



Comparative analysis of draft genome assemblies developed from whole genome sequences of two *Hyaloperonospora brassicae* isolate samples differing in field virulence on *Brassica napus*

Ming Pei You^a, Javed Akhtar^b, Meenakshi Mittal^b, Martin J. Barbetti^{a,*}, Solomon Maina^{a,c}, Surinder S. Banga^{b,*}

^a School of Agriculture and Environment and the UWA Institute of Agriculture, University of Western Australia, 35 Stirling Highway, Crawley, WA, 6009, Australia.

^b Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, 141004 Punjab, India

^c Australian Grains Genebank, Agriculture Victoria, 110 Natimuk Road, Horsham, Australia

ARTICLE INFO

Keywords:

Downy mildew
Hyaloperonospora brassicae
Genome sequence
Crucifers
Brassicaceae
Rapeseed
Indian mustard

ABSTRACT

Hyaloperonospora brassicae causes downy mildew, a major disease of Brassicaceae species. We sequenced the genomes of two *H. brassicae* isolates of high (Sample B) and low (Sample C) virulence. Sequencing reads were first assembled *de novo* with software's SOAPdenovo2, ABySS V2.1 and Velvet V1.1 and later combined to create meta-assemblies with genome sizes of 72.762 and 76.950Mb and predicted gene densities of 1628 and 1644 /Mb, respectively. We could annotate 12,255 and 13,030 genes with high proportions (91–92%) of complete BUSCOs for Sample B and C, respectively. Comparative analysis revealed conserved and varied molecular machinery underlying the physiological specialisation and infection capabilities. BLAST analysis against PHI gene database suggested a relatively higher loss of genes for virulence and pathogenicity in Sample C compared to Sample B, reflecting pathogen evolution through differential rates of mutation and selection. These studies will enable identification and monitoring of *H. brassicae* virulence factors prevailing in-field.

1. Introduction

Downy mildew (*Hyaloperonospora brassicae*) is a major foliar disease on oilseed and other Brassicaceae species [1,2]. This disease is endemic across Brassica-growing regions worldwide [3]. Severe outbreaks of downy mildew occur in India on mustard (*B. juncea*) [4] and across southern Australia in oilseed rape (*B. napus*), with worst-affected crops showing up to 55% leaves diseased, 15% leaf area lost to lesions and 13% leaf area collapsed [5,6]. Estimated annual losses up to AUS\$13 million were reported in Western Australia alone [5,6]. Across southern Australia, downy mildew has become increasingly severe on seedlings and young plants over the past 15 years, likely fostered by increasing temperatures consequent from climate change [5,7]. The definitive taxonomy of downy mildews (*Peronosporaceae*) remains challenging. Initially, taxonomic classification for *Peronospora* was based on conidial measurements and specialisation on host species [8,9]. While Constantinescu and Fatehi [10] used morphological and molecular studies to divide *Peronospora* into *Peronospora* s. str., *Hyaloperonospora* and *Perofascia*, multi-molecular studies by Choi et al. and Göker et al.

[11–13] and Voglmayer [14] highlighted genetic differences within *H. parasitica* and suggested this genus may be further divided. Subsequently, Göker et al. [15] reclassified the downy mildew pathogen of brassicas as *H. brassicae* (syn. *Hyaloperonospora parasitica*). *H. brassicae* shows a high degree of physiological specialisation, including distinct physiologic races or pathotypes, for example, Natti et al. [16] and Coelho et al. [2] for isolates from *B. oleracea*; Nashaat and Rawlinson [17] for isolates on *B. napus* ssp. *oleifera*; and Nashaat et al. [4,18] for isolates on *B. napus*. More recently, Mohammed et al. [19] identified appropriate host differentials and characterised eight distinct pathotypes of *H. brassicae* and defined the phylogenetic relationships for *H. brassicae* in Australia. Pathotypes and phylogenetic variation not only determine the severity of downy mildew epidemics in *Brassica* spp. [19], but are also crucial to the identification of appropriate host resistances and their effective deployment [19]. However, importantly, Mohammed et al. [6,20] highlighted high levels of pathotype-independent resistance sources across diverse Brassicaceae to *H. brassicae*, particularly within *R. sativus*, *B. carinata*, *B. juncea*, *B. oleracea* and *Crambe abyssinica*. Increasing the level of virulence (i.e., disease-producing power) of

* Corresponding authors.

E-mail addresses: martin.barbetti@uwa.edu.au (M.J. Barbetti), nppbg@pau.edu (S.S. Banga).

<https://doi.org/10.1016/j.btr.2021.e00653>

Received 21 January 2021; Received in revised form 28 May 2021; Accepted 16 June 2021

Available online 22 June 2021

2215-017X/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

H. brassicae isolates may challenge such resistance sources [19]. Mohammed et al. [19] were the first to highlight the variation in levels of virulence in *H. brassicae* isolates in Australia. They also showed that while isolates were most virulent on their *Brassica* species of origin, they were still virulent on other related *Brassica* species, as shown earlier by Chang et al. [21], McMeekin [22], Dickinson and Greenhalgh [23], Sherriff and Lucas [24], Silue et al. [1] and Vicente et al. [25]. Mohammed et al. [19] also showed that this increased virulence was associated with a corresponding increased severity of downy mildew disease of rapeseed across southern Australia. In addition, there is a historical molecular evidence confirming the existence of genetic differences within *H. brassicae* (e.g., Choi et al. [11]; Göker et al. [13]; Voglmayr [14]). *H. brassicae* isolates can be either homothallic or heterothallic [26] and outcrossing of homothallic isolates involving recombination of virulence attributes is probably important for increasing variation for virulence [27]. While there have been investigations of other downy mildews involving linkage mapping, genome-wide association mapping, genome sequencing or enrichment sequencing [28], this is not the case in relation to virulence of *H. brassicae*.

This paper reports *de novo* whole-genome assemblies and analysis of two contrasting Australian *H. brassicae* samples, of high versus low virulence. We also show differences between the biological processes (biosynthetic and catabolic process, cell communication, cellular component organization and localization) and cellular components (catalytic complex, coated membrane, endomembrane system, endoplasmic reticulum membrane, and envelope) between lower and more virulent isolates. To the best of our knowledge, this is the first report of a genome assembly for this organism.

2. Materials and methods

2.1. Pathogen isolates

Two separate collections of *Hyaloperonospora brassicae* isolates, named Samples B and C, were used for the present studies. **Sample B** comprised 5 isolates of *H. brassicae* from Victoria, Australia, from commercial rapeseed fields that showed high levels of leaf infection (ranging from 10 to 45% and with a mean % disease incidence of 29.5 (i.e., situations showing high virulence of isolates). **Sample C** comprised 6 isolates of *H. brassicae* of which 5 were from New South Wales, Australia, and one from Western Australia. These samples were collected from commercial rapeseed fields with lower levels of leaf infection (0 to 20% and with a mean % disease incidence of 8.2) compared to the Sample B (i.e., situations showing a much lower virulence of isolates) (Table 1). These isolates were obtained as part of an Australia-wide survey of foliar diseases of rapeseed in 2018, where 80 isolates of *H. brassicae* were obtained. Samples B and C were selected to best represent a range of high and low virulence. As virulence level was the sole criterion for selection of the two *H. brassicae* samples, isolates used were not pathotyped.

2.2. DNA extraction and PCR conditions

DNA was extracted from 80 isolates of *H. brassicae* (collected in 2018 from 39 locations across southern Australia), using the procedure of Cenis [29]. This same procedure was used previously by Mohammed et al. [19] for *H. brassicae*. Briefly, mycelia from leaf samples were harvested directly from infested leaves using tweezers under a dissecting microscope. Mycelia were then macerated using Precellys Evolution Homogeniser (Bertin TECHNOLOGIES) in 300 µl of extraction buffer [200 mM Tris-HCl (pH 8.5), 250 mM NaCl, 25 mM EDTA], 0.5% (wt/vol) sodium dodecyl sulphate (SDS)]. Then, 150 µL of 3 M sodium acetate (pH 5.2) was added and mixed well by pipetting gently and maintained at 20°C for 10 min, centrifuging at 13,000 rpm for 10 min. Subsequently, the supernatant was transferred to a fresh tube before

Table 1

Details of *Hyaloperonospora brassicae* isolates and their Samples B [with high virulence, (B)] and C [with low virulence, (C)]. (% leaf infection range) as used in these studies.

Field survey site number	Australian state	Isolate Sample (B or C) and % leaf infection range	GPS location	
			S (latitude)	E (longitude)
167	Victoria	B (35-40%)	-36.59325	144.7731333
167	Victoria	B(35-40%)	-36.59325	144.7731333
168	Victoria	B(10-45%)	-36.59056667	144.70565
172	Victoria	B(20-40%)	-36.10483333	145.20465
177	Victoria	B(10-20%)	-35.92498333	145.3684
186	New South Wales	C(5-10%)	-35.16766667	146.9054333
193	New South Wales	C(10-20%)	-35.5147	146.7008667
201	New South Wales	C (0-20%)	-36.06776667	146.1856333
218	New South Wales	C(0-15%)	-36.1562	145.9855833
230	New South Wales	C(4-5%)	-34.96745	147.5596833
239	Western Australia	C(0-10%)	-31.166	116.461

adding an equal volume of cold isopropanol (450 µL) and held at room temperature for at least 15min and then centrifuged for 13,000 rpm for 10 min and the supernatant was discarded. Precipitated DNA pellet was washed with 70% (vol:vol, ethanol:water), air dried, then resuspended in 30–50 µl of TE buffer. Quantity and the quality of the extracted DNA was determined using a NanoDrop 1000 Spectrophotometer (Thermo Scientific). DNA and then stored at 4°C. DNA was subjected to PCR using a master mix of a total volume of (50 µL) that contained 0.2 µM of each primer (primers ITS1-O (5'-CGG AAG GAT CAT TAC CAC-3' [30]) and ITS4-H (5'-TCC TCC GCT TAT TAA TAT GC-3' [12], a modification of ITS4.ITS1-O was chosen due to its specificity that resolves the need for additional amplification of host ITS rDNA [15, 30]. PCR assays were undertaken as follows: Initial denaturation 94°C for 2 min, followed by 35 cycles at 94°C for 1 min, with annealing gradient temperature set at range 50°C–60°C (in this study annealing temperature of 53°C was selected) for 1 min and extension at 72°C for 2 min; followed by a final extension step at 72°C for 7 min and then held at 4°C. PCR products were subjected to agarose gel electrophoresis at 60 mV for 120 min on a 1% (w/v) agarose gel containing 0.1% GelRed™Biotium Inc. (United States) and then visualized under UV light. The amplified PCR products of 80 isolates were subjected to Sanger genome sequencing and library preparation by Macrogen Inc. (Korea) and sequences analysed using Geneious. The final sequences were subjected to NCBI BLAST to confirm the species and percentage nucleotide relationships. DNA from highly similar sequences was selected as a representative of geographical regions of Victoria (5 isolates), hereafter named **Sample B**, and **Sample C** [New South Wales (5 isolates) plus one from Western Australia]. Samples B and C were individually pooled together in each tube (sample) separately for library preparation. Shotgun genome sequencing and library preparation were performed by Macrogen Inc. (Korea).

2.3. De novo genome assemblies

Our bioinformatics approach involved constructing separate but multiple *de novo* genome assemblies, three each for the two samples (Samples B and C) by using different algorithms. Two sets of *de novo* assemblies were then merged separately, using Metassembler reconciliation algorithm to produce one best assembly for each sample.

We used 100 bp paired-end reads with about 100x genome coverage for developing genome assemblies. The raw reads of each sample were first error-corrected with software Trimmomatic and then quality

checked with FASTQC program. Software KmerGenieV1.7 [31] was next used to select the most optimum k-mer length for constructing the genome assembly. Subsequently, we used three softwares, SOAPdenovo2 [32, 33], ABySSV2.1 [34] and Velvet V1.1 [35], for developing three *de novo* assemblies for each of the two samples. Software settings used for these algorithms were SOAPdenovo2 (all -s -K 87 -R -F YES), ABySS (k=87), Velvet (k-mer 87 -ins_length 300 -exp_cov 50). Software GapFillerV1.1 (m 20 -o 2 -r 0.7 -n 10 -d 50 -t 10 -T1 -I 1) was used to improve the assemblies and scaffolding was performed using SSPACE V3.0 (-x 0 -m 20 -o 10 -k 3 -a 0.70) [36]. We next used the software tool Metassembler [37] with default settings to merge three sets of *de novo* assemblies into two meta-assemblies. The development of meta-assemblies was facilitated through pairwise progressive merger, using cross species mate paired libraries for each Sample (A or B) of *H. brassicae* isolates. Mate-pair libraries were generated *in silico* by using software Cross-mates as implemented in Cross-species Scaffolding [38]. Genome assembly of *Hyaloperonospora arabidopsidis* (protists.ensembl.org/Hyaloperonospora_arabidopsidis/Info/Annotation/#assembly; GCA_000173235.2), hereafter Sample A, was used as a reference [39]. The completeness of the meta-assemblies, one each for Sample A and B, was assessed with software BUSCO [40] using Alveolata dataset of BUSCOs.

2.4. Gene prediction

We predicted protein-coding genes of *H. brassicae* isolates using GeneMarkS v2 with default settings and revised gene models ("Hyaar1_GeneModels_FilteredModels3_nt.fasta") available at JGI website (<https://phyocosm.jgi.doe.gov/Hyaar1/Hyaar1.home.html>). Pseudogenes were detected from genome sequences (meta-assembly scaffolds), both at exon and gene levels. Predicted genes compared favourably with the gene finders that employed supervised annotation in the reference set of *H. arabidopsidis*. We also conducted a whole-genome BLAST analysis against the pathogen-host interaction (PHI) gene database [41] to identify potential virulence-associated genes. For this, a set of all PHI-base accessions with associated phenotypes and protein sequence, was downloaded (<http://www.phi-base.org>) and gene sets from Samples B and C were used to run a BLASTX search against PHI-base sequence set with e-value <10⁻³. The distribution of BLAST2GO annotations of putative genes from *H. arabidopsidis* and *H. brassicae* were also computed [42].

2.5. Functional genome analysis

The assembled genome sequences were functionally annotated using the OmicsBox pipeline. This involved querying of predicted protein-coding sequences against Oomycete Non-redundant protein database from NCBI using BLASTX with default settings (e-10⁻³). This analysis assigned a protein ID/accession number to each positive hit. The resulting sequences were run through a mapping process that used assigned accessions to retrieve gene names, then we sought their specific functions in the GO database. The final step in this pipeline was to run annotation analysis. This step helped assign the most specific GO terms to each sequence with respect to their cellular location, role in biological processes, and their molecular functions. The sequences with assigned GO terms were subsequently used for further analysis. Our focus was on the genomic variation between two groups of isolates in terms of their virulence (i.e., Samples B and C). Therefore, sequences showing similarity to various virulence-related genes were further analysed.

3. Results

3.1. Illumina sequencing

Produced data from Sample B were 78,607,680 reads and total read bases 7.9G bp and, for Sample C, reads were 91,282,136 with total read

bases of 9.2G bp. GC content and Q30 score for Sample B were 47.98 percent and 92.4 percent, and for Sample C, it was 47.75 percent and 92.7 percent, respectively. The raw reads were first curated by discarding low-quality reads and removing adaptor sequences using the software Trimmomatic. Total clean reads for Samples B and C were 76,082,958 and 88,547,096, respectively. These were used for further analysis (Table 2).

3.2. De novo and meta-assemblies of *H. brassicae*

We used three state-of-the-art de Bruijn graph assemblers, SOAPdenovo2, ABySS, and Velvet, to develop three sets of the genome assemblies that produced varied numbers and sizes of scaffolds. Scaffolds were constructed from contigs produced by Velvet for both samples, using softwares SSAPCE and GapFiller. This software facilitated filling of 1,505 (14.11%) gaps out of 10,665 in Sample B and 1,310 (11.60%) out of 11,395 in Sample C. SOAdenovo2 produced 447,236 and 425,174 scaffolds for Samples B and C, respectively, at k-mer 87. Software KmerGenieV1.7 was used to select 87 as the best k-mer length for *de novo* genome assembly (S Fig. 1). The longest scaffolds had 47,638bp for Sample B and 58,419bp for Sample C. ABySS assembled Samples B and C had scaffolds producing 198,569 and 194,799, respectively. The N50 value of assemblies were 8,505 bp for Sample B and 8,796bp for sample C. The assembler ABySS constructed longest scaffolds. These were 136,672bp for Sample B and 197,135bp for Sample C. Velvet assembled 296,308 scaffolds for Sample B and 218,453 for Sample C. The longest scaffold in the assembly of Sample B was 58,632 bp and of Sample C was 59,541bp (Table 3). We meta-assembled three *de novo* assemblies separately for each sample, as described earlier. This step dramatically improved the assemblies, by detection and integration of poorly sequenced regions. Sample B had 6,438 scaffolds with an N50 of 23,533bp, the largest scaffolds being 154,210bp (Table 3). Sample C had 6,470 scaffolds with an N50 of 24,471bp, the longest scaffolds being 197,135bp in length. Total length of Sample B was 72,162,632bp compared to 76,950,397bp for Sample C. The average scaffold length of Sample B was 11,209bp and Sample C was 11,893bp. The criteria used to select the best assemblies were total genome size expected for genome, lower number of scaffolds, larger N50 value and fewer non-ATGC characters. The N50 value of both samples was approximately 24 Kb, showing a reasonably good quality assembly for downstream analysis. Genome assemblies are available on GenBank as BioProject PRJNA540582 with accession numbers SAMN11547018 and SAMN11547017. Assembly parameters were also assessed with Quast [43] to determine contig sizes, N50, and assembly size (Supplementary Table 1). The reliability of the assemblies and genome annotation was also obvious from high proportion of the Benchmarking Universal Single-Copy Orthologs's (BUSCO) found in the predicted gene models, including 91%-92% as complete BUSCO genes (Fig.1).

3.3. Annotation of protein-coding genes

We annotated the meta-assemblies of two Australian *H. brassicae* samples for protein-coding genes. We used *Ab initio* gene prediction after protein homology with the trained optimal parameters with GeneMarkS. This allowed prediction of 36,819 genes in Sample B and 40,346 genes in Sample C (Table 4), on the basis of the reference genome of *H. arabidopsidis*. Gene densities were 1425/Mb and 1435/Mb for Samples B and C, respectively. For Sample B, 24,055 genes were blasted with Oomycete Non-redundant Protein Database (NCBI); out of which, 16,996 were mapped, 12,255 were annotated. For Sample C, 26,362 genes were blasted, out of a total 40,346 predicted genes with Oomycete Non-redundant Protein Database. Of these, 18,570 genes were mapped, of which 13,030 and 5,540 genes revealed significant and non-significant annotations, respectively (Table 5). We also plotted Venn Diagram comparing *H. arabidopsidis* genes found in the genome assemblies of *H. brassicae* Samples B (with high virulence) and C (with low

Table 2

Genomic reads stats of *Hyaloperonospora brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

Sample ID	Total reads bases (bp)	Total raw reads	GC (%)	AT (%)	Q20 (%)	Q30 (%)	Clean reads
Sample B	7,939,375,680	78,607,680	47.98	52.02	95.44	92.4	76,082,958
Sample C	9,219,495,736	91,282,136	47.75	52.25	95.63	92.7	88,547,096

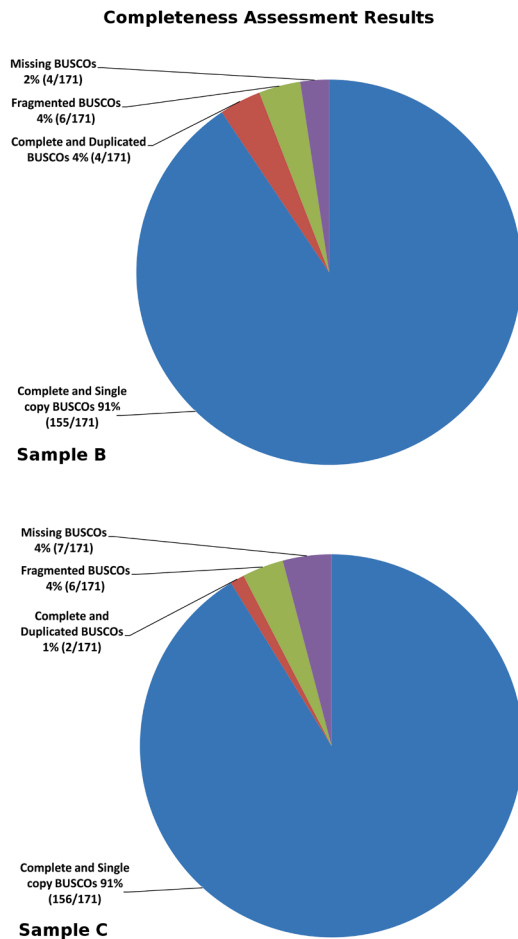


Fig. 1. Assessment of genome assembly and annotation completeness with software BUSCO for *Hyaloperonospora brassicae* Samples B [with high virulence] and *H. brassicae* Sample C [with low virulence].

Table 3

Final whole genome assembly features of *Hyaloperonospora brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)], by Metassembler.

Assembly parameters	Sample B	Sample C
Total genome size (bp)	72,162,632	76,950,397
Number of scaffold	6,438	6,470
Minimum scaffold length (bp)	500	500
Maximum scaffold length (bp)	154,210	197,135
Average scaffold length (bp)	11,209	11,893
Total number of non-ATGC characters	779,011	664,485
N50 value (bp)	23,533	24,471
GC (%)	47.33	47.42

virulence) (Fig. 2). Almost 33.0 percent of the genes were common across all three assemblies and 16.7 percent of *H. arabidopsidis* genes were not found in *H. brassicae* assemblies.

Whole genome BLAST analysis conducted against the pathogen-host interaction (PHI) gene database facilitated identification of 6,969 (Sample B) and 7,116 (Sample C) genes putatively homologous to genes

Table 4

Details of protein-coding genes in *Hyaloperonospora brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

Protein-coding genes details:	Sample B	Sample C
Total size (bp)	22,614,046	24,541,500
Total protein-coding genes(≥250bp)	36,819	40,346
Gene density (number of genes per Mb)	1,628	1,644
Annotation details:		
Blasted coding genes	24,055	26,362
Mapped of blasted coding genes	16,996	18,570
Annotated of mapped coding genes	12,255	13,030

Table 5

Classes of genes identified in *Hyaloperonospora brassicae* Sample B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)] as compared with *Hyaloperonospora arabidopsidis*.

Gene Category/Sample ID	<i>H. arabidopsidis</i>	Sample B	Sample C
RXLr effectors	94	382	490
Elicitors	12	22	19
Necrosis-inducing proteins	4	6	6
Pectate lyase	9	19	17
Pectinesterases	7	9	8
Glucanases/Cellulases	30	61	53
CRN proteins	ND	2	2
Cutinases	2	ND	ND
Serine-protease inhibitors	2	5	7
Short cysteine-rich proteins (SCR)	5	6	8
Peptidases/Metallopeptidases	105	171	185
Nudix domain containing protein	8	7	6
Other Pathogenic genes	38	27	29
Defence responsive genes	8	7	6
Avirulence (Avh) protein	6	5	4

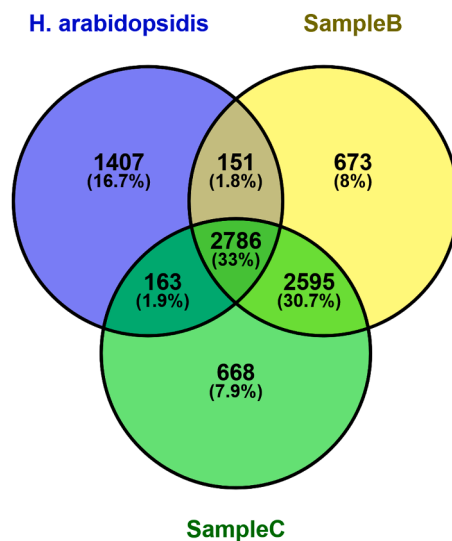


Fig. 2. Venn Diagram comparing *Hyaloperonospora arabidopsidis* genes found in the genome assemblies of *H. brassicae* Sample B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)]. Almost 33 percent of the genes were common in all three assemblies. 16.7% of *H. arabidopsidis* were not found in *H. brassicae* assemblies.

in the database. Best fits with five categories were detected. More than half of the genes, in both isolate Samples, were represented by “virulence” (i.e., disease-producing power) category followed by “pathogenicity” (i.e., ability to be pathogenic) (Fig. 3). Although most identified PHI tags were common among the two isolates, 190 and 271 tags were unique to Samples B and C, respectively. Sample C showed a marginally higher number of genes with reduced virulence compared to Sample B. Number of effectors or genes for avirulence was also higher in Sample C. Virulence-associated genes were further classified based on their specific function in the host-pathogen interaction and virulence/pathogenicity (Fig. 4). As expected, Samples B and C shared more common genes (95) among themselves compared to only 83 with *H. arabidopsidis*.

3.4. Genomic Ontology (GO) annotations of Australian *H. brassicae* sequences

BLAST alignment of the assembled sequences conducted against the oomycete database in NCBI highlighted almost 65% sequences with positive BLAST hits in both Samples, and about 70% of those sequences were mapped but only 50% of positive BLAST hits could be linked to any Gene Ontology term after annotation. The remaining sequences could not be assigned to any GO term (Table 5). The sequences belonging to “virulence/pathogenicity” GO entries were extracted and classified on the basis of their specific function in the host-pathogen interaction and virulence/pathogenicity. The predicted gene models were analysed to identify gene functions. From the query sets of more than 37,000 genes, 12-13,000 genes were successfully annotated. GO terms for the annotated genes were placed into three basic categories: biological processes (BP), molecular function (MF), and cellular components (CC). Both *H. Brassicae* Samples (B and C) showed almost similar molecular functions (MF). However, our samples were distinctively different from the reference, *H. arabidopsidis* (classified as Sample A) in terms of carbohydrate derivative binding, catalytic activity on a protein, drug binding, heterocyclic compound binding, hydrolase activity, ion binding, organic cyclic compound binding, small molecule binding and transferase activity (Fig. 5). The two samples also differed distinctively for biological processes and cellular components. In BP, biosynthetic process, cellular metabolic process, nitrogen compound metabolic process, and primary metabolic processes were distinctively higher in less virulent Sample C with lower virulence/pathogenicity. Differences between Samples B and C were greater than between Samples C and A (Fig. 6). In CC, most of the GO-categories were represented equally among Samples B and C, but a few GO-classes such as host cellular components and intracellular anatomical structure were distinctively higher in Sample C. Although both the isolate samples did not differ significantly for most their cellular components (Fig. 7) or molecular functions (Fig. 5), that they did differ significantly for important biological processes and cellular

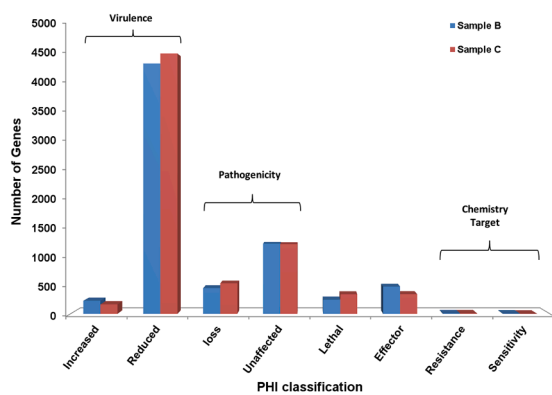


Fig. 3. Distribution of the Pathogen-Host Interaction (PHI) genes *Hyaloperonospora arabisidopsidis* in the *Hyaloperonospora brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

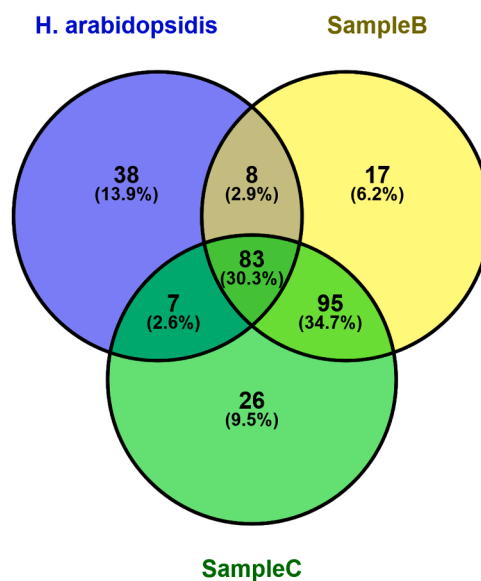


Fig. 4. Venn Diagram showing overlap of virulence associated gene sets of *Hyaloperonospora arabisidopsidis* with those of *H. brassicae* Sample B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

components was a key finding. This finding opens the way to identify *H. brassicae* pathotypes prevailing in the field and to detect early any virulence changes within *H. brassicae* populations.

4. Discussion

We believe that these studies are not only the first genome assemblies of *H. brassicae*, but, more importantly, the first comparing a virulent versus a less virulent sample. Together, they provide new understanding about the *H. brassicae* genome, its metabolic pathways and genes encoding proteins of relevance for virulence. While the two pathogen samples were similar for molecular functions, both biological processes (biosynthetic and catabolic process, cell communication, and cellular component organization and localization) and cellular components (catalytic complex, coated membrane, endomembrane system, endoplasmic reticulum membrane, and envelope) were distinctively different between the more virulent Sample B than the less virulent Sample C. Importantly, the genome sequences presented in the current study constitute an important resource for future functional genomic analyses, fostering DNA barcoding of *H. brassicae* isolates and the development of genetic markers (e.g. SNPs and short repeats) that could be used to assess the population diversity and to tag genes for virulence in the pathogen population by comparing highly virulent and non-virulent isolates.

Oomycetes like *Phytophthora* spp. and *Hyaloperonospora* are causal organisms for many serious plant diseases, often with catastrophic consequences for crop productivity. Genome sequences of many Oomycetes are now available (e.g., Tyler et al. [44]; Hane et al. [45]; David et al. [46]). Tools for comparative analyses of Oomycete genomes are also now available to identify conserved and biologically important proteins (e.g., Rujirawat et al. [47]). These have provided new insights into the biology and evolution of virulence/pathogenicity. In this study, we present two genome assemblies for two samples of *H. brassicae* isolated from commercial rapeseed fields in Australia. Using SOAPdenovo2, ABySS V2.1, and Velvet V1.1 to perform *de novo* assembly and create a meta-genome assembly, using software Metassembler. We generated 6,438 scaffolds for Sample B with an N50 of 23,533 with the longest scaffold being 154,210 bp. In contrast, 6,470 scaffolds were produced for Sample C with an N50 of 24,471 bp and highest scaffold length 197,135 bp. Metassembler has been widely used to develop meta-genome assemblies [48–55]. The genome size of Sample C was

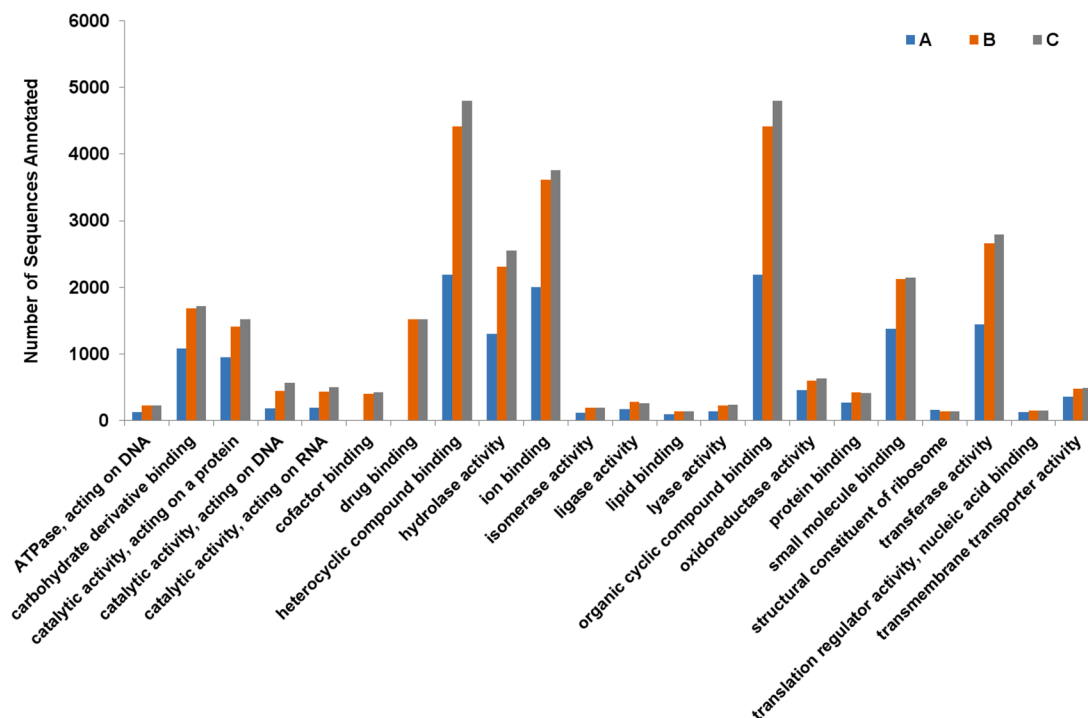


Fig. 5. Distribution of BLAST2GO annotations of putative genes from *Hyaloperonospora arabisidopsis* and *H. brassicae*. The graph shows level 3 annotations for top 20 molecular functions for *H. arabisidopsis* (A) and *H. brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

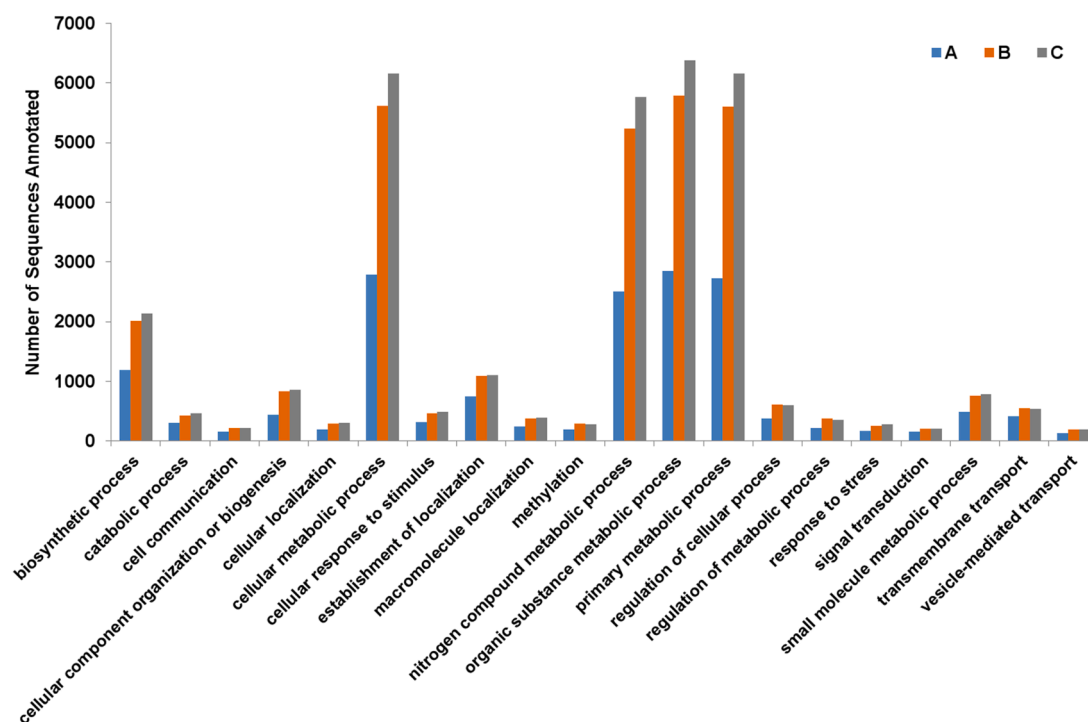


Fig. 6. Distribution of BLAST2GO annotations of putative genes from *Hyaloperonospora arabisidopsis* and *H. brassicae*. The graph shows level 3 annotations for top 20 biological processes for *H. arabisidopsis* (A) and *H. brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

76.95Mb compared to 72.16Mb for Sample B. The N50 value of both samples approximated 24 Kb, indicating good quality for further downstream analysis. Using reference sequences for *H. arabisidopsis*, we predicted 36,819 and 40,346 protein-coding genes, for Samples B and C, respectively, including those associated with plant-pathogen interactions. Of the predicted genes only 12,255 and 13,030 could be

mapped and annotated for Samples B and C, respectively. We achieved good assembly continuity using short-read data and annotated over 91%-92% of BUSCO Alveolata dataset set of single-copy orthologs, with assemblies relatively contiguous and with a negligible number of duplicated BUSCO genes. Comparative analysis of the genomic data obtained from two *H. brassicae* isolates revealed conserved and varied

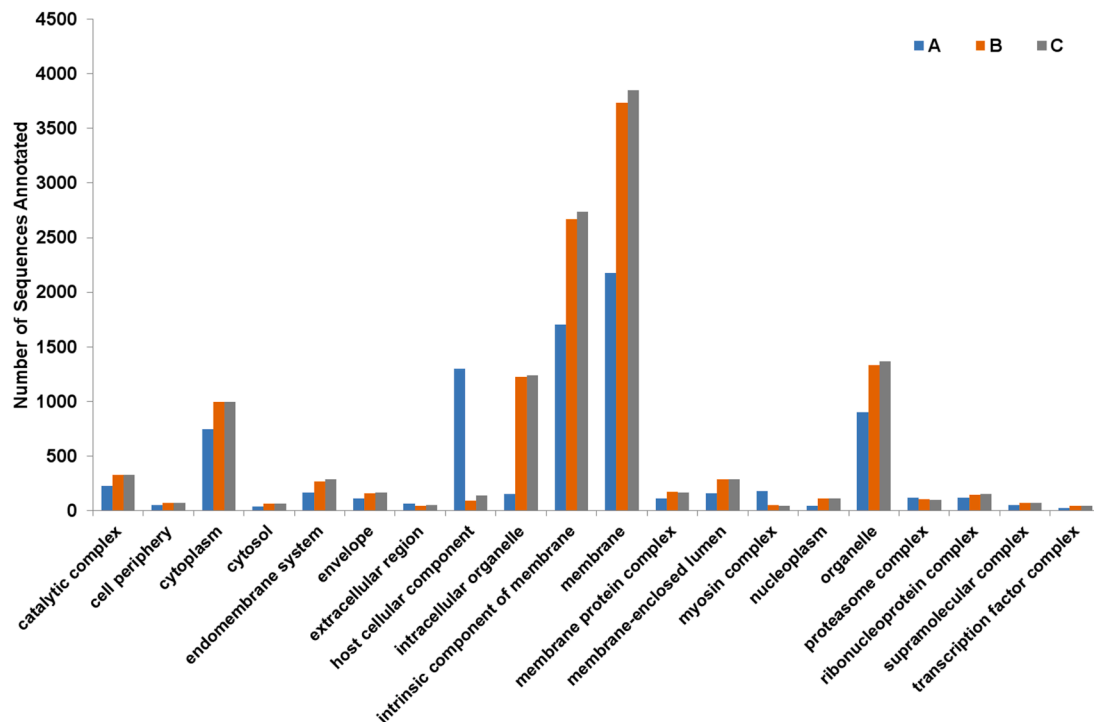


Fig. 7. Distribution of BLAST2GO annotations of putative genes from *Hyaloperonospora arabidopsidis* and *H. brassicae*. The graph shows level 3 annotations for top 20 cellular components for *H. arabidopsidis* (A) and *H. brassicae* Samples B [with high virulence, (B)] and *H. brassicae* Sample C [with low virulence, (C)].

molecular machinery, underlying their physiological specialisation and infection capabilities. Several gene classes were identified. These included necrosis-inducing proteins (NLPs) produced by fungal pathogens that cause cell death while stimulating the plant's immune reaction (e.g., Feng et al. [56]). Oome et al. [57] showed that “downy mildew pathogen *H. arabidopsidis* encodes 10 noncytotoxic NLPs (HaNLPs) that do not cause necrosis” that “act as potent activators of the plant immune system in *Arabidopsis thaliana*.” Many gene classes we identified differed significantly between the model pathogen *H. arabidopsidis* and *H. brassicae*, in both identified class and numbers within each class. Of particular interest were the RXLR effectors that are primarily associated with the suppression of host immunity and play a major role in virulence for downy mildew species [58]. The predicted RXLR effectors and elicitors were more in the *H. brassicae* samples B and C compared with *H. arabidopsidis*. Zhang et al. [59] also reported RXLR effectors in *P. sojae* isolates, and of these, 42 core RXLR effectors were considered important for infection. In our study, the number of elicitors in both *H. brassicae* Samples B and C was higher than in *H. arabidopsidis*. Elicitors are pathogen signal metabolites recognized by plant cells, and most constitute Pathogen-Associated Molecular Patterns (PAMPs) (e.g., Fawke et al. [60]).

Plant pathogens must overcome host resistance to successfully infect plants [61] and, toward this outcome, fungal pathogens produce pectin lyases (PL) to catalyse the depolymerisation of esterified pectin by a β -elimination mechanism [62]. The contribution of PL to infection by *H. brassicae* remains largely unknown, similar to other oomycete pathogens, despite PL having been extensively studied in major fungal pathogens (e.g., Fu et al. [63]). Pectinesterases are involved in plant tissue maceration and soft-rotting, and a few cellulases (glucanases/cellulases) of plant-pathogenic fungi have been shown to be involved in pathogenicity (e.g., Eshel et al. [64]). A further example is the fungal pathogen *Alternaria alternata* that produces an endocellulase that is an important factor in disease development in persimmon fruit [64]. *H. arabidopsidis* secretes proteins and fibrillar β -1,3-glucans that binds its germ tubes to the substratum [65], likely helping resist detachment by wind or rain and to help protect against desiccation [66].

Plant pathogens rely on virulence factors (effectors) to modulate host immunity and ensure progressive infection. The CRNs (i.e., CRinkling and Necrosis) are among the highly conserved protein families among secreted proteins. CRN proteins were first identified in the plant pathogenic Oomycetes where these are secreted and translocated inside host cells through a conserved N-terminal domain. Study of CRN functions may allow a better understanding of CRN effector biology and the processes contributing to host susceptibility and immunity [67]. Further, CRN effectors exhibit various pathogenic functions, including induction of Programmed Cell Death (PCD) and suppression of PCD via PAMP-triggered immunity or/and effector-triggered immunity [68]. Additionally, peptidases/metallopeptidases, enzymes that cleave peptide bonds, yielding proteins and peptides [69], also regulate/activate or inactivate target substrates via proteolysis [70]. Peptidases play an important role in plant-pathogen interactions, including nutrient acquisition and catabolic activities and can influence the outcome of plant-pathogen interactions. The first examples of avirulence proteins were from oomycete species such as *Melampsora lini*, *Phytophthora sojae*, *P. infestans*, and *H. arabidopsidis* [71]. Avirulence genes in pathogens directly or indirectly encode molecules that are recognized by plant receptors encoded by major resistance genes (R genes) triggering plant immunity. Avirulence proteins (Avh) are effectors involved in pathogenicity [72].

5. Conclusions

The genome assemblies presented in this study constitute critical genetic resources that can spur further research into the mechanisms of virulence caused by *H. brassicae*, a destructive but less understood pathogen of many crop species. Our genome sequences, while differing significantly to those of *H. arabidopsidis*, did show a large degree of consistency with those of *H. arabidopsidis*, at least to the extent these find use in population genomic studies for the identification of virulence haplotypes and alleles. Comparative analysis of annotated protein-coding genes obtained from two *H. brassicae* isolates revealed conserved and varied molecular machinery underlying the physiological specialisation

and infection capabilities of this important pathogen. In particular, *H. brassicae* Samples A and B differed in terms of biological processes for their cellular components. Our studies will pave the way for early detection of any virulence changes within the pathogen populations, and perhaps even for molecular identification of pathotypes of this pathogen. Such information is crucial for *H. brassicae* as there is not only virulence/pathotype diversity among its populations, but that such diversity is rapidly evolving [19]. Further, genome sequencing or enrichment analysis may also be used to identify alleles associated with particular virulence/avirulence effects. However, while linkage mapping or genome-wide association mapping can enable more rapid generation of cultivars resistant to *H. brassicae*, more virulent newly evolving downy mildew genotypes will keep emerging [28]. Hence, developing markers for virulence/pathotype classification would allow rapid molecular identification and monitoring of the prevalent virulence potentials existing within pathogen populations in-field [28].

Credit authorship contribution statement

Ming Pei You: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review and editing (e.g., designed studies, carried out the DNA isolation and purification and jointly wrote the paper). **Javed Akhtar:** Methodology, Investigation, Data curation, Formal analysis (e.g., undertook genome assemblies and annotation and helped with their interpretation). **Meenakshi Mittal:** Methodology, Investigation, Data curation, Formal analysis (e.g., undertook genome assemblies and annotation and helped with their interpretation). **Martin J. Barbetti:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review and editing (e.g., designed studies and jointly wrote the paper). **Solomon Maina:** Investigation, Conceptualization and Methodology as relates to initial DNA isolation and purification. **Surinder S. Banga:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review and editing (e.g., undertook genome assemblies and annotation and their interpretation and jointly wrote the paper).

Declaration of Competing Interest

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

Data availability statement

Genome assemblies have been deposited in GenBank via BioProject SUB5539449.

CRedit authorship contribution statement

Ming Pei You: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Javed Akhtar:** Methodology, Investigation, Data curation, Formal analysis. **Meenakshi Mittal:** Methodology, Investigation, Data curation, Formal analysis. **Martin J. Barbetti:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Solomon Maina:** Investigation, Conceptualization, Methodology. **Surinder S. Banga:** Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing.

Acknowledgements

This work was supported equally by the University of Western Australia, Australia, and Punjab Agricultural University, India. Thanks, are also due to anonymous reviewers for the helpful suggestions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.btre.2021.e00653.

References

- [1] D. Silué, N.I. Nashaat, Y. Tirilly, Differential responses of *Brassica oleracea* and *B. rapa* accessions to seven isolates of *Peronospora parasitica* at the cotyledon stage, *Plant Dis* 80 (1996) 142–144.
- [2] P.S. Coelho, J.G. Vicente, A.A. Monteiro, E.B. Holub, Pathotypic diversity of *Hyaloperonospora brassicae* collected from *Brassica oleracea*, *Eur. J. Plant Pathol.* 134 (2012) 763–771.
- [3] P. Kaur, K. Sivasithamparam, H. Li, M.J. Barbetti, Pre-inoculation with *Hyaloperonospora parasitica* reduces incubation time and increases severity of disease caused by *Albugo candida* in a *Brassica juncea* variety resistant to downy mildew, *J. Gen. Plant Pathol.* 77 (2011) 101–106.
- [4] N.I. Nashaat, A. Heran, R.P. Awasthi, S.J. Kolte, Differential response and genes for resistance to *Peronospora parasitica* (downy mildew) in *Brassica juncea* (mustard), *Plant Breed* 123 (2004) 512–515.
- [5] A.E. Mohammed, M.P. You, M.J. Barbetti, Temperature and plant age drive downy mildew disease epidemics on oilseed *Brassica napus* and *B. juncea*, *Eur. J. Plant Pathol.* 151 (2018) 703–711.
- [6] A.E. Mohammed, M.P. You, M.J. Barbetti, Resistances to downy mildew (*Hyaloperonospora brassicae*) in diverse *Brassicaceae* offer new disease management opportunities for oilseed and vegetable crucifer industries, *Eur. J. Plant Pathol.* 153 (2019) 67–81.
- [7] M.J. Barbetti, S. Banga, P.A. Salisbury, Challenges for crop production and management from pathogen biodiversity and diseases under current and future climate scenarios: Case study with oilseed Brassicas, *Field Crops Res* 127 (2012) 225–240.
- [8] E. Gäumann, Über die Formen der *Peronospora parasitica* (Pers.) Fries, *Beihefte zum Botanische Centralblatt* 35 (1918) 395–533.
- [9] E. Gäumann, Beiträge zu einer Monographie der Gattung *Peronospora* Corda, *Beiträge zur Kryptogamenflora der Schweiz* 5 (1923) 1–360.
- [10] O. Constantinescu, I.J. Fateh, *Peronospora*-like fungi (Chromista, *Peronosporales*) parasitic on *Brassicaceae* and related hosts, *Nova Hedwigia* 74 (2002) 291–338.
- [11] Y.-J. Choi, S.-B. Hong, H.-D. Shin, Sequence analysis of the ITS region of rDNA was used to investigate the level of genetic diversity occurring within *Hyaloperonospora parasitica*, and to show the relationship between phylogenies of these fungi and their hosts (*Brassicaceae*), *Mycol. Res.* 107 (2003) 1314–1322.
- [12] M. Göker, A. Riethmüller, H. Voglmayr, M. Weiß, F. Oberwinkler, Phylogeny of *Hyaloperonospora* based on nuclear ribosomal internal transcribed spacer sequences, *Mycol. Prog.* 3 (2004) 83–94.
- [13] M. Göker, H. Voglmayr, A. Riethmüller, M. Weiß, F. Oberwinkler, Taxonomic aspects of *Peronosporaceae* inferred from Bayesian molecular phylogenetics, *Can. J. Bot.* 81 (2003) 672–683.
- [14] H. Voglmayr, Phylogenetic relationships of *Peronospora* and related genera based on nuclear ribosomal ITS sequences, *Mycol. Res.* 107 (2003) 1132–1142.
- [15] M. Göker, H. Voglmayr, G.G. Blázquez, F. Oberwinkler, Species delimitation in downy mildews: the case of *Hyaloperonospora* in the light of nuclear ribosomal ITS and LSU sequences, *Fungal Biol.* 113 (2009) 308–325.
- [16] J.J. Natti, M.H. Dickson, J.D. Aktin, Resistance of *Brassica oleracea* varieties to downy mildew, *Phytopathology* 57 (1967) 144–147.
- [17] N.I. Nashaat, C.J. Rawlinson, The response of oilseed rape (*Brassica napus* ssp. *oleifera*) accessions with different glucosinolate and erucic acid contents to four isolates of *Peronospora parasitica* (downy mildew) and the identification of new sources of resistance, *Plant Pathol.* 43 (1994) 278–285.
- [18] N.I. Nashaat, A. Heran, S.E. Mitchell, R.P. Awasthi, New genes for resistance to downy mildew (*Peronospora parasitica*) in oilseed rape (*Brassica napus* ssp. *oleifera*), *Plant Pathol.* 46 (1997) 964–968.
- [19] A.E. Mohammed, M.P. You, H.F.D. Al-lami, M.J. Barbetti, Pathotypes and phylogenetic variation determine downy mildew epidemics in *Brassica* spp. in Australia, *Plant Pathol.* 67 (2018) 1514–1527.
- [20] A. Mohammed, M.P. You, M.J. Barbetti, New resistances offer opportunity for effective management of the downy mildew (*Hyaloperonospora parasitica*) threat to canola, *Crop Past. Sci.* 68 (2017) 234–242.
- [21] I.H. Chang, A preliminary study on the physiological differentiation of the downy mildews (*Peronospora parasitica* (Pers.) Fr.) of Chinese cabbage and other cruciferous vegetables in the vicinity of Peking and Tientsin, *Acta Phytopathol. Sin.* 7 (1964) 33–44.
- [22] D. McMeekin, Other hosts for *Peronospora parasitica* from cabbage and radish, *Phytopathology* 59 (1969) 693–696.
- [23] C. Dickinson, J. Greenhalgh, Host range and taxonomy of *Peronospora* on crucifers, *Trans. Brit. Mycol. Soc.* 69 (1977) 111–116.
- [24] C. Sherriff, J.A. Lucas, The host range of isolates of downy mildew, *Peronospora parasitica*, from *Brassica* crop species, *Plant Pathol.* 39 (1990) 77–91.
- [25] J.G. Vicente, N.D. Gunn, L. Bailey, D.A.C. Pink, E.B. Holub, Genetics of resistance to downy mildew in *Brassica oleracea* and breeding towards durable disease control for UK vegetable production, *Plant Pathol.* 61 (2012) 600–609.
- [26] C. Sherriff, J.A. Lucas, Heterothallism and homothallism in *Peronospora parasitica*, *Mycol. Res.* 92 (1989) 311–316.
- [27] N.D. Gunn, J. Byrne, E.B. Holub, Outcrossing of two homothallic isolates of *Peronospora parasitica* and segregation of avirulence matching six resistance loci

- in *Arabidopsis thaliana*. In: (eds). Spencer-Phillips PTN, Gisi U, Lebeda A. *Advances in Downy Mildew Research*, Springer, Dordrecht, Netherlands, 2002, pp. 185–188.
- [28] O. Spring, J. Gomez-Zeledon, D. Hadziabdic, R.N. Trigiano, M. Thines, A. Lebeda, Biological characteristics and assessment of virulence diversity in pathosystems of economically important biotrophic Oomycetes, *Crit. Rev. Plant Sci.* 37 (2018) 439–495.
- [29] J.L. Cenis, Rapid extraction of fungal DNA for PCR amplification, *Nucleic Acids Res.* 20 (1992) 2380.
- [30] M. Bachofer, *Molekular biologische Populations studien an *Plasmopara halstedii*, dem Falschen Mehltau der Sonnenblume*, Ph.D. thesis, University of Hohenheim, Stuttgart, Germany, 2004.
- [31] R. Chikhi, P.I. Medvedev, Informed and automated k-mer size selection for genome assembly, *Bioinformatics* 30 (1) (2014) 31–37.
- [32] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, et al., Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler, *GigaSci* 4 (2015).
- [33] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, et al., SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler, *GigaSci* (2012) 1.
- [34] S.D. Jackman, B.P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S.A. Hammond, et al., ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter, *Genome Res.* 27 (2017) 768–777.
- [35] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Research* 18 (2008) 821–829.
- [36] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, *Bioinformatics* 27 (2011) 578–579.
- [37] A.H. Wences, M.C. Schatz, Metassembler: merging and optimizing de novo genome assemblies, *Genome Biol* 16 (2015) 207.
- [38] J.H. Grau, T. Hackl, K.-P. Koepfli, M. Hofreiter, Improving draft genome contiguity with reference-derived in silico mate-pair libraries, *GigaSci* 7 (2018).
- [39] I. Baxter, J.N. Brazelton, D. Yu, Y.S. Huang, B. Lahner, E. Yakubova, et al., A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *ATHTK1;1*, *PLoS Genet* 6 (2010), e1001193.
- [40] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* 31 (2015) 3210–3212.
- [41] M. Urban, A. Cuzick, K. Rutherford, A. Irvine, H. Pedro, R. Pant, et al., PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database, *Nuc. Acid Res.* 45 (2017) 604–610.
- [42] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics* 21 (2005) 3674–3676.
- [43] G. Alexey, S. Vladislav, V. Nikolay, T. Glenn, QUAST: quality assessment tool for genome assemblies, *Bioinformatics* 29 (2013) 1072–1075.
- [44] B.M. Tyler, S. Tripathy, X. Zhang, P. Dehal, R.H. Jiang, A. Aerts, et al., *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis, *Science* 313 (2006) 1261–1266.
- [45] J.K. Hane, R.G.T. Lowe, P.S. Solomon, K.-C. Tan, C.L. Schoch, J.W. Spatafora, et al., Dothideomycete-Plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*, *Plant Cell* 19 (2007) 3347–3368.
- [46] J.S. David, P. Panda, E.S. Von Stowasser, M. González, R. Hill, C. Sambles, et al., Genome sequencing of oomycete isolates from Chile supports the New Zealand origin of *Phytophthora kernoviae* and makes available the first *Nothophytophthora* sp. genome, *Mol. Plant Pathol.* 20 (2019) 423–431.
- [47] T. Rujirawat, P. Patumcharoenpol, W. Kittichotirat, T. Krajaejun, Oomycete Gene Table: An online database for comparative genomic analyses of the oomycete microorganisms, *Database* (2019) 2019.
- [48] B. Davaasuren, J. Yamagishi, D. Mizushima, S. Narantsatsral, D. Otgonsuren, P. Myagmarsuren, B. Battsetseg, B. Battur, N. Inoue, K. Suganuma, Draft genome sequence of *Trypanosoma equipertum* strain IVM-t1, *Microbiol Resour Announc* 8 (2019), <https://doi.org/10.1128/MRA.01119-18> e01119-18.
- [49] C. Fischer, S. Koblmüller, C. Börger, et al., Genome sequences of *Tropheus moorii* and *Petrochromis trewavasae*, two eco-morphologically divergent cichlid fishes endemic to Lake Tanganyika, *Sci. Rep.* 11 (2021) 4309, <https://doi.org/10.1038/s41598-021-81030-z>.
- [50] A. Kumar, P. Mishra, R. Maurya, A.K. Mishra, V.K. Gupta, P.W. Ramteke, S. S. Marla, Improved draft genome sequence of a monoteliosporic culture of the Karnal bunt (*Tilletia indica*) pathogen of wheat, *Genome Announc.* 6 (2018) e00015–e00018, <https://doi.org/10.1128/genomeA.00015-18>.
- [51] S.S. Marla, P. Mishra, R. Maurya, M. Singh, D.P. Wankhede, A. Kumar, M.C. Yadav, N. Subbarao, S.K. Singh, R. Kumar, Refinement of Draft Genome Assemblies of *Pigeonpea (Cajanus cajan)*, *Front. Genet.* 11 (2020), 607432, <https://doi.org/10.3389/fgene.2020.607432>.
- [52] F. Moreno-Avitia, L. Lozano, J. Utrilla, F. Bolívar, A. Escalante, Draft genome sequence of *Pseudomonas chlororaphis* ATCC 9446, a nonpathogenic bacterium with bioremediation and industrial potential, *Genome Announc.* 5 (2017), <https://doi.org/10.1128/genomeA.00474-17> e00474-17.
- [53] F. Salvà-Serra, H.E. Jakobsson, A. Busquets, M. Gomila, D. Jaén-Luchoro, C. Seguí, F. Aliaga-Lozano, E. García-Valdés, J. Lalucat, E.R.B. Moore, A. Bennasar-Figueras, Genome sequences of two naphthalene-degrading strains of *Pseudomonas balearica*, isolated from polluted marine sediment and from an oil refinery site, *Genome Announc.* 5 (2017) e00116–e00117, <https://doi.org/10.1128/genomeA.00116-17>.
- [54] J.L. Steffani-Vallejo, C. Zuñiga, P. Cruz-Morales, L. Lozano, M. Morales, C. Licon-Cassani, S. Revah, J. Utrilla, Draft genome sequence of *Sphingobacterium* sp. CZ-UAM, isolated from a methanotrophic consortium, *Genome Announc.* 5 (2017), <https://doi.org/10.1128/genomeA.00792-17> e00792-17.
- [55] J. Takayama, S. Tadaka, K. Yano, et al., Construction and integration of three de novo Japanese human genome assemblies toward a population-specific reference, *Nat Commun.* 12 (2021) 226, <https://doi.org/10.1038/s41467-020-20146-8>.
- [56] B.-Z. Feng, X.-P. Zhu, L. Fu, R.-F. Lv, D. Storey, Tooley, et al., Characterization of necrosis-inducing NLP proteins in *Phytophthora capsici*, *BMC Plant Biology* 14 (2014) 126.
- [57] S. Oome, T.M. Raaymakers, A. Cabral, S. Samwel, H. Böhm, I. Albert, et al., Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in *Arabidopsis*, *Proc. Nat. Acad. Sci. USA* 111 (2014) 16955–16960.
- [58] R.G. Anderson, D. Deb, K. Fedkenheuer, J.M. McDowell, Recent progress in RXLR effector research, *Mol. Plant-Microbe Interact.* 28 (2015) 1063–1072.
- [59] X. Zhang, B. Liu, F. Zou, D. Shen, Z. Yin, R. Wang, et al., Whole genome re-sequencing reveals natural variation and adaptive evolution of *Phytophthora sojae*, *Front. Microbiol* 10 (2019) 2792.
- [60] S. Fawke, M. Doumane, S. Schornack, Oomycete interactions with plants: Infection strategies and resistance principles, *Microbiol, Mol. Biol. Rev.* 79 (2015) 263–280.
- [61] L. Atanasova, M. Dubej, M. Grujić, M. Gudmundsson, C. Lorenz, M. Sandgren, et al., Evolution and functional characterization of pectate lyase PEL12, a member of a highly expanded *Clonostachys rosea* polysaccharide lyase 1 family, *BMC Microbiol.* 18 (2018) 178.
- [62] A. Lara-Márquez, M.G. Zavala-Páramo, E. López-Romero, N. Calderón-Cortés, R. López-Gómez, U. Conejo-Saucedo, et al., Cloning and characterization of a pectin lyase gene from *Colletotrichum lindemuthianum* and comparative phylogenetic/structural analyses with genes from phytopathogenic and saprophytic/opportunistic microorganisms, *BMC Microbiol.* 11 (2011) 260.
- [63] L. Fu, C. Zhu, X. Ding, X. Yang, P.F. Morris, B.M. Tyler, et al., Characterization of cell-death-inducing members of the pectate lyase gene family in *Phytophthora capsici* and their contributions to infection of pepper, *Mol. Plant-Microbe Int.* 28 (2015) 766–775.
- [64] D. Eshel, I. Miyara, T. Ailing, A. Dinoor, D. Prusky, pH regulates endoglucanase expression and virulence of *Alternaria alternata* in *Persimmon* fruit, *Mol. Plant-Microbe Int.* 15 (2002) 774–779.
- [65] R. Carzaniga, P. Bowyer, R.J. O’Connell, Production of extracellular matrices during development of infection structures by the downy mildew *Peronospora parasitica*, *New Phytol.* 149 (2001) 83–93.
- [66] H.S. Judelson, A.M.V. Ah-Fong, Exchanges at the Plant-Oomycete interface that influence disease, *Plant Physiol.* 179 (2019) 1198–1211.
- [67] T.M.M.M. Amaro, G.J.A. Thilliez, G.B. Motion, E. Huitema, A perspective on CRN proteins in the genomics age: Evolution, classification, delivery and function revisited, *Front. Plant Sci.* (2017) 1–12, 99.
- [68] L. Liu, L. Xu, Q. Jia, R. Pan, R. Oelmüller, W. Zhang, et al., Arms race: diverse effector proteins with conserved motifs, *Plant Signal Behav.* 14 (2019), 1557008.
- [69] Y.A.A.H. Neto, N.G.d.R. Garzon, R. Peduzzi, H. Cabral, Specificity of peptidases secreted by filamentous fungi, *Bioengineered* 9 (2018) 30–37.
- [70] R.G. Lowe, O. McCorkelle, M. Bleackley, C. Collins, P. Faou, S. Mathivanan, M. Anderson, Extracellular peptidases of the cereal pathogen *Fusarium graminearum*, *Front. Plant Sci.* 6 (2015) 962.
- [71] K. Bailey, V. Cevik, N. Holton, J. Byrne-Richardson, K.H. Sohn, M. Coates, et al., Molecular cloning of ATR5(Emoy2) from *Hyaloperonospora arabidopsidis*, an avirulence determinant that triggers RPP5-mediated defense in *Arabidopsis*, *Mol Plant Microbe Interact.* 24 (2011) 827–838.
- [72] T. Rouxel, M.-H. Balesdent, Avirulence Genes, *Encyclopedia of Life Sciences (ELS)*, John Wiley & Sons, 2010. Ltd, Chichester, UK.