A global pangenome for the wheat fungal pathogen *Pyrenophora tritici-repentis* and prediction of effector protein structural homology

Paula Moolhuijzen[1*], Pao Theen See[1], Gongjun Shi[2], Harold R. Powell[3], James Cockram[4], Lise N. Jørgensen[5], Hamida Benslimane[6], Stephen E. Strelkov[7], Judith Turner[8], Zhaohui Liu[2*] and Caroline S. Moffat[1]

[1] Centre for Crop Disease and Management, School of Molecular and Life Sciences, Curtin University, Bentley, Western Australia, Australia

[2] Department of Plant Pathology, North Dakota State University, Fargo, North Dakota, USA

[3] Department of Life Sciences, Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, England, United Kingdom

[4] NIAB, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, United Kingdom

[5] Department of Agroecology, Aarhus University, Slagelse, Denmark

[6] Département de Botanique, Ecole Nationale Supérieure Agronomique (ENSA), Hassan Badi, El-Harrach, Algiers, Algeria

[7] Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada

[8] Fera Science Ltd., York, YO41 1LZ, United Kingdom

* Correspondence: Zhaohui Liu, zhh.liu@ndsu.edu * Correspondence: Paula Moolhuijzen, Paula.Moolhuijzen@curtin.edu.au

Keywords: Necrotrophic fungi, plant pathogen, toxin

Repositories: The sequenced genomes have been deposited in the DDBJ/ENA/GenBank under accession numbers JAAFOX000000000, JAHCSW000000000, JAHCYZ000000000, NRDI02000000, PSOO00000000-PSOU00000000 and RXHK00000000-RXHN00000000.

Abstract

The adaptive potential of plant fungal pathogens is largely governed by the gene content of a species, comprised of core and ancillary genes across the pathogen isolate repertoire. To approximate the complete gene repertoire of a globally significant crop fungal pathogen, a pan genomic analysis was undertaken for *Pyrenophora tritici-repentis* (Ptr), the causal agent of tan (or yellow) spot disease in wheat.

In this study, fifteen new Ptr genomes were sequenced, assembled and annotated, including isolates from three races not previously sequenced. Together with eleven previously published Ptr genomes, a pangenome for twenty-six Ptr isolates from Australia, Europe, North Africa and America, representing nearly all known races, revealed a conserved core-gene content of 56% and presents a new Ptr resource for searching natural homologues using remote protein structural homology. Here, we identify for the first time a nonsynonymous mutation in the Ptr effector gene *ToxB*, multiple copies of *toxb*, a distant natural *Pyrenophora* homologue of a known *Parastagonopora nodorum* effector, and clear genomic break points for the *ToxA* effector horizontal transfer region.

This comprehensive genomic analysis of Ptr races includes nine isolates sequenced via long read technologies. Accordingly, these resources provide a more complete representation of the species, and serve as a resource to monitor variations potentially involved in pathogenicity.

Author Notes

All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary data files and fifteen supplementary figures are available with the online version of this article.

Impact Statement

Our *Pyrenophora tritici-repentis* (Ptr) pangenome study provides resources and analyses for the identification of pathogen virulence factors, of high importance to microbial research. Key findings include: 1) Analysis of eleven new sequenced (with three new races not previously available) and previously published isolates, 26 genomes in total, representing the near complete Ptr race set for known effector production collected from Australia, Europe, North Africa and the Americas. 2) We show that although Ptr has low core gene conservation, the whole genome divergence of other wheat pathogens was greater. 3) The new PacBio sequenced genomes provide unambiguous genomic break points for the large *ToxA* effector horizontal transfer region, which is only present in ToxA producing races. 4) A new web-based Ptr resource for searching *in silico* remote protein structural homology is presented, and a distant natural *Pyrenophora* protein homologue of a known effector from another wheat pathogen is identified for the first time.

Data Summary

The sources and genomic sequences used throughout this study have been deposited in the National Centre for Biotechnology Information (NCBI), under the assembly accession numbers provided in Tables 1 and 2 (available in the online version of this article). The new M4 resource for protein structural homology is freely available through the BackPhyre web-portal URL, http://www.sbg.bio.ic.ac.uk/phyre2/.

Introduction

Tan (or yellow) spot, caused by the necrotrophic fungal pathogen *Pyrenophora tritici-repentis* [(Died.) Drechs.] (abbreviated to Ptr), can occur on both bread wheat (*Triticum aestivum* L.) and durum wheat (*T. turgidum* subsp. *durum* L.). A globally significant disease of economic importance (Murray GM and Brennan JP 2009, Benslimane, Lamari et al. 2011), tan spot can reduce crop production with up to 31% yield losses reported (Bhathal, Loughman et al. 2003).

During infection, necrotrophic fungi secrete necrotrophic effectors (NEs) that interact with the corresponding sensitivity genes in the host wheat lines (Friesen, Zhang et al. 2008, Ciuffetti, Manning et al. 2010, Faris, Liu et al. 2013, Downie, Lin et al. 2021). To date, Ptr has three known NEs, Ptr ToxA, Ptr ToxB and Ptr ToxC that produce either necrosis or chlorosis symptoms on their sensitive wheat genotypes (Ciuffetti, Tuori et al. 1997, Strelkov, Lamari et al. 1999). It is the different combinations (and absence) of these NEs that have been used to define different Ptr races (Lamari and Strelkov 2010, Faris, Liu et al. 2013), including race 1 (Ptr ToxA and Ptr ToxC), race 2 (Ptr ToxA), race 3 (Ptr ToxC), race 4 (no Ptr ToxA, Ptr ToxB or Ptr ToxC), race 5 (Ptr ToxB), race 6 (Ptr ToxB and Ptr ToxC), race 7 (Ptr ToxA and Ptr ToxB) and race 8 (Ptr ToxA, Ptr ToxB and Ptr ToxC). However, there are reports of isolates beyond the current classification (Ali S, Gurung S et al. 2010, Benslimane, Lamari et al. 2011, Kamel, Cherif et al. 2019). AR CrossB10 from North Dakota, USA, was such an isolate that produces both Ptr ToxC with an unknown effector, which has been recently sequenced (Kariyawasam, Wyatt et al. 2021) and will subsequently be referred to here as "race unknown" (Moolhuijzen, See et al. 2018). In addition to the three known effectors, the presence of novel NEs has been suggested in several studies (Tuori, Wolpert et al. 1995, Andrie, Pandelova et al. 2007, Ali S, Gurung S et al. 2010, Rybak, See et al. 2017, See, Marathamuthu et al. 2018). These findings make the sequencing of new isolates a priority to capture and understand the complete gene repertoire for Ptr.

Genome sequencing projects for fungal pathogens using single molecule long reads, such as PacBio and Oxford Nanopore technologies, have significantly improved our understanding of pathogen genomes, as they allow near complete genome assembly. In particular, the wheat fungal pathogens *Fusarium graminearum* (cause of fusarium head blight), Ptr, *Parastagonospora nodrum* (Sn, cause of Septoria nodorum blotch) and *Zymoseptoria tritici* (Zt, cause of Septoria tritici blotch) are known to have highly variable genomes characterised by gene loss and duplication events as well as large-scale genome rearrangements (Manning, Pandelova et al. 2013, Richards, Wyatt et al. 2017, Moolhuijzen, See et al. 2018, Badet, Oggenfuss et al. 2020, Alouane, Rimbert et al. 2021, Bertazzoni, Jones et al. 2021). To understand the genome composition of a species, the protein-coding genes from all available isolates are clustered based on the sequence identity of conserved protein domains into core (genes shared by all isolates) and ancillary/accessory (genes absent in one or more isolates) groups. The union of the core and accessory groups for the collection of isolates is then referred to as the pangenome, which is larger than the genome of any one individual (Vernikos, Medini et al. 2015). Depending on the number of sequenced isolates, associations to distinct habitats and phenotypes may then be detected within a pathogen species (Vernikos, Medini et al. 2015).

In this study, fifteen new Ptr isolates collected from Europe (Denmark, Germany and the United Kingdom), North Africa (Algeria and Tunisia) and the Americas (Brazil, Canada and the USA) were sequenced, assembled and annotated, for comparative analysis with eleven previously published Australian and North American Ptr isolates (Manning, Pandelova et al. 2013, Moolhuijzen, See et al. 2018, Moolhuijzen, See et al. 2019). A total of 26 annotated Ptr genomes, which represent nearly all known Ptr races, are presented here for a pangenome analysis to determine whole genome phylogeny and sequence variations in relation to core and ancillary genes. Ptr proteins are then further explored *in silico* to identify remote natural structural homology between different necrotrophic fungal species (not acquired by a horizontal gene transfer).

Materials and methods

Isolate collection and DNA extraction

Ptr isolates were collected from Algeria (Alg130 and Alg215), Brazil (Biotrigo9-1), Canada (90-2), Denmark (EW306-2-1, EW4-4, and EW7m1), Germany (SN001A, SN001C and SN002B), USA (86-124 and Ls13-192), United Kingdom (CC142) and Tunisia (T199 and T205). All isolates were collected from bread wheat (*T. aestivum* L.), except Alg215 which was collected from durum wheat (*T. turgidum* subsp. *durum* L.). Fungi were grown on V8-PDA agar according to (Moffat, See et al. 2014). Genomic DNA was extracted using a BioSprint 15 DNA Plant Kit (Qiagen, Hilden, Germany) with some modifications. Briefly, DNA was extracted using the BioSprint 15 automated workstation, according to the manufacturer’s instructions, from 3-day old mycelia grown in Fries 3 medium (Moffat, See et al. 2014). DNA was further treated with 50 µg/ml of RNase enzyme (Qiagen, Hilden, Germany) for 1 h followed by phenol/chloroform extraction, precipitation with sodium acetate and ethanol, and finally resuspension in Tris-EDTA buffer.

Isolate pathotyping

Ptr isolates were pathotyped for race classification through infection assays of differential wheat genotypes differing in their specific effector sensitivities. The wheat genotypes used were Glenlea (Ptr ToxA-sensitive), 6B662 (Ptr ToxB-sensitive), 6B365 (Ptr ToxC-sensitive) and Auburn or Salamouni (insensitive to all three effectors).

Two-week-old wheat (*T. aestivum* L.) seedlings were inoculated by spraying conidia onto the whole plants evenly at a rate of 3,000 conidia/ml and grown at 20 °C under a 12-h day/night cycle in a controlled growth chamber (Moffat, See et al. 2014). The second leaves were harvested 7-days post-inoculation, visually inspected for symptoms (Lamari, Sayoud et al. 1995) and photographed. The inoculation experiments were repeated twice with three replicate plants per wheat genotype.

Ptr isolate sequencing and genome assembly

Genomic DNA from four Ptr isolates was sequenced using the PacBio Sequel system, 90-2 (Novogene, China), Biotrigo9-1 (Novogene, USA), Ls13-192 and 86-124 (Mayo Clinic, Minnesota, USA). Error correction and *de novo* genome assembly of PacBio reads was completed with Canu version v2.1.1 (Koren, Walenz et al. 2017) with the following options (genomeSize=43, useGrid=TRUE, maxThreads=28, merylThreads=28, ovlThreads=28 ovlMerThreshold=500 and gridOptionsOBTOVL="--cpus-per-task=28) on computer resources (Broadwell Intel Xeon cores, 100 Gb/s Omni-Path interconnect and 128GB of memory per compute node) at Pawsey Supercomputing Centre, Perth, Western Australia. Previously generated Illumina 150 bp paired end DNA sequence reads of 86-124 genomic DNA (Moolhuijzen, See et al. 2018) and Biotrigo9-1 Illumina sequence (this study) were aligned to the contigs using BWA V0.7.17-r1188 (Li H and Durbin R 2009), and the sorted alignment bam files then used for further base error corrections using Pilon v1.24 (Walker, Abeel et al. 2014).

The genomic DNA for an additional 11 Ptr isolates (EW306-2-1, EW4-4, and EW7m1, SN001A, SN001C, SN002B, CC142, Alg130, Alg215, T199 and T205) was sequenced using Illumina Hi-Seq 150 bp pair-end reads by the Australian Genome Research Facility (AGRF). Isolate sequence data was quality checked with FASTQC (Andrews 2011), trimmed for poor quality, ambiguous bases and adapters using Skewer (Jiang, Lei et al. 2014) and Trimmomatic v0.22 (Bolger, Lohse et al. 2014) with a read head crop of 6 bp and minimum length of 100 bp. *De novo* genome assembly was undertaken using SPAdes version v3.10.0 (Bankevich, Nurk et al. 2012).

Gene prediction and functional annotation

Ptr sequenced genomes were soft masked for low complexity, as well as known transposable elements, using RepeatMasker (RM) (Chen 2004) v. open-4.0.6 with rmblastn version 2.2.27+ on RepBase (Kohany O, Gentles AJ et al. 2006) RM database version 20150807 (taxon=fungi). *Ab-initio* gene predictions were made with GeneMark-ES v4.33 (--ES --fungus --cores 16) (Borodovsky and Lomsadze 2011) and CodingQuarry v1.2 Pathogen Mode (PM) (Testa, Hane et al. 2015), assisted by RNA-Seq (Moolhuijzen, See et al. 2018) genome alignments using TopHat2 (Kim and Salzberg 2011) for a minimum intron size of 10 bp. The Ptr M4 and Pt-1C-BFP reference proteins (Manning, Pandelova et al. 2013, Moolhuijzen, See et al. 2018) were aligned using Exonerate v2.2.0 (--minintron 10 --maxintron 3000) protein2genome mode (Slater and Birney 2005). Gene annotations were assigned from BLASTX (v2.2.26) (Shiryev, Papadopoulos et al. 2007) searches against Uniref90 (Oct 13, 2020), NCBI Refseq (taxon=Pezizomycotina) (Oct 13, 2020) and InterProScan v5.17-56 (Quevillon, Silventoinen et al. 2005) protein databases. Sequence domains were assigned by RPS-BLAST (v2.2.26) against Pfam v33.1, Smart v6.0 and CDD v3.19 databases. The blast protein and domain searches were then summarised using AutoFACT v3.4 (Koski LB, Gray MW et al. 2005). Proteins were screened for a signal peptide using SignalP v5.0b (Petersen TN, Brunak S et al. 2011). Effector predictions were conducted on proteins with signal peptides using EffectorP v3.0 (Sperschneider, Dodds et al. 2018, Sperschneider and Dodds 2021). To ensure the same prediction methods were used for comparative analyses, SignalP V5.0b and EffectorP v3.0 (Sperschneider, Dodds et al. 2018, Sperschneider and Dodds 2021) were used to update the effector gene predictions on all the publicly available isolate genomes (Supplementary data 1). All predicted proteins were also ranked using Predector v1.1 (Jones, Rozano et al. 2021) (Supplementary data 1). Gene completeness was accessed using BUSCO v3, lineage fungi (Seppey, Manni et al. 2019).

Comparative genomics

To conduct comparative analyses across the Class Ascomycota, publicly available isolate genomes were downloaded from the National Centre for Biotechnology Information (NCBI) GenBank. These included *Bipolaris* (*B. cookei*, *B. maydis*, *B. sorokiniana*, *B. zeicola*), *Leptosphaeria* (*L. maculans*), *Parastagonospora* (*P. nodorum*), *Pyrenophora* (*Pyrenophora teres* f. *teres*, *Pyrenophora teres* f. *maculata*, *Pyrenophora seminiperda*) and *Zymoseptoria* (*Z. tritici*) (Syme, Tan et al. 2016, McDonald, Ahren et al. 2017, Richards, Wyatt et al. 2017, Syme, Martin et al. 2018, Badet, Oggenfuss et al. 2020) (Supplementary data 2). The published genomes of *P. tritici-repentis* isolates Pt-1C-BFP, DW5, DW7, SD20 (Manning, Pandelova et al. 2013), Ptr134, Ptr239, Ptr11137, Ptr5213, M4, 86-124 (Moolhuijzen, See et al. 2018), AR CrossB10 (Kariyawasam, Wyatt et al. 2021) and V1 (Moolhuijzen, See et al. 2019) were also included for analysis.

Genome nucleotide pairwise distance was calculated with Phylonium v1.5 (Klotzl and Haubold 2020) with two-pass enabled and 100 bootstrap matrices. Whole genome phylogenetic trees were constructed using Phylip 1:3.695-1 (Retief 2000), consensus program v3.695 on 100 Kitsch and neighbour joining v3.695 trees. The tree was then visualised using FigTree v1.4.4. Genomic nucleotide regions were compared between isolates using NUCmer v3.1 (Delcher, Salzberg et al. 2003) and Easyfig v 2.2.3 (Sullivan, Petty et al. 2011).

To determine the presence, copy number and percent identity of all genes in Ptr, the gene nucleotide sequences from all 26 isolates were aligned to all 26 genomes using GMAP version 2021-05-27 with options "-f 2 -t 48 -n 300 --max-intronlength-middle=1000 --max-intronlength-ends=1000 --fulllength --trim-end-exons=0 --alt-start-codons --canonical-mode=1 -- --max-deletionlength=20". Isolate mRNA Pearson correlations and predicted effector protein lengths and scores were analysed using R v4.0.3 (Team" 2021) using the R packages corrplot v0.84, ggplot2 v3.3.3, ggridges v0.5.3 and pheatmap v1.0.12. The analysis and data are available in v1.3.1093 RStudio (RStudio-Team 2020) markdown notebook https://github.com/ccdmb/PTR-60.

Isolate reads were aligned to the isolate M4 reference genome using BWA 0.7.14-r1138 and coverage (10 kb windows) was calculated using BedTools (genomecov) v2.17.0 on SamTools v 0.1.19-96b5f2294a sorted bam files. Regions of absence were then plotted using Circos v0.69-3 and R v3.5.1, bioconductor package chromPlot v1.10.0.

Protein orthologous clustering and effector analysis

Predicted protein data for all available Ptr isolates were clustered using OrthoFinder v2.5.2 (Emms and Kelly 2015). The predicted effector groups (with signal peptides) were then screened for three-dimensional (3-D) protein model predictions using the Protein Homology/analogY Recognition Engine V 2.0 Phyre2 (Kelley, Mezulis et al. 2015) batch processing mode. The predicted models were superimposed on the best ranked template to find the largest subset of atoms within an approximate threshold of 3.5 Å, which was adjusted based on the size of the aligned proteins using iMol (Rotkiewicz 2007). Protein sequences with high confidence (Phyre^2 ≥ 90%) predicted 3-D protein models were also searched against the Plant Host Interactions database (PHI-base) of known pathogenic phenotypes (Urban, Cuzick et al. 2017), at an expected value threshold of ≤ 1e-10 for significant alignments. Hidden Markov Model (HMM) libraries were created for the whole genome of Ptr isolate M4, which has been made publicly available through the online resource BackPhyre, Imperial College, London (Kelley, Mezulis et al. 2015).

Results

PacBio genome sequencing, assembly and annotation of four *P. tritici-repentis* isolates

A total of four Ptr genomes comprising two race 4 (lack all three known Ptr effectors) isolates (North Dakota (USA) isolate Ls13-192 (Guo, Shi et al. 2020) and Canadian isolate 90-2 (Lamari, Gilbert et al. 1998)), and two race 2 (producing Ptr ToxA only) isolates (Brazilian isolate Biotrigo9-1 (Bertagnolli, Ferreira et al. 2019) and Canadian isolate 86-124 (Lamari and Bernier 1989)) were sequenced using PacBio technology, assembled and protein-coding genes were predicted for comparative analysis.

The assembled Ptr genomes ranged in size from 37.56 Mb to 42.19 Mb (Table 1) and of these, the known effector producing isolates (86-124 and Biotrigo9-1) had a size comparable to previously PacBio sequenced genomes (M4 and DW5) (Moolhuijzen, See et al. 2018, Moolhuijzen, See et al. 2020). The race 4 isolate not producing known effectors, Ls13-192, had the smallest genome size at 37.56 Mb, at least 2 Mb smaller than all the known effector-producing isolate genomes, but similar to Pt-1C-BFP which was sequenced prior to the availability of third generation long read technologies and which lacks some representation of repeat/complex genomic regions (Manning, Pandelova et al. 2013). Our four new assemblies were more fragmented than the previously assembled genomes M4 and DW5 (Moolhuijzen, See et al. 2018, Moolhuijzen, See et al. 2020). In particular race 4 isolate 90-2 was fragmented into 162 contigs, over twice as many contigs as compared to race 2 isolate Biotrigo9-1 and race 4 isolate Ls13-192. The four genome assemblies had a BUSCO quantitative assessment greater than 98.9% for completeness with respect to gene content (Supplementary Fig. S1).

The number of predicted protein-encoding genes for our new PacBio sequenced genome assemblies ranged from 12,816 (Ls13-192) to 14,394 (Biotrigo9-1) (Table 1). The number of predicted protein effectors for race 2 isolates 86-124 and Biotrigo9-1 and race 4 isolate Ls13-192 was lower than the numbers predicted for race 1 isolate M4 and race 5 isolate DW5. However, race 4 isolate 90-2 had the highest number of proteins predicted as effectors, due to a higher gene copy number identified later in the protein clustering analysis. Furthermore, in the race 4 isolates a single *toxb* (found in non-pathogenic Ptr isolates and having no toxic activity (Kim and Strelkov 2007, Figueroa Betts, Manning et al. 2011) was detected in Ls13-192 on contig 4 (113,627-113,893 bp) and an exact *toxb* duplication event was detected in 90-2 on contig 37 (termed here *toxb1*, 15,199-15,465 bp) and on contig 42 (termed *toxb2*, 15,135-15,401 bp). The *toxb* genes appeared close to a contig end. Ls13-192 contig 4 and 90-2 contigs 37 and 42 have contig assembly sizes 3,110,128 Mb, 116 kb and 87 kb, respectively. No *toxb* gene coding region, protein or nucleotide sequence variations were identified (Supplementary Fig. S2 and S3). *ToxA* was identified in race 2 isolates 86-124 (contig 17, 764,135-764,722 bp) and Biotrigo9-1 (contig 7, 1,370,173-1,370,760 bp), no gene coding region, nucleotide or protein sequence variations were found. The Ptr-specific hairpin element (PtrHp1) *ToxA* 3'UTR insertion previously identified (Moolhuijzen, See et al. 2018) was not detected in the *ToxA* 3'UTR region of these genomes.

The four new assembled and annotated genomes Ls13-192, 86-124, 90-2 and Biotrigo9-1 have been deposited in NCBI GenBank and can be found under accession numbers JAHCSW000000000, NRDI02000000, JAAFOX000000000 and JAHCYZ000000000, respectively.

Table 1. Summary statistics for our four PacBio sequenced Ptr genome assemblies, compared with those of two previously published Ptr assemblies.

	86-124	**Biotrigo9-1**	**Ls13-192**	**90-2**	***M4**	***DW5**
Race	2	2	4	4	1	5
Known effector	A	A	-	-	AC	B
			toxb	toxb x2		
Source	Canada	Brazil	USA	Canada	Australia	USA
GenBank accession	NRDI02	JAHCYZ00	JAHCSW00	JAAFOX00	NQIK00	MUXC02
Number of contigs	139	75	72	162	50	60

Total contigs length (Mb)	41.15	42.19	37.56	39.71	40.92	40.87
Mean contig size (kb)	296.04	562.57	521.68	245.18	998.09	681.19
Median contig size (bp)	23,098	34,534	32,389	20,792	32,745	31,213
Longest contig (Mb)	3.92	10.08	7.54	7.30	9.91	8.11
Shortest contig (bp)	3,180	8,676	1,765	2,050	3,304	2,843
[a] Contigs > 10 kb	113 (81.29 %)	74 (98.67 %)	69 (95.83 %)	52 (93.3 %)	38 (92.68 %)	39 (65.00 %)
[a] Contigs > 100 kb	41 (29.50 %)	18 (24.00 %)	18 (25.00 %)	39 (24.07 %)	11 (26.83 %)	17 (28.33 %)
[a] Contigs > 1 Mb	14 (10.07 %)	12 (16.00 %)	12 (16.67 %)	9 (5.56 %)	10 (24.39 %)	12 (20.00 %)
N50	1,684,023	3,177,932	2,530,800	1,794,835	3,658,030	3,133,851
L50	9	5	5	5	4	5
N80	623,938	1,969,426	1,691,594	567,608	2,765,034	2,129,786
L80	17	10	10	17	8	10
Genes	14,158	14,394	12,816	14,227	15,459	14,276
Total protein (aa) length (Mb)	6.81	6.97	5.95	5.85	6.90	5.95
[b] Predicted effectors	178 (1.2%)	169 (1.1%)	189 (1.4%)	380 (2.6%)	291(1.8%)	314 (2.1%)

*Previously published in NCBI GenBank. [a] Percentage of contigs over the displayed length is shown in brackets. [b] EffectorP V3 predictions ≥ 0.7, the percentage of genes predicted an effector is shown in brackets. N50 and N80 is the sequence length of the shortest contig at 50% and 80% of the total genome length. L50 and L80 is the count of smallest number of contigs whose length sum makes up 50% and 80% of the genome size, respectively.

Illumina genome sequencing, assembly and annotation of eleven *P. tritici-repentis* isolates

Whole genome Illumina sequencing and assembly was then undertaken for eleven new Ptr genomes comprised of isolates from Denmark (EW306-2-1, EW4-4 and EW7m1), Germany (SN001A, SN001C and SN002B), United Kingdom (CC142), Algeria (Alg130 and Alg215) and Tunisia (T199 and T205). The assembled Ptr genomes ranged in size from 34.15 Mb to 35.18 Mb (Table 2), comparable to previous Illumina Ptr isolate assembly sizes (Moolhuijzen, See et al. 2018).

The number of predicted protein-encoding genes ranged between 12,237 to 12,539 for the assembled genomes. Of these, 279 to 300 effectors were predicted with a probability score ≥ 0.7. *ToxA* was identified in isolates T199, Alg215, CC142, EW3061-2-1, EW4-4, SN001C and SN001B, and *ToxB* was identified in the Alg130 genome. *ToxA* and *ToxB* were both detected in T199 and Alg215 genomes, however, the Alg215 *ToxB* sequence was partial (due to a sequence inversion in the 5' end of the gene), truncated by 33 amino acid residues in the protein N-terminus which includes the encoded signal peptide (amino acid positions 1 to 22) (Supplementary Fig. S4). Furthermore, a single nonsynonymous substitution (I > R, residue position 17) was detected. Neither *ToxA* or *ToxB* were detected in isolates EW7m1, SN001A and T205.

The Ptr-specific hairpin element (PtrHp1) *ToxA* 3' UTR insertion previously identified in isolates EW306-2-1 and EW4-4 (Moolhuijzen, See et al. 2018) was also detected in *ToxA* 3' UTR for our United Kingdom isolate CC142, but not in the remaining North African *ToxA* isolates T199 and Alg215 (Table 2).

The plant infection assays on the wheat differential lines confirmed CC142, EW306-2-1 and EW4-4 as race 1 isolates (producing ToxA and ToxC), EW7m1 and SN001A as race 3 isolates (producing ToxC), Alg130 as a race 5 isolate (producing ToxB), T199 as a race 7 isolate (producing ToxA and ToxB) and T205 as a race 4 isolate (no ToxA, ToxB or ToxC production) (Supplementary Fig. S5 and S6). Due to the truncated *ToxB* gene in isolate Alg215 and a weaker chlorosis phenotype on the ToxB wheat differential lines, Alg215 has been provisionally classified as a race 8 isolate (producing ToxA, ToxB and ToxC) (Table 2). The SN001C and SN002B isolates could not be tested for race classification because the colonies sporulated poorly; nonetheless, *ToxA* was present and *ToxB* was absent in the genome sequence for both isolates. As ToxC production in SN001C and SN002B remains unknown, they could be race 1 or 2.

All the assembled and annotated genomes have been deposited in NCBI GenBank and can be found under accession numbers PSOO00000000-PSOU00000000 and RXHK00000000-RXHN00000000.

Table 2. Illumina sequenced genome assemblies of 11 new Ptr isolates. Table shows isolate source, race and *de novo* assembly statistics.

Isolate	CC142	EW306-2-1	EW4-4	EW7m1	SN001A	SN001C	SN002B	Alg130	T199	T205	Alg215
GenBank accession	PSOU00000000	PSOT00000000	PSOS00000000	PSOR00000000	PSOQ00000000	PSOP00000000	PSOO00000000	RXHN00000000	RXHM00000000	RXHL00000000	RXHK00000000
Source	United Kingdom	Denmark	Denmark	Denmark	Germany	Germany	Germany	Algeria	Tunisia	Tunisia	Algeria
Year collected	-	2015	2015	2015	2016	2016	2016	2016	2016	2016	2015
Race	1	1	1	3	3	n.d.	n.d.	5	7	4	[a]8
Known effectors	ToxA, ToxC	ToxA, ToxC	ToxA. ToxC	ToxC	ToxC	[b]ToxA	[b]ToxA	ToxB	ToxA, ToxB	None	ToxA, [c]ToxB, ToxC
Locus ID	10965	03130	05320	-	-	12604	05700	11547	11003; 11565		05415
ToxA-PtrHp1	Present	Absent	Present	Absent	Absent	Present	Absent	Absent	Absent	Absent	Absent
Contigs	2,398	2,592	2,355	2,409	2,369	2,487	2,956	3,748	3,609	3,314	3,374
Total length (Mb)	34.34	34.54	34.37	34.23	34.15	34.30	35.16	35.18	34.68	34.43	34.87
Mean contig size	14,323	13,326	14,595	14,210	14,417	13,791	11,895	9,388	9,610	10,390	10,335
Longest contig	205,419	291,678	233,712	233,813	258,315	188,750	233,762	272,310	309,714	272,252	342,897
N50	47,343	48,368	48,593	48,975	49,129	45,477	48,749	52,525	54,960	53,235	62,752
L50	213	202	206	199	202	221	216	189	176	181	160
Genes	12,347	12,499	12,323	12,429	12,311	12,392	12,472	12,475	12,358	12,237	12,539
Total CDS length (Mb)	15.73	15.79	15.72	15.67	15.63	15.7	15.76	16.27	16.18	16.11	16.37

[d] Predicted Effectors	279 (2.26%)	289 (2.31%)	287 (2.33%)	284 (2.28%)	281 (2.28%)	287 (2.32%)	291 (2.33%)	300 (2.40%)	297 (2.40%)	286 (2.34%)	291 (2.32%)

[a] Provisionally assigned as race 8. [b] Not determined; colonies were not viable for spore production. [c] Partial sequence that is truncated and contains a synonymous SNP, [d] EffectorP3.0 score ≥ 0.7, the percentage of genes predicted an effector is shown in brackets.

Whole genome comparative analyses

Whole genome phylogenetic analysis of the 26 Ptr isolates, sourced from the major wheat growing regions in the Americas, Australia, Europe and North Africa (Fig. 1A), showed distinct clades for European and North African geographic locations (Fig. 1B). Surprisingly, isolate Alg215 from North Africa did not cluster with the remaining North African isolates. On genome alignment to the reference genome of isolate M4, a large 1 Mb distal region on M4 contig 1 and many smaller regions were absent in Alg130, T199 and T205 but present in Alg215 (Supplementary Fig. S7). Furthermore, branches for race 4 (that do not produce ToxA, ToxB or ToxC) isolates (SD20, 90-2 and Ls13-192) had the greatest phylogenetic distances from the known effector producing isolate groups, while race 4 T205 and SD20 (both Illumina sequence) did not cluster. In particular, isolates SD20 (USA) and 90-2 (Canada) were more distant than the isolate Ls13-192 (USA).

Whole genome phylogenetic analysis of Ptr and related ascomycete fungal species clustered into four distinct clades for *Bipolaris* spp., *P. nodorum*, *P. tritici-repentis* and *P. teres* (Fig. 1C). A lower phylogenetic divergence within the individual *Pyrenophora* species (Ptm, Ptr and Ptt) was observed as compared with Bs, Pn and Zt isolates (Supplementary Fig. S8).

To observe regions of absence across the assembled genomes, regions ≥ 10 kb absent for the Ptr isolates were plotted against the reference M4 genome (Fig. 1D). The large horizonal transferred region for ToxA on chr6 was present in all ToxA producing isolates and absent in ToxA non-producing isolates. For the previously reported large Ptr *ToxA* horizontal transfer region, believed to have come from *P. nodorum* (Friesen, Stukenbrock et al. 2006, Manning, Pandelova et al. 2013, Moolhuijzen, See et al. 2018), clear break points on M4 chr6 at the 1,645,874 bp and 1,774,022 bp positions

(128 kb insertion) could be determined between isolates producing and not producing ToxA (Fig. 1D and Supplementary Fig. S9). The flanking regions of the breakpoints were highly conserved between the all aligned isolates (Supplementary Fig. S9). A region on chr1 near the 1.47 Mb position was found absent in all non-ToxC producing isolates and the unknown race (ToxC producing) when only looking at long read assemblies (Fig 1D and Supplementary Fig. S10 (plot on left hand side)). The race 4 isolates had more regions of absence, particularly in the distal ends of chromosome 2. A greater number of absent regions were obtained for Illumina sequenced assemblies (Supplementary Fig. S10, plot on right hand side). Regions of variation appear mostly associated with chromosome telomeres and centromeres. In particular, the distal region on M4 chr10 the equivalent of race 5 isolate DW5 chr11 (1,752,563-2,152,826 bp) was mostly unique as compared with races 1, 2, 4 and the unknown race, with fragmented alignments dispersed throughout the last 100 kb of the chromosome surrounding Ptr *ToxB2* (2,152,563-2,152,826 bp) (Supplementary Fig. S11).

Fig. 1. Whole genome analysis of Ptr isolates. A) Geographic source and number of Ptr isolates currently available and analysed. Legend shows the number of isolates. B) Whole genome phylogenetic tree of Ptr isolates from Illumina sequencing (Alg130, T199, T205, Alg215, CC142, EW306-2-1, EW4-4, EW7m1, DW7, Pt-1C-BFP, Ptr239, Ptr11137, Ptr5213, SD20, SN001A, SN001C, SN002B), PacBio Technologies (86-124, 90-2, AR CrossB10, Biotrigo9-1, DW5, Ls13-192, M4 and V1) and Oxford Nanopore Technologies (Ptr134). The unrooted neighbour joining phylogenetic tree displays clades for the European (violet) and North Africa (tan) isolates. Geographic source of the other isolates are Australian (blue), USA (green), Canada (red) and

Brazil (purple). The race 4 isolates (Ls13-192, 90-2 and SD20) have the greatest distance from the clade of known effector producing isolates. C) Unrooted Neighbour joining phylogenetic tree for Ptr (purple clade), *Pyrenophora teres* (*P. teres* f. *maculata* (Ptm) and *P. teres* f. *teres* (Ptt)) (orange clade), *Bipolaris* (*B. sorokiniana* (Bs1-3 and Q7399), *B. maydis* (Bm-ATCC and Bm-C5) and *Bipolaris zeicola* (Bz) (green clade), *Parastagonospora nodorum* (Sn4, Sn15 and Sn79) (yellow clade), *Leptosphaeria maculans* (Lm) and *Zymoseptoria tritici* (Zt) isolates. The branches for race 4 isolates not producing known effectors (Ls13-192, 90-2 and SD20) are highlighted (blue) within the Ptr clade. D) Circular plots show 10 kb regions of absence plotted for the Ptr isolates genomes sequenced using long-read technologies (PacBio and Oxford Nanopore Technology) as compared with the chromosomes of the reference Ptr genome of isolate M4. Isolates are coloured by race. Three regions of interest are highlighted in grey and zoomed at 20x for chromosome 2 and chromosome 1, and 40x for chromosome 9.

P. tritici-repentis **mRNA sequence alignment to whole genomes**

To ensure a comprehensive search of Ptr genes in the pangenome, predicted mRNA sequences from all isolates were aligned to all the genomes at greater than 90% sequence identity and 90% coverage. The number of alignments and greatest percent identity for each locus were recorded to determine isolate correlations (Fig. 2). Although a closer correlation by gene percent sequence identity could be determined for isolates that were Illumina or PacBio sequenced, a distinct grouping for Alg130, T199 and T205, and a grouping of the European isolates, was evident. Furthermore, the race 4 isolates 90-2 and SD20 were less correlated to all the remaining isolates (Fig 2A). Based on gene counts (copy number) three distinct groups were observed, for long read-sequenced, European Illumina-sequenced and Australian/North-African/North-American Illumina-sequenced isolates (Fig 2B). However, the three race 4 isolates (Ls13-192, 90-2 and SD20) were outliers.

Fig. 2. Ptr pangenome predicted mRNA correlation plots for gene sequence percentage identity (A) and gene copy number (B). Ptr isolates from Illumina sequencing (Alg130, T199, T205, Alg215, CC142, EW306-2-1, EW4-4, EW7m1, DW7, Pt-1C-BFP, Ptr239, Ptr11137, Ptr5213, SD20, SN001A, SN001C, SN002B), PacBio Technologies (86-124 (Ptr86-124), 90-2 (Ptr90-2), AR CrossB10, Biotrigo9-1, DW5, Ls13-192, M4 and V1) and Oxford Nanopore Technologies (Ptr134).

All genes were then filtered for presence/absence variation between ToxC-producing isolates [race 1 (Pt-1C-BFP, CC142, EW4_4 and EW306-2-1), race 3 (EW7m1 and SN001A), race unknown (AR CrossB10) and provisional race 8 (Alg215)] and non-ToxC producing isolates [race 2 (86-124 and Biorigo9-1), race 4 (T205, Ls13-192, 90-2 and SD20), race 5 (DW5 and DW7) and race 7 (T199)] to identify genes that may be related to ToxC production. When only PacBio sequenced genomes were queried, a gene cluster of 16 genes from isolate M4 mRNAs 12743 to 12761 (proteins KAF7566087-KAF7566105) positioned on M4 chromosome 9 within 101,367 - 138,426 bp and 14 single loci genes outside of the cluster were found present in the ToxC producing races (races 1 and unknown) and absent in races not producing ToxC (races 2, 4 and 5) (Supplementary Fig. S12). The region was however absent for the race 1 Oxford Nanopore technology (ONT)-sequenced isolate Ptr134 (Fig. 1D). None of the 30 genes found to be specific to ToxC-producing isolates (based on PacBio technology) had an identified signal peptide or appeared to be part of any predicted biosynthetic gene cluster (Table 3 and Supplementary data 3). A search of the pathogen-host interaction database PHI-base (Urban, Cuzick et al. 2017), which provides expertly curated molecular and biological information on genes proven to affect the outcome of pathogen-host interactions, did however identify four proteins with significant alignments to proteins with classified reduced virulence and lethal phenotypes. The following proteins with reduced virulence phenotype were described as being an indoleacetamide hydrolase (iaaH) involved in auxin biosynthesis and plant hormone metabolism (P06618) in *Pseudomonas savastanoi*, a Non-Ribosomal Protein Synthase (NRPS) (A0A024CHY2) in *Pseudomonas cichorii* and an AMP binding protein (E3QPY3) in *Colletotrichum graminicola*. The protein I1RXA5, classified with a lethal phenotype in *F. graminearum*, appears to be a transcription factor (homeobox).

Table 3. Ptr predicted mRNA sequences identified specific to ToxC-producing isolates (Pacbio sequenced) and PHI-base results.

M4 GenBank accession	Chr.	Strand	Gene position	M4 mRNA	Protein ID	Description	Length (aa)	PHI base ID	Expected value
CM025795	Chr1	+	4192655-4193250	mRNA_1649	KAF7577409	hypothetical protein	161		
CM025796	Chr2	-	5052985-5053659	mRNA_5712	[c]KAF7575752	hypothetical protein	224		
CM025797	Chr3	+	94019-97874	mRNA_5765	KAF7572524	transmembrane Dimer-Tnp-hAT domain containing protein	1104	[a] A0A024CHY2	6.00E-18
CM025797	Chr3	+	104828-106828	mRNA_5768	KAF7572527	hypothetical protein	666	[a] E3QPY3	2.00E-56
CM025797	Chr3	-	1062456-1062785	mRNA_6121	KAF7572880	hypothetical protein	109		
CM025797	Chr3	+	1234950-1235279	mRNA_6173	KAF7572932	hypothetical protein	109		
CM025799	Chr5	+	1432573-1432928	mRNA_8853	KAF7570515	hypothetical protein	99		
CM025799	Chr5	+	1504991-1506668	mRNA_8883	KAF7570545	hypothetical protein	518		
CM025799	Chr5	+	3350956..3351804	mRNA_9590	KAF7571252	DDE-3 multi-domain protein	282	[b] I1RXA5	1.00E-55
CM025800	Chr6	-	157070-158024	mRNA_9652	KAF7568901	hypothetical protein	260		
CM025800	Chr6	-	227855-229037	mRNA_9675	KAF7568924	hypothetical protein	292		
CM025800	Chr6	+	536251-538210	mRNA_9782	[c]KAF7569031	Amidase domain containing protein	535	[a] P06618	1.00E-14
CM025800	Chr6	+	1257276-1261804	mRNA_10029	KAF7569278	hypothetical	1489		

						protein	
CM025803	Chr9	-	101367-102082	mRNA_12743	KAF7566087	hypothetical protein	166
CM025803	Chr9	+	103558-104538	mRNA_12744	KAF7566088	hypothetical protein	326
CM025803	Chr9	-	108927-109433	mRNA_12746	[cd]KAF7566090	hypothetical protein	153
CM025803	Chr9	-	109617-109871	mRNA_12747	[d]KAF7566091	hypothetical protein	84
CM025803	Chr9	-	112305-113390	mRNA_12748	[d]KAF7566092	hypothetical protein	346
CM025803	Chr9	+	114593-114844	mRNA_12749	[d]KAF7566093	hypothetical protein	83
CM025803	Chr9	-	115206-116625	mRNA_12750	[d]KAF7566094	hypothetical protein	454
CM025803	Chr9	+	117388-118112	mRNA_12751	[d]KAF7566095	hypothetical protein	225
CM025803	Chr9	+	118451..119788	mRNA_12752	[d]KAF7566096	hypothetical protein	445
CM025803	Chr9	-	120249-121379	mRNA_12753	KAF7566097	Methyltransf-18 multi-domain protein	376
CM025803	Chr9	+	125936-126175	mRNA_12755	[d]KAF7566099	hypothetical protein	60
CM025803	Chr9	-	127211-127757	mRNA_12756	KAF7566100	hypothetical protein	159
CM025803	Chr9	+	128397-131000	mRNA_12757	KAF7566101	Cwf-Cwc-15 domain containing protein	867
CM025803	Chr9	-	131128-131483	mRNA_12758	KAF7566102	hypothetical protein	99
CM025803	Chr9	-	134706-135131	mRNA_12759	[d]KAF7566103	hypothetical	91

						protein	
CM025803	Chr9	-	137602-138426	mRNA_12761	KAF7566105	hypothetical protein	274
CM025804	Chr10	-	2199594-2200463	mRNA_14375	KAF7565231.1	hypothetical protein	265

PHI-base phenotype classifications [a] reduced virulence, [b] lethal. [c] M4 mRNA *in planta*. [d] M4 mRNA *in vitro*.

The genes specific for ToxC-producing isolates (that were PacBio-sequenced) were also searched in previous published *in planta* and *in vitro* RNA-seq data (Moolhuijzen, See et al. 2018) and most of the gene cluster (KAF7566087-KAF7566105) had *in vitro* transcription support (Supplementary Fig. S13). Only the hypothetical transmembrane protein (KAF7575752), amidase containing protein (KAF7569031) and two other hypothetical proteins had *in planta* transcription support during Ptr infection (Table 3).

When all sequenced isolates were considered, only a single locus for a transmembrane protein, an integral membrane component, was identified core to all ToxC-producing isolates, represented by the M4 protein (KAF7575752) on chromosome 2 position 5,052,985-5,053,659 bp (Fig. 1D). This gene was recently identified as *ToxC1*, a gene required but not sufficient for ToxC production in Ptr (Shi, Kariyawasam et al. 2022). A less stringent search for *ToxC1* in all isolates detected the presence of *ToxC1* in the race 2 isolate Biotrigo9-1 genome, which was disrupted by a large insertion of 5,348 bp, positioned at 45,946 to 51,292 bp on contig 12, which disrupted the *ToxC1* protein coding region in the 582-583 bp position. Examination of the 2 kb gene flanking regions of all genomes indicated a further large insertion downstream of the gene in Biotrigo9-1 (Fig. 3). The two large insertions do not have a similar sequence identity, with the insertion downstream of *ToxC1* carrying Gypsy retrotransposon transposable elements (TEs) and the *ToxC1* insertion carrying Copia retrotransposon TEs informed by flanking long terminal repeats (LTRs) (Supplementary Fig. S14).

Fig. 3. Ptr isolate M4 *ToxC1* locus and 2 kb flanking sequence region alignment to twelve other Ptr ToxC producing isolates. The Biotrigo9-1 *ToxC1* region has two

large insertions within and downstream of *ToxC1*. Nucleotide sequence alignments (blue) between the *ToxC1* region for Ptr isolates (top to bottom M4, AR CrossB10, Biotrigo9-1, Pt-1C-BFP, V1, Ptr134, Alg215, CC142, EW306-2-1, EW7m1, SN0001A, SN0001C and SN0002B) (black lines). The M4 genes are shown as red arrows. The light blue alignment segments are regions of low identity among the isolates, while the crossed regions indicate a repeat region in each sequence.

Analyses of core and ancillary gene sets / protein clusters

To determine core and ancillary protein groups in Ptr, a total of 335,037 predicted protein coding genes from this study and published genomes downloaded from NCBI (see Material and Methods) were clustered. Out of the total number, 331,581 proteins clustered into 14,960 orthologous groups and 3,456 singletons, representing a pangenome for Ptr. A total of 8,509 groups were core (56%) (with all isolates present) and 7,170 orthogroups (47%) consisted entirely of single-copy genes (Supplementary data 4). Overall, for the PacBio sequenced isolates, race 4 isolate 90-2 had a low percentage of single copy genes (78%) (higher percentage of gene duplications), similar to M4 (76%) (Supplementary data 4). The percentage of single copy genes for the PacBio sequenced genomes ranged from 76% for M4 to 95% for AR CrossB10.

Across the Ptr pangenome (core and ancillary genes), 32,257 (9.6%) genes had a signal peptide of which just over one third (11,911 genes) were predicted to be effectors (EffectorP 3.0 default probability score ≥ 0.5). The EffectorP 3.0 effector probability scores for ToxA and ToxB were 0.702 and 0.93, respectively. The effectors *ToxA* and *ToxB*/*toxb* were identified in orthologous protein groups OG0011421 and OG0011851, respectively.

All predicted effectors protein sequences were then clustered into 738 orthogroups, of which five groups were isolate specific (containing paralogous genes) and 187 were singletons (a single gene). Of the 738 effector orthogroups, only 119 (16%) were core to all isolates and of the core orthogroups 25 orthogroups (21%) had 100% sequence identity. Of the non-core effector groups, 62 orthogroups were absent in the race 4 isolates T205, Ls13-192, 90-2 and SD20.

A comparison of predicted effectors from orthogroups with race 4 absent to those with race 4 present found that the average protein length was shorter (T-test, Wilcoxon adj. *P-value* 2.9e-294) and the effector probability scores were higher (T-test, Wilcoxon adj. *P-value* 1.8-28) (Supplementary Fig. S15).

Protein tertiary structure analysis of predicted effectors

To identify protein tertiary structure homology, predicted effectors were screened using remote homology detection methods against known protein structures to build 3-D models. Of these, 147 proteins had predicted high confidence tertiary models based on published tertiary protein structures (Phyre2 confidence ≥ 90% and alignment coverage ≥ 90%) (Supplementary data 5). Of the high confidence proteins, a total of 48 and 19 had annotated hydrolase and binding functions, respectively. Five were annotated as effectors, which included Ptr ToxA necrosis effector (KAF7569451) with 100% sequence identity to the Protein Database (PDB) crystal protein structure of ToxA 1ZLD and four elicitor proteins, hrip2 (KAF7578077, KAF7575054, KAF7570798 and KAF757229) based on the crystal structure from *Magnaphorthe oryzae* (PDB 5FID) with sequence identities ranging between 23 - 26%.

The 147 predicted effector proteins with a confident protein tertiary model were then searched against Phi-Base (Urban, Cuzick et al. 2017). A total of 34 proteins had known Phi-Base pathogenicity or reduced virulence hits, of which 11 were plant avirulence determinants, which included ToxA (Supplementary data 5).

To enable the capture of genes that may have been filtered out previously (that may not have a predicted signal peptide), whole genome HMM libraries of M4 were generated for screening using BackPhyre (Kelley, Mezulis et al. 2015). Effector related protein structures were then selected from toxins available in the RCSB PDB for Ptr ToxB (2MM0), toxb (2MM2), ToxA (1ZLD) and SnTox3 (6WES) to identify any other structural homologues and orthologues, respectively. No structural paralogues for ToxA or ToxB were identified in isolate M4 (with confidence levels ≥ 20.0); however, an orthologous structure was identified for SnTox3 with 58% alignment coverage (46 - 138 amino acids) to M4 (protein accession KAF7577476) (104 - 195 amino acids in the alignment) with a confidence score of 95.5 and 34% protein sequence identity (Fig. 4A). This indicated a high confidence that the match between KAF7577476 and the PnTox3 template is a true homology that adopts the overall protein fold and that the core protein is modelled at a high accuracy (2-4 A° from the native, true structure). The 3-D protein structures for SnTox3 (Fig. 4B) and predicted structure for KAF7577476 (Fig. 4C) were then structurally aligned and superimposed with a root mean square distance (RMSD) of 1.14 A° (Fig. 4D).

Fig. 4. The predicted protein sequence and structural alignments of SnTox3 and the isolate M4 protein KAF7577476. A) Multiple protein sequence alignment of SnTox3, Ptt CAA9973983.1 (W1-1), Ptm CAA9957881.1 (SG1) and Ptr KAF7577476 (M4). The Kex2 motif conservation is shown boxed in red. Only four cysteine residues were conserved across the four species (black asterisks) and those not conserved (red asterisks) are shown below the alignment for *P. nodorum* and above for the *Pyrenophora* spp. B) The known 3-D protein structure for SnTox3 (PDB 6WES). C) The 3-D structure for KAF7577476 as predicted by Phyre2. D) Superimposed

structural alignment (yellow) of SnTox3 and KAF7577476 with an RSDM of 1.14 A°.

A TBLASTN sequence search of the M4 isolate predicted protein KAF7577476 against all the Ptr genomes found evidence that the gene encoding this predicted protein is present in all isolates. The KAF7577476 protein sequence was then also searched against the genomes of the related barley necrotrophic fungal pathogens *P. teres* f. *teres* isolate W1-1 (Ptt) and *P. teres* f. *maculata* isolate SG1 (Ptm) (Syme, Martin et al. 2018) and high identity orthologues were also identified: CAA9973983.1 (isolate W1-1) and CAA9957881.1 (isolate SG1), respectively (Fig. 4A). An automated and combinative method for ranking top candidate effector proteins (Predector) (Jones, Rozano et al. 2021) ranked Ptr KAF7577476 in the 262th position, Ptt CAA9973983 as the top candidate (number one) and Ptm CAA9957881 in the 56th position with Predector scores of 1.9, 3.9 and 2.7, respectively (Supplementary data S1). The SnKex2 cleavage motif LSKR (69 - 72 amino acids) of SnTox3 (Outram, Solomon et al. 2021) aligned to AKEL protein residues in the three *Pyrenophora* spp. (Ptr, Ptm and Ptt), where the residue positioned before the cleavage site (P1) is expected to be exclusively an arginine (Arg, R) (Outram, Solomon et al. 2021) (Fig. 4A). Furthermore, the *Pyrenophora* sequences appeared to possess only 4 of the 6 cysteine residues, that form three disulphide bonds (Osbourn 2010), conserved with SnTox3. The predicted apoplastic effector scores for SnTox3, KAF7577476 (Ptr), CAA9973983 (Ptt) and CAA9957881 (Ptm) were 0.573, 0.572, 0.691 and 0.765, respectively.

Discussion

P. tritici-repentis pangenome analysis

In this study, we present the pangenome of 26 Ptr isolates with a near complete representation of the eight known race categories. Our 15 newly assembled and annotated genomes, along with the 11 previously published genomes, represent a global pangenome of Ptr for major wheat growing regions, with close and distant proximity to the origin of wheat domestication in the Fertile Crescent of western Asia. The repertoire of the known Ptr genes (Moolhuijzen, See et al. 2018) was expanded by 31%, represented by 18,416 non-redundant sequences. This expansion of genes is also observed in other plant fungal species, where a pangenome analysis of 20 *F. graminearum* isolates resulted in a 32% gene expansion over the reference isolate (Alouane, Rimbert et al. 2021). The 56% conservation of core orthogroups in Ptr identified here is similar in magnitude to a recent 19 isolate pangenome analysis of the wheat pathogen *Z. tritici* (Badet, Oggenfuss et al. 2020), which found 60% of gene orthogroups were core.

A number of Ascomycete genomes, such as Pn, Ptt and Zt, have ‘two speed genomes’, where the genome is compartmentalised into gene-poor AT-rich regions and can have accessory chromosomes. In contrast, Ptr does not appear to have accessory chromosomes and has a GC-equilibrated genome (Dong, Raffaele et al. 2015, Testa, Oliver et al. 2016, Bertazzoni, Williams et al. 2018, Moolhuijzen, See et al. 2018, Syme, Martin et al. 2018). The whole genome phylogenetic analysis clearly showed greater isolate phylogenetic distances within Bs, Pn and Zt isolates as compared with the *Pyrenophora* spp. (Ptm, Ptr and Ptt). However, even with comparatively low phylogenetic distances within Ptr, distinct clades could be detected based on geographic locations. The only exception was isolate Alg215 from Algeria, which clustered with the Australian and American isolates, sharing a large sub-telomeric region in common. This sequence variation, plus a disrupted *ToxB*, set Alg215 apart from the other isolates collected from North Africa. Despite the low whole-genome phylogenetic distances in Ptr, a lower percentage of core orthogroups (56%) was found compared with a previous analyses of 11 isolates (PacBio and Illumina sequenced), which indicated 69% core orthologous groups (Moolhuijzen, See et al. 2018). This suggests that not only has the pangenome complexity risen with an increase in the numbers of isolates sequenced (as expected), but that an increased divergence in Ptr conserved protein domains is apparent.

Although in this analysis only a single gene was identified as specific to all ToxC producing isolates (*ToxC1*), PacBio sequencing identified a potential gene cluster of interest which would be near impossible to identify in Illumina-sequenced genomes, due to the repetitive nature of the region. Interestingly, our analysis found no putative effectors that were core to all isolates, again indicating a large variability within Ptr for this type of gene. Recently, *ToxC1* was functionally validated using a gene knockout approach (Shi, Kariyawasam et al. 2022), where it was found to be required, but not sufficient, for ToxC production. In our study, no clear gene cluster for a secondary metabolite or Ribosomally synthesized and post-translationally modified peptides (RiPPs) was identified, in part due to the positioning of the *ToxC* locus within the complex subtelomeric region of chromosome 2 (Shi, Kariyawasam et al. 2022), which despite long read sequencing still remains a problematic region to resolve. The presence of *ToxC1* in a non-ToxC producing isolate (Biotrigo9-1) was surprising, and raises more questions regarding the evolution and/or origin of ToxC production. It is possible that the large *ToxC1* insertion by a LTR retrotransposon has disrupted the production of ToxC in Biotrigo9-1 and that those remaining gene(s) involved in ToxC production are present.

The divergence of Ptr race 4 isolates (that do not produce the known effectors on wheat) from isolates that produce known effectors was clearly shown, except for T205. The genome sizes and gene duplication rates of the two race 4 isolates (Ls13-192 and 90-2) also revealed a complexity that was unexpected. Race 4 was first described 30 years ago (Lamari and Bernier 1989, Lamari and Bernier 1991) as a nec$^-$ chl$^-$ pathotype (avirulent) on the set of differential wheat lines, and has since been reported but not as frequently as the other races from collections of Ptr across different wheat growing regions. A recent study by Guo et al. (2020) showed that despite the inability of race 4 isolates to induce tan spot symptoms on the differential wheat lines, four race 4 isolates (Ls13-14, Ls13-86, Ls13-192 and Ls13-198) from North Dakota in the USA induced varying degrees of disease reactions upon inoculation on tetraploid (durum) wheats (Guo, Shi et al. 2020). This may well provide an explanation for the observed distinction of Ls13-192 from the other race 4 isolates (90-2 and SD20) in the whole genome phylogenetic clustering and gene correlation analyses, since unlike Ls13-192, SD20 and 90-2 have not been reported to be virulent on durum wheat. Furthermore, as *ToxA* and *ToxB* were absent, race 4 isolate T205 is unlike the new virulence type that lacked ToxA and ToxB gene expression on bread wheat differentials but produced necrosis in durum wheat (Benslimane 2018).

While it was unexpected that a race 4 isolate (90-2) had the highest percentage of genes predicted as effectors, this appeared to be the result of a genome-wide expansion of gene copies (which included predicted effectors). It is possible that although the predicted effectors in race 4 isolates may have not have a pathogenic role in bread wheat, they may play a role in another system.

We report, for the first time, an identical *toxb* (non-toxic homologue of *ToxB*) copy in a race 4 genome (2 genes in 90-2). As each *toxb* are on separate contigs, it is not possible to identify if they are co-located. We can, however, speculate that based on the difficulty in assembling the *toxb* regions, they may lie in a subtelomeric chromosome location similar to the multicopy *ToxB*, which was shown to be nested in the complex subtelomeric chromosomal regions of the DW5 genome (Moolhuijzen, See et al. 2020). It is increasingly believed that effector genes are located in transposon-rich and gene-sparse subtelomeric regions of the pathogen genome, allowing opportunity for gene duplication events and thereby contributing to the evolution of virulence diversity. We also show no conservation between the different races for the *ToxB* locus or flanking regions. The sequence variation in the chromosomal centromeric and telomeric regions shown in our whole genome alignments, indicates that these regions are indeed hot spots for diversity. It is furthermore interesting that one of the North Africa isolates (Alg 215) had a truncated *ToxB* gene with a nonsynonymous mutation within the coding region, which may have resulted in a weak chlorosis phenotype on the wheat differential lines. We believe that this is the first report of a *ToxB* nonsynonymous mutation in Ptr. The large *ToxA* horizontal transfer region previously identified (Manning, Pandelova et al. 2013, Moolhuijzen, See et al. 2018) was shown to be absent in all non-ToxA producing isolates and clear insertion breakpoints were identified in all ToxA producing isolates.

In this study, a pangenome approach was undertaken to approximate the complete gene repertoire of the species to capture all gene variations (percentage identity and copy number) and identify candidate genes specific for ToxC-producing races.

A new *Pyrenophora* resource to identify protein structural homologues

Pathogenic fungi possess large effector repertoires that are dominated by hundreds of small secreted proteins only related by protein tertiary structures (3-D structure) (de Guillen, Ortiz-Vallejo et al. 2015). The prediction of new effector candidates that are not the result of horizontal gene transfer is therefore complicated.

To conduct a comprehensive whole genome search of protein tertiary structures an *in silico* screening was employed using BackPhyre (Kelley, Mezulis et al. 2015). We present here the first necrotrophic fungal pathogen publicly available through BackPhyre (Kelley, Mezulis et al. 2015) for effector and other protein tertiary structure searches, providing further annotation evidence for a number of hypothetical genes. In this pangenome screen of proteins, no other ToxA or ToxB-like paralogues were identified based on structural similarity in Ptr.

Overall, the use of protein three-dimensional structure modelling improved the identification of a number of proteins which included effector candidates potentially involved in pathogenicity.

In silico protein structural analysis reveals a natural homologue to *SnTox3* in *Pyrenophora*

We report here for the first time a distant *SnTox3* natural homologue in *Pyrenophora*. We showed conserved structural homology between SnTox3 and *Pyrenophora* proteins that lacked conservation in the R residue position of the Kex2 motif (LXXR) and the full set of cysteine residues forming the three disulphide bonds in SnTox3.

SnTox3 is a pro-domain containing effector, where the signal peptide and pro-domain are removed (cleaved by the Kex2 protease) to produce a more potent protein that activates host cell death (Snn3) (Outram, Sung et al. 2021). The Kex2 cleavage motif (LXXR) has the following residue preferences, a Leucine (L, Leu) or any other aliphatic residue, any residue X as it does not interact with Kex2, Lysine (K, Lys) but has other possible residues Lys > Arginine (R, Arg) >Threonine (T, Thr) > Proline (P, Pro) > Glutamic acid (E, Glu) > Isoleucine (I, Ile) (X) and exclusively an arginine (R, Arg) before the cleavage site (Outram, Solomon et al. 2021). Here we found the conserved *Pyrenophora* motif (AKEL) did not conformed to the Kex2 cleavage motif (LXXR) in two residue positions that included the exclusive arginine residue.

Interestingly, the Ptr structural homologue to SnTox3 was in all isolate races, unlike the non-active *toxb* that only occurs in non-pathogenic race 4 isolates (not producing known effectors). As no *in planta* gene expression for the Ptr homologue of *SnTox3* was detected and the protein sequence had a low effector prediction ranking, we believe it may not be an effector candidate in the wheat-pathogen system. However, conversely in Ptt, as the structural homologue to *SnTox3* is expressed during barley infection (Moolhuijzen, Lawrence et al. 2021) and was ranked as the top candidate effector, we believe further investigation is warranted. Here, we propose that the identification of a SnTox3 structural homologue in *Pyrenophora* (Ptm, Ptr and Ptt) could be part of a structurally defined family and are phylogenetically related to *SnTox3*, as observed for the *M. oryzae* Avirulence (Avrs) and ToxB (MAX-effector proteins) (de Guillen, Ortiz-Vallejo et al. 2015).

In conclusion, the new genomic resources presented here improve the pangenome representation of Ptr and provide putative effector candidates based on structural modelling and ranking specific to effector producing isolates. These resources can be used to monitor Ptr variations potentially involved in pathogenicity. As Ptr is commonly shown to infect wheat in combination with other necrotrophic pathogens (Justesen, Corsi et al. 2021), the future ability to simultaneously monitor such changes in multiple necrotrophic species may enhance pathogen monitoring activities within a wider framework of crop protection activities.

Funding

This work was generously supported through co-investment by the Grains Research and Development Corporation (GRDC) and Curtin University (project code CUR00023), as well as the Australian Government National Collaborative Research Infrastructure Strategy and Education Investment Fund Super Science Initiative. This project was also supported by the Agriculture and Food Research Initiative competitive grants program (award number 2016-67014-24806) and the National Institute of Food and Agriculture, United States Department of Agriculture (USDA) Hatch project (ND02234) to ZL. JC, JT and LJ were supported by the 'Efectawheat' project funded within the framework of the 2nd call ERA-NET for Coordinating Plant Sciences by British Biological Sciences Research Council (BBSRC) grant BB/N00518X/1 to JC and the Danish Council of Strategic Research grant case number 5147-00002B to LJ. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgements

We thank the Australian grain growers for their continued support of research through the Grains Research and Development Corporation (GRDC) and the Australian

Government National Collaborative Research Infrastructure Strategy (NCRIS) for providing access to Pawsey Supercomputing under a National Computational Merit Allocation Scheme (NCMAS), Nectar Research and Pawsey Nimbus Cloud resources. We would also like to acknowledge Professor Richard Oliver who was key in the setting up the collaborators in this study as part of the EfectaWheat project.

Competing interests

The authors declare no competing interests.

Author contributions

Conceptualisation ZL and CSM; methodology, PM, PTS and HP; formal analysis, PM, PTS, GS and HP; investigation, PM; project resources JC, JT, SS, HB and LJ; writing - original draft preparation, PM; writing – review and editing, PM, PTS, ZL, GS, HB, JC, LJ, SS and CM. All authors have read and agreed to the published version of the manuscript.

Data availability

All data generated or analyzed during this study are included and can be accessed in this published article (and its supplementary files). The sequence data has been deposited in the DDBJ/ENA/GenBank under accession numbers JAAFOX000000000, JAHCSW000000000, JAHCYZ000000000, NRDI02000000 (version 2), PSOO00000000-PSOU00000000 and RXHK00000000-RXHN00000000.

References

Ali S, Gurung S and Adhikari TB (2010). "Identification and Characterization of Novel Isolates of Pyrenophora tritici-repentis from Arkansas." APS **94**(2): 229-235.

Alouane, T., H. Rimbert, J. Bormann, G. A. Gonzalez-Montiel, S. Loesgen, W. Schafer, M. Freitag, T. Langin and L. Bonhomme (2021). "Comparative Genomics of Eight Fusarium graminearum Strains with Contrasting Aggressiveness Reveals an Expanded Open Pangenome and Extended Effector Content Signatures." Int J Mol Sci **22**(12).

Andrews, S. (2011). "FastQC." Retrieved 2016, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Andrie, R. M., I. Pandelova and L. M. Ciuffetti (2007). "A Combination of Phenotypic and Genotypic Characterization Strengthens Pyrenophora tritici-repentis Race Identification." Phytopathology **97**(6): 694-701.

Badet, T., U. Oggenfuss, L. Abraham, B. A. McDonald and D. Croll (2020). "A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici." BMC Biol **18**(1): 12.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev and P. A. Pevzner (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." J Comput Biol **19**(5): 455-477.

Benslimane, H. (2018). "Virulence Phenotyping and Molecular Characterization of a New Virulence Type of Pyrenophora tritici-repentis the Causal Agent of Tan Spot." Plant Pathol J **34**(2): 139-142.

Benslimane, H., L. Lamari, A. Benbelkacem, R. Sayoud and Z. Bouznad (2011). "Distribution of races of Pyrenophora tritici-repentis in Algeria and identication of a new virulence type." Phytopathologia Mediterranea **50**(2): 203-211.

Bertagnolli, V. V., J. R. Ferreira, Z. H. Liu, A. C. Rosa and C. C. Deuner (2019). "Phenotypical and genotypical characterization of Pyrenophora tritici-repentis races in Brazil." European Journal of Plant Pathology **154**(4): 995-1007.

Bertazzoni, S., D. A. B. Jones, H. T. Phan, K. C. Tan and J. K. Hane (2021). "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." BMC Genomics **22**(1): 382.

Bertazzoni, S., A. H. Williams, D. A. Jones, R. A. Syme, K. C. Tan and J. K. Hane (2018). "Accessories Make the Outfit: Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi." Molecular Plant-Microbe Interactions **31**(8): 779-788.

Bhathal, J. S., R. Loughman and J. Speijers (2003). "Yield reduction in wheat in relation to leaf disease from yellow (tan) spot and septoria nodorum blotch." European Journal of Plant Pathology **109**(5): 435-443.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Borodovsky, M. and A. Lomsadze (2011). "Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES." Curr Protoc Bioinformatics **Chapter 4**: Unit 4 6 1-10.

Chen, N. (2004). "Using RepeatMasker to identify repetitive elements in genomic sequences." Curr Protoc Bioinformatics **Chapter 4**: Unit 4 10.

Ciuffetti, L. M., V. A. Manning, I. Pandelova, M. F. Betts and J. P. Martinez (2010). "Host-selective toxins, Ptr ToxA and Ptr ToxB, as necrotrophic effectors in the Pyrenophora tritici-repentis-wheat interaction." New Phytol **187**(4): 911-919.

Ciuffetti, L. M., R. P. Tuori and J. M. Gaventa (1997). "A single gene encodes a selective toxin causal to the development of tan spot of wheat." Plant Cell **9**(2): 135-144.

de Guillen, K., D. Ortiz-Vallejo, J. Gracy, E. Fournier, T. Kroj and A. Padilla (2015). "Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi." PLoS Pathog **11**(10): e1005228.

Delcher, A. L., S. L. Salzberg and A. M. Phillippy (2003). "Using MUMmer to identify similar regions in large sequence sets." Curr Protoc Bioinformatics **Chapter 10**: Unit 10 13.

Dong, S., S. Raffaele and S. Kamoun (2015). "The two-speed genomes of filamentous pathogens: waltz with plants." Curr Opin Genet Dev **35**: 57-65.

Downie, R. C., M. Lin, B. Corsi, A. Ficke, M. Lillemo, R. P. Oliver, H. T. T. Phan, K. C. Tan and J. Cockram (2021). "Septoria Nodorum Blotch of Wheat: Disease Management and Resistance Breeding in the Face of Shifting Disease Dynamics and a Changing Environment." Phytopathology **111**(6): 906-920.

Emms, D. M. and S. Kelly (2015). "OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy." Genome Biol **16**: 157.

Faris, J. D., Z. Liu and S. S. Xu (2013). "Genetics of tan spot resistance in wheat." Theor Appl Genet **126**(9): 2197-2217.

Figueroa Betts, M., V. A. Manning, K. B. Cardwell, I. Pandelova and L. M. Ciuffetti (2011). "The importance of the N-terminus for activity of Ptr ToxB, a chlorosis-inducing host-selective toxin produced by Pyrenophora tritici-repentis." Physiological and Molecular Plant Pathology **75**(4): 138-145.

Friesen, T. L., E. H. Stukenbrock, Z. Liu, S. Meinhardt, H. Ling, J. D. Faris, J. B. Rasmussen, P. S. Solomon, B. A. McDonald and R. P. Oliver (2006). "Emergence of a new disease as a result of interspecific virulence gene transfer." Nat Genet **38**(8): 953-956.

Friesen, T. L., Z. Zhang, P. S. Solomon, R. P. Oliver and J. D. Faris (2008). "Characterization of the interaction of a novel Stagonospora nodorum host-selective toxin with a wheat susceptibility gene." Plant Physiol **146**(2): 682-693.

Guo, J., G. Shi, A. Kalil, A. Friskop, E. Elias, S. S. Xu, J. D. Faris and Z. Liu (2020). "Pyrenophora tritici-repentis Race 4 Isolates Cause Disease on Tetraploid Wheat." Phytopathology **110**(11): 1781-1790.

Jiang, H., R. Lei, S. W. Ding and S. Zhu (2014). "Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads." BMC Bioinformatics **15**: 182.

Jones, D. A. B., L. Rozano, J. W. Debler, R. L. Mancera, P. M. Moolhuijzen and J. K. Hane (2021). "An automated and combinative method for the predictive ranking of candidate effector proteins of fungal plant pathogens." Sci Rep **11**(1): 19731.

Justesen, A. F., B. Corsi, A. Ficke, L. Hartl, S. Holdgate, L. N. Jørgensen, M. Lillemo, M. Lin, I. J. Mackay, V. Mohler, M. Stadlmeier, K.-C. Tan, J. Turner, R. P. Oliver and J. Cockram (2021). "Hidden in plain sight: a molecular field survey of three wheat leaf blotch fungal diseases in North-Western Europe shows co-infection is widespread." European Journal of Plant Pathology **160**(4): 949-962.

Kamel, S., M. Cherif, M. Hafez, T. Despins and R. Aboukhaddour (2019). "Pyrenophora tritici-repentis in Tunisia: Race Structure and Effector Genes." Front Plant Sci **10**: 1562.

Kariyawasam, G. K., N. Wyatt, G. Shi, S. Liu, C. Yan, Y. Ma, S. Zhong, J. B. Rasmussen, P. Moolhuijzen, C. S. Moffat, T. L. Friesen and Z. Liu (2021). "A genome-wide genetic linkage map and reference quality genome sequence for a new race in the wheat pathogen Pyrenophora tritici-repentis." Fungal Genet Biol **152**: 103571.

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nat Protoc **10**(6): 845-858.

Kim, D. and S. L. Salzberg (2011). "TopHat-Fusion: an algorithm for discovery of novel fusion transcripts." Genome Biol **12**(8): R72.

Kim, Y. M. and S. E. Strelkov (2007). "Heterologous expression and activity of Ptr ToxB from virulent and avirulent isolates of Pyrenophora tritici-repentis." Canadian Journal of Plant Pathology **29**(3): 232-242.

Klotzl, F. and B. Haubold (2020). "Phylonium: fast estimation of evolutionary distances from large samples of similar genomes." Bioinformatics **36**(7): 2040-2046.

Kohany O, Gentles AJ, Hankus L and Jurka J (2006). "Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor." BMC Bioinformatics **7**: 474.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman and A. M. Phillippy (2017). "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation." Genome Res **27**(5): 722-736.

Koski LB, Gray MW, Lang BF and Burger G (2005). "AutoFACT: an automatic functional annotation and classification tool." BMC Bioinformatics **6**: 151.

Lamari, L. and C. C. Bernier (1989). "Virulence of isolates of Pyrenophora tritici-repentis on 11 wheat cultivars and cytology of the differential host reactions." Canadian Journal of Plant Pathology **September 1989**: 284-290.

Lamari, L. and C. C. Bernier (1991). "Genetics of Tan Necrosis and Extensive Chlorosis in Tan Spot of Wheat Caused by Pyrenophora-Tritici-Repentis." Phytopathology **81**(10): 1092-1095.

Lamari, L., J. Gilbert and A. Tekauz (1998). "Race differentiation in Pyrenophora tritici-repentis and survey of physiologic variation in western Canada." Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie **20**(4): 396-400.

Lamari, L., R. Sayoud, M. Boulif and C. C. Bernier (1995). "Identification of a new race in Pyrenophora tritici-repentis: implications for the current pathotype classification system." Canadian Journal of Plant Pathology **17**(4): 312-318.

Lamari, L. and S. E. Strelkov (2010). "The wheat/Pyrenophora tritici-repentis interaction: progress towards an understanding of tan spot disease." Canadian Journal of Plant Pathology **32**(1): 4-10.

Li H and Durbin R (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.
Manning, V. A., I. Pandelova, B. Dhillon, L. J. Wilhelm, S. B. Goodwin, A. M. Berlin, M. Figueroa, M. Freitag, J. K. Hane, B. Henrissat, W. H. Holman, C. D. Kodira, J. Martin, R. P. Oliver, B. Robbertse, W. Schackwitz, D. C. Schwartz, J. W. Spatafora, B. G. Turgeon, C. Yandava, S. Young, S. Zhou, Q. Zeng, I. V. Grigoriev, L. J. Ma and L. M. Ciuffetti (2013). "Comparative genomics of a plant-pathogenic fungus, Pyrenophora tritici-repentis, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence." G3 (Bethesda) **3**(1): 41-63.
McDonald, M. C., D. Ahren, S. Simpfendorfer, A. Milgate and P. S. Solomon (2017). "The discovery of the virulence gene ToxA in the wheat and barley pathogen Bipolaris sorokiniana." Mol Plant Pathol **19**(2): 432-439.
Moffat, C. S., P. T. See and R. P. Oliver (2014). "Generation of a ToxA knockout strain of the wheat tan spot pathogen Pyrenophora tritici-repentis." Mol Plant Pathol **15**(9): 918-926.
Moolhuijzen, P., J. Lawrence and S. R. Ellwood (2021). "Potentiators of disease during barley infection by Pyrenophora teres f. teres in a susceptible interaction." Mol Plant Microbe Interact.
Moolhuijzen, P., P. T. See, J. K. Hane, G. Shi, Z. Liu, R. P. Oliver and C. S. Moffat (2018). "Comparative genomics of the wheat fungal pathogen Pyrenophora tritici-repentis reveals chromosomal variations and genome plasticity." BMC Genomics **19**(1): 279.
Moolhuijzen, P., P. T. See and C. S. Moffat (2018). "Exploration of wheat and pathogen transcriptomes during tan spot infection." BMC Res Notes **11**(1): 907.
Moolhuijzen, P., P. T. See and C. S. Moffat (2019). "A new PacBio genome sequence of an Australian Pyrenophora tritici-repentis race 1 isolate." BMC Res Notes **12**(1): 642.
Moolhuijzen, P., P. T. See and C. S. Moffat (2020). "PacBio genome sequencing reveals new insights into the genomic organisation of the multi-copy ToxB gene of the wheat fungal pathogen Pyrenophora tritici-repentis." BMC Genomics **21**(1): 645.
Moolhuijzen, P. M., P. T. See, R. P. Oliver and C. S. Moffat (2018). "Genomic distribution of a novel Pyrenophora tritici-repentis ToxA insertion element." PLoS One **13**(10): e0206586.
Murray GM and Brennan JP (2009). "Estimating disease losses to the Australian wheat industry." Australasian Plant Pathology **38**(6): 558–570.
Osbourn, A. (2010). "Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation." Trends Genet **26**(10): 449-457.
Outram, M. A., P. S. Solomon and S. J. Williams (2021). "Pro-domain processing of fungal effector proteins from plant pathogens." PLoS Pathog **17**(10): e1010000.
Outram, M. A., Y. C. Sung, D. Yu, B. Dagvadorj, S. A. Rima, D. A. Jones, D. J. Ericsson, J. Sperschneider, P. S. Solomon, B. Kobe and S. J. Williams (2021). "The crystal structure of SnTox3 from the necrotrophic fungus Parastagonospora nodorum reveals a unique effector fold and provides insight into Snn3 recognition and pro-domain protease processing of fungal effectors." New Phytol.
Petersen TN, Brunak S, von Heijne G and Nielsen H (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nature Methods **8**(10): 785-786.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez (2005). "InterProScan: protein domains identifier." Nucleic Acids Res **33**(Web Server issue): W116-120.

Retief, J. D. (2000). "Phylogenetic analysis using PHYLIP." Methods Mol Biol **132**: 243-258.

Richards, J. K., N. A. Wyatt, Z. Liu, J. D. Faris and T. L. Friesen (2017). "Reference Quality Genome Assemblies of Three Parastagonospora nodorum Isolates Differing in Virulence on Wheat." G3 (Bethesda).

Rotkiewicz, P. (2007). iMol Molecular Visualization Program. http://www.pirx.com/iMol

RStudio-Team. (2020). "RStudio: Integrated Development Environment for R." 1.3.1093. from http://www.rstudio.com/.

Rybak, K., P. T. See, H. T. Phan, R. A. Syme, C. S. Moffat, R. P. Oliver and K. C. Tan (2017). "A functionally conserved Zn2 Cys6 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 **18**(3): 420-434.

See, P. T., K. A. Marathamuthu, E. M. Iagallo, R. P. Oliver and C. S. Moffat (2018). "Evaluating the importance of the tan spot ToxA-Tsn1 interaction in Australian wheat varieties." Plant Pathology **67**(5): 1066-1075.

Seppey, M., M. Manni and E. M. Zdobnov (2019). "BUSCO: Assessing Genome Assembly and Annotation Completeness." Methods Mol Biol **1962**: 227-245.

Shi, G. S., G. Kariyawasam, S. Liu, Y. Leng, S. Zhong, S. Ali, P. Moolhuijzen, C. Moffat, J. B. Rasmussen, T. L. Friesen, J. D. Faris and Z. Liu (2022). "A conserved hypothetical gene is required but not sufficient for Ptr ToxC production in Pyrenophora tritici-repentis." Mol Plant Microbe Interact.

Shiryev, S. A., J. S. Papadopoulos, A. A. Schaffer and R. Agarwala (2007). "Improved BLAST searches using longer words for protein seeding." Bioinformatics **23**(21): 2949-2951.

Slater, G. S. and E. Birney (2005). "Automated generation of heuristics for biological sequence comparison." BMC Bioinformatics **6**: 31.

Sperschneider, J. and P. Dodds (2021). "EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes." Mol Plant Microbe Interact.

Sperschneider, J., P. N. Dodds, D. M. Gardiner, K. B. Singh and J. M. Taylor (2018). "Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0." Mol Plant Pathol **19**(9): 2094-2110.

Strelkov, S. E., L. Lamari and G. M. Ballance (1999). "Characterization of a host-specific protein toxin (Ptr ToxB) from Pyrenophora tritici-repentis." Molecular Plant-Microbe Interactions **12**(8): 728-732.

Sullivan, M. J., N. K. Petty and S. A. Beatson (2011). "Easyfig: a genome comparison visualizer." Bioinformatics **27**(7): 1009-1010.

Syme, R. A., A. Martin, N. A. Wyatt, J. A. Lawrence, M. J. Muria-Gonzalez, T. L. Friesen and S. R. Ellwood (2018). "Transposable Element Genomic Fissuring in Pyrenophora teres Is Associated With Genome Expansion and Dynamics of Host-Pathogen Genetic Interactions." Front Genet **9**: 130.

Syme, R. A., K. C. Tan, J. K. Hane, K. Dodhia, T. Stoll, M. Hastie, E. Furuki, S. R. Ellwood, A. H. Williams, Y. F. Tan, A. C. Testa, J. J. Gorman and R. P. Oliver (2016). "Comprehensive Annotation of the Parastagonospora nodorum Reference

Genome Using Next-Generation Genomics, Transcriptomics and Proteogenomics." PLoS One **11**(2): e0147221.
Team", R. C. (2021). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.
Testa, A. C., J. K. Hane, S. R. Ellwood and R. P. Oliver (2015). "CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts." BMC Genomics **16**: 170.
Testa, A. C., R. P. Oliver and J. K. Hane (2016). "OcculterCut: A Comprehensive Survey of AT-Rich Regions in Fungal Genomes." Genome Biol Evol **8**(6): 2044-2064.
Tuori, R. P., T. J. Wolpert and L. M. Ciuffetti (1995). "Purification and immunological characterization of toxic components from cultures of Pyrenophora tritici-repentis." Mol Plant Microbe Interact **8**(1): 41-48.
Urban, M., A. Cuzick, K. Rutherford, A. Irvine, H. Pedro, R. Pant, V. Sadanadan, L. Khamari, S. Billal, S. Mohanty and K. E. Hammond-Kosack (2017). "PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database." Nucleic Acids Res **45**(D1): D604-D610.
Vernikos, G., D. Medini, D. R. Riley and H. Tettelin (2015). "Ten years of pan-genome analyses." Curr Opin Microbiol **23**: 148-154.
Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young and A. M. Earl (2014). "Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement." PLoS One **9**(11): e112963.

Figure legends

Fig. 1. Whole genome analysis of Ptr isolates. A) Geographic source and number of Ptr isolates currently available and analysed. Legend shows the number of isolates. B) Whole genome phylogenetic tree of Ptr isolates from Illumina sequencing (Alg130, T199, T205, Alg215, CC142, EW306-2-1, EW4-4, EW7m1, DW7, Pt-1C-BFP, Ptr239, Ptr11137, Ptr5213, SD20, SN001A, SN001C, SN002B), PacBio Technologies (86-124, 90-2, AR CrossB10, Biotrigo9-1, DW5, Ls13-192, M4 and V1) and Oxford Nanopore Technologies (Ptr134). The unrooted neighbour joining phylogenetic tree displays clades for the European (violet) and North Africa (tan) isolates. Geographic source of the other isolates are Australian (blue), USA (green), Canada (red) and Brazil (purple). The race 4 isolates (Ls13-192, 90-2 and SD20) have the greatest

distance from the clade of known effector producing isolates. C) Unrooted Neighbour joining phylogenetic tree for Ptr (purple clade), *Pyrenophora teres* (*P. teres* f. *maculata* (Ptm) and *P. teres* f. *teres* (Ptt)) (orange clade), *Bipolaris* (*B. sorokiniana* (Bs1-3 and Q7399), *B. maydis* (Bm-ATCC and Bm-C5) and *Bipolaris zeicola* (Bz) (green clade), *Parastagonospora nodorum* (Sn4, Sn15 and Sn79) (yellow clade), *Leptosphaeria maculans* (Lm) and *Zymoseptoria tritici* (Zt) isolates. The branches for race 4 isolates not producing known effectors (Ls13-192, 90-2 and SD20) are highlighted (blue) within the Ptr clade. D) Circular plots show 10 kb regions of absence plotted for the Ptr isolates genomes sequenced using long-read technologies (PacBio and Oxford Nanopore Technology) as compared with the chromosomes of the reference Ptr genome of isolate M4. Isolates are coloured by race. Three regions of interest are highlighted in grey and zoomed at 20x for chromosome 2 and chromosome 1, and 40x for chromosome 9.

Fig. 2. Ptr pangenome predicted mRNA correlation plots for gene sequence percentage identity (A) and gene copy number (B). Ptr isolates from Illumina sequencing (Alg130, T199, T205, Alg215, CC142, EW306-2-1, EW4-4, EW7m1, DW7, Pt-1C-BFP, Ptr239, Ptr11137, Ptr5213, SD20, SN001A, SN001C, SN002B), PacBio Technologies (86-124 (Ptr86-124), 90-2 (Ptr90-2), AR CrossB10, Biotrigo9-1, DW5, Ls13-192, M4 and V1) and Oxford Nanopore Technologies (Ptr134).

Fig. 3. Ptr isolate M4 *ToxC1* locus and 2 kb flanking sequence region alignment to twelve other Ptr ToxC producing isolates. The Biotrigo9-1 *ToxC1* region has two large insertions within and downstream of *ToxC1*. Nucleotide sequence alignments (blue) between the *ToxC1* region for Ptr isolates (top to bottom M4, AR CrossB10,

Biotrigo9-1, Pt-1C-BFP, V1, Ptr134, Alg215, CC142, EW306-2-1, EW7m1, SN0001A, SN0001C and SN0002B) (black lines). The M4 genes are shown as red arrows. The light blue alignment segments are regions of low identity among the isolates, while the crossed regions indicate a repeat region in each sequence.

Fig. 4. The predicted protein sequence and structural alignments of SnTox3 and the isolate M4 protein KAF7577476. A) Multiple protein sequence alignment of SnTox3, Ptt CAA9973983.1 (W1-1), Ptm CAA9957881.1 (SG1) and Ptr KAF7577476 (M4). The Kex2 motif conservation is shown boxed in red. Only four cysteine residues were conserved across the four species (black asterisks) and those not conserved (red asterisks) are shown below the alignment for *P. nodorum* and above for the *Pyrenophora* spp. B) The known 3-D protein structure for SnTox3 (PDB 6WES). C) The 3-D structure for KAF7577476 as predicted by Phyre2. D) Superimposed structural alignment (yellow) of SnTox3 and KAF7577476 with an RSDM of 1.14 A°.

Supporting Information

Supplementary data

Supplementary data 1 XLSX. Predicted effector genes for *Pyrenophora tritici-repentis* genomes.

Supplementary data 2 PDF. List of URLs for publicly available isolate genomes downloaded from NCBI for this study.

Supplementary data 3 XLSX. Predicted biosynthetic gene clusters for *Pyrenophora tritici-repentis* genomes.

Supplementary data 4 XLSX. Orthologous protein clusters for *Pyrenophora tritici-repentis* genomes.

Supplementary data 5 XLSX. Phyre2 three-dimensional protein modelling for *Pyrenophora tritici-repentis* predicted effector proteins.

Supplementary figures

Supplementary Fig. S1. BUSCO quantitative assessment of the completeness of genome assemblies in terms of expected gene content.

Supplementary Fig. S2. Ptr *ToxB* and *toxb* nucleotide sequence alignments.

Supplementary Fig. S3. Ptr ToxB and toxb protein sequence alignments.

Supplementary Fig. S4. Ptr Alg215 isolate partial *ToxB* sequence alignments show Alg215 *ToxB* is truncated at the 5' end of the sequence. A) The first 162 nucleotide bases of Alg215 scaffold 03337 sequence aligned to the *ToxB* coding sequence (CDS) (99-261 bp). A *ToxB* single nucleotide polymorphism (SNP; at the 149 bp position) shows a thiamine nucleotide change to guanine (T > G). B) The Alg215 *ToxB* region protein translated (1-94 aa) aligned to the ToxB protein sequence (1-87 bp). A nonsynonymous amino acid residue change (I>R) at ToxB residue position 50 is beyond the ToxB signal peptide cleavage site between amino acid positions 23 and 24.

Supplementary Fig. S5. Ptr plant leaf infection assays to identify isolate ToxA production by the development of necrosis symptoms on the differential wheat cultivar Glenlea (left hand side) and ToxB production by chlorosis symptoms on the differential wheat line 6B662 (sensitive) (right hand side).

Supplementary Fig. S6. Ptr plant infection assays to confirm no symptoms on response to the differential wheat line Auburn (insensitive) (left hand side) and ToxC production by chlorosis symptoms on the differential wheat line 6B365 (right hand side).

Supplementary Fig. S7. Closer examination a large 1 Mb distal region on isolate M4 contig 1 and many smaller regions were absent in isolates Alg130, T199 and T205 but present in Alg215.

Supplementary Fig. S8. Phylogenetic analysis of publicly available ascomycete genomes downloaded from NCBI.

Supplementary Fig. S9. Ptr isolate M4 ToxA horizontal transfer and flanking genomic region (400 kb) alignments for race 1, 2, 4, 5 and unknown races. Break points are displayed for the large 128 kb insertion in the M4 isolate. The *ToxA* horizontal transfer region is absent in all non-ToxA producing races (race 4, 5 and unknown) and present in ToxA producing races (race 1 and 2).

Supplementary Fig. S10. Circular plots show 10 kb regions of absence for Ptr isolates as compared to M4, coloured by race. The left plot shows assembled isolate genomes sequenced from long-read technologies, PacBio and Oxford Nanopore Technology. The right plot displays all the genomes.

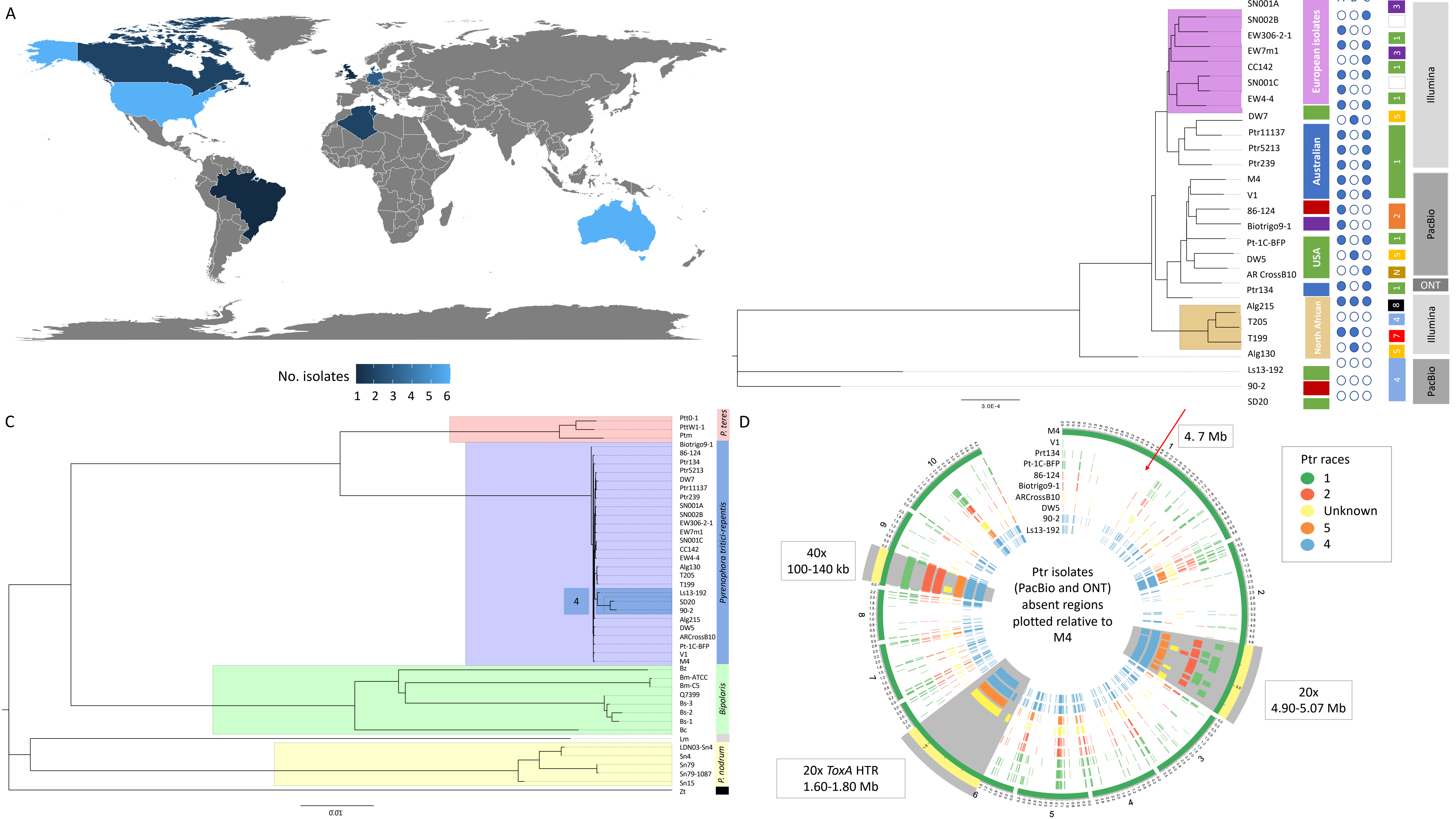
Supplementary Fig. S11. DW5 chromosome 11 *ToxB2* and flanking genomic region alignments for race 1, 2, 4, 5 and unknown races. Slide 1, EasyFig Blastn alignments for 400 kb region. Top to bottom DW5 aligned to race 4 (90-2 and Ls13-192), race unknown (AR CrossB10), race 2 (86-124 and Biotrigo9-1) and race 1 (V1, M4 and Ptr134). Slide 2, NUCmer sequence dot plot for 200 kb region.

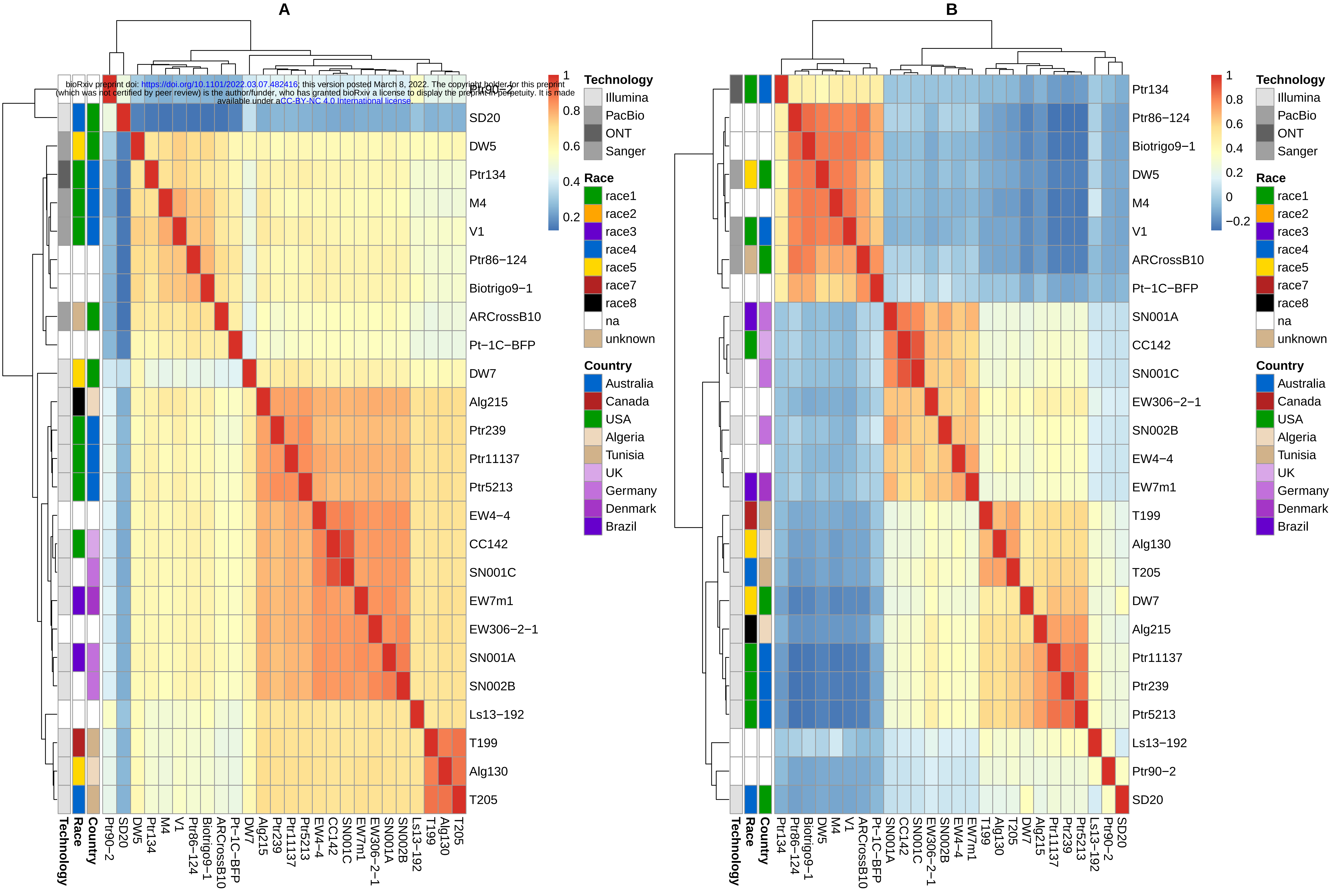
Supplementary Fig. S12. Ptr loci found only in PacBio sequenced ToxC-producing isolates. Loci are absent in non-ToxC producing Ptr isolates Ls13-192 (race 4), 86-124 and Biotrigo9-1 (race 2), DW5 (race 5) and are present in ToxC producing isolates AR CrossB10 (AR, race unknown), V1 and M4 (race 1) isolates. M4 isolate had multiple gene copies, while some genes were absent in Pt-1C-BFP (race 1).

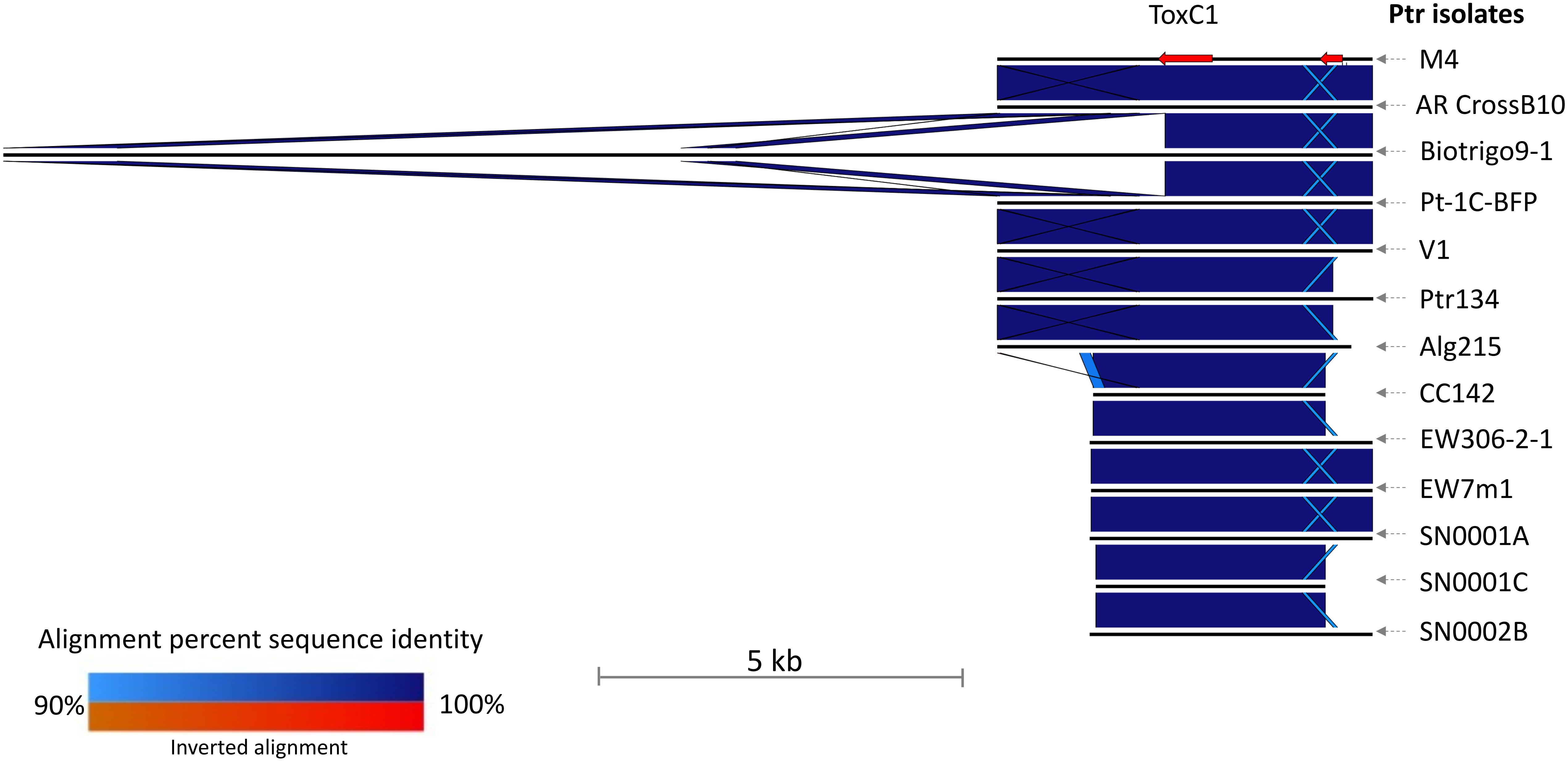
Supplementary Fig. S13. RNA expression for *Pyrenophora tritici-repentis* isolate M4 for the M4 ToxC producing isolate specific gene cluster 64 kb region (red). The read coverage and alignments show RNA expression *in vitro* (top) and *in planta* (below) on chromosome 9 (CM025803.1).

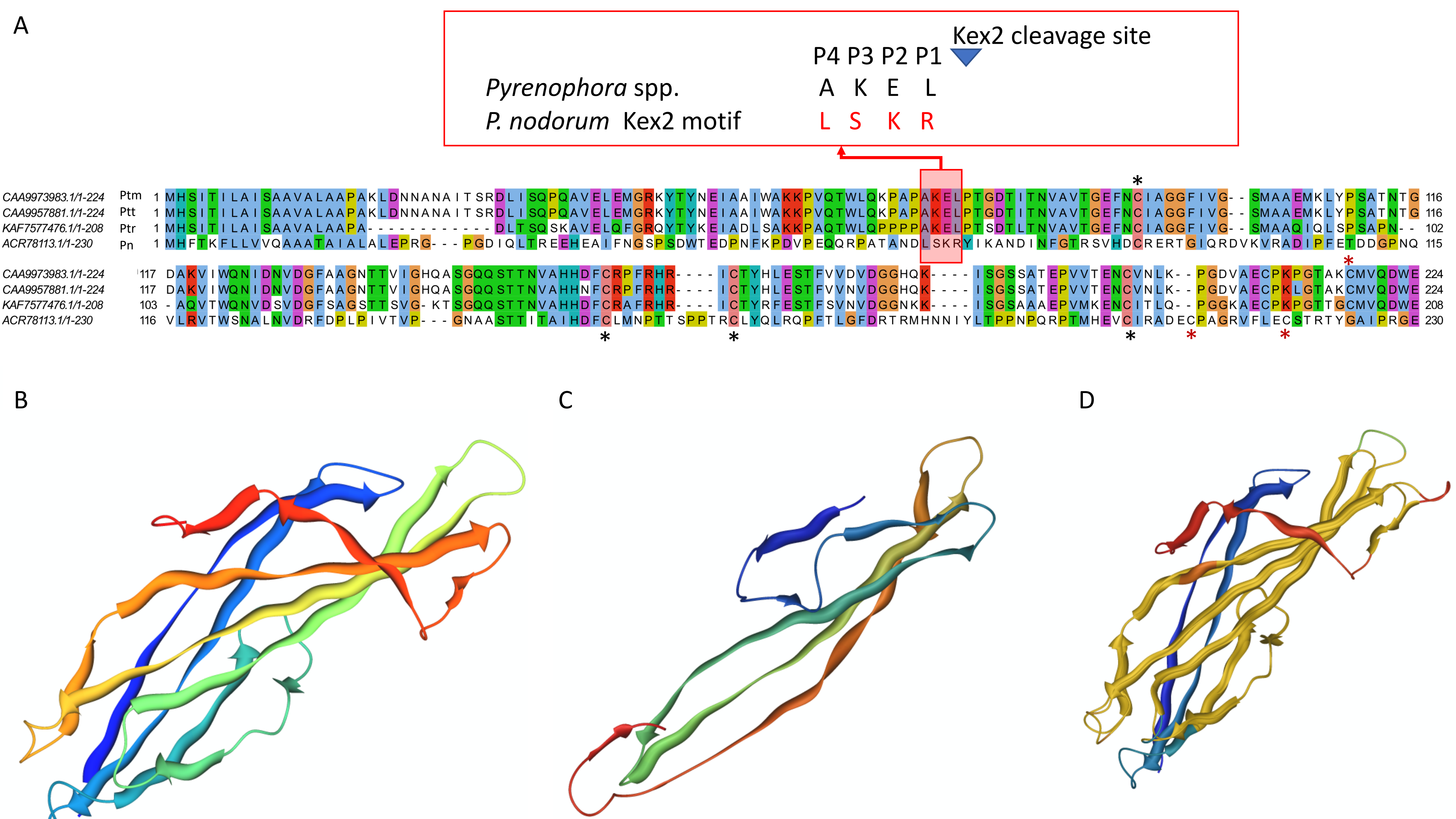
Supplementary Fig. S14. Ptr isolate Biotrigo9-1 *ToxC1* genomic region on contig 12 (43,362-60,409 bp) shows *ToxC1* is disrupted by a large single insertion (5,348 bp in size) positioned from 45,946 to 51,292 bp that carries a nested long terminal repeat (LTR) retrotransposon and the motifs for the Copia retrotransposon transposable element (TE).

Supplementary Fig. S15. Ptr pangenome predicted effectors specific to race 4 and specific to known effector producing isolates (non-race 4). A) Boxplot of protein lengths. B) Boxplot of effector probability scores. C) Violin plot of lengths. D) Violin plot of effector probability scores.


A
No. isolates
1 2 3 4 5 6
SN001A
SN002B
EW306-2-1
EW7m1
CC142
SN001C
EW4-4
DW7
Ptr11137
Ptr5213
Ptr239
M4
V1
86-124
Biotrigo9-1
Pt-1C-BFP
DW5
AR CrossB10
Ptr134
Alg215
T205
T199
Alg130
Ls13-192
90-2
SD20
European isolates
Australian
USA
North African
Illumina
PacBio
ONT
Illumina
PacBio
3.0E-4
C
Ptt0-1
PttW1-1
Ptm
Biotrigo9-1
86-124
Ptr134
Ptr5213
DW7
Ptr11137
Ptr239
SN001A
SN002B
EW306-2-1
EW7m1
SN001C
CC142
EW4-4
Alg130
T205
T199
Ls13-192
SD20
90-2
Alg215
DW5
ARCrossB10
Pt-1C-BFP
V1
M4
Bz
Bm-ATCC
Bm-C5
Q7399
Bs-3
Bs-2
Bs-1
Bc
Lm
LDN03-Sn4
Sn4
Sn79
Sn79-1087
Sn15
Zt
4
P. teres
Pyrenophora tritici-repentis
Bipolaris
P. nodrum
0.01
D
M4
V1
Prt134
Pt-1C-BFP
86-124
Biotrigo9-1
ARCrossB10
DW5
90-2
Ls13-192
4. 7 Mb
Ptr races
1
2
Unknown
5
4
40x
100-140 kb
Ptr isolates
(PacBio and ONT)
absent regions
plotted relative to
M4
20x
4.90-5.07 Mb
20x ToxA HTR
1.60-1.80 Mb


A
B
Technology
Illumina
PacBio
ONT
Sanger
Race
race1
race2
race3
race4
race5
race7
race8
na
unknown
Country
Australia
Canada
USA
Algeria
Tunisia
UK
Germany
Denmark
Brazil
Ptr90−2
SD20
DW5
Ptr134
M4
V1
Ptr86−124
Biotrigo9−1
ARCrossB10
Pt−1C−BFP
DW7
Alg215
Ptr239
Ptr11137
Ptr5213
EW4−4
CC142
SN001C
EW7m1
EW306−2−1
SN001A
SN002B
Ls13−192
T199
Alg130
T205


ToxC1
Ptr isolates
M4
AR CrossB10
Biotrigo9-1
Pt-1C-BFP
V1
Ptr134
Alg215
CC142
EW306-2-1
EW7m1
SN0001A
SN0001C
SN0002B
Alignment percent sequence identity
90%
100%
Inverted alignment
5 kb


A
Kex2 cleavage site
P4 P3 P2 P1
Pyrenophora spp. A K E L
P. nodorum Kex2 motif L S K R
CAA9973983.1/1-224 Ptm 1 MHSITILAISAAVALAAPAKLDNNANAITSRDLISQPQAVELEMGRKYTYNEIAAIWAKKPVQTWLQKPAPAKELPTGDTITNVAVTGEFNCIAGGFIVG--SMAAEMKLYPSATNTG 116
CAA9957881.1/1-224 Ptt 1 MHSITILAISAAVALAAPAKLDNNANAITSRDLISQPQAVELEMGRKYTYNEIAAIWAKKPVQTWLQKPAPAKELPTGDTITNVAVTGEFNCIAGGFIVG--SMAAEMKLYPSATNTG 116
KAF7577476.1/1-208 Ptr 1 MHSITILAISAAVALAAPA------------DLTSQSKAVELQFGRQYTYKEIADLSAKKPAQTWLQPPPPAKELPTSDTLTNVAVTGEFNCIAGGFIVG--SMAAQIQLSPSAPN-- 102
ACR78113.1/1-230 Pn 1 MHFTKFLLVVQAAATAIALALEPRG---PGDIQLTREEHEAIFNGSPSDWTEDPNFKPDVPEQQRPATANDLSKRYIKANDINFGTRSVHDCRERTGIQRDVKVRADIPFETDDGPNQ 115
CAA9973983.1/1-224 117 DAKVIWQNIDNVDGFAAGNTTVIGHQASGQQSTTNVAHHDFCRPFRHR----ICTYHLESTFVVDVDGGHQK----ISGSSATEPVVTENCVNLK--PGDVAECPKPGTAKCMVQDWE 224
CAA9957881.1/1-224 117 DAKVIWQNIDNVDGFAAGNTTVIGHQASGQQSTTNVAHHNFCRPFRHR----ICTYHLESTFVVNVDGGHQK----ISGSSATEPVVTENCVNLK--PGDVAECPKLGTAKCMVQDWE 224
KAF7577476.1/1-208 103 -AQVTWQNVDSVDGFSAGSTTSVG-KTSGQQSTTNVAHHDFCRAFRHR----ICTYRFESTFSVNVDGGNKK----ISGSAAAEPVMKENCITLQ--PGGKAECPKPGTTGCMVQDWE 208
ACR78113.1/1-230 116 VLRVTWSNALNVDRFDPLPIVTVP---GNAASTTITAIHDFCLMNPTTSPPTRCLYQLRQPFTLGFDRTRMHNNIYLTPPNPQRPTMHEVCIRADECPAGRVFLECSTRTYGAIPRGE 230
B
C
D