Wheat Panache: A pangenome graph database representing presence–absence variation across sixteen bread wheat genomes

Philipp Bayer, Jakob Petereit, Éloi Durant, Cécile Monat, Mathieu Rouard, Haifei Hu, Brett Chapman, Chengdao Li, Shifeng Cheng, Jacqueline Batley, et al.

HAL Id: hal-04003970

https://hal.inrae.fr/hal-04003970v1

Submitted on 24 Feb 2023

ORIGINAL RESEARCH

Wheat Panache: A pangenome graph database representing presence–absence variation across sixteen bread wheat genomes

Philipp E. Bayer[1] | **Jakob Petereit**[1] | **Éloi Durant**[2,3,4,5] | **Cécile Monat**[3] | **Mathieu Rouard**[4,5] | **Haifei Hu**[6] | **Brett Chapman**[6] | **Chengdao Li**[6] | **Shifeng Cheng**[7] | **Jacqueline Batley**[1] | **David Edwards**[1]

[1]School of Biological Sciences, The Univ. of Western Australia, Perth 6009, Australia

[2]DIADE, Univ. of Montpellier, CIRAD, IRD, Montpellier 34830, France

[3]Syngenta Seeds S.A.S., 12 chemin de l'Hobit, Saint-Sauveur 31790, France

[4]Bioversity International, Parc Scientifique Agropolis II, Montpellier 34397, France

[5]French Institute of Bioinformatics (IFB)—South Green Bioinformatics Platform, Bioversity, CIRAD, INRAE, IRD, Montpellier 34398, France

[6]Western Crop Genetics Alliance, Murdoch Univ., 90 South Street, Murdoch 6150, Australia

[7]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

Correspondence
David Edwards, School of Biological Sciences, The Univ. of Western Australia, Perth 6009, Australia.
Email: dave.edwards@uwa.edu.au

Assigned to Associate Editor Henry Nguyen.

Abstract

Bread wheat (*Triticum aestivum* L.) is one of humanity's most important staple crops, characterized by a large and complex genome with a high level of gene presence–absence variation (PAV) between cultivars, hampering genomic approaches for crop improvement. With the growing global population and the increasing impact of climate change on crop yield, there is an urgent need to apply genomic approaches to accelerate wheat breeding. With recent advances in DNA sequencing technology, a growing number of high-quality reference genomes are becoming available, reflecting the genetic content of a diverse range of cultivars. However, information on the presence or absence of genomic regions has been hard to visualize and interrogate because of the size of these genomes and the lack of suitable bioinformatics tools. To address this limitation, we have produced a wheat pangenome graph maintained within an online database to facilitate interrogation and comparison of wheat cultivar genomes. The database allows users to visualize regions of the pangenome to assess PAV between bread wheat genomes.

1 | INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is one of the most widely grown crops, yet there is a significant challenge to increase yield to meet the projected demands of a growing world

---

Abbreviations: IWGSC, The International Wheat Genome Sequencing Consortium; PAV, presence–absence variation.

population. With predictions of climate-change-related yield losses ranging from 17 to 31% by the middle of the 21st century (Obembe et al., 2021), improved genomics-based breeding approaches are required to produce climate change-ready wheat cultivars.

Wheat genomics has made rapid advances in recent years with the first draft genome assembly produced in 2014 (The International Wheat Genome Sequencing Consortium [IWGSC], 2014) based on the shotgun sequencing of isolated chromosome arms (Berkman et al., 2011, 2012; Lai et al., 2015). A first near-complete assembly of the 'Chinese Spring' was produced in 2017 (Zimin et al., 2017) with a final reference genome assembly available in 2018 (IWGSC, 2018). This reference assembly was rapidly followed by assemblies of 15 additional cultivars from global breeding programs (Walkowiak et al., 2020).

The increasing availability of reference genome assemblies made it clear that there is significant presence–absence variation (PAV) between individuals (Bayer et al., 2020; Golicz et al., 2016, 2020; Hurgobin & Edwards, 2017). This insight has led to the production of pangenomes that reflect the gene content of a species rather than an individual (Bayer et al., 2021; Franco et al., 2020; Golicz et al., 2016; Jensen et al., 2020; Montenegro et al., 2017; Rijzaani et al., 2021; Ruperao et al., 2021; Song et al., 2020; Zhao et al., 2020). Pangenomes are now available for several plant species; the first bread wheat pangenome representing the gene content of 16 bread wheat cultivars was published in 2017 (Montenegro et al., 2017). This wheat pangenome was assembled using an iterative mapping approach, which efficiently identified new gene space and called gene presence or absence between individuals. This kind of pangenome is, however, limited in that the physical location of the new gene space can be difficult to determine with accuracy. With the availability of multiple whole-genome references, this limitation may be addressed through the production of a graph-based pangenome. Graph-based pangenomes have recently become popular thanks to the graph data structure, which can accurately represent the physical locations of genomic and structural variants with minimal reference bias with tools such as vg (Hickey et al., 2020), seqwish (Garrison & Guarracino, 2022), minigraph (Li et al., 2020), and PHG (Jensen et al., 2020) being successfully applied to build variation, sequence, or haplotype graphs.

A major limitation of pangenome graphs is that few tools are available to visualize these complex graph structures. Genome visualization tools, such as GBrowse (Donlin, 2009), JBrowse2 (Buels et al., 2016), or Circos (Krzywinski et al., 2009), are designed to display information relative to a linear reference genome, not a graph of several genomes, while graph viewers, such as Bandage (Wick et al., 2015), or pangenome viewers, such as ODGI (Guarracino et al., 2021),

Core Ideas

- Graph pangenomes represent more genomic variants than reference genomes.
- We present a wheat graph pangenome based on 16 public assemblies.
- We present Wheat Panache, an online visual representation of this graph.
- Wheat Panache lets users search the graph for presence–absence variants.
- We also distribute the graph preindexed for Giraffe utilization.

focus on visualizing the graph itself but display little other information such as genome annotations.

Panache is a recent pangenome visualization tool that can process linearized assembly graphs and display shared regions as a web-based dynamic heatmap (Durant et al., 2021). Panache has so far only been applied to visualize PAV in the banana (*Musa acuminata* Colla) pangenome (Rijzaani et al., 2021) but has the potential to be expanded to other species even for crop genomes as large as wheat. Here, we present a graph pangenome representing 16 bread wheat cultivars hosted within a public Wheat Panache database with a new web-based browser for visualizing genomic regions across the wheat pangenome along with the graph formatted for minimap2 (Li, 2018) and Giraffe (Jouni et al., 2021). This tool offers researchers and breeders the ability to assess genome variation between these cultivars, mining the diversity present in this large and complex genome.

2 | MATERIALS AND METHODS

We used publicly available genome assemblies including 15 high-quality bread wheat genome assemblies (Walkowiak et al., 2020) and the IWGSC v1 Chinese Spring assembly (IWGSC, 2018) to assemble a graph using minigraph v0.14 (Li et al., 2020). To optimize assembly, we used *k*-mers that appear <100 times (-f.1) for the graph assembly and assembled the graph genome by genome starting with IWGSC v1 followed by alphabetical order of cultivar names and ending with the spelt [*Triticum* aestivum L. subsp. *spelta* (L.) Thell.] PI190962 assembly.

All assemblies were aligned with the final graph using minimap2 v2.18 (Li, 2018) and alignments were converted to BED format. The main graph was linearized using gfatools gfa2bed v0.4 with default parameters (https://github.com/lh3/gfatools/releases) and merged with all minimap2 alignments using bedtools v2.30.0 multiinter (Quinlan & Hall, 2010). The
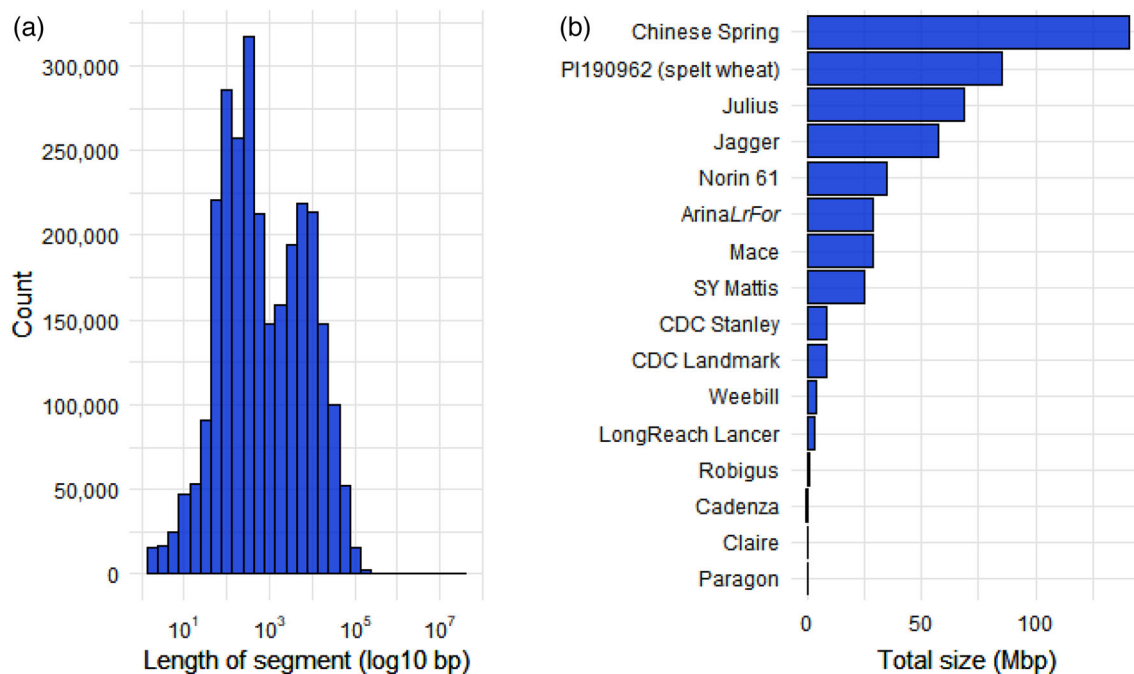
FIGURE 1 (a) Bar chart showing the distribution of the size of all assembly graph segments (log scale). (b) Total size of unique segments per cultivar in megabase pairs (Mbp). PI190962 is a line of species *Triticum spelta*, Chinese Spring is the reference cultivar of *T. aestivum*

resulting blocks were intersected with the IWGSC gene annotation using bedtools v2.30.0 intersect.

The data was converted to Panache JSON format and a Panache instance was set up to serve the data (Durant et al., 2021). To make the display feasible on a regular workstation, we retained only blocks overlapping with Chinese Spring genes and then merged adjacent blocks if they showed identical PAV behavior across all individuals.

3 | RESULTS AND DISCUSSION

3.1 | A wheat graph pangenome

We constructed a graph pangenome using 16 high-quality wheat genome assemblies representing the global variation of modern bread wheat cultivars. The assembled graph had a total size of 15.8 Gbp compared with the founder genome assembly sizes of 13.9–14.2 Gbp (Walkowiak et al., 2020). After aligning all genomes back to the graph, these 15.8 Gbp were split up into 2,791,482 segments present in at least one individual. The segments had an average size of 5.6 Mbp (median, 498 bp) ranging from 2 bp to 37.6 Mbp (Figure 1a). Realignment of the 16 genome assemblies to the graph revealed that out of the 2.7 million segments, 542,711 (19%) segments were present in all individuals (total size, 10.2 Gbp (65%) ranging from 2 bp to 4.9 Mbp; average size, 19 Kbp) with the remaining 2,248,771 segments

(total size, 5.6 Gbp) being present in a median of eight individuals with an average size of 2 Kbp (Supplemental Figure 1). The 10,437 segments (0.4% of all segments) with a total length of 19.9 Mbp (average length, 1.9 Kbp) were not covered by any genome assembly during the realignment step, probably because these segments were too small or too repetitive.

Interestingly, the cultivar with the most unique segments was the reference cultivar Chinese Spring, with 158,503 (7%) of segments with a total size of 140.5 Mbp being only present in Chinese Spring (Figure 1b). These 158,503 segments contain 2,216 genes present only in the Chinese Spring reference assembly. We searched for these genes in the IWGSC functional annotation (IWGSC, 2018). We compared the 2,216 genes present only in Chinese Spring with the IWGSC functional annotation and found genes known to be highly variable such as transposable elements or transposable element candidates (235 genes [10.6%]; $P < .05$ Chi-squared test) and disease resistance genes carrying an NB-ARC domain (PF00931, 107 [5%]; $P < .05$ Chi-squared test). These results indicate that these genes have been lost in modern cultivars relative to Chinese Spring.

This may be due to the genomic distance between Chinese Spring and the other cultivars, consistent with previous observations (Montenegro et al., 2017), and reflecting Chinese Spring's age (collected around 1900) and its lack of agronomic characters that were selected for in modern cultivars (Sears & Miller, 1985). The distance between the Chinese

FIGURE 2 Wheat Panache screenshot showing an *Aegilops ventricosa* introgression at the beginning of chromosome 2 in cultivars Stanley, Jagger, Mace, and SY Mattis (Gao et al., 2021; Keilwagen et al., 2022). Black boxes were added to show the region missing in cultivars where the introgression replaced parts of chromosome 2A. The graph assembly started with the IWGSC v1 assembly leading to linearized regions following the same naming scheme as the IWGSC v1.0 assembly (chr1A_part1, chr1A_part2, chr2A_part1, …). CS, 'Chinese Spring'. Shown here is the beginning of the first part of chr2A. Black blocks are gene models. White regions correspond to regions that are present in the graph but contain no genes

Spring assembly and the 15 other assemblies is also supported by 1.2 Gbp of the graph in 901,475 segments not being present in Chinese Spring but in at least one other cultivar, reflecting the complex history of introgressions in modern bread wheat (Keilwagen et al., 2022; Walkowiak et al., 2020). We aligned the IWGSC v1 gene annotation for Chinese Spring (IWGSC, 2018) back to the graph by intersecting the linearized graph with gene positions. We found a position in the graph for 110,790 (100%) genes confirming that the graph assembly contains all gene models of the IWGSC assembly.

We compared the wheat graph pangenome with the earlier published iterative-mapping-based wheat pangenome (Montenegro et al., 2017); this wheat pangenome contains 51,460 (32%) genes lost in at least one individual and 109,071 (68%) genes present in all individuals. We intersected the graph pangenome with the IWGSC annotation to count which IWGSC genes are lost in at least one of the assemblies. Within the 16 individuals, 47,454 (31%) genes are lost in at least one individual and 104,270 (69%) genes are present in all individuals. Even though our graph pangenome contains different accessions than the Montenegro et al. (2017) pangenome, the wheat graph pangenome gene numbers are remarkably similar, indicating general patterns of gene variation occur-

ring in both sets of cultivars relative to the Chinese Spring reference.

3.2 | The Wheat Panache web portal

Using this graph, we built a web-based Panache instance (Durant et al., 2021), allowing users to visualize regions or genes of interest for presence or absence across the chosen wheat cultivars. The webserver is online (http://www.appliedbioinformatics.com.au/wheat_panache).

Wheat Panache displays a linear version of the pangenome graph subdivided into blocks based on the presence or absence of the selected individuals. A block is defined to have no internal PAV and to contain at least one gene. Blocks are named based on the pseudomolecule they originated in, and, as we started the assembly with the IWGSC assembly, most blocks (1,890,035 out of 2,791,483 blocks, 67%) are named after their position in the IWGSC assembly.

The interface displays the linearized pangenome as a chain of such graph segments with one horizontal track per cultivar (Figure 2). Coordinates are based on the pangenome graph assembly. Genes are represented as black dots above blocks, and hovering over a gene reveals its coordinates within the

assembly and exon structure. Three summary tracks below the cultivar tracks show which blocks are core or variable based on a user-definable threshold, how long the block is, and how often the block is repeated within Panache. Users can zoom into blocks or search for 'hollow areas' (areas of consecutive absence based on a user-defined threshold) using the Hollow Area Finder, which is a convenient way to automatically focus on large PAV areas. Users can sort the cultivars alphanumerically, by gene presence or absence status, or by a phylogeny based on Mash v2.3 (Ondov et al., 2016). The graph assembly displayed in Wheat Panache, including a version preindexed for vg v1.37.0 Giraffe (Jouni et al., 2021) is available online (https://doi.org/10.5281/zenodo.6085239) (Bayer et al., 2022), allowing for downstream analyses of the population graph.

In summary, we present the first wheat graph pangenome assembly based on 16 cultivars with an online visual representation of the graph within the Panache visualization tool. The graph assembly will be a valuable tool for wheat genomics researchers looking for a more accurate reference assembly. The web platform Panache allows users to interrogate this graph and search for structural variants around regions of interest. We plan to incorporate new wheat genome assemblies into Wheat Panache as they are being released and to update the Wheat Panache instance as new versions of Panache are being released.

DATA AVAILABILITY STATEMENT

The graph assembly displayed in Wheat Panache, including a version pre-indexed for vg v1.37.0′s Giraffe (Jouni et al., 2021) is available at https://doi.org/10.5281/zenodo.6085239 (Bayer et al., 2022).

AUTHOR CONTRIBUTIONS

Philipp E. Bayer: Conceptualization; Writing – original draft; Writing – review & editing. Jakob Petereit: Writing – review & editing. Éloi Durant: Visualization; Writing – review & editing. Cécile Monat: Writing – review & editing. Mathieu Rouard: Visualization; Writing – review & editing. Haifei Hu: Writing – review & editing. Brett Chapman: Writing – review & editing. Chengdao Li: Writing – review & editing. Shifeng Chen: Writing – review & editing. Jacqueline Batley: Writing – review & editing. David Edwards: Project administration; Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Philipp E. Bayer https://orcid.org/0000-0001-8530-3067
Jakob Petereit https://orcid.org/0000-0003-2159-0380
Éloi Durant https://orcid.org/0000-0002-2734-4327
Cécile Monat https://orcid.org/0000-0002-0574-3976
Mathieu Rouard https://orcid.org/0000-0003-0284-1885
Haifei Hu https://orcid.org/0000-0003-1070-213X
Brett Chapman https://orcid.org/0000-0003-4484-2253
Chengdao Li https://orcid.org/0000-0002-9653-2700
Jacqueline Batley https://orcid.org/0000-0002-5391-5824
David Edwards https://orcid.org/0000-0001-7599-6760

REFERENCES

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., & Edwards, D. (2020). Plant pan-genomes are the new reference. *Nature Plants*, *6*, 914–920. https://doi.org/10.1038/s41477-020-0733-0

Bayer, P. E., Petereit, J., Durant, E., Monat, C., Rouard, M., Hu, H., Chapman, B., Li, C., Batley, J., & Edwards, D. (2022). Bread wheat genomes graph pangenome. https://doi.org/10.5281/zenodo.6085239

Bayer, P. E., Scheben, A., Golicz, A. A., Yuan, Y., Faure, S., Lee, H., Chawla, H. S., Anderson, R., Bancroft, I., Raman, H., Lim, Y. P., Robbens, S., Jiang, L., Liu, S., Barker, M. S., Schranz, M. E., Wang, X., King, G. J., Pires, J. C., … Edwards, D. (2021). Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal*, *19*, 2488–2500. https://doi.org/10.1111/pbi.13674

Berkman, P. J., Skarshewski, A., Lorenc, M. T., Lai, K., Duran, C., Ling, E. Y. S., Stiller, J., Smits, L., Imelfort, M., Manoli, S., McKenzie, M., Kubaláková, M., Šimková, H., Batley, J., Fleury, D., Doležel, J., & Edwards, D. (2011). Sequencing and assembly of low copy and genic regions of isolated Triticum aestivum chromosome arm 7DS. *Plant Biotechnology Journal*, *9*, 768–775. https://doi.org/10.1111/j.1467-7652.2010.00587.x

Berkman, P. J., Skarshewski, A., Manoli, S., Lorenc, M. T., Stiller, J., Smits, L., Lai, K., Campbell, E., Kubaláková, M., Šimková, H., Batley, J., Doležel, J., Hernandez, P., & Edwards, D. (2012). Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theoretical and Applied Genetics*, *124*, 423–432. https://doi.org/10.1007/s00122-011-1717-2

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., Goodstein, D. M., Elsik, C. G., Lewis, S. E., Stein, L., & Holmes, I. H. (2016). JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biology*, *17*, 66. https://doi.org/10.1186/s13059-016-0924-1

Donlin, M. J. (2009). Using the generic genome browser (GBrowse). *Current Protocols in Bioinformatics*, *28*, 9.9.1–9.9.25. https://doi.org/10.1002/0471250953.bi0909s28

Durant, É., Sabot, F., Conte, M., & Rouard, M. (2021). Panache: A web browser-based viewer for linearized pangenomes. *Bioinformatics*, *37*, 4556–4558. https://doi.org/10.1093/bioinformatics/btab688

Franco, J. A. V., Gage, J. L., Bradbury, P. J., Johnson, L. C., Miller, Z. R., Buckler, E. S., & Romay, M. C. (2020). A maize practical

haplotype graph leverages diverse NAM assemblies. *bioRxiv*, 2020. 2008.2031.268425. https://doi.org/10.1101/2020.08.31.268425

Gao, L., Koo, D. - H., Juliana, P., Rife, T., Singh, D., Lemes Da Silva, C., Lux, T., Dorn, K. M., Clinesmith, M., Silva, P., Wang, X., Spannagl, M., Monat, C., Friebe, B., Steuernagel, B., Muehlbauer, G. J., Walkowiak, S., Pozniak, C., Singh, R., … Poland, J. (2021). The Aegilops ventricosa 2NvS segment in bread wheat: Cytology, genomics and breeding. *Theoretical and Applied Genetics*, *134*, 529–542. https://doi.org/10.1007/s00122-020-03712-y

Garrison, E., & Guarracino, A. (2022). Unbiased pangenome graphs. bioRxiv, 2022.2002.2014.480413. https://doi.org/10.1101/2022.02.14.480413

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., Chan, C. K. K., Severn-Ellis, A., McCombie, W. R., Parkin, I. A. P., Paterson, A. H., Pires, J. C., Sharpe, A. G., Tang, H., Teakle, G. R., Town, C. D., Batley, J., & Edwards, D. (2016). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*, *7*, 13390. https://doi.org/10.1038/ncomms13390

Golicz, A. A., Bayer, P. E., Bhalla, P. L., Batley, J., & Edwards, D. (2020). Pangenomics comes of age: From bacteria to plant and animal applications. *Trends in Genetics*, *36*, 132–145. https://doi.org/10.1016/j.tig.2019.11.006

Guarracino, A., Heumos, S., Nahnsen, S., Prins, P., & Garrison, E. (2021). ODGI: Understanding pangenome graphs. bioRxiv, 2021.11.10.467921. https://doi.org/10.1101/2021.11.10.467921

Hickey, G., Heller, D., Monlong, J., Sibbesen, J. A., Sirén, J., Eizenga, J., Dawson, E. T., Garrison, E., Novak, A. M., & Paten, B. (2020). Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biology*, *21*, 35. https://doi.org/10.1186/s13059-020-1941-7

Hurgobin, B., & Edwards, D. (2017). SNP discovery using a pangenome: Has the single reference approach become obsolete? *Biology*, *6*, 21. https://doi.org/10.3390/biology6010021

The International Wheat Genome Sequencing Consortium (IWGSC). (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, *345*, 125178. https://doi.org/10.1126/science.1251788

The International Wheat Genome Sequencing Consortium (IWGSC). (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*, eaar7191. https://doi.org/10.1126/science.aar7191

Jensen, S. E., Charles, J. R., Muleta, K., Bradbury, P. J., Casstevens, T., Deshpande, S. P., Gore, M. A., Gupta, R., Ilut, D. C., Johnson, L., Lozano, R., Miller, Z., Ramu, P., Rathore, A., Romay, M. C., Upadhyaya, H. D., Varshney, R. K., Morris, G. P., Pressoir, G., … Ramstein, G. P. (2020). A sorghum practical haplotype graph facilitates genome-wide imputation and cost-effective genomic prediction. *The Plant Genome*, *13*, e20009. https://doi.org/10.1002/tpg2.20009

Keilwagen, J., Lehnert, H., Berner, T., Badaeva, E., Himmelbach, A., Börner, A., & Kilian, B. (2022). Detecting major introgressions in wheat and their putative origins using coverage analysis. *Scientific Reports*, *12*, 1908. https://doi.org/10.1038/s41598-022-05865-w

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109

Lai, K., Lorenc, M. T., Lee, H. C., Berkman, P. J., Bayer, P. E., Visendi, P., Ruperao, P., Fitzgerald, T. L., Zander, M., Chan, C. - K. K., Manoli, S., Stiller, J., Batley, J., & Edwards, D. (2015). Identification and characterization of more than 4 million intervarietal SNP s across the group 7 chromosomes of bread wheat. *Plant Biotechnology Journal*, *13*, 97–104. https://doi.org/10.1111/pbi.12240

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Feng, X., & Chu, C. (2020). The design and construction of reference pangenome graphs with minigraph. *Genome Biology*, *21*, 265. https://doi.org/10.1186/s13059-020-02168-z

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C. - K. K., Visendi, P., Lai, K., Doležel, J., Batley, J., & Edwards, D. (2017). The pangenome of hexaploid bread wheat. *The Plant Journal*, *90*, 1007–1013. https://doi.org/10.1111/tpj.13515

Obembe, O. S., Hendricks, N. P., & Tack, J. (2021). Decreased wheat production in the USA from climate change driven by yield losses rather than crop abandonment. *Plos One*, *16*, e0252067. https://doi.org/10.1371/journal.pone.0252067

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*, 132. https://doi.org/10.1186/s13059-016-0997-x

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rijzaani, H., Bayer, P. E., Rouard, M., Doležel, J., Batley, J., & Edwards, D. (2021). The pangenome of banana highlights differences between genera and genomes. *The Plant Genome*, *15*, e20100. https://doi.org/10.1002/tpg2.20100

Ruperao, P., Thirunavukkarasu, N., Gandham, P., Selvanayagam, S., Govindaraj, M., Nebie, B., Manyasa, E., Gupta, R., Das, R. R., Odeny, D. A., Gandhi, H., Edwards, D., Deshpande, S. P., & Rathore, A. (2021). Sorghum pan-genome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain. *Frontiers in Plant Science*, *12*, 963. https://doi.org/10.3389/fpls.2021.666342

Sears, E., & Miller, T. (1985). The history of Chinese Spring wheat. *Cereal Research Communication*, *13*, 261–263.

Sirén, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C., Sibbesen, J. A., Hickey, G., Chang, P.-C., Carroll, A., Gupta, N., Gabriel, S., Blackwell, T. W., Ratan, A., Taylor, K. D., Rich, S. S., Rotter, J. I., Haussler, D., Garrison, E., & Paten, B. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science*, *374*(6574), abg8871. https://doi.org/10.1126/science.abg8871

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., & Zhou, R. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants*, *6*, 34–45. https://doi.org/10.1038/s41477-019-0577-7

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., Ramirez-Gonzalez, R. H., Kolodziej, M. C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J. N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., … Pozniak, C. J. (2020). Multiple wheat genomes reveal global

variation in modern breeding. *Nature*, *588*, 277–283. https://doi.org/10.1038/s41586-020-2961-x

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, *18*, 1946–1954. https://doi.org/10.1093/bioinformatics/btv383

Zhao, J., Bayer, P. E., Ruperao, P., Saxena, R. K., Khan, A. W., Golicz, A. A., Nguyen, H. T., Batley, J., Edwards, D., & Varshney, R. K. (2020). Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnology Journal*, *31*, 3350–3352. https://doi.org/10.1111/pbi.13354

Zimin, A. V., Puiu, D., Hall, R., Kingan, S., Clavijo, B. J., & Salzberg, S. L. (2017). The first near-complete assembly of the hexaploid bread wheat genome, Triticum aestivum. *Gigascience*, *6*, gix097. https://doi.org/10.1093/gigascience/gix097

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bayer, P. E., Petereit, J., Durant, É., Monat, C., Rouard, M., Hu, H., Chapman, B., Li, C., Cheng, S., Batley, J., & Edwards, D. (2022). Wheat Panache: A pangenome graph database representing presence–absence variation across sixteen bread wheat genomes. *The Plant Genome*, *15*, e20221. https://doi.org/10.1002/tpg2.20221