Extending the Curio Genomics Platform for the Wheat Research Community DNA Sequences to Analysis Results Presented April 30th, 2020 to the IWGSC Community Author: Shawn Quinn (CTO, Curio Genomics)

CURIO



Extending the Curio Genomics Platform for the Wheat Research Community: **DNA Sequences to Analysis Results**

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

Presentation Overview

- What is Curio
- Demonstration of Several Crop/Wheat Research Examples •
 - DNA-Seq and RNA-Seq
 - Read Mapping Analysis
 - Variant Analysis
 - Coverage 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





DNA-Seq Hexaploid Wheat: Read Mapping and Visualization

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

Hexaploid Wheat: Read Mapping

 Incorporates Chinese Spring Wheat (*Triticum aestivum*) reference assembly from the IWGSC (i.e. the International Wheat Genome Sequencing Consortium)

Dashboard / All Proj	Dashboard / All Projects / Project Details						
		Align Paired-End Sequence Files ×					
Project Detail	S	Select Type:	Bowtie 2: Paired Alignr	nent	•	More Options -	
General Inform	ation	Assembly:	IWGSC WGA 1.0 (Chines	e Spring Wheat)			
Project	t Name:						
	and the second se	Alignment	Read Trimming: 📀	 Do not perform any read trimmi Trim the reads before aligning 	ing		
5	species:	Pre-Processing	Dead Length: 0	Deed Leastha			
General	I Notes:	Scoring	Reau Lengui.	Discard short reads			
		Paired FASTQ	Down Sampling: 😧	Process all reads			
				Process a sampling (%) of the r	reads		
	_			Process a fixed number of read	Is		
All Files Raw	Sequences		UMI Processing: 🕑	Reads do not contain UMI/UMT	īs -		
Start An Align	ment			Reads contain unique molecula	ar ids/tags	Create Paired FASTQ File	
File Name				Cancel	Start Alignment	1	
SampleA-R	2.fastq.gz		None		2010-12-00-17	:23 > Start an Alignment -	
SampleA-R	1.fastq.gz		2 Comple	eted,	2019-12-06 17	:23 Start an Alignment -	
SampleA Pa	Paired 16 Comp		leted,	2019-12-06 17	:00 > Start an Alignment -		

Viewing 1-3

Hexaploid Wheat: Read Mapping

- Incorporates Chinese Spring Wheat (*Triticum aestivum*) reference assembly from the IWGSC (i.e. the International Wheat Genome Sequencing Consortium)
- 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



			_	
		×		
				More Options 🗸
t		•		
t				
nt ent				
t) Dor	ot perform any read trimming			
) Trim	the reads before aligning			
t:	0	?		
1:	Don't trim end of reads	?		
d:	0	?		
S:	All Bases	?		
Proc	ess reads of any length		ed	
) Disc	ard short reads			
Pro	ess all reads		7:23	Start an Alignment
) Proc	ess a sampling (%) of the reads		7.03	Start an Alignment
) Proc	ess a fixed number of reads		1.25	
Reads do not contain UMI/UMTs			7:00	► Start an Alignment -
Rea	ds contain unique molecular ids/tags	;		
				Viewing 1-3
	Cancei Start Align	iment		

Hexaploid Wheat: Alignment Visualization

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



D	isplay Options	No
	Show gene and	exon
4	Show reference	genor
	Show nucleotide differences exce	cove ed 20
	Exclude reads a	nd ba
	Group reads by consensus of 60	UMI (%
	Remove read du position and CIG	iplicat SAR

Reads (Showing 583 alignments of 777, excluded 1 of low quality, removed 615 duplicates)					
< 18,685,750 bp	18,686,000 bp	18,686,250 bp	18,686,500 bp	18,686,750 b	2,71

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



Display Options	No
Show gene and	exon
Show reference	genor
 Show nucleotide differences excert 	e cove ed 20
Exclude reads a	nd bas
Group reads by consensus of 60	UMI (()%
Remove read du position and CIG	uplicate GAR

Reads (Showing 567 alignments of 740, excluded 40 of low quality, removed 613 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 M	arks Sharing		
 Show gene and Show reference Show nucleotide differences exce Exclude reads an Group reads by consensus of 60 	exon positions <u>f</u> genome bases coverage and ed 20% nd bases with qu UMI (perfect ma	rom IWGSC (HC) IWGSC High Confide IWGSC Low Confide uality under 99.9% atches), requiring	 Show individual ence Genes (1.1) ence Genes (1.1) Highlight read in Color forward and 	read details hsus sequence information hed bases isertions ind reverse reads differently
Remove read du position and CIG	iplicates at the s GAR	ame alignment		



			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



C	isplay Options	No
	Show gene and	exon p
•	Show reference	genon
•	Show nucleotide differences exce	ed 20
¥	Exclude reads a	nd bas
	Group reads by consensus of 60	UMI (p %
	Remove read du	plicat

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

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



- 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



Display Options	No
Show gene and	exon
Show reference	genor
 Show nucleotide differences exce 	cove ed 20
Exclude reads a	nd bas
Group reads by consensus of 60	UMI (f %

Reads (Showing 83 alignments, excluded 17 of low quality, removed 49 duplicat	les) 🗖 🗖		glutenin Q
< 21,714,500 bp	377 bp 21,714,600 bp	21,714,700 bp	Download PNG Image Download PDF Document
			 Download SVG Vector Image Download Reference Sequence
CSW10_chr1A_21714434-21714810.fasta - Notepad File Edit Format View Help chr1A: 21714434 - 21714810 AGGAGAAGGAGGAGGAGGATCTATGGTAGCCGGCCGT CGTTGCAGCAGTAGCAGACGTCGCCGCGCGGACGAG TCGCATGCACTTGCCCGTCCCTTGCCAGCACTTG CCCCGCACCGCTGTTGTCGCCGCCCCCACTGCAC CACTAGTTGGCTCTTATCGAACACTTACGGATGA	GTCGCTCTTGTGCAGGCGAGGAGGGT ATTGCAGCGGCCATGGACGTAGCCCA ACCATGCACTCTGCCCTCAGCACGTC ATATGAAGAACGAGATTTGAACAAAA GAAGGATGGCCATGAAGAACAGTTTC	GGTGCGCCGTCGT GGGCCTGGCAGCA GCACACCAGCTCG TTCATCGAAGATT TTGCCCATCTTGT	 Download Consensus Sequence Download Read Coverage Download Read Consensus Download Minor Alleles

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

Remove read duplication and CIGAR

Hexaploid Wheat: Alignment Visualization

- 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)

tes and Marks	Sharing				
Search Comme	ents	Q		0 -	_
Jpdated 2019 De	c 20 15:45:54. Posted 2019 [Dec 20 14:42:57	C	<u>ش</u>	
ozygous SNPs she sample.	own here in this Mut5 sample	appear to be heterozy	ygous in ti	he wild	
Updated 2019 De	c 20 14:49:15. Posted 2019 [Dec 20 14:48:24	ľ	<u>ش</u>	
atic variants in chr	2A of the Mut5 and Mut7 san	nples.			
Updated 2019 De	c 20 15:06:26. Posted 2019 [Dec 20 15:05:36	C	<u>ش</u>	-
			,692-762,	285,9	۹
bp	762,285	,900 bp			
					-
	I	т			
		1			
		• • • • • • • • • • • • • • • • • • •			



DNA-Seq Hexaploid Wheat: Variant Analysis

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

Hexaploid 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



		More Options -
	×	
Conomo Chinaco Enring Whaat (v1.1)	_	
Senome - Chinese Sphing Wheat (VI.I)		
plicates at the same position and CIGAR		
Phred:	30 (99.9%)	
10	Reads	
se quality data in heterozygous calls equencies only in heterozygous calls		
	n%	
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	Lud View Alignments -
186,030,770	2019-12-18 16:21	Lind View Alignments -

Hexaploid 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
- Automatically track the provenance of every file...



Hexaploid 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
- Automatically track the provenance of every file...
- ...including the specific options that went into any analysis

roje	ect Details	4 Summary	Туре:			More Option
Gen	eral Information	🗋 File Details	Variant Analysis Status:			
	Project Name:	📽 Sharing	Initiated By:			
		🕸 Job Details	squinn (Shawn Quinn)			
	Species:	Processing	Parameter Name	Parameter Value		
	General Notes:	Lill Charts	GENOME_FEATURE_FILE	IWGSC - High Confidence Genes - Chinese Spring Wheat (v1.1)		
			MAX_RARE_ALLELE_FREQUENCY	0.0		
			MIN_COVERAGE	2		
			MIN_RARE_ALLELE_FREQUENCY	0.0		
			MIN_VARIANT_FREQUENCY	0.1		
All Fi	iles Raw Sequences		OUTPUT_PROJECT_FILE	CultivarE-WT - BWA-Aln (Analyzed 2019-12-20 19:53 UTC) - High Sensitivity / HC Only		
•	Available Actions -		PERFORM_RARE_ALLELE_DETECTION	false		
	File Name		QUALITY_THRESHOLD	20	adatad	
	File Nallie		REMOVE_DUPLICATE_READS	false	Juateu	
	CultivarE-WT - BWA-Alr		SINGLE_PROJECT_FILE	CultivarE-WT - BWA-Aln (Aligned 2019-12-19 20:28 UTC)	20 05:59	Lul View Alignments 👻
	CultivarE-Mut5 - BWA-A		UMI_PROCESSING_ENABLED	false	20 05:50	💵 View Alignments 👻
			USE_QUALITY_DATA_IN_HETEROZYGOUS_CALLS	true		
	CultivarE-Mut7 - BWA-A				20 04:08	Lill View Alignments 👻
	CultivarE-WT - Bowtie2		Analysis Version Info:		18 21:55	Lul View Alignments 👻
			Curio Variant Detection Analysis, Version:	1.4.1, Updated: October 14,		

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





Hexaploid Wheat: Variant Analysis

	Compare	Notes and Marks	Sharing
an	d exon annotatio	ns from	Show only variants not present in the control (i.e. somatic)
w	ithin known codir ith predicted con:	ig regions 😧 🛛 🖉	Show only variants with control coverage of at least 5 reads
S	e alt coverage is		Match homozygous SNVs to related heterozygous SNVs in control 🕢
nc	zygous or hetero	ezygous	Show all variants found within selected sample files Show only variants common to all selected sample files
n r d (are alleles on the variant's a	Iternate	Show file coverage track of variants detected per position
ed	position the vari	ant's	

Hexaploid 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 • **6**B 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 Sharin	g	
	Q		
File Name		Records	Last Updated
ultivarE-Mut5 - i:03 UTC)	BWA-Aln (Analyzed 2019-12-20	376,711	2019 December 20, Friday 10:56:25 UTC-5
ultivarE-Mut7 - 5:03 UTC)	BWA-Aln (Analyzed 2019-12-20	306,472	2019 December 20, Friday 10:55:12 UTC-5
ultivarE-WT - E 5:23 UTC) - Me	WA-Aln (Analyzed 2019-12-20 edium Sensitivity	5,859,463	2019 December 20, Friday 11:41:27

Viewing 1-8 of 8

Hexaploid 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	ks	Sharing
an	d exon annotatio	ns from	•	Show only variants not present in the control (i.e. somatic)
w	ithin known codi	ng regions 😧 sequence of		Show only variants with control coverage of at least 5 reads
)S	e alt coverage is	not at least	•	Match homozygous SNVs to related heterozygous SNVs in control @
no	ozygous or heter	ozygous	•	Show all variants found within selected sample files Show only variants common to all selected sample files
nr d(are alleles on the variant's a	Iternate		Show file coverage track of variants detected per position
ed	position the vari	iant's		

K R Y Y R M Y Y S R W S K S R

MYYSRASKSR

- 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



MWKWRWK

SRRM

Sample File #1 Variant Count: 40

Sample File #2 Variant Count: 34

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

Hexaploid Wheat: Variant Analysis

_						
	Compare	Notes and Marks		Sharing		
n	d exon annota	ations from	/	Show only somatic)	variants not present in the control (i.e.	
w	ithin known <u>c</u>	oding regions 🔞 🛛 🖟		Show only	variants with control coverage of at leas	st
w	ith predicted	Exonic Regions				
		Intronic Regions			zygous SNVs to related heterozygous	
S	e alt covera	Coding Regions 🗸			trol 🚱	
nc	ozygous or t	Intragenic Regions			iants found within selected sample file ariants common to all selected sample	s e
			15	(UTR)		
1 [are alleles	5 UTRS			verage track of variants detected per	
d (on the varia	3 UTRS				
		Splice Regions				
ed	position the	Upstream/Downstre	an	n Regions		
		Upstream Regions				
		Downstream Region	S			
		Intergenic Regions				
		,		Ð	chr6A:1-618,079,261 Q	
		60 bp				
6				D D		c
				R R		3
				_		_
0	c <mark>K</mark>	S K R	ĸ	<mark>S</mark> Y	YR RY WSW	s
					_	
						-

- 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



Hexaploid Wheat: Variant Analysis

	Compare No	tes and Marks		Sharing
an	d exon annotations f	rom (1	Show only variants not present in the control (i.e. somatic)
w w	ithin known coding re ith predicted consequ	egions 😧 🛛	1	Show only variants with control coverage of at least 5 reads
nc	nymous Substitution) ast	1	Match homozygous SNVs to related heterozygous SNVs in control 🚱
m	ature Stop Codon)		۲	Show all variants found within selected sample files
nat Co	or Codon Change)		0	Show only variants common to all selected sample files
(/	All of the Above) 🗸			Show file coverage track of variants detected per position
yn en	onymous Terminator t Mutation))		

- 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)



Hexaploid Wheat: Variant Analysis

re notatio n codi ed cor age is heter ant's <u>c</u>	Notes and Mark ons from ing regions (2) insequence of not at least ozygous general type General Type Alternate Allele Genomic Local Predicted Cons	ks	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
notatio n codi ed cor age is heter ant's <u>c</u>	ons from Ing regions (2) Insequence of Inot at least OZYGOUS General Type Alternate Allele Genomic Locat Predicted Cons		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
n codi ed cor age is heter ant's g he va	ing regions insequence of not at least ozygous general type General Type Alternate Allele Genomic Locat Predicted Cons		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
age is heter ant's <u>c</u> he va	not at least ozygous general type General Type Alternate Allele Genomic Locat Predicted Cons	 • •<	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
heten	ozygous general type General Type Alternate Allele Genomic Locat Predicted Cons	• • • • tion	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
ant's_g he va	general type General Type Alternate Allele Genomic Locat Predicted Cons	■ • tion	files Show file coverage track of variants detected per position
ant's_g	General type General Type Alternate Allele Genomic Locat Predicted Cons	■ • tion	Show file coverage track of variants detected per position
he va	General Type Alternate Allele Genomic Locat Predicted Cons	✓	
	i realetea een	seau	ience
0		ooqu	
9,260 R	bp К S	К	R W R S
	KS	K	K R R Y W S
		K S	



DNA-Seq Hexaploid Wheat: Coverage Analysis

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

Hexaploid Wheat: Coverage Analysis

 Measure library / kit exome or custom target area key coverage metrics



Display	Options	С

Coverage and Chromoso

- Limit GC content to p deviations from the av
- Show on a column se average coverage de
- Show on a line series percentage of covere

Feature/Region Details:

- Search gene and exo
- Show on a column se coverage depth
- Show on a line series covered positions

overage Summary	Targets & Chromo	osomes Feature	e/Region Details
Key Metrics ?			
Features/Regions A	nalyz 749,233 fe	eatures	90 bases)
Total Aligned Seq	uences: 7,612,616	reads (1,090,140,19	
	Target	Regions Padded	Regions Padded +/-
	Features/Regions	+/- 100 bp	200 bp
Genomic Positions	170,215,801	248,743,228	296,596,680
Analyzed	positions	positions	positions
Covered Genomic	106,153,560	154,477,995	176,997,919
Positions	positions (62.36%)	positions (62.1%)	positions (59.68%)
Overlapping	3,182,536 reads	3,730,547 reads	3,954,530 reads
Reads	(41.81%)	(49%)	(51.95%)
Overlapping	329,259,294 bases	492,632,910	558,434,526 bases
Sequenced Bases	(30.2%)	bases (45.19%)	(51.23%)

ompare	Notes and Marks	Sharing
ome Summar	y:	Shared Settings:
ositions withi verage	n 5 standard	Show metrics for the exact genomic range of each feature
eries each chi epth	romosome's	Show metrics for the range of each feature padded +/- 100 bp
s each chromo d feature pos	osome's sitions	Show metrics for the range of each feature padded +/- 200 bp
on positions fr	rom IWGSC (HC)	
eries each fea	ature's average	
s each feature	e's percentage of	



- Measure library / kit exome • or custom target area key coverage metrics
- Compare and visualize • multiple samples simultaneously





Hexaploid Wheat: Coverage Analysis

mpare Notes and Marks Shari	ng		
×			
	Records	Last Updated	•
(Bowtie Local, Very Sensitive) - No	749,233	2019 December 9, Monday 16:03:43 UTC-5	
(Bowtie Local, Very Fast) - No Qual	749,233	2019 December 9, Monday 16:02:52 UTC-5	
(Bowtie End to End, Very Sensitive) -	749,233	2019 December 9, Monday 16:04:23 UTC-5	
(Bowtie End to End, Very Fast) - No	749,233	2019 December 9, Monday 16:06:35 UTC-5	-

Viewing 1-5

- Measure library / kit exome • or custom target area key coverage metrics
- Compare and visualize • multiple samples simultaneously
- Access individual gene or • transcript coverage metrics leveraging IWGSC annotation sets



Display Options С Coverage and Chromos Limit GC content to p deviations Show on a column se average coverage de Show on a line series percentage of covere Feature/Region Details:

- Search gene and exc
- Exclude features with positions covered
- Show on a column see coverage depth
- Show on a line series covered positions



Hexaploid Wheat: Coverage Analysis

ompare	Notes and Marks	Sharing
ome Summa	ary:	Shared Settings:
ositions with	nin 2 standard	 Show metrics for the exact genomic range of each feature
eries each ch epth	nromosome's	 Show metrics for the range of each feature padded +/- 100 bp
s each chron ed feature po	nosome's ositions	Show metrics for the range of each feature padded +/- 200 bp
on positions	from IWGSC (HC)	
n less than 1	% of base	
eries each fe	ature's average	
s each featui	re's percentage of	

		glutenin	Q
Percent	Hover over a co more details or	TRAESCS1A02G008000 (chr1A) low molecular weight glutenin subunit TRAESCS1A02G010900 (chr1A) low molecular weight glutenin subunit	
. ද 80% C		TRAESCS1A02G039600 (chr1A) high molecular weight glutenin subunit	
40% 40%		TRAESCS1A02G317500 (chr1A) high molecular weight glutenin subunit TRAESCS1B02G011600 (chr1B) low molecular weight glutenin subunit	
ALN, Defaults) – HC Genes e Local, Very Fast) – HC Genes			



RNA-Seq Hexaploid Wheat : 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 / All Projects / Proje	ct Details			_	
Project Details	Align Multiple Seque	Align Multiple Sequence Files ×			
	Select Type:	TopHat 2: Paired Alignm	nent	•	More Options •
General Information	Assembly:	IWGSC WGA 1.0 (Chinese	e Spring Wheat)		
Project Name: Species:	Alignment	Genomic Features:	 Align reads to entire genome Align reads to a transcriptome 		
General Notes:	Transcriptome	Feature File:	Select Genome Feature Set	•	
	Scoring	Read Alignment:	Transcriptome Alignment Only	• ?	
	Paired FASTQ	Prefilter Reads:	Include Multihit Reads	• ?	
All Files Raw Sequences		Max Hits:	20	?	
Start Multiple Alignments		Max Multihits:	20	?	● Create Paired FASTQ File -
File Name					
SRR1542417.1 Paired			Cancel Sta	2:5 rt Alignment	51 Start an Alignment -
SRR1542416.1 Paired				4:3	Start an Alignment
SRR1542415.1 Paired		4 Com	pleted,	2019-12-11 14:3	Start an Alignment
SRR1542414.1 Paired		4 Com	pleted,	2019-12-09 17:3	37 ► Start an Alignment -
SRR1542413.1 Paired		4 Com	pleted,	2019-12-11 16:1	9 Start an Alignment -
SRR1542412.1 Paired		4 Com	pleted,	2019-12-12 13:5	51 🕨 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

Dashbo	ard / All Projects / Project	t Details			_	
		Align Multiple Sequence Files ×				
Proje	ect Details	Colort Type:	Tapliat 2: Daired Alian	nont		More Options -
Car	Linformation	Select Type.	TopHat 2. Paired Alight	nent		
Gen	eral information	Assembly:	IWGSC WGA 1.0 (Chines	e Spring Wheat)		
	Project Name:					
	Oracian	Alignment	Genomic Features:	 Align reads to entire genome Align reads to a transcriptome 		
	Species:	Pre-Processing				
	General Notes:	Transcriptome	Feature File:	Select Genome Feature Set	•	
		Scoring	Community Genon	ne Feature Sets Confidence Genes - Chinese Spring V	Wheat (v1 1)	
		A Daired FASTO	IWGSC - Low C	onfidence Genes - Chinese Spring V	Wheat (v1.1)	
	_	* Failed FASTQ	IWGSC - Whole	Genome - Chinese Spring Wheat (v	v1.1)	
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	npleted,	2019-12-12 13:	:51 > Start an Alignment -

Display Options

Show gene and exor Show reference gen

Show nucleotide cov

Exclude reads and b

Group reads by UMI

Remove read duplica

consensus of 60%

differences exceed 2

- 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



Reads (Showin

00 bp

PS 04 84 44 05	position &	n and CIGAR
466 alignments of 1,736, excluded 1,040	of low quality)	
		2 889 hp
158,617,000 bp	158,617,500 bp	158,618,000 t
		TraesCS2B02G183300

RNA-Seq Read Mapping Visualization

lotes and Marks	Sharing	
n positions from IWG ome bases erage and highlight v 0% ases with quality und (perfect matches), re ates at the same alig	SC (HC) when der 99% equiring	 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



ReadS (Showing 466 alignments of 1,736, excluded 1,040 of low quality)





RNA-Seq Read Mapping Visualization

otes and Marks	Sharing	
positions from IWGS ome bases IWGSC erage and IWGSC	SC (HC) Show individ C High Confidence Genes (1.1) C Low Confidence Genes (1.1)	hed bases
ases with quality und (perfect matches), re	equiring	d insertions d and reverse reads differently
tes at the same aligr	nment	



RNA-Seq Hexaploid Wheat: Expression Analysis

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

Bread Wheat: Expression Analysis

 Analyze feature expression utilizing the Chinese Spring Wheat reference assembly

Dashboard / All Projects / Project Details				
Project Details	Start Multiple Expression A	Analyses		
General Informatio	Genome Feature Set: 0	IWGSC - High Co		
Project Nan	Feature to Count: 🕑	Gene		
Speci	Exon Matching: 🕑	 Count only reads Count reads that of 		
General Not	es:	Allow reads to counter to counter the second sec		
	De-duplication: 🕑	Remove potential		
	Minimum Quality: 🕑			
All Files Raw Sequence Raw Raw Sequence Raw Sequence Raw	uences	_		
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	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	Lul View Alignments -
198,975,170	2019-12-16 13:05	Lill 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	Lul View Alignments 👻
203,941,764	2019-12-14 08:25	Ltd 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	t Details				
Proj€	ect Details	Start Multiple Expression A	nalyses			
Gen	eral Information	Genome Feature Set: 😧	IWGSC - High Co			
	Project Name:	Feature to Count: 😧	Community Genome			
	Creation	Exon Matching: 😧	IWGSC - Low Con			
	General Notes:	Multi Feature Overlap: 🚱	 Allow reads to course 			
		De-duplication: 😧	Remove potential			
		Minimum Quality: 🕑				
All F	iles Raw Sequences					
•	Available Actions -	Delete Files				
	File Name					
	SRR1542407.1 Paired (A	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 (A	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	Lul View Alignments 👻
198,975,170	2019-12-16 13:05	Lul 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	Lill 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

Dashbo	Dashboard / All Projects / Project Details						
Proje	ect Details	Start Multiple Expression A	nalyses		×	More Options -	
Genome Feature Set: IWGSC - High Co			IWGSC - High Confidence	High Confidence Genes - Chinese Spring Wheat (v1.1)			
	Drainet Name:	Feature to Count: 😧	Gene	•			
Project Name: Exon Matching: • Gene Transcript Exon Species: • Count reads that		Gene Transcript Exon Count reads that overlap	e script n nt reads that overlap introns or exons of the gene				
	General Notes:	Multi Feature Overlap: 🕑	 Allow reads to count toward 	ards all features they overl	ap with		
		De-duplication: 😧	Remove potential read dependencies	uplicates at the same posi	tion and CIGAR		
		Minimum Quality: 😧			Phred: 20 (99%)		
All F	iles Raw Sequences Available Actions -	Delete Files	C	ancel Start Expres	ssion Analysis		
	File Name			Reads Processed	▼ Last Updated		
Ø	SRR1542407.1 Paired (A	Aligned 2019-12-13 21:40 UTC)		193,694,736	2019-12-16 17:27	Lul View Alignments 👻	
	SRR1542404.1 Paired (A	Aligned 2019-12-13 21:40 UTC)		198,975,170	2019-12-16 13:05	Lul View Alignments 👻	
	SRR1542405.1 Paired (A	Aligned 2019-12-13 21:40 UTC)		184,418,714	2019-12-16 11:02	Lud View Alignments 👻	
	SRR1542406.1 Paired (A	Aligned 2019-12-13 21:40 UTC)		184,129,219	2019-12-16 07:01	Luu View Alignments 👻	
	SRR1542408.1 Paired (A	Aligned 2019-12-13 21:40 UTC)		165,312,410	2019-12-16 02:07	Luu View Alignments 👻	
	SRR1542409.1 Paired (A	Aligned 2019-12-12 19:53 UTC)		203,941,764	2019-12-14 08:25	Lill View Alignments 👻	

Reads Processed	▼ Last Updated	
193,694,736	2019-12-16 17:27	Lud View Alignments 👻
198,975,170	2019-12-16 13:05	Lint View Alignments 👻
184,418,714	2019-12-16 11:02	Lint View Alignments 👻
184,129,219	2019-12-16 07:01	Lint View Alignments 👻
165,312,410	2019-12-16 02:07	Lind View Alignments 👻
203,941,764	2019-12-14 08:25	🔟 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	Start a Differential Expression Analysis			
Gen	eral Information		Analysis Name:	Effects of drought a	nd heat s	
				E.g. 'Expression chan	ige in radi	
	Project Name:	Expe	eriment Design	Advanced Options		
	Species:	Defin	e Sample Groups:			
	General Notes:		▲ Set Group		File Nar	
			Group 1	Y	SRR15	
			Group 1	T	SRR15	
			Group 2	T	SRR15	
All F	All Files Raw Sequences		Group 2	T	SRR15	
Тур	be here		Group 3	T	SRR15	
	▲ File Name		Group 3	¥	SRR15	
	SRR1542404 - CK		Group 4	T	SRR15	
	SRR1542405 - CK		Group 4	T	SRR15	
	SRR1542406 - DS1h					
	SRR1542407 - DS1h					
	SRR1542408 - DS6h				Canc	
	SDD1542400 DS65			Expression 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

Dashbo	oard / All Projects / Projec	Start a Differential B	Expression Analysis			×		
Proje	ect Details	Analysis Name:	Effects of drought and h	neat stress on bread wheat		?		More Options -
			E.g. 'Expression change in	n radiation exposed tissue'				
Ger	neral information	Experiment Design	Advanced Options					
	Project Name:	DESeq2 Options:						
	Species:							
	General Notes:	Μ	linimum Expression Count:	1	?			
		Required Samples	with Minimum Expression:	1	?			
		Independent Filte	ering Significance Cutoff: 🛿	0.1				
All F	iles Raw Sequences	Count ⁻	Transformation Blinding: 0	Use experiment design information w calculating transformed counts	/hen			
Тур	pe here			 Blind the transformation counts to the experimental design (i.e. unsupervise 	e ed)			
	File Name							
	SRR1542404 - CK	edgeR Options:					16 🔟 Vie	ew Analysis Results
	SRR1542405 - CK	Minimum	Counts Per Million (CPM):	1.0	?		54 Lui Vie	ew Analysis Results
	SRR1542406 - DS1h	Required Sa	mples with Minimum CPM:	1	?		09 🔟 Vie	ew Analysis Results
	SRR1542407 - DS1h	Differe	ential Testing Procedure: 0	Use GLM likelihood ratio test			12 🔟 Vie	ew Analysis Results
	SRR1542408 - DS6h		-	Use GLM quasi-likelihood (QL) F-tes	t		56 🔟 Vie	ew Analysis Results

Bread Wheat: Differential Expression

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





Display/Filter Options

Notes Sharino

- 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





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







DNA-Seq: Tetraploid Wheat

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

Tetraploid Wheat: Read Mapping

- Also incorporates the Svevo • (*Triticum turgidum*) reference assembly from the IDWGSC (i.e. the International Durum Wheat Genome Sequencing Consortium)
- Instantly browse any portion • of the IDWGSC reference sequence and visualize read alignments of samples of any size
- Dynamically leverage gene/transcript/exon annotations for Svevo from Ensembl



Display Options	N
Show gene and	exon
Show reference	geno
Show nucleotide	cove
differences exce	ed 20
Exclude reads a	nd ba
Group reads by	UMI (

position and CIGAR

consensus of 60%

Reads (Showing 120	alignments, excluded 398 of low quality, rem	noved 3 duplicates)	
<			- 54
572,000 bp	721,572,100 bp	721,572,200 bp	





Tetraploid Wheat: Variant Analysis

- Call and analyze variants • leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and • visualize variant density across multiple samples
- Dynamically filter to different • genomic region types based on annotations published by the IDWGSC via Ensembl



Variant Analysis Tets Paired - vs Svevo - varcalls phred 20 (Analyzed 2020-01-01 21:33 UTC)



	Compare	Notes and Marks	Sharing
and exon annotations from			Show only variants not present in the control (i.e. somatic)
within known coding regions with predicted consequence of		ng regions 😧 🛛 🗷 sequence of	Show only variants with control coverage of at least 1 reads
)S(e alt coverage is		Match homozygous SNVs to related heterozygous SNVs in control 🚱
nc	bzygous or hetero	e ozygous	Show all variants found within selected sample files Show only variants common to all selected sample files
n r d (are alleles on the variant's g	eneral type	Show file coverage track of variants detected per position
ed	l position the vari	ant's	

Tetraploid Wheat: Variant Analysis

- Call and analyze variants • leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and • visualize variant density across multiple samples
- Dynamically filter to different • genomic region types based on annotations published by the IDWGSC via Ensembl
- Access predicted biological • consequences based on calculated amino acid shifts



Variant Analysis Tets Paired - vs Svevo - varcalls phred 20 (Analyzed 2020-01-01 21:33 UTC)



Tetraploid Wheat: Variant Analysis

- Call and analyze variants • leveraging the Svevo 1.0 reference assembly
- Simultaneously compare and • visualize variant density across multiple samples
- Dynamically filter to different genomic region types based on annotations published by the IDWGSC via Ensembl
- Access predicted biological consequences based on calculated amino acid shifts
- Filter down to consequences • of interest



Variant Analysis Tets Paired - vs Svevo - varcalls phred 20 (Analyzed 2020-01-01 21:33 UTC)





In Conclusion

Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar

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 wheat and related species, along with 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) •
- International Durum Wheat Genome Sequencing Consortium (IDWGSC)

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.



Extending the Curio Genomics Platform for the Wheat Research Community / IWGSC Webinar