July 2018

Project Report No. 630

Using field pathogenomics to study wheat yellow rust dispersal and population dynamics at a national and international scale

Diane G.O. Saunders

John Innes Centre

This is the final report of a 36-month project (2140023105) that started in July 2015. The work was funded by BBSRC and a contract for £45,000 from AHDB.

CONTENTS

1. ABSTRACT 1

2. INTRODUCTION 2

3. MATERIALS AND METHODS 3

3.1. Identification of single nucleotide polymorphisms (SNPs) 3

3.2. DNA extraction 4

3.3. Amplification of SNP sites 4

3.3.1. Design of oligonucleotides 4

3.3.2. Amplicon library preparation 4

3.4. KASP assays 5

3.4.1. Preparation of positive controls 5

3.4.2. KASPar assays 5

3.5. Analysis of amplicon data 5

4. RESULTS 6

4.1. Identification of SNP markers to distinguish PST genetic groups 6

4.1.1. Initial testing of SNP markers to distinguish PST genetic groups 7

4.1.2. Testing different methodology for SNP markers analysis 8

4.1.3. Utilising the Illumina custom amplicon strategy 9

4.1.4. Analysis of KAPA library preparations 9

4.1.5. KASPar assays 10

4.2. Genotyping field samples using KASPar assays 11

5. DISCUSSION 12

6. REFERENCES 13

1. Abstract

Traditionally, surveillance of rust fungal pathogens in agroecosystems has hinged on field biology and lengthy race-pathotype surveys to provide phenotypic information on pathogen diversity. However, assessments of genotypic diversity are not included routinely and, when employed, are restricted to just a handful of markers. In 2013, we initiated the development of a new approach called 'field pathogenomics' for pathogen population surveillance. This technique is based on high-resolution data, acquired using next-generation sequencing of field samples of yellow rust (*Puccinia striiformis* f. sp *tritici* – PST) infected wheat and triticale. While effectively capturing pathogen diversity, the analysis of infected host tissue can also be leveraged to assess the genotype of the host. The overall aim of this project was to apply this new, rapid gene-sequencing technology to the surveillance of wheat yellow rust, in collaboration with the UK Cereal Pathogen Virulence Survey (UKCPVS), and undertake comprehensive global population genetic analyses of this important plant pathogen. Within this large multi-institute and international project, funding from AHDB was provided to specifically contribute to further development of the field-pathogenomics method. The intention was to provide a simple, rapid genotyping method to track PST movement through the UKCPVS framework. To this aim, we identified a series of single nucleotide polymorphisms (SNPs) that could be used as genetic markers to differentiate the major wheat PST races currently in the UK. The next step will be to expand this set of initial marker regions to ensure they can detect any novel PST races that may enter the UK in future. In the meantime, the current markers will be integrated into the UKCPVS. In the long-term, this will ensure the UKCPVS is more predictive and better fulfils its main objective of gaining an early indication of new races and predicting durability of wheat varieties.

2. Introduction

Wheat is a critical staple providing 20% of the calories and over 25% of the protein consumed by humankind [1]. Wheat yellow rust disease, caused by the fungus *Puccinia striiformis* f. sp *tritici* (PST), is an historical and continuing threat to wheat production worldwide, reducing yields by up to 50% in untreated crops. In the last decade, new PST races have emerged that are capable of adapting to warmer temperatures, have expanded virulence profiles, and are more aggressive than previously characterized races [2]. More recently, PST has re-emerged as a major constraint on UK agriculture as races arising in Europe have overcome many of the major resistance genes in European germplasm. For instance, in 2011 a PST race group collectively called “Warrior” (based on the virulence of one of the initial variants of this group to the UK wheat variety Warrior) emerged as a serious threat to wheat production. The devastating impact of this disease gives a deep sense of urgency to breeders, farmers and the community at large to improve surveillance and employ comprehensive genetic analyses to understand the molecular basis of PST pathogenicity.

Traditionally, surveillance of rust fungal pathogens in agroecosystems has hinged on field biology and lengthy race pathotype surveys to provide phenotypic information on pathogen diversity. However, assessments of genotypic diversity are not included routinely and when employed are restricted to just a handful of markers such as simple sequence repeats (SSRs) or amplified fragment length polymorphisms (AFLPs) [3]. In the UK, the UK Physiologic Race Survey (later re-named UK Cereal Pathogen Virulence Survey: UKCPVS) was set up in 1967, largely in response to an unexpected and devastating epidemic of PST, due to the emergence of a new virulent pathotype. Since then, the UKCPVS has reported a steady succession of changes in pathogenicity in the UK yellow rust population, impacting on the varieties being grown by farmers and the new sources of resistance introduced by breeders. However, despite the success of the UKCPVS and our understanding of the phenotypic diversity of wheat yellow rust, we know almost nothing about its genetic diversity. The time-consuming nature of traditional surveillance methods also limits the number of PST isolates assessed each year. For instance, in the UK, a target number of twenty-five PST samples are tested each year, specifically focusing on those collected from wheat varieties with a previous record of good field resistance. With new wheat yellow rust pathotypes arising on susceptible varieties by mutation, recombination or through exotic incursions, it is unlikely that a new, rare pathotype would be detected in a timely fashion by the current surveillance system due to the low number of samples that can be assessed when reliant solely on traditional pathology. Furthermore, phenotypic characterization of pathogen isolates is limited to a snapshot of differential wheat varieties and cannot detect gain of virulence to varieties not included in the differential set.

We recently developed an approach called “field pathogenomics” for pathogen population surveillance based on high-resolution transcriptome data acquired directly from field samples of PST-infected wheat and triticale [4, 5]. This revealed a complete shift in the wheat yellow rust population with a number of new exotic races recently entering the UK and displacing the previous

population. Whilst effectively capturing pathogen diversity, RNA-seq of infected host tissue can also be leveraged to assess the genotype of the host. This is possible due to the availability of gene-specific wheat SNP markers [6, 7]. Furthermore, the approach uses RNA-seq data that is generated directly from pathogen-infected leaves collected in the field, providing a unique opportunity to characterize the pathogen population and its host directly in their natural environment. The assessment of genotypic diversity is not included routinely within rust pathogen surveillance activities. Our new RNA-seq based approach enables the rapid generation of high-resolution genotypic data for integration into pathogen surveillance activities that is vital to improve our understanding of the genetic sub-structure within a population [8].

The overall aim of this project was to apply rapid gene-sequencing technology to the surveillance of PST and undertake comprehensive global population genetic analyses of this important plant pathogen. Within this large multi-institute and international project, funding from AHDB was provided to specifically contribute to further development of the “field pathogenomics” method to provide a simple, rapid genotyping method that can be used to track PST movement through the UKCPVS framework in the future.

3. Materials and methods

3.1. Identification of single nucleotide polymorphisms (SNPs)

RNA-seq data was collated for a total of 296 PST-infected wheat, triticale and rye samples collected between 2013 and 2015 generated as described previously [4]. The FASTX-Toolkit (version 0.0.13.2) was used for quality filtering and trimmomatic (version 0.33) for removing the adapters. The 101-bp (HiSeq) pair-end reads were then aligned to the PST-130 assembly [9] using the TopHat package (version 2.0.11) and Bowtie alignment program (version 2.2.1) with default parameters and an expected mean inner distance between mate pairs of 200 [10, 11]. SNP-calling for PST was performed with SAMtools (version 0.1.19) [12], considering only the sites with a minimum depth of coverage of 20x. Homokaryotic and heterokaryotic SNP sites that induced synonymous and non-synonymous substitutions were identified using SnpEff, version 3.6 [13]. The genetic sub-structuration of PST isolates was assessed using multivariate analysis using discriminant analyses of principal components (DAPC) implemented in the Adegenet package in the R environment as described previously [8]. Synonymous SNP sites were then identified that were identical between isolates within a genetic group and yet different to all PST isolates within all other genetic groups. Several thresholds were tested by varying the percentage of isolates allowed to be different at a particular SNP site within a specific genetic group (Supplemental Table 1).

3.2. DNA extraction

DNA was extracted from PST-infected plant samples collected in the field using the MyTaq™ Extract-PCR kit, as described by the manufacturer (Bioline). In short, 20 μl of Buffer A, 20 μl of Buffer B and 70 μl of water was added to a total of 10-20mg of tissue that was disrupted using a homogenizer. Samples were then incubated for 5 minutes at 75ºC followed by 10 minutes at 95ºC. After extraction, DNA was purified using Agencourt® AMPure® XP magnetic beads (Beckman Coulter). DNA was added to the same volume of AMPure XP magnetic beads and incubated for 15 minutes at room temperature. The resulting solution was then placed in a magnetic rack and incubated for a further 5 minutes. Following removal of the supernatant, two washes were performed with 200 μl of 80% Ethanol. Finally, 52μl of water was added and the sample incubated for 2 minutes at room temperature. Samples were incubated on the magnetic rack for 15 minutes and the supernatant containing DNA removed. DNA was quantified using a Qubit fluorometer and PicoGreen quantification regents (Invitrogen).

3.3. Amplification of SNP sites

3.3.1. Design of oligonucleotides

To amplify specific regions of the PST genome that contained SNPs that can be used to distinguish the major genetic groups of PST in the UK primers were designed to amplify 400-500 bp regions surrounding each SNP site (Supplemental Table 2). Oligonucleotides were ordered from Eurofins. Primers were designed with a similar Tm and an average Tm of 57.12 °C was used to amplify all regions in a single PCR reaction using 5ng/μl of DNA and DreamTaq Green polymerase (ThemoFisher Scientific).

3.3.2. Amplicon library preparation

Two strategies for library preparation were trialled: (i) the KAPA HTP/LTP library preparation kit (Roche) and (ii) the Illumina TruSeq custom amplicon protocol. For the KAPA library preparation first amplicons were prepared using primers designed in house and then the resulting product used to generate sequencing libraries following the manufactures' instructions (Roche). Unique NEXTflex® DNA Barcodes (Bioo Scientific) were added to each sample in order to identify specific samples later in the sequence analysis. For the Illumina TruSeq custom amplicon method this utilised primers custom designed and optimised by Illumina. A total of 25 ng/μl input DNA was used and libraries prepared following the manufacturer's instructions. Adapters were added as with the KAPA libraries. Finally, a clean-up step with Agencourt® AMPure® XP magnetic beads, as previously described for DNA extraction, was performed and libraries were normalised as indicated by the manufacturer (Illumina).

3.4. KASP assays

3.4.1. Preparation of positive controls

DNA was generated from 4 PST isolates, one from each of the 4 genetic groups (group 1, group 3, group 4 and group 5-1) (Table 1) to use as positive controls. First, wheat plants from the susceptible variety (Vuka) were grown under controlled environment conditions as described previously [4]. Four days after sowing, plants were treated with 25 ml of growth regulator (Maleic Hydrozide 0.2g/l). To facilitate inoculation, spores were heat-activated at 40°C for 5 minutes and then re-suspended in Novec 7100 (1 mg/mL) prior to inoculation. After 14 days, wheat plants were spray inoculated with the heat-activated spores from each of the PST isolates independently and kept in the dark at 10 °C for two days under high humidity conditions before returning the plants to controlled environment conditions. Infected tissue was collected 14 days post-inoculation and stored in RNAlater® solution at 4°C (Life technologies, UK). DNA was then extracted from these samples, as previously explained, to be used in subsequent KASPar assays.

Table 1. PST isolates used as positive controls.

Genetic Group	Isolate name	Year of collection	Origin
1	F18	2014	UK
3	15/151	2015	UK
4	14/7	2014	UK
5-1	14/106	2014	UK

3.4.2. KASPar assays

For each marker, three primers were designed using Primer3plus software (version 4.0.0.), one specific for each possible base at that position and a third common primer. Each of the specific primers was designed carrying standard FAM or HEX compatible tails (FAM tail: 5' GAAGGTGACCAAGTTCATGCT 3'; HEX tail: 5' GAAGGTCGGAGTCAACGGATT 3') and with the target SNP at the 3' end. Oligonucleotides were ordered from Sigma and primer mixes were as recommended by the manufacturer [46 µL dH_2O, 30 µL common primer (100 µM), and 12 µL each tailed primer (100 µM); LGC Genomics]. Assays were carried out as described previously [4] with the following modifications: each reaction contained 2.5 µl of the V4 2xKaspar mix, 0.07 µl of primer mix and 2.5 µl of sample DNA (5-30 ng/µl). PCR cycling was performed in an Eppendorf Mastercycler pro 384 and 384-well optically clear plates (Roche) were read on a Tecan Safire plate reader. Data analysis was performed manually using Klustercaller software (version 2.22.0.5, LGC).

3.5. Analysis of amplicon data

The amplicon libraries were sequenced on the Illumina HiSeq 2500 machine at the Earlham Institute, UK. The FASTX-Toolkit (version 0.0.13.2) was used for quality filtering and trimmomatic (version 0.33) for removing the adapters. The pair-end reads were aligned to the PST-130 assembly [9] using

bowtie2 (version 2.1.0) and then SAMtools (version 0.1.19) was used for SNP-calling, considering sites that had a depth of coverage of at least 20x. The nucleotide positions for the 98 SNP markers were extracted to determine the most likely genetic group for a particular sample by scoring 1 every time the base at that position matched one from a specific genetic group. Samples were assigned to the genetic group with the highest score.

4. Results

4.1. Identification of SNP markers to distinguish PST genetic groups

Five different genetic groups of PST were found to infect wheat in the UK in recent years. Four of these were identified and described in 2013 (Groups 1 - 4; Hubbard *et al.*, 2015) and a fifth race was identified in 2014 that was assigned to the Kranich race (Group 5-1). To allow new PST field samples to be rapidly assigned to one of these five genetic groups, we decided to focus on identifying a limited set of genetic markers, rather than carrying out full transcriptomic sequencing as done previously [4]. To this aim, we analysed data from more than 600 PST isolates collected in the UK during 2013, 2014 and 2015 to identify a number of single nucleotide polymorphisms (SNPs) specific for each genetic group. First, we used representative isolates from 2013 (only collected within the UK), then we included all PST isolates from 2014 and 2015 (including isolates from outside the UK). As a number of PST isolates were not well assigned to a certain genetic group and/or had lower amounts of data from the pathogen we focused on a core set of 296 well defined PST isolates for identifying genetic markers. Using this dataset all synonymous SNP sites were extracted for each PST isolate and then compared to identify specific markers for each of the five genetic groups. The parameters for selecting a marker that could be used to define a genetic group were as following:

1. The SNP had to be present in all populations
2. The SNP should have coverage in all PST isolates
3. The SNP should have coverage within each genetic group
4. The SNP marker should be different between at least two genetic groups

However, the high variability that exists between isolates within a single genetic group [4, 8] made it difficult to identify consensus markers within and between populations. Furthermore, low level sequencing errors may complicate this analysis further. Therefore, we included a margin of error in the analysis during marker selection, while setting a threshold to make sure that selected markers were sufficiently reliable. A vast number of specific markers were found for genetic group 2, which was expected since PST isolates within this genetic group are mainly found on triticale or rye and not on wheat like all PST isolates in other genetic groups. Likewise, an abundance of markers could be identified for genetic groups 1, 3 and 5-1. However, it was difficult to identify specific markers for genetic group 4 due to the high degree of genetic variability within this genetic group [8] (Figure 1 a,b).

Markers were selected considering both homozygous and heterozygous positions and varying the threshold for missing data. A threshold of 50% of missing data across all PST isolates within a population group was accepted, when no sequencing errors were present. The amount of missing data was decreased to 30% when 10% when sequencing errors were considered in the analysis. A set of 98 markers were obtained and then tested computationally using RNA-seq data from all 600 samples from the Field Pathogenomics project, with a high rate of success: only samples with low quality data or those not well assigned to any specific genetic group could not be identified.

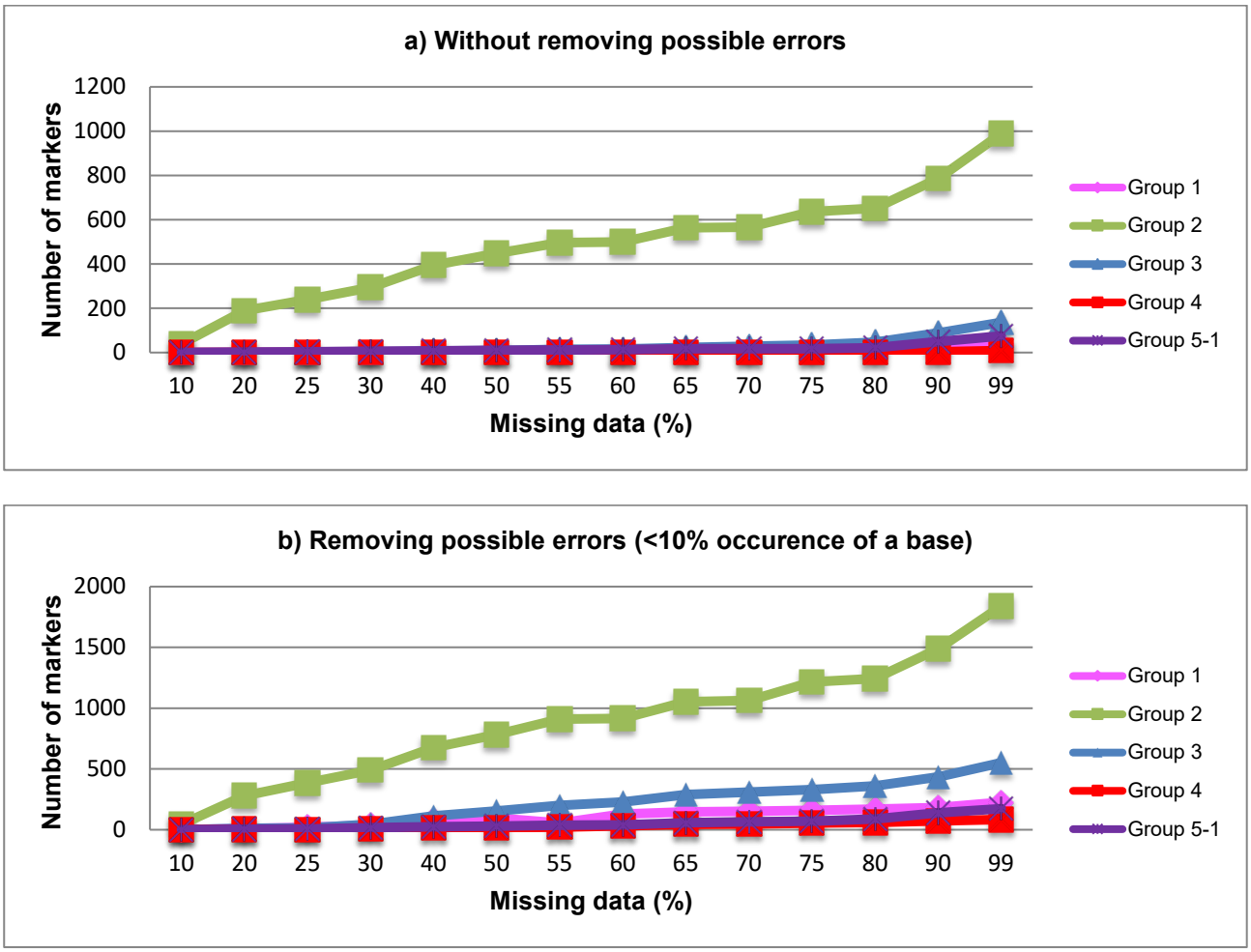

Figure 1. Number of SNP markers identified increased as stringency of missing data filter was reduced. Number of markers compared to percentage of missing data (a) without considering sequencing error and (b) considering 10% of missing data as error.

4.1.1. Initial testing of SNP markers to distinguish PST genetic groups

KASPar assays were developed to test a subset of the genetic markers identified above. A total of 11 markers that were able to differentiate the PST populations were selected for this initial analysis (Table 2). Three known control isolates were used, one for each genetic group tested (Group 1: 13/26, Group 3: 13/182, Group 4: 13/36) and forty-two PST-infected field samples from 2015 were analysed using this smaller set of markers (Table 2). When combined, the markers were clearly able

to distinguish each of the genetic groups analysed. This indicates that this set of markers is able to differentiate the different PST genetic groups evaluated, although further analysis is required to validate the entire set of 98 markers.

Table 2. Provisional analysis of 11 KASPar assays showed that these were sufficient for distinguishing the three major wheat PST races in the UK (Groups 1, 3 and 4). KASPar analysis of a set of markers using 3 isolates whose genetic group is known. Red shading refers to an assay assigned as Y:Y, green to X:Y and blue to X:X.

	Contig	Position	Allele X	Allele Y	Group 1 isolate 13/26	Group 3 isolate 13/182	Group 4 isolate 13/36
Assay 1	PST130_10127	1130	G	A	Y:Y	X:Y	X:Y
Assay 2	PST130_10441	476	C	T	Y:Y	X:Y	X:Y
Assay 3	PST130_7670	3524	C	T	X:X	Y:Y	X:Y
Assay 4	PST130_10982	2778	G	A	X:X	Y:Y	X:Y
Assay 5	PST130_7291	1846	G	A	Y:Y	X:Y	X:Y
Assay 6	PST130_11355	649	A	C	X:X	Y:Y	X:Y
Assay 7	PST130_10126	6216	C	A	*	X:Y	*
Assay 8	PST130_16961	1098	C	T	*	*	*
Assay 9	PST130_22010	172	C	T	Y:Y	X:X	X:X
Assay 10	PST130_28344	2372	A	G	Y:Y	Y:Y	Y:Y
Assay 11	PST130_8160	11876	A	G	*	X:Y	X:Y

4.1.2. Testing different methodology for SNP markers analysis

Three different methods were used to evaluate the above 98 SNP markers. Two of these methods were based on next-generation sequence analysis using two different strategies for library preparation, and the third method utilised KASPar assays to detect single SNPs. In order to evaluate these different approaches, we first selected a set of PST samples whose genetic group was known to utilise as positive controls in the analysis (Table 3). The selected samples consisted of PST-infected wheat leaves that were collected from the field in 2016 and previously subjected to full transcriptome sequencing and subsequent genetic analysis.

Table 3. Number of PST-infected samples used from each genetic group.

Genetic Group	KAPA libraries	Illumina libraries
1	13	8
3	8	3
4	36	33
5-1	8	4

4.1.3. Utilising the Illumina custom amplicon strategy

To evaluate the Illumina custom amplicon method a pool of primers were designed by Illumina that amplified 100 bp surrounding the SNP site of interest. A total of 91 libraries were prepared, of which 49 were from PST-infected samples that had been previously genotyped using RNA-seq analysis and thus their genetic group was known. The other 42 were samples collected from the 2016 field trials. Using this technique, out of the 49 samples where we had previously determined the genetic group (Table 3), we could identify the genetic group correctly for all 44 samples that produced sufficient data for analysis. The remaining 5 samples had very low concentrations of DNA and therefore failed to generate sufficient sequencing data. Out of the 42 that were collected during the field trials, we could define the genetic group for all samples, which all belonged to Group 4. For most samples, we used between 50-60 of the 98 markers to identify the genetic group. In fact, we found that just using a small subset of SNPs was enough to correctly identify the genetic group (Figure 2).

		GROUP 1			GROUP 3			GROUP 4			GROUP 5_1 (KRANICH)		
	Group	16.0164	F18	16.0127	16.0676	16.0401	16.0402	16.0669	16.0203	15/157	16.0183	16.0186	14/106
10441_476	1,5_1	TT	TT	TT	CT	CT	-	CT	CT	CT	TT	TT	TT
10441_749	1,5_1	TT	TT	TT	CT	-	CT	CT	CT	CT	TT	TT	TT
6921_3712	2, 3	TT	TT	TT	CT	CT	CT	TT	TT	TT	TT	TT	TT
7364_28843	1	AT	AT	AT	TT	TT	TT	TT	TT	TT	TT	TT	TT
7446_5852	2,5_1	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	AG	AG
8552_2672	ALL	AA	AA	AA	AG	AG	AG	AG	AG	AG	AA	AA	AA
7237_8357	ALL	GG	GG	GG	CC	CC	CC	CG	CG	CG	GG	GG	CG
7291_1846	1, 2	AA	AA	AA	AG	AG	AG	AG	AG	AG	AG	AG	AG
7582_7250	2,4	CT	CT	CT	CT	CT	CT	CC	CC	CC	CT	CT	CT
7582_8369	2,4	AG	AG	AG	AG	AG	AG	GG	GG	GG	GG	GG	GG
9877_1344	2, 3	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
7341_4368	1,5_1	AT	AT	AT	TT	TT	TT	-	TT	TT	AT	AT	AT
7341_4770	3	CT	-	CT	TT	TT	-	CT	CT	CT	-	CT	-

Figure 2. A small subset of 13 SNP markers were sufficient to correctly identify the genetic group for 12 samples that spanned the four wheat PST genetic groups in the UK. Example of 12 samples that were sequenced whose genetic group was known for 13 markers.

4.1.4. Analysis of KAPA library preparations

To evaluate the KAPA library preparation technique, we first designed primers to amplify a ~400 bp region surrounding each of the 98 amplicons, with two of these amplicons containing 2 SNP markers per amplicon. Of these 96 primer pairs, 62 amplified the expected sized fragment. The remaining primer pairs (34) did not produce a PCR product or produced non-specific products as indicated by multiple bands following gel electrophoresis (Figure 3; Supplemental Table 3). The 62 successful primer pairs were then used to amplify the corresponding regions from the PST-infected samples prepared as positive controls above. The resulting amplicons were then used to make KAPA sequencing libraries. However, as the Illumina strategy was run in parallel and was extremely successful in assigning PST isolates to the correct genetic groups, this strategy was not pursued beyond library preparation.

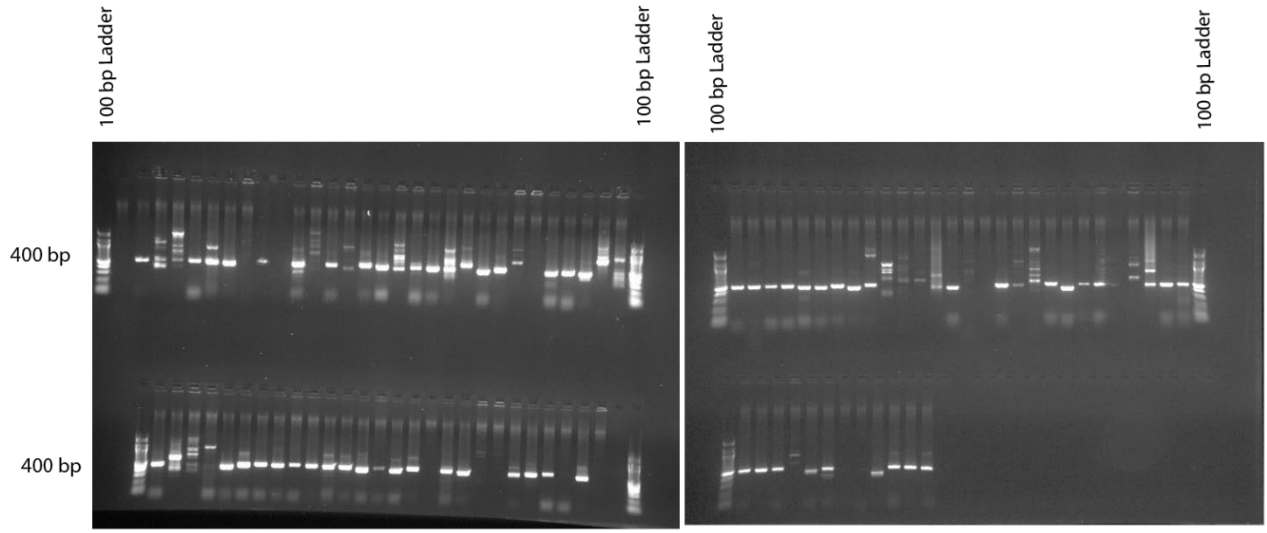

Figure 3. Sixty-two of the ninety-six primer pairs tested amplified successfully. Gel of PCRs to test the 96 amplicons designed, Tm 57.12°C. PCR reactions were loaded for each amplicon in the order described in the Supplemental Table 3.

4.1.5. KASPar assays

As a more cost-effective method we also designed and tested KASPar assays specific to each SNP site as an alternative strategy to define PST genetic groups. First, primers were designed for all 98 SNP positions. However, primers could not be designed to amplify all 98 specific SNPs due to the high level of GC content or repeats in some regions. Primers were thus designed for a subset of 36 SNP markers. A total of 5 samples were utilised as positive controls for testing the KASPar assays that represented the four genetic groups of PST that infect wheat: PST isolates F18 and F22 (both Group 1), PST isolate 15/151 (Group 3), PST isolate 14/7 (group 4) and PST isolate 14/106 (Group 5-1). Initial analysis revealed that many of the KASPar assays were inconsistent (Figure 4). Therefore, we selected four primer triplets that could be used to reliably differentiate between the four genetic groups: CTPST163, 167, 16 and 18 (Figure 5). Once the primers had been validated using positive controls, the four primers were further evaluated using additional PST-infected samples that were also subjected to full transcriptome sequencing. The KASPar primers were very effective in distinguishing PST genetic groups, with identical results compared to the Illumina sequencing. For instance, out of the 49 samples sequenced using the amplicon Illumina sequencing method, 10 were also tested in KASPar analysis, and gave the same result.

	PRIMER	CLPST160	CLPST163	CLPST167	CLPST171	PST-16	PST-17	PST-18	PST-23	PST-29	PST-19
	Marker	7670_3437	10982_2778	7291_1846	11355_649	7364_28843	753_1139	8291_7564	19742_674	7237_13346	8552_2672
	SNP	C3524T	G2778A	G1846A	A649C	t28843a	c1139a	a7564g	c674t	g13346c	A2672G
	Specific	ALL	ALL	Group 1	ALL	Group 1	Group 1, 5-1	Group 5-1	ALL	ALL	Group 3
Expected	Group 4	XY	XY	XY	XY	XX	XY	XX	XY	XY	XY
	Group 3	XX	YY	XY	YY	XX	XY	XX	YY	YY	XY
	Group 1	YY	XX	YY	XX	XY	XX	XX	XX	XX	XX
	Group 1	YY	XX	YY	XX	XY	XX	XX	XX	XX	XX
KASP results	Group 4	XX	XY	XY	XX	XX	XX	XX	XX	XY	-
	Group 3	YY	YY	XY	XY	XX	XX	XX	XX	XX	-
	Group 1	XX	XX	YY	XX	XY	XX	XX	XX	XY	-
	Group 1	XX	XX	YY	XX	XY	XX	XX	XX	XY	-

Figure 4. Many of the KASPar assays gave inconsistent results. Example of various primer tests for some of the selected markers, using control samples. The cells were highlighted in green if the expected and KASP results for the base being calling matched.

	F18	F22	15/151	14//7	14/109
CLPST163	XX	XX	YY	XY	XX
CLPST167	YY	YY	XY	XY	XY
PST-16	XY	XY	XX	XX	XX
PST-18	XX	XX	XX	XX	XY

Figure 5. Four primer triplets could be used to reliably differentiate between the four wheat PST genetic groups. Expected base calling for the selected positive control samples for the four selected primers.

4.2. Genotyping field samples using KASPar assays

As the Illumina amplicon strategy and KASPar assays were both effective in assigning PST genetic groups we chose to genotype a wider range of field samples using the quicker and more economical KASPar method. We selected a total of 6 samples to use as positive controls in each assay that represented each of the four PST genetic groups that infect wheat and one negative control with wheat DNA (Figure 6). The PST isolates selected as positive controls were F18, F22 (both Group 1), 15/151 (Group 3), 14/7 (group 4) and 14/106 (Group 5-1).

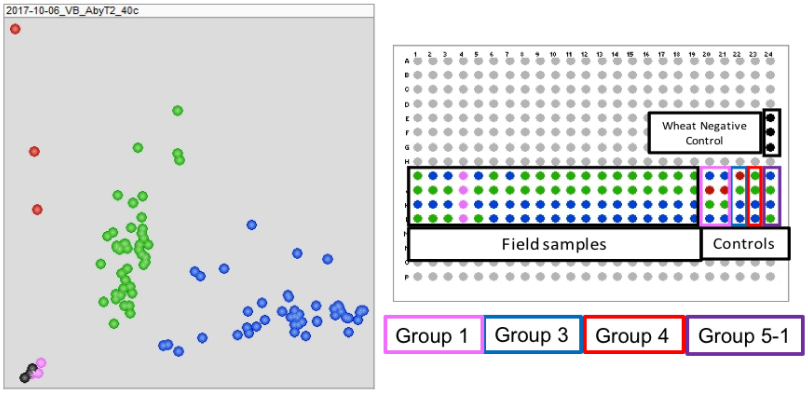

Figure 6. KASPar analysis of 19 PST-infected field samples and 5 positive controls illustrates the ability to clearly distinguish the 4 major wheat PST races found in the UK. The positive controls consisted of two from group 1, one from group 3, one for group 4 and one for group 5-1. A negative control of wheat DNA was also included. In the scatter plot, red dots refer to samples assigned as Y:Y, green to X:Y and blue to X:X.

Next, we used 5 validated KASPar assays to examine 85 field samples from field trials carried out in the 2015-2016 season and 374 field samples from the 2016-2017 season to evaluate the distribution of PST genetic groups in a field across the wheat growing season. Figure 7 illustrates how the PST genetic groups identified throughout the growing season changed in a single field in Taunton, Devon, sampled in 2016. However, when we genotyped the samples collected from the same field in the following growing season, we found that all samples belonged to the same genetic group: Group 4. This was a trend that was evident at all sites assessed in 2017, where all 374 samples that were genotyped belonged to genetic group 4 and reflects the increasing dominance of genetic group 4 across the UK [8].

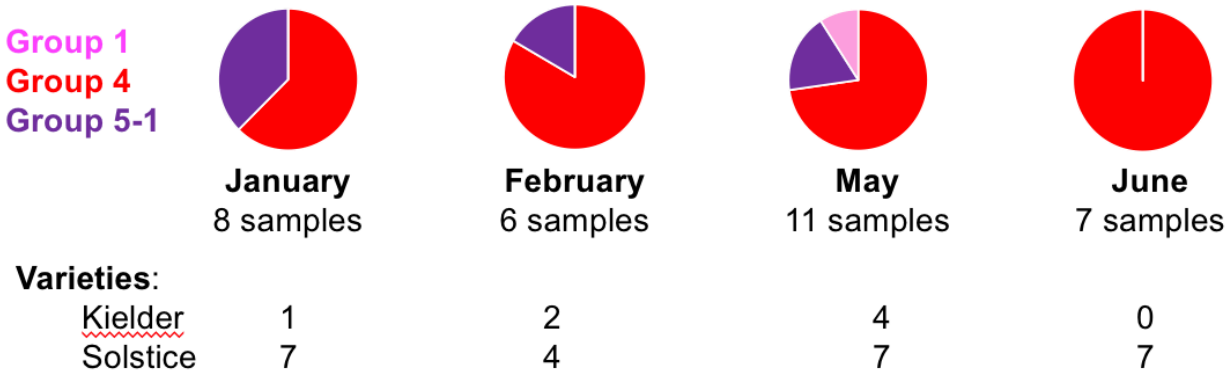

Figure 7. Group 4 PST isolates increased in prevalence throughout the 2015-2016 growing season at the Taunton field trial site. Distribution of samples collected from each genetic group throughout the 2016 wheat growing season at a single field trial site. Each field trial site consisted of two wheat varieties (Kielder and Solstice) and the number of samples taken from each variety are indicated.

5. Discussion

Traditionally, the surveillance of rust fungal pathogens in agroecosystems has hinged on field biology and race pathotype surveys to provide phenotypic information on pathogen diversity [14]. However, assessments of genotypic diversity are not included routinely and when employed are restricted to just a handful of markers such as simple sequence repeats (SSRs) or amplified fragment length polymorphisms (AFLPs) [3]. Our field pathogenomics approach, which is reliant on full transcriptome sequencing of PST-infected field samples, enables the integration of high-resolution genotypic data into pathogen surveillance activities [4]. Our approach also captures the PST population directly from the field and negates any biases that might be caused by purification and multiplication of the pathogen in the laboratory, a lengthy process that can impose artificial selection on the pathogen. However, full transcriptome sequencing is costly. Here, we have utilised the data generated in the wider field pathogenomics project to develop a series of SNP markers that can be used as an initial quick and cheap test to assign PST isolates to one of the five major PST races found within the UK. This is initially still dependant on a small number of markers that were specifically designed to assess

the distribution of the current UK-specific PST races. However, the next step will be to expand this set of initial marker regions to ensure they can detect any novel PST races that may enter the UK in future. In the meantime, the current markers will be integrated into the UKCPVS. In the long-term this will ensure the UKCPVS is more predictive and better fulfils its main objective of gaining an early indication of new races and predicting durability of wheat varieties. For instance, generating a better understanding of the underlying variation in the PST population and how this will affect resistance breeding is essential for the strategic planning required to produce high-yielding wheat varieties which will be resistant when deployed in 10 years (considering the time between initial crossing and variety release). Furthermore, in the future it will enable the UKCPVS to rapidly monitor development of potential fungicide resistance mutations that would then provide early warning to agronomists and farmers of potential resistance emerging within UK PST populations. This would then place the UK at the forefront of the most progressive pathogen surveillance mechanisms and enable the UKCPVS to better understand increasingly complex pathogen populations.

6. References

1. **Food and Agriculture Organization of the United Nations** [http://faostat.fao.org]
2. Hovmoller MS, Walter S, Justesen AF: **Escalating threat of wheat rusts.** *Science* 2010, **329:**369.
3. Kolmer JA: **Tracking wheat rust on a continental scale.** *Curr Opin Plant Biol* 2005, **8:**441-449.
4. Hubbard A, Lewis CM, Yoshida K, Ramirez-Gonzalez RH, de Vallavieille-Pope C, Thomas J, Kamoun S, Bayles R, Uauy C, Saunders DG: **Field pathogenomics reveals the emergence of a diverse wheat yellow rust population.** *Genome Biol* 2015, **16:**23.
5. Bueno-Sancho V, Bunting DCE, Yanes LJ, Yoshida K, Saunders DGO: **Field Pathogenomics: An Advanced Tool for Wheat Rust Surveillance.** *Methods Mol Biol* 2017, **1659:**13-28.
6. Rodriguez-Guerra R, Ramirez-Rueda MT, de la Vega OM, Simpson J: **Variation in genotype, pathotype and anastomosis groups of *Colletotrichum lindemuthianum* isolates from Mexico.** *Plant Pathology* 2003, **52:**228-235.
7. Wang SC, Wong DB, Forrest K, Allen A, Chao SM, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, et al: **Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array.** *Plant Biotechnology Journal* 2014, **12:**787-796.
8. Bueno-Sancho V, Persoons A, Hubbard A, Cabrera-Quio LE, Lewis CM, Corredor-Moreno P, Bunting DCE, Ali S, Chng S, Hodson DP, et al: **Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity.** *Genome Biol Evol* 2017, **9:**3282-3296.
9. Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, Jurka J, Michelmore RW, Dubcovsky J: **Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust.** *PLoS One* 2011, **6:**e24230.
10. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7:**562-578.
11. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.

12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.
13. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3.** *Fly (Austin)* 2012, **6:**80-92.
14. Park R, Fetch T, Hodson D, Jin Y, Nazari K, Prashar M, Pretorius Z: **International surveillance of wheat rust pathogens: progress and challenges.** *Euphytica* 2011, **179:**109-117.