & Pathania, R. (2019). Efflux pump inhibitors for bacterial pathogens: From bench to bedside. *The Indian Journal of Medical Research*, *149*(2), 129–145. https://doi.org/10.4103/ijmr.IJMR_2079_17

Skurnik, D., Roux, D., Aschard, H., Cattoir, V., Yoder-Himes, D., Lory, S., & Pier, G. B. (2013). A Comprehensive Analysis of In Vitro and In Vivo Genetic Fitness of Pseudomonas aeruginosa Using High-Throughput Sequencing of Transposon Libraries. *PLoS Pathogens*, *9*(9), e1003582. https://doi.org/10.1371/journal.ppat.1003582

Sun, Q., Wu, W., Qian, W., Hu, J., Fang, R., & He, C. (2003). High-quality mutant libraries of *Xanthomonas oryzae* pv. *Oryzae* and *X. campestris* pv. *Campestris* generated by an efficient transposon mutagenesis system. *FEMS Microbiology Letters*, *226*(1), 145–150. https://doi.org/10.1016/S0378-1097(03)00583-4

The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

TPP Overview—TRANSIT v1.1.2 documentation. (n.d.). Retrieved 12 August 2024, from https://transit2.readthedocs.io/en/stable/tpp.html#mapping-to-genomes-with-multiple-contigs

Transit/CHANGELOG.md at master · mad-lab/transit. (n.d.). GitHub. Retrieved 2 September 2024, from https://github.com/mad-lab/transit/blob/master/CHANGELOG.md

van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, *6*(10), 767–772. https://doi.org/10.1038/nmeth.1377

Vilhena, C., Kaganovitch, E., Shin, J. Y., Grünberger, A., Behr, S., Kristoficova, I., Brameyer, S., Kohlheyer, D., & Jung, K. (2018). A Single-Cell View of the BtsSR/YpdAB Pyruvate Sensing Network in Escherichia coli and Its Biological Relevance. *Journal of Bacteriology*, *200*(1), e00536-17. https://doi.org/10.1128/JB.00536-17

Vismans, G., Spooren, J., Pieterse, C. M. J., Bakker, P. A. H. M., & Berendsen, R. L. (2021). Soil-Borne Legacies of Disease in Arabidopsis thaliana. *Methods in Molecular Biology (Clifton, N.J.)*, *2232*, 209–218. https://doi.org/10.1007/978-1-0716-1040-4_17

Vismans, G., Van Bentum, S., Spooren, J., Song, Y., Goossens, P., Valls, J., Snoek, B. L., Thiombiano, B., Schilder, M., Dong, L., Bouwmeester, H. J., Pétriacq, P., Pieterse, C. M. J., Bakker, P. A. H. M., & Berendsen, R. L. (2022). Coumarin biosynthesis genes are required after foliar pathogen infection for the creation of a microbial soil-borne legacy that primes plants for SA-dependent defenses. *Scientific Reports*, *12*(1), 22473. https://doi.org/10.1038/s41598-022-26551-x

Voogdt, C. G. P., Tripathi, S., Bassler, S. O., McKeithen-Mead, S. A., Guiberson, E. R., Koumoutsi, A., Bravo, A. M., Buie, C., Zimmermann, M., Sonnenburg, J. L., Typas, A., Deutschbauer, A. M., Shiver, A. L., & Huang, K. C. (2024). Randomly barcoded transposon mutant libraries for gut commensals II: Applying libraries for functional genetics. *Cell Reports*, *43*(1), 113519. https://doi.org/10.1016/j.celrep.2023.113519

Wheatley, R. M., Ford, B. L., Li, L., Aroney, S. T. N., Knights, H. E., Ledermann, R., East, A. K., Ramachandran, V. K., & Poole, P. S. (2020). Lifestyle adaptations of Rhizobium from rhizosphere to symbiosis. *Proceedings of the National Academy of Sciences*, *117*(38), 23823–23834. https://doi.org/10.1073/pnas.2009094117

Wheatley, R. M., Ramachandran, V. K., Geddes, B. A., Perry, B. J., Yost, C. K., & Poole, P. S. (2017). Role of O_2 in the Growth of Rhizobium leguminosarum bv. Viciae 3841 on Glucose and Succinate. *Journal of Bacteriology*, *199*(1). https://doi.org/10.1128/JB.00572-16

Xu, D., & Zhang, Y. (2009). Generating Triangulated Macromolecular Surfaces by Euclidean Distance Transform. *PLoS ONE*, *4*(12), e8140. https://doi.org/10.1371/journal.pone.0008140

Xu, K., Lin, L., Shen, D., Chou, S.-H., & Qian, G. (2021). Clp is a “busy” transcription factor in the bacterial warrior, *Lysobacter enzymogenes*. *Computational and Structural Biotechnology Journal*, *19*, 3564–3572. https://doi.org/10.1016/j.csbj.2021.06.020

Yuan, J., Zhao, J., Wen, T., Zhao, M., Li, R., Goossens, P., Huang, Q., Bai, Y., Vivanco, J. M., Kowalchuk, G. A., Berendsen, R. L., & Shen, Q. (2018). Root exudates drive the soil-borne legacy of aboveground pathogen infection. *Microbiome*, *6*(1), 156. https://doi.org/10.1186/s40168-018-0537-x

Supplemental

Primer sequences

Table 8. *Primers for checking transposon insertions*

Name	Sequence (5'->3')	Purpose	Source /Reference
pSAM-Tnase_Fw (BF009)	ATGAGTTTCCGGACTCTGCC	Amplify transposase gene	This study
pSAM-Tnase_Rv (BF010)	TTAACAGCTGCAAACACCGC	Amplify transposase gene	This study
KanR_Fw (BF041)	GATTGCACGCAGGTTCTCCG	Amplify kanamycin resistance gene inside transposon	This study
KanR_Rv (BF042)	TCCTGATCGACAAGACCGGC	Amplify kanamycin resistance gene inside transposon	This study
pSAM_INS_Fw (BF043)	TCCTGATCGACAAGACCGGC	Amplify mariner transposon	This study
pSAM_INS_Rv (BF044)	ACTCAGGAGAGCGTTCACCG	Amplify mariner transposon	This study

Table 9*. Primers INSeq library preparation*

Name	Sequence (5'->3')	Purpose	Source /Reference
BioSamA (BF034)	Bio-TEG-CGGTTCGCTTGCTGTCCATAAAAC	Linear PCR	(Perry & Yost, 2014)
M12_top (BF037)	CTGTCCGTTCCGACTACCCTCCCGAC	Mme1 digestion	(Goodman et al., 2011)
M12_bottom (BF038)	GTCGGGAGGGTAGTCGGAACGGACAG	Mme1 digestion	(Goodman et al., 2011)
INSeq-MOD-Adpt-Top (BF047)	ATCCACGGTAGCATCAAATGCGGATA	LibAdapt, Amplify insertion sites	(Knights et al., (2024), unpublished data)
INSeq-MOD-Adpt-Bot (BF048)	TATCCGCATTTGATGCTACCGTGGATNN	LibAdapt, Amplify insertion sites	(Knights et al., (2024), unpublished data)
INSeq-Novo-Final-For (BF045)	ATAAAACCGCCCAGTCTACTCGAGGG	Final PCR	(Knights et al., (2024), unpublished data)
INSeq-Novo-Final-Rev	TATCCGCATTTGATGCTACCGTGGAT	Final PCR	(Knights et al., (2024),

(BF046)			unpublished data)
INSeq-Final-Test-Rev (BF049)	ATAAGTCCCCGGTCTTCGTATGCC	Troubleshooting: Test DNA presence on beads. Replaces BF046.	This study

Scripts INSeq analysis

Trimming reads - AdapTrimming_outside_map_EK.sh

```
#!/bin/bash
set -e

# adds conda initialisation
source /opt/miniconda3/etc/profile.d/conda.sh
#make sure your paired files are named like "Sample1_1_raw.fq" and "Sample1_2_raw.fq".
"Sample1_" will be the "basename"
for fq1 in RawReads/*1_raw.fq
do
        echo "working with $fq1"
        base=$(basename $fq1 _1_raw.fq)

        echo "base is $base"

        mkdir -p Trimmed_Reads/TL_TS_${base}
        mkdir -p Trimmed_Reads/Trimmed_${base}

        fq1=RawReads/${base}_1_raw.fq
        fq2=RawReads/${base}_2_raw.fq

        fq1cut=Trimmed_Reads/Trimmed_${base}/Trimmed_${base}_1_cut.fq
        fq2cut=Trimmed_Reads/Trimmed_${base}/Trimmed_${base}_2_cut.fq
        fq2cutRV=Trimmed_Reads/Trimmed_${base}/Trimmed_${base}_2_cutRV.fq

        fq1TL=Trimmed_Reads/TL_TS_${base}/TL_TS_${base}_1_TL.fq
        fq2TL=Trimmed_Reads/TL_TS_${base}/TL_TS_${base}_2_TL.fq
        fq1TS=Trimmed_Reads/TL_TS_${base}/TL_TS_${base}_1_TS.fq
        fq2TS=Trimmed_Reads/TL_TS_${base}/TL_TS_${base}_2_TS.fq

        conda activate /home/elianca/.conda/envs/cutadapt

        cutadapt  -a
CATTTGATGC...ACAGGTTGGATGATAAGTCCCCGGTCTTCGTATGCCGTCTTCTGCTTGGCGCGCCCTCGAG
TAGACTGGGCGGTTTTAT -A
ATAAAACCGCCCAGTCTACTCGAGGGCGCGCCAAGCAGAAGACGGCATACGAAGACCGGGGACTTATCAT
CCAACCTGT...GCATCAAATG --revcomp --discard-untrimmed -m 25 -M 27 --too-long-output $fq1TL
--too-long-paired-output $fq2TL --too-short-output $fq1TS --too-short-paired-output $fq2TS -q 20
-Q 20 -o $fq1cut -p $fq2cut $fq1 $fq2

        conda activate /home/elianca/.conda/envs/bbmap-env
```

```
        reformat.sh in=$fq2cut out=$fq2cutRV rcomp

        cat $fq1cut $fq2cutRV > Trimmed_Reads/Trimmed_${base}/${base}_comb_cut.fq

        done
```

Mapping to reference genome - TPP_mapping_1mm_fixed_EK.sh

```
#!/bin/bash
#GOOD script!!! flags
set -e

# adds conda initialisation
source /opt/miniconda3/etc/profile.d/conda.sh
# run inside map 1mm_analysis
for fq1 in ../Trimmed_Reads/Trimmed_*/*comb_cut.fq  #output Trimming script =
../Trimmed_Reads/Trimmed_${base}/${base}_comb_cut.fq
do
        echo "working with $fq1"
        base=$(basename $fq1 _comb_cut.fq)

        echo "base is $base"

        mkdir -p TPP_output_${base}

        #fq1Trimmed=Trimmed_${base}/${base}_comb_cut.fq
        fqTrimmed="../../Trimmed_Reads/Trimmed_${base}/${base}_comb_cut.fq"
        fqTPP="${base}_comb_TPP"

        conda activate /opt/miniconda3/envs/transit-env

        cd TPP_output_${base}

  echo "Executing TPP with inputs: $fq1, $fqTPP"
# 1mm allowed
        python3 /opt/miniconda3/envs/transit-env/bin/tpp -bwa /usr/bin/bwa -ref
/media/TB4/Brandon/inseq/seqdata/batch1/WCS2014-23Ref/Ref_fixed/WCS2014-
23_OneChr_Ref_fixed.fna -reads1 "$fqTrimmed" -output "$fqTPP" -protocol Mme1 -primer
TACCGTGGATG -bwa-alg aln -flags "-l 13 -n 1"
        cd ..
done
```

Hidden Markov Model (HMM) – Transit_HMM_betageom_fixed_EK.sh

```
#!/bin/bash

set -e

# adds conda initialisation
source /opt/miniconda3/etc/profile.d/conda.sh
#run inside the map where you want the output, e.g. Transit_betageon
```

```
for fq1 in ../TPP_output_*/*comb_TPP.wig  #output TPP_mapping_script = "${base}_comb_TPP"
under which a .wig file is created
do
        echo "working with $fq1"
        base=$(basename $fq1 _comb_TPP.wig)

        echo "base is $base"

        mkdir -p Transit_output_${base}

        #fq1Trimmed=Trimmed_${base}/${base}_comb_cut.fq
        fqWig="../TPP_output_${base}/${base}_comb_TPP.wig"
        fqHMM="Transit_output_${base}/${base}_TranOutB.txt"
        RefGenomeAnno="/media/TB4/Brandon/inseq/seqdata/batch1/WCS2014-
23Ref/Ref_fixed/WCS2014-23_OneChr_fixed_Anno.prot_table"
        conda activate /opt/miniconda3/envs/transit-env

        echo "Executing Transit with input: $fqWig"

        python3 /opt/miniconda3/envs/transit-env/bin/transit hmm "$fqWig"
"$RefGenomeAnno" "$fqHMM" -iN 10 -iC 10 -n betageom

done
```

Adding confidence scores to HMM - Add_conf_HMM_EK3.sh

```
#!/bin/bash
set -e

# adds conda initialisation
source /opt/miniconda3/etc/profile.d/conda.sh

for fq1 in *_TranOutB_genes.txt
do
        echo "Calculating the confidences of $fq1"
        base=$(basename $fq1 _TranOutB_genes.txt)

        echo "base is $base"

        mkdir -p Transit_b_conf

        HMM_comf="Transit_b_conf/${base}_genes_conf.txt"

        python3 ~/HMM_conf.py "$fq1" > "$HMM_comf"

done
```

HMM_conf.py = script from: (Transit/CHANGELOG.Md at Master · Mad-Lab/Transit, n.d.)

Statement regarding GenAI

For writing my report I sometimes used GenAI tool "OpenAI Chat-GPT 3.5 (2024)" to lookup English grammar options and incidental to reformulate/restructure sentences. For the INSeq analysis I used the same GenAI tool to look up the meanings of coding errors during the trouble shooting process. Comparable to how you would use Google to look up these errors.