Plant Pathology (2016) **65**, 767–781 Doi: 10.1111/ppa.12469

Comparative analysis of expressed CRN and RXLR effectors from two *Plasmopara* species causing grapevine and sunflower downy mildew

P. Mestre[ab]*, S. Carrere[cd], J. Gouzy[cd], M.-C. Piron[ab], D. Tourvieille de Labrouhe[e], P. Vincourt[cd], F. Delmotte[fg] and L. Godiard[cd]*

[a]*INRA, UMR 1131 Santé de la Vigne et Qualité du Vin, F-68000, Colmar;* [b]*Université de Strasbourg, UMR 1131 Santé de la Vigne et Qualité du Vin, F-68000 Colmar;* [c]*INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, F-31326 Castanet-Tolosan;* [d]*CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, F-31326 Castanet-Tolosan;* [e]*Développement d'outils de phénotypage 'Maladies', UE PHACC 1375, INRA, 5 Chemin de Beaulieu, F-63039 Clermont-Ferrand Cedex 2;* [f]*Institut des Sciences de la Vigne et du Vin, UMR 1065 Santé et Agroécologie du Vignoble, INRA, F-33883 Villenave d'Ornon; and* [g]*ISVV, UMR 1065 SAVE, Université de Bordeaux, F-33883 Villenave d'Ornon France*

Plasmopara halstedii and *Plasmopara viticola* are the causative agents of downy mildew of two economically important crops, sunflower and grapevine, respectively. These phylogenetically related oomycetes are obligate biotrophs belonging to Peronosporales but have different pathological profiles. Despite the economic importance of diseases caused by both *Plasmopara*, genomic resources are very limited and their effector repertoire is unknown. In this study parallel transcriptome sequencing of *P. halstedii* and *P. viticola* was performed, and a *Plasmopara* species cDNA database (PlasmoparaSp) released containing 46 000 clusters. In oomycetes, two classes of effectors are translocated into the host cytoplasm, the RXLRs and the CRNs. The PlasmoparaSp database was screened in order to identify the repertoire of expressed effectors used by both *Plasmopara* species; about 50 putative RXLR and 60 CRN were identified for each. These effectors were compared within both species and with seven publicly available oomycete species representative of Peronosporales and Albuginales. Sequence analyses revealed the presence of 55 RXLR families, 12 of them shared by both *Plasmopara*, and 19 showing amino acid conservation with predicted peptides from at least one oomycete species. Analyses of *Plasmopara* spp. CRN C-terminal variable regions revealed sequence conservation inside *Plasmopara* spp. and across oomycetes, with the exception of *Hyaloperonospora arabidopsidis*. Finally, the analyses confirmed the presence of eight CRN C-terminal domains described in *Phytophthora infestans* and identified CRN effectors showing similarity to serine proteases. Species-specific effectors were identified that may be involved in host specificity, as well as effectors conserved among oomycetes that might play important roles in the pathogen biology.

Keywords: cDNA sequencing, downy mildew, grapevine, oomycete effectors, *Plasmopara*, sunflower

Introduction

Successful infection of plants by pathogens such as bacteria, fungi or oomycetes, relies on pathogenicity factors called effectors that modify the metabolism of the host to their benefit (Bozkurt *et al.*, 2012). Oomycetes constitute a distinct phylogenetic lineage of stramenopile eukaryotes closely related to photosynthetic brown algae and diatoms. The Peronosporales, containing 1300 species, forms the largest and most important group of oomycetes. Many species in the Peronosporales are pathogenic on plants, causing diseases on hundreds of species, including the devastating potato late blight pathogen *Phytophthora infestans* (Thines & Kamoun, 2010). The Peronosporales include oomycetes with different life styles: biotrophs, requiring a living host plant to accomplish their life cycle (*Bremia*, *Hyaloperonospora*), necrotrophs (*Pythium*) and hemibiotrophs (*Phytophthora* spp.), whose life cycle combines biotrophic and necrotrophic stages. Obligate biotrophic oomycetes, causing downy mildew, constitute the largest group in the Peronosporales, with 800 species grouped into eight genera including the crop pathogens *Bremia*, *Hyaloperonospora*, *Peronospora*, *Plasmopara* and *Pseudoperonospora*. Important plant pathogenic oomycetes are also found in the Albuginales, which is composed exclusively of biotrophs that have wide host ranges, such as *Albugo candida* (Thines & Kamoun, 2010).

Plasmopara halstedii and *Plasmopara viticola* are the causative agents of downy mildew of two economically important crops, sunflower and grapevine, respectively (Viennot-Bourgin, 1949). They are phylogenetically clo-

*E-mail: laurence.godiard@toulouse.inra.fr; pere.mestre@colmar.inra.fr

Published online 20 November 2015

Plant Pathology

sely related obligate biotrophs belonging to Peronosporales and have a similar genome size, estimated at approximately 110 Mb (Voglmayr & Greilhuber, 1998; Gascuel *et al.*, 2015). Both *Plasmopara* spp. are invasive plant pathogens in Europe, of American origin (Ahmed *et al.*, 2012; Fontaine *et al.*, 2013). They have a narrow host range, restricted to some Compositae plants for *P. halstedii* and to the *Vitis* genus for *P. viticola*, and induce similar leaf disease symptoms consisting of distaining of infected regions and sporulation on the lower leaf side (Viennot-Bourgin, 1949). There are 36 identified *P. halstedii* pathotypes worldwide, named by an international system for pathotype nomenclature (Gascuel *et al.*, 2015), 16 of which are found in France today (Ahmed *et al.*, 2012). The adaptation of *P. viticola* to grapevine resistance has resulted in the emergence of virulent isolates in Europe (Peressotti *et al.*, 2010; Delmotte *et al.*, 2014). Besides differences in host range, *P. halstedii* infects sunflower germlings and young plantlets through roots, hypocotyls and leaves (Gascuel *et al.*, 2015), whereas *P. viticola* infects grapevine leaves, tendrils and fruits. They both produce haustoria in infected cells and share similar life cycles, with primary spring and secondary summer infections and dissemination through motile biflagellate zoospores (Viennot-Bourgin, 1949). Comparing effectors from both *Plasmopara* spp. should help to understand their role in pathogen biology and host adaptation and to identify conserved effectors putatively involved in essential virulence functions.

Oomycete effectors are divided into apoplastic effectors that function in the extracellular space between the pathogen and the host cell, and cytoplasmic effectors, which are translocated into the cytoplasm of the plant cell targeting different cellular compartments (Bozkurt *et al.*, 2012; Stam *et al.*, 2013). Two major classes of cytoplasmic oomycete effectors have been described. They are characterized by the presence of conserved amino acid motifs in their N-terminal region: the RXLR class, containing an RXLR followed by (D)EER motif, and the Crinkler (CRN) class, containing an LXLFLAK motif (Torto *et al.*, 2003; Rehmany *et al.*, 2005). The RXLR-(D)EER and LXLFLAK motifs have been described as essential for the translocation of effector proteins into the plant cell (Whisson *et al.*, 2007; Schornack *et al.*, 2010). Although RXLR and CRN effectors are expected to present in their N-terminal region a signal peptide allowing them to exit the oomycete cell, there are reports of CRNs that do not contain a canonical signal peptide (Stam *et al.*, 2013).

CRN effectors present a characteristic modular structure composed of an N-terminal conserved region including the LXLFLAK motif and extending until an HVLVVVP motif, followed by variable C-terminal domains. CRNs are secreted proteins initially described as inducing crinkling and necrosis following their ectopic expression *in planta* (Torto *et al.*, 2003). The necrotic activity of several CRNs has been shown to be dependent on their targeting to the nucleus of the host plant (Schornack *et al.*, 2010). Haas *et al.* (2009) described 36 types of CRN C-terminal domains and eight additional singletons in *Phytophthora infestans*, most of which are also found in other *Phytophthora* species (Stam *et al.*, 2013). However, most of these domains do not seem to correspond to any known function, and the role of most CRN proteins is still unknown, with the exception of *P. infestans* CRN8, which shows kinase activity *in planta* and is important for full virulence (van Damme *et al.*, 2012), and two CRN effectors from *Phytophthora sojae* that were shown to be involved in pathogen virulence and in the induction or suppression of plant cell death (Liu *et al.*, 2011).

While CRNs have been described in all oomycete species, RXLR effectors *sensu stricto* seem to be restricted to the Peronosporales group (Schornack *et al.*, 2010). An alternative QXLR translocation motif has been described in secreted proteins of *Pseudoperonospora cubensis* (Tian *et al.*, 2011) and specific *in planta* recognition has been shown for GKLR secreted proteins of *Bremia lactucae* (Stassen *et al.*, 2013). Outside of Peronosporales, a small group of secreted proteins of *Albugo candida* presenting a RXL variant motif have been proposed as putative RXLRs (Links *et al.*, 2011). Several RXLRs are recognized by plant resistance (*R*) genes or act as suppressors of plant innate immunity (Vleeshouwers *et al.*, 2008; Oh *et al.*, 2009; Fabro *et al.*, 2011). Indeed most of the oomycete Avr proteins characterized belong to the RXLR effector family (Jiang & Tyler, 2012). Genomic sequences are available for four hemibiotrophic *Phytophthora* species, *P. infestans*, *P. ramorum*, *P. sojae* and *P. capsici* (Tyler *et al.*, 2006; Haas *et al.*, 2009; Lamour *et al.*, 2012) as well as the necrotroph *Pythium ultimum* (Levesque *et al.*, 2010). Available genomic resources from obligate biotrophic oomycetes include the genome sequence of *Hyaloperonospora arabidopsidis*, (Baxter *et al.*, 2010), a draft assembly of the *P. cubensis* genome (Tian *et al.*, 2011), transcriptome-based resources in *B. lactucae* (Stassen *et al.*, 2012) and, outside the Peronosporales, genomic sequences from *A. candida* and *A. laibachii* (Kemen *et al.*, 2011; Links *et al.*, 2011). Increasing the number of obligate biotrophic oomycetes for which genomic resources are available should contribute to a better understanding of their biology and may lead to the identification of effector genes linked to this particular life style.

A preliminary search for *P. halstedii* effectors in an EST library derived from infected sunflower resulted in the identification of five RXLR and 15 CRN putative effectors (As-sadi *et al.*, 2011). Sequencing of 1543 cDNA clones of an *in vitro* germinated spore library of *P. viticola* resulted in the identification of two putative RXLR effectors that are expressed upon infection (Mestre *et al.*, 2012). Nevertheless, despite the economic importance of both pathogens, genomic resources for *P. halstedii* and *P. viticola* are limited; in particular their effector repertoire is unknown.

This study presents parallel transcriptome sequencing from inoculated plant material and *in vitro* germinated spores of *P. halstedii* and *P. viticola*, and the release of a

Plasmopara species cDNA database (PlasmoparaSp). An *in silico* search was performed on the database in order to provide an inventory of the putative RXLR and CRN effectors of *P. halstedii* and *P. viticola*. These effectors were compared within both *Plasmopara* and with seven representative sequenced oomycete species.

Materials and methods

Pathogen materials and sample preparation

Sunflower (*Helianthus annuus*) line GB susceptible to all pathotypes of *P. halstedii* was infected with pathotypes 100, 304, 703 and 710 in confined growth chambers according to Mouzeyar *et al.* (1993, Supplementary References in Data S1). About 100 infected sunflower hypocotyls per pathotype were harvested when sporulation was abundant on cotyledons, between 11 and 15 days after infection. Hypocotyls were immediately frozen in liquid nitrogen and stored at −80°C. Sporulating cotyledons were placed in sterile water, in order to release zoosporangia. *In vitro* germinated *P. halstedii* zoospores were obtained as described in Riemann *et al.* (2002, Suppl. References in Data S1), using an incubation time of 4 h before addition of 10 mM NaCl for 1 h. After checking zoospore germination with a microscope, zoospores were collected by centrifugation as described by Mestre *et al.* (2012) and frozen in liquid nitrogen before storage at −80°C.

Plasmopara viticola isolates SC and SL (Peressotti *et al.*, 2010) were used to inoculate the susceptible *Vitis vinifera* 'Muscat Ottonel'. For the preparation of cDNA libraries, *P. viticola* infected leaves and germinated zoospores were obtained as described by Mestre *et al.* (2012). *Plasmopara viticola*-infected leaves were sampled 48 h post-infection.

RNA extraction and cDNA synthesis

Total RNA from germinated zoospores was extracted using the RNeasy Plant Mini kit (QIAGEN) following the manufacturer's instructions. Total RNA from infected sunflower hypocotyls was extracted using the RNeasy Plant Midi kit (QIAGEN). Total RNA from infected grapevine leaves was extracted according to Zeng & Yang (2002, Suppl. References in Data S1). Residual genomic DNA was removed by DNase treatment of RNAs using the Ambion Turbo-DNAfree kit (Life Technologies). RNA quantity and quality were checked as described in As-sadi *et al.* (2011).

Clontech SMARTer cDNA synthesis kit (Clontech) was used to synthesize all double-stranded cDNAs. For *P. halstedii*, cDNAs were synthesized from 4 μg total RNA by long distance (LD)-PCR using 18, 20 or 23 cycles of amplification depending on the amplification tests performed on each of the eight samples. For *P. viticola*, cDNAs were synthesized from 1 μg of total RNA by LD-PCR using 19 cycles of amplification.

cDNA sequencing and sequence assembly

Eight *P. halstedii* and four *P. viticola* libraries each consisting of 11 to 18 μg ds cDNAs were sequenced at Genoscope (Evry, France). The sequencing runs were performed on a Roche 454 GS-FLX Titanium sequencer following the manufacturer's recommendations. Initial cleaning of the sequences consisted of removing *in silico* the incorporated adaptors used during cDNA synthesis (CAP and polyA adaptors). Ribosomal sequences and reads shorter than 40 nucleotides were removed. Clustering was carried out for each species separately with a modified version of TGICL (Pertea *et al.*, 2003, Suppl. References in Data S1) named TGICL++. Briefly, the TGICL++ package was optimized to accommodate very large data sets. Using NRCL and TCLUST, tools available in the TGICL package, the TGICL++ pipeline performed successive clustering steps, being very strict at first then increasingly permissive. The starting parameters were a match identity of 97% with an overlap of at least 100 bases. Parameters for final assembly were an overlap length cut-off superior to 40 bp and an overlap percentage identity cut-off superior to 97%. For *P. halstedii* sequences, TGICL (-p 97 -l 40) was run on the cleaned data from the eight libraries generated in this study merged with *H. annuus* and *P. halstedii* sequences available from public domains (August 2010), and from HP database (As-sadi *et al.*, 2011). For *P. viticola* sequences, the same process was used for the clean-up of sequences from the four libraries merged with *V. vinifera* sequences available from public domains (August 2010) and from a *P. viticola* germinated zoospore library (Mestre *et al.*, 2012).

Preliminary draft genomic sequences were obtained following Illumina HiSeq paired-end sequencing of genomic DNA at 100× coverage. Genome assemblies were done using a combination of VELVET (Zerbino and Birney, 2008, Suppl. References in Data S1), SOAPDENOVO and SOAPGAPCLOSER (Luo *et al.*, 2012, Suppl. References in Data S1) software. Assemblies featured as follows, after removing clusters smaller than 500 nt: *P. halstedii* (Num: 7276, N50 bp: 56 667, N50num: 376), *P. viticola* (Num: 48 820, N50 bp: 5535, N50num: 5139). Comparison of cDNA sequences to draft genomic sequences was performed with GENOMETHREADER v. 1.3.1 (-dpminintronlen 39 -exdrop 1 -gcmincoverage 50).

Sequence analysis

All BLAST (Altschul *et al.*, 1997, Suppl. References in Data S1) analyses were performed on the PlasmoparaSp portal (https://www.heliagene.org/PlasmoparaSpecies), which provides tools for exploring gene function and protein families and includes the databases listed in Table S1.

PSI-TBLASTN (PSSM) was performed independently on each *Plasmopara* database using the annotated sequences for oomycete RXLR and CRN effectors available at NCBI in March 2010 as models and BLASTPGP software (v. 2.2.26) with the following arguments: Expectation value (e): 1e–4, Filter query sequence with SEG (F): False, e-value threshold for inclusion in multipass model (h): 1e–4, Maximum number of passes to use in MULTIPASS v. (j): 10. PATSCAN (Dsouza *et al.*, 1997, Suppl. References in Data S1) analyses were performed on the PlasmoparaSp database on the GETORF predicted longest peptides from the two *Plasmopara* spp. using the pattern ^M 25…50 (R,H,Q)X(L,Y,F,M)(R,Q,G) and allowing only one position to be different from the canonical RXLR. Additionally, the pattern ^M 30…90 (D,E)EER was used to identify proteins carrying the dEER motif.

Effector families inside *Plasmopara* spp. were established using a BLASTN (v. 2.0MP-WashU) analysis on the PlasmoparaSp database using the effector set from each species as a query. Entries producing hits (*E* value <10e–3, score >100) were considered as belonging to the same family. Effector families among oomycetes were established using a TBLASTN analysis on the predicted transcriptome of seven oomycete species using the effector set from each *Plasmopara* species as a query. Entries producing hits (*E* value <10e–4) were considered as belonging to the same family.

Prediction of signal peptides and transmembrane domains was done with SIGNALP v. 4.1 and TMHMM v. 2.0, respectively, at

CBS (http://www.cbs.dtu.dk/services). Structural homology prediction was performed using the HHpred server (http://toolkit.tuebingen.mpg.de/hhpred) with default parameters. The probability reported by HHpred is the probability that the database match is homologous to some part of the query sequence, and can be interpreted literally (Soding *et al.*, 2005, Suppl. References in Data S1). Sequence alignments were performed with MUSCLE implemented in SEAVIEW (Gouy *et al.*, 2010, Suppl. References in Data S1) and identities were highlighted using BOXSHADE at Swiss EMBnet (http://www.ch.embnet.org/). Phylogenetic trees were constructed using PHYML after alignment curation with GBLOCKS, with LG model and 100 bootstrap repetitions, performed at http://www.phylogeny.fr (Dereeper *et al.*, 2008, Suppl. References in Data S1).

All reads obtained are available at SRA (Sequence Read Archive), accession numbers SRR1141084 to SRR1141091 for *P. halstedii*, and SRR1141092, SRR1141093, SRR1138574 and SRR1138575 for *P. viticola*.

Results

cDNA sequencing: statistics and clustering

To create genomic resources for the identification of expressed effectors, cDNA was sequenced from both *Plasmopara* spp. Total RNA was collected from inoculated plant material and from *in vitro* germinated spores, the latest stage of the pathogen life cycle that can be obtained *in vitro*. Four pathotypes were used for *P. halstedii* (100, 304, 703 and 710) and two isolates for *P. viticola* (SL and SC).

cDNA sequencing using Roche 454 and cleaning of reads resulted in 843 116 and 996 983 reads from *P. halstedii* and *P. viticola* tissues, respectively. Detailed results are presented in Table S2. Because of the presence of plant sequences in the cDNA, the available sunflower and grapevine cDNA or mRNA sequences from NCBI were integrated in the assembly. The assembly also included 144 *P. halstedii* EST and mRNA sequences present in GenBank (January 2009) as well as *Plasmopara* spp. EST sequences obtained previously (285 766 sequences from *P. halstedii* and 482 sequences from *P. viticola*; As-sadi *et al.*, 2011; Mestre *et al.*, 2012). Clustering was performed separately for each *Plasmopara* species.

Clustering produced 80 354 contigs and 130 678 singletons for *P. halstedii*, with 89·6% of transcripts clustered. For *P. viticola,* 67 164 contigs and 98 992 singletons were obtained, with 92·8% of transcripts forming contigs. Libraries were named *P. halstedii*+ and *P. viticola*+ to indicate the presence of sequences from plant and pathogen origin. Consensus sequences smaller than 100 nt were discarded. The mean sizes were 447 and 682 nt for *P. halstedii* and *P. viticola,* respectively. The size distributions of contigs and singletons for each library are presented in Figure 1a, b. The contigs and singletons, designated hereafter as clusters, were annotated as Plhal or Plvit followed by a six-digit number. FRAMEDP was run to predict peptides from clusters using the Swiss-Prot database as a reference library. A total of 118 976 and 109 539 peptides longer than 29 amino acids were predicted from *P. halstedii*+ and *P. viticola*+ libraries, respectively, with mean sizes 128 and 183 aa (Fig. 1a, b).

Constructing PlasmoparaSp database

In order to create a transcriptomic resource containing only *P. halstedii* and *P. viticola* clusters, a curated database was constructed from *P. halstedii*+ and *P. viticola*+ libraries. To select for oomycete clusters, preliminary draft genomic assemblies from both *Plasmopara* spp. were used, derived from Illumina paired-end sequencing of spore genomic DNA. Sequences were compared to draft genomic assemblies and those not producing hits were discarded. This step might lead to the elimination of some *bona fide* pathogen sequences, but it ensures having a set of sequences with unambiguous oomycete origin. Finally, selected sequences were compared to grapevine and sunflower genomic sequences and those producing significant hits were discarded. From the initial amount of 377 188 plant and *Plasmopara* spp. clusters from *P. halstedii*+ and *P. viticola*+ databases, 136 233 (90 019 singletons and 46 214 contigs) fulfilled such criteria and were considered as expressed sequences from either oomycete. For quality reasons, only contigs were included in the database.

The selected 46 214 clusters constituted a new database, named PlasmoparaSp (https://www.heliagene.org/PlasmoparaSpecies). The mean cluster size is 621 nt. In total, 28 862 peptides were predicted, 17 417 from *P. halstedii* and 11 445 from *P. viticola*, with a mean size of 163 aa (Fig. 1c).

The PlasmoparaSp portal gives a summary of information on each cluster and its predicted peptide (sizes, numbers and sequences of reads, protein patterns detected by INTERPROSCAN) as well as BLAST results obtained with reference libraries (Fig. 2). It allows the user to search for motifs and to undertake analyses on selected databases.

Identification of *Plasmopara* spp. candidate effector proteins

The following searches were performed in order to identify *Plasmopara* spp. candidate effector proteins: (i) PSI-BLAST of the whole data set against a library of oomycete effector sequences; and (ii) a motif search on the PlasmoparaSp database using PATSCAN.

The procedure for identification of candidate RXLR effector proteins is described in Figure 3a. The PATSCAN search took account of the different variations of the RXLR motif allowing cell entry as well as the QXLR motif found in *P. cubensis*. The pattern (R,H,Q)X(L,Y,F,M)(R,Q,G) was used, allowing only one position to be different from the canonical RXLR (Fig. 3a). In addition, proteins carrying the dEER motif in the first 90 amino acids were identified using the pattern (D,E)EER. The motif GKLR found in *B. lactucae* was also searched for. Output sequences from PSI-BLAST and PATSCAN searches were screened for completeness, presence of signal peptide, absence of transmembrane domains, and selected

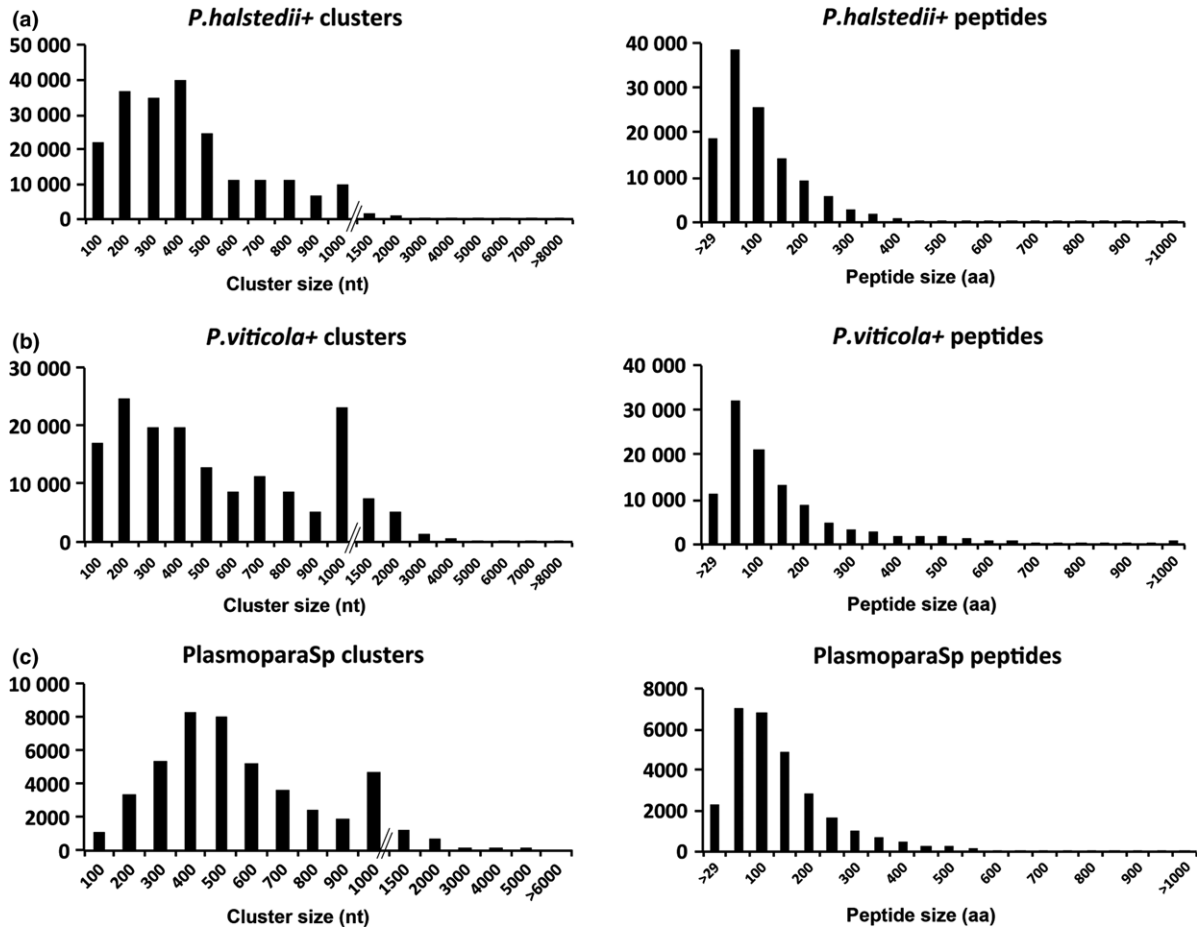


Figure 1 Distribution of clusters and peptides in *Plasmopara* spp. databases. Size distribution of clusters (left) and peptides (right) from *P. halstedii+* (a), *P. viticola+* (b) and PlasmoparaSp (c) databases. Diagonal stripes show scale break in the cluster size axis.

based on their sizes. The identity of the sequences was confirmed by comparing them with draft genomic assemblies. Finally, comparison with available oomycete sequences allowed identification of eventual false positives in the form of strong hits to core/structural proteins and/or proteins too large (>400 aa).

Results of the search are presented in Table 1 and Figure 3b. Forty-nine RXLR were identified for *P. halstedii* and 45 for *P. viticola*, with average sizes of 147 and 200 aa, respectively. For both species, the PatScan-based strategy was more successful than psi-blast. Eleven *P. halstedii* and nine *P. viticola* RXLRs were identified using both methods, making them strong candidate RXLR genes.

The procedure for identification of candidate CRN effectors is described in Figure 3c. It is the same as for the RXLRs but excluding the search for a signal peptide. Search for CRNs resulted in 54 candidate genes for *P. halstedii* and 60 for *P. viticola*, with average sizes of 254 and 223 aa, respectively (Table 1; Fig. 3d). A considerable number of identified proteins were relatively small (Table 1), appearing to encode a small C-terminal motif. The psi-blast strategy was more efficient for the identification of CRN effectors.

Accessions corresponding to RXLR and CRN effectors are listed in Table S3.

Analysis of RXLR effectors

To compare the RXLR genes identified for both *Plasmopara* spp., a blastn analysis (*E* value < 10e–3, score >100) was performed on the PlasmoparaSp database using the set of effector genes from each species as a query. This should also identify other putative RXLR genes not detected by PatScan and/or psi-blast searches. *Plasmopara viticola* RXLR search provided 212 sequences, 199 from *P. viticola* and 13 from *P. halstedii*, forming 29 groups on the basis of nucleotide sequence identity. *Plasmopara halstedii* RXLR search provided 157 sequences, 138 from *P. halstedii* and 19 from *P. viticola*, forming 33 groups. The exclusivity of the groups, i.e. absence of sequences belonging to more than one group, was verified manually. Each of these groups will hereafter be referred to as a family (Table S4). Several RXLR families contained only members from one species, whilst 12 families contained RXLR effectors from both *Plasmopara* spp. Thus species-specific RXLR

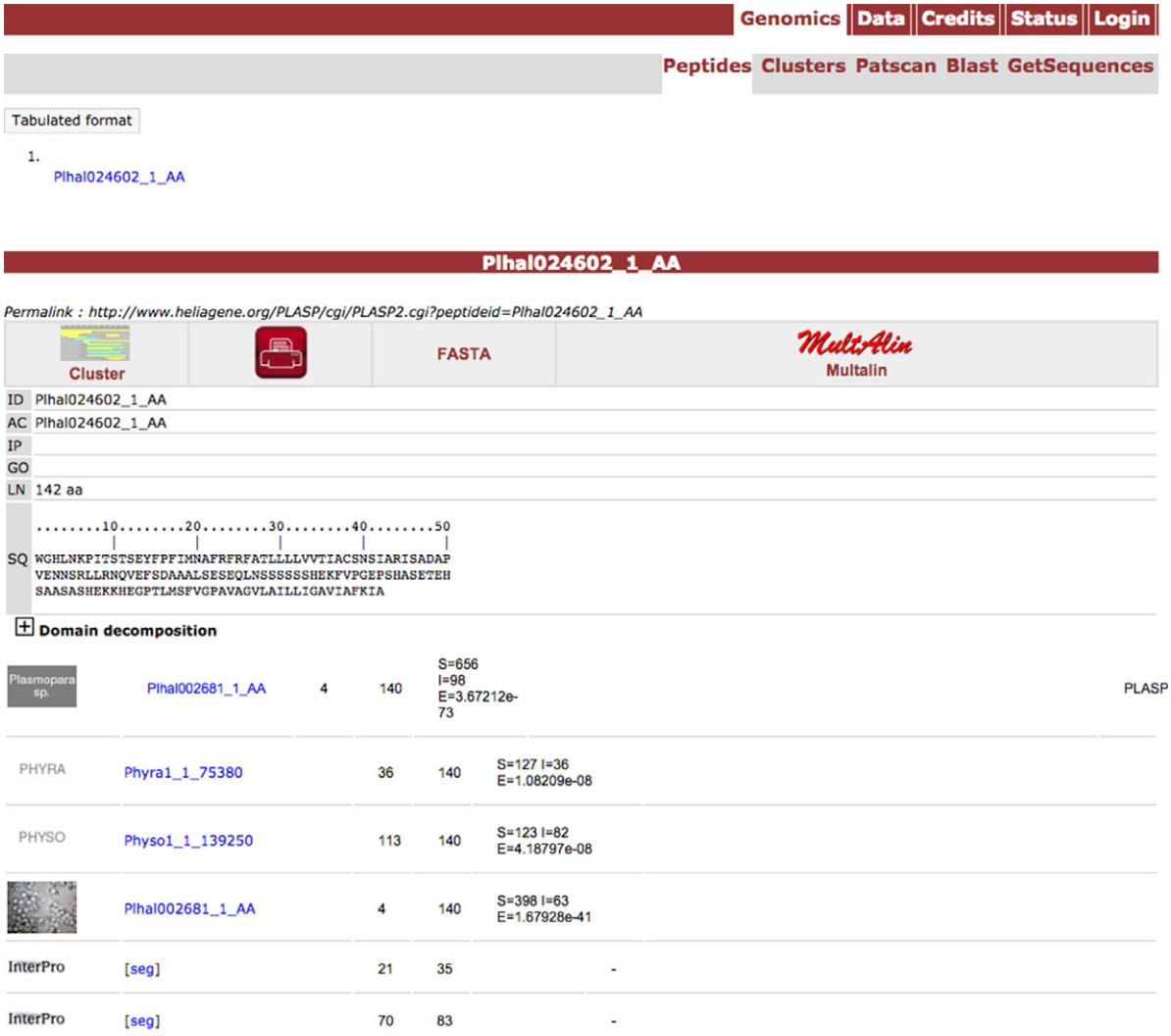


Figure 2 The PlasmoparaSp database. Screenshot of an entry from the PlasmoparaSp database showing the different information and tools available.

effectors as well as effectors conserved in both *Plasmopara* species were identified.

To assess the presence of the *Plasmopara* spp. RXLR families across oomycetes, a TBLASTN analysis (*E* value < 10e–4) was performed on the predicted transcriptome of seven oomycete species. For the individual genes, 28 out of 45 *P. viticola* and 26 out of 49 *P. halstedii* proteins showed hits with at least one other oomycete (Table 2). Nine out of 43 families specific to each *Plasmopara* spp. showed hits with other oomycetes. Ten out of 12 families with members from both *Plasmopara* spp. presented hits to at least one *Phytophthora* species, four had hits to *P. cubensis*, three to *H. arabidopsidis* one to *P. ultimum* and none to *A. candida*. Interestingly, three families (PlvitRxL10-PlhalRxL30, PlvitRxL15, and PlvitRxL58) were reduced in *P. halstedii* compared to other oomycetes (Table S5). The RXLR effector complement from *Plasmopara* spp. seems to be more closely related to *Phytophthora* spp. than to other oomycetes.

To perform detailed analysis of conservation of *Plasmopara* spp. RXLR families, the sequences of the predicted RXLR transcripts were corrected using the draft genomic assembly and the encoded proteins were used in a BLASTP analysis (*E* value < 10e–10) against all oomycete species (including *Plasmopara* spp.). Analysis of sequence variability revealed RXLR effectors conserved between oomycete species. Plvit007953 and Plhal024201, belonging to family PlvitRxL11-PlhalRxL39, showed similarity to *P. sojae* Avr1b (both 41% identity and 93% coverage), and Plvit007530, belonging to family PlvitRxL10-PlhalRxL30, was slightly similar to *P. sojae* Avh147 (28% identity and 52% coverage). It is worth noting that Plvit000250, a member of family PlvitRxL56, is predicted to contain a Nudix-hydrolase domain but did not show hits to *P. sojae* Avr3b, containing the same domain. The best conservation levels were obtained with *Plasmopara* spp. families showing few hits to *Phytophthora* spp. Figure 4 shows the phylogenetic analysis for

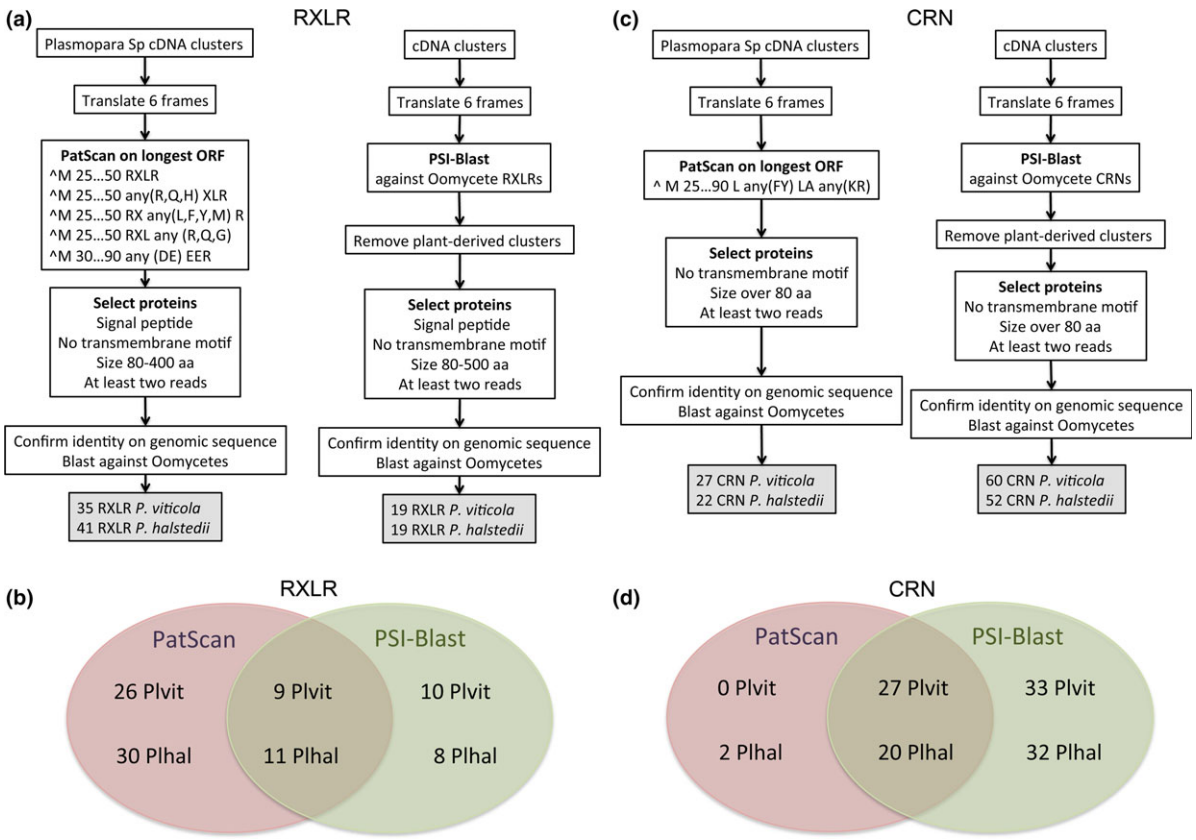


Figure 3 Identification of *Plasmopara* spp. RXLR and CRN effectors. (a) Flowchart for the identification of RXLR effectors. (b) Results of the identification of RXLR effectors using strategies described in (a). (c) Flowchart for the identification of CRN effectors. (d) Results of the identification of CRN effectors from searches described in (c).

two of such families. For all families analysed, the effectors from both *Plasmopara* spp. clustered together inside each family, separated from those from other oomycetes (Fig. 4), bringing together the two closely related species, as one would expect from conserved proteins.

Analysis of CRN effectors

CRN effectors from both *Plasmopara* species were used as queries in TBLASTN analysis (*E* value < 10e–4) against the PlasmoparaSp transcriptome and the predicted transcriptome of the same oomycete species used for the analysis of RXLR effectors. Results revealed hits with all oomycetes (Table S6), but most hits corresponded to the N-terminal part of the protein including the conserved LXLFLAK motif and the DWL domain. To gain better knowledge of the level of conservation of *Plasmopara* spp. CRNs among oomycete species, the sequences of CRN proteins were cut in two parts: a conserved, N-terminal portion, from the start codon to the HVLVVVP motif, and a C-terminal portion starting after HVLVVVP. For further analysis, proteins containing at least 100 aa C-terminal to the HVLVVVP motif were kept, resulting in 15 and 24 CRN effectors, respectively, for *P. viticola* and *P. halstedii*. The C-terminal parts of these proteins were used in a BLASTP analysis (*E* value < 10e–10) against all oomycete species, including *Plasmopara* spp. All 15 *P. viticola* proteins and 23 out of 24 *P. halstedii* proteins showed hits with at least one other oomycete (Table 3). Numbers of genes from each oomycete species showing similarity to the *Plasmopara* spp. CRNs are shown in Table S7. The CRN complement from *Plasmopara* spp. appears to be closely related to *P. infestans* and *P. cubensis*. Interestingly, no similar proteins were found in the obligate biotroph *H. arabidopsidis*.

The *Plasmopara* spp. CRN C-terminal regions were screened for the presence of conserved domains as defined by Haas *et al.* (2009) using BLASTP (*E* value < 10e–10). Eight among 36 described C-terminal domains (DBE, DBF, DFB, DN5, DN17, DO, DXZ and newD2) and two out of eight described singleton regions (SN4 and SN6) were found in the *Plasmopara* spp. C-terminal regions, with some proteins presenting domain combinations (Table 4). The DFB domain has not previously been reported in biotrophic oomycetes.

Analysis of conserved CRN C-terminal regions

Plasmopara spp. C-terminal regions showing strong similarity (*E* value < 10e–80) with proteins from at least one other oomycete were selected, and their level of conser-

vation among oomycetes was studied. Strong similarity was defined as a BLASTP output with an *E* value < 10e–80, which considering the size of these proteins, corresponded to >60% identity and >90% coverage of the query sequence. Motif and structural homology searches were performed in parallel, to gain insight on the putative function of these proteins. Most *Plasmopara* spp. CRN C-terminal regions presented strong hits with *P. infestans* and *P. cubensis* CRN proteins, three were highly similar to *P. ramorum* and *P. sojae* proteins and one showed a strong hit with *P. ultimum* (Table 5). It is worth noting that results obtained when comparing both *Plasmopara* spp. against each other must be taken with care, because the sequences may not be complete and thus the percentage of similarity may be misevaluated. The analysis revealed a high level of conservation for selected CRN effectors.

Table 1 Candidate RXLR and CRN effectors from *Plasmopara* spp.

Effector	*P. halstedii*	*P. viticola*
RXLR		
Total	49	45
Method		
PatScan	41	35
PSI-Blast	19	19
Common	11	9
Motif		
RXLR	12	8
RXLR-EER	8	5
(QH)XLR	0	2
(QH)XLR-EER	1	3
RX(FYM)R	0	0
RX(FYM)R-EER	0	0
RXL(QG)	5	2
RXL(QG)-EER	4	7
dEER	14	8
GKLR	0	0
Other	5	10
Size		
Average	147	200
Max	366	480
Min	80	84
CRN		
Total	54	60
Method		
PatScan	22	27
PSI-Blast	52	60
Common	20	27
Motif		
LxLFLAK	15	31
LxLFL(S/G)K	5	0
LxLFLAR	4	25
LxLYLAK	11	4
LxLSLAK	19	0
Size		
Average	254	223
Max	648	625
Min	80	88
Size distribution	28 < 200 aa	42 < 200 aa
	21 < 150 aa	21 < 150 aa
	15 < 130 aa	7 < 130 aa

Structural similarity searches using HHpred revealed putative functions for some *Plasmopara* spp. C-terminal regions, and by extension for some of the domains identified by Haas *et al.* (2009). As expected, CRNs containing the DBF domain showed a high level of structural similarity to protein kinases (>100 hits with 100% probability). CRNs containing the SN4 and SN6 singletons showed structural similarity to ATP-binding proteins (79 and 100 hits with >99% probability, respectively). The DN17 domain showed low structural similarity to restriction endonucleases (three hits with >95% probability), but the presence of coiled-coils in the hit proteins could invalidate the observation. Finally, CRNs containing the DN5 domain showed structural similarity to serine proteases from different species, from bacteria to human (19 hits with >95% probability).

Interest was then centred on entries Plhal000298Cter and Plhal000572Cter, possessing the DBF domain and thus showing structural similarity to protein kinases. Closer examination of both entries revealed that they were almost identical, so Plhal000298Cter was kept for further analysis. Figure S1a shows an alignment of oomycete full-length proteins similar to Plhal000298Cter (BLASTP *E* < 10e–10). Unlike that observed for RXLR conserved proteins, *P. viticola* and *P. halstedii* sequences did not cluster together following phylogenetic analysis (Fig. S1b). Apart from the high level of conservation of the C-terminal region from Plhal000298, it is worth noting that whilst the *P. halstedii*, *P. viticola* and *P. infestans* entries correspond to predicted CRN proteins, the remaining entries lack the typical CRN N-terminal domain. Manual examination of the sequences upstream of the predicted genes for *P. sojae*, *P. ultimum* and *A. candida* confirmed that misannotation was not responsible for the absence of conserved N-terminal CRN motifs. Thus CRN C-terminal domains were found

Table 2 Conservation of *Plasmopara* spp. RXLR effectors among oomycetes. Table shows the number of RXLR genes from *P. halstedii* (Plhal) and *P. viticola* (Plvit) showing hits to the proteome of different oomycetes following TBLASTN (*E* value < 10e–4)

Query	Plvit[a]	Plhal	Phinf	Phram	Phsoj	Hyara	Pscub	Pyult	Alcan	No hits
Plvit	45[b]	24	18	14	16	6	10	3	2	17
Plhal	14	49	16	6	10	4	2	2	0	23

[a]Plvit: *Plasmopara viticola*, Plhal: *Plasmopara halstedii*, Phinf: *Phytophthora infestans*, Phram: *Phytophthora ramorum*, Phsoj: *Phytophthora sojae*, Hyara: *Hyaloperonospora arabidopsidis*, Pscub: *Pseudoperonospora cubensis*, Pyult: *Pythium ultimum*, Alcan: *Albugo candida*.
[b]Shaded boxes indicate the total number of genes from each *Plasmopara* species used as query.

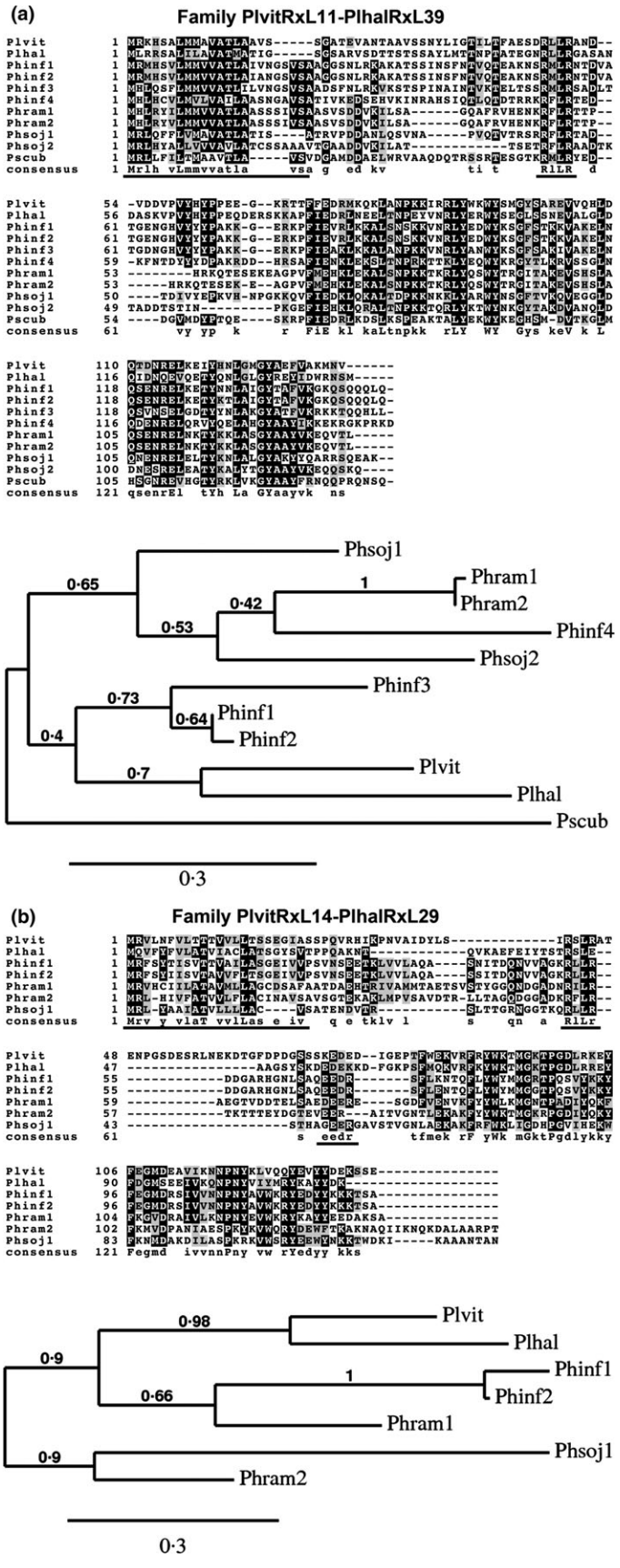

(a)
Family PlvitRxL11-PlhalRxL39
Plvit
Plhal
Phinf1
Phinf2
Phinf3
Phinf4
Phram1
Phram2
Phsoj1
Phsoj2
Pscub
consensus
Phsoj1
Phram1
Phram2
Phinf4
Phsoj2
Phinf3
Phinf1
Phinf2
Plvit
Plhal
Pscub
0·65
1
0·42
0·53
0·73
0·64
0·4
0·7
0·3
(b)
Family PlvitRxL14-PlhalRxL29
Plvit
Plhal
Phinf1
Phinf2
Phram1
Phram2
Phsoj1
consensus
Plvit
Plhal
Phinf1
Phinf2
Phram1
Phram2
Phsoj1
0·98
0·9
1
0·66
0·9
0·3

Figure 4 Conserved oomycete RXLR effector genes. Alignment of two families of conserved RXLR effectors and resulting phylogenetic trees. Predicted signal peptide, RXLR and dEER motifs are underlined. Black shading indicates identity; grey shading indicates similarity. Numbers at branches indicate bootstrap support based on 100 replicates. Oomycete abbreviations as in Table 2. (a) Family PlvitRxL11-PlhalRxL39. *Plvit*: Plvit007953, *Plhal*: Plhal024201, *Phinf1*: PITG_07594T0, *Phinf2*: PITG_18609T0, *Phinf3*: PITG_07597T0, *Phinf4*: PITG_07587T0, *Phinf5*: PITG_22828T0, *Phram1*: Phyra1_1_87087, *Phram2*: Phyra1_1_79397, *Phsoj1*: Physo1_1_135627, *Phsoj2*: Physo1_1_135626, *Pscub*: AHJF01013746.1. (b) Family PlvitRxL14-PlhalRxL29. *Plvit*: Plvit008290, *Plhal*: Plhal008929, *Phinf1*: PITG_04279T0, *Phinf2*: PITG_19526T0, *Phinf3*: PITG_19528T0, *Phram1*: Phyra1_1_84626, *Phram2*: Phyra1_1_84627, *Phsoj1*: Physo1_1_136266.

Table 3 Conservation of the C-terminal domain of *Plasmopara* spp. CRN effectors among oomycetes. Table shows number of CRN genes from *P. halstedii* (Plhal) and *P. viticola* (Plvit) showing hits to the proteome of different oomycetes following BLASTP (*E* value < 10e–10) using the C-terminal part of the protein.

Query	Plvit	Plhal	Phinf	Phram	Phsoj	Hyara	Pscub	Pyult	Alcan	No hits
Plvit	15[a]	14	15	4	8	0	15	9	2	0
Plhal	13	24	17	7	18	0	21	15	3	1

[a]Shaded boxes indicate the total number of genes from each *Plasmopara* species used as query.
Oomycete abbreviations as in Table 2.

Table 4 Effector domains and domain combinations found in *Plasmopara* spp. CRN proteins. The number of CRN proteins found for each domain is indicated (– indicates domain absent). Domains as defined in Haas *et al.* (2009).

Domain	*P. viticola*	*P. halstedii*
DBE	3	2
DBF	1	3
DFB	–	1
DN5	–	1
DN17	3	2
DO	1	2
DXZ	1	2
newD2	–	1
SN4	2	–
SN6	2	1
SN6-DO	1	–

that may be present either as independent genes or as part of a CRN in different oomycetes.

The focus then turned to entry Plhal003764Cter, which contains domain DN5 and shows structural homology to serine proteases. Alignment of C-terminal regions from oomycete proteins similar to Plhal03764C-ter (BLASTP $E < 10e{-}10$), which form a large family in *P. infestans*, revealed important conservation between species (Fig. 5a, b). Alignment of selected CRNs with proteases to which they show structural homology allowed the identification of a putative catalytic triad and revealed conserved motifs characteristic of trypsin-like serine proteases (Fig. 5c, d). One of the *P. cubensis* and both *P. sojae* entries were missing predicted catalytic sites. The results reveal the existence of a family of CRN effectors with trypsin-like protease motifs, suggesting that proteolysis may play a role in the interference with plant processes during oomycete infection.

Discussion

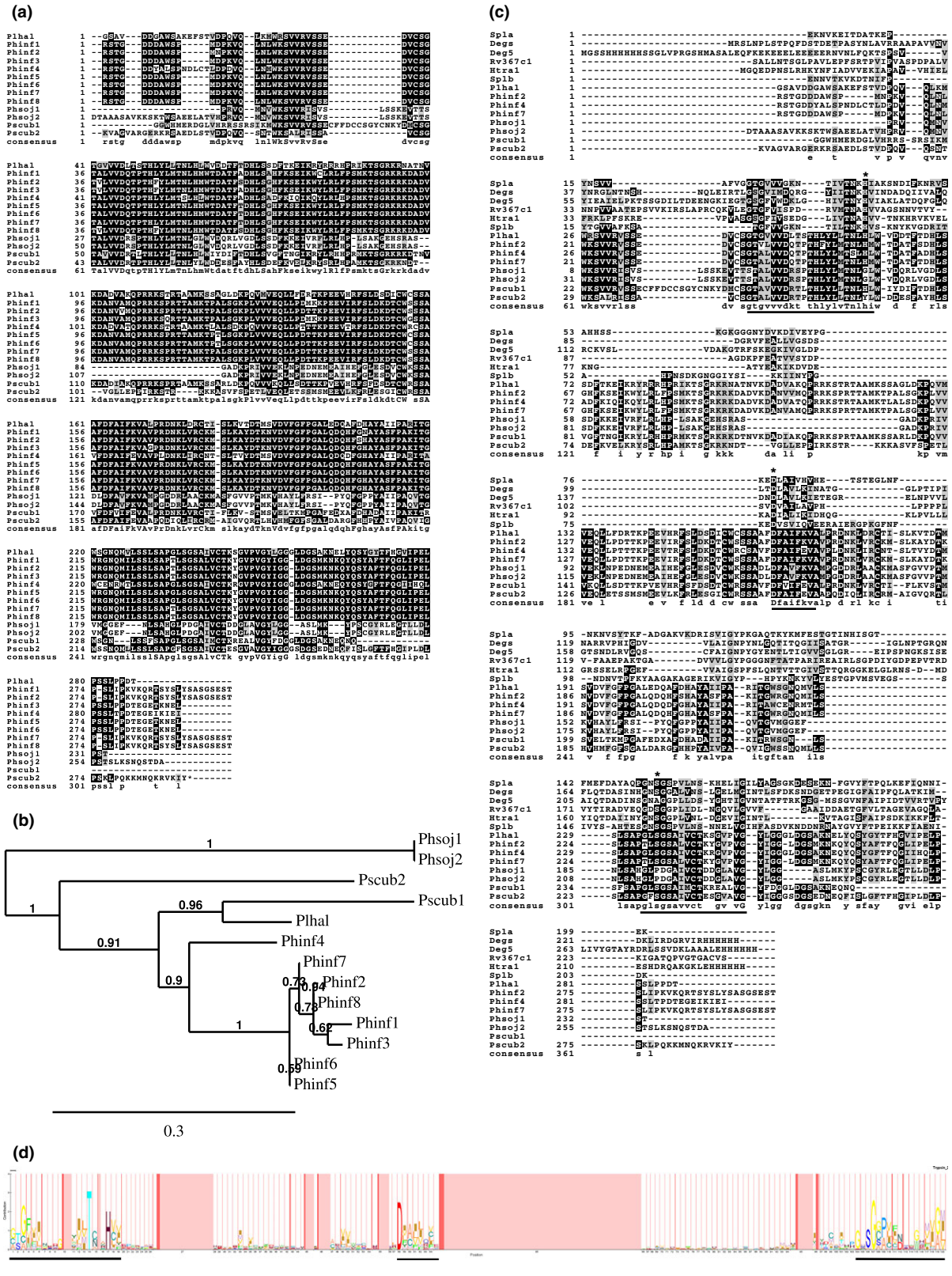
Oomycete RXLR and CRN effectors are reported as being involved in pathogenicity and host adaptation. Unveiling the effector complement from oomycete species is pivotal not only for understanding the biology of each species but also to increase knowledge about oomycete evolution and host specificity. This study described the identification of RXLR and CRN effectors from two obligate biotrophic oomycetes belonging to the genus *Plasmopara*, as well as the creation of a *Plasmopara* spp. database as a portal for sequence analysis. Comparing *Plasmopara* spp. sequences with other oomycetes resulted in the identification of eight families of RXLR effectors conserved among the Peronosporales, with two RXLRs, Plvit007950 and Plvit007530, showing similarity to *P. sojae* Avr1b and Avh147, respectively. The same analysis revealed 16 CRN effectors conserved among oomycetes. Finally, structural homology analyses identified putative functions for four C-terminal domains of CRN effectors.

Similar amounts of effector sequences were found for both *Plasmopara* species. This similarity is surprising, taking into account the fact that infected tissues from both pathosystems were sampled at two different infection stages. Although it cannot be ruled out that this is due to chance alone, it is tempting to speculate that, because this is looking at cDNA sequences, the amount of identified effectors would be representative of the set of expressed effectors in *Plasmopara* spp. It has been reported that a low percentage of RXLR effector sequences are highly or moderately expressed upon infection, with numbers of expressed RXLRs varying from 18 for *H. arabidopsidis* to 79 for *P. infestans* (Haas *et al.*, 2009; Cabral *et al.*, 2011; Wang *et al.*, 2011). The 49 and 45 RXLRs found, respectively, in *P. halstedii* and *P. viticola* are thus consistent with previous reports.

Table 5 *Plasmopara* spp. CRN protein C-terminal regions showing strong similarity to proteins from other oomycetes[a]. Best hits for each oomycete species are shown (cut-off *E* <10e–10). Note that no similar proteins were found from *Hyaloperonospora arabidopsidis*.

	P. halstedii/P. viticola		*P. infestans*		*P. ramorum*		*P. sojae*		*P. cubensis*		*P. ultimum*		*A. candida*		
CRN gene ID	ID	*E* value	PITG ID	*E* value	Phyra1_1 ID	*E* value	Physo1_1 ID	*E* value	AHJF0 ID	*E* value	PYU1 ID	*E* value	AC2VRR ID	*E* value	Cter motif
Phal000095Cter	Plvit018398	2·4e–32	NS	NS	NS	NS	136850	1·4e–99	1002671	1·4e–40	T012907	5·3e–36	NS	NS	DO
Phal000298Cter	Plvit000837[b]	4·8e–156	12619T0	0	NS	NS	NS	NS	1003702	1·4e–53	T004047	1·6e–47	s00137g84	8·9e–39	DBF
Phal000548Cter	Plvit040601	1·2e–50	19373T0	6·1e–113	72036	4·8e–85	108741	8·4e–105	1000666	3·3e–141	T012526	0	NS	NS	DXZ
Phal000572Cter	Plvit000837	3·3e–162	12612T0	0	NS	NS	NS	NS	1003702	1·5e–57	T004047	8·0e–48	s00137g84	2·4e–39	DBF
Phal001471Cter	Plvit001983	3·4e–35	18847T0	6·1e–154	82635	3·7e–124	139470	3·4e–37	1014014	1·4e–101	T012535	7·8e–56	NS	NS	DN17
Phal003764Cter	Plvit008262	8·2e–32	18133T0	3·8e–127	NS	NS	137868	1·5e–50	1000669	3·4e–83	NS	NS	NS	NS	DN5
Phal029836Cter	Plvit011188	1·6e–20	14267T0	2·8e–86	84114	0	NS	NS	1008522	3·0e–117	T010807	5·9e–28	s00022g353	1·7e–19	newD2
Phal040004Cter	NS	NS	18833T0	1·2e–108	NS	NS	139086	1·6e–25	NS	NS	NS	NS	NS	NS	DFB
Phal057705Cter	Plvit018398	2·6e–32	NS	NS	NS	NS	136850	1·7e–100	1002671	1·4e–40	T012907	6·4e–35	NS	NS	DO
Plvit000058Cter	Plhal006299	1·2e–92	23275T0	0	NS	NS	NS	NS	1009129	7·2e–168	T004343	2·6e–50	NS	NS	SN4
Plvit000085Cter	Plhal001695	3·4e–116	23285T0	2·9e–61	NS	NS	141063	1·5e–21	1001313	2·8e–99	NS	NS	s00025g7	2·6e–13	DO-SN6
Plvit000145Cter	Plhal006299	6·7e–91	23275T0	0	NS	NS	NS	NS	1009129	4·6e–157	T004343	2·9e–50	NS	NS	SN4
Plvit000293Cter	Plhal039675	2·1e–34	23275T0	9·9e–109	NS	NS	NS	NS	1009129	1·1e–107	T004343	7·6e–28	NS	NS	NA
Plvit002233Cter	Plhal000645	7·1e–39	22848T0	1·5e–33	77369	6·6e–33	132663	1·3e–35	1023367	5·3e–82	T012535	5·2e–21	NS	NS	DN17
Plvit003112Cter	Plhal000645	9·6e–42	22848T0	3·5e–35	77369	2·9e–34	132663	5·5e–37	1023367	3·6e–83	T012535	5·2e–22	NS	NS	DN17
Plvit003904Cter	Plhal040906	2·7e–34	23275T0	8·8e–85	NS	NS	NS	NS	1009129	3·4e–81	T004343	4·9e–21	NS	NS	NA

[a]*Plasmopara halstedii, Plasmopara viticola, Phytophthora infestans, Phytophthora ramorum, Phytophthora sojae, Pseudoperonospora cubensis, Pythium ultimum, Albugo candida*.
[b]Shaded boxes highlight proteins with *E* < 10e–80, which considering the size of the proteins corresponded to >60% identity and >90% coverage of the query sequence.


(a)
(b)
Phsoj1
Phsoj2
Pscub2
Pscub1
Plhal
Phinf4
Phinf7
Phinf2
Phinf8
Phinf1
Phinf3
Phinf6
Phinf5
0.3
(c)
(d)

However, it must be kept in mind that BLAST searches with the complete effector set revealed additional sequences from *Plasmopara* spp., which coded for incomplete proteins and eventually could enlarge the list of expressed effectors. It is also worth noting that by selecting the transcripts against a preliminary pathogen genome assembly, some sequences that would be missing in the genomic assembly may have been lost, as happened with some CRN genes (Table S3). A deeper transcriptome sequencing effort would be required to unveil the actual set of *Plasmopara* spp. expressed effectors.

The RXLR motif is involved in translocation inside the host cell. In the search for RXLR effectors, variations of the RXLR motif allowing protein translocation were included, as described by Kale *et al.* (2010). Candidate proteins in both *Plasmopara* spp. containing the RXL(Q, G) or HXLR motifs were found, but no protein with the RX(F,Y,M)R motif was found, suggesting that replacing the leucine (L) at the third position may have adverse consequences for the function of the protein. The search found several effectors from *P. viticola* containing the QXLR motif, which had been described only in effectors from *P. cubensis* (Tian *et al.*, 2011). However, a parallel PATSCAN search (^M 25…50 GKLR) could not find proteins containing the GKLR motif described in effectors from *B. lactucae* (Stassen *et al.*, 2013). Several of the putative effector proteins possessed only a dEER motif and no obvious RXLR motif; these effectors could work similarly to ATR5, the *H. arabidopsidis* protein recognized by RPP5, which possesses a dEER motif and is translocated into the host cell without an RXLR motif (Bailey *et al.*, 2011).

RXLR effectors from *Plasmopara* spp. were grouped into families based on nucleotide sequence identity. Forty-three out of the 55 families identified were specific for each species, whilst 12 families (22%) presented members from both *Plasmopara* spp., suggesting that the latter could play a role in the advent of the *Plasmopara* lineage. Comparing the *Plasmopara* spp. RXLR families to other plant pathogenic oomycetes resulted in the identification of RXLRs conserved between *Plasmopara* spp., *Phytophthora* spp. and *P. cubensis*, suggesting that these effectors develop important functions in the biology of these species. Other families were present only in *P. viticola* or *P. halstedii*, suggesting a possible role of such effectors on host specificity. It is worth noting that the complement of RXLR effectors from *P. halstedii* seems reduced compared to other oomycetes; indeed, three families with several members in *Phytophthora* spp. and *P. viticola* are absent or highly reduced in *P. halstedii*.

CRNs are expected to possess a signal peptide allowing them to exit the oomycete cell, but Stam *et al.* (2013) reported that 58% of the *Phytophthora capsici* predicted CRNs were missing a canonical signal peptide. SIGNALP failed to predict a signal peptide for 80% of *P. halstedii* and 52% of *P. viticola* putative CRNs. Interestingly, comparing the consensus sequence corresponding to the expected signal peptide of *Plasmopara* spp. and *P. infestans* CRNs revealed a high level of consensus conservation (Fig. S2), suggesting a functional significance. An important number of CRN effectors were small proteins (61% smaller than 200 aa) and did not seem to code for an effector domain. A similar observation was reported for *P. capsici*, where only 35% of predicted CRN genes coded for an effector domain (Stam *et al.*, 2013). The presence of small genes on cDNA sequencing data suggests that they may have functional significance.

Analysis of C-terminal effector domains of *P. infestans* CRN proteins identified 36 domains (Haas *et al.*, 2009), 27 of which have been also found in the closely related species *P. capsici* and *P. sojae* (Stam *et al.*, 2013). Eight of these domains were found in the C-terminus of *Plasmopara* spp. CRNs, seven of which were already found in the obligate biotroph *P. cubensis*, plus the DFB domain, which was never found in biotrophic oomycetes (Stam *et al.*, 2013). With the exception of domain DBF, encoding a protein kinase, most CRN effector domains defined by Haas *et al.* (2009) are of unknown function. Structural homology searches revealed that SN4 and SN6 singletons are putative ATP-binding proteins. The proteins resulting from the searches presented a variety of functions, from hydrolases to proteasome subunits, so the precise role of CRN effectors with predicted ATP-binding motifs remains to be elucidated. CRNs containing the DN5 domain showed structural homology to serine proteases, and alignment of similar oomycete CRN effectors to serine proteases from diverse organisms ranging from human to bacteria revealed sequence conservation around the predicted protease catalytic residues. Effector proteins showing protease activity have been reported for different plant

Figure 5 A family of CRN effectors with putative serine protease function. (a) Alignment of the C-terminal effector domain (C-terminal to HVLVVVP) of oomycete CRN effectors showing structural similarity to serine proteases. The complete sequence of Phsoj1 lacks the N-terminal CRN motifs, while Phsoj2 appears to have an HVLVVVP-like motif. *Pseudoperonospora cubensis* predicted proteins were incomplete, so it is not known if they code for a CRN protein. (b) Phylogenetic tree built using the alignments shown in (a). Numbers at branches indicate bootstrap support based on 100 replicates. (c) Alignment of oomycete proteins to serine proteases from different species, to which they show structural homology. Asterisks mark protease catalytic residues. Conserved areas around catalytic residues are underlined. Several *Phytophthora infestans* proteins have been removed for the sake of clarity. (d) HMM logo for conservation profile of the Trypsin_2 family of trypsin-like peptidases (PF13365). Underlined regions correspond to underlined sequences in (c). Image taken from http://pfam.sanger.ac.uk/family/PF13365. Position 65 in logo has been removed for clarity. Oomycete abbreviations and shading as in Figure 4. *Plhal*: Plhal003764, *Phinf1*: PITG_09043T0, *Phinf2*: PITG_09051T0, *Phinf3*: PITG_09053T0, *Phinf4*: PITG_18133T0, *Phinf5*: PITG_18140T0, *Phinf6*: PITG_18148T0, *Phinf7*: PITG_21034T0, *Phinf8*: PITG_23182T0, *Phsoj1*: PHYSODRAFT_440636, *Phsoj2*: Physo1_1_137868, *Pscub1*: AHJF01000669.1, *Pscub2*: AHJF01004600.1, *Spla*: Spla serine protease of *Staphylococcus aureus*, *DegS*: *Arabidopsis thaliana* serine protease Deg5, *DegS*: *Escherichia coli* serine protease Deg5, *Rv367c1*: Rv3671c protease from *Mycobacterium tuberculosis*, *Htra1*: human serine protease Htra1, *Splb*: serine protease Splb from *S. aureus*.

pathogens. AvrPi-ta from the fungus *Magnaporthe oryzae* is a metalloprotease (Jia *et al.*, 2000); in bacteria, AvrRpt2 and AvrPphB from *Pseudomonas syringae* and RipP2 (formerly known as PopP2) from *Ralstonia solanacearum* are cysteine proteases (Shao *et al.*, 2002; Axtell *et al.*, 2003; Tasset *et al.*, 2010). Interestingly, like most oomycete CRN proteins (Schornack *et al.*, 2010; van Damme *et al.*, 2012; Stam *et al.*, 2013), RipP2 is also targeted to the nucleus (Deslandes *et al.*, 2003). Although the protease activity of CRN proteins needs to be confirmed, observations here suggest that nuclear targeted proteases could be a tool widely used by plant pathogens to colonize their hosts.

There is evidence suggesting that CRN diversification occurs through recombination between CRN families through the LXLFLAK and HVLVVVP motifs (Haas *et al.*, 2009). In the search for proteins similar to the C-terminal domain of *Plasmopara* spp. CRN effectors, some cases were found of highly similar proteins lacking the N-terminal modules characteristic of CRNs, i.e. missing the LXLFLAK and HVLVVVP motifs (see Fig. S1 for an example). This observation opens up the possibility that CRN proteins with new functions may appear by 'recruitment' of cellular functions by N-terminal CRN modules.

No proteins were found from *H. arabidopsidis* similar to the effector domain of CRN from both *Plasmopara* spp. The same result was obtained with the reciprocal search; searching the PlasmoparaSp database with the 32 predicted CRNs from *H. arabidopsidis* produced only one hit. *Hyaloperonospora arabidopsidis* and both *Plasmopara* spp. are obligate biotrophs belonging to the monophyletic group of the downy mildews. CRN effectors are present in different groups of oomycetes and they have been proposed as being oomycete ancestral effectors that diversified later in evolutionary history (Schornack *et al.*, 2010). It could thus be expected that comparing *Plasmopara* spp. CRN effectors with other oomycetes would result in closer sequences inside the downy mildews than in other species. However, observations reveal that this is not the case, suggesting that drastic diversification of effectors played an important role on specialization inside the downy mildews.

In summary, the search for CRN and RXLR effectors from two obligatory biotrophic oomycetes from the genus *Plasmopara* resulted in the identification of both conserved and species-specific effectors. Although the data set may be incomplete, effectors specific to each species are candidates that may be involved in host specificity. The differences observed between effector sets could also be due to the stage when the infected tissues were harvested: sunflower hypocotyls at early sporulation on cotyledons for *P. halstedii*; infected grapevine leaves before sporulation for *P. viticola*. The finding of CRN and RXLR effectors conserved between oomycete species suggest that they might play important roles in the pathogen biology. The usefulness of effector genes for the identification of resistance genes has been reported (Vleeshouwers *et al.*, 2008; Oh *et al.*, 2009), and several authors have suggested using conserved effectors in the search for durable resistance against oomycetes (Michelmore, 2003; Birch *et al.*, 2008). Thus, the conserved effectors identified in this study could contribute to the identification of potentially durable resistance to these devastating pathogens.

Acknowledgements

The authors thank Julie Poulain (Génoscope, Evry, France) for 454 cDNA sequencing and Sylvain Raffaele (LIPM, Toulouse) for critical reading of the manuscript. They acknowledge Jeanne Tourvieille (INRA Clermont-Ferrand, France) for *P. halstedii* inoculations, Amandine Bordat and Marie-Claude Boniface (LIPM, Toulouse) for help in collecting plant material, and Olivier Catrice (LIPM Toulouse) for estimation of genome sizes of *Plasmopara* spp. We are indebted to Genotoul, Toulouse, for allowing access to informatics facilities. Sequencing funded by INRA AIP Bioressources 2009. Work performed at LIPM was supported by the French Laboratory of Excellence project 'TULIP' (ANR-10-LABX-41; ANR-11-IDEX-0002-02).

References

Ahmed S, Tourvieille de Labrouhe D, Delmotte F, 2012. Emerging virulence arising from hybridisation facilitated by multiple introductions of the sunflower downy mildew pathogen *Plasmopara halstedii*. *Fungal Genetics and Biology* **49**, 847–55.

As-sadi F, Carrere S, Gascuel Q *et al.*, 2011. Transcriptomic analysis of the interaction between *Helianthus annuus* and its obligate parasite *Plasmopara halstedii* shows single nucleotide polymorphisms in CRN sequences. *BMC Genomics* **12**, 498.

Axtell MJ, Chisholm ST, Dahlbeck D, Staskawicz BJ, 2003. Genetic and molecular evidence that the *Pseudomonas syringae* type III effector protein AvrRpt2 is a cysteine protease. *Molecular Microbiology* **49**, 1537–46.

Bailey K, Cevik V, Holton N *et al.*, 2011. Molecular cloning of ATR5 (Emoy2) from *Hyaloperonospora arabidopsidis*, an avirulence determinant that triggers RPP5-mediated defense in *Arabidopsis*. *Molecular Plant–Microbe Interactions* **24**, 827–38.

Baxter L, Tripathy S, Ishaque N *et al.*, 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* **330**, 1549–51.

Birch P, Boevink P, Gilroy E *et al.*, 2008. Oomycete RXLR effectors: delivery, functional redundancy and durable disease resistance. *Current Opinion in Plant Biology* **11**, 373–9.

Bozkurt TO, Schornack S, Banfield MJ, Kamoun S, 2012. Oomycetes, effectors, and all that jazz. *Current Opinion in Plant Biology* **15**, 483–92.

Cabral A, Stassen JH, Seidl MF *et al.*, 2011. Identification of *Hyaloperonospora arabidopsidis* transcript sequences expressed during infection reveals isolate-specific effectors. *PLoS ONE* **6**, e19328.

van Damme M, Bozkurt TO, Cakir C *et al.*, 2012. The Irish potato famine pathogen *Phytophthora infestans* translocates the CRN8 kinase into host plant cells. *PLoS Pathogens* **8**, e1002875.

Delmotte F, Mestre P, Schneider C *et al.*, 2014. Rapid and multiregional adaptation to host partial resistance in a plant pathogenic oomycete: evidence from European populations of *Plasmopara viticola*, the causal agent of grapevine downy mildew. *Infection Genetics and Evolution* **27**, 500–8.

Deslandes L, Olivier J, Peeters N *et al.*, 2003. Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proceedings of the National Academy of Sciences, USA* **100**, 8024–9.

Fabro G, Steinbrenner J, Coates M *et al.*, 2011. Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. *PLoS Pathogens* **7**, e1002348.

Fontaine MC, Austerlitz F, Giraud T *et al.*, 2013. Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *Plasmopara viticola*. *Molecular Ecology* **22**, 2771–86.

Gascuel Q, Martinez Y, Boniface M-C *et al.*, 2015. The sunflower downy mildew pathogen *Plasmopara halstedii*. *Molecular Plant Pathology* **16**, 109–22.

Haas BJ, Kamoun S, Zody MC *et al.*, 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–8.

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B, 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO Journal* **19**, 4004–14.

Jiang RHY, Tyler BM, 2012. Mechanisms and evolution of virulence in oomycetes. *Annual Review of Phytopathology* **50**, 295–318.

Kale S, Gu B, Capelluto D *et al.*, 2010. External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell* **142**, 284–95.

Kemen E, Gardiner A, Schultz-Larsen T *et al.*, 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biology* **9**, e1001094.

Lamour KH, Mudge J, Gobena D *et al.*, 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Molecular Plant–Microbe Interactions* **25**, 1350–60.

Levesque CA, Brouwer H, Cano L *et al.*, 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biology* **11**, R73.

Links MG, Holub E, Jiang RHY *et al.*, 2011. De novo sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics* **12**, 503.

Liu T, Ye W, Ru Y *et al.*, 2011. Two host cytoplasmic effectors are required for pathogenesis of *Phytophthora sojae* by suppression of host defenses. *Plant Physiology* **155**, 490–501.

Mestre P, Piron MC, Merdinoglu D, 2012. Identification of effector genes from the phytopathogenic oomycete *Plasmopara viticola* through the analysis of gene expression in germinated zoospores. *Fungal Biology* **116**, 825–35.

Michelmore RW, 2003. The impact zone: genomics and breeding for durable disease resistance. *Current Opinion in Plant Biology* **6**, 397–404.

Oh S-K, Young C, Lee M *et al.*, 2009. *In planta* expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein Rpi-blb2. *The Plant Cell* **21**, 2928–47.

Peressotti E, Wiedemann-Merdinoglu S, Delmotte F *et al.*, 2010. Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. *BMC Plant Biology* **10**, 147.

Rehmany AP, Gordon A, Rose LE *et al.*, 2005. Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. *The Plant Cell* **17**, 1839–50.

Schornack S, van Damme M, Bozkurt TO *et al.*, 2010. Ancient class of translocated oomycete effectors targets the host nucleus. *Proceedings of the National Academy of Sciences, USA* **107**, 17421–6.

Shao F, Merritt PM, Bao ZQ, Innes RW, Dixon JE, 2002. A *Yersinia* effector and a pseudomonas avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* **109**, 575–88.

Stam R, Jupe J, Howden AJM *et al.*, 2013. Identification and characterisation of CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. *PLoS ONE* **8**, e59517.

Stassen JHM, Seidl MF, Vergeer PWJ *et al.*, 2012. Effector identification in the lettuce downy mildew *Bremia lactucae* by massively parallel transcriptome sequencing. *Molecular Plant Pathology* **13**, 719–31.

Stassen JHM, den Boer E, Vergeer PWJ *et al.*, 2013. Specific *in planta* recognition of two GKLR proteins of the downy mildew *Bremia lactucae* revealed in a large effector screen in lettuce. *Molecular Plant–Microbe Interactions* **26**, 1259–70.

Tasset C, Bernoux M, Jauneau A *et al.*, 2010. Autoacetylation of the *Ralstonia solanacearum* effector PopP2 targets a lysine residue essential for RRS1-R-mediated immunity in *Arabidopsis*. *PLoS Pathogens* **6**, e1001202.

Thines M, Kamoun S, 2010. Oomycete–plant coevolution: recent advances and future prospects. *Current Opinion in Plant Biology* **13**, 427–33.

Tian M, Win J, Savory E *et al.*, 2011. 454 genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. *Molecular Plant–Microbe Interactions* **24**, 543–53.

Torto TA, Li SA, Styer A *et al.*, 2003. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Research* **13**, 1675–85.

Tyler BM, Tripathy S, Zhang X *et al.*, 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261–6.

Viennot-Bourgin G, 1949. *Les Champignons Parasites des Plantes Cultivées*. Paris, France: Masson 597 and Cie, Librairies de l'Académie de Médecine.

Vleeshouwers VGAA, Rietman H, Krenek P *et al.*, 2008. Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. *PLoS ONE* **3**, e2875.

Voglmayr H, Greilhuber J, 1998. Genome size determination in Peronosporales (Oomycota) by Feulgen image analysis. *Fungal Genetics and Biology* **25**, 181–95.

Wang Q, Han C, Ferreira AO *et al.*, 2011. Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. *The Plant Cell* **23**, 2064–86.

Whisson SC, Boevink PC, Moleleki L *et al.*, 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**, 115–9.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Data S1. Supplementary references (Materials and Methods).

Figure S1 Conserved *Plasmopara* spp. CRN C-terminal domains.

Figure S2 Conserved N-terminal domain in oomycete CRN effectors.

Table S1 Sequence databases included in the analysis.

Table S2 Number of raw and cleaned sequences obtained by 454 GS-FLX pyrosequencing of cDNA from *Plasmopara halstedii* (four pathotypes) and *Plasmopara viticola* (two isolates) germinated spores and infected plant tissues.

Table S3 *Plasmopara viticola*+ and *Plasmopara halstedii*+ accessions corresponding to RXLR and CRN effectors.

Table S4 *Plasmopara* spp. RXLR effector families.

Table S5 Comparison of *Plasmopara* spp. RXLR effector families with other oomycetes.

Table S6 Comparison of *Plasmopara* spp. CRN effectors with other oomycetes.

Table S7 Comparison of the C-terminal domain of *Plasmopara* spp. CRN effectors with other oomycetes.