ECFG15
ROME • ITALY 2020
15th European Conference on Fungal Genetics
Poster # C2-36

Draft genome assemblies of a global collection of *Pyrenophora tritici-repentis* (tan spot of wheat)

Ryan Gourlie[1], Rodrigo Ortega Polo[1], Kaveh Ghanbarnia[1], Mohamed Hafez[1], Raja Ragupathy[1], Fouad Daayf[2], Stephen Strelkov[3], and Reem Aboukhaddour[1]

[1]Agriculture and Agri-Food Canada, Lethbridge Research and Development Center, Lethbridge, AB, Canada; [2]University of Manitoba, Department of Plant Science, Winnipeg, MB, Canada; [3]University of Alberta, Department of Agricultural, Food and Nutritional Science, Edmonton, AB, Canada

Agriculture and Agri-Food Canada

Introduction

The fungus *Pyrenophora tritici-repentis* (Ptr) causes tan spot, a destructive foliar wheat disease worldwide. Eight races of Ptr have been identified based on their ability to produce combinations of three necrotrophic effectors: Ptr ToxA, Ptr ToxB and Ptr ToxC. In this study, the complete haploid asexual genomes 37 isolates representing all known races from various geographical regions were sequenced on the Illumina HiSeq X platform and two isolates out of that pool were also sequenced using PacBio RS II to generate hybrid assemblies to be used as reference genomes. De novo assemblies were generated with Illumina reads using the programs: Shovill with SPAdes[1,2], Shovill with MEGAHIT[2,3], SOAPdenovo2[4], and CLC Genomics Workbench 12[5]. Genome annotation was carried out with the FunGap pipeline[6]. Parameters such as N50, number of contigs, number of gaps ('N' insertions), and number of annotated genes were used to compare the de novo assembly algorithms for this dataset. To our knowledge, this is the most comprehensive collection of global Ptr genomes assembled, and preliminary results from pan-genome analysis are presented.

Materials and Methods

Fungal isolates, DNA extraction, and sequencing

Ptr isolates in this study were collected from Canada (21), Algeria (3), Azerbaijan (8), Syria (3), and Tunisia (5) (Table 1) DNA for Illumina HiSeq X sequencing (all isolates, 150bp paired-end) was extracted using 'Genomic-tip 20/G' (Qiagen) and for (two isolates) PacBio RS II sequencing using 'Genomic-tip 100/G' (Qiagen). Sequencing was performed by Genome Quebec.

Table 1. Races of Ptr and number sequenced (+ = present; - = absent)

Race	PtrTox			Number sequenced
	A	B	C	
1	+	-	+	10
2	+	-	-	6
3	-	-	+	4
4	-	-	-	3
5	-	+	-	8
6	-	+	+	3
7	+	+	-	2
8	+	+	+	3
novel	-	+	-	1
Total				40

De novo assembly and annotation

A workflow is shown in Figure 1. Briefly, four assembly programs were assessed for their suitability to create de novo assemblies using Illumina reads. Hybrid assemblies were created using PacBio reads assembled with Flye[7,8] and then polished with Illumina reads using Pilon[9]. Gene annotations were done using the FunGAP pipeline utilizing publicly available RNA reads[10]. Assemblies were assessed for quality based on N50, number of contigs, number of gaps, number of annotated genes, and BUSCO[11] completeness (ascomyocota gene set v3.0.2)

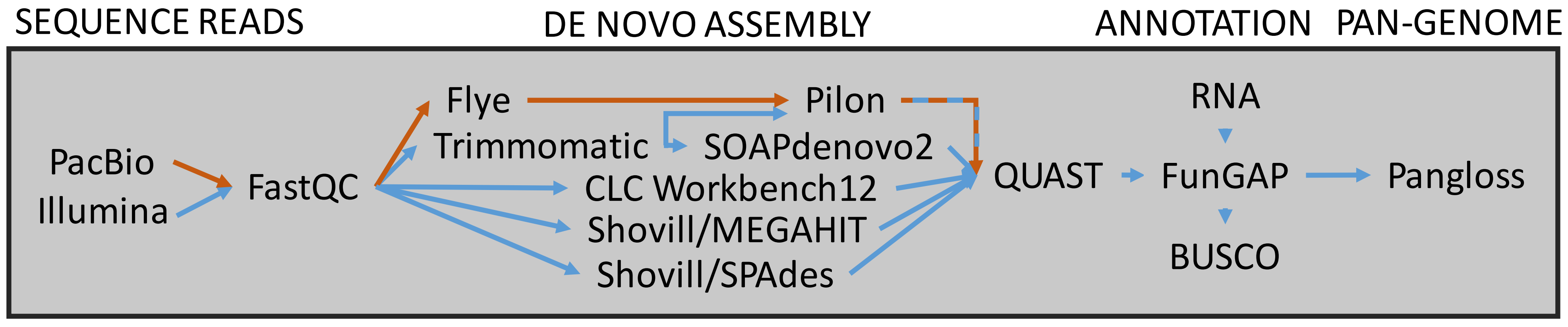

Figure 1. Workflow diagram for Ptr assemblies and pan-genome (blue=short; orange=long reads).

Pan-genome

The Pangloss[12] pipeline was used to perform pan-genome analysis of the Ptr SPAdes assemblies using default settings (with --no_pred). We included a reference genome of a race 1 isolate assembled by the Broad Institute[13] and retrieved from GenBank. Figures visualized with R (v3.4.3) and Bandage (v.0.8.1)[14].

Results

De novo hybrid assemblies using PacBio + Illumina reads

Assemblies of the PacBio reads have almost captured near-chromosome level resolution of the Ptr genomes (Figure 2). The two Flye assemblies (isolates D308 and I73-1) had 70 and 39 contigs with N50s of 3,667,238 and 3,646,749 bp respectively.

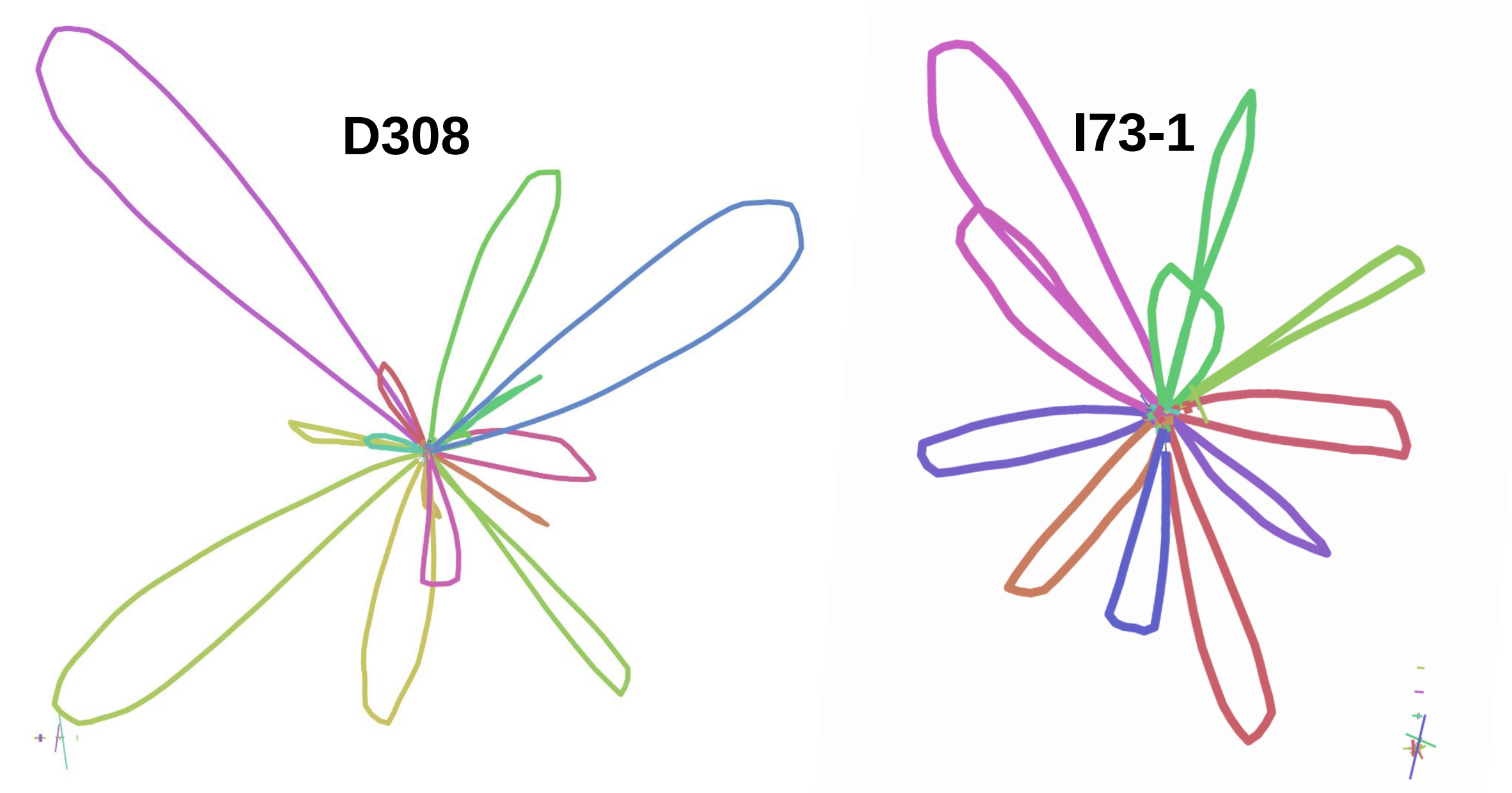

Figure 2. Flye assembly graphs from PacBio reads for two Ptr isolates.

De novo assembly of Illumina reads

Shovill with SPAdes with Shovill created assemblies with high N50 values and low number of contigs while having no gaps (Table 2). FunGAP predicted similar numbers of genes for each assembly program, generally SOAP assemblies had the fewest genes. All programs for all isolates (one exception) produced assemblies with >99% completeness using the ascomycota BUSCO gene set (v.3.0.2), there was minor variability in percentage of BUSCO genes with fragmentation (0.4% – 0.8%).

Results

Table 2. Summary statistics of four assembly programs using Illumina reads for five Ptr isolates.

Isolate	Location	Race	Tox	Assembler	QUAST			FunGAP	BUSCO
					N50 (bp)	contigs	N's	genes	%
90-2	Alberta/Sask	4	-	CLC Workbench12	170,424	2,220	~27K	13,005	99.4
90-2	Alberta/Sask	4	-	MEGAHIT (Shovill)	87,309	37,952	0	13,112	99.5
90-2	Alberta/Sask	4	-	SPAdes (Shovill)	287,769	3,872	0	13,011	99.5
90-2	Alberta/Sask	4	-	SOAPdenovo2	294,236	9,296	~16K	12,976	99.5
AB88-2	Alberta	2	A	CLC Workbench12	69,421	2,782	~18K	13,086	99.6
AB88-2	Alberta	2	A	MEGAHIT (Shovill)	46,933	34,163	0	13,045	99.7
AB88-2	Alberta	2	A	SPAdes (Shovill)	80,901	6,504	0	13,010	99.5
AB88-2	Alberta	2	A	SOAPdenovo2	82,444	4,374	~10K	12,885	99.5
ASC1	Manitoba	1	AC	CLC Workbench12	64,418	2,859	~10K	13,004	99.4
ASC1	Manitoba	1	AC	MEGAHIT (Shovill)	43,607	32,968	0	13,089	99.4
ASC1	Manitoba	1	AC	SPAdes (Shovill)	78,535	6,518	0	13,089	99.5
ASC1	Manitoba	1	AC	SOAPdenovo2	92,353	4,412	~15K	11,430	89.4
AZ35-5	Azerbaijan	5	B	CLC Workbench12	66,400	2,974	~25K	13,248	99.6
AZ35-5	Azerbaijan	5	B	MEGAHIT (Shovill)	43,709	38,454	0	13,091	99.7
AZ35-5	Azerbaijan	5	B	SPAdes (Shovill)	77,908	7,229	0	13,214	99.6
AZ35-5	Azerbaijan	5	B	SOAPdenovo2	79,529	5,661	~10K	13,127	99.6
I72-1	Syria	3	C	CLC Workbench12	54,919	2,663	~23K	12,893	99.6
I72-1	Syria	3	C	MEGAHIT (Shovill)	40,319	24,387	0	12,948	99.5
I72-1	Syria	3	C	SPAdes (Shovill)	63,650	6,744	0	12,886	99.6
I72-1	Syria	3	C	SOAPdenovo2	65,951	4,573	~11K	12,904	99.6

Annotation and Pan-genome

The average number of coding genes predicted by FunGAP was 13,028 for the 38 isolates presented (37 seq + 1 GenBank). The core genome is comprised of 10,601 genes with 20,315 unique genes (Figure 3). Of the 9,714 genes in the accessory pan-genome 4,678 (48.2%) are present as singletons. Based on the extrapolated growth curve (Figure 4) and the large number of accessory genes present in the pan-genome, Ptr has an open genome with an average increase of ~200 new genes per additional genome.

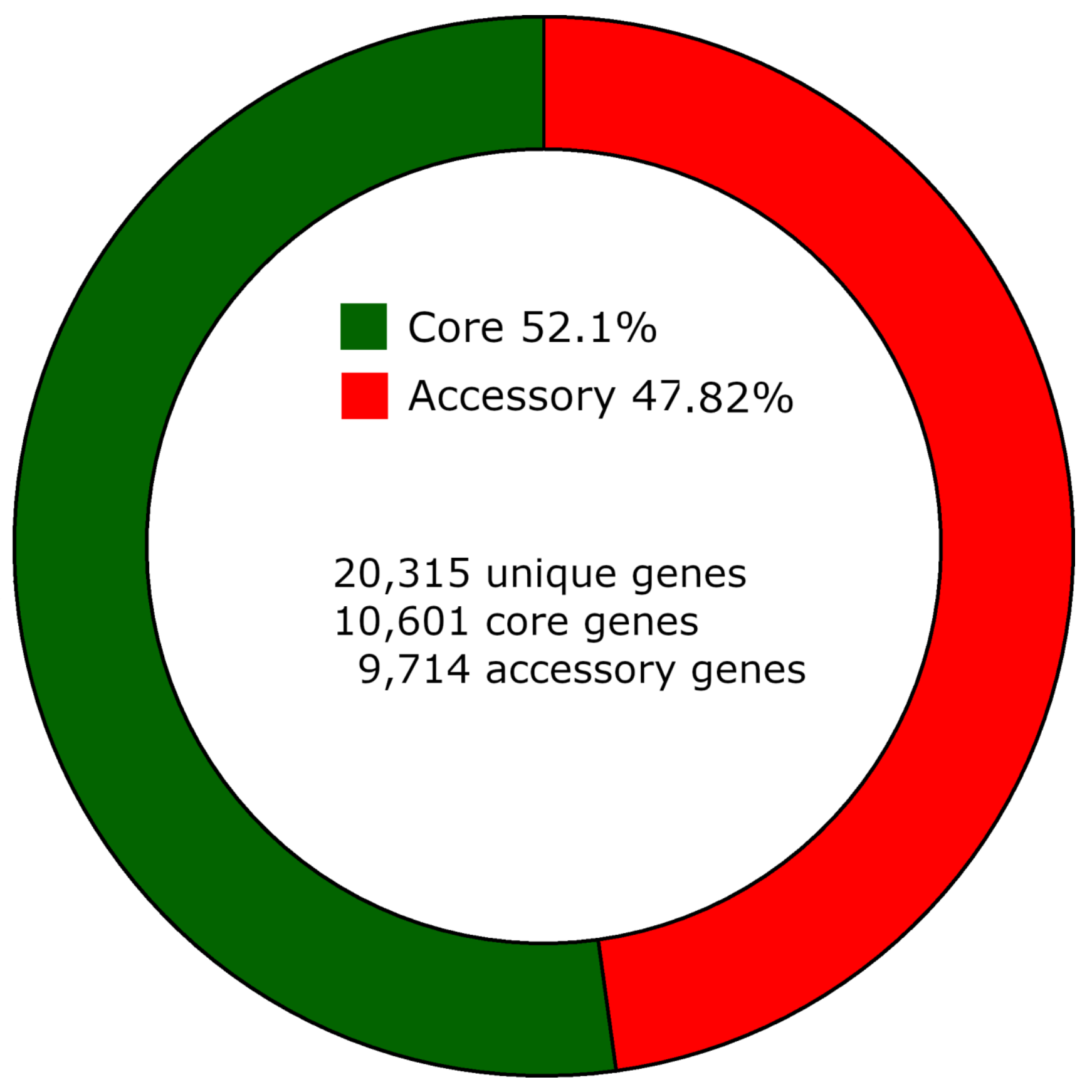

Figure 3. Core and accessory genes of the Ptr pan-genome based on 38 isolates.

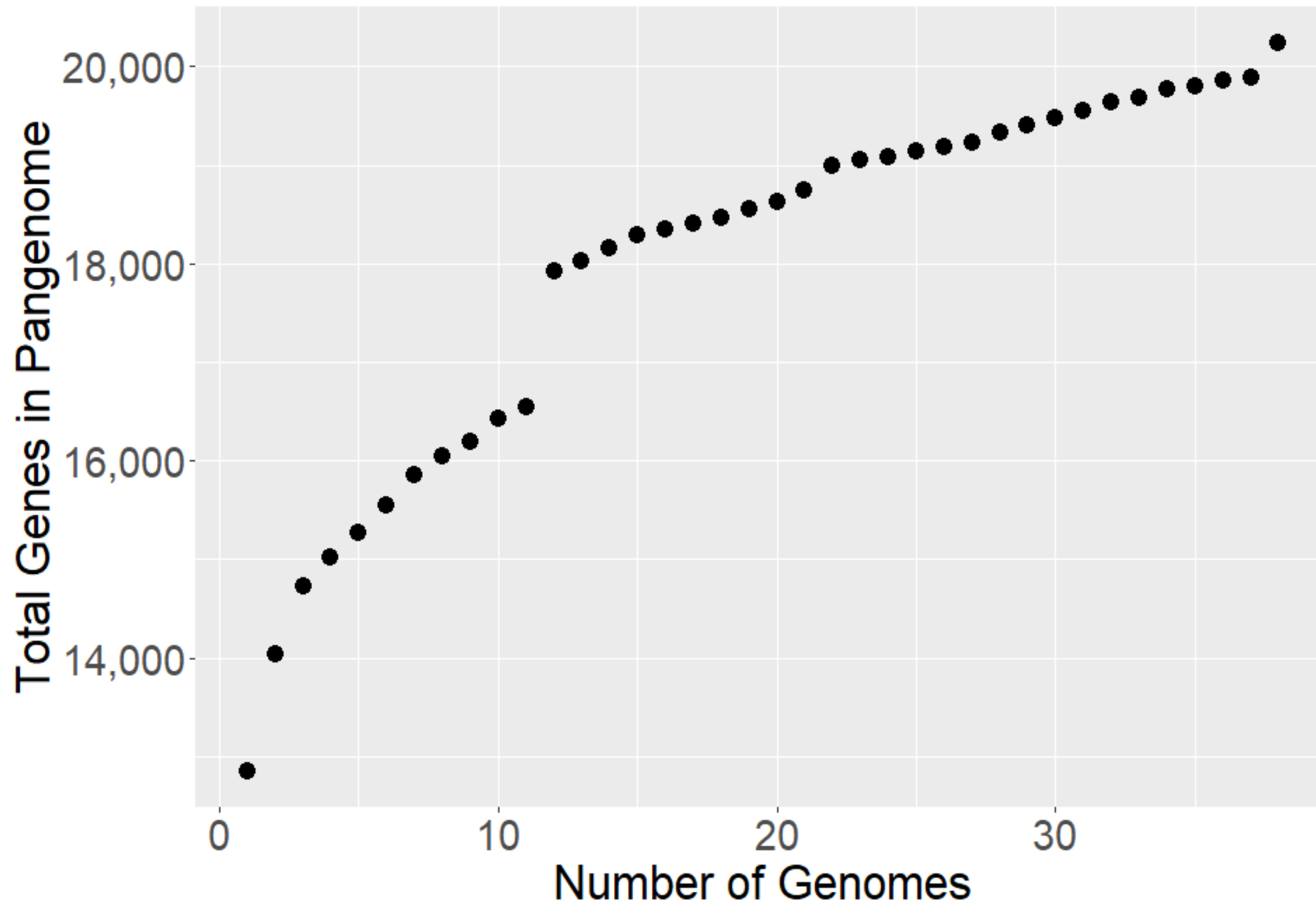

Figure 4. Number of genes in Ptr pan-genome as number of genomes included increases. The curve continues to increase suggesting Ptr has an open genome.

Conclusions

- For Illumina reads, Shovill with SPAdes created high quality de novo assemblies based on N50, contigs, gaps, predicted gene sets, and BUSCO completeness and is recommended for future Ptr de novo assemblies.
- Assemblies using PacBio reads are near-chromosome level resolution.
- Ptr appears to have an open genome, this is not surprising as Ptr is a widespread plant pathogen infecting wheat all around the globe.
- Ptr has a large accessory pan-genome (47.8% of genes) with 48.2% of the accessory genes present as singletons. These results are similar to the pan-genome of a closely related genus *Parastagonospora* sp.[15] where ~60% of the genes were accessory.
- Future studies will attempt to understand the genome structural organization among the different Ptr races.

References

[1]Bankevich et al., 2012. *Journal of Computation Biology*, 19(5), 455-477; [2]Seemann, 2019. github.com/tseemann/shovill; [3]Li et al., 2015. *Bioinformatics*, 31(10), 1674-1676; [4]Luo et al., 2012. *Gigascience*, 1(1), 18; [5]Qiagen, 2018; [6]Min et al., 2017. Bioinformatics 33(18), 2936-2937; [7]Kolmogorov et al., 2019. *Nature Biotechnology* 37(5), 540–546; [8]Lin et al., 2016. *Proceedings of the National Academy of Sciences*, 113(52), E8396-E8405; [9]Walker et al., 2014. *PloS One*, 9(11), e112963; [10]Moolhuijen et al., 2018. *BMC Research Notes*, 11(1), 907-909; [11]Simão et al., 2015. *Bioinformatics* 31(19), 3210-3212; [12]McCarthy & Fitzpatrick, 2019. *Genes* 10(7), 521; [13]Manning et al., 2013. *G3* 3(1), 41-63; [14]Wick et al., 2015. Bioinformatics 31(20), 3350-3352;[15]Syme et al., 2018. *Genome Biology and Evolution* 10(9), 2443-2457

Agriculture and Agri-Food Canada
Agriculture Funding Consortium
Alberta Wheat Commission
Sask Wheat Development Commission
Canada