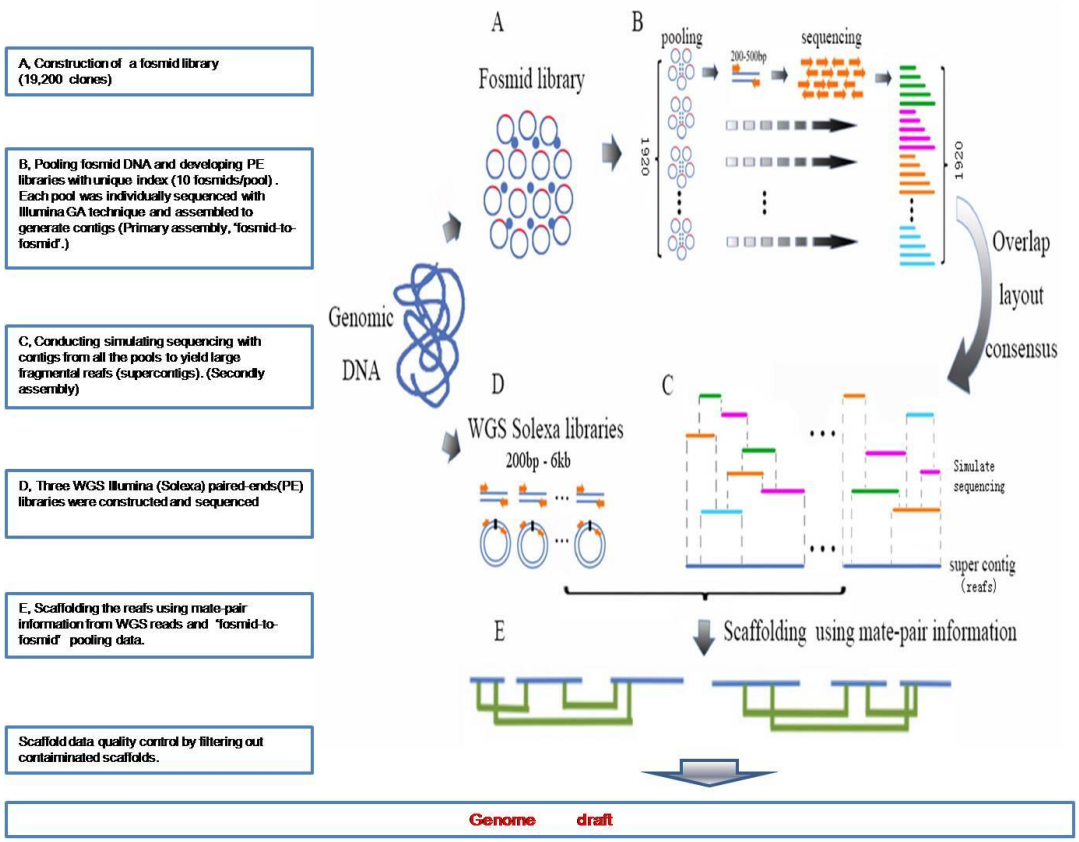
Supplementary Figures



Supplementary Fig. S1 | Diagram of *Pst* genome sequencing and assembly using a 'fosmid to fosmid' strategy.

Fosmid pooling and sequencing: Fosmid librarywas constructed according to Kim *et al* and randomly divided to clone pools (10 fosmids/pool). In total, 1,920 pools containing 19,200 fosmid clones were created for Illumina sequencing. For each pool with unique index, one 170-bp and one 500-bp insert libraries were constructed. Libraries were then bar-coded and paired-end sequenced in a lane on the Illumina sequencing platform.

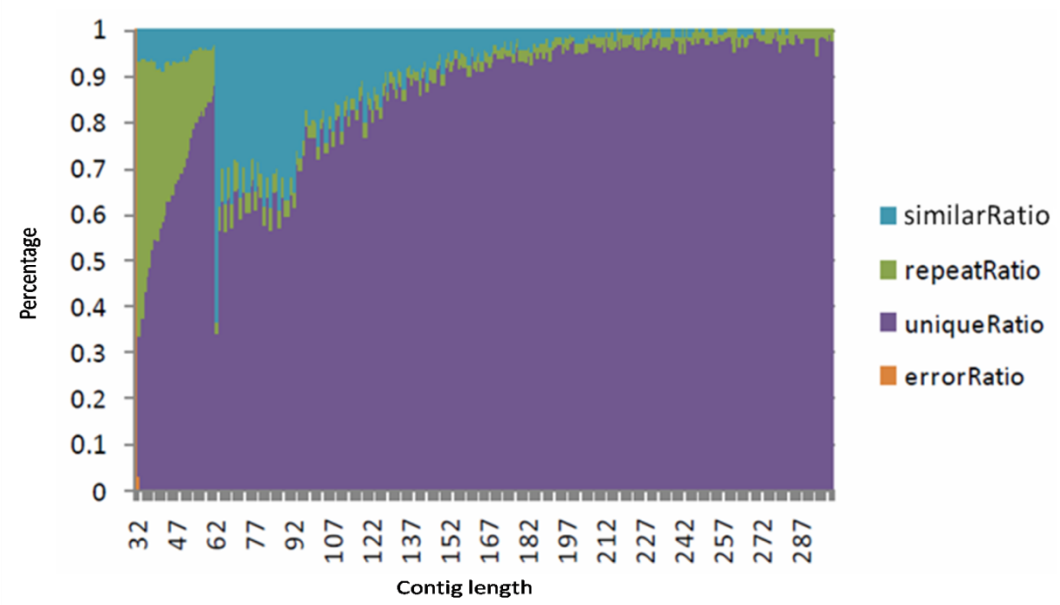
The primary assembly for each pool independently: First, the reads in a pool were assembled using SOAPdenovo into contigs and scaffolds via various combination of parameters with *k*-mer set from 23 to 63. As a result, a contig N50 of 4.7-kb and scaffold N50 of 21.4-kb were achieved after gap filling by combining assemblies from all the 1,920 pools.

Secondary assembly to create large reafs (super-contigs): The resulted contigs (total length of 728.14-Mb) with a N50 size of 4.7-kb were then taken as 'simulate reads' for secondary assembly (simulating sequencing) to generate large reafs using Celera assembler, which is an Overlap Layout Consensus (OLC) assembler.

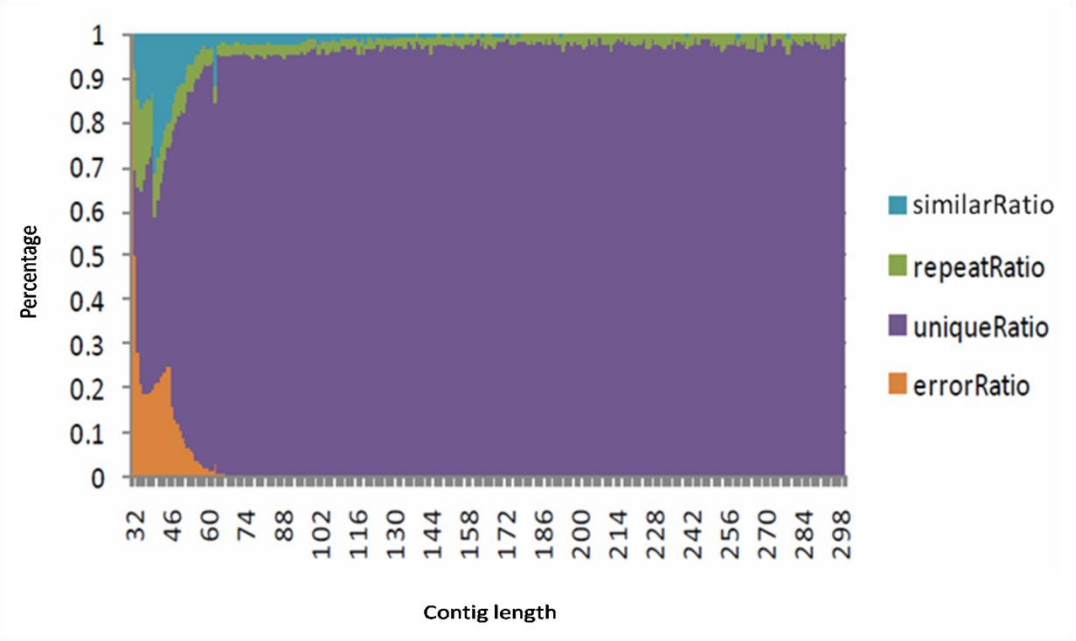
Scaffolding of all reafs: The virtual paired-end reads artificially extracted from assembled scaffolds of each pool with insert size from 50-bp to 37.5-kb, and the paired-end reads of Illumina sequencing for WGS PE libraries (200-bp to 6-kb inserts) were used for scaffolding of all reafs by SOAPdenovo_scaffolder.

The assembled scaffolds were searched against nt database to remove contaminations from plant, bacteria, and insects. The final combined length of assembled scaffolds was 130.7-Mb. The N50 of scaffolds and contigs reached 125-kb and 18-kb, respectively.

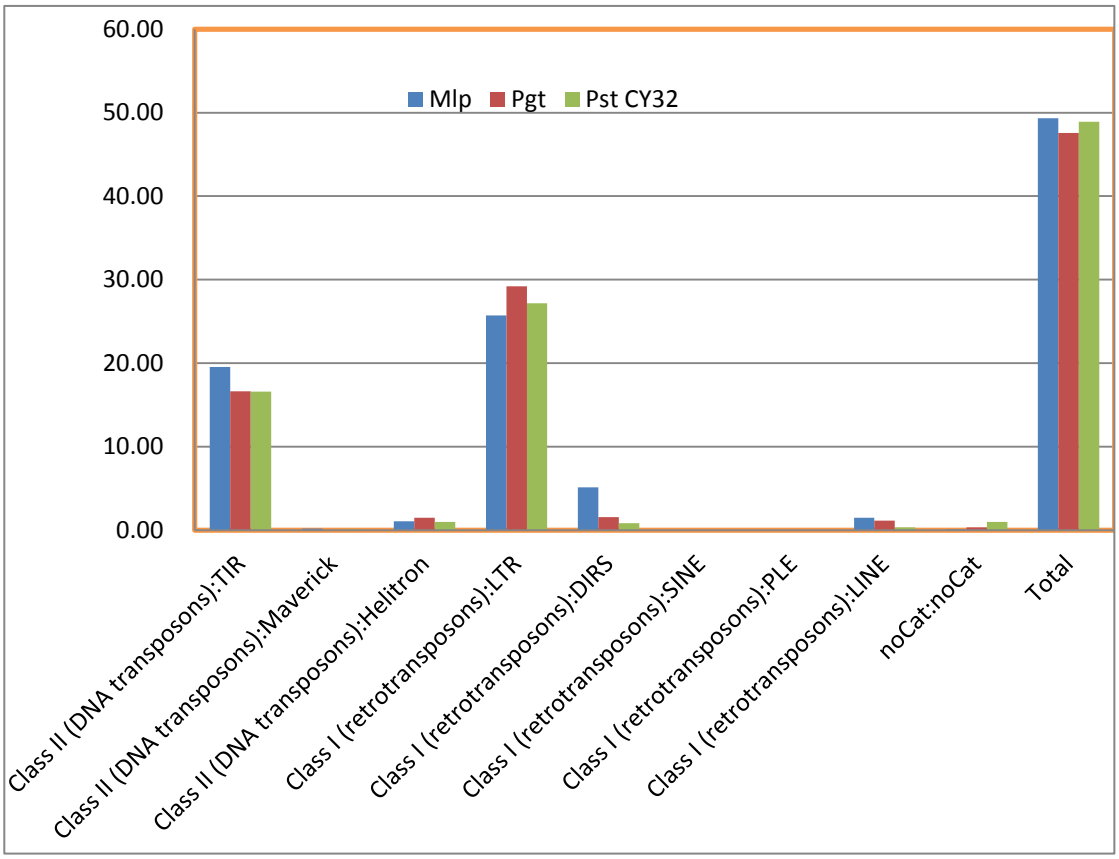
a)



b)

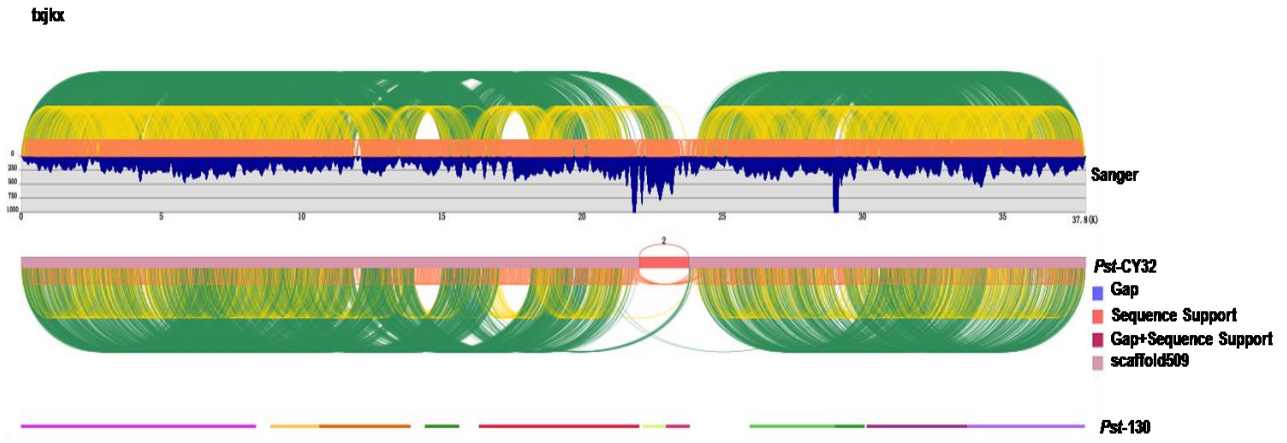


Supplementary Fig. S2 | Similar contig analysis of the *Pst* genome and human genome. By comparing the results of contig analysis for the *Pst* (a) and human (b) genomes, a manifest trend can be observed that the ratio of similar contigs in the *Pst* genome is significantly higher than that in the human genome, indicating a higher heterozygosity in the *Pst* genome. For estimating the ratio of different contig types with different lengths, the AvgCvg of depth of <0.1 fold and >1.8 folds was considered as error and repeats, respectively. Comparison of same length contigs was done for all contigs with depth of 1.8 folds ≧ AvgCvg ≧0.1 folds. Contigs with similarity of ≧0.95 were selected as similar contigs, the remaining ones were selected as unique contigs.

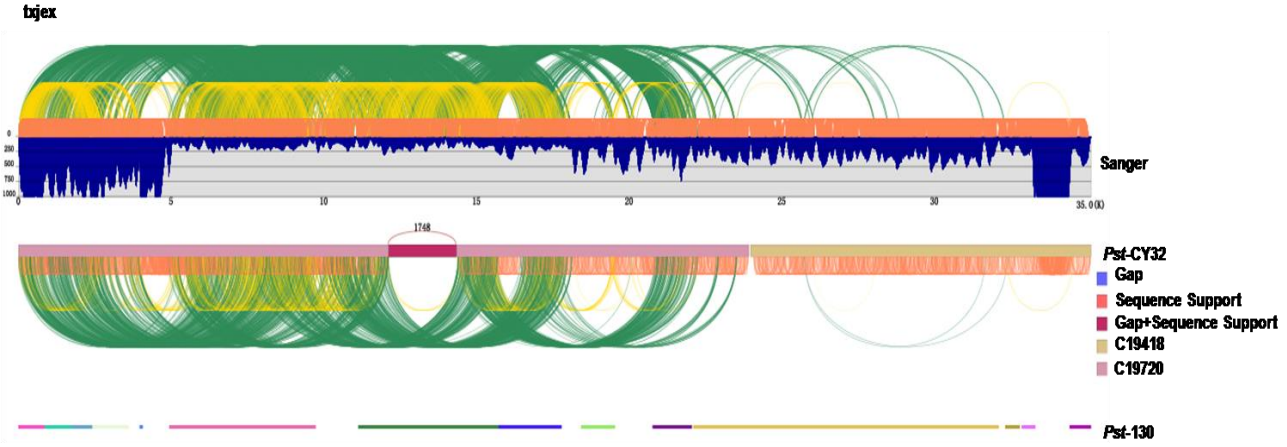


Supplementary Fig. S3 | Distribution of repetitive sequence elements in three rust fungal genomes. The result shows similar distribution patterns for the percentages of repetitive sequences in the genomes of *Puccinia striiformis* f. sp. *tritici* (*Pst*), *Puccinia graminis* (*Pgt*) and *Melampsora larici-populina* (*Mlp*), and all of them consist of high proportion for repetitive sequences. The repetitive sequences were identified by knowledge-based or de novo prediction with the REPET pipeline**.**

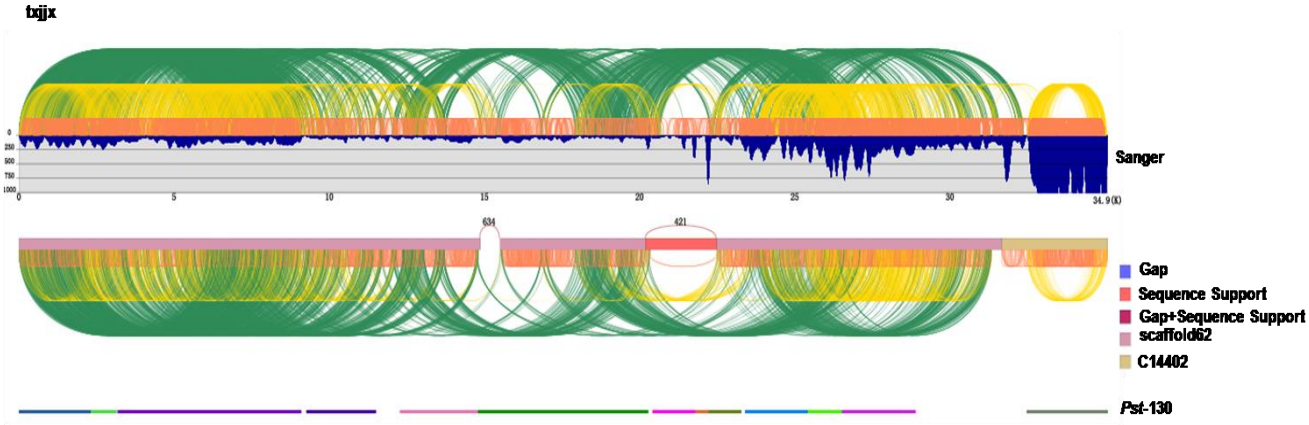
a)


txjkx
Sanger
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
scaffold509
Pst-130

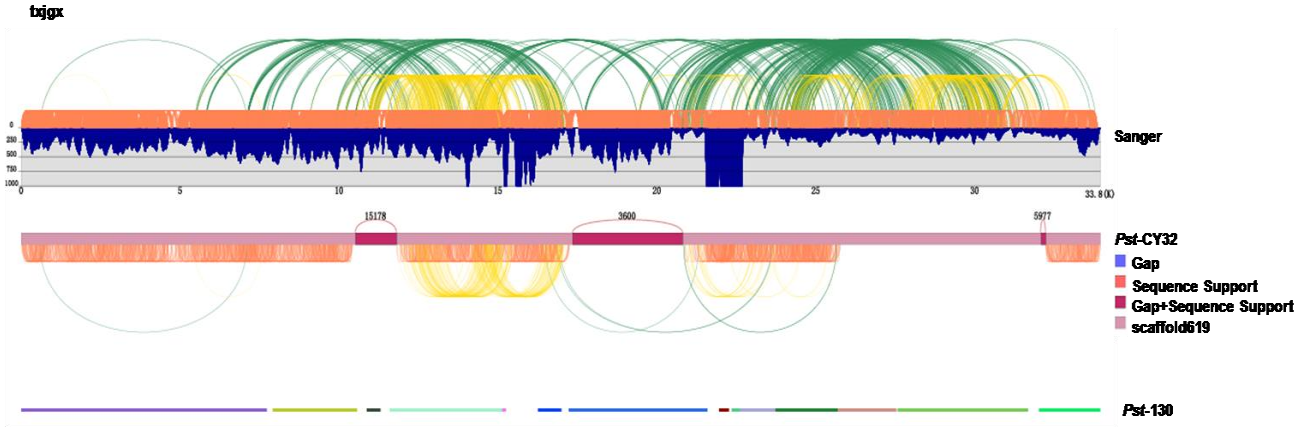
b)


txjex
Sanger
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
C19418
C19720
Pst-130

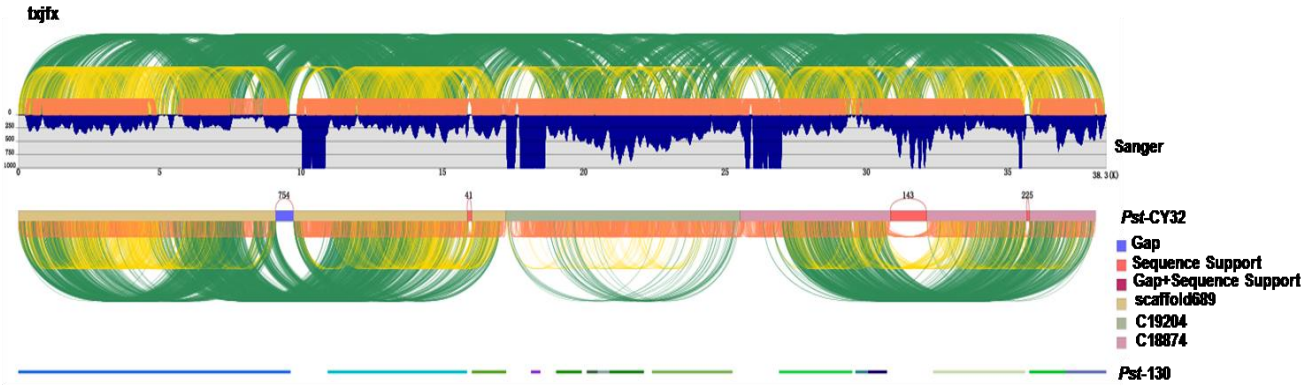
c)


txjjx
Sanger
Gap
Sequence Support
Gap+Sequence Support
scaffold62
C14402
Pst-130

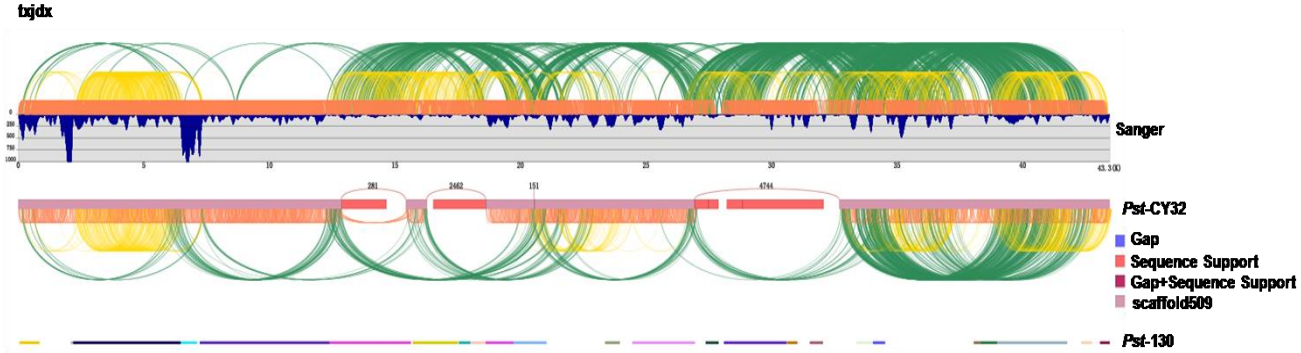
d)


txjgx
Sanger
15178
3600
5977
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
scaffold619
Pst-130

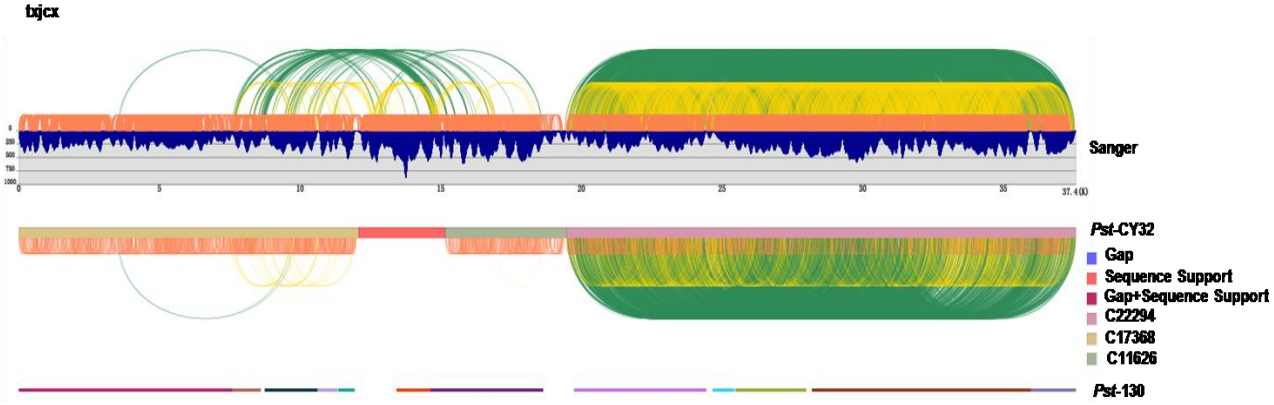
e)


txjfx
Sanger
754
41
143
225
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
scaffold689
C19204
C18874
Pst-130

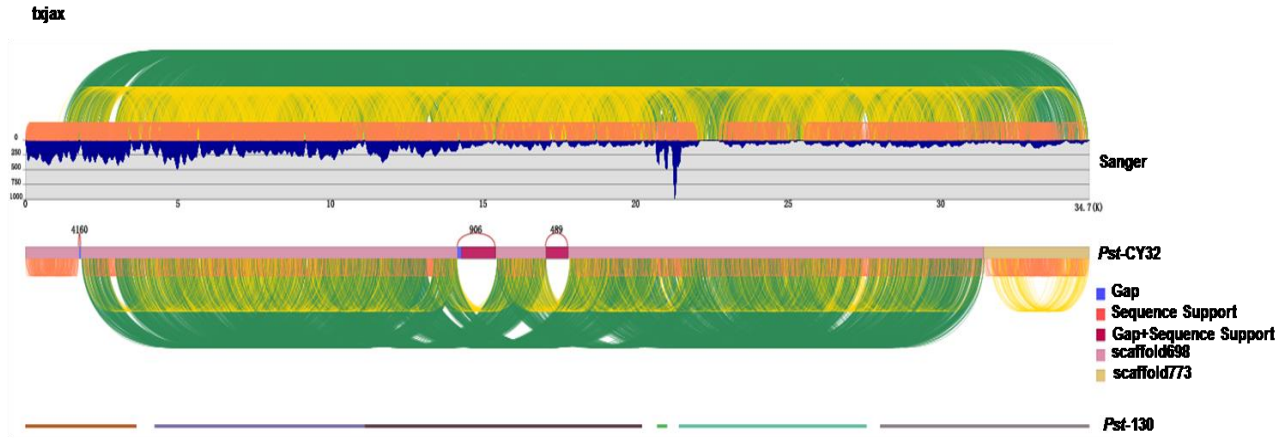
f)


txjdx
Sanger
281
2462
151
4744
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
scaffold509
Pst-130

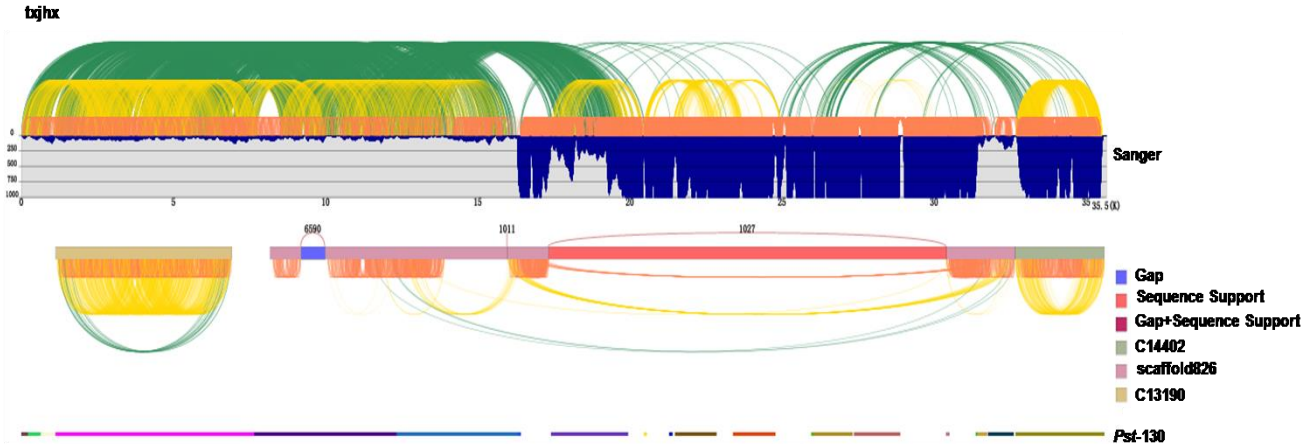
g)

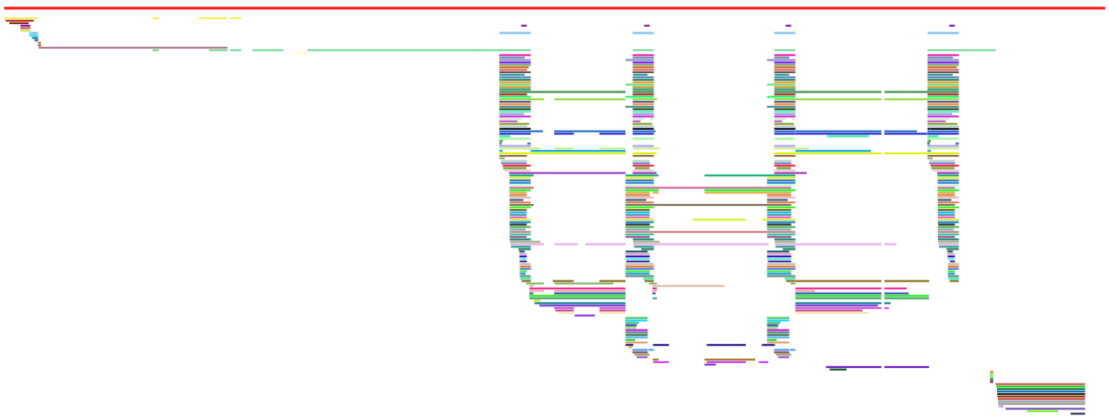

txjcx
Sanger
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
C22294
C17368
C11626
Pst-130

h)


txjax
Sanger
4160
906
489
Pst-CY32
Gap
Sequence Support
Gap+Sequence Support
scaffold698
scaffold773
Pst-130

i)



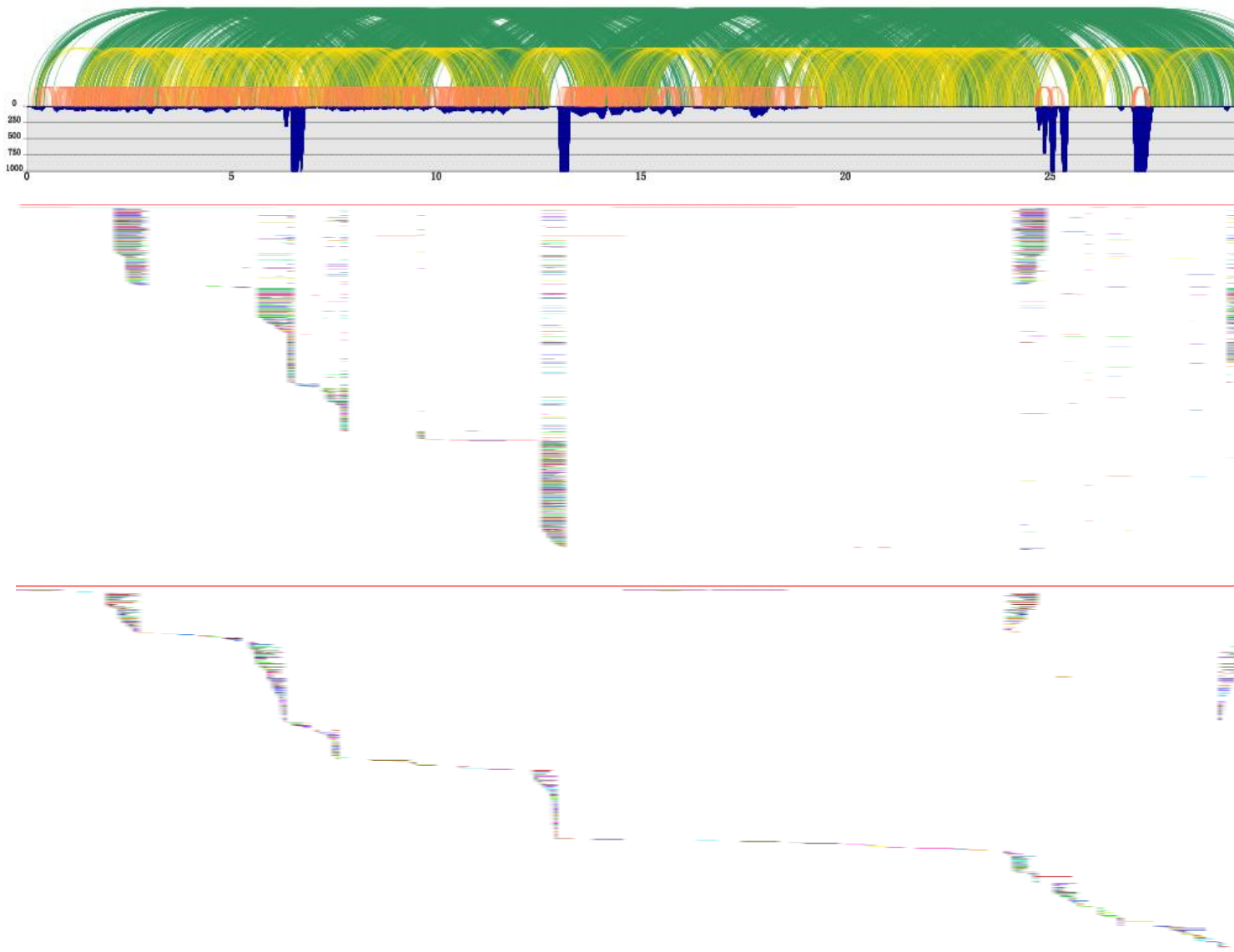


Pst-CY32

Pst-130

j)

txjbx



Paired-end supporting data from the 6-kb PE library.

Paired-end supported data from the 2-kb PE library.

Paired-end supported data from the<1-kb PE library.

Coverage and depth of Sanger sequencing.

Unfiltered different contigs.

Supplementary Fig. S4 | Evaluation of the *Pst*-CY32 genome assembly by mapping to fully Sanger sequenced 10 fosmid clones. The quality of *Pst*-CY32 assembly was assessed by mapping scaffolds to fully sequenced fosmids and paired-end (PE) supporting data. Matching contigs from the *Pst*-130 genome were included to show the improved integrity of *Pst*-130 assembly. Validation results demonstrate that the *Pst*-CR32 assembly is highly confident (Average matched ratio reached 96%). In comparison with the *Pst*-130 assembly generated with conventional WGS Illumina sequencing data, the integrity increased on average 7.2 folds in fragment numbers. Some regions which could not be assembled well were represented in PE and/or sequence supported data, indicating a relative rational confidence. Some complicated regions, such as fosmid_Txjbx, may be difficult to assemble due to highly repetitive sequences or SV regions related to genome heterozygosity.

a) **Fosmid_Txjkx.** Scaffold 475 of *Pst*-CY32 matched to Txjkx except for one 1,743-bp PE and sequence supported region (likely a SV existed). The same region consists of 11contigs in the *Pst*-130 assembly using WGS Illumina sequencing data only.

b) **Fosmid_Txjex.** ScaffoldsC19720 (with one 1,748-bpPE and sequence supported gap) and C19418 of *Pst*-CY32 matched to the region of Txjex that consists of 14 contigs in the *Pst*-130 assembly.

c) **Fosmid_Txjjx.** Scaffolds62 (with two PE supported gaps) and C14402 of *Pst*-CY32 matched to Txjjx. The same region consists of 13 contigs in the*Pst*-130 assembly.

d) **Fosmid_Txjgx.** Scaffold 619(with three sequence supported gaps) covers the entire length of Txjgx. The same region consists of 15 contigs in the*Pst*-130 genome assembled with WGS Illumina sequencing data.

e) **Fosmid_Txjfx.** Scaffolds 689 (two PE supported gaps), C19204, and C18874 (two PE supported gaps) of *Pst*-CY32 matched to the region of fosmid_Txjfx that is represented by 15 contigs in the*Pst*-130 assembly.

f) **Fosmid_Txjdx.** Scaffold509 with three PE or sequence supported gaps matched to the region of fosmid_Txjdx that has 24 contigs in the *Pst*-130 assembly.

g) **Fosmid_Txjcx.** Scaffolds C22294, C17368, and C11626 of *Pst*-CY32 matched to the region of fosmid_Txjcx that has 12contigs in the*Pst*-130 assembly.

h) **Fosmid_Txjax.** Scaffolds 698 (with three gaps) and Scaffolds 773 of *Pst*-CY32 matched to the region of fosmid_Txjax that has 6 contigs in the *Pst*-130 assembly.

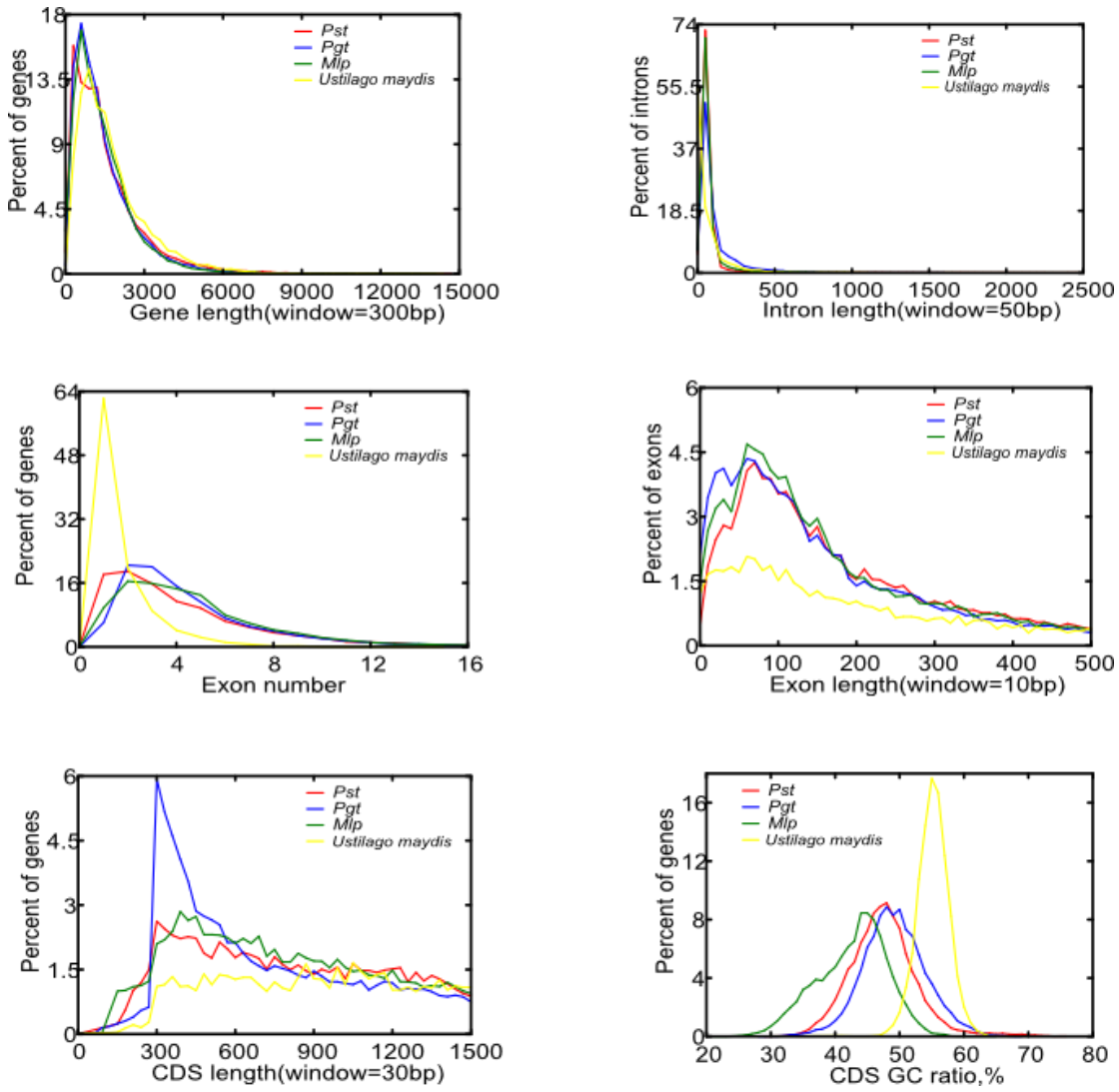
i) **Fosmid_Txjhx.** Scaffolds 826, C14402, and C13190 ofPst-CY32 matched to fosmid_Txjhx. Scaffold 826 have a 1K insertion and a 12K deletion in the 36309 to 37320 region which were supported by PE sequencing data. The same region consists of 19 contigs in the Pst-130 assembly. The dissection for unfiltered contigs of *Pst*-CY32 and *Pst*-130 matched to fosmid_Txjhx shows that both the 1k insertion and 12k deletion regions are covered with unfiltered reads and have abnormal cover depth. Failure to assemble these unfiltered contigs resulted from its complexity, manifestly indicating it may be SV and high repetitive region.
Red line showed correlative fosmid sequence.

j) **Fosmid_Txjbx.** No matching scaffolds of *Pst*-CY32 or contigs of *Pst*-130 aligned well to fosmid_Txjbx. However, sequences from unfiltered contigs/scaffolds matched to small regions of Txjbx, of which some had abnormally high cover depth, indicating that these areas consist of highly repetitive or heterozygous sequences.

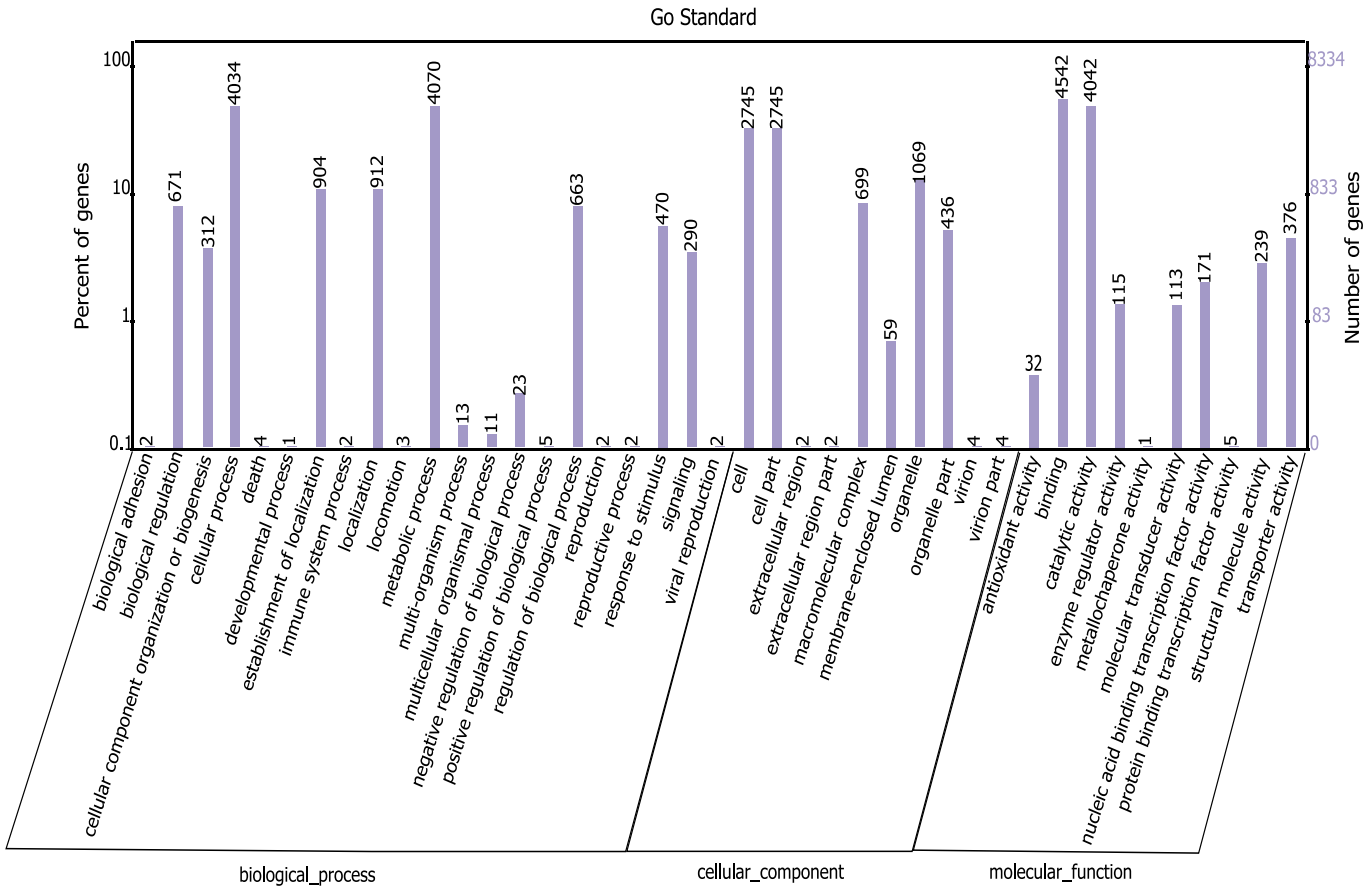
Data from the *Pst*-CY32 genome include contigs and scaffolds assembled with the 'fosmid to fosmid' approach, Illumina reads from 'fosmid to fosmid' sequencing for pair-end supporting analysis (PE library insert size 500-bp; average read length 93-bp; total PE sequence: 123,240,156-bp), and WGS Illumina reads from pair-end libraries with average insert size of 6-kb (864,513,496-bp), 2-kb (1256,790,128-bp), and <1-kb (5,490,267,014-bp). Sequencing data of the *Pst*-130 genome was cited from NCBI [14].

The labels on the figure left are:
Sanger: randomly selected *Pst*-CY32 fosmids fully sequenced by Sanger sequencing.

***Pst*-CY32**: matching scaffolds in the *Pst*-CY32 genome assembly.
Gap: sequence gap within a scaffold;
Sequence support: absent sequence based on mapping fosmid with sequence assembled in other scaffold or contig;
Gap+sequence support: gap with sequence support.
***Pst*-130**: contigs of *Pst*-130 from WGS Illumina sequencing. Different colors represent different contigs.
High PE support density reflects high assembly confidence. Existed PE and sequence supported gap indicated this region may have SV (Structure variation) derived from genome heterozygosity or highly repetitive fragments.

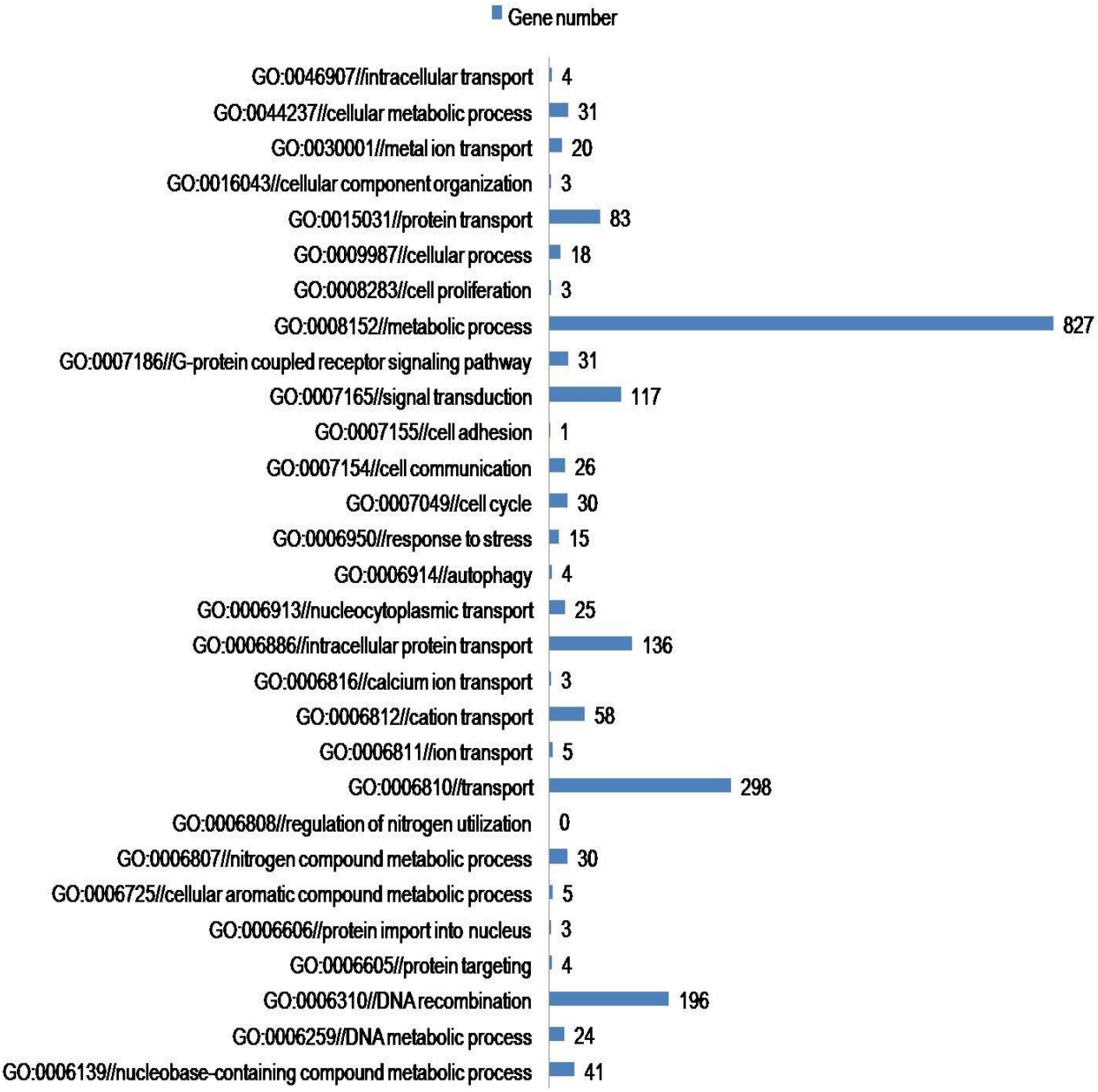


Supplementary Fig. S5 | Comparison of the gene parameters among fungal genomes of three rust and common corn smut. Six gene parameters were chosen for analysis: Plots of gene lengths, intron lengths, Exon numbers, Exon lengths, CDS lengths, and CDS GC ratios. No distinct differences were seen for *Pst* compared to the other rust fungi, suggesting that the *Pst* gene annotation is of high quality.

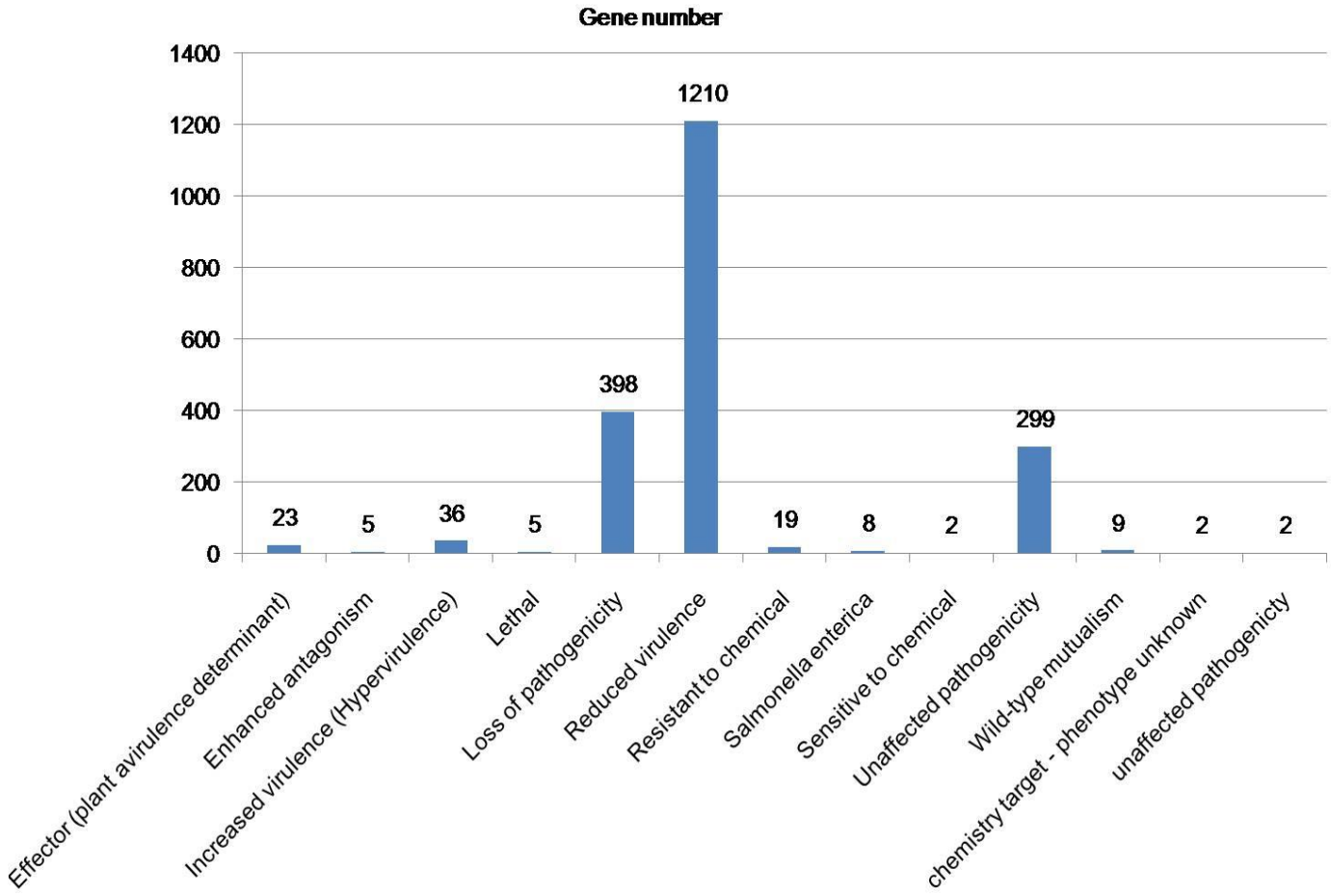


Supplementary Fig. S6 | Functional categories and distribution of predicted *Pst* genes covered by GO database.

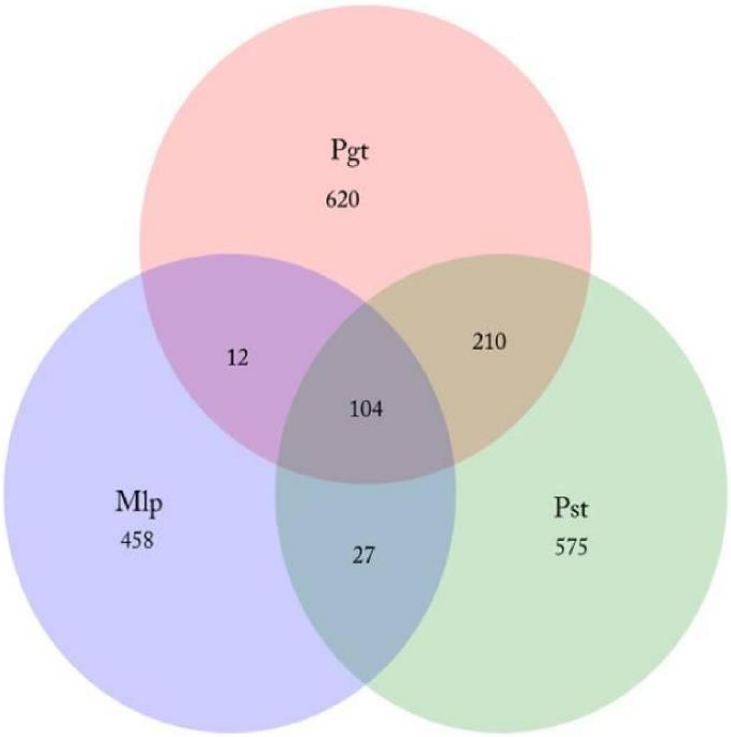
The figure shows the numbers of predicted *Pst* genes belonging to different functional groups based on the GO database. Only about 33% of the predicted *Pst* genes (8,334 genes in total) have homologues with known functions in GO database. Among them, the groups of cellular process, metabolic process, binding, and catalytic activity are mostly in abundance.



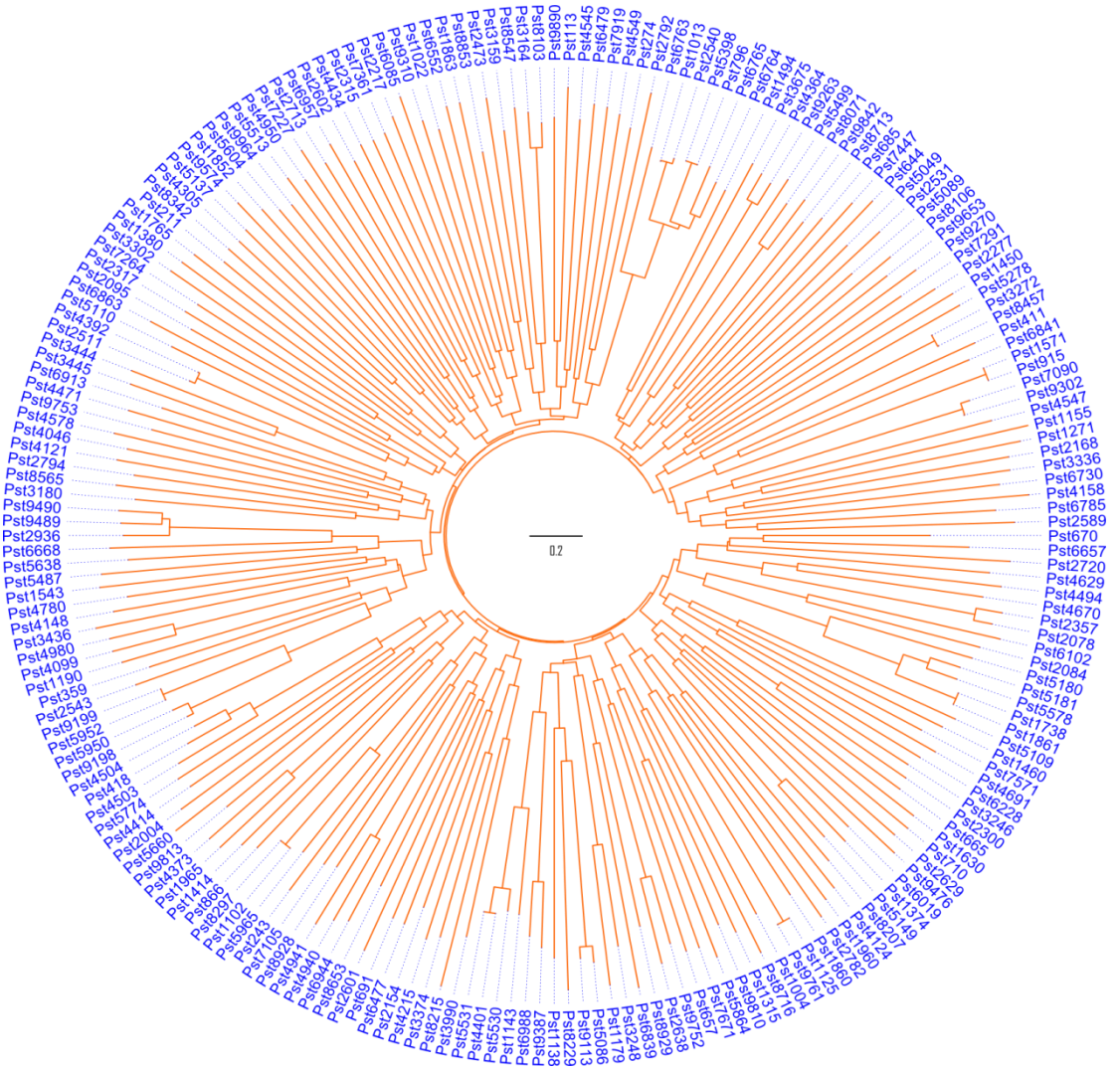
Supplementary Fig. S7 | Functional categories and distribution of predicted *Pst* genes covered by PAMGO database. The result presented the numbers of predicted *Pst* genes belonging to different functional groups based on the PAMGO database. Remarkably *Pst* is found to be abundant in genes involved in DNA recombination, which might be related to its evolutionary adaptation to environmental stresses.



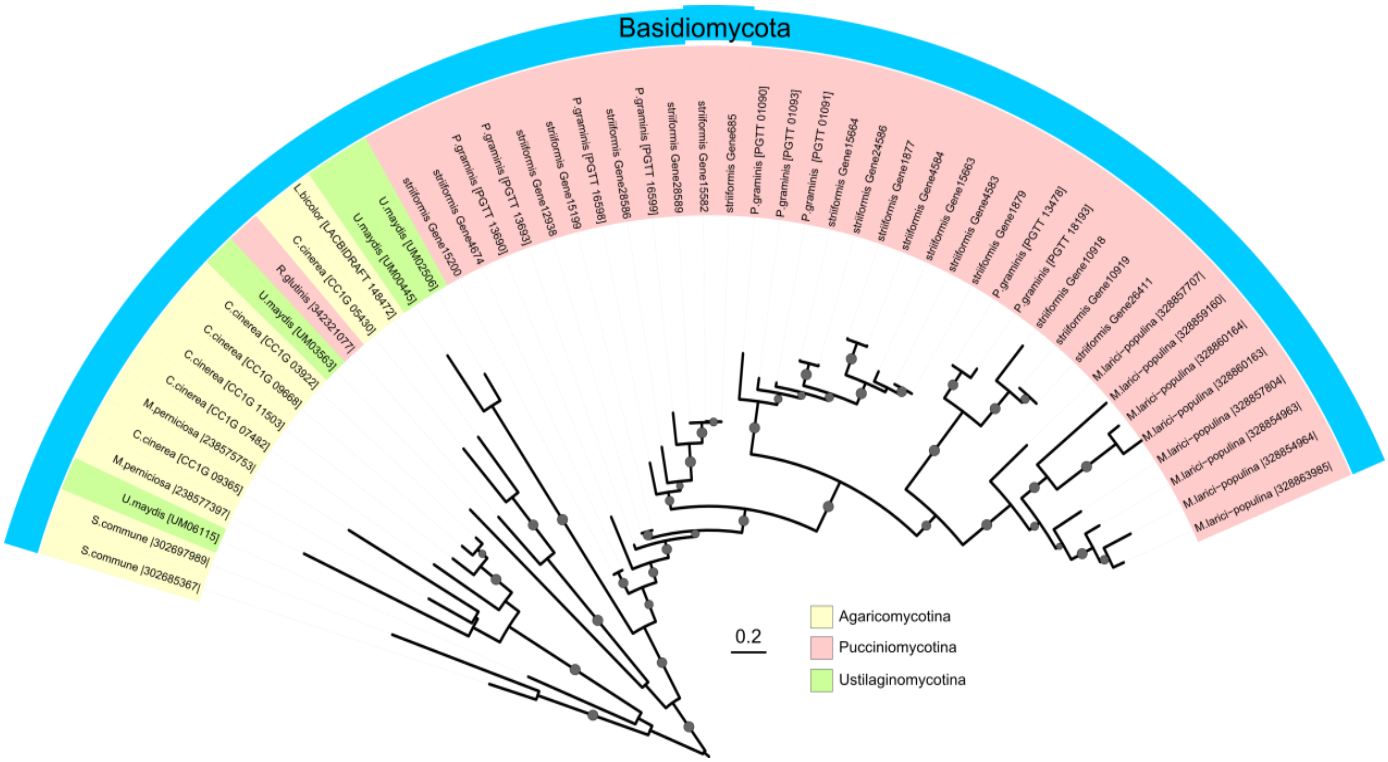
Supplementary Fig. S8 | Functional categories and distribution of predicted *Pst* genes covered by PHI database. The figure shows the numbers of predicted *Pst* genes belonging to different functional groups based on the PHI database. Only about 8% of the *Pst* genes (2,018 genes in total) are found to have the homologues with known function in PHI database, reflecting inadequate investigation of gene functions in rust pathogens. More than half of them are genes related to reduced virulence or loss of pathogenicity. These genes may play important role in the infection and development of rust pathogen.



Supplementary Fig. S9 | Venn diagram of putative secreted protein (SP) gene families among *Pst, Pgt* and *Mlp*. Orthologous SP gene families were merged into multiple species orthologous groups by MultiParanoid.



Supplementary Fig. S10 | Phylogenic tree of 198 randomly selected putative *Pst* SP genes established according to the CDS similarity. No obvious clustering of SPs within the *Pst* secretome was found in this analysis, and the SPs showed great sequence diversity with few members grouping in 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. S11 | Family expansion of cutinase genes in three rust fungi. Multiple alignments of amino acid sequences were constructed using COBALT (www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?link_loc=-BlastHomeAd) and manually edited. Maximum likelihood (ML) phylogenies were estimated by using program PhyML3.0 with the LG model. The reliability of internal branches was evaluated based on SH-aLRT supports. The midpoint rooted base tree was drawn using Interactive Tree Of Life Version 2.1.1 (itol.embl.de). The p-values of approximate likelihood ratios (SH-aLRT) plotted as circle marks on the branches and circle size is proportional to the p-values. Scale bars correspond to 0.2 amino acid substitutions per site. L. bicolor: Laccaria bicolor; C. cinerea: Coprinopsis cinere; M. perniciosa: Moniliophthora perniciosa; S. commune: Schizophyllum commune.

Supplementary Tables

Supplementary Table S1. Sequencing data used in the 17 k-mer analysis

PE Library	Insert size (bp)	Read length (bp)	Total sequence (Mb)	Q20 (%)
XN-PE1	224	75, 75	1323.33	91.61
XN-PE2	224	75, 75	1,238.37	79.69
XN-PE3	311	44, 44	1417.21	93.27
XN-PE4	515	75, 75	2403.55	83.66

Supplementary Table S2. Genome size estimation by 17 k-mer depth distribution

K-mer number	Pkdepth	Estimated genome size (bp)	Used base	Used reads
3,215,565,577	16	200,972,843	4.565E+09	84,323,652
3,215,565,577	30	107,185,519	4.565E+09	84,323,652

Supplementary Table S3. Statistics of repeat contig rates given expected contig depth as 17

	Number	Length (bp)	Repeat contigs	Repeat length (bp)	Number ratio (%)	Length ratio (%)
Short contig*	1,419,146	73,209,974	315,016	13,337,659	22.19	18.21
Long contig	226,022	61,535,829	4,169	769,263	1.84	1.25
Total contig	1,645,168	134,745,803	319,185	14,106,922	19.4	10.46

*Short contig ≤100-bp; Long contig≧100-bp.

Supplementary Table S4. Statistics of heterozygous contigs*

	Number	Length (bp)	Heterozygous number	Heterozygous length (bp)	Number ratio (%)	Length ratio (%)
Short contig	1,419,146	73,209,974	451,475	28,408,974	31.81	38.8
Long contig	226,022	61,535,829	105,379	20,861,903	46.62	33.9
Total contig	1,645,168	134,745,803	556,854	49,270,877	33.84	36.56

*Contigs with depth between 4 and 12 were considered as heterozygous fragments

Supplementary Table S5. Genome coverage of TE families (REPET consensus sequences) in *Pst* CY32.

TE component	TE component length	Coverage relative to genome (%)
Class II (DNA transposons):TIR	18,261,748	16.60
Class II (DNA transposons):Maverick	129,063	0.12
Class II (DNA transposons): Helitron	1,085,922	0.99
Class I (retrotransposons):LTR	29,885,541	27.17
Class I (retrotransposons):DIRS	945,078	0.86
Class I (retrotransposons):SINE	0	0
Class I (retrotransposons):PLE	0	0
Class I (retrotransposons):LINE	371,650	0.34
noCat: noCat *	1,101,404	1.00
Total	53,799,116	48.91

* Consensus sequences without any known structure or similarity were classified as “NoCat”

Supplementary Table S6. Genome coverage of TE families (REPET consensus sequences) in *Pgt*.

TE component	TE component length	Coverage relative to genome (%)
Class II (DNA transposons):TIR	14,758,804	16.65
Class II (DNA transposons):Maverick	506	0.00
Class II (DNA transposons):Helitron	1,339,916	1.51
Class I (retrotransposons):LTR	25,901,650	29.22
Class I (retrotransposons):DIRS	1,385,587	1.56
Class I (retrotransposons):SINE	16,492	0.02
Class I (retrotransposons):PLE	410	0.00
Class I (retrotransposons):LINE	1,026,360	1.16
noCat: noCat	310,170	0.35
Total	42,160,698	47.56

* Consensus sequences without any known structure or similarity were classified as “NoCat”

Supplementary Table S7. Genome coverage of TE families (REPET consensus sequences) in *Mlp*.

TE component	TE component length	Coverage relative to genome (%)
Class II (DNA transposons):TIR	19,789,406	19.57
Class II (DNA transposons):Maverick	181,956	0.18
Class II (DNA transposons):Helitron	1,087,075	1.07
Class I (retrotransposons):LTR	26,009,398	25.72
Class I (retrotransposons):DIRS	5,199,656	5.14
Class I (retrotransposons):SINE	12,383	0.01
Class I (retrotransposons):PLE	1,960	0.00
Class I (retrotransposons):LINE	1,526,312	1.51
noCat: noCat	131,580	0.13
Total	49,885,177	49.33

* Consensus sequences without any known structure or similarity were classified as “NoCat”

Supplementary Table S8.Statistics of assembled *Pst*-CY32 scaffolds matching to 10 fully Sanger-sequenced fosmids.

Fosmid	Length (bp)	Match length (bp)	Match Ratio (%)	Reads coverage (bp)
txjax	34,727	34,429	99.14	34,236
txjbx	30,801	24,814	80.56	22,664
txjcx	37,427	37,060	99.02	37,331
txjdx	43,291	42,117	97.29	43,189
txjex	34,991	34,788	99.42	34,990
txjfx	38,327	37,166	96.97	37,657
txjgx	33,824	33,727	99.71	33,823
txjhx	35,527	34,296	96.54	35,307
txjjx	34,949	34,591	98.98	34,946
txjkx	37,773	36,984	97.91	37,807
Average	36,250	35,707	98.53	36,022

Supplementary Table S9. Coverage of ESTs in the *Pst*-CY32 assembly.

EST Dataset (bp)	Number	Total Length (bp)	Bases covered by assembly (%)	Sequences covered by assembly (%)		>90% sequences in one scaffold		>50% sequence in one scaffold
						Number	Percent	Number Percent
>100	1,861	955,468	92.84	98.44	1,350	72.54	1,804	96.94
>200	1,808	946,058	92.94	98.51	1,334	73.78	1,753	96.96
>500	1,023	648,732	94.34	98.92	884	86.41	999	97.65
>1000	39	45,402	95.71	100	34	87.18	39	100

Supplementary Table S10. Statistics of the completeness of the genome based on 248 CEGs #

	Prots	%Completeness	-	Total	Average	%Ortho
Complete	226	91.13	-	458	2.03	71.24
Group 1	55	83.33	-	119	2.16	76.36
Group 2	52	92.86	-	96	1.85	63.46
Group 3	59	96.72	-	124	2.10	77.97
Group 4	60	92.31	-	119	1.98	66.67
Partial	233	93.95	-	522	2.24	76.82
Group 1	58	87.88	-	130	2.24	79.31
Group 2	53	94.64	-	106	2.00	67.92
Group 3	60	98.36	-	141	2.35	81.67
Group 4	62	95.38	-	145	2.34	77.42

These results are based on the set of genes selected by Genis Parra et al.[21]

Key:

Prots = number of 248 ultra-conserved CEGs present in genome

%Completeness = percentage of 248 ultra-conserved CEGs present

Total = total number of CEGs present including putative orthologs

Average = average number of orthologs per CEG

%Ortho = percentage of detected CEGS that have more than 1 ortholog

Supplementary Table S11. Main cell wall hydrolyzing enzymes of plant biotrophic, hemibitrophic and necrotrophic fungal pathogens

	Cutinase	**Pectinases**				**Cellulase**			**Hemicellulases**	
Fungi		Pectate lyase	Pectin lyase	Pectin esterase	Polygalacturonase	EG	CBH	BG	Xylanase	Mannanase
P. striiformis	18	1	1	9	2	33	9	2	11	23
P. graminis	9	0	3	7	1	19	8	2	7	15
M. larici-populina	8	0	3	3	3	20	3	2	6	11
U. maydis	4	0	1	1	1	15	0	3	5	1
B. graminis	2	0	0	0	0	4	0	1	0	7
F. graminearum	12	7	9	3	6	33	3	20	26	8
Laccaria bicolor	1	0	0	3	7	3	0	2	0	1

Description：EG, endo-1,4-β-D-glucanase; BG, β-1,4- glucosidase; CBH, exo-1,4-β-D-glucannase

Supplementary Table S12. Virulence of *Pst* strains used for re-sequencing

wheat differentials	containing resistant genes	PK-CDRD (Pakistan）	Hu09-2 (Hungary)	104E137A- (Australia)	CY23 (China)	*Pst*-78 (US)	CY32 (China)	*Pst*-130 (US)
AVS/6*Yr1	*Yr1*	R	R	R	R	R	S	R
AVS/6*Yr2	*Yr2*	R	R	R	R	S	S	S
AVS/6*Yr5	*Yr5*	R	R	R	R	R	R	R
AVS/6*Yr6	*Yr6*	S	S	S	S	S	S	S
AVS/6*Yr7	*Yr7*	S	S	S	R	S	S	S
AVS/6*Yr8	*Yr8*	S	S	S	S	S	S	S
AVS/6*Yr9	*Yr9*	R	R	R	R	S	S	S
AVS/6*Yr10	*Yr10*	R	R	R	R	R	S	S
AVS/6*Yr15	*Yr15*	R	R	R	R	R	S	R
AVS/6*Yr17	*Yr17*	R	R	R	R	S	R	R
AVS/6*Yr18	*Yr18*	S	S	S	R	R	R	S
AVS/6*Yr24	*Yr24*	R	R	R	R	R	R	R
AVS/6*Yr26	*Yr26*	S	R	S	R	R	R	R
AVS/6*Yr27	*Yr27*	R	R	R	R	R	R	R
AVS/6*Yr28	*Yr28*	R	R	S	R	R	S	R
AVS/6*Yr29	*Yr29*	S	R	R	R	R	R	R
AVS/6*YrSp	*YrSp*	R	S	R	R	R	S	S

* R,resistant (avirulent); S, susceptable (virulent). Virulence of *Pst*-130 was cited from literature.

Supplementary Table S13. Depth and coverage statistics of re-sequenced isolates and *Pst*-130

Isolates	Map_ratio	Unmap_ratio	Cover_ratio*	Average depth
CY23	84.39	4.04	95.43	28.94
PK-CDRD	84.32	4.11	95.35	27.3
Pst-78	84.33	4.1	95.36	25.09
Hu09-2	84.32	4.11	95.35	22.8
104E137A-	84.38	4.05	95.42	22.24
Pst-130※	83.64	4.79	94.58	24.83

* Cover_ratio was calculated for the ratio of reads covered length excluding gaps in reference genome assembly (gap-ratio, 11.57%).

※Sequencing data of *Pst*-130 was downloaded from NCBI.

Supplementary Table S14. InDel types revealed by re-sequencing.

Strain	Total InDel	Homo*		Hetero		Gene Region
		Num	Rate (%)	Num	Rate (%)	CDS
CY23	1,665	935	56.16	730	43.84	316
104E137A-	1,990	929	46.68	1,061	53.32	168
PK-CDRD	1,799	957	53.2	842	46.8	154
Pst-78	1,849	914	49.43	935	50.57	160
Hu09-2	2,011	958	47.64	1,053	52.36	177

*****Homo, homological SNP; Hetero, heterozygous SNP; CDS, coding sequence.

Supplementary Table S15. SV types revealed by re-sequencing.

Strain	Total	Insertion		Deletion		Complex	
		Number	Rate（%）	Number	Rate (%)	Number	Rate (%)
CY23	631	0	0	631	100	0	0
104E137A-	695	221	31.8	474	68.2	0	0
PK-CDRD	809	178	22	629	77.75	2	0.25
Pst-78	838	211	25.18	627	74.82	0	0
Hu09-2	823	143	17.38	680	82.62	0	0

Supplementary Note 1

Estimation of the *Pst* genome size by k-mer analysis

Sequence data from 4 pair-end (PE) libraries (Supplementary Table S1) were used for the k-mer 17 frequency distribution analysis. The size of the genome was estimated by establishing the frequency of occurrence of each 17-bp k-mer (a unique sequence of k (that is, 17) nucleotides in length) within the genomic sequence data set (from the 224-bp, 311-bp and 515-bp libraries). Genome size was estimated using a modification of the Lander–Waterman algorithm, where the haploid genome length in base pairs is $G = (N \times (L - K + 1) - B)/D$, where N is the read length sequenced in base pairs, L is the mean length of sequence reads, K is the k-mer length (17 bp), B is the number of k-mers occurring less than four times and D is the peak value of k-mer.

The 17 K-mer depth distribution of WGS Illumina reads presented two peaks (Figure 2A). The estimated genome size of *Pst* was approximately 200- or 107-Mb (Supplementary Table S2) depending on whether 16 or 30 was considered as the main peak, indicating that the two haploid nuclei of dikaryotic urediniospores differ significantly in genome sequences.

Supplementary Note 2

Estimation of the *Pst* genome size by contig frequency analysis

Contigs were preliminarily assembled by using SOAPdenovo 31-mer parameters. Based on the statistic results for assembled contigs length and depth, contigs longer than 1,000 bp were selected for calculating average depth and used as the average depth of the genome (AvgCvg: average coverage). Two peaks of Cvg (7 and 17) were identified in the distribution of contig length and depth (Figure 2B). Consider that the *Pst* genome has heterozygous phenomenon, the expected depth of contigs would be 17. Contigs with depth ≧34 are the repetitive contigs. Given the species has heterozygous nature, its contig repetitive level is acceptable (Supplementary Table S3) and has a certain number of heterozygous regions with a combined length of approximately 24.6 Mb (49,270,877/2). The length of the non-heterozygous regions accounts for approximately 85.4 Mb (134,745,803 - 49,270,877-bp), including repetitive regions of 14.1 Mb and unique regions of 57.2 Mb (Supplementary Table S4). Therefore, the haploid genome size of *Pst* was estimated to be approximately **110 Mb** (24.5 +85.4 =109.9 Mb).

The results here together with that of Fig. 2, Supplementary Fig. S2, and Supplementary Note 1 indicate the dikaryotic *Pst* genome is highly heterozygous, its genome size is estimated to be approximately 110 Mb.