DISS. ETH NO. 29476

Resistance gene identification through advanced host-pathogen genomics in the *Xanthomonas*-ryegrass pathosystem

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

FLORIAN EUGÈNE GOETTELMANN

Master en Sciences, Technologies, Santé, Université de Nice-Sophia Antipolis

born on 13.06.1993

accepted on the recommendation of

Prof. Dr. Bruno Studer, examiner

Dr. Roland Kölliker, co-examiner

Prof. Dr. Jens Boch, co-examiner

2023

Table of contents

Summary

Grasslands are crucial ecosystems that provide various ecosystem services and are a key component in livestock production. In these systems, Italian ryegrass (*Lolium multiflorum* Lam.), is among the most widely used species due to its high yield, fast ground cover, and great nutritional quality. Its high adaptation to mowing makes it a species of choice for fodder in the form of hay and silage. One of the main diseases that affect *L. multiflorum* is bacterial wilt, caused by *Xanthomonas translucens* pv. *graminis* (*Xtg*), which can lead to severe yield losses. To date, breeding for resistant cultivars is the only feasible way for an efficient and durable control of the disease. Cultivars with increased resistance have been obtained by recurrent phenotypic selection, but due to the outbreeding nature of ryegrasses, cultivars are highly heterozygous, and susceptibility still occurs. Precisely identifying genetic loci involved in *Xtg* resistance in order to specifically target them in genomics-assisted breeding would be crucial to develop new cultivars that are highly resistant to *Xtg*. Furthermore, identifying molecular components that are involved in the host-pathogen interaction is essential to better understand the disease and develop new ways to control infection. Therefore, this thesis aimed at identifying resistance genes in *L. multiflorum* and genes involved in pathogenicity in *Xtg* by using advanced genomic-based analyses.

Chapter 1 starts with a general introduction about the *L. multiflorum* – *Xtg* pathosystem and the state of the art about their host-pathogen interaction. It then details the currently available means to control the disease and the importance of genomics-based tools to better understand this interaction and identify sources of resistance in order to breed for resistant cultivars.

Chapter 2 describes a first comparative analysis within *X. translucens* aimed at characterizing the species and identifying pathovar-specific features. Indeed, like most *Xanthomonas* species, *X. translucens* is divided into pathovars, defined by their host range. The species contains pathovars that infect cereals, as well as pathovars that infect forage grasses, like *Xtg*. In this study, we produced complete genome sequences for all eleven pathotype strains of *X. translucens* using long read sequencing technology, enabling a comprehensive comparative genomic analysis within the species. This revealed that the species is genetically separated into three distinct clades, with *Xtg* being closely related to the four other forage grasses-infecting pathovars. A total of 2,181 genes were found to be shared by all pathovars, while 190, 588 and 168 genes were exclusive to each clade, respectively. The sequenced *Xtg* strain was found to have the largest set of genes in the species, and a high number of pathovar-specific genes. Furthermore, the *Xtg* strain lacked many virulence features commonly present in other strains of the species, highlighting the particularity of *Xtg*. This study laid the foundations for further in-depth analyses within the *X. translucens* species.

Chapter 3 reports the production of four additional complete genome sequences for *Xtg* strains, allowing a comprehensive comparative genomic analysis of *Xtg*. This revealed that while *Xtg* strains were highly related, with 99.9 to 100% average nucleotide identity, they displayed a high genome plasticity, with many chromosomal rearrangements between strains. This plasticity is likely due to the presence of many mobile genetic elements, with 413 to 457 insertion/excision elements identified per *Xtg* strain compared to 24 to 123 in other pathovars of the species. Genes that were specific to *Xtg* were identified, including *XopE* and *XopX* class effectors, a unique set of minor pilins of the type IV pilus, 17 TonB-dependent receptors, and *gumP*, a gene associated with the *gum* xanthan synthesis gene cluster. These constitute a set of potential virulence factors that could be involved in the pathogenicity of *Xtg*.

Chapter 4 describes a bulked segregant analysis approach aimed at fine-mapping a previously identified quantitative trait locus for resistance to *Xtg*. To this end, a large (n = 7,484) population was screened and the 750 most resistant and 761 most susceptible individuals were selected to form two pools. These pools were sequenced by whole-genome deep sequencing, allowing a high resolution for fine-mapping. This revealed a 300 kbp region that was highly associated with *Xtg* resistance and contained many genes with a potential role in disease resistance, including genes encoding for Pik2-, RGA4- and RGA5-like disease resistance proteins, cysteine-rich kinases, and serine protease inhibitors. This region and the underlying genes and genetic markers hold great potential to breed cultivars with increased resistance.

Finally, **Chapter 5** discusses the results obtained in this work and expands on the future perspectives to validate the identified genes for resistance to *Xtg*, define markers for marker-assisted selection, and identify new sources of resistance for a more efficient and durable control of *Xtg*.

Résumé

Les prairies sont des écosystèmes cruciaux qui fournissent divers services écosystémiques et constituent un élément clé de la production animale. Dans ces systèmes, le ray-grass italien (*Lolium multiflorum* Lam.) est l'une des espèces les plus utilisées en raison de son rendement élevé, de sa couverture rapide du sol et de sa grande qualité nutritionnelle. Sa grande adaptation à la fauche en fait une espèce de choix pour le fourrage sous forme de foin et d'ensilage. L'une des principales maladies affectant *L. multiflorum* est le flétrissement bactérien, causé par *Xanthomonas translucens* pv. *graminis* (*Xtg*), qui peut entraîner de graves pertes de rendement. À ce jour, la sélection de cultivars résistants est le seul moyen envisageable pour lutter efficacement et durablement contre la maladie. Des cultivars présentant une résistance accrue ont été obtenus par sélection phénotypique récurrente, mais en raison de la nature allogame des ray-grass, les cultivars sont fortement hétérozygotes et la sensibilité à la maladie persiste. L'identification précise des loci génétiques impliqués dans la résistance à *Xtg* afin de les cibler spécifiquement dans la sélection assistée par la génomique serait cruciale pour développer de nouveaux cultivars hautement résistants à *Xtg*. En outre, l'identification des composants moléculaires impliqués dans l'interaction hôte-pathogène est essentielle pour mieux comprendre la maladie et développer de nouvelles méthodes de contrôle de l'infection. Par conséquent, cette thèse vise à identifier des gènes de résistance chez *L. multiflorum* et des gènes impliqués dans la pathogénicité de *Xtg* en utilisant des analyses génomiques avancées.

Le **Chapitre 1** commence par une introduction générale sur le pathosystème *L. multiflorum* – *Xtg* et l'état de l'art sur leur interaction hôte-pathogène. Il détaille ensuite les moyens actuellement disponibles pour lutter contre la maladie et l'importance des outils génomiques pour mieux comprendre cette interaction et identifier des sources de résistance afin de sélectionner des cultivars résistants.

Le **Chapitre 2** décrit une première analyse comparative au sein de *X. translucens* visant à caractériser l'espèce et à identifier les caractéristiques spécifiques à chaque pathovar. En effet, comme la plupart des espèces de *Xanthomonas*, *X. translucens* est divisé en pathovars, définis par leur gamme d'hôtes. L'espèce contient des pathovars qui infectent les céréales, ainsi que des pathovars qui infectent les graminées fourragères, comme *Xtg*. Dans cette étude, nous avons produit des séquences génomiques complètes pour les onze souches pathotypes de *X. translucens* en utilisant une technologie de séquençage à longue lectures, ce qui a permis une analyse génomique comparative complète au sein de l'espèce. Cette analyse a révélé que l'espèce est génétiquement séparée en trois clades distincts, *Xtg* étant très proche des quatre autres pathovars infectant les graminées fourragères. Au total, 2 181 gènes étaient partagés par tous les pathovars, tandis que 190, 588 et 168 gènes étaient exclusifs à chaque clade, respectivement. La souche de

Xtg séquencée possédait le plus grand nombre de gènes de l'espèce, ainsi qu'un grand nombre de gènes spécifiques au pathovar. En outre, de nombreuses caractéristiques de virulence communément présentes dans d'autres souches de l'espèce étaient absentes de la souche de *Xtg*, ce qui souligne la particularité de *Xtg*.

Le **Chapitre 3** fait état de la production de quatre séquences génomiques complètes supplémentaires pour des souches de *Xtg*, ce qui a permis une analyse génomique comparative complète du pathovar. Cette analyse a révélé que si les souches *Xtg* étaient très proches, avec une identité nucléotidique moyenne de 99,9 à 100 %, elles présentaient une grande plasticité génomique, avec de nombreux réarrangements chromosomiques entre les souches. Cette plasticité est probablement due à la présence de nombreux éléments génétiques mobiles, avec 413 à 457 éléments d'insertion/excision identifiés par souche de *Xtg*, contre 24 à 123 dans d'autres pathovars de l'espèce. Des gènes spécifiques à *Xtg* ont été identifiés, notamment des effecteurs de classe *XopE* et *XopX*, un ensemble unique de pilines mineures du pilus de type IV, 17 récepteurs dépendant de TonB, et *gumP*, un gène associé au cluster de gènes de synthèse de xanthane. Ces gènes constituent un ensemble de facteurs de virulence potentiels qui pourraient être impliqués dans la pathogénicité de *Xtg*.

Le **Chapitre 4** décrit une approche d'analyse de ségrégants en mélange visant à affiner la localisation d'un locus de trait quantitatif pour la résistance à *Xtg* précédemment identifié. A cette fin, une grande population (n = 7 484) a été criblée et les 750 individus les plus résistants et les 761 individus les plus sensibles ont été sélectionnés pour former deux pools. Ces pools ont été séquencés par séquençage profond du génome entier, ce qui a permis d'obtenir une résolution élevée. Cela a révélé une région de 300 kbp fortement associée à la résistance à *Xtg* contenant de nombreux gènes ayant un rôle potentiel dans la résistance à la maladie, y compris des gènes codant pour des protéines de résistance à la maladie de type Pik2-, RGA4- et RGA5-like, des kinases riches en cystéine, et des inhibiteurs de protéases à sérine. Cette région et les gènes et marqueurs génétiques sous-jacents offrent un grand potentiel pour la sélection de cultivars présentant une résistance accrue.

Enfin, le **Chapitre 5** discute des résultats obtenus dans ce travail et développe les perspectives futures pour valider les gènes identifiés pour la résistance à *Xtg*, définir des marqueurs pour la sélection assistée par marqueurs, et identifier de nouvelles sources de résistance pour un contrôle plus efficace et durable de *Xtg*.

1. General introduction

Grasslands are among the largest ecosystems in the world and cover a very large part of the Earth's surface. According to the Food and Agriculture Organization (FAO), permanent meadows and pastures represent 67% of the global agricultural area and 24% of the global land area in 2020 (FAO, 2016). They are the main source of feed for ruminants, and as such are a key component of livestock production. Furthermore, permanent grasslands are mostly found in marginal agricultural land that is not suitable for other crops. Because of this, grass-fed livestock makes an efficient use of this land to provide animal products without competition with non-animal food production, contrary to concentrate-fed livestock. Additionally, temporary grasslands included in crop rotations are a significant component of arable crop systems and can provide benefits to the following crops by improving soil structure and soil biodiversity (Hoeffner et al., 2021). Overall, grasslands are important in improving soil quality, preventing soil erosion, storing carbon in the soil, and as a habitat for wildlife (Zhao et al., 2020).

The world's human population keeps growing, estimated to reach 9.77 billion by 2050 (United Nations Department of Economic and Social Affairs, 2018), and with it the demand for animal products will increase (OECD and FAO, 2018). Furthermore, concerns have been raised about the role of animal production in driving climate change and other environmental impacts, with ruminant livestock having the largest impact on greenhouse gas emissions, land and energy use, acidification, and eutrophication (Clark and Tilman, 2017). In turn, climate change has a direct effect on agricultural production systems through changes in temperature and precipitations, which can lead to increased drought, as well as an increased frequency of other extreme weather events that eventually affect productivity (Hopkins and Del Prado, 2007; Emadodin et al., 2021). Moreover, climate change may render crops more susceptible to pathogens, as it can promote conditions that are favorable to their development, shift or expand their geographical range, and impact the ability of plants to resist them (Hunjan and Lore, 2020). Therefore, developing livestock production practices that are sustainable and resilient to climate change represents one of the most important challenges that modern agriculture is facing. One key component of this development of sustainable grassland-based livestock production is the use of highly productive forage species such as ryegrasses.

1.1. *Lolium multiflorum*

Ryegrasses (*Lolium* spp.) are members of the Poaceae family, which also contains many major cereal crops such as rice (*Oryza sativa* L.), wheat (*Triticum* spp.) and barley (*Hordeum vulgare* L.). Within Poaceae, the genus *Lolium* is found in the Poeae tribe, together with many grass genera such as *Avena* (oats), *Dactilys* (orchardgrasses or cocksfoot grasses), *Poa* (bluegrasses), *Phleum* (timothy), or *Festuca* (fescues) (Soreng et al., 2015). Species of the genera *Lolium* and *Festuca* are very close genetically, and form together the *Festuca-Lolium* complex (Yamada et al., 2005). Naturally occurring hybrids of both

species, which combine the high yields and nutritional quality of ryegrasses with the resilience of fescues against abiotic stresses, can be found, classified under the *x Festulolium* nothogenus (Ghesquière et al., 2010).

Ryegrasses such as Italian ryegrass (*L. multiflorum* Lam.) and perennial ryegrass (*L. perenne* L.) are among the most widely used species in temperate grasslands due to their high yield potential, fast ground cover, adaptation to grazing and mowing, and great forage quality features such as palatability, digestibility and high nutritional value (Frame, 1991; Bernard et al., 2002). While *L. perenne* is highly tolerant to grazing and is therefore generally grown on pastures, *L. multiflorum* is rather grown for fodder because of its low tolerance to grazing and to summer and winter stresses. They are often used in combination with other grasses or with legumes such as red and white clover (*Trifolium pratense* L., *T. repens* L.) or lucerne (*Medicago sativa* L.). These grass-legume combinations have shown numerous advantages, including an increase of the nitrogen uptake of both the grass and the legume, compared to monocultures of each (Nyfeler et al., 2011; Suter et al., 2015).

L. multiflorum is a cool-season annual grass native to southern Europe. It has a lower persistence and stress tolerance than *L. perenne*, but a higher yield potential and a faster ground cover. The *L. multiflorum* species is divided into two subspecies: *L. multiflorum* ssp. *italicum* (Italian ryegrass) and *L. multiflorum* ssp. *multiflorum* (Westerwold ryegrass). The former is biennial to pluriannual and requires vernalization and long days to produce reproductive tillers, while the latter is annual, with higher yields but lower persistency (Humphreys et al., 2010).

Goals in breeding of *L. multiflorum* are focused on increasing dry matter and seed yield, improving nutritional quality (*i.e.* better digestibility, better palatability, higher content in carbohydrates, proteins and lipids, as well as other nutrients), and improving resistance to biotic and abiotic stresses. Diseases that can affect ryegrasses' yields and quality include rusts caused by *Puccinia* spp., snow mold caused by *Microdochium* spp. and *Typhula* spp., leaf spots caused by *Bipolaris* spp. and *Drechslera* spp., leaf blotch caused by *Rhynchosporium* spp., and bacterial wilt caused by *Xanthomonas translucens* pv. *graminis* (Michel et al., 2000).

L. multiflorum is an outbreeding species, and as a result, it is bred as highly heterozygous populations. The most common strategy in breeding *L. multiflorum* is recurrent phenotypic selection, consisting of multiple cycles of evaluating a population and selecting the best individuals to cross for the next generation (Humphreys et al., 2010; Nay et al., 2022). After several cycles, the best individuals are selected, cloned, and planted in the field for a polycross where each plant will get pollinated by multiple other plants. The resulting progeny is then tested in multisite trials and may result in a new synthetic variety that will be released to the market.

More recently, the progress of genetic tools allowed the development of breeding techniques, such as marker-assisted selection (MAS), that can assist in targeting specific genetic loci associated with a phenotype of interest (Collard and Mackill, 2007). MAS is of particular interest in selecting for traits that are difficult or expensive to evaluate, when several genes need to be selected at once, or when selecting for or against recessive alleles (Xu and Crouch, 2008). It is of particular interest in outbreeding species such as *L. multiflorum*, where the high heterozygosity maintains undesirable recessive alleles, hindering the ability of classical selection to fix desirable alleles.

Due to its impact on *L. multiflorum*, resistance to *Xanthomonas* has been integrated as a main criteria for selection in breeding schemes such as that performed at Agroscope Reckenholz, Switzerland, with plants being inoculated at each cycle of recurrent selection (Humphreys et al., 2010).

1.2. *Xanthomonas* spp.

The genus *Xanthomonas* comprises at least 20 species of Gram-negative gamma-proteobacteria that are rod shaped, yellow pigmented, obligate aerobe, and many of them possess a single polar flagellum which renders them motile (Vauterin et al., 1995; Ryan et al., 2011). Most *Xanthomonas* species are phytopathogenic, causing spots, streaks, necrosis, gummosis, wilting, rots and dieback diseases on leaves, stems or fruits (Fig. 1). Their host range covers at least 124 monocotyledonous and 268 dicotyledonous plant species, including many cultivated species (Leyns et al., 1984). They are among the most damaging pathogens in agriculture, causing severe economic losses throughout the world (Leyns, 1993). Members of the genus *Xanthomonas* generally have a narrow host range and *Xanthomonas* species are further subdivided into pathovars according to their host range. Most of them persist as epiphytes on the plant surface and enter the plant via wounds, cracks or natural openings like stomata, hydathodes, lenticels, and nectaries (Rudolph, 1993). After penetration into the host plant, the bacteria start to develop within the parenchymal intercellular space, and then reach the xylem, allowing them to spread through the whole plant (Rudolph, 1993). *Xanthomonas* species produce an extracellular polysaccharide (EPS) called xanthan, which is characteristic to the genus (Rudolph et al., 1994). As it is secreted into the extracellular space, it leads to a mucoid slime, which can sometimes be extruded from the lesions of the plant (Fig. 1B). Due to its properties, it is used as a thickening agent and stabilizer in food products, among other uses. This slime also has a characteristic yellow color due the pigment xanthomonadin produced by the bacteria. This pigment however does not seem to have an importance in the pathogenesis, but might play a role in protection against UV light and oxidative stress, thus promoting epiphytic survival (Rajagopal et al., 1997; Poplawsky et al., 2000; He et al., 2020).

Figure 1 Symptoms caused by different *Xanthomonas* spp. in their respective plant host. **a)** *X. oryzae* causing bacterial blight in rice. **b**) *X. campestris* pv. *musacearum* causing banana *Xanthomonas* wilt **c)** *X. citri* in citrus. **d**) *X. axonopodis* pv. *mangiferaindicae* causing mango black spot disease. **e)** Four *Xanthomonas* spp. are associated with bacterial spot disease in tomato and pepper: *X. cynarae* pv. *gardneri*, *X. euvesicatoria*, *X. perforans* and *X. vesicatoria*. **f)** *X. translucens* pv. *graminis* causing bacterial wilt in Italian ryegrass (left) compared to a healthy uninfected plant (right). Adapted from (Timilsina et al., 2020).

1.2.1. *Xanthomonas translucens*

Xanthomonas translucens is a species that mainly infects Poaceae, although strains infecting pistachio (*Pistacia vera* L.) have recently been classified as a pathovar of *X. translucens* (Giblot-Ducray et al., 2009). To date, eleven pathovars of *X. translucens* have been defined based on their host range, and two major groups can be distinguished. The "translucens" group comprises pathovars infecting cereals, such as rye (*Secale cereale* L.), wheat, and barley (pv. *cerealis*, *hordei*, *secalis*, *translucens*, and *undulosa*), as well as a pathovar infecting pistachio (pv. *pistaciae*), while the "graminis" group contains pathovars infecting forage grasses (pv. *arrhenatheri*, *graminis*, *phlei*, *phleipratensis*, and *poae*) (Sapkota et al., 2020). Recent genome sequencing data supports this distinction but suggests pv. *cerealis* may be genetically distant from the "translucens" group and constitute a third group (Rademaker et al., 2006; Peng et al., 2016; Langlois et al., 2017). Furthermore, strains isolated from pistachio and classified as pv. *pistaciae* were found to be separated in two subgroups, referred to as subgroups A and B. However, molecular phylogenies found that strains of subgroup A were closely related to other strains of the "translucens" group, but strains of subgroup B were closely related to pv. *cerealis*, suggesting these two subgroups may not be directly related (Facelli et al., 2005; Marefat et al., 2006; Giblot-Ducray et al., 2009). So far, phylogenies of *X. translucens* have been based on a few DNA sequences such as the *gyrB* gene, and no whole genome sequence-based phylogenies have been performed to confirm the genetic relationships between pathovars of the species.

1.2.2. *Xanthomonas translucens* pv. *graminis*

Xanthomonas translucens pv. *graminis* (*Xtg*), initially named *X. graminis*, was first described in 1975 as the bacterial agent causing wilt in forage grasses (Egli et al., 1975). It was later renamed to *X. campestris* pv. *graminis* (Young et al., 1978) and then to *X. translucens* pv. *graminis* (Vauterin et al., 1995). Later, three other pathovars of *X. translucens* causing wilt in forage grasses were distinguished from *Xtg*, according to their host range (Egli and Schmidt, 1982). Pathovars *arrhenatheri*, *phlei*, and *poae* are restricted to the genera *Arrhenatherum*, *Phleum*, and *Poa*, respectively, while *Xtg* has a wide host range, and can infect many grasses from genera such as *Agrostis*, *Alopecurus*, *Dactylis*, *Deschampsia*, *Festuca*, *Lolium*, *Phalaris*, *Phleum*, *Poa* and *Trisetum*, with *Lolium* and *Festuca* being the most susceptible both in the greenhouse and in the field (Egli et al., 1975; Egli and Schmidt, 1982; Hersemann, 2016). A fifth pathovar, *X. t.* pv. *phleipratensis*, has been found to infect forage grasses, however the disease manifests as leaf spots and not as wilting (Stead, 1989). Of all the forage grasses-infecting *X. translucens* pathovars, *Xtg* is also the most widespread. It was first identified in France, Germany and Switzerland (Egli et al., 1975), but was later found in Holland (Labruyere, 1975), England (Wilkins and Exley, 1977), Belgium (Leyns et al., 1981; Van Den Mooter et al., 1981), Scotland (Channon and Hissett, 1984), Norway (Sletten, 1989), New Zealand (Egli and Schmidt, 1982; Falloon and Hume, 1988), and north America (Roberts et al., 1985).

No seed or soil transmission of the disease has been observed, and spraying the leaves with a bacterial solution resulted in limited infection (Leyns et al., 1988). Masses of *Xtg* were observed by electron microscopy around stomata after spraying, however no symptoms appeared on the plant (Wang and Sletten, 1995) and another study using electron microscopy found no particular concentration around stomata after spraying (De Cleene, 1989). This suggests that penetration of *Xtg* through natural openings may occur but is not the main way of entry. *Xtg* transmission is considered to occur mostly via wounds (Egli et al., 1975; Channon and Hissett, 1984; De Cleene, 1989; Wang and Sletten, 1995). Being cut frequently for harvesting makes *L. multiflorum* particularly exposed to the disease. Each mowing creates wounds, which can be used by the pathogen as an entry point into the plant, with the number of infected plants increasing throughout the growing season after each cut. The range of susceptibility of *L. multiflorum* to *Xtg* is very wide, but Westerwold ryegrasses seem to be most susceptible (Schmidt and Nuesch, 1980), while tetraploid cultivars seem less susceptible than diploid cultivars (Paul and Freudenstein, 1989). As it is one of the main species grown for forage, infection of *L. multiflorum* by *Xtg* is an important concern. It causes variable but serious yield losses, accounting for up to 47% in some susceptible cultivars (Gondran and Betin, 1989). About 40% yield loss has been observed for the later cuts in susceptible cultivars, corresponding to about 20% yield loss across the entire growing season (Schmidt and Nuesch, 1980).

After infection by *Xtg*, young leaves start to curl and wither (Fig. 1F). Wilting starts at the tip of the leaf and progresses towards the base of the plant. As *Xtg* colonizes the xylem, yellow stripes can be seen along the vascular system. In favorable conditions, the plant may die within a few days (Leyns, 1993). Given the significant impact of the disease on *L. multiflorum*, it is essential to better understand the underlying molecular components involved in the infection in order to devise ways to control the disease.

1.3. Molecular mechanisms of virulence in *Xtg*

Successful infection by pathogenic bacteria involves adhesion to the plant surface, invasion of the host tissue, nutrient acquisition, and suppression of plant defense responses. Each of these steps is mediated by a set of virulence factors, such as adhesins and exopolysaccharides that promote adhesion and biofilm formation, degradative enzymes that break down host tissues, or effectors that manipulate host processes.

1.3.1. Adhesion, biofilm formation and motility

Prior to infection, Xanthomonads survive on the plant by attaching to the leaf surface (Rudolph, 1993; Mhedbi-Hajri et al., 2011). For example, *X. oryzae* pathovars have been shown to adhere to trichomes in rice, while *Xtg* did not, in both rice and *L. multiflorum* (De Cleene, 1989). Masses of *Xtg* cells could be seen in epidermal depressions and even on the whole leaf, however, cell surface adhesion in *Xtg* has not been further investigated so far. Following attachment, the bacteria aggregate to form a biofilm, which allows them to resist biotic and abiotic stresses (Fazeli-Nasab et al., 2022). Attachment and biofilm formation are mediated by EPS and extracellular structures embedded in the bacterial membrane called adhesins (Mhedbi-Hajri et al., 2011). Furthermore, it was also shown that lipopolysaccharides may play a role in adhesion and biofilm formation (Li and Wang, 2011; Petrocelli et al., 2012).

Adhesins are grouped in two major classes: fimbrial adhesins (or pili), consisting in heteropolymers, and non-fimbrial adhesins, consisting in a single protein or homotrimers (Gerlach and Hensel, 2007). Mutants of both fimbrial and non-fimbrial adhesins showed a reduced biofilm formation in *Xylella fastidiosa*, indicating they both play a role in this process (Feil et al., 2007). Fimbrial adhesins mediate adhesion to the host through specific interaction with receptors on the host cell surface, but were also suggested to be involved in cell to cell aggregation (Feil et al., 2007). Non-fimbrial adhesins include filamentous hemagglutinins, the autotransporter-like protein YapH, or other proteins such as XadA or XadB and are involved in binding to the host surface (Feil et al., 2007; Ryan et al., 2011).

Together with adhesins, EPS such as xanthan are a key component in adhesion and biofilm formation and in epiphytic survival (Kemp et al., 2004; Rigano et al., 2007). Xanthan is a heteropolysaccharide consisting of a cellulose backbone with trisaccharide side-chains of mannose-glucuronate-mannose. Its

production is controlled by the *gum* gene cluster, consisting of twelve genes (*gumB* to *gumM*), responsible for its formation and excretion (Katzen et al., 1998). In *X. translucens*, the *gum* cluster has been found, however it lacks the *gumG* gene (Katzen et al., 1998; Wichmann et al., 2013; Falahi Charkhabi et al., 2017). In *Xtg*, xanthan is considered to be responsible for the wilting in forage grasses, as the accumulation of EPS results in the clogging of the xylem and prevents water uptake from the soil (Rudolph, 1993). It was also shown that EPS preparations extracted from *X. oryzae* exhibit some toxicity and induced wilting in rice and other non-host plants, without the presence of the bacteria, indicating that they might play a role in pathogenicity (Kuo et al., 1970). Finally, it has also been suggested EPS might play a role in suppressing plant immunity by chelating calcium ions (Aslam et al., 2008).

Following attachment and biofilm formation, the bacteria will penetrate the plant via wounds or natural openings. As the *Xanthomonas* genus lacks active mechanisms for penetration, motility towards these entry points is crucial in the infection. This motility is achieved mainly by the polar flagellum, which has been suggested to also play a role in biofilm formation (Athinuwat et al., 2018). Interestingly, *Xtg* strains, unlike closely related *X. translucens* pathovars, do not have a flagellum and were shown to lack swimming motility (Hersemann et al., 2017). As *Xtg* is spread mostly through mowing tools, flagellar motility might not be crucial to its lifestyle, as the flagellum plays a role in penetration, but not in the multiplication *in planta*. Nonetheless, one component playing a role in motility as well as adhesion to the host surface and biofilm formation is the type IV pilus (T4P, Dunger et al., 2016). The T4P is a transmembrane protein complex related to the type II secretion system (T2SS, see below). The substrate translocated by the T4P is a polymer of the major pilin PilA and minor pilins, forming a filament that can be extended and retracted. The extension, attachment, and retraction of the filament produces a movement called twitching motility. In most *Xanthomonas* genomes, the gene encoding for PilA is found in a cluster with structural genes of the T4P. In *X. oryzae* pv. *oryzicola*, mutations of the *pilQ*, *pilT* and *pilZ* structural genes induced a slight reduction of virulence, while mutation of *pilM* caused a severe loss of virulence (Wang et al., 2007). Additionally, *Xanthomonas* spp. also carry an operon containing genes encoding for the minor pilins PilE, PilX, PilW, PilV, FimT and the anti-retraction factor PilY1. In *Xtg*, these minor pilin genes, along with the gene encoding for the major pilin PilA, have been found to be highly divergent compared to other *X. translucens* pathovars (Hersemann et al., 2017). As extracellular features such as the flagellum and pilins may act as elicitors of plant defenses, the particular structure of the pilins, as well as the absence of a flagellum, could be involved in the avoidance of plant defenses in *Xtg*.

1.3.2. Secretion systems

Most virulence factors from bacteria are secreted into the apoplast or the plant cell via a range of mechanisms, mostly thanks to secretion systems (Büttner and Bonas, 2010; Alvarez-Martinez et al., 2021). These systems are complex molecular machineries embedded in the bacterial membrane that

allow the secretion of molecules. These secretion systems are classified into six main types (I, II, III, IV, V and VI). The type I secretion system is involved in the secretion of various molecules, including adhesins, toxins, and degradative enzymes (Spitz et al., 2019). The type IV and VI secretion systems are involved in competition with other microorganisms through the secretion of antimicrobial compounds into the cells of bacteria as well as eukaryotes such as amoeba (Sgro et al., 2019; Choi et al., 2020). The type V secretion system encompasses different transmembrane transporters involved in the secretion of non-fimbrial adhesins (Fan et al., 2016). As they are arguably the most important in virulence, the type II and type III secretion systems are the most studied in *Xanthomonas* species.

1.3.3. Type III secretion system (T3SS)

Bacteria of the *Xanthomonas* genus are able to directly secrete effector proteins into the host cell. These effectors contribute to the pathogen's survival and virulence by acting on different mechanisms, such as host transcription regulation, host immunity, or nutrient acquisition. In Gram-negative phytopathogenic bacteria like *Xanthomonas* spp., most effectors are secreted into the plant cell via the type III secretion system (T3SS) (Büttner and Bonas, 2010; Galán et al., 2014). These type III effector proteins (T3E) are generally referred to as "*Xanthomonas outer proteins"* (Xop), with 53 families categorized from XopA to XopBA (White et al., 2009, listed at http://www.xanthomonas.org/). Furthermore, some T3E are named after their ability to cause avirulence, such as AvrBs1 to AvrBs4. The T3E repertoire is highly variable between species, pathovars, and even strains. For example, 24 T3E were identified in *X. axonopodis* pv. *citri*, 32 in *X. oryzae* pv. *oryzae* strain KACC10331, and 37 in strains MAFF 311018 and PXO99 (Büttner and Bonas, 2010). However, although some strains can possess a large repertoire of T3E, all of them may not contribute to pathogenicity. For example, in *X. campestris* pv. *vesicatoria*, a genetic screen identified seven novel T3E, but after disrupting these genes only the XopN mutant showed a significant reduction in growth in pepper and tomato (Roden et al., 2004). A comparative analysis in *X. translucens* identified *Xtg*-specific effector proteins including a YopT-like cysteine protease and a XopJ class effector protein, which were tested for their role in virulence of *Xtg* (Hersemann et al., 2017). However, single and double knock-out mutants for these genes showed no significant difference in virulence compared to the wild type strain.

Additionally, a specific class of T3E called transcription activator-like (TAL) effectors were identified in *Xanthomonas* and are characteristic of the genus (Boch and Bonas, 2010). These effectors act as transcription activators and induce the transcription of plant genes that will facilitate the infection of the pathogen. For example, in rice, many genes of the SWEET family, encoding sucrose transporters, were found to be targeted by *X. oryzae* pv. *oryzae* TAL effectors (Streubel et al., 2013a). This causes their activation and the overproduction of transporters which results in the export of sugars to the apoplast that the bacteria use as a source of nutrition. Additionally, many resistance genes against *Xanthomonas* species directly or indirectly interact with TAL effectors to trigger a specific defense response (Zhang

et al., 2022). The best known example is AvrBs3, the first TAL effector to be characterized, which specifically interacts with the *Bs3* gene in pepper, which leads to a hypersensitive response (Bonas et al., 1989; Boch and Bonas, 2010). However, no TAL effector has yet been identified in *Xtg* (Wichmann et al., 2013).

Most of the components of the T3SS machinery are encoded by a large gene cluster called *hrp* (hypersensitive response and pathogenicity), found either on the chromosome or a plasmid. In *Xtg*, a unique noncanonical T3SS was revealed thanks to a first draft genome sequence of the Xtg29 isolate obtained by shotgun sequencing (Wichmann et al., 2013). Later, a similar *hrp* cluster was found in genome sequences of pv. *arrhenatheri*, *cerealis*, *phlei*, *poae*, *translucens* and *undulosa* (Hersemann et al., 2016; Falahi Charkhabi et al., 2017; Pesce et al., 2017; Shah et al., 2021). In this cluster, most components of the T3SS were found, however the *hrpF* and *hpaF* genes that are involved in effector translocation in some *Xanthomonas* species were not found. Moreover, the disposition of the regulating elements *hrpG* and *hrpX* is unique, as they are found within the *hrp* gene cluster, while in other Xanthomonads these genes are found outside of the cluster (Wichmann et al., 2013). *Xtg* mutants for the T3SS structural and regulating genes *hrpG*, *hrpE* and *hrcR* were significantly impaired for their ability to cause disease symptoms, even though their development in the plant was not hindered, showing that the T3SS is required for the induction of symptoms, but not for the survival of *Xtg* (Wichmann et al., 2013). This, together with the fact that some virulence was still observed in mutants, suggests that other virulence factors, such as extracellular polysaccharides or degradative enzymes secreted by the T2SS could play a crucial role in the infection of *L. multiflorum* by *Xtg*.

1.3.4. Type II secretion system (T2SS)

After penetration, the bacteria multiply in the mesophyll, often invading the entire parenchyma, including the protoxylem lacuna and the adjacent xylem parenchyma cells (Masuch et al., 1989). They ultimately reach the xylem vessels, allowing long-distance transport of the bacteria. During the infection process, dissolution of primary cell wall was observed, allowing the bacteria to pass through the resulting openings (Masuch et al., 1989). Such dissolution is mainly the result of cell wall-degrading enzymes that are secreted by the T2SS into the host apoplast (Jha et al., 2005). These enzymes include cellulases, pectate lyases, α-amylases, lipases, and proteases. In *Xanthomonas* spp., two distinct T2SS gene clusters have been identified: the *xps* cluster that is conserved across the genus, and the *xcs* cluster that is specific to some species or pathovars such as *X. campestris* and *X. axonopodis* pv. *citri* (Szczesny et al., 2010). The T2SS and the secreted enzymes play a crucial role in the virulence of *Xanthomonas*. In *X. campestris* pv. *campestris*, protease- and amylase-deficient mutants showed reduced virulence, and in *X. oryzae* pv. *oryzae*, lipase- and xylanase-deficient double mutants showed severe virulence deficiency (Dow et al., 1990; Rajeshwari et al., 2005a; Lin et al., 2021). In *Xtg*, a *xps* T2SS was found, and a comparative genomic analysis identified nine genes containing signal peptides that were exclusive

to *Xtg*, including two glycerophosphodiester phosphodiesterase and an alpha-galactosidase, which may be translocated by the T2SS (Hersemann et al., 2017).

1.4. Control of bacterial wilt

As disease dispersal is promoted by mowing with contaminated equipment, disinfection of mowing tools is advised. Dispersal being more severe under wet conditions, it is recommended to mow in dry conditions (Leyns et al., 1987; Schmidt, 1988a, 1989). However, this limits the propagation of the disease to a small extent only, and disease dispersal still occurs. Using antibiotics could allow for control of *Xtg*, but due to the rise of antibiotic resistant bacteria that could cause problems for animal and human medicine, their use is not recommended. Other biocides such as copper could help control bacterial wilt, however they need to be applied after each mowing, which may not be economically viable (Lamichhane et al., 2018). Moreover, they can be phytotoxic and damaging to the environment, and their use in grasslands is prohibited in many countries.

Biological control was considered as a means to control *Xtg* using *Pseudomonas fluorescens* and *Erwinia herbicola*, which have been found to reduce bacterial wilt when inoculated simultaneously with *Xtg* (Schmidt, 1988c). However, biological control with these epiphytic bacteria and other bacteria found on the phylloplane was limited, working only when the infection was weak, and with concentrations up to 100 times that of *Xtg* needed for sufficient control both in the greenhouse and *in vitro* (Schmidt, 1988c, 1988b). Application of *P. fluorescens* on mowing tools reduced yield losses under high disease pressure in the most susceptible cultivars, however this effect was short-lived (Schmidt, 1988a).

Therefore, breeding for resistant cultivars is currently the only way for efficient and durable control of *Xtg* in *L. multiflorum*. So far, breeding programs have been based on recurrent phenotypic selection of resistant individuals following infection with *Xtg.* Some resistance in *L. multiflorum* was obtained, and cultivars with increased resistance to *Xtg* are available (Suter et al., 2021). However, since *L. multiflorum* is an allogamous species, cultivars are highly heterozygous. Dominant resistance alleles can mask susceptibility alleles, and homozygosity of these alleles may occur even after several cycles of recurrent selection. Using one highly aggressive strain of *Xtg* for breeding has shown a greatly increased resistance in the next generation in meadow fescue (*Festuca pratensis*), however improvement became stagnant afterwards (Michel, 2001). In recurrent selection of *F. pratensis*, significant improvement was obtained in the first selection cycle, but progress decreased and stagnated after four to six cycles, without complete resistance being achieved (Boller et al., 2001). In order to breed resistant cultivars in an efficient manner and fix resistance alleles in *L. multiflorum* cultivars, genomics-assisted breeding methods such as MAS would greatly help. However, these techniques require a comprehensive understanding of disease resistance and the genetic factors that govern it.

1.4.1. Plant immunity

Plant immunity has been described in the zig-zag model as made of two types of defenses (Fig. 2; Jones and Dangl, 2006; Ngou et al., 2022). The first type is a general broad spectrum defense made of transmembrane pattern recognition receptors (PRRs) that recognize common pathogen-associated molecular patterns (PAMPs) such as flagellin, elongation factor Tu or fungal chitin (Felix et al., 1999; Kunze et al., 2004; Wan et al., 2008). After recognition of PAMPs, the PRRs induce a signaling cascade that leads to the activation of mitogen-activated protein (MAP) kinases, fluxes of Ca^{2+} and other ions, generation of reactive oxygen species (ROS), deposition of callose on the cell wall, and activation of immune-related genes (Boller and Felix, 2009). This is referred to as pathogen-triggered immunity (PTI), which is a basal immune response, considered slow, short and weak. Pathogens have evolved to produce effector proteins, generally secreted into the plant cell or the apoplast, that overcome this resistance by manipulating the plant's structure or functions, inducing effector-triggered susceptibility (ETS) (Abramovitch et al., 2006; Grant et al., 2006; Kay and Bonas, 2009). These effectors can for example suppress PTI by acting directly on PRRs or on other defense mechanisms like MAP kinases, alter the host cell's metabolism, facilitate the nutrition or the dispersal of the pathogen, etc (Macho and Zipfel, 2015). In turn, plants also evolved receptors under the control of major R genes that mostly encode nucleotide-binding leucine rich repeat (NLR) proteins that recognize effectors and trigger the second type of resistance, effector-triggered immunity (ETI). In this particular case where the effector does not serve its virulence purpose anymore and rather triggers plant immunity, these effectors are called avirulence factors. This specific recognition of avirulence factors usually leads to the programmed death of the infected cells, called hypersensitive response (HR) and to resistance generalized to the whole plant, systematic acquired resistance (SAR), which prevents the pathogen from spreading and protects the plant from future infections. There is a constant arms race between pathogens and plants, the former developing ways to suppress or evade ETI and other defense mechanisms, the latter developing new ways to recognize the pathogen and prevent infection (Jones and Dangl, 2006; Anderson et al., 2010).

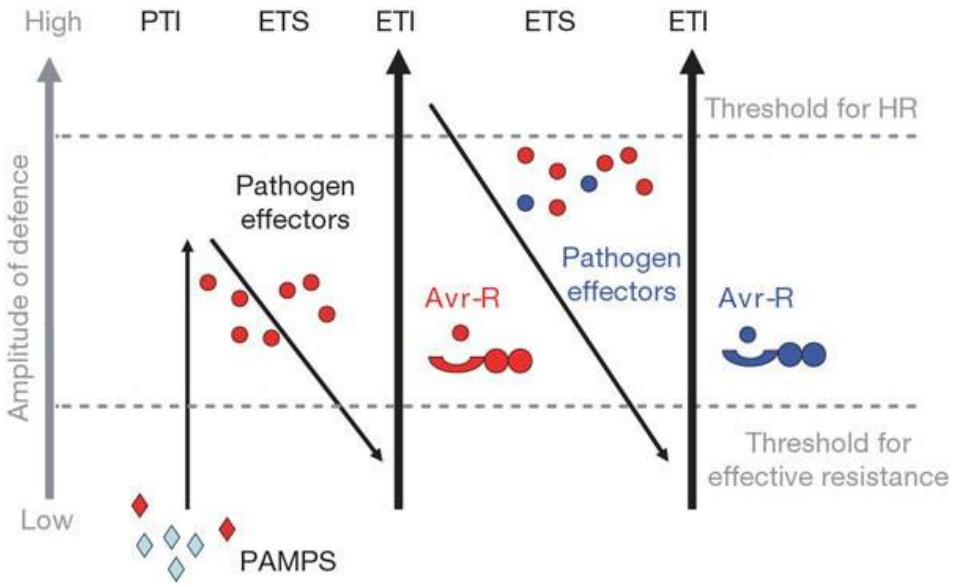


Figure 2 The zig zag model proposed by (Jones and Dangl, 2006). Pathogen-associated molecular patterns (PAMPs) are recognized by pattern recognition receptors (PRRs), leading to pattern-triggered immunity (PTI). Pathogen effectors act on PTI and other processes to induce effector-triggered susceptibility (ETS). These effectors may in turn become avirulent factors by being recognized by R proteins from the plant host, leading to effector-trigger immunity (ETI), often characterized by the hypersensitive response (HR). An evolutionary arms race occurs between the pathogen and the host, the former evolving new effectors to trigger ETS, and the latter evolving new receptors to trigger ETI and HR.

This zig zag model has however been criticized, as the distinctions between PAMPs and effectors; between PRRs and R proteins; and between PTI and ETI are not entirely clear (Thomma et al., 2011). While PAMPs are widely conserved across genera, effectors are specific to a genus or species. However, many groups of effectors, such as the LysM effectors that suppress chitin-triggered immunity, are widespread in fungal pathogens, and could thus be considered PAMPs (de Jonge and Thomma, 2009). Effectors were initially equated to proteins secreted into the host cell by the type III secretion system (T3SS), however, the definition rapidly became broader to include all secreted molecules that affect the host cell structure or function (Hogenhout et al., 2009; Win et al., 2012). Therefore, many PAMPs that are secreted in the apoplast or the plant cell and affect its structure or function may also be considered effectors. Regarding plant resistance, PRRs are typically cell surface receptor kinases that recognize PAMPs in the apoplast, while R genes encode mainly for NLR proteins that recognize effectors in the cytoplasm. However, in tomato, the cell surface receptor Ve1 was shown to lead to resistance following recognition of the Ave1 effector, which is conserved in many species of *Verticillium* (Song et al., 2018). Therefore, Ave1 may be considered both a PAMP and an effector, and Ve1 both a PRR and a R protein. Finally, while PTI is considered a weak response, ETI is considered a strong response, often leading to HR and SAR. However, both pathways share many signaling components, and many PAMPs, such as the *Erwinia amylovora* harpin, the *Phytophthora parasitica* CBEL, and the flg22 flagellin fragment have been shown to trigger HR (Wei et al., 1992; Khatib et al., 2004; Naito et al., 2007). Furthermore, PAMP recognition has also been shown to lead to SAR in *A. thaliana* (Mishina and Zeier, 2007). This shows that the zig-zag model is likely an oversimplification of the complex interactions between hosts and

pathogens. Nonetheless, while this model should be refined, it has provided a strong foundation to better understand plant-pathogen interactions and remains a valuable conceptual tool.

1.4.2. Identification of resistance genes and QTL

To identify genes or quantitative trait loci (QTL) involved in disease resistance, researchers generally rely on QTL mapping based on specific mapping populations (Glazier et al., 2002; Miedaner et al., 2020). In most cases, a biparental population is constructed by crossing a resistant parent to an unrelated susceptible parent. The resulting F_1 or the subsequent F_2 will be segregating for resistance, allowing to distinguish resistant individuals from susceptible individuals. Genetic markers that are highly associated with resistance can then be identified, allowing the positional cloning of the gene responsible for resistance. In some cases, F_3 or further generations can be used, the F_2 individuals can be self-fertilized to produce recombinant inbred lines, or double-haploid lines can be produced, which facilitate the fine-mapping of QTL (Takuno et al., 2012; Filiault et al., 2017). The production of recombinant inbred lines is however not possible in *Lolium* spp., as they are self-incompatible. Until the development of DNA sequencing, such studies relied on various markers, such as RFLPs, AFLPs or SSRs. This resulted in low marker densities, which hindered the precise identification of a small locus linked to resistance. More recently, QTL mapping studies rely on sequencing-based markers, mainly SNPs, which allow for a much higher marker resolution. Still, QTL mapping studies are often unable to narrow down the QTL to a small chromosomal region, as other factors influence these studies, such as the heritability of the trait or the population size. In this regard, bulked segregant analysis (BSA) allows to screen large populations by sequencing pools of DNA extracted from a large number of individuals in order to fine-map QTL more efficiently (Michelmore et al., 1991).

1.4.3. Resistance to *Xanthomonas* spp.

To date, more than a hundred of genes, candidate genes, or QTL for resistance to *Xanthomonas* spp. have been identified in various species such as banana (*Musa* spp.), bean (*Phaseolus vulgaris* L.), cassava (*Manihot esculenta*), citrus (*Citrus* spp.), cotton (*Gossypium* spp.), mango (*Mangifera indica* L.), pepper (*Capsicum annuum*), populus (*Populus* spp.), prunus (*Prunus* spp.), rice (*Oryza* spp.), soybean (*Glycine max* L. Merr.), strawberry (*Fragaria* spp.), tomato (*Solanum lycopersicum*), walnut (*Juglans regia* L.) and wheat (listed at http://www.xanthomonas.org/). However, a large portion of these QTL have not been fine-mapped, and only a small number of genes have been identified and characterized so far. In wheat, the use of F_3 populations derived from reciprocal crosses between two susceptible and three resistant parents allowed to identify five loci for resistance to bacterial leaf streak caused by *X. translucens* pv. *undulosa*, named *Bls1* to *Bls5* (Duveiller et al., 1992). However, this study did not allow to genetically map these loci. A further study conducting association mapping in 566 wheat accessions allowed to define five QTL for resistance to bacterial leaf streak, corresponding to the five

previously-identified loci, facilitating breeding for resistant wheat cultivars (Adhikari et al., 2012). However, the underlying genes have not been characterized to date, and the resistance conferred by these loci is only partial (Adhikari et al., 2012; Kandel et al., 2012). In the pepper - *X. euvesicatoria* pathosystem, many resistance genes linked to avirulence genes from the pathogen have been identified. In particular, the *Bs1* to *Bs4* genes, which interact respectively with the *AvrBs1* to *AvrBs4* avirulence factors from the pathogen and trigger HR (Stall et al., 2009). While *Bs1* has not been cloned so far, *Bs2*, *Bs3* and *Bs4* have been cloned and characterized. The *Bs2* and *Bs4* genes encode NLR proteins that bind to *AvrBs2* and *AvrBs4*, respectively, while *Bs3* encodes a flavin monooxygenase and acts as an executor R gene. Executor R genes are genes that have evolved to possess promoters with regions similar to the binding regions of TAL effectors, resulting in their activation by these TAL effectors, which will lead to HR (Ji et al., 2022). No clear similarity between the known executor genes have been identified, and their mode of action is largely unknown, but likely involves a signal transduction leading to HR, with different pathways depending on the gene. Based on these identified genes, markers have been developed and are used in MAS to select them more efficiently (Truong et al., 2011; Holdsworth and Mazourek, 2015; Barka and Lee, 2020). Finally, the most studied pathosystem in *Xanthomonas* pathology is arguably the rice - *X. oryzae* pv. *oryzae* pathosystem, in which as many as 47 resistance genes were identified, named *Xa1* to *Xa47*. Though most of these genes have not been cloned nor characterized, many similar genes could be characterized (Joshi et al., 2020). First, the *Xa3*/*Xa6*/*Xa26*, *Xa21* and *Xa4* genes encode for transmembrane receptor-like kinases (RLK), which respectively recognize *AvrXa3*, *RaxX*, and an undiscovered *Avr* gene. The *Xa13*, *Xa25* and *Xa41* genes encode SWEET proteins that export sugars to the apoplast (respectively SWEET11, SWEET13 and SWEET14). These genes act more as susceptibility genes, as they are the targets of many effectors such as PthXo1 to PthXo3, TalC or TalF. However, naturally occurring variants of these genes prevent the action of effectors, allowing them to act as resistance genes. The *Xa10*, *Xa23* and *Xa27* genes encode for executor R genes, which are triggered by their respective TAL effectors AvrXa10, AvrXa23 and AvrXa27. Finally, *Xa1* and its allelic genes *Xa2*/*Xa14*/*Xa31*/*Xa45*, and *Xa47* are *NLR* genes that bind to various TAL effectors. However, the resistance conferred by *Xa1* and its allelic genes can be suppressed by interfering TAL effectors (Ji et al., 2020). This shows the importance of QTL mapping to identify resistance genes, and the same concepts have been previously applied to identify sources of resistance to *Xtg* in *L. multiflorum*.

1.4.4. Genomics-based *Xtg* resistance gene discovery in *L. multiflorum*

The first basis for understanding the genetic architecture of *L. multiflorum* consisted in linkage maps that have been constructed in *L. multiflorum* using RFLP, RAPD (Hayward et al., 1998), AFLP (Studer et al., 2006), SSR (Hirata et al., 2006), EST-derived CAPS (Miura et al., 2007), and DArT (Bartoš et al., 2011) markers. These linkage maps were then combined, along with *L. perenne* maps, into one

consensus linkage map for *Lolium* spp. using EST-derived SSR markers as anchor markers (Studer et al., 2010). These maps served as a first basis for identifying genomic regions associated with a trait of interest and for the development of markers to be used in MAS.

Later, a first draft genome sequence was produced for *L. perenne*, providing further insights into the genome structure and gene content of *Lolium* species (Byrne et al., 2015). The first draft genome assembly for *L. multiflorum* was made available in 2018, allowing for more in-depth analyses in the species, and a first comparison with *L. perenne* and other Poaceae (Knorst et al., 2018b). Furthermore, this genome assembly was produced by sequencing genomic DNA from M2289, an individual from advanced breeding germplasm that is resistant to *Xtg*, providing a first basis to determine the genetic loci responsible for *Xtg* resistance.

To identify these genetic loci, a first study used a segregating F_1 population (hereafter referred to as Xtg-ART) resulting from a cross between the M2289 resistant individual, and an individual from the highly susceptible cultivar Adret (Studer et al., 2006). Using AFLP and SSR markers, a major QTL explaining 43 to 84% of phenotypic variation for resistance was identified on linkage group (LG) 4, together with a few minor QTL with limited positioning precision on LG1, 4, 5 and 6, explaining 2.9 to 10.8% of phenotypic variation for resistance. However, further characterization of the QTL and the development of markers for MAS was hindered by the predominantly anonymous markers and the lack of sequence information for *L. multiflorum* at the time.

Later, analyses of transcriptional response to *Xtg* infection using cDNA-AFLP (Rechsteiner et al., 2006) and a cDNA microarray (Wichmann et al., 2011a) showed that many genes were differentially expressed in *L. multiflorum* after infection by *Xtg*, and that this differential expression was very different between a susceptible and a resistant genotype. These studies allowed to identify genes potentially involved in the interaction between *L. multiflorum* and *Xtg*.

The cDNA-AFLP approach was able to identify many differentially expressed transcript-derived fragments (TDF) in a resistant genotype. TDFs showing similarity to ethylene-responsive element-binding proteins were found to be upregulated in infected plants. Ethylene acts as a hormone in plants and is involved in many mechanisms of defense against biotic and abiotic stress (Broekaert et al., 2006; Lin et al., 2009). Thus, the upregulation of these TDFs in *L. multiflorum* suggests they may be involved in the interaction with *Xtg*. Another upregulated TDF showed homology to a gene expressed in *Vitis vinifera* after infection by *Xylella fastidiosa*, a member of the Xanthomonadacea family. TDFs showing some, but not significant homology were also found to be upregulated. One showed homology to an ATP/GTP-binding site motif A, which is characteristic for the NLR region observed in most resistance genes (Dangl and Jones, 2001). Another showed homology to a terpene synthase. Terpenes are secondary metabolites that have a role in defense mechanisms against pathogens (Wittstock and

Gershenzon, 2002). However, this cDNA-AFLP approach was limited by the small number and the short length of the TDFs and only one resistant genotype was used.

The second approach, using a cDNA microarray, allowed for a more comprehensive transcriptome analysis with a larger number of transcripts using both a resistant and a susceptible genotype. Lsi1, belonging to the nodulin26-like major intrinsic protein family of aquaporins, which was found to be up-regulated in citrus after infection by *Xanthomonas axonopodis* pv. *axonopodis* (Cernadas et al., 2008) was upregulated in *Lolium multiflorum* after infection by *Xtg*. Lsi1 is known to be involved in silicon uptake in rice and barley, which might have an important role in defense. Plants supplied with silicon produce defense compounds in response to fungal infection, and have enhanced activity of defense mechanisms (Ma and Yamaji, 2006). Treating wheat with silicon showed a 50.2% reduction of the chlorotic area on leaves infected by *X. translucens* pv. *undulosa* compared to non-treated plants (Silva et al., 2010). Another upregulated gene showed sequence similarity to proteins of the germin-like family, which are involved in broad spectrum basal defence (Manosalva et al., 2009). This gene also showed sequence similarity to an oxalate oxidase in wheat. Germin-like proteins exhibit oxalate oxidase activity or superoxide dismutase activity, both resulting in the production of H_2O_2, a ROS involved in stress response (Bernier and Berna, 2001). Oxalate oxidases have already been associated with defense against *Xanthomonas* spp., as a gene encoding an oxalate oxidase in rice was associated with a putative QTL on chromosome 3 accounting for 10.3% of resistance against *X. oryzae* pv. *oryzae* (Ramalingam et al., 2003). Furthermore, chromosome 3 displays synteny with LG4 of *Lolium multiflorum*, where the major QTL for resistance against *Xtg* was identified (Devos, 2005). In *Lolium perenne*, an oxalate oxidase (*Lp*OXO) was also found in a similar location as the major QTL on LG4 (Dracatos et al., 2009). Synaptobrevin-like proteins, vesicle-associated membrane proteins involved in vesicle trafficking were also upregulated in the resistant genotype. One of them mapped to LG4 and was also upregulated in tomato and citrus infected with *Xanthomonas* spp. (Cernadas et al., 2008). The generation of ROS during pathogen response is damaging to the cell membrane and vesicle trafficking is important in membrane repair and in suppressing ROS-induced cell death, but also for cell wall appositions, called papillae, which prevent pathogens from penetrating the cell (An et al., 2006). Flowering gene MFS18 was also upregulated after *Xtg* inoculation in the cDNA microarray study. Flowering time has been shown to be strongly correlated with disease resistance, with many defense pathways altering the transition to flowering, and it has been hypothesized that flowering genes may have a dual role in both flowering and defense (Lyons et al., 2015). An ankyrin repeat protein was also upregulated after infection by *Xtg*. Ankyrin repeat proteins, such as NPR1 in *Arabidopsis thaliana* are known to be involved in growth, development, and in SAR as well as in ETI after activation by the defense hormone salicylic acid (Cao et al., 1997; Pajerowska-Mukhtar et al., 2013). In pepper, an ankyrin repeat-containing zinc finger protein was upregulated in response to inoculation of *X. axonopodis* pv. *glycines*, as well as cold and salt stresses (Seong et al., 2007). Interestingly, this differential expression in *Lolium multiflorum*

following infection by *Xtg* is very similar to differential expression in PTI, triggered by the recognition of PAMPs (Wichmann et al., 2011a).

Recently, a BSA study using pooled DNA sequencing of a segregating mapping population allowed to identify scaffolds mapping to LG4 containing potential candidate genes for resistance to *Xtg* (Knorst et al., 2018a). Among them, one gene encoding a serine/threonine-protein kinase-like protein CCR4 (CRINKLY4) related protein was found. Phosphorylation cascades, such as the MAP kinase (MAPK) cascade, with Ser/Thr kinases being at the top of the cascade, are crucial in signal transduction to activate plant defenses (Meng and Zhang, 2013). In tomato, the *Pto* gene encodes a Ser/Thr kinase conferring resistance to *Pseudomonas syringae* strains carrying the *avPto* gene which acts as a suppressor of the MAPK cascade (Martin et al., 1993; He et al., 2006). In *L. multiflorum*, CRINKLY4 could be involved in detection of *Xtg* and the initiation of plant defense. A gene encoding for a nodulin MtN3 family protein, also known as SWEET17 (Sugars Will Eventually be Exported Transporters), was also identified. SWEET17 is a sugar vacuolar transporter (Chardon et al., 2013), and could be a target of *Xtg* to take advantage of the plant's sugar for its growth, as is the case for other SWEET transporters targeted by many pathogens (Chen, 2014). In rice, *OsSWEET11*, also known as *Xa13* or *Os8N3* is a target of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) and is up-regulated by the TAL effector *PthXo1* (Chu et al., 2006; Yang et al., 2006), and mutations in its promoter leads to resistance against *Xoo* (Chu, 2006). Intrestingly, OsSWEET12, OsSWEET13 (Xa25), OsSWEET14, and OsSWEET15, four other SWEET proteins closely related to OsSWEET11 have also been found to be potential targets of *Xoo* (Streubel et al., 2013a). However, in rice, this is achieved by the use of TAL effectors by *Xoo*, which are absent in *Xtg* (Wichmann et al., 2013). A BAK1-interacting receptor-like kinase 1 was also found. BAK1 is a leucine-rich repeat (LRR) receptor-like kinase that is involved in signaling by flagellin-sensitive 2 (FLS2 ; Gómez-Gómez and Boller, 2000) and elongation factor Tu receptor (EFR ; Zipfel et al., 2006), two PRRs sensing flagellin and elongation factor Tu, respectively (Chinchilla et al., 2007). BAK1 forms a heterodimer with LRR receptor kinases like FLS2 after ligand binding, and becomes phosphorylated, which is a key initial step in signal transduction (Schulze et al., 2010). However, *Xtg* lacks a flagellum (Hersemann et al., 2017), thus it does not seem plausible that BAK1 would be involved in the interaction between *Xtg* and *L. multiflorum* through FLS2. Other candidate genes included a cysteine-rich receptor-like kinase (RLK) 26, known to be involved in pathogen response (Chen et al., 2004), and a toll interleukin 1 (TIR)-NLR disease resistance protein with no reported function (Meyers et al., 2003). However, only 44 of the most resistant and the most susceptible F1 individuals were used in this analysis, limiting the potential for narrowing down the QTL. Another limit of this study was the lack of sequence annotation for most of the candidate scaffolds, due primarily to the small length of the scaffolds and the fragmented reference genome assembly of *L. multiflorum*, obtained by shotgun sequencing of the parents of the Xtg-ART population (Knorst et al., 2018b).

The presence of a major QTL generally suggests the existence of one or a few major resistance genes, usually leading to a race-specific resistance based on gene-for-gene interactions between bacterial effector genes and plant resistance genes. To test this hypothesis, 62 *L. multiflorum* genotypes were inoculated with six different *Xtg* strains (Wichmann et al., 2011b). Significant differences were observed among the *Xtg* isolates, however no significant *L. multiflorum* genotype by *Xtg* genotype interaction was observed, suggesting the absence of race-specific resistance in the genotypes investigated, and that bacterial wilt resistance is effective across a broad range of bacterial isolates and plant genotypes.(Wichmann et al., 2011b). Together, the lack of race-specificity in defense against *Xtg*, and the quantitative nature of the observed segregation of resistance in the Xtg-ART population suggest that the resistance to *Xtg* is likely a quantitative broad-spectrum resistance similar to PTI, and may be influenced by multiple genes, rather than a specific resistance under the control of one major gene.

1.5. Thesis outline

Although first insights into *Xtg* resistance have been uncovered, with a region on LG4 found to contain a major QTL, this region remains large, preventing the precise identification of candidate genes. Furthermore, no marker tightly linked to resistance could be clearly defined for the development of markers to be used in MAS. To be the most efficient, MAS requires the identification of markers in close linkage to resistance genes. To this end, the first aim of this work was to narrow down the QTL region using BSA coupled with high throughput whole-genome sequencing to define a small region linked to *Xtg* resistance and uncover underlying candidate genes. Additionally, identifying virulence factors from the pathogen is essential in understanding how the disease occurs as well as in developing resistant cultivars. While previous comparative analyses provided crucial knowledge about potential genes involved in virulence of *Xtg*, these studies were based on incomplete draft genomes, preventing more comprehensive analyses. Therefore, the second aim was to produce high-quality complete genome sequences for all pathovars of *X. translucens* and multiple strains of *Xtg* in order to perform comparative analyses and identify a set of genes that could be crucial to the pathogenicity of *Xtg*. Overall, the aim of this work was to provide essential knowledge and genomic resources to better characterize the *L. multiflorum* – *Xtg* interaction and identify sources of resistance to breed resistant cultivars more efficiently.

Chapter 2 reports the production of high-quality complete genome sequences of representative strains for each pathovar ox the *X. translucens* species, allowing for a comprehensive comparative analysis within the species. **Chapter 3** describes an in-depth comparative analysis focused on *Xtg* using additional complete genome sequences for four strains of *Xtg*, which allowed for the identification of a remarkable genome plasticity within the pathovar, as well as conserved features exclusive to *Xtg* that may be crucial in pathogenicity. Then, **Chapter 4** presents the use of a high-resolution bulked segregant analysis in a *L. multiflorum* F_2 population segregating for resistance, which allowed to refine the

previously identified QTL and define a ~300 kbp region that was highly associated with resistance to *Xtg*. In this region, multiple genes known to be involved in resistance to pathogens can be found and constitute strong candidate genes for resistance to *Xtg*. Finally, **Chapter 5** concludes this thesis with a general discussion of the findings described in the previous chapters, focusing on future perspectives to validate the identified candidate genes, define markers for MAS, and identify new sources of resistance to *Xtg*.

2. Complete genome assemblies of all *Xanthomonas translucens* pathotype strains reveal three genetically distinct clades

Florian Goettelmann[1], Veronica Roman-Reyna[2], Sébastien Cunnac[4], Jonathan M. Jacobs[2], Claude Bragard[3], Bruno Studer[1], Ralf Koebnik[4], Roland Kölliker[1*]

[1]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

[2]Department of Plant Pathology, The Ohio State University, Columbus, OH, USA

[3]Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium

[4]Plant Health Institute of Montpellier, University of Montpellier, Cirad, INRAE, Institut Agro, IRD, Montpellier, France

Research article: Published in Frontiers in Microbiology, 2022.

Available at: https://doi.org/10.3389/fmicb.2021.817815

Keywords: *Xanthomonas translucens*, complete genomes, phylogeny, host adaptation, comparative genomics, virulence factors.

2.1. Abstract

The *Xanthomonas translucens* species comprises phytopathogenic bacteria that can cause serious damage to cereals and to forage grasses. So far, genomic resources for *X. translucens* were limited, which hindered further understanding of the host-pathogen interactions at the molecular level and the development of disease-resistant cultivars. To this end, we complemented the available complete genome sequence of the *X. translucens* pv. *translucens* pathotype strain DSM 18974 by sequencing the genomes of all the other ten *X. translucens* pathotype strains using PacBio long-read technology and assembled complete genome sequences. Phylogeny based on average nucleotide identity revealed three distinct clades within the species, which we propose to classify as clades Xt-I, Xt-II and Xt-III. In addition to 2,181 core *X. translucens* genes, a total of 190, 588 and 168 genes were found to be exclusive to each clade, respectively. Moreover, 29 non-TAL and 21 TAL type III effector classes were found, and clade- or strain-specific effectors were identified. Further investigation of these genes could help to identify genes that are critically involved in pathogenicity and/or host adaptation, setting the grounds for the development of new resistant cultivars.

2.2. Introduction

Xanthomonas translucens is a species of Gram-negative phytopathogenic bacteria which causes serious damage in gramineous plants. To date, a total of eleven pathovars have been defined based on their host range, each represented by one strain referred to as the pathotype strain and deposited in the appropriate

bacterial strain collections. Two major groups are historically distinguished within the species: the "translucens" group and the "graminis" group. The translucens group consists of pathovars *cerealis*, *hordei*, *secalis*, *translucens* and *undulosa*, which cause leaf streak and black chaff in economically important cereals such as wheat (*Triticum* spp.) and barley (*Hordeum vulgare*) (Sapkota et al., 2020). The graminis group consists of pathovars *arrhenatheri*, *graminis*, *phlei*, *phleipratensis* and *poae*, which cause bacterial wilt in forage grasses (Egli et al., 1975; Egli and Schmidt, 1982; Stead, 1989). Although recent genome sequence data loosely supported this phenotypic classification (Peng et al., 2016; Hersemann et al., 2017; Langlois et al., 2017; Shah et al., 2021), these groups are not clearly defined based on the genetic relationships between the strains. Moreover, genomic studies have previously suggested that pv. *cerealis* could be genetically distinct from the other "translucens" group pathovars (Bragard et al., 1997; Rademaker et al., 2006; Peng et al., 2016; Langlois et al., 2017). Furthermore, *Xanthomonas* strains causing dieback in pistachio (*Pistacia vera* L.) were found to belong to the *X. translucens* species, a species so far thought to infect only monocotyledonous plants (Giblot-Ducray et al., 2009). These strains were classified as pv. *pistaciae*, in which two groups were distinguished by rep-PCR and *gyrB*-based phylogeny, and are referred to as groups A and B (Facelli et al., 2005; Marefat et al., 2006; Giblot-Ducray et al., 2009). The group A strain was found to be closely related to pathovars *translucens*, *secalis* and *undulosa* while the group B strain was closest to pv. *cerealis*, suggesting that the ability to infect pistachio may have evolved two separate times in *X. translucens*.

X. translucens pathovars were initially classified solely based on their host range as pathovars of the *X. campestris* species. However, a classification based on DNA-DNA hybridization was proposed by Vauterin et al. (1995), who amended these pathovars of *X. campestris* to the species level, forming the *X. translucens* species. Nonetheless, though these pathovars are now considered pathovars of the *X. translucens* species, the definition of these pathovars is still based on their host range as before. Phylogeny is crucial to better understand the evolutionary history of each pathovar, in order to determine how each strain has adapted to infect its hosts. This knowledge is key in defining targets for breeding resistant cultivars. The current definition of pathovars of *X. translucens* might not properly reflect these processes of host adaptation, and a better definition based on the genetic relationships between the strains would improve the comprehension of the host-pathogen relationships. Recent advances in sequencing technology now allow for the rapid sequencing of complete genomes of bacterial strains, which provide a better basis for phylogenetic analyses. At the time of this study, complete genome sequences were publicly available for pv. *cerealis* strain NXtc01 (Shah et al., 2019), pv. *translucens* strains DSM 18974 (Jaenicke et al., 2016), XtKm7, XtKm8, XtKm9, XtKm33, XtKm34 (Shah et al., 2021) and pv. *undulosa* strains XT4699 (Peng et al., 2016), ICMP 11055 (Falahi Charkhabi et al., 2017), LW16,P3 (Peng et al., 2019), XtFa1, XtLr8, XtKm12 and XtKm15 (Shah et al., 2021). Though the number of available complete genome sequences is continuously growing, these sequences represent

only three pathovars to date, with a complete genome sequence being available for only one pathotype strain.

High-quality genome sequences also help in the identification of genes that are directly linked to the pathogen's virulence and host range. Indeed, successful infection by pathogenic bacteria is mediated by a set of virulence factors, including degradative enzymes and effector proteins. These are often host-specific, and the repertoire of virulence factors that a bacterial strain possesses defines which hosts it is able to infect. In *Xanthomonas* species, virulence factors generally depend on the type II and type III secretion systems (T2SS and T3SS, respectively). These secretion systems allow the bacteria to export virulence factors in order to enable or facilitate their proliferation and survival in the host by targeting specific host components (Büttner and Bonas, 2010; Alvarez-Martinez et al., 2021). Identifying which virulence factors are involved in the pathogen's virulence and which host components are the targets of these virulence factors is crucial, as these host components can then be the focus of resistance breeding.

The T3SS is encoded by the "hypersensitive reaction and pathogenicity" (*hrp*) gene cluster (Bonas et al., 1991). Some *hrp* genes that are highly conserved are referred to as "*hrp* conserved" (*hrc*) genes (Bogdanove et al., 1996). Moreover, some genes in the cluster that are involved in, but not necessary to the host-plant interaction are called "*hrp*-associated" (*hpa*) genes. In the *X. translucens* species, the core *hrp* cluster consists of 23 genes, with eight *hrp* genes, eleven *hrc* genes and four *hpa* genes (Wichmann et al., 2013; Pesce et al., 2017).

This secretion system injects effector proteins into the host cell. In *Xanthomonas*, these effectors are generally called "*Xanthomonas* outer proteins" (Xop), with 53 classes from XopA to XopBA (White et al., 2009). Other effectors are named according to their avirulence characteristics, causing hypersensitive response in the host, such as AvrBs1 to AvrBs3. These effectors are key virulence factors, as, when translocated into the host cell, they are able to target different pathways of the host, allowing the pathogen for example to acquire nutrients or to evade or suppress host defenses. A specific type of effectors secreted by the T3SS are the "transcription activator-like" (TAL) effectors. Their amino acid sequences contain highly conserved repetitive sequences of ~34 amino acids, with only the 12th and 13th residues being hypervariable and referred to as the "repeat variable di-residue" (RVD). The RVD array of each TAL effector allows it to bind to specific nucleotide sequences in the host DNA, thus activating the expression of the neighboring genes to the pathogen's advantage (Streubel et al., 2017; Wang et al., 2017b).

Due to their repetitive sequences, TAL effector genes are difficult to assemble using short-read sequencing technology. Complete genome sequences based on either long-read sequencing, very high coverage of short-read sequencing, or a mixture of both, are thus necessary to properly identify these effectors. Such high-quality genome sequences allowed the identification of eight and five TAL effector genes in pv. *translucens* strains DSM 18974 and UPB886, respectively (Jaenicke et al., 2016; Roman-

Reyna et al., 2020), eight and seven in pv. *undulosa* strains XT4699 and ICMP 11055, respectively (Peng et al., 2016; Falahi Charkhabi et al., 2017), and two in pv. *cerealis* strains CFBP 2541 and NXtc01 (Pesce et al., 2015; Shah et al., 2019). However, of these, only four were functionally characterized and found to play a role in virulence to date (Falahi Charkhabi et al., 2017; Peng et al., 2019; Shah et al., 2019). No TAL effector has yet been identified in other pathovars of *X. translucens*.

Similar to the T3SS, the T2SS is responsible for the export of virulence factors, most of which are cell wall-degradative enzymes, into the host apoplasm (Jha et al., 2005). In *Xanthomonas*, two types of clusters encoding the T2SS can be found. The *xps* cluster is conserved across *Xanthomonas* species, while the *xcs* cluster is only found in some species such as *X. citri* and *X. campestris* (Szczesny et al., 2010).

In addition to the T2SS and T3SS, the type IV and type VI secretion systems (T4SS and T6SS, respectively) also secrete proteins that may affect the virulence of the pathogen. However, while the T2SS and the T3SS target host components, the T4SS and the T6SS are involved in defense against microbial predators such as amoeba, as well as in competition with other microorganisms (Büttner and Bonas, 2010; Alvarez-Martinez et al., 2021). The T4SS is evolutionarily related to bacterial conjugation systems, and is involved in competition with other bacteria by injecting protein effectors or protein-DNA complexes into their cells (Sgro et al., 2019). The T6SS is related to the tail of bacteriophages, and similarly to the T4SS, is able to inject effector proteins into prokaryotic as well as eukaryotic cells (Bayer-Santos et al., 2019). In the Xanthomonadales order, three subtypes of T6SS have been found: subtypes 1, 3 and 4. Moreover, subtype 3 is further subdivided into subgroups 3*, 3** and 3***. Although not directly related to virulence, the T4SS and T6SS could be key elements in the proliferation and survival of the bacteria on the host plant (Souza et al., 2015; Choi et al., 2020).

In this study, in order to complement the available complete genome sequence of the pv. *translucens* pathotype strain DSM 18974 (Jaenicke et al., 2016), we sequenced whole genomes of all other ten pathotype strains of the *X. translucens* species, as well as a representative strain of the *X. translucens* pv. *pistaciae* group B, to produce high-quality genome sequences. Using these, we built a phylogeny of these strains to clarify the taxonomy of the *X. translucens* species. We then scrutinized the genomes for major virulence features of these strains to identify genes that might be important for pathogenicity and in defining their host range.

2.3. Materials and methods

2.3.1. Bacterial strains, growth conditions and DNA extraction

Relevant data for all *Xanthomonas translucens* strains used in this study are listed in Table 1. The genome sequence of *X. translucens* pv. *translucens* strain DSM 18974 was retrieved from the NCBI

GenBank database (accession number LT604072). Strains LMG 726, LMG 727, LMG 728, LMG 730, LMG 843 and UPB458 were grown at 28°C on YDC agar medium (2 % dextrose, 1 % yeast extract, 2 % CaCO3, 1.5 % agar) for 48 h. Bacteria were then dissolved in 10 mL washing buffer (50 mM TRIS-HCl pH 8.0, 50 mM EDTA pH 8.0, 150 mM NaCl). The genomic DNA was then extracted with the NucleoSpin® Microbial DNA kit (Macherey Nagel, Duren, Germany), according to the manufacturer's recommendations. Strains CFBP 2055, CFBP 2539, CFBP 2541 and CFBP 8304 were grown at 28°C on PSA medium (0.5 % peptone, 2 % sucrose, 1.5 % agar) for 24h. Bacteria were then resuspended in 10 mM $MgCl_2$ and diluted to an optical density at 600 nm of 1.0. Cells from 2 ml were harvested by centrifugation, washed once with 10 mM $MgCl_2$, and genomic DNA was isolated using QIAGEN Genomic tip 100/G (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Genomic DNA from strain ICMP 16317 was extracted following a standard phenol/chloroform method (Booher et al., 2015).

2.3.2. Sequencing, genome assembly and annotation

Library preparation and DNA sequencing of strains LMG 726, LMG 727, LMG 728, LMG 730, LMG 843 and UPB458 was done at the Functional Genomics Center Zurich. For these strains, as well as for ICMP 16317, libraries were prepared and multiplexed with the PacBio SMRTbell® Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) according to the published protocol (https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-Multiplexed-Microbial-Libraries-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf). Multiplex library preparation in pools of eight strains, including strains CFBP 2055, CFBP 2539, CFBP 2541 and CFBP 8304, and simultaneous sequencing of eight strains on one SMRTCell was conducted at the GENTYANE genotyping platform (INRA Clermont-Ferrand, France). All strains were sequenced with the PacBio Sequel technology.

Genomic sequences were then *de novo* assembled with flye 2.7 for strains CFBP 2055, CFBP 2539, CFBP 2541 and CFBP 8304, flye 2.8.1 for LMG 726, LMG 728, LMG 730 and UPB458, flye 2.9 for LMG 727 and LMG 843, and HGAP 4 for ICMP 16317 (Chin et al., 2013; Kolmogorov et al., 2019). Assemblies produced with flye were done using the "--plasmids --iterations 2" parameters. Genomes were functionally annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) or with Prokka 1.13 (Seemann, 2014).

2.3.3. Phylogeny and comparative analysis

Average nucleotide identity (ANI) was calculated using pyani 0.2.11 with default parameters (Pritchard et al., 2016), and a phylogeny dendrogram was constructed using Ward's hierarchical clustering method in R version 4.1.0. The gene content of the strains used in this study was compared with Roary 3.7.0

using the Prokka annotation with default parameters (Page et al., 2015). To identify genomic rearrangements and conserved genomic regions within the three clades that we defined, we compared the genomic structure of the complete genome sequences used in this study with Mauve v20150226 (Darling, 2004).

Presence of T2SS, T3SS and T4SS was determined by tBLASTn using the amino acid sequences of the main components of each cluster retrieved from UniProt as query (Supplementary Dataset 1), and the genomic sequences of each strain as subject, with a 0.01 e-value threshold. The presence of each secretion system was then validated by looking for the presence of the cluster in the PGAP annotation at the predicted locus. Presence of T6SS was determined using SecReT6 3.0 with default settings (Li et al., 2015). The T3SS gene cluster was analyzed further by comparing the sequences of the PGAP-annotated genes using Clinker 0.0.21 with default parameters (Gilchrist and Chooi, 2021). In strains CFBP 8304 and LMG 726, the *xopF* gene found in the cluster was incorrectly annotated with PGAP, but was correctly annotated with Prokka, and was thus manually corrected in the PGAP annotation for the cluster comparison. The same was true for the *xopM* gene in strains UPB458, CFBP 8304 and LMG 728.

Presence of type III effectors was determined by BLASTp using the amino acid sequences of the effectors retrieved from http://xanthomonas.org/ as query (Supplementary Dataset 2), and the amino acid sequences of the genes annotated with Prokka as subject, with a 0.01 e-value threshold. Only hits that had >30 % identity over >70 % query sequence length were retained. To validate the presence of each effector, the amino acid sequences of the selected genes were then extracted and used as query for a BLASTp against the type III effectors sequences as subject. Hits that had >30 % identity over >70 % query sequence length were then considered to be putative type III effectors, as discussed in this article. The presence of TAL effectors, their RVD sequence, and their classification were determined using AnnoTALE 1.5 (Grau et al., 2016).

2.4. Results

2.4.1. Genome assembly

For all the sequenced strains, sequence coverages between 81- and 404-fold were obtained. This allowed to assemble complete genome sequences, consisting of one single circular chromosome for all strains (Table 1). Genome sizes ranged from 4,386,175 bp for ICMP 16317 to 4,902,099 bp for LMG 843. In strain CFBP 2055, an additional circular contig of 46,036 pb was assembled. A comparison by BLAST to the NCBI non-redundant nucleotide sequences database showed homology to plasmid sequences. This contig was thus considered to represent a plasmid of pv. *undulosa* strain CFBP 2055.

Table 1 Xanthomonas translucens pathotype strains used in this study and characteristics of the obtained genome assemblies.

Pathovar	Strain	Country	Isolated from	Origin[1]	Coverage	Assembly length	Plasmid length	Accession
arrhenatheri	LMG 727	Switzerland	*Arrhenatherum elatius*	BCCM	103	4,843,101	NA	CP086333
cerealis	CFBP 2541	USA	*Bromus inermis*	CFBP	404	4,504,942	NA	CP074364
graminis	LMG 726	Switzerland	*Dactylis glomerata*	BCCM	81	4,678,781	NA	CP076254
hordei	UPB458[2]	India	*Hordeum vulgare*	BCCM	94	4,679,124	NA	CP076249
phlei	LMG 730	Norway	*Phleum pratense*	BCCM	87	4,569,024	NA	CP076251
phleipratensis	LMG 843	USA	*Phleum pratense*	BCCM	121	4,902,099	NA	CP086332
pistaciae (group A)	CFBP 8304	Australia	*Pistacia vera*	CFBP	301	4,599,174	NA	CP074365
pistaciae (group B)	ICMP 16317	Australia	*Pistacia vera*	ICMP	161	4,386,175	NA	CP083804
poae	LMG 728	Switzerland	*Poa trivialis*	BCCM	86	4,792,655	NA	CP076250
secalis	CFBP 2539	Canada	*Secale cereale*	CFBP	186	4,565,955	NA	CP074363
translucens	DSM 18974[3]	USA	*Hordeum vulgare*	DSMZ	223	4,715,357	NA	LT604072
undulosa	CFBP 2055	Canada	*Triticum turgidum* subsp. *durum*	CFBP	379	4,653,288	46,036	CP074361-CP074362

[1] DSMZ: German Collection of Microorganisms and Cell Cultures; CFBP: Collection of Plant Pathogenic Bacteria; ICMP: International Collection of Microorganisms from Plants; BCCM: Belgian Coordinated Collections of Microorganisms; UPB: Collection of plant pathogenic bacteria, Earth&Life Institute, UCLouvain.
[2] UPB458 is the name of the strain in the UPB collection from where it was obtained and corresponds to LMG 737 in the BCCM collection.
[3] Jaenicke et al., 2016

2.4.2. Phylogeny and comparative analysis

Phylogeny based on ANI revealed that though all strains shared >95 % ANI, three distinct groups could be observed, with ANI values above 97 % (Fig. 1). The first group consisted of strains CFBP 2055, CFBP 2539, UPB458, DSM 18794 and CFBP 8304, the second consisted of strains ICMP 16317 and CFBP 2541, and the third consisted of strains LMG 730, LMG 843, LMG 728, LMG 727 and LMG 726. These three groups will thereafter be referred to as clades Xt-I, Xt-II and Xt-III, respectively. These three clades could also be observed when including all the publicly available complete genome sequences of *X. translucens* strains, with ANI values above 97 % as well (Suppl. Fig 1). Interestingly, the two strains regarded as pv. *pistaciae* were genetically distinct and were found in two separate clades.

Comparison of the genomic structure within each clade showed 34 to 48 locally collinear blocks (LCB) in clade Xt-I (Fig. 2A), eleven LCB in clade Xt-II (Fig. 2B), and 175 to 312 LCB in clade Xt-III (Fig. 2C). These LCB correspond to genomic regions that are conserved between the compared strains, showing no rearrangement. Thus, the low number of LCB in clade Xt-II showed that there are very few genomic rearrangements between the two strains of the clade, while there are more rearrangements in clade Xt-I, and the most in clade Xt-III.

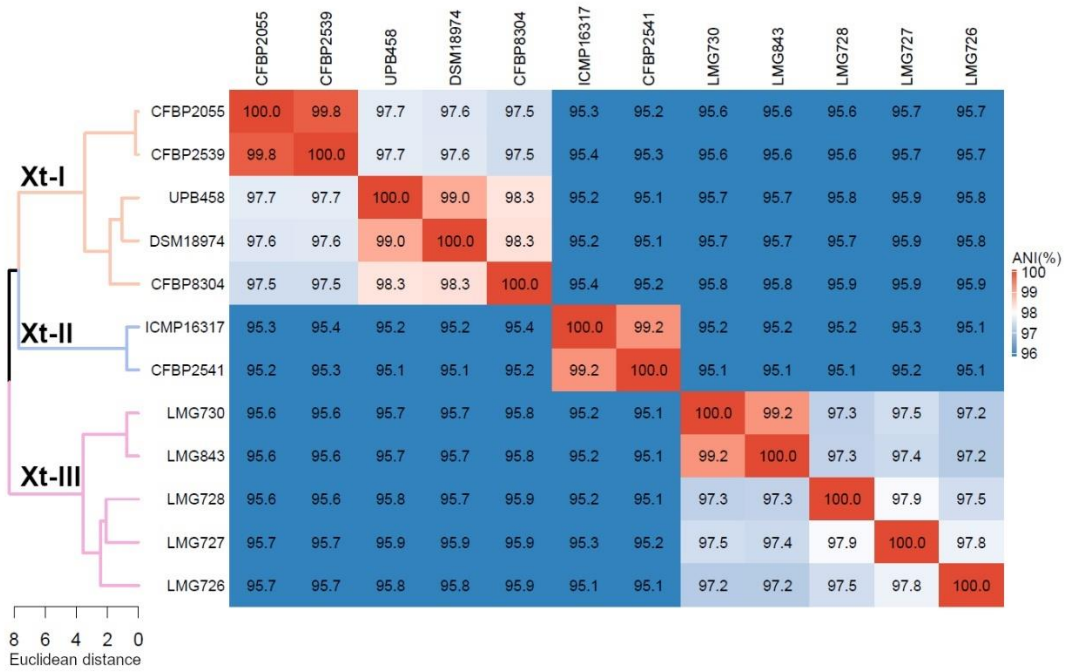


Figure 1 Average nucleotide identity (ANI) of X. translucens pathotype strains and ANI-based phylogeny constructed with Ward's hierarchical clustering method. Distance in the dendrogram represents ANI dissimilarity between nodes. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III. ANI is depicted as a gradient from blue (< 96%) to white (98%) to red (100%).

The number of genes found in each strain with the Prokka annotation ranged from 3,735 genes in ICMP 16317 to 4,262 genes in LMG 726. The pangenome of the twelve *X. translucens* strains used in this study consisted of 9,772 genes, while the core genome consisted of 2,181 genes (Fig. 3). A total of 190 genes were exclusive to clade Xt-I, 588 genes to clade Xt-II, and 168 genes to clade Xt-III. Moreover, a total of 3,681 genes were exclusive to one strain only, ranging from 48 genes in strain CFBP 2539 to 966 genes in LMG 726.

In all strains, a *xps* T2SS and a T3SS were identified, while no *xcs* T2SS was identified in any strain (Fig. 4A). A T6SS-i3*** was identified in all strains of clades Xt-I and Xt-II, while a T6SS-i4 was identified in strains CFBP 2055, CFBP 2539, UPB458, ICMP 16317, CFBP 2541 and LMG 843. A T4SS was identified in strains DSM 18974, CFBP 8304, LMG 730, LMG 728 and LMG 727 which all lacked a T6SS-i4. The strain LMG 726 was the only one that did not possess a T4SS or a T6SS.

Direct comparison of the *hrp* cluster showed that all strains share the same genetic organization of the cluster (Fig. 5). Moreover, most of the main components of the T3SS, from *hrcC* to *hrpD*, as well as *hpaH* and *xopF* were very conserved, with >80 % identity across the species. However, the *hrpE* structural component was the most variable, with as little as 60 % identity within clades Xt-I and Xt-III (Fig. 6). In strains LMG 728 and LMG 726, an additional gene was found between *hrpX* and *hrcT*, in opposite directions in each strain, but showing 100 % identity. This gene showed a high identity with insertion sequence (IS) 5 family transposases found in other *Xanthomonas* species by BLASTx against

the NCBI non-redundant protein sequences database. Moreover, in strains LMG 727 and LMG 726, one additional gene was predicted between *hpaB* and *hrpG*, with 61 % identity between the two strains. However, no similarity to a known gene was found by BLASTx. Additionally, *hpaC* was found to be truncated in DSM 18974 due to an early stop codon, but still showed between 71 and 86 % identity with the other strains (Fig. 6).

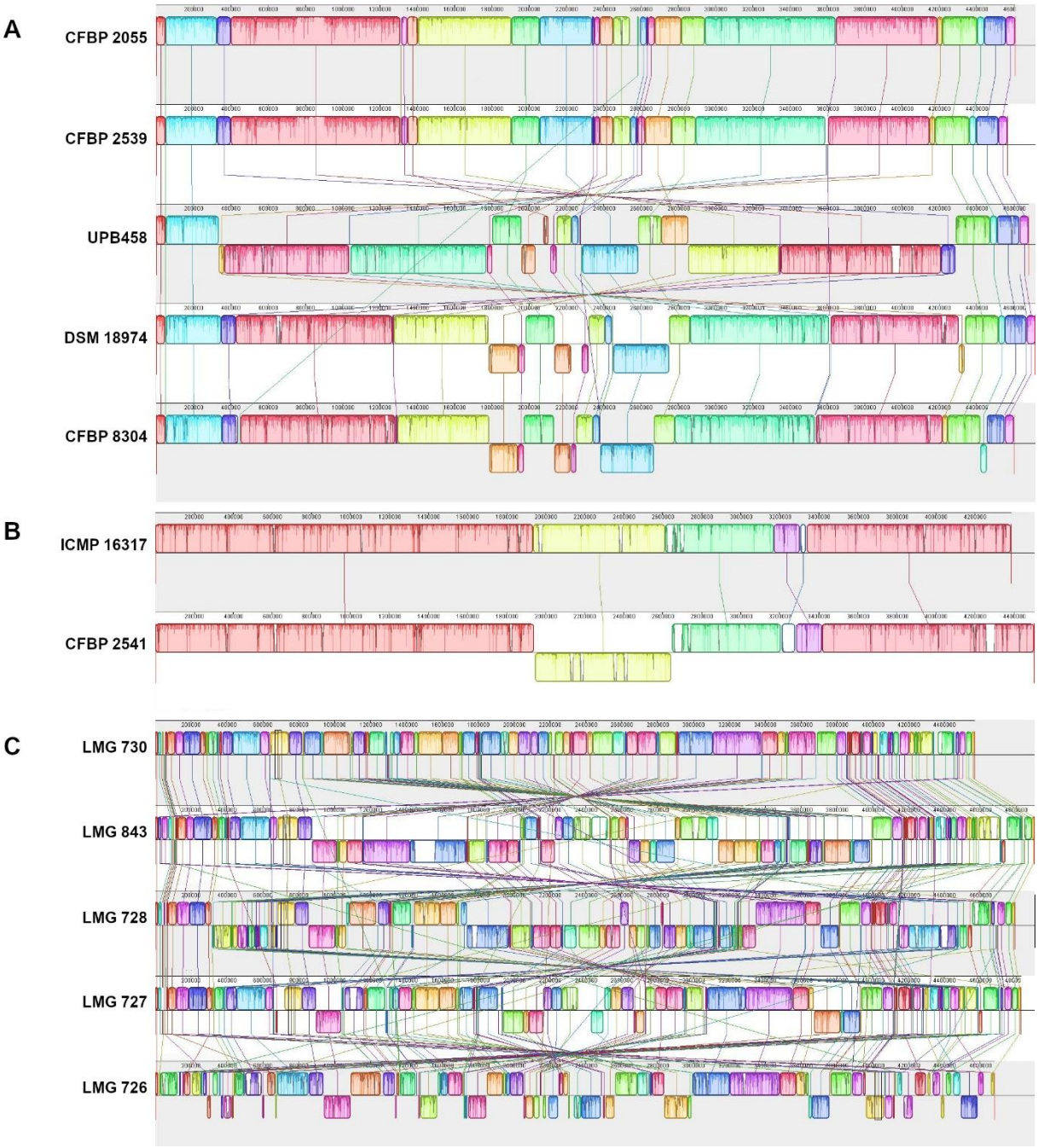


Figure 2 Pairwise comparisons of the genomic structure within the three X. translucens clades with progressive Mauve (Darling, 2004). **(A)** Clade Xt-I, **(B)** clade Xt-II, **(C)** clade Xt-III. Colors represent conserved genomic regions (locally collinear blocks, LCBs), i.e., regions with no rearrangement across all the compared genome sequences. Lines between strains link LCBs that are orthologous between two genome sequences. LCBs found on the bottom part represent regions that are in reverse orientation compared to the reference. In each comparison, the sequence on top is used as reference

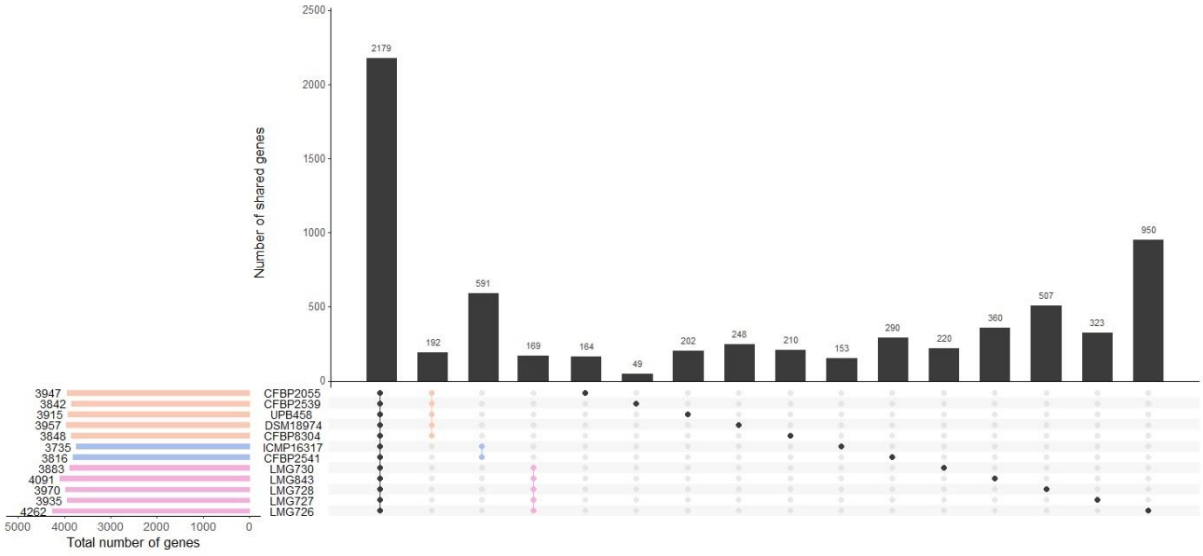


Figure 3 Gene content of the 12 X. translucens pathotype strains used in this study. Vertical bars show the number of genes exclusive to the strains marked in the matrix below. The first group represents the core genome of the species. The second, third, and fourth groups represent genes that are exclusive to the three clades, as indicated by their respective color. Other groups represent genes that are exclusive to one strain. Horizontal bars show the total number of genes in each strain. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III.

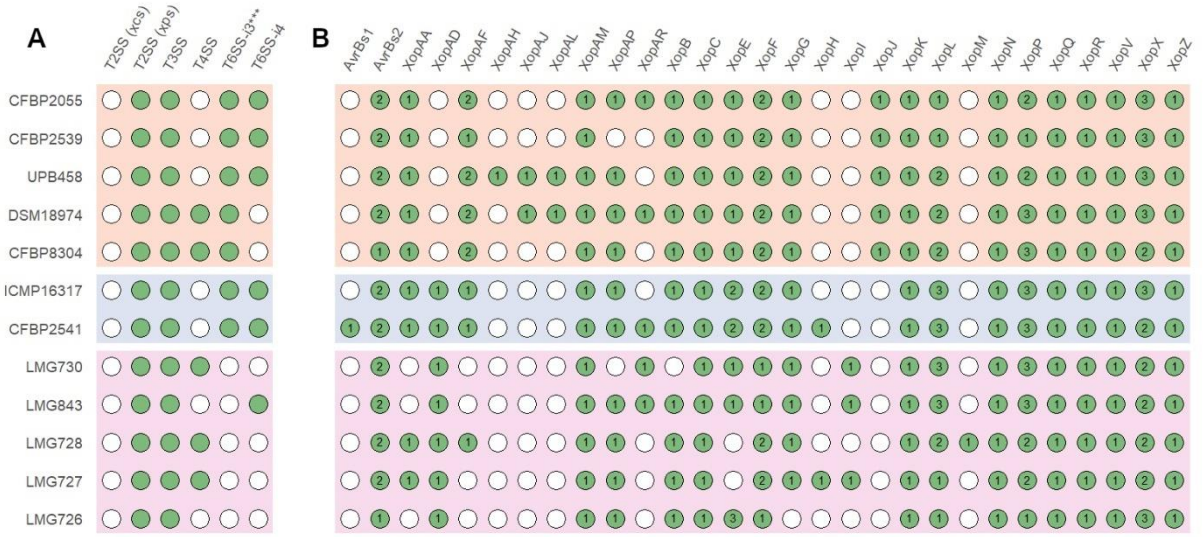


Figure 4 (A) Presence of the xps and xcs type II secretion systems (T2SS), type III secretion system (T3SS), type IV secretion system (T4SS), subtype 3*** type VI secretion system (T6SS-i3***), and subtype 4 type VI secretion system (T6SS-i4) gene clusters in X. translucens pathotype strains. **(B)** Presence of putative type III effectors. Green circles indicate presence, and white circles indicate absence. Numbers denote the number of putative effectors of each class. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III

The T3SS-associated components upstream and downstream of the core *hrp* cluster (*hpaT*, *hpa1*, *xopM* and *hgiB*) showed more variability. The putative translocon *hpaT* was very conserved between the two clade Xt-II strains and showed between 63 and 89 % identity in clade Xt-I (Fig. 6). It was much more variable in clade Xt-III, where it showed between 39 and 67 % identity. The second putative translocon component, *hpa1*, was very conserved among the two clade Xt-II strains (98 % identity), and among the UPB458, DSM 18974 and CFBP 8304 strains in clade Xt-I. It was, however, more dissimilar between these three strains and the other two of the clade, with 45 to 47 % identity. Within the Xt-III clade, there

was a high variability, with 50 to 79 % identity between strains. Moreover, one additional gene was predicted between *hpaT* and *hpa1* in UPB458, but no similarity to any known gene was found by BLASTx.

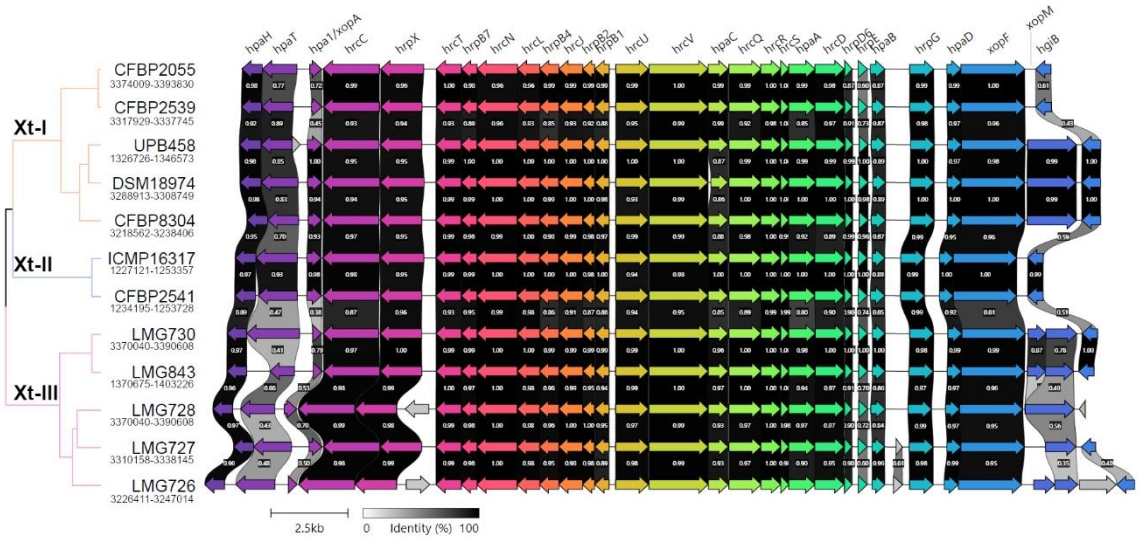


Figure 5 Pairwise comparisons of the genetic content of the type III secretion system cluster in the X. translucens pathotype strains obtained with Clinker (Gilchrist and Chooi, 2021). The phylogenetic tree is based on ANI as shown in Figure 1, orange: Clade Xt-I, blue: clade Xt-II, pink: clade Xt-III. Colors of the arrows represent groups of similar genes. Clusters for strains ICMP 16317, CFBP 2541, and LMG 727 are found in opposite direction and were reversed in the figure.

In clade Xt-I, the XopM effector was only present in UPB458, DSM 18974 and CFBP 8304 and was 99 % identical between the three strains. The effector was also found in LMG 728, where it was 92 to 93 % identical to the ones found in clade Xt-I. In the other four strains of clade Xt-III, the gene was disrupted by early stop codons. The *hgiB* gene was also very variable, with as low as 43 % identity within clade Xt-I, and 37 % identity within clade Xt-III, but with 99 % identity between the two clade Xt-II strains. In LMG 728, it was found to be truncated. Furthermore, an additional gene was found between *xopM* and *hgiB* in LMG 726, which showed 100 % identity to an IS4 family transposase by BLASTx.

Together with XopF and XopM, a total of 29 putative non-TAL type III effector classes were found to be present in the twelve *X. translucens* strains used in this study, ranging from 21 effectors in LMG 726 to 31 in CFBP 2541 and DSM 18974 (Fig. 4B). Among the predicted type III effectors, AvrBs2, as well as effectors of classes XopC, XopF, XopK, XopL, XopN, XopP, XopQ, XopR, XopV, XopX, XopZ and XopAM were found in all strains. Effectors of classes XopB, XopG, XopE, XopAA and XopAF were conserved in all strains of clades Xt-I and Xt-II, but present in only some strains of clade Xt-III. XopJ was found only in clade Xt-I, while XopAD was found only in clades Xt-II and Xt-III. Interestingly, the *xopM* gene located downstream of the *hrp* cluster was not found in this analysis, the XopM effector identified in LMG 728 being a different one. However, this XopM was found to be present in some strains by visual inspection of the gene clusters (Fig. 5).

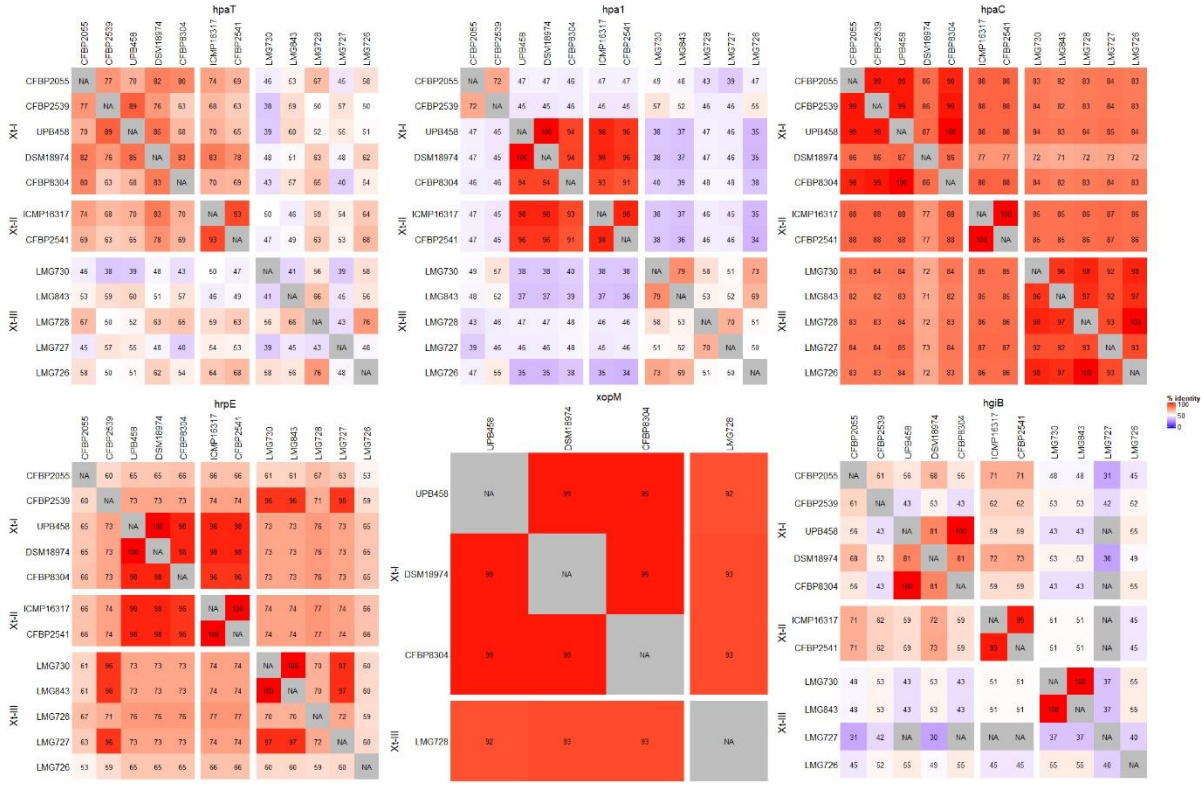


Figure 6 Pairwise comparisons of the nucleotide sequences of the most variable components of the type III secretion system across all X. translucens pathotype strains. For xopM, only UPB458, DSM 18974, CFBP 8304, and LMG 728 were considered, as they were the only strains where a complete gene was identified.

A total of 21 TAL effector classes were identified in the twelve strains used in this study, with up to eight in DSM 18974 (Table 2). Clade Xt-I was the clade with the most TAL effectors. However, the TAL effector repertoires of the clade were very diverse, with only TalDA being present in all strains of the clade. Interestingly, TalDA was also found in ICMP 16317. Clade Xt-II and Xt-III strains had a much smaller set of TAL effectors, ranging from zero to three, with most of them being exclusive to one strain, except for TalIT, which were found in both LMG 843 and LMG 727. Additionally, a potential pseudo-TAL effector was identified in LMG 727, with only two repeats. However, the second repeat is only 19 amino acids long and many stop codons were found in the N-terminus. LMG 726 was the only strain in which no TAL effectors were identified.

Table 2 Classes and repeat variable diresidue (RVD) sequences of transcription activator-like (TAL) effectors found in *Xanthomonas translucens* pathotype strains.

TAL class[1]	Strain	RVD sequence	Chromosome position (location:strand)
TalCT	CFBP 2055	HN-HD-HD-HD-NI-NI-NI-HN-HD-HD-NN-NN-NI-NN-HD	1,324,426-1,327,651:1
	CFBP 2539	NN-HD-HD-HD-NI-NI-NI-HN-HD-HD-NN-NN-NI-NN-HD	1,325,514-1,328,739:1
	DSM 18974	NN-HD-HD-HD-NI-NI-NI-NN-HD-HD-NN-NN-NI-NN-HD	671,336-674,561:1
	UPB458	NN-HD-HD-HD-NI-NI-NI-NN-HD-HD-NN-NN-NI-NN-HD	3,936,481-3,939,712:-1
TalCU	DSM 18974	NG-HD-HD-HN-NG-NI-HG-HG-HD-ND-NN-NN-NI-NH-QD	3,565,727-3,568,955:-1
TalCV	DSM 18974	NG-NN-HD-HD-NN-NI-HG-HD-ND-HG-NI-NN-HD	3,562,396-3,565,408:-1
	UPB458	NG-NN-HD-HD-NN-NI-HG-HD-ND-HG-NI-NN-HD	1,087,442-1,090,460:1
TalCW	DSM 18974	NN-NI-HN-HD-NI-NH-NG-HN-HD-HD-HD-NI-QD	660,311-663,329:1

TalCX	CFBP 8304	NN-HD-NG-NN-HN-KG-NI-HD-NI-NK-HD-HD-HD-HD-NI-HN-NH-HD-QD	674,110-677,491:1
	DSM 18974	NN-HD-NG-NI-HN-KG-NI-HD-NI-NH-NG-NN-HD-HD-NI-NN-NI-HD-QD	663,647-667,295:1
	UPB458	NN-QD-NG-NN-HN-KG-NI-HD-NI-NH-NG-HN-HD-HD-NI-NN-HD	3,939,840-3,943,281:-1
TalCY	DSM 18974	NI-NG-HN-NN-HD-NG-ND-NK-QD-NH-QD	668,858-671,207:1
TalCZ	CFBP 2055	NH-NN-HD-NN-HD-NH-HD-YK-NG-NH-Y*-HD-NN-NI-NG-QD	1,916,521-1,919,857:-1
	CFBP 2539	NH-NN-HD-NN-HD-NH-HD-YK-NG-NH-Y*-HD-NN-NI-NG-QD	1,917,023-1,920,359:-1
	DSM 18974	NH-NN-HD-NN-HD-NH-HD-YK-NG-NH-Y*-HD-NN-NI-NG-QD	1,995,271-1,998,607:-1
TalDA	CFBP 2055	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-QD-NH-HD	593,184-596,364:1
	CFBP 2539	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-QD-NH-HD	594,296-597,476:1
	CFBP 8304	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NK-QD-NH-QD	653,260-656,587:1
	DSM 18974	NN-HD-NG-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-HD-NH-HD	627,243-630,573:1
	ICMP 16317	HD-YD-NI-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-QD-NH-HD	586,137-589,323:1
	UPB458	NN-HD-NG-NG-NG-NN-YK-NG-HD-NG-NG-ND-NG-HD-NH-HD	4,001,560-4,004,884:-1
TalDC	CFBP 2055	NN-NG-HD-HD-HD-KG-NN-Y*-NG-HD-HD-QD-HN	1,321,073-1,324,088:1
TalDD	CFBP 2055	NN-HD-NG-NN-HN-KG-NI-HD-NI-NN-HD-HN-HD-HD-NI-HN-HD-QD	612,552-616,092:1
	CFBP 2539	NN-HD-NG-NN-HN-KG-NI-HD-NI-NN-HD-HN-HD-HD-NI-HN-HD-QD	613,664-617,204:1
	CFBP 2539	NN-HD-NG-NN-HN-KG-NI-HD-NI-NN-HD-HD-NN-NN-NI-HN-HD	1,321,735-1,325,176:1
TalDE	CFBP 2539	NN-HD-NG-NN-HN-HN-NI-NI-NI-NH-NN-HD-NN-NH-HD-HD	1,717,212-1,720,551:1
	CFBP 2055	NN-HD-NG-NN-HN-HN-NI-NI-NI-NH-NN-HD-NN-NH-HD-HD	1,716,713-1,720,052:1
TalDF	CFBP 2055	HD-HN-HN-HD-NH-NH-HG-HD-ND-NN-Y*-NG-HD-NI-NH-NG-HD-HN	1,711,799-1,715,336:1
	CFBP 2539	HD-HN-HN-HD-NH-NH-HG-HD-KG-NN-Y*-NG-HD-NI-NH-NG-HD-HN	1,712,298-1,715,835:1
TalDO	CFBP 2541	NN-NN-KI-NN-HD-NG-HD-NG-NG-NK-HD-HD-NN-QD-NG-QD	2,549,947-2,553,289:1
TalDP	CFBP 2541	NS-KI-NI-HD-NK-GI-HD-NK-HD-NN-HD-NK	590,471-593,369:1
TalJU	LMG 843	NS-NG-HD-HD-NN	1,670,023-1,673,728:-1
TalJS	LMG 727	NN-NG-NN-NG-HD-NK-NG-NI-DD-NK-HD-NG-NN-NI-NG-NN-HD-HD-QD	3,139,389-3,143,094:1
	LMG 843	NN-NG-NN-NG-HD-NK-NG-NK-DD-NK-DD-NG-NN-NI-NG-NN-HD-HD-QD	2,289,805-2,292,265:1
TalIY	UPB 458	NI-NG-HN-NK-HD-NH-HN-HD-HD-HD-HD-QD	1,816,261-1,819,174:-1
TalJB	CFBP 8304	NH-NN-HE-NK-HD-NK-HD-YK-NG-NH-Y*-HE-NI-NI-NG-QD	1,986,200-1,989,539:-1
TalJT	LMG 730	NN-NN-HK-HK-HD-HN-HD-NN	2,152,104-2,154,627:1
TalJV	LMG 728	NN-NI-HD-HD-HD-HD-KT-NG-NN-NN-KT-NI-NN-HD-NG-NG-NN-HD-NK	2,784,954-2,788,575:-1
TalJW	LMG 728	HK-HD-HN-NI-NG-HD-HN-NI-NG-HD-NG-NN-HN-HD-NG-NN-NI-HD-NG-NN-HD-HD-QD	2,798,245-2,802,361:-1
TalJX	LMG 728	NK-NG-NI-HD-NG-NN-NG-HD-NK-N*-NK-HD-NN-HD-NG-NN-NI-NG-HD-HD-NN-HD-HD-NN-HD-HD-QD	1,903,396-1,907,938:-1
Pseudo TAL	LMG 727	NI-NN	2,236,495-2,238,688:1

1 TAL effector classes based on AnnoTALE (Grau et al., 2016).
* : 13th residue is missing

2.5. Discussion

In this study, we generated high-quality complete genome sequences of all pathotype strains of *X. translucens* and make them available as a community resource for in-depth comparative genome analyses within one of the most important pathogenic bacterial species. These are the first complete genome sequences for strains of the pathovars *arrhenatheri*, *graminis*, *hordei*, *phlei*, *phleipratensis*,

pistaciae, *poae* and *secalis*. These resources complement the complete genome of the pv. *translucens* pathotype strain DSM 18974 (Jaenicke et al., 2016), as well as the already available genome sequences of pv. *cerealis, translucens* and *undulosa* strains.

Phylogeny based on ANI revealed that three genetically distinct groups can be identified, with members of each clade being less than 96% identical to members of the two other clades. This is in contrast with the usual distinction of only two groups, the "translucens" group and the "graminis" group but is in line with the previous studies suggesting that pv. *cerealis* could be genetically distinct from the other pathovars (Peng et al., 2016; Langlois et al., 2017; Shah et al., 2019). Based on this phylogeny, we propose to classify these groups as clade Xt-I, containing pathovars *hordei, translucens, undulosa*, and *secalis*; clade Xt-II, containing pv. *cerealis*; and clade Xt-III, containing pathovars *arrhenatheri*, *graminis*, *phlei*, *phleipratensis*, and *poae*.

Moreover, the two strains of pv. *pistaciae* were grouped in two different clades, with the group A strain found in clade Xt-I and the group B strain found in clade Xt-II, the two strains being only 95.4 % identical. This confirmed the previous phylogeny based on *gyrB* sequences where the group A strain was more closely related to pathovars *translucens, secalis* and *undulosa*, while the group B strain was closest to pv. *cerealis* (Giblot-Ducray et al., 2009). There is clear evidence that these strains are not directly related, although they share the same host.

These results raise some limitations of the classical pathovar classification used in *Xanthomonas*. Indeed, the pathotype strains of pathovars *secalis* and *undulosa* are 99.8 % identical and share similar genomic organization and virulence features. As they also have a similar host range, this could lead to them being considered as the same taxonomic entity. On the other hand, the two groups of pv. *pistaciae* are very different genetically and considering them under the same pathovar could hinder the better understanding of their respective biology. Despite these limitations, the pathovar classification was surprisingly robust for the other pathovars. Nonetheless, this work suggests that this classification, currently based solely on pathogenicity tests, should be rethought to better reflect the genetic relationships between pathovars and their evolutionary history.

Within the three clades identified by the ANI-based phylogeny, we found 190, 588 and 168 genes that were specific to clade Xt-I, Xt-II and Xt-III, respectively, as well as 48 to 966 genes that were strain-specific. However, these numbers are probably biased by the small number of strains included in the comparison and including more strains would result in a smaller number of clade- and strain-specific genes. Nonetheless, these constitute a valuable list of genes that could shape the host range of each clade and/or strain.

All strains had an *xps* T2SS and no *xcs* T2SS, as well as a T3SS with a similar genetic organisation, although a few genes of the *hrp* cluster were variable in sequence between strains. Interestingly, strains

that possessed a T4SS did not possess a T6SS-i4, and vice-versa. As these two secretion systems are both involved in antimicrobial activity, it is possible that their function is redundant in *X. translucens* and only one of them is required to play this role. However, the presence or absence of these two secretion systems does not necessarily reflect the phylogenetic relationships between the strains investigated in this study. Additionally, a T6SS-i3*** was found in all strains of clades Xt-I and Xt-II. However, it was previously hypothesized that this subgroup of T6SS could be non-functional due to the lack of a PAAR, and could be complemented by the presence of another subtype of T6SS (Bayer-Santos et al., 2019). Interestingly, no T4SS or T6SS were identified in strain LMG 726 which confirms previous research that showed that many pv. *graminis* strains lack a T6SS, with only strains Xtg2, Xtg9, Xtg10 and NCPPB3709 found to harbor one (Hersemann et al., 2017).

We have identified a total of 13 type III non-TAL effector classes that could constitute a core set of effectors in *X. translucens*. Moreover, XopAD was not found in strains of clade Xt-I in our analysis, but was previously identified in strain DSM 18974 in different analyses, indicating that it could also be part of the *X. translucens* core set of type III effectors (Peng et al., 2016; Koebnik et al., 2021; Shah et al., 2021). Five additional effector classes constitute a core set of effectors in clades Xt-I and Xt-II, as well as XopJ, which was specific to clade Xt-I. Some effector classes were specific to one strain, such as AvrBs1 in pv. *cerealis* strain CFBP 2541, or XopAH in pv. *hordei* strain UPB458. Furthermore, XopAJ and XopAL class effectors were found in both pv. *translucens* and *hordei*. These effectors could have a role in the host specificity of each clade and/or pathovar. However, no type III effector has yet been functionally characterized in *X. translucens*, and the effectors identified in this study will need to be functionally validated to confirm their role in the pathogenicity of each pathovar.

High-quality genome assemblies allowed for the identification of the first TAL effector to be reported in grass-infecting *X. translucens* strains, as well as in pathovars *secalis*, *hordei* and *pistaciae*. No TAL effector was found in pv. *graminis* in our analysis, confirming previous research where no TAL effector was identified in draft genomes of the pathovar (Wichmann et al., 2013). The TAL effectors identified in pv. *undulosa* correspond to those previously identified in other strains of this pathovar, except for Xt4699-Tal3, for which no similar TAL effector was found, as was the case in strain ICMP 11055 (Peng et al., 2016; Falahi Charkhabi et al., 2017). The TAL effector of class TalDC that we identified has the same RVD sequence as Xt4699-Tal8, and a similar sequence to ICMP 11055-Tal4b. Additionally, the TAL effector of class TalDD identified in CFBP 2055 and CFBP 2539 has a similar RVD sequence to ICMP 11055-Tal2. Furthermore, the TAL effectors identified in pv. *cerealis* strain CFBP 2541 correspond to the TAL effector previously identified in strain Nxtc01, with the TAL effector of class TalDP having a similar sequence to Nxtc01-Tal1 (Shah et al., 2019). As these four TAL effectors have previously been shown to have a role in virulence, the corresponding effectors we identified in this study could play a similar role in their respective strain (Falahi Charkhabi et al., 2017; Peng et al., 2019; Shah et al., 2019). However, these are the only TAL effectors that have been functionally characterized in

X. translucens to date. Nonetheless, as most of the TAL effectors identified in this study are clade- or strain-specific, they could be essential components of host adaptation. Indeed, TAL effectors activate the transcription of plant genes by binding to their promoter region. As these sequences vary between plant species, TAL effectors must be adapted to target the specific sequence of that promoter in their host and thus reflect processes of co-adaptation between the bacteria and its host (Jacques et al., 2016). However, the lack of TAL effectors in pv. *graminis* indicates that other mechanisms play a role in *X. translucens* host speciation.

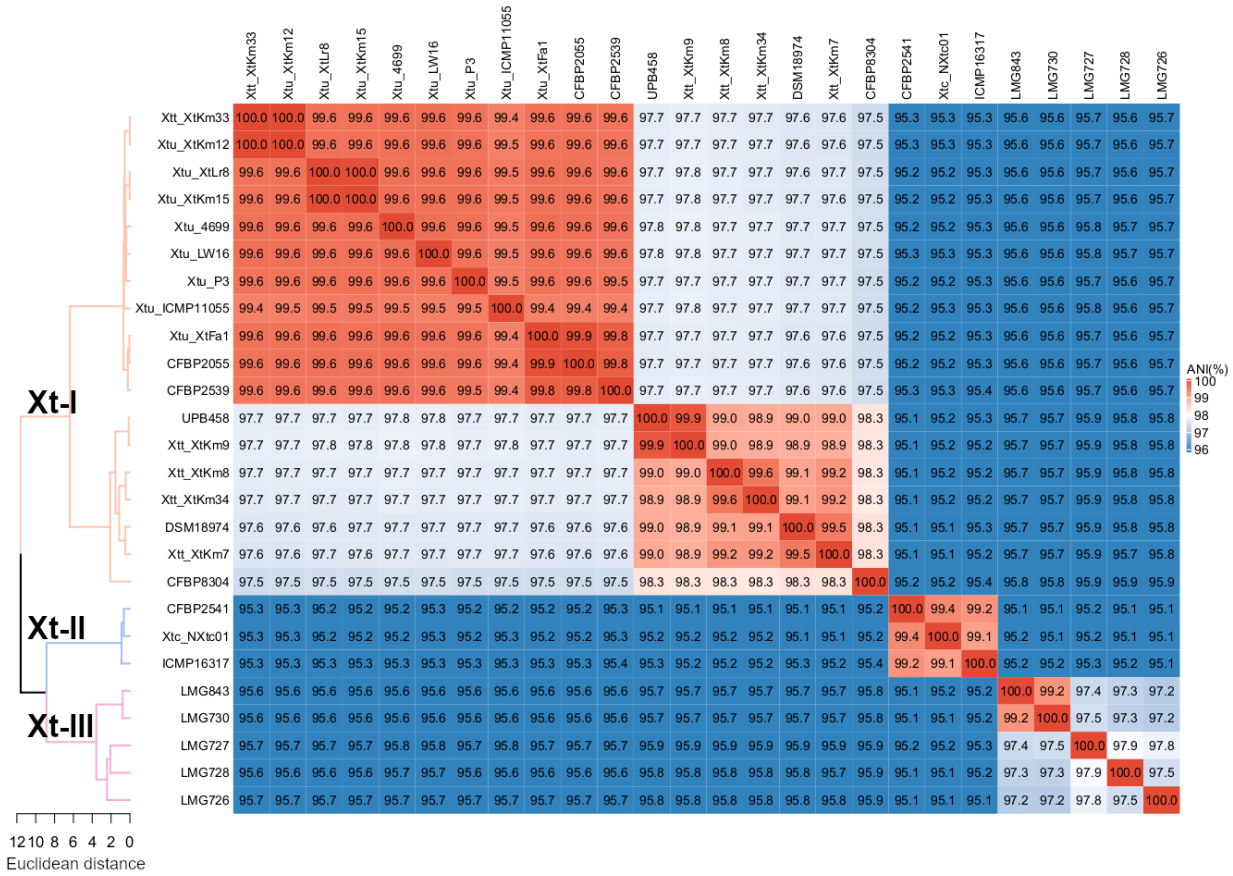
The pv. *pistaciae* group A pathotype strain shared a very similar genomic organization, T3SS cluster organization and gene identity, as well as type III effector and TAL effector repertoires with the pathotype strains of pv. *translucens* and *hordei*. The same was true for the group B pv. *pistaciae* strain and the pv. *cerealis* pathotype strain, which were even more closely related. This could indicate that a very small set of genes is responsible for their ability to infect pistachio, as a previous study showed pv. *translucens* strain DAR 35705 was unable to cause symptoms in pistachio and other *Anacardiaceae*, while pv. *pistaciae* strains of both groups were pathogenic in *Anacardiaceae* and *Poaceae* (Marefat et al., 2006). However, no type III effector or TAL effector exclusive to either strain of pv. *pistaciae* has been identified in this study.

Although it had the biggest set of total genes, the pv. *graminis* strain LMG 726 was found to lack many virulence features such as the T4SS, the T6SS as well as TAL effectors, and had the smallest set of type III effectors. This is surprising, as this pathovar is known to be the most widespread and virulent grass-infecting *X. translucens* pathovar, with the largest host range in clade Xt-III. Indeed, while pv. *phlei* and pv. *phleipratensis*, pv. *arrhenatheri*, and pv. *poae* are restricted to the genera *Phleum*, *Arrhenatherum*, and *Poa*, respectively, pv. *graminis* can infect many grasses from genera such as *Agrostis*, *Alopecurus*, *Dactylis*, *Deschampsia*, *Festuca*, *Lolium*, *Phalaris*, *Phleum*, *Poa* and *Trisetum* (Egli et al., 1975; Egli and Schmidt, 1982). Nonetheless, these results go along previous research that showed that strains of pv. *graminis* also lack a flagellum and have a distinct type IV pilus compared to other *X. translucens* pathovars (Hersemann et al., 2017). It was hypothesized that since pv. *graminis* is usually spread by mowing tools, the lack of a flagellum might not hinder its ability to spread in the plant, as flagellar motility is mostly necessary for the bacteria to reach points of entry in the plant. Similarly, it might encounter less competition with epiphytic microorganisms, and the presence of a T4SS and/or a T6SS might not be crucial for its survival. Additionally, as these features can act as elicitors of plant defense, their absence in pv. *graminis* could help to evade such defense mechanisms – an intriguing hypothesis which, however, has not yet been tested.

In conclusion, our study substantially increased the number of complete genome sequences available for *X. translucens*, providing high-quality genomic resources for all pathovars of the species. These sequences constitute a valuable basis for future studies investigating the phylogenetic relationships

between *X. translucens* pathovars and other *Xanthomonas* species, as well as their key genetic features. The virulence features of *X. translucens* that we identified will help to better understand the biology of each clade and/or pathovar, and processes of adaptation to their respective hosts. Inclusion of high-quality genome sequences of additional strains of each pathovar in further comparative genomics studies will help in refining this list, laying the foundations for the development of new resistant cereal crop and forage grass cultivars.

2.6. Supplementary material



Supplementary Figure 1 Average nucleotide identity (ANI) of all currently available complete genome sequences of *Xanthomonas translucens* strains and ANI-based phylogeny constructed with Ward's hierarchical clustering method. Distance in the dendrogram represents ANI dissimilarity between nodes. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III. ANI is depicted as a gradient from blue (<96 %) to white (98%) to red (100%).

3. High genomic plasticity and unique features of *Xanthomonas translucens* pv. *graminis* revealed through comparative analysis of complete genome sequences

Florian Goettelmann[1], Ralf Koebnik[2], Veronica Roman-Reyna[3], Bruno Studer[1], Roland Kölliker[1*]

[1]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland
[2]Plant Health Institute of Montpellier, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France
[3]Department of Plant Pathology and Environmental Microbiology, The Pennsylvania State University, University Park, PA, United States.

Research article: Published in BMC Genomics, 2023.

Available at: https://doi.org/10.1186/s12864-023-09855-8

Keywords: *Xanthomonas translucens* pv. *graminis*, complete genomes, phylogeny, host adaptation, comparative genomics, virulence factors.

3.1. Abstract

Xanthomonas translucens pv. *graminis* (*Xtg*) is a major bacterial pathogen of economically important forage grasses, causing severe yield losses. So far, genomic resources for this pathovar consisted mostly of draft genome sequences, and only one complete genome sequence was available, preventing comprehensive comparative genomic analyses. Such comparative analyses are essential in understanding the mechanisms involved in the virulence of pathogens and to identify virulence factors involved in pathogenicity. In this study, we produced high-quality, complete genome sequences of four strains of *Xtg*, complementing the recently obtained complete genome sequence of the *Xtg* pathotype strain. These genomic resources allowed for a comprehensive comparative analysis, which revealed a high genomic plasticity with many chromosomal rearrangements, although the strains were highly related. A high number of transposases were exclusively found in *Xtg* and corresponded to 413 to 457 insertion/excision transposable elements per strain. These mobile genetic elements are likely to be involved in the observed genomic plasticity and may play an important role in the adaptation of *Xtg*. The pathovar was found to lack a type IV secretion system, and it possessed the smallest set of type III effectors in the species. However, three XopE and XopX family effectors were found, while in the other pathovars of the species two or less were present. Additional genes that were specific to the pathovar were identified, including a unique set of minor pilins of the type IV pilus, 17 TonB-dependent receptors (TBDRs), and 11 degradative enzymes. These results suggest a high adaptability of *Xtg*, conferred by the abundance of mobile genetic elements, which could play a crucial role in pathogen adaptation. The

large amount of such elements in *Xtg* compared to other pathovars of the species could, at least partially, explain its high virulence and broad host range. Conserved features that were specific to *Xtg* were identified, and further investigation will help to determine genes that are essential to pathogenicity and host adaptation of *Xtg*.

3.2. Introduction

Bacterial wilt of forage grasses, caused by *Xanthomonas translucens* pv. *graminis* (*Xtg*), is one of the main diseases of forage grasses in temperate grasslands (Egli et al., 1975). *Xtg* is a member of the recently-defined clade Xt-III of *X. translucens*, which also contains four other forage grass infecting pathovars (pv. *arrhenatheri*, *phlei*, *phleipratensis* and *poae*) and is distinguished from clades Xt-I and Xt-II, which contain mainly cereal-infecting pathovars such as pv. *translucens* and *cerealis*, respectively (Sapkota et al., 2020; Goettelmann et al., 2022). While most *X. translucens* pathovars are generally restricted to a small host range, *Xtg* is able to infect many grass genera, including *Agrostis*, *Alopecurus*, *Dactylis*, *Deschampsia*, *Festuca*, *Lolium*, *Phalaris*, *Phleum*, *Poa*, and *Trisetum* (Egli et al., 1975; Egli and Schmidt, 1982). It is also the most widespread, being found in most of Europe, America, and New Zealand (Egli and Schmidt, 1982; Roberts et al., 1985; Falloon and Hume, 1988).

In contrast to cereal crops such as wheat, where *X. translucens* is considered a seedborne pathogen, no seed or soil transmission of *Xtg* has been observed on forage grasses so far (Leyns et al., 1988; Duveiller et al., 1992). Here, the bacteria are mainly spread by contaminated mowing tools and penetrate via plant wounds. While some *X. translucens* pathovars such as pv. *undulosa* are restricted to the non-vascular tissue, all forage grass infecting *X- translucens* pathovars spread in the xylem, resulting in a disruption of the flow of water and nutrients (Duveiller et al., 1992; Rudolph, 1993). The disease affects mainly Italian ryegrass (*Lolium multiflorum* Lam.), which is cut several times during the growing season for harvesting. After penetration, the bacteria spread in the xylem, causing a disruption of water and nutrient absorption. The symptoms manifest as curling and wilting of the leaves and tillers, and eventually the death of the plant (Leyns, 1993; Rudolph, 1993). Infection by the pathogen can lead to serious yield losses, estimated to reach up to 40% under unfavourable conditions (Schmidt and Nuesch, 1980). To control the disease in an efficient manner, it is essential to understand the mechanisms involved in the pathogenicity of *Xtg*.

In plant pathogenic bacteria, successful infection of the plant host generally involves virulence factors that allow the bacteria to adhere to the plant surface, invade the host tissue, acquire nutrients, and suppress plant defence mechanisms. In Xanthomonads, some of the most important virulence factors are effectors, in particular type III effectors (T3E), that are secreted into the plant host cells by the type III secretion system (T3SS). These T3E play a major role in pathogenicity by specifically targeting different pathways in the host cell. A total of 53 classes of T3E have been defined, generally referred to

as "*Xanthomonas* outer proteins" (Xop) (White et al., 2009). Other T3E can cause a hypersensitive response in their plant host and are referred to by this characteristic, like AvrBs1 and AvrBs2. Additionally, transcription activator-like (TAL) effectors are specific T3E that consist of repetitive sequences that allow them to bind to specific nucleotide sequences in the host genome and manipulate the expression of specific genes (Boch and Bonas, 2010; Zhang et al., 2022). However, while TAL effectors are characteristic to the *Xanthomonas* genus, none have ever been identified in *Xtg* (Wichmann et al., 2013; Goettelmann et al., 2022).

Other secretion systems also play an important role in the infection of *Xanthomonas* spp., including the type II, IV and VI secretion systems (T2SS, T4SS and T6SS, respectively). The T2SS is mostly involved in the secretion of plant cell wall degradative enzymes in the apoplast, allowing the bacteria to invade the plant host, as well as acquire the resulting degradation products as a source of nutrients (Jha et al., 2005). Two types of T2SS can be found in *Xanthomonas* spp., the *xps* T2SS, which is conserved across the genus, and the *xcs* T2SS, which is only found in some species (Szczesny et al., 2010). To date, no *xcs* T2SS was found in *X. translucens* (Goettelmann et al., 2022).

The T4SS and T6SS have a similar function and are involved in defence against microbial predators and competition with other microorganisms (Büttner and Bonas, 2010; Alvarez-Martinez et al., 2021). Three subtypes of the T6SS have been found in *Xanthomonas* spp., subtypes 1, 3 and 4, and subtype 3 is further divided into subgroups 3*, 3** and 3***. Previously, no T4SS has been identified in *Xtg*, and T6SS genes were identified only in strains Xtg2, Xtg9, Xtg10 and NCPPB 3709 (Hersemann et al., 2017; Goettelmann et al., 2022).

To reliably identify virulence factors in plant pathogenic bacteria, complete and high-quality genome sequences are essential, as they provide an accurate representation of the genes present in each strain. Such genome sequences enable comparative genomics, which can reveal conserved or divergent genes of a pathovar and help define genes that are crucial to pathogenicity. Previously, draft genome sequences of *Xtg* strains CFBP 2053, ICMP 6431, NCPPB 3709, Xtg2, Xtg9 and Xtg29 allowed a first comparative analysis, which revealed specific features of the pathovar, such as the lack of a flagellum and a distinct type IV pilus (Hersemann et al., 2017). Hypothetical proteins specific to *Xtg* were also identified, including genes with predicted functions in nutrient acquisition, regulatory mechanisms, virulence, adhesion, and motility. Effector proteins that were specific to *Xtg* included a YopT-like cysteine protease and a XopJ class effector protein, which were tested for their role in virulence. However, single and double knock-out mutants for these genes showed no significant difference in virulence with the wild type strain (Hersemann et al., 2017).

At that time, genomic resources for *X. translucens* pathovars consisted of highly fragmented and incomplete draft genome sequences based on short read sequencing, which might have prevented the identification of important genes. Furthermore, these genome assemblies fail to assemble complex

repetitive sequences such as TAL effectors, which may have hindered their identification. Long read sequencing technologies now allow to assemble complete bacterial genome sequences, allowing for a comprehensive overview of the gene content within these genomes, including TAL effectors (Peng et al., 2016; Erkes et al., 2023). Recently, high-quality, complete genome sequences were produced for all pathotype strains of *X. translucens*, which allowed for a more comprehensive comparative analysis (Goettelmann et al., 2022). This revealed that the *Xtg* pathotype strain LMG 726 lacked many virulence features, such as a T4SS and T6SS, and had the smallest set of T3E. However, the inclusion of only one strain of *Xtg* did not allow for an in-depth comparative analysis. In this study, we aimed at producing high-quality complete genome sequences for four additional strains of *Xtg*, perform a comprehensive comparative genomic analysis of the pathovar, and identify unique features that could play a role in its virulence and host specificity.

3.3. Materials and methods

3.3.1. Bacterial strains, growth conditions and DNA extraction

The *Xtg* strains NCPPB 3709, Xtg2, Xtg9 and Xtg29 were selected to be sequenced in this study (Table 1). These strains were grown at 28°C on YDC agar medium (2% dextrose, 1% yeast extract, 2% CaCO3, 1.5% agar) for 48 h. Bacteria were then dissolved in 10 mL washing buffer (50 mM TRIS-HCl pH 8.0, 50 mM EDTA pH 8.0, 150 mM NaCl). Genomic DNA was then extracted with the NucleoSpin® Microbial DNA kit (Macherey Nagel, Duren, Germany), according to the manufacturer's recommendations.

3.3.2. Sequencing, genome assembly and annotation

Library preparation and DNA sequencing was performed at the Functional Genomics Center Zurich. Libraries were prepared and multiplexed with the PacBio SMRTbell® Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) according to the published protocol (Procedure & checklist – preparing multiplexed microbial libraries using SMRTbell express template prep kit 2.0, 2021). After shearing and size selection to 7-10 kb, 1 µL of the non-size-selected library was spiked in the final solution, for a final volume of 10 µL, to allow for sequencing of potential plasmids. Libraries were then sequenced on the PacBio Sequel II platform.

Genomic sequences were then *de novo* assembled with flye 2.7 for the strains Xtg2, Xtg9 and Xtg29 using the "--plasmids --iterations 2" parameters (Kolmogorov et al., 2019). Since the assembly of strain NCPPB 3709 did not result in a complete circular chromosome, we used canu with default parameters for this strain (Koren et al., 2017)

The assembled circular chromosomes were rotated to start with the *dnaA* gene sequence for comparability. Genomes were functionally annotated with Prokka 1.14.6 with default parameters (Seemann, 2014).

3.3.3. Phylogeny and comparative analysis

To determine phylogeny within clade Xt-III, we used all available complete genome sequences of the clade, as well as the genome sequences of pv. *translucens* strain DSM 18974 and pv. *cerealis* strain CFBP 2541 as representatives of clades Xt-I and Xt-II, respectively (Table 1). Average nucleotide identity (ANI) and alignment fraction (AF) were calculated using the anib subcommand of pyani 0.2.12 with default parameters (Pritchard et al., 2016), and a phylogeny dendrogram was constructed using Ward's hierarchical clustering method in R version 4.2.0 (R Core Team, 2022). To identify genomic rearrangements and conserved genomic regions within *Xtg*, the genomic structure of the five *Xtg* genome sequences were compared with Mauve v20150226 (Darling, 2004).

For a comparative analysis of the gene content of *Xtg* compared to other *X. translucens* strains, all currently available complete genome sequences of the *X. translucens* species were retrieved and annotated with Prokka (Table 1, Suppl. Table 1). The gene content within the *X. translucens* species was compared with Roary 3.13.0 using default parameters (Page et al., 2015). This allowed to predict singletons, genes that are exclusive to *Xtg*, by selecting genes that were only found in *Xtg* strains in our dataset. The pangenome and the predicted singleton contents of *Xtg* strains were visualized in Venn diagrams with jvenn (Bardou et al., 2014). Clusters of orthologous groups (COG)-based functional characterization of the gene content of *Xtg* was performed using eggNOG-mapper v2 (Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021). As the COG analysis revealed a high number of transposable elements in *Xtg*, these were further characterized using the mobileOG-db tool in the Proksee online software, with default parameters (Brown et al., 2022; Grant et al., 2023). Additionally, putative horizontal transfer events were identified using the Alien Hunter tool in Proksee (Vernikos and Parkhill, 2006). The presence of signal peptides or non-classical secretion were predicted using SignalP 6.0 and SecretomeP 2.0, respectively (Bendtsen et al., 2005; Teufel et al., 2022).

Minor pilins identified in the singleton analysis were investigated with InterProScan by identifying the characteristic pilin N-terminal transmembrane domain, as well as with the PilFind online tool (Imam et al., 2011; Jones et al., 2014). To compare the two filamentous hemagglutinin singleton genes identified with similar previously-identified genes, the nucleotide sequence of these genes in the Xtg29 draft genome sequence in which they were identified was extracted (Hersemann et al., 2017). These sequences were then used as query in a BLASTn analysis using the nucleotide sequences of the genes identified in this study or the whole genome sequences of *Xtg* strains as subject. The DNA sequences of these gene clusters were compared with Clinker v0.0.27 (Gilchrist and Chooi, 2021).

Presence of T2SS, T3SS, T4SS and type III effectors was determined as detailed by (Goettelmann et al., 2022). Presence of T6SS was determined using SecReT6 3.0 with default settings (Li et al., 2015). The presence of TAL effectors was determined using AnnoTALE 1.5 (Grau et al., 2016).

3.4. Results

3.4.1. Features of the *Xtg* genome sequences

Each strain was assembled to one circular complete chromosome. The chromosome sequences were of comparable size, ranging from 4,673,077 bp for Xtg29 to 4,734,182 bp for Xtg9. The sequence coverage ranged from 64- to 89-fold for all the sequenced strains (Table 1). Additionally, a circular contig sequence of 44,971 bp was assembled in strain NCPPB 3709. A comparison by BLASTn to the NCBI non-redundant nucleotide sequences database showed homology to plasmid sequences. This contig was thus considered to represent a plasmid of this strain.

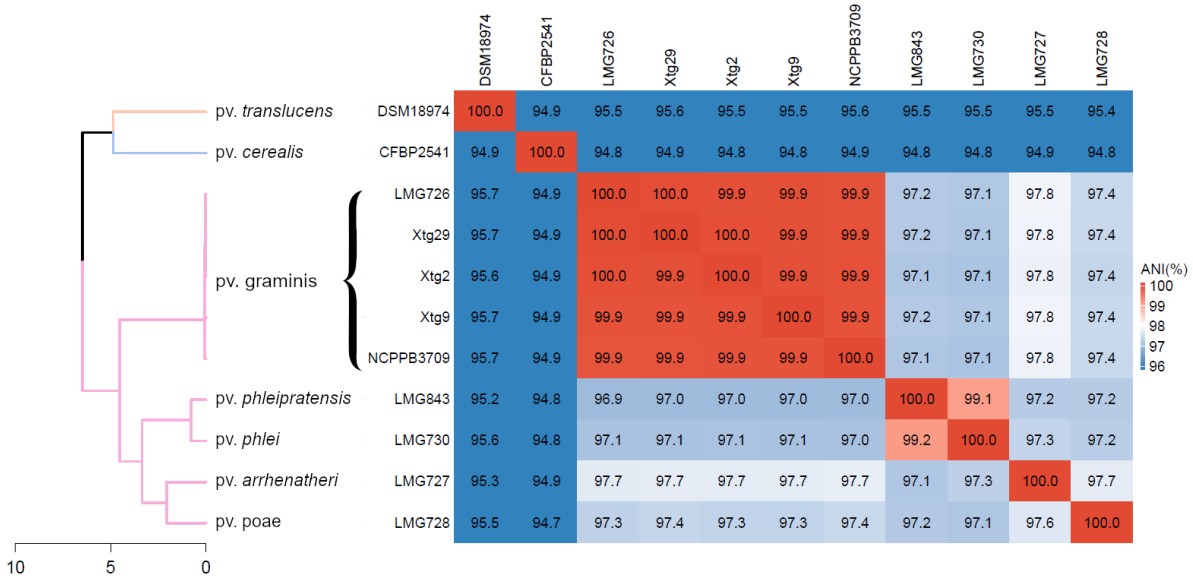


Figure 1 Average nucleotide identity (ANI)-based phylogeny of *Xanthomonas translucens* clade Xt-III, including all complete genome sequences currently available for the clade, as well as DSM 18974 and CFBP 2541 as representatives of clades Xt-I and Xt-II, respectively. Phylogeny was constructed with Ward's hierarchical clustering method. ANI is depicted as a gradient from blue (< 96%) to white (98%) to red (100%). Distance in the dendrogram represents dissimilarity between nodes. Orange: clade Xt-I, blue: clade Xt-II, pink: clade Xt-III.

3.4.2. *Xtg* displays a high genomic plasticity

Phylogeny based on ANI revealed that all strains of *Xtg* were very similar, with 99.9% to 100% ANI (Figure 1). The alignment fraction was >95% between *Xtg* strains, further supporting their high similarity (Suppl. figure 1). Comparison of the genomic structure within *Xtg* showed 65 to 124 locally collinear blocks (LCB), genomic regions that show no rearrangements between the compared strains (Figure 2). This showed that although *Xtg* strains are very close genetically, a high level of rearrangements can be seen within the pathovar.

Table 1 *Xanthomonas translucens* strains used for comparative genomic analysis and characteristics of the genome assemblies. The pv. *translucens* and pv. *cerealis* pathotype strains DSM 18974 and CFBP 2541 were included as representatives of clades Xt-I and Xt-II, respectively.

Pathovar	Strain[1]	Country	Isolated from	Origin[2]	Coverage	Chromosome length (bp)	Plasmid length (bp)	Accession	Reference
graminis	NCPPB 3709	Norway	*Lolium multiflorum*	NCPPB	89	4,732,427	44,971	CP118161-CP118162	This study
	Xtg2	Switzerland	*Lolium multiflorum*	(Kölliker et al., 2006)	78	4,722,726	NA	CP076253	This study
	Xtg9	Switzerland	*Lolium multiflorum*	(Kölliker et al., 2006)	84	4,734,182	NA	CP076252	This study
	Xtg29	Switzerland	*Lolium multiflorum*	(Kölliker et al., 2006)	64	4,673,077	NA	CP076257	This study
	LMG 726	Switzerland	*Lolium multiflorum*	BCCM	81	4,678,781	NA	CP076254	(Goettelmann et al., 2022)
arrhenatheri	**LMG 727**	Switzerland	*Arrhenatherum elatius*	BCCM	103	4,843,101	NA	CP086333	(Goettelmann et al., 2022)
poae	**LMG 728**	Switzerland	*Poa trivialis*	BCCM	86	4,792,655	NA	CP076250	(Goettelmann et al., 2022)
phlei	**LMG 730**	Norway	*Phleum pratense*	BCCM	87	4,569,024	NA	CP076251	(Goettelmann et al., 2022)
phleipratensis	**LMG 843**	USA	*Phleum pratense*	BCCM	121	4,902,099	NA	CP086332	(Goettelmann et al., 2022)
cerealis	**CFBP 2541**	USA	*Bromus inermis*	CFBP	404	4,504,942	NA	CP074364	(Goettelmann et al., 2022)
translucens	**DSM 18974**	USA	*Hordeum vulgare*	DSMZ	223	4,715,357	NA	LT604072	(Jaenicke et al., 2016)

[1]Pathotype strains are marked in bold.

[2]NCPPB: National Collection of Plant Pathogenic Bacteria (Fera Science Ltd., Sand Hutton, United Kingdom); BCCM: Belgian Coordinated Collections of Microorganisms; CFBP: Collection of Plant Pathogenic Bacteria; DSMZ: German Collection of Microorganisms and Cell Cultures.

[3]TBDRs: TonB-dependent receptors

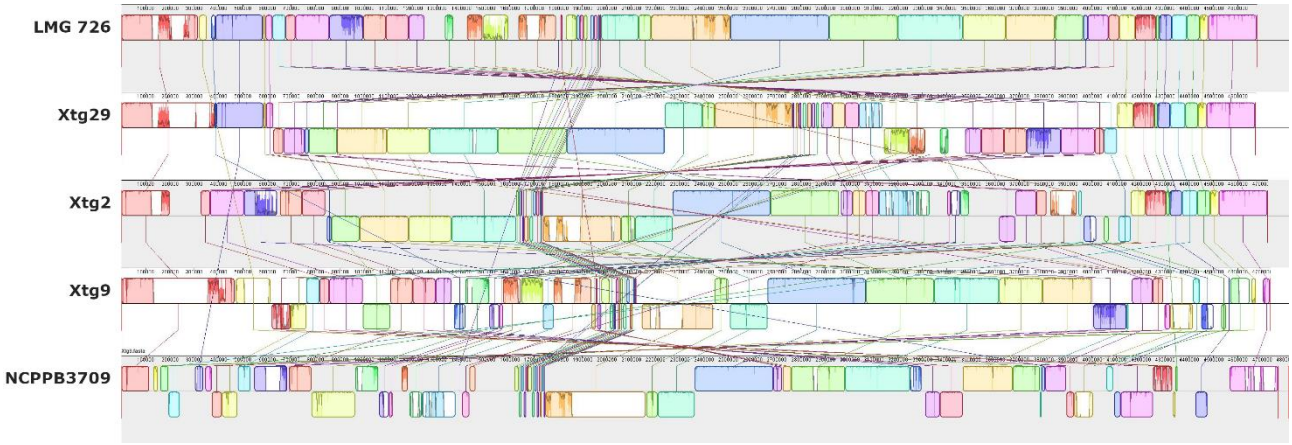


Figure 2 Pairwise comparisons of the chromosome structure within *Xanthomonas translucens* pv. *graminis* using all five currently available complete genome sequences of the pathovar. The LMG 726 genome sequence at the top was used as a reference. Colours represent locally colinear blocks (LCB), regions that show no rearrangement across all the compared genome sequences. Synteny is displayed by a line that links orthologous LCBs. LCBs found on the bottom part represent regions that are found in reverse orientation compared to the reference.

The number of genes found in each strain of *Xtg* was similar, ranging from 4,408 in Xtg9 to 4,445 in NCPPB 3709 (Table 2). The pangenome, i.e., the total number of genes, of the five *Xtg* strains used in this study consisted of 4,931 genes, while the core genome, i.e., the set of genes shared by all strains, consisted of 3,739 genes (Fig. 3A). Moreover, a total of 1,409 predicted singletons, genes that were not found in any other pathovar of the *X. translucens* species, were identified, with 536 being shared by all *Xtg* strains (Fig. 3B). A total of 433 predicted singletons were found to be strain-specific, while 131 were shared by two strains, 164 were shared by three strains, and 145 were shared by four strains. The total number of predicted singletons per strain ranged from 856 in Xtg9 to 922 in Xtg29 (Table 2).

The functional characterization based on COG (Fig. 4) showed that many genes within *Xtg* could not be functionally annotated (13-15%) or were of unknown function (15-16%). The largest category with a known function was replication, recombination, and repair, with 733 to 781 genes of this category per strain (16-18%). Furthermore, this category was also largely represented in predicted singletons (48-53%), ranging from 424 genes of this category in NCPPB 3709 to 453 in Xtg9. Investigation of these genes revealed that most of them were annotated as transposases, suggesting a high transposase activity in *Xtg*. Further investigation of mobile genetic elements in the *Xtg* genome sequences showed an overrepresentation of integration/excision elements, corresponding with the previously identified transposases, with 413 to 457 of such elements found per strain, compared to 24 to 123 in other strains of the species (Table 2). The number of other types of genetic mobile elements was similar between *Xtg* and other pathovars. Additionally, *Xtg* strains displayed a higher number of putative horizontal gene transfer events, with 99 to 124 predicted putative events, compared to 49 to 62 in the rest of the species (Table 2).

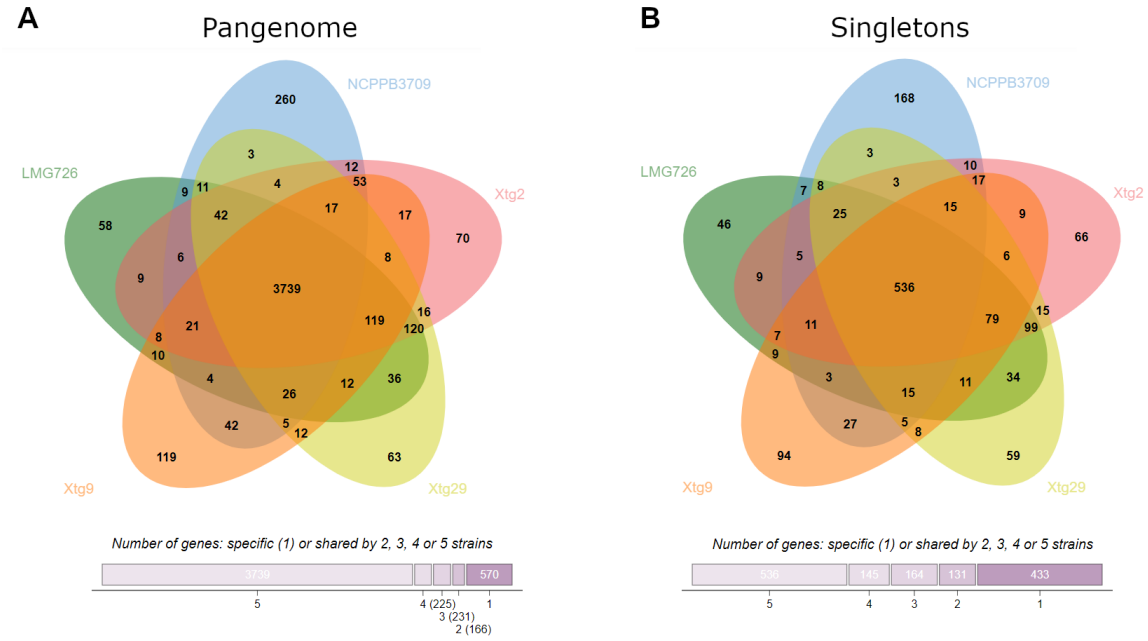


Figure 3 A) Pangenome (i.e., total gene content) of the five *Xanthomonas translucens* pv. *graminis* strains investigated, and **B)** content in predicted singletons (i.e., genes that were exclusive to *X.t.* pv. *graminis* strains). Numbers indicate the number of genes shared by specific sets of strains. Bottom bars indicate the number of genes that were by two, three, four, or five strains, or that were specific to one strain.

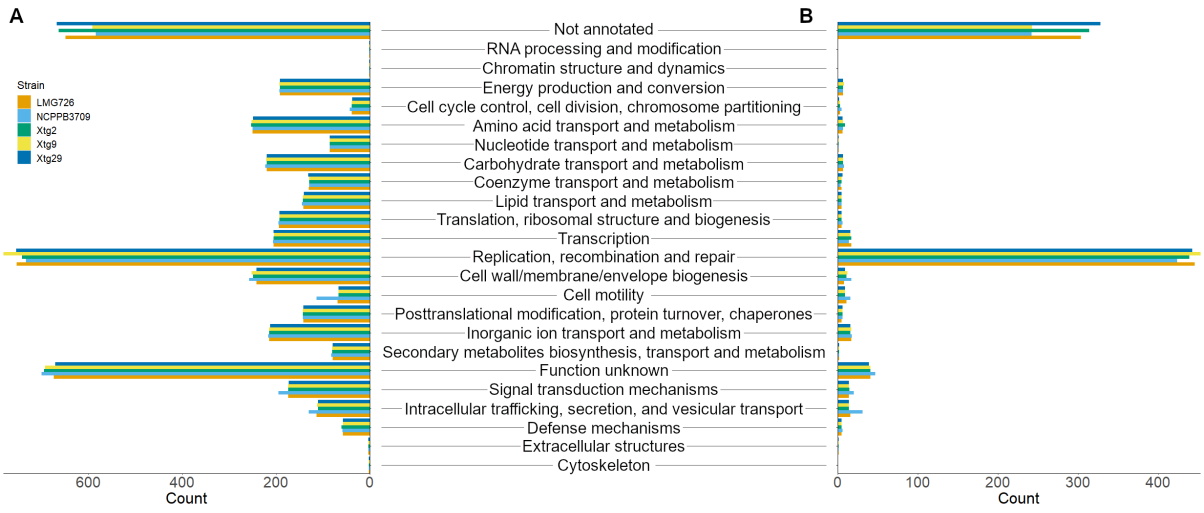


Figure 4 Clusters of orthologous groups (COG)-based functional characterization of A) the total gene content of the five *X. translucens* pv. *graminis* strains investigated, and B) predicted singletons, genes that were exclusive to *X. t.* pv. *graminis* strains.

3.4.3. Unique features of *Xtg*

Investigation of secretion systems showed that all *Xtg* strains have the *xps* T2SS and the T3SS but not the *xcs* T2SS and T4SS (Fig. 5). Furthermore, a T6SS-i4 was found in NCPPB 3709, Xtg2 and Xtg9, but no T6SS could be found in LMG 726 and Xtg29. The type III effector (T3E) repertoire of *Xtg* was comparable among the strains of the pathovar, though only one effector of the XopP family was found

Table 2 Gene content of the *Xanthomonas translucens* strains compared in this study.

Pathovar	Strain[1]	Total number of genes	Predicted singletons	Functionally annotated singletons	Functionally annotated singletons, excluding transposases	Integration/ excision elements	Replication/ recombination/ repair elements	Phage elements	Stability/transfer/ defence elements	Putative horizontal gene transfer events	TBDRs[3]
graminis	NCPPB 3709	4,445	859	647	222	413	37	16	17	108	76
	Xtg2	4,443	915	631	189	433	34	16	17	124	76
	Xtg9	4,408	856	641	184	457	34	16	17	103	76
	Xtg29	4,422	922	627	182	427	33	15	16	99	74
	LMG 726	4,415	904	627	181	428	34	15	17	111	77
arrhenatheri	**LMG 727**	3,998	NA	NA	NA	57	35	15	19	49	63
poae	**LMG 728**	4,067	NA	NA	NA	118	34	15	12	58	62
phlei	**LMG 730**	3,957	NA	NA	NA	123	33	16	18	49	64
phleipratensis	**LMG 843**	4,158	NA	NA	NA	108	35	37	17	62	67
cerealis	**CFBP 2541**	3,886	NA	NA	NA	24	35	20	22	51	51
translucens	**DSM 18974**	4,022	NA	NA	NA	103	36	18	26	57	62

[1]Pathotype strains are marked in bold.

[2]NA: Not applicable

[3]TBDRs: TonB-dependent receptors

in LMG 726 and Xtg29, compared to two in the other strains, and a XopL family effector was identified in LMG 726 only. When comparing to other pathovars, differences could be seen in the number of effectors of some families, such as AvrBs2 and XopF effectors, with only one found in *Xtg*, compared to two in most other pathovars, though only one XopF effector was found in LMG 730 and LMG 843 as well (Fig. 5). No XopG effector could be identified in *Xtg*, while all other pathovars possessed one. Moreover, aside from one gene found in LMG 726, no XopL effector was found in the other four *Xtg* strains, when other pathovars possessed one to three effectors of this family. While all other pathovars of clade Xt-III possessed exactly two XopX effectors, three could be found in *Xtg*, as was the case for pv. *translucens* (Xt-I) and pv. *cerealis* (Xt-II). Finally, three genes of the XopE family were identified in all *Xtg* strains, while only two were found in CFBP 2541, one in LMG 730, LMG 843 as well as DSM 18974, and none were found in LMG 728 and LMG 727. Overall, the *Xtg* strains displayed the smallest T3E set, ranging from 20 to 21 per strain, compared to 24 to 26 in other pathovars of clade Xt-III and 31 in pv. *translucens* and pv. *cerealis*. No TAL effector could be identified in any *Xtg* strain.

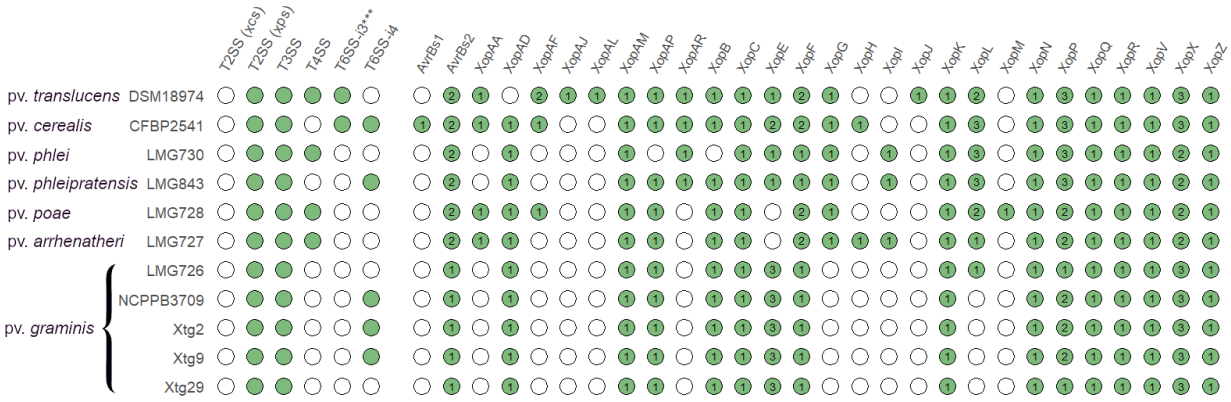


Figure 5 Presence of the *xps* and *xcs* type II secretion systems (T2SS), type III secretion system (T3SS), type IV secretion system (T4SS), subtype 3*** type VI secretion system (T6SS-i3***), and subtype 4 type VI secretion system (T6SS-i4) gene clusters, as well as putative type III effectors in the genome sequences of *Xanthomonas translucens* strains. Green circles indicate presence and white circles indicate absence. Numbers denote the number of putative effectors of each family.

Among the 856 to 922 predicted singletons observed in *Xtg* strains, 627 to 647 were functionally annotated (Table 2). After excluding transposases, 181 to 222 annotated predicted singletons could be identified, of which 148 were shared by all strains (Suppl. Table 2). Among these predicted singletons, genes that could have a role in the pathogenicity of *Xtg* were further investigated. As secreted or cell surface proteins are often involved in interaction with other organisms, we focused on genes that harbored a predicted secretion signal or a putative non-classical secretion Among these, genes, two neighboring encoding for a filamentous hemagglutinin family outer membrane proteinsprotein, a YadA family autotransporter adhesin, ten TonB-dependent receptors (TBDRs) and GumP, a protein associated with the xanthan synthesis gene cluster gum), and 11 plant cell wall degradative enzymes, including two subtilases, two esterases, a serine hydrolase, a glycerophosphodiester phosphodiesterase, a pectate lyase, a phospholipase, an alpha-galactosidase, a beta-galactosidase, and an aminopeptidase, were found. Of

the 17 identified TBDRs, five were annotated as diverse iron-siderophore receptors, five as oar-like proteins, three as vitamin B12 transporters, and one as a colicin receptor. However, while TBDRs were overrepresented in *Xtg*, ranging from 74 to 77 per strain, compared to 51 to 67 in other pathotype strains of the species (Table 1), many genes with similar annotations were also found among the TBDR genes that were conserved between pathovars, suggesting a functional redundancy of these TBDRs.

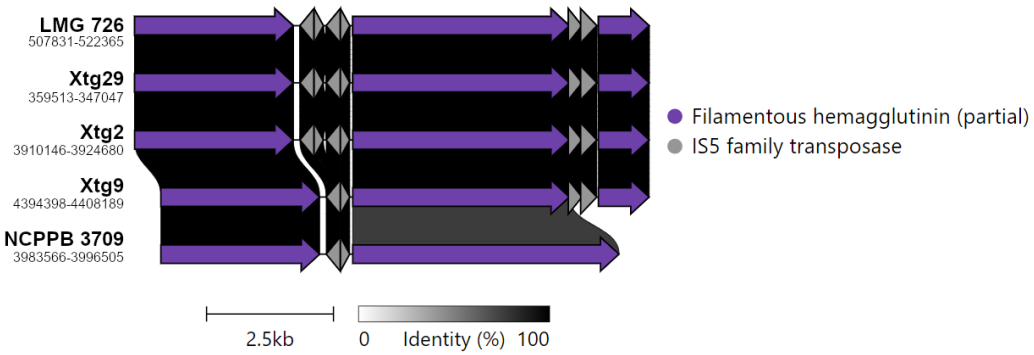


Figure 6 Pairwise comparisons of the filamentous hemagglutinin locus within *Xanthomonas translucens* pv. *graminis*. Colors denote similar genes. Numbers below strain names refer to the position of the gene cluster in this genome sequence.

Additionally, a second gene predicted as encoding a filamentous hemagglutinin gene was found neighbouring the first one, being shared by all strains except NCPPB 3709. Further investigation of the locus containing the two genes revealed the presence of transposases in between them, suggesting it may be a single gene that was truncated by the insertion of these transposable elements. A BLASTn comparison with similar partial filamentous hemagglutinin genes that were previously identified in a draft genome sequence of Xtg29 revealed they corresponded to the same locus. One of the previously identified genes corresponded to the 5'-most gene that was identified in the singleton analysis, while the second gene corresponded to an additional smaller gene downstream of the two other genes we identified, which was present in all strains except NCPPB 3709 (Figure 6). A BLASTx comparison of the hypothetical full-length gene, consisting of the combined partial sequences, revealed a high identity to other filamentous hemagglutinin genes from *Xanthomonas* spp., further confirming this locus corresponds to a single gene that was truncated.

Furthermore, a cluster containing five minor pilins of the type IV pilus (T4P) *fimT*, *pilV*, *pilW*, *pilX*, and *pilE*, as well as the anti-retraction factor *pilY1* was also found. This cluster is also present in other *X. translucens* pathovars, but the cluster found in *Xtg* is highly divergent from the one found in other pathovars, and these genes were thus predicted as singletons by the roary pipeline (Figure 7A). An in-depth analysis of the minor pilins revealed that the *pilX* gene lacked the characteristic N-terminal methylation domain and class III signal peptidase cleavage site in any *X. translucens* strains investigated, suggesting it may not be functional. In contrast, these domains were present in *pilE*, *pilV* and *fimT* for all strains. When investigating *pilW*, the gene was predicted as encoding a putative pilin only in *Xtg*, pv. *phleipratensis* strain LMG 843 and pv. *poae* strain LMG 728 (Figure 7B). However, the conserved N-

terminal phenylalanine residue was present only in LMG 843 and was replaced by a leucine in LMG 728 and all *Xtg* strains.

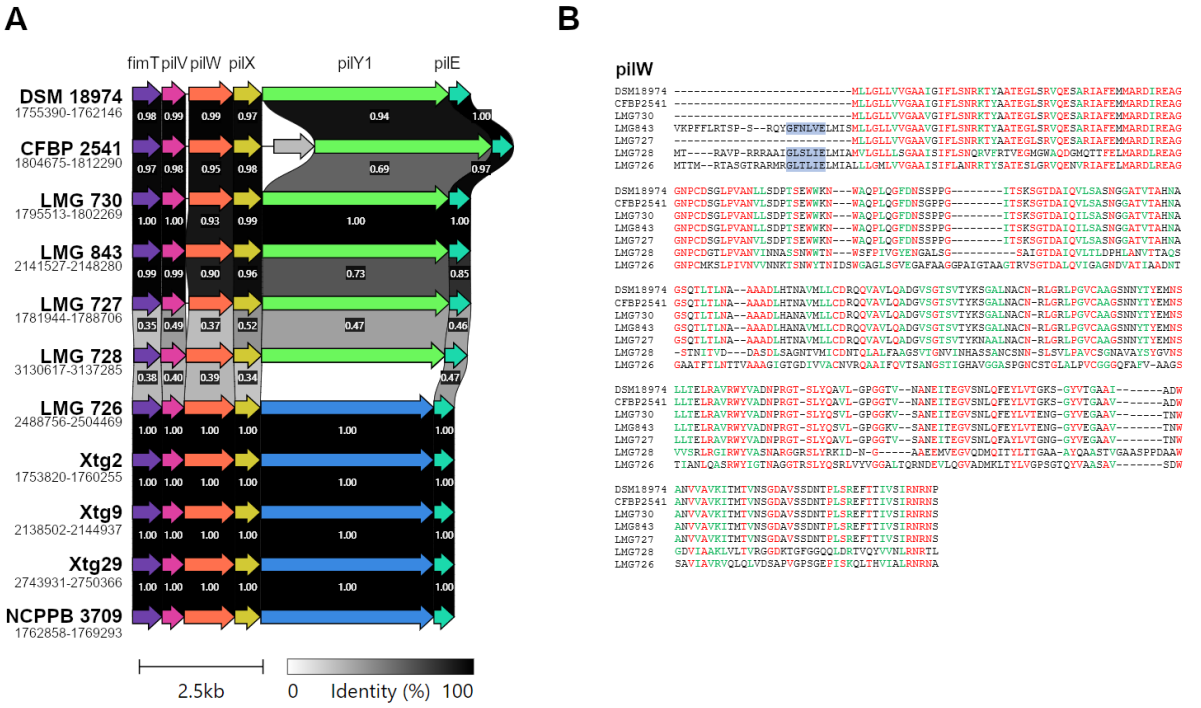


Figure 7 A) Pairwise comparisons of the minor pilins gene cluster between strains of *Xanthomonas translucens* (i.e. *X.t.* pv. *translucens* DSM 18974, pv. *cerealis* CFBP 2541, pv. *phlei* LMG 730, pv. *phlepratensis* LMG 843, pv. *arrhenatheri* LMG 727, pv. *poae* LMG 728, and pv. *graminis* LMG 726, Xtg2, Xtg9, Xtg29 and NCPPB 3709). Colors denote similar genes. The *pilY1* gene is represented in different colours in pv. *graminis* strains compared to other strains as its sequence was too divergent and they were considered as different genes by the software. Numbers below strain names refer to the position of the gene cluster in this genome sequence. **B)** Alignment of the amino acid sequences of pilW. Only the pv. *graminis* strain LMG 726 is represented as the sequences were identical in all five pv. *graminis* strains. Colours represent amino acids that were conserved among at least six out of seven strains (red) or amino acids with similar properties that were found in at least six out of seven strains (green). Highlighted in blue is the class III signal peptidase cleavage site that is characteristic to pilins.

3.5. Discussion

Comparative analysis of five *Xtg* strains revealed a very high genome plasticity within the pathovar compared to all other pathovars of *X. translucens*. While the strains showed 99.9 to 100% nucleotide identity, ANI compares conserved sequences, thus the genome sequences may still differ in lost or acquired sequences, as well as in genome structure. Indeed, the analysis of conserved genomic regions showed a high level of rearrangement between the five strains with 65 to 125 LCBs. Additionally, regions without any link to a similar block can be seen, which may be due to the presence of duplicated sequences, which are not considered by the MAUVE software.

This is the first report of such a high level of rearrangements within *Xtg*, which could be observed thanks to the generation of complete genome sequences that allowed a clear observation of the genome organization in each strain. In comparison, a similar analysis with data published by Shah et al. (Shah et al., 2021) and using complete genome sequences of four pv. *translucens* and eight pv. *undulosa* strains

with ANI values as low as 98.9 and 99.5%, respectively, showed a considerably lower amount of rearrangements, with 53 to 67 and 39 to 67 LCB (38, data not shown). This is in line with a previous comparative analysis within the *X. translucens* species that showed that clade Xt-III, which contains *Xtg*, was the clade showing the most rearrangements (Goettelmann et al., 2022). Additional genome sequences for other pathovars of clade Xt-III would help to determine whether this high level of plasticity is exclusive to *Xtg* or whether it is a characteristic of the entire clade.

The singleton analysis revealed 1,409 genes that were exclusive to the *Xtg* strains, with 536 of them being found in all five strains. Functional characterization revealed that most of these predicted singletons were annotated as transposases, and further analysis of mobile genetic elements also revealed a high amount of insertion/excision elements in the genomes of *Xtg* strains, corresponding to the identified transposases. Such a high number of transposable elements in *Xtg* could explain the high genomic plasticity that was observed, as they are genetic elements that are mobile within the genome and can facilitate gene duplications, chromosomal recombinations and horizontal transfer (Gray, 2000; Vandecraen et al., 2017). A high genome plasticity has previously been found in many *Xanthomonas* spp., and was suggested to play an important role in host adaptation (Timilsina et al., 2020). A comparably large number of insertion sequences was reported in *X. oryzae* pv. *oryzae*, where these numerous mobile elements were suggested to play a role in race differentiation (Ochiai et al., 2005). Insertion elements can also cause the loss of gene function, and virulence can emerge from the inactivation of avirulence genes (Kang et al., 2001; Inami et al., 2012). Furthermore, transposable elements have been suggested to play a role in the movement of "passenger genes" in *Xanthomonas* spp., including virulence genes, contributing to their diversification and spread (Ferreira et al., 2015; Gochez et al., 2018). As such, mobile genetic elements can play a crucial role in pathogen adaptation, and the large amount of such elements in *Xtg* compared to other pathovars of the species could, at least partially, explain its high virulence and broad host range. Investigating all the identified transposable elements, their site of insertion, and the genes flanking them would be crucial in understanding their role in pathogenicity and adaptation of *Xtg* and should be tackled in the future.

The total number of genes found in each strain of the pathovar was comparable, confirming the high number of genes already identified in the *Xtg* pathotype strain LMG 726. When compared with other pathotype strains, LMG 726 exhibited the highest number of genes of the species, as well as the highest number of genes that were exclusive to the strain (Goettelmann et al., 2022). While in this previous comparative analysis, as much as 950 genes were found to be exclusive to LMG 726, this number was influenced by the strains included in the comparison. The finding of 1,409 genes exclusive to *Xtg*, with 536 shared by all strains, using all available complete genome sequences of *X. translucens* strains is most likely a better estimation of the number of genes specific to *Xtg*. However, this unusually high amount seems to mainly be explained by the abundance of transposases in the pathovar. Nonetheless,

many predicted singletons that could play a role in the pathogenicity and host range of *Xtg* have been uncovered.

First, a cluster containing five minor pilins and the anti-retraction factor *pilY1* was found to greatly differ between *Xtg* and other pathovars of the species. These pilins are components of the T4P, which mediates surface adhesion, biofilm formation and twitching motility (Dunger et al., 2014, 2016). The T4P consists of a filamentous polymer of the major pilin PilA, and minor pilins may be incorporated into the pilus. Further investigation of minor pilins of this cluster revealed that only *pilE*, *pilV* and *fimT* may be functional in all pathovars of the species. For *pilW*, the N-terminal cleavage and methylation domain that is characteristic to pilins and required for their processing was present only in pvs. *graminis*, *poae*, and *phleipratensis*, though the conserved phenylalanine that undergoes methylation after cleavage was only present in pv. *phleipratensis* and was substituted by a leucine in the other two pathovars. However, it was previously shown that although the phenylalanine residue is highly conserved in pilins, its substitution by another amino acid did not affect secretion, cleavage and methylation of the pilin and subsequent T4P assembly (Strom and Lory, 1991). It is thus plausible that *pilW* is still functional in pvs. *graminis* and *poae*. The six identified genes, as well as the major pilin PilA were previously shown to differ significantly in Xtg29 compared to strains of other *X. translucens* pathovars, while the genes encoding the main components of the T4P machinery were highly conserved (Hersemann et al., 2017). That this divergent cluster is conserved within *Xtg* could indicate a role in the virulence of the pathovar. Indeed, the T4P is an extracellular feature of the bacteria that may be detected by plant defense mechanisms. The high sequence variation in the minor pilins, as well as the major pilin PilA may allow *Xtg* to avoid plant defenses while mediating surface adhesion and biofilm formation.

Among other predicted singleton genes, two filamentous hemagglutinin genes were identified. However, further investigation revealed that they likely correspond to a single gene in which multiple transposases were inserted. Moreover, these genes corresponded to the two partial filamentous hemagglutinin genes that were previously reported as predicted singletons of *Xtg* (Hersemann et al., 2017). The N-terminus of filamentous hemagglutinins possesses a signal peptide, as well as a two partner system domain that are involved in their secretion to the periplasm and the apoplasm, respectively (Villarino Romero et al., 2014; Guérin et al., 2017). Whether the gene encoding for the N-terminal part of the protein is still functional remains to be determined, though it is unlikely, as essential domains may be found in the C-terminal part of the gene. In *X. axonopodis* pv. *citri*, filamentous hemagglutinins were shown to be crucial to virulence by promoting plant surface adhesion and biofilm formation but acted as elicitors of plant defense in *X. campestris* pv. *vesicatoria* (Gottig et al., 2009; Choi et al., 2013). Thus, the loss or truncation of this gene in *Xtg* may help it evade recognition by the plant host.

Additionally, 17 predicted singleton genes coding for TBDRs were found, with ten being shared by all *Xtg* strains. TBDRs are known for their role in iron and vitamin B12 uptake and are mostly regulated by the ferric uptake regulator gene (Postle and Kadner, 2003). Mutants of this regulator were unable to induce disease symptoms in *X. o.* pv. *oryzae* which was hypothesized to be due to an inability to cope with the oxidative stress conditions encountered during the infection, emphasizing the role of TBDRs in virulence (Subramoni and Sonti, 2005). Moreover, in *X campestris* pv. *campestris*, some TBDRs were found to be activated by HrpG and HrpX, regulatory proteins of the type III secretion system, and the SuxA TBDR was found to play a major role in pathogenicity by acting on sucrose import (Blanvillain et al., 2007). A comparative analysis across 226 eubacterial genome sequences showed that TBDRs were overrepresented in *Xanthomonas* spp. compared to other bacteria, ranging from 36 TBDRs in *X. oryzae* pv. *oryzae* to 68 in *X. axonopodis* pv. *citri* (Blanvillain et al., 2007). In this study, similar numbers of TBDRs were found in all strains of *X. translucens*, with an overrepresentation in *Xtg*. The identified predicted singletons were mainly annotated with functions related to iron and vitamin B12 uptake, and similar genes were found to be shared across the species. It is thus unclear if these 17 predicted singletons serve a specific function in the pathogenicity of *Xtg*, or if they are functionally redundant with other TBDRs.

A total of 11 degradative enzymes were also found in singletons that were shared by all *Xtg* strains and were predicted to be secreted. Such enzymes are often secreted by the T2SS and involved in the degradation of the plant cell wall (Jha et al., 2005; Büttner and Bonas, 2010). These include cellulases, xylanases, pectate lyases, and polygalacturonases, which degrade the main polysaccharide cell wall constituents, as well as various proteases and lipases. *X. oryzae* pv. *oryzae* mutants lacking a T2SS were shown to be virulence deficient and multiple enzymes secreted by the T2SS were shown to be necessary for virulence (Ray et al., 2000; Rajeshwari et al., 2005b; Hu et al., 2007). As *Xtg* mutants lacking a T3SS showed a drastic reduction in symptoms, but still survived in the plant, it was hypothesized that other virulence factors may be important in *Xtg* (Wichmann et al., 2013). Therefore, the T2SS and the enzymes it secretes may play a crucial role in the *in planta* growth of *Xtg*.

Finally, a gene encoding for a YadA family autotransporter adhesin was predicted as a singleton shared by all *Xtg* strains. Such adhesins showing a high similarity to YadA from *Yersinia* spp. have been shown to mediate cell to cell aggregation, biofilm formation, and adhesion to the host cell surface (Feil et al., 2007; Totsika et al., 2012). In *Xanthomonas oryzae* pv. *oryzae*, mutations in the genes encoding for the YadA-like adhesins XadA and XadB were shown to result in reduced surface attachment and virulence after surface inoculation (Ray et al., 2002; Das et al., 2009). In *Xtg*, this adhesin may therefore play a similar role in attachment to the host.

An in-depth analysis of the secretion systems within *Xtg* showed that a T6SS-i4 was present in strains Xtg2, Xtg9 and NCPPB 3709 and not in LMG 726 and Xtg29, confirming previous reports investigating

these strains (Hersemann et al., 2017). It was previously hypothesized that the T4SS and T6SS could act as elicitors of plant defense, and that their absence in LMG 726 could help it evade plant defenses (Goettelmann et al., 2022). However, the presence of a T6SS in Xtg2, Xtg9 and NCPPB 3709 may disprove this hypothesis. Nonetheless, its absence in Xtg29 and LMG 726 indicates that it is not crucial to the pathogen's survival and virulence. Conversely, the *xps* T2SS and the T3SS were found to be conserved, suggesting they may be crucial to the pathovar.

The T3E repertoire within the pathovar was largely conserved, except for only one XopP family effector being present in strains Xtg2, Xtg9 and NCPPB 3709, compared to two in LMG 726 and Xtg29, as well as one XopL family effector being present only in the LMG 726 strain. This XopL effector could have been conserved in LMG 726, or it might have been recently acquired through horizontal transfer. Nonetheless, as it is unique to LMG 726, it is likely not crucial to the lifecycle of *Xtg*. Rather, the absence of XopL effectors may have been beneficial to other strains of *Xtg*, as they all lacked such effectors, while all other pathovars of the species possessed at least one and up to three. Similarly, several effectors appear to have been lost in *Xtg*, such as AvrBs2, XopF, or XopG, as the pathovar possesses fewer of these effectors than other pathovars. Furthermore, a comparable loss of a XopP effector may have occurred in strains LMG 726 and Xtg29. Interestingly, *Xtg* possesses the smallest set of T3E in the species, which may be explained by such a loss of effectors during evolution. Finally, no TAL effectors were identified, further confirming the absence of such effectors in the pathovar (Wichmann et al., 2013; Goettelmann et al., 2022). This suggests that these T3E were either unnecessary or even deleterious to the pathovar. As these effectors could trigger plant defense, their loss may have contributed to the success of *Xtg* as a pathogen of many grass species. On the other hand, three effectors of the XopE family were present in *Xtg*, while only two, one, or none were found in other pathovars. Additionally, three XopX effectors were found in *Xtg*, while other pathovars of clade Xt-III possessed only two. XopE effectors belong to the HopX family, which contains transglutaminases and possess a catalytic site that may be involved in proteolysis (Nimchuk et al., 2007). The XopE2 effector was shown to suppress hypersensitive response in *Nicotiana* spp. and to inhibit pattern-triggered immunity (Lin et al., 2011; Popov et al., 2016). XopX effectors were found to be required for virulence in *X. euvesicatoria* on tomato and pepper by interfering with plant immune responses (Metz et al., 2005; Stork et al., 2015). This suggests that XopE and XopX effectors may play a role in the pathogenicity of *Xtg* by suppressing plant defences.

Phylogeny based on ANI showed that all five available *Xtg* strains were highly identical, with 99.9 to 100% ANI. A previous analysis of the genetic diversity of *Xtg* using AFLP markers also showed that multiple strains from Switzerland, France and Belgium displayed a very low diversity (Kölliker et al., 2006). In other pathovars of the species, a higher diversity was observed by previous phylogenies, with only 99% ANI within pv *undulosa* and pv. *translucens* (Shah et al., 2021). Although these strains were from a more diverse origin than in *Xtg*, with strains from North and South America as well as Iran, there

was still a high diversity when comparing strains of a similar origin. This further highlights the particularity of the low diversity observed within *Xtg*. Still, as aside from one Norwegian strain, all *Xtg* strains compared in this study originate from Switzerland, obtaining additional complete genome sequences from strains of more diverse origins would help to better determine the genomic diversity existing within the pathovar.

Overall, many common features of *X. translucens* were lacking in *Xtg* and seem to have been lost over the course of evolution, as they were present in the rest of the species. No T4SS was found in the pathovar and a T6SS was found only in strains NCPPB 3709, Xtg2 and Xtg9. It had the smallest set of T3E and no TAL effectors were identified. Previous research also showed that although a nearly complete flagellar gene cluster is found in strain NCPPB 3709, such a cluster was absent in all other strains investigated (Hersemann et al., 2017). In this regard, the high mobile genetic element activity observed in *Xtg* could have played an important role in this loss of virulence features through transposable element insertion within a gene, or by facilitating chromosomal region deletions (Kang et al., 2001; Inami et al., 2012; Siguier et al., 2014). Furthermore, a highly divergent T4P was found, both in terms of the amino acid sequence of the pilins, as well as in the presence or absence of functional minor pilins, compared to the rest of the species. This loss or divergence of main virulence features in *Xtg* may be responsible for its ability to infect a large range of plant hosts, by allowing the bacteria to evade plant defense mechanisms.

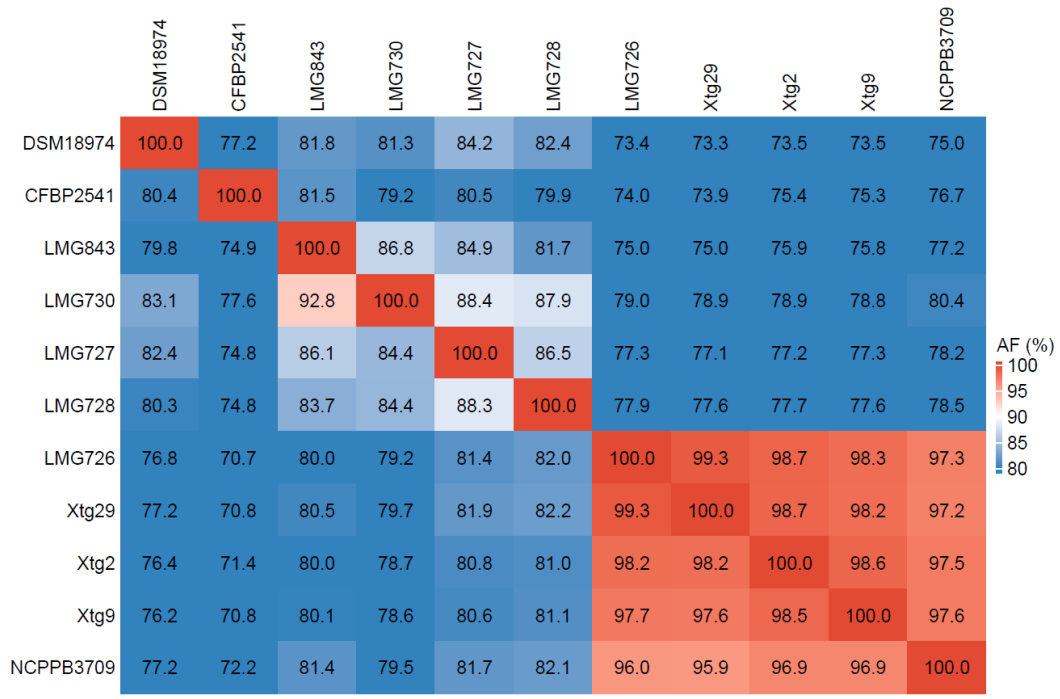
Our study provided new high-quality complete genome sequences for *X. translucens Xtg*, allowing for an in-depth comparative genome analysis within the pathovar. This revealed that *Xtg* exhibits a remarkable genome plasticity, likely related to an unusually high amount of transposases, which could explain its success as a critical pathogen of a large range of forage grasses. A set of potential virulence factors of the pathovar were identified, and a comprehensive analysis of these genes will contribute to better understand its virulence and host range, providing a basis for the development of new cultivars of forage grasses with increased resistance to the disease.

3.6. Supplementary material

Supplementary table 1 *Xanthomonas translucens* strains included in the gene content analysis and characteristics of the genome assemblies.

Clade	Pathovar	Strain	Country	Isolated from	Chromosome length (bp)	Plasmid length (bp)	Accession	Reference
Xt-I	*translucens*	B1FA	USA	*Hordeum vulgare*	4,661,643	NA	CP090000	(Schachterle et al., 2022)
	translucens	B8GF	USA	*Hordeum vulgare*	4,708,077	NA	CP089999	(Schachterle et al., 2022)
	translucens	CIX43	USA	*Hordeum vulgare*	4,664,501	36,413	CP072988-CP072989	(Heiden et al., 2022)
	translucens	CIX95	USA	*Hordeum vulgare*	4,647,206	NA	CP072990	(Heiden et al., 2022)
	translucens	XtKm7	Iran	*Hordeum vulgare*	4,577,861	NA	CP064005	(Shah et al., 2021)
	translucens	XtKm8	Iran	*Hordeum vulgare*	4,792,950	NA	CP064004	(Shah et al., 2021)
	translucens	XtKm9	Iran	*Hordeum vulgare*	4,689,955	NA	CP064003	(Shah et al., 2021)
	translucens	XtKm34	Iran	*Hordeum vulgare*	4,680,513	NA	CP064001	(Shah et al., 2021)
	hordei	UPB458	India	*Hordeum vulgare*	4,679,124	NA	CP076249	(Goettelmann et al., 2022)
	undulosa	Xtu 4699	USA	*Triticum* spp.	4,561,137	NA	CP008714	(Peng et al., 2016)
	undulosa	ICMP 11055	Iran	*Triticum* spp.	4,761,583	NA	CP009750	(Falahi Charkhabi et al., 2017)
	undulosa	LW16	USA	*Triticum* spp.	4,746,074	NA	CP043540	(Peng et al., 2016)
	undulosa	P3	USA	*Triticum* spp.	4,618,583	NA	CP043500	(Peng et al., 2016)
	undulosa	XtFa1	Iran	*Triticum* spp.	4,605,208	NA	CP063996	(Shah et al., 2021)
	undulosa	XtKm12	Iran	*Triticum* spp.	4,581,137	NA	CP064000	(Shah et al., 2021)
	undulosa	XtKm15	Iran	*Lolium perenne*	4,560,646	41,956 45,639	CP063997-CP063998-CP063999	(Shah et al., 2021)
	undulosa	XtKm33	Iran	*Hordeum vulgare*	4,626,215	NA	CP064002	(Shah et al., 2021)
	undulosa	XtLr8	Iran	*Triticum* spp.	4,563,212	45,351 40,770	CP063993-CP063994-CP063995	(Shah et al., 2021)
	undulosa	CFBP 2055	Canada	*Triticum* spp.	4,607,252	46,036	CP074361-CP074362	(Goettelmann et al., 2022)
	undulosa	MAI5034	Uruguay	*Triticum* spp.	4,625,916	NA	CP089584	(Clavijo et al., 2022)
	undulosa	UPB513	Mexico	*Triticum* spp.	4,613,657	90,481	CP096570-CP096569	Toth et al, unpublished
	secalis	CFBP 2539	Canada	*Secale cereale*	4,565,955	NA	CP074363	(Goettelmann et al., 2022)
	pistaciae group A	CFBP 8304	Australia	*Pistacia vera*	4,599,174	NA	CP074365	(Goettelmann et al., 2022)
Xt-II	*cerealis*	NXtc01	China	*Triticum* spp.	4,622,298	NA	CP038228	(Shah et al., 2019)
	pistaciae group B	ICMP 16317	Australia	*Pistacia vera*	4,386,175	NA	CP083804	(Goettelmann et al., 2022)

Supplementary table 2 Predicted singletons shared by all *Xanthomonas translucens* pv. *graminis*, excluding transposable elements. Available at: https://doi.org/10.1186/s12864-023-09855-8



Supplementary figure 1 Alignment fraction (AF) used in each comparison for the ANI calculations.

4. High-resolution bulked segregant analysis enables candidate gene identification for bacterial wilt resistance in Italian ryegrass

Goettelmann Florian[1], Chen Yutang[1], Knorst Verena[1], Yates Steven[1], Copetti Dario[1,2], Studer Bruno[1], Kölliker Roland[1*]

[1] Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland
[2] Arizona Genomics Institute, The University of Arizona, Tucson AZ, USA

Research article: In preparation

Keywords: Bulked segregant analysis (BSA), disease resistance, bacterial wilt, *Xanthomonas translucens* pv. *graminis*, Italian ryegrass, *Lolium multiflorum* Lam.

4.1. Abstract

Bacterial wilt, caused by *Xanthomonas translucens* pv. *graminis* (*Xtg*), is a serious disease of economically important forage grasses, including Italian ryegrass (*Lolium multiflorum* Lam.). A major QTL for resistance to *Xtg* was previously identified, but the precise location as well as the genetic factors underlying the resistance are yet to be determined. To this end, we applied a bulked segregant analysis (BSA) approach, using whole-genome deep sequencing of pools of the most resistant and most susceptible individuals of a large (n = 7,484) biparental F_2 population segregating for resistance to *Xtg*. Using chromosome-level genome sequences as references, we were able to define a ~300 kb region highly associated to resistance on chromosome 4. Further investigation of this region revealed multiple genes with a known role in disease resistance, including genes encoding for Pik2-like disease resistance proteins, cysteine-rich kinases, and RGA4- and RGA5-like disease resistance proteins. Investigation of allele frequencies in the pools and comparative genome analysis in the grandparents of the F_2 population revealed that some of these genes contain variants with allele frequencies that correspond to the expected heterozygosity in the resistant grandparent. This study emphasizes the efficacy combining BSA studies in very large populations with whole genome deep sequencing and high-quality genome sequences to pinpoint regions associated with a binary trait of interest and accurately define a small set of candidate genes. Furthermore, markers identified in this region hold significant potential for marker-assisted breeding strategies to breed resistance to *Xtg* in *L. multiflorum* cultivars more efficiently.

4.2. Introduction

Italian ryegrass (*Lolium multiflorum* Lam.) is a widely cultivated forage species, known for its high yield and nutritional value as well as its fast ground cover (Frame, 1991; Humphreys et al., 2010). As such, it provides an important feed source as hay or silage for ruminant livestock production in temperate climates. However, it can be affected by diseases, the most important being bacterial wilt caused by *Xanthomonas translucens* pv. *graminis* (*Xtg*, Egli et al., 1975). The bacteria mainly penetrate through wounds of the plant and spread in the xylem vessels, disrupting the flow of water and nutrients (Rudolph, 1993). This results in wilting and yellowing of the leaves, stunted growth and eventually the death of the plant, causing serious yield losses. Such losses are reported to reach up to 40 % under unfavorable conditions (Schmidt and Nuesch, 1980; Gondran and Betin, 1989). Moreover, the global impact of the disease may become more severe due to climate change, which could increase the occurrence of conditions that are favorable to the pathogen, expand its geographical range, or impact host-pathogen interactions (Hunjan and Lore, 2020).

Currently, the only efficient, durable, and economically feasible way to manage the disease is to breed for resistant cultivars. So far, breeding efforts to develop *Xtg* resistant cultivars have been based on recurrent phenotypic selection of resistant individuals following infection with *Xtg*. This has led to cultivars with an increased resistance to the disease (Suter et al., 2021). However, due to the allogamous nature of ryegrasses, these cultivars are genetically heterogeneous, and classical breeding approaches based on phenotypic observation fail to fix dominant alleles, as they can mask susceptibility alleles which are thus maintained in the population. For example, in recurrent selection of *Festuca pratensis*, some level of resistance to *Xtg* was achieved, but quickly stagnated, and complete resistance could not be attained (Boller et al., 2001; Michel, 2001). In this objective, genomics-assisted breeding would greatly aid in identifying and incorporating resistance alleles into elite *L. multiflorum* cultivars by targeting specific genetic markers associated with resistance, making the breeding process more efficient. However, such methods require a thorough understanding of the disease resistance and its genetic control.

With the rapid progress in high throughput sequencing technologies in recent years, determining which genetic loci govern a specific phenotype has become increasingly effective and precise. However, such studies, like genome-wide association studies, commonly rely on obtaining both the genotype and phenotype of a large sample of individuals, ranging from hundreds to thousands, which often results in very high costs and efforts (Uffelmann et al., 2021). In order to reduce costs, bulked segregant analysis (BSA) can be used to identify molecular markers associated with a phenotype of interest in a biparental population segregating for this phenotype using pools of the most extreme phenotypes in this population. The method was first proposed in 1991 and has

since undergone significant development in particular thanks to the rise of next generation sequencing (NGS), which allows for high marker densities at a low cost, and led to the development of different downstream statistical analysis methods (Michelmore et al., 1991; Li and Xu, 2022).

To statistically determine genetic loci linked to a phenotype of interest, BSA studies generally rely on the ΔSNP index or the G statistics (Magwene et al., 2011; Takagi et al., 2013). However, sequencing data introduces noise associated with many factors, including sequencing errors, varying sequencing coverage or uneven marker density. In order to reduce this noise, many downstream analyses have focused on developing smoothing methods based on sliding windows. The tricube ΔSNP (t-ΔSNP) index and the G' statistics, smoothed versions of the ΔSNP index and G statistics, respectively, are based on a weighted average computed by the Nadaraya-Watson tricube smoothing kernel across a sliding window of a set base pair length (Nadaraya, 1964; Watson, 1964). Furthermore, Euclidean distance (ED)-based methods, as well as their smoothed version, ED^4, have proven to be very powerful at detecting QTL, even outperforming the previous methods (de la Fuente Cantó and Vigouroux, 2022). The implementation of these statistical methods in BSA studies allows to reliably and precisely identify genetic loci associated with a phenotype of interest.

To identify sources of resistance to *Xtg* in *L. multiflorum*, a first study used a segregating F_1 population (hereafter referred to as Xtg-ART) resulting from a cross between an individual of the susceptible cultivar 'Adret' and M2289, a resistant individual from advanced breeding germplasm (Studer et al., 2006). This study was the first to identify a major QTL explaining 43 to 84% of phenotypic variation for *Xtg* resistance on linkage group (LG) 4. However, further characterization of the QTL was hindered by the predominantly anonymous markers and the lack of genome sequence information for *L. multiflorum* at the time. Later, the assembly of a draft genome sequence of M2289 allowed to apply a BSA approach using shotgun sequencing of pools of the 44 most resistant and the 44 most susceptible individuals within the Xtg-ART population (Knorst et al., 2018a, 2018b). This study was able to confirm the presence of the QTL on LG 4 and highlighted the efficacy of BSA in fine-mapping the QTL for resistance. However, the small number of individuals available, together with the low depth of the sequencing output and the fragmented nature of the M2289 draft genome sequence did not allow for further fine-mapping of the QTL region and determination of the underlying genes. Recently the high-quality chromosome-level genome assembly of the cultivar 'Rabiosa' for *L. multiflorum* was made available, providing a strong basis for further genomics-based analyses in *L. multiflorum* (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030979885.1/). Furthermore, the high-throughput sequencing platforms currently available allow for a high coverage of the whole

genome sequence, resulting in high marker densities and more accurate allele frequency estimations, improving the effectiveness and accuracy of BSA studies.

The goal of this study was to define candidate genes for resistance to *Xtg* by applying a BSA approach based on high coverage whole-genome sequencing of pools of susceptible and resistant individuals in an F_2 population derived from the Xtg-ART population, hereafter referred to as the Xtg-ART-F_2 population. We aimed at fine-mapping the major QTL on LG 4, identifying candidate resistance genes, and defining markers associated with resistance to be used in marker-assisted breeding.

4.3. Results

4.3.1. Resistance to bacterial wilt is segregating in the Xtg-ART-F_2 population

The disease progress in the 7,484 Xtg-ART-F_2 individuals in the greenhouse followed the expectations, with first symptoms being visible at 14 days post infection (dpi) and becoming more severe over time (Fig. 1). Since the natural death of older leaves was challenging to distinguish from *Xtg* symptoms, most healthy plants were assigned a score of two at 21 and 28 dpi. Cutting back the plants after the third scoring allowed to make this distinction between natural leaf senescence and *Xtg* symptoms and at 49 dpi 1,165 (15.5%) plants showed no symptoms, while

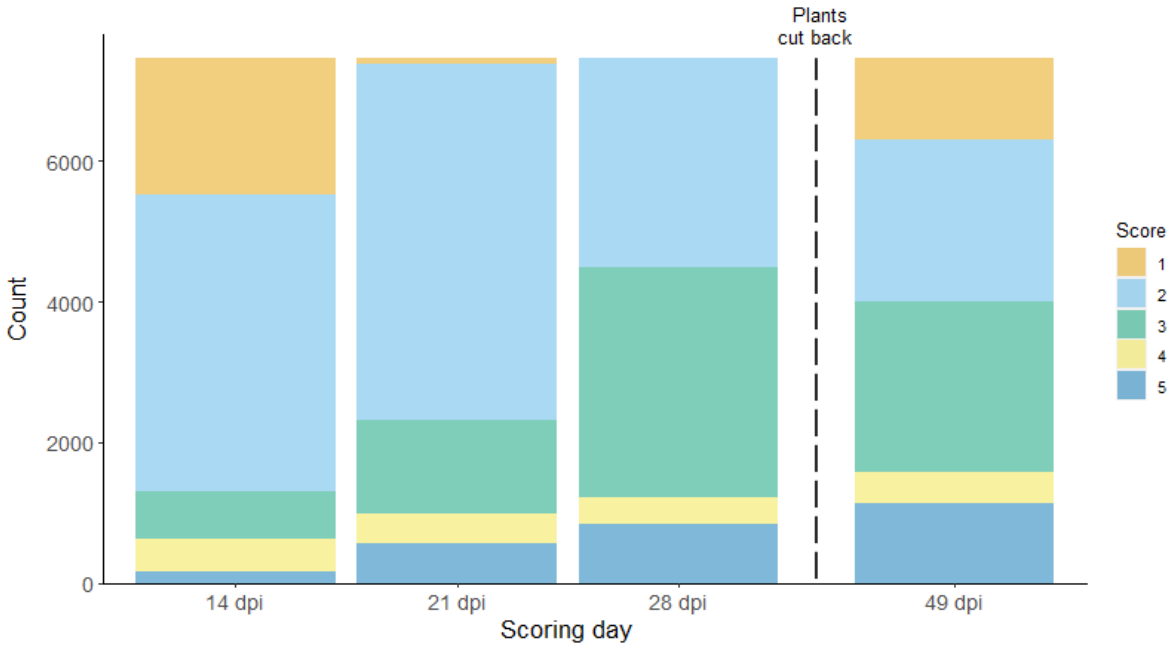


Figure 1 Distribution of the bacterial wilt disease scores in 7,484 individuals of the *L. multiflorum* Xtg-ART-F_2 population at 14, 21 and 28 days post inoculation with *Xtg* (dpi), as well as at 49 dpi, three weeks after the plants were cut back and allowed to regrow. A score of 1 corresponds to a healthy plant showing no symptoms (orange), 2 corresponds to mild symptoms (light blue), 3 corresponds to intermediate symptoms (green), 4 corresponds to severe symptoms (yellow), and 5 corresponds to a dead plant (dark blue).

1,140 (15.2%) had died. From these two sets of plants, the most resistant and the most susceptible individuals were selected, resulting in a total of 750 resistant plants and 761 susceptible plants that were used to form two pools that were sequenced.

4.3.2. Obtaining chromosome-level genome sequences of the Xtg-ART-F_2 grandparents

In order to obtain a chromosome-level genome sequence that is representative of the resistant individuals of the Xtg-ART-F_2 population, the draft genome assembly of the resistant grandparent M2289 (Knorst et al., 2018b), which consists of 129,579 scaffolds, was anchored to the Rabiosa genome sequence (Chen et al., unpublished). A total of 78,219 scaffolds were successfully anchored to the seven pseudochromosomes of the Rabiosa assembly, and 3,912 scaffolds from the M2289 assembly were anchored to 829 larger scaffolds that are not mapped to pseudochromosomes in the Rabiosa assembly. In addition, 47,448 scaffolds from the M2289 assembly could not be anchored and were excluded from further analysis. This chromosome-level genome sequence of M2289 will hereafter be referred to as "M2289v2".

To compare M2289v2 with the genome sequence of the susceptible grandparent of the Xtg-ART-F_2 population, a similar approach was applied to the Adret draft genome assembly, which consists of 117,277 scaffolds, to produce the "Adretv2" genome sequence (Knorst et al., 2018a). A total of 72,888 scaffolds were anchored to pseudochromosomes, 3,596 scaffolds were anchored to 795 larger scaffolds from the Rabiosa assembly, and 40,793 scaffolds were left untouched.

4.3.3. A 300 kbp region of chromosome 4 is highly associated with resistance

A total of 606,760 Gb of DNA sequence information was obtained by whole genome sequencing of the two pools (average = 303,378 Gb per pool), corresponding to an average of 120X coverage of the haploid genome sequence per pool, considering an estimated genome sequence size of 2.5 Gb (Copetti et al., 2021, Chen et al., unpublished). After trimming and filtering reads, a total of 490,386 Gb remained (average = 245,193 Gb per pool), corresponding to an average of 98.1 x coverage of the haploid genome sequence per pool. The mapping of the reads to the M2289v2 genome sequence resulted in an average alignment rate of 34.13 %. The variant calling yielded a total of 7,965,619 biallelic variants, including 7,068,640 SNPs and 896,979 INDELs. After filtering for total depth and quality, a total of 4,239,024 variants remained for the subsequent analysis. Additionally, in order to have a comparison with a better quality and more complete genome sequence, the Rabiosa genome sequence was also used as a reference, which resulted in an average of 87.77 % alignment rate, yielding a total of 18,765,678 biallelic variants, including

15,170,193 SNPs and 3,595,485 INDELs. After filtering, 7,525,443 variants remained for the subsequent analysis.

The association of each variant to resistance was determined with the |ΔSNP| index, the G value and the ED value showed a clear association to resistance on pseudochromosome 4 with all three approaches, and with both reference genome sequences, with a high peak spanning most of the pseudochromosome (Figure 2). However, a high level of noise can be seen, with fluctuations due to sequencing errors or the stochastic nature of sequencing, which result in different alleles at polymorphic loci being sequenced at varying depths. This results in some regions showing a high association to resistance, however, these did not correspond to a large peak shape expected in a biparental population. Furthermore, these fluctuations hindered the precise identification of a region where association to resistance was the highest on pseudochromosome 4.

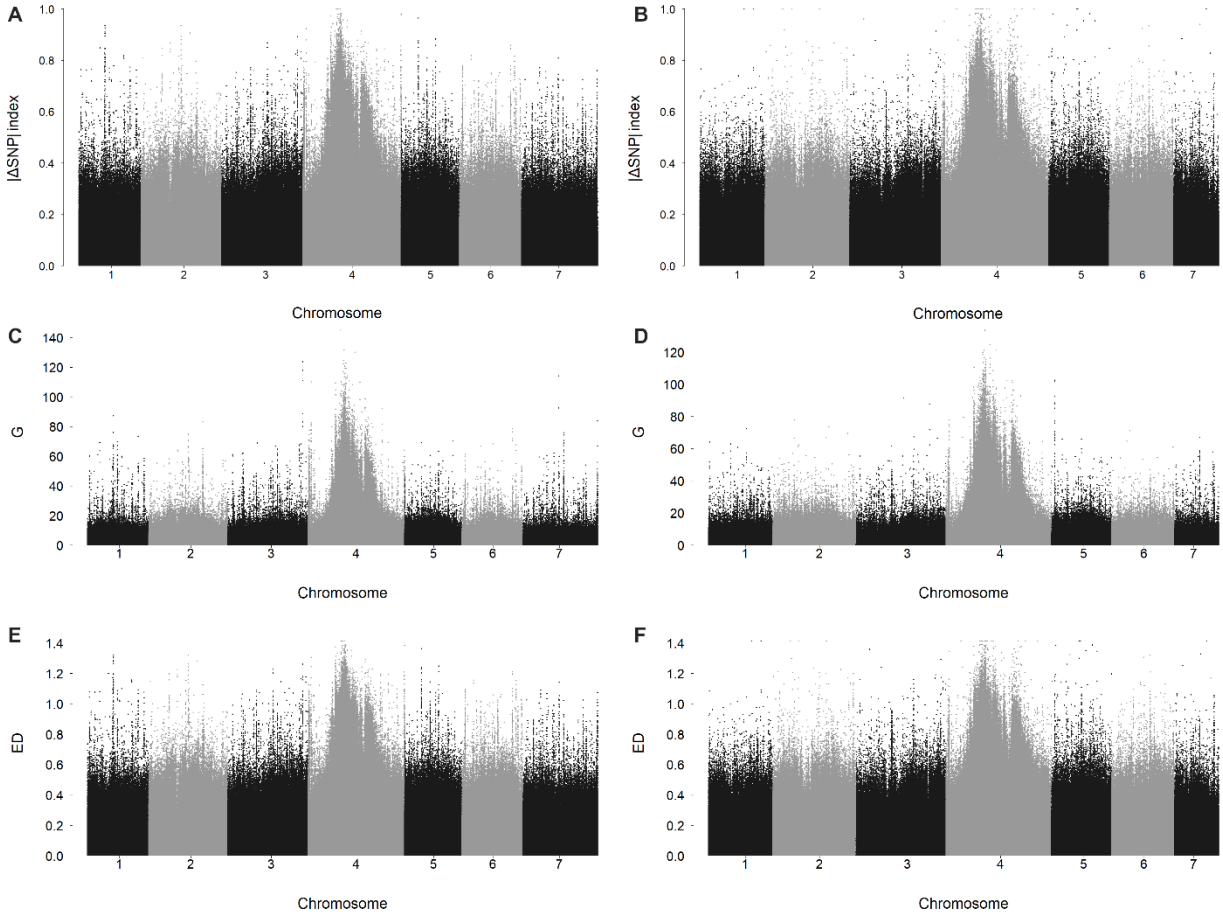


Figure 2 Association of each variant to *Xanthomonas translucens* pv. *graminis* resistance across all seven *Lolium multiflorum* chromosomes calculated with the |ΔSNP| index (A, B), the G value (C, D) or the Euclidean Distance (ED) (E, F) with M2289v2 (A, C, E) or Rabiosa (B, D, F) as a reference genome sequence.

To increase the signal to noise ratio as well as to refine the peak region, smoothing methods based on sliding windows were applied to obtain the t-|ΔSNP| index, the G' value and the ED^4 value. These methods effectively reduced noise and allowed to distinguish a clear peak on chromosome

4, which exceeded the 99% confidence interval for the t-|ΔSNP| index and the 99% false discovery rate threshold for the G' value (Figure 3). With all approaches, the strongest association to resistance was found at a similar location on chromosome 4, and no association to resistance was found on any other chromosome. With the ED^4 approach, a sharp peak could be seen, spanning approximately from 28,800,000 bp to 28,900,000 bp in the M2289v2 genome sequence, and this region was selected to be further investigated (Figure 3 E, F). In Rabiosa, similar results were observed, and a sharp peak was observed from approximately 163,280,000 bp to 163,400,000 bp with the ED^4 approach (Suppl. Figure 1).

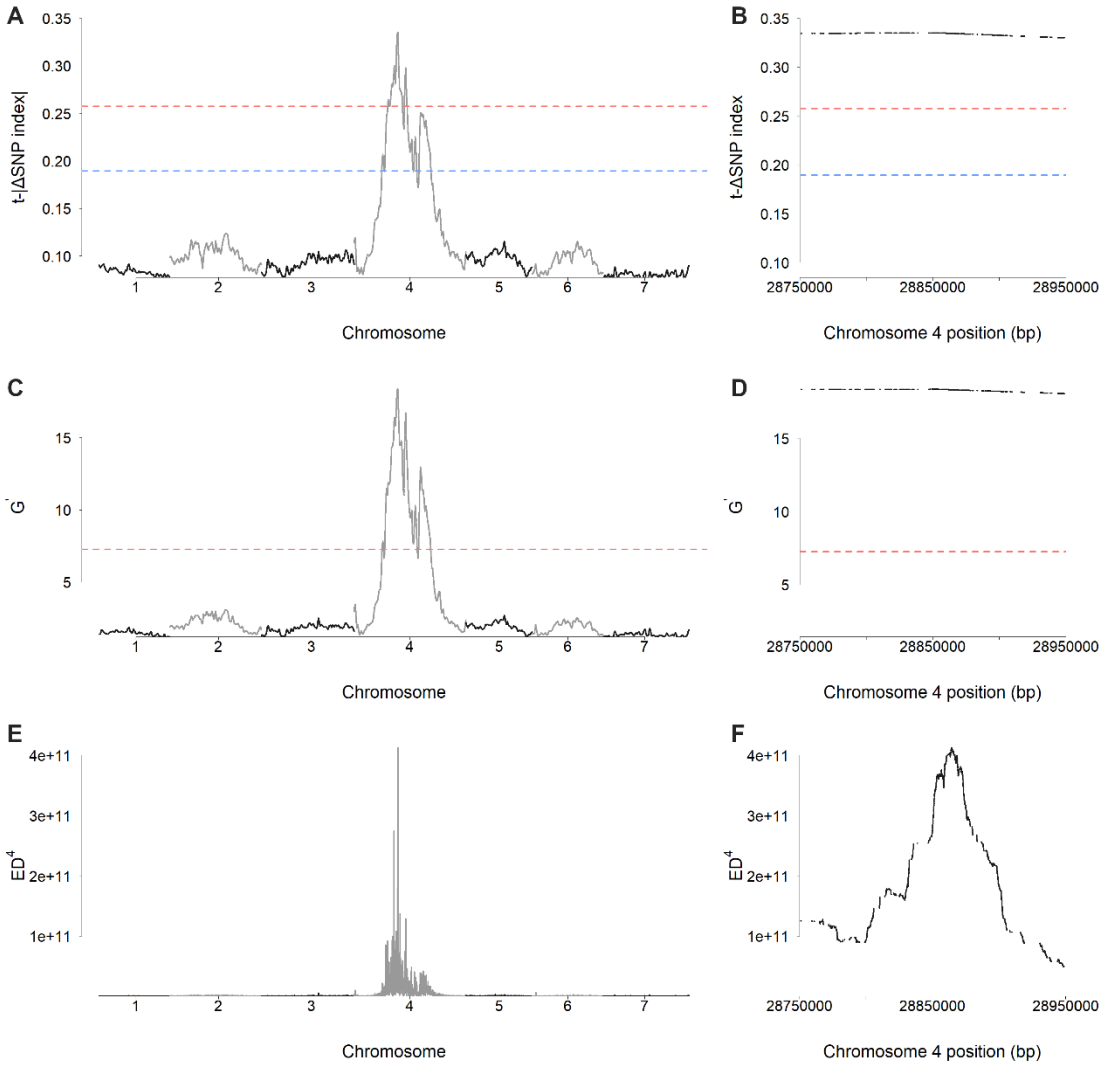


Figure 3 Association of each variant to *Xanthomonas translucens* pv. *graminis* resistance in *Lolium multiflorum* calculated with the t-|ΔSNP| index (A, B), the G' value (C, D) and the ED^4 value (E, F), with M2289v2 as a reference genome sequence, on all chromosomes (A, C, E) or the region with the highest association to resistance on chromosome 4 (B, D, F). Dotted blue lines represent the 95% confidence interval for the t-|ΔSNP| index, and dotted red lines represent the 99% confidence interval for the t-|ΔSNP| index or the false discovery rate threshold of 1% for the G' value.

A BLASTn-based comparison of the annotated gene sequences within these two regions was able to determine that the region identified in Rabiosa corresponded to a portion of the region identified in M2289v2. For a better comparability of the two regions, the larger region in Rabiosa, corresponding to the M2289v2 region, was defined, spanning from 163,094,000 bp to 163,410,000 bp and selected for further analysis. These regions were investigated in more detail and revealed a cluster of genes that could potentially be linked to disease resistance.

4.3.4. Multiple candidate resistance genes were found in the candidate region

A total of twelve and nine annotated genes found in the candidate regions of M2289v2 and Rabiosa, respectively, showed similarity to genes involved in disease resistance, and were defined as candidate genes for resistance to *Xtg* (Table 1). These included genes showing similarity to Pik-2-like and resistance gene analog (RGA)4-like disease resistance proteins, as well as two cysteine-rich kinases (CRK) that were found in both genome sequences. The candidate region in the M2289v2 assembly also contained two RGA5-like disease resistance proteins, while in the Rabiosa assembly, four serine protease inhibitor (SERPIN) genes were found. Additionally, in Rabiosa, one gene showed homology to a retrotransposon, and one to a histone deacetylase, which could not be linked to disease resistance. Furthermore, eight and four genes, in M2289v2 and Rabiosa respectively, could not be annotated as no match to any known gene was found. Moreover, in M2289v2, the region contained 17 gaps and spanned ~100 kb large, whereas the corresponding region in the Rabiosa genome sequence spanned ~300 kb. This indicates that additional genes could be present in this region but cannot be found in the M2289v2 assembly.

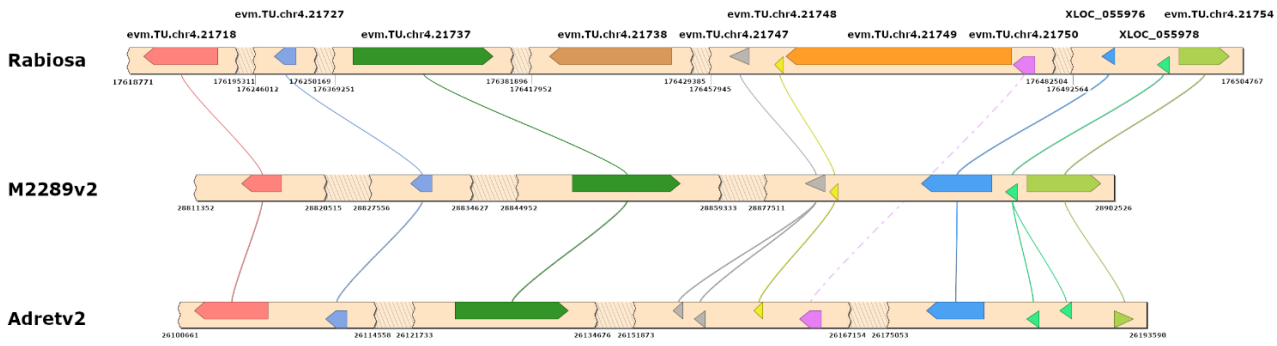


Figure 4 Comparison of the candidate genes composition within the candidate region across the genome sequences of Rabiosa, the resistant grandparent M2289v2, and the susceptible grandparent Adretv2. Genes are represented as bars, with the arrowhead representing the 3' end of the gene. Synteny between genome sequences is displayed as lines connecting homologous genes. Regions with no genes are shown as hatched breaks for improved clarity.

The Pik-2-like disease resistance protein gene evm.TU.chr4.21718 from Rabiosa was homologous to the region in M2289v2 where seven Pik-2-like genes were found, indicating that they may correspond to a single gene, homologous to the one found in Rabiosa (Figure 4). Additionally, regions displaying homology to the two genes that showed similarity to RGA5-like disease resistance proteins in M2289v2 were found in Rabiosa, although they were not predicted as genes in the Rabiosa annotation (Figure 4). Overall, the comparison of these candidate genes between M2289v2 and Rabiosa, as well as with the updated genome sequence of the susceptible grandparent, Adretv2, showed that most genes are conserved between the three genome sequences. Still, no homology to the evm.TU.chr4.21738 and evm.TU.chr4.21749 genes from Rabiosa was found in M2289v2 nor Adretv2, and homology to evm.TU.chr4.21750 was found only in Adretv2.

Table 1 Genes found within the candidate region, defined from 28,800,000 bp to 28,900,000 bp on pseudochromosome 4 of the M2289v2 genome sequence, and from 163,090,000 bp to 163,410,000 bp on pseudochromosome 4 of the Rabiosa genome sequence, that were annotated as a potential disease resistance gene.

Annotation	M2289v2 gene ID	Start (bp)	End (bp)	Rabiosa gene ID	Start (bp)	End (bp)
Pik-2-like	XLOC_040221	28,810,090	28,810,476	evm.TU.chr4.21718	163,094,711	163,100,016
	XLOC_040220	28,811,063	28,811,263			
	XLOC_040219	28,811,461	28,811,664			
	XLOC_040218	28,812,525	28,812,767			
	XLOC_000927	28,814,543	28,815,826			
	XLOC_000926	28,816,808	28,816,882			
	XLOC_021628	28,818,511	28,819,200			
CRK6	XLOC_017025	28,830,345	28,831,876	evm.TU.chr4.21727	163,153,306	163,154,874
CRK8	XLOC_023208	28,853,351	28,854,622	evm.TU.chr4.21737	163,276,545	163,286,601
				evm.TU.chr4.21738	163,325,147	163,333,991
SERPIN				evm.TU.chr4.21747	163,365,239	163,366,661
				evm.TU.chr4.21748	163,368,491	163,369,111
				evm.TU.chr4.21749	163,369,308	163,385,542
				evm.TU.chr4.21750	163,385,641	163,387,209
RGA5-like	XLOC_055978	28,892,028	28,892,210			
	XLOC_055976	28,895,207	28,895,509			
RGA4-like	XLOC_055977	28,899,358	28,901,523	evm.TU.chr4.21754	163,406,159	163,409,767

A total of 3,213 variants were found in the candidate region using M2289v2 as a reference, and 3,177 with Rabiosa. Of these, 880 and 1,010, respectively, had a variant that was absent or present in very low frequency in the susceptible pool and the Adret susceptible grandparent, corresponding to an absence or low frequency expected of a dominant resistance allele in a susceptible individual or population. Using M2289v2 as a reference, 101 of these variants were found in coding sequences of 11 genes, and 54 were found to induce a nonsynonymous amino acid change, in a total of eight genes (Figure 5A). The allele frequencies in XLOC_000927, XLOC_021628 and XLOC_017025 indicate that these three genes are heterozygous at these loci in the M2289 grandparent. The remaining five genes, however, seem to be homozygous in

M2289. With the Rabiosa genome sequence as a reference, 74 variants with a resistance allele were found in a coding sequence, in a total of six genes, and 47 would result in a nonsynonymous amino acid change, found in the same six genes (Figure 5B). Most of these variants were homozygous in the M2289 resistant grandparent, with only eight being heterozygous in evm.TU.chr421718, evm.TU.chr421747 and evm.TU.chr421750.

Figure 5 Frequencies of resistance alleles for variants causing amino acid changes in annotated genes within the candidate region. Frequencies are represented within the resistant and susceptible pools, as well as within the M2289 and Adret grandparents, with both the M2289v2 (A) and the Rabiosa (B) genome sequences. Gray negative bars denote an absence of reads for this variant.

4.4. Discussion

We were able to define a ~300 kbp genome region that was highly associated with resistance to bacterial wilt in *L. multiflorum*. This region contained a small number of candidate resistance genes with known roles in disease resistance. Consistent results were found with three different statistics for BSA data analysis, supporting the robustness of the applied methodology. It also demonstrates the efficacy of BSA based on whole genome sequencing of pools of contrasting phenotypes to identify candidate genes associated with a binary trait of interest in a *L. multiflorum* biparental F_2 population.

While previous work based on a first draft of the M2289 genome sequence was able to identify genomic regions associated with resistance mapping to chromosome 4, the very fragmented

nature of this genome assembly, as well as the lack of reliable information on the position of each scaffold on chromosomes, did not allow for a precise location of the resistance in the genome (Knorst et al., 2018a; Goettelmann et al., 2021). The recent development of a high-quality chromosome-level genome sequence of Rabiosa allowed to produce the updated chromosome-level genome sequence M2289v2 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030979885.1/). Though this genome sequence is still fragmented, it allowed to reliably anchor scaffolds on a pseudochromosome, which enabled a clear visualization of the association to bacterial wilt resistance across the genome and allowed to narrow down the location of the QTL on chromosome 4. Furthermore, the sliding window-based statistical approaches used in this study required highly contiguous genome sequences and could not have been applied with the M2289 draft genome sequence, in which the largest scaffold is 43,903 bp, much smaller than the window size of 10^6 we used (Knorst et al., 2018b).

The high population size coupled with a whole genome sequencing approach at high coverage was able to optimize most of the factors on which the success of BSA studies rely, such as the population and pool sizes, sequencing depth, and marker density (Magwene et al., 2011; de la Fuente Cantó and Vigouroux, 2022; Huang et al., 2022). In BSA studies that have been performed so far, the sizes of the investigated populations were generally small, often around 200 individuals, though they can range from <100 to >10,000 individuals (reviewed in Huang et al., 2022). Pool sizes were also variable, ranging from only five individuals per pool to more than 400. In our study, the investigation of a total of 7,484 F2 individuals, and the selection of 750 and 761 individuals for the resistant and susceptible pool, respectively, is thus relatively high and allowed for both a high detection power and mapping resolution. Another crucial factor for BSA studies is the number of recombination events. This is of particular importance as *Lolium* species have been shown to exhibit a low recombination rate, in particular in chromosome 4 (Kopecký et al., 2010; King et al., 2013). Nonetheless, the use of an F_2 population and the high number of individuals per pool allowed to effectively address this factor.

Using both the M2289v2 and the Rabiosa genome sequences as reference showed consistent results, with the highest association to resistance found in a similar location in both genome sequences. The genes found within the two regions were comparable, both in terms of presence of the genes, as well as their order. Still, some genes found in Rabiosa were not found in M2289v2. While they may be absent from the M2289 genome, this may also be explained by the fragmented nature of this genome sequence, with many gaps present in the candidate region. These genes may either be missing from the assembly or may not have been anchored to the proper locus when producing the M2289v2 genome sequence. Furthermore, the Rabiosa genotype shows some susceptibility to *Xtg* and may not contain the resistance source (data not shown). Therefore, the anchoring of the M2289 genome sequence to Rabiosa may not allow to properly access the

resistance source in M2289. In human genetics, where one genome sequence is universally used as a reference, it has been shown that many population-specific DNA sequences were missing from this reference, preventing the identification of genes of interest in these populations (Sherman et al., 2019). This highlights the need for high-quality, complete genome assemblies that are related to the investigated population in association studies. In this regard, obtaining the complete DNA sequence of resistant individuals at the candidate region would be crucial to accurately determine genes that could be responsible for resistance to *Xtg*.

Nonetheless, the present study was able to identify multiple genes that are known to be involved in disease resistance. The most interesting candidates within the genome region associated with *Xtg* resistance include XLOC_000927, XLOC_021628 and evm.TU.chr4.21718, annotated as Pik-2-like disease resistance proteins. Pik proteins are nucleotide-binding leucine-rich repeat (NLR) proteins that were shown to be involved in resistance to many strains of rice blast (*Magnaporthe grisea*), recognizing the AVR-Pik proteins from the pathogen and triggering a specific defense response (Ashikawa et al., 2012; Wang et al., 2017a). However, the evm.TU.chr4.21718 gene from Rabiosa showed homology to the M2289v2 region containing seven genes annotated as Pik-2-like disease resistance proteins, suggesting these seven genes may correspond to a single gene in M2289v2. The prediction of multiple smaller genes at this locus in M2289v2 is likely due to the fragmented nature of the genome sequence, which is supported by the fact that this gene region is found on three separate scaffolds in the original draft assembly, from which the annotation was produced. Another candidate is XLOC_017025, annotated as a cysteine-rich receptor-like kinase (CRK) 6. These proteins are transmembrane proteins involved in the detection of biotic and abiotic stresses, leading to programmed cell death (Chen et al., 2004). In *A. thaliana*, the overexpression of *CRK6* and other *CRK* genes led to an enhanced activation of pathogen-associated molecular pattern-triggered immunity and some CRK have been shown to mediate pattern-triggered immunity (Ederli et al., 2011; Yeh et al., 2015). Finally, evm.TU.chr4.21747 and evm.TU.chr4.21750, two candidate genes of the serine protease inhibitor (SERPIN) superfamily, were found. SERPINs act on many cell processes, and some members of the family have been shown to have a role in disease resistance in *A. thaliana* and barley by inhibiting proteinases from the pathogen (Pekkarinen et al., 2007; Laluk and Mengiste, 2011; Quilis et al., 2014). Moreover, while the allele frequencies of variants found in these genes were not in line with the expected heterozygous frequency in the M2289 parent, the candidate region also contained genes encoding for CRK8, a member of the CRK family like CRK6. Additionally, two other SERPINs in Rabiosa, that could be involved in resistance were also found. Lastly, the candidate region also contained genes of the RGA4- and RGA5-like protein families. These NLR proteins act together in mediating disease resistance to *M. grisea* in rice (César et al., 2014).

The identified candidate region provides valuable insight into the potential loci responsible for resistance to *Xtg*. Interestingly, this region contains mostly genes with known functions in disease resistance. As such genes are often found in clusters containing other resistance genes as well as paralogs resulting from the duplication and divergence of some genes, it is plausible that multiple genes within this region could play a role in resistance to *Xtg* (Michelmore and Meyers, 1998; Sekhwal et al., 2015). This emphasizes the need for a more comprehensive overview of the genes present in the candidate region in resistant individuals in order to better understand the mechanisms involved in resistance to *Xtg*.

Ultimately, this study allowed to identify genetic variants that were highly associated with resistance to *Xtg* and could be used as diagnostic markers for marker-assisted selection. In wheat, markers associated with the *Fhb1* QTL for resistance to *Fusarium* head blight are routinely used in breeding programs, significantly improving resistance to the disease in new cultivars (Anderson et al., 2007). In rice, marker-assisted selection allowed the introgression of the *Xa38*, *Xa21* and *xa13* resistance genes against the bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae*, as well as the *Pi2* and *Pi54* resistance genes against rice blast caused by *M. oryzae*, into elite rice cultivars (Swathi et al., 2019; Yugander et al., 2019; Sagar et al., 2020). Although further validation is necessary to confirm their effectiveness, the identified variants hold significant potential as molecular markers to be used in marker-assisted breeding applications, laying the foundations for an efficient selection of bacterial wilt resistance in *L. multiflorum* cultivars.

4.5. Materials and Methods

4.5.1. Plant material, bacterial material, inoculation

A total of 143 F_1 individuals from the Xtg-ART population (Studer et al., 2006) were planted in isolation tents for a polycross in late summer of 2017 and an average of 6,000 seeds were harvested from each F1 plant in early summer of 2018. Self-pollination was identified in the F1 plants using a fully informative single sequence repeat (SSR) marker, and 131 F1 plants were confirmed to come from cross-pollination and their seeds were selected for the experiment. An equal number of 61 seeds per F1 plant (7,991 seeds in total) were planted in 96 30x45 mm pot trays filled with soil and grown in the greenhouse. From these, 7,484 individual plants were established and formed the Xtg-ART-F_2 population. After a month, plant material was harvested from each F2 plant and stored individually at -80°C. The plants were allowed to grow back for a month before being inoculated.

For the inoculation, bacteria from the highly virulent Xtg29 strain (Kölliker et al., 2006) were grown at 28°C on YDC agar medium (2% dextrose, 1% yeast extract, 2% $CaCO_3$, 1.5% agar) for

48 h. Bacteria were then dissolved in a physiological solution (0.877% NaCl) and the OD_{580} was adjusted to 0.3, corresponding to a concentration of 10^8 CFU/mL.

The plants were inoculated by dipping scissors in the bacterial solution and cutting them 4 cm above soil. Each plant was then scored at 14, 21 and 28 days post infection (dpi). The plants were then cut back and allowed to grow back, and a last scoring was performed at 49 dpi. Scorings were performed on a scale of one to five, where one corresponds to a healthy plant showing no symptoms and five corresponds to a dead plant.

At the last scoring, 1,165 plants showed no symptoms and 1,140 were dead (Figure 1). From these, to select truly resistant and susceptible plants, inconsistent scores across the scoring period were filtered out. For the resistant individuals, plants showing a score of three or above anytime during the scoring period were excluded, and a total of 750 plants remained. For the susceptible individuals, plants showing a disease score of three and above at early stages (14 and 21 dpi) and of four and above at a later stage (28 dpi) were selected, accounting for a total of 761 plants.

4.5.2. DNA extraction and Sequencing

DNA was extracted from the previously harvested plant material using the NucleoSpin® Plant II DNA kit (Macherey Nagel, Duren, Germany). One leaf cutting of approximately three centimeters per plant was used. For a better quality and accuracy of the sequencing, each pool was divided into three sequencing libraries of an equal number of individuals (~250) that were sequenced independently and pooled after sequencing. Sequencing libraries were prepared using the Illumina TruSeq Nano DNA Library Preparation, (Illumina, San Francisco, CA, USA) and sequenced using 150 bp paired-end reads on one lane of an S4 flowcell using the Illumina NovaSeq 6000 platform by the Functional Genomics Center Zurich, Zurich, Switzerland.

4.5.3. Production of chromosome-level genome sequences, alignment and variant calling

For the alignment of the sequencing reads, a chromosome-level version of the M2289 resistant grandparent, M2289v2, was produced by anchoring the scaffolds of the published M2289 draft genome sequence to the recently-obtained high-quality assembly of the *L. multiflorum* cultivar 'Rabiosa' (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030979885.1/) using RagTag 'scaffold' v2.1.0 with default parameters (Knorst et al., 2018b; Alonge et al., 2019, 2021). The same approach was used to anchor the draft genome sequence of the Adret susceptible grandparent, resulting in the chromosome-level Adretv2 genome sequence (Knorst et al., 2018a). Additionally, the subsequent analysis was also applied using the Rabiosa genome sequence as reference.

The obtained reads were trimmed to remove adapter sequences and low-quality bases with trimmomatic v0.35 using TruSeq adapter sequence information and options "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:70" (Bolger et al., 2014). The trimmed reads were then aligned to the M2289v2 or the Rabiosa genome sequence with bowtie2 v2.3.4.2 using parameter "--no-unal" (Langmead and Salzberg, 2012). The resulting alignments were converted into binary alignment files and sorted using samtools v1.9 (Danecek et al., 2021). Variants were identified using samtools 'mpileup' and bcftools 'call' with option "-t DPR" for depth information (Danecek et al., 2021).

The resulting variant call file was then loaded into R version 4.1.0 with the 'vcfr' package v1.12.0 (R Core Team, 2022; Knaus and Grünwald, 2017). Loci with a QUAL score lower than 50 were excluded, and only biallelic loci were retained. Variants were filtered for read numbers >50 and <200 (summed across both pools).

4.5.4. Identification and characterization of candidate genes

For each variant, the Euclidean distance (ED) between the resistant and the susceptible pool was calculated as defined by Hill et al., 2013, modified for biallelic loci as follows:

$$ED = \sqrt{(Ref_{Res} - Ref_{Sus})^2 + (Alt_{Res} - Alt_{Sus})^2} \qquad \text{(equation 1)}$$

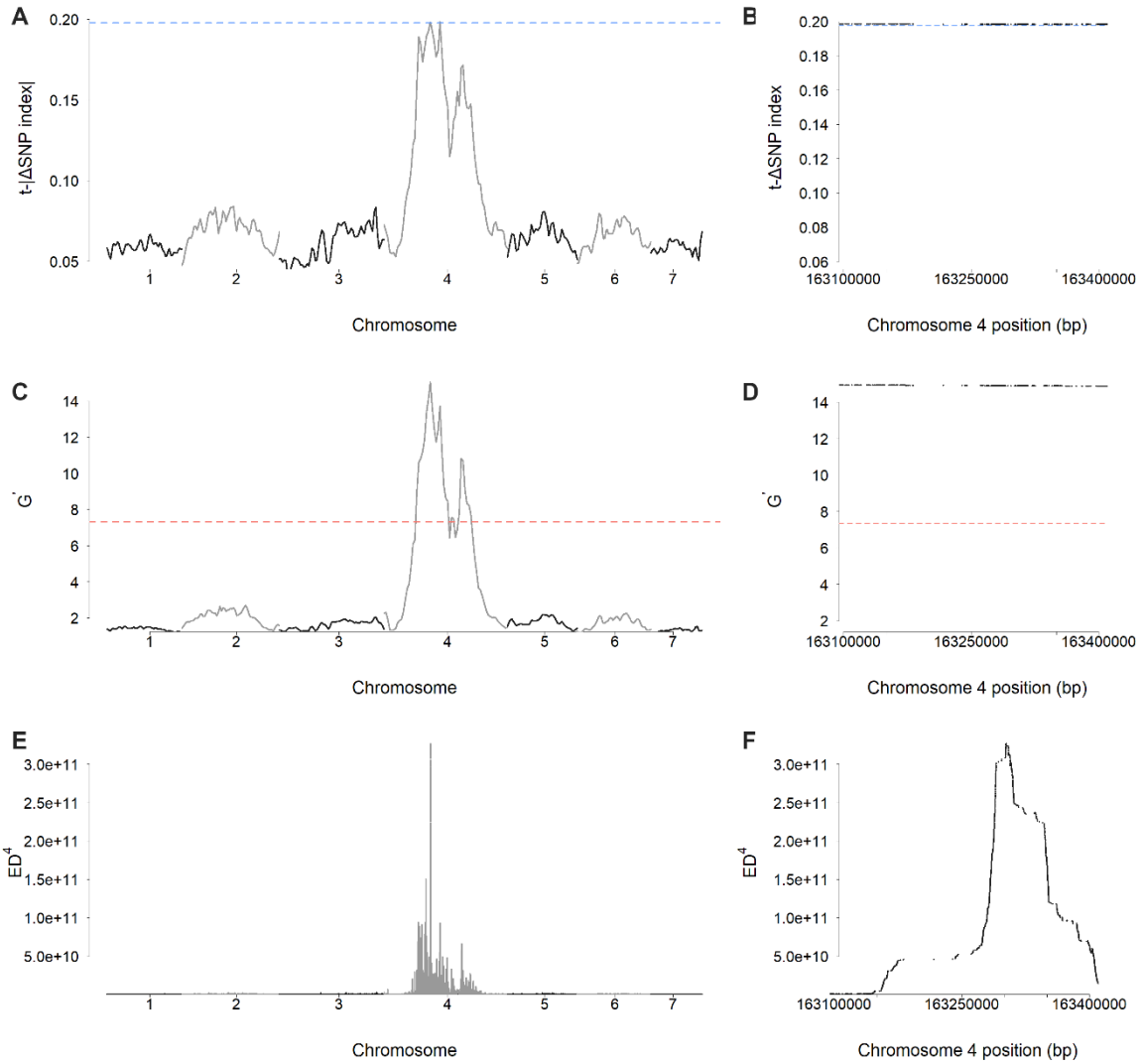
where Ref_{Res} and Ref_{Sus} are the frequencies of the reference allele in the resistant pool and in the susceptible pool, respectively, and Alt_{Res} and Alt_{Sus} are the frequencies of the alternative allele in the resistant pool and in the susceptible pool, respectively. To increase the signal to noise ratio, the ED^4 was calculated for each variant as the fourth power of the sum of ED within a window of 1,000 consecutive variants, centred around this variant. As the ΔSNP approach requires the reference genome used to be of a parent of the population, we used the absolute value of the ΔSNP (|ΔSNP|) in order to compare the analysis with M2289v2 and Rabiosa as reference genomes, as proposed by Omboki et al. (2018). The |ΔSNP| index and the G value, as well as their corresponding smoothed values, the t-|ΔSNP| index and G', were calculated with the QTLseqR package, using a sliding window size of 10^6 for the M2289v2 analysis and $5x10^6$ for the Rabiosa analysis, as this genome sequence is larger and more complete (Mansfeld and Grumet, 2018). The results were then visualized in Manhattan plots with the 'qqman' R package (Turner, 2018)

Candidate genes were defined by determining the genes present in the identified regions in the genome annotations of both M2289v2 and Rabiosa. The regions identified with both reference genome sequences were then compared by performing a BLASTn using the candidate gene sequences as query and the genome sequences as subject. This showed that the peak region in Rabiosa corresponds to a subset of the region identified in M2289v2. For a better comparison, the

larger region in Rabiosa corresponding to the region in M2289v2 was selected, and the list of genes appended accordingly. The identified genes were then annotated by performing a BLASTx with the gene nucleotide sequences as query and the TAIR10 *A. thaliana* or the NCBI nr protein databases as subject (Berardini et al., 2015; Sayers et al., 2022). Genes that showed homology to genes known to be involved in disease resistance were defined as candidate genes. The synteny between M2289v2 and Rabiosa was determined with a BLASTn-based approach using the SimpleSynteny online tool with the sequences of candidate genes in Rabiosa, as this assembly is more complete, as well as two additional genes that were identified in M2289v2 only (Table 1, Veltri et al., 2016).

In order to further characterize the candidate genes, raw reads from the M2289 and Adret grandparents of the XTG-ART-F_2 population were aligned to the M2289v2 and Rabiosa reference genome sequences and variants were determined in the same manner as before (Knorst et al., 2018a, 2018b). Candidate genes containing variants were defined and the allele frequencies of the variants investigated. As it is expected that the resistance allele is absent in susceptible individuals, loci where one allele had a frequency < 0.2 in the susceptible pools and the Adret susceptible grandparent were identified, and this allele was then defined as a resistance allele. Loci where such variants would result in a change in amino acid were then determined. The frequency of resistance alleles in these loci was then investigated in both pools, M2289 and Adret.

4.6. Supplementary material



Supplementary figure 1 Association of each variant locus to *Xanthomonas translucens* pv. *graminis* resistance in *Lolium multiflorum* calculated with the t-|ΔSNP| index (A, B), the G' value (C, D) and the ED^4 value (E, F), with Rabiosa as a reference genome sequence, on all chromosomes (A, C, E) or the region with the highest association to resistance on chromosome 4 (B, D, F). Dotted blue lines represent the 95% confidence interval for the t-|ΔSNP| index, and dotted red lines represent the 99% confidence interval for the t-|ΔSNP| index or the false discovery rate threshold of 1% for the G' value.

Supplementary Table 1 Genes found within the candidate region, defined from 28,800,000 bp to 28,900,000 bp on pseudochromosome 4 of the M2289v2 genome sequence, and from 163,090,000 bp to 163,410,000 bp in the Rabiosa genome sequence, that could not be annotated or could not be linked to disease resistance.

Genome sequence	Gene ID	Annotation[1]	Start (bp)	End (bp)
M2289v2	XLOC_040217	NA	28,813,107	28,814,194
	XLOC_036956	NA	28,834,593	28,835,078
	XLOC_036955	NA	28,835,563	28,835,649
	XLOC_036954	NA	28,838,096	28,838,188
	XLOC_036953	NA	28,838,292	28,838,354
	XLOC_036097	NA	28,841,401	28,841,562
	XLOC_000518	NA	28,848,401	28,848,547
	XLOC_023207	NA	28,851,730	28,852,380
Rabiosa	evm.TU.chr4.21722	retrotransposon	163,122,267	163,126,439
	evm.TU.chr4.21729	histone deacetylase	163,162,987	163,164,200
	evm.TU.chr4.21734	NA	163,227,754	163,228,353
	evm.TU.chr4.21743	NA	163,343,823	163,344,968
	evm.TU.chr4.21751	NA	163,389,342	163,394,185
	evm.TU.chr4.21753	NA	163,398,533	163,405,464

[1]NA: no homology to a known gene was found.

5. General discussion

Since the discovery of bacterial wilt in 1975, researchers and breeders have sought to develop *L. multiflorum* cultivars with increased resistance to the disease, however, complete resistance was never attained. During the following decades, first insights into a genomic region that may be responsible for resistance have been reported (Studer et al., 2006; Knorst et al., 2018a). Using a bulked segregant analysis coupled with high coverage whole genome sequencing, a 300 kbp region was defined to be highly associated with resistance **(Chapter 4)**. However, challenges remain in the path towards the development and commercialization of new resistant cultivars. First, candidate loci need to be functionally validated before they can be considered for breeding.

5.1. Strategies for candidate gene validation

Association studies and QTL mapping have become extensively used and have proven successful in uncovering candidate genes related to phenotypes of interest such a disease resistance (Young, 1996; Gangurde et al., 2022). However, the identified candidate genes require to be functionally validated and characterized in order to be used in breeding, as well as to give further insight as to how the disease resistance occurs. To validate candidate genes, generating mutations, ideally knockouts, for these genes is one of the most important approaches. Knocking out a gene involved in resistance to a pathogen would be expected to result in a more susceptible plant, thus substantiating the role of this gene in disease resistance. For example, in rice, knockout mutants of the newly identified *Xa47* gene were significantly more susceptible to *X. oryzae* pv. *oryzae* (*Xoo*) than the wild-type genotype, highlighting the role of *Xa47* in resistance to the pathogen (Lu et al., 2022b). To achieve such knockout mutations, one possible approach is the use of gene editing, mainly through the use of CRISPR/Cas-based systems (Jinek et al., 2012; Nidhi et al., 2021). This approach involves the use of guide RNA with a sequence that is specific to the targeted gene, together with a Cas nuclease, such as Cas9. This leads the nuclease to generate double strand breaks within the gene, which often leads to small insertions or deletions at the cleavage site, disrupting the gene. Currently, CRISPR/Cas-based genome editing is one of the main tools for validation of candidate genes, as it allows to specifically target a gene of interest to introduce a mutation (Gurumurthy et al., 2016; Chao et al., 2019; Curtin et al., 2020). However, this method requires optimized protocols to be implemented in an efficient and reliable manner, as a tissue culture step is necessary, with genetic transformation of calli followed by regeneration of explants (Altpeter et al., 2016). Both of these steps represent major challenges in genetic transformation of plants, in particular in monocotyledonous plants, for which transformation and regeneration efficiencies are rather low and often dependent on the genotype (Shrawat and Lörz,

2006). While targeted gene editing has been achieved in *Lolium* spp., a low editing efficiency has been observed so far (Zhang et al., 2020; Grogg et al., 2022).

A more accessible approach to generate mutations in candidate genes consists in using a Targeting Induced Local Lesions IN Genomes (TILLING) approach (McCallum et al., 2000; Tsai et al., 2011). This method involves the generation of a mutant population through random mutagenesis, followed by a screening of the individuals that harbor a mutation in the candidate gene. Random mutagenesis consists in the introduction of mutations using physical or chemical agents, and results in populations harboring numerous mutations throughout their genome (Bado et al., 2015; EFSA Panel on Genetically Modified Organisms (GMO), 2021). A TILLING method has successfully been implemented in *Lolium* spp., although some challenges remain (Manzanares et al., 2016). First, obtaining plants that are homozygous for the obtained mutation is difficult, as ryegrasses are mostly self-incompatible. In the case of *Xtg* resistance, generating a heterozygous plant with a mutation in the allele associated to resistance, while the other non-mutated allele is an allele associated with susceptibility would be possible, and prevent the need for a homozygous plant. This would require knowledge of the allele sequences from both haplotypes in the resistant parent and which allele is associated with resistance to ensure this allele is mutated through amplicon sequencing such as that performed in current TILLING by sequencing approaches (Tsai et al., 2011). Still, the crossing of the first mutant generation (M_1) would be required to avoid chimerism, as M_1 plants may carry cells that are not mutated, or harbor different mutations. Additionally, a method of producing homozygous plants similar to that proposed by Ukai (2010) could be devised. Selecting multiple plants with a mutation in a candidate gene in the M_1 generation and producing an M_2 progeny from these individuals in a polycross would then result in a population with heterozygous individuals with one wild type allele and one mutated allele, and homozygous individuals with two wild type alleles or two mutated alleles. Screening the latter for resistance to *Xtg* would thus support the validation of the role of this gene if all mutants share a similar phenotype. However, such method would require multiple M_1 plants with a mutation in genes of interest and may therefore be laborious and time-consuming.

Finally, validation of candidate genes may be achieved by means of overexpression or silencing of these genes (Baulcombe, 2004). Overexpression involves the introduction of the candidate gene under the control of a strong constitutive promoter in a susceptible individual, while silencing based on RNA interference involves the downregulation of the candidate gene in a resistant individual. The gain of resistance in the first case, or a loss of resistance in the second case would provide strong evidence for the role of the candidate in resistance to *Xtg*. However, both approaches rely on genetic transformation, which has been achieved in *Lolium* spp., but is still a challenging and time-consuming process, with a low transformation efficiency, as discussed above. An additional method for silencing that can be applied without relying on transgenesis is

virus-induced gene silencing (VIGS), which has been previously applied in *L. temulentum* using the barley stripe mosaic virus as a vector (Martin et al., 2013). Still, this method has not been further applied to *Lolium* spp. and may be arduous to use in validation of candidate genes for *Xtg* resistance.

While approaches based on genetic transformation such as gene editing-based knockouts, overexpression, or silencing would be ideal, they are not easily applicable to *L. multiflorum*. As a consequence, TILLING stands out as the most straightforward approach to validate the candidate genes identified in **(Chapter 4)**, and such an approach is currently being investigated.

5.2. Marker-assisted selection

While validation of the identified candidate genes would provide valuable information as to how the plant interacts with *Xtg* and allow to select for resistance more efficiently, such a validation is not strictly necessary to breed resistant cultivars. Indeed, thanks to the high coverage enabled by the high-throughput sequencing technologies used, many markers linked to *Xtg* resistance have been identified in the candidate region. Such markers can be very powerful in breeding for marker-assisted selection (MAS; Collard and Mackill, 2007; Bernardo, 2008). MAS allows to select plants according to their genotype rather than their phenotype, facilitating the identification and selection of desirable traits that may be difficult to assess. Many markers have been developed to select for traits of interest, such as markers linked to the *Lr23* gene for resistance to leaf rust in wheat, or to the *Xa38, Xa21 and xa13* genes for resistance to bacterial blight in rice (Chen et al., 2001; Swathi et al., 2019; Yugander et al., 2019; Sagar et al., 2020).

In the case of *Xtg* resistance, MAS would allow to fix the source of resistance, i.e., obtaining a population where all individuals are homozygous for the resistance allele. This fixation is difficult due to the outbreeding nature of ryegrasses, resulting in a high heterozygosity in populations. The dominant resistance allele can mask the recessive allele that does not offer resistance, and the latter allele is thus maintained in the population, leading to the presence of homozygous susceptible individuals, even in cultivars with increased resistance. Therefore, MAS would allow to identify individuals harboring alleles linked to susceptibility, which would not be possible by phenotypic selection, in order to remove them from the breeding population and prevent susceptibility to *Xtg*. Many types of markers have been used for MAS over the years, such as RAPD, RFLP and AFLP, and may still be used to this day, but they have largely been replaced by more efficient markers. Nowadays, SNP markers have become the most commonly used markers in MAS, as they are the most common form of genetic variation, found in high quantities and densities throughout the genome, and can be easily detected (Pandurangan et al., 2022; Song et al., 2023). Different approaches exist to genotype SNPs, including low to medium throughput

allele-specific PCR-based methods such as kompetitive allele-specific PCR (KASP) assays or high-throughput methods such as SNP arrays or genotyping-by-sequencing. KASP markers have the advantage of being flexible, easy to use and affordable compared to GBS or SNP arrays (He et al., 2014; Pandurangan et al., 2022). Therefore, KASP markers based on the SNPs identified in the candidate region for *Xtg* resistance in **(Chapter 4)** should be developed and validated for their ability to predict *Xtg* resistance (Semagn et al., 2014; Platten et al., 2019). Validated markers may then be applied routinely in MAS in breeding populations to introgress or fix *Xtg* resistance in new cultivars.

5.3. New sources of resistance for resistance management

Although the identified candidate genes and markers hold great potential to fix resistance in elite *L. multiflorum* cultivars and control the disease in an efficient manner, it is crucial to carefully manage the use of these cultivars to avoid the emergence of new strains that could overcome the resistance. Indeed, in rice, cultivars harboring the *Xa-4* gene conferring resistance to bacterial blight caused by *Xoo* were introduced in the early 1970s in the Philippines (Mew, 1992). This effectively decreased infections at first, but rapidly led to the emergence of a new race of the pathogen that was able to infect modern cultivars possessing the *Xa-4* gene. In current *L. multiflorum* cultivars with increased but not complete resistance to *Xtg*, the presence of susceptible individuals allows for a decreased selection pressure, which may delay the evolution of new virulent strains. Fixing *Xtg* resistance in highly resistant cultivars may therefore increase this selection pressure and lead to new virulent races of *Xtg*. Considering the possible high adaptability of *Xtg* that could be conferred by its genome plasticity, as reported in **(Chapter 3)**, it is essential to quickly diversify the approaches deployed to control the disease. Furthermore, the resistance identified in the M2289 individual is not a complete resistance and may vary depending on environmental factors. Indeed, in the first study where this QTL was identified, it accounted to as high as 84% of phenotypic variance for *Xtg* resistance, but only 43% in a similar experiment the previous year (Studer et al., 2006). In this regard, identifying additional sources of resistance to incorporate into breeding programs would help to increase both the efficiency and durability of *Xtg* resistance in *L. multiflorum* cultivars. These new sources of resistance may be incorporated in new cultivars through recurrent phenotypic selection, with the cultivation of mixtures of cultivars harboring different resistance genes being an efficient way to manage disease spread and the breakdown of resistance by the pathogen (de Vallavieille-Pope, 2004). Additionally, MAS allows to pyramid multiple resistance sources at once into one cultivar, resulting in cultivars with long-lasting, efficient and wide-ranging disease resistance (Dormatey et al., 2020).

The resistance source identified in this work originates from the breeding program of Agroscope Reckenholz, Switzerland, where resistance to *Xtg* is a main selection criterion, with plants being

screened for resistance before being incorporated into the breeding population (Humphreys et al., 2010). While this may not be the focus of other breeding programs, some of them do incorporate screening for *Xtg* resistance during the selection process and may hold great potential to identify new sources of resistance to the disease (Humphreys et al., 2010). Nevertheless, even germplasm coming from breeding programs without selection for *Xtg* resistance constitute a major source of genetic diversity where resistance genes could be found. Furthermore, wild populations of *L. multiflorum* coming from unmanaged and semi-managed grasslands may harbor unique resistance genes that have not yet been exploited in breeding programs. Indeed, in a study evaluating the performance of Swiss ecotypes from permanent meadows, it was found that some displayed a high resistance to *Xtg*, showing that natural selection likely has led to the selection of resistance (Boller et al., 2009). Finally, close relatives of *L. multiflorum*, such as *L. perenne*, *L. remotum*, *L. rigidum*, *L. temulentum*, or even *Festuca* spp. constitute an additional source of genetic diversity to inspect. Indeed, wild relatives have proven to be an invaluable resource for improving crops, in particular for disease and pest resistance (Migicovsky and Myles, 2017). Investigating these diverse populations using advanced genomic methods coupled with high throughput sequencing technologies, would greatly aid in the precise identification of new sources of resistance to *Xtg*.

5.3.1. Improved methods for resistance gene discovery

To identify resistance genes, classical QTL mapping is performed using biparental populations segregating for the phenotype of interest, such as the Xtg-ART and the Xtg-ART-F_2 population used in **(Chapter 4)**. While QTL mapping is a powerful method, it suffers from two major limitations (Korte and Farlow, 2013; Pascual et al., 2015). First, the analysis is limited to the allelic diversity coming from the two parents, which accounts for only a part of the genetic variation in the species. Secondly, it relies on only a few recombination events that occur in the development of the population, which limits the resolution of the QTL mapping. These two limitations are overcome in genome-wide association studies (GWAS), which allow to investigate diverse natural and breeding populations, and to exploit historical recombination events (Monnot et al., 2021; Demirjian et al., 2023). For example, a GWAS using 421 rice accessions was able to identify 13 QTL associated with resistance to *Xoo* (Lu et al., 2022a). In *L. perenne*, a GWAS based on 197 genotypes allowed for the identification of two candidate genes related to leaf growth under drought stress (Jaškūnė et al., 2020). However, one limitation of GWAS is the effect of population structure, which can lead to false positive associations and a reduced statistical power to detect true associations (Korte and Farlow, 2013; Vilhjálmsson and Nordborg, 2013). This effect can be reduced with statistical methods that account for population structure, such as mixed linear models that incorporate population structure as a fixed effect and kinship among individuals as a random effect (Yu et al., 2006; Zhang et al., 2010). Still, these methods may lead

to overcorrection for population structure and relatedness, thus increasing the number of false negatives (Zhao et al., 2011; Gyawali et al., 2016; Li et al., 2019).

Another way to address this issue of population structure, as well as the limited genetic diversity of QTL mapping in biparental populations, is to construct advanced mapping populations such as nested association mapping (NAM) or multi-parent advanced generation intercross (MAGIC) populations (Cavanagh et al., 2008; Yu et al., 2008). In NAM populations, multiple founder parents are crossed with a single reference parent, which harbors the phenotype of interest. Such a population was established in *L. multiflorum* using 23 founder genotypes crossed with one common founder and allowed for the identification of several QTL for resistance to stem rust (Kölliker et al., 2022). In MAGIC designs, multiple founder parents are intercrossed multiple times, resulting in highly diverse populations with genetic material from all founder parents. These advanced population designs increase genetic diversity compared to biparental populations, reduce the effects of population structure and allow for an increased statistical power to identify genetic loci that control a phenotype of interest with a high precision and resolution (Bandillo et al., 2013; Xu et al., 2017). Constructing such a population using *L. multiflorum* genotypes resistant to *Xtg* identified from diverse origins would be a valuable approach to uncover additional sources of resistance to *Xtg*. Nevertheless, QTL mapping in biparental populations, advanced mapping populations, or GWAS are all powerful methods, with each their advantages and limitations, and no method is superior to the other. Careful evaluation of the aims relative to the available material is necessary to determine the most appropriate method in a given situation.

In addition to the identification of existing resistance in diverse populations, another approach is the use of random mutagenesis to generate resistance in susceptible plants. Mutations in specific genes may confer resistance to a pathogen, and screening mutant populations allow for the identification of individuals harboring such mutations. To this day, 3,402 varieties resulting from mutagenesis are registered in the mutant variety database of the Food and Agriculture Organization and the International Atomic Energy Agency (accessible at https://www.iaea.org/resources/databases/mutant-varieties-database, accessed on 29/05/2023), and of these, 543 are registered as having been improved for resistance to biotic stresses through mutagenesis. Among these, 143 are varieties of rice, in which mutagenesis is advantageous due to its small genome (Viana et al., 2019). This has led to the development of cultivars that are resistant to rice blast caused by *Magnaporthe grisea* or bacterial blight caused by *Xoo* (Nakai et al., 1990; Jung et al., 2005; Busungu et al., 2016). In ryegrasses, the use of a mutagenized TILLING population to identify individuals with a phenotype of interest has been previously investigated in *L. perenne*, where 100,000 seeds were chemically mutagenized and screened for various traits related to abiotic stress tolerance (Manzanares et al., 2016). This allowed for the identification of 49 mutants that were tolerant to salt, 22 to osmotic stress, and 16 that were

insensitive to the stress-signaling hormone abscisic acid, holding potential for the development of cultivars with increased tolerance to salt, drought, or other stresses. Generating mutagenized *L. multiflorum* populations from susceptible populations and screening them for individuals exhibiting increased resistance to *Xtg* could thus allow for the identification of mutants of interest that could be introduced in breeding programs.

Transcriptomic analyses can be an additional option to identify, as well as validate candidate genes, and can give insights regarding how the host and pathogen interact. Genes that are expressed differentially at specific time points after infection, or in resistant genotypes compared to susceptible genotypes, may be of importance in the response to *Xtg*. A previous analysis in *L. multiflorum* investigating the transcriptome of a resistant genotype at different time point after inoculation, as well as another transcriptomic analysis investigating the differential gene expression of a resistant and a susceptible genotype at different time points were able to identify genes of interest (Rechsteiner et al., 2006; Wichmann et al., 2011a). However, the lack of genome information, together with the limited number of genes that could be investigated at the time, could not allow for a comprehensive understanding into the host-pathogen interaction. Current sequencing technologies now offer a high resolution in transcriptomic analysis, allowing for whole-transcriptome analyses (Stark et al., 2019; Tyagi et al., 2022). Such methods were applied in many crops, and allowed for the identification of candidate genes related to disease resistance (Gao et al., 2021; Meng et al., 2021; Geng et al., 2022). A similar approach was attempted in this doctoral work, with the aim of comparing the gene expression of one resistant genotype and a susceptible genotype at four different time points after infection (Bürki, 2021). However, this analysis did not yield satisfying results, and no candidate gene with a role in disease resistance could be identified. Further investigation revealed that the two selected genotypes shared similar allele frequencies in the region identified in **(Chapter 4)**, suggesting that both genotypes were resistant. Moreover, disease symptoms were very mild even in positive controls, which may be due to a low starting inoculum, or an increased basal resistance due to previous infections by pests in the greenhouse. Nonetheless, undertaking such an experiment after selection of genotypes using genetic markers and a more stringent control of experimental conditions would be a valuable approach in identifying and validating sources of resistance to *Xtg*.

5.3.2. Investigating bacterial virulence to breed resistant plants

Studying bacterial virulence factors, in particular effectors, is also a crucial element in breeding for disease resistance, as it provides knowledge about how the pathogen is able to infect its host and helps to identify ways to control the disease. This therefore requires a comprehensive overview of the effectors used by the pathogen to infect its host. While a set of 22 putative type III effectors (T3E) of *Xtg* was identified in **(Chapter 2)** and **(Chapter 3)**, confirming their

secretion by the T3SS is necessary. Moreover, the BLAST-based approach used in this work has limitations. First, putative effectors were identified solely based on their homology to known effectors, which may not signify a similar function. Secondly, this approach may miss unique effectors that could be present in *Xtg*. To accurately define the effectome of *Xtg*, different approaches could be investigated. Thanks to the increasing knowledge about T3E, similar amino acid compositions of amino-terminal regions were identified, which enabled the development of bioinformatic models of the secretion signals of T3E that allow to predict putative T3E (Guttman et al., 2002; Wagner et al., 2022a, 2022b). Additionally, methods to screen proteins secreted by the T3SS *in vivo* have been devised, such as an assay using insertions of the C-terminal part of the AvrRpt2 T3E from *Pseudomonas syringae*, which induces hypersensitive response in *A. thaliana* (Guttman et al., 2002). The observation of hypersensitive response would thus mean that the insertion of this C-terminal part, which lacks a secretion signal, occurred in a T3E that possesses one. This method allowed the identification of 13 effectors of *P. syringae*. A similar screen was developed in *Xcv* using the AvrBs2 effector devoid of its secretion signal and identified seven T3E (Roden et al., 2004). As no avirulence factor has yet been identified in *Xtg*, such a screen is not directly transferable to the *Xtg* – *L. multiflorum* pathosystem. However, using a heterologous system such as those described above, in which *Xtg* genes are fused to the reporter avirulence gene may allow to identify new *Xtg* effectors.

The effectors and conserved features of *Xtg* identified in **(Chapter 3)** represent an important list of potential virulence factors that could explain the pathogenicity and host range of the pathovar. However, to confirm their role in the virulence of *Xtg*, it is crucial to functionally validate these genes. Compared to *Lolium* spp., targeted gene knockouts are easier to produce in bacteria such as *Xanthomonas* spp. So far, this has been achieved through the use of site-directed mutagenesis, and more recently with CRISPR-Cas-based genome editing approaches (Sauer et al., 2016; Yan et al., 2023). Furthermore, as *Xanthomonas* spp. can possess endogenous CRISPR-Cas systems, such systems may even be repurposed for easier gene editing (Jiang et al., 2022). Knockouts mutants allowed for example to confirm the important role of the effectors XopA and XopN in virulence of *X. campestris* pv. *vesicatoria* (*Xcv*) in pepper and tomato (Noël et al., 2002; Roden et al., 2004). In *Xtg*, knockout mutants for the *hrpE* and *hrcR* structural genes of the T3SS, as well as a mutant for the *hrpG* regulator showed a significant reduction in disease symptoms, confirming the role of the T3SS in infection by *Xtg* (Wichmann et al., 2013). Generating similar mutants for each of the 20 conserved type III effectors found in *Xtg* would give further insights as to their role in virulence.

However, such studies on individual effectors may overlook the cumulative role of each effector, as well as functional redundancy, and effectome-wide studies would provide a more comprehensive overview of the role of effectors in *Xtg* (Arroyo-Velez et al., 2020). In *P. syringae*,

the DC3000Δ36E strain is a polymutant strain in which each of the 36 type III effectors from the pathogen were deleted (Wei et al., 2015). The development of strains possessing pairs of effectors derived from this effectorless strain allowed to observe the interplay between effectors (Wei et al., 2018). While six effectors acted as avirulence determinants, triggering cell death, eight other effectors suppressed this avirulence when paired with any of the six effectors. Furthermore, 36 strains carrying only a single effector of the 36 type III effectors found in the wild-type bacteria were developed from DC3000Δ36E (Ruiz-Bedoya et al., 2023). While each strain was unable to infect *A. thaliana* independently, the 36 strains together were as virulent as the wild-type strain, further highlighting the cooperative behavior of type III effectors. Developing such an effectorless strain of *Xtg* or mutating more than one effector at a time would thus be a valuable approach in understanding the interplay between each effector of the pathovar.

Following the identification *Xtg* effectors, characterizing their mode of action in the plant would be crucial to understand the molecular basis of the host-pathogen interaction. Effectors are proteins that generally act on host processes by targeting one or multiple host proteins, as well as nucleic acids or metabolites. In rice, many genes of the *OsSWEET* family, encoding for sugar transporters, are the target of TAL effectors from *Xoo*, which bind to their promoter sequence (Streubel et al., 2013b). Using a genome editing approach, rice lines mutated in the promoter sequences of five *OsSWEET* genes showed robust resistance to various strains of the pathogen (Oliva et al., 2019). A similar result was obtained in citrus by editing the *CsLOB1* gene, which is involved in susceptibility to *X. citri* pv. *citri* (Jia et al., 2017). In *L. multiflorum*, the identification of targets of these effectors would thus provide valuable information in the pursuit of breeding resistant cultivars, and targeting such genes involved in susceptibility to *Xtg* using gene editing or TILLING approaches may render the plant resistant (van Schie and Takken, 2014; Borrelli et al., 2018). Moreover, identifying natural variants of these genes in breeding and natural populations with approaches such as ecoTILLING, that applies the principles of TILLING in natural populations, could allow to identify existing resistance to *Xtg* (Comai et al., 2004). Additionally, resistance genes from plants can act in reaction to the detection of effectors, and the investigation of plant proteins from resistant cultivars that interact with effectors may uncover plant proteins that act in resistance to *Xtg*. To identify plant susceptibility genes or resistance genes, determining the host proteins that interact with effectors is possible using methods that identify protein-protein interactions, such as yeast two-hybrid (Young, 1998), bimolecular fluorescence complementation (Kerppola, 2006) or co-immunoprecipitation (Lin and Lai, 2017). Using yeast two-hybrid screens showed that the pepper aldehyde dehydrogenase CaALDH1 as well as the heat shock protein CaHSP70a interact with the XopJ family effector AvrBsT from *Xcv*, triggering hypersensitive response (Kim and Hwang, 2015a, 2015b). For effectors that bind to the host DNA, chromatin immunoprecipitation assays allow to determine the binding site of

the effector in the plant genome (Gade and Kalvakolanu, 2012). Such a method allowed to confirm that the AvrBs3 TAL effector from *Xcv* binds to the promoter region of the *upa20* transcription factor from pepper (Kay et al., 2007). These approaches are key in the understanding which molecular components are involved in the host-pathogen dialogue. Identifying plant proteins that interact with the T3E of *Xtg* would therefore provide valuable insight for breeding *Xtg* resistance.

5.4. High-quality genome sequences, an important tool for breeding

Genomic resources play a pivotal role in identifying sources of disease resistance for breeding, as they enable or facilitate many of the genomic-based approaches discussed above. Furthermore, in these approaches, the quality of the available genomic resources is a very important aspect. Draft genome sequences have proven to be valuable resources, and allowed the identification of genes of interest, such as candidate genes linked to drought tolerance in *L. perenne*, or potential virulence factors of *Xtg* (Hersemann et al., 2017; Jaškūnė et al., 2020). However, draft genome sequences have many limitations (Palmer and McCombie, 2002).

First, they generally consist of many contigs of various sizes without precise information as to their location in the genome. In a first analysis of the data from **(Chapter 4)** using the M2289 draft genome sequence as a reference, the position of scaffolds was estimated based on synteny to the barley genome sequence (Goettelmann et al., 2021). While a clear association to resistance could be seen on chromosome 4, the observed peak showed an inconsistent pattern towards the beginning of the chromosome, likely due to a different structure of chromosome 4 in *L. multiflorum* compared to its homolog in barley. Further analysis using the chromosome-level Rabiosa genome sequence as a reference showed a more consistent pattern. In blueberry (*Vaccinium corymbosum*), the re-analysis of data from a previous GWAS based on a draft genome sequence using a newly produced complete genome showed that the significant SNPs identified with the draft genome sequence were scattered throughout the unplaced scaffolds, leading to possible misinterpretations as to how many loci are involved in the trait variation (Benevenuto et al., 2019). Furthermore, the analysis using this draft genome sequence did not allow to accurately locate these SNPs in the genome. Conversely, the use of a high-quality, chromosome-level genome sequence allowed to observe two clear peaks with a precise location and to investigate the underlying genes. Additionally, this genome fragmentation and lack of positional information of scaffolds hinders the observation of structural variations such as large insertions or deletions and chromosomal rearrangements (Yang, 2020). In *X. translucens*, the production of complete genome sequences allowed to compare the chromosomal structure between pathovars and strains **(Chapter 2 and 3)**. This showed that clade Xt-III, and in particular *Xtg*, displayed a high genome plasticity compared to the rest of the species.

Secondly, while these genome sequences provide a good representation of the gene content within the genome, as evidenced by the 80.5% of complete BUSCOs in the M2289 draft genome sequence, they are not entirely comprehensive (Knorst et al., 2018b). Genes contained in regions that were not assembled will be missing from the final genome sequence, and genes that are present in the assembly may be fragmented due to the high number of small scaffolds. In the M2289 draft genome sequence, while scaffolds were larger than 2 kb, the N50 scaffold length was 5,036 bp and the median scaffold length was 3,628 bp. This likely results in genes being contained in more than one scaffold, with sequence information missing, hindering their annotation. This is evidenced by the 11% of missing BUSCOs and the 8.5% of fragmented BUSCOs in the draft M2289 genome assembly. In *Xtg*, previous draft genome sequences consisted of 249 to 369 contigs, with a genome size of 4.2 Mb and 3,519 to 3,553 predicted coding genes, while the complete genome assemblies obtained in **(Chapter 2)** and **(Chapter 3)** have a size of 4.7 Mb, with 4,408 to 4,445 predicted genes (Hersemann et al., 2017). In a survey of bacterial draft genome sequences in the GenBank database, varying levels of gene fragmentation were observed, reaching up to 81.4% of fragmented genes, and this gene fragmentation correlated with genome sequence fragmentation (Klassen and Currie, 2012). An in-depth analysis of *Streptomyces* spp. genome sequences revealed that gene fragmentation was particularly enriched in some protein classes with highly repetitive structures. In *Xanthomonas* spp., this is particularly important in the case of TAL effectors, which harbor repetitive sequences that make their assembly complex and often impossible with short read sequencing data (Erkes et al., 2023). This highlights the limitations of draft genome sequences to accurately determine the gene content of individuals or strains, and the importance of complete genome sequences for comparative analyses.

Lastly, these genome sequences are derived from a single individual. However, one individual cannot be representative of the whole genetic diversity present in the species, or a population under study. While the M2289 genome sequence is representative of a high proportion of the diversity within the Xtg-ART-F_2 population, it is lacking many alleles found in the Adret parent. Moreover, as *L. multiflorum* is outbreeding, its genome is highly heterozygous, and a haploid genome sequence such as the M2289 draft genome sequence also lacks alleles found in the Xtg-ART-F_2 population. Furthermore, the production of the chromosome-level M2289v2 genome sequence was done based on synteny to the Rabiosa genome sequence, which likely lacks the source of resistance, possibly hindering the anchoring of the scaffold containing it.

As a consequence, the alignment of reads to a reference genome sequence introduces reference bias, as reads containing DNA sequences that are not present in the reference may align improperly, or not at all (Martiniano et al., 2020). In blueberry, using a complete genome assembly revealed loci of interest that were not detected in the first analysis with a draft genome

sequence (Benevenuto et al., 2019). Furthermore, genes found near significant SNPs were found to be incomplete in the draft genome sequence, with no functional annotation, while the complete genome sequence provided complete gene sequences with existing orthologs or sequence annotation. This highlights the importance of high-quality, complete genome sequences that are representative of the studied individuals, populations, or species to accurately identify novel genes that may play a role in host-pathogen interactions.

With the development of long-read sequencing from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), as well as additional methods for gap closure such as mate-pair sequencing or optical mapping, producing high-quality, complete genome sequences has become increasingly accessible. Furthermore, these tools and additional approaches such as trio binning allow for the production of phased, haplotype-resolved diploid genome assemblies, providing a comprehensive overview of the genetic diversity present in an individual (Tewhey et al., 2011; Koren et al., 2018). In *L. multiflorum*, a first nearly complete, high-quality draft diploid genome assembly revealed a high intergenic sequence variation between alleles, further highlighting the importance of diploid genome sequences (Copetti et al., 2021).

Additionally, to solve the limitations of reference genome-based analyses, or when such a genome sequence is not available, reference-free analyses have been developed (Mehrab et al., 2020; Neves et al., 2020; Voichek and Weigel, 2020). These analyses are based on the production of small subsequences from sequencing reads, followed by the identification of subsequences that are specific to genotypes displaying the phenotype of interest. The resulting subsequences may then be aligned to an existing reference genome sequence, or *de novo* assembled in order to produce DNA sequences that are associated with the phenotype of interest, in which genes can be identified. In potato, a reference-free analysis of sequencing data from pools of individuals of a biparental population that are resistant or susceptible to potato wart disease, as well as the resistant and susceptible parents allowed to narrow down a resistance QTL to a 260 kb region (Prodhomme et al., 2019). Such an approach was investigated in **(Chapter 4)**, however the high linkage disequilibrium in the Xtg-ART-F_2 population prevented the identification of a small DNA sequence linked to *Xtg* resistance. Nonetheless, this method may be applied using a more diverse sample with a lower population structure.

Going further than the development of complete genome sequences, pangenomes have recently emerged as new genomic resources, solving many of the limitations of reference genome sequences (Wang et al., 2022; Liao et al., 2023). These sequences offer a comprehensive representation of global genetic diversity within a species and allow to fully exploit this diversity in association studies. While reference genome-based analyses rely on SNPs and small insertions and deletions, pangenomes also allow to investigate large insertions and deletions, as well as

chromosome rearrangements. Pangenome-based analysis pipelines have been developed to perform association studies and allowed to effectively reduce reference bias and identify genetic loci of interest (Garrison et al., 2018; Sirén et al., 2021; Ebler et al., 2022). In tomato (*Solanum lycopersicum*), using a pangenome constructed from 838 genome sequences allowed the identification of variants linked to fruit soluble solid content, highlighting the power of pangenome-based analyses to identify genes of interest (Zhou et al., 2022). Going forward, genomic resources are likely to become more complete, more accurate, and therefore allow more efficient analyses of the genetic control of traits of interest. Such genomic resources are and will be crucial tools in the hands of researchers and breeders in order to develop crops that are more productive, more resistant to biotic and abiotic stresses, and allow for a more sustainable agriculture.

Bacterial wilt of forage grasses is an important disease that threatens livestock production and needs to be controlled in an efficient manner. The advanced genomic approaches presented in this thesis have allowed to 1) identify a narrow region with a high association to *Xtg* resistance, harboring multiple genes that represent good candidate genes to investigate further 2) comprehensively characterize the *Xtg* genome sequence, revealing a high genome plasticity within the pathovar, and 3) identify potential virulence factors of *Xtg* that may explain its high virulence. Markers found within the identified candidate region carry substantial potential for fixing the resistance to *Xtg* in breeding populations and new cultivars. Overall, this work provides valuable knowledge about the *L. multiflorum* – *Xtg* interaction and is an important step towards the development of *L. multiflorum* cultivars with a durable resistance to *Xtg*.

6. Bibliography

Abramovitch, R. B., Anderson, J. C., and Martin, G. B. (2006). Bacterial elicitation and evasion of plant innate immunity. *Nat Rev Mol Cell Biol* 7, 601–611. doi: 10.1038/nrm1984.

Adhikari, T. B., Gurung, S., Hansen, J. M., Jackson, E. W., and Bonman, J. M. (2012). Association mapping of quantitative trait loci in spring wheat landraces conferring resistance to bacterial leaf streak and spot blotch. *The Plant Genome* 5. doi: 10.3835/plantgenome2011.12.0032.

Alonge, M., Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., Lippman, Z. B., et al. (2021). Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. Plant Biology doi: 10.1101/2021.11.18.469135.

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., et al. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* 20, 224. doi: 10.1186/s13059-019-1829-6.

Altpeter, F., Springer, N. M., Bartley, L. E., Blechl, A. E., Brutnell, T. P., Citovsky, V., et al. (2016). Advancing crop transformation in the era of genome editing. *The Plant Cell* 28, 1510–1520. doi: 10.1105/tpc.16.00196.

Alvarez-Martinez, C. E., Sgro, G. G., Araujo, G. G., Paiva, M. R. N., Matsuyama, B. Y., Guzzo, C. R., et al. (2021). Secrete or perish: The role of secretion systems in *Xanthomonas* biology. *Computational and Structural Biotechnology Journal* 19, 279–302. doi: 10.1016/j.csbj.2020.12.020.

An, Q., Ehlers, K., Kogel, K.-H., van Bel, A. J. E., and Hückelhoven, R. (2006). Multivesicular compartments proliferate in susceptible and resistant MLA12-barley leaves in response to infection by the biotrophic powdery mildew fungus. *New Phytologist* 172, 563–576. doi: 10.1111/j.1469-8137.2006.01844.x.

Anderson, J. A., Chao, S., and Liu, S. (2007). Molecular breeding using a major QTL for fusarium head blight resistance in wheat. *Crop Science* 47, S-112-S-119. doi: 10.2135/cropsci2007.04.0006IPBS.

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., and Singh, K. B. (2010). Plants versus pathogens: an evolutionary arms race. *Funct Plant Biol* 37, 499–512. doi: 10.1071/FP09304.

Arroyo-Velez, N., González-Fuente, M., Peeters, N., Lauber, E., and Noël, L. D. (2020). From effectors to effectomes: Are functional studies of individual effectors enough to decipher plant pathogen infectious strategies? *PLoS Pathog* 16, e1009059. doi: 10.1371/journal.ppat.1009059.

Ashikawa, I., Hayashi, N., Abe, F., Wu, J., and Matsumoto, T. (2012). Characterization of the rice blast resistance gene Pik cloned from Kanto51. *Mol Breeding* 30, 485–494. doi: 10.1007/s11032-011-9638-y.

Aslam, S. N., Newman, M.-A., Erbs, G., Morrissey, K. L., Chinchilla, D., Boller, T., et al. (2008). Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Current Biology* 18, 1078–1083. doi: 10.1016/j.cub.2008.06.061.

Athinuwat, D., Brooks, S., Burr, T. J., and Prathuangwong, S. (2018). Flagella and pili of *Xanthomonas axonopodis* pv. *glycines* are associated with motility, biofilm formation and virulence on soybean. *Journal of Phytopathology* 166, 590–600. doi: 10.1111/jph.12722.

Bado, S., Forster, B. P., Nielen, S., Ali, A. M., Lagoda, P. J. L., Till, B. J., et al. (2015). "Plant mutation breeding: current progress and future assessment," in *Plant Breeding Reviews: Volume 39* (John Wiley & Sons, Ltd), 23–88. doi: 10.1002/9781119107743.ch02.

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6, 11. doi: 10.1186/1939-8433-6-11.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15, 293. doi: 10.1186/1471-2105-15-293.

Barka, G. D., and Lee, J. (2020). Molecular marker development and gene cloning for diverse disease resistance in pepper (*Capsicum annuum* l.): current status and prospects. *Plant Breeding and Biotechnology* 8, 89–113. doi: 10.9787/PBB.2020.8.2.89.

Bartoš, J., Sandve, S. R., Kölliker, R., Kopecký, D., Christelová, P., Stočes, Š., et al. (2011). Genetic mapping of DArT markers in the *Festuca–Lolium* complex and their use in freezing tolerance association analysis. *Theoretical and Applied Genetics* 122, 1133–1147. doi: 10.1007/s00122-010-1518-z.

Baulcombe, D. (2004). RNA silencing in plants. *Nature* 431, 356–363. doi: 10.1038/nature02874.

Bayer-Santos, E., Ceseti, L. de M., Farah, C. S., and Alvarez-Martinez, C. E. (2019). Distribution, function and regulation of type 6 secretion systems of Xanthomonadales. *Front. Microbiol.* 10, 1635. doi: 10.3389/fmicb.2019.01635.

Bendtsen, J. D., Kiemer, L., Fausbøll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiology* 5, 58. doi: 10.1186/1471-2180-5-58.

Benevenuto, J., Ferrão, L. F. V., Amadeu, R. R., and Munoz, P. (2019). How can a high-quality genome assembly help plant breeders? *Gigascience* 8, giz068. doi: 10.1093/gigascience/giz068.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *genesis* 53, 474–485. doi: 10.1002/dvg.22877.

Bernard, J. K., West, J. W., and Trammell, D. S. (2002). Effect of replacing corn silage with annual ryegrass silage on nutrient digestibility, intake, and milk yield for lactating dairy cows. *Journal of dairy science* 85, 2277–2282.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131.

Bernier, F., and Berna, A. (2001). Germins and germin-like proteins: plant do-all proteins. But what do they do exactly? *Plant physiology and biochemistry* 39, 545–554.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., et al. (2007). Plant carbohydrate scavenging through tonb-dependent receptors: a feature shared by

phytopathogenic and aquatic bacteria. *PLoS One* 2, e224. doi: 10.1371/journal.pone.0000224.

Boch, J., and Bonas, U. (2010). *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* 48, 419–436.

Bogdanove, A. J., Beer, S. V., Bonas, U., Boucher, C. A., Collmer, A., Coplin, D. L., et al. (1996). Unified nomenclature for broadly conserved *hrp* genes of phytopathogenic bacteria. *Molecular Microbiology* 20, 681–683. doi: 10.1046/j.1365-2958.1996.5731077.x.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.

Boller, B., Peter-Schmid, M. K. I., Tresch, E., Tanner, P., and Schubiger, F. X. (2009). Ecotypes of Italian ryegrass from Swiss permanent grassland outperform current recommended cultivars. *Euphytica* 170, 53–65. doi: 10.1007/s10681-009-9963-y.

Boller, B., Tanner, P., Schubriger, F., and Streckeisen, P. (2001). "Selecting meadow fescue ecotypes for reduced susceptibility to bacterial wilt.," in *Breeding for stress tolerance in fodder crops and amenity grasses* Proceedings of the 23rd Meeting of the Fodder Crops and Amenity Grasses Section of EUCARPIA, Azores, Portugal, 1-4 October 2000., ed.European Association for Research on Plant Breeding (Terceira Island, Azores: Department of Agricultural Sciences), 103–107.

Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346.

Bonas, U., Schulte, R., Fenselau, S., Mainsavage, G. V., Staskawicz, B. J., and Stall, R. E. (1991). Isolation of a gene cluster from *Xanthomonas campestris* pv. *vesicatoria* that determines pathogenicity and the hypersensitive response on and the hypersensitive response on pepper and tomato. *Mol. Plant-Microbe Interact.* 4, 81–88.

Bonas, U., Stall, R. E., and Staskawicz, B. (1989). Genetic and structural characterization of the avirulence gene avrBs3 from *Xanthomonas campestris* pv. *vesicatoria*. *Mol Gen Genet* 218, 127–136. doi: 10.1007/BF00330575.

Booher, N. J., Carpenter, S. C. D., Sebra, R. P., Wang, L., Salzberg, S. L., Leach, J. E., et al. (2015). Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. *Microbial Genomics* 1, e000032. doi: 10.1099/mgen.0.000032.

Borrelli, V. M. G., Brambilla, V., Rogowsky, P., Marocco, A., and Lanubile, A. (2018). The enhancement of plant disease resistance using CRISPR/Cas9 technology. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01245.

Bragard, C., Singer, E., Alizadeh, A., Vauterin, L., Maraite, H., and Swings, J. (1997). *Xanthomonas translucens* from small grains: diversity and phytopathological relevance. *Phytopathology*® 87, 1111–1117. doi: 10.1094/PHYTO.1997.87.11.1111.

Broekaert, W. F., Delauré, S. L., De Bolle, M. F. C., and Cammue, B. P. A. (2006). The role of ethylene in host-pathogen interactions. *Annu Rev Phytopathol* 44, 393–416. doi: 10.1146/annurev.phyto.44.070505.143440.

Brown, C. L., Mullet, J., Hindi, F., Stoll, J. E., Gupta, S., Choi, M., et al. (2022). mobileOG-db: a manually curated database of protein families mediating the life cycle of bacterial mobile genetic elements. *Applied and Environmental Microbiology* 88, e00991-22. doi: 10.1128/aem.00991-22.

Bürki, F. (2021). Transcriptomic analysis of the *Lolium multiflorum* Lam. – *Xanthomonas translucens* pv. *graminis* interaction. Master's Thesis. ETH Zurich.

Busungu, C., Taura, S., Sakagami, J.-I., and Ichitani, K. (2016). Identification and linkage analysis of a new rice bacterial blight resistance gene from XM14, a mutant line from IR24. *Breed. Sci.* 66, 636–645. doi: 10.1270/jsbbs.16062.

Büttner, D., and Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* 34, 107–133. doi: 10.1111/j.1574-6976.2009.00192.x.

Byrne, S. L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., et al. (2015). A synteny-based draft genome sequence of the forage grass Lolium perenne. *The Plant Journal* 84, 816–826. doi: 10.1111/tpj.13037.

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* 38, 5825–5829. doi: 10.1093/molbev/msab293.

Cao, H., Glazebrook, J., Clarke, J. D., Volko, S., and Dong, X. (1997). The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* 88, 57–63.

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* 11, 215–221. doi: 10.1016/j.pbi.2008.01.002.

Cernadas, R. A., Camillo, L. R., and Benedetti, C. E. (2008). Transcriptional analysis of the sweet orange interaction with the citrus canker pathogens *Xanthomonas axonopodis* pv. *citri* and *Xanthomonas axonopodis* pv. *aurantifolii*. *Molecular Plant Pathology* 9, 609–631. doi: 10.1111/j.1364-3703.2008.00486.x.

Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., et al. (2014). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J* 33, 1941–1959. doi: 10.15252/embj.201487923.

Channon, A. G., and Hissett, R. (1984). The incidence of bacterial wilt caused by *Xanthomonas campestris* pv *graminis* in pasture grasses in the West of Scotland. *Plant Pathology* 33, 113–121. doi: 10.1111/j.1365-3059.1984.tb00594.x.

Chao, T., Liu, Z., Zhang, Y., Zhang, L., Huang, R., He, L., et al. (2019). Precise and rapid validation of candidate gene by allele specific knockout with CRISPR/Cas9 in wild mice. *Frontiers in Genetics* 10.

Chardon, F., Bedu, M., Calenge, F., Klemens, P. A. W., Spinner, L., Clement, G., et al. (2013). Leaf fructose content is controlled by the vacuolar transporter SWEET17 in arabidopsis. *Current Biology* 23, 697–702. doi: 10.1016/j.cub.2013.03.021.

Chen, K., Fan, B., Du, L., and Chen, Z. (2004). Activation of hypersensitive cell death by pathogen-induced receptor-like protein kinases from Arabidopsis. *Plant molecular biology* 56, 271–283.

Chen, L.-Q. (2014). SWEET sugar transporters for phloem transport and pathogen nutrition. *New Phytologist* 201, 1150–1155. doi: 10.1111/nph.12445.

Chen, S., Xu, C. G., Lin, X. H., and Zhang, Q. (2001). Improving bacterial blight resistance of '6078', an elite restorer line of hybrid rice, by molecular marker-assisted selection. *Plant Breeding* 120, 133–137. doi: 10.1046/j.1439-0523.2001.00559.x.

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10, 563–569. doi: 10.1038/nmeth.2474.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J. D. G., et al. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* 448, 497–500. doi: 10.1038/nature05999.

Choi, H. W., Kim, D. S., Kim, N. H., Jung, H. W., Ham, J. H., and Hwang, B. K. (2013). *Xanthomonas* filamentous hemagglutinin-like protein Fha1 interacts with pepper hypersensitive-induced reaction protein CaHIR1 and functions as a virulence factor in host plants. *Mol Plant Microbe Interact* 26, 1441–1454. doi: 10.1094/MPMI-07-13-0204-R.

Choi, Y., Kim, N., Mannaa, M., Kim, H., Park, J., Jung, H., et al. (2020). Characterization of type VI secretion system in *Xanthomonas oryzae* pv. *oryzae* and its role in virulence to rice. *Plant Pathol J* 36, 289–296. doi: 10.5423/PPJ.NT.02.2020.0026.

Chu, Z. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes & Development* 20, 1250–1255. doi: 10.1101/gad.1416306.

Chu, Z., Fu, B., Yang, H., Xu, C., Li, Z., Sanchez, A., et al. (2006). Targeting xa13, a recessive gene for bacterial blight resistance in rice. *Theoretical and Applied Genetics* 112, 455–461. doi: 10.1007/s00122-005-0145-6.

Clark, M., and Tilman, D. (2017). Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. *Environmental Research Letters* 12, 064016. doi: 10.1088/1748-9326/aa6cd5.

Clavijo, F., Barrera, C., Benčič, A., Croce, V., Jacobs, J. M., Bernal, A. J., et al. (2022). Complete genome sequence resource for *Xanthomonas translucens* pv. *undulosa* MAI5034, a wheat pathogen from Uruguay. *Phytopathology*® 112, 2036–2039. doi: 10.1094/PHYTO-01-22-0025-A.

Collard, B. C. Y., and Mackill, D. J. (2007). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 557–572. doi: 10.1098/rstb.2007.2170.

Collinge, D. B. (2009). Cell wall appositions: the first line of defence. *Journal of Experimental Botany* 60, 351–352. doi: 10.1093/jxb/erp001.

Comai, L., Young, K., Till, B. J., Reynolds, S. H., Greene, E. A., Codomo, C. A., et al. (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *The Plant Journal* 37, 778–786. doi: 10.1111/j.0960-7412.2003.01999.x.

Copetti, D., Yates, S. A., Vogt, M. M., Russo, G., Grieder, C., Kölliker, R., et al. (2021). Evidence for high intergenic sequence variation in heterozygous Italian ryegrass (*Lolium multiflorum* Lam.) genome revealed by a high-quality draft diploid genome assembly. Genomics doi: 10.1101/2021.05.05.442707.

Curtin, S. J., Tiffin, P., and Young, N. D. (2020). "The use of CRISPR/Cas9 as a reverse genetics tool to validate genome-wide association candidates," in *The Model Legume Medicago truncatula* (John Wiley & Sons, Ltd), 882–886. doi: 10.1002/9781119409144.ch110.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi: 10.1093/gigascience/giab008.

Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *nature* 411, 826.

Darling, A. C. E. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14, 1394–1403. doi: 10.1101/gr.2289704.

Das, A., Rangaraj, N., and Sonti, R. V. (2009). Multiple adhesin-like functions of *Xanthomonas oryzae* pv. *oryzae* are involved in promoting leaf attachment, entry, and virulence on rice. *MPMI* 22, 73–85. doi: 10.1094/MPMI-22-1-0073.

De Cleene, M. (1989). Scanning electron microscopy of the establishment of compatible and incompatible *Xanthomonas campestris* pathovars on the leaf surface of Italian ryegrass and maize. *EPPO Bulletin* 19, 81–88.

de Jonge, R., and Thomma, B. P. H. J. (2009). Fungal LysM effectors: extinguishers of host immunity? *Trends in Microbiology* 17, 151–157. doi: 10.1016/j.tim.2009.01.002.

de la Fuente Cantó, C., and Vigouroux, Y. (2022). Evaluation of nine statistics to identify QTLs in bulk segregant analysis using next generation sequencing approaches. *BMC Genomics* 23, 490. doi: 10.1186/s12864-022-08718-y.

de Vallavieille-Pope, C. (2004). Management of disease resistance diversity of cultivars of a species in single fields: controlling epidemics. *Comptes Rendus Biologies* 327, 611–620. doi: 10.1016/j.crvi.2003.11.014.

Demirjian, C., Vailleau, F., Berthomé, R., and Roux, F. (2023). Genome-wide association studies in plant pathosystems: success or failure? *Trends in Plant Science* 28, 471–485. doi: 10.1016/j.tplants.2022.11.006.

Devos, K. M. (2005). Updating the 'Crop Circle.' *Current Opinion in Plant Biology* 8, 155–162. doi: 10.1016/j.pbi.2005.01.005.

Dormatey, R., Sun, C., Ali, K., Coulter, J. A., Bi, Z., and Bai, J. (2020). Gene pyramiding for sustainable crop improvement against biotic and abiotic stresses. *Agronomy* 10, 1255. doi: 10.3390/agronomy10091255.

Dow, J. M., Clarke, B. R., Milligan, D. E., Tang, J. L., and Daniels, M. J. (1990). Extracellular proteases from *Xanthomonas campestris* pv. *campestris*, the black rot pathogen. *Applied and Environmental Microbiology* 56, 2994–2998. doi: 10.1128/aem.56.10.2994-2998.1990.

Dracatos, P. M., Cogan, N. O., Sawbridge, T. I., Gendall, A. R., Smith, K. F., Spangenberg, G. C., et al. (2009). Molecular characterisation and genetic mapping of candidate genes for qualitative disease resistance in perennial ryegrass (*Lolium perenne* L.). *BMC Plant Biology* 9, 62. doi: 10.1186/1471-2229-9-62.

Dunger, G., Guzzo, C. R., Andrade, M. O., Jones, J. B., and Farah, C. S. (2014). *Xanthomonas citri* subsp. *citri* type iv pilus is required for twitching motility, biofilm development, and adherence. *MPMI* 27, 1132–1147. doi: 10.1094/MPMI-06-14-0184-R.

Dunger, G., Llontop, E., Guzzo, C. R., and Farah, C. S. (2016). The *Xanthomonas* type IV pilus. *Current Opinion in Microbiology* 30, 88–97. doi: 10.1016/j.mib.2016.01.007.

Duveiller, E., van Ginkel, M., and Thijssen, M. (1992). Genetic analysis of resistance to bacterial leaf streak caused by *Xanthomonas campestris* pv. *undulosa* in bread wheat. *Euphytica* 66, 35–43. doi: 10.1007/BF00023506.

Ebler, J., Ebert, P., Clarke, W. E., Rausch, T., Audano, P. A., Houwaart, T., et al. (2022). Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. *Nat Genet* 54, 518–525. doi: 10.1038/s41588-022-01043-w.

Ederli, L., Madeo, L., Calderini, O., Gehring, C., Moretti, C., Buonaurio, R., et al. (2011). The *Arabidopsis thaliana* cysteine-rich receptor-like kinase CRK20 modulates host responses to *Pseudomonas syringae* pv. *tomato* DC3000 infection. *Journal of Plant Physiology* 168, 1784–1794. doi: 10.1016/j.jplph.2011.05.018.

EFSA Panel on Genetically Modified Organisms (GMO), Mullins, E., Bresson, J., Dalmay, T., Dewhurst, I. C., Epstein, M. M., et al. (2021). In vivo and in vitro random mutagenesis techniques in plants. *EFS2* 19. doi: 10.2903/j.efsa.2021.6611.

Egli, T., Goto, M., and Schmidt, D. (1975). Bacterial wilt, a new forage grass disease. *Journal of Phytopathology* 82, 111–121. doi: 10.1111/j.1439-0434.1975.tb02825.x.

Egli, T., and Schmidt, D. (1982). Pathogenic variation among the causal agents of bacterial wilt of forage grasses. *Journal of Phytopathology* 104, 138–150. doi: 10.1111/j.1439-0434.1982.tb00520.x.

Emadodin, I., Corral, D. E. F., Reinsch, T., Kluß, C., and Taube, F. (2021). Climate change effects on temperate grassland and its implication for forage production: a case study from northern germany. *Agriculture* 11, 232. doi: 10.3390/agriculture11030232.

Erkes, A., Grove, R. P., Žarković, M., Krautwurst, S., Koebnik, R., Morgan, R. D., et al. (2023). Assembling highly repetitive *Xanthomonas* TALomes using Oxford Nanopore sequencing. *BMC Genomics* 24, 151. doi: 10.1186/s12864-023-09228-1.

Facelli, E., Taylor, C., Scott, E., Fegan, M., Huys, G., Noble, R. D., et al. (2005). Identification of the causal agent of pistachio dieback in Australia. *Eur J Plant Pathol* 112, 155–165. doi: 10.1007/s10658-005-3120-9.

Falahi Charkhabi, N., Booher, N. J., Peng, Z., Wang, L., Rahimian, H., Shams-Bakhsh, M., et al. (2017). Complete genome sequencing and targeted mutagenesis reveal virulence contributions of Tal2 and Tal4b of *Xanthomonas translucens* pv. *undulosa* ICMP11055 in bacterial leaf streak of wheat. *Front. Microbiol.* 8, 1488. doi: 10.3389/fmicb.2017.01488.

Falloon, R. E., and Hume, D. E. (1988). Productivity and persistence of prairie grass (*Bromus willdenowii Kunth*) 1. Effects of the head smut fungus Ustilago bullata Berk. *Grass and forage science* 43, 179–184.

Fan, E., Chauhan, N., Udatha, D. B. R. K. G., Leo, J. C., and Linke, D. (2016). Type V secretion systems in bacteria. *Microbiology Spectrum* 4, 4.1.10. doi: 10.1128/microbiolspec.VMBF-0009-2015.

FAO (2016). FAOSTAT. Available at: http://www.fao.org/faostat/en/?#data/EL/visualize [Accessed November 9, 2018].

Fazeli-Nasab, B., Sayyed, R. Z., Mojahed, L. S., Rahmani, A. F., Ghafari, M., Antonius, S., et al. (2022). Biofilm production: A strategic mechanism for survival of microbes under stress conditions. *Biocatalysis and Agricultural Biotechnology* 42, 102337. doi: 10.1016/j.bcab.2022.102337.

Feil, H., Feil, W. S., and Lindow, S. E. (2007). Contribution of fimbrial and afimbrial adhesins of *Xylella fastidiosa* to attachment to surfaces and virulence to grape. *Phytopathology*® 97, 318–324. doi: 10.1094/PHYTO-97-3-0318.

Felix, G., Duran, J. D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal* 18, 265–276.

Ferreira, R. M., de Oliveira, A. C. P., Moreira, L. M., Belasque, J., Gourbeyre, E., Siguier, P., et al. (2015). A TALE of Transposition: Tn3-Like Transposons Play a Major Role in the Spread of Pathogenicity Determinants of *Xanthomonas citri* and Other *Xanthomonads*. *mBio* 6, e02505-14. doi: 10.1128/mBio.02505-14.

Filiault, D. L., Seymour, D. K., Maruthachalam, R., and Maloof, J. N. (2017). The generation of doubled haploid lines for QTL mapping. *Methods Mol Biol* 1610, 39–57. doi: 10.1007/978-1-4939-7003-2_4.

Frame, J. (1991). Herbage production and quality of a range of secondary grass species at five rates of fertilizer nitrogen application. *Grass and Forage Science* 46, 139–151.

Gade, P., and Kalvakolanu, D. V. (2012). Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. *Methods Mol Biol* 809, 85–104. doi: 10.1007/978-1-61779-376-9_6.

Galán, J. E., Lara-Tejero, M., Marlovits, T. C., and Wagner, S. (2014). Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. *Annual Review of Microbiology* 68, 415–438. doi: 10.1146/annurev-micro-092412-155725.

Gangurde, S. S., Xavier, A., Naik, Y. D., Jha, U. C., Rangari, S. K., Kumar, R., et al. (2022). Two decades of association mapping: Insights on disease resistance in major crops. *Frontiers in Plant Science* 13.

Gao, S., Wang, F., Niran, J., Li, N., Yin, Y., Yu, C., et al. (2021). Transcriptome analysis reveals defense-related genes and pathways against *Xanthomonas campestris* pv. *vesicatoria* in pepper (*Capsicum annuum* L.). *PLOS ONE* 16, e0240279. doi: 10.1371/journal.pone.0240279.

Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., et al. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat Biotechnol* 36, 875–879. doi: 10.1038/nbt.4227.

Geng, X., Gao, Z., Zhao, L., Zhang, S., Wu, J., Yang, Q., et al. (2022). Comparative transcriptome analysis of resistant and susceptible wheat in response to *Rhizoctonia cerealis*. *BMC Plant Biology* 22, 235. doi: 10.1186/s12870-022-03584-y.

Gerlach, R. G., and Hensel, M. (2007). Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. *International Journal of Medical Microbiology* 297, 401–415. doi: 10.1016/j.ijmm.2007.03.017.

Ghesquière, M., Humphreys, M. W., and Zwierzykowski, Z. (2010). "Festulolium," in *Fodder Crops and Amenity Grasses* Handbook of Plant Breeding., eds. B. Boller, U. K. Posselt, and F. Veronesi (New York, NY: Springer), 288–311. doi: 10.1007/978-1-4419-0760-8_12.

Giblot-Ducray, D., Marefat, A., Gillings, M. R., Parkinson, N. M., Bowman, J. P., Ophel-Keller, K., et al. (2009). Proposal of *Xanthomonas translucens* pv. *pistaciae* pv. nov., pathogenic to pistachio (*Pistacia vera*). *Systematic and Applied Microbiology* 32, 549–557. doi: 10.1016/j.syapm.2009.08.001.

Gilchrist, C. L. M., and Chooi, Y.-H. (2021). clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475. doi: 10.1093/bioinformatics/btab007.

Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science* 298, 2345–2349. doi: 10.1126/science.1076641.

Gochez, A. M., Huguet-Tapia, J. C., Minsavage, G. V., Shantaraj, D., Jalan, N., Strauß, A., et al. (2018). Pacbio sequencing of copper-tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of transcription activator-like effectors among *X. citri* strains. *BMC Genomics* 19, 16. doi: 10.1186/s12864-017-4408-9.

Goettelmann, F., Copetti, D., Yates, S., Studer, B., and Kölliker, R. (2021). Pooled sequencing reveals multiple regions involved in resistance to bacterial wilt in Italian ryegrass. in, eds. S. Hartmann, S. Bachmann-Pfabe, U. Feuerstein, B. Julier, R. Kölliker, D. Kopecký, et al. (Palacký University), 62. doi: 10.5507/vup.21.24459677.14.

Goettelmann, F., Roman-Reyna, V., Cunnac, S., Jacobs, J. M., Bragard, C., Studer, B., et al. (2022). Complete genome assemblies of all *Xanthomonas translucens* pathotype strains reveal three genetically distinct clades. *Frontiers in Microbiology* 12.

Gómez-Gómez, L., and Boller, T. (2000). FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Molecular cell* 5, 1003–1011.

Gondran, J., and Betin, M. (1989). Sensibilité de cultivars de raygrass ďltalie à *Xanthomonas campestris* pv. *graminis* . *EPPO Bulletin* 19, 149–155. doi: 10.1111/j.1365-2338.1989.tb00142.x.

Gottig, N., Garavaglia, B. S., Garofalo, C. G., Orellano, E. G., and Ottado, J. (2009). A filamentous hemagglutinin-like protein of *Xanthomonas axonopodis* pv. *citri*, the phytopathogen responsible for citrus canker, is involved in bacterial virulence. *PLOS ONE* 4, e4358. doi: 10.1371/journal.pone.0004358.

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C., et al. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, gkad326. doi: 10.1093/nar/gkad326.

Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M., and Dangl, J. L. (2006). Subterfuge and manipulation: Type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* 60, 425–449. doi: 10.1146/annurev.micro.60.080805.142251.

Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., et al. (2016). AnnoTALE: bioinformatics tools for identification, annotation and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci Rep* 6, 21077. doi: 10.1038/srep21077.

Gray, Y. H. M. (2000). It takes two transposons to tango:transposable-element-mediated chromosomal rearrangements. *Trends in Genetics* 16, 461–468. doi: 10.1016/S0168-9525(00)02104-1.

Grogg, D., Rohner, M., Yates, S., Manzanares, C., Bull, S. E., Dalton, S., et al. (2022). Callus induction from diverse explants and genotypes enables robust transformation of perennial ryegrass (*Lolium perenne* L.). *Plants* 11, 2054. doi: 10.3390/plants11152054.

Guérin, J., Bigot, S., Schneider, R., Buchanan, S. K., and Jacob-Dubuisson, F. (2017). Two-partner secretion: combining efficiency and simplicity in the secretion of large proteins for bacteria-host and bacteria-bacteria interactions. *Frontiers in Cellular and Infection Microbiology* 7.

Gurumurthy, C. B., Grati, M., Ohtsuka, M., Schilit, S. L. P., Quadros, R. M., and Liu, X. Z. (2016). CRISPR: a versatile tool for both forward and reverse genetics research. *Hum Genet* 135, 971–976. doi: 10.1007/s00439-016-1704-4.

Guttman, D. S., Vinatzer, B. A., Sarkar, S. F., Ranall, M. V., Kettler, G., and Greenberg, J. T. (2002). A functional screen for the type III (hrp) secretome of the plant pathogen *Pseudomonas syringae*. *Science* 295, 1722–1726. doi: 10.1126/science.295.5560.1722.

Gyawali, S., Harrington, M., Durkin, J., Horner, K., Parkin, I. A. P., Hegedus, D. D., et al. (2016). Microsatellite markers used for genome-wide association mapping of partial resistance to Sclerotinia sclerotiorum in a world collection of Brassica napus. *Mol Breeding* 36, 72. doi: 10.1007/s11032-016-0496-5.

Hayward, M. D., Forster, J. W., Jones, J. G., Dolstra, O., Evans, C., McAdam, N. J., et al. (1998). Genetic analysis of Lolium. I. Identification of linkage groups and the establishment of a genetic map. *Plant breeding* 117, 451–455.

He, C., Holme, J., and Anthony, J. (2014). SNP genotyping: the KASP assay. *Methods Mol Biol* 1145, 75–86. doi: 10.1007/978-1-4939-0446-4_7.

He, P., Shan, L., Lin, N.-C., Martin, G. B., Kemmerling, B., Nürnberger, T., et al. (2006). Specific bacterial suppressors of MAMP signaling ipstream of MAPKKK in *Arabidopsis* Innate Immunity. *Cell* 125, 563–575. doi: 10.1016/j.cell.2006.02.047.

He, Y.-W., Cao, X.-Q., and Poplawsky, A. R. (2020). Chemical structure, biological roles, biosynthesis and regulation of the yellow xanthomonadin pigments in the phytopathogenic genus *Xanthomonas*. *MPMI* 33, 705–714. doi: 10.1094/MPMI-11-19-0326-CR.

Heiden, N., Roman-Reyna, V., Curland, R. D., Dill-Macky, R., and Jacobs, J. (2022). Comparative genomics of barley-infecting *Xanthomonas translucens* shows overall genetic similarity but globally-distributed virulence factor diversity. *Phytopathology*®.

Hersemann, L. (2016). Characterization of virulence factors in *Xanthomonas translucens* pv. *graminis*. Available at: https://www.zora.uzh.ch/id/eprint/129341/ [Accessed August 23, 2019].

Hersemann, L., Wibberg, D., Blom, J., Goesmann, A., Widmer, F., Vorhölter, F.-J., et al. (2017). Comparative genomics of host adaptive traits in *Xanthomonas translucens* pv. *graminis*. *BMC Genomics* 18, 35. doi: 10.1186/s12864-016-3422-7.

Hersemann, L., Wibberg, D., Widmer, F., Vorhölter, F.-J., and Kölliker, R. (2016). Draft genome sequences of three *Xanthomonas translucens* pathovar reference strains (pv. *arrhenatheri*, pv. *poae* and pv. *phlei*) with different specificities for forage grasses. *Standards in Genomic Sciences* 11. doi: 10.1186/s40793-016-0170-x.

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorsi, B., Su, Y.-C., and Yost, H. J. (2013). MMAPPR: Mutation Mapping Analysis Pipeline for Pooled RNA-seq. *Genome Res.* 23, 687–697. doi: 10.1101/gr.146936.112.

Hirata, M., Cai, H., Inoue, M., Yuyama, N., Miura, Y., Komatsu, T., et al. (2006). Development of simple sequence repeat (SSR) markers and construction of an SSR-based linkage map in Italian ryegrass (*Lolium multiflorum* Lam.). *Theoretical and Applied Genetics* 113, 270–279. doi: 10.1007/s00122-006-0292-4.

Hoeffner, K., Beylich, A., Chabbi, A., Cluzeau, D., Dascalu, D., Graefe, U., et al. (2021). Legacy effects of temporary grassland in annual crop rotation on soil ecosystem services. *Science of The Total Environment* 780, 146140. doi: 10.1016/j.scitotenv.2021.146140.

Hogenhout, S. A., Van der Hoorn, R. A. L., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *MPMI* 22, 115–122. doi: 10.1094/MPMI-22-2-0115.

Holdsworth, W. L., and Mazourek, M. (2015). Development of user-friendly markers for the pvr1 and Bs3 disease resistance genes in pepper. *Mol Breeding* 35, 28. doi: 10.1007/s11032-015-0260-2.

Hopkins, A., and Del Prado, A. (2007). Implications of climate change for grassland in Europe: impacts, adaptations and mitigation options: a review. *Grass and Forage Science* 62, 118–126. doi: 10.1111/j.1365-2494.2007.00575.x.

Hu, J., Qian, W., and He, C. (2007). The *Xanthomonas oryzae* pv. *oryzae* eglXoB endoglucanase gene is required for virulence to rice. *FEMS Microbiology Letters* 269, 273–279. doi: 10.1111/j.1574-6968.2007.00638.x.

Huang, L., Tang, W., and Wu, W. (2022). Optimization of BSA-seq experiment for QTL mapping. *G3 Genes|Genomes|Genetics* 12, jkab370. doi: 10.1093/g3journal/jkab370.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47, D309–D314. doi: 10.1093/nar/gky1085.

Humphreys, M., Feuerstein, U., Vandewalle, M., and Baert, J. (2010). "Ryegrasses," in *Fodder Crops and Amenity Grasses* (New York, NY: Springer New York), 211–260. doi: 10.1007/978-1-4419-0760-8_10.

Hunjan, M. S., and Lore, J. S. (2020). “Climate change: impact on plant pathogens, diseases, and their management,” in *Crop Protection Under Changing Climate*, eds. K. Jabran, S. Florentine, and B. S. Chauhan (Cham: Springer International Publishing), 85–100. doi: 10.1007/978-3-030-46111-9_4.

Imam, S., Chen, Z., Roos, D. S., and Pohlschröder, M. (2011). Identification of surprisingly diverse type IV pili, across a broad range of Gram-positive bacteria. *PLoS ONE* 6, e28919. doi: 10.1371/journal.pone.0028919.

Inami, K., Yoshioka-Akiyama, C., Morita, Y., Yamasaki, M., Teraoka, T., and Arie, T. (2012). A genetic mechanism for emergence of races in *Fusarium oxysporum* f. sp. *lycopersici*: inactivation of avirulence gene AVR1 by transposon insertion. *PLOS ONE* 7, e44101. doi: 10.1371/journal.pone.0044101.

Jacques, M.-A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., et al. (2016). Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*. *Annu. Rev. Phytopathol.* 54, 163–187. doi: 10.1146/annurev-phyto-080615-100147.

Jaenicke, S., Bunk, B., Wibberg, D., Spröer, C., Hersemann, L., Blom, J., et al. (2016). Complete genome sequence of the barley pathogen *Xanthomonas translucens* pv. *translucens* DSM 18974 ^{T} (ATCC 19319 ^{T}). *Genome Announc* 4, e01334-16. doi: 10.1128/genomeA.01334-16.

Jaškūnė, K., Aleliūnas, A., Statkevičiūtė, G., Kemešytė, V., Studer, B., and Yates, S. (2020). Genome-wide association study to identify candidate loci for biomass formation under water deficit in perennial ryegrass. *Frontiers in Plant Science* 11.

Jha, G., Rajeshwari, R., and Sonti, R. V. (2005). Bacterial type two secretion system secreted proteins: double-edged swords for plant pathogens. *MPMI* 18, 891–898. doi: 10.1094/MPMI-18-0891.

Ji, C., Ji, Z., Liu, B., Cheng, H., Liu, H., Liu, S., et al. (2020). Xa1 allelic R genes activate rice blight resistance suppressed by interfering TAL effectors. *Plant Communications* 1, 100087. doi: 10.1016/j.xplc.2020.100087.

Ji, Z., Guo, W., Chen, X., Wang, C., and Zhao, K. (2022). Plant executor genes. *Int J Mol Sci* 23, 1524. doi: 10.3390/ijms23031524.

Jia, H., Zhang, Y., Orbović, V., Xu, J., White, F. F., Jones, J. B., et al. (2017). Genome editing of the disease susceptibility gene CsLOB1 in citrus confers resistance to citrus canker. *Plant Biotechnology Journal* 15, 817–823. doi: 10.1111/pbi.12677.

Jiang, D., Zhang, D., Li, S., Liang, Y., Zhang, Q., Qin, X., et al. (2022). Highly efficient genome editing in *Xanthomonas oryzae* pv. *oryzae* through repurposing the endogenous type I-C CRISPR-Cas system. *Molecular Plant Pathology* 23, 583–594. doi: 10.1111/mpp.13178.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829.

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031.

Joshi, J. B., Arul, L., Ramalingam, J., and Uthandi, S. (2020). Advances in the *Xoo*-rice pathosystem interaction and its exploitation in disease management. *J Biosci* 45, 112. doi: 10.1007/s12038-020-00085-8.

Jung, Y.-H., Lee, J.-H., Agrawal, G. K., Rakwal, R., Kim, J.-A., Shim, J.-K., et al. (2005). The rice (*Oryza sativa*) blast lesion mimic mutant, blm, may confer resistance to blast pathogens by triggering multiple defense-associated signaling pathways. *Plant Physiology and Biochemistry* 43, 397–406. doi: 10.1016/j.plaphy.2005.03.002.

Kandel, Y. R., Glover, K. D., Tande, C. A., and Osborne, L. E. (2012). Evaluation of spring wheat germplasm for resistance to bacterial leaf streak caused by *Xanthomonas campestris* pv. *translucens*. *Plant Disease* 96, 1743–1748. doi: 10.1094/PDIS-03-12-0303-RE.

Kang, S., Lebrun, M. H., Farrall, L., and Valent, B. (2001). Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. *MPMI* 14, 671–674. doi: 10.1094/MPMI.2001.14.5.671.

Katzen, F., Ferreiro, D. U., Oddo, C. G., Ielmini, M. V., Becker, A., Pühler, A., et al. (1998). Xanthomonas campestris pv. campestrisgum Mutants: Effects on Xanthan Biosynthesis and Plant Virulence. *Journal of Bacteriology* 180, 1607–1617. doi: 10.1128/JB.180.7.1607-1617.1998.

Kay, S., and Bonas, U. (2009). How *Xanthomonas* type III effectors manipulate the host plant. *Current Opinion in Microbiology* 12, 37–43. doi: 10.1016/j.mib.2008.12.006.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648–651. doi: 10.1126/science.1144956.

Kemp, B. P., Horne, J., Bryant, A., and Cooper, R. M. (2004). Xanthomonas axonopodis pv. manihotis gumD gene is essential for EPS production and pathogenicity and enhances epiphytic survival on cassava (Manihot esculenta). *Physiological and Molecular Plant Pathology* 64, 209–218. doi: 10.1016/j.pmpp.2004.08.007.

Kerppola, T. K. (2006). Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat Protoc* 1, 1278–1286. doi: 10.1038/nprot.2006.201.

Khatib, M., Lafitte, C., Esquerré-Tugayé, M.-T., Bottin, A., and Rickauer, M. (2004). The CBEL elicitor of *Phytophthora parasitica* var. *nicotianae* activates defence in *Arabidopsis thaliana* via three different signalling pathways. *New Phytologist* 162, 501–510. doi: 10.1111/j.1469-8137.2004.01043.x.

Kim, N. H., and Hwang, B. K. (2015a). Pepper aldehyde dehydrogenase CaALDH1 interacts with *Xanthomonas* effector AvrBsT and promotes effector-triggered cell death and defence responses. *Journal of Experimental Botany* 66, 3367–3380. doi: 10.1093/jxb/erv147.

Kim, N. H., and Hwang, B. K. (2015b). Pepper heat shock protein 70a interacts with the type iii effector AvrBsT and triggers plant cell death and immunity. *Plant Physiology* 167, 307–322. doi: 10.1104/pp.114.253898.

King, J., Armstead, I., Harper, J., Ramsey, L., Snape, J., Waugh, R., et al. (2013). Exploitation of interspecific diversity for monocot crop improvement. *Heredity* 110, 475–483. doi: 10.1038/hdy.2012.116.

Klassen, J. L., and Currie, C. R. (2012). Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. *BMC Genomics* 13, 14. doi: 10.1186/1471-2164-13-14.

Knaus, B. J., and Grünwald, N. J. (2017). VCFR : a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17, 44–53. doi: 10.1111/1755-0998.12549.

Knorst, V., Byrne, S., Yates, S., Asp, T., Widmer, F., Studer, B., et al. (2018a). Pooled DNA sequencing to identify SNPs associated with a major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). *Theoretical and Applied Genetics* 132, 947–958. doi: 10.1007/s00122-018-3250-z.

Knorst, V., Yates, S., Byrne, S., Asp, T., Widmer, F., Studer, B., et al. (2018b). First assembly of the gene-space of *Lolium multiflorum* and comparison to other Poaceae genomes. *Grassland Science* 65, 125–134. doi: 10.1111/grs.12225.

Koebnik, R., Burokiene, D., Bragard, C., Chang, C., Saux, M. F.-L., Kölliker, R., et al. (2021). The complete genome sequence of *Xanthomonas theicola*, the causal agent of canker on tea plants, reveals novel secretion systems in clade-1 xanthomonads. *Phytopathology*® 111, 611–616. doi: 10.1094/PHYTO-07-20-0273-SC.

Kölliker, R., Frey, L. A., Goettelmann, F., and Studer, B. (2022). Unravelling the genetic control of disease resistance in outbreeding forage crop species. doi: 10.3929/ethz-b-000552829.

Kölliker, R., Kraehenbuehl, R., Boller, B., and Widmer, F. (2006). Genetic diversity and pathogenicity of the grass pathogen *Xanthomonas translucens* pv. *graminis*. *Systematic and Applied Microbiology* 29, 109–119. doi: 10.1016/j.syapm.2005.07.004.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37, 540–546. doi: 10.1038/s41587-019-0072-8.

Kopecký, D., Havránková, M., Loureiro, J., Castro, S., Lukaszewski, A. J., Bartoš, J., et al. (2010). Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. *Cytogenet Genome Res* 129, 162–172. doi: 10.1159/000313379.

Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., et al. (2018). De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* 36, 1174–1182. doi: 10.1038/nbt.4277.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116.

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 29. doi: 10.1186/1746-4811-9-29.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor tu elicits innate immunity in Arabidopsis plants. *The Plant Cell* 16, 3496–3507. doi: 10.1105/tpc.104.026765.

Kuo, T., Lin, B., and Li, C. (1970). Bacterial leaf blight of rice plant. III. Phytotoxic polysaccharides produced by *Xanthomonas oryzae*. *Botanical Bulletin Academia Sinica* 11, 46–54.

Labruyere, R. (1975). *Xanthomonas graminis* bij grassen. *Jaarverslag Instituut voor Plantenziektenkundig Onderzoek 1975*, 24.

Laluk, K., and Mengiste, T. (2011). The *Arabidopsis* extracellular UNUSUAL SERINE PROTEASE INHIBITOR functions in resistance to necrotrophic fungi and insect herbivory: *Arabidopsis* UPI in resistance to necrotrophs and herbivory. *The Plant Journal* 68, 480–494. doi: 10.1111/j.1365-313X.2011.04702.x.

Lamichhane, J. R., Osdaghi, E., Behlau, F., Köhl, J., Jones, J. B., and Aubertot, J.-N. (2018). Thirteen decades of antimicrobial copper compounds applied in agriculture. A review. *Agron. Sustain. Dev.* 38, 28. doi: 10.1007/s13593-018-0503-9.

Langlois, P. A., Snelling, J., Hamilton, J. P., Bragard, C., Koebnik, R., Verdier, V., et al. (2017). Characterization of the *Xanthomonas translucens* complex using draft genomes, comparative genomics, phylogenetic analysis, and diagnostic lamp assays. *Phytopathology*® 107, 519–527. doi: 10.1094/PHYTO-08-16-0286-R.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. doi: 10.1038/nmeth.1923.

Leyns, F. (1993). "Xanthomonas campestris pv. graminis: cause of bacterial wilt of forage grasses," in *Xanthomonas* (Edited by J. G. Swings & E. L. Civerolo. London: Chapman & Hall.), p55-57.

Leyns, F., Cleene, M. D., Bogaert, G. V., Van De Vijver, L., and Ley, J. D. (1987). Some physiological aspects of bacterial wilt of forage grasses caused by *Xanthomonas campestris* pv. *graminis*. *Journal of Phytopathology* 120, 130–135.

Leyns, F., De Cleene, M., Swings, J.-G., and De Ley, J. (1984). The host range of the genus *Xanthomonas*. *The Botanical Review* 50, 308–356. doi: 10.1007/BF02862635.

Leyns, F., De Cleene, M., Van Bogaert, G., Van de Wiele, A., and De Ley, J. (1988). Preliminary investigations about the mode of transmission and spread of *Xanthomonas campestris* pv. *graminis* on forage grasses. *Journal of Phytopathology* 122, 76–88.

Leyns, F., Van Den Mooter, M., Swings, J., De Cleene, M., and De Ley, J. (1981). Distribution of *Xanthomonas campestris* pv. *graminis* in fields of forage grasses in northern Belgium. *Parasitica* 37, 131–133.

Li, J., and Wang, N. (2011). The wxacO gene of *Xanthomonas citri* ssp. *citri* encodes a protein with a role in lipopolysaccharide biosynthesis, biofilm formation, stress tolerance and virulence. *Molecular Plant Pathology* 12, 381–396. doi: 10.1111/j.1364-3703.2010.00681.x.

Li, J., Yao, Y., Xu, H. H., Hao, L., Deng, Z., Rajakumar, K., et al. (2015). SecReT6: a web-based resource for type VI secretion systems found in bacteria: Genomics update. *Environ Microbiol* 17, 2196–2202. doi: 10.1111/1462-2920.12794.

Li, L., Peng, Z., Mao, X., Wang, J., Chang, X., Reynolds, M., et al. (2019). Genome-wide association study reveals genomic regions controlling root and shoot traits at late growth stages in wheat. *Ann Bot* 124, 993–1006. doi: 10.1093/aob/mcz041.

Li, Z., and Xu, Y. (2022). Bulk segregation analysis in the NGS era: a review of its teenage years. *The Plant Journal* 109, 1355–1374. doi: 10.1111/tpj.15646.

Liao, W.-W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., et al. (2023). A draft human pangenome reference. *Nature* 617, 312–324. doi: 10.1038/s41586-023-05896-x.

Lin, J.-S., and Lai, E.-M. (2017). "Protein–protein interactions: co-immunoprecipitation," in *Bacterial Protein Secretion Systems: Methods and Protocols* Methods in Molecular Biology., eds. L. Journet and E. Cascales (New York, NY: Springer), 211–219. doi: 10.1007/978-1-4939-7033-9_17.

Lin, R.-H., Peng, C.-W., Lin, Y.-C., and Peng, H.-L. (2011). The XopE2 effector protein of *Xanthomonas campestris* pv. *vesicatoria* is involved in virulence and in the suppression of the hypersensitive response. *Botanical Studies* 52.

Lin, Y., Liao, Y.-Y., Huang, R.-X., Li, A.-Z., An, S.-Q., Tang, J.-L., et al. (2021). Extracellular amylase is required for full virulence and regulated by the global posttranscriptional regulator rsma in *Xanthomonas campestris* pathovar *campestris*. *Phytopathology*® 111, 1104–1113. doi: 10.1094/PHYTO-08-20-0372-R.

Lin, Z., Zhong, S., and Grierson, D. (2009). Recent advances in ethylene research. *Journal of Experimental Botany* 60, 3311–3336. doi: 10.1093/jxb/erp204.

Lu, J., Li, Q., Wang, C., Wang, M., Zeng, D., Zhang, F., et al. (2022a). Identification of quantitative trait loci associated with resistance to *Xanthomonas oryzae* pv. *oryzae* pathotypes prevalent in South China. *The Crop Journal* 10, 498–507. doi: 10.1016/j.cj.2021.05.009.

Lu, Y., Zhong, Q., Xiao, S., Wang, B., Ke, X., Zhang, Y., et al. (2022b). A new NLR disease resistance gene *Xa47* confers durable and broad-spectrum resistance to bacterial blight in rice. *Frontiers in Plant Science* 13.

Lyons, R., Rusu, A., Stiller, J., Powell, J., Manners, J. M., and Kazan, K. (2015). Investigating the association between flowering time and defense in the *Arabidopsis thaliana-Fusarium oxysporum interaction*. *PLOS ONE* 10, e0127699. doi: 10.1371/journal.pone.0127699.

Ma, J. F., and Yamaji, N. (2006). Silicon uptake and accumulation in higher plants. *Trends in Plant Science* 11, 392–397. doi: 10.1016/j.tplants.2006.06.007.

Macho, A. P., and Zipfel, C. (2015). Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. *Current Opinion in Microbiology* 23, 14–22. doi: 10.1016/j.mib.2014.10.009.

Magwene, P. M., Willis, J. H., and Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput Biol* 7, e1002255. doi: 10.1371/journal.pcbi.1002255.

Manosalva, P. M., Davidson, R. M., Liu, B., Zhu, X., Hulbert, S. H., Leung, H., et al. (2009). A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiology* 149, 286–296. doi: 10.1104/pp.108.128348.

Mansfeld, B. N., and Grumet, R. (2018). QTLseqr: an R package for bulk segregant analysis with next-generation sequencing. *The Plant Genome* 11, 180006. doi: 10.3835/plantgenome2018.01.0006.

Manzanares, C., Yates, S., Ruckle, M., Nay, M., and Studer, B. (2016). TILLING in forage grasses for gene discovery and breeding improvement. *New Biotechnology* 33, 594–603. doi: 10.1016/j.nbt.2016.02.009.

Marefat, A., Scott, E. S., Ophel-Keller, K., and Sedgley, M. (2006). Genetic, phenotypic and pathogenic diversity among xanthomonads isolated from pistachio (*Pistacia vera*) in Australia. *Plant Pathology* 55, 639–649. doi: 10.1111/j.1365-3059.2006.01437.x.

Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Frary, A., and Ganal, M. W. (1993). Conferring disease resistance in tomato.

Martin, R. C., Glover-Cutter, K., Martin, R. R., and Dombrowski, J. E. (2013). Virus induced gene silencing in *Lolium temulentum*. *Plant Cell Tiss Organ Cult* 113, 163–171. doi: 10.1007/s11240-012-0257-z.

Martiniano, R., Garrison, E., Jones, E. R., Manica, A., and Durbin, R. (2020). Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. *Genome Biology* 21, 250. doi: 10.1186/s13059-020-02160-7.

Masuch, G., Schoene, K., and Paul, V. H. (1989). Histological investigations on the pathogenesis of *Xanthomonas campestris* pv. *graminis* to *Lolium multiflorum* . *EPPO Bulletin* 19, 73–80. doi: 10.1111/j.1365-2338.1989.tb00131.x.

McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol* 123, 439–442. doi: 10.1104/pp.123.2.439.

Mehrab, Z., Mobin, J., Tahmid, I. A., Pachter, L., and Rahman, A. (2020). Reference-free association mapping from sequencing reads using k-mers. *Bio Protoc* 10, e3815. doi: 10.21769/BioProtoc.3815.

Meng, H., Sun, M., Jiang, Z., Liu, Y., Sun, Y., Liu, D., et al. (2021). Comparative transcriptome analysis reveals resistant and susceptible genes in tobacco cultivars in response to infection by *Phytophthora nicotianae*. *Sci Rep* 11, 809. doi: 10.1038/s41598-020-80280-7.

Meng, X., and Zhang, S. (2013). MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology* 51, 245–266. doi: 10.1146/annurev-phyto-082712-102314.

Metz, M., Dahlbeck, D., Morales, C. Q., Sady, B. A., Clark, E. T., and Staskawicz, B. J. (2005). The conserved *Xanthomonas campestris* pv. *vesicatoria* effector protein XopX is a virulence factor and suppresses host defense in *Nicotiana benthamiana*. *The Plant Journal* 41, 801–814. doi: 10.1111/j.1365-313X.2005.02338.x.

Mew, T. W. (1992). Changes in race frequency of *Xanthomonas oryzae* pv. *oryzae* in response to rice cultivars planted in the Philippines. *Plant Dis.* 76, 1029. doi: 10.1094/PD-76-1029.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308.

Mhedbi-Hajri, N., Jacques, M.-A., and Koebnik, R. (2011). "Adhesion mechanisms of plant-pathogenic Xanthomonadaceae," in *Bacterial Adhesion: Chemistry, Biology and Physics* Advances in Experimental Medicine and Biology., eds. D. Linke and A. Goldman (Dordrecht: Springer Netherlands), 71–89. doi: 10.1007/978-94-007-0940-9_5.

Michel, V., Schori, A., Mosimann, E., Boller, B., and Schubiger, F. (2000). Maladies des graminées et des légumineuses fourragères. *Revue suisse d'agriculture* 32.

Michel, V. V. (2001). Interactions between *Xanthomonas campestris* pv. *graminis* strains and meadow fescue and italian rye grass cultivars. *Plant Disease* 85, 538–542. doi: 10.1094/PDIS.2001.85.5.538.

Michelmore, R. W., and Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8, 1113–1130. doi: 10.1101/gr.8.11.1113.

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the national academy of sciences* 88, 9828–9832.

Miedaner, T., Boeven, A. L. G.-C., Gaikpa, D. S., Kistner, M. B., and Grote, C. P. (2020). Genomics-assisted breeding for quantitative disease resistances in small-grain cereals and maize. *International Journal of Molecular Sciences* 21, 9717. doi: 10.3390/ijms21249717.

Migicovsky, Z., and Myles, S. (2017). Exploiting wild relatives for genomics-assisted breeding of perennial crops. *Frontiers in Plant Science* 8.

Mishina, T. E., and Zeier, J. (2007). Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in Arabidopsis. *The Plant Journal* 50, 500–513. doi: 10.1111/j.1365-313X.2007.03067.x.

Miura, Y., Hirata, M., and Fujimori, M. (2007). Mapping of EST-derived CAPS markers in Italian ryegrass (Lolium multiflorum Lam.). *Plant Breeding* 126, 353–360. doi: 10.1111/j.1439-0523.2007.01386.x.

Monnot, S., Desaint, H., Mary-Huard, T., Moreau, L., Schurdi-Levraud, V., and Boissot, N. (2021). Deciphering the genetic architecture of plant virus resistance by GWAS, state of the art and potential advances. *Cells* 10, 3080. doi: 10.3390/cells10113080.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.* 9, 141–142. doi: 10.1137/1109020.

Naito, K., Ishiga, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2007). N-terminal domain including conserved flg22 is required for flagellin-induced hypersensitive cell death in *Arabidopsis thaliana*. *J Gen Plant Pathol* 73, 281–285. doi: 10.1007/s10327-007-0017-9.

Nakai, H., Nakamura, K., Kuwahara, S., and Saito, M. (1990). A new gene, developed through mutagenesis, for resistance of rice to bacterial leaf blight (*Xanthomonas campestris* pv. *oryzae*). *J. Agric. Sci.* 114, 219–224. doi: 10.1017/S0021859600072233.

Nay, M., Tanner, P., and Christoph Grieder (2022). Bastardraigras vereint die besten Eigenschaften von Englischem und Italienischem Raigras. doi: 10.34776/AFS13-151.

Neves, C. J., Matzrafi, M., Thiele, M., Lorant, A., Mesgaran, M. B., and Stetter, M. G. (2020). Male linked genomic region determines sex in dioecious *Amaranthus palmeri*. *Journal of Heredity* 111, 606–612. doi: 10.1093/jhered/esaa047.

Ngou, B. P. M., Ding, P., and Jones, J. D. G. (2022). Thirty years of resistance: Zig-zag through the plant immune system. *The Plant Cell* 34, 1447–1478. doi: 10.1093/plcell/koac041.

Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., et al. (2021). Novel CRISPR–Cas systems: an updated review of the current achievements, applications, and future research perspectives. *Int J Mol Sci* 22, 3327. doi: 10.3390/ijms22073327.

Nimchuk, Z. L., Fisher, E. J., Desveaux, D., Chang, J. H., and Dangl, J. L. (2007). The HopX (AvrPphE) family of *Pseudomonas syringae* type III effectors require a catalytic triad and a novel N-terminal domain for function. *MPMI* 20, 346–357. doi: 10.1094/MPMI-20-4-0346.

Noël, L., Thieme, F., Nennstiel, D., and Bonas, U. (2002). Two novel type III-secreted proteins of *Xanthomonas campestris* pv. *vesicatoria* are encoded within the hrp pathogenicity island. *J Bacteriol* 184, 1340–1348. doi: 10.1128/JB.184.5.1340-1348.2002.

Nyfeler, D., Huguenin-Elie, O., Suter, M., Frossard, E., and Lüscher, A. (2011). Grass–legume mixtures can yield more nitrogen than legume pure stands due to mutual stimulation of nitrogen uptake from symbiotic and non-symbiotic sources. *Agriculture, Ecosystems & Environment* 140, 155–163. doi: 10.1016/j.agee.2010.11.022.

Ochiai, H., Inoue, Y., Takeya, M., Sasaki, A., and Kaku, H. (2005). Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. *JARQ* 39, 275–287. doi: 10.6090/jarq.39.275.

OECD and FAO (2018). *OECD-FAO Agricultural Outlook 2018-2027*. OECD doi: 10.1787/agr_outlook-2018-en.

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J. C., Perez-Quintero, A., Li, T., et al. (2019). Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nat Biotechnol* 37, 1344–1350. doi: 10.1038/s41587-019-0267-z.

Omboki, R. B., Zheng, Y., Chen, Z., Guan, H., Tang, W., Huang, L., et al. (2018). Pooled mapping of quantitative trait loci conferring male sterility-conditioned glume split in rice (*Oryza sativa*). *Plant Breeding* 137, 848–856. doi: 10.1111/pbr.12643.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421.

Pajerowska-Mukhtar, K. M., Emerine, D. K., and Mukhtar, M. S. (2013). Tell me more: roles of NPRs in plant immunity. *Trends in Plant Science* 18, 402–411. doi: 10.1016/j.tplants.2013.04.004.

Palmer, L. E., and McCombie, W. R. (2002). On the importance of being finished. *Genome Biology* 3, comment2010.1. doi: 10.1186/gb-2002-3-10-comment2010.

Pandurangan, S., Workman, C., Nilsen, K., and Kumar, S. (2022). “Introduction to marker-assisted selection in wheat Breeding,” in *Accelerated Breeding of Cereal Crops* Springer Protocols Handbooks., eds. A. Bilichak and J. D. Laurie (New York, NY: Springer US), 77–117. doi: 10.1007/978-1-0716-1526-3_3.

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J.-P., et al. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnology Journal* 13, 565–577. doi: 10.1111/pbi.12282.

Paul, V. H., and Freudenstein, K. H. (1989). Susceptibility of cultivars of Italian ryegrass to *Xanthomonas campestris* pv. *graminis*. *EPPO Bulletin* 19, 165–171. doi: 10.1111/j.1365-2338.1989.tb00145.x.

Pekkarinen, A. I., Longstaff, C., and Jones, B. L. (2007). Kinetics of the inhibition of *Fusarium* serine proteinases by barley (*Hordeum vulgare* L.) inhibitors. *J. Agric. Food Chem.* 55, 2736–2742. doi: 10.1021/jf0631777.

Peng, Z., Hu, Y., Xie, J., Potnis, N., Akhunova, A., Jones, J., et al. (2016). Long read and single molecule DNA sequencing simplifies genome assembly and TAL effector gene analysis of *Xanthomonas translucens*. *BMC Genomics* 17, 21. doi: 10.1186/s12864-015-2348-9.

Peng, Z., Hu, Y., Zhang, J., Huguet-Tapia, J. C., Block, A. K., Park, S., et al. (2019). *Xanthomonas translucens* commandeers the host rate-limiting step in ABA biosynthesis for disease susceptibility. *Proc Natl Acad Sci USA* 116, 20938–20946. doi: 10.1073/pnas.1911660116.

Pesce, C., Bolot, S., Cunnac, S., Portier, P., Fischer-Le Saux, M., Jacques, M.-A., et al. (2015). High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain cfbp 2541. *Genome Announc* 3, e01574-14. doi: 10.1128/genomeA.01574-14.

Pesce, C., Jacobs, J. M., Berthelot, E., Perret, M., Vancheva, T., Bragard, C., et al. (2017). Comparative genomics identifies a novel conserved protein, hpaT, in proteobacterial type III secretion systems that do not possess the putative translocon protein hrpF. *Front. Microbiol.* 8, 1177. doi: 10.3389/fmicb.2017.01177.

Petrocelli, S., Tondo, M. L., Daurelio, L. D., and Orellano, E. G. (2012). Modifications of </i>Xanthomonas axonopodis</i> pv. </i>citri</i> lipopolysaccharide affect the basal response and the virulence process during citrus canker. *PLOS ONE* 7, e40051. doi: 10.1371/journal.pone.0040051.

Platten, J. D., Cobb, J. N., and Zantua, R. E. (2019). Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. *PLOS ONE* 14, e0210529. doi: 10.1371/journal.pone.0210529.

Poplawsky, A. R., Urban, S. C., and Chun, W. (2000). Biological role of xanthomonadin pigments in *Xanthomonas campestris* pv. *campestris*. *Applied and environmental microbiology* 66, 5123–5127.

Popov, G., Fraiture, M., Brunner, F., and Sessa, G. (2016). Multiple *Xanthomonas euvesicatoria* Type III effectors inhibit flg22-triggered immunity. *MPMI* 29, 651–660. doi: 10.1094/MPMI-07-16-0137-R.

Postle, K., and Kadner, R. J. (2003). Touch and go: tying TonB to transport. *Mol Microbiol* 49, 869–882. doi: 10.1046/j.1365-2958.2003.03629.x.

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H.

Procedure & checklist – preparing multiplexed microbial libraries using SMRTbell express template prep kit 2.0 (2021). Available at: https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93-Preparing-Multiplexed-Microbial-Libraries-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf.

Prodhomme, C., Esselink, D., Borm, T., Visser, R. G. F., van Eck, H. J., and Vossen, J. H. (2019). Comparative Subsequence Sets Analysis (CoSSA) is a robust approach to identify haplotype specific SNPs; mapping and pedigree analysis of a potato wart disease resistance gene Sen3. *Plant Methods* 15, 60. doi: 10.1186/s13007-019-0445-5.

Quilis, J., López-García, B., Meynard, D., Guiderdoni, E., and San Segundo, B. (2014). Inducible expression of a fusion gene encoding two proteinase inhibitors leads to insect and pathogen resistance in transgenic rice. *Plant Biotechnol J* 12, 367–377. doi: 10.1111/pbi.12143.

R Core Team (2022). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Available at: https://www.R-project.org/.

Rademaker, J. L. W., Norman, D. J., Forster, R. L., Louws, F. J., Schultz, M. H., and de Bruijn, F. J. (2006). Classification and identification of *Xanthomonas translucens* isolates, including those pathogenic to ornamental asparagus. *Phytopathology*® 96, 876–884. doi: 10.1094/PHYTO-96-0876.

Rajagopal, L., Sundari, C. S., Balasubramanian, D., and Sonti, R. V. (1997). The bacterial pigment xanthomonadin offers protection against photodamage. *FEBS letters* 415, 125–128.

Rajeshwari, R., Jha, G., and Sonti, R. V. (2005a). Role of an in planta-expressed xylanase of *Xanthomonas oryzae* pv. *oryzae* in promoting virulence on rice. *MPMI* 18, 830–837. doi: 10.1094/MPMI-18-0830.

Rajeshwari, R., Jha, G., and Sonti, R. V. (2005b). Role of an in planta-expressed xylanase of *Xanthomonas oryzae* pv. *oryzae* in promoting virulence on rice. *MPMI* 18, 830–837. doi: 10.1094/MPMI-18-0830.

Ramalingam, J., Vera Cruz, C. M., Kukreja, K., Chittoor, J. M., Wu, J.-L., Lee, S. W., et al. (2003). Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. *Molecular Plant-Microbe Interactions* 16, 14–24.

Ray, S. K., Rajeshwari, R., Sharma, Y., and Sonti, R. V. (2002). A high-molecular-weight outer membrane protein of *Xanthomonas oryzae* pv. *oryzae* exhibits similarity to non-fimbrial adhesins of animal pathogenic bacteria and is required for optimum virulence. *Molecular Microbiology* 46, 637–647. doi: 10.1046/j.1365-2958.2002.03188.x.

Ray, S. K., Rajeshwari, R., and Sonti, R. V. (2000). Mutants of *Xanthomonas oryzae* pv. *oryzae* deficient in general secretory pathway are virulence deficient and unable to secrete xylanase. *MPMI* 13, 394–401. doi: 10.1094/MPMI.2000.13.4.394.

Rechsteiner, M. P., Widmer, F., and Kolliker, R. (2006). Expression profiling of Italian ryegrass (*Lolium multiflorum* Lam.) during infection with the bacterial wilt inducing pathogen *Xanthomonas translucens* pv. *graminis*. *Plant Breeding* 125, 43–51. doi: 10.1111/j.1439-0523.2006.01170.x.

Rigano, L. A., Siciliano, F., Enrique, R., Sendín, L., Filippone, P., Torres, P. S., et al. (2007). Biofilm formation, epiphytic fitness, and canker development in *Xanthomonas axonopodis* pv. *citri*. *MPMI* 20, 1222–1230. doi: 10.1094/MPMI-20-10-1222.

Roberts, D. L., Vargas Jr, J. M., and Detweiler, R. (1985). Occurrence of bacterial wilt on *Poa annua* and other turfgrasses. *Phytopathology* 75, 1289.

Roden, J. A., Belt, B., Ross, J. B., Tachibana, T., Vargas, J., and Mudgett, M. B. (2004). A genetic screen to isolate type III effectors translocated into pepper cells during *Xanthomonas* infection. *Proceedings of the National Academy of Sciences* 101, 16624–16629. doi: 10.1073/pnas.0407383101.

Roman-Reyna, V., Luna, E. K., Pesce, C., Vancheva, T., Chang, C., Ziegle, J., et al. (2020). Genome resource of barley bacterial blight and leaf streak pathogen *Xanthomonas translucens* pv. *translucens* strain UPB886. *Plant Disease* 104, 13–15. doi: 10.1094/PDIS-05-19-1103-A.

Rudolph, K. (1993). "Infection of the plant by *Xanthomonas*," in Xanthomonas, eds. J. G. Swings and E. L. Civerolo (Dordrecht: Springer Netherlands), 193–264. doi: 10.1007/978-94-011-1526-1_4.

Rudolph, K. W. E., Gross, M., Ebrahim-Nesbat, F., Nöllenburg, M., Zomorodian, A., Wydra, K., et al. (1994). "The role of extracellular polysaccharides as virulence factors for phytopathogenic pseudomonads and xanthomonads," in *Molecular Mechanisms of Bacterial Virulence* Developments in Plant Pathology., eds. C. I. Kado and J. H. Crosa (Dordrecht: Springer Netherlands), 357–378. doi: 10.1007/978-94-011-0746-4_25.

Ruiz-Bedoya, T., Wang, P. W., Desveaux, D., and Guttman, D. S. (2023). Cooperative virulence via the collective action of secreted pathogen effectors. *Nat Microbiol* 8, 640–650. doi: 10.1038/s41564-023-01328-8.

Ryan, R. P., Vorhölter, F.-J., Potnis, N., Jones, J. B., Van Sluys, M.-A., Bogdanove, A. J., et al. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nat Rev Microbiol* 9, 344–355. doi: 10.1038/nrmicro2558.

Sagar, V., Dhawan, G., Gopala Krishnan, S., Vinod, K. K., Ellur, R. K., Mondal, K. K., et al. (2020). Marker assisted introgression of genes governing resistance to bacterial blight and blast diseases into an elite Basmati rice variety, 'Pusa Basmati 1509.' *Euphytica* 216, 16. doi: 10.1007/s10681-019-2549-4.

Sapkota, S., Mergoum, M., and Liu, Z. (2020). The translucens group of *Xanthomonas translucens*: Complicated and important pathogens causing bacterial leaf streak on cereals. *Molecular Plant Pathology* 21, 291–302. doi: 10.1111/mpp.12909.

Sauer, N. J., Mozoruk, J., Miller, R. B., Warburg, Z. J., Walker, K. A., Beetham, P. R., et al. (2016). Oligonucleotide-directed mutagenesis for precision gene editing. *Plant Biotechnology Journal* 14, 496–502. doi: 10.1111/pbi.12496.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50, D20–D26. doi: 10.1093/nar/gkab1112.

Schachterle, J. K., Shi, G., Baldwin, T., and Liu, Z. (2022). Complete genomes of two *Xanthomonas translucens* pv. *translucens* strains isolated from barley in North Dakota. *Microbiology Resource Announcements* 11, e00010-22. doi: 10.1128/mra.00010-22.

Schmidt, D. (1988a). Le flétrissement bactérien des graminées fourragères : essais pour limiter la dispersion de la maladie lors du fauchage. *Rev. Suisse Agric* 20, 351–357.

Schmidt, D. (1988b). Prevention of bacterial wilt of grasses by phylloplane bacteria. *Journal of Phytopathology* 122, 253–260. doi: 10.1111/j.1439-0434.1988.tb01014.x.

Schmidt, D. (1988c). *Pseudomonas fluorescen*s and *Erwinia herbicola* reduce wilt of grasses caused by *Xanthomonas campestris* pv. *graminis*. *Journal of Phytopathology* 122, 245–252. doi: 10.1111/j.1439-0434.1988.tb01013.x.

Schmidt, D. (1989). Epidemiological aspects of bacterial wilt of fodder grasses. *EPPO Bulletin* 19, 89–95. doi: 10.1111/j.1365-2338.1989.tb00133.x.

Schmidt, D., and Nuesch, B. (1980). Resistance to bacterial wilt (*Xanthomonas graminis*) increases yield and persistency of *Lolium multiflorum*. *EPPO Bulletin* 10, 335–339. doi: 10.1111/j.1365-2338.1980.tb02707.x.

Schulze, B., Mentzel, T., Jehle, A. K., Mueller, K., Beeler, S., Boller, T., et al. (2010). Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. *Journal of Biological Chemistry* 285, 9444–9451. doi: 10.1074/jbc.M109.096842.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153.

Sekhwal, M., Li, P., Lam, I., Wang, X., Cloutier, S., and You, F. (2015). Disease resistance gene analogs (RGAs) in plants. *IJMS* 16, 19248–19290. doi: 10.3390/ijms160819248.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breeding* 33, 1–14. doi: 10.1007/s11032-013-9917-x.

Seong, E.-S., Choi, D.-I., Cho, H.-S., Lim, C.-K., Cho, H.-J., and Wang, M.-H. (2007). Characterization of a stress-responsive ankyrin repeat-containing zinc finger protein of capsicum annuum (CaKR1). *BMB Reports* 40, 952–958. doi: 10.5483/BMBRep.2007.40.6.952.

Sgro, G. G., Oka, G. U., Souza, D. P., Cenens, W., Bayer-Santos, E., Matsuyama, B. Y., et al. (2019). Bacteria-killing type IV secretion systems. *Front. Microbiol.* 10, 1078. doi: 10.3389/fmicb.2019.01078.

Shah, S. M. A., Haq, F., Ma, W., Xu, X., Wang, S., Xu, Z., et al. (2019). Tal1NXtc01 in *Xanthomonas translucens* pv. *cerealis* contributes to virulence in bacterial leaf streak of wheat. *Front. Microbiol.* 10, 2040. doi: 10.3389/fmicb.2019.02040.

Shah, S. M. A., Khojasteh, M., Wang, Q., Taghavi, S. M., Xu, Z., Khodaygan, P., et al. (2021). Genomics-enabled novel insight into the pathovar-specific population structure of the bacterial leaf streak pathogen *Xanthomonas translucens* in small grain cereals. *Front. Microbiol.* 12, 674952. doi: 10.3389/fmicb.2021.674952.

Sherman, R. M., Forman, J., Antonescu, V., Puiu, D., Daya, M., Rafaels, N., et al. (2019). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat Genet* 51, 30–35. doi: 10.1038/s41588-018-0273-y.

Shrawat, A. K., and Lörz, H. (2006). Agrobacterium-mediated transformation of cereals: a promising approach crossing barriers. *Plant Biotechnology Journal* 4, 575–603. doi: 10.1111/j.1467-7652.2006.00209.x.

Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* 38, 865–891. doi: 10.1111/1574-6976.12067.

Silva, I. T., Rodrigues, F. Á., Oliveira, J. R., Pereira, S. C., Andrade, C. C. L., Silveira, P. R., et al. (2010). Wheat resistance to bacterial leaf streak mediated by silicon. *Journal of Phytopathology* 158, 253–262. doi: 10.1111/j.1439-0434.2009.01610.x.

Sirén, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C., et al. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* 374, abg8871. doi: 10.1126/science.abg8871.

Sletten, A. (1989). Detection methods for *Xanthomonas campestris* pv. *graminis* on forage grasses. *EPPO Bulletin* 19, 69–72.

Song, L., Wang, R., Yang, X., Zhang, A., and Liu, D. (2023). Molecular markers and their applications in marker-assisted selection (MAS) in bread wheat (*Triticum aestivum* L.). *Agriculture* 13, 642. doi: 10.3390/agriculture13030642.

Song, Y., Liu, L., Wang, Y., Valkenburg, D., Zhang, X., Zhu, L., et al. (2018). Transfer of tomato immune receptor Ve1 confers Ave1-dependent *Verticillium* resistance in tobacco and cotton. *Plant Biotechnol J* 16, 638–648. doi: 10.1111/pbi.12804.

Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Zuloaga, F. O., Judziewicz, E. J., et al. (2015). A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution* 53, 117–137. doi: 10.1111/jse.12150.

Souza, D. P., Oka, G. U., Alvarez-Martinez, C. E., Bisson-Filho, A. W., Dunger, G., Hobeika, L., et al. (2015). Bacterial killing via a type IV secretion system. *Nat Commun* 6, 6453. doi: 10.1038/ncomms7453.

Spitz, O., Erenburg, I. N., Beer, T., Kanonenberg, K., Holland, I. B., and Schmitt, L. (2019). Type I secretion systems—one mechanism for all? *Microbiology Spectrum* 7, 7.2.12. doi: 10.1128/microbiolspec.PSIB-0003-2018.

Stall, R. E., Jones, J. B., and Minsavage, G. V. (2009). Durability of resistance in tomato and pepper to Xanthomonads causing bacterial spot. *Annual Review of Phytopathology* 47, 265–284. doi: 10.1146/annurev-phyto-080508-081752.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat Rev Genet* 20, 631–656. doi: 10.1038/s41576-019-0150-2.

Stead, D. E. (1989). Grouping of *Xanthomonas campestris* pathovars of cereals and grasses by fatty acid profiling 1. *EPPO bulletin* 19, 57–68.

Stork, W., Kim, J.-G., and Mudgett, M. B. (2015). Functional analysis of plant defense suppression and activation by the *Xanthomonas* core type III effector XopX. *MPMI* 28, 180–194. doi: 10.1094/MPMI-09-14-0263-R.

Streubel, J., Baum, H., Grau, J., Stuttman, J., and Boch, J. (2017). Dissection of TALE-dependent gene activation reveals that they induce transcription cooperatively and in both orientations. *PLOS ONE* 12, e0173580. doi: 10.1371/journal.pone.0173580.

Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013a). Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytologist* 200, 808–819. doi: 10.1111/nph.12411.

Streubel, J., Pesce, C., Hutin, M., Koebnik, R., Boch, J., and Szurek, B. (2013b). Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytologist* 200, 808–819. doi: 10.1111/nph.12411.

Strom, M. S., and Lory, S. (1991). Amino acid substitutions in pilin of *Pseudomonas aeruginosa*. Effect on leader peptide cleavage, amino-terminal methylation, and pilus assembly. *Journal of Biological Chemistry* 266, 1656–1664. doi: 10.1016/S0021-9258(18)52345-0.

Studer, B., Boller, B., Herrmann, D., Bauer, E., Posselt, U. K., Widmer, F., et al. (2006). Genetic mapping reveals a single major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). *Theoretical and Applied Genetics* 113, 661–671. doi: 10.1007/s00122-006-0330-2.

Studer, B., Kölliker, R., Muylle, H., Asp, T., Frei, U., Roldán-Ruiz, I., et al. (2010). EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (Lolium spp.). *BMC Plant Biology* 10, 177. doi: 10.1186/1471-2229-10-177.

Subramoni, S., and Sonti, R. V. (2005). Growth deficiency of a *Xanthomonas oryzae* pv. *oryzae fur* mutant in rice leaves is rescued by ascorbic acid supplementation. *MPMI* 18, 644–651. doi: 10.1094/MPMI-18-0644.

Suter, D., Frick, R., and Hirschi, H. (2021). Sortenprüfung mit Italienischem Raigras: Zwei Neuempfehlungen für die Schweizer Landwirtschaft. doi: 10.34776/AFS12-128.

Suter, M., Connolly, J., Finn, J. A., Loges, R., Kirwan, L., Sebastià, M.-T., et al. (2015). Nitrogen yield advantage from grass–legume mixtures is robust over a wide range of legume proportions and environmental conditions. *Global Change Biology* 21, 2424–2438. doi: 10.1111/gcb.12880.

Swathi, G., Durga Rani, Ch. V., Md, J., Madhav, M. S., Vanisree, S., Anuradha, Ch., et al. (2019). Marker-assisted introgression of the major bacterial blight resistance genes, Xa21 and xa13, and blast resistance gene, Pi54, into the popular rice variety, JGL1798. *Mol Breeding* 39, 58. doi: 10.1007/s11032-019-0950-2.

Szczesny, R., Jordan, M., Schramm, C., Schulz, S., Cogez, V., Bonas, U., et al. (2010). Functional characterization of the Xcs and Xps type II secretion systems from the plant pathogenic

bacterium *Xanthomonas campestris* pv *vesicatoria*. *New Phytologist* 187, 983–1002. doi: 10.1111/j.1469-8137.2010.03312.x.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal* 74, 174–183. doi: 10.1111/tpj.12105.

Takuno, S., Terauchi, R., and Innan, H. (2012). The power of QTL mapping with RILs. *PLOS ONE* 7, e46545. doi: 10.1371/journal.pone.0046545.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3.

Tewhey, R., Bansal, V., Torkamani, A., Topol, E. J., and Schork, N. J. (2011). The importance of phase information for human genomics. *Nat Rev Genet* 12, 215–223. doi: 10.1038/nrg2950.

Thomma, B. P. H. J., Nürnberger, T., and Joosten, M. H. A. J. (2011). Of PAMPs and Effectors: the blurred PTI-ETI dichotomy. *The Plant Cell* 23, 4–15. doi: 10.1105/tpc.110.082602.

Timilsina, S., Potnis, N., Newberry, E. A., Liyanapathiranage, P., Iruegas-Bocardo, F., White, F. F., et al. (2020). Xanthomonas diversity, virulence and plant–pathogen interactions. *Nat Rev Microbiol*. doi: 10.1038/s41579-020-0361-8.

Totsika, M., Wells, T. J., Beloin, C., Valle, J., Allsopp, L. P., King, N. P., et al. (2012). Molecular characterization of the EhaG and UpaG trimeric autotransporter proteins from pathogenic *Escherichia coli*. *Appl Environ Microbiol* 78, 2179–2189. doi: 10.1128/AEM.06680-11.

Truong, H. T. H., Kim, K.-T., Kim, S., Cho, M.-C., Kim, H.-R., and Woo, J.-G. (2011). Development of gene-based markers for the Bs2 bacterial spot resistance gene for marker-assisted selection in pepper (*Capsicum* spp.). *Hortic. Environ. Biotechnol.* 52, 65–73. doi: 10.1007/s13580-011-0142-4.

Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K. J., et al. (2011). Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiology* 156, 1257–1268. doi: 10.1104/pp.110.169748.

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *JOSS* 3, 731. doi: 10.21105/joss.00731.

Tyagi, P., Singh, D., Mathur, S., Singh, A., and Ranjan, R. (2022). Upcoming progress of transcriptomics studies on plants: An overview. *Front Plant Sci* 13, 1030890. doi: 10.3389/fpls.2022.1030890.

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat Rev Methods Primers* 1, 1–21. doi: 10.1038/s43586-021-00056-9.

Ukai, Y. (2010). Theoretical studies on population development design in mutation breeding of allogamous plants. *Breed. Sci.* 60, 267–278. doi: 10.1270/jsbbs.60.267.

United Nations Department of Economic and Social Affairs (2018). World population prospects: the 2017 revision. Available at:

https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html [Accessed October 19, 2018].

Van Den Mooter, M., Swings, J., De Cleene, M., Leyns, F., and De Ley, J. (1981). Isolation and identification of *Xanthomonas campestris* pv. *graminis* from forage grasses in Belgium. *Parasitica*, 23–8.

van Schie, C. C. N., and Takken, F. L. W. (2014). Susceptibility Genes 101: How to Be a Good Host. *Annual Review of Phytopathology* 52, 551–581. doi: 10.1146/annurev-phyto-102313-045854.

Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. *Critical Reviews in Microbiology* 43, 709–730. doi: 10.1080/1040841X.2017.1303661.

Vauterin, L., Hoste, B., Kersters, K., and Swings, J. (1995). Reclassification of *Xanthomonas*. *International Journal of Systematic Bacteriology* 45, 472–489. doi: 10.1099/00207713-45-3-472.

Veltri, D., Wight, M. M., and Crouch, J. A. (2016). SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. *Nucleic Acids Res* 44, W41–W45. doi: 10.1093/nar/gkw330.

Vernikos, G. S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22, 2196–2203. doi: 10.1093/bioinformatics/btl369.

Viana, V. E., Pegoraro, C., Busanello, C., and Costa de Oliveira, A. (2019). Mutagenesis in rice: the basis for breeding a new super plant. *Frontiers in Plant Science* 10.

Vilhjálmsson, B. J., and Nordborg, M. (2013). The nature of confounding in genome-wide association studies. *Nat Rev Genet* 14, 1–2. doi: 10.1038/nrg3382.

Villarino Romero, R., Osicka, R., and Sebo, P. (2014). Filamentous hemagglutinin of *Bordetella pertussis*: a key adhesin with immunomodulatory properties? *Future Microbiology* 9, 1339–1360. doi: 10.2217/fmb.14.77.

Voichek, Y., and Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nat Genet* 52, 534–540. doi: 10.1038/s41588-020-0612-7.

Wagner, N., Alburquerque, M., Ecker, N., Dotan, E., Zerah, B., Pena, M. M., et al. (2022a). Natural language processing approach to model the secretion signal of type III effectors. *Frontiers in Plant Science* 13.

Wagner, N., Avram, O., Gold-Binshtok, D., Zerah, B., Teper, D., and Pupko, T. (2022b). Effectidor: an automated machine-learning-based web server for the prediction of type-III secretion system effectors. *Bioinformatics* 38, 2341–2343. doi: 10.1093/bioinformatics/btac087.

Wan, J., Zhang, X.-C., and Stacey, G. (2008). Chitin signaling and plant disease resistance. *Plant Signaling & Behavior* 3, 831–833. doi: 10.4161/psb.3.10.5916.

Wang, B., Ebbole, D. J., and Wang, Z. (2017a). The arms race between *Magnaporthe oryzae* and rice: Diversity and interaction of </i> Avr</i> and </i>R</i> genes. *Journal of Integrative Agriculture* 16, 2746–2760. doi: 10.1016/S2095-3119(17)61746-5.

Wang, H. M., and Sletten, A. (1995). Infection biology of bacterial wilt of forage grasses. *Journal of Phytopathology* 143, 141–145.

Wang, L., Makino, S., Subedee, A., and Bogdanove, A. J. (2007). Novel candidate virulence factors in rice pathogen *Xanthomonas oryzae* pv. *oryzicola* as revealed by mutational analysis. *Applied and Environmental Microbiology* 73, 8023–8027. doi: 10.1128/AEM.01414-07.

Wang, L., Rinaldi, F. C., Singh, P., Doyle, E. L., Dubrow, Z. E., Tran, T. T., et al. (2017b). TAL effectors drive transcription bidirectionally in plants. *Molecular Plant* 10, 285–296. doi: 10.1016/j.molp.2016.12.002.

Wang, T., Antonacci-Fulton, L., Howe, K., Lawson, H. A., Lucas, J. K., Phillippy, A. M., et al. (2022). The Human Pangenome Project: a global resource to map genomic diversity. *Nature* 604, 437–446. doi: 10.1038/s41586-022-04601-8.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics; Series A* 26.

Wei, H.-L., Chakravarthy, S., Mathieu, J., Helmann, T. C., Stodghill, P., Swingle, B., et al. (2015). *Pseudomonas syringae* pv. *tomato* DC3000 type III secretion effector polymutants reveal an interplay between HopAD1 and AvrPtoB. *Cell Host & Microbe* 17, 752–762. doi: 10.1016/j.chom.2015.05.007.

Wei, H.-L., Zhang, W., and Collmer, A. (2018). Modular study of the type III effector repertoire in pseudomonas syringae pv. tomato DC3000 reveals a matrix of effector interplay in pathogenesis. *Cell Reports* 23, 1630–1638. doi: 10.1016/j.celrep.2018.04.037.

Wei, Z. M., Laby, R. J., Zumoff, C. H., Bauer, D. W., He, S. Y., Collmer, A., et al. (1992). Harpin, elicitor of the hypersensitive response produced by the plant pathogen Erwinia amylovora. *Science* 257, 85–88. doi: 10.1126/science.1621099.

White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Molecular Plant Pathology* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x.

Wichmann, F., Asp, T., Widmer, F., and Kölliker, R. (2011a). Transcriptional responses of Italian ryegrass during interaction with *Xanthomonas translucens* pv. *graminis* reveal novel candidate genes for bacterial wilt resistance. *Theoretical and Applied Genetics* 122, 567–579. doi: 10.1007/s00122-010-1470-y.

Wichmann, F., Müller Hug, B., Widmer, F., Boller, B., Studer, B., and Kölliker, R. (2011b). Phenotypic and molecular genetic characterization indicate no major race-specific interactions between *Xanthomonas translucens* pv. *graminis* and Lolium multiflorum: Bacterial wilt of Italian ryegrass. *Plant Pathology* 60, 314–324. doi: 10.1111/j.1365-3059.2010.02373.x.

Wichmann, F., Vorhölter, F.-J., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Molecular Plant Pathology* 14, 576–588. doi: 10.1111/mpp.12030.

Wilkins, P., and Exley, J. (1977). Bacterial wilt of ryegrass in Britain. *Plant Pathology* 26, 99.

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D. G. O., Yoshida, K., Dong, S., et al. (2012). Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol* 77, 235–247. doi: 10.1101/sqb.2012.77.015933.

Wittstock, U., and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. *Current Opinion in Plant Biology* 5, 300–307. doi: 10.1016/S1369-5266(02)00264-9.

Xu, Y., and Crouch, J. H. (2008). Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science* 48, 391–407. doi: 10.2135/cropsci2007.04.0191.

Xu, Y., Li, P., Yang, Z., and Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal* 5, 175–184. doi: 10.1016/j.cj.2016.06.003.

Yamada, T., Forster, J. W., Humphreys, M. W., and Takamizo, T. (2005). Genetics and molecular breeding in Lolium/Festuca grass species complex. *Grassland Science* 51, 89–106. doi: 10.1111/j.1744-697X.2005.00024.x.

Yan, F., Wang, J., Zhang, S., Lu, Z., Li, S., Ji, Z., et al. (2023). CRISPR/FnCas12a-mediated efficient multiplex and iterative genome editing in bacterial plant pathogens without donor DNA templates. *PLoS Pathog* 19, e1010961. doi: 10.1371/journal.ppat.1010961.

Yang, B., Sugio, A., and White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proceedings of the National Academy of Sciences* 103, 10503–10508.

Yang, L. (2020). A practical guide for structural variation detection in human genome. *Curr Protoc Hum Genet* 107, e103. doi: 10.1002/cphg.103.

Yeh, Y.-H., Chang, Y.-H., Huang, P.-Y., Huang, J.-B., and Zimmerli, L. (2015). Enhanced *Arabidopsis* pattern-triggered immunity by overexpression of cysteine-rich receptor-like kinases. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00322.

Young, J. M., Dye, D. W., Bradbury, J. F., Panagopoulos, C. G., and Robbs, C. F. (1978). A proposed nomenclature and classification for plant pathogenic bacteria. *New Zealand Journal of Agricultural Research* 21, 153–177. doi: 10.1080/00288233.1978.10427397.

Young, K. H. (1998). Yeast two-hybrid: so many interactions, (in) so little time…. *Biology of Reproduction* 58, 302–311. doi: 10.1095/biolreprod58.2.302.

Young, N. D. (1996). QTL mapping and quantitative disease resistance in plants. *Annu Rev Phytopathol* 34, 479–501. doi: 10.1146/annurev.phyto.34.1.479.

Yu, J., Holland, J. B., McMullen, M. D., and Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* 178, 539–551. doi: 10.1534/genetics.107.074245.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38, 203–208. doi: 10.1038/ng1702.

Yugander, A., Sundaram, R. M., Singh, K., Prasad, M. S., Hari Prasad, A. S., Madhav, M. S., et al. (2019). Marker assisted introgression of a major bacterial blight resistance gene, Xa38

into a rice maintainer line, APMS 6B. *Indian Phytopathology* 72, 35–41. doi: 10.1007/s42360-018-00111-8.

Zhang, B., Han, X., Yuan, W., and Zhang, H. (2022). TALEs as double-edged swords in plant–pathogen interactions: Progress, challenges, and perspectives. *Plant Communications* 3, 100318. doi: 10.1016/j.xplc.2022.100318.

Zhang, Y., Ran, Y., Nagy, I., Lenk, I., Qiu, J.-L., Asp, T., et al. (2020). Targeted mutagenesis in ryegrass (*Lolium* spp.) using the CRISPR/Cas9 system. *Plant Biotechnology Journal* 18, 1854–1856. doi: 10.1111/pbi.13359.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42, 355–360. doi: 10.1038/ng.546.

Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2, 467. doi: 10.1038/ncomms1467.

Zhao, Y., Liu, Z., and Wu, J. (2020). Grassland ecosystem services: a systematic review of research advances and future directions. *Landscape Ecol* 35, 793–814. doi: 10.1007/s10980-020-00980-3.

Zhou, Y., Zhang, Z., Bao, Z., Li, H., Lyu, Y., Zan, Y., et al. (2022). Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* 606, 527–534. doi: 10.1038/s41586-022-04808-9.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., Boller, T., et al. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts agrobacterium-mediated transformation. *Cell* 125, 749–760. doi: 10.1016/j.cell.2006.03.037.

7. Acknowledgements

First, I would like to dedicate this thesis to my amazing wife, Joanna Olivi-Goettelmann, without whom I would not be where I am today, and who supported me in many ways during this doctorate, whether by emotionally supporting me, or giving her scientific advice on my experiments. I also dedicate this work to my grandfather, Georges Fabre, who transmitted his love of science to me and was always there for me.

I give my sincere thanks to Prof. Dr. Bruno Studer for giving me the opportunity to conduct this doctoral thesis in the Molecular Plant Breeding group, and for always doing his best to support me. I express my deepest gratitude to Dr. Roland Kölliker, who was the greatest supervisor I could have hoped for. You were always open, trusting, supporting, and made the difficult times I encountered much easier to bear. Thanks to the both of you for your great scientific input, which pushed me towards the results I present in this thesis.

I thank Prof. Dr. Jens Boch for accepting to be part of my thesis committee and to be a co-examiner for my doctoral examination, and always giving excellent scientific advice.

Thanks to everyone from the Molecular Plant Breeding group, which was a great group to do my doctorate with. I particularly thank Esther Ravelhofer-Alge, who was always helpful and supportive, and a great person to talk to. Many thanks also to Verena, who helped me a lot at the beginning of my doctorate and gave great advice, to Dr. Steven Yates, who taught me many things about bioinformatics that allowed me to do most of my work and helped me a lot, to Dr. Dario Copetti and Yutang who also helped a lot in my analysis, and to Max, Daniel and Carles who gave great input. Thanks to Fabienne Bürki who was a great student to supervise.

I thank Dr. Franco Widmer for allowing me to conduct my work in collaboration with the Molecular Ecology group at Agroscope Reckenholz, and all the Mökis: Aaron, Chiara, Johanna, Florian, and Maria, as well as Miguel who also had the MPB-Möki double hat like me, with whom I had great times at Reckenholz and outside. Thanks to Tabea and Urs for the great help in the lab.

I thank Dr. Ralf Koebnik, Dr. Claude Bragard and Dr. Jonathan Jacobs who were great people to work with on the *Xanthomonas* sequencing project. I also particularly thank Dr. Veronica Roman-Reyna who taught me many things that were instrumental in my analysis. I also thank all the people from the EuroXanth COST action I interacted with at the EuroXanth conference in Lednice and at the training school in Toulouse.

Finally, I thank my friends: Ziad, Jordan, Nicolas and Cypriane, who I could always talk to, with whom I spent great times, and who were a great source of comfort.

8. Curriculum vitae

Personal Information

Florian Goettelmann

Email: florian.goettelmann@gmail.com

Born on 13/06/1993

Work Experience

October 2018 - June 2023	ETH Zurich – Doctoral student Resistance gene identification through advanced host-pathogen genomics in the Xanthomonas-ryegrass pathosystem
March - August 2018	CPIE Iles de Lérins et Pays d'Azur – Civic service
January - June 2017	Institut Sophia Agrobiotech – 2nd Master's Thesis Evolution of plant defense compounds in the fresh plant - decomposed litter continuum
February - June 2016	Institut Sophia Agrobiotech – 1st Master's Thesis Functional analysis of plant targets of effectors in the parasitism of *Meloidogyne incognita*

Education

October 2018 – June 2023	ETH Zurich - Doctorate
2015-2017	Université de Nice-Sophia Antipolis – Master's degree in Environmental Health and Biology – Plant protection and agroenvironment
2011-2015	Université de Nice-Sophia Antipolis – Bachelor's degree in Life and Health biology – Organisms and Ecosystems Biology

Publications

Goettelmann, F., Roman-Reyna, V., Cunnac, S., Jacobs, J. M., Bragard, C., Studer, B., et al. (2022). Complete Genome Assemblies of All Xanthomonas translucens Pathotype Strains Reveal Three Genetically Distinct Clades. Frontiers in Microbiology 12.

Conference articles, presentations, and posters

Goettelmann F, Copetti D, Yates S, Studer B, Kölliker R. Pooled Sequencing Reveals Multiple Regions Involved in Resistance to Bacterial Wilt in Italian Ryegrass. Conference lecture presented at: 34th Meeting of the EUCARPIA Fodder Crops and Amenity Grasses Section, Freising, Germany, September 6–8, 2021

Goettelmann, F., Copetti, D., Yates, S., Studer, B., and Kölliker, R. (2021). Pooled sequencing reveals genome regions involved in resistance to bacterial wilt in Italian ryegrass. in, eds. S. Hartmann, S. Bachmann-Pfabe, U. Feuerstein, B. Julier, R. Kölliker, D. Kopecký, et al. (Palacký University), 62. doi: 10.5507/vup.21.24459677.14.

Goettelmann F, Copetti D, Yates S, Studer B, Kölliker R. Pooled sequencing identifies candidate genes for resistance to *Xanthomonas transluscens* pv. *graminis* in *Lolium multiflorum*. Conference lecture presented at: 4th Annual Conference of the EuroXanth COST Action Integrating Science on Xanthomonadaceae for integrated plant disease management in Europe (EuroXanth 2021), Belgrade, Serbia, June 28–30, 2021

Goettelmann F, Roman-Reyna V, Jacobs J M, Bragard C, Studer B, Koebnik R, Kölliker R Comparative genomics to understand host range in *Xanthomonas translucens*. Poster presented at: 4th Annual Conference of the EuroXanth COST Action Integrating Science on Xanthomonadaceae for integrated plant disease management in Europe (EuroXanth 2021), Belgrade, Serbia, June 28–30, 2021

Goettelmann F, Kölliker R, Studer B. Identifiying resistance genes in the *Lolium multiflorum* - *Xanthomonas translucens* pv. *graminis* pathosystem. Poster presented at: Plant Response to Environment across Scales, Symposium of the Zurich-Basel Plant Science Center, 2019 Dec 11, Zurich, Switzerland.

Goettelmann F, Kölliker R, Studer B. Identifiying resistance genes in the *Lolium multiflorum* - *Xanthomonas translucens* pv. *graminis* pathosystem. Poster presented at: 3rd Annual Conference of the EuroXanth COST Action, 2019 Sep 9-11, Lednice, Czech Republic.