

RESEARCH ARTICLE

Open Access



# Chromosome-level genome assembly and manually-curated proteome of model necrotroph *Parastagonospora nodorum* Sn15 reveals a genome-wide trove of candidate effector homologs, and redundancy of virulence-related functions within an accessory chromosome

Stefania Bertazzoni<sup>1</sup>, Darcy A. B. Jones<sup>1</sup>, Huyen T. Phan<sup>1\*</sup>, Kar-Chun Tan<sup>1\*</sup> and James K. Hane<sup>1,2\*</sup> 

## Abstract

**Background:** The fungus *Parastagonospora nodorum* causes septoria nodorum blotch (SNB) of wheat (*Triticum aestivum*) and is a model species for necrotrophic plant pathogens. The genome assembly of reference isolate Sn15 was first reported in 2007. *P. nodorum* infection is promoted by its production of proteinaceous necrotrophic effectors, three of which are characterised – ToxA, Tox1 and Tox3.

**Results:** A chromosome-scale genome assembly of *P. nodorum* Australian reference isolate Sn15, which combined long read sequencing, optical mapping and manual curation, produced 23 chromosomes with 21 chromosomes possessing both telomeres. New transcriptome data were combined with fungal-specific gene prediction techniques and manual curation to produce a high-quality predicted gene annotation dataset, which comprises 13,869 high confidence genes, and an additional 2534 lower confidence genes retained to assist pathogenicity effector discovery. Comparison to a panel of 31 internationally-sourced isolates identified multiple hotspots within the Sn15 genome for mutation or presence-absence variation, which was used to enhance subsequent effector prediction. Effector prediction resulted in 257 candidates, of which 98 higher-ranked candidates were selected for in-depth analysis and revealed a wealth of functions related to pathogenicity. Additionally, 11 out of the 98 candidates also exhibited orthology conservation patterns that suggested lateral gene transfer with other cereal-pathogenic fungal species. Analysis of the pan-genome indicated the smallest chromosome of 0.4 Mbp length to be an accessory chromosome (AC23). AC23 was notably absent from an avirulent isolate and is predominated by mutation hotspots with an increase in non-synonymous mutations relative to other chromosomes. Surprisingly, AC23 was deficient in effector candidates, but contained several predicted genes with redundant pathogenicity-related functions.

\* Correspondence: [Huyenphan.Phan@curtin.edu.au](mailto:Huyenphan.Phan@curtin.edu.au); [Kar-Chun.Tan@curtin.edu.au](mailto:Kar-Chun.Tan@curtin.edu.au); [James.Hane@curtin.edu.au](mailto:James.Hane@curtin.edu.au)

<sup>1</sup>Centre for Crop & Disease Management, Curtin University, Perth, Australia  
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions:** We present an updated series of genomic resources for *P. nodorum* Sn15 – an important reference isolate and model necrotroph – with a comprehensive survey of its predicted pathogenicity content.

## Background

The fungus *Parastagonospora nodorum* causes septoria nodorum blotch (SNB) of wheat (*Triticum aestivum*) and is a model species for necrotrophic plant pathogens. In order to provide insight on the evolutionary history and gene repertoire of this pathogen, a genome assembly of *Parastagonospora nodorum* model isolate Sn15 was first reported in 2007 [1]. This used Sanger shotgun sequencing of a genomic BAC library which produced a 37.5 Mbp draft genome reference with 108 scaffolds and 10,762 genes. It was the first species among the class Dothideomycetes for which a whole-genome reference was available [1] and has been used as a model species for cereal necrotrophs. This draft genome resource contributed to the discovery of three proteinaceous necrotrophic effectors (NEs) corresponding to known gene loci - ToxA [2], Tox1 [3] and Tox3 [4] - which are major host-specific virulence determinants in *P. nodorum*. The presence of additional NEs have been detected via their interaction with quantitative trait loci (QTL) corresponding to host sensitivity loci, but the genes encoding these effectors have not yet been identified [5–11] and others may have not yet been uncovered. In order to discover novel effectors in *P. nodorum* and in other fungal plant pathogens, it is important to ensure that the genome assembly and gene annotations are as accurate and reliable as possible.

Recent advances in long-read genome sequencing technologies, and established genetic and physical mapping techniques, have made whole-chromosome assembly of microbial genomes readily achievable [12–18]. Three decades ago, chromosome size and number estimates via pulsed-field gel electrophoresis (PFGE) of 11 *P. nodorum* isolates had estimated a range from 14 to 19, totalling 28 to 32 Mbp, ranging from 0.4 to 3.5 Mbp in length, with the smallest observed only in wheat and barley-infecting isolates [19]. The *P. nodorum* Sn15 genome assembly was progressively improved over subsequent years. It was updated in 2013 reducing the number of scaffolds from 108 to 91 [20], and again in 2016 with revised gene annotations that were supported by protein and transcriptome alignments and manual curation [21–23]. Leveraging these resources, comprehensive analyses of its genomic landscape and genome-based processes contributing to pathogenic adaptations have extended to transposable elements (TE) and gene repeats [1, 24, 25], repeat-induced point mutations (RIP) [24–26], mesosynteny [27], and multiple comparative genomics studies [14, 15, 23, 28, 29]. Initially, the Sn15

reference isolate was compared to a hyper-virulent isolate (Sn4) and a non-aggressive isolate (Sn79–1087) lacking known effector genes *ToxA*, *Tox3* and *Tox1* [20, 23]. These newly gained information and resources have played a vital role in studying important pathogenicity gene candidates. Subsequent comparison to an international panel (across 10 countries) of 22 *P. nodorum* and 10 *Parastagonospora avenae* isolates indicated presence-absence variation (PAV) - with notable absences in the ‘avirulent’ Sn79–1087 isolate assembly - of known effector loci and of large regions (i.e. scaffolds 44, 45, 46 and 51) [23], which was supplemented by a predictive analysis of accessory chromosome (AC) or region (AR) sequence properties (scaffolds 50 and 69) [29]. These large PAV regions were indicative of ACs/ARs that are associated with host-specific virulence in numerous fungal species [30], but this could not be confirmed with an unfinished genome assembly. In 2018, long-read-based genome assemblies were generated for 3 *P. nodorum* isolates (Sn4, Sn79–1087 and Sn2000) with 22 to 24 contigs [15]. Analysis of the Sn4 genome revealed that ‘contig23’ (~0.48 Mbp) was absent in Sn79–1087 and therefore considered an AC. This study also used transcriptome data from the Sn15 reference isolate to ‘auto-annotate’ genes in Sn4, and subsequently trained gene prediction software on the Sn4 annotations, which was used to perform in silico prediction of genes in the remaining isolates [15]. A follow up study in 2019 compared these four assemblies to NGS-based assemblies for a panel of 197 isolates from the United States, highlighted widespread diversifying selection within predicted effector loci and across the AC Sn4 contig23, reinforced the impact of the known *ToxA*, *Tox1* and *Tox3* effectors, and predicted 17 candidate effector loci with high levels of diversifying selection [31].

The recent updates to *P. nodorum* genomic resources enable consideration of the genomic landscape and sequence features which are relevant to pathogenicity or adaptation at the chromosome-scale, such as repeat-rich regions and mutation hotspots [26, 30, 32]. Long-read-based methods have significantly improved genome assembly of these previously challenging regions [12, 16, 18, 33, 34] and scaffold lengths can be further improved with genome-finishing techniques including optical restriction [14, 35] or chromosome interaction mapping [36]. In fungal pathogens with “two-speed” genomes, repeat-rich regions typically accumulate mutations more rapidly than conserved gene-rich regions [26], leading to compartmentalisation of pathogen genomes into stable

GC-equilibrated regions and AT-rich ‘mutation hotspots’, which can include pathogenicity-associated ACs or ARs [30]. For pathogenicity loci not residing within ACs, growing evidence supports their frequent location in sub-telomeric mutation hotspots [37–39] which may also be ARs. The segregation bias of certain gene functions to the sub-telomere may be associated with the role of heterochromatin found at sub-telomere region in regulating gene expression during infection [39] and protection of the core genome from interspersions of sub-telomeric heterochromatin [40].

The presence or absence of effector genes, or ACs/ARs that contain them, can determine host/cultivar-specific virulence for several pathogen species [30]. Bioinformatic methods for effector prediction are usually of a reductive nature, which filter the complete gene set down to a candidate effector subset based upon multiple criteria [41]. These methods typically require effector gene annotations not to have been missed in the complete gene set (at either assembly or gene prediction steps), and directly benefit from the proper application of transcriptome data to gene annotation, which for gene-dense genomes like those of fungi can pose a technical challenge [42]. In this study, we present an updated chromosome-scale genome assembly for *P. nodorum* reference isolate Sn15, combining long-read data and optical mapping to arrive at a near complete telomere-to-telomere assembly of 23 chromosomes. Sn15 gene annotations have also been updated integrating new transcriptome data and extensive manual curation, which will ensure its reliability and ongoing utility as a model necrotroph. Insights from comparative genomics analysis is presented for comparisons of the Sn15 reference isolate versus the Sn4, Sn2000 and Sn79–1087 long-read assemblies, and an international panel of NGS-based assemblies for 28 other *Parastagonospora* isolates. This has highlighted mutation hotspots and locational biases across the 23 chromosomes of Sn15, including a 0.4 Mbp accessory chromosome and several telomeric ARs. New effector gene predictions for Sn15 are also provided, integrating the wealth of past data for Sn15 with new data including PAV and diversifying selection across the international pan-genome. These aggregated resources for *P. nodorum* Sn15 will offer novel research opportunities and serve as a useful tool to enhance ongoing efforts to breed for crop disease resistance.

## Results

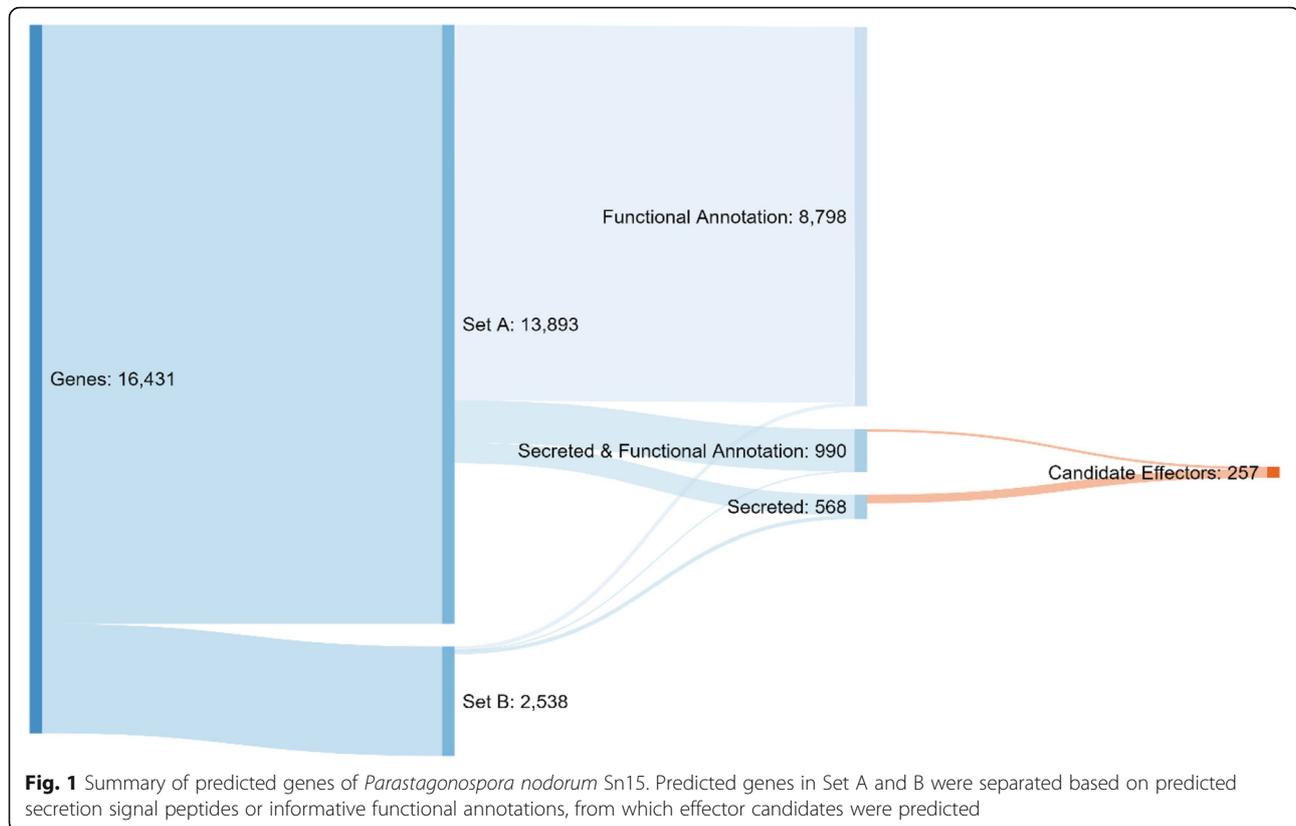
### A chromosome-level reference genome assembly for *P. nodorum* Sn15

In order to complete the Sn15 assembly, a combination of long read sequencing using PacBio technology and optical mapping were used. PacBio DNA sequencing generated 368,822 raw reads of 50 bp to 41 Kbp in

length at ~71X coverage. Self-correction resulted in 118,028 corrected reads, totalling 1.31 Gbp with an average length of 10 Kbp. Corrected reads were assembled into a draft assembly of 36 gapless contigs ranging from 3.5 Mbp to 37 Kbp, with a total length of 37.4 Mbp at 33.6X coverage. Only 844 corrected reads (0.7%) were not assembled. One of the 36 contigs corresponded to the previously published mitochondrial DNA sequence [GenBank: [EU053989](#)] and was discarded. An optical map produced 23 maps with an estimated total length of 39.26 Mbp (Supplementary Text 1). Thirty out of the 35 assembled contigs (36.88 Mbp) aligned to the 23 optical maps. The 5 contigs that did not align were short (38 to 120 Kbp, or 0.84% of the contig assembly) and highly repetitive, with no predicted genes. The curated scaffolds of contigs aligned to the 23 optical maps - subsequently referred to as ‘chromosomes’ - were numbered in descending size order based on the physical lengths predicted by the optical map (Supplementary Table 5). Fourteen chromosomes contained no gaps, and 8 gaps were added to join non-overlapping contigs within chromosomes 2, 3, 4, 6, 7, 10 and 20. Terminal ‘TTAGG’ tandem repeats indicating telomeres were observed at both ends of 21 chromosomes, with 2 having a single telomere. New repetitive regions comprised ~0.4% of the assembly. The new assembly had 286 fewer gaps than the previous version [20, 23] and there was a ~4 Kbp increase in the average length of AT-rich regions, a reduction of incompletely assembled AT-rich regions (-46) and an increase in fully assembled AT-rich regions (+33) (Supplementary Table 3).

### A revised set of gene annotations aggregated from multiple sources of evidence, including new *in planta* RNA-seq, fungal-specific gene finding software and manual curation

An estimate of the representation of the core gene content in the updated Sn15 assembly via BUSCO (v5.1.2) indicated 99.1% completeness versus the “fungi” dataset (fungi\_odb10, 2020-09-10). The combination of various gene prediction methods (see methods), incorporating recently published *in vitro* and *in planta* RNA-seq data [23, 43], fungal-specific gene prediction software, and manual curation, resulted in 16,431 predicted genes. This gene set was split into two subsets: a higher confidence set (Set A), and a lower confidence set to allow more sensitivity for subsequent pathogenicity gene predictions (Set B) (Fig. 1, Table 1A). Set A included 13,893 high confidence genes models with higher levels of support, whereas set B contained 2538 putative genes with either shorter coding sequence length or less RNA-seq support (Table 1B). Compared to the previously published annotation [20, 23], average gene length decreased by 70 bp and gene density increased by 2.8 genes per



Mbp (Table 1A). Set B annotations were on average length 4 times shorter than those of Set A and in 86% of cases were a single exon (Table 1A). Of the 16,341 genes, 9788 were informatively functionally annotated (i.e. a conserved domain), and 990 of these also had a predicted secretion signal peptide (Fig. 1). The predicted secretome comprised 1568 genes of which 257 (1.5% of total genes and 25.3% of the secretome) were effector candidates (Fig. 1, Table 1). Across the Sn15 genome, gene density was inversely correlated with density of repetitive DNA (Fig. 2), with genes distributed at a relatively even density (~450 Mbp) except for accessory chromosome 23 (AC23 which was gene sparse (~380 Mbp) (Fig. 2). A lower proportion (36.7%) of loci were assigned functional annotations within AC23, which was 13% less than average. The known necrotrophic effector genes *ToxA*, *Tox1* and *Tox3* were all located within sub-telomeric regions of chromosome 4, 10 and 11 respectively, with *ToxA* also notably residing in the middle of a large (~570 Kbp) repeat-rich region (Fig. 2).

### Comparative genomics

In comparisons of the Sn15 genome to alternate isolates, the Sn15 genome exhibited multiple large PAV regions (Fig. 2, Supplementary Table 6, Supplementary Table 7). Prior pan-genome and in silico studies using the previously published Sn15 assembly as its reference genome

had indicated scaffolds 44, 45, 50 and 51 as regions of the genome with PAVs [20, 23, 29] (Supplementary Table 8).

Scaffold 50 corresponded to a sub-telomeric region of chromosome 8 (Supplementary Table 8). New reports of additional variable regions derived from this study include regions of chromosomes 7, 8 and 10. Chromosome 7 contained a ~455 Kbp region that is potentially duplicated in some isolates, but is represented in single copy in the current Sn15 assembly. Chromosomes 8 and 10 contained ~88 Kbp and ~10 Kbp PAV regions respectively. The PAV on chromosome 10 contained no genes, and the PAV on chromosome 8 contained 27 genes (Supplementary Table 9) but did not contain any predicted effector candidates (Supplementary Table 4).

Former scaffolds 44 and 45 corresponded to the ~444 Kbp AC23 of this study (Supplementary Table 8). Pan-genome alignment of Sn15 chromosomes with other *P. nodorum* and *P. avenae* isolates indicated that chromosome 23 was absent in *P. avenae* and the non-aggressive *P. nodorum* isolate Sn79–1087 (Fig. 2), which suggested that it lacked genes required for viability and was an accessory chromosome. In contrast, the majority of other “core” chromosomes were well conserved across *Parastagonospora* spp. AC23 also exhibited higher overall levels of non-synonymous mutations indicating diversification across this population relative to the Sn15

**Table 1** Summary of new gene annotations of *P. nodorum* reference isolate Sn15. A) Comparison of high-confidence Set A and low confidence Set B to previous annotation versions and B) summary of data supporting gene annotations

A) Summary	Previous studies [20, 23]	This study (Set A)		This study (Set B)				
Number of genes	13,569	13,869		2534				
Number of mRNAs	13,944	14,160		2557				
Average gene length, bp	1558	1488		373				
Number of exons	36,557	36,447		3070				
Average exons/gene	2.6	2.6		1.2				
Average exon length (bp)	556	539		299				
Gene density (genes / Mbp)	369	372		440 (Set A + Set B)				
B) Supporting Data	Loci	FPKM > 50	FPKM > 5	FPKM < 5	SignalP	SignalP+ EffectorP	Region not in previous assembly	Fuctional annotation
<b>Set A</b>	13,869	4033	9968	4030	1488	340	19	9505
In previous studies [20, 23]	13,663	3742	9633	4030	1463	323		9456
Not in previous studies [20, 23]	206	291	335	0	25	17		49
<b>Set B</b>	2534	220	2508	26	97	61	23	75
In previous study [1]	117	5	117	0	19	6		0
New or modified genes	2417	215	2391	26	77	55		75

reference isolate (Supplementary Table 10). However the mutation profile of AC23 contained two regions separated by a large repeat island – each side corresponding to scaffolds 44 and 45 of the previously Sn15 assembly [20, 23] - which exhibited distinctly different mutation rates (Fig. 2). Comparison of AC23 to other Sn15 chromosomes did not indicate that it had originated from duplication of core chromosomes (Supplementary Figure 1), however homologous (non-repetitive) regions in *Pyrenophora tritici-repentis* [14, 44] and *Bipolaris* spp. [28, 29] genomes tended to be located in sub-telomeric regions (Supplementary Figure 2).

The previous scaffold 51 corresponded to a ~ 74 kbp region within a repeat-rich sub-telomeric region of chromosome 4 (Supplementary Table 8), which also contained the effector gene *ToxA*. This sub-telomeric region had below average GC content, correspondingly high repeat content (~ 18.3% higher than the genome average), increased mutation density and less than half of the average gene density (Supplementary Table 6). The 9 predicted loci within this region had an average DN/DS of 1.9, more than double the genome average (Supplementary Table 6). Alignment of this region between the Sn15 assembly presented in this study, and the long read assemblies of Sn4, Sn2000 and Sn79–1087, showed structural variations that may indicate that breakage-fusion bridge (BFB)-mediated rearrangements (distal translocations between chromosomes lacking telomere caps) may have occurred in one or more of

these isolates (Supplementary Figure 3) [45]. Comparisons of this region to corresponding regions containing *ToxA* homologs in related species *Pyrenophora tritici-repentis* [14, 44], *Bipolaris maydis* [29] and *B. sorokiniana* [28], indicated further chromosome structure diversity. The sub-telomeric *ToxA* region of chromosome 4 in *P. nodorum* appeared to be consistent with *Bipolaris* spp. where it was also found in sub-telomeric locations. In contrast, the *P. tritici-repentis ToxA* -containing chromosome appeared to be a product of the breakage of the *P. nodorum ToxA* region, followed by chromosome fusions resulting this region being flanked by sequences corresponding to *P. nodorum* chromosomes 14 and 19 (Fig. 3).

## Discussion

### The chromosome-level assembly for *P. nodorum* reference isolate Sn15 improved detection of pathogenicity gene-rich regions

The new chromosome-level genome assembly of *P. nodorum* Sn15 created by this study has established the correct number of chromosomes for this pathogen, which was previously underestimated by PFGE to range from 14 to 19 [19], and is consistent with 22–23 observed in assemblies of other isolates [15, 31]. PFGE fragment resolution accuracy requires at least ~ 1% difference in chromosome size [46], meaning 6 out of the 23 assembled sequences were within a potentially unresolvable size range (Supplementary Table 5). The



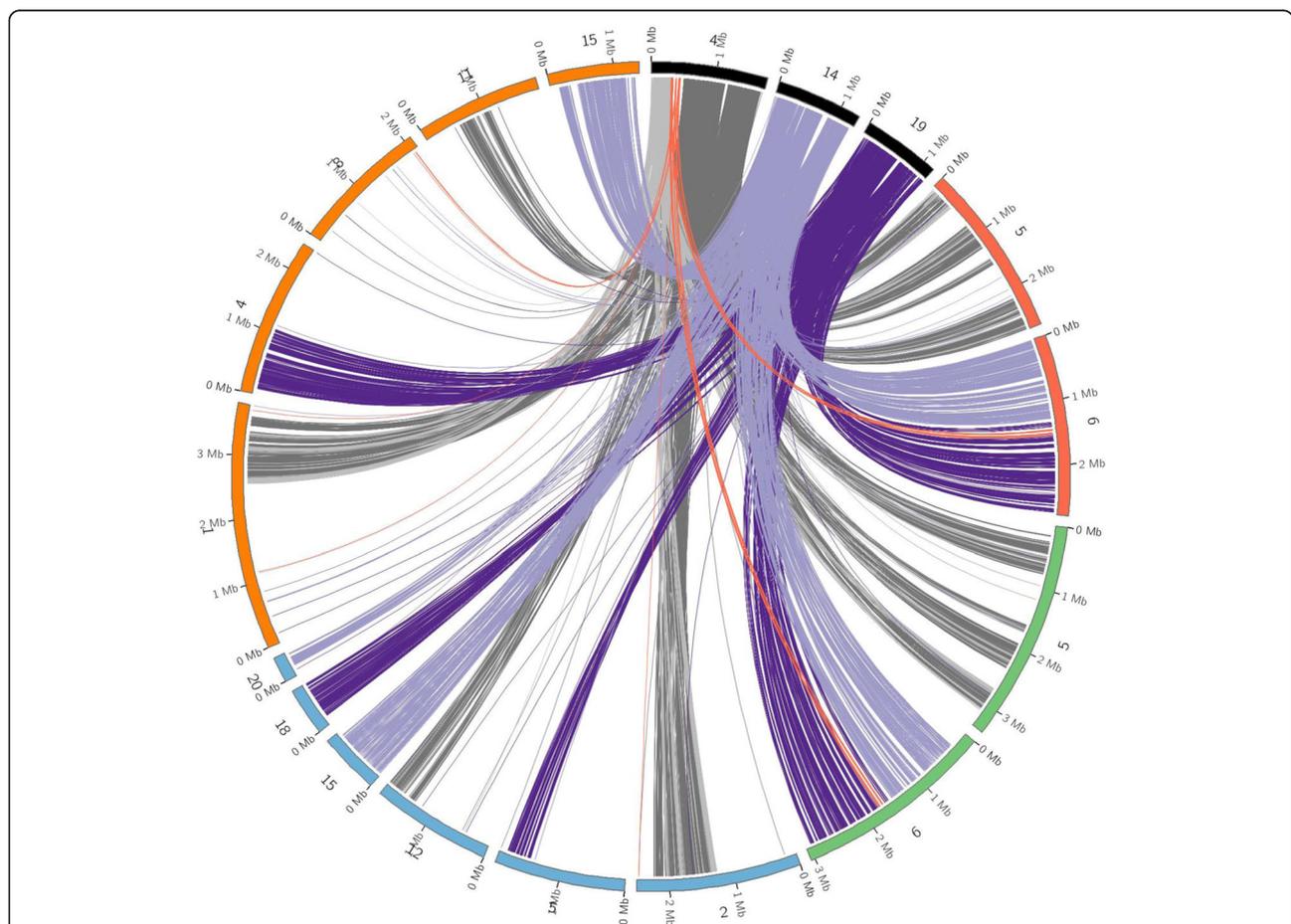
(See figure on previous page.)

**Fig. 2** Sequence comparisons of the new genome assembly of the *Parastagonospora nodorum* Sn15 reference isolate with alternate *P. nodorum* isolates and *P. avenae* isolates, within 50 Kbp windows, for: **a** Presence-absence variation (PAV) indicated by percent coverage of MUMmer matches (green), **b** SNP density (red), and **c** the ratio of non-synonymous to synonymous SNP mutations (DN/DS) relative to Sn15 (purple). Rings indicate (in inwards order): i) Sn15 chromosome (black); ii) loci predicted by EffectorP and score from 0 to 1 (dark green); iii) gene presence (blue); iv) AT-rich regions (orange); v) repeat regions (red); vi) average SNP mutation density from (b) (orange); vii) average DN/DS from (c) (purple); viii) PAV versus alternate isolates Sn4, Sn79–1087 and Sn2000; ix) *P. nodorum* isolate draft assemblies; x) *P. avenae* isolate draft assemblies. *P. nodorum* Sn15 accessory chromosome 23 (AC23) has been highlighted with regions corresponding to scaffolds 44 (yellow) and 45 (red), previously reported in Syme et al. 2018 to be conditionally-dispensable and under positive selection

difference in chromosome number between the two studies therefore is justified. This study also presents 21 out of 23 chromosomes with both telomeric ends and 14 gapless chromosomes. In addition, the new chromosome-level genome assembly for Sn15 was also supported by transcriptome data and manually curated gene annotations, and related bioinformatic resources for the *P. nodorum* Sn15 reference isolate were updated,

enhancing these important resource for studying molecular host-pathogen interactions and for effector discovery [41, 47, 48].

Chromosome-level analysis of the genomic landscape can enable detection of compartmentalised mutation ‘hot spots’ that may contain pathogenicity loci - a commonly reported feature for “two-speed” genomes which have been broadly affected by transposon activity and



**Fig. 3** Sequence similarity comparisons between *ToxA*-containing and related sequences of (A, black) *P. nodorum* Sn15 (chromosomes 4, 14 and 19); (B, red) *Pyrenophora tritici-repentis* BFP (chromosomes 5 and 6); (C, green) *Pyrenophora tritici-repentis* M4 (chromosomes 5 and 6); *Bipolaris maydis* (blue) (scaffolds 2, 5, 12, 15, 18 and 20); and *Bipolaris sorokiniana* CS10 (orange) (chromosomes 1, 4, 8, 1 and 15). Matches with *P. nodorum* Sn15 chromosome 4 are coloured grey, with the *ToxA*-containing region highlighted in red, and matches with *P. nodorum* Sn15 chromosomes 14 and 19 and coloured light and dark purple respectively

repeat-induced point mutation (RIP) [24, 26]. As genome assemblies have been improved towards chromosome-scale representation, there have also been several reports of pathogenicity genes within sub-telomeric locations in other pathosystems [15, 17, 37, 38, 40]. Thus, the new Sn15 assembly presented new opportunities to predict novel pathogenicity-related genes within the ‘two-speed’ regions that are repeat-rich or conditionally-dispensable [26, 30, 32]. Presumably, the apparent bias of pathogenicity gene locations within mini-chromosome or sub-telomeric regions could be associated with BFB formation [45] and mesosyntenic rearrangements [27, 32] between chromosome termini. Indeed, pan-genome comparisons indicated that the 0.4 Mbp AC23 was an accessory chromosome, with gene content relevant to pathogenicity (see below). The updated Sn15 assembly also highlighted additional regions not present in previous assembly versions, which comprised ~152 Kbp of repetitive DNA (0.4% of the genome) and 86,468 bp of non-repetitive DNA. While these represented a very small proportion of the genome, they may have special significance for plant pathology as they are more likely to contain effector or other pathogenicity genes. The relative placement of these regions in the genomic landscape was also important in assessing their likely roles in pathogenicity adaptation [26, 30, 32] as a parameter for effector prediction [41].

#### Candidate effector genes were derived from extensive gene annotation data for *P. nodorum* Sn15

Considerable efforts have been made across previous studies to ensure the reliability and ongoing applicability to plant pathology research of the annotated gene set for *P. nodorum* Sn15 [1, 20–23, 42], particularly for the purpose of effector and pathogenicity gene discovery. The revised Sn15 gene set includes a primary set of 13,893 genes (Set A) and a lower confidence set of 2538 (Set B) which was retained to enhance the sensitivity and capacity of effector gene predictions. The total number of predicted genes has increased since the previous annotation version [23], and is also higher than the currently reported average across the Ascomycota [49]. However we note general trends across all species, that while reducing assembly fragmentation can reduce the total number of predicted genes [50], the addition of significantly improved transcriptome data [51] or gene prediction methods [42] can increase this number. Functional annotations were assigned to 59.5% of predicted genes (Set A + B, excluding non-specific features e.g. coiled-coils, intrinsic disorder).

Across the whole genome, 257 effector candidate genes were predicted (Fig. 1, Supplementary Table 4), a number comparable to similar fungal pathogen genome surveys [41]. Effector candidate genes exhibited the

typical features expected of effectors, including: secretion, low molecular weight, cysteine richness, diversifying selection, association with mutation hotspots, and where functional annotations were assigned these had a common pathogenicity-related theme (Supplementary Table 4). Secretion was predicted for 1558 genes (9.5% of Set A + B), of which 257 (16.5% of predicted secretome) were effector candidates and 12 were predicted to localise to the chloroplast (including the confirmed effector *ToxA*) (Supplementary Table 4). Effector candidate loci were typically found within either 5–10 or 20–25 Kbp of AT-rich regions, which was not the case across the whole gene set (Supplementary Table 4, Supplementary Table 11). This is consistent with reports of RIP and effector location bias within AT-rich mutation hotspots [26]. The *ToxA* locus was 4039 bp and *Tox3* was 1860 bp from their nearest respective AT-rich regions. The *Tox1* locus was located >200 Kbp from its nearest AT-rich region, however all 3 effector loci were also located within sub-telomeric regions (Fig. 2). This association between telomeres and effector-rich mutation hotspots is also reported in other pathogen species [30, 32]. Comparison of orthologs between the Australian reference isolate Sn15 and the US isolate Sn4 [15], indicated 14 out of the 17 previously published Sn4 candidate effectors were also predicted among the Sn15 candidates (Supplementary Table 12). These Sn4 candidates – which included *Tox1* – were previously reported to exhibit diversifying selection that was specific to one of the 2 major US sub-populations [31].

#### Functionally-redundant genes may be associated with potential pathogenic properties of accessory chromosome 23

Surprisingly, AC23 which exhibited typical characteristics of ACs [15, 30, 31, 52] – and may correspond to anecdotal reports of a ~0.4 Mbp AC specific to *P. nodorum* wheat and barley-infecting isolates [53] – had a relatively low density of effector candidate loci (Supplementary Table 4, Supplementary Table 10). Six candidate effector loci were predicted on AC23 in a previous study [23], with two of these (SNOG\_16226 and SNOG\_16236) re-predicted (with ranked scores of 10 and 9 respectively) in the more stringent predictions of this study. AC23 also encoded multiple genes with other pathogenicity-related and/or redundant functions, which may indicate tandem duplications or multiple BFB events. These functions included: Ulp1 protease (SNOG\_16274, SNOG\_16214), RING/FYVE/PHD-type zinc finger proteins (SNOG\_16310, SNOG\_16333), valyl-trna synthase (SNOG\_16268, SNOG\_16213, SNOG\_16211), and UstYa-like protein (mycotoxin biosynthesis) (SNOG\_16357) (Supplementary Table 13). Ulp1 protease is involved in the modification of SMT3, a

ubiquitin-like protein of the SUMO family which suppresses MIF2 mutations. MIF2 is a centromere protein that regulates stability of di-centromeric minichromosomes in baker's yeast [54]. Its presence on AC23 is notable given that AC23 is a mini-chromosome and is therefore more likely to be unstable. UstYa-like proteins are involved in the secondary metabolite synthesis of cyclic peptide mycotoxins including ustiloxin and chyclochlorotine [55], however the products of many remain unknown. FYVE domain zinc finger proteins reportedly may bind to phospholipid PI3P [56], which could potentially facilitate host cell uptake. The genes and functions listed above represent candidate pathogenicity loci residing on AC23 which are of high importance for further investigation.

#### A trove of effector and pathogenicity gene homologs were predicted among candidate effector-loci

We previously observed that the deletion of *ToxA*, *1* and *3* in *P. nodorum* SN15 resulted in a mutant that retained near-WT level of virulence on most commercially adopted wheat varieties [57, 58]. This suggested that SN15 that lacked *ToxA*, *Tox1* and *Tox3* may have produced undiscovered effectors or other virulence factors to functionally compensated for the loss of these major NE genes [59, 60]. In addition, biochemical and genetic characterisation of US *P. nodorum* isolates identified evidence of other NEs [5–11]. This prompted us to apply a bioinformatic approach to predict for NE candidates in the near-complete SN15 genome that are relevant to the Australian cereal industry. From the prediction analysis, *ToxA*, *Tox1* and *Tox3* ranked highly among the top 98 Sn15 candidates with ranked scores of 5 and above (Supplementary Table 10). While remaining candidates are unconfirmed, among these we observed a wealth of assigned functions or matches strongly suggesting roles in pathogenicity. *SNOG\_13622* and *SNOG\_08876* encode for CFEM domain proteins, which have roles in iron acquisition and several of which have been reported with roles in virulence [61]. *SNOG\_42372* and *SNOG\_07772* encode for chitin-binding LysM domain proteins which offer protection from PTI in the host [62]. *SNOG\_07596* encodes a thaumatin-like protein, which when produced by host plants are pathogenesis-related (PR) proteins involved in defence, however fungal homologs have also been reported with roles in virulence [63]. *SNOG\_03746* encodes a knottin-like protein. Knottins are cytotoxins that are best represented by snake and arachnid venoms, with the first fungal report of a knottin in the poplar rust *Melampsora larici-populina* [64]. *SNOG\_30910* encodes a homolog to phospholipase A2 - which cleaves sn-2 acyl bond between 2 phospholipids that releases arachidonic acid and lysophosphatidic acid - and is also a common domain in spider, insect and snake venoms

that disrupt cell membranes [65]. *SNOG\_00200* encodes a product similar to *Alternaria alternata* allergen 1 (AA1-like). The AA1-like family [66] contains the *V. dahliae* effector PevD1, which binds the host thaumatin PR5 [67]. *SNOG\_00182*, *SNOG\_02182*, and *SNOG\_16063* encode ribotoxins, which have a conserved sarcin/ricin loop (SRL) structure that cleaves specific sequences in the host rRNA, leading to ribosome inactivation and cell death by apoptosis [68]. *SNOG\_13722* encodes a cerato-platanin, which induces phytoalexin synthesis and causes necrosis [69]. *SNOG\_06012* encodes a protein similar to gamma crystallin/yeast killer toxin, which is a pore-forming cytotoxin [70]. *SNOG\_01218* encodes a subtilisin, a serine protease family that is frequently reported in fungi to promote virulence [71]. *SNOG\_03959* encodes a protein similar to a cyclophilin-like/peptidyl-prolyl cis-trans isomerase (PPIase) [72], which in humans is well known for interfering with the immunosuppressive drug cyclosporin A, but is widespread across eukaryotes and has been reported as virulence determinants in several fungi including: *Leptosphaeria* spp., *Botrytis cinerea*, *Cryphonectria parasitica*, *Puccinia triticina*, *M. oryzae*, and *Lhelliinus sulphurascens*, as well as various oomycete species of the *Phytophthora* genus [73]. *SNOG\_08289* encodes a pectin/pectate lyase, which are reported in many fungi to promote virulence [74]. *SNOG\_11034* encodes a protein similar to Egh16, an appressorially-located virulence factor of *Blumeria graminis* f. sp. *hordei* with broadly conserved homologs across several pathogenic fungal species [75]. *SNOG\_15608* encodes a cutinase, which may be involved in host surface penetration [76]. *SNOG\_02399*, *SNOG\_03334*, *SNOG\_40970*, *SNOG\_08150*, *SNOG\_04779* all encode for proteins with lipid interacting domains. *SNOG\_11842* encodes a Hce2 effector homolog - which is named after Homologs of *C. fulvum* ECP2, a necrosis inducing effector. There are 3 defined classes of proteins with Hce2 domains, of which *SNOG\_11842* belongs to class I, the smallest and most common class [77]. Many of the above candidates with pathogenicity-related functional annotations are also expressed higher *in planta* (IP) relative to *in vitro* (IV) by a factor of 5, however the *Tox3* IP:IV is only 2 indicating this lower values may also be relevant in host-pathogen interactions. A lower-ranked candidate (*SNOG\_06459*) with a ranked score below 5 is also mentioned here as it encoded a cerato-ulmin homolog. Cerato-ulmin is a hydrophobin, which is not a functional class normally reported to be directly involved in pathogenicity, but has been reported as a potential virulence factor in dutch elm disease [78]. Its mode of action is not like a typical effector however, as its role appears to be to protect spores from desiccation, which leads to increased spore survivability and transmission.

### Multiple effector candidate loci were predicted to be laterally-transferred with other cereal-pathogenic fungal species

Of the 98 highly-ranked effector candidates, 11 showed a conservation pattern indicating potential lateral transfer when compared to a panel of whole gene sets of > 150 fungal species (Supplementary Table 4) [79]. This included *SNOG\_16571* (*ToxA*), *SNOG\_20078* (*Tox1*), *SNOG\_13622* (CFEM domain), *SNOG\_15952* (ribotoxin-like), *SNOG\_00152*, *SNOG\_01658*, *SNOG\_20100*, *SNOG\_08426*, *SNOG\_07039*, *SNOG\_00726*, and *SNOG\_14618*. These had rare orthology relationships indicating potential lateral gene transfer (LGT) with *Pyrenophora* spp., *Setosphaeria turcica*, *Alternaria brassicicola*, *Verticillium dahliae*, *Leptosphaeria maculans* and *Coccidioides immitis*. Aside from *SNOG\_13622*, *SNOG\_15952*, and known effectors *ToxA* and *Tox1*, this group of effector candidates had no predicted functional annotations. As expected, *SNOG\_08981* (*Tox3*) was not included in this set and has so far been reported to have no known homologs.

### Conclusions

The *P. nodorum* isolate Sn15 was the first representative of the class Dothideomycetes with a genomic survey report [1], and has since become an important reference and model necrotroph with a significant set of accumulated genomic, transcriptomic, proteomic and bioinformatic resources supporting its genome and gene data [15, 20–25, 31, 43]. This study updates these resources in the context of a chromosome-scale assembly, identifying genome features relevant to pathogenicity i.e. sub-telomeric regions, accessory chromosomes and mutations hotspots. This has provided genomic context to subsequent predictions of candidate genes encoding effectors and other pathogenicity factors. In contrast to the earliest Sn15 genome study, effector candidates were supported by a wealth of functional annotation and comparative genomics data indicating strong homology to known effectors and other pathogenicity genes. This study is an important step forward for the further characterisation of *P. nodorum* chromosome structure and its role in pathogenicity, particularly in highly mutable and potentially effector-rich regions of the genome including AC23. Additionally, the increased representation of repeat-rich regions and provision of curated gene annotations within them, is of high value to ongoing efforts to characterise and understand fungal effectors. We anticipate future studies will utilise the effector predictions provided in this study to confirm new novel *P. nodorum* effectors, and potentially discover a role for AC23 in promoting virulence.

### Methods

#### Genome sequencing and assembly

Genomic DNA of *P. nodorum* (*syn. Phaeosphaeria nodorum*, *Stagonospora nodorum*, *Leptosphaeria*.

*nodorum*, *Septoria nodorum*) strain Sn15 [20] – originally isolated in Western Australia by the Dept. Primary Industries and Regional Development (DPIRD: Agriculture & Food) – was sequenced via Pacific Biosciences P5-C3 chemistry with 4 SMRT cells, at the Génome Québec Innovation Centre (McGill University, Montreal, QC, Canada). The longest 25% of reads were self-corrected and assembled using Canu v1.0 (–pacbio-raw, expected genome size 39 Mbp) [13]. Assembly base-calls were corrected with Pilon v1.16 [80] using Illumina reads [20] which were mapped to the assembly with Bowtie2 v2.3.3.1 [81]. Mitochondrial contigs assembled by the above methods were identical to a previously published Sn15 mtDNA [GenBank: EU053989] [1] therefore the old mtDNA record was not updated by this new assembly..

Optical maps were used to order and orient the Canu-assembled Sn15 contigs into a complete genome. Sn15 protoplasts were extracted from hyphae as per Solomon et al [82], which was adjusted to 1x10e8 with GMB (0.125 M EDTA pH 8, 0.9 M sorbitol) at 42 °C. Protoplasts were added 1:1 to 1% low melt agarose (SeaPlaque GTG in 2% sorbitol and 50 mM EDTA) and poured into Plug Mold (Bio-Rad Laboratories, Munich, Germany) and set at 4 °C for 30 min. The plug was added to 5 ml Proteinase K solution (1 mg/ml Proteinase K, 100 mM EDTA pH 8.0, 0.2% Na deoxycholate, 10 mM Tris pH 8.0 and 1% N-lauroyl sarcosine) and incubated at 50 °C overnight, then added to sterile wash buffer (20 mM Tris pH 8, 50 mM EDTA pH 8) for 4 h changing the solution every hour. Clean plugs were transferred into 0.5 M EDTA at 9.5 pH and stored at 4 °C until shipment at room temperature. High molecular weight DNA was extracted from protoplasts as per Syme et al [23] and digested with *SpeI*, resulting in 63,440 fragments with an average size of ~ 315 Kbp. Optical maps were generated and manually curated with MapSolver™ (OpGen, MD, USA). Contig joins were made by inserting a 100 bp unknown (N) gap. Where the optical map indicated contig mis-assemblies, potential breakpoints were inspected for a localised drop in aligned read coverage. Chromosome scaffolds were numbered in descending size order based on the estimated physical lengths derived from the optical map. Chromosome scaffolds were assessed by Quast v5.2 [83], BUSCO 5.1.2 [84] and by coverage depth of alignments of raw and corrected SMRT reads by bwa-mem (0.7.17-r1188, –x pacbio) [85] via SAMtools [86] and BEDtools [87]. The assembled Sn15 chromosome scaffolds of this study were compared to previously published assembly versions (Supplementary

Table 1) [15, 23] with MUMmer 3.0 (nucmer -max-match, show-cords) [88] and alignments were visualised with Dot [89]. The SN15 reference genome data is available under BioProject: PRJNA686477. The updated SN15 genome assembly is deposited under [Genome: GCA\_016801405.1/ASM1680140v1] and [NUC: CP069023.1 - CP069045.1].

### Annotation of genome features

Repetitive DNA regions within the Sn15 genome assembly were analysed by three methods. The presence and overall proportion of AT-rich regions were calculated with OcculterCut v1.1 [26]. Annotation of repeat regions was performed using RepeatMasker 4.0.6 (sensitive mode, rmbblastn version 2.2.27+) [90] in four separate analyses, using: a) a published set of de novo repeats derived from a previous Sn15 genome assembly [24], b) RepBase (taxon “Fungi”) [91], c) LTRharvest of the GenomeTools suite [92], and d) a newly predicted set of de novo repeats generated by RepeatModeler v1.0.8 [93] (–engine ncbi -pa 15). Subsequent repeat analyses requiring a repeat-masked input used the output derived from the new de novo repeat dataset (Supplementary Table 2, Supplementary Table 3). Tandem repeats were predicted using Tandem Repeat Finder (Parameters: Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 180, MaxPeriod = 2000) [94]. Telomere regions were identified by terminal “TTAGGG” tandem repeats [95].

Protein-coding gene loci were annotated incorporating multiple transcriptome datasets from previous studies [23, 43]. RNA-seq reads from a prior study [43] were trimmed with cutadapt v.9.1 (paired end mode, --quality-cutoff = 30, --minimum-length 25, -n 3) [96] and de-duplicated with khmer v2.0 and screed v0.9 (normalize-by-median.py, -C 30 -M 100e9) [97, 98]. Fungal RNA-seq reads were derived from a mixed *in planta* library during early infection (3 dpi) when known effectors are maximally expressed [60]. Fungal reads were separated from wheat sequences with BBsplit v36.11 (BBmap, Seal v36.11, -Xmx200g) [99] to screen against the *Triticum aestivum* assembly TGAC v1.30 (GCA\_900067645.1) [100]. Filtered reads were mapped to the new Sn15 assembly with STAR v2.5.2b (align-Reads, --outSAMstrandField intronMotif --outFilterIntronMotifs) [101]. Transcript assembly was performed with Trinity v2.2.0 (--seqType fa --trinity\_complete --full\_cleanup --jaccard\_clip) [102]. RNAseq reads were aligned to the SN15 assembly with TopHat v2.2.6.0 (defaults) [103]. Relative expression was quantified with Cufflinks [104] and Stringtie v1.3.3b (params -m 50 -B -e -p 8) [105].

A final set of gene annotations was generated by combining annotations from multiple sources. Initial

transcriptome-based predictions were made with PASA v2.0.2 [106], incorporating: Trinity transcripts; open-reading frames generated with Transdecoder v2.0.1 [102]; CodingQuarry (CQ) v2.0 predictions based on TopHat outputs [42]; a second round of predictions generated using Coding-Quarry “Pathogen Mode” (CQPM) (*A. testa*, 2016) within regions between the initial CQ predictions. De novo prediction was performed with GeneMark-ES v4.32 (--ES, --fungus) [107]. Previously published gene annotations for Sn15 were aligned to the new assembly with AAT r03052011 [108] using CDS features (--dds '-f 100 -i 20 -o 75 -p 70 -a 2000' --filter '-c 10' --gap2 '-x 1') and protein sequences (--dps '-f 100 -i 30 -a 200' --filter '-c 10' --nap '-x 10'). All predicted gene annotations described above were assigned relative weight scores (AAT protein mapping 1, EST 5, AAT CDS mapping, GeneMark-ES 1, CQ/CQPM 10, transdecoder 10, PASA 9) and were then integrated into a single annotation set via EvidenceModeler (EVM) v1.1.1 [106]. Every locus of the EVM annotation set was manually curated using Webapollo [109] alongside supporting evidence from the various prediction methods described above, InterProScan domains aligned to the genome [110], aligned RNAseq reads [23, 43] and annotated repeat features (see above). New loci were manually annotated within intergenic regions if supported by RNAseq alignments. The resulting set of manually curated annotations were filtered (Table 2) for either: 1) orthologous best hit to the 13,690 gene models from the previously published annotation [20, 23] or; 2) coding regions (CDS) of >300 bp in length and with RNAseq read coverage of >50 fragments per kilobase of transcripts per million mapped reads (FPKM). This filtered set of manually curated genes is subsequently referred to as the primary gene set (Set A). The remaining predictions that failed this filter were retained as a secondary gene set (Set B) if CDS length was >90 bp and RNAseq depth was >5 FPKM or if homologous to a previously annotated gene [1, 20, 23]. To be consistent with previous publications on *P. nodorum* genomics, in this study gene annotations corresponding to loci that had been numbered in previous studies [1, 20, 23] retained their previous locus number despite non-sequential order along the new assembly. New annotations not corresponding to previously annotated loci were numbered from SNOG\_40000 onwards.

Various software and databases were used to assign functional annotations for both gene sets (A and B). OcculterCut v1.1 [26] predicted AT-rich regions and distances to the nearest AT-rich region for each locus. InterProScan (5.27–66.0) [110] was used to generate a broad range of functional annotations (InterPro, Pfam, Gene3D, Superfamily, MobiDB). PHibase v4.2 [48] was searched to assign homology to known effectors. SignalP

**Table 2** Criteria used to predict the primary gene prediction set (Set A) for *P. nodorum* Sn15, and the secondary set (Set B) which contains low-confidence gene predictions for the purpose of extracting a small proportion of strong effector candidates

Criteria	Set A (high confidence)	Set B (putative)
CDS length	> 300 bp / > 100 aa	(> 90 bp / > 30 aa)
Gene expression	AND	AND
Orthology/Homology	> 50 FPKM OR Orthology (reciprocal best hit) to previous annotation [20, 23]	> 5 FPKM OR Homology to previous annotation set [20, 23]

v4 [111]. was used to predict extracellular secretion, EffectorP v2.0 [47] was used to predict effector functions, and Localizer v1.0 (–e mode) [112] was used to predict potential host-cell sub-cellular localisation. The dbCAN r07/20/2017 [113] and AntiFam r3.0 databases were both searched using hmmsearch (--cut\_ga) [114] to predict carbohydrate-active enzymes (CAZymes) and pseudogenes respectively.

### Comparative pan-genomics

The new Sn15 genome reference was compared with draft assemblies of 18 *P. nodorum* isolates [20, 23] and 10 *P. avenae* isolates [BioProject: PRJNA476481], as well as long-read assemblies of 3 isolates of *P. nodorum*: Sn79–1087 [BioProject: PRJNA398070; Genome: GCA\_002267025.1], Sn4 [BioProject: PRJNA398070; Genome: GCA\_002267045.1] and Sn2000 [BioProject: PRJNA398070; Genome: GCA\_002267005.1] [15, 31] (Supplementary Table 1). Whole-genome alignments and variant calling was performed via MUMmer v3 (nucmer --max-match, show-coords -T -H -r) [88] and the percent coverage of matches of each isolate relative to Sn15 was calculated within 50 Kbp windows via BEDtools v2.26.0 (makewindows, coverage) [87]. PAVs were calculated from nucmer (delta-filter –1) alignments via BEDtools v2.26.0 (genomecov -bga) [87]. SNPs were calculated from nucmer alignments (show-snps -rLHTC) [88] and SNP density was calculated in 10 Kbp windows via BEDtools v2.26.0 (genomecov -bga) [87]. SNPs were analysed with SnpEff [115] relative to the new Sn15 gene annotations and the non-synonymous/synonymous mutation (Dn/Ds) ratio was calculated for every Sn15 gene both: 1) versus individual isolates and 2) averaged over all isolate comparisons. For visualisation by CIRCOS v 0.69–3 [116], Dn/Ds ratios were averaged across all genes within 50 Kbp windows via BEDtools v2.26.0 (map -c 4 -o mean) [87].

### Prediction of necrotrophic effector (NE) candidate gene loci

Putative effector genes were predicted based on the ranking of cumulative scores assigned from effector-associated gene or protein properties, as has been reported in previous studies [20, 23]. In this study the features (Supplementary Table 4) used to assign scores

were: predicted secretion by SignalP 4.1 [111] (1 point); molecular weight < 30 kDa (1 point); %cysteines > 4% (1 point); DN:DS > 1.5 (1 point); distance of 0–5 from an AT-rich region as predicted by OcculterCut [26] (1 point); EffectorP [47] score > 0.9 (1 point); an *in planta* to *in vitro* differential expression ratio (IP:IV) > 5 (2 points) or < 1 (–2 points); sub-telomeric location within genome (within 500 Kbp of sequence end, 2 points); presence (–3 points) or absence (3 points) of ortholog in low-virulence isolate Sn79–1087; presence of ortholog predicted as an effector candidate in US isolate Sn4 [31] (3 points); assigned an effector/toxin-like functional annotation (3 points) and; predicted lateral-gene transfer event with a fungal pathogen species [79] (3 points).

### Abbreviations

AC: Accessory chromosome; AR: Accessory region; BFB: Breakage-fusion bridge; EDTA: Ethylenediaminetetraacetic acid; IP: *In planta*; IV: *In vitro*; P5-C3: 5th generation polymerase, 3rd generation chemistry; PAV: Presence-absence variation; PFGE: Pulsed-field gel electrophoresis; RIP: Repeat-induced point mutation; SMRT: Single-molecule real-time; SNB: *Septoria nodorum* blotch; SNP: Single nucleotide polymorphism

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07699-8>.

**Additional file 1: Supplementary Figure 1.** Comparison of non-repetitive regions of accessory chromosome 23 (AC23, red) to other Sn15 chromosomes (black), indicating that it is not the product of duplication of a core, sister chromosome. The GC content of AC23 is indicated by the linear plot, and local repeat density is indicated in the heat map below (red). Nucleotide matches > 200 bp are indicated by grey arcs

**Additional file 2: Supplementary Figure 2.** Comparison of non-repetitive regions of accessory chromosome 23 (AC23, black) of > 100 bp in length (grey arcs) to *P. tritici-repentis* BFP chromosomes 1, 3, 4, and 11 (red), *P. tritici-repentis* M4 chromosomes 1, 3, 4, 6 and 10 (green), *Bipolaris maydis* scaffold 16 (blue) and *B. sorokiniana* chromosomes 2, 4, and 9. This comparison indicated a trend of telomeric proximity in the relative matching regions of related species.

**Additional file 3: Supplementary Figure 3.** Alignment of locally collinear blocks (LCBs) via Mauve, indicating large sections of similarity with structural rearrangements between Chromosome 4 of *P. nodorum* isolate Sn15 (top) and corresponding chromosomal sequences of isolates Sn4, Sn2000, and Sn79–1087, presented at the whole chromosome level (A) and within ~ 700–800 Kb of the telomere (B).

**Additional file 4: Supplementary Table 1.** Summary of draft (A) and high-quality (B) genome assemblies of *Parastagonospora* spp. alternate isolates used in this study for comparative genomics versus the Australian reference isolate Sn15.

**Additional file 5: Supplementary Table 2.** Comparison of repetitive sequence masking in the new *P. nodorum* Sn15 genome assembly using 3 different repeat libraries applied sequentially in 3 iterations.

**Additional file 6: Supplementary Table 3.** De novo repeat sequences predicted within the *Parastagonospora nodorum* Sn15 genome assembly.

**Additional file 7: Supplementary Table 4.** Summary of *Parastagonospora nodorum* gene properties and predicted effector candidate genes.

**Additional file 8: Supplementary Table 5.** Summary of assembled sequence lengths in the new *P. nodorum* Sn15 genome assembly, and estimates of their potential to be unresolved by PFGE comparing a 1% size error range to the size difference with the next longest sequence.

**Additional file 9: Supplementary Table 6.** Properties of selected large regions of the Sn15 assembly exhibiting presence absence variation (PAV) across the *Parastagonospora* population.

**Additional file 10: Supplementary Table 7.** Presence-absence variation (PAV) matrix for comparison of *Parastagonospora nodorum* Sn15 genes versus all other *Parastagonospora* spp. isolates included in this study.

**Additional file 11: Supplementary Table 8.** Summary of scaffold sequences from Syme et al. 2013 corresponding to chromosomes of new optical map-assisted long-read genome assembly for *Parastagonospora nodorum* Sn15

**Additional file 12: Supplementary Table 9.** Summary of genes and their functional annotations within the Chromosome 8 PAV region.

**Additional file 13: Supplementary Table 10.** Summary of average SNP density and DN/DS selection metrics across the *Parastagonospora* spp. population, relative to the *P. nodorum* Sn15 reference genome assembly. )

**Additional file 14: Supplementary Table 11.** Summary of gene and effector candidate gene distances from nearest AT-rich regions in the *P. nodorum* Sn15 assembly.

**Additional file 15: Supplementary Table 12.** Orthology comparison between gene predictions of *P. nodorum* Sn15 and Sn4, indicating Sn15 orthologs to Sn4 effector candidates under diversifying selection identified in Richards et al. 2019.

**Additional file 16: Supplementary Table 13.** Summary of gene content and functional annotation for *P. nodorum* Sn15 accessory chromosome 23 (AC23).

**Additional file 17: Supplementary Text 1.** Notes on the *P. nodorum* Sn15 genome assembly and the integration of optical mapping data.

#### Acknowledgements

We thank Prof. Richard Oliver for his expertise and guidance, and Ms. Julie Lawrence for her assistance with laboratory work relating to optical mapping.

#### Authors' contributions

KCT and JKH conceived the experiment and supervised the research. SB performed the majority of bioinformatics analysis, with assistance from DABJ and JKH. SB and JKH drafted the manuscript and JKH, HTP and KCT substantially revised the manuscript. All authors read and approved the manuscript.

#### Funding

This study was supported by the Centre for Crop and Disease Management, a joint initiative of Curtin University and the Grains Research and Development Corporation (Research Grant CUR00023). This research was undertaken with the assistance of resources and services from the Pawsey Supercomputing Centre and the National Computational Infrastructure (NCI), which is supported by the Australian Government (Research Grant Y95).

#### Availability of data and materials

The SN15 reference genome and protein data is available under BioProject: PRJNA686477. The updated SN15 genome assembly sequence consisting of 23 chromosomes is deposited under [Genome: GCA\_016801405.1/

ASM1680140v1; NUC: CP069023.1 - CP069045.1]. *P. nodorum* Sn15 transcriptome data was previously deposited under BioProject:PRJNA632579. Alternate isolate data for *P. nodorum* and *P. avenae* was previously deposited under BioProject:PRJNA47648. Alternate reference isolate data was obtained for Sn79–1087 from BioProject:PRJNA398070 / GCA\_002267025.1, for Sn4 from BioProject:PRJNA398070 / GCA\_002267045.1, and for Sn2000 from BioProject:PRJNA398070 / GCA\_002267005.1.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Centre for Crop & Disease Management, Curtin University, Perth, Australia.

<sup>2</sup>Curtin Institute for Computation, Curtin University, Perth, Australia.

Received: 22 October 2020 Accepted: 11 May 2021

Published online: 25 May 2021

#### References

- Hane JK, Lowe RG, Solomon PS, Tan K-C, Schoch CL, Spatafora JW, et al. Dothideomycete–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell*. 2007;19(11):3347–68. <https://doi.org/10.1105/tpc.107.052829>.
- Liu Z, Friesen TL, Ling H, Meinhardt SW, Oliver RP, Rasmussen JB, et al. The Tsn1–ToxA interaction in the wheat–*Stagonospora nodorum* pathosystem parallels that of the wheat–tan spot system. *Genome*. 2006;49(10):1265–73. <https://doi.org/10.1139/g06-088>.
- Liu Z, Zhang Z, Faris JD, Oliver RP, Syme R, McDonald MC, et al. The cysteine rich Necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring Snn1. *PLoS Pathog*. 2012;8(1):e1002467. <https://doi.org/10.1371/journal.ppat.1002467>.
- Liu Z, Faris JD, Oliver RP, Tan K-C, Solomon PS, McDonald MC, et al. SnTox3 acts in effector triggered susceptibility to induce disease on wheat carrying the Snn3 gene. *PLoS Pathog*. 2009;5(9):e1000581. <https://doi.org/10.1371/journal.ppat.1000581>.
- Abeyssekara NS, Friesen TL, Keller B, Faris JD. Identification and characterization of a novel host–toxin interaction in the wheat–*Stagonospora nodorum* pathosystem. *Theor Appl Genet*. 2009;120(1):117–26. <https://doi.org/10.1007/s00122-009-1163-6>.
- Friesen TL, Chu C, Xu SS, Faris JD. SnTox5–Snn5: a novel *Stagonospora nodorum* effector–wheat gene interaction and its relationship with the SnToxA–Tsn1 and SnTox3–Snn3–B1 interactions. *Mol Plant Pathol*. 2012; 13(9):1101–9. <https://doi.org/10.1111/j.1364-3703.2012.00819.x>.
- Friesen TL, Meinhardt SW, Faris JD. The *Stagonospora nodorum*–wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene-for-gene manner. *Plant J*. 2007;51(4):681–92. <https://doi.org/10.1111/j.1365-3113.X.2007.03166.x>.
- Gao Y, Faris J, Liu Z, Kim Y, Syme R, Oliver R, et al. Identification and characterization of the SnTox6–Snn6 interaction in the *Parastagonospora nodorum*–wheat pathosystem. *Mol Plant-Microbe Interact*. 2015;28(5):615–25. <https://doi.org/10.1094/MPMI-12-14-0396-R>.
- Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol*. 2015;66(1):513–45. <https://doi.org/10.1146/annurev-arplant-043014-114623>.
- McDonald MC, Solomon PS. Just the surface: advances in the discovery and characterization of necrotrophic wheat effectors. *Curr Opin Microbiol*. 2018; 46:14–8. <https://doi.org/10.1016/j.mib.2018.01.019>.
- Shi G, Friesen TL, Saini J, Xu SS, Rasmussen JB, Faris JD. The wheat Snn7 gene confers susceptibility on recognition of the *Parastagonospora nodorum* necrotrophic effector SnTox7. *Plant Genome*. 2015;8(2):1–10.
- van Dam P, Fokkens L, Ayukawa Y, van der Gragt M, ter Horst A, Brankovics B, et al. A mobile pathogenicity chromosome in *Fusarium oxysporum* for

- infection of multiple cucurbit species. *Sci Rep.* 2017;7(1):9042. <https://doi.org/10.1038/s41598-017-07995-y>.
13. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36. <https://doi.org/10.1101/gr.215087.116>.
  14. Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, et al. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics.* 2018;19(1):279. <https://doi.org/10.1186/s12864-018-4680-3>.
  15. Richards JK, Wyatt NA, Liu Z, Faris JD, Friesen TL. Reference quality genome assemblies of three *Parastagonospora nodorum* isolates differing in virulence on wheat. *G3.* 2018;8(2):393.
  16. Van Kan JA, Stassen JH, Mosbach A, Van Der Lee TA, Faino L, Farmer AD, et al. A gapless genome sequence of the fungus *Botrytis cinerea*. *Mol Plant Pathol.* 2017;18(1):75–89. <https://doi.org/10.1111/mpp.12384>.
  17. Wyatt NA, Richards JK, Brueggeman RS, Friesen TL. A comparative genomic analysis of the barley pathogen *Pyrenophora teres f. teres* identifies Subtelomeric regions as drivers of virulence. *Mol Plant-Microbe Interact.* 2020;33(2):173–88. <https://doi.org/10.1094/MPMI-05-19-0128-R>.
  18. Derbyshire M, Denton-Giles M, Hegedus D, Seifbarghy S, Rollins J, van Kan J, et al. The complete genome sequence of the phytopathogenic fungus *Sclerotinia sclerotiorum* reveals insights into the genome architecture of broad host range pathogens. *Genome Biol Evol.* 2017;9(3):593–618. <https://doi.org/10.1093/gbe/evx030>.
  19. Cooley RN, Caten CE. Variation in electrophoretic karyotype between strains of *Septoria nodorum*. *Mol Gen Genet.* 1991;228(1–2):17–23. <https://doi.org/10.1007/BF00282442>.
  20. Syme RA, Hane JK, Friesen TL, Oliver RP. Resequencing and comparative genomics of *Stagonospora nodorum*: sectional gene absence and effector discovery. *G3.* 2013;3(6):959–69.
  21. Bringans S, Hane JK, Casey T, Tan K-C, Lipscombe R, Solomon PS, et al. Deep proteogenomics; high throughput gene validation by multidimensional liquid chromatography and mass spectrometry of proteins from the fungal wheat pathogen *Stagonospora nodorum*. *BMC Bioinformatics.* 2009;10(1):301. <https://doi.org/10.1186/1471-2105-10-301>.
  22. Ipcho SV, Hane JK, Antoni EA, Ahren D, Henrissat B, Friesen TL, et al. Transcriptome analysis of *Stagonospora nodorum*: gene models, effectors, metabolism and pantothenate dispensability. *Mol Plant Pathol.* 2012;13(6):531–45. <https://doi.org/10.1111/j.1364-3703.2011.00770.x>.
  23. Syme RA, Tan K-C, Rybak K, Friesen TL, McDonald BA, Oliver RP, et al. Pan-*Parastagonospora* comparative genome analysis—effector prediction and genome evolution. *Genome Biol Evol.* 2018;10(9):2443–57. <https://doi.org/10.1093/gbe/evy192>.
  24. Hane JK, Oliver RP. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics.* 2008;9(1):478. <https://doi.org/10.1186/1471-2105-9-478>.
  25. Hane JK, Oliver RP. In silico reversal of repeat-induced point mutation (RIP) identifies the origins of repeat families and uncovers obscured duplicated genes. *BMC Genomics.* 2010;11(1):655. <https://doi.org/10.1186/1471-2164-11-655>.
  26. Testa AC, Oliver RP, Hane JK. OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. *Genome Biol Evol.* 2016;8(6):2044–64. <https://doi.org/10.1093/gbe/eww121>.
  27. Hane JK, Rouxel T, Howlett BJ, Kema GHJ, Goodwin SB, Oliver RP. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* 2011;12(5):R45. <https://doi.org/10.1186/gb-2011-12-5-r45>.
  28. McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, et al. Transposon-mediated horizontal transfer of the host-specific virulence protein ToxA between three fungal wheat pathogens. *mBio.* 2019;10(5):e01515.
  29. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* Fungi. *PLoS Pathog.* 2012;8(12):e1003037. <https://doi.org/10.1371/journal.ppat.1003037>.
  30. Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan K-C, Hane JK. Accessories make the outfit: accessory chromosomes and other dispensable DNA regions in plant-pathogenic Fungi. *Mol Plant-Microbe Interact.* 2018;31(8):779–88. <https://doi.org/10.1094/MPMI-06-17-0135-FI>.
  31. Richards JK, Stukenbrock EH, Carpenter J, Liu Z, Cowger C, Faris JD, et al. Local adaptation drives the diversification of effectors in the fungal wheat pathogen *Parastagonospora nodorum* in the United States. *PLoS Genet.* 2019;15(10):e1008223. <https://doi.org/10.1371/journal.pgen.1008223>.
  32. Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog.* 2012;8(4):e1002608. <https://doi.org/10.1371/journal.ppat.1002608>.
  33. Gardiner DM, Benfield AH, Stiller J, Stephen S, Aitken K, Liu C, et al. A high-resolution genetic map of the cereal crown rot pathogen *Fusarium pseudograminearum* provides a near-complete genome assembly. *Mol Plant Pathol.* 2018;19(1):217–26. <https://doi.org/10.1111/mpp.12519>.
  34. Plissonneau C, Hartmann FE, Croll D. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* 2018;16(1):1–16.
  35. Mendelowitz L, Pop M. Computational methods for optical mapping. *Gigascience.* 2014;3(1):33. <https://doi.org/10.1186/2047-217X-3-33>.
  36. Burkhardt AK, Childs KL, Wang J, Ramon ML, Martin FN. Assembly, annotation, and comparison of *Macrophomina phaseolina* isolates from strawberry and other hosts. *BMC Genomics.* 2019;20(1):802. <https://doi.org/10.1186/s12864-019-6168-1>.
  37. Farman ML. Telomeres in the rice blast fungus *Magnaporthe oryzae*: the world of the end as we know it. *FEMS Microbiol Lett.* 2007;273(2):125–32. <https://doi.org/10.1111/j.1574-6968.2007.00812.x>.
  38. Rehmeier C, Li W, Kusaba M, Kim Y-S, Brown D, Staben C, et al. Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Res.* 2006;34(17):4685–701. <https://doi.org/10.1093/nar/gkl588>.
  39. Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, et al. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet.* 2014;10(3):e1004227. <https://doi.org/10.1371/journal.pgen.1004227>.
  40. Tashiro S, Nishihara Y, Kugou K, Ohta K, Kanoh J. Subtelomeres constitute a safeguard for gene expression and chromosome homeostasis. *Nucleic Acids Res.* 2017;45(18):10333–49. <https://doi.org/10.1093/nar/gkx780>.
  41. Jones DA, Bertazzoni S, Turo CJ, Syme RA, Hane JK. Bioinformatic prediction of plant-pathogenicity effector proteins of fungi. *Curr Opin Microbiol.* 2018;46:43–9. <https://doi.org/10.1016/j.mib.2018.01.017>.
  42. Testa AC, Hane JK, Ellwood SR, Oliver RP. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics.* 2015;16(1):170. <https://doi.org/10.1186/s12864-015-1344-4>.
  43. Jones DAB, John E, Rybak K, Phan HTT, Singh KB, Lin S-Y, et al. A specific fungal transcription factor controls effector gene expression and orchestrates the establishment of the necrotrophic pathogen lifestyle on wheat. *Sci Rep.* 2019;9(1):15884. <https://doi.org/10.1038/s41598-019-52444-7>.
  44. Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, et al. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3.* 2013;3(1):41–63.
  45. McClintock B. The fusion of broken ends of chromosomes following nuclear fusion. *Proc Natl Acad Sci.* 1942;28(11):458–63. <https://doi.org/10.1073/pnas.28.11.458>.
  46. Ferris MM, Yan X, Habberset RC, Shou Y, Lemanski CL, Jett JH, et al. Performance assessment of DNA fragment sizing by high-sensitivity flow cytometry and pulsed-field gel electrophoresis. *J Clin Microbiol.* 2004;42(5):1965–76. <https://doi.org/10.1128/JCM.42.5.1965-1976.2004>.
  47. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol.* 2018;19(9):2094–110. <https://doi.org/10.1111/mpp.12682>.
  48. Urban M, Czucka A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base: the pathogen–host interactions database. *Nucleic Acids Res.* 2020;48(D1):D613–20. <https://doi.org/10.1093/nar/gkz904>.
  49. Mohanta TK, Bae H. The diversity of fungal genome. *Biol Proced Online.* 2015;17(1):8–8. <https://doi.org/10.1186/s12575-015-0020-z>.
  50. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput Biol.* 2014;10(12):e1003998. <https://doi.org/10.1371/journal.pcbi.1003998>.
  51. Magrini V, Gao X, Rosa BA, McGrath S, Zhang X, Hallsworth-Pepin K, et al. Improving eukaryotic genome annotation using single molecule mRNA sequencing. *BMC Genomics.* 2018;19(1):172. <https://doi.org/10.1186/s12864-018-4555-7>.

52. Syme RA, Martin A, Wyatt NA, Lawrence JA, Muria-Gonzalez MJ, Friesen TL, et al. Transposable element genomic fissuring in *Pyrenophora teres* is associated with genome expansion and dynamics of host-pathogen genetic interactions. *Front Genet.* 2018;9:130. <https://doi.org/10.3389/fgene.2018.00130>.
53. Zolan ME. Chromosome-length polymorphism in fungi. *Microbiol Rev.* 1995; 59(4):686–98. <https://doi.org/10.1128/MR.59.4.686-698.1995>.
54. Mossessova E, Lima CD. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell.* 2000;5(5):865–76. [https://doi.org/10.1016/S1097-2765\(00\)80326-3](https://doi.org/10.1016/S1097-2765(00)80326-3).
55. Nagano N, Umemura M, Izumikawa M, Kawano J, Ishii T, Kikuchi M, et al. Class of cyclic ribosomal peptide synthetic genes in filamentous fungi. *Fungal Genet Biol.* 2016;86:58–70. <https://doi.org/10.1016/j.fgb.2015.12.010>.
56. Stahelin RV, Long F, Diravivam K, Bruzik KS, Murray D, Cho W. Phosphatidylinositol 3-phosphate induces the membrane penetration of the FYVE domains of Vps27p and Hrs. *J Biol Chem.* 2002;277(29):26379–88. <https://doi.org/10.1074/jbc.M201106200>.
57. Phan HTT, Rybak K, Bertazzoni S, Furuki E, Dinglasan E, Hickey LT, et al. Novel sources of resistance to *Septoria nodorum* blotch in the Vavilov wheat collection identified by genome-wide association studies. *Theor Appl Genet.* 2018;131(6):1223–38. <https://doi.org/10.1007/s00122-018-3073-y>.
58. Tan K-C, Phan HTT, Rybak K, John E, Chooi YH, Solomon PS, et al. Functional redundancy of necrotrophic effectors – consequences for exploitation for breeding. *Front Plant Sci.* 2015;6:501.
59. Phan HTT, Rybak K, Furuki E, Breen S, Solomon PS, Oliver RP, et al. Differential effector gene expression underpins epistasis in a plant fungal disease. *Plant J.* 2016;87(4):343–54. <https://doi.org/10.1111/tpj.13203>.
60. Rybak K, See PT, Phan HT, Syme RA, Moffat CS, Oliver RP, et al. A functionally conserved Zn (2) Cys (6) binuclear cluster transcription factor class regulates necrotrophic effector gene expression and host-specific virulence of two major Pleosporales fungal pathogens of wheat. *Mol Plant Pathol.* 2017;18(3):420–34. <https://doi.org/10.1111/mpp.12511>.
61. Zhu W, Wei W, Wu Y, Zhou Y, Peng F, Zhang S, et al. BcCFEM1, a CFEM domain-containing protein with putative GPI-anchored site, is involved in pathogenicity, conidial production, and stress tolerance in *Botrytis cinerea*. *Front Microbiol.* 2017;8:1807. <https://doi.org/10.3389/fmicb.2017.01807>.
62. Kombrink A, Thomma BPHJ. LysM effectors: secreted proteins supporting fungal life. *PLoS Pathog.* 2013;9(12):e1003769. <https://doi.org/10.1371/journal.ppat.1003769>.
63. Zhang J, Wang F, Liang F, Zhang Y, Ma L, Wang H, et al. Functional analysis of a pathogenesis-related thaumatin-like protein gene TaLR35PR5 from wheat induced by leaf rust fungus. *BMC Plant Biol.* 2018;18(1):76. <https://doi.org/10.1186/s12870-018-1297-2>.
64. de Guillen K, Lorrain C, Tsan P, Barthe P, Petre B, Saveleva N, et al. Structural genomics applied to the rust fungus *Melampsora larici-Populina* reveals two candidate effector proteins adopting cystine knot and NTF2-like protein folds. *Sci Rep.* 2019;9(1):18084. <https://doi.org/10.1038/s41598-019-53816-9>.
65. Burke JE, Dennis EA. Phospholipase A2 structure/function, mechanism, and signaling. *J Lipid Res.* 2009;50(Suppl (Suppl)):S237–42.
66. Chruszcz M, Chapman MD, Osinski T, Solberg R, Demas M, Porebski PJ, et al. *Alternaria alternata* allergen Alt a 1: a unique  $\beta$ -barrel protein dimer found exclusively in fungi. *J Allergy Clin Immunol.* 2012;130(1):241–247.e249.
67. Zhang Y, Gao Y, Liang Y, Dong Y, Yang X, Qiu D. *Verticillium dahliae* PevD1, an Alt a 1-like protein, targets cotton PR5-like protein and promotes fungal infection. *J Exp Bot.* 2019;70(2):613–26. <https://doi.org/10.1093/jxb/ery351>.
68. Olombrada M, Peña C, Rodríguez-Galán O, Klingauf-Nerurkar P, Portugal-Calisto D, Oborská-Oplová M, et al. The ribotoxin  $\alpha$ -sarcin can cleave the sarcin/ricin loop on late 60S pre-ribosomes. *Nucleic Acids Res.* 2020;48(11): 6210–22. <https://doi.org/10.1093/nar/gkaa315>.
69. Baccelli I. Cerato-platanin family proteins: one function for multiple biological roles? *Front Plant Sci.* 2015;5:769.
70. Antuch W, Güntert P, Wüthrich K. Ancestral  $\beta$ -crystallin precursor structure in a yeast killer toxin. *Nat Struct Biol.* 1996;3(8):662–5. <https://doi.org/10.1038/nsb0896-662>.
71. Figueiredo J, Sousa Silva M, Figueiredo A. Subtilisin-like proteases in plant defence: the past, the present and beyond. *Mol Plant Pathol.* 2018;19(4): 1017–28. <https://doi.org/10.1111/mpp.12567>.
72. Singh K, Winter M, Zouhar M, Ryšánek P. Cyclophilins: less studied proteins with critical roles in pathogenesis. *Phytopathology.* 2017;108(1):6–14. <https://doi.org/10.1094/PHYTO-05-17-0167-RWW>.
73. Viaud MC, Balhadère PV, Talbot NJ. A Magnaporthe grisea cyclophilin acts as a virulence determinant during plant infection. *Plant Cell.* 2002;14(4):917–30. <https://doi.org/10.1105/tpc.010389>.
74. Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics.* 2013;14(1):274. <https://doi.org/10.1186/1471-2164-14-274>.
75. Grell MN, Mouritzen P, Giese H. A *Blumeria graminis* gene family encoding proteins with a C-terminal variable region with homologues in pathogenic fungi. *Gene.* 2003;311:181–92. [https://doi.org/10.1016/S0378-1119\(03\)00610-3](https://doi.org/10.1016/S0378-1119(03)00610-3).
76. Schäfer W. The role of cutinase in fungal pathogenicity. *Trends Microbiol.* 1993;1(2):69–71. [https://doi.org/10.1016/0966-842X\(93\)90037-R](https://doi.org/10.1016/0966-842X(93)90037-R).
77. Zhang M, Xie S, Zhao Y, Meng X, Song L, Feng H, et al. Hce2 domain-containing effectors contribute to the full virulence of *Valsa mali* in a redundant manner. *Mol Plant Pathol.* 2019;20(6):843–56. <https://doi.org/10.1111/mpp.12796>.
78. Temple B, Horgen PA, Bernier L, Hintz WE. Cerato-ulmin, a hydrophobin secreted by the causal agents of Dutch elm disease, is a parasitic fitness factor. *Fungal Genet Biol.* 1997;22(1):39–53. <https://doi.org/10.1006/fgbi.1997.0991>.
79. Hane JK. Effector-like ROG Predictions. <https://effectordb.com/Igt-effector-predictions-summary/> Accessed 20 Oct 2020.
80. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014;9(11):e112963. <https://doi.org/10.1371/journal.pone.0112963>.
81. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
82. Solomon PS, Thomas SW, Spanu P, Oliver RP. The utilisation of di/tripeptides by *Stagonospora nodorum* is dispensable for wheat infection. *Physiol Mol Plant Pathol.* 2003;63(4):191–9. <https://doi.org/10.1016/j.pmp.2003.12.003>.
83. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086>.
84. Seppy M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol.* 1962;2019:227–45.
85. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997*; 2013.
86. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16): 2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
87. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
88. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
89. Dot. <https://github.com/dnanexus/dot>. Accessed 20 Oct 2020.
90. RepeatMasker Open-4.0. <http://www.repeatmasker.org>. Accessed 20 Oct 2020.
91. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015;6(1):11. <https://doi.org/10.1186/s13100-015-0041-9>.
92. Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform.* 2013;10(3):645–56. <https://doi.org/10.1109/TCBB.2013.68>.
93. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci.* 2020;117(17):9451–7. <https://doi.org/10.1073/pnas.1921046117>.
94. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>.
95. Qi X, Li Y, Honda S, Hoffmann S, Marz M, Mosig A, et al. The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Res.* 2013;41(1):450–62. <https://doi.org/10.1093/nar/gks980>.
96. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):3.

97. Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv preprint arXiv:12034802; 2012.
98. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, et al. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res*. 2015;4:900.
99. Bushnell B. *BBMap: a fast, accurate, splice-aware aligner*. Berkeley: Lawrence Berkeley National Lab.(LBNL); 2014.
100. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res*. 2017;27(5):885–96. <https://doi.org/10.1101/gr.217117.116>.
101. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
102. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494–512. <https://doi.org/10.1038/nprot.2013.084>.
103. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
104. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc*. 2012;7(3):562–78. <https://doi.org/10.1038/nprot.2012.016>.
105. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650–67. <https://doi.org/10.1038/nprot.2016.095>.
106. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol*. 2008;9(1):R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
107. Ter-Hovhannisyantsyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008;18(12):1979–90. <https://doi.org/10.1101/gr.081612.108>.
108. Huang X, Adams MD, Zhou H, Kerlavage AR. A tool for analyzing and annotating genomic sequences. *Genomics*. 1997;46(1):37–45. <https://doi.org/10.1006/geno.1997.4984>.
109. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, et al. Web Apollo: a web-based genomic annotation editing platform. *Genome Biol*. 2013;14(8):R93. <https://doi.org/10.1186/gb-2013-14-8-r93>.
110. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40. <https://doi.org/10.1093/bioinformatics/btu031>.
111. Nielsen H. Predicting secretory proteins with SignalP. In: *Protein Function Prediction*: Springer; 2017. p. 59–73.
112. Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci Rep*. 2017;7(1):1–14.
113. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012;40(W1):W445–51. <https://doi.org/10.1093/nar/gks479>.
114. Eberhardt RY, Haft DH, Punta M, Martin M, O'Donovan C, Bateman A. AntiFam: a tool to help identify spurious ORFs in protein annotation Database 2012; 2012.
115. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695>.
116. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45. <https://doi.org/10.1101/gr.092759.109>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

