

# Moving targets: rapid evolution of oomycete effectors

Darren M. Soanes and Nicholas J. Talbot

School of Biosciences, Geoffrey Pope Building, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK

**Plant pathogenic microbes secrete proteins known as effectors, which enter the cytoplasm of plant cells and suppress host defences. Known effectors in oomycete pathogens possess an RXLR–EER motif in their amino acid sequence that is necessary for transport of the effector into a host plant cell. A large number of putative effectors have now been identified in oomycete genomes, the sequences of which show evidence of diversifying selection at their C terminus. Here, we describe recent progress in characterizing RXLR–EER effectors and discuss why so many of these rapidly evolving proteins are encoded by the genomes of plant pathogenic oomycetes.**

## Plant pathogens produce effectors to suppress plant defences

Microorganisms that cause diseases of plants have a devastating effect on many important crops [1]. Therefore, it is essential to understand how they infect plant tissues and proliferate within them. Plants have a range of defence mechanisms that protect them from microbial attack. These include the production of antimicrobial compounds and proteins (such as chitinases and glucanases that damage the cell wall of the invading pathogen), the release of reactive oxygen species and the strengthening of the plant cell walls by deposition of callose. Plant defence responses are activated by the recognition of conserved microbial proteins (known as pathogen-associated molecular patterns, or PAMPs) that are presented at the cell surface or secreted by invading microbes [2]. For a pathogen to invade a host plant successfully, it must suppress these plant defences. Bacterial pathogens achieve this by directly injecting proteins into the plant cytoplasm using the type-III secretion system [3]. These proteins, known as effectors, suppress plant defences but in some cases, might elicit a phenomenon known as the hypersensitive response (HR). During HR, plant cells at the site of microbial invasion undergo programmed cell death, which prevents further proliferation of the phytopathogen within the plant because during the early stages of infection, the infecting microbe requires living plant tissue [4]. Other bacterial effectors also suppress HR. Proteins that elicit HR are known as avirulence (Avr) factors and are recognized (normally indirectly) by corresponding resistance (R) genes in the host plant – the so called ‘gene-for-gene’ hypothesis [5]. Several Avr genes have been discovered in fungal and oomycete pathogens that encode products recognized by R proteins that are situated in the plant cytoplasm, indicating that there is a

mechanism of transporting fungal and oomycete effectors into plant cells [6]. Little is known about this mechanism in fungi, but recent research on oomycete phytopathogens has identified a large number of potential effectors, identified motifs in their amino acid sequence that are necessary for their entry into plant cells and discovered information relating to the evolution of effectors in these species [7–9].

## The RXLR–EER motif is essential for translocation of effectors into plant cells

Although they share morphological features with some fungal plant pathogens, oomycetes are heterokonts (stramenopiles) that belong to the kingdom Chromalveolata and, therefore, are more closely related to diatoms and brown algae. Phytopathogenic oomycetes are responsible for economically important diseases, such as late blight of potato and sudden oak death caused by *Phytophthora infestans* and *Phytophthora ramorum*, respectively. The downy mildew *Hyaloperonospora arabidopsidis* (formerly *Hyaloperonospora parasitica*) is a natural pathogen of the genetically tractable plant *Arabidopsis thaliana* and, therefore, is a useful model to study host–pathogen interactions and the molecular basis of effector function [1].

Four oomycete Avr genes have been identified and cloned: two from *H. arabidopsidis* [10,11], one from the soybean pathogen *Phytophthora sojae* [12] and one from *P. infestans* [13]. These proteins are recognized by R genes located in the plant cytoplasm, consistent with the hypothesis that the proteins encoded by the Avr genes are translocated into the plant cytoplasm. Sequence alignment of the oomycete Avr proteins showed a conserved RXLR motif within 32 amino acids of the signal peptide. This motif resembles, in both sequence and position, the RXLX(E/D/Q) motif required for the translocation of proteins from the malaria parasite *Plasmodium falciparum* into host erythrocytes [14]. Indeed, the oomycete RXLR motif is functional in *P. falciparum* when the motif is positioned in the N-terminal region of the protein following the signal peptide [15]. Three of the four Avr proteins also have an EER motif less than 25 residues downstream of the RXLR motif, and it seems likely that the RXLR motif defines a region of 25–30 amino acids that are required for host-cell targeting. To test whether the RXLR–EER motif is necessary for translocation of AVR proteins into plant cells, Avr3a from *P. infestans* has been used as a reporter for translocation because it triggers HR following recognition within potato cells that contain the R3a resistance protein [7]. Mutant *P. infestans* strains were constructed containing Avr3a genes in which the RXLR and EER motifs had been replaced singly and in combination by alanine residues and also by KMIK–DDK (a change that conserved the physico-chemical properties of the protein). These mutant strains failed

to elicit HR. Because recognition of Avr3a by R3a previously had been shown to be independent of the RXLR–EER motif [16], it was hypothesized that both the RXLR and EER motifs are necessary for translocation of proteins into host cells. Labelling of both wild-type and mutant forms of Avr3a with monomeric fluorescent red protein showed that the RXLR–EER motif was not necessary for secretion of protein from the haustoria, indicating that the signal peptide was sufficient for this purpose. To investigate the role of the RXLR–EER motif in protein translocation, the N-terminal domain of Avr3a containing the signal peptide and RXLR–EER motif was fused to the *Escherichia coli gusA* gene that encodes  $\beta$ -glucuronidase (GUS), an enzyme whose localization can be visualized by the addition of an appropriate substrate (X-gluc or 5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucuronic acid), which is catalysed by GUS to generate a blue dye. When potato was infected with *P. infestans* strains expressing this construct, GUS activity was observed in potato cells adjacent to haustoria. No GUS activity was seen in potato cells infected with *P. infestans* expressing a construct in which GUS was fused to the N-terminal domain of Avr3a in which the RXLR–EER motif had been replaced with alanine residues. These results enabled the authors to propose a hypothesis in which the signal peptide sequence is sufficient for oomycete effectors to be secreted from haustoria but the RXLR–EER motif is necessary for translocation of effectors into plant cells [7]. Strong support for this idea has come from a recent study of the effector protein Avr1b from *P. sojae* [17]. Particle-bombardment studies have shown that the RXLR–EER motif is both necessary and sufficient for host-cell targeting of Avr1b in the absence of the pathogen. This indicates that no additional pathogen-encoded cellular machinery is necessary for host-cell delivery. In fact, the RXLR–EER motif was shown to be sufficient to enable delivery of green fluorescent protein (GFP) into host plant cells [17]. The *P. falciparum* host-cell-targeting motif also functioned in these experiments to bring about delivery into soybean cells, underlining the conservation in host-cell targeting between the oomycetes and the evolutionarily related, but ecologically distinct, apicomplexan malaria pathogen.

### Evidence of positive selection at the C terminus of effector proteins

The availability of genome sequences from *P. sojae*, *P. ramorum* and *H. parasitica* [18] has enabled potential effectors to be identified from these species based on the presence of RXLR. Genome sequences were scanned for protein-encoding genes that have a signal peptide, followed by an RXLR motif within 30–60 amino acids of the N terminus of the protein. Using this methodology, Win and colleagues identified 672 candidate RXLR effectors from *P. sojae*, 531 from *P. ramorum* and 149 from *H. arabidopsidis* [8]. In a separate study, Jiang and co-workers identified 375 candidate RXLR–EER effectors from *P. ramorum* and 396 from *P. sojae* using a mixture of recursive BLAST and Hidden-Markov Model approaches [9]. The latter approach might underestimate the number of effectors because some experimentally verified effectors lack the EER motif. Both studies have shown, however,

that the sequences of candidate effectors are highly divergent. Only a small number of sequences (6%), for instance, had significant similarity ( $e < 10^{-5}$ ) to known proteins from non-oomycete species. The use of BLAST to compare the sets of predicted RXLR effectors from each species of oomycete showed that only 14 of the 149 *H. arabidopsidis* genes showed similarity to effectors from either of the *Phytophthora* species. There was a greater overlap between the *Phytophthora* species, although this was limited to 36% of the *P. ramorum* and 22% of the *P. sojae* candidate effectors [8]. In several *H. arabidopsidis* effectors, the RXLR motif overlapped with a LXLFLAK motif. These effectors showed similarity to members of the Crinkler (Crn) family, which is a distinct and intriguing group of cytoplasmic effectors discovered in *P. infestans*, members of which do not have RXLR motifs in the *Phytophthora* species [19]. This provides more evidence for a large divergence between *H. arabidopsidis* and *Phytophthora* effector sequences. Jiang and colleagues, using recursive BLAST, placed more than 90% of the RXLR–EER effectors from *P. sojae* and *P. ramorum* into one large superfamily, which nevertheless showed high sequence diversity, indicating rapid evolution of this gene family [9]. Despite the diversity of sequences, the authors identified several conserved C-terminal motifs (named W, Y and L) that were found in more than half of the effectors. These motifs were usually arranged as a module and repeated up to eight times. Interestingly, the Avr1b effector of *P. sojae* also contains conserved motifs (K, W and Y) in the C terminus, and a recent study has shown that these are essential for the avirulence interaction between Avr1b and the Rps1b resistance protein in soybean [20]. Furthermore, the Avr1b protein has been shown to suppress programmed cell death triggered by the mouse BAX protein in soybean, yeast and *Nicotiana benthamiana* cells, and this requires the W and Y motifs. It seems likely that the conserved C-terminal motifs, therefore, have an important role in modulation of host-cell signalling and perhaps suppression of HR. When considered together, these studies have demonstrated that a large pool of potential effectors exists within each species of oomycete, with a large number of effectors playing a part in host-defence suppression, because at least half of oomycete effectors contain W and Y motifs [20]. Furthermore, there is considerable sequence diversity both within and between these groups of genes.

The requirement of a large pool of effectors, the sequences of which are evolving rapidly, is consistent with their role within the plant cytoplasm. Host R genes are under strong selective pressure to recognize effector-gene products, triggering HR and, thus, halting the spread of the pathogen. Therefore, there is evolutionary pressure for the sequence of the effector to change so that the protein is no longer recognized by the R gene product. As a consequence, there is also selective pressure for the sequences of R genes to diversify so that new effectors can be recognized. Such co-evolution between Avr and R genes has already been demonstrated in the interaction between *H. arabidopsidis* and *A. thaliana*, with a high degree of polymorphism evident in the amino acid sequence encoded by both the Avr gene *ATR13* and the corresponding R gene *RPP13* [11]. Similarly, evidence for positive selection acting on the

C terminus of *Phytophthora* effectors was first observed during the isolation and characterization of the Avr1b protein of *P. sojae* [12]. Evidence for diversifying (positive) selection can be found by comparing the nucleotide sequences of sets of orthologues from several closely related species or a set of paralogues from the same species. The number of synonymous mutations ( $d_S$  – mutations that do not result in a change in amino acid sequence) and non-synonymous mutations ( $d_N$  – mutations that result in a change in amino acid sequence) are determined. Genes that are under selective pressure to undergo protein sequence change have a  $d_N:d_S$  ratio of  $>1$ . Determining  $d_N:d_S$  ratios over the whole gene can provide some evidence for positive selection, but more sensitivity is obtained by using maximum-likelihood methods to identify positions within a gene (individual structural domains, for instance) that have a  $d_N:d_S$  ratio of  $>1$  [16,21]. Evidence of diversifying selection was found in 70 of 99 paralogue groups of RXLR effectors from three oomycete species, and the overwhelming majority of sites showing positive selection were located in the C terminus of effectors [8–10,17,20]. This is consistent with a modular model in which the N terminus is responsible for secretion of the effector from haustoria and its translocation into plant cells, whereas the C terminus is responsible for modulating plant defences and other biological functions within plant cells.

### Future perspectives

The identification of RXLR motifs in known oomycete Avr genes has enabled researchers to identify a large pool of potential effectors from several oomycete species. Although known effectors could potentially be used as targets to manufacture chemicals that prevent infection of plants, in practice the large number of these genes carried by each pathogen and their probably rapid evolution means that anti-microbial agents produced in this way are unlikely to provide long-term control of a given pathogen. The mechanism of entry of effectors into plant cells is currently unknown, but it now seems unlikely that a pathogen-manufactured apparatus mediates effector delivery in the same way as the type-III secretion system in bacteria [17]. Alternatively, a plant translocation system, perhaps endocytic or pinocytotic, could be co-opted for effector delivery by means of an RXLR–EER interaction. In addition, although the whole RXLR–EER motif is necessary for effector translocation in potato infected by *P. infestans*, other oomycete effectors lack the EER portion of the motif – so is this motif dispensable in some situations? Exploring the nature of the effector protein delivery mechanism in oomycetes is likely to be one of the most important challenges in molecular plant pathology and is one of several outstanding questions that need to be addressed (Box 1).

Interestingly, although some avirulence genes have been identified in fungi, no conserved motifs have been found in their sequences, and searches for RXLR motifs have proved fruitless so far (Box 1). The fact that oomycete effector genes exhibit diversifying selection, however, provides a means of identifying potential fungal effectors, assuming that fungal effectors are also delivered into plant

### Box 1. Outstanding questions

One of the most intriguing questions to emerge from recent studies of oomycete plant pathogens is ‘Why do there seem to be so many oomycete effectors?’ Bacterial pathogens of both plants and animals typically encode between 30 and 60 type-III effector proteins, but oomycetes seem to contain  $>300$  such proteins, based on initial genome analysis, and perhaps as many as 600 in certain cases [8,9]. What could be the function of so many effectors? Does this mean, for instance, that effectors also serve morphological functions in oomycetes, perhaps enabling formation of the invasive haustoria by which oomycetes feed within infected plant hosts or facilitating cell-to-cell movement of the pathogen? Formation of haustoria clearly requires alteration of plant cytoskeletal organization, in addition to suppression of plant defences. How is this accomplished, and does it require oomycete mimics of plant signalling proteins, as observed among effectors of some human pathogenic bacteria (see, for example, Ref. [22])? How many effectors are required for developmental regulation and how many for modulation of host defences? How much redundancy exists in effector function, and are there effectors that are absolutely crucial for plant infection? Are there patterns in the types of effectors that have subsequently become recognized by plants as avirulence proteins? The opportunity to determine the full spectrum of biological functions fulfilled by oomycete effectors and their evolutionary relatedness among distinct pathogens is an exciting prospect.

Another important question is whether plant pathogenic fungi require a large repertoire of effector proteins. Fungi share many features with oomycetes; they are also filamentous, osmotrophic eukaryotic microbes that form infection structures and haustoria. In addition, pathogenic fungi can also possess large sets of secreted proteins, although their number varies greatly among different pathogenic species and no clear host-cell targeting motif has yet been identified from secreted fungal proteins [23]. It will, therefore, be important to determine whether effector proteins fulfil the same range of functions in fungi as in oomycete pathogens, or whether the capacity of fungi to use secondary metabolites to modulate host metabolism and plant defence signalling is, instead, even more important than previously realized [23].

cells and are, thus, subject to the same evolutionary pressures to avoid recognition. With the advent of next-generation DNA-sequencing technologies, it is becoming viable to sequence several different strains of the same fungal species when a reference genome is available. This will enable researchers to identify genes that show evidence of a high degree of positive selective pressure, which could potentially lead to a set of likely effector candidates involved directly in plant–pathogen interactions. A new era of investigating eukaryotic effector function in plant diseases is progressing rapidly and is likely to provide new insight into the molecular arms race between plant hosts and invading microbes.

### References

- 1 Kamoun, S. (2003) Molecular genetics of pathogenic oomycetes. *Eukaryot. Cell* 2, 191–199
- 2 Arnold, D.L. *et al.* (2007) Evolution of microbial virulence – the benefits of stress. *Trends Genet.* 23, 293–300
- 3 Alfano, J.R. and Collmer, A. (2004) Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.* 42, 385–414
- 4 Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nature* 444, 323–329
- 5 Flor, H.H. (1971) Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9, 275–296
- 6 Birch, P.R. *et al.* (2006) Trafficking arms: oomycete effectors enter host plant cells. *Trends Microbiol.* 14, 8–11

- 7 Whisson, S.C. *et al.* (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450, 115–118
- 8 Win, J. *et al.* (2007) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* 19, 2349–2369
- 9 Jiang, R.H. *et al.* (2008) RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4874–4879
- 10 Allen, R.L. *et al.* (2004) Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* 306, 1957–1960
- 11 Rehmany, A.P. *et al.* (2005) Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two *Arabidopsis* lines. *Plant Cell* 17, 1839–1850
- 12 Shan, W. *et al.* (2004) The Avr1b locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene Rps1b. *Mol. Plant Microbe Interact.* 17, 394–403
- 13 Armstrong, M.R. *et al.* (2005) An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7766–7771
- 14 Hiller, N.L. *et al.* (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* 306, 1934–1937
- 15 Bhattacharjee, S. *et al.* (2006) The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog.* 2, e50
- 16 Bos, J.I. *et al.* (2006) The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant J.* 48, 165–176
- 17 Dou, D. *et al.* (2008) RXLR-mediated entry of *Phytophthora sojae* Avr1b into soybean cells does not require pathogen-encoded machinery. *Plant Cell*, DOI: 10.1105/tpc.107.056093
- 18 Tyler, B.M. *et al.* (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313, 1261–1266
- 19 Win, J. *et al.* (2006) Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. *Fungal Genet. Biol.* 43, 20–33
- 20 Dou, D. *et al.* (2008) Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell* 20, 1118–1133
- 21 Yang, Z. *et al.* (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449
- 22 Alto, N.M. *et al.* (2006) Identification of a bacterial type III effector family with G protein mimicry functions. *Cell* 124, 133–145
- 23 Soanes, D.M. *et al.* (2007) Insights from sequencing fungal and oomycete genomes: what can we learn about plant disease and the evolution of pathogenicity? *Plant Cell* 19, 3318–3326

0966-842X/\$ – see front matter © 2008 Elsevier Ltd. All rights reserved.  
doi:10.1016/j.tim.2008.08.002 Available online 25 September 2008

## Letters

# Time to remove the model organism blinkers

Julian Parkhill

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

In this issue, Whitworth [1] suggests that the primary function of bacterial comparative genomics is to generate context for model organisms and that, according to his calculations, we now have about enough bacterial genomes, and valuable resources could be better applied elsewhere. However, this analysis rests on the underlying assumption that most microbiologists work only on a limited number of individual organisms and look at the world through these model organism blinkers. Although the study of model organisms has value for the in-depth understanding of a specific microbial system, it should be remembered that the choice of model organisms was ultimately determined by historical contingency. Model organisms are generally only important because they are well studied and we have, therefore, invested in them, not because they are intrinsically more interesting or more important than any other organism. Exclusively studying model organisms blinds us to one of the most remarkable aspects of the microbial world; its incredible, almost overwhelming, diversity. Continued (and expanded), large-scale sequencing of bacteria will enable us to explore this diversity, both at the level of the biosphere and at the level of the species.

Bacterial diversity within the environment (and that includes our own bodies) is enormous, and for the most part unstudied and uncultured [2]. To suggest that the only

reason for sequencing these unknown organisms is to give us context for our pet laboratory strains is misguided. These organisms are worth exploring for their own sake, to understand their biology and how it affects our environment and our health. Genome sequence will be the fundamental enabler of these studies, especially for uncultivated organisms. For example, metagenomic approaches have shown us that previously unknown bacterial rhodopsins are actually ubiquitous and important light-harvesting pigments in the oceans [3], and sequence-based population sampling has identified the potential for very common health problems, such as obesity, to be affected by our resident microbiota [4]. Understanding the contribution of individual microorganisms to these processes will require many more genome sequences.

Within-species diversity can be very great or very little, but at both ends of the spectrum, continued sequencing will enable us to explore important biological and medical problems. One of the fundamental model organisms for microbiologists has historically been *Escherichia coli* and, probably, more is known about this organism on a molecular level than any other. However, genomics has shown us that different strains of *E. coli* can differ from each other by the presence and absence of many hundreds of genes, with significant effects on host range, pathogenicity, nutrient utilization and core biochemistry [5]. Whereas Jacques Monod said that ‘What is true for *E. coli* is true for the elephant’, a genomic perspective tells

Corresponding author: Parkhill, J. (parkhill@sanger.ac.uk).