Molecular genetics and evolution of disease resistance in cereals

Simon G. Krattinger and Beat Keller
Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

Acknowledgements

References

Key words: cereals, disease resistance, leucine-rich repeat receptor (NLR), nucleotide-binding, pathogen, quantitative disease resistance.

Summary

Cereal crops produce a large part of the globally consumed food and feed. Because of the constant presence of devastating pathogens, the molecular characterization of disease resistance is a major research area and highly relevant for breeding. There has been recent and accelerating progress in the understanding of three distinct resistance mechanisms in cereals: resistance conferred by plasma membrane-localized receptor proteins; race-specific resistance conferred by intracellular immune receptors; and quantitative disease resistance. Intracellular immune receptors provide a particularly rich source for evolutionary studies, and have, for example, resulted in the recent discovery of a novel detection mechanism based on integrated decoy domains. Evolutionary studies have also revealed the origins of active resistance genes in both wild progenitors of today’s cereals as well as in cultivated forms. In addition, independent evolution of orthologous genes in related cereals has resulted in resistance to different pathogen species. Quantitative resistance genes have been best characterized in wheat. The quantitative resistance genes identified so far in wheat encode transporter proteins or unusual kinase proteins. The recent discoveries in these three different resistance mechanisms have contributed to the basic molecular understanding of cereal immunity against pathogens and have suggested novel applications for resistance breeding.

I. Introduction

Plants live in close association with a vast number of diverse microbial organisms, including viruses, bacteria and fungi. For the plant, these interactions can have different outcomes, ranging from beneficial, as in the case of rhizobacteria that fix nitrogen, to harmful, as in the case of pathogens. Hence, plants have evolved amazing strategies that allow them to perceive harmful microbes without compromising the ability to associate with symbionts. Damage caused by plant pathogens is of particular importance in...
agriculture, where diseases are a major reason for crop losses. In addition, secondary metabolites produced by certain fungal pathogens, for example *Fusarium* and *Aspergillus*, can have toxic effects on livestock and humans. Cereal crops, including wheat, rice, maize, barley, millet, sorghum, oat and rye, play a pivotal role in global food and feed production. Indeed, most calories consumed today are directly or indirectly derived from cereals: directly through the consumption of cereal grains and indirectly through animal feed. All cereals are close relatives that belong to the grass family (Poaceae) and share their last common ancestor at 45–60 million yr ago (The International Brachypodium Initiative, 2010). It is estimated that, on average, 10–15% of the global crop production is lost to plant diseases (Chakraborty & Newton, 2011; Fisher et al., 2012). For cereals alone, a 10% loss amounts to a total of c. 300 million tons each year (the statistics division of the Food and Agricultural Organization of the United Nations (FAOstat)). If all cereal diseases could be eliminated, this would result in additional food for 1.7 billion people, assuming a consumption of c. 170 kg of cereals per capita per year (Food and Agricultural Organization of the United Nations (FAO)). A better understanding of the molecular genetics and evolution of natural disease resistance in cereals is therefore an important prerequisite to continuously improve cereal cultivars for effective field resistance.

### II. Immunity in plants

Plants lack an adaptive immune system comparable with that found in mammals. Instead, plants rely solely on innate immune mechanisms that are mostly based on the perception of pathogen-derived ligands by plant receptor proteins. Intensive research efforts during the past two decades have led to the identification of a plethora of immune receptors in plants, which can be broadly classified into two distinct types according to their subcellular localization: (1) plasma membrane-localized receptors with an extracellular ligand-binding domain; and (2) intracellular immune receptors (Jones & Dangl, 2006; Dodds & Rathjen, 2010; Thomma et al., 2011; Cook et al., 2015). Most surface-localized immune receptor proteins perceive the presence of pathogen structures in the apoplast. These can comprise highly conserved microbial signatures that are characteristic for entire pathogen classes. Examples of such ‘pathogen-associated molecular patterns’ (PAMPs) are the bacterial flagella or the fungal cell wall component chitin (Jones & Dangl, 2006; Dodds & Rathjen, 2010). Other plasma membrane-localized immune receptors perceive the presence of ligands that are specific to a single or a few related pathogen species (Thomma et al., 2011; Cook et al., 2015), and others again are able to bind host plant-derived molecules that are released during pathogen penetration. The recognition of an appropriate ligand by the corresponding receptor protein activates a defense response whose magnitude can range from a weak basal resistance to death of the attacked cell. We discuss examples of surface-localized immune receptors in cereals in Section III.

To successfully colonize a host plant, pathogens rely on a set of virulence effectors that is tailored to specifically disarm components of the host plant’s immune response. Virulence effectors are often small secreted proteins that can modify or degrade host proteins involved in basal immunity. The magnitude and diversity of a pathogen’s effector set defines its host range, that is the number of host species it can infect (Schulze-Lefert & Panstruga, 2011). Although some pathogens have a very narrow host range, others are able to infect many different species. Many virulence effectors are injected into the plant cell and therefore circumvent perception by surface-localized immune receptors.

In response to cytoplasmic virulence effectors, plants have evolved a second line of defense. This second tier is formed by intracellular immune receptors mostly belonging to the conserved protein family of nucleotide-binding, leucine-rich repeat receptor (NLR) proteins (Dodds & Rathjen, 2010). Direct or indirect effector perception by NLRs generally triggers a strong hypersensitive response (HR) that often leads to the death of the infected cell. It was mainly research performed in the dicotyledonous model plant *Arabidopsis thaliana* (mouse-ear-cress) that led to the initial discovery and characterization of surface-localized immune receptors and NLR proteins. Numerous studies have documented that these proteins also play an important role in cereals and we discuss recent insights into NLR evolution in cereals in Sections IV and V.

In addition, observations made in crops, particularly in cereals, have revealed resistance mechanisms that are different from those described above. Such resistance mechanisms are often collectively referred to as ‘quantitative’. Some quantitative disease resistance (QR) genes provide durable, albeit partial, field resistance against most or all races of a pathogen (St Clair, 2010; Ellis et al., 2014; Niks et al., 2015). The term ‘durable’ refers to disease resistance that is effective over a long period of time in an environment favorable to the disease (Johnson, 1984), and the fact that QR genes often confer resistance to all races of a pathogen species, and sometimes even against multiple pathogens, is referred to as ‘broad spectrum’. The durability, broad-spectrum specificity and lack of HR clearly differentiate such QR genes from NLR-triggered immunity. A number of public cereal breeding programs nowadays favor this resistance type, but there are still many breeding efforts based on NLR-based disease resistance. Repeated phenotypic observations by experienced pathologists in different locations and over multiple years are necessary to reliably quantify the sometimes subtle phenotypic effects of single QR genes. This can explain why this particular resistance type has so far mainly been described in crops and not in model plants. In contrast with its importance in breeding, research on the genetic, molecular and evolutionary basis of QR is still in its infancy.

### III. Receptors in the plant plasma membrane – the importance of monitoring the extracellular space

The extracellular environment plays a central role in host–pathogen interactions because it usually is the first point of contact between the two organisms. It is therefore not surprising that the extracellular space is closely monitored for the presence of pathogen-derived ‘non-self’ signatures by cell surface-localized receptor-like kinases (RLKs) and receptor-like proteins (RLPs) (Jones & Dangl, 2006; Dodds & Rathjen, 2010; Thomma et al., 2011; Cook et al., 2015). Similar receptor proteins, named Toll-like receptors (TLRs), are also involved in innate immunity in
humans (Nurnberger et al., 2004). In contrast with humans, however, RLKs and RLPs massively increased in number during plant evolution, highlighting the importance of the innate immune system in plants. Although humans only have 10 TLR genes, more than 100 genes encoding RLKs and RLPs are usually found in plant genomes (Schwessinger & Ronald, 2012). One of the first plant RLKs identified was XA21 of rice (Song et al., 1995) which confers resistance against most strains of the agronomically important bacterial blight disease (Xanthomonas oryzae pv. oryzae). Xa21 consists of an extracellular leucine-rich repeat (LRR) domain and a cytoplasmic kinase domain. LRR domains have been shown to be involved in the perception of proteinaceous elicitors. A recent key paper indeed identified a sulfated Xanthomonas peptide named ‘required for activation of XA21’ (RaxX) as an elicitor of XA21. Interestingly, the closest homolog of RaxX was not found in other bacteria (Pruitt et al., 2015). Instead, RaxX shows homology to a sulfated signaling peptide of Arabidopsis, which led to the conclusion that RaxX might mimic plant proteins involved in basal immunity.

Although extracellular LRRs mainly perceive peptide elicitors, other non-LRR domains found in RLKs and RLPs recognize different elicitor classes (Schwessinger & Ronald, 2012). For example, the extracellular domains of the rice RLP CEBiP (chitin elicitor-binding protein) and the RLK CERK1 contain LysM motifs. These two proteins cooperatively perceive the fungal cell wall component chitin (Kaku et al., 2006; Shimizu et al., 2010). In addition, several wall-associated receptor-like kinases (WAKs) have been shown recently to confer resistance against fungal disease in rice and maize. Fine mapping of the head smut (Sporisorium reilianum) resistance locus qHSR1 in maize identified ZmWAK as the causal resistance protein (Zuo et al., 2015). Head smut is a soil-borne fungus that infects maize roots and subsequently spreads to above-ground parts where symptoms develop. Interestingly, ZmWAK did not repress root penetration of the fungus and maize lines with and without qHSR1 showed similar levels of fungal growth in roots. Instead, ZmWAK was mainly expressed in the mesocotyl, where it repressed spread of the fungus to the above-ground parts. It is possible that strong expression of ZmWAK in roots would compromise the ability of maize plants to interact with beneficial arbuscular mycorrhizal fungi and that is why maize evolved this amazing resistance mechanism that allows for a certain degree of head smut infection. Similar to qHSR1, Hurni et al. (2015) identified the WAK-encoding gene Hn1 in maize that confers resistance against the fungal disease northern corn leaf blight (Esroehilum turcicum). In addition to maize, several WAKs in rice have recently been reported to positively or negatively influence basal defense against the fungal rice blast disease (Magnaporthe oryzae) (Li et al., 2009; Delteil et al., 2016). Interestingly, although only 27 WAK-like genes are found in the model plant Arabidopsis, the rice genome contains > 120 WAK-like genes.

Kinase domains can be classified into RD and non-RD based on the presence of a conserved arginine residue (R) at the catalytic site. The fusion of WAK domains to non-RD kinase domains is specific to monocots (de Oliveira et al., 2014). Although RD kinases are often associated with developmental processes, non-RD kinases are typically found in plant immune receptors (Dardick et al., 2012). Hence, the WAK gene family in monocots has expanded, diversified and might play a very important role in fungal disease resistance, specifically in cereals. Delteil et al. (2016) hypothesized that the rice blast resistance-conferring WAK proteins OsWAK14, OsWAK91, OsWAK92 and OsWAK112d trigger a basal defense response on perception of the fungal cell wall component chitin.

Some RLKs and RLPs perceive highly conserved ligands and can confer resistance against a broad range of a particular group of pathogen. The transfer of such immune receptors between species could therefore enhance the durability of disease resistance. Indeed, the functional transfer of RLKs has been demonstrated recently in several cases. The rice XA21 protein, for example, is functional in banana, where it confers resistance against the bacterial banana Xanthomonas wilt disease (Tripathi et al., 2014). Likewise, the Arabidopsis LRR-RLK EFR, which perceives a bacterial elongation factor, is functional in transgenic wheat against the halo blight-conferring bacterial pathogen Pseudomonas syringae pv. oryzae (Schoonbeek et al., 2015). These examples provide the first experimental evidence that the transfer of surface-localized immune receptors across species is a feasible strategy to improve disease resistance in cereals. In addition, targeted modifications of endogenous RLKs and RLPs through genome editing might also be used to increase their resistance spectra or durability. The molecular differences between susceptible and resistant alleles might be used as a starting point to identify critical domains and amino acids for recognition.

**IV. Inside the cell – the role of intracellular immune receptors in cereals**

The involvement of intracellular receptors, mostly NLRs, in effector recognition is probably the best studied area of plant immunity. A complete summary of NLR-based immunity in cereals would go beyond the scope of this review. Rather, we focus on the evolutionary mechanisms that have shaped NLR specificity in cereals. Specifically, we discuss NLR evolution in the Triticeae species wheat, barley and rye. This tribe of the grass family is particularly interesting for evolutionary studies because it contains several closely related crop species and different ploidy levels.

Most secreted virulence effectors of pathogens show high rates of diversifying selection. This has been shown, for example, for the two closely related powdery mildew formae speciales of wheat and barley, where candidate effector genes exhibit a higher rate of diversifying selection than most other genes (Wicker et al., 2013). The rapid evolution of effectors allows the pathogen to escape recognition by the cognate NLR. This evolutionary process can have dramatic consequences for cereal production and usually results in rapid breakdown of NLR-based disease resistance in the field. Ironically, the most popular NLR genes are most prone to resistance breakdown because they are used in many cereal cultivars grown over large areas, factors that, in turn, favor rapid adaptation and spread of virulent pathogen races. A fateful example was the emergence of new and highly virulent stem rust races in wheat, once the popular stem rust resistance gene Sr31 was overcome. The Sr31-breaking race was first observed in Uganda in 1999 and was
therefore termed Ug99 (Pretorius et al., 2000). Ug99 spread over East Africa into the Middle East within only a few years. In this Ug99 lineage, rapid diversification and adaptation to additional genes was later observed (Singh et al., 2015). In response to the rapid evolution of pathogen effectors, many NLR genes are also under strong diversifying selection, which results in the typical arms race between host and pathogen.

Generally, one would expect that the pathogen would have an evolutionary advantage over the host plant because microbial organisms have relatively short generation times and produce enormous numbers of spores (Fetch & McCallum, 2014). Assuming an equal mutation rate in both host and pathogen, we would expect that the pathogen would evolve much more rapidly than the host plant. So, how does the host manage to not lose ground in the constant tug of war with the pathogen, and why have all the resistance genes not been overcome long since? An interesting and exciting solution to this evolutionary conundrum, resulting in NLR-IDs.

Fig. 1 Different models of indirect nucleotide-binding, leucine-rich repeat receptor (NLR) effector perception. (a) Microbial effectors (red) target and modify components of the host’s basal defense response (blue). In the absence of an effector-recognizing NLR, this results in suppressed basal immunity. (b) In the guard model, the status of the basal defense component (guard) is monitored by an NLR protein (gray). Changes to the guard introduced by virulence effectors are perceived by the NLR protein which triggers a hypersensitive response (HR). (c) The decoy (purple) has often arisen through duplication of a basal defense gene. The decoy, however, has lost its original function and its sole purpose is to trap effectors. In the classical decoy model, the decoy and the corresponding NLR protein are encoded by two separate genes. (d) Recent reports have shown that decoy domains can be directly integrated into the corresponding NLR, resulting in NLR-IDs.

Two recent whole-genome comparisons in different plant species have revealed that ‘integrated domain’ NLRs (NLR-IDs) are frequently found in many plant species. Between 3.5% and 10% of all NLRs contain additional integrated domains (Kroj et al., 2016; Sarris et al., 2016). In addition to the HMA domain, Sarris et al. (2016) identified 264 additional integrated domains in a total of 750 NLR-IDs. Seventeen integrated domains were found in at least two of the five cereal species included in their analysis (Table 1). Some of these integrated domains have a known function in basal immunity, such as WRKY or protein kinase domains. Other domains found in these NLR-IDs have so far not been linked to plant immunity. The current assumption is that proteins similar to the integrated domain play a role in basal defense and, consequently, that a knowledge of integrated domains with unknown function could lead to the identification of novel effector targets. The systematic analysis of NLR-IDs could therefore lead to the possibility of the identification of further effector targets and to a better understanding of host–pathogen interactions. Kroj et al. (2016) experimentally tested this hypothesis by using the BED zinc finger domain which is frequently found in NLR-IDs in different plant species. No direct involvement of BED domain-containing proteins in disease resistance has been demonstrated so far. The protein encoded by the rice ZBED gene contains three BED domains which show homology to the respective integrated domains in rice NLR-IDs. Interestingly, rice lines overexpressing containing protein Pi21 of rice resulted in durable and broad-spectrum rice blast resistance, and it is tempting to hypothesize that wild-type Pi21 plays a role in basal defense (Fukuoka et al., 2009). By fusing an HMA domain to an NLR, rice plants might have evolved a neat trick to trap HMA-targeting effectors.
Table 1  Integrated domains found in at least two of the five studied cereal species by Sarris et al. (2016): barley (Hordeum vulgare), rice (Oryza sativa), sorghum (Sorghum bicolor), wheat (Triticum aestivum) and maize (Zea mays)

<table>
<thead>
<tr>
<th>Domain description</th>
<th>Cereal species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein kinase domain</td>
<td>Barley (6), rice (1), sorghum (3), wheat (13), maize (1)</td>
</tr>
<tr>
<td>Protein tyrosine kinase</td>
<td>Barley (1), rice (1), sorghum (1), wheat (1)</td>
</tr>
<tr>
<td>Jacalin-like lectin domain</td>
<td>Barley (2), rice (3), sorghum (1), wheat (2)</td>
</tr>
<tr>
<td>Thioredoxin</td>
<td>Barley (1), rice (1), sorghum (2), wheat (3)</td>
</tr>
<tr>
<td>WRKY DNA-binding domain</td>
<td>Barley (1), sorghum (4), wheat (2)</td>
</tr>
<tr>
<td>BED zinc finger</td>
<td>Barley (1), rice (2), wheat (1)</td>
</tr>
<tr>
<td>WD domain, G-beta repeat</td>
<td>Rice (1), sorghum (3)</td>
</tr>
<tr>
<td>B3 DNA-binding domain</td>
<td>Barley (1), rice (1)</td>
</tr>
<tr>
<td>VQ motif</td>
<td>Rice (1), sorghum (1), wheat (1)</td>
</tr>
<tr>
<td>FNIP repeat</td>
<td>Barley (1), sorghum (2)</td>
</tr>
<tr>
<td>Kelch motif</td>
<td>Barley (2), wheat (4)</td>
</tr>
<tr>
<td>Protein phosphatase 2C</td>
<td>Barley (1), wheat (2)</td>
</tr>
<tr>
<td>Cleavage site for pathogenic effector</td>
<td>Barley (1), rice (1)</td>
</tr>
<tr>
<td>Glutaredoxin</td>
<td>Barley (3), sorghum (1)</td>
</tr>
<tr>
<td>Phloem protein 2</td>
<td>Sorghum (1), maize (1)</td>
</tr>
<tr>
<td>DDE superfamily endonuclease</td>
<td>Barley (2), wheat (1)</td>
</tr>
<tr>
<td>Exo70 exocyst complex subunit</td>
<td>Barley (1), wheat (1)</td>
</tr>
</tbody>
</table>

The numbers in parentheses indicate the number of nucleotide-binding, leucine-rich repeat receptor proteins with the respective integrated domain (NLR-IDs).

ZBED were more resistant to rice blast disease, whereas ZBED mutants were more susceptible. Hence, Kroj et al. (2016) successfully used information on integrated domains to establish a role of ZBED proteins in disease resistance.

V. NLR evolution in Triticeae

In this section, we focus on the evolution of NLR specificities in the grass tribe Triticeae. The availability of closely related species with different ploidy levels, as well as wild and domesticated forms, represents a powerful system to address the evolution of NLR specificities.

1. The MLA-like family of NLR proteins

An interesting example that illustrates the diversification of NLR genes is the mildew A (Mla)-like family found in Triticeae. Originally, Mla was described as an allelic series of resistance genes in barley and > 30 functional alleles with different specificities against barley powdery mildew strains (Blumeria graminis f.sp. hordei) were identified molecularly (Wei et al., 1999; Seeholzer et al., 2010). Subsequently, a functional ortholog of the barley Mla gene was found in the diploid wheat Triticum monococcum. TmMLA1 showed 78% amino acid identity with the barley HvMLA1 protein and conferred race-specific resistance against wheat powdery mildew (B. graminis f.sp. tritici) in a transient assay (Jordan et al., 2011). TmMLA1, however, showed no effect against barley powdery mildew in barley. Even more remarkably, two additional Mla orthologs, Sr33 and Sr50, in wheat provide resistance against fungal stem rust disease (Puccinia graminis f.sp. tritici) (Periyannan et al., 2013; Mago et al., 2015). Sr50 was introgressed from rye into cultivated wheat and Sr33 stems from the diploid wild wheat progenitor Aegilops tauschii. Mildew and rust fungi are only distantly related and they belong to different phyla. Although wheat, barley and rye shared their last common ancestor c. 9 million yr ago (Middleton et al., 2014), powdery mildew and rusts split c. 400 million yr ago (Taylor & Berbee, 2006). The protein sequences of Sr33, TmMLA1 and Sr50 blast to the same MLA family member in the barley genome, indicating that the different pathogen specificities of Sr33, TmMLA1 and Sr50 arose after wheat and barley shared their last common ancestor. These findings demonstrate that different members of the same NLR family can rapidly diversify and adapt to perceive effectors of distantly related fungal pathogens (Fig. 2). Whether the perception of virulence effectors by the MLA family is based on a direct or indirect recognition is still unknown, and it is possible that the rapid adaptation to different pathogens is the result of indirect recognition events (Fig. 1). Furthermore, even effectors with little sequence homology can have very similar protein structures and hence might be recognized by the same or related NLR immune receptors (Ellis, 2016).

2. The wheat leaf rust resistance genes Lr1 and Lr10 originated in wheat progenitors

The long agricultural history of wheat has produced an enormous genetic diversity within this species. Much of this diversity is conserved in genebanks which contain more than half a million accessions (Mitrofanova, 2012). This material represents a unique possibility to study the evolution and diversity of specific genes. In particular, there are wild and domesticated species and genotypes, landraces and elite cultivars, all representing relatively recent evolutionary events. In addition, very similar genomes are present in species of different ploidy levels. Tapping the diversity stored in genebanks allows comparative approaches and the study of evolutionary events at the molecular level. Bread wheat (Triticum aestivum) is a hexaploid species which resulted from two hybridization events (Marcussen et al., 2014). First, two wild species, Triticum urartu and a now extinct relative of Aegilops speltoides, hybridized and formed wild tetraploid wheat (T. turgidum ssp. dicoccoides). After domestication of this wild form to cultivated tetraploid wheat (T. turgidum ssp. dicoccoides), a second hybridization occurred with the wild grass Ae. tauschii resulting in the hexaploid bread wheat (Fig. 3).

Many resistance genes have been introgressed from wild species into cultivated wheat (Baum et al., 1992). For example, the Lr21 leaf rust resistance gene was introgressed from Ae. tauschii using a synthetic wheat. Extensive sequence analyses revealed that the functional Lr21 allele is a chimera of two non-functional Lr21 haplotypes (H1 and H2) (Huang et al., 2003, 2009). In bread wheat, only inactive lrl alleles are present, but an active resistance gene could be experimentally reconstituted by intramolecular recombination of the inactive H1 and H2 haplotypes. This provides an indication that novel, active
resistance genes might have originally evolved in wild species by recombination of existing, non-functional haplotypes. Such events are in agreement with the proposed model for the recycling of resistance gene sequences resulting in the evolution of molecular diversity (Holub, 2001).

In contrast with Lr21 and other wheat NLR genes, the origin of a large number of resistance genes first described in hexaploid wheat is not known. Their origin should be determined by comparative analyses and allele mining of diverse germplasm. The wheat leaf rust resistance gene Lr1 was first described by Ausemus.

Fig. 2 Closely related members of the mildew A (MLA)-like immune receptor family perceive effector proteins of distantly related fungal pathogens. Left: phylogenetic tree of MLA-like family members from diploid wheat (Tm), hexaploid wheat (Ta) and barley (Hv). Mildew-perceiving members are indicated in gray and nucleotide-binding, leucine-rich repeat receptor (NLR) proteins recognizing stem rust are marked in brown. The leucine-rich repeat (LRR) domain was used to construct the tree and the wheat Pm3A protein was used as outgroup to root the tree. Bootstrap numbers at the forks indicate how many times the sequences to the right of the fork occurred in the same group out of 100 trees. For barley, only three of the > 30 known functional MLA proteins were included in the tree. Right: schematic representation of the divergent evolution of different MLA-like family members providing resistance to different fungal pathogens. Divergence times in million yr ago (Ma) are based on the reports of Taylor & Berbee (2006), Wicker et al. (2013) and Middleton et al. (2014).

Fig. 3 Evolutionary history of different wheat disease resistance genes. The times of the different hybridization events are based on Middleton et al. (2014). Ma, million years ago.
et al. (1946) in the hexaploid wheat cultivar Malakoff (Dyck & Samborski, 1968). *Lr1* is located on chromosome 5D of wheat and Ling et al. (2004) screened >200 accessions of *Ae. tauschii* to search for an *Lr1*-type gene in the D-genome donor of bread wheat. Using well-characterized leaf rust isolates, five different *Ae. tauschii* accessions were identified which carried a resistance gene that was functionally identical to *Lr1* in bread wheat and mapped to the same genetic region as *Lr1* on chromosome 5D, indicating an orthologous resistance gene. Physical mapping and comparison with the later isolated *Lr1* gene (Cloutier et al., 2007) revealed that a specific polymorphic segment of 605 bp, encoding LRRs 9–15, distinguished both the hexaploid wheat *Lr1* as well as the *Ae. tauschii* ortholog from the susceptible allele of *Lr1* (Qiu et al., 2007). There are only very few polymorphisms between the *Lr1* gene in hexaploid wheat and the orthologous *Lr1* genes in the five *Ae. tauschii* accessions with the *Lr1* resistance phenotype, and it is likely that the *Lr1* gene in hexaploid wheat evolved in the diploid *Ae. tauschii* and was later introducted to bread wheat by gene flow based on hybridization.

Similar studies have been performed on the origin of the *Lr10* leaf rust resistance gene in wheat. It is located on chromosome arm 1AS and was cloned from bread cv Thatcher *Lr10* (Feuillet et al., 2003). A diversity study on 20 diploid (*T. urartu*; A-genome donor) and tetraploid (wild and domesticated) wheat lines revealed that all carried a sequence homologous to *Lr10* (Lourete et al., 2009). It has been shown previously that the hexaploid wheat genepool has two haplotypes at the *Lr10* locus. One originated from a complex deletion event in the original functional haplotype, followed by a large inversion, and therefore is evolutionary younger. This deletion/inversion haplotype in hexaploid wheat was found in an identical form in the A genomes of diploid and tetraploid wheat, and is therefore also an ancient haplotype (Isidore et al., 2005). Thus, identical presence/absence haplotypes were present at the *Lr10* locus at all three ploidy levels. Interestingly, the *Lr10* coding sequence in the tetraploid wheat cultivar Altar was identical to the bread wheat *Lr10* gene, with only two sequence polymorphisms in the intronic region. Using virus-induced gene silencing, it was found that, in addition to cultivar Altar, there were functional *Lr10* genes in the two tetraploid cultivars Russello and Bufalo that are of Italian origin (Lourete et al., 2009). These studies demonstrated ancient diversity of the *Lr10* locus and the presence of an active *Lr10* gene in tetraploid wheat. The evolutionary stability of a presence/absence polymorphism of the *Lr10* gene suggests a balanced polymorphism and maintenance of both haplotypes in the genepool, as also suggested for other resistance genes, such as *Rpm1* in Arabidopsis (Tian et al., 2003). In conclusion, the *Lr10* gene was most probably already present in a tetraploid wheat progenitor (Fig. 3).

The studies on *Lr1* and *Lr10* both revealed orthologous functional genes in diploid and tetraploid bread wheat progenitor species. Thus, these two genes are most probably ancient and have been introduced into the bread wheat genepool, either in the original hybridization resulting in hexaploid wheat or by gene flow (Fig. 3). Clearly, this is one evolutionary route of bread wheat resistance gene evolution, but, as the next case study shows, is not the only one.

3. The evolutionary history of functional *Pm3* alleles in relation to wheat evolution and domestication

The *Pm3* powdery mildew resistance in hexaploid wheat germplasm was described originally as a multi-allelic series of 10 resistance alleles, *Pm3a–j* (McIntosh et al., 2008), each having a characteristic race specificity. The gene is located on chromosome arm 1AS in a similar region of the chromosome arm as the *Lr10* gene described above. The *Pm3* alleles were identified in cultivars originating from all over the world, with no clear center of origin. After map-based cloning of the first allele *Pm3b* (Yahiaoui et al., 2004), the remaining alleles were isolated by PCR amplification and were found to be true alleles of the same gene in a cluster of *Pm3*-like genes (Srichumpa et al., 2005; Yahiaoui et al., 2006). Three of the alleles (*Pm3b*, *Pm3i* and *Pm3j*) turned out to be identical to *Pm3d*, *Pm3c* and *Pm3b*, respectively.

The seven alleles *Pm3a–g* all encode classical NLR immune receptors and the encoded proteins are highly similar to each other. Polymorphisms were either restricted to a few single amino acid polymorphisms, mostly in the LRR region, or caused by short polymorphic blocks encoded by gene sequences probably derived from gene conversion events (Yahiaoui et al., 2006). Large-scale allele mining in diverse wheat germplasm consisting of >2000 accessions, identified by both a focused identification of germplasm strategy as well as geographical considerations, revealed a total of nine additional functional *Pm3* alleles (*Pm3f–h*) (Bhullar et al., 2009, 2010b). In addition, these alleles are highly similar in amino acid sequence, indicating a very recent common ancestor in hexaploid wheat. Indeed, a rough estimate of the allelic divergence time suggested that they could have diverged as recently as a few thousand years ago, that is, after formation of hexaploid wheat. Based on the relatively low sequence diversity of these 16 alleles, a *Pm3* consensus sequence could be derived (Yahiaoui et al., 2006; Bhullar et al., 2010b). Interestingly, a gene that is identical to this consensus sequence exists in the bread wheat genepool, for example, in the Chinese landrace ‘Chinese Spring’, which is powdery mildew susceptible. This *Pm3* allele was called *Pm3CS* and is non-functional, that is, it does not confer resistance to any of the tested powdery mildew isolates.

Allele mining in 375 accessions representing wild and domesticated tetraploid wheat species, including 126 accessions from a highly diverse collection assembled by Ozkan et al. (2003, 2005), revealed considerable diversity of *Pm3* alleles in this genepool (Yahiaoui et al., 2006, 2009). With the exception of one haplotype, haplotypes in tetraploid wheat were all different from haplotypes in hexaploid wheat, indicating an independent evolution of *Pm3* in hexaploid wheat after domestication. Only one new, functionally active resistance allele, *Pm3k*, was identified in tetraploid genotypes. The *Pm3k* sequence is not present in the bread wheat genepool analyzed so far (Bhullar et al., 2009; Yahiaoui et al., 2009). Thus, none of the 16 functional *Pm3* alleles in bread wheat corresponds to the only active *Pm3* allele in tetraploid wheat, confirming independent evolution in the two species. As described above, there is a natural, susceptible allele in the bread wheat genotype Chinese Spring, *Pm3CS*, which also corresponds to the consensus sequence of all functional *Pm3* alleles. Interestingly, the
Pm3CS allele was also found in a total of six tetraploid wheat lines and is actually the only common Pm3 allele between the tetraploid and hexaploid wheat gene pools that diverged only 10,000 yr ago (Yahiaoui et al., 2006). Moreover, the four wild tetraploid wheat genotypes which contain Pm3CS originate from south-eastern Turkey, from sites which are geographically close or identical to the genotypes which contain Pm3CS (Yahiaoui and hexaploid wheat gene pools that diverged only 10,000 yr ago and is actually the only common allele was also found in a total of six tetraploid wheat lines. Possibly, at this time, this was a functional allele providing resistance to powdery mildew and was only later overcome once large plant populations containing this allele were grown in early agriculture after bread wheat domestication. A wild tetraploid wheat line carrying this allele contributed the Pm3 genomic region to some of the first domesticated emmer genotypes. Furthermore, one of these early domesticated emmer wheats with Pm3CS was involved in the hybridization(s) with Ae. tauschii, resulting in a hexaploid wheat line carrying Pm3CS. Based on this template allele, mutation and gene conversion events then resulted in a larger set of new haplotypes, some of which were active against powdery mildew and therefore selected by farmers and breeders.

In summary, the functional allelic diversity of Pm3 in bread wheat originated only after bread wheat domestication, and therefore the evolution of active Pm3 genes is distinct from that of the ancient resistance genes Lr1 and Lr10 (Fig. 3).

4. Modeling of Pm3 NLR protein structure reveals distinct evolutionary histories in tetraploid and hexaploid wheat

Three-dimensional modeling of the Pm3 protein was used to determine the possible location of polymorphic amino acid residues compared with the susceptible consensus sequence (Sela et al., 2014). This analysis of possible structural consequences of diversity suggested distinct evolutionary histories of Pm3 proteins in tetraploid and hexaploid wheat. The structural model proposed a very long LRR 24 repeat which separated the two regions of highest variability in the two wheat species. Most interspecies polymorphic amino acids occur in LRRs 19–24 in the core motif LxxLxLxxN/C and also in the loop regions of LRRs 20, 22 and 22a which connect the core motifs (Fig. 4). By contrast, the within-species polymorphism hotspot in Pm3 variants of hexaploid wheat is mostly confined to the core motif in the LRRs 25–28, specifically at solvent exposed residues. This pattern could reflect differential selection pressure on Pm3 in the two wheat species. With the exception of Pm3k, the Pm3 alleles in wild tetraploid wheat do not give resistance to current mildew races. However, it is possible that they were functional before domestication and the location of the polymorphisms indicates selection pressure during that time. After domestication, all the tetraploid Pm3 genes were overcome because of the rapidly increasing acreage of wheat and the large pathogen populations. The new Pm3 resistance alleles would then have evolved as a consequence of mutations in different LRR regions. This model implies that both LRRs 19–24, as well as 25–29, are involved in specific binding of the avirulence gene product or a putative guardée/decoy.

In the last 2 yr, there has been significant progress in the characterization of the genetic basis of mildew interactions on the pathogen side. Although all the work described above on the identification and characterization of Pm3 alleles in tetraploid and hexaploid wheat was performed with mildew from bread wheat, there is some evidence for mildew specialization on the two different wheat species (Ben-David et al., 2016; Menardo et al. 2016). The first Avr gene recognized by a Pm3 allele has been cloned recently (AvrPm3a/f, recognized by Pm3a and Pm3f) (Bourras et al., 2015) and has been found to encode a typical secreted protein. The molecular isolation of this Avr gene provides a
start to test the interactions suggested by structural modeling. The protein modeling data indicate that molecular diversity studies can generate hypotheses for molecular functional studies, demonstrating the potential of such studies which goes beyond a simple catalog of diversity.

5. Conservation of orthologous gene function after long divergence times in evolution

It is still an open question how frequently orthologous genes in different cereal species evolve in parallel over a long time to confer resistance against similar pathogens. The Mla locus in barley and wheat represents such a case in which the same orthologous genes in different species confer resistance (Jordan et al., 2011; Periyannan et al., 2013; Mago et al., 2015). Another well-studied example relates to the Pm3 locus. It was found that the rye gene Pm8 is an ortholog of wheat Pm3 (Hurni et al., 2013). Pm8 is active in wheat, but the global use of this gene, which forms part of the yield-increasing 1BL/1RS translocation, has resulted in the occurrence of many virulent races. It is surprising that orthologous genes have a conserved function after many millions of years of divergence, as resistance genes and resistance loci in general are very dynamic and can evolve rapidly.

The functional conservation of orthologous resistance genes also has implications on our understanding of interactions with the individual mildew forms that are specialized on the different hosts. The powdery mildew form specialized on rye cannot infect wheat. Nevertheless, the conservation of activity of Pm8 towards wheat powdery mildew suggests that the molecular activity recognized by Pm8 is functionally conserved in the two specialized mildews. Therefore, a comparative molecular study of the Pm3/Pm8 interaction with mildew might reveal evolutionary conserved elements that are essential for the pathogen and that could ultimately be used to develop more durable resistance. Such pathogen-informed strategies could ideally complement a whole set of approaches focusing on resistance management on the host side, which have been described in two recent excellent reviews (Burdon et al., 2014; Zhan et al., 2015).

6. Evolutionary differentiation of NLR genes can result in unexpected side effects: hybrid necrosis and resistance suppression

Plant immune receptor-type proteins have been implicated in hybrid necrosis, which occurs when two incompatible genotypes are crossed, resulting in evolutionary steps towards plant speciation (Bomblies & Weigel, 2007; Bomblies et al., 2007; Ispolatov & Doebeli, 2009; Chae et al., 2014). Hybrid necrosis was also observed in crosses of some wheat genotypes. The two wheat genes Ne1 and Ne2 occur in different allelic forms in the genepool and, when combined in the same genotype, result in hybrid necrosis. Therefore, genepools with either the Ne1 or Ne2 gene are on an evolutionary trajectory to evolve into different species. Based on genetic crosses and mutational studies, Zhang et al. (2016) have concluded recently that Lr13 and the Ne2m allele may be the same gene and, based on the work in Arabidopsis, it is tempting to speculate that Lr13 encodes an NLR-type protein. Given the rapid progress in wheat genome sequencing, the cloning of Lr13 seems to be within reach in the next few years.

In addition to negative interactions, the model of Ispolatov & Doebeli (2009) also predicts a class of hybrids with a weakened resistance response. Interestingly, it has been observed previously that the introgression of resistance genes from lower to higher ploidy levels frequently results in loss of resistance (Hsam & Zeller, 2002; McIntosh et al., 2011). It has been speculated that this could be a result of suppression interactions by non-functional gene homoeologs of the introgressed active resistance genes. This would result in a dominant negative genetic interaction. These hypotheses have been experimentally tested for the rye Pm8 resistance gene, which was found to be suppressed in certain wheat genotypes, but not in others. It was found that suppression depends on Pm3 alleles (Hurni et al., 2014). Suppression did not occur at the transcriptional or translational level, but seemed to be caused by protein interactions resulting in protein dimers or multimers that were incompetent for the induction of the immune response. Interestingly, this suppression was also observed among some, but not all, of the Pm3 alleles when they were present in hybrid plants, but also in transgenic plants homozygous for two different Pm3 alleles (Stirnweis et al., 2014). Thus, the products of cereal NLR genes, being functional genes or non-functional homologs, can contribute to a variety of molecular interactions which result in plant phenotypes of evolutionary and agricultural importance. The molecular and structural basis of such interactions remains unclear and represents one of the very attractive research topics in the field of NLR genes.

VI. Quantitative disease resistance – why plasma membrane-localized receptors and NLRs are not sufficient to explain disease resistance in cereals

Quantitative disease resistance (QR) is defined as host plant resistance that is partial and conferred by the joint effect of several genes (St Clair, 2010; Niks et al., 2015). Although there is no causal link between the completeness of resistance gene action and durability, it has been found that partial QR genes are often more durable than NLR-based immunity (Lagudah, 2011; Ellis et al., 2014). It is not surprising that very little is known about partial QR in model plants and that the bulk of our knowledge on QR stems from cereals. The partial effects of QR genes can be difficult to score and can depend on environmental conditions, as well as the developmental stage of the plant. Breeders and pathologists, however, have exploited QR during decades of cereal breeding. Access to breeding records and disease monitoring programs provides us with exact and rich information on cereal cultivars that were grown over long periods and on large areas without resistance breakdown. It is important to note that, despite the joint action of several partial resistance genes in one cultivar, single QR genes can be mendelized and their effects scored as single genes in near-isogenic backgrounds. One example is the recessive rice blast resistance gene pi21 that has been discussed previously. Pi21 was originally identified in a quantitative trait locus (QTL) study as one of several QR genes in the Japanese upland rice cultivar
Owarihatamochi (Fukuoka & Okuno, 2001). Through repeated backcrosses, pi21 was transferred into a near-isogenic background which allowed the assessment of the phenotypic effect of pi21 as a single gene. Near-isogenic lines with the loss-of-function pi21 allele showed smaller rice blast lesions than lines with the functional, susceptible Pi21 allele (Fukuoka et al., 2009). QTL analyses are statistical approaches and it is not possible to draw any conclusions from QTL peaks with regard to the nature of the protein responsible for the resistance phenotype. For example, it is feasible that membrane-localized RLKs and RLPs can be identified in QTL studies, and there are even examples of race-specific NLR genes that were identified in QTL studies (Paillard et al., 2012; Zhang et al., 2015). However, recent research has led to the identification of QR genes that code for unusual resistance proteins. In the next two sections, we discuss three examples of such unusual resistance proteins.

1. Spontaneous mutation events in transporter genes led to the emergence of durable, multi-pathogen resistance after wheat domestication

Hexaploid bread wheat contains a small, but important, group of genes that confer resistance against multiple biotrophic fungal pathogens (Ellis et al., 2014). Two of these genes, Lr34 and Lr67, have been cloned recently and encode for membrane-localized transporter proteins: Lr34 for a full-size ATP-binding cassette (ABC) transporter (Krattinger et al., 2009) and Lr67 for a hexose transporter (Moore et al., 2015). Both genes confer race non-specific and partial resistance against the three wheat rusts and powdery mildew. Lr34 can be considered as one of the most durable sources of partial disease resistance known in cereals. A very interesting observation was that the resistant versions of both Lr34 and Lr67 only differed by two critical amino acid changes from their susceptible protein versions. The susceptible version of Lr67 showed a high affinity to glucose in yeast uptake experiments, whereas glucose uptake was abolished in the resistant hexose transporter version. The substrate of the Lr34 ABC transporter is still unknown.

Very interestingly, the resistant alleles of both genes were only found in cultivated hexaploid bread wheat, but not in wild wheat progenitors (Krattinger et al., 2013; Moore et al., 2015). This indicates that the resistant versions of both transporters evolved only after wheat domestication, most probably as a result of spontaneous sequence changes (Fig. 3). Selection by ancient farmers might therefore have played an important role in preserving these rare mutation events for modern breeding. It is possible that these rare alleles conferred a selective disadvantage in natural populations and that they would therefore rapidly disappear again. Only under human cultivation might such gene variants be beneficial. Based on the observations on Lr34 and Lr67, we hypothesize that similar specific sequence changes in other transporter proteins might have the same effect. This suggests the exciting possibility to artificially ‘create’ durable multi-pathogen alleles through targeted mutagenesis. It is therefore essential to obtain a better understanding of how exactly these mutations affect the substrate specificities in the two transporters. Interestingly, Lr34-like or Lr67-like disease resistances have not been described in diploid cereals, such as barley, rice or maize. Transformation experiments, however, have shown that Lr67 and Lr34 can be functionally transferred into barley (Lr67/Lr34) and rice (Lr34), where they confer partial resistance against diseases specific to barley (barley leaf rust, barley powdery mildew) and rice (rice blast), respectively (Risk et al., 2013; Moore et al., 2015; Krattinger et al., 2016). These results indicate that all components required for this type of disease resistance are present in diploid cereals as well.

2. Yr36 – a wheat broad-spectrum resistance gene that was not incorporated into modern cultivars

Another interesting example of a partial, broad-spectrum resistance gene is Yr36, which was originally identified in wild tetraploid emmer wheat (Triticum turgidum ssp. dicoccoides) (Uauy et al., 2005). Yr36 confers race non-specific resistance against fungal stripe rust disease and encodes a WHEAT KINASE START1 (WKS1) protein with an N-terminal non-ND kinase domain fused to a START lipid-binding domain at the C-terminus (Fu et al., 2009). Interestingly, there are two major splice variants of this gene, resulting in a full-length WKS1.1 protein and WKS1.2 with a complete kinase but a truncated START domain. Only the WKS1.1, but not WKS1.2, variant conferred resistance to stripe rust (Gou et al., 2015). After stripe rust infection, the WKS1.1 splice variant increased in abundance compared with the alternative variant. WKS1.1 interacted with the thylakoid-associated ascorbate peroxidase (tAPX) in the chloroplast. Ascorbate peroxidases are involved in the detoxification of peroxides. Gou et al. (2015) concluded that the interaction of WKS1.1 and tAPX might reduce the cell’s ability to detoxify reactive oxygen species, which could trigger an HR. As WKS1.1-mediated signaling is much slower than a typical NLR-triggered HR, Yr36-containing wheat plants only show a partial resistance phenotype. Interestingly, Yr36 was not introgressed into the cultivated durum and bread wheat genepool, but was only found in wild emmer. Huang et al. (2016) studied the distribution of Yr36 in 435 wild emmer accessions and only found the gene in accessions from the southern regions of the Fertile Crescent, whereas accessions from the northern parts were absent for Yr36. Wheat domestication is thought to have occurred in the northern parts of the Fertile Crescent, which would explain the absence of Yr36 in domesticated wheats. Yr36 therefore represents an example of an important resistance gene that escaped incorporation into the modern genepool through domestication and breeding. This highlights the importance of mining wild progenitors of cultivated cereals and landraces for new sources of durable and broad-spectrum disease resistance ( Tanksley & McCouch, 1997).

VII. Conclusions: evolutionary studies of cereal resistance genes contribute to both basic research and breeding

Studies on the origin of resistance genes can identify geographic regions and/or gene pools that are promising for the identification of new functional resistance genes. For example, the diversity
studies on Pm3 revealed that both Turkey as well as the Himalayan mountain range are regions with high diversity of this gene (Bhullar et al., 2009, 2010a,b). Furthermore, for genes that evolved after domestication, the bread wheat genepool is most promising for the identification of additional alleles, with wild or domesticated relatives being less promising. Nevertheless, it is also possible to find functional diversity there, as shown by the identification of the rye Pm8 as an orthologue of Pm3, or the functional orthologs of Mla in wheat and rye. Wild relatives are attractive to explore for ancient genes, because a higher level of allelic diversity can be expected in wild grass genepools. Thus, it is highly important to isolate as many resistance genes as possible in the near future to be able to perform allele mining in diverse genepools. Given the recent rapid progress in wheat genome sequencing and the development of new technology based on rapid mutant identification (Steuer Nagel et al., 2016), it can be expected that many more resistance genes will be cloned in the next few years, resulting in large opportunities for gene studies and combination breeding and gene stacking. It has been shown that a transgenic use of overexpressed Pm3 alleles in the field can increase resistance (Brunner et al., 2011, 2012), but the real test for agricultural use would be the growth of such transgenic lines on a large scale to determine a possible pathogen adaptation.

The observation of very recent events resulting in novel, active resistance genes is also encouraging for emerging approaches to develop resistance genes with new specificities by gene modification. In cases such as Pm3, Lr67 or Lr34, the short time period of domestication and the low natural mutation frequencies might not have been sufficient to evolve all possible useful variants of these genes based on the available ‘template’ sequences that made it through the bottleneck of domestication. Thus, it can reasonably be assumed that, based on a better molecular understanding, novel genes can be designed. Indeed, the use of sequence information obtained from natural diversity studies has allowed Stirnweis et al. (2014) to make improved versions of known Pm3 genes in the laboratory by modifying only two amino acids. Finally, there are some immediate applications of several of the recent findings in practical breeding. For example, mutation screens for the removal of suppressor genes which have possibly accumulated during evolution could awaken ‘sleeping’ resistance genes which might be highly useful. Systematic work in this direction is very promising and has not yet been undertaken at a larger scale.

Acknowledgements

We would like to thank Dr Thomas Wicker and Dr Teresa Koller for support with the preparation of the figures. This work was supported by an Ambizione grant from the Swiss National Science Foundation to S.G.K. and a grant from the same funding source to B.K. (#310030_163260).

References


