Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio

Presented at the 2020 Plant and Animal Genome Conference Author: Shawn Quinn (CTO, Curio Genomics)

CURIO



Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio

Presentation Overview

- What is Curio
- Sample Analysis with Chinese Spring Wheat
 - Read Alignment and Visualization
 - Handling Large Chromosomes
 - Mapping Algorithms & Coverage
 - Variant Analysis
 - Expression Analysis
- Looking Ahead
- Acknowledgements



What is Curio?

- Modern big data management and genomic analysis platform, fully web-based, collaboration ready
- Supports <u>both</u> bioinformatic processing and scientific interpretive analysis
- Provides scalable data processing and interactive data visualizations using real-time databases and clustering technologies
- Designed for extensibility to continuously support new analysis methods, data types, etc.
- Includes complex crop research solutions, including tetraploid and hexaploid wheat DNA-Seq and RNA-Seq analysis





Bread Wheat DNA-Seq: Read Mapping and Navigation

Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

©2020, Curio Genomics

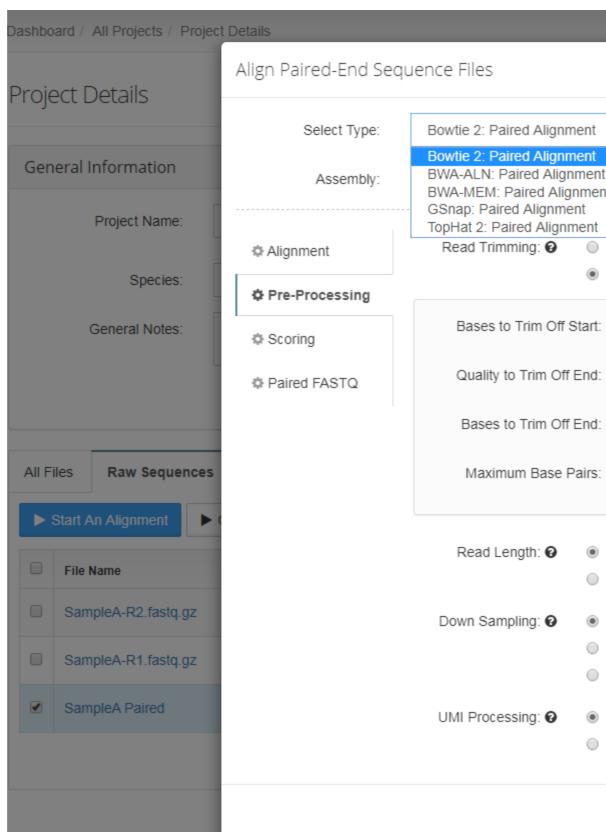
Incorporates Chinese Spring • Wheat (*Triticum aestivum*) reference assembly from the IWGSC

Dashbo	ard / All Projects / Project	Details				
Proie	ect Details	Align Paired-End Sec	quence Files		×	More Options -
			Bowtie 2: Paired Alignn	nent	•	
Gen	eral Information	Assembly:	IWGSC WGA 1.0 (Chines	e Spring Wheat)		
Project Name:		Alignment	Read Trimming: 🕑	Do not perform any read trimm	nina	
	Species:	Pre-Processing		 Trim the reads before aligning 		
General Notes:		Scoring	Read Length: 🕑	 Process reads of any length Discard short reads 		
		Paired FASTQ	Down Sampling: 🛛	 Process all reads Process a sampling (%) of the 	reads	
				 Process a fixed number of read 		
All F	iles Raw Sequences		UMI Processing: 🛛	Reads do not contain UMI/UM	Ts	
	Start An Alignment			Reads contain unique molecul	ar ids/tags	Create Paired FASTQ File
	File Name			Cancel	Start Alignment	
	SampleA-R2.fastq.gz		NUIC		2013-12-00 17:23	► Start an Alignment -
	SampleA-R1.fastq.gz		2 Comple	eted,	2019-12-06 17:23	► Start an Alignment -
SampleA Paired 16 Co		16 Compl	leted,	2019-12-06 17:00	► Start an Alignment -	

Bread Wheat: Read Mapping

Viewing 1-3

- **Incorporates Chinese Spring** • Wheat (*Triticum aestivum*) reference assembly from the **IWGSC**
- Multiple read mapping • algorithms with pre-built indexes that are deployed and ready on a computational cluster
- Experiment with various • alignment and read processing options without requiring any pipeline configuration

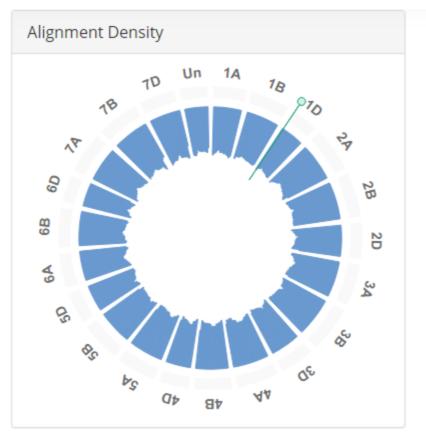


Bread Wheat: Read Mapping

			_	
		×		
				More Options 🗸
t		•		
t				
nt ent				
t) Dor	not perform any read trimming			
	the reads before aligning			
t:	0	?		
1:	Don't trim end of reads	?		
d:	0	?		
S:	All Bases	?		
				create Paired FASTQ File
	ess reads of any length		ed	
) Disc	ard short reads			
Pro	ess all reads		7:23	► Start an Alignment -
) Proc	cess a sampling (%) of the reads		7:23	► Start an Alignment -
Process a fixed number of reads			1.25	Start an Aighment
Reads do not contain UMI/UMTs			7:00	► Start an Alignment -
Reads contain unique molecular ids/tags				
				Viewing 1-3
	Cancel Start Align	iment		

Bread Wheat: Alignment Visualization

 Quickly browse and visualize reads from samples of any size, anywhere in the genome

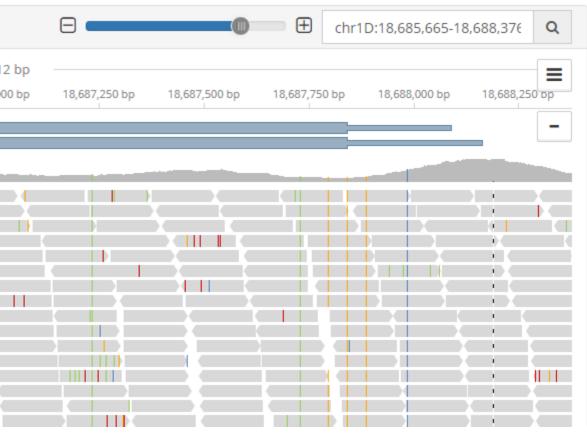


Display Options N						
Show gene	and exon					
Show refere	ence geno					
Show nucle differences						
Exclude real	ds and ba					
Group reads	-					
Remove real	d duplicat					

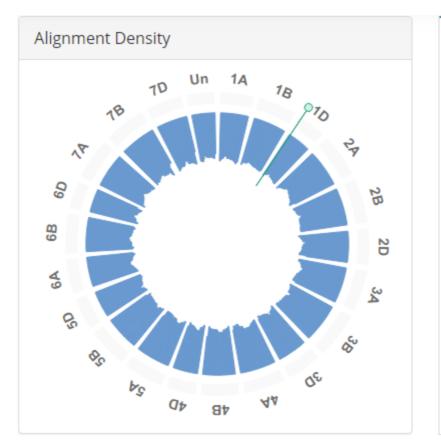
position and CIGAR

 2,712 18,685,750 bp 18,686,000 bp 18,686,250 bp 18,686,500 bp 18,686,750 bp 18,686,750 bp 18,687,00 	Reads (Showir	ng 583 alignments of 7	77, excluded 1 of lov	v quality, removed	d 615 duplica	tes)	
	< 18,685,750 bp	18,686,000 bp	18,686,250 bp	18,686,500	bp 18,	586,750 bp	

otes and Marks Sharing	
a positions from IWGSC (HC) ome bases erage and highlight when 0% ases with quality under 90% (perfect matches), requiring ates at the same alignment	 Show individual read details Show read consensus sequence information Highlight mismatched bases Highlight read insertions Color forward and reverse reads differently

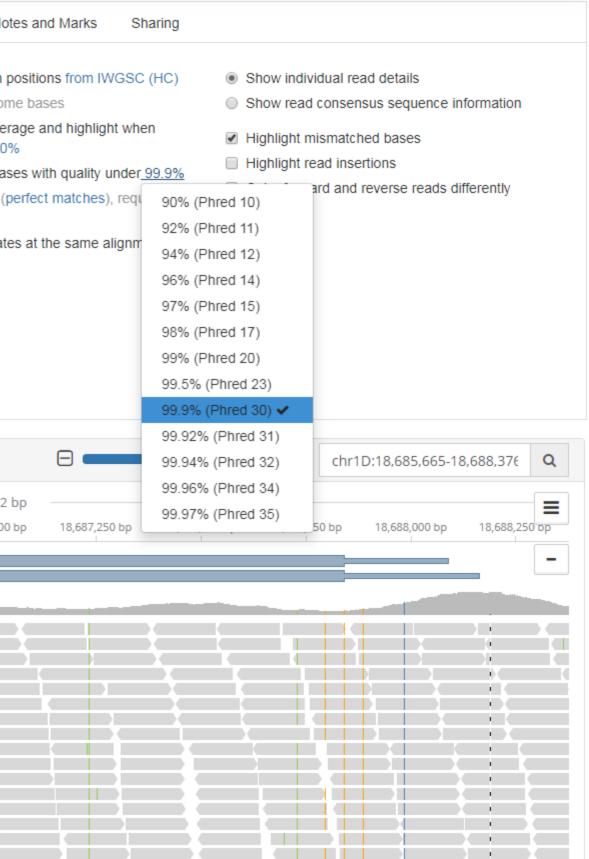


- Quickly browse and visualize • reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, • etc. on the fly

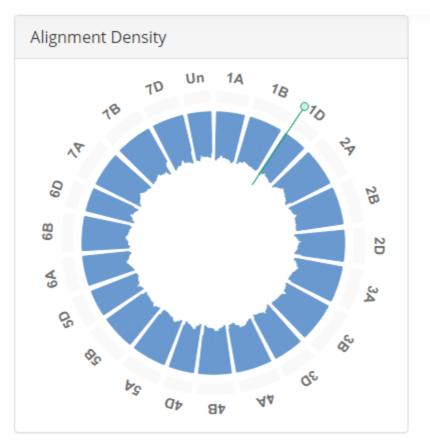


Di	splay Options	No
1	Show gene and	exon p
1	Show reference	genon
1	Show nucleotide differences exce	
1	Exclude reads a	nd bas
	Group reads by consensus of 60	
1	Remove read du position and CIG	

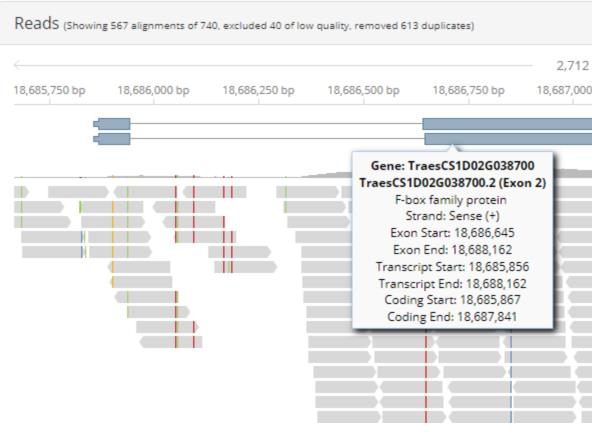
Reads (Showin	g 567 alignments of 7	40, excluded 40 of lov	v quality, removed 61	3 duplicates)	
(18,685,750 bp	18,686,000 bp	18,686,250 bp	18,686,500 bp	18,686,750 bp	2,712



- Quickly browse and visualize • reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, • etc. on the fly
- Access both the transcript and • functional annotations from the IWGSC inline

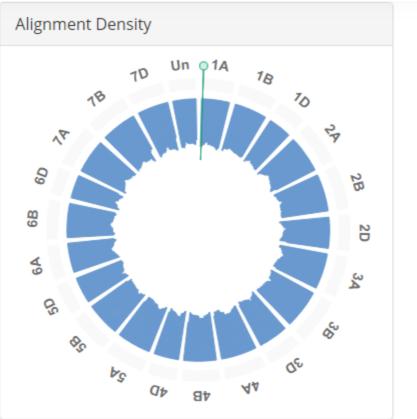


Display Options	Notes and Marks Sharing		
 Show gene and Show reference Show nucleotide differences exce 	coverage and IWGSC Low Conf	● Show individual fidence Genes (1.1)	I read details Isus sequence information hed bases
 Exclude reads and bases with quality under 99.9% Group reads by UMI (perfect matches), requiring consensus of 60% 		 Highlight read in Color forward and 	nsertions nd reverse reads differently
Remove read du position and CIG	iplicates at the same alignment GAR		



			Ð	chr1D:18,685	,665-18,688,37€	Q
2 bp)0 bp	18,687,250 bp	18,687,500 bp	18,687,75	0 bp 18,688,0 '	00 bp 18,688,2	50 pp

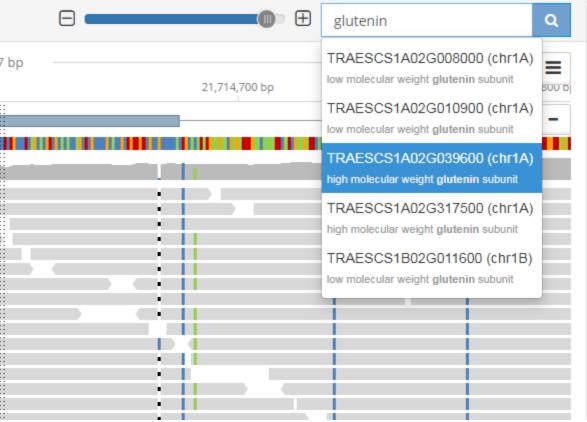
- Quickly browse and visualize • reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, • etc. on the fly
- Access both the transcript and • functional annotations from the IWGSC inline
- Quickly search and navigate • the genome using the IWGSC annotation data as well



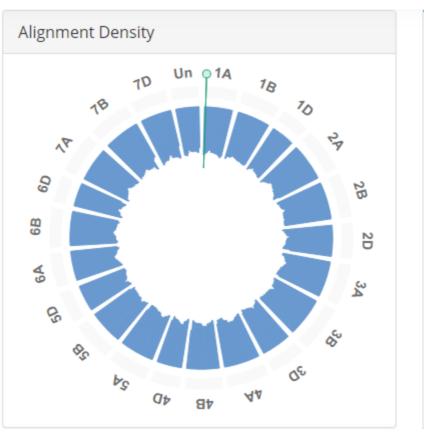
Display Options	No
Show gene and	
 Show reference Show nucleotide differences exce 	cover
Exclude reads and	
Group reads by consensus of 60	
Remove read du	plicate

Reads (Showing 83 alignments, excluded 17 of low quality, removed 49 duplicates)						
<	377 k					
21,714,500 bp 21,714,600 bp						

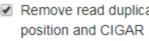
play Options	Notes and Marks	Sharing	
Show gene and Show reference	exon positions from IW	GSC (HC)	 Show individual read details Show read consensus sequence information
Show nucleotide differences exce	e coverage and highlight ed 20%	when	Highlight mismatched bases
Exclude reads and bases with quality under 99.9% Group reads by UMI (perfect matches), requiring consensus of 60%			Highlight read insertions
			Color forward and reverse reads differently
Remove read du position and CIG	iplicates at the same ali GAR	gnment	



- Quickly browse and visualize • reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, etc. on the fly
- Access both the transcript and • functional annotations from the IWGSC inline
- Quickly search and navigate • the genome using the IWGSC annotation data as well
- Conveniently access reference • or consensus sequences, and read coverage details

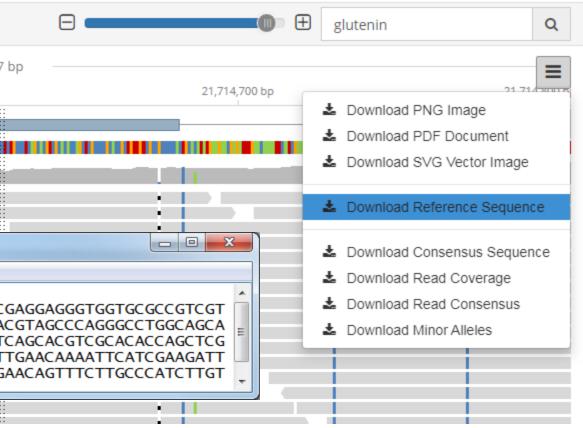


Di	splay Options	Not
1	Show gene and	exon p
1	Show reference	genom
1	Show nucleotide differences exce	
1	Exclude reads a	nd bas
	Group reads by consensus of 60	
	Domovo rood du	nlianta

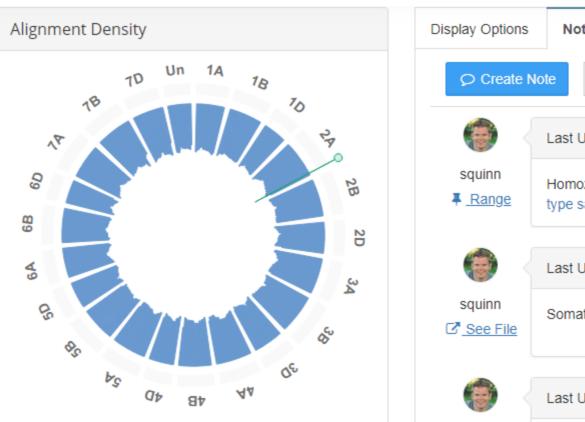


Reads (Showing 83 alignments, excluded 17 of low quality, removed 49 duplicates)							
<i>(</i>		377					
	21,714,500 bp	21,714,600 bp					
	, 						
_							
	CSW10_chr1A_21714434-21714810.fasta - Notepad						
	<u>File Edit Format View Help</u>						
	<pre>bchr1A:21714434-21714810 AGGAGAAGGAGGAGGAGGATCTATGGTAGCCGGCCG</pre>	TGTCGCTCTTGTGCAGGC					
	CGTTGCAGCAGTAGCAGACGTCGCCGCGGACGA	GATTGCAGCGGCCATGGA					
	TCGCATGCACTTGCCCGTCCCTTGCCAGCACTT CCCCGCACCGCTGTTGTCGCCGCCCCCACTGCA						
	CACTAGTTGGCTCTTATCGAACACTTACGGATG						

 Show individual read details Show read consensus sequence information Highlight mismatched bases Highlight read insertions Color forward and reverse reads differently



- Quickly browse and visualize • reads from samples of any size, anywhere in the genome
- Adjust filters, PCR processing, • etc. on the fly
- Access both the transcript and • functional annotations from the IWGSC inline
- Quickly search and navigate • the genome using the IWGSC annotation data as well
- Conveniently access reference • or consensus sequences, and read coverage details
- Navigate and collaborate through annotated comments



ReadS (Showing 295 alignments, excluded 28 of low quality) 62,285,700 bp 762.285.800 bp

Sharing				
ents	Q		\$ •	_
c 20 15:45:54. Posted 2019 D	ec 20 14:42:57	Ø	<u>ش</u> (£	
own here in this Mut5 sample	appear to be heterozy	ygous in t	the wild	
c 20 14:49:15. Posted 2019 D	ec 20 14:48:24	Ø	<u>ش</u>	
2A of the Mut5 and Mut7 sam	ples.			
c 20 15:06:26. Posted 2019 D	ec 20 15:05:36	Ø	<u>ش</u>	•
		,692-762	,285,9	Q
762,285,9	900 bp			-
I	I			
	ents 20 15:45:54. Posted 2019 D own here in this Mut5 sample 20 14:49:15. Posted 2019 D 2A of the Mut5 and Mut7 sam 20 15:06:26. Posted 2019 D 762,285,9 762,285,9	ents Q 20 15:45:54. Posted 2019 Dec 20 14:42:57 own here in this Mut5 sample appear to be heterozy 20 14:49:15. Posted 2019 Dec 20 14:48:24 2A of the Mut5 and Mut7 samples. 20 15:06:26. Posted 2019 Dec 20 15:05:36 (m) (m) (m) (m) <	ents Q 20 15:45:54. Posted 2019 Dec 20 14:42:57 won here in this Mut5 sample appear to be heterozygous in the sample appeare to be heterozygous in theterozygous i	ents Q

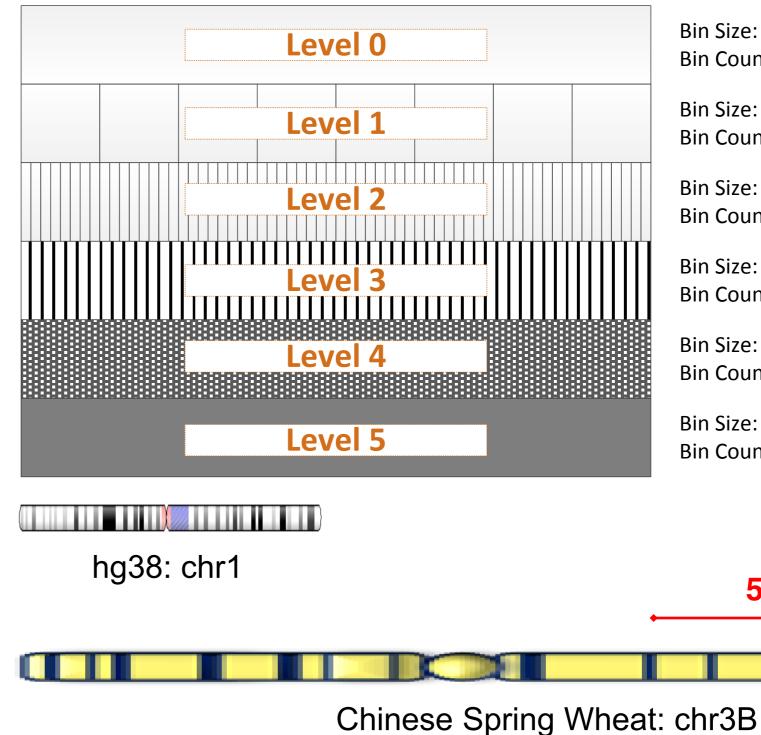


Handling Large Chromosomes

Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

Standard Bin Strategy

- Binning strategy used in BAM index files (i.e. a "BAI" file)
- 6 levels deep and maximum bin size of: 536 million bases
- Largest human chromosome (chr1): 249 million bases
- Largest CSW chromosome (chr3B): 837 million bases
- Forces a split chromosome approach or causes various tool compatibility issues



Bin Size: 536,870,912 bps Bin Count: 1

Bin Size: 67,108,864 bps Bin Count: 8

Bin Size: 8,388,608 bps Bin Count: 64

Bin Size: 1,048,576 bps Bin Count: 512

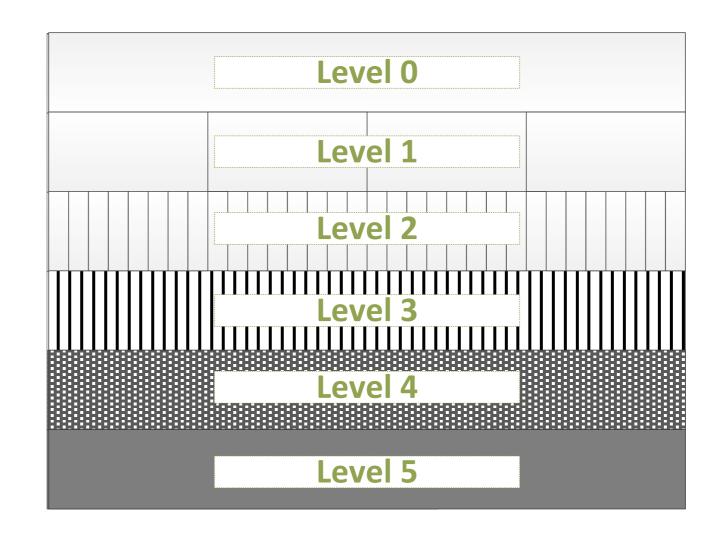
Bin Size: 131,072 bps Bin Count: 4,096

Bin Size: 16,384 bps Bin Count: 32,768

56% too large

A Better Approach

- Dynamic binning strategy based on species & assembly
- For wheat utilize a strategy • based on the coordinate-sorted index specification with a minimum bit shift of 16
- Maintain 6 levels of depth & the number of bins per level, but use the first half of each level
- Max bin size: **1 billion bases**
- Largest CSW chromosome (chr3B): 837 million bases





Chinese Spring Wheat: chr3B

Bin Size: 1,073,741,824 bps Bin Count: 1

Bin Size: 268,435,456 bps Bin Count: 4 of 8

Bin Size: 33,554,432 bps Bin Count: 32 of 64

Bin Size: 4,194,304 bps Bin Count: 256 of 512

Bin Size: 524,288 bps Bin Count: 2,048 of 4,096

Bin Size: 65,536 bps Bin Count: 16,384 of 32,768

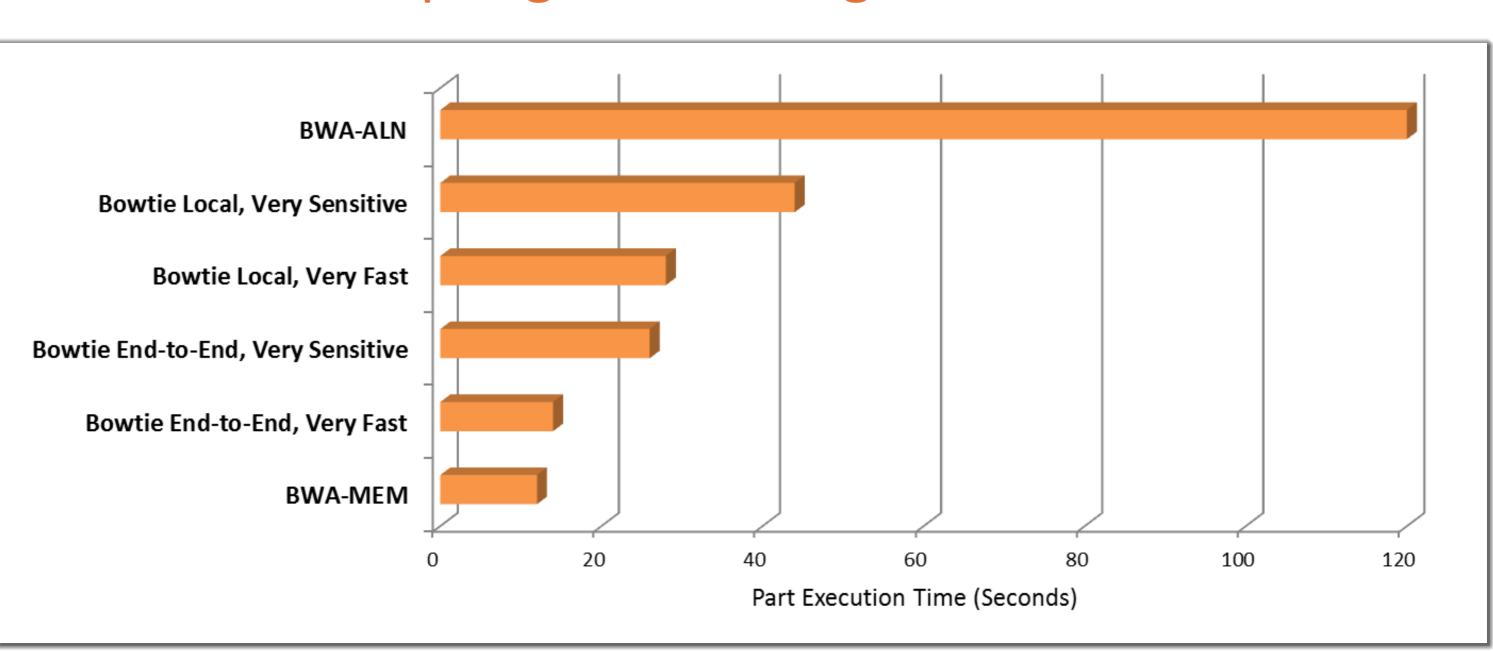


DNA-Seq Bread Wheat: Alignment Algorithm Coverage Impacts

Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020



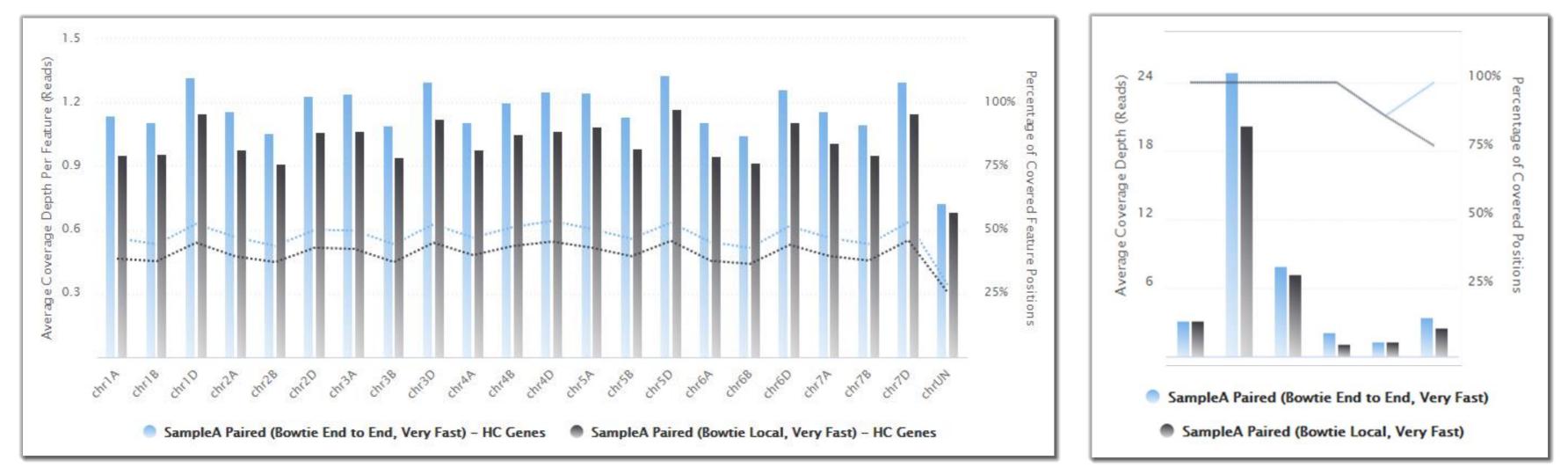
Chinese Spring Wheat: Aligner Performance



Note: the time needed for both the "bwa aln" and "bwa sampe" commands are included in the BWA-ALN metric shown above.



Bowtie2 "End to End" vs "Local" Algorithm, Exome Coverage



Average Exome Coverage Across all Chromosomes

©2020, Curio Genomics

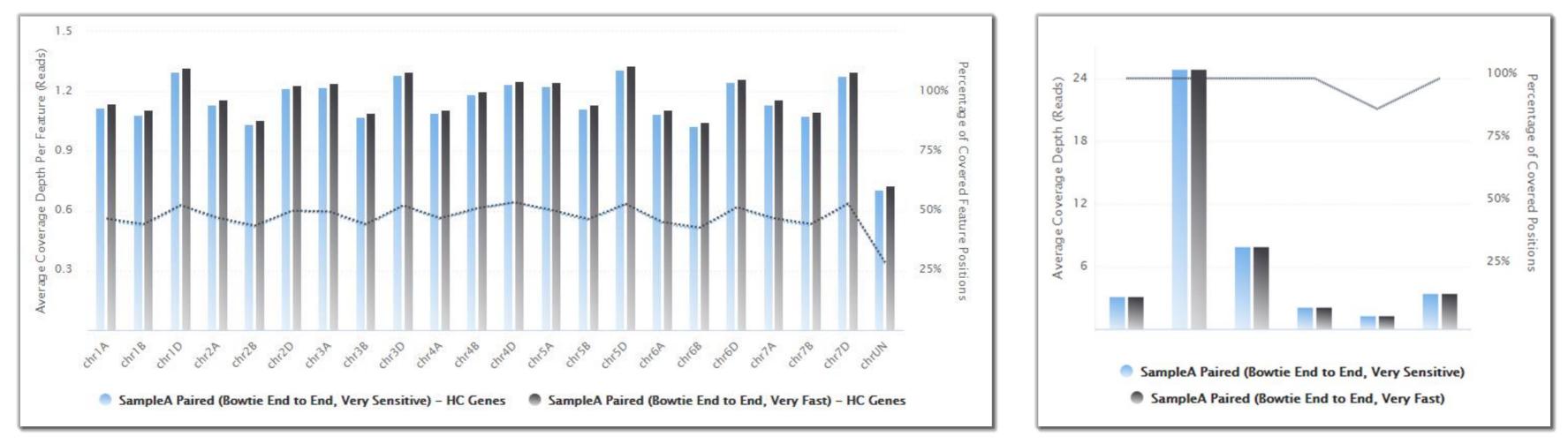
Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

Exons of Single Gene

Note: L-lactate dehydrogenase gene (TRAESCS1A02G238700) shown here



Bowtie2 "Very Sensitive" vs "Very Fast" Presets, Exome Coverage



Average Exome Coverage Across all Chromosomes

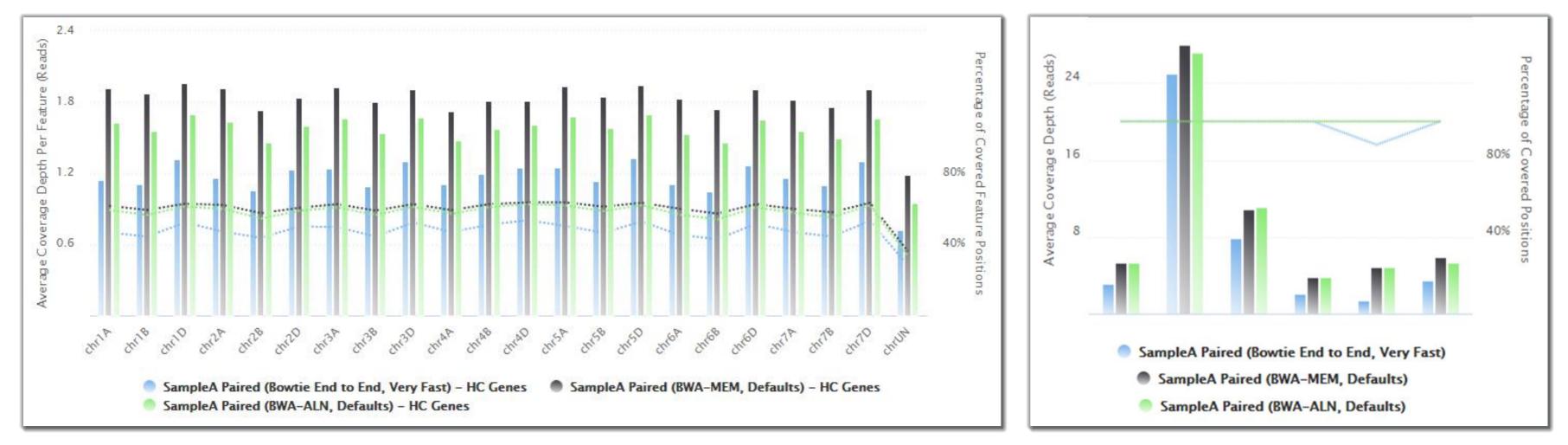
Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

Exons of Single Gene

Note: L-lactate dehydrogenase gene (TRAESCS1A02G238700) shown here



Aligner Algorithms, IWGSC "HC" Genes, Default Quality Filtering



Average Exome Coverage Across all Chromosomes

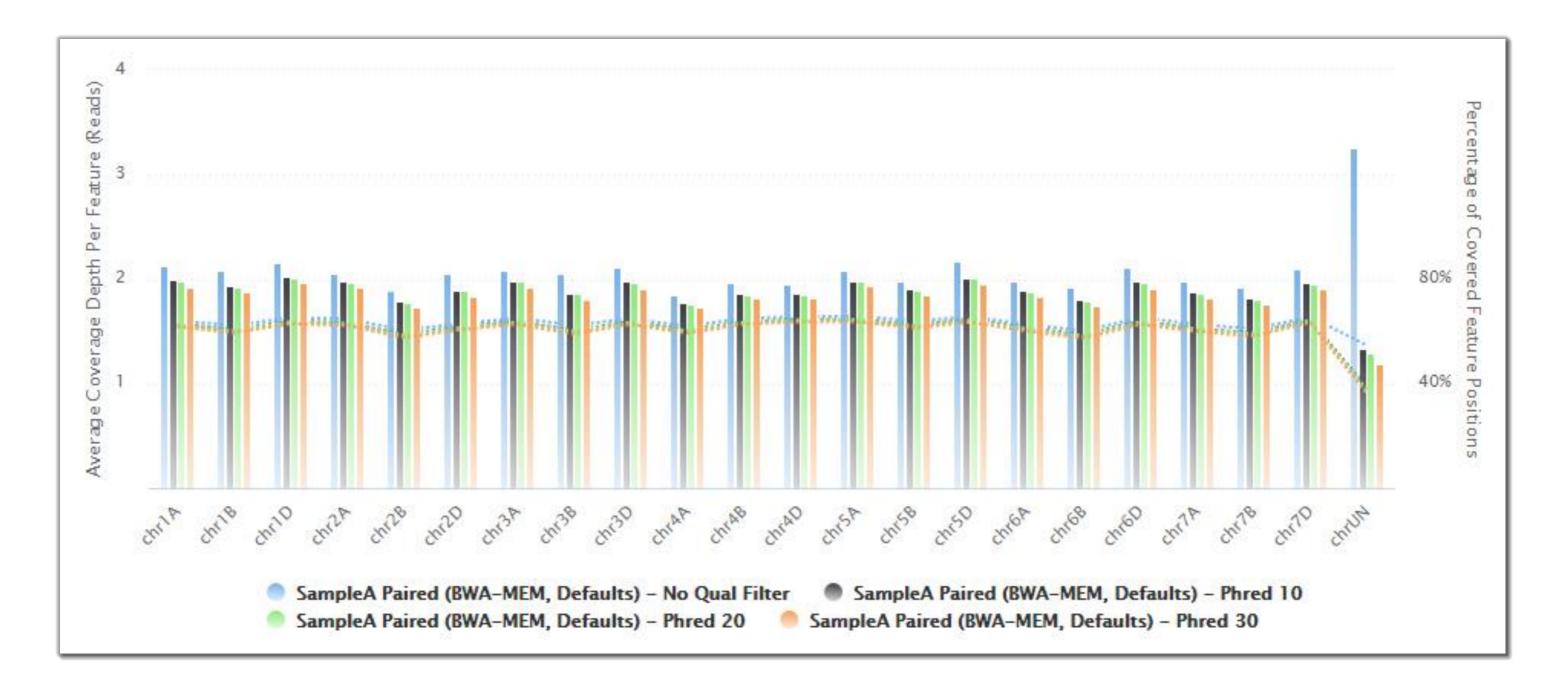
Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

Exons of Single Gene

Note: L-lactate dehydrogenase gene (TRAESCS1A02G238700) shown here



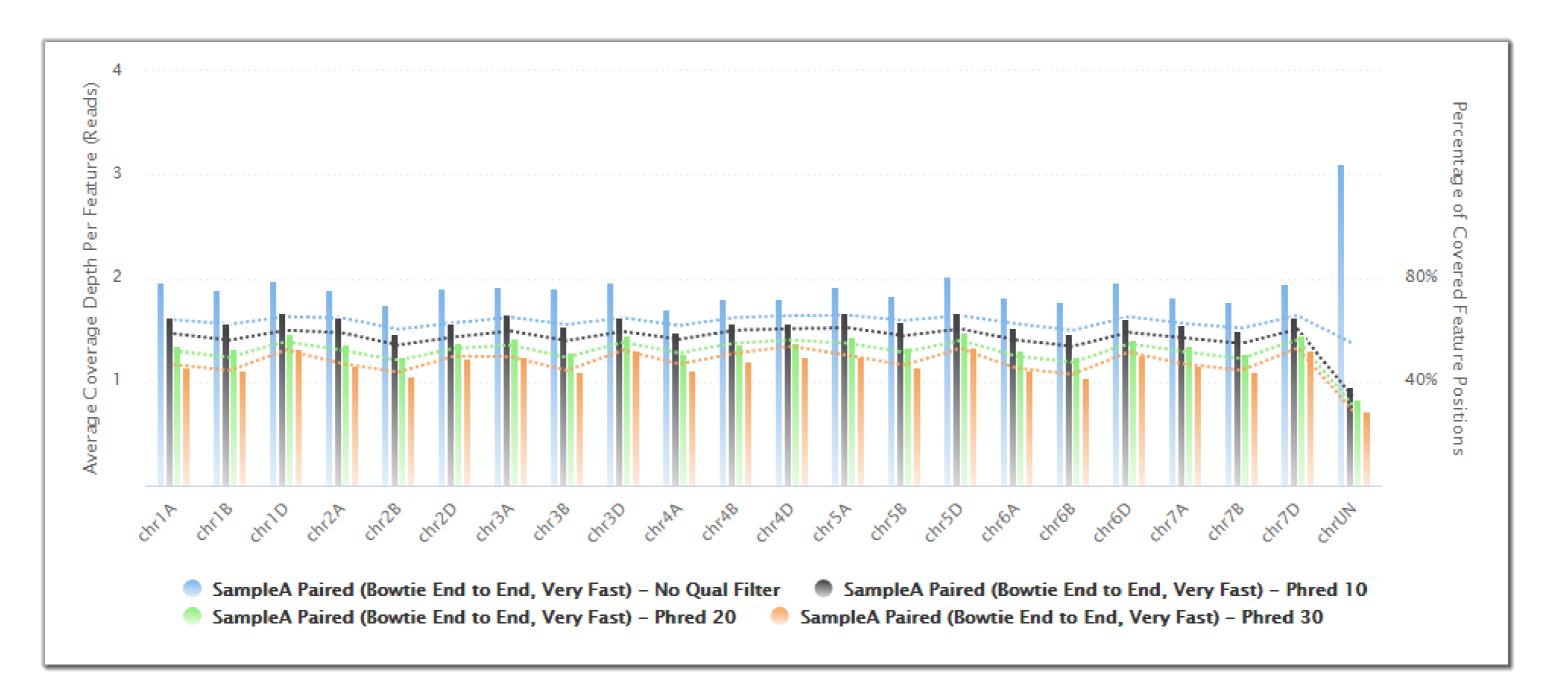
BWA-MEM Read Mapping Quality Impact on Coverage



Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

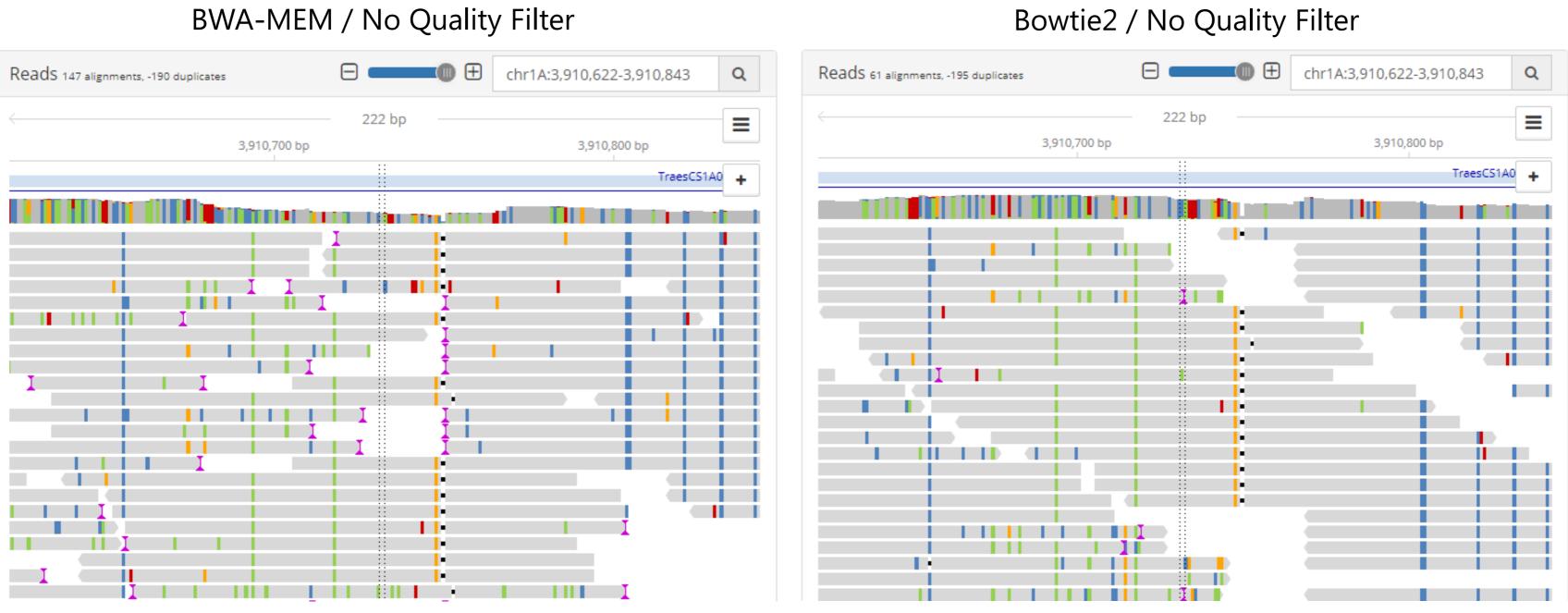


Bowtie2 Read Mapping Quality Impact on Coverage



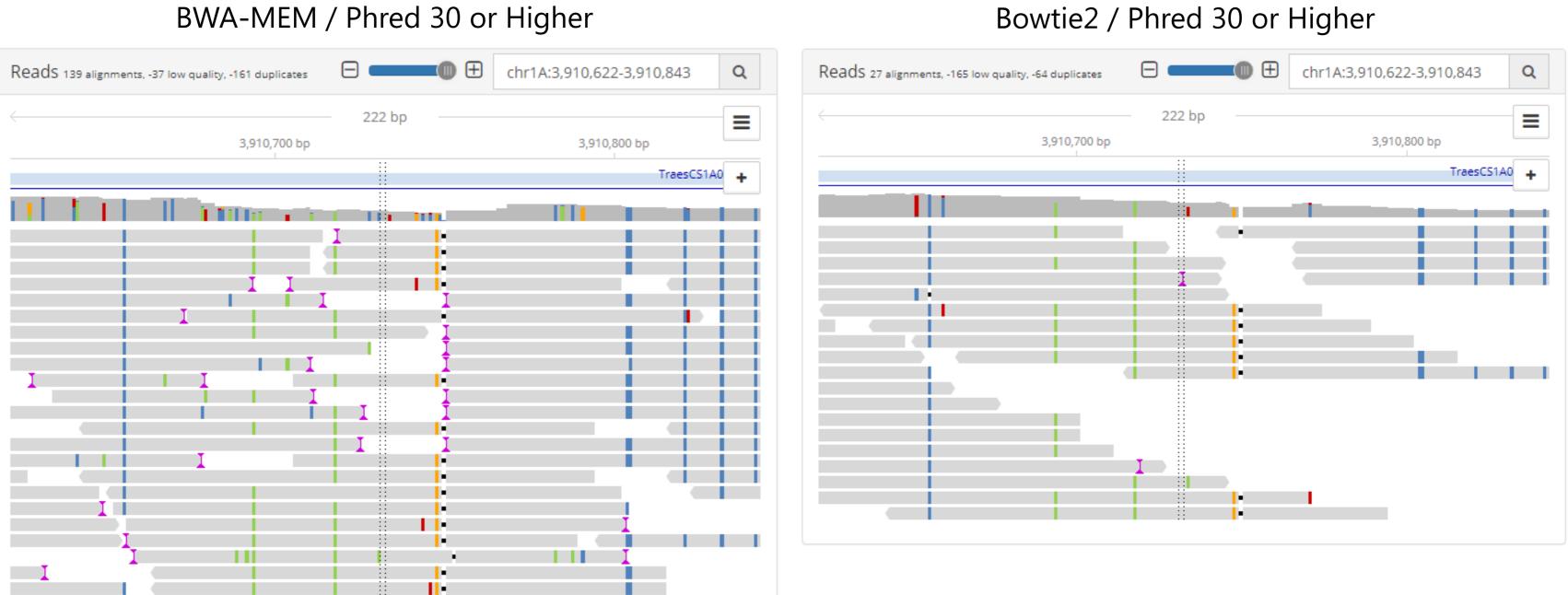
Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

- BWA-MEM and Bowtie2 with no quality filter applied
- Visualizing low quality base calls within aligned reads •



Bread Wheat: Coverage Analysis

- BWA-MEM and Bowtie2 with a quality filter set to 99% (i.e. Phred 30)
- Visualizing impacts of different mapping algorithm approaches to alignment quality •

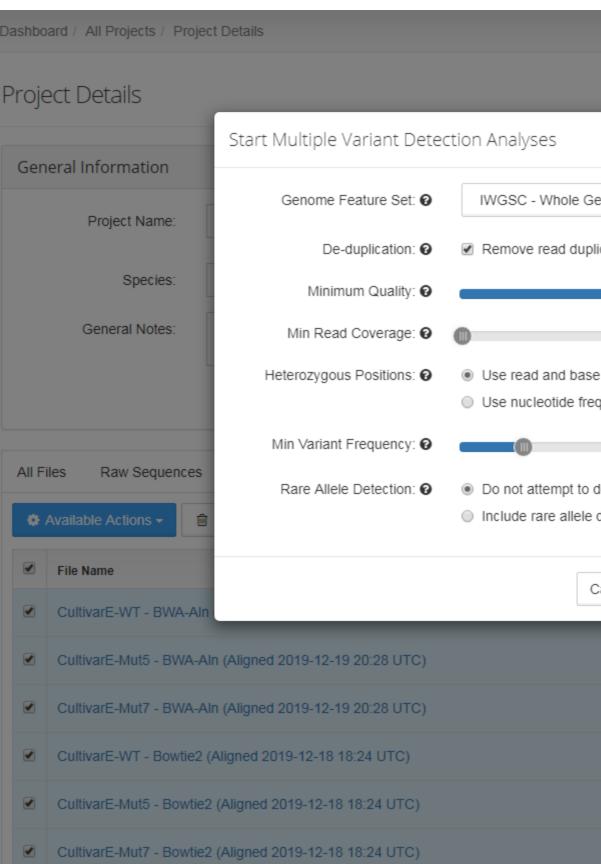


Bread Wheat: Coverage Analysis



DNA-Seq Bread Wheat: Variant Analysis

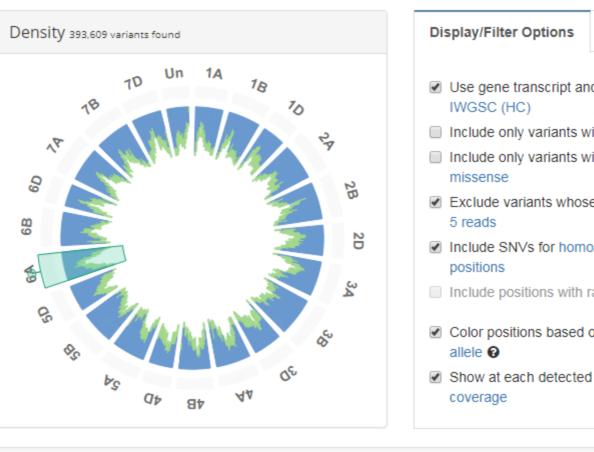
- Call variants leveraging • the Chinese Spring Wheat reference assembly and related annotations from the IWGSC
- Operate on one or • multiple samples of any read depth simultaneously
- Conveniently adjust for • sensitivity and specificity

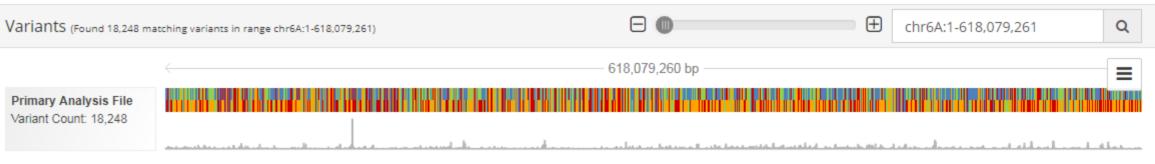


Bread Wheat: Variant Analysis

		More Options -
	×	
Genome - Chinese Spring Wheat (v1.1)	_	
Senome - Chinese Spring Wheat (V1.1)		
plicates at the same position and CIGAR		
Phred:	30 (99.9%)	
10	Reads	
se quality data in heterozygous calls equencies only in heterozygous calls		
	10%	
o detect rare alleles		
e detection in the analysis		
Cancel Start Variant Detection A	nalysis	
	D 05:59	Lind View Alignments 👻
224,364,008	2019-12-20 05:50	Ltd View Alignments 👻
186,030,770	2019-12-20 04:08	Ltd View Alignments 👻
237,564,990	2019-12-18 21:55	Lint View Alignments 👻
224,364,008	2019-12-18 17:01	Land View Alignments -
186,030,770	2019-12-18 16:21	Lind View Alignments -

Browse and visualize variants • across the entire CSW genome Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only





Bread Wheat: Variant Analysis

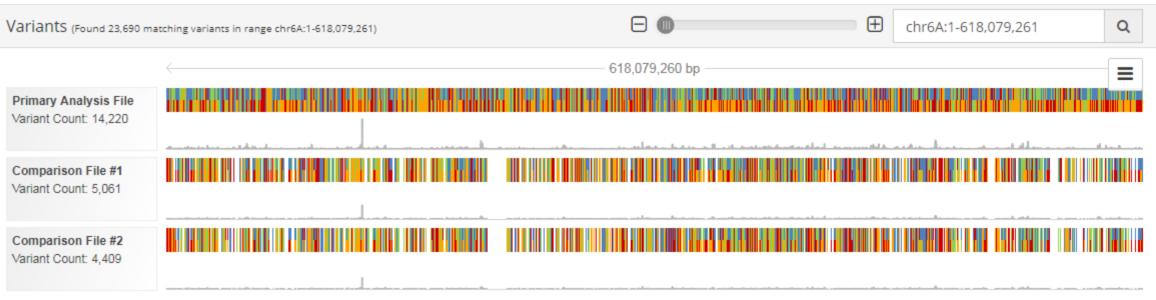
irks Sharing
Show only variants not present in the control (i.e. somatic)
 Show only variants with control coverage of at least 5 reads
Match homozygous SNVs to related heterozygous SNVs in control
 Show all variants found within selected sample files Show only variants common to all selected sample files
Show file coverage track of variants detected per position

Bread Wheat: Variant Analysis

- Browse and visualize variants • across the entire CSW genome
- Simultaneously compare and • visualize multiple samples

Density 393,609 variants found Display/Filter Options 7D Type here 1_B 18 Туре N Q_{9} 2B 1 Sample • 6B 20 1 Sample • ŝ С Sample • 10 **4**B





Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

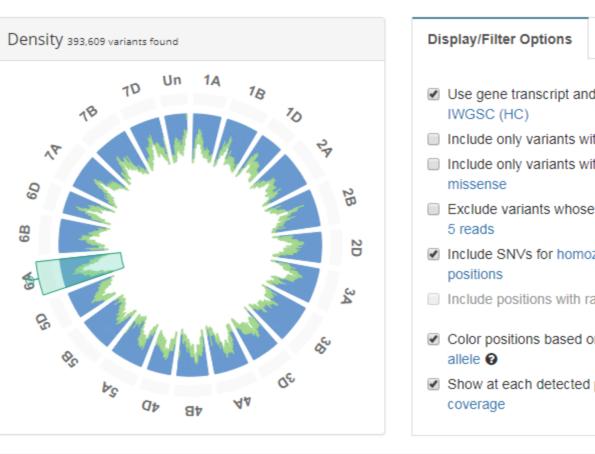
More Options -

Compare	Notes and Marks Sharing		
	Q		
File Name		Records	Last Updated
cultivarE-Mut5 - I 5:03 UTC)	BWA-Aln (Analyzed 2019-12-20	376,711	2019 December 20, Friday 10:56:25 UTC-5
cultivarE-Mut7 - 5:03 UTC)	BWA-Aln (Analyzed 2019-12-20	306,472	2019 December 20, Friday 10:55:12 UTC-5
cultivarE-WT - B 6:23 UTC) - Me	WA-Aln (Analyzed 2019-12-20 dium Sensitivity	5,859,463	2019 December 20, Friday 11:41:27

Viewing 1-8 of 8

Bread Wheat: Variant Analysis

- Browse and visualize variants • across the entire CSW genome
- Simultaneously compare and • visualize multiple samples
- Use one or more samples as a • control to exclude variants present in the wild type



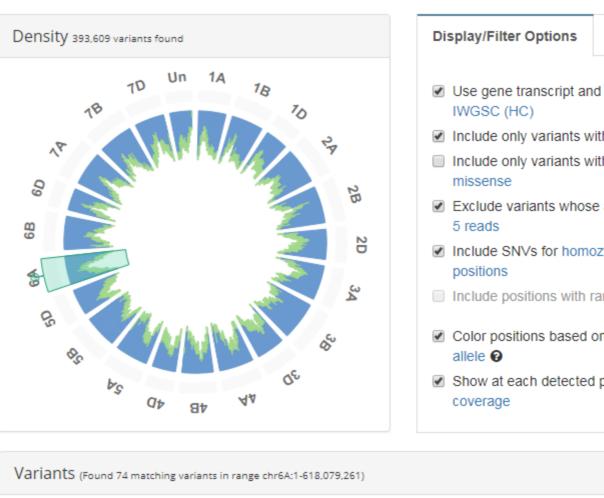
Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only



;	Compare	Notes and Mar	rks	Sharing
an	d exon annotatio	ns from	•	Show only variants not present in the control (i.e. somatic)
	ithin known codi			Show only variants with control coverage of at least 5 reads
	e alt coverage is		•	Match homozygous SNVs to related heterozygous SNVs in control 🚱
m	ozygous or heter	ozygous	•	Show all variants found within selected sample files Show only variants common to all selected sample files
	are alleles	Iternete		Show file coverage track of variants detected per position
	on the variant's a I position the vari			Postaer
teo	position the vari	iant's		

Bread Wheat: Variant Analysis

- Browse and visualize variants • across the entire CSW genome
- Simultaneously compare and • visualize multiple samples
- Use one or more samples as a • control to exclude variants present in the wild type
- Focus on genomic regions of • interest based on IWGSC gene transcript annotations



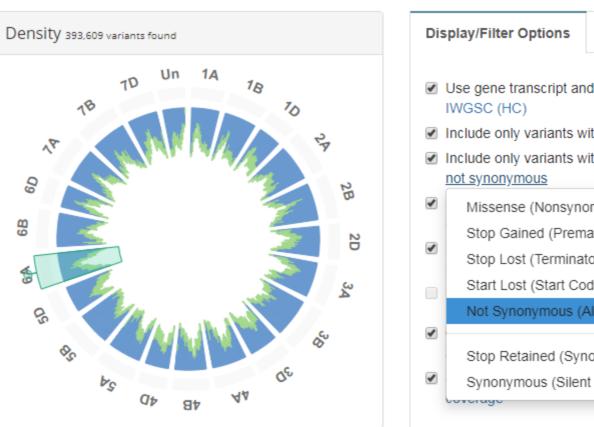


Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

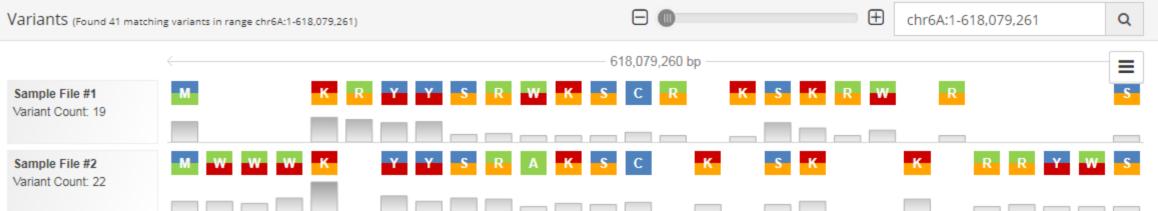
	Compare	Notes and Marks	Sharing
	id exon annot		Show only variants not present in the control (i.e. somatic)
5 W	/ithin known <u>c</u>	oding regions 🛛 🖉	Show only variants with control coverage of at least
5 W	ith predicted	Exonic Regions	
		Intronic Regions	zygous SNVs to related heterozygous trol 🚱
os	e alt covera	Coding Regions 🗸	
m/	ozygous or t	Intragenic Regions	iants found within selected sample files
III	bzygous of f	Untranslated Regions	ariants common to all selected sample s (UTR)
h ı	rare alleles	5' UTRs	
		3' UTRs	verage track of variants detected per
ed on the varia		Splice Regions	
		Upstream/Downstrear	m Regions
tec	position the	Upstream Regions	in regions
	_	Downstream Regions	>
		Intergenic Regions	t chr6A:1-618,079,261 Q
			CHIGA. 1-616,079,261
		60 bp	
2 (C R R Y	K R S K Y R W	R R YYR SYYY S
2	C K	S K R K	SYYR RY WSWS

Bread Wheat: Variant Analysis

- Browse and visualize variants • across the entire CSW genome
- Simultaneously compare and • visualize multiple samples
- Use one or more samples as a • control to exclude variants present in the wild type
- Focus on genomic regions of • interest based on IWGSC gene transcript annotations
- Access predicted biological • consequences based on calculated amino acid shifts



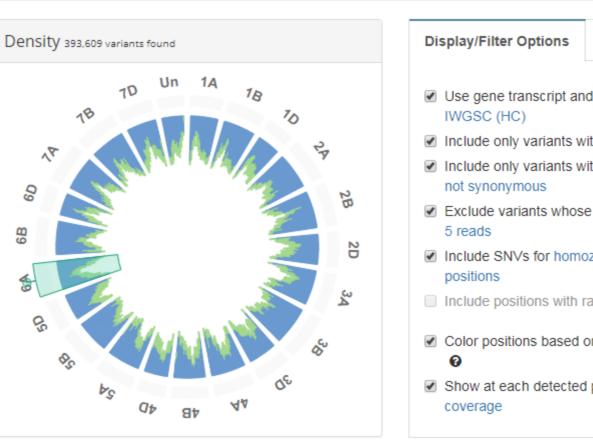
Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only



Compare Notes	and Marks	Sharing
xon annotations from		
	\$	Show only variants not present in the control (i.e. somatic)
		Show only variants with control coverage of at least 5 reads
mous Substitution)	⊘ ast	Match homozygous SNVs to related heterozygous SNVs in control @
re Stop Codon)	۲	Show all variants found within selected sample files
Codon Change) n Change)	•	Show only variants common to all selected sample files
of the Above) 🗸		Show file coverage track of variants detected per position
ymous Terminator)		
lutation)		
	predicted consequen mous Substitution) re Stop Codon) Codon Change) n Change) of the Above) ✓	predicted consequence of mous Substitution) ast re Stop Codon) Codon Change) n Change) of the Above) ✔

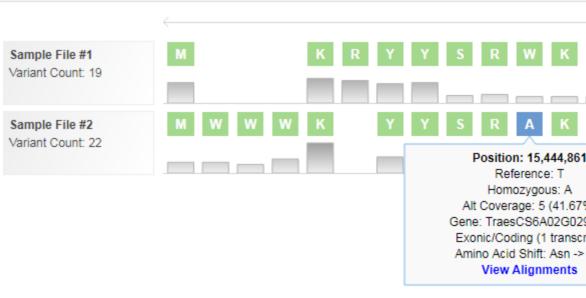
Bread Wheat: Variant Analysis

- Browse and visualize variants . across the entire CSW genome
- Simultaneously compare and • visualize multiple samples
- Use one or more samples as a • control to exclude variants present in the wild type
- Focus on genomic regions of • interest based on IWGSC gene transcript annotations
- Access predicted biological • consequences based on calculated amino acid shifts
- Interactive visualizations provide deeper access to the information behind each call



Variant Analysis CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only

Variants (Found 41 matching variants in range chr6A:1-618,079,261)



5	Compare	Notes and	Marks Sharing
an	d exon anno	otations from	 Show only variants not present in the control (i.e. somatic)
		coding regions	E roada
		ge is not at least	Match homozygous SNVs to related heterozygous SNVs in control O
		-	Show all variants found within selected sample files
		ieterozygous	 Show only variants common to all selected sample files
	rare alleles		Show file coverage track of variants detected per position
ed	on the varia	nt's general type	position
teo	position the	General T Alternate	
		Genomic	
		Predicted	Consequence
		0	chr6A:1-618,079,261 Q
	618,079	260 bp	
K	S C	R	SKRW R
K	S C	К	S K K R R Y W S
,86	1		
GO2	'%) 29200 ≎ript) ⊳ Tyr		



Bread Wheat RNA-Seq: Read Mapping and Visualization

Bread Wheat: RNA-Seq Read Mapping

 Conveniently select to align to either the full Chinese
 Spring Wheat genome or a transcriptome

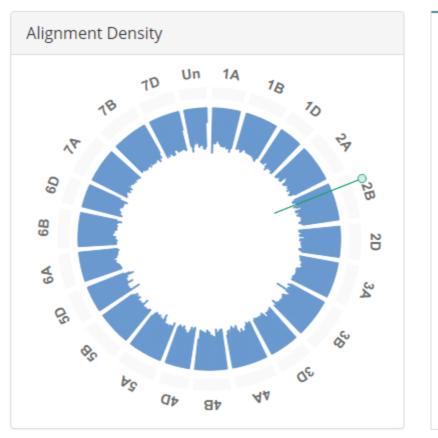
Dashboard / A	II Projects / Project	Details					
Align Multiple Sequence Files × Project Details							Mars Ontings
Project Details		Select Type:	Type: TopHat 2: Paired Alignment				More Options -
General Information Assembly:		Assembly:	IWGSC WGA 1.0 (Chinese Spring Wheat)				
Ρ	Project Name:		Genomic Features:	Align reads to entire genome			
	Species:	Alignment Pre-Processing	Contrine r cataleo.	 Align reads to a transcriptome 			
Ge	eneral Notes:	Transcriptome	Feature File:	Select Genome Feature Set	•		
	L	& Scoring	Read Alignment:	Transcriptome Alignment Only	▼ ?		
		Paired FASTQ	Prefilter Reads:	Include Multihit Reads	▼ ?		
All Files	Raw Sequences		Max Hits:	20	?		
► Start Mu	ultiple Alignments		Max Multihits:	20	?	• Create P	aired FASTQ File 🝷
🗹 🔻 File	Name						
SRR1	542417.1 Paired			Cancel > Start	Alignment	2:51 🕞 s	Start an Alignment 👻
SRR1	542416.1 Paired		_			/4:30 ► S	Start an Alignment 👻
SRR1	542415.1 Paired		4 Con	npleted,	2019-12-11 1	14:30 > S	Start an Alignment 👻
SRR1	542414.1 Paired		4 Con	npleted,	2019-12-09 1	17:37 🕨 S	Start an Alignment 👻
SRR1	542413.1 Paired		4 Con	npleted,	2019-12-11 1	16:19 🕞 s	Start an Alignment 🚽
SRR1	542412.1 Paired		4 Con	npleted,	2019-12-12 1	13:51 > S	Start an Alignment 👻

Bread Wheat: RNA-Seq Read Mapping

- Conveniently select to align to either the full Chinese
 Spring Wheat genome or a transcriptome
- Transcriptome is dynamically assembled on demand using selected annotation set from the IWGSC

Dashboard / All Projects / Project Details						
Project Details		Align Multiple Sequence Files			×	
		Colort Type:	TopHat 2: Paired Alignment			More Options -
General Information		Select Type:	TopHat 2. Paired Alight	nent		
		Assembly:	IWGSC WGA 1.0 (Chinese Spring Wheat)			
	Project Name:					
Species:		Alignment	Genomic Features:	 Align reads to entire genome Align reads to a transcriptome 		
		Pre-Processing				
General Notes:		Transcriptome	Feature File:	Select Genome Feature Set	•	
		Scoring	Community Genome Feature Sets IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)			
		Paired FASTQ	IWGSC - Low Confidence Genes - Chinese Spring Wheat (v1.1) IWGSC - Whole Genome - Chinese Spring Wheat (v1.1)			
		₩ Falleu FASTQ				
All F	iles Raw Sequences		Max Hits:	20	?	
Start Multiple Alignments			Max Multihits:	20	?	• Create Paired FASTQ File
	▼ File Name					
	SRR1542417.1 Paired				2.	:51 ► Start an Alignment -
				Cancel	Start Alignment	
	SRR1542416.1 Paired				4:	30 Start an Alignment -
	SRR1542415.1 Paired		4 Con	npleted,	2019-12-11 14:	:30 > Start an Alignment -
	SRR1542414.1 Paired		4 Con	npleted,	2019-12-09 17:	:37 Start an Alignment -
	SRR1542413.1 Paired		4 Con	npleted,	2019-12-11 16:	:19 Start an Alignment -
SRR1542412.1 Paired		4 Con	4 Completed,		:51 > Start an Alignment -	

- Conveniently select to align • to either the full Chinese Spring Wheat genome or a transcriptome
- Transcriptome is • dynamically assembled on demand using selected annotation set from the IWGSC
- Quickly visualize and • navigate RNA-Seq data from samples of any size



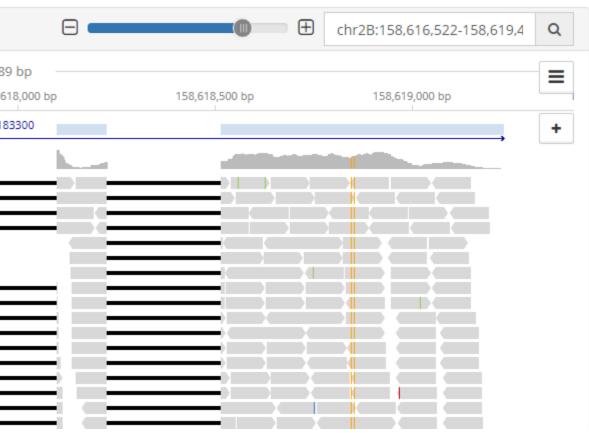
DI	splay Options	N
1	Show gene and	exon
1	Show reference	geno
1	Show nucleotide	
	differences exce	ed 2(
1	Exclude reads a	nd ba

- Group reads by UMI consensus of 60%
- Remove read duplication position and CIGAR

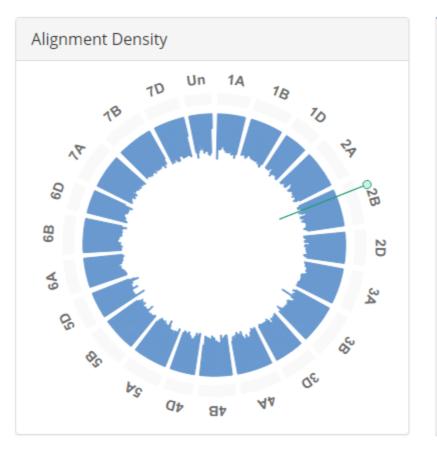
Reads (Showing 466 alignments of 1,736, excluded 1,040 of low quality)						
<			2,88			
00 bp	158,617,000 bp	158,617,500 bp	158,6			
			TraesCS2B02G1			

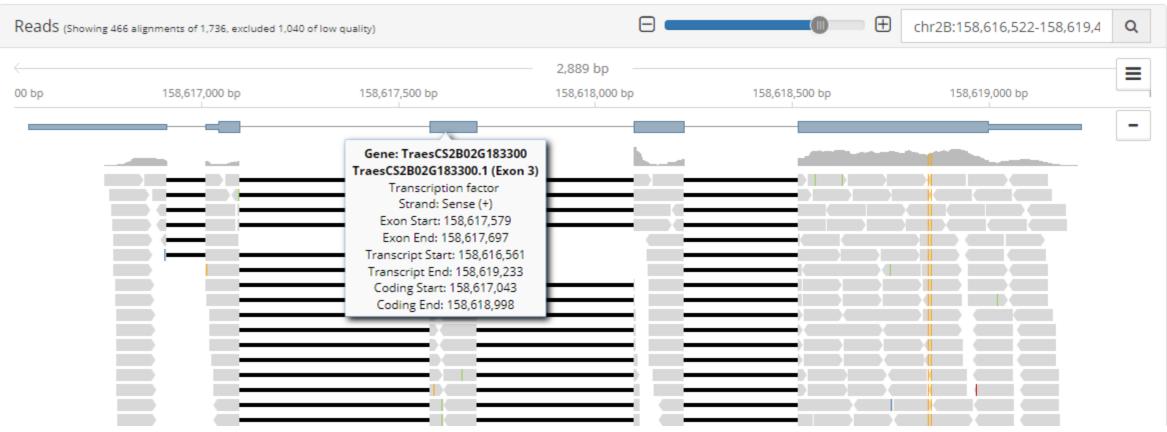
RNA-Seq Read Mapping Visualization

otes and Marks	Sharing	
a positions from IWG ome bases erage and highlight 0% ases with quality ur (perfect matches), ates at the same ali	when nder 99% requiring	 Show individual read details Show read consensus sequence information Highlight mismatched bases Highlight read insertions Color forward and reverse reads differently



- Conveniently select to align • to either the full Chinese Spring Wheat genome or a transcriptome
- Transcriptome is • dynamically assembled on demand using selected annotation set from the IWGSC
- Quickly visualize and • navigate RNA-Seq data from samples of any size
- Reference IWGSC transcript • and functional annotations on the fly





RNA-Seq Read Mapping Visualization

Di	splay Options	Notes and M	larks Sha	aring		
1	Show gene and Show reference Show nucleotide differences exce	genome bases coverage and	IWGSC Hig	h Confidenc	● Show individual r ce Genes (1.1) ✓ e Genes (1.1)	ead details hsus sequence information hed bases
	Exclude reads and Group reads by consensus of 60	UMI (perfect ma	-	9%	 Highlight read ins Color forward and 	ertions d reverse reads differently
	Remove read du position and CIG		same alignmer	nt		

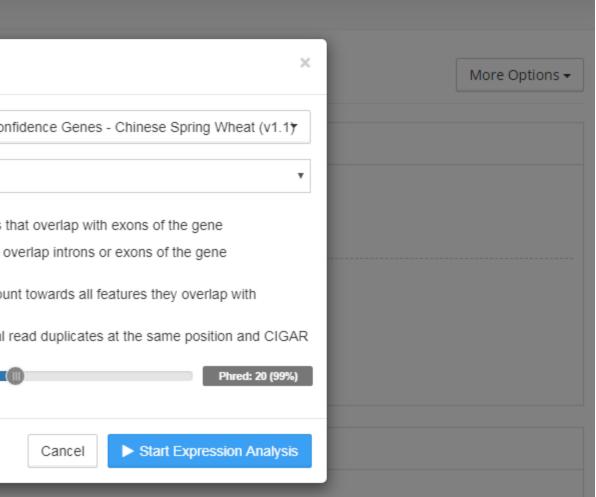


Bread Wheat RNA-Seq: Expression Analysis

Bread Wheat: Expression Analysis

 Analyze feature expression utilizing the Chinese Spring Wheat reference assembly

Dashbo	Dashboard / All Projects / Project Details						
Proje	ect Details	Start Multiple Expression A	nalyses				
Gen	eral Information	Genome Feature Set: 0	IWGSC - High Co				
	Project Name:	Feature to Count: 📀	Gene				
	Species:	Exon Matching: 🥹	 Count only reads Count reads that of 				
	General Notes:	Multi Feature Overlap: 😧	Allow reads to counter the second				
		De-duplication: 😧	Remove potential				
		Minimum Quality: 🕑					
All F		Delete Files					
	File Name						
	SRR1542407.1 Paired (A	Aligned 2019-12-13 21:40 UTC)					
	SRR1542404.1 Paired (A	Aligned 2019-12-13 21:40 UTC)					
	SRR1542405.1 Paired (A	Aligned 2019-12-13 21:40 UTC)					
	SRR1542406.1 Paired (A	Aligned 2019-12-13 21:40 UTC)					
	SRR1542408.1 Paired (A	Aligned 2019-12-13 21:40 UTC)					
	SRR1542409.1 Paired (A	Aligned 2019-12-12 19:53 UTC)					

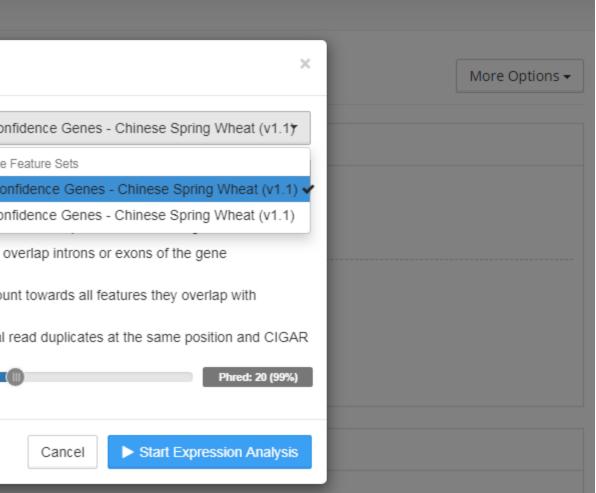


Reads Processed	▼ Last Updated	
193,694,736	2019-12-16 17:27	Lill View Alignments 👻
198,975,170	2019-12-16 13:05	Lul View Alignments 👻
184,418,714	2019-12-16 11:02	Lill View Alignments 👻
184,129,219	2019-12-16 07:01	Lill View Alignments 👻
165,312,410	2019-12-16 02:07	Lill View Alignments 👻
203,941,764	2019-12-14 08:25	Lint View Alignments 👻

Bread Wheat: Expression Analysis

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC "High Confidence" or "Low Confidence" annotation sets

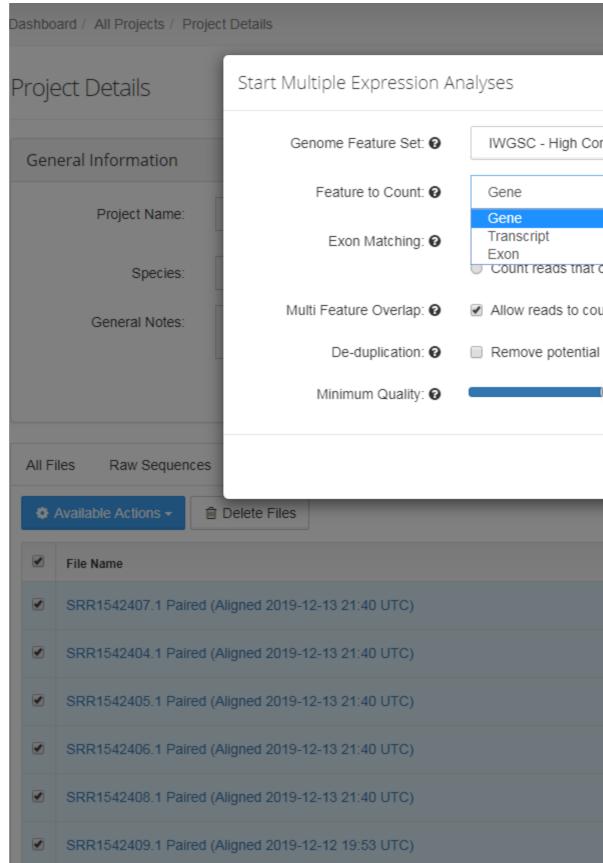
Dashbo	oard / All Projects / Projec	ct Details	
Proje	ect Details	Start Multiple Expression A	nalyses
Gen	eral Information	Genome Feature Set: 🕑	IWGSC - High Col
	Project Name:	Feature to Count: @	IWGSC - High Co
		Exon Matching: 😧	IWGSC - Low Con
	Species:		Count reads that of
	General Notes:	Multi Feature Overlap: 📀	Allow reads to council
		De-duplication: 🚱	Remove potential
		Minimum Quality: 🕑	
All F		Delete Files	
	File Name		
	SRR1542407.1 Paired (/	Aligned 2019-12-13 21:40 UTC)	
	SRR1542404.1 Paired (/	Aligned 2019-12-13 21:40 UTC)	
	SRR1542405.1 Paired (/	Aligned 2019-12-13 21:40 UTC)	
	SRR1542406.1 Paired (/	Aligned 2019-12-13 21:40 UTC)	
	SRR1542408.1 Paired (/	Aligned 2019-12-13 21:40 UTC)	
	SRR1542409.1 Paired (/	Aligned 2019-12-12 19:53 UTC)	



Reads Processed	 Last Updated 	
193,694,736	2019-12-16 17:27	Lud View Alignments 👻
198,975,170	2019-12-16 13:05	Lud View Alignments 👻
184,418,714	2019-12-16 11:02	Lud View Alignments 👻
184,129,219	2019-12-16 07:01	Lill View Alignments 👻
165,312,410	2019-12-16 02:07	Lul View Alignments 👻
203,941,764	2019-12-14 08:25	Lul View Alignments 👻

Bread Wheat: Expression Analysis

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC "High Confidence" or "Low Confidence" annotation sets
- Measure expression levels of genes, transcripts, or individual exons



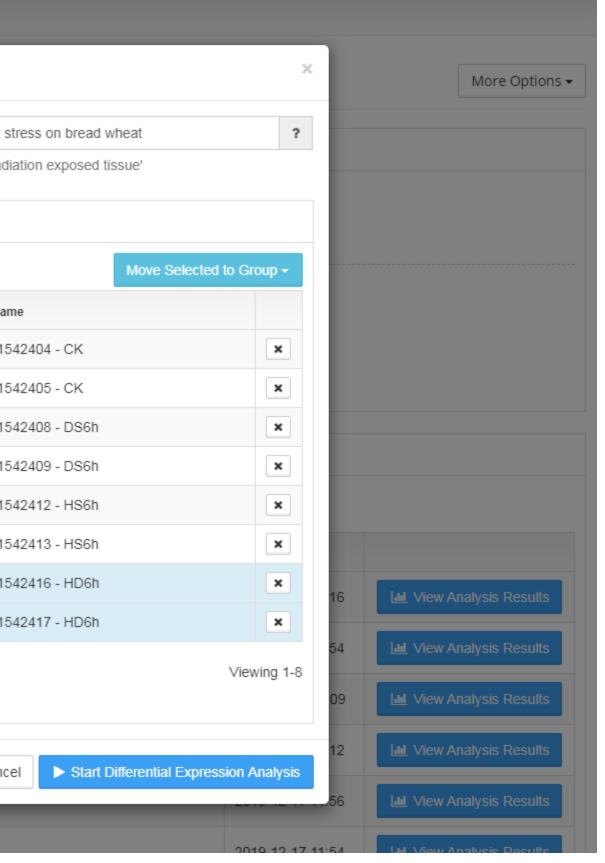
×	More Options 🗸
onfidence Genes - Chinese Spring Wheat (v1.1)	
•	
overlap introns or exons or the gene	
unt towards all features they overlap with I read duplicates at the same position and CIGAR	
(III) Phred: 20 (99%)	
Cancel Start Expression Analysis	

Reads Processed	 Last Updated 	
193,694,736	2019-12-16 17:27	Lud View Alignments 👻
198,975,170	2019-12-16 13:05	Lud View Alignments 👻
184,418,714	2019-12-16 11:02	Lill View Alignments 👻
184,129,219	2019-12-16 07:01	Lul View Alignments 👻
165,312,410	2019-12-16 02:07	Lul View Alignments 👻
203,941,764	2019-12-14 08:25	Lul View Alignments 👻

Bread Wheat: Differential Expression

- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC "High Confidence" or "Low Confidence" annotation sets
- Measure expression levels of genes, transcripts, or individual exons
- Group samples for a differential expression analysis...

Dashboard / All Projects / Project Details					
Project Details	Start a Differential Expression Analysis				
General Information				nd heat s	
Project Name:	Expe	eriment Design	E.g. 'Expression chan	ge in rau	
Species:	Defin	e Sample Groups:			
General Notes:		Set Group		File Na	
		Group 1	•	SRR15	
		Group 1	Ŧ	SRR15	
		Group 2	¥	SRR15	
All Files Raw Sequences		Group 2	T	SRR15	
Type here		Group 3	Ŧ	SRR15	
File Name		Group 3	T	SRR15	
SRR1542404 - CK		Group 4	Y	SRR15	
		Group 4	T	SRR15	
SRR1542405 - CK					
SRR1542406 - DS1h					
SRR1542407 - DS1h					
SRR1542408 - DS6h			Lapi content in	Canc	
B SDD1542400 DS6b			Evoression An	alveie	



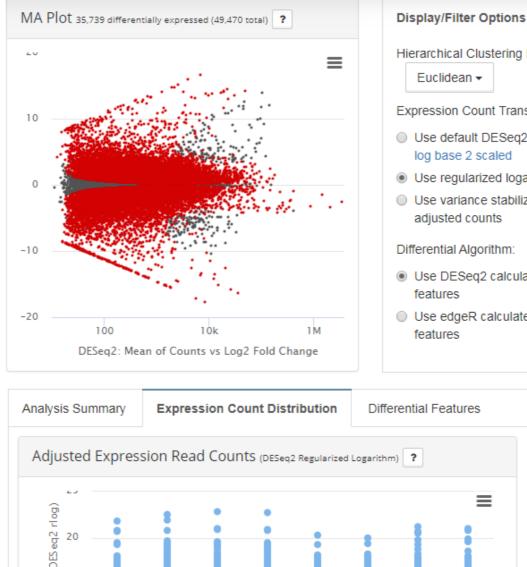
Bread Wheat: Differential Expression

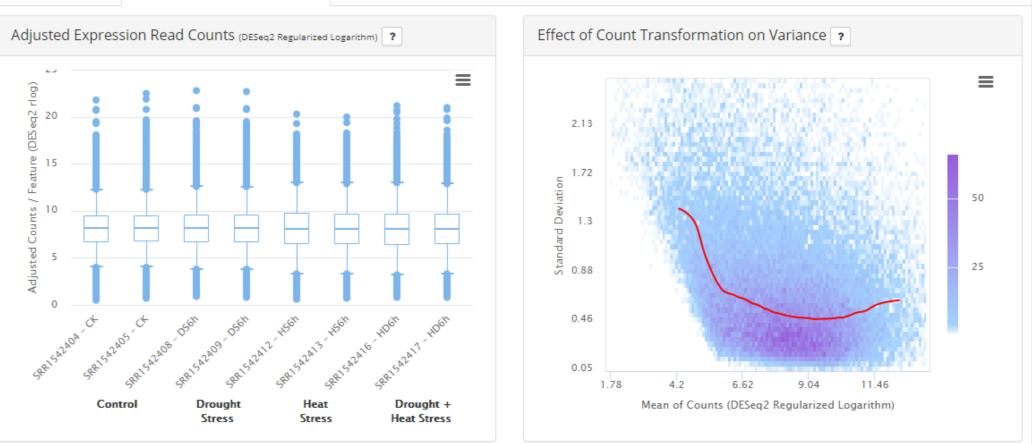
- Analyze feature expression utilizing the Chinese Spring Wheat reference assembly
- Leverage either the IWGSC "High Confidence" or "Low Confidence" annotation sets
- Measure expression levels of genes, transcripts, or individual exons
- Group samples for a differential expression analysis...
- …and conveniently tune industry standard 'R' based algorithms

Dashboard / All Projects / Projec Start a Differential Expression Analysis					×			
Proje	ect Details	Analysis Name:		eat stress on bread wheat		?		More Options -
Ger	neral Information		L.g. Expression enangen					
		Experiment Design	Advanced Options					
	Project Name:	DESeq2 Options:						
	Species:							
	General Notes:	М	inimum Expression Count:	1	?			
	_	Required Samples	with Minimum Expression:	1	?			
		Independent Filte	ring Significance Cutoff: 0	0.1				
All F	Files Raw Sequences	Count	Transformation Blinding: 0	 Use experiment design information wh calculating transformed counts Blind the transformation counts to the experimental design (i.e. unsupervised) 				
	File Name							
	SRR1542404 - CK	edgeR Options:					16 🔟 View	Analysis Results
	SRR1542405 - CK	Minimum	Counts Per Million (CPM):	1.0	?		54 Luu View	Analysis Results
	SRR1542406 - DS1h	Dequired Sa	mples with Minimum CPM:	1	?		09 Lui View	Analysis Results
0	SRR1542407 - DS1h		ential Testing Procedure: 0	 Use GLM likelihood ratio test 	r		12 🔟 View	Analysis Results
	SRR1542408 - DS6h		,	 Use GLM quasi-likelihood (QL) F-test 			56 🔟 View	/ Analysis Results

Bread Wheat: Differential Expression

 Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets





Notes Sharing

- Hierarchical Clustering Distance Measure:
- Expression Count Transformation:
- Use default DESeq2 normalized expression counts
- Use regularized logarithm (rlog) adjusted counts
- Use variance stabilizing transformations (VST)
- Use DESeq2 calculated differentially expressed
- Use edgeR calculated differentially expressed

Sort and Show Differential Features With:

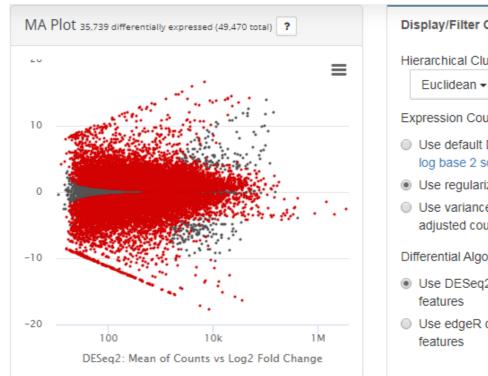
- Strongest up or down regulation (log2 fold change, absolute)
- Strongest down-regulation (log2 fold change, ascending)
- Strongest up-regulation (log2 fold change, descending)

Maximum Features To Hierarchically Cluster



Bread Wheat: Differential Expression

- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets
- Validate sample distances through principal component analysis





Display/Filter Options

Notes Sharing

- Hierarchical Clustering Distance Measure:
- Expression Count Transformation:
- Use default DESeq2 normalized expression counts log base 2 scaled
- Use regularized logarithm (rlog) adjusted counts
- Use variance stabilizing transformations (VST) adjusted counts
- Differential Algorithm:
- Use DESeq2 calculated differentially expressed
- Use edgeR calculated differentially expressed

Sort and Show Differential Features With

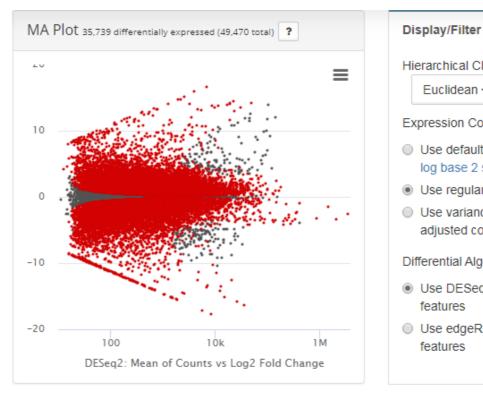
- Strongest up or down regulation (log2 fold change, absolute)
- Strongest down-regulation (log2 fold change, ascending)
- Strongest up-regulation (log2 fold change, descending)

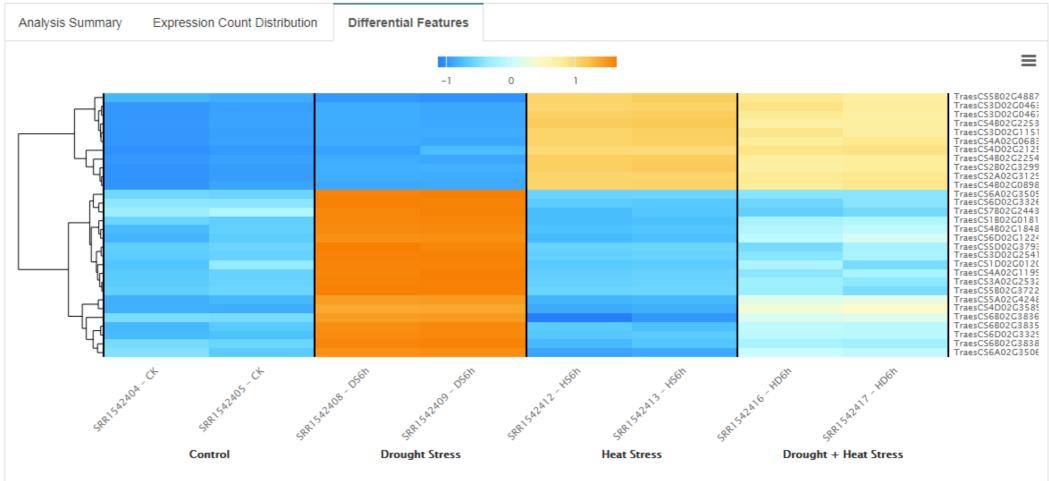
Maximum Features To Hierarchically Cluster



Bread Wheat: Differential Expression

- Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets
- Validate sample distances through principal component analysis
- Highlight up or down regulated genes through interactive heat maps





Display/Filter Options

Notes Sharing

Hierarchical Clustering Distance Measure

Expression Count Transformation

- Use default DESeq2 normalized expression counts log base 2 scaled
- Use regularized logarithm (rlog) adjusted counts
- Use variance stabilizing transformations (VST) adjusted counts

Differential Algorithm:

- Use DESeq2 calculated differentially expressed
- Use edgeR calculated differentially expressed

Sort and Show Differential Features With

- Strongest up or down regulation (log2 fold change, absolute)
- Strongest down-regulation (log2 fold change, ascending)
- Strongest up-regulation (log2 fold change, descending)

Maximum Features To Hierarchically Cluster



Bread Wheat: Differential Expression

 Visualize expression levels utilizing transcriptome-based alignments leveraging IWGSC annotation sets

- Validate sample distances through principal component analysis
- Highlight up or down regulated genes through interactive heat maps
- Compare and contrast different • industry standard algorithms in real-time







In Conclusion

Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020

Looking Ahead

- Incorporation of the recently released IWGSC
 2.0 reference assembly
- Addition of the soon-to-be released IWGSC 2.0 transcript and functional annotations
- Consideration of other standard hexaploid and tetraploid wheat assemblies and related annotation sets
- Support for custom reference assemblies
- Additional crop-research-specific analysis types and interpretive visualizations



Acknowledgements

Special thanks to collaborators at:

- Arbor Biosciences •
- John Innes Centre (JIC) •
- French National Institute for Agricultural Research (INRA) •
- University of Adelaide, Plant Genomics Centre •

And the collective efforts of the:

International Wheat Genome Sequencing Consortium (IWGSC) •

Differential expression analysis utilized FASTQ files published as part of the following study:

Liu Z, et al. Temporal transcriptome profiling reveals expression partitioning of homeologous • genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.) BMC Plant Biol. 2015;15:152. doi: 10.1186/s12870-015-0511-8.



Alignment, Analysis, and Visualization of Wheat NGS Samples in Curio / Presented at PAG 2020