

ARTICLE

Received 27 Mar 2013 | Accepted 26 Sep 2013 | Published 23 Oct 2013

DOI: 10.1038/ncomms3673

OPEN

# High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus

Wenming Zheng<sup>1,2,\*</sup>, Lili Huang<sup>1,\*</sup>, Jinqun Huang<sup>3,\*</sup>, Xiaojie Wang<sup>1</sup>, Xianming Chen<sup>4</sup>, Jie Zhao<sup>1</sup>, Jun Guo<sup>1</sup>, Hua Zhuang<sup>1</sup>, Chuangzhao Qiu<sup>3</sup>, Jie Liu<sup>1</sup>, Huiquan Liu<sup>1</sup>, Xuelling Huang<sup>1</sup>, Guoliang Pei<sup>1</sup>, Gangming Zhan<sup>1</sup>, Chunlei Tang<sup>1</sup>, Yulin Cheng<sup>1</sup>, Minjie Liu<sup>1</sup>, Jinshan Zhang<sup>1</sup>, Zhongtao Zhao<sup>1</sup>, Shijie Zhang<sup>1</sup>, Qingmei Han<sup>1</sup>, Dejun Han<sup>1</sup>, Hongchang Zhang<sup>1</sup>, Jing Zhao<sup>1</sup>, Xiaoning Gao<sup>1</sup>, Jianfeng Wang<sup>1</sup>, Peixiang Ni<sup>3</sup>, Wei Dong<sup>3</sup>, Linfeng Yang<sup>3</sup>, Huanming Yang<sup>3</sup>, Jin-Rong Xu<sup>5</sup>, Gengyun Zhang<sup>3</sup> & Zhensheng Kang<sup>1</sup>

Stripe rust, caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*), is one of the most destructive diseases of wheat. Here we report a 110-Mb draft sequence of *Pst* isolate CY32, obtained using a 'fosmid-to-fosmid' strategy, to better understand its race evolution and pathogenesis. The *Pst* genome is highly heterozygous and contains 25,288 protein-coding genes. Compared with non-obligate fungal pathogens, *Pst* has a more diverse gene composition and more genes encoding secreted proteins. Re-sequencing analysis indicates significant genetic variation among six isolates collected from different continents. Approximately 35% of SNPs are in the coding sequence regions, and half of them are non-synonymous. High genetic diversity in *Pst* suggests that sexual reproduction has an important role in the origin of different regional races. Our results show the effectiveness of the 'fosmid-to-fosmid' strategy for sequencing dikaryotic genomes and the feasibility of genome analysis to understand race evolution in *Pst* and other obligate pathogens.

<sup>1</sup>State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China. <sup>2</sup>Key Laboratory of Physiology, Ecology and Genetic Improvement of Food Crop in Henan Province and College of Life Sciences, Henan Agricultural University, Zhengzhou, Henan Province 450002, China. <sup>3</sup>State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China. <sup>4</sup>USDA-ARS and Department of Plant Pathology, Washington State University, Pullman, Washington 99164-6430, USA. <sup>5</sup>Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Z.K. (email: kangzs@nwsuaf.edu.cn) or to G.Z. (email: zhanggengyun@genomics.org.cn) or to J.-R.X. (email: jinrong@purdue.edu).

Stem, leaf and stripe rust diseases of wheat caused by *Puccinia graminis* f. sp. *tritici* (*Pgt*), *P. triticina* (*Pt*) and *P. striiformis* f. sp. *tritici* (*Pst*) have caused significant yield losses during epidemics that have resulted in famines throughout human history. They continue to threaten worldwide wheat production and global food security<sup>1–3</sup>. To prevent wheat rust epidemics, an international consortium known as the Borlaug Global Rust Initiative (<http://www.globalrust.org/>) was established. Of the three wheat rust diseases, stripe rust occurs most frequently in the United States, primarily due to infection during the early growing season, long-distance air-borne spore dispersal and a recent adaptation to warmer temperatures<sup>4,5</sup>. Although *Pst* populations in most areas around the world have low genetic diversity due to clonal asexual reproduction, high levels of genetic diversity have been observed in some *Pst* populations from western China and central Asia<sup>6,7</sup>. Recently, *Berberis* species were reported to serve as the alternate host of *Pst* in western China and United States<sup>8,9</sup>. Therefore, sexual reproduction may be an important factor accounting for genetic variation in *Pst*, and regions in central Asia may serve as the centres of origin for new races.

Better understanding of virulence variations and mechanisms of rust fungus–wheat interactions is necessary for the development and deployment of more effective and durable resistant cultivars<sup>10</sup>. However, wheat rust fungi are obligate biotrophic pathogens that cannot be cultured on artificial media. Most of their life stages, including urediniospores that are commonly used for aetiological and evolutionary biology studies, are dikaryotic (Fig. 1). Therefore, it is extremely difficult to conduct molecular studies and functional characterizations of genes in rust fungi<sup>10,11</sup>. Nevertheless, recent technological advances have enabled the use of genomic approaches to study virulence variation, development and evolution in rust and other obligate biotrophic fungi<sup>12,13</sup>.

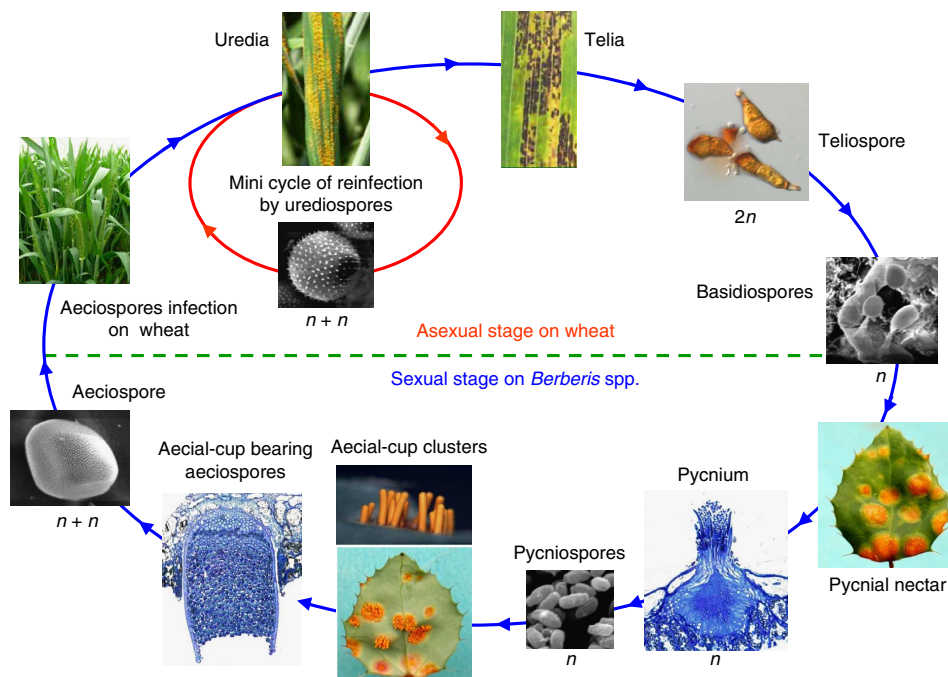
Among the rust fungi that have been sequenced, the genome sequences of *Pgt* and *Melampsora larici-populina* (*Mlp*) were generated by Sanger sequencing<sup>12</sup>. In 2011, a rough draft

sequence of a US isolate of *Pst* was generated with short reads from next-generation sequencing (NGS)<sup>14</sup>. All of these rust genome projects used genomic DNA isolated from dikaryotic (N + N) urediniospores for sequencing analyses. The two nuclei in urediniospores are highly heterozygous in rust fungi, a characteristic that has confounded genome assembly by conventional approaches, especially with short NGS reads. The coverage and contig integrity of the released *Pst* draft sequence (contig N50 of 5 kb) were obviously hampered by the heterozygosity<sup>14</sup>.

In this study, we generate a significantly improved draft genome (~110 Mb) of a Chinese *Pst* isolate, CY32, using a ‘fosmid-to-fosmid’ sequencing strategy that combines the use of a fosmid library and Illumina sequencing. This isolate has been the most prevalent virulent strain in China since 1991 and caused severe yield losses in the 2002 wheat stripe rust epidemic<sup>15</sup>. The assembled genome sequence of CY32 is 130.7 Mb, which is approximately twice as large as the released *Pst*-130 genome (64.8 Mb)<sup>14</sup>. Our results show that the ‘fosmid-to-fosmid’ strategy is more effective than the conventional whole-genome shotgun (WGS) approach for sequencing dikaryotic rust fungi. In addition, we resequence five prevalent *Pst* isolates collected from different geographical locations. We identify a large number of SNPs using whole-genome comparative analyses between *Pst*-CY32 and individual re-sequenced isolates. Approximately 35% of SNPs are in the coding sequence (CDS) regions and half of them are non-synonymous, indicating significant genetic variation among these *Pst* isolates. Our re-sequencing data implies a significant role for sexual reproduction in evolution of *Pst* races.

## Results

**Genome sequencing and assembly.** A genomic DNA library consisting of 19,200 fosmid clones with an average insert size of 36 kb was constructed (~6.3-fold genome coverage) and randomly pooled into 1,920 pools (10 clones per pool). Each of the



**Figure 1 | Life cycle of *Puccinia striiformis* Westend f. sp. *tritici*.** In most of the wheat-growing regions worldwide, urediniospores are the only inoculum for the initial and recurrent infection of wheat plants, which is known as the mini cycle of *Pst*. Although the alternate host was identified recently, the role of sexual reproduction in genetic variation and race evolution of *Pst* is largely unknown.

barcoded pools (each with a unique index) was sequenced to ~228-fold depth using Illumina GA and assembled separately using SOAP *de novo*<sup>16,17</sup> ('fosmid-to-fosmid'). The primary assembled contigs were used for simulating sequencing and merged into large fragmental reefs (supercontigs) with the overlap layout consensus (OLC)<sup>18</sup> to generate the secondary assembly. Self-to-self whole-genome alignment with LASTZ and sequencing depth information were used to remove redundancy in the assembly<sup>19</sup>. The WGS sequencing approach was employed to sequence paired-ends (PE) genomic libraries that had an average insert size of 200 bp to 6 kb with Illumina GA. The resulting Illumina PE reads (average of 67 bp) in combination with mate pair information from fosmid-pooling reads were used for scaffolding of all the reefs (supercontigs) after filtering out low-quality reads. Because urediniospores used for DNA isolation were harvested from infected wheat leaves instead of sterile cultures, the assembled scaffolds were searched against the nucleotide database to remove possible contamination from plant, bacteria and insects (Supplementary Fig. S1). The combined length of assembled scaffolds is 130.7 Mb. The N50 of

scaffolds and contigs reach 125 kb and 18 kb, respectively, which is a significant improvement compared with the *Pst* draft assembled solely with WGS Illumina reads (Table 1).

The *Pst* genome size was estimated with Illumina sequencing data of three WGS PE libraries (with average insert sizes of 224, 311 and 515 bp, respectively) based on k-mer (k-base fragment) analysis<sup>20</sup> and the distribution of contig length and depth. Two peaks (16 and 30) were identified in the 17 k-mer depth distribution of WGS Illumina reads (Fig. 2a), and similarly, two peaks of coverage (Cvg) (7 and 17) were identified in the distribution of contig length and depth (Fig. 2b), indicating that the two haploid nuclei of *Pst* urediniospores differ significantly in genome sequences. Considering the heterozygous nature of dikaryotic urediniospores, the haploid genome size of *Pst* was estimated to be ~110 Mb (Supplementary Fig. S2, Supplementary Tables S1–S4, Supplementary Notes 1 and 2).

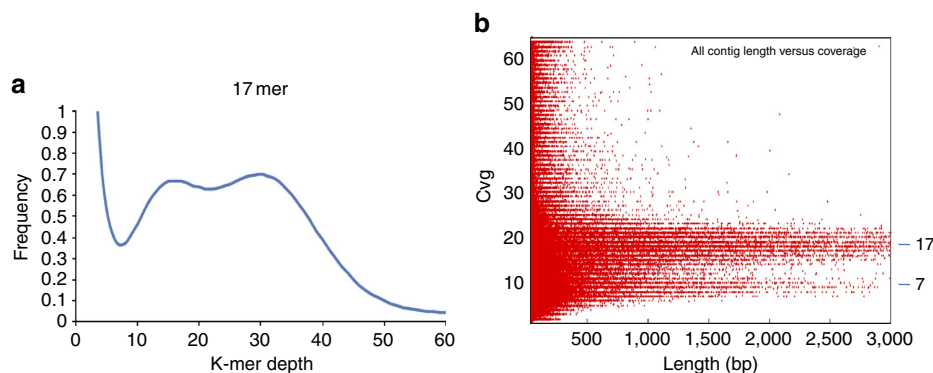
Approximately 49% of the *Pst* genome assembly consists of transposable elements (TEs) or repetitive sequences. The class I LTR retro-elements and class II terminal inverted repeat (TIR) DNA transposons are among the most abundant ones and account for ~27% and ~17% of the *Pst* genome, respectively. The same pipeline was used to identify the TE or repetitive sequences in the other two sequenced rust fungi, *Pgt* and *Mlp*. We found that the proportion of TEs or repeats in the *Pgt* and *Mlp* genomes was ~48% and ~49%, respectively (Supplementary Tables S5–S7). Whereas the *Mlp* genome is enriched for TIR class II DNA transposons and DIRS class I retrotransposons, the *Pgt* genome has more LTR class I retrotransposons (Supplementary Fig. S3). The role of these TEs in the evolution of rust genomes requires further investigation.

**Validation of sequence assembly.** To validate the 'fosmid-to-fosmid' assembly of the *Pst*-CY32 genome, 10 randomly selected fosmids were individually sequenced from end to end by Sanger sequencing. The mapped ratio of assembled scaffolds of Illumina reads to the Sanger-sequenced fosmids reached 98.53% (Supplementary Table S8). Nine of them were well covered by assembled scaffolds of *Pst*-CY32. The other Sanger-sequenced fosmid contained highly repetitive sequences or SV (structural variation) regions, which was revealed by alignment with unfiltered contigs and Illumina paired-end sequences (Supplementary Fig. S4). In comparison with the *Pst*-130 assembly solely using WGS Illumina sequences, the 'fosmid-to-fosmid' strategy significantly improved the assembly integrity of the highly heterozygous *Pst* genome. For example, scaffold 475 of *Pst*-CY32 contains only

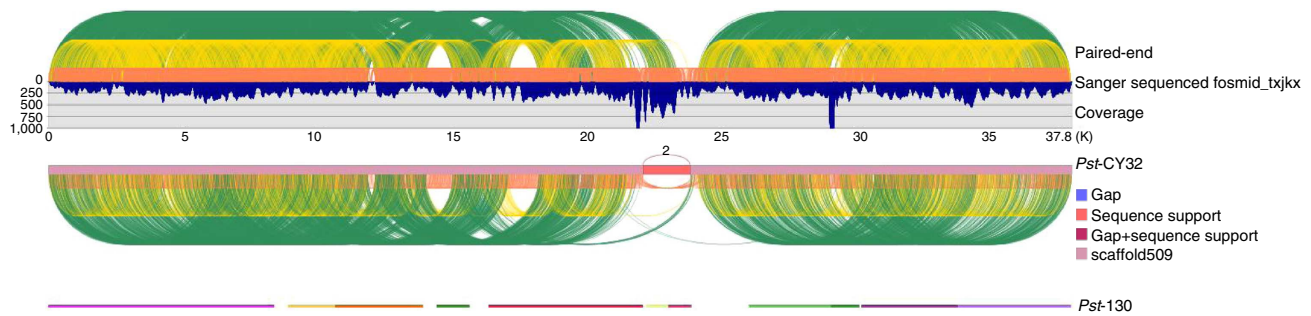
**Table 1 | Features of the *Pst*-CY32 dikaryotic genome draft.**

Estimated genome size	110 Mb
Total number of base pairs within assembled scaffolds	130,657,047
N50 length in bp; total number > 2 kb in length	125,324; 3,596
N90 length in bp; total number > N90 length	14,360; 1,560
Total number of assembled scaffolds	4,283
Total number of base pairs within assembled contigs	115,537,624
N50 length in base pairs; total number > 2 kb in length	18,007; 9,743
N90 length in base pairs; total number > N90 length	4,099; 6,964
Total number of assembled contigs	12,833
GC content of whole genome (%)	44.77
Repetitive sequences (%)	48.91
Proportion of genome that is coding (exonic; including introns) (%)	33.59
Number of putative coding genes	25,288
Gene size (mean base pairs; contain introns)	1,735.43
Average number of exon per gene	4.35
Average exon length (bp)	297.83
Average intron length (bp)	101.41
GC content in coding regions (%)	47.65
Number of transfer RNAs	864

N50: 50% of all nucleotides in the assembly within scaffolds/contigs of  $\geq 125$  kb/18 kb.  
N90: 90% of all nucleotides in the assembly are within scaffolds/contigs of  $\geq 14$  kb/4 kb.



**Figure 2 | The K-mer distribution and contig frequency analysis.** (a) Seventeen k-mer depth distribution of whole-genome Illumina reads. Two peaks at 16 and 30 were identified. (b) Two peaks of Cvg 7 and 17 were identified in the distribution of contig length and depth. These results indicated that the two haploid nuclei of *Pst* urediniospore are highly heterozygous. The haploid genome size can be estimated using a modified Lander–Waterman algorithm based on K-mer analysis (Supplementary Notes 1 and 2). K-mer, a unique sequence of *k* nucleotides long; Cvg, coverage.



**Figure 3 | Evaluation of assembly integrity.** Scaffold 475 of the *Pst*-CY32 assembly has only one PE supported gap, and it matches to one region of fosmid clone\_txjxk that was fully sequenced by Sanger sequencing. The same region consists of 11 contigs in the *Pst*-130 assembly generated solely by WGS Illumina sequencing. Therefore, the fosmid-to-fosmid strategy greatly improved assembly integrity. High PE (paired end) support density indicates high assembly confidence. PE supported gaps are regions that may have SVs derived from genome heterozygosity. Green, yellow and orange curves show paired-end supports from sequences of 6-, 2- and <1-kb insert libraries, respectively.

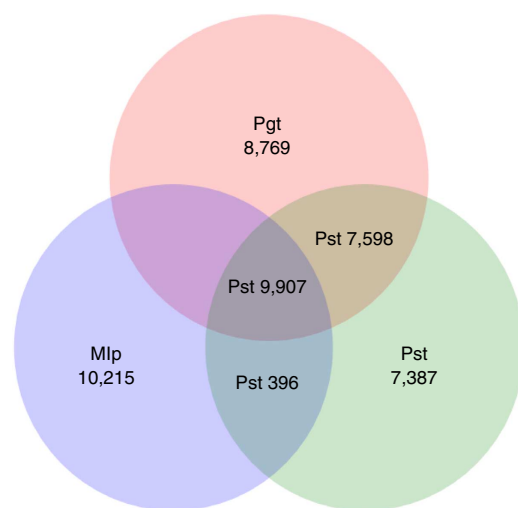
one PE sequence-supported gap in the region that matched fosmid\_txjxk and 11 contigs of the *Pst*-130 assembly (Fig. 3). The same phenomenon was observed in other Sanger-sequenced fosmids. On average, the ratio of *Pst*-130 contigs to matching *Pst*-CY32 scaffolds is 7.2 (Supplementary Fig. S4).

A total of 1,861 expressed sequence tags (ESTs) from CY32 complementary DNA libraries constructed with RNA isolated from urediniospores, germ tubes and haustoria also were used to validate the assembly coverage. For ESTs longer than 500 bp, the coverage by the genome assembly reached 98.92% (Supplementary Table S9). Reciprocal blasting analysis of *Pst*-coding sequences against those of *Pgt* and *Mlp* revealed that over 70% of the predicted *Pst* genes (17,901 genes) have support in the other two rust genomes (Fig. 4). These results reflect a relatively high confidence for the assembly.

Analysis of the ultra-conserved core eukaryotic genes (CEGs) by CEGMA<sup>21</sup> indicates that the *Pst*-CY32 genome assembly has 215 complete and 7 partial models of 248 CEGs (~94%), indicating a high degree of completeness of the genome draft (Supplementary Table S10).

**Comparative analysis with the *Pst*-130 genome sequence.** In comparison with the *Pst*-130 genome (64.8 Mb) assembled solely with WGS Illumina reads<sup>14</sup>, the *Pst*-CY32 genome generated by the ‘fosmid-to-fosmid’ method increased the N50 value from 5 kb to 18 kb for contigs and from 5 kb to 125 kb for scaffolds. The estimated genome size of *Pst* was increased to 110 Mb. More than 5.3 Mb of scaffolds that contain 1,519 predicted genes of *Pst*-CY32 were absent in the published *Pst*-130 genome (*e*-value  $1e^{-10}$ ).

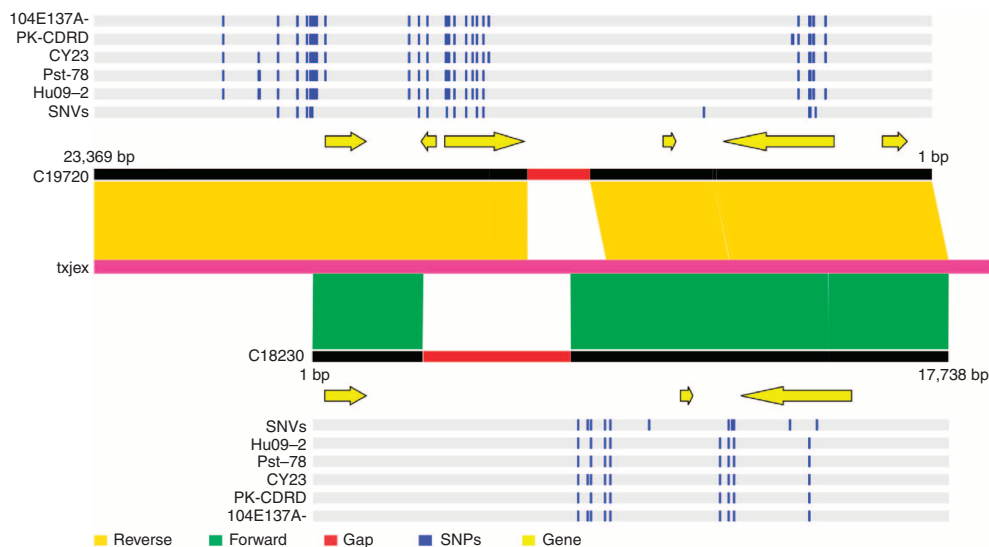
**Heterozygous features in the genome of dikaryotic *Pst*.** High coverage and deep reads ( $26\times$ ) allowed the identification of 83,736 heterozygous single-nucleotide variants (SNVs) within the sequenced *Pst*-CY32 genome (Supplementary Data 1). The heterozygosity rate ( $0.68 \times 10^{-3}$ ) was ~1.6 times higher than that of human ( $0.43 \times 10^{-3}$ )<sup>22,23</sup>. Therefore, the two nuclei in urediniospores of *Pst*-CY32 must be highly heterozygous. In comparison with the estimated genome size, the current assembly is significantly larger and possesses more than 20-Mb assembly ‘redundancy’. Systematic analyses of sequences of highly similar scaffolds revealed that a set of 1,059 genes was present in the overlapping regions of scaffolds with different kinds of overlapping relationships (Supplementary Data 2). The heterozygous regions and distinctive SNV genotypes of the two nuclei of *Pst*-CY32 were revealed by aligning sequences of assembly



**Figure 4 | Reciprocal Blast analysis of *Pst* coding sequences against *Pgt* and *Mlp*.** Venn diagram showing reciprocal Blast search results for orthologous coding sequences of *Pst*, *Pgt* and *Mlp* (*e*-value of  $1e^{-7}$ ). For genes with multiple alternative transcripts, the transcript with the best alignment was selected.

scaffolds with matching Sanger-sequenced fosmids. Some of the predicted genes were located in the heterozygous regions of the dikaryotic *Pst* genome, implying that these ‘redundancy’ genes may result from heterozygous features in the dikaryotic *Pst* genome (Fig. 5). The ‘fosmid-to-fosmid’ strategy allowed us to independently sequence and assemble heterozygous regions that were represented in different fosmid pools, which is an advantage for investigating highly heterozygous genomic regions of dikaryotic rust fungi.

**Annotation and comparison of three rust fungal genomes.** By combining automated gene calling and EST-based prediction, a total of 25,288 genes were predicted to encode peptides > 50 amino acids. Whereas 1,059 of them are likely ‘redundancy genes’ related to the heterozygous nature and assembly method (Supplementary Data 2), 403 genes are truncated genes (Supplementary Data 3). The coding region, including introns, accounts for 33.59% of the *Pst* genome assembly (Table 1). To assess the quality of gene prediction, we compared the length distribution of genes, CDS, exons and introns, and the



**Figure 5 | Heterozygous regions in the *Pst* genome.** Heterozygous regions were identified by aligning the sequence of Txjex, a fosmid fully sequenced by Sanger sequencing with c19720 and c18230, two highly similar scaffolds of the *Pst*-CY32 assembly generated with Illumina sequences from fosmid pools 7PB\_Index52\_Index26/20PB\_Index47\_Index8 and 14PB\_Index44\_Index29\_Index6\_Index35, respectively. SNPs between the reference *Pst*-CY32 genome and re-sequenced strains Hu09-2, *Pst*-78, CY23, PK-CDRD or 104E137A- are marked with blue vertical lines. SNV, heterozygous SNV in the dikaryotic *Pst*-CY32 genome.

distribution of exon number per *Pst* gene with those of related plant pathogenic fungi, including two other rust and common corn smut<sup>12,24</sup>. *Pst* was found to be similar to other rust fungi in all of the major parameters analysed (Supplementary Fig. S5).

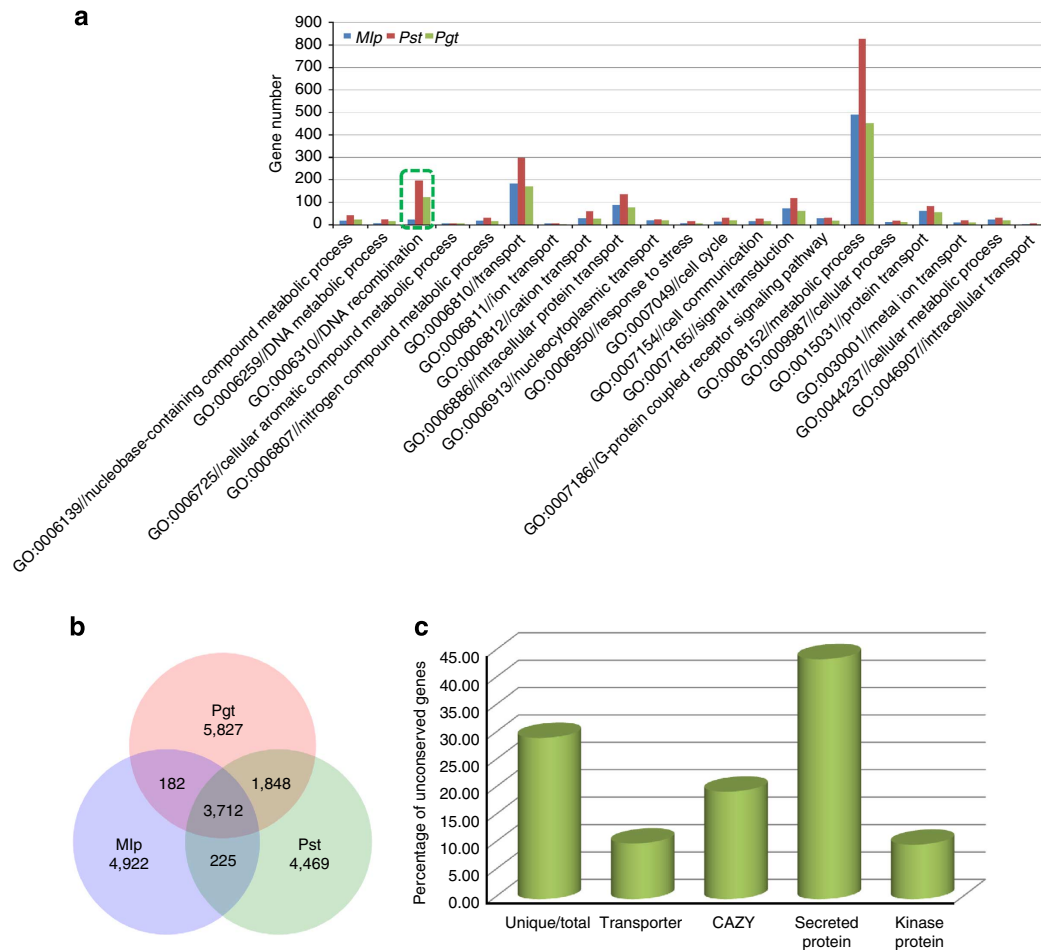
Among all of the predicted *Pst* genes, only 8,334 (32.96%), 2,067 (8.17%) and 2,018 (7.98%) of them had homologues with known functions in the GO, PAMGO, and PHI databases, respectively (Supplementary Figs S6–S8). Among the genes belonging to different PAMGO categories in three rust fungi, the two wheat rust fungi contained more genes involved in DNA recombination, but the gene contents in other categories were similar to those in *Mlp* (Fig. 6a). We used Hcluster\_sg<sup>25</sup> to identify orthologues among three rust fungi. A total of 5,560 orthologous clusters were identified in two wheat rust fungi, but only 3,937 clusters were common to *Pst* and *Mlp*. The numbers of species-specific gene clusters were 4,469, 5,827 and 4,922 for *Pst*, *Pgt* and *Mlp*, respectively (Fig. 6b). These results are consistent with the closer relationship of *Pgt* and *Pst* than to *Mlp*. Among the genes unique to *Pst* in comparison with *Pgt* and *Mlp*, only 14.43% (1,066 genes) and 3.38% (250 genes) of them have homologues with known functions in the GO and PAMGO databases, reflecting inadequate investigation of gene functions in rust pathogens. Interestingly, most of known homologues belong to metabolic pathways and transport in both databases. When they were categorized into transporters, secreted proteins (SPs), kinases and carbohydrate-active enzymes (CAZY), the proportion of SP genes is higher than the others (Fig. 6c), consistent with observations in *Mlp* and *Pgt*<sup>12</sup> and suggesting the importance of SPs in virulence of rust pathogens. This phenomenon also was observed in other obligate plant pathogenic fungi, such as the powdery mildew fungus<sup>13</sup>.

**Secretome.** A total of 2,092 proteins encoded by the *Pst* genome were predicted by SignalP, TargetP and TMHMM<sup>26–28</sup>, which accounts for 8.3% of the total number of proteins, and were grouped into 916 families. Similarly, the genomes of *Pgt* and *Mlp* are predicted to encode 1,459 and 1,178 SPs, respectively, accounting for 8.2% and 7.1% of the total number of *Pgt* and

*Mlp* proteins. Among the putative SPs of three rust pathogens, *Pst* shared 341 families with the other two rust pathogens, and more than 62% of the predicted SP families were lineage-specific for *Pst* (Supplementary Fig. S9). Furthermore, over 92% of them had no distinct orthologues in non-rust fungi, including *Ustilago maydis*, *Blumeria graminis*, *Magnaporthe oryzae* and *Fusarium graminearum*<sup>13,24,29,30</sup>. In cluster analyses, only a few predicted SPs could be grouped into small families, suggesting a substantial diversity of the *Pst* secretome and the lack of sequence similarity or conservation in fungal avirulence or effector genes (Supplementary Fig. S10).

A total of 491 SPs were distributed in 100 putative physical clusters of 2–18 members (Supplementary Data 4). Twenty-seven putative SPs in *Pst* are homologous to seven haustorially expressed SPs (HESPs) of *M. lini*<sup>31</sup>, including Hesp-897, Hesp-735, Hesp-379, Hesp-379-like and Hesp-C49. Eighteen SPs in *Pst* are homologous to the rust transferred protein 1 (RTP1)<sup>32</sup> in other rust pathogens (Supplementary Data 5). A search for known candidate effector motifs indicated that 1,052 of the *Pst* SP genes had the 'Y/F/WxC' motif of the powdery mildew fungus<sup>33</sup> and 114 of them had the RXLR motif reported in oomycete effectors<sup>34</sup> (Supplementary Data 6). However, these two motifs occurred at a similar frequency in non-SPs (11,267 and 3,742 of them with the 'Y/F/WxC' and 'RXLR' motifs, respectively). Therefore, whether *Pst* effectors use these sequence elements for entry into plant cells is questionable.

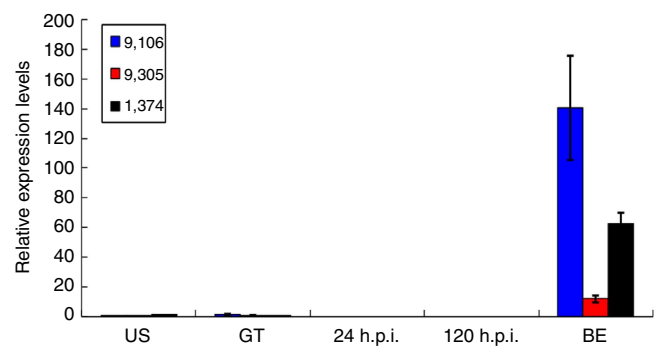
The genomes of the obligate biotroph *B. graminis*, facultative biotroph *U. maydis*, necrotrophic pathogen *F. graminearum* and symbiont *Laccaria bicolor*<sup>35</sup> were found to have 442, 426, 1,049, and 1,161 SP genes, respectively, by the same scanning methodology. They account for 7.5%, 6.2%, 7.6% and 5.2% of total protein-coding genes in *B. graminis*, *U. maydis*, *F. graminearum* and *L. bicolor*, respectively. Therefore, *Pst* and *Pgt* appear to have a relatively larger secretome (higher percentage) compared with that of other plant-associated fungi. Several putative *Pst* effector genes were found to be specifically expressed during sexual reproduction (Fig. 7, Supplementary Data 7 and 8), suggesting that rust fungi with alternate hosts may transmit some specific SPs to cope with infection of the



**Figure 6 | Functional categories and comparative analysis of predicted *Pst* genes.** (a) Functional categories and distribution of predicted *Pst* genes based on the PAMGO database. Among the 2,067 *Pst* genes categorized with the PAMGO database, major groups with 5 genes or more are presented. The most abundant functional groups belong to metabolic process, transport and DNA recombination. (b) Venn diagram showing orthologous gene clusters among the *Pst*, *Pgt* and *Mlp* genomes. For genes with multiple alternative transcripts, the transcript with the best alignment was selected. Orthologous gene families were merged into multiple species orthologous groups by MultiParanoid. (c) The result showed the percentage of *Pst*-specific genes in the total number of genes belonging to each gene category.

primary and alternate host plants. In comparison with *B. graminis*, *U. maydis* and *L. bicolor*, several subcategories of cell wall-hydrolyzing enzyme (CWHE) genes, particularly pectin esterase and mannanase genes, appeared to be expanded in *Pst* and *Pgt* (Supplementary Table S11). Previous studies have shown that the cellulose-degrading capacity is similar among pathogens, but the ability to cope with soft tissue components, for example, hemicellulose and pectin, may vary<sup>36</sup>, which is consistent with our observation that expanded CWHEs in rust secretomes are mainly hemicellulases and pectinesterases. Interestingly, cutinase genes also were expanded in *Pst*. Although they may not be involved in wheat infection (germ tubes of *Pst* urediniospores invade wheat through stomata), cutinases may have an important role in *Berberis* infection because germ tubes produced by basidiospores directly penetrate epidermal cells of *Berberis* leaves<sup>37</sup>.

**Transporters.** A total of 512 transporter genes, including 7 sugar and 12 amino acid transporter genes, were present in the *Pst* genome. Four of seven *Pst* genes encoding putative nucleotide sugar transporters were confirmed by qRT-PCR analysis to be highly expressed *in planta*. In contrast, two hexose transporters that are closely related to *U. fabae* HXT1 were constitutively expressed in urediniospores, germ tubes and other infection



**Figure 7 | Assays for three putative *Pst* SP genes specifically upregulated during infection of *Berberis* leaves.** *Pst* genes striiformis\_9106, 9305, and 1374 are putative SP genes that are specifically upregulated at the pycnial stage in the infected *Berberis* leaves. RNA samples were isolated from urediniospores (US), germ tubes (GT), infected wheat leaves at 24 h post inoculation (h.p.i.) or 120 h.p.i. and infected berberis (*Berberis shensiiana* Ahrendt) leaves (BE) at 11 days post inoculation (11 d.p.i.), respectively. The relative expression levels were calculated by the comparative Ct method with the elongation factor-1 gene of *Pst* as the internal standard. Means and s.e. were calculated from three replicates. Error bars represent the s.d. of the mean.

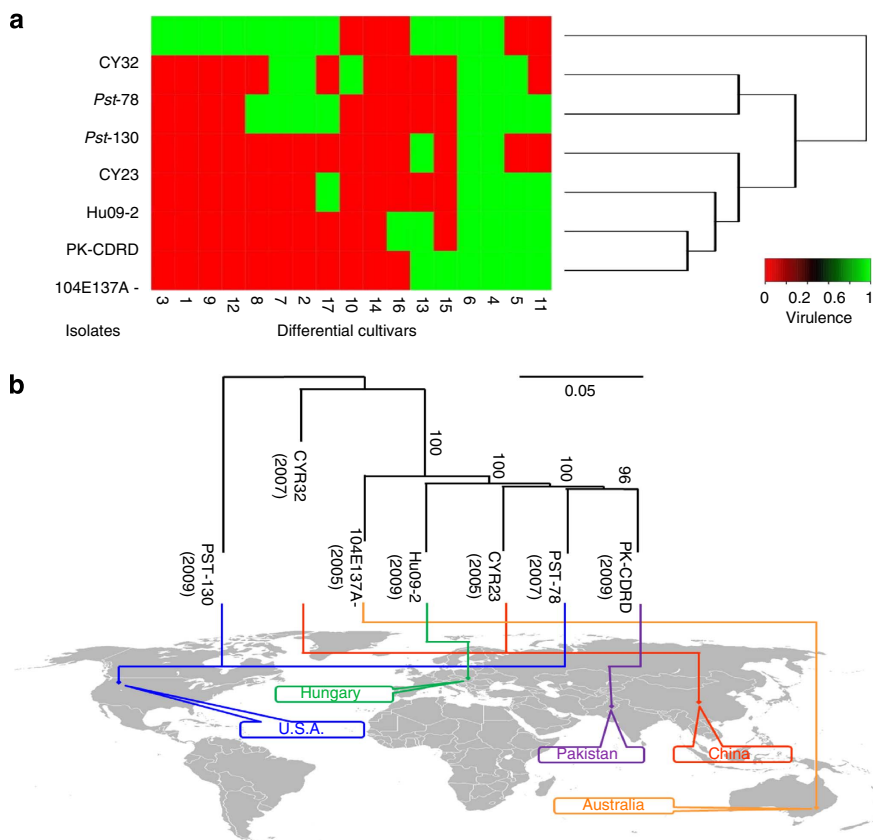
structures (Supplementary Data 7 and 8), which contrasts with the expression profiles of hexose transporter genes in *U. fabae*<sup>38</sup> and *M. larici-populina*<sup>12</sup>. The *Pst* genome also lacks a distinct orthologue of the *U. maydis* *srt1* sucrose transporter gene, which is similar to observations in *Pgt* and *Mlp*<sup>12,39</sup>. Thus, the rust fungi may convert sucrose to hexose by secreted invertase, and haustoria then take up hexose from host cells as the major carbohydrate source<sup>38,40</sup>.

Another notable adaptation was observed in the amino acid-polyamine-organocation superfamily. Whereas the yeast genome contains only nine amino acid transporter genes, the *Pst* genome contains 15, indicating an expansion of the amino acid transporter family that may be related to the defect of *Pst* in nitrogen metabolism. Noticeably, the lack of a nitrite reductase gene in *Pst* will make it incapable of nitrate assimilation, which is consistent with observations in *Pgt* and *Mlp*<sup>12</sup>. Therefore, the uptake of amino acids from host cells by haustoria is essential for protein synthesis in *Pst*. Data from qRT-PCR analysis confirmed that 8 of 12 predicted amino acid transporter genes are highly expressed during *Pst* infection of wheat or *Berberis* plants (Supplementary Data 7 and 8). Six of them are highly similar to PIG27 of *U. fabae*, which is specific for lysine and histidine<sup>41</sup>. Two others share homology with AAT3p and PIG2 of *U. fabae*<sup>42,43</sup>. Their upregulated expression in infected plant tissues supported the hypothesis that amino acid transporters are responsible for the uptake of amino acids from plant cells by haustoria<sup>37</sup>.

**Re-sequencing of five additional *Pst* isolates.** To explore the intra-specific race evolution, five additional *Pst* isolates from different geographical locations were selected for re-sequencing

(Supplementary Table S12). Based on their virulence on 17 differential wheat cultivars (Fig. 8a), *Pst*-CY32 appeared to be more closely related to *Pst*-130 (ref. 44) than to CY23 and other *Pst* isolates. Their genomes were re-sequenced at 22-29X by the WGS strategy to a genome coverage of over 95% (Supplementary Table S13). In comparison with *Pst*-CY32, more than 100,000 SNPs were identified in the five re-sequenced *Pst* isolates and 81,000 in *Pst*-130 (ref. 14). The number of SNPs is similar for individual isolates, and heterozygous SNPs occupied more than 80% of SNPs in the genomes of all six of these *Pst* isolates, which may be related to their heterozygous dikaryotic genomes (Table 2). Based on re-sequencing data, rich SNP distribution in the heterozygous regions of *Pst*-CY32 could be observed, and genotypes of these five isolates are distinctive in these regions, further indicating the heterozygous property of the *Pst* genome (Fig. 5). The number of SNPs in gene-coding regions (cSNPs) accounts for ~35% of total SNPs. Among them, half were non-synonymous mutations that may impact virulence, race specificity and other biological traits and may be responsible for adaptation to host and environmental changes.

Phylogenetic trees of these *Pst* isolates were constructed based on total SNPs and cSNPs. The topological features of the cSNP and SNP trees are similar. The *Pst* isolates collected from different locations exhibited significant genetic variation based on genome-wide cSNPs and could be classified into different branches (Fig. 8b). These results indicate that similar races may possess unexpected levels of genetic variation at the genome level. However, the SNP- or sequence-based phylogenetic tree differed from the virulence-based tree. Therefore, the evolutionary relations of *Pst* isolates based on SNPs were not correlated with their virulence features. For example, isolate PK-CDRD is



**Figure 8 | Relationship among seven *Puccinia striiformis* f. sp. tritici isolates.** (a) Heat map and race relationship of 7 *Pst* isolates according to their virulence on 17 differential cultivars. The susceptible and resistant types were recorded as 1 and 0, respectively. Heatmap.2 was used for clustering. (b) Phylogenetic tree of 7 *Pst* isolates constructed with their cSNPs using Mega5-nj. The origin and isolation date of each isolate were marked.

**Table 2 | Number and type of SNPs in five re-sequenced *Pst* isolates and *Pst*-130.**

Sample Name	Total SNP	Homo		Hetero		Gene region	
		Number	Rate (%)	Number	Rate (%)	CDS	
						Non_Syn	Syn
<i>Pst</i> -CY23	107,491	18,466	17.18	89,025	82.82	19,120	19,403
104E137A-	102,786	16,138	15.70	86,648	84.30	18,494	18,687
PK-CDRD	108,785	18,670	17.16	90,115	82.84	19,529	19,789
<i>Pst</i> -78	106,374	18,086	17.00	88,288	83.00	19,121	19,252
Hu09-2	104,300	16,913	16.22	87,387	83.78	18,890	18,991
<i>Pst</i> -130	81,001	14,434	17.82	66,567	82.18	15,131	13,827

Homo, homological SNP; Hetero, heterozygous SNP; CDS, coding sequence; Non\_Syn, non-synonymous mutation; Syn, synonymous mutation. Reads of *Pst*-130 were downloaded from NCBI.

more similar to 104E137A- in virulence than it is to *Pst*-78 in genome variation. Furthermore, these intra-species isolates cannot be grouped in the phylogenetic tree based on their geographical origins. Isolate *Pst*-78, a prevalent race in the United States, shared most recent ancestors with isolate PK-CDRD from Pakistan. Isolate *Pst*-130 from North America was more closely related to the Chinese isolate CY32 than to other *Pst* isolates. These results indicate that sexual reproduction and genetic recombination may be more important than expected for the origin of these globally dispersed isolates.

We also investigated the distribution of InDels and SVs among five re-sequenced *Pst* isolates (Supplementary Tables S14 and S15). In general, all of them are rich in InDels (1,863 on average) and SVs (759 on average), although they vary in the abundance. These results further indicate that significant genetic variation exists among these *Pst* isolates. Interestingly, isolate CY23, a relatively old Chinese race with a limited host range, has almost twice as many InDels as other *Pst* isolates in comparison with *Pst*-CY32 and contains only deletion-type SVs, suggesting that CY23 may differ from other *Pst* isolates in variations associated with race evolution and virulence. When 12 candidate *Pst* effector genes were selected for sequencing analysis, one of them, Gene\_5778, was found to contain a 3-bp deletion in the CDS region (deletion of AGT at 346–348) in isolate CY23.

## Discussion

The stripe rust fungus is one of the most important pathogens of wheat in the world. Although haploid basidiospores were recently found to infect *Berberis* species<sup>8,9</sup>, *Pst* exists in the dikaryotic phase during all stages of wheat infection (Fig. 1). Unlike diploid organisms that contain pairs of homologous chromosomes, two nuclei in dikaryotic urediniospores formed by plasmogamy between two compatible haploid strains may be distantly related and highly heterozygous because the resulting diploids may be sterile or of low fertility. In comparison with the human genome, the *Pst*-CY32 genome is indeed more heterozygous. The dikaryotic nature of *Pst* made it extremely difficult for genome assembly, particularly with short WGS reads generated by NGS techniques. Among three rust genomes that have been published, the contig N50 for the *Pst*-130 assembly was only 5 kb (ref. 14). For the *Pgt* and *Mlp* genomes, improved assembly was achieved with longer reads from Sanger sequencing, which is much more expensive than NGS. However, over 90 Mb of the Sanger reads could not be assembled into the released 88.6-Mb *Pgt* genome (<http://www.broadinstitute.org>), likely due to genome heterozygosity. The current *de novo* genome assembly of *Pst*-CY32 has

scaffold N50 of 125 kb, which is significantly improved over the assembly based solely on WGS sequencing.

In comparison with non-rust fungal pathogens, *Pst*, *Pgt* and *Mlp* all have larger and more complex secretomes. Some of the expanded families of secreted proteins in *Pst* or other rust fungi may have important roles in fungal–plant interactions. For example, one cutinase family expanded in three rust fungi has 18, 9, and 8 members in *Pst*, *Pgt* and *Mlp*, respectively (Supplementary Fig. S11). These cutinase genes may be important for the infection of leaves of the woody alternate host plant. Among the *Pst*-specific genes with homologues of known functions, the most abundant category, with over 250 members, is related to DNA replication and repair, which is consistent with the report in *Pgt* and *Mlp*<sup>12</sup>. Some of these genes may be related to evolutionary adaptation to environmental stresses, such as strong UV light during airborne dispersal.

During the last century, spontaneous mutations were assumed to be the major source of genetic variation responsible for the emergence of new virulent races<sup>3,45</sup>. Because the alternate host was not known until 2010 (ref. 8), studies of the role of sexual reproduction in genetic variation and race evolution have been neglected in *Pst*. The re-sequencing data generated here show that similar *Pst* races or isolates from different continents possessed an unexpectedly large amount of genetic variation at the genome level based on SNP and InDel analyses. The phylogenetic relationship among these *Pst* isolates is not related to their virulence features on differential cultivars or to their geographical origins. All these data from comparative genome analyses suggest that genetic recombination contributes significantly to genome variations in *Pst* and may be more important than expected for the origin of globally dispersed isolates. Consistent with this inference, high levels of genetic diversity were observed in *Pst* populations from western China and central Asia<sup>7,46</sup>, where susceptible *Berberis* species are widely distributed<sup>9</sup>. Therefore, the role of sexual reproduction in genetic diversity and global epidemics of *Pst* should be considered important and further assessed by sequencing additional *Pst* isolates and performing population-level comparative analyses.

## Methods

**Strains and DNA preparation.** A single-spore isolate of *Pst*-CY32 (ref. 15) was reproduced on seedlings of wheat cultivar Mingxian169. Five additional *Pst* strains isolated from Shaanxi province of China, USA, Pakistan, Hungary, and Australia were assayed for their virulence on 17 international differential wheat cultivars<sup>45,47</sup> and used for re-sequencing. Urediniospores were harvested and used for DNA isolation by the CTAB/SDS method. High molecular weight genomic DNA was isolated with the Genra Puregene Cell Kit (Qiagen) and separated on a 1% agarose gel with a Bio-Rad CHEF-DR II electrophoresis apparatus. The 35–45 kb DNA fraction was isolated from the CHEF gel and used for fosmid library construction<sup>48</sup>.

**Genome sequencing and gene prediction.** The dikaryotic *Pst*-CY32 strain was sequenced by Illumina GA and assembled by a ‘fosmid-to-fosmid’ strategy using SOAP *denovo*<sup>16,17</sup> and Celera Assembler<sup>18</sup> (Supplementary Fig. S1). The protein-coding genes were predicted with a combination of homology-based prediction (GeneWise (wise2-2.0), GeneMark-ES (version 2.3a), Augustus (version 2.5) and EST-based prediction. The resulting predictions were integrated using Glean (glean-1-0-1)<sup>49,50</sup>. The integrated gene collection was searched against the TE database to remove TE-related fragments, and genes encoding peptides shorter than 50 amino acids were filtered out (Supplementary Data 9).

Reciprocal blast analysis of *Pst*, *Pgt* and *Mlp* coding sequences was conducted with BlastP V2.2.23 (e-value 1e–7). Hcluster\_sg V0.5.0 was used for gene family clustering<sup>51,52</sup>. Solar V0.6.9 was used to remove redundant members (Match rate  $\geq 0.33$ )<sup>20</sup>. TEs and tandem repeats were identified by *de novo* and knowledge based (Repeat Update, v. 17.11) strategies for all three rust fungi with REPET<sup>53</sup> (<http://urgi.versailles.inra.fr/index.php/urgi/Tools/REPET>).

**SNV detection.** To assess the heterozygosity rate and its distribution, high-quality reads (average quality score of >20) from WGS sequencing of three short-insert libraries (224-bp inserts, total clean reads of 3,322 Mb) were realigned to the



reference assembly with SOAPaligner (version 2.21)<sup>23</sup>. The probabilities of every possible genotype for each position on the reference genome were calculated, and a statistical model based on Bayesian theory and the Illumina quality system was used to call SNVs. The allelic sequence with the highest probability was used as the reference sequence, and heterozygous SNVs were called if other alleles also had high probability<sup>23,54</sup>.

**Secretome and subcellular targeting prediction.** SPs of *Pst*, *Pgt*, *Mlp*, *Um*, *Fg*, *Lb* and *Bg* were identified by the same methodology<sup>12</sup>. SignalP3.0 was used to predict signal peptides. Tmhmm-2.0c and Targetp-1.1 were used to predict proteins with transmembrane domains and targeting to mitochondria, respectively. GPI (glycosylphosphatidyl inositol)—anchor proteins were identified with big-PI FungalPredictor<sup>26–28</sup>.

**Re-sequencing and genetic variation analysis.** PE (paired ends) libraries with 500-bp inserts for every *Pst* isolate were sequenced using Illumina technology by WGS strategy. High-quality reads were compared with the reference sequence of *Pst*-CY32 by SOAPaligner (SOAP\_snp and SOAP\_InDel) to identify highly confident SNPs and InDels. Phylogenetic trees of *Pst* isolates based on total SNPs and cSNPs were constructed with TreeBeST-1.9.2 (PHYML) and NJ (bootstraps: 100). Virulence clustering of the seven *Pst* isolates was done using Heatmap.2.

## References

- Brown, J. K. M. & Hovmöller, M. S. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537–541 (2002).
- Fisher, M. C. *et al.* Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186–194 (2012).
- Kolmer, J. A. Tracking wheat rust on a continental scale. *Curr. Opin. Plant Biol.* **8**, 441–449 (2005).
- Milus, E. A., Kristensen, K. & Hovmöller, M. S. Evidence for increased aggressiveness in a recent widespread strain of *Puccinia striiformis* f. sp. *tritici* causing stripe rust of wheat. *Phytopathology* **99**, 6 (2009).
- Hovmöller, M. S., Walter, S. & Justesen, A. F. Escalating threat of wheat rusts. *Science* **329**, 369 (2010).
- Shan, W., Chen, S., Kang, Z., Wu, L. & Li, Z. Genetic diversity in *Puccinia striiformis* Westend. f.sp. *tritici* revealed by pathogen genome-specific repetitive sequence. *Can. J. Bot.* **76**, 587–595 (1998).
- Hovmöller, M. S., Sørensen, C. K., Walter, S. & Justesen, A. F. Diversity of *Puccinia striiformis* on cereals and grasses. *Annu. Rev. Phytopathol.* **49**, 197–217 (2011).
- Jin, Y., Szabo, L. J. & Carson, M. Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* as an alternate host. *Phytopathology* **100**, 432–435 (2010).
- Zhao, J., Zhang, H.-C., Yao, J.-N., Huang, L.-L. & Kang, Z.-S. Confirmation of *Berberis* spp. as alternate hosts of *Puccinia striiformis* f. sp. *tritici* on wheat in China. *Mycosystema* **30**, 895–900 (2011).
- Webb, C. A. & Fellers, J. P. Cereal rust fungi genomics and the pursuit of virulence and avirulence factors. *FEMS Microbiol. Lett.* **264**, 1–7 (2006).
- Line, R. F. Stripe rust of wheat and barley in North America: a retrospective historical review 1. *Annu. Rev. Phytopathol.* **40**, 75–118 (2002).
- Duplessis, S. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl Acad. Sci.* **108**, 9166–9171 (2011).
- Spanu, P. D. *et al.* Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**, 1543–1546 (2010).
- Cantu, D. *et al.* Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS One* **6**, e24230 (2011).
- Wan, A., Liren, W., Jin, S. & Wang, B. Wang Baotong Discovery and studies on CY32, a new race of *Puccinia striiformis* f. sp. *tritici* in China. *Acta Phytopylacica Sinica* **30**, 6 (2003).
- Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
- Li, R. *et al.* *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272.
- Denisov, G. *et al.* Consensus generation and variant detection by Celera Assembler. *Bioinformatics* **24**, 1035–1040 (2008).
- Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Pennsylvania State Univ (2007).
- Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**, 311–317 (2010).
- Parra, G., Bradnam, K. & Korff, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
- Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65 (2008).
- Li, R. *et al.* SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **19**, 1124–1132 (2009).
- Kamper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101 (2006).
- Alexeyenko, A., Tamas, I., Liu, G. & Sonnhammer, E. L. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* **22**, e9–e15 (2006).
- Dyrlov, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
- Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).
- Saunders, D. G. O. *et al.* Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One* **7**, e29847 (2012).
- Dean, R. A. *et al.* The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).
- Cuomo, C. A. *et al.* The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**, 1400–1402 (2007).
- Catanzariti, A. M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A. & Ellis, G. J. Haustorially-expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243–256 (2006).
- Kemen, E. *et al.* Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant-Microbe Interact* **18**, 1130–1139 (2005).
- Godfrey, D. *et al.* Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics* **11**, 317 (2010).
- Morgan, W. & Kamoun, S. RXLR effectors of plant pathogenic oomycetes. *Curr. Opin. Microbiol.* **10**, 332–338 (2007).
- Martin, F. *et al.* The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88–92 (2008).
- Amselem, J. *et al.* Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**, e1002230 (2011).
- Kemen, E. & Jones, J. D. G. Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends Plant Sci.* **17**, 448–457 (2012).
- Voegelé, R. T., Struck, C., Hahn, M. & Mendgen, K. The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proc. Natl Acad. Sci.* **98**, 8133–8138 (2001).
- Wahl, R., Wippel, K., Goos, S., Kämper, J. & Sauer, N. A novel high-affinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis*. *PLoS Biol.* **8**, e1000303 (2010).
- Spanu, P. D. The genomics of obligate (and nonobligate) biotrophs. *Annu. Rev. Phytopathol.* **50**, 91–109 (2012).
- Struck, C., Ernst, M. & Hahn, M. Characterization of a developmentally regulated amino acid transporter (AAT1p) of the rust fungus *Uromyces fabae*. *Mol. Plant Pathol.* **3**, 23–30 (2002).
- Struck, C., Mueller, E., Martin, H. & Lohaus, G. The *Uromyces fabae* UfAAT3 gene encodes a general amino acid permease that prefers uptake of in planta scarce amino acids. *Mol. Plant Pathol.* **5**, 183–189 (2004).
- Hahn, M., Neef, U., Struck, C., Göttfert, M. & Mendgen, K. A putative amino acid transporter is specifically expressed in haustoria of the rust fungus *Uromyces fabae*. *Mol. Plant Microbe Interact.* **10**, 438–445 (1997).
- Chen, X., Penman, L., Wan, A. & Cheng, P. Virulence races of *Puccinia striiformis* f. sp. *tritici* in 2006 and 2007 and development of wheat stripe rust and distributions, dynamics, and evolutionary relationships of races from 2000 to 2007 in the United States. *Can. J. Plant Pathol.* **32**, 315–333 (2010).
- Wellings, C. R. *Puccinia striiformis* in Australia: a review of the incursion, evolution, and adaptation of stripe rust in the period 1979–2006. *Aust. J. Agric. Res.* **58**, 567–575 (2007).
- Ali, S., Leconte, M., Walker, A.-S., Enjalbert, J. & de Vallavieille-Pope, C. Reduction in the sex ability of worldwide clonal populations of *Puccinia striiformis* f.sp. *tritici*. *Fungal Genet. Biol.* **47**, 828–838 (2010).
- Chen, X. M. & Line, R. F. Inheritance of stripe rust resistance in wheat cultivars used to differentiate races of *Puccinia striiformis* in North America. *Phytopathology* **82**, 633–637 (1992).
- Kim, U.-J., Shizuya, H., de Jong, P. J., Birren, B. & Simon, M. I. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20**, 1083–1085 (1992).
- Varshney, R. K. *et al.* Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotech.* **30**, 83–89 (2012).
- Elsik, C. G. *et al.* Creating a honey bee consensus gene set. *Genome Biol.* **8**, R13 (2007).
- Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* **34**, D572–D580.
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Flutre, T., Duprat, E., Feuillet, C. & Quesneville, H. Considering transposable element diversification in *de novo* annotation approaches. *PLoS One* **6**, e16526 (2011).
- Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949 (2012).

## Acknowledgements

This study was supported financially by the National Basic Research Program of China (No. 2013CB127700, No. 2012CB114001), the earmarked fund for Modern Agro-industry Technology Research System (No. CARS-3-1-11), the Nature Science Foundation of China (No. 30930064) and the 111 Project from the Ministry of Education of China (B07049). We thank L. Dunkle for proofreading this manuscript.

## Authors contributions

W.Z., L.H., J.H. and X.W. contributed equally to this work. Z.K., W.Z., L.H., J.-R.X. and G. Zhang managed the project. Z.K., W.Z., L.H., J.-R.X., G. Zhang, X.C., H.Y. designed the experiments and led the data analysis. H.Z., W.Z., Jie Zhao, J.G., G. Zhan, Q.H., X.G. and X.C. prepared the DNA samples. J.H., C.Q., P.N., L.Y., W.D. and W.M.Z. performed sequencing and assembly. J.H. and C.Q. performed genome annotation and repeat analysis. X.W., J.G., J.L., H.L., X.H., G.P., L.H., C.Q., J.H. and W.Z. performed comparative genomics. X.W., J.L., J.G., X.H., Jie Zhao, C.T., Y.C., M.L., J. Zhang, Z.Z., S.Z., D.H., J.W., H.Z. and Jing Zhao. performed the virulence identification and gene expression experiments. Z.K., W.Z., J.-R.X., G. Zhang, L.H., X.W., J.L., J.G., X.H. and J.Z. wrote the paper.

## Additional information

**Accession codes:** Sequence data have been deposited in GenBank nucleotide core database under the accession number AONS00000000.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Zheng, W. *et al.* High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat. Commun.* 4:2673 doi: 10.1038/ncomms3673 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>