A new reference genome shows the one-speed genome structure of the barley pathogen *Ramularia collo-cygni*

Remco Stam[1*§], Martin Münsterkötter[4,5*], Saurabh Dilip Pophaly[2%],Like Fokkens[6], Hind Sghyer[1], Ulrich Güldener[3], Ralph Hückelhoven[1], Michael Hess[1]

[1]Chair of Phytopathology, [2]Section of Population Genetics, [3] Department of Bioinformatics

School of Life Sciences Weihenstephan, Technische Universität München, Germany

[4]Functional Genomics and Bioinformatics, University of Sopron, Hungary

[5]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Germany

[6]Molecular Plant Pathology, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands.

* contributed equally to this work

§ correspondence:

Remco Stam: stam@wzw.tum.de

% current address: Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Sweden & Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universität München, Germany

Abstract

Ramularia leaf spot has recently emerged as a major threat to barley production world-wide, causing 25% yield loss in many barley growing regions. Here we provide a new reference genome of the causal agent, the Dothideomycete *Ramularia collo-cygni.* The assembly of 32 Mb consists of 78 scaffolds. We used RNA-seq to identify 11,622 genes of which 1303 and 282 are coding for predicted secreted proteins and putative effectors respectively.

The pathogen separated from its nearest sequenced relative, *Zymoseptoria tritici* about 27 million years ago. We calculated the divergence of the two species on protein level and see remarkably high synonymous and nonsynonymous divergence. Unlike in many other plant pathogens, the comparisons of transposable elements and gene distributions, show a very homogeneous genome for *R. collo-cygni*. We see no evidence for higher selective pressure on putative effectors or other secreted proteins and repetitive sequences are spread evenly across the scaffolds. These findings could be associated to the predominantly endophytic life-style of the pathogen. We hypothesize that *R. collo-cygni* only recently became pathogenic and that therefore its genome does not yet show the typical pathogen characteristics. Due to its high scaffold length and improved CDS annotations, our new reference sequence provides a valuable resource for the community for future comparative genomics and population genetics studies.

Introduction

The filamentous ascomycete fungus *Ramularia collo-cygni* was first described in 1893 as *Ophiocladium hordei* (Cavara 1893). It is the biotic agent of Ramularia Leaf Spot (RLS) (Oxley & Havis 2004), a disease typically occurring late in the growing season on the upper canopy (Salamati & Reitan 2006). Since the mid-1980's it has become the major pathogen in many barley growing regions worldwide and quickly developed resistance to major fungicides (Matusinsky et al. 2011; Havis et al. 2015; Piotrowska et al. 2017). It can now be detected in barley samples worldwide (Havis et al. 2015) and in infected fields it is estimated to cause losses around 25% of the yield potential through a significant decrease of kernel size and quality (Harvey 2002).

A draft genome assembly of *R. collo-cygni* strain DK05, isolated in Denmark, had been published previously (McGrann et al. 2016). Like its closest sequenced relative *Zymoseptoria tritici, R. collo-cygni* has few plant cell wall degrading enzyme genes and a large number of gene clusters associated with secondary metabolite production. These findings are thought to be linked to the relatively long period of asymptomatic growth of both pathogens inside the host.

We present an independent draft genome of a strain isolated in southern Uruguay. The significantly increased scaffold size enabled several analyses related to genome architecture. Moreover, our improved annotation allows for more reliable identification of genes that are under positive selection and may be involved in pathogenicity.

Results

Genome properties

Combining short distance with long jumping distance libraries, we obtained a ~32 Mb assembly of *R. collo-cygni* isolate Urug2, in 78 scaffolds with an N50 scaffold size of 2.1 Mb, compared to 576 scaffolds and an N50 of 0.21 in the previous DK05 assembly (Table 1). Dot plot analysis comparing both Urug2 to DK05, shows strong linearity suggesting little to no over-assembly (Figure S1). To bolster gene annotations we sequenced mRNA isolated from *R. collo-cygni* grown under six axenic conditions to generate a diverse set of transcripts. Using these data mapped to the genome, the annotation was manually corrected gene by gene, yielding a more reliable set of gene models (example shown in Figure S2). The curated *R. collo-cygni* genome contains 11,622 protein coding genes of which 11,125 show expression evidence (95.7%, Table S1). We predicted 1303 secreted proteins ranging in length from 41 to 3256 amino-acids, representing around 9% of the predicted proteome, among which 282 effector candidates genes (putative effectors) (2%) (Table S2). These numbers are an increase over the previous *R. collo-cygni* assembly, that contained 1053 putative secreted proteins and 150 effector candidates.

The general genome features are comparable to other Dothideomycete plant pathogens with available genome sequences (*Parastagonospora nodorum* SN15 (Hane et al. 2007; Syme et al. 2013), *Pyrenophora tritici-repentis* BFP-ToxAC (Manning et al. 2013) *, Pyrenophora teres f. teres* 0-1 *(Wyatt et al. 2018)* and *Zymoseptoria tritici* IPO323 (Goodwin et al. 2011) (Table 1). We estimated genome completeness using BUSCO (version 3.0.1). 96.8% of the genes were detected as complete and single-copy in the *R. collo-cygni* Urug2 assembly. This exceeds the completeness of *R. collo-cygni DK05* (93.9%) and *Z. tritici IPO323* (92.5%) and is similar to *Pyrenophora teres f.teres* 0-1 (Table S3).

Gene expression analysis

We saw similar levels of gene expression in all six tested media, with a slightly larger fraction of genes expressed in Barley Straw Agar (BSA), a host-mimicking agar medium compared to neutral or pH-adjusted media (Table S1. Figure S3A). We found a larger number of differentially expressed genes (DEGs) (> 4-fold change) in BSA (Figure S3B). We expected that gene expression in BSA most closely resembles infection of the plant. Indeed, we found that the fraction of putative secreted proteins and effector candidates is two times higher in the BSA differentially expressed genes than in the genome as a whole (resp. 22 and 5%).

Strong divergence from *Zymoseptoria tritici*

We reconstructed a phylogenetic tree of *R. collo-cygni*, with 15 more closely and three more distantly related species (Figure 1) and confirm that *R. collo-cygni* falls within the Mycosphaerellaceae clade of the Dothideomycete class. *R. collo-cygni* (Rc) diverged from the closest sequenced relative, *Z. tritici* (Zt) 27 million years ago. To gain insight in the differentiation of *R. collo-cygni* from *Z. tritici*, we calculated the ratio of non-synonymous over synonymous substitutions (dN/dS) (Figure 2). The very high dS indicate that these two species have significantly diverged since the split. When comparing the dN/dS between the two species for the putative secreted proteins and putative effectors, we find that the putative secreted proteins show a slightly higher dN/dS ratio than non secreted proteins (Dunn's multiple comparisons Kruskal-Wallis test, $p = 4.7*10^{-16}$). For the effectors this difference is not significant ($p = 0.16$). Effectors and secreted proteins also show no significant differences ($p=0.2$). In terms of absolute values and outliers, there are no putative effectors that stand out. Similar results can be observed when comparing differentially expressed genes on BSA. As mentioned above, these genes are hypothesised to be important for virulence on barley, yet there are no significant differences in dN/dS between these BSA up- or down regulated genes and non-DEGs or between up or down regulated genes in general (S Figure 3C, Dunn's multiple comparisons Kruskal-Wallis test, $p > 0.01$,).

The *R. collo-cycni* genome is relatively repeat poor and homogeneous.

We compared the content of non-coding sequences and repeat sequences like DNA transposons and other transposable elements (TEs) (Table S4). In terms of repeat sequence content, *R. collo-cygni* is placed at the low end of the spectrum amongst Dothideomycetes. Only 6% of the genome consists of TEs, whereas in *P. tritici-repentis*, *P. teres f. teres and Z. tritici* this is 21, 38 and 17%, respectively. Next we compared the distance of predicted genes to its nearest repeat sequence as well as the general intergenic distance. Close association of genes to TEs and large intergenic distance for regions with high effector content are features of a so-called "two-speed-genome" and often associated with accelerated evolution (Raffaele & Kamoun 2012). Figure 3A shows that intergenic distances are not differently distributed between putative effectors genes or other genes and the mean values for the difference to the nearest TE for putative effectors and putative secreted proteins are not significantly different from the distances for not secreted proteins (effector: 3': 1739 bp, 5': 1646 bp, secreted 3': 1659 bp, 5' 1715 bp, non-secreted: 3' 1724, 5' 1727) (Dunn's multiple comparisons Kruskal-Wallis test, $p > 0.1$).

Lastly, we also do not find a significant correlation between the number of TEs per kb/scaffold or the number of effectors or secreted proteins (Spearman rho, $p > 0.01$, Figure 3C).

Discussion

A first draft genome of *R. collo-cygni* (isolate DK05) had been available since 2016. The data suggested a genetic composition that might at least partially explain the lifestyle of *R. collo-cygni*, which is characterized by a long endophytic phase throughout the life cycle of the host and an intense parasitic phase during crop senescence (McGrann et al. 2016). We generated an independent draft genome and annotation for another isolate (Urug2) to get better insights in the *R. collo-cygni* genome structure. We assembled the 32 Mb genome into only 78 scaffolds with *11,622* high confidence genes. Our expression data greatly helped with gene annotations and provide interesting insights in genes expressed under different axenic conditions. This will help researchers to verify target gene candidates for functional studies, yet to truly understand gene expression during the infection process, additional RNA-Seq from infected plant tissue will be required.

Our sequencing approach allowed for comparative studies and confirmed that unlike many other pathogens *R. collo-cygni* did not undergo any genome expansions since it diverged from its nearest sequenced sister species 27 million years ago. Unlike what can be seen between certain fungal and oomycete species, where the numbers of genes in some effector families differ up to an order of magnitude (Stam et al. 2013) the numbers of putative effectors in *R. collo-cygni* are comparable with that of related fungi.

We performed pairwise comparisons of the coding sequences of *R. collo-cygni* and the related wheat pathogen *Z. tritici*. The dN/dS ratio has a simple and intuitive interpretation of selection pressure, but comes with limitations, especially when dS is high (Kryazhimskiy & Plotkin 2008). However, our analyses provide interesting insights. Contrary to the phenomenon observed in a large number of other plant pathogens, we see little evidence for accelerated evolution of secreted proteins, putative effectors or genes that are likely differentially expressed during infection, between *R. collo-cygni* and *Z. tritici*. This is in stark contrast to for example *Colletotrichum species,* where high dN/dS of effectors was associated with the switch from endophytic to parasitic lifestyle (Hacquard et al. 2016). This however, leaves the possibility that this switch is still ongoing *in R. collo-cygni.* The species can also infect other graminaceous hosts, but with less severe symptoms it often appears endophytic *(*Kaczmarek *et al.* 2017*)*

From Z. tritici, R. collo-cygni's nearest sequenced relative, we know that rapid pathogen evolution can often be associated with high repeat content of the genome (Poppe et al. 2015), or close physical association of TEs with putative effector genes, which results in a so-called "two-speed" genome architecture. Also for other

Dothideomycetes like *Parastagonospora nodorum* (Richards et al. 2018) and *P. teres f. teres* (Wyatt et al. 2018) this two speed genome is evident. In *P. tritici-repentis* TE content has even be directly associated with the pathogenicity of the strains (Manning et al. 2013). In *R. collo-cygni* repeat content is low and secreted proteins or putative effectors are not closely associated with TEs. Other examples of typical “one-speed-genome” pathogen is the biotrophic barley pathogen *Blumeria graminis* (Frantzeskakis et al. 2018). However, that species is relatively unrelated and has a very different lifestyle. Comparing the mechanisms that drive evolution of pathogenicity in these two diverse one-speed-genome barley pathogens will be particularly interesting. Also, additional investigation is TEs and the relatedness to host, host specificity and aggressiveness as a pathogen in *R. collo-cygni* and other Dothideomycetes will likely teach us more on how this diverse class of cereal pathogens arose and became successful. Our new reference genome and improved annotation provides a starting point for doing so.

Methods

Genome assembly and annotation

A detailed description of the genome sequencing strategy, assembly and annotation can be found in the supplementary files. In short: Sequencing was done by Eurofins Genomics GmbH, Germany, using a short distance library (SD) (insert size, 500 bp, paired-end sequencing 2x150 bp) and a long jumping distance library (LJD) (jumping distance 8 kbp, paired-end sequencing 2x300 bp). RNA-seq was performed using TruSeq Rapid PE Cluster (PE-402-4001) and the TruSeq Rapid SBS (FC-402-4001) Kits.

The assembly was performed by ALLPATHS-LG (Gnerre et al. 2011) using ~100 fold SD and ~30 fold LJD coverage. Gene models were generated by 1) Fgenesh (Salamov & Solovyev 2000)s 2) GeneMark-ES (Ter-Hovhannisyan et al. 2008) and 3) AUGUSTUS (Stanke et al. 2006). RNA-seq transcripts were assembled using Trinity (Grabherr et al. 2011). Gene models were visualized in Gbrowse (Donlin 2009), allowing manual validation of coding sequences. The best fitting model per locus was selected manually and gene structures were adjusted by splitting or fusing of gene models or redefining exon-intron boundaries if necessary.

The protein coding genes were analyzed and functionally annotated using the PEDANT system (Walter et al. 2008). We combined SecretomeP1.1, (Bendtsen et al. 2004) SignalP3 (Bendtsen et al. 2004b) and SignalP4.0 (Petersen et al. 2011) with TargetP1.1 (Emanuelsson et al. 2007), TMHMM2.0 (Krogh et al. 2001) to predict secreted proteins. For effector prediction, this set was submitted to EffectorP 2.0 (Sperschneider et al. 2016). Repetitive elements were identified using the RepeatModeler pipeline (http://www.repeatmasker.org/RepeatModeler/), which is based on RepeatScout (Price et al. 2005).

Further analyses

Orthologous genes for 10 main housekeeping genes we selected from the proteomes of *Urug2,* 15 closely related species and three more distal species that are available from the FunyBASE (Marthey et al. 2008). We performed a ClustalW pairwise alignment and multiple alingment with standard parameters for the combined proteins of all species (*An= Aspergillus nidulans; Bc = Botrytis cinerea; Bg = Blumeria graminis; Cb = Cercospora beticola; Cg = Colletotrichum graminicola; Ds= Dothistroma septosporum; Fg= Fusarium graminearum; Ff= Fusarium fujikuroi; Mf= Pseudocercospora fijiensis; Mo=Magnaporthe grisea; Pi= Piriformospora indica; Pr= Pyrenophora tritici-repentis; Pt= Pyrenophora teres f. teres; Rc= Ramularia collo-cygni; Sn= Parastagonospora nodorum; Sm= Sphaerulina musiva; Um = Ustalago maydis; Zt= Zymoseptoria*

tritici; Outgroup = *Laccaria bicolor;*) using Mega X (Kumar et al. 2018) Dating of the Zt – Rc split was done using the RelTime method (Tamura et al. 2012), where Lb was used as an outgroup and the data were calibrated using the Um to An split (Taylor & Berbee 2006).

Orthologous genes to all single copy genes were identified in the Zt proteome (BLASTP e-value: 10^{-10}), only reciprocal matches were used and globally aligned using t-coffee (default parameters) (Table S5). Amino acids were replaced by codons from the cds sequence using pal2nal (Suyama et al. 2006). The dN/dS ratios were calculated with PAML (Yn00 command) (Yang & Nielsen 2000). Intergenic distances and TE distances were calculated using bedtools (closest) (Quinlan & Hall 2010). Genome alignments between the assembly of Urug2 and that of DK05 were inferred using nucmer (with --maxmatch, otherwise default options) from the MUMmer package (version 3.23) and alignments that span more than 1 kb and are more than 90% identical in sequence were plotted using Gnuplot.

Acknowledgements

We thank C. Wurmser (Chair of Animal Breeding TUM) for support with sequencing. Special thanks goes to Carolin Hutter and Regina Dittebrand for technical assistance. This work was in part financially supported by grants from the Bavarian State Ministry of Food, Agriculture and Forestry (Projects KL/12/07 and KL/08/07) (to MH, RH), the Bavarian State Ministry of the Environment and Consumer Protection in frame of the project network BayKlimaFit (to RH, subproject 10) the Hungarian Széchenyi 2020 Programme (grant number GINOP-2.3.2-15-2016-00052) (to MM). This work was supported by the German Research Foundation (DFG) and the Technical University of Munich within the Open Access Publishing Funding Programme.

Data availability

The genome assembly, annotation data are available from the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena/data/view/FJUY01000001-FJUY01000078). The RNA-seq data has been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE109135.

Figure Legends

Figure 1: Phylogenetic tree showing the relationship between Ramularia collo-cygni and 16 other fungi.
A) Maximum likelihood phylogenetic tree based on the analysis of 10 housekeeping genes. Reported on the nodes are bootstrap values for 1000 bootstraps. B) Timetree results. Reported on the nodes are the divergence times in mya.
Bc = Botrytis cinerea; Bg = Blumeria graminis; Cg = Colletotrichum graminicola; Cb = Cercospora beticola; Ds= Dothistroma septosporum; Fg= Fusarium graminearum; Ff= Fusarium fujikuroi; Mf= Pseudocercospora fijiensis; Mo=Magnaporthe grisea; Pr= Pyrenophora tritici-repentis; Pt= Pyrenophora teres f. teres; Rc= Ramularia collo-cygni; Sn= Parastagonospora nodorum; Sm= Sphaerulina musiva; Zt= Zymoseptoria tritici
An= Aspergillus nidulans; Um = Ustalago maydis;Pi= Piriformospora indica; Outgroup = Laccaria bicolor;

Figure 2: dN/dS between *R. collo-cygni* and *Z. tritici*
A) Scatter plot of the dS (x axis) against dN (y axis) for each predicted protein in a pairwise comparison between *R. collo-cygni* and *Z. tritici.* B) Violin diagrams of the dN/dS ratio for predicted proteins in a pairwise comparison between *R. collo-cygni* and *Z. tritici.* Data coloured based on whether the proteins are predicted to be putatively secreted proteins, putative effectors or other, non-secreted, proteins. Horizontal bar depicts the median.

Figure 3: Intergenic distance and TEs in *R. collo-cygni*
A) Density plot of the intergenic distances on the 5' end (x axis) against the 3' end (y axis) for each predicted gene on the genome. Distances for all genes are binned in hexagons and coloured using a blue to red scale, red indicating the largest amount of genes per hexagon. Putative effectors are plotted as separate black dots. B) Violin diagrams of the distance to the nearest TE for each gene coding for predicted non-secreted (grey), putative secreted (yellow) or effector (green) proteins. C) Density of putative effector encoding genes (top, green) and transposable elements (bottom, grey) plotted in 10 kb non-overlapping sliding windows along the genome (x axis). Density is defined as number of basepairs that is part of a putative effector or TE in the window (y axis).

Table 1: Summary statistics for our Urug2 genome, as well as other related fungi. Genome statistics were taken from the respective publications for each organism. The number of putative secreted proteins and effector candidates was recalculated for each species except DK05, using our newest pipeline. *DK05 gene/protein models were not publicly available.

Figure S1: Dot plot representation of the comparison of the previously published *R. collo-cygni* DK05 genome and our Urug2 genome. Only alignments that span more than 1 kb in both the query and the reference, and are more than 90% identical are included. Dots are coloured according to % identity as indicated in the colour bar to the right.

Figure S2: Genome browser screenshot showing the alternative gene models called using gene predictors and the accuracy that the RNA-seq brings in choosing the right model. The gene in the red box matches the RNA-Seq data, the alternative call is discarded.

Figure S3: A) The number of genes with expression evidence (fkpm > 1) under different axenic conditions. B) The number of genes that show higher or lower expression values (log2-fold change) under different axenic conditions when compared with standard growth medium. C) Differences in dN/dS between genes that have higher (magenta) and lower (green) expression values on BSA or no changes (grey) compared to other axenic media.

Figure S4 Density of putative effector encoding genes (green) and transposable elements (grey) for scaffolds 21-34 plotted in 10 kb non-overlapping sliding windows along the genome (x axis). Density is defined as number of basepairs that is part of a putative effector or TE in the window (y axis). Generally very few effectors are found on these scaffolds and association with TEs is not evident.

Table S1: Summary of the RNA-seq data. FPKM values are reported for each gene in different media and can be used to calculate differential expression.

Table S2: Gene identifiers of all predicted putative secreted proteins and putative effectors

Table S3: Summary statistics for BUSCO output.

Table S4: Comparison of transposable and repeat elements in *R. collo-cygni* and related species.

Table S5: *R. collo-cygni* and *Z. tritici* orthologs and dN/dS calculations

References

Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. 2004. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. PEDS. 17:349–356. doi: 10.1093/protein/gzh037.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S ren, Bendtsen JD. 2004. Improved Prediction of Signal Peptides: SignalP 3.0. J. Mol. Biol. 340:783–795.

Cavara F. 1893. Über einige parasitische Pilze auf dem Getreide. Z. Für Pflanzenkrankh. 3:16–26.

Donlin MJ. 2009. Using the generic genome browser (GBrowse). Curr. Protoc. Bioinforma. 9.9. 1-9.9. 25.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2:953 – 71.

Frantzeskakis L et al. 2018. Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics. 19:381. doi: 10.1186/s12864-018-4750-6.

Gnerre S et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. 108:1513–1518. doi: 10.1073/pnas.1017351108.

Goodwin SB et al. 2011. Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis Malik, HS, editor. PLoS Genet. 7:e1002070. doi: 10.1371/journal.pgen.1002070.

Grabherr MG et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29:644–652. doi: 10.1038/nbt.1883.

Hacquard S et al. 2016. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. Nat. Commun. 7:ncomms11362. doi: 10.1038/ncomms11362.

Hane JK et al. 2007. Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen Stagonospora nodorum. Plant Cell. 19:3347–3368. doi: tpc.107.052829 [pii] 10.1105/tpc.107.052829.

Harvey IC. 2002. Epidemiology and control of leaf and awn spot of barley caused by *Ramularia collo-cygni*. N. Z. Plant Prot. 331–335.

Havis ND et al. 2015. *Ramularia collo-cygni*--An Emerging Pathogen of Barley Crops. Phytopathology. 105:895–904. doi: 10.1094/phyto-11-14-0337-fi.

Kaczmarek M et al. 2017. Infection strategy of *Ramularia collo-cygni* and development of ramularia leaf spot on barley and alternative graminaceous hosts. Plant Pathol. 66:45–55. doi: 10.1111/ppa.12552.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580. doi: 10.1006/jmbi.2000.4315.

Kryazhimskiy S, Plotkin JB. 2008. The Population Genetics of dN/dS. PLoS Genet. 4:e1000304. doi: 10.1371/journal.pgen.1000304.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35:1547–1549. doi: 10.1093/molbev/msy096.

Manning VA et al. 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 Genes Genomes Genet. 3:41–63. doi: 10.1534/g3.112.004044.

Marthey S et al. 2008. FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics. 9:456.

Matusinsky P et al. 2011. Impact of the seed-borne stage of *Ramularia collo-cygni* in barley seed. J. Plant Pathol. 93:679–689. doi: 10.4454/jpp.v93i3.3650.

McGrann GR et al. 2016. The genome of the emerging barley pathogen *Ramularia collo-cygni*. BMC Genomics. 17:584.

Oxley SJP, Havis ND. 2004. Development of R*amularia collo-cygni* on spring barley and its impact on yield. In: Vol. 2004 pp. 147–152.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods. 8:785–786.

Piotrowska MJ, Fountaine JM, Ennos RA, Kaczmarek M, Burnett FJ. 2017. Characterisation of *Ramularia collo-cygni* laboratory mutants resistant to succinate dehydrogenase inhibitors. Pest Manag. Sci. 73:1187–1196. doi: 10.1002/ps.4442.

Poppe S, Dorsheimer L, Happel P, Stukenbrock EH. 2015. Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici* (Mycosphaerella graminicola). PLOS Pathog. 11:e1005055. doi: 10.1371/journal.ppat.1005055.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinforma. Oxf. Engl. 21 Suppl 1:i351-358. doi: 10.1093/bioinformatics/bti1018.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842. doi: 10.1093/bioinformatics/btq033.

Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat. Rev. Microbiol. 10:417–30. doi: 10.1038/nrmicro2790.

Richards JK, Wyatt NA, Liu Z, Faris JD, Friesen TL. 2018. Reference Quality Genome Assemblies of Three *Parastagonospora nodorum* Isolates Differing in Virulence on Wheat. G3 Genes Genomes Genet. 8:393–399. doi: 10.1534/g3.117.300462.

Salamati S, Reitan L. 2006. *Ramularia collo-cygni* on spring barley, an overview of its biology and epidemiology. In: pp. 19–35.

Salamov AA, Solovyev VV. 2000. Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10:516–522.

Smit, AFA. 2013. *RepeatMasker Open-4.0.* http://www.repeatmasker.org.

Sperschneider J et al. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761.

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

Stanke M et al. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612. doi: 10.1093/nar/gkl315.

Syme RA, Hane JK, Friesen TL, Oliver RP. 2013. Resequencing and comparative genomics of *Stagonospora nodorum*: sectional gene absence and effector discovery. G3 Genes Genomes Genet. 3:959–969.

Tamura K et al. 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. 109:19333–19338. doi: 10.1073/pnas.1213199109.

Tamura T, Akutsu T. 2007. Subcellular location prediction of proteins using support vector machines with alignment of block sequences utilizing amino acid composition. BMC Bioinformatics. 8:466.

Taylor JW, Berbee ML. 2006. Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia. 98:838–849. doi: 10.1080/15572536.2006.11832614.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18:1979–1990.

Walter MC et al. 2008. PEDANT covers all complete RefSeq genomes. Nucleic Acids Res. 37:D408–D411.

Wyatt NA, Richards JK, Brueggeman RS, Friesen TL. 2018. Reference Assembly and Annotation of the *Pyrenophora teres f. teres* Isolate 0-1. G3 Genes Genomes Genet. 8:1–8. doi: 10.1534/g3.117.300196.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 17:32–43.

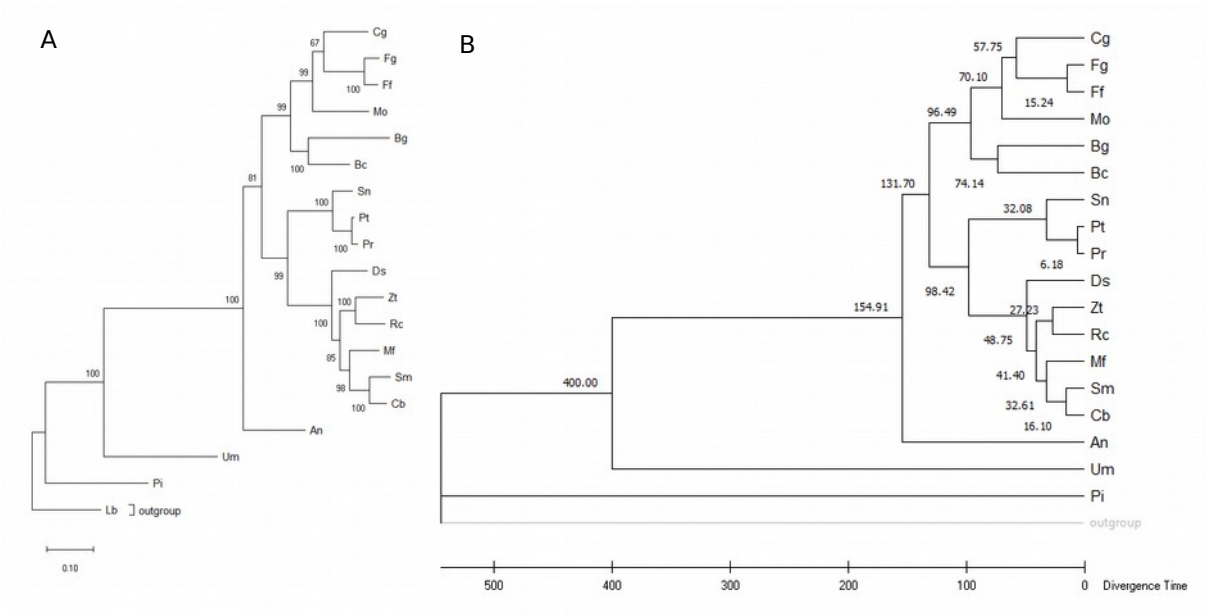

Figure 1: Phylogenetic tree showing the relationship between Ramularia collo-cygni and 16 other fungi. *A) Maximum likelihood phylogenetic tree based on the analysis of 10 housekeeping genes. Reported on the nodes are bootstrap values for 1000 bootstraps. B) Timetree results. Reported on the nodes are the divergence times in million years.*

Bc = Botrytis cinerea; Bg = Blumeria graminis; Cg = Colletotrichum graminicola; Cb = Cercospora beticola; Ds= Dothistroma septosporum; Fg= Fusarium graminearum; Ff= Fusarium fujikuroi; Mf= Pseudocercospora fijiensis; Mo=Magnaporthe grisea; Pr= Pyrenophora tritici-repentis; Pt= Pyrenophora teres f. teres; Rc= Ramularia collo-cygni; Sn= Parastagonospora nodorum; Sm= Sphaerulina musiva; Zt= Zymoseptoria tritici
An= Aspergillus nidulans; Um = Ustalago maydis;Pi= Piriformospora indica; Outgroup = Laccaria bicolor

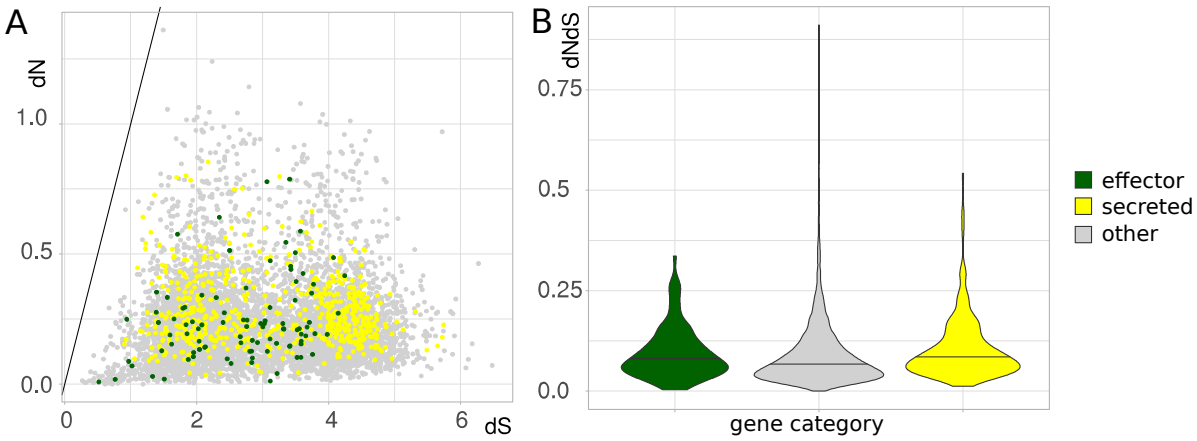

Figure 2: dN/dS between *R. collo-cygni* and *Z. tritici*

A) Scatter plot of the dS (x axis) against dN (y axis) for each predicted protein in a pairwise comparison between *R. collo-cygni* and *Z. tritici.* B) Violin diagrams of the dN/dS ratio for predicted proteins in a pairwise comparison between *R. collo-cygni* and *Z. tritici.* Data coloured based on whether the proteins are predicted to be putatively secreted proteins, putative effectors or other, non-secreted, proteins. Horizontal bar depicts the median.

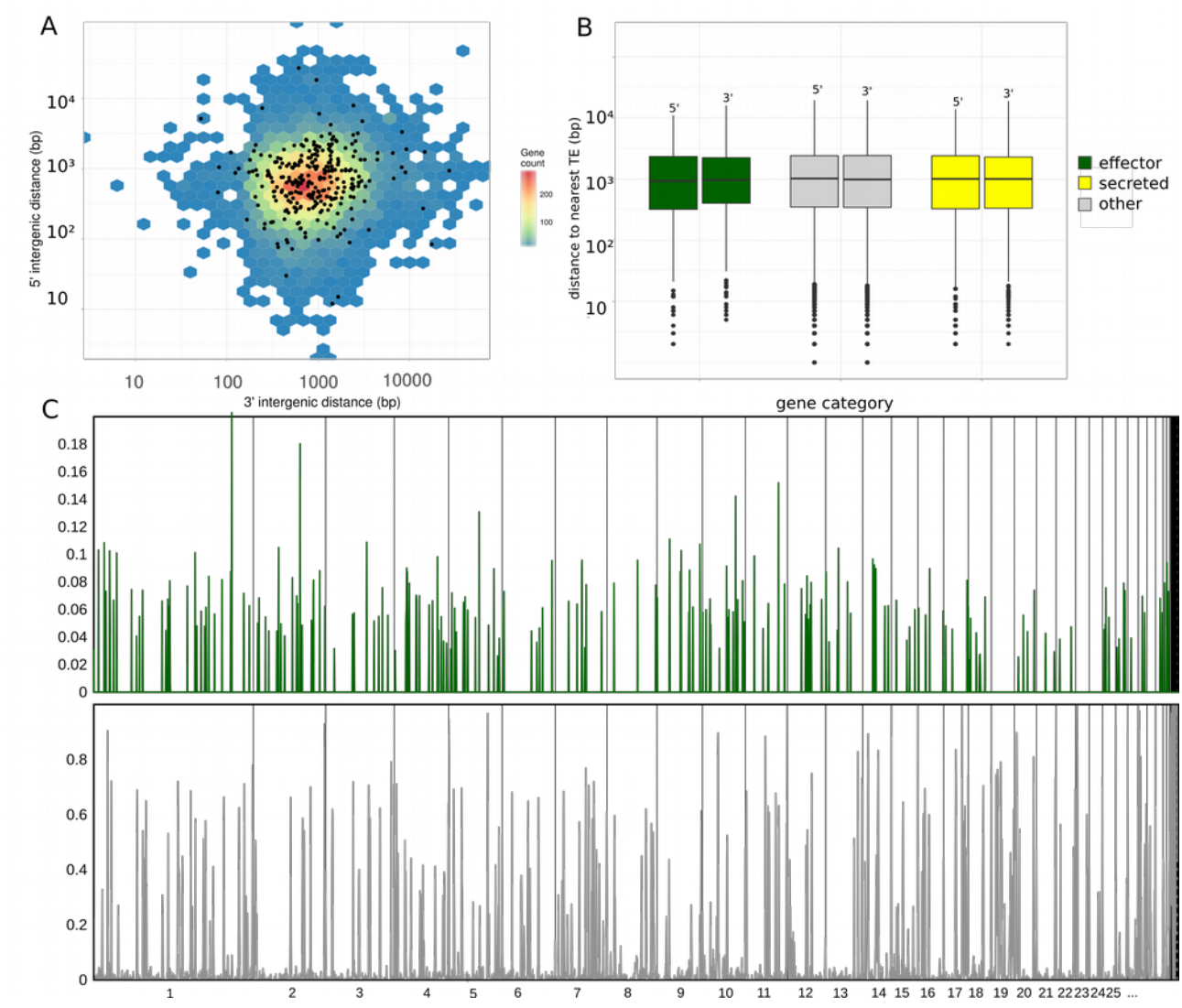

Figure 3: Intergenic distance and TEs in *R. collo-cygni*

A) Density plot of the intergenic distances on the 5' end (x axis) against the 3' end (y axis) for each predicted gene on the genome. Distances for all genes are binned in hexagons and coloured using a blue to red scale, red indicating the largest amount of genes per hexagon. Putative effectors are plotted as separate black dots. B) Violin diagrams of the distance to the nearest TE for each gene coding for predicted non-secreted (grey), putative secreted (yellow) or effector (green) proteins. C) Density of putative effector encoding genes (top, green) and transposable elements (bottom, grey) plotted in 10 kb non-overlapping sliding windows along the genome (x axis). Density is defined as number of basepairs that is part of a putative effector or TE in the window (y axis).

Sheet1

	Ramularia collo-cygni	***Ramularia collo-cygni***	***Phaeosphaeria nodorum***
Isolate	***Urug2***	*DK05 Rcc001*	*SN15*
Genome size (Mb)	***32.3***	*30.3*	*37.2*
Chromosomes	***n.d.***	*n.d.*	*n.d.*
Scaffolds	***78***	*576*	*107*
N50 scaffold (Mb)	***2.1***	*0.21*	*0.17*
GC-content (%)	***49.7***	*51.4*	*50.2*
Protein coding genes	***11622***	*11617*	*14885*
Gene density	***383***	*384*	*402*
Total secreted Protein	**1303**	1053*	2172
*Number of predicted effectors**	**282**	150*	625

	Pyrenophora tritici-repentis	***Pyrenophora teres-teres***	***Zymoseptoria tritici***
Isolate	*BRP-ToxAC*	*0-1*	*IPO323*
Genome size (Mb)	*37.8*	*46.5*	*39.7*
Chromosomes	*n.d.*	*n.d.*	*21*
Scaffolds	*47*	*86*	*21*
N50 scaffold (Mb)	*1.9*	*1.73*	*Finished*
GC-content (%)	*51*	*46*	*51.7*
Protein coding genes	*12141*	*11541*	*10933*
Gene density	*321*	*248*	*276*
Total secreted Protein	1433	1357	1252
*Number of predicted effectors**	382	337	380