Host-specificity factors in plant pathogenic fungi

Jiming Li, Ben Cornelissen, Martijn Rep*

Keywords:
Host specificity
Virulence genes
Avirulence genes
Plant pathogenic fungi

Abstract

Fortunately, no fungus can cause disease on all plant species, and although some plant-pathogenic fungi have quite a broad host range, most are highly limited in the range of plant species or even cultivars that they cause disease in. The mechanisms of host specificity have been extensively studied in many plant-pathogenic fungi, especially in fungal pathogens causing disease on economically important crops. Specifically, genes involved in host specificity have been identified during the last few decades. In this overview, we describe and discuss these host-specificity genes. These genes encode avirulence (Avr) proteins, proteinaceous host-specific toxins or secondary metabolites. We discuss the genomic context of these genes, their expression, polymorphism, horizontal transfer and involvement in pathogenesis.

1. Introduction

Plant diseases caused by fungi are recognized as a major threat to food security (Doelemann et al., 2017; Fisher et al., 2012; Pennisi, 2010). For example, wheat stem rust caused by Puccinia graminis, Asian soybean rust caused by Phakopsora pachyrhizi, rice blast caused by Magnaporthe oryzae, and banana black sigatoka caused by Mycosphaerella fijiensis have resulted in serious yield losses in human history (Pennisi, 2010).

Most plant pathogenic fungal species have a narrow range of plant species in which they cause disease, a phenomenon we here call 'host species specificity'. The collective host range of some fungal species, such as Fusarium oxysporum (F. oxysporum), can be very large, but then individual strains are often limited to infect one or a few plant species only (Borah et al., 2018; Pietro et al., 2003). Based on host range, strains within such fungal species are commonly classified into different pathotypes or formae speciales. In addition, fungal species or formae speciales are sometimes divided into different races depending on the particular cultivars of a plant species that they are able to infect. This phenomenon we here call 'host cultivar specificity'. For example, F. oxysporum is classified into more than 100 formae speciales based on host species specificity, including the tomato-infecting strain F. oxysporum forma specialis (f. sp.) lycopersici (Fol), while Fol itself is subdivided into three races based on host cultivar specificity (Takken and Rep, 2010). The molecular basis of host cultivar specificity has been extensively studied in many plant pathogenic fungi, while the molecular basis of host species specificity is less well understood (de Wit, 2016; Lanver et al., 2017; Lo Presti et al., 2015; Prasad et al., 2019; Selin et al., 2016; Yan and Talbot, 2016). In this review, the term 'host specificity' includes both host species specificity and host cultivar specificity.

To understand the genetic basis of host specificity in plant pathogenic fungi, the emergence of a molecular understanding of plant immunity over the last three decades has been essential. To defend themselves against fungal pathogens, plants have evolved two layers of immunity (Jones and Dangl, 2006). The first layer of immunity responds to pathogen-associated molecular patterns (PAMPs) common to many microbes, including non-pathogens, and this defense system is called PAMP-triggered immunity (PTI) (Jones and Dangl, 2006). To suppress PTI responses, pathogens secrete molecules called effectors to facilitate colonization. These effectors are commonly secreted proteins but can also be metabolites. Some effectors target plant susceptibility (S) proteins, resulting in effector-triggered susceptibility (ETS) (van Schie and Takken, 2014). For example, necrotrophic effectors of Stagonospora nodorum (S. nodorum) are able to interact with wheat susceptibility gene products (Oliver et al., 2012). The second layer of defense comprises plant resistance (R) proteins that directly or indirectly recognize pathogen-produced effectors, resulting in effector-triggered immunity (ETI) (Jones and Dangl, 2006). Among identified R proteins, most are nucleotide-binding leucine-rich repeat proteins (NLRs) (Kourelis and van der Hoorn, 2018). To evade effector recognition by plant R proteins, pathogens can undergo loss or mutation of the corresponding effector genes (Jones and Dangl, 2006; Wang and Wang, 2018). As a result, pathogens and plants are evolving in a perpetual arms race. Most effector genes are located in repeat-rich regions. For example, all known effectors genes in Leptosphaeria maculans (L.
2.1.1. *F. oxysporum f. sp. lycopersici* (Fol)

The *I-3* protein, was identified in Fol as well (Rep et al., 2004). The resistance gene *AVR2* and *AVR3* are (Houterman et al., 2009; Rep et al., 2004). For activation of 1-2-mediated resistance, not only Avr2 is required, but also a Fol protein called Secreted in xylem 5 (Six5). Like Avr2, Six5 is required for full virulence (Houterman et al., 2009; Ma et al., 2015). All three Fol AVR genes encode small, secreted proteins with multiple cysteines, and are located on a single accessory chromosome with high density of repetitive elements (Ma et al., 2010; Schmidt et al., 2013).

This arms race between pathogens and plants results in relatively fast evolution of effectors, R and S proteins (Jones and Dangl, 2006; Wang and Wang, 2018). In general, effectors promote virulence and are therefore virulence factors. Effectors that are recognized by R proteins are (also) called avirulence (Avr) factors (Avr proteins) (Stotz et al., 2014). These avirulence and virulence factors determining resistance and susceptibility of plants, respectively, are considered to be host-specificity factors and have been identified in many plant pathogenic fungi (de Wit, 2016; Lanver et al., 2017; Prasad et al., 2019; Selin et al., 2016; Toruño et al., 2016; Yan and Talbot, 2016). The virulence factors reviewed here are known as proteinaceous toxins, therefore they are referred as proteinaceous host-specific toxins hereafter. In addition to pathogen-secreted proteins that can determine host specificity, secondary metabolites can also act as host-specificity determinants, such as the host-selective toxins (HSTs) of *Alternaria alternata* and *Verticillium dahliae* (Chen et al., 2018; Tsuge et al., 2013; Zhang et al., 2019).

Here we first describe host-specificity factors in plant pathogenic fungi, including avirulence proteins, proteinaceous host-specific toxins and secondary metabolites. Then, involvement of horizontal transfer and the genomic context in the acquisition and evolution of host-specificity genes is addressed.

2. Avirulence proteins determining host specificity

Avirulence proteins determining host cultivar specificity have been identified in various fungal species (Table 1). Below we provide a brief overview of these cases, including involvement in pathogenesis, polymorphism and interaction with the corresponding resistance proteins.

2.1. *Fusarium oxysporum*

*F. oxysporum* is a soil-borne (presumably) asexual fungal species complex, which causes vascular wilt and root rot (Edel-Hermann and Lecomte, 2019). Avirulence genes in *F. oxysporum* determining host cultivar specificity have been identified in several *forme speciales*, including Fol (Takken and Rep, 2010). *F. oxysporum f. sp. melonis* (Schmidt et al., 2016) and *F. oxysporum f. sp. niveum* (Niu et al., 2016). Here, we will review the avirulence (AVR) genes identified in Fol and *F. oxysporum f. sp. melonis*, which infect tomato and melon, respectively.

2.1.1. *F. oxysporum f. sp. lycopersici* (Fol)

*F. oxysporum f. sp. lycopersici* (Fol) can be divided into three races based on their capability to infect tomato cultivars containing different resistance genes to Fol (Takken and Rep, 2010). Race 1 contains three AVR genes, notably AVR1, AVR2 and AVR3. The protein encoded by resistance gene *I* recognizes the product of AVR1, upon which the immune system is activated in the plant (Houterman et al., 2008). Avr1 also suppresses recognition of Avr2 and Avr3 by resistance proteins I-2 and I-3, respectively (Houterman et al., 2008). Race 2 evolved from race 1 by deletion of a chromosomal region containing AVR1, likely due to a recombination event between two TE bordering the fragment (Biju et al., 2017). The I-2 resistance gene was introduced into tomato cultivars to protect them against race 2. The I-2 protein recognizes Avr2 (Houterman et al., 2009; Ma et al., 2015). Single point mutations in AVR2 subsequently emerged such that the gene product was no longer recognized by I-2, resulting in race 3 (Houterman et al., 2009). Resistance gene I-3 against race 3 was introduced in tomato cultivars, and the corresponding AVR3 gene, the product of which is recognized by the I-3 protein, was identified in Fol as well (Rep et al., 2004).

Among the three Fol AVR genes known, *AVR1* is not required for full virulence on susceptible hosts (Houterman et al., 2008), whereas *AVR2* and *AVR3* are (Houterman et al., 2009; Rep et al., 2004). For activation of 1-2-mediated resistance, not only Avr2 is required, but also a Fol protein called Secreted in xylem 5 (Six5). Like Avr2, Six5 is required for full virulence (Houterman et al., 2009; Ma et al., 2015). All three Fol AVR genes encode small, secreted proteins with multiple cysteines, and are located on a single accessory chromosome with high density of repetitive elements (Ma et al., 2010; Schmidt et al., 2013).

2.1.2. *F. oxysporum f. sp. melonis*

*F. oxysporum f. sp. melonis* is divided into race 0, race 1, race 2, and race 1,2. So far, only one avirulence gene has been identified, *AvrFom2*, whose product is recognized by the protein encoded by the melon *R* gene *Fom2* (Schmidt et al., 2016). *AvrFom2* is a small secreted protein with two cysteine residues and without recognizable domains (Schmidt et al., 2016). However, it does show an overall low similarity to ToxA of *Pyrenophora tritici-repentis* (discussed below), and the cysteine residues that form the characteristic cysteine knot in ToxA are conserved (Schmidt et al., 2016). *AvrFom2* is located in a lineage-specific region of the Fom001 genome and resides close to transposons (Schmidt et al., 2016; van Dam et al., 2017). The gene is absent in race 2 isolates (Schmidt et al., 2016).

2.2. *Cladosporium fulvum*

*C. fulvum* (Passalora fulva) is a non-obligate biotrophic fungal species and the causal agent of tomato leaf mold (de Wit, 2016). Already in the 1970s, it was found that the gene-for-gene relation between tomato and *C. fulvum* is based on the interaction of specific fungal products with specific resistance proteins in tomato (van Dijkman and Kaars Sijpsteijn, 1973).

The first fungal Avr protein was identified in *C. fulvum* in 1991 (van Kan, 1991). Until now, ten AVR genes of *C. fulvum*, which encode small cysteine-rich proteins, have been identified, notably *Avr2* (Luderer et al., 2002), *Avr4* (Joosten et al., 1994), *Avr4E* (Westrink et al., 2004), *Avr5* (Mesarich et al., 2014), and *Avr9* (van Kan, 1991), *Ecp1* (Laugé et al., 1997), *Ecp2-I* (Laugé et al., 1997), *Ecp4* (Laugé et al., 2000), *Ecp5* (Laugé et al., 2000) and *Ecp6* (Bolton et al., 2008). Recognition of the encoded proteins in tomato is mediated by the cognate Cf resistance genes (CF-2, CF-4, CF-4E, CF-5 and CF-9, Cl-Ecp1, Cl-Ecp2-2, Cl-Ecp4, Cl-Ecp5 and Cl-Ecp6, respectively. These *C. fulvum* Avr proteins and other apoplastic effector proteins have been extensively reviewed previously (de Wit, 2016; Rivas and Thomas, 2005; Stergiopoulos and de Wit, 2009). In a recent study, nine newly identified small secreted proteins were found to be recognized by specific wild accessions of tomato, but the corresponding Cf immune receptor genes in these accessions are still unknown (Mesarich et al., 2018).

All *C. fulvum* Avr proteins identified are less than 300 amino acids in size and contain an even number of at least four cysteine residues (Mesarich et al., 2018, 2014). A virulence function for *Avr2*, *Avr4*, *Avr5* and *Ecp6* has been demonstrated (Mesarich et al., 2014; Stergiopoulos and de Wit, 2009). To avoid recognition by R proteins, several types of sequence modifications have occurred in *C. fulvum* AVR genes including gene deletions, gene disruption by insertion of a transposon-like element, and nonsynonymous amino acid substitutions (Stergiopoulos and de Wit, 2009). The frequency of such mutations may have been enhanced by proximity of these AVR genes to repetitive elements (de Wit et al., 2012).

2.3. *Leptosphaeria maculans*

*L. maculans* is a hemi-biotrophic ascomycete responsible for stem canker of oilseed rape (Petit-Houdenot and Fudal, 2017). To date, eight AVR genes from *L. maculans* have been identified, *AvrLm1*, *AvrLm2*, *AvrLm3*, *AvrLm4–7*, *AvrLm5–9* (AvrLm11), *AvrLm6*, *AvrLm10* and *AvrLm11*, and all are located in repeat-rich, gene-poor genomic regions
### Table 1
Host-specificity factors in plant pathogenic fungi.

<table>
<thead>
<tr>
<th>Fungal species</th>
<th>Host</th>
<th>Host-specificity genes</th>
<th>Corresponding resistance or susceptibility genes</th>
<th>Genome location of host-specificity genes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Fusarium oxysporum</em> f. sp. <em>lycopersici</em></td>
<td>Tomato</td>
<td>AVR1, AVR2, AVR3</td>
<td>I-1, I-2, I-3</td>
<td>Accessory chromosome</td>
<td>(Ma et al., 2010; Takken and Rep, 2010)</td>
</tr>
<tr>
<td><em>Fusarium oxysporum</em> f. sp. <em>melonis</em></td>
<td>Melon</td>
<td>AvrFom2</td>
<td></td>
<td>Close to repetitive elements</td>
<td>(Schmidt et al., 2016)</td>
</tr>
<tr>
<td><em>Cladosporium fulvum</em> (<em>Passalora fulva</em>)</td>
<td>Tomato</td>
<td>AvrLm1, AvrLm2, AvrLm3, AvrLm4–7, AvrLm5-9, AvrLm6, AvrLm10 and AvrLm11</td>
<td></td>
<td>Repeat-rich regions</td>
<td>(Ghanbarian et al., 2018; Petit-Houdenot et al., 2019; Plissonneau et al., 2018; Rouxel and Balesdent, 2017)</td>
</tr>
<tr>
<td><em>M. oryzae</em></td>
<td>Grass species</td>
<td>Avr-Pi-54, Avr-Pi-9, Avr-Pa, Avr-Pi, Avr-Pi/km/kp, Avr-Pi-ta1/2/3, Avr-Pi-tz, Avr-Pi-CO39, Avr-CO39</td>
<td>Pi-54, Pi-9, Pi-a, Pi-Pi/km/kp, Pi-ta1/2/3, Pi-tz, Pi-CO39, Pi33</td>
<td>Close to repetitive elements</td>
<td>(Fernandez and Orth, 2018)</td>
</tr>
<tr>
<td><em>Rhynchosporium secalis</em></td>
<td>Barley</td>
<td>Nip1</td>
<td></td>
<td>Present on a large chromosome</td>
<td>(Hahn, 1993; Mohd-Assaad et al., 2019)</td>
</tr>
<tr>
<td><em>Melampsora lini</em></td>
<td>Flax</td>
<td>AvrLs67, AvrM, AvrP123, AvrP4, AvrL2-A, and AvrM14-A</td>
<td></td>
<td>Mostly located at recombination hot-spots</td>
<td>(Lorrain et al., 2019; Petre et al., 2014)</td>
</tr>
<tr>
<td><em>Puccinia graminis</em> f. sp. <em>tritici</em></td>
<td>Wheat</td>
<td>AvrSr35; AvrPm3b2/c2, AvrPm3d3</td>
<td>Sr35, Sr50</td>
<td>Close to repetitive elements</td>
<td>(Chen et al., 2017; Salcedo et al., 2019)</td>
</tr>
<tr>
<td><em>Blumeria graminis</em> f. sp. <em>hordei</em></td>
<td>Barley</td>
<td>AvrK1, AvrA10, AvrA1, AvrA7, AvrA9, AvrA22</td>
<td></td>
<td>Close to repetitive elements</td>
<td>(Bourras et al., 2018; Saur et al., 2019)</td>
</tr>
<tr>
<td><em>Blumeria graminis</em> f. sp. <em>tritici</em></td>
<td>Wheat</td>
<td>AvrPm3a2/b2/c2, AvrPm3b2/c2, AvrPm3d3</td>
<td></td>
<td>Close to repetitive elements</td>
<td>(Bourras et al., 2019, 2018)</td>
</tr>
<tr>
<td><em>V. dahliae</em></td>
<td>Tomato</td>
<td>Avr1</td>
<td></td>
<td>Close to repetitive elements</td>
<td>(Bourras et al., 2019)</td>
</tr>
<tr>
<td><strong>Avirulence proteins determining host cultivar specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M. oryzae</em></td>
<td>Grass species</td>
<td>Pw1, PwT3, PwT4</td>
<td>Rwe1 and Rwe4</td>
<td>–</td>
<td>(Inose et al., 2017; Sweigard et al., 1995; Takabayashi et al., 2002)</td>
</tr>
<tr>
<td><em>B. graminis</em> f. sp. <em>tritici</em></td>
<td>Wheat</td>
<td>AvrPm3b2/c2, AvrPm3d3</td>
<td>Pm3b, Pm3c, Pm3d</td>
<td>Close to repetitive elements</td>
<td>(Bourras et al., 2019)</td>
</tr>
<tr>
<td><strong>Proteinaceous host-specific toxins determining host cultivar specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Stagonospora nodorum</em></td>
<td>Wheat</td>
<td>SnTox1, SnTox2, SrTox3</td>
<td>Sn1, Sn3, Sn3</td>
<td>Repeat-rich regions</td>
<td>(Friesen et al., 2006; Liu et al., 2012, 2009)</td>
</tr>
<tr>
<td><em>Pyrenophora tritici-repentis</em></td>
<td>Wheat</td>
<td>ToxA and ToxB</td>
<td>Tn1 and Tc2</td>
<td>Close to repetitive elements</td>
<td>(Friesen et al., 2006; Martinez et al., 2004)</td>
</tr>
<tr>
<td><strong>Secondary metabolites determining host cultivar specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cochliobolus carbonum</em></td>
<td>Corn</td>
<td>HC-toxin</td>
<td>hm/hm</td>
<td>Repeat-rich region</td>
<td>(Walton, 2006)</td>
</tr>
<tr>
<td><em>C. heterostrophus</em></td>
<td>Corn</td>
<td>T-toxin</td>
<td>Texas <em>male sterile cytoplasm</em> (cms-T)</td>
<td>Repeat-rich regions</td>
<td>(Inderbitzin et al., 2010; Yang et al., 1996)</td>
</tr>
<tr>
<td><strong>Secondary metabolites determining host species specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Verticillium dahliae</em></td>
<td>Cotton, olive and okra</td>
<td>VdJS5 and VdBP6</td>
<td>–</td>
<td>Lineage-specific genomic region</td>
<td>(Chen et al., 2018; Zhang et al., 2019)</td>
</tr>
<tr>
<td><em>Alternaria alternata</em></td>
<td>A number of crops</td>
<td>AM-toxin, AF-toxin, AK-toxin, ACT-toxin, ACR-toxin, AAL-toxin, AT-toxin</td>
<td>–</td>
<td>Accessory chromosome</td>
<td>(Akimitsu et al., 2014; Tsuge et al., 2013)</td>
</tr>
</tbody>
</table>
M. oryzae genes located near the end of chromosomes and/or are surrounded by transposons (Wang et al., 2017). Most genes are located on a dispensable mini-chromosome, which is frequently lost during meiosis (Balesdent et al., 2013).

AVR genes of L. maculans are commonly inactivated by the RPi mechanism which usually targets duplicated sequences (Daverdin et al., 2012). In addition, gene deletion (such as of AvrLm1 and AvrLm11) (Balesdent et al., 2013; Gout et al., 2006) nucleotide polymorphisms (SNPs) (such as in AvrLm7-4 and AvrLm2) (Ghanbarnia et al., 2015; Parlangue et al., 2009), terminal repeat insertion, single base pair deletions, and low expression in planta have contributed to regaining virulence on oilseed rape (Daverdin et al., 2012).

Complex interactions between AVR proteins and R proteins have been demonstrated in L. maculans. For instance, AvrLm1 can be recognized by both Rlm1 and Rlm3 (Gout et al., 2006; Larkan et al., 2013). Similarly, AvrLm4-7 is recognized by Rm4 and Rm7 (Parlangue et al., 2009). Like in Foli, where Avr1 suppresses I-2 and I-3-mediated resistance, AvrLm4-7 suppresses recognition of AvrLm3 and AvrLm5-9 by Rm3 and Rm9, respectively (Ghanbarnia et al., 2018; Plissonneau et al., 2016). AvrLm3 is highly conserved in the fungal population despite its sub-telomeric location, implying its importance for fitness (Plissonneau et al., 2016).

2.4. Magnaporthe oryzae

M. oryzae causes disease on more than 50 grass species, but individual strains can only infect one or a few species. Both host cultivar specificity and host species specificity factors have been identified in M. oryzae (Inoue et al., 2017; Wang et al., 2017), including the Avr-Pita family (Orbach et al., 2000), Avr1-CO39 (Farman et al., 2002; Farman and Leong, 1998), Pwl2 (Sweigard et al., 1995), PWT3/4 (Inoue et al., 2017) and Ace1 (Böhmer et al., 2004). Several Avr factors contribute to host species specificity. Pwl2 functions as a host species specificity factor preventing infection of weeping lovegrass (Sweigard et al., 1995), while PWT3 and PWT4 from Avena-infecting isolates are recognized by the corresponding resistance proteins Rwt3 and Rwt4 in wheat, respectively (Inoue et al., 2017; Takabayashi et al., 2002).

Most M. oryzae AVR genes encode small secreted proteins (less than 200 aa), excepted ACE1 which encodes a polypeptide synthesize/peptide synthetase of 4035 aa potentially involved in the biosynthesis of a secondary metabolite (Wang et al., 2017). The three-dimensional structures of Avr1-CO39 and Avr-Pia have been determined (de Guillen et al., 2015). These two effectors, together with Avr-Piz-t and ToxB, an Avr1-CO39 family (Orbach et al., 2000), Avr1-CO39 (Farman et al., 2002; Farman and Leong, 1998), Pwl2 (Sweigard et al., 1995), PWT3/4 (Inoue et al., 2017) and Ace1 (Böhmer et al., 2004), Several Avr factors contribute to host species specificity. Pwl2 functions as a host species specificity factor preventing infection of weeping lovegrass (Sweigard et al., 1995), while PWT3 and PWT4 from Avena-infecting isolates are recognized by the corresponding resistance proteins Rwt3 and Rwt4 in wheat, respectively (Inoue et al., 2017; Takabayashi et al., 2002).

Most M. oryzae AVR genes encode small secreted proteins (less than 200 aa), excepted ACE1 which encodes a polypeptide synthesize/peptide synthetase of 4035 aa potentially involved in the biosynthesis of a secondary metabolite (Wang et al., 2017). The three-dimensional structures of Avr1-CO39 and Avr-Pia have been determined (de Guillen et al., 2015). These two effectors, together with Avr-Piz-t and ToxB, an effector from the wheat tan spot pathogen Pyrenophora tritici-repentis (Pr), have the same structures, and they are named MAX-effectors (Magnaporthe Avrs and ToxB like) (de Guillen et al., 2015). Most AVR genes are located near the end of chromosomes and/or are surrounded by transposons (Wang et al., 2017).

Diverse mechanisms contribute to loss of the avirulence function of effectors in M. oryzae. For example, point mutations and insertions in and deletions of Avr-Pita permit the fungus to avoid triggering a resistance response mediated by Pita (Orbach et al., 2000). In addition, transposon insertion in Avr-Pi9 (Wu et al., 2015), Ace1 (Fudal et al., 2005), Avr-Pi-zt (Li et al., 2009) and Avr-Pib (Zhang et al., 2015), and segmental deletion in Avr-Pib (Zhang et al., 2015) render the pathogen virulent.

The molecular interactions of seven pairs of M. oryzae AVR and rice R proteins have been studied, including Pi-ta/Avr-Pita, Piz-t/Avr-Piz-t, Pik/Avr-Pik, Pia/Avr-Pia, Pi-CO39/Avr1-CO39, PiS4/Avr-PiS4, and Pii/Avr-Pii (Wang et al., 2017). Three modes of interaction have been demonstrated, including a direct interaction between one R and one AVR protein, i.e., PiS4/Avr-PiS4 (Ray et al., 2016), two R proteins interacting with one AVR protein, i.e., Pi-k1/Pi-k2/Avr-Pik (Aishikawa et al., 2008), and two cooperating R proteins recognizing two AVR proteins, i.e., RGBA RGBA5/Avr-Pia/Avr1-CO39 (Cesari et al., 2013; Okuyama et al., 2011).

2.5. Rhynchosporium secalis

R. secalis is the causal agent of leaf scald on barley. Three low molecular weight necrosis-inducing peptides (NIPs), designated Nip1 to Nip3, function as non-specific toxins on barley (Wevelsiepe et al., 1991). Nip1 is also a race-specific elicitor of defense responses in barley cultivars carrying the resistance gene Rss1. The amino acid sequence encoded by NIP1 contains a secretory signal peptide and a cysteine-rich mature protein of 60 residues (Rohe et al., 1995). Strains of R. secalis virulent on Rss1-containing plants either lack NIP1 or carry alleles with point mutations that translate into single amino acid substitutions (Rohe et al., 1995; Schürch et al., 2004). NIP1 was lost with a high frequency (45%) among 614 isolates from different geographic populations on four continents, and 14 types of DNA polymorphisms were found, indicating diversifying selection (Schürch et al., 2004). A recent study shows that the NIP1 gene family evolved mainly through point mutations and copy number variation (Mohd-Assaad et al., 2019). NIP1 is present on a large chromosome that is not likely to be dispensable (von Felten et al., 2011).

2.6. Melampsora lini

The flax rust fungus M. lini is an obligate biotrophic basidiomycete that infects flax (Linum usitatissimum) and other species of the genus Linum (Lawrence et al., 2007). To date, avirulence genes have been identified in six loci in M. lini, encoding AvrL567 (Dodds et al., 2004), AvrM (Catanzariti et al., 2006a), AvrP123 (Catanzariti et al., 2006b), AvrP4 (Catanzariti et al., 2006b), AvrL2-A (Anderson et al., 2016) and AvrM14-A (Anderson et al., 2016), respectively. AvrL567-A is recognized by resistance proteins L5, L6 and L7, whereas AvrL567-B is recognized most strongly by L5, weakly by L6 and not at all by L7 (Dodds et al., 2004). The gene encoding AvrL567-C co-segregates with the virulence phenotype, and this version is not recognized by L5, L6, or L7 (Dodds et al., 2004). AvrP4 and AvrP123 are cysteine-rich proteins, whereas AvrM does not contain cysteine residues at all. AvrL2-A and AvrM14-A were identified by map-based cloning (Anderson et al., 2016). AvrM14-A is not related to AvrM and is recognized by both the flax M1 and M4 resistance proteins (Anderson et al., 2016). AvrM14-A shows homology with the nudix hydrolase superfamily and is the first rust avirulence protein for which a biochemical function could be predicted from the protein sequence (Anderson et al., 2016). The AvrL2 protein family has no homology to proteins with known function.

The three-dimensional structures of AvrL567, AvrM and AvrP have been determined. They exhibit completely different structures, but they all display surface polymorphic residues involved in the recognition by the flax resistance proteins L5/L6/L7, M and P, respectively (Lorrain et al., 2019; Ve et al., 2013; Wang et al., 2007; Zhang et al., 2018). Except for AvrL2-A, all avirulence genes identified are located at recombination hot-spots (Anderson et al., 2016). AvrL2-A has been proposed to be located in a centromere or other heterochromatin repeat-rich region (Anderson et al., 2016). All the identified genes have undergone diversifying selection (Anderson et al., 2016; Barrett et al., 2009; Ellis et al., 2007).

2.7. Puccinia graminis

Puccinia graminis f. sp. tritici (Pgt) causes wheat stem rust, which has posed a threat to wheat production recently (Singh et al., 2011). Two avirulence factors have been identified in Pgt, AvrSr35 and AvrSr50, which are recognized by resistance proteins Sr35 and Sr50, respectively (Chen et al., 2017; Salcedo et al., 2017). AvrSr50 encodes a 132-amino
acid protein which interacts with Sr50 directly (Chen et al., 2017; Salcedo et al., 2017). The origin of isolates virulent on Sr50-containing plants is associated with the insertion of a miniature inverted transposable element (MITE) in AvrSr35 (Salcedo et al., 2017). A switch to virulence towards Sr50 was due to the exchange of a whole chromosome between two haploid nuclei, resulting in loss of the avirulence allele (Chen et al., 2017). In addition, the Pgt protein PTAUSPE-10-1 causes cell death in a host line carrying resistance gene Sr22. Therefore, PTAUSPE-10-1 might be the avirulence factor corresponding to Sr22 (Upadhyaya et al., 2014). Through mutagenic genomics approaches, AvrSr27 has also been identified, and identification of AvrSr5 is underway (Dodds et al., 2019).

2.8. Blumeria graminis

2.8.1. B. graminis f. sp. hordei (Bgh)

B. graminis f. sp. hordei (Bgh) causes powdery mildew on barley, and interacts with its host in a gene-for-gene manner (Zhang et al., 2005). So far, seven avirulence genes have been identified in Bgh, namely AvrK1, AvrA1, AvrA13, AvrA7, AvrA9, AvrA10 and AvrA22, corresponding to barley resistance genes mlk1, mla13, mla7, mla9, mla10 and mla22, respectively (Lu et al., 2016; Ridout et al., 2006; Saur et al., 2019). AvrA7, AvrA9, AvrA10, and AvrA22 were identified in a recent study and these proteins interact directly with their respective plant R proteins (Saur et al., 2019). Virulence of isolates is predominately associated with non-synonymous SNPs and loss of expression of avirulence genes (Saur et al., 2019). AvrA10 and AvrA22 share homology and associated with non-synonymous SNPs and loss of expression of avirulence genes (Saur et al., 2019). Although the gene is required for the defoliation of cotton and olive. Application of this heterologously produced protein to cotton seedlings also induced defoliation, indicating that the protein is directly responsible for the defoliation symptoms.

3. Virulence genes determining host cultivar specificity

3.1. Stagonospora nodorum

The ascomycete fungus S. nodorum (also known as Parastagonospora nodorum) is a major necrotrophic pathogen of wheat causing leaf and glume blotch (Liu et al., 2009). So far, three S. nodorum genes encoding host selective toxins, notably SnToxA (Liu et al., 2012), SnTox1 (Liu et al., 2009) and SnTox3 (Liu et al., 2009) have been identified. The genes reside on three different chromosomes (Liu et al., 2012). The toxins induce cell death and necrosis as an outcome of their interaction with their cognate dominant susceptibility gene products (ToxA-Tsn1, Tox1-Snn1 and Tox3-Snn3). Below these host-specific interactions are summarized.

3.1.1. SnTox1-Snn1

Although the SnTox1-Snn1 interaction was the first to be characterized (Liu et al., 2004), the toxin gene was identified only in 2012 (Liu et al., 2012). The mature SnTox1 contains 100 amino acids including 16 cysteine residues, all predicted to be involved in disulfide bridges necessary for the activity/stability of the protein. The C-terminus shows similarity to the chitin-binding domain of Avr4 of C. fulvum (van Esse et al., 2007). Later, it was shown that SnTox1 binds chitin of the fungal cell wall, protecting the pathogen from chitinase degradation (Liu et al., 2016). SnTox1 is highly expressed at 3 days post infection (dpi), which correlates with the onset of necrotic lesion development. Different from other effector genes in S. nodorum, which are located in gene poor and repeat-rich regions, SnTox1 is located in a gene-rich region and no obvious repeats or AT-rich sequences were identified within the 300 kb region containing SnTox1. Among 159 Sn isolates from around the globe, 11 Tox1 isoforms were found, suggesting diversifying selection on SnTox1 (Liu et al., 2012).

3.1.2. SnToxA-Tsn1 and PtrToxA-Tsn1

SnToxA is highly similar to PtrToxA identified in P. triticis-repentis (Ciuffetti et al., 1997), having only two amino acid differences, and both are recognized by the wheat protein Tsn1 (Faris et al., 2010). Mature SnToxA is a 13.2 kDa protein containing two cysteine residues as well as an RGD-containing vitronectin-like motif that is present in a solvent-exposed loop in the active protein (van Esse et al., 2007). The regions upstream and downstream of SnToxA contain repetitive, AT-rich regions, but these AT-rich regions have not been found in the corresponding flanking regions of P. triticis-repentis (Friesen et al., 2006). By sequencing 95 S. nodorum ToxA and 54 P. triticis-repentis ToxA species, such as tomato (Klosterman et al., 2009). So far, identification of only one avirulence protein, Ave1, has been published. Ave1 is a 134 aa secreted protein recognized by Ve1 in tomato, and expression of Ave1 is induced during host colonization (de Jonge et al., 2012). Intriguingly, no SNP was found in 85 Ave1 alleles from Verticillium strains isolated from various host plants and different geographical locations (de Jonge et al., 2012). The presence of numerous Ave1 orthologs in plants, absence of orthologs in fungi other than Fol, Colletotrichum higginsianum, and Cercospora beticola, and the association of Ave1 with a flexible genomic region containing various TEs suggest that Verticillium acquired Ave1 from plants through horizontal gene transfer (HGT) (de Jonge et al., 2012). In a recently study, Ave1 was shown to play a role in niche colonization by suppressing microbes with antagonistic activities (Snelders et al., 2020).

In another study of V. dahliae (personal communication with Jinling Li, unpublished data), a duplicated defoliation-specific gene, encoding a small secreted protein, was found. It was demonstrated that this effector gene is required for the defoliation of cotton and olive. Application of this heterologously produced protein to cotton seedlings also induced defoliation, indicating that the protein is directly responsible for the defoliation symptoms.

2.9. Verticillium dahliae

V. dahliae is an asexual soil-borne, xylem-invading plant pathogen that causes vascular wilt diseases in over 200 dicotyledonous plant species such as tomato (Klosterman et al., 2009). So far, identification of only one avirulence protein, Ave1, has been published. Ave1 is a 134 aa secreted protein recognized by Ve1 in tomato, and expression of Ave1 is induced during host colonization (de Jonge et al., 2012). Intriguingly, no SNP was found in 85 Ave1 alleles from Verticillium strains isolated from various host plants and different geographical locations (de Jonge et al., 2012). The presence of numerous Ave1 orthologs in plants, absence of orthologs in fungi other than Fol, Colletotrichum higginsianum, and Cercospora beticola, and the association of Ave1 with a flexible genomic region containing various TEs suggest that Verticillium acquired Ave1 from plants through horizontal gene transfer (HGT) (de Jonge et al., 2012). In a recently study, Ave1 was shown to play a role in niche colonization by suppressing microbes with antagonistic activities (Snelders et al., 2020).

In another study of V. dahliae (personal communication with Jinling Li, unpublished data), a duplicated defoliation-specific gene, encoding a small secreted protein, was found. It was demonstrated that this effector gene is required for the defoliation of cotton and olive. Application of this heterologously produced protein to cotton seedlings also induced defoliation, indicating that the protein is directly responsible for the defoliation symptoms.
amplicons from geographically diverse populations, 11 haplotypes were found in *S. nodorum* and only one haplotype in *P. tritici-repentis* (Friesen et al., 2006). The gene has likely undergone HGT from *S. nodorum* to *P. tritici-repentis* based on the following observations: (1) the presence of an almost identical 11 kb region containing SnToxA/PtrToxA in both species; (2) the presence of high sequence diversity of SnToxA in *S. nodorum* and monomorphism in Ptr; (3) absence of PtrToxA in related species; (4) the recent emergence of tan spot (Friesen et al., 2006; Liu et al., 2006).

### 3.1.3. SnTox3-Snn3
SnTox3 was identified by partial purification and sequencing of the protein (Liu et al., 2009). SnTox3 encodes a 230 amino acid pre-protein consisting of a 20 amino acid signal sequence and a predicted pro-domain of approximately 30 amino acids, resulting in a mature protein of ~18 kDa (Liu et al., 2009). SnTox3 contains six cysteine residues, each being predicted to be involved in the formation of a disulphide bridge critical to the structure and function of the protein (Liu et al., 2009). Like SnTox1, SnTox3 is highly expressed at 3 dpi when lesions start to develop. SnTox3 interacts with a wheat pathogenicity-related-1 protein (Breen et al., 2016). SnTox3 is also flanked by AT-rich sequences, containing long terminal repeat retrotransposons. By sequencing SnTox3 from 245 isolates, eleven haplotypes were identified resulting in four amino acid polymorphisms (Liu et al., 2009).

### 3.2. *Pyrenophora tritici-repentis* (*P. tritici-repentis*)

The ascomycete *Pyrenophora tritici-repentis* (Ptr) causes tan spot and chlorosis on wheat. Two Ptr host-specific genes encoding host-specific toxins have been identified, including *Ptr-ToxA* (Ciuffetti et al., 1997) and *Ptr-ToxB* (Strelkov et al., 1999). Ptr-ToxA and Ptr-ToxB interact specifically with the products of the host susceptibility genes *Tsn1* and *Tsc2*, respectively. It is likely that *Ptr-ToxA* originated from *S. nodorum* (see above) (Friesen et al., 2006). Ptr-ToxA is a single domain protein having a β-sandwich fold with two antiparallel β-sheets composed of four strands each enclosing the hydrophobic core (Sarma et al., 2005). Unlike *Ptr-ToxA*, which is present as a single copy in the genome, *Ptr-ToxB* is present in multiple copies in the genome of some races of *Ptr*, and the number of copies is proportional to virulence (Martinez et al., 2004). *Ptr-ToxB* encodes a 64 amino acid host-selective toxin (Martinez et al., 2001), contains four cysteine residues involved in the formation of two disulphide bridges (Nyaroko et al., 2014) and is a heat-stable protein (Strelkov et al., 1999). All *Ptr-ToxB* loci are associated with retrotransposons (Martinez et al., 2004). *Ptr-tox* is a related single copy gene from a non-pathogenic strain and shares 86% identity with *Ptr-ToxB* (Martinez et al., 2004). Both *Ptr-ToxB* and *Ptr-tox* adopt a β-sandwich fold stabilized by two disulphide bonds, but differ in the dynamics of one sandwich half. The absence of toxic activity of *Ptr-tox* is attributable to the more open structure close to one disulphide bond, higher flexibility, and different residues in an exposed loop of *Ptr-tox* (Nyaroko et al., 2014).

### 4. Secondary metabolites determining host specificity

#### 4.1. *Cochliobolus carbonum*

*Cochliobolus carbonum* (*C. carbonum*) causes Northern Corn Leaf Spot and is virulent on *hm/hm* corn (Johal and Briggs, 1992). The virulence of the fungus is due to production of HC-toxin, a cyclic tetrapeptide (Panaccione et al., 1992). *Hm1* and *Hm2* both encode a carboxyl reductase that inactivates HC-toxin (Johal and Briggs, 1992). It has been hypothesized that HC-toxin alters the expression of plant defense genes by inhibiting histone deacetylases (Ransom and Walton, 1997; Walton, 2006). The production of HC-toxin is governed by the single locus *TOX2* (Panaccione et al., 1992). It contains *HST1*, encoding the 570 kDa non-ribosomal peptide synthetase key enzyme. *TOX2* is duplicated in toxin-producing isolates of the fungus, but is completely absent from the genomes of toxin-non-producing isolates (Panaccione et al., 1992). Disruption of all copies of *HST1* resulted in abolished HC-toxin production and loss of host-selective pathogenicity. Other genes involved in the processing of biosynthetic intermediates and possibly secretion include *TOXA* (major facilitator superfamily transporter) (Pitkin et al., 1996), *TOXC* (fatty acid synthase subunit) (Ahn and Walton, 1997) and *TOXF* (amino acid transaminase) (Cheng et al., 1999). Except for one copy of the pathway-specific regulator gene *TOXE*, all other TOX2 genes are located in an approximately 600 kb repeat-rich region (Ahn and Walton, 1996; Condon et al., 2013).

#### 4.2. *Cochliobolus heterostrophus*

*Cochliobolus heterostrophus* (*C. heterostrophus*) causes Southern Corn Leaf Blight. There are two known races, race T and race O, with race T producing a HST called T-toxin. Race T is highly virulent on corn carrying Texas male sterile cytoplasm, while Race O only shows mild virulence. So far, nine genes have been shown to be required for T-toxin production (Baker et al., 2006; Inderbitzin et al., 2010; Rose et al., 2002; Yang et al., 1996), and these genes are located at two unlinked loci, designated as *ToxA1* and *ToxB1*. Appropriately, these genes are absent in Race O. The nine known *Tox1* genes encode two polyketide synthases (PKS), a decarboxylase, five dehydrogenases and an unknown protein (Inderbitzin et al., 2010). The genes do not reside in a single cluster, but reside alone or in small groups in four distinct AT-rich regions (Inderbitzin et al., 2010). Together, these *Tox1* regions comprise less than 5% of the 1.2 Mb of race T-specific DNA (Inderbitzin et al., 2010).

#### 4.3. *Verticillium dahliae*

As mentioned above, *V. dahliae* is a vascular wilt pathogen which can infect nearly 200 plant species. However, it causes defoliation in a few hosts only, including cotton, olive and okra (Milgroom et al., 2016). Recently, by using comparative genomics, seven genes associated with the defoliation pathotype, *VdDf1-VdDf7*, were discovered in a lineage-specific genomic region (G-LSR2) of *V. dahliae* Vd991 (Chen et al., 2018; Zhang et al., 2019). *VdDf5* and *VdDf6* are critical for the defoliation phenotype (Zhang et al., 2019). The VdDfs are involved in the production of NAE 12:0, that cause defoliation either by altering abscisic acid sensitivity, hormone disruption or sensitivity to the pathogen (Zhang et al., 2019). Phylogenetic analysis of the region comprising all seven protein-coding genes suggests that G-LSR2 has been acquired from *F. oxysporum* *f. sp. vasinfectum* through horizontal transfer (Chen et al., 2018).

#### 4.4. *Alternaria alternata*

*Alternaria alternata* (*A. alternata*) is a ubiquitous, mostly saprophytic fungus present in dead plant material but is also known as a weak pathogen causing opportunistic diseases in a number of crops (Akimitsu et al., 2014). HSTs in *A. alternata* have been extensively reviewed before (Akimitsu et al., 2014; Meena et al., 2017). So far, there are seven known diseases caused by *A. alternata* in which HSTs are responsible for pathogenesis. Accordingly, *A. alternata* is classified into seven different pathotypes, each producing a distinct host-specific toxin: AM-toxin (apple pathotype), AF-toxin (strawberry pathotype), AK-toxin (Japanese pear pathotype), ACT-toxin (tangerine pathotype), ACR-toxin (rough lemon pathotype), AAL-toxin (tomato pathotype), or AT-toxin (tobacco pathotype). Interestingly, HST genes in *A. alternata* are located on conditionally dispensable chromosomes (CDCs) as gene clusters (Hu et al., 2012). Chemical structures of HSTs from six pathotypes have been determined, excluding that of AT-toxin of the tobacco pathotype. AAL-toxin is a structural analog of sphingolipid precursors, and thereby efficiently inhibits eukaryotic ceramide synthases. Appropriately,
5. Horizontal gene and chromosome transfer and host specificity

HGT or horizontal chromosome transfer (HCT) has likely occurred in several plant pathogenic fungi, including Colletotrichum gloeosporioides (He et al., 1998), S. nodorum/P. tritic-repentis (Friesen et al., 2006), F. oxysporum (Ma et al., 2010; van Dam et al., 2017; Vlaardingerbroek et al., 2016) and A. alternata (Akimitsu et al., 2014). The outbreak of tan spot disease on wheat in 1941 was caused by P. tritici-repentis. The high virulence of P. tritici-repentis is due to the presence of a HST, ToxA, and, as mentioned above, it is likely that the ToxA gene was transferred from S. nodorum to P. tritici-repentis in a recent event (Friesen et al., 2006).

In F. oxysporum, HCT of pathogenicity chromosomes has been shown in three formae speciales: Fol (Ma et al., 2010; Vlaardingerbroek et al., 2016), F. oxysporum f. sp. radicis-cucumerinum (Forc) (van Dam et al., 2017) and F. oxysporum f. sp. melonis (Forc) (Li et al., 2020). In all three cases, following transfer of an accessory chromosome containing effector genes from a pathogenic strain to non-pathogenic F. oxysporum, the recipient strain becomes pathogenic to the host species of the chromosome donor strain.

As mentioned above, in A. alternata, all HST-encoding genes are located on CDCs (Mehrabi et al., 2011). It has been suggested that these CDCs are transferable between different pathotypes in A. alternata (Akagi et al., 2009). Indeed, HCT between different pathotypes has been demonstrated through protoplast fusion experiments. For example, by fusion of a tomato pathotype with a strawberry pathotype, the resulting strain was found to be pathogenic on both tomato and strawberry (Akagi et al., 2009).

Often, virulence or host-determining genes are clustered, so HGT or HCT can cause previously non-pathogenic microbes to become pathogenic, or pathogens to expand or change host range. In A. alternata and F. oxysporum, genes determining host range – a secondary metabolite gene cluster or virulence genes together with transcription factors – are located on a single chromosome, therefore single chromosome transfer is sufficient to expand host range. This may be an important mechanism for asexual fungi to generate genetic variation and adapt to a changing environment.

6. Potential role of genome localization of host-specificity genes in adaptation

From the findings discussed above, host-specificity genes appear to be predominately located in AT-rich isochromes (such as avirulence genes in L. maculans and AvrPm3b1/2, AvrPm3b2/2, and AvrPm3b3 in B. graminis f. sp. tritici), and/or TE-rich lineage-specific chromosomes (such as AVR1, AVR2, and AVR3 in Fol). These repeat-rich genomic compartments are believed to evolve more rapidly than other parts of the genome. Based on this, the concept of a ‘two-speed genome’ has been proposed (Croll and McDonald, 2012; Raffaele and Kamoun, 2012). The gene-poor, repeat-rich genomic compartment could serve as a ‘cradle’ for adaptive evolution (Croll and McDonald, 2012). Evidence for a higher diversification rate has been found for almost all the host-specificity genes listed above. For example, in L. maculans, avirulence genes are located in AT-rich isochromes and these genes have an exceptionally high mutation rate due to RIP (Rouxel and Balesdent, 2017).

Modern crop management practices have accelerated the arms race between pathogens and plants (Möller and Stukenbrock, 2017). Cultivars with new resistance genes may be introduced in each growing season, which poses great pressure on pathogens to rapidly adapt to the new cultivar. Since avirulence factors that are recognized by resistance proteins are generally located in repeat-rich genomic regions, a high frequency of point mutations, deletions, duplications, silencing or rearrangement of avirulence genes can result in rapid emergence of strains that evade recognition by resistance proteins. If loss of or changes in avirulence genes have no or little fitness cost, these strains will quickly become dominant in the pathogen population. The compartmentalization of a genome thus allows pathogens to harbor fast-evolving genes without affecting the stability of core genes.

7. Conclusions and perspectives

Both proteins and secondary metabolites can determine host specificity in plant pathogenic fungi. With an increased rate of publication of pathogen genome sequences, the rate of discovery of genes determining host cultivar-specificity will likewise increase, by comparing strains of different races within a species or forma specialis. Identification of host species-specificity factors may require more effort, depending on the fungal species involved. Regarding the latter, it would be interesting to test the hypothesis of Schulze-Lefert and Panstruga, who proposed that PTI plays a major role in non-host resistance of evolutionary divergent non-host plant species, while ETI would be more dominant as a non-host resistance mechanism in more closely related plant species (Schulze-Lefert and Panstruga, 2011). To test whether ETI indeed plays a dominant role in non-host resistance of closely related plant species, we propose to identify effectors eliciting an ETI response in non-host plant species closely related to host species in fungal pathogens such as M. oryzae or F. oxysporum.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

J.L. was financially supported by the China Scholarship Council program (File number: 201504910768). The authors thank Slavica Janevska for providing comments on the manuscript and both reviewers for their constructive, critical and detailed comments.

Author contributions

Conceptualization: JL and MR; Investigation: JL; Writing of the original draft: JL; Review & editing: JL, BC, and MR.

References


Rose, M.S., Yun, S.-H., Asvarak, T., Lu, S.-W., Yoder, O.C., Turgeon, B.G., 2002. A family of tomato expansins are induced by treatment with HC-toxin or infection by the filamentous fungus Cochliobolus car-


