Whole-genome sequencing of multiple isolates of *Puccinia triticina* reveals asexual lineages evolving by recurrent mutations

John P. Fellers ^{1,*}, Sharadha Sakthikumar,² Fei He ³, Katie McRell,³ Guus Bakkeren ⁴, Christina A. Cuomo ², and James A. Kolmer⁵

¹USDA-ARS, Hard Winter Wheat Genetics Research Unit, Manhattan, KS 66506, USA

²Broad Institute and MIT, Cambridge, MA 02142, USA

³Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA

⁴Agriculture and Agri Food Canada, Summerland, BC V0H1Z0, USA

⁵USDA-ARS, Cereal Disease Laboratory, St. Paul, MN 55108, USA

*Corresponding author: USDA-ARS, Department of Plant Pathology, 4008 Throckmorton Hall, 1712 Claflin Drive, Manhattan, KS 66506, USA. Email: john.fellers@usda.gov

Abstract

The wheat leaf rust fungus, *Puccinia triticina* Erikss., is a worldwide pathogen of tetraploid durum and hexaploid wheat. Many races of *P. triticina* differ for virulence to specific leaf rust resistance genes and are found in most wheat-growing regions of the world. Wheat cultivars with effective leaf rust resistance exert selection pressure on *P. triticina* populations for virulent race types. The objectives of this study were to examine whole-genome sequence data of 121 *P. triticina* isolates and to gain insight into race evolution. The collection included isolates comprising of many different race phenotypes collected worldwide from common and durum wheat. One isolate from wild wheat relative *Aegilops speltoides* and two from *Ae. cylindrica* were also included for comparison. Based on 121,907 informative variants identified relative to the reference Race 1-1 genome, isolates were clustered into 11 major lineages with 100% bootstrap support. The isolates were also grouped based on variation in 1311 predicted secreted protein genes. In gene-coding regions, all groups had high ratios of non-synonymous to synonymous mutations and nonsense to readthrough mutations. Grouping of isolates based on two main variation principle components for either genome-wide variation or variation just within the secreted protein genes, indicated similar groupings. Variants were distributed across the entire genome, not just within the secreted protein genes. Our results suggest that recurrent mutation and selection play a major role in differentiation within the clonal lineages.

Keywords: wheat leaf rust; *Puccinia triticina*; genomic relationship; pathogen adaptation

Introduction

Rust diseases of cereal crops are widespread throughout grain-growing regions of the world and are capable of causing severe yield loss (Saari and Prescott 1985). Of the three rust diseases on wheat (Kolmer et al. 2019), leaf rust, caused by *Puccinia triticina* Erikss., is the most widespread and the leading cause of yield loss on a worldwide basis (Savary et al. 2019). In the U.S., leaf rust is found in the soft red winter wheat region of eastern and southern states, the hard red winter wheat region of the southern Great Plains and in the hard red spring wheat region of the northern Great Plains (Kolmer 2019). In 2007, a severe leaf rust epidemic occurred in the Great Plains region and Kansas and Minnesota losses were estimated at 14 and 7%, respectively (Kolmer et al. 2009).

Like the other cereal rusts, *P. triticina* is a heteroecious and macrocyclic rust requiring two taxonomically unrelated hosts to complete the sexual cycle with five distinct spore stages (Kolmer et al. 2019). The principle hosts of *P. triticina* are common hexaploid wheat (*Triticum aestivum* L.); tetraploid durum wheat

(*T. turgidum* L. ssp. *durum*); and wild emmer wheat (*T. turgidum* ssp. *dicoccoides*; Bolton et al. 2008). Infections of *P. triticina* are also found on *Aegilops cylindrica* L. (goatgrass) in the southern Great Plains. Dikaryotic urediniospores are asexually produced on susceptible host plants and can cycle indefinitely. Urediniospores are wind-blown and deposited in wheat fields by precipitation events then germinate and penetrate the host plants through the stomata. Haustoria form in the mesophyll cells below the epidermal and palisade layers of cells. New uredinia erupt on the upper leaf surface within 7–10 days after the initial infection. The sexual cycle proceeds with karyogamy, fusion of two haploid nuclei, and the formation of two-celled, dikaryotic teliospores in the old overwintering stage uredinia (subsequently called telia). In the spring, reduction division meiosis derived teliospores undergo a brief diploid phase, then each cell germinates to produce four haploid basidiospores that infect alternate hosts. The sexual cycle is completed by the production of haploid spermatogonia and spermatia on the upper leaf surface. Fertilization results in

dikaryotic aeciospores in aecial cups on the lower leaf surface which are then wind-dispersed and infect the uredinial-telial hosts. The most compatible alternate host of *P. tritricina* is common meadow rue (*Thalictrum speciosissimum* L.) found in southern Europe and West Asia (Jackson and Mains 1921; Saari et al. 1968). The native North American *Thalictrum* species have limited susceptibility to basidiospore infection. Populations of *P. tritricina* in North America (Ordóñez and Kolmer 2009) and worldwide (Kolmer et al. 2019) have high levels of linkage disequilibrium and higher than expected heterozygosity for microsatellite or simple sequence repeat (SSR) loci, which strongly indicates clonal reproduction.

P. tritricina interacts with wheat in a gene-for-gene manner (Samborski and Dyck 1968, 1976) as described by Flor (1942, 1971). Specific wheat leaf rust resistance genes interact with specific effectors in avirulent isolates to produce incompatible infection types, which range from minute hypersensitive necrotic flecks to small uredinia that are surrounded by necrosis or chlorosis (Bolton et al. 2008). Isolates that lack corresponding effectors and are virulent to the specific leaf rust resistance gene, produce large uredinia without chlorosis or necrosis and can produce hundreds of urediniospores daily. Since the mid-1940s (Kolmer 2019) wheat breeding programs in the Great Plains region have released wheat cultivars with leaf rust resistance genes that provide resistance to most of the *P. tritricina* population. However, within a few years after release, virulent genotypes of *P. tritricina* arise and increase, rendering the cultivars susceptible. As a result of the constant selection for virulent genotypes, many different races of *P. tritricina* are found in the U.S. (Kolmer 2019) and worldwide (Kolmer et al. 2019). *P. tritricina* races are determined using differential isogenic wheat genotypes differing by single leaf rust resistance genes and characterizing avirulent/virulent infection types produced by single-uredinial isolates (Bolton et al. 2008).

In the absence of the sexual cycle, the source of new races in *P. tritricina* in North America has been assumed to be somatic mutations within urediniospores. Some isolates of *P. tritricina* have been shown to be heterozygous for avirulence/virulence alleles to specific leaf rust resistance genes (Samborski and Dyck 1968, 1976; Statler 1977, 1979, 1982, 2000). Mutations at heterozygous loci can lead to loss of recognition by the host resistance protein and result in a compatible infection type (Kolmer and Dyck 1994). *P. tritricina* can survive the winter in a large area of the southern U.S. winter wheat region, resulting in a large population in which mutation and selection of virulent genotypes can occur.

A reference genome sequence of Race 1-1, designated as BBBDB in the *P. tritricina* nomenclature system (Long and Kolmer 1989) recently produced (Cuomo et al. 2017). Race 1-1 is avirulent to most leaf rust resistance genes and is heterozygous for avirulence/virulence at a number of effector loci (Samborski and Dyck 1968). The genome contains 135 Mb of DNA, 51% mobile and repetitive elements, with an estimated 14,800 genes. The reference genome is very useful for alignment of sequence reads for identifying and assessing genetic polymorphisms in populations of *P. tritricina* using whole-genome sequencing of sample isolates. In this study, 121 *P. tritricina* isolates, representing different race phenotypes collected from common wheat in North America and Europe, worldwide durum wheat, and wheat relatives *Ae. cylindrica* and *Ae. speltoides*, were sequenced and compared with the BBBDB_race 1-1 reference genome. The objectives of the research were to determine the phylogenetic lineage relationships in a diverse collection of *P. tritricina*, characterize the genetic variation within lineages, and gain insight into the genetic basis of variation leading to the development of new virulent isolates.

Methods

Isolate phenotyping

Single uredinial isolates of *P. tritricina* were increased on susceptible seedlings of *T. aestivum* variety “Little Club” (CI 4066; Clark et al. 1926), and then inoculated onto 7-day old seedlings of Thatcher near-isogenic differential lines that differ for single leaf rust resistance genes. Thatcher lines with genes *Lr1*, *Lr2a*, *Lr2c*, *Lr3*, *Lr9*, *Lr16*, *Lr24*, *Lr26*, *Lr3ka*, *Lr11*, *Lr17*, *Lr30*, *LrB*, *Lr10*, *Lr14a*, *Lr18*, *Lr3bg*, *Lr14b*, *Lr20*, *Lr21*, *Lr28*, and *Lr39* were used to characterize the isolates for virulence. Protocols for rust inoculation, incubation, and collection of urediniospores, and greenhouse conditions were as previously described (Kolmer and Hughes 2018). The isolates were rated for virulence to the individual Thatcher differential lines 10–12 days post-inoculation. Isolates with infection types 3–4 (moderate to large uredinia lacking necrosis or chlorosis) were considered virulent, and infection types 0–2+ (immune response with no sign of infection to moderate-size uredinia with prominent chlorosis) were considered avirulent. Each isolate was given a 20-digit binary number based on the avirulent/virulent response to the 20 Thatcher differential lines.

DNA extraction

Seed of *T. aestivum* L. variety “Little Club” was planted in six 10" × 10" × 2" aluminum cake pans containing Metro Mix 360 soil media (SunGro). At the two-leaf stage, seedlings were inoculated with 25 mg of urediniospores from each isolate by suspending spores in 3 ml of Soltrol 170 paraffin oil (Phillips Petroleum, Bartlesville, OK, USA). Seedlings were fogged with an atomizer using 40 psi of compressed air and placed overnight in a dew chamber (I-36DL, Percival Scientific, Perry, IA, USA) at 100% humidity and 20–22°C with lights off. Plants were placed back into a controlled environment chamber for 16 hours/8 hours day/night light cycle and 20°C. After 11–14 days, spores were collected using a cyclonic Kramer-Collins spore collector (Tallgrass Solutions, Inc. Manhattan, KS, USA). Urediniospores were germinated overnight by evenly scattering 0.5 g of spores over 300 ml of 1X germination solution contained in 20 × 30 cm Pyrex® baking pan (Webb et al. 2006). DNA was isolated as described in Cuomo et al. (2017). Spore mats were skimmed off the surface of the germination solution, washed 3X with ddH₂O, dried in a Buchner funnel, and stored at –80°C. From each mat, 350 mg of tissue was ground using liquid N₂. DNA was isolated using the OmniPrep™ DNA isolation Kit according to the recommended protocol for large tissue samples (Q-Biosciences, St. Louis MO, USA, Cuomo et al. 2017).

Isolate sequencing

DNA samples were sent to the Broad Institute (Cambridge, MA, USA) for sequencing and analysis. From each sample, 200 bp insert Illumina (Illumina, San Diego, CA, USA) libraries were constructed and ligated with a unique index barcode. All libraries were pooled and sequenced with eight lanes of HiSeq 2000, 101 base-paired end reads.

Variance analysis

Illumina reads were aligned against *P. tritricina* Race 1-1 assembly version 2 (BBBDB; NCBI accession PRJNA36323) using BWA (Version: 0.5.9). Alignments were filtered requiring a mapping quality of 30 using SAMtools view (Li et al. 2009; Li and Durbin 2010). Regions with high insertion/deletion counts, which could be misaligned, were identified using GATK (version 2.1.9; McKenna et al. 2010) RealignerTargetCreator and realigned using

GATK IndelRealigner. Single-nucleotide polymorphisms (SNPs) were identified using GATK UnifiedGenotyper and filtered using GATK VariantFiltration (version 3) best practices for hard filters (QD < 2.0, MQ < 30.0, FS > 60.0, HS > [$\mu + 2\sigma$] MQRankSum < -12.5, ReadPosRankSum < -8.0), except that Haplotype Score (HS) was filtered above the mean HS plus two standard deviations. A single table of SNP calls in all strains was constructed, and at positions where no call was present in a strain the sequence data was re-evaluated to predict homozygous reference genotype calls. Using the pileup output from SAMtools, homozygous reference positions were called requiring an 8X depth of coverage and reference allele frequency >80%. Proteins with secretions signals were identified using SignalP V4.0 from the version 2 genome gene predictions (Petersen et al. 2011; Cuomo et al. 2017).

Phylogenetic analysis

Phylogenetic relationship based on heterozygous and homozygous SNP differences were inferred using PAUP Maximum Parsimony and “hetequal” substitution model (Swafford 1998, 2002) with 1000 bootstrap replications. PLINK1.9 was used for the calculation of principal components, and identity-by-descent and identity-by-state (IBD/IBS). VCFtools (<https://vcftools.github.io/>) was used to calculate F_{ST} and diversity nucleotide π . The Wilcoxon rank-sum test was used to determine significance.

Results

The genomes of 121 isolates of *P. tritici* were included in this study (Supplementary Table S1). Eighty-five isolates were from North America, with 65 different virulence phenotypes; 21 durum wheat isolates from world-wide collections with three different virulence phenotypes; and 12 isolates from European common wheat with 11 different virulence phenotypes. A single *Ae. speltoides* isolate collected in Israel, and two *Ae. cylindrica* (common goatgrass) isolates from North America were also included. For each isolate the virulence phenotype, country of origin, sequence statistics, avirulence/virulence formula, and references to the previous characterization by earlier marker systems is listed in Supplementary Table S1. The *Ae. speltoides* isolate (Pt_ISR_850) had the lowest alignment to the reference genome at 78%, all other isolates had >90% alignment. Average depth of coverage across the reference genome ranged from 9X (isolate 11US045_2 TNBG) to 95X (isolate 11US220_3 MFNSB).

In the absence of a phased genome, a maximum parsimony approach was used to model transitions between homozygous and heterozygous SNPs. Inspection of the phylogenetic tree confirmed that clades corresponded to groups of isolates with very similar race designations, validating this approach. Using 121,907 informative SNPs from sites with $\leq 5\%$ missing calls, the isolates could be clustered into eleven major groups with 100% support. The isolates from North America (NA) clustered into six discrete groups (Figure 1). The genome reference BBBD_B_Race 1-1, collected in 1954 (Ordóñez and Kolmer 2009), was in the NA1 group with three other isolates. Isolate 04NE356, with the similar virulence phenotype BBBDJ and believed to be the most current isolate most similar to Race 1-1, was collected 50 years later and differed by only 4447 SNPs when compared to the reference genome. The NA2 group consisted of eleven isolates with ten different virulence phenotypes, all avirulent to the isogenic *Lr2a* line, and virulent to lines with *Lr2c* and *Lr28* (Supplementary Table S2). The 35 NA3 group isolates consisted of 25 phenotypes, all virulent to genes *LrB*, *Lr3bg*, *Lr17*, and avirulent to *Lr28*. These

virulence phenotypes were first found in large numbers in the mid 1990's after the release of *Lr37* in hard red winter cultivars in the southern Great Plains (Ordóñez and Kolmer 2009; Kolmer 2019). The two NA4 group isolates were originally collected from *Ae. cylindrica* in the southern Great Plains and were both virulent to *Lr1*, *Lr2a*, *Lr2c*, *Lr17*, *Lr28*, and avirulent to *Lr3*. The NA5 group consisted of 32 isolates with 23 different virulence phenotypes and were avirulent to lines with genes *LrB*, *Lr17*, and virulent *Lr28*. Two isolates PBJL_PRTUS6 and PBJSF_84VAPA, both avirulent to *Lr2a*, and virulent to *Lr2c*, *Lr11*, and *Lr17*, were highly related and placed in the NA6 group. A single isolate, TCTDL_03VA190, did not cluster in any other group.

Thirteen isolates from Europe (EU) with 12 different virulence phenotypes were all virulent to *LrB*, avirulent to *Lr28*, and varied for virulence to the other *Lr* genes (Supplementary Table S2, Figure 1). The European isolates were also more diverse for SNP genotypes, but were grouped discretely from all NA group isolates. DBBQJ_09AZ was collected from durum wheat in Arizona, but grouped with the European isolates since it is highly related to those isolates for virulence and SNP genotype. All other durum wheat isolates have distinct SNP genotypes compared to the isolates from hexaploid wheat. The seven Ethiopian durum (ETH_DURUM) isolates are distinct for SNP genotype compared to the other durum isolates. These isolates are all avirulent to the differential line recurrent parent cultivar, Thatcher and thus were not phenotyped. Durum wheat isolates collected in California, Mexico, Argentina, Spain, Ethiopia, France, and Israel (DURUM) are closely related for SNP genotype, distinct from all other isolates, are all virulent to *Lr14b* and *Lr20* (Supplementary Table S2), and avirulent to most of the other differential set *Lr* genes. A single isolate, Pt_ISR_850, collected in Israel from diploid wheat progenitor *Ae. speltoides*, had a distinct SNP genotype and did not cluster with isolates from any other group. Isolate SCPBJ_ISR173, also collected in Israel but from common wheat, also did not group with any other isolates in any of the genotype groups but was closest for genotype to three European isolates.

Excluding Pt_ISR_850 from *Ae. speltoides*, isolates in NA4 had the lowest average number of SNPs (310,033) relative to Race 1-1 and ETH-Durum group isolates had the highest average number of SNPs (450,544; Table 1). NA4 Isolates also had the lowest π value, 2.29×10^{-3} , and isolates in the ETH-durum group had the highest π value, 3.33×10^{-3} . Common wheat isolate groups had higher levels of SNP heterozygosity ranging from 76% for NA3 to 97% for NA1. The international and Ethiopian durum groups had lower levels of SNP heterozygosity at 56% (Table 1). Even though the isolate from *Ae. speltoides* had the highest number of SNPs (996,187), only 35% were heterozygous. The large majority of isolate SNPs occurred in the nonexpressing intergenic regions ranging from 78.5% for the *Ae. speltoides* isolate to 84.6% for NA4 isolates (Table 1). The ratios for all groups of nonsynonymous to synonymous SNPs within the expressed regions were >1.0, and >4.0 for nonsense to read through SNPs (Table 1).

Principle component analysis (PCA) and groupings were performed on the isolates. The initial analysis grouped the isolates in a two-dimensional PCA plot simply based on virulence to the 22 Thatcher *Lr* isolines (Figure 2, Supplementary Table S1). The isolates did not fall into discrete nonoverlapping clusters. However, there were regions of the PCA plot where most of the isolates were from one group. Of the NA groups, isolates in NA1, NA2, NA4, and NA5 were grouped more closely in the right side of the plot with positive first axis values, while isolates in NA3 and NA6 all had negative first axis values. Isolates in the EU group were diverse for virulence and were found in both sides of the

Figure 1 PAUP maximum parsimony analysis of SNPs across the genome of *P. triticina* isolates with 1000 bootstrap replicates. Total SNPs included were 134,940, of which 121,907 were parsimony-informative. Markers used were 95% informative at each site.

Table 1 Genome-wide SNP diversity of *P. triticina* isolates in groups from North America (NA) and Europe (EU), two groups virulent to durum wheat, and the wheat progenitor *Aegilops speltoides*

Population	Average total SNPs	Average diversity π (nucleotide)	Average intergenic SNPs (%)	Average SNP heterozygosity (%)	Average nonsynonymous/synonymous SNPs	Average nonsense/read-through SNPs
NA1	321,326	2.37×10^{-3}	84.0	97	1.659 (21,404/12,899)	4.237 (603/142)
NA2	422,333	3.12×10^{-3}	83.7	87	1.632 (28,383/17,389)	4.586 (796/173)
NA3	342,861	2.53×10^{-3}	84.2	76	1.657 (22,594/13,630)	4.311 (657/152)
NA4	310,033	2.29×10^{-3}	84.6	66	1.702 (20,426/12,000)	4.992 (646/129)
NA5	322,011	2.38×10^{-3}	84.1	90	1.654 (21,478/12,980)	4.345 (628/145)
NA6	346,721	2.56×10^{-3}	84.2	81	1.669 (22,974/13,765)	4.369 (680/155)
Intl-Durum	416,611	3.08×10^{-3}	84.0	56	1.645 (27,469/16,696)	4.258 (738/174)
ETH-Durum	450,544	3.33×10^{-3}	84.1	56	1.657 (29,843/18,010)	4.254 (791/186)
Europe (EU)	413,951	3.06×10^{-3}	83.8	81	1.635 (27,723/16,949)	4.355 (794/182)
<i>Ae. speltoides</i>	996,187	7.36×10^{-3}	78.4	35	1.155 (69,440/60,101)	4.045(1,505/372)

plot. Of the NA groups with more than ten isolates, NA3 had the smallest average virulence difference (number of different virulence changes) at 3.50, and isolates in NA5 had the largest at 4.85 (Table 2). Isolates in the International Durum group were very closely related to virulence, differing on average by only 1.03. The average virulence difference overall in the 121 isolates was 7.90.

The total genomic SNP variation (intronic and intergenic) was then compared to SNPs occurring only in predicted secreted protein genes, this to examine if the genetic relationship between isolates were due to differences based on total genomic variation and/or if relationships were based on variation that occurred in genes involved with pathogenicity and virulence specificity. Isolates in ETH-Durum and the *Ae. speltoides* isolate were not included due to their high divergence from the other isolate groups. Principle component variation based on the total 343,000 SNPs across the genome were PC1=35.72%, PC2=11.54%, PC3=9.12%, PC4=5.54%, and PC5=2.11%. The two-dimensional principal coordinate plot for PC1 and PC2 of the total SNPs across the genome (Figure 3A) grouped the isolates in the same manner as in the maximum parsimony analysis. Analysis of variation in 1311 predicted secreted

protein genes was based on 16,000 SNPs within 2,000bp 5' of the start codon, open reading frame, and 2,000bp 3' of the stop codon. Variation in principle components were PC1=38.73%, PC2=13.00%, PC3=8.74%, PC4=6.02%, and PC5=2.18%. Isolates were clustered in a similar manner as the total genomic data (Figure 3B) but values were reversed on the Y-axis.

Median SNP diversity (π) was calculated for all genes and compared to predicted secreted protein genes. Secreted protein genes were less diverse than all other genes (Wilcoxon rank-sum test P -value = 1.19×10^{-11} ; Figure 4A). Individual genes were evaluated for diversity (π , window size of 1 Kb) to determine if certain predicted secreted protein genes were under a higher level of selection or variation. Secreted protein gene diversity was scattered evenly amongst the ranked list. Using the top 10% of total genes (1477), 65 secreted peptide genes were found to be the most variable (Supplementary Table S3). Next, the median F_{st} , measuring population differentiation and divergence, was determined for the whole genome and secreted proteins. In this analysis, F_{st} was significantly higher (Wilcoxon rank-sum test P -value = 2.80×10^{-12}) in secreted proteins than overall genome diversity (Figure 4B).

Figure 2 Principal coordinate plot based on *P. triticina* isolate virulence to 20 leaf rust resistance gene differential Thatcher near isogenic lines. Isolates are from North American (NA) and European (EU) groups and two groups virulent to durum wheat.

Table 2 Virulence diversity of *P. triticina* isolates in SNP genotype groups from North America (NA), Europe (EU), and two groups virulent to durum wheat

Group	No. of isolates	No. of virulence phenotypes	Average virulence difference
NA1	4	3	3.30
NA2	11	10	3.84
NA3	35	25	3.50
NA4	2	2	1.00
NA5	32	23	4.85
NA6	2	2.0	—
EU	13	12	4.63
Intl Durum	13	3	1.03
ETH Durum	8	—	—

Discussion

The lineage groups in North America are highly associated with avirulence/virulence to specific leaf rust resistance genes. In the absence of sexual recombination, clonal reproduction of

urediniospores has maintained the association between molecular genotype and virulence. In a previous study based on allelic variation at 23 SSR loci (Ordóñez and Kolmer 2009), isolates were placed into the NA1, NA2, NA3, NA4, and NA5 groups, and a separate group of durum wheat isolates. The NA1, NA2, NA4, and NA5 groups are long-standing in North America, as isolates with the characteristic virulence can be found dating to the 1950s (Johnston et al. 1968). NA3 isolates first become common in the mid-1990s after hard red winter wheat cultivars with *Lr37* were widely grown in the southern Great Plains (Ordóñez and Kolmer 2009). In the current U.S. *P. triticina* population, NA3 isolates account for the large majority of all isolates (Kolmer 2019). Two isolates were assigned to different NA groups based on SNP genotype compared to SSR genotypes (Ordóñez and Kolmer 2009). PBJSF_84VA was previously placed in the NA2 group based on SSR genotype, however, the complete genome sequence data placed this isolate with isolate PBJL_PRTUS6 in the NA6 group. Similarly, isolate TCTDL_03VA was placed in NA1 based on SSR genotype, however, based on complete genome sequence this isolate was not closely related to any other isolates in this study. Isolate Durum_09AZ103A was collected from durum wheat in Arizona in 2009. Surprisingly this isolate had an SNP genotype that placed it with isolates from Europe, indicating the high probability of this isolate being a recent introduction. The isolates from Europe in this study were distinct for SNP genotype from all of the other North American isolates. In a previous study, these same European isolates were also very distinct for SSR genotype compared to the North American isolates (Kolmer et al. 2019).

In this study, all of the isolates were sequenced using short-read technology. When BBBD_B_Race 1-1 was re-sequenced and aligned back to the genome reference sequence, 273,385 SNPs were identified. The reference genome (Cuomo et al. 2017) was assembled using the Arachne (HybridAssemble; Jaffe et al. 2003) program, and it is likely that since the genome is dikaryotic, significant regions from each nucleus were collapsed into one contig

Figure 3 Principal component (PC) analysis of 121 *P. triticina* isolates representing populations from North America, Europe, South America and from different wheat classes and species. Data is based on (A) ~343,000 whole-genome SNPs (PC Variation: PC1 = 35.72%, PC2 = 11.54%, PC3 = 9.12%, PC4 = 5.54%, and PC5 = 2.11%); or (B) ~16,000 SNPs found in 1311 predicted secreted protein genes (PC Variation: PC1 = 38.73%, PC2 = 13.00%, PC3 = 8.74%, PC4 = 6.02%, and PC5 = 2.18%).

Figure 4 (A) Median diversity (π) for SNPs within all of the genes, then SNPs within predicted secreted protein genes. (B) Median Fst measuring population substructure and divergence for genome-wide SNPs and within predicted secreted protein genes.

and the genome represents more of a haploid genome. The high number of BBDB_Race 1-1 SNPs from the re-alignment to the reference genome is likely due to sequence from each genome being aligned to a contig region and the GATK program identifying differences caused by the two genomes. The high number of SNPs of the re-sequenced BBDB_Race 1-1 and the other isolates in the NA1 group relative to the reference genome is most probably caused by the assembly of the two genomes into a single genome. The high SNP heterozygosity of the groups from common wheat may also be a result of the genomes being collapsed into a single genome. Wu et al. (2017) re-sequenced 20 isolates of *P. tritica* from Australia that were aligned with the reference genome. The 20 isolates averaged 404,690 SNPs with 87% heterozygosity relative to the reference genome. Recently, the genome assembly of *P. graminis* f.sp. *tritici* was sorted into two groups representing the two nuclei. Using long reads, chromosome association, and specific algorithms, Feng et al. (2019) were able to sort contigs into their respective genomes. Kolmer et al. (2020) noted much lower heterozygosity on the order to 8–12% of SNPs in a large collection of *P. tritica* isolates that were characterized with the genotype-by-sequence (GBS) approach. The isolate reads (100bp, single end) in the GBS study were also aligned with the BBDB_Race 1-1 reference genome, however, the alignment, genotype data, VCF file, and haplotypes were extracted using different software programs. Differences in the mapping, SNP calling, threshold levels for heterozygotes, genome coverage, regions of the genome sequenced, depth of coverage, and assembly programs may also

contribute to the discrepancies in results between studies. Interestingly, both groups of isolates from durum wheat and the single isolate from *A. speltoide*s had much lower levels of SNP heterozygosity compared to the isolates from common wheat.

Results of the whole genome sequence confirmed the genetic relationships between the diploid wheat progenitor *A. speltoide*s isolate and tetraploid durum and hexaploid common wheat *P. tritica* forms. The maximum parsimony analysis indicated that the *A. speltoide*s isolate was most closely related to the ETH-Durum isolates, followed by the worldwide durum wheat isolates, and then common wheat isolates. In a coalescence analysis based on sequence at 15 loci, Liu et al. (2014) determined that the *A. speltoide*s form was ancestral, followed by ETH-Durum isolates, and then by common wheat and durum wheat isolates. Kolmer et al. (2020) using a GBS molecular clock approach also found the same phylogenetic relationships between the different forms of *P. tritica*. The ETH-Durum isolates are highly virulent to durum wheat, but are avirulent to almost all common wheat cultivars and have been found only in Ethiopia (Kolmer and Acevedo 2016). Ethiopia is a secondary center of origin of tetraploid wheat, where landrace cultivars of emmer and durum wheat are still grown (Eticha et al. 2006). The ETH-Durum isolates may be a remnant of the *P. tritica* population that existed before the widespread cultivation of hexaploid common wheat cultivars. The unique host environment of the landrace tetraploid wheats may allow the ETH-Durum isolates to survive in Ethiopia. The two NA4 isolates have virulence phenotypes that are found on the diploid *A. cylindrica* in the southern Great Plains of the U.S. The *P. tritica* population in Europe were previously determined to be highly diverse for SSR genotype groups (Kolmer et al. 2012a). Somatic recombination has been hypothesized as a source of variation in cereal rust fungi that reproduce clonally (Park et al. 1999; Park and Wellings 2012) and Feng et al. (2019) have recently shown the origin of Ug99 in *P. graminis* *tritici* was due to the exchange of nuclei.

Host populations are important drivers of the differentiation of plant pathogens at a species level and also for forms and genotypes within a species (Gladiux et al. 2017). The evolution of wheat from diploid progenitors to tetraploid and hexaploid cultivars was tracked by *P. tritica* (Liu et al. 2014). *P. tritica* isolates virulent to hexaploid common wheat have many different virulence phenotypes selected by leaf rust resistance genes (Kolmer 2019). Sequence variation in genes associated with pathogenicity and gene-for-gene interactions could possibly be used to differentiate pathogen genotypes for virulence to host resistance genes. However, even though isolates in the different NA groups differ for characteristic virulence to *Lr2a*, *Lr3*, *Lr3bg*, *Lr17*, *Lr28*, and *LrB*, there was no difference in the overall genetic relationship between isolates with groupings based on total genomic SNPs compared to groupings based on SNPs located in predicted secreted proteins associated with effector avirulence genes. Likewise, there was little difference in the nucleotide diversity and genetic similarity in relation to phenotypic similarity of total genomic SNPs and SNPs in secreted peptide genes. Because the current *P. tritica* population reproduces clonally, genomic SNPs associated with an NA group may not necessarily be linked with an effector gene, because mutations that occur throughout the genome are essentially linked. Similarly, the SNPs in secreted protein genes may be widely distributed in the genome and not specific to effectors that interact with isolate specific *Lr* genes.

P. tritica has a well-defined population structure with isolates clustering into distinct clades influenced by selection on resistance genes, host species, and geographic location (Ordóñez and Kolmer 2007, 2009; Szabo and Kolmer 2007; Kolmer et al. 2012a and 2012b; Kolmer and Acevedo 2016). In most of the isolate collection regions, the alternate host is not present, however, in

Europe and the Middle East it is present lending to the possibility of sexual recombination. Using molecular markers, previous work showed that *P. triticina* populations were derived from asexual, clonal propagation with high levels of linkage disequilibrium (Kolmer 1992; Liu and Kolmer 1998). Testing 15 genomic regions, including house-keeping genes, flanking regions around SSRs, and hypervariable regions, neutrality tests showed that evolution was neutral (Liu et al. 2014). However, it was concluded that gene flow was reciprocal between common wheat and cultivated durum *P. triticina* forms, but unidirectional between common wheat and Ethiopian durum forms. In the absence of sexual recombination, mutations in noncoding and coding regions in *P. triticina* have accumulated, resulting in high levels of SNP and SSR heterozygosity (Kolmer et al. 2019), genetically differentiated groups, and a consistent association between virulence/avirulence and molecular genotypes. This accumulation of genetic mutations, that would otherwise be removed with recombination, is similar to what has been seen in *Magnaporthe* (Gladiux et al. 2017). Despite the lack of sexual recombination with the resultant accumulation of mutations in coding regions, different virulent phenotypes of *P. triticina* have continued to adapt to host resistance genes in both durum and common wheat cultivars and spread throughout the world. The forces of recurrent mutation, genetic drift, and selection by host resistance genes have together differentiated and diversified populations of *P. triticina* worldwide.

Authors' Contributions

J.P.F., K.M., F.H., and J.A.K. planned and developed experiments, isolated DNA, analyzed results, and co-wrote the manuscript. S.S. and C.A.C. planned and developed experiments, sequenced the samples, analyzed results, and co-wrote the manuscript. G.B. provided data and co-wrote the manuscript.

Acknowledgments

The authors thank the Broad Institute Genomics Platform for generating all sequence data for this study. The authors would like to thank Robert Bowden and Myron Bruce for their assistance in preparation of the manuscript and technical guidance.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Funding

This work was funded through USDA CSREE awards 2008-35600-04693 and 2009-65109-05916 and USDA-CRIS project 3020-21000-10-00D.

Data availability

DNA sequence data were deposited at NCBI as BioSample PRJNA77809, and individual race SRA run numbers are listed in Supplementary Table S1.

Supplementary material is available at G3 online.

Conflicts of interest

None declared.

Literature cited

- Bolton MD, Kolmer JA, Garvin DF. 2008. Wheat leaf rust caused by *Puccinia triticina*. *Mol Plant Pathol*. 9:563–575.
- Clark JA, Love HH, Gaines EF. 1926. Registration of standard wheat varieties. *Agronj*. 18:920–922.
- Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, et al. 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3 (Bethesda)*. 7:361–376.
- Eticha F, Belay G, Bekele E. 2006. Species diversity in wheat landrace populations from two regions of Ethiopia. *Genet Resour Crop Evol*. 53:387–393.
- Feng L, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, et al. 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nat Commun* 10:5068.
- Flor HH. 1942. Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology*. 32:653–669.
- Flor HH. 1971. Current status of gene-for-gene concept. *Ann Rev Phytopathol*. 9:27.
- Gladiux P, Ravel S, Rieux A, Cros-Arteil S, Adreit H, et al. 2017. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. *mBio*. 9:e01806-17.
- Jackson HS, Mains EB. 1921. Aecial stage of the orange leaf rust of wheat, *Puccinia triticina* Erikss. *J Agric Res*. 22:151–172.
- Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, et al. 2003. Whole-genome sequence assembly for mammalian genomes: *Arachne 2*. *Genome Res*. 13:91–96.
- Johnston CO, Caldwell RM, Compton LE, Browder LE. 1968. Physiologic races of *Puccinia recondita* f. sp. *tritici* in the United States from 1926 through 1960. *USDA Tech Bull*. 1393:1–18.
- Kolmer JA. 1992. Virulence heterozygosity and gametic phase disequilibrium in two populations of *Puccinia recondita* (wheat leaf rust fungus). *Heredity*. 68:505–513.
- Kolmer JA. 2019. Virulence of *Puccinia triticina*, the wheat leaf rust fungus, in the United States in 2017. *Plant Dis*. 103:2113–2120.
- Kolmer JA, Acevedo MA. 2016. Genetically divergent types of the wheat leaf rust fungus *Puccinia triticina* in Ethiopia, a center of tetraploid wheat diversity. *Phytopathology*. 106:380–385.
- Kolmer JA, Dyck PL. 1994. Gene expression in the *Triticum aestivum*-*Puccinia recondita* f.sp. *tritici* gene-for-gene model. *Phytopathology*. 84:437–440.
- Kolmer JA, Hanzalova A, Goyeau H, Bayles R, Morgounov A. 2012a. Genetic differentiation of the wheat leaf rust fungus *Puccinia triticina* in Europe. *Plant Path*. 62:21–31.
- Kolmer JA, Herman A, Ordoñez ME, German S, Morgounov A, et al. 2020. Endemic and panglobal genetic groups, and divergence of host-associated forms in worldwide collections of the wheat leaf rust fungus *Puccinia triticina* as determined by genotyping by sequencing. *Heredity (Edinb)*. 124:397–409.
- Kolmer JA, Hughes ME. 2018. Physiologic specialization of *Puccinia triticina* on wheat in the United States in 2016. *Plant Dis*. 102:1066–1071.
- Kolmer JA, Long DL, Hughes ME. 2009. Physiologic specialization of *Puccinia triticina* on wheat in the United States in 2007. *Plant Dis*. 93:538–544.
- Kolmer JA, Long DL, Hughes ME. 2012b. Physiologic specialization of *Puccinia triticina* on wheat in the United States in 2010. *Plant Dis*. 96:1216–1221.
- Kolmer JA, Long DL, Hughes ME. 2013. Physiologic specialization of *Puccinia triticina* on wheat in the United States in 2011. *Plant Dis*. 97:1103–1108.
- Kolmer JA, Ordoñez ME, German S, Morgounov A, Pretorius Z, et al. 2019. Multilocus genotypes of the wheat leaf rust fungus *Puccinia*

- triticina in worldwide regions indicate past and current long-distance migration. *Phytopathology*. 109:1453–1463.
- Kolmer JA, Ordoñez ME, Manisterski J, Anikster Y. 2011. Genetic differentiation of *Puccinia triticina* populations in the Middle East and genetic similarity with populations in Central Asia. *Phytopathology*. 101:870–877.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26:589–595.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25: 2078–2079.
- Liu JQ, Kolmer JA. 1998. Molecular and virulence diversity and linkage disequilibria in asexual and sexual populations of the wheat leaf rust fungus, *Puccinia recondita*. *Genome*. 41:832–840.
- Liu M, Rodrigue N, Kolmer J. 2014. Population divergence in the wheat leaf rust fungus *Puccinia triticina* is correlated with wheat evolution. *Heredity (Edinb)*. 112:443–453.
- Long DL, Kolmer JA. 1989. A North America System of nomenclature for *Puccinia triticina*. *Phytopathology*. 79:525–529.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20: 1297–1303.
- Ordóñez ME, Kolmer JA. 2007. Simple sequence repeat diversity of a world-wide collection of *Puccinia triticina* from durum wheat. *Phytopathology*. 97:574–583.
- Ordóñez ME, Kolmer JA. 2009. Differentiation of molecular genotypes and virulence phenotypes of *Puccinia triticina* from common wheat in North America. *Phytopathology*. 99:750–758.
- Park R, Burdon J, Jahoor A. 1999. Evidence for somatic hybridisation in the leaf rust pathogen of wheat (*Puccinia recondita* f. sp. *tritici*). *Mycological Res*. 103:715–723.
- Park RF, Wellings CR. 2012. Somatic hybridization in the Uredinales. *Annu Rev Phytopathol*. 50:219–239.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 8:785–786.
- Saari EE, Young HC, Kernkamp MF. 1968. Infection of North American *Thalictrum* ssp. with *Puccinia recondita* f.sp. *tritici*. *Phytopathology*. 58:939–943.
- Saari EE, Prescott J. 1985. World Distribution In Relation To Economic Losses. Mexico: CIMMYT.
- Samborski DJ, Dyck PL. 1968. Inheritance of virulence in wheat leaf rust on the standard differential wheat varieties. *Can J Genet Cytol*. 10:24–32.
- Samborski DJ, Dyck PL. 1976. Inheritance of virulence in *Puccinia recondita* on six backcross lines of wheat with single genes for resistance to leaf rust. *Can J Bot*. 54:1666–1671.
- Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, et al. 2019. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol*. 3:430–439.
- Statler GD. 1977. Inheritance of virulence of culture 73-47 *Puccinia recondita*. *Phytopathology*. 77:906–908.
- Statler GD. 1979. Inheritance of pathogenicity of culture 70-1, Race 1, of *Puccinia recondita* *tritici*. *Phytopathology*. 69:661.
- Statler GD. 1982. Inheritance of virulence of *Puccinia recondita* f.sp. *tritici* on durum and spring wheat cultivars. *Phytopathology*. 72: 210–213.
- Statler GD. 2000. Inheritance of virulence of *Puccinia triticina* culture X47, the F₁ of the cross 71-112 x 70-1. *Can J Plant Pathol*. 22: 276–279.
- Swafford DL. 1998. PAUP. Phylogenetic Analysis Using Parsimony and Other Methods. Sunderland, MA: Sinauer Associates.
- Swafford DL. 2002. PAUP. Phylogenetic Analysis Using Parsimony and Other Methods. Version 4.0b10. Sunderland, MA: Sinauer Associates.
- Szabo LJ, Kolmer JA. 2007. Development of simple sequence repeat markers for the plant pathogenic rust fungus *Puccinia triticina*. *Mol Ecol*. 7:710.
- Webb CA, Szabo LJ, Bakkeren G, Staples R, Garry C, et al. 2006. Using biolistics with *Puccinia triticina* as a method of transient expression and insertional mutagenesis. *Funct Integr Genomics*. 6: 250–260.
- Wu JQ, Sakthikumar S, Dong C, Zhang P, Cuomo CA, et al. 2017. Comparative genomics integrated with association analysis identifies candidate effector genes corresponding to *Lr20* in phenotype-paired *Puccinia triticina* isolates from Australia. *Front Plant Sci*. 8:148.

Communicating editor: J. E. Stajich